An Improved Cross-Domain Sentiment Classification using L_{1/2} Penalty Logistic Regression

J. Karthikeyan, P.G Scholar Dept. of computer Science and Engineering, Al-Ameen Engineering College, Erode-638104, Tamilnadu, India.

Abstract— Sentiment Classification is used in applications such as polarity analysis of reviews, summarization of user comments and domain adaptation. Each domain has different sentiment words and is costly to annotate data for each new domain. In Cross-Domain Sentiment Classification, the features or words that appear in the source domain do not always appear in the target domain. So a classifier trained on one domain might not perform well on a different domain because it fails to learn the sentiment of the unseen words. One solution to this issue is to use a thesaurus which groups different words that express the same sentiment. Hence, feature expansion is required to augment a feature vector with additional related features selected for the thesaurus to reduce the mismatch between features. The proposed method generates a thesaurus that is aware to the sentiment of words expressed in different domains. It uses both labeled as well as unlabeled data available from the source domains and unlabeled data from the target domain. Then it makes use of created thesaurus to expand feature vectors at train and test times in an $L_{1/2}$ Regularized binary classifier. The $L_{1/2}$ Regularization outperforms L1 Regularized logistic regression based binary classifier and results in better prediction of sentiments in Cross-Domains.

Keywords—Cross-Domains; Sentiment classification; Feature Exappsion; L_1 Regularization; $L_{1/2}$ Regularization.

I. INTRODUCTION

The main goal of data mining is to effectively handle huge data, obtain actionable pattern and learn perceptive information. Various methods, tools and algorithms were developed for handling huge amount of data. The sudden increase of social media has created many opportunities for people to publicly vote their opinions. Because social media is widely used for a variety of purposes, huge amounts of usergenerated data exist and can be made obtainable for data mining.

Sentiment analysis evaluates people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics, and their attributes. Users state their opinions about products or services they get through in blog posts, shopping sites, or review sites. Reviews on an extensive range of goods are available on the Web such as, books (amazon.com), hotels (tripadvisor.com), movies (imdb.com), automobiles (caranddriver.com), and restaurants (yelp.com). It is valuable for both the consumers as well as for the producers E. Suresh, Faculty Dept. of Computer Science and Engineering, Al-Ameen Engineering College, Erode-638104, Tamilnadu, India.

to know what common public feel about a particular product or service.

The main objective of Sentiment analysis is to automatically mine opinions expressed in the user-generated reviews. Sentiment analysis and opinion mining make producers to know about consumer's opinion about product, brand insight, new product awareness, and reputation supervision. Sentiment investigation is tough because languages used to generate contents are indistinct. Performance assessment of sentiment analysis is another challenge because of the lack of ground precision.

Automatic document level sentiment categorization is the task of categorizing a given review with reverence to the sentiment stated by the creator of the review. Sentiment categorization has been useful in several tasks such as opinion extraction, opinion summarization, appropriate marketing and advertise analysis. For instance, in an opinion summarization system it is valuable to categorize every review into positive or negative sentiments and then produce a synopsis for each opinion type for a particular product. Supervised learning algorithms that need labeled data have been effectively used to construct sentiment classifiers for a given field.

However, sentiment is expressed in a different way in different fields, and it is expensive to interpret data for each new domain in which one would like to apply a sentiment classifier. For instance, in the electronics domain the words 'durable' and 'glow' are used to express positive attitude, while 'costly' and 'short battery time' points to negative attitude. On the other hand, in the books domain the words 'exciting' and 'thriller' express positive attitude, whereas the words 'boring' and 'long' usually express negative attitude. A classifier trained on single domain might not do well on a different domain because it fails to learn the attitude of the unnoticed words. The Cross-Domain Sentiment Classification concentrates on the problem of training a classifier from one or more domains (source domains) and applying the trained classifier on a different domain (target domain). A crossdomain opinion classification system should identify which source domain features are associated to which target domain features. Next, it needs a learning structure to add in the information regarding the relatedness of source and target domain features.

The section 2 gives a brief overview of referred related works. The proposed work is explained in section 3. Section 4 analyses the experimental results.

II. RELATED WORK

The automatic examination of user generated contents such as online news, reviews, blogs and tweets are very important for tasks such as mass sentiment evaluation, corporate status calculation, political orientation classification, stock market forecast and community opinion study. Pang.B et al.[9] discuss the new challenges raised by sentiment sensitive applications. Sentiment classification systems can be broadly categorized into single-domain and cross-domain classifiers based upon the domains from which they are trained on and subsequently applied to.

A. Single-Domain Sentiment Classification

In single-domain sentiment classification, a classifier is trained using labeled data annotated from the domain in which it is applied. Pang.B et al. [8] examined whether it is adequate to treat sentiment classification simply as a special case of topic-based categorization or whether special sentimentclassification methods need to be developed. This approach used three standard algorithms: Naive Bayes classification, Maximum Entropy Classification, and Support Vector Machines (SVMs) for Sentiment Classification. In topic-based classification, all three classifiers have been reported to attain accuracies of 90% and above for particular categories. This shows that sentiment categorization is harder than topic classification.

Turney.P.D [11] measured the co-occurrences between a word and a set of manually selected positive words (e.g., good, nice, excellent and so on) and negative words (e.g., bad, nasty, poor and so on) using pointwise mutual information to calculate the attitude of a word.

Kanayama.H et al. [7] proposed an approach to build a domain-oriented sentiment dictionary to identify the words that express a particular sentiment in a given domain. By construction, a domain specific dictionary considers sentiment orientation of words in a particular domain. Therefore, this method cannot be used to categorize sentiment in a different domain.

Aue.A et al.[1] reported a number of experimental tests on domain adaptation of sentiment classifiers. They used a group of nine classifiers to train a sentiment classifier. However, most of these tests were unable to do better than a simple baseline classifier that is trained using all labeled data for all domains. They admit the challenges involved in crossdomain sentiment classification and put forward the possibilities of using unlabeled data to improve performance.

Blitzer.J et al. [3] examined the task of domain adaptation. Experimental threat minimization suggests wellknown learning guarantees when training and test data come from the same domain. In the real world, people often wish to adapt a classifier from a source domain with a large amount of training data to different target domain with very little training data. This work proposed consistent convergence bounds for algorithms that minimize a convex combination of source and target experimental threat.

Daume.H et al. [6] proposed a semi-supervised (labeled data in source, and both labeled and unlabeled data in target) expansion to a well-known supervised domain adaptation approach. This semisupervised approach to domain adaptation is extremely easy to employ, and can be applied as a preprocessing step to any supervised learner. However, despite their simplicity and experimental success, it is not theoretically obvious why these algorithms do so well. Compared to single-domain sentiment classification, cross-domain sentiment classification has recently received awareness with the advancement in the field of domain adaptation.

B. Cross-Domain Sentiment Classification

Sentiments are expressed in different ways in different domains, and interpreting corpora for each domain of interest is not practical. Blitzer.J et al.[2] proposed Structural Correspondence Learning-Mutual Information (SCL-MI) algorithm for minimizing the relative error due to adaptation between domains by an average of 30% over the original Structural Correspondence Learning (SCL) algorithm. This work gives a measure of domain similarity that correlates well with the potential for adaptation of a classifier from one domain to another. This measure could for instance be used to select a small set of domains to annotate whose trained classifiers would transfer well to many other domains.

Pan.S.J et al.[8] developed a general solution to sentiment classification when the system do not have any labels in a target domain but have some labeled data in source domain. In this cross-domain sentiment classification setting, to bridge the gap between the domains, a Spectral Feature Alignment (SFA) algorithm is proposed to arrange domainspecific words from different domains into unified clusters, with the help of domain independent words as a bridge. In this way, the clusters can be used to reduce the gap between domain-specific words of the two domains, which can be used to train sentiment classifiers in the target domain correctly.

A primary problem when applying a sentiment classifier trained on a particular domain to classify reviews on a different domain is that words (hence features) that appear in the target domain do not always appear in the trained model. Bollegala.D et al. [4] proposed a method to overcome this problem in cross-domain sentiment classification. In this work, first a sentiment sensitive distributional thesaurus is created using labeled data for the source domains and unlabeled data for both source and target domains. Next, feature vectors are expanded using created thesaurus during train and test times in a L_1 regularized logistic regression based binary classifier.

C. Regression based classification

Zou.H et al.[13] compares the regularization methods for linear models. The $L_{1/2}$ Regularization overcomes the limitations of the Lasso Regression (or L_1 Regression). The $L_{1/2}$ combines shrinkage and variable selection, and in addition encourages grouping of variables: groups of highly correlated variables tend to be selected together, where the Lasso would only select one variable of the group. Thus an $L_{1/2}$ regularized logistic regression based classifier would outperform the L_1 regularized logistic regression based binary classifier.

III. PROPOSED MODEL FOR SENTIMENT CLASSIFICATION IN CROSS-DOMAINS

The following sections describe the modules of the proposed work.

A. Lexical and Sentiment Elements Generation

The proposed method first splits each review into individual sentences and then part-of-speech (POS) tagging and lemmatization is applied to these sentences using the RASP system. The feature sparseness is minimized by lemmatization process. The nouns, adjectives, verbs and adverbs are filtered out based on the POS tags. Adjectives are found to be good indicators of sentiment. The Proposed system models review as a bag of words and selects unigrams and bigrams from each sentence. The unigrams and bigrams are collectively as lexical elements. The Lexical elements can be created from both source domain labeled reviews as well as unlabelled reviews from source and target domains.

The sentiment elements are created from each source domain labeled review, by appending the label of the review to each lexical element generated from that review. '*P' is appended to the lexical elements to denote positive sentiment and '*N' is appended to the lexical elements to denote negative sentiment.

B. Pointwise Mutual Information Evaluation

The Proposed system represents a lexical or sentiment element u by a feature vector u, where each lexical or sentiment element w that co-occurs with u in a review sentence contributes a feature to u. The value of the feature w in vector u is denoted by f(u,w). The vector u can be seen as a representation of the distribution of an element u over the set of elements that co-occur with u in the reviews. The distributional hypothesis states that words that have similar distributions are semantically similar f(u,w) is the pointwise mutual information (PMI) between a lexical element u and feature w. The pointwise mutual information, f(u,w) is calculated using the following equation.

$$f(\mathbf{u},\mathbf{w}) = \log \left[\frac{\frac{c(u,w)}{N}}{\sum_{i=1}^{n} c(i,w)} \frac{\sum_{j=1}^{m} c(u,j)}{N} X \frac{\sum_{j=1}^{m} c(u,j)}{N} \right]$$
(1)

Here, c (u, w) is the number of review sentences in which a lexical element u and a feature w co-occur, n is the total number of lexical elements, m is the total number of features and N = $\sum_{i=1}^{n} \sum_{j=1}^{m} c(i, j)$.

C. Relatedness value Evaluation

For any two lexical or sentiment elements u and v (represented by feature vectors u and v, respectively), the proposed method computes the relatedness score T(v,u) of the element v to the element u using the following equation.

$$T(\mathbf{v},\mathbf{u}) = \frac{\sum_{w \in \{x \mid f(v,x) > 0\}} f(u,w)}{\sum_{w \in \{x \mid f(u,x) > 0\}} f(u,w)}$$
(2)

Note that PMI values can be negative. To avoid considering negative pointwise mutual information values, the system considers only positive weights. Relatedness is an asymmetric measure and the relatedness T(v,u) of an element v to another element u is not necessarily equal to T(u,v), the relatedness of u to v. In cross-domain sentiment classification the source and target domains are not symmetric.

By using the relatedness measure a sentiment sensitive thesaurus is generated in which for each lexical element u the system lists up lexical elements v that co-occur with v (i.e. f(u, v) > 0) in the descending order of the relatedness score values T(v,u). The lexical elements alone are retained in the sentiment sensitive thesaurus because when predicting the sentiment label for target reviews (at test time) one cannot generate sentiment elements from those (unlabeled) reviews; therefore, it is not required to find expansion candidates for sentiment elements.

However, the relatedness values between the lexical elements listed in the sentiment-sensitive thesaurus are computed using co-occurrences with both lexical and sentiment elements, and, therefore, the expansion candidates selected for the lexical elements in the target domain reviews are sensitive to sentiment labels assigned to reviews in the source domain. To construct the sentiment sensitive thesaurus, the system must compute pairwise relatedness values using equation (2) for numerous lexical elements.

D. Ranking Score Calculation

A fundamental crisis in cross-domain sentiment classification is that features that appear in the source domains do not always appear in the target domain. Therefore, even if the classifier is trained using labeled data from the source domains, the trained model cannot be readily used to classify test instances in the target domain. To overcome this problem, the proposed system uses a feature expansion method suggested by Bollegala.D et al. [4] where the system augments a feature vector with additional related features selected from the sentiment-sensitive thesaurus created.

First, following the bag-of-words model, the proposed system models a review d using the set $\{w_1, w_2, w_3, \ldots, w_N\}$, where the elements w_i are either unigrams or bigrams that appear in the review d. Then it represents a review d by a real-valued term-frequency vector d, where the value of the j-th element d_j is set to the total number of occurrences of the unigram or bigram w_j in the review d. The system calculates ranking score score (u_i,d) for each base entry in the thesaurus using the equation

score
$$(u_i,d) = \frac{\sum_{j=1}^{N} d_j T(w_j, u_i)}{\sum_{l=1}^{N} d_l}$$
 (3)

According to the equation (3), given a review d, a base entry u_i will have a high ranking score if there are many words w_j in the review d that are also listed as neighbours for the base entry u_i in the sentiment-sensitive thesaurus. To expand a vector, d, for a review d, the system first ranks the base entries, u_i using the ranking score and select the top k ranked base entries. Let us denote the r-th ranked $(1 \le r \le k)$ base entry for a review d by v_d^r . Second, extend the original set of unigrams and bigrams $\{w_1, w_2, w_3, \dots, w_N\}$ by the base entries $\{v_d^1, v_d^2, \dots, v_d^r\}$ to create a new vector d' with dimensions corresponding to $\{w_1, w_2, w_3, \dots, w_N, v_d^1, v_d^2, \dots, v_d^r\}$ for a review d. The values of the extended vector d' are set as follows: The values of the first N dimensions that correspond to unigrams and bigrams w_i that occur in the review d are set to d_i, their frequency in d. The subsequent k dimensions that correspond to the top ranked base entries for the review d are weighted according to their ranking score. Specifically, the system sets the value of the r-th ranked base entry v_d^r to 1/r.

By using the inverse rank as the feature value for expanded features, the system only takes into account the relative ranking of base entries and at the same time assign feature values lower than that for the original features. Note that the score of a base entry depends on a review d. Therefore, the system selects different base entries as additional features for expanding different reviews. By adjusting the value of k, the number of base entries used for expanding a review, one can change the size of this latent space onto which the feature vectors are mapped.

E. L_{1/2} Regularized Classification

Logistic Regression is used to estimate a model for predicting the future outcome. Ordinary least squares (OLS) regression does not perform well with respect to both prediction accuracy and model complexity. OLS regression may result in highly variable estimates of the regression coefficients in the presence of collinearity or when the number of predictors (P) is large relative to the number of observations (N). Ridge regression reduces this variability by shrinking the coefficients, resulting in more prediction accuracy. In Ridge regression, the coefficients are shrunken towards zero, but will never become exactly zero. So, when the number of predictors is large, Ridge regression will not provide a sparse model that is easy to interpret. Subset selection, on the other hand, does provide interpretable models, but does not reduce the variability of the estimates of the coefficients. In Ridge regression, the sum of squares of the coefficients is constrained as follows:

$$L^{\text{ridge}}(\beta_1,\ldots,\beta_P) = \|\mathbf{y}-\boldsymbol{\sum}_{j=1}^P \beta_j \quad X_j\|^2 + \lambda_2 \sum_{j=1}^P \beta_j^2 \tag{4}$$

with N the number of observations, P the number of predictor variables, β_j , j = 1, ..., P, the regression coefficients, and λ_2 the Ridge penalty parameter, and where $\|\cdot\|^2$ denotes the squared Euclidean norm.

The Lasso was developed to improve both prediction accuracy and model interpretability by combining the nice features of Ridge regression and subset selection. The Lasso reduces the variability of the estimates by shrinking the coefficients and at the same time produces interpretable models by shrinking some coefficients to exactly zero. In terms of prediction accuracy and interpretability, the Lasso outperforms Ridge regression and subset selection for data with a small to moderate number of moderate-sized effects; subset selection performs the best with a small number of large effects, and Ridge regression performs the best with a large number of small effects. The Lasso constrains the sum of the absolute values of the coefficients:

$$L^{\text{lasso}}\left(\beta_{1},\ldots,\beta_{P}\right) = \|y-\sum_{j=1}^{P}\beta_{j} \quad X_{j}\|^{2} + \lambda_{1}\sum_{j=1}^{P}\text{sign}\left(\beta_{j}^{2}\right)\beta_{j}$$
(5)

with λ_1 the Lasso penalty.

Recently, the $L_{1/2}$ was introduced to overcome the limitations of the Lasso in some situations. The $L_{1/2}$ also combines shrinkage and variable selection, and in addition encourages grouping of variables: groups of highly correlated variables tend to be selected together, where the Lasso would only select one variable of the group. Also, in the case P » N, Lasso algorithms are limited because at most N variables can be selected. Whenever Ridge regression improves on OLS, the $L_{1/2}$ will improve the Lasso. Ridge regression, the Lasso, and the $L_{1/2}$ are regularization methods for linear models. The sparse logistic regression model based on the $L_{1/2}$ penalty has the form:

$$\beta^{1/2} = \operatorname{argmin} \{ l(\beta|D) + \lambda \sum_{i=1}^{p} |\beta_{i}|^{1/2} \}$$
(6)

where $\lambda > 0$ is a tuning parameter and P(B) is a regularization term. Thus an $L_{1/2}$ regularized logistic regression based binary classifier would outperform the L_1 regularized logistic regression based binary classifier. Using the extended vectors d' to represent reviews, the system trains a binary classifier from the source domain labeled reviews to predict positive and negative sentiment in reviews. Once an $L_{1/2}$ regularized logistic regression based binary classifier is trained, it can be used to predict the sentiment of a target domain review.

IV. EXPERIMENTAL RESULTS

The Dataset used by the proposed method and the experimental results obtained are analyzed in the following sub-sections.

A. Dataset Description

The proposed system uses the cross-domain sentiment classification dataset prepared by Blitzer et al (2007) to compare the proposed method against existing work on cross-domain sentiment classification. This dataset consists of Amazon product reviews for four different product types: books, DVDs, electronics and kitchen appliances. Each review is assigned with a rating (0-5 stars), a reviewer name and location, a product name, a review title and date, and the review text. Reviews with rating > 3 are labeled as positive, whereas those with rating < 3 are labeled as negative. The overall structure of this benchmark dataset is shown in Table 1.

TABLE I. NO OF REVIEWS IN AMAZON PRODUCT REVIEW DATASET

Domain	Positive	Negative	Unlabeled
Kitchen	800	800	16746
DVDs	800	800	34377
Electronics	800	800	13116
Books	800	800	5947

For each domain, there are 1000 positive and 1000 negative examples, the same balanced composition as the polarity dataset constructed by Pang et al (2002). The dataset also contains some unlabeled reviews for the four domains.

This benchmark dataset has been used in much previous work on cross-domain sentiment classification and by evaluating on it one can directly compare the proposed method against existing approaches. The proposed work randomly selects 800 positive and 800 negative labeled reviews from each domain as training instances (total number of training instances are 1600*4 = 6400), and the remainder is used for testing (total number of test instances are 400*4 = 1600).

To conduct experiments, the proposed system selects each domain in turn as the target domain, with one or more other domains as sources. The system creates a sentiment sensitive thesaurus using labeled data from the source domain and unlabeled data from source and target domains. Then it uses this thesaurus to expand the labeled feature vectors (train instances) from the source domains and train an $L_{1/2}$ regularized logistic regression-based binary classifier.

B. Performance Evaluation

The Classification Accuracy on target domain is used as a performance metric for evaluation. It is the fraction of the correctly classified target domain reviews from the total number of reviews in the target domain, and is defined as follows:

$$Accuracy = \frac{No \ of \ correctly \ classified}{Total \ no \ of \ reviews} \tag{7}$$

C. Implementation

Using python $L_{1/2}$ regularized classifier has been implemented and analyzed. The $L_{1/2}$ regularized Classifier improves the accuracy of predicting the sentiments in target domain. The Thesaurus file output is shown below.

🔄 thesau	rus - WordPad		23
$\cdot \mathbf{X} \cdot \cdot \cdot 1$	2 3 4	5 6 7 8	• 1 • 9
unit	unit,1.000000	time,0.249333	
	good,0.218506	great,0.20172	0
	support,0.19650		
	problem, 0.18221		
		thing,0.13611	
	line,0.135968	technical,0.1	345 -
•	III		P
	100%	\odot $$ \bigcirc $ \bigcirc$	÷:

Fig. 1. Thesaurus File Generated

Using the above generated thesaurus file the feature vectors are expanded.

E	C v train_vects - WordPad
	Home View
1	<u>3</u> · · · 1 · · · 2 · · · 3 · · · 4 · · · 5 · · · 6 · · · 7 · · · 8 · · · 9 · · ·10 · · ·11 · ·12 · · ·13 · · ·14 · · ·15
	1 BASE=trite:0.007519 BASE=disappointment:0.001548
	BASE=exercise:0.001279 BASE=huge:0.008264 BASE=successful:0.001563 BASE=sound_system:0.001692
	BASE=reasonable_price:0.001558 BASE=hundred:0.002681 skip:1.000000 BASE=depth_character:0.001124 BASE=later:0.004975
	BASE=wonderful_movie:0.005525 BASE=certain:0.001340
	BASE=great_quality:0.005952 BASE=big_fan:0.001064

Fig. 2. Feature Vector Expansion

The $L_{1/2}$ classifier gives higher accuracy than L_1 Classifier. The $L_{1/2}$ classifier improves the accuracy of predicting the sentiments by including the penalties of both ridge regression and lasso regression.

V. CONCLUSION

The proposed system develops a cross-domain sentiment classification system that uses $L_{1/2}$ regularized Classification model. To overcome the feature mismatch problem in cross-domain sentiment classification, it uses labeled data from multiple source domains and unlabeled data from source and target domains to compute the relatedness of features and construct a sentiment sensitive thesaurus. The thesaurus is used to expand feature vectors during train and test times for a binary classifier. A relevant subset of the features is selected using $L_{1/2}$ regularization. Once an $L_{1/2}$ regularized logistic regression based binary classifier is trained, it is used to predict the sentiment of a target domain review. The prediction accuracy of cross-domain sentiment classification is improved by $L_{1/2}$ classifier. It outperforms prediction accuracy of L_1 classification.

ACKNOWLEDGMENT

I thank my guide and Head of the Department for their valuable information and encouragement. I thank everyone who supported me in developing this system.

REFERENCES

- [1] Aue.A and Gamon.M (2005), 'Customizing Sentiment Classifiers to New Domains: A Case Study', Technical report, Microsoft Research.
- [2] Blitzer.J, Dredze.M, and Pereira.F (2007), 'Biographies, Bollywood, Boom-Boxes and Blenders: Domain Adaptation for Sentiment Classification', Proc. 45th Ann. Meeting of the Assoc. Computational Linguistics (ACL '07) pp. 440-447.
- [3] Blitzer.J, Crammer.K, Kulesza.A, Pereira.F, and Wortman.J (2008), 'Learning Bounds for Domain Adaptation', Proc. Advances in Neural Information Processing Systems Conf. (NIPS '08). pp. 1-8.
- [4] Bollegala.D, Weir.D, Caroll.J (2013), 'Cross-Domain Sentiment Classification using a Sentiment Sensitive Thesaurus', IEEE Transcations on Knowledge and Data Engineering, Vol.25, No.8 pp. 1719-1731.
- [5] Briscoe.T, Carroll.J, and Watson.R (2006), 'The Second Release of the RASP System', Proc. COLING/ACL Interactive Presentation Sessions Conf. pp.77-80.
- [6] Daum'e III.H, Abhishek.K, Avishek.S(2010), 'Frustratingly Easy Semi-Supervised Domain Adaptation', Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing, ACL 2010 pp. 53–59.
- [7] Kanayama.H and Nasukawa.T, "Fully Automatic Lexicon Expansion for Domain-Oriented Sentiment Analysis," Proc. Joint Conf. Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP '06), pp. 355-363, 2006.
- [8] Pan.S.J, Ni.X, Sun.J.T, Yang.Q, and Chen.Z (2010), 'Cross-Domain Sentiment Classification via Spectral Feature Alignment', Proc. 19th Int'l Conf. World Wide Web (WWW '10).
- [9] Pang.B, Lee.L, and Vaithyanathan.S (2002), 'Thumbs Up? Sentiment Classification Using Machine Learning Techniques', Proc. ACL-02 Conf. Empirical Methods in Natural Language Processing (EMNLP '02) pp. 79-86.
- [10] Pang.B and Lee.L (2008), 'Opinion Mining and Sentiment Analysis', Foundations and Trends in Information Retrieval, vol. 2, nos. 1/2 pp. 1-135.
- [11] Turney.P.D (2002) 'Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews', Proc.

International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 3 Issue 2, February - 2014

40th Ann. Meeting on Assoc. for Computational Linguistics (ACL '02) pp. 417-424.

- [12] Yong Liang, Cheng Liu, Xin-Ze Luan, Kwong-Sak Leung, Tak-Ming Chan, Zong-Ben Xu and Hai Zhang (2013), "Sparse logistic regression with a L_{1/2} penalty for gene selection in cancer classification", BMC Bioinformatics
- [13] Zou.H., Hastie.T (2005), 'Regularization and variable selection via the Elastic Net', Journal of the Royal Statistical Society, Series B pp.301–320.

