



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

Sentiment and Intent Analysis for Business Intelligence

Lalit Pathak¹, Sanjana Mulchandani², Mayuri Kate³, Kajal Khatwani⁴, Sujata Khedkar⁵

U.G. Student, Department of Computer Engineering, VESIT Engineering College, Mumbai, Maharashtra, India¹

U.G. Student, Department of Computer Engineering, VESIT Engineering College, Mumbai, Maharashtra, India²

U.G. Student, Department of Computer Engineering, VESIT Engineering College, Mumbai, Maharashtra, India³

U.G. Student, Department of Computer Engineering, VESIT Engineering College, Mumbai, Maharashtra, India⁴

Associate Professor, Department of Computer Engineering, VESIT Engineering College, Mumbai, Maharashtra, India⁵

ABSTRACT: Intent mining is the field of extracting and studying people's opinions, sentiments, attitudes, and emotions expressed in different digital form in form of reviews on e-commerce and other social networking sites. Therefore, our aim is on extracting features that enable an effective classification of PIs from user's reviews.

The prime objective of our system for consumer is to predict the user's intent and provide an overview of sentimental analysis of a particular service or product, the user tends to use or buy also to discover the demand for the brands and services. The main aim of our system is to know how our product is perceived in the business environment. This helps us to make the marketing strategy also beneficial for the growth of the business. In order to implement this, we have obtained the reviews by scraping the amazon website, Quora reviews, twitter database.

KEYWORDS: Intent mining, social media mining, natural language processing, text classification, purchase intent, sentiment analysis, emoticons polarity, feature extraction, text classification.

I. INTRODUCTION

Intent mining is also known as sentiment analyser which involves building a system to gather and examine opinions about the product made in reviews. A system that analyses the customer reviews and find out important information from the reviews is very beneficial to the user. Online data is popularly used for extracting public opinions about brands, products, services. Thus, the art of data mining helps us to extract the opinions about the products and services as well as the summary of the item. The outcome of this system will empower marketers and business owners to derive knowledge from large amounts of online data and to formulate decisions, converting data into actionable knowledge. It will also help customers to know more about the product or service that they intend to buy, thereby making it easier for them to select a low-cost and better-quality product.

The Framework is designed for:

- i) The Framework will summarize the intents of the users graphically as positive, negative and neutral and also provides rating for all the products.
- ii) The framework is able to visualize the results via interactive and intuitive charts and graphs for both managers and customers.
- iii) It can provide suggestion to the users about particular product or service.
- iv) It can suggest business intellectuals for any changes or recommendations in existing product and service.
- v) Analyses the purchase intents and displays the information in dashboard.
- vi) To Displaytop five intents or tweets.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 1, January 2018

II. RELATED WORK

Several tools already exist to help companies extract and analyse information from blogs about large-scale trends in customers' opinions about products. However, are limited to polarity evaluation or mood classification according to a limited set of emotions. The existing system does not provide sentiment based intent analysis that can aid both customers and business owners. Intent Analysis goes a level deeper than sentiment analysis and gives an idea of whether a string of text is a complaint, a suggestion or a purchase/consumption intent. Most tools do not consider emoticons for intent and sentiment analysis. The proposed system takes into consideration various emoticons, highlighted and emphasized texts used in reviews and accordingly assigns polarity to the reviews.

III. EXISTING SYSTEM

Several tools already exist to help companies to extract and analyse information from blogs about customers' opinions about products and services. However, are limited to polarity evaluation or mood classification according to a limited set of emotions. The existing system does not provide sentiment based intent analysis that can aid both customers and business owners. Intent Analysis goes a level deeper than sentiment analysis and gives an idea of whether a string of text is a complaint, a suggestion or purchase/consumption intent. Most tools do not consider emoticons for intent and sentiment analysis. The proposed system takes into consideration various emoticons, highlighted and emphasized texts used in reviews and accordingly assign polarity to the reviews. The motivation behind this project was primarily an interest in extracting and analysing data from various sources and social networking sites to make it easier for business stakeholders to understand about the products and services. That may include the entire sales cycle, from winning new customers, to servicing and tracking existing customers, to providing post-sales services. This will also help businesses to gain insights into consumer behaviour and suggest areas that need improvement.

IV. METHODS AND TOOLS

A. *Key Phrase Extraction* - We can't directly take the text available on the internet for processing as it is simply a string of characters. The primary step involves cleaning the input text so that processing can be done in later phases. We are using the Natural Language Toolkit for this. The "nltk" provides most of the tools that is required for text cleaning and processing.

1) *Parsing of text*: This involves identifying words and sentences in the text, which is identified by the spaces, punctuations and the other non-alphanumeric characters. At first step, the entire text is split into sentences by noting the location of the punctuation marks viz. ".", "?", "!"

2) *Tokenization*: Then, whole of the text or each of the sentence is converted into tokenized words. This converts each sentence into a list of words. It is to be noted that punctuations like ",", ";", etc. form individual tokens. Converting the text into a list of tokens is important because this helps applying the tools of natural language to the text.

3) *Part of Speech Tagging*: After the text, has been parsed and tokenized, the part of speech of each tokenized word is identified using the tools of Natural Language Toolkit. This identifies the words as nouns, adjectives, verbs, determiners etc.

4) *Listing candidate keywords*: A set of all the candidate keywords are created. For this, visit all the words in the text and then choose the ones that are noun or adjectives preceding the noun. The final output is consisting of the set of candidate keywords that have the potential of becoming a keyword defining the text in a way or the other.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

5) *Identifying all the co-occurring words*: Further task is to identify the interconnection of that word with the other words in the text. For this, a window is selected on N words, usually N ranges between 2 and 10. For a particular candidate word, if another candidate word is found within a window of N words, then the two words are defined to co-occur and thus interconnected.

6) *Scoring of the candidate keywords*: The scoring of the keywords depends upon the pattern of occurrence of the word in the text. In our approach, we used a blend of frequency of the occurrence of the word and also frequency with which a word has co-occurred with other words.

7) *From Keywords to Key-phrases*: It is observed that the keywords generated now are all unigrams, but obviously, phrases convey central idea better the unigrams. We can use these generated keywords to extract phrases from the text. To extract the phrases, firstly all the possible phrases are extracted from the text.

B. Emoticons Extraction for sentiment analysis:

1) *Detecting the polarity of a sentence*: We shall use SVM along with unigram features for words in combination with sentiment dictionary like SentiWordNet

2) *Detecting the polarity of an Emoji*: we shall use the ESR for obtaining the polarity of emojis.

3) *Comparison of polarities*: In order to compare sentiment polarity of a sentence, enhancement of sentiment (i.e., polarity of emoji with polarity of sentence), or modification of sentiment, comparing emoji polarity with sentence polarity. The last step is to automatically label sentences with their usages like sentiment expression.

C. Classification of review as intent or non-intent:

1) *Feature Extraction*: We are going to choose features for designing an efficient classification model. A Purchase Intent post consists of Intent and corresponding Usable product. Let us consider an example where 'buy' is the consumption intent and 'phone' is the consumable object (e.g. "A good phone to buy is iPhone 6"). We use the parse tree for analysing this component.

2) *Purchase Object categories*: An intent for purchasing is expressed in terms of a product or a service. Therefore, identifying whether the product or service in a review is usable or not is vital for classification of Intents. A usable product is described using a noun phrase (NP) (e.g. "Which bag should I buy?"). The purchase intent for the given example is "buy" and purchase object is "bag".

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

V. CLASSIFICATION OF REVIEWS

We have classified reviews as shown in following figure.

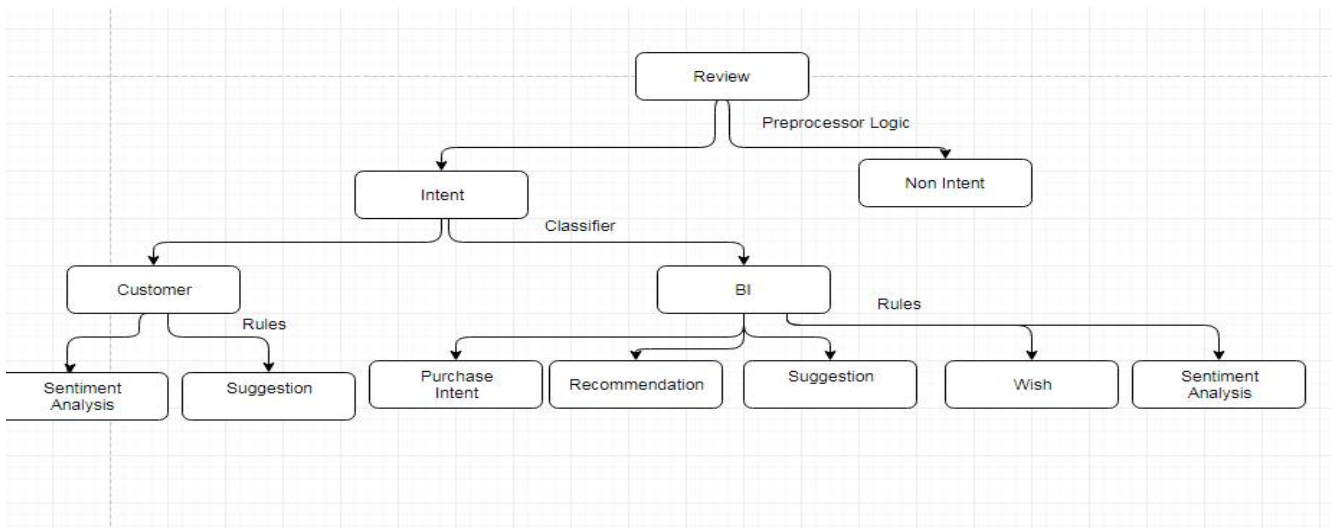


Fig 1.Intent Classification

VI. TOOLS

- SVM classifier provided by LIBSVM toolkit.
- Amazon Mechanical Turk (AMT) to recruit human annotators.
- PCFG parser in Stanford NLP tool.
- WordNet.
- Freebase.
- Dependency parser provided by Stanford NLP.
- NLTK.
- AlchemyAPI.
- Tweepy.

VII. FUTURE SCOPE

We would like to achieve the Domain independent output in our system. We are planning to expand our system for different services. Many times, the reviews given by user are fake. So, we will try to find out the sarcasm in reviews and Handling the sarcasm in reviews for sentiment analysis by studying and implementing user. Also, we are trying to implement our system in time analysis manner, so that consumer and business owner both can get the benefit from that. Business owner can decide the quantity for the production by analysing the purchase intent and consumer can also get the of timeline analysis to know that time period in which product has been on boost. Also, we are trying to integrate the intent of text and emoticons and give the final rating based on polarity of both text and emoticons.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 6, Issue 1, January 2018

VIII. RESULTS

The following section contains the screenshots of our system output. In figure 2 we are showing the timeline analysis for the product year wise. Graph shows the year and sold products ratio. Also in figure 3 we are give the suggestions to business persons. In the figure 4 we are giving top 10 most helpful reviews for user. In figure 5 we are giving aggregate sentiment analysis to user. In figure 6 we are trying to show the comparative study of two products. In figure 7 we are giving intent analysis to user and in figure 8 we are giving purchase intent to business people.

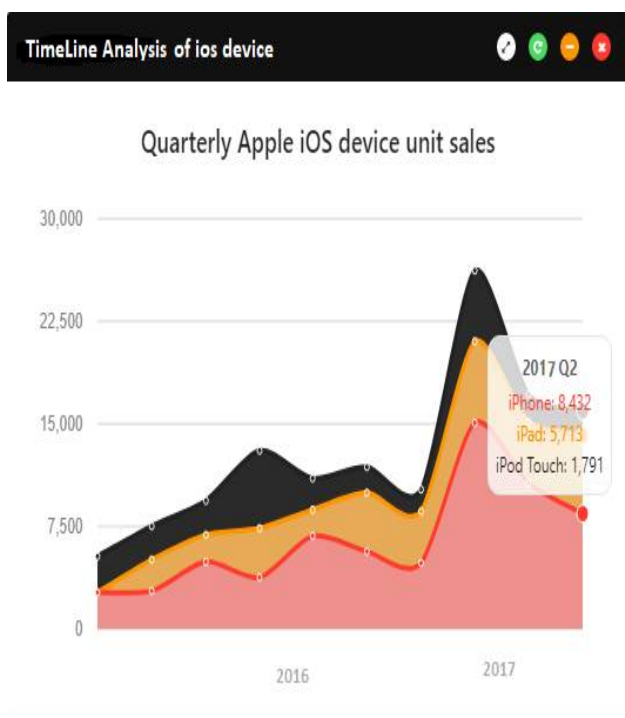


Fig 2. TimeLine analysis



Fig 3. Suggestions for BI

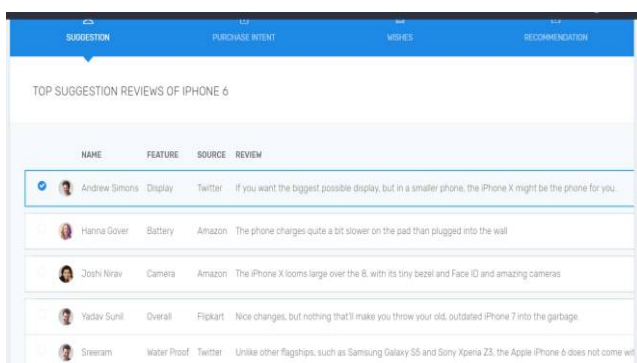


Fig4. Top 5 reviews



Fig 5. Sentiment analysis

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

Comparative Study of OS

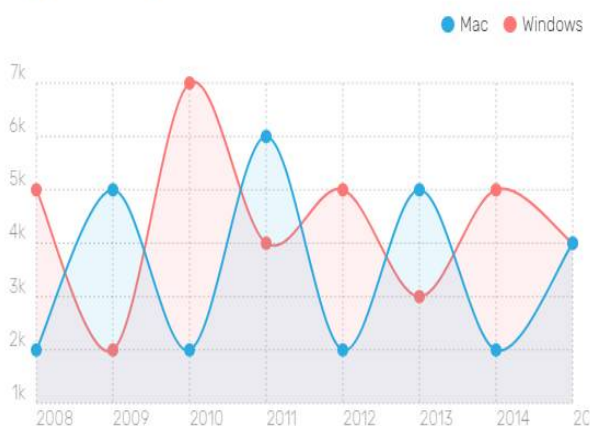


Fig 6. Comparative study

Intent Analysis



Fig 7. Intent analysis

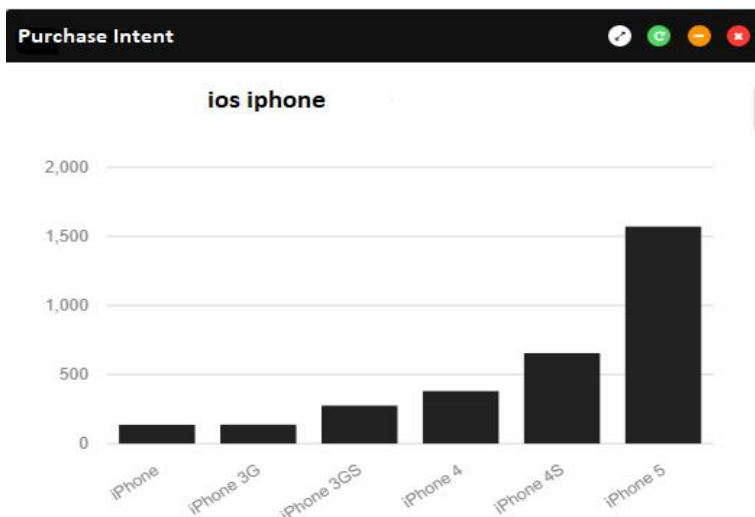


Fig 8. Purchase intent

IX. APPLICATIONS

There are many fields in which Intent Mining may find its use.

- 1) Before launching a new product/service the adoption Intent of users can be mined. By finding the Consumption Intent of a large number of users in a city, warehouses can be prepared and delivery facilities can be arranged in



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

accurate manner. It can also be used in predicting the demand of a product/service before starting a business in a particular area, based on web search data of potential users.

2) It can be used in the field of Targeted Marketing.

3) Intent Mining might be used in detection of fraud and criminal intent.

4) Intent Mining may be used in analysis of the response of a new policy introduced by the government.

5) Predicting the inflow of tourists and timely preparation in a tourist destination could be a potential use of Intent mining.

X. CONCLUSION

We analyse the reviews based on user's intentions. The system also generates reports for the manufacturers as well as customers. We are developing a framework for both managers and customers of system to help analyse unstructured review text and get to the core of features of the products and understand their performance with various users and get the positive aspects as well as the negative aspects of the product. This can be effectively utilized by the managers to enhance their market reach and improve their products. A dashboard of all the products is created wherein the user can view the product and its ratings based on reviews, also the user can get to know about the features of the product. Thus, it would be easier for users looking out for a specific set of features, to make a correct choice of a product. Thus, due to feature based product analysis, manufacturers can get to know the proper reason behind the failure of their product. Manufacturers can identify the features which were not found to be up to the mark based on the reviews, and thus work upon them to deliver a better product. Since our analysis is totally review based, a product's popularity would be a reflection of the product's quality. With declining popularity, users can look for better products which serve the same purpose. The system also gives product recommendation to customer. An amateur using this, is not misguided by flashy advertisements or false promotional tactics, but instead would become aware about the actual quality of the product with the help of our system, which is based on genuine reviews by the people about the product.

REFERENCES

- 1) Aditi Sharan, Siddiqi, Sifatullah, "Keyword and key phrase extraction techniques: A literature review", International Journal of Computer Applications" 109, (2015).
- 2) Chang, C.-C., and Lin, C.-J. 2011. Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST).
- 3) Ashkan, A., and Clarke, C. L. 2009. Term-based commercial intent analysis. In SIGIR, 800–801.
- 4) Zhu, X.; Ghahramani, Z.; Lafferty, J.; et al. 2003. Semi-supervised learning using gaussian fields and harmonic functions. In ICML, volume 3, 912–919.
- 5) Li Dong, Furu Wei, YajuanDuan, Xiaohua Liu, Ming Zhou, and Ke Xu. 2013. The automated acquisition of suggestions from tweets. In Marie desJardins and Michael L. Littman, editors, AAAI. AAAI Press.
- 6) Akbani, R.; Kwek, S.; and Japkowicz, N. 2004. Applying support vector machines to imbalanced datasets. In Proceedings of the 15th European Conference on Machine Learning (ECML), 39–50.
- 7) Li Dong, Furu Wei, YajuanDuan, The Automated Acquisition of Suggestions from Tweets, AAAI Conference.
- 8) Caroline Brun and Caroline Hagege. 2013. Suggestion mining: Detecting suggestions for improvement in user's comments. Research in Computing Science.
- 9) Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for intent mining. In Proceedings of the 5th Conference on Language Resources and Evaluation. (LREC06, pages 417–422.)
- 10) Ashkan, A., and Clarke, C. L. 2009. Term-based commercial intent analysis. In SIGIR, 800–801.
- 11) Beitzel, S. M.; Jensen, E. C.; Lewis, D. D.; Chowdhury, A.; and Frieder, O. 2007. Automatic classification of web queries using very large unlabeled query logs. ACM Trans. Inf. Syst. 25(2).
- 12) Belkin, M.; Matveeva, I.; and Niyogi, P. 2004. Regularization and semi-supervised learning on large graphs. In COLT, 624–638.