Text Analysis of Teaching Evaluation Based on Machine Learning

Xin Hu [1+], Yanfei Yang[1], Xinlin Wu[2], Yan Li [1]

[1] Basic Department, Air Force Early Warning Academy, Wuhan, China

[2] Teaching and Research Guarantee Center, Air Force Early Warning Academy, Wuhan, China

Abstract. The traditional teaching quality evaluation methods of colleges and universities have been unable to meet the informatization and modern teaching modes in terms of accuracy and implementation efficiency. Therefore, for the problem of evaluating teaching quality in colleges and universities, this paper proposes a sentiment analysis method for teaching evaluation text based on machine learning. This article establishes a teaching evaluation feature dictionary, reduces the dimensionality of attribute features through mining analysis, and extracts the features most relevant to teacher evaluation. In addition, the support vector machine algorithms with linear kernel, polynomial kernel and radial basis kernel are used to classify the sentiment of the text data in teaching evaluation to judge the sentiment tendency of evaluation. The experimental results show that the support vector machine radial basis kernel has the best effect on the classification of teaching evaluation text data, which can enable teachers to accurately obtain feedback information for evaluation, so that they can adjust their teaching work in time to assist teaching decisions and improve teaching quality.

Keywords: machine learning; teaching evaluation; support vector machine; text analysis

1. Introduction

In recent years, the teaching quality of teachers in various schools has attracted great attention of the society, and the scientific, objective and fair evaluation of a teacher's teaching quality has become a hot topic. In terms of accuracy and implementation efficiency, the traditional teaching quality evaluation method in colleges and universities has been unable to meet the requirements of information and modern teaching mode [1]. At present, many colleges and universities are establishing and improving teachers' teaching quality evaluation methods, trying to use modern teaching evaluation methods including AHP, multiple linear regression, and neural network models. Among them, the fuzzy comprehensive evaluation model [2] constructed by the analytic hierarchy process can remove subjective components as much as possible on the basis of expert knowledge and subjective experience, and use reasonable and logical mathematical methods to reasonably determine the weight of evaluation indicators. The teaching quality evaluation model [3] based on BP neural network can fit the training data well, has good prediction effect, and can realize a comprehensive teaching quality evaluation system. However, the neural network is prone to local extremum and cannot obtain the global optimal value, which leads to large error between the teaching evaluation results and the actual results and relatively long time consuming.

These existing evaluation methods are mainly based on the scoring mechanism of objective evaluation [4], and the text information in the subjective evaluation as an important content of teaching evaluation has not been effectively used. The sentiment analysis [5] technology in text mining can effectively summarize the text information of teaching evaluation, so that teachers can accurately obtain the evaluation opinions on the characteristics of each evaluation index of teaching to assist teaching decision-making and improve the quality of higher education teaching. This article focuses on the use of text mining technology, especially

---

+ Corresponding author. Tel.: +13871494002; fax: +027 85965633

E-mail address: sstcookie@163.com.

short text sentiment analysis technology based on machine learning, through the establishment of a professional dictionary for teaching evaluation, and the use of word2vec language model [6], combined with support vector machine [7] algorithm to classify text evaluation information, to provide decision support for teaching management.

2. Text Analysis Algorithm Based on Machine Learning

Machine learning is a technology that uses computers to simulate human learning behaviors, so that machines also have cognitive and comprehension abilities. From the learning method, machine learning can be mainly divided into supervised learning and unsupervised learning. The main difference lies in whether the training data set tells the machine the correct answer. The training set that gives the correct answer is called supervised learning, which is mainly used for classification prediction and regression prediction. Supervised learning techniques mainly include logistic regression, decision tree algorithm, SVM algorithm and so on. If the training set does not specify the correct answer, it is unsupervised learning. Common algorithms include k-means clustering analysis, system clustering and so on.

2.1. Support Vector Machine

In view of the good performance of support vector machines in classification and regression estimation, this paper chooses to use support vector machine algorithm as a tool for text analysis. Support vector machine was proposed by Cortes et al [8] based on statistical theory in 1995. It is a small sample learning method. Its main idea is to find a hyperplane as a two-class segmentation in a high-dimensional complex space to ensure the smallest classification error. SVM uses kernel functions to map low-dimensional nonlinear data to high-dimensional feature space to become linearly separable, and its application in text classification has achieved good results [9]. In the classification problem, suppose there is a training data set $\{(x_i, y_i)\}_{i=1}^n$, where $x_i \in R^m$ represents m-dimensional data characteristics, $y_i \in \{-1, +1\}$ represents the data category, and n is the number of data. SVM solves two classification problems by minimizing the following optimization problems.

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^n \xi_i \tag{1}$$

$$\text{s.t.} \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \tag{2}$$

$$\xi_i \geq 0, i = 1, \cdots, n. \tag{3}$$

where $w \in R^m$ is the normal vector of the hyperplane, b is the threshold, ξ_i is the introduced relaxation variable, and $C > 0$ is the penalty factor. Using Lagrangian multiplier method and kernel technique, the dual programming of the optimization problem is obtained:

$$\max \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j k(x_i, x_j) \tag{4}$$

$$\text{s.t.} \quad \sum_{i=1}^n \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C, i = 1, \cdots, n \tag{5}$$

where α_i is the Lagrangian multiplier, and the sample component corresponding to α_i whose value is greater than 0 is the support vector. The decision function of SVM can be expressed as:

$$f(x) = \text{sgn}(\sum_{i=1}^n \alpha_i y_i k(x_i, x) + b) \tag{6}$$

where $k(x_i, x)$ is the kernel function that satisfies the Mercer condition, sgn represents a binary function, and its value is $\{-1, +1\}$.

2.2. Kernel Function and its Parameter Selection

In SVM, the processing method for the case of nonlinear data is to select a kernel function and solve the problem of the linear inseparability of unstructured data in the original space by mapping the data to a high-dimensional space. The commonly used kernel functions mainly include linear kernel, polynomial kernel,

radial basis kernel and so on. This paper selects the linear kernel, polynomial kernel and radial basis kernel to classify the sentiment evaluation text data and compare their classification performance.

The linear kernel function is mainly used in the case of linear separability. It finds the optimal linear classifier in the original space and has the advantages of few parameters and fast speed. Its kernel function is expressed as:

$$k(x_i, x) = (x \cdot x_i) \tag{7}$$

The polynomial kernel function is a global kernel function, allowing data points that are far apart to have an influence on the value of the kernel function. The kernel function is expressed as:

$$k(x_i, x) = q[(x \cdot x_i) + D], q > 0 \tag{8}$$

where D is the bias coefficient, which is a constant, and q is the power exponent of the polynomial kernel function. The larger the parameter q is, the higher the dimension of the mapping is, and the greater the computation will be. When q is too large, the learning complexity will be too high, and the "over-fitting phenomenon" is prone to occur.

The form of radial basis kernel function is:

$$k(x_i, x) = \exp(-\frac{\|x - x_i\|^2}{2\delta^2}), \delta > 0 \tag{9}$$

where δ is the bandwidth of the radial basic kernel, which determines the scope of the function. Gaussian radial basis kernel function has good anti-interference ability to the noise points in the data, but due to its strong locality, its anti-interference ability weakens with the increase of parameter δ.

Kernel functions have their own specific parameters and must be adjusted to achieve the optimal performance of classification. This paper uses the K-fold cross validation method [10] to train and adjust the test data and select the optimal parameters. In this method, the test data is divided into K subsamples, one of which is reserved as test data, and the remaining $K-1$ samples are used as training data. Each subsample is verified once, and the cross validation is repeated iteratively K times, and the set of parameters with the highest classification accuracy is selected as the optimal parameter. In this paper, $K = 10$ was selected.

3. Sentiment Analysis Method of Teaching Evaluation Text Based on Machine Learning

The number of words evaluated by supervisory experts on teachers' texts is mostly between 25 and 150 words, with sparse content and high feature dimensions. They are short texts and are suitable for judgment by text sentiment analysis. This article first establishes the teaching evaluation domain dictionary, preprocesses the evaluation data, then uses SVM to analyze the emotional polarity of the evaluation data, and finally outputs and evaluates the classification results.

3.1. Teaching Evaluation Feature Dictionary

Dictionary is an important tool and basic task of text sentiment analysis. The quality of the dictionary directly affects the results of sentiment analysis. So far, in Chinese text sentiment analysis, there is no complete dictionary that can cover all vocabulary. In the evaluation of Chinese teaching, the phenomenon that many vocabulary used in teaching evaluation cannot be retrieved by the evaluator often results in incomplete teaching evaluation. At the same time, different evaluators often use different words to describe the same attribute characteristics, which causes semantic confusion. In view of the specialty of Chinese teaching evaluation, this article uses the "Harbin Institute of Technology Information Retrieval Research Center Synonym Cilin Extended Version"as the basic dictionary, and uses Bi-Gram, Tri-Gram language models and manual annotation methods to establish the teaching evaluation feature dictionary, so as to improve the accuracy and comprehensiveness of vocabulary retrieval.

The teaching evaluation feature dictionary established in this paper contains two parts: the first part is to build a dictionary of characteristic words for teaching evaluation that can not be retrieved in the basic dictionary. The characteristic words include teaching posture, head-up rate, class design, blackboard writing

and case teaching; In the second part, a thesaurus of teaching evaluation is established for synonyms. See Table 1 for examples.

Table 1: Examples of Synonym Dictionary for Teaching Evaluation

Original word	Replaced word
Some students doze off	Enhance class interaction
Class atmosphere is not active enough	Enhance class interaction
Not paying enough attention to students	Enhance class interaction

3.2. Feature Selection

The purpose of feature item selection is to select from the preliminary feature set some highly contributed words that can express the actual content of the comment as the feature item set, the core of which is the calculation of feature item weights. The calculation method used in this paper is the TFIDF method: the TFIDF value is used to measure the ability of a word or phrase to represent a text, which is affected by the word frequency and the inverse document frequency (IDF).

In order to facilitate the selection of feature items, the original text is first preprocessed. After a series of preprocessing including the removal of stop words and part-of-speech tagging, the word2vec language model is used to train word vectors, and each word is mapped into a k-dimensional vector. Since the attributes of teaching evaluation are often related to nouns, the nouns and noun phrases are selected as candidates for evaluation. The nouns and noun phrases with relatively few occurrences (word frequency less than 2) in the corpus are filtered out, and the remaining nouns and noun phrases are calculated Cosine similarity. The K-means algorithm is used to cluster nouns and noun phrases with high similarity into the same attribute feature cluster, and then the final attribute feature cluster is determined by manual screening as the evaluation object. Finally the emotional tendency of each attribute feature word is marked and divided into positive and negative sentiment tendencies, and the weight of each attribute characteristic word is defined by TFIDF.

3.3. Evaluation Index

In the evaluation of sentiment analysis model, the evaluation index usually adopts accuracy rate P, recall rate R and comprehensive classification rate $F1$. The calculation formulas are:

$$P = \frac{TP}{TP+FP}, \quad R = \frac{TP}{TP+FN}, \quad F1 = \frac{2 \times P \times R}{P+R}. \tag{10}$$

This paper uses macro average accuracy rate (P_{macro}), macro average recall rate (R_{macro}) and macro average comprehensive classification rate ($F1_{macro}$) to evaluate the overall classification performance of Chinese sentiment analysis methods. The corresponding calculation formulas are:

$$P_{macro} = \frac{1}{n}\sum_{i=1}^{n} P_i, \quad R_{macro} = \frac{1}{n}\sum_{i=1}^{n} R_i, \quad F1_{macro} = \frac{2 \times P_{macro} \times R_{macro}}{P_{macro} + R_{macro}}. \tag{11}$$

4. Empirical Analysis

4.1. Data Sources and Experimental Procedures

The data used in this article comes from the Chinese text evaluation data of teachers by the supervising experts in the teaching management system of Air Force Early Warning Academy, from which 581 evaluation data are extracted, 15 irrelevant data are deleted, and the remaining 566 data are used as corpus. 100 of them were selected as test data and the remaining 466 as training data. The whole experiment steps are as follows:

Step 1. Data input. First, extract the evaluation data of the Chinese short texts of the teachers by the supervising experts, filter the texts, and remove the irrelevant texts (such as "adjust the class today", etc.) to generate a review corpus.

Step 2. Data preprocessing. Word segmentation is done through the Chinese word segmentation library "jieba" of Python. In terms of dictionaries, the "Harbin Institute of Technology Information Retrieval Research Center Synonym Cilin Extended Version" and the teaching evaluation feature dictionary established in this article are used to remove stop words and perform part-of-speech tagging (individual

special words are marked manually). The processed data is in the form of "word + part of speech" with "bar" as the unit. The words are separated by spaces and saved as TXT format.

Step 3. Feature extraction. Use the method in Section 3.2 to extract attribute feature words and define feature weights. The teaching evaluation data is represented as a vector in a high-dimensional space, one vector corresponds to one evaluation, and each dimension of the vector corresponds to one attribute characteristic of the evaluation.

Step 4. Classifier classification. The SVM algorithm of linear kernel, polynomial kernel and radial basis kernel are applied to train the emotion classifier and classify the test data.

Step 5. Performance evaluation. Output classification results and analyze them.

4.2. Result Analysis

The kernel function uses the linear kernel, polynomial kernel and radial basis kernel SVM as the classifier respectively, and uses the K-fold cross validation method to select the optimal parameters and emotionally classify the test data to compare their classification performance. The experimental results obtained are shown in Table 2.

Table 2: Comparison of classification results under different kernel functions

Kernel function	Optimal parameters	P_{macro}	R_{macro}	$F1_{macro}$
Linear kernel	$C = 15$	83.45	90.12	87.11
Polynomial kernel	$q = 3, C = 9$	92.08	85.91	88.04
radial basis kernel	$\delta = 0.4, C = 5$	91.77	89.04	90.27

It can be seen from Table 2 that when the optimal parameters of the SVM classifier are selected, the macro average classification accuracy of the polynomial kernel and the radial base kernel is relatively equal, and that of the linear kernel is poor; The macro average recognition recall rate of linear kernel and radial basis kernel is high, and that of polynomial kernel is poor. while the radial basis kernel has relatively high macro average accuracy rate and macro average recall rate, the macro average comprehensive classification rate is the highest and the performance reaches the best. Among the three kernel functions, the radial basis kernel has the best effect on the classification performance of teaching evaluation text data, and is relatively stable. Therefore, when using SVM to classify unstructured text data, the kernel function can give priority to radial basis kernel.

5. Conclusion

Teaching evaluation text data is an important content in teaching evaluation. This paper proposes a method of sentiment analysis for teaching evaluation text based on machine learning, establishes a dictionary of teaching evaluation features, and provides a classification basis for sentiment analysis. It evaluates the unstructured text evaluation data provided by the supervision experts, and the support vector machine's sentiment classification performance under the linear kernel, polynomial kernel and radial basis kernel, which effectively solves the problem of nonlinear text classification in teaching evaluation. Because the sparseness of text data in teaching evaluation and the quality of evaluation data will directly affect the performance of sentiment classification, how to better extract the hidden features in teaching evaluation text data, reduce the dependence of classification performance on evaluation data, and optimize classification algorithms are the focus of future research.

6. References

[1] COHEN J, BROWN M. Teaching quality across school settings[J]. New educator,2016,12(2):191-218.

[2] Dong Guoyu, Qi Yingchun. Application of AHP in the evaluation of classroom teaching quality in colleges and universities [J]. China Adult Education, 2017 (9): 53-56

[3] Zheng Yong, Chen Yan. Teaching quality evaluation model of college teachers based on BP neural network [J]. Journal of Chongqing University of Technology, 2015, 29 (1): 85-90.

[4]   Pang Guoying, Liu Jun, Wei Jie. Indicators and weight values of the teaching evaluation system of college teachers [J]. Journal of Military Communications Institute, 2015, 17 (5): 83-86.

[5]   LIU B. Sentiment analysis and opinion mining [M]. *San Francisco：Morgan & Claypool*，2012.

[6]   SU Z, XU H, ZHANG D,et al. Chinese sentiment classification using a neural network tool:word2vec [C]// *Proceedings of International Conference on Multisensor Fusion and Information Integration for Intelligent Systems*. Beijing: IEEE,014:1-6.

[7]   PENG S, HU Q, CHEN Y, et al. Improved support vector machine algorithm for heterogeneous data [J]. Pattern recognition,2015,48(6):2072-2083.

[8]   CORTES C, VAPNIK V N. Support-vector networks[J]. Machine Learning, 1995, 20(3):273-297.

[9]   Wei Jingjing, Wu Xiaoyin. Research and implementation of multi-level sentiment analysis for e-commerce product reviews [J]. Software, 2013, 34 (9): 65-67.

[10]   CORANI G, BENAVOLI A. A Bayesian approach for comparing cross-validated algorithms on multiple data sets [J]. Machine learning, 2015,100(2):285-304.