A Public Opinion Analysis System Based on Emotion Analysis in Linyi

Zhou Jiaqian*, Wang Jihou, Li Qintao, Wen Yulian, Zhang Yan, Lin Peiguang
School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China

*ww17853131657@163.com

Abstract. This article, the official news site of Linyi five services in Shanghai, Yiwu GanZhou Shenzhen national logistics hub of news as the data source, through Word2Vec and construction LSTM emotion classification model, with positive or negative emotion in general categories, calculating to analyze its emotional value, from the emotional category and time series analysis and word frequency vector to Linyi public opinion analysis of logistics hub

keyword: logistics hinge, public opinion analysis, sentiment analysis, Linyi, deep learning.

1. Introduction
Linyi trade service-oriented country logistics hub, located in Linyi mallcore, is the National Development and Reform Commission and the Ministry of Transport in September 2019 established the first joint one of 23 national logistics hub, aimed at improving the logistics hub of the state function-driven radiation ability and industrial supporting ability, strengthen the connectivity and collaboration between the national logistics hub, form the hub + + channel network of new modern logistics service system, driving the development of various factors of an industry cluster, fostering regional new economic growth pole.

At present, there are altogether 5 business service hubs in China, which are located in Shanghai, Jinhua, Zhejiang, Yiwu, JiangXi, GanZhou, Linyi, Shandong, and Shenzhen, Guangdong.

To sum up, scholars discussed the topic of public opinion from the two dimensions of communication and management, network public opinion structure, etc. Therefore, this paper will integrate public opinion and emotion in Linyi, select the news content of the official website, and provide theoretical guidance for public opinion analysis in Linyi from the following three aspects.
1) The emotion classification model was constructed based on Word2Vec and LSTM methods to analyze the emotion of official news in a more fine-grained way.

2) Time series method is used to analyze the importance of Linyi logistics hub.

3) Carry out word frequency statistics and analyze the similarities and differences between Linyi logistics hub and other different regions.

2. Related work
By crawling news data, emotion analysis mined the evolution characteristics of public emotion, and then predicted the behavior change trend of various subject areas, such as Liu tracking the movie comments on Twitter and using OpinionParser system to analyze the positive and negative opinions of each movie, so as to predict the movie box office[1]. Bermingham et al. took the Irish general election as an example and predicted the political election results by monitoring the public sentiment on Twitter[2]. Ragini et al. combed the disaster data on Twitter, classified the disaster data by a machine learning algorithm, analyzed the emotional trend of the affected people, and then proposed disaster emergency response strategies.

Domestic scholars are more committed to discussing the theories and methods of emotion analysis on Chinese microblog. For example, Han Ping et al. proposed an emotion analysis model e-disan based on emotion fusion and multidimensional self-attention mechanism[3]. Amy, etc. Use a variety of text characteristic, based on the SVM and CRF characteristics of different combination, make the emotional analysis effect optimal[4] in addition, for the use of public opinion analysis method focusing on emotional dictionary[5] convolution neural network[6] and so on, and explore the weibo emotional change trend, to penetrate algorithm in solving practical case study of literature are rare.

Public opinion analysis covers communication, sociology, management, communication mathematics, library and information science, and other disciplines, with diverse research perspectives. Foreign scholars put forward the model of public opinion evolution stage and used a complex network, random network, and other methods to build the dynamic network model of public opinion communication[7]. Emotion analysis is the extraction, detection, and classification of the theme views, feelings, and attitudes, which was first proposed by Nasukawa et al.[8]. The main tasks of affective analysis include subjective and objective analysis[9], affective tendency analysis (affective classification), viewpoint information extraction, comment mining, etc., among which affective classification is the most commonly used.

3. Model
This paper makes a comparative analysis of the development prospect of Linyi logistics hub by crawling the news of the logistics hub on the official website, carrying out emotional analysis, word frequency statistics, and time analysis.
3.1 Data cleaning

The data preprocessing of this paper is mainly to clean, screen, remove noise and irrelevant content from the original data of published news, so as to obtain high-quality data and make the subsequent emotional analysis more accurate.

First of all, the data should be cleaned and the stock comments with missing or high missing ratios should be deleted. Then, the precise mode of jieba word segmentation tool is used to divide the text data, remove line breaks, delete irrelevant vectors to remove noise, and create a word dictionary, returning the index of each word, word vector, and the word index corresponding to each sentence, laying the foundation for the subsequent emotional analysis. Jieba participles come in three modes: precision mode, full mode, and search engine mode. Among them, the accurate mode can be used to slice sentences in the most accurate way, which is often used for emotion analysis.

3.2 Sentiment analysis

In essence, emotion classification is actually a three-classification task. We made a mark for each title, namely, \( W \in \{P, N, O\} \), (\( P \) stands for positive, \( N \) stands for negative, \( O \) stands for neutral). Then the label for the positive sample is 1, the label for the neutral sample is 0, and the label for the negative sample is -1. The process is shown in Fig.2:
The text analysis of the title is divided into two modules: data preprocessing and emotion analysis. In the "Preprocessing" module, we first cleaned the header text, removed some dirty data and noise, and then used the jieba tool for word segmentation and dictionary creation.

When we processed the title, we chose the exact model of jieba word segmentation and the CBOW training model of word2vec to generate the word vector and return the index of each word, word vector, and the word index corresponding to each sentence. The process is as follows:

1. Create a mapping of words to indexes
2. Create a mapping of words to vectors
3. Transform the training test dictionary

Preprocessing module, we will be more than 10 indexes and word frequency vector was recorded, then, we used the LSTM layer to extract text information characteristics, emotional analysis, extract the characteristics of the information in the input neurons for emotion classification characteristic value, the final output in the output layer using softmax function emotional categories, positive emotions is 1, the ordinary is 0, negative emotions to 1, the final assignment into emotional weight, and calculate the value of each title, the specific calculation process is shown below.

\[ f_t = \sigma(W_f [h_{t-1}, x_t] + b_f) \]  \hspace{1cm} (1)
\[ i_t = \sigma(W_i [h_{t-1}, x_t] + b_i) \]  \hspace{1cm} (2)
\[ \tilde{C}_t = \tanh(W_c [h_{t-1}, x_t] + b_c) \]  \hspace{1cm} (3)
\[ C_t = f_t C_{t-1} + i_t \tilde{C}_t \]  \hspace{1cm} (4)
\[ o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \]  \hspace{1cm} (5)

Among them, \( \sigma \) says the sigmoid function, \( \tanh \) is the hyperbolic tangent function, \( f_t \) and \( i_t \) represent the forgotten the door and enter the door, \( C_t \) said a moment before state, \( C_t \) said t moment, at this point that corresponds to review \( C_t \) said this is emotional category label \( \in (-1,0, 1) \), \( o_t \) for the emotional category label calculation.
output, at this point \( o_t \) represent the emotional value of each title, \( o_t \in [0, 1] \), the to 1 title says this is intensely emotional, on the contrary, that title has said this is an emotion for a little proportion.

3.3. Word frequency analysis
Word frequency is of high frequency in mass text "key words", this paper USES wordcloud package, to crawl five trade service-oriented country logistics hub of news information for word processing and word frequency statistics, get a lot of "key words", then according to the number of occurrences of keywords ranking, the more significant "key words", appear in all the content of the higher the frequency, and the heat needs words.

4. Experiments and Results

4.1. Data set
Based on the search contents in CCTV1, People's Daily, People.com.cn, Guangming.com, Chinanenews.com, Ministry of Commerce, China Media Group, and CnR.cn, we designed python crawler and collected the relevant press releases from 2017 to 2020 of a total of 5 domestic business service hubs, including Linyi. We have access to the Linyi trade service-oriented country logistics hub reports 155, Shanghai trade service-oriented country logistics hub reports 715, Jinhua, Yiwu trade service-oriented country logistics hub article 38, GanZhou in Jiangxi province trade service-oriented country logistics hub related to article 54, trade service-oriented country logistics hub in Shenzhen, Guangdong reported 79 articles, as shown in Tab.1.

| City         | Linyi | Shanghai | Shenzhen | GanZhou | Yiwu |
|--------------|-------|----------|----------|---------|------|
| original     | 29.4% | 35.9%    | 41.7%    | 34.4%   | 26.3%|
| reprinted    | 70.6% | 64.1%    | 58.3%    | 65.6%   | 73.7%|
| number of articles | 155 | 715      | 79       | 54      | 38   |

Tab.1 Statistics of the number of news articles in each city

4.2. Data distribution analysis
It can be seen from Tab.1 that The Times of Linyi's news reports are far more than those of YiWu, GanZhou, and Shenzhen, but relatively speaking, Linyi's news accounts for more reprints. As shown in Fig.3, within the time period from 2017 to present, YiWu and GanZhou there are different forms of state-run media reports, and YiWu has focused on after entering the first national list of logistics hub construction, Linyi state-run media interview directly is relatively small, mainly the Linyi local media reports, only a small number of media reports, at the provincial level in central Europe in the report on the trains other trade service-oriented country logistics hub.
4.3. Sentiment analysis

Face our crawl data sets, can be divided into "title" and "content" of two parts, the title as a summary of an article, very concise the core content of extract the most important of all, we first of all, from the news title of each logistics hub, text analysis, the title of the news of crawl sentiment analysis. The output results are shown in Tab.2:

| Title                                                                 | Emotional categories | Emotional weight  |
|----------------------------------------------------------------------|----------------------|-------------------|
| Let villagers harvest more happiness                                 | 1                    | 0.994217476       |
| Investing more than 50 million to build online wisdom Why Linyi Mall | 1                    | 0.998356925       |
| becomes an "empty shelf"                                             |                      |                   |
| Linyi, Shandong live with goods "with fire" mall economy             | 0                    | 0.653022813       |
| ...                                                                  | ...                  | ...              |

Tab.2 Statistics of affective categories in different regions

4.3.1. Affective categorization analysis

For the exported content, we used the COUNTIF function in the table document to analyze the "emotion categories" of each region. The results are as follows:

| Place   | Linyi | Shenzhen | Yiwu | GanZhou | ShangHai |
|---------|-------|----------|------|---------|----------|
| Positive| 74.8% | 53.2%    | 65.8%| 43.4%   | 61.2%    |
| Ordinary| 22.6% | 46.8%    | 34.2%| 54.7%   | 37%      |
| Negative| 2.6%  | 0%       | 0%   | 1.9%    | 1.8%     |

Tab.3 Statistics of affective categories in different regions
It can be seen from Tab.3 that the emotional categories of the five cities are mainly positive, while the negative ones only account for a small number or almost none. Relatively speaking, Linyi has the largest proportion of "positive" emotions, indicating that the government and people generally hold a positive attitude towards the establishment and development prospect of the national logistics hub of business and service.

We conducted a statistical analysis of the negative headlines in each city.

As shown in Tab.4, the main negative factors in GanZhou come from unhappy consumers in the express delivery process. The main negative factors in Linyi come from online shopping scams and other aspects related to financial interests; Shanghai's negative factors focus on "safety", "waste", "broken promise", "cost" and so on.

4.3.2. Affective intensity contrast

In order to more intuitively extract the enthusiasm of each logistics hub city, we summarized the descending statistics of the emotional weight of each title in different cities in the emotional analysis, as shown in the figure below:

| City     | Title                                                                 |
|----------|-----------------------------------------------------------------------|
| GanZhou  | False sign-in, post loss... How can express delivery make consumers "unhappy"? |
| Linyi    | Two swindlers who release false information about big truck drivers harm others and harm themselves |
|          | New online payment scam: Customer service can't provide "gift card" where to spend |
| ShangHai | Two couriers were poisoned when hazardous chemicals were sent by Courier, the Shanghai Minhang Procuratorate suggested |
|          | From the field to the table, food waste is appalling! We should still advocate saving food after bumper harvests! |
|          | Domestic steel PMI fell to 52.0% in September, and the industry sentiment has eased somewhat |

**Tab.4 Negative headlines in cities**

![Fig.4 YiWu emotional level map](image_url)

![Fig.5 GanZhou Emotional level map](image_url)
It can be seen from Fig. 4-8 that the emotional part takes up a large part of the overall situation in any city. Among the news headlines in Linyi, emotional headlines accounted for the largest proportion of the five cities, and the data of emotional strength and weakness accounted for the least, indicating that the country and the people have great expectations for the construction of Linyi logistics hub.

4.4. Word frequency

In this paper, we output the words in front of the word frequency of news headlines in five cities for analysis.

|          | GanZhou | Shenzhen | ShangHai | Linyi  | Yiwu   |
|----------|---------|----------|----------|--------|--------|
| Jiangxi  | Shenzhen| Shanghai | shandong | yiwu   |
| trains   | logistics| logistics| linyi    | zhejiang|
| the Belt and Road | The international area | logistics | The Belt and Road |
| China and the eu hub | area | The development of mall | The world |
|          | The customs | new | struggle | business |

Tab.5 Statistics of frequent words in each city

As shown in Tab.5, the hot words in each region are different according to local conditions, and the emphasis of logistics hubs in each location is also different. For example, Yiwu in Zhejiang province has prominent "foreign trade", "e-commerce", "import" and so on, which can be corresponding to YiWu's famous "small commodity wholesale market". GanZhou focuses on "Network media"; Compared with other cities' logistics hubs, Shenzhen's "international" and "customs" are quite conspicuous. ShangHai's "enterprises" are relatively prominent; In contrast, Linyi focuses on "mall","
"construction" and "struggle", "help" and so on, relative to ShangHai and Shenzhen's profound economic foundation, Linyi's development and construction are relatively weak, although there is a certain gap, it can be used as Linyi logistics hub construction direction.

5. Conclusion
In the context of big data, emotion analysis is one of the research hotspots in various fields at present. This paper takes the public sentiment analysis of "Linyi Logistics hub" as an example, collects news headlines and contents as data sources, builds an emotional analysis model of public sentiment analysis of logistics hubs based on Word2Vec and LSTM, and analyzes the emotional preference and time distribution of five logistics hubs. Wordcloud is used to export the news information of five national logistics hubs of commercial service for word segmentation and word frequency statistics, analyze the hot words of each logistics hub, discuss the similarities and differences, advantages and disadvantages between Linyi logistics hubs and other regions, and put forward some suggestions. In the future, the data volume will be further increased, and relevant problems will be discussed and analyzed in a more comprehensive way to improve accuracy and comprehensiveness.

References
[1] Bing L 2012 Sentiment analysis and opinion mining Synthesis Lectures on Human Language Technologies 5 pp 1-167
[2] Be R and Smeaton A 2011 On using Twitter to monitor olitical sentiment and predict election results SAAIP pp 2-10
[3] Pin H, Jiahui S, Cheng F and Yunfei Jia 2015 An emotion analysis based on emotion fusion and multi-dimensional self-attention mechanism Journal of Computer Applications 32 pp 978-981
[4] Tingting L and Donghong J 2015 Microblog emotion analysis based on SVM and CRF multi-feature combination Application Research of Computers 32 pp 978-981
[5] Lu A and Lin W 2017 An analysis on the evolution of Public opinions on Weibo for emergencies integrating themes and emotional characteristics Library And Information Service 15 pp 120-12
[6] Haitao Z, Dan W, Hailing X and Siyang S 2018 Research on the Classification of Weibo Public Opinion Emotions based on convolutional neural network Journal of The China Society for Scientific and Technical Information 37 pp 47-54
[7] Yulia T, Hao L, Heng J and William W 2013 Evolution of communities on Twitter and the role of their leaders during emergencies IEEE pp 727-733
[8] Yi J, Nasukawa T, Bunescu R and Niblack W 2003 Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques IEEE pp 427-434
[9] Wiebe J, Bruce R and O’hara T 1999 Development and use of a gold-standard data set for subjectivity classifications Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics pp 246-253