Research on Online Comment Negative Emotion Warning Based on Machine Learning from the Perspective of Big Data

Zhi-xian LU¹, Run-xin HE² and Jin-hai Li³

¹School of Management, Jiangsu University, Zhenjiang, China
²Baidu USA LLC, 1195 Bordeaux Dr, Sunnyvale, CA94089, USA
³Taizhou University, Taizhou, China

Keywords: Big data, Online review, Machine learning, Crisis early warning.

Abstract. With the rapid development of information technology, the amount of data has grown exponentially. The birth of cloud computing directly sent us into the era of big data. In view of the uncontrollable factors of online commentary, how to effectively mine the hidden consumer experience and control the negative experience communication from massive online reviews has become the key to building a dynamic “green” network environment in current big data era. This paper is based on machine learning to mine online comments and proposes an online commentary emotional early warning model, which realizes online emotional monitoring based on intelligent information processing to a certain extent.

Introduction

With the rapid development of numerical simulation and artificial intelligence [1] [2], machine learning has been widely concerned by scholars. The emotional experience contained in online comments is a weak signal, which is not easy to be directly detected. Therefore, it is difficult to find the effective solution to mine the emotional tendency contained in online comments in an intelligent way, master customer experience in real time and prevent the spread of negative emotions in the era of big data. Because previous research mainly relied on rule review to conduct emotional analysis. On the one hand, this study how to extract the emotional tendency of online censorship through machine learning analysis so as to reduce the interference of subjective factors, and ultimately realize the automation of emotional recognition. On the other hand, this research achieves the prevention of emotional crisis by establishing an emotional early warning model to track the emotions of online censorship in real time.

In recent years, Chinese and foreign scholars have carried out relevant research work based on different text mining methods from different granularity of words and sentences. Wen bin et al. [4] extracted 22 negative adverbs from HowNet and used relevant rules to change the emotional tendency of sentences so as to improve the accuracy of emotional evaluation. Shi wei et al. [5] found the correlation between different feature evaluations and different emotion classes from the perspective of sentence level to document level, and improved the construction of online comment analysis system based on fuzzy semantics. Chen jianmei et al. [6] used the method of combining syntactic analysis, emotional vocabulary ontology database, rule database and instance database, and concluded that the combination of the above three can solve the general problem of emotional evaluation.

In the field of emotional mining of online comments, li jinhai et al. [7] mapped feature words in online comments into different levels through domain ontology, and constructed hierarchical product attribute mining model and emotional tendency analysis model by using the combination of different levels of conceptual feature words. Cui xuelian et al. [8] proposed an unsupervised online comment emotion classification model based on theme similarity to improve the accuracy and stability of unsupervised emotion classification, and effectively distinguish the emotional tendency of comments.

Using machine learning in the field of online comment analysis, scholars have tried to explore from different perspectives and achieved some results. Peng danlei et al. [9] used SVM and LSTM...
methods to build models for the commodity reviews on jingdong, and better extracted the hidden emotional information in the word vectors through complex nonlinear calculations. Liu xiaotong et al. [10] constructed Word2vec + SVM word vector classifier by combining deep learning and machine learning, and optimized the classification efficiency.

To sum up, there are still some limitations of online reviews research. Firstly, the current online reviews intention mining is mainly focused on the semantic rules, while there is less research based on machine learning. Secondly, relevant scholars is to use machine learning comments mining, but the precision of the fuzzy model based training sample size limitations, time consuming, and model, the result and is expected to a certain degree of deviation. Thirdly, after judging the emotional tendencies of reviews, online reviews mining model is seldom established.

Based on this, this paper focuses on USES machine learning and adopts FastTest[11] word classifier model, which is more advanced than Word2vec. Through the training model, it is more efficient to dig out the emotional tendency contained in online comments. By constructing the emotional warning model of online comments, it can provide useful advice for effectively avoiding negative experience propagation.

Construction of Negative Emotion Warning Model

Machine Learning

Machine learning is beneficial for computers to automatically learn rules from data and use them to predict unknown data. The steps of applying machine learning to online comment mining mainly include beginning, Chinese word segmentation, feature extraction, feature selection, classification model and output results. The overall process is shown in the figure.

![Figure 1. Online comment analysis process based on machine learning.](image)

**FastTest Classification Algorithm**

1. FastTest architecture
   
   FastTest has been a word vector and text categorization tool opened by Facebook AI Research since 2016. The model consists of three parts namely, model architecture, layered Softmax and n-gram features. FastTest greatly reduces the training time while saving the classification effect.

2. Softmax (Hierarchical softmax)
   
   Because the softmax function expression is: \( F(z_j) = \frac{e^{z_j}}{\sum_{i=1}^{n} e^{z_i}} \), under the condition of the class very much, however, softmax is used to calculate the price is very big, its time complexity is \( O(kh) \), the k is the number of categories, h is the dimension of text representation. The time complexity of the hoffman tree-based layer softmax is \( O(h; \log_2(k)) \), based on Huffman tree from the root node to start looking for, and the greater the weight in the Huffman tree nodes closer to the root node, which further accelerated the search speed.

3. n-gram
   
   N-gram is an algorithm based on language model. The basic idea is to conduct the sliding window operation of size n in order of byte in the text content and finally form the sequence of byte
fragments with length N. If we have a sequence of M words (a sentence), we want to calculate the probability $P(w_1, w_2, ..., w_m)$, according to the chain rule: $P(w_1, w_2, ..., w_m) = P(w_1)P(w_2|w_1)P(w_3|w_1, w_2) ... P(w_m|w_1, w_2, ..., w_{m-1})$, using the Markov chain hypothesis, that is, the current word is only related to the first few finite words, so it is necessary to simplify the calculation: $P(w_i|w_1, ..., w_{i-1}) = P(w_i|w_{i-n+1}, ..., w_{i-1})$. For the n-gram model, it is assumed that $C(w_1, w_2, ..., w_m)$ represents the number of occurrences of n-gram, $w_1, w_2, w_m$ in the training corpus. N is the total number of words in the corpus, and then $P(w_i|w_1, ..., w_{i-1}) = P(w_i|w_{i-n+1}, ..., w_{i-1})$.

**Construct the Emotional Warning Model of Online Comments**

(1) Warning model structure composition

Online comment emotion recognition warning is divided into four steps from the process, and seven functional modules from the functional composition. The specific operation mode is shown in the figure.

![Diagram of the emotional warning model](image)

(2) Effective index calculation

In this paper, the mean value of F1 value under the 20 fine-grained emotional dimensions evaluated by users is taken as the evaluation index. The specific calculation method is as follows:

$$F_{1, \text{score}} = \frac{\sum_{i=1}^{20} F_{1, \text{score}(i)}}{20}$$  \hspace{1cm} (1)

Where, $F_{1, \text{score}(i)}$ is Macro F1 value under the corresponding fine-grained emotional dimension, and $F_{1, \text{score}}$ is the mean value of Macro F1 value under the 20 fine-grained emotional dimensions, that is, the final result evaluation index. $F_{1, \text{score}(i)}$ can be obtained by the following calculation:

$$F_{1, \text{score}(i)} = \frac{\sum_{j=1}^{4} 2 \times \frac{\text{precision}_{(j)}^{(i)} \times \text{recall}_{(j)}^{(i)}}{\text{precision}_{(j)}^{(i)} + \text{recall}_{(j)}^{(i)}}}{4}$$  \hspace{1cm} (2)

$$\text{precision}_{(j)}^{(i)} = \frac{TP_{(j)}^{(i)}}{TP_{(j)}^{(i)} + FP_{(j)}^{(i)}}$$  \hspace{1cm} (3)
\[
\text{recall}^{(i)}_{(j)} = \frac{TP^{(i)}_{(j)}}{TP^{(i)}_{(j)} + FN^{(i)}_{(j)}}
\]  

(4)

The evaluation index of n classification is divided into n binary classification problems for solution. According to the \(TP^{(i)}_{(j)}\), \(FP^{(i)}_{(j)}\), \(FN^{(i)}_{(j)}\) of each binary classification evaluation, the corresponding accuracy and recall rate are calculated, and finally \(F_{1\_score}\) is determined.

(3) P\(_c\) determine

Emotional alertness determines whether measures should be taken to block negative experience diffusion. The emotion category of this paper is divided into no crisis (0-0.25), false crisis (0.25-0.35), potential crisis (0.35-0.8) and persistent crisis (>0.8).

(4) Crisis prevention

Crisis prevention is directly related to whether can timely find consumers to comment on the emotional tendencies, according to the mechanism of crisis identification based on emotions predicted values and the index value difference in time to determine the crisis level, clear whether crisis control, and according to the nature of crisis based on crisis plan strategy library to choose reasonable crisis measures, to prevent the crisis from spreading. The overall process of crisis plan selection and control mechanism is shown in the figure below:

![Figure 3. The overall process of crisis plan selection and control mechanism.](image)

**Experimental Research**

This article provides a set of high quality of huge amounts of data, a total of 20 fine-grained elements contains six categories of emotional tendency. Besides, this paper is established under the fine-grained elements of labeling algorithm to emotional mining of user reviews and determine the emotional intensity. In addition, it can build online reviews of crisis early warning mechanism model, make the implement of real-time monitoring of word of mouth public opinion and effectively avoid reputation risk and loss of the user.

(1) The project structure

![Figure 4. The project structure.](image)

(2) Experimental conclusion and analysis

The online comments are finely divided into six evaluation indexes according to the hierarchical method: location, service, price, environment, dishes and other six evaluation indexes. Through experiments, emotional values of each dimension are obtained as follows:
Table 1. Emotional values of each dimension.

| The first layer | The second layer                         | Emotional value (0-1) |
|-----------------|-----------------------------------------|----------------------|
| Location        | Traffic convenience                      | 0.5235               |
|                 | Distance from business district          | 0.4323               |
|                 | Easy to find                            | 0.5656               |
| service         | Wait time                               | 0.5015               |
|                 | Waiter's attitude                        | 0.6764               |
|                 | Parking convenience                      | 0.5673               |
|                 | Serving speed                            | 0.5697               |
| price           | Price level                             | 0.6244               |
|                 | Cost-effective                          | 0.5619               |
|                 | Discount                                | 0.5770               |
| environment     | Decoration                              | 0.5527               |
|                 | Noise                                   | 0.5713               |
|                 | Space                                   | 0.5320               |
|                 | Cleaness                                | 0.5563               |
| dishes          | Portion                                 | 0.5503               |
|                 | Taste                                   | 0.6104               |
|                 | Look                                    | 0.4332               |
|                 | Recommendation                          | 0.4040               |
| others          | Overall experience                      | 0.5310               |
|                 | Willing to consume again                | 0.5469               |

Through the calculation rules of emotional indicators above, the final $F_{1,\text{score}} = 0.5449$ is obtained, it is a potential crisis. According to the crisis management mechanism, on the one hand, further effective monitoring of online comments should be carried out to achieve real-time extraction of emotional experience, prevent it from evolving into a continuous crisis, and regulate the network order. On the other hand, this research adopts to find the right "breakthrough", reduce the risk of crisis, strive to transform the potential crisis into no crisis, promote the spread of positive emotional experience, and build a "green" network environment.

Summary

In the era of big data, online comments are instant, transparent and spread rapidly. The spread of negative emotional experience would be accompanied by a serious crisis of trust. Effective mining of online reviews, grasping the emotional experience of netizens, and putting forward early warning measures for negative experience have certain research value. From the perspective of machine learning and text analysis, this paper introduces a fast test classifier model, which greatly shortens the time of emotional value analysis. Such research and testing is conducive to timely grasp the emotional experience of netizens’ comments, through the construction of emotional early warning model for online comments. This study has gained some negative early warning experience and taken preventive measures to achieve the ultimate goal of standardizing the network order. However, due to the immaturity of machine learning development, there may be some differences in the debugging, testing, data set size and selection of the classification model based on the different experimental environment, which need to be further optimized.
References

[1] He Runxin, Gonzalez Humberto. Numerical Synthesis of Pontryagin Optimal Control Minimizers Using Sampling-Based Methods[C]//IEEE 56th Annual Conference on Decision and Control (CDC). Melbourne, Australia: IEEE CDC, 2017:733-738.

[2] He Runxin, Gonzalz Humberto. Zoned HVAC Control via PDE-Constrained Optimization[C]// Proceedings of the American Control Conference (ACC).Boston USA ACC, 2016(7):587-592.

[3] Turney, Peter D. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002.

[4] Wen Bin, He Tingting, Luo Le, et al. Research on text emotion classification based on semantic understanding [J]. Computer science, 2010, 37(6):261-264.

[5] Shi Wei, Wang Hongwei, He Shaoyi. Semantic analysis of Chinese online comments [J]. Journal of intelligence, 2013, 32(8):860-867.

[6] Chen Jianmei, Lin Hongfei. Construction of Chinese emotional knowledge base [J]. Journal of intelligence, 2009(4):492-498.

[7] Li Jinhai, He Youshi, Ma Yunlei. Hierarchical mining of online comment information based on domain ontology [J]. Systems engineering, 2016(10):43-51.

[8] Cui Xuelean, Nailisa, Liu Xiaojun. Emotional analysis of online comments based on theme similarity [J]. Journal of systems management, 2018, 27(05):24-30.

[9] Peng Danlei, Gu Lize, Sun Bin. Emotional analysis of commodity reviews based on SVM and LSTM [J]. Software, 2019, 40(01):49-53.

[10] Liu Xiaotong, Tian Dagang. Emotional analysis of online comments integrating deep learning and machine learning [J]. Software guide, 2019, 18(02):7-10.

[11] Joulin A, Grave E, Bojanowski P, et al. Bag of tricks for efficient text classification[J]. arXiv preprint arXiv:1607.01759, 2016.