Analysis and Evaluation of Engineering Job Demand Based on Big Data Technology

Chun Wang

Yiwu Industrial and Commercial College, Yiwu, China

Correspondence should be addressed to Chun Wang; spring@ywicc.edu.cn

Received 25 July 2022; Revised 24 September 2022; Accepted 28 September 2022; Published 10 October 2022

Academic Editor: Lianhui Li

Copyright © 2022 Chun Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper carried out an analysis and evaluation research of engineering job demand based on big data technology. By collecting the semistructured or unstructured recruitment information of engineering positions on the recruitment website, text mining technology is used to mine the knowledge model hidden in the market by building a relatively perfect Dictionary of professional skills of engineering positions. Based on the large sample data, a comprehensive, multidimensional and high-precision post-demand characteristic model is constructed. The model can not only interpret the existing recruitment market and elaborate the specific skill needs of different positions but also predict and analyze the model and estimate the required skills in combination with the specific post-responsibility characteristics, which can not only enable candidates to submit resumes reasonably, Moreover, it provides suggestions for colleges and universities or other relevant employment institutions to have an accurate, comprehensive and in-depth insight into market demand and carry out effective talent training for all units. Through the analysis of the experimental results, it is proved that the established engineering post-demand analysis model is consistent with the actual situation, and has certain explanatory significance for the current economic and social recruitment phenomenon, and the model has reference value.

1. Introduction

1.1. Research Background. Since the 21st century, the country has paid more and more attention to the cultivation of talents and has invested more and more in education. People have also paid more attention to the improvement of their own and the cultural level of the next generation, which has led to the continuous expansion of the enrollment scale of colleges and Universities [1–4]. In recent years, the number of graduates in the country has increased linearly. These phenomena have resulted in an increasing number of people waiting for employment, but the employment positions are limited and becoming saturated. The employment situation is getting more and more serious [5–7]. Therefore, it is very necessary to carry out the research of engineering job demand analysis and evaluation.

Nowadays, online recruitment has become the most popular recruitment method in the talent recruitment department of enterprises [8–13]. Online recruitment is not limited by time and space and has low cost and fast efficiency. Enterprises only need to publish recruitment information in the web page, which can be seen by candidates all over the world, covering a huge area.

Online recruitment announcements contain a large amount of text information. Generally speaking, the recruitment information includes the job description, responsibility requirements, salary and welfare, and other recruitment related information provided by the enterprise, as well as the enterprise capital injection, business operation, enterprise scale, workplace, and other information [12, 13]. However, these information are in the form of text in the web page, which basically belongs to the unstructured or semistructured form. The traditional statistical analysis method for structured data is not suitable.

1.2. Research Meaning. Based on the knowledge of computer, statistics, informatics, and other disciplines, this paper comprehensively uses statistical analysis, text mining, machine learning, and other methods to mine the job
recruitment information of engineering job, aiming to find out the hidden related information from the extracted demand characteristics, and further expand the application of text mining technology in real life.

From the perspective of research methods, the data studied in this paper is from online recruitment data, which is different from the data obtained by single research and questionnaire survey. The data obtained from the online recruitment website is that the recruitment enterprises are located in all major provinces in China, covering all enterprise sizes, from private enterprises with dozens of people to state-owned enterprises and listed companies with hundreds of thousands of employees. The enterprise field involves all aspects, It completely makes up for the one-sided conclusion brought by the simple questionnaire survey. Secondly, this paper uses Chinese text mining technology to intelligently and efficiently extract key feature words from a large number of recruitment information, which greatly saves the cost and time of manual coding and analysis.

From the research results, this paper uses Chinese text mining technology to mine and analyze the demand characteristics of recruitment information. From the results, we can get the corresponding conclusions: for colleges and universities, on the one hand, the training plan for specific majors should be formulated in combination with the real talent demand of the market and the teaching advantages of colleges and universities, so as to provide a reference for cultivating all-round and high-quality applied talents, on the other hand, The school can set up a special employment department to pay attention to the recruitment information in the society at any time. Once there is a position suitable for students to apply, the employment information will be forwarded to students in time so that students can more accurately locate their employment direction; for students, in the face of massive recruitment information, students can quickly locate key information, find their own suitable positions for application, and consciously cultivate their professional qualities in daily learning and life, aiming at a certain type of enterprise or a specific position in advance; for enterprises, making recruitment announcements based on job requirements will be more conducive to the delivery of resumes by suitable candidates, and enterprises can also recruit qualified talents.

2. Related Theory and Technology

2.1. Text Mining Definition. Text mining involves a wide range of fields. According to different knowledge structure systems and research directions, scholars’ understanding and definition of text mining will be different, and the core technologies in various fields are also different. However, generally speaking, the definitions proposed by experts in any field are basically interpreted on the basis of text data, text information, and text knowledge definitions. Text data is composed of natural language text sets that people can understand but can not be fully used, so text data also has the fuzziness and ambiguity of language itself; text information refers to the processing of text data through some methods and means, encoding the data into a set of formatted data that can be recognized by the computer, which is structured and unambiguous; Text knowledge refers to extracting useful knowledge or models from text data.

Big data technology [13–15] can provide a reliable tool for this study. Based on the definitions of text data, text information and text knowledge [16–20], this paper summarizes the definition of text mining. Text mining refers to the process of extracting knowledge structures that are unknown in advance, but in fact are potential, and can also be extracted, utilized, and understood from a large number of text data. Of course, text data exists in the form of unstructured or semistructured, so text mining can be seen as an extension of traditional data mining in text information processing.

2.2. Text Mining Process. The premise of the text mining process [21–25] is to collect text data with research value and significance and start text mining after cleaning and sorting out the “impurities” in the text data. Figure 1 below is a typical whole process of text mining.

2.2.1. Text Collection. This stage is mainly to collect, mine, and sort out the text data to be studied by the task. Data is displayed in the form of text, that is, most of them exist in semistructured or unstructured form. There are various forms of text collection. The simplest form is to copy and paste existing content to create text data, but in most cases, the object of research will be in the web page. For the data that exists on the web page, you can use a crawler to obtain it, use search engine technology to retrieve the required information, analyze web page elements and structures, grab data through web crawler technology, and establish a text database for research. However, there will be some irrelevant information in web pages, such as navigation, advertising, copyright, and other content, which is commonly referred to as “noise.” Before crawling data, it is necessary to set certain text extraction rules, which play an important role in purifying data. Through the setting of rules, useful information is retained, and unimportant disturbing information is discarded, so as to achieve the purpose of purifying data.

2.2.2. Chinese/English Word Segmentation. Because text mining technology [26–29] mainly deals with unstructured or semistructured text data, the existing traditional computer recognition methods are difficult to understand the semantics of the natural language, so the text data cannot be directly applied by computer after being collected, and appropriate processing should be carried out to extract metadata that can represent its characteristics, save it in a structured form, and form a text feature library. For English documents, stemming and lemmatization are required, and sometimes the two methods are directly referred to as stemming. Both of these methods are aimed at recognizing the transformation form of words to get the basic word form. However, there is still a certain difference between the two. The main purpose of stemming is to obtain word stems by using writing rules, while lemmatization is to obtain the
reduce the computing space and save costs. So far, researchers at home and abroad have proposed a variety of models to represent texts, including vector space model (VSM), probability model, N-gram model, and hybrid model.

2.2.4. Text Feature Selection. When the amount of data is large, the amount of feature word segmentation obtained also increases correspondingly. Using spatial vectors for representation can sometimes be as high as tens of thousands of dimensions, but such a large feature data set is not necessarily useful for the research and analysis process and even leads to a significant reduction in the efficiency of calculation. It is very beneficial to find text feature sets with low dimensions but little difference from the analysis results of the original data set. The text feature is a search process, the purpose of which is to find feature subsets that can represent the original data set but have much lower dimensions. Information gain, cross entropy, mutual information, word frequency method, document frequency method, statistics, and evidence weight are common methods for text feature selection.

2.2.5. Pattern or Knowledge Mining. After selecting the text features, we use appropriate analysis methods to mine knowledge and structural models for the obtained feature sets. Text classification, clustering, correlation analysis, association rule analysis, and so on are common text mining task methods.

2.2.6. Result Evaluation. The models or knowledge obtained are not all useful. Some models are not consistent with the actual situation and have little significance. Therefore, corresponding evaluation rules need to be used to judge which models are useful and which are not. Among them, precision and recall are the most common evaluation indicators.

2.2.7. Mode or Knowledge Output. This is the last step in the process of text mining. For the obtained model or knowledge output that meets the requirements.

2.3. Association Rule Technology

2.3.1. Association Rule Implementation Process. Association rule mining has now become one of the important technologies in the field of data mining. Correlation means that there is a certain law between the values of two or more variables. Data association is a kind of important, potential but discoverable knowledge in a database. The purpose of association rules is to find out the hidden association network in the database. For unstructured or semistructured text, after transforming it into structured text feature vectors in a certain way, we can find text frequent patterns or text association rules in large-scale text sets.
2.3.2. Apriori Algorithm. This paper uses the classical frequent itemset mining algorithm Apriori algorithm in traditional association rules to mine. This algorithm is the most influential Boolean association rule mining algorithm at present, and it is also the basic algorithm of most frequent itemset mining algorithms of width first search type.

3. Data Collection and Preprocessing

3.1. Results of Data Collection. The data in this article comes from three websites: Zhaopin Limited, 51job, and Lagou. Retrieved the positions of engineering job nationwide from the website; the text data fields include job title, company name, enterprise-scale, work location, enterprise nature, financing, salary, job description, and job requirements. The research object is the text data of 30,000 valid recruitment analyst positions I obtained from three professional websites.

3.2. Data Preprocessing

3.2.1. Data Cleaning. The text data obtained from web pages also contain a lot of “impurities;” for example, special symbols, useless numbers, etc.; sometimes, the same company will publish recruitment information on different websites simultaneously, and data duplication will occur. The overlapping text data needs to be cleared. In addition, some entries that are not meant for recruitment classification are also deleted. Similar to some companies, there are no specific requirements in the recruitment announcement for engineering job positions, just a simple introduction to the corporate culture. Such job postings are not helpful for us to capture the characteristic words of engineering job positions and must be deleted.

3.2.2. Chinese Word Participle. This paper uses the jieba tokenizer in R software for participle, where the participle engine uses a hybrid model. This model combines the advantages of the maximum probability method and the hidden Markov model, and the participle effect is better. Then, the keyword extraction technology based on the TF-IDF algorithm is used to obtain words with higher frequency in the text than in other texts, and such words are extracted as characteristic keywords.

Part of the participle results are listed as follows:

Before participle: job requirements: (1) more than three years of work experience, has experience in user growth, and engineering job is preferred; (2) have a relatively in-depth understanding of the Internet industry and have specific strategic thinking; (3) have more strong of team collaboration ability and communication ability, thinking active, and learning ability strong; (4) stable and meticulous, able to withstand certain work pressure; (5) proficient data warehouse; proficient in data query languages such as hive/SQL and statistical analysis software such as SAS/R; familiar with scripting languages such as Python/Shell; familiar with the Linux environment and common commands.

After participle: job/requirement/1./Three Years/More Than/Work/Experience/Has/Users/Growth/And/Data/Analytics/Experience/Is/Preferred/2./Hava/Internet/Industry/Hava/Comparison/In-depth/Understanding/Hava/Certain/of/Strategy/Thinking/3./Hava/More than/Strong/of/Team/Collaboration/Ability/And/Communication/Ability/Thinking/Active/Learning/Ability/Strong/4./Stable/Meticulous/Able to Withstand/Certain/Work/Pressure/5./Proficient/Data/Warehouse/Proficient/Hive/SQL/etc/Data/Query Language/And/SAS/R/etc/Statistics/Analysis/Software/Familiar/Python/Shell/etc/Scripting/Language/Familiar/Linux/Environment/and/Common Commands.

3.2.3. Text Stop Word Filtering. As seen from the above-given participle results, many words are not helpful for us to extract the feature words of engineering job positions. For example: of, have, who, right, and equivalent conjunctions, prepositions, auxiliary words, and other function words. Some words frequently appear in the text but do not affect the analysis results, such as job, requirement, or possess. After filtering by using stop words, the final participle results are as follows:

Three Years/More Than/Work/Experience/Has/Users/Growth/And/Data/Analytics/Experience/Is/Preferred/Hava/Internet/Industry/Hava/Comparison/In-depth/Understanding/Hava/Certain/of/Strategy/Thinking/Hava/More than/Strong/of/Team/Collaboration/Ability/And/Communication/Ability/Thinking/Active/Learning/Ability/Strong/Stable/Meticulous/Able to Withstand/Certain/Work/Pressure/Proficient/Data/Warehouse/Proficient/Hive/SQL/etc/Data/Query Language/And/SAS/R/etc/Statistics/Analysis/Software/Familiar/Python/Shell/etc/Scripting/Language/Familiar/Linux/Environment/and/Common Commands.

It can be seen from the above-given participle results that the text after participle has become a collection of words, which is conducive to quantitative analysis, but there are still unsatisfactory places. For example, “strategic thinking,” “teamwork ability,” and “communication ability” are a whole; once the split makes the subsequent feature extraction inaccurate. Therefore, a dictionary of professional skills in the field of engineering job is constructed in this paper. The dictionary is constructed by searching the professional catalogs in colleges and universities. And, the professional terms involved in the analysis and summary of the recruitment information are summarized and sorted to generate a dictionary. The obtained keywords can be further selected and refined through professional dictionaries to mine text features deeply.

4. Empirical Analysis of Engineering Job Demand

4.1. Distribution of Work Locations. After the feature extraction of the acquired text data, a statistical analysis of the geographic location of the recruiting unit was carried out. Figure 2 is distinguished by the different shades of color.
Provinces with darker colors indicate larger demand, while lighter provinces have smaller demand to show the demand for an engineering job in different provinces. Overall, engineering job positions are concentrated in Beijing, Guangdong, Shanghai, and Zhejiang where the proportions are 49.33%, 18.21%, 11.16%, and 11.05%, respectively. These four provinces and cities are the most advanced domestic economic and technological development areas. All have superior geographical environment and resource allocation advantages, so many Internet giants are located in these significant areas. For example, famous Internet companies such as Alibaba, Tencent, Huawei, JD.com, and Baidu have significantly increased the demand for an engineering job. But other provinces also have some demand, but the number is not much, only 10.45% of the total.

Although Beijing, Shanghai, Guangzhou, Shenzhen, and Hangzhou have great demand for engineering job, they all have their characteristics. Various job-hunting positions in Beijing bring more development opportunities and have relatively high salaries. But it also comes with the same price: more overtime hours and times and more significant life pressure. The entire industry in Shanghai is relatively competitive, and people’s spending power is also very strong. Today’s Shenzhen is a metropolis where our country’s strategic emerging industries and cutting-edge technology companies gather. Salaries are higher than those in other parts of the country, and there is a steady increase. However, people are under considerable living pressure due to the high rent and living standards. However, in Guangzhou, the cost of living is lower, and opportunities and salaries are less than those in Beijing, Shanghai, and Shenzhen. But less overtime and less stress. These cities have their own advantages, but they are all ideal places for career development.

In general, our country’s engineering job position is concentrated in economically developed areas such as Beijing, Shanghai, Guangzhou, Shenzhen, and Hangzhou. Therefore, job-seekers considering a job in engineering job in the future can consider this area with rapid economic development and high demand for this position. But often, where there are more opportunities, the competition will be greater, and the pressure will be more significant. Therefore, making more adequate preparations and improving your professional ability to obtain the ideal job position and the salary level is necessary.

We conducted an in-depth analysis of Beijing, Shanghai, Guangzhou, Shenzhen, and Hangzhou, cities with many engineering job positions. It is found that the proportion of its engineering job in the urban area of each city is also very different.

Engineering job positions are concentrated in Beijing and Shenzhen and are only distributed in individual areas. The engineering job positions in Beijing are mainly distributed in Chaoyang Area and Haidian Area; The distribution of engineering job positions in Shenzhen is primarily concentrated in Nanshan Area; engineering job positions in Hangzhou, Shanghai, and Guangzhou are relatively scattered. Almost every urban area has the distribution of its engineering job posts, but its distribution is not uniform. Hangzhou is mainly concentrated in Gongshu Area and Xihu Area. Shanghai is primarily located in Pudong New Area, while Guangzhou is mainly in Tianhe Area.

4.2. The Situation of Corporate Financing. Most companies need external financing in the process of growth. Generally speaking, the more mature the company is, the more financing it will take. Angel investment rounds are generally projects in the early stage of the company’s startup. At this time, the company has a preliminary product prototype, business model, and Core users, some of whom do not have a complete product and business plans. The subsequent A rounds, B rounds, C rounds, D rounds, and above-given are all the external financing needs of enterprises from losses to profits and gradually mature until they are about to go public. However, some small private enterprises do not require external financing due to the small scale of their development. As far as listed companies are concerned, they can conduct internal financing by issuing additional shares and stock shares and usually do not need external financing. As can be seen from Figure 3, in the recruitment of engineering job positions, it can be seen from the financing status of enterprises that for small enterprises that have not yet raised funds or are in angel round financing, the proportion is tiny, accounting for only 4.3% %, listed companies, C rounds, D rounds and above, and enterprises that do not need financing account for 75% of the total. That is to say; the more mature companies need to recruit engineering job; this is also consistent with the job responsibilities of engineering job. For small companies, vast amounts of data will not be generated in daily business activities. Through
relatively simple statistics, Analytical methods are enough to make decisions, which is why there is not much demand for an engineering job in small businesses.

4.3. Distribution of Job Classes. The job responsibilities in the characteristic vocabulary are subdivided into six classes and 17 positions, as shown in Table 1. Among all position postings, the most technical positions are technical, with a total of 8,759 jobs, accounting for 28.08% of the total. Followed by operation class (7602, 24.37%), marketing class (7573, 24.27%), design class (4963, 15.91%), functional class (2301, 7.38%), and production class (1311 items, 4.20%). During data collection, the most in-demand jobs for an engineering job in the talent recruitment market include marketing, development, operations, and web design positions.

5. Conclusions

The imbalance between the supply and demand of social talents has always been a major problem in the development process of our country. With the continuous expansion of the enrollment scale of colleges and universities, the number of college graduates is increasing year by year. A large number of fresh graduates are facing employment problems, while a large number of enterprises cannot recruit talents that meet the requirements. The main reasons for the above difficulties are the disconnection between the professional talents cultivated in Colleges and universities and the real market demand, the unreasonable professional curriculum system, and the lack of accurate insight into the talent market demand.

In order to solve the above problems, this paper takes the recruitment of Engineering posts as an example, collects more than 30000 recruitment information on the recruitment network, uses text mining technology to build a multi-dimensional postdemand feature analysis model, uses natural language processing technology and machine learning to create a relatively perfect skill dictionary for engineering posts and analyzes the market demand and employment skills from multiple perspectives. This method makes up for the time-consuming defects of traditional research, manual statistical analysis, and other methods and realizes the rapid, efficient, and intelligent mining of postdemand characteristics, especially in the case of a large amount of data and

![Figure 3: The situation of corporate financing.](image)

Table 1: Recruitment position category statistics.

<table>
<thead>
<tr>
<th>Position class</th>
<th>Specific position</th>
<th>Frequency</th>
<th>Position class</th>
<th>Specific position</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Production class</td>
<td></td>
<td></td>
<td>Marketing and selling class</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Production</td>
<td>1311</td>
<td>Selling</td>
<td></td>
<td>5825</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Marketing</td>
<td></td>
<td>1748</td>
</tr>
<tr>
<td>Technical class</td>
<td>Develop</td>
<td>8759</td>
<td>Operation class</td>
<td></td>
<td>7602</td>
</tr>
<tr>
<td></td>
<td>Test</td>
<td>5235</td>
<td>Operation</td>
<td></td>
<td>2823</td>
</tr>
<tr>
<td></td>
<td>Operation and maintenance</td>
<td>1499</td>
<td>Customer service</td>
<td></td>
<td>2122</td>
</tr>
<tr>
<td></td>
<td>Optimization</td>
<td>1732</td>
<td>Edit</td>
<td></td>
<td>1519</td>
</tr>
<tr>
<td></td>
<td></td>
<td>353</td>
<td>Promote</td>
<td></td>
<td>821</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Data manage</td>
<td></td>
<td>317</td>
</tr>
<tr>
<td>Functional</td>
<td>Administrative</td>
<td>2301</td>
<td>Design class</td>
<td></td>
<td>4963</td>
</tr>
<tr>
<td></td>
<td>Finance</td>
<td>1316</td>
<td>Web page</td>
<td></td>
<td>2888</td>
</tr>
<tr>
<td></td>
<td>Human resources</td>
<td>822</td>
<td>Visual</td>
<td></td>
<td>2075</td>
</tr>
<tr>
<td></td>
<td></td>
<td>163</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td></td>
<td>31198</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
complex structure. In the past, the data used in the mining and research of the demand characteristics of recruitment positions were obtained through questionnaires and surveys. The amount of data obtained by these methods is small, the coverage area is limited, and there is no applicability. The data in this paper is from the recruitment announcement information data of engineering positions in the three major recruitment websites in China. The recruitment enterprises are all over the country, but there is no lack of zero demand for engineering positions in some regions. In short, the data is persuasive.

Compared with the traditional analysis methods, this paper has the following characteristics in the construction of the model and the selection of methods: the research data comes from online recruitment websites, and the enterprises that publish recruitment announcements are distributed nationwide, with a large amount of data, a more perfect structure and more persuasive; in terms of data processing, by building a relatively perfect Dictionary of professional skills of data analysts and using Chinese word segmentation technology for text data preprocessing, we can greatly give play to the existing computer technology to process text data efficiently and intelligently through algorithms and machine learning.

Different from the previous data mining, which only aims at structured data, this paper starts with semistructured or unstructured text information, deeply excavates the hidden knowledge model in the text, and makes the output of the results more meaningful; in terms of engineering job technology, text mining algorithm, statistical analysis method, correlation analysis, association rule analysis, and other methods are used to mine the postdemand characteristics and the implicit relationship between the characteristics.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[23] L. Li, C. Mao, H. Sun, Y. Yuan, and B. Lei, “Digital twin driven green performance evaluation methodology of intelligent manufacturing: hybrid model based on fuzzy rough-sets AHP,


