Quantitative Study of Commodity Review Text Based On Natural Language Processing

[1]School of Computer and Communication, Hunan Institute of Engineering Xiangtan 411104, China
[2]School of Finance and Information, Ningbo University of Finance and EconomicsNingbo, 315300,China

Zhou Zi-kai[1]
Email: 2916446385@qq.com
TangZi-pei[2]
41423804@qq.com
Zeng Sai-feng[1]
Email: 156081946@qq.com
TangZhi-hang[1*]
Email:zhtang@hnie.edu.cn

--**ABSTRACT**--

Choosing trusted reviews is the primary way that consumers get a true feeling about buying products online. At present, many scholars study the authenticity and usefulness of shopping reviews, but it is difficult for us to effectively quantify review texts. This paper designs a quantitative method of commodity reviews based on Baidu natural language processing API interface. According to the technical means of Python Crawler, the reviewtext is extracted from the shoppingweb page, and the C-F model is established to analyze the review text quantitatively by calling the natural language processing technology of Baidu AI open platform. In this approach, the problems of natural language quantification and trust can be solved effectively.

I. OVERVIEW

Since its inception in 1995, Chinese e-commerce has been experienced more than 20 years. While people want to make a purchase decision online, they will browse other customers' past shopping comments. This traditional method is mainly based on consumers' own ideas to take a decision, and the individual subjectivity is relatively strong[1]. Therefore, it greatly narrows the convenience of online shopping. Because physical objects are limited by space, consumers can't get the real experience. Therefore, their buying psychology is easily affected by the evaluation of others. What's more, comments are sometimes watered down by sellers who "buy the comments"[2]. In this case, the consumers' experience is further reduced. So choosing reviews with a high degree of trust is the first and foremost way for consumers to get a true feeling for online shopping. In order to filter trusted comments, this article provides a call to Baidu AI's natural language processing API interface to achieve the review text, and the python language is easy to understand and the code is short and concise, which is very friendly to beginners.Natural Language Processing (NLP) is widely used in human-computer interaction (HCI), but its analysis and recognition function have not been fully utilized in commodity review[3].This paper uses the interface of Baidu natural language processing API and the method of establishing C-F model to quantify comments for the first time [4], which realizes the fast selection of comments and the ranking of comments with high credibility during shopping, and provides users with reliable decision information. In addition, the basic process of dealing with online shopping comments is expounded.

II. REVIEW TEXTS CRAWL

To begin quantifying the review text, this article first get the experimental data, which are obtained from the website using the python crawler.

2.1Key parameter acquisition

Log to JD.com first, and take random JD.com stores as an example.
The website is:
URL=https://item.jd.com/11244282628.html.
Click on preview on the left to see that the comment text is in content under fetchJSON.Open headers next to it, record the cookie , callback, referer, and user-Agent values, and you are ready to design the code.

2.2 Python Crawler Design

The design environment uses the PyCharm compiler, the Anaconda compiler version 3.8, which needs to call the request library, requests the web page, calls the json Parsing Library, parses the request json file, transforms the json-formatted data into the dictionary, or converting strings into dictionaries for further comment; calling Random Library to set the request time for the requested page; #-*-Coding: utf-8 to set the encoding format GB18030; Please return comments content, time creation time, color productcolor, and set productsizein comments; use the format function to page comments; call the sys library to save output directly to Shuju.csv In a CSV file.

Figure 1.The result of data crawling.

III. INTRODUCTION OF AI OPEN PLATFORM ON BAIDU

Artificial intelligenceis the core ofBaidu's open platform of AI [5], which can drive force of the new round of technological and industrial revolution interacting with the real economy, and the realization of technological breakthroughs and ecological prosperity. It gives all walks of life intelligent upgrade, application landing and industrial services [6], to achieve talent training and AI Service Society. Through the open API of natural language processing services, developers can call the interface to achieve the information data processing with highly accuracy and reliability.

3.1 Baidu API call design

First log in Baidu smart cloud, create an application, fill in the application name, application type, and interface selection of Natural Language Processing, Text Audit; fill in the application description immediately click to create, in the refresh page to get APP, API Key and Secret Key.

3.1.1 Access token

To read the complete authentication mechanism [7] in the novice guide, we need to use the address to the authorization service when invoking the API:
https://aip.baidubce.com/oauth/2.0/token

With the APP, API Key, and Secret Key parameters obtained in the URL, POST the request and return the access parameter:

- **grant_type**: 必须参数，固定为client_credentials;
- **client_id**: 必须参数，应用的API Key;
- **client_secret**: 必须参数，应用的Secret Key;

Figure 2.Request URL parameter diagram.


```
"access_token":"24.f8158e624934f6d3f67fe89b86f7fc9
```

Figure 3.The Access token gets the result display.

3.1.2 Sentiment profiling API calls

Read the introduction for the sentiment profiling API, which has the HTTP method of the request and the request header:URL:https://aip.baidubce.com/rpc/2.0/nlp/v1/sentiment_classify.
We need to take access_token value at the end.


```
[{"positive_prob": 0.999994, "confidence": 0.999986, "negative_prob": 6.39982e-06, "sentiment": 2}]]
```

Figure 4. Shows the result of the sentiment profiling API call.

3.1.3 Call to the dialog emotion recognition API

Read the introduction for the sentiment profiling API, which has the HTTP method of the request and the request header:
URL:https://aip.baidubce.com/rpc/2.0/nlp/v1/sentiment_classify

We need to takeaccess_token value at the end.


```
"prob": 0.891293, "label": "optimistic"},
```

Figure 5. Shows the result of the call to the dialog emotion recognition API.

3.1.4 Text audit API interface calls

Read the introduction for the text auditing API, which contains the HTTP method requested and the request header:
URL:https://aip.baidubce.com/rest/2.0/solution/v1/text_censor/v2/user_defined
Weneed to include the APP_ID, API_KEY and SECRET_KEY at the end.


```
"msg": "疑似存在恶意推广不合规", "type": 12, "conclusion": "疑似"}], "conclusion": "疑似"}
```

Figure 6.Shows the result of the text review API call.

IV. CREDIBILITYMETHOD

Credibility [8] is the degree of belief that a thing or phenomenon is true based on experience, which is subjective and empirical, and its accuracy is difficult to grasp. The credibility method is also called C-F model. In 1975, E.H.Shortliffe of Stanford University and others proposed an uncertain reasoning model after investigating the non-probabilistic and non-formal reasoning process, which based on reliability representation and called the basic method of uncertainty reasoning [9]. It was first used in hematological diagnosis expert system MYCIN in 1976. It is a very simple and effective reasoning method in uncertainty reasoning. At present, many successful expert systems[10] are based on this method. In this method, CF (Certainty Factor) is used as uncertainty measure, and by calculating CF(H,E), the quantitative support degree of evidence E to hypothesis H is discussed.E for evidence, H for hypothesis, and E to assume that CF(H,E) for the rule corresponding to the knowledge intensity.

(1) MB (measure belief, MB) is the trust growth, which indicates the degree to which conclusion H is true because of evidence matching the prerequisite E.
(2) MD (measure disbelief, MD) is called the increase in distrust, and it represents the increase in distrust of conclusion H. due to evidence matching the premise E.
(3) MB, MD satisfy the law of mutual exclusion, that is, when $MB(H,E) > 0$, $MD(H,E) = 0$; when $MD(H,E) > 0$, $MB(H.E) = 0$. So $CF(H)$ can be represented as following:
(4) The uncertainty reasoning in the model is based on the uncertain initial evidence, and through the use of relevant uncertainty reasoning rules, the conclusion is finally deduced and the credibility value of the conclusion is obtained. Conclusion reliability of $CF(H)$ is calculated by the following formula:
(5) The Deterministic algorithm of the evidence areas following:

(6) Experimental analysis of review credibility evaluation: A random comment: "This is a high-quality goods, iphone is really high-end atmospheric grade. Its touch is a very good. It feels very safe because of its lightness. The sub-operation is expected to be very smooth, the screen is very clear, no impurities, quick response. Appropriate size and weight, and it's very easy to use. It was a nice shopping experience, and thelogisticswere really fast. Appearance Material: metal material. Screen Effect: clear and smooth.Runningspeed: very smooth. ThinDegree: Suitable. OtherFeature: Just the right size and weight.

By calling Baidu's natural language processing API, the following information is obtained: The probability of getting rid of the default favorable comments is set to a priori probability $P(H)=0.71$, and a posterior probability $P(H|E)$ is the confidence given by Baidu:

Table 1. Comment parameter values

	Affective disposition analysis	Dialogue emotion recognition	Text Review
Analyze the tag	Positive emotion	Like	Malicious promotion
Degree of confidence	0.99	0.89	0.92
P (H|E)	0.99	089	0.08
P (H)	0.71	0.71	0.71
CF(H,E)	0.96	0.62	-0.88
CF(E)	0.99	0.89	0.92

Using the formula of the C-F method, the credibility of the composite evidence is calculated to be 0.84:

$CF(H1)=0.96*MAX\{0,CF(E1)\}=0.95$
$CF(H2)=0.62*MAX\{0,CF(E2)\}=0.55$
$CF(H3)=-0.88*MAX\{0,CF(E3)\}= -0.81$
$CF(H1,2)=CF(H1)+CF(H2)-CF(H1)*CF(H2)=0.97$
$CF((H1,2)3)=CF(H1,2)+CF(H3)/1-min$ 〔(H1,2)|,|CF(H3)|〕 $=0.84$

V. SUMMARY

This article uses python to randomly grab review texts from a Web site in JD.com, and calls Baidu AI's natural language processing API to input review texts, output tagged review texts and confidence values, the uncertainty reasoning model is established by using C-F method, the confidence factors of each feature are derived, and then the composite feature confidence rate is calculated. After quantifying, the trust rate can be ranked, showing the high frequency of trust rate to consumers first, which can reduce the interference of "brush single favorable comment", and can reduce the screening time, improve the shopping experience. While reviews of shopping text are a major part, commenting on images and videos is also important, allowing researchers to combine more of the AI's methods with features to further improve the accuracy of the

selection, providing users with higher quality reviews helps them analyze the product objectively and rationally, thus avoiding being misled[11].

REFERENCES

[1]. Zeng Xiaoqin, Yu Hong. Sentiment analysis of commodity review text based on Python J. Computer Literacy and technology, 2020,1608:181-183.

[2]. Zhu Caiwei. The criminal law analysis of the behavior of letter-printing. Zhejiang Gongshang University, 2020.

[3]. The Gardener. The development and application of natural language processing technology in the era of artificial intelligence J. Office automation, 2019,2410:63-64.

[4]. GaoXin, Yang Ruyan, Guan Xiaofei, Wu Yong. A quantitative model of credibility for wine evaluation by WINE EXPERTS AND ITS APPLICATION J. Proceedings of the Yunnan Agricultural University, 2014,2902:235-240.

[5]. XuFangfang. Baidu brain: Born into the future. China Science and Technology Awards, 201811:50-53 + 82.

[6]. Wang Qiurong, Yu Zhihong. Take on the responsibility of AI for the benefit of mankind — A visit to vice president of Baidu Group, vice director of national engineering laboratory of deep learning technology and application, Wu Tianj. Journal of Sustainable Development Economics, 202004:21-23.

[7]. HaoErwei. Research on implementation technology of online translation software based on Baidu Cloud Service J. World of digital communications, 201912:106 + 7.

[8]. Shi Minghui, Zhou Changle, Wu Qingfeng, Wu Yun, Zhang Zhifeng. Uncertainty reasoning based on credibility and its neural network implementation of J. Computer Applications Research, 200701:241-243 + 312.

[9]. Chen Xiaoyu. Research and implementation of uncertainty inference engine based on Expert System J. Manufacturing automation, 2011,3318:78-81.

[10]. Zhang Yudong, Wu Le Nan, Wang Shuihua. Overview of Expert System Development J.

Computer engineering and applications, 2010,4619:43-47.

[11]. Ni Jinling. A discourse strategy for product description in e-commerce platform environment — A case study of Amazon platform appliances. Journal of Lanzhou Institute of Education, 2019, 3506: 133-135.

Biographies and Photographs

Tang Zhihang, male, professor, doctor, post doctor of South China University of Technology, young backbone teacher of Hunan Province, tutor of master's degree, academic leader of computer application of university level application characteristics discipline, mainly engaged in data mining, intelligent decision-making, big data and other research. He has published more than 50 academic papers in academic journals at home and abroad, of which more than 30 have been included in SCI and EI; He has presided over more than 10 provincial and ministerial projects such as China Post doctoral Science Foundation, Hunan Natural Science Foundation, and Hunan Postgraduate Teaching Platform Project; Instruct students to obtain 2 national undergraduate innovation and entrepreneurship training programs; Instruct students to win one third prize of Hunan Challenge Cup; Instructed students to win one bronze prize in Hunan University Students Entrepreneurship Competition for Youth Creation.