International Journal of Advance Research in Science and Engineering
Volume No.06, Issue No. 12, December 2017
www.ijarse.com
IJARSE
ISSN: 2319-8354

SENTIMENTAL ANALYSIS TEXT MINING USING FOR SOCIAL MEDIA

Mr. M. Yuvaraja[1], Dr. S. Thavamani[2]

[1] Assistant Professor in Department of Computer Science, Bishop Ambrose College, Coimbatore, Tamil Nadu, and India.

[2] Associate Professor in Department of Computer Applications, Sri Ramakrishna College for Arts and Science (Formerly SNR Sons College), Coimbatore, Tamil Nadu, and India

ABSTRACT

A social network is a social structure made up of a set of social actors and a set of the dynamic ties between these actors. The social network perspective provides a set of methods for analyzing as well as a variety of theories explaining the patterns observed in these structures. The study of these structures uses social network analysis to identify local and global patterns, locate influential entities and examine network dynamics. These foreground topics can give potential interpretations of the sentiment variations. To Future enhance the readability of the mined reasons. We select the most representative tweets for foreground topics and develop another generative model called Reason candidate and Background LDA to rank them with respect to their popularity within the variation period.

I.INTRODUCTION

Sentiment analysis performs on specific domain to achieve higher level of accuracy. The feature vector used in sentiment analysis contains a bag of words which are limited and should be specific to particular domain (domain can be consider as book, hotel, electronics etc.). However sentiment expressed differently in different domains and it is costly to annotate data for each new domain in which we would like to apply a sentiment classifier. Hence the solution can be to perform cross domain sentiment analysis but the problem is that classifier trained in one domain may not work well when applied to other domain due to mismatch between domain specific words. So before applying trained classifier on target domain some techniques must be applied like feature vector expansion, finding relatedness among the words of source and target domain, etc. Cross-domain classification is nothing but to make a sentiment analysis from domain specific to generalize. A different technique gives different analysis, result and accuracy which depend on the documents, domain taken into consideration for classification. The main task in sentiment classification is to determine the polarity of the comments as positive, negative or objective. It can be done at different levels such as word/phrase levels, sentence level and document level. Sentiment can be expressed in text, in different ways.

The document level or at the sentence level does not provide the necessary detail needed opinions on all aspects of the entity which is needed in many applications, to obtain these details; we need to go to the aspect level. Aspect-level SA aims to classify the sentiment with respect to the specific aspects of entities. The first step is to identify the entities and their aspects. The opinion holders can give different opinions for different aspects of the same entity like this sentence ‘‘the voice quality of this phone is not good, but the battery life is long’’. This survey tackles the first two kinds of SA. The data sets used in SA are an important issue in this field. The main sources of data are from the product reviews. These reviews are important to the business holders as they can take business decisions according to the analysis results of user’s opinions about their products. Sentiment analysis is used to extract this data and produce a summarized result. Basically sentiment analysis is to classify the polarity of text in document or sentence whether the opinion expressed is positive, negative or neutral. Sentiment classification techniques can be divided into following approach. Supervised machine learning techniques are used for classified document or sentences into finite set of class i.e. into positive, negative and objective. Training data set is available for all kind of classes. An optimal scenario will allow for the algorithm to correctly determine the class labels for unseen instances. This requires the learning algorithm to generalize from the training data to unseen situations in a "reasonable" way. Unsupervised machine learning techniques don’t use training data set for classification. Semantic Orientation also provides to generate accurate result for classification. Point wise mutual information (PMI) is also one of the unsupervised classification methods for sentiment analysis.

II. PROBLEM IDENTIFICATION

Unsupervised Cross-domain Sentiment Classification is the task of adapting sentiment classifier trained on a particular domain (source domain), to a different domain (target domain), without requiring any labelled data for the target domain .By adapting an existing sentiment classifier to previously unseen target domains, we can avoid the cost for manual data annotation for the target domain. We model this problem as embedding learning, and construct here objective functions that capture: (a) distributional properties of pivots (i.e. common features that appear in both source and target domains),(b) label constrains in the source domain documents, and,(c) geometric properties in the unlabeled documents in both source and target domains.

III. REVIEW OF LITERATURE

3.1 Learning similarity metrics for event identification in social media

Social media sites (e.g., Flickr, YouTube, and Facebook) are a popular distribution outlet for users looking to share their experiences and interests on the Web. These sites host substantial amounts of user-contributed materials (e.g., photographs, videos, and textual content) for a wide variety of real-world events of different type and scale. By automatically identifying these events and their associated user-contributed social media documents, which is the focus of this paper, we can enable event browsing and search in state-of-the-art search engines. To address this problem, we exploit the rich “context” associated with social media content, including user-provided annotations and automatically generated information (e.g., content creation time). Using this rich context, which includes both textual and non-textual features, we can define appropriate document similarity metrics to enable online clustering of media to events. We explore a variety of techniques for learning multi-feature similarity metrics for social media documents in a principled manner. Our evaluation results suggest that our approach identifies events, and their associated social media documents, more effectively than the state-of-the-art strategies on which we build.

3.2 MODELLING PUBLIC MOOD AND EMOTION: TWITTER SENTIMENT AND SOCIO-ECONOMIC PHENOMENA

Micro blogging is an increasingly popular form of communication on the web. Micro blog posts, commonly known as tweets, are extremely short in comparison to regular blog posts, being at most 140 characters in length. A recent analysis of the Twitter network revealed a variegated mosaic of uses (Java et al. 2007), including a) daily chatter, e.g., posting what one is currently doing, b) conversations, i.e., directing tweets to specific users in their community of followers, c) information sharing, e.g., posting links to web pages, and d) news reporting, e.g., commentary on news and current affairs. Despite the diversity of uses emerging from such a simple communication channel, it has been noted that tweets normally tend to fall in one of two different content camps: users that micro blog about themselves and those that use micro blogging primarily to share information (Mor Naaman 2010). In both cases, tweets can convey information about the mood state of their authors. In the former case, mood expressions are evident by an explicit “sharing of subjectivity” (Crawford 2008), e.g. “I am feeling sad”. In other cases, even when a user is not specifically micro blogging about their personal emotive status, the message can reflect their mood.

3.3 PARAMETER ESTIMATION FOR TEXT ANALYSIS

Presents parameter estimation methods common with discrete probability distributions, which is of particular interest in text modelling. Starting with maximum likelihood, a posterior and Bayesian estimation, central concepts like conjugate distributions and Bayesian networks are reviewed. As an application, the model of latent Dirich let allocation (LDA) is explained in detail with a full derivation of an approximate inference algorithm based on Gibbs sampling, including a discussion of Dirichlet hyper parameter esti

This technical note is intended to review the foundations of Bayesian parameter estimation in the discrete domain, which is necessary to understand the inner workings of topic-based text analysis approaches like probabilistic latent semantic analysis (PLSA) latent Dirichlet allocation (LDA) [BNJ02] and other mixture models of count data. Other very good introductory work on topic models (e.g., [StGr07]) skips details of algorithms and other background for clarity of presentation. A Bayesian network forms a directed acyclical graph (DAG) with nodes that correspond to random variables and edges that correspond to conditional probability distributions, where the condition variable at the origin of an edge is called a parent node and the dependent variable at the end of the edge a child node. The double circle around the variable w~={wn} denotes an evidence node, i.e., a variable that is (assumed) observed, and the surrounding plate indicates the N i.i.d. samples. The unknown variables ~p and α~ can be distinguished into a multivariate parameter.

3.4 OPINION MINING AND SENTIMENT ANALYSIS

An important part of our information-gathering behavior has always been to find out what other people think. This survey covers techniques and approaches that promise to directly enable opinion-oriented information seeking systems. Our focus is on methods that seek to address the new challenges raised by sentiment aware applications, as compared to those that are already present in more traditional fact-based analysis. We include material on summarization of evaluative text and on broader issues regarding privacy, manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to.

3.5 RATED ASPECT SUMMARIZATION OF SHORT COMMENTS

As Web 2.0 applications become increasingly popular, more and more people express their opinions on the Web in various ways in real time. Such wide coverage of topics and abundance of users make the Web an extremely valuable source for mining people's opinions about all kinds of topics. This thesis focuses on the problem of opinion integration and summarization whose goal is to better support user digestion of huge amounts of opinions for an arbitrary topic. To systematically study this problem, we have identified three important dimensions of opinion analysis: separation of aspects (or subtopics) of opinions, understanding of sentiments, and assessment of quality of opinions. These dimensions form three key components in an integrated opinion summarization system. Accordingly, this thesis makes contributions in proposing novel and general computational techniques for three synergistic tasks: (1) integrating relevant opinions from all kinds of Web 2.0 sources and organizing them along different aspects of the topic which not only serves as a semantic grouping of opinions but also facilitates user navigation into the huge opinion space; (2) inferring the sentiments in the opinions with respect to different aspects and different opinion holders, so as to provide the users with a more detailed and informed multi-perspective view of the opinions; and (3) improving the prediction of opinion quality which critically decides the usefulness of the information extracted from the opinions. We focus on general and robust methods which require minimal human supervision so as to make the automated methods applicable to a wide range of topics and scalable to large amounts of opinions

.

3.6 MINING ASSOCIATION RULES BETWEEN SETS OF ITEMS IN LARGE DATABASES

We are given a large database of customer transactions. Each transaction consists of items purchased by a customer in a visit. We present an efficient algorithm that generates all significant association rules between items in the database. The algorithm incorporates buffer management and novel estimation and pruning techniques. We also present results of applying this algorithm to sales data obtained from a large retailing company, which shows the effectiveness of the algorithm.

Consider a supermarket with a large collection of items. Typical business decisions that the management of the supermarket has to make include what to put on sale, how to design coupons How to place merchandise on shelves in order to The work reported in this paper could be viewed as a step towards enhancing databases with functionalities to process queries. Find all rules that have \Diet Coke" as consequent. These rules may help plan what the store should do to boost the sale of Diet Coke. Find all rules that have \bagels" in the antecedent. These rules may help determine what products may be impacted if the store discontinues selling bagels.

Find all the rules relating items located on shelves A and B in the store. These rules may help shelf planning by determining if the sale of items on shelf A is related to the sale of items on shelf B.

3.7 LEARNING DECISION TREES

An important tool for evaluation and comparison of classifiers when the operating characteristics (i.e. class distribution and cost parameters) are not known at training time. Usually, each classifier is characterized by its estimated true and false positive rates and is represented by a single point in the ROC diagram. In this paper, we show how a single decision tree can represent a set of classifiers by choosing different labelling of its leaves, or equivalently, an ordering on the leaves. To the best of our knowledge, this is the first probabilistic splitting criterion that is not based on weighted average impurity. However, in many situations, not every misclassification has the same consequences, and problem-dependent misclassification costs have to be taken into account. If the cost parameters are not known at training time, Receiver Operating Characteristic (ROC) analysis can be applied ROC analysis provides tools to distinguish classifiers that are optimal under some class and cost distributions from classifiers that are always sub-optimal, and to select the optimal classifier once the cost parameters are known. ROC analysis for two classes is based on plotting the true-positive rate (TPR) on the y-axis and the false-positive rate (FPR) on the x-axis. This gives a point for each classifier. A curve is obtained because, given two classifiers, we can obtain as many derived classifiers as we want along the segment that connects them, just by voting them with different weights.

3.8 A MULTISESSION-BASED MULTIDIMENSIONAL MODEL

The problem of how to specify changes in multidimensional databases. These changes may be motivated by evolutions of user requirements as well as changes of operational sources. The multi version-based multidimensional model we provide supports both data and structure changes. The approach consists in storing star versions according to relevant structure changes whereas data changes are recorded through dimension instances and fact instances in a star version. The model is able to integrate mapping functions to populate multi version-based multidimensional databases.

On-Line Analytical Processing (OLAP) has emerged to support multidimensional data analysis by providing manipulations through aggregations of data drawn from various transactional databases. This approach is often based on a Multidimensional Data Base (MDB). A MDB schema [1] is composed of a fact (subject of analysis) and dimensions (axes of analysis). A fact contains indicators or measures. A measure is the data item of interest. As mentioned in [2], fact data reflect the dynamic aspect whereas dimension data represent more static information. However, sources (transactional databases) may evolve and these changes have an impact on structures and contents of the MDB built on them. In the same way, user requirement evolutions may induce schema changes; eg. to create a new dimension or a new “dimension member” [3], to add a new measure,…Changes occur on dimensions as well as facts. This paper addresses the problem of how to specify changes in a MDB. The changes may be related to contents as well as schema structures. Our work is not limited to represent the mapping data into the most recent version of the schema. We intend to keep trace of changes of multidimensional structures.

IV. PROPOSED METHODLOGY

SENTIMENT CLASSIFICATION ALGORITHMS

4.1 SUPPORT VECTOR MACHINES

Sentiment analysis is treated as a classification task as it classifies the orientation of a text into either positive or negative. Support Vector Machines (SVM) are the most favored supervised learning method of sentiment classification because of their consistently robust performances in natural language processing. As a discriminative classifier, SVM do not require any prior probabilities or assumptions about the training data as does a generative classifier such as Naive Bayes classifiers. Rather, the key idea behind a binary SVM is to find a decision surface in the feature space that will separate positive and negative training examples. In the case of opinion detection, this means separating subjective from objective examples [10]. Based on the intuition there are no examples near the decision surface, the classification decision is less uncertain and has good generalization capability. SVM undergoes 1) Pre-Processing, 2) Datasets Description, 3)Feature Extraction, 4)Feature Selection, 5)Text Classification Method Selection, 6) Effectiveness Measure. 1) Pre-Processing: The datasets will go through the pre-processing task of the text documents such as tokenization, stop word removal, low-ercase conversion and stemming. Tokenization is the procedure of splitting a text into words, phrases, or other meaningful parts, namely tokens. Stop words are the words that are commonly encountered in texts without dependency to a particular topic such as conjunctions, prepositions, etc. 2) Datasets Description: Overall description of the data used in the project or system. 3) Feature Extraction: feature extraction is the process of transforming the input data into set of features. The performance of the machine learning process depends heavily on its features so it is crucial to choose the extract features. 4) Feature Selection: There are filter, wrapper, and embedded approaches for Feature selection. In the experiments, filter methods were used due to classifier independence and relatively low computation time of the filters.

Chi-Square: Filter attributes evaluation of Chi-
Square weight features were used to select informative features and ranking method was also applied in order to remove irrelevant features. One of the most popular feature selection approaches is CHI2. In statistics, the CHI2 test is used to examine independence of two events. The events, X and Y, are assumed to be independent

$$p(XY) = p(X)p(Y)$$

5) Text Classification Method Selection: Support Vector Machine (SVM) has been chosen for the classification in the experiments. The support-vector machines are a learning machine for two-group classification problems introduced by C. Cortes and V. Vapnik. It is used to classify the texts as positives or negatives. SVM works well for text classification due to its advantages such as its potential to handle large features. Another advantage is SVM is robust when there is a sparse set of examples and also because most of the problem are linearly separable. 6) Effectiveness Measure: Four effective measures that have been used in this study are based on confusion matrix output, which are True Positive (TP), False Positive(FP), True Negative (TN), and False Negative(FN).

•Precision(P) = TP/(TP+FP)
•Recall(R) = TP/(TP+FN)
•Accuracy(A) = (TP+TN)/(TP + TN + FP +FN)
•AUC (Area under the (ROC) Curve) = 1/2.((T- P/(TP+FN))+(TN/(TN+FP))
•F-Measure(Micro-averaging) = 2.(P.R)/(P+R)

4.2 LINEAR DISCRIMINANT ANALYSIS

Linear Discriminant Analysis (LDA) as an alternative to SVM. Discriminative classifiers seek to find a decision boundary that maximizes certain measure of separation between classes. The earliest of such methods, Fisher's Linear Discriminant Analysis (Fukunaga (1990)), tries to find a linear combination of input variables that discriminates best between the two class distributions (estimated from available data). The classical LDA approach proved to be extremely useful in practice, and it has been successfully applied in many situations where the underlying assumptions (about the class distributions) for the LDA approach do not hold. For instance, LDA is often used to discriminate between the class distributions with different covariance matrices (where an optimal decision boundary is known to be nonlinear, i.e. quadratic). Practical attractiveness of LDA can be explained by its low model complexity, and its ability to capture the essential characteristics of the data distributions (mean and covariance) from finite training data, and then estimating the decision boundary using these global characteristics of the data. As a result, LDA benefits from feature combinations that produce the highest separation between classes. Therefore, classical LDA has been used and re-discovered in many recent learning techniques.

LDA can be derived from simple probabilistic models which model the class conditional distribution of the data $P(X|y=k)$ for each class k. Predictions can then be obtained by using Bayes' rule:

$$P(y=k|X) = \frac{P(X|y=k)P(y=k)}{P(X)} = \frac{P(X|y=k)P(y=k)}{\sum_l P(X|y=l) \cdot P(y=l)}$$

More specifically, for linear discriminant analysis, is modelled as a multivariate Gaussian distribution with density:

$$p(X|y=k) = \frac{1}{(2\pi)^n|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(X-\mu_k)^t\Sigma_k^{-1}(X-\mu_k)\right)$$

In order to find a good projection vector, we need to define a measure of separation between the projections. The mean vector of each class in x and y feature space is

$$\mu_i = \frac{1}{N_i}\sum_{x\in\omega_i} x \quad \text{and} \quad \tilde{\mu}_i = \frac{1}{N_i}\sum_{y\in\omega_i} y = \frac{1}{N_i}\sum_{x\in\omega_i} w^T x = w^T\mu_i$$

The solution proposed by Fisher is to maximize a function that represents the difference between the means, normalized by a measure of the within-class scatter n For each class we define the scatter, an equivalent of the variance, as

$$\tilde{s}_i^2 = \sum_{y\in\omega_i}(y-\tilde{\mu}_i)^2$$

where $\mathbf{S_i}^1$ and $\mathbf{S_i}^2$ the quantity is called the within-class scatter of the projected examples.

The Fisher linear discriminant is defined as the linear function wTx that maximizes the criterion function

$$J(w) = \frac{|\tilde{\mu}_1 - \tilde{\mu}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2}$$

4.3 STOCHASTIC AGREEMENT REGULARIZATION

Stochastic Agreement Regularization (SAR) is one of the most important work in multi-view sentiment classification. It models a probabilistic agreement framework based on minimizing the Bhattacharyya distance between models trained using two different views. They regularize the models from each view by constraining the amount by which they permit them to disagree on unlabeled instances from a theoretical model. Their co-regularized objective which has to be minimized, where Li for i = 1, 2 are the standard regularized log likelihood losses of the models p1 and p2, Eu[B(p1(θ1), p2(θ1))] is the expected Bhattacharyya distance between the predictions of the two models on the unlabeled data, and c is a constant defining the relative weight of the unlabeled dataset both on random views of unigrams and random views of bigrams and take its results as baselines [11].

4.4 IMPROVED GRADIENT BOOSTING ALGORITHM

Improved Gradient Boosting Algorithm is based on the co-training algorithm with agreement, but instead of just taking into account unlabeled examples with similar predictions from both classifiers to update the set of labeled examples such as in Wan (2009), we impose that only the examples with highest confidence upon agreement are added to the labeled list. Basically, CCC takes two main inputs: a set of labeled examples from one domain (L), the source domain, and a set of unlabeled examples from another domain (U), the target domain. After training on the source domain, both classifiers classify unlabeled documents from the target domain. If both classifiers agree on their predictions, the unlabeled document is added to an agree list for each classifier with the categorization label and the classification confidence. Finally, the P positive (subjective) and N negative (objective) documents with higher confidence values are selected from each agree list and transferred from the set of unlabeled documents to the labeled set.

In many supervised learning problems one has an output variable *y* and a vector of input variables *x* connected via a joint probability distribution P (X,Y). Using a training set $\{(x_1,y_1)\ldots\ldots(x_n,y_n)\}$ of known values of *x* and corresponding values of *y*, the goal is to find an approximation F(x) to a function F(x) that minimizes the expected value of some specified loss function

L(y F(X)):

$$\hat{F} = \arg\min_{F} \mathbb{E}_{x,y}[L(y, F(x))].$$

The gradient boosting method assumes a real-valued *y* and seeks an approximation **F(X)** in the form of a **h(X)** weighted sum of functions from some class , called base (or weak) learners:

$$F(x) = \sum_{i=1}^{M} \gamma_i h_i(x) + \text{const}.$$

In accordance with the empirical risk minimization principle, the method tries to find an approximation F(X) that minimizes the average value of the loss function on the training set. It does so by starting with a model, consisting of a constant function $F_0(X)$, and incrementally expanding it in a greedy fashion:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma),$$

$$F_m(x) = F_{m-1}(x) + \arg\min_{h \in \mathcal{H}} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + h_m(x_i)).$$

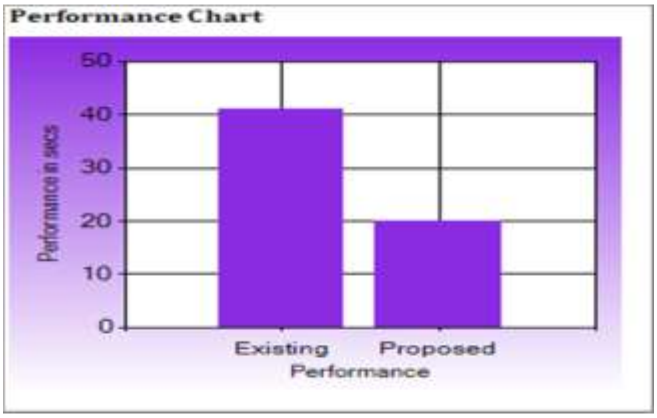
where h € H is a base learner function.

Unfortunately, choosing the best function *h* at each step for an arbitrary loss function *L* is a computationally infeasible optimization problem in general. Therefore, we will restrict to a simplification.

The idea is to apply a steepest descent step to this minimization problem. If we considered the continuous case, i.e. H the set of arbitrary differentiable functions on R, we would update the model in accordance with the following equations

$$F_m(x) = F_{m-1}(x) - \gamma_m \sum_{i=1}^{n} \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i)),$$
$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) - \gamma \nabla_{F_{m-1}} L(y_i, F_{m-1}(x_i))\right),$$

where the derivatives are taken with respect to the functions F_i for I € {1…..m}. In the discrete case however, i.e. the set h is finite, we will choose the candidate function *h* closest to the gradient of *L* for which the coefficient *γ* may then be calculated with the aid of line search the above equations. Note that this approach is a heuristic and will therefore not yield an exact solution to the given problem, yet a satisfactory approximation.



Hence the time taken for the execution has been formulated in the above mentioned chart. The proposed system taken less number of time. While comparing with the existing system, the proposed system works well

V.CONCLUSION

These foreground topics can give potential interpretations of the sentiment variations. To Future enhance the readability of the mined reasons. We select the most representative tweets for foreground topics and develop another generative model called Reason candidate and Background LDA to rank them with respect to their popularity within the variation period.

REFERENCES

[1.] 1 H. Becker, M. Naaman, and L. Gravano, *"Learning similarity met-rics for event identification in social media,"* in Proc. 3rd ACM WSDM, Macau, China, 2010.

[2.] 2 Modelling public mood and emotion: Twitter sentiment and socio-economic phenomena J. Bollen, H. Mao, and A. Pepe, *"Modeling public mood and emo-tion: Twitter sentiment and socio-economic phenomena,"* inProc. 5th Int. AAAI Conf. Weblogs Social Media, Barcelona, Spain, 2011.

[3.] 3 Parameter estimation for text analysis G. Heinrich, *"Parameter estimation for text analysis,"* Fraunhofer IGD, Darmstadt, Germany, Univ. Leipzig, Leipzig, Germany, Tech. Rep., 2009.

[4.] 4 Rated aspect summarization of short comments Y.Lu,C.Zhai, and N.Sundaresan, *"Rated aspect summarization Of short comments,"* in WWW2009, 2009,pp.131–140

[5.] 5 Sentiment-oriented contextual advertising T.-K.FanandC.-H.Chang, *"Sentiment-oriented contextual advertising,"* Knowledge and Information Systems, vol.23,no.3,pp. 321–344,2010

[6.] 6 Mining and Summarizing Customer Reviews M.Hu and B.Liu, *"Mining and summarizing customer reviews,"* in KDD2004, 2004,pp.168–177.

[7.] 7. Foundations of Statistical Natural Language Processing.

[8.] C.D.Manning and H.Schutze, *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press, 2002.

[9.] 8. Mining Association Rules between Sets of Items in Large Databases.R. Agrawal, T. Imielinski, and A. Swami. *Mining association rules between sets of items in large database*.In Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'93), pp:207-216, Washington, DC, May 1993.

[10.] 9. Learning Decision Trees Using the Area under the ROC Curve Cesar Ferri-Ram´ ırez, Peter A. Flach, and Jose Hernandez-Orallo. *Learning decision trees using the area under the roc curve*. In Proceedings of the Nineteenth International Conference on Machine Learning, pages 139–146, Morgan Kaufmann, 2002.

[11.] 10. Deliver smarter products and services by unifying software development and IT operations.

[12.] IBM, *"Deliver smarter products and se*rvices by unifying software development and IT operations," Issue Date : Sep. 2009, ISBN RAW14175-USEN-00.

[13.] 11. Logical Design of Data Warehouses from XML

[14.] M. Banek, Z. Skocir, and B. Vrdoljak.*Logical Design of Data Warehouses from XML* . In ConTEL '05: Proceedings of the 8th international conference on Telecommunications, volume 1, pages 289–295, 2005.

[15.] 12. A multisession-based multidimensional model M. Body, M. Miquel, Y. B´edard, and A. Tchounikine. *A multidimensional and multi version structure for OLAP applications*. In DOLAP '02: Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP, pages 1–6, New York, NY, USA,2002. ACM.

[16.] 13. Discovering business intelligence from online product reviews: A rule-induction framework.

[17.] W. Chung and H. Chen. *Web-Based Business Intelligence Systems: A Review and Case Studies*. In G. Adomavicius and A. Gupta, editors, Business Computing, volume 3, chapter 14, pages 373–396. Emerald Group Publishing, 2009.

[18.] 14. Transaction Management for a Main-Memory Database

[19.] P. Burte, B. Aleman-meza, D. B. Weatherly, R. Wu, S. Professor, and J. A. Miller. *Transaction Management for a Main-Memory Database*. The 38th Annual South eastern ACM Conference, Athens, Georgia, pages 263–268, January 2001.

Mr. M. Yuvaraja MCA., (M.Phil)., is an Assistant Professor in Department of Computer Science, Bishop Ambrose College , Coimbatore, Tamil Nadu, and India. He has a teaching experience of 3 years in the field of Computer science. He has presented 2 Papers in International and National Conferences and he has published 2 international Journals. He pursuing M.Phil Degree in SNR Sons College, under Bharathiar University, Coimbatore.

Dr. S. Thavamani M.Sc., M.Phil., Ph.D., is an Associate Professor in Department of Computer Applications, Sri Ramakrishna College for Arts and Science (Formerly SNR Sons College), Coimbatore, Tamil Nadu, and India. She has a teaching experience of 17 years in the field of Computer science. Her area of Specialization is Distributed Computing and P2P Networks. She has presented more than 18 Papers in various International and National Conferences and she has published 11 international Journals. She had authored two books 1.Introduction to Data Structures, 2. Introduction to Assembly of Computer Architecture. She is currently a supervisor for M.Phil and Ph.D research works of