Method Development of Evaluating Government Public Service Performance Based on Big Data Technology Using Social Media Data as Evaluation Data

Zhan Yun*
Sichuan University,
Chengdu, Sichuan, China

Keywords: Social Media Data, Government Public Services, Performance Evaluation Methods.

Abstract. In the era of big data, it is of great significance to develop a method to evaluate the performance of government public services based on big data and technical application. This paper takes social media data as the evaluation data. It uses the natural language processing technology of big data to establish a set of special, comprehensive and systematic new methods for government public service performance evaluation. They include: (1) method of collecting social media data based on Python web crawler technology; (2) evaluation index screening method based on word frequency analysis technology; (3) method of determining index weight based on discriminant rate technique; (4) method of calculating performance score based on sentiment analysis; (5) visualization method of public service performance based on performance matrix; (6) method of mining the influencing factors of public service performance based on semantic network analysis technology. The contribution of this study is that it is of great value to improve and innovate the evaluation method of government public service performance, and has a strong promotion significance.

1 INTRODUCTION

With the advent and development of the era of big data, as a special category of big data, social media data appears widely. It includes data from content sharing websites, forums, blogs and microblogs (Farzindar & Inkpen 2020). It is characterized by easy access, more democratic, real-time generation and high level of interaction, free and so on (Arnaboldi, Coget 2016, Gao, Barbier, Goolsby 2011, Yates 2010). These advantages determine that it is of great significance to develop a set of methods to evaluate the performance of government public services using social media data. Previously, the academic community has discussed the theory and method of using social media data to evaluate the performance of public services (Bamberger 2019, Netzer, Feldman, Goldenberg, et al. 2012, Nguyen, Quan, Phan 2014). In recent years, some scholars have studied from practice. Fabra-Mata and Mygind used Twitter data to assess Norway 's contribution to the Colombian peace process (Fabra-Mata, Mygind 2019). Ceron and Negri used Twitter data to evaluate Italian policies on job market reform and school reform (Ceron, Negri 2016). Deborah and Michel used Twitter data to evaluate the performance of university public services in an Italian region (Agostino, Arnaboldi 2017). Guo and Mei used social media data to evaluate the spatial distribution and performance of mass sports facilities (Guo, Mei 2020). Existing studies have made useful explorations for the use of social media data to carry out performance evaluation, but there are simplification and unsystematic problems.

In view of this, this paper attempts to develop a set of comprehensive and systematic method to evaluate the performance of government public services using social media data. This method involves the whole process from data collection and collation to influencing factor mining. Finally, this paper reflects on the contributions and limitations of this method.

Yun, Z.
Method Development of Evaluating Government Public Service Performance Based on Big Data Technology Using Social Media Data as Evaluation Data.
DOI: 10.5220/0012069300003624
In *Proceedings of the 2nd International Conference on Public Management and Big Data Analysis (PMBDA 2022)*, pages 18-22
ISBN: 978-989-758-658-3

2 DEVELOP A METHOD TO EVALUATE GOVERNMENT PUBLIC SERVICE PERFORMANCE USING SOCIAL MEDIA DATA

2.1 Evaluation Data Collection

For the collection of evaluation data, first of all, the appropriate social media should be selected based on how widely it is used and relevance to government public services. The second is to crawl the data. There are two ways to crawl data: One is to write a web crawler program with python. Most social media platforms open their own APIs and can call python 's built-in encapsulation module requests method to crawl the required data with API. This paper takes China 's Sina Weibo as an example. After determining the search term, you can use the advanced search function of Weibo to obtain important information such as URLs and cookies, and then use python3.8 to send requests through the request library. The parsed web page data is stored in the local computer in the csv format. The microblog collection process is shown in Fig. 1. Another way is to use the data collection platform on the market, we usually only need to enter the keywords to get the relevant data. Their disadvantage is that they cannot be customized at will.

"Request" makes a request >> Get Weibo page response >> Parsing web content >> Download and save text data in csv format

Figure 1: Process of Collecting Sina Weibo Data (own-drawn).

2.2 Screening of Evaluation Indicators Based on Word Frequency Analysis

(1) Feature word extraction based on TF-IDF. TF-IDF is a weighted method, which mainly solves the problem of low word frequency and high importance. TF is the word frequency, that is, the number of times a word appears in the text; IDF is the reverse text frequency, a measure of the general importance of words; TF-IDF is the TF value multiplied by the IDF value. Their formulas are as follows:

$$TF_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

$$IDF_i = log\frac{|D|}{|\{j:t_i \in d_j\}|+1} \tag{2}$$

$$\text{TF-IDF}=\text{TF*IDF} \tag{3}$$

Select the top 500 words in TF-IDF, delete the words that obviously do not meet the evaluation characteristics, delete the numerals, verbs, emotional color words, and merge the words of regional characteristics into ' XX ' place name.

(2) Word vector acquisition based on Word2vec. To build a word vector model requires corpus training, the crawled data as a corpus or the use of existing corpus, through Python call word2vec word vector model training. Use the API interface of Gensim module to load Word2Vec and set the word vector dimension. The dimension represents the characteristics of words. The more features, the greater the discrimination of words. However, too high dimension setting may lead to errors due to insufficient computer CPU and too large dimension, which leads to the relationship between words too dilute. Thus, large corpus is generally set to 300-500 dimensions, small specific areas of the corpus is generally 200-300 dimensions.

(3) Construct evaluation index based on K-Means clustering. The K-Means algorithm uses Euclidean distance as the similarity index. The smaller the Euclidean distance, the higher the similarity of the two words. The idea of word clustering using k-Means algorithm is as follows: 1) k points are randomly selected as the clustering center; 2) Calculate the distance from each word to each cluster center; 3) Each point is divided into the nearest cluster center to form k clusters ; 4) recalculate the centroid of each cluster ; 5) Repeat the above steps until the position of the centroid does not change or the set number of iterations is reached. The core index of the elbow method is SSE (sum of the squared errors). The relationship between SSE and k is the shape of an elbow, and the k value corresponding to this elbow is the true clustering number of the data, the formula is as follows:

$$SSE = \sum_{i=1}^{k} \sum_{p \in c_i} |p - m_i|^2 \tag{4}$$

Among them, SSE is the clustering error, which represents the quality of the clustering effect. c_i represents the i-th cluster, p represents the sample point of c_i, and m_i represents the centroid of c_i.

2.3 Weight of Indicators Based on Discriminant Rate

The weight of the index can be designed by discriminant ratio. The discriminant rate determines its importance by calculating the proportion of a subset in the total set. It can judge the amount of discussion of each indicator in the discussion of the evaluated object in social media, and quantify the importance of specific indicators on social media platforms in the form of percentages. The calculation method is shown as follows:

$$\text{Relative importance} = \frac{\text{Total number of indicators discussed}}{\text{Total data volume}} * 100\% \quad (5)$$

Firstly, the index data set is extracted. The machine learning method is used to classify the text and eliminate the data unrelated to the evaluation index. At present, the traditional machine learning methods used in empirical research are SVM and TextCNN. This article also recommends a relatively simple and convenient method. It is through the ROST CM6 software keyword extraction, and then for loop respectively traverse text and keywords to achieve text extraction. The details are shown in Fig. 2.

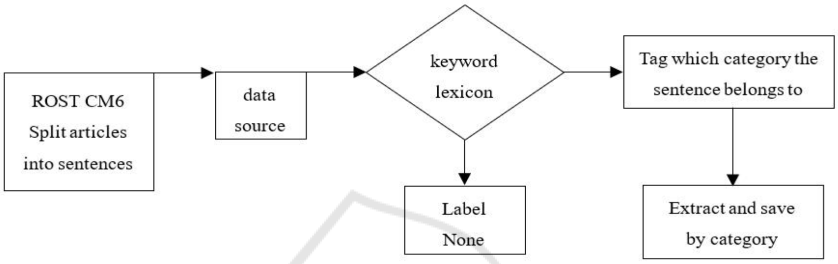

Figure 2: The process of sentiment sentence extraction (own-drawn).

Secondly, the relative importance analysis. After the classification is completed, the amount of data for each indicator can be obtained, and a clear discriminant ratio result can be obtained only by performing a percentage conversion according to the formula. In order to complete the relative importance analysis of the first-level indicators, it is necessary to convert on the basis of the discriminant rate of each second-level indicator. Because the data is constant, it is only necessary to accumulate the discriminant rate of the second-level indicators under each first-level indicator. In order to better assess the performance of government public services, a distinction should be made between official and non-official accounts in the weighting process.

2.4 Public Service Performance Score Calculation Based on Sentiment Analysis

Sentiment analysis can be used to obtain the performance score of each index, because sentiment analysis can score emotional vocabulary. At present, some institutions have established some sentiment analysis platforms or modules based on deep learning, such as the sentiment analysis module of Baidu AI open platform. The module in a number of vertical classes (cars, restaurants, hotels, etc.) emotional orientation analysis accuracy of more than 95 %. From the operation point of view, only need to call the AipNlp module, and then write the file can realize the sentiment analysis of each sentence in the file.

For the sake of objectivity, the calculation of emotional score is based on the score of unofficial account. The emotional scores of the secondary indicators of government public services are first calculated. The first-level indicator scores and the overall performance emotional scores can be weighted by the emotional scores of the secondary indicators. The calculation formula is as follows:

$$ES = \sum_{i=1}^{n}(S_i N_i) \quad (6)$$

S_i is the weight of the second-level index i, that is, the unofficial discriminant ratio of the index; N_i is the unofficial sentiment value of the second-level indicator; ES is the overall performance sentiment score for government public services.

2.5 Public Service Performance Visualization Based on Performance Matrix

For the purpose of diagnosis or result use, it is also necessary to visualize the obtained evaluation results. Drawing on the practice of Deborah and Michel (Nguyen, Quan, Phan 2014), this paper visualizes the performance of government public services by developing a performance matrix.

The matrix is visualized in the form of horizontal and vertical coordinates (as shown in Fig. 3). The horizontal axis of the matrix is the emotional score, and the vertical axis is the unofficial account ratio. These axes are centered at the median value 0.5 of the sentiment value and the average of non-official account ratios. Each point in the matrix represents a discussion topic (second level evaluation index). Four quadrants can be determined in this way, which correspond to four regions: high performance region, risk region, potential high-performance region and vigilance region. They represent: services that governments do best, services that require immediate government intervention, services that require greater government advocacy, services that require government surveillance. Through the performance matrix, we can not only intuitively observe the effectiveness of public services, but also help to determine the priority of improvement actions.

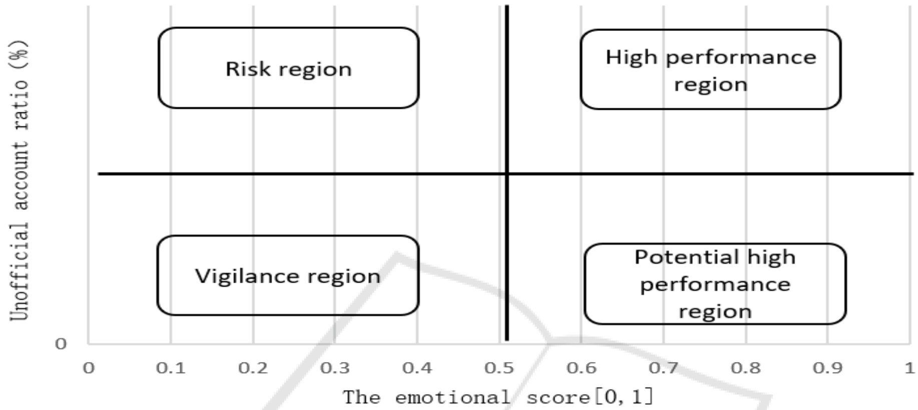

Figure 3: Government Public Service Performance Matrix (own-drawn).

2.6 Analysis of Factors Affecting Public Service Performance Based on Semantic Network Analysis

In order to have a deeper understanding of the performance evaluation results of government public services, it is also necessary to explore the influencing factors hidden behind the performance scores. Semantic network analysis in big data analysis provides a solution for the analysis of influencing factors. Semantic network analysis is carried out after the word frequency analysis of the comment text. The word frequency analysis has been completed in the index system part, so it only needs to be carried out based on the previous work. When performing semantic network analysis of government public services, the following process can be followed: selecting unofficial account comments - classifying positive and negative entries - counting word frequency separately - drawing word cloud maps separately - drawing semantic network maps separately - observing and summarizing the links between feature words. In operation, the first step is to use the jieba word segmentation module of Python to cut the positive and negative entries and delete the stop words, and then use the collections word frequency statistics library for word frequency statistics, and then use the easy word cloud 3.0 to draw the word cloud map. The second step is to use the ROST tool to import the positive and negative text data of each index, and draw the semantic network diagram.

3 CONCLUSIONS

Based on the grasp of big data methods and public sector performance evaluation methods, this paper develops a set of methods for evaluating government public service performance using social media data through expert consultation and combining existing research results. It includes the methods for collection and collation of media data, the evaluation index screening method based on word frequency analysis, the index weight establishment method based on discriminant rate, the performance score calculation method based on sentiment analysis, the public service performance visualization based on performance matrix and the public service performance influencing factors mining method based on semantic network diagram.

However, the current method still has room for improvement: Firstly, due to the complexity and confusion of the data, it may be difficult to achieve good elimination and retention when cleaning. Secondly, in terms of data analysis, there may be more appropriate and effective ways to obtain assessment results and explore influencing factors. These limitations need to be further studied in the future.

REFERENCES

Anna Atefeh Farzindar & Diana Inkpen. Natural Language Processing for Social Media, Third Edition [M]. San Rafael: Morgan & Claypool Publishers, 2020: 2-3.

Arnaboldi M and Coget J. Social media and business: We’ve been asking the wrong question! [J]. Organizational Dynamics, 2016, 45(1), 47-54.

Bamberger M. Integrating big data into the monitoring and evaluation of development programmes [EB/OL]. [2019-11-15]. http:// unglobalpulse.org/ sites/default/ files/IntegratingBigData_intoMEDP_web_UNGP.pdf.

Ceron, A., Negri, F. The “Social Side” of Public Policy: Monitoring Online Public Opinion and Its Mobilization During the Policy Cycle[J]. Policy & Internet, 2016, 8(02): 131-147.

Dave Yates, Scott Paquette. Emergency knowledge management and social media technologies: A case study of the 2010 Haitian earthquake[J]. Proceedings of the American Society for Information Science and Technology, 2010, 47(1), 1-9.

Deborah Agostino, Michela Arnaboldi. Social media data used in the measurement of public services effectiveness: Empirical evidence from Twitter in higher education institutions[J]. Public Policy and Administration, 2017, 32(04), 296-322.

Gao H, Barbier G and Goolsby R. Harnessing the crowdsourcing power of social media for disaster relief[J]. IEEE Intelligent Systems, 2011, 26(3), 10-14.

Guo, Qi & Mei, Hongyuan. Research on spatial distribution and performance evaluation of mass sports facilities based on big data of social media: A case study of harbin[C]. CAADRIA, 2020, 537-546.

Javier Fabra-Mata and Jesper Mygind. Big data in evaluation: Experiences from using Twitter analysis to evaluate Norway’s contribution to the peace process in Colombia[J]. Evaluation, 2019, 25(1), 6-22.

Netzer O, Feldman R, Goldenberg J, et al. Mine your own business: Market-structure surveillance through text mining[J]. Marketing Science, 2012, 31(03): 521-543.

Nguyen T T, Quan T T, Phan T T. Sentiment search: An emerging trend on social media monitoring systems[J]. Aslib Journal of Information Management, 2014, 66(05):553-580.