Precise Position Intelligent Matching System of Online Recruitment Platform Based on Data Mining Technology

Zhen Gao*
Sanya Aviation and Tourism College, Sanya, Hainan, China

*Corresponding author e-mail: gaozhen@hnasatc.edu.cn

Abstract. With the rapid development of Internet technology and computer technology, network applications have been developed more and more, and have penetrated into all walks of life in society. The emergence of the networking of the talent market has made the scale of online recruitment increase, and the amount of data on the Internet has become larger and larger, and online recruitment has become the main channel for corporate recruitment. Therefore, how to use the massive online recruitment data to quickly and accurately find the corresponding information and explore the hidden knowledge mode is a very valuable research topic. Data mining (DM) is a technology for data analysis for large amounts of data. It can discover hidden, hidden, and potentially useful knowledge hidden in the data from the vague, noisy, and random mass data, and build relevant Model, realize prediction, etc. The characteristics of data mining technology (DMT) are very suitable for the analysis of online recruitment information, research on large amounts of information, and find out the knowledge in it for decision support. This article aims to study the accurate job matching system of the online recruitment platform based on DMT. Based on the analysis of the advantages of online recruitment, related DMT and the design principles of the online recruitment platform system, the data collected by Weka DM tools are analyzed. Analyzing and getting useful job positions is just to provide job seekers and corporate-related recruiters with useful job information. The experimental results show that the online recruitment platform system can complete the collection of online recruitment position information, and can realize the DM function, which has good practical application value.

Keywords: Data Mining Technology, Online Recruitment Platform, Precise Job Matching, Intelligent Matching

1. Introduction
As the social economy develops faster and faster, more and more data are generated, and the era of big data has come [1-2]. Due to the rapid growth of data, traditional data statistics and query methods can no longer meet people’s needs. At present, people pay more attention to how to analyze and process massive amounts of data in order to discover potential and valuable data sets from the data, and based
on the characteristics of the data set find the correlation between the data, and then provides a certain reference for people to make scientific decisions [3-4].

With the explosive growth of recruitment and job search information, the amount of data from both the job posting party and the demand side has become larger and larger. Under the background of this data volume explosion, the traditional headhunting based on manual screening is difficult for channels to meet the current requirements for efficiency and accuracy [5-6]. With the rapid development of network technology and the popularization of smart mobile terminals (mobile phones and tablet computers, etc.), Internet recruitment methods are recognized and welcomed by more and more job seekers and enterprises due to their convenience and speed [7-8]. At present, various recruitment websites are emerging on the Internet in an endless stream, and the networked social mode also provides a good human resources platform.

Based on the analysis of the advantages of online recruitment, related DM technologies and the principles of online recruitment platform system design, this article analyzes the collected data through Weka DM tools, and obtains useful recruitment positions only for job seekers and corporate-related recruitment staff provide useful recruitment information. The experimental results show that the online recruitment platform system can complete the collection of online recruitment position information, and can realize the DM function, which has good practical application value.

2. Research on the Precise Job Intelligent Matching System of the Online Recruitment Platform Based on DMT

2.1 Advantages of Online Recruitment
(1) Not limited by time and space

The Internet-based online recruitment method allows users to update various information such as resume, position, recruitment process, recruitment progress, etc. anytime and anywhere, completely getting rid of the constraints of time and space, and facilitating the time and space of both employers and job applicants [9-10]. Recruiting units can search the online talent information database selectively according to their own requirements for recruited talents, so that they can grasp the initiative of recruiting talents, and can get rid of geographic restrictions to select outstanding talents from different regions: and Job seekers only need to sit in front of a computer or carry a mobile phone with them to continuously check new recruitment information and deliver their resumes to their favorite positions in a timely manner, creating more job opportunities for individuals.

(2) Relatively low cost

Job seekers on online recruitment platforms no longer need to spend time running around, printing a lot of paper resumes. All they need to do is simply log on to the talent website, browse different positions published by the company in all walks of life, and complete the transfer of their resume with a few clicks of the mouse. In this way, job seekers can not only save a lot of printing and copying costs of paper resumes and certificates, but also save transportation expenses, and no longer have to travel around to participate in various offline job fairs before the interview: For employers, the cost of online recruitment will be lower than traditional methods. Most companies that recruit talents through the Internet have registered on a specific talent website and become members of that website. These general-purpose or industry-specific talent websites can provide employers with functions such as job posting, talent search, and provision of intermediary services [11-12].

(3) High recruitment efficiency

With the convenience of the Internet, recruitment and job hunting information can be quickly and quickly transmitted online, even if the submitted information is changed, it will be updated quickly, and the information dissemination efficiency is reliable and efficient. For job seekers, this not only solves the problem of geographical span, but also enables them to access a large amount of real-time updated information in the online recruitment system, and apply for jobs according to their intentions according to the type of job and other conditions through the search function. In addition, the paperless way of submitting talent resumes will also reduce the laborious job hunting and some unnecessary
troubles. At the same time, the employer can directly compare the resumes on the computer when reviewing the resumes, thereby greatly reducing labor costs.

2.2 DMT

(1) Classification analysis

The input set of classification analysis is a set of records and several kinds of marks. The classification first assigns a mark to each record, that is, records divided according to the mark, and then checks these calibrated records to describe the attributes and characteristics of these records. The main purpose of classification is to classify an event or a certain object into an appropriate category. The classification establishes a prediction model by analyzing existing historical data for analyzing and predicting future historical data. The data that we need to use when we build a predictive model are collectively called the training set, and the data in the training set can be considered as data generated from historical data or obtained from experiments.

(2) Cluster analysis

The clustering method is to divide all the data in a database into a series of subsets with special meaning, namely clustering. In cluster analysis, a group of similar data objects together form a cluster, so the objects in the same cluster are very close, but there are big differences between objects in different clusters. In practical applications, each data object in a cluster is often treated as a collection as a whole.

(3) Neural network

The neural network itself is a neural network that originated from the human body and biology. It is usually a set of linked input or output units. Each link must have a weight closely related to it. For non-linear data, or data with noise, DM based on neural network is more suitable. The method it uses is mainly to cluster and analyze the data first, and then calculate the weights through the bandit method, so as to better fit the data. The application of neural network is very extensive, and it has been fully used in large-scale database analysis and modeling tools, so large-scale DM and modeling tools based on neural network are also very popular.

(4) Decision tree

A decision tree is a tree structure that can represent a decision set. It can only be widely used in the analysis of historical data, and it can also be widely used in the prediction of future data. The establishment of a corresponding rule, the use of decision tree rule classification and calculation techniques can be used to establish a corresponding data model, this kind of algorithm that needs to be used is generally called an inductive algorithm. The first goal of using these tools is to collect all the corresponding data in a database to generate a rule and corresponding decision tree, and then the new data can be analyzed and predicted. The process of creating a decision tree is a process of continuously segmenting data, and each segmentation must correspond to a node. When performing segmentation, the difference between the groups is the largest, and the difference within the group is the smallest.

2.3 System Design Principles of Online Recruitment Platform

(1) The principle of applicability

Applicability is the most basic design principle of the system, and it is not easy to achieve this requirement. There are two key points to consider: one is to fully understand the characteristics of system participants, and the other is that system designers must have a high sense of responsibility and provide good support for platform design. The designer has strengthened the demand research work, held multiple demand analysis meetings before the design, analyzed the results through the report, and designed a system with high applicability.

(2) The principle of integrity

The system should be an organic whole. Although the principle of divide and conquer is adopted in the realization process, at the macro level, the system should remain as a whole. The whole system is a complete part for users. Users only need to interact with the operation interface, without caring about
the implementation details and corresponding components of the system, and should ensure the integrity of the design level and the integrity of the operation level.

(3) Workflow principles

Workflow refers to the staff who need to complete all related work and work process interactions carried out by a certain business. This system mainly consists of two parts, the network recruitment information collection system and the DM system, the two are an organic combination. The data must be downloaded from the network recruitment information collection system to the local, and then the related work of DM can be carried out.

(4) Strategic principles

When designing the system, it is necessary to pay attention to the early acquisition of system requirements, and it is necessary to conduct on-site demand surveys. Based on the principle of top-down and bottom-up demand acquisition, the elements of information management should be fully considered in the demand design to effectively grasp the strategic goal of the system design should be taken as the guiding ideology. The core requirements of the system should be analyzed in depth and modularized to reflect their main business processes to ensure that the results of the design are consistent with the strategic goals.

3. Experiment

3.1 Research Purpose

Online recruitment has gradually become an important application field in the development of informatization, and its construction and application have been more widely popularized. Various countries and regions have different levels of recruitment platforms, and a large number of databases have been created. In these databases, many rich rules and knowledge are often hidden in large amounts of data. The continuous development of information technology makes the historical data of various systems continue to grow, and people urgently need to convert these data into useful information and knowledge. If we can fully analyze and use these data and mine the hidden knowledge, then this knowledge will help the government make more accurate and effective decisions and better serve the public.

3.2 Algorithm Analysis

Measurement method of numeric attribute data

Since different variables often have different units or different degrees of difference, this will cause abnormal calculation results. In order to eliminate the influence of different dimensions and the value of the variable itself, before the clustering step of the data set, the standardization process should be carried out first, and then the dissimilarity between the objects should be calculated, so as to achieve a better clustering effect. Here, we introduce the following commonly used standardization methods:

(1) Min-max standardization

\[ (x_{jl})' = \frac{x_{jl} - \min(x_l)}{\max(x_l) - \min(x_l)} \]  

Among them, \( x_{jl} \) represents the value of the j-th data object under the l-th attribute, and \( \max(x_l) \) and \( \min(x_l) \) respectively represent the maximum and minimum values of the l-th attribute in the data set, such that \( (x_{jl})' \in [0,1] \).

(2) Z-score standardization

\[ (x_{jl})' = \frac{x_{jl} - x_l}{s_j} \]
4. Discussion
According to the online recruitment position information collection system, we collected data from the three-level recruitment website. The main categories are computer-related categories, with a total of 87,584 records. This paper is mainly for research purposes, so the amount of data collected is limited, but the data comes from three different mainstream online recruitment websites, which are still representative.

4.1 Cluster Analysis
(1) Cluster analysis based on academic qualifications
Performing cluster analysis through academic qualifications can make a macro analysis of job seekers in online positions. For job seekers or corporate recruiters, the results are very useful data. Using Weka's own K-means clustering algorithm to cluster the collected data based on academic qualifications, the results are shown in Table 1.

| Education   | High school and below | Junior college | Undergraduate | Graduate | Total |
|-------------|-----------------------|----------------|---------------|----------|-------|
| Number      | 5164                  | 21306          | 37444         | 646      | 64560 |
| Proportion  | 7%                    | 34%            | 57%           | 2%       | 100%  |

From the result data, we can see that among the computer-related job advertisements, the demand for undergraduate and junior college students is the largest, with the number of junior college students approaching 34%, undergraduate students reaching 57%, and the demand for other levels is low. It can be seen that with the continuous development of computer technology and software technology, the structure of recruiters has also changed. In the past, computer-related industries with high thresholds are now in increasing demand for junior college students. This is due to the continuous reduction of computer technology requirements and small and medium-sized enterprises in order to save human capital.

(2) Cluster analysis based on salary
Salary level is an important indicator to measure the recruitment market. Under normal circumstances, when the relationship between supply and demand changes, it will be directly reflected in the level of work. Due to the company's own reasons, many recruitment websites do not directly write out the price of the job position, and the salary-column of many recruitment advertisements says "negotiable", so the available amount of collected data is reduced. The clustering results based on wages by Weka are shown in Figure 1.
Figure 1. Salary result storage graph

From the clustering results, we can see that with the continuous development of computer technology, the overall salary level of computer-related industries is in a downward trend. The salary with the largest proportion is at the level of 1500-5000 yuan monthly salary, but it is higher than 8000 yuan monthly salary. The demand for regional talents is also large. This may be caused by the lack of talents with higher competitiveness.

(3) Cluster analysis based on job category

The data collected in this article is the job information of computer-related categories. The research on the current distribution of job categories is a good guide for job seekers and corporate recruiters. You can understand which job vacancies are the largest and what positions are currently available in the market. The demand is greatest. The result of clustering is shown in Figure 2.

Figure 2. Job category result storage map
4.2 Correlation Analysis

In order to better understand the relationship between the job category and the job applicant's profession, we conduct a correlation analysis on the collected data. The division of professional categories is also distinguished by the professional input options provided in the website recruitment. The results obtained based on the Weka correlation algorithm are shown in Table 2.

|                | computer software | Computer hardware | computer network | computer application | Computer image production |
|----------------|-------------------|-------------------|------------------|----------------------|--------------------------|
| computer software | 1                 |                   |                  |                      |                          |
| Computer hardware  | 0.322             | 1                 |                  |                      |                          |
| computer network   | 0.531             | 0.471             | 1                |                      |                          |
| computer application | 0.624       | 0.327             | 0.615            | 1                    |                          |
| Computer image production | 0.212 | 0.108 | 0.304 | 0.357 | 1 |

From the above analysis results, we can see that computer software has the highest correlation coefficient with computing applications, while computer hardware has the lowest correlation with computer image production. Therefore, we can intuitively understand the interrelationship between different disciplines.

5. Conclusions

The advantage of Internet recruitment over traditional recruitment lies in the high speed of the Internet and the massive amount of information, breaking through the limitations of time and space, and becoming the fastest growing field in human resource management. With the development of computer technology and information technology, job recruitment through the Internet is also increasing. Combining the latest computer and network technology, this article analyzes information collection and DM in detail, and discusses the design principles of the system. Compared with traditional systems, the online recruitment platform system has stronger real-time processing capabilities and lower operating costs, information sharing is higher, easier to maintain, etc., which can greatly improve the efficiency of DM and greatly reduce the workload of management personnel.

Acknowledgments

Key project of philosophy and social science planning project (ideological and political special project) of Hainan Province in 2018 (Project No.: hnsz2018-46)

References

[1] Joseph S R, Hlomani H, Letsholo K. DM Algorithms: An Overview. Neuroscience, 2016, 12(3):719-43.
[2] Lu H, Setiono R, Liu H. Effective DM using neural networks. Knowledge & Data Engineering IEEE Transactions on, 2016, 8(6):957-961.
[3] A survey of DM and social network analysis based anomaly detection techniques. Egyptian Informatics Journal, 2016, 17(2):199-216.
[4] Bakon M, Oliveira I, Perissin D, et al. A DM Approach for Multivariate Outlier Detection in Postprocessing of Multitemporal InSAR Results. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2017, 10(6):2791-2798.
[5] Sinharay, Sandip. An NCME Instructional Module on DM Methods for Classification and Regression. Educational Measurement: Issues and Practice, 2016, 35(3):38-54.
[6] T Pérez-Palacios, Caballero D, Antequera T, et al. Optimization of MRI Acquisition and Texture Analysis to Predict Physico-chemical Parameters of Loins by DM. Food & Bioprocess Technology, 2017, 10(4):1-9.
[7] Lu H, Setiono R, Liu H. Effective DM using neural networks. Knowledge & Data Engineering
IEEE Transactions on, 2016, 8(6):957-961.

[8] Chen X, Min L, Zhou Y, et al. A Truthful Incentive Mechanism for Online Recruitment in Mobile Crowd Sensing System. Sensors, 2017, 17(1):79.

[9] Jenkins T E, Chapman E M, Bryant M. Bio-inspired online variable recruitment control of fluidic artificial muscles. Smart Materials and Structures, 2016, 25(12):125016.

[10] Venzin M. Keys to Online Recruitment, Engagement And Retention. The Membership Management Report, 2020, 16(12):3-3.

[11] Vidros S, Kalias C, Kambourakis G. Online recruitment services: another playground for fraudsters. Computer Fraud & Security, 2016, 2016(3):8-13.

[12] Rattani A, Johns A. Collaborative Partnerships and Gatekeepers in Online Research Recruitment. American Journal of Bioethics, 2017, 17(3):27-29.