

# **Sentiment Mining Using Machine Learning Optimization**

A Thesis

Presented to

The Academic Faculty

by

**MOHAMMED ARIF KHAN**

In Partial Fulfilment

of the Requirements for the M.Tech Degree of



**DEPARTMENT OF MATHEMATICS  
INDIAN INSTITUTE OF TECHNOLOGY PATNA.**

**APRIL 2015**

**Copyright © Mohammed Arif Khan2015**

Dedicated to,  
My Mother  
**Smt. Bismillah Bano**  
My Father  
**Shri. Nathu Mohammed**  
&  
All Teachers  
who have taught me

## **Acknowledgements**

For the initiation, supervision and all insightful discussions related to this thesis work, I would foremost like to convey my sincerest gratitude towards my supervisors Dr. Asif Ekbal, Assistant Professor, Department of Computer Science and Engineering, at Indian Institute of Technology Patna and Prof. Dr. Johannes Furnkranz, Professor, Knowledge Engineering Group, Department of Informatics at Technische Universitat Darmstadt, Germany. Without their constant guidance the completion of this project would not have been possible. I would also like to convey my sincerest thanks to Eneldo Loza Mencia, Post Doctoral Fellow, Knowledge Engineering Group, Departments of Informatics at Technische Universitat Darmstadt, Germany.

I would further like to acknowledge the help extended by the staff of the KE group, staff of Computer Science Department at IIT Patna and the staff of International Office at TU Darmstadt for making my stay comfortable at Darmstadt.

I would be failing in my duty if I do not thank German Academic Exchange Service (DAAD) for providing fellowship throughout my stay in Germany.

I would also like to mention the constant support received from my family, who have been a guiding light for me throughout my life. In the end I cannot express the deep indebtedness of all the knowledge that I have gained from all my previous teachers at various levels of my education.

–Mohammed Arif Khan

**Name of the MTech Specialization:** MATHEMATICS AND COMPUTING

**Name of the Student:** MOHAMMED ARIF KHAN

**Title of the Thesis:** Sentiment Mining Using Machine Learning Optimization

**Date:** .....

### DECLARATION

It is hereby certified that

1. the work contained in this thesis has not been submitted to any other institute for any degree or diploma.
2. the thesis has been written as per guidelines provided by the Institute.
3. ethical norms and guidelines have been conformed while writing this thesis.
4. whenever any material (data, models, figures, and text) from other sources have been used, due credit to them has been provided by citing them in text of the thesis, giving their details in references, and taking permission from copyright owners of the sources, whenever necessary.
5. the thesis has been checked by an anti-plagiarism tool and the report has been attached with this thesis.

.....  
Signature of the Student

.....  
Signature of Supervisor(s)

## THESIS CERTIFICATE

This is to certify that the thesis titled **Sentiment Mining Using Machine Learning Optimization**, submitted by **Mohammed Arif Khan**, to the Indian Institute of Technology, Patna, for the award of the degree of **Master of Technology**, is a bonafide record of the research work done by him under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

**Prof. Dr. Johannes Furnkranz**  
Research Guide  
Professor  
KE Group, Dept. of Informatics  
TU Darmstadt Germany, D-64289

**Dr. Asif Ekbal**  
Research Guide  
Assistant Professor  
Dept. of Computer Science & Eng.  
IIT-Patna, 800 013

Place: Patna

Date: April, 2015

## **Abstract**

In recent times, there has been a huge interest to mine and understand the sentiments that people are communicating in social media. In this thesis, sentiment analysis has approached through supervised machine learning for both multi-class and multi-label classification. Feature and parameter combinations have significant effect to the performance of the classifier for any machine learning technique. For multi-class classification, we used Random Forest and Support Vector Machines (i.e. SMO, LibSVM and LibLinear) for solving the problem of feature selection and problem of simultaneous feature selection and parameter optimization. Multi-label learners, namely Label Powerset, Binary Relevance, Classifier Chains and Calibrated Label Ranking each with decision trees (J48) and Support Vector Machine (LibSVM) as base learner were used for multi-label sentiment classification. We used SemEval-14 data sets for multi-class sentiment analysis and two different data sets (Emotion and CAL500) were used for multi-label sentiment analysis. Evaluation results show that the proposed approach attains the 17 % improvement in F-measure in case of feature selection and 20.10 % improvement in F-measure for simultaneous feature selection and parameter optimization. In all the experiments, the optimal feature subset obtained through an optimization technique (NSGA-II) performs reasonably better than the baseline systems, developed with random feature subsets.

# TABLE OF CONTENTS

<b>LIST OF TABLES</b>	<b>viii</b>
<b>LIST OF FIGURES</b>	<b>ix</b>
<b>ABBREVIATIONS</b>	<b>x</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Sentiment Analysis . . . . .	2
1.1.1 Multi-class Sentiment Analysis . . . . .	3
1.1.2 Multi-label Sentiment Analysis . . . . .	3
1.2 Multi-objective Optimization . . . . .	4
1.3 Classifiers Ensemble . . . . .	4
1.4 Organization of The Thesis . . . . .	5
<b>2 Review of Prior Works</b>	<b>6</b>
2.1 Sentiment Analysis . . . . .	6
2.2 Multi-label Classification . . . . .	6
2.3 Feature Selection . . . . .	7
<b>3 Methodology and Approach</b>	<b>9</b>
3.1 Framework for Multi-objective Optimization for Feature Selection . . . . .	9
3.1.1 Multi-objective Formulation of Feature Selection Problem . . . . .	9
3.1.2 Selection of Objectives . . . . .	9
3.1.3 Chromosome Representation and Population Initialization . . . . .	10
3.1.4 Fitness Computation . . . . .	10
3.1.5 Other Operators . . . . .	11
3.1.6 Selection of Final Solution . . . . .	11
3.2 Framework for Multi-objective Optimization for Simultaneous Feature Selection and Parameter Optimization . . . . .	12
3.3 Multi-class Sentiment Analysis . . . . .	12

3.3.1	Features used . . . . .	13
3.3.2	Framework for Multi-class Sentiment Analysis . . . . .	18
3.4	Multi-label Sentiment Analysis . . . . .	19
3.4.1	Features used . . . . .	19
3.4.2	Framework for Multi-label Sentiment Analysis . . . . .	19
<b>4</b>	<b>Experiments and Result Discussion</b>	<b>21</b>
4.1	Experiments for Multi-class Sentiment Analysis . . . . .	21
4.1.1	Data sets . . . . .	21
4.1.2	Feature Selection using Multi-objective Optimization . . . . .	21
4.1.3	Simultaneous Feature Selection and Parameter Optimization using Multi-objective Optimization . . . . .	25
4.2	Experiments for Multi-label Sentiment Analysis . . . . .	28
4.2.1	Data sets . . . . .	28
4.2.2	Feature Selection using Multi-objective Optimization for Emotion Data . . . . .	28
4.2.3	Feature Selection using Multi-objective Optimization for CAL500 Data . . . . .	31
<b>5</b>	<b>Conclusion and Future Work</b>	<b>39</b>
	<b>References</b>	<b>42</b>



## LIST OF TABLES

1.1	Illustrative example: Multi-class Classification . . . . .	3
1.2	Illustrative example: Multi-label Classification . . . . .	3
3.1	Illustrative example: Unigram Computation . . . . .	15
3.2	Illustrative example: Bigram Computation . . . . .	15
3.3	POS Tags . . . . .	17
4.1	Multi-class Data Sets . . . . .	21
4.2	Feature Sets of S28 . . . . .	22
4.3	Results for Exp-PR1 . . . . .	23
4.4	Confusion Matrix-PR . . . . .	24
4.5	Results for Exp-ANF1 . . . . .	24
4.6	Confusion Matrix-ANF . . . . .	25
4.7	Feature Sets of S55 . . . . .	35
4.8	Results for Exp-PR2 . . . . .	36
4.9	Results for Exp-ANF2 . . . . .	37
4.10	Optimized Parameters . . . . .	37
4.11	Multi-label Data Sets . . . . .	37
4.12	Results with Emotion Data . . . . .	38
4.13	Results with CAL500 Data . . . . .	38

## LIST OF FIGURES

1.1	Growth of data generated by Twitter . . . . .	2
3.1	Chromosome representation for NSGA-II based feature selection . .	11
3.2	Chromosome representation for NSGA-II based simultaneous feature selection and parameter optimization . . . . .	12
3.3	System Architecture . . . . .	20
4.1	Population representation for Emotion data with J48 . . . . .	29
4.2	Population representation for Emotion data with LibSVM . . . . .	30
4.3	Population representation for CAL500 data with J48 . . . . .	32
4.4	Population representation for CAL500 data with LibSVM . . . . .	33

## ABBREVIATIONS

<b>BLL</b>	Bing Liu's Lexicon
<b>BR</b>	Binary Relevance
<b>CC</b>	Classifier Chains
<b>CLR</b>	Calibrated Label Ranking
<b>Exp-ANF1</b>	Experiment with accuracy and number of features as objective functions for S28
<b>Exp-ANF2</b>	Experiment with accuracy and number of features as objective functions for S55
<b>Exp-PR1</b>	Experiment with precision and recall as objective functions for S28
<b>Exp-PR2</b>	Experiment with precision and recall as objective functions for S55
<b>FS</b>	Feature Selection
<b>LP</b>	Label Powerset
<b>MOO</b>	Multi-Objective Optimization
<b>NLP</b>	Natural Language Processing
<b>NSGA-II</b>	Non dominated Sorting Genetic Algorithm II
<b>NSP</b>	Ngram Statistics Package
<b>PO</b>	Parameter Optimization
<b>POS</b>	Part-of-Speech
<b>SMO</b>	Sequential Minimal Optimization
<b>SVM</b>	Support Vector Machine
<b>SWN</b>	Senti Word Net

# CHAPTER 1

## Introduction

Social media has grown enormously over the last decade. Due to extensive use of social media like Twitter, it generates huge data per day (see Figure 1.1 <sup>1</sup>). People's opinion on certain things are always important but finding insights manually from these huge data is almost impossible so there is an obvious need of automatic systems that can find the important hidden insight from these large data sets. In this thesis, we addressed this problem through sentiment analysis and our contributions are summarized as below:

- To the best of our knowledge, this work is the first to propose to use multi-objective optimization for both feature selection and simultaneous feature selection and parameter optimization in the domain of sentiment analysis.
- We show that our multi-objective optimization based approach for feature selection out perform compare to baseline models. We performed experiments for feature selection using three different classifiers, namely, Sequential Minimal Optimization (SMO), LibSVM and LibLinear for multi-class classification on SemEval-14 data sets; and using four multi-label learner, namely Label Powerset, Binary Relevance, Classifier Chains and Calibrated Label Ranking each with J48 and LibSVM as base learner for multi-label sentiment classification on two different data sets (Emotion and CAL500).
- We proved that along with feature selection, parameter optimization is also contributes towards better performance for a classifier. We performed experiments for simultaneous feature selection & parameter optimization for three different classifiers, namely, Random Forest, LibSVM and LibLinear for multi-class classification on SemEval-14 data sets

---

<sup>1</sup><http://www.theguardian.com/technology/blog/2010/mar/05/twitter-ten-billionth-tweet>

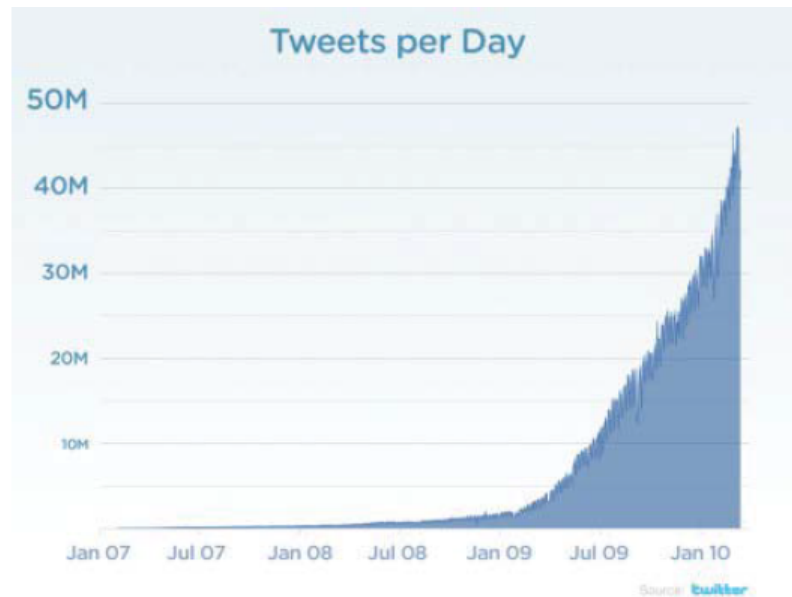


Figure 1.1: Growth of data generated by Twitter

## 1.1 Sentiment Analysis

Sentiment Analysis is the process of detecting the polarity of text. In other words, it determines whether a piece of writing is positive, negative or neutral. It is also known as Opinion Mining. Bing Liu [1] defines it in following way:

“A Sentiment (or opinion) is a quadruple, (g, s, h, t), where g is the sentiment (or opinion) target, s is the sentiment about the target, h is the opinion holder and t is the time when the opinion was expressed.”

Sentiment Analysis is used to determine sentiment on a variety of levels like entire document, a phrase or an individual word. It is one of the major tasks in Natural Language Processing (NLP) which is known by various names such as sentiment analysis, opinion mining, opinion extraction, subjectivity analysis, affect analysis, emotion analysis, review mining and sentiment mining etc’.

Sentiment analysis, a multi-disciplinary area covering natural language processing, data mining and machine learning, is a task that extract emotions from texts. This is an active research area and has been used with different applications such as financial prediction, evaluating customer feedbacks and understanding opinions etc. Machine learning and Lexicon-based approaches are two common approaches for sentiment classification. In this thesis we have considered two types of sentiment analysis as follows:

### 1.1.1 Multi-class Sentiment Analysis

Considering classification problem if a document belongs to one class and total number of classes are more than one then this type of classification is known as multi-class classification. Here we extracted sentiments from twitter data so we can term this type of classification as multi-class sentiment analysis. An example of Multi-class Sentiment Analysis is shown in Table 1.1

Table 1.1: Illustrative example: Multi-class Classification

S.N.	Documents (Tweets)	Class
1.	Apple CEO Tim Cook apologized for the error-ridden map.	Negative
2.	I love Tom Boonen. Forget the whole birth/citizen issue.	Positive
3.	Tryna make sure everything go as planned Saturday.	Neutral

### 1.1.2 Multi-label Sentiment Analysis

Multi-label classification is concerned with categorizing instances into multiple classes (labels), while the associated classes are not exclusive. Each associated class of an instance is called a label. Many real-world classification problems involve multiple labels. In multi-class classification, each sample belongs to one label only; whereas in multi-label classification, each sample may associate with multiple labels. For example, in text categorization, a document can belong to the categories of piracy, copyright and software. Similarly, in bioinformatics, a gene may be associated with the functions of transcription, metabolism and protein synthesis. Identifying sentiment from multi-label data is termed as Multi-label Sentiment Analysis. An example of Multi-label classification is shown in Table 1.2.

Table 1.2: Illustrative example: Multi-label Classification

S.N.	Documents (Tweets)	Labels
1.	Lacey teen dies after house fire: LACEY, Wash. A 17-year-old boy critically injured after a house fire in Lacey.	Injury Fire
2.	Shelton woman shot, in critical condition.	Injury Shooting
3.	Something tells me that Adam crashed and burned after that press conference.	Injury Fire Crash

## 1.2 Multi-objective Optimization

Simultaneously optimizing more than one objective is known as Multi-objective Optimization(MOO) [2]. In general the objectives have trade-off between two or more conflicting objectives. For example minimizing cost and maximizing comfort while buying a car. In practical problems, there can be more than two objectives. In mathematical terms, a multi-objective optimization problem can be formally stated as:

Find the vector of decision variables

$$x = [x_1, x_2, \dots, x_n]^T \quad (1.1)$$

which will satisfy the m inequality constraints:

$$g_i(x) \geq 0, i = 1, 2, \dots, m. \quad (1.2)$$

and the p equality constraints

$$h_j(x) = 0, j = 1, 2, \dots, p. \quad (1.3)$$

and simultaneously optimizes M objective functions

$$f_1(x), f_2(x), \dots, f_M(x). \quad (1.4)$$

Best solution(s) are selected based upon the Pareto optimal concept. A candidate is Pareto optimal iff:

- (i) It is atleast as good as all other candidates for all objectives, and
- (ii) It is better than all other candidates for atleast one objective.

## 1.3 Classifiers Ensemble

In ensemble method [3], we construct a set of classifiers by varying the training data, classifiers and parameters of the classifiers etc and then it is used to predict class label of previously unseen records by combining predictions made by multiple classifiers.

An ensemble itself is a supervised learning algorithm, because it can be trained and then used to make predictions. The trained ensemble represents a single hypothesis. Generally, ensembles provide better results if models have significant diversity.

## **1.4 Organization of The Thesis**

This chapter provides a background for the topics covered in this thesis. We provided a description of Sentiment Analysis, Multi-objective optimization and classifier ensemble process. Rest of the chapters are organised as follows: in next chapter we provide review of prior works related to sentiment analysis, multi-label classification and feature selection. In Chapter 3, we discuss the methodologies including the features sets, framework for sentiment analysis and framework for MOO for both multi-class and multi-label sentiment analysis. And Chapter 4 reports the details of data sets we used for the experiments and evaluation results. Finally in Chapter 5, we conclude with the future direction of research.



# CHAPTER 2

## Review of Prior Works

This chapter contains the details of the previous works done on sentiment analysis, multi-label classification and feature selection.

### 2.1 Sentiment Analysis

In recent years, there has been a huge interest to mine and understand the sentiments. Alec et al. (2009) [4] provide one of the first studies on sentiment analysis on microblogging websites which describes a distant supervised based approach for sentiment classification using hash tags in tweets. Kevin et al. (2011) [5] propose an approach to sentiment analysis in twitter using Part-of-Speech (PoS) tagged n-gram features and some twitter specific hash tags. Pak et al. (2010) [6] propose a method to automatically create a training corpus using emoticons . Agarwal et al. (2011) [7] used tree kernel decision tree that made use of the features such as kernel decision, Part-of-Speech information, lexicon based features and several other features. Dmitry et al. (2010) [8] observed that hash tags and smileys work good for sentiment analysis. Finn (2011) [9] concluded that the AFINN word list performs slightly better than ANEW (Affective Norms for English Words) in Twitter sentiment analysis. N-gram features, Part-of-Speech (PoS) features, emoticons features and opinion words are the most widely used features for developing sentiment analyser. Support Vector Machine (SVM) <sup>1</sup> and Naive Bayes <sup>2</sup> are two most popular machine learning algorithms that were used for training the models.

### 2.2 Multi-label Classification

In multi-label classification, the instances are associated with a set of labels. There are two main types of approaches for multi-label classification, algorithm adaptation meth-

---

<sup>1</sup>[www.support-vector-machines.org/](http://www.support-vector-machines.org/)

<sup>2</sup><http://www.statsoft.com/textbook/naive-bayes-classifier>

ods and problem transformation methods [10]. In this work we used problem transformation methods as multi-label learners. A problem transformation method transforms a multi-label classification problem into one or more single-label problems. The problem transformation methods viz. Binary Relevance, Label Powerset, Classifier Chains and Calibrated Label Ranking used in this work for multi-label sentiment analysis.

Binary Relevance (BR) [11] transforms a multi-label classification problem to multiple binary classification problems. Here we train one classifier for each label. Basically, binary relevance is an ensemble of binary classifiers, where each classifier predicts if an instance belongs to one specific class or not and then the union of all classes (labels) that were predicted is taken as the multi-label output.

In Label Powerset (LP)[12] all the possible subset of a label set are considered as a class and hence multi-label classification problem transforms into a multi-class classification problem.

The basic idea of Classifier Chains (CC) [13] algorithm is to transform the multi-label classification problem into a chain of binary classification problems, in which subsequent binary classifiers of the chain is built upon the predictions of preceding ones.

Calibrated Label Ranking (CLR) [14] transforms the multi-label learning problem into the label ranking problem, in which ranking among labels is fulfilled by techniques of pairwise comparison. It uses pairwise comparison and have one extra dummy label (virtual label) which separates the relevant labels from irrelevant labels.

Shuhua et al. (2014) [15] purposed the first work in multi-label sentiment classification in which they compared total 11 state-of-the-art multi-label classification methods on two microblogs datasets using 8 evaluation metrics.

## 2.3 Feature Selection

Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) are widely used techniques for feature selection. Beside these, Document Frequency, Information Gain, Gain Ratio, CHI statistic and Relief-F Algorithm were also used for feature selection process. Yue Lu (2011) [16] proposed a novel optimization framework that pro-

vides a unified and principled way to combine different sources of information. Xia Hu et al. (2013) [17] presented a mathematical optimization formulation that incorporates the sentiment and emotional contagion theories into the supervised learning process. Yiming Zhao (2013)[18] proposed an algorithm performing well in both accuracy and efficiency. Liu et al. (2007) [19] study the effect of various feature selection schemes on imbalanced data. Dave et al. (2003) [20] obtained an evaluation of linguistic and statistical measures, as well as weighting schemes to improve feature selection.

# CHAPTER 3

## Methodology and Approach

### 3.1 Framework for Multi-objective Optimization for Feature Selection

#### 3.1.1 Multi-objective Formulation of Feature Selection Problem

The Multi-objective Optimization (MOO) [2] can be described as follows:

Find the vectors

$$x^* = [x_1^*, x_2^*, \dots, x_N^*]^T \quad (3.1)$$

of decision variables that simultaneously optimize the M objective values

$$f_1(x), f_2(x), \dots, f_M(x); N \geq M, N > 1 \text{ and } M > 1 \quad (3.2)$$

while satisfying the constraints, if any.

The feature selection problem for sentiment analysis under multi-objective framework is stated as follows:

$$\text{maximize } P_1(F), P_2(F), \text{ where } P_1, P_2, \dots \in (\text{recall}, \text{precision}, F - \text{measure}, \dots). \quad (3.3)$$

#### 3.1.2 Selection of Objectives

In MOO, we require to determine the objective functions which are contradictory in nature. For multi-class sentiment analysis we are using precision, recall, accuracy and number of features as the objective functions.

Precision is defined as the fraction of number of true positives (TP) to the the sum of number of true positives and number of false positives (FP) and recall is defined as the fraction of number of true positives to the sum of number of true positives and number of false negatives (FN).

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP})$$

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FN})$$

$$\text{Accuracy} = (\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN})$$

In multi-label sentiment analysis, we used hamming loss, subset accuracy and micro-averaged F-measure as objectives. Hamming loss is the fraction of the wrong labels to the total number of labels. This is a loss function, so the optimal value is zero. Subset accuracy is defined as- if the entire set of predicted labels for an instance strictly match with the true set of labels, then the subset accuracy is 1 otherwise it is 0. Micro-averaged F-measure is a harmonic mean of precision and recall which gives each sample label pair an equal contribution to the overall metric.

### **3.1.3 Chromosome Representation and Population Initialization**

If total number of features is N, then the length of chromosome will be N. For example, let N=12 (i.e. total 12 different features are available). The chromosome represents the use of 7 features for constructing a classifier (first, third, fourth, seventh, tenth, eleventh and twelfth ). The value of each bit in chromosome is randomly initialized to either 0 or 1. Here, if the  $i^{\text{th}}$  position of a chromosome is 1 then the  $i^{\text{th}}$  feature participates in constructing the classifier else not. If the population size is P then all the P number of chromosomes of this population are initialized in the same way. A chromosome representation is shown in Figure 3.1 [21].

### **3.1.4 Fitness Computation**

For the fitness computation, the following procedure was executed.

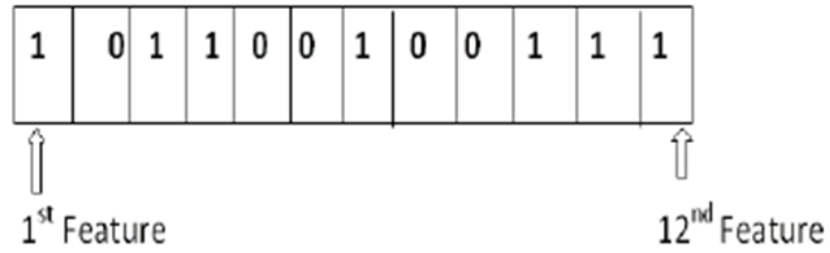


Figure 3.1: Chromosome representation for NSGA-II based feature selection

- (i) Let there are  $F$  number of features present in a particular chromosome (i.e. there are total  $F$  number of 1's present in that chromosome).
- (ii) Construct a classifier with only these  $F$  features.
- (iii) Now, test this classifier with test data and obtain objective values, for example, recall and precision values. The objective is to optimize these values of objective functions using the search capability of NSGA-II.

### 3.1.5 Other Operators

After fitness computation, we used binary tournament selection [2] as in NSGA-II, followed by crossover [2] and mutation [2]. The most characteristic part of NSGA-II is its elitism operation [2], where the non-dominated solutions among the parent and child populations are propagated to the next generation. The near pareto optimal strings of the last generation provide the different solutions to the feature selection problem.

### 3.1.6 Selection of Final Solution

Multi-objective optimization algorithm provides a large number of non-dominated solutions [2] on the final pareto optimal front. Although each of these solutions are equally important but sometimes the user may require a single solution. Hence, we developed a

method to select a single solution from the set of solutions for the multi-class sentiment analysis.

For every solution from the final pareto optimal front, (i) Classifier is trained using the feature present in that particular solution (ii) tested this classifier on test data and (iii) obtained the F-measure value using the value of the recall and the precision. Finally, we selected the solution with maximum F-measure value.

### 3.2 Framework for Multi-objective Optimization for Simultaneous Feature Selection and Parameter Optimization

To develop a framework for multi-objective optimization for simultaneous feature and parameter selection, we updated the framework of multi-objective feature selection problem (details mentioned in section 3.1). The only difference between these two frameworks is in the chromosome representation. Here chromosome representation have two parts, first one consists of the features value and second part consists of the parameter values as shown in Figure 3.2

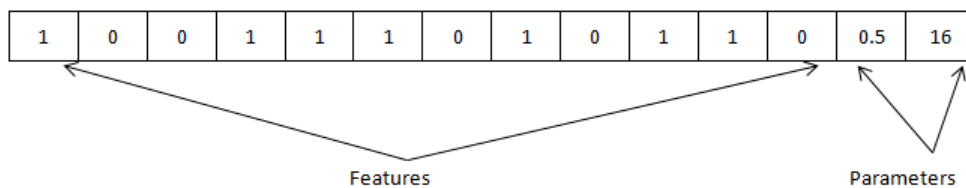


Figure 3.2: Chromosome representation for NSGA-II based simultaneous feature selection and parameter optimization

### 3.3 Multi-class Sentiment Analysis

This section describe features used for training the classifiers and framework for multi-class sentiment analysis.

### 3.3.1 Features used

For a classification problem feature selection plays an important role. We used a set of 55 features. These features are categorized in the following six categories.

#### 3.3.1.1 Emoticon Features

**3.3.1.1.1 Positive Smiley** It is a common habit that people represents emoticons through verity of smiles. A smiley present in a tweet directly represents its sentiment. A feature is defined that identifies whether the positive smiley(s) is/are present or not in a tweet. We used a set of positive smiles available at this web page. <sup>1</sup>

**3.3.1.1.2 Negative Smiley** Similar to positive smiley, we also obtained a set of negative smileys from the same source. The value of this feature is set to “yes” or “no” which ensures the presence of at least one negative emoticon.

**3.3.1.1.3 Last Token Smiley** This feature indicates whether the last token in a tweet is a smiley or not.

#### 3.3.1.2 Lexicon Features

We used three automatically created sentiment lexicons, namely NRC Hash tag Sentiment Lexicon [22], Sentiment140 Lexicon [22] and Bing Liu Lexicon [23].

**3.3.1.2.1 NRC Hash tag Sentiment Lexicon** A set of hash tags, with their scores, kindly provided by [22] was used to calculate the value of NRC Lexicon. We obtained the following features from this Lexicon:

(i) A feature value was set to +1, 0 or -1 based upon the value of its overall score which was obtained as the sum of score of all the words of a tweet. While taking the score as zero for those words which are not present in the NRC Lexicon list.

(ii) A feature value generated based upon the number of the words present in a tweet having their score greater than zero.

---

<sup>1</sup><http://www.datagenetics.com/blog/october52012/index.html>



(iii) A feature value generated based upon the number of the words present in a tweet having their score less than zero.

(iv) The maximum value of polarity among all the tokens of a tweet.

(v) The minimum value of polarity among all the tokens of a tweet.

(vi) The polarity value of last token which have non zero polarity.

**3.3.1.2.2 Sentiment140 Lexicon** In this lexicon, the individual scores of the tokens have been calculated based on the number of the tweets in which these tokens co-occurred with the positive or the negative emoticons. For every tweet in the data set, the following features were computed using the sentiment score  $score(w)$  of each token  $w$  in the tweet:

(i) The total score =  $\sum_{w \in tweet} score(w)$ .

(ii) The number of tokens in the tweet with  $score(w) > 0$

(iii) The number of tokens in the tweet with  $score(w) < 0$

(iv) The maximal score of any token in the tweet =  $\max_{w \in tweet} score(w)$

(v) The minimal score of any token in the tweet =  $\min_{w \in tweet} score(w)$

(vi) The score of the last positive token (  $score(w) > 0$  ) in the tweet

**3.3.1.2.3 Bing Liu's Lexicon** From Bing Liu's Lexicon we obtained the following two features:

(i) The number of words of a tweet present in the BLL's (Bing Liu Lexicon) word list of the positive lexicons.

(ii) The number of words of a tweet present in the BLL's (Bing Liu Lexicon) word list of the negative lexicons.

### 3.3.1.3 N-gram Features

We generated unigrams and bigrams from N-gram Statistics Package (NSP) [24].

**3.3.1.3.1 NSP Unigram** This feature generates a bit string of length equal to number of words present in training data and in test data for training a model and to test the model respectively. Presence of any word denoted by 1. It have 0 if a word is not present in a particular tweet. Example of unigrams computation is shown in Table 3.1

Table 3.1: Illustrative example: Unigram Computation

S.N.	Tweets	Unigrams						Unigram Vector
		he	is	fine	sad	this	fake	
1.	he is fine	1	1	1	0	0	0	111000
2.	he is sad	1	1	0	1	0	0	110100
3.	this is fake	0	1	0	0	1	1	010011

**3.3.1.3.2 NSP Bigram** Similar to unigrams, bigrams for entire training data and test data were generated using Ngram Statistics Package [24]. Bigrams computation is shown in Table 3.2.

Table 3.2: Illustrative example: Bigram Computation

S.N.	Tweets	Bigrams					Bigram Vector
		he is	is fine	is sad	this is	is fake	
1.	he is fine	1	1	1	0	0	111000
2.	he is sad	1	1	0	1	0	110100
3.	this is fake	0	1	0	0	1	010011

### 3.3.1.4 SentiWordNet Features

To implement SWN Positive words, SWN Negative words, SWN Neutral words and SWN Polarity features, we used a SentiWordNet dictionary [25] which provides positive and negative sentiment score for 1.17 hundred thousands words.

**3.3.1.4.1 SWN Positive words** This feature equates with the number of words present in a tweet for which positive sentiment score is greater than negative sentiment score, while the words of a tweet which were not present in the SentiWordNet dictionary were ignored.

**3.3.1.4.2 SWN Negative words** Similar to SWN positive words, SWN negative words are also used as a feature which denote the number of words of a tweet for

which negative sentiment score is greater than the positive sentiment score (ignoring the words not available in SentiWordNet dictionary).

**3.3.1.4.3 SWN Neutral words** Number of neutral words for a tweet obtained by the following formula:

SWN Neutral words = (number of words in a tweet) - (number of SWN positive words + number of SWN negative words).

**3.3.1.4.4 SWN Polarity** For every tweet a polarity score is assigned based upon the positive score and negative score. Let “x” denotes the sum of positive sentiment score of all the words of a tweet and “y” denotes the sum of negative sentiment score of all the words of a tweets. If  $(x-y) > 0.5$  then polarity will be positive. If  $(x-y) < -0.5$  then polarity will be negative. If  $(x-y)$  lies between -0.5 to 0.5 then neutral polarity is assigned.

### **3.3.1.5 Part-of-Speech (POS) Features**

Chris Nicholls and Fei Song [26] found that different PoS categories contribute to sentiment in varying degrees. To extract the part of speech of every word present in a tweet, we used the CMU ARK Tagger<sup>2</sup>. It provides the Part-of-Speech of a word along with its confidence value in between 0 to 1. Table 3.3 shows all the used PoS tags.

### **3.3.1.6 Miscellaneous Features**

Beside above five categories of features the following features are also used:

**3.3.1.6.1 Hash Count** This feature gives the number of hash (#) present in a tweet.

**3.3.1.6.2 Tweet Length** Generally a lengthy tweet have more number of the stop words. This feature counts the number of words present in a particular tweet.

---

<sup>2</sup><http://www.ark.cs.cmu.edu/TweetNLP/>

Table 3.3: POS Tags

S. N	Tag	Description	Example
1	CN	Common Noun	books, someone
2	Pro	Pronoun	I, you, he, she
3	NP	Nominal + Possessive	books/â someone/âs
4	PN	Proper Noun	lebron usa iPad
5	PNP	Proper Noun + Possessive	Americaâs
6	NV	Nominal + Verbal	heâs bookâll iono
7	PNV	Proper Noun + Verbal	Markâll
8	VC	Verb Coupla	might gonna
9	A	Adjective	active, best, brave
10	AD	Adverb	actively, boldly, successfully
11	!	Interjection	Ahh, Alas, Bingo
12	D	Determiner	the, teh, its, itâs
13	P	Preposition	at, for, in, into
14	CC	Coordinating Conjunction	and, n, &, +, BUT
15	VP	Verb Particle	out off Up UP
16	EP	Existential Predeterminer	both
17	EPV	Existential Predeterminer + Verbal	thereâs, allâs
18	#	Hash	#FIFA, #CIKM2014
19	@	AT-mention	@modi, @ravi
20	DM	Discourse Marker	so, right, okay
21	U	URL-email	www.google.com/students
22	E	Emoticon	:) :( :D
23	\$	Numeral	1st, 10th, 3rd
24	,	Punctuaion	%, &, *, ', , '

**3.3.1.6.3 Initial Cap** The words starting with capitalized character contributes more towards classifying it.

**3.3.1.6.4 Capital Characters** It gives the fraction of characters which are in upper case, i.e. the number of characters in upper case divided by the total number of characters in a tweet.

**3.3.1.6.5 All Cap Words** As the text written in capital letters express sentiment strongly. This feature gives the number of words having all characters in upper case in a particular tweet.

**3.3.1.6.6 Negation** Any form of “not” present in a sentence revert the sentiment polarity. We manually collected all possible form of words with use of not. Example-

“can’t”, “won’t”, “shouldn’t” etc. This feature value is set to “yes” or “no” based on the presence of any form of not.

**3.3.1.6.7 Stop Words** Hassan Saif et al. (2012) [27] observed that the classifiers learned with stop words outperform those learned without the stop words. So we used this as a feature and assumed that if the number of stop words is greater than 20 % (in terms of the number of words) then the tweet have more possibility of being neutral.

**3.3.1.6.8 Elongated Words** To represent strong emotions, people use to repeat same character more than two times. Examples are: “Happpppy”, “cooooool” etc. Value of this feature is assigned based on the presence of at least one elongated word.

**3.3.1.6.9 Last Token** The value of this feature is assigned based on the presence of “?” or “!” in the last word. If the last token of the tweet contains “?” or “!” then feature value is set to “yes” otherwise “no”.

### **3.3.2 Framework for Multi-class Sentiment Analysis**

The data has to be pre-processed before being used for actual machine learning training. Each tweet is processed to extract only those relevant parts that are useful for sentiment classification. For example, we deleted tweets which don’t have any label in training data; symbols and punctuation markers are filtered out; URLs are replaced by the word URL; words starting with digits are filtered out etc. Each tweet is then passed through the ARK tagger developed by CMU<sup>3</sup> for tokenization and Part-of-Speech (POS) tagging.

From pre-processed tweets we build the required training and test arff files and then string to word vector and reorder filters were applied. A training data file thus obtained, is used to train a classifier and then same model applied on test data. For training a classifier, we used four different algorithms- Sequential Minimal Optimization (SMO)[28], LibSVM [29], LibLINEAR [30] and Random Forest [31]. Model evaluation on test data provides some parameters (eg. Precision, Recall, Accuracy, F-measure etc.) as a

---

<sup>3</sup><http://www.ark.cs.cmu.edu/TweetNLP/>

result which are supplied to NSGA-II (Non-dominated Sorting Genetic Algorithm).

In one experiment precision and recall and in another experiment accuracy and number of features were given as inputs to MOO's framework where tournament selection, mutation, population decoding, population evaluation and non dominated sorting operations were used. A combined system architecture of sentiment analyser and MOO's framework is shown in Figure 3.3

## 3.4 Multi-label Sentiment Analysis

### 3.4.1 Features used

In multi-label sentiment classification, we used two different data sets (details in section 4.2.1). Both data sets have different features. A total of 72 features were used with Emotion data set which includes 8 Rhythmic features [32] and 64 Timbre features [32]. More details on these features are available on link <sup>4</sup>. CAL500 data set has a total of 68 features including Semantic features [33] and Music feature [33]. More details regarding these features is available on link <sup>5</sup>.

### 3.4.2 Framework for Multi-label Sentiment Analysis

The framework for multi-label sentiment analysis is almost same as the framework for multi-class sentiment analysis which we discussed in section 3.3.2. The minor up-gradation in this framework is discussed here. In the framework for multi-label sentiment analysis no filters were required and for training a model it require a multi-label learner along with a base classifier. To implement multi-label learner we used Mulan [34] which is a Java library for multi-label classification.

---

<sup>4</sup><http://lpis.csd.auth.gr/publications/tsoumakas-ismir08.pdf>

<sup>5</sup><http://cseweb.ucsd.edu/~datorres/bibs/IEEE-ASLP.pdf>

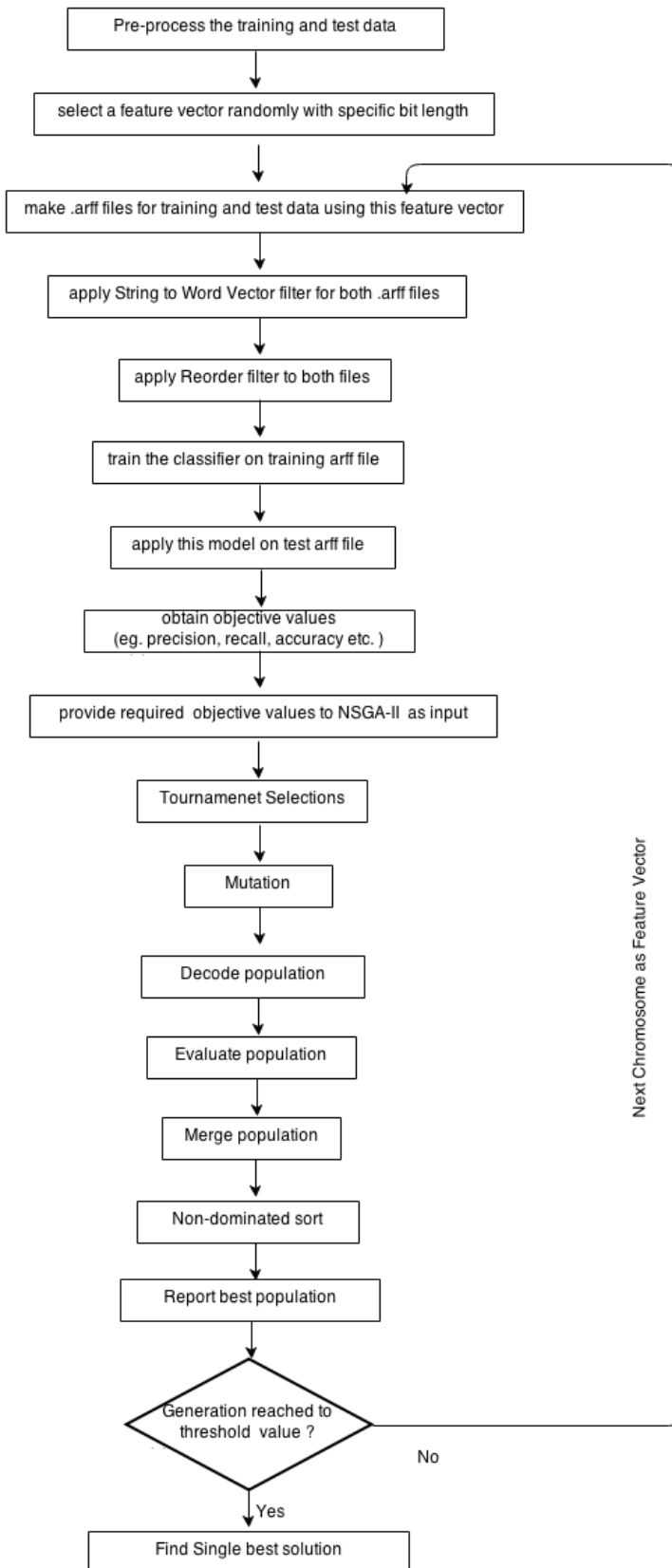


Figure 3.3: System Architecture

# CHAPTER 4

## Experiments and Result Discussion

### 4.1 Experiments for Multi-class Sentiment Analysis

#### 4.1.1 Data sets

We used the data sets from SemEval-2014<sup>1</sup>. The data sets contain 8223 classified tweets in training data and 8987 tweets in test data. The details of data sets according to classes are given in Table 4.1.

Table 4.1: Multi-class Data Sets

S.N.	Data Set	Class			Total
		Positive	Negative	Neutral	
1	Training	3071	1210	3942	8223
2	Test	3506	1541	3940	8987

#### 4.1.2 Feature Selection using Multi-objective Optimization

##### 4.1.2.1 Feature sets

We used different feature sets as shown in Table 4.2. The sets are defined as follows:

**S28:** is a set of 28 features (these features are listed in Table 4.2 ).

**R11:** is a subset having 11 features randomly selected from S28

**R14:** is a subset having 14 features randomly selected from S28

**R18:** is a subset having 18 features randomly selected from S28

**OS28:** Set is different-different optimal sets (obtained by NSGA-II) for three classifiers.

---

<sup>1</sup><http://alt.qcri.org/semEval2014/task9/>



Table 4.2: Feature Sets of S28

Feature No.	Features	Feature sets									
		R11	R14	R18	S28	OS28 for Exp-PR1			OS28 for Exp-ANF1		
						SMO	LS	LL	SMO	LS	LL
0	Hash Count	1	0	0	1	0	0	0	0	1	1
1	Tweet Length	0	1	1	1	0	0	1	0	0	0
2	Init Cap	0	0	1	1	1	0	1	1	1	0
3	Percent Cap	0	1	0	1	0	1	0	0	1	0
4	All Upper Word	1	0	1	1	0	0	0	0	1	0
5	Not Present	0	1	0	1	1	0	1	0	0	0
6	Positive Smiley	1	1	1	1	1	1	1	1	1	1
7	Negative Smiley	0	1	0	1	0	0	1	0	0	0
8	Stop Words	0	0	1	1	0	1	0	0	0	1
9	Elongated Words	1	0	0	1	1	1	1	1	0	1
10	Last Token	0	1	1	1	1	0	0	1	0	1
11	SWN Positive Words	0	1	1	1	1	1	0	0	0	1
12	SWN Negative Words	1	0	1	1	0	1	0	1	1	0
13	SWN Neutral Words	0	1	0	1	1	0	1	1	0	0
14	SWN Polarity	1	0	1	1	1	0	1	1	0	1
15	LexNRC	1	1	0	1	1	1	1	1	1	0
16	Positive LexNRC Token	0	0	1	1	1	1	0	0	1	0
17	Negative LexNRC Token	0	0	1	1	0	1	1	0	0	0
18	BLL Positive Words	1	1	0	1	1	1	1	1	1	1
19	BLL Negative Words	0	0	1	1	1	1	1	1	1	1
20	Adjective	1	0	1	1	1	0	1	1	0	1
21	Adverb	0	1	0	1	0	1	1	1	0	0
22	Interjection	0	1	1	1	0	0	0	0	0	0
23	Preposition	1	0	1	1	0	1	1	1	0	0
24	Coordinating Conjunction	0	0	1	1	0	1	0	0	0	0
25	URL-Email	0	1	1	1	0	0	1	1	0	0
26	NSP Unigram	1	0	1	1	1	1	1	0	1	0
27	NSP Bigram	0	1	0	1	0	1	1	0	0	0
Number of features		11	14	18	28	14	16	18	14	11	10

LS=LibSVM, LL=LibLINEAR

#### 4.1.2.2 Results

To build a Sentiment Analyser, we performed the following two experiments to optimize the objectives. In both the experiments R11,R14 and R18 sets were used as baseline models. In these experiments, MOO parameters were used as follows:

Population size = 32

Number of generations = 10

Number of objective functions = 2

Number of binary variables = 28

Probability of crossover of binary variable = 0.98

Probability of mutation of binary variable = 0.00357

Seed for random number generator = 0.5896

**4.1.2.2.1 Experiments using Precision and Recall as Objective Functions** In this experiment, we used recall and precision as two different objectives for optimization through NSGA-II. Generally precision and recall are contradictory in nature. F-measure is a harmonic mean of precision and recall and our target is to increase F-measure hence precision and recall are best suitable objectives. Table 4.3 shows all the results for this experiment.

Table 4.3: Results for Exp-PR1

S.N.	Feature Set	Parameters	Classifier		
			SMO	LibSVM	LibLINEAR
1.	R11	P	64.00	51.20	63.90
		R	63.50	56.40	63.30
		F1	63.74	53.67	63.59
2.	R14	P	62.20	50.10	62.90
		R	62.20	55.10	62.40
		F1	62.20	52.48	62.64
3.	R18	P	63.50	62.40	61.80
		R	63.10	50.70	62.10
		F1	63.30	55.94	61.94
4.	S28	P	63.70	64.50	64.50
		R	63.50	55.30	63.80
		F1	63.60	59.54	64.14
5.	OS28	P	65.10	68.50	66.00
		R	64.40	56.60	63.50
		F1	64.75	61.98	64.72

All parameters are in percentage, P= Precision, R=Recall, F1=F-measure

As shown in Table 4.3, F-measure (in percentage) for baseline models are 63.74, 62.20, 63.30 (with SMO), 53.67, 52.48, 55.94 (for LibSVM) and 63.59, 62.64, 61.94 (for LibLINEAR) for R11,R14 and R18 respectively. F-measure for optimal subset (OS28) have the improvements of 1.01 %, 2.55 %, 1.45 %, 1.15 % with SMO, 8.31 %, 9.50 %, 6.04 %, 2.44 % with LibSVM and 1.13 %, 2.08 %, 2.78 %, 0.58 % with LibLINEAR from its R11, R14, R18 and S28 feature sets respectively.

**4.1.2.2.1.1 Result after Classifiers Ensemble** From pareto optimal solutions we used 10 SMO, 5 LibSVM and 5 LibLINEAR classifiers for ensembling. As a result we obtained F-measure  $F1 = 65.40\%$ . Resultant confusion matrix is shown in Table 4.4

**4.1.2.2.2 Experiments using Accuracy and Number of Features as Objective Functions** In another experiment, accuracy and number of feature were used as two differ-

Table 4.4: Confusion Matrix-PR

Actual Class	Predicted Class		
	Positive	Negative	Neutral
Positive	1966	156	1384
Negative	248	546	1384
Neutral	456	176	3308

ent objectives, in which target is to find maximum accuracy with minimum number of feature being used. A detailed result sheet is shown in Table 4.5.

Table 4.5: Results for Exp-ANF1

S.N.	Feature Set	Para- meters	Classifiers		
			SMO	LibSVM	LibLINEAR
1.	R11	A	63.47	56.39	63.35
		NF	11	11	11
		F1	62.30	49.60	62.30
2.	R14	A	62.17	55.12	62.37
		NF	14	14	14
		F1	61.20	47.90	61.00
3.	R18	A	63.08	50.65	62.06
		NF	18	18	18
		F1	62.30	42.80	61.80
4.	S28	A	63.49	55.29	63.77
		NF	28	28	28
		F1	62.80	48.10	63.50
5.	OS28	A	64.14	56.48	63.55
		NF	14	11	10
		F1	63.40	49.70	62.50

A = Accuracy (in %), NF= number of features, F1= F-measure (in %)

F-measure (in percentage) in this experiment for baseline models are 62.30, 61.20, 62.30 (with SMO); 49.60, 47.90, 42.80 (for LibSVM) and 62.30, 61.00, 61.80 (for LibLINEAR) for R11,R14 and R18 respectively. F-measure for optimal subset (OS28) have the improvements of 1.10%, 2.20%, 1.10% with SMO; 0.10%, 1.80%, 6.90%, with LibSVM and 0.20%, 1.50%, 0.70% with LibLINEAR from its R11, R14 and R18 feature sets respectively.

**4.1.2.2.1 Result after Classifiers Ensemble** For this experiment also, we used 10 SMO, 5 LibSVM and 5 LibLINEAR classifiers for ensembling from pareto optimal solutions. As a result we obtained F-measure F1 = 64.80%. Resultant confusion matrix is shown in Table 4.6

Table 4.6: Confusion Matrix-ANF

Actual Class	Predicted Class		
	Positive	Negative	Neutral
Positive	1969	167	1370
Negative	261	538	742
Neutral	489	184	3267

### 4.1.3 Simultaneous Feature Selection and Parameter Optimization using Multi-objective Optimization

Generally a classifier has many parameters whose values heavily influence the performance of a classifier. Hence, like feature selection, determining the appropriate values of parameters for a classifier is another important key issue. In these experiments along with feature selection we optimized parameters values as number of trees<sup>2</sup> for Random Forest, cost and gamma parameter<sup>3</sup> for LibSVM and cost parameter<sup>4</sup> for LibLINEAR.

#### 4.1.3.1 Feature sets

We used different feature sets as shown in Table 4.7. The sets are defined as follows:

**S55:** is a set of 55 features (these features are listed in Table 4.7 ).

**R20:** is a subset having 20 features randomly selected from S55.

**R30:** is a subset having 30 features randomly selected from S55.

**R40:** is a subset having 40 features randomly selected from S55.

**OS55 with FS & PO:** Set is different-different optimal sets (obtained by NSGA-II for both feature selection and parameter optimization) for three classifiers.

**OS55 with default parameters:** Set is different-different optimal sets (obtained by NSGA-II with default parameters) for three classifiers.

<sup>2</sup><http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/RandomForest.html>

<sup>3</sup><http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

<sup>4</sup><http://www.csie.ntu.edu.tw/~cjlin/papers/liblinear.pdf>

#### 4.1.3.2 Results

To build a sentiment analyser while considering both feature selection as well as parameter optimization, we performed the following two experiments to optimize the objectives. In both the experiments, we used R20, R30 and R40 sets as baseline models. MOO parameters for these experiments were set as follows:

Population size = 32

Number of generations = 20

Number of objective functions = 2

Number of real variables = 2

Probability of crossover of real variable = 0.98

Probability of mutation of real variable = 0.50

Distribution index for crossover = 14

Distribution index for mutation = 38

Number of binary variables = 55

**4.1.3.2.1 Experiment using Precision and Recall as Objective Functions** Recall and precision were used as two objectives for optimization through NSGA-II. We used F-measure to select one best solution among all of the pareto optimal solutions.

As shown in Table 4.8, F-measure (in percentage) for baseline models are 52.30, 50.40, 53.60 (with Random Forest); 50.60, 31.00, 46.20 (for LibSVM); 50.10, 50.40, 51.00 (for LibLINEAR) for R20, R30 and R40 respectively. F-measure for the full set S55 is 55.10, 39.10 and 53.40 for Random Forest, LibSVM and LibLINEAR respectively. While considering both feature selection (FS) and parameter optimization (PO), we obtained the F-measure 59.30, 59.10 and 57.70 for optimal subset (OS55 with FS and PO) with respect to Random Forest, LibSVM and LibLINEAR. Further we used same OS55 set with default parameters then F-measure were obtained as 55.50, 53.90 and for 58.30 Random Forest, LibSVM and LibLINEAR respectively. Due to only feature selection process we gained the percentage improvement of 00.40, 14.80 and 04.90 for Random Forest, LibSVM and LibLINEAR respectively while considering both feature selection (FS) and parameter optimization (PO) we gained the percentage improvement of 04.20, 20.00 and 04.30 for Random Forest, LibSVM and LibLINEAR respectively. Simultaneous feature selection and parameter optimization provided bet-

ter results for Random Forest and LibSVM while in case of LibLINEAR result obtained with default parameters is better. Optimal values obtained through NSGA-II for each classifier are shown in Table 4.10.

**4.1.3.2.2 Experiments using Accuracy and Number of Features as Objective Functions** To obtain maximum accuracy with minimum number of feature being used, we selected accuracy and number of feature as two objectives. A detailed result sheet is shown in Table 4.9.

As shown in Table 4.9, F-measure (in percentage) for baseline models are 52.30, 50.40, 53.60 (with Random Forest); 46.20, 31.00, 50.60 (for LibSVM); 54.10, 50.30, 53.10 (for LibLINEAR) for R20, R30 and R40 respectively. F-measure for the full set S55 is 55.10, 39.10, 49.90 for Random Forest, LibSVM and LibLINEAR respectively. While considering both feature selection (FS) and parameter optimization (PO), we obtained the F-measure 59.70, 59.20 and 55.50 for optimal subset (OS55 with FS and PO) with respect to Random Forest, LibSVM and LibLINEAR. Further we used same OS55 set with default parameters then F-measure were obtained as 55.50, 56.10 and 58.00 for Random Forest, LibSVM and LibLINEAR respectively. Due to only feature selection process we gained the percentage improvement of 00.40, 17.00 and 08.10 for Random Forest, LibSVM and LibLINEAR respectively while considering both feature selection (FS) and parameter optimization (PO) we gained the percentage improvement of 04.60, 20.10 and 05.60 for Random Forest, LibSVM and LibLINEAR respectively. In case of Random Forest and LibSVM simultaneous feature selection and parameter optimization provided better results while in case of LibLINEAR result obtained with default parameters is better. For the optimal subset (OS55) number of features were used as 31,23 and 26 for Random Forest, LibSVM and LibLINEAR respectively. Optimal values obtained through NSGA-II for each classifier are shown in Table 4.10.

## 4.2 Experiments for Multi-label Sentiment Analysis

### 4.2.1 Data sets

We used two data sets for conducting experiments with multi-label sentiment classification. The data sets obtained from <sup>5</sup>. Emotion data set contains a total of 593 instances and there are total 6 labels in it. CAL500 data set have a total of 502 instances and 174 labels. The details of both the data sets are given in Table 4.11.

### 4.2.2 Feature Selection using Multi-objective Optimization for Emotion Data

We performed experiments with two different base learner each with LP, BR, CC and CLR multi-label learner to find a optimal feature subset from a set of 72 features. Here in both the experiments, we considered hamming loss and subset accuracy as our objectives for MOO. we used MOO parameters as follows:

Population size = 32

Number of generations = 20

Number of objective functions = 2

Number of binary variables = 72

Probability of crossover of binary variable = 0.98

Probability of mutation of binary variable = 0.00032

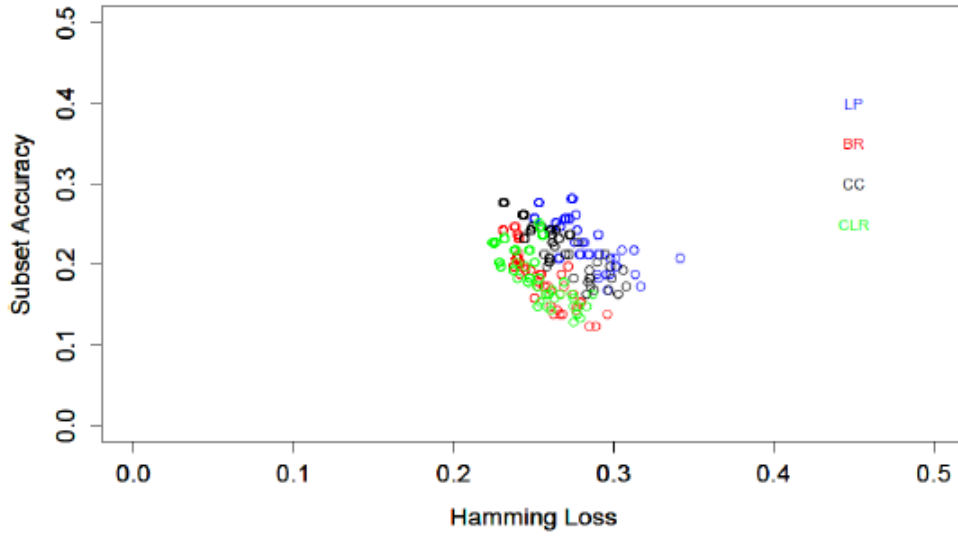
Seed for random number generator = 0.5896

#### 4.2.2.1 Experiments with J48 as base learner

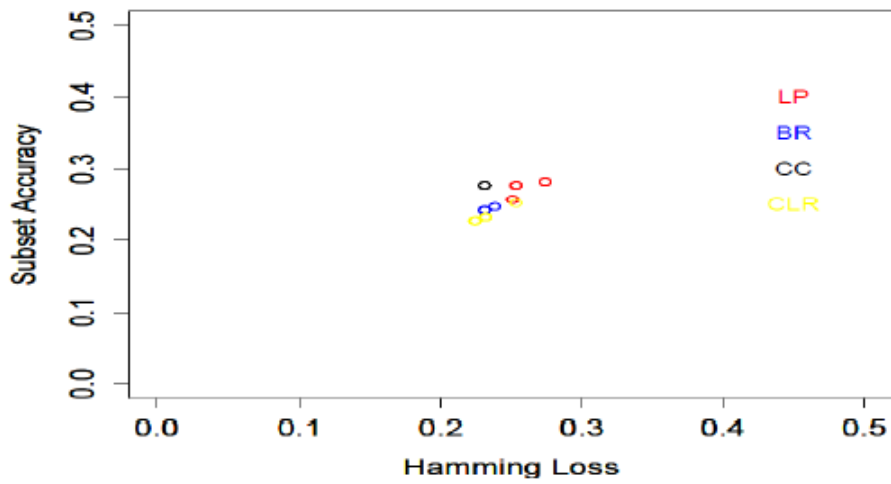
In this experiment, we used LP, BR, CC and CLR multi-label learner with J48 as base learner. Figure 4.1 shows the population representation for all population as well as best population selected by NSGA-II algorithm. Here CLR provided minimum hamming loss while best value of subset accuracy obtained for LP.

---

<sup>5</sup>[//mulan.sourceforge.net/datasets-mlc.html](http://mulan.sourceforge.net/datasets-mlc.html)



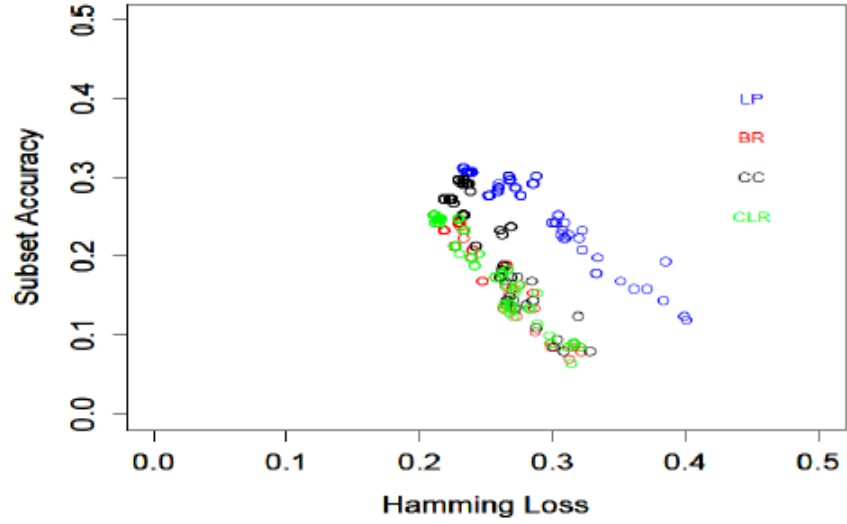
(a) All population for Emotion data with J48



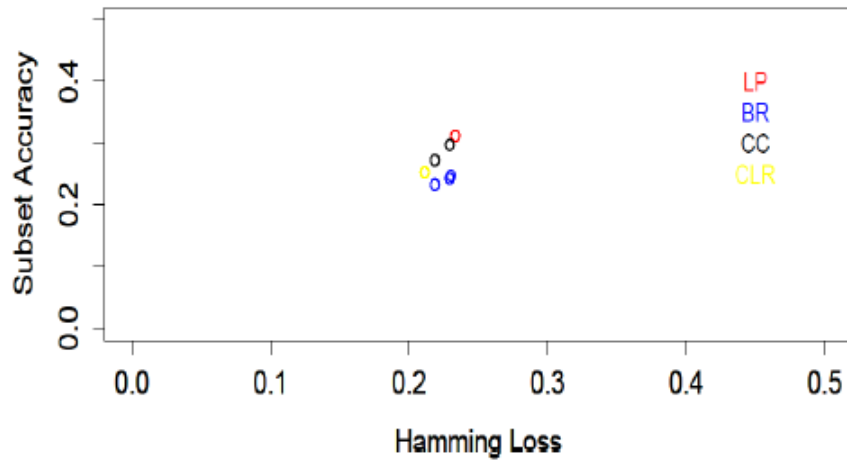
(b) Best population for Emotion data with J48

Figure 4.1: Population representation for Emotion data with J48





(a) All population for Emotion data with LibSVM



(b) Best population for Emotion data with LibSVM

Figure 4.2: Population representation for Emotion data with LibSVM

#### 4.2.2.2 Experiments with LibSVM as base learner

We carried out this experiment with LibSVM as base learner along with four multi-label learner viz. LP, BR, CC and CLR. Population representation for all population as well as best population selected by NSGA-II algorithm is shown in Figure 4.2. Here CLR provided minimum hamming loss while best value of subset accuracy obtained for LP.

### 4.2.2.3 Conclusion

We selected two best solution from best population, one corresponding to hamming loss and another corresponding to subset accuracy for each multi-label learner. Table 4.12 shows that if solution is required with minimum hamming loss then (21.12, 25.25, 38) is the best optimal value for (hamming loss, subset accuracy, number of features) obtained by CLR with LibSVM and if we require solution with maximum subset accuracy then (23.35, 31.19, 39) is the best optimal solution for (hamming loss, subset accuracy, number of features) obtained by LP with LibSVM.

## 4.2.3 Feature Selection using Multi-objective Optimization for CAL500

### Data

To find an optimal feature subset from a set of 68 features we performed experiments with two different base learner i.e. J48 and LibSVM. Here in both the experiments, we considered hamming loss and micro-averaged F-measure as our objectives for MOO. we used MOO parameters as follows:

Population size = 32

Number of generations = 20

Number of objective functions = 2

Number of binary variables = 68

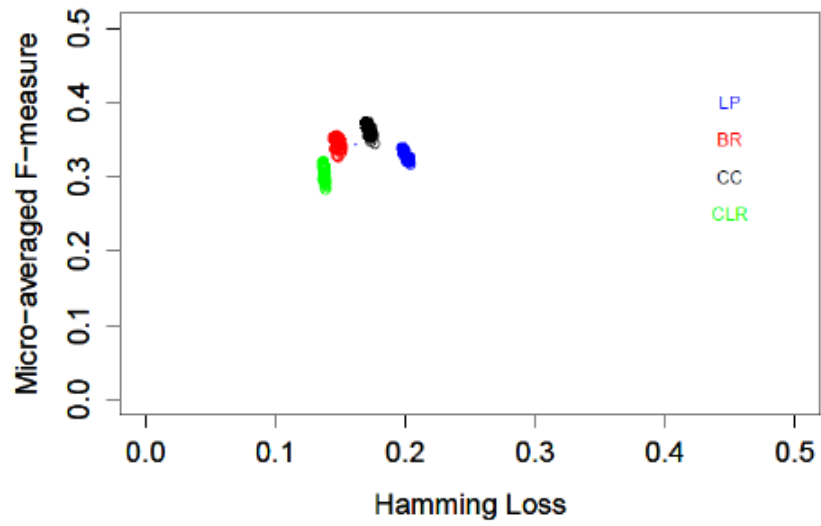
Probability of crossover of binary variable = 0.98

Probability of mutation of binary variable = 0.00032

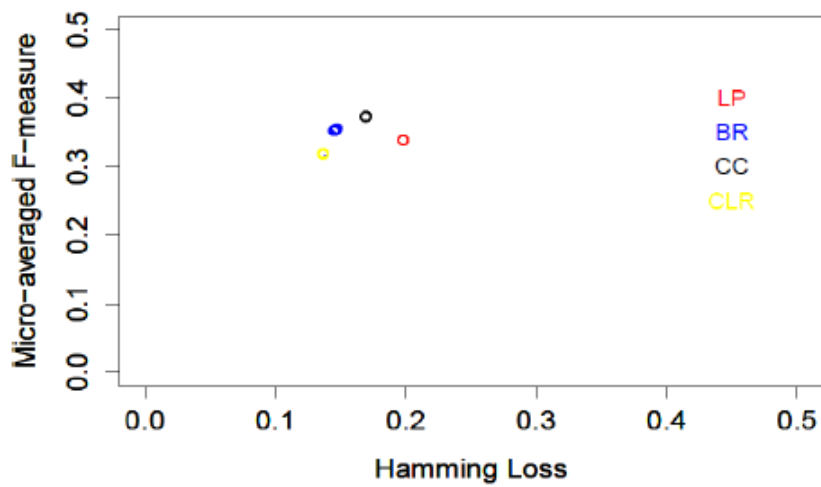
Seed for random number generator = 0.5896

### 4.2.3.1 Experiments with J48 as base learner

In this experiment, we used LP, BR, CC and CLR multi-label learner with J48 as base learner. Figure 4.3 shows the population representation for all population as well as best population selected by NSGA-II algorithm. Here CLR provided minimum hamming loss while best value of micro-averaged F-measure obtained for CC.



(a) All population for CAL500 data with J48



(b) Best population for CAL500 data with J48

Figure 4.3: Population representation for CAL500 data with J48

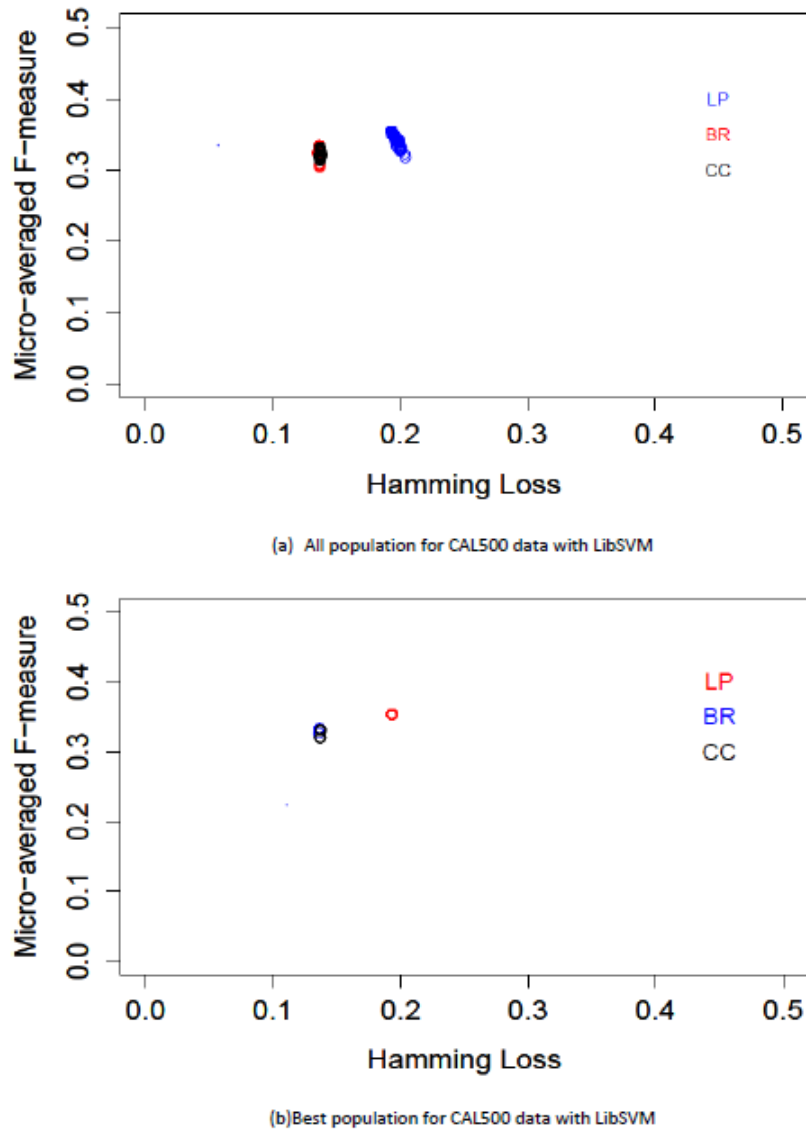


Figure 4.4: Population representation for CAL500 data with LibSVM

#### 4.2.3.2 Experiments with LibSVM as base learner

We carried out this experiment with LibSVM as base learner along with three multi-label learner viz. LP, BR and CC. Population representation for all population as well as best population selected by NSGA-II algorithm is shown in Figure 4.4. Here both BR and CC provided minimum hamming loss while best value of micro-averaged F-measure obtained for LP.

### 4.2.3.3 Conclusion

We selected two best solution from best population, one corresponding to hamming loss and another corresponding to micro-averaged F-measure for each multi-label learner. Table 4.13 shows that if solution is required with minimum hamming loss then (13.61, 32.63, 26) is the best optimal value for (hamming loss, micro-averaged F-measure, number of features) obtained by BR with LibSVM and if we require solution with maximum micro-averaged F-measure then (16.96, 37.34, 26) is the best optimal solution for (hamming loss, micro-averaged F-measure, number of features) obtained by CC with J48.

Table 4.7: Feature Sets of S55

Feature No	Features	Feature sets									
		R20	R30	R40	S55	OS55 for Exp-PR2			OS55 for Exp-ANF2		
						Random Forest	LibSVM	LibLINEAR	Random Forest	LibSVM	LibLINEAR
0	hash_count	1	0	1	1	0	0	0	0	0	0
1	tweet_length	1	0	1	1	0	1	0	0	0	0
2	init_cap	0	1	1	1	1	1	1	1	1	1
3	percent_capital	0	1	0	1	0	1	0	1	1	1
4	all_upper_word	0	0	1	1	1	0	1	0	0	0
5	not_present	0	1	0	1	1	0	0	0	0	0
6	psmiley	1	0	1	1	0	0	0	1	0	1
7	nsmiley	0	1	1	1	1	1	1	1	1	1
8	LastTokenSmiley	0	1	1	1	1	0	0	1	0	0
9	stop_words	1	0	1	1	0	1	1	0	1	1
10	elongated_words	0	1	1	1	1	0	0	0	1	1
11	last_token	0	0	1	1	1	0	0	0	1	0
12	SWN_positive_token_count	1	0	0	1	1	1	0	1	0	0
13	SWN_negative_token_count	0	1	1	1	1	1	1	0	1	1
14	SWN_neutral_token_count	0	0	1	1	1	0	0	1	0	0
15	SWN_polarity	1	1	0	1	1	1	1	1	1	1
16	LexNRC	0	1	1	1	1	1	1	0	1	1
17	positiveLexNRC_token	1	0	1	1	0	1	0	0	1	0
18	negativeLexNRC_token	0	1	0	1	0	0	1	1	0	0
19	maxLexNRC	0	0	1	1	1	1	1	1	1	1
20	minLexNRC	1	0	1	1	1	1	1	1	1	1
21	last_nonzero_scoreNRC	0	1	1	1	0	0	1	0	0	0
22	Lex140	1	0	1	1	0	0	0	0	0	0
23	positiveLex140_token	0	1	0	1	1	0	0	0	0	0
24	negativeLex140_token	0	1	1	1	0	0	0	1	0	1
25	maxLex140	1	0	1	1	0	0	1	0	0	0
26	minLex140	0	0	1	1	1	0	0	1	0	0
27	last_nonzero_score140	0	1	1	1	0	0	0	0	0	0
28	BLL_Positive_words	1	0	1	1	1	1	1	1	1	1
29	BLL_Negative_words	0	1	0	1	1	1	1	1	1	1
30	common_noun	0	1	1	1	1	0	1	1	1	0
31	pronoun	1	1	0	1	0	1	0	0	1	0
32	nominal_possessive	0	0	1	1	1	0	0	0	1	1
33	proper_noun	0	1	1	1	0	0	0	0	0	0
34	proper_noun_possessive	1	1	0	1	1	1	1	1	0	1
35	nominal_verbal	0	0	1	1	1	0	1	1	0	0
36	proper_noun_verbal	0	1	0	1	0	0	1	1	0	1
37	verb_coupla	1	0	1	1	1	0	0	1	0	0
38	adjective	0	1	1	1	1	0	1	1	0	1
39	adverb	0	1	0	1	0	1	0	1	1	0
40	injection	0	0	1	1	0	1	0	0	1	0
41	determiner	1	0	1	1	1	0	0	1	0	0
42	preposition	0	1	1	1	1	0	1	0	0	1
43	conditional_conjunction	0	0	1	1	0	1	1	0	1	1
44	verb_particle	1	1	0	1	0	0	0	0	0	0
45	existential_predeterminer	0	1	1	1	0	0	0	0	0	0
46	existential_predeterminer_verbal	0	1	0	1	0	0	0	1	0	0
47	number_of_hash	1	0	1	1	0	0	0	1	0	1
48	AT_mention	1	0	1	1	1	1	1	1	0	1
49	discourse_marker	0	1	1	1	1	1	1	1	0	0
50	url_email	1	0	1	1	1	0	0	1	0	0
51	emoticon	0	1	0	1	1	1	1	1	1	1
52	numeral	0	1	1	1	1	1	1	1	1	1
53	punctuaion	0	1	0	1	1	1	0	0	0	1
54	other_POS	1	0	1	1	1	1	1	1	1	1
Number of features		20	30	40	55	33	25	26	31	23	26

Table 4.8: Results for Exp-PR2

S. N.	Feature Sets	Parameters	Classifiers		
			Random Forest	LibSVM	Lib-LINEAR
1	R20	P	52.10	60.40	49.90
		R	52.70	56.30	53.20
		F1	52.30	50.60	50.10
2	R30	P	50.50	52.30	52.90
		R	50.80	44.60	52.10
		F1	50.40	31.00	50.40
3	R40	P	53.50	58.70	50.50
		R	54.10	53.50	52.20
		F1	53.60	46.20	51.00
4	S55	P	55.10	57.90	53.50
		R	55.60	49.20	55.10
		F1	55.10	39.10	53.40
5	OS55 with FS & PO	P	62.20	61.00	62.40
		R	61.00	60.60	60.30
		F1	59.30	59.10	57.70
6	OS55 default paameters	P	55.50	59.40	62.30
		R	56.10	57.30	60.60
		F1	55.50	53.90	58.30
% improvement due to FS			00.40	14.80	04.90
% improvement due to FS and PO			04.20	20.00	04.30

P=Precision, R=Recall, F1=F-measure (all in %) FS= Feature Selection, PO= Parameter Optimization

Table 4.9: Results for Exp-ANF2

S. N.	Feature Sets	Parameters	Classifiers		
			Random Forest	LibSVM	Lib-LINEAR
1	R20	A	52.65	56.33	53.32
		NF	20	20	20
		F1	52.30	50.60	53.10
2	R30	A	50.84	44.64	52.00
		NF	30	30	30
		F1	50.40	31.00	50.30
3	R40	A	54.11	53.49	55.20
		NF	40	40	40
		F1	53.60	46.20	54.10
4	S55	A	55.61	49.2	54.16
		NF	55	55	55
		F1	55.10	39.10	49.90
5	OS55 with FS & PO	A	61.32	60.58	59.06
		NF	31	23	26
		F1	59.70	59.20	55.50
6	OS55 default paameters	A	56.08	58.53	60.24
		NF	31	23	26
		F1	55.50	56.10	58.00
% improvement due to FS			00.40	17.00	08.10
% improvement due to FS and PO			04.60	20.10	05.60

A=Accuracy (in %), NF=Number of Features, F1=F-measure (in %), FS= Feature Selection, PO= Parameter Optimization

Table 4.10: Optimized Parameters

S.N.	Classifier	Parameters	Experiment	
			Exp-PR2	Exp-ANF2
1.	Random Forest	Trees	862, 853	1990
2.	LibSVM	Cost	$2^{14}$	$2^{14}$
		Gamma	$2^{(-10)}$	$2^{(-9)}$
3.	LibLINEAR	Cost	4	$2^{(-2)}, 2^{(-7)}, 2^{(-8)}$

Table 4.11: Multi-label Data Sets

S.N.	Data Sets	Instances	Labels
1.	Emotion	593	6
2.	CAL500	502	174



Table 4.12: Results with Emotion Data

Base Classifier	Multi-label learner	Optimal values biased towards	
		Hamming Loss	Subset Accuracy
J48	LP	(25.08, 25.74, 37)	(27.39, 28.22, 39)
	BR	(23.10, 24.26, 38)	(23.84, 24.75, 32)
	CC	(23.10, 27.72, 32)	
	CLR	(22.44, 22.77, 34)	(25.33, 25.25, 38)
LibSVM	LP	(23.35, 31.19, 39)	
	BR	(21.86, 23.27, 39)	(23.02, 24.75, 34)
	CC	(21.86, 27.23, 39)	(22.94, 29.70, 38)
	CLR	(21.12, 25.25, 38)	

Here (x, y, z)= (Hamming Loss, Subset Accuracy, number of features).

Table 4.13: Results with CAL500 Data

Base Classifier	Multi-label learner	Optimal values biased towards	
		Hamming Loss	Micro averaged F-measure
J48	LP	(19.79, 33.91, 35)	
	BR	(14.48, 35.28, 26)	(14.73, 35.62, 30)
	CC	(16.93, 37.22, 25)	(16.96, 37.34, 26)
	CLR	(13.67, 31.73, 26)	(13.68, 31.95, 24)
LibSVM	LP	(19.29, 35.40, 27)	(19.32, 35.41, 28)
	BR	(13.61, 32.63, 26)	(13.66, 33.35, 29)
	CC	(13.70, 32.03, 37)	(13.74, 33.10, 36)

Here (x, y, z)= (Hamming Loss, Micro-averaged F-measure, number of features).

# CHAPTER 5

## Conclusion and Future Work

In this work, we have posed the problem of appropriate feature selection for multi-class and multi-label sentiment analysis as a MOO problem with three different data sets and obtained 17 % improvement in F-measure for LibSVM. We solved this crucial problem with a technique, based on a popular multi-objective evolutionary algorithm, NSGA-II. We performed feature selection for Random Forest and three well-known implementations of SVM, viz. SMO, LibSVM and LibLinear. Comparing these classifiers we observed that SMO based system outperform the other three methods.

In second phase, we have posed the problem of simultaneous feature selection and parameter optimization for multi-class sentiment analysis as a MOO problem with SemEval-14 data sets and obtained 20.10 % improvement in F-measure for LibSVM.

In future, we would like to add more features and then both feature selection and parameter optimization will be considered as MOO problem for multi-label sentiment analysis. We would like to consider more objectives along with precision, recall, F-measure, hamming loss and subset accuracy and also with different algorithm adaptation method like MLkNN.

## REFERENCES

- [1] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan Claypool Publishers, 2012.
- [2] K. Deb, *Multi-objective Optimization Using Evolutionary Algorithms*. John Wiley and Sons, Ltd, 2001.
- [3] T. G. Dietterich, “Ensemble methods in machine learning,” in *Multiple classifier systems*. Springer, 2000, pp. 1–15.
- [4] R. B. Alec Go and L. Huang, “Twitter sentiment classification using distant supervision,” Univ. stanford, Tech. Rep., 2009.
- [5] N. S. Kevin Gimpel, “Part-of-speech tagging for twitter: Annotation, features, and experiments,” in *Proceeding HLT '11 Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2 Pages 42-47*, 2011.
- [6] A. Pak and P. Paroubek, “Part-of-speech tagging for twitter: Annotation, features, and experiments,” in *Seventh International Conference on Language Resources and Evaluation, LREC 2010*, 2010.
- [7] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, “Sentiment analysis of twitter data,” in *Proceedings of the Workshop on Languages in Social Media*. Association for Computational Linguistics, 2011, pp. 30–38.
- [8] O. T. Dmitry Davidov and A. Rappoport, “Enhanced sentiment learning using twitter hashtags and smileys,” in *Coling 2010: Poster Volume, pages 241–249*, Beijing, August 2010.
- [9] F. Å. Nielsen, “Afinn,” Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Tech. Rep., March 2011.
- [10] K. Tsoumakas and Vlahavas, *Data Mining and Knowledge Discovery Handbook*. Springer, 2010.
- [11] S. X. . B. C. Boutell M.R., Luo J., “Learning multi-label scene classification,” in *Pattern Recognition*, 2004.
- [12] I. K. Grigorios TSOUMAKAS, “Multi-label classification: An overview,” in *International Journal of Data Warehousing and Mining*, 2004.
- [13] G. H. Jesse READ, Bernhard PFAHRINGER and E. FRANK, “Classifier chains for multi-label classification,” in *Machine Learning*, 2011.
- [14] J. Fürnkranz, E. Hüllermeier, E. L. Mencía, and K. Brinker, “Multilabel classification via calibrated label ranking,” vol. 73, no. 2. Springer, 2008, pp. 133–153.

- [15] S. M. Liu and J.-H. Chen, “A multi-label classification based approach for sentiment classification,” vol. 42, no. 3. Elsevier, 2015, pp. 1083–1093.
- [16] Y. LU, “Opinion integration and summarization,” Tech. Rep., 2011.
- [17] X. Hu, L. Tang, J. Tang, and H. Liu, “Exploiting social relations for sentiment analysis in microblogging,” in *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 537–546.
- [18] Y. Zhao, “Text sentiment analysis algorithm optimization and platform development in social network,” in *Computational Intelligence and Design (ISCID), 2013 Sixth International Symposium on (Volume:1 )*, Hangzhou, Oct. 2013.
- [19] Y. Liu and H. T. Loh, *Natural Language Processing and Text Mining*. Springer London, 2007.
- [20] S. L. Kushal Dave and D. M. Pennock, “Mining the peanut gallery: Opinion extraction and semantic classification of product reviews,” in *WWW2003*, Budapest, Hungary, May 2003.
- [21] A. Ekbal, S. Saha, and C. S. Garbe, “Feature selection using multiobjective optimization for named entity recognition,” in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 1937–1940.
- [22] S. Mohammad, S. Kiritchenko, and X. Zhu, “Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets,” in *Proceedings of the seventh international workshop on Semantic Evaluation Exercises (SemEval-2013)*, Atlanta, Georgia, USA, June 2013.
- [23] X. Ding, B. Liu, and P. S. Yu, “A holistic lexicon-based approach to opinion mining,” in *Proceedings of the 2008 International Conference on Web Search and Data Mining*. ACM, 2008, pp. 231–240.
- [24] Banerjee and Pedersen, “The design, implementation, and use of the ngram statistics package,” in *Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, February 2003.
- [25] S. B. Andrea Esuli and F. Sebastiani, “Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining,” in *Seventh conference on International Language Resources and Evaluation LREC’10*, Valletta, Malta, May 2010.
- [26] F. S. CHRIS NICHOLLS, “Improving sentiment analysis with part-of-speech weighting,” in *Eighth International Conference on Machine Learning and Cybernetics*, Baoding, July 2009.
- [27] Y. H. Hassan Saif and H. Alani, “Semantic sentiment analysis of twitter,” in *11th international conference on The Semantic Web ISWC’12, Volume I*, Heidelberg, 2012.
- [28] J. Platt *et al.*, “Fast training of support vector machines using sequential minimal optimization,” vol. 3. Cambridge, MA, 1999.
- [29] C.-C. Chang and C.-J. Lin, “Libsvm: A library for support vector machines,” in *ACM Transactions on Intelligent Systems and Technology*, 2011.

- [30] K.-W. C. Rong-En Fan and C.-J. Hsieh, “Liblinear: A library for large linear classification,” in *Journal of Machine Learning Research*, 2008.
- [31] L. Breiman, “Random forests,” in *Machine Learning*, 2001.
- [32] G. T. Konstantinos Trohidis and G. Kalliris, “Multi-label classification of music into emotions,” in *ISMIR*, 2008.
- [33] L. B. Douglas Turnbull and D. Torres, “Semantic annotation and retrieval of music and sound effects,” in *IEEE Transation on Audio, Speech and Language Processing*, 2008.
- [34] E. S.-X. Grigorios Tsoumakas and J. Vilcek, “Mulan: A java library for multi-label learning,” in *Journal of Machine Learning Research*, 2011.