An Analysis of Opinion Mining Research Works Based on Language, Writing Style and Feature Selection Parameters

Jasleen Kaur[1,2]
[1]Assistant Professor
[2]Research Scholar
[1]Shroff S. R. Rotary Institute of Chemical Technology
[2]Uka Tarsadia University,
[1]Ankleshwar, Gujarat, India
[2]Bardoli, Gujarat, India
sidhurukku@yahoo.com
Dr.Jatinderkumar R. Saini[1,2]
[1]Associate Professor
[2]Research Supervisor
[1]Narmada College of Computer Application
[2]Uka Tarsadia University,
[1] Bharuch, Gujarat, India
[2]Bardoli, Gujarat, India
saini_expert@yahoo.com

---ABSTRACT---

Different writing styles, either formal or informal, can be adopted to present the written text. A Piece of text may contain a lot of emotional states, feelings or ideas presented through the means of words and means of Language. Various techniques and methods are present in the field of Opinion Mining and Sentiment Analysis to extract the emotions from text. This paper presents analysis of Formal and Informal text pieces written in different International Languages in the field of Opinion Mining and Sentiment Analysis. This paper presents a study and analysis of differences of approaches used for Opinion Mining and Sentiment Analysis for both cases. Formal and Informal Text Pieces are present in 8 different International Languages (English, Chinese, Arabic, Malaysian, Spanish, Turkish, Persian , Korean).In this study Formal text pieces ,in form of poetry, proverbs ,essay and documents, and Informal text in form of micro blogs, chats, emails and SMS, are analyzed. Maximum performance for Opinion Mining in case of Informal Text is achieved in Arabic Language and for formal text; maximum accuracy is obtained in Persian and Turkish Language. In this study 4 different Feature selection parameters (IG, TF-IDF, n-gram, MI and MMI) were analyzed in order to find emotional states associated with written text. It was found that parameter, IG and TF-IDF, were experimented by Researchers maximum number of times and IG outperformed all other feature selectors.

Keywords - **Emotion Detection, Formal Text, Informal Text, Opinion Mining, Sentiment Analysis, Information Gain**

I. INTRODUCTION

Language is one of medium for communicating your views or messages. Written Text is one good source for expressing your ideas, emotions and feelings. Languages not only used for communication but also impart emotion associated with it. Feelings can be easily expressed in form of writing. Human Being has a power to feel different kinds of emotion because Life of every human being is filled with a lot of emotions. Joy, fear, anger and sadness are few emotional states that a person encounters in day to day life. And using computer, the categorization of text in these emotional states is known as Sentimental analysis/ Emotion Detection. Sentiment Classification is classifying the text according to the sentimental information associated with the text.

Gathering feeling or emotions associated with text is known as Opinion Mining. Opinion Mining is extracting the opinions from text. Mining Opinions associated with text can be useful to know the experience of user about a place, about any event or any product. Opinion Mining can be applicable to any kind of text. There is a minor difference between Sentiment Analysis and emotion detection. In general, Sentiment Analysis divides text into two binary states (positive/ negative) whereas Emotion Detection uses larger set of emotions for division of text. Number of emotional states like joy, fear, anger, brief, surprise or disgust is encountered in day to day life. The terms “Sentiment Analysis” and “Opinion Mining” represents the same field of study. These are synonyms and can be used interchangeable. Both of these areas can be considered as a sub area of subjectivity analysis as stated by Pang and Lee [6].

Text can be written using two writing styles: formal and informal writing style. Formal writing consists of Poetry, novels, plays, government/ official documents. And informal text consists of chat room data, short message on social media, SMS. As Literary arts contains a lot of emotions, these literature pieces especially poems can be used for task of Sentiment Classification which is very challenging in computational point of view. And secondly, short messages like tweets, face book status, are also become the useful source for Opinion Mining and their sentiment analysis. Because of length constraints in these kinds of messages, Opinion Mining is very difficult on this dataset.

2. LITERATURE SURVEY

A lot of work had been done on field of Opinion Mining and Sentiment Classification problem. Pang and Lee [6] discussed in detail various challenges that need to be dealt with while performing Sentiment Analysis/ Opinion Mining on different kinds of Data. Opinion Mining/Sentiment Analysis. Different types of datasets (blogs, movie reviews, chatting data, and micro blogging sites) can be used for Opinion Mining as discussed by Kaur and Saini [13]. Literary Pieces, lyrics and unsolicited bulk mails [12] can be mined for views/ feelings/emotions. Liu B. [5] presents different granularity levels of Opinion Mining task. Opinion Mining can be done at Document Level, Feature level and Sentence Level. Approaches used for Opinion Mining and Sentiment Classification are broadly categorized into Supervised and Unsupervised learning. Different Supervised Learning such as Support Vector Machine (SVM) [1,3,9,11,16,17,22,23,25], Naïve Bayes (NB) [1,3,4,17,19,23], K-Nearest Neighbor (KNN) [14,23,26], Maximum Entropy (ME) [3], Winnow Classifier [21] and Centroid[21] were experimented by different research on different kinds of dataset.

In this paper discussion involves two different types of writing styles: formal and informal. Similarly corpus for experimentation can be divided into: formal text corpus and informal text corpus. Brief surveys about both styles are presented here.

2.1. Formal text corpus

Formal Writing Style is followed in following pieces of text like literary arts (poems, novels, essay, plays etc.), official documents, and legal documents. Literary Arts are the one of the complex example of formal text. A brief survey of classification in formal text is presented below:

Barros et.al [15] tried to automatically categorize poems based on their emotional content. For this experiment, they have used a Quevedo’s poetry written in Spanish. A reference classification of the same (Bleuca’s Categorization) is also used during the experimentation. They had done the work in two parts: 1) to check the original manual classification could be distinguished in terms of sentiment reflected by poems. 2) Exploring automatic learning techniques to produce better results with this dataset. Decision Tree is built using Weka toolset for classification problem. The Accuracy of this classifier is 56.22%, which is increased to 75.13% by using resample filter. This experiment is done to determine whether a classifier with information about emotions detected in a given Quevedo’s poem can able to reproduce Bleuca’s Categorization.

Hamidi et.al [11] proposed a meter classification system for Persian poems based on features extracted from uttered poem. In the first stage, the utterance has been segmented into syllables using three features, pitch frequency and modified energy of each frame of the utterance and its temporal variations. In the second stage, each syllable is classified into long syllable and short syllable classes. In this stage, the classifier is an SVM classifier with radial basis function kernel and employed features are the syllable temporal duration, zero crossing rate and PARCOR coefficients of each syllable. The system has been evaluated on 136 poetries utterances from 12 Persian meter styles gathered from 8 speakers, using k-fold evaluation strategy. The results show 91% accuracy in three top meter style choices of the system.Support Vector Machine (SVM) based method is used to differentiate bold-and-unconstrained style from graceful-and-restrained style of poetry as presented in He [25]. In this work, a piece of poetry is expressed using Vector Space Model (VSM) first, and then information gain is used to select the poetry's feature terms. SVM-based method is used to divide the style of poetry by analyzing the influence of feature numbers and feature items for poetry style. The performance of the proposed method has been evaluated by a series of experiments; 10-fold cross validation an average accuracy 88.6% is achieved.

Jamal et.al [16] represents classification of Malay pantun using Support Vector Machines (SVM). Pantun is traditional Malay poetry. The capability of SVM through Radial Basic Function (RBF) and linear kernel functions are implemented to classify pantun by theme, as well as poetry or non-poetry. A total of 1500 pantun are divided into 10 themes with 214 Malaysian folklore documents used as the training and testing datasets. TF-IDF used for both classification experiments. The highest average percentage of 58.44% accuracy was found for the classification of poetry by theme. The results of each experiment showed that the linear kernel achieved a better percentage of average accuracy compared to the RBF kernel.Kumar and Minz [23], author works to find the best classification algorithms among the K-nearest neighbor (KNN), Naïve Bayesian (NB) and Support Vector Machine (SVM) with reduced features for classification of poems. Information Gain Ratio is used for feature selection. The results show that SVM has maximum accuracy (93.25 %) using 20 % top ranked features.

Alsharif et.al [17] tried to classify Arabic poetry according to emotion associated with it. The problem was treated as a text categorization problem, classifying poems into four classes: Retha, Ghazal, Heja and Fakhr. Four machine learning algorithms are compared: Naïve Bayes, SVM, VFI (Voting Feature Intervals) and Hyperpipes. The best precision achieved was 79% using Hyperpipes with non-stemmed, non-rooted, mutually deducted feature vectors containing 2000 features. Can et.al [9] investigated two fundamentally different machine learning text categorization methods, Support Vector Machines (SVM) and Naïve Bayes (NB), for categorization of Ottoman poems according to their poets and time periods. Dataset comprises of the collected works (divans) of ten different Ottoman poets. The Result shows that SVM, with almost 90% accuracy, is a more accurate classifier compared to NB in categorization tasks.

Li et.al [14] applied K-Nearest Neighbor (KNN) algorithm for text categorization to essays. In this paper, each essay is represented by the Vector Space Model (VSM). After removing the stop words, the words, phrases and arguments as features of the essays chosen and the value of each vector is expressed by the term frequency and inversed document frequency (TF-IDF) weight. The TF and information fain (IG) methods are used to select features by predetermined thresholds. Experiments on CET4 essays in the Chinese Learner English Corpus (CLEC) show accuracy above 76% is achieved.

Noah and Ismail [19] presented an experimental study on automatic Classification of Malay proverbs using Naïve Bayesian algorithm. One thousand training and testing dataset which were classified into five categories: family, life, destiny, social and knowledge are used for experiment. Two types of testing have been conducted; testing on dataset with stop words and dataset with no stop words by using three cases of Malay proverbs, i.e., proverb alone, proverb with meaning and proverb with the meaning and example sentences. The results showed that a maximum of 72.2 and 68.2% of accuracy have been achieved respectively by the Multinomial model and the Multivariate Bernoulli for the dataset with no stop words using proverb with the meaning and example sentences. Tan and Zhang [21] presented an empirical study of Sentiment categorization on Chinese documents. Four feature selection methods (MI, IG, CHI and DF) and five learning methods (Centroid Classifier, K-Nearest Neighbour, Winnow Classifier, Naive Bayes and SVM) are investigated on a Chinese sentiment corpus with a size of 1021 documents. The experimental results indicate that IG performs the best for sentimental terms selection and SVM exhibits the best performance for Sentiment Classification.

2.2. Informal Text Corpus

Large amount of informal text is present on World Wide Web. With the advent of World Wide Web this text is increasing day by day. Social media users can freely express their feelings, views, opinions and emotion on social networking site. A brief literature survey of Sentiment Analysis of text present on World Wide Web is presented below:

Cho and Kang [20] proposed a new approach to Sentiment Classification at paragraph length using contextual information. Contextual information such as keywords, the position of the sentence, and the flow of sentiment are computed in texts of multiple sentence length. They have considered four domains for this experiment which includes consumer product, travel, food and movie. To construct a test data set, texts corresponding to these classes are collected from Social Network Service such as Twitter, Face book and Me2Day. A feature vector for a given text is constructed from the contextual information and is then classified by the Support Vector Machine (SVM) classifier as positive, negative or neutral. Method performs well in classifying the sentiments expressed in the multiple texts of social media. The results reported by this experimentation is 0.85(F-score) for positive class, 0.76 for negative class and 0.70 for neutral class.

Samsudin et.al [18] performed opinion mining on online messages on face book, twitter present in Malaysian language. In Malaysia, online messages are written in mixed languages known as ‘Bahasa Rojak. This study introduced a Malay Mixed Text Normalization Approach (MyTNA) and a feature selection technique based on Immune Network System (FS-INS) in the opinion mining process using machine learning approach. The purpose of MyTNA was to normalize noisy texts in online messages. In addition, FS-INS will automatically select relevant features for the opinion mining process. Several experiments involving 1000 positive movies feedback and 1000 negative movies feedback was conducted. The results show that accuracy values of opinion mining using Naïve Bayes (NB), k-Nearest Neighbour (kNN) and Sequential Minimal Optimization (SMO) increase after the introduction of MyTNA and FS-INS.

Bagheri et.al [4] considers the problem of Sentiment Classification for online customer reviews in Persian language. Another common problem of Persian text is word spacing. In Persian in addition to white space as interwords space, an intra-word space called pseudo-space separates word's part. One more noticeable challenge in customer reviews in Persian language is that of utilizing many informal or colloquial in text. The proposed model is based on Persian language and is employed Naive Bayes learning algorithm for classification. And also presented a new feature selection method based on the mutual information method to extract the best feature collection from the initial extracted features. Finally they evaluate the performance of the model on a manually gathered collection of cell phone reviews, where the results show the effectiveness of the proposed model.

Pang et.al [3] present a work based on classic topic classification techniques. The proposed approach aims to test whether a selected group of machine learning algorithms can produce good result when Opinion Mining is perceived as document level, associated with two topics: positive and negative. He presented the results using Naive Bayes, maximum entropy and Support Vector Machine algorithms and shown the good results as comparable to other ranging from 71 to 85% depending on the method and test data sets.

Ding et.al [24] proposed a holistic lexicon-based approach to solving the problem by exploiting external evidences and linguistic conventions of natural language expressions. This approach allows the system to handle opinion words that are context dependent, which cause major difficulties for existing algorithms. A system, called Opinion Observer, based on the proposed technique has been implemented. Experimental results using a benchmark product review data set and some additional reviews show that the proposed technique is highly effective. It outperforms existing methods significantly with Average F score- is 0.90.

Poria et.al [22] introduced a novel paradigm to concept-level sentiment analysis that merges linguistics, common-sense computing, and machine learning for improving the accuracy of tasks such as polarity detection. A SVM and ELM classifiers were trained, over the training portion of the movie review dataset, using the sentence feature set and it was found that ELM outperformed SVM in terms of both accuracy and training time. Accuracy obtained was 67.35% accuracy with ELM and 65.67% on movie review dataset and obtained 72.00% accuracy with ELM on second dataset but a much lower accuracy with SVM.

Silva [1] presented a novel approach which analyzes the email flow of emotions by extracting multiple sentiments at sentence level and categorizing using emotion based dictionaries. A multi class approach that assigns a label to each email considering the overall flow of sentiments found inside of it.SVM turns out to be the most stable learning algorithm, and also provided the best accuracy (62 %).

3. OPINION MINING ANALYSIS IN DIFFERENT LANGUAGES

Number of factors affects the performance of Opinion Mining problem. Data corpus, feature selection algorithm, corpus size, language involved and writing style are few of them. Kaur and Saini [2] had discussed about influence of writing style on sentiment classification problem. Two different writing styles i.e. Formal and Informal writing style are present in the literature. Formal and writing styles are entirely different from each other. Formal writing style is adopted in formal communication like research documents, whereas informal writing style is used in our day to day communication like SMS. Formal writing style is adopted in official documents, research papers, literary arts (i.e. poetry, novels, plays and stories). And Informal writing style includes casual English sentences or phrases without any constrain. This type of style is adopted in every day to day conversation, real time data (like tweets, face book statuses etc), emails, chatting data etc.

This formal or Informal text written in which language also affects the accuracy of opinion mining problem. To find the opinion associated with any document, word or sentence written in any language requires the use of various natural language processing tasks. In general, Stemming, POS tagger, word sense disambiguation, stop word removal are the various preprocessing tasks that need to be done on the corpus for sentiment analysis. Each Language has its own character set and its characteristics, depending on that, preprocessing step may vary. Sentiment classification problem become challenging for sarcastic languages and resource scare languages. Different techniques and approaches used for mining Opinions written in different International Languages are shown in table –I. A lot of research work in field of opinion mining can be found in the literature for Languages for which resources, like tools for preprocessing task and corpus, are available. English and Chinese (& its Variations) are the languages which are widely chosen by Researchers for Opinion Mining problem because of availability of lot resources for the same. Whereas there are many other languages like Malaysian and Indian Languages, for which Natural language processing world is still far away.

Various Feature selection algorithms are present in the literature as shown in Table –I. Information Gain (IG), Mutual Information (MI), Modified Mutual Information (MMI), Term Frequency, and Inverse Document Frequency(TF-IDF) and n-gram(unigram and bigram) were experimented by various researchers as shown in Table-I. IG tends to perform better for sentimental term selection in both the writing styles, Formal and Informal, irrespective of language used.

TABLE I. DIFFERENT WORKS IN FIELD OF OPINION MINING IN DIFFERENT LANGUAGES

Language	Writing style	Author	Approach	Feature selection/ Algo	Corpus	Performance
English	Formal	Kumar and Minz [23]	SVM,NB, KNN	IG, TF-IDF	www.poetseer.org www.poetry.org	Accuracy:93.25%
		Li [14]	KNN	IG,TF-IDF	CLEC	Accuracy:76%
	Informal	Silva [1]	NB, Decision Trees, AdaBoost, SVM, Random Forest	-	Enron corpus	Accuracy:62%
		Pang [3]	SVM,NB,ME	Unigram, bigram	Movie review corpus	Accuracy: 78.7 (unigram), 77.1 (bigram)
		Poria et.al [22]	SVM,ELM	Common Knowledge, negation,	Movie review , product review dataset	Accuract:67.35(movie dataset), 72.00 (product reviews)
		Ding et.al [24]	Lexicon based approach	Opinion Observer	Movie reviews	F-score: 0.90
Chinese	Formal	He [25]	SVM	IG	Song-Ci	Accuracy:88.6%
		Tan and Zhang [21]	SVM, KNN, Winnow, centroid	MI.,IG,CHI, DF	Chinese documents	F-score: 0.8664
	Informal	Lee and Renganathan [26]	ME	-	http://product.it168.com	81.65
Arabic	Formal	Alsharif [17]	SVM,NB,VFI, Hyperpies	Mutualdeduction function	Arabic poetry	Precision:0.791
	Informal	Al-Kabi et.al.[10]	NB,SVM,KNN	TF	Arabic reviews	Accuracy: 96.9%
Malaysian	Formal	Jamal [16]	SVM-RBF and LFK	TF-IDF	Malay Pantum	Accuracy:58.44%
		Noah and Ismail [19]	NB(multinomial and multivariate Bernoulli model)	IG	Malay Proverbs	Accuracy: 72.2 and 68.2% (respectively in each model)
	Informal	Samsudin et.al [18]	KNN	FS-INS	Online messages	Accuracy : 69.09 %
Spanish	Formal	Barros [15]	Decision Tree	IG	Quevedo's poem	Accuracy : 56.22% and after resample 75%
	Informal	Sidorov et.al[8]	SVM, NB, Decision Tree	-	Spanish Tweets	Accuracy :81.6 %
Persian	Formal	Hamidi [11]	SVM-RBF	zero crossing rate, pitch frequency	Persian poems	Accuracy: 91%
	Informal	Bagheri [4]	NB	MI,TFV,MMI	Customer reviews on Cell phones	Accuracy :85%
Turkish	Formal	Can [9]	SVM,NB	Style Marker (TOL,TYL,MWF,TWC)	Divans, poems written in ottoman	Accuracy:90%
	Informal	Esra [7]	SVM,NB,Decision Tree,		Turkish Tweets	Accuracy: 86.49%
Korean	Informal	Cho and Kang [20]	SVM	TF-IDF	Twitter, face book, me2day	F-Score: 0.85

Abbreviation: SVM(Support Vector Machine), NB(Naive Bayes), ME(Maximum Entropy), IG(Information Gain), TOL-TYL (Token & Type Length), TWC(Two-word Collocations), MFW(Most Frequent Words),MI (Mutual Info), DF (Document Frequency), CLEC(Chinese English Learner Corpus), TFV (Term frequency variance), MMI (Modified version of Mutual Information),TF-IDF(Term frequency Inverse Document Frequency), ELM(Extreme Learning Machine).

4. FINDINGS

Various Techniques used by authors for classification problem are shown in Table-I. Support Vector Machine, Naive Bayes, K-Nearest Neighbour, Decision Tree and various other approaches (Hyperpipes, Winnow, Centroid, VFI, Random Forest, ELM) were used for classifying sentiments in both the writing styles (Formal and Informal Text).Opinion mining task done in 8 International Languages (English, Chinese, Arabic, Malaysian, Spanish, Persian, Turkish, Korean) is considered for analysis. Performance of Classification can be measured in accuracy, F-measure, Recall and Precision. As Accuracy is specified in percentage and remaining three are in decimal format, so all decimal numbers are converted directly into percentage. Performance Accuracy obtained in these 8 languages in field of Opinion Mining (irrespective of Techniques used) with the assumption stated above is shown in Table-II.

TABLE I. PERFORMANCE COMPARISON OF DIFFERENT LANGUAGES IN FORMAL AND INFORMAL WRITING STYLE

Sr. No.	Language	Accuracy	
		Formal	Informal
1	English	84.62	74.52
2	Chinese	87.72	81.65
3	Arabic	79.00	96.90
4	Malaysian	66.28	69.09
5	Spanish	65.61	81.60
6	Persian	91.00	85.00
7	Turkish	90.00	86.49
8	Korean	-	85.00

Graphical Representation of Table-II is shown in Fig. 1. Maximum average Accuracy in field of Opinion Mining is achieved in Arabic, Persian and Turkish Language. But very little work is done in Persian and Turkish.

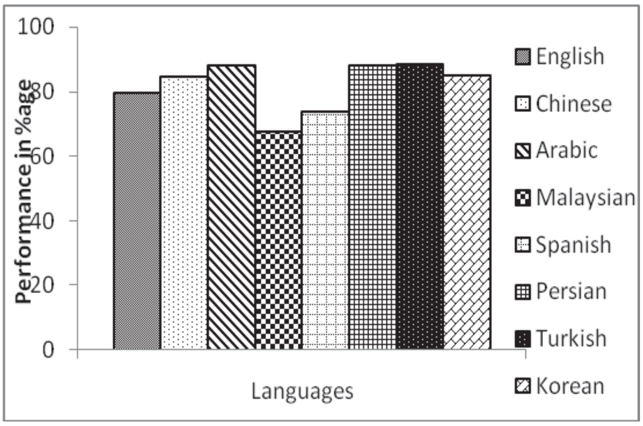

Fig. 1. Performance achieved in different languages.

Fig. 2 shows accuracy achieved in sentiment classification task in both writing style (Formal and Informal) written using 8 different languages. Maximum performance for Opinion Mining in case of Informal Text is achieved in Arabic Language and for formal text; maximum accuracy is obtained in Persian and Turkish Language.

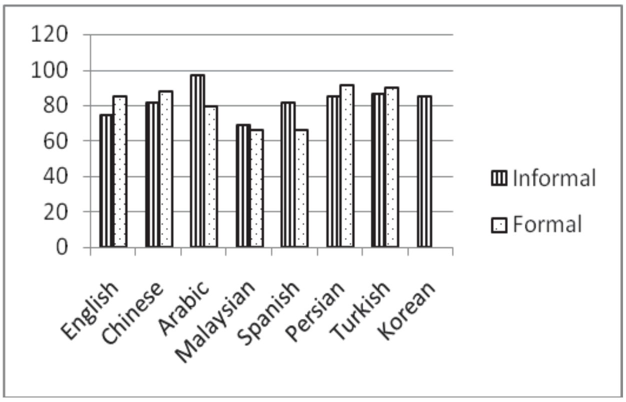

Fig. 2. Comparison of performance in Formal and Informal writing style

Performance of Opinion Mining Task also depends on type feature selection method used. Different feature selection methods are experimented and proposed by the authors. Few of them are IG, TF-IDF, n-gram, MI and MMI, and others (includes CHI, FS-INS, DF). Table III provides information regarding different feature selection methods used to find the sentimental states from text written in different languages.

TABLE II. DIFFERENT FEATURE SELECTION METHODS USED IN OPINION MINING

Sr. No.	Feature Selection	Usage in Opinion Mining
1	IG	6
2	TF-IDF	6
3	N-gram	2
4	MI and MMI	3
5	others	3

Graphical representation of Table III is shown in Fig 3. It shows IG and TF-IDF are used maximum number of times for find the emotion associated with the text written in any language and IG proves to be the best for find sentimental states associated with the text.

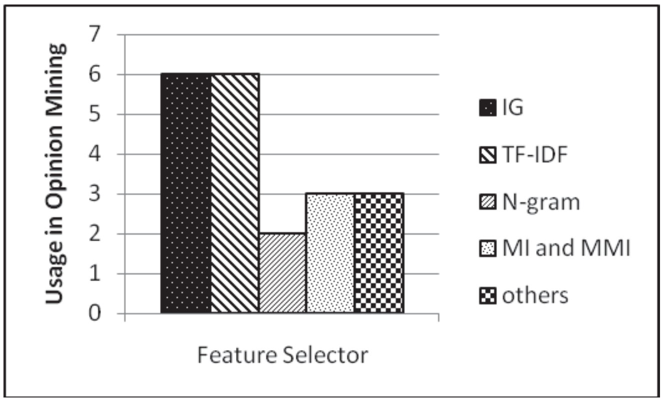

Fig. 3. Usage of different Feature selection parameters in field of Opinion Mining

5. CONCLUSION

Computational Linguist, Natural Language Processing and Text Analysis plays major role in field of Opinion Mining. Finding emotions associated with piece of text whether formal or informal, written in any language involve better understanding of Natural Language Processing by Machine, from which computational World is still very far. To increase the performance of Opinion Mining task requires better understanding of features of language being used. It was found that maximum accuracy is obtained, in field of Opinion mining, with text written in Arabic, Persian and Turkish. In case of Formal, Persian and Turkish language text performed well and for Informal Arabic text, classifier shows maximum accuracy. Feature selection parameter, which is used for find the emotional states associated with a written text, IG outperformed all other feature selectors. And For contextual dependence, the N-gram (unigram, bigram) approach for feature selection was used in case of Formal text especially poetry.

6. References

[1] Silva A. V., "Classifying Emails by Flow of Emotions" accessed from https://wiki.engr.illinois.edu/download/attachments/200017637/dm_report.pdf?version=1&modificationDate=1336540369000 on July 2014.

[2] Kaur J. and Saini J.R. ,"Emotion Detection and Sentiment Analysis in Text Corpus: A Differential Study with Informal and Formal Writing Styles" paper published in International Journal of Computer Application, ISSN (0975 – 8887), Volume 101, Issue 9, September 2014, pp.1-9.

[3] Pang B., Lee L., and Vaithyanathan S., "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," Proceedings Of The Conference On Empirical Methods In Natural Language Processing, 2002, pp. 79-86.

[4] Bagheri, A. , Saraee, M. de Jong, F."Sentiment Classification in Persian: Introducing a mutual information-based method for feature selection" 21st Iranian Conference on Electrical Engineering (ICEE), May 2013, pp.1-6.

[5] Liu B.," Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012.

[6] Pang B. and Lee L., "Opinion Mining and Sentiment Analysis", Foundations and Trends in Information Retrieval, vol. 2, no. 1-2,2008, pp. 1–135

[7] Esra A.S "Aspect Based Opinion Mining On Turkish Tweets", Thesis Submitted In July, 2012.

[8] Sidorov G., Jimenez S.M., Jimenez V.F., Gelbukh A.F., Sanchez N.A.C., Velasquez F., Rangel I.D., Guerra S.S., Trevino A., and Gordon J., "Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets", 11th Mexican International Conference on Artificial Intelligence, MICAI 2012, San Luis Potosí, Mexico, October 27 – November 4, 2012,pp. 1-14.

[9] Can E.F., Can F., Duygulu P., Kalpakli M.," Automatic Categorization of Ottoman Literary Texts by Poet and Time Period", Computer and Information Sciences-II, 2012, pp. 51-57.

[10] Al-Kabi M.N., Alsmadi I.M., Gigieh A.H., Wahsheh H.A., Haidar M.M., "Opinion Mining and Analysis for Arabic Language" , (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 5, 2014,pp. 181-195.

[11] Hamidi S. , Razzazi F., Ghaemmaghami M.P., "Automatic meter classification in Persian poetries using Support Vector Machines" IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), Dec. 2009, pp.563 – 567.

[12] Saini J. R., "Polarity Determination using Opinion Mining in Stocks and Shares-advertising Unsolicited Bulk e-mails", International Journal of Engineering Innovation & Research (ISSN: 2277-5668), vol. 1, no. 2, March 2012, pp. 86-92.

[13] Kaur J. and Saini J. R., "On classifying sentiments and mining opinions" paper published in International Journal of Emerging Technologies in Computational and Applied Sciences(IJETCAS), ISSN 2279-0047, Volume 9, Issue 3, August 2014, pp.210-214.

[14] Li B., Lu J., Yao J. M., Zhu Q. M., "Automated Essay Scoring Using the KNN Algorithm" International Conference on Computer Science and Software Engineering, vol.1, 12-14 Dec. 2008, pp.735-738.

[15] Barros L., Rodriguez P., Ortigosa A. "Automatic Classification of Literature Pieces by Emotion Detection A Study on Quevedo's Poetry" , Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on ,2-5 Sept. 2013, pp.141-146.

[16] Jamal N., Mohd M. and Noah S.A., "Poetry Classification Using Support Vector Machines" Journal of Computer Science, vol. 8, no.9, 2012, pp.1441-1446.

[17] Alsharif O., Alshamaa D. and Ghneim N., "Emotion Classification in Arabic Poetry using Machine Learning" International Journal of Computer Applications , volume 5, no.16, March 2013, pp. 10-15.

[18] Samsudin N., Hamdan A.R., Puteh M., Nazri M.Z.A. "Mining Opinion in Online Messages" International Journal of Advanced Computer Science and Applications, vol. 4, No. 8, pp.19-23, 2013.

[19] S.A. Noah and F. Ismail, 2008. "Automatic Classifications of Malay Proverbs Using Naive Bayesian Algorithm" Information Technology Journal, vol. 7, 2008, pp. 1016-1022.

[20] Cho S.H. and Kang B.H., "Statistical Text Analysis and Sentiment Classification in Social Media", 2012 IEEE International Conference on Systems, Man, and Cybernetics, COEX, Seoul, Korea, October 14-17, 2012.

[21] Tan S. and Zhang J. "An empirical study of Sentiment Analysis for Chinese documents" Expert Systems with Applications: An International Journal, Volume 34 Issue 4, May, 2008, pp. 2622-2629.

[22] Poria S., Cambriab E., Wintersteinc G., Huanga G.B., "Sentic Patterns: Dependency-Based Rules for Concept-Level Sentiment Analysis", Knowledge-Based Systems, 2014, pp. 1–32.

[23] Kumar V. and Minz S., "Poem classification using machine learning" Proceedings published in Advances Volume 236, 2014, pp. 675-682.

[24] Ding X, Liu B., Yu P.S., "A Holistic Lexicon-Based Approach to Opinion Mining" Proceedings of the 2008 International Conference on Web Search and Data Mining published by ACM, February 11-12, 2008, Palo Alto, California, USA. 2008.

[25] He Z.S., Liang W.T., Li L.Y., Tian Y.F., "SVM-Based Classification Method for Poetry Style" International Conference on Machine Learning and Cybernetics,volume-5,2007, pp. 2936 – 2940.

[26] Lee H.Y. and Renganathan H. "Chinese Sentiment Analysis Using Maximum Entropy" Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP), IJCNLP 2011, pp.89–93.

Jasleen Kaur had done Bachelor of Technology, Computer Science and Engineering from Guru Teg Bahadur Khalsa Institute of Engineering Technology, Malout, Punjab and Master of Technology (Computer Engineering) from Punjabi University, Patiala, Punjab. She has published 10 papers in various International Journals Presently she is working as Assistant Professor at Shroff SR Rotary Institute of Chemical Technology, Ankleshwar, Gujarat.

Dr. Jatinderkumar R. Saini is Ph.D. from VNSGU, Surat. He secured First Rank in all three years of MCA and has been awarded Gold Medals for this. Besides being University Topper, he is IBM Certified Database Associate (DB2) as well as IBM Certified Associate Developer (RAD). Associated with more than 50 countries, he has been the Member of Program Committee for more than 50 International Conferences (including those by IEEE) and Editorial Board Member or Reviewer for more than 30 International Journals (including many those with Thomson Reuters Impact Factor). He has more than 55 research paper publications and nearly 20 presentations in reputed International and National Conferences and Journals. He is member of ISTE, IETE, ISG and CSI. Currently he is working as Associate Professor and Director I/C at Narmada College of Computer Application, Bharuch, Gujarat, India. He is also Director (Information Technology) at Gujarat Technological University, Ahmedabad (GTU)'s A-B Innovation Sankul.