ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 2, February 2018

Ensemble Learning as Opinion Mining Approach: A Survey

[1] Padmapani P. Tribhuvan, [2] Sunil G. Bhirud, [3] Ratnadeep R. Deshmukh
[1] DIEMS, Aurangabad, [2] VJTI, Mumbai, [3] Dr. B.A.M.U., Aurangabad.

Abstract: **User generated data on the web has many research challenges and thus has attracted many researchers. As result, many new disciplines are evolved and opinion mining is one of them. Opinion mining deals with reviews expressed on the web. Opinion mining analyses views, sentiments, opinions, attitudes and emotions expressed in reviews. There are different approaches to solve the problem of opinion mining. Ensemble learning is one of the paradigms of machine learning in which multiple learners are used to solve the same problem. It has been used in different types of application efficiently and effectively. In the discipline of opinion mining, different ensembles are proposed by researchers. This survey focuses on opinion mining using ensemble learning approach. We discussed different ensembles used to solve problem of opinion mining.**

Keywords: **- Aspect Level Sentiment Analysis, Ensemble Learning, Feature Based Opinion Mining, Machine Learning, Opinion Mining, Sentiment Classification Sentiment Analysis.**

I. INTRODUCTION

In the age of Big Data, data on web is increasing rapidly. There are many challenges in managing and analyzing this data. As a result, many new disciplines, like text summarization, information retrieval, machine learning, text mining, opinion mining, recommendation systems etc. are evolved. 'Opinion mining can be defined as a discipline of computational linguistics and information retrieval which is concerned not with the topic a document is about but the opinion it expresses' [1]. It gives whether reviewer expresses positive, negative or neutral opinion. Opinion mining is also known as sentiment analysis. Opinion mining is achieved by using techniques of Natural Language Processing, Information Retrieval, and Machine Learning or by combining these techniques.

There are three different evaluative tasks of opinion mining: 1. Sentiment classification, 2. Feature-based opinion mining and summarization, 3. Comparative sentence and relation mining [1]. Sentiment classification is a classification problem that classifies a whole document into positive, negative or neutral class. Feature-based opinion mining and summarization find outs what are the features about which opinion expressed and which type opinion expressed. It also generates a summary. Comparative sentence and relation mining compares one object with one or more similar objects.

Opinion mining can applied on variety of data. Following are domains of opinion mining applications:

1. Product Reviews
2. News Articles
3. Movie Ratings
4. Twitter Data
5. Blogs
6. Microblogs etc.

Recommendation systems are evolved from the concept of opinion mining and they are widely used today in day today life.

The rest of the paper is organized as follows: Section II we discussed Sentiment Classification. Section III is about Feature Based Opinion Mining. In Section II and Section III we discussed opinion mining process with respect to the product review domain rather than in general. Section IV deals with Ensemble Learning, Section V is Opinion Mining using Ensemble Learning and Section VI concludes the paper.

II. SENTIMENT CLASSIFICATION

Sentiment classification is opinion mining task at document level. This is considered as simple classification problem. The problem of sentiment classification can be defined as follows:

Given a document d, classify the document in positive, negative or neutral.

Following approaches are used to accomplish sentiment classification task.

1. Semantic Orientation Approaches
2. Machine Learning Approaches

Document level opinion mining or sentiment classification assumes that a single document is expressing a review of single entity which is not the case every time. A single document may describe more than on entity.

One more problem is that even if document is classified as positive i.e. document expresses positive opinion about an entity, it doesn't mean that all features of that entity is lie by reviewer. Therefore, feature level analysis is needed.

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 2, February 2018

III. FEATURE BASED OPINION MINING

Feature based opinion mining is one of the opinion mining tasks. It is also known as Aspect Level Sentiment Analysis. In general, it needs to perform three tasks (as shown in Fig.1): 1. Feature Extraction, 2 . Classification / Polarity Identification and 3. Summarization.

The problem of Feature Based Opinion Mining can be defined as follows:

An opinion in feature based opinion mining task is triple (f, o, p) where f is feature, o is opinion word and p is polarity (positive, negative or neutral). Feature based opinion mining find outs such opinions from set of review documents and summarizes the findings.

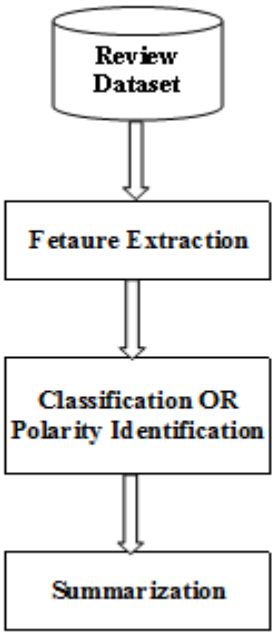

Fig1. : Feature Based Opinion Mining Tasks

A. Feature Extraction

In feature extraction, from a review features about which opinion is expressed is extracted. It is also known as Aspect detection or Aspect Extraction. There are different approaches for feature extraction. These approaches can be classified as follows:

1. Frequency-based Approaches
2. Relation-based Approaches
3. Machine Learning Approaches
4. Hybrid Approaches

B. Classification / Polarity Identification

In classification or polarity identification, feature wise opinions are classified into either into positive or negative. Following are the approaches that are used for classification:

1. Lexicon-based Approaches
2. Relation-based Approaches
3. Machine Learning Approaches
4. Hybrid Approaches

C. Summarization

In summarization task, findings of previous two tasks are summarized. There are different machine approaches to solve the problem of feature based opinion mining. Ensemble learning is one of the paradigms of machine learning. In this paper, we focused on this approach.

IV. ENSEMBLE LEARNING

Ensemble learning is a machine learning paradigm where learners are trained to solve a problem. An ensemble has different base learners' also known as weak learners. These are generated by base learning algorithms. Ensemble learning boosts these weak base learners. Fig2. Shows the general architecture of ensemble. Training data is used to train the different machine learning classification models. Then output of all models is combined using principles like Majority Vote principle. Then final class is predicted. To construct ensemble a number of base learners are generated and then they are combined to use. Following are three ensemble methods:

A. Boosting

It is an ensemble for reducing bias and variance. In [2], boosting ensemble is introduced. Boosting is used to convert weak learners to strong learners. Prediction of each weak learner is combined to form strong learner. Decision stamp, margin-maximizing classification algorithm, AdaBoost (Adaptive Boosting), Gradient Tree Boosting and XGBoost are used for boosting algorithm. Boosting algorithms are most widely used in data science.

B. Bagging

Bagging predictors is a method for generating multiple versions of a predictor and using these to get an aggregated predictor [3]. It is also known as Bootstrap aggregating. It is designed to improve the stability and accuracy. Bagging avoids over fitting. It reduces variance.

C. Stacking

It is also called as stack generalization. Stacking ensembles diverse and strong set of learners. Random forest is example stacking ensemble. Random forests are an effective tool in prediction [4].

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)
Vol 3, Issue 2, February 2018

In [5], Kaur and Kaur recommended not to use any of the ensemble methods with Naïve Bayes, Logistic regression and Voted Feature Interval learners.

V. OPINION MINING USING ENSEMBLE LEARNING

This section discusses how ensemble learning is used for opinion mining by the researchers. Fersini et al. proposed ensemble method for sentiment classification. This method is based on Bayesian Model averaging. They considered two issues. One is the identification of a suitable weighting schema for training the supervised classifiers and another one is the comparison of BMA with state of the art approaches, i.e. baseline classifiers and traditional ensembles [6]. In [7], Wang et al. conducted a comparative assessment of the performance ensemble methods. They used five base learners namely Naive Bayes, Maximum Entropy, Decision Tree, K Nearest Neighbor, and Support Vector Machine. Out of Random Subspace, Bagging and Boosting they find that result of Random Subspace is better. Silva et al. applied classifier ensemble on Twitter micro blogging for opinion mining. They classified Tweets into positive class or negative class. To improve classification accuracy they proposed an ensemble classifier using Multinomial Naive Bayes, SVM, Random Forest, and Logistic Regression [8].

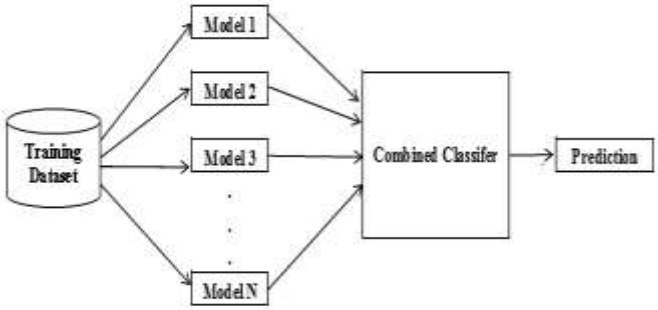

Fig2. : General Architecture of Ensemble

In [9], Hassan et al. proposed a bootstrap ensemble framework for opinion mining of Tweets. Their proposed approach, as compared to various comparison tools and algorithms, is more accurate and balanced in its predictions across sentiment classes. The framework controls class imbalance, sparsity issues. In [10], Xia at el. applied a framework to sentiment classification task. They created two feature sets and applied three types of ensemble methods, namely the fixed combination, weighted combination and meta-classifier combination. Ensemble is formed using naïve Bayes, maximum entropy and support vector machine. Poria at el. proposed a multimodel framework for sentiment analysis of video contents. They extract sentiment from visual modality using SVM, audio modality using SVM and text modality using hybrid classifier CNN-SVM [11].

In [12], AL-Sharuee at el. proposed completely automatic approach for sentiment analysis. They used unsupervised ensemble learning. They used modified k-means as base classifier of ensemble. The proposed approach gives better the clustering performance in term of accuracy, stability and generalization ability.

In [13], Appel at el. proposed two specific aggregation semantics, namely fuzzy-majority based on Induced Ordered Weighted Averaging and compensatory based on Uninorm operators. They used ensemble learning for sentence level sentiment classification.

In [14], Akhtar at el. proposed various deep learning models based on Convolutional Neural Network (CNN), Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). They develop ensemble of these deep learning models and Support Vector Regression (SVR) based supervised model. In [15], Wan and Gao proposed an ensemble for Airline Service Industry Twitter data sentiment classification. This ensemble is constructed using Naive Bayes, Bayesian Network, SVM, C4.5 Decision Tree and Random Forest algorithms. It is based on Majority Vote principle. Alnashwan at el. proposed an ensemble model based on the meta-level features. This model constructed using four learners Support Vector Machine, Bayes Point Machine, Logistic Regression and Decision Forest). It is applied on Twitter data [16].

VI. CONCLUSION

The problem of opinion mining can be solved using different approaches. Machine learning approach is more popular approach. Ensemble learning is one of the machine learning paradigms which is also used to solve problem of opinion mining. As compare to other approaches ensemble learning gives better performance in term of accuracy and stability. In this paper, we discussed some of the ensembles proposed for opinion mining.

REFERENCES

[1] Liu, B., 2007. Web data mining: exploring hyperlinks, contents, and usage data. Springer Science & Business Media.

[2] Schapire, Robert E. "The strength of weak learnability." Machine learning 5, no. 2 (1990): 197-227.

[3] Breiman, Leo. "Bagging predictors." Machine learning 24, no. 2 (1996): 123-140.

[4] Breiman, Leo. "Random forests." Machine learning 45, no. 1 (2001): 5-32.

[5] Kaur, Arvinder, and Kamaldeep Kaur. "Performance analysis of ensemble learning for predicting defects in open source software." In Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on, pp. 219-225. IEEE, 2014.

[6] Fersini, Elisabetta, Enza Messina, and Federico Alberto Pozzi. "Sentiment analysis: Bayesian ensemble learning." Decision support systems 68 (2014): 26-38.

[7] Wang, Gang, Jianshan Sun, Jian Ma, Kaiquan Xu, and Jibao Gu. "Sentiment classification: The contribution of ensemble learning." Decision support systems 57 (2014): 77-93.

[8] Da Silva, Nadia FF, Eduardo R. Hruschka, and Estevam R. Hruschka Jr. "Tweet sentiment analysis with classifier ensembles." Decision Support Systems 66 (2014): 170-179.

[9] Hassan, Ammar, Ahmed Abbasi, and Daniel Zeng. "Twitter sentiment analysis: A bootstrap ensemble framework." In Social Computing (SocialCom), 2013 International Conference on, pp. 357-364. IEEE, 2013.

[10] Xia, Rui, Chengqing Zong, and Shoushan Li. "Ensemble of feature sets and classification algorithms for sentiment classification." Information Sciences 181, no. 6 (2011): 1138-1152.

[11] Poria, Soujanya, Haiyun Peng, Amir Hussain, Newton Howard, and Erik Cambria. "Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis." Neurocomputing 261 (2017): 217-230.

[12] AL-Sharuee, Murtadha Talib, Fei Liu, and Mahardhika Pratama. "An Automatic Contextual Analysis and Clustering Classifiers Ensemble approach to Sentiment Analysis." arXiv preprint arXiv:1705.10130 (2017).

[13] Appel, Orestes, Francisco Chiclana, Jenny Carter, and Hamido Fujita. "IOWA & Cross-ratio Uninorm operators as aggregation tools in sentiment analysis and ensemble methods." In Fuzzy Systems (FUZZ-IEEE), 2017 IEEE International Conference on, pp. 1-6. IEEE, 2017.

[14] Akhtar, Md Shad, Abhishek Kumar, Deepanway Ghosal, Asif Ekbal, and Pushpak Bhattacharyya. "A Multilayer Perceptron based Ensemble Technique for Fine-grained Financial Sentiment Analysis." In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 540-546. 2017.

[15] Wan, Yun, and Qigang Gao. "An ensemble sentiment classification system of twitter data for airline services analysis." In Data Mining Workshop (ICDMW), 2015 IEEE International Conference on, pp. 1318-1325. IEEE, 2015.

[16] Alnashwan, Rana, Adrian P. O'Riordan, Humphrey Sorensen, and Cathal Hoare. "Improving sentiment analysis through ensemble learning of meta-level features." In KDWEB 2016: 2nd International Workshop on Knowledge Discovery on the Web. Sun SITE Central Europe (CEUR)/RWTH Aachen University, 2016.