# The Study on Automated Evaluation of Online Discussion Quality Based on Semantics

Wen-Yan Qin[1], Shuang Qi[2], Shu-Yue Zhou[1], and Chao-Jun Xu[1]
[1] School of Education Science, Nanjing Normal University, Nanjing, Jiangsu, China
[2] Affiliated to Hangzhou No.14 High School, Hangzhou, Zhejiang, China
Email: 267440207@qq.com

*Abstract*—**High-quality discussions are of great significance to promote learning and achieve teaching goals. In this study, content analysis is used to perform semantic analysis on the content of the discussion posted by learners on online learning platform from the perspective of semantic similarity. The semantic similarity of different topics under the course is weighted to obtain the online discussion quality score of the learner. Then the evaluation results from experts discussed with the learner are divided into four levels. The Kappa consistency test and the McNemar paired chi-square test are performed. The experimental results show that there is consistency between the two evaluation methods, and the accuracy rates of the two ranked lower rankings are better. Therefore, the automated evaluation of the online discussion quality of learners based on semantics will play an indispensable role in the selection and sticking of comments on online learning platforms.**

*Index Terms*—**automated evaluation, discussion quality, semantic analysis, semantic similarity**

## I. Introduction

Online discussion is an interactive link of online learning [1]. Learners complete the discussion process by posting and replying in discussion area, forum, chat room, etc, of online teaching platform. Text is the main form of discussion. The online learning platform has inherent advantage of recording interactive content. The course discussion area can record learners interactive behaviors during the learning process and retain all learners interactive content. It has become the focus of attention of scholars that how to excavate valuable information from the discussion content, and use it as a data source to automatically evaluate the quality of discussion of learners, and make it a part of course evaluation [2]-[4].

Up to now, there have been many modes of online discussion quality analysis [5]-[7]. And almost every mode pays attention to participation, interaction mode and interaction level. The automated evaluation of the quality of online discussion by learners enables learners to clearly understand their own learning situation without help of teachers and researchers, and allows teachers to easily and conveniently understand learners' mastery of course content, thereby improving teaching. It also helps teachers discover the personality differences of learner interaction.

Automated evaluation of text quality has been studied in the field of automated essay scoring [8]-[12]. Although there are guidelines for writing and evaluating articles, this is not the case for forum posts, because different users may consider different quality standards when scoring [13]. Kim's group found that the relationship between a student's posting behavior and student's score can be automatically assessed. The main indicators they used are numbers of student posts, average post length, and average number of replies to the post [14]; Feng's group and Kim's group described a system for finding the most authoritative answers in forum threads, using speech act analysis and author's credibility as classification features, and calculating text quality according to the authors automated quality classification scheme [15]. These studies are evaluation methods based on basic statistics, using objective quantitative data related to posts in discussion area as basic parameters. The scope of evaluation objects is relatively limited. The number of interactions is a necessary condition for interaction quality but not a sufficient condition [16].

With development of natural language processing and big data technology, some studies have used some classification techniques to automatically evaluate the quality of forum posts [17]. Markus Weimer et al. proposed an algorithm for automatically evaluating the quality of forum posts. Using a classification technique, experiments were conducted on five feature classes: surface features, vocabulary features, syntactic features, forum features, and similarity features, to achieve online automated evaluation of discussion quality [18]. In addition, one of the present authors Chao-Jun Xu and Xi-Xiang You studied the question-and-answer pairs in Q&A community. They used semantic analysis technology to extract keywords for answers, and the LDA topic model to conduct topic mining on answers. With distribution of topics and judgement of topic words, they get the subject content in answer [19]. Jia-Ying Chen et al. proposed a recommendation algorithm fused with semantic analysis feature extraction. Through the knowledge graph entity recognition link technology, the item attribute feature entity and its associated entities are extracted in knowledge base according to the text information content. And one uses it to analyze the fine-

Manuscript received February 15, 2022; revised May 3, 2022.



doi: 10.18178/ijlt.8.2.80-85

grained features of users and items, and the vector representation of users and items based on features. Thus, the recommendation to target users can be completed [20].

From the perspective of topic-based semantics, this work focuses on the semantics of post content. We use content analysis to analyze learner's post content and course subject content text. Then we dig out the topic of the discussion content and the learner's views implicit in it.

By calculating the semantic similarity between two texts, we evaluate the quality of learners' discussion content, and perform Kappa consistency test and McNemar paired chi-square test with expert's evaluation results to verify effectiveness of semantic-based automated evaluation method for the quality of online learners' discussions. It is found that the final discussion quality score obtained by learner according to the weighted average of text semantic similarity is consistent with expert evaluation result. It shows that two evaluation methods have no obvious evaluation tendency, which proves the effectiveness of semantic-based automated evaluation of learner online discussion quality, so the research results of this paper have certain credibility and reference value.

## II. Method

### A. Feature Words Selection and Representation of Online Discussion

In order to select feature items that actually contribute to the calculation of text similarity, we select a subset from the original feature set to reflect the theme content of original text according to some criteria. The commonly used feature item selection methods mainly include document frequency method based on feature item frequency statistics, mutual information, information gain, etc. [21]-[24]. The object of feature selection in this work is the discussion content and the subject content of course. The discussion content of learners is all around the same topic under the same course. There must be a certain amount of same or similar vocabulary in discussion content of different learners. The course content of different chapters belong to the same course, and there must be the same or similar vocabulary.

Therefore, this paper discards the traditional feature selection algorithm's attention to the relationship between feature words and document set, and focuses on the interior of a single document. Combining part of speech, word length, and word frequency, the filter conditions are set to select characteristic words with greater contribution to text.

### B. Calculation of Text Semantic Similarity

In the real world, every word has multiple semantics. The different semantics of the word correspond to different "concepts". Therefore, the semantic relevance between different words can be reflected by the semantic relevance between "sememe". In order to achieve this, we adopts the similarity algorithm based on "HowNet". Unlike "TongYi Cilin", "HowNet" builds a semantic network based on world knowledge. Instead of using a tree structure, it uses a network structure to describe knowledge.

We assume that the higher the similarity between the discussion text of a learner under a certain course teaching topic and course topic content text is, the higher the quality of learner's discussion under the topic is [25]. Therefore, in this experiment, the online discussion quality score is transformed into a comprehensive score of semantic similarity between learner's course content and discussion content under a certain course. The comprehensive score is obtained by the weighted summation of semantic similarity of text under each topic. Assuming that a course has topics in total, the overall text similarity of a learner A in course is given by

$$Sim_A = \sum_{k=1}^{e} Sim_{Ak}\, \gamma_k, \tag{1}$$

where $Sim_{Ak}$ is the text similarity of learner A under $k$-th topic, and $\gamma_k$ is the weight of $k$-th topic under the course which satisfies the following condition,

$$\sum_{k=1}^{e} \gamma_{\mathrm{k}} = 1. \tag{2}$$

By default, we think that the weight of each topic is the same, that is, value of $\gamma_k$ is obtained by average weighting. If the importance of each topic to the course is not the same, teacher can assign corresponding weight according to the importance of topic.

Next, we calculate the text similarity $Sim_{Ak}$ of learner A under the $k$-th topic of the course. The discussion text of learner A under the topic is $d_x$ which has $m$ non-repeated feature words $(t_1, t_2, \ldots, t_m)$. The course knowledge point text corresponding to this topic is $d_y$ which has n non-repeated feature words $(t_1, t_2, \ldots, t_n)$. Then the similarity of $d_x$ and $d_y$ is given by

$$Sim_{Ak} = Sim_{(x,y)} = \frac{\sum_{i=1}^{m} Sim(t_i)_{max} \times w_{xi} \times w_{yj}}{m}. \tag{3}$$

We obtain one similarity value for each feature word in the text $d_y$, and the maximum value in similarity set is taken as maximum similarity $Sim(t_i)_{max}$ of feature words. Its function form is given by

$$Sim(t_i)_{max} = \max_{i=1\ldots m}\left(Sim(t_i, t_1), \ldots, Sim(t_i, t_n)\right). \tag{4}$$

The specific method to calculate maximum similarity is as shown in Fig. 1.

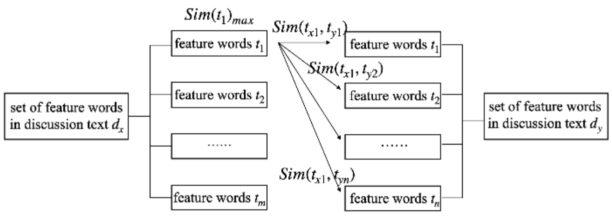


Figure 1. Selection strategy for maximum similarity of feature words.

And the maximum similarity $Sim(t_1)_{max}$ of feature word $t_1$ is the maximum value in $Sim(t_{x1}, t_{y1})$, $Sim(t_{x1}, t_{y2})$,……, $Sim(t_{x1}, t_{yn})$.

To be noted, the calculation for maximum similarity of feature words in this experiment does not adopt the method of maximum weight matching, but allows the feature words in the course content to "multiple use of a word". This method takes into account the feature of course content. For courses, knowledge points are specific and limited. If three feature words in discussion text of a learner are most similar to a fixed core feature word, and this core feature word can only be used once, this approach is obviously unreasonable.

In (3), $w_{xi}$ is the weight of feature word $t_i$ in text $d_x$. Its function form is given by

$$w_{xi} = \frac{f_i}{\sum_{i=1}^{m} f_i}, \tag{5}$$

where $f_i$ is the frequency of feature word $t_i$ in text $d_x$. $w_{yj}$ is the weight of feature word $t_j$ in text $d_y$. $t_j$ is the feature word in $d_y$ corresponding to feature word $t_i$ when the maximum similarity of $t_i$ is taken. The function form of $w_{yj}$ is given by

$$w_{yj} = \frac{f_j}{\sum_{i=1}^{n} j_i}. \tag{6}$$

The idea of this algorithm is as following. First, we calculate the similarity between each feature word in learner's discussion text and all feature words in course content text, and we take the maximum value. At the same time, we multiply the maximum value by the weight of two corresponding feature words in two texts when the maximum value is taken. The product of three is regarded as contribution of feature words in discussion text. The contribution of each feature word in the discussion text is calculated in turn, and finally the average contribution of feature word in discussion text is calculated. Using this result, we discuss the text similarity between discussion content and course content.

When the maximum similarity of feature word $Sim(t_i)_{max}$ is multiplied by the weight in discussion text $w_{xj}$ and the weight in course text $w_{yj}$ in (3), the importance of corresponding feature word in discussion text and that in course text are taken into consideration, respectively. For instance, if a certain feature word is very similar to the core knowledge feature word, but the learner has a lot of discussion feature words, and this certain feature word only appears once, we think the importance of this feature word to the discussion text is not significant. In other words, in the feature word set of knowledge points, the importance of each feature word is distinguished. The higher the frequency of feature words, the higher their importance. And this importance is positively related to the contribution of corresponding feature words in discussion text.

According to (3), the ranges of similarity of feature word $Sim(t_i)_{max}$, weight of feature words in discussion text $w_{xi}$, and weight of feature words in course text $w_{yj}$, are between 0 and 1.

Therefore, the contribution of feature words must also be between 0 and 1. The learner's text similarity in a certain chapter of course $Sim_{(x,y)}$ is the average of contribution of all feature words. The overall text similarity of learner under course $Sim_{(x,y)}$ is the weighted average of text similarity of all chapters, so it must be a number between 0 and 1.

And we think the closer to 1 final text similarity is, the higher quality of learner's discussion under the course is.

In addition, it should be noted that the quality of discussion among learners in different courses is not comparable. Although the semantic similarity algorithm is the same in theory, due to the large differences in content of different courses, the reference system of semantic similarity of learners under different courses is also different. Therefore, the comparison of the quality of discussion between different learners should be based on premise of the same course [26].

### C. Experiment Design

This experiment selected the relevant data of "Modern Educational Technology" course offered by associate professor Chao-Jun Xu from School of Educational Science in Nanjing Normal University on Note100 online education platform (http://www.note100.cn) in the first semester of 2017~2018. It includes text data of discussion content of 47 students in teaching class under 4 topics and text data of course content of chapters. Totally, we download 1371 discussions, including 203 main posts and 1168 reply posts. After removing invalid data, the total number of remaining discussion data is 1196, including 194 main posts and 1002 reply posts.

This experiment is divided into four parts. First, we obtain the discussion content and course content data of learners from the online education website; through the above text data collection process, course content text data has been exported.

Then we preprocess the original text, including word segmentation and removal of stop words. The Chinese word segmenter in this study is Ansj word segmenter which is a high-precision Chinese word segmentation tool implemented in the Java language. It can be used for personal name recognition, place name recognition, organizational structure name recognition, multi-level part-of-speech tagging, keyword extraction, fingerprint extraction and other fields. Ansj has four-word segmentation modes, namely basic word segmentation, precise word segmentation, NLP word segmentation, and index-oriented word segmentation. We chose the precise word segmentation mode. After word segmentation, the text has been represented as a set of words. In a word-based retrieval system, stop words refer to words that appear too frequently and have no substantial meaning, or words whose frequency is too low to represent the topic of text. However, there are still some words in these words that are not helpful for expressing topic. In order to eliminate the influence of these words on the theme of expression, we process word sets of two texts to remove stop words. There are many stop word lists, but there is no widely recognized stop word list. By comparing with multiple stop word lists, we choose a relatively comprehensive stop word list to remove stop words from the word sets of two texts.

Then, according to terms such as part of speech, word frequency, and word length, we filter out the characteristic words that can represent the topic of discussion content text and the course content text respectively.

Finally, we calculate the similarity between two text feature word sets, and use the similarity calculation result as basis for the evaluation of learner's discussion quality, so as to realize automatic evaluation of learner's online discussion quality.

## III. Results and Discussion

We calculate the feature word set of course content text and learner's discussion content text according to the algorithm of text semantic similarity, and obtain the scores of 47 students in class under 4 topics in “Modern Educational Technology” course. The 4 scores are weighted and averaged to get final discussion quality score. The specific scores are shown in Table I.

TABLE I. Discuss Quality Experiment Score

| Serial Number | Experiment Score | Serial number | Experiment score |
|---|---|---|---|
| 1 | 93.83 | 25 | 66.79 |
| 2 | 65.67 | 26 | 77.24 |
| 3 | 64.97 | 27 | 33.63 |
| 4 | 72.01 | 28 | 74.36 |
| 5 | 82.63 | 29 | 72.93 |
| 6 | 69.37 | 30 | 68.32 |
| 7 | 67.84 | 31 | 69.70 |
| 8 | 69.78 | 32 | 69.86 |
| 9 | 62.96 | 33 | 62.97 |
| 10 | 56.83 | 34 | 79.76 |
| 11 | 80.21 | 35 | 71.79 |
| 12 | 72.39 | 36 | 71.03 |
| 13 | 69.96 | 37 | 68.99 |
| 14 | 74.89 | 38 | 65.95 |
| 15 | 67.10 | 39 | 75.22 |
| 16 | 78.61 | 40 | 75.98 |
| 17 | 71.14 | 41 | 70.58 |
| 18 | 77.65 | 42 | 67.25 |
| 19 | 81.72 | 43 | 69.90 |
| 20 | 58.62 | 44 | 80.80 |
| 21 | 67.84 | 45 | 69.42 |
| 22 | 65.22 | 46 | 74.65 |
| 23 | 70.77 | 47 | 37.20 |
| 24 | 69.30 | | |

In order to verify the validity of students’ discussion quality scores in experiment, teacher scored the discussion of 47 students in class on 4 topics. Then we averagely weight the scores under 4 topics to get expert scores of students’ discussion quality. The experimental scores and expert scores for 47 students are divided into four levels, A, B, C, and D, corresponding to $1^{st}$-$5^{th}$, $6^{th}$-$15^{th}$, $16^{th}$-$35^{th}$ and $36^{th}$-$47^{th}$, respectively, as shown in Table II.

TABLE II. Rating Level and Ranking Corresponding Information

| Rating level | Ranking range |
|---|---|
| A | 1-5 |
| B | 6-15 |
| C | 16-35 |
| D | 36-47 |

In order to check the consistency between discussion quality score calculated according to the semantic similarity of the text and the expert evaluation results, we conduct a Kappa consistency test on graded experimental scoring results and expert scoring results in SPSS. The results are shown in Table III.

TABLE III. Kappa Consistency Test Results

| | Value | Asymptotic standardized error | Approximate $T$ | Approximate significance |
|---|---|---|---|---|
| Measure of agreement Kappa | .420 | .101 | 4.662 | .000 |
| N of valid cases | 47 | | | |

TABLE IV. McNemar paired chi-square test results

| | Value | df | Asymptotic Sig. (2-sided) |
|---|---|---|---|
| McNemar-Bowker test | | 4 | .973 |
| N of valid cases | 47 | | |

TABLE V. RATING LEVEL AND ACCURACY RATE INFORMATION

| Rating level | Accuracy rate (%) |
|---|---|
| A | 40.0 |
| B | 58.0 |
| C | 75.0 |
| D | 87.0 |

As shown in Table III, the Kappa value is 0.420 and P value is 0.000. Since P value is less than 0.05, we reject H0 hypothesis and accept H1 hypothesis, and consider experimental scoring results are consistent with expert scoring results. According to results of Kappa consistency test, we conjecture that there is consistency between experimental scoring results and expert scoring results.

Then we verify whether these two evaluation methods have obvious evaluation tendency, and perform McNemar paired chi-square test on experimental scoring results and expert scoring results. The results are shown in Table IV.

The null hypothesis of McNemar paired chi-square test is as the following: there is no difference between the results of two evaluation methods, and P value is 0.973, which is obviously greater than 0.05. Therefore, the null hypothesis is accepted, that is, there is no difference between experimental scoring results and expert scoring results, that is, there is no certain This evaluation method has an obvious tendency of high or low evaluation.

In the experimental verification results, this teaching class has 47 students. 29 of them have accurate classification of grades. The accuracy rate is about 64.7%. The accuracy rates of each grade are shown in Table V.

As shown in Table V, using the research method of this experiment, the accuracy of evaluation of learners in four levels is quite different. For "A" and "B" two levels, the accuracy rate is less than 60%, and the accuracy is poor. The accuracy rate for two levels of "C" and "D" are both above 75%, and the accuracy is good. The reason for this phenomenon may be that the length of the discussion texts of the students with grades "C" and "D" at bottom of the ranking is different, and the difference is large, because they are not familiar with the course content. Thus, there are less key words related to the course content in their discussion texts. Therefore, the distinction of their discussion text are large in calculation. In the case of students in top "A" and "B" grades, it is more difficult to distinguish.

## IV. SUMMARY

We have found that the final discussion quality score calculated according to text semantic similarity is consistent with expert evaluation results. Two evaluation methods have no obvious evaluation tendency. In addition, the evaluation accuracy for lower two levels is better. Therefore, the semantic-based method for evaluating the quality of learner online discussion has certain credibility and reference value.

In this work, we have calculated the semantic similarity between course content and its corresponding discussion content from perspective of text semantic similarity. The result is used as a basis for evaluating the quality of learners' online discussion. We have also proposed an online discussion quality evaluation method based on semantic similarity to quantify discussion quality which is difficult to measure abstractly. By setting up specific processing strategies according to research needs, objective results can be automatically calculated. The semantic-based automated evaluation of learner online discussion quality monitors learners' learning dynamics to a certain extent. Teachers can make teaching adjustments in time to adapt to learners' personalized learning, and it also provides an effective way for learners' online intelligent question and answer. Online discussion is just a situation in context of “big data and education”. If there is an intervention strategy for discussion incentives and intelligent evaluation is used to improve scientificity and timeliness of classroom teaching evaluation, the quality of online discussion will be greatly improved.

Note100 (http://www.note100.cn) online education platform includes online discussion information for the entire subject, but this research is limited by time and other resources. Only the discussion text data under the relevant topics of this course is selected to verify the effectiveness of the automated evaluation of online discussion quality. The sample data is small and coverage is not broad enough. Because of the unstructured data generated in the process of teaching and learning, machines cannot accurately analyze and process educational big data. In the future, we will collect more data, including but not limited to this subject, and further verify whether the results of automated evaluation experiments of different subjects are similar to the results of expert evaluation.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Chao-Jun Xu conducted the research; Wen-Yan Qin and Shuang Qi preformed the calculation and analyzation of the data; Shu-Yue Zhou collected the data; Wen-Yan Qin wrote the paper. All authors had approved the final version of this paper.

## ACKNOWLEDGMENT

This work was supported by Fund project “Research on the Construction Method and Application Model of Basic Education Knowledge Graph”, fund number: BCA190087.

## REFERENCES

[1] A. W. Bates. (2005). Technology, e-learning and Distance Education (2nd ed.). Routledge. [Online]. Available: https://doi.org/10.4324/9780203463772

[2] D. Bernhard and I. Gurevych, "Answering learners' questions by retrieving question paraphrases from social Q&A sites," in *Proc. Third Workshop on Innovative Use of NLP for Building Educational Applications*, 2008, pp. 44-52.
[3] A. C. Graesser, Z. Cai, B. Morgan, *et al*., "Assessment with computer agents that engage in conversational dialogues and trialogues with learners," *Computers in Human Behavior*, vol. 76, pp. 607-616, 2017.
[4] T. O'Riordan, D. E. Millard, and J. Schulz, "Is critical thinking happening? Testing content analysis schemes applied to MOOC discussion forums," *Computer Applications in Engineering Education*, vol. 29, no. 4, pp. 690-709, 2021.
[5] R. M. Marra, J. L. Moore, and A. K. Klimczak, "Content analysis of online discussion forums: A comparative analysis of protocols," *Educational Technology Research and Development*, vol. 52, no. 2, p. 23, 2004.
[6] C. M. Chen, M. C. Li, W. C. Chang, *et al*., "Developing a Topic Analysis Instant Feedback System to facilitate asynchronous online discussion effectiveness," *Computers & Education*, vol. 163, 2021.
[7] L. Williams and M. Lahman, "Online discussion, student engagement, and critical thinking," *Journal of Political Science Education*, vol. 7, no. 2, pp. 143-162, 2011.
[8] Shermis, D. Mark, and J. Burstein, Eds., *Handbook of Automated Essay Evaluation: Current Applications and New Directions*, Routledge, 2013.
[9] T. Zesch, M. Wojatzki, and D. Scholten-Akoun, "Task-independent features for automated essay grading," in *Proc. Tenth Workshop on Innovative use of NLP for Building Educational Applications*, 2015, pp. 224-232.
[10] I. Persing and V. Ng, "Modeling argument strength in student essays," in *Proc. 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015, pp. 543-552.
[11] I. Persing, A. Davis, and V. Ng, "Modeling organization in student essays," in *Proc. Conference on Empirical Methods in Natural Language Processing*, 2010, pp. 229-239.
[12] K. Taghipour and H. T. Ng, "A neural approach to automated essay scoring, in *Proc. Conference on Empirical Methods in Natural Language Processing*, 2016, pp. 1882-1891.
[13] Z. Jiang, Z. Shu-fang, G. Wei-li, L. Xue, "TFLA: A quality analysis framework for user generated contents," *Acta Electronica Sinica*, vol. 46, no. 9, pp. 2201-2206, 2018.
[14] J. Kim, E. Shaw, D. Feng, C. Beal, and E. Hovy, "Modeling and assessing student activities in on-line discussions," in *Proc. Workshop on Educational Data Mining at the Conference of the American Asociation of Artificial Intelligence (AAAI-06)*, Boston, MA, 2006.
[15] S. M. Kim, P. Pantel, T. Chklovski, and M. Penneacchiotti, "Automatically assessing review helpfulness," in *Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Sydney, Australia, July 2006, pp. 423-430.
[16] L. Mei, Y. Juan, and L. Yingqun, "The analysis of interbehavior in peer to peer reciprocal learning," *China Educational Technology* vol. 5, pp. 91-97, 2016.
[17] K. F. Lui, S. C. Li, and S. O. Choy, "An evaluation of automatic text categorization in online discussion analysis," in *Proc. IEEE International Conference on Advanced Learning Technologies*, 2007, pp. 205-209.
[18] M. Weimer, I. Gurevych, and M. Muhlhauser, "Automatically assessing the post quality in online discussions on software," in *Proc. ACL 2007 Demo and Poster Sessions*, Prague, June 2007, pp. 125-128.
[19] X. Chao-jun and F. Xiao-min, "Application of LDA in semantic tagging of web educational resource," *Higher Education of Sciences*, vol. 3, pp. 61-67, 2019.
[20] C. Jiaying, Y. Jiong, and Y. Xingyao, "A feature extraction based recommender algorithm fusing semantic analysis," *Journal of Computer Research and Development*, vol. 57, no. 3, pp. 562-575, 2020.
[21] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependence, max- relevance, and min-redundancy," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226-1238, 2005.
[22] J. Hua, D. T. Waibhav, and R. D. Eduward, "Performance of feature selection methods in the classification of high-dimension data," *Pattern Recognition*, vol. 42, no. 7, pp. 409-424, 2009.
[23] J. Lee and D. W. Kim, "Mutual Information-based Multi-label feature selection using interaction information," *Expert Systems with Applications*, vol. 42, no. 4, pp. 2013-2025, 2015.
[24] H. Liu, J. Sun, L. Liu, and H. Zhang, "Feature selection with dynamic mutual information," *Pattern Recognition*, vol. 42, no. 7, pp. 1330-1339, 2013.
[25] Y. Shuo, W. Wei-ya, and L. You-quan, "Research on text clustering based on semantic feature extraction," *Computer Technology and Development*, vol. 30, no. 3, pp. 46-50, 2020.
[26] Z. Yong-qiang, "Research and implementation of personalized book recommendation in university libraries based on text similarity comparison," *Journal of Qinghai Normal University (Natural Science)*, vol. 35, no. 3, pp. 85-91, 2019.



**Wen-Yan Qin,** was born in Lvliang, Shanxi Province, China, in July 16th 1998. She had undergraduate education in the major of Instructional Technology in Shanxi Datong University from 2015 to 2019. She got the bachelor degree in science in 2019.

Ms. Qin is currently a master course student in Nanjing Normal University from 2019, majored in Instructional Technology. She is expected to have the master degree of science in Instructional Technology in June 2022.