SOCIETY FOR SCIENCE AND EDUCATION UNITED KINGDOM
TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE
VOLUME 3, ISSUE 1
ISSN 2054-7390

Opinion Mining Using Sequence Labelling

[1]Vijiyalakshmi S and [2]Vijaya M S
PSGR Krishnammal College for women, Coimbatore, India
[1]vijiyalakshmi1886@gmail.com; [2]msvijaya@grgsact.com

ABSTRACT

Opinion mining aims to determine the attitude of a person by identifying and extracting subjective information. The attitude is the judgement, evaluation or emotional state of the person towards a product, or service or a person. An essential task in opinion mining is to classify the polarity of a review at the document, sentence, or feature level whether the expressed opinion is positive, negative or neutral. The main objective of this research work is to formulate opinion mining task as sequence labelling and to build the models for classifying the opinion about the product Kingston Pen drive review as positive or negative. The performances are evaluated and the comparative results are analyzed and reported.

Keywords: Conditional Random Fields; Classification; Opinion mining; Prediction; Sequence labeling; Training.

1 Introduction

Social network analysis has emerged as a key technique in modern sociology and has become a popular topic of study in areas like Business and Economics, Geography, Information science, Organizational studies, social psychology, Sociolinguistics. For example, SNA has been used in epidemiology to understand the pattern of human contacts that cause the spread of diseases in a population. SNA can be used as a tool for market analysis based on opinions about products or brand to market products and services. SNA can also be an effective tool for mass surveillance. For example to determine whether or not a particular individual have criminal tendencies.

Opinion Mining or Sentiment analysis involves construction of a system to explore user's opinions made in blog posts, comments, reviews or tweets, about the product, policy or a topic. Nowadays, the exponential increase in the Internet usage and exchange of user's opinion is the inspiration for Opinion Mining. The Web is a huge repository of prearranged and unstructured data. The analysis of this data to extract the user's opinion and sentiment is a challenging task.

Opinion mining is important for businesses and organizations to know the opinions on their products, competitor products and services. It is very helpful for a company to automatically extract consumer intents from public forums like Blogs, review sites etc., which enable them to spend less expenditure on their market research activities. They can also find comparative opinions in these forums for related products or services. Opinion mining is also important for individuals in finding peers opinion when purchasing a product or subscribing to a service. Opinion mining helps to creates awareness among individuals about the product and features effectiveness and drawbacks.

DOI: 10.14738/tmlai.31.870
Publication Date: 24th February, 2015
URL: http://dx.doi.org/10.14738/tmlai.31.870

Using opinion mining, a review can be evaluated at various levels like document level, sentence level and feature level. In document level evaluation, whole review is classified into either positive or negative depending upon the opinion expressed in that review. When review is evaluated at sentence level, each sentence in a review is classified into either positive or negative. Whereas feature level or feature based opinion mining gives outline which feature of product is liked or disliked by reviewer.

Various techniques and methodologies have been developed to classify opinions. A lot of research works have been carried out in opinion mining. Based on the study of various literatures available on opinion mining a brief report is presented below.

Harb et al. [7] used two sets of seed words with positive and negative semantic orientations to perform blog classification. Google search engine was used to create association rules. Total number of positive and negative adjectives was counted in a document to classify the documents. Authors achieved 0.717 F1 score identifying positive documents and 0.622 F1 score identifying negative documents.

Taboada et al. [8] executed movie review dataset for lexicon-based method to perform sentiment classification. For classification positive and negative words dictionaries were used and semantic orientation calculator (SO-CAL) was built that incorporate intensifiers and negation words. This approach achieved 59.6% to 76.4% accuracy on 1900 documents.

An unsupervised learning algorithm was proposed by Turney et al. [9], used semantic orientation of phrases with adjectives/adverbs for review classification. The approach first extracts phrases with adjectives/adverbs; the phrase’s semantic orientation was estimated using PMI-IR depending on average semantic orientation of phrases, the review is either recommended Thumbs up or not recommended Thumbs down. Experiment used 410 reviews on various topics leading to an average accuracy of 74%.

Turney et al. [10] used “Poor” and “Excellent” seed words to calculate the semantic orientation for the movie review domain, point wise mutual information method is used to calculate the semantic orientation. The sentiment orientation of a document was calculated as the average semantic orientation of all such phrases. 66% accuracy was achieved for the movie review domain.

Valarmathi et al. [11] analysed a method to create exclusive lists from a document’s extracted words. Corpus of words created after exclude list was based on Singular Value Decomposition (SVD) scores. Classification and Regression Trees (CART) and Bayes Net with 10 fold cross validation determined classification accuracy as 76% and 78.667% respectively.

Kabinsingha et al [12] considered movie ratings. Data mining was applied to movie classification. Movies are rated into PG, PG-13 and R in the prototype. The 240 prototype movies from IMDb were used. Data was divided into testing and training set with four fold cross validation. Among various movie attributes like actors, actress, directors, budget, genre and producers, total number of selected attributes was 8 depending on movie genres and words used in movies corresponding to a decision used by most film rating organizations. The prototype was based on the Weka used decision tree (J48). Experiments achieved 80% to 88% precision for all tested rating.

Nasukawa and Yi [13] illustrated an approach to extract sentiments from sentences that contain opinions for specific subjects from a document, instead of classifying the whole document into positive or negative. Authors first identified sentiment expressions in different texts, the polarity and strength of the expressions, and whether the expressions indicate a positive or negative opinion towards a subject. Author used a particular subject of interest and manually defined a sentiment lexicon for identification. The prototype system achieved high precision of 75% to 95% correctness depending on the data. The test was performed on different web pages and news articles.

In this research work a new approach based on sequence labelling for opinion mining is proposed wherein POS tags and opinion tags are used for generating the predictive model.

The rest of the paper is organized as follows; Section II discusses about opinion mining using sequence labeling. Section III presents Conditional Random Fields. Section IV describes experimental results. Finally, Section V gives conclusion and scope for future work.

2 Opinion Mining using Sequence Labeling

Sequence labeling is the simplest subclass of structured prediction problems. In sequence labeling, the most likely one among all the possible label sequences is predicted for a given input. Although sequence labeling is the simplest subclass, a lot of real-world tasks are modeled as problems of this simplest subclass. Many models have been proposed for sequence labeling tasks, such as Hidden Markov Models (HMM), Conditional Random Fields (CRF), Max-Margin Markov Networks and others. These models have been applied to lots of practical tasks in natural language processing (NLP), bioinformatics, speech recognition, and so on. And they have shown great success in recent years. In real-world tasks, it is often needed to cascade multiple predictions. A cascade of predictions here means the situation in which some of predictions are made based upon the results of other predictions.

In opinion mining using sequence labeling approach, conditional random fields are used for learning the prediction model. Conditional random fields is a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for structured prediction. This can be particularly important for opinion mining on product reviews. CRF is used to encode known relationships between reviews opinion and construct consistent interpretation of the reviews. In this approach, conditional random fields predict the sequences of labels for a given input sequences. Here, the reviews are considered as input sequences and POS tags and opinion tags are used as output labels. CST (Center for SprogTeknologi) online tagger is used for performing POS tagging and opinion tagging is done manually. The essentials tasks of POS tagging and Opinion Tagging are described below.

2.1 POS Tagging

Part-of-Speech (POS) tagging is the process of assigning part-of-speech tags to words in a review. A part-of-speech tag is a grammatical category such as verbs, nouns, adjectives, adverbs, and so on. POS tagging is necessary to determine the opinion words. It can be done manually or with the help of POS tagger. In this research work, collected reviews are given as an input to the POS tagger that tags all the words in the review. CST's tagger is used to tag the words in the reviews. CST's tagger is an online tool for POS tagging. CST's tagger is one of the first and most widely used English pos tagger, employs rule-based algorithms. In CST's tagger the following processes are used for tagging the reviews.

Segmentation: Each review is segmented into lines.

Tokenisation: CST's tokeniser segments each line into words, numbers and punctuations.

POS-Tagging: The POS-tagger automatically assigns word class information to each word in a review, whether it is a noun, a verb, etc.

Example of POS tagging is given below.

Input Review: ***the kingston pen drive speed is good***

Tagged Output: ***the*/DT *kingston*/JJ *pen*/NN *drive*/NN *speed/*NN *is*/VBZ good/JJ**

Here, each word gets a tag of POS such as NN (noun word), JJ (adjective word) etc.

2.2 Opinion Tagging

Here for each review opinion tags are assigned. The goal is to extract product entities from reviews which also include the opinion polarities. Opinion tags consist of product polarities and its entities. The product entities are divided into four categories components, functions, features and opinions.

In opinion tagging, there are three types of tags to define each word: entity tag, position tag and opinion tag. The category name of a product entity is used as the entity tag. For a word which is not an entity, tag ’B’ is used to represent it. An entity can be a single word or a phrase. For the phrase entity, position is assigned to each word in the phrase. Any word of a phrase has three possible positions: the beginning of the phrase, the middle of the phrase and the end of phrase. The tags ’Feature-B’, ’Feature-M’ and ’Feature-E’ are used to indicate the three positions respectively. For ”opinion” entities, tags ’P’ and ’N’ are used to represent positive opinion and negative opinion respectively. These tags are called opinion tags. Thus with all of above defined tags, opinion tagging is performed for the reviews in order to prepare training dataset.

Opinion tagging for a sample review is given below:

The(B) speed(Featue-B) is(B) good(Opinion-P) and(B) its(B) ease(Feature-B) of(Feature-M) use (Feature-E) is(B) satisfying(Opinion-P).

In this review, ’speed’ and ’ease of use’ are both features of the pen drive and ’ease of use’ is a phrase, so ’Feature-B’, ’Feature-M’ and ’Feature-E’ are added to specify the position of each word in the phrase. ’Good’ is a positive opinion expressed on the feature ’speed’ and so the tag ’Opinion-P’ is assigned. For all other words which do not belong to any categories of entity, the tag ’B’ are assigned. In this manner, the task of opining tagging is done manually in order to transform the opinion mining task as sequence labeling task.

The modeling of opinion classification includes tagging the reviews, training the CRF model and evaluating the model based on the test data. The various stages of the proposed implementation are described in section 4.

3 Conditional Random Fields

Conditional Random Fields (CRF) are a probabilistic framework for labelling and segmenting sequential data, based on conditional probability. CRFs define conditional probability distributions p(Y/X) of label sequences Y given observations sequences X. Conditional models are used to label a novel observation sequence x* by selecting the label sequence y* that maximizes the conditional probability p(y*|x*). For the input sequence x = x1.. xn and the label sequence y = y1 .. yn, a CRF is specified by local feature vector f and a corresponding weight vector λ. Each local feature f is either a state feature s(yi, x, i) of the label at position i or a transition feature t (yi-1,yi,x,i) of the observation sequence x and the labels at positions i and i −1 in the label sequence. The probability of a particular label sequence y given observation sequence x is the normalized product of potential functions, of the form

$$e^{\sum_j \lambda_j t_j (y_{i-1}, y_i, x, i) + \sum_k \mu_k s_k (y_i, x, i)}$$

Typically, features depend on the inputs around the given position and may also depend on global properties of the input. The CRF's global feature vector for input sequence x and label sequence y is given by

$$F(y, x) = \sum_{i=1}^{n} f_j(y_{i-1}, y_i, x_i, i)$$

where i ranges over input positions where each $f_j(y_i\text{-}1,y_i,x,i)$ is either a state function $s(y_i\text{-}1,y_i,x,i)$ or a transition function $t(y_i\text{-}1,y_i,x,i)$. Any positive conditional distribution p(Y/X) that obeys the Markov property

$$p(Y_i / \{Y_j\}_{j \neq i,} X) = p(Y_i / Y_{i-1}, Y_{i-1}, X)$$

can be written as the probability of a label sequence y, given an observation sequence x.

$$P(y|x,\lambda) = \left(\frac{1}{Z(x)}\right) e^{\sum \lambda_j F_j (y, x)}$$

where Z(X) is a normalization factor and is given by

$$Z_\lambda (X) = \sum_y \exp(\lambda F(y, x))$$

CRF training is performed by maximizing the log-likelihood L(λ) of a given training set T = { (x_i, y_i) } i=1,..n. For a CRF, the log likelihood is given by

$$L(\lambda) = \sum_k \log P_\lambda (y_k / x_k)$$

$$L(\lambda) = \sum_k [\lambda F(y_k, x_k) - \log Z_\lambda (x_k)]$$

The maximum likelihood parameters λ are computed using an iterative technique such as iterative scaling or gradient-based methods.

4 Experiments and Results

Two independent experiments, one based on sequence labelling and the other based on classification approach, have been carried out for implementing opinion mining. Tweets about product reviews are collected manually and the training datasets are developed. In classification based opinion mining, word features are used for preparing the training dataset whereas in sequence labelling approach, POS tags and opinion tags are used. Classification algorithms such as Naïve Bayes, Maximum Entropy, and Decision Tree are employed to build classification based opinion prediction models.

4.1 Data Collection

The tweets or reviews about Kingston pen drive are collected from amazon.com. Both positive and negative reviews are collected separately by crawling raw review data. The meaningful reviews are collected from the HTML pages of the product related web pages. A total of 85 positive reviews and 85 negative reviews have been collected and used for further processing.

4.2 Classification Based Opinion Mining

For this experiment, the reviews are tokenized and the white spaces are removed during preprocessing. A dictionary is created with all the words in the reviews. Dictionary is also known as associative arrays or hash table. All the words in the reviews are considered as features. This method of feature extraction is word feature method. Then finally the opinion prediction models are learnt using naive bayes, maximum entropy and decision tree through implementing algorithms in Python integrated with NLTK. The performance of the classifiers is evaluated and the results are analyzed. The results of classifiers in terms of predictive accuracy, precision, recall and F-score are shown in Table 1 and depicted in Figure1.

Table 1: Evaluation measures of classification

Classifier/ Evaulation Criteria	Precision	Recall	F-score	Accuracy (%)
NB	0.67	0.92	0.78	73
ME	1.0	0.58	0.73	79
DT	0.75	0.91	0.83	81

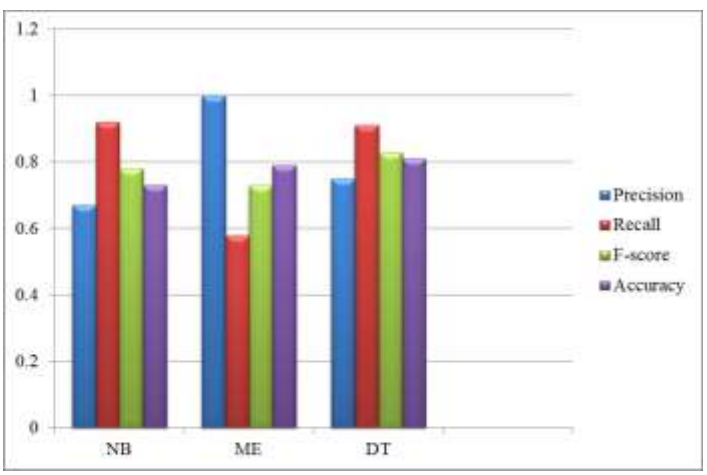

Figure 1: Comparison of classifiers

4.3 Sequence Labeling Based Opinion Mining

The second experiment is implemented using CRF++ under Linux operating system. For all the product reviews POS tagging and opinion tagging are done as described in section 2. In CRF++ the training file and the test file must be in column format. Training and test file consists of multiple tokens. A token consist of multiple columns. Each token must be represented in one line, with the column separated by white space. The training and test dataset are prepared in three column format as required by CRF++.

The first column contains words of reviews. The second column contains POS tags and the third column is opinion tags. Each review is separated by a blank line. The three column format of the training dataset is shown below.

Tweets	POS Tags	Opinion Tags
I	PRP	B
Like	IN	Opinion-P
That	IN	B
It	PRP	B
Is	VBZ	B
Small	JJ	Feature-B
And	CC	B
Compact	JJ	Feature-B
.	.	B

The model is created in CRF++ by using the template. The template file helps to extract the features from the reviews, POS tags and opinion tags. Based on the training file, the template creates a model by extracting features. The training dataset are given as an input to generate a model. In order to evaluate the CRF based opinion prediction model, the test dataset is prepared in three column format similar to training dataset. The CRF++ generates opinion tags for the reviews in test dataset and the predicted output will be displayed in fouth column. The evaluation measures of sequence labeling based opinion prediction model are shown in Table 2. The result of CRF++ is shown in Figure2.

Table 2: Results of CRF

Classifier/ Evaluation Criteria	Accuracy (%)	Precision	Recall	F-Score
CRF	**0.890**	**0.896**	**0.886**	**0.891**

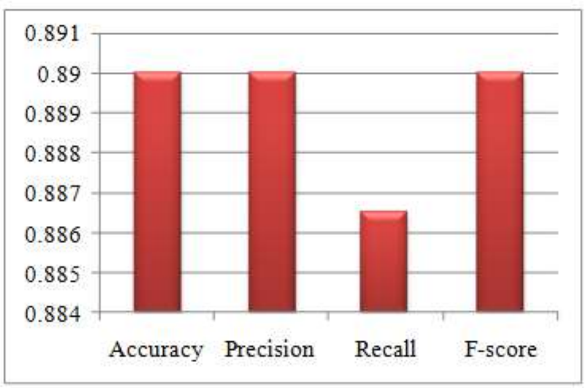

Figure2. Evaluation measures of CRF

4.4 Comparative Analysis

The comparative analysis of two experiments has been carried out and the comparative results indicate that the performance of the model based on conditional random field is better when compared to naïve bayes, maximum entropy and decision tree. Comparative results are summarized in Table 3 and depicted in Figure3.

Table 3: Comparative Results of sequence labeling and classification approaches

Classifier/ Accuracy	NB	ME	DT	CRF
Accuracy (%)	73	79	81	89
Precision	0.67	1.0	0.75	0.89
Recall	0.92	0.58	0.91	0.88
F-score	0.78	0.73	0.83	0.89

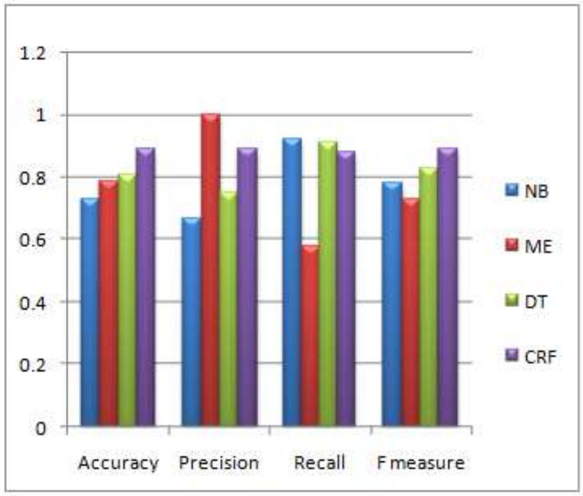

Figure3: Comparison of sequence labeling and classification approaches

From the comparative analysis, it is observed that the sequence labeling model show better accuracy and efficient in terms of precision, recall and F-score. Predictive accuracy plays vital role in real time predictions. For opinion prediction task, the comparative results shows that the sequence labeling based opinion mining approach is more supportive for predicting the review as positive or negative when compared to the classification based approach.

5 Conclusion and Future Work

This paper describes the application of sequence labeling approach to opinion prediction task and compares the results with common classification approach. The sequence labeling approach for opinion classification is implemented in CRF++ for conditional random fields. The implementation of classification based opinion mining is carried out in python language with NLTK library. Various classifiers naïve bayes, maximum entropy and decision tree have been used for building the classifiers. The performances are evaluated based on accuracy, precision, recall and F measure. It is concluded that the sequence labeling approach out performs well when compared to normal classification approach for opinion mining. In future work opinion of the product can be extended to also predict the neutral opinion of the reviews.

REFERENCES

[1] Jiawei Han, MichelineKamber, *Data Mining: Concepts and Techniques*. second edition.

[2] Mitchell T, *Machine learning*. Ed McGraw-Hill International edition.

[3] Dave, K., Lawrence, S., Pennock, *D.M.: Mining the peanut gallery: opinion extraction and semantic classification of product reviews*. In: 12th International Conference on World Wide Web, 2002, pp. 519–528.

[4] Eibe Frank, Ian H. Witten, *DataMining – Practical Machine Learning Tools and Techniques*. 2005.

[5] E. Brill. *Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging*. Comput. Linguist., 1995, 21:p.543–565.

[6] Xia, Y. Q., Xu, R. F., Wong, K. F., and Zheng, F. *The unified collocation framework for opinion mining*. Proc. IEEE Int. Conf. on Machine Learning and Cybernetics, 2007.

[7] A. Harb, M. Planti, G. Dray, M. Roche, Fran, o. Trousset, and P. Poncelet, *Web opinion mining: how to extract opinions from blogs?*. 2008.

[8] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, *Lexicon-based methods for sentiment analysis*. Comput.Linguist., 2011, vol.37, pp. 267-307.

[9] Ting-Chun Peng; Chia-Chun Shih, *Using Chinese part-of-speech patterns for sentiment phrase identification and opinion extraction in user generated reviews.* Digital Information Management.

[10] Turney, P., and Littman, M. L., *Unsupervised learning of semantic orientation.* Hundred-billion-word corpus, 2002.

[11] Valarmathi, B., and Palanisamy, V. *Opinion Mining Classification Using Key Word Summarization Based on Singular Value Decomposition*. International Journal on computer science and Engineering, 2011, Vol.3, No.1, pp.212-215.

[12] S.Kabinsingha,S., Chindasorn,C., and Chantrapornchai.A, *Movie Rating Approach and Application Based on Data Mining*. International Journal of Engineering and Innovative Technology, 2012, Vol. 2, No.1, pp.77-83.

[13] Tetsuya Nasukawa, Jeonghee Yi, *Sentiment Analysis: Capturing Favorability Using Natural Language* Processing, K-CAP'03.

[14] Bing Liu, *Sentiment Analysis and Opinion Mining.* Morgan & Claypool Publishers, 2012.

[15] Lee L. Pang, B and S. Vaithyanathan. *Thumbs up? Sentiment classification using machine learning techniques.* In EMNLP, 2002, pp 79–86.

[16] John, L., Andrew, M., Fernando.P, *Conditional random fields: probabilistic models for segmenting and labeling sequence data.* In: International Conference on Machine Learning, 2001, pp. 282–289.

[17] Luole Qi and Li Chen, *A Linear-Chain CRF-Based Learning Approach for Web Opinion Mining.* WISE 2010, LNCS 6488, 2010, pp. 128–141.

[18] http://www.amazon.com/Kingston-Digital-8GB DataTravelerDrive/dp/B004TS1J18/ref=sr_1_3?ie=UTF8&qid=1408526920&sr=8-3&keywords=kingston+pendrive

[19] NL Processor– Text Analysis Toolkit 2000. http://www.infogistics.com/textanalysis.html

[20] Scaffidi, C., Bierhoff, K., Chang, E., Felker, M., Ng, H., Jin, C.: *Red Opal: product feature scoring from reviews.* In: 8th ACM Conference on Electronic Commerce, 2007, pp. 182–191.

[21] Wilson, T., Hoffmann, P., Soma sundaran, S., Kessler, J., Wiebe, J., Choi, Y., Cardie, C., Riloff, E., Patwardhan, S.: *OpinionFinder: a system for subjectivity analysis.* In: HLT/EMNLP on Interactive Demonstrations, 2005, pp. 34–35.

[22] Arun K. Pujari, *Data Mining Techniques.* Universities of Press, 2010.