English Language Teaching; Vol. 14, No. 1; 2021
ISSN 1916-4742 E-ISSN 1916-4750
Published by Canadian Center of Science and Education

A Methodological Review of Machine Learning in Applied Linguistics

Zhiqing Lin[1]

[1] Faculty of English Language and Culture, Guangdong University of Foreign Studies, Guangzhou, China

Correspondence: Zhiqing Lin, Faculty of English Language and Culture, Guangdong University of Foreign Studies, No. 2 North Baiyun Avenue, Guangzhou, China.

Received: October 10, 2020 Accepted: November 21, 2020 Online Published: December 23, 2020

doi: 10.5539/elt.v14n1p74 URL: https://doi.org/10.5539/elt.v14n1p74

Abstract

The traditional linear regression in applied linguistics (AL) suffers from the drawbacks arising from the strict assumptions namely: linearity, and normality, etc. More advanced methods are needed to overcome the shortcomings of the traditional method and grapple with intricate linguistic problems. However, there is no previous review on the applications of machine learning (ML) in AL, the introduction of interpretable ML, and related practical software. This paper addresses these gaps by reviewing the representative algorithms of ML in AL. The result shows that ML is applicable in AL and enjoys a promising future. It goes further to discuss the applications of interpretable ML for reporting the results in AL. Finally, it ends with the recommendations of the practical programming languages, software, and platforms to implement ML for researchers in AL to foster the interdisciplinary studies between AL and ML.

Keywords: applied linguistics, machine learning, linear regression, interdisciplinary studies

1. Introduction

The past few years have witnessed the increasing awareness on the importance of statistical methods in linguistics (Khany & Tazik, 2019; Nikitina & Furuoka, 2018; Norris et al., 2015). The reason might be the fact that statistical approaches play a vital role in investigating the variables in linguistics. One of the most commonly used methods is linear regression. But this algorithm suffers from the handicaps of strict assumptions including normality, linearity, and homoscedasticity (Plonsky & Ghanbar, 2018). This gives rise to the applications of more advanced algorithms to tackle more complicated linguistic problems. These new technologies, to a large extent, are represented by machine learning (ML). However, this is no systematic review of the applications of ML in applied linguistics (AL). Most of the reviews on the methods employed in AL are on traditional methods, for example, linear regression. Besides, how to report and interpret the result of ML model remains elusive to most of the researchers in applied linguistics. The introduction of corresponding ways to exploit the ML is needed. Therefore, this paper attempts to fill in these gaps by summarizing the applications of ML in AL, introducing the interpretable ML, and presenting suggested approaches to implement ML for researchers in AL.

2. Literature Review

2.1 Definitions of Keywords

Machine learning refers to the process of figuring out the underlying pattern of data by computers automatically instead of designing any man-made rule presumably. ML can be classified into two categories: supervised learning and unsupervised learning. The definition of AL adopted in this essay refers to the studies on language and language-relevant problems in which people use or learn languages as what was defined by Lei and Liu (2019). This essay principally focuses on the studies on the topic of language teaching, language learning, language testing, corpus linguistics, psycholinguistics, and phonetics. These topics will be discussed in detail in section 4.

2.2 Previous Studies on Traditional Methods

There are several studies on the statistics used in linguistics research. Plonsky and Oswald (2017) did a systematic review of the multiple regressions in the second language (L2) research with comparisons to ANOVA. Concerning linear mix-effects models, Meteyard and Davies (2020) conducted a study on this topic. They stated

the concern over the application of this method in psycholinguistics. Nicklin and Plonsky (2020) delineated the problem of outliers in L2 research and they summarized the present methods adopted to deal with outliers. As for factor analysis, Plonsky and Gonulal (2015) did a review of the exploratory factor analysis in linguistic research. Lindstromberg (2016) reviewed inferential statistics in English language teaching research and emphasized the unsuitability of *p* values. Norris (2015) conducted a review on the statistical significance testing in L2 research and he figured out the problem of statistical testing and argued for the directions to reform. King and Mackey (2016) summarized the developmental trend of the research methodology employed in L2 research. Paquot and Plonsky (2017) conducted a review on the quantitative methods in corpus linguistics, including ANOVA, factor analysis, and resampling. Moreover, researchers, for example, Norouzian (2020), also researched the sample size planning problem in second language research. Nikitina and Furuoka (2018) discussed the application of quantile regression with bootstrapping to deal with the non-normal data and outliers. These reviews and research made great contributions to our understanding of the methodology in applied linguistics. But the problem is that these studies focus on traditional methods exclusively. The problem of strict assumptions has not been solved and the demand for more advanced methods to tackle more complicated problems in applied linguistics has not been met. This leaves the room to investigate machine learning in applied linguistics.

2.3 Brief Introductions to ML

What follows below is the introduction of the typical algorithms in ML including logistics, K nearest neighbor, Bayesian model, support vector machine, random forests, XGBoost, clustering, and neural networks. Logistics is actually a generalized linear model. K nearest neighbor is somewhat regarded as the simplest algorithms in ML for its main idea lies in calculating the geometric distance and then finding out the best result. The Bayesian model is founded on probability. Support vector machine is the representative algorithm in ML before 2005. Its algorithmic idea is the projection of original data into higher dimensional space in which the data can be effectively tackled. Random forests and XGBoost algorithms are two major algorithms in ensemble learning. The former is based on bagging and the latter is based on boosting. The neural network is the imitation of the human brain and it is renamed deep learning in recent years, which stands for the state-of-the-art technology in AI. It has several different versions including recurrent neural networks, long short-term neural networks, convolutional neural networks, etc. (Barredo Arrieta et al., 2020).

2.4 Motivations to Review the Applications of ML in AL

The rationale for reviewing the applications of ML in AL can be summarized as follows. The first motivation has something to do with the drawbacks of linear regression. The traditional linear regression has several flaws, including normality assumption, linearity assumption, collinearity, and robustness.

First of all, linear regression is based on the normality assumption. This assumption comes from the theory of central limited theorem, but it turns out to be problematic when the data deviate from the normal distribution. It has been a long time since scholars argued for more advanced methods that do not rely on normal distribution for linguistic studies (Nicklin & Plonsky, 2020). According to a review by Hu and Plonsky (2019), there are a large number of L2 studies that did not follow assumption-check strictly. This means that the research on L2 without the normality-check may be problematic to some extent.

In addition, linear regression is based on the linearity assumption, which may depart from the nonlinear development pattern according to the Complex Dynamic Systems Theory in AL (Lesonen et al., 2020). It cannot wrestle with nonlinear data. More importantly, linear regression fails to calculate the importance of variables when predictors are collinear. Some variables have to be deleted in this case (Tomaschek et al., 2018). This poses a problem to the research on applied linguistics (Wurm & Fisicaro, 2014). Besides, the traditional methods such as linear regression are sensitive to the influence of outliers which may lead to some mistakes in the final results. Moreover, linear regression is built up upon the ideal condition in which the independent variables are calculated without any interaction effect. The final disadvantage of linear regression is the side effect brought by significant testing. We cannot make the final decision on whether one variable has any relationship with another simply by significance testing. Table 1 is the summary of the comparisons between traditional linear regression and ML-based methods

Table 1. Comparisons between traditional linear regression and machine learning

	Linear regression	ML-based methods
Normality assumption	Yes	No
Linearity assumption	Yes	No
Collinearity problem	Yes	No
Outlier sensitivity	Yes	No
Interpretability	Easy to interpret	Relative hard to interpret
Accuracy	Relatively low	Relatively high

Another motivation for this synthesis resides in the fact that ML, particularly the neural network, was criticized for its opacity. The applications of ML in AL, for example, automatic scoring, were blamed for ML-based systems cannot be interpreted. A review on how to explain the results of ML is necessary. Aside from the aforementioned two reasons, the third motivation of this review is justified by the learning curve of ML. Due to the complexity of ML, linguistic researchers may have difficulty in implementing the ML to solve the problems in AL. For this reason, this thesis will also cover the most user-friendly programming language, software, and platform for researchers in AL.

With these three motivations taken into account, this thesis is going to answer the following questions:

Question 1: What contributions has machine learning made to different branches of AL?

Question 2: What about the interpretability of ML in AL?

Question 3: What is the suitable approach for the researchers in AL to make use of ML?

3. Research Methodology

3.1 Data Collection

3.1.1 Inclusion Criteria

The inclusion criteria are listed as follows: First, ML should be adopted to solve relevant problems. Second, the problem should have something to do with applied linguistics. Third, those essays which can shed light on how to apply ML to AL are also included.

3.1.2 Literature Research

Google scholar and web of science are utilized to collect the data. First of all, the author tries to search the keywords related to ML in Web of Science with the list of linguistic journals ranked according to their impact. The ranking was done by Web of Science and further information about this can be checked online. These keywords include K nearest neighbor, naïve bay and Bayesian networks, support vector machine, random forests, XGBoost, neural networks, clustering, machine learning, data mining, and artificial intelligence. In order to capture all the related literature, google scholar is also used to cover relevant literature. Moreover, the exemplary studies from adjacent disciplines are also added, but these studies only account for a very small proportion. The reason for the inclusion is that these studies can enlighten us on how to carry out cross-disciplinary studies between AL and ML. Most of the papers are from the Social Science Citation Index (SSCI) journals.

3.1.3 Exclusion Criteria

As this thesis focuses on the applications of ML in AL, the papers should be related to both ML and AL at the same time. If the essay concentrates on only one aspect, it will be eliminated.

3.2 Data Coding and Grouping

After the iteration of researching and eliminating, all the essays are read by the author one by one. As there are many branches of AL, the results are grouped based on similarity. Some of the branches of AL are with no application according to the search results and these branches will not be reviewed.

4. Results

4.1 Results of Question 1

Table 2 shows the selected essays which comply with the aforementioned criteria. The similar branches of AL are grouped as one item. Most of them will be discussed in detail and some will not for brevity.

Table 2. Representative research

Applied linguistic branches	Representative studies
Language teaching Language learning	(Charitopoulos et al., 2020; Crossley, 2013; Kelly et al., 2018; Pliakos et al., 2019; Rico-Juan et al., 2019; Wiechmann & Kerz, 2014; Yang & Li, 2018)
Language testing	(Almond et al., 2007; Chapelle & Chung, 2010; Cui et al., 2016; O. Kang & Johnson, 2018; Kumar & Boulanger, 2020; Latifi & Gierl, 2020; Liu & Cheng, 2017; Man et al., 2019; Rudner, 2016; Shin & Gierl, 2020; Tomasevic et al., 2020; Zheng et al., 2020; Zopluoglu, 2019)
Second language research	(Gudmestad et al., 2013; Norouzian et al., 2018; Papi & Teimouri, 2014; Warschauer et al., 2019)
Corpus linguistics	(Ballier et al., 2020; Deshors 2020a, 2020b; Deshors & Gries, 2016; Fonteyn & Nini, 2020; Frey, 2020; Her & Tang, 2020; Hilpert, 2016; H. Kang & Yang, 2020; Meurers & Dickinson, 2017; Shawar & Atwell, 2005; Sung et al., 2015; Th Gries, 2020; van Halteren et al., 2005; Wan et al., 2019; Xiao & Sun, 2020)
Clinical linguistics	(Armstrong et al., 2018; Fergadiotis et al., 2016; Geetha et al., 2000; Gillespie et al., 2018; Keshet, 2018; Reed & Wu, 2013; Vasquez-Correa et al., 2018)
Psycholinguistics Neurolinguistics	(Fromont et al., 2020; Heikel et al., 2018; Monner et al., 2013; Munsell et al., 2019; Pearl & Enverga, 2014)
Phonetics	(Al-Tamimi & Khattab, 2018; Arnhold & Kyrolainen, 2017; Bybee & De Souza, 2019; Charalabopoulou et al., 2011; Howell et al., 2017; Litman et al., 2018; Przybyla & Teisseyre, 2014; Xie et al., 2019)

4.1.1 Language Teaching and Language Learning

In language teaching and learning, the contribution of ML is manifested in technology-enhanced teaching and learning. A study conducted by Crossley (2013) systematically sketched the picture of researching second language writing through computational tools and ML. Wiechmann and Kerz (2014) tapped the potential of applying ensemble learning to fit the compositions written by German. It was proven that decision trees and random forest algorithm can help us understand the writing. Kelly et al. (2018) focused on the questions in the real-world classroom. He built up a model by ML to automatically measure the authenticity of questions in the classroom. Yang and Li (2018) employed backpropagation neural networks to estimate student's performance and identified the factors that may influence their final achievement. Pliakos et al. (2019) proposed a system that combined machine learning and item response theory to address the cold start problem in adaptive learning. Rico-Juan et al. (2019) researched peer-assessment in the classroom with ML to detect the discrepancy between numerical scores and feedback. Charitopoulos et al. (2020) summarized the applications of data mining algorithms which showed us how these algorithms can be applied to language teaching in the future. Although some of the examples are from the adjacent area, they set up very quintessential examples for language teaching and learning activities. Further studies on the interdisciplinary studies between ML and language teaching can draw on the valuable experience from these cases. In conclusion, ML may render it possible to boost language teaching and learning by ML-enhanced systems.

4.1.2 Language Testing

Concerning language testing, the applications of ML in language testing can be summarized in several aspects: test administration, automatic essay scoring, cognitive diagnostic assessment. Chapelle and Chung (2010) conducted a comprehensive overview of the possibility of applying natural language processing and speech recognition in language testing. Regarding the test administrations, Rudner (2016) conducted a study on the Bayesian probability theory and found that it can be used to furnish students with feedback no matter whether the test is unidimensional or not. Man et al. (2019) set up a system that can detect test fraud by both supervised ML and unsupervised ML and Zopluoglu (2019) also developed a detecting system by XGBoost. As for automatic scoring, Latifi and Gierl (2020) adopted ML and natural language processing techniques to

automatically predict the score by Coh-Metrix features. Shin and Gierl (2020) applied convolutional neural networks to the essay scoring system and the result showed that neural networks can outperform traditional methods. O. Kang and Johnson (2018) created a model by ML algorithm to predict the English oral proficiency with suprasegmental features. Tomasevic et al. (2020) conducted a review on the application of supervised machine learning in predicting student's test performance. Litman et al. (2018) did a broad review on the approach, challenges, and opportunities that new technology, namely speech recognition based on ML has brought to second language speaking assessment. Kumar and Boulanger (2020) introduced how deep learning can be applied in automatic essay scoring with the example of the Kaggle competition. Zheng et al. (2020) took advantage of ML to develop an adaptive test system. In terms of cognitive diagnostic assessment (CDA), Almond et al. (2007) introduced how the Bayesian networks can be applied to model CDA. Liu and Cheng (2017) applied support vector machines to CDA in the small sample size context and Cui et al. (2016) implemented CDA by the neural network. From what has been discussed above, it can be learned that automatic essay scoring and automatic speech scoring which is based on the ML technology is likely to be a new trend in the future and student's competence might be diagnosed more and more precisely with the advance of ML.

4.1.3 Second Language Research

In second language research (L2), probably, the most frequently mentioned ML algorithm is Bayesian analysis. Gudmestad et al. (2013) delineated how Bayesian analysis can be applied in second language acquisition. Another study that is similar to Gudmestad was done by Norouzian et al. (2018). In their research, Bayesian networks were shown as a powerful tool in second language research and they argued for a Bayesian revolution. Apart from the Bayesian networks, clustering was proven very useful in second language research. Papi and Teimouri (2014) grouped the language learner's motivations by clustering. Warschauer et al. (2019) introduced how clustering algorithms can boost vocabulary learning. Crowther et al. (2020) systematically reviewed the applications of clustering algorithms in L2 research. What can be learned from above is that Bayesian analysis has been exploited by linguists in second language research. Clustering can help us group students into different categories based on which geared language teaching and learning might be possible.

4.1.4 Corpus Linguistics

Within the scope of corpus linguistics, perhaps, the most frequently adopted algorithm is random forests. A methodical review of random forest on corpus linguistics was done by Th Gries (2020). Fonteyn and Nini (2020) employed the random forest algorithm and conditional inference trees to analyze gerunds. Deshors (2020a) applied random forests to investigate the contextualized past tense and the interactions between variables. Deshors argued that this method can overcome the assumption of normality as what had already been mentioned in the literature review. In 2020, he did another research by random forests to investigate multi-speaker interactions. Frey (2020) provided a very comprehensive overlook of the algorithms of ML in corpus linguistics. Apart from random forests, clustering was introduced in corpus linguistics by Hilpert (2016). Moreover, Sung et al. (2015) applied support vector machines to classify the readability of second language reading texts with reference to the Common European Framework of Reference (CEFR). H. Kang and Yang (2020) quantified the political bias using machine learning. Ballier et al. (2020) reviewed a Kaggle competition which employed machine learning algorithms and natural language processing techniques to automatically score essays. It seems that the random forest algorithm is the most frequently used algorithm in corpus linguistics and other algorithms of ML also show great potentials.

4.1.5 Clinical Linguistics

In the field of clinical linguistics, the contribution of ML is mainly embodied in the application of diagnosis of language disorder. Geetha et al. (2000) employed artificial neural networks to classify childhood disfluencies using neural networks with 92% accuracy. Logistics is applied by Reed and Wu (2013) for risk factor modeling. Fergadiotis et al. (2016) classified different paraphasic errors by a series of ML algorithms. Keshet (2018) exploited ML algorithms to build up an automatic speech recognition system for pathology researchers. Vasquez-Correa et al. (2018) utilized ML algorithms (both linear and nonlinear methods) to automatically evaluate the patients with dysarthria. Armstrong et al. (2018) built up a regressor to predict the language difficulties by both ML and traditional linear regression. Gillespie et al. (2018) applied ML to identify the affective state change for grownups with aphasia by acoustic features. Xie et al. (2019) took advantage of ML to handle the data of neurophysiological responses. It is obvious that ML can complement traditional methods and improve the accuracy of the diagnosis of language disorder.

4.1.6 Psycholinguistics and Neurolinguistics

In the area of psycholinguistics and neurolinguistics, one of the most important contributions ML has made is the implications of the neural network framework for language processing models. Concretely speaking, the architecture in neural networks can shed light on the language learning models, especially in bilingual studies. Zhao and Li (2010), for example, talked about the bilingual lexical interactions from the perspective of neural networks. Monner et al. (2013) explained the language learning phenomenon from the perspective of neural networks. Frank (2020) discussed how the recurrent neural networks can enlighten us on multilingual sentence processing models. It can be learned from above that ML algorithms, especially neural networks, have gradually received attention from the researchers specializing in psycholinguistics or neurolinguistics. One plausible reason may be that the neural network is the convergent point where both psycholinguists or neurolinguistics and computer scientists are focusing on. And the researchers from both sides seek to draw on the strength from each other. Furthermore, magnetic resonance imaging (MRI) is currently adopted by neurolinguistics. It is possible that the research on the diagnosis of language disorder and Alzheimer's problem by MRI can be enhanced with help of ML. One typical example was done by Basaia et al. (2019). Heikel et al. (2018) recorded the evolution of the neurocognitive process by a machine learning method called multivariate pattern analysis. Pearl and Enverga (2014) developed a mind-print-based system by machine learning to identify the mental state. Munsell et al. (2019) applied machine learning algorithms to predict the performance of naming in temporal lobe epilepsy. Fromont et al. (2020) applied random forests to model the individual data and found that language exposure and proficiency were the most important predictive variables. All in all, the neural network and ML algorithms may show a bright future in psycholinguistics and neurolinguistics.

4.1.7 Phonetics

In phonetics, the major contribution of ML is mainly embodied in acoustic feature importance ranking and automatic speech recognition system. Al-Tamimi and Khattab (2018) employed both random forests and linear mixed models to find out the most predictive indicators for distinguishing different acoustic stops. Przybyla and Teisseyre (2014) analyzed the utterances to train a regressor to predict the speaker's background by several ML algorithms. The results showed that random forests and k nearest neighbor algorithm outperformed other algorithms. Arnhold and Kyrolainen (2017) investigated the focus marking by both random forests and the generalized additive mixed algorithm with the spotlight on the variable importance. This can help us understand phonetics and develop a speech scoring system that can be applied either in language testing or clinical linguistics, etc. Support vector machine and linear discriminant analysis are employed by Howell et al. (2017) to train a classifier by different kinds of speeches. Bybee and De Souza (2019) analyzed the vowel duration in two different constructions by random forest analysis based on conditional inference trees. It can be seen from above that ML can also be applied to phonetics as long as the original acoustic data can be digitalized. After that, researchers in phonetics can establish a model by ML to solve related problems.

4.2 Results of Question 2

After training a classifier or a regressor, the accuracy or confusion matrix marks can be obtained. But sometimes we are far more interested in the importance of input variables. This is a problem on the interpretability of ML. Admittedly, the applications of ML in AL, for example, automatic scoring, was criticized for the drawback that the system cannot be interpreted. Therefore, the problem of interpretability will be discussed in this part.

To begin with, the random forest algorithm might be the most popular algorithm adopted by researchers in AL. The underlying reason might be the fact that the information of feature importance can be informed. More importantly, it enjoys great suitability whether the data follow the normal distribution or not. It is still applicable when predictors are collinear and works for the data with large numbers of predictors and limited samples (Matsuki et al., 2016). Interaction effect will also be taken into consideration by random forests (Baumann & Winter, 2018). Here are some linguistic studies by random forest algorithm. Her and Tang (2020), for example, ranked the feature importance by random forest to understand the predictive power of input variables. There are also some other similar cases, such as Deshors (2020b), and Wiechmann and Kerz (2014). As for decision trees, the result of decision trees can be visualized. This can help us understand how the system works. For example, Fromont et al. (2020) illustrated the effect of individual variability by visualizing decision trees.

However, neural networks cannot be explained as easily as decision trees or random forests. It is for this reason that some automatic scoring system based-on ML in applied linguistics was criticized. As a matter of fact, there is an alternative method called Shap value (Ribeiro et al., 2016) which can explain neural networks. Actually, Shap can explain any classifier or regressor. Frey (2020) had introduced this method in his doctoral dissertation on corpus linguistics. With this method, the automatic scoring system by deep learning can be validated by

linguistic practitioners on the one hand. It can also, on the other hand, offer precious guidance for us. This method is hardly adopted by linguists but it was already applied in other science disciplines. Further information on explainable ML could be learned from the paper by Barredo Arrieta et al. (2020). Plots of partial dependence and individual conditional expectation are also very useful methods to peek inside the black-box. Further information can be found in the essay by Adadi and Berrada (2018). Studies on AL in the future should put the interpretable ML into full use.

4.3 Results of Question 3

After reviewing the applications and interpretability of ML, the following question is how to make it possible for all the linguistic researchers to have access to these techniques. This question will be answered from three aspects: programming languages, software, and platforms for ML.

To begin with, the most recommended programming language for ML should be Python language followed by R language. Python is ranked as the third most popular language in computer science and it is gaining more and more popularity for its simplicity and convenience. It is open freely to the public. R language is also very popular in both computer science and AL (Mizumoto & Plonsky, 2016). As for the library for implementing the ML algorithms, Sklearn definitely is the best candidate for it enjoys numerous mighty functions for all kinds of algorithms (Hao & Ho, 2019) and Keras based on TensorFlow is a flexible and powerful tool to carry out neural networks (Pang et al., 2019). Pytorch is also attracting more and more users in recent years, especially in the academic circle. The recommended platform for R language is RStudio. The recommended platform for Python language is Spyder or Jupyter notebook supported by Anaconda.

5. Conclusions

As for the contributions ML has made to AL, they can be further summarized as follows: computer-assisted language learning and language teaching, identification, diagnosis, automatic scoring, and feature importance ranking. Language learning and language teaching will be more adaptive and personalized with help of ML. It can also help us identify the test fraud automatically and provide fined-grained information about the student's ability. Automatic scoring based on ML makes it possible to increase the reliability and validity of scoring in language testing. The diagnostic system based on ML plays an important role in clinical linguistics because ML makes it possible to diagnose the patient with the language disorder problem within a short period of time. Finally, ML algorithms show greater suitability than traditional linear regression. Random forests, for example, can deal with complex data types notwithstanding the normality, linearity, and collinearity assumptions. And the interaction between variables can also be investigated. Further studies in AL should make use of interpretable ML to gain more information from data. From the viewpoint of question type, the problems that can be solved by ML include regression problems, classification problems, clustering problems, and dimensionality reduction problems. Researchers in AL should pay more attention to these four kinds of problems in which interdisciplinary studies between ML and AL can be carried out. It seems that ML is applicable in most branches of AL. The data in AL include digital numbers, natural language, pictures, and other objects as long as they can be digitalized. And it might help us solve AL problems towards the trend of automatization. Most importantly, ML might continue to help researchers in AL delve into and deal with the more perplexing linguistic problems that traditional methods cannot solve (Gass et al., 2020). All in all, the application of ML in AL holds an advantage over traditional methods for its superiority in accuracy and flexibility.

6. Limitations and Directions for Further Studies

Regardless of various efforts to circumvent possible flaws, this thesis still has the following weaknesses. Primarily, there is no clear cut between different sub-fields of AL branches. The essay assigned to one branch of AL can also be classified into another. Studies in the future can overcome this defect by setting up clear and fine-grained standards. Furthermore, this essay focuses on the finite range of linguistic journals. Some related meeting papers and doctoral dissertations may be missed although some typical studies from the outside circle of AL are included. Finally, this thesis concentrates on limited branches of AL. But this does not mean that ML is inapplicable to other branches. Brown et al. (2014), for example, applied random forests to solve pragmatic problems. Studies in the future can sweep all the possibly related papers to depict a more panoramic picture.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access, 6*, 52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052

Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J.-D. (2007). Modeling diagnostic assessments with bayesian networks. *Journal of Educational Measurement, 44*(4), 341-359.

https://doi.org/10.1111/j.1745-3984.2007.00043.x

Al-Tamimi, J., & Khattab, G. (2018). Acoustic correlates of the voicing contrast in lebanese arabic singleton and geminate stops. *Journal of Phonetics, 71*, 306-325. https://doi.org/10.1016/j.wocn.2018.09.010

Armstrong, R., Symons, M., Scott, J. G., Arnott, W. L., Copland, D. A., McMahon, K. L., & Whitehouse, A. J. O. (2018). Predicting language difficulties in middle childhood from early developmental milestones: A comparison of traditional regression and machine learning techniques. *Journal of Speech Language and Hearing Research, 61*(8). https://doi.org/10.1044/2018_Jslhr-L-17-0210

Arnhold, A., & Kyrolainen, A. J. (2017). Modelling the interplay of multiple cues in prosodic focus marking. *Laboratory Phonology, 8*(1). https://doi.org/10.5334/labphon.78

Ballier, N., Canu, S., Petitjean, C., Gasso, G., Balhana, C., Alexopoulou, T., & Gaillat, T. (2020). Machine learning for learner english: A plea for creating learner data challenges. *International Journal of Learner Corpus Research*. https://doi.org/10.1075/ijlcr.18012.bal

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., . . . Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion, 58*, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012

Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., & Filippi, M. (2019). Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks. *NeuroImage: Clinical, 21*, 101645. https://doi.org/10.1016/j.nicl.2018.101645

Baumann, S., & Winter, B. (2018). What makes a word prominent? Predicting untrained german listeners' perceptual judgments. *Journal of Phonetics, 70*, 20-38. https://doi.org/10.1016/j.wocn.2018.05.004

Brown, L., Winter, B., Idemaru, K., & Grawunder, S. (2014). Phonetics and politeness: Perceiving korean honorific and non-honorific speech through phonetic cues. *Journal of Pragmatics, 66*, 45-60. https://doi.org/10.1016/j.pragma.2014.02.011

Bybee, J., & De Souza, R. N. (2019). Vowel duration in english adjectives in attributive and predicative constructions. *Language and Cognition, 11*(4), 555-581. https://doi.org/10.1017/langcog.2019.32

Chapelle, C. A., & Chung, Y.-R. (2010). The promise of nlp and speech processing technologies in language assessment. *Language Testing, 27*(3), 301-315. https://doi.org/10.1177/0265532210364405

Charalabopoulou, F., Stafylakis, T., & Mikros, G. K. (2011). Developing a scoring algorithm for automatic pronunciation assessment of modern greek. *Journal of Quantitative Linguistics, 18*(1), 1-22. https://doi.org/10.1080/09296174.2011.533586

Charitopoulos, A., Rangoussi, M., & Koulouriotis, D. (2020). On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010-2018. *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-020-00200-8

Crossley, S. A. (2013). Advancing research in second language writing through computational tools and machine learning techniques: A research agenda. *Language Teaching, 46*(2), 256-271. https://doi.org/10.1017/S0261444812000547

Crowther, D., Kim, S., Lee, J., Lim, J., & Loewen, S. (2020). Methodological synthesis of cluster analysis in second language research. *Language Learning*. https://doi.org/10.1111/lang.12428

Cui, Y., Gierl, M., & Guo, Q. (2016). Statistical classification for cognitive diagnostic assessment: An artificial neural network approach. *Educational Psychology, 36*(6), 1065-1082. https://doi.org/10.1080/01443410.2015.1062078

Deshors, S. C. (2020a). Contextualizing past tenses in l2: Combined effects and interactions in the present perfect versus simple past alternation. *Applied Linguistics*. https://doi.org/10.1093/applin/amaa017

Deshors, S. C. (2020b). English as a lingua franca: A random forests approach to particle placement in multi-speaker interactions. *International Journal of Applied Linguistics, 30*(2), 214-231. https://doi.org/10.1111/ijal.12275

Deshors, S. C., & Gries, S. T. (2016). Profiling verb complementation constructions across new englishes a two-step random forests analysis of ing vs. To complements. *International Journal of Corpus Linguistics, 21*(2), 192-218. https://doi.org/10.1075/ijcl.21.2.03des

Fergadiotis, G., Gorman, K., & Bedrick, S. (2016). Algorithmic classification of five characteristic types of

paraphasias. *American Journal of Speech-Language Pathology, 25*(4), S776-S787. https://doi.org/10.1044/2016_AJSLP-15-0147

Fonteyn, L., & Nini, A. (2020). Individuality in syntactic variation: An investigation of the seventeenth-century gerund alternation. *Cognitive Linguistics, 31*(2), 279-308. https://doi.org/10.1515/cog-2019-0040

Frank, S. L. (2020). Toward computational models of multilingual sentence processing. *Language Learning*. https://doi.org/10.1111/lang.12406

Frey, J. C. (2020). *Using data mining to repurpose german language corpora. An evaluation of data-driven analysis methods for corpus linguistics* (unpublished doctoral dissertation). University of Bologna, Bologna, Italy.

Fromont, L. A., Royle, P., & Steinhauer, K. (2020). Growing random forests reveals that exposure and proficiency best account for individual variability in l2 (and l1) brain potentials for syntax and semantics. *Brain and Language, 204*. https://doi.org/10.1016/j.bandl.2020.104770

Gass, S., Loewen, S., & Plonsky, L. (2020). Coming of age: The past, present, and future of quantitative SLA research. *Language Teaching*, 1-14. https://doi.org/10.1017/S0261444819000430

Geetha, Y. V., Pratibha, K., Ashok, R., & Ravindra, S. K. (2000). Classification of childhood disfluencies using neural networks. *Journal of Fluency Disorders, 25*(2), 99-117. https://doi.org/10.1016/S0094-730X(99)00029-7

Gillespie, S., Laures-Gore, J., Moore, E., Farina, M., Russell, S., & Haaland, B. (2018). Identification of affective state change in adults with aphasia using speech acoustics. *Journal of Speech Language and Hearing Research, 61*(12), 2906-2916. https://doi.org/10.1044/2018_JSLHR-S-17-0057

Gudmestad, A., House, L., & Geeslin, K. L. (2013). What a bayesian analysis can do for sla: New tools for the sociolinguistic study of subject expression in l2 spanish. *Language Learning, 63*(3), 371-399. https://doi.org/10.1111/lang.12006

Hao, J., & Ho, T. K. (2019). Machine learning made easy: A review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics, 44*(3), 348-361. https://doi.org/10.3102/1076998619832248

Heikel, E., Sassenhagen, J., & Fiebach, C. J. (2018). Time-generalized multivariate analysis of eeg responses reveals a cascading architecture of semantic mismatch processing. *Brain and Language, 184*, 43-53. https://doi.org/10.1016/j.bandl.2018.06.007

Her, O.-S., & Tang, M. (2020). A statistical explanation of the distribution of sortal classifiers in languages of the world via computational classifiers. *Journal of Quantitative Linguistics, 27*(2), 93-113. https://doi.org/10.1080/09296174.2018.1523777

Hilpert, M. (2016). Cluster analysis for corpus linguistics. *International Journal of Corpus Linguistics, 21*(4), 581-585. https://doi.org/10.1075/ijcl.21.4.07hil

Howell, J., Rooth, M., & Wagner, M. (2017). Acoustic classification of focus: On the web and in the lab. *Laboratory Phonology, 8*(1). https://doi.org/10.5334/labphon.8

Hu, Y., & Plonsky, L. (2019). Statistical assumptions in l2 research: A systematic review. *Second Language Research*, 0267658319877433. https://doi.org/10.1177/0267658319877433

Kang, H., & Yang, J. (2020). Quantifying perceived political bias of newspapers through a document classification technique. *Journal of Quantitative Linguistics*, 1-24. https://doi.org/10.1080/09296174.2020.1771136

Kang, O., & Johnson, D. (2018). The roles of suprasegmental features in predicting english oral proficiency with an automated system. *Language Assessment Quarterly, 15*(2), 150-168. https://doi.org/10.1080/15434303.2018.1451531

Kelly, S., Olney, A. M., Donnelly, P., Nystrand, M., & D'Mello, S. K. (2018). Automatically measuring question authenticity in real-world classrooms. *Educational Researcher, 47*(7), 451-464. https://doi.org/10.3102/0013189X18785613

Keshet, J. (2018). Automatic speech recognition: A primer for speech-language pathology researchers. *International Journal of Speech-Language Pathology, 20*(6), 599-609. https://doi.org/10.1080/17549507.2018.1510033

Khany, R., & Tazik, K. (2019). Levels of statistical use in applied linguistics research articles: From 1986 to 2015. *Journal of Quantitative Linguistics, 26*(1), 48-65. https://doi.org/10.1080/09296174.2017.1421498

King, K. A., & Mackey, A. (2016). Research methodology in second language studies: Trends, concerns, and new directions. *The Modern Language Journal, 100*(S1), 209-227. https://doi.org/10.1111/modl.12309

Kumar, V. S., & Boulanger, D. (2020). Automated essay scoring and the deep learning black box: How are rubric scores determined? *International Journal of Artificial Intelligence in Education*. https://doi.org/10.1007/s40593-020-00211-5

Latifi, S., & Gierl, M. (2020). Automated scoring of junior and senior high essays using coh-metrix features: Implications for large-scale language testing. *Language Testing*. https://doi.org/10.1177/0265532220929918

Lei, L., & Liu, D. (2019). Research trends in applied linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics, 40*(3), 540-561. https://doi.org/10.1093/applin/amy003

Lesonen, S., Steinkrauss, R., Suni, M., & Verspoor, M. (2020). Dynamic usage-based principles in the development of L2 finnish evaluative constructions. *Applied Linguistics*. https://doi.org/10.1093/applin/amaa030

Lindstromberg, S. (2016). Inferential statistics in language teaching research: A review and ways forward. *Language Teaching Research, 20*(6), 741-768. https://doi.org/10.1177/1362168816649979

Litman, D., Strik, H., & Lim, G. S. (2018). Speech technologies and the assessment of second language speaking: Approaches, challenges, and opportunities. *Language Assessment Quarterly, 15*(3), 294-309. https://doi.org/10.1080/15434303.2018.1472265

Liu, C., & Cheng, Y. (2017). An application of the support vector machine for attribute-by-attribute classification in cognitive diagnosis. *Applied Psychological Measurement, 42*(1), 58-72. https://doi.org/10.1177/0146621617712246

Man, K., Harring, J. R., & Sinharay, S. (2019). Use of data mining methods to detect test fraud. *Journal of Educational Measurement, 56*(2), 251-279. https://doi.org/10.1111/jedm.12208

Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The random forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading, 20*(1), 20-33. https://doi.org/10.1080/10888438.2015.1107073

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language, 112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Meurers, D., & Dickinson, M. (2017). Evidence and interpretation in language learning research: Opportunities for collaboration with computational linguistics. *Language Learning, 67*(S1), 66-95. https://doi.org/10.1111/lang.12233

Mizumoto, A., & Plonsky, L. (2016). R as a lingua franca: Advantages of using r for quantitative research in applied linguistics. *Applied Linguistics, 37*(2), 284-291. https://doi.org/10.1093/applin/amv025

Monner, D., Vatz, K., Morini, G., Hwang, S.-O., & DeKeyser, R. (2013). A neural network model of the effects of entrenchment and memory development on grammatical gender learning. *Bilingualism: Language and Cognition, 16*(2), 246-265. https://doi.org/10.1017/S1366728912000454

Munsell, B. C., Wu, G., Fridriksson, J., Thayer, K., Mofrad, N., Desisto, N., . . . Bonilha, L. (2019). Relationship between neuronal network architecture and naming performance in temporal lobe epilepsy: A connectome based approach using machine learning. *Brain and Language, 193*, 45-57. https://doi.org/10.1016/j.bandl.2017.08.006

Nicklin, C., & Plonsky, L. (2020). Outliers in l2 research in applied linguistics: A synthesis and data re-analysis. *Annual Review of Applied Linguistics, 40*, 26-55. https://doi.org/10.1017/S0267190520000057

Nikitina, L., & Furuoka, F. (2018). Expanding the methodological arsenal of applied linguistics with a robust statistical procedure. *Applied Linguistics, 39*(3), 422-428. https://doi.org/10.1093/applin/amx026

Norouzian, R. (2020). Sample size planning in quantitative l2 research: A pragmatic approach. *Studies in Second Language Acquisition, 42*(4), 849-870. https://doi.org/10.1017/S0272263120000017

Norouzian, R., de Miranda, M., & Plonsky, L. (2018). The bayesian revolution in second language research: An applied approach. *Language Learning, 68*(4), 1032-1075. https://doi.org/10.1111/lang.12310

Norris, J. M. (2015). Statistical significance testing in second language research: Basic problems and suggestions

for reform. *Language Learning, 65*(S1), 97-126. https://doi.org/10.1111/lang.12114

Norris, J. M., Ross, S. J., & Schoonen, R. (2015). Improving second language quantitative research. *Language Learning, 65*(S1), 1-8. https://doi.org/10.1111/lang.12110

Pang, B., Nijkamp, E., & Wu, Y. N. (2019). Deep learning with tensorflow: A review. *Journal of Educational and Behavioral Statistics, 45*(2), 227-248. https://doi.org/10.3102/1076998619872761

Papi, M., & Teimouri, Y. (2014). Language learner motivational types: A cluster analysis study. *Language Learning, 64*(3), 493-525. https://doi.org/10.1111/lang.12065

Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research, 3*(1), 61-94. https://doi.org/10.1075/ijlcr.3.1.03paq

Pearl, L. S., & Enverga, I. (2014). Can you read my mindprint? Automatically identifying mental states from language text using deeper linguistic features. *Interaction Studies, 15*(3), 359-387. https://doi.org/10.1075/is.15.3.01pea

Pliakos, K., Joo, S.-H., Park, J. Y., Cornillie, F., Vens, C., & Van den Noortgate, W. (2019). Integrating machine learning into item response theory for addressing the cold start problem in adaptive learning systems. *Computers & Education, 137*, 91-103. https://doi.org/10.1016/j.compedu.2019.04.009

Plonsky, L., & Ghanbar, H. (2018). Multiple regression in l2 research: A methodological synthesis and guide to interpreting R2 values. *The Modern Language Journal, 102*(4), 713-731. https://doi.org/10.1111/modl.12509

Plonsky, L., & Gonulal, T. (2015). Methodological synthesis in quantitative l2 research: A review of reviews and a case study of exploratory factor analysis. *Language Learning, 65*(S1), 9-36. https://doi.org/10.1111/lang.12111

Plonsky, L., & Oswald, F. L. (2017). Multiple regression as a flexible alternative to anova in l2 research. *Studies in Second Language Acquisition, 39*(3), 579-592. https://doi.org/10.1017/S0272263116000231

Przybyla, P., & Teisseyre, P. (2014). Analysing utterances in polish parliament to predict speaker's background. *Journal of Quantitative Linguistics, 21*(4), 350-376. https://doi.org/10.1080/09296174.2014.944330

Reed, P., & Wu, Y. Q. (2013). Logistic regression for risk factor modelling in stuttering research. *Journal of Fluency Disorders, 38*(2), 88-101. https://doi.org/10.1016/j.jfludis.2012.09.003

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should i trust you?: Explaining the predictions of any classifier. Paper presented at the *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, USA. https://doi.org/10.1145/2939672.2939778

Rico-Juan, J. R., Gallego, A.-J., & Calvo-Zaragoza, J. (2019). Automatic detection of inconsistencies between numerical scores and textual feedback in peer-assessment processes with machine learning. *Computers & Education, 140*, 103609. https://doi.org/10.1016/j.compedu.2019.103609

Rudner, L. (2016). Accuracy of bayes and logistic regression subscale probabilities for educational and certification tests. *Practical Assessment, Research, and Evaluation, 21*. https://doi.org/10.7275/q7zz-d655

Shawar, B. A., & Atwell, E. S. (2005). Using corpora in machine-learning chatbot systems. *International Journal of Corpus Linguistics*, 489-516. https://doi.org/10.1075/ijcl.10.4.06sha

Shin, J., & Gierl, M. J. (2020). More efficient processes for creating automated essay scoring frameworks: A demonstration of two algorithms. *Language Testing*, 0265532220937830. https://doi.org/10.1177/0265532220937830

Sung, Y. T., Lin, W. C., Dyson, S. B., Chang, K. E., & Chen, Y. C. (2015). Leveling l2 texts through readability: Combining multilevel linguistic features with the cefr. *Modern Language Journal, 99*(2), 371-391. https://doi.org/10.1111/modl.12213

Th Gries, S. (2020). On classification trees and random forests in corpus linguistics: Some words of caution and suggestions for improvement. *Corpus Linguistics and Linguistic Theory, 16*(3). https://doi.org/10.1515/cllt-2018-0078

Tomaschek, F., Hendrix, P., & Baayen, R. H. (2018). Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics, 71*, 249-267. https://doi.org/10.1016/j.wocn.2018.09.004

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining

techniques for student exam performance prediction. *Computers & Education, 143*, 103676. https://doi.org/10.1016/j.compedu.2019.103676

van Halteren, H., Baayen, H., Tweedie, F., Haverkort, M., & Neijt, A. (2005). New machine learning methods demonstrate the existence of a human stylome. *Journal of Quantitative Linguistics, 12*(1), 65-77. https://doi.org/10.1080/09296170500055350

Vasquez-Correa, J. C., Orozco-Arroyave, J. R., Bocklet, T., & Noth, E. (2018). Towards an automatic evaluation of the dysarthria level of patients with parkinson's disease. *Journal of Communication Disorders, 76*, 21-36. https://doi.org/10.1016/j.jcomdis.2018.08.002

Wan, M. Y., Fang, A. C., & Huang, C. R. (2019). The discriminativeness of internal syntactic representations in automatic genre classification. *Journal of Quantitative Linguistics*. https://doi.org/10.1080/09296174.2019.1663655

Warschauer, M., Yim, S., Lee, H., & Zheng, B. B. (2019). Recent contributions of data mining to language learning research. *Annual Review of Applied Linguistics, 39*, 93-112. https://doi.org/10.1017/S0267190519000023

Wiechmann, D., & Kerz, E. (2014). Cue reliance in l2 written production. *Language Learning, 64*(2), 343-364. https://doi.org/10.1111/lang.12047

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language, 72*, 37-48. https://doi.org/10.1016/j.jml.2013.12.003

Xiao, W., & Sun, S. (2020). Dynamic lexical features of phd theses across disciplines: A text mining approach. *Journal of Quantitative Linguistics, 27*(2), 114-133. https://doi.org/10.1080/09296174.2018.1531618

Xie, Z. L., Reetzke, R., & Chandrasekaran, B. (2019). Machine learning approaches to analyze speech-evoked neurophysiological responses. *Journal of Speech Language and Hearing Research, 62*(3), 587-601. https://doi.org/10.1044/2018_JSLHR-S-ASTM-18-0244

Yang, F., & Li, F. W. B. (2018). Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. *Computers & Education, 123*, 97-108. https://doi.org/10.1016/j.compedu.2018.04.006

Zhao, X. W., & Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *International Journal of Bilingual Education and Bilingualism, 13*(5), 505-524. https://doi.org/10.1080/13670050.2010.488284

Zheng, Y., Cheon, H., & Katz, C. M. (2020). Using machine learning methods to develop a short tree-based adaptive classification test: Case study with a high-dimensional item pool and imbalanced data. *Applied Psychological Measurement, 44*(7-8), 499-514. https://doi.org/10.1177/0146621620931198

Zopluoglu, C. (2019). Detecting examinees with item preknowledge in large-scale testing using extreme gradient boosting (XGBoost). *Educational and Psychological Measurement, 79*(5), 931-961. https://doi.org/10.1177/0013164419839439