ISSN 1913-8989

# COMPUTER AND INFORMATION SCIENCE

Vol. 1, No. 1
February 2008

CCSE

Canadian Center of Science and Education

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 1, No. 1

February 2008

# Contents

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# A Model for Voice-activated Expression Editing

David D. Langan
School Computer & Information Sciences, University of South Alabama
307 University Blvd N, Mobile, AL 36688, USA
Tel: 1-251-460-6390 E-mail: david.langan@gmail.com
Thomas F. Hain
School Computer & Information Sciences, University of South Alabama
307 University Blvd N, Mobile, AL 36688, USA
Tel: 1-251-460-6390 E-mail: thain@usouthal.edu
W. Christopher Camery
Lockheed Martin Simulation, Training, and Support,
12506 Lake Underhill Rd., Orlando, FL 32825, USA
E-mail: chris.camery@gmail.com

**Abstract**

Code entry and editing by manually disabled computer programmers is difficult in conventional GUI-based development environments, since these rely heavily on mouse and keyboard use. The advent of accurate and responsive speech recognition technologies has made the speech user interface (SUI) a viable option for input. While high level programming constructs can be entered in a fairly natural way us/ing voice, expressions may be very complex in structure, and may use a wide variety of variables, literals, punctuation marks and operator symbols. A SUI-based syntax-directed editor was previously presented, but its interface model did not include expression editing in its model and implementation. This paper fills this gap by presenting a SUI-based model for entering and editing programming language expressions. A Java implementation was evaluated by a group of programmers to verify the functionality and to test the ease of use of the model. The results of this evaluation are presented and discussed.

**Keywords:** Speech user interface, Expression editor, Integrated development environment, Accessibility

## 1. Introduction

For programmers, entering and editing code in a development environment involves the manual operation of input devices such as the keyboard and mouse. This provides a daunting, if not impossible challenge for people with manual disabilities ranging from repetitive strain injuries (RSI) to more severe disabilities. A 2006 report from the Bureau of Labor Statistics (USDL, 2007). reported that those who suffered from carpal syndrome experienced a median of 27 days of missed work. These statistics provide evidence that programmers other than just the manually handicapped could profit from a form of input other than the mouse and keyboard. The relatively recent advent of accurate and responsive speech recognition technologies has made speech user interfaces (SUIs) a viable option for program entry. A SUI-based development environment offers accessibility to programming for manually disabled programmers. There is also the possibility that if programming by voice is sufficiently intuitive and fast, users who are not disabled may at times choose this option if the programming environment is multi-modal. What is needed for SUIs is a "standard" for their use such as now exists for GUIs.

### *1.1 Editing Code by Voice*

The term *expression* in the context of expression editing is used to mean a collection of variables, function calls, and operators that are combined in a logical manner according to a language specification. Expressions are syntactic elements of programming languages found, for example, as the predicate of *if*, *while* or *for* statements, as right-hand sides of assignment statements, or as function parameters. Expressions, more than most other constructs of programming languages, can be quite long and intricate.

Here we will discuss three existing models for voice-driven programming: Voice Code (NRCC, 2005c), Happy Hands (Hennessey, 2005b), and Voice-Activated Syntax-Directed Editor (VASDE) (Hubble, 2006 and Hubble 2007). Voice

Code and Happy Hands integrate expression editing into the overall coding framework, using a dictation mode. As will be shown, there are distinct inadequacies in this approach. VASDE takes the view that expression editing is inherently different from producing program code, being at a much higher level of abstraction, and defers expression editing altogether (provisionally using GUI input in the interim). This paper presents a model to fill this gap.

1.1.1 Voice Code

While entering text by voice may work well for dictating natural language documents, the Voice Code documentation (NRCC, 2005a) contains a very good example of why dictating code is so difficult, and uses an expression to demonstrate this point: "Programming languages were never meant to be spoken. Consequently, it is very difficult to dictate code.

For example, to dictate the following C++ code:

```
if (currRecNum < maxOffSet)
{
    ^
}
```

the programmer might have to say something like this:

*if open-paren Charlie uniform romeo romeo cap romeo echo Charlie cap-November uniform mike less-than maxbegin-capitalize begin-no-space off set end-capitalize end-no-space close-paren new-line open-brace new-line new-line close-brace up-arrow tab-key*" (NRCC, 2005a).

The Voice Code team decided on a more natural-language approach to dictating the above expression, in this example they proposed something like "*if current record number is less than max offset do the following"* (NRCC, 2005a). The Voice Code project treats expressions in a very general way. The only support for expressions in Voice Code are simple tools for inserting and navigating general balanced expressions. Commands for inserting expressions include the following:

*between <exprName>*

*<exprName> pair*

*empty <exprName>* (NRCC, 2005b)

The semantics of these commands are explained as follows:

The *"between <exprName>"* and *"<exprName> pair"* forms are used to type the balanced expression and put the cursor inside the expression. The *"empty <exprName>"* form types the balanced expression but puts the cursor right after it. (NRCC, 2005b).

For navigating, there are commands such as "*jump out*" and "*jump back out*" which place the cursor just outside of the expression currently being edited. As Voice Code is designed to deal with any programming language, it does not include any form of code assistance (such as Microsoft's Intellisense) for any particular language. Although Voice Code currently supports only Python as of November 2006, support for other languages, including C++, Java, and Perl is planned to be added in the future. Voice Code does not replace the GUI paradigm, but rather augments it. Voice Code is built as an extension to the text editor Emacs, and thus is merely a speech layer built atop a traditional GUI (NRCC, 2005c).

1.1.2 Happy Hands

Sean Hennessey created the Happy Hands Java Speech Editor (Hennessey, 2005b). This editor deals only with Java as the editing language, and therefore has Java-specific features. Happy Hands is of interest due to its similarities with Hubbell's VASDE, both of which were developed concurrently. The similarity of Happy Hands to VASDE, makes it a relevant starting point to examine in the design of an expression editor (Hennessey, 2005b).

The expression editor of Happy Hands, called the "expression transcriber", allows the creation of expressions by reading, from left to right, the elements that make up the expression. It is also possible to specify that an operator will be to the left of the current variable. However, Happy Hands does not provide a way to remove a part of an expression. The Happy Hands documentation itself admits "Happy Hands does not have a good way to remove elements from expressions, other than using the backspace key. Since the expression is quickly re-parsed and active transcribers are updated, this is perfectly valid the result is no different." (sic) (Hennessey, 2005a).

While in Happy Hands, the user must use the mouse or keyboard to select the element he or she wishes to edit, in VASDE the selection of an existing expression to edit is done entirely by voice.

According to the Happy Hands documentation, expressions are implemented as "for the most part binary trees, with some lists too" (Hennessey, 2005a). Whenever an expression element is being edited the existing element is overwritten

unless an operator is spoken, in which case new nodes will be added to the binary tree, one for each side of the binary operator.

While both Happy Hands and Voice Code make it easy to enter expressions by voice, editing them is entirely too difficult.

1.1.3 VASDE

Hubbell (Hubble, 2006 and Hubble 2007) presented a model and implementation called VASDE (Voice Activated Syntax-Directed Editor) to create and edit Java programs. He takes the position that entering code provides an opportunity for semi-automation by the fact that computer languages have a rigid syntax for high level constructs. The syntax-directed approach used by Hubbell uses a different paradigm than conventional text editing. Instead of typing a code block or statement, there are separate editors for each type of statement or block. For example, to create a for-loop, a dialog will appear and the user is required to enter the initial condition, the condition for continuation, and post-loop action into text fields. The code for the loop is then generated. Using this paradigm, programmers create and edit structures and statements rather than editing text directly.

The tests and evaluation of the VASDE editor yielded favorable results (Hubble, 2006 and Hubble 2007), and showed the viability of both the syntax-driven approach, as well as the use of voice for the task of programming. However, the low level task of expression creation and editing was bypassed in the initial research. It was felt that expressions were too "small" and complicated to be naturally supported by the language syntax driven approach that had been appropriate for higher level constructs.

## 2. Proposed Model

Expressions, across a spectrum of modern high level programming languages, tend to be more alike than different. The Voice Expression Editor (VEE) model proposed here is based on Java syntax, but can be easily ported to other high-level object-oriented programming languages. The Java Language Specification (JLS) (Gosling, 2000) states that expressions are composed of syntactic elements, including variables, operators, casts, literals, class names, method invocations, and subexpressions.

The VEE-model proposed here enforces the restrictions imposed by the JLS (e.g., two binary operators may not be adjacent). However, the editing process does not structurally conform to the formal JLS definition of an expression wherein each node in the syntax tree is an individual syntactic element (Shavor, 2003). For example, the simple expression `3*4` consists of a three-node tree with the root being `*`, and its children being `3` and `4`. While this syntax description lends itself nicely to the parsing and code generation, it proves to be awkward as the basis for by-voice expression editing, leading to unnecessarily long navigation paths. VEE "flattens" this syntax tree, creating a subexpression only when a term is surrounded by parentheses, or is an element of a list (e.g., an argument to a method having multiple arguments). VEE would view the expression, `3*4` as a single node containing three elements. The individual elements within a node are contained within a "slot." In the previous example, the expression would have 3 slots. Elements are dynamically allocated to the slots from left-to-right.

Expression editing involves several identifiable sub-tasks, namely (a) navigation, (b) selection/insertion/deletion, and (c) data entry. The sections that follow will discuss each of these in detail.

### *2.1 The Navigation Process*

When an expression is being edited by voice, it is appropriate to add visual clues to facilitate the navigation process, such as the automatic numbering of all slots, and the highlighting of slots that are currently selected. A special expression-editor—distinct from the program-editor—is invoked for this purpose. The two editors communicate in the sense that the program-editor, when required, invokes the expression-editor (potentially pre-loading an existing expression), and the expression-editor when terminated returns control to the program-editor, passing it the created or edited expression.

The expression-editor will recursively invoke itself when a subexpression (i.e., a subnode in the VEE syntax tree) needs to be edited, initially loaded with that subexpression. In each case, a command-word such as *done* or *OK* is used to terminate an instance of the expression-editor to return to the calling context. This relationship may be seen in the left side of the Figure 1. The role of the clipboard is explained below.

### *2.2 Selection/Insertion/Deletion*

Expressions are composed of elements, each of which is dynamically (from left-to-right) allocated to a slot. A slot may contain one of the following element types:

a keyword (e.g., `this, true, false, new`)

an operator,

a literal,

an identifier (variable, or type),

an expression within parentheses (other than a parameter list), and

an element of a list (elements separated by '**,**' or '**;**').

The last two element types represent "subexpressions".

We first discuss the selection, insertion, and deletion of expression elements in general terms, and then discuss the insertion of each of the specific element types.

2.2.1 Selection

With a mouse or keyboard, one can click on, or arrow to, any desired element within an expression prior to performing some desired operation. In the VEE-model, we allow direct access to elements within an expression by the use of the automatically generated and displayed slot numbers. The numbers are ordered from left to right.

In the expression **`(y+1)+3*add(2*x,8)`**, there are 7 slots (numbered 1 to 7), containing **`y+1`**, **`+`**, **`3`**, **`*`**, **`add`**, **`2*x`**, and **`8`**. The displayed slot numbers enable selection of a slot by saying *select* and then the slot number (e.g., *select 4* to select the slot containing **`*`**), or a range of slots by saying *select A to B* (e.g., *select 2 to 3*).

In this model, the editing process is hierarchical in the sense that selecting a subexpression leads to invoking a new instance of the editor. In the above example, the slot **`y+1`** would be edited within its own editor. One would say *select 1* to select the slot, and then say *edit* to recursively enter a new instance of the expression-editor (subexpression-editor) to work on just **`y+1`**.

2.2.2 Insertion

All expression elements are inserted immediately after the currently selected slot. Slot zero is always present at the beginning of any expression (even an empty one), although it has no visible content. Therefore, to add something to the beginning of an expression, slot zero should be selected.

If the expression resulting from an insertion would be invalid, a blank slot will be automatically inserted. For example, if the expression being edited is **`2+4`** (with 3 slots), and the slot containing **`2`** is selected, and the user inserts a **`-`** operator (by saying *minus*), then the resulting expression will be **`2-_+4`** (having 5 slots) because two binary operators cannot appear adjacent to one another. Here, the expression editor creates blank numbered slots containing a placeholder, _.

2.2.3 Deletion

To delete a slot, or a range of slots, the user would say *select* to pick the desired slot(s) and then say *delete*. The slots are immediately removed from the expression and are placed in the recent-clipboard (discussed below).

If a deletion would create an invalid expression, blank slots (each containing a placeholder, _) will be inserted as needed. For example, if the user were to delete the **`+`** operator from **`1+2`**, the resulting expression would be **`1_2`**, with slot 1 as the selected slot.

*2.3 Insertion/Editing of Specific Element Types*

2.3.1 Insertion of keywords and operators

The insertion of keywords (after the selected slot) is accomplished by simply saying the keyword, e.g., *null*, *false*, *true*, *this*. The insertion of both unary and binary, logical and numeric operators are similarly invoked by words which are recognized by the speech engine and accepted by the editor. For example, to insert **`+`**, one would say *plus*. A sampling of other recognized words are *minus*, *preincrement*, *postdecrement*, *times* (or *multiply*), *divide, logical-and, binary-and*, *mod* (or *modulus*).

The expression **`1+2*3`** can be entered by saying *one plus two times three*, pausing between each word until the "recognizer busy" indicator (discussed below) disappears.

2.3.2 Literals

Numeric literals are inserted by speaking the number that is to appear in the expression. For integers the actual number must be spoken. For example, if the user wanted to enter the number 136751, then they would say *one hundred thirty six thousand seven hundred fifty one*. In the case of a float like 3.14 the user would say *three point one four*. Character and String constants can be entered by saying *character* or *string* to enter a special entry mode. In the case of *string*, the user enters a dictation mode that he exists by saying *done*.

2.3.3 Subexpressions

To create a subexpression, the user can select a slot or range of slots and say *subexpression* or *parentheses*. The selected slot or slots will be converted into a single subexpression slot, surrounded by parentheses. If the current expression is **`3+2*x`** (having 5 slots), and the user wanted **`3+2`** to be a subexpression, then the user would say *select one through*

*three* to select the appropriate slots, and then say *subexpression*. The resulting expression would be **(3+2)*x,** (consisting of three slots).

When a slot containing a subexpression is selected, a new invocation of the expression-editor is generated when the user says *edit*. This editor is preloaded with the subexpression (minus the parentheses, if it is a parenthesized subexpression), and can be edited as an independent unit (with its own set of slots.) Once the *OK* command is given, this (sub)expression-editor is terminated, and in the invoking editor, the original subexpression is replaced by the edited subexpression. This process can be repeated recursively for expressions with nested parentheses, as shown in Figure.

*2.4 Insertion of Identifiers*

This topic is sufficiently complex to warrant a separate section on its own. While numeric literals and binary operators can be entered by speaking the name of the number or the operator, identifiers cannot be dictated in this manner. Variables, methods, classes, and packages are not guaranteed to be English words, or even to be pronounceable. We make use of the fact that the VASDE program-editor requires identifiers to be declared before use, and that these identifiers, within a given scope, can therefore be known by the expression-editor. Identifier entry into expressions is done by selection rather than dictation. The challenge, therefore, was to find a method whereby identifiers are presented to the programmer in a navigable manner, so that the choices at any given time are not overwhelming. We call the mechanism we use a "clipboard" window.

2.4.1 The Clipboard

The clipboard window is used to enter text into expressions. If the text entries are package names, class names, method names, or variable names, a spoken command (e.g., *package*, *imports*, *class*, *method*, *variable*) is issued to generate and display a context-sensitive numbered list of applicable entries—reminiscent of "intellisense". Loading the clipboard can be facilitated by the interaction of VEE with the program-editor to determine the in-scope items. The item to be entered into the expression is selected and inserted by saying the associated number. This approach minimizes required vocal bandwidth, and the possibility of speech misrecognition. While this obviates the need for mouse and keyboard (a necessary goal), it incidentally provides a memory aid for the programmer. Also, by comparison, this method of choice avoids the possibility of misspelled names (a common problem with the keyboard).

The clipboard also forms the by-voice replacement of the ubiquitous cut-and-paste clipboard found on editors utilizing mouse/keyboard interaction. If the spoken command is *recent*, a list of previously selected fragments (individual slots, or slot ranges) that were deleted or copied from an expression are listed, and can be pasted into the current expression.

We can think of the clipboard as having different versions, depending on the invoking command, i.e., the type of entry to be inserted into the expression. Some versions may invoke additional versions, for example selecting a method may then require selecting a class, etc. The clipboard intelligently navigates such hierarchical structures, as will be shown below. Figure 2 shows a more detailed view of how the clipboard would work internally and relative to the expression (subexpression) editor for access to names.

Other available clipboard navigation commands include *back* and *cancel*. The *back* command does what one would assume: it takes the user back one step, effectively undoing the previous command. The *cancel* command immediately closes the clipboard and returns control to the expression editor without modifying the expression.

We now give details of the clipboard versions, and their interactions. Initially we describe the selection process for each version, and then talk about insertion into expressions.

2.4.2 Variables

If the user wants to enter a (a) local variable, (b) instance variable (possibly inherited), (c) class-variable (possibly inherited), or (d) particular data-field of an object, one would enter the clipboard by saying *variable*. The clipboard is now titled "Variable Clipboard," and is put into a state of "Accepting Variables." Because variables can be instances of a class, the user must be able to select not only the variable itself, but any available fields of that variable. If a variable contains accessible fields, that variable will appear in the clipboard in two forms: the variable name and the variable name followed by a dot (e.g., **x** and **x.**). If we assume that **x** is a local variable that has an available data member **y**, the user would see a menu that offered both **x** and **x.**. If he now selects **x** from the menu and says *OK*, he would return to the expression-editor with an **x**. On the other hand, if he selects **x.** he would see a new menu of available variables that would include **y** which he could then select to return **x.y** to the expression editor. This navigation applies recursively to fields of fields, etc.

2.4.3 Methods:

If the user wants to insert a method call, the user would enter the clipboard by saying *method*. The clipboard is now titled "Method Clipboard," and is put into a state of "Accepting Methods." The clipboard would display a numbered list of all methods (possibly inherited), including overloads, available in the current context, i.e., the current class. For

example, if the enclosing class has two (possibly inherited, but visible) methods with signature **thePerson(INT)**, and **thePerson(INT,INT)**, both will appear in the clipboard numbered list.

If the user wants a method from a different class, or one that can be invoked by a variable, they would say the appropriate command (*class* or *variable*). This would load the clipboard with identifiers associated with that class or variable. Thus, the clipboard can at different times offer available package names, class names, variable/field names or method invocations (with an appropriate formal parameter list). When a clipboard is titled as "Method Clipboard," what is meant is not necessarily "a clipboard that is currently displaying methods," but rather "a clipboard whose task is to allow the user to select a method."

An example may help to illustrate the creation of a method call. If the user wants to enter the method call **x.f(2)**, where **x** is a local variable, the user would say *method* to invoke the method clipboard. Upon entry, the clipboard would offer the instance and class methods appropriate to the calling context (this might not include **f**). The desired method is an invocation by a local variable, so the user next says *variable* to view a list of variables (which would include the local variables.) The user would then be offered **x**, and after he selects it and says *OK*, he would now see a list of methods available to **x**. This list would include **f(INT)** and possibly other versions of **f** (if it had been overloaded.) After selecting **f(INT)** and saying *OK*, the programmer would be returned to the expression-editor with `x.f(INT)` inserted in the appropriate slot.

Sometimes, the desired method will require a fully qualified name that includes the package name. Hence one may say *package* to select a package name from among those available. While packages in Java cannot contain other packages, they can have names that give the appearance of nested packages (e.g., **javax.swing** and **javax.sound** may appear to both be contained in the **javax** package). Packages with such names may have very long names, and there are numerous packages in the Java standard library. To make them more manageable, and to avoid showing several hundred packages to the user at once, the clipboard treats packages as if they were nested. For example, to select the package **javax.swing**, the user will first select **javax.** and then select **swing**.

The net effect of the clipboard nesting variables, classes, and packages, is that a user may select a particular item based on its fully-qualified name. For example, to select the method **javax.swing.JOptionPane.showInputDialog(OBJECT)**, the user will first say *method* to invoke the method clipboard. The user will then say *package* to view a list of packages. He will select the partial package name **javax.** , say *OK*, and then select the package itself **swing,** and say *OK* again. A list of classes belonging to the javax.swing package will now be displayed. The user will then select the class `JOptionPane`, whereupon its methods will be displayed, and the user will select the method **showInputDialog(OBJECT)**. The entire command sequence required to enter the fully qualified method in the example would be something like : *method*, *package*, *8, OK, 4, OK, 90, OK, 21*" (where *8* is the menu-number corresponding to **javax.**, *4* is the number corresponding to **swing**, *90* refers to **JOptionPane** and *21* refers to **showInputDialog(OBJECT)**).

To prevent the user from having to select each part of a package name every time it is to be used, the clipboard keeps track of packages used and adds them to an "imports" clipboard, which can be accessed by saying *imports*. While this only saves one selection in the case of **javax.swing**, the user might be spared several selections in the case of the longer names.

2.4.4 Classes

For subexpressions used in casts, a type (class) is required, and the user would enter the clipboard by saying *class*. The clipboard is put into a state of "Accepting Classes". The clipboard will list both inner classes, and their dotted versions (if necessary). The behavior after selecting an item is much like that of variables and methods. The mechanism for full qualification of classes is like that described for methods.

**3. VEE an Implementation**

The VASDE project focused on programming in "the-large" and did not provide a mechanism for the voice input of expressions. The VEE model was proposed to rectify this missing component. The VEE implementation (VASDE Expression Editor) was written in Java as an Eclipse (Budinsky, 2004) plug-in so that it would function with VASDE. VASDE was modified to use VEE anytime a user needed to enter or edit an expression, and VEE when it terminated would return the expression to VASDE.

The VEE implementation followed the model presented in the previous section. The expression-editor was implemented using a dialog as shown in Figure. Here one sees the stack nature of the expression editor. In this case, the user entered the editor with the expression **10+(x+(18-4))*2** (which has 5 slots). The user then selected slot 3, said *edit*, and entered the editing of the subexpression **x+(18-4)** (which has 3 slots). The user then selected slot 3, **18-4** (also having 3 slots,) and requested an *edit*. Finally he selected slot 3 of this subexpression, where he could use *delete* to remove the **4** and then use some other command (e.g., *five*) to insert a new element, **5**.

The upper portion of the expression editor shows the stack, while beneath that, the current (sub)expression is shown with its slot numbers, and the selected slot(s) being highlighted. After entering/editing a subexpression one may say *OK* to return to the previous subexpression in the stack, with changes being reflected in that expression. Finally, the last *OK* would return the programmer to the program-editor, with the final edited expression.

The VEE clipboard is called upon from the expression editor and is used to enter variables, methods, class names, package names, and recently used text. That clipboard shows a numbered collection of items that may be selected as described in the previous section (see Figure 4).

## 4. Evaluation

Five people with experience in C++ or Java were selected to test and evaluate the effectiveness of VEE as a voice-driven expression editor (Camery, 2006). The total evaluation took about three hours. The first phase of the evaluation required the evaluators to use the voice-training features of the voice-recognition engine (in our case IBM Via-Voice version 9) to train it to recognize their speech patterns. In the second phase, the evaluators were asked to dictate and edit given expressions into the SpeechPad utility program that was supplied with IBM Via-Voice. It was not required that the users complete this task, but only that they understood what would be involved. The purpose of this was to have them simply experiment with the expression editing task based only on dictation. In phase 3, the evaluators were exposed to a guided tutorial on the use of VASDE and all of the major features of VEE. During this time they were free to ask questions. After the tutorial was completed they entered phase 4, wherein they were asked to perform a series of VEE tasks on their own. The evaluation task was designed to use all features of VEE, and to allow the evaluator to see what it would be like to use the editor to create a program. In this phase, the evaluator was free to ask for help using VASDE (since they were not asked to evaluate VASDE itself,) but they were asked to complete the VEE task without assistance. This task took about an hour. Finally, in phase 5, the evaluators were asked to fill out an evaluation questionnaire regarding their thoughts on, and experiences with, VEE. The questionnaire included five 5-point Likert-scale questions regarding the evaluator's opinion of various aspects of the editor. At the end of the questionnaire, the evaluators were invited to make any additional comments regarding various issues with the model or its implementation.

The questions elicited the evaluator's overall opinion of the editor, the effectiveness of the use of slots, the effectiveness of the editor compared with alternative methods of voice input, and the degree to which the recognition problems of the underlying speech engine affected the evaluator's opinion of VEE.

The results of the 5-point Likert-scale (wherein 1 represents the least support for the statement, and 5 represents the strongest degree of agreement with the statement) questions are given in Table 1.

### *4.1 Evaluator Response to Ease of Use (Q1/Q5)*

While the overall ease of use of the editor was rated poorly, every evaluator responded that the poor performance of the speech engine limited the effectiveness of the editor, and led to frustrations while using VEE. The most common problem was that the speech engine misrecognized words. This was not found to be a problem for most of the main screen's functions (with the exception of entering numeric literals). The clipboard screen was found to have the most misrecognition errors, but these were again related to numbers. The recognizer would too often misrecognize a number being spoken (e.g. the number 190 being recognized as 119). This clearly frustrated several evaluators. One evaluator commented that the speech engine was "a definite weak point in the process".

The rate at which some evaluators initially attempted to enter data in the main editor window was too fast for the speech engine to recognize. Once the evaluators began to speak more slowly, there were far fewer misrecognition errors. The red "recognizer busy" indicator was of great help to these people.

One evaluator stated that the ability to say the names of methods and variables to select them from the clipboard would be nice, as opposed to selecting them by number. Another evaluator expressed the desire to be able to declare a new variable while inside of the expression editor.

It should be noted that these tests were conducted using the IBM Via-Voice (version 9). Subsequent to these tests, the latest version of Dragon Naturally Speaking was tested for voice input and was found to be both faster and considerably more accurate. In particular it handled the input of numbers far better than IBM Via-Voice.

### *4.2 VEE Compared to Simple Dictation (Q2/Q4)*

Every evaluator stated that they would rather use VEE than use a traditional text editor with a set of voice macros. Several evaluators noted that while entering an expression in VEE may take longer, editing expressions was fairly painless.

However, no evaluator seemed particularly enthusiastic about using voice to enter expressions, and in fact, question 4 received the lowest score of all the questions. The explanation for this, however, may lie in the universal opinion that the underlying voice engine used was largely at fault (question 5).

The responses from all evaluators indicate that while keyboard input is preferable to voice input for entering and manipulating expressions, if they *had* to use voice input, they would want to use something like VEE. This observation tends to support the main goal of this project, namely to provide a good voice-activated editor for the manually disabled. None of the five evaluators was manually disabled. It is possible that a programmer with such a disability would have an entirely different perspective.

*4.3 Evaluator Response to the Use of Slots (Q3)*

It is interesting that, while most evaluators responded fairly negatively regarding the use of voice as an viable means of entering expressions, they rated the effectiveness of the use of slots quite highly. Several evaluators commented that they would like to have the ability to break expressions into slots when using a standard text editor in order to more easily visualize and manage complex expressions.

In the open-ended portion of the evaluation, the evaluators stated that the use of slots is of great benefit to editing expressions by voice. One evaluator noted that it organizes expressions by syntactic elements, which is more appropriate for programmers than a simple text editor that uses only characters and whitespace. Another evaluator commented that "if corrections are necessary, the use of slots makes it easy to go back and edit things individually". It was concluded that the use of slots appears to be an effective method of organizing expressions.

## 5. Results and Conclusions

The results gathered here clearly reflect that, while the VEE implementation has some problems, these problems were based on the use of a poor speech engine, and not on the model itself. Later experiments showed that much better performance was available using the Dragon Naturally Speaking (version 9) speech engine. It is quite likely that the evaluators would have experienced a significantly lower level of frustration had this engine been used.

The VEE model for voice expression-editing is somewhat complex in its approach to the task, but the task itself seems to be inherently difficult to "explain" using words. One possible explanation for the weak evaluation of VEE may be that the evaluation process was too short thus not allowing the programmers to master the paradigm required to use VEE as compared to using a keyboard. It is speculated that with additional practice programmers would become more proficient in its use.

The authors and evaluators found the slot-approach to be fairly intuitive for the task for expression editing. This suggests that, while a voice-activated expression editor based on VEE might not be attractive to the general population of programmers, it might very nicely fit into a voice-activated editor (such as VASDE) for manually disabled programmers for whom the use of a keyboard and/or mouse is problematic.

## References

Budinsky, Frank (2004). *The Eclipse Modeling Framework*, Addison Wesley

Camery, W. Christopher (December 2006). VEE: The VASDE Expression Editor. *Master's Project*, University of South Alabama, Mobile, AL.

Gosling et. al. (2000). *Java Language Specification*, Second Edition. http://java.sun.com/docs/books/jls/second_edition/html/expressions.doc.html. Sun Microsystems.

Hennessey, Sean (2005a). Using the Happy Hands Java Speech Editor. http://www.hdm.com/resources/HappyHands_Java/manual/index.html.

Hennessey, Sean (2005b). The Happy Hands Java Speech Editor. http://www.hdm.com/resources/HappyHands_Java/manual/index.html.

Hubbell, T. J., Langan, D. , and Hain, T. F. (2006). A Voice-activated Syntax-directed Editor for Manually Disabled Programmers, *Eighth International ACM SIGACESS Conference on Computers and Accessibility*, Portland, Oregon, pages 205–212.

Hubbell, Thomas (2005). Voice-Activated Syntax-Directed Editing. *Master's Thesis*, University of South Alabama, Mobile, AL.

NRCC (2005a). Why code dictation is so hard. http://voicecode.iit.nrc.ca/VoiceCode/public/wiki.cgi?Why_code_dictation_is_so_hard. National Research Council of Canada.

NRCC (2005b). Dictating General Balanced Expressions. http://voicecode.iit.nrc.ca/VCode_1_Doc/public/wiki.cgi?obj = DictatingGeneralBalancedExpressions. National Research Council of Canada.

NRCC (2005c). Voice Code. http://voicecode.iit.nrc.ca/VoiceCode/public/ywiki.cgi. National Research Council of Canada.

Shavor, et. al. (2003). *The Java™ Developer's Guide to Eclipse*. Addison Wesley.

USDL (2007). Nonfatal Occupational Injuries and Illnesses Requiring Days Away From Work, 2006. *Bureau of Labor Statistics*. http://www.bls.gov/news.release/osh2.nr0.htm.

Table 1. Table of Evaluation Questionnaire

| | E1 | E2 | E3 | E4 | E5 | Mean |
|---|---|---|---|---|---|---|
| Q1 – Overall VEE was easy to use. | 3 | 2 | 4 | 3 | 2 | 2.8 |
| Q2 – VEE was easy to use compared to dictation. | 5 | 5 | 4 | 4 | 4 | 4.4 |
| Q3 – The use of slots was effective for the given task. | 5 | 4 | 5 | 4 | 4 | 4.4 |
| Q4 – Voice was as an effective means of entering expressions. | 2 | 2 | 3 | 3 | 3 | 2.6 |
| Q5 – Issues with speech engine limited the effectiveness of the VEE editor. | 5 | 5 | 5 | 5 | 5 | 5.0 |

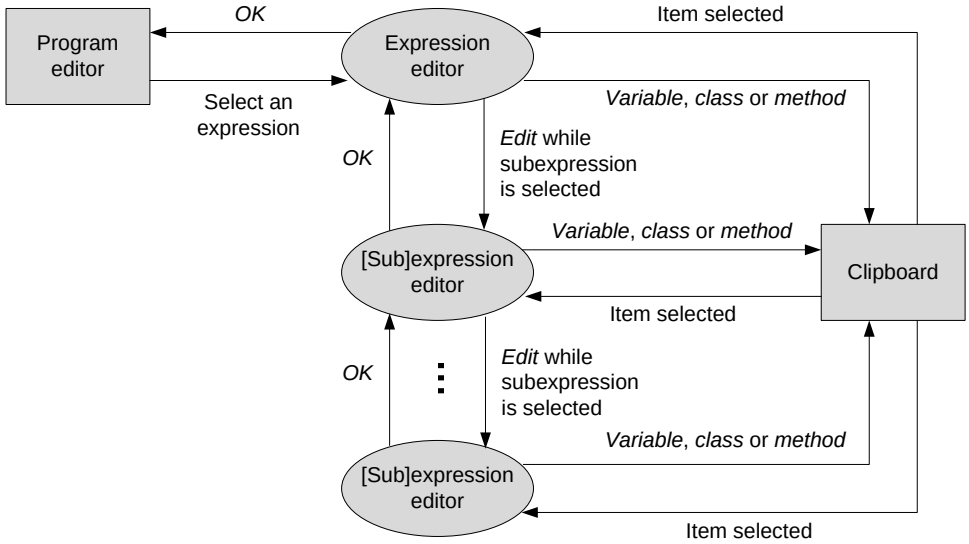



Figure 1. VEE Control Flow Model

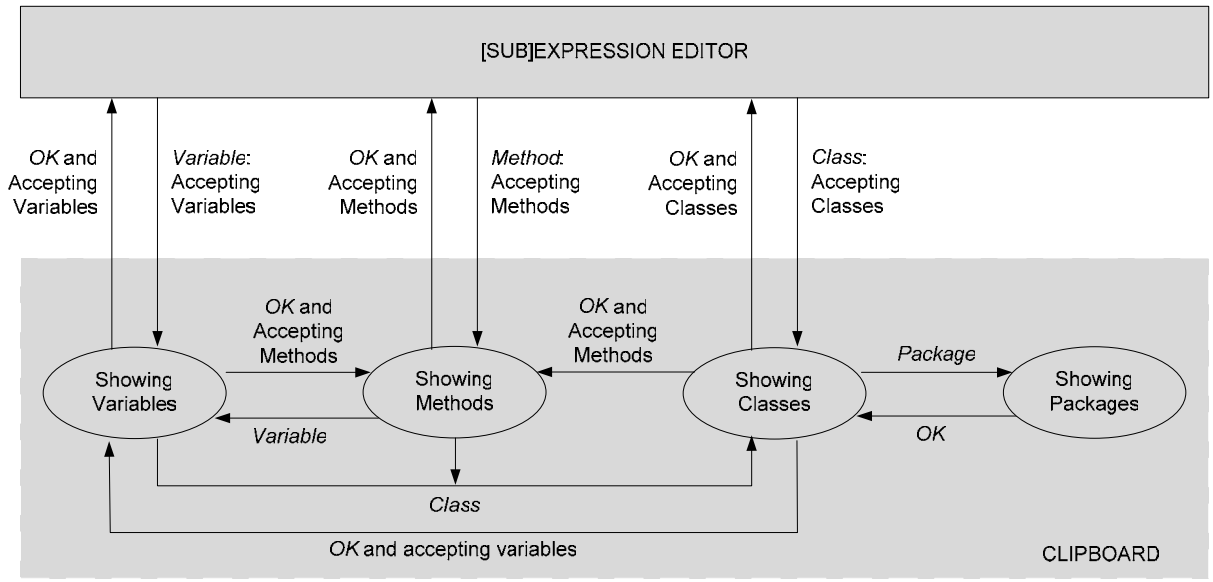



Figure 2. Clipboard States for Variable, Method and Class Entry

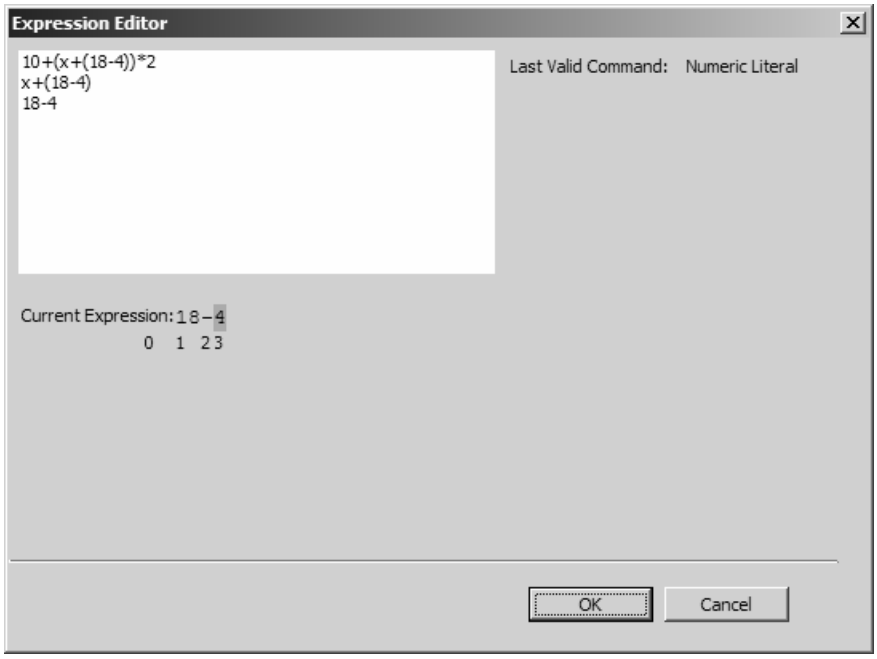



Figure 3. Screen Shot of Expression Editor

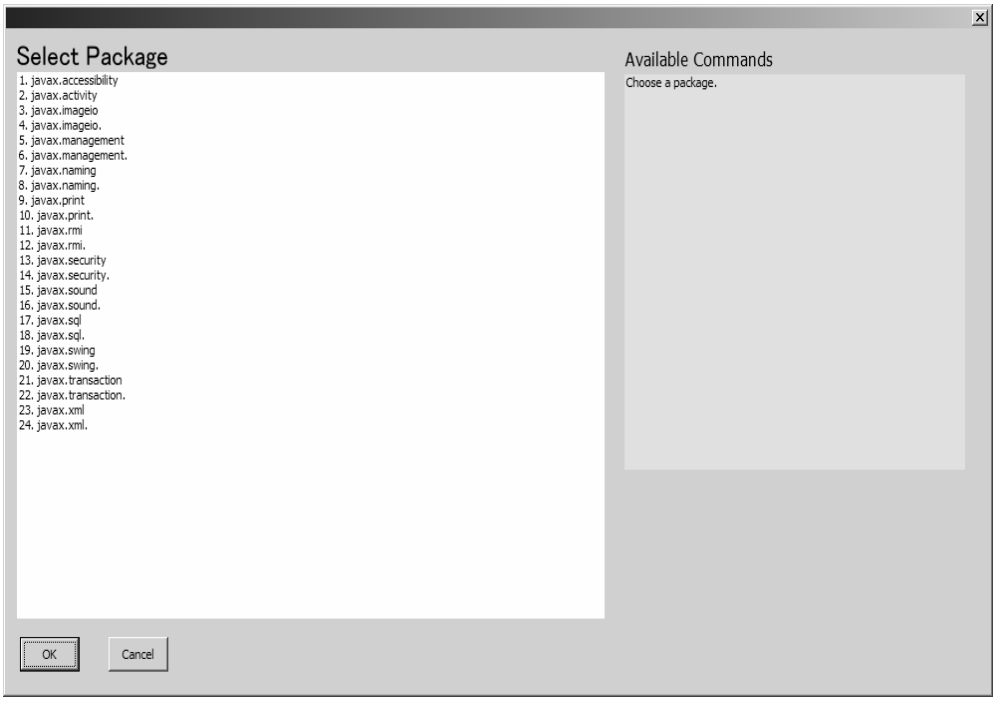



Figure 4. Screen Shot of Package Clipboard

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# Speedy Algorithm of Public Traffic Route Selection Based on Adaptive Backbone Network

Xiaowen Xu
Management School, Xi'an Jiaotong University, Xi'an 710049, China
E-mail: xiaowen@mail.xjtu.edu.cn
Jiayin Wang
School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China
E-mail: wangjiayin@mail.xjtu.edu.cn

*The research is supported by National Natural Science Foundation of China (No. 70702030) and National Undergraduate Innovation Experimental Project of China (No. 610762).*

**Abstract**

The inquiry service of public traffic routes is the important part of urban information service, which core is the public traffic route selection algorithm. However the traditional public traffic route selection algorithms have high time complexities and cannot support the inquiry of multiple changes. In this article, we put forward the speedy algorithm of public traffic route selection based on adaptive backbone network. In this algorithm, if the public traffic routes which pass certain public traffic station exceed or equal a certain value, so the station is defined as the backbone station, and the backbone stations and the public routes which pass them compose backbones network. If changes are limited at the backbone stations, so we can realize multiple changes inquiry and reduce computation in 10% of traditional algorithms through changing the certain value to adjust the backbone network.

**Keywords:** Adaptive backbone network, Backbone stations, Tabu Search, Public traffic route selection, Speedy algorithm

## 1. Introduction

The inquiry service of public traffic routes is the important measure to convenient for people's daily life and enhance the operation efficiency of public traffic route network, and it is also the important composing of urban information service, which core is the public traffic route selection algorithm. At present, traditional public traffic route selection algorithms include Dijkstra algorithm, Floyd algorithm or their ameliorations (Lu, 2001, p.68-70, Wang, 2007, p.63-67, Yue, 1999, p.209-212, Zhu, 2007, p.121-122 & Xu, 2005, p.16-17), which time complexities respectively are $O(|V|^2)$ and $O(|V|^3)$ (Thomas, 2007, p.324-330), but when a large of public traffic stations exist, higher time complexity will induce that inquiry time far exceeds what inquirer hopes. At the same time, these two algorithms cannot support route inquiry of multiple changes and cannot realize the first aim of least change times (Yang, 2000, p.87-91 & Ma, 2004, p.38-44). Aiming at problems existing in traditional algorithms, in this article, we put forward the speedy algorithm of public traffic route selection based on adaptive backbone network. This algorithm induces computation quantity and expedites computation speed through the backbone network technique, and can support the route inquiry of multiple changes through the adaptability of backbone network.

## 2. Adaptive backbone network

Suppose that the public traffic stations are the crunodes in the public traffic network map and all public traffic routes between any two public traffic stations are the directional routes, so the public traffic network map can be noted as $G=(V,E,L)$, where, V is the set of all public traffic stations, E is the set of all directional public traffic routes between any two public traffic stations, and L is the set of all public traffic routes.

Taking the public traffic route layout of the main city zone in Beijing, the main city zone in Beijing has 3957 public traffic stations (No.0001-No.3957) and 520 public routes (No.001-No.520). Suppose that $v_i$ represents the public traffic station which number is $i$, $l_j$ represents the public traffic route which number is j, so $V=\{v_1,v_2,\cdots,v_{3957}\}$ and $L=\{l_1,l_2,\cdots,l_{520}\}$. And if station $v_i$ and station $v_j\,(i,j\in[1,3957])$ are located on the uplink (or downlink) at the public route $l_k\,(k\in[1,520])$ and the next station of $v_i$ is $v_j$, so the directional route $(i,j)$ is one side in map $G$.

Every public traffic route must pass some pubic traffic stations, and every public traffic station must at one public traffic route at least, and suppose that the set of all public traffic routes which pass station $v_i$ is $L(v_i)$, so the set of all stations which pass station $v_i$ on all the public traffic routes is $V(L(v_i))$.

Suppose that the station $v_i$ is the backbone station with n adaptive degree which fulfills $L(v_i) \geq n\,(n \in N)$ and is noted as $v_i^n$, $V^n = \{v_i^n \mid i \in [1,3957] \wedge L(v_i) \geq n\}$ is the set of all backbone stations with n adaptive degree, $E^n$ is the set of all directional public traffic routes between any two backbone stations with n adaptive degree, $L^n$ is the set of public routes which pass the backbone station with n adaptive degree, and the sub-map of the public traffic network map which only includes all backbone stations with n adaptive degree is the backbone network with n adaptive degree, which is noted as $G^n = (V^n, E^n)$.

To establish the backbone network with n adaptive degree should possess two conditions, one is which can cover most public traffic routes and the other is that the purpose station which is reachable on the public traffic route also can be reachable on the backbone network with huge probability. Suppose that the reachable probability $\eta$ between any two stations on the backbone network is the square of the ratio between the sum of station which $L^n$ covers and the gross of public traffic stations, so Table 1 can be obtained through computation.

From Table 1, when $n = 20$, $|V^n| = 335$ and $L^n = 505$, i.e. the backbone stations which only occupy 8.47% of the gross of public traffic station cover 97.12% of public traffic routes, and the reachable probability between any two stations on the backbone network achieves 96.25%. Therefore, the reachability among most stations can be actualized through the backbone network and changes among backbone stations, and suppose that all changes can be actualized only at backbone stations in this article, thus it only needs to compute backbone stations not all public traffic stations, and when the gross of public traffic stations is large, this method can effectively reduce computation quantity and get approximate optimal solution. Suppose that the limitation of inquiry may induce the route obtained is not the optimal route, i.e. the approximate optimal solution, and considering the actual situations that the backbone stations are the zones that ten or tens of public traffic routes gather and the zones with dense human streams, so the route strange inquirer gets is convenient for memory and identification. If the station inquired cannot be reached through the backbone network, the backbone network needs to be extended. The so-called adaptive backbone network is to dynamically control the scale of backbone network according to actual needs of inquiry, which can not only ensure the solvability, but also achieve the speediness of solution.

The basic idea to establish adaptive backbone network algorithm is that the inquirer gives that start station $v_{start}$ and the purpose station $v_{end}$, and if $n = \min(L(v_{start}), L(v_{end}))$ and go through all public traffic stations, the station $v_i$ which fulfills $L(v_i) \geq n$ is the backbone station with n adaptive degree, so the backbone network $L(v_i) \geq n$ is confirmed.

**3. Algorithm description**

The basic idea of route selection algorithm based on the backbone network is that inquirer gives any start station $v_{start}$ and purpose station $v_{end}$, firstly compute $L(v_{start})$ and $L(v_{end})$, judge whether $L(v_{start}) \cap L(v_{end})$ is $\phi$, and if it is not $\phi$, so $v_{start}$ and $v_{end}$ can be reached directly, so record the reachable public traffic route, or else $v_{start}$ and $v_{end}$ can not be reached directly, and the change is needed.

If $v_{start}$ and $v_{end}$ can not be reached directly, we should consider one time change, compute the set $V_B(L(v_{start}))$ of all backbone stations on those public traffic routes which pass the station $v_{start}$ and the set $V_B(L(v_{end}))$ of all backbone stations on those public traffic routes which pass the station $v_{end}$, judge whether $V_B(L(v_{start})) \cap V_B(L(v_{end}))$ is $\phi$, and if it is not $\phi$, so $v_{start}$ and $v_{end}$ can be reached through one time change at some backbone station, so record the reachable public traffic route with one time change, or else $v_{start}$ and $v_{end}$ can not be reached through one time change, and multiple changes are needed.

If $v_{start}$ and $v_{end}$ can not be reached through one time change, we should consider two times changes, firstly compute the set $V_{start} = V_B(L(v_{start}))$ of backbone stations that the start station can reach directly, then compute the set $V_{end} = V_B(L(v_{end}))$ of backbone stations which can reach the purpose station directly, and let $v_i \in V_{start}$ and $v_j \in V_{end}$ at random, judge whether $v_i$ and $v_j$ can be reached directly, i.e. look for $l_k$ and make $v_i \in V(l_k) \cap v_j \in V(l_k)$, and if

one or more $l_k$ can be found, it indicates $v_{start}$ and $v_{end}$ can be reached through two times changes, so record the reachable public traffic route with two times change, or else $v_{start}$ and $v_{end}$ can not be reached through two times change.

If $v_{start}$ and $v_{end}$ can not be reached through two times change, i.e. any station in $v_{start}$ and any station in $v_{end}$ can not be reached directly, we can consider three times changes, and for this situation the recursion is used. Let $v_i \in V_{start}$ and $v_j \in V_{end}$ at random, judge whether $v_i$ and $v_j$ can be reached directly, and the algorithm that computes whether two stations can be reached through one time change has been given. If $v_i$ and $v_j$ can be reached through one time change, it indicates that $v_{start}$ and $v_{end}$ can be reached through three times changes, and if any station in $v_i$ and any station in $v_j$ can not be reached through one time change, so it indicates that $v_{start}$ and $v_{end}$ can not be reached through three times changes. Through the recursion in turn, any two stations can be reached through some times changes.

The steps of the algorithm are as follows:

The inputs of the algorithm include the start station $v_{start}$ and the end station $v_{end}$. The outputs of the algorithm include the sets of reachable routes.

Step 1: Suppose that cyclic variable $i=1$, variable $L(v_{start})=L(v_{end})=\phi$, take $l_i$ in L, and if $v_{start} \in V(l_i)$, so $L(v_{start})=L(v_{start})\cup\{l_i\}$, and if $v_{end} \in V(l_i)$, so $L(v_{end})=L(v_{end})\cup\{l_i\}$, and if $i \neq 520$, so $i=i+1$ and go to Step 1.

Step 2: If $L(v_{start})\cap L(v_{end}) \neq \phi$, so record these reachable routes and exit.

Step 3: Suppose that cyclic variable $i=1$, variable $V_B(L(v_{start}))=\phi$, take $l_i$ in $L(v_{start})$, and let $V_B(L(v_{start}))=V_B(L(v_{start}))\cup\{v_j \mid v_j \in V(l_i)\cap v_j \in V^n\}$ (i.e. merge all backbone stations on route $l_i$ into $V_B(L(v_{start}))$, where $V^n$ is confirmed by the algorithm established by the backbone network.), if $i<|L(v_{start})|$ and $i=i+1$ then go to Step 3.

Step 4: Suppose that cyclic variable $i=1$, variable $V_B(L(v_{end}))=\phi$, take $l_i$ in $L(v_{end})$, and let $V_B(L(v_{end}))=V_B(L(v_{end}))\cup\{v_j \mid v_j \in V(l_i)\cap v_j \in V^n\}$ (i.e. merge all backbone stations on route $l_i$ into $V_B(L(v_{end}))$, where $V^n$ is confirmed by the algorithm established by the backbone network.), if $i<|L(v_{end})|$ and $i=i+1$ then go to Step 4.

Step 5: If $V_B(L(v_{start}))\cap V_B(L(v_{end})) \neq \phi$, so record these reachable combined routes through one time change and exit.

Step 6: Take $v_i$ in $V_B(L(v_{start}))$ at random, and take $v_j$ in $V_B(L(v_{end}))$ at random, let $v_{start}=v_i$ and $v_{end}=v_j$, i.e. take $v_i$ and $v_j$ as the new start station and the new purpose station, recursively transfer the algorithms in Step 1 and Step 2, and if $v_i$ and $v_j$ can be reached directly, so $v_{start}$ and $v_{end}$ can be reached through two times changes, export route selection information and exit.

Step 7: Take $v_i$ in $V_B(L(v_{start}))$ at random, and take $v_j$ in $V_B(L(v_{end}))$ at random, let $v_{start}=v_i$ and $v_{end}=v_j$, i.e. take $v_i$ and $v_j$ as the new start station and the new purpose station, recursively transfer the algorithms in Step 3, Step 4, and Step 5, and if $v_i$ and $v_j$ can be reached directly, so $v_{start}$ and $v_{end}$ can be reached through three times changes, export route selection information and exit, recursively transfer until the feasible solution with least change times can be obtained.

The algorithm will get some feasible route combinations with "least change times", so we can evaluate these routes according to inquirer's actual needs and select one group or several groups of optimal route combinations which can fulfill inquirer's requirement.

## 4. Conclusion

In this article, we put forward the speedy algorithm of public traffic route selection based on adaptive backbone network. In this algorithm, if the public traffic routes which pass certain public traffic station exceed or equal a certain value, so the station is defined as the backbone station, and the backbone stations and the public routes which pass them compose

the sub-map of public traffic network map, i.e. the backbones network. The algorithm supposes that change can only be actualized at the backbone station, and because backbone stations almost cover all pubic traffic routes, so most public traffic stations can be reached under that hypothesis. Because the quantity of backbone stations generally don't exceed 10% of the public traffic station gross, so the algorithm only needs to compute backbone stations not all public traffic stations when computing changes, and the computation quantity can be reduced fully comparing with traditional algorithms, and the speedy inquiry can be actualized. At the same time, the algorithm can actualize the inquiry with multiple changes through recursive computation, which can better fulfill inquirer's demand. Considering the solvability of the route inquiry among any public traffic stations, this article puts forward the method to dynamically control the scale of backbone network, i.e. adaptive backbone network, accordingly gives attention to the solvability and the speediness of the solution.

**References**

Lu, Zhenbo & Huangwei. (2001). City Public Transport Guiding System: Applications of GIS in Public Transport Management. *Journal of Highway and Transportation Research and Development.* Vol.18. No.5. p.68-70.

Ma, Lianghe, Liu, Xinbin & Liao, Daqing. (2004). The Shortest Path of Public Transport Net and Related Optimal Taking Bus Problem. *Mathematics in Practice and Theory.* Vol.34. No.6. p.38-44.

Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest & Clifford Stein. (2007). *Introduction of Algorithms (2nd Edition)*. Beijing: Machine Industry Press. p.324-330.

Wang, Shixiang & Rao, Weiya. (2007). The Data Structure of Inquiry for Public Transport Route and the Realization of Its Algorithm in Major City. *Computer Systems Applications.* No.9. p.63-67.

Xubing & Xie, Shiyi. (2005). The Implementation of Bus Change Algorithm Based on Station Priority. *Fujian Computer.* No.7. p.16-17.

Yang, Xinmiao, Wanghui & Ma, Wenteng. (2000). GIS-Based Public Transit Passenger Route Choice Model. *Journal of Southeast University (Natural Science Edition).* Vol.30. No.6. p.87-91.

Yueyang & Gong, Jianya. (1999). An Efficient Implementation of Shortest Path Algorithm Based on Dijkstra Algorithm. *Journal of Wuhan Technical University of Surveying and Mapping (WTUSM).* Vol.24. No.9. p.209-212.

Zhu, Jiangyun & Wang, Yukun. (2007). The Optimal Route Algorithm Based on Least Change Times. *Fujian Computer.* No.3. p.121-122.

Symbol table.

| | |
|---|---|
| $V$ | The set of all public traffic stations |
| $v_i$ | The public traffic station which number is i |
| $L$ | The set of all public traffic routes |
| $l_j$ | The public traffic route which number is j |
| $\|A\|$ | The degree of set A, A is the set of public traffic stations or routes |
| $V(A)$ | The set of all public traffic stations at A, A is the public traffic station or the set of public traffic station |
| $V_B(A)$ | The set of all backbone stations at A, A is the public traffic station or the set of public traffic station |
| $L(v_i)$ | The set of public routes which pass station Vi |
| $v_i^n$ | The backbone station with n adaptive degree which number is i |
| $V^n$ | The set of all backbone stations with n adaptive degree |
| $L^n$ | The set of public routes which pass the backbone station with n adaptive degree |
| $E^n$ | The set of all directional public traffic routes between any two backbone stations with n adaptive degree |

| | |
|---|---|
| $v_{start}$ | The start station which inquirer gives |
| $v_{end}$ | The purpose station which inquirer gives |

Table 1. The reachable probability of any two stations in the backbone network

| $n$ | $\left\|V^n\right\|$ | $L^n$ | $\eta^n$ |
|---|---|---|---|
| 10 | 846 | 516 | 99.14% |
| 11 | 807 | 516 | 99.14% |
| 12 | 704 | 515 | 99.14% |
| 13 | 657 | 515 | 99.14% |
| 14 | 559 | 514 | 99.04% |
| 15 | 528 | 513 | 98.64% |
| 16 | 466 | 510 | 97.24% |
| 17 | 442 | 510 | 97.24% |
| 18 | 398 | 508 | 96.74% |
| 19 | 373 | 505 | 96.25% |
| 20 | 335 | 505 | 96.25% |

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# Cognitive Radio

Kevin Curran, Gary O'Callaghan, Fionnuala McCullough
University of Ulster, Northern Ireland, UK
Tel: +44 2871 375565 E-mail: kj.curran@ulster.ac.uk

**Abstract**

Cognitive radio is where wireless devices are sufficiently computationally intelligent about radio resources and related computer-to-computer communications to detect user communications needs as a function of use context, and provide radio resources and wireless services most appropriate to those needs. Its roots came from the discovery by regulatory bodies in various countries which found that most of the radio frequency spectrum was inefficiently utilised. For example, cellular network bands are overloaded in most parts of the world, but amateur radio and paging frequencies are not. Independent studies concluded that spectrum utilization depends strongly on time and place. Moreover, fixed spectrum allocation prevents rarely used frequencies from being used by unlicensed users, even when their transmissions would not interfere at all with the assigned service. This was the reason for allowing unlicensed users to utilize licensed bands whenever it would not cause any interference.

Cognitive radio can monitor, observe, and react to events in a specified environment. The fundamental benefits are the ability to self-adapt, self-manage, and self-optimize under normal conditions, and the ability to self-diagnose and self-heal when unusual problems arise e.g. base stations can automatically adjust operating range (e.g., by controlling transmit power, receiver sensitivity, data rate, frequency channel selection) to compensate for neighboring base stations that fail. In contrast to many wireless systems where static frequency channels are pre-selected in advance and tedious frequency planning is mandatory to minimize interference from equipment operating in adjacent coverage areas. CR can achieve a performance that is close to a licensed band even when using an unlicensed band due primarily to its inherent ability to switch to best channels. This paper provides an overview of Cognitive Radio.

**Keywords:** Cognitive radio, Software defined radio, Wireless, Networks

## 1. Introduction

The term cognitive radio (CR) was invented by Joseph Mitola (Mitola et al., 1999). CR refers to a software based radio system which is capable of sensing factors in its environment such as geographical location and the RF characteristics of other radio devices in the same locale, the CR device can then alter its power, frequency, modulation and other parameters to dynamically reuse available radio spectrum (Mannion, 2004). CR builds on the invention of Software Defined Radio (SDR), whereby for example the software in a cell phone defines the parameters of operation when the user moves from place to place. In the US the development CR is led by the Federal Communication Commission (FCC), who deal with commercial applications, and the Defence Advanced Research Projects Agency (DARPA), who deal with military applications (Mannion, 2004).

Cognitive radio can monitor, observe, and react to events in a specified environment. The fundamental benefits are the ability to self-adapt, self-manage, and self-optimize under normal conditions, and the ability to self-diagnose and self-heal when unusual problems arise e.g. base stations can automatically adjust operating range (e.g., by controlling transmit power, receiver sensitivity, data rate, frequency channel selection) to compensate for neighboring base stations that fail. In contrast to many wireless systems where static frequency channels are pre-selected in advance and tedious frequency planning is mandatory to minimize interference from equipment operating in adjacent coverage areas. CR can achieve a performance that is close to a licensed band even when using an unlicensed band due primarily to its inherent ability to switch to best channels (Devroye et al., 2006). This paper will present at overview of cognitive radio but first we introduce the concepts of licensed and unlicensed frequency bands and software defined radio.

## 2. Licensed and Unlicensed Frequency Bands

It can be extremely difficult to locate radio spectrum and even harder to find spectrum that is available worldwide. Spectrum allocation in the international community is controlled by multiple regulatory bodies such as the FCC in U.S., MKK in Japan and CEPT in Europe). The Licensed bands are generally not prone to interference problems but premium licensed bands can be costly. For example, 2.5 GHz, 3.5 GHz (outside U.S.) and the 4.9 GHz (public safety) are popular licensed bands for wireless access e.g., Wi-Max, wireless DOCSIS. The license-free frequency bands however are

widely employed and they provide dynamic, opportunistic access to spectrum for a limited time period. There are benefits such as large-scale frequency planning is avoided and ad-hoc networks also become possible. The 5 GHz band is increasingly used to support wireless backhaul transmissions e.g., multihop mesh networks. It is becoming increasingly deregulated and in the UK market forces are starting to have a bigger say in the use and allocation of radio spectrum e.g. the UK recently deregulated 865-868 MHz band for RFID use.

In the US the FCC assigns users to specific frequencies which include AM, FM radio, shortwave, citizens band, VHF and UHF for TV and bands for cellular phones, GPS, air traffic control, security alarms and radio controlled keys. With the development of new wireless devices the demand on the radio spectrum will only continue to increase (Ashley, 2006). Legacy hardware can exacerbate this shortage, for example vacuum tube TV sets in the 1950's forced new transistor based model to receive only VHF signals. Such hardware related inflexibility is now addressed by software based wireless designs. The next generation wireless technology called Software Defined Radio (SDR) uses embedded signal-processing algorithms to sift out weak signals and reconfigurable code structures to receive and transmit new radio protocols (Ashley, 2006). SDR code and other programmable radio frequency front end interface technologies on laptops could receive TV signals. If fitted with an SDR RF card, for instance, a laptop could then be used as a cellular handset or base station. CR advances this situation by allowing the SDR system to reconfigure its analog RF output, and incorporate self awareness, transmission protocols and etiquette. This would free the user a controlling network and their contract fees (Ashley, 2006).

In 2004, the FCC commissioned a Notice of Proposed Rule Making for how CR could be realized. This coincided with other work to open up new radio spectrum to commercial use, e.g. opening spectrum between 3.1 and 10.6 GHz to commercial use for ultra wideband signalling. In their Spectrum Policy Taskforce report of 2002 the FCC recommended a migration towards a "policy based" solution allowing an opportunity to seamlessly move across spectrum using policy engines to check whether this is permissible (Mannion, 2004). The FCC has identified much potential in allowing the 6MHz wide band in the UHF band currently assigned for television to be opened up for to secondary markets for last mile data access. This would give the 54 Mbps available with WiFi but the bandwidth delivered would be useful especially in rural areas (Mannion, 2004). It is worth noting that the entire radio spectrum up to 100 GHz only 5 - 10% is in use at any given instant (Bing, 2004).

Software Defined Radio will allow the industry to find more creative and efficient use of the airwaves, which in turn, will continue to bring benefits to consumers however software must have sufficient controls so that devices cannot be modified to operate outside FCC-approved parameters. In September 2005, Cisco received the first FCC certification for 802.11a SDR.

## 3. Cognitive Radio

Software defined radio (SDR) is seen as a key technology for future wireless communications (Bing, 2004). SDR hardware is largely reconfigurable and reprogrammable by software. This opens the way for new services and prolongs the lifespan of wireless devices. A device with flexible radio hardware is capable of running different applications, much like PC hardware. A high degree of reconfiguration requires support for different system functionalities. Multi-band systems support more than one frequency band and multi-homing systems support more than one radio standard or service. An example of SDR technology is Motorola's Canopy product (Note 1). Cognitive Radio (CR) is a concept with which a network or wireless node changes its transmission parameters without interfering with licensed users. The alteration of these parameters is based on several factors found in the external and internal radio environment. Examples of these factors are the radio frequency spectrum, user behaviour and network state. The main functions that a Cognitive Radio can perform are:

(1) Spectrum Sensing – Detecting unused spectrum and sharing it without interference to others.

(2) Spectrum Management – Capturing the best spectrum to meet user requirements.

(3) Spectrum Mobility – a cognitive radio user exchanges the frequency of the operation.

(4) Spectrum Sharing – Provides a spectrum scheduling method.

Smart radio that leverages on SDR hardware to perform intelligent and opportunistic sharing of unlicensed radio spectrum employs real-time interaction with its environment to determine transmitter parameters such as frequency, power, modulation and learns when to operate and when to interrupt service. Smart radio must rely on intelligent protocols to adapt spectrum use in response to location and operating environment ultimately leads to "plug and play" wireless systems which require minimum or virtually no manual intervention or frequency management during deployment.

Potential military applications for CR are being developed by DARPA in their XG (next generation communications) in order to allow multiple users to share spectrum in a way that coexists and complements sharing protocols used in WiFi technologies (Mannion, 2004; Bing, 2004). CR is capable of reconfiguring its communication functions based on prior experience, so in the case of car radio, CR would allow the device to build up a database of propagation characteristics, signal strength of different transmission bands throughout car journeys. It could then use this information to decide how

best to transmit at different times of the day and in different places. This dynamic use of bandwidth will free up the radio spectrum and allow for more RF options which are more dependable and considerably cheaper than at present (Ashley, 2006).

In traditional wireless systems most of the intelligence to make the network operate efficiently resides centrally in the network. In CR by contrast this intelligence would reside in the handsets, laptops or wireless devices. A cognitive controlled subsystem controls the SDR allowing a CR unit to detect RF networking opportunities wherever it finds itself. It is envisioned that CR devices will exist in networks of other CR devices with each contributing to and taking information from a central database which contains the information necessary to transmit effectively in a particular location at a particular time (Ashley, 2006). Although the communications and wireless industries are now only beginning to take notice of cognitive radio, the U.S. government is already interested in the possibilities that CR offers. Cognitive radios are thought to be a powerful tool for explaining and solving general and specific spectrum access issues, an example of this is locating an open frequency and using it. It improves current spectrum use e.g. fills in the unused spectrum and leaves spectrum that is already in use alone and can improve wireless database performance with increased user throughput and system consistency. The FCC in the US is seeking to remove regulatory impediments in order to continue development and deployment of cognitive radios. Cognitive radios are a possible solution to interoperability issues that are present in public safety organisations especially first responders.

## 4. Conclusion

Cognitive radio has the potential to allow multidimensional reuse of spectrum in space, frequency and time and remove the spectrum and bandwidth limitations that slowed broadband wireless development around the world. This new software is closely related to Software- Defined Radio (SDR). SDR is the software found within a cell phone which operates in real time as the user goes from location to location. Cognitive Radio is more intelligent however as it is a radio that is aware of and can sense its environment; learn from its environment, and perform functions that best serve its users. Although Cognitive Radio can be described as software defined radio, there are problems in designing high quality spectrum sensing devices and algorithms for exchanging spectrum sensing data between nodes. Applications of spectrum sensing include emergency networks, Wireless LANs, higher throughput and transmission distance extensions.

Cognitive Radio is poised to take advantage of the increasing deregulation of radio spectrum to provide high-speed broadband services. It can potentially lead to virtually unlimited wireless bandwidth when spectrum is used and reused more efficiently and co-operatively and ultimately increase capacity and efficiency as a direct result of being able to switch between momentarily idle channels in different portions of radio spectrum for short period of usage. This will solve two key problems in multihop networks which are the ability to operate on multiple non-interfering channels removes the bandwidth penalty associated with single-channel multihop systems and the ability to sense for spectrum availability and switch to different channels dynamically provides an excellent solution to wireless Denial of Service (DoS) attacks and network intrusion problems. The technical foundations for Cognitive Radio have been established with the development of Wireless LANs, where dynamic frequency selection and transmit power control are core features. Software functions like filtering, band selection and interference mitigation need to be further developed into this model however many technical hurdles still remain to be overcome such as determining an acceptable level of interference with other radios in the same locale so that it can stay within an acceptable threshold.

## References

Ashley, S. (2006). Cognitive Radio, *Scientific American*, Vol. 294, No. 3, March 2006.

Bing, B. (2005). *All in a Broadband Wireless Access Network*, December 2005, ISBN: 0976675218.

Devroye, N., Shyy, D. and Dunyak, J. (2006). Limits on Communications in a Cognitive Radio Channel, *IEEE Communications Magazine*, Vol. 44, No. 6, pp. 44-49 June 2006.

Mannion, P. (2004). EETimes - *Sharing spectrum the smarter way*, 4th of May 2004.

http://www.eetimes.com/showArticle.jhtml? article ID=18700443.

Mitola, J. and Maguire, G. (1999). Cognitive Radio: Making Software Radios More Personal, *IEEE Personal Communications*, Vol. 6, No. 4, 1999.

Tanenbaum, A.S. (2006). *Computer Networks*, Fifth edition, Prentice Hall.

## Notes

Note 1. http://motorola.canopywireless.com/

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# The Design of Switched Reluctance Motor Controller Based on DSP

Zhen Guo, Xiaoyan Zhou, Suzhen Wang
School of Electrical Engineering, Qingdao Technological University, Qingdao 266520, China
Tel: 86-532-88782001 E-mail: qs2004b@163.com

**Abstract**

In this paper, a new design method of the three-phase 12/8 Switched Reluctance Motor controller is presented, including hardware design and software design. It provides a novel power converter which can accelerate winding discharging and optimal current waveforms. Simulation results of the Switched Reluctance Motor are given which show the system can achieve expected performances.

**Keywords:** Switched Reluctance Motor (SRM), Controller, Novel power converter

## 1. Introduction

Switched Reluctance Motor (SRM) has an intrinsic simplicity and low cost that make them well suited to many applications. Furthermore, the motor has a high robustness due to the ability to operate with the loss of one or more motor phase and are thus well suited to operate in harsh industrial environment. However, the motor has many drawbacks due to the motor's doubly salient structure as well as highly nonuniform torque output and magnetization characteristics. The double salient structure leads to the inability to excite the motor using conventional ac motor waveforms and thus the inability to apply well established ac motor rotating field theory to the motor. Additionally, the highly nonlinear magnetization characteristics of the motor entail that the control of the motor is complex. Hence, the problem of the SRM controlling has been an ongoing area of research (P.J.Lawrenson, 1992).

The Switched Reluctance Driver is a new type of variable-speed drive, it is very important to design perfect and reliable controller which is mainly made up with microcomputer controller, power converter, drive circuit, sensors. This paper presents the design method of the three-phase 12/8 SRM controller. The simulation and experimental results are given.

## 2. Primary Equations of SRM

Although SRM operation appears simple, an accurate analysis of the motor's behavior requires a relatively complex, mathematical approach. The instantaneous voltage across the terminal of a single phase of an SRM winding is related to the flux linked in the winding by Faraday's law, so voltage equation is described as

$$v = iR_m + \frac{d\phi}{dt} \tag{1}$$

where $v$ is the terminal voltage, $i$ is the phase current, $R_m$ is the motor resistance, and $\phi$ is the flux linked by the winding.

Because of the double salient construction of the SRM and its magnetic effects, in general the flux linked in one phase varies as a function of rotor position $\theta$ and the motor current. Thus, equation (1) can be expanded as

$$v = iR_m + \frac{\partial\phi}{\partial i}\frac{di}{dt} + \frac{\partial\phi}{\partial\theta}\frac{d\theta}{dt} \tag{2}$$

Often, SRM analysis proceeds under the assumption that the motor remains unsaturated during operation. When magnetic saturation is neglected, the relationship from flux to current is given by

$$\phi = L(\theta)i \tag{3}$$

where the motor inductance $L$ varies only as a function of motor angle. Defining $W_c$ is stored field co-energy, then

$$W_c = \frac{i^2}{2}L(\theta) \tag{4}$$

and the torque equation is

$$T = \frac{i^2}{2}\frac{dL}{d\theta} \tag{5}$$

Equation (5) suggests that positive torque is produced when the motor inductance is rising as the shaft angle is increasing, $\frac{dL}{d\theta} > 0$. Similarly, a negative torque is produced by supplying SRM winding current while $\frac{dL}{d\theta} < 0$.

## 3. Hardware Description

Due to the advantages of the TMS320F240 in controlling motors (TMS320C2XX), this paper choose TMS320F240 as main control unit. TMS320F240 has nine independent PWM channels, providing maximum flexibility for SRM. TMS320F240 has dual ADC model with eight channels in each ADC. Therefore, phase currents can be read simultaneously. A current controlled SRM drive is implemented by using Evaluation Module (EVM) for TMS320F240 interfaced with a motor interface board and power converter board. The block diagram of the SRM driving control system is given by Figure 1. The following sections describe the major component in detail.

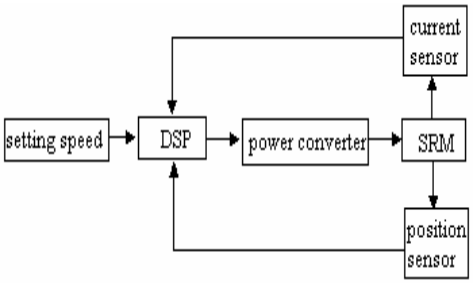



Figure 1. Block diagram of the SRM driving control system

### *3.1 Power converter*

The power inverter is an important issue in SRM control because it largely delicates how the motor can be controlled. SRM do not require bi-directional current like other common ac motors. Therefore, unipolar converters are used as the power converter for SRM drivers. Figure 2 shows the converter that is used in this paper. This particular converter is a modified version of a classical SRM converter. The switches and diodes are rated for the supply voltage with any required safety factor. During motoring operation, both power switching devices are turned on for a particular phase. The amount of current flowing through the SRM winding is regulated by switching on or off power device-GTR, which connect each SRM phase to a DC bus. With DC/DC converter, it can accelerate winding discharging and optimal current waveforms by regulating the voltage value of $U_d$. This converter provides independent control for each phase and consequently phase overlap operation can be implemented easily.

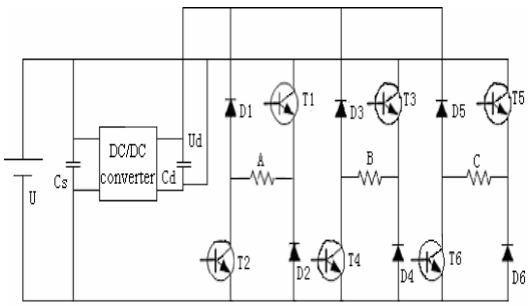



Figure 2. Three-phase power converter

### *3.2 Current and position sensing*

Different currents can be sensed for control purposes. In the particular implementation, phase currents are sensed to implement current controlled SRM drive. TMS320F240 has a dual ADC that enable a user to simultaneously sample and convert two variables. Total time required for the sample is 6.6μs. The phase currents are sensed at every 40μs to implement a 25kHZ current loop. The new PWM duty cycles obtained from current information is loaded at the beginning of a PWM cycle. This is achieved by programming ACTR and SACTR control registers of TMS320F240.

Measuring the rotor position of SRM is very important. Shaft position information is provided using an 8-slot, slotted disk connected to the rotor shaft and three opto-couplers mounted to the stator housing in this paper. The opto-couplers are nominally located 30º apart from each other along the circumference of the disk. Each opto-output is connected to both a capture input and a digital I/O input of TMS320F240. The capture inputs are used once the motor is running, and the digital I/O inputs are used for estimating initial rotor position and for starting the SRM.

*3.3 Microcomputer controller*

SRM control is described in terms of "low-speed" and "high-speed". Low-speed operation is typically characterized by the ability to arbitrarily control the current to any desired value. This mode of operation is called current chopping mode. As the motor's speed increases, it becomes increasingly difficult to regulate the current because of a reduced amount of time for the commutation interval, so the method of choice of the turn-on and turn-off is employed, This mode of operation is called the single-pulse mode. So the ultimate performance of the SRM is determined by the exact choice of the turn-on and turn-off angles and the magnitude of the phase current. The block diagram of the microcomputer controller is shown in Figure 3. A highly efficient, variable speed drive can be achieved with SRM by using voltage PWM with closed loop position control. The torque command is executed by regulating the current in the inner loop. The turn on timing and the total conduction period determines efficiency and other performance characteristics. The reference current is determined from the load characteristics. The controller needs current feedback information from each of motor phase.

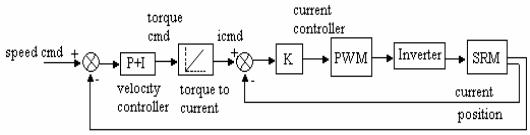



Figure 3. Block diagram of microcomputer controller

When the motor is in the situation of "low-speed", chopping current mode is used. The chopping current wave is shown by Figure 4, where $I_m$ is the chopping value, $\Delta I$ is the range of the fluctuation.

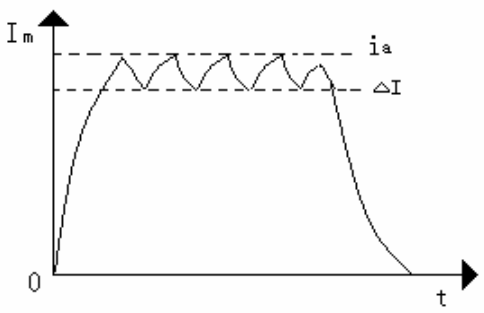



Figure 4. Chopping current wave

When the motor is in the situation of "high-speed", the single-pulse method is used. In order to detect the rotor position of SRM more accurately, frequency doubling technology is employed. The information of position transferred by shaft sensor is treated into 15º pulse signal for the three-phase 12/8 SRM. 15º pulse signal can be turned into 256 single-pulse, so the motor can be controlled more accurately. The relation between Q1 and Q2 is shown in Figure 5.

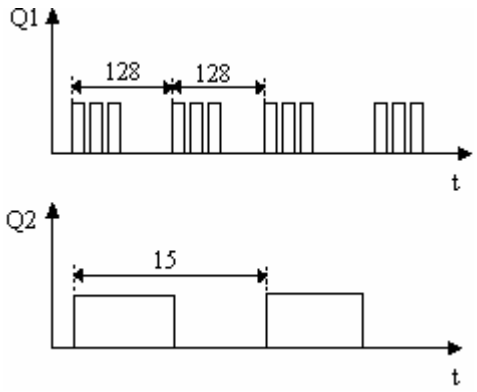



Figure 5. Q1 signal and Q2 signal

**4. Software Design**

The software written on TMS320F240 EVM implements all the control blocks described in previous sections. There are two loops in the run routines. Firstly, the current loop regulates phase current according to the reference. Secondly, the position loop checks the position and determines the proper rotation direction and proper commutation sequence.

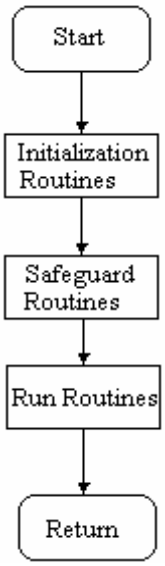



Figure 6. Flowchart of SRM controller

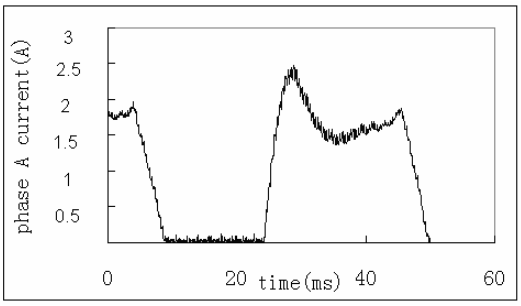
Figure 6 shows the structure of the SRM control software based on TMS320F240. It includes two parts. One is monitoring system, the other is controlling system that can accomplish some calculation in one periodic time, then provide control direction to the motor. Initialization routines mainly include DSP setup, event manager initialization, SRM algorithm initialization. Run routines include time interrupt service routine, capture interrupt service routine.

**5. Simulation and experimental results**

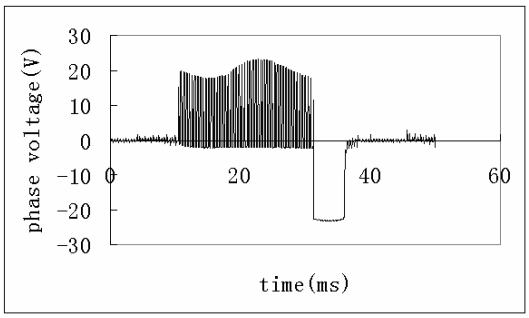
Based on the equations mentioned in section 2, simulation results for the three-phase 12/8 SRM are shown in Figure 7. The characteristics of the SRM used are given by Table 1.

Table 1. SRM parameters

| | |
|---|---|
| Number of phase, m | 3 |
| Number of stator poles, $N_s$ | 12 |
| Number of rotor poles, $N_R$ | 8 |
| Nominal phase resistance, $R_m$ | 8Ω |
| Nominal aligned inductance, $L_a$ | 9.69mH |
| Nominal unaligned inductance, $L_u$ | 1.07mH |
| DC bus voltage, $V_{bus}$ | 200V |




(a) Phase A current wave




(b) Phase torque wave

Figure 7. SRM speed 1500r/m simulation wave

Figure 7 describes the characteristics of current and torque of SRM in deferent speed. Different control mode is used in different work condition. By varying turn-on and turn-off angle can obtain different current wave and torque wave.

## 6. Conclusion

This paper analyzes the structure of the switched reluctance motor controller and provides design information in detail. With the novel power inverter employed, winding discharging can be accelerated and current waveforms can be optimized.

The presented work demonstrates that the measured phase current and voltage, together with the machine parameters, can be used to obtain an ideal controlled method of SRM under different conditions.

## References

P.J.Lawrenson. (1992). Switched Reluctance Drives: A Prospective.Proc.ICEM.

TMS320C2XX. User's Guide, Texas Instruments.

F.Filicori et al. (1993). Modeling and Control Strategies for a Variable Reluctance Direct-Drive Motor. *IEEE Tran*.on IE-40 No.1, 2.

Caio.A.Ferreira et al. (1995). Detailed Design of a 30-kW Switched Reluctance Starter/Generator System for a Gas Turbine Engine Application. *IEEE Tran*. On IA-31, No.3, May/June.

TJE Miller. (1993). *Switched Reluctance Motors and Their Control.* Magna Physics Publishing and Clarendon Press, Oxford.

TMS320C24x. (1997). *DSP Controllers-Reference Set*: Vol.1, Texas Instruments Inc.

B.K.Bose, T.J.E. Miller. (1986). "Microcontroller control of switched reluctance motor", *IEEE* Transanctions on Industry Applications, Vol. IA-22, July/August, pp.708-715.

P.J.Lawrenson, J.M.Stephenson, P.T.Blenkinsop, J.Corda,and N.N.Fulton. (1980). "Variable speed switched reluctance motors," *IEE Proc*.,vol.127, pt. B, no. 4, pp.253-265.

R.M.Davis, W.F.Ray, and R.J.Blake. (1981). "Inverter drive for switched reluctance motor: Circuits and component ratings," *IEE Proc*., vol.128, pt. B, no.2, pp. 126-136, Mar.

J.V.Byrne and F.Devitt. (1985). "Design and performance of a saturable variable reluctance servo motor," in *Proc. Motorcon* conf., Chicago, il, Oct., pp.139-146.

J. Ish-Shalom and D.G. Manzer. (1985). "Commutation and control of step motors," in *Proc.14th symp. Incremental Motion, Contr. Syst. and Devices*, June, pp: 283-292.

T.J.Miller. "Nonlinear analysis of the switched reluctance motor drive," General Electric Rep.84CRD079.

M.Ilic'-Spong, R.Marino, S. M. Peresada, and D. G. Taylor. (1987). "Feedback linearizing control of switched reluctance motors," *IEEE Tranns. Automat. Contr*., vol. AC-32, PP.371-379, May.

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# Flow Regime Classification Using Artificial Neural Network Trained on Electrical Capacitance Tomography Sensor Data

Khursiah Zainal-Mokhtar
School of Electrical and Electronic, Engineering Campus University Science Malaysia
14300 Nibong Tebal, Penang, Malaysia
Tel: 006-0195653288 E-mail: ven_conmigo1010@yahoo.com
Junita Mohamad-Saleh
School of Electrical and Electronic, Engineering Campus University Science Malaysia
14300 Nibong Tebal, Penang, Malaysia
Tel: 006-0194732732 E-mail: jms@eng.usm.my
Hafizah Talib
School of Electrical and Electronic, Engineering Campus University Science Malaysia
14300 Nibong Tebal, Penang, Malaysia
Tel: 006-0134181851 E-mail: smiley_ahaks@yahoo.com
Najwan Osman-Ali
School of Electrical and Electronic, Engineering Campus University Science Malaysia
14300 Nibong Tebal, Penang, Malaysia
Tel: 006-0195435865 E-mail: ee73620@eng.usm.my

*The research is fully funded by Ministry of Science, Technology and Innovation (MOSTI), Malaysia under eScienceFund grant.*

**Abstract**

The main goal of the presented work is to analyse the performance of the Multi-Layer Perceptron (MLP) neural network for flow regime classification based on sets of simulated Electrical Capacitance Tomography (ECT) data. Normalised ECT data have been used to separately train several MLPs employing various commonly used back-propagation learning algorithms, namely the Levenberg-Marquardt (LM), Quasi-Newton (QN) and Resilient-Backpropagation (RP), to classify the gas-oil flow regimes. The performances of the MLPs have been analysed based on their correct classification percentage (CCP). The results demonstrate the feasibility of using MLP, and the superiority of LM algorithm for flow regime classification based on ECT data.

**Keywords:** Multi-Layer Perceptron, Back-propagation algorithm, Electrical Capacitance Tomography, flow regime, Classification

## 1. Introduction

Knowledge of flow regimes is essential particularly for investigation involving multi-component flows such as in oil production. Many types of flow regime could interchangeably form along the pipeline during oil transportation. Since pipelines are normally concealed, a special technique is required in order to obtain information regarding flow regimes form in a pipeline.

Tomographic method was first applied in medical imaging in 1970s when hospital doctors used Computerised Tomography (CT) machine to visualise internal organs of a human body. Nowadays, there are various other tomography systems and ECT is one of the commonly used sensing methods in industries (Beck and Williams, 1996). ECT sensor is non-invasive in nature since it does not need to be inserted inside the pipeline, in which case, may intrude the material permittivity distribution. Moreover, it is non-hazardous since it involves electrical lines unlike CT machines which deal with radiation. Figure 1 illustrates a basic ECT system consisting of the sensors, data acquisition system and a computer system (Gamio et al., 2005). Using this technique, multiple electrodes are mounted equally around the periphery of an

insulating pipe vessel. These electrodes are sensitive to the dielectric constants of materials and the capacitance variations occur between combinations of electrodes. Measurements between electrode pairs give the capacitance values.

Conventionally, the capacitance values were used to determine the permittivity distribution with the aid of an image reconstruction algorithm. Then, the type of flow regime can be determined from the reconstructed image. However, image reconstruction method is too time-consuming particularly, when online process control is of main concern (Mohamad-Saleh and Hoyle, 2002). In addition, reconstructed images usually have low fidelity and thus, flow process parameter estimations based on the images are typically imprecise. Although rather accurate reconstructed images can be obtained by employing artificial neural network (Marashdeh et al., 2006) as the reconstruction mechanism, the reconstruction time is still a problem (Mohamad-Saleh and Hoyle, 2002).

Historically, the development of ANN technology was motivated by human brain's ability to solve very complex tasks. ANNs comprise of two main elements; the neurons or processing elements (PEs) and the connections between the PEs which have weight values. Figure 2 shows a schematic diagram of a basic PE model (Haykin, 1999).

PEs are normally arranged in layers, where at least three layers of PEs (input, hidden and output layers) are required to construct an ANN which is capable of solving non-linear problems. A variant of the three-layered ANN architecture is the Multi-Layer Perceptron (MLP) neural network as shown in Figure 3 (Haykin, 1999). In the figure, *x*, *w* and *y* represent the inputs, weights and output values, respectively. In classification and non-linear function approximation, MLP is the most common and popular ANN ever used today (Ventari and Lampinen, 2000; Sterjovski et al., 2005).

ANNs are not programmed. Instead, the supervised type of ANNs such as the MLP, learn from given sets of input-output examples, establishing a mapping between inputs and outputs by updating their weight connections, to become 'intelligent' (Bishop, 1994). They 'learn' to predict the correct class for a given sets of inputs based on a suitable learning algorithm. This work investigated the most suitable learning algorithm for MLP in the quest for flow regime classification using ECT data of various flow patterns.

The presented work is concerned with flow regime classification without the need for imaging. The direct approach involves artificial neural networks (ANNs) which have been trained to classify various type of the flow regime based on the sets of ECT data. The following section describes procedures for preparing a set of simulated data and presents the ECT sensor model used in the investigation. Subsequently, the development of an 'intelligent' MLP in classifying flow regimes is explained. The results of MLP performances based on various employed learning algorithms are then presented and discussed.

## 2. Development of the MLP Classification System

Sensors for an ECT system are simple electrodes. Figure 4 shows the ECT sensor's parameter dimensions designed for this investigation. The inner pipe (R1), outer pipe (R2) and screen wall (R3) radii have been chosen to be 1, 1.2 and 1.4 units, respectively. The measurements of unit coordinate are used to facilitate generic mapping of parameter dimensions. The primary sensor (β) and guard electrodes (α) have subtended angles of $22^o$ and $2.5^o$, respectively, making gaps of $8^o$ between each adjacent primary electrodes.

In this research, the ECT sensor model is chosen to have 12 electrodes since this is a common number of sensors used for ECT systems. With 12 electrodes, 66 distinct differences in capacitance between all possible pairs of electrodes can be obtained. This number is based on the equation below

$$m = \frac{N(N-1)}{2} \tag{1}$$

where N is the number of electrodes. It has to be noted that an ECT sensor creates the highest sensitivity around the edge of a sensing area which is nearer to the sensor electrodes, and the sensitivity tends to decrease towards the centre of the sensing area which is farther from the electrodes (Xie et al., 1992).

Figure 5 shows a schematic diagram of six flow regimes focused in this work. An ECT simulator (Spink, 1996) has been used to generate ECT data for different flow patterns of all investigated flow regimes based on the defined dimensions of the ECT sensor parameters. Every set of the ECT data corresponding to each flow pattern are normalized using,

$$C_{i,j} = \frac{C_{i,j}^m - C_{i,j}^l}{C_{i,j}^h - C_{i,j}^l} \tag{2}$$

where $C_{i,j}^m$ is the measured capacitance, $C_{i,j}^l$ is the capacitance when the pipe is full with the lower permittivity material, and $C_{i,j}^h$ is the capacitance when the pipe is full with the higher permittivity material.

The numbers of flow patterns generated for each flow regime are given in Table 1. Empty and full pipe regimes have only one flow pattern each. This is because only one type of material (i.e. gas or oil) exists for both flow regimes and fully occupies the pipeline. Unlike the full or empty regimes, the stratified has many different patterns since it can be simulated using different oil heights as well as different tilted angle. The same with stratified regime, bubble flow patterns can be simulated for various bubble radii, as well as different bubble locations and hence, a rather large number of patterns for bubble flow. The annular and core flow patterns are just the opposite of each other where a larger number of these flow patterns can be simulated by varying their radii sizes.

Once all data have been generated and normalized, they are divided into training, validation and testing sets of data based on 8:1:1 ratio, respectively. Then, the process of MLP development commences where MLP is put into training process. Figure 6 illustrates the MLP training procedure, which has been accordingly executed using MATLAB®.

Initially, the best transfer functions for the MLPs' hidden and output PEs are investigated by employing the training procedures. Since the output PEs should give either '1' or '0', the most suitable transfer function to be applied to them is the logarithmic sigmoid (log). As for the hidden PEs, either log or hyperbolic tangent (tanh) sigmoid transfer function can be applied. Hence, in the initial stage of the development, for a fixed number of hidden PEs, two investigations have been carried out, involving the combinations of log-log and tanh-log for the hidden and output PEs, respectively.

The next stage is to develop "intelligent" MLP classifiers for flow regime classification. In this case, three different learning algorithms; Levenberg-Marquardt (LM), Quasi-Newton (QN) and Resilient-Backpropagation (RP), have been investigated to determine the most suitable algorithm for the application. These algorithms are utilised because they are the commonly used algorithms for classification application. In doing so, for each learning algorithm, the number of hidden PEs for MLPs is increased (i.e. using the network growing approach), and their CCPs are recorded. The best algorithm is selected based on the highest CCP achieved by the MLP at a certain number of hidden PEs for that algorithm to ensure the network has learned the training process.

## 3. Results and Discussion

The CCPs for both investigations have been obtained as depicted in Table 2. The results show that the tanh-log combination of transfer functions gives higher CCP compared to log-log, although the difference is rather small.

Figure 7, 8 and 9 illustrate the results for CCP versus the number of hidden neurons for test dataset using LM, QN and RP, respectively. Clearly, from the plot in Figure 7, it can be seen that the LM algorithm is capable of training the MLP to obtain 100% CCP starting at 11 hidden PEs. This shows that LM algorithm has the ability to avoid local minima traps, a problem that always occurs in ANN training. Obviously, compared to other two training algorithms, the LM algorithm thus appears to be the best algorithm. This may be because LM algorithm has a robust minimum-searching strategy and the capability to jump out of the local minima (Bishop, 1994). Meanwhile, the highest CCP achieved for both QN and RP algorithms are 99.3%. This shows that the MLP trained with the LM algorithm classifies the test data better than the MLP trained with the QN and RP algorithms.

Table 3 shows the CCP for each of the flow regimes based on the test dataset. The full and empty regimes give 0% CCP because these flow patterns have been included only in the training data, and none in the test data since there is only one flow pattern for each regime. Among the flow regimes, stratified gives the highest CCP of about 43.75%. This could be because stratified regime has the largest number of patterns in the training datasets (612 patterns) and hence, the MLP could classify stratified patterns the best. In addition, it is easier to identify a stratified pattern because the interface of differing permittivity (i.e. gas and oil) can be identified at the edge of the pipe, where the sensitivity of the sensing area is the highest.

The second highest CCP is 43.72%, obtained from the bubble regime, which has the second largest number of patterns in the training dataset.

Although the core and annular regimes have the same number of patterns in the training dataset, their CCP differs due to the differing sensitivity effect they contribute to the ECT sensor. Based on the results, the MLP can identify the core flow patterns better than annular.

## 4. Conclusion

The work proposed a direct method for flow regime classification of gas-oil flows using MLP neural network as the classifier, based on simulated ECT data. In developing the MLP classifier for the task, the LM, QN and RP learning techniques have been investigated to determine the most suitable algorithm for the task of classifying six different flow regimes. The results demonstrated superiority of MLP when trained with the LM algorithm due to a 100% correct classification of unseen test data. Detailed analyses of each flow regime shows that the MLP could best identify stratified flow patterns. This could be because most of the training set consists of ECT data corresponding to this regime. In the future, it would be interesting to investigate the effect of simulated data replication on the MLP performance, particularly for ECT research where only limited number of patterns can be obtained for some classes such as with the

empty and full regimes. Such investigation results could prove useful for other classification application that has bias (i.e. different) numbers of training sets for different classes.

**References**

Beck, M.S. and Williams, R.A. (1996). Process Tomography: A European innovation and its application. *Flow Measurement and Instrumentation*, 7, pp. 215-224.

Bishop, C.M. (1994). Neural Network and Their Application. *Review of Scientific Instruments*, 65(6), pp. 1803- 1832.

Gamio, J.C., Castro, J., Rivera, L., Alamila, J., Garcia-Nocetti, F. and Aguilar L. (2005). Visualisation of gas-oil two-phase flows in pressurised pipes using electrical capacitance tomography. *Flow Measurement and Instrumentation*, 16(2-3), 2005, pp 129- 134.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. (1st ed.). Englewood Cliffs: Prentice Hall, (Chapter 1).

Marashdeh, Q., Warsito, W., Fan, L-S and Teixera F.L. (2006). A nonlinear image reconstruction technique for ECT using a combined neural network approach. *Measurement Science and Technology*, 17, pp. 2097-2103.

Mohamah-Saleh, J. and Hoyle, B.S. (2002). Determination of multi-component flow process parameters based on electrical capacitance tomography data using artificial neural networks. *Measurement Science and Technology*, 13, pp. 1815-1821.

Spink, D.M. (1996). Direct finite element solution for the capacitance, conductance or inductance, and force in linear electrostatic and magnetostatic problems. *The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 15(3), pp. 70-84.

Sterjovski, Z., Dolan, A., Carpenter, K.R, Dunne, D.P. and Norrish, J. (2005). Artificial neural networks for modeling the mechanical properties of steels in various applications. *Journal of Materials Processing Technology*, 170, pp. 536-544.

Ventari, A. and Lampinen, J. (2000). Bayesian MLP neural networks for image analysis. *Pattern Recognition Letters*, 21, pp. 1183-1181.

Xie, C.G., Huang, S.M., Hoyle, B.S., Thorn, R., Lenn, C., Snowden, D. and Beck, M.S. (1992). Electrical Capacitance Tomography for Flow Imaging. *IEE Proceedings – G*, 139(1), pp. 89-98.

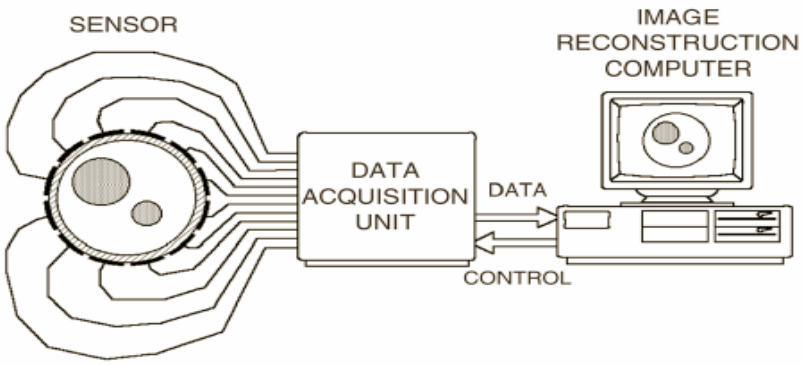



Figure 1. A basic ECT system

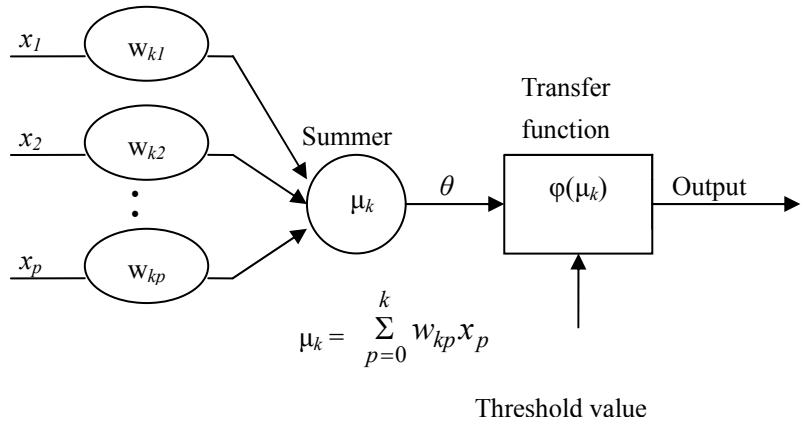



Figure 2. A basic PE model

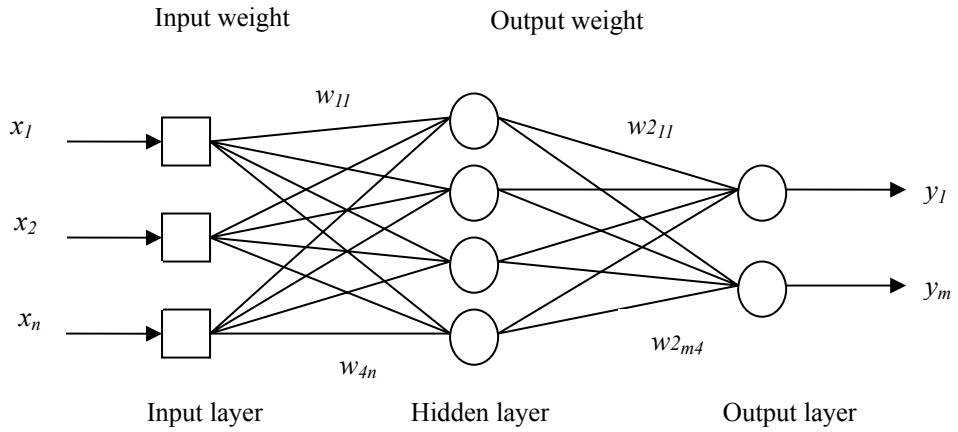



Figure 3. A schematic diagram of a Multi-Layer Perceptron (MLP) neural network

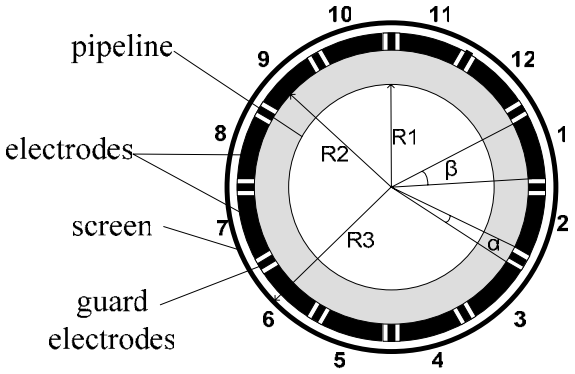
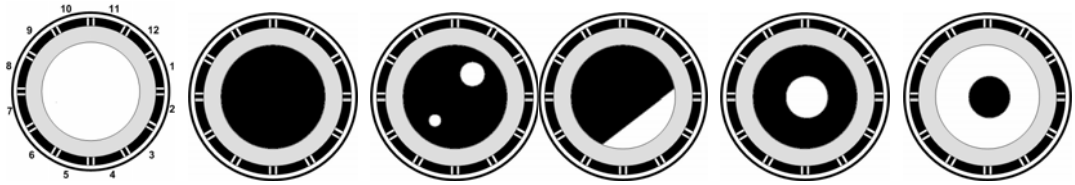



Figure 4. ECT sensor model



(a) (b) (c) (d) (e) (f)

Figure 5. Schematic diagram of the flow regimes to be classified:
(a) empty (b) full (c) bubblen (d) stratified (e) annular (f) core

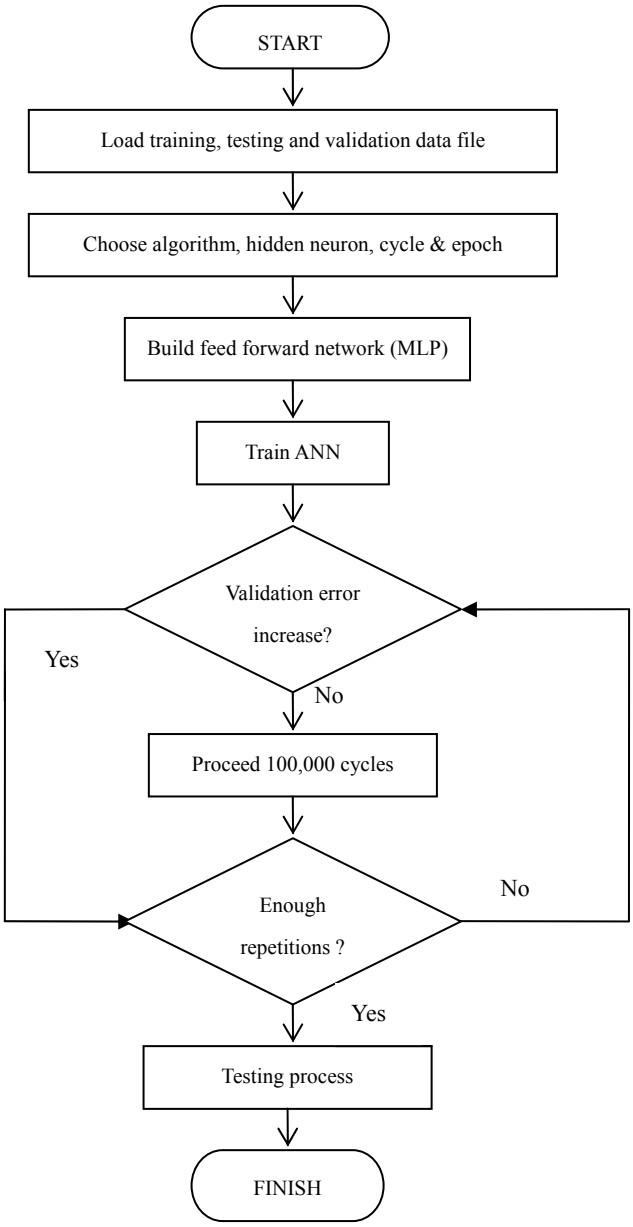



Figure 6. Flowchart of MLP training process

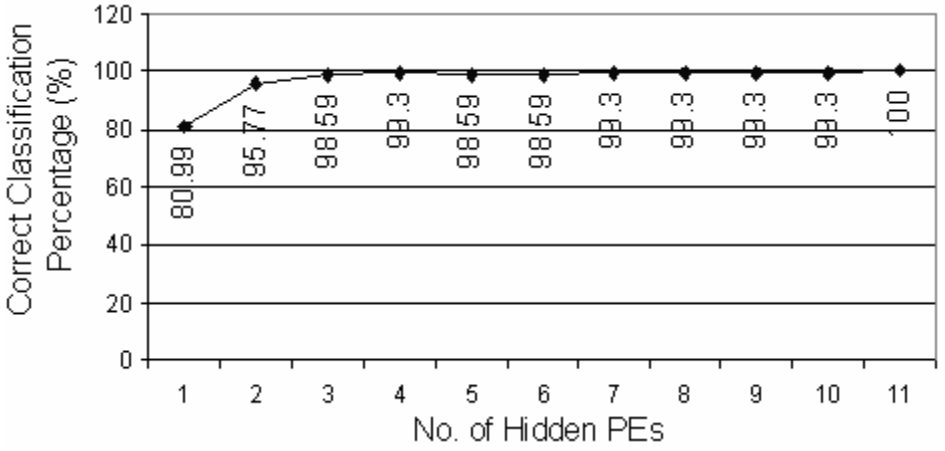



Figure 7. CCP for test dataset based on LM algorithm at various numbers of hidden PEs

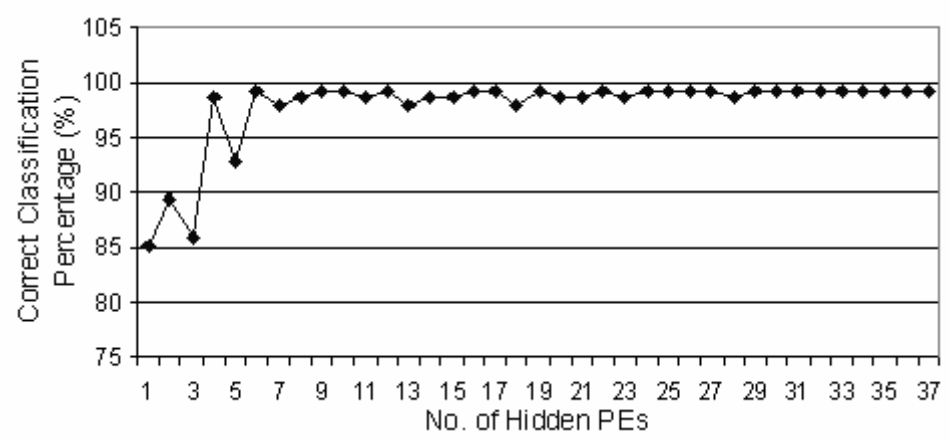



Figure 8. CCP for test dataset based on QN algorithm at various numbers of hidden PEs

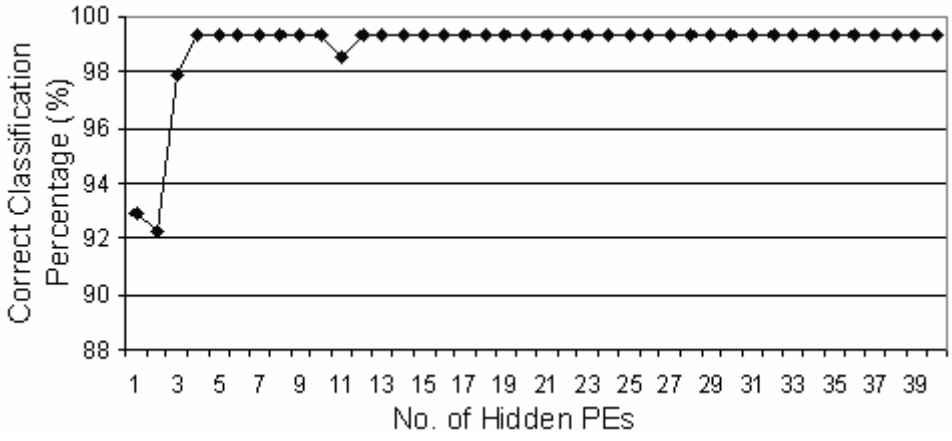



Figure 9. CCP for test dataset based on RP algorithm at various numbers of hidden PEs

Table 1. The numbers of for each flow regime

| Flow Regime | No. flow patterns |
|---|---|
| Full | 1 |
| Stratified | 612 |
| Bubble | 520 |
| Core | 99 |
| Annular | 99 |
| Empty | 1 |
| Total | 1332 |

Table 2. Performance of MLP for different combinations of transfer function

| Combination of transfer function | CCP (%) |
|---|---|
| Log-Log | 97.2 |
| Tanh-Log | 98.6 |

Table 3. CPP of each flow regime for the best MLP based on test datasets

| CCP (%) | | | | | |
|---|---|---|---|---|---|
| Full | Stratified | Bubble | Core | Annular | Empty |
| 0 | 43.75 | 43.72 | 6.79 | 5.74 | 0 |

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# A Study on the Issue of Chinese Character in Ruby on Rails and Relevant Solving Scheme

Ying Li
School of Management, Tianjin Polytechnic University, Tianjin 300387, China
E-mail: liying@smtjpu.net

**Abstract**

As Ruby on rails springs out in China, it wins greater attentions from developers at once. Focusing on the Chinese character issue appeared in the Ruby on rails development frame, this paper puts forward specific solving scheme.

**Keywords:** Ruby, Rails, Chinese character

## 1. Introduction

Although Ruby has already emerged for a period of time, it has not won popular except in Japan before the year 2000 because it does not have the English file. Since the appearance of Rails frame in 2000, the Ruby has become popular in Web development. The Rails is a frame edited by pure Ruby. It serves as a powerful support for Web development, such as data reflection, MVC mode, Web services, and safety. And these functions can be realized easier than similar products. But the development of Ruby on rails is still at its first step in China. Up till now, there is still no perfect scheme for solving the Chinese character issue. It is a critical weakness for domestic programmers who are engaged in developing Chinese interface.

Since we have already determined to use Ruby on rails to finish some projects, we should try our best to achieve this goal. As a matter of fact, most programmers are persistent and never give up. Therefore, focusing on the Chinese character issue emerged in program development, this paper puts forward specific solving scheme, with the hope of being helpful references for all programmers.

## 2. The code of rails

As the base of rails, although the ruby language supports the Unicode character string, its class libraries do not support the Unicode. In other words, in dealing with one Chinese character, Ruby fails to identify that two bytes under the Unicode format mean one character.

### *2.1 The relationship between utf8 and Unicode*

The utf8 means Unicode Translation Format. Unicode has many different formats, such as utf8 and utf16, in transmission. In fact, the utf8 is not a friendly format for Chinese character. Why? In the utf8, one Chinese character has three bytes. But one English ascii character has one bytes. As a result, Chinese characters will occupy a large space and their numbers can not be deduced merely by their bytes. In contrast, the utf16 is better. Each character has two bytes. However, the utf16 and even the utf32 are seldom used. The utf8 is generally recognized.

### *2.2 Converse the code*

If we require for the ‘iconv’ database, we can do:

Iconv. conv (“utf8”, “GBK”, “……”). Turn the characters under the GBK mode into the ones under the utf8 mode. As the source characters have illegal code, we ask iconv to neglect it. Otherwise, iconv will order to stop.

### *2.3 The Chinese characters under the utf8 mode*

Calculate the number of the characters.

In the ruby environment, we can do:

Require ‘jcode’

$KCODE='u' # or $KCODE='UTF8'

They are equal.

Get certain characters.

“How do you do (in Chinese)” [0,1] # random code

“How do you do (in Chinese)”. Scan (/./) [0,1]. join # “you”

**3. The Chinese character issue in the rails database**

Under the common situation, change the MySQL file C:\Program Files\MySQL\MySQL Server 5.0\my.ini. Make two changes. default-character-set=utf8. Then restart MySQL (windows service). This step can also be realized by instance wizard. By this way, the Chinese character issue existed in most rails databases can be solved. But there are still some special situations.

As database mysql uses utf-8 to store Chinese characters, the SQLyog Enterprise shows a series of random codes. Because the Chinese characters are turned into utf8 mode as we store them. And they are still in this mode as they are on the screen. At this time, we can use this order as follow.

Set names ‘gbk’;

Here, Chinese characters are changed into gbk mode from utf8 mode. We can read the Chinese characters at last.

For another instance, if we set source file code as UTF-8 mode, the database content is normal in web pages. But if we require the database content by phpMyAdmin and mysql-front, it shows random code. And under the control platform mode, the database content still shows random code. In practice, we find the solving method as follow.

(1) For MySql: set the character as utf8 mode

(2) Write the following codes into application.rb

before filter: configure charsets

def configure charsets

@response.headers["Content-Type"] = "text/html; charset=utf-8"

# Set connection charset. MySQL 4.0 doesn’t support this so it

# will throw an error, MySQL 4.1 needs this

suppress(ActiveRecord::StatementInvalid) do

ActiveRecord::Base.connection.execute 'SET NAMES UTF8'

end

end

(3) Write following codes into environment.rb

$KCODE = 'u'

require 'jcode'

**4. The random code issue in uploading and downloading files in rails**

For Chinese users, the default code of files is gbk. In other words, the file name in utf8 mode is random code for us. That is why as we upload or download files, we usually come across random code. Therefore, we should converse the random code.

Generally speaking, ruby is not good at providing with support for Unicode, which in a sense blocks the development. Therefore, this issue has to turn to other tools. In general, we are merely familiar with the utf8 mode and the gbk mode. As long as the database, connection way, source code, and page files are in utf8 mode, the work will become easy. But recently as we download files, we still find the random code issue. Then we can use the additional database Iconv in ruby. It is simple.

Iconv. new (to, from)

It is a class method. Generate and back from to the new converter. To and from respectively refers to the after-conversion character and the before-conversion character.

Iconv#iconv(str)

Start to converse characters and return the result.

Firstly, we define two global variable in environment.rb

UTF8_TO_GBK = Iconv.new "gbk", "utf-8"

GBK_TO_UTF8 = Iconv.new "utf-8", "gbk"

Then as we download files, we use “send_file”. The file name can be shown in Chinese character. But we still have to turn the utf8 mode into the gbk mode. We use:

:filename => UTF8_TO_GBK.iconv(filename)

As a result, it shows Chinese character. As far as uploading files with Chinese names is concerned, the main problem is the random code of local files stored in the service. We have to ensure that the local files are in GBK mode as we read or store them.

(1) Uploading files

filename=file.original_filename

File.open("#{RAILS_ROOT}/documents/#{UTF8_TO_GBK.iconv(filename)}", "wb")

By this way, the file names uploaded on the service are shown in Chinese characters.

(2) Downloading files

filename=Document.find(params[:id]).name

send_file(“#{RAILS_ROOT}/documents/#{UTF8_TO_GBK.iconv(filename)}”,:filename=>UTF8_TO_GBK.iconv(filename))

By this way, the file names as we read or store files are shown in Chinese characters.

**5. Conclusion**

Here this paper puts forward a way of solving the Chinese character issue in programming by rails. Comparing with JAVA, Ruby on rails is a new frame and it is still immature. Especially in China, it is still on its initial stage and few articles can serve as valuable reference. As a matter of fact, the Chinese character issue in rails is originated from the unpopularity of Chinese and the lag-behind technology of China. Rails and ruby have taken references from java in many ways. But they sustain their own special advantages. It is believed that more and more people will use rails to write web programs.

**References**

Fan, Kai. (2006). The miracle created by Ruby. *Software World*. No.21.

Fikko. (2005). Ruby on Rails. *Programmer*. No.9.

Meng, Yan. (2005). Ruby in Rails: the wonderful new Web development technology. *Programmer*. No.9.

Meng, Yan & Zhu, Haiyan. (2006). How will Java meet new challenges. *Programmer*. No.11.

Zhang, Xin & Kui, Gang. (2002). Initial Ruby language. *Internet Information World*. No.12.

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# Incorporating the Innovation Attributes Introduced by Rogers' Theory into Theory of Reasoned Action: An Examination of Internet Banking Adoption in Yemen

Ali Hussein Saleh Zolait
Faculty of Business and Accountancy, Department of Business
Sana'a University – Amran University Branch
P.O. Box 41000, Khamer, Republic of Yemen
Tel: 967-01-228795 E-mail: alicomyemen@yahoo.com
Ainin Sulaiman
Faculty of Business & Accountancy, University of Malaya
50603 Kuala Lumpur, Malaysia
Tel: 603-79673800 E-mail: ainins@um.edu.my

**Abstract**

The causal/effect of seven salient beliefs and an individual's attitude and norms, all of which lead to form a person's Behavioural Intention (BI), are not well documented in the context of Internet Banking (IB). The attitudinal belief, represented by five innovation attributes, together with the normative belief, represented by two types of interaction channels, were extracted in accordance with Rogers' (1995) and Ajzen's (1991) theories and literature. The study proposes a conceptual framework of an individual's behavioural intention determinants to adopt IB and tests it using a path analysis of the Ordinary Least Squares (OLS). The results support the argument that attitude, relative advantage/compatibility, observability, ease of use and mass media interaction are the key determinants of BI to use IB.

**Keywords:** Internet Banking, Behavioural Intention, Attitude, Norms, Theory of Reasoned Action, Developing Countries

## 1. Introduction

The consumers' adoption of Internet banking (IB) has received wide attention from researchers in different contexts such as Black et al. (2001), Polatoglu & Ekin (2001), Tan & Teo (2000) among others. Also, incorporating the Theory of Reasoned Action (TRA) and the Diffusion of Innovation (DOI) theory is a new research practice. Therefore, this study examines the determinants of IB in light of both Rogers' (1995) and Ajzen & Fishbein's (1980) theories in an integrated model. Rogers (1995) suggested five important characteristics of an innovation that influences its adoption. These are relative advantage, compatibility, complexity, observability and trialability. Taylor and Todd (1995) utilized these attributes as indicators of attitude in TPB. More to the point, Rogers (1995) suggested two types of interaction channels that accelerate the diffusion of innovation. These are word-of-mouth (WOM) and mass media (MM) interaction. Previous adoption studies that utilized the Theory of Reasoned Action (TRA) used the word-of-mouth referent but not that of mass media (Taylor and Todd, 1995). This study has expanded upon the types of referent describing the normative belief of an innovation, which may be assumed to affect its adoption.

## 2. Literature Review

Internet banking can be defined as the provision of banking services by a bank to its customers over the Internet (Daniel, 1999). In recent years, IB has been one of the major developments in the financial service sector. Floh and Treiblmaiera (2006) reported that over the last five years IB was the fastest growing Internet activity in the U.S. and in Germany the number of online accounts increased almost tenfold, with 40% of all accounts now being online. IB's literature shows that several scholars have used several theoretical models to study IB adoption. This study has two key objectives; firstly, to investigate the factors that influence the adoption of IB and, secondly, to propose an incorporated theoretical method that can be used as a reference for future studies of innovation diffusion in the field of MIS. This study commenced with a revision of the main theoretical frameworks commonly used for analyzing the adoption of

innovations in MIS. These are Innovation Diffusion Theory (Rogers, 1995; Moore & Benbasat, 1991), Theory of Reasoned Action (Fishbein & Ajzen, 1975), Technology Acceptance Model (Davis, 1989; Venkatesh & Davis, 2000), Theory of Planned Behaviour (Ajzen, 1991), and Decomposed Theory of Planned Behaviour (Taylor & Todd, 1995). A comprehensive review of the IS literature on Internet banking (IB) adoption was conducted and research found an absence of studies exploring the adoption of IB in light of Rogers' diffusion of innovation (DOI). In addition, none of the previous studies attempted to identify the prominent predictors utilizing an integrated framework based on human psychology behaviour and innovation characteristics. This study fills this gap by introducing a conceptual framework merging the TRA into Rogers' DOI.

### *2.1 The Theory of Reasoned Action (TRA)*

Ajzen and Fishbein developed the TRA in 1967 and 1980. It is designed to explain human behaviour (Ajzen and Fishbein, 1980) and consists of two factors that affect behavioural intentions; attitude towards behaviour and subjective norms.

#### 2.1.1 Behavioural Intention (BI)

Behavioural intentions are regarded according to Armitage and Christian (2003) as an individual's decision to follow a course of action, as well as an index of how hard people are willing to try and perform the behaviour (Fishbein & Ajzen, 1975). Theoretically, Ajzen & Fishbein (1980), proposed in the TRA that attitudes and subjective norms (SN) affect BI. Accordingly, the influence of attitude on behaviour is mediated through behavioural intentions. Many researchers like Armitage and Christian (2003) use a BI construct as a dependent variable, assuming that intentions are sufficiently predictive of behaviour and consistently lead to behaviour. For instance, the Technology Acceptance Model (TAM) hypothesizes that the actual use of technology is affected by the BI which is itself affected by the attitude towards use. Similarly, in the TRA, the effects of attitude and SN on behaviour are thus mediated by BI (Ajzen and Fishbein, 1980).

#### 2.1.2 Attitude

Attitude had been assumed to be predictive of behaviour in many psychological studies, for instance Armitage and Christian (2003) who defined it as "the individual's overall positive or negative evaluations of behaviour". It is an important determinant in the information system studies which influence the intention to adopt the system. It was proposed by multiple theories including TRA, TPB and also was utilized by Davis' (1993) TAM to examine user acceptance of computer technology. Furthermore Taylor and Todd (1995) employed attitude to understand the usage of information technology which was found to be an influential element for intention behaviour. Hence, attitude seems to be a person's evaluation or general feeling of favourableness or unfavourableness to use Internet banking services.

#### 2.1.3 Subjective Norms

The influence of social environment on BI is the second normative component in the TRA which according to Ajzen and Fishbein, (1980) concern a persons' perception that most people who are important to him think he should or should not use Internet banking services. Pavlou and Chai (2002) related the issue of social influence to Hofstede's dimension of collectivism in which individuals are integrated into groups and form their judgments based on group norms. The normative influence, according to Bearden et al., (1986) occurs when individuals conform to the expectations of others. Similarly, the informational-based normative influence, according to Rogers (1995, p.199) occurs when potential adopters are aware of an innovation and are motivated to try it. Empirically in this study, normative beliefs are determined by indicating, "The extent to which a referent would expect a potential adopter to adopt internet banking. In fact, the TRA built on that specific salient belief influences behavioural perceptions and subsequent actual behaviour (Ajzen & Fishbein, 1980). There are two types of belief in the TRA that affect two perceptual constructs: behavioural beliefs that influence attitudes, and normative beliefs that affect subjective norm. In turn, these two perceptual constructs determine behavioural intentions and actual behaviour. These salient beliefs are discussed in section 2.2 in the light of Rogers (1995) theory.

### *2.2 Rogers' Diffusion of Innovation*

Rogers (1995) theory, introduces four main elements in the diffusion of innovations. They are; innovation attributes, communicated channels, time and social system. The first elements asserted that the rate of adoption of innovations is impacted by five factors: relative advantage, compatibility, trialability, observability, and complexity (Rogers, 1995). Working in an IS context, Moore and Benbasat (1991) examined the influence of these attributes on attitude. All factors except for complexity are generally positively correlated with the rate of adoption. The second element asserting two communication channels, which according to Rogers' (1995) theory will affect the dissemination of the innovation to others, are the interpersonal influence (word-of-mouth) and the mass media that the individual possesses within the 'innovation decision process'.

#### 2.2.1 Factors of Innovation Attributes

In terms of innovation attributes, an individual's perception about the innovation attributes according to Rogers (995)

review is the first main element in the diffusion of innovation. In line with Rogers (995) concepts, Internet banking should show attractive characteristics to customers which in turn lead to maximize the rate of adoption by them. Rogers (995) recommended some characteristics like; (1) Relative advantage, (2) Compatibility, (3) Complexity, (4) Trialability and (5) Observability. Thus, researchers have to evaluate IB characteristics as seen by those people within the social system of the banking industry to understand the trend of adoption and also the adoption rate. Innovation Attributes were utilized by many authors of IS including Moore and Benbasat (1991), Taylor & Todd (1995a, b), Sarel and Marmorstein (2003) to study innovations diffusion in the IS context. Some authors like Black et al. (2001), Polatoglu & Ekin (2001), Tan & Teo (2000), and Al-Sabbag & Mola (2004) have applied Rogers' variables to IB where different models were discussed. To digress, Rogers' (1995) literature proposed that innovations which are perceived by individuals as having greater relative advantage, compatibility, trialability, observability, and less complexity will have a greater adoption rate than other innovations. To look more closely, this study will elaborate and discuss the issue of these attributes in the following sections.

I. Relative Advantage (RA)

In light of Rogers' (1995) review, the RA of IB is defined here as to what extent an individual perceives IB as being better than the idea it supersedes. Advantages of IB is often expressed as effectiveness, time and effort savings, immediacy of the reward or as decrease of discomfort and social prestige. The construct of relative advantage according to Mattila et al., (2003) can be seen differently in the context of different innovations and different consumers. The value of IB arises and is formed from lowering the transaction costs for both customer (i.e. Lichtenstein and Williamson, 2006; Floh and Treiblmaier, 2006) and banks. It is also derived from the globularity of the medium, i.e. individuals can freely transact and get access to their local bank current account when they are overseas. Lichtenstein and Williamson, (2006) reported that the consumer also considers whether the perceived relative advantages of Internet banking, when compared with other forms of banking, outweigh the perceived risks and costs. In this line, Mattila et al., (2003) reported that the relative advantage gained, compatibility of services with adopters existing values, turned out to be the most significant predictors of adoption in IB. Accordingly, Sarel and Marmorstein (2003) pointed out that when both relative advantage and felt need are low, marketers must make efforts to increase the perceived value of the benefits and issues. Perceived usefulness according to the TAM introduced by Davis' (1989) has a direct affect on attitudes. In some research both relative advantage and compatibility compound together and form constructs grouping all the relevant items.

II. Ease of Use vs. Complexity (EOU)

This attribute has many Synonyms like Usability or Complexity and ease of use. It was frequently cited in literature MIS and found closely linked to an individual's perception on the complexity of practicing the introduced innovation. In this line, Complexity, defined by Rogers' (1995) as "how difficult or easy an innovation appears to an individual. Accordingly, some innovations are easy to understand, communicate and use at first glimpse, others are more complicated and require a long time to diffuse. Davis (1989) in his TAM model demonstrated that the perceived ease of use directly affects attitudes. A low level of complexity or a high level of ease of use lead to higher adoption rates (Rogers, 1995). In other words, complexity increases rejection rates (Rogers, 1995; Sarel and Marmorstein, 2003). Rogers illustrated a negative relationship between complexity and adoption rates. Additional diffusion studies confirm the relationships posited by Rogers. For instance, Tan and Teo's (2000) study demonstrated a negative and insignificant relationship.

III. Compatibility (COM)

Compatibility concerns whether or not IB, as an innovative channel, is compatible with the individual's values and experiences. In this line, innovation is more likely to be adopted when it is compatible with individuals (Rogers, 1995). This argument was supported by the meta-analysis of innovation adoption conducted by Tornatzky and Klein (1982). In previous studies, compatibility appears to have a significant impact on willingness to adopt (i.e. Sarel and Marmorstein, 2003). In this study, respondents were asked about three IB values addressing the assumption of whether IB fits their work; style, mechanism and preferences.

IV. Observability (OBS)

Observability of an innovation like IB according to Rogers (1995) describes the extent to which IB is visible for others to observe and communicate the benefits. However, this definition, in the context of IB, will be considered cautiously because observability of IB might turn to privacy and security issues. Of course, some banking innovations like ATMs which can be seen on the street, or in hyper markets and stores may make this technology more observable than Internet banking which is conducted (indoors) inside one's office or home. In the USA, Kolodinsky and Hogarth (2001) examined the adoption of four electronic banking methods, by which they found observability is only associated with an increased probability of adopting phone banking. In our case observability also describes the degree to which the service can be observed being successfully used (Lichtenstein and Williamson, 2006).

V. Trialability (TR)

Rogers (1995) argues that potential adopters who are allowed to experiment with an innovation will feel more comfortable with it and are more likely to adopt it. Sometimes, trialability according to Kolodinsky and Hogarth (2001) provides customers the ability to evaluate innovation benefits. Consequently, if consumers are given the opportunity to try the innovation certain fears of the unknown and the inability to use can be reduced. In respect to IB, Tan and Teo's (2000) study of Internet users also supported the importance of trialability. Similarly, Chung and Paynter (2002) found that lack of prior use of IB inhibited consumer adoption.

2.2.2 Factors of Rogers' Communication of Diffusion

The second main element in the diffusion of innovation is the communication channel (Rogers, 1995). A communication channel is the means by which individuals learn about Internet Banking. A second area of research involved how norms affect diffusion. In previous studies, mass media channels were more effective in creating knowledge of innovations, whereas word-of-mouth channels were more effective in forming and changing attitudes toward a new idea, and thereby influencing the decision to adopt or reject a new idea. Most individuals evaluate an innovation, not on the basis of scientific research by experts, but through the subjective evaluations of near-peers who have adopted the innovation. This study will examine the potency of the mass media and the word-of-mouth communication in encouraging the diffusion of an innovation by affecting a persons' subjective norms (Zolait & Ainin, 2008).

I. Word-of-Mouth (WOM)

It was argued that in word-of-mouth learning, not only do people learn from a small number of people but that these people also tend to be closer to them (in some sense) than the average person in the population. This is what is called "learning from neighbours (Banerjee and Fudenberg 2004). In this study learning by word-of-mouth communication stands for the logic that, online banking will not be viewed by most respondents as an exciting innovation. Therefore; word-of-mouth learning and communicating IB is assumed to increase the adoption rate of IB. In a previous study on marketing online banking services conducted by Sarel and Marmorstein, (2003) they highlighted that poor word-of-mouth communication contributed to the weak adoption rate. This area presents one of the most critical obstacles to adoption. This study looks into the word-of-mouth by addressing the influence of peer, family, friend and bank's staff on an individual's overall Subjective Norm. Furthermore, Rogers (2003) highlighted that interpersonal communications (word-of-mouth) provide a more effective means of persuading individuals of the benefits of a new innovation.

II. Mass Media (MM)

Mass Media is referred to here as a means of public communication which reaches a large audience. Kreps and Thornton (1992) pointed that media extends people's ability to communicate, to speak to others far away, to hear messages, and to see images that would be unavailable without media. Rogers (2003) reported that, "mass media channels are usually the most rapid and efficient means of informing potential adopters about the existence of an innovation - that is, to create awareness-knowledge". In other words, mass media's most powerful effect on diffusion is that it spreads knowledge of innovations to a large audience rapidly (Rogers, 1995 p.285). Khalifa & Cheng (2002) and Zolait & Ainin (2008) argued that the media, as a source of social influence, can play an important role in the individual's intention formation and it also contributes to exposure.

## 3. Methodology

In terms of Conceptual Framework, adoption can be conceptualized as behavioural responses of individuals to two motivated forces of attitudinal and normative beliefs. As such, the TRA by Ajzen (1991) suggests that BI can be predicted from the individual attitude and subjective norms. In turn, attitude can be predicted from Rogers' (1995) five attributes of innovation as found in the literature of Taylor (1995). These are relative advantage (RA), Ease of Use (EOU), Observability (OBS), Compatibility (Com), and Trialability (TR). In addition, SN can be predicted using the type of communication channel by which individuals interact with the introduced innovation, identified by Rogers (1995) as word-of-mouth and mass media. The particular theoretical perspective adopted here is from TRA and DOI (Ajzen, 1991; Rogers, 1995; Taylor, 1995 and Moore & Benbasat, 1991). The conceptual framework is shown in figure 1 below;

*Study Design*, In order to operationalize and test the proposed conceptual framework a multi-phase research design was adopted. First, literature was reviewed and the data relevant to potentially significant variables was collected. Second, an exploratory factor analysis (FA) was performed on the underlying factor structure of variables. Then the content of factors and items loading was analysed to ensure content, construct, and criterion validity and reliability of factors extracted. Third, the proposed framework was operationalized into a testable model and hypotheses pertaining to the relationships in model variables were developed and tested using regression. Finally, procedures of path analysis approached the Ordinary Least Square (OLS) were performed to assess the overall fit of the model.

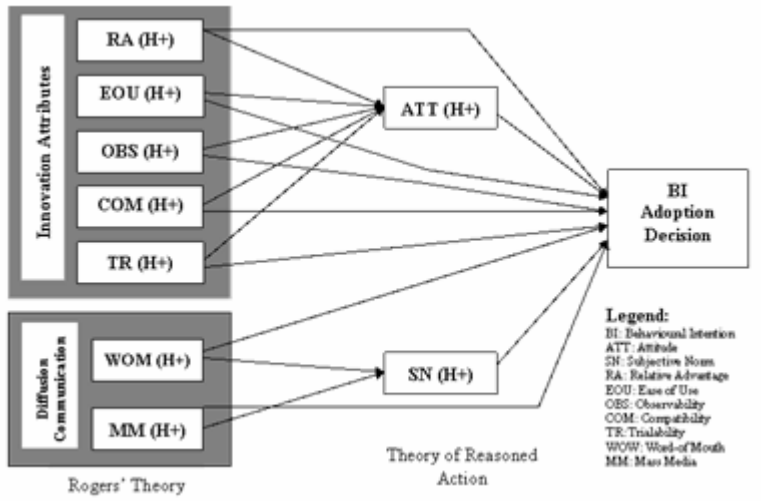



Figure 1. Conceptual Framework of Behavioural Intention Determinants

*Instrument*, this study utilized previous studies of adoption in developing the appropriate instrument for the data collection. The aim was to ensure the face validity of the scales intended to measure these variables. The final, refined instrument consists of two groups of variables. The first group addressed Rogers' variables with seven variables intended to measure an individual's attitudinal belief and normative beliefs formed by an individual's perception on IB attributes and interaction via both the WOM and MM diffusion channel. The second group deals with the TRA main variables intended to measure, intention, attitude, and subjective norms. Both measures used a 7-point Likert scale. Face and construct validity were established during the adaptation and factorability procedures. To mitigate the responses bias pre-tested questionnaires were not included in the final instrument.

*Research Hypotheses*, the behavioural intention (BI) to use IB is determined by two theoretical constructs following Ajzen's (1991) TRA which are individuals' attitude and subjective norm. This is in the direct relationship of causal and effect. In addition, the indirect causal and effect relationships link both IB characteristics as behavioural belief and the two types of normative belief of communication channel to the BI construct. Therefore, it is expected that attitude is influenced positively by IB characteristics as well as the subjective Norm by referents' channel. The relationships aforementioned in the conceptual framework are summarized in the following hypotheses:

$H_1$: Individual's intention to use Internet banking increases as

A) Individual's attitude on IB increases;

B) Individual's subjective norm on IB increases;

C) Individual's perception on relative advantage/compatibility of IB increases;

D) Individual's perception on the ease of use of IB increases;

E) Individual's perception on IB trialability increases;

F) Individual's perception on IB observability increases;

G) Individual's interaction about IB through mass media increases;

H) Individual's interaction about IB through word-of-mouth increases;

$H_2$: Individual's attitude towards using Internet banking increases as

I) Individual's perception on relative advantage/compatibility of IB increases;

J) Individual's perception on the ease of use of IB increases;

K) Individual's perception on IB trialability increases;

L) Individual's perception on IB observability increases;

$H_3$: Individual's subjective Norm on using Internet banking increases as

M) Individual's interaction through mass media increases;

N) Individual's word-of-mouth interaction increases

**4. Sample Plan and Sample Profile**

Customers of banks in Yemen are the population of this study, although it is not possible to clearly identify the total population in the banking field. Also because of time and convenience, the 369 respondents are selected randomly as

the research sample. The sample in this study targeted customers who are categorized as holders of bank accounts. 1000 questionnaires were issued and self-administrated to 17 bank headquarters in Sana'a city. There were 369 valid questionnaires returned and the response rate was 52 %. The sample profile of the respondents is shown in Table (1)

Table 1. Respondents Demographic Profile

| Variable | Value | Freq. | % |
|---|---|---|---|
| Gender | Male | 302 | 81.8 |
| | Female | 67 | 18.2 |
| Age | Twenties (19-29 Year) | 135 | 36.6 |
| | Thirties (30-39 Year) | 147 | 39.8 |
| | Forties (40-49 Year) | 74 | 20.1 |
| | Older (=> 50 Year) | 13 | 3.5 |
| Marital Status | Single | 86 | 23.3 |
| | Married with children | 228 | 61.8 |
| | Married without children | 55 | 14.9 |
| Nationality | Yemeni | 350 | 94.9 |
| | Non-Yemeni | 19 | 5.1 |
| Resident Area | Sana'a Area | 290 | 78.6 |
| | Other Areas | 79 | 21.4 |
| Personal Income | Less than 30001 Y.R | 55 | 14.9 |
| | 30001-60000 Y.R | 111 | 30.1 |
| | 60001-120000 Y.R | 140 | 37.9 |
| | 120001-180000 Y.R | 27 | 7.3 |
| | Above 180001 Y.R | 36 | 9.8 |
| Profession (Job) | Managerial work | 132 | 35.8 |
| | Clarks | 65 | 17.6 |
| | Specialists | 43 | 11.7 |
| | Technicians | 31 | 8.4 |
| | Agricultures | 5 | 1.4 |
| | Engineers | 27 | 7.3 |
| | Handcraft | 5 | 1.4 |
| | Simple professional | 12 | 3.3 |
| | Other | 49 | 13.3 |
| Total | | 369 | 100.0 |

| Variable | Value | Freq. | % |
|---|---|---|---|
| Sector | Public sector | 91 | 24.7 |
| | Private sector | 216 | 58.5 |
| | Individual business | 62 | 16.8 |
| Education | Preparatory level & < | 31 | 8.4 |
| | Secondary & diploma | 86 | 23.3 |
| | Undergraduate | 203 | 55.0 |
| | Postgraduate & Professional | 49 | 13.3 |
| Residence Ownership | Own | 154 | 41.7 |
| | Family house | 63 | 17.1 |
| | Own with mortgage | 12 | 3.3 |
| | Rent | 126 | 34.1 |
| | Given for services | 12 | 3.3 |
| | Others | 2 | .5 |
| Business Nature | Manufacturing | 28 | 7.6 |
| | Services | 83 | 22.5 |
| | Government | 26 | 7.0 |
| | Commercial | 99 | 26.8 |
| | Banking & Finance | 127 | 34.4 |
| | Others | 6 | 1.6 |
| Household Income | less than 40001 | 31 | 8.4 |
| | 40001- 80000 Y.R | 90 | 24.4 |
| | 80001-120000 Y.R | 78 | 21.1 |
| | 120001- 160000 Y.R | 66 | 17.9 |
| | 160001-200000 | 27 | 7.3 |
| | 200001-240000 Y.R | 29 | 7.9 |
| | Above 240001 | 48 | 13.0 |
| Total | | 369 | 100.0 |

**5. Data Analysis**

An explanatory factor analysis with Factor Axis and varimax rotation was performed to ensure the discriminant convergent validity. The Table displayed variables belonging to the same factor grouped together to form the operational factor aforementioned in the framework. Particularly, each factor items were examined cautiously; only items with consistent meaning were retained for measuring the factor while other items deemed not reliable were excluded from further analysis. Because of this overriding concern with the interpretability of the factors, the analysis suggested that some factors must be purified accordingly. The items dropped from their respective factors were: (EOU05), (EOU06) and (OBS07). The study's instrument and the purified factors are displayed in Table 2 and 3.

Group 1

Table 2. Coding, Items, and Reliability Test of Behavioural Belief Constructs

| Factor included | / Items | Coefficient Alpha | Reference |
|---|---|---|---|
| Attitude (ATT) | | 0.91 | (Fishbein & Ajzen,1975; Ajzen & Fishbein, 1980) |
| ATT01: | IB services are a good idea. | | |
| ATT02: | IB is a wise idea. | | |
| ATT03: | I like the idea of using the IB services. | | |
| ATT04: | Using the IB services would be a pleasant experience. | | |
| Relative advantage (RA) | | 0.93 | (Moore & Benbasat, 1991; Karahanna et al. (1999) |
| RA1: | IB would enable me to accomplish my tasks more quickly | | |
| RA2: | IB would improve the quality of my work | | |
| RA3: | IB would enhance my effectiveness on the job | | |
| RA4: | IB would make my job easier | | |
| RA5: | IB gives me greater control over my work | | |
| Complexity (EOU) | | 0.93 | (Moore & Benbasat, 1991; Karahanna et al., 1999; Tan & Teo, 2000; Wang et al., 2003) |
| EOU 1: | Learning to operate IB would be easy for me | | |
| EOU 2: | Overall, If I were to use IB, it would be easy to use | | |
| EOU 3: | It would be easy for me to become skilful at using IB. | | |
| EOU 4: | I believe that it is easy to get IB to do what I want it to do. | | |
| Compatibility (COMPT) | | 0.92 | Benbasat, 1991; Karahanna et al., 1999; Tan &Teo, 2000) |
| COM1: | IB would be compatible with most aspects of my work. | | |
| COM2: | IB would fit my work style | | |
| COM3: | IB would fit well with the way I like to work. | | |
| Trialability (TR) | | 0.88 | Moore & Benbasat, 1991; Karahanna et al. 1999; Tan&Teo 2000; Brown, et al. (2004 ) |
| TR01:<br>TR02:<br>TR03: | I want to be able to use IB on a trial basis.<br>I want to be able to properly try out IB.<br>I want to be permitted to use IB, on a trial basis long enough to see what it can do. | | |
| Observability (OBS) | | 0.79 | Karahann, et al., |

| | | | |
|---|---|---|---|
| OBS1: | I will use IB when many use it. | | (1999) |
| OBS2: | I will use IB when I have seen others using IB. | | |
| OBS3: | I will use IB as soon as I get to know about it. | | |
| OBS4: | I will use IB if this service becomes popular. | | |
| OBS5: | I will wait until other customers start to use IB. | | |
| OBS6: | I will use IB when other people have successful experience of using it. | | |

Group 2

Table 3. Coding, Items, and Reliability Test of Normative Belief Constructs

| Factor included | Items | Coefficient Alpha | Reference |
|---|---|---|---|
| Subjective Norm (SN) | | 0.93 | |
| SN1 | Most people who are important to me would think that I should use IB to get bank services | | Taylor & Todd (1995b) Shih & Fang (2004) |
| SN2 | The people who influence my decisions would think that I should use IB. | | |
| SN3 | Most people who are important to me would think that I should try out the bank's website to get access to the bank IB. | | |
| SN4 | The people who influence my decisions would think that I should try out the bank's website to get access to the bank | | |
| SN5 | Most people who are important to me would think that using IB is a good idea. | | |
| SN6 | Most people who are important to me would think I should use IB. | | |
| Personal Norms (PR) | | 0.94 | |
| (MCPER1)* | Peers /colleagues think I should use IB and I will do what peer/colleagues suggest I do. | | |
| MCPER2* | Peers/colleagues think I should try out IB and I will do what peer/colleagues suggest I do. | | |
| MCLEDR3* | Opinion leaders think I should use IB and I will do what leaders suggest I do. | | |
| MCLEDR4* | Opinion leaders think I should try out IB and I will do what leaders suggest I do. | | |
| MCEMPLY* | Bank's employees think I should use IB and I will do what bank's people suggest I do. | | |
| MCEMPLY6* | Bank's employees think I should try out IB and I will do what bank's people suggest I do. | | |
| Media Norms MM | | 0.86 | |
| MCMEDIA1* | Media suggests using IB is good idea and I will do what the media suggest. | | |
| MCMEDIA2* | Media consistently recommend using IB services and I will do what the media suggest. | | |
| MCPRFS3* | For my profession, it is advisable to use Internet Banking services and I will do what it suggests. | | |
| MCMEDIA3 | I read /saw news reports that using IB is a good way of managing my | | |

| | | | |
|---|---|---|---|
| * | bank account and I will do what this media suggest. | | |

** Normative Belief measured using Theoretical approach (a belief-based measure)*

*5.1 Factor Analysis*

To test factorability and reliability of research constructs, the purification procedures of Factor Analysis (FA), item to total correlation and Cronbach's Alpha analysis were performed in this study. Two groups of FA were conducted and the results displayed in Tables 4 and 5. Factors with an eigenvalue greater than 1.0 were retained and the cut-off value of factor loading was greater than 0.5. Accordingly, the set of items comprising IB attributes (Behavioural Belief) construct were subjected to Principal Factor Analysis (PFA) and the solution was rotated using Varimax criterion. Table 4 showing the result of PFA, reveals five distinctive factors underlying an individuals' behavioural belief with respect to the use of IB.

Group 1

Table 4. PFA Result: Factors Underlying Behavioural Belief of IB

| | Factor | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| RA01 | .653 | | | | |
| RA02 | .740 | | | | |
| RA03 | .780 | | | | |
| RA04 | .761 | | | | |
| RA05 | .760 | | | | |
| COM01 | .654 | | | | |
| COM02 | .770 | | | | |
| COM03 | .776 | | | | |
| OBS01 | | .707 | | | |
| OBS02 | | .739 | | | |
| OBS04 | | .750 | | | |
| OBS05 | | .732 | | | |
| OBS06 | | .764 | | | |
| EOU01 | | | .744 | | |
| EOU02 | | | .810 | | |
| EOU03 | | | .747 | | |
| EOU04 | | | .535 | | |
| ATT1 | | | | .701 | |
| ATT2 | | | | .663 | |
| ATT3 | | | | .735 | |
| ATT4 | | | | .654 | |
| TRA01 | | | | | .773 |
| TRA02 | | | | | .811 |
| TRA03 | | | | | .710 |
| OBS03 | | | | | |
| Eigenvalue | 10.749 | 4.409 | 1.384 | 1.156 | 1.065 |
| Variance explained | 42.995 | 17.635 | 5.535 | 4.623 | 4.258 |
| Cronbach's Alpha | 0.94 | 0.87 | 0.93 | 0.91 | 0.88 |
| (a) Total Variance Extracted by the five factors 75 %; KMO = 0.928; Barlett's Test <.001<br>(b) Extraction Method: : Principal Axis Factoring;<br>(c) Rotation Method: Varimax with Kaiser Normalization. | | | | | |

Items RA01, RA02, RA03, RA04, RA05, COM01, COM02 and COM03 loaded on what were named as

"advantageous". It is obvious from the loading that the aforementioned items are highly correlated with this factor. This solution is in agreement with previous studies conducted by authors including Moore & Benbasat, 1991; Taylor & Todd, (1995 a,b); Mattila, (2003) and Tornatzky & Klein (1982). Item OBS03 did not appear in the rotated matrix as it is not related to any construct. Similarly, the set of items comprising Normative Belief construct were subjected to PFA and the solution was rotated using Varimax criterion. Table 5 shows the result of FA, which reveals three distinctive factors underlying an individual's normative belief.

Groups 2

Table 5. PFA Result: Factors Underlying Normative Belief of IB

| | Factor | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| SN01 | .769 | | |
| SN02 | .778 | | |
| SN03 | .787 | | |
| SN04 | .779 | | |
| SN05 | .598 | | |
| SN06 | .717 | | |
| MCPER1 | | .724 | |
| MCPER2 | | .681 | |
| MCLEDR3 | | .794 | |
| MCLEDR4 | | .802 | |
| MCEMPLY5 | | .718 | |
| MCEMPLY6 | | .654 | |
| MCMEDIA1 | | | .840 |
| MCMEDIA2 | | | .838 |
| MCPRFS3 | | | .520 |
| MCMEDIA3 | | | .818 |
| Eigenvalue | 9.180 | 1.699 | 1.301 |
| Variance explained | 57.372 | 10.617 | 8.133 |
| Cronbach's Alpha | 0.93 | 0.94 | 0.86 |
| (a) Total Variance Extracted by the three factors 76 %; KMO = 0.923; Barlett's Test <.001<br>(b) Extraction Method: Principal Axis Factoring;<br>(c) Rotation Method: Varimax with Kaiser Normalization. | | | |

Items SN01, SN02, SN03, SN04, SN05 and SN06 loaded on what Ajzen (1985) named as the "subjective norm". It is obvious from the loading that the aforementioned items are highly correlated with this factor. Items MCPER1, MCPER2, MCLEDR3, MCLEDR4, MCEMPLY5, and MCEMPLY6 discriminate themselves and converged in what was named as "word-of-mouth referents". It is obvious from the loading that items are highly correlated with this factor. Lastly, items MCMEDIA1, MCMEDIA2, MCPRFS3, and MCMEDIA3 were loaded on the study named "mass media referents".

*5.2 Path Analysis*

A path analytical approach using the Ordinary Least Squares (OLS) technique was utilized to test the proposed model as recommended by Cohen & Cohen (1983) and is shown in Figure (2). Furthermore, to test for mediation, Baron and Kenny (1986) proposed a four step approach in which several regression analyses were conducted and significance of the coefficients was examined at each step. A series of multiple regression and correlation operations (see Appendix A) were performed due to the specification of the operational model shown in figure (1). The regression beta weights being used as the estimate of the path coefficients**.**

Table 6. Results of Multiple Linear Regression: BI as Dependent Variable

| Predictor Variable | Unstandardised Coefficients | | Standardised Coefficients | t |
|---|---|---|---|---|
| | B | Std. Error | Beta | |
| IV1 - ATT | .763 | .056 | .571 | 13.630* |
| IV2 - SN | .031 | .033 | .041 | .943 |
| IV3 - RAC | .095 | .033 | .139 | 2.885** |
| IV4 - OBS | -.080 | .032 | -.082 | -2.481** |
| IV5 - EOU | .184 | .051 | .158 | 3.616* |
| IV6 - TR | -.046 | .059 | -.025 | -.773 |
| IV7 - WOM | .000 | .004 | .002 | .053 |
| IV8 - MM | .012 | .007 | .066 | 1.742*** |
| R: | .857 | | | |
| $R^2$: | .735 | | | |
| Adjusted $R^2$: | .729 | | | |

| | Analysis of Variance | | | |
|---|---|---|---|---|
| | DF | Sum of Squares | Mean Square | F | Significance of F |
| Regression | 8 | 17586.615 | 2198.327 | 124.577 | .000 |
| Residual | 360 | 6352.691 | 17.646 | | |

*$P$ <.001, **P <.05, ***P < .10

The relationships among the variables in the recursive model depicted in series equations as follows;

$X_{RAC} = e_{RAC}$

$X_{OBS} = e_{OBS}$

$X_{EOU} = e_{EOU}$

$X_{TR} = e_{TR}$

$X_{PR} = e_{PR}$

$X_{MM} = e_{MM}$

$X_{ATT}$ = P ATT RAC XRAC + P ATT OBS XOBS + P ATT EOU XEOU + PATT TR XTR + $e_{ATT}$

$X_{SN}$ = PSNPR XPR + P SN MM XMM + $e_{SN}$

$X_{BI}$ = P BI RAC XRAC + PBIRXR + PIAXA + PA2X2 + PI3X3 + PN5X5 + PINXN + PI6X6 + PI7X7 + PI8X8 + PI9X9 + PICXC + PCI10X10 + $e_I$

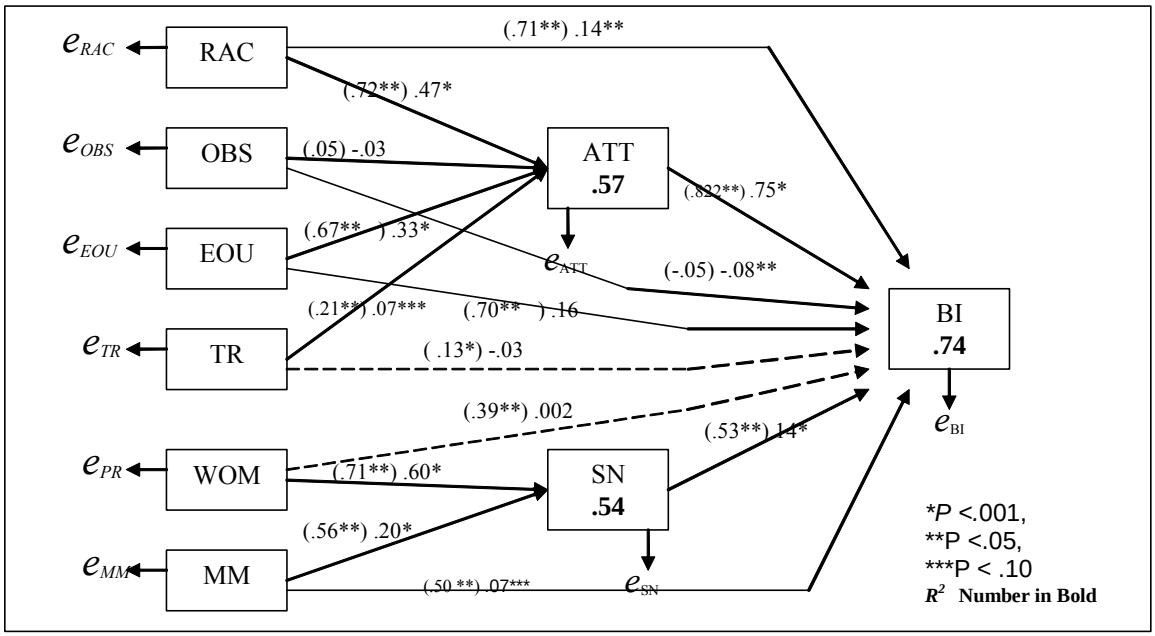



Figure 2. Full Effect Model of Causal Path Findings via OLS

Numbers in Parenthesis indicate zero-order correlation and other numbers are path coefficients.

## 6. Findings & Discussion

The path coefficient reveals that different factors also exert indirect influences on behavioural intention through either attitude or subjective norm. First, Rogers' (1995) attributes in Internet banking is viewed as a salient behavioural belief that RAC, EOU, OBS and TR directly influences a customer's attitude, and indirectly with exception of TR affects a customer's behavioural intentions to use IB services. Secondly, Rogers' (1995) diffusion channel in IB viewed as a salient normative belief that WOM and MM directly influences a customer's subjective norm. Also the mass media channel exerted a positive effect on an individual's intention at $P < .10$ while the word-of-mouth showed an insignificant effect on BI. Thirdly, an individual's Subjective Norm of IB is related strongly to the individual's mass media based-interaction compared to the individual's word-of-mouth interaction.

Can IB attributes from Rogers' theory of innovation be linked empirically to an individual's intention to adopt IB? This study of 369 bank customers provides a positive answer to this research question, when the intention is considered as a dependent variable and the behavioural and normative beliefs derived from Rogers' adoption variables as independent variables for attitude and subjective norms as in the Theory of Reasoned Action. Therefore; the results also support the proposed conceptual framework that intention can be explained clearly by behavioural and normative beliefs variables. All the paths proposed by the integrated model were supported with the exception of insignificant links of both TR and WOM to the BI. Also, all the hypotheses were supported except for the hypotheses linking TR and WOM to the BI. Therefore; Determinants of Individuals' Intention to Use IB, seem to be the research variables of attitude and subjective which have a direct influence on intention to use IB. This key finding supports the argument. In contrary to Tan and Teo's (2000) study, the relationship between perceived EOU of using Internet banking services and both attitude and intentions to adopt such services was supported. Tan and Teo's (2000) claimed that the insignificant result is due more to the sample's characteristics of Singapore Internet users rather than the inappropriateness of the measure. The Determinant of Individuals' Attitude to use IB, Result that was not expected is the moderate and inverse relationship between the individual's attitude and IB observability as well as IB trialability. This finding points to the existence of a more complex relationship. Findings show that, enabling the observability attribute for innovations like IB is not desirable if the intention is to increase the adoption rate. One explanation could be due to the sensitive nature of banking and specifically IB. The second explanation is that when potential adopters are given the chance to observe IB functions, they become inverse thinking on whether IB is secure or not. They may also be concerned with privacy issues. Determinants of Individuals' SN to use IB, Research has shown that the concrete person's Subjective Norm is developed through communication exchanges about the innovation with through word-of-mouth (i.e. peers, staff and opinion leaders more than through Mass Media. In line with Sarel and Marmorstein's (2003) study, banks need to examine current communication tactics and identify more effective ways to communicate benefits of IB. New approaches to address these problems need to be considered.

## 7. Generalizability

The data set of the sample split into two samples, the hold-out subsample (192 cases) and the analysis subsample (177

cases). The purpose of validation analysis is to test the generalizability of the regression analysis Model to the population represented by the sample in the analysis.

Table 7. Split Sample Validation Analysis: Validating Regression Results (Determinants and Models)

| Variable Entered | Full Model Sample (n=369) | | | Sample 1 Split = 1(n=192) | | | Sample 2 Split = 0 (n=177) | | |
|---|---|---|---|---|---|---|---|---|---|
| DV – Intention | Beta | t | *p*. | Beta | t | *p* | Beta | t | *p* |
| *F* | | 252.10 | .000 | | 61.956 | .000 | | 127.14 | .000 |
| (Constant) | | 1.304 | .193 | | .578 | 564 | | 1.401 | .163 |
| IV1 - ATT | .571 | 13.630 | .000 | .531 | 8.842 | .000 | .632 | 10.362 | .000 |
| IV2 - SN | .041 | .943 | .347 | .066 | 1.021 | .309 | .002 | .031 | .975 |
| IV3 - RAC | .139 | 2.885 | .004 | .167 | 2.425 | .016 | .102 | 1.476 | .142 |
| IV4 - OBS | -.082 | -2.481 | .014 | -.087 | -1.896 | .060 | -.075 | -1.531 | .128 |
| IV5 - EOU | .158 | 3.616 | .000 | .126 | 1.918 | .057 | .189 | 3.164 | .002 |
| IV6 - TR | -.025 | -.773 | .440 | -.004 | -.085 | .933 | -.052 | -1.134 | .258 |
| IV7 - PR | .002 | .053 | .958 | -.011 | -.168 | .867 | .015 | .256 | .799 |
| IV8 - MM | .066 | 1.742 | .082 | .096 | 1.774 | .078 | .035 | .640 | .523 |
| Summary Table | | | | | | | | | |
| *Multiple R* | .86 | | | .86 | | | .86 | | |
| $R^2$ | .74 | | | .73 | | | .75 | | |
| *Adjusted* $R^2$ | .73 | | | .72 | | | .73 | | |
| *SE* | 4.20 | | | 4.31 | | | 4.14 | | |

## 8. Conclusion

As a topic for further research this study concurs with Ajzen (1991) who encourages the exploration of additional variables and regards the theory of planned behaviour as "open to the inclusion of additional predictors if it can be shown that they capture a significant proportion of the variance in intention or behaviour after the theory's current variables have been taken into account" (p. 199).

The primary benefit of this study is as a contribution to knowledge in the area of diffusion of innovation in developing countries. It emerged there is a need to incorporate the attributes of innovation together with the channel by which these attributes are communicated to the social network.

"We make a living by what we get, but we make a life by what we give"

Sir Winston Churchill

## References

Ajzen, I. (1991). The theory of planned behaviour, Organizational Behaviour and Human Decision Processes, 50, 179–211.

Ajzen, I. (1985), From Intentions to Action: A Theory of Planned Behaviour, in: Kuhl, J. & Beckman, J. (Eds.), Action control: From cognitions to behaviours. New York: Springer-Verlag.

Ajzen, I. and Fishbein, M. (1980). Understanding Attitudes and Predicting Social Behaviour, Englewood Cliffs, NJ: Prentice-Hall, Inc.

Al-Sabbagh, I. and Molla, A (2004). Adoption and Use of Internet Banking in the Sultanate of Oman: An Exploratory Study, *Journal of Internet Banking and Commerce*; Web Archive, Vol. 9, No. 2

Armitage, C.J and Christian (2003). From Attitudes to Behaviour: Basic and Applied Research on the Theory of Planned Behaviour. *Current Psychology: Developmental, Learning, Personality, Social*. Fall 2003, Vol. 22, No. 3, pp. 187–195.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology,* 51, 1173–1182.

Banerjee, A. and Fudenberg, D (2004). Word-of-mouth learning, *Games and Economic Behavior*, No 46, pp 1–22.

Bearden, W.O., Calcich, S.E., Netemeyer, R., & Teel, J.E (1986). An Exploratory Investigation of Consumer

Innovativeness and Interpersonal Influences , *Advances in Consumer Research*, Vol. 13, No. 1, pp 77-82.

Black, N. J., Lockett, A., Winklhofer, H. & Ennew, C. (2001). The Adoption of Internet Financial Services: a Qualitative Study, *International Journal of Retail & Distribution Management*, Vol. 29, No, 8 pp. 390-398.

Chung, W. and Paynter, J. (2002). An Evaluation of Internet Banking in New Zealand, IEEE Proceedings of the 35th Hawaii International Conference on System Sciences – HICSS, IEEE, Conference on 7-10 Jan 2002, PP. 2410 – 2419

Cohen, J. & Cohen, P. (1983), Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences, (2nd ed), New Jersey, Lawrence Erlbaum Associates.

Daniel, E. (1999). Provision of Electronic Banking in the UK and the Republic of Ireland, International *Journal of Bank Marketing*, Vol. 17, No. 2, pp. 72-82.

Davis, F. (1993). User Acceptance of Computer technology: System Characteristics, User Perceptions and Behaviour Characteristics. International Man-Machine Studies 38, 475-487.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly*, 13, 319±340.

Fishbein, M. and Ajzen, I (1975). Belief, attitude, intention, and behaviour: An introduction to theory and research. Reading, MA:Addison-Wesley.

Floh,A and Treiblmaier, H (2006). What Keeps The E-Banking Customer Loyal? A Multigroup Analysis of the Moderating Role of Consumer Characteristics on E-Loyalty in the Financial Service Industry, *Journal of Electronic Commerce Research*, Vol. 7, No.2, pp 97-110.

Khalifa, M and Cheng, S (2002). Adoption of Mobile Commerce: Role of Exposure, Proceedings of the 35th Hawaii International Conference on System Sciences – 2002, [Online] Available: http://www.hicss.hawaii.edu/HICSS_35/HICSSpapers/PDFdocuments/CLWMC01.pdf (Jan 10, 2006)

Kolodinsky,J and Hogarth, J.M (2001). The Adoption of Electronic Banking Technologies by American Consumers, *Consumer Interests Annual*, Vol. 47, [Online] Available: http://www.consumerinterests.org/files/public/Kolodinsky,_Hogarth.pdf (August 20, 2006)

Kreps, G. L., & Thornton, B. C. (1992). Health Communication Theory & Practice, Prospect Heights, IL: Waveland Press.

Lichtenstein, S. & Williamson, K. (2006). Consumer Adoption of Internet Banking: Understanding Consumer Adoption of Internet Banking An interpretive study in the Australian Banking context, *Journal of Electronic Commerce Research*, Vol. 7, No.2, pp.50-66

Mattila, M., Karjaluoto, H., & Pento, T. (2003). Internet Banking Adoption among Mature Customers: Early Majority or Laggards, *Journal of Services Marketing*, Vol. 17, No 5 pp. 514-528

Mattila, M (2003), Factors Affecting the Adoption of Mobile Banking Services, *Journal of Internet Banking and Commerce*, Vol. 8, No. 1, [Online] Available: http://www.arraydev.com/commerce/JIBC/articles.htm (July 11, 2006)

Moore, G.C. and Benbasat, I. (1991), Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation, *Information Systems Research*, Vol. 2, No. 3, pp. 192-222.

Pavlou & Chai (2002). What drives electronic commerce across culture? A cross-cultural empirical investigation of the theory of planned behaviour, *Journal of Electronic Commerce Research*, Vol. 3, No. 4, pp 240-253.

Polatoglu, V.N. & Ekin, S. (2001). An Empirical Investigation of the Turkish Consumers' Acceptance of Internet Banking Services, *The International Journal of Bank Marketing*, Vol. 19, No. 4, pp. 156-165.

Rogers, E. M. (1995), Diffusion of Innovations (4th ed), New York, Free Press.

Rogers, E. (2003) Diffusion of innovations, (5th ed.). New York: Free Press.

Sarel, D. and Marmorstein, H. (2003). Marketing Online Banking Services: The Voice of the Customer, *Journal of Financial Services Marketing*, Vol. 8, No. 2, pp.106–118.

Tan, M. and Teo, T. S. H. (2000). Factors Influencing the Adoption of Internet Banking, *Journal for association of information system*, Vol.1, No. 5, [Online] Available: WWW.isworld.org (March 11, 2006)

Taylor, S. and Todd, P. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research,* Vol. 6, No.2, PP. 144-176.

Taylor, S. and Todd, P. (1995b). Decomposition and Crossover Effects in the Theory of Planned Behaviour: A Study of Consumer Adoption Intentions, *International Journal of Research in Marketing*, Vol. 12, No. 2, pp. 137-155.

Tornatzky, L.G., Klein, K.J. (1982). Innovation Characteristics and Innovation Adoption-Implementation: A

Meta-Analysis of Findings, IEEE Transactions on Engineering Management, New York: Vol. 29, No. 1, pp. 28-45.

Venkatesh, V., and Davis, F.D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies, *Management Science*, Vol. 45, No. 2, pp. 186-204.

Lichtenstein, S and Williamson, K (2006). Understanding Consumer Adoption of Internet Banking: An Interpretive study in the Australian Banking Context, *Journal of Electronic Commerce Research*, Vol. 7, No.2, pp. 50-66.

Zolait, A.H.S & Ainin, S. (2008). Evidence That Normative Beliefs is a Multidimensional Construct: Personal and Mass Media Referents, Proceedings of the 9th IBIMA International Conference on Information Management in Modern Organizations: Trends & Challenges, Marrakech-Morocco, 4-6 January 2008, PP. 1114 – 1123 (ISBN: 0-9753393-8-9)

**Appendices**

Appendix A

Multiple Regressions

| DV1 – Attitude (ATT) | SE | Beta | t | F | *p* | *Hypotheses Testing* |
|---|---|---|---|---|---|---|
| Independent Variables | | | | | | |
| | | | | 118,286 | 0.000 | |
| Constant | 1.037 | | 5.363 | | 0.001 | |
| IV1–Relative Advantage/Compatibility | .027 | .466 | 8.764 | | 0.000 | Supported |
| (RAC) | .030 | -.026 | -.627 | | 0.531 | Rejected |
| IV2-Oservability (OBS) | .045 | .325 | 6.230 | | 0.000 | Supported |
| IV3-Ease of Use (EOU) | .056 | .068 | 1.653 | | 0.099 | Rejected |
| IV4–Trialability (TR) | | | | | | |
| Model Summary | | | | | | |
| | R | $R^2$ | Adj. $R^2$ | Durbin-Watson | | |
| | .752(a) | .565 | .560 | 2.010 | | |
| | | | | | | |
| DV2 – Subjective Norms (SN) | SE | Beta | T | F | *P* | |
| Independent Variables | | | | | | |
| | | | | 210.169 | 0.000 | |
| Constant | .786 | | 13.774 | | 0.000 | |
| IV1 - Personal Referent (PR) | .006 | .596 | 13.334 | | 0.000 | Supported |
| IV2 - Media Referent (MM) | .011 | .196 | 4.396 | | 0.000 | Supported |
| Model Summary | | | | | | |
| | R | $R^2$ | Adj. $R^2$ | Durbin-Watson | | |
| | 0.731 | 0.535 | 0.532 | 1.794 | | |
| | | | | | | |
| DV3 – Behavioural Intention | SE | Beta | T | F | *P* | |
| (BI)Independent Variable | | | | | | |
| | | | | 406.029 | 0.000 | |
| Constant | .900 | | .319 | | 0.750 | |
| IV1 – Attitude (ATT) | .046 | .749 | 21.808 | | 0.000 | Supported |
| IV2 – Subjective Norms (SN) | .026 | .138 | 4.011 | | 0.000 | Supported |
| Model Summary | | | | | | |

| | R | $R^2$ | Adj. $R^2$ | Durbin-Watson | |
|---|---|---|---|---|---|
| | .830 | .689 | .688 | 2.136 | |

**P > .05 ** p >.1*

Correlation

| Variables | BI | ATT | SN | PR | MM | RACOMPT | OBSERVABLITY | EASEOFUSE | TRIALABLITY |
|---|---|---|---|---|---|---|---|---|---|
| BI | 1 | | .534(**) | .394(**) | .499(**) | .707(**) | -.045 | .700(**) | .130(*) |
| ATT | .822(**) | 1 | .530(**) | .384(**) | .482(**) | .716(**) | .050 | .672(**) | .211(**) |
| SN | .534(**) | .530(**) | 1 | .714(**) | .555(**) | .562(**) | .057 | .532(**) | .199(**) |
| PR | .394(**) | .384(**) | .714(**) | 1 | .603(**) | .481(**) | .218(**) | .383(**) | .261(**) |
| MM | .499(**) | .482(**) | .555(**) | .603(**) | 1 | .581(**) | .169(**) | .460(**) | .231(**) |
| RACOMPT | .707(**) | .716(**) | .562(**) | .481(**) | .581(**) | 1 | .140(**) | .726(**) | .265(**) |
| OBSERVABLITY | -.045 | .050 | .057 | .218(**) | .169(**) | .140(**) | 1 | -.076 | .512(**) |
| EASEOFUSE | .700(**) | .672(**) | .532(**) | .383(**) | .460(**) | .726(**) | -.076 | 1 | .099 |
| TRIALABLITY | .130(*) | .211(**) | .199(**) | .261(**) | .231(**) | .265(**) | .512(**) | .099 | 1 |

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# Application of Image Recognition Technology in Sports Competition

Fengzhi Zhao & Feng Bao
Applied Technical College, Daqing Petroleum Institute, Qinhuangdao 066004, China

*The research is supported by Educational Commission of Heilongjiang Province of China (2004.3).*

**Abstract**

This article introduces the design principle and implementation method of the automatic judgment software of terminal in sports competition. Image recognition technology makes the competition management level achieve equality, justice, precision and high efficiency, This actualizing automatic judgment by means of recognition model for sports imaging, which is based on the principle of radio frequency identification (RFID).

**Keywords:** Image Recognition, Judgment management software, Sports imaging, RFID

## 1. Introduction

With the rapid development of computer technology and internet, the electric information technology has applications in every walk of life and particularly the sports domain. Some scholars apply high technology measures taking the computer technology as their basis in physical training. This revolutionary change has affected the whole athletic sports. In traditional sports competition, this information gathering mode relied on the personnel giving out judgments through visual judgement then inputs the date recorded into the computer. This process wastes time and energies and may produce various wrong data because of various artificial mistakes which can influence the reliability of information gathered. However, the automatic identification technology can provide automatic transmission of information and automatic identification through computer, and increase the speed and veracity of judgement, make people cast off multifarious works of statistical identification, this eliminate the interference of human judgement, and highly enhances the efficiency and level of sports judgment, and really achieves equity, justice, and a high efficiency for competition.

Presently, the national and domestic sporting competitions, so with computer and electric equipment becoming the important management measures particularly for the track meets which make competition more justice, high efficiency, exact and without errors, however these systems cost very expensively (usually about millions of yuan RMB). General universities, middle schools, elementary schools and other grass roots units can not support these large investments at all, and their sports technology is still low, and most of them have no electronic equipment for training and competition. To fulfill the needs of training and competition of grass roots athletic meeting, in this article we adopt video frequency and radio frequency identification technology to realize automatization of operation which needs personnel in the past, and this system we designed has many characters such as high precision, low costs, convenient use and will be more accurate for providing good judgements in sporting events .—whether a large meet or a grass-roots meet. All will be fair in the world of sports.

## 2. Design principle of system

### *2.1 Main operation principle*

When the competition begins, the starting camera gathers the light and smog produced by the starting gun and inputs them into the computer to be the sign point of timing start. In the competition, the tracking camera is fixed beside the finish line, which follows and gathers the images of every competition, and watches problems occurring in the process of competition. The terminal camera is fixed beside the finish line, which shoots the images of terminal sprint process of every competition following inclined angle of vertical plane. To the competition category by course, firstly, the image gathering system records and stores the video information when every group of athlete passes terminal, then the computer automatically distinguishes the video image that the chest of every athlete achieves the terminal according to the course and compute the competition achievement of every athlete. To the competition category without course, because every athlete takes the sign of RFID, so the RFID reader can transmit the information that every athlete passes the terminal every time to the computer and the computer will record the athletes' identity information and the time that the athletes pass the terminal every time. When the competition ends, the computer will automatically compare with the information gathered by the starting camera, and give out every athlete's grade and place.

The whole system is composed by image gathering system, image identification processing system, athlete data

management system, RFID radio data processing system and network system. Hereinto, the image gathering system includes camera, video gathering card and video gathering and processing software, which mainly gathers the video information (firing screen) of the starting and the video information of the terminal in every course to the computer and stores them according to their respective course appellation and group. The image identification processing system applies the image processing software to establish image processing platform, and its main function is to identify and time the video images of sprint process gathered by terminal camera. When the competition grades are very close in short distance competition, the grades can be judged through the single frame in a sequence by the video return technology, then the data of place and time affirmed will be transmitted to central management computer for output and storage. The RFID radio data processing system receives the data transmitted by RFID reader and completes lap record and totalizing timing in the middle and long distance competitions. Finally the athlete data management system applies system management software to implement schedule arrangement, grade processing, place ranking, data output and storage for the competition.

*2.2 Design of software module*

The software design of this system includes video image gathering system, video image processing system and radio information processing system.

The video image gathering system has functions including starting shoot, stopping shoot, storing video image and video return.

The video image processing system has functions including establishing athlete information, inputting checking list, video processing and timing, picking up final list, clearing checking list, initializing system and setting up system.

(1) Establishing athlete information. Input athletes' basic information (athlete number, name, sex, group and course) to the database. It also includes some sub-functions such as adding, deleting, inquiring, printing and saving.

(2) Inputting checking list. In the competition, according to starting checking information pick up or add every athlete's information such as number, name, course and group from the database to the checking list, and offer information for the timing processing module.

(3) Video processing and timing. Move the video files saved as "course + group" by the video gathering system to the computer which will deal with these images. First, confirm the starting time through identifying firing screen, then form instantaneous continuous narrow digital images of sprint, move the vertical time axis line to certain part (regulated by the competition rule) of the mobile, push mouse key to confirm, so the computer will read out grades and automatically rank, finally, store videos after processing to the hard disk or mobile disk. This function also includes quick orientation of video, single frame return, course identification, grade computation, results printing and saving data.

(4) Clearing checking list. Completely delete all information in the present checking list and prepare for the next event.

The radio information processing system includes functions of establish initial information of every athlete, starting reader, lap record processing and comprehensive timing processing.

(1) Establishing initial information of every athlete. Establish corresponding relations between digital information on the RFID sign and every athlete's number in every group.

(2) Comprehensive timing processing. Record the totalizing time that athlete passes the terminal every time. When the competition ends, compute every athlete's grade through comparing with the starting time confirmed by the video identification processing.

## 3. Key technology and implementation method

To realize automatic judgment through image identification technology, we need solve following several key problems including (1) how to confirm the starting time, (2) how to identify the sprint video of athlete, (3) how to integrate the videos gathered by the starting camera and the terminal camera, (4) how to compute athletes' grades, (5) how to timing in the middle and long distance events?

Hereinto, the most pivotal technologies are the confirmation of starting time and the identification of sports image which will be discussed in the following text.

*3.1 Confirmation of starting time*

In the initial stage of system design, all methods to confirm the starting time are to add a switch quantity, then transmit the analog signals to the central signal controller through communication cable, and the controller receives and deals with the electric signal of starting point and transform it into digital signal and input it into the computer which will start timer of zero clock, so the starting time is confirmed. This method can timely and exactly obtains the starting signal, but to enhance the reliability of the system, we need buy equipments with high performance and equip a circuit communication cable along the playground, so the costs are very expensive and it is not fit for grass roots games. Through large numbers of experiments, we can find that the video of firing screen also can timely and exactly confirm

the starting time, and the costs are cheap and the use is convenient (Figure 1 is a group of firing video gathered where the first frame is the image to confirm the starting time).

The implementation method can be described as follows. First, obtain the background image in the initial situation of firing screen, then make subtracting operation between present input image and the saved background image and obtain the difference image (seen in Figure 2). According the number of pixel point that the grey degree is bigger than certain threshold value in the difference image, we can confirm the first frame in the firing screen and take is as the start of timing. The concrete implementation flow is seen in Figure 3.

*3.2 Video identification principle of athletes' sprint*

3.2.1 Basic principle of video identification

(1) Adopting the Gaussian filtering method to filter the serial frame difference of image, strengthening the robustness of frame difference to the yawp and automatically separating the sprint area and the background.

(2) According to three frames of image in the image sequence, effectively solving the sheltered problem the former frame and the later frame of sprint through taking the edge superposition of the image grey in the first and last frames difference as the edge of the middle frame.

(3) Adopting joint area operator, taking the joint area which area is smaller than certain threshold value as the background, accordingly eliminating big background yawp and labeling sports multi-area.

(4) To every sports area labeled, adopting outside profile tracking arithmetic to automatically position the initial profile of mobile object, and adopting improved dynamic profile shrinking arithmetic to exactly confirm the outside profile of the object and extract the profile of sports multi-objects.

The checking model of sprint area can be described as follows.

$$Df(x,y)_{[t,t+1]} = G(x,y) \times |f(x,y,t+1) - f(x,y,t)|$$

$$BDf(x,y)_{[t,t+1]} = \begin{cases} 1 & if \quad Df(x,y)_{[t,t+1]} > \tau \\ 0 & otherwise \end{cases}$$

Where,

$$\tau = mid(\tau_1, \tau_2, \tau_3, \tau_4)$$

$$\tau_i = \frac{1}{N_{M_i}} \left[ c \bullet \sum_{(u,v \in M_i)} Df(u,v)_{[t,t+1]} \right].$$

3.2.2 Profile identification model of sports multi-objects

The dynamic profile model can be described as follows.

$$v(s) = (x(s), y(s)) \qquad [s \in [0,1]]$$

And the energy function in the dynamic profile can be described as follows.

$$E_{total} = \int_0^1 E((v(s))ds = \int_0^1 [E_{\text{int}}(v(s)) + E_{ext}(v(s))]ds$$

$$E_{total} = (v(s)) = (\alpha(s)) |v_s(s)|^2 + \beta(s) |v_{ss}(s)|^2)$$

$$E_{ext}(v(s)) = -\gamma |\forall f(v)|^2$$

*3.3 Timing in middle & long distance competition events by RFID technology*

In the middle and long distance competition events, because of too many athletes, the competition rule without course is adopted. The usage of video identification technology has following problems. The first problem is that the technology only can confirm the grade that the athlete passes terminal, and can not confirm concrete athlete. The second one is that it is difficult to identify retaining lap. To solve above problems, this system adopts the technology of RFID.

The basic composing of RFID includes tag, reader and antenna. The tag is composed by coupling component and chips, and every tag possesses one and only electric coding attached on the object marking object. The reader can read (sometimes read-in) the information of tag, which can be designed as hand type or fixed type. The antenna transmits radio signals between tag and reader.

The key technology is to tie the radio card (tag) on the athlete's shoestring and place little shim block with mini-antenna on the terminal of the event (seen in Figure 4). The reader can read and identify the athletes' data saved in the radio card through an untouched way accordingly the automatic identification of athletes' information can be achieved. Connecting the reader with the computer, the information read is transmitted to the computer for the next processing.

When athlete passes this shim block every time, the timing system will receive ID number sent by radio card on the athlete and record the present time and passed lap number.

RFID is a sort of automatic untouched identification technology, which automatically identifies the object through radio signal and acquires relative data, and because the identification needs not manual work, so it can work under various environments. The technology of RFID can identify mobile object with high speed and multiple tags at the same time, and the operation is quick and convenient.

**4. Conclusion**

In this article, we discuss the key technology and implementation method that we adopt image identification technology to actualize the automatic judgment in the athletic sports, and the practice indicates that the resolution of computer image position in this system is 25 frames per second and the time resolving is 4% second which is synchronized with the identification menu. The grade time is very exact according with the standard of International Association of Athletics Federations (IAAF). The image identification timing judgment at terminal in the event can be implemented timely, delayed or repeatedly, which doesn't influence the starting time and can ensure the event is implemented successfully. The competition process is automatically managed by the program of computer and the judgment personnel are very few, so the computer can automatically complete the mission of terminal judgment. The video material of the whole competition can be saved automatically, and the disputed grades or problems can be discussed conveniently, and various man-made mistakes in the competition can be eliminated, so the judgment level of the competition can be enhanced and the justice, speediness, nicety and high efficiency of the competition can be ensured.

**References**

Cai, Wandong. (2000). *Multimedia Communication Technology.* Xian: Publishing House of Xi′an University of Electronic Science and Technology.2000.10. p.50-99.

Cao, Zhijin & Tang, Huiming. (2004). Identification of Auto Type in Video Image. *Computer Engineering Application.* No.24. p.226-228.

D.Zhong and S. F. Chang. (2001). Structure Analysis of Sports Video Using Domain Models (C). *IEEE Conference on Multimedia and Exhibition.* Tokyo, Japan. Aug, 2001. p.22-25.

Gunilla Borgefors. (1986). Distance Transformations in Digital Images. *Computer Vision, Graphics, and Image Processing.* No.34. p.344-371.

Wangqian. (1999). *Judgment Manual of Athletics Competition.* Beijing: China People's Publishing House of Sports. p.81-140.

Initial Firing Screen

Image of First Frame

Figure 1. Changes of Firing Screen Video

Image of Difference Value

Figure 2. Image of Difference Value

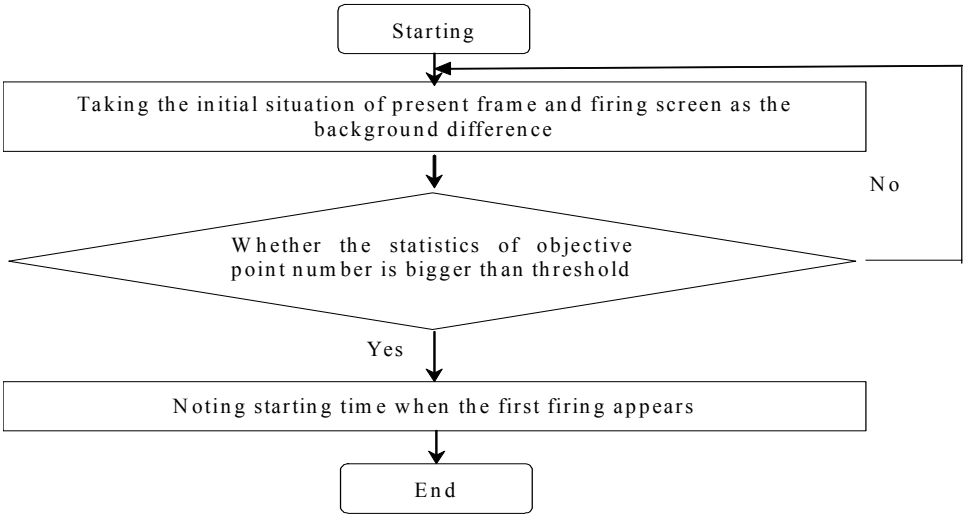



Figure 3. Flow of Starting Time Identification

Figure 4. Little Shim Block of REID Mini Antenna

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 1, No. 1

February 2008

# Polychotomiser for Case-based Reasoning beyond the Traditional Bayesian Classification Approach

Dino Isa

Faculty of Engineering and Computer Science

University of Nottingham, Malaysia Campus, Semenyih, Malaysia.

Tel: 603-89248116 E-mail: dino.isa@nottingham.edu.my

Lam Hong Lee

Faculty of Engineering and Computer Science

University of Nottingham, Malaysia Campus, Semenyih, Malaysia.

Tel: 603 89248141 E-mail: kcx4lhl@nottingham.edu.my

V.P. Kallimani

Faculty of Engineering and Computer Science

University of Nottingham, Malaysia Campus, Semenyih, Malaysia.

Tel: 603 89248141 E-mail: VP.Kallimani@nottingham.edu.my

R. Prasad

Faculty of Engineering and Computer Science

University of Nottingham, Malaysia Campus, Semenyih, Malaysia

Tel: 603-89248116 E-mail: Rajprasad.Rajkumar@nottingham.edu.my

**Abstract**

This work implements an enhanced Bayesian classifier with better performance as compared to the ordinary naïve Bayes classifier when used with domains and datasets of varying characteristics. Text classification is an active and on-going research field of Artificial Intelligence (AI). Text classification is defined as the task of learning methods for categorising collections of electronic text documents into their annotated classes, based on its contents. An increasing number of statistical approaches have been developed for text classification, including k-nearest neighbor classification, naïve Bayes classification, decision tree, rules induction, and the algorithm implementing the structural risk minimisation theory called the support vector machine. Among the approaches used in these applications, naïve Bayes classifiers have been widely used because of its simplicity. However this generative method has been reported to be less accurate than the discriminative methods such as SVM. Some researches have proven that the naïve Bayes classifier performs surprisingly well in many other domains with certain specialised characteristics. The main aim of this work is to quantify the weakness of traditional naïve Bayes classification and introduce an enhance Bayesian classification approach with additional innovative techniques to perform better than the traditional naïve Bayes classifier. Our research goal is to develop an enhanced Bayesian probabilistic classifier by introducing different tournament structures ranking algorithms along with a high relevance keywords extraction facility and an accurately calculated weighting factors facility. These were done to improve the performance of the classification tasks for specific datasets with different characteristics. Other researches have used general datasets, such as Reuters-21578 and 20_newsgroups to validate the performance of their classifiers. Our approach is easily adapted to datasets with different characteristics in terms of the degree of similarity between classes, multi-categorised documents, and different dataset organisations. As previously mentioned we introduce several techniques such as tournament structures ranking algorithms, higher relevance keyword extraction, and automatically computed document dependent (ACDD) weighting factors. Each technique has unique response while been implemented in datasets with different characteristics but has shown to give outstanding performance in most cases. We have successfully optimised our techniques for individual datasets with different characteristics based on our experimental results.

**Keywords:** Text Classification, Bayes Theorem, Bayesian Filtering, Probability, Case-Based Reasoning

## 1. Introduction

Document classification is defined as the task of learning methods for categorising collections of electronic documents into their annotated classes, based on its contents. For several decades now, document classification in the form of text classification systems have been widely implemented in numerous applications such as spam filtering. (Sahami et.al., 1998; Cunningham et. al., 2003; Delany, Cunningham & Coyle, 2005), e-mails categorising (Kamens, 2005), directories maintenance, and ontology mapping (Su, 2002), contributed by the extensive and active researches. An increasing number of statistical approaches have been developed to document classification, including k-nearest-neighbor classification, naïve Bayes classification, support vector machines, decision tree induction, rule induction, maximum entropy, artificial neural network, etc.

Each of the document classification schemes has their own properties. The decision tree induction algorithm and rule induction algorithm are simple to understand and interpret after a brief explanation. However, these algorithms do not work well when the number of distinguishing features is large (Quinlan, 1993). k-nearest neighbor algorithm is easy to implement and shows its effectiveness in a variety of problem domains (Han, Karypis & Kumar, 1999). A major drawback of the k-NN algorithm is computationally intensive, especially when the size of the training set grows (Han, Karypis & Kumar, 1999). Support vector machine can be used as a discriminative document classification method which has been shown to be more accurate in classification tasks (Joachims, 1998; Chakrabakti, Roy & Soundalgekar, 2003). The high accuracy of SVM is due to the implementation of structural risk minimisation which entails finding a hyper-plane which guarantees the low error plus an ability to learn which is independent of the dimensionality of the feature space (Joachims, 1998). However, the usage of SVMs is not popular in many real world applications due to its convoluted training and categorising algorithms (Chakrabakti, Roy & Soundalgekar, 2003).

Among these approaches, naïve Bayes text classification has been widely used because of its simplicity in both training and classifying stage although this generative method has been reported less accurate than discriminative methods such as SVM (Joachims, 1998; Chakrabakti, Roy & Soundalgekar, 2003). However, some researches have proven that naïve Bayes classification approach provides intuitive simple text generation models and performs surprisingly well in many other domains, under some specific conditions (McCallum & Nigam, 2003). A naive Bayes classifier is a simple probabilistic classifier based on Bayes' Theorem with strong independence assumptions but this assumption severely limits its applicability. Depending on the precise nature of the probability model, naive Bayes classifiers can be trained very efficiently and requires a relatively small amount of training data to estimate the parameters necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

Bayesian Filtering is a probabilistic inference approach which is typically implemented in mail repositories to remove spam e-mails (Sahami et.al., 1998; Delany, Cunningham & Coyle, 2005; Cunningham et. al., 2003; Delany et. al., 2004). Bayesian filtering is a highly effective approach in the task of classifying data with a very low number of false positives. It overcomes the obstacles faced by more static technologies. The traditional Bayesian Filtering approach is implemented as dichotomiser, a two-category classifier, which classifies e-mails into spam and legitimate, by calculating the overall probability of the text body of the document. However, our research has emphasised in developing an extended version of Bayesian classifier which is able to handle multiple categories classification tasks, and guarantee minimum error rate classifications. Our proposed Bayesian probabilistic classifier has been implemented in conjunction with a self-organising map (Hartley et. al., 2006) in a case-based reasoning system which contribute to an efficient and time saving case retrieval process. Figure 1 illustrates the block diagram of our proposed case-based reasoning system's structure, which is enhanced with the extended Bayesian classifier in this work.

In the context of classification, Bayes theorem emphasises that the probability that a particular document is annotated to a particular category, given that the document contains certain words in it, is equal to the probability of finding those certain words in that particular category, times the probability that any document is annotated to that category, divided by the probability of finding those words in any document.

$$\Pr(\text{Category} \mid \text{Word}) = \frac{\Pr(Word \mid Category).\Pr(Category)}{\Pr(Word)}$$

Each kind of text documents contains words which are given probabilities based on its number of occurrence within that particular kind of documents. Bayesian filtering is predicated on the idea that spam e-mails can be filtered out based on the probability that certain words will correctly identify a piece of e-mail as spam while other words will correctly identify a piece of e-mail as legitimate. At the basic level, a Bayesian filter examines a set of e-mails that have been categorised to be spam and legitimate, and compares the content in both categories in order to build a database of words and their occurrence. The list of words occurrence is used to identify or predict future e-mails as spam or legitimate, according to the probability of words from the whole body of the text message of an e-mail.

Many researches have been carried out to implement Bayesian probability theories in case-based reasoning approach for case retrieval purposes. Bayesian network is one of the popular mechanism which many research groups have

investigated and developed new approaches for case matching algorithm in case-based reasoning (Myllymaki & Tirri, 1993; Myllymaki & Tirri, 1994; Kontkanen et. al., 1997; Kontkanen et. al., 1998). In this paper we emphasis on the Bayesian filtering approach to be implemented as an extensive version of classifier and acts as a part of our proposed case-based reasoning's case retrieval algorithm.

In order to perform the task as a probabilistic classifier at the front end of case retrieval algorithm, the traditional Bayesian filtering approach needs to be extended to handle different types of multiple categorised data efficiently. The solutions database in a knowledge repository can be divided into multiple categories according to their similarity of properties, attributes and features. This classification leads to the efficient and time-saving case retrieval since the solution cases are well-organised by the self-organising map. A solutions database can be classified into linear multi-dimensional organisation, or hierarchical organisation, or even the integration of the above methods. Therefore, our classifier is designed to be highly adaptable to these databases with different organisations.

The proposed Bayesian classifier in our research is equipped with some specialised algorithms to ensure the high performance in handling different types of knowledge domain with unique characteristics and guarantee a low error rate classification. An extendable rank classification algorithm and a series of tournament structures algorithms have been designed and implemented in our proposed system. Besides, the polychotomiser, which represents the multiple categories classifier, is equipped with some extra techniques to guarantee the high accuracy of classification tasks.

## 2. The Classification Algorithms

### *2.1 Rank Classification Algorithm*

The rank classification algorithm slightly modifies the traditional Bayesian filtering probability calculation in order to support multi-category dataset. The probability value for a document A to be annotated to a category C is computed as Pr(C|A). As an assumption that we have a category list as Cat1, Cat2, Cat3, Cat4, Cat5, ……., CatN, thus, each document has N associated probability values, where document X will have Pr(Cat1|X), Pr(Cat2|X), Pr(Cat3|X), Pr(Cat4|X), Pr(Cat5|X), ……, Pr(CatN|X). The probability value of an input document to be annotated to a particular category is calculated by considering all the categories together in a single round. The rank classification algorithm is then sorts all the probability values for each category and selects the highest score as the most likely category to be annotated.

A multi-category Bayesian classifier needs to fulfill some conditions to guarantee the high accuracy of classification. One of the conditions is that the classifier needs to be trained on an approximately equal number of training documents for each category due to the imbalance of sizes between categories may cause great risk in misclassifications. The greater the imbalance, the worse problems occur. This is due to the prior probability Pr(Category), is computed as 1/the total number of category. This equation is made based on the assumption that training set is perfectly balanced with the same number of training documents for each category. Furthermore, even the number of training documents for each category is balanced, the problems still occur since the size of each training documents is not approximately equal to each other. Therefore, as the solution for the problem of imbalance of training dataset, the prior probability for each category, Pr(Category), is transformed from the ordinary 1/the total number of category, to the equation:

$$\Pr(\text{Category}) = \frac{Total\_of\_Words\_in\_Category}{Total\_of\_Words\_in\_Training\_Dataset} = \frac{Size\_of\_Category}{Size\_of\_Training\_Dataset}$$

### *2.2 Round Robin Tournament Algorithm*

Tournament structures are possible to be implemented in the classification algorithms to handle multi-category classification. Previous researches proved that the round robin tournament algorithm can performs the spam e-mails filtering beyond the ordinary binary classification (Kamens, 2005). In a round robin tournament algorithm, the calculation of the Bayesian probability values is performed only between two categories. It is a looping binary classification algorithm, and each competitor plays against all the others an equal number of times, typically once. The round robin tournament algorithm contributes to a relatively simple and complete competition among all the categories, and the process iterates until every category has compete against all the others.

The structure of the round robin tournament algorithm in our proposed system is a "Host and Guest" concept. Firstly, all the categories are randomly sorted. The first category will be the host for the initial round and plays against all the others sequentially which ranked below it as guests. At the second round, the second category will become the host for the second round and those categories which are ranked below it will compete against the host. The process iterates until all categories have played against the others an equal number of time. There are some methods available to determine the final winner after the iterative calculation processes complete. One of the methods is the winning category of each match is awarded with a score, typically 1, and the loser is not awarded with any score, or in the other words, score 0 is awarded. The scores from every match of a particular category are added together after the competition until the calculation has completed. The category with the highest score will be the overall winner, which represents the right category for the input document to be annotated.

There is a situation where dilemma occurs in determining the right category of an input document when more than one category which have the same highest final score. This situation can be avoided by awarding the two competing categories of each match with the score which is equal to their probability value computed from the binary classification. With this method, the final highest score for every category is rarely to be the same.

As the result, the round robin tournament has an improved ability to distinguish similar categories, since it is a looping binary classification algorithm. The binary classification algorithm can easily differentiate between two similar categories that both have great differ from the others. It is smart enough to isolate two categories temporarily and perform the probability values calculation without considering other parties. However, the iterative binary classification process consumes a relatively long time compare with other algorithms.

*2.3 Single Elimination Tournament Algorithm*

By comparing with other algorithms, the single elimination tournament structure has some restrictions. Firstly, most often the number of competitors is fixed as a power of two. Somehow, in the situation that the number of participants is not a power of two, typically the highest-rated competitors from the previous accomplishment will be advanced to the second round without joining any match in the first round. Besides, seeding is recommended as a pre-process to prevent highest-rated competitors being scheduled to face each other in the early stages of the competition. The seeds ranking process can be executed by using the rank classification algorithm or the round robin tournament algorithm. Therefore, the single-elimination algorithm is more suitable to be implemented at the back-end of an integrated algorithm.

As similar to the round robin tournament algorithm, single-elimination algorithm also performs the probability values calculation in the form of binary competition for every match. In the first round, we play the best competitor against the worst, and the second best against the second worst, and so on for the rests. Brackets are set up, so that the top two seeds could not possibly meet until the final round, none of the top four can meet before the semifinals, and so on. This concept is applicable in the following rounds until the overall winner is representing by the winner of the final round.

In contrast to round robin tournament structure, as rounds progress, the successive rounds of the single elimination tournament structure halve the number of remaining competitors by progressing the winners from the previous round to the next round and eliminating the losers. Single-elimination tournament algorithm is suitable to be implemented in the domains which have a large number of categories. Somehow, since this algorithm is also a binary classification based algorithm, it has great ability in handling the classification tasks which involve categories with high degree of similarities.

*2.4 Swiss System Tournament Algorithm*

In our proposed system, the Swiss system tournament algorithm can be implemented independently or at the back-end of a hybrid algorithm. The initial seeding of a Swiss system tournament is not a compulsory as the single-elimination tournament algorithm, but it is recommended. The competing categories are then divided into two parts, the top half which is paired with the bottom half. As an example, if there are eight categories in the classifier, first category is paired with fifth category; the second is paired with the sixth, the third is paired with the seventh, and so on.

After the first round of the competition, the winners from the first round will plays against the winners, and the losers will plays against the losers. As similar to the round robin tournament, the winning category of each match is awarded with a score, typically 1, and the loser is not awarded with any score. In further rounds, each competitor will be pitted against another competitor who has the same score. Modifications are then made to prevent competitors from meeting each other twice.

In contrast to round robin tournament, the Swiss system tournament algorithm can determine the top ranked and bottom ranked competitors with fewer rounds, although the middle rankings are unreliable. By the way, we only have interest on the final overall winner. Therefore, the Swiss system is applicable in the classification tasks of our system with large number of categories, and similar to other binary classification based algorithms, it is suitable to be implemented in the classification tasks which the domain contains the categories with high degree of similarities.

However, the number of competing categories has becomes the biggest obstruction for the Swiss system tournament algorithm. As similar to the round robin tournament, the Swiss system tournament algorithm has the potential in facing a dilemma in determining the annotated category of an input document, or the final winner. It may happen that two or more categories have the same highest and perfect score, won all the games but never faced each other. Therefore, the ordinary algorithm needs to be slightly modified in terms of the number of rounds played. To determine a clear overall winner, we have applied the same concept with single-elimination tournament algorithm in terms of number of rounds that is the base 2 logarithm of the number of competitors rounded up.

## 3. The Low Error Rate Classification Techniques

*3.1 Keywords Counting Methods- Multinomial and Multivariate Technique*

In our proposed system, as similar to other approaches, the classifier must be trained in advance so that it can build up all the probability values for every recognised keyword to be annotated to every category. The knowledge engineers or the domain experts must manually organise the training dataset which contains a reasonable number of training files for each

category, and then the classifier will accordingly adjusts the words' probabilities in every category of the database. In the training phase, we first need to analyse the training files by extracting all the words to generate a list of words occurrence frequency. Our prototype system, which is developed by using JAVA, analyses the plain text documents from the training set to generate a list of words occurrence frequency for each category, and the list of words occurrence frequency for every category is stored independently by using a data structure, as a TreeMap of String word -> Integer frequency. The occurrence of a particular word in the list of words occurrence for a particular category is given by the value regards to the total occurrence of that particular word in all the training files from that particular category, so called multinomial method.

The multinomial keywords counting method is suitable to be implemented for the training dataset which has the approximately balance sizes of categories. If the training dataset is under the imbalance condition in terms of the size of each category, problems may occur since that each indicative probability score is very low, and the problems may become serious as the greater imbalance of training data. This problem can be solved by collecting training files which has the approximately same size.

Multivariate keywords counting method is an alternative method which has better performance as compared to the multinomial method under the condition of imbalance training dataset. The multivariate method calculates the words occurrence frequency for every category based on the number of training files which contain the particular word in a particular category. For example, if a particular word is found in a number of training files which have been categorised under a particular category, the occurrence of that particular word in the list of words occurrence for the particular category is given by the value regards to the total number of training files which contain the particular word in that particular category.

However, the multivariate method is not suitable to be implemented in the cases which the categories are very similar to each other. The results from our experiments in handling dataset with high degree of similarity show that multinomial method performs better than multivariate method. This can be concluded as the training files of similar categories contain most of the same keywords with each other. The frequency of usage of a particular keyword in different categories is the key factor to differentiate these similar categories. Therefore, the classifiers which handle categories with high degree of similarities are preferably to be implemented with the multinomial keywords counting method.

### *3.2 High Relevance Keywords Extraction Facility*

The ordinary Bayesian filtering approach takes the whole message into account to identify whether an e-mail is spam or legitimate (Sahami et. al., 1998). Improving on this classification method, we have proposed, for our text classifier, a technique which identifies the high relevance keywords during the fly of the classification process, not from a pre-defined keywords dictionary. When the system has analysed the input document or input query the input text is stored into a class Scanner instead of String since Scanner can parse primitive types and strings using regular expressions. During the calculation of the probability values, each individual word from Scanner is extracted to calculate for the probability to be annotated to each of the categories. However, only words which have relatively high probability value, or "Important" to be annotated to a particular category compare with the others will be taken into account. The degree of "Importance" of keywords can be adjusted by setting a threshold. Based on our experiments, the greater the threshold, the better the classifier performs. However, the performance decreases when the threshold reaches a certain limit, which is the saturation point of the degree of difference. The saturation point of the degree of difference is domain-dependent.

### *3.3 ACDD Weighting Factors Facility*

The research has found that in certain situations, certain categories have high number of misclassified documents. In other words, a large number of documents are likely to be misclassified in a wrong category. This problem may be caused by reasons such as the imbalance training datasets and the improper organisation of the training database resulting in the mis-training of the system.

This work has proposed a solution for the problem mentioned above, which is the implementation of the automatically computed document dependent (ACDD) weighting factors to the probability values of the input data. Different values of weighting factor are applicable to different categories in the classification tasks. Categories which have high intensity of correctly classified data will be awarded with a relatively small value of weighting factor. On the other hand, relatively large value of weighting factors will be awarded to categories which their documents are always been misclassified as others. For the ACDD weighting factors calculation session, a process which is similar to the calculation of probability values of input documents, is executed by getting another set of data, namely ACDD weighting factors retrieval dataset as the input. Therefore, to implement the ACDD weighting factors method, an initial training dataset, and an ACDD weighting factors retrieval dataset are needed for the pre-process for the classifier before it can start its classification tasks.

The ACDD weighting factors retrieval dataset can be organised by the domain experts, or extracted from the training dataset. The system is initially trained with the training dataset before performs the weighting factors computation. The ACDD weighting factors is calculated by loading the ACDD weighting factors retrieval dataset as the input data and the

results are recorded for the purpose of generating the weighting factors for every category. Weighting factor of a particular category is computed based on the total number of documents from the ACDD weighting factors retrieval dataset which are annotated to it. The more documents are annotated to the category, the smaller the value of its weighting factor.

After the ACDD weighting factors computation, every category has been awarded by a unique weighting factor. During the classification process of the input data, the probability values of the document are calculated based on the same training dataset and the same algorithm and methods. The probability values of a document to be annotated to each category are added with their own unique weighting factor before the system determines the annotated category of the input document.

Results from our experiments proved that the ACDD weighting factors implementation successfully reduces the misclassification rates. However, among the disadvantages of this approach is that it consumes a relatively long processing time compare with other approaches, and required a complex algorithm to compute the weighting factors automatically. The system needs to process the ACDD weighting factor retrieval dataset after the training session to compute the weighting factors for every category. Besides, a complex and iterative algorithm is needed to determine the formula and the multiplier for the computation of the weighting factors for every category, which need to be tested iteratively until a set of optimum ACDD weighting factors for every category is generated.

## 4. The Experimental Results

The objective of these evaluations is to determine whether our proposed approaches and methods resulted in better classification accuracy and performance compare with the ordinary version, which will greatly contribute as the front-end classifier for case retrieval stage of case-based reasoning system.

The evaluations of our prototype system are executed by applying different kinds of knowledge domain, different probability values calculation algorithms and different low error rate classification methods. As mentioned in the sections above, every classification algorithm and low error rate classification method has different performance, depending on the characteristics of the knowledge domain. Therefore, to have an optimum Bayesian classifier, the flexibility of the classifier is needed by selecting right techniques to the right knowledge domain.

### *4.1 Experiment 1: Dataset with Low Degree Similarity and Imbalanced Category.*

The dataset of variants of vehicle is tested by our prototype system for the evaluation of classification performance in handling the case with categories which have low degree of similarity. Our selected dataset contains four categories of vehicles: Aircrafts, Boats, Cars, and Trains. All the four categories are easily to be differentiated and every category has a large number of their own unique keywords. We have collected 90 documents for each category, with the total of 360 documents in the entire dataset. 30 documents from each category are extracted randomly to build the training dataset for the classifier. The rest of 60 documents for each category are remained as the testing dataset to test the classifier.

Initially, we have performed the experiment by implementing different classification algorithms: the rank classification, round robin tournament, single-elimination tournament and Swiss system tournament. The goal of this experiment is to compare the performance of these algorithms. Figure 2 illustrates the comparison chart of the performances of each algorithm in this experiment.

The results illustrated in Figure 2 show that the rank classification algorithm performs the best among the other classification algorithms, with the classification rate of 88.89%. The rank algorithm is a direct classification approach which performs the probability values calculations in one round and consumes the shortest time as compared to the others.

The round robin tournament algorithm performs the best among the tournament structure based algorithms since it contributes to a relatively simple and complete competition among the categories. The classification rate of round robin tournament algorithm is 79.50%. However, the complete competition among the categories is due to the round robin tournament algorithm executes the binary classification process iteratively. Therefore, relatively long time consumption is required by this algorithm.

The single-elimination tournament algorithm has the similar performance with the Swiss system tournament algorithm, where the classification rate is 76.82% for single-elimination and 76.44% for Swiss system. However, both of these algorithms need a pre-process for the initial seeds ranking, which is a compulsory for the single-elimination tournament algorithm but is an optional for the Swiss system. Hence, the results show that the integrated versions are not performing as good as the independent versions of classification algorithm in terms of performance and time consumption.

Another experiment is also been carried out to justify the contribution of the low error rate classification methods. By applying these methods to the classifier which implements the rank classification as the basic algorithm, a comparison of the performance of these low error rate classification methods is illustrated as Figure3.

The results illustrated in Figure 3 above show that the keywords occurrence counting based on the multivariate method performs better with the classification rate of 88.89%, while the keywords occurrence counting based on the multinomial

method only contributes to a classification rate of 64.94%. The reason for the relatively high misclassification rate of multinomial method is due to the imbalance of the size of categories in the training dataset. Problems may occur due to the high frequency of occurrence for certain keywords in big size categories which may lead to the classifier misclassified most of the input documents to them.

Besides our prediction for the better performance of the implementation of High Relevance Keywords Extraction method in overall, the HRKE also performs better while threshold for the "Importance" of keywords increase. The High Relevance Keywords Extraction method contributes to 89.08% classification rate with the threshold is set at 40%, and 95.21% with the threshold is set at 90%, and 97.92% with the threshold is set at 95%. This method culls out all the common and disregarded words for every category, and also the words which have the similar effect to the classification task. Therefore, the confusion during the classification task for the input documents is reduced and this will leads to a higher classification rate.

*4.2 Experiment 2: Dataset with High Degree Similarity and Balanced Categories*

Another dataset which has been tested by our prototype system is the dataset of mathematics topics, which has a high degree of categories' similarity. Due to these topics are the sub-topics of mathematics subject, the common mathematical terms are widely used by the documents from these topics. There are only a few specific and unique terms to differentiate each topic, therefore the degree of similarities for the categories from this dataset is relatively high compare with the Vehicles dataset. 10 files for each category have been organised as the training database, while the testing dataset contains 240 files for testing purposes.

As similar to the previous experiment, we have tested our classifier with the proposed classification algorithms. We have executed each of these under the basic condition by not implementing any low error rate classification methods for the goal of pure comparison of the performance between the classification algorithms. The chart illustrated in Figure 4 shows the comparison of the performance of each algorithm in this experiment.

The pattern of this comparison is different from Experiment 1. The rank classification algorithm contributes to a lower classification rate of 80.42% since it calculates the probability values of every category for an input data in one round and takes all categories into account. This will lead to a great confusion to the classification tasks since all the categories are similar to each other in terms of words usage. In such a situation, binary classification based algorithms are able to overcome this problem. Therefore, all the tournament structure based algorithms perform better than the rank classification algorithm, with the same classification rate of 81.25% which is slightly higher than the rank classification algorithm. Although all the three tournament structure based algorithms in this experiment have the same classification rates co-incidentally, the patterns of classification for each of them are different.

The following test is to figure out the comparison of performance of the low error rate classification methods. The rank classification algorithm has been chosen as the basic classification algorithm. The chart in Figure 5 illustrates the comparison of performance of every low error rate classification methods for this high degree of similarity and balanced dataset.

From the comparison chart illustrated in Figure 5, we concluded that for the classifier to handle dataset with high degree of categories' similarity, the multinomial keywords counting method performs better than the multivariate method, which is 80.42% compare to 77.92%. This is due to majority of the documents in every category contain the same keywords since the degree of similarity is high. To differentiate these highly similar categories, the keywords counting based on the number of keywords occurrence in all the training files of the categories is required since different topics may have different frequency of usage for a particular keyword.

The High Relevance Keywords Extraction method does not act the same pattern with Experiment 1. In this experiment, every word in the documents may bring great effect to the classification tasks. The increment of the threshold leads to the less words to be considered. However, the common and disregarded words can be culled out with a low threshold, which can slightly increase the classification rate of the classifier. When the value of the threshold increases, the important keywords for the classification task are not taken into account. Therefore, the value of threshold in such a situation should not over the saturation point of the degree of difference.

The most significant method which we have discovered in our research is the ACDD weighting factors implementation. Weighting factor is a variable component used in calculations that allows for a margin of error above the minimum error on a measurement, or produces a desired result. This method contributes to 88.90% of classification rate, compares to the others which are lower than it. This is due to the weighting factors are able to reduce the intensity of misclassified data for some "popular" categories, and contributes to the increment of the classification accuracy. The highest classification rate from this experiment by implementing the ACDD weighting factors is 88.90%. Experiment 1 scored the highest classification rate at 97.92% by implementing the HRKE facility, which can be considered as an almost perfect classification task. By referring back to the number of training files for both experiments, Experiment 2 takes the training dataset which contains 10 training files for each category while Experiment 1 has 30 training files for each category in the

training dataset. From here we can conclude that this system is acting like other Artificial Intelligence application, "The More It Learns, The Smarter It Works".

## 5. Conclusion and Future Works

The case retrieval approach for a case-based reasoning system through the implementation of the Bayesian Filtering technique at the front-end, in conjunction with a self-organising map at the back-end, has been proposed and developed by our research group. This approach takes the advantages of Bayesian probability theorem which may greatly enhance the performance of ordinary case retrieval algorithms. Besides, we introduce a series of classification algorithms and low error rate classification methods to develop an extensive version of Bayesian probability classifier. The results from our experiments show that the Bayesian probability classifier is required to be flexible and optimised for different datasets. This can be done by implementing different algorithms and methods to match the requirements of different characteristics of variety of dataset. We have established the optimal requirements for the classification tasks for specific dataset with different characteristics. The experimental results show that the properties of all the proposed classification algorithms and also the techniques used for the low error rate classification can be optimised for any domain of different characteristics. In the future, our research group is emphasising on enhancing the ability and performance of our existing prototype by introducing some facilities, such as natural language processing approach to handle sentences classification, not restricted to individual keywords classification. Our group hopes to extend the advantages of Bayesian probability theorem, not restricted to categorise data, but towards more AI applications such as in sensor monitoring (Isa et. al., 2007).

## References

Aamodt, A., Plasa, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AICom - Artificial Intelligence Communications, IOS Press, Vol. 7: 1,* pp. 39-59.

Agre, G., Koprinska, I. (1996).Case-Based Refinement of Knowledge-Based Neural Network. *In Proceedings of the International Conference,Intelligent Systems: A Semiotic Perspective", October 20-23, Gaithersburg, MD, USA*.

Bergmann, R., Richter, M. M. Schmitt, S. Stahl, A. Vollrath, I. (2001). Utility-Oriented Matching: A New Research Direction for Case-Based Reasoning. *In Proceeding of the 9 th German Workshop on Case-Based Reasoning, GWCBR'01, Baden-Baden, Germany*.

Block, S., Medin, D., & Osherson, D. (2002). Probability from Similarity. Northwestern University; Rice University.

Burkhard, H. D.Richter, M. M. On the Notion of Similarity in Case-Based Reasoning and Fussy Theory. Soft computing in case based reasoning, *Springer-Verlag,* London, 2000.

Chakrabarti, S. Roy, S.& Soundalgekar, M. V. (2003). Fast and Accurate Text Classification via Multiple Linear Discriminant Projection. The VLDB Journal The International Journal on Very Large Data Bases*,* pp. 170-185.

Cunningham, P. Nowlan, N. Delany, S. J. & Haahr, M. (2003). A Case-Based Approach in Spam Filtering that Can track Concept Drift. *In The ICCBR'03 Workshop on Long-Lived CBR Systems, Trondheim, Norway*.

Delany, S.J., Cunningham, P. Tsymbal, A. & Coyle, L. (2004). A Case-Based Technique for Tracking Concept Drift in Spam Filtering. *Journal of Knowledge Based Systems, 18 (4-5), Elsevier,* pp. 187-195.

Delany, S. J. Cunningham, P. & Coyle, L. (2005). An Assessment of Case-Based Reasoning for Spam Filtering. *Artificial Intelligence Review Journal, Volume 24, Numbers 3-4*, pp. 359-378.

Flach, P. A., Gyftodimos, E. & Lachiche, N. (2002). Probabilistic Reasoning with Terms. University of Bristol, Loius Pasteur University, Strasbourg.

Golding A.R., Rosenbloom, P. S. (1996). Improving Accuracy by Combining Rule-Based and Case-Based Reasoning. *Artificial Intelligence 87*, pp. 215-254.

Han, E. H., Karypis, G., & Kumar, V. (1999). Text Categorisation Using Weight Adjusted k-Nearest Neighbour Classification. Department of Computer Science and Engineering, *Army HPC Research Center*, University of Minnesota.

Hartley, M., Isa, D., Kallimani, V. P., & Lee, L. H. (2006). A Domain Knowledge Preserving in Process Engineeringusing Self-Organising Concept. Intelligent System Group, Faculty of Engineering and Computer Science, University of Nottingham Malaysia Campus, Malaysia.

Isa, D., Prasad, R., Lee, L. H., & Kallimani, V. P. (2007). Oil and Gas Pipeline Acoustic Sensor Monitoring System. Intelligent System Group, Faculty of Engineering and Computer Science, University of Nottingham Malaysia Campus, Malaysia.

Jacsynski, M., Trousse, B. (1998). An Object-Oriented Framework for the Design and the Implementation of Case-Based Reasoners. *In 6Th German Workshop on Case-Based Reasoning, Berlin, Germany*.

Joachims, T. (1998). Text Categorisation with Support Vector Machines: Learning with Many Relevant Features. *In Machine Learning: ECML-98, Tenth European Conference on Machine Learning,* pp. 137--142.

Kamens, B. (2005). Bayesian Filtering: Beyond Binary Classification. Fog Creek Software, Inc.

Kontkanen, P., Myllymaki, P., Silander, T., Tirri, H. (1997). A Bayesian Approach for Retrieving Relevant Cases. *In Proceedings of the EXPERSYS-97 Conference), Sunderland, UK*, pp 67--72.

Kontkanen, P., Myllymaki, P., Silander, T., Tirri, H. (1998). On Bayesian Case Matching. *In B. Smyth & P. Cunningham (Eds.), Advances in case-based reasoning, proceedings of the 4th european workshop (EWCBR-98), Vol. 1488*, pp. 13--24.

Lees, B., Corchado, J. (1997). Case-Based Reasoning in a Hybrid Agent-Oriented System, Department of Computing & Information Systems, University of Paisley, Paisley, Scotland.

McCallum, A., Nigam, K. (2003). A Comparison of Event Models for Naïve Bayes Text Classification. *Journal of Machine Learning Research 3*, pp. 1265-1287.

Myllymaki, P., Tirri, H. (1993). Bayesian Case-Based Reasoning With Neural Networks. *In Proceeding of the IEEE International Conf. on Neural Networks, San Francisco*.

Myllymaki, P., Tirri, H. (1994). Massively Parallel Case-Based Reasoning with Probabilistic Similarity Metrics. *Proceedings of the First European Workshop on Case-Based Reasoning*, pp 48--53.

Nigam, K., Lafferty, J., & McCallum, A. (1999). Using Maximum Entropy for Text Classification. *IJCAI-99 Workshop on Machine Learning for Information Filtering,* pp. 61—67.

Quinlan, J. R. (1993). *C4.5: Program for Machine Learning*. Morgan Kaufmann, San Mateo, Ca.

Sahami, M., Dumais, S., Heckerman, D., & Horvits, E. (1998). A Bayesian Approach to Filtering Junk E-Mail. *In AAAI-98 Workshop on Learning for Text Categorisation*.

Schiaffino, S. N., Amandi, A. (2000). User profiling with Case-Based Reasoning and Bayesian Networks, *IBERAMIA-SBIA 2000 Open Discussion Track*.

Su, X. (2002). A Text Categorisation Perspective for Ontology Mapping. Department of Computer and Information Science, Norweigian University of Science and Technology, Norway.

Von Wangenheim, C. G. (2000). Case-Based Reasoning – A Short Introduction, Universidade do Vale do Itajai.

Weber, R., Aha, D. W., Snadhu, N., Munos-Avila, H. (2001). A Textual Case-Based Reasoning Framework for Knowledge Management Applications. *In Proceedings of the Ninth German Workshop on Case-Based Reasoning*.

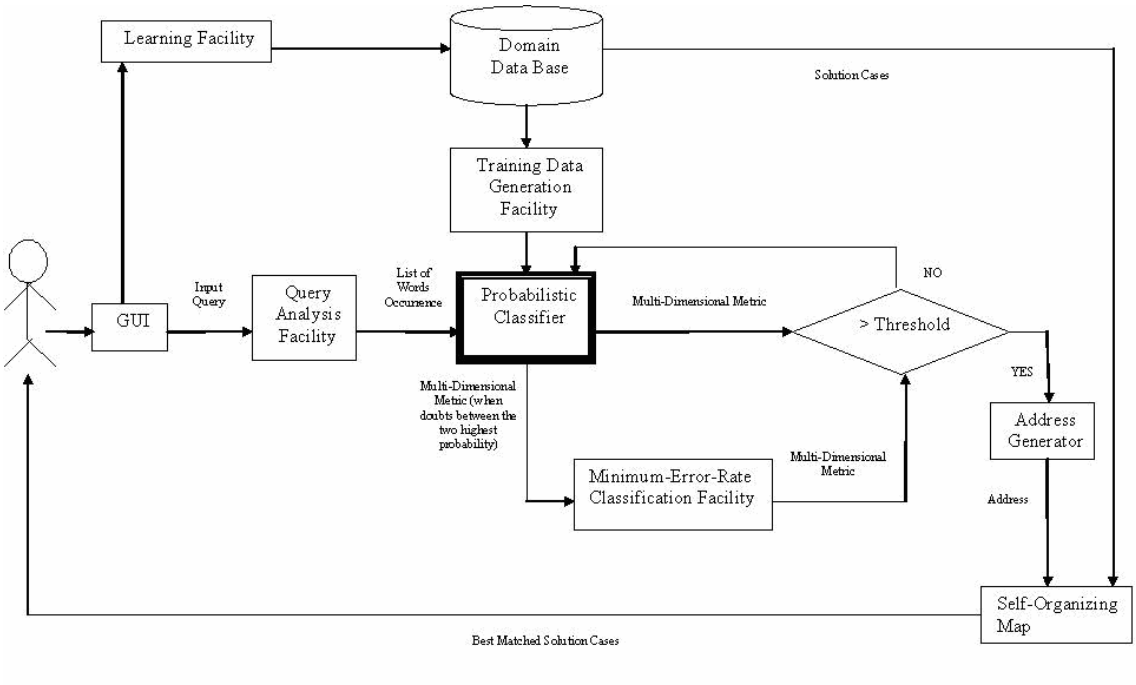



Figure 1. Proposed Enhanced CBR System's Block Diagram

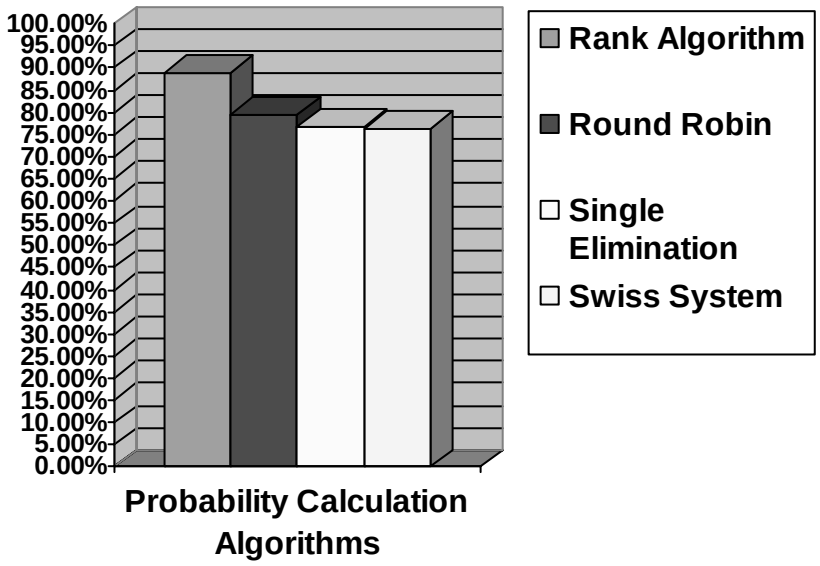



Figure 2. Comparison Chart for Performances of Different Classification Algorithms in Experiment 1

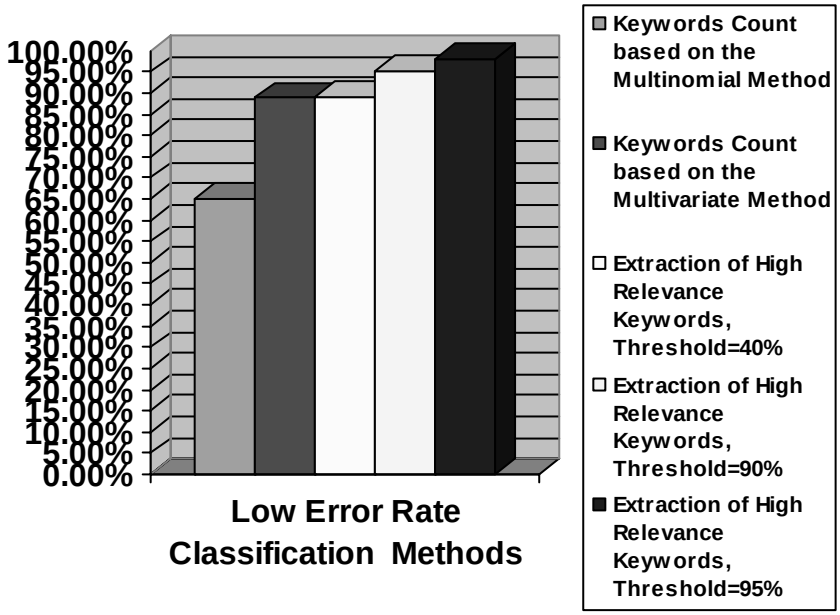



Figure 3. Comparison Chart for Performances of Different Low Error Rate Classification Methods in Experiment 1

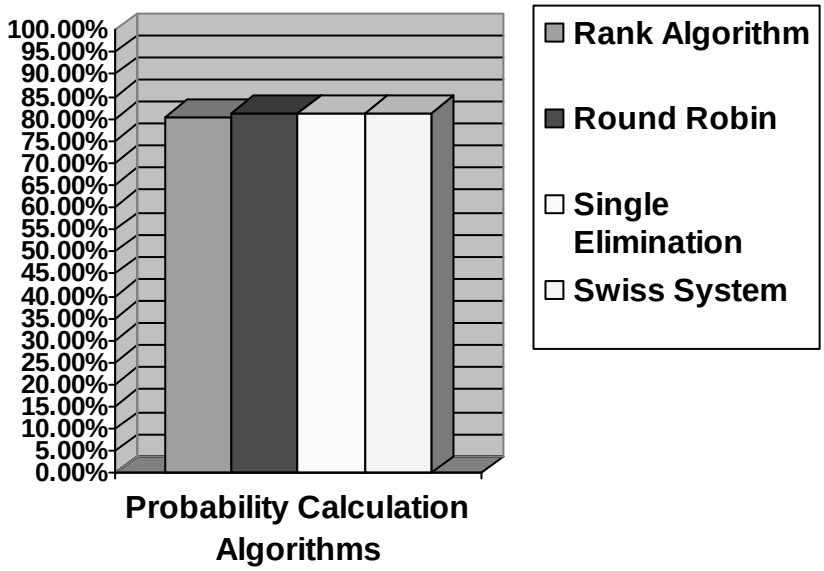



Figure 4. Comparison Chart for Performances of Different Classification Algorithm in Experiment 2

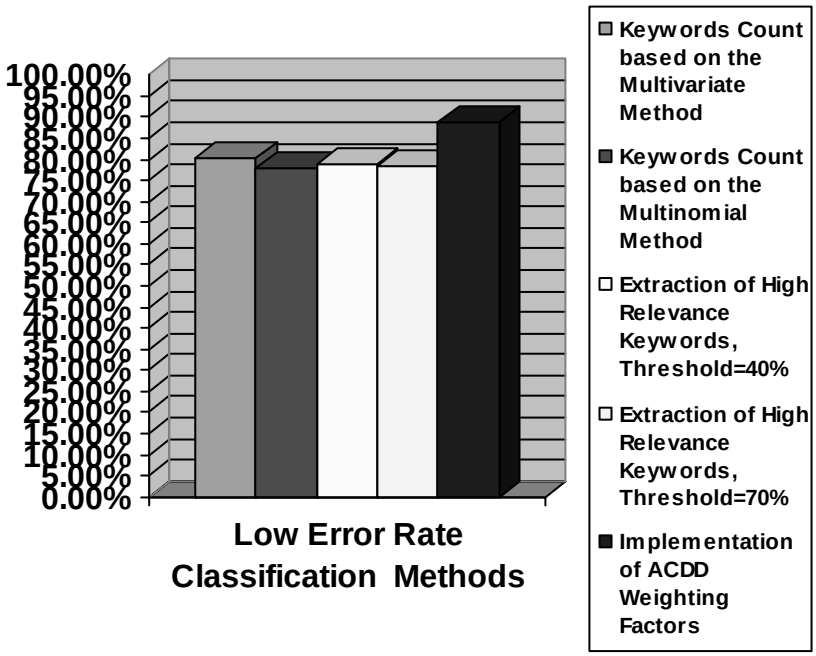



Figure 5. Comparison Chart for Performances of Different Low Error Rate Classification Methods in Experiment 2

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 1, No. 1

February 2008

# Disruption Tolerant Networking

Kevin Curran & James Knox

University of Ulster, Northern Ireland, UK

Tel: 44-2871 375565 E-mail: kj.curran@ulster.ac.uk

**Abstract**

Disruption Tolerant Networking (DTN) is a technology that will provide network services for environments so extreme that no end-to-end path exists through a network. Disruption Tolerant Networking hopes to tackle the problem of communicating in areas where due to various factors normal means of communication have limited success and are unreliable. DTN is aimed specifically at networks which are subject to frequent and long lasting disruptions that destroy or severely degrade normal communications. This paper provides an overview.

**Keywords:** Disruption tolerant networking, Computer networks, Communications, Wireless networks

## 1. Introduction

Delay Tolerant Networking (DTN) is an approach to computer network architecture that seeks to address the technical issues in mobile or extreme environments that lack continuous network connectivity. In a DTN, asynchronous variable-length messages (called bundles) are routed in a store and forward manner between participating nodes over varied network transport technologies. Disruption Tolerant Networking is important due to the importance of managing network congestion. Congestion is caused by heavy traffic load which can result in long delays in data transmission due to retransmission because of data loss. Currently end to end congestion control is handled by the TCP protocol that prevents the network from collapsing but network degradation does occur when the network becomes congested. TCP/IP works well when there is no disruption to end-to-end communication. Disruption Tolerant Networking addresses weaknesses in TCP such as implementing congestion control mechanisms where each router can make decisions based on local information such as accepting a bundle of data from another router. This local information may include storage availability and the risk of accepting the data bundle based on previous experience, which may have had adverse effects on the network resulting in loss of data. This paper provides a review of Delay Tolerant Networking.

## 2. Delay Tolerant Networking

Disruption Tolerant Networking allows messages to pass through the network with successive responsibilities, rather than the traditional end-to-end acknowledgement scheme (Riehn, 2004). The main thrust behind Disruption Tolerant Networking is survivability and resilience. The aim is to have a network that will work in environments where continuous end-to-end connectivity cannot be assumed. This could be due to environmental factors, underwater, in desert conditions, deep space, or during warfare. There are many benefits in such a system for example, in places that suffer from disaster or where normal communications are destroyed (Grover and Tipper, 2005).

One of the problems with TCP is that it has a short time-out counter, therefore if no data is sent within a short period the circuit is shut down. This makes it unsuitable as a protocol in deep space as end-to-end connectivity would not be possible. Disruption Tolerant Networking is better suited as it is delay tolerant. Disruption Tolerant networking organises information flow into bundles. Messages pass through networks with successive responsibility unlike the current TCP structure that uses end to end acknowledgements. Intelligence is moved into the network to allow each network to make best choice on delivery of bundles via the optimal path.

Flow control is used to ensure that the destination can handle all incoming data. Congestion control is used when buffers in routers are under pressure due to the limitations of the buffer size. It limits the data loss within a network by inducing flow control which is automatically triggered when there is a difference in the arrival rate and transmission rate of data within the network. When the rates differ TCP will send an acknowledgement to the source which will reduce the transmission rate until transmission rate is equal to the arrival rate. This works well for the internet with end to end connectivity however where end to end connectivity cannot be guaranteed due to external factors this causes problems as automatic flow control and congestion control cannot work under the same set of rules. Unreliable connections make it difficult to ascertain when action is required to limit flow control or when to initialise congestion control (Grover and Tipper, 2005). The only way around this problem with DTN is to initiate congestion control by disregarding bundles due to resource depletion and returning bundles if, the management of local buffer space is under

threat. In this case bundles would be returned to their destination address until the problem was rectified. The management of this system would require the sending application to issue differing TTLs with every bundle being sent depending on their importance and the conveyance of this bundle would be on the basis of the time to live being met. The downside of this would be to accept all bundles with a large TTL; however this has the disadvantage of depleting router buffer space for a long time and all the inherent problems this brings with it.

To help reduce the problems with TTLs and bundles being dropped, DTN implements a custodial transfer system to allocate preference to certain types of bundles that meet certain criteria, allowing for a more efficient system with the conveyance of high priority bundles at the expense of others. As high priority bundles would be forwarded faster than lower priority bundles it would mean that they would have a far better chance of reaching their destination before being timed out, this means that end to end latency would be lower allowing the added benefit of reducing the TTL at every point on the path and therefore increasing the delivery time, and reducing congestion. If a bundle meets the requirements of the congestion control algorithm which is applied to all bundles and if the inbound bundle is flagged for custody transfer then a custody acceptance message is sent back to the current custodian. This causes the bundle to be removed from the custodian's buffer freeing up buffer space and relieving congestion control (see Figure 1).

If a custody transfer bundle is refused then a custody refusal message is sent back to the current custodian along the originals sender route, this retriggers the resending of the bundle usually via a different route. At the same time, the absence of a custodial acceptance message triggers congestion control. Here congestion pressure at the current custodian remains unrelieved because the custodial bundle remains in the custodian's buffers, causing net space to increase over time (Burleigh et al., 2004). The problem with this is how one decides on the custodial timeout because unlike TCP/IP that has an end to end connection it is far harder to gauge the time of a round trip using the custodian method. However the benefit of the DTN based network systems is that unlike the protocol TCP/IP which requires a continuous end to end connection in order to work, the Custodial transfer system allows for breaks to occur in the connection of the end to end path which will not affect the data transmission as illustrated in Figure 2.

For example, if node A was trying to communicate with node E through nodes B, C and D then with the TCP/IP protocol this would not be possible as there is no continuous connection. However with DTN node C acts like a waterwheel first scooping the data from node B then releasing it to node D. This allows for there to be a break in connection between nodes C and D while the connection between B & C is open. Therefore it allows connections to be maintained even though connections to B & C and C & D were not opened concurrently. Thus the network integrity is maintained for bundle transfer to take place without the need for continuous end to end connection.

## 3. Conclusion

DTN based network systems are ideal for Interplanetary Networks which allow for long-haul communication capabilities where a continuous end to end link is not sustainable. Other places where DTN can be used include spacecraft, military/tactical, some forms of disaster response, underwater, and some form of ad-hoc sensor/actuator networks (Akyildiz, 2003). It may also include Internet connectivity in places where performance may suffer such as developing parts of the world.

## References

Akyildiz, I., Akan, O., Chen, C., Fang, J., and Su, W. (2003). InterPlaNetary Internet: State-of-the-Art and Research Challenges, *Computer Networks Journal*, Vol. 43, Issue 2, pp. 75-113, October 2003.

Riehn, M. (2004). Disruption Tolerant Networking. Available at:
http://www.darpa.mil/sto/solicitations/DTN/proposers.htm Accessed [10th March 2007].

Burleigh, S., Jennings, E. and Schoolcraft, J. (2007) Autonomous Congestion Control in Delay-Tolerant Networks, American Institute of Aeronautics and Astronautics.

http://pdf.aiaa.org/preview/CDReadyMSPOPS06_1317/PV2006_5970.pdf [15th March 2007].

Grover, W. and Tipper, D. (2005). Design and Operation of Survivable Networks, Journal of Network and Systems Management, *Special Issue on Designing and Managing Networks and Service Reliability*, Vol. 13, No. 1, pp. 7-13.

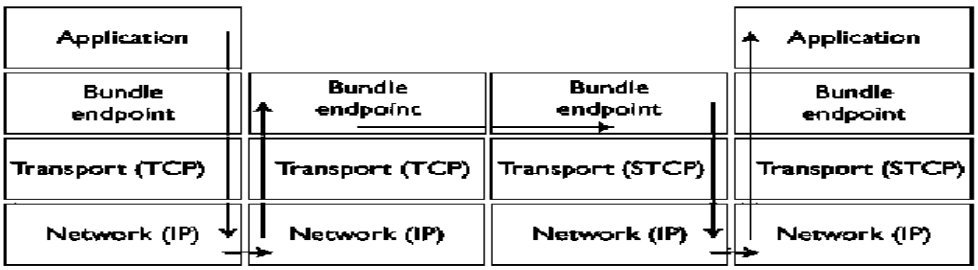



Figure 1. High priority

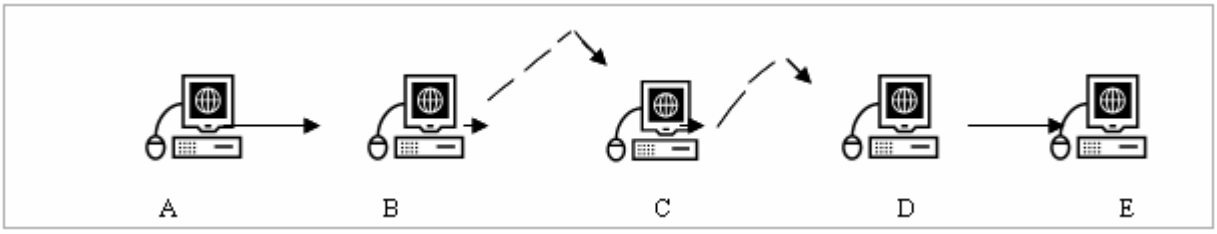



Figure 2. Custodial Transfer System

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# The Social Dynamic of Health Disclosure

# Who Do Patients Tell and When?

Nasriah Zakaria
School of Computer Sciences, Room 726 Building DG31
Universiti Sains Malaysia, 11800 Minden, Penang, Malaysia
Tel: 6(019)4106826 E-mail: nasriah@cs.usm.my
Jeffrey M. Stanton
School of Information Studies, Hinds Hall 316
Syracuse University Syracuse, NY USA 13244-4100
Tel: 1(315) 443-2879 E-mail: jmstanto@syr.edu

**Abstract**

This paper focuses on social dynamics among disclosure communities in terms of balancing disclosure and privacy. Why does a man tell his mother but not his father when facing a medical crisis? This is an example of the core question in this study. To better understand these situations, we need to identify the disclosure activities that take place within the patient and their disclosure community when dealing with illness, and explore the different ways patients balance the need to disclose and the need to be private. The outcome of this study is the development of a conceptual framework (in the form of conceptual map) that describes the social dynamic of health disclosure for chronically ill patients. Concepts like balancing act, which is being closed (private) and at the same time being open (disclosure), timing of disclosure, and role of disclosure emerged as three patterns that emerge during the disclosure process.

**Keywords:** Disclosure, Health Information, Health Communication

## 1. Introduction

### *1.1 The Concept of Self-Disclosure*

"The concept of disclosure is clearly related to notions of privacy and secrecy. If privacy concerns keeping things hidden, and secrets are the specific messages chosen not to be shared, then disclosure is the process that grants access to private things and to secrets" (Rosenfeld, 2000, page 6).

The interrelationship between privacy and disclosure is at the heart of this study. The main assumption is that patients balance the need for disclosure and the need for privacy when dealing with health information relative to their disclosure communities (DC). Petronio (1991) developed the boundary metaphor to describe how humans build protective walls around their sensitive information and how they balance disclosure and privacy.

## 2. Literature Review and Background Theory

### *2.1 Self Disclosure*

Self-disclosure is defined as "any message about self that one communicates to another" (Cozby, 1973). The topic has attracted scholars from diverse fields – communications, information studies, social psychology, sociology, and psychology. Northouse and Northouse (1998) presented five variables in health communication, where disclosure is one of the central variables that determine the effectiveness of communication. The authors discussed disclosure as a variable since the presence (or absence) of disclosure facilitates (or impedes) effective health communication among patients and medical teams.

Self-disclosure is a communication process whereby people transmit personal information to others; in exchange they may receive feedback or help from others. These messages about self may consist of facts or of personal thoughts, feelings, and experiences, and may refer to the present or the past. Empirical studies on the nature of the self-disclosure

message (Cozby, 1973; Wheeles & Grotz, 1976) identified five attributes: intention, amount, valence, honesty and depth. Intention is concerned with an individual's willingness to disclose to others. Amount refers to the quantity of information exchanged. Valence indicates the positive or negative content of the message; if a person discloses positive content, the valence was positive, and the valence is negative when the content is perceived to be negative. Honesty refers to the factual accuracy of the message while depth refers to the level of intimacy of the message conveyed. (As an example, prognosis information has more depth than general health information).

Most of the earlier works on self-disclosure were in the form of reviews and lacked a theoretical framework (Berg & Derlega, 1987, page 2). However, in recent years, researchers have begun to develop theories that attempt to predict and/or explain self-disclosure behaviors. Among these theories are Communication Boundary Management and Communication Privacy Management (Petronio, 2002), and Uncertainty Reduction Theory (Berger & Calabrese, 1975). These theories looked at disclosure under three major themes: (a) the role of self-disclosure in treating psychological distress (Berg & Derlega, 1987), (b) the role of self-disclosure in building, maintaining or destroying relationships, and (c) self-disclosure as a personality/intrapersonal trait. This study takes the second approach, exploring the formation, maintenance, and disintegration of relationships (specifically, disclosure of health information among patients and other people). Before venturing into a discussion of the role of self-disclosure in relationships, however, a brief description of the body of empirical studies that have been done in the therapy and personality areas is in order, since these works provide background in the overall disclosure research.

Early research in self-disclosure focused on the benefits of self-disclosure to well-being and health – mainly psychological or mental health. Sydney Jourard, a psychologist who studied self-disclosure in the context of mental health conditions (1964, 1968, and 1977), found strong indications that self-disclosure promotes healthy interpersonal relationships. A great deal of Jourard's research looked at the self-disclosure variable from a relational perspective, wherein he examined how disclosure helps individuals in general. He found that the process of self-disclosure is linearly related to an individual's mental health. Without disclosure, individuals feel lonely and estranged from the social world; with it they can interact and become close with others. However, several other studies (Chaikin & Derlega, 1974a, 1974b; Cozby, 1973) challenged Jourard's conclusions, postulating a curvilinear relationship between self-disclosure and mental health condition: both extremely low and extremely high levels of self-disclosure correlated with a low level of personality adjustment. In other words, people with problems reveal either a great deal more, or a great deal less information about themselves than "normal" people, a finding which has contributed to treatment of psychological distress.

Derlega and Chaikin (1974, 1977) proposed two factors that mediate between self-disclosure and health conditions: the nature of the relationship, and situational context. The nature of the relationship comes into play because research indicates that people tend to disclose to those with whom they have a close relationship, such as spouse, immediate family and close friends. People also evaluate the situational context in deciding whether and what to disclose to others. For example, a person who knows that he will need a great deal of time off for chemotherapy will disclose his medical state to his manager and immediate subordinate so someone can take over his responsibilities at work. However, he may not disclose to his employers if his medical condition is mild and under control. In a later compilation of self-disclosure work, Rosenfeld (2000) suggested four factors that influence a disclosure decision: nature of relationship, risk-to-benefit ratio, situational context and reciprocity. Nature of relationship and situation have already been described. The risk to benefit ratio refers to an individual's assessment of the benefits of self-expression, self-clarification, and catharsis as compared to the risk of rejection, humiliation, and possible avoidance by other people. Reciprocity proposes that once a sensitive disclosure is made, the recipient is expected to respond by providing some piece of sensitive information of his own to the discloser, either then or at a later time.

The research that views self-disclosure as a personality trait assumes that self-disclosure is an individual choice. Under this assumption, people constantly reassess the risk-to-benefit ratio before making a disclosure. Research in this area has identified various specific benefits and risks and suggests that individuals do make their own decisions on what information to disclose and what information to withhold from others (Wenburg & Wilmot, 1973). In the context of health communication, benefits of self-disclosure are catharsis, receipt of feedback from others, and help with problem solving (Cozby, 1973; Northouse & Northouse, 1998). It can also improve patient care: Petronio and her colleagues (2004) found that family members who accompanied patients during medical interviews disclosed patients' personal information in order to clarify, correct or refute the physician's statements. Identified risks of inappropriate disclosure include maladjustment (Chaikin & Derlega, 1974a; Wortman et al., 1976), social stigma of showing weakness and vulnerability (Petronio, 2000), and the loss of personal relationships (Northouse & Northouse, 1998). For example, Cline and McKenzie (2000) found that HIV patients who disclosed their status faced embarrassment and humiliation because of HIV's perceived association with illegal drug use and homosexuality. Block (2000) found that some cultures regard certain medical conditions (particularly mental disorders) as taboo, and patients who disclose such illnesses to other people usually get negative reactions.

Researchers who studied self-disclosure as an interpersonal or social relationship aspect found several ways in which self-disclosure can enhance social relationships. Altman and Taylor's (1973) Social Penetration Theory (SPT) explains the relationship between self-disclosure and intimacy by proposing that a high level of disclosure increases the intimacy level between discloser and recipient. In addition, SPT also explains the formation, maintenance, and disintegration of relationships. Uncertainty Reduction Theory, on the other hand, proposes that the function of disclosure is to reduce human uncertainty under various circumstances. For an example, when the victim of sexual assault discloses the identity of the culprit, it reduces the uncertainty for her family and friends.

Disclosure reciprocity is an important concept in understanding why and how people in a relationship disclose to each other. As stated above, reciprocity means that the recipient of a disclosure is expected to exchange similar information about himself, either immediately or at some later (but not too distant) time. This reciprocity may serve to express liking for and trust in the discloser, to accord with social norms, or to follow or imitate the behavior of discloser. Dindia (2000) performed research in existing work on self-disclosure in social relationship aspects and found that reciprocity occurs at all stages of a relationship, from the beginning or introductory stages to more intimate phases.

Scholars such as Westin (1967, 2003) and Altman (1977) studied self-disclosure from the other side, in terms of privacy. In fact, Altman's work linked the privacy concept closely to self-disclosure and his work was followed by extensive research by Derlega & Chaikin (1977) and by Petronio (1991, 2000). Petronio relied on the "boundary" metaphor originated by Altman and from there developed her Communication Boundary Management (CBM) theory. CBM offers a new perspective on self-disclosure in two ways. First, Petronio (1991, 2000) defined the concept of "boundary" as the border surrounding private information. The boundary is actually a psychological construct created by human beings in order to manage their information. Many previous research studies viewed self-disclosure as simply revealing information about oneself to others (Berg & Derlega, 1987), but Petronio's work looked at both disclosure (revealing) and privacy (not revealing) at the same time. The second important contribution of CBM was the dialectical perspective of self-disclosure and privacy, wherein humans constantly try to balance openness with privacy in order to achieve, respectively, relational intimacy or autonomy.

In sum, then, the literature lays out three approaches to studying self-disclosure: the role of self-disclosure in psychotherapy, the role of self-disclosure in social relationships, and the personality or intrapersonal aspect. Self-disclosure has been shown to benefit physical health, mental health, interpersonal communication, and social relationships. Because the main focus of my research is on social relationships (second aspect), we look at the roles of each party – the transmitter of information and the recipient – rather than at the personality trait or mental health aspect. The social relationship aspect of disclosure can shed light on how disclosure can create, maintain or disintegrate social bonds among individuals. We wanted to see how disclosure activities contribute to the balancing act of openness versus privacy about health information. In a social relationship perspective, disclosure occurs based on what other people think, say and do. For example, one person may decide to disclose her problem when she is surrounded by people who are talking about their problems. Another person may choose to disclose to her friends because her friends ask her to clarify gossip that they have been hearing.

Many theories have been advanced to explain how disclosure functions in relationships as well as how individual disclosure decisions are made. We selected Communication Boundary Management Theory (CBM) as the foundation for this study in order to help guide the exploration of disclosure dynamics among three stakeholder groups: patient, disclosure community and medical team. The reasons for our choice of CBM as the basis for this study are that it deals extensively with the issue of balancing disclosure and privacy in relationships, and has been used to explain various contemporary health communication issues.

### *2.2 Self-Disclosure in Health Care Situations*

Numerous studies have examined disclosure of health information. These studies looked at the content, frequency and consequences of self-disclosure in different medical situations and raised different issues of disclosure, but frequently ignored the balance between disclosure and privacy issues (Cozby, 1973; Petronio, 1991). Wiener et al. (1996) explored the process of disclosure and consequences among HIV patients and their support network. It examined the caregiver-patient dyad it did not discriminate between the different roles of caregivers during the coping period, thus we are unable to conclude what specific roles are crucial for effective communication with the dyad. Another study (Baird, McConachie & Scrutton, 2000) explored how family and caregivers cope with cerebral palsy, and found that support network members were dissatisfied with the information disclosure activities. That study also determined that factors such as the age of the child, the timing of diagnosis and the severity of physical disability influenced information disclosure. The study suggested "best practice" guidelines for disclosing the diagnosis of an illness to the support network. Contro et al. (2002) showed that some family members such as siblings were often not included during the information disclosure process; this in turn caused the siblings to feel "left out" of the coping process.

A few studies pointed out the lack of empirical work in understanding self-disclosure among medical teams and family in chronic illness situations (Bradley et al., 2001; Contro et al., 2002). Families also made it clear that they wanted more

information from the medical teams in order to be able to provide the appropriate support for the patient (Bradley et al., 2001; Holroyd, Turnbull, Wolf, 2002). Other recent research on self-disclosure in healthcare examined different situations, contexts and issues as well as the nature of relationships. Topics examined include the process of disclosure (e.g. content and frequency) and its consequences (Holroyd, Turnbull, Wolf, 2002), the dyad relationship during disclosure (Wiener et al., 1996, Nielsen. 1998), practices for disclosing illness (Baird, McConachie & Scrutton, 2000), disclosure of chronic illnesses (Bradley et al., 2001; Contro et al., 2002), functional perspectives on health disclosure (Derlega, Winstead, Folk-Barron, 2000), privacy and disclosure (Welch & McKenzie, 2000), and the influence of illness, relationship and information-seeking on disclosure (Brann, 2003). Recently, self-disclosure research in healthcare has begun to pay attention to the larger group of people involved in patient care – friends, employers, spiritual groups and self-help groups. Petronio and her colleagues (2004) explored the role of these "informal advocates" by examining the self-disclosure behavior of family and friends when they were present during a patient's visits with the physician and/or medical team.

All the empirical work mentioned above informs the question of disclosure in various medical situations. However, it is clear that little work has been done on the social dynamics in disclosure communities when facing illness: as such, it is a good candidate for exploration using grounded theory, which is ideal for discovering conceptual model in areas about which little is known.

## 3. Method

This study explored the dynamics of disclosure communities by examining data from patients, medical team members, and disclosure community (family, sexual partners, neighbors, spiritual advisors, etc). Semi-structured interviews were conducted to gather stories regarding patients' experiences in deciding on communication rules for the disclosure process during important health events. Semi-structured interviews provide a beneficial and productive method of gathering data for this research because of the sensitivity of the topic, the likely complexity of the social dynamics, and the absence of structured interviews protocol to this topic area. Because of the need to speak with individuals who recently experienced an important health event, we employed the snowball-sampling technique which asks one respondent to refer another person who might also be willing to share his or her past disclosure experiences in a health care context.

The sampling was helpful in locating individuals who met the eligibility criteria: i.e., centered on a patient who had been admitted to the hospital for treatment of an illnesses, and who were thus likely to have encountered HIPAA-driven privacy policies while receiving treatment. These patients could also share their experiences in interacting with their disclosure communities during those times. We also conducted semi-structured interviews with medical team members – doctors and nurses – to elicit their experiences managing patients' disclosure needs and preferences. Finally, the third group of respondents comprised individuals who shared and managed a patient's health information when the patient faced illness. These individuals included family members, friends, sexual partners, non-profit organization members, and other support group members. The results of this study describe the social dynamics of the disclosure communities and provide understanding of how patients decide whether or not to tell their health information to others.

The overall goal was to develop a new conceptual model that explicates social dynamics of disclosure communities through the use of grounded theory (Glaser & Strauss, 1967; Strauss & Corbin, 1990). One of the best methodological approaches to building such explanations is "grounded theory." Grounded theory is a systematic approach to data collection and analysis that aids the researcher in seeing emergent patterns in narrative data (Glaser & Strauss, 1967; Strauss & Corbin, 1990). Currently, there are two modes of thinking in doing grounded theory: "Glaserian" and "Straussian." The difference between the two lies in how each of them views the source data.

Glaser believed in developing emergent theory from the data itself, while Strauss questioned the data initially, before conceptualizing or abstracting it to a higher level. Glaser's positivist background makes him objectivist toward data; he is a neutral observer. Strauss brought his social science background, wherein he gave voice to the respondents. He emphasized the interaction between researcher and respondent and added an interpretive stance when building data.

In this study, we viewed the disclosure topic using the dialectic view of disclosure vs. privacy. This means that we had some preconceived ideas about the subject matter, so whenever we had data at hand, we had some biases. In addition, our overall conceptual framework was influenced by the constructs offered in CBM theory. For these reasons we chose to use Strauss and Corbin's version of grounded theory approach, which admits of the presence of an existing theory.

From the data, we developed a conceptual model able to explain the social dynamics of disclosure derived from what respondents say about the situations they had experienced. During data collection, we simultaneously performed various data analysis activities such as writing memos and creating conceptual labels or codes. Codes coalesced into categories which eventually became concepts that fit the situation and helped to explicate the process that respondents were going through (Glaser & Strauss, 1967). The final product is a conceptual model which explains the emerging concepts and the relations among them, illuminating the social dynamics of disclosure communities.

**4. Results**

The respondents for this study comprised medical patients, medical survivors, physicians, nurses, social workers, family members and clergy. All twenty-three people have been directly involved in patient care either in hospitals, nursing homes or residential. Table 1 shows the overview of the respondents sample for this research.

One of the important tasks in grounded theory approach is to reduce the conceptual findings to one core category, which represents the main storyline of the research. The storyline for this study revolves around the social dynamics in disclosure communities. The conceptual map (Figure 1) explains and describes two important factors – stages of illness and role of disclosure community (DC) – and shows how they influenced the balancing act between disclosure and privacy of health information

The map is divided into three sections. The arrows indicate the flow of disclosure reported by patients, supported by DC responses. The first box on the left, "Balancing act," represents an individual's decision process in choosing to be open or closed about his health condition. In this research, the data demonstrated that individuals make this decision by considering both respondent's and receiver's social interactions.

Four arrows flow from the balancing act. Three of them lead into boxes that represent the three stages of illness (diagnosis/checkups, treatment/surgery, and post-treatment), indicating the influence that illness stage has on disclosure behavior. For example, a person chooses to disclose her illness when she is the recovery stage because she accepts her illness then, when she did not before. Other individuals choose to be private even in the recovery stage because they are still distressed with their current situation. The fourth arrow leads to the role of disclosure community (DC), indicating the influence of the three possible roles found in members of the DC: social support, advocate and medically literate.

From each of these four factors that influence a patient's disclosure decisions – the three stages of illness and the role of DC – one or two arrows lead out, to typical open behaviors associated with that factor (e.g., asking for prayers), and/or to typical closed behaviors (e.g. refusing to talk about it). Note that "Special Roles" leads only to open behaviors, indicating that those roles only served to encourage open behavior

The other important dimension in this conceptual map is the time factor (arrow at the bottom of the Figure 1). Chronic illnesses span a long period of time and this time span influences the evolution of an individual's disclosure process. For instance, the OPEN and CLOSED disclosure behaviors indicated by the boxes on the far right change across time as the individual is constantly balancing being open or closed to others The outcomes of these disclosure decisions (e.g., negative reaction from boss) feed back into the system, thus influencing future decisions; this iterative process is represented by the feedback arrows from these behaviors into the balancing act box.

The three stages of illness – diagnosis and checkups (T1), treatments (T2), and post-treatment (T3) – occur across time (time). Based on the limited data, the post-treatment stage comprises of several possible outcomes: recovery, still suffering ("battling") or dying. For example, some of the respondents suffered their illness for more than twenty years, and most of that time they were in the treatment stage, while other respondents suffered for 10 years, mostly in the post-treatment stage. Even though each person suffered illness for a different length of time, they went through each of the three stages at least once during the disease course.

The conceptual map gives an overview of typical disclosure behaviors. Overall, during the first stage of illness – diagnosis and checkups – respondents discussed their medical diagnosis in detail with others. Most respondents reported that they were open to people who were sympathetic and supported them emotionally, or to those who helped them with decision-making and who provided them with information. A few respondents, however, felt upset and confused with the situation, and so tried not to discuss their health issue with many people during this stage. In the second or treatment and surgery stage, some respondents were compelled to disclose their health condition at work and in other social settings. Some had to take time off for treatment, some had to quit their job and some asked for the community to pray for them during this critical stage. Others remained silent, avoided seeing others and/or provided only minimum information to others as a way to keep their health information private.

The third stage could be recovery, still suffering ("battling") or dying. In this stage, many respondents felt a responsibility or desire to use their first-hand experiences, to tell others about their illness in order to educate and give social support. Some of them even became advocates for new patients by accompanying them to doctor's appointments and dealing directly with the medical team. Others chose to remain private about certain aspects of their illness, either because they were still upset or felt they didn't have enough information about their prognosis. All three stages of illness can yield either open or closed behaviors, indicating that individuals each perform their own balancing act in choosing open and closed behavior.

The special roles played by members of the DC indicate the social interaction that takes place between patient and DC and which influences the disclosure behavior. Four broad categories of DC roles emerged (see "Special Roles in DC" at upper left in Figure 1). Note that these special roles of DC only influence "Open" behavior; the special roles only appeared as factors that enhance the disclosure activities. There was not enough data to build up categories of "negative

roles" that might encourage closed behavior. This is indicated by an arrow from Special Roles to open behavior, but no arrow from Special Roles to closed behavior. On the right side of the diagram, bubbles inside the categories of "Open" and "Closed" summarize the different behaviors found in the data. Feedback arrows from these open and closed behaviors to the balancing act box indicate the constant reevaluation of disclosure decisions, as mentioned above in the discussion on Time as a factor.

The essence of the conceptual map is to show that disclosure processes occur across time and that an iterative balancing between disclosure (open) and privacy (closed) takes place, in which decisions are made not once but again and again. The balancing for each individual occurs across time, and is influenced by each stage of illness that they pass through.

The researcher's (first author) role as an interviewer fits into this map. The qualitative nature of this study means that the researcher is involved in the social dynamics of the disclosure process, since the respondents had to disclose to the researcher (first author) about their disclosure behaviors with others. Through snowball sampling technique, trust is established through the use of a mutual social network; this allowed the respondents to accept the interviewer as part of their disclosure community at that point in time. In addition, given the sensitive and emotional nature of the topic, the researcher becomes a social supporter to the respondents for the duration of the interview (for example, the researcher made sure that she listened carefully and showed sympathy). The researcher, then, by being an active, respectful listener genuinely interested in the respondent's experience, falls into the "Special Roles" category, specifically the "social supporter" subcategory. The relationship is admittedly brief, but nevertheless clearly present. The researcher also has developed friendship relationship with a few of the respondents and asked their well being from time to time.

## 5. Conclusion of study

Information privacy is a term that has many meanings to people. Most people will say they are open about their situation and that they have nothing to hide, but in reality individuals make conscious decisions which information to share (or not to share) with others. For medical patients, health information lets them understand what is going on and helps them feel more confident about informing selected individuals. Patients gained a positive outcome when they disclosed to others suggesting that the importance of a disclosure community to patient well-being. The lessons learned from this study support the idea that patients need to be able reach out to others by sharing their information so that others know how and when to provide help to patients.

The medical teams also play an important role in the disclosure process, which is to provide timely and accurate information to their patients without discrimination. The medical team must be aware that the information that they provide can help to inform, clarify and provide assurance to patients and their DC. In addition, the medical team needs to know how to explain the medical situation in general terms so that patients are able to fully understand their health situation and explain it to others.

Finally, this study emphasizes that awareness of health communication among the three parties is a crucial aspect of quality health care. The medical team who serve as the information broker can foster an atmosphere of openness by providing good information to patients so that the patient can disseminate information to their DC. This way, DC will be able to provide the appropriate support to patients. In addition, the medical team also can be proactive, giving ideas and guidelines to the DC on how they can directly and indirectly help patients in order to improve the quality of care.

## References

Altman, I. (1977). Privacy regulation: Culturally universal or culturally specific? *Journal of Social Issues,* 33 (3), 66-84.

Altman, I. & Taylor, D. (1973). *Social Penetration: The Development of Interpersonal Relationships*. NewYork: Holt, Rinehart and Winston.

Baird, G., McConachie, H., Scrutton, D. (2000). Parents' perceptions of disclosure of the diagnosis of cerebral palsy. *Archives of disease in childhood,* 83(6), 475-480.

Berger, C. & Calabrese, R. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research,* 1, 99-112.

Berg, J.H., Derlega, V.J. (1987). Chapter 1: Themes in the Study of Self-Disclosure. In V.J. Derlega & J.H. Berg (Eds), *Self Disclosure: Theory, Research, and Therapy* (pp. 1-7). Plenum Press, NY.

Block, S.D. (2000). Assessing and Managing Stress in the Terminally Ill Patient. *Annals of Internal Medicine,* 132,209-218.

Bradley, E.H., Hallemeier, A.G., Fried, T.R., Johnson-Hurzeler, R., Cherly S., Kasl, S.V., Horwitz, S.M. (2001). Documentation of discussions about prognosis with terminally patients. *The American Journal of Medicine*, 111(3), 218-223.

Brann, B.M.S. (2003). Assessing the influence of illness, relationships, and information-seeking strategies on physicians' disclosure of confidential health information to relatives of patients. *Dissertation Abstracts International. 64*(07A), 2309. (UMI No. AAI3097381.

Chaikin, A.L. & Derlega V.J. (1974a). Liking for the norm-breaker in self-disclosure. *Journal of Personality,* 42, 117-129.

Chaikin, A.L. & Derlega V.J. (1974b). Variables affecting the appropriateness of self-disclosure. *Journal of Consulting and Clinical Psychology,* 42s, 588-593.

Contro, N., Larson, J., Scofield, S., Sourkes, B., Cohen, H. (2002). Family Perspectives on the quality of pediatric palliative care. *Archives of pediatrics & adolescent medicine,* 56(1), 14-19.Cozby, P.C. (1973). Self-disclosure: A literature review. *Psychological Bulletin,* 79(2), 73-91.

Derlega, V.J., Chaikin, A.L. (1977). Privacy and self-disclosure in social relationships. Journal of Social Issues, 33(3), 102-117.

Dindia, K. (2000). Chapter 2: Sex Differences in Self-Disclosure, Reciprocity of Self Disclosure, and Self-Disclosure and Liking: 3 Meta Analyses Reviewed. In S. Petronio (Ed.), *Balancing the secrets of Private Disclosures* (pp. 21-35). Mahwah, NJ: Lawrence Erlbaum Associates.

Glaser, B.G., Strauss, A.L. (1967). *Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine.

Holroyd, S., Turnbull, Q., Wolf, A.M. (2002). What are patients and their families told about the diagnosis dementia? Results of a family survey. *International journal of geriatric psychiatry,* 17(3), 218-221.

Nielsen, S.E. (1998). Confidentiality in the nurse-patient relationship (Ethics). *MAI, 36*(04), 1067(UMI No. AAG1388817).

Jourard, S.M. (1964). *The transparent self*. Princeton, NJ: Van Nostrand Reinhold.

Jourard, S.M. (1968). *Disclosing man to himself*. Princeton, NJ: Van Nostrand Reinhold.

Jourard, S.M. (1977). *The transparent self* (2nd ed). Princeton, NJ: Van Nostrand Reinhold.

Northouse L.L, Northouse, P.G. (1998). *Health Communication: Strategies for Health Professionals* (3rd ed). Appleton & Lange.

Rosenfeld, L.B. (2000). Overview of the Ways Privacy, Secrecy and Disclosure Are Balanced in Today's Society. In S. Petronio (Ed.), *Balancing the secrets of Private Disclosures* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Petronio, S. (1991). Communication Boundary Management: a theoretical model of managing disclosure of private information between marital couples. *Communication Theory,* 1,311-335.

Petronio, S. (2000). Chapter 3: The Boundaries of Privacy: Praxies of Everyday Life. In S. Petronio (Ed.), *Balancing the secrets of Private Disclosures* (pp. 37-49). Mahwah, NJ: Lawrence Erlbaum Associates.

Petronio, S. (2002). *Boundaries of Privacy: Dialectics of Disclosure.* SUNY series in communication studies.

Petronio, S. Sargent, J., Andea, L. Reganis, P. Cichoki, D. (2004). Family and friends as healthcare advocates: Dilemmas of confidentiality and privacy. *Journal of Social and Personal Relationships.*21 (1):33-52.

Strauss, A., Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* Sage Publications Inc.

Welch Cline, R.J., McKenzie, N.L. (2000). Chapter 5: Dilemmas of Disclosure in the Age of HIV/AIDS: Balancing Privacy and Protection in the Health Care Context. In S. Petronio (Ed.), *Balancing the secrets of Private Disclosures* (pp. 71-82). Mahwah, NJ: Lawrence Erlbaum Associates.

Wenburg, J.R., Wilmot, W.W. (1973). *The personal communications process*. New York: Wiley.Westin, A.F (1967). *Privacy and Freedom*. New York: Atheneum.

Westin, A.F. (2003). Social and Political Dimensions of Privacy. *Journal of Social Issues,* 59(2), 431-453.

Wortman C.B., Adesman, P., Herman, E. & Greenberg, R. (1976). Self-disclosure: An Attributional Perspective, *Journal of Personality and Social Psychology,* 33, 256-266.

Wheeles, L.R. & Grotz, J. (1976). Conceptualization and Measurement of reported self-disclosure. *Human Communication Research,* 2(4), 238-246.

Wiener, L.S., Battles, H.B., Heilman, N., Sigelman, C.K., Pizzo, P.A. (1996). Factors associated with disclosure of diagnosis to children HIV/AIDs. *Pediatric AIDS and HIV infection,* 7(5), 310-324.

Table 1. Overview of Respondents

| Research Respondents (Total: 23 people) | |
|---|---|
| Medical Patient/Survivor (9) | Patients (5)<br>Diabetes (1)<br>Chronic Obstructive Pulmonary Disease (1)<br>Heart Disease and Depression (1)<br>Renal Failure (1)<br>Chronic Fatigue Syndrome (1)<br><br>Survivors (4)<br>Breast Cancer Survivor (3)<br>Thyroid Cancer Survivor (1) |
| Disclosure Community (8) | Family (6)<br>Daughter (5)<br>Daughter and Legal Guardian (1)<br><br>Non-family (2)<br>Spiritual (2) |
| Medical Team (6) | Medical Professionals (5)<br>Physicians (4)<br>Nurse (1)<br>Other (1)<br>Social Worker (1) |

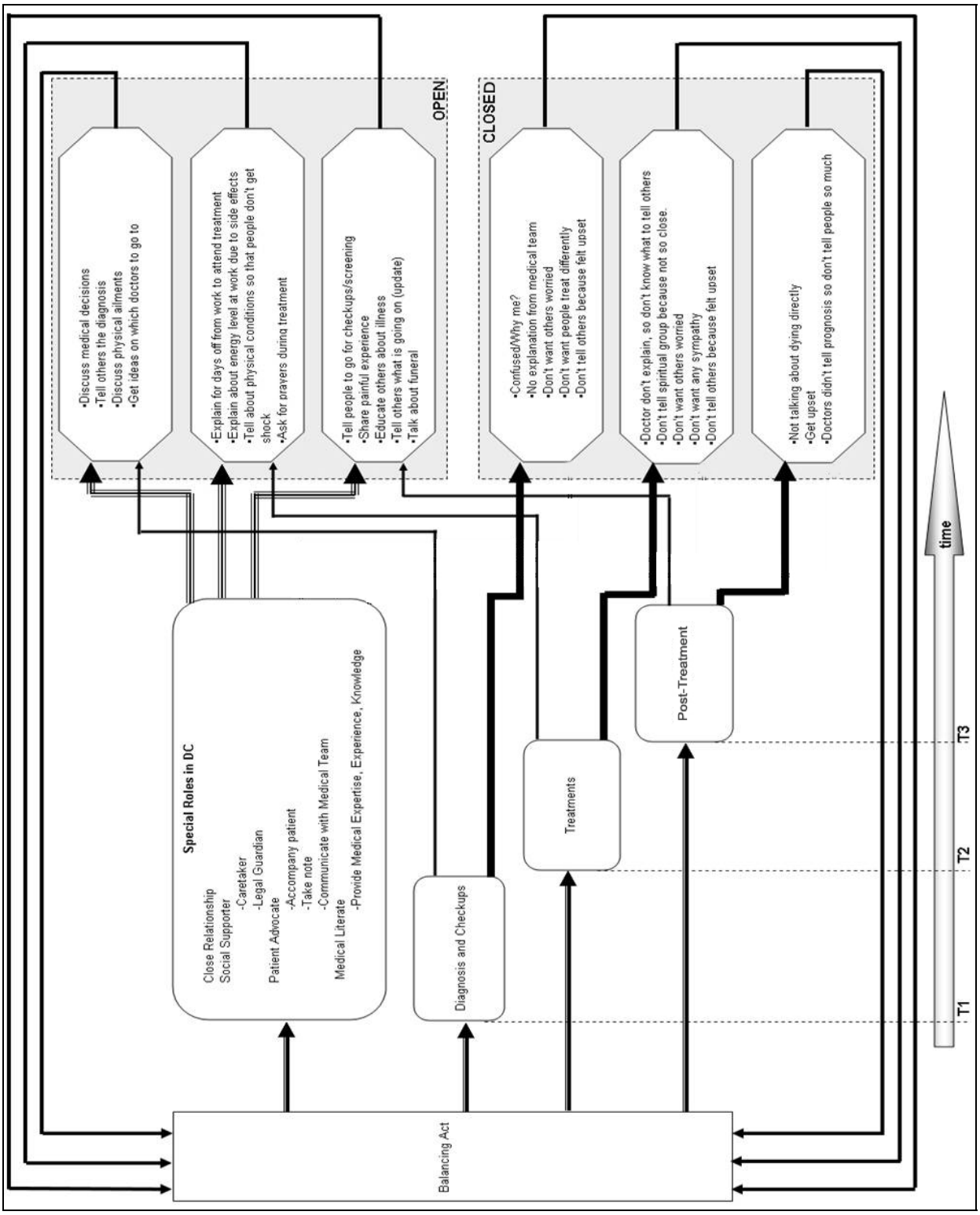



Figure 1. Conceptual Map for Social dynamics in disclosure communities

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# Peak Load Shifting Distribution of Multi-zone Fuzzy Group Decision-Making

Yang Li, Jia Liu, Liqun Gao, Zhi Kong
School of Information Science and Technology, Northeastern University
3-11 Wen Hua Road, Shenyang 110004, China
E-mail: liyang@ise.neu.edu.cn

**Abstract**

On the basis of different expert knowledge structure, from the greatest factors in power limiting distribution, using the method combining Analytical Hierarchy Process (AHP) and fuzzy set theory, a fuzzy comprehensive group decision model for multi-objects and multi-zones power limiting distribution in peak load shifting is built up. The problem of power limiting distribution between multi-zones is researched. The result manifests this model is reasonable and applicable.

**Keywords:** AHP, Fuzzy set, Group decision, Power limiting distribution

## 1. Current situation of foreign funds in China insurance industry

With the quick development of economy, the power shortage circumstance is appeared in most eastern and central regions in our country. To alleviate the complexion of power shortage, national and regional electric capacities carry out large-scale invest item for extending generate electricity capacity (including SanXia item). But the increasing speed of power consumption is more larger than the one of extension of power basic equipments, the amount of power shortage in each region is large, although some economy and policy means were carried out to control peak load shifting, for the consideration of power grid safety, so power limiting control for some regions in electro-apex is essential. It is an urgent solving and having actual meaning problem that how to scientifically and rationally distribute limited power limiting capacity in each region and make the security of power grid of each region highest and loss least, according to power limiting criterion and safety requirement of each region, through the comprehensive analysis of benefit on society, economy and environment.

Peak load shifting distribution is an important task and it needs knowledge and experiment of many persons. At the same time, it is a little fuzzy. So the method of combining fuzzy sets theory with hierarchy analysis is used to discuss the multi-objective group decision of multi-zone peak load shifting, according to the systemic analysis of society, economy, resource and environment. Therefore, power limiting distribution belongs to nonlinear multi-objective group decision category.

## 2. Construct Peak Load Shifting Distribution Decision System

### *2.1Power limiting distribution guide line system and hierarchy structure building*

The reasonable time distribution of peak load shifting in multi-zone is for the sake of power capacity standard and least loss based on the grid security. In this aspect, there are many factors, which are related to range and time of power limiting, also the conditions of social economy, such as population, industry configuration, parameter of national economy and so on. Some of them can be showed in economic quantity. Such as population in power limiting area, number of electricity enterprises in power limiting area and gross product in power limiting area, etc. Furthermore, it is hard to estimate some potential indirect benefit like effect on economy, adjustment of economy structure, damage for environment. The necessity of peak load shifting control is usually appraised by economic benefit. And these always affect the distribution of power limiting time. So it has some deficiency. Thus, the guide line system and hierarchy structure (Fig 1.) should be established according to the parameters picking up from the factors affecting power supply limiting like society, economy and so on. This will be good to the reasonable distribution of power limiting time and least loss. And the law of clear conception, abundant information, easy calculation and show justice is necessary.

### *2.2 Building and determining relative membership degree functions of each parameter*

Decision-making parameters always have four types in multi-objective decision problem. “a” is “excellent with large”, “b” is “excellent with small”, “c” is “fixed value” and “d” is “inter-zone”.

The type of "fixed value" parameter is a kind of parameter that the excellent value is a constant. Inter-zone means that the parameter value in some fixed region is excellent. The purpose of decision parameter classification is for compare among same parameters.

According to the difference between decision parameter types, parameter set F can be classified.

$$F = \bigcup_{i=1}^{4} F_i, \quad F_i \cap F_j = \Phi \ (i, j = 1, 2, 3, 4, i \neq j).$$

In this equation, $F_i$ ( $i = 1, 2, 3, 4$) is the kind of a, b, c and d respectively, and $\Phi$ is null.

Because of the difference between multi-objective dimensions, it needs to standardize the decision matrix M. And then, the eigenvalue of each decision parameter in relative state can be got. The expression of standardization is as follows.

$$a_{\max} = \max\{a_{ij} : 1 \le i \le n, 1 \le j \le r\} \tag{1}$$

$$a_{\min} = \min\{a_{ij} : 1 \le i \le n, 1 \le j \le r\} \tag{2}$$

($r_{ij}$ is relative membership degree of parameter $j$ in zone $i$, *max* and *min* is the symbol for taking maximum and minimum respectively, $\alpha_{ij}$ is the objective value of parameter $j$($r$ in all) in zone $i$($n$ in all) and have character of fuzzy by effect of value and statistical error.)

For type a,

$$r_{ij} = \frac{a_{ij} - a_{\min}}{a_{\max} - a_{\min}} \ (1 \le i \le n, 1 \le j \le r, f_k \in F_1) \tag{3}$$

For type b,

$$r_{ij} = \frac{a_{\max} - a_{ij}}{a_{\max} - a_{\min}} \ (1 \le i \le n, 1 \le j \le r, f_k \in F_2) \tag{4}$$

For type c,

$$r_{ij} = \begin{cases} 1.0, & a_{ij} = a^0 \\ 1.0 - \dfrac{|a_{ij} - a^0|}{\max\{a_{ij} - a^0\}}, & a_{ij} \neq a^0 \end{cases} \qquad (1 \le i \le n, 1 \le j \le r, f_k \in F_3) \tag{5}$$

$a^0$ is the optimal constant of parameter $f_k$.

For typed,

$$r_{ij} = \begin{cases} 1.0 - \dfrac{|b_1 - a_{ij}|}{\max\{b_1 - a_{ij}, a_{ij} - b_2\}}, & a_{ij} < b_1 \\ 1.0, & b_1 \le a_{ij} \le b_2 \\ 1.0 - \dfrac{|a_{ij} - b_2|}{\max\{b_1 - a_{ij}, a_{ij} - b_2\}}, & a_{ij} > b_2 \end{cases} \quad (1 \le i \le n, 1 \le j \le r, f_k \in F_4) \tag{6}$$

$[b_1, b_2]$ is the optimal region value of parameter $f_k$.

The decision eigenvalue matrix can be taken from the equations above.

From Figure 1, the experts get that the Power limiting region whose population are more, electro-enterprise are more, the total value of produce are more, economic loss of Power limiting are more, proportion of using electricity peak are more, the average temperature is more, the proportion of ice air condition are less, the distributed proportion of Power Limiting distribution for Peak Load Shifting are more.

Thus, these factors are all "b", except the scale of reconcilable peak enterprises and ice air condition are "a".

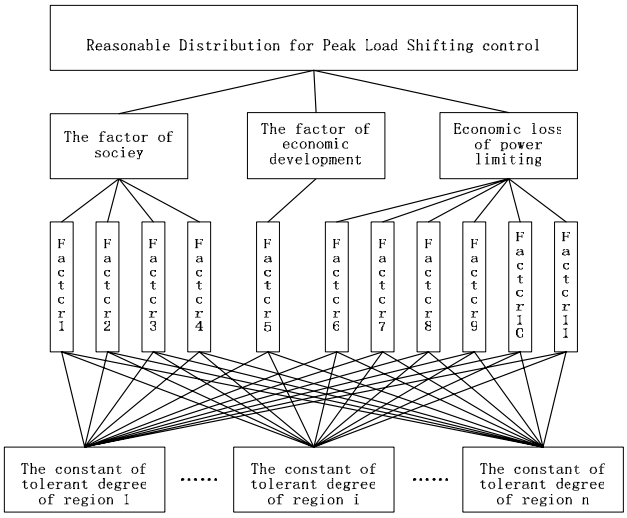



Figure 1. Decision system of Peak Load Shifting Power Limiting Distribution

factor1-The population of region

factor2- The quantity of electro-enterprise

factor3-The capacity of using power

factor4-The proportion of electro-enterprise that can be power-off

factor5-The total value of produce

factor6-The capacity of power limiting

factor7-The economic loss of power liming

factor8-The total population

factor9-The proportion of peak for using power

factor10-The average temperature

factor11-The proportion of ice air condition

## 3. Parameter Relative Weight Value Calculation

As a multi-objective decision method for combination of quality and quantity, hierarchy analysis has been applied extensively. But there are few examples about how to solve the group decision problem with hierarchy analysis method. In this paper, weighted geometry average group vector sort method based on different expert knowledge structure, has been used to deal with the hobby of some experts in multi-zone peak load shifting controlling time distribution and the problem of group hobby from knowledge structure.

### *3.1 The choice of expert group*

The time of peak load shifting is mainly affected by the factors in Figure 1. Thus, the composition of expert group is always peak load shifting control expert and electric power programming expert, etc. Also including electric utility manage expert, grid economy expert and grid security expert, who are accomplished in operation, theory and scientific research respectively. Obviously, the reasonable decision weight of each expert for each factor in Figure 1 is different.

### *3.2 The weight vector calculation of judging matrix for each expert*

In case of the number of expert is s and the number of reasonable peak load shifting control distribution factor or parameter is n,

$$P = \{P_1, P_2, \cdots P_n\} \tag{7}$$

The important degree should be judged by every expert though all effecting factors. If $V$ is the important degree judged matrix for factors building by expert *i*,

$$V = \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{s1} & \cdots & v_{sn} \end{pmatrix} \tag{8}$$

$V$ follows AHP method. And the weight vector of factors judged by expert $i$ can be taken according to the judged matrix.

Normalized the judged matrix $V$, $V'$ can be taken.

$$V' = \begin{pmatrix} v'_{11} & \cdots & v'_{1n} \\ \vdots & \ddots & \vdots \\ v'_{s1} & \cdots & v_{sn} \end{pmatrix} \tag{9}$$

$$v'_{ij} = \frac{v_{ij}}{\sum_{k=1}^{n} v_{ik}}, i = 1,2,\cdots\cdots,s \tag{10}$$

*3.3 The synthesis of expert weight vector*

During group decision, there are two methods of composing each decision maker's weight vector: one method is to compose on the base of judgment matrix; another method is to compose on the base of weight vector. Weight vector composition mainly adopts arithmetic weighted mean composition, and that model thinks over the relative important degree between composition elements, and permeates formula in the form of weight value, which rationalizes evaluation process. Because of experts' difference in knowledge, experience, and ability, their relative importance in decision may different. Supposing $s$ experts' relative importance is all same.

Then the method of arithmetic weighted mean composition is: according to the important degree's weight vector of $n$ influencing factors arrived at separately by s experts

$$V'_i = (v'_{i1}, v'_{i2}, \cdots\cdots, v'_{in})^T \quad i = 1,2,\cdots\cdots,s \tag{11}$$

Afterward, according to nether formula calculate all indicators' relative weight vector of weighted geometric mean group sorting vector $\overline{\omega} = (\omega_1, \omega_2, \omega_3, \cdots\cdots, \omega_n)^T$:

$$\lambda_1 = \lambda_2 = \cdots\cdots = \lambda_s = 1/s \tag{12}$$

$$\theta_j = \sum_{k=1}^{s} (v'_{kj} \times \lambda_k) \tag{13}$$

$$\omega_j = \theta_j \Big/ \sum_{i=1}^{n} \theta_i \ , j = 1,2,3,\cdots\cdots,n \tag{14}$$

$\lambda_1, \cdots\cdots \lambda_s$ correspond weight coefficient of decision maker ability level.

Finally, calculate standard deviation of $\omega_j$

$$\sigma_j = \sqrt{\frac{1}{s-1} \sum_{k=1}^{s} (\omega_{jk} - \omega_j)^2} \tag{15}$$

when $\sigma_j < \varepsilon \ (\varepsilon \in [0.5,1])$ ,we think group judgment acceptable, and feedback the indicator absolute weight vector of every decision maker to decision makers for their reference; if multiple decision makers all accept above weight vector, then calculation finishes. Otherwise, let decision maker offer amend judgment. Repeat many times in this manner, and up still all decision makers obtain satisfying weight vector.

**4. Optimized Distribution Decision Calculation of Multi-objective Power Limiting Time Distribution**

For the hierarchy model structure in Figure 1, the fuzzy synthesized decision model of multi-objective power limiting time distribution in multi-zone is established as follows.

$$Z = R \circ W = \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{pmatrix} \circ \begin{pmatrix} \omega_1 \\ \vdots \\ \omega_n \end{pmatrix} \tag{16}$$

In the equation above, $R$ is the relative membership degree matrix, $r_{ij}$ means the relative membership degree of the

parameter *j* in zone *i* of *m*, *W* is the parameter weight vector determined by group decision theory, $\omega_j$ is the weight of parameter *j* of *n* under group decision, *Z* is the scale vector of zone power limiting time distribution. "∘" is the fuzzy arithmetic operators. The scale of power limiting distribution of each region can be taken according to equation (16).

## 5. Research of Examples

According to the data from the charge department of Liaoning province, a peak load shifting distribution in six regions is planned to operate in summer of 2005. The social economy of six regions is shown in Table I. The loss is in Table II.

If there are three decision-maker who participate in the decision of peak load shifting controlling distribution, the result can be got using the model above. (The calculation process is omit)

Table 1. The Parameter of Social Economy

| regions | Population(10000) | Electricity consumption number | Electricity capacity(Mkw) | Reconcilable peak scale (%) | Product gross(1000rmb) |
|---|---|---|---|---|---|
| 1 | 56.42 | 308 | 47.28 | 17.83 | 476.52 |
| 2 | 23.18 | 106 | 21.53 | 52.12 | 180.63 |
| 3 | 89.13 | 416 | 62.52 | 41.63 | 725.20 |
| 4 | 10.67 | 89 | 18.62 | 71.63 | 97.65 |
| 5 | 36.21 | 218 | 32.15 | 31.25 | 207.26 |
| 6 | 43.56 | 229 | 29.56 | 52.06 | 168.88 |

Table 2. The Economic Loss of Power Limiting

| region | Power limiting capacity(Mkw) | Economic loss of power limiting(1000rmb) | Population(myriad human) | Electricity consumption peak scale(%) | Average temperature | Ice air condition scale(%) |
|---|---|---|---|---|---|---|
| 1 | 41.07 | 28.71 | 50.26 | 67.12 | 32.5 | 25.6 |
| 2 | 16.52 | 8.28 | 19.17 | 40.06 | 31.2 | 30.1 |
| 3 | 55.73 | 34.62 | 83.54 | 62.56 | 32.7 | 25.7 |
| 4 | 17.64 | 9.43 | 8.99 | 36.11 | 29.8 | 35.2 |
| 5 | 27.35 | 10.52 | 34.04 | 50.12 | 30.6 | 22.6 |
| 6 | 23.06 | 9.31 | 39.56 | 54.58 | 33.3 | 21.7 |

1) Indicator absolute weight vector of decision maker 1 is

$\bar{\omega}_1$ = {0.0483, 0.1690, 0.0549, 0.0554, 0.0978, 0.0052, 0.1949, 0.1093, 0.0918, 0.0602, 0.1132}

2) Indicator absolute weight vector of decision maker 2 is

$\bar{\omega}_2$ = {0.0623, 0.346, 0.0650, 0.0574, 0.1324, 0.0064, 0.1529, 0.0925, 0.0744, 0.1432, 0.0789}

3) Indicator absolute weight vector of decision maker 3 is

$\bar{\omega}_3$ = {0.0924, 0.2084, 0.0386, 0.0621, 0.1105, 0.0116, 0.0988, 0.0863, 0.1242, 0.0428, 0.1243}

4) Indicator weight vector of group decision is

$\bar{\omega}_{群} = \frac{1}{3}\sum_{i=1}^{3}\omega_i$ and normalized

= {0.0705, 0.1677, 0.0557, 0.0598, 0.1162, 0.0073, 0.1404, 0.0954, 0.0938, 0.0900, 0.1032}

5) Every region's Power Limiting Distribution for Peak Load Shifting control proportions are:

$Z_1$=7.83%, $Z_2$=19.75%, $Z_3$=4.87%, $Z_4$=43.52%, $Z_5$=12.98%, $Z_6$=11.05%

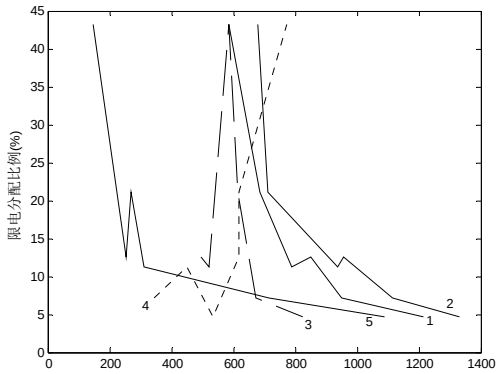



Figure 2. corresponding curve schematic diagram of social economic indicator and Power Limiting Distribution in every region

Population-1

The quantity of electro-enterprise-2

The capacity of using electricity-3

The proportion of moving-peak enterprise-4

The total value of produce-5

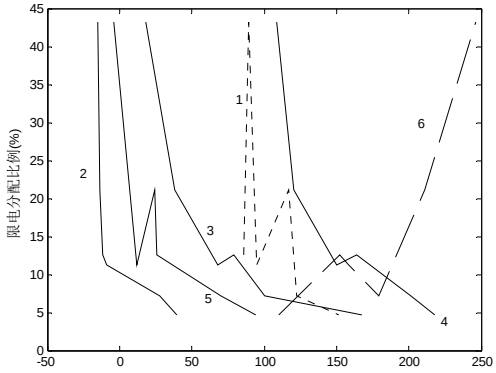



Figure 3. corresponding curve schematic diagram of the economic loss indicator of Power limiting and Power limiting Distribution proportion

Limited capacity-1

The economic loss of Power limiting-2

Population-3

Proportion of using electricity peak-4

Average temperature-5

The proportion of ice air condition-6

## 6. Conclusion

It can be seen from the corresponding curve schematic diagram of various decision indicators and Power Limiting Distribution proportion in above region (Figure 2; Figure 3):

With the change of social economic indicators and the economic loss of power limiting in various regions, the final Power Limiting Distribution proportion also changes: the Power limiting region whose population are more, electro-enterprise are more, the total value of produce are larger, economic loss of Power limiting is larger, proportion of using electricity peak is bigger, the average temperature is higher, the proportion of ice air condition is smaller, is distributed smaller proportion of Power Limiting distribution for Peak .

In a word, power limiting distribution is a complex multi-object group decision problem, the decision of power limiting

time distribution is not only related to technology, economy and power limiting criterion but also to the supplying safety of power grid and the stability of society, so the research of decision of power limiting time distribution is more complex. Analyzing from main factors effecting to region power limiting time distribution, criterion system of region power limiting distribution is built up in this paper, and the model and method of large-scale systematical multi-object group decision combining hierarchy analysis and fuzzy comprehensive evaluation is proposed also, and satisfied research results of actual example is obtained. But in the future it should be more strengthened on the society evaluation of power limiting distribution of peak load shifting control, economical evaluation, environmental evaluation, decision method of power limiting uncertain distribution and its reliability research.

**References**

Chiclana F, Herrera F & Herrera Viedma E (1998). Integrating three representation models in fuzzy multipurpose decision making based on fuzzy preference relations. *Fuzzy Sets and Systems*, 9, 33-48

Delgado M, Herrera F Herrera Viedma E (1998).Combining numerical and linguistic information in group decision making. *Information Sciences*, 10, 177-194

H.C.Huang, R.C.Wang & J.G. Hsieh (2002). A new artificial intelligent peak power load forecaster based on non-fixed neural networks. *Electr.Power Energy Syst*, 24, 245-250

Herrera F & Herrera Viedma E (2000). Linguistic decision analysis: steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems*, 11, 67-82

Herrera F & Herrera Viedma E (2000). Choice functions and mechanisms for Linguistic preference relation. *European Journal of Operational Research*, 20, 223-239

J. Ma, Z.P. Fan & L.H. Huang (1999). A subjective and objective integrated approach to determine attributes weights. *European Journal of Operational Research*, 12, 397-404

Marmol AM, Puerto J & Fernanndez FR (1998). The use of partial information on weights in multi-attribute decision problem. *Journal of Multi-attribute Decision Analysis*, 7, 322-329

M. Krunic, K.I. Slavisa & N. Rajakovic (2000). An improved neural network application for short term load forecasting in power systems. *Electr. Mach. Power Syst*, 28, 703-721

M.R.G. Al-Shakrchi & M.M. Ghulaim (2000). Short term load forecasting for Baghdad electricity region. *Electr. Mach. Power Syst*, 28, 355–371

Noel Bryson & Ayodele Mobolurin (1999). An action learning evaluation procedure for multiple criteria decision making problems. *European Journal of Operational Research*, 6, 379-386,

Shamsuddin Ahmed (2005). Seasonal models of peak electric load demand. *Technological Forecasting and Social Change*, 72, 609-622

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# Mapping of Power Transmission Lines on Malaysian Highways Using UPM-APSB's AISA Airborne Hyperspectral Imaging System

Hj. Kamaruzaman Jusoff
Forest Geospatial Information & Survey Lab, Lebuh Silikon
Faculty of Forestry, Universiti Putra Malaysia
Serdang 43400 Selangor, Malaysia
Tel: +60-3-8946 7176 E-mail: kamaruz@aeroscan.biz
Norsuzila Ya'acob
Faculty of Electrical Engineering, Universiti Teknologi MARA
UiTM 40450 Shah Alam, Malaysia
Tel: +60-3-5544 2000 E-mail: norsuzilayaacob@yahoo.com

**Abstract**

Power transmission lines routes mapping is an important technique for locating power transmission line routes and towers on mountain/hilltops to assist viewing of their impacts on the environment, operations and allocation of public utilities. A study was therefore conducted to map the power transmission lines within Bukit Lanjan PLUS highway. The main objective of this study is to assess the capability of airborne hyperspectral sensing for mapping of power transmission. By using ENVI software, the airborne hyperspectral imaging data was enhanced using convolution filtering technique using band 3 which produced a gray scale image which appeared clearer and sharper. The spectral reflectance curves were acquired for each power line which showed the same spectrum characteristics in curve or the reflectance energy. This is because of the same power lines composition material for all power lines. Ground verification was done by comparing the UPM-APSB's AISA Global Positioning System (GPS) coordinates readings with ground GPS coordinates readings of the power transmission lines footings. The ground verification result from the two matching power transmission line footings showed that the accuracy of power lines identification was acceptable. This study implies that airborne hyperspectral imagers are powerful tools for mapping and spotting of suitable large transmission towers and lines.

**Keywords:** Power transmission lines, Airborne spectrometry, Spectral signature, Routing

## 1. Introduction

Power transmission lines are electrical lines that typically carry high voltage and have to traverse the length and breadth of the country, for evacuation of power from generating stations to load centers and beneficiary states, the topographical and geographical nature of the terrains play significant influence in the project cost and implementation time. Hence, it is important to determine power transmission lines routes spotting. Unmonitored power transmission lines such as when tree grow too close to power lines, it is potential threat to electrical system reliability and safety, resulting in unnecessary power outages and interruptions in electrical service to customers or may cause forest fires when the bamboo tops hit the wire lines. The potential environmental impacts from the construction and operation of transmission lines can be minimized once precise location of power transmission lines tower footings can be determined. Many of the issues relating to changes in land value resettlement and loss of productivity can be dealt with easily. For instance, sites with cultural or historical importance that might fall in the transmission line route or even the encroachment into precious ecological forested areas and valuable lands may be totally avoided (Kamaruzaman, 2004a).

Power transmission lines corridor are traverse along the linear length through physiographic features, land use pattern, types of habitation in and around the vicinity by supported of the transmission tower which is cover large distances. Due to the force of gravity, power lines tend to sag. This initial sag increases with line temperature because the conducting material, of which the line is made, expands, effectively lengthening the line. A small increase in line length produces a large, and potentially hazardous, increase in sag. Sagging power lines will contact vegetation and short

circuit. This causes power interruptions and forest fires. On the other hand, the suitable sites for new transmission lines have been getting restricted, because of development of a rural areas and the growing concern over environmental issues. Power transmission lines should avoid the main settlements but there is a possibility that they pass over some houses scattered over the hill, grazing and pastureland and terrace farm this is cause by not well spotted during the planning stage. In lieu of the above problems, an urgent need for a remote sensing data is important to aid the spotting of power transmission footings on relatively high spots such as over the hills maintaining similar height. In Malaysia, the only alternative remote sensing data source is the UPM-APSB's AISA airborne hyperspectral imaging which carries an optical sensor and proven to be useful for different applications (Kamaruzaman, 2004b, 2004c; and Kamaruzaman, 2008).

The general objective of this study is to assess the capability of airborne hyperspectral sensing for TNB power transmission lines mapping and spotting in Bukit Lanjan, Selangor. Meanwhile, the specific objective is to precisely locate and map the power transmission line route over the thick dense forested hills mountains and highway.

## 2. Methods and materials

### *2.1 Description of study area*

The study was conducted in Bukit Lanjan (Figure 1), situated in the state of Selangor and approximately 47 km from Kuala Lumpur's capital city. It lies roughly between latitudes 3º10'40'' and 3º11'02'' North and longitudes 101º35'50'' and 101º36'26'' East. Bukit Lanjan is almost in the midst of affluence Petaling Jaya, is the highest lone peak in the Klang Valley. There is a small patch of pristine tropical forest at the peak. The road in Bukit Lanjan is currently under the maintenance of The North South Expressway (PLUS).

### *2.2 Equipment and software*

#### 2.2.1 Description of UPM-APSB's AISA Sensor

UPM-APSB's AISA is a solid-state, commercially produced inexpensive hyper spectral push-broom imaging instrument to recording remote sensing images over a large spectrum wavelengths from the visible (VIS) to near infrared (NIR) which is 400-1000 nm. Images have a ground pixel size of 1 m x 1 m with a flight altitude of 1,000 m a.s.l and a constant flight speed of 120 knots (Kamaruzaman, 2004a). It is designed to provide a near real-time, frequent, repetitive, accurate and reliable push-broom instrument that acquire images in 288 registered, contiguous narrow spectral band passes such that for each element it is possible to derive a complete reflectance spectrum. The UPM-APSB's AISA hyperspectral imager is a complete system that consists of a compact hyperspectral imager head, miniature Global Positioning System (GPS)/Integrated Navigation System (INS) sensor for precise positioning, data acquisition unit and CaliGeo pre-processing software. This small portable instrument, with a total weight of only 15 kg was mounted on an aluminium metal plate that is compatible with a standard aerial camera mount, available in any fixed wing aircrafts such as that of a Pan Malaysia Air Transport (PMAT) Short SkyVan SC7, a Sabah Air GAF Nomad N22B or a RMAF C402B. Swath width is 360 pixels and field of view (FOV) in cross track direction 20° which makes ground resolution from 1 km altitude approximately 1 m at a flight speed 120 knots (60 m/s).

The versatile graphical user interface (GUI) provides flexible, easy-to-use instrument with several efficient operating modes and features that may be changed during flight within seconds. The data is stored as a default to a large capacity hard disk, which is providing higher frame rates compared to traditional tape storage based systems. The refractive properties of the two opposing prisms allow for a linear projection of light onto the charged coupled device (CCD) array. The two-dimensional array consists of a spatial axis of 364 detectors, and a spectral axis of 286 detectors. The UPM-APSB's AISA sensor system with 20 pixels per swath for downwelling irradiance system was acquired via a fiber optic irradiance sensor (FODIS) on the N22B aircraft. The FODIS allows for the concurrent measurement of downwelling and upwelling radiance by the UPM-APSB's AISA sensor head. The calibration of the FODIS coupled with the UPM-APSB's AISA sensor allows for the calculation of apparent at-platform reflectance. Normally, downwelling irradiance system is needed so that the upwelling and downwelling measurements can be compared directly (Figure 1).

UPM-APSB's AISA is capable of collecting data within a spectral range of 430 to 900 nm. Although UPM-APSB's AISA is capable of collecting up to 286 spectral channels within this range, the data rate associated with the short integration times (sampling rates) required of the sensor in most operational/flight modes, limits the number of channels. Spectral resolution is important for detecting fine spectral features that can identify specific materials. The full spectral mode, however, is useful for acquiring 286 band spectral signatures of specific targets that can be used to generate pure end members as well as for band selection purposes. Current operational collection configurations range from 10 to 70 spectral bands depending on the aircraft speed, altitude and mission goals**.** Table 1 shows the spectral and spatial resolutions achievable when holding ground speed constant, in this case 120 Knots (*Table* 1).

2.2.2 Softwares: The Caligeo and ENVI Version 4.0

CaliGeo is a software package designed to process raw UPM-APSB's AISA sensor data quickly and accurately to a format that can be read using a data analysis package and analyzed using the latest data processing and analysis methods into the final data products. The process follows few simple steps that included radiometric correction, geometric correction, rectification and geo-referencing. After these steps were completed, the data were visualized on a user-friendly interface that runs interactively. This means that the data can be viewed without a need to transfer data and large files from one program to another reducing the amount of inconvenience for the user. Most of the processing is relatively straightforward and automatic, and the graphical user interface (GUI) makes it easy for any one to learn and to use. After CaliGeo processing the data was then analyzed using ENVI.

ENVI software version 4.0 which is a unique approach and user friendly to search for images spectral signatures, in addition to the geospatial capabilities, which result in an effective identification, visualization, spatial and spectral data reduction and management tool with integrated decision-making capabilities. ENVI is used to find hidden targets (including sub-pixel targets), identify terrain features, visualize 3D terrain, and perform line-of-sight analysis. ENVI is also the undisputed leader in hyperspectral image analysis, providing the only environment capable of fully utilizing the feature identification power of hyperspectral data. Easily ingest panchromatic, multispectral, hyperspectral, radar, elevation images, or vector GIS data.

*2.3 Airborne* data acquisition and data pre-processing

The calibration flight by UPM's FGISL/Aeroscan Precision (M) Sdn.Bhd. took place on 19 February 2004 in Bukit Lanjan using a SC-7 aircraft with flight altitude of 1,000 m a.s.l and 1m x 1m ground spatial resolution. Mission profiles were planned using high-resolution digital maps of the operating area. The flight path lines (FPL) were identified using a specialized GPS software and 20 operational bands ranging between 438.8 nm to 894.1 nm wavelengths including visible light (red, green, blue) and near infrared were selected and configured. Data delivery over some the selected Area of Interest (AOI's) i.e Bukit Lanjan was accomplished within 24 hours of completion of data acquisition.

The UPM-APSB's AISA imaging spectrometer was configured to measure 20 spectral bands. The data was first pre-processed using a CaliGeo software (a plug in of the latest ENVI version 4.0) for a calibrated and provides for the automatic geometric and radiometric correction, rectification, mosaicking, and calculation of radiance or apparent at-platform reflectance (FODIS ratio). The program uses the GPS and attitude information from the INS to perform the geometric, geo-referencing and mosaicking operations. Automated batch processing provides for rapid turnaround times for data delivery.

Geometric correction is an error on its image between the actual image coordinates and the ideal image coordinates. The distortions are in form of internal distortion resulting from the geometry of the sensor and external distortion due to the altitude of the sensor or the shape of the object. Meanwhile, radiometric correction is to correct for the varying factors such as scene illumination, atmospheric conditions, viewing geometry and instrument response. Radiometric correction also detects and measures the radiant energy, either as separate wavelengths or integrated over a broad wavelength band, and the interaction of radiation with matter in such ways as absorption, reflectance and emission. All the three types of radiometric correction i.e., radiometric correction due to sensor sensitivity, sun angle and topography, atmospheric correction due to absorption and scattering were corrected on-board the aircraft during image pre-processing using the Caligeo software.

Rectification is the process by which the geometry of an image is made planimetric using GCP's to transform the geometry of an image, so that each pixel corresponds to a position in a real world coordinate system. However, UPM-APSB's AISA sensor is unique in the sense that no GCP's were required with the available GPS/INS unit on-board the aircraft to automatically calculate the position of the aircraft.

*2.4 Ground verification*

Ground data verification was conducted to determine the mapping accuracy and exact location of the power transmission lines on the ground using a handheld differential GPS.

*2.5 Image processing and final output*

The UPM-APSB's AISA data were processed digitally using a user friendly ENVI version 4.0 to develop the image spectral signatures in addition to the geospatial capabilities, which result in an effective identification, visualization, spatial and spectral data reduction and management tool with integrated decision-making capabilities. The data were subjected to the minimum noise fraction transformation, pixel purity index, n-dimension visualizer, identification; spectral angle mapper and mixture tuned matched filtering processing. The final output is the AeroMAP$^{TM}$ product that shows the location and routing of power transmission lines tower footings which is useful for developing a systematic management of the distribution and position of power transmission lines.

## 3. Results and discussion

### *3.1 Band Combination without Image Enhancement*

Generally, there are 20 different bands in UPM-APSB's AISA data that can be combined to produce false color composite images. Figure 2 shows the pre-processed image of power transmission lines using band combination 19-11-2 (RGB) without enhancement. From the raw image, there is no indication that the power transmission lines are detected.

There are three sets of false color composites selected consisting of band combinations 3-3-3, 8-8-8 and 9-9-9. Contrast Stretching and Quick Filter with band combinations was applied to adjust the color or gray scale range of a selected image. By applying this technique, the computer display's dynamic range can be filled and the selected contrast of the image can be improved as shown in Figure 3. There are several default stretching options consisting of Linear 0–255, Linear 2%, Gaussian, Equalization, and Square Root. The image showing the power transmission lines routings and footings in the gray scale appear sharper and clearer compared to the other band combinations.

### *3.2 Spectral library profiles*

By using the ENVI software, the image spectral reflectance curves for each power transmission lines were acquired (Figure 4). From the spectral library profiles, the spectral reflectance of each power lines shows a relatively high spectral signature profile which is between the frequency ranges of 450-900 nm. This is due to the aluminium material used by the power transmission lines which has a shiny surface that reflects every visible and near infrared wavelength range (Baskar, 2000). (Figure 2, Figure 3, Figure 4).

### *3.3 Ground truth/verification*

The ground verification was done by comparing the UPM-APSB's AISA Differential GPS coordinates readings with the handheld ground DGPS coordinates of power transmission lines footings. From the airborne hyperspectral data, the locations of power transmission lines footings lies between latitudes 3º 10' 57'' and 3º 10' 58'' North and longitudes 101º 36' 4'' and 101º 36'5'' East. Meanwhile, the ground GPS readings for No. 1 footing is latitude 3º 10' 57.4'' North and longitude 101º 36' 04.9'' East, No. 2 footing is latitude 3º 10' 57.3'' North and longitude 101º 36' 04.3'' East (Figure 5). The ground verification result from both power transmission lines footings show that the accuracy is acceptable and was used to support the capability of airborne hyperspectral data in mapping power transmission lines routing (Figure 5)

## 4. Conclusions and recommendation

From the study carried out on mapping of power transmission lines routing and spotting using UPM-APSB's AISA sensor in Bukit Lanjan PLUS highway, the following conclusions can be derived from this study, namely (i) airborne hyperspectral imaging can locate and map the power transmission lines, and (ii) Image enhancement filtering using convolution technique with band 3 produced gray scale image was found to be the best technique for power transmission lines mapping. UPM-APSB's AISA hyperspectral imagery should be used for power transmission lines mapping over thick dense forested hills and highway.

## References

Bhaskar, S.D. (2000). *Sub-pixel analysis of urban surface material: A case study of Perth, Western Australia.* (Student Prize Competition) 24rth-29th July, 2000 IEEE/IGARSS, Honolulu, USA.

Kamaruzaman Jusoff, (2004a). AISA Airborne Hyperspectral Imaging for Highway Transportation Planning and Management. Poster paper presented at *International Seminar Geospatial* 2004, Istana Hotel, Kuala Lumpur. 8p.

Kamaruzaman Jusoff, (2004b). The Potential of Utilizing Hyperspectral Remotely Sensed Data Integrated Coastal Management with Special Emphasis on Airborne Imaging Spectrometry. Invited paper presented *the KUSTEM 3rd Annual on Science, Sustainability & Management*, 4-5 May, 2004, Primula Beach Resort, Kuala Terengganu. 6p.

Kamaruzaman Jusoff, (2004c). AISA Airborne Hyperspectral Imaging for Precision Agriculture and Forestry. An oral paper presented at *4th KUSTEM Annual Seminar Sustainability Science and Management Meeting Challenges in Sustainability Agrotechnology*, 2-3 May, 2004, Primula Beach Resort, Kuala Terengganu. 6p.

Kamaruzaman Jusoff, (2008). Search and Rescue (SAR) Operations for the Missing Bell 206 Long Ranger. Helicopter in Sarawak, Malaysia Using UPM-APSB's AISA Airborne Hyperspectral Imaging Data. *An Int. J. Disaster Prevention & Management* (In Press).

Table 1. Spectral and spatial resolutions with ground speed of 120 knots

| Altitude | Spatial Resolution | Numbers of spectral bands |
|---|---|---|
| 1,000 m (3,280 ft) | 1.0 m | 20 |
| 1,500 m (4,920 ft) | 1.5 m | 26 |
| 2,000 m (6,560 ft) | 2.0 m | 34 |
| 2,500 m (8,200 ft) | 2.5 m | 55 |
| 3,000 m (9,840 ft) | 3.0 m | 58 |
| 4,000 m (13,120 ft) | 4.0 m | 70 |

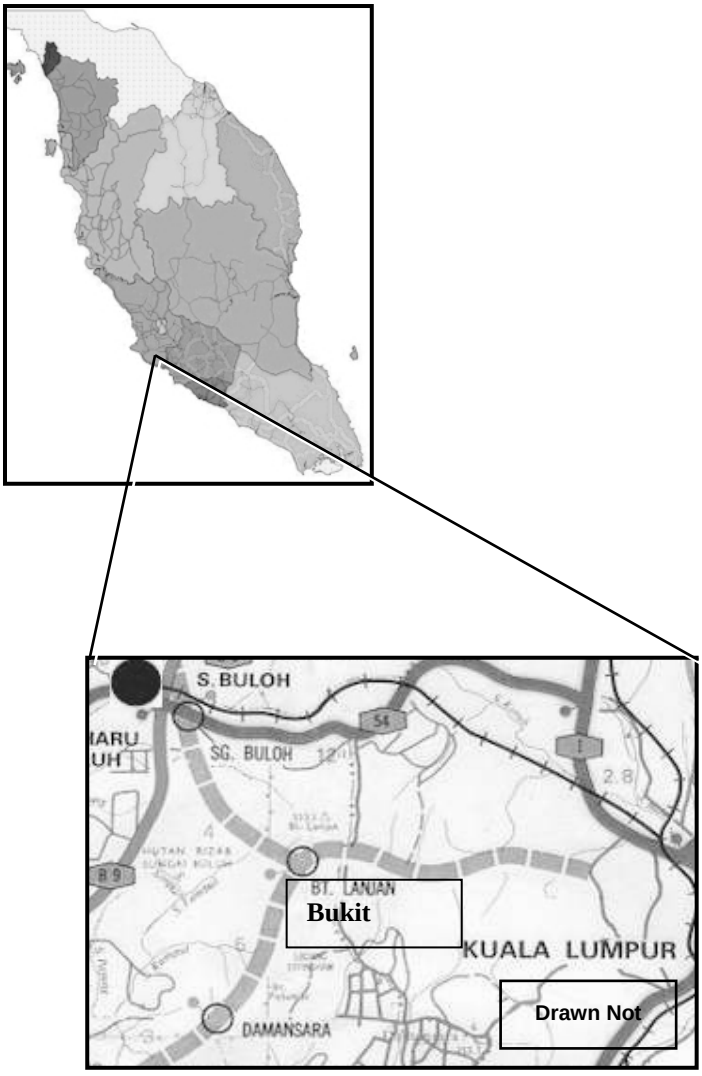



Figure 1. A map of Peninsular Malaysia showing the location of the study site

Figure 2. Pre-processed image band combination 19-11-2 (RGB) without enhancement

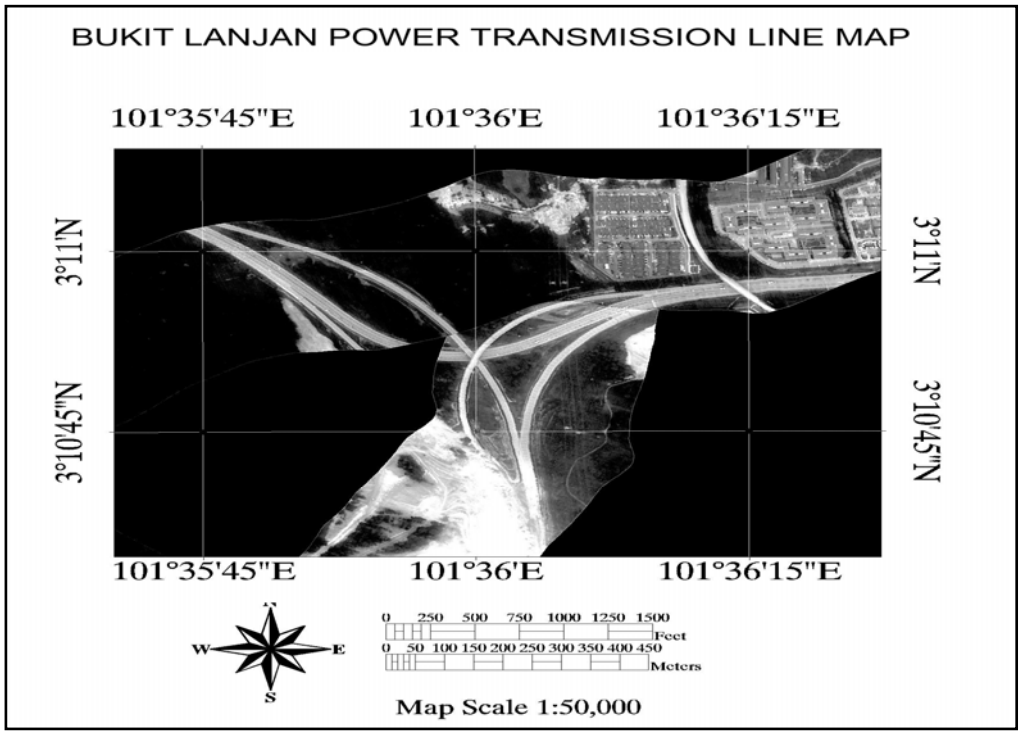



Figure 3. Image enhancement and filtering using convolution technique

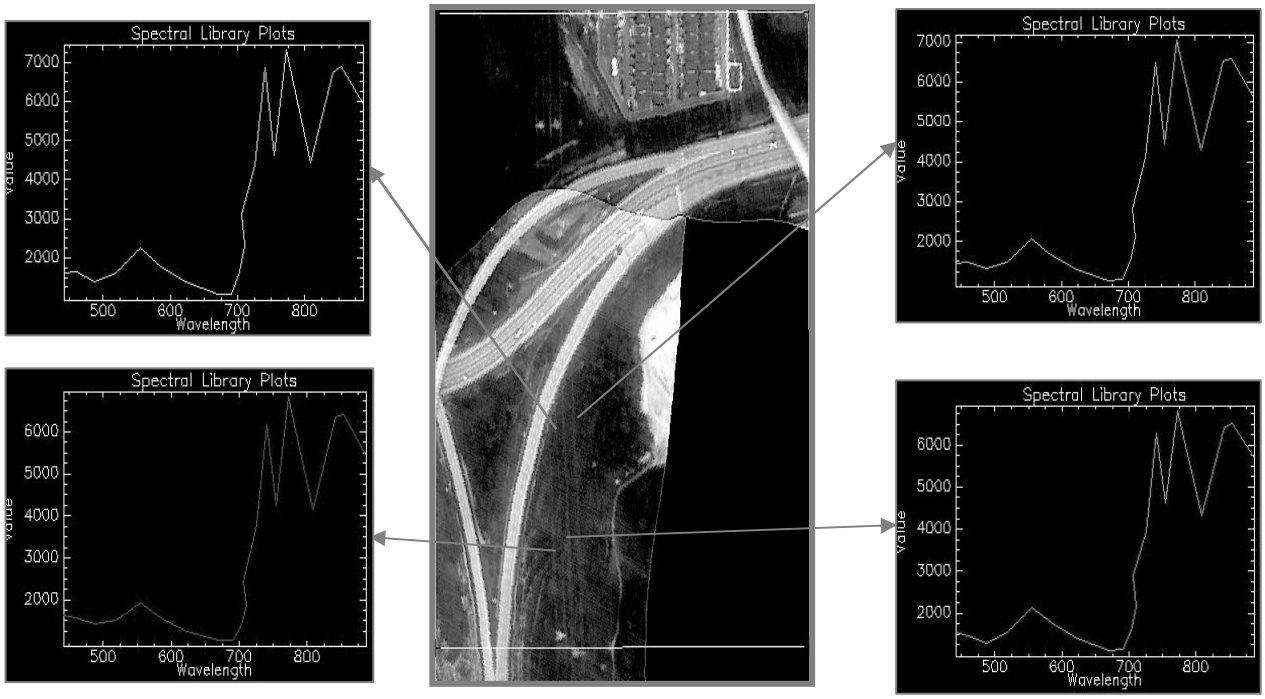



Figure 4. Image Spectral Signatures Developed and Defined
for Power Transmission lines Features in Bukit Lanjan PLUS Highway

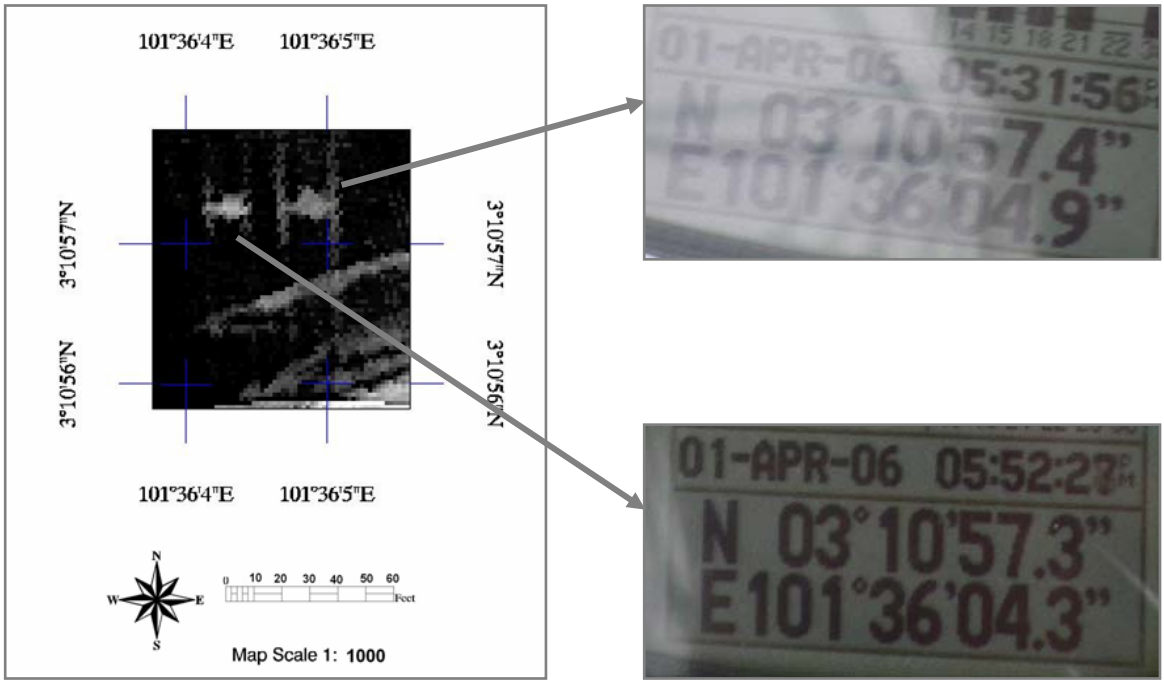



Figure 5. Field DGPS Readings of Power Transmission Lines Footings

Computer and Information Science
www.ccsenet.org/journal.html
Vol. 1, No. 1
February 2008

# Modelling the Dynamics of Internet Adoption

Mohd Nazri Mahmud
School of Electrical and Electronic Engineering, Universiti Sains Malaysia
Engineering Campus, Seri Ampangan, 14300, Nibong Tebal, Penang, Malaysia
Tel: 6-04-5996059 E-mail: nazriee@eng.usm.my

**Abstract**

The adoption of the Internet in societies is becoming more important as more and more public and business services are delivered via the Internet. Understanding the dynamics of the adoption is vital in developing policies to stimulate greater adoption. This paper highlights the dynamics of internet adoption in a general society based on the segmentation of the society into subcultures according to their respective affinities towards Internet adoption. System dynamics model are built to reveal the dynamics.

**Keywords:** Internet Adoption, System dynamics

## 1. Introduction

The internet is prevalent as a medium of service delivery for public and businesses. The success of e-government initiatives in public service delivery for example depends on the adoption of the Internet among the citizen. Policy makers are finding ways to increase the adoption especially in developing countries where adoption mostly concentrates in the urban areas. It is important to recognize the difference among the society in terms of their affinity towards Internet adoption. This study takes a segmented view of the society to assist in developing a system dynamic model which is used in investigating the dynamics of the adoption. The system dynamics modeling approach is employed here in order to give a holistic view of the dynamics of the Internet adoption business.

## 2. Segmentation of the society

### *2.1 Categorisation of potential adopters*

The categorisation of adopters encompasses demographic, socioeconomic and regional characteristics that have been commonly included in Internet demand studies (Deaton, 1992), (McCraken, 1988). Such a market-based segmentation implies an underlying assumption founded on standard economic theory of utility maximisation that concerns with individual's demand for goods (Liebowitz and Margolis, 1994).

Generally, cost-conscious consumers make their Internet subscription decision by comparing costs and benefits from such a connection based on utility maximisation under income constraint. This means it is a choice between spending their disposable income for optional needs to acquire Internet connection or spend on other optional needs such as entertainment (InfoSoc Malaysia, 2000). Subscription is made if the utility of an Internet connection exceeds the utility from other alternatives such as a post-mail or a fax. Utility is also a function of many factors such as content appropriateness and security. This decision behaviour of the potential Internet users is an important dynamic to be captured in the model.

The standard utility perspective also recognises the elasticity of demand. Price elasticity depends on the existence of substitutes, the importance of the product or service in the consumer’s budget and the time period under consideration. (Villasis, 1996, p.71). It also relates to other types of elasticities namely income, cross-price and supply elasticities. Cross-price elasticity highlights the important dynamics between prices of complements and substitutes (e.g. computers, access lines, post-mail, faxes) and subscription to Internet.

This utility-based view of the users means that subscription to Internet is not a one-way process. There is a possibility of attrition especially when such subscriptions are based on dual-pricing system, whereby charges comprise of two components: access and usage. For example, a moderate Internet user may subscribe on the basis of low usage fees but discontinue later on because their low usage levels do not justify the high access fees (Lemon and Winer, 1995).

Apart from the consumers, firms are also important actors in the Internet market. These are the Internet service providers (ISPs), Application Service Providers (ASPs) and Content Service Providers (CSPs). The first one provides connectivity while the other two provides applications such as financial services and content such as digital movies.

Drawing upon the theory of the firm, these providers follow the standard profit-maximizing and price-taking behaviour

by optimising their supply (Bental and Spiegel, 1995). This price taking behaviour is visible in developed Internet market such as the US where – unlike the less developed market such as Africa- the providers exceeds several hundreds (ITU, 2001). Thus, the Internet market increasingly becoming close to perfectly competitive condition. This theory and evidence inform us the competitive dynamics of the Internet market. With increasing market potential, the model shall exhibit the proliferation of ISPs, ASPs and CSPs.

*2.2 Causal loop diagrams*

The society is divided into eight segments based on their characteristics with respect to Internet adoption. This is shown in Table 1 below. Modeling the factors in the subsystems involves the formation of Causal Loop Diagrams to represent the dynamics and causality between variables. The Causal Loop Diagram for the Imitators subsystem is given in Figure 1.

The existing subscribers influence the potential "imitators" through "word-of-mouth". As the population of subscribers increases, the "word-of-mouth" effect increases and causes the subscription rate to increase even further. The degree of influence also increases with subscriber base as "network externalities" increases the perceived value of the Internet when more of their close friends and families join the network. This represents a positive reinforcing feedback loop, labeled as "word-of-mouth" in the diagram.

On the other hand, if the population of imitators is finite and/or when the rate of increase in the number of potential imitators is lower than the subscription rate, an increase in the latter would reduces the population of the former. This in turn reduces the subscription rate and thus provides the balancing feedback loop to the subsystem.

Realistically, not all of these encounters involve existing subscribers. To take this into account, the numbers of encounters with subscribers is determined by the probability of encountering an existing subscriber. Similarly, the probability of encountering a subscriber is also determined by the total number of existing subscribers and the country's total population. Not all of these encounters results in Internet subscription. This is determined by the Influential Encounter fraction; a probability that an encounter will result in subscription.

The second subsystem models the Internet cost conscious segment and focuses on modeling the effects of economies of scale, competition, Internet cost and income. The Causal Loop Diagram is given in Figure 2. As the subscriber population increases, the marginal cost to provide Internet service decreases due to economies of scale. This reduces the Internet cost to consumers and increases the number of subscriptions, which increases the total subscriber population.

Similarly, as the subscriber population increases, the Internet sales revenue also increases. Such a signal of market potential attracts more Service Provider to enter the industry and increases the competition that drives the price down. Consequently, a reduction in price encourages more cost-conscious consumers to subscribe. The dotted lines in the diagram represent delayed effects as it takes time for a new entrant to enter and affect the prices.

The Internet cost is affected by both the economies of scale and competitive forces. The value is the minimum of the two as some competitors may give substantial price reduction even if they have not acquired the scale economies. When there is no competition, the only provider is the state-owned incumbent that by default has a substantial scale advantage to reduce the price. In the absence of a monopoly, the resulting competitive forces drive the price down.

However, the number of service provider cannot rise infinitely. As the number of competitor increases, sales revenue for individual provider decreases. This reduces the number of providers that can profitable compete in the market. Thus, we can expect to see the maximum number of providers as the market matures.

Similar to the previous subsystem, if the population of Internet Cost Conscious is finite and/or when its rate of increase is lower than the subscription rate, the number of subscription by this segment will approach saturation. This in turn reduces the subscription rate and thus provides the balancing feedback loop to the system.

Similar to the previous subsystem, if the population of Internet Cost Conscious is finite and/or when its rate of increase is lower than the subscription rate, the number of subscription by this segment will approach saturation. This in turn reduces the subscription rate and thus provides the balancing feedback loop to the system.

The third subsystem involves dynamics that resembles that of Internet Cost Conscious subsystem. The disposable income for this group needs to be lower than ICC depending on the context. The main Causal Loop Diagram is given in Figure 3 below.

The Causal Loop Diagram for the fourth subsystem is given in Figure 4 below. Again, the subsystem mimics the previous two subsystems albeit a much lower disposable income for this group. The fifth subsystem which represents the dynamics in the underserved areas focuses mainly on the capacity acquisition in the presence of a waiting list. The subsystem considers only those who can afford to subscribe. Poor people from this area are put under another group that needs collective access and special training program. The Causal Loop Diagram is given in Figure 5.

Due to infrastructure unavailability, the orders are put on a waiting list and labeled as "waiters". The greater the

population in this area, the higher the number of households without access lines. New order increases depending upon the fraction of the household placing orders. These new orders become new waiters who add to the existing waiting list. The longer the waiting list and the rate of increase in new order, the higher the order forecast would be. When the financial resources are abundant, the order forecast dominates the decision on the number of capacity to be added. In many cases, countries face capital constraints to providing access in remote areas [28]. The more capacity added, the more additional capacity available for serving waiters. However, higher number of waiters served results in lower capacity left to fulfill new orders.

The final subsystem highlights the dynamics that occur in the educational pipeline that can have long term influence on the Internet penetration level. The higher the educational achievement translates to higher potential of becoming Internet users. An Early exposure to Internet in the educational sector also increases the likelihood of future usage. The causal loop diagram given in Figure 6 also captures the dynamics that lead to the formation of the group who are considered as marginalised in ICT. The formation of the marginalised group occurs when children either do not attend any formal education and become illiterate or drop out of school early without undergoing any special program to acquire any IT skill. Thus, they are unable to use the services provided via the Internet. The higher the population of marginalsied society, the less capable a country is with respect to the knowledge economy and this causes a cycle of poverty.

## 3. Case application

This section shows how the model can be tailored to a specific country's application and build confidence in its applicability to policy experimentation. The platform model has been applied to the Malaysian context.

The Internet started in Malaysia in 1988 when the Malaysian Institute of Microelectronic Systems (MIMOS) set up a university computer network called Rangkaian Komputer Malaysia ("Malaysian Computer Network), RangKom, which consisted of 4 dial-up lines to Australia, Holland, Korea, and the US. It offered e-mail and news group services. This dialup infrastructure was replaced by a satellite link to the US, and Malaysia MIMOS established the Joint Advanced Integrated Networking (JARING) as an Internet Service Provider (ISP) and remained the sole provider until Telekom Malaysia's TMnet entered in 1996. The market remained a duopoly until 2000 when additional licenses were issued (ITU, 2001).

There is a great concern within the Malaysian government about the "Digital Divide", particularly in differences in access between urban and rural area. The limiting factor is the low level of personal computers penetration. The cost of Internet access including a PC purchase is 88% of the rural household disposable income for optional needs. The biggest bottleneck however is the fixed telephone lines penetration. 37% of households do not have telephone and in 2000 the waiting list was 98,000 (ITU, 2001). It is too costly to serve those still without home telephones. Based on the information and detailed data, the model was simulated and the following penetration characteristics in Figure 7 was observed between 1990 to 2002.

The model follows closely the level of penetration up to 1997 when a substantial underestimation occurs. The reason is that the model does not take into account any policy intervention during the time period. The government regulated the price for Internet Access to be lower than the voice telephone calls. Such a subsidy through partial-state-owned incumbent, Telekom Malaysia, result in the proliferation of Internet Subscription in Malaysia more than what it would have been under market forces (or failure). This example highlights the important transition step from laboratory to decision table. A modeler needs to incorporate the contextual dynamics that can have a long term impact into the basic model. This price-setting intervention is incorporated as a step input starting at the corresponding year. Resimulation results in an improved fit until year 2000 when the model again underestimates the growth. See Figure 8.

A modeler continues to check with policy makers if there were any policy intervention that must be included in the model. In the case of Malaysia, there were many policies and programmes launched in 1999 that improved the penetration. (Note 1)

### *3.1 What-if scenario*

A model can only be applied to test scenarios and policy alternatives when the end user has confidence in the behaviour of the model. This does not mean that the model must achieve 100% historical fit. This section serves only to illustrate how the model can be used for policy experiments. It does not guarantee accuracy as many dynamics are reserved for future development. For example, lowering the price of Internet by subsidies can increase the level of penetration but also create other consequences. Policy makers and analysts can use the system dynamics model to see the impact of such policy on other areas of policy (e.g. competition policy). Figure 9 shows an example of such a use.

The model shows that by 2003, the Malaysian Internet industry could have at least eight service providers. On the other hand, the model shows that due to price regulation 1996, there are only 2 providers by 2003. This is because the revenue per service provider is lower at lower price and do not attract many firms to enter the market. In fact, until 2000 the Malaysian market remained a duopoly with Telekom Malaysia controlling 70% of the market (ITU, 2001).

## 4. Conclusion

The study builds upon the existing literature to reveal the factors, components and the underlying dynamics that involve in the Internet penetration process. In doing so, it shows how the dynamics and complexity of the interactions can be modeled to give a holistic view of the situation to policy makers. In particular, the Causal-Loop diagramming and time-series plot helps understanding of the process and consequences. The dynamic behaviour of a system can be understood by looking at the time plots and relating the time plots to the causal loop diagram.

As the study seeks policy application, the model has been taken from a generic level to a country specific application in order to show how the transition can gain acceptance by tailoring the model to include contextual dynamics. Due to the holistic view of the approach, a system dynamic model is capable of showing how policy can interact with one another as exemplified by price-lowering subsidies and competition policy. Interaction across many elements also presents policy makers with visual representation of the possible consequences. Therefore, this modeling approach and the generic model have future potential for policy application. However, the model must include more case applications in order to gain confidence. In particular, the parameter estimation can be improved further by including more time-series data. There are variables such as demand elasticity which I did not have access to and can be included in the future.

In further research, it would be useful to conduct a random survey of both Internet users and those not connected to better understand the reasoning behind consumer choices. Additionally, it is useful to incorporate user retention and attrition dynamics to the model.

## References

Bental, B and M. Spiegel. (1995). Network Competition, Product Quality, and Market Coverage in the Presence of Network Externalities. *Journal of Industrial Economics* 43(2). pp. 197-208.

Deaton, A. (1992). *Understanding Consumption*. Oxford: Clarendon Press.

InfoSoc Malaysia. (2000). Discussion paper on Access and Equity: Benchmarking for Progress. Available: http://www.nitc.org.my/resources/INFOSOCMalaysiaRapReport.pdf

ITU. (2001). Internet Diffusion in South East Asia. *ITU Internet Case Studies*. Geneva.

Lemon, Katherine, R.S. Winer. (1995). *A Model of Customer Retention for New-to-the World Product and Services*. Working Paper, Duke University, Durham, NC.

Liebowitz, S.J. and S.E. Margolis. (1994). “Network Externality: An Un- common Tragedy,” *The Journal of Economic Perspectives*, 8: 133-50.

McCraken, G. (1988). *Culture and Consumption: New Approaches to the Symbolic Character of Consumer Goods and Activities*. Indianapolis: Indiana University Press.

Villasis, S. J. (1996). An Optimal Pricing Mechanism for Internets End User's [On-line], University of Idaho. Available: http://www.uidaho.edu/~econ/svthes.pdf

## Notes

Note 1. In October 1999, a PC Ownership Campaign was introduced with government loan and allowing citizen to withdraw their retirement (Employee Provident Fund) contribution to purchase PCs. 'One Home One PC’ campaign offers lower-priced PCs at special fairs. ISps have offered lower-priced PC upon signing up for Internet Service.

Table 1. Potential Subscribers Categories

| Categories | Characteristics |
|---|---|
| Immediate Adopters | The first group of people to subscribe when the Internet was first available. It represents some of university graduates, researchers and teleworkers. They see the value for money of the Internet |
| Imitators | Subscribe to the Internet after being influenced by the “innovators” through social encounters. They either were not aware of the Internet or did not see the value until many people around them subscribe. Able and/or willing to spend their money on computers and the Internet after being convinced on the benefits. |

| | |
|---|---|
| Internet-Cost-Conscious | Possess access lines and computers but not yet subscribe to Internet due to an unfavourable perceived value of the Internet and/or under income constraint |
| Computer-Cost-Conscious | Have access lines but do not have computers due to unfavourable perceived value of them and/or are under income constraint. Usually entertainment supersedes PC purchase. They occasionally use Internet in Cybercafes. |
| Access-Line-Cost-Conscio us | Have no access line or a computer due to an unfavourable perceived value on both and under income constraint. |
| Underserved Areas | People living in remote areas currently not covered by access infrastructure. Consists of a medium-income, low-income and severely poor people. Unlike the last two, the first group can afford the access line, computer and Internet such as local shop owners and estate plantation owners. |
| Education Pipeline | Those who are still in school either at primary, secondary or tertiary level. |
| Marginalised Society | Those with no formal education, illiterate and some primary school drop-outs. |

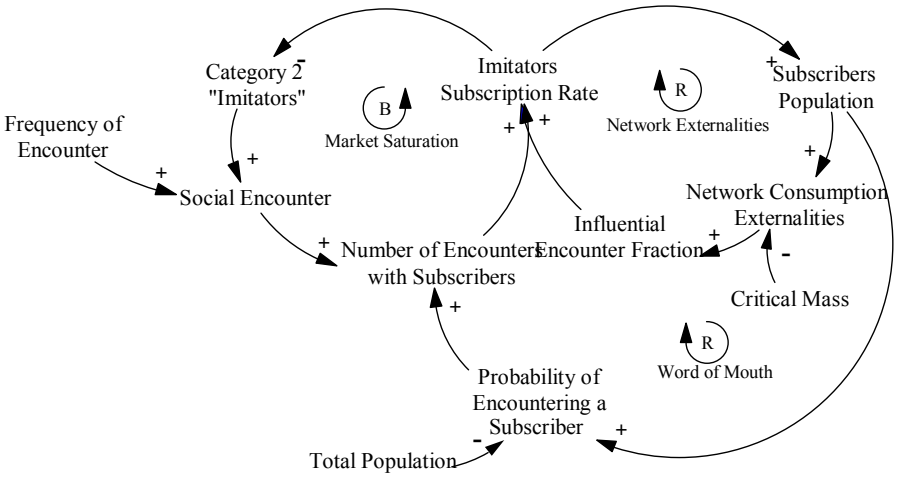



Figure 1. The imitators Subsystem

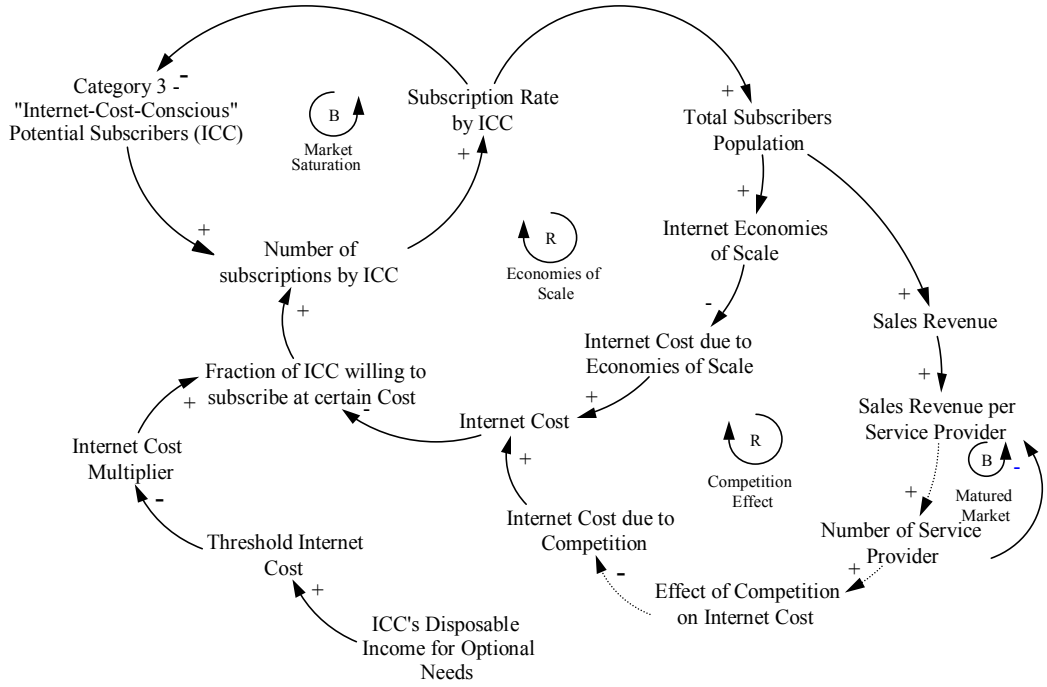



Figure 2. The Internet Cost Conscious Subsystem

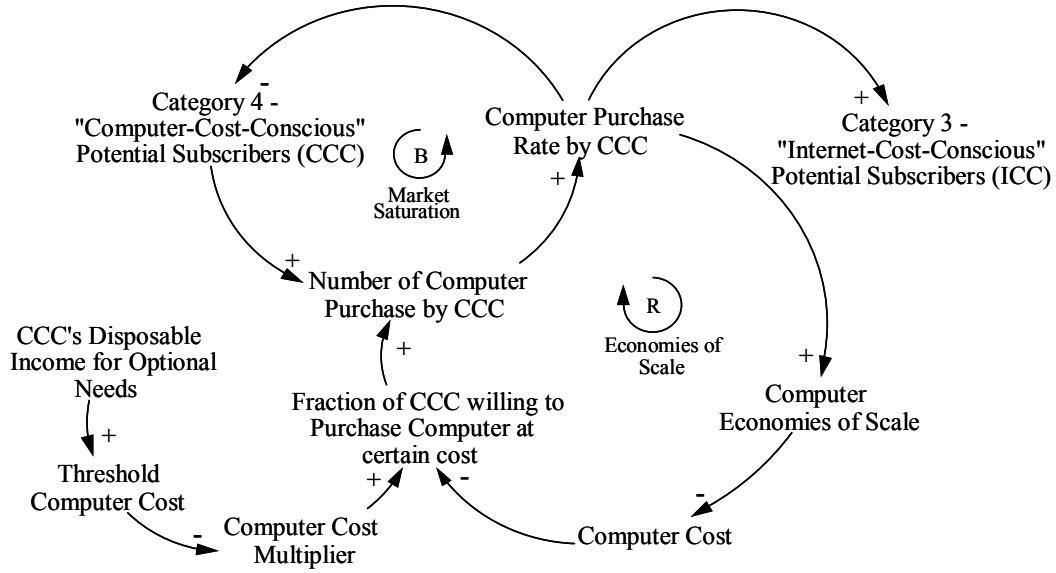



Figure 3. The Computer Cost Conscious Subsystem

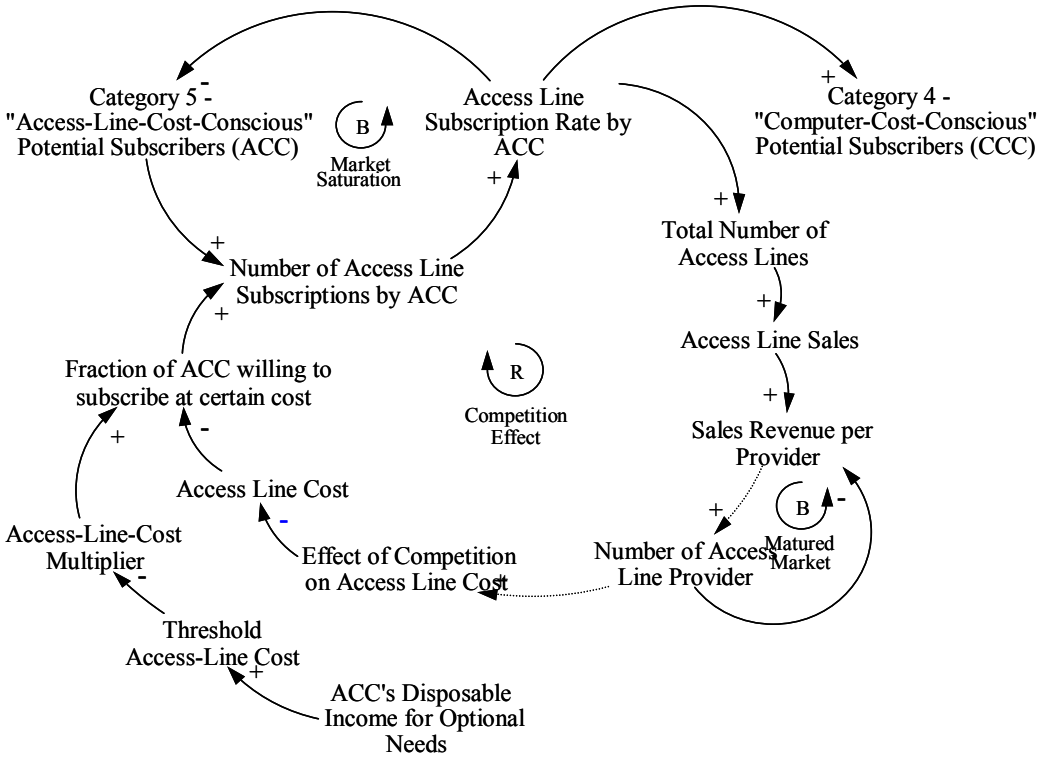



Figure 4. The Access Line Cost Conscious Subsystem

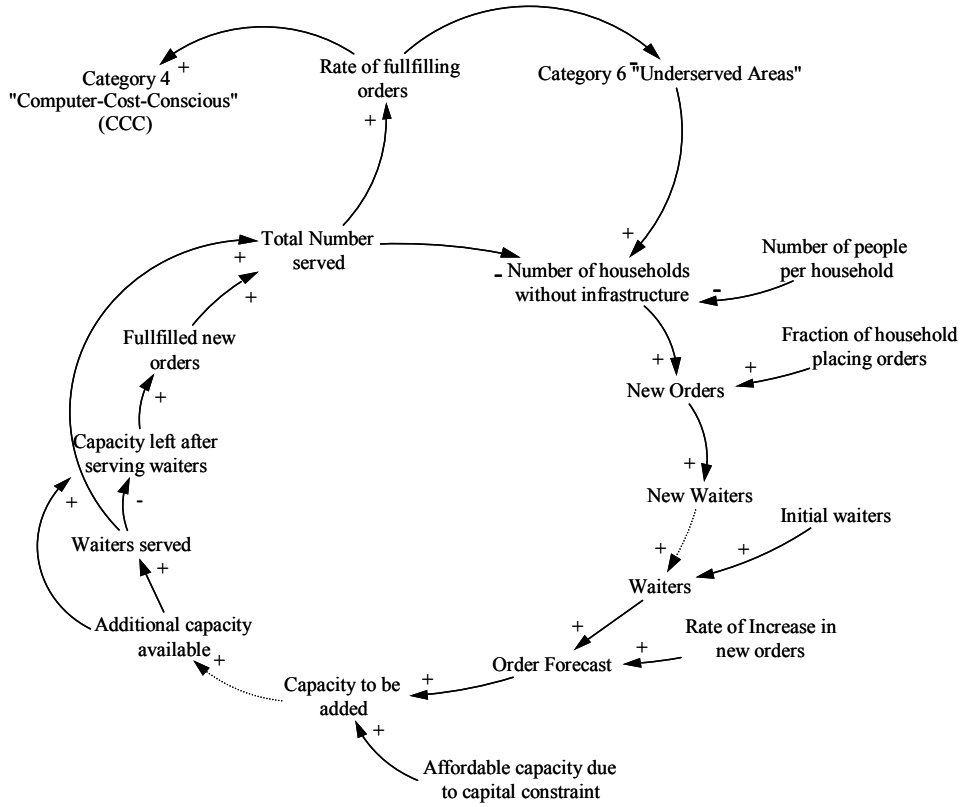



Figure 5. The Underserved Area Subsystem

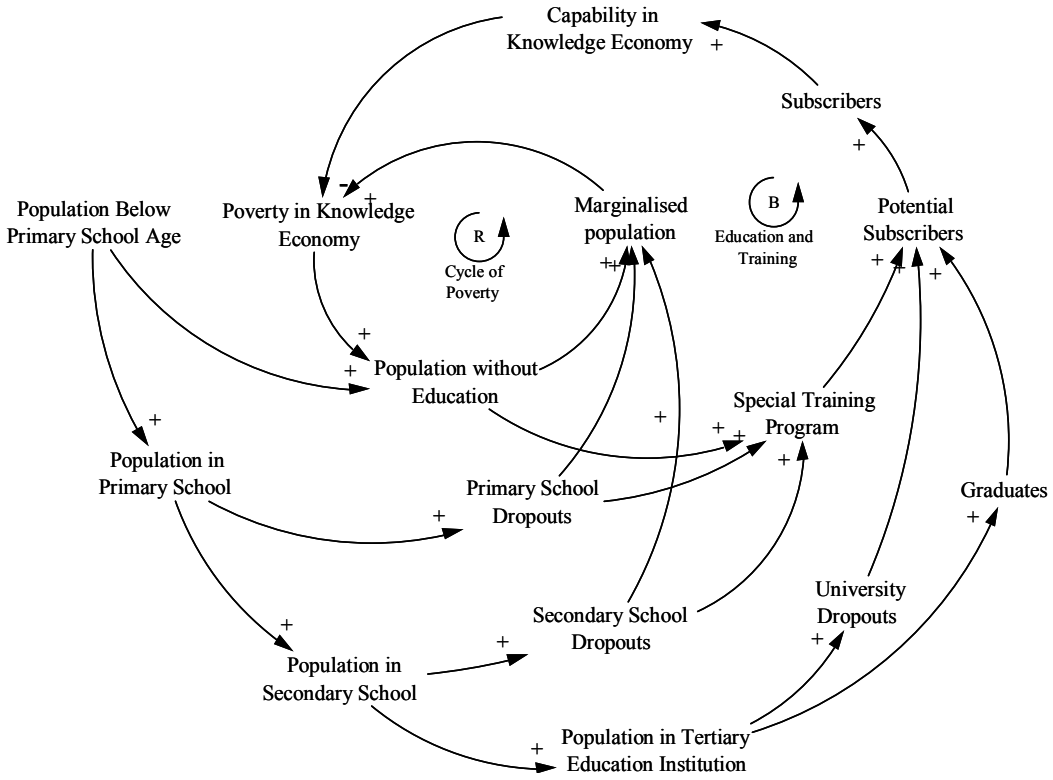



Figure 6. The Educational Subsystem

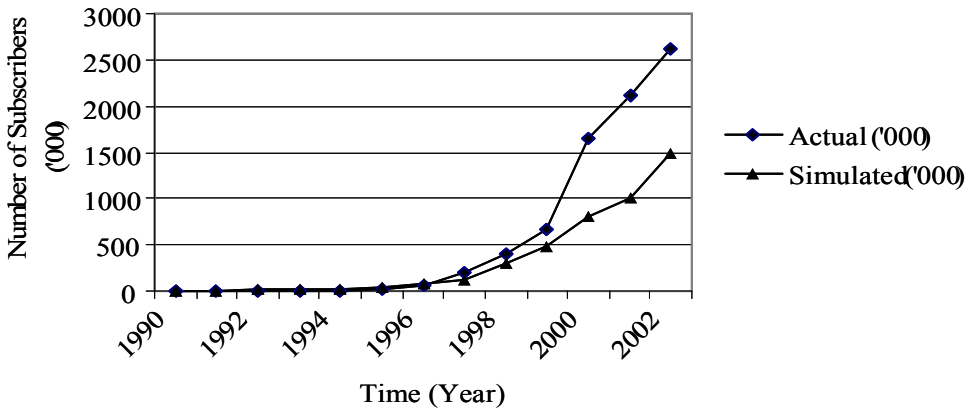



Figure 7. The Malaysian Internet Penetration

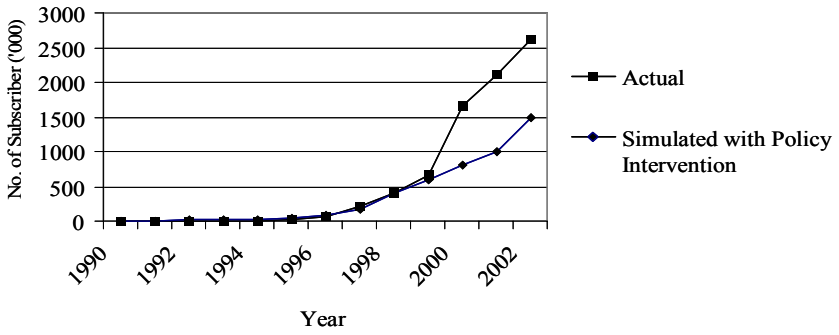



Figure 8. The Malaysian Internet Penetration under Policy Intervention

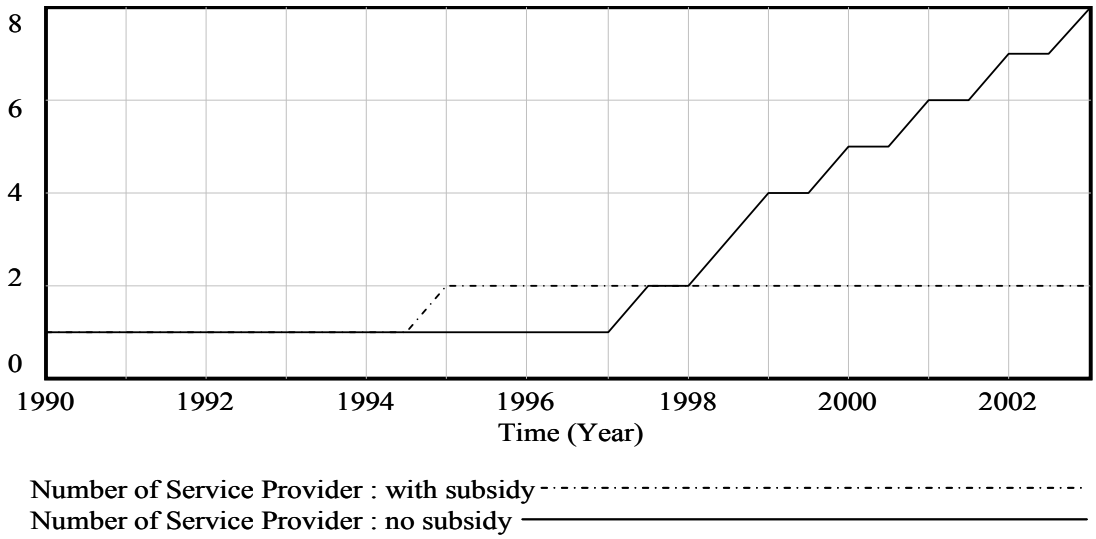



Figure 9. The Number of Service Providers in the Malaysian Internet Market

Computer and Information Science

www.ccsenet.org/journal.html

Vol. 1, No. 1

February 2008

# Autonomic Vehicles

Kevin Curran & James Lee

Intelligent Systems Research Centre

Faculty of Engineering, University of Ulster, Northern Ireland, UK

Tel: +44 2871 375565 E-mail: kj.curran@ulster.ac.uk

**Abstract**

If you thought intelligent cars capable of negotiating a city centre with no driver behind the wheel was just a fantasy of Hollywood – think again. Throughout the history of robotics, robots have been designed to fulfil a multitude of tasks, from opening doors to flying space shuttles to the moon. In this paper, we explain the evolution of robotics in the field of cars and how robotics can be used to make the humans redundant in controlling a car.

**Keywords:** Autonomic vehicles, Autonomous systems, Robotics

## 1. The evolution of robotics through cars

Perhaps the touchstone or founder of the idea of a robot driving a car would have to be the immense success of the computer that is capable of playing sports games such as chess as this opened up the field of progressing artificial intelligence to perform a wide range of tasks (Taney et al., 2005). There are still several leaps of technology to be implemented into robotic autonomous cars before we are capable of creating cars which can navigate through race tracks by themselves akin to the times of professional competitors.

The DARPA Grand Challenge is the de facto annual competition for autonomous vehicles where autonomous cars are required to navigate through a gruelling desert track in the Mojave Desert, Nevada. The first challenge was in 2004 but no car completed the course. In 2006, an autonomous car completed along with several other cars which finished in the allotted time given. This is a great feat in the development of technology for autonomous vehicles on the road but it was not done overnight. There has been years of research into developing these systems. A difficult part is to create an obstacle avoidance system which is capable of detecting not only the immobile but also the mobile objects that will be approaching it. (Graefe & Jacobs, 1991) for instance discuss how this might be achieved by encoding 7 subtasks into the systems recognition system. These are as follows:

(1) Detecting faster vehicles at a distance when they are approaching from behind.

(2) Tracking the detected vehicle while they continue to approach from behind.

(3) Detecting a passing vehicle a second time when it first enters the field of view of a forward-looking camera.

(4) Tracking the vehicle in the image after it has passed the robot, until it is so far ahead that its backside is completely visible.

(5) Changing to a different tracking mode and further tracking the vehicle in the image until its distance exceeds a certain safety limit.

(6) Instantiating and updating a spatio-temporal model of the passing vehicle while it is being tracked to determine its relative speed and distance.

(7) Coarse classification of the passing vehicle (truck, car, motor cycle) while it is being tracked (Graefe & Jacobs, 1991).

The arrows representing the sensors of the robot detecting the car on its approach towards it and past it along with the surrounding objects are depicted in Figure 1.

One of the necessities in robotic vehicles is the development of evolvable intelligence that can de designed and implemented into the functionality of the vehicles. An example is the recent project taken on by the University of Essex led by Dr Simon Lucas of the Department of Computer Science who intend to take the evolution of the artificial intelligence of automated vehicles to the next stage. The proposed system employs an artificial intelligence that will allow the robotic car to essentially think for itself and be able to adapt automatically to changes in surroundings without needing a new subroutine uploaded (Note 1).

Throughout the development of autonomous vehicles there have been several developments in the principles used to create a fully functional artificial intelligence. Some of these principals include invariant features, invariant matching,

localisation and map building (Pradalier and Sekhavat, 2002). These principals are necessary to create a working artificial intelligence which is able to travel from one location to the other by itself which are the stepping stones in an ever evolving artificial intelligence race to develop the first completely autonomous vehicle for common use. When creating an autonomous vehicle there are two essentials, hardware and software. While the software needed to operate the machinery is going to play an integral part of the functionality of the vehicle if it does not have the appropriate equipment to operate then it will not function.

An excellent example of the evolution of the artificial intelligence used in cars is that of Julian Togelius who has created an autonomous vehicle who can drive aggressively through a race track without being programmed how to drive. No human has told the cars how to drive, or indeed how to manoeuvre aggressively so that competitors crash into walls. Instead, evolution strategies (similar to genetic algorithms) were used to evolve neural networks that control the cars (Togelius, 2006). The creation of this technology was developed by firstly using random networks to see which ones operate properly. When these networks are found they are then recreated with minor changes such as in the evolution of a sentient being. The experimental vehicle started of only being able to drive on a standard track using collision detection software. An image of the sort of track used is shown in Figure 4.

As can be seen, the track is extremely simple in design so that the vehicle can be tested to see if the networks being used are functioning within the set parameters. After a training period, the vehicle is allowed to learn other more complicated tracks without any intervention by allowing the artificial intelligence control of the positioning of the sensors so that is more accurate. Problems still exist when two vehicles are racing on a track in that whenever a car is rewarded for being in front of another while racing, the cars behaviour becomes quite aggressive and attempts to force the other off the track (Togelius and Lucas, 2005).

## 2. Conclusion

There have been immeasurable advancements in the field of robotics in the past from the simple artificial intelligence that could play chess to the advanced ones that can navigate a car through a race and adapt to different situations without the intervention of a user. There may indeed be numerous autonomous vehicles sooner than we think.

## References

Tanev, I.; Joachimczak, M.; Hemmi, H.; Shimohara, K. (2005). Evolution of the driving styles of anticipatory agent remotely operating a scaled model of racing car, Evolutionary Computation, 2005. *The 2005 IEEE Congress* on, Vol.2, Iss: 2-5, pp: 1891-1898, Vol. 2, Sept. 2005.

Graefe, V.; Jacobs, U. (1991). Detection of passing vehicles by a robot car driver Intelligent Robots and Systems, *Proceedings IROS, IEEE/RSJ International Workshop*, Vol, Issue, 3-5 Nov 1991 pp: 391-396, Vol.1.

Graefe, V. (1992). Visual Recognition of Traffic Situations by a Robot Car Driver, Intelligent Control and Instrumentation, 1992. *SICICI, Proceedings, Singapore International Conference* on, Volume 1, Issue, 17-21 Feb 1992, pp: 4 – 9.

Pradalier, C., Sekhavat, S. (2002). Concurrent Matching, Localization and Map Building Using Invariant Features. Intelligent Robots and System, 2002. *IEEE/RSJ International Conference* on, Volume: 1, pp: 514- 520, Vol. 1, September 30 - October 4 2002.

Togelius, J. and Lucas, S. M. (2005). Evolving controllers for simulated car racing. In *Proceedings of the Congress on Evolutionary Computation*, 2005.

Togelius, J. and Lucas, S. M. (2006) Arms races and car races. In *Proceedings of Parallel Problem Solving from Nature*. Springer, 2006.

## Notes

Note 1. http://www.essex.ac.uk/news/2007/nr20070320.htm

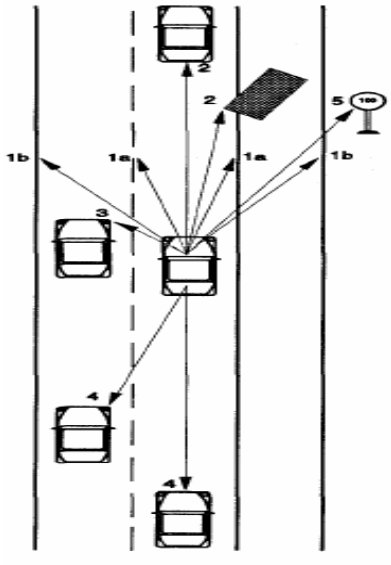



Figure 1. Robotic Sensors (Graefe, 1992)

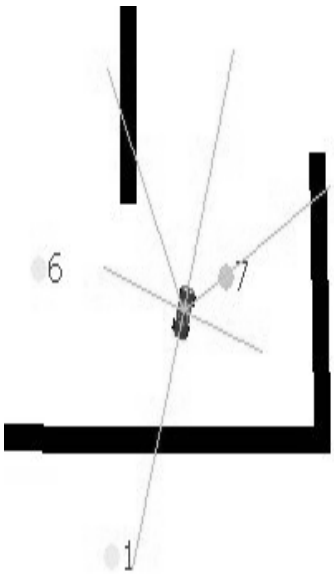



Figure 2. Track (Togelius, 2006)

A journal archived in Library and Archives Canada
A journal indexed in CANADIANA (The National Bibliography)
A journal indexed in AMICUS
A leading journal in computer and information science

# Computer and Information Science

Quarterly
Publisher Canadian Center of Science and Education
Address PO Box 1087, Toronto, Ontario M3A 2Y7
Telephone (416) 585-8198
Fax (416) 585-7247
E-mail CIS@ccsenet.org
Website www.ccsenet.org
Printer William Printing Inc.
Price CAD.$ 20.00

ISSN1913-8989