

STACKING ENSEMBLE MODEL FOR POLARITY CLASSIFICATION IN FEATURE BASED OPINION MINING

Padmapani P. Tribhuvan

Department of Computer Science Engineering,
Deogiri Institute of Engineering and Management Studies, Aurangabad, India.
padmapanitribhuvan@dietms.org

Sunil G. Bhirud

Department of Computer Engineering and IT,
Veermata Jijabai Technological Institute, Mumbai, India.

Ratnadeep R.Deshmukh

Department of CS and IT,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India.

Abstract - We proposed stacking ensemble model to solve the problem of feature-based opinion mining. We used Naive Bayes, Support Vector Machine and K-Nearest Neighbor as base learner and Support Vector Machine as Meta Classifier. Using domain knowledge the dataset consists of Feature-Opinion-Negation triple is created and trained using the proposed stacking ensemble model. The proposed model predicts feature based opinion polarity identification of 4096 laptop product reviews with 92.5315% accuracy.

Keywords: Stacking; Ensemble Learning; Opinion Mining; Sentiment Analysis; Naive Bayes; Support Vector Machine; K-Nearest Neighbor

1. Introduction

As customer reviews are user generated data, its analysis is very critical task. Opinion mining is the discipline using which these reviews can be analyzed efficiently and effectively. Feature-based opinion mining is one of the tasks of opinion mining. It is also known as Aspect Based Sentiment Analysis. The problem of Feature-based opinion mining can be solved by using different approaches. Machine learning is one the most popular approaches. The machine learning approaches can be classified as supervised and unsupervised. To improve the performance of supervised learning models different classifiers are combined to solve the problem. This is called as ensemble learning. There are three types of ensembles namely bagging, boosting and stacking. In this paper we have proposed the stacking ensemble learning model to solve the problem of feature-based opinion mining.

Stacking is also called as stack generalization. In stacking, different base learners are trained on same dataset. The Meta classifier is trained using output of the base learners. Stacking combines heterogeneous classifiers and gives improved performance.

Stacking ensemble combines multiple classifiers via meta-classifier. The proposed ensemble has two levels. The base level i.e. level 0 consists of three classifiers, namely Naive Bayes, Support Vector Machine, and K-Nearest Neighbor. At level 1 we used Support Vector Machine as meta-classifier.

The paper contributes System architecture for feature based opinion mining and summarization using stacking ensemble. This paper is organized as follows. Section-2 discusses related work, Section-3 gives the overview of the proposed model, Section-4 discusses the experiments and results and Section-5 concludes the paper with future directions.

2. Related Work

There are many ensembles which are proposed for opinion mining. Wang et al. used five base learners namely Naive Bayes, Maximum Entropy, Decision Tree, K Nearest Neighbor, and Support Vector Machine to create an ensemble [1]. In [2], Hassan et al. proposed framework that controls class imbalance, sparsity issues. They proposed bootstrap ensemble framework for opinion mining of Tweets. In [3], Wan and Gao proposed an voting 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. Silva et al. proposed an ensemble to improve classification accuracy. They applied classifier ensemble on Twitter micro blogging for opinion mining. They proposed ensemble classifier using Multinomial Naive Bayes, SVM,

Random Forest, and Logistic Regression [4]. AL-Sharuee et al. proposed unsupervised ensemble learning model. They used modified k-means as base classifier of ensemble. They proposed approach which is completely automatic approach for sentiment analysis [5]. Alnashwan et al. proposed an ensemble model for Twitter data which is based on the meta-level features. This ensemble model constructed using four learners Support Vector Machine, Bayes Point Machine, Logistic Regression and Decision Forest [6].

3. The Overview of Proposed Stacking Ensemble Model

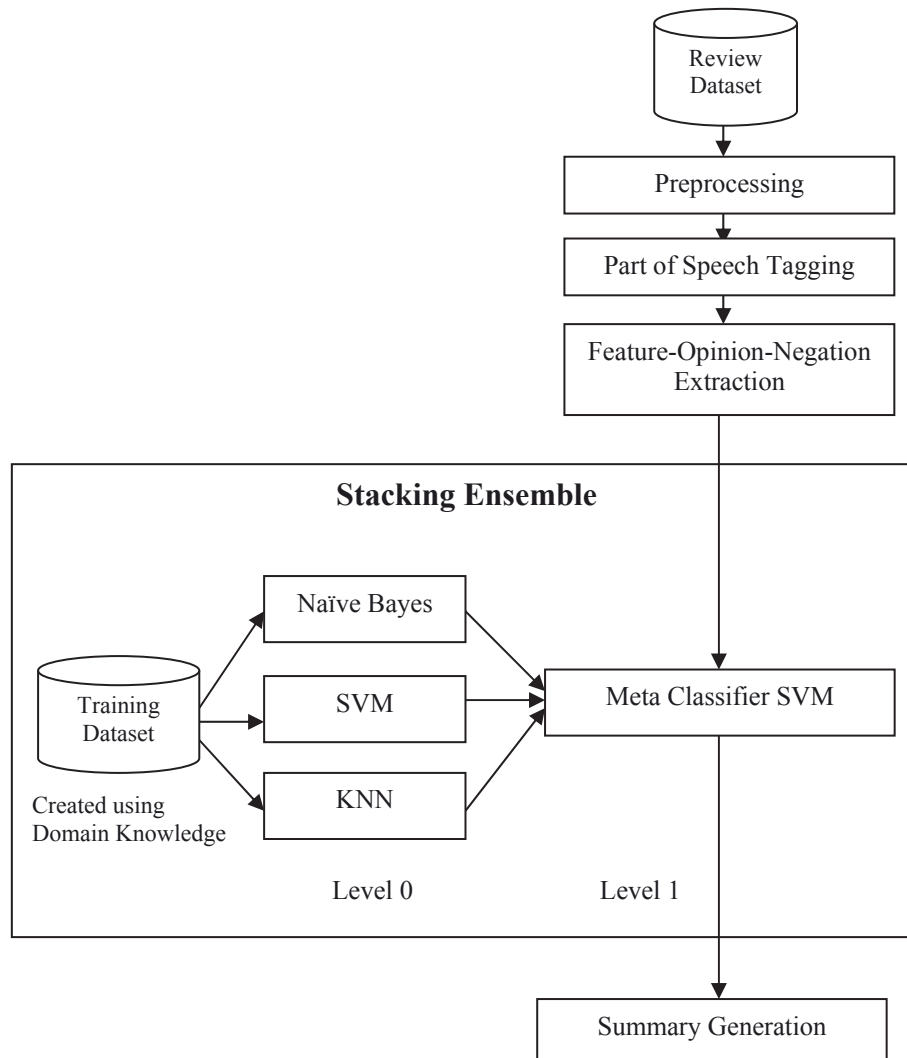


Figure1. System Architecture for Feature Based Opinion Mining using Stacking Ensemble

In this section, we present our method. Proposed system architecture is shown in Figure 1. First we create a dataset based on domain knowledge. The Feature-Opinion-Negation dataset consists of five different attributes namely, ID, Feature, Opinion Word, Number of Negation Words and Polarity. Here polarity is dependent variable and can have value either positive or negative. To create a dataset we considered all possible product features about which opinion can be expressed. Depending on the feature we find out different possible opinion words that can be used for expressing an opinion. Number of negation words attribute is categorical attribute with two categories Even and Odd. We considered that if the number of negation words in sentence are even then the polarity will be the polarity of opinion word in the sentence. If the number of negation words in sentence is odd, then the polarity will be opposite polarity as that of the opinion word in the sentence. Therefore for each feature and opinion word pair we created two records one with even number of negation word and another with odd number of negation word.

Once a Feature-Opinion-Negation Dataset is created using domain knowledge, we train Stacking Ensemble model for Opinion Polarity Classification. Then this model is applied on the product reviews. As product reviews are user generated data, they are unstructured in nature. To convert these product reviews into structured data we apply basic Natural Language Processing techniques. First in Data Preprocessing, we remove unwanted part like, reviewer id, reviewer name etc. from the product reviews. Then we separated sentences in

product review. The number of negation words in the each sentence is count. Part-of-Speech tagging is applied on each sentence. All nouns are extracted as opinion words and adjectives as product feature from each sentence. Based on these, we created Feature-Opinion-Negation dataset for the product reviews.

After creating Feature-Opinion-Negation dataset for product reviews, we applied trained Stacking ensemble model on these dataset for polarity classification. Lastly we generate summary which gives how many positive and negative reviews are expressed on a particular product feature in the product reviews.

4. Experiments

4.1 Creation of Feature-Opinion-Negation Dataset

We created a Feature-Opinion-Negation Dataset for Laptop. The dataset consists of 2062 records. To create dataset we consider 44 features of laptop. 36 different opinion words are considered features. Using domain knowledge Feature-Opinion pair and its polarity is found out. After these we consider two possibilities, one where number of negation words is even and another where number of negation words is odd. As explain in previous section we considered the final polarity. Table 1 shows first 10 records in Feature-Opinion-Negation Dataset we created. In the first record Feature-Opinion Pair is batteri-long. We know that having a long battery life of laptop is positive opinion but number of negation words are odd so negative polarity is considered. Similarly for second record Feature-Opinion Pair is batteri-long. As the number of negation words is even so positive polarity is considered.

4.2 Training Ensemble for Opinion Polarity Classification

We used stacking ensemble for polarity classification. Stacking ensemble is created using Naive Bayes, Support Vector Machine and K-Nearest Neighbor classifiers. We used SVM as Meta classifier.

4.3 Review Dataset and POS Tagging

The review dataset consists of 4096 laptop reviews. In preprocessing, sentences are separated. After preprocessing we got 19278 sentences. Then we performed part-of-speech tagging using Stanford POS tagger and extracted Feature-Opinion-Negation from the sentences. Using this information we created dataset for predictions. The created dataset is consists of 186787 records of 19278 sentences.

Table 1: Feature-Opinion-Negation Dataset (First 10 Records)

ID	Feature	Opinion	Negation	Class
1	batteri	long	odd	negative
2	batteri	long	even	positive
3	batteri	short	odd	positive
4	batteri	short	even	negative
5	batteri	good	odd	negative
6	batteri	good	even	positive
7	batteri	bad	odd	positive
8	batteri	bad	even	negative
9	batteri	nice	odd	negative
10	batteri	nice	even	positive

4.4 Evaluation Metrics

To evaluate proposed model we used following evaluation metrics:

1. Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positive observations.
2. Recall: Recall is the ratio of correctly predicted positive observations to the all observations.
3. Accuracy: Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations

4.5 Result

Table 2 shows the result of proposed model and the table is graphically represented in Figure 2 and Figure 3.

Table 2: Result of Proposed Model

Performance Measure/ Classifier	Precision	Recall	Accuracy
NB	0.602	0.599	60.1843
KNN	0.1	0.101	9.6023
SVM	0.722	0.725	72.2599
Stacking (SVM)	0.926	0.925	92.5315

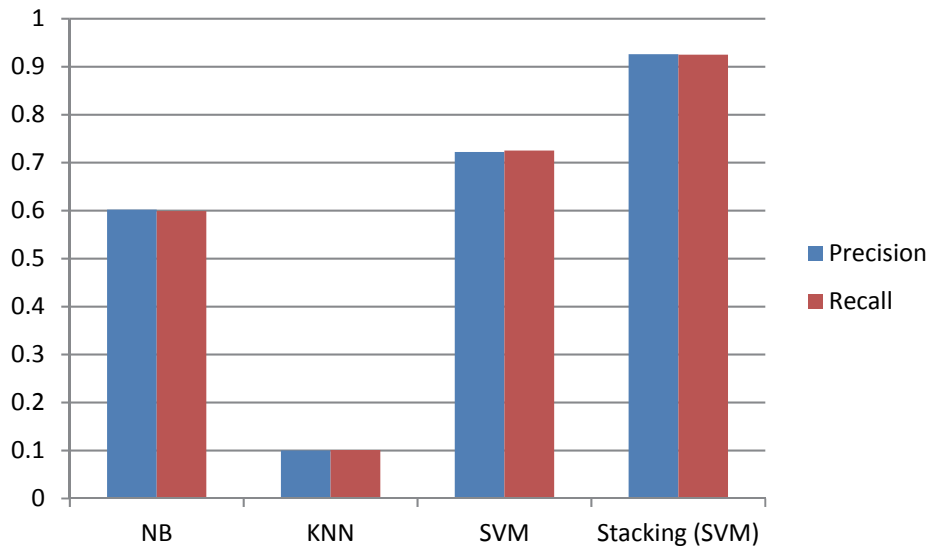


Figure 2: Recall and Precision of Proposed Model

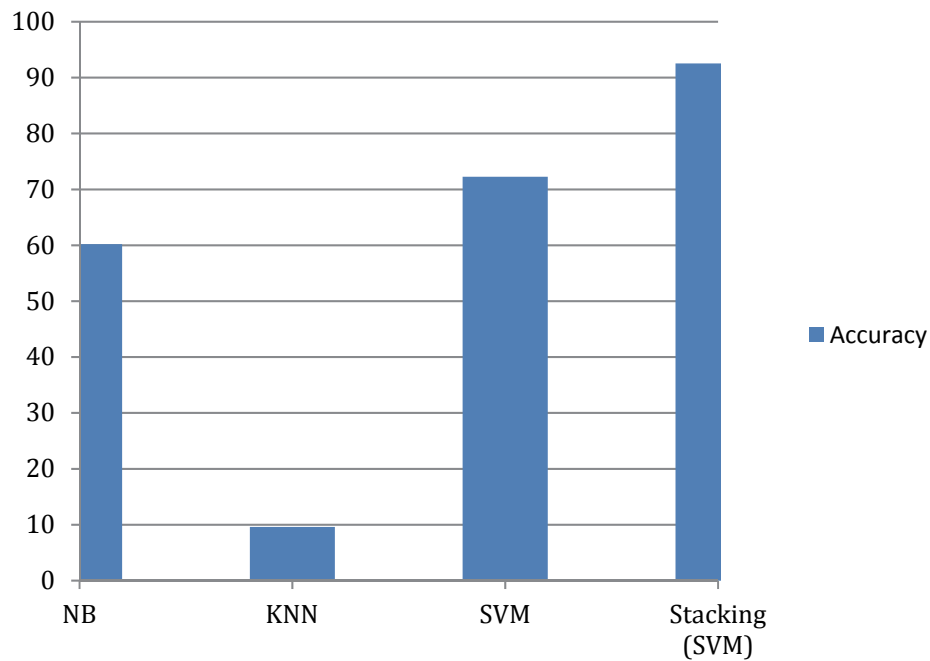


Figure 3: Accuracy of Proposed Model

5. Conclusions and Future Work

The result of proposed model shows that stacking ensemble improves the accuracy of predictions. The KNN has given less accuracy as compared to NB and SVM. The SVM has given more accuracy than other two. The ensemble model has increased the accuracy about 20%.

We have used the model only on Laptop reviews, the model can be applied to different product reviews as well. As the accuracy of prediction is totally depend on the training dataset, one has to prepare training dataset very carefully based on domain knowledge.

References

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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).
- [6] 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.