*Retraction*

# Retracted: Research on the Relationship between Intelligent Analysis and Weight of Keywords in English Test Questions

## Wireless Communications and Mobile Computing

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

## References

[1] J. Qiu, "Research on the Relationship between Intelligent Analysis and Weight of Keywords in English Test Questions," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 3480746, 11 pages, 2022.

WILEY | Hindawi

*Research Article*

# Research on the Relationship between Intelligent Analysis and Weight of Keywords in English Test Questions

**Jiaosheng Qiu**

*School of Foreign Languages, Taishan University, Tai'an, Shandong 271000, China*

Correspondence should be addressed to Jiaosheng Qiu; qjs@tsu.edu.cn

With the rapid development of educational informationization, Internet of Things, and other technologies, English education has been paid special attention to, and all aspects such as educational model, learning behavior, teaching philosophy, and teaching evaluation have been greatly influenced by educational informationization. Based on the experience of practical education, this paper explores and studies the connotation and characteristics of English test questions and its influence and application on modern English test questions. This paper constructs a systematic method to extract keywords from English test questions. It perfects the fair and reasonable index of English keywords, establishes the weight system, discusses the relationship of keywords, and conducts academic research on vocabulary, word frequency and word position, emphatically adopts BayesNet algorithm to extract keywords, and realizes the evaluation system of English test keywords based on intelligent analysis and weight relationship. The results show that (1) selecting the calculation method and weight relationship suitable for the text system to carry out intelligent analysis, the weight ratio exceeds 65%; that is, the text keyword retrieval is successful. (2) The average accuracy (%), average recall (%) and average $F$-measure (%) in weighted names are almost less than 70%. Only the BayesNet algorithm has 72.3% weight analysis in keyword extraction in reading comprehension. (3) KEA algorithm, PAT TERR algorithm, and BayesNet algorithm take 0-2.8 s, 0-2.6 s, and 0-2.1 s, respectively, and the BayesNet algorithm takes the shortest time. The calculation time of users is greatly saved. (4) According to the calculation results of CPT model, the sum of the weights of the three algorithms is equal to 1, and the BayesNet algorithm is dominant in extracting keywords with a weight analysis of 0.529 in verb translation.

## 1. Introduction

With the popularization of educational informationization and the continuous expansion of the scale of the Internet, the improvement of the education system makes the students receive more perfect and standardized cultural education. Nowadays, the powerful functions and functions of the Internet are more convenient for people to learn, and educational resources can be gradually informationized and intelligent with the progress of the times. In view of the research on English test questions, we can use new technologies to upgrade the test questions to a new level. In this paper, we choose the correct keywords from English test questions, express the text content and the main idea of the article, and make the setting of test questions and the arrangement of investigation objects more reasonable. We have made an in-depth study on how to extract keywords, collected relevant data, and literature for reference research. Literature [1] introduces the function of corpus in assisting students' cognitive schema construction and introduces the application of corpus in English novel selection and reading teaching in senior high schools. The development [2] is beneficial to help learners construct cognitive schema of the language, cultivate innovative thinking, and learn effectively. Literature [3] cultivates students' language ability in international communication and studies the cultural imbalance of teaching materials to students' language learning and cultural construction. Literature [4] countermeasures to improve English textbooks in primary and secondary schools and compiles and penetrates the core literacy of English subjects. Literature [5] realizes the alignment of text, sentence, and chunk, which shows the value of medical English teaching and Chinese-English translation of medical literature. In

literature [6], research and the development of corpus in vocabulary analysis of textbooks are applied to the analysis of English majors' textbooks. Literature [7] cultivate applied English talents and promote students to digest English knowledge. Literature [8] examines the development and evolution of the influence of culture in the English world and introduces the words related to material culture that attracted the attention of the English world earlier. Literature [9] deeply studies the discourse, analyzes and refines the core literacy training value of listening discourse, and promotes the achievement of the core literacy training goal. Literature [10] promotes the acquisition of reading information and adds reading information matching questions in English tests. Literature [11] promotes the teaching and testing of English reading in senior high schools and tests the validity of reading comprehension tests. Literature [12] implements the cultivation of core literacy of English subject and improves teachers' own core literacy and literacy teaching ability. Literature [13] analyzes the problems and challenges in cultivating thinking quality in English and puts forward ways and methods to develop thinking quality. Literature [14] judges the practicality of English and the significance of English education and further discusses the significance of implementing English education in basic education. Literature [15] discusses the training strategies of students' core literacy based on English reading teaching, which improves students' pragmatic and language sense.

## 2. Key Word Extraction Technology of English Test Questions

*2.1. Text Mining-Related Technologies.* The key words in the text are extracted and analyzed quantitatively and qualitatively and realize the visualization of key words in test questions. The four stages are text preprocessing stage, feature extraction and reduction stage, learning and knowledge pattern extraction stage, and knowledge model evaluation stage. The process flow is shown in Figure 1:

*2.2. Selection of Feature Keywords.* Feature keywords are easy to identify, find, and distinguish. The content of the test text is processed in batches and analyzed in structure to achieve the choice of word features. Using words, words, or phrases as text feature items is helpful to improve the operation efficiency of subsequent algorithms. The commonly used evaluation methods of text feature items include word frequency, mutual information, information gain, and expected cross-entropy. For feature words, the above evaluation function is as follows:

(1) Word frequency is as follows: the number of times feature items appear in the text

(2) Mutual information is as follows:

$$MI(t, c) = \log \frac{P(t/c)}{P(t)} \approx \log \frac{A \times N}{(A + D) \times (A + B)}. \quad (1)$$

Among them, the total number of texts in the corpus is denoted as $N$, the text category including feature words $t$ is

denoted as $C$, the conditional probability including feature words $t$ and belonging to category $c$ is denoted as $A$, the conditional probability including feature words $t$ but not belonging to category $c$ is denoted as $B$, and the conditional probability belonging to category $c$ but not including feature words $t$ is denoted as $D$.

(3) Expected cross entropy is as follows:

$$EC(t) = P(t) \sum_{i=1}^{m} P(c_i|t) \log P(c_i|t), \quad (2)$$

where the number of text categories is $m$, $P(t)$ represents the frequency of text containing feature words $t$ in the corpus, and $P(\text{cit})$ represents the conditional probability of containing feature words $t$ and belonging to ci category.

(4) Information gain is as follows:

$$\begin{aligned} IG(t) = P(t) \sum_{i=1}^{m} P(ci|t) \log P(ci|t) \\ + P(\bar{t}) \sum_{i=1}^{m} P(c_i|\bar{t}) \log P(c_i|\bar{t}). \end{aligned} \quad (3)$$

The number of text categories is $M$, $P(T)$ represents the frequency of text containing feature words $T$ in the corpus, $P(\text{CIT})$ represents the frequency of text containing feature words $T$ and belonging to CI class, $P(T)$ represents the frequency of text not containing feature words $T$, and $P(\text{CIT})$ represents the frequency of text not containing feature words but belonging to class.

*2.3. Common Keyword Extraction Algorithm*

*2.3.1. KEA Algorithm.* KEA algorithm [16] is as follows: it is a keyword extraction algorithm based on Naive Bayesian model, and it is a mathematical analytical method. The method is divided into two stages: training stage and extraction stage, and the processing flow is shown in Figure 2:

The KEA algorithm is suitable for well-structured English text keyword extraction but cannot support cross-language, not universal. The accuracy of keyword extraction is greatly reduced, which is not conducive to the correct judgment of users. The calculation formula of KEA extraction keywords is as follows:

$$TF \times IDF = \frac{\text{freq}(P, D)}{\text{size}(D)} \times -\log_2 \frac{\text{df}(P)}{N}. \quad (4)$$

Naive Bayesian model [17] is as follows: the formula for selecting keywords after calculating the score ranking of candidate keywords is as follows:

$$P[\text{yes}] = \frac{Y}{Y + N} P_{TF \times IDF}[t|yes] P_{\text{distance}}[d|yes]. \quad (5)$$
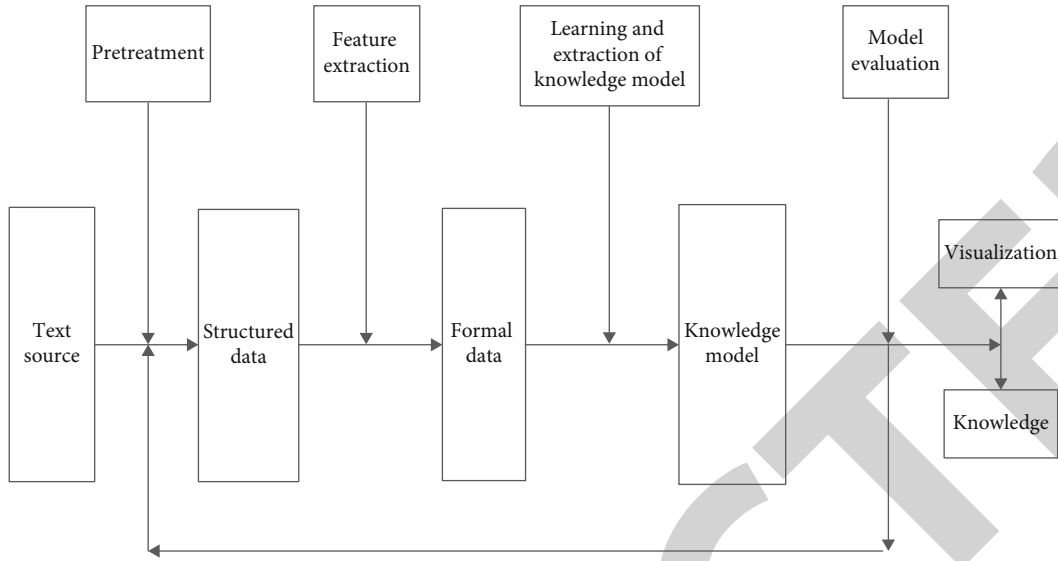
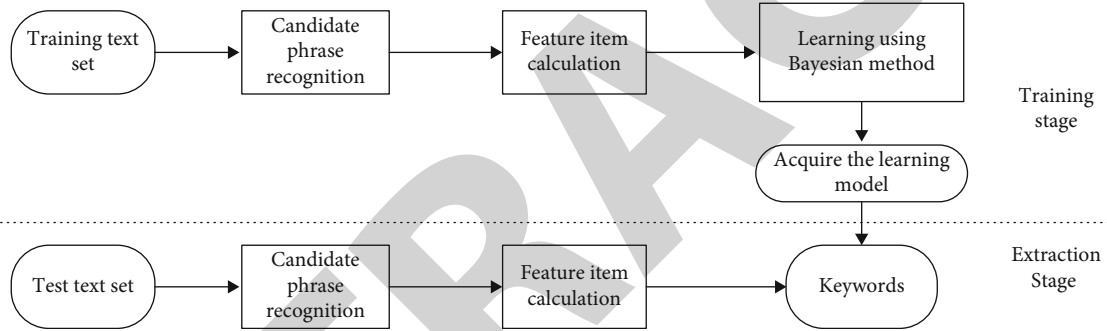FIGURE 1: Schematic diagram of text mining process.



FIGURE 2: Schematic diagram of KEA algorithm flow.

*2.3.2. PAT TREE Algorithm.* PAT TERT algorithm [18] is as follows: it is a good index algorithm in information retrieval. The basic data structure is a binary index tree index [19]. First, the string to be queried is transformed into a bit string in binary format, and the path from root node to each leaf node represents the index bit string. In the binary index tree, the internal node is used to store the index path, and the leaf node is used to index the string information, so that the text can be retrieved quickly. The retrieval process against the input query string $S$ is the process of finding the path in the PAT TERT, as shown in Figure 3:

Figure 3 shows the PAT TREE constructed from the first eight-bit characters of the binary string $S = 01100010110110$. If the string to be queried starts from the root node and matches the first bit 0 of find, it turns to the left node, matches the second bit 0, and continues to turn to the left node, and the third bit 1 turns to the right node. At this time, the internal node value reached is 5, which indicates the difference caused by the fifth bit difference of the query string. The fifth bit is 0; so, choose to turn right and output the leaf node of the result $S$ string. Explain that find exists in $S$ and the query was successful.

*2.3.3. BayesNet Method.* BayesNT algorithm [20] is as follows: taking the prior knowledge as the starting point, the sample data is trained and then transformed into a probability model BayesNet. It includes two parts: BayesNet algorithm and CPT parameters.

*(1) Learning BayesNet Structure.* The process of establishing topological results is to address the appropriate exponential coefficient Nijk to maximize the joint probability $P(G, D)$. The formula is as follows:

$$P(G, D) = P(G) \prod_{i=1}^{n} g(i, \pi_i). \tag{6}$$

Among them,

$$g(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} N_{ijk}!. \tag{7}$$

*2.3.4. CPT Parameter Learning.* CPT [21] parameter learning is as follows: on the premise of establishing the network
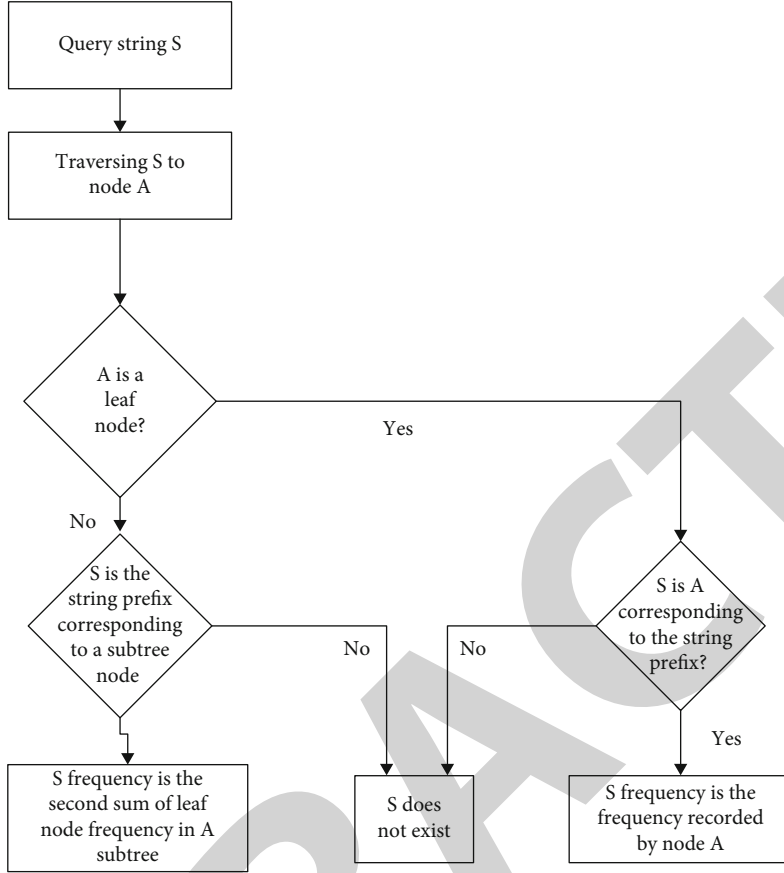
FIGURE 3: A flow chart of PAT TREE retrieval string.

structure, every node in the sample space is tested again.

$$P(\theta_{ij}|D,\xi) = \text{Dirichlet}\left(\theta_{ij}|\partial_{ij1} + N_{ij1}, \partial_{ij2} + N_{ij2}, \cdots, \partial_{ijr} + N_{ijr}\right)$$
$$= \frac{\Gamma(\partial_{ij} + N_{ij})}{\prod_{k=1}^{r}(\partial_{ijr} + N_{ijr})} \prod_{k=1}^{r} \theta_{ijk}^{\partial ijk-1}.$$

(8)

There are usually two ways to simplify or simplify Bayes-Net structure.

(1) Artificial subjective determination, which is generally carried out under the experience of experts, can determine the topological structure according to the causal relationship between nodes

(2) Topological structure is studied by training sample data set. This method is a NP problem academically. At present, there are many algorithms for this kind of learning, and the classic ones are two methods based on search scoring: K2 [22] and MCMC [23]

*2.3.5. Genetic Algorithm.* Genetic algorithm [24] is as follows: it draws on Darwin's theory of biological evolution. It is a randomized search method for the optimal solution evolved from the evolution law of survival of the fittest and

survival of the fittest. This algorithm is put forward by the professor. Its important feature is that it adopts the probabilistic search method for the optimal solution, which can automatically obtain and guide the search space, and can adaptively adjust the search party without making rules. The genetic algorithm has been widely used in various fields because of its many advantages. The mathematical model is as follows:

$$\max/\min f(x)(9-1),$$
$$x \in R(9-2), \qquad (9)$$
$$R \in U(9-3),$$

where formula (9) is the objective function, formula (9-2) and formula (9-3) are constraints, $U$ represents the basic set, and $R$ is a subset of $U$.

*(1) The Basic Idea of the Genetic Algorithm.* There are two difficulties in keyword extraction in English test text: first, the content of test text is short and the feature items are sparse; so, the traditional text keyword extraction algorithm based on statistical model or semantics cannot select complete and effective feature items; secondly, the text of the test questions is diverse, and the forms are changeable; so, it is
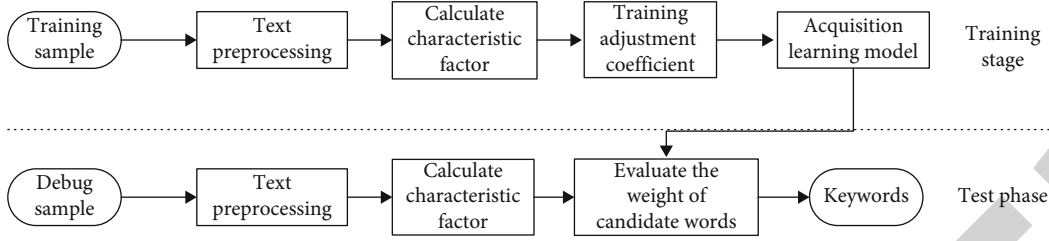
Figure 4: Framework diagram of the keyword extraction algorithm based on adaptation.

impossible to realize the weight evaluation compared with the traditional keyword extraction algorithm.

For the first question, based on the algorithm, combined with the structural features of the test text and English semantic features, position factor, word length factor, and word cooccurrence factor are proposed, which makes up for the incomplete defect of selecting text feature items only by word frequency factor. The introduction of multiple feature weight factors greatly increases the calculation time consumption of the algorithm.

To solve the second problem, four feature item weight adjustment coefficients are introduced into the feature item weight evaluation function to adapt to different types of test text. The keyword extraction algorithm has two stages: training stage and testing stage, as shown in Figure 4.

(2) *Genetic Operation Selection Operator Operation*

Select operator [25] is as follows: it selects the individuals whose fitness is closest to the threshold and inherits them to the next generation by directly inheriting or pairing and crossing, which embodies the evolutionary law of survival of the fittest in evolution. The calculation formula is as follows:

$$P(t_i) = \frac{\text{fitness}(t_i)}{\sum_{i=1}^{n} \text{fitness}(t_i)}, \quad (10)$$

where the number of individuals in the population is $n$, the fitness of individual ti is denoted as fitness, and the probability of ti being selected is denoted as $P$ (ti).

*2.4. Summary of the Experiment in This Chapter.* This chapter focuses on five commonly used keyword extraction algorithms: KEA algorithm, PAT TERT algorithm, BayesNet keyword weight allocation method, and further compares the weights, and analyzes the advantages and disadvantages of data model-based method, word meaning-based method, and intelligent learning-based method, which prepares for further research.

## 3. Keyword Extraction Method Based on Feedback Statistics

Based on the data statistics and understanding of user feedback, the BayesNet algorithm is obviously higher in accuracy and efficiency than other algorithms in KEA algorithm, PAT TERT algorithm, and BayesNet keyword weight allocation method, which provides users with a comprehensive analysis and systematic method.

*3.1. Weight Analysis of Key Words in English Test Questions.* There are many factors that affect the importance of keywords, most of which are adopted at present, including (1) the position where words appear, (2) frequency of words, and (3) the frequency of cooccurrence of words and some synonyms.

*3.2. Research on Classification Algorithm.* Statistical analysis method of classification variables is based on logistic regression. The function is as follows:

$$h_\theta(x) = g\left(\theta^T x\right) = \frac{1}{1 + e^{-\theta^{Tx}}}, \quad (11)$$

$$g(z) = \frac{1}{1 + e^{-z}}. \quad (12)$$

Logistic regression is a binary classification method that judges whether the result belongs or not. The formula is as follows:

$$p(y = 1|x;\theta) = h_\theta(x), \quad (13)$$

$$p(y = 0|x;\theta) = 1 - h_\theta(x). \quad (14)$$

The results of finding the maximum likelihood estimation formula are as follows:

$$\theta_j := \theta_j + \partial\left(y^{(i)} - h_\theta\left(x^i\right)\right)x^i_j. \quad (15)$$

The logistic regression model is as follows:

$$P(Y = 1|x) = \pi(x) = \frac{1}{1 + e^{-g(x)}}. \quad (16)$$

The logical occurrence probability formula is as follows:

(1) Define the conditional probability of no event as

$$P(Y = 0|x) = 1 - P(Y = 1|x) = 1 - \frac{e^{g(x)}}{1 + e^{g(x)}}$$
$$= \frac{1}{1 + e^{g(x)}}. \quad (17)$$

(2) Define the probability ratio of occurring events to nonoccurring events as

$$\frac{P(x=1|x)}{P(x=0|x)} = \frac{p}{1-p} = e^{g(x)}. \tag{18}$$

This ratio is called odds. There are $0 < p < 1$ and odds $> 0$. That is, take the logarithm of odds.

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \sum_{l-1}^{k-1} \beta_{jl} D_{jl} + \beta_p x_p. \tag{19}$$

*3.3. BayesNet Algorithm.* The BayesNet algorithm is a mathematical model based on probabilistic reasoning. The BayesNet algorithm is to solve the problem of uncertainty and incompleteness. BayesNet consists of two components defining a set of directed acyclic graphs and conditional probability tables. For each variable, BayesNet has a conditional probability table (CPT), and the CPT of variable $Y$ describes the conditional distribution. The BayesNet mathematical model is as follows:

$$B = (V, E, P). \tag{20}$$

Among them:

$$\text{Random variable set}: V = \{V_1, V_2, \cdots, V_n\}, \tag{21}$$

$$\text{A set of directed edges}: E = \{V_i V_j | V_i, V_j \in V\}. \tag{22}$$

Conditional probability distribution set, that is, conditional probability table is

$$P = \{P(V_i | V_1, V_2, \cdots, V_{i-1}, V \in V)\}. \tag{23}$$

*3.3.1. Establishment of the BayesNet Algorithm.* According to the location and frequency of keywords in English test questions, a network database is formed to split and count, so as to achieve the integrity of the data and realizes the expression of the collection.

*3.3.2. Establishment of the BayesNet Initial Model to CPT.* The SimpleEstimator algorithm is used to calculate the conditional distribution of variables, and the proportion of different keywords in the test text is realized. The formula is as follows:

$$P(x_i = k | pa(x_i) = j) = \frac{N_{ijk} + N'_{ijk}}{N_{ij} + N'_{ij}}. \tag{24}$$

*3.3.3. Calculation of User Feedback Statistics.* Select the key words to express the theme of the article in the test questions for analysis and find that the frequency of key words will be significantly higher than other words. It is explained that whether a word is a keyword or not can be judged according to the frequency of its occurrence. The sampling formula is as follows:

$$n \geq \left(\frac{k}{a}\right)^2 p(1-p). \tag{25}$$

*3.4. Summary of This Chapter.* This chapter solves the problem that the text style of test questions is difficult to understand and the meaning of synonyms is difficult to judge by various methods and introduces a judgment model to adapt to different kinds of keyword evaluation of test questions.

## 4. Experimental Analysis of Keywords in English Test Questions and Setting of the Evaluation System

*4.1. Algorithm Comparison.* For the comparison of keyword extraction methods in English test questions, there are great differences in extraction efficiency in different stages and methods, so that the analysis results are slow and the accuracy rate is reduced. Comparison of the weight of three algorithms in reading comprehension, cloze test and test translation is as follows: the comparison table is shown in Figure 5.

It can be seen from Figure 5 that the average accuracy (%), average recall (%), and average -measure (%) of the BayesNet algorithm are obviously higher than the KEA algorithm and PAT TREE algorithm with the highest value of 72.3. That is to say, the BayesNet algorithm has achieved high efficiency in extracting keywords from English test questions and achieved effective extraction of keywords with high accuracy.

*4.2. Correct Rate Comparison.* Through the frequency and position of keywords and the change of test questions, to judge the correctness of keyword extraction, the experimental results are shown in Figures 6 and 7:

It can be seen from Figure 6 that the correct rate weight of the BayesNet algorithm for keyword extraction in reading comprehension is between 3.7 and 5.9, which is significantly higher than the correct rate of the KEA algorithm and PAT TERR algorithm (2.1-5.1), which shows that the BayesNet algorithm is efficient in keyword extraction in reading comprehension.

As shown in Figure 7, the weight of accuracy in cloze keyword extraction is 9.5 in the BayesNet algorithm, which is obviously higher than the maximum value of 5.8 in the KEA algorithm and 5.5 in the PAT TERR algorithm; that is, the BayesNet algorithm has the highest accuracy in cloze keyword extraction.

As can be seen from Figure 8, the BayesNet algorithm (3.1-8.9), KEA algorithm (3.2-5.3), and PAT TERR algorithm (3.3-6.2) are the weights of keyword accuracy in translated test questions; that is, the BayesNet algorithm has the greatest interval in translated test questions, which is more conducive to comprehensive and detailed keyword extraction.

*4.3. Efficiency Comparison.* In order to analyze the extraction time of keywords and the extraction speed of multiclass text features, the experimental results of the three algorithms are analyzed according to keyword weights, as shown in Figure 8:
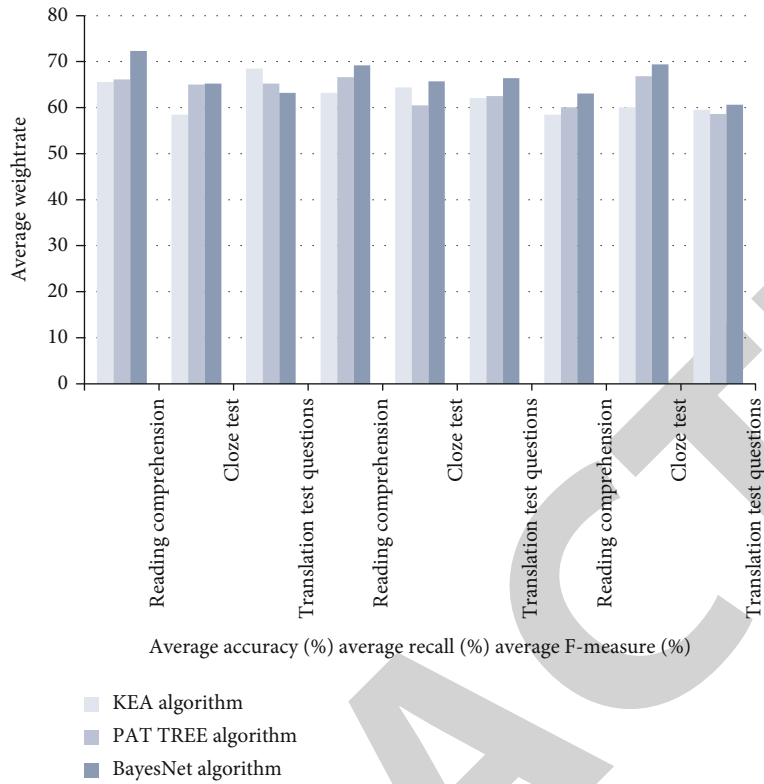
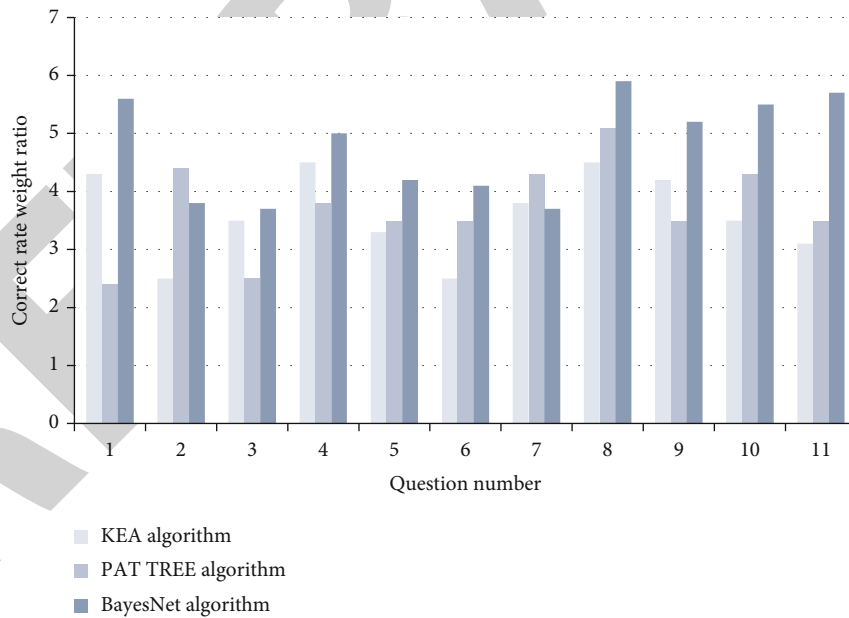Figure 5: Comparison table of keyword extraction results based on three algorithms.



Figure 6: Weight diagram of keyword accuracy rate in reading comprehension.

As can be seen from Figure 9, with the increase of the number of texts, the time used to extract keywords in the three algorithms also increases, showing a positive correlation. Among the 0-16 texts, KEA algorithm, PAT TERR algorithm, and BayesNet algorithm take 0-2.8 s, 0-2.6 s, and 0-2.1 s, respectively, and the BayesNet algorithm takes the shortest time.

4.4. Comparative Analysis of Multifeature Keyword Weights. Because each word shows different meanings in different articles and sentences, according to the weight analysis of nouns, verbs, adjectives, and adverbs of words, the weight analysis of words by different algorithms (judged according to the initial model CPT) is shown in Figure 9.
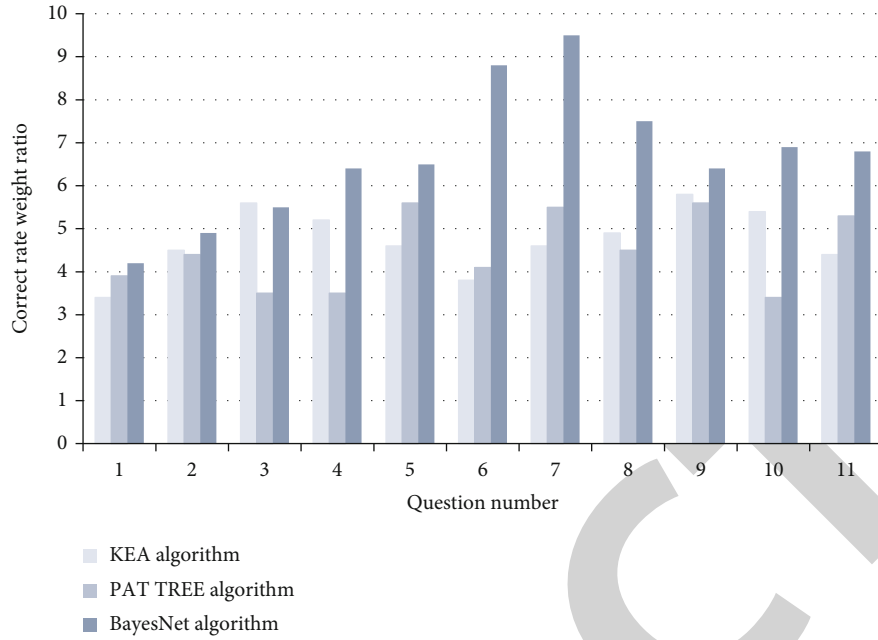
KEA algorithm
PAT TREE algorithm
BayesNet algorithm

FIGURE 7: Weight diagram of keyword accuracy rate in cloze test.



KEA algorithm
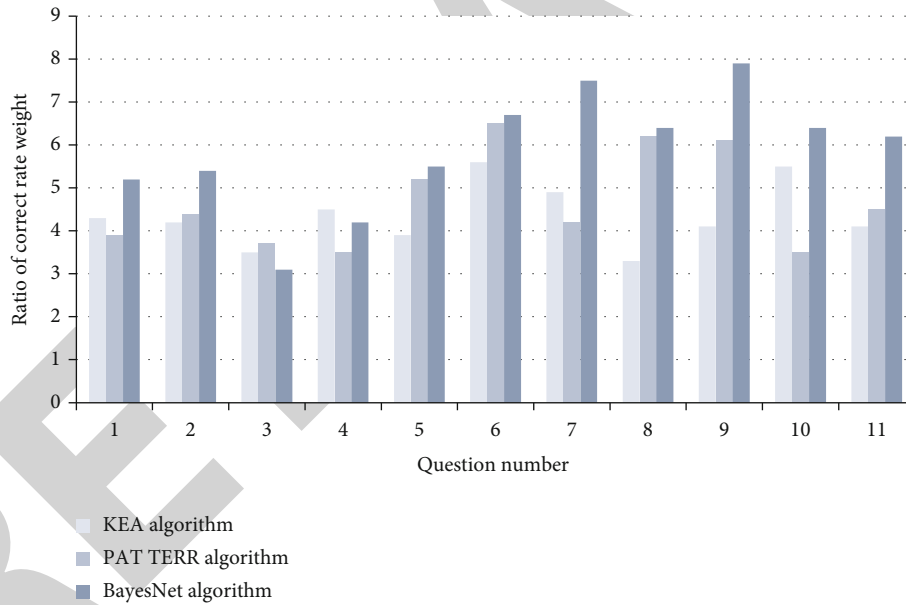PAT TERR algorithm
BayesNet algorithm

FIGURE 8: Comparison of the correct rate weights of keywords extracted by three algorithms in translation test questions.

It can be seen from Figure 10 that the three algorithms mainly extract verb keywords, among which the BayesNet algorithm ranks first with 0.549 data in verb CPT weight coefficient, but the KEA algorithm has the highest 0.361 data in noun CPT weight coefficient; so, it is known that various algorithms have different proportions under different word meanings.

*4.5. Summary of This Chapter.* This chapter carries on the experiment analysis from the different test question type, the correct rate, the efficiency, and the CPT model weight relations and obtains the experiment result, judges the differ-

ent test question corresponds to the different extraction method the experiment foundation, and obtains the verification. Because each word shows different meanings in different articles and sentences, different keyword extraction methods are used to adapt to different test questions, and various algorithms are used for data analysis.

## 5. Conclusion

With the development of education and the popularization of the Internet, parents, teachers, and students attach importance to learning, and English education has obviously
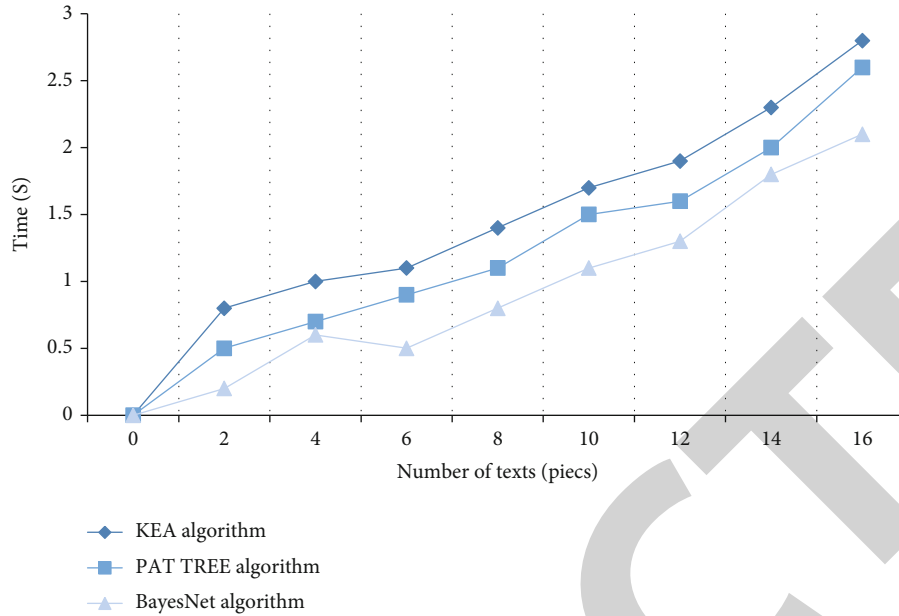
Figure 9: The usage time of three algorithms to extract keywords under different number of test questions.
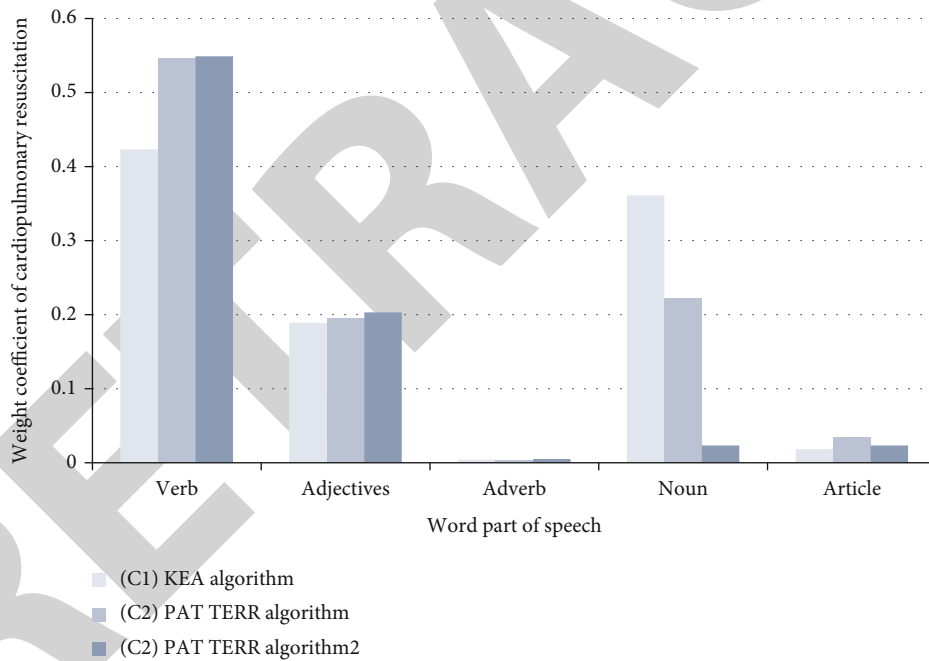


Figure 10: CPT coefficients.

become an indispensable part, using appropriate keywords to examine the students' learning situation and the formation of the teacher's key knowledge explanation system, so that students can better adapt to learning and improve the efficiency of students' learning. From simple vocabulary to the application of key words in test questions, we can adjust the teaching content by understanding key words in order to achieve a better educational environment. We make a suitable keyword extraction system for English test questions, research and develop a set of important evaluation indica-

tors for English keywords, and realize the characteristics of important keywords through the reaction of data.

According to the research results and work characteristics, we can summarize the following points: (1) establish a set of important indicators to judge keywords. Through the weight relationship of keywords, the frequency of vocabulary appearance, and the constant changes of test questions, we can judge whether the word is a keyword and analyze it, reduce the uncertainty, and then judge it as a keyword, which is the increase of data crying vocabulary and

comprehensively evaluate the keywords of the article. (2) Using dynamic analysis and weight relation to study the extraction of English keywords, it enables students to fully grasp the key words of English test questions and improve their learning efficiency. In order to adapt to various test questions, structured comparison and judgment are conducive to the accuracy of data, increase the grasp of vocabulary and questions, greatly facilitate the understanding of test questions, and improve the efficiency of dealing with problems. (3) BayesNet algorithm, KEA algorithm, and PAT TERR algorithm are used, and the commonly used keywords are extracted and verified, which greatly improves the wrong judgment of keywords, puts forward the semantic judgment and recognition function, and improves the accuracy of keywords in the test text. (4) Choosing the right method reduces the blindness and error in the extraction process and evaluates the test questions to realize the simplicity and rapidity of core vocabulary and execution mode.

In the future, the following aspects need to be improved: (1) because of the variability of English words in the test questions, there will be errors in obtaining keywords, which will affect the accuracy of keyword extraction; so, we can further optimize the algorithm. (2) The imbalance of weight ratio is easy to occur in the weight relationship, which will lead to the omission and deficiency of some keywords extraction, make the overall understanding of the test text appear errors and misjudgments, and reduce the weight reputation. (3) The popularization of Internet, the promotion of intelligent technology, the slowdown of calculation speed, and the untimely update of database lead to the omission of key words in test questions; that is, the timely update of the system makes the information comprehensive and accurate. (4) The phenomenon of synonyms is not considered. It slows down the extraction speed and increases the wrong judgment of keywords and the wrong understanding of test questions. That is, we need to consider the dynamic analysis from many aspects to achieve the accuracy of the results.

## Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

## Acknowledgments

## References

[1] Z. Yan, "Application of corpus in senior high school English novel selection and reading teaching," *Journal of Shaoguan University*, vol. 37, no. 11, p. 4, 2016.

[2] X. Hongliang, "Research on improving English learning effectiveness based on corpus," *Journal of Fuyang Normal University: Social Science Edition*, vol. 2, p. 5, 2016.

[3] L. Yanhong, J. Z. Lawrence, and M. Stephen, "Textbook culture research based on National Planning College English Textbook Corpus," *Foreign Language Circle*, vol. 6, pp. 85–93, 2015.

[4] W. Chi and Y. Yuntian, "40 years of English teaching materials construction in primary and secondary schools in China," *Curriculum, Teaching Materials and Teaching Methods*, vol. 38, no. 9, pp. 8–13, 2018.

[5] H. Kaibao and L. Yi, "Research on the establishment and application of contemporary English textbook corpus," *Audiovisual Teaching of Foreign Languages*, vol. 3, pp. 34–39, 2016.

[6] T. Jieyi, "Corpus-assisted lexical chunk analysis in English textbooks-taking verb chunks as an example," *Journal of Shaoguan University*, vol. 33, no. 3, pp. 189–193, 2012.

[7] C. Caihong, "Research on the cultural character of English discipline literacy," *Educational Theory and Practice*, vol. 38, no. 8, pp. 47–49, 2018.

[8] S. Bin, "A study on the influence of Zhejiang cultural keywords in the English-speaking world-from the perspective of cultural omics," *Zhejiang Journal*, vol. 2, pp. 201–207, 2017.

[9] L. Huifang, "English listening and speaking teaching in middle schools directing to the cultivation of core literacy," *Teaching and Management*, vol. 34, pp. 51–54, 2019.

[10] L. Jieyuan and Q. Luxia, "A study on backwash effect of new English reading questions in Guangdong college entrance examination," *English Teachers*, vol. 10, pp. 2–15, 2012.

[11] H. Juan, "A study on the content validity of English reading comprehension questions in 2017 college entrance examination," *Education Measurement and Evaluation*, vol. 4, pp. 17–24, 2018.

[12] G. Baoxian and Z. Jianzhong, "How to cultivate the core literacy of English in classroom teaching," *Curriculum, Teaching Materials and Teaching Methods*, vol. 39, no. 4, pp. 66–71, 2019.

[13] C. Zehang, W. Qiang, and Q. Xiaofang, "On the thinking quality and its development path in the core literacy of English discipline," *Curriculum, Teaching Materials and Teaching Methods*, vol. 39, no. 1, pp. 91–98, 2019.

[14] C. Xiaotang, "Thoughts on the adjustment of current English education policy," *Curriculum, Teaching Materials and Teaching Methods*, vol. 34, no. 5, pp. 58–64, 2014.

[15] L. Mingyuan and P. Huaqing, "Cultivation of students' core literacy in English reading teaching," *Teaching and Management*, vol. 28, pp. 58–60, 2016.

[16] Z. Zhuoping, "The study on metaphors in Robert Frost's poems," *Overseas English*, vol. 13, no. 1, pp. 222–225, 2019.

[17] L. Ying, W. Wenqin, and X. Lixin, "A diachronic study on the content validity of English reading comprehension test in Jiangsu college entrance examination (2009-2015)," *Educational Measurement and Evaluation on Edition*, vol. 2, pp. 23–29, 2016.

[18] W. Qiang, "From comprehensive language application ability to core literacy of English discipline," *English Teacher*, vol. 16, pp. 6-7, 2015.

[19] X. Wang, "Analysis of the core literacy thought of Bloom's cognitive goal taxonomy (revised edition)," *Modern Basic Education Research*, vol. 29, no. 1, pp. 80–86, 2018.

[20] X. Guming, "Connotation analysis of language ability as the core literacy of English discipline," *Journal of Xingyi Normal University for Nationalities*, vol. 1, pp. 103–109, 2018.

[21] X. Tao, Y. Jiang, and L. Xia, "Construction of core literacy model of students in compulsory education in China," *Journal of Beijing Normal University (Social Science Edition)*, vol. 1, pp. 5–11, 2013.

[22] Y. Hui, "Design of English teaching activities led by the core literacy of "thinking quality"," *Educational Theory and Practice*, vol. 37, no. 29, pp. 47–49, 2017.

[23] Z. Ming, "On English writing teaching in senior high schools from the perspective of learners," *Teaching and Management*, vol. 13, pp. 48–51, 2018.

[24] Z. Fen, "Exploration of in-depth teaching mode based on the improvement of English core literacy," *Teaching and Management*, vol. 30, pp. 102–104, 2018.

[25] R. Meiqin and W. Chaoling, "Research on English teaching design in senior high schools pointing to the core literacy of disciplines," *Global Education Outlook*, vol. 49, no. 7, pp. 79–91, 2020.