I J C T A, 9(26) 2016, pp. 351-356

Opinion Mining Using Sentence Level on Movie Reviews

F. Jabez Samuel[1] and M. Lovelin Ponn Felciah[2]

ABSTRACT

Opinion mining is one of the most difficulttasks of the field of information retrieval. Research community has been publish a number of articles on this topic but animportant increase in interest has been observed during the past decade particularly after the launchof several online social networks. This paper shows how opinion mining is applied to movies reviews. Much research on textual data processing focused on mining and retrieval of factual information, like information retrieval, Internet search, text classification, text clustering and related text mining and natural language giving out tasks. Opinions are objectiveterminologyrelating people's sentiments, appraisals/ feelings to entities, events and their property. A meaning of opinion is very broad. Internet or world wide web (WWW) is a source of huge amount of information where people express their opinion on anything at movie review sites, internet forums different discussion groups, web blogs etc. Opinion plays a very important role in decision making process. It is very difficult for the customer to manually go through all the review about movies. In this paper, it is proposed to extract the feature set from review and the review are confidential as positive or negative using Naïve Bayes, Fuzzy Lattice reasoning classifier.

Keywords: Opinion mining, Sentiment Analysis, Reviews, Sentence level, Naive Bayes classifier, Naive Bayes classifier.

1. INTRODUCTION

Opinions are central to nearly all individual activities and are key influencers of our behaviors. Our beliefs and perceptions of actuality, and the choice we make, are, to a considerable degree, conditioned upon how others see and evaluate the world. For this motive, when we need to make a choice we often seek out the opinions of others. This is not only right for individuals but also true for organization. Opinions and its related concepts such as sentiments, evaluations, attitude, and emotion are the subject of study of sentiment analysis [11] and opinion mining.

Opinion mining is the field of study that analyzepeople'sappraisals, attitudes, opinions, sentiment, evaluations and emotions towards entities such as , services, organization, individuals, events, topics, issues and their attributes. It represents a large problem space. There are also a lot of names and slightly special tasks, *e.g.*, sentiment analysis, opinion mining, opinion extraction [9], sentiment mining, subjectivity study, concern analysis, emotion analysis, review mining, etc.While in industry, the period sentiment analysis is more usually used, but in university both sentiment analysis and opinion mining are frequently in employment. Sentimentanalysis and opinion mining mostlyfocus on opinions which express or imply positive or negative sentiments.

[1] Research Scholar, Computer Science, Bishop Heber College, Tiruchirapalli, Tamil Nadu, India.

[2] Assistant Professor, Computer Applications, Bishop Heber College, Tiruchirapalli, Tamil Nadu, India.

2. OPINION MINING AND METHODOLOGY

2.1 Different Levels of Opinion Mining

2.1.1 Document level Opinion Mining

Normally Opinions are not stored in full document [17], unless it is some organizational feedback report. But client share their opinions in blogs, forums whichare not in text form. Hence for customer opinion mining article level mining is not appropriate. It is much useful for high level or formal feedback or sentiment mining

2.1.2 Sentence level Opinion Mining

In this technique, individual sentences, bearing sentiments in the text are considered for cataloging .In sentence level Opinion Mining; the relations between sentences are calculated. The same document level classificationsmethod can be applied to the sentence level classification the subjective sentences contain opinion words which help in formative the sentiment about the entity. After that sentence arrangement is done into positive and negative classes.

2.1.3 Phrase/feature level Opinion Mining

Opinion Mining can be based on some exact phrases. The phrase level [14] sentiment classification is a great deal more precise in identify opinions. In this technique the phrases that contain opinion words are found out and anexpression level categorization is done. But in some other cases, where contextual relations matters, the result may not be accurate.

2.2 Challenges of Opinion Mining

Identifying synonyms

Different words or phrases can be used to refer to the same value of the object. So, such words should be identified and grouped together. It is a difficult task to recognize these words. A lot of study is required to be done on this issue as it has not been much addressed in the past.

Identifying positive or negative opinion

Correspondingly detecting an opinion as positive, negative or neutral can be a difficult task in opinion mining. A word could be calculated positive in one situation and negative in another situation. This can be tricky to calculate as a sentence can be measured negative because of the use of negative words in it.

2.3 Research Areas in Opinion Mining

- Client feedback for Individual movie
- Identification of highly rated experts
- Overall positive and negative relatives at paragraph level
- improving the accuracy of algorithm for opinion detection
- Ranking of best section or best sentence based on best feature
- Continuous Improvement of the algorithms for opinion detection Decrease the human exertion needed to analyze contents

- Sentiment classification
- Reduction of human effort needed to analyze content

2.4 Methodology

Reviews were from an Internet [14] Movie Database archiverestarts, the Moviesreviews [17], their positive/ negative classification is extracted automatically from ratings, as specified by reviewer. The dataset includes only reviews where stars indicate movie rating or a statistical system. This study used a subset of 200 positive/200 negative opinions.

Naive Bayes Classifier

Itclassifier is a probabilistic classifier based on application of Bayes' theorem (Bayesian statistics) with strong (naive) independence assumptions. A descriptive term for underlying probability model is an "self-determining feature model". A naive Bayes classifierassume that a specific class feature's presence is unrelated to the presence (or absence) of other features. The prospect model for a classifier is a conditional model over a dependent class variable C with a limited outcomes number or classes, conditional on several feature variables F_1 through F_n

$$P(C|F_1...F_n)$$

The issue is that if features number is large or when a feature takes on many values number, then such a model being based on probability tables is infeasible. Using Bayes' theorem

$$p(C|F_1,...,F_n) = \frac{p(C)\,p(F_1,...,F_n|C)}{p(F_1,...,F_n)}$$

Fuzzy lattice reasoning (FLR)

FLR classifier induces rules from training data by increasing a rule's diagonal size to a maximum threshold D_{crit}. FLR is a leaderfollower classifier, learning rapidly in training data's single pass- through. Input data presentation order is significant. The FLR classifier can set out learning without a priori knowledge; but the latter can be supplied to FLR classifier as initial rules set. Rules total number to be learned is not known a priori but, determined on-line during learning. Further FLR classifier training using additional training data, does not delete earlier learning. Specifically, retraining FLR classifier with new data set either enhances previously learned rules or creates new rules. There is only a single parameter to tune, which is maximum threshold size D_{crit}, regulating learning granularity.

FLR Training Algorithm

S0: The first input (a_0, C_0) is memorized. At an instant, there are *c* Known Classes *C*1, ...*Cc* memorized in memory, initially $c = 0$.

S1: Present next input (*ai*, *Ck*), $i = 1, ..., m$ to initial "set" family of rules

S2: If no rules are "set" then

Store input (*ai*, *CK*),

$c = c + 1$,

Go to S1.

Else

Compute $k(a_0, ai)$, $i = 1, ..., c$ of the "set" rules.

S3: Competition among "set" rules:Winner is rule (aJ, CJ) so that $J = \text{argmax}\{k(a_0, ai)$

$\}, i = 1, ..., c.$

S4: The Assimilation Condition:Both

$Z(ai(\vee aJ) \leq \rho$ and $Ci = CJ.$

S5: If Assimilation Condition is satisfied then

Replace aJ by $a_0 \vee_a J$

Else

"reset" the winner (aJ, CJ), Go to S2

3. EXPERIMENT AND RESULTS

The proposed method has been implemented using .NET Technology. Experiments are conducted for sentiment classification[2] using online movie review data. 500 instances (250 positive and 250 negative) were used for evaluation. Following Tables and Fig give the cataloging accuracy, precision and recall for the various classifiers used for classifying the opinion into positive or negative. It is seen from Figure 1, that the classification accuracy achieved by FLR is much better than the Naïve Bayes. Naïve Bayes achieves 14 to 15.34% better classification accuracy than the other classifiers

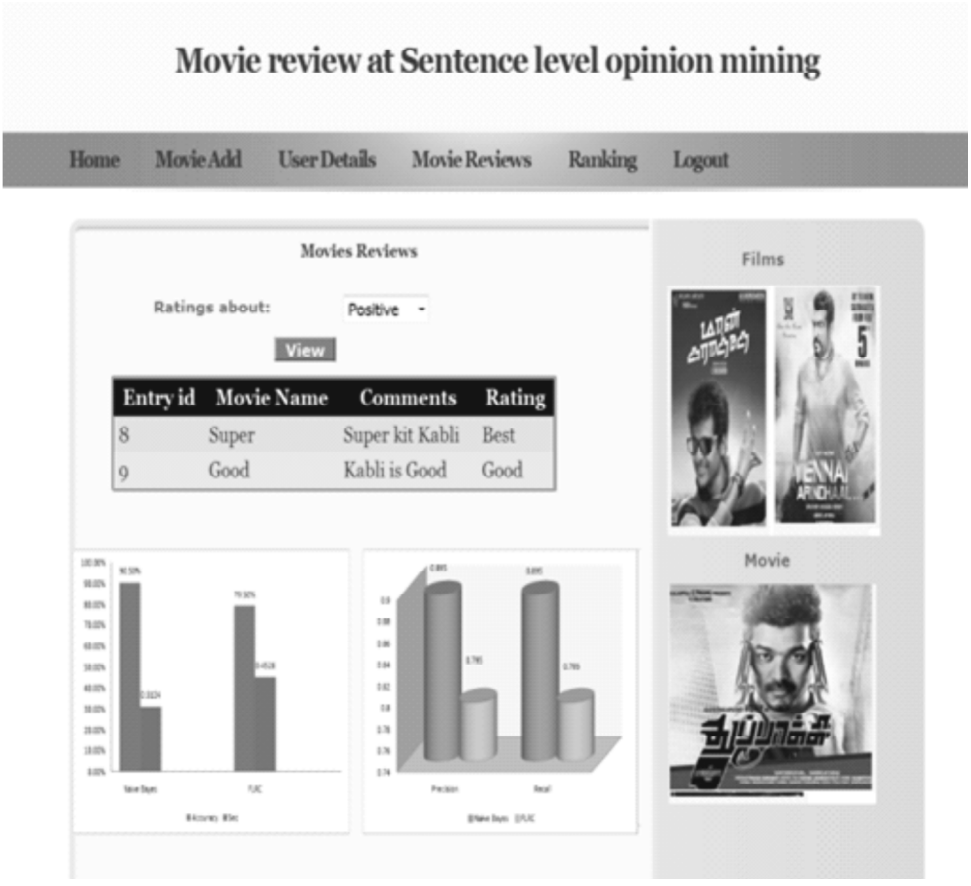

Figure 1: Classification Accuracy and Second

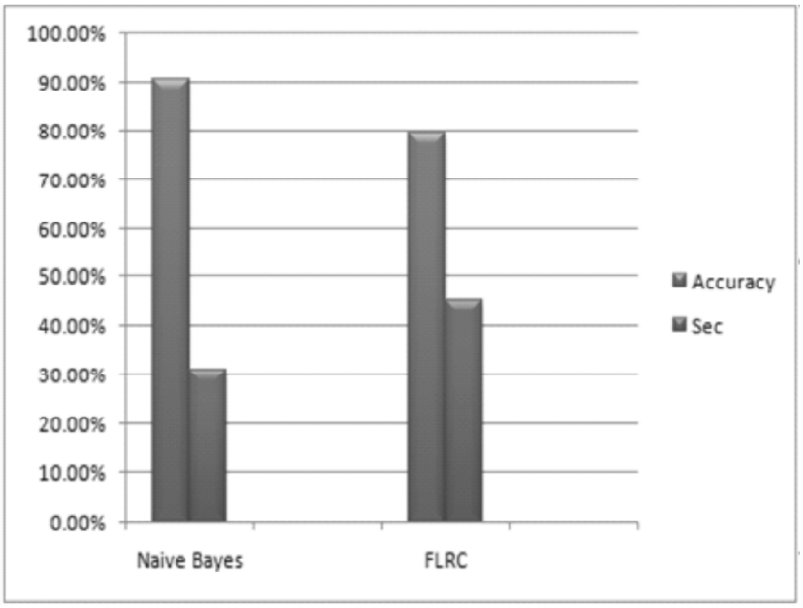

Chart 1: Classification Accuracy and SecFor Various Classifiers Used

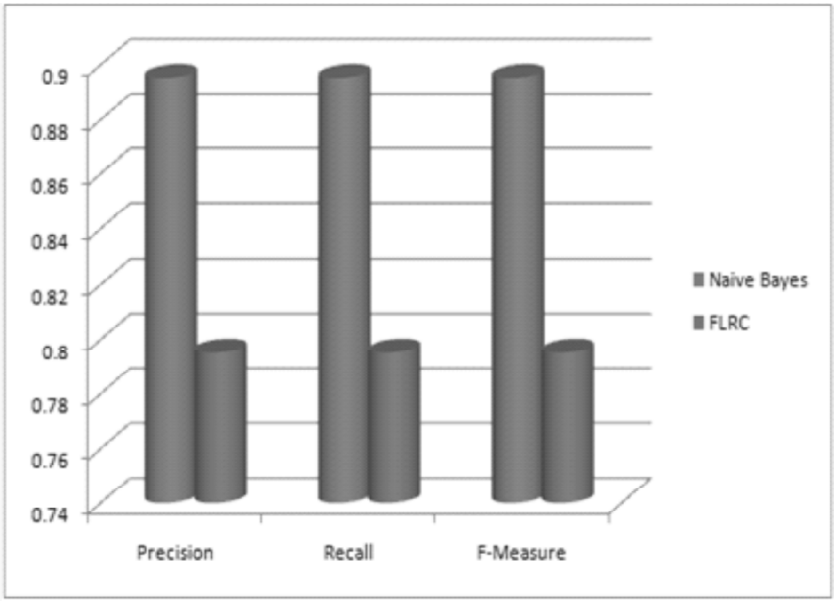

Chart 2: Precision and Recall

Table 1
Classification Accuracy AndSecond For Various Classifiers

Technique used	*Classification Accuracy*	*RMSE*
Naive Bayes	90.5%	0.3124
FLRC	79.5%	0.4528

Table 2
Precision and Recall Values

Technique used	*Precision*	*Recall*	*F-Measure*
Naive Bayes	0.895	0.895	0.895
FLRC	0.795	0.795	0.795

5. CONCLUSION

This paper describe work on mining opinions from unstructured documents. Sentiment analysis [10] has grown to be one of the most active research areas. It has thus become a necessity to collect and studyopinions on the Web. Throughthis literature review, the related works done to solve thisproblem could be studied. Although Many solutions have been proposed to classify sentiments[2] of online reviews, a fully automated and highlyefficient system has not been introduced till now. In this research we propose way to automatically classify movie reviews in terms of positive, negative and neutral classes using hidden markov model approach. The focus was on extracting relations between movie reviews and opinion expressions. Opinion in movie reviews is analyzed/classified as positive/negative. Features are extracted from reviews using Inverse document frequency and reviews are classified through use of the Naïve Bayes [7], and FLR classifier. Experimental results show that Naïve Bayes achieve the best classification. Further investigation based on supervised learning is to be undertaken for improving the classification.

REFERENCES

[1] Peter Turney. "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews." *In Proceedings of the 40th Annual Meeting of the Associationfor Computational Linguistics (ACL)*, 417-424, 2002.

[2] B. Pang, L. Lee, S. Vaithyanathan, "Thumbs up? Sentiment Classification using MachineLearning Techniques," *Proceedings of the 2002 Conference on Empirical Methods inNatural Language Processing*,79-86, 2002.

[3] Aarati Mahadik and Asha Bharambe "Aspect Based opinion Mining and Ranking: Survey" (2015) *International Journal of Current Engineering and Technology*5(6), 3589-3592, 2015.

[4] B. Pang, L. Lee,"Seeing stars: Exploiting class relationships for sentiment categorizationwith respect to rating scales," *Proceedings of the ACL*, 2005.

[13] J. S. Kessler and N. Nicolov, "Targeting sentiment expressions through supervised ranking of linguistic configurations," in *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, San Jose, California, USA, 90-97, 2009.

[6] Tetsuya Nasukawa, and Jeonghee Yi, "Sentiment Analysis: Capturing Favorability UsingNatural Language Processing" *K-CAP*, October, 70-77, 2003.

[7] The Stanford Natural Language Processing Group (http://nlp.stanford.edu/software/lexparser.

[8] Kanayama Hiroshi, Nasukawa Tetsuya and Watanabe Hideo. "Deeper Sentiment Analysis Using Translation Technology", 4-5.

[9] Kushal Dave, Steve Lawrence, and David M. Pennock, "Mining the peanut gallery: Opinion Extraction and semantic classification of product reviews," *in Proceedings of the 12thInternational World Wide Web Conference (WWW)*, 519-528, 2003.

[10] Bing Liu, "Sentiment Analysis and Opinion Mining, Morgan" *& Claypool Publishers"*, 1-167, 2012.

[11] Bo Pang and Lillian Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," *in Proceedings of the 42th Annual Meeting ofthe Association for Computational Linguistics (ACL)*, 1-8, 2004.

[12] Livia Polanyi and Annie Zaenen. "Contextual valence shifters". *In Proceedings of the AAAI Symposium on Exploring Attitude and Affect in Text: Theories and Applications* (publishedas AAAI technical report SS-04-07), 1-6, 2004.

[13] Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, Daniel M. Ogilvie, and associates."The General Inquirer: A Computer Approach to Content Analysis", The MIT Press,1-651, 1966.

[14] Richa Sharma, Shweta Nigam and Rekha Jain (2014) "Mining of product reviews at aspect level". *Internatinal Journal in Foundations of Computer Science & Technology*(IJFCST) **4(3)**, 87-95, 2014.

[15] Beineke, Philip, Trevor Hastie, Christopher Manning, and Shivakumar Vaithyanathan. "Exploring sentiment summarization", *In AAAI Spring Symposium on Exploring Attitudeand Affect in Text: Theories and Applications (*AAAI tech report SS-04-07).

[16] W. Jin, H. H. Ho, and R. K. Srihari, "OpinionMiner: a novel machine learning system for web opinion mining and extraction," presented at the Proceedings of the 15th ACM SIGKDD *international conference on Knowledge discovery and data mining*, Paris, France 1195-1204, 2009.

[17] Richa Sharma, Shweta Nigam and Rekha Jain (2014)" Opinion mining of Movie Reviews At Document level". *Internatinal Journal in Foundations of Computer Science & Technology (IJFCST)* **4(3)**, 13-21, 2014.