

Evaluation and Research on Students' Comprehensive Quality Based on Text Mining

Xue Yanru^{1, a}, Huang Lei^{2, b}

¹*School of Economic and Management, Beijing Jiaotong University, Beijing 100044, China.*

²*School of Economic and Management, Beijing Jiaotong University, Beijing 100044, China.*

^a*xueyanrubyy@163.com*

Keywords: Comprehensive Quality; Text Mining; Support Vector Machine; Word Cloud;

Abstract: This paper analyzes the construction principles of the comprehensive quality evaluation index system and constructs a comprehensive quality evaluation index system for college students. It uses text mining technology to analyze the actual interview records, such as text segmentation, clustering and data mining, finally describes the main factors affecting the overall quality of college students.

1. Introduction

With the development of society, the society needs higher requirements for the comprehensive quality of talents. Colleges and universities should strengthen the cultivation of comprehensive quality of college students, optimize the comprehensive quality evaluation system, and pay attention to the development of students. However, in some colleges and universities, the evaluation system of students' comprehensive quality is still not perfect, resulting in the low comprehensive quality of students.

In order to grasp the overall quality of students, this paper bases on the actual interview records, using text segmentation, clustering and mining to analysis the records, and finally obtain the main factors affecting the overall quality of college students, so as to get more recommendations for future guidance.

2. Literature Review

2.1 Students' Comprehensive Quality Evaluation System

Due to the development of big data, technological innovation and social development are closely related to the rapid development of professional skills in various disciplines. Therefore, some colleges and universities are eager for quick success, and inevitably pay too much attention to the cultivation of college students' ability to adapt to future professional careers. For example, most colleges and universities kept the academic evaluation as the main evaluation criteria, not paid enough attention to students' ideological education and mental health. Therefore, the simplification of university graduates' comprehensive quality evaluation index has become a common and primary education in colleges and universities. On the one hand, most colleges and universities have not

paid enough attention to the construction of the comprehensive quality evaluation index system for college graduates; on the other hand, the current comprehensive evaluation system for students lacks orientation, objectivity and comprehensiveness.

The overall quality of college students includes four aspects: ethical, cultural, professional and physical. The higher the degree of combination, the higher the quality. Therefore, many scholars focus the research of the comprehensive quality of students, hoping to find some management methods to improve the overall quality of college students. They grasp the relationship between the comprehensive quality indicators of students and the quality of employment, so that school administrators can better track the development of students, and educate students to improve the overall quality.

2.2 Research of Text Mining Technology

Text mining mainly accomplishes the task of discovering hidden knowledge and patterns from a large number of documents. The objects generally processed are massive, heterogeneous, and distributed documents. The data processed by traditional data mining is structured and stored in a database, and the documents are semi-structured or unstructured. Therefore, the primary problem facing text mining is how to represent text in a computer so that it contains enough information to reflect the characteristics of the text.

Text representation and feature extraction are fundamental issues in the field of text mining. Currently, vector space models are commonly used to generate text vectors to represent unstructured text data. However, the high-dimensional text vector obtained directly by word segmentation and word frequency statistics is represented as text, which not only brings huge computational cost to subsequent tasks such as text comprehension, but also accuracy is affected. Therefore, effective text feature selection and compression methods are necessary for dimensionality reduction. At present, the research on text representation mainly focuses on the text representation model method and feature selection algorithm. The basic unit used to represent text is often referred to as the feature item of the text.

3. Construction of Students' Comprehensive Quality Evaluation System

The comprehensive quality evaluation of college students is based on certain educational values or educational goals, collecting the characterization information in the fields of college students' ideological and moral, academic level, ability literacy and physical and mental conditions, and making value judgments on the comprehensive quality of college students. The evaluation system is based on the existence of market economy value, maximizing the quality of quality education, and achieving the requirements for the comprehensive development of college students and the reform and development of higher education. The optimization of college students' comprehensive quality evaluation system and the allocation of factors can effectively exert the guiding, stimulating, corrective and management roles of evaluation.

3.1 Evaluation indicators

According to the literature research, according to the analytic hierarchy process, the following hierarchical structure of the comprehensive quality evaluation indicators can be constructed as follows:

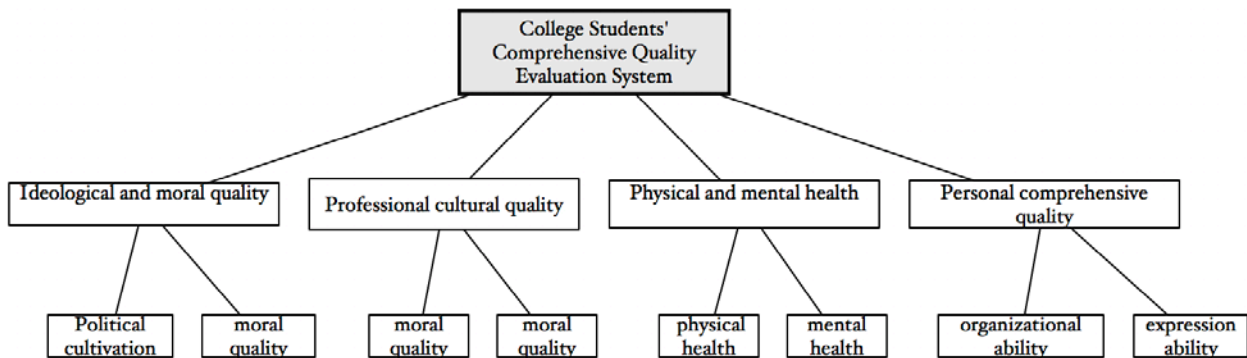


Fig.1 The comprehensive quality evaluation indicators

4. The Application of Text Mining Model

4.1 Construct Chinese word cloud

In order to visually obtain the subject's emotion tendency, we choose the word cloud to analyze.

The word cloud is a visually prominent representation of keywords with high frequency in the text. The color images of various shapes are formed by the rendering of keywords, so that the meaning of the text data can be visually recognized.

The main steps in word cloud are as follows:

(1) Chinese word segmentation: The text data is processed through Jieba word segmentation, mainly for word segmentation and removal of stop words;

(2) Calculate the frequency at which each word appears in the text: a hash table that generates the frequency of the words based on the frequency of occurrence of each word in the text;

(3) Generate words on the word cloud layout according to the corresponding word frequency: mainly use word cloud in python to generate word cloud;

(4) Complete the coloring of the lyrics on the word cloud.

4.3 Divide emotional polarity.
The visitor's personality is more introverted and slightly unstable, interpersonal relationship is tense, usually has anxiety and nervousness, and lacks self-confidence; Learning has not recovered from the state of high school, lack of motivation to learn, and the ability to do hands-on and self-study research is also weak. The treatment of emotions in daily life is not very timely, easy to emotional, affecting daily life.

4.2 Divide emotional polarity

4.2.1 Data preparation and preprocessing

The emotional polarity of the user's question can be obtained from the analysis of the problem set. Mainly through the following to reduce data noise and data size.

The treatment of the student's answer is mainly through the following scheme to reduce data noise and data size.

The first step: data preparation. The main job of this step is to obtain data and various dictionaries, mainly including:

1) Stop word dictionary: This article uses the Chinese stop word list of 1208 stop words published by the Chinese Natural Language Processing Open Platform of the Chinese Academy of Sciences, and removes the stop words in the comments; 2) Positive and negative corpus: This article

uses the hotel commentary corpus derived from Chinese emotion mining, which is 7000 forward and 3,000 negative;3) Verification set: The answer to the student question.

The second step: data preprocessing. The work carried out in this step includes:

(1) Word segmentation and removal of stop words: The word segmentation operation uses the Jieba participle, and the removal of the stop word is based on the list of stop words in step 1. (2) Training word vector: This article uses Word2Vec to convert the corpus into a word vector. (3) Standardization: In order to unify all feature vectors into one magnitude, data needs to be standardized in advance. (4) Dimensionality reduction: According to the PCA results, it is found that the first 100 dimensions can cover more than 95% of the variance, so the dimension of the word vector can be reduced to 100 dimensions;

4.2.2 Building the model

This paper adopts the SVM method. In general, it is a two-class classification model. The basic model is defined as a linear classifier with a large interval on the feature space. The learning strategy is that the interval is larger and can be converted into a solution to a convex quadratic programming problem.

The obtained sentiment analysis results are shown in the figure:

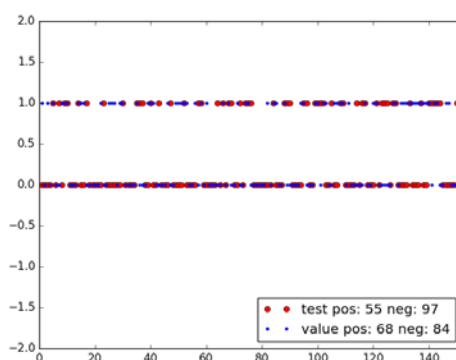


Fig.2 The obtained sentiment analysis results

As can be seen from the above figure:

(1) The above figure shows the test results and actual results of the corpus, and the accuracy rate is around 0.6, indicating that the prediction accuracy of this model needs to be improved;

(2) In addition, since the training set and the test set are from different fields, there is reason to think that the training set is not enough. In the future, the training set can be considered to improve the accuracy.

4.3 Calculating emotional scores

4.3.1 Data preparation and preprocessing

The first step: data preparation. The main job of this step is to obtain data and various dictionaries, including:

(1) Emotional Dictionary: This article uses the Boson NLP Emotional Dictionary, which removes the stop words without emotional color in the dictionary, and excludes the influence of stop words and Chinese part of speech on the accuracy of the model. (2) Negative word dictionary: The appearance of negative words will change the emotional turn of sentences; (3) Degree adverbs dictionary: This article uses the words of CNKI's sentiment analysis. In this dictionary, each degree

adverb has a corresponding emotional score. (4) Stop word dictionary: This paper adopts the Chinese stop word list of 1208 stop words published by the Chinese Natural Language Processing Open Platform of the Chinese Academy of Sciences, and removes the stop words in the comments;(2) The second step: data preprocessing. The work done in this step includes: word segmentation and removal of stop words. The word segmentation operation uses the Jieba participle, and the removal of the stop word is based on the stop word list in step one.

4.3.2 Building the model

Step 1: the previous process completed the word segmentation work, and now we have obtained the dictionary of words after removing the stop words. Through the sentiment dictionary, the negative dictionary and the degree adverb dictionary, the words in the Word dictionary are stored and labeled according to the emotional words, the no-degree and the degree adverb categories, and the emotional word dictionary, the negative word dictionary and the degree adverb dictionary are obtained.

Step 2: Calculate sentence score: Defining an emotional word group, all negative words and degree adverbs between the two emotional words and the latter emotional words of the two emotions constitute an emotional phrase, for example, the comment may appear "not very satisfied", which is not a negative word Very adverbs of degree, satisfaction is an emotional word, then the score of this emotional word phrase is: $\text{score(not)} * 1 * \text{score(very)} * \text{score(satisfaction)}$, where 1 represents 1 negative word, and the score of each word is Available in the dictionary. When calculating the emotional score of a comment, the period is used as a sign to end the sentence. If a text contains multiple sentences, the average of the scores of the emotions of all sentences is calculated.

The radar chart made after the score of the obtained text is as shown below:

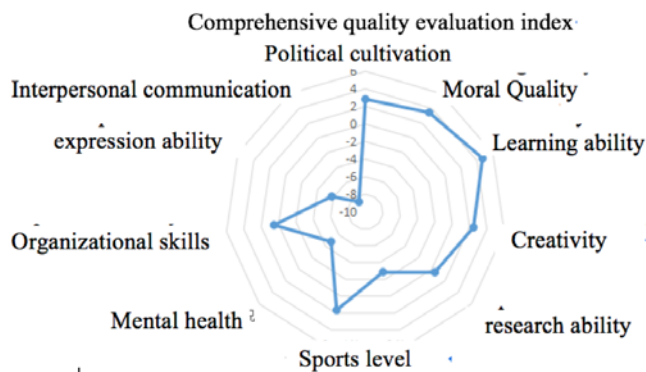


Fig.3 The radar chart made

The student's scores in political accomplishment, moral quality, professional study are relatively high, indicating that the student's thoughts are relatively correct and have certain learning ability, but in terms of comprehensive personality qualities such as organizational ability, expressive ability and interpersonal skills. And the emotional control ability is weak, and it is easy to produce negative emotions and be affected. Some data scores have abnormal values, and the scores are relatively low, which may be related to the presence of a large number of adverbs in the sentence.

5. Conclusion

The introduction of text mining analysis method into the evaluation of students' comprehensive quality has certain practical significance. The word cloud can directly conclude the psychological

state of the students; the division of emotional polarity and emotional scores can be more quantitative to evaluate the quality of students in all aspects.

References

- [1] Liu Lijing. *Construction and Analysis of the Comprehensive Quality Evaluation Index System for University Graduates* [J]. Beijing Information Vocational & Technical College General Ability Teaching Department; 2013-30
- [2] Xiong Lin, Yang Yuhong. *Analysis of College Students' Comprehensive Quality—Taking Civil Engineering College of Southwest Jiaotong University as an Example* [J]. Southwest Jiaotong University; 2017-07
- [3] Xu Hui. *Optimization of College Students' Comprehensive Quality Evaluation System—Based on the Perspective of Improving Students' Quality* [J]. Jinan Engineering Vocational and Technical College; 2016-08
- [4] Meng Xianghong, Li Wei. *Construction of Students' Comprehensive Quality Evaluation Index System under the Transformation of Colleges and Universities—Taking Jilin Agricultural Science and Technology College as an Example* [J]. Jinan Engineering Vocational and Technical College; 2014-11
- [5] Lu Peng, Zhuang Min. *Analysis and Prospect of Text Feature Extraction Research* [J]. Information Science; 2016, 24
- [6] Feng Xiaoqin, He Hong. *Evaluation Method of College Students' Comprehensive Quality Based on Kmeans Clustering* [J]. Journal of Chinese People's Liberation Army, 2017(5)