



Using Social Big Data and Neural Network Algorithms to Evaluate the Quality of Talent Training in Colleges and Universities

Shaoying Chen, Zhenyu Huang^(✉), and Zhe Chen

Zhanjiang University of Science and Technology, Zhanjiang, China
{ zhenyuhuang, chenzhe }@zjktju.edu.cn

Abstract. This study captures the social texts of 48 undergraduate colleges and universities in Guangdong Province, uses text sentiment analysis and text classification technology to conduct research on the data, and builds an evaluation model for the quality of college talent training, so as to explore the public's influence on higher education talent training in the network society. Quality attitudes and suggestions, discussing whether this technology can be transformed into: the possibility of relevant universities and departments to provide commercial services corresponding to public opinion monitoring and decision support.

Keywords: Talent training in colleges and universities · Neural network algorithm · Sentiment analysis

1 Introduction

With the gradual acceleration of China's internationalization process, in order to improve the efficiency of running schools and the quality of teaching, domestic colleges and universities are carrying out educational reforms, and the cultivation of higher education talent quality has become a key link in teaching quality management. Educational data mining can realize the collection, processing and analysis of a large amount of high-quality educational data, so as to provide decision support for teaching work. Based on the social evaluation data of the quality of higher education talent training in Guangdong Province, this study uses text sentiment analysis and text classification technology to conduct research on the data, so as to explore the attitudes and suggestions of the network society towards the quality of higher education talent training. According to the accuracy of the results discuss the implementation path of providing commercial services related to public opinion monitoring and decision support to relevant customer groups with application value.

At present, most of the study on the quality of higher education talent training in China is focused on the professional construction of schools, teacher team construction and student training methods in colleges and universities. Based on experience at home and abroad, the talent training quality evaluation system is optimized, and there is little continuous improvement of the talent training program based on the feedback of the

training quality effect; the data sources used in the research are basically through questionnaires, expert interviews and other channels, and these data have large subjective errors, and it is difficult to provide accurate empirical support for the research. Therefore, the current studies of the evaluation method oriented to the quality and effect of talent training in colleges and universities using accurate analysis of independent and objective data are relatively scarce. In this paper, through the indiscriminate collection and analysis of network information and network social data, we carried out a statistical analysis of massive big data texts, and the neural network model based on two-way circular attention is introduced to solve the association problem between high-frequency words and emotional keywords in the text. And integrate the attention model to assign a higher weight to the domain keywords, and further improve the quality evaluation system of talent training in colleges and universities in the end.

2 Methodology

The paper uses the content analysis method of network text and the comparative research method. Content analysis is a scientific research method for objective, systematic and quantitative analysis of literature content. The advantage of this method is that it can transform fragmented and interactive information into systematic quantitative data, so as which is helpful for information mining, processing and knowledge extraction. Study uses a crawler program written in Python to obtain social text data, uses sentiment dictionary and neural network algorithm to perform sentiment judgment on the text after word segmentation, and obtains text measurement, and finally establishes a talent training quality evaluation system to analyze and evaluate the data.

There are two main sources of data for this study: one is social text data, which comes from one of the largest social platforms in China: Baidu Tieba. The other is the evaluation data of the talent training quality of 48 provincial undergraduate colleges and universities in Guangdong Province by the Guangdong Provincial Department of Education in the “Innovation and Strengthening Schools” project.

This study tries to answer the following questions through this experiment:

Q1: Can the texts obtained through social media comprehensively demonstrate the quality of talent training in higher education?

Q2: What are the characteristics and regularity of talent training quality among different universities in Guangdong Province?

3 Data Acquisition and Data Processing

3.1 Acquisition of Social Network Texts

In text analysis tasks, it is often necessary to use a large number of text data sets to complete the training of the algorithm, thereby improving the performance of the algorithm. It is an independent brand of Baidu and the world’s leading Chinese community. The community effectively uses specific keywords to gather interested Internet users together, which provides great convenience for our colleges to capture the social texts

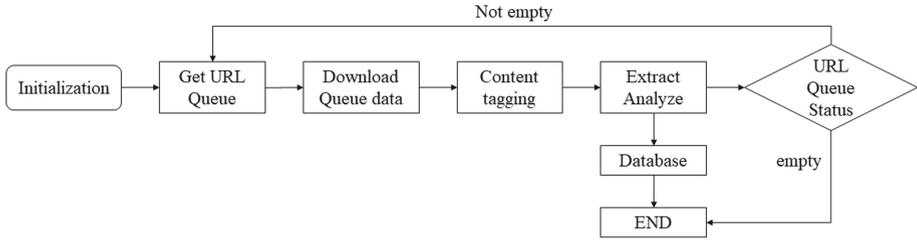


Fig. 1. Baidu Tieba text acquisition process

Table 1. Dictionary of Social Evaluation Indicators (Part)

Primary Indicator	Secondary Indicator	Keywords
Practical teaching	Proportion of practical (training) teaching credits	training
		practice
	Construction of practical teaching platform	off-campus practice
		social practice
	Innovation and Entrepreneurship Education	start business
		innovation

of students from different colleges and universities in Guangdong Province. stand. In the process of data crawling, users can log in by means of crawler and configure the application key locally to obtain the data. The flow chart of Baidu Tieba data collection is as Fig. 1.

After obtaining the data, we perform preliminary processing on the data: After the web page text data crawled by the web crawler is parsed, it may still contain many HTML tags. These useless HTML tags will affect the subsequent text analysis procedure, so it needs to be removed. In this paper, in the tag removal task, the HTML tags in the web page data are filtered by matching regular expressions. In addition, the corresponding webpage label characters can also be removed by string matching, or by using the method of wildcard text replacement to remove labels.

3.2 Construction of Social Evaluation Index and Dictionary

After completing the construction of the social evaluation index system for the quality of higher education talent training, on this basis, further construct the relevant index dictionary for each type of index. The words of each index dictionary must have the characteristics that can represent this category of indicators. The various index dictionaries are shown in the Table 1.

We divide the values into three groups. The first group is the special score of “talent training” selected by the Guangdong Provincial Department of Education in the work of “innovating and strengthening schools”. The second and the last group of data are the emotional judgment score before and after training of the neural network algorithm.

Table 2. Guangdong University Social Text Sentiment Analysis Score Sheet (Part)

University	Innovation and strong school score	Algorithm score (before training)	Algorithm score (after training)
Guangzhou Industrial and Commercial College	23.12	118.89	184
Guangdong Institute of Technology	25.21	34.3	-1602
Beijing Normal University, Zhuhai	29.89	452.1	1646

By establishing the keywords in the social evaluation index dictionary, the subject text in the massive text is located, and the emotional tendency (positive, negative, neutral) of the subject text is derived through the analysis of the sentiment dictionary + convolutional neural network algorithm. Sentiment analysis scores are obtained on the basis of the system (Table 2).

The scoring value of the neural network emotion algorithm submitted by the experimenters for training with the artificial judgment method is lower than the P value (0.048) of the special score of “talent training” selected by the Guangdong Provincial Department of Education in the work of “innovating and strengthening schools” (0.048), which is lower than 0.05. The correlation is significant, there is a negative correlation, but it is not very obvious - this conclusion answers the Q1.

4 Empirical Results and Discussion

By crawling 48 school social platforms, we obtained millions of text data (the average number of text characters per school is about 70,000, and the total number of text characters is about 3,360,000). Sentiment analysis was carried out on the content of each text sentence in Tieba. We formulated keywords according to the detailed indicators in the expert assessment and evaluation of the “Innovative and Strong School Project” conducted by the Guangdong Provincial Department of Education every year, and screened out the corresponding text sentences. The method collects, organizes and displays the emotional states marked by stickers.

4.1 Emotional Judgments Retrieved by Keywords

In this study, taking Tieba as an example, through program settings, Baidu Tieba in each school obtains the content of the first 30 pages. Judging from the number of records obtained, most colleges and universities can obtain about 100–200 texts Sentence records, only 7 colleges and universities have less than 100 records, accounting for 14.5%. Among all 48 colleges and universities, Guangdong University of Foreign Studies has the most records of 401, while Jiaying College has the least, only 68 records.

Table 3. Emotional Judgment Score Summary (Part)

University	Teacher teaching ability		
	Ability promotion		
	teaching	classroom	class
Beijing Institute of Technology Zhuhai College	-15	-1	-11
Beijing Normal University, Zhuhai	63	5	8
University of Electronic Science and Technology of China Zhongshan College	12	-1	-40
Dongguan University of Technology City College	16	0	0
Guangdong Baiyun University	3	-3	-33
Guangdong University of Finance and Economics	0	0	-35

The talent training quality evaluation system includes 6 first-level indicators and 17 s-level indicators. According to these indicators, a total of 32 keywords are selected as the analysis objects. The number of records containing keywords can be used to determine the key areas of concern of Tieba users to the talent training system in colleges and universities. Through the keyword search and summary, it is found that “postgraduate entrance examination” is the most popular keyword in college post bars, with 2,231 occurrences, followed by “communication” 1,192 times. In addition, the occurrences of the keywords “internship” and “course” are respectively. It has reached 966 times and 720 times, showing that Tieba users are very concerned about the content of college entrance examination rate, communication, practice and curriculum setting. In addition, teachers’ teaching ability and the reform of the credit system are also one of the focuses. Its representative keywords “teaching”, “class” and “credits” have more than 500 retrieval records (Table 3).

4.2 Indicator Analysis

4.2.1 Major and Curriculum Construction

Under the secondary indicator of “professional and curriculum construction planning”, the sentiment analysis on the keyword “course” is the most comprehensive among 48 colleges and universities. The values in the following table represent the positive and negative emotions of each college for the keyword. The comparison of the number of records, that is, “(number of positive emotions - number of negative emotions) * number of posts”, from the overall situation, the positive and negative emotions of the curriculum construction planning in colleges and universities are almost half, reflecting that the users of social platforms are satisfied with the college curriculum settings. Degrees vary. Beijing Normal University Zhuhai Branch has the highest level of satisfaction (180), followed by Zhuhai College of Jilin University (150), Guangzhou Institute of Physical Education (96), Guangdong University of Foreign Studies (91), South China University of Technology Guangzhou College (80); negative emotions The strongest

are Guangdong Institute of Petrochemical Technology (−140), Guangdong Institute of Technology (−112), and Southern College of Sun Yat-sen University (−95).

4.2.2 Teacher Teaching Ability

In the indicator of teachers' teaching ability, the emotions of Tieba users under the indicator of "teaching" are generally positive, and there are 23 schools with more positive emotions than negative emotions, namely: Guangdong Technical Teachers College (273), Guangzhou Medical College University (147), Guangdong Baiyun College (144); 15 schools have more negative emotions than positive ones, namely: Guangzhou Business School (−84), Guangdong University of Foreign Studies Nanguo Business School (−77), Guangdong Medical University University (−48).

The information retrieved by the keyword "classroom" is relatively small, which is not enough to support analysis and judgment. The content of the "teaching" and "class" indicators overlap. We conducted regression analysis on the two sets of data, and the results showed that the P value (0.005) is lower than 0.05, the correlation is significant, but not absolutely similar. It is speculated that the two indicators cannot be replaced, but they are highly correlated, and follow-up research will explore the correlation logic.

4.2.3 Practical Teaching

Regarding the negative evaluation of school teaching management, the sentiments of users on social platforms still tend to be neutral. The number of text sentences is small and not statistically significant. Practical and innovative courses are closely related to students' students, so get a lot of text sentences. Under the keyword "practice", there are 20 schools with more positive emotions and 23 schools with more negative emotions; under the keyword "innovation", 23 schools have more positive emotions, but only 15 schools have most negative emotions.

In the search results of keywords such as practical teaching, the number of samples for some keywords is very small. On the one hand, the school will mention this part of the content in the official publicity, and the students are only acting as obedience, there is no interest and enthusiasm for this, and it will not be discussed in private. On the other hand, it may be that the school's work in this area is perfunctory, and students have not been able to gain a positive perception of it.

4.2.4 Talent Training Reform and Effect

"Credits" and "Outcomes" are the most concerned aspects of students and are closely related to their interests. The sentiment texts under the keyword "credits" have 20 positive schools and 23 negative ones, which are basically the same; under the "achievement" keyword, there are 23 positive sentiment texts and 15 negative ones, with the majority of positive sentiments.

4.2.5 Quality Assurance System Construction Effectiveness

Among the first-level indicators of “Quality Assurance System Construction and Effectiveness”, only the number of texts obtained by the keywords “Registration” and “Postgraduate entrance examination” meets statistical significance. Due to time and resource constraints, other keywords are not scientifically set because of unscientific settings. Zero adjustments can be made in subsequent studies. There are 18 schools with positive emotional texts under the keyword “Registration”, and only 15 schools with negative sentiments, but the overall result is negative, indicating that students are dissatisfied with the school’s regulations when faced with problems related to registration and graduation; There are as many as 40 schools with positive emotional texts under the keyword “postgraduate entrance examination”, and only 5 schools with negative emotional texts, indicating that students are more emotionally positive in the face of problems related to their own future.

The results of the keyword search in this part of the subject competition are very few, indicating that the work of students, which the school claims to attach great importance to, has actually become a competition for individual “students” and teachers, and most students will not gain a sense of participation from it. Or the benefit, which is also not discussed in private. Graduation thesis (design) is something that every student must participate in, but at least the keyword search results, the direct inference is that both the school and the students are perfunctory about the graduation thesis, and will not become a topic of private discussion for students.

4.2.6 Degree Awarding Quality

We only set up one keyword in “degree awarding quality”, 15 schools are evidenced by the emotional text under the “degree” keyword, and 28 schools have the most negative emotions, indicating that most students have a policy on the school’s degree regulations. Not satisfied, this is the highest dissatisfaction among the indicators set in this study. It is speculated that there is still much room for improvement in the degree management work of many colleges and universities.

Generally speaking, the positive emotions of text sentences in general medical colleges are greater than the negative emotions. It is speculated that it may be because the academic management of medical colleges is stricter than that of ordinary colleges, and some publicity and supervision work is well implemented. The negative emotions in the text sentences of Guangdong Peizheng College are very large. After a random inspection, it was found that most of them were caused by the sale of advertisements, which did not have statistical significance.

5 Conclusion

Looking at the demands of students from social networks, compared with teaching practice and student competitions, college students are more interested in graduation and further study. However, it may also be that students are more inclined to communicate topics related to further study and graduation on social platforms. As for whether it is a topic interest issue or a platform selection issue, further analysis and research are needed.

One of the most critical results of this study is that there is a significant negative correlation between the evaluation results of the quality of talent training by the competent departments of universities in Guangdong Province and the evaluation results of social network texts. The understanding of this result is divided into two aspects: one reason is that the administrative departments of colleges and universities have different observation perspectives from college students, and the information they obtained has a large error; the other reason is that the university is in accordance with the higher authorities. Students' demands are often left behind at the time, which result in dissatisfaction. Regardless of the above situation, there may be a big difference between the information announced by the Guangdong Provincial Higher Education Authority and the actual situation. For students and parents, the distortion of the information can be misleading in choosing schools and careers. In view of this, the quality of higher education talents using text analysis on social platforms shows great commercial value, and there are various ways to realize commercial value, such as: providing paid consulting services, membership-based data usage, and free announcements (e.g.: advertising income) etc. At the same time, we suggest that the Guangdong Provincial Department of Education can further refine the indicators in the process of formulating the optimization standards for the quality evaluation of talent training in colleges and universities, especially taking the satisfaction and emotional appeals of college students into consideration to the implementation of the school's plan.

Acknowledgements. We are grateful for the financial and technical support provided by the following projects:

Guangdong Province's 2020 College Ideological and Political Education Project (2020GXSZ110).

The 5th of non-funded scientific and technological research projects of Zhanjiang City Science and Technology Bureau in 2021 (2021B01428).

Guangdong Province 2021 General Universities Recognized Scientific Research Project - General Universities Young Innovative Talents Project (2021KQNCX141).

References

1. Blaya C (2019) Cyberhate: a review and content analysis of intervention strategies. *Aggress Violent Behav* 45:163–172
2. Coates H (2005) The value of student engagement for higher education quality assurance. *Qual High Educ* 11(1):25–36
3. Harrewijn A, Van Der Molen MJ, van Vliet IM, Tissier RL, Westenberg PM (2018) Behavioral and EEG responses to social evaluation: a two-generation family study on social anxiety. *NeuroImage Clin* 17:549–562
4. Hou K, Guo M, Li X (2020) Study in the big data talent training oriented reform of information management and information system major under the management category. *J Phys Conf Ser* 1616(1):012025. IOP Publishing
5. Lindgren BM, Lundman B, Graneheim UH (2020) Abstraction and interpretation during the qualitative content analysis process. *Int J Nurs Stud* 108:103632
6. Romero C, Ventura S (2013) Data mining in education. *Wiley Interdiscip Rev Data Min Knowl Discov* 3(1):12–27

7. Romero C, Ventura S (2020) Educational data mining and learning analytics: an updated survey. *Wiley Interdiscip Rev Data Min Knowl Discov* 10(3):e1355
8. Wang L (2014) Internationalization with Chinese characteristics: the changing discourse of internationalization in China. *Chin Educ Soc* 47(1):7–26
9. Yadav A, Vishwakarma DK (2019) Sentiment analysis using deep learning architectures: a review. *Artif Intell Rev* 53(6):4335–4385. <https://doi.org/10.1007/s10462-019-09794-5>
10. Yang L, Li Y, Wang J, Sherratt RS (2020) Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access* 8:23522–23530

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

