

Identifying and Ranking Dominating Product Features using NLP Technique

Sonali D. Ingale
Dept. of CS and IT
Dr. B. A. M. University,
Aurangabad-431004, India

Ratnadeep R. Deshmukh
Dept. of CS and IT
Dr. B. A. M. University,
Aurangabad-431004, India

ABSTRACT

As internet usage increased users use internet not only to access and search information but also at the same time able to spread and publish own idea, sentiments, knowledge via different number of websites. Different websites encourage their user to write their views in the form of electronic text. This system increasing user-written electronic text in the world of internet, large numbers of user opinions are available on World Wide Web. User review contains important information, which is beneficial for customer as well as retailer. These reviews are in scattered format so extracting important data from this large corpus is time consuming work. Here developing a system which will automatically identify and rank the product features. The Stanford parser is used for identify product features. Sentence level sentiment classification is used for identify sentiment of each sentence separately, Sentiment Classifier is used for classifying each sentence, and finally a probabilistic ranking algorithm is used to rank the product features.

General Terms

Sentiment classification, Sentence level sentiment classification, Natural Language Processing, probabilistic aspect ranking

Keywords

Sentiment classification, Sentence level sentiment classification, probabilistic aspect ranking.

1. INTRODUCTION

Sentiment Analysis also referred as sentiment mining, opinion mining, review mining, sentiment mapping, and emotion analysis etc. It is a field of research which study people's emotions, sentiments, appraisals, opinions about particular entity like organization, product, issue, and its assets [1].

Sentiment Analysis and Classification uses various techniques which are mainly divided in to two approaches they are Machine Learning and Lexicon based approach[2]. This two main approaches again sub divided into different categories like Machine learning is divided into supervised, unsupervised. Lexicon based is divided into dictionary based and corpus based approach [3].

For most of people "What other person believe "is always important side while decision making. In past days before putting money into any product people used to ask judgment to their family, friend circle and colleagues and then they take the decision [4]. In today's world there is a boom of World Wide Web, huge amount of data is available on internet so while purchasing a product rather than asking to particular person, customer take decisions by analyzing electronic text [5].

Customers will give more attention to particular product features while purchasing particular product. Generally a product has multiple features like if consider domain mobile, then it consist of features like camera, bluetooth , wifi ,battery etc[6].If particular go in practical way then reading all review and then finding the product important features is time consuming and lengthy process. Here need an automatic approach which will classify product features and rank them [7].

2. IMPLEMENTATION

In this system using Machine learning approach which is divided into supervised ,unsupervised and semi supervised methods[8].Here supervised classification is being implemented which requires training and testing dataset. So the training and testing dataset is downloaded from <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>,



Fig 1: How sentiment classification required for neutral sentence

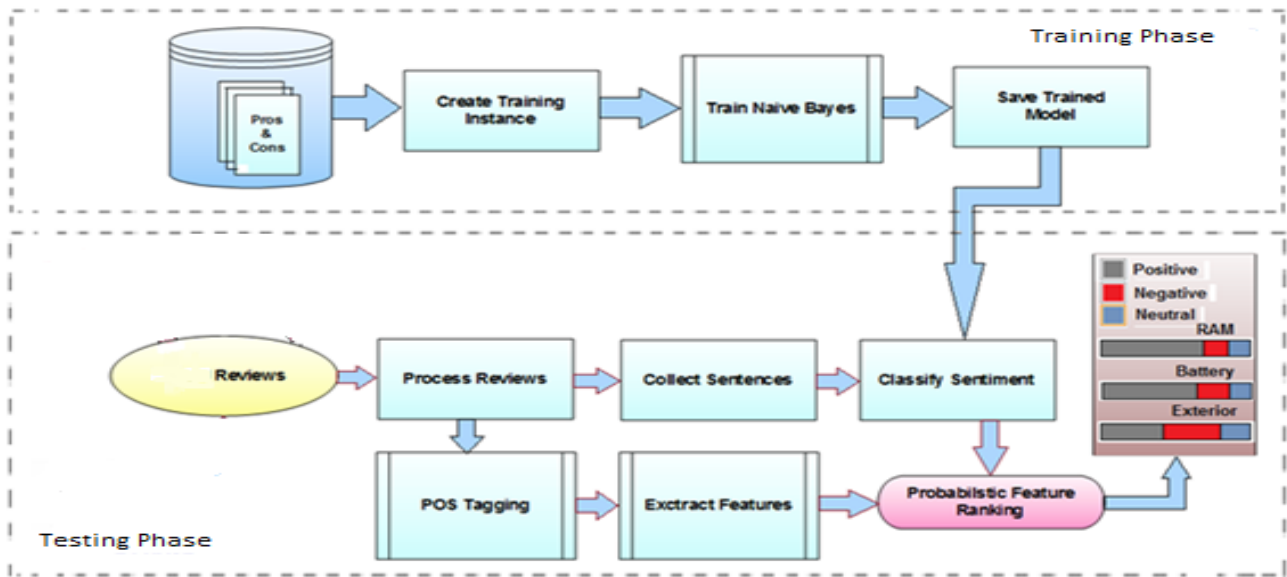


Fig 2: System Architecture

training dataset consist of total 6828 positive and negative words. Testing data set contain free text reviews written by user towards particular product[8]. Figure2 shows system architecture which is divided into two phase .Training phase and testing phase. one of the main phase is feature identification towards particular product.

2.1 Preprocessing of reviews

Here converting json files into xml files. So it will be easy to parse the sentences using Stanford parser.

2.2 Feature identification

To identify features of particular product author search out the noun terms from the free text reviews. for that purpose split the reviews in the sentence and then parse each sentence using the Stanford parser .Stanford parser is used by natural language processing to parse a sentence into part of speech tagging[9].

Example: Consider the below sentence

This Mobile has excellent camera.

Tagging

This/DT Mobile/NNP has/VBZ excellent/JJ camera/NN ./.

Parse

(ROOT

(S

(NP (DT This) (NNP Mobile))

(VP (VBZ has)

(NP (JJ excellent) (NN camera)))

(. .))

Typed dependencies

det(Mobile-2, This-1)

nsubj(has-3, Mobile-2)

root(ROOT-0, has-3)

amod(camera-5, excellent-4)

doobj(has-3, camera-5)

3. SENTIMENT CLASSIFICATION

Sentiment classifier is used for prediction of sentiment on each feature. Here using Pros and Cons review which contain positive and negative sentiment words, this reviews are useful to train opinion classifier, it will help further for free text reviews testing[10]. Free text review contains more than one feature. Here author consider sentiment expression which edits corresponding feature.

eg.consider the term "excellent" in the opinion "The mobile has excellent battery "for the feature "battery".

as the excellent is positive sentiment for feature battery. so it is categorized as positive sentence. For every positive and negative statement a weight is assigned. Here a threshold is set for neutral sentence as it has greater influence on result of system [11].

3.1 Naïve Bayes Classifier

It is a simple classifier, works well with text classification. Bayes classifier apply strong independence theorem. Here each attribute is considered as independent class variable. It applied to particular class problem, mainly to those in which phrased as related to an object with a different division. It is more accurate, simple and speedy classifier [12]. It considers each attribute equal to take final decision, which is computational efficient with compared to different sentiment classifiers.

3.2 Multinomial Naive Bayes Classifier

Particular situation in which multiple occurrences of particular word is have more effect in text classification then multinomial classification is used. For example it used in topic classification. Below formula shows how Multinomial Naïve Bayes classifier works

$$\hat{P}(t|c) = \frac{T_{ct} + 1}{\sum_{t' \in V} (T_{ct'} + 1)} = \frac{T_{ct} + 1}{(\sum_{t' \in V} T_{ct'}) + B'}$$

where Tct is the number of occurrences of t in training dataset from class c including more than one occurrence of a term in a document is number of terms in the vocabulary[13].

4. ASPECT RANKING

A simple bubble sorting formula is used for ranking of features, before ranking features here developer applied some weight to the positive, negative and neutral sentiment features. For positive 3 number is assigned as it has higher preference

,for neutral 2,and negative 1.Neutral will gives effect on positive and negative terms so it assign second importance. Figure 3 shows graphical representation of ranking features for domain mobile.

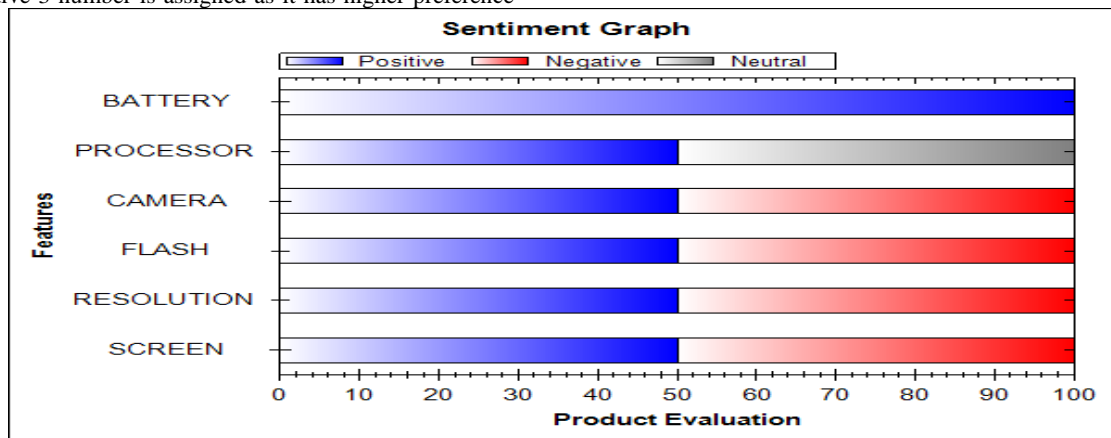


Fig 3:Graphical result of aspect ranking

5. RESULT ANALYSIS

Training Dataset consist of Total 6824 words in which Pros word count is 2012 and cons word count is 4812.Table1 shows details about testing dataset.

Total 8 products of 4 different domains like Camera, Printer, TV and Mobile were taken for implementation of system. Top 10 features of different domain products are displayed using graph for different features accuracy of aspect ranking for each product is displayed in Table 2.total average accuracy for aspect ranking is 0.78 that is 78%.

Table2: performance of sentence level sentiment Classification by product (AR) Aspect Ranking

Domain	A.R
Nikon Coolpix s9200	0.9
Canon 60D	0.71
Hp5550	1.0
Panasonic VIERA	0.7
Samsung S4	0.8
LG 840G	0.7
Nokia 520	0.7
MotoG	0.8
Average	0.78

Table1: Statistics of data Corpus, # indicates the Number of Reviews/Sentences

Product Name	Domain	#of reviews	#of sentences
Nikon Coolpix s9200	Camera	107	505
Canon 60D		17	175
Hp5550	Printer	186	976
Panasonic VIERA	TV	244	3502
Samsung S4	Mobile	149	2294
LG 840G		767	3891
Nokia 520		882	4367
MotoG		906	3657
	Total	3258	19367

6. CONCLUSION

Mobile, Camera, Printer, TV are dominating product in now days so author has taken their reviews for sentiment mining. The Stanford parser is used for identify product features. Sentence level sentiment classification is used for identify sentiment of each sentence separately, and finally a

probabilistic ranking algorithm is used to rank the product features and ranking the features of the product. Here author performed product feature ranking using sentence level classification. Particular feature performance improvement can be obtained by using the product feature ranking. Major of work is done in the field on sentence level, document level sentiment analysis ,but still lot of major problems and

challenges are present like , grammatical mistakes in English ,fake reviews, single word with different meaning ,recently there is a trend change in expressing feelings , user write comments in the format of smiley, images ,etc.so sentiment mining will be always a ongoing research field for present and future researchers.

7. ACKNOWLEDGMENTS

I am highly indebted to Dr. R.R.Deshmukh, Professor and Head of the Department, Computer Science and Information Technology,Dr.Babasaheb Ambedkar Marathwada University Aurangabad ,for his encouragement towards the completion of this project.This project has been supported by Department of Computer Science and IT Dr.Babasaheb Ambedkar Marathwada University, Aurangabad. The views expressed here are those of the authors only.

8. REFERENCES

- [1] Singh and Vivek Kumar, “A clustering and opinion mining approach to socio-political analysis of the blogosphere”. Computational Intelligence and Computing Research (ICCR), 2010 IEEE
- [2] A. Ghose and P. G. Ipeirotis, “Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics,” IEEE Trans. Knowl. Data Eng., vol. 23, no. 10, pp. 1498–1512. Sept. 2010.
- [3] Zheng-Jun Zha, Member, IEEE, Jianxing Yu, Jinhui Tang, Member, IEEE, Meng Wang, Member, IEEE, And Tat-Seng Chua, “Product Aspect Ranking And Its Applications”, IEEE Transactions On Knowledge And Data Engineering, Vol. 26, No. 5, May 2014.
- [4] M.Govindarajan ,Romina M, “A Survey of Classification Methods and Applications for Sentiment Analysis ”, The International Journal Of Engineering And Science (IJES) ||Volume||2 ||Issue|| 12||Pages|| 11-15||2013|| ISSN(e): 2319 – 1813 ISSN(p): 2319 – 1805.
- [5] Neha S. Joshi, Suhasini A. Itkat, “A Survey on Feature Level Sentiment Analysis”, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5422-5425.
- [6] G.Vinodhini, R.M.Chandrasekaran, “Sentiment Analysis and Opinion Mining: A Survey”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 6, June 2012.
- [7] Ms.Sonali. D.Ingale, Dr.R.R.Deshmukh, “A Survey on sentiment mining”, International Journal on Recent and Innovation Trends in Computing and Communication Volume: 2 Issue: 11 ISSN: 2321-8169.
- [8] Bing Xiang, Liang Zhou, “Improving Twitter Sentiment Analysis with Topic-Based Mixture Modeling and Semi-Supervised Training”, Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 434–439, Baltimore, Maryland, USA, June 23-25 2014.
- [9] N, Anwer and A, Rashid “Feature Based Opinion Mining of Online Free Format Customer Reviews Using Frequency Distribution and Bayesian Statistics” Networked Computing and Advanced Information Management (NCM), 2010 Sixth International Conference on 16-18 Aug. 2010.
- [10] M Fan, G WU “Opinion Summarization of Customer comments” International conference on Applied Physics and Industrial Engineering in 2012.
- [11] N. M. Shelke, S. Deshpande and V. Thakre “Survey of Techniques for Opinion Mining” International Journal of Computer Applications (0975 – 8887) Volume 57– No.13, November 2012.
- [12] Bing Liu, Sentiment Analysis and Opinion Mining, Morgan and Claypool Publishers, May 2012.p.18-19,27-28,44-45,47,90-101.
- [13] Bing Liu, “Sentiment Analysis and Opinion Mining”, EBook “www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf”.