CONVERTER MAGAZINE
Volume 2021, No. 6

Research on Information Extraction and Evaluation Methods on Product Quality Safety in E-commerce Comments

Zhaocai Jiang[1], Anny Bian[2], Ning Sun[1], Li Song[1]
[1]*China National Institute of Standardization, Beijing 100101, China*
[2]*Beijing InsightWorks Info Tech CO., LTD, Beijing 100102, China*

Abstract

The massive volume of consumer reviews on the e-commerce platform can be used as an important reference for market surveillance and productive enterprises to detect product defect and enhance product quality and safety. Thus, product safety evaluation based on these reviews can contribute to a better understanding of industrial safety and quality dynamics, as well as critical defects in the field. In this paper, a model is developed to measure the quality and safety of targeted consumer products based on consumer reviews on e-commerce platforms from two dimensions, namely, quantitative word-of-mouth, which raises concern among consumers, and, content word-of-mouth, which reflects product fault.

Keywords: *Quantitative Word-of-Mouth, Content Word-of-Mouth, Product Safety Evaluation*

I. Introduction

In recent years, with the popularization and development of computer Internet technology, the way users shop and consume has seen tremendous changes. Featuring high convenience and high cost performance, e-commerce has gradually become popular and grown into one of the most mainstream transaction modes, which to a certain extent has replaced the role of traditional offline business. It is mentioned by "The 13th Five Year Plan on E-commerce", by 2020, China's e-commerce transaction volume will exceed RMB40 trillion and the total online retail sales will reach about RMB10 trillion, with the relevant practitioners exceeding 50 million[1]. However, the safety issues of e-commerce sales products are emerging, as the consequence of the vigorous development of e-commerce. These issues may not only hinder the development of our e-commerce, but also bring immeasurable losses to the personal and property safety of consumers.

After completing the transaction on the e-commerce platform, many consumers will comment on their experience of purchasing goods, sharing the experience after using the goods and receiving the service. E-commerce comments have also become one of the distinctive features of the e-commerce platform. To a certain degree, e-commerce comments have broken through the barriers of information asymmetry between merchants and consumers, and increased the awareness and recognition of audiences toward the e-commerce platform. Meanwhile, e-commerce comments, as a feedback mechanism, promote a word-of-mouth interactive network of such three main bodies as e-commerce websites, merchants and consumers, forming a unique three-party interactive network system. On the one hand, e-commerce comments can significantly affect consumers' purchasing decisions. For example, online consumers can make their own purchase plans by measuring other people's comments on a certain product. On the other hand, they can also provide multi-dimensional data support, such as understanding consumer preferences to upgrade product or eliminating defects to improve quality, so as to help relevant quality control departments dig out clues of consumer product defects. That said, online product comments for a certain product are usually massive and uneven in quality. Therefore, a high cost of time and energy is required to filter out valuable information; at the same time, click farming of some merchants also confuses consumers and affect the normal decision-making. As a result, it is highly relevant for the public, enterprises and market supervision departments to carry out research on consumers' comments on word of mouth of e-commerce platforms.

II. **Product Quality Safety Information Extraction from Comments of E-commerce**

E-commerce has witnessed momentous progress of many years in the era of big data, thereby accumulating a huge amount of data resources about industry, users, product quality, logistics, word of mouth, etc. The product safety information in the e-commerce comments proposed in this paper refers to the online comments submitted by consumers about fault performance or injuries that occurred during the use of the product.

2.1 Data sample survey

According to data observation, the description of product evaluation data released by consumers at the e-commerce platform contains clues about product defects and safety risks. For example, in 2019, there were more than 4.8 million pieces of comments on the display side of electric kettles that were sold on the two major domestic e-commerce platforms of Tmall and Jingdong, of which about 100,000 can meet the defect clues; there were more than 2 million pieces of comments on the display side of power bank products, of which about 40,000 pieces can meet the defect clues; and there were about 3.5 million pieces of comments on the display side of hair dryers, of which about 80,000 pieces can meet the defect clues. The number of defective clues in the comments of the above-mentioned products in e-commerce accounts for around 2% of the total, which shines the spotlight on the universal issues of concerned industry, reflects the safety level of the industry.

The point is that not all such categorized products in the e-commerce comments of product safety information have the similar proportion. For example, during the e-commerce comment data monitoring carried out on the chair lifts in 2018 at the beginning of this paper, there were more than 2 million pieces of comments, but of which only over 100 can meet product safety information, which are low-quality clue data. Therefore, it is necessary to carry out more systematic data research before conducting e-commerce comment analysis on certain types of products.

2.2 Data extraction method

In recent years, due to the simple structure of e-commerce comments, the relatively clear and brief description of the problems that arise, and the large amount of data, many experts and scholars have adopted sentiment analysis based on e-commerce comments to research on product quality evaluation, product purchase rate and satisfaction. For example, Yao Zhi'an[2]Analysis and Research on Sentiment Polarity of E-commerce Comments Based on Deep Neural Network, helps consumers and merchants grasp the product quality by mining the emotional tendency of comments. This paper mainly summarizes and takes clustering analysis on the texts of product fault performance and injury situations in the e-commerce comments, and applies pre-training and fine-tuning to dig out fault performance data in e-commerce comments. The main steps are: first train a language model (pre-training) in more than ten thousand of corpora, and then apply the data of this paper to fine-tuning the parameters of the pre-training model, and fit the model parameters to the data features of this paper, thereby improving the accuracy and robustness of the model.

2.3 Main analysis methods

Today China's e-commerce platforms employ the utility evaluation of online commodity comments, which ranks the commodity comments according to the number of likes and comments obtained by each comment. The higher the number of likes and comments, the ranking of the comment is higher. However, this evaluation mechanism still has a certain degree of one-sidedness in the effectiveness analysis of the description of product quality. For example, Deng Lingmin[3], an expert, put forward that the evaluation model should be established based on product comment factors such as commenter's experience value, rating levels, rating extremes, comment length, whether the comment has pictures, and the number of comment responses to evaluate user satisfaction and product promotion effects. This paper takes the number of e-commerce comments, ratings, and the emotional recognition of the audience triggered by the comments as factors. For example, the number of likes or comments to moderate and negative comments indicates to a certain extent that those comments have their value to or have been approved by other consumers. Taking the product comment of the JD e-commerce platform as an example, the indicators that can reflect the magnitude factors include the total number of comments to the product, the number of moderate comments, the number of negative comments, the number of likes to moderate and negative comments, and the number of comments to moderate and negative comments.

III. Design of Product Safety Information Analysis Model for E-Commerce Comments

The word of mouth evaluation model studied in this paper is designed based on such two dimensions as the magnitude and the content word of mouth of e-commerce comments. It is calculated by weighting and summing such two evaluation indicators as the magnitude and the content and to measure the safety word of mouth evaluation of target products and industries on e-commerce platforms.

3.1 Design and analysis of magnitude dimension

Since e-commerce comments have become an important reference for potential consumers to make purchase decisions, illegal merchants have falsified online comments, making consumers confused about product quality, and adversely affecting consumers’ judgment and accuracy of comments[4]. In order to avoid the data deviation caused by this, the magnitude dimension indicators of the evaluation model in this paper are designed to refer to the rating of the product comments by the platform itself, including the total number of comments, the number of moderate comments, and the number of negative comments; on this basis, indicators that can further reflect the emotional recognition of the data, such as the total number of likes and total comments corresponding to the moderate and negative comments, are used as the indicator vector. Taking the JD e-commerce platform as an example, the total number of comments for a consumer product in the product sales page is N, then the number of moderate and negative comments will affect the product's word of mouth in magnitude. In addition to the number of moderate and negative comments, the number of likes and comments to each moderate or negative comment can indicate the user's degree of recognition of the comment. The higher the recognition of the comment is, the greater it impacts on the word of mouth of consumer products. Therefore, the total number of comments and likes to each moderate or negative comment are also used as the indicator vector in the model.

$$I(QW) = \big(1 - a(N_{medium} + c \times C_{medium} + d \times T_{medium})/N \\ - b(N_{negative} + c \times C_{negative} + d \times T_{negative})/N\big) \times 100\%$$

Among which, N is the total number of comments, N_{medium} is the number of moderate comments, C_{medium} is the total number of comments to moderate comments, T_{medium} is the total number of likes to moderate comments, $N_{negative}$ is the number of negative comments, $C_{negative}$ is the total number of comments to negative comments, and $T_{negative}$ is the total number of likes to negative comments. (a, b, c, d) is the corresponding weight of each variable. After the weight of each dimension is determined, the magnitude word-of-mouth indicator is obtained through calculation.

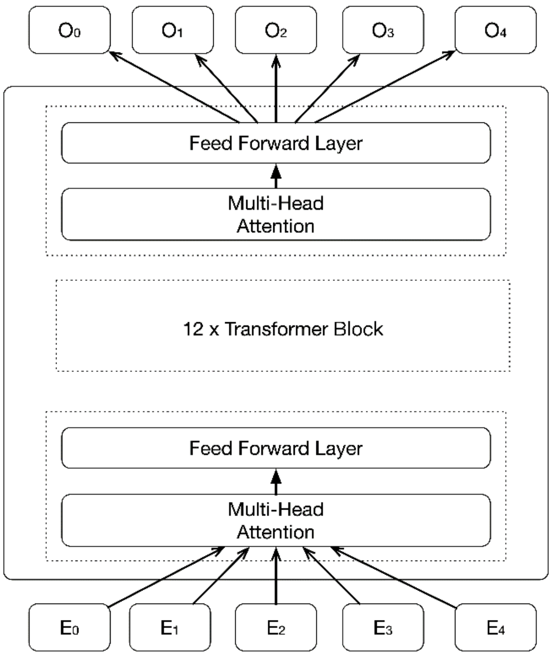
3.2 Design and analysis of content word of mouth

Content word-of-mouth, referring to consumer products of different categories and characteristics, defines sensitive keywords related to fault performance and injury situations at the e-commerce comment content level. And it divides the keywords into two levels according to their sensitivity, namely, "sensitive keywords related to primary safety issues" and "sensitive keywords related to secondary performance”. For example, "electric leakage", "scald", "burst", etc. in relation to electric kettles are defined as "sensitive keywords related to primary safety issues” since these issues may bring about personal and property injuries and other potential safety hazards; while "slow water boiling", "peculiar smell", "big noise”, etc. are defined as "sensitive keywords related to secondary performance” since these issues involved are less likely to cause some injuries. This paper is to search and locate all the collected comments involving defined sensitive keyword groups, and establish the mapping between contents and sensitive keywords, so as to count the frequency of each keyword hit. On the contrary, because the description of product fault performance varies greatly in different contexts, model training based solely on the shallow word frequency of the text cannot achieve the expected results. Therefore, this paper, by using pre-training and fine-tuning, brings word vector with characteristics of speech and semantics to large model data, adding multi-dimensional features to model training to enhance model robustness and generalization capabilities. This paper is set to use the BERT language model to encode the text data into a vector, and then output the fault performance characteristics of the data through the vector feature map.

3.2.1 Application of BERT model in fault performance identification

The BERT language model uses a multi-layer Transformer structure as the main framework[5]. On one hand, it makes use of multi-heads self-attention mechanism, combined with the two-way context relationship, to detect the deeper relationship of the text in the e-commerce comment sentence. On the other hand, using multi-heads self-attention mechanism can efficiently and concurrently do trainings. Its advantage is that the amount of data is larger and easier to converge than the LSTM model training in the same time. BERT makes an innovative use of Mask Language Model (MLM)[3] and Next Sentence Prediction (NSP) to build multi-task training goals. Therefore, compared with left-to-right autoregressive models (LSTM, CNN, etc.), BERT can "see" and use contextual information at the same time. The structure chart is shown in Figure 1:

Figure1: BERTmodel structure chart



BERT inputs two sentences into the model at the same time as one piece of training data at a time, marking the beginning and end of the sentence with "[CLS]" and "[SEP]", and separating the two sentences with "[SEP]". And then it uses the "[MASK]" mark to encode the sentence, and concatenate the encoded vectors to form the initial input vector of BERT. Its model structure is as shown in the following equations:

$$H_0 = EW_e + W_p \tag{1}$$

$$H_l = transformer\_block(H_0) \tag{2}$$

$$O_u = softmax(H_l W_e^T) \tag{3}$$

Among which, W_e is the unique identification of words, and W_p is the locations of words. l is the layer of Transformer, and $l = 12$. H_lH_l is the hidden layer of l, and O_u is the output of the model.

Using a pre-training language model is beneficial to that the semantic relationship between text contexts can be obtained with less computational overhead. There is no need to train model parameters from scratch, and only needs to fine-tuning the existing data in the trained BERT language model. At the same time, in order to fit the characteristics of the multi-label classification task, this paper modifies the model output, and map the output layer O_u to the dimension L. L is the number of labels. As a result, it can explicitly indicate the correct fit degree of each dimension vector to the label, and at the same time convert the semantic feature of the text into the label feature of the multi-label task.

3.2.2 Product safety multi-label model experiment and result analysis

The experimental data in this paper are mainly from Internet e-commerce comment data. After filtration, there are about 120,000 pieces of eligible data, involving such electronic products as power banks, electric kettles, hair dryers, etc. Each piece of data contains text information and label information, and the number of labels is not less than one. The data sample is shown in Table 1.

Table 1 Internet e-commerce comment data

Commodity category	Text data	Label
Electric kettles	No more than two stars were given and it had the smell of plastic. The lid was not tight at all, and the water leaked. But the heat preservation effect was good, and the noise was low. The water boiled very quickly, there was not much noise, but the smell of plastic was quite big. I didn’t know how the heat preservation effect was, so I would review it after a few days! The water boiled very quickly,, but the noise was a little loud. On the whole, it was very satisfactory.	Peculiar smell; water leakage; noise
Electric kettles	Inner wall material:316 material for cups: it's OK, a bit hot. Water boiling speed: very fast. In order to test the speed of boiling water, I boiled water intermittently about 6 times. 1. The smell of plastic was still there. Netizens said that adding something to burn it was just to cover up the smell of plastic but cannot be removed. However, the water didn’t have the smell of plastic. 2. There were small black spots on the bottom, which can not be washed clean with water, corroded.	Peculiar smell; easy to corrode; poor quality
Electric kettles	The base was unstable, the boiling water noise was loud, the surface plastic was not good, and there were black things in the inner pot. Therefore, it was not very satisfactory.	Base fault; noise
Electric kettles	Bad, there was water leaking from the bottom. And the lid was a bit loose, the workmanship inside was very rough, and the	Water leakage; poor quality

	inner pot had many burrs and was very shaved.	of the lid; scratches	

In order to fully fit the diversity of e-commerce evaluation data, this paper increases the number of training rounds to 10, and adjusts the learning rate to 5e-5. The Transformers structure layers of bert-base-chinese model are set to 12, with the vector of hidden layers as 768, so the experimental result is as shown in Table 1. This paper is to use equations (4)(5)(6) to evaluate the model's ability. TP_c represents the correct number of defect labels predicted; FP_c represents the actual number of wrong labels, but the predicted result is the correct number of labels; TN_c represents the number of incorrectly predicted defect labels; FN_c represents the actual number of correct labels, but the predicted result is the incorrect number of labels.

$$Pre = \frac{\sum_c TP_c}{\sum_c TP_c + \sum_c FP_c} \tag{4}$$

$$Recall = \frac{\sum_c TP_c}{\sum_c TP_c + \sum_c FN_c} \tag{5}$$

$$F1 = \frac{2Pre \cdot Recall}{Pre + Recall} \tag{6}$$

Table 2 Table of experimental training parameters

Parameter	Description	Result
Epoch	Round	10
Learning_rate	Learning rate	5e-5
Model	Pre-training model	bert-base-chinese
Warmup_steps	Warm-up steps	500
Threshold	Threshold value	0.7

In order to verify the effectiveness of the pre-training language model method, this paper, based on e-commerce comment data, makes a comparison analysis on traditional machine learning methods and deep learning models based on LSTM. The result is as shown in Table 3. It is shown that compared with the machine learning method, the pre-training language model based on BERT has made SVM improved by 9%, and compared with the deep learning LSTM model, the correct rate has been improved by 4%.

Table3 Comparative experiment result

Model method	Precision	Recall	F-1
SVM	0.69	0.71	0.70
LSTM	0.74	0.77	0.75
BERT	0.78	0.80	0.79

3.2.3 Application of product safety-orientated information in model design

By using pre-training and fine-tuning, this paper extracts comments on fault performance of a certain type of consumer product at the e-commerce platform. In fact, a single piece of comment information in the actual data may correspond to one or more fail-safe labels. Therefore, this paper takes electric kettles as an example. The data sample is shown in Table 4.

Table 4: Table of experimental data examples

Commodity category	E-commerce comment data	Fault label	Content rating of word of mouth
Electric kettles	Negative comment!!! It was broken the first time I used it. Because the first time I	Water leakage,	Primary

	used it, I put a little bit of water to boil to remove the peculiar smell. As a result, the kettle rang and the water leaked and burned my fingers. After contacting the after-sales service, I was asked to provide photo and video.	scald	Safety
Electric kettles	After waiting for a long time, I finally bought it this time. The design is very user-friendly. The water boiled quickly but the noise was very loud. The first time the water was heated, it would have an unpleasant smell. It is very convenient to carry and is an artifact of business trips!	Noise, peculiar smell	Secondary Performance
Electric kettles	It's very convenient to use, but it had an unpleasant smell. The product description said it was automatically powered off, but it was not at all. I had to watch it while boiling water, which was a bit fooled.	Peculiar smell, failure to switchover	Primary Safety
Electric kettles	The product was too poor and shoddy, and my hand was scratched by the spout, which was too sharp. I went to wash the electric kettle as soon as it arrived. The result was a trick. Is the purpose of leaving a spout for no reason to cut the user's hand?	Sharp edge, scratch	Primary Safety

This paper quantifies the product comments for which the mapping has been established through the calculation model of the content word of mouth, and calculates the comments of the brand product in terms of product faults.

$$I(CW) = \left(1 - p \times Norm(N_{keyword1}) - q \times Norm(N_{keyword2})\right) \times 100\%$$

Among which, $N_{keyword1}$ is the total number hit by sensitive keywords related to primary safety issues, and $N_{keyword2}$ is the total number hit by sensitive keywords related to secondary performance; (p, q) is the weight corresponding to each variable. In order to clearly distinguish between product safety faults and performance faults, the weight value p of the primary keyword in this model is higher than that q of the secondary keyword.

IV. Application of Product Safety Information Evaluation Model

The total value of e-commerce word-of-mouth rating for the consumer product is calculated by assigning the magnitude word-of-mouth rating and the content word-of-mouth rating with corresponding weights when carrying out the analysis on product safety evaluation based on word of mouth comments by consumers in the e-commerce platform. In the foregoing steps, the magnitude word-of-mouth rating $\mathrm{I(QW)}$and content word-of-mouth rating $\mathrm{I(CW)}$ have been calculated. Next, weight of magnitude word-of-mouth ratings is assigned$\mathrm{m}$, and weight of content word-of-mouth ratings is assigned$(1-\mathrm{m})$, and they are summed to calculate the total value $\mathrm{I(W)}$ of word-of-mouth rating of the target consumer product as shown in Table 5. Through numerical comparison, specific investigations can be conducted on the quality and safety evaluation of various brand products in e-commerce comments.

ISSN: 0010-8189

Table 5 Table of product quality and safety evaluation for electric kettles

Magnitude dimension		Brand A	Brand B	Brand C	Brand D	Brand E	Industry Average
Primary safety label	Failure to switchover	1.25%	0.00%	3.04%	1.35%	0.57%	0.07%
	Scald	0.00%	0.00%	0.00%	0.00%	0.00%	0.06%
	Water leakage	1.25%	1.79%	0.00%	0.00%	0.10%	0.20%
	Burst	1.25%	0.00%	0.00%	0.00%	0.00%	0.01%
	Scratch and sharp edge	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%
Secondary performance label	Easy-stick pot (healthy pot)	0.00%	0.00%	0.00%	0.00%	0.00%	0.03%
	Loud noise	0.62%	0.00%	0.00%	0.00%	0.00%	0.15%
	Easy to rust	0.00%	0.00%	0.00%	0.00%	0.10%	0.03%
	Slow heating	6.88%	0.00%	0.00%	0.45%	0.10%	0.20%
	Peculiar smell	1.25%	8.93%	8.11%	8.11%	7.36%	0.91%
	Easy to corrode	0.00%	0.00%	0.00%	0.00%	0.38%	0.09%

While evaluating and calculating the word of mouth of safety on single-brand consumer products in e-commerce, this model is used to carry out product quality and safety evaluation and analysis on all brand products involved in the target industry, and calculate the industry average of product quality and safety in this industry (such as Table 5), which can clearly reflect the overall product safety of the target industry. On the basis of calculating the industry average, the analysis of the safety faults of brand products that are higher than the industry average can evidently help identify the industry brands with low quality and high-risks.

V. Conclusion

This paper presents the research on product safety evaluation method based on comments of word of mouth at e-commerce platforms, with product names and consumer comments of Jingdong and Tmall e-commerce platforms as study factors. With the help of pre-training and fine-tuning, it adopts pre-training language model to mine fault performance data from comments in e-commerce and makes an emotional scale analysis on the number of comments from consumers and audiences of those comments, so as to evaluate the word of mouth of products and the related industry in terms of quality safety. On the basis of the word-of-mouth of the e-commerce platform itself, the emotional influence of the audience is strengthened. The text data is encoded in vector form, and then the fault label of the content data is output through vector feature and content rating is carried out. It is shown that this paper effectively highlights the product safety information in the word-of-mouth text, ensures the integrity of the data, and avoids the missing of fault performance. The evaluation results can serve as a reference for follow-up product defect investigation and industry quality safety investigation.

With the huge data source of comments from e-commerce platform consumers, the research method of this paper can be further optimized, such as increasing the word-of-mouth content ratings of fault labels, refining the primary safety-related contents on the basis of the existing primary safety key words and secondary performance key words to distinguish high risks from medium and low risks, exploring the factors of product sales to reduce the deviation of fault mention rate (which has implications on the evaluation method) caused by unbalanced product sales.

Acknowledgements

The projects funded this research include "Research on the Safety Evaluation Method of Typical Consumer Product Defects Based on E-commerce Comments" (Project No.: 282020Y-7511), a dean fund project of China National Institute of Standardization, and “Research on the Safety Evaluation Method of Typical Consumer Product Defects Based on E-commerce Comments " (Project No: 2020MK161), a technology planning project of the State Administration for Market Regulation.

References

[1] E-commerce and Information Dept., MOC The 13th Five Year Plan on E-commerce(December 29, 2016)
[2] Z.A. Yao, “Analysis and Research on Sentiment Polarity of E-commerce Comments Based on Deep Neural Network,” Shanghai Normal University, June 2020.
[3] L.M. Deng, Y. Xu, “Research on E-commerce Evaluation under the Influence of Online Word of Mouth,” Journal of Library and Data, vol. 1, no. 2, pp. 78-75.
[4] C.H. Zhang, J.J. Liu, H. Ran, “Research on E-commerce Credit Evaluation Method Based on Hybrid Algorithm of Bayesian Network and Neural Network,” Information Science, pp. 81-87, February 2020.
[5] L T. Wilson, “Cloze procedure: A new tool for measuring readability,” Journalism Bulletin, vol. 30, no. 4, pp. 415–433, 1953.
[6] Z.C. Jiang, N. Sun, L. Song, F.Z. Zhang, “Research on Clue Analysis of Consumer Product Defects Based on E-commerce Evaluation Information,” Standard Science, no. 11, pp. 140-144, 2019.
[7] J.A.K. Suykens, J. Vandewalle, “Least Squares Support Vector Machine Classifiers,” Neural Processing Letters, vol. 9, no. 3, pp. 293-300, 1999.
[8] S. Hochreiter, J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

ISSN: 0010-8189