

RestoQ – Aspect Based Sentiment Analysis

Ashish Chandra Jha ^{1,†}
REVA University
Bangalore, India
Ashish.ba05@reva.edu.in

Lalit Agarwal
REVA University
Bangalore, India
Lalit.ba05@reva.edu.in

Sanjeev Kumar Jha ^{2,†}
REVA University
Bangalore, India
Sanjeev.ba05@reva.edu.in

[†]A.C.J. and S.K.J. contributed equally to this work.

Abstract – People love experimenting on their food with different tastes. And when it's about visiting the restaurant or ordering food online, they will definitely look for the reviews which will talk about the aspects like services, ambience and cost along with food quality. This food problem is not a single day problem, it's getting repeated everyday. Some end up with positive reviews and few end up with negative or neutral reviews. In this work a framework is developed called 'RestoQ', which uses text analytics for sentiment analysis at the aspect level to discover and rank the restaurants. The framework analyzes the reviews for the sentiments across four aspects – price, food quality, service quality and ambience. Unsupervised lexicon-based classifier and a naïve Bayesian classifier are used to evaluate and score the sentiments at aspect level. The final score will be a combined sum of each score for the review, which requires further work rank the aspects based on reviews. Surprisingly unsupervised method out performs the supervised method. It is proposed to extend the work with context based methods using word2vect and LSTM.

I. INTRODUCTION

Before ordering food or booking a table in any restaurant consumers generally check the reviews of the places. Online food ordering sites like Zomato, Food Panda and UberEats do sentiment analysis of the reviews given by the customers and give a rating for these restaurants. The majority of current sentiment analysis approaches try to detect the overall polarity of the reviews or sentence regardless of the target entities (e.g. restaurants) and their aspects (e.g. services, ambience and cost along with food quality). Aspect Based Sentiment Analysis is fine grained sentiment analysis. A sentence may contain multiple opinions about different entities and we need to find each of them. This has to be analysed by model and should give insights. In this work, the research findings of such a system are presented.

II. LITERATURE SURVEY

Sentiment analysis is one of the fastest growing research areas in computer science, making it challenging to keep track of all the activities in the area. It is a case of natural language processing which could mark the emotion or mood of the people about any specific product by analysis. It is a process of automatic extraction of features by mode of notions of others about specific product, services or experience. [1]

Customers as well as ecommerce companies (online food ordering in this case) are looking for the reviews of the restaurants to order food or to check their customer satisfaction ratio. A lot of research has been done on Sentiment analysis on restaurants and their reviews. Reviews are considered to be positive, negative or neutral on the overall score of the sentence. To some extent it is very useful and many customers are using it before ordering their food on daily basis [2][3].

Unlike document level sentiment classification task, aspect based sentiment analysis is a more fine-grained classification task. It aims at identifying the sentiment polarity (e.g. positive, negative and neutral) of one specific

aspect in its context sentence. For example, given a sentence "great food but the service was dreadful" the sentiment polarity for aspects "food" and "service" are positive and negative respectively [4].

Aspect Based Sentiment Analysis (ABSA) was introduced as a shared task for the first time in the context of SemEval in 2014; SemEval2014 Task 41 (SE-ABSA14) provided datasets of English reviews annotated at the sentence level with aspect terms (e.g., "mouse", "pizza") and their polarity for the laptop and restaurant domains, as well as coarser aspect categories (e.g., "food") and their polarity only for restaurants (Pontiki et al., 2014). SemEval-2015 Task 122 (SE-ABSA15) built upon SE-ABSA14 and consolidated its subtasks into a unified framework in which all the identified constituents of the expressed opinions (i.e., aspects, opinion target expressions and sentiment polarities) meet a set of guidelines and are linked to each other within sentence-level tuples (Pontiki et al., 2015) [5][6][7].

Aspect Based Sentiment Analysis poses several challenges in processing text data, is a popular area of research in this direction. Several challenges which has not been addressed and people are trying to do some research are implicit aspect detection, mapping aspect words to categories, resolving anaphora references etc. Researchers combine techniques from common sense rules, unsupervised supervised and semi supervised techniques to perform these tasks.

Aspect Based Sentiment Analysis has been done for this particular topic by various researchers [8] [9] [10]. In this paper, the four aspects depending on which the comments will be reviewed. It has been seen that there are lots of aspects which affects the overall sentiment of the review. For example, in restaurants, people give review based on food quality, services, ambience and price. In this work restaurants will categorized based on the customer reviews. The goal is to determine the sentiment expressed toward each aspect on restaurant of Bangalore in English language.

The problem of aspect-based sentiment analysis deals with classifying sentiments (negative, neutral, positive) for a given aspect in a sentence. A traditional sentiment classification task involves treating the entire sentence as a text document and classifying sentiments based on all the words [11].

Labeling of data is a little difficult task to perform automatically. Most of the researcher who are working on new dataset used to label the data manually. The lack of labeled data has led to several researchers to explore unsupervised learning techniques to learn both aspects and their sentiments expressed in plain text. Particularly the fact that aspects are normally described by opinion words and opinion words in turn will have a target aspect can be used to iteratively expand the sentiment and aspect lexicon. The expansion is done with the help of rules to associate aspects and sentiment [12][13].

In this paper, we are trying to do aspect based sentiment analysis on restaurant reviews data from an online food delivery site (Zomato). We have introduced a system based on Text Analytics on the reviews using Supervised Machine Learning with a Naïve Bayes algorithm and unsupervised Machine Learning with Lexicon based algorithm for scoring sentiments

III. RESEARCH METHODOLOGY

In this section, we will explore the different techniques, methods, and features used in this experiment. We will divide the section into two sections: data exploration and pre-processing and model building. Model building is further divided into supervised ML and Unsupervised ML.

Data access: 2000 restaurant across Bangalore along with their reviews has been collected. The data is from an online food ordering company i.e. Zomato. Labeling of data is the hard part of any new research and is done manually. Here the reviews have been labeled based on restaurants name, aspects and sentiment. Positive, negative and neutral sentiments have been used as the three classes.

Data Exploration: Data has the solution to every problem. But one must know how to use that data. Data exploration gives the ability to summarize the main characteristics of a data set, including its size, accuracy, initial patterns, null values, outlier values and missing values. It can use a combination of manual methods and automated tools such as data visualizations, charts, and initial reports to explore the data.

Data Preprocessing: It is the most vital part of any analysis. Considering few important preprocessing steps, below mentioned techniques have been used

- **Stopwords Removal** - Stopwords are the meaningless and repeated words which do not contribute to the semantic of the statement. It should be removed.
- **Symbol Removal** – Reviews generally contain symbols like @, #, \$ with no contribution towards analyzing the sentiments. So, it should be removed.
- **Contractions and Annotation Removal** – The contractions and annotation like shouldn't should be removed with 'should not'
- **Normalization** - Normalization stands for making the word or sentence case insensitive. Data should be normalized.
- **Exploration** – It is to check the word frequency of the corpus. It gives the idea of what the document is about. We check the word frequency by TF-IDF model. Words with high frequency can be seen using word cloud. In addition conditional exploration, based on sentiments and aspects word cloud has been made.
- For sentiments three word cloud, one for positive, one negative and one for neutral emotions. For aspect four word cloud has been made based on food quality, services, ambience and cost.

Model Building: The model will be created four times with different strategies. Here combination of 90%-10%, 80%-20%, 70%-30% and 60%-40% train/test split along with 10 fold cross validation has been used.

Starting with the supervised learning model then tried unsupervised learning model has been created to compare which algorithm will perform better.

In Supervised Machine Learning stage, label data is used for building classifier using Naive Bayes algorithm. Naive Bayesian algorithm is a probabilistic ML algorithm, which assumes independence among the features.

In Unsupervised Machine Learning stage, label data is used for building classifier using Lexicon based algorithm. The lexicon based approach is based on the assumption that the contextual sentiment orientation is the sum of the sentiment orientation of each word or phrase.

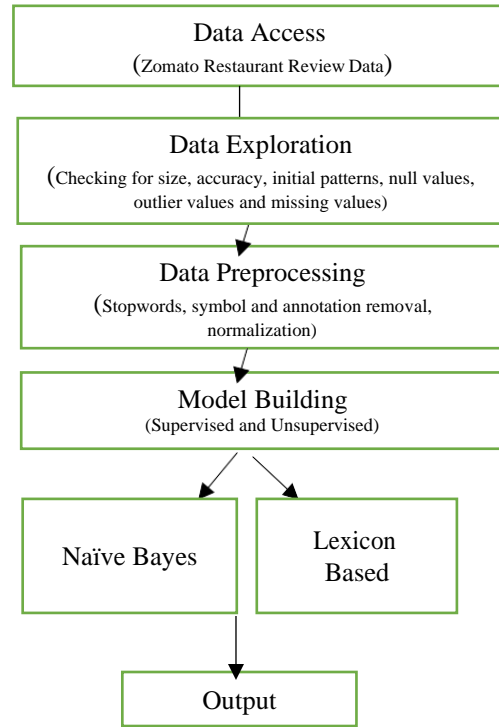


FIG.1: PROPOSED APPROACH

IV. RESULTS AND DISCUSSION

Supervised (Lexicon based model) and Unsupervised (Naive Bayes model) has been created at different train test split. For each instant, accuracy has been captured and report is mentioned below:

Table 1. Experimental results at review level

Training	Test	Accuracy	
		Lexicon	Naive Bayes
90%	10%	71%	72%
80%	20%	68%	65%
70%	30%	70%	65%
60%	40%	68%	63%

The classification accuracy of all the models are consistent with the results published in literature and hence support the methodology used in this research. From the result, it clearly shows, the unsupervised learning lexicon-based model performs better than supervised learning technique using Naive Bayes. Accuracy at different train test split given an idea of optimum split scoring highest accuracy.

V. CONCLUSIONS

This paper covers the Aspect Based Sentiment Analysis on restaurants reviews dataset for Bangalore restaurants. The ABSA task consists of four aspects namely food quality, services, ambience and cost. For each aspect, sentiments have been analyzed. Supervised and Unsupervised machine learning has been used.

The proposed approaches achieved very good results. The algorithm successfully able to analyze the aspects of the sentiments. Further the restaurants are ranked based on the over all score and the positive score, which can be used by consumers for selection of restaurants. It is proposed to carry out a context-based analysis of the sentiments using word2vec and LSTM to test the improvements in the accuracies.

REFERENCES

- [1] "A Study on Sentiment Analysis: Methods and Tools by Abhishek Kaushik1, Anchal Kaushik2, Sudhanshu Naithani3 <https://pdfs.semanticscholar.org/c151/dfad8c1bf88b0afc716758c77d533ded7dd0.pdf>"
- [2] "Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews <https://arxiv.org/ftp/arxiv/papers/1709/1709.08698.pdf>"
- [3] "SENTIMENT ANALYSIS OF RESTAURANT REVIEWS USING HYBRID CLASSIFICATION METHOD

- http://iraj.in/journal/journal_file/journal_pdf/4-54-140014488817-23.pdf
- [4] "Sethia, A., & Bhattacharyya, P. Aspect Based Sentiment Analysis-A Survey. Accessed on August 2019; http://www.cfitl.iitb.ac.in/resources/surveys/aspect-based-sentiment-analysis_survey.pdf"
- [5] "Bhoi, A., & Joshi, S. (2018). Various Approaches to Aspect-based Sentiment Analysis. arXiv preprint arXiv:1805.01984"
- [6] "Brychcin, T., Konkol, M., & Steinberger, J. (2014, August). Uwb: Machine learning approach to aspect-based sentiment analysis. In Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014) (pp. 817-822) <http://alt.qcri.org/semeval2014/cdrom/pdf/SemEval2014145.pdf>
- [7] "Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks Binxuan Huang, Yanglan Ou and Kathleen M. Carley <https://arxiv.org/pdf/1804.06536.pdf>"
- [8] 8. "Pontiki, M., Galanis, D., Papageorgiou, H., Androutopoulos, I., Manandhar, S., Mohammad, A. S., & Hoste, V. (2016, June). Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016) (pp. 19-30). <https://www.aclweb.org/anthology/S15-2082>"
- [9] "SemEval-2016 Task 5: Aspect Based Sentiment Analysis Maria Pontiki*1, Dimitrios Galanis1, Haris Papageorgiou1, Ion Androutopoulos1,2, Suresh Manandhar3, Mohammad AL-Smadi4, Mahmoud Al-Ayyoub4, Yanyan Zhao5, Bing Qin5, Orphée De Clercq6, Véronique Hoste6, Marianna Apidianaki7, Xavier Tannier7, Natalia Loukachevitch8, Evgeny Kotelnikov9, Nuria Bel10, Salud Maria Jiménez-Zafra11, Gülşen Eryigit12 <https://www.aclweb.org/anthology/S16-1002.pdf>"
- [10] "SemEval-2015 Task 12: Aspect Based Sentiment Analysis Maria Pontiki*, Dimitrios Galanis*, Haris Papageorgiou*, Suresh Manandhar±, Ion Androutopoulos* * <https://www.aclweb.org/anthology/S15-2082.pdf>"
- [11] "Various Approaches to Aspect-based Sentiment Analysis Amlaan Bhoi Department of Computer Science University of Illinois at Chicago Chicago, IL, USA abhoi3@uic.edu Sandeep Joshi Department of Computer Science University of Illinois at Chicago Chicago, IL, USA sjoshi37@uic.edu . <https://arxiv.org/pdf/1805.01984.pdf>"
- [12] "http://UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis Toma's Brychcin in NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzita 8, 306 14 Plzeň Czech Republic brychcin@kiv.zcu.cz Michal Konkol NTIS – New Technologies for the Information Society, Faculty of Applied Sciences, University of West Bohemia, Univerzita 8, 306 14 Plzeň Czech Republic konkol@kiv.zcu.cz Josef Steinberger Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzita 8, 306 14 Plzeň Czech Republic jstein@kiv.zcu.cz <https://www.aclweb.org/anthology/S14-2145.pdf>"
- [13] "Semi-supervised Aspect Based Sentiment Analysis for Movies using Review Filtering Deepa Ananda Deepan Naorema <https://www.sciencedirect.com/science/article/pii/S1877050916300850>"