# Sentiment Analysis System on Automobile Customer Comments

Zhikai Kang[1]

[1]Office of Integrated Media Interface, Department of Media Technology and Art, School of Mechatronics Engineering, Harbin Institute of Technology, Harbin, Heilongjiang Province, China

**Keywords:** Text Sentiment Analysis; Computer Application; Automotive Field; Customer Comments Analysis.

**Abstract.** In recent years, the technology of text sentiment analysis develops rapidly and becomes a hot research topic in the field of Natural Language Processing. Now it is applied in increasing number of fields, include e-commerce website, shopping website and so on. In this paper, the application of sentiment analysis in automobile field is studied. Based on that technology, a sentiment analysis system on automobile customer comments is established. Related data are collected from online automobile forums, and then processed. Text sentiment analysis technique is used to extract evaluation objects and evaluative terms, and then analyze emotional tendencies. At the end of evaluation, the system achieves good results.

## Introduction

With the continuous development of economy, more and more families are considering buying a car. For common families, buying a car is an important and relatively expensive thing. So it is important to choose a car which has suitable price and quality.

This thesis trying to analyze the application of text sentiment analysis technique in automobile field, and establish a sentiment analysis system on automobile customer comments to provide guidance for automobile customers. This task is divided into three stages: obtaining and preliminary processing data in automobile field; extracting evaluation objects and evaluative terms from data preliminary processed; analyzing the emotional tendencies of users.

The construction of automobile review sentiment analysis system, on one hand, can help users to understand relevant automotive information; on the other hand, can help enterprises to learn users' emotions, and explore accurate and efficient emotion recognition method for online automobile reviews. In that way, automobile companies can understand consumer psychology, improve marketing strategy, and finally improve corporate image and increase enterprise profit.

## Relevant Work

With the continuous development of Internet technology, Internet users have changed from passive receivers to active information providers. In that case, a large number of evaluation information on people, things and articles is created. This information is of great research value.

Text sentiment analysis technology [2], is a kind of technology which uses methods like Natural Language Processing, statistics and machine learning techniques to analyze the subjective attitude, sentiment orientation, or the polarity of views in the text. With the development of text sentiment analysis technology, the application field of big data research is enlarging. The text sentiment analysis technique is now applied in systems like micro-blog food map and hot topic sentiment analysis. There's no application system on the field of automobile consumption.

In the practical application of text sentiment analysis, text preprocessing is the first step. After processing original product reviews, the noise of the text will be reduced; the accuracy of later text analysis will be improved. Word segmentation technology is the basis of Natural Language Processing; part of speech tagging is used in text classification. In this paper, the language cloud (LTP) of Harbin Institute of Technology is used to accomplish this task. LTP, or Language Technology Platform, is a set of Chinese language processing system established by Social Computing and Information Retrieval Center of Harbin Institute of Technology after ten-years' research and

development. LTP provides functions such as word segmentation, part of speech tagging, syntactic analysis and so on.

The development of visualization technology makes it easier for people to find the value of big data. Thus, visualization technology is used in the representing of final results of the system, which enables users to find analysis results conveniently.

## The Construction of Sentiment Analysis System on Automobile Customer Comments

**System framework and flow chart.** According to the research purpose of this paper, main functions of the automobile review sentiment analysis system are as following. When a user inputs the name of car in the search text box, the system will go to get information on comments of that car on Internet, and then preprocess data by filtering unnecessary information, word segmentation, POS tagging and syntactic commentary. Then evaluation objects and assessment words are extracted to analyze the emotional tendency. Finally, the overall evaluation of that car is represented through visual method. The flow chart of the system is as follows:

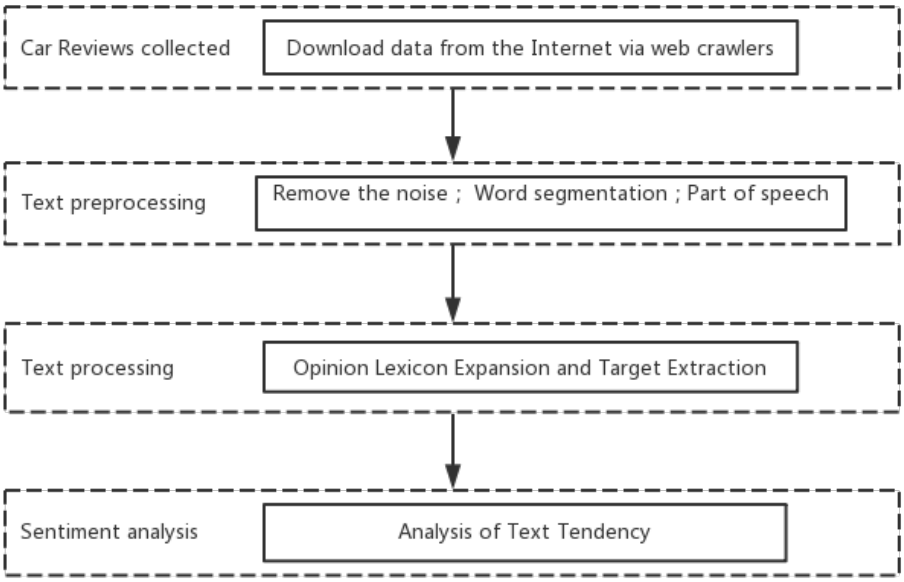


Figure 3-1 Flow chart of automobile comments system

There are several modules in this system, which are the basic information search module, the comment and attribute extraction module, the emotion analysis module and the interactive design module. The following four parts are introduced afterwards.

Basic information search module: the main function of that module is to get relevant information on the car user typed from the Internet, and then show the information it collected.

Comment and attribute extraction module: comment and attribute extraction module is one of the most important modules. It is the main evaluation object of consumer review and emotional tendency.

Sentiment analysis module: the main function of sentiment analysis module is to evaluate each attribute extracted from the extraction module, to determine the polarity of each attribute, and then calculate the percentage of positive and negative emotions as the final result.

Interactive design module: interactive design module is responsible for designing interfaces and interactive processes of the whole system.

**Extracting evaluation objects and evaluative terms.** The method used in this system is a two-way communication algorithm. The method uses a strategy based on association rules. [3] Before start, a set of seed words, which includes typical evaluation and assessment words, is needed to carry out the extraction. Researches do not need to worry about the size of the set. Through experiment, it is found that even a very small set of evaluation words can have a high recall rate.

Concrete realization of two-way communication:

Building the seed set of emotional words;

Mining evaluation object according to seed words and adding the evaluation object into the set of evaluation words;

Finding corresponding emotion words in the given samples according to the relationship between the evaluation object with emotional words, and the characteristics of the evaluation object; then adding the emotional word into the set if it is not in the emotional set;

Recording the relationship between feature words and emotional words in (2) (3);

Repeating steps (2) - (4), until a certain number of iterations is realized, or no more new feature words and emotional words are found.

In the process of task processing, there will be a lot of words that do not meet the requirements. In this paper, we use two methods to reduce the noise. First, word frequency information is used to filter information. NN and NP with less word frequency in the corpus are filtered. Secondly, PMI algorithm based on network mining is used to filter information; the PMI values of a and b are as following:

$$PMI\ a\text{-}b = Nab / (Na \times Nb) \quad (3\text{-}1)$$

Among them, $N_{ab}$ means text data which contains both a and b; $N_a$ means text data which contains a, $N_b$ means text data which contains b. As it can be seen from the formula, statistical thinking is applied in the calculation. PMI calculation is based on that assumption: the more co-occurrence of two words, the greater of their link will be. In theory, statistical effect improves with the increasing number of text data. PMI value should be more accurate if there's more text data. In order to get enough information, this paper selects the search results of Baidu as corpus. The process is as following: for each field, selecting the most representative word $w_a$, and calculating PMI value between the candidate evaluation object W and the related field of $w_a$. The greater of the value, the stronger correlation will be, and W will be more likely to become an evaluation object. Finally, the method of threshold is used to filter information. In the system evaluation afterwards, these methods achieve good results.

**Evaluation of emotional tendencies.** In this part, the author uses the dictionary method to evaluate the tendency of emotion. Dictionary method [4] means to judge the polarity of a word by using relationships between words in the dictionary. Tian-fang Yao [5] and other scholars establish their artificial polarity dictionary to make sentiment analysis on auto comment text. In this system, the attribute of evaluation and assessment of words are selected to judge the polarity of emotional words. Then researchers summarize the total number of words with positive and negative emotions, and calculate the percentages of positive and negative emotion words as the final result.

In this module, the attribute evaluation tables are scanned in turn, until all the attributes are covered. First, the system needs to read a record and extract adjectives in the record; then match adjectives with words in the dictionary. If the match is unsuccessful, the system will consider that the adjective cannot be used as the evaluation of that attribute. Then the emotional analysis is cancelled. If the matching is successful, the system needs to judge whether the emotional polarity of that word is positive. If it is positive, 1 score will be added in the evaluation result of that attribute. Otherwise, the system needs to judge whether the emotional polarity of that word is negative. If it is negative, 1 score will be reduced in the evaluation result of that attribute. Finally, after scanning, the ratio of positive and negative emotions of each attribute is calculated, and then the emotion value is obtained.

**System evaluation.** Two-way communication algorithm is used to extract evaluation objects and evaluative words. Three indicators are adopted to evaluate information, namely Precision Rate, Recall Rate and F value.

Precision rate = the number of correct evaluation objects / the total number of evaluation objects selected by this method (3-2)

Recall rate = the number of correct evaluation objects / the number of hand-classified evaluation objects (3-3)

The F value is the harmonic mean of precision rate and recall rate; it is a comprehensive index of the two indicators, which is used to reflect the overall situation:

$$F\ Value = 2*P*R / (P+R) \quad (3\text{-}4)$$

Precision rate is used to reflect the classification accuracy of the system from the perspective of quality; recall rate is used to investigate the classification completeness of the system from the perspective of quantity. The two indicators supplement each other, and make a more comprehensive reflection on the effect of the algorithm from two different aspects.

Automobile field

| Automobile field | | |
|---|---|---|
| | Evaluation object | Evaluative terms |
| Precision rate | 77.35% | 78.26% |
| Recall rate | 74.76% | 76.41% |
| The value of F | 76.033% | 77.32% |

According to above data, it can be seen that, the effect of evaluation words and objects obtained through two-way communication algorithm is relatively ideal, which can achieve expected results.

## Design of Interactive System

**Interactive process of the system.** This system is a vertical search site with web application which can help users to choose cars. The relatively simple home page of the site is a search box. Users can search information on relevant models of cars through this search box.

The home page of the site is in a relatively simple style. After the user typing text in the search box and clicking the search icon on the right side, the website will jump to the page of basic information of that car. Then, if the user click the analyze button and attribute button, the analysis results of that automobile will be shown. The analysis results show the evaluation results on each attribute of that car; the percentage represents the degree of users' evaluation on that attribute. There is a positive evaluation by default. On the other hand, there's an overall label of that car. The label is displayed as text cloud, allowing users to intuitively understand the characteristics of the car. Sample results are shown in the following figure:

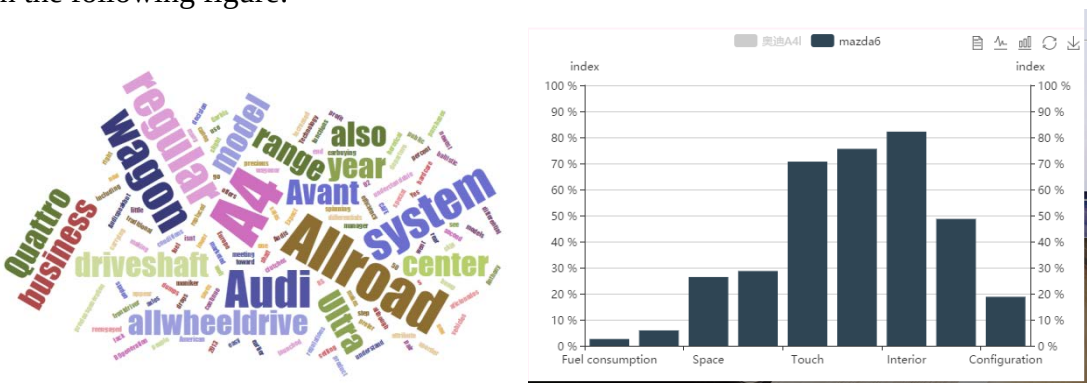


Figure 4-3 Sentimental analysis result of comments on automobile

Through visualization technology, data can be displayed clearly, which enables users to get information they wanted easily.

**Overall evaluation of user experience.** After completing the design of the whole system, it is necessary to test the functions of that system. Because the system can be accessed through online website, some users are invited to visit the site. The contents of the test include accessible of the site, information display after users typing search text, accuracy of analysis results and interactive experience. The system is tested by some students in our department. Then users input in car models in the search box. In that process, it is found that some car models cannot be searched, and the error information page can be seen. This part needs further processing. For car models which can be searched, basic information can be displayed. Afterwards, analyses on the evaluation tendencies of car models are made. Then related buttons are clicked to display search results.

In the whole operation process, the overall performance of the system is good. Some points in interactive experience still need to be improved, including the handling of errors, and the overall running speed of system.

## Conclusions

In this paper, sentiment analysis is used the automobile field by analyzing user comments. A user orientated evaluation system for automobiles is designed and implemented. Through related experiments, it is shown that this system realizes the collecting and displaying of basic information of automobiles, the analysis of users' emotional tendencies on automobiles, and the visualized representing of relevant results, which is convenient for users to obtain relevant information. The operation of the whole system is basically normal; relative functions can also be realized.

Next, we need to constantly improve and enrich functions of the system, like improving the search function of car model, improving operation speed, and developing a variety of methods to extract car review texts and judge emotional tendencies.

## References

[1] G. Qiu, et al., Opinion word expansion and target extraction through double propagation, J. Computational linguistics. 37 (2011) 9-27.

[2] Y.Y. Zhao, B. Qin, T. Liu, Text Semimetal Analysis, J. Journal of Software. 21 (2010) 1834-1848.

[3] M.S. Ni, H.F. Lin, Mining of commodity reviews based on association rules and polarity analysis. The Third National Conference on Information Retrieval and Content Security, Vol. 635, 2007

[4] M.Q. Hu, B. Liu, Mining and summarizing customer reviews. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2004.

[5]D.C. Wei, T.F. Yao, Analysis on Chinese semantic polarity sentences and methods of opinion extraction, J. Journal of Computer Applications. 26 (2006).