

**1.ABSTRACT**

With the increase in popularity of online review sites comes a corresponding need for tools capable of extracting the information most important to the user from the plain text data.

Sentiment Analysis is a widely addressed Natural Language Processing task wherein the semantic orientation of a text unit is adjudged. However, a major challenge in Sentiment Analysis is the identiﬁcation of entities towards which the opinion is expressed. Aspect based Sentiment Analysis consists of two tasks. The ﬁrst part involves the extraction of the aspect term from a sentence and secondly the polarity of the opinion corresponding to that aspect is adjudged.

This project is the implementation for Unsupervised and Supervised Methods for Aspect-Based Sentiment Analysis [1] of restaurant reviews dataset. Some Unsupervised and supervised methods are proposed, implemented and evaluated. We focus on determining the aspect terms existing in each sentence, finding out their polarities, detecting the categories of the sentence and the polarity of each category.

**2. INTRODUCTION**

**2.1  General Introduction to the Topic**

Sentiment analysis is increasingly viewed as a vital task both from an academic and a commercial standpoint. The majority of current approaches, however, attempt to detect the overall polarity of a sentence, paragraph, or text span, regardless of the entities mentioned (e.g. restaurants) and their aspects (e.g. food, service).

By contrast, this task is concerned with aspect based sentiment analysis (ABSA), where the goal is to identify the aspects of given target entities and the sentiment expressed towards each aspect.

Eg. The food was good, but the ambiance was poor.

Here, we have 2 aspects “food”, and “ambiance”. Also, polarities associated with both

aspects are “positive”, and “negative” respectively.

We will use 2 labelled datasets for this problem, and will find aspects, corresponding

polarities, and visualize them.

**2.2  Organization**

**Impetus**is a software solutions, products and services company headquartered in Los Gatos, USA with development centers in NOIDA, Indore, Gurgaon, and Bengaluru, India.

With more than 1600 employees globally, Impetus is focused on creating new ways of analyzing data for businesses—helping them gain key business insights across the enterprise.

 They bring together a unique mix of Data Science capabilities and technology expertise across the Big Data ecosystem including Hadoop, NoSQL, NewSQL, MPP databases, machine learning, and innovative visualization.

**2.3  Area of Computer Science**

Aspect Based Sentiment Analysis (ABSA) falls under the Machine Learning Branch of Computer Science Applications.

Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome. Machine learning is so pervasive today that you probably use it dozens of times a day without knowing it.

Many researchers also think it is the best way to make progress towards human-level AI. In this class, you will learn about the most effective machine learning techniques, and gain practice implementing them and getting them to work for yourself.

**2.4  Hardware and Software Requirements**

* R studio
* TensorFlow
* CoreNLP

**3. PROBLEM DEFINITION**

2.1 Dataset

The data set provided by SemEval is a subset of Ganu et at(2009)[4]. It is in the XML format, and has separate ﬁles for Laptop and Restaurant reviews. The training data contains about 500 reviews for each domains, which is 1386 sentences for Restaurants and 1403 sentences for Laptops. The image attached below shows a part the XML ﬁle of the Restaurant data.



Fig.1: Snipptet of the Restaurant dataset XML ﬁle

For reach sentence, we have a target attribute which lists the aspect term, and a corresponding polarity attribute. The distribution of positive and negative sentiments in both the dataset is provided below:

|  |  |  |
| --- | --- | --- |
| Domain | Positive | Negative |
| Restaurant | 1198 | 408 |

**4.  OBJECTIVE(S)**

Datasets consisting of customer reviews with human-authored annotations identifying the mentioned aspects of the target entities and the sentiment polarity of each aspect are given.

The goal is to identify the aspects of given target entities and the sentiment expressed towards each aspect. Eg. The food was good, but the ambiance was poor.

Here, we have 2 aspects “food”, and “ambiance”. Also, polarities associated with both

aspects are “positive”, and “negative” respectively.

We will use 2 labelled datasets for this problem, and will find aspects, corresponding

polarities, and visualise them.

5.**BACKGROUND**

Various approaches have been adopted to identify aspects from sentences. Bing Lui et al. used frequency of noun phrases, followed by a redundancy pruning to identity the feature corresponding to a review[6].Yejin Choi et al. performed semantic tagging using conditional random ﬁelds with features based on Capitalization, syntactic chunking to extract sources of opinions from texts[3].

The best performing one uses a Conditional Random Field with features extracted using named entity recognition, POS tagging and parsing.

We try to augment this approach by using features not only based upon text processing techniques, but also on vector embeddings of words and sentences. The motivation behind this being that the number of candidate aspect words of restaurant domain is rather restrictive. The task of polarity detection was addressed using various classiﬁcation techniques like Naive Bayes, SVM etc.

**6. METHODOLOGY**

In particular, the task consists of the following subtasks:

**6.1 Aspect Term Extraction**

Given a set of sentences with pre-identified entities (e.g., restaurants), identify the aspect terms present in the sentence and return a list containing all the distinct aspect terms. An aspect term names a particular aspect of the target entity.

For example, "I liked the service and the staff, but not the food”, “The food was nothing much, but I loved the staff”. Multi-word aspect terms (e.g., “hard disk”) should be treated as single terms (e.g., in “The hard disk is very noisy” the only aspect term is “hard disk”).

**6.2 Aspect Term Polarity**

For a given set of aspect terms within a sentence, determine whether the polarity of each aspect term is positive, negative, neutral or conflict (i.e., both positive and negative).

For example:

“I loved their fajitas” → {fajitas: positive}

“I hated their fajitas, but their salads were great” → {fajitas: negative, salads: positive}

“The fajitas are their first plate” → {fajitas: neutral}

“The fajitas were great to taste, but not to see” → {fajitas: conflict}

**6.3 Aspect Category Detection**

Given a predefined set of aspect categories (e.g., price, food), identify the aspect categories discussed in a given sentence. Aspect categories are typically coarser than the aspect terms of Subtask 1, and they do not necessarily occur as terms in the given sentence.

For example, given the set of aspect categories {food, service, price, ambience, anecdotes/miscellaneous}:

“The restaurant was too expensive”  → {price}

“The restaurant was expensive, but the menu was great” → {price, food}

**6.4 Aspect Category Polarity**

Given a set of pre-identified aspect categories (e.g., {food, price}), determine the polarity (positive, negative, neutral or conflict) of each aspect category.

For example:

“The restaurant was too expensive” → {price: negative}

“The restaurant was expensive, but the menu was great” → {price: negative, food: positive}

**7. IMPLEMENTATION DETAILS**

The source code can be found at github : <https://github.com/shashwat1995> .

**8. PROJECT SCHEDULE**

* *January 2018*
* Study Part of speech Tagging in NLP
* *February 2018*
* Implementation of unsupervised ABSA NLP program
* *March 2018*
* Implementation of supervised ABSA NLP program
* Documentation
* *April 2018*
* Submission of report & evaluation
* Study DeepLearning Technology and begin implementation
* *May 2018*
* ABSA implementation using a DeepLearning Approach ( TensorFlow)
* *June 2018*
* Submission of report & evaluation  
    
     
    
     
    
     
    
    
    
    
    
    
  **9.REFERENCES**
* G. Ganu, N. Elhadad, and A. Marian, “Beyond the stars: Improving rating predictions using review text content”. Proceedings of the 12th International Workshop on the Web and Databases, Providence, Rhode Island, 2009.
* M. Hu and B. Liu, “Mining and summarizing customer reviews”. Proceedings of the 10th KDD, pp. 168–177, Seattle, WA, 2004.
* S.-M. Kim and E. Hovy, “Extracting opinions, opinion holders, and topics expressed in online news media text”. Proceedings of the Workshop on Sentiment and Subjectivity in Text, pp. 1– 8, Sydney, Australia, 2006.
* B. Liu, Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. Morgan & Claypool, 2012.
* S. Moghaddam and M. Ester, “Opinion digger: an unsupervised opinion miner from unstructured product reviews”. Proceedings of the 19th CIKM, pp. 1825–1828, Toronto, ON, 2010.
* M. Tsytsarau and T. Palpanas. “Survey on mining subjective data on the web”. Data Mining and Knowledge Discovery, 24(3):478–514, 2012.
* Z. Zhai, B. Liu, H. Xu, and P. Jia. “Clustering product features for opinion mining”.  Proceedings of the 4th International Conference of WSDM, pp. 347–354, Hong Kong, 2011.
* S. Brody and N. Elhadad. “An unsupervised aspect-sentiment model for online reviews”. Proceedings of NAACL, pages 804–812, Los Angeles, CA, 2010.  
    
    
    
    
    
    
    
    
    
     
    
     
    
     
    
    
  PROJECT DETAILS

|  |  |  |  |
| --- | --- | --- | --- |
| *Student Details* | | | |
| **Student Name** | **SHASHWAT GUPTA** | | |
| Register Number | 140905047 | Section / Roll No | C - 06 |
| Email Address | [Shashwat2211@gmail.com](mailto:Shashwat2211@gmail.com) | Phone No (M) | 9620193750 |
|  | | | |
| *Project Details* | | | |
| **Project Title** | Aspect based Sentiment Analysis using NLP (R) and Deep Learning | | |
| Project Duration | 5 Months | Date of reporting | 22 - Jan - 2018 |
|  |  | | |
| *Organization Details* | | | |
| **Organization Name** | **Impetus Infotech India Pvt. Ltd.** | | |
| Full postal address with pin code | SDF No. K-13 to 16, NSEZ, Phase-II, NOIDA - 201305 (U.P.) India | | |
| Website address | Impetus.com | | |
|  |  | | |
| *External Guide Details* | | | |
| **Name of the Guide** | **Mr. Vivek Gupta** | | |
| Designation | Senior Analytics Engineer | | |
| Full contact address with pin code | Data Science Practice , SDF No. K-13 to 16, NSEZ, Phase-II, NOIDA - 201305 (U.P.) India | | |
| Email address | [Vivek.gupta@impetus.co.in](mailto:Vivek.gupta@impetus.co.in) | Phone No (M) | 9910464101 |
|  |  | | |
| *Internal Guide Details* | | | |
| **Faculty Name** | **Mrs. Archana Praveen Kumar** | | |
| Full contact address with pin code | Dept of Computer Science & Engg, Manipal Institute of Technology, Manipal – 576 104 (Karnataka State), INDIA | | |
| Email address | [Archana.kumar@manipal.edu](mailto:Archana.kumar@manipal.edu) | | |