

ISAR: Implicit Sentiment Analysis of User Reviews

Chaitali Chandankhede¹, Pratik Devle², Abhijit Waskar³, Nikhil Chopdekar⁴, Sunil Patil⁵

Department of Information Technology, MIT Pune, India

Email: chaitali.chandankhede@mitpune.edu.in, pratik.devle@gmail.com, abhijitwaskar@gmail.com, nikhil.chopdekar93@gmail.com, sunilpatil143128@gmail.com

Abstract—Sentiment analysis is the process of analyzing the text about a topic written in a natural language and classify them as positive or negative based on the human sentiments, opinions expressed in it. Due to the increasing growth in use of social media (e.g. reviews, forum, blogs, Twitter and postings in social network sites) on the Web, users now have many opportunities to express their opinions about a product or topic. Users express their opinion through the reviews. These reviews are used by the individuals and organizations for decision making purpose. It is impossible to read and extract user opinions from such huge number of reviews manually. To solve such problem an automated opinion mining approach is required. It is difficult for a user to read and understand all the reviews. Relevant and important information about these establishments should be fetched from reviews and presented to user in a summarized manner.

Current approaches for opinion mining, attempts to detect the polarity of a sentence, paragraph or text regardless of the aspects mentioned in it. In this paper we proposed an aspect based approach for opinion mining which uses aggregate score of opinion words and aspect table together for opinion mining process. The main motive of the system is to develop an opinion mining application with improved accuracy by following an implicit approach.

Keywords—Pre-processing; sentiment; SentiWordNet; aspect; implicit; explicit; tokenization.

I. INTRODUCTION

Everyday many users purchase products, book bus tickets, buy products and services through web. Users also share their views about products, restaurants, news articles etc. on web in the form of comments, reviews, blogs etc. Many users read review information given on the internet to take decisions such as purchasing products, watching movies, going to restaurants etc. Reviews contain users opinion about different types of product. It is difficult for web users to read and comprehend contents from very large number of reviews.

Sentiment analysis refers to the use of natural language processing, text analysis to identify and obtain subjective information in source materials. Sentiment analysis is largely applied to reviews and social media for various kinds of applications, ranging from customer service to marketing. Opinion mining is the process of identifying users opinion about movie, hotel, and product from reviews

[3][5]. Opinion mining involves classification of users expressed opinion into positive or negative polarity [1] [3]. Opinion summarization is a process of representation of review information in short and summarized form. It also involves selecting important aspects and representing related expressed opinions from reviews. There are many applications of opinion mining such as decision making, recommendation systems, feedback analysis etc. It is one of the popular research areas in text mining and natural language processing.

II. ASPECT BASED OPINION MINING

Aspect can be defined as the subject of the review such as the product like “cell phone” or it can be the products attribute like “display”, “sound”. An aspect that appears as noun or noun phrase in a sentence is known as the explicit aspect whereas the aspect which is implied in the sentence is called implicit aspects [6]. Aspect engineering is an extremely basic and essential task for Opinion Mining. Converting a piece of text to an aspect vector is the key step in any data driven approach to Opinion. Without knowing aspects in a corpus, the mined opinions will be of little use.

Majority of the work has been done in finding explicit aspects from online customer reviews. Some had completely ignored implicit aspect identification [11][12]. Hu & Liu had partially addressed implicit aspect identification by applying the same methods used for explicit aspect identification [13].

Mostly reviews are categorized into reviews with explicit aspect and with implicit aspect. Explicit aspects are first identified and extracted as the first step in aspect extraction as they are explicitly specified in the reviews. This would indirectly help in the identification of implicit aspects that are hidden or implied in a review sentence.

Reviews with implicit aspects are unable to convey the noun of the sentence explicitly. The most likely aspect for an opinion word occurring in an implicit sentence can be obtained using the association between opinion phrases and explicit aspects. Implicit aspects can also be identified by clustering explicit features. It is then labeled by a representative term. Co-occurrence matrix can be used

which records the frequent association of the explicit aspect and opinion phrases that occur together. Association rules can be applied to select opinion and aspect pairs that frequently appear together.

III. PROPOSED SYSTEM

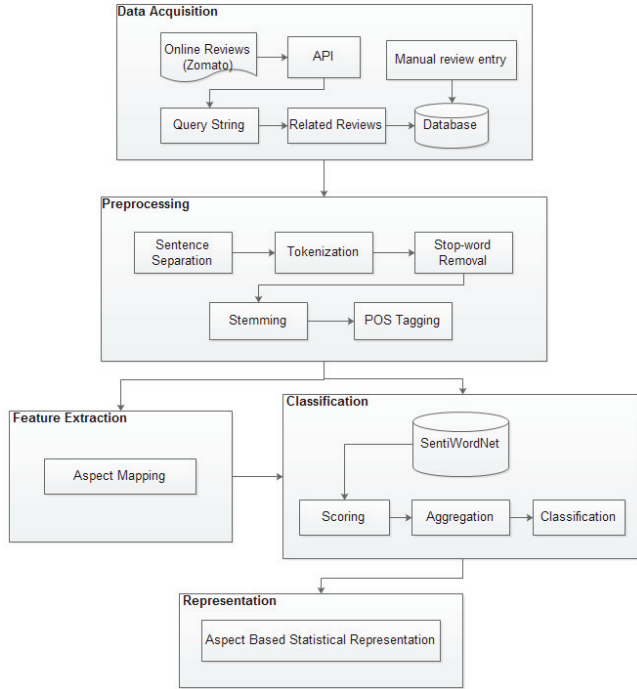


Figure 1: Proposed System Architecture

The proposed system architecture for our system process is given in the Figure 1. The system consists four modules as Data Acquisition, Preprocessing, Feature Extraction, and Classification. We have done experimental study on restaurant dataset. Reviews from Zomato website are captured and preprocessed followed by classification. Also there is provision to enter reviews manually through our web portal. The details about implementation is discussed in section III. The flow of the system is also discussed in Figure 2. The emphasis is on implicit reviews to improve the result.

IV. METHODOLOGY

Figure 1 shows the System Architecture. The system consists of following modules:

A. Data Acquisition

Data acquisition refers to acquiring the user reviews from one or many sources. These sources of reviews should be reliable and sufficient in number. The sources of reviews used are listed below:

a) **Zomato:** Zomato API was used to fetch user reviews in real time. Reviews were fetched for some selected restaurants in the city.

b) **Mannal review entry:** User had the facility to manually enter his/her review at any time in the system. A special textbox was provided on the restaurants page for this purpose.

B. Pre-processing

Preprocessing refers to cleaning of textual data in order to avoid excess processing overhead in further processing. Pre-processing involves various steps such as sentence separation, tokenization, special character removal, stop words removal, stemming, POS tagging etc [2] [10]. OpenNLP libraries and WordNet were used for pre-processing activities.

C. Feature Extraction

Five main aspects were used on which restaurants were supposed to be analyzed. Those were food, service, hygiene, ambiance and price. A set of factual words (synonyms) were maintained for each aspect for e.g. cost, rate were synonyms of price [5] [6] [9]. Opinion to aspect mapping was done by maintaining a set of aspect related words like delicious and tasty mapped to food and dim and dull mapped to ambiance. Whenever these keywords were encountered they were simply mapped to the respective aspect.

D. Classification

The SentiWordNet dictionary is used for scoring the opinion words. Positive and negative scores are assigned to the words. The positive and negative scores are separately aggregated and recorded.

E. Representation

The statistics were represented in a table. JS Charts were used for representing the statistics in form of bar graph, line graph and pie chart for each restaurant.

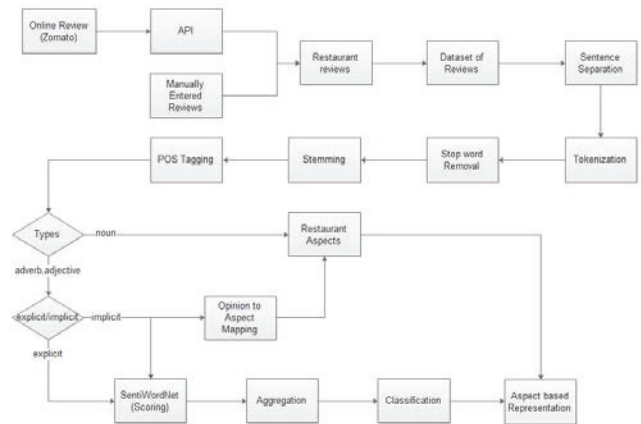


Figure 2: Dataflow Diagram

The reviews are obtained from multiple sources as mentioned above. These reviews go through several pre-processing algorithms like sentence separation, tokenization, stop-word removal, stemming and POS tagging. The nouns are usually the aspects, so we finalized five main aspects such as Food, Staff, Service, Ambiance and Rates. Adverbs and adjectives are usually words expressing opinions. Synonyms and words related to these aspects were stored separately, for e.g. 'delicacies' 'delicious' 'tasty' mapped to food and 'cost' 'price' 'expensive' mapped to rates. The reviews then go through the mapping function which maps the opinion words to the relevant aspect. These reviews then go to SentiWordNet dictionary where positive and negative scores are assigned to each opinion word which in-turn points to some aspect. The positive and negative scores are aggregated separately for each aspect. This summarized result is displayed to the user in form of various graphs and charts.

V. IMPLEMENTATION DETAILS

1. Working Principle:

We considered the user a reviewer or a visitor and administrator as person in charge of adding new restaurants and maintaining database.

a) *User interaction:* User interaction with the system is not a challenging task. The user can search restaurants and view the details of restaurants. User can also post his opinion about the restaurant which he visited earlier.

b) *Administrator interaction:* Administrator plays an important role in the system. It is the responsibility of the administrator for adding new restaurants in the restaurants database. Administrator is also responsible for keeping the information about restaurants up to date. He can do modification in the data if need be. It is the job of the administrator to keep the web portal running and functioning.

2. Assumptions:

- Users have basic knowledge of English language and can write meaningful sentences.
- Users will not use abbreviations like btw, fyi, ,, :-), etc. in the reviews.
- User avoids sarcastic comments in the reviews.
- User does not post fake reviews on the restaurants.

Various Technologies used in implementation of the system are as follows:

- Java, HTML, CSS, JavaScript
- Tomcat
- MySQL
- SentiWordNet

SentiWordNet is one of the popular lexical resources for sentiment analysis. SentiWordNet is a database that contains word with its polarity score for positive or negative sentiment based on its part of speech [7]. By calculating scores of different words in review we get polarity of review text. SentiWordNet provides the sentiment value of the words for different senses like noun, verb, adjective and adverbs. The sentiment values ranges between 0-1.

3. Aspect Table

The aspect table given in Figure 3 and 5 shows all 5 aspects namely Food, Staff, Service, Ambience and Rates along with their aggregated positive and negative scores. The positive scores and negative scores of each aspect are separately aggregated. The formula for the same is as follows:

Aggregate = Sum of the positive scores of a single aspect / Total no of positive scores of that aspect

Similarly for negative scores,

Aggregate = Sum of the negative scores of a single aspect / Total no of negative scores of that aspect

VI. EXPERIMENTAL RESULTS AND OBSERVATIONS

For sentiment scoring and classification we used SentiWordNet dictionary [7]. SentiWordNet is an enhanced lexical resource explicitly devised for supporting sentiment classification and opinion mining applications. This dictionary was used to compare the encountered opinion words and assign positive and negative scores accordingly. The positive and negative scores are aggregated separately for each aspect.

The results for explicit reviews for restaurant Where else café are shown below:

Sr. NO.	Restaurant Id	Product Aspect	+ve	-ve
1	6506018	Food	1.367107239993822	-0.32346328697948257
2	6506018	Staff	0.0	0.0
3	6506018	Service	0.0	0.0
4	6506018	Ambience	0.011210444266583842	0.0
5	6506018	Rates	0.0	0.007736801656600991

Figure 3: Statistical Result for Explicit Reviews



Figure 4: Aspect-wise Ratings for Explicit Reviews

As shown in the Figure 4, aspect wise ratings for explicit reviews are calculated. From the graph it is observed that most of the aspects are showing rating 0, reviews are either implicit or there is no review for the aspect. Except for aspect Food for the tested dataset, other aspects like Staff, Service, Ambience and Rates shows neutral ratings. It becomes difficult to identify overall polarity. So there is need to identify implicit aspects.

The results for implicit reviews for restaurant Where else café are shown below:

Sr. NO.	Restaurant Id	Product Aspect	+ve	-ve
1	6506018	Food	2.361662898146302	-0.3373907850255881
2	6506018	Staff	0.5190783381088225	-0.16420573454259454
3	6506018	Service	1.8988607678908918	-0.05023081293442885
4	6506018	Ambience	1.6575324024931282	0.0
5	6506018	Rates	0.0	0.015473603313201982

Figure 5: Statistical Result for Implicit Reviews

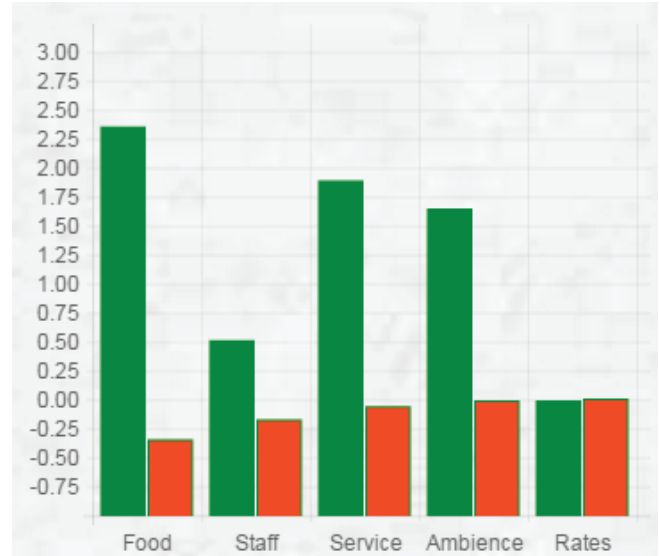


Figure 6: Aspect-wise Ratings for Implicit Reviews

Figure 6 shows ratings for implicit reviews. From the figure it is observed using implicit approach, all the aspects like Food, Staff, Service and Ambience are assessed for the same dataset which is used for explicit approach. It improves the overall accuracy of the existing system.

The following graph shows the overall positive and negative ratings by following the implicit approach:

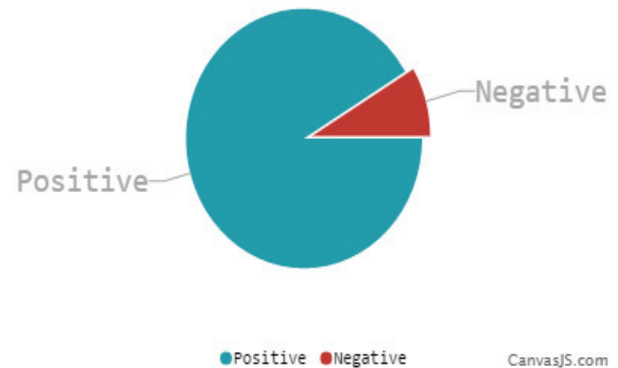


Figure 7: Overall Positive and Negative Ratings

As shown in Figure 7, the positive and negative scores for respective restaurant are aggregated in order to get an overall idea about the quality of restaurant.

VII.CONCLUSION AND FUTURE WORK

An aspect-based opinion mining technique allows us to analyze opinions thoroughly and provides a more convenient way for decision making. This project deals with reviews from restaurants. We have successfully built a system which would take into consideration aspect level reviews by following an implicit approach. We have

successfully extracted user reviews with the use of popular foodie website Zomato by using API. Pre-processing of these extracted reviews is done by using WordNet, Stanford NLP libraries. The accuracy of the system has been increased by following an implicit approach over explicit. For better understanding of the quality of restaurants, we have provided the results to user in statistical manner.

The future work includes increasing the scope of the system so that it could work on other important review datasets like real-estate, stock-market and E-commerce. Also adding features like comparing restaurants and maintaining profile of reviewers would increase the functionality of the system. Treatment of sarcastic and fake reviews would decrease the count of neutral reviews.

VIII. REFERENCES

- [1] Shibily Joseph, "A Syntactic Approach for Aspect Based Opinion", 9th International Conference on Semantic Computing 2015, (IEEE ICSC 2015).
- [2] Satarupa Guha, Aditya Joshi, Vasudeva Varma, "SIEL: Aspect Based Sentiment Analysis in Reviews", 9th International Workshop on Semantic Evaluation, Denver, Colorado, June 4-5, 2015.
- [3] Jingbo Zhu, Member, IEEE, HuizhenWang, Muhua Zhu, Benjamin K. Tsou, Member, IEEE, and Matthew Ma, Senior Member, IEEE, "Aspect-Based Opinion Polling from Customer Reviews", IEEE Transactions on Affective Computing, Vol. 2, No. 1, January-March 2011.
- [4] Ana-Maria Popescu and Oren Etzioni, "Extracting Product Features and Opinions from Reviews",
- [5] Miao Fan and Guoshi Wu, "Aspect Opinion Mining on Customer Reviews", International Conference on ICCE2011, AISC 112, pp. 27-33, Springer-Verlag Berlin Heidelberg 2011.
- [6] Mily Lal, Kavita Asnani, "Implicit Aspects Identification Techniques for Mining Opinions: A Survey", International Journal of Computer Applications (0975 – 8887) Volume 98– No.4, July 2014.
- [7] Andrea Esuli, Fabrizio Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining".
- [8] Ellen Riloff, Ashequl Qadir, "Sarcasm as Contrast between a Positive Sentiment and Negative Situation", Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 704–714, Seattle, Washington, USA, 18-21 October 2013.
- [9] Aditya Bhardwaj, Yogendra Narayan, Vanraj, Pawan, Maitreyee Dutta, "Sentiment Analysis for Indian Stock Market Prediction Using Sensex and Nifty", 4th International Conference on Eco-friendly Computing and Communication Systems, Elsevier 2015.
- [10] Dr. S. Vijayarani, Ms. J. Ilamathi, Ms. Nithya, "Preprocessing Techniques for Text Mining - An Overview", International Journal of Computer Science & Communication Networks Vol 5(1), 7-16.
- [11] Hu, M. and Liu, B. 2004. Mining and summarizing customer reviews. International Conference on Knowledge Discovery and Data Mining (ICDM), KDD'04, August 22–25, 2004, Seattle, Washington, USA. Copyright 2004 ACM.
- [12] Popescu, A. and Etzioni, "Extracting product features and opinions from reviews", In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)2005.
- [13] Liu, B., Hu, M., Cheng, J., "Opinion observer: analyzing and comparing opinions on the web. In WWW International Journal of Computer Applications (0975 – 8887) Volume 98– No.4, July 2014 '05: Proceedings of the 14th international conference on World Wide Web. ACM, New York, NY, USA, 342–351.
- [14] Abd. Samad Hasan Basari, Burairah Hussin, I. Gede Pramudya Ananta, Junta Zeniarja, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization", Procedia Engineering 53 (2013) 453 – 462.