Design and Implementation of Visual Employment Recommendation System

Gengsheng Huang

Department of Electronic & Information, Guangdong Vocational Institute of Public Administration, Guangzhou, Guangdong, 510800, China

Corresponding author’s e-mail: smart0306@foxmail.com

ABSTRACT: This paper studies the design and implementation of a visual employment recommendation system based on deep learning. The system mainly uses web crawler technology, word segmentation system jieba, Django framework, data analysis algorithm Kmeans, data visualization Echarts framework and other technologies. The system can make two-way precise recommendation for enterprises and fresh graduates to improve employment matching.

1. INTRODUCTION:

The research content of this paper is to design a visual employment recommendation system based on artificial intelligence and big data analysis. First, use web crawler technology to collect job information on common recruitment websites, then clean, process, and analyze the data, and visualize the analysis results on the front end. The Python language is used to design the crawler for data collection. Based on the Scrapy framework, the Python code is simple and has good text processing capabilities, which is convenient for web content extraction. In terms of database selection, since most recruitment information is semi-structured and unstructured data, the non-relational database MongoDB was chosen to store the data.

After the data is stored, use the Chinese word segmentation technology Jieba to process the pre-processed recruitment information, and extract the core words that belong to the job requirements for analysis. Then, through the Django framework, complete the code implementation of core functions such as registration, login, recruitment data list display, recruitment data analysis chart display, and job search recommendation, and perform system testing. The technical route is shown in Figure 1.

(1) Based on deep learning, multi-view analysis, collaborative filtering and other technical and theoretical foundations, research and establish the key technical routes of the system, including the extraction of employability demand profile based on deep learning, student capability profile based on multi-view analysis, and student performance Similarity modeling, bias-based collaborative filtering algorithm, etc. [1].

(2) R&D based on the Django framework, designed a visual employment recommendation system, and visualized the analysis results on the front end. Through this system, two-way accurate recommendation can be made to enterprises and fresh graduates, and the degree of employment matching can be improved.
2. SYSTEM DESIGN:
The system architecture design is shown in Figure 2: The presentation layer is the interface for user interaction. In this system, the presentation layer mainly visualizes recruitment data in intuitive ways such as charts. The business logic layer is mainly responsible for various operations on the data, including data acquisition, data import, data processing, data analysis, etc. The data imported into the system is processed after data analysis, and the final analysis results are filed. The form is saved in the data access layer. The data access layer is also called the persistence layer, which implements access to the database. Since most of the data exists in an unstructured form, the data is stored in the MongoDB database.
3. KEY TECHNOLOGY:

3.1. Web Crawler
In this era of big data, the importance of web crawling technology on the Internet cannot be ignored. Information collection is an indispensable basis for data analysis and data mining. Web crawlers are also called web robots [2], which is a program that automatically collects and organizes data and information on the Internet in place of people according to certain rules, simulates the behavior of humans visiting websites, and automatically crawls related programs on the Internet or Script and then use certain rules to extract valuable data. The schematic diagram of the web crawler is shown in Figure 3 below. The data source of this system comes from mainstream online recruitment websites, such as BOSS direct recruitment network, Lagou network, 51job.com, Zhaopin Recruitment, ChinaHR.com, etc. It can capture the distribution of positions in different regions and different positions according to job requirements Work experience requirements, academic requirements and salary.

![Figure 3 Schematic diagram of web crawler technology](image)

This system adopts Scrapy technology. Scrapy is a high-level, fast and open source web crawling framework for crawling websites and extracting structured data from pages. For other unstructured recruitment text information, word segmentation and stop word processing are required. Since most of the recruitment information is in Chinese text, it is generally necessary to use the word segmenter jieba to complete the word segmentation of the Chinese text. As a third-party word segmentation tool for natural language processing for Chinese, jieba is a probabilistic language segmentation model. It can fully segment Chinese text, and then find the segmentation scheme S from all the results obtained from the segmentation, so that P (S) The largest, jieba supports three word segmentation modes: full mode, precise mode and search engine mode, as well as traditional Chinese word segmentation and custom dictionary [3].

3.2. Echarts framework
The full name of ECharts is EnterpriseCharts, which is a data visualization tool under Baidu. It is an open source, web-based cross-platform framework. It not only has a high-performance graphics renderer, but also a declarative visual design language core that enables it Define the built-in chart type, and has high scalability [4].

3.3. Django framework
The difference between the Django framework is that it splits into three parts: Model (model), Template (template) and View (view), which is the MTV framework. Model: Responsible for the mapping (ORM) between business objects and the database, View: Responsible for the interaction with the user (page), Controller: Accept the user's input to call the model and view to complete the user's request. The working principle of Django is shown in Figure 4:
3.4 K-means algorithm

Clustering refers to the process of separating data items with similar characteristics [5]. In this way, all member objects in the same subset have similar attributes. Different from classification methods, clustering methods emphasize more on automaticity, and gather data into clusters automatically on data sets without category marking. The information in this process is mainly based on the characteristics of data and the interconnection between data. The K-Means algorithm [6] is described as follows: Firstly, K objects are randomly selected from the database, serving as the center of mass of the initial K clusters. Then, for the remaining data objects in the database, which are calculated their distance from each centroid, and the object is assigned to the cluster with the minimum distance value according to the distance value. Finally, we calculate the centroid of all clusters again. The new centroid is reassigned according to the distance value.

The error square and criterion function are used to evaluate clustering performance for K-means clustering algorithm. Given the data set X, only the description attribute is contained, not the category attribute. Suppose X contains K clustering subset X 1 , X 2 , ..., X K . The samples in each cluster subset were respectively n 1 , n 2 , ..., n k ; The mean representative points (also called cluster centers) of each cluster subset are m 1 , m 2 , ..., m k . E is the sum of squared error of all objects in the database, p is the data object, and m i is the average value of cluster X i . E can maximizes the independence and compactness of the results. The equation is shown in (1).

\[
E = \sum_{i=1}^{k} \sum_{p \in X_i} \| p - m_i \|^2
\]

(1)

4. SYSTEM FUNCTION:

The visual employment recommendation system adopts B/S architecture and is divided into server-side and client-side. The client includes three types of users of students, enterprises and universities with different roles and functions. By sharing database information, they can fully coordinate and play the role of various users in the system, so as to provide guarantee for the effective development of employment recommendation services for college students. (As shown in Figure 5) The server side is mainly two modules of background management and database management. (As shown in Figure 6)
5. Summary
To sum up, the visual employment recommendation system based on deep learning refers to the use of web crawler technology, word segmentation system jieba, Django framework, data analysis algorithm Kmeans, data visualization Echarts framework and other technologies to provide Job seekers recommend suitable companies and recommend suitable employees to the company, thereby saving time and achieving a win-win situation for job seekers and the company.

Acknowledgments
Fund Project Number: Key Scientific Research Projects of Universities in Guangdong Province(project number:2019GKQNCX022,2019GWQNCX018,2018GkQNCX125).

REFERENCES
[1] Huang Gengsheng.Research on employment recommendation system based on deep learning[J].COMPUTER PROGRAMMING SKILLS&MAINTENANCE, 2019, (11):49-52.
[2] ZENG jian-rong,Implementation Technology and Application of Web Crawler for Multi-data Sources [J].COMPUTER SCIENCE. 2019 ,(05):304-309
[3] ZHU Yong-zhi, JING Jing. Chinese Word Segmentation Technology based on Python Language[J]. Communications Technology, 2019, 52(07): 1612-1619

[4] Brenck Florian, Michel-Backofen Achim, Katzer Christian, Smykalla Nathan, Bott Constantin, Koch Christian, Sander Michael. StudyAlert: From ECharts to ModernMessengers[J]. Studies in health technology and informatics, 2019, 264: 1643

[5] Xun Yaling, Zhang Jifu, Qin Xiao, et al. (2017) Data Partitioning in Frequent Item set Mining on Hadoop Clusters[J]. IEEE Transactions on Parallel and Distributed Systems, V28(1): 101-114.

[6] Liangjun Zhang, Zhe Fan. (2016) Hadoop Practice of Big Data Analysis and Mining[M]. Mechanical Industry Press.