Sentiment Analysis Through Evaluation of logistics support company Customer Reviews

MD RABIUL AWAL SHUVO
Department of Software Engineering, Daffodil International University

ABSTRACT

The development of modern logistics has become a powerful and powerful enterprise for the construction of smart cities. Supervision of fine suppliers can help change sales, and the development of logistics companies can improve the city's public goods. In this study, we plan an optimal assessment of logistics service providers mainly through a combination of real-time analysis perspective and traditional assessment perspective. With the help of sensitivity analysis, this version aims to extract male or female characteristics of service delivery quality from user opinions and to obtain information about customers' use of certain products. By improving the SERVQUAL version, we are creating a new logistics product quality measurement system. The tool used the assessment tool to my content rating data by removing features and creating family members in content rating and rating. We also use opinion-based evaluation, with features shown to clearly compare employees' world-class work. Experimental analysis shows that the developed models and methods are very accurate.

Keywords*: smart city; sentiment analysis; logistics service quality; feature extraction*

1.Introduction

As the infrastructure at the bottom degree of society, logistics not exceptional connects manufacturing and consumption but also supports the pillar industries and extraordinary industries of the city. currently, the rapid popularity of purchasing has extensively. Promoted the improvement of the logistics issuer enterprise. Logistics offerings, which are vital components of online buying, are without delay associated with consumer satisfaction and affect purchase intentions [1,2]. To awareness on improving their non-public carrier stage and to attract extra customers, many e-trade agencies usually outsource their logistics offerings to one/3-party agencies due to the truth they may be now not middle aggressive. but, customers have the propensity to treat the whole way of purchasing as a whole, and their attitude in the route of the fine of logistics offerings without delay impacts their attitudes inside the route of the issuer and commodity pleasant, thereby affecting the earnings of commodities [3]. Consumers' attitudes closer to logistics service have ended up being a vital element affecting the improvement of e-trade.

In the current logistics service, the competition between logistics service companies is very fierce and the price structure of logistics services is becoming more and more the same. Choosing a high-service shipping company is a concern of customers [4,5]. To build trust with customers, most e-commerce businesses pay close attention to customer feedback on quality delivery. Most studies allow users to evaluate the quality of logistics services by creating a benchmarking system. According to the evaluation model [6-9], the quality level of the logistics service can be evaluated by analyzing the scores of the users and appropriate logistics service can be offered to the customers. However, good logistics services have many aspects, and different users will have different

experiences. Index-based evaluation can only score and evaluate the quality of logistics services. Chapter However, it is difficult to use certain factors to explain the advantages and disadvantages of users Chapter. Sentiment analysis, also known as sentiment mining and directional analysis, is an important branch of natural language processing. The main function of analytical thinking is to assist people who use computer resources to obtain, organize, and analyze information about the subject, including intelligent analysis of data files, processing, writing content and reasons [10]. Methods based on sentiment analysis have been used in many fields such as social networking [11], e-commerce [12], online demographic analysis [13], and others. In the operation of logistics services, e-commerce platforms have collected a lot of customer evaluation data on the quality of logistics services. Reviews often describe various aspects of logistics services in the literature that can affect the quality. It is difficult to manually process data on such a large scale; therefore, automated machine analysis methods should be used. With the help of thinking technology research, the emotional characteristics of the words related to logistics service quality can be extracted to help the platform analyze user behavior regarding various logistics services and explain the quality of logistics services. This method is very important to help users choose the best logistics service and improve the service level of businesses.

2 Related Works

This chapter focuses on user sentiment mining research, especially for sentiment analysis in online services. Various machine learning studies and dictionary-oriented methods were also researched and analyzed to form the basis of this study.

2.1 Customer Opinion Mining in online services

The development of technology and the widespread use of social media have created opportunities to obtain information from unstructured data. Opinion mining in big data is used to classify the opinions of customers with different opinions and to measure the needs of people. At the same time, thought mining has produced surprising results according to many reviews on the web. Customers, restaurants, schools, hospitals, resorts, etc. where they share their views on products and services. The value of user reviews, comments or ratings for a particular product or service is their opinion, judgment, opinion or opinion about their quality, appearance or price. Thoughts can be positive or negative depending on the person's point of view.

One of the limitations of machine learning-based methods is that it depends on the size of the training data, which must be tagged and large enough. However, registration information is often scarce, especially in some narrow specialties. Research groups often have to spend time and money collecting data.

2.2 Machine learning-based customer sentiment analysis

Sentiment and thoughts are the subject of interest and research for many researchers (Akter et al., 2016; Lugović et al., 2016). Therefore, there is a difference of opinion. Depending on the nature of the emotions, emotions can be divided into two groups: positive emotions and negative emotions. Depending on how and where we express them, we can divide emotions into six categories: happiness, sadness, anger, surprise, disgust, and fear. Under the influence of different situations, different environments, and different emotions, people's emotions sometimes get mixed up, mixed, and combined. This makes its owner the owner of other thoughts.

In general, emotional analysis is defined as "the scientific study of thoughts, feelings and emotions represented in texts" (Nagpal et al., 2020). In other words, opinion mining, as a way to get the opinions of the people who created a particular document, has recently become the most popular research topic in relation to relationship (Pang et al., 2008; Ohana et al., 2009). With the rise of social media such as reviews, forums, and social media, opinion polls have grown in importance.

2.3 A classification algorithm

SVM is a machine learning taxonomy that uses kernel functions to map the location of nonlinearly allocated data points to new locations with misclassifications. We refer the reader to Burges (1998) for an explanation of SVM and its design principles. Detailed information on the use of this model for text classification can be found in Joachims (2002).

SVM is essentially an optimization problem; The purpose of the algorithm is to find the location of F and the overall decision f that minimizes the classification error of F. Let the model set $\{(x_1, y_1),(x_2,y_2),...(x_f,y_f)\}$ and $x_i \in R^n$ in two ways as follows: $y_i \in \{-1,1\}$, class anchor tag x_i (-1 for class I, 1 for class II). We have the vector xi in space of the equation hyperplane: $x_i.w + b = 0$

Set $f(x_i) = \text{sign}(x_i.w + b) = \begin{cases} +1, xi.w + b > 0 \\ -1, xi.w + b < 0 \end{cases}$

Reference	Algorithm / Classifier	Data collection	Accuracy	Comparism Proposed Methodology
Bang Nguyen	Support Vector Machine	Social Media	81%	SVM 0.89%
Van-Ho Nguyen	Naive bayes	Social Media	49%	Naive bayes 0.52%
Thanh Ho	Decision Tree	social Media	89%	KNN 0.75%,
Burhan BİLEN	SVM	Social Media	**N/A**	SVM 0.84%
Fahrettin HORASAN	Random Forest	Website	0.87 %	Random Forest 0.87%
Chao Wei	LSTM	websites	90%	LSTM 90%,
Dingkai Zhang	Random forest.	website	0.87%	Random Forest 0.87%

Table 1 Comparative Analysis and Summary

3. MATERIALS AND METHODS

The collection of data is the initial step in the pre-processing process. As a basis, I began by gathering information through the Facebook Review section, Logistics department website review. I am collected around 10,000 data through different logistics website.

Following that, each process must use the dataset. Since each of the algorithms we use behaves differently, as we have already demonstrated, we must preprocess the data in order to make it fit into our model correctly.

Neural Network:

Neural networks are an artificial intelligence method that teaches computers to process data inspired by the human brain. It is a machine learning process called deep learning that uses cells or neurons connected in a hierarchical structure similar to the human brain. It creates an adaptive system that computers can use to learn from their mistakes and continually improve. Therefore, neural networks try to solve complex problems such as gathering information or recognizing faces with higher accuracy. A simple neural network has three layers of artificial neurons interconnected:

Input Layer: Information from the outside world enters the neural network through the input layer. Input nodes process, analyze or classify data and transmit it to the next layer.

Hidden Layers: hidden layers provide input from the input layer or other hidden layers. Neural networks can have many hidden layers. Each layer analyzes the output of the previous layer, processes it further and forwards it to the next layer.

Output Layer: Output operation provides the final result of all data processed by the neural network. It can be single or multiple. For example, if we have a binary classification problem (yes/no), the output operation will output a 1 or a 0. However, if we have a multi-class classification problem, the output method has multiple outputs.

LSTM

LSTM networks are deep learning neural networks that allow continuous data processing. It is a special type of neural network that can solve the gradient problem solved by RNNs. LSTM was developed by Hochreiter and Schmidhuber to solve problems caused by traditional RNN and machine learning algorithms. LSTMs can be implemented in Python using the Keras library. For example, you remember the previous event while watching a video, or you know what happened in the previous chapter while reading a book. RNNs work similarly; they remember past information and use it to form current ideas. The disadvantage of RNNs is that they cannot remember long-term prospects due to loss. LSTMs are specifically designed to avoid long-term dependency issues.

The RNN approach is embedded in LSTM by adding LSTM cells to the RNN architecture. LSTMs take longer to calculate than RNNs because LSTMs store weights. On long data, LSTM may outperform RNN because the LSTM cell has a self-loop. There are three main elements in the LSTM processing stage: the memory gate, the input gate, and the output gate. Figure 3 is the LSTM architecture.

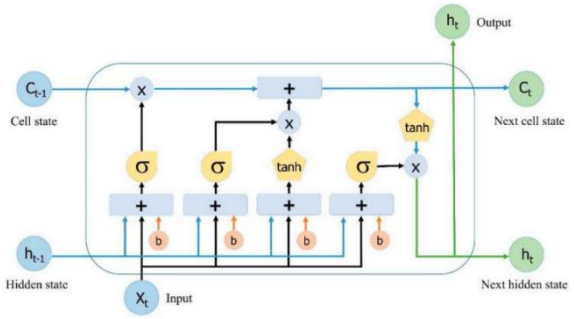

Figure 1. LSTM Architecture

The formula for the forget gate (f_t) is written in formula 1.

$f_t = \sigma(W_f .[h_{t-1} , x_t] + b_f)$.

Furthermore, the two parts are combined to update the

C_t. $i_t = \sigma(W_i .[h_{t-1} , x_t] + b_i)$

$\tilde{C}_t = tanh(W_c .[h_{t-1} , x_t] + b_c)$

For the formula to update the old cell state (Ct1) to the new cell state (Ct) in formula 4.

$c_t = (i_t . \tilde{C}_t + f_t . c_{t-1})$

Second, pass cell state (C) in formula 6 through tanh to produce a value between -1 and 1. Then multiply it by the sigmoid gate, resulting in the decided value.

$o_t = \sigma(W_o .[h_{t-1} , x_t] + b_o)$

$h_t = o_t .tanh(c_t)$

CNN

Convolutional neural networks were first used in image processing or computer vision. In 2015, research was conducted on Natural Language Processing (NLP) using CNNs to classify text. Apply convolutional techniques to sentences, paragraphs or whole text for text classification using CNNs. Deconvolution techniques work by splitting a matrix of text representations into windows or filters. Filters are added together to create a new representation of the text, which can be called a custom map. The maximum value is derived from each specification using the maximum pooling technique. The CNN architecture is shown in Figure 4, including sentence representation, convolutional layers, maximum pooling, full coupling, dropout, and softmax. Sentence representation is the sentence entered in matrix form, this matrix is used as input to the convolution layer. Convolutional layers are made with filters that create feature maps.

Convolutional layers also set parameters. The most important thing is the number of cores and the core size.

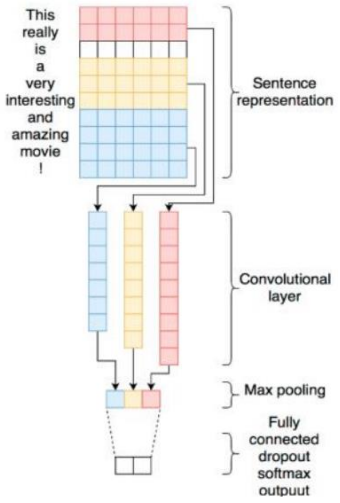

Figure 2. CNN Architecture

n of the data. SVM is recognized for working well with problems that have a lot of characteristics and little training data. SVM may, however, be memory and computationally-intensive for large datasets. The parameters and choice of the kernel may also have an impact on SVM. Numerous practical applications of SVM are available, such as bioinformatics, text classification, and image classification. SVM may be used with other machine learning algorithms to improve the performance of the system. KNN KNN (K-Nearest Neighbors) is an illustration of a supervised learning algorithm used for classification and regression analysis. KNN is a non-parametric algorithm, therefore it doesn't base its decisions on how the data are distributed. The Manhattan distance or the Euclidean distance are two distance measures that the KNN technique uses to identify the K data points that are closest to the query point. The K nearest neighbors are then used to identify the class or value of the query point. A majority of the query point's K nearest neighbors determine how to classify it. If K = 5 and the three nearest neighbors are in class A and the other two are in class, the query point is placed in class A, for example. In a regression, the query point's value is the mean or median of the K nearest neighbors. KNN has the advantage of being clear-cut and easy to understand. When the training set contains a lot of data as well as when the dataset is little, KNN may be effective. KNN may, however, be sensitive to the data distribution, K number of neighbors, and the choice of distance measure. KNN provides a wide range of practical applications, such as text and picture classification, recommendation systems, and anomaly detection.

RANDOM FOREST

The random forest distribution method of co-learning uses multiple decision trees to improve the robustness and accuracy of the model. Random Forest is a supervised learning method for classification and regression analysis. Section Harassment Each decision tree is constructed from a variable feature set and a random subset of training examples. As a result, overfitting is reduced and the capability of the model is improved. When a new sample is wanted to be classified, it is sent to all decision trees in the forest and the most supported class is selected as the protected class for the sample limit. There are many advantages to comparing random forest paths with self-determining trees. It is not easy to overfit and can accept noise and missing data. It also has the ability to manage files that are large in size and have many variables.

4. RESULTS AND DISCUSSION

Distribution of information is determined by our standards. We have selected five classification methods that are frequently used for this situation. We use CNN, Logistic Regression, RNN and LSTM algorithms. We tested each release strategy during prototype training to see which worked best.

Algorithm	Test Score	F1-Score	Accuracy
Simple Neural Network	0.37	0.98	0.99
CNN	0.33	0.84	0.85
LSTM	0.31	0.65	0.86
Logistic Regression	0.74	0.79	0.77

Table 2. Different Algorithm Result

5. Conclusion

Test mining was developed by limiting information to non-trivial guesses to understand subjects' normal thoughts about these thoughts. The method I propose captures both the user's conclusions and the original content for these metrics. Therefore, this strategy removes all guesswork by making them a priority and providing clear insights into who will buy how much. The purpose of the plan is to answer questions about the drivers of each strategy registered in the database and to examine the generality of that strategy. Here I used the ML group technique to find a hypothetical research on data from customer reviews. Logistic regression provides high precision to predict controls I tried my idea with a set of controls and it worked. I write regular reviews for 5-star carriers and will test our strategy with more brands in the future. I always use my approach to check various fields to check the suggested method. It should be noted that while the instruction-based techniques used in hypothetical events can identify underlying assumptions, the interpretation can sometimes be ambiguous, such as inconsistency or generalization in understanding. Future testing should try to answer these problems.

6. REFERENCES

[1] Nguyen, Bang, Van-Ho Nguyen, and Thanh Ho. "Sentiment analysis of customer feedbacks in online food ordering services." *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy* 12.2 (2021): 46-59.

[2] Singla, Zeenia, Sukhchandan Randhawa, and Sushma Jain. "Statistical and sentiment analysis of consumer product reviews." 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, 2017.

[3] Tang, Yachun. "Research on Word Vector Training Method Based on Improved Skip-Gram Algorithm." *Advances in Multimedia* 2022 (2022).

[4] Asgari-Chenaghlu, Meysam. *Word Vector Representation, Word2vec, Glove, and Many More Explained*. Diss. University of Tabriz, 2017.

[5] Nguyen, Bang, Van-Ho Nguyen, and Thanh Ho. "Sentiment analysis of customer feedbacks in online food ordering services." *Business Systems Research: International journal of the Society for Advancing Innovation and Research in Economy* 12.2 (2021): 46-59.

[6] Zhang, Shaozhong, et al. "Research on Logistics Service Assessment for Smart City: A Users' Review Sentiment Analysis Approach." *Electronics* 11.23 (2022): 4018.

[7] Bhatt, Aashutosh, et al. "Amazon review classification and sentiment analysis." *International Journal of Computer Science and Information Technologies* 6.6 (2015): 5107-5110..

[8] Laksono, Rachmawan Adi, et al. "Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes." *2019 12th international conference on information & communication technology and system (ICTS)*. IEEE, 2019..

[9] Sharma, Tamanna, Anu Bajaj, and Om Prakash Sangwan. "Deep learning approaches for textual sentiment analysis." *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines*. IGI Global, 2022. 256-267.

[10] BİLEN, Burhan, and Fahrettin HORASAN. "LSTM network based sentiment analysis for customer reviews." *Politeknik Dergisi* 25.3 (2021): 959-966.

[11] Nandwani, Pansy, and Rupali Verma. "A review on sentiment analysis and emotion detection from text." *Social Network Analysis and Mining* 11.1 (2021): 81.

[12] Singla, Zeenia, Sukhchandan Randhawa, and Sushma Jain. "Statistical and sentiment analysis of consumer product reviews." *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2017.

[13] Alzahrani, Mohammad Eid, et al. "Developing an intelligent system with deep learning algorithms for sentiment analysis of e-commerce product reviews." *Computational Intelligence and Neuroscience* 2022 (2022).

[14] Alsubari, Saleh Nagi, et al. "Development of integrated neural network model for identification of fake reviews in E-commerce using multidomain datasets." *Applied Bionics and Biomechanics* 2021 (2021).

[15] Elmurngi, E. I., & Gherbi, A. (2018). Unfair reviews detection on amazon reviews using sentiment analysis with supervised learning techniques. *J. Comput. Sci.*, *14*(5), 714-726.

[16] Dharma, Arie Satia, and Yosua Giat Raja Saragih. "Comparison of Feature Extraction Methods on Sentiment Analysis in Hotel Reviews." *Sinkron: jurnal dan penelitian teknik informatika* 7.4 (2022): 2349-2354.

[17] Vaishanvi, S., et al. "Product Recommendation Using Sentiment Analysis." *2022 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE, 2022.