International Journal of Engineering & Technology, 7 (3.3) (2018) 87-89

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Sentiment Analysis for Product Recommendation Using Random Forest

Gayatri Khanvilkar[1]*, Prof. Deepali Vora[2]

[1] *Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India.*
[2] *Department of Information Technology, Vidyalankar Institute of Technology, Mumbai, India.*
**Corresponding author E-mail: khanvilkarg7@gmail.com*

Abstract

Analysis of sentiments is to analyze the natural language and to find the emotions, express by the human beings. The idea behind sentiment analysis is to determine polarity of textual opinion given by person. Sentiment Analysis is useful in product recommendations. Based on the reviews given by the user; the products can be recommended to another user. Major product websites are using sentiment analysis to understand the popularity and problems with the product. Sentiment analysis mainly formulated as two class classification problem, positive and negative. Sentiment analysis using ordinal classification gives more clear idea about sentiments. The proposed system determines polarity of reviews given by users, using ordinal classification. The system will give polarity using machine learning algorithms SVM and Random Forest. The achieved polarity will be used to provide recommendation to users.

Keywords: *Natural Language Processing; Ordinal classification; product recommendation; Random Forest; Sentiment Analysis;*

1. Introduction

Sentimental Analysis is nothing but the task of Natural Language Processing. It observes the attitude of customer behind the comments. Sentiment analysis is a method of identifying sentiments in text. [1] Sentiment analysis is computational methodology of extracting sentiments from text, speech or dataset. It can classify emotions, attitude, opinion and subjective impression into polarity. [2] Researchers and decision makers understand the approach of the people using consumer sentiment analysis and can make decisions accordingly. Business analysis application can be developed by using this technique. [3]

Social network development and popularity is increasing day by day. A growing number of users prefer to order online products and prefer to share their experiences on social networks. Searching for appropriate product online is a difficult task. Recommender system can help users by providing suggestions.

Recommender system creates recommendation list. There are three Recommendation system approaches, content based, collaborative and hybrid approach. The content-based approach considers the information of an item and the user's profile and the recommendation of items is based on the user's preferences. The collaborative-based approach analyses the user behaviour and preferences and find the same preferences among people. It is well known that collaborative-based techniques are normally more accurate than content-based techniques. The hybrid approach combines both methods.

User's feedback is important tool in recommendation system. The recommendation systems gives suggestion based on the user history and on the user's profile, but nowadays, the recommendations are starting to explore making more suggestions based on sentiment analysis. [4] [5]

2. Purpose of study

The primary goal of this study is to provide recommendation list based on sentiment analysis. Classification of sentiments in scale of -5 to+5 using machine learning algorithms and ordinal classification approach. Providing recommendation system which will offer personalized recommendation experience to uses; based on sentiment scale and user information.

3. Work Done in Sentiment Analysis and Recommendation System

Sentiment analysis in NLP (Natural Language Processing) has become a major area. There can be three approaches to perform sentiment analysis, the corpus-based approach using machine learning, the lexicon-based approach using a word dictionary and a hybrid-based method which combines both approaches. In machine learning approach SVM, KNN, and Naive Bayes are popular algorithms; where Random Forest has different characteristics, which can be useful for sentiment analysis. There are different inferences related to sentiment analysis are mention in table 1. A recommendation system identifies and provides recommended content or digital items for users. There are three Recommendation system approaches, content based, collaborative and hybrid approach. Nowadays, recommendation system also focuses on use of sentiment analysis to improve recommendation result. Related work in recommendation system based on sentiment analysis is mentioned in table 2.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

4. Findings from Study

Nowadays, the recommendation systems are starting to explore making more suggestions based on sentiment analysis. Sentiment analysis using ordinal classification helps to get clear idea about sentiments of writer. Ordinal classification is nothing but the ratings which have a natural order.

5. Proposed System

The proposed system predicts polarity of opinions given by user and gives recommendation list. The system will do Sentiment Analysis using Machine Learning Algorithms like Random Forest. The proposed system will work as shown in block diagram. The step wise flow of system given in fig. 1.

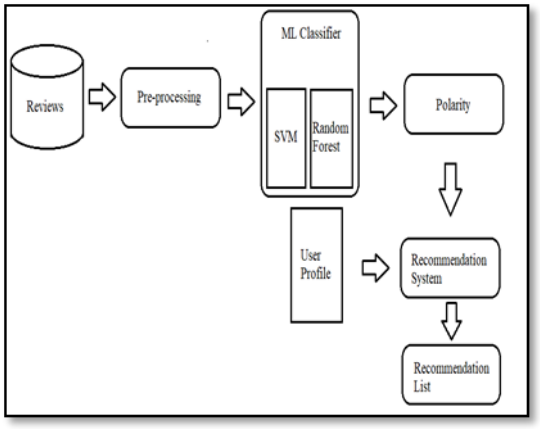


Fig 1: Block diagram of Proposed System

Methodology:

Step 1: Language Standardization

The collected data is Multilingual data. So, for further operations need to translate that data to standard language i.e. English.

Step 2: Preprocessing

Process of cleaning and preparing data as input to classifier is known as preprocessing. In this many subtasks are included like tokenization, stop word and punctuation removal, streaming etc.

Step 3: Apply Machine Learning algorithm

Apply Machine Learning algorithms and perform classification on dataset. In that use Random Forest algorithm and SVM.

Step 4: Polarity

From applying Random Forest and SVM on dataset, proposed system gives the polarity of sentiments.

Step 5: Recommendation system

Consider Sentiment analysis and user profile information for recommendation of product.

6. Conclusion

People are now a day more attracted towards Social media and online shopping. People share their thoughts, opinion on social networking websites. While shopping online, customer check reviews given by other users. These are main sources of people's Sentiments.

Sentiment analysis is very important in making business decisions. Using sentiment analysis decision maker get the customer's need and weak points of product and accordingly take decision for improvement. Sentiment Analysis can be useful in product recommendations. Based on the reviews given by the user on product websites as well as on social networking sites and user's information; the products can be recommended to user.

Ordinal classification can help to get better accuracy about people's opinion on any subject or product. SVM is popular for its prediction and classification accuracy where random forest is effective for not only accuracy but also for robustness. In this system SVM and Random Forest machine learning algorithm will help to improve Sentiment analysis for Product Recommendation using Multiclass classification.

References

[1] Bhavitha, B. K., Anisha P. Rodrigues, and Niranjan N. Chiplunkar. "Comparative study of machine learning techniques in sentimental analysis." *Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference on*. IEEE, 2017.

[2] Hegde, Yashaswini, and S. K. Padma. "Sentiment Analysis Using Random Forest Ensemble for Mobile Product Reviews in Kannada." *Advance Computing Conference (IACC), 2017 IEEE 7th International*. IEEE, 2017.

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

[4] Rosa, Renata L., Demsteneso Z. Rodriguez, and Graca Bressan. "Music recommendation system based on user's sentiments extracted from social networks." *IEEE Transactions on Consumer Electronics* 61.3 (2015): 359-367.

[5] Zheng, Xiaoyao, et al. "A tourism destination recommender system using users' sentiment and temporal dynamics." *Journal of Intelligent Information Systems* (2018): 1-22.

[6] Parmar, Hitesh, Sanjay Bhanderi, and Glory Shah. "Sentiment Mining of Movie Reviews using Random Forest with Tuned Hyperparameters." (2014).

[7] Kuzey, Cemil, Ali Uyar, and Dursun Delen. "An Investigation of the Factors Influencing Cost System Functionality Using Decision Trees, Support Vector Machines and Logistic Regression." (2018).

[8] Rosenthal, Sara, Noura Farra, and Preslav Nakov. "SemEval-2017 task 4: Sentiment analysis in Twitter." Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017). 2017.

[9] Kethavath, Shivaraju. Classification of Sentiment Analysis on Tweets using Machine Learning Techniques. Diss. 2015.

[10] Soni, Rishabh, and K. James Mathai. "Improved Twitter Sentiment Prediction through Cluster-then-Predict Model." arXiv preprint arXiv:1509.02437 (2015).

Table 1: Work done in Sentiment Analysis

Sr no.	Paper	Algorithm	Key findings	Inferences
1	Comparative study of machine learning techniques in sentimental analysis [1]	Random Forest, SVM	Random Forest Classifier shows the result with greater accuracy and performance.	RF requires high processing power and training time
2	An ensemble sentiment classification system of twitter data for airline services analysis. [2]	Naive Bayes, SVM, Bayesian Network, C4.5 Decision Tree and Random Forest	In sentiment classification, features can be unigrams, bigrams, trigrams and more. Accuracy Evaluation Based on F-measure-Recall, Precision, F-measure, Error Rate.	Only the texts of the tweets are considered and other information like the user's information, retweet and other factors are also potentially useful.
3	Sentiment Analysis Using Random Forest Ensemble for Mobile Product Reviews in Kannada [3]	Lexicon based approach, Random forest	Accuracy of sentiment analysis is depending on preprocessing and sentiment extraction. Also, accuracy	Working on large corpus is not possible because of limited number of reviews available in Kannada.

Sr no.	Paper	Algorithm	Key findings	Inferences
			of classifier is depending on feature selection and efficiency of classification algorithm.	
4	Sentiment Mining of Movie Reviews using Random Forest with Tuned Hyperparameters [6]	Random Forest, SVM, Naïve Bayes and Maximum Entropy	Random Forest classifier provides two types of randomness, first is with respect to data and second is with respect to features. Random Forest is considered to be an accurate and robust classifier.	It is very time-consuming and complicated because it is necessary to pay special attention to the tuning of hyperparameters in random jungle because it must be tuned manually
5	An Investigation of the Factors Influencing Cost System Functionality Using Decision Trees, Support Vector Machines and Logistic Regression." (2018) [7]	C5, CART, CHIAD, SVM	reasons for DT popularity which include (1) intuitiveness, (2) expressiveness, (3) transparency, (4) efficiency, (5) robustness, (6) accuracy, and (7) deploy ability.	data was collected by using questionnaire survey; thus, it is subjective which reflects the perceptions of the respondents.
6	SemEval-2017 task 4: Sentiment analysis in Twitter. [8]	CNN and LSTM, Maximum Entropy, Logistic Regression, Random Forest, Naïve Bayes classifier, and Conditional Random Fields	A five point scale which is in line with product ratings occurring in the corporate world. Five-point scale is nothing but ordinal classification.	Irony and emotion detection are the challenging task in sentiment analysis which are not considered.
7	Classification of Sentiment Analysis on Tweets using Machine Learning Techniques [9]	Svm, Decision tree, Nave Bayes classifier, Multinomial Nave Bayes algorithm	POS Tagging is used to tokenize and Tag the words further more for naming elements.	can't discover spelling mistakes
8	Improved Twitter Sentiment Prediction through Cluster-then-Predict Model. [10]	CART, SVM, logistic Regression, K-means clustering and Random Forest	A hybrid mechanism- 'Cluster-then predict Model' improve accuracy of predicting twitter sentiment	Sentiment values for each tweet are employed by the workers on the basis of Amazon's Mechanical Turk (MTurk).

Table 2: Work done in Recommendation system based on Sentiment Analysis

Sr no.	Paper	Approach/model	Key findings	Inference
1	Music recommendation system based on user's sentiments extracted from social networks. [4]	Hybrid approach	User's profile and sentiment analysis are important factors in recommendation system.	Sentiment analysis classifies sentences with positive, neutral or negative intensities, and do not recognize emotion differences according to user profiles.
2	A tourism destination recommender system using users' sentiment and temporal dynamics. [5]	SVD++, HTF, SVD, TopicMD	Recommendation models which are based on text reviews and temporal influences.	To ensure the fairness of the experiment, only conducted experiments using optimal λ settings for different models.