Multiobjective Optimization for Aspect Based Sentiment Analysis

A Thesis

Presented to

The Academic Faculty

by

AMIT KUMAR

In Partial Fulfilment

of the Requirements for the M.Tech Degree in

Mathematics & Computing

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY PATNA

MAY 2016

Certificate of Approval

Date:

Certified that the thesis entitled **Multiobjective Optimization for Aspect Based Sentiment Analysis** submitted by **Amit Kumar** to Indian Institute of Technology Patna, for the award of the degree of M.Tech has been accepted by the examination committee and that the student has successfully defended the thesis in the viva-voce examination held today.

(Supervisor) (External Examiner) (Internal Examiner)

Date:

DECLARATION

I certify that

a. The work contained in this thesis is original and has been done by myself under the general supervision of my supervisors.

b. The work has not been submitted to any other institute for degree or diploma.

c. I have followed the institute norms and guidelines and abide by the regulation as given in the Ethical Code of Conduct of the institute.

d. Whenever I have used materials (data, theory and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the reference section.

5. The thesis document has been thoroughly checked to exclude plagiarism.

....

Signature of the Student
Roll No. 1411MC02

THESIS CERTIFICATE

This is to certify that the thesis entitled **Multiobjective Optimization for Aspect Based Sentiment Analysis**, submitted by **Amit Kumar** to Indian Institute of Technology Patna, is a record of bonafide research work under our supervision and we consider it worthy of consideration for the degree of Master of Technology of this Institute. This work or a part has not been submitted to any university/institution for the award of degree/diploma. The thesis is free from plagiarized material.

Dr. Asif Ekbal **Dr. Chris Biemann**

Place: Patna

Date: 13th May 2016

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my Master's project advisors **Dr. Asif Ekbal, Assistant Professor, Dept. of Computer Science and Engineering, IIT Patna** and **Prof. Chris Biemann, Language Technology group, Technische Universität Darmstadt** for their excellent guidance and advice. Their guidance helped me a lot in all the time of my research and writing of this thesis. I would also like to thank German academic exchange service (**DAAD**) for providing me scholarship to conduct my thesis work at Technische Universität Darmstadt, Germany.

A special thanks to **Sarah Kohail**, **Deepak Kumar Gupta** and **Md Shad Akhtar** who constantly withhold me from being lost in the vast domain of research. They always cleared my doubts and gave me the direction whenever I got stuck during the work.

I would be failing in my duty if I do not thank **Prof. Pushpak Bhattacharyya, Hon'ble Director, IIT Patna** and **Dr. Sriparna Saha, Assistant Professor, Dept. of Computer Science and Engineering, IIT Patna** whose lectures in Machine Learning helped me to get to the fundamental of the subject. I also thank all the faculty members of Department of Mathematics and Department of Computer Science and Engineering. I thank all the members of Language Technology(LT) group at TU Darmstädt and NLP researchers at IIT Patna for their kind supports.

Last but not the least, I thank my family: my parents, my loving sister, brother and my friends who have always supported, loved and believed in me.

Amit

ABSTRACT

KEYWORDS: Aspect Based Sentiment Analysis (ABSA), Sentiment Analysis, Feature Selection, MOO Framework.

The number of user generated online contents has increased too much in the recent past. The substantial growth of e-commerce has led to an enormous number of reviews for a product or service. This provides proper information to the users to take a fully informed decision on whether to acquire the service and/or products or not. Each of these reviews may describe about the different features of the products. Hence, the customer has to come across a large number of reviews before he/she can arrive to a fully informed decision on whether to buy the product or not. With enormous data, comes the problem of finding the precise information required by the user. In this thesis we present the technique for automatic feature selection for aspect term extraction, opinion target expression detection and sentiment classification. As it is not a priori clear that features from which of our techniques are more useful for a particular task or language, so feature selection is desired. Optimization techniques are widely used in areas such as economics, engineering and have shown promise in human language technology (HLT). We treated feature selection problem as a multi-objective optimization problem. We come up with a novel approach based on multiobjective optimization (MOO) and Machine Learning, and the approach is planned to solve the problems of a well-known information extraction problem, Aspect Based Sentiment Analysis. We have done our experiments on the benchmark datasets of SemEval-2014[1] and SemEval-2016[2] Aspect Based Sentiment Analysis tasks. Comparisons of the performance of our system with baselines and other existing systems show that our system achieves encouraging accuracies with reduced set of features in all kind of settings.

[1]http://alt.qcri.org/semeval2014/task4-/index.php?id=data-and-tools
[2]http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ABSA	Aspect based Sentiment analysis
CRF	Conditional Random Fields
DT	Distributional Thesaurus
ML	Machine Learning
MOO	Multi objective optimization
NLP	Natural Language Processing
OTE	Opinion target expression
SMO	Sequential minimal optimization
SOO	Single objective optimization

CHAPTER 1

INTRODUCTION

These days people are more attracted towards internet. They are using internet very frequently and even this is true for activities relevant to business and commerce. The number of e-commerce portals has grown tremendously. This has made the lives of people more comfortable. They prefer to spend more time on internet for buying a product or to acquire any service instead of going anywhere personally. The number of forums, blogs, review sites etc. are increasing day by day, as many people search for the suggestions from the fellow users before taking any decision on whether to buy the product or to acquire any service or not. Actually, these forums, blogs, review sites etc. are very helpful for both the manufacturers as well as the customers. After getting feedbacks of the customers, manufacturers can improve or enhance the quality of their products by focusing on the components on which they are lacking. Customers can take the service, which is more related to them. To develop automated technique for mining relevant information is thus, very important.

1.1 Natural Language Processing (NLP)

Natural Language Processing (NLP) is a branch of Artificial Intelligence which has the potential to accept the human speech as it is conveyed. The advancement of NLP applications is very challenging because a computer can understand the human's language only when it is spoken in some programming language, which is precise, unambiguous and highly structured. Human speech, however, is not always precise – it is often full of ambiguity and the linguistic structure depends on region, context etc.. Natural language generation is also a challenging problem in NLP that includes generating computer's response in natural language.

In many applications we use the concept of NLP. In reality, any task which deals with the text is an application of NLP. The most frequent applications that use NLP are Deep Analytics, Sentiment Analysis, Machine Translation, Question Answering, Summarization and many more.

The aim of NLP is to train computers to understand human dialogue in order to do away with machine languages like perl, python or C all together. With NLP, machines would be able to directly understand human language and speech. In our task we focus on an NLP application namely Aspect Based Sentiment Analysis for multiple languages.

1.2 Sentiment Analysis

Sentiment analysis, also called polarity identification, is the area of study that analyzes human's opinions, sentiments, evaluations, appraisals, attitudes and emotions towards entities such as products, services, organizations, individuals, issues, events, topics and their attributes [3, 10].

Sentiment analysis deals with analyzing emotions, feelings, the attitude of a speaker or a writer from a given piece of text. "Sentiment analysis or opinion mining refers to the application of natural language processing, computational linguistics, and text analytics to identify and extract subjective information in source materials" [49]. It involves analyzing of individual's behavior, likes and dislikes towards the target entity. There is no definite definition of "Sentiments", but in general they are considered as thoughts, views and attitude of an individual arising mainly based on the emotion instead of a reason. Sentiments are considered as the manifestation of our feelings and emotions. This area of computer science analyzes and predicts the hidden information stored in the text. Sentiment analysis categorizes the sentiment into subjective and objective nature. Subjective text contains some opinion whereas objective text doesn't contain any opinion.
Some examples-

1. **Subjective** - Virat Kohali is a superb batsman.
   (this sentence contains an opinion, 'superb'indicates the feeling and hence it is subjective.)
2. **Objective** - The Sun rises in the east.
   (This sentence doesn't contain any opinion, just a simple information and hence

it is objective.)

The subjective text can be classified into 3 categories based on the sentiments expressed in the text.

1. **Positive** - India is a beautiful country.
2. **Negative** - The movie was horrible.
3. **Neutral** - Food was okay, nothing great. (This sentence does not have any positive or negative polarity so it is neutral .)

Many NLP researchers have already published application oriented research papers in this area. For example, in [4], The authors proposed a sentiment model to predict sales performance. To rank products and merchants, reviews were used in [6]. In [5], the connections between the NFL betting line and public opinions in blogs and twitter were studied. In [7], The authors linked between twitter sentiment and public opinion polls.

1.3 Aspect Based Sentiment Analysis(ABSA)

Sentiment analysis is the task of identifying the sentiments (positive, negative or neutral) of the users based on the opinions and emotions expressed in the reviews written either for a particular product or service or any of its aspects (or feature/attribute). Classification of the sentiment at sentence and document level doesn't satisfy the need of user's requirement. They need the more precise information *i.e.* sentiment related to particular aspects (or features) of the product.

An *aspect* refers to an attribute or a component of the target entity that has been commented in a text.
Example:-

*"**Food** is good but the **service** is poor."*

In the above review there are two aspect terms, *Food* and *service*. Sentiment for these two aspect terms are contrasting in nature. For the first aspect term we have positive review but for the second we have negative review.

The tasks of ABSA are to extract all the relevant aspects from the text for which opinions have been expressed and classify these opinions into negative, positive or neutral [2, 3] category. Aspect terms can affect the sentiment polarity within same domain. For example, in restaurant domain, *cheap* is generally positive if we take *food* as the aspect term, but it indicates negative polarity when *ambiance* or *decor* acts as the aspect term [8].

ABSA was introuced in SemEval-2014 shared task [27]. In this shared task[1], for a given set of sentences with pre-identified entities, we have to identify all the aspect terms present in the sentence and return a list containing all the aspect terms. Then, for the given set of aspect terms within a sentence, we have to identify sentiment for each of the aspect terms whether it is positive, negative, neutral or conflict (i.e., both negative and positive).
Example:-

*"I liked the **service** and the **staff**, but not the **food**."*

service : Positive, staff : Positive, food : Negative

In SemEval-2016 shared task[2] ABSA, the organizers extended their previous SemEval-2014 shared task [27] task ABSA and introduced the concept of Opinion Target Expression (OTE). There are three subtasks, Aspect Category detection, OTE detection, Sentiment Classification.

Aspect Category : Aspect Category consists of *E#A* pair. We have to identify each entity E and attribute A pair for which an opinion is expressed in the given text. E and A should be chosen from predefined inventories of entity types (e.g. RESTAURANT, FOOD) and attribute labels (e.g. PRICE, QUALITY).

Example:-

"Everything is always cooked to perfection, the service is excellent, the decor cool and understated."

FOOD#QUALITY, SERVICE#GENERAL, AMBIENCE#GENERAL

Here FOOD#QUALITY, SERVICE#GENERAL and AMBIENCE#GENERAL are the

[1] http://alt.qcri.org/semeval2014/task4/
[2] http://alt.qcri.org/semeval2016/task5/

aspect category.

Opininon Target Expression (OTE) : OTE is the linguistic expression used in the given text that refers to the reviewed entity E of each E#A pair. If entity is not explicitly mentioned then the value will be "NULL". Example:-

"Everything is always cooked to perfection, the ***service*** *is excellent, the* ***decor*** *cool and understated."*

Opinion target="NULL" category="FOOD#QUALITY" from="0" to="0"

Opinion target="service" category="SERVICE#GENERAL" from="47" to="54"

Opinion target="decor" category="AMBIENCE#GENERAL" from="73" to="78"

Sentiment Classification : We have identified polarity for each of the textitE#A, OTE tuple.

Example:-

"Everything is always cooked to perfection, the service is excellent, the decor cool and understated."

Opinion target="NULL" category="FOOD#QUALITY" *polarity="positive"* from="0" to="0"

Opinion target="service" category="SERVICE#GENERAL" *polarity="positive"* from="47" to="54"

Opinion target="decor" category="AMBIENCE#GENERAL" *polarity="positive"* from="73" to="78"

1.4 Hypotheses and Objectives

The objective of Aspect based Sentiment Analysis (ABSA) is to produce feature based opinion summary from multiple reviews. It primarily focuses on mining relevant information from the large number of online reviews available for a popular product or service. We hypothesize that we can improve upon the results of the aspect term extraction and sentiment classification by incorporating the techniques of lexical expansion of text and optimization. Feature selection will be proposed based on the concept of MOO.

Hypotheses

- Lexical expansion can be helpful in finding the aspect term, OTE and sentiment classification.
- Feature selection can be effective to achieve improved performance.
- MOO techniques can be effectively used for finding features for aspect term extraction, OTE extraction and aspect based sentiment classification.

Objectives

- Develop a framework using MOO.
- Studies on the feasibility of the proposed approach for solving aspect term extraction, OTE extraction and sentiment classification problems.
- Adapting specific modules such as feature selection [9] for the particular task.
- Development of a high accurate aspect term extraction, OTE detection and sentiment classification system.

Our main focus is to develop a framework using multi-objective optimization (MOO) and study the feasibility of the proposed approach for solving the information extraction problems, particularly aspect based sentiment analysis. Various specific modules such as feature selection and parameter optimization using MOO are to be developed.

This thesis is structured as follows: In Chapter 2 we discuss the related work and literature survey that is relevant to our work. We explain research methods and techniques employed in this work in Chapter 3. Feature selection using MOO is discussed there in detail. In Chapter 4 experimental results are reported and discussed along with various details and specifications of experiments. Results of various experiments are analysed and discussed in detail. Chapter 5 concludes the present work and outline the further possibilities of improvement.

CHAPTER 2

Literature and Research Review

Sentiment analysis has got popularity among NLP researchers from last decade. A lot of works have already been done for English language, but for languages other than English have just begun. Extracting sentiments through lexicon and dictionary based approaches has given way to machine learning techniques by incorporating syntactic and semantic features. In this chapter, we discuss the previous work done in the field of sentiment analysis.

Syntactic Approaches : In [13], The authors used syntactic approach for sentiment classification using n-Grams. Traditional n-Gram along with POS information have been used to perform machine learning by the authors. They used Naive Bayes, Maximum Entropy and Support Vector Machines for three fold cross validation. They tried different combinations of n-Gram approach in their experiments, such as unigrams+bigrams, unigrams+POS, most frequent unigrams etc.. They achieved on the conclusion that incorporating the frequency of matched n-gram might be a feature. They obtained maxiumum of 82.9 % accuracy in all their experiments in the presence of unigram using SVM.

Semantics/Pattern Mining : Semantic approaches in combination with part of speech learning has also been very popular in determining the sentiment of a text. In [10] and [17], authors have performed binary classification using this approach. Different NLP techniques like word sense disambiguation, chunking, n-gram have been used to perform binary polarity classification in [14] and [18]. Ohana et. al. [15] and Saggion et. al. [16] used sentiwordnet for identifying the sentiment of a given text. They computed negative and positive score for the given review and based on the high score, polarity has been assigned. They also extracted features and used machine learning techniques to identify the polarity. Turney [10] and McKeown [35] used part of speech information in finding the sentiments. Lawrence et. al. [1] came up with their own scoring function which was based on probability. For negation handling, they used lexical substitution and for deciding the label of the review, rainbow classifier was used.

Features/Machine Learning : Many of the works have also been done using machine learning for finding the sentiments. For performing supervised machine learning pang [13] deduced some of the features. The performance of feature based learning is better in comparison to the traditional semantics and syntactic approaches.

Sentiment analysis can be done on three levels:

1. **Document level :** In document level sentiment classification, the whole document focuses on a single object and contains opinion from a single opinion holder. The task is to determine the overall sentiment orientation of the document. In Pang [13] and Turney [10], the authors have done sentiment classification on document level.

2. **Sentence level :** A document consists of many sentences. Here each of the sentence is analyzed individually and classified as negative, positive or objective. It is assumed that a sentence contains only one opinion. The task is to identify whether the sentence contains any opinion or not, if yes then find the opinion. Many significant works have been performed by Wiebe et. al. [21], Yu et. al. [22], Hu et. al. [2] and Kim [11] with respect to the sentence level sentiment classification.

3. **Aspect or Phrase level :** In aspect level sentiment classification, our task is to extract all of the object features i.e. aspect(s) that have been commented in the review and then find the sentiment around each of the aspect. Wilson et. al. [20], Agarwal et. al. [19] have done significant works on phrase level.

Literature indicates that aspect-based sentiment analysis has got the attention of NLP researchers in the recent past. Previously, sentiment classification focused primarily on document [10] and sentence [11] level. But, this information is not enough for people seeking opinions for particular product features. They are interested in specific product's features. This is a topic of aspect-based sentiment analysis [33]. Earlier approaches for aspect term extraction are based on frequently used noun and noun phrases [2, 12, 23]. In [2], The authors proposed the method which identifies frequently present noun phrases from the text based on association rule mining. This type of approach works well when high frequency of aspects are strongly co-related with certain types of words (e.g. Noun), but many times fails when frequency of the terms, which used as the aspects are low. Supervised machine learning techniques [24, 25] are being widely used with the emergence of various labelled datasets. Some other techniques for extracting aspect terms include manually specified subset of the Wikipedia category [31] hierarchy, semantically motivated technique [10] and unsupervised clustering technique [12]. Phrase dependency tree [26] is also helpful in aspect term extraction.

Most of the features used for aspect term extraction or sentiment classification exploit syntactic, lexical or semantic level features. The features used for a domain often fail to perform well for the other domains. The method for automatic feature selection for these two tasks have not been attempted so far.

CHAPTER 3

Research Methods

In this chapter we discuss the various techniques what we have used for our problems "Aspect Term Extraction (ATE)", "Opinion Target Expression (OTE)" extraction and "Sentiment Classification". We discuss the lexical acquisition features Distributional Thesaurus (DT) in section 3.1 and Unsupervised POS tagging in section 3.2. In section 3.3, we discuss the lexical expansion methods and results, what we used for resource poor languages in our experiment. After that, we discuss MOO, sequence learner CRF Suite, SMO and data-preprocessing in the upcoming sections. Last section describes the feature sets for different tasks. The significance of these techniques in our research have also been highlighted.

3.1 Distributional Thesaurus (DT)

Thesaurus lists the synonyms words or the related concepts. Like the major thesaurus, DT also groups the similar words (sometimes antonym also) based on the computation of distributional similarity index. For the high frequency words which occur in corpora (on which the DT model is built), the most similar words are computed over the similar context in the DT corpora, which implement the distributional hypothesis [38]. This automatically induced lexical resource [37] is used for lexical expansion of text by virtually expanding every content word in the text with the list of most similar words from the DT corpora. The authors used the concept of DT in [37] for lexical expansion of the given text in Word Sense Disambiguation (WSD) problem.

Approaches based on matching sense definitions from dictionaries with contexts of ambiguous terms to assign correct sense go through a type of “lexical gap problem". This problem arises when we do not have enough words to express the certain meaning or sense. To solve this problem, the authors in [37] produce the lexical expansion with help of distributional thesaurus. They lexically expanded each of the words with help of DT, which was constructed by contextual information from unlabelled corpora. By

Language	Token	Similar DT Words		
English	Good	Great	Excellent	Nice
English	India	Italy	Africa	China
Dutch	snel	gauw	vlug	spoedig
Dutch	wonderlijk	merkwaardig	vreemd	bizar
French	ample	large	vaste	véritable
French	fatigué	fatiguée	content	déçu
Spanish	humanitario	solidario	político	comunitario
Spanish	cantidad	cifra	cuantía	proporción

Table 3.1: Example of Distributional Thesaurus (DT)

this way, the authors in [37] created the approach using DT for WSD problem, which performs better than the traditional approaches. We are incorporating this technique in our task, which is inspired by the above discussed work. An example of lexically expanded tokens can be seen in Table 3.1

3.2 Unsupervised PoS Tagging

As we know that, to assign PoS tags to a word is an important step in many higher level NLP tasks. All the supervised methods for PoS tagging learn tag probabilities from pre-exist manually tagged corpus. Hence its reachability and application is limited to resource-rich languages. It is very difficult to develop efficient supervised PoS tagger for resource poor and minority languages. It means that unavailability of efficient PoS tagger for a minority language affects the other NLP tasks for that language.

Unsupervised PoS tagging is a technique which does not require a pre-existing manually tagged corpus to construct the tagging model, hence suitable for minority and resource poor languages. Unsupervised PoS tagging employes the idea of bootstrapping. These techniques build their training model based on untagged corpus and method itself generates the number of tags. These techniques observe the words's pattern in the untagged corpus and extract different PoS tags from it. Word local contexts play an important role in inducing PoS categories. For example, words like sing, dance or play occur in similar contexts, while words like I, we or they occur in different ones. Unsupervised PoS tagger produces slightly different categories as what is assumed by a linguistically motivated PoS tagger.

In [36], the authors implemented a system which takes a reasonable amount of tokenized and unlabelled text without PoS information mentioned as input and induces number of word clusters. The Chinese Whispers, a graph clustering algorithm is used in two stages which is based on contextual similarity. The size of feature words, target words and window are the main parameters of the algorithm. In first stage, the most frequent target words within a context of the most frequent feature words, which appear either to the left or right of a target word, are clustered based on their context statistics. In second stage, pairwise similarity scores computed by the number of shared neighbours between two words in context window of size four. Then, the final clustering is formed by combining the cluster obtained in these two stages.

Algorithm 1 : Chinese Whispers, a graph clustering algorithm

```
for all v_i ∈ V
        class(v_i) = i
end for
for it = 1 to no. of iteration do
      for all v ∈ V, randomised order do
         class(v) = predominant class in neigh(v)
      end for
end for
return partition P induced by class labels
```

For extracting the unsupervised POS tag, we tokenized the text before applying the script[1]. This step has to be repeated for other languages as well.
Example:

The|76 wine|3 list|169 is|238 interesting|4 and|231 has|282
many|217 good|4 values|1 .|337

The number beside each of the token represents the corresponding unsupervised POS tag.

3.3 Lexical Expansion

Based on distributional hypothesis, we induce the lexical expansion for sentiment words. Our observation is that for unseen tokens and limited number of lexicons, this expansion can provide a useful backoff technique, also cf. [43].

[1] http://wortschatz.uni-leipzig.de/ cbiemann/software/unsupos.html

Language	Seed Lexicon			Induced Lexicon	Common Entries
	Positive	Negative	Neutral		
English	2005	4789	-	12953	4120
Dutch	3314	5923	-	8496	2992
French	9338	10339	5993	18308	7636
Spanish	2175	1737	7869	12480	4306

Table 3.2: Expansion statistics of induced lexicons. Common entries indicate the number of words, occur in both seed lexicon and induced lexicon.

For all the languages, we construct the list of sentiment lexicons using an external corpus and seed sentiment lexicon. We use Bing Liu's lexicon [2] for English, VU sentiment lexicon[2] for French, Dutch and Spanish as seed lexicons. To induce the lexicons, we get the top 100 DT expansion of each token in the seed lexicon. Then we define candidate terms based on two criteria: a) atleast present in the expansion of 10 seed terms and b) having corpus frequency more than 50 in the background corpus (English[3], French[4], Spanish[5], Dutch[6]). We calculate the normalized negative, positive and neutral score for each token similar to [40] and inspired by [35]. We assume that words tend to be semantically more similar to words of same sentiment. So, words which appear more number of times in positive (negative/neutral), get higher positive (negative/neutral) sentiment score. Here we calculate only normalized positive, negative and neutral score rather than assigning a polarity class as done in [40]. The following two factors influence the amount of induced lexicon: (i) number of words having expansion in the seed lexicon and (ii) pruned threshold value in obtaining the induced lexicon. High threshold value for accepting the candidate term and all the seed lexicons not having expansion shorten the amount of induced lexicon. The statistics of the induced lexicons have been provided in Table 3.2.

3.4 Multi-objective optimization (MOO)

To understand Multi-Objective Optimization (MOO) problems, at first we must understand that what is Single Objective Optimization (SOO) problems and what is the

[2]https://github.com/opener-project/VU-sentiment-lexicon
[3]https://snap.stanford.edu/data/web-Amazon.html
[4]http://wacky.sslmit.unibo.it/doku.php?id=corpora
[5]http://corporafromtheweb.org/escow14/
[6]http://corporafromtheweb.org/nlcow14/

difference between single objective optimization and multi-objective optimization problems. Suppose we have to travel from Patna to Delhi. We make the list of all possible tickets and select the cheapest one. So here we are only optimizing the price, our only one objective function is to minimize the travelling cost. This is single objective optimization problem. But, at the same time if we want to minimize the travelling time also then we have to pay more. Then, we have two objective functions travelling time and travelling cost. Even in this problem we can add more constraints , like we have to travel some particular day , within some particular time etc.. The later problem is multi objective optimization problem. So, in single objective optimization we have only one objective function but in multi objective optimization we have more than one objective functions. In our daily life we have to face such situations where we have to deal with more than one objective functions. A general multi objective optimization (MOO) problem is shown in Equation 3.1 .

$$\begin{aligned}
&\text{Minimize/Maximize } f_m(x), m = 1, ..., M \\
&\qquad\qquad \text{subject to } g_j(x) \geq 0, j = 1, ..., J; (inequality\ constraints) \\
&\qquad\qquad\qquad h_k(x) = 0, k = 1, ..., K; (equality\ constraints) \\
&\qquad\qquad\qquad x_i^L \leq x_i \leq x_i^U, i = 1, ..., N
\end{aligned} \tag{3.1}$$

The solution of above general equation is x, a vector $(x_1, x_2...., x_N)^T$ of size N, where $x_i; i = 1, ..., N$ represents the decision variables. The last set of constraints are called variable bounds, constrainting each of the decision variable x_i to take a value between x_i^L, lower value and x_i^U, upper value. They are called bounded variables. The above specified optimization problem has J number of inequality constraints and K number of equality constraints. $g_i(x)$ and $h_k(x)$ represent the constraint functions. The solution x which does not satisfy all of the $(J + K)$ number of constraints and all of the $2N$ variable bounds is called an infeasible solution. On the contrary, the solution y which satisfies all variable bounds and constraints is called a feasible solution. The set of all possible feasible solutions is called the feasible region and represented by S.

In the above specified general optimization problem, For expression $f_m(x)$, m can take any possible value between 1 to M. That means here more than one objective function can be defined and for each value of m, we can either minimize or maximize the function. Here unlike the single objective optimization problem, the objective function can be from the multidimensional space, in addition to the usual decision variables space that is same in both type of optimization problem. In MOO, mapping takes place between an $N - dimensional$ vector solution to $M - dimensional$ objective vector, while in SOO problem, mapping takes place between $N - dimensional$ solution vector and $single - dimensional$ objective vector.

3.4.1 MOO Algorithms

The above specified multi-objective optimization (MOO) problem in Eq. 3.1 can be stated in other language as : If M objective functions $f_1(x), f_2(x), ..., f_M(x)$ have been given then to find vectors x of decision variables and satisfy the given constraints if there is any.

Domination plays an important part in multi-objective optimization (MOO) problem. A solution $x^{(1)}$ is said to dominate the other solution $x^{(2)}$, if the following conditions are true:

1. The solution $x^{(1)}$ is no worse than $x^{(2)}$ in all objectives.
2. The solution $x^{(1)}$ is strictly better than $x^{(2)}$ in at least one objective.

The set of all the non dominated solutions is called non dominated set and the globally pareto optimal set is called the non dominated set of entire search space.

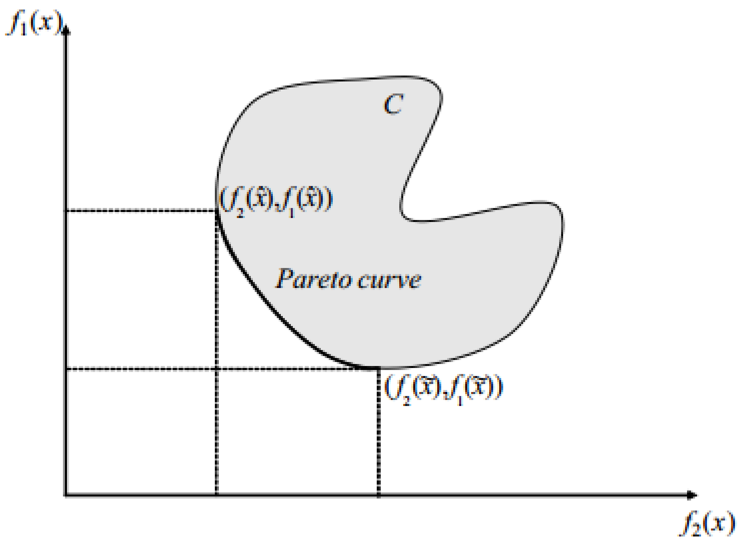

Figure 3.1: Pareto optimal front [42]

MOO Performance Measures: In MOO, we find a set of solutions. The solution set should converge as close to the true pareto optimal front as possible and it should maintain the diversity [28]. The first condition secures the optimality of the solution and the second condition secures the diversity of the solution. In literature a lot of approaches exist to solve MOO problems [28]. They use the concept of aggregating, population based pareto and non-pareto techniques. In aggregating techniques, all the objective functions are usually combined into one function using weighting or goal based methods. We will discuss one of the existing approach which we have used for our problems in finding the best possible solutions for aspect term extraction and sentiment classification.

Non-dominated Sorting Genetic Algorithm (NSGA-II): In our Experiment, for multi-objective optimization we have chosen NSGA-II [28] because it provides several benefits over others:

1. Low computation cost $O(MN^2)$, where no. of objective functions = M and Population size = N
2. High elitism (In each iteration keep the best individual from the parent and child population.)
3. Diversity in population without having to specify any parameter manually
4. Easy to use

NSGA-II Algorithm : This Algorithm is based on the concept of Genetic Algorithms (GA). In recent days, GA is known to be more effective than the traditional optimization methods such as goal programming, weighted metrics method. The flow chart of this algorithm has been shown in Fig. 3.3.

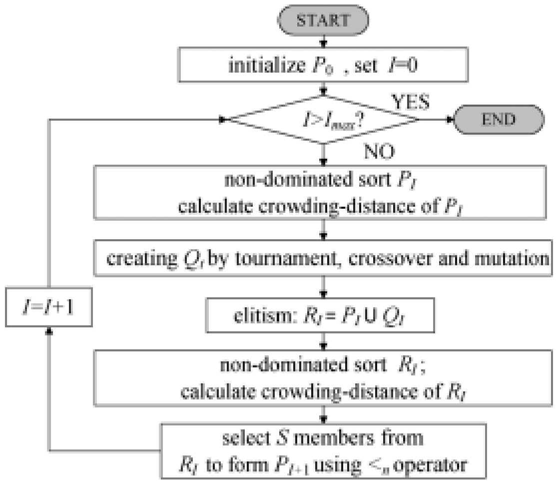

Figure 3.2: Flow Chart of NSGAII [41]

Algorithm 2 : Non-dominated Sorting Genetic Algorithm-II (NSGA-II)

$R_t = P_t \cup Q_t$
F = fast - non - dominated - sort(R_t)
P_{t+1} = $\emptyset$ and i = 1
while $|P_{t+1}|$ + $|F_i| \leq N$ do
    crowding - distance - assignment(F_i)
    $P_{t+1} = P_{t+1} \cap F_i$
    i = i + 1
end while
Sort($F_i <_n$)
P_{t+1} = $P_{t+1} \cap F_i$
Q_{t+1} = make - new - pop(P_{t+1})
t = t + 1

This algorithm uses the search capability of GA and tries to minimize the fitness functions (*i.e.* objective functions). It starts with creating the parent population P_0 of size N. Each of the candidates in the population is called chromosome. For each of the chromosome, fitness function is computed. By applying binary tournament selection, recombination and mutation operators on parent population P_0, a child population Q_0 of size N is created. In t^{th} iteration, a combined population of parent and child popularion R_t = P_t + Q_t is formed. The size of combined population R_t is $2N$. All the candidates of R_t are sorted according to non-dominated sorting algorithm. The non dominated set is F_1, F_2, F_3... according to order (best to worse). If the size of F_1 is smaller than N, then all of the solutions of F_1 is included in P_{t+1}. The remaining members of population P_{t+1} is selected from next subsequent non dominated fronts according to their ranking. To select exactly N solutions, the solutions of the last included front are sorted according to crowding distance operator and the best among them are selected to fulfill the exact number of N solutions in P_{t+1}. Then the new population P_{t+1} is used for selection, crossover and mutation to create a child population Q_{t+1} of size N. In this way parent population and child population are created until the specified number of generations or the stopping criteria is met as shown in Algorithm 1.

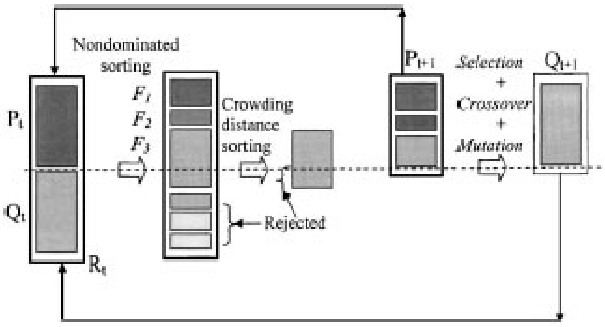

Figure 3.3: Non-dominated Sorting Algorithm Procedure [28]

3.5 CRF Suite

CRF Suite[7] is a fast implementation of Conditional Random Field (CRFs) for labelling sequential data. Conditional Random Field (CRF) [34] is a robust statistical classifier widely used for building the probabilistic models to segment and label the sequence data. The conditional probability of a state sequence $s =< s_1, s_2, \ldots, s_T >$ given an observation sequence $o =< o_1, o_2, \ldots, o_T >$ is calculated as:

$$P_{\wedge}(s|o) = \frac{1}{Z_o} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_k \times f_k(s_{t-1}, s_t, o, t)),$$

where, $f_k(s_{t-1}, s_t, o, t)$ is a feature function whose weight λ_k, is to be learned via training. The values of the feature functions may range between $-\infty, \ldots + \infty$, but typically they are binary. To make all conditional probabilities sum up to 1, we must calculate the normalization factor,

$$Z_o = \sum_{s} \exp(\sum_{t=1}^{T} \sum_{k=1}^{K} \lambda_k \times f_k(s_{t-1}, s_t, o, t)),$$

which as in Hidden Markov Model (HMM), can be obtained efficiently by dynamic programming.

[7] http://www.chokkan.org/software/crfsuite/

A feature function $f_k(s_{t-1}, s_t, o, t)$ takes the value of 0 for most cases and is only set to be 1, when s_{t-1}, s_t are certain states and the observation has certain properties.

3.6 Sequential Minimal Optimization (SMO)

The machine learning and computer vision research communities have started to enjoy adoption of SVMs. However, it has not got so much popularity among engineering community. There can be two possible reasons: first, training in SVM is slow for large problems. Second, the training algorithms in SVM are slow and sometimes hard to implement. The new SVM algorithm, Sequential Minimal Optimization (SMO) [39] is simple, easy to implement, faster and better scaling properties than the previous one. Previous SVM learning algorithms use numerical quadratic programming (QP) as an inner loop, whereas SMO uses an analytic QP step. SMO consumes most of the time in evaluating the decision function instead of performing QP, it can exploit for sparse datasets, with either binary or non binary input data.

The quadratic problem to train SVM is shown below:

$$\begin{aligned} \max_{\alpha} W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} y_i y_j k(\vec{x_i}, \vec{x_j}) \alpha_i \alpha_j, \\ 0 \leq \alpha_i \leq C, \forall i, \\ \sum_{i=1}^{l} y_i \alpha_i = 0 \end{aligned} \tag{3.2}$$

The problem in above equation is solved by SMO algorithm. A point is optimal iff the Karush-Kuhn-Tucker *(KKT)* conditions are fulfilled and $Q_{ij} = y_i y_j k(\vec{x_i}, \vec{x_j})$ is positive semi-definite. That point may be non-unique and non-isolated optimum. The *KKT* conditions are simple:

$$\begin{aligned} \alpha_i = 0 \Rightarrow y_i f(\vec{x_i}) \geq 1, \\ 0 < \alpha_i < C \Rightarrow y_i f(\vec{x_i}) = 1, \\ \alpha_i = C \Rightarrow y_i f(\vec{x_i}) \leq 1, \end{aligned} \tag{3.3}$$

SMO solves any QP task without any extra storage and calling the iterative numerical subroutine for each sub-problem. At each step, it selects to solve smallest possible optimization problem. Here the smallest optimization problem involves two lagrange multipliers and lagrange multipliers must obey the linear equality constraint.

The advantage of SMO is due to the fact that two lagrange multipliers can be solved analytically. Thus, it avoids the entire inner iteration of numerical QP optimization. The SMO algorithm may be less susceptible to numerical precision problems.

3.7 Data Pre-processing

We use the datasets provided in the SemEval-2014[8] and SemEval-2016 [9] shared task. We pre-process these data and remove all the XML tags. We run Stanford CoreNLP[10] suite in order to extract the information such as lemma, part of speech (PoS) and named entity (NE) in English language. For language other than English , we use Universal parser[11] for tokenization and parsing. Since aspect-term and opinion target expression (OTE) detection are sequence labelling problems, it is necessary to identify the boundary of any aspect term and OTE properly. We follow the standard BIO notation where **"B-ASP"**, **"I-ASP"** and **"O"** represent the beginning, intermediate and outside token of a multi-word OTE or aspect term.
e.g. Here "Battery life" is the aspect term.

"Battery	life	is	very	good	."
B-ASP	I-ASP	O	O	O	O

3.8 Feature Extraction

3.8.1 Features for Aspect term and OTE extraction

We identify and implement the following set of features for aspect term and opinion target expression (OTE) extraction tasks.

1. **Word and local context :** Context words of a token play a significant role in determining the aspect term. They give a very effective information. We use

[8] http://alt.qcri.org/semeval2014/task4-/index.php?id=data-and-tools
[9] http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools
[10] http://nlp.stanford.edu/software/corenlp.shtml
[11] http://www.undl.org/unlsys/uparser/UP.htm

the current token, its lower case form, succeeding five words and preceding five words of the current token as the features.

2. **Part of Speech (PoS) information:** Generally an aspect term is a noun, verb, adverb or adjective. Latest research [30] shows that 60-70% of the aspect term belongs to explicit noun category. Thus, PoS information of the token can be very useful in extracting the aspect term. We use PoS information of current, preceding two and following tokens as the features. We use Stanford CoreNLP[12] for English language and Universal parser[13] for language other than English for extracting the PoS information.

3. **Head word :** Generally, aspect terms belong to the category of noun phrase. The head word of the noun phrase is used as the feature.

4. **Head word PoS :** We use PoS information of the head word as the feature for our model.

5. **Prefix and Suffix :** Suffix and Prefix of fixed length character sequences are trimmed from each token and used as the features for our model. Word suffixes are the fixed number of letters stripped from rightmost position in word and prefixes are the letters stripped from leftmost of the word. Here we use prefix (up to 3-character), Suffix (upto 3- character) of context tokens [-1,0,1] and prefix, suffix of 4 -character of current token as the features.

6. **Frequent Aspect Term :** We construct a list of frequently occurring OTEs from the training set. An OTE is considered to be frequent if it appears atleast four times in the training corpus. We define a binary feature for the presence or absence of extracted OTEs.

7. **Dependency Relation :** Here we define two features: we consider the dependency relation as the feature where current token is the modifier and governer.

8. **Character n-grams :** Character n-gram is a contiguous sequence of n character extracted from a given token. We extract character n-grams of length two (bigram), three (trigram), four (quad) of the current token and use them as features in our model.

9. **Orthographic feature :** This feature checks whether the current token starts with the capitalized letter or not.

10. **DT features :** Distributional thesaurus gives the lexicon expansion of the token based on similar context. In [37], the authors have used it for lexical expansion of text by virtually expanding every content word in the text with the list of most similar words from the DT. It is very helpful in unseen texts. We obtain top 5 DTs of current token and top 3 DTs of context tokens [-2,-1,1,2] as the features.

11. **Expansion Score :** OTEs and aspect terms have opinion around them. Opinions are regularly lexicalized with words found in sentiment lexicons. We compute sentiment score based on our induced lexicons by considering the window size of

[12] www.nlp.stanford.edu/software/corenlp.shtml
[13] http://www.undl.org/unlsys/uparser/UP htm

10 (preceding 5 and following 5 of the current token). We use expansion score of context tokens [-2..2] as the features for our model.

12. **Unsupervised PoS tag :** Unsupervised PoS tag of context tokens [-2..2] are used as the features in our model.

13. **Semantic Orientation (SO) Score :** *SO* score [35] is the measurement of negative or positive sentiment expressed in a phrase. *SO* score of each token is computed with Point wise mutual information (PMI) as follows:

$$SO(w) = PMI(w, prev) - PMI(w, nrev)$$

    Here PMI is measurement of association of token w with respect to negative $nrev$ or positive $prev$ reviews. We calculate *SO* score from training data of each domain respectively. A negative *SO* score of a token means that this token is more related to negative than positive reviews.

We additionally extract the following set of features only for English language.

1. **Chunk information :** A text can have multi-word aspect term or OTE. To identify the boundaries of these multi words, we use chunk information of context tokens[-1,0,1] as features.

2. **Lemma :** Lemmatization trims the inflectional forms and derivationally related forms of a token to a common base form. We use lemma of the current token as the feature.

3. **WordNet :** Tokens from the same lexical category those are roughly synonymous are grouped into synsets in WordNet [29]. We extract top four noun synsets of each token and use as feature.

4. **Named Entity Information :** We extract named entity information of the current token with Stanford CoreNLP tool, and use the NERsequence labels in BIO-scheme as features.

5. **SentiWordNet Lexicon[14] :** SentiWord net is one of the most popular sentiment lexicons . We calculate sentiment score of all words that appear in surrounding context (previous-5 and next-5) of the target token. We use SentiWord net score of current token along with the score of surrounding context tokens [-2..2] as features.

e.g. “Judging from previous posts this used to be a good ***place*** , but not any longer."

[14] http://sentiwordnet.isti.cnr.it/

Feature	Value
Local context	used to be a good place , but not any longer
POS	pos[-2]=DT, pos[-1]=JJ, pos[0]=NN, pos[1]=,, pos[2]=CC
HeadWord	1, NN
Prefix	g, p, ,, go, pl, goo, pla
Suffix	d, e, ,, od, ce, ood, ace
Freq. Aspect term List	1, 1
Dependency	Head - mark, cop, det, amod Modifier - xcomp
Character n-gram	Bigram - pl,la, ac,ce, Trigram - pla, lac, ace, Quadgram- plac, lace
DT	dt[-2] = another, aa, some, dt[-1] = great, decent, nice, dt[0] = refuge, chances, precedence, dt[1] = -, ..., –, dt[2] = although, and, ,but
Expansion Score	Sc[-2] = 2, Sc[-1] = 2, Sc[0] = 3, Sc[1] = 2, Sc[2] = 1
SentiWord Net Score	Se[-2] = 6, Se[-1] = 0, Se[0] = 5, Se[1] = 6, Se[2] = 6
Unsupervised POS	unpos[-2]=150, unpos[-1]=4, unpos[0]=3, unpos[1]=98, unpos[2]=16
Chunk Information	chnk[-1]=I-NP, chnk[0]=I-NP, chnk[1]=O
NER, lemma, Orthographic	O, place, 0
Noun Synset	location, cognition

Table 3.3: An example of feature extraction for OTE extraction

3.8.2 Features for Sentiment Classification - SemEval 2014

1. **Aspect term and context :** We convert the actual forms of aspect terms in lower case character and use it as a feature along with the actual aspect terms. The polarity orientation of aspect term heavily depends on local context words where it appears. We also include succeeding five and preceding five of the aspect term to provide the contextual information.

2. **Lexicon :** Sentiment lexicons are the useful resources, which provide important information for predicting the sentiment. For computing lexicon sentiment score we consider the preceding five and following five tokens of the aspect term. We use following set of lexicons.
   - **MPQA :** We take the help of MPQA subjectivity lexicon [32] which contains a list of words denoting the negative, positive and neutral sentiments.
   - **Bing Liu Lexicon :** For each token in training and test set we define the values in the following way: -1 for negative; 1 for positive and 2 for those do not appear in Bing Liu lexicons [45]. Then, we define two features:
     1. We calculate sum of sentiment score of all the words that appear in context of target aspect term and use as a feature.
     2. We also compute the sum of the sentiment scores of only those words which have *direct dependency relation* with the target aspect term.
   - **SentiWordNet Lexicon**[15] **:** This is one of the most widely used lexicons for sentiment analysis. We compute sentiment score of all words that appear in surrounding context (previous-5 and next-5) of the target aspect term.
   - **Other Lexicons :** Apart from above mentioned lexicons we also use AFINN [47], NRC Hashtag, Sentiment 140 [48] and NRC Emotion [46] lexicons for calculating the score and use them as features.

3. **Domain-Specific Words :** In general lexicons, there are no coverage of all the sentiment words which are used in domain specific. Example: *yummy*, *over cooked*, *mouth watering* are some of the sentiment words for restaurant domain. We hand-made a list of words from general intuition and an online site[16] that

[15] http://sentiwordnet.isti.cnr.it/
[16] http://world-food-and-wine.com/describing-food

describes food. We define feature value by taking score 1 for positive, -1 for negative and 2 for those that do not appear in the list. We compute the value based on local contexts (upto previous-5 and next-5) of the aspect term.

4. **Lexical Expansion Score :** We compute expansion score based on our induced lexicons score (described in Sec 3.3) by considering the window size of 10 (preceding 5 and following 5 of the current token).

e.g.

"But the ***staff*** was so horrible to us."

Feature	Value
Aspect term	staff
Local context	NA NA NA But the was so horrible to us
Bing Lexicon	-1,5
SentiWord, MPQA, Domain sp.	0, negative, -1
Sentiment 140 (Uni,Bigram)	-2.398,-3.993
NRC Hashtag (Uni,Bigram)	-3.0289998,-2.475
AFINN, NRC Emotion, Induced Score	-3.0,-1,0.41981673

Table 3.4: An example of feature extraction for sentiment classification

3.8.3 Features for Sentiment Classification - SemEval 2016

For Sentiment classification task for SemEval-2016[17] in English language we use unigram, bigram and score based on above mentioned lexicons as features. In Dutch language we use unigram, bigram and induced lexicon score (described in Sec 3.3) as features.

[17]http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools

CHAPTER 4

Experiments

In this chapter results of various information extraction tasks have been addressed. We use different feature combinations, different fitness functions (for optimization) in different tasks, which have been discussed in detail in the respective section.

4.1 Datasets

Twitter is the most common micro-blogging site on the web, and we use it to gather tweets that express sentiment about popular topics. Public streaming Twitter API is generally used to download tweets.

For experiment with twitter texts we have used benchmark datasets, SemEval-2014[1] and SemEval-2016 [2] shared task, aspect based sentiment analysis. We discuss the statistics of datasets in their respective sections.

4.1.1 SemEval-2014 Dataset

In SemEval-2014, The organizers provided datasets for two domains namely, restaurant and laptop domain in English language only. They provided 3044 and 3045 reviews for training in restaurant and laptop domain respectively. For testing purpose they provided 800 reviews in both the domains. We establish our experimental set up in both the domains. The statistics of training and test set is provided in Table 4.1.

[1]http://alt.qcri.org/semeval2014/task4-/index.php?id=data-and-tools
[2]http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools

Data Set	Restaurant		Laptop	
	Training Set	Test Set	Training Set	Test Set
No. of Reviews	3044	800	3045	800
No. of Aspect terms	3699	1134	2358	654

Table 4.1: Statistics of SemEval-2014 Data

4.1.2 SemEval-2016 Dataset

In SemEval-2016, The organizers provided datasets in multi languages such as English, Dutch, French, Spanish etc. in various domains. We establish our experimental set up in English, Dutch, French, Spanish languages in restaurant domain only. They provided 2000, 1722, 1664, 2070 reviews for training and 676, 575, 667, 881 reviews for testing in English, Dutch, French and Spanish languages respectively. Table 4.2 depicts the statistics of training and test set in all the languages.

Data Set	Training Set		Test Set	
	No. of Reviews	No. of OTE	No. of Reviews	No. of OTE
English	2000	1762	676	612
Dutch	1722	1240	575	373
French	1664	1673	667	650
Spanish	2070	1884	881	713

Table 4.2: Statistics of SemEval-2016 Data

4.2 Performance Measure

We evaluate performance of our system with the help of script provided in SemEval-2014[3] and SemEval-2016[4] shared task. The script results the system's performance in terms of *precision*, *recall* and *F1-measure* for aspect term and OTE extraction and *accuracy* for sentiment classification.

To understand these *functions*, let us take an example:

Consider a document retrieval system which returns 30 documents for a query for

[3]http://alt.qcri.org/semeval2014/task4/

[4]http://alt.qcri.org/semeval2016/task5/

which actually 40 documents are relevant. In those 30 documents, only 20 documents are relevant. Then, the *recall* value of this system is 20/40 and *precision* is 20/30. These *functions* can be mathematically defined as:
precision is fraction of retrieved information that are relevant to the query.

$$precision = \frac{|relevant\ \ information| \cap |retrieved\ \ information|}{|retrieved\ \ information|}$$

Recall is fraction of information that are relevant to the query that are successfully retrieved.

$$recall = \frac{|relevant\ \ information| \cap |retrieved\ \ information|}{|relevant\ \ information|}$$

F1-measure is the measurement of correctness of the information retrieval system. It is actually the harmonic mean of *precision* and *recall*.

$$F1 - measure = \frac{2*recall*precision}{recall+precision}$$

Accuracy is the measurement of correctness of the system. It measures how many instances correctly classified out of total instances.

$$Accuracy = \frac{Correctlyclassifiedinstances}{Totalinstances}$$

4.3 Results and Discussion (SemEval-2014 DataSet)

4.3.1 Aspect Term Extraction Task

We performed various experiments with different feature combinations for both the domains. In the first experiment, we used only standard features. Standards features are all the features which we discussed in Sec 3.8.1 except induced expansion score and unsupervised PoS tag. In the second Experiment, we added DT in the standard features and in the third experiment we added unsupervised PoS tag in whole set of features. The results with different feature combinations are shown in Table 4.3 and 4.4 for restaurant and laptop domain respectively. We can see that after adding non standard features the performance of system is increasing.

Restaurant	Std.Fea	Std.Fea + DT	Std.Fea+DT+UnPOS
No. of Features	68	83	88
F1-measure	78.69	79.54	79.53
Precision	82.29	83.09	83.30
Recall	75.39	76.27	76.10

Table 4.3: Aspect term extraction in restaurant domain (SemEval-2014)

Laptop	Std.Fea	Std.Fea + DT	Std.Fea+DT+UnPOS
No. of Features	69	84	89
F1-measure	71.25	71.59	71.66
Precision	82.79	84.26	83.36
Recall	62.53	62.23	62.84

Table 4.4: Aspect term extraction in laptop domain (SemEval-2014)

As discussed earlier, we also used multi-objective optimization for our experiments. In MOO, we have taken *number of generation* = 20, *population size* = 64 and developed two models:

- Model 1: In first model, we have two fitness functions: *number of features* and *F1-measure* value. Our objective is to maximize the *F1-measure* value and minimize the *number of features*. Here our motive is to develop a good system with much less complexity. The results of model 1 have been provided in Table 4.5 for both the domains.
- Model 2: This model we developed with two fitness functions: *precision* and *recall* value. Our objective is to maximize the value of both the functions *i.e. precision* and *recall* value. Here our motive is to develop a good system without caring about the complexity of the system. Results have been provided in Table 4.6 for both the domains.

In model 1, for restaurant domain before MOO for 88 features we have the *F1-measure* value of 79.53%. After optimization only for 43 features we got the *F1-measure* value of 79.92%. In laptop domain, before MOO for 89 number of features we have the *F1-measure* value 71.66% and after optimization only for 24 number of features we got 70.56% *F1-measure* value. We can notice that with much less set of features we are getting the performance of the system, comparable to the older system.

In model 2, for restaurant domain before MOO we have *precision* value 83.30% and *recall* value 76.10%. After applying MOO, we got 83.78% *precision* and 76.98% *recall*

Data Set	Restaurant		Laptop	
	Before MOO	After MOO	Before MOO	After MOO
No. of Features	88	43	89	24
F1-measure	79.53	79.92	71.66	70.56

Table 4.5: After MOO (Model 1) - Aspect term extraction (SemEval-2014)

value. For laptop domain before MOO, we have 83.36% *precision* and 62.84% *recall* value. After MOO, we got 83.76% *precision* and 64.67% *recall* value respectively. We can see that after applying MOO we are getting increment in both function values. Here, the optimization algorithm is performing well for both the domains.

Data Set	Restaurant		Laptop	
	Before MOO	After MOO	Before MOO	After MOO
Precision	83.30	83.78	83.36	83.76
Recall	76.10	76.98	62.84	64.67

Table 4.6: After MOO (Model 2) - Aspect term extraction (SemEval-2014)

In restaurant domain, total number of gold aspect terms were 1134. Our best system predicted 1042 aspect terms, out of which 873 are correctly predicted. In laptop domain, total number of actual aspect term was 654. Our best system predicted 505 aspect terms, out of which 423 are correctly predicted. This result has been shown in Table 4.7.

Domain	Gold aspect term	System aspect term	Correct aspect term
Restaurant	1134	1042	873
Laptop	654	505	423

Table 4.7: No. of Aspect terms extracted (SemEval-2014)

4.3.2 Sentiment Classification Task

For sentiment classification we performed several experiments. In the first experiment we used only standard features, features that are discussed in Sec 3.8.2 except induced lexicon score. In the second experiment we added induced lexicon score. We established both the experiments for both the domains. The results of these experiments have been provided in Table 4.8. In other experiments, we have also found out the accuracy of each of the class, before and after adding induced lexicon score, which has been shown in Table 4.9.

Data Set	Restaurant		Laptop	
	Std. Fea	Std.Fea+Exp.Score	Std.Fea	Std.Fea+Exp.Score
No. of Features	37	38	37	38
Accuracy	72.48	72.13	61.31	61.01

Table 4.8: Result of sentiment classification before MOO (SemEval-2014)

Data Set	Restaurant		Laptop	
	Std.Fea	Std.Fea+Exp.Score	Std.Fea	Std.Fea+Exp.Score
Positive	88.59	88.73	76.83	76.53
Negative	54.59	54.08	64.84	64.84
Neutral	34.69	32.65	32.54	31.95
Conflict	14.28	14.28	06.25	06.25

Table 4.9: Accuracy of each class before MOO (SemEval-2014)

We used multi-objective optimization in these experiments with *number of generation* = 10 and *population size* = 32 and developed two models:

- Model 1: In first model, we have two fitness functions: *number of features* and *accuracy* of the system. Our objective is to minimize *number of features* and maximize *accuracy* of the system. The intuition behind this experiment that we want to develop a system with less complexity. The result of this model has been shown in Table 4.10 for both the domains.

Data Set	Restaurant		Laptop	
	Before MOO	After MOO	Before MOO	After MOO
No. of Features	38	20	38	12
Accuracy	72.13	76.89	61.01	64.67

Table 4.10: Model 1 : Result of sentiment classification after MOO (SemEval-2014)

- Model 2: In second model, we have four fitness functions : *accuracy* for each the class (positive, negative, neutral and conflict). We have maximized all the four objective functions. Our motive is to develop a system with fine performance without caring about the complexity. The results of model 2 have been shown in Table 4.11.

Data Set	Restaurant		Laptop	
	Before MOO	After MOO	Before MOO	After MOO
Positive	88.73	89.56	76.53	78.29
Negative	54.08	60.20	64.84	70.31
Neutral	32.65	39.79	31.95	40.82
Conflict	14.28	0.00	06.25	12.50

Table 4.11: Model 2 : Accuracy of each class after MOO (SemEval-2014)

In first model, for restaurant domain before MOO for 38 features, our system results in 72.13% accuracy . But, after optimization only for 20 features, we got the accuracy of 76.89%. In laptop domain, for 38 number of features our system results in 61.01% accuracy. But after MOO only for 12 features we got the accuracy of 64.67%. We can notice that we are getting 4.76% increment in restaurant domain and 3.66% increment in laptop with much less number of features. Here again the optimization algorithm performs very well and proves its usability.

For second model, we can see the result in Table 4.11. We can notice that after applying MOO we are getting increment in accuracy for each of the class except the 'conflict' class in restaurant domain.

4.4 Results and Discussion (SemEval-2016 DataSet)

4.4.1 Opinion Target Expression (OTE) Extraction Task

In SemEval-2016[5], the organizers provided the datasets in multi-language to extract the OTEs. We establish our experimental set up in four languages namely English, Spanish, French and Dutch. We performed the same experiments which we have discussed in Section 4.3.1. Results with different features combination for English, Spanish, French and Dutch have been shown in Table 4.12, 4.13, 4.14 and 4.15 respectively.

We applied the MOO and developed the same two models with *number of generation* = 20 and *population size* = 64, which we discussed in Section 4.3.1 with same motive. The results of model 1 for all the four languages are shown in Table 4.16,

[5]http://alt.qcri.org/semeval2016/task5/

English	Std.Fea	Std.Fea + DT	Std.Fea+DT+UnPOS
No. of Features	63	78	83
F1-measure	66.90	67.98	68.45
Precision	74.40	75.60	75.54
Recall	60.78	61.76	62.58

Table 4.12: OTE in English language (SemEval-2016)

Spanish	Std.Fea	Std.Fea + DT	Std.Fea+DT+UnPOS
No. of Features	47	62	67
F1-measure	68.90	68.72	69.73
Precision	71.84	76.23	76.85
Recall	66.19	62.55	63.81

Table 4.13: OTE in Spanish language (SemEval-2016)

French	Std.Fea	Std.Fea + DT	Std.Fea+DT+UnPOS
No. of Features	47	62	67
F1-measure	68.97	70.18	69.64
Precision	70.64	71.81	70.85
Recall	67.38	68.61	68.46

Table 4.14: OTE in French language (SemEval-2016)

Dutch	Std.Fea	Std.Fea + DT	Std.Fea+DT+UnPOS
No. of Features	47	62	67
F1-measure	63.37	62.63	64.37
Precision	64.88	64.22	66.10
Recall	61.93	61.12	62.73

Table 4.15: OTE in Dutch language (SemEval-2016)

where we have minimized the *number of features* and maximized *F1-measure* value. The results of model 2 for all the four languages are shown in Table 4.17, where we have maximized both *precision* and *recall* value.

	English	Spanish	French	Dutch
No. of Features	24	26	11	15
F1-measure	68.21	70.33	67.85	65.01

Table 4.16: Model 1 : After MOO - OTE (SemEval-2016)

	English	Spanish	French	Dutch
Precision	75.84	74.28	70.91	66.57
Recall	62.09	69.28	66.76	62.46

Table 4.17: Model 2 : After MOO - OTE (SemEval-2016)

Results with total number of gold OTEs, our best system predicted OTEs and total number of common OTEs for all the four languages are shown in Table 4.18.

Language	Gold aspect term	System aspect term	Correct aspect term
English	612	507	383
Spanish	713	665	494
French	650	621	446
Dutch	373	353	236

Table 4.18: No. of OTEs extracted (SemEval-2016)

4.4.2 Sentiment Classification Task

For SemEval-2016 data, we have done sentiment classification in only two languages, English and Dutch. For both the languages, we have performed several experiments with different features combination which have been shown in Table 4.19. In other experiment we computed the accuracy for each of the class, which have been shown in Table 4.20. Since, the number of features for this task in both the languages are too less, we didn't apply MOO.

Data Set	English		Dutch	
	Std. Fea	Std.Fea+Exp.Score	Uni+Bi-gram	Uni+Bi+Exp.Score
Cross Validation	79.20	79.41	74.40	74.46
Accuracy	81.83	81.84	73.73	74.87

Table 4.19: Result of sentiment classification (SemEval-2016)

Data Set	English		Dutch	
	Std. Fea	Std.Fea+Exp.Score	Uni+Bi-gram	Uni+Bi+Exp.Score
Positive	88.70	88.54	82.92	83.73
Negative	76.47	76.96	68.72	70.61
Neutral	11.36	11.36	03.03	03.03

Table 4.20: Result of sentiment classification (SemEval-2016)

4.5 Comparative Analysis

As explained in previous section, we performed experiments on SemEval-2014 [6] and SemEval-2016 [7] datasets. Many teams participated in these tasks. We did the comparison of performance of our system with those systems. Our performance is satisfactory. Comparison of aspect term and OTE extraction tasks are shown in Table 4.21. We are at third position in English datasets among all the participants. In non-English language we are at first position among all the teams. In these languages, we are 3.18%, 4.87% and 13.24% ahead in Spanish, French and Dutch respectively than the other top team who participated in the challenge, which is significantly very high. These comparisons show that our technique is very good for non-English language. Comparison of sentiment classification task is shown in Table 4.22. We did well in this task also.

4.6 Feature Selection

The performance of any classification problem fully depends on the features, which we are using for the datasets. Feature selection is an approach, widely used in machine learning for selecting the most relevant set of features for building the robust training models. It is also known as attribute selection, variable selection, feature reduction,

[6] http://alt.qcri.org/semeval2014/task4-/index.php?id=data-and-tools
[7] http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools

Datasets	Rank	Diff. from top ranked team
Restaurant (SemEval-2014)	3/28	3.78
Laptop (SemEval-2014)	3/27	1.56
English (SemEval-2016)	3/19	3.89
Spanish (SemEval-2016)	1/4	-
French (SemEval-2016)	1/2	-
Dutch (SemEval-2016)	1/2	-

Table 4.21: Comparison of aspect-term and OTE extraction task

Datasets	Rank	Diff. from top ranked team
Restaurant (SemEval-2014)	5/31	4.06
Laptop (SemEval-2014)	6/31	5.04
English (SemEval-2016)	8/27	6.28
Dutch (SemEval-2016)	4/4	2.94

Table 4.22: Comparison of sentiment classification task

variable subset selection or dimensionality reduction. Feature selection is one of the well known optimization problem in machine learning which selects an appropriate subset of features. It helps in improving the performance of a classifier by removing the most irrelevant set of features. We have shown feature selection for SemEval-2014 aspect based sentiment analysis task in table 4.21. Feature selection for both the models have been shown. *Model 1 (M1)* and *Model 2 (M2)* are the same model that we have discussed in Section 4.3.1. Feature selection for SemEval-2016 opinion target expression extraction task for both the models have been shown in Table 4.22. Feature selection for SemEval-2014 sentiment classification task has been shown in Table 4.23.

Features	Restaurant		Laptop	
	Model 1	Model 2	Model 1	Model 2
Word and Context	-3,-1..3	-3,-1..1,3	-1,0,3	2
Char n-gram(2,3,4-gram)	2,3	4	2,4	✓
Chunk information(-1,0,1)	0,1	-1,1	✓	✓
Dependency		✓		✓
Dependency (Head)	✓	✓		✓
Dependency (Modifier)	✓	✓	✓	✓
Ortho (digit)	-	-	✓	✓
Distri. Thesaurus(-2,-1,0,1,2)	-2..0	-2,0,2	0	-2,0
Frq. Aspct List1		✓		
Freq. Aspct List2	✓	✓	✓	✓
Head Word				
Head Word Token			-	-
Head Word Token|POS			-	-
Head Word POS		✓	✓	✓
Head Word|POS	-	-		✓
Lemma	-	-		
Lower Case		✓		
NER (-1,0,1)	0		-1	-1,0
Ortho		✓	✓	
PMI (-1,0,1)		✓	0,1	0,1
Rounded PMI (-1,0,1)	-1	0	1	-1,1
POS (-2,-1,0,1,2)	-2..0	-2,0	-2,1	-2..1
1-Char Prefix(-1,0,1)	1	-1,1		0
2-Char Prefix(-1,0,1)		-1		
3-Char Prefix(-1,0,1)		1	-1	0
Expansion score(-2,-1,0,1,2)	✓	✓	2	-1,1
Round. Exp.Score(-2,-1,0,1,2)		0,2	1	0,2
SentiWord(-2,-1,0,1,2)	0,2	-2..0,2		-1,2
Round. SentiWord(-2,-1,0,1,2)	-2,2	0,2	-1	
StopWord	✓	✓		
1-Char Suffix(-1,0,1)	✓	✓		0
2-Char Suffix(-1,0,1)	-1,0	-1		
3-Char Suffix(-1,0,1)	✓	-1,0		-1,0
Noun Synset	✓	✓		✓
Unsupervised POS(-2,-1,0,1,2)	-1,2	-1..2		2

Table 4.23: Feature selection of aspect term Extraction - 2014

Language	Obj. Function	Word and Context	Char n-gram(2,3,4-gram)	Chunk information(-1,0,1)	Dependency (Head)	Dependency (Modifier)	Distri. Thesaurus(-2,-1,0,1,2)	Frq. Aspct List1	Freq. Aspct List2	Head Word	Head Word POS	Lemma	Lower Case	NER	Ortho	POS (-2,-1,0,1,2)	1-Char Prefix(-1,0,1)	2-Char Prefix(-1,0,1)	3-Char Prefix(-1,0,1)	4-Char Prefix	Expansion score(-2,-1,0,1,2)	Round. Exp.Score(-2,-1,0,1,2)	SentiWord(-2,-1,0,1,2)	Round. SentiWord(-2,-1,0,1,2)	1-Char Suffix(-1,0,1)	2-Char Suffix(-1,0,1)	3-Char Suffix(-1,0,1)	4-Char Suffix	Noun Synset	Unsupervised POS(-2,-1,0,1,2)
Eng	M1	0,3,4	2..4		✓		-2..1						✓		✓	-2,0,2	0,1		-1			-1,1,2		2	-1					
	M2	-4..-2,1,2,4,5	4	0,-1	✓		-2..0,2	✓				✓		✓	✓	-1..2	0,1	✓	0,1	✓	-1,1,-2	-1,1	1	-2	✓	0	✓		✓	-2,0,2
Dutch	M1	-1,1		-			-1,0					-		-		-2,-1			-1,0	✓	-2	-2	-	-		-1	1		-	-1,0
	M2	-4,-2,-1,1		-	✓		-2..0			✓	✓	-	✓	-	✓	-2,0,1		0,1	1	✓	-2,-1,1	-2	-	-	0,1	✓	1		-	-1,0
Spani	M1	-5,-1,0	2,3	-		✓	-1,0	✓	✓			-		-	✓	-2..0	0	1	-1	✓	-1	-2,1	-	-	0	1	-1		-	-1,0
	M2	-4,-3,5	✓	-			-1		✓			-	✓	-	✓	-2,-1	0,1	-1	-1,0	✓	-1,0	-2,-1,1	-	-	0,1	1	0		-	-1..1
Frenc	M1	-3,-1		-			-2,0					-		-		-2,-1	1	0,1				2	-	-		-1			-	
	M2	-5,-3..0,2	2,3	-		✓	-1,0		✓	✓	✓	-		-		-2,-1,1	✓	0	-1,0		-2,1,2	-2..0	-	-	-1	-1		✓	-	-1

Table 4.24: Feature selection of OTE - SemEval 2016

Features	Restaurant		Laptop	
	Model 1	Model 2	Model 1	Model 2
Word and Context	-4..3	-4,-2..1	-4,-2..3,5	-5,-2..4
Bi-gram	2		2	-2,2
Bing Lexicon	✓	✓	✓	✓
Bing Direct Lexicon	✓	✓	✓	✓
SentiWord	✓	✓		✓
PMI	✓	✓		✓
MPQA	✓	✓	✓	✓
Domain Specific Word	✓	✓	-	-
Sentiment-140 lexicon (Unigram,Bigram)	✓	Unigram		
NRC Hashtag lexicon (Unigram,Bigram)	✓	Bigram		
AFINN lexicon				
NRC Emotion	✓		✓	
Expansion Score		✓		

Table 4.25: Feature selection of sentiment classification - 2014

CHAPTER 5

Conclusion and Outlook

In the present work, we propose aspect term extraction, opinion target expression detection and sentiment classification systems by incorporating the technique of unsupervised lexical acquisition and multi-objective optimization. We assume that using features from unsupervised lexical acquisition resources and instead of selecting the features heuristically to train a classifier, feature selection using MOO can be an alternative way to deal with scarcity of training data for resource poor or minority language.

We made several experiments for both the tasks and came up with the conclusion that features from unsupervised lexical acquisition technique are very useful specially for resource poor languages (in our case Spanish, French and Dutch). Feature selection using MOO is more fruitful approach, even with very less number of features we are getting the performance of the system comparable to the earlier system. By applying MOO, we are getting very good increment in performance of the system in sentiment classification task, while in aspect term extraction task, we are getting better results in four out of six datasets. Other interesting and important characteristic of proposed system is that it makes use of mostly language-independent features that can be easily derived for almost all the languages without any knowledge of them a priori. Thus, a variety of languages are benefited by our proposed technique.

In future we would like to include more language independent features in our available feature set from various existing resources and tools. Instead of selecting only one best feature set by a classification system using MOO algorithm, ensembling several classification systems where each of them is developed using different classification technique and/or different feature sets can be explored.

REFERENCES

[1] K. Dave, S. Lawrence and D. M. Pennock. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. In *Proceedings of the 12th International Conference on World Wide Web*, Budapest, Hungary, pages 519-528, 2003.

[2] M. Hu and B. Liu: Mining and summarizing customer reviews. In *Proceedings of the 10th KDD*, Seattle, USA, pages 168-177, 2004.

[3] B. Liu: Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers, Volume 5, pages 1-167, 2012.

[4] Y. Liu, X. Huang, A. An and X. Yu. ARSA: a sentimentaware model for predicting sales performance using blogs. In *Proceedings of ACM SIGIR Confrence on Research and Development in Information Retrieval (SIGIR-2007)*, Amsterdam, Netherlands, pages 607-614, 2007.

[5] Y. Hong and S. Skiena. The Wisdom of Bookies? Sentiment Analysis vs. the NFL Point Spread. In *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM-2010)*, Washington, DC, pages 251-254, 2010.

[6] M. McGlohon, N. Glance and Z. Reiter. Star Quality: Aggregating Reviews to Rank Products and Merchants. In *Proceedings of 4th International AAAI Conference on Weblogs and Social Media*, Washington, pages 114-121, 2010.

[7] B. O'Connor, R. Balasubramanyan, B. R. Routledge and N. A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of 4th International AAAI Conference on Weblogs and Social Media*, Washington, pages 122-129, 2010.

[8] S. Brody and N. Elhadad: An unsupervised aspect-sentiment model for online reviews. In *Proceedings of NAACL*, Los Angeles, US, pages 804-812, 2010.

[9] A. Ekbal and S. Saha. A multiobjective simulated annealing approach for classifier ensemble: Named entity recognition in indian languages as case studies. *Expert Syst. Appl. (ISSN: 0957-4174)*, Volume 38(12), pages 14760-14772, 2011.

[10] P. D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of 40th Annual Meeting Computational Linguistics*, Philadelphia, pages 417-424, 2002.

[11] S. M. Kim and E. Hovy. Determining the sentiment of opinions. In *Proceedings of the 20th international conference on Computational Linguistics*, Geneva, pages 1367-1373, 2004.

[12] A. M. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Vancouver, pages 339-346, 2005.

[13] B. Pang, L. Lee and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Stroudsburg, PA, USA, pages 79-86, 2002.

[14] T. Nasukawa and J. Yi. Sentiment analysis: capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, New York, NY, USA, pages 70-77, 2003.

[15] B. Ohana and B. Tierney. Sentiment classification of reviews using sentiwordnet. In *9th IT&T Conference*, Dublin, Ireland, 2009.

[16] H. Saggion and A. Funk. Interpreting sentiwordnet for opinion classification. In *7th International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta, 2010.

[17] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato and V. S. Subrahmanian. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, USA, 2007.

[18] K. Bloom, N. Garg and S. Argamon. Extracting appraisal expressions. In *Proceedings of Human Language Technologies/North American Association of Computational Linguists*, Rochester, pages 308-315, 2007.

[19] A. Agarwal, F. Biadsy and K. R. Mckeown. Contextual phrase-level polarity analysis using lexical affect scoring and syntactic n-grams. In *Proceedings of the 12th Conference of the European Chapter of the ACL 2009*, Athens, Greece, pages 24-32, 2009.

[20] T. Wilson, J. Wiebe and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of HLT-EMNLP*, Vancouver, pages 347-354, 2005.

[21] J. M. Wiebe, R. F. Bruce and T. P. O'Hara. Development and use of a gold-standard data set for subjectivity classifications. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics (ACL '99)*, Stroudsburg, PA, USA, pages 246-253, 1999.

[22] H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the conference on Empirical methods in natural language processing (EMNLP '03)*, Stroudsburg, PA, USA, pages 129-136, 2003.

[23] S. B. Goldensohn, K. Hannan, R. McDonald, T. Neylon, G. A. Reis, and J. Reynar. Building a sentiment summarizer for local service reviews. In *Proceedings of WWW2008 workshop on NLP in the Information Explosion Era*, Beijing, China, pages 14-23, 2008.

[24] L. Zhuang, F. Jing, and X. Zhu. Movie review mining and summarization. In *Proceedings of ACM International Conference on Information and Knowledge Management (CIKM06)*, Arlington, Virginia, USA, pages 43-50, 2006.

[25] A. Mukherjee and B. Liu. Aspect Extraction through Semi-Supervised Modeling. In *Proceedings of 50th Anunal Meeting of Association for Computational Linguistics (ACL-12)*, Jeju Island, Korea, pages 339-348, 2012.

[26] Y. Wu, Q. Zhang, X. Huang and L. Wu. Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Singapore, pages 1533-1541, 2009.

[27] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos and S. Manandhar. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), at the 25th International Conference on Computational Linguistics (COLING 2014)*, Dublin, Ireland, pages 27-35, 2014.

[28] K. Deb. Multi-objective optimization using evolutionary algorithms. *John Wiley and Sons, Ltd*, England, 2001.

[29] G. A. Miller. Wordnet: a lexical database for english. *Communications of the ACM (ISSN: 0001-0782)*, Volume 38(11), pages 39-41, 1995.

[30] B. Liu. Web data mining: exploring hyperlinks, contents, and usage data. *Springer-Verlag New York, Inc.*, 2006.

[31] A. Fahrni and M. Klenner. Old wine or warm beer: Target-specific sentiment analysis of adjectives. In *Proceedings of the Symposium on Affective Language in Human and Machine (AISB 2008)*, Aberdeen, Scotland, pages 60-63, 2008.

[32] J. Wiebe and R. Mihalcea. Word sense and subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, Sydney, Australia, pages 1065-1072, 2006.

[33] S. Moghaddam and M. Ester. Aspect-based opinion mining from product reviews. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Informat*, Portland, Oregon, USA, page 1184, 2012.

[34] J. D. Lafferty, A. McCallum and F. C. N. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the 18th International Conference on Machine Learning*, San Francisco, CA, USA, pages 282-289, 2001.

[35] V. Hatzivassiloglou and K. R. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the 35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, Madrid, Spain, pages 174-181, 1997.

[36] C. Biemann. Unsupervised part-of-speech tagging in the large. *Research on Language and Computation (ISSN: 1570-7075)*, Volume 7(2-4), pages 101-135, 2009.

[37] C. Biemann and M. Riedl. From global to local similarities: A graph-based contextualization method using distributional thesauri. In *Proceedings of the 8th Workshop on TextGraphs in conjunction with EMNLP*, Seattle, USA, pages 39-43, 2013.

[38] Z. S. Harris. Methods in structural linguistics. *Wiley Online Library*, Volume 54(3), 2009.

[39] J. C. Platt. Fast Training of Support Vector Machines Using Sequential Minimal Optimization. *Advances in Kernel Methods, MIT Press*, pages 185-208, 1999.

[40] A. Kumar, S. Kohail, A. Ekbal and C. Biemann. IIT-TUDA: System for Sentiment Analysis in Indian Languages using Lexical Acquisition. *Third International Conference, MIKE 2015*, Hyderabad, India, pages 684-693, 2015.

[41] C. J. Ye and M. X. Huang. Multi-Objective Optimal Power Flow Considering Transient Stability Based on Parallel NSGA-II. *IEEE Transaction on power system*, Volume 30(2), pages 857-866,2015.

[42] M. Caramia and P. D. Olmo. Multi-objective Management in Freight Logistics Increasing Capacity, Service Level and Safety with Optimization Algorithms. *London:Springer-Verlag*, 2008.

[43] Govind, A. Ekbal and C. Biemann. Multiobjective optimization and unsupervised lexical acquisition for named entity recognition and classification. In *Proceedings the 11th International Conference on Natural Language Processing (ICON)*, Goa, India, 2014.

[44] M. Salameh, S. Mohammad and S. Kiritchenko. Sentiment after translation: A case-study on arabic social media posts. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Denver, Colorado, pages 767-777, 2015.

[45] X. Ding, B. Liu and P.S. Yu. A holistic lexicon-based approach to opinion mining.In *Proceedings of the International Conference on Web Search and Data Mining*, WSDM 2008, Newyork, pages 231-240, 2008.

[46] S. M Mohammad and P. D. Turney. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, Volume 29(3), pages 436-465, 2013.

[47] F. Nielsen. A new anew: Evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, Heraklion, Greece, pages 93-98, 2011.

[48] X. Zhu, S. Kiritchenko and S. M. Mohammad. Nrc-canada-2014: Recent improvements in the sentiment analysis of tweets. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin, Ireland, pages 443-447, 2014.

[49] https://en.wikipedia.org/wiki/Sentiment_analysis.