Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 1132758, 12 pages
https://doi.org/10.1155/2021/1132758

Research Article

Demand Analysis of Online Chinese Behavior Expression in Wireless Sensor Network

Zheng Liu [1,2]

[1]*School of Journalism, Fudan University, Shanghai 200433, China*
[2]*School of Humanities & Communications, Zhejiang Gongshang University, Hangzhou 310018 Zhejiang, China*

Correspondence should be addressed to Zheng Liu; lz@mail.zjgsu.edu.cn

Received 1 July 2021; Revised 14 August 2021; Accepted 4 September 2021; Published 27 September 2021

Academic Editor: Zhihan Lv

Due to the common progress and interdependence of wireless sensor networks and language, Chinese semantic analysis under wireless sensor networks has become more and more important. Although there are many research results on wireless networks and Chinese semantics, there are few researches on the influence and relationship between them. Wireless sensor networks have strong application relevance, and the key technologies that need to be solved are also different for different application backgrounds. In order to reveal the basic laws and development trends of online Chinese semantic behavior expression in the context of wireless sensor networks, this paper adopts big data analysis methods and semantic model analysis methods and constructs semantic analysis models through PLSA method calculations, so that the λ construction process conforms to this research topic. Research the accuracy and applicability of the semantic analysis model. Through word extraction of 1.05 million word data of 1,103 documents on Baidu Tieba, HowNet, and citeulike websites, the data set was integrated into a data set, and the PLSA model was verified with this data set. In addition, through the construction of the wireless sensor network, the semantic analysis results in the expression of Chinese behavior are obtained. The results show that the accuracy of the data set extracted from 1103 documents increases with the increase of the number of documents. Second, after using the PLSA model to perform semantic analysis on the data set, the accuracy of the data set is improved. Compared with traditional semantic analysis, the model and the big data analysis framework have obvious advantages. With the continuous development of Internet big data, the big data methods used to count Chinese semantics are also constantly updated, and their efficiency is constantly improving. These updated semantic analysis models and statistical methods are constantly eliminating the uncertainty of modern online Chinese. The basic laws and development trends of statistical Chinese semantics also provide new application scenarios for online Chinese behavior. It also laid a ladder for subsequent scholars.

1. Introduction

1.1. Background and Significance. Wireless sensor network is an application-related network system. At present, there is no unified software, hardware, and network protocol standards. For different application backgrounds, the key issues that need to be considered are different. The current development of information science and technology keeps pace with the times and is constantly updated, which has brought tremendous changes to the work and life patterns of human society. With the popularization of personal computers and the development of computer network technology, the Internet has become an indispensable tool worldwide. The Internet has also evolved from basic e-mails and news forums, etc., using text as the carrier of information and data to communicate with today's multiple carriers of text, pictures, audio, and video for more real-time and visual communication [1].

Context is essential for language understanding, and it is especially important for word meaning understanding, because it always affects the direction and content of online understanding of word meaning and can eliminate the uncertainty, variability, and relativity of word meaning, so that the meaning of the word is specific; there is uniqueness in discourse. From the perspective of big data, the essence of language is symbol, and the essence of symbol is data [2]. As

we all know, the three major elements of language include phonetics, semantics, and grammar, especially online Chinese, which has a wide range of aspects, complex types, and diverse semantic expressions. It is precisely because of these uncertain characteristics that big data, the Internet, and cloud computing are more needed to find the collocation and application structure of these words' semantics [3–5]. Based on the results of previous studies, this paper analyzes the semantics of online Chinese behavior expression from the perspective of big data Internet, so as to provide more abundant application scenarios for online Chinese in the future.

1.2. Related Work. Due to the importance of analyzing Chinese semantic expression under big data, many expert teams at home and abroad have carried out various researches on it and achieved good results. For example, Qiu Lin and Lu Jie start from the perspective of Chinese and personality, and through the analysis of Chinese microblog, conclude that the expression of language in personality has both universal and special parts [6]. Mou et al. from the perspective of e-commerce social language, through the literature method, the relevant research after 2016 is evaluated [7]. Sun and Wang proposed a more convenient image retrieval framework, which can clearly show the differences between images, tags, and semantics in the follow-up experiments, and we obtained very positive results [8].

In the research of semantic analysis of Chinese behavior expression, the establishment of semantic analysis model is a good method, which can optimize the traditional analysis methods and improve the quality of semantic analysis results, so it is widely used. For example, Benedetti et al. established a new computational method on the basis of specific knowledge base (such as Wikipedia), which was called context semantic analysis (CSA). The performance of this analysis method was not only better than the traditional method, but also enriched the semantic quality [9]. Ji applied the automatic lancaster semantic analysis system (USAS) to Chinese research, and the analysis results were compared with the results of the English version efficiency of the USAS system and the parts to be improved [10]. Zhan et al. established a sparse discrimination model with the help of ksvd algorithm and Kun algorithm and finally generated the optimal dictionary, which was applied to improve the level of video semantic analysis in video monitoring, which has certain reference value [11–13].

This paper analyzes and summarizes the advantages and disadvantages of the above domestic and foreign scholars' language analysis methods, gives full play to the advantages, eliminates the disadvantages, and is applied to the semantic analysis of online Chinese behavior expression under the background of big data Internet, so as to solve the problems of wide range, complex types, and diverse semantic expression of online Chinese.

1.3. Innovation. In order to solve the semantic analysis of online Chinese behavior expression under the background of big data Internet, this paper will use literature method to collect relevant professional knowledge such as semantic analysis and big data model; second, select the part beneficial to the research content of semantic analysis model and big data analysis framework, compare KD and PLSA methods, and use PLSA method as the experimental research in this paper methods: finally, under the establishment of lambda framework and PLSA model, the semantics of online Chinese behavior expression is analyzed. Through the experimental study of 1103 documents with 1.05 million words on Baidu Post Bar, HowNet, and Citeulike websites, the results show that the accuracy of word recommendation in this dataset is directly proportional to the number of documents by calculating the average absolute error of 0.636 and 0.596. The basic laws and development trends of statistical Chinese semantics also provide new application scenarios for online Chinese behavior. On the basis of lambda architecture, the performance of PLSA model is better, with the growth of subsequent users and word resources advantages of PLSA model analysis and big data statistics will be more obvious.

2. Big Data Analysis and Semantic Analysis of Online Chinese Behavior Expression

2.1. Wireless Sensor Network. As an information acquisition method, the scale of wireless sensor network can be very large (tens of thousands of sensor nodes), and nodes can dynamically join and exit the network, which can well meet the needs of experimental modal analysis multichannel data collection. The wireless sensor system mainly consists of four parts: sensor node, convergence node, Internet, and remote management center. The wireless sensor network structure is shown in Figure 1. The data is transmitted to the sink node in a multihop manner in the network, and the sink node stores and responds to the data and then forwards the data to the remote control center through the Internet or satellite, and the control center performs high-level management and monitoring of the network [14].

The sensor node is composed of four basic components: perception module, data processing module, communication module, and power supply, as shown in Figure 2.

The sensor collects information based on physical principles, converts it into digital signals, and then processes or stores them accordingly and sends them out by the wireless communication module. The whole process is powered by the power supply. In reality, wireless communication consumes much more energy than others. The sum of the energy of the two processes.

2.2. Semantic Analysis. Lexical pragmatics emphasizes the temporality, flexibility, and context dependence of word meaning in use and focuses on the dynamic change of static word meaning in use. This research idea is in line with the dynamic nature of language and helps to better reveal understanding of the meaning of a word, and the rules of its use will help to further enrich the content, methods, and theories of the study of the meaning of a word. The meaning of semantic analysis is the meaning behind a word or sentence. The generalized language analysis is to study and analyze the influence of the external environment on the language used

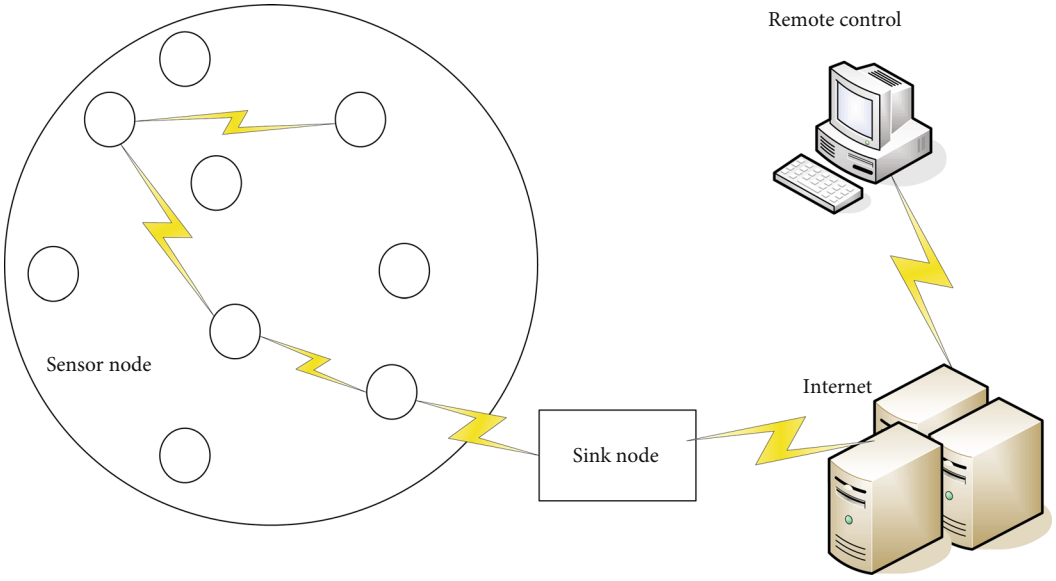

Figure 1: Network structure.

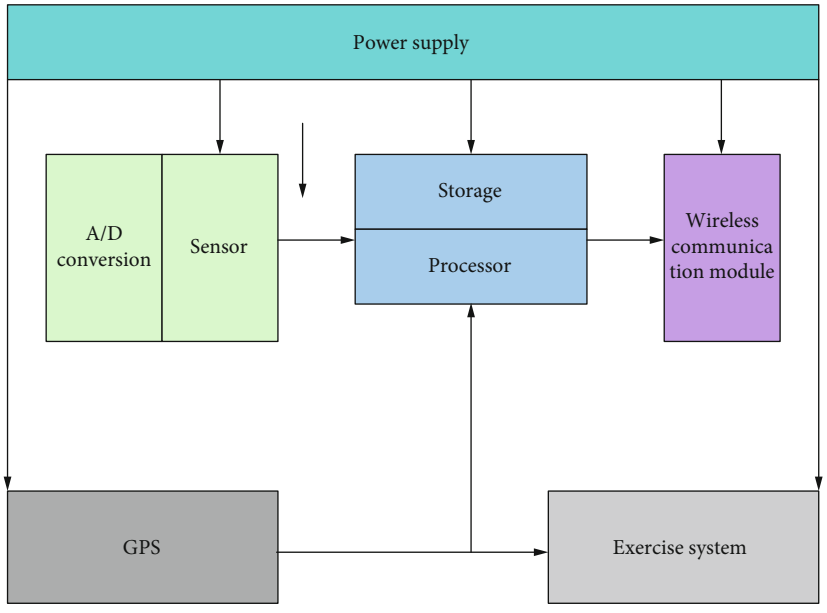

Figure 2: Node structure.

by the language users; the narrow sense language analysis is a formalized expression method of the meaning of the sentence according to the sentence's grammatical structure and the phoneme rules of each word [15]. Besides, the meaning of the sentence is not only the input analysis of syntax of line, and syntax also involves phrases and single words. The meaning is contained in a sentence or paragraph. For example, the word "painting" is known to all that it is a painting or painting behavior. They have the same spelling and form, and the single meaning is different. In the context of big data Internet, besides describing the event occurrence scenario, semantic analysis can also study the segmentation or sentence formation of words and conjunctions. Python NLTK library transforms the broken text into many phrases or special characters, then processes them to be abstract, and then put them into a large database for algorithm analysis [16].

2.3. Semantic Analysis Model

(1) Kernel dependency semantic analysis model (KD)

KD mainly uses core dependency to express sentence semantics. In essence, any sentence has core words, which

include verbs, nouns, and adjectives. Dependency is used to control these core words to express the frame semantics of sentences [17, 18]. For example, the National Health Commission said it would carry out disinfection in public places. There are two verbs in this sentence: express and carry on. The dependency of KD is these two verbs. Use these two verbs to control the whole sentence. The word "carrying on" represents the executor, place, person, or thing to be executed. Although the word "National Health Commission" is the subject of "expression" and has a master-slave relationship with "progress", in KD model, only "proceeding" can be identified as the main core word, and the core dependency in each sentence is the corpus information.

(2) Probabilistic latent semantic analysis (PLSA)

PLSA model is a faceted model of "word hidden topic document." It finds the topic information hidden in the text by identifying individual words in the text and then uses the probability method to explain the relationship among words, topics, and documents, so as to get the semantics behind the text [19, 20].

(1) The formula of the section model is as follows:

$$P(d_i, w_j) = P(d_i)P(w_j|d_i)\,; P(w_j|d_i) = \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i). \tag{1}$$

$P(z_k/d_i)$ is the probability of mapping between hidden topic z_k and document d_i.

$P(w_i/z_k)$ is the probability projection from z_k to w_i.

The posterior probabilities $P(w_i/z_k)$ and $P(z_k/d_i)$ are calculated by using the section model, and the maximum likelihood function estimation method is used to estimate these two parameters [21, 22]. The function expression is as follows:

$$\begin{aligned} L &= \sum_{i=1}^{N}\sum_{j=1}^{M} n(d_i, w_j) \log P(d_i, w_j) \\ &= \sum_{i=1}^{N} n(d_i)\left[\log p(d_i) + \sum_{j=1}^{M} \frac{n(d_i, w_j)}{n(d_i)} \log \sum_{k=1}^{K} P(w_j|z_k)P(z_k|d_i)\right]. \end{aligned} \tag{2}$$

(2) EM algorithm

EM algorithm is divided into two steps: E and M. E is to calculate the posterior probability of hidden variables, and M is to update the parameter value according to the expected value of probability maximization likelihood function obtained by E [23]. The calculation formula of step E is as follows:

$$P(z_k|d_i, w_j) = \frac{P(w_j|z_k)P(z_k|d_i)}{\sum_{l=1}^{K} P(w_j|z_l)P(z_l|d_i)}. \tag{3}$$

The M-step calculation formula is as follows:

$$E[L^C] = \sum_{i=1}^{N}\sum_{j=1}^{M} n(d_i, w_j) \sum_{k=1}^{K} P(z_k|d_i, w_j) \log [P(w_j|z_k)P(z_k|d_i)]. \tag{4}$$

Two constraints are considered $\sum_{j=1}^{M} P(w_j|z_k) = 1$and $\sum_{j=1}^{M} P(z_k|d_i) = 1$, after introducing Lagrange multiplier method, we get the following results:

$$H = E[L^c] + \sum_{k=1}^{K} \tau_k \left(1 - \sum_{j=1}^{M} P(w_i|z_k)\right) + \sum_{i=1}^{N} \rho_i \left(1 - \sum_{k=1}^{K} P(z_k|d_i)\right). \tag{5}$$

To maximize $\boldsymbol{h}$, the partial derivatives of the two parameters are the maximum likelihood estimators of the parameters obtained by solving the equation are as follows [24]:

$$P(w_j|z_k) = \frac{\sum_{i=1}^{N} n(d_w, w_j)P(z_k|d_i, w_j)}{\sum_{m=1}^{M}\sum_{i=1}^{N} n(d_i, w_m)P(z_k|d_i, w_m)}, \tag{6}$$

$$P(d_i|z_k) = \frac{\sum_{j=1}^{M} n(d_i, w_j)P(z_k|d_i, w_j)}{n(d_i)}. \tag{7}$$

In order to improve the universality of the partial model, PLSA must also test the calculated model data with a TEM algorithm, so as to control the overfitting problem that may occur in the calculation process [25]. The objective function of the algorithm is as follows:

$$\begin{aligned} F_\beta = &-\beta \sum_{i=1}^{N}\sum_{j=1}^{M} n(d_i, w_j) \sum_{k=1}^{K} \overrightarrow{P}(z_k\,; d_i, w_j) \log [P(d_i|z_k)P(w_j|z_k)P(z_k)] \\ &+ \sum_{i=1}^{N}\sum_{j=1}^{M} n(d_i, w_j) \sum_{k=1}^{K} P(z_k\,; d_i, w_j) \log P(z_k\,; d_i, w_j). \end{aligned} \tag{8}$$

This step is equivalent to M step in EM algorithm. According to the new function formula, the posterior probability of step E is deduced as follows:

$$\begin{aligned} \overrightarrow{P}(z_k\,; d_i, w_j) &= \frac{[P(z_k)P(d_i|z_k)P(w_j|z_k)]^\beta}{\sum_l [P(z_l)P(d_i|z_l)P(w_j|z_l)]^\beta} \\ &= \frac{[P(z_k d_i)P(w_j|z_k)]^\beta}{\sum_l [P(z_l|d_i)P(w_j|z_l)]^\beta}. \end{aligned} \tag{9}$$

According to this probability, we can draw the following conclusions: (1) $\boldsymbol{\beta} \longleftarrow \mathbf{1}$, execute EM algorithm; (2) $\boldsymbol{\beta} \longleftarrow$

Table 1: The number of five types of documents.

Category	Economics	Law	Politics	Entertainment	Finance
Number of documents	291	193	178	252	189
Number of sentences	3201	1544	1246	3276	1323

$\boldsymbol{\eta\beta}$ ($\boldsymbol{\beta} < \mathbf{1}$), and execute one iteration of TEM. (3) As long as the objective function value calculated by TEM is still decreasing, use this $\boldsymbol{\beta}$ value to continue iteration, otherwise, turn to the second step (4) when changing the $\boldsymbol{\beta}$ value, the program performance cannot be improved, and the $\boldsymbol{\beta}$ value is the maximum value [26].

2.4. Big Data Analysis. Big data is actually a memory bank of the Internet. In this memory, we can find, extract, and excavate valuable information and knowledge, find out the development trend, and provide strong theoretical basis for demanders to make more wise decisions. This is the essence of big data analysis [27, 28]. Therefore, when doing big data analysis, we should not only focus on the display data but also establish a scientific big data analysis framework according to their own needs, so as to automatically analyze valuable information. The big data analysis framework applied in this paper is lambda architecture.

Lambda architecture is an architecture that can meet the key characteristics of real-time big data systems. It can not only meet the computing needs of low latency but also have the ability to process full data. It can integrate offline computing and real-time computing and integrate a series of architecture principles such as immutability and read-write separation.

An important prerequisite for the construction of a big data analysis framework is whether there is a spatial correlation between variables. Normally, before the Lambda index is measured, a correlation weight model based on big data must be constructed first. The construction principles are as follows:

$$w = \begin{cases} 1, & \text{When } i \text{ is adjacent to } j, \\ 0, & \text{When } i \text{ and } j \text{ are not adjacent,} \end{cases} \tag{10}$$

$$\begin{aligned} I &= \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(xj - \bar{x})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(xi - \bar{x})^2} \\ &= \frac{n\sum_{i=1}^{n}\sum_{i\neq j}^{n} w_{ij}(x_i - \bar{x})(xj - \bar{x})}{S^2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}. \end{aligned} \tag{11}$$

The spatial dependence model expression is as follows:

$$\mathrm{y} = \alpha \mathrm{W}\mathrm{y} + \beta \mathrm{X} + \varepsilon. \tag{12}$$

The expression of the constructed spatial error model is as follows:

$$\mathrm{y} = \beta \mathrm{X} + \varepsilon, \tag{13}$$

$$\varepsilon = \lambda \mathrm{W} + \xi, \tag{14}$$

$$(\mathrm{In}\text{-}\alpha \mathrm{W})\mathrm{y} = (\mathrm{In}\text{-}\alpha \mathrm{W})\mathrm{X}\beta + \varepsilon. \tag{15}$$

Table 2: Frequency of different types of CEC.

Clause types	Locational	Directional	Possessive	Total
Number	697	206	9	884
Percentage	71.4	22.4	0.8	112

(15) It is the general expression form of spatial vocabulary analysis. The vocabulary analysis model is mainly used to reflect the influence of critical vocabulary on the vocabulary

$$\mathrm{y} = \alpha \mathrm{W}\mathrm{y} + \beta \mathrm{X}\text{-}\alpha \mathrm{W}\beta \mathrm{X} + \varepsilon. \tag{16}$$

Equation (16) can also be expressed as

$$\mathrm{y} = \beta \mathrm{X} + (\mathrm{In}\text{-}\alpha \mathrm{W})^{-1}\varepsilon. \tag{17}$$

The simplified general expression form of the lexical analysis model is

$$\mathrm{y} = \alpha W y + \beta_1 X - W\beta_2 X + \varepsilon. \tag{18}$$

Among them, $\boldsymbol{WX}$ reflects the situation where explanatory variables are added to the spatial matrix, which is used to reflect the influence of neighboring variables on the lexical dependent variable.

3. Semantic Analysis Experiment of Online Chinese Behavior Expression under the Background of Big Data

According to the previous introduction of semantic analysis model and big data analysis framework, this chapter will adopt the results of the PLSA model and lambda architecture method. The data are from 1103 posts and texts on Baidu Post Bar, HowNet, and Citeulike websites, with 1.05 million words. The contents can be divided into various types, including economy, law, finance, politics, and entertainment. The number of each type is shown in Table 1.

(1) Text preprocessing

Each different type of document is transformed into the corresponding word packet information, the words with high frequency are removed, and the remaining words are stored according to different types to form a dictionary containing 880000 words. In this way, most of these documents can be represented by words in dictionaries. The percentages in Table 2 show that positional CEC accounted for 71.4%, their directional CEC ranked second, and possessive CEC had the lowest frequency.

Table 3: Resources for sers and recommendation index.

Recommended resource type	Economics	Law	Politics	Entertainment	Finance
Recommended resource serial number	31	35	26	13	22
Recommendation index	5	5	4	4	3

Table 4: Cost model distribution.

Model	Target field resources	Cost	Original model		Self-study		ER
			Participle	Part of speech	Participle	Part of speech	
Basic model		0	89.43	82.58	90.01	83.61	5.69
	NR(T)	0	89.5	83.57	90.84	84.88	7.8
Dictionary annotation	3 K(T)	5 h	91.59	86.19	92.52	87.33	8.12
	ORACLE(T)	oo	92.76	88.53	93.66	89.57	9
	300(S)	5 h	92.25	86.52	92.99	87.51	7.19
Sentence tagging	600(S)	10 h	92.85	87.79	93.47	88.67	7.07
	900(s)	15 h	93.19	88.19	93.81	88.99	6.63
Combination of dictionary and sentence annotation	**3 K(T) + 300(S)**	10 h	93.15	88.2	93.66	88.87	5.51
	3 K(T) + 600(S)	15 h	93.64	88.93	94.27	89.53	5.25

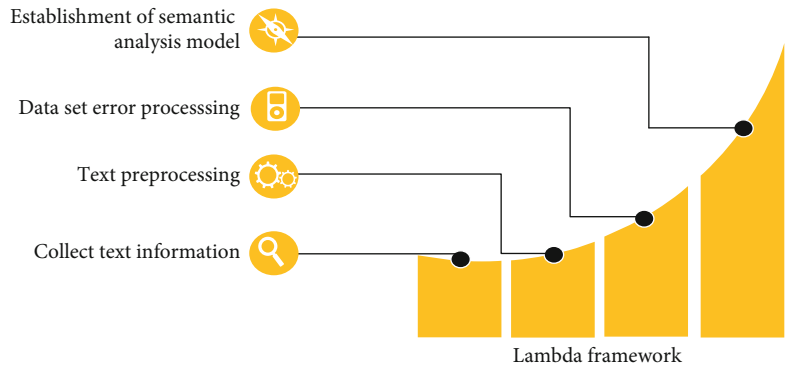

Figure 3: About the lambda framework constructed in this experiment.

For the measurement of dictionary quality, the average absolute error (MAE) of statistical measurement method is used in this experiment. The smaller the value, the higher the accuracy of the words recommended by the dictionary. MAE is calculated as follows:

$$p_1{}^* = \frac{2k}{k+1} + \frac{2c_1 + c_2 + 3et + 2et\zeta}{3}, \tag{19}$$

$$\mathrm{MAE} = \frac{\left(\sum_{i=1}^{N} |p_i - q_i|\right)}{N}. \tag{20}$$

$\boldsymbol{P_I}$ is the predicted score of the $\boldsymbol{i}$-th item, q_i is the actual score of the $\boldsymbol{i}$-th item, and N is the total number of words tested.

The dictionary is named data set N, 5 and 10 documents are randomly selected, the dictionary is used to recommend these documents, and the accuracy is 0.636 and 0.596, respectively. It can be seen that the more the number of documents, the higher the accuracy of the words recommended by the dictionary.

(2) Online Chinese semantic model

Ten users were selected to participate in the study, and the topics they were interested in were selected in the dictionary. The articles selected from Baidu Post Bar, HowNet, and citeulike websites were used as references. The serial number of each user's clicking articles was recorded to get the value of "resource result." The results are shown in Table 3:

The results show that the subject words provided by the dictionary can make users find the content with high scores and fit in each website. Second, the website will find the resources with the highest number of articles clicked by users in a keyword in the dictionary, then find the most similar resource combination of the resources, and recommend them to the similar keyword search content.

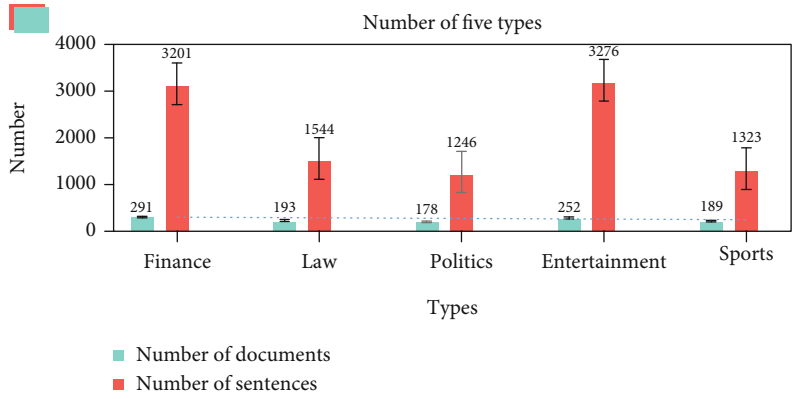

Figure 4: Number of five types.

Table 5: User resource text relationship table.

Resource type	Economics	Law	Politics	Entertainment	Finance
User similarity	0.5000	0.0001	0.5445	0.9286	0.1716
Text accuracy	0.2926	0.5742	0.6631	0.9287	0.3795
Text similarity index	1.0013	0.9972	0.9667	0.9691	0.9943

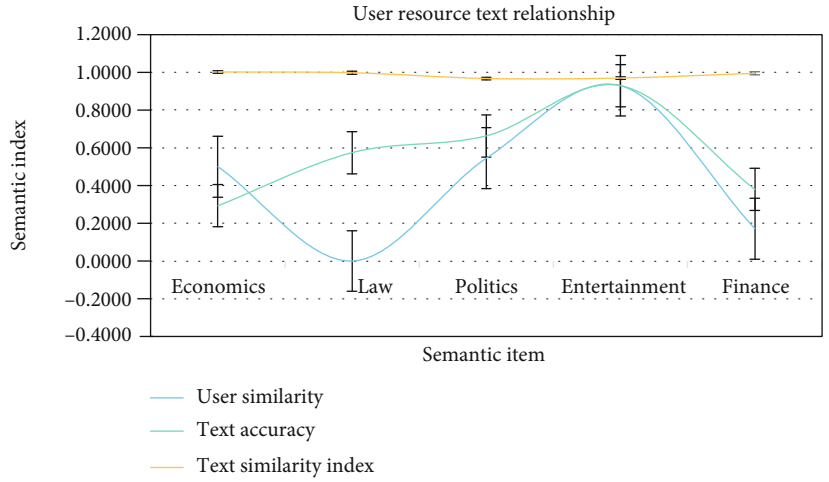

Figure 5: User resource text relationship.

From the cost data in Table 4, it is shown that dictionary annotation and sentence annotation can equivalently enhance the domain transplantation ability of the joint analysis model of word segmentation and part-of-speech tagging based on the same cost.

(3) Lambda method establishes big data analysis framework

Lambda framework consists of three parts: batch processing layer, accelerating processing layer, and merging layer. In this experiment, the MAE algorithm is used to process the error of the 830000 word dataset ***n*** converted from 1103 documents, and then the PLSA model is used to merge the dataset and the real-time resource serial number to return to the website. Thus, an effective semantic analysis model is established. The specific process is shown in Figure 3.

4. Semantic Analysis of Online Chinese Behavior Expression under the Background of Big Data Internet

4.1. Experimental Results. The research on traditional Chinese mainly focuses on grammar, words, sentences, and so on. A large number of data will produce certain errors in the research, and the research progress is also very slow. After the rise of big data analysis, it provides a lot of new analysis methods for Chinese behavior expression. The data of this research are from 1103 posts and texts on Baidu Post Bar, HowNet, and citeulike websites, with 1.05 million

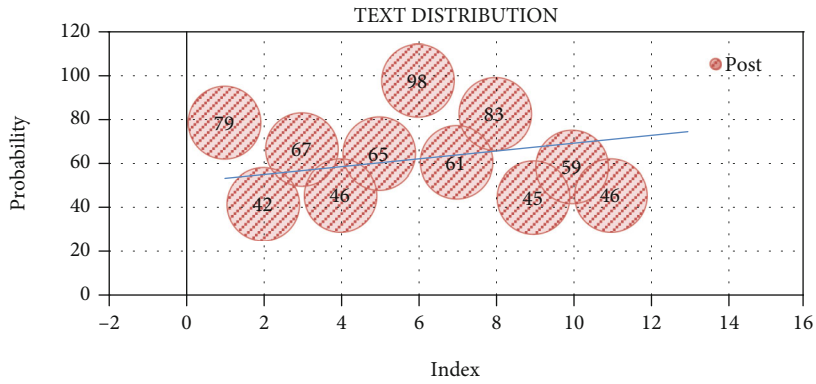

Figure 6: Text distribution.

words. The contents can be divided into various types, including economy, law, finance, politics, and entertainment, as shown in Figure 4.

In this paper, the lambda framework and PLSA model are used, and the final "user resource text" relationship is shown in Table 5 and Figure 5.

For the number of words and texts used by users, each point represents a post. According to the scatter Figure 6, it can be seen that the posts found by words with different contents are gathered together.

Through the EM algorithm of the PLSA model, we can get the accurate probability of finding the topic of the article. From Table 6 and Figure 7, we can see that most of the sections of the document are correct with only a small part of error. In order to solve the problem of repetition of similar subject words, the PLSA model makes topics merge at different levels, avoiding the problem of overfitting the content of the article, while maintaining a strong ability of topic discovery, as shown in Table 6 and Figure 7.

This experiment verifies the effectiveness of big data analysis in Chinese semantic analysis. In the case of small experimental samples, the recommendation performance of the model is better. With the increase of the number of subsequent users and word resources, the relationship table will become more extensive, and the advantages of PLSA model analysis and big data statistics will be more obvious.

4.2. Online Chinese Behavior Expression in the Context of Big Data Internet. The efficiency of Chinese semantic research based on big data statistical methods is getting higher and higher. The development of the Internet also provides new application scenarios for online Chinese behavior, and data is an efficient and intelligent research tool.

As shown in Figure 8, for word segmentation, in the corpus of big data, the system will combine large-scale characters into useful phrases according to the phrase arrangement rules provided by the semantic model in accordance with Chinese understanding. The corpus of big data will mark the part of speech of its own words, and the part of speech number of each phrase is different. When users in different regions use different systems to search for words, the system will automatically match

Table 6: Calculation results of β value and P value of test text.

Type	Economics	Law	Politics	Entertainment	Finance
β value	0.4898	0.5206	0.4704	0.4212	0.4722
PLSA (P)	0.7985	0.7080	0.8829	0.6174	0.8972

according to the part-of-speech number behind the input words.

As shown in Figure 9, for syntactic applications, grammatical dependency analysis uses semantic analysis to analyze the syntax of the entire sentence through the dependent core verbs in the sentence. The predicate verbs of the core words dominate the behavior of the subject and object, such as the school mobilizes students. In the sentence of the internship conference, the core predicate verb in the sentence is "hold," the subject is "school," the object is "conference," and the object modifier is "mobilize students for internship." After the sentence is split, it can be clearly seen. It is the "school" that held the "conference," not the "mobilization" of the "conference." Therefore, the core predicate verbs will not be dominated by the subject, object, and object complement words.

As shown in Figure 10, for the semantic position, in the article or paragraph, the position of the word may affect its importance in the article. The subject words generally appear in the beginning and the end, and most of the keywords appear in the text. According to the type and position of the words, the search satisfaction of the article can be improved. In the process of intelligent translation, information search, and Q&A, the system can automatically identify the semantics of phrases or phrases in articles or paragraphs and find content in the same field according to these semantic types. Take the common reference paper search, for example, enter a topic, and the system will automatically recommend papers on that topic and papers that match the content of the topic.

As shown in Figure 11, for text error correction, in addition to recognizing typos or typos, it also needs to automatically correct errors. For Chinese, some Chinese characters or phrases that can be used as independent individuals will be regarded as grammatical errors if they do not meet the semantic composition rules after insertion. However, for

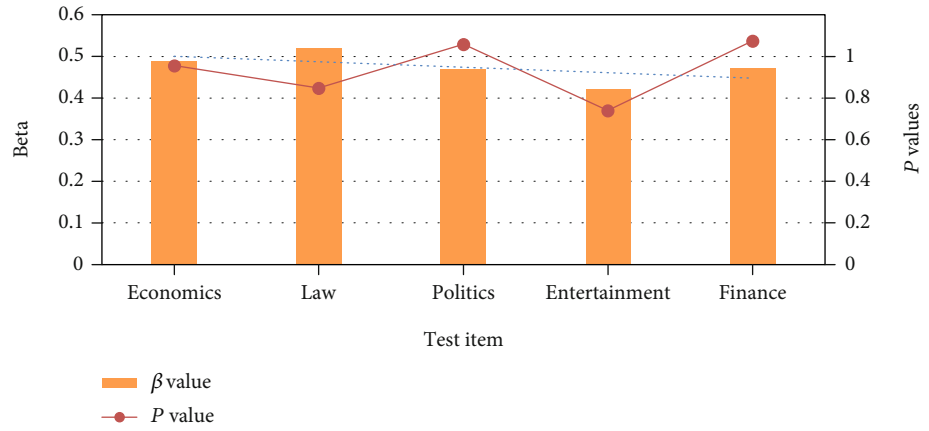

Figure 7: Beta and P values of test text.

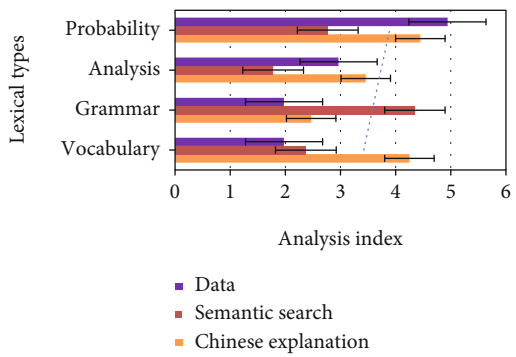

Figure 8: The corpus of big data.

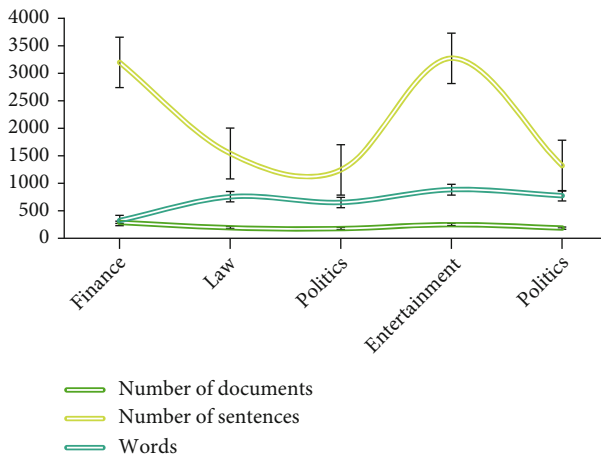

Figure 9: The syntax of the whole sentence.

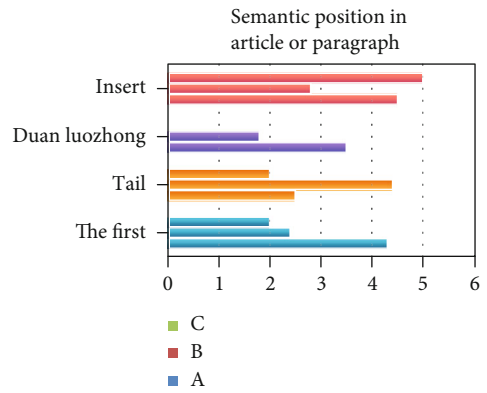

Figure 10: Semantic position in article or paragraph.

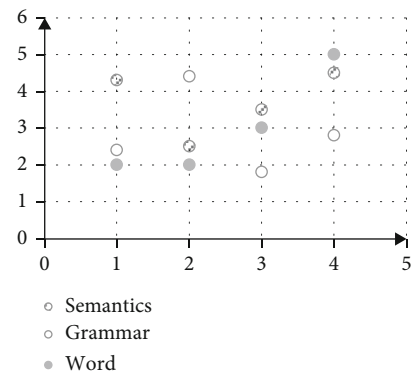

Figure 11: Error correction is not only the recognition of wrong words or phrases.

users whose native language is Chinese, the probability of sentence text error correction is very small. It is all about correcting and replacing typos.

As shown in Figure 12, the application of articles needs to extract tags. Each document will have its own unique tags. These tags are the system's classification standard for the content of the document, which is equivalent to the keywords in the paper. Of course, the tags here words or phrases that are closely related to the topic and content of the article and can run through the full text. Tags allow users to quickly grasp the main content of documents, especially in fields with strict professional knowledge such as technology, law, and finance, which are widely used. Text fitting similarity has different understandings in different fields, and the most used scenario for text fitting is the deduplication of a large amount of text. In a popular topic or field, there are a lot of texts with similar content. At this time, the fitting degree processing of these texts can greatly improve the time to

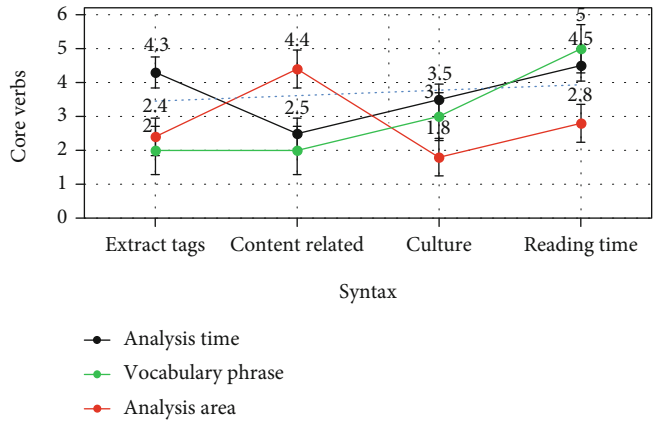

Figure 12: Dependent core verbs in the sentence to analyze the syntax.

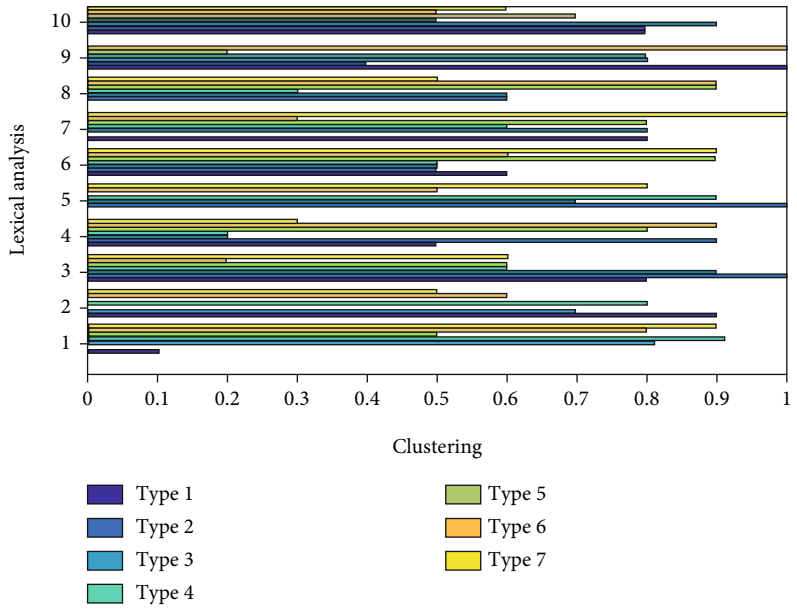

Figure 13: Topic analysis model.

read the text. For example, "how to improve work efficiency" and "how to improve work efficiency," search for these two sentences in the system, and the system will get instructions on the theme of "improve work efficiency." It is just that different users have different permutations and combinations of words, and the essence is still the same.

As shown in Figure 13, for the topic analysis model, this model is mainly a statistical model of the topic words in the article. If some specific words frequently appear in the article, it means that the article must contain this topic. Generally speaking, an article contains more than one topic, and the frequency of specific words in each topic is also different. For example, "The United States will no longer grant sanctions exemptions to certain countries and regions for importing Iranian oil in order to completely ban Iranian oil exports." Seeing the words describing the country, it can be divided into political topics, and the appearance of the two words oil and exports can also be divided into economic topics. The topic model is mainly a method of generating a topic based on the meaning of words. Each article selects a topic area based on the probability of the words in the article and then reselects a topic word from the field, that is, only one article is generated after going back and forth. Text clustering is mainly to classify documents according to specific topics. The computer automatically reads the content of the documents and assigns them to the technical system of the corresponding category. The computer reads the documents of each topic in advance and extracts the features to sort them into the resource library, then recognizes the content of the new document, and classifies it into a unified topic according to the specific subject words in the new document. The similarity of documents of the same kind is greater but different, the document similarity of the class is small, and the documents with high similarity are archived as the same class.

5. Conclusions

For the large scale and variety of Chinese language, big data can just contain it. From the perspective of big data and the Internet, we should observe the trend of contemporary Chinese application, use the information characteristics of big data to drive the direction of language research, solve linguistic problems, discover the linguistic laws that were not noticed or could not be studied in the past, and solve the "hard bones" left in Chinese grammar, which is the most urgent work at present. Therefore, this paper starts from the semantic analysis of Chinese behavior expression, adopts the average absolute error calculation and the construction of lambda framework, optimizes the data set, and reduces the performance degradation of semantic analysis model caused by big data clutter. In data processing, it uses a variety of algorithms to constantly check and eliminate errors, effectively avoiding confusion of subject information and repetition of similar topics. The data-based method provides a "user resource text" diagram, which lays a foundation for research. With the support of corpus resources, people can distinguish grammatical differences more carefully. In addition, the database can better help us to study the relationship between human language laws and cognitive laws.

The shortcoming of this paper is that in the establishment of the online Chinese model of the network under the LEACH routing protocol, the node energy consumption analysis assumes that the number of nodes in the cluster is always constant. However, the number of nodes in the network is constantly decreasing over time. Yes, the energy consumption model has changed, and the frequency of other nodes serving as cluster heads increases, which will accelerate the death of the entire network, so a dynamic energy consumption model needs to be established.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there is no conflict of interest with any financial organizations regarding the material reported in this manuscript.

Acknowledgments

This work was supported by the projects of the Social Science and Humanity on Young Fund of the Ministry of Education "Research on the Institutional History of French Communication under the New Cultural History Paradigm" (19YJC860031).

References

[1] Z. Lv, "The security of internet of drones," *Computer Communications*, vol. 148, pp. 208–214, 2019.

[2] Q.-Y. Yu and J.-R. Pei, "A two-way parallel query correction approach based on semantic analysis and reverse hidden Markov model in Chinese information processing system," *Journal of Information Hiding & Multimedia Signal Processing*, vol. 8, no. 6, pp. 1257–1266, 2017.

[3] X. Fu, X. Sun, H. Wu, L. Cui, and J. Z. Huang, "Weakly supervised topic sentiment joint model with word embeddings," *Knowledge-Based Systems*, vol. 147, no. 1, pp. 43–54, 2018.

[4] H. Xu and Z. Chen, "Review of Zhang (2014): sadness expressions in English and Chinese: Corpus linguistic contrastive semantic analysis," *Languages in Contrast*, vol. 16, no. 2, pp. 285–288, 2016.

[5] Z. Lv, Y. Han, A. K. Singh, G. Manogaran, and H. Lv, "Trustworthiness in industrial iot systems based on artificial intelligence," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 1496–1504, 2021.

[6] L. Qiu, J. Lu, J. Ramsay, S. Yang, W. Qu, and T. Zhu, "Personality expression in Chinese language use," *International Journal of Psychology*, vol. 52, no. 6, pp. 463–472, 2017.

[7] J. Mou, Y. Cui, and K. Kurcz, "Bibliometric and visualized analysis of research on major e-commerce journals using citespace," *Journal of Electronic Commerce Research*, vol. 20, no. 4, pp. 219–237, 2019.

[8] M. Sun, X. Wang, and B. Chang, "Chinese computational linguistics and natural language processing based on naturally annotated big data," in *Lecture Notes in Computer Science*, vol. 8202, no. 4, pp. 9–43, 2016.

[9] F. Benedetti, D. Beneventano, S. Bergamaschi, and G. Simonini, "Computing inter-document similarity with context semantic analysis," *Information Systems*, vol. 80, no. 2, pp. 136–147, 2018.

[10] M. Ji, "A quantitative semantic analysis of Chinese environmental media discourse," *Corpus Linguistics and Linguistic Theory*, vol. 32, no. 2, pp. 387–403, 2018.

[11] Y. Zhan, S. Dai, Q. Mao, L. Liu, and W. Sheng, "A video semantic analysis method based on kernel discriminative sparse representation and weighted KNN," *Computer Journal*, vol. 58, no. 6, pp. 1360–1372, 2015.

[12] L. Chen and W. Liu, "Chinese text content extraction method based on nltk," *Application of computer system*, vol. 28, no. 1, pp. 277–280, 2019.

[13] S. Selot, N. Tripathi, and A. S. Zadgaonkar, "Neural network model for semantic analysis of Sanskrit text," *International Journal of Natural Computing Research*, vol. 7, no. 1, pp. 1–14, 2018.

[14] J. Yan, Y. Meng, X. Yang, X. Luo, and X. Guan, "Privacy-preserving localization for underwater sensor networks via deep reinforcement learning," *IEEE Transactions on Information Forensics and Security*, 2021.

[15] Q. Jiang, F. Shao, W. Lin, K. Gu, G. Jiang, and H. Sun, "Optimizing multistage discriminative dictionaries for blind image quality assessment," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2035–2048, 2018.

[16] Z. P. Cai, Z. B. He, X. Guan, and Y. S. Li, "Collective data-sanitization for preventing sensitive information inference attacks in social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 15, no. 4, pp. 1–590, 2016.

[17] N. Rodriguez and S. Rojas–Galeano, "Discovering feature relevancy and dependency by kernel-guided probabilistic model-building evolution," *Biodata Mining*, vol. 10, no. 1, pp. 12-13, 2017.

[18] K. Winson-Geideman, "Sentiments and semantics: a review of the content analysis literature in the era of big data," *Journal of Real Estate Literature*, vol. 26, no. 1, pp. 1–12, 2018.

[19] B. Liu, "Text sentiment analysis based on CBOW model and deep learning in big data environment," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 2, pp. 451–458, 2020.

[20] L. Hou, "Analysis of internet user behavior under the background of big data," *Digital technology and Application*, vol. 37, no. 3, pp. 227-228, 2019.

[21] N. Li, W. Luo, K. Yang, F. Zhuang, Q. He, and Z. Shi, "Self-organizing weighted incremental probabilistic latent semantic analysis," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 12, pp. 1987–1998, 2018.

[22] R. C. Deo, X. Wen, and F. Qi, "A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset," *Applied Energy*, vol. 168, no. 15, pp. 568–593, 2016.

[23] Z. Xie, Z. Sun, L. Jin, H. Ni, and T. Lyons, "Learning spatial-semantic context with fully convolutional recurrent network for online handwritten Chinese text recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1903–1917, 2018.

[24] W. Yuan, P. Deng, T. Taleb, J. Wan, and C. Bi, "An unlicensed taxi identification model based on big data analysis," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1703–1713, 2016.

[25] R. Patan and M. R. Babu, "A novel performance aware real-time data handling for big data platforms on lambda architecture," *International Journal of Computer Aided Engineering and Technology*, vol. 10, no. 4, pp. 418–430, 2018.

[26] T. Numnonda, "A real-time recommendation engine using lambda architecture," *Artificial Life and Robotics*, vol. 23, no. 2, pp. 249–254, 2018.

[27] Y. Li, J. Zhao, Z. Lv, and J. Li, "Medical image fusion method by deep learning," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 21–29, 2021.

[28] M. Jahanbakht, W. Xiang, L. Hanzo, and M. R. Azghadi, "Internet of underwater things and big marine data analytics – a comprehensive survey, IEEE communications surveys & tutorials," *Second Quarter*, vol. 23, no. 2, pp. 904–956, 2021.