Sentiment Analysis for Kannada using Mobile Product Reviews : A Case Study

Yashaswini Hegde
SJCE, Mysore
yhegde@gmail.com

S.K. Padma
Department of Information Science
SJCE, Mysore
padmask@gmail.com

***Abstract*---Sentiment Analysis (SA) is a very popular research area in the field of text mining as its computational capabilities have found a lot of research applications. Sentiment Prediction,Subjectivity Detection,Text summarization,Sentiment Summarization for Opinions etc. are some example applications. There are many research studies in the area of SA in different languages. However, Kannada SA has not been explored extensively and in particular, for the analysis of product reviews. In this paper, a case study of Kannada SA for mobile product reviews is proposed as there are many user generated Kannada product reviews available online. In this approach a lexicon based method for aspect extraction has been developed. Furthermore, the Naive Bayes classification model is applied to analyze the polarity of the sentiment due to its computational simplicity and stochastic robustness. This is the first attempt in Kannada to the best of author's knowledge. Therefore, a customized corpus has been developed. The weekly reviews from the column 'Gadget Loka' by U.B Pavanaja are considered to develop this corpus. The source for this is the famous Kannada news paper 'Prajavani'. The preliminary results indicate this approach is an efficient technique for Kannada SA.**

***Index Terms*---Sentiment Analysis,Product Reviews, Kannada, NLP, ಕನ್ನಡ, ಭಾವನೆಗಳ ವಿಶ್ಲೇಷಣಾ ತಂತ್ರ**

1. Introduction

Sentiment Analysis (SA) is a computational methodology of identifying,characterizing,extracting the sentiment contents such as attitudes,emotions,opinions,subjective impressions in the text,speech or databases. SA uses Natural Language Processing (NLP),Statistics and Machine learning techniques. SA is also called Opinion Mining, almost as a synonym. [2] [3] [4]. Sentiment Analysis is possible at document,sentence or feature/aspect level and it includes the study of

- Subjectivity Detection (detects opinionated text)
- Sentiment Prediction (predicts the positive or negative polarity of the text) and
- Aspect Based Sentiment Summarization (star ratings/scores of features of the product, summarized as sentiment)
- Text summarization for Opinions(summarizes the text content)
- Contrastive Viewpoint Summarization (contradicting opinions are emphasized)
- Product Feature Extraction (product features are extracted from the review)
- Detecting opinion Spam (Identifies fake or bogus opinion from reviews.)

Regular opinions and Comparative opinions are the two types of sentiment opinions [5]. "A regular opinion is a sentiment about a particular entity or an aspect of the entity" [1], for e.g.,"Moto G is a good business phone", where as "the comparative opinion compares multiple entities based on their shared aspects" [1], e.g., “Moto E as good as a business phone like Moto G at lower price”. There are two ways to approach SA,

- Machine learning approach: Uses classifiers like Naive Bayes,Maximum entropy and Support vector machine.
- Lexicon based: Uses sentiment dictionary with opinion words to match them with the data to determine positive or negative sentiments. There are some issues using sentiment lexicons [1] which are listed below.
  * There are phrases and idioms which give implied meaning other than sentiment words.
  * Sentiment lexicon is required but not a must for sentiment analysis. Because sometimes same word can impart negative and positive meaning as in, "This phone sucks" and "This phone really sucks".
  * Sentences containing sentiment lexicons may not express any sentiment. e.g,"How good is this phone?".
  * Sarcastic sentences are very hard to analyze though they contain sentiment words. e.g, "What a great phone!"
  * Objective sentences without sentiment words can imply factual informative opinions. e.g, "The phone cover is loose, it comes out"

  There are three methods to construct a sentiment lexicon: manual construction, corpus-based methods and dictionary-based methods.
  * Manual extractions of lexicons
  * Corpus-based methods: Corpus based techniques rely on syntactic patterns in large corpora. Corpus-based methods can produce opinion words with relatively high accuracy. Most of these corpus based methods need very large labeled training data,to find the orientation of domain specific opinion words.
  * Dictionary-based methods: Opinion words with known polarities are manually collected,later WordNet dictionary is searched for their antonyms and synonyms to grow this word-set bigger.

978-1-4799-8047-5/15/$31.00 ©2015 IEEE

Sentiment analysis also can be viewed as an NLP problem; co reference resolution, negation handling, and word sense disambiguation are the problems to be solved in NLP, do also appear as problems in SA. But SA is limited in that sense compared to NLP, only a few aspects of sentences or documents are sufficient to analyze.

Finally research in this field either addresses classification of sentiments or detecting subjectivity of the documents which may not be sufficient for all applications. Hence it is required to do much more deeper analysis which is a big challenge in SA, generally, and also in Kannada NLP and SA.

In section no 2 we are providing a survey of sentiment analysis in general, in Section no 3, sentiment analysis in Kannada, in section no 4 our approach of Kannada Product reviews using sentiment analysis. And finally in 5th section, evaluation of the algorithm is discussed.

2. Related Background

The man kind is seeing a huge amount of data with opinion in the social media and on the Web. Hence it is not a surprise, if the growth of sentiment analysis and the social media go hand in hand on the internet,Web. Many researchers have been contributing to this area of research, hence in this section, we are mainly considering a survey of classification and aspect based sentiment analysis, some of the works in other Indian languages and finally survey of Kannada sentiment analysis.

A. *Classification based Sentiment Analysis*

Classification of sentiment is generally considered as classification problem of two-class, a positive and negative. Normally product reviews are taken as training and testing data. Hence sentiment classification is a problem of text classification and any existing classification algorithms like Naive Bayes,SVM are used in literature. Joachims, 1999 [9], Shawe-Taylor et al., 2000 [12], Pang et al.,2002 [8] are the pioneers who have worked in this area. There are many papers in this area starting with Dave, Lawrence and Pennock, 2003 [11] who have worked with words in positive and negative reviews. Pang and Lee, 2004 [7] used minimum cut algorithm on a graph to determine classification of sentiments. Xia and Zong, 2010 used syntactic relations [10]. Contextual valence and sentiment shifters (words like don't) are used for classification in Kennedy et al., 2006 [13], Li et al., 2010 [6]. Li, Zhang et al., 2009, used a non-negative matrix factorization method [14]. Semi supervised algorithms are used in Dasgupta et al., 2009 [15], Li et al., 2011 [16], Zhou et al., 2010 [17]. Lexicon based self supervised algorithm is explored in Qiu et al., 2009 [18]. A dependency tree based learning method using conditional random fields was proposed by Nakagawa et al., 2010 [19]. Graph based hashtag method was proposed by Wang et al., 2011 [20]. Bespalov et al., 2011 [21], used latent n-gram analysis for supervised sentiment classification. Yulan, He, Lin and Alani 2011, have worked in cross-domain classification problem; to recognize opinion topics they worked on joint topic modeling and then bridged them [22]. Bollegala et al., 2011 [23], worked on a technique to create thesaurus automatically, for sentiment sensitive words with unlabeled and labeled data from multiple sources to determine word associations, that gives similar kind of sentiments in different domains. Those words are used to train the classifier later. Wei and Pal in 2010 [24] proposed a structural correspondence learning (SCL) method, for cross-language sentiment analysis, since a direct machine translation is difficult.

B. *Aspect based Sentiment Analysis*

In aspect based sentiment analysis, unlike document based one,"the opinion target is decomposed into entity and its aspects", according to Pang[7]. Researchers have identified two major tasks in this type of SA namely **Aspect Extraction** which extracts aspects to evaluate **Aspect sentiment classification** which determines opinions such as positive, negative, or neutral for different aspects [1]. In Aspect sentiment classification again we have two approaches - **supervised learning** and **lexicon based**.
Wei and Gulla in 2010 [25], have proposed hierarchical classification model, and addressed the coverage of aspect's interest in the sentence. Jiang et al., 2011 [26], used dependency parser to generate a set of aspects for classification.
Supervised learning based SA can do better at document level than at sentence level hence it is not used in much applications. Lexicon based-are normally unsupervised. This technique uses sentiment lexicons, composite expressions, rules of opinions, sentiment shifters, but-clauses etc. to determine the sentiment orientation. Generally this approach has four steps according to Ding, Liu and Yu, 2008 [27], namely

- Mark sentiment words and phrases
- Apply sentiment shifters
- Mark sentiment words and phrases
- Aggregate opinions

C. *Sentiment Analysis in other Indian Languages*

Many researchers have worked in other Indian Languages like Hindi, Tamil and Malayalam. Among them Aditya Joshi et al., 2010 [29], have studied different approaches for Hindi sentiment analysis taking Hindi movie reviews as a case study. They have tried to translate Hindi to English to train the classifiers in their first approach. However in their second approach they have created a lexical resource Hindi-SentiWordNet and implemented a majority score based strategy for classification. Mittal and Agarwal in 2013, worked out Hindi sentimental analysis, with impact of the negation and discourse rules [28]. Sneha Mulatkar in 2014 [30], worked on Hindi sentiment classification with an emphasis on sense based features than words based features. Sudhakar and Bensraj in 2014 [31], have worked in sentiment analysis for Tamil. They have considered Tamil Dooradarshan news and developed Fuzzy Neural Network (FNN) to classify a sentence based on emotions. Janardhanan P.S. Nair et al., 2014[32], have worked on Malayalam text, which is domain-specific, to extract the mood at sentence-level. The process of mood extraction, included POS tagging, pattern extraction from the input sentence which contains adjective, adverb etc., as they determine the mood of the sentence.

D. Work Related to Kannada Sentiment Analysis

There are a very few research studies that are related to Kannada sentiment analysis. The reported work of R. Jayashree et al., in 2013 [33], is an effort to test the paragraph level text classification, using Naïve Bayesian methods for Kannada language. They have used Minimum term frequency, stop word identification and elimination methods to achieve dimensionality reduction. They have tested their algorithm on a custom built corpus called TDIL (Technology for Development of Indian Languages), a comprehensive Kannada text resource developed by Central Institute of Indian Languages(CIIL).

3. Kannada Sentiment Analysis - Issues

Like in all sentiment analysis of other languages, in kannada also the main issue is to figure out the sentiment/opinion word because obviously they are instrumental to sentiment analysis. (All the subsequent Kannada transliterations are 'Baraha' [35] encripted.)

In kannada 'chennaagide' (ಚೆನ್ನಾಗಿದೆ), 'truptikaravaagide' (ತೃಪ್ತಿಕರವಾಗಿದೆ), 'adbhutavaagide' (ಅದ್ಭುತವಾಗಿದೆ), 'uttamavaagide' (ಉತ್ತಮವಾಗಿದೆ), 'atyuttama' (ಅತ್ಯುತ್ತಮ), 'tumbaa chennaagide' (ತುಂಬಾ ಚೆನ್ನಾಗಿದೆ)etc. gives **positive opinion** and 'chennaagilla' (ಚೆನ್ನಾಗಿಲ್ಲ), 'truptikaravaagilla' (ತೃಪ್ತಿಕರವಾಗಿಲ್ಲ), 'uttamavaagilla' (ಉತ್ತಮವಾಗಿಲ್ಲ), 'aShTakkaShTe' (ಅಷ್ಟಕಷ್ಟೆ), 'samarpakavaagilla' (ಸಮರ್ಪಕವಾಗಿಲ್ಲ), 'hELikoLLuvaMtilla' (ಹೇಳಿಕೊಳ್ಳುವಂತಿಲ್ಲ) etc. gives **negative opinion**. "A list of such words and phrases is called a sentiment lexicon (or opinion lexicon)" according to Bing.[1].

Though the sentiment words are very important in identifying the polarity of the opinion they are just not sufficient for the following reasons.

- The sentiment lexicons may mislead the polarities. : eg: "emtha adbhtavaada phonu" (ಎಂಥಾ ಅದ್ಭುತವಾದ ಫೋನು !) may give a negative meaning. This kind of statements are difficult to handle. But generally reviews don't carry this sort of sarcasm.
- Presence of statements which would imply some polarity. : eg: " samsung avara app gaLa mElE beraLu aaDisidare vishESha pratikriye neeDuttade"; (ಸಾಂಸ್ಯಂಗ್ ಅವರ ಆಪ್ ಗಳ ಮೇಲೆ ಬೆರಳು ಆಡಿಸಿದರೆ ವಿಶೇಷ ಪ್ರತಿಕ್ರಿಯೆ ನೀಡುತ್ತದೆ.) "ee phone nalli ellaa namUneya aaTagalannu aDetaDeyillade aaDabahudu" (ಈ ಫೋನ್ ನಲ್ಲಿ ಎಲ್ಲಾ ನಮೂನೆಯ ಆಟಗಳನ್ನು ಅಡೆತಡೆಯಿಲ್ಲದೆ ಆಡಬಹುದು.) both gives positive opinion yet not using any sentiment lexicon.
- There are some sentences which uses positive/negative lexicons but impart opposite polarity. :eg: "ee phone chennaagide aadare dubaari" (ಈ ಫೋನ್ ಚೆನ್ನಾಗಿದೆ ಆದರೆ ದುಬಾರಿ), "olleya business phone aadare idu neeDuva savalattige bele jaasti" (ಒಳ್ಳೆಯ ಬಿಸಿನೆಸ್ ಫೋನ್ ಆದರೆ ಇದು ನೀಡುವ ಸವಲತ್ತಿಗೆ ಬೆಲೆ ಜಾಸ್ತಿ), "ear phone, bluetooth, camera guNamatta atyttamavalladiddarU idu oLLeya business phone" (ಇಯರ್ ಫೋನ್, ಬ್ಲೂಟೂತ್, ಕ್ಯಾಮೆರ ಗುಣಮಟ್ಟ ಅತ್ಯುತ್ತಮವಲ್ಲದಿದ್ದರೂ ಇದು ಒಳ್ಳೆಯ ಬಿಸಿನೆಸ್ ಫೋನ್) etc.
- Presence of comparative statements. :eg: "samsung, iphone ge paipOTi neeDaballudu" (ಸಾಂಸ್ಯಂಗ್, ಐಫೋನ್ ಗೆ ಪೈಪೋಟಿ ನೀಡಬಲ್ಲದು), "idE shReNiya nokia lumiya gaLige hOlisidare idara kaaryaacharaNa vyavasthe uttama ennabahudu" (ಇದೇ ಶ್ರೇಣಿಯ ನೋಕಿಯ ಲ್ಯುಮಿಯಗಳಿಗೆ ಹೋಲಿಸಿದರೆ ಇದರ ಕಾರ್ಯಾಚರಣ ವ್ಯವಸ್ಥೆ ಉತ್ತಮ ಎನ್ನಬಹುದು) etc.

all of these are challenging work to figure out the polarity of the statement

Apart from all these generic issues , there are some specific ones. For example in all reviews we may not get either positive or negative review but a conditional. For eg:"neevu iphone koLLuva yOcaneyalliddare adakkiMta kaDIme beleya samsung galaxi alpha vannu koLLabahudu" (ನೀವು ಐಫೋನ್ ಕೊಳ್ಳುವ ಯೋಚನೆಯಲ್ಲಿದ್ದರೆ ಅದಕ್ಕಿಂತ ಕಡಿಮೆ ಬೆಲೆಯ ಸ್ಯಾಂಸಂಗ್ ಗಾಲಾಕ್ಸಿ ಆಲ್ಫವನ್ನು ಕೊಳ್ಳಬಹುದು). Such kind of statements neither give positive nor negative polarity on either samsung galaxi alpha or iphone. It rather targets the user. Classification of such sentiments are also challenging. And also cross language reviews and multi class sentiments.

4. Our Approach

In our approach of Kannada SA, we have chosen a lexicon based approach where we have used lexicon entity models for aspect extraction. We have implemented a computationally simple and statistically robust, Naive Bayes classifier(algorithm) which is used to classify the sentiments of the mobile product reviews in Kannada. We have considered mobile, weekly product reviews -'GadgetLoka' [34] from famous Kannada daily 'Prajavani' by U.B Pavanaja. We used this to create the corpus in this study. We have trained the base classifier, with Kannada text as input, represented in unicode format. This approach is applicable for both categorical and continuous aspect values. The input model is 'model of entity' and is defined as follows.

Model of entity: An entity E_i is represented as a finite set of aspects A_i = {'a_1', 'a_2', ..., 'a_n'} and its values V_i = { v_1', 'v_2', ..., 'v_n'}. For example in our case they are A_i={ವೇಗ(ಗಿಗಾಹರ್ಟ್ಸ್), ಹೃದಯ, ...} and V_i= {2.5,ನಾಲ್ಕು,....}. The entire set is given in table 4.1.

The entity model generated like this, serves as the training set for the Naive Bayes classifier. It predicts the polarity of the sentiment of the new entity model extracted from the mobile product review. We estimated conditional probabilities Pr(A|B) by $n_c|n$, where n_c and n are the number of times AB and B occurred respectively in training data. This can lead to erroneous estimation if n_c = 0. To avoid this a non zero prior estimate 'p' is used to compute Pr(A|B) (as in equation 1). The number m is the measure of confidence on the prior estimate p. With this the probability equation becomes

$$Pr(A|B) = \frac{(n_c + m * p)}{(n + m)} \tag{1}$$

In our training set, since there are more number of positive records than negatives, we have set the value of p = 2/3 for positive class i.e ಕೊಳ್ಳಬಹುದು, p = 1/3 for negative class i.e ಕೊಳ್ಳುವುದು ಬೇಡ and m=1. This algorithm 1 is developed in Python. This algorithm scales linearly.

Table 4.1
Aspects and their values of different phones: A sample input

Aspects/Features	Phones	
	One Plus	**Philips Xenium W6610**
ವೇಗ(ಗಿಗಾಹರ್ಟ್ಸ್)	2.5	1.3
ಹೃದಯ	ನಾಲ್ಕು	ನಾಲ್ಕು
ಮೆಮೊರಿ(ಗಿಗಾಬೈಟ್)	೧೬	೧ + ೪
ಕಾರ್ಯಾಚರಣ ವ್ಯವಸ್ಥೆ	ಆಂಡ್ರಾಯಿಡ್ 4.4.4 ಸಯನೋಜನ್	ಆಂಡ್ರಾಯಿಡ್ 4.2
ಪಿಕ್ಸೆಲ್ ರೆಸೊಲೂಶನ್	1●8● x 192●	54● x 96●
ಪರದೆ/ಟಚ್ ರೆಸ್ಪಾನ್ಸ್	ಸ್ಪರ್ಶಸಂವೇದಿ	ಸ್ಪರ್ಶಸಂವೇದಿ
ನೆಟ್‌ವರ್ಕ್ ಬೆಂಬಲ	-	2ಜಿ
ಕ್ಯಾಮೆರಾ(ಮೆಗಾಪಿಕ್ಸೆಲ್)	13	8
ವಿಡಿಯೊ	ಹೈಡೆಫಿನಿಶನ್	ಹೈಡೆಫಿನಿಶನ್ ಇಲ್ಲ
ವೈಫೈ	ಇದೆ	ಇದೆ
ಬ್ಲೂಟೂತ್ ಸಂಪರ್ಕ	ಇದೆ	ಇದೆ
ಯುಎಸ್‌ಬಿ	ಆನ್ ದ ಗೋ ಇದೆ	ಇದೆ
ಜಿಪಿಎಸ್	ಇದೆ	ಇದೆ
ಬ್ಯಾಟರಿ(mAh)	31●●	53●●
ಗಾತ್ರ(ಮಿ.ಮೀ.)	152.9 x 75.9 x 8.9	145.4 x 74.1 x 11.4
ತೂಕ (ಗ್ರಾಂ)	162	2●●
ವಿನ್ಯಾಸ/ರಚನೆ	ತುಂಬ ಚೆನ್ನಾಗಿದೆ	ಚೆನ್ನಾಗಿದೆ
ಎಫ್‌ಎಂ ರೇಡಿಯೊ	ಇಲ್ಲ	ಇದೆ
ಸಂಗೀತ ಮತ್ತು ಧ್ವನಿಯ ಗುಣಮಟ್ಟ/ಆಡಿಯೊ	ಚೆನ್ನಾಗಿಲ್ಲ	ಉತ್ತಮವಾಗಿಲ್ಲ
ಇಯರ್‌ಫೋನ್	ಇಲ್ಲ	ಅಷ್ಟಕ್ಕಷ್ಟೆ
ಆಟ ಆಡುವ ಅನುಭವ	ಚೆನ್ನಾಗಿದೆ	ಅಡೆತಡೆ ಅನ್ನಿಸಲಿಲ್ಲ
ಕ್ಯಾಮೆರಾ ಗುಣಮಟ್ಟ	ಚೆನ್ನಾಗಿದೆ	ಅಷ್ಟಕ್ಕಷ್ಟೆ
ವಿಡಿಯೊ ವೀಕ್ಷಣೆ/ಗುಣಮಟ್ಟ	ಅತ್ಯುತ್ತಮ	ಪರವಾಗಿಲ್ಲ
ಬೆಲೆ	21,373	15,6●●
ಕ್ಲಾಸ್	**ಕೊಳ್ಳಬಹುದು**	**ಕೊಳ್ಳುವುದು ಬೇಡ**

Algorithm 1: **NB-Classifier algorithm for Kannada**

procedure NB-KAN
    Identify Kannada lexicons from mobile phones reviews in Kannada.
    Create flat files as record data type with identified lexicons as features for each type of mobile phones.
    Generate Kannada features and its associated values in unicode as training sets.
    Compute prior probabilities and class conditional probabilities for the training set.
    Apply Naive Bayes classifier algorithm to figure out post probability of the test set.
    Predict the class of the test set of mobile using computed post probability.
end procedure
end

A. *Limitation of the Algorithm*

Some of the limitation of this algorithm are:

1) Multi class is not yet addressed.
2) Comparative and conditional statements not considered yet.
3) Not handling sentiment shifts.

5. Evaluation of NB-Classifier Algorithm

We have evaluated the efficiency of our algorithm using our Kannada product review corpus with 5 fold cross validation method. In this approach we have divided the data into five sets and during each run we have taken one part for testing and the rest for training the model. We have used the evaluation measures that are standard and popular which are defined using confusion and cost matrix. The confusion and cost matrix is given in table 5.1.

A. *Performance Measures*

1) **Accuracy:** is a measure of predictive capability of the model.

$$Accuracy = \frac{(c_{TP} + c_{TN})}{(c_{TP} + c_{FN} + c_{FP} + c_{TN})} \tag{2}$$

Accuracy has a limitation, it gives good accuracy even if the classifier fails to identify the objects belong to any one of the class.

2) **Cost sensitive measures:** are

Precision is also called as Positive predictive value.

$$Precision(PPV) = \frac{c_{TP}}{(c_{TP} + c_{FP})} \tag{3}$$

$$Sensitivity(Recall) = \frac{c_{TP}}{(c_{TP} + c_{FN})} \tag{4}$$

$$Specificity = \frac{c_{TN}}{(c_{TN} + c_{FP})} \tag{5}$$

$$NegativePredictiveValue(NPV) = \frac{c_{TN}}{(c_{TN} + c_{FN})} \tag{6}$$

$$F - Measure = \frac{(2c_{TP})}{(2c_{TP} + c_{FN} + c_{FP})} \tag{7}$$

All the above mentioned evaluation measures are developed in python.

Table 5.1
CONFUSION AND COST MATRIX

	Predicted Class		
	c(i/j)	Class=**Yes**	Class=**No**
Actual Class	Class=**Yes**	C(Y/Y) c_{TP}	C(N/Y) c_{FN}
	Class=**No**	C(Y/N) c_{FP}	C(N/N) c_{TN}

6. RESULTS

The output of our classifier includes aspect entity, its values, its individual positive and negative probabilities, missing values if any, and final cross conditional probabilities. Our algorithm finally identifies the polarity of the sentiment which is same as the class label. The snapshot of the output is given in the table 6.1. From this snapshot we can see that the identified polarity of the test vector is negative.

Table 6.1
SNAPSHOT OF THE OUTPUT

......
v= ಏನೇನೂ ಚೆನ್ನಾಗಿಲ್ಲ probp= 0.333333333333 prodpos= 1.34046677987e-10
v= ಏನೇನೂ ಚೆನ್ನಾಗಿಲ್ಲ probn= 0.4 prodneg= 1.02597629011e-07
no key in pos computation
no key in pos computation 2.97881506638e-11
v= ಚೆನ್ನಾಗಿದೆ probn= 0.4 prodneg= 4.10390516044e-08
finl probb= 2.97881506638e-11 4.10390516044e-08
the test vector belongs to negative class

The confusion matrix obtained as an output of our classifier is tabulated in the table 6.2.

Table 6.2
CONFUSION MATRIX OF OUR ALGORITHM CONSIDERING 20 POSITIVE AND NEGATIVE TRAINING SET

	Predicted Classs		
	c(i/j)	Class=**Yes**	Class=**No**
Actual Class	Class=**Yes**	c_{TP}=15	c_{FN}=5
	Class=**No**	c_{FP}=9	c_{TN}=11

Table 6.3
PERFORMANCE MEASURES OF CLASSIFIERS

Measure	Evaluation Result in %
Accuracy	65
Precision(p)	62.5
Sensitivity(Recall)	75
Specificity	55
NPV	68.75
F-Measure	68.2

The performance measures are computed using the values in the table 6.2 and equations 2, 3, 4, 5, 6, 7 and are tabulated in the table 6.3. These values are the average of each kth run.

The table 6.4 gives the comparison of our results with the results of other languages based on the results published by the authors. However a proper comparison of apple to apple is difficult due to different data sets and different methods/algorithms and also different performance measures used by the other authors.

Table 6.4
COMPARATIVE STUDY OF OUR RESULTS WITH RESULTS OF OTHER LANGUAGES

Language	Data Set	Method	Results
Hindi	Hindi SentiWordNet	Negation and Discourse rules	80.21
Hindi	Hindi Movies with HindiSentiWordNet	RapidMiner with SVM	78.14
Tamil	Tamil Doordarshan news	Fuzzy Neural Network	55.78
Kannada	Mobile Reviews	Naive Bayes	65

7. CONCLUSION

Our results indicate that the lexicon based aspect extraction with Naive Bayes sentiment classifier works efficiently for Kannada (unicode) sentiment analysis. From the results, we can see that our method is performing with 65 % accuracy, 62.5% precision (PPV), 75% (sensitivity)recall, 55% specificity, 68.75% NPV and 68.2 % F-Measure. Our final objective is to apply this method for Kannada text summarization. However initially we have chosen a simple but robust method. Our future research warrants addressing some of the limitations of the algorithm described in section 4A with improved accuracy. Moreover, in the future we propose to apply other sophisticated methods of aspect extractions that are used in natural language processing (with larger corpora).

ACKNOWLEDGEMENTS

We would like to thank Dr.U.B.Pavanaja for his series of review articles published in 'Prajavani' Kannada daily. We also would like to thank Dr.Shubha Kadambe for providing valuable suggestions.

References

[1] Bing Liu, Sentiment Analysis and Opinion Mining, 2012.
[2] Vohra,Teriaya "A COMPARATIVE STUDY OF SENTIMENT ANALYSIS TECHNIQUES "
[3] Medhat W et al., Sentiment analysis algorithms and applications: A survey, Ain Shams Eng J (2014), http://dx.doi.org/10.1016/j.asej.2014.04.011
[4] Mullen, T., Collier, N. (2004). “Sentiment analysis using support vector machines with diverse information sources”. In Proceedings EMNLP’04,pp.412-418.
[5] Jindal, Nitin and Bing Liu. Mining comparative sentences and relations. in Proceedings of National Conf. on Artificial Intelligence (AAAI-2006). 2006b.
[6] Li, Shasha, Chin-Yew Lin, Young-In Song, and Zhoujun Li. Comparable entity mining from comparative questions. in Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010).2010.
[7] Pang, Bo and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. in Proceedings of Meeting of the Association for Computational Linguistics (ACL-2004).2004.
[8] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?:sentiment classification using machine learning techniques. in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP-2002). 2002.
[9] Joachims, Thorsten. Making large-Scale SVM Learning Practical, in Advances in Kernel Methods - Support Vector Learning, B. Schölkopf, C.Burges, and A. Smola, Editors. 1999, MIT press.
[10] Xia, Rui and Chengqing Zong. Exploring the use of word relation features for sentiment classification. in Proceedings of Coling 2010: Poster Volume. 2010.
[11] Dave, Kushal, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews in Proceedings of International Conference on World Wide Web (WWW-2003). 2003.
[12] Shawe-Taylor, John and Nello Cristianini. Support Vector Machines, 2000, Cambridge University Press.
[13] Kennedy, Alistair and Diana Inkpen. Sentiment classification of movie reviews using contextual valence shifters. Computational Intelligence,2006. 22(2): p. 110-125.
[14] Li, Tao, Yi Zhang, and Vikas Sindhwani. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. in Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL-2009). 2009.
[15] Dasgupta, Sajib and Vincent Ng. Mine the easy, classify the hard: a semi-supervised approach to automatic sentiment classification. in Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP (ACL-2009). 2009.
[16] Li, Shoushan, Zhongqing Wang, Guodong Zhou, and Sophia Yat Mei Lee. Semi-Supervised Learning for Imbalanced Sentiment Classification. in Proceedings of International Joint Conference on Artificial Intelligence (IJCAI-2011). 2011.
[17] Zhou, Shusen, Qingcai Chen, and Xiaolong Wang. Active deep networks for semi-supervised sentiment classification. in Proceedings of Coling 2010: Poster Volume. 2010.
[18] Qiu, Likun, Weish Zhang, Changjian Hu, and Kai Zhao. Selc: a self-supervised model for sentiment classification. in Proceeding of the 18th ACM conference on Information and knowledge management (CIKM-2009). 2009.
[19] Nakagawa, Tetsuji, Kentaro Inui, and Sadao Kurohashi. Dependency tree-based sentiment classification using CRFs with hidden variables. in Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL (HAACL-2010).2010.
[20] Wang, Xiaolong, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach. in Proceeding of the ACM conference on Information and knowledge management (CIKM-2011). 2011.
[21] Bespalov, Dmitriy, Bing Bai, Yanjun Qi, and Ali Shokoufandeh. Sentiment classification based on supervised latent n-gram analysis. in Proceeding of the ACM conference on Information and knowledge management (CIKM-2011). 2011.
[22] He, Yulan, Chenghua Lin, and Harith Alani. Automatically extracting polarity-bearing topics for cross-domain sentiment classification. in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011). 2011.
[23] Bollegala, Danushka, David Weir, and John Carroll. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011). 2011.
[24] Wei, Bin and Christopher Pal. Cross lingual adaptation: an experiment on sentiment classifications. in Proceedings of the ACL 2010 Conference Short Papers (ACL-2010). 2010.
[25] Wei and Jon Atle Gulla. Sentiment learning on product reviews via sentiment ontology tree. in Proceedings of Annual Meeting of the Association for Computational Linguistics (ACL-2010). 2010.
[26] Jiang, Long, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent twitter sentiment classification. in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-2011). 2011.
[27] Ding, Xiaowen, Bing Liu, and Philip S. Yu. A holistic lexicon-based approach to opinion mining. in Proceedings of the Conference on Web Search and Web Data Mining (WSDM-2008). 2008.
[28] Mittal and Agarwal. Sentiment Analysis of Hindi Review based on Negation and Discourse RelationInternational Joint Conference on Natural Language Processing, pages 45–50,Nagoya, Japan, 14-18 October 2013
[29] Aditya Joshi,Balamurali A R, Pushpak Bhattacharyya. A Fall-back Strategy for Sentiment Analysis in Hindi: a Case Study, Proceedings of ICON 2010: 8th International Conference on Natural Language Processing Macmillan Publishers, India. Also accessible from http://ltrc.iiit.ac.in/proceedings/ICON-2010.
[30] Sneha Mulatkar. Sentiment Classification In Hindi,INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 3, ISSUE 5, MAY, ISSN 2277-86162014
[31] B.Sudhakar and R.Bensraj. An Efficient Sentence-based Sentiment Analys is for Expressive Text-to-speech using Fuzzy Neural Network , Research Journal of Applied Sciences, Engineering and Technology 8(3): 378-386, 2014 ISSN: 2040-7459; e-ISSN: 2040-7467
[32] Neethu Mohandas, Janardhanan P.S. Nair, Govindaru V., "Domain Specific Sentence Level Mood Extraction from Malayalam Text", ICACC, 2012, 2012 International Conference on Advances in Computing and Communications, 2012 International Conference on Advances in Computing and Communications 2012, pp. 78-81, doi:10.1109/ICACC.2012.16
[33] Jayashree R, Srikantamurthy K, Basavaraj S Anami. Suitability of Naïve Bayesian Methods for Paragraph Level Text Classification in the Kannada Language using Dimensionality Reduction Technique, International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 5, September 2013
[34] http://www.prajavani.net/columns/ಗ್ಯಾಜೆಟ್-ಲೋಕ-ಯುಬಿ-ಪವನಜ
[35] http://www.baraha.com/