Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1170622, 10 pages
https://doi.org/10.1155/2021/1170622

Research Article

Intelligent Recognition and Teaching of English Fuzzy Texts Based on Fuzzy Computing and Big Data

Ling Liu [1] and Sang-Bing Tsai [2]

[1] *School of Foreign Languages, Gannan University, Ganzhou, Jiangxi 341000, China*
[2] *Regional Green Economy Development Research Center, School of Business, Wuyi University, China*

Correspondence should be addressed to Ling Liu; benbenbianbian1615@163.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 2 June 2021; Revised 20 June 2021; Accepted 1 July 2021; Published 12 July 2021

Academic Editor: Yuanpeng Zhang

In this paper, we conduct in-depth research and analysis on the intelligent recognition and teaching of English fuzzy text through parallel projection and region expansion. Multisense Soft Cluster Vector (MSCVec), a multisense word vector model based on nonnegative matrix decomposition and sparse soft clustering, is constructed. The MSCVec model is a monolingual word vector model, which uses nonnegative matrix decomposition of positive point mutual information between words and contexts to extract low-rank expressions of mixed semantics of multisense words and then uses sparse. It uses the nonnegative matrix decomposition of the positive pointwise mutual information between words and contexts to extract the low-rank expressions of the mixed semantics of the polysemous words and then uses the sparse soft clustering algorithm to partition the multiple word senses of the polysemous words and also obtains the global sense of the polysemous word affiliation distribution; the specific polysemous word cluster classes are determined based on the negative mean log-likelihood of the global affiliation between the contextual semantics and the polysemous words, and finally, the polysemous word vectors are learned using the Fast text model under the extended dictionary word set. The advantage of the MSCVec model is that it is an unsupervised learning process without any knowledge base, and the substring representation in the model ensures the generation of unregistered word vectors; in addition, the global affiliation of the MSCVec model can also expect polysemantic word vectors to single word vectors. Compared with the traditional static word vectors, MSCVec shows excellent results in both word similarity and downstream text classification task experiments. The two sets of features are then fused and extended into new semantic features, and similarity classification experiments and stack generalization experiments are designed for comparison. In the cross-lingual sentence-level similarity detection task, SCLVec cross-lingual word vector lexical-level features outperform MSCVec multisense word vector features as the input embedding layer; deep semantic sentence-level features trained by twin recurrent neural networks outperform the semantic features of twin convolutional neural networks; extensions of traditional statistical features can effectively improve cross-lingual similarity detection performance, especially cross-lingual topic model (BL-LDA); the stack generalization integration approach maximizes the error rate of the underlying classifier and improves the detection accuracy.

1. Introduction

The important means of education informatization is to apply information technology and network technology to education to realize the mode of “Internet + education” [1]. Education informatization covers various aspects such as education management, education process, and education resources. In terms of educational resources, in addition to physical/digital forms such as paper and digital textbooks, tutorial materials, teaching guides, and practice problems, there is a large amount of data, information, and resources in the network that can provide support for education informatization. In web-based education, learning resources can be shared and students can have a more generalized learning space. In the new informatization education platform, the learning place has changed from the fixed classroom and fixed class time to anytime and anywhere teaching, and the learning means has changed from traditional blackboard and paper books to multimedia video and audio and electronic courseware, etc. The education form is getting richer and richer, and network education starts to play an increasingly important role. According to different concepts and objects, online education can be divided into online education, distance education, adaptive education, virtual classroom, and other forms.

In addition to the above roles of text similarity detection, its direct application has considerable practical value [2]. For example, it has a significant role in protecting the intellectual property rights of electronic texts and combating illegal copying and plagiarism of academic results. Since its launch, the "Academic Misconduct Detection System" of China Knowledge Network has blocked the publication of articles with high repetition rate at the source of academic results; the CrossCheck antiplagiarism literature detection system developed by the International Publishing Links Association has minimized the textual repetition rate of English publications [3]. The above two systems are both text similarity-based detection systems, which shows that the technology has important application value and social significance for combating paper forgery, correcting academic culture, and promoting independent innovation. Unfortunately, most of the successful text similarity detection systems are monolingual, and cross-lingual similarity detection systems have been tried, but the detection effect is poor [4]. The frequency-based word packet model constructs text feature engineering only by the frequency or number of words appearing in the document which mainly leads to the advantage of words with higher frequency or number in the text data in the text feature extraction process.

This thesis investigates data mining methods and application research for personalized learning, providing data-driven solutions. The research objects are learners, learning resources, and personalized learning mechanisms. Learners are broadly defined as students who are engaged in learning activities (including online students and offline students whose academic activities are recorded). Learning resources are the practice questions (including daily homework questions, test questions, and general classroom exercises) that have been recorded by various products and systems and are available to students for learning. Learning strategies are broadly defined as those that can support and ensure learners perform personalized learning recommendation strategies, focusing on, for example, online practice question recommendation algorithms. Although the relevant research on relation extraction is relatively adequate, based on the above analysis and introduction, the extraction of semantic relations from short English texts is highly differentiated from the classical relation extraction task. Studying this topic is of great importance both at the algorithmic level and at the application level.

2. Current Status of Research

Calzada et al. proposed the LexVec model, which uses a sampling-based approach to decompose the PPMI matrix and reconstructs the loss function to penalize excessive word-frequency cooccurrence errors [5]. PPMI-SVD, on the other hand, uses the truncated singular value decomposition (SVD) technique to achieve dimensionality reduction, aimed at finding the maximum features to describe the semantics of words and the retained dimensional features to approximate the PPMI matrix composed of word-context information [6]. Charitopoulos et al. point out that the SVD rank reduction algorithm does not guarantee the nonnegativity of the matrix decomposition, so nonnegative matrix factorization (NMF) becomes a sensible choice for decomposing the PPMI, which guarantees the nonnegativity of the dimensional approximate reduction of the semantic relationship of the word-context information and is more consistent with the semantic relationship hypothesis [7]. Raveh et al. demonstrated that matrix factorization methods are consistent with the word vectors of neural network language models in terms of task performance, such that PPMI-SVD matrix decomposition is equivalent to skip-gram (SGNS) based on negative sampling [8]. The neural network language model has an inherent advantage over factorization methods, which have more flexible super parameter settings [9].

With the development of the Internet, the dramatic increase of all kinds of information, especially textual information, has brought great challenges to the management of knowledge and the mining of relationships among knowledge [10]. There is an urgent need to retrieve the target information more quickly and efficiently and even expect intelligent access to all information related to the target [11]. To meet this demand, knowledge bases are created, in which ontologies are one of the means to manage knowledge more conveniently and to explore the relationships between knowledge [12]. As creating ontologies manually is not very practical in ontology engineering, it is time-consuming to build ontologies manually especially when the amount of data is large. While semiautomatic construction is difficult to reconcile with manual construction and also suffers from subjective problems, many scholars have started to try to develop automatic ontology creation methods [13]. For example, Levi et al. proposed a novel event-based ontology construction method that extracts domain ontologies mainly from unstructured text data [14]. We know that there is rich knowledge in many data, such as text and databases [15]. The automatic ontology construction method automatically extracts knowledge from these data to conform to a specific format (ontology). However, the results are always unsatisfactory because the ontologies generated by the automatic process are not accurate and of low quality, so automatic construction is still a challenge that is not yet perfect.

In contrast, semiautomatic construction is based on a combination of manual and automatic methods, which needs to coordinate the balance between manual and machine, and the semiautomatic approach will encounter the problems existing in both manual and automatic [16]. The automatic ontology construction approach can effectively extract knowledge from a large amount of text to build ontologies based on such knowledge, but the difficulty of automatic construction is that it requires the builder to be knowledgeable about the domain and the developer to have a proficiency in the key technologies involved in the construction [17].

Whether semiautomatic or automatic construction, the key technologies involved mainly include knowledge representation, entity recognition, relationship extraction, graph storage, relationship inference, entity disambiguation, and other technologies. Building an ontology requires systematic integration of these technologies to form a complete set of processes. In traditional semiautomatic construction methods, it is necessary to rely on linguistic tools or artificially construct many features (including lexical and semantic features) and then train a machine learning model for each task (e.g., entity recognition and relation extraction). The disadvantage of this approach is that the applicability of the model is low and the performance of the model depends heavily on the quality of the features. With the development of deep learning, neural network-based methods have been proposed, but these methods still have some shortcomings, such as the extraction of multiple relationships and the problem of overlapping entity labels. Given the shortcomings in the key techniques of the above ontology construction methods, this paper conducts an in-depth study and proposes a solution with better performance.

3. Analysis of Parallel Projection and Region-Expanding English Fuzzy Text Intelligent Recognition Teaching

3.1. English Fuzzy Text Intelligent Recognition Algorithm Design. Text mining is a mathematical modeling approach to characterize the important relationship between internal rules and the semantic level of text by using mathematical statistics as a general idea. However, as nonnumerical information, text cannot be directly used as input data for computer modeling languages to obtain results. Therefore, converting text-based data into computer-readable numerical data becomes a prerequisite and foundation for text mining. On the one hand, the bag-of-words model based on word frequency constructs text feature engineering only by the frequency or frequency of words appearing in the document, which leads to the fact that word items with higher frequency or frequency in text data will have a natural advantage in the text feature extraction process [18]. When extracting features from documents, using frequency alone as the measure of word item importance can result in the absence of key information in the text. In text data, certain words or phrases may occur infrequently, but still contain key information of the document. In general, the frequency of deactivated words in the text is much higher than that of keywords that can reflect the theme of the document, so removing deactivated words will significantly improve the accuracy of feature extraction; on the other hand, when the volume of text data is large, this high-dimensional sparse matrix representation will bring a great difficulty to the model operation.

The training data consists of historical observations of the explanatory variables and the corresponding response variables. The model can predict the values of the response variables corresponding to explanatory variables that are not in the training data. The goal of the regression problem is to predict the continuous values of the response variables. Several linear regression models will be learned, and the training data, modeling and learning algorithms, and evaluation of the effectiveness of each method will be presented later.

The word frequency-inverse document frequency model, abbreviated as TF-IDF model, which combines the word frequency of words in text data with their inverse document frequency, started as a metric to show the ranking function of search engine user query results and has now become one of the important methods for information retrieval and text feature extraction. The mathematical representation of the model is shown in

$$\mathrm{tf} + \mathrm{idf} = \mathrm{tf} \times \mathrm{idf}, \tag{1}$$

where tf denotes the word frequency of a text unit, idf denotes its inverse document frequency, so the tf-idf value of the text unit is the product of the two. First, the word frequency tf is the frequency value of a word in a particular document; the frequency value has a variety of different ways to represent, either its absolute frequency value or frequency value or can be a binary variable, that is, the word in the document is expressed as 1, and vice versa is 0. Suppose there is a word set w and document set D, as shown in equation (2), then the frequency value of the word w, in document D, which is the tf value is shown in equation (3).

$$w = \{w_1, w_2, w_3, \cdots, w_k\}, \tag{2}$$

$$D = \{D_1, D_2, D_3, \cdots, D_k\}, \tag{3}$$

$$\mathrm{tf}(w_i, D_i) = N_{w_1,D_j} \sum_{n=1}^{K} N_{w_n,D_j} N, \tag{4}$$

where N_{w_1,D_j} denotes the frequency of text units w, in document D, and the denominator denotes the total frequency of each word in the dictionary in document D_i. Second, the inverse document frequency idf is the inverse of the document frequency of each text unit, obtained by dividing the total number of documents by the number of documents containing the specified text unit and then taking the logarithm of the resulting quotient. The idf value is calculated as shown in

$$\mathrm{idf}(w_i) = \ln \frac{|D|}{\left|\{D_i : D_j \in w_j\} - 1\right|}. \tag{5}$$

The TF-IDF model based on the bag-of-words model achieves feature extraction of documents based on statistical word frequency combined with inverse document frequency. However, text data belongs to sequence-type information, and the position relationship of text units has a very important influence on semantics, so it is impossible to understand the meaning of text content accurately by relying solely on frequency-based features. In particular, in natural language processing tasks, the preceding and following texts are very closely related, so each word, phrase, or sentence cannot be analyzed in isolation [19]. Exploring text feature extraction models based on sequence information, combining frequency characteristics with text unit location information, has become an important research direction for document feature engineering construction. In this paper, we adopt the document word vector training model based on an artificial neural network and combine it with the TF-IDF model to propose the tf-idf weighted average word vector to represent document features, as shown in Figure 1.

The key point of the fitting process is whether the values of the parameters such as the weight term and bias term in the model satisfy the corresponding conditions. The proposed training algorithm achieves automatic learning of the relevant parameters, and the model can automatically modify the values of the weight terms and bias terms by training on the sample data, thus solving the problem of poor generality of human settings. The algorithm first initializes the values of the parameters to 0 and then modifies them using the perceptron iteration rules, which are shown in

$$w_i = w_i - \eta(t - y)x_i, \tag{6}$$

$$b = b - \eta(t - y). \tag{7}$$

w_i is the weight term corresponding to the input information x_i, b is the bias term, t is the actual value of the training sample, y is the output value calculated by the model, and n is the learning rate, which is used to adjust the magnitude of each weight adjustment. In the training algorithm, the model calculates the output value corresponding to each training data in turn and adjusts the weight term accordingly using the above iteration rule. By calculating the sample data in turn and iteratively modifying the weight terms, the training set data are processed iteratively to finally obtain the relevant parameters of this perceptron model, to achieve an accurate fit to the target function.

When used as an independent unit, a perceptron can fit a linear function; when used as a component in an artificial neural network, a perceptron is called a neuron. These neurons are interconnected according to certain rules to form a complete neural network model. Each neuron in the model is computed in the same way as the perceptron. The structure of a classical fully connected neural network is shown in Figure 2.

The neurons in the neural network model are laid out in a layer structure. The leftmost layer in Figure 2 is the input layer with three nodes, which is responsible for receiving the input data; the rightmost layer is the output layer with two nodes, which is responsible for the output of the model; the layer between the input layer and the output layer is called the hidden layer with four nodes, which is responsible for the intermediate transmission of the data. The complexity of our model is relatively low, and the performance of the general configuration can meet the demand; of course, the higher the performance the faster the time. In the fully connected neural network model shown above, all neurons in adjacent layers are connected, i.e., each neuron in layer IV is connected to all neurons in layer N-1, and the output of the neuron in layer N-1 is the input of the neuron in layer IV; meanwhile, there is no connection between neurons in the same layer, and each neuron is individually numbered, and the operation of this neural network model as the algorithm of this neural network model is described as follows. First, the output value of the nodes in the input layer is the input vector X. Therefore, the dimension of the input vector should be the same as the number of neurons in the input layer; second, after obtaining the output value of each neuron in the input layer, the output value of each node in the hidden layer can be calculated. Take note of the hidden layer as an example, its calculation method is shown in

$$a = f\left(W^T \otimes X\right). \tag{8}$$

After the above steps, the artificial neural network model can obtain the predicted values of the sample data input to the model, thus achieving the purpose of supervised machine learning. In addition to the most basic three-layer fully connected neural network model, many variants such as deep neural networks, convolutional neural networks, and recurrent neural networks have emerged after a long period of research and development. The number of hidden layers is changed from a single layer to multiple layers, and with the increase of the number of hidden layers, the fitting of the algorithm model will show better results, but the difficulty of the model training and the computing time will increase exponentially, so there are high requirements on the computer hardware conditions. In deep learning with the deep neural network as the core, the number of hidden layers should be controlled within 10 layers in general. A sentence contains a lot of information, but it cannot have a corresponding meaning when separated by a word live phrase.

The N-gram method is based on the hidden Markov chain assumption that the text information at the current position is only relevant to a finite number of text units before that position, without having to take all the previous contents into account [20]. According to such an assumption, a large-scale reduction of the training data can be achieved, thus improving the training efficiency of the model. In the actual training data construction process, not only the content before the specified position will be considered but also the text data after that position, and the window sizes before and after will be kept consistent. In the N-gram method, the size of the window N is artificially set and is, therefore, one of the hyperparameters of the model. For general text data, the best empirical value for the window size is 5, i.e., the information of 5 text units before and after the specified position is considered.

The process of constructing the model training data in this paper uses the 5-gram method, in which the word vector of the specified text unit generated using the method in the previous subsection is used as the sample input feature, and the word vector of the text unit whose distance from the unit is within the window size (5) is used as the label of the sample to construct the model training data. In this subsection, to provide a brief explanation of the algorithm principle, only the first sentence of the sample text data is used as an example, and a 1-gram is used for the construction of the training

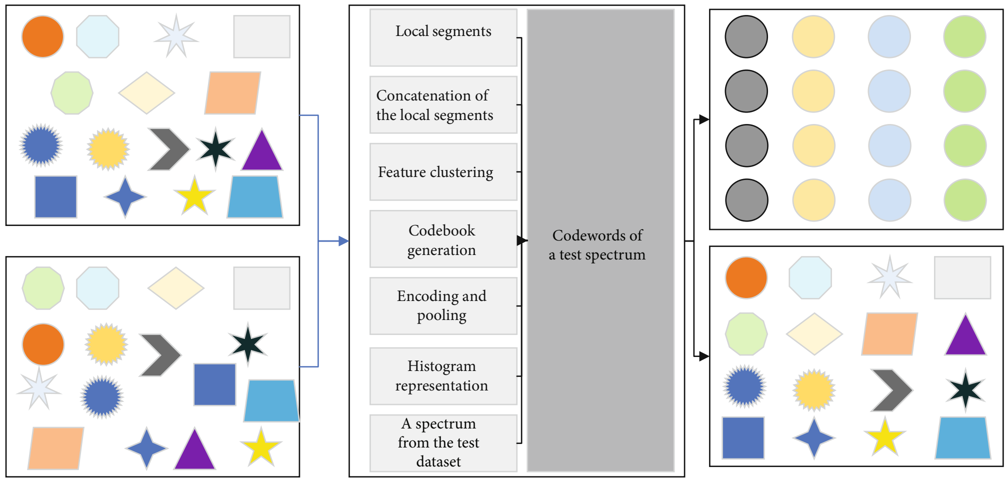

FIGURE 1: The framework of the bag-of-words model.

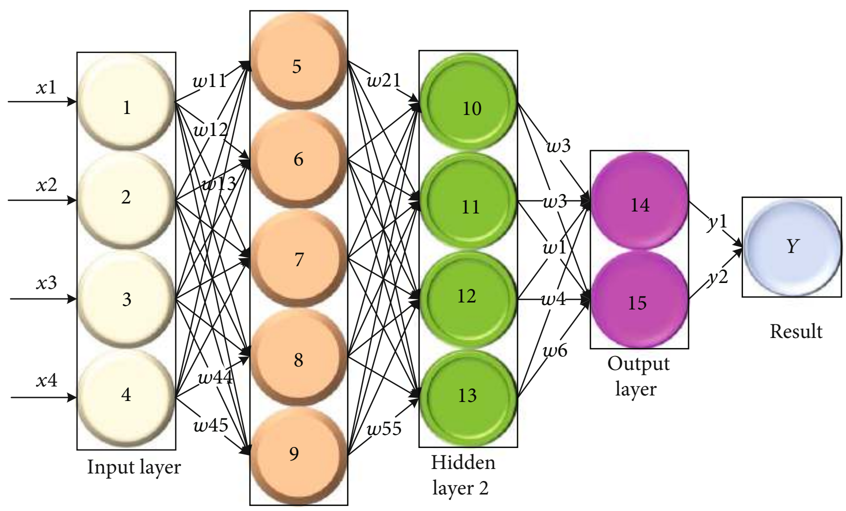

FIGURE 2: Structural diagram of fully connected neural network.

data, and the construction process and the generated results are shown in Figure 3.

The advantages of converting computer-unreadable text data into computer-readable digital information are mainly easier to process and analyze. According to the model of this paper, the purpose of constructing document feature engineering in terms of sentences is achieved. The model combines the principles of artificial neural networks and their algorithms to achieve feature extraction of documents by creating word vectors for independent words. The advantage of the proposed method in this paper is that it not only preserves the frequency information of text units in the document but also combines the location information between different text units, thus enabling the model to show a better fit when dealing with sequence-type text data. The document feature extraction work is the prerequisite and foundation of the higher-order text mining task, and the process converts computer-unreadable text-based data into computer-readable numerical information. The construction of document feature engineering lays the foundation for the key information extraction based on text distance and the intelligent recognition algorithm in the later paper.

3.2. Parallel Projection and Regional Extension ELT Design. The smart classroom environment of this educational experiment integrates an interactive whiteboard, computer, projector, tablet PC, etc., where the tablet PC is one for each student (in line with the list of devices recommended by the North Teachers' University for the smart classroom environment). The interactive whiteboard is integrated with the interactive classroom platform, which contains a variety of common teaching toolkits, such as graphing tools (e.g., straightedge, semicircle, and coordinate system), brush tools, and baseboard tools (e.g., field grid, essay grid, and English alphabet grid), and has real-time synchronized screen dynamic operation, grouping, review, mutual evaluation, quizzing (all, group), randomly naming answers, voting, and student explanation. It also has functions such as sending and receiving assignments and recording screens. When the teacher initiates these learning activities, the students

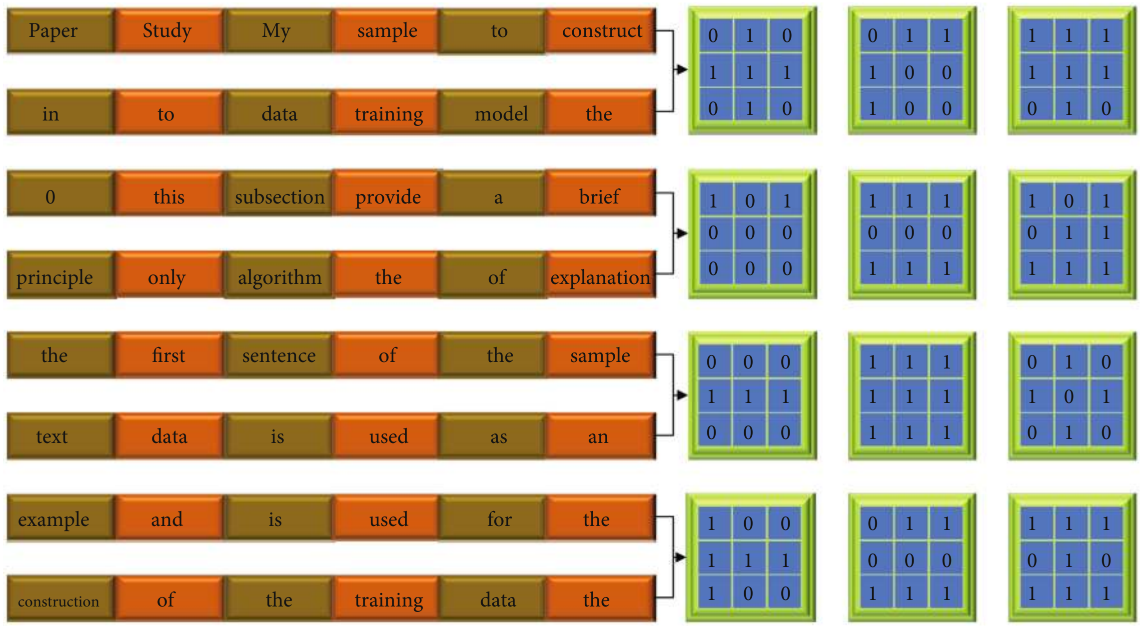

Figure 3: Construction process of training data.

respond on the tablet. These interactive features enable students to engage deeply in classroom learning, and in this experiment, they will also be used to support deep learning strategies such as lecture (both teacher-led and student-led), discussion, presentation, mutual evaluation, and critique [21]. Students can study at various kinds of free time. In addition, the interactive classroom platform can record students' answer processes to subjective questions, which is very helpful for in-depth analysis of students' cognitive and thinking processes. During the experiment, they will be used as materials for teachers or students to discuss, interassess, and critique.

In addition, students' learning time, learning progress, self-study and participation, microvideo viewing, cooperative and mutual learning questioning, assessment of learning, and reflection in this smart classroom can be recorded, analyzed, and fed back (mainly through the Creative Learning Platform). These data are presented to teachers and students in the form of visual reports, which provide them with the opportunity to grasp students' learning status and the situation in real time and provide data sources for teachers of this experiment to monitor students' learning status and achieve teacher data inspiration in decision flexibility. Compared to traditional machine learning algorithms, neural networks usually require more data, at least thousands of millions of labeled samples. In contrast, many machine learning problems can be solved well with less data if other algorithms are used. In addition, this smart classroom is equipped with six general whiteboards, two on either side of the interactive whiteboard (front side of the classroom) and the remaining four on both sides of the classroom, which provides conditions for this experiment to conduct multiple groups at the same time to carry out effective deep learning activities such as collaboration, inquiry, discussion, and presentation.

During class, teachers were asked to use an interactive whiteboard (integrated with the Interactive Classroom Platform) for instruction, and students in both classes were asked to use personal tablets (integrated with the Creative Learning Platform) for learning. In both the experimental and control classes, the teacher facilitated student lectures (both teacher-led and student-led), discussions, presentations, mutual assessments, and reviews through the interactive classroom platform, and both classes provided approximately the same amount of time for students to conduct their investigations. The content is sent from the interactive classroom platform to the students' tablets [22]. During this time, students are supported by their tablets and the four whiteboards on either side of the classroom to engage in independent learning, group inquiry, collaborative work, and discussion and sharing. During this time, teachers act as supervisors, monitoring students' learning by walking around and viewing their learning activities and progress data in the Creativity Academy platform, determining whether individual students/groups need individual prompts or tutoring, moral support, and more difficult/easy learning tasks, and reviewing the appropriateness of the personalized learning tasks pushed by the platform. In addition, the teacher also acts as a guide and facilitator, providing tips, guidance, encouragement, or assistance to students with these needs, either through pushing them through the platform or by visiting them for consultation. This "interactive" flexible teaching model is shown in Figure 4.

According to the above teaching model, the teacher completes the deep learning teaching design in the Creative Learning Platform, in which the control class follows the original design and the experimental class follows the requirements of the deep learning sheet (see the subsection on intelligent classroom environment design) so that the concept of the deep learning sheet is integrated into the Creative Learning Platform of the experimental class, and it will play the role of a learning scaffold in the student learning process. This educational experiment is an exploration of

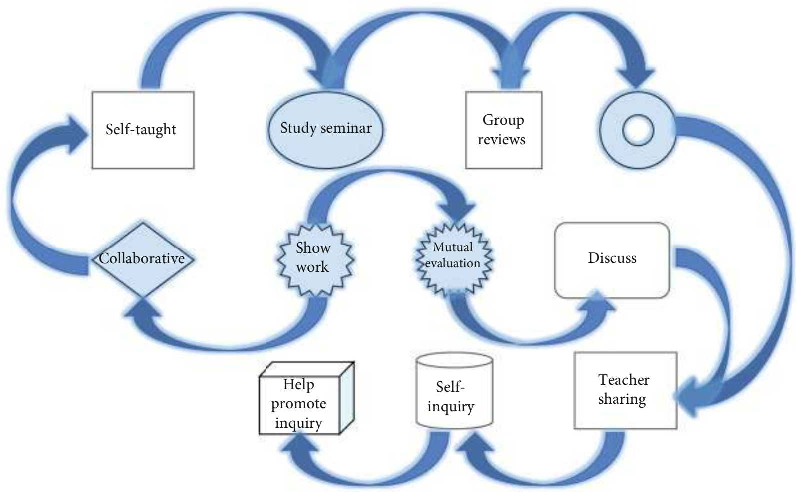

FIGURE 4: "Interactive" flexible teaching model.

the role of the scaffold, the learning sheet, in facilitating deep learning. Among other things, the SCS is an assessment of the quality of the smart classroom from the perspective of students' perceptions. This is necessary because even if the hardware and software of the smart classroom environment are the same, students' perceptions of it will be different, and such perceptions will influence the way students participate in the classroom and the learning strategies they adopt, thus directly or indirectly affecting their deep learning outcomes. This is because traditional text similarity algorithms, which focus on the similarity calculation of the text itself, therefore require a lot of normalization processes, such as normalization of numbers and Arabic numerals, and normalization of English units. Afterward, each student in both classes learned in the smart classroom environment of their respective classes. In both the experimental and control classes, the teacher presented the same learning objectives and topics and suggested reference materials in advance on the interactive whiteboard. The teacher briefed the students on these objectives, topics, and materials before the class began and followed the requirements in the Classroom Design above [23]. At the beginning of each independent inquiry, the teacher explicitly informed them that they could learn in any way they liked, but they could not leave the classroom, and students in the pilot class were additionally informed that they could complete the tasks on the in-depth learning sheet in the order they preferred (only those tasks required for this independent inquiry). Students were required to mark each task they completed in the learning process map (with frequent reminders from the teacher) to document their learning path.

The data collected by the SCS scale were analyzed by independent sample t-test and the results are shown in Figure 5. Overall, students in the experimental class perceived a higher level of quality in the smart classroom ($M = 3.4884$, $\mathrm{SD} = 0.72121$) than those in the control class ($M = 3.4703$, $\mathrm{SD} = 0.68382$). A closer look at the results of the ten-factor test shows (it is feasible to delve further into the dimensions of factors and domains2) that students in both classes perceived their smart classroom environment to be not significantly different across the ten-factor dimensions (see Figure 5).

The key information extraction based on the text ranking algorithm achieves the purpose of reducing the document content and lays the foundation for the fuzzy identification of the content categories of academic documents in this subsection. The parsing of text data content can be generally divided into precise recognition and fuzzy recognition. Precise recognition is mainly applied to tasks with small text data size (generally not more than a few tens of characters) and fixed patterns of content to be recognized; fuzzy recognition is mainly applied to tasks with medium or large text data size and no fixed patterns of content to be recognized. For the research object and the research objective of this paper, the fuzzy recognition method based on edit distance will be used to intelligently identify the content categories of academic documents.

4. Analysis of Results

4.1. Analysis of Identification Results. Different document feature engineering models will be used to identify the filtered document dataset item by item according to the content categories of academic literature on cutting layout issues, and the text classification model evaluation system will be used to evaluate the test results and the model performance related to them. First, an ANN-based document feature model is used to identify the target type (O), and the results are shown in Figure 6 for journal 1 as an example (see Figure 6).

To evaluate the quality of the extracted relations, two sets of tests were conducted in this set of experiments: accuracy test and coverage test. To evaluate the accuracy of the extracted relations, we refer to the evaluation method of the YAGO system, and for each relation category, we randomly sample 200 relation tuples and check their accuracy manually. The coverage test is used to measure whether the extracted relations exist in the existing English knowledge graph. If not, it indicates that PNRE can extract new relations that cannot be covered by existing methods to complement the existing English knowledge graph. Within Broad point,

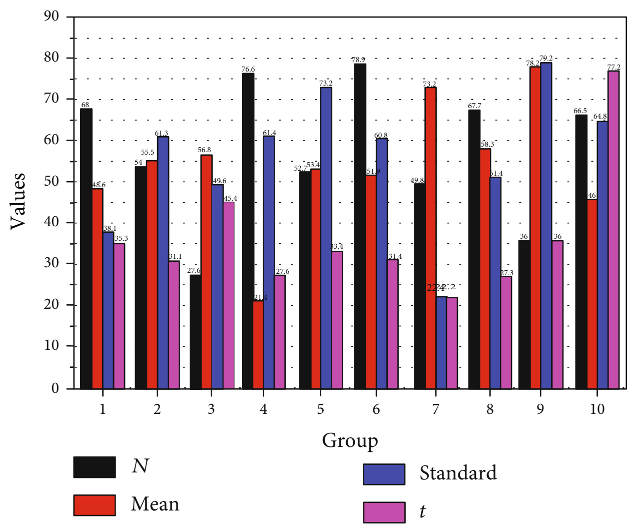

Figure 5: Descriptive statistics and independent sample t-test results.

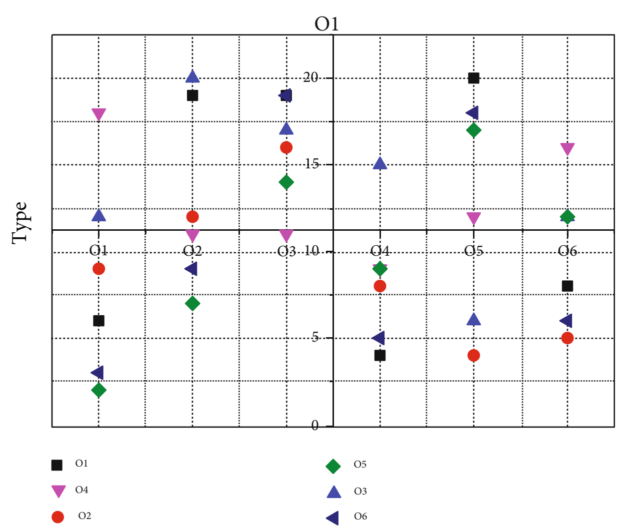

Figure 6: Confusion matrix of test results.

topic model has been successfully applied in many aspects, such as text classification features, relevance calculation, ctr prediction, precise ad targeting, and matrix decomposition. Specifically, based on the topic model, the topic distribution of texts and users can be calculated and used as features of pact and relevance, and it can also be used as a matrix decomposition method for dimensionality reduction, recommendation, etc.

When O1 and O2 are used as positive classes, all the performance indexes achieve good results, which indicates that the proposed method has a better recognition effect for the literature whose objective type is input minimization and output maximization; when O3 is used as a positive class, all the other three indexes achieve poor results except accuracy, which indicates that the proposed method has poor recognition effect for the literature whose objective type is multiobjective planning. To obtain the overall recognition performance of this paper's model for the target constraint, a term to be recognized, the weighted average of each subclass is taken as the result, and the results show that this paper's method shows good overall performance for the recognition of the target constraint in journal 1. Following the same method, the target constraint recognition is tested for the literature in the remaining journals, and the results are shown in Figure 7.

The proposed method in this paper shows good performance in the accuracy of target constraint recognition of literature, and the overall recognition accuracy of each database and each journal is above 70%. When the journals are used as the test unit, the performance of the method in this paper is poor in recognizing the target constraints of the documents published in journals 3, 4, and 5. The worst results were found in journal 3, with the lowest values of 57.14%, 57.94%, and 55.37%, respectively. When the database is used as the test unit, the proposed method shows good recognition performance in the task of document target constraint recognition, and the results of each performance index of each database are above 75.00%, and most of the indexes are above 80.00%, which proves the effectiveness and feasibility of this method.

To prove that the research framework of this paper has the same effectiveness for the remaining document recognition tasks, this paper completed the testing work for all recognition tasks and produced the results. In this subsection, the recognition test results of the problem dimension and set features are shown, and the results are analyzed accordingly. From the above comparison results, due to the corresponding improvement of the text feature extraction model, the overall research scheme proposed in this paper achieves better results in each recognition task, and the weighted average value of the $F1$ value of each database is above 70%, with the lowest being 70.17% and the highest being 86.78%, and all of them are significantly higher than the bag-of-words model and TF-IDF model, and the comprehensive acceptability. The overall acceptability was 81.19%, which reached the research goal of this paper.

4.2. Analysis of Teaching Results. According to the concept of the R-SPQ-2F scale, this effect of deep learning sheets on deep and superficial learning equations stems from both strategy means and motivation. As can be seen in Figure 8, the difference between the experimental and control classes in terms of deep strategy means (DS) was significant and had a moderate treatment effect level, and the difference in terms of deep motivation (DM) was also significant and close to a moderate effect level. Although the difference in the superficial learning strategy between these two classes was not significant, the differences between its two components were not insignificant; specifically, the difference between the experimental and control classes in the superficial strategy means (SS) was 0.892, which was not significant; however, the difference in superficial motivation (SM), which was significant $(F(1, 84) = 4.470,\ p = 0.037 < 0.05)$, had an actual effect value of 0.051, which was at the small effect level (see Figure 8).

The automatic construction of ontologies by libraries will save a lot of social costs and will be the focus of ontology

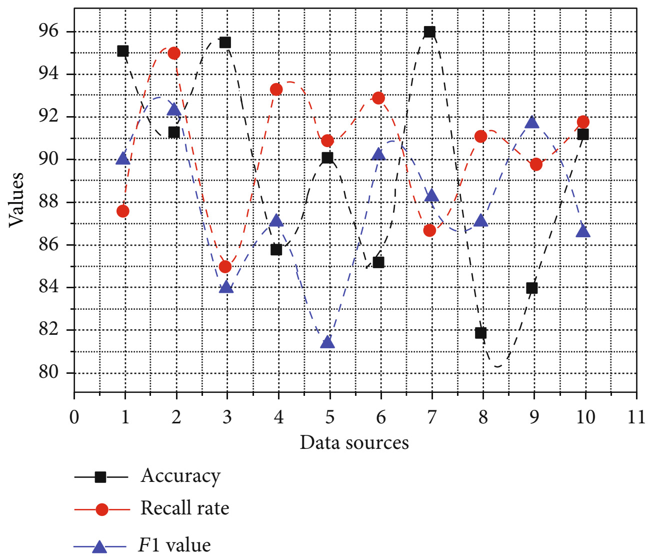

Figure 7: Target type recognition test results.

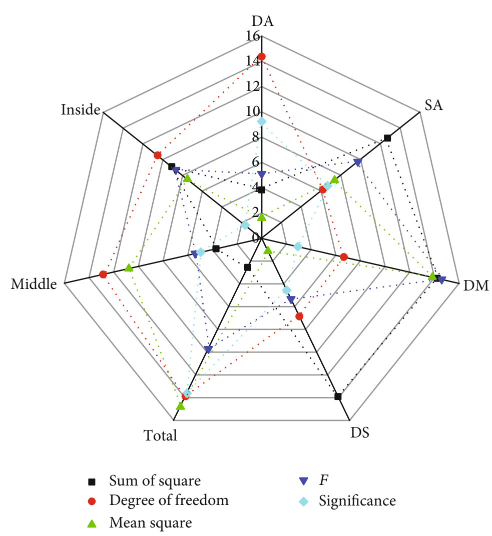

Figure 8: Results of one-way ANOVA for learning strategy.

construction now and in the future. In this paper, we summarize the mainstream ontology construction methods in the world today. In this paper, we summarize the mainstream automatic ontology construction methods in the world today and the main development directions of future ontology construction technologies are presented. To exclude the influence of prior creative experiences on students' creativity development, covariates were used as covariates in this substudy for the analysis of covariance. Because the K-DOCS scale, which is used to capture prior creative experiences, and the CDWR scale, which is used to assess creativity levels, use different scoring systems (i.e., they have different scales), they need to be standardized before the covariance analysis. Given the inability of linear transformations to eliminate the effect of the scale on covariance, this substudy used Z-scores to standardize the two types of data.

Due to the relatively small sample size of the experiment, the residuals of the previously created empirical data needed to be tested for chi-square and normal distribution after standardization to determine whether analysis of covariance could be performed. The Levene chi-square test was calculated to be $F(1, 84) = 1.794$, $p = 0.184$, and the Shapiro-Wilk test for normal distribution was $p = 0.188$, both of which were not significantly different ($p > 0.05$), indicating that the experimental and control samples had homogeneous variances and were normally distributed, thus allowing for covariance between the previously created empirical data and the test data. In addition, due to the small sample size of this experiment, the actual effect sizes of the deep learning sheets need to be reported when reporting the results of the covariance analysis. With the support of the smart classroom environment, the deep learning sheets promote the development of deep learning skills (e.g., creativity) by increasing students' deep engagement in learning and guiding them to adopt deep learning strategies. Following this logic, this substudy constructs regression models with the smart classroom, classroom engagement, level of mindful flow, and deep learning strategies, which are the focus of the deep learning sheets, as independent variables, and the level of creativity as the dependent variable. Given the influence of prior creative experiences on creativity development, the present substudy also used them as independent variables in the multiple regression analysis.

5. Conclusion

Document key information extraction based on text ranking algorithm and text content fuzzy recognition based on edit distance is proposed. Due to a large amount of text data in academic literature, the adoption of the global matching recognition method will cause the problems of long solving time and low recognition efficiency, so this paper adopts a step-by-step recognition method: firstly, the key information is extracted from the original document, and secondly, the extracted information is matched for recognition of related items. Through the above method, the research objective of this paper is finally achieved. The overall research framework constructed in this paper was tested on the document dataset, the test results were analyzed in detail, and the analysis results verified the effectiveness of the method in this paper. Meanwhile, to make the research work of this paper have certain practicality, a prototype system of intelligent mining of academic documents is developed to realize the modularization, integration, and automation of the research methods in each stage.

Data Availability

All information is within the paper.

Conflicts of Interest

No competing interests exist concerning this study.

References

[1] S. Puri and S. P. Singh, “A fuzzy matching based image classification system for printed and handwritten text documents,” *Journal of Information Technology Research*, vol. 13, no. 2, pp. 155–194, 2020.

[2] F. Ren and Y. Bao, “A review on human-computer interaction and intelligent robots,” *International Journal of Information Technology & Decision Making*, vol. 19, no. 1, pp. 5–47, 2020.

[3] W. Liu, H. L. Ma, and A. Walsh, “Advance in photonic crystal solar cells,” *Renewable and Sustainable Energy Reviews*, vol. 116, p. 109436, 2019.

[4] X. Zhang, C. Zang, H. L. Ma, and Z. J. Wang, “Study on removing calcium carbonate plug from near wellbore by high-power ultrasonic treatment,” *Ultrasonics Sonochemistry*, vol. 62, p. 104515, 2020.

[5] H. L. Ma, X. Zhang, F. F. Ju, and S. B. Tsai, “A study on curing kinetics of nano-phase modified epoxy resin,” *Scientific Reports*, vol. 8, no. 1, p. 3045, 2018.

[6] K. I. Frank, M. Hibbard, M. Shucksmith et al., “Comparative rural planning cultures,” *Planning Theory & Practice*, vol. 21, no. 5, pp. 769–795, 2020.

[7] A. Charitopoulos, M. Rangoussi, and D. Koulouriotis, “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*, vol. 30, no. 3, pp. 371–430, 2020.

[8] E. Raveh, Y. Ofek, R. Bekkerman, and H. Cohen, “Applying big data visualization to detect trends in 30 years of performance reports,” *Evaluation*, vol. 26, no. 4, pp. 516–540, 2020.

[9] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, “Effects of residence time and heating rate on gasification of petroleum residue,” *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.

[10] H. L. Ma and S. B. Tsai, “Design of research on performance of a new iridium coordination compound for the detection of Hg^{2+},” *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, p. 1232, 2017.

[11] L. Y. Mo, W. H. Z. Sun, S. Jiang et al., “Removal of colloidal precipitation plugging with high-power ultrasound,” *Ultrasonics Sonochemistry*, vol. 69, p. 105259, 2020.

[12] D. Xu and H. Ma, “Degradation of rhodamine B in water by ultrasound-assisted TiO_2 photocatalysis,” *Journal of Cleaner Production*, vol. 313, article 127758, 2021.

[13] M. Middell, “From universal history to transregional perspectives: the challenge of the cultural and spatial turn to world and global history in the 1970s and today,” *Cultural History*, vol. 9, no. 2, pp. 241–264, 2020.

[14] D. Levi and P. Tucker, “‘J after J. Ruskin’: line in the art teaching of John Ruskin and Ebenezer Cooke,” *Journal of Art Historiography*, vol. 22, pp. 1–16, 2020.

[15] D. Gao, Y. Liu, Z. Guo et al., “A study on optimization of CBM water drainage by well-test deconvolution in the early development stage,” *Water*, vol. 10, no. 7, p. 929, 2018.

[16] S. B. Tsai and H. Ma, “A research on preparation and application of the monolithic catalyst with interconnecting pore structure,” *Scientific Reports*, vol. 8, no. 1, p. 16605, 2018.

[17] J. Xie and H. Ma, “Application of improved APO algorithm in vulnerability assessment and reconstruction of microgrid,” *IOP Conference Series: Earth and Environmental Science*, vol. 108, no. 5, article 052109, 2018.

[18] K. Chakraborty, S. Bhattacharyya, and R. Bag, “A survey of sentiment analysis from social media data,” *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 450–464, 2020.

[19] J. Prüfer and P. Prüfer, “Data science for entrepreneurship research: studying demand dynamics for entrepreneurial skills in the Netherlands,” *Small Business Economics*, vol. 55, no. 3, pp. 651–672, 2020.

[20] T. Chaiechi, “Sustainable tropical cities: a scoping review of multidisciplinary methods for urban planning,” *eTropic: Electronic Journal of Studies in the Tropics*, vol. 19, no. 2, pp. 25–51, 2020.

[21] R. S. Liévanos, “Racialised uneven development and multiple exposure: sea-level rise and high-risk neighbourhoods in Stockton, CA,” *Cambridge Journal of Regions, Economy and Society*, vol. 13, no. 2, pp. 381–404, 2020.

[22] Z. K. Al, “Military and politics in Syria (1946-1963),” *Syria Studies*, vol. 12, no. 1, pp. 52–202, 2020.