Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 3 (2023)

Aspect Level Opinion Mining for Hotel Text Reviews using Traditional Machine Learning Approaches

Manisha Prasad[1*], Krishnandan Prasad[1]

[1]Department of Mathematics, Patliputra University, Patna, Bihar.
*Corresponding author(s). E-mail(s): manishaprasad0909@gmail.com;

Abstract

The customers intend to write their opinion about the hotel they visit in order to provide suggestions, complains, or appreciate the services. These opinions have been expressed in terms of text on the online platforms. The hotel industries need to process and analyze these opinions to improve their business in the right direction. The challenges such as computational efficiency to process a huge size data ore of opinions and ability to learn from the continuously generated data. This article proposes a machine learning based opinion mining method that captures the intention of the customer as well as the opinions with respect to the specific aspects of the hotel entities. The proposed method has a three stage pipeline that includes pre-processing of the textual opinions, feature extraction from the pre-processed opinions, and classification of the features values into its correct ratings and various sub-ratings. The experiments have been carried out on 4.5M reviews of HotelRec dataset. The results obtained through the proposed method has outperformed the existing baseline methods.

Keywords: Opinion Mining, Aspect level, Hotel reviews, Ensemble learning.

1 Introduction

The expansion of social media is serving as a supplement to the Internet as the most accessible and economical source of information. The online blogs on web- sites, reviews on different platforms, tweets, posts on various social media, and discussions are scanned to glean user opinions. Generally, these opinions are written in text using various languages. While examining a person's behaviour, or sentiments, it is crucial to consider their attitude, perspectives, feelings, and opinions. The study of these feelings about any object has been termed as opining mining [24]. The term object represents the target entity present in the textual content. An object can be a constituent of some attributes or a set of components. A text can be categorized into various opinion that serves different agenda to gain profit. In the hospitality industry, text reviews are one of the sources of information to get user experience. If any customer visits any hotel, then an online feedback form is provided to that user to write a detailed review about their experiences. The user is allowed to mention all the key aspects such as cleanliness, business service, staff behaviour, location, food, rooms, etc. The user is also allowed to give ratings corresponding to the text review as well as for each key aspect. The ratings provided by the different users are stored to calculate the average rating of the hotel. The average hotel rating helps new customers to find their best solution for a stay. Similarly, the reviews and ratings help hotel administration to improve their business.

Opinion mining automatically classifies the customer text review into positive, negative, neutral, etc. types of sentiments. An online text review is unstructured in nature and due to natural language content, it requires lots of effort to understand the opinion of each user about the hotel. The opinion of each user has been assigned a value known as a rating. Hotel ratings play an important role in judging the hotels for any individual. It influences the user's decision and signifies the credibility of the hotel. The hotel rating can depend on various factors like reviews, location, average cost of a person, cuisines and many other factors. The travelers or commuters (users) face struggles to find hotels which meet their expectations of the stay and hotel rating could be one solution as their deciding factor. It has been observed that the user may forget to provide the ratings for the hotel but may provide the review for the same. This leads to another challenge to predict the ratings based on the review provided by the customer. Since every customer may have a different level of knowledge of speaking and writing the languages and it may impact the understanding of the reviews as well. These challenges of natural language require a method to understand each and every review provided by the different customers. It gives the motivation of the present work to predict the ratings corresponding to the reviews provided by the customer for the hotels. The goal is to make an adaptable model that analyses the customer reviews and predicts the hotel rating using opinion mining. The rating of the hotel is the average sum of all the ratings provided by the customers for the respective hotel. The better ratings and positive reviews play an important role in increasing the revenue of the hotels, whereas the negative reviews or low rating gives the feedback to the hotel to improve their services.

This work has proposed an ensemble learning based rating prediction method that provides the rating based on the text review provided by the customer. The text review has gone through the text cleansing first and then it has been passed through the natural language processing (NLP) steps. The term- frequency (TF), inverse document frequency (IDF), and bag-of-words (BOW) techniques have been used to develop the feature vector of the corresponding hotel text reviews. This feature vector has been provided to the different machine learning classifiers to predict the rating of the hotel on HotelRec dataset [3]. The majority voting technique of ensemble learning has been used for the prediction of the ratings. The proposed work also predicts the aspect level ratings for each customer review. The aspect level hotel ratings help hotels to know opinions about particular aspects of the hotel. The contribution of the proposed work is as follows:

1. Each hotel review has been cleaned using NLP techniques and the feature vectors have been obtained using TF-IDF.
2. The ensemble learning based proposed method uses the 6 different classifiers to predict the hotel rating based on the review provided by the customer.
3. The proposed method predicts hotel rating and various sub-rating for each customer text review.

The rest of the paper has been organized as follows: The related works have been discussed in Section 2. The Section 3 describes the dataset and Section 4 elaborates the proposed method. Section 5 provides the result and discussion of the experimental work and Section 6 concludes this work.

2 Related Works

Nowadays, opinion mining has become well acknowledged among researchers, companies, organizations and governments. The existing methods of opinion mining generally solve either core tasks or sub-tasks [22]. The opinion mining on core task solves document level, sentence level and aspect level opinion classification problems [24]. On the other hand, the opinion mining on sub-task solves multimodel and multi domain opinion classification problem. Traditional methods of opinion mining [4] used supervised learning techniques to develop a classifier for Twitter sentiment. Researchers developed a feature combination strategy in addition to the frequently utilized text features to enhance classification performance [14]. Moreover, tweet messages contain a few remote supervised characteristics. Many research employ these enormous quantities of noisy labelled tweets as training data or as an auxiliary source for classifying Twitter sentiment [21]. For training word embedding, which is a combination of noisy-labelled resource and knowledge base, our method combines distant supervised information and lexical knowledge, in contrast to earlier works.

A strategy to transforming text data into low-dimension emotional space has been put out by Luo et al. [16]. They have small, annotated words with distinct, obvious meanings. When giving words weight based on emotional tags, they used two alternative methods. The messages are divided into various groups depending on the determined overall emotional weight of all emotional tags. A multi-view sentiment analysis dataset including a collection of image- text pairs with manual annotation that was gathered from Twitter has been proposed by Niu et al. [19]. Their sentiment analysis methodology can be divided into two categories: lexicon- based and statistic learning. While different machine learning approaches are used with specific textual features in statistic learning, lexicon-based analysis considers a set of opinion words or phrases that have a specified sentiment score. Using the IMDb dataset, Tripathy et al. [23] performed sentiment categorization using the n-gram machine learning technique. For classification using n-gram approaches such as uni- gram, bigram, trigram, unigram + bigram, bigram + trigram, they used Naive Bayes, maximum entropy, support vector machine (SVM), and stochastic gradient descent. They discover that SVM with unigram+ bigram has produced the best results of all these methods.

Several multi-label classifications on sentiment classification have been proposed by Liu and Chen [15]. The authors claim that issue transformation and algorithm adaptation are the two primary methods by which the multi-label classification process completes the task. When a problem is transformed, it is split up into several single-label problems. The trained classifier first produces predictions at a single label and then translates them to many labels during testing. The system learns from these converted single-label data during training. Chinese comments can be categorized using word2vec and SVM, according to a proposal made by Zhang et al. [25]. They mostly used a two-part strategy. In the first section, they took the word2vec tool into consideration to group comparable features in order to capture the semantic aspects in the chosen domain. Finally, to create the training data, the lexicon-based and POS-based feature selection methodologies were used. Internet travel review platforms like TripAdvisor produce a significant amount of text-based data in the form of online travel review data [13]. Researchers and practitioners can accurately comprehend visitors' travel preferences and demands with the aid of online review text data [1, 6]. Potential tourists' decisions are significantly influenced by the views presented in user-generated comments [5]. Big data's properties have made the knowledge extraction process more challenging. For large data applications, the issue of how to convert data into worthwhile information has become critical [17]. Prior studies on online reviews have mostly ignored the language of online reviews and instead concentrated on the quantitative ratings offered on the website [7]. Ratings are unable to provide any information on the particular product features that customers like or dislike, as this is usually stated in the review text [10]. Also, the vast volume of review data offered on travel web review sites overwhelms many customers. Similar queries have been posed by researchers in other domains.

A city's rating score is insufficient to provide accurate information, according to Ali et al. [12], despite the fact that urban traffic congestion is rapidly expanding. Nonetheless, comments or tweets may assist visitors and traffic management in comprehending all aspects of the city. In order to assist users in identifying the primary information and emotions present in the review text, an efficient procedure must be established.

Consumer decision-making is thought to be significantly influenced by human emotions and emotional thinking. As a result, mining the implications of online travel review texts using sentiment analysis is a successful technique. Dictionary- based methods [1], Machine learning methods, Deep learning methods [2, 20], and hybrids of the aforementioned methods [8, 11] are the several types of text sentiment analysis techniques. Dictionary-based systems, according to Alaei et al. [1], depend on the use of sentiment dictionaries and rule sets. Their article makes the argument that as these methods are unable to keep up with the big data era's rapid increase in data volume, more potent automated methods for sentiment analysis must be developed. To reach their full potential, deep learning techniques typically need a lot of training data, which typically calls for pricey class labelling [9]. The most popular machine learning techniques in the context of sentiment analysis for the tourism industry are SVM and Naive Bayes. SVM and Naive Bayes require fewer class annotations to train the model than neural networks do [1]. The majority of studies on the topic have demonstrated [18] that compared to other machine learning techniques, SVM-based sentiment analysis of text yields better results. Kirilenko et al. [12] assessed whether several automatic classifier types are appropriate for common applications in the tourism, hotel, and marketing research settings by comparing them to people and automatic text sentiment analysis classifiers. According to the article, computer classifiers perform worse than humans on challenging and noisy datasets. So, it may be inferred that in order to enable the analysis of particular data, the current sentiment analysis technology needs to be enhanced.

3 Dataset Description

Many people post hotel reviews on internet directories every day to express their thoughts on a variety of topics, including their overall experience, the service, and the location (e.g., Booking, TripAdvisor). TripAdvisor is the largest online travel site in the world with roughly 1.4 million hotels. As a result, we constructed our dataset HotelRec on hotel reviews found on TripAdvisor [3]. The present work has collected the HotelRec dataset from the public domain. The complete dataset has 50 million customer reviews. The present study has only considered 4.5 million customer reviews. The number of reviews for each rating categories are not equal. The data is biased towards two polarity and majority of the ratings (81.2%) are either 4 or 5. The distribution of the rating in the dataset are illustrated in Fig 1.

Each customer review has different number of words and it has been seen that it may vary from 50 – 1000 words. Since the number of words in each customer review is different so, a fixed length of feature vector has to be developed to solve the problem. The Fig. 2 shows the word length frequency distribution among the customer reviews in the dataset. It has been found that the majority of the customer reviews have length approx 200.

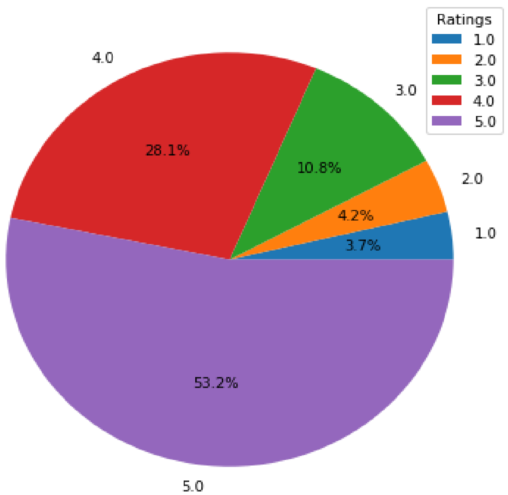

Fig. 1 Data distribution among various opinion values of the dataset.

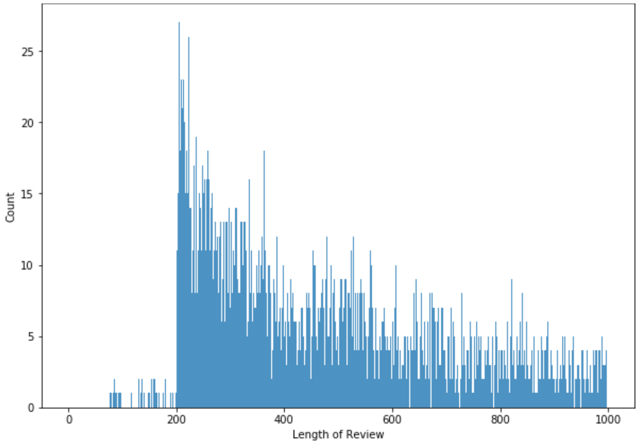

Fig. 2 Statistics of the length of each review in the dataset.

4 Methodology

This work proposes a novel method to predict the opinion of a customer about the various aspects of the hotel through the customer text review. Each textual content has been preprocessed as per the English NLP. Initially, each word and character present in any customer review has been converted into lower case to remove the redundancy. The special characters, hypertext markup language tags, and numbers has been removed using various pre-processing techniques. The lemmatization technique has been applied to each customer review to bring every word to its root word present in the English dictionary. All the unique words present in the entire dataset has been collected together to build a single vocabulary.

4.1 Feature extraction

The textual review has to be converted into numeric format so that the features can be extracted in order to find the opinion. In order to do so, the pre-processed customer review has gone through text embedding techniques. The text embedding techniques convert the entire customer review into numeric format. In the proposed study, TF-IDF word embedding technique has been used. The TF counts the frequency of each unique word in every customer review. The TF value for every word has been computed using Equation (1).

$$TF_{word} = \frac{count\ of\ a\ word}{Total\ number\ of\ words} \tag{1}$$

where, TF_{word} is a single value for every word. The IDF value of any word shows how frequent that word is present in a customer review. The IDF value of any word has been computed using Equation (2).

$$IDF_{word} = log \frac{Total\ number\ of\ reviews}{Total\ number\ of\ reviews\ having\ that\ word} \quad (2)$$

The TF-IDF value of any word is a product of TF_{word} and IDF_{word} values. Each word present in the customer review has been replaced by the TF-IDF value of that word.

4.2 Opinion Mining

Each customer text review has been converted into numeric format using TF- IDF word embedding technique. The machine learning techniques take each numeric format customer review and predicts the opinion corresponding to it. The proposed study developed majority voting-based ensemble learning technique to provide the overall opinion of the customer review and 8 different aspect level sub-opinions. The ensemble learning method uses multivariate linear regression, SVM, Decision tree, Random forest, Naive Bayes, and K-NN machine learning techniques to predict opinion of the customer and eight different aspect level opinion. A total of 9 predictions have been made for a single customer review.

5 Results and Discussion

The performance of the proposed opinion mining method has been evaluated on the dataset mentioned in Section 3. The dataset has been divided in two sets—train set and test set in the ratio of 7:3. The performance of the proposed method has been evaluated in terms of *precision*, *recall*, *accuracy*, and *F1-score* performance metrics. These metrics are explained as follows:

$$Precision = TP / TP + FP \quad (3)$$

$$Recall = TP / TP + FN \quad (4)$$

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (5)$$

$$\text{F1 score} = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \qquad (6)$$

The experiment results have been computed on various Machine learning classifiers used in the proposed ensemble learning technique. The performance of the opinion mining on HotelRec dataset with various classifiers are presented in Table 1.

Table 1 Comparison of the opinion mining performances with various machine learning classifiers

Classifier	Dataset	Precision	Recall	Accuracy	F1-measure
Multivariate Linear regression	HotelRec	70.11	69.31	60.1	69.70
SVM	HotelRec	87.09	86.17	81.3	86.62
Decision tree	HotelRec	76.53	75.13	68.2	75.82
Random forest	HotelRec	78.77	77.68	70.1	78.22
Naive Bayes	HotelRec	74.72	73.91	65.9	74.31
K-NN	HotelRec	77.64	76.23	69.3	76.93
Ensemble learning	HotelRec	89.71	88.12	82.7	88.90

5.1 Discussion

The existing opinion mining techniques use very few Machine learning classifiers, whereas the proposed study has used a majority voting scheme to predict the opinion from each customer review. The results from Table 1 show that the F1 score of ensemble learning technique has outperformed the existing various machine learning classifiers.

6 Conclusion and Future Work

This work presented a majority voting scheme based ensemble learning opinion mining method to predict the sentiment of each customer review from the HotelRec dataset. The customer review has been preprocessed using NLP techniques and the TF-IDF values of each word has been computed to extract features from the text review. The majority voting schema among various machine learning methods has been developed to retrieve a strong decision value from the proposed method. The results of the proposed method show the robust performance of the proposed method. The present work can be extended in the various domains of Opinion mining to retrieve various aspect level sentiment of the textual content. An attempt will be made in future to extract features from the trainable network to improve the performance of the opinion mining.

References

[1] Alaei, A. R., Becken, S., and Stantic, B. (2019). Sentiment analysis in tourism: capitalizing on big data. *Journal of travel research*, 58(2):175–191.

[2] An, H.-w. and Moon, N. (2022). Design of Recommendation system for tourist spot using sentiment analysis based on cnn-lstm. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–11.

[3] Antognini, D. and Faltings, B. (2020). Hotelrec: a novel very large-scale hotel recommendation dataset. *arXiv preprint arXiv:2002.06854*.

[4] Appel, O., Chiclana, F., Carter, J., and Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge- Based Systems*, 108:110–124.

[5] Asghar, M. Z., Kundi, F. M., Ahmad, S., Khan, A., and Khan, F. (2018). T-saf: Twitter sentiment analysis framework using a hybrid classification scheme. *Expert Systems*, 35(1):e12233.

[6] Fang, B., Ye, Q., Kucukusta, D., and Law, R. (2016). Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52:498–506.

[7] Guo, Y., Barnes, S. J., and Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism management*, 59:467–483.

[8] Handhika, T., Fahrurozi, A., Sari, I., Lestari, D. P., Zen, R. I. M., et al. (2019). Hybrid method for sentiment analysis using homogeneous ensemble

classifier. In *2019 2nd International Conference of Computer and Informatics Engineering (IC2IE)*, pages 232–236. IEEE.

[9] Hinterstoisser, S., Lepetit, V., Wohlhart, P., and Konolige, K. (2018). On pre-trained image features and synthetic images for deep learning. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0.

[10] Kauffmann, E., Peral, J., Gil, D., Ferr´andez, A., Sellers, R., and Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, 90:523–537.

[11] Kim, K., Park, O.-j., Yun, S., and Yun, H. (2017). What makes tourists feel negatively about tourism destinations? application of hybrid text mining methodology to smart destination management. *Technological Forecasting and Social Change*, 123:362–369.

[12] Kirilenko, A. P., Stepchenkova, S. O., Kim, H., and Li, X. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research*, 57(8):1012–1025.

[13] Li, J., Xu, L., Tang, L., Wang, S., and Li, L. (2018). Big data in tourism research: A literature review. *Tourism management*, 68:301–323.

[14] Liu, M., Zhang, L., Hu, H., Nie, L., and Dai, J. (2016). A classification model for semantic entailment recognition with feature combination. *Neurocomputing*, 208:127–135.

[15] Liu, S. M. and Chen, J.-H. (2015). A multi-label classification based approach for sentiment classification. *Expert Systems with Applications*, 42(3):1083–1093.

[16] Luo, B., Zeng, J., and Duan, J. (2016). Emotion space model for classifying opinions in stock message board. *Expert Systems with Applications*, 44:138–146.

[17] Mariani, M., Baggio, R., Fuchs, M., and H¨oepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality*

Management, 30(12):3514–3554.

[18] Neidhardt, J., R¨ummele, N., and Werthner, H. (2016). Can we predict your sentiments by listening to your peers? In *Information and Commu- nication Technologies in Tourism 2016: Proceedings of the International Conference in Bilbao, Spain, February 2-5, 2016*, pages 593–603. Springer.

[19] Niu, T., Zhu, S., Pang, L., and El Saddik, A. (2016). Sentiment analy- sis on multi-view social data. In *MultiMedia Modeling: 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016, Proceedings, Part II 22*, pages 15–27. Springer.

[20] Pan, D., Yuan, J., Li, L., and Sheng, D. (2019). Deep neural network-based classification model for sentiment analysis. In *2019 6th International Con- ference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, pages 1–4. IEEE.

[21] Poria, S., Cambria, E., and Gelbukh, A. (2015). Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pages 2539–2544.

[22] Ravi, K. and Ravi, V. (2015). A survey on opinion mining and senti- ment analysis: tasks, approaches and applications. *Knowledge-based systems*, 89:14–46.

[23] Tripathy, A., Agrawal, A., and Rath, S. K. (2016). Classification of sen- timent reviews using n-gram machine learning approach. *Expert Systems with Applications*, 57:117–126.

[24] Yadav, A. and Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6):4335–4385.

[25] Zhang, D., Xu, H., Su, Z., and Xu, Y. (2015). Chinese comments sen- timent classification based on word2vec and svmperf. *Expert Systems with Applications*, 42(4):1857–1863.