

Design Approach for Opinion Mining In Hotel Review Using SVM with Particle Swarm Optimization (PSO)

Sameeksha Jain

Harshit Sanwal

jainsameeksha.1999@gmail.com

Abstract:Opinion mining has emerged as an active domain among the research fraternity because an enormous amount of heterogeneous user data is continuously increasing every day via www, viz., e-commerce websites, social networks, discussion forums, blogs etc. This paper presented opinion mining and summarization of hotel reviews on the web. For opinion classification of hotel reviews we used SVM with Particle swarm optimization (PSO) algorithms Intentions are expressed in a different way with different vocabulary, short forms, and jargon making the data massive and disorganized. The proposed approach is termed sentiment polarity that automatically prepares a sentiment dataset for training and testing to extract unbiased opinions of hotel services from reviews. A comparative analysis was established with compliment Naïve Bayes and Composite hypercubes on iterated random projections to discover a suitable SVM with Particle swarm optimization(PSO) for the classification component of the proposed approach.

Keywords:Text Mining, Opinion Mining, Particle swarm optimization (PSO), SVM(Support Vector machine).

I. INTRODUCTION

Opinions, speculations, emotions and evaluations typically reveal the states of different from class; they include opinionative narrow data expressed during a language which is compiled of subjective statement [2]. "Sentiment analysis task is to identify people's belief, assessments, thoughts, appraisals, sentiments, and feelings towards instances like services, outcomes In this paper, we have chosen textual data in the form of hotel reviews for sentiment analysis with opinion mining from customer perspectives. Sentiment analysis uses the techniques of natural language processing and computational linguistics to automate the classification of sentiments generated from reviews. Hotels provide satisfaction, security, comfort, luxury and lodging services for travellers and people on vacation. Mining hotel reviews is desirable to gain deeper knowledge of customer expectations and support effective management of customer relationships. It would enable the hotel managers to have a good understanding of customer needs,

Discover areas for further improvement and improve service quality. The hotel reviews are provided exclusively by customers who have made reservations at a particular hotel. Customers post feedback about hotels which include hygiene, quality of food, location, customer service quality and hospitality exhibited by hotel staff. Moreover, sentiment analysis of hotel reviews is crucial to understand hidden patterns generated by data that would help to effectively improve performance [1].

SVM with Particle swarm optimization (PSO) are a broadly exploit device for taxonomy. Multiple effectual carrying out present for fitting a two-class SVM model or replica. The consumer has to give values for the tuning parameters. The SVM was first recommended via Vapnik and has since charmed a higher degree of curiosity in the machine learning research Community. Multiple hot lessons have reported which the SVM with Particle swarm optimization(PSO) are normally are proficient of delivering higher recital in terms of taxonomy correctness than the other data taxonomy algorithms [1]. SVM with Particle swarm optimization(PSO) are is a supervised learning replica. This replica is related with a learning algorithm that analyzes the data and identifies the pattern for classification. SVMs are a group of respective approaches for supervised learning, suitable to both taxonomy and regression issues. A SVM with Particle swarm optimization(PSO) are selects an extrememargin hyper plane which lies in a transformed input space and splitting the case classes, while maximizing the example to the near unreservedly divided instance. Parameters of the solution hyper plane are derived from a quadratic programming optimization issue [2]. In this paper, we present opinion mining of hotel reviews based on machine learning approach and SentiWordNet [10] based approach. We also present sentence extraction based opinion summarization of hotel reviews. Section II contains related work, in section III we present our proposed approach section IV contains experiments performed and results obtained. Section V contains conclusion.

II. OPINION MINING

Opinion is the view of a person representing their sentiments, beliefs or judgments in regard to a matter of importance in a particular context and is normally considered to be subjective in nature. Studies show that

more than facts, opinions of stakeholders greatly influence decision making of individuals as well as communities like governments and organizations. Opinion mining and sentiment analysis, the terms that are used interchangeably these days is a field of text data mining that involves extraction of opinions from evaluative texts and classification of the opinion's polarity as being positive or negative based on the orientation of the text results following the computational treatment of opinions expressed towards the key features [7]. Since opinions are expressed in human language, Natural Language Processing (NLP) techniques are mostly employed in conjunction with KDD methods for various stages of opinion mining like opinionated statement detection, feature identification, opinion extraction, polarity determination and opinion summarization. From among the lexicon based approaches and machine learning approaches, supervised machine learning techniques based on algorithms like Support Vector Machine (SVM), Naïve Bayes (NB), K Nearest Neighbor (KNN) and Maximum Entropy etc. that uses large number of labeled training data are commonly employed to determine polarity for the purpose of classification [8]. The survey paper [7] aptly explains SVM, NB, NN and KNN classifiers as follows: SVM classifier works best for classifying sparse text data by defining rectilinear partitions in the data set and divides the set into different classes. The best partition plane is determined by the maximum normal distance between the data sets. The NB classifier is the most commonly used text mining classifier which uses Bayes theorem to calculate the possibility of the given label belong to a particular feature, $p(l/f)$ using the below formula $P(l/f) = (P(l) * P(f/l))/P(f)$ (1) Where $P(l)$ is the possibility of occurrence of a label in the dataset, $P(f/l)$ is the possibility that a given feature belongs to a particular label. $P(f)$ is the occurrence of a particular feature in the data set. If the features such as $f_1, f_2, f_3 \dots f_n$ are not dependent on one another then equation (1) becomes $P(l/f) = (P(l) * P(f_1/l) * P(f_2/l) \dots P(f_n/l))/P(f)$ Neural Network classifier employs multiple layer of neurons as a medium of classification where each neuron takes the word frequencies of the dataset as input. They are also associated with a weight for calculation of its input function. The output of each layer of neuron is back propagated to its other layers as training mechanism. The classifier predicts purely based on the input set, the weight and the trained neurons. KNN classifier employs an indexing mechanism for the training data sets. To classify a document it calculates the similarity of the document with the training set index and uses the k-nearest by measuring the similarity by functions such as Euclidean distance. Final results of OM heavily depend on the preprocessing or preparing the data before classification, representation of the text suitable for classification and the classifier used. Main tasks involved in data preprocessing are tokenizing – separating the

sentences into words, removal of stop words prepositions, and pro-nouns, etc. that does not give any additional meaning to the documents, stemming – converting various grammatical forms of words into root word, and generating n grams [9]. References [10] and [11] point out that identification of the appropriate stop words to be removed has an impact on the quality of final classification. Feature based opinion mining is yet another dimension that has been analyzed in the works of [12][13][14] Opinion mining is being used now in retail industry for product reviews and recommender systems, service industry like education, health and tourism, governmental sectors – public opinion on policies, taxes, candidates and entertainment sector such as movie reviews etc. Opinion mining has been used to evaluate and classify student feedback from SMS as discussed in [15]. The author has developed three models, the base model with the necessary operators for classification, the second level model for data preprocessing and the last for performing sentiment analysis involving reading text resources, parsing SMS texts and categorizing text containing students' feedback.

III. RELATED WORK

The opinion mining is large field of mining of the reviews depend on the textual data and now a day there are vast amount of data for mining the opinions and reviews, there are some techniques that are previously applied in the hybrid methods with use of the SVM with the PSO and ACO.

Arti, Dubey et al [1] In this paper illustrates the views of 'Random Forest' that has primary result on comprehensive accuracy of the interpretation or analysis. This proposed methodology has an accuracy of 81.69% for classification. So the comparison between totally different techniques and proposed method shows that proposed techniques is preferable in essential analysis measurable attributes of correctly classifying out of all examples (tweets), specificity, F-score and Area Under Curve.

Rajput, V. S et al. [2] The proposed paper has given a comparative study of naïve bayes and SVM on the opinions of the reviewers of the stock market. No system has been created for sentiment analysis in the share market. Thus, new field is chosen and worked upon and its result can helps the user to take better decisions in the field of stock market.

Zvarevashe, K., et al [3] The proposed framework is termed sentiment polarity that automatically prepares a sentiment dataset for training and testing to extract unbiased opinions of hotel services from reviews. A comparative analysis was established with Naïve Bayes multinomial, sequential minimal optimization, compliment Naïve Bayes and Composite hypercubes on

iterated random projections to discover a suitable machine learning algorithm for the classification component of the framework.

Cambria et al. [4] disputed the interchange of these concepts by classifying opinion mining as polarity detection and sentiment analysis as focusing on emotion recognition. The opinion mining system only needs to understand polarity that can be positive, negative or neutral sentiments depending on the nature of sentences expressed in a review

M.S. Akhtar et al[5]The process of detecting polarity is strongly linked to analysing sentiments on a particular subject. Most researches on sentiment analysis are focused on descriptive data.

Karthikayini, T., et al[6] In this paper, we presented a complete system to analyze huge dataset generated from Amazon Product reviews. First, the datasets were cleaned using data processing techniques. The fine-tuned data is then used as an input to identify the sentiments. The Datumbbox and NLTK APIs were called in the code to classify the sentiments. In our study, we found that very limited data has been polarized which cannot be sufficient for making any business decisions. The existing machine learning algorithms implemented in both the tools had many research flaws such as language variability, word intensity and poor handling of negation statements

IV. PROPOSED METHODOLOGY

Opinion mining is identification of user's opinion about particular topic from reviews. It is classification of a review text as positive or negative opinion polarity review. Opinion summarization is a process of finding most important aspects about topic and related opinion sentences from reviews to represent a summary. Our proposed architecture performs opinion mining and summarization of hotel reviews is shown in Figure 1. The proposed system consist of three modules review text retrieval, classification and summarization. Reviews about hotel are retrieved from review websites such as www.tripadvisor.com by web crawling techniques. Review text is classified as positive or negative review using machine learning classifiers or SentiWordNet based algorithm. Classified review text is pre-processed and sentence scores are calculated. Finally most informative and context relevant sentences represented in summary. The conceptual view of the intuition model begins with the feedback collection. Customers respond to questionnaires concerning their feelings about services received from the selected hotels. This can be done in a number of ways, for example opening a web portal through which customers can drop comments. The next step will be to label the comments based on intuition. This will be done by human agents who simply read the comments and assign labels based on perceptions. Once

data are transformed to a desire format, the next step will be to convert the labelled text to feature vectors through the use of filters. This will make it easier to implement a classification algorithm for training and testing of data. The next step involves the selection of an appropriate classification algorithm whilst the last step is the training and testing of the selected algorithm on dataset and capturing of results The research reported in this paper was done using the sentimental polarity based model. the proposed model begins with the elicitation of opinions which is the step skipped because we used the raw dataset. Customers respond to questionnaires concerning services received from the selected hotels through an appropriate user interface. The next step will be to label the comments based on sentiment polarity score using a sentiment polarity algorithm. The score obtained will determine whether a comment is positive, negative or neutral. Once the data are transformed, the next step is to convert the labelled text to feature vectors through the use of filters.

The next step involves the selection of a suitable classification algorithm. The last step is the training and testing of the selected classification algorithm and capturing of results. The distinguishing property of labelling is that automatic, it does not involve human intervention and it is quite consistent in labelling sentiments. However, the proposed relies heavily on human intervention to label sentiments which sometimes may not be consistent and the labelling process is intrinsically laborious and time demanding. We have proposed a recommendation system for the opinions given by the user. In general, the customers are not aware about the ongoing things. They give their reviews as positive and negative. But, few cases are of neutral. The response over few things can't be determined and they are marked as neutral. Thus, for the processing of the views, SVM with Particle swarm optimization(PSO) are with Particle swarm optimization(PSO) are is implemented. SVM is supervised learning based algorithm which optimizes the results. Steps in the proposed algorithm:

1. The data is pre-processed and cleaned by removing the unnecessary and redundant terms.
2. Opinions are checked by the Disparity of the sentiments. The opinions are further processed by passing them into SVM with Particle swarm optimization (PSO) are.
3. The features that are passed to SVM with Particle swarm optimization(PSO) are used for training and testing the data. 66% of the information is trained. 33% of the information is then tested.
4. The final results are recommended to the users and misperception matrix is generated.
5. The effectiveness factors like precision, kappa are calculated are results are displayed.

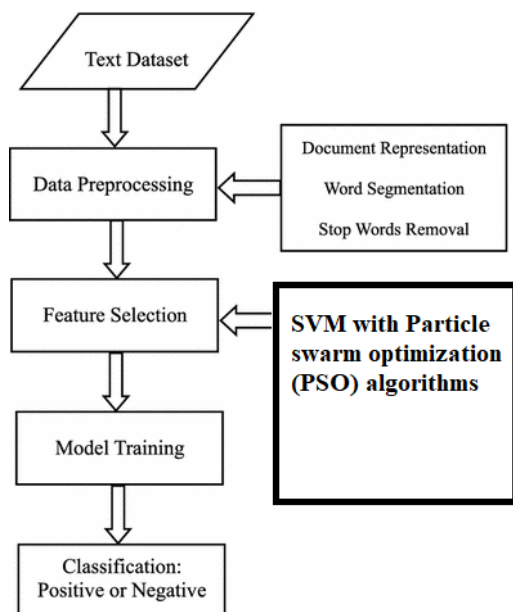


Figure 1 : proposed system flow chart

Step 1. Obtain the reviews for hotel review site

Step 2. Parse the json data

Step 3. Extract one review text from the json data object

Step 4. Process the extracted data by removing duplicates, punctuations, numbers, special characters and extra blank spaces

Step 5. To determine the polarity using natural language processing tool kit invoke the API. The API requires input of type string. Pass the processed review text to the API

Step 6. Extract the sentiment / polarity from the API response using JSON parser and store it. Extracted data contains have the polarity value, positive or negative or neutral

Step 7. For each review text in the json data object obtained on step 2, repeat the process from step 3 to Step 6

Step 8. Plot a pie chart using the output data

Feature selection the major objective of feature selection (FS) is to conclude a negligible feature subset from a issues domain while retentive an appropriately high accuracy in representing the unique characteristics. In several real world issues FS is a must cause to the luxuriance of loud, irrelevant or misleading characteristics. For example, through eliminating these elements, learning from data methods can advantage too. The helpfulness of a characteristic or characteristic subset is laid down through both its redundancy and relevancy. A characteristic is said to be occasional if it is prognostic of the verdict feature(s), otherwise it is incompatible. Classification rules that are generated in learning phase

are stored for the performance evaluation of the dataset created. In this phase, the testing set generated in data splitting module is used as input to evaluate the performance. The outcome of this phase is additional promoted to next phase classifier performance evaluator. We labelled the hotel features in the dataset using a sentiment polarity software written in Python with the TextBlob which is a library for processing textual data. The scores obtained from the sentiment polarity were then used to automatically label the data. After performing all the necessary processing steps, including labelling and filtering, the dataset was split into two subsets to create testing and training datasets. We used four classification algorithms which are SVM with Particle swarm optimization(PSO), Sequential minimal optimization (SMO), Compliment Naïve Bayes (CNB) and Composite hypercubes on iterated random projections (CHIRP) to train and test the dataset. the comparative results obtained after experimentation. Optimization by using review data of sentiment analysis of ethics in the social media surfing from 200 review data in Indonesian text consisting of 100 positive reviews and 100 negative reviews. The resulting model gets the accuracy result of first test done by using a Naive Bayes namely 70.70% accuracy values, then tested again with the second test by combining SVM with Particle Swarm Optimization so that it is obtained the best values with 90.00% accuracy. it can be concluded that a sentiment analysis test using a Naive Bayes will get better result if we combine SVM(Support Vector machine) with Particle Swarm Optimization so that we can solve the problem of sentiment analysis by choosing best hotels to be more accurate.

V. CONCLUSION

Opinion mining is a text classification problem where review text document is classified into classes as positive or negative opinion review. Machine learning approach and resource based approach for opinion classification of review text can be used. The proposed algorithm worked on the views/ opinions of the reviewers on the hotel review. hotel review in an important field where views of the users matter. The views of the experts affect a lot to the traders who want to enter into the market. we used term frequency and relevance scoring method to represent most informative sentences in summary. This classified and summarized review information will assist users in decision making about hotels. The unsupervised and supervised learning depend methods help to find the results in a better way. The comparative study is done to justify the results. Further, optimization methods can be applied in sequence to get improved results.

Reference

- [1]. Arti, Dubey, K. P., & Agrawal, S. (2019). An Opinion Mining for Indian Premier League

- Using Machine Learning Techniques. 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU). doi:10.1109/iot-siu.2019.8777472.
- [2]. Rajput, V. S., &Dubey, S. M. (2016). Stock market sentiment analysis based on machine learning. 2016 2nd International Conference on Next Generation Computing Technologies (NGCT). doi:10.1109/ngct.2016.7877468
- [3]. Zvarevashe, K., &Olugbara, O. O. (2018). A framework for sentiment analysis with opinion mining of hotel reviews. 2018 Conference on Information Communications Technology and Society (ICTAS). doi:10.1109/ictas.2018.8368746
- [4]. E. Cambria, B. Schuller, Y. Xia and C. Havasi, "New avenues in opinion mining and sentiment analysis". IEEE Intelligent Systems, vol. 28, no. 2, pp. 15-21, 2013.
- [5]. M.S. Akhtar, D.k Gupta, A. Ekbal and P. Bhattacharyya. "Feature selection and ensemble construction: a two-step method for aspect based sentiment analysis." Knowledge-Based Systems, vol. 125, pp. 116-135, 2017.
- [6]. Karthikayini, T., &Srinath, N. K. (2017). Comparative Polarity Analysis on Amazon Product Reviews Using Existing Machine Learning Algorithms. 2017 2nd International Conference on Computational Systems and Information Technology for Sustainable Solution (CSITSS). doi:10.1109/csitss.2017.8447660.
- [7]. Raut, V. B., &Londhe, D. D. (2014). Opinion Mining and Summarization of Hotel Reviews. 2014 International Conference on Computational Intelligence and Communication Networks. doi:10.1109/cicn.2014.126.
- [8]. Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, GenChi Lu, and Emery Jou "Movie Rating and Review Summarization in Mobile Environment", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, Vol. 42, No. 3, May 2012, pp.397-406.
- [9]. Elena Lloret, Alexandra Balahur, José M. Gómez, Andrés Montoyo, Manuel Palomar, "Towards a unified framework for opinion retrieval, mining and summarization" Journal of Intelligent Information Systems Springer 2012, pp.711- 747.
- [10]. Alexandra Balahur, MijailKabadjov, Josef Steinberger, Ralf Steinberger, Andrés Montoyo, "Challenges and solutions in the opinion summarization", Journal of Intelligent Information Systems Springer 2012, pp.375-398.
- [11]. Abd El-Jawad, M. H., Hodhod, R., & Omar, Y. M. K. (2018). Sentiment Analysis of Social Media Networks Using Machine Learning. 2018 14th International Computer Engineering Conference (ICENCO). doi:10.1109/icenco.2018.8636124.
- [12]. Dhanalakshmi, V., Bino, D., &Saravanan, A. M. (2016). Opinion mining from student feedback data using supervised learning algorithms. 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC). doi:10.1109/icbdsc.2016.7460390.
- [13]. Trisha Patel, "Sentiment Analysis of Parents Feedback for Educational Institutes", International Journal of Innovative and Emerging Research in Engineering, Volume 2, Issue 3, 2015.
- [14]. Padmapani and Tribhuvan, "A Peer Review of Feature Based Opinion Mining and Summarization", International Journal of Computer Science and Information Technologies, Vol. 5 (1), 247-250, 2014.
- [15]. Chee Kian Leong, "Mining sentiments in SMS texts for teaching evaluation", Journal of Expert Systems with Applications, Vol. 39, 2584– 2589, 2012. 16. Chien-wen Shen, "Learning in massive open online courses: Evidence from social media mining", Journal of Computers in Human Behavior, vol. 51 568–577, 2015.

Submitted By -

HarshitSanwal(1601DMBCS00937)

SameekshaJain(1601DMBCS00909)