Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 6, No. 2, 2023

Research on Mining Talent Demand for E-commerce Majors based on LDA Topic Model

Jiale Fu [1, 2], Hongyan Li [1, 2], Yanxia Zhao [3, 4, *], Run Zhang [1, 2], Hejing Zhang [1, 2], Taotao Pang [1]

[1] College of International Business, Zhejiang Yuexiu University, Shaoxing, 312069, China
[2] Shaoxing Key Laboratory of Intelligent Monitoring and Prevention of Smart City, Shaoxing, 312069, China
[3] Finance Office, Zhejiang Yuexiu University, Shaoxing, 312069, China
[4] School of Public Administration, Zhejiang Gongshang University, 310018, China

[*] **Corresponding author**: Yanxia Zhao (Email: 20212015@zyufl.edu.cn)

Abstract: [Purpose] Based on the text mining method, this study analyzes the job demand for domestic e-commerce industry positions in the Internet-oriented recruitment data, promotes the matching of e-commerce positions and talents, and promotes the good construction of the employment environment in the domestic e-commerce industry.[Method] Use the LDA(Latent dirichlet allocation) topic model for job competency requirements to calculate the similarity of the word segmentation results for professional talent requirements, determine the optimal number of topics, and output the visualization results of the LDA topic model.[Conclusion] In eastern China, there is a high demand for e-commerce jobs, and the majority of these positions only call for 1-3 years of experience. The basic requirement for the majority of jobs is a college degree. The Director of Operations and Director of Network Operations focus on e-commerce operations and management, the Network Promotion Specialist focuses on e-commerce advertising and product promotion, the Data Analyst and Market Analyst focus on data analysis and market research, and the Sales Representative focuses on sales. There is a correlation between the salary of the positions and education, experience and region; the higher the education, the more experience and the higher the level of regional development, the higher the salary level.

Keywords: LDA Topic Model; Text Mining; Talent Demand Analysis; Text Similarity.

1. Introduction

The employment issue has always been one of the most concerning issues for the country and the people. With the popularity of big data, the massive data on the Internet lacks of sufficient talents to turn it into a driving force for economic growth, resulting in a surge in demand for talents in all industries. Network recruitment has gradually replaced traditional recruitment methods due to its advantages of convenient and fast recruitment information release, wide information dissemination, low cost, and rapid information update, and has become the mainstream way for job seekers to obtain job information. However, online recruitment also has its problems, such as difficulty for job seekers to accurately search for job information that meets their core competitive advantages, ideal salary, or work, low authenticity of recruitment information, poor information confidentiality, and imperfect service systems of recruitment websites. However, online recruitment also has some problems, such as job seekers' difficulty in accurately searching for job information that suits their core competitive advantages, ideal salary or job requirements, low authenticity of recruitment information, poor information confidentiality, and imperfect service systems of recruitment websites[1].

In recent years, text classification algorithms have become one of the key technologies for processing massive data [3]. In response to the above problems, this article obtains the demand for job talents through text mining, which helps to promote the development of the e-commerce employment environment [4]. This article conducts descriptive statistical analysis on the information of recruitment positions and draws visual graphics, and then conducts data visualization analysis on the relationship between various characteristics and salaries from the urban distribution of required positions, experience and education requirements, etc. It also conducts word cloud visualization operations on processed text information to compare the differences in ability requirements between various industries. An LDA topic model for job competency requirements was established based on TF-IDF, and the similarity of the word segmentation results for professional talent requirements was calculated. Finally, determine the optimal number of topics and output the visualization results of the LDA topic model.

2. Relative Work

Text mining was first proposed by Mete et al.[5] during the research of knowledge discovery in text data (KDT). Xu et al. [6] conducted text mining on product reviews on e-commerce websites, categorizing text information into negative and positive, and calculating trust values using D-S evidence theory to help consumers choose better products. Cavique et al.[6] used Airbnb customer reviews in Lisbon as an example, conducting text mining with three main themes: host services, physical factors, and location. Spreafico et al.[8]conducted text mining on circular economy information incorporated a dependency pattern and introduced TRIZ to classify the information during information analysis. Orea-Giner et al.[9] conducted text mining on TripAdvisor's review data and made the results better by quantifying the relationship between data and considering the emotional connection between robots and guests. Manikandan et al.[10] conducted text mining on the treatment of bipolar disorder and repositioned the drug ketamine by combining gene expression data, providing a new approach for treating bipolar disorder. Waghmare et al.[11] conducted text mining on articles from library and information journals over the past decade and combined word

cloud analysis to conduct a comparative analysis of the themes of different journals.

In recent years, more and more scholars have made improvements and innovations in algorithms or research ideas on the basis of previous text mining and analysis. Shuang Li et al. [12] creatively proposed an effective method combining text mining, association rule mining and Bayesian network to deeply mine and use the massive coal mine safety accident case text data, so as to achieve effective identification of coal mine safety risk factors and explore the mechanism of interaction between risk factors and their importance. Hui Xu et al. [13] used deep learning to automatically classify and predict the cause of accidents by the textual features and text mining methods that quickly extract the cause information to optimize accident record analysis. Aqsa Rehman et al. [14] adopts VADER to analyze the subjectivity and polarity score of tweets, a topic model was also created using the LDA algorithm to determine the themes that were talked about on Twitter the most. The models have been constructed and evaluated using Word2Vec to capture the semantic relationships between words and LSTM and RNN sequential model for sentiment analysis.

Then, with the advent of the Internet era, a large number of online recruitment information flooded the Internet, which has attracted many scholars to discuss and study. He Müller B et al.[15] conducted text-mining research on recruitment data and combined it with vertical search engines and ontology screening to build a job recommendation search engine for fresh graduates. Spada I [16] designed an analysis system that can intelligently mine the demand for professional talents using unstructured data by analyzing the specific needs of e-commerce positions. Li J et al. [17] used financial recruitment data to construct a professional skill dictionary, built a network model based on job and professional skills, and analyzed the relationship between jobs and professional skills using complex network tools. Fang F et al.[18] took the recruitment information for data analysis positions as an example, classified the skills requirements in different regions, and combined the clustering results of ability requirements text with salary, company size, and other information for analysis. Li Y et al.[19] collects recruitment information from recruitment websites, analyzes the talent needs of companies and positions, explores the cities with the highest demand for talent, job types, and career paths, and makes recommendations for cultivating students in universities.

3. Data Presentation and Statistical Analysis

3.1. Data Collection and Cleaning

This article selects the recruitment data of the e-commerce industry on mainstream domestic recruitment websites as the research data and cleans the data after crawling. Data cleaning refers to the removal, filling, and conversion of data to eliminate useless, missing, and erroneous data, thereby improving the quality and accuracy of the data. The data cleaning in this article mainly includes the following steps:

(1) ***Removing useless data***: The dataset used in this article contains some irrelevant fields that are not useful for analysis, such as search keywords, publishing time, and hyperlinks. Therefore, we need to remove the data in these fields and delete some serial and misplaced data to complete the cleaning process of useless data.

(2) ***Repetitive value processing***: When companies recruit relevant talents, they may use multiple recruitment websites to post similar recruitment information, resulting in a large amount of repetitive data in the recruitment data. To improve the accuracy and reliability of the data, it is necessary to eliminate or merge these duplicate data. It can ensure the uniqueness and consistency of data.

(3) ***Missing value processing***: There are some empty values in some fields in the dataset, such as company nature and company industry. The absence of such data is due to the failure to capture them during the collection process, so these data are directly deleted. However, there are a large number of blank values in some fields, such as job category and number of recruits. The absence of these fields is due to the fact that they are not required fields on recruitment websites, and companies do not fill them in when posting job listings. Therefore, for these missing values, this article has adopted a new category named "unlimited" to avoid deleting too much useful data and affecting subsequent analysis and modeling results.

3.2. Data Normalization

When an enterprise publishes job information on its website, the lack of a unified data entry standard may lead to inconsistencies in the format and units of the collected data. For example, the position salary field may appear in different forms such as "10,000/month", "yuan/month", "10,000/year", "1,000/month", and "yuan/day". To facilitate subsequent data processing and analysis, we need to unify them into the same format unit.

3.3. Chinese Word Segmentation

This article uses the precise mode of the jieba library for Chinese word segmentation operations for the ability requirements and job content fields in recruitment information. However, due to the large number of e-commerce-related vocabulary in the competency requirements description, For example, "website operation specialist" "SEO engineer" etc. The word segmentation of the Jieba library will cut these words into "website" "operation" "specialist" and "SEO" "engineer", which will destroy the original meaning of the words and affect the subsequent analysis results. Therefore, in order to reduce the impact of this phenomenon on subsequent analysis, this article adds a custom dictionary of relevant professional terms during word segmentation, which greatly improves the accuracy of the word segmentation results.

3.4. Remove Stop Words

Text after-word segmentation usually contains multiple phrases or words, but some word segmentation results do not need to be considered. They have no relevance to the text content and have no practical significance, such as the commonly used connecting words "de" "zai" and "shi". In order to reduce the dimensionality of text analysis and reduce the impact on keyword frequency, it is necessary to eliminate these words after Chinese word segmentation, which is why we need to remove stopwords. Removing stop words can improve the efficiency of text analysis, reduce the difficulty and time cost of analysis, and make the analysis results more accurate.

This article combines well-known Chinese stopword lists such as Baidu, Harbin Institute of Technology, and Sichuan University's Machine Intelligence Laboratory to complete the processing of stopwords. In addition, this article also defines

a stopword list for some words that are not meaningful for capability analysis, such as "work address" "welfare benefits" "position responsibilities" "college degree or above", in hopes of obtaining a better data preprocessing result.

3.5. Data Visualization Analysis

3.5.1. Distribution of Job Demand for E-Commerce Across the Country

This article conducted a statistical analysis of company names and locations, and obtained the urban demand conditions for the positions of chief operating officer, network operations director or manager, network promotion specialist, and sales representative across the country. As shown in Figure 1, according to the color scale, the color on the left represents the number of people, blue indicates low demand, red indicates high demand, and no color indicates no demand. By observing the distribution of urban demand for four types of jobs, it can be found that cities with large job demands are mainly distributed in the eastern region of China. The demand for jobs in the central region is lower than that in the eastern region, but the demand is still relatively large.

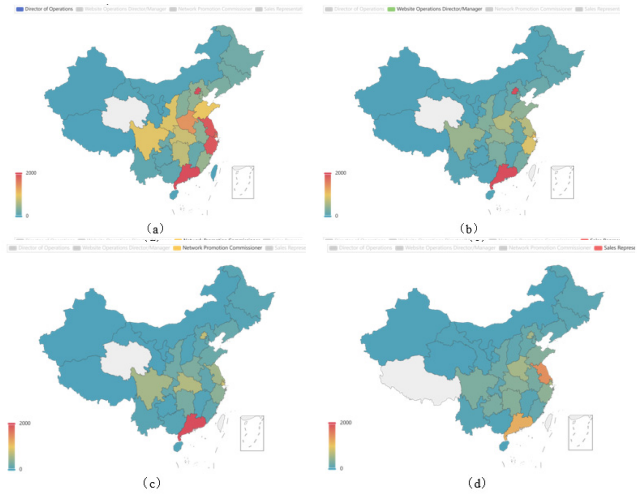

Figure 1. Heat Map of City Demand Distribution for E-commerce Jobs Nationwide

3.5.2. Distribution of Educational Requirements

As shown in Figure 2, it can be seen that the highest degree requirement among these four types of positions is a junior college degree, while the number of positions requiring a bachelor's degree and no degree requirement is at the same level. These three types of degree requirements constitute the largest proportion of degree requirements: 91.55% for the director of operations, 91.75% for the director or manager of network operations, 82.99% for the network promotion specialist, and 88% for the sales representative. It can be concluded that compared to the requirements for the positions of operations director and network operations director or manager, the employment threshold for network promotion specialists and sales representatives is lower, and the educational requirements for job applicants are not high.

This also means that for job seekers without high academic qualifications, online promotion specialists and sales representatives are suitable options. Of course, we need to be clear that although academic qualifications have a significant impact on job seekers' employment and career development, they are not the only determining factor. Therefore, job seekers need to continuously improve their comprehensive quality based on a solid grasp of professional knowledge to cope with the ever-changing workplace challenges.

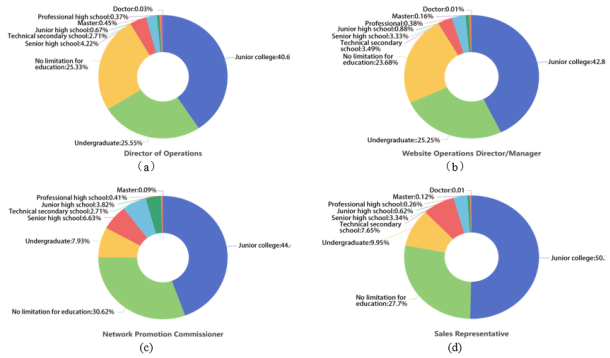

Figure 2. Distribution of education requirements for e-commerce positions using a rose diagram

3.5.3. Experience Requirement Analysis

In recruitment, job seekers' work experience is an aspect that enterprises value more. Experienced job seekers are usually more favored by employers. Through statistical analysis of the experience requirements fields for the four types of positions in Figure 3, we found that the proportion of positions requiring no experience was 33.68%, 39.04%, 53.77%, and 65.59% for the positions of operations director, network operations director or manager, network promotion specialist, and sales representative, respectively. Among them, the experience requirements for network promotion specialists and sales representatives are relatively low, while those for operations directors and network operations directors or managers are relatively high.

It should be noted that the number of positions requiring no experience is the largest. This is mainly related to the characteristics of e-commerce jobs, as the entry threshold for such jobs is relatively low, and anyone can try them without requiring too much experience. The work is easy to get started and fast to learn. However, for some high-level positions, such as the director of operations and the director or manager of network operations, a certain amount of work experience is still required to be better qualified for the job. Therefore, the experience requirements for job seekers in this type of position are mainly 1-3 years of experience, and the number of positions with 3-5 years of experience and 5-10 years of experience is far greater than that of network promotion specialists and sales representatives.

3.5.4. Post Salary Distribution

This article conducts a visual analysis of the salary and educational requirements for these four types of positions. As shown in Figure 4, the highest salaries for these four types of positions are concentrated in the "master's degree" category, followed by the "doctoral degree" category. The salary situation has been steadily rising from "college degree" to "bachelor degree". In general, job salaries depend on the level of education. However, due to the particularity of e-commerce positions, this article found that the salary for "no academic requirements" is also not low, indicating that the entry threshold for such positions is low. The salary for "PhD" is lower than that for "Master", and this article speculates that the industry's position requires a lower requirement for "PhD" degree applicants. Due to the industry's more opportunities for promotion, "PhD" degree applicants only come to this industry as a stepping stone for upward promotion, so the initial salary is lower.

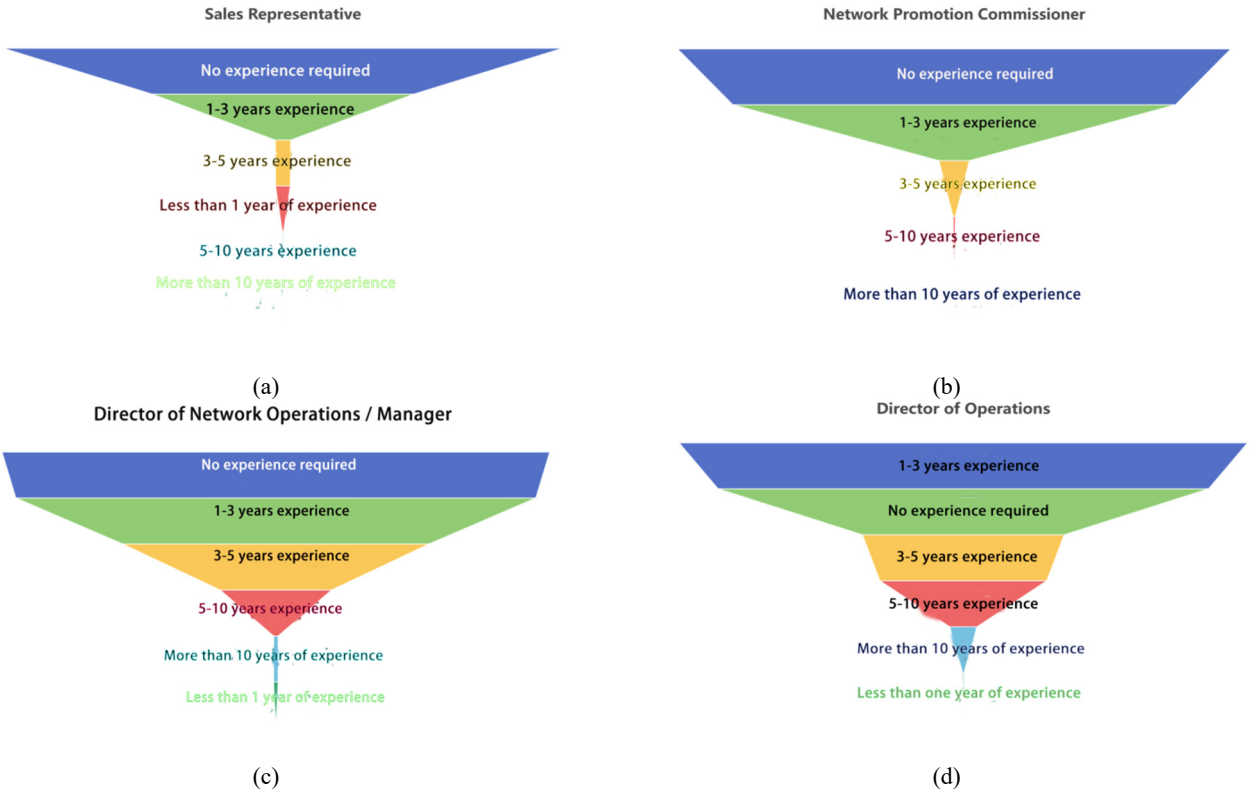

Figure 3. Funnel diagram of job experience requirements for e-commerce positions

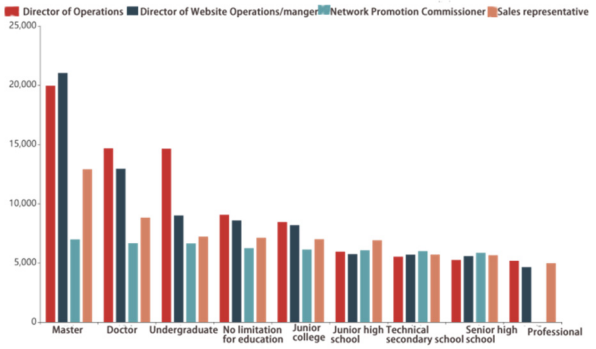

Figure 4. Bar chart of e-commerce job salary and educational requirements

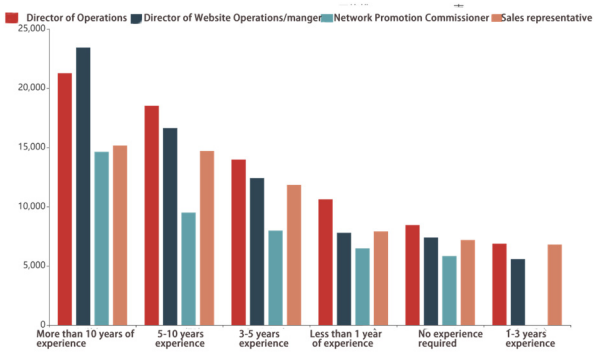

Figure 5. Bar chart of salary and work experience requirements for e-commerce positions

In addition, this article also compares and analyzes the salary and work experience requirements of these four types of positions. As shown in Figure 5, the salary peaks of these four types of positions are all located in the demand for positions with "more than 10 years of experience". The salary from "3-5 years of experience" to "more than 10 years of experience" is in a steady rise stage, indicating that as the working age increases, the salary will also increase. However, the salary situation of "no experience required" and "1-3 years of experience" is quite unique. We speculate that this is due to the large number of job openings in the e-commerce industry, and the fact that the salary composition of the e-commerce industry is mostly composed of performance commissions, so many companies are using higher salary gimmicks to attract job seekers. The lower salary for the "1-3 years of experience" requirement should be the actual salary, excluding commission, which leads to this phenomenon.

3.5.5. Distribution of Job Competency Requirements

Through data preprocessing on the field of work ability requirements, text data is cleaned and segmented. The original text data contains a large number of words that are not relevant to the job requirements topic. Here is an example to illustrate.

We only keep the Chinese characters and English characters. Next, we perform word segmentation on the cleaned text and then filter out the stop words.

Based on the analysis of keywords and competency requirements for different positions, it can be seen that the main keywords for the positions of Operations Director and Network Operations Director or Manager are operations, responsibility, management, team, etc., and the main competency requirements are operational capabilities. This is because these two positions are responsible for the operation and management of the entire e-commerce platform, including product launch, promotion, marketing, and operational data analysis. Therefore, job seekers for these two positions need to have strong operational and management skills, as well as good teamwork skills.

The main competency requirements for network promotion specialists and sales representatives are promotion and sales. This is because both positions are responsible for promoting and selling e-commerce platforms, including developing

promotion strategies, executing promotion plans, and communicating with customers. Therefore, job seekers in these two positions, need to have strong promotion and sales skills, as well as good communication and customer service skills.

4. LDA Topic Model

LDA (Latent dirichlet allocation) [20] topic model is a commonly used text analysis method that can be used for topic analysis and topic mining on large-scale text data. This model is based on the idea of probabilistic graphical models, in which each document is viewed as a mixture of multiple topics, each of which is composed of multiple words. In this model, both topics and words are hidden variables, which are inferred and learned through observed text data.

LDA has a three-layer generative Bayesian network structure, which includes the probability distribution relationships between words, documents, and the overall document. The structure is a document layer, topic layer, and feature word layer, as shown in **Figure 6.** Schematic diagram of the topology structure of hidden topics in LDA model. Its basic method includes two steps: topic inference and parameter learning. Topic inference refers to inferring the topic distribution of each document and the distribution of each word in each topic based on given text data. Parameter learning refers to learning the parameter values of topic distribution and word distribution by maximizing the likelihood function, thereby obtaining the optimal topic model.

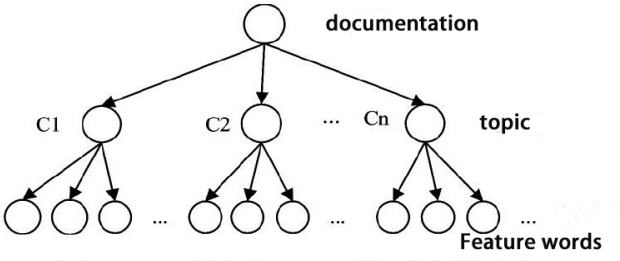

Figure 6. Schematic diagram of the topology structure of hidden topics in LDA model

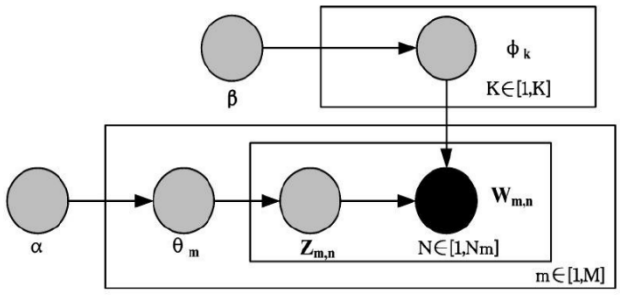

Figure 7. LDA Document Generation Flowchart

The LDA probabilistic topic model is a directed probability graph model, and its process of generating text is shown in **Figure 7.** LDA Document Generation Flowchart. First, the number of words in the text is obtained N according to the Poisson distribution P, and then the topic distribution θ_m of the document is generated according to the Dirichlet prior parameters α of the topic multinomial distribution θ_m for each document topic. The topic of the n th word in the document m is sampled from the topic multinomial distribution θ_m, and the word distribution φ_k corresponding to the topic $Z_{m,n}$ is generated by sampling from the Dirichlet prior parameters β of the word multinomial distribution under each topic. Finally, the word is sampled from the word multinomial distribution φ_k to output the word $W_{m,n}$. Where θ_m represents the probability distribution of the document topic, K represents the number of topics in the document set, and M represents the number of texts in the document set.

LDA The core formula for building a topic model is as follows:

$$P(w|d) = P(\omega|t) * P(t|d) \qquad (2-3) \qquad (1)$$

Intuitively, the formula uses Topic as the intermediate layer and utilizes current θ_d and φ_t to calculate d the probability of the word appearing in the document. Where, $P(t|d)$ is calculated by θ_d and $P(\omega|t)$ is calculated by φ_t. In fact, we can use current θ_d and φ_t to calculate the corresponding $P(w|d)$ to any topic for a word in a document, and then update the topic corresponding to the word based on these results. If the update changes the topic corresponding to the vocabulary, it will in turn affect θ_d and φ_t. Since the topic distribution of text is a simple mapping of the text vector space, it is only necessary to vectorize the LDA model text, then compare their similarity, calculate and output similar text results and similarity.

5. Job Demand Analysis based on LDA Model

5.1. Establishment of LDA Topic Model

When using the LDA topic model for text analysis, first import the Chinese word segmentation result document that has completed the data preprocessing process, and combine all the words in the preprocessed text data into a document search term matrix, which is to construct a dictionary.

To improve the accuracy and reliability of topic extraction, this article adopts the use of the TF-IDF algorithm to calculate the weight of each word. TF-IDF algorithm can calculate the weight of each word according to its frequency of appearance in the document and the document frequency in the entire corpus, thereby highlighting the importance of keywords in the text. By combining the TF-IDF algorithm to build corpus, we can reduce the impact of stop words and reduce their influence on topic extraction, while also reducing the influence of ignored or low-weighted words on topic extraction, improving the performance and stability of the topic model.

To implement the TF-IDF algorithm, we use the models. TfidfModel method in the gensim library to create TF-IDF documents and convert them into a document collection. Then, you can use the gensim. models. ldamodel. LdaModel method in the gensim library to create an LDA object. When creating a LDA object, it is necessary to determine the number of topics (num_topics). This parameter can be determined using confusion and consistency scores, thereby improving the accuracy and reliability of the topic model.

Finally, we use the LDA topic model to extract topics from text data, and visualize the results using visualization tools such as pyLDAvis. Through the construction and optimization of LDA topic models, we can better mine the topic information in text data, providing a more reliable foundation for subsequent text analysis and decision-making.

5.2. Determination of the Optimal Number of Topics

Perplexity and coherence are commonly used data metrics for evaluating LDA topic models. Among them, confusion is an indicator used to evaluate language models, representing the model's ability to predict new data. The lower the confusion, the better the model's fit to the sample data. As the number of topics increases, the level of confusion tends to decrease first and then increase.

This experiment combines the interpretability of topics, application scenarios, and other influencing factors to select the optimal values for two metrics and perform model training and evaluation to output the optimal topic model. The specific process is as follows: Define a function to calculate confusion and consistency, set the number of topics from 1 to 9, construct LDA topic models for each number of topics, and output a line chart of the relevant scores. See Table 1

Table 1. The influence of different number of topics N on the score of e-commerce positions

job	Number of topics	perplexity	consistency	job	Number of topics	perplexity	consistency
Director of Operations	1	-13.5935	0.5169	Network Operations Director /Manager	1	-13.1916	0.5271
	2	-13.8038	0.4821		2	-13.4786	0.4692
	3	-13.9172	0.5852		3	−13. 6418	0.5034
	4	-14.1254	0.5822		4	−13. 8199	0.4953
	5	-14.1930	0.6460		5	-13.9563	0.5505
	6	-14.1920	0.5132		6	-14.0763	0.6153
	7	-14.2657	0.5779		7	-14.0877	0.5788
	8	-14.3304	0.6310		8	-14.0372	0.5617
	9	-14.4075	0.6765		9	-14.2771	0.6502
Network Promotion Commissioner	1	-11.6362	0.5776	Sales Representative	1	-14.3360	0.4363
	2	-11.8874	0.6505		2	-14.4477	0.5082
	3	-11.9915	0.7187		3	-14.8883	0.4664
	4	-12.0413	0.5811		4	-14.9716	0.5628
	5	-12.1857	0.6694		5	-15.0236	0.5052
	6	-12.2288	0.5687		6	-15.1178	0.5528
	7	-12.5271	0.5162		7	-15.7581	0.6054
	8	-12.4436	0.5955		8	-15.4795	0.4610
	9	-12.3595	0.6115		9	-15.8231	0.5196

Taking the "Chief Operating Officer" as an example, when N=6, the degree of subject confusion increases, and the model may have overfitting. Therefore, this article selects N=5 and N=9, which are the lowest points of confusion and highest consistency before and after the increase in confusion, respectively.

Subfigure (a) to (b) in Figure 8 present the visualization effects of LDA topic models with N=5 and N=9, respectively. Although the confusion score for the "Chief Operating Officer" position is the lowest and the consistency score is the highest when the number of topics is N=9, there is overlap between topics 1, 2, 3, and 5, indicating that the topic division at N=9 is not effective. Although there is also overlap between themes 1 and 2 when N=5, the division effect is still better when N=5 compared to N=9. Therefore, it can be preliminarily determined that the model overfits when N=6, resulting in poor performance of the topic model. Therefore, choosing N=5 as the optimal number of topics for text mining of competency requirements for the "Chief Operating Officer" position.

Similarly, a comparative analysis of the other three types of positions was conducted to determine the optimal number of topics for these four types of positions, which were: operations director (N=5), network operations director/ manager (N=6), network promotion specialist (N=3), and sales representative (N=7).

5.3. Visualization Analysis of LDA Topic Model

In the visualization diagram of the LDA model, the circles on the left represent topics, with larger circles indicating a higher probability of the topic in the document. The distance between circles indicates the similarity between topics. On the right are the words corresponding to the topics, with longer bars indicating a higher frequency of occurrence of the word. The red portion represents the distribution of terms for the selected topic. To measure the influence of words on topics, we use the parameter λ. When λ When it approaches 1, the words that appear more frequently under the topic have a stronger correlation with the topic; On the contrary, when λ as the value approaches 0, the more unique and exclusive words under the topic have a stronger correlation with the topic. The role of this parameter is to help us determine the correlation between topics and words, thereby better understanding and analyzing the visualization results of the LDA model. (c) to (f) in Figure 8 presents the LDA visualization results of recruitment requirements for various positions in e-commerce.

Next, we will conduct a textual topic visualization analysis of the recruitment capability requirements for four types of positions. We used the jieba Chinese word segmentation

method and, on this basis, manually removed words that were unrelated to the topic or of little significance to obtain the visualization results of the LDA topic model.

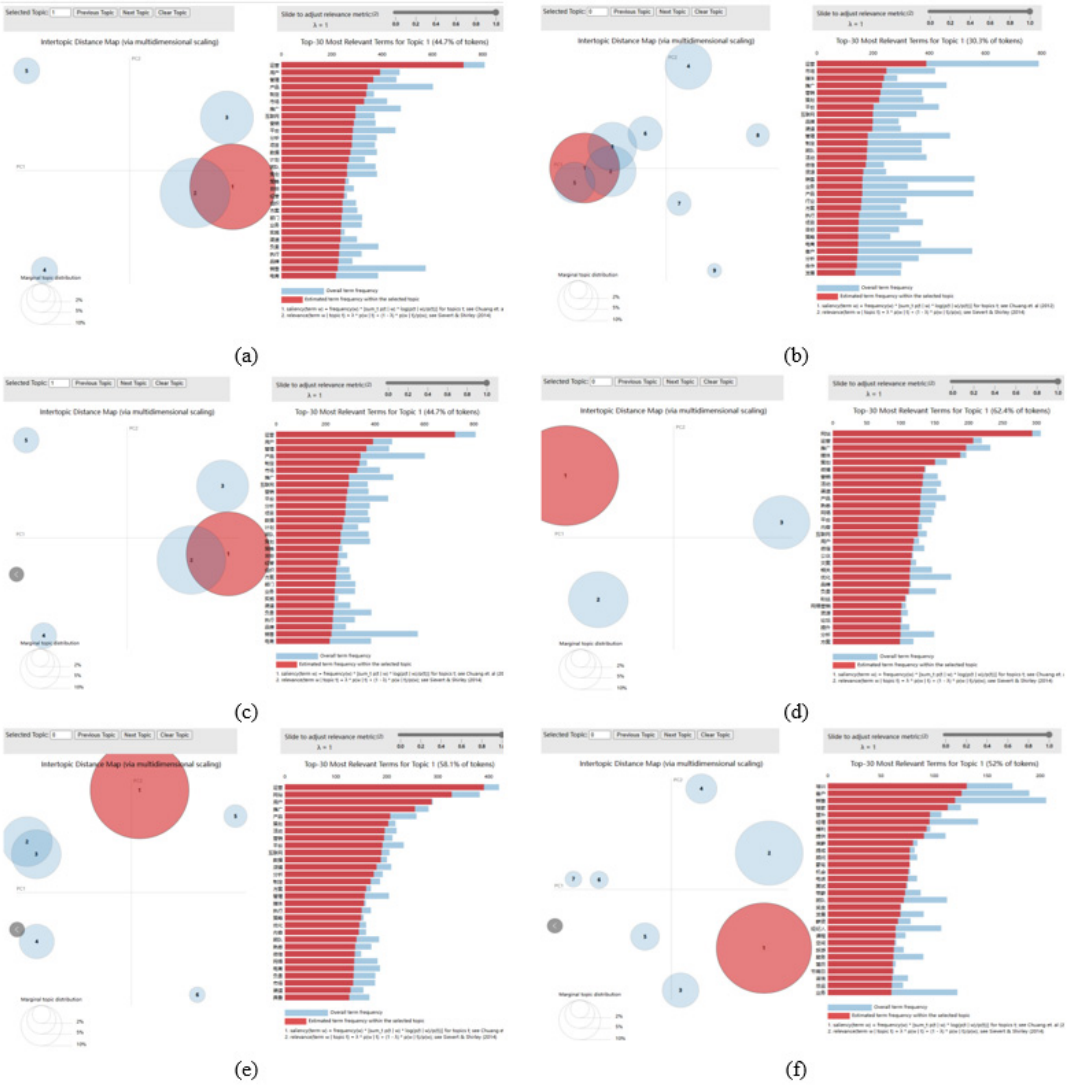

Figure 8. Visualization of LDA topic model

Table 2. Distribution of keywords of e-commerce job recruitment ability

job	thematic	byword									
Director of Operations	1	scheduled service	subscribers	Management	Produkte	develop	Markets	Promotion	Internet	Marketing	Platform
	2	customers	Sales	Stores	Benefits	Taobao	Customer Service	Maintenance	Consulting	SERVICES	E-commerce
	3	devise	Sites	Make	Copywriting	Creative	Page	Media	Graphic Design	Editorial	Art
Network Operations Director /Manager	1	scheduled service	Website	Users	Promotions	products	Planning	Events	Marketing	Analytics	Strategy
	2	sales	Foreign Trade	English	Clients	Paid	Commission	National	Amazon	Training	Welfare
	3	customers	Sales	service	inquiry	Recruitment	Customer Service	Communication	Mandarin	Customers	Love
Network Promotion Commissioner	1	node	Operation	Promotion	Media	Planning	Microblogging	Marketing	Internet	Produkte	Events
	2	bid	Placement	Accounts	Advertisements	Optimization	information flow	SEM	Search	Data	Strategies
	3	sales	Clientele	Training	Achievements	Telephony	Affinity	Challenges	Engines	Programs	Markets
Sales Representative	1	cultivate	Customers	Sales	Chain	Manager	Consultant	Commission	Client Relationships	Consulting	Business
	2	offerings	Promotion	Planning	Sales	Related	analyze	Marketing	Brokerage	Responsibility	Clients
	3	storefront	Manager	Elite	Business	Director	Leasing	Senior	E-commerce	Reserve	Training

For the first three topics of each position, we selected 10 keywords for examples, as shown in Table 2 Taking the "Director of Operations" position as an example, selecting N=5 to establish an LDA topic model will generate five topic words, each of which will be composed of several weighted words. When the number of topics N=5, the generated topic weight distribution is shown in Table 3. By looking at the distribution of subject term composition, one can have a preliminary understanding of each type of subject term. These subject terms are representative and can reflect the meaning of the topic and the distribution of vocabulary. Topic analysis can be conducted through these different weighted words.

Table 3. Keyword weight distribution of the director of operation (N=5)

thematic	Word weight distribution
1	0.011*"scheduled service" + 0.008*"subscribers" + 0.008*"offerings" + 0.008*"devise" + 0.007*"node" + 0.007*"generalization" + 0.007*"flat-roofed building" + 0.006*"planner" + 0.006*"store" + 0.006*"the Internet" + 0.005*"digital" + 0.005*"maneuver" + 0.005*"e-commerce" + 0.005*"market" + 0.005*"media" + 0.005*"analyze" + 0.004*"demand" + 0.004*"make superior" + 0.004*"programmatic" + 0.004*"WeChat"
2	0.007* "rendering" + 0.006* "insurance" + 0.006* "customer" + 0.005* "only" + 0.005* "health insurance" + 0.004* "benefits" + 0.004* "channeler" + 0.004* "gratuity" + 0.004* "vacation" + 0.004* "voice" + 0.004* " Mobile Games" + 0.004*"Progress" + 0.004*"Maternity" + 0.004*"Pension" + 0.004*"Elegance" + 0.004*"Paid" + 0.003*"Tianjin" + 0.003*"Worker's Compensation Insurance" + 0.003*"Bereavement Leave" + 0.003*"Provident Fund"
3	0.007* "Management" + 0.006* "Sales" + 0.005* "Operations" + 0.005* "Operating" + 0.005* "Markets" + 0.005* "Programs" + 0.005* "Projects" + 0.005* "Sectors" + 0.004* "Organizations" + 0.004* "Customers" + 0.004* "Enterprises" + 0.004* "Developing " + 0.004*"Coordinate" + 0.004*"Operate" + 0.004*"Assist" + 0.004*"Responsible" + 0.003*"Train" + 0.003*"Regional" + 0.003*"Target" + 0.003*"Team"
4	0.011*"Purchasing" + 0.010*"Warehouse" + 0.008*"Order" + 0.008*"Logistics" + 0.008*"Accounting" + 0.008*"Finance" + 0.007*"English" + 0.007*"Amazon" + 0.007*"Vendor" + 0.006*"Reconciliation" + 0.006*"Stock" + 0.005*" Shipping" + 0.005*"Amazon" + 0.005*"Sizzling" + 0.005*"Goods" + 0.005*"After-sales" + 0.005*"Office" + 0.005*"Proficient" + 0.004*"Responding" + 0.004*"Stats"
5	0.009*"Customer" + 0.007*"Sales" + 0.006*"Time" + 0.005*"Benefits" + 0.005*"Commission" + 0.005*"Base Salary" + 0.005*"Design" + 0.004*"Offer" + 0.004*"Call" + 0.004*"Advancement" + 0.004*"Paid" + 0.004*"Dual Vacation " + 0.004*"Salary" + 0.004*"Network" + 0.004*"Bonus" + 0.004*"Insurance" + 0.004*"Statutory" + 0.004*"Holiday" + 0.004*"Treatment" + 0.003*"Software"

6. Summary and Outlook

This article mainly analyzes the demand for e-commerce positions, educational requirements, work experience requirements, and salary conditions. The results showed that there was a large demand for e-commerce jobs in the eastern region, while cities in the central region such as Sichuan, Shaanxi, Hubei, and Henan also had a large demand for e-commerce jobs. In terms of educational requirements, a college degree is the basic required educational background, while the number of bachelor's degrees and other qualifications is not limited. In terms of work experience, experienced job seekers are more likely to be favored by employers. In terms of salary, this article found that the salary of e-commerce positions depends on the level of education. In addition, the peak salaries for all four job categories are all located in the demand for jobs with "more than 10 years of experience", and the salary will continue to increase as the working age increases. Therefore, job seekers should choose a position and development direction that suits them based on their own situation, and improve their competitiveness by constantly enhancing their sales, operations, promotion, and other abilities.

In the future, we can further explore the demand for professional talents in different types of e-commerce enterprises within the industry and provide more effective support for e-commerce industry recruitment by combining them with actual situations to put forward corresponding suggestions. For example, we can analyze different types of e-commerce enterprises (such as B2B, B2C) to understand their similarities and differences in professional talent demand. At the same time, we can also compare the professional talent demand in this industry with the actual supply situation to explore the reasons for the contradiction between supply and demand and structural imbalances, and put forward corresponding policy recommendations to provide more powerful support for the development of the e-commerce industry.

References

[1] Fernandez R M, Rubineau B. Network recruitment and the glass ceiling: Evidence from two firms[J]. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2019, 5(3): 88-102.

[2] Fellnhofer K. Visualised bibliometric mapping on smart specialisation: A co-citation analysis[J]. International Journal of Knowledge-Based Development, 2018, 9(1): 76-99.

[3] Glänzel W, Debackere K. Various aspects of interdisciplinarity in research and how to quantify and measure those[J]. Scientometrics, 2022, 127(9): 5551-5569.

[4] Wei J, Xu Y. The Application of LDA Model in the Analysis of Job Talent Demand under Big Data Technology[C]//2022 International Conference on Artificial Intelligence in Everything (AIE). IEEE, 2022: 301-305.

[5] Mete M, Yuruk N, Xu X, et al. Knowledge discovery in textual databases: A concept-association mining approach[J]. Data Engineering: Mining, Information and Intelligence, 2010: 225-243.

[6] Xu Haiping et al. Evaluating Online Products Using Text Mining: A Reliable Evidence-Based Approach[J]. International Journal of Semantic Computing, 2022, 16(04).

[7] Cavique Mariana et al. Examining Airbnb guest satisfaction tendencies: a text mining approach[J]. Current Issues in Tourism, 2022, 25(22): 3607-3622.

[8] Spreafico C, Spreafico M. Using text mining to retrieve information about circular economy[J]. Computers in Industry, 2021, 132: 103525.

[9] Orea-Giner A, Fuentes-Moraleda L, Villacé-Molinero T, et al. Does the implementation of robots in hotels influence the overall TripAdvisor rating? A text mining analysis from the industry 5.0 approach[J]. Tourism Management, 2022, 93: 104586.

[10] Manikandan S, Misra S, McCalla S. Drug Repositioning Ketamine as a New Treatment for Bipolar Disorder Using Text Mining[J]. BioChem, 2021, 2(1): 1-7.

[11] Waghmare P. Text mining and word cloud analysis of the articles published in selected library and information science journals over the past decade[J]. International Journal of Information Dissemination & Technology, 2021, 11(3).

[12] Li S, You M, Li D, et al. Identifying coal mine safety production risk factors by employing text mining and Bayesian network techniques[J]. Process safety and environmental protection, 2022, 162: 1067-1081.

[13] Xu H, Liu Y, Shu C M, et al. Cause analysis of hot work accidents based on text mining and deep learning[J]. Journal of Loss Prevention in the Process Industries, 2022, 76: 104747.

[14] Rehman A, Aslam N, Abid K, et al. The Impact of COVID-19 on E-Learning: Context-Based Sentiment Analysis Discourse Using Text Mining[J]. 2023.

[15] Müller B, Poley C, Pössel J, et al. Livivo–the vertical search engine for life sciences[J]. Datenbank-Spektrum, 2017, 17: 29-34.

[16] Spada I, Chiarello F, Barandoni S, et al. Are universities ready to deliver digital skills and competences? A text mining-based case study of marketing courses in Italy[J]. Technological Forecasting and Social Change, 2022, 182: 121869.

[17] Li J, Yang M, Liu C, et al. Listen to the Companies: Exploring BIM Job Competency Requirements by Text Mining of Recruitment Information in China[J]. Journal of Construction Engineering and Management, 2023, 149(9): 04023076.

[18] Fang F, Zhou Y. A study on recruitment of data analyst based on text mining and visualization technology[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1952(4): 042017.

[19] Li Y, Chen X, Mao T, et al. User portrait for archival talents based on recruitment[C]//2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA). IEEE, 2021: 116-120.

[20] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of machine Learning research, 2003, 3(Jan): 993-1022.