**Major Project Report**

**On**

**Aspect Based Sentiment Analysis**

**In requirements for the complete degree of**

Bachelor of Technology

In

Computer Science and Engineering

**SUBMITTED BY:**

**Anamika(14507)**

**Sheweta(14509)**

**Sahil Singh(14544)**

**Utkarsh Raj(14565)**

**Shantanu Shukla(14566)**

****

**COMPUTER SCIENCE AND ENGINEERING**

**DEPARTMENT OFNATIONAL INSTITUTE TECHNOLOGY**

**HAMIRPUR-177005**

**DECEMBER 2017**

****

**CERTIFICATE**

This is to certify that the project which is being submitted in report titled “Aspect Based Sentiment Analysis” in partial fulfillment of the requirements for the completion of Bachelor of Technology Degree in Computer Science and Engineering and provided to the department at National Institute of Technology, Hamirpur is an genuine proof of our efforts carried under the guidance of Dr. Pardeep Singh, Assistant Professor, Department Computer Science and Engineering, National Institute of Technology, Hamirpur.

The matter submitted in the report has not been proposed by us for the award of any other degree elsewhere.

Anamika (14507)

Sheweta (14509)

SahilRangra(14544)

UtkarshRaj(14565)

Shantanu Shukla(14566)

This is to certify that the above statement made by the candidates is correct to the best of our knowledge.

Dr.Pardeep Singh

Supervisor

CSE Department

**ACKNOWLEDGEMENT**

We consider our self privileged to express gratitude and respect towards all those who guided us through this project. It is with our hearty gratitude that we acknowledge their contributions to this project. We heartily express our sincere gratitude to Dr. Pardeep Singh for his unflinching support and guidance, valuable suggestions and expert advice. His words of wisdom and expertise in subject matter were of immense help throughout the duration of this project.We also take the opportunity to express our deepest thanks to the faculty of Department of Computer Science and Engineering, NIT Hamirpur for helping us by providing necessary knowledge base and resources. Last but not the least we thank our parents and all our mates whose words of encouragement and criticism have molded this report in its present form.

Anamika(14507)

Sheweta(14509)

Sahil Rangra(14544)

Utkarsh Raj(14565)

Shantanu Shukla(14566)

I

**Abstract**

Sentiment Analysis is hot field of exploration under Opinion Mining Domain. It is a fast growing research field nowadays. It refers to the study of different perspectives, approaches, emotions and assessment of people for different products and services provided by different organizations all around the world. With the advent of e-commerce, the demand for opinion mining and sentiment analysis has reached new heights. It is extensively being used for decision making in various organizations and even by individuals. Sentiment analysis can be based on explicit or implicit aspects. Explicit aspects have been studied and analyzed extensively by majority of scientists and researchers. But the field of implicit aspects remain untouched. A very low progress has been made in this field. Our paper tries to cover both explicit and implicit aspect based sentiment analysis.

II

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Table of contents** | |  |
| **Title** |  |  |  | **Page No** |
| *Certificate* | | | |  |
| *Acknowledgement* | | | | I |
| *Abstract* | | | | II |
| *Table of Contents* | | | | III |
| *List of Figures* | | | | V |
| *List of Tables* | | | | VI |
| **Chapter 1: Introduction** | | | | **1** |
| **Chapter 2: Related Work and Background** | | | | **3** |
| **Chapter 3: Phases of the Project** | | | | **4** |
| **Chapter 4: Method** | | | | **6** |
| **4.1. Corpus for Aspect Extraction6** | | | |  |
| **4.2. Pre-Processing** | | | | **6** |
| **4.3. Aspect Term Extraction** | | | | **7** |
| **4.4. Tasks to be accomplished** | | | | **9** |
| 4.4.1. | Aspect Polarity Detection | | | 9 |
| 4.4.2. | Aspect Categorization | | | 10 |

4.4.3. Sentiment Calculation 10

|  |  |
| --- | --- |
| **Chapter 5: Evaluation Measures and Result** | **11** |
| **Chapter 6: Conclusion and Future Work** | **13** |
| **REFERENCES** | **14** |

IV

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **List of Figures** |  |  |
| **Figure** |  | **Title** | | **Page no.** |
| Fig. 3.1 | Aspect Word Extraction | | | 4 |
| Fig. 4.2 | Aspect Sentence Extraction | | | 4 |
| Fig. 5.3 | Dependency Parse Tree | | | 4 |
| Fig. 7.4 | Dependency Tree | | | 4 |
| Fig. 2.1 | Dependency Tree | | | 6 |
| Fig. 9.2 | Dependency Tree | | | 7 |
| Fig. 2.3 | Aspect Extractor Code | | | 9 |

V

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **List of Tables** | |  |
|  |  |  |  |  |
| **Table** | **Title** | | | **Page no.** |
| Table. 5.1 | Contingency Matrix | | | 12 |

VI

**CHAPTER -1**

**INTRODUCTION**

Sentiment analysis has widely approached the world as the hard hitting reform in commercial as well academic standpoints. Although most of the approaches used these days to use the words in the sentence to the polarity of the sentence, all the paragraphs and also the entire article. The aspect used to judge the sentiment can be as follows – battery, screen, food, service. The task associated with ABSA, the whole goal is to get the aspects from the present entities, use the sentiments related to those aspects.

The project we have worked upon contains three main sub modules namely: extraction of words, collection of aspect terms, and a sentiment polarity calculation. The first module tries to find out single- and multi-word terms naming aspects of the item held under discussion. We have named these terms as aspect terms . The second sub module tries to group similar aspect terms (e.g., ‘price’, ‘food’) , depending on the user behavior and other rules and regulations (e.g., the screen size on which ABSA system will be shown). The third sub module of our project estimates the aggregate value of the sentiments per aspect term or group of aspect terms.

The datasets were collected differently from various sources for all of the above mentioned sub modules. ASBA can be of two types: implicit or explicit. Aspects which are explicit also known as Explicit Aspect are the ones that indicate the target value in the line held under discussion or being operated upon. The indirect expression of an aspect is called implicit aspect clue(IAC) for e.g,. about the aspects: appearance and price of the entity: camera.

This type of aspect extraction (Explicit Aspects) is being studied extensively and there are multiple ways proposed for this. But the field of implicit aspects is more or less untouched. A very low progress has been made in this field. The task of implicit aspect extraction is considered to be a little tough but very important as the phenomenon of implicit aspects happens to be discussed in every document under operation.

For example, the following excerpt has been adopted from (Hu and Liu, 2004) , it uses only implicit aspect: This phone has got the best ever quality one could ever get. The phone has all the features one would require in a mobile: It has got a weight of a feather i.e. light weight, it is attractive and neat. I found quite manageable and easy to change; very convenient to navigate. Here, the word “light weight” is weighable quality of the phone; the words “neat” and “attractive” refers to the looks; the compound

“manageable” to the user interface;

1

The penultimate Phrase “easy to change” is referring to its functionality, which can be used in terms of the phrase for mobile and its aspects as calculated by the evaluation based on the discussed criteria.

2

**CHAPTER -2**

**ASSOCIATED WORKS**

Aspect extraction from reviewed and talked upon text was initiated by Hu and Liu (Hu and Liu, 2004), who are responsible for pioneering research in this field with classifications on implicit and explicit aspects. However it was only based on preliminary research based on statistical analysis. Popescu and Etzioni (Popescu and Etzioni, 2005) and Blair-Goldensohn (Blair-Goldensohn et al., 2008) later improved on the work done by Hu and Liu. The assumption of Popescu and Etzioni is based on the product class being known a priori. The method proposed by them first detects on the noun or noun phrase and check if that is a feature of the product. Scaffidi et al. (Scaffidi et al., 2007) proposed a method that uses a language model to identify features of the products. Their method was based on the assumption that the features of the products are talked about in product reviews with more frequency than in text in general language. Method proposed however seemed to be imprecise as the aspects received were filled with noise.

Extraction of aspects can be seen as a problem of extraction from general language, where we generally used the process of sequential labels. Hidden Markov Models(HMM) and Conditional Random Fields (CRF) (Lafferty et al., 2001). Popescu and Etzioni (Popescu and Etzioni, 2005) first grasped on such type of extraction which increases the accuracy of polarity detection. A clustered method was proposed by Su (Su et al., 2008) in order to map the aspects meaning associated to their corresponding aspects terms.

3

**CHAPTER-3**

**PHASES OF THE PROJECT**

**Phase1: Data Collection and Aspect term extraction**

The dataset which was used for the explicit and implicit aspect extraction was SemEval 2016 database[7]. Aspect term extraction is the phase where we extracted the Aspect terms from the reviews. There is a given group of sentences having prior known entities e.g., hotels, then we try to find the aspect terms. Then we output a list of all the different aspect terms which provides a name for a specific target object.

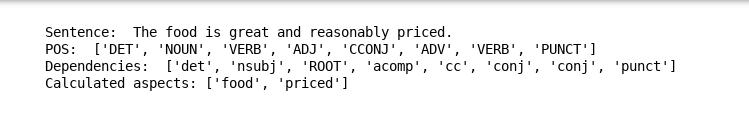


Fig. 3.1 Aspect Term Extraction

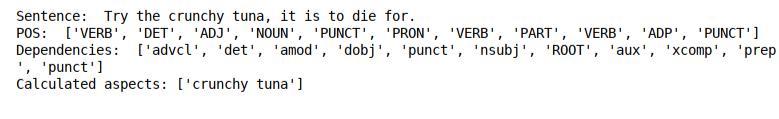


Fig. 3.2 Aspect Term Extraction

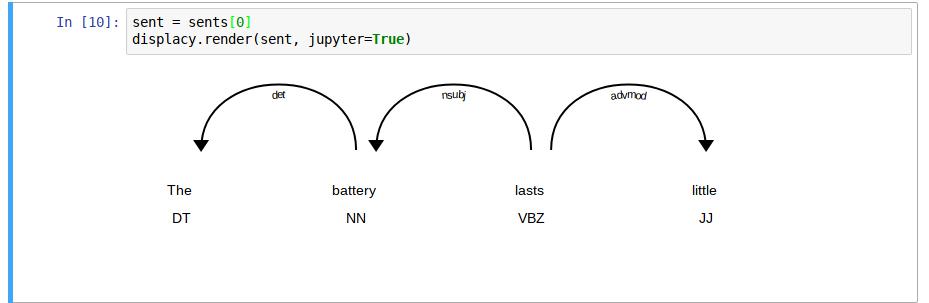


Fig. 3.3 Dependency Tree

4

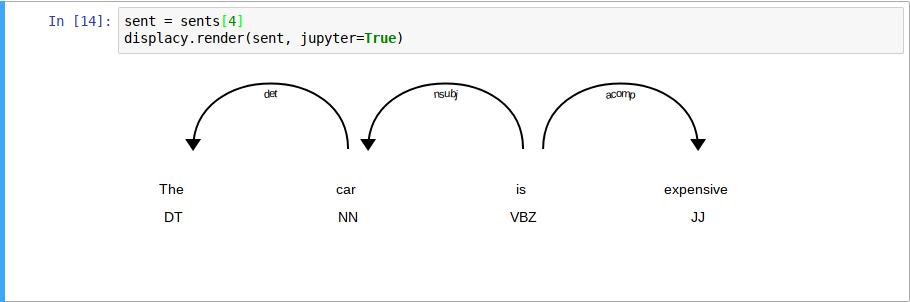


Fig. 3.4 Dependency Tree

**Phase 2: Aspect Polarity Detection**

Aspect Polarity Detection involves classification of all aspect terms as the polarity – negative, neutral, conflict or positive.

**Phase 3: Aspect Categorization**

We have set of aspect categories, e.g., food, price. We classify the found aspects into one of these categories.

**Phase 4: Sentiment Calculation**

Given the aspect categories, we find the dichotomy - negative, neutral and positive or either conflict of them.

5

**CHAPTER -4**

**METHOD**

**4.1 Database for Aspect Extraction**

To start our sentiment analysis process, the corpus or the dataset that is used is SemEval Dataset 2014[6].

**4.2 Preprocessing**

Aspect Extraction is totally dependant upon the pre-processing. Pre- Processing includes majorly four steps . Firstly, the data corpus received is converted to a huge list of sentences i.e. segmentation is performed. Secondly, these sentences are tokenized using Spacy tokenizer [7]. The third step , includes part of speech tagging (POS) for each sentence formed in the list using Spacy Part of Speech Tagger[8]. A dependency tree for each line in the database using Spacy Dependency Parser [9] is created.

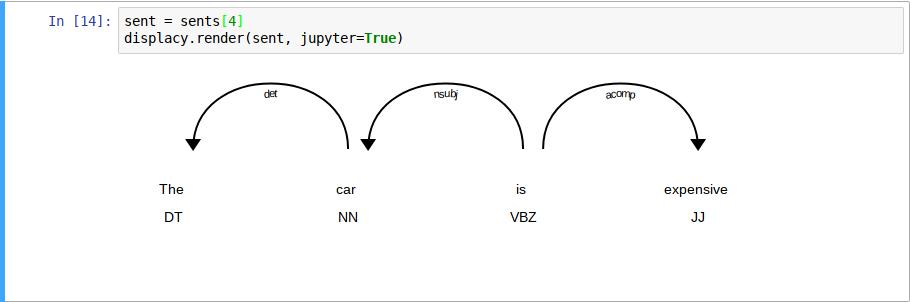


Fig. 5.1 Dependency Tree

6

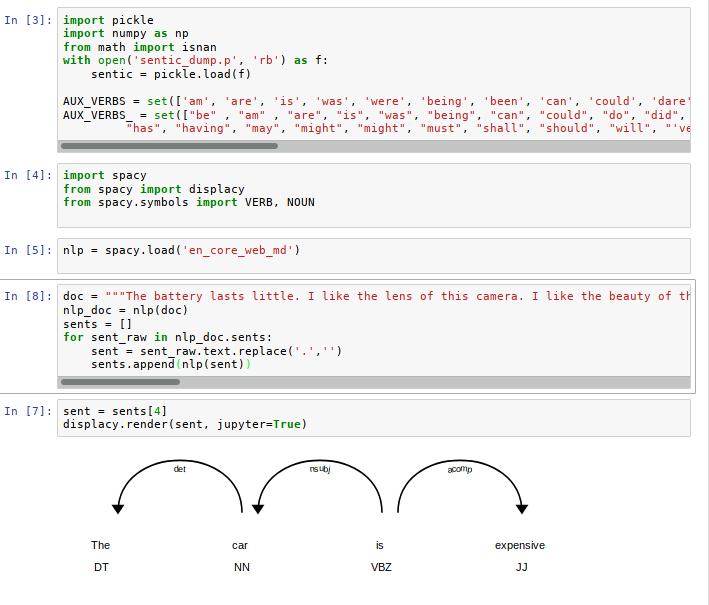


Fig. 5.2 Dependency Tree

**4.3 Aspect Term Extraction**

Some rules have been devised to extract the aspect terms from the corpus in the previous works done in this field[7]. We have used the same rule based approach which formulates the widely accepted rules as follows:

Implementation: Our system is based on 2 general rules:

1) For a sentence where subject verb is present.

2 For sentence where a subject verb is not present.

Rules: For a sentence where subject verb is present.

7

1. if x is having any modifier that exists in database, then x is taken out.
2. If no secondary verb in line and x modified by adjective or adverb then both h & x are aspects.
3. x is having a direct object associationship with noun n,
   1. if n not present in database then n is aspect term.
   2. else n is in senticnet then n is associated with n1(Noun) and n1 is aspect term.
4. x is related to t1 with open association and x-x1 should exist in opinion lexicon.

a) Then x-x1 is extracted as aspect.

b) If n (Noun) connected to x1 then n is aspect term

1. If x is associated with copular then x is an aspect.
2. if noun, h then h is an aspect.

For sentence where a subject verb is not present.

1. h, adverb is with clausal association with x,
   * + 1. then h is extracted as adverb
2. h is associated with a x which is a noun – a prepositional association,
   * 1. h, x are both aspects
3. If h is in direct association with x,
   1. then x will be aspect term

We can extract aspect terms by using given set of rules from reviews. The result from this task is a list of aspect terms which we have stored in a dictionary which has sentence-id as key.

8



Fig. 5.3 Aspect Extractor Code

**4.4 Tasks to be accomplished**

**4.4.1 Aspect Polarity Detection**

Now for the next phase of our project , polarity of the terms which are aspects – negative, positive, conflict, neutral.

9

**4.4.2. Aspect Categorization**

We now know the predefined set of aspect categories, so now we have to identify the categories of the aspects provide we are given each aspect terms.

There were five categories that were used for the aspect terms of the data we have used. These five for the data of restaurants are, [ food, service, ambience, price, miscellaneous ]

**4.4.3 Sentiment Calculation**

We now , in this step, will calculate the polarity of aspects present in the aspect categories. The steps are as follows: We will use thee calculated polarities of the aspect terms from the previous steps, and categories from categorization step, and calculate the polarities of each category.

10

**CHAPTER - 5**

**EVALUATIONS MEASURES & RESULT**

Method uses set G with human annotators with the texts where the true positives is found, false positives, false negatives and the precision, recall are defined as usually.

Accuracy defined as follows:

Aspect terms TP is given both by evaluated terms and the human annotators. Some are given by human annotators but not calculated by evaluation. The contrary is also taken into account for the results.

From the table we can see:

= *80*

= *782*

= *394*

= *16*

Based on that we can calculate other values:

= /( + ) = 0.17

= /( + ) = 0.83

= ( + )/( + + + ) = 0.677

11

|  |  |  |
| --- | --- | --- |
| Total terms(1176) | Real aspect term | Real non-aspect term |
|  |  |  |
| Predicted aspect term | 80 | 394 |
|  |  |  |
| Predicted non-aspect term | 16 | 782 |
|  |  |  |

Table 5.1 Contingency Matrix

12

**CHAPTER – 6**

**CONCLUSION AND FUTURE WORK**

The project tries to labour upon the depths of Aspect Based Sentiment Analysis. The work has been explored only in the range of explicit aspects but in our project we have tried to explore the implicit aspects as well. We have inferred from our journey of process that implicit aspect based sentiment analysis is quite tough but important. As a very less work has been done in this field, this field needs to be explored to quite an extent. Though our project has done well for explicit aspects, there is wide scope of improvement for implicit ones. In his project we tried extracting and summarizing restaurant reviews based natural language processing methods. The objective was to perform aspect extraction and segmentation. Without knowing the aspects, the opinions are useless for operation under opinion mining. For aspect and opinion extraction, we have used Stanford coreNLP POS tagger. Lemmatization is used for aspect reduction. The project is divided into four phases and the opinion-aspect pairs thus generated as a result of these processes are used for identifying implicit aspects.

Although the methodology used here seems to be quite reasonable in identifying implicit and explicit aspects but some undesirable errors still exist. Some of the errors are caused by incorrect identification of explicit aspects and opinion words. The errors have affected the implicit aspect based extraction quite difficult as compared to explicit one. Thus, the field of implicit aspects is still to be put under a huge knife for exploitation.

13

**REFERENCES**

1. Minqing Hu and Bing Liu,Mining Opinion Features in Customer Reviews.International Conference on Knowledge Discovery and Data Mining (ICDM).
2. Sasha Blair-Goldensohn, Kerry Hannan, Ryan McDonald, Tyler Neylon, George A. Reis, and Jeff Reynar. 2008. Building a sentiment summarizer for local service reviews. In Proceedings of WWW-2008 workshop on NLP in the Information Explosion Era, page 14.
3. Dave, K., Lawrence, S., and Pennock, D., 2003. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. WWW-2003.
4. Bing Liu. 2012. Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers. [5]SoujanyaPoria, Erik Cambria, Lun-Wei Ku, Chen Gui. A Rule-Based Approach to Aspect Extraction from Product Reviews

[6]SemEval 2014 testing dataset <http://metashare.ilsp.gr:8080/repository/browse/semeval-2014-absa-test-data-phase-a/3a238694b10a11e382fe842b2b6a04d73fe269392a55406b90b5315f33a99abd/>

1. <https://spacy.io/usage/linguistic-features#section-tokenization>
2. <https://spacy.io/usage/linguistic-features#section-pos-tagging> [9]<https://spacy.io/usage/linguistic-features#section-dependency-parse>

14