# Sentiment Analysis of Product Reviews on E-Commerce Platforms Based on NLP: Take HUAWEI P40 as an Example

Rui Yang, Yuzhu Xie, Jie Sun, Ling Ding, Zhemiao Zhang and Jianjun Wang*

*School of Sciences, Sichuan Agricultural University, Ya'an, Sichuan,625014, China*

Keywords: Natural Language Processing, HUAWEI, Sentiment Analysis, Online Shopping Review.

Abstract: As the boom and popularity of e-commerce greatly influenced people's consumption patterns and socio-economic structures, the study of product reviews has great significance for guiding purchases. Most of these product reviews are subjective and non-structural. Natural Language Processing (NLP) is an effective method to address communication between humans and computers. It is commonly used to convert unstructured language into structured language for computer recognition and processing to solve problems that are difficult to solve manually. In this paper, we put out the previous HUAWEI Mall reviews about HUAWEI P40 model mobile phone as the data source for natural language processing. And then, we use a CNN model based on the language of Python to perform sentiment analysis of consumers' attitudes about this phone. The results illustrate that consumers' concerns about the HUAWEI P40 phone are generally divided into the following categories: photography, battery life, communication, running speed, display effect and appearance, and their attitudes are generally positive and optimistic. In other word, the negative sentiment towards this phone is not significant.

## 1 INTRODUCTION

### 1.1 Research Significance

In recent years, in the background of "big data" era, with the popularity of online shopping, Taobao, Jingdong and other major e-commerce platforms are flourishing and competing fiercely (TAO, ZHANG, SHI, WEI 2020, ZENG, YU 2020, Feng 2019, Li 2019). They are not only a platform for selling goods, but also a platform for consumers to exchange product use experience as well as quality evaluation. Therefore, many online users often choose to check the information and evaluation of a certain product before buying or using it, and different consumers have different performance tendencies for the same type of products.

In order to improve the quality of customer service, in addition to price wars, it is increasingly important to understand the needs of customers and listen to their voices, an important way is data mining and sentiment analysis of consumers' text reviews (TAO, ZHANG, SHI, WEI 2020). Getting these information also can improve the competitiveness of the corresponding products, and can be used as a guide for consumers' purchases, thus it is of great practical significance to help consumers' choices become more rational.

### 1.2 Current Research Status at Home and Abroad

Text sentiment analysis is a multidisciplinary research field, which belongs to the scope of computer linguistics. It has a wide range of applications in computer science as well as management, political science, economics, sociology, and the booming of business in the industry has further expanded its applications (TAO, ZHANG, SHI, WEI 2020). Already in the 21st century, as a new research direction, this field has gradually become active and attracted increasingly more researchers. In the application of sentiment analysis, "user comment analysis and decision making" is one of the most frequently used applications of sentiment analysis technology (Feng 2019). Many domestic and foreign research institutions have developed sentiment analysis systems in various fields according to the specific needs in real life to help users analyze and make decisions on large amount of information. Many

commercial systems and applications based on sentiment analysis have emerged. For example, Google Shopping, a representative platform based on product reviews, provides users with product search and price comparison services for online purchasing platforms. OpinionE Q allows commercial organizations and individuals to customize product analytics services on demand (Li 2019). However, due to the increasingly swelling problem of the diversity of sentiment research objects and the complexity of sentiment tasks. Sentiment objects range from positive and negative tendency comments on products and services to categorizing user and topic sentiments in social media categories, with more diverse expressions and a greater variety of emotions, and a corresponding shift in research content, including more attention to the user's information as well as the change of the emotion for the user (Li 2019, Ziang, Wang, Li, Lévy, Nie, Jurafsky, Ng 2017, Lan, Chen, Goodman, Gimpel, Sharma, ALBERT 2020, Young, Hazarika, Poria, Cambria 2018, Rush 2018, Vaswani, Shazeer, Parmar, Uszkoreit, Jones, Gomez, Kaiser, Polosukhin 2017). On the other hand, consumers' comments on short texts are diverse in their expression of emotions, either in a straightforward manner or through rhetorical devices or even irony. The multiple forms of expression of emotions such as irony make the analysis inaccurate. This project applies a CNN model with fewer parameters but better performance to text data for sentiment analysis, and uses the analysis results to evaluate the attributes of products. This will help merchants to improve their products and consumers to choose more suitable products for themselves.

## 2 METHODS

### 2.1 Data Collection and Pre-processing

At present, online shopping had become one of the important ways for people to trade and consume. By the end of 2020, Taobao and Jingdong two major e-commerce platforms had reached 1.2 billion active purchasing users in the past year. Due to consumer preference, digital enthusiasts tended to purchase electronic products from official channels. As of July 2021, as the P40 model phone's main and official sales channel, more than 30,000 after-sales reviews had been accumulated in the HUAWEI Mall. If the entire model phone's review data were crawled and studied, the process would be very tedious, resulting in unnecessary human, financial, and material waste. Therefore, this paper randomly selected 500 real reviews of HUAWEI P40 model mobile phones as the basis of the research. The study first selected 10,000 reviews data of other mobile phone products with known labels to train the CNN model. The consumers' attitudes of "positive", "negative" or "other" are determined based on the content of the reviews of different labels.

For text data, text data usually contained a lot of noise that is not relevant to the task, so it should be Pre-processed such as removing URL links, removing emoji emoji, converting English upper case to lower case and converting Chinese traditional to simplified, etc.

Simultaneously, 492 valid questionnaires were collected from online questionnaire survey of HUAWEI mobile phone users. The results showed that the users' concern about HUAWEI p40 mobile phone are mainly focused on six aspects, such as running speed, photography, battery life, communication, appearance and display effect. According to the above six tags, the dataset is classified by Python in six labels to get six review sets.

### 2.2 CNN Model Building

In the field of computer deep learning, CNN convolutional neural network is one of the most successful application of deep learning algorithm. Consisting of neurons with learnable weights and bias constants, each neuron receives some input and does some dot product computation to output a score for each classification.
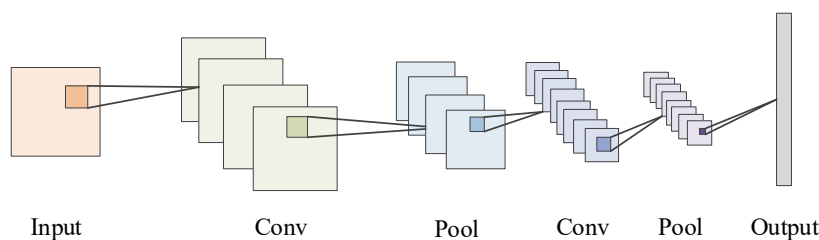


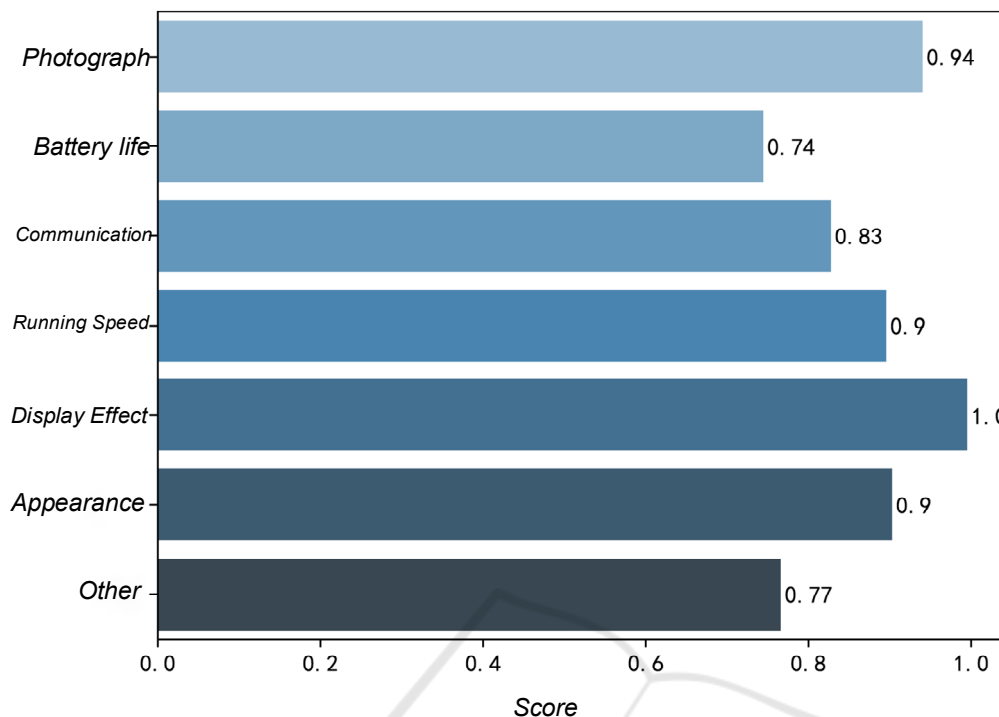Figure 1: CNN convolutional neural network model.

Figure 2: HUAWEI Mall HUAWEI P40 review sentiment analysis values (mean).

## 2.3 Model Training and Data Processing

Due to the non-full connectivity and weight sharing characteristics of CNN neural network, which makes its complexity lower and the number of weights less, so CNN convolutional neural network is chosen for the review sentiment analysis of HUAWEI P40 mobile phone. The text in the review set was Tokenized, and a vocabulary table was built. Then the Token sequence was converted into a word embedding matrix. The word embedding was then fed into the CNN neural network for training. When training the model, the number of traversals could control the training time of the model. The smaller the number of traversals, the shorter the training time, but the training effect would also decrease. The learning rate could be adjusted between 0.001 and 0.0001, and different learning rate models would also have different training effects. After multiple adjustments, we found that setting the number of traversals to 30 and a learning rate of 0.001 has the best training effect. At the end of the final training, the accuracy of the CNN model was 94%, and the model was considered good and the next step of data processing could be performed.

The 500 data exported from HUAWEI Mall were imported into the trained model for sentiment analysis, and the following results were obtained

## 3 ANALYSIS OF THE RESULTS

### 3.1 The Overall Attitude of Consumers Towards Each Label Shows a Positive Trend

The sentiment analysis of HUAWEI P40 reviews in HUAWEI Mall showed that the results were divided into positive and negative scores, representing whether consumers had a positive or negative attitude towards the phone. Both positive and negative scores ranged from 0 to 1, when the score was greater than 0.5, it was considered that consumers have an optimistic and positive attitude towards this feature of the phone. when the score was less than 0.5, consumers had an unfavorable attitude. Meanwhile, the closer the value was to 0 or 1, the more absolute consumer attitude was.

The results of the sentiment analysis showed that the sentiment scores about the HUAWEI P40 phone concerned by online consumers were greater than 0.5 in all 6 aspects, which means that consumers were more positive about this phone and generally

Table 1: Evaluation indicators Number of choices Percentage (492 surveys).

| Evaluation index | Select number | Proportion |
| --- | --- | --- |
| Apearance | 457 | 0.9289 |
| Running Speed | 442 | 0.8984 |
| Photograph | 376 | 0.7642 |
| Battery life | 257 | 0.5224 |
| HongMeng system | 141 | 0.2866 |
| Communication | 137 | 0.2785 |
| Cost performance | 122 | 0.2480 |
| Effect of display | 105 | 0.2134 |
| Hand feeling | 97 | 0.1972 |
| Others | 70 | 0.1423 |

satisfied with it.

Among them, consumers think the best was the "display effect" of HUAWEI P40, with a value of 0.999 and an image display of 1. Secondly, people were generally satisfied with the "appearance" and "running speed" of this phone, with a value of 0.9.The lowest rated label was "battery life" for the HUAWEI P40, which indicated that consumers have higher expectations for the battery life of this phone.

## 3.2 Consumers are More Diversified in Terms of what they Care about HUAWEI Phones

As of July 1, 2021, Mac Form collected 492 reviews of the HUAWEI P40 model phone.

With multiple choices available to respondents, the relevant high-frequency words are shown in Table 1.

As shown in the table, the most frequent words are "appearance" and " running speed", indicating that consumers were more concerned about the appearance of the phone on the outside and the speed of the phone on the inside. From the comparison of the results of sentiment analysis, consumers were more optimistic about these two features.

Secondly, the emergence of the terms "Cost performance" and "Hongmeng system" is also noteworthy. Consumers who buy HUAWEI phones were also more concerned about the price/performance ratio of domestic phones and whether to use the domestic operating system Hongmeng, with a certain subjective atiitude and personal preference.

## 4 CONCLUSION

This paper used natural language processing technology to carry out machine learning through CNN model, in order to evaluate the products on e-commerce platform, which taked HUAWEI P40 model mobile phone in HUAWEI Mall as an example. The relative reviews were classified and scored by tags, finally review sentiment analysis is conducted. The sentiment analysis initially reveal consumers' sentiment attitude towards the phone. There are two conclusions and recommendations.

(1) Consumers focus on 6 aspects of HUAWEI phones, such as "appearance", "running

speed" and "photography", and the proportion of these three items is greater than 50%, indicating that consumer preferences are more convergent. Mobile phone manufacturers can improve and stabilize the performance of these three aspects to achieve a better user experience.

(2) Consumers' preferences for HUAWEI P40 mobile phone's 6 performance items

all showed a positive trend, indicating that users' trust and love of this phone were high. This phenomenn indicates that users had a high degree of trust and affection for this phone, and the user stickiness was high at this time. Producers can accelerate the development of new byproducts, which is to further expand the market and enhance consumers' purchasing power.

The limitation of this paper is mainly the limitation of the analysis sample.

The majority of reviews were filled in before consumers had fully experienced the product features, and it is possible that with the passage of time, the consumer attitudes might still fluctuate over time. Secondly, the survey sample was limited by the number of HUAWEI P40 users, which made it difficult to accurately collect questions, the questionnaire respondents were HUAWEI mobile phone users. Although it can largely represent the attitude of HUAWEI product consumers towards HUAWEI mobile phones, the classification of this label was still not precise enough and may not reflect 100% of consumers' concerns about the

HUAWEI P40. With the further promotion of HUAWEI P40 mobile phone in the market, we can use more diversified e-commerce platforms to collect different samples for analysis in a stratified manner.

# REFERENCES

A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A N. Gomez, L Kaiser, I Polosukhin. Attention Is All You Need[D].2017.

Feng XZ. Sentiment Analysis of Commodity Reviews Based on Deep Neural Network[D]. Beijing University Of Posts And Telecommunications,2019.

Li J Q. Sentiment Analysis Research and Application of Online Shopping Review Data [D]. Chengdu University of Technology,2019.

Rush, Alexander. The Annotated Transformer[A]. Association for Computational Linguistics. Proceedings of Workshop for NLP Open Source Software (NLP-OSS) [C]. Melbourne, Australia. 2018.52–60.

TAO Y C,ZHANG X Q, SHI L,WEI L. R esearch on Multi-feature Fusion Method for Short Text Sentiment Analysis [J]. Journal of Chinese Computer Systems,2020,41(06):1126-1132.

T Young, D Hazarika, S Poria, E Cambria. Recent Trends in Deep Learning Based Natural Language Processing[D].2018.

ZENG X Q, YU H. Sentiment Analysis of Merchandises' Comment Based on Python [J]. Computer Knowledge and Technology ,2020,16(08):181-183.

Ziang X, Sida I. Wang, J Li, D Lévy, A Nie, D Jurafsky, A Y. Ng. Data Noising as Smoothing in Neural Network Language Models[D].2017.

Z Z Lan, M D Chen, S Goodman, K Gimpel, P Sharma, R Soricut. ALBERT: A Lite BERT for Self-supervised Learning of Language Representations[D].2020.