

Design of Intelligent English Writing Self-evaluation Auxiliary System

Man Liu

School of Foreign Languages, Changchun Institute of Technology, Changchun, Jilin 130012, China

E-mail: manliu_lm@aliyun.com

Keywords: English writing, self-evaluation, auxiliary system, writing teaching

Received: May 23, 2019

Since the reform and opening up, the exchanges between China and the world have become more and more frequent. English, as a widely used international language, plays an important role in international exchanges. English teaching includes five aspects, listening, speaking, reading, writing and translation. Writing teaching is very important but difficult. In order to improve students' autonomous writing ability, this paper briefly introduced the real-time multi-writing teaching mode and designed an automatic scoring algorithm of writing self-evaluation auxiliary system, random sampling based Bayesian classification and combinational algorithm. One thousand CET-4 and CET-6 compositions from Chinese Learner English Corpus (CLEC) were evaluated, and the scoring effect of Bayesian classification algorithm was also tested. The results showed that the accuracy rate, recall rate and F value of the proposed algorithm was better than that of Bayesian classification algorithm under 150 feature extraction dimensions, the two algorithms had improved scoring effect under the optimal feature extraction dimensions, and the improvement amplitude of the algorithm proposed in this study was larger. In summary, the random sampling based Bayesian classification and combinational algorithm is effective and feasible as an automatic scoring algorithm of writing self-evaluation auxiliary system.

Povzetek: Za boljše učenje angleščine je za kitajske študente razvita vrsta pripomočkov v obliki informacijskih storitev.

1 Introduction

As the economic globalization deepens, the communication between China and other countries is more and more frequent, and the most frequently used language is English. English teaching includes five aspects, listening, speaking, reading, writing and translation, among which writing teaching is the most important and difficult part [1]. English writing ability can be improved through a large number of writing exercises and comments of teachers. However, the ratio of the number of teachers to the number of students is very small in China. Teachers can not provide guidance to all students. If students lack guidance about writing exercises, the improvement effect will be greatly reduced. Because of the rapid development of computer and natural language research, a computer intelligence based self-correction system has been developed [2]. Despite the fact that the writing language is changeable and the effect is not ideal in practice, it is still of great help to lighten the burden on teachers and improve students' English writing level. Deng [3] put forward a design method of cloud service platform based intelligent English writing auxiliary teaching system, constructed the overall design model of English writing teaching system, improved the intelligence level of the English writing teaching system using Cloud-P2P fusion model, and found that the English writing teaching system had favourable adaptability, learning ability and reliability. Li [4] selected two parallel classes as the object, one was taught by the traditional business English writing teaching mode and the other was guided by computer assisted technology. The teaching

quality was evaluated after one academic term, and it was found that computer assisted technology had positive effect on business English writing. Tsai [5] applied the blackboard course management system in English essay writing teaching and found that the teaching result of the experimental group was superior to that of the control group after two academic years. The questionnaire result suggested that most of the students had positive learning result, indicating the teaching mode could improve the effectiveness of English writing learning. This paper briefly introduced the real-time multi-writing teaching mode and designed the automatic scoring algorithm of the writing self-evaluation assistant system. One thousand CET-4 and CET-6 compositions from Chinese Learner English Corpus (CLEC) were scored. The scoring effect of the Bayesian classification algorithm was also tested and compared with the automatic scoring algorithm.

2 English writing teaching mode

The traditional teaching mode in China is mostly “duck-stuffing”, which is similar to assembly line. Most colleges and universities regard English writing teaching as a part of English teaching, or as a subsidiary part, and moreover it is not paid much attentions to because of time and energy waste. The traditional teaching of English writing is usually conducted in the classroom, but teachers usually do not pay attention to whether students understand or not and only provide students with template sentences and simple explanation [6].

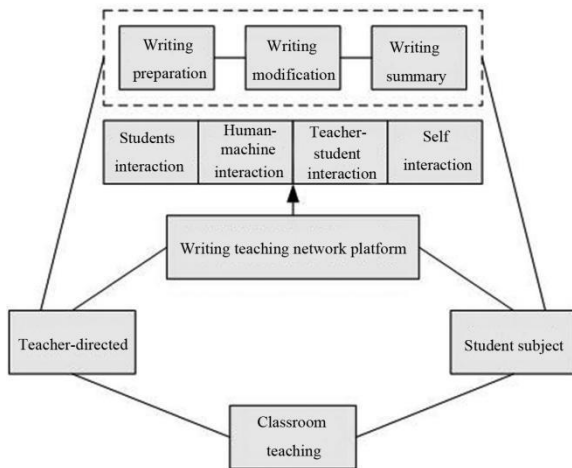


Figure 1: The flow of real-time multi-writing teaching.

Guided by the theory of constructivism [7], a new teaching mode, real-time multi-writing teaching mode [8] has been proposed, and its flow chart is shown in Figure 1. The whole writing process is divided into 3 parts, writing preparation, writing modification, and writing summarization. Teachers and students participate in the whole process. Teachers take students as the center to teach in the classroom and on the network platform. Classroom teaching follows the principle of student-oriented to implement the traditional teaching mode. On the network platform, student interaction, human-computer interaction, teacher-student interaction and self-interaction can be achieved because of the convenient Internet. Besides the Internet, the achievement of the above interaction also relies on the writing self-evaluation system. Through the objective evaluation of computer and based on the evaluation of students and teachers, comments and review comments are obtained.

3 Writing self-evaluation system

3.1 General structure of the system

The general structure of the system is shown in Figure 2. The system in this paper is a Web system based on B/S mode [9]. The client used by user runs in the browser, while the business function of the system runs on the server. The overall structure of the system is divided into 3 parts, user interface layer, business logic layer and data layer. The user interface is the web browser; the business logic layer contains all the functions of the system, and the data layer contains the data needed to run the system. Automatic scoring is the main function of the system.

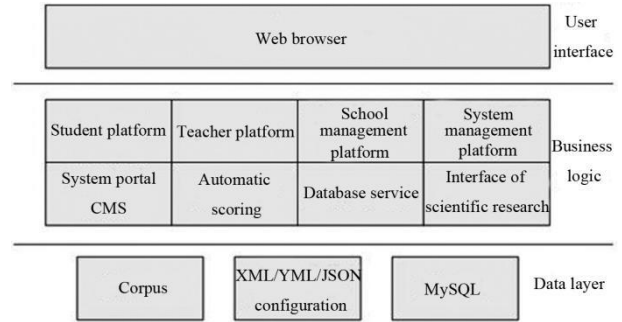


Figure 2: The overall structure of the system.

3.2 Automatic scoring

3.2.1 Feature extraction

The features of compositions with different scores needed to be extracted through training corpus before use to facilitate the classification of compositions to be tested [10]. In this study, the information gain method was used to extract the features of compositions. The expression of the information gain method [11] is:

$$IG(t) = -\sum_{i=1}^k P(a_i) \log(P(a_i)) + P(t) \sum_{i=1}^k P(a_i|t) \log P(a_i|t) + P(\bar{t}) \sum_{i=1}^k P(a_i|\bar{t}) \log P(a_i|\bar{t}) \quad (1)$$

where t is the feature of adjacent binary phrase, a_i is the set of compositions with the i -th score, \bar{t} stands for the condition in case of absence of feature t , $P(a_i)$ stands for the possibility of score a_i in the training corpus, $P(t)$ stands for the possibility of composition containing feature t in the training corpus, $P(a_i|t)$ stands for the possibility of composition containing feature t and with score a_i , $P(\bar{t})$ stands for the possibility of composition not containing feature t , $P(a_i|\bar{t})$ stands for possibility of composition which is scored as a_i but not contains feature t , k stands for the number of score grade, 4 here (grade 1: 1 ~ 5 points; grade 2: 6 ~ 9 points; grade 3: 10 ~ 13 points; grade 4: 13 ~ 14 points).

3.2.2 Random sampling and Bayesian classification based composition scoring algorithms

The flow of the scoring algorithm [12] is shown in Figure 3. $Y = \arg \max_{1 \leq i \leq m} I(\alpha = y_i)$ (2), where α refers to different score grades, y_i refers to the classification result of a random sampling, $I(\bullet)$ refers to indicator function, 1 if the parameter is true and 0 if not.

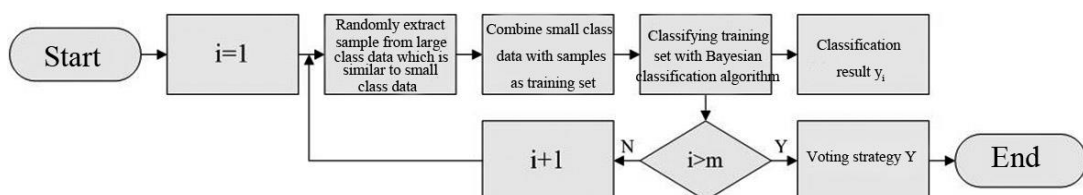


Figure 3: The flow of random sampling and Bayesian classification based composition scoring algorithms.

The calculation formula of Bayesian classification based algorithm [13] is:

$$P(b|a_i) = \prod_{k=1}^V \frac{P(\theta_k|a_i)^{N_k}}{N_k!} \quad (3)$$

$$P(\theta_j|a_i) = \frac{N_{ji} + 1}{N_{a_i} + |V|} \quad (4)$$

where b refers to the spatial vector of an English composition, $b = \{\theta_1, \theta_2, \dots, \theta_i, \dots, \theta_m\}$, m refers to the number of features of adjacent binary phrase, θ_i refers to the weight of the i -th feature in composition b , V refers to the number of binary phase features, N_k refers to the number of times of the k -th feature appearing in b , $P(b|a_i)$ refers to the possibility of a composition obtaining some score, $P(\theta_j|a_i)$ refers to the possibility of feature θ_j in a composition which is scored as a_i , N_{ji} refers to the number of times of the j -th feature appearing in a composition which is scored as the j -th feature, and N_{a_i} refers to the number of all the features of a composition which is scored as a_i .

4 Automatic scoring test of the writing self-evaluation system

4.1 Testing method

Five hundred of CET-4 compositions which involved two themes and 500 CET-6 compositions with which involved two themes were selected from CLEC [14]. Before the test, the compositions were scored and classified into four score grades according to the scoring criteria of CET-4 and CET-6 compositions. Based on the binary phrase features, the compositions which involved four themes were scored under 150 feature extraction dimensions, i.e., the 1000 compositions were classified into four score grades using the algorithm proposed in this study. Then the compositions were scored after the automatic

calculation of optimal feature dimensions. The whole system program ran on a server in a lab. The server was configured with Windows 7 system, I7 processor and 16G memory. In order to increase the persuasiveness of the results, the test results of the Bayesian classification algorithm based scoring system was selected for comparison. The test method was the same as the system proposed in this study.

4.2 Evaluation criteria

The automatic scoring effect of the writing self-evaluation system was evaluated in the aspects of accuracy rate, recall rate and F value [15]. The accuracy rate could be calculated using the following formula: the accuracy rate = correctly recognized number/total recognized number, where correctly recognized number is the number of correctly classified compositions based on above scoring algorithm and the total recognized number is the total number of compositions identified by the scoring algorithm.

The recall rate could be calculated using the following formula: recall rate = correctly recognized number/actually existed number, where actually existed number refers to the number of compositions which actually existed and ought to be recognized.

F value (comprehensive evaluation index) could be calculated using the follow formula: F value = $2 \times$ accuracy rate \times recall rate / (accuracy rate + recall rate).

4.3 Testing results

As shown in Table 1, the accuracy rate, recall rate and F value of the scoring algorithm were about 14%, 28% and 22% higher than those of Bayesian classification algorithm in scoring different themes of compositions. The accuracy rate, recall rate and F value of the Bayesian classification algorithm were 0.739, 0.661 and 0.690 respectively; the accuracy rate, recall rate and F value of the algorithm proposed in this study were 0.845, 0.850 and 0.844 respectively. It indicated that the algorithm proposed in this study was superior to the Bayesian classification algorithm in scoring compositions, and all

Theme	Bayesian classification algorithm			Random sampling based Bayesian classification algorithm		
	Accuracy rate (P)	Recall rate (R)	F value	Accuracy rate (P)	Recall rate (R)	F value
Theme 1 (CET-4)	0.719	0.626	0.660	0.849	0.881	0.862
Theme 2 (CET-4)	0.750	0.679	0.708	0.838	0.833	0.832
Theme 3 (CET-6)	0.759	0.710	0.730	0.836	0.831	0.831
Theme 4 (CET 6)	0.726	0.630	0.661	0.857	0.853	0.851
Average value	0.739	0.661	0.690	0.845	0.850	0.844

Table 1: The scoring test results of the two algorithms under 150 feature extraction dimensions.

Theme	Optimal feature dimension number	Bayesian classification algorithm			Random sampling based Bayesian classification algorithm		
		Accuracy rate (P)	Recall rate (R)	F value	Accuracy rate (P)	Recall rate (R)	F value
Theme 1 (CET-4)	390	0.856	0.846	0.838	0.990	0.989	0.989
Theme 2 (CET-4)	350	0.784	0.778	0.763	0.989	0.988	0.988
Theme 3 (CET 6)	710	0.878	0.866	0.857	1	0.999	0.999
Theme 4 (CET 6)	530	0.844	0.795	0.739	0.987	0.986	0.986
Average value		0.841	0.821	0.799	0.992	0.991	0.991

Table 2. The scoring test results of the two algorithms under optimal feature extraction dimensions.

the indicators were above 80%. Based on the binary phrase features extracted from the compositions, the algorithm could accurately classify the tested composition into the corresponding score grade and make accurate and reasonable evaluation on the content of compositions based on the score grade.

As shown in Table 2, the optimal feature extraction dimension of theme 1, 2, 3 and 4 compositions was 390, 350, 710 and 530 respectively. Considering the evaluation index values of the four themes, the accuracy rate, recall rate and F value of Bayesian classification algorithm were 0.841, 0.821 and 0.799 respectively under the optimal feature extraction dimension, and the accuracy rate, recall rate and F value of the algorithm proposed in this study were 0.992, 0.991 and 0.991 respectively.

As shown in Figure 4, the accuracy, recall rate and F value of the algorithm under the optimal feature extraction dimensions were better than those under 150 feature extraction dimensions. Moreover the three indicators of the algorithm proposed in this study were always the best, and the improvement of the algorithm was greater after changing feature extraction dimensions. That is to say, after the application of optimal feature extraction dimension, the algorithm proposed in this study could classify the compositions more accurately according to the binary phrase features, and make more accurate and reasonable evaluation according to the score grade.

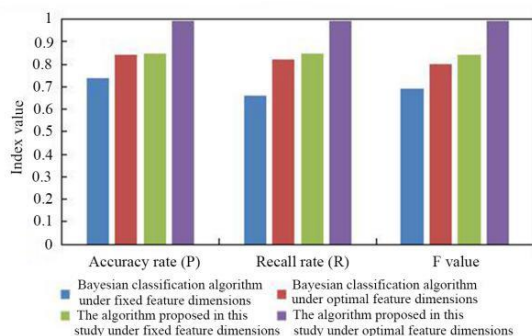


Figure 4: The comparison of testing results of the two algorithms under the fixed and optimal feature extraction dimensions.

5 Conclusion

This paper briefly introduced the real-time multi-writing teaching model and designed an automatic scoring algorithm for the writing self-evaluation system. Then 1000 CET-4 and CET-6 compositions in CLEC were scored according to four grades. As a comparison, the scoring effect of the Bayesian classification algorithm was tested. The results showed that the accuracy rate, recall rate and F value of the Bayesian classification algorithm were 0.739, 0.661 and 0.690 respectively, and the corresponding data of the algorithm proposed in this study were 0.845, 0.850 and 0.844 respectively, indicating that the scoring effect of the algorithm proposed in this study was superior to that of the Bayesian classification algorithm. Under the optimal feature extraction dimensions, the accuracy rate, recall rate and F value of the Bayesian classification algorithm were 0.841, 0.821 and 0.799 respectively and those of the algorithm proposed in this study were 0.992, 0.991 and 0.991 respectively, which were improved compared to under the fixed feature extraction dimensions. Moreover the improvement amplitude of the algorithm proposed in this study was larger.

Acknowledgement

This study was supported by General Project of Higher Education Department of Education Department of Jilin Province in 2018, Research on English “Four-in-One” Teaching Reform in College English under the Background of Internet+, Key Project of Teaching Reform of Changchun University of Engineering, June 2018 to June 2020, under researching.

References

[1] Nattapong J, Rattanavich S (2015). The effects of computer-assisted instruction based on top-level structure method in English reading and writing abilities of Thai EFL students. *English Language*

- Teaching*, 8(11), pp. 231. <https://doi.org/10.5539/elt.v8n11p231>
- [2] Xia M, Zhang Y, Zhang C (2017). A TAM-based approach to explore the effect of online experience on destination image: A smartphone user's perspective. *Journal of Destination Marketing & Management*. <https://doi.org/10.1016/j.jdmm.2017.05.002>
- [3] Deng L (2018). Design of English writing assisted instruction teaching system based on intelligent cloud service platform. *International Conference on Intelligent Transportation, Big Data & Smart City*. IEEE Computer Society, pp. 287-290. <https://doi.org/10.1109/ICITBS.2018.00080>
- [4] Li X (2018) Influence of computer-aided instruction model on business English writing teaching effect. *International Journal of Emerging Technologies in Learning*, 13(3), pp. 197.
- [5] Tsai Y R (2015). Applying the technology acceptance model (TAM) to explore the effects of a course management system (CMS)-assisted EFL writing instruction. *Calico Journal*, 32(1). <https://doi.org/10.1558/cj.v32i1.153-171>
- [6] Tarhini A, Hassouna M, Abbasi M S, et al. (2015). Towards the acceptance of RSS to support learning: an empirical study to validate the technology acceptance model in Lebanon. *Electronic Journal of e-Learning*, 13(1), pp. 30-41.
- [7] Li L (2016). Design and implementation of higher vocational English writing and training testing system. *International Conference on Materials Engineering, Manufacturing Technology and Control*.
- [8] Chapela M E G (2016). A corpus-based analysis of post-auxiliary ellipsis in modern English: methodological and theoretical issues.
- [9] Ohta R (2017). The impact of an automated evaluation system on student-writing performance. *Kate Bulletin*, 22, pp. 23-33.
- [10] Ahsen M E, Ayvaci M, Raghunathan S (2017). When algorithmic predictions use human-generated data: a bias-aware classification algorithm for breast cancer diagnosis. *Social Science Electronic Publishing*. <https://doi.org/10.2139/ssrn.3087467>
- [11] Alcázar V, Fernández S, Borrajo D, et al. (2015). Using random sampling trees for automated planning. *Ai Communications*, 28(4), pp. 665-681.
- [12] Myttenaere A D, Golden B, Grand B L, et al. (2015). Study of a bias in the offline evaluation of a recommendation algorithm. *Computer Science*, 79(3), pp. 263-72. [https://doi.org/10.1016/0002-9416\(81\)90074-9](https://doi.org/10.1016/0002-9416(81)90074-9)
- [13] Hao J, Niu K, Meng Z, et al. (2017). A collaborative filtering recommendation algorithm based on score classification. *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*. Springer, Cham, pp. 435-445. https://doi.org/10.1007/978-3-319-72395-2_40
- [14] Santos D P D, André C P L F D C (2017). Automatic selection of learning bias for active sampling. *Intelligent Systems*. IEEE, Recife, Brazil, pp. 55-60. <https://doi.org/10.1109/BRACIS.2016.021>
- [15] Zoph B, Yuret D, May J, et al. (2016). Transfer learning for low-resource neural machine translation. *Conference on Empirical Methods in Natural Language Processing*, pp. 1568-1575.

