

Members' Behavior in Virtual Learning Community: A Study Using Data Mining Approach

Xiaokang Li¹, Yu Nie², Min Chen³, Xiaoqing Liu⁴ & Xiaolei Liu¹

¹ Department of Medical Care Management, Xijing Hospital, the Fourth Military Medical University, Xi'an, Shaanxi, China

² Department of Outpatient, Wuhan Institute of Ordnance Officer

³ Department of Digestive Diseases, Xijing Hospital, the Four Military Medical University, Xi'an, Shaanxi, China

⁴ Department of Politics Theory, Faculty of Social Science, the Fourth Military Medical University, Xi'an, Shaanxi, China

Correspondence: Xiaolei Liu, Department of Medical Care Management, Xijing Hospital, the Fourth Military Medical University, No.127 Changle West Road, Xincheng District, Xi'an, Shaanxi, China. Tel: 86-029-8477-5023; Fax: 86-029-8477-3143. E-mail: ken1010169@sina.com

Received: April 21, 2014

Accepted: May 4, 2014

Online Published: August 6, 2014

doi:10.5539/cis.v7n1p1

URL: <http://dx.doi.org/10.5539/cis.v7n4p1>

Abstract

Purpose: With the development of information technology, online virtual learning community is on its way to become an important approach for people to construction and sharing of knowledge. Researches on virtual learning community are not only important to the establishment and management of virtual learning community itself, but are helpful for people's quest for the future development of online learning. However, current researches related to the virtual learning community are in inadequacy, and especially the application of quantitative analysis method for research is rarely seen. Using quantitative analysis method of data mining to study members' behavior in online learning communities. **Method:** In this article, the discussion data (posts) from five online English virtual learning communities in China are sampled and collected. These data were processed according to a series of guidelines to obtain proper data documents, and these data documents were opened under Waikato Environment for Knowledge Analysis and then carried out preprocessing. Next, the module of association rule learning in Waikato Environment Knowledge Analysis were used to perform mining on these processed data, and obtained a series of potential behavior rules in these communities. The partial rules have been listed in the article with their meaning analyzed. **Findings:** The result shows that in this setting it is feasible to apply the association rule learning to virtual learning community. **Value:** It provides approaches and lays the foundation for future relevant studies.

Keywords: association rule learning, behaviour, data mining, online learning, virtual learning, community

1. Introduction

1.1 Virtual Learning Communities and Forums

With the development of information technology (IT), more and more people begin to study on the Internet. In recent years, the application of online cooperation and telecommunication tools to study and education has drawn great interest among people. A new approach to e-learning has emerged with the concept of virtual learning communities (Gaudioso and Talavera, 2006). At present, learners have established spontaneously or non-spontaneously a mass of virtual learning communities of various subjects and professions.

Virtual learning community has emerged as knowledge and information hubs. It is a group of people who gather in cyberspace with the intention of pursuing learning goals (Daniel, McCalla, and Schwier, 2008). It is learning community based not on actual geography, but on shared purpose or interests (Kalogiannakis, 2004). Via technology, learners from different regions gather together in the virtual learning community and establish their formal or informal groups.

Although there still remain many questions such like the viability and veracity of online learning (Song,

Singleton, Hill, and Koh, 2003), the advantages of online learning become increasingly obvious: extensibility, availability, faster speed, lower price, etc. Communication and interaction between online members results in the improvement of the knowledge of each participant in the community and contributes to the development of the knowledge within the domain.

As a new type of information and telecommunication technology nowadays, the online forum undergoes an increasingly expanding application. In the domain of learning and education, it is a tool for promoting conversational modes of learning (Thomas, 2002) and is a significant component of online learning (Marra, Moore, and Klimczak, 2004). There is a large amount of useful data produced in the forum of virtual learning community. Analysis and understanding on these data are helpful for us to identify various interaction modes and their features in the virtual learning communities, while recognition and analysis on these modes and features can assist us in developing a better tool supporting effective interaction, thereby improving the effect and efficiency of online sharing and learning of knowledge, as well as helping us know the social behavior in the virtual learning community and its potential valuable connection.

1.2 Relevant Researches

It is no doubt that the behavior mode of online community members and their relations are complicated and diversified, on which researchers have carried out an amount of work.

Most of the existing researches on user behavior in online communities rely on one of two approaches. Researchers either examine user actions, with basic or complex, if any, analysis of the text or content they post, or collect user opinion through surveys (Hara, Bonk, and Angeli, 2000; Li and Huang, 2008; Marra et al, 2004; Chen and Chang, 2004; Wever, Schellens, Valcke, and Keer, 2006). In the analysis approach, so far, the focus is on the periphery metadata of the messages, such as thread structures, author attributes and their relationships (Jeong, 2003; Marra et al, 2004; Xia, Zhai, Tan, Lu, and Mei, 2008). Very little in-depth analysis of the text posted by users is done. In the survey approach, the sampling bias and response bias may be hard to overcome (Grandcolas, Rettie and Marusenko, 2003; Xia, et al, 2008). In either approach, or even when the survey data can be pooled with user action data, much of the information embedded in text has not been fully taken advantage of. Given the vast amount and increasing availability of user generated text data, and the progress of data mining techniques, the time is ripe to add data mining as an important tool to help study user behavior.

In this article, we are in an attempt to use the association rule theory in data mining and sociological method to mine and analyze the forum data in the virtual learning communities.

1.3 Association Rule Mining

Data mining technology is a rising interdisciplinary field, and is emerging and developing in the course of accumulation of data and active demand for information and knowledge in market competition. It is hoped that people can find valuable information contained in data by data mining technology, and then find undiscovered knowledge. The information and knowledge “contained” in data can not be determined by the priori knowledge and experience, and is helpful for decision making.

Data mining is the process of extracting desirable knowledge or interesting patterns from existing databases for specific purposes (Hong, Kuo, and Chi, 1999). In other words, its function and purpose is to generate a computation or enumeration of expression for fact (Han and Kamber, 2000). Most data mining methods are based on tried and tested techniques from machine learning, pattern recognition, and statistics: classification, clustering, regression, and so forth (Fayyad, Shapiro, and Smyth, 1996; Kantardzie, 2005). The method adopted in this article is association rule learning.

As an important branch of data mining, the association rule learning was first proposed by Agrawal and other researchers in 1993 (Agrawal, Imielinski, and Swami, 1993). It is a popular and well researched method for discovering interesting relations between variables and data. With the in-depth research of related work, association rule has been widely used in many areas including commerce, computer science, medicine, etc (Imberman, Domanski, and Thompson, 2002). At present, the application of association rule learning in the field of education and learning is at the early stage. Association rule learning is a successful and important mining task, aiming at uncovering all frequent patterns among sets of data attributes (Zaki and Ogihara, 1998). The association rule in a detailed manner will be discussed in the following passage.

2. Materials and Methods

2.1 Data Collection

The online English learning community was chosen as the research object of the article. In China nowadays,

English learning is very popular in China. Nearly all learners, from pupils to college students, need to learn English. Five online English learning communities were chosen out of relatively popular and successful domestic ones (Qu, 2010; Meng, 2010). See Table 1 for detailed information about them. The communities' registers are from all places of the country, mainly are college students in all grades aging from 18 to 24. Since they are keen English learners, a vivid and rustling virtual community has been originated by their spontaneous joining, through which experiences are exchanged, puzzles are explored, resources are shared and information is searched out. All the data in this article are from these five websites, which are managed well and have many active and vigorous registered users, thereby they can better present the characteristics of domestic virtual English learning community.

Table 1. The five popular virtual English learning communities for college students in China

Name	Website	Number of texts selected
Xiaomuchong	http://emuch.net/bbs/forumdisplay.php?fid=126&page=1&type=110	85
Online Reading Garden	http://www.readfree.net/bbs/thread.php?fid=27	152
Aguo	http://www.agpr.net/bbs/thread.php?fid=43.html	58
Dingxiangyuan	http://www.dxy.cn/bbs/post/page?bid=11&sty=1&tpg=1&s=52&age=0	128
Renren	http://bbs.rrenglish.com/forum-10-1.html	77
Total	5 websites	500 texts

2.2 Principles of Association Rule Learning

In the field of data mining, the association rule learning is a popular and well-studied method for finding the relations among different variables in the large-size database. Shapiro and Frawley (1991) described analyzing and presenting strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawal and Srikant (1994) introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-sale (POS) system in supermarkets. For instance, the rule found in the scales data of a supermarket, {onion, tomato} \rightarrow {beef} (Support=30%, Confidence=60%) will reveal that if a buyer has purchased both onions and tomatoes, then he is likely to buy beef. This kind of information can be availed as the foundation of business behavioural decisions. Apart from the instance of analysis on the shopping basket of supermarket, the association rules are also applied to the fields of webpage mining, invasion detection and bioinformatics today. The core of association rule learning is the algorithm of association rules. Up to now, many algorithms used for generating association rules have been raised. Some famous ones include Apriori, Eclat, and FP-Growth. Apriori is the best-known algorithm to mine association rules. In the field of computer science, Apriori algorithm is a classic algorithm for the application of association rule learning. Apriori is designed to operate to databases containing transactions.

Support and Confidence exist accompanied with association rules, to some extent; the two are the necessary supplement for the rules. In the abovementioned case, sales data analysis for supermarkets: Support=30% refers the probability of simultaneous emerging onion, tomato and beef in all sales business. Suppose the probability in appearing onion, tomato and beef at the same time is quite low, it means there being little relation among onion, tomato and beef; to the contrary, frequent probability reflects the knowledge that onion, tomato and beef are always associated with turn out to be the common sense. Confidence=60% means the probability in appearing beef under the circumstance of there being onion and tomato in all sales business. 100% Confidence level indicates that beef will definitely appear when onion and tomato appear. Therefore, there is every reason for bundling sale. Excessive low Confidence level conveys that onion and tomato are not closely related with beef.

For more detailed knowledge of association rules, please refer to:

http://en.wikipedia.org/wiki/Association_rule_learning and

http://wapedia.mobi/en/Association_rule_learning (The wapedia reference seems to offer little that is not in the text of the paper. Which came first? Is there a copyright issue?)

2.3 Preprocessing of the Sample Data — Comment on the Threads

The posts posted in these virtual communities were collected during the period from August 15 to October 15 in 2011. A large number of students use the Internet frequently for learning purposes during the period. Apart from the advertisements and those posts irrelevant with English learning, there are 500 posts in total. In these posts, some call for help, some communicate, some publish opinions, and some share interesting ideas. Then, information in these posts was processed, in which information reflecting the potential behavior characteristics of the online learning members was picked and collected. To be specific, in this step, we read carefully the sample post and marked the key information in each post according to pre-formulated post evaluation system which contains about 10 indexes (Li, 2009; Sun and Mao, 2003). Table 2 shows the detailed information of this system. The meaning of each index is illustrated as follows:

Language_attitude refers to the tone of the learner expressed in his/her post. Description refers to whether the information described in the post can be understood easily. Length refers to information quantity (character number) contained in the post. having three value ranges: 0-100 characters (short), 101-300 characters (medium), and 301-maximum (long). Type refers to the categories of post contents, including knowledge sharing, help seeking, emotion expression, interpersonal communication, etc. Click refers to the browsing quantity of a post by people, and it has several value ranges from zero to the maximum value. Reply refers to the replying times by people on one post, and it also has several value ranges from zero to the maximum value. Source refers to where the post comes from, including reprinting, self-compilation, and other ways. Correlation refers to whether the post content is relevant to the community it is in and to the post title. This index is proposed for solving the common problem that the post content is not consistent with the forums it is in and with its title. This index is proposed for solving the common problem in many online forums or discussion groups that the post content is not consistent with the forum or discussion group it is in and with its title. Form_of_expression refers to in what way a post is expressed, including character, picture, video, hypertext link, etc. Result refers to whether a result is obtained after a discussion on the subject one post involves among online learners. Content refers to the type of the theme one post involves, and in this research, it includes vocabulary, grammar, writing, listening, etc.

Table 2. Post evaluation system

Index	Detailed Content
Language_attitude	good, normal, bad
Description	clear, normal, unclear
Length	short(0_100), medium(101_300), long(301_max)
Type	knowledge sharing, help seeking, discussion, emotion expression, interpersonal communication, others
Click	quantity of clicks (0_100,101_200,201_500,501_max)
Reply	quantity of replies (0_20,21_50,51_200,201_max)
Source	reprint, self-compilation, others
Correlation	yes, no
Form_of_expression	character, image, video, hypertext link, f_complex (comprehensive form) have a result, have no result
Result	vocabulary, grammar, writing, listening, tongue, reading, entertainment,
Content	translation, methodology, others, c_complex (comprehensive content)

We can see from Table 2 that the entire index system can be classified into two types: the numeric index and the nominal index. Numeric index includes Click and Reply, and the rest part belongs to nominal index. For numeric indexes, the quantity was divided into some ranges of value to convert them into nominal data. This is a very important procedure in the application of Waikato Environment of Knowledge Analysis (WEKA). WEKA (Frank, Hall, and Holmes, et al) is famous software for data mining at present. WEKA (Version 3.5.8) is used to mine the information in the text.

In the process of reading and collecting of the post texts, it is found that nearly all texts, based on the post evaluation system, share the same value in such three items as “Language attitude”, “Description” and “Correlation”. In a bid to avoid producing invalid or meaningless rules, these three attributes were deleted from

WEKA before mining.

2.4 Some Settings in WEKA before Mining

Parameter setting is an important step for data mining in using WEKA. Various results will be shown if setting is done in accordance with different requirements. For an association rule like $L \rightarrow R$, Support and Confidence are usually used to evaluate its importance. Support is used to estimate the probability of L and R appear at the same time in the basket, while Confidence are usually used to estimate the probability of R appears in the basket when L appears. Association rules aim at producing rules with high Support and Confidence level. Several similar measures of Confidence are set in WEKA to measure the association degree of rules, which are Lift (sometimes written as Improvement), Leverage and Conviction respectively. Simply speaking, as far as the association rule learning is concerned, the larger the values of these indexes are, the more reliable the rules are.

In the research, we plan to explore such association rules whose Support falls between 20% and 100%, and Lift value exceeds 1.5 and ranks top 100. LowerBoundMinSupport and upperBoundMinSupport are set as 0.2 and 1 respectively, metricType as Lift, minMetric as 1.5, numRules as 100, and other options remain as default.

For more knowledge of WEKA functions and usage, please refer to <http://www.cs.waikato.ac.nz/ml/weka/>

3. Mining Results & Analysis

One particularly important point is that this is an exploratory investigation, which establishes the general potential of data mining of discussion postings – not particular findings to be taken from these discussion postings, which are a particular case.

Mining configurations in WAKE were set and 100 rules which were best supported and most credible are displayed. Here we only list and analyze rules which we think are of much importance and meaning in this setting. See Table 3 for the detailed information.

Table 3. Part of mining results and index related to association degree

No.	Mining results (rules)	Confidence	Lift	Leverage	Conviction
1	Type=help_seeking ==> Reply=0_20 Source=other	1.00	2.52	0.23	103.22
2	Type=help_seeking ==> Reply=0_20 Form_of_expression=character	0.95	1.72	0.15	7.66
3	Type=help_seeking ==> Length=0_100 Reply=0_20 Source=other	0.92	2.52	0.21	7.73
4	Type=help_seeking ==> Length=0_100 Source=other	0.92	2.50	0.21	7.70
5	Length=0_100 ==> Type=Knowledge_sharing Source=reprint Form_of_expression=f_complex	0.87	2.11	0.18	4.36
6	Type=knowledge_sharing ==> Source=reprint	0.96	1.71	0.23	10.30
7	Length=0_100 Source=reprint ==> Form_of_expression=f_complex	0.88	2.00	0.17	4.47
8	Length=0_100 ==> Type=knowledge_sharing Form_of_expression=f_complex	0.88	1.99	0.17	4.32
9	Length=0_100 ==> Type=knowledge_sharing Reply=0_20 Source=reprint	0.94	1.75	0.16	7.47

1. Type =help_seeking ==> Reply=0_20 Source=other

2. Type = help_seeking ==> Reply=0_20 Form_of_Expression=character

The two rules indicate that, with respect to the posts belonging to help seeking, few helping hands are stretched out. Two reasons are found to contribute to the phenomenon: the first is the ones seeking help – the problems they raised are generally inappropriate, either too general or too troublesome. It's the major cause. For example, some learners want other learners to provide a certain kind of leaning software. Others are in want of information sources which are too private or scarce. The second is, from an objective perspective, the problems put forward by help-seekers are too difficult to be answered by people, which reminds us that, we'd better avoid raising "invalid" problems in order to get satisfying answers or help.

3. Type = help_seeking ==> Length=0_100 Reply=0_20 Source=other

4. Type =help_seeking ==> Length=0_100 Source=other

The posts seeking help are mostly expressed in characters and not too long. The two rules above tell us that characters (texts) remain the main expression way to seek help in online learning. In fact, not only the behavior of seeking help, but also other kind of behaviors is expressed in characters (texts). Online interaction overwhelmingly takes place by means of discourse and participants interact by means of verbal languages, usually typed on a keyboard and read as text on a computer screen (Herring, 2004). For this reason, when we raise questions in virtual learning communities on the Internet, brief and terse wording is the best choice, while long and wordy description often makes people tiring, thereby being ignored.

5. Length =0_100 Type=knowledge_sharing ==> Source= reprint Form_of_Expression = f_complex

6. Type =knowledge_sharing ==> Source=reprint

The above two rules tell us that the posts of knowledge sharing are generally from reprint. When people feel like sharing some knowledge, experience or information, they are inclined to make use of the resources available from others or other places, which implicates the online learners' lack in originality spirit to some extent. This is a very important problem. Thank to IT, people now can easily acquire and reveal information, and can search information of any topic in the search engine. Too much "copy" and "paste" were used in one's articles, which are very helpful to improve the learning efficiency. However, the ability of independent thinking is weakened unconsciously on the other hand. The problem now should arouse wide concerns.

7. Length = 0_50 Source = reprint ==> Form_of_expression=f_complex

The rule states the posts reprinted are generally not only in characters but in various ways, such as picture, link or video.

8. Length = 0_50 Type=knowledge_sharing ==> Form_of_expression=f_complex

The rule shows knowledge-sharing posts are not only expressed in characters but in various ways, such as picture, link or video.

It can be drawn from the two rules above that diversified expression ways are very popular in online learning. With the support of IT, people can easily post a thread containing characters, pictures, videos and/or super links, which provide much more information, and are more interesting and more easily understood by people than pure characters.

9. Length=0_50 Type=knowledge_sharing ==> Reply=0_20 Source=reprint

The rule indicates that only a fraction of learners give comments or feedbacks on posts which fall into knowledge sharing. This does not mean people are not interested in such posts. On the contrary, posts of this category are much popular with learners. In fact, people are focused more on browsing or downloading the information than on posting their viewpoints. We also find that many learners in the online learning community always lurk conscientiously: they seek others' help and use the information provided by others, but seldom help others or have interactions with the other learners, and they rarely participate in the discussion on a certain topic in the community. They seldom post a deeply meaningful and far from short contribution which shows that they were ruminating while they lurked. They gain much while devote less. Nonnecke, Andrews, and Preece (2006) described the nonpublic participation within an online community and name it as lurking. He also examined the nature of lurking, why people lurk and the differences in attitudes between lurkers and posters. In our opinion, it is not favorable for the development of online community that members lurk for a long time. Measures should be taken by the website administrator or guide to activate the lurkers.

Only part of rules are listed in the above contents. As a matter of fact, many other interesting rules could be

gotten by changing the value range of parameters in the setting procedure. One point is the change of value range of parameters should be within an appropriate range. Interesting rules refer to effective, novel, and potential-for-use ones that can be understood finally. Some rules are only significant in statistics, but not in practical application, which is not understandable.

4. Conclusion

In the article, a quantitative research on the members' behaviors in the virtual learning community by using the data mining technology are made, and some meaningful rules are concluded. From the viewpoint of the whole process and the final result of research, it is feasible to apply the data mining technology to the virtual learning community. Compared with other methods, association rule learning is more prone to get the features of the virtual learning community. The research result helps the researchers have a rational recognition on the members' behaviors in the virtual learning community, and makes the traits of virtual learning community in the minds of the researchers and community members clear, distinct and regular. The article conducts a probe into the virtual learning community by a quantitative research, and has laid a foundation for the future work.

The research has still many limitations. In our setting, the research object is relative simple, and the evaluation index system of posts is not very sound. More over, explanations of some mining results need for further discussion. In the future research, we will further improve this technology and apply it to virtual communities of other types.

Acknowledgements

We thank the official web site of WEKA (<http://www.cs.waikato.ac.nz/ml/weka/>) for providing the software used in this paper.

References

- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of International Conference on Very Large Data Bases (VLDB'94)*, 487-499, Santiago, Chile.
- Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of ACM-SIGMOD International Conference on Management of Data (SIGMOD'93)*, 207-216, Washington, DC. <http://dx.doi.org/10.1145/170035.170072>
- Chen, H., & Chang, C. (2004). The analysis of an open online discussion forum in science. In L. Cantoni, & C. McLoughlin (Eds.), *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications* (pp. 4125-4131). Chesapeake, VA.
- Daniel, B. K., McCalla, G. I., & Schwier, R. A. (2008). Social network analysis techniques: Implications for information and knowledge sharing in virtual learning communities. *International Journal of Advanced Media and Communication*, 2(1), 20-34. <http://dx.doi.org/10.1504/IJAMC.2008.016212>
- Eibe, F., Mark, H., & Geoffrey et al. (2010). Weka-A Machine Learning Workbench for Data Mining. 2010, Part 8, 1269-1277. *Data Mining and Knowledge Discovery Handbook*. http://dx.doi.org/10.1007/978-0-387-09823-4_66
- Fayyad, U., Shapiro, G. P., & Smyth, P. (1996). Knowledge discovery and data mining: towards a unifying framework. *2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 82-88. Portland, Oregon.
- Gaudiose, E., & Talavera, L. (2006). Data mining to support tutoring in virtual learning communities: experiences and challenges. In C. Romero, & S. Ventura (Eds.), *Data mining in e-learning* (pp. 207-226). Southampton, UK: Wit Press.
- Grandcolas U., Rettie, R., & Marusenko K. (2003). Web survey bias: sample or mode effect? *Journal of marketing management*, 19(5-6), 541-561. <http://dx.doi.org/10.1080/0267257X.2003.9728225>
- Han, J., & Kamber, M. (2000). *Data Mining: Concepts and Techniques*, 3-43. San Francisco: Morgan Kaufman.
- Hara, N., Bonk, C. J., & Angeli, C. (2000). Content analysis of online discussion in an applied educational psychology course. *Instructional Science*, 28(2), 115-152. <http://dx.doi.org/10.1023/A:1003764722829>
- Herring, S. C. (2004). Computer-mediated discourse analysis – an approach to researching online behavior. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for virtual communities in the service of learning* (pp. 338-367). Cambridge: Cambridge University Press.
- Hong, T. P., Kuo, C. S., & Chi, S. C. (1999). Mining association rules from quantitative data. *Intelligent Data*

- Analysis*, 3(5), 363-376. <http://dx.doi.org/10.3233/IDA-1999-3504>
- Imberman, S. P., Domanski, B., & Thompson, H. W. (2002). Using dependency/association rules to find indications for computed tomography in a head trauma dataset. *Artificial Intelligence in Medicine*, 26(1), 55-68. [http://dx.doi.org/10.1016/S0933-3657\(02\)00052-0](http://dx.doi.org/10.1016/S0933-3657(02)00052-0)
- Jeong, A. C. (2003). The sequential analysis of group interaction and critical thinking in online threaded discussion. *The American Journal of Distance Education*, 17(1), 25-43. http://dx.doi.org/10.1207/S15389286AJDE1701_3
- Kalogiannakis, M. (2004). A virtual learning environment for the French physics teachers. *Education and Information Technology*, 9(4), 345-353. <http://dx.doi.org/10.1023/B:EAIT.0000045292.59105.a2>
- Kantardzie, M. (2005). Data mining: concepts, models, methods, and algorithms. *Journal of Computing and Information Science in Engineering*, 5(4), 394-395.
- Li, Y., & Huang, R. (2008). An analysis of online interaction discourse in knowledge forum. *Proceedings of the 2008 International Conference on Computer Science and Software Engineering*, 45-48. Wuhan, China. <http://dx.doi.org/10.1109/CSSE.2008.34>
- Li, Z. R. (2009). The three standard of a good post based on “Archive World”. *Archive Management*, 177, 74-75.
- Marra, R. M., Moore, J. L., & Klimczak, A. K. (2004). Content analysis of online discussion forums: A comparative analysis of protocols. *Educational Technology Research and Development*, 52(2), 23-40. <http://dx.doi.org/10.1007/BF02504837>
- Meng, T. (2010). Dingxiangyuan: From establishing a business to “double-win”. *China Health Human Research*, 62-65.
- Nonnecke, B., Andrews, D., & Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1), 7-20. <http://dx.doi.org/10.1007/s10660-006-5985-x>
- Qu, B. Q. (2010). Scientific cooperation and rethinking--take “xiaomuchong” as an example. *Science and Technology Management and Research*, 215-218.
- Shapiro, G. P., & Frawley, W. J. (1991). *Knowledge Discovery in Databases*, 229-248. Cambridge, AAAI/MIT Press.
- Song, L., Singleton, E. S., Hill, J. R., & Koh, M. H. (2003). Improving online learning: Student perceptions of useful and challenging characteristics. *The Internet and Higher Education*, 7(1), 59-70. <http://dx.doi.org/10.1016/j.iheduc.2003.11.003>
- Sun, Y., & Mao, B. (2003). Behavior of Members in virtual community based on data mining. *Computers Applications*, 23(1), 50-53.
- Thomas, M. J. W. (2002). Learning within incoherent structures: the space of online discussion forums. *Journal of Computer Assisted Learning*, 18(3), 351-366. <http://dx.doi.org/10.1046/j.0266-4909.2002.03800.x>
- Wever, B. D., Schellens, T., Valcke, M., & Keer, H. V. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, 46(1), 6-28. <http://dx.doi.org/10.1016/j.compedu.2005.04.005>
- Xia, M., Zhai, C., Tan, B., Lu, Y., & Mei, Q. (2008). “You Are What You Write” – understanding user online behavior through text mining. In CHI 2009 workshop of Social Mediating Technologies
- Zaki, M. J., & Ogihara, M. (1998). *Theoretical Foundations of Associations Rules*. Retrieved April 15, 2014, from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.6006&rep=rep1&type=pdf>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).