Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MINING MOVIE REVIEWS – AN EVALUATION

[1]R.SATHEESH KUMAR, [2]Dr.S.VIJAYAN

[1]Assistant Professor ,Department of Information Technology , Hindusthan College of Engineering & Technology , Coimbatore , Tamilnadu , INDIA

[2]Principal, Surya Engineering College, Perundurai, Erode, Tamilnadu, INDIA

E-mail: [1]satheeshpkd@gmail.com , [2]svijayansurya@gmail.com

ABSTRACT

Much research on textual information processing focused on mining and retrieval of factual information, like information retrieval, Web search, text classification, text clustering and related text mining and natural language processing tasks. Opinions are subjective expressions describing people's sentiments, appraisals/feelings to entities, events and their properties. A definition of opinion is very broad. In this paper, it is proposed to extract the feature set from reviews and the reviews are classified as positive or negative using Naïve Bayes, Ada Boost and Fuzzy Lattice reasoning classifier**.**

Keywords: *Opinion Mining, Sentiment Analysis, Naïve Bayes, Ada Boost, Fuzzy Lattice Reasoning Classifier*

1. INTRODUCTION

The computational study of opinions, sentiments and emotions expressed in text is called Sentiment analysis/opinion mining [1]. Usually, opinions are expressed on anything including products, services, individuals, organizations, events, topics. The term object denotes a target entity to be commented upon. An object has a components set and an attributes set, each having its own sub-components and attributes. So object can be hierarchically decomposed based on part-of relation.

Opinions gain importance due to the fact that whenever people make decisions, they need to hear others' opinions and it is the same for organizations. But, only limited computational studies on opinions existed prior to the Web as there was little opinionated text (text with opinions/sentiments) available. Initial research started with sentiment and subjectivity classification, treating the issue as a text classification problem. Currently, Sentiment classification deals with opinionated document (such as product reviews) and classifies a sentence as expressing positive or negative opinions. Subjectivity classification determines whether a sentence is subjective/objective. But many real-life applications need a detailed analysis as users want to know opinions [2] on subjects.

An opinion holder is a person/organization expressing opinion. In product reviews and blogs, opinion holders are generally the post's authors. Opinion holders are important in news articles as they often clearly state that the person/organization holds a specific opinion [3]. An opinion on a feature f (or object o) is a positive/negative view/appraisal on f (or o) from opinion holders. Positive and negative are called opinion orientations.

Opinions shoulder a major role in decision-making [4]. When people have to make a choice, they are hear others' opinions and if it involves consuming valuable resources like time and/or money, people strongly rely on peers' past experiences. The shift from a read-only to a read-write Web gave people new tools allowing them to create/share their own contents, ideas, and opinions with millions of people on the World Wide Web in a timely and cost-efficient way,. The chance to capture public's opinions about product preferences, social events, political movements, marketing campaigns, and company strategies created high interest in both the scientific community and business world. It raised more interest regarding emerging challenges and in the latter for remarkable fallouts in marketing/financial market prediction.

It is extremely difficult to mine opinions/sentiments from natural language as it

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

involves a deep understanding of languages explicit/implicit, regular/irregular, syntactical and semantic rules [5]. Present approaches rely on text parts where opinions/sentiments are clearly expressed like polarity terms, affect words and co-occurrence frequencies. But, opinions/sentiments are usually conveyed through latent semantics, which ensure that syntactical approaches are ineffective.

A basic technology in present opinion-mining and sentiment-analysis applications is that classification includes regression and ranking [6]. The reason for classification's importance is that many interesting problems can be formulated as applying classification/regression/ranking to textual units. Binary classification labeling of an opinionated document which expresses an overall positive/negative opinion is called sentiment polarity classification or polarity classification. With increasing availability of opinion-rich resources like online reviews and personal blogs, many opportunities/challenges come up as people now can, and do, actively use information technology to seek out/understand others opinions. A sudden eruption of opinion mining and sentiment analysis activity which deals with opinion, sentiment and subjectivity's computational treatment has occurred due to a direct response to interest in new systems dealing directly with opinions as first-class objects.

In this paper, opinion in movie reviews is analysed and classified as positive or negative. The features are extracted from the reviews using Inverse document frequency and the reviews are classified using the Naïve Bayes, Ada Boost and FLR classifier.

2. RELATED WORKS

Poirier et al [7] explored 2 differing opinion extraction methods. The first relied on machine learning technique based Naive Bayesian classifier whereas the second applied NLP techniques to process opinions and build dictionaries which determine a comment's polarity based on its words. Both approaches were evaluated with contents from flixster.com. The results proved that using a low-level NLP approach with a small corpus led to good training: a lexicon building cost and negation detection designing remained reasonable. When the corpus was large, ML approach could be deployed easily.

Jakob et al [8] evaluated whether an anaphora resolution algorithm could improve a baseline opinion mining system's performance. Based on two different anaphora resolution systems, an analysis was presented. Experiments on a movie review corpus demonstrated that unsupervised anaphora resolution algorithm greatly improved target extraction in opinions.

De Freitas et al [9] proposed/evaluated methods to identify Portuguese user generated reviews polarity based on features described in domain ontologies. This method includes 4 steps. To being with, the algorithm received a preprocessed set of reviews as input, after identification of aspects in reviews using ontology terminology. Polarity measurement is based on a lexicon of tagged positive/negative/neutral opinion words. Finally, module tuples with object features and polarity are generated in opinion mining.

A movie-rating and review-summarization system in a mobile environment was designed and developed by Liu et al [10]. Movie-rating information is based on sentiment-classification results. Condensed movie review descriptions are generated from a feature-based summarization. The authors propose a new approach based on latent semantic analysis (LSA) for product features identification. Also, summary size was reduced, based on product features from LSA. Both sentiment-classification accuracy and system response time were taken into consideration in system designing. Rating and review-summarization system is flexible to be extended to other product-review domains easily.

Ghose et al [11] explored review text's multiple aspects like subjectivity levels, varied readability measures and spelling errors extent for text-based features identification. Additionally, multiple reviewer-level features like past reviews average usefulness and reviewers self-disclosed identity measures were examined. Econometric analysis reveals that reviews extent of subjectivity, informativeness, readability, and linguistic correctness matters in influencing sales and its perceived usefulness. Reviews mixing objective, and highly subjective sentences are linked negatively to product sales, as compared to reviews which include only subjective/objective information. But, such reviews are rated as informative (helpful) by users. Usage of Random Forest based classifiers demonstrated that review impact on sales and perceived usefulness is capable

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of being accurately predicted. Three broad categories relative importance: 'reviewer-related' features, 'review subjectivity' features, and 'review readability' features, when examined revealed that using any of the 3 feature sets lead to statistically equal performance as when using all features.

Pak et al [12] focused on using Twitter, a popular micro-blogging platform, for sentiment analysis where automatic collection of a corpus for sentiment analysis and opinion mining was revealed. Linguistic analysis was done on collected corpus and discovered phenomena explained. A sentiment classifier was built with the corpus capable of determining a document's positive/negative/neutral sentiments. Evaluations prove the proposed techniques to be efficient, performing better than previous methods.

A method to automatically create a reference corpus for training text classification procedures was proposed and evaluated by Sarmento et al [13]. This related to political opinions being mined in user-generated content. It includes compiling a highly opinionated user posted comment collection of an on-line newspaper. Then, manually-crafted high-precision rules sets were defined and used, backed by a large sentiment-lexicon to identify comments in sentences which expressed opinions on political entities. Finally, opinions are propagated to remaining sentences of the comment mentioning same entities, thereby increasing opinion-bearing sentences number and variety. Results revealed that most rules identified negative opinions with high precision, with these being safely propagated to remaining sentences in the comment in 100% cases. Due to irony based problems, identification precision drops for positive opinions, but many rules still reach high precision. Propagation of Positive opinions propagation is correct in 77% of cases, and errors here are due to irony and polarity inversion throughout the comment.

3. METHODOLOGY

Movie reviews data set by Pang and Lee (2004) [14] with 2,000 movie reviews: 1,000 positive and 1,000 negative evaluated classification algorithms. This data set's earlier version having 700 positive and 700 negative reviews was used in Pang et al. (2002) [15]. Reviews were from an Internet Movie Database (IMDb) archive rec.arts.movies.reviews. 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 numerical system. This investigation used a subset of 200 positive/200 negative opinions.

A Naive Bayes classifier [16] 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 "independent feature model". A naive Bayes classifier assumes that a specific class feature's presence is unrelated to the presence (or absence) of other features. The probability 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.

$$p(C \mid F_1, \ldots, F_n)$$

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

$$p(C \mid F_1, \ldots, F_n) = \frac{p(C)\, p(F_1, \ldots, F_n \mid C)}{p(F_1, \ldots, F_n)}$$

AdaBoost [17] uses a weighted classifiers sequence, each forced on learning a different data aspect, to generate a final, comprehensive classifier, which having high probability outperforms misclassification error rate of an individual classifier. The basic steps of the algorithm

1: Initialize weights $w_i = \frac{1}{n}$

2: for m = 1 to M do

3: fit y = $h_m(x)$ as the base weighted classifier using w_i and d

4: let

$$W\_(h_m) = \sum_{i=1}^{N} w_i I\{y_i h_m(x_i) = -1\} \; and \;\; \alpha_m = \log\left(\frac{1 - W - (h)}{W - (h)}\right)$$

5:

$$w_i = w_i \exp\{\alpha_m I\{y_i \neq h_m(x_i)\}\} \text{ scaled to sum to one } \forall i \in \{1, \ldots, N\}$$

6: end for

AdaBoost algorithm [18] is an iterative procedure combining many *weak* classifiers to approximate the Bayes classifier $C_*(\boldsymbol{x})$. Starting

Journal of Theoretical and Applied Information Technology
20th October 2013. Vol. 56 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

with unweighted training sample, AdaBoost builds a classifier. For example a classification tree producing class labels. When a training data point is misclassified, training data point's weight is increased. A second classifier built with new weights is not equal anymore. Again, misclassified training data have weights boosted with the process being repeated. One can build 500/1000 classifiers thus. Each classifier is assigned a score and final classifier defined as a linear combination of all stage classifiers.

Fuzzy lattice reasoning (FLR) classifier induces rules from training data by increasing a rule's diagonal size to a maximum threshold $D_{crit.}$ FLR is a leader-follower classifier [19], 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.

The 4 essentials of fuzzy lattice reasoning are,

First, according to Assimilation Condition, rule induction may be effected by replacing a hyperbox A_J by a larger hyperbox$ai \vee A_J$. **Second**, a rule $A_l \rightarrow Cl$, l = 1,. . . ,L defines a fuzzy set $k(x \le A_l)$ in the family of hyperboxes so that hyperbox A_l corresponds to core of fuzzy set $k(x \le A_l)$. **Third**, fuzzy lattice reasoning deals with semantics in 2 different senses: (1) Occam razor semantics as explained above, and (2) non-numeric data, e.g. structured data(graphs), etc., can be accommodated in constituent lattice. **Fourth**, FLR classifier deals with a missing data value in a constituent lattice Liby replacing a missing datum with lattice interval [a,b] so that vi([a,b]) = vi(hi(a)) +vi(b) ffi 0.

FLR Training Algorithm [19]

S0. The first input ($a0$, $C0$) is memorized. At an instant, there are c Known Classes $C1$, . . . ,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 $k(a0, ai)$, $i = 1, . . . , c$ of the "set" rules.
S3. Competition among "set" rules:
Winner is rule (aJ,CJ) so that $J = \mathrm{argmax}\{k(a0, ai)\}$, $i = 1, . . . , c$.
S4. The Assimilation Condition:
Both $Z(ai \vee aJ) \le \rho$and $Ci = CJ$.
S5. If Assimilation Condition is satisfied then
Replace $aJ$by $a0 \vee aJ$.
Else
"reset" the winner (aJ,CJ), Go to S2.

4. RESULTS

Experiments are conducted for sentiment classification using online movie review data. 400 instances (200 positive and 200 negative) were used for evaluation. Following Tables and Figures give the classification 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 Naïve Bayes is much better than that of Ada Boost and FLR. Naïve Bayes achieves 14 to 15.34% better classification accuracy than the other classifiers.

Table 1: Classification Accuracy And RMSE For Various Classifiers Used

Technique used	Classification Accuracy	RMSE
Naive Bayes	88.5%	0.3124
Ada Boost	73.25%	0.4274
FLRC	79.5%	0.4528

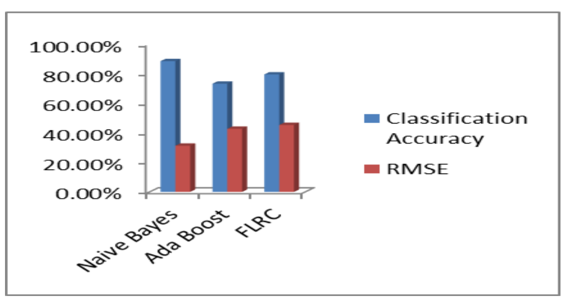

Figure 1: Classification Accuracy And RNSE For Various Classifiers Used

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 3: Precision And Recall Values

Technique used	Precision	Recall	F-Measure
Naive Bayes	0.887	0.885	0.885
Ada Boost	0.733	0.733	0.732
FLRC	0.795	0.795	0.795

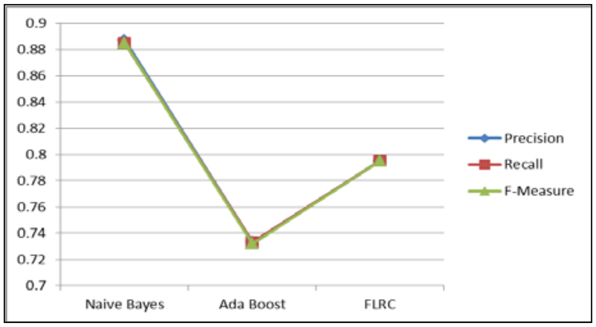

Figure 2: Precision And Recall

5. CONCLUSION

This paper described work on mining opinions from unstructured documents. 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, Ada Boost 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] Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, *2*, 568.

[2] Chen, H., & Zimbra, D. (2010). AI and opinion mining. *Intelligent Systems, IEEE*, *25*(3), 74-80.

[3] Kim, S. M., & Hovy, E. (2006, July). Extracting opinions, opinion holders, and topics expressed in online news media text. In Proceedings of the Workshop on Sentiment and Subjectivity in Text (pp. 1-8). Association for Computational Linguistics.

[4] Cleland-Huang, J., & Mobasher, B. (2008, May). Using data mining and recommender systems to scale up the requirements process. In Proceedings of the 2nd international workshop on Ultra-large-scale software-intensive systems (pp. 3-6). ACM.

[5] Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In Proceedings of the National Conference on Artificial Intelligence (pp. 755-760). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

[6] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, *2*(1-2), 1-135.

[7] Poirier, D., Bothorel, C., De Neef, E. G., & Boullé, M. (2011). Automating opinion analysis in film reviews: the case of statistic versus linguistic approach. In Affective Computing and Sentiment Analysis (pp. 125-140). Springer Netherlands.

[8] Jakob, N., & Gurevych, I. (2010, July). Using anaphora resolution to improve opinion target identification in movie reviews. In Proceedings of the ACL 2010 Conference Short Papers (pp. 263-268). Association for Computational Linguistics.

[9] Freitas, L. A., & Vieira, R. (2013, May). Ontology based feature level opinion mining for portuguese reviews. In Proceedings of the 22nd international conference on World Wide Web companion (pp. 367-370). International World Wide Web Conferences Steering Committee.

[10] Liu, C. L., Hsaio, W. H., Lee, C. H., Lu, G. C., & Jou, E. (2012). Movie rating and review summarization in mobile environment. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 42(3), 397-407.

[11]Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. Knowledge and Data Engineering, IEEE Transactions on, 23(10), 1498-1512.

[12] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of LREC (Vol. 2010).

[13] Sarmento, L., Carvalho, P., Silva, M. J., & de Oliveira, E. (2009, November). Automatic creation of a reference corpus for political opinion mining in user-generated content. In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion (pp. 29-36). ACM.

[14]Bo Pang and Lillian Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of ACL 2004, pp.271-278.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[15]Bo Pang, Lillian Lee and Shivakumar Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP 2002, pp.79-86.

[16] Culp, M., Johnson, K., & Michailides, G. (2006). ada: An r package for stochastic boosting. *Journal of Statistical Software*, *17*(2), 9.

[17] Zhu, J., Zou, H., Rosset, S., & Hastie, T. (2009). Multi-class adaboost.*Statistics and Its*.

[18] Kaburlasos, V. G., Athanasiadis, I. N., & Mitkas, P. A. (2007). Fuzzy lattice reasoning (FLR) classifier and its application for ambient ozone estimation.*International Journal of Approximate Reasoning*, *45*(1), 152-188.

[19] Khezeli, Y. J., & Nezamabadi-pour, H. (2012). Fuzzy lattice reasoning for pattern classification using a new positive valuation function. *Advances in Fuzzy Systems*, *2012*, 14.