CS584: TripAdvisor Reviews Analysis and Prediction of Attribute values

Krina Uday Shah[1], Varun Reddy Doddipalli[1]
[1]Stevens Institute of Technology
kshah119@stevens.edu, vdoddipa@stevens.edu

Abstract

Sentiment analysis is essential for detecting and understanding customer emotions. Companies use these to understand the feelings of their customers and thus to improve customer services. In our project we aim to collect user reviews from Tripadvisor website. Tripadvisor offers online hotel bookings for transportation, accommodation, travel experiences and restaurants. We target those customers who have booked their hotel in NYC and extracted their reviews for our analysis. From the model trained, we try to obtain the estimated values for the rating attributes from the user text reviews which in turn would help the business partner of the hotel realize the quality of service that the hotel provides and which specific areas can be improved.

1 Introduction

Analysis of user comments and reviews can help businesses in understanding how their customers are feeling about their products and services, which in turn provides deep insights to major stakeholders in the business on how to improve specific areas of products and services.

TripAdvisor is a travel company that assists its customers in finding the best rates for their hotel stay as well as booking tickets for their trip. One of the services it offers is their comprehensive hotel booking suite which enables its users to not only view hotels based on location, cost, cleanliness, and various other factors but also review the stay of other travelers at those hotels. The users are prompted to write a text-based review of more than 200 characters and provide an overall rating as well as a rating for cleanliness, rooms, and location as part of their review. Users can read thousands of reviews left by other users for a specific hotel before making their choice. These reviews are not only useful for other users, but they provide several insights to major stakeholders for the hotels which might help them improve the quality of their services.

TripAdvisor sticks to five main ratings for a specific hotel, namely cleanliness, rooms, value or price, services and location of the hotel, along with an overall rating for the hotel. However, it is not necessary that the guests are always looking for these specific services in the hotel. Adversely, the review left by the user might include more details about services which they might be unhappy about, however the overall numerical rating does not provide any information regarding the details of those services.

For example, A guest might be satisfied with the cleanliness of the hotel, their room size as well as the location, but they might be extremely unhappy with other services such as food or value for money[4]. The guests might express these concerns in their text review and change the overall rating for the hotel, but this numerical rating does not provide enough information to the Hotel's management team to make changes or improve their services.

Our Project aims to predict the overall rating as well as individual categories, room rating, value rating, location rating, cleanliness rating and sleep quality rating based on the review text provided by the user. This will help to analyze the sentiment of each individual category. These categories not only help the user's narrow down their search for their perfect stay, but also helps the businesses to ascertain which services need to be improved in order to increase customer satisfaction, and bring in more business into their respective hotels[1].

2 Background

We reviewed the work detailed in the paper by Hsiu-Yuan Tsao and Ming-Yi Chen "The asymmetric effect of review valence on numerical rating"[5], where the authors have conducted a sentiment analysis via text mining, using self-developed computer programs to retrieve a data set from the TripAdvisor website. This study finds there is an asymmetric relationship between review valence or the verbal review text and numerical rating. The authors further find brand strength to have an important moderating role. For a stronger brand, negative review content will have a greater impact on numerical ratings than positive review content, while for a weaker brand, positive review content will have a greater impact on numerical ratings than negative review content.

Therefore, the overall rating that is provided to a hotel is not a reliable measure of services offered by a specific hotel branch or customer satisfaction. The authors mention that assuming verbal review text is symmetrically related to the numerical rating might be a false one, since brand image is a significant factor that customers consider while writing these reviews on TripAdvisor. Similarly, other factors or services offered by a specific hotel might not be considered while providing their independent overall rating to the hotel. The authors further conclude that marketers could adopt sentiment analysis via text mining of online reviews as a valid measure or predictor of consumer satisfaction or numerical ratings. Strong brands should direct more attention to negative reviews, because in such reviews the negative impact transcends the positive. In contrast, weak brands should aim to exploit as many positive reviews as possible to minimize the impact of any negative reviews.

We noted that part of the "Brand Image" of the Hotel is simply just one of the factors that might affect the Review valence and overall rating. Other factors would include the services offered by the specific Hotel Branch, such as the quality of food and dining services, gym and fitness services, staff politeness, etc. All of these keywords can be identified and a sentiment analysis would provide us with more insights as to whether the customer reviewing the hotel had a positive or negative experience on these specific factors. This might in turn help us to bridge the gap between the review valence and the overall rating provided by TripAdvisor.

We also reviewed the work by Kudakwashe Zvarevashe and Oludayo O. Olugbara "A framework for sentiment analysis with opinion mining of hotel reviews"[7]. Because people frequently express their thoughts in complex and sometimes difficult-to-understand ways, computerized text data labeling is tough. The labeling procedure is time-consuming, and mislabeled datasets frequently result in wrong choices. In the paper, authors provide a framework for sentiment analysis with opinion mining for hotel customer feedback in this research. The majority of hotel review datasets are unlabeled, which creates a lot of effort for researchers in terms of text data pre-processing. Furthermore, sentiment datasets are frequently domain sensitive and difficult to construct, as sentiments include feelings such as emotions, attitudes, and opinions that are frequently rich with idioms, onomatopoeias, homophones, phonemes, alliterations, and acronyms. The framework proposed suggests that sentiment polarity is a suggested system that automatically prepares a sentiment dataset for training and testing in order to extract unbiased hotel service judgments from reviews. To find an appropriate machine learning algorithm for the framework's classification component, a comparison analysis was established using Naive Bayes multinomial, sequential minimal optimization, complement Naive Bayes and Composite hypercubes on iterated random projections.

Furthermore, we reviewed the work by Jin Zheng and Limin Zheng "A hybrid bi-directional recurrent neural network attention based-model for text Classification"[6]. This paper discusses how convolution and recurrent neural networks are important for text classification but when it comes to multi-class text classification for fine-grained sentiment analysis it's still a challenge. The paper put forward BRCAN model, which combines bidirectional LSTM and CNN model with attention layer. It was concluded that BRCAN worked better than the state-of-the-art model.

3 Approach

Given the user review text, is it possible to predict not only the overall rating but also the rating of individual categories. Here our approach was to implement multi-class text classification using neural networks where given the user review text can we predict overall rating and various individual ratings. Additionally, we were also curious about the different categories/services in hotels that customers generally mention in their review for a Hotel. There are various categories many users would like to know about a particular hotel, such as its services, location, cleanliness, sleep quality, and its value. So, we are interested in what are the typical categories customers mention in their review and which are the categories that are not mentioned in the review.

For prediction, as mentioned in the paper we reviewed[6], we were trying to implement a multi-layer bidirectional recurrent convolution neural network. It was mentioned in the paper that it works better than the traditional classification models and thus, we were keen to implement it and view the results.

4 Methodology

A. Data Set:

Our source of data is : `http://www.cs.cmu.edu/~jiweil/html/hotel-review.html` The data was in JSON format, so it had to be converted in CSV format for the ease of use in our scenario and we extracted only those attributes that were necessary for our analysis. Our source variable is review text given by the user and our target variable was the overall rating, service rating, cleanliness rating, value rating, location rating and sleep rating. Following Figure 1 shows a snippet of our dataset.

	text	overall	service	cleanliness	value	location	sleep_quality
0	Stayed in a king suite for 11 nights and yes i...	5.0	5.0	5.0	5.0	5.0	5.0
1	On every visit to NYC, the Hotel Beacon is the...	5.0	5.0	5.0	5.0	5.0	5.0
2	This is a great property in Midtown. We two di...	4.0	4.0	5.0	4.0	5.0	4.0
3	The Andaz is a nice hotel in a central locatio...	4.0	5.0	5.0	5.0	5.0	5.0
4	I have stayed at each of the US Andaz properti...	4.0	4.0	5.0	3.0	5.0	5.0

Figure 1: Data Set

Column	Description
Text	The user describes his/her experience about their stay in a particular hotel
Overall	On the scale of 1-5, the overall rating provided by the user
Service	Rating about the service provided of the hotel on the scale of 1-5
Cleanliness	Rating about the cleanliness of the hotel on the scale of 1-5
Value	Rating about how does user felt about the price for a stay on scale of 1-5
Location	Rating about the location of the hotel on the scale of 1-5
Sleep Quality	Rating about the sleep quality on the scale of 1-5

B. Prepossessing of Data:

We had few data rows where the rating of individual categories was not mentioned. The value was 0. While building the model, we also consider the chances where the user might not provide the value of any individual rating and thus also consider 0 as a value. Initially, we expanded words such as “ain’t” as “am not”, “can’t” as “cannot” , etc. Then we passed our text to a series of pre-processing such as removal of punctuation, URL and numbers. Further some reviews were truncated based on the number of words used. In our use-case, we removed reviews less than 50 words and more that 1000 words. The removal of lower length reviews helps increase our model performance and removing the higher length reviews makes sure that our model size is in check without consuming lot of computing resource.

To speed up the process of tokenization and lemmatization of the text, we created multiple threads using multiprocessing libraries. Here, we use the SPACY[3] model to load English vocab.

Given that we had over 5,000,000 records to train, we cannot store the whole data at once in the memory, so we had to utilize the Data generators provided by Tensorflow [2]
Passed source data x-train and y-test as well as target data/labels y-train and y-test to our sequence generation function along with other parameters such as tokenizer(tokenized words), batch size(length of batch size) and text window(length of input text). We padded the input text so that each text length is of the same length. This function returns sequence output of batch tokens, overall rating value, service rating value, cleanliness rating value, value rating value, location rating value and sleep quality rating value.

C. Model Implementation :

Model 1 : Multi Layer Bi-directional Recurrent Convolution Neural Network was built for multi-class text classification. The model is a combination of two layers Bi-directional LSTM neural network followed by two layers of convolution neural network.

Bi-directional LSTM Neural Network is able to learn long-term dependency and as it is able to process the input in both forward and reverse direction, it will help in classification by using the future context. Additionally, CNN is able to extract local and deep features from the natural language.

Our model consists of one input layer, followed by an embedding layer. Embedding layer is used to transform word/text into vector format. It captures the morphological, syntactic, and semantic information of words by mapping words from a lexicon to a corresponding vector of actual values. The parameters passed in embedding layers where

```
0    stay mini family vacation include visit museum...
1    staff friendly , honest helpful . respond quic...
2    hotel right door reliant center .. walk door /...
3    person encounter check - desk rude . party boo...
4    Holiday Inn Reliant Park super place stay . ex...
5    stay Friday night business area , mile street ...
6    overall wrong hotel . service good room pretty...
7    property desparately need - . room 1 degree ac...
8    hotel locate inner loop ( central area ) Houst...
9    time Extended Stay America , pleased need . ac...
Name: processed_text, dtype: object
```

Figure 2: Preprocessed Data

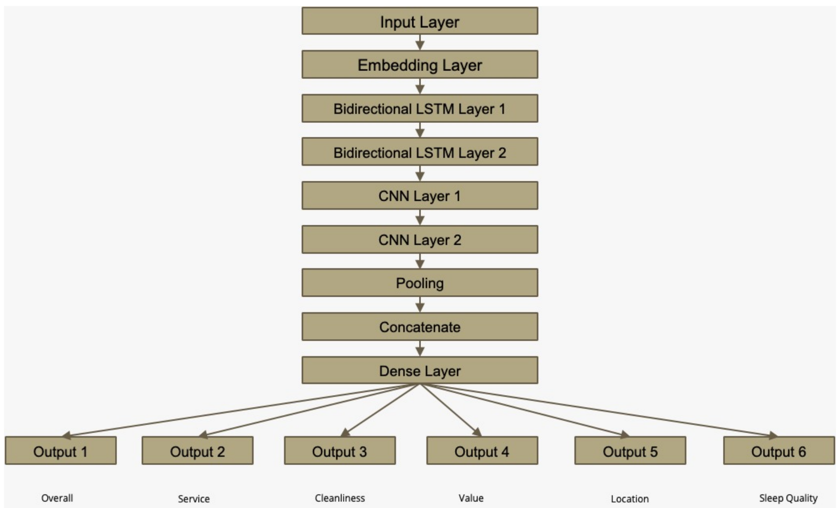

Figure 3: Model Flowchart

the source vocab size along with embedding dimension. After getting the embedding, the data was passed into two stacked layers of bi-direction LSTM model where return-state was set to False and return-sequece was set to True. As we need the output of all the hidden state of all the time steps we set return-sequence as True and as in any RNN model, it tends to capture too much of last words in the final vector and thus we do not need that and thus it was set to False. Further, CNN is used to capture the most influential n-grams of different semantic aspects from text. Thus, the output was then passed to two layers of 1 dimension convolution layer which was eventually concatenated. We used a global max pooling layer. The output activation function used was soft-max and six outputs were generated for each rating : overall rating, service rating, cleanliness rating, value rating, location rating and sleep quality rating.

Model 2:

We also develop the same multi-layer Bi-directional recurrent convolution neural network with an attention layer. Attention layer helps neural networks in memorizing large sequences of data.

The core idea of the attention layer is to make a direct connection to focus on a particular part of the source sequence. The basic architecture is the same as that of model 1 but we added a attention layer from LSTM layer to CNN layer. Refer Figure 6.

5 Results

Following are the results of Model 1 and Model 2 :

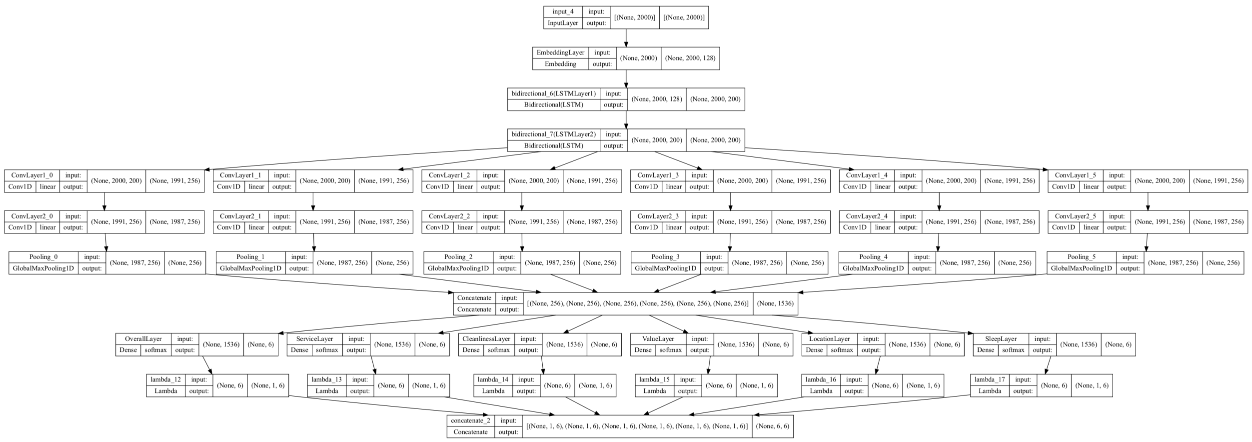

Figure 4: Model 1: Architecture without Attention

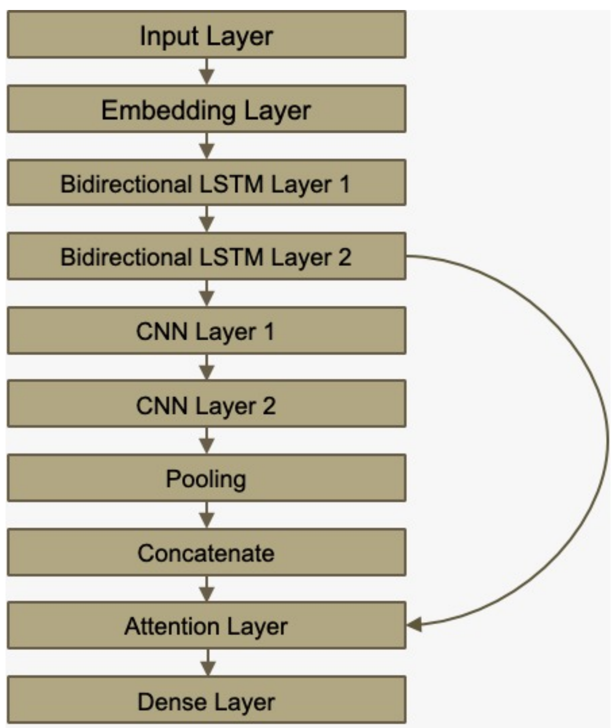

Figure 5: Model 2 flowchart

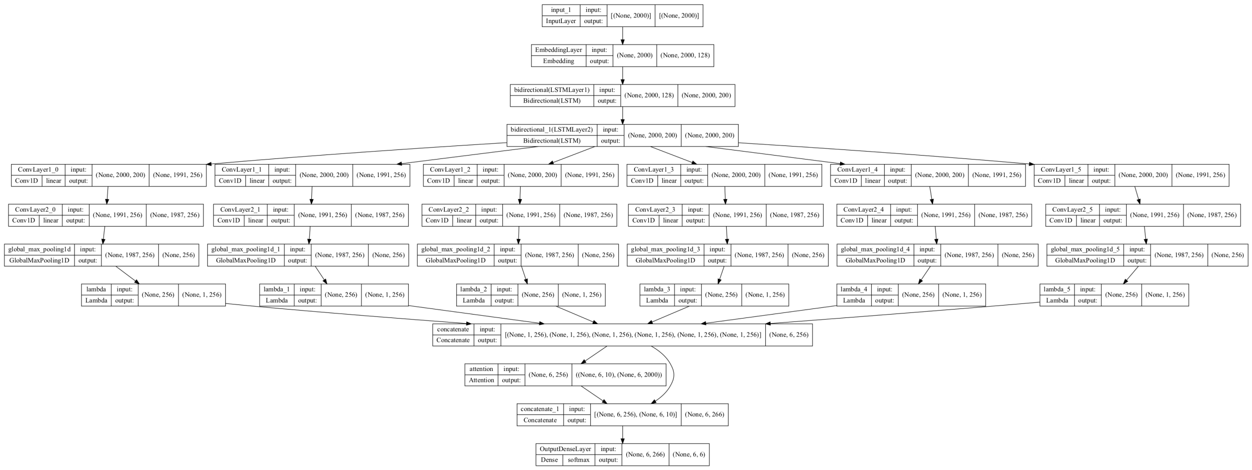

Figure 6: Model 2: Architecture with Attention

Model	Train Accuracy	Validation Accuracy
Model without attention layer	70.75	68.96
Model with attention layer	74.24	72.25

Also, to determine the most and least categories mentioned in the review about a particular hotel, we plotted a graph for each category where the rating is 0. As mentioned in data prepossessing, value 0 indicates a customer who has not given a rating for a particular individual category.

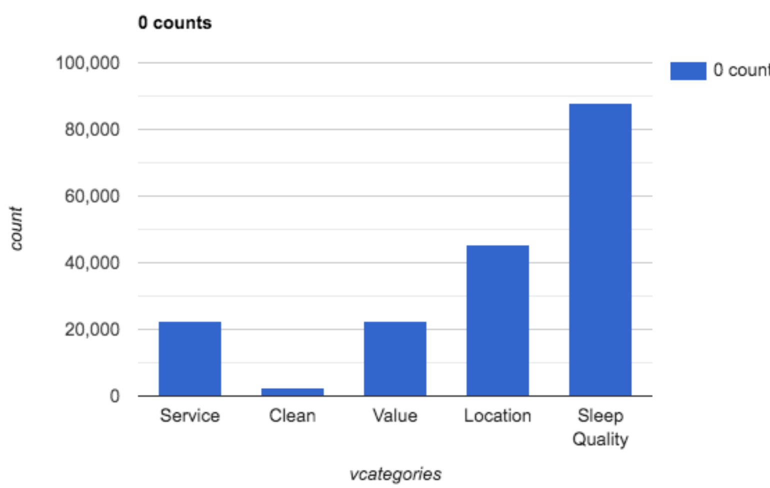

Figure 7: Zero count

6 Analysis Results

Comparing between different models we can see that the Model with attention layer outperforms the Vanilla model. The attention mechanism is used to train the weighted linear combinations of vectors so that more important features can get higher weights which results in better results.

Additionally, customers not mentioning the individual rating suggest that he/she might have not even mentioned the same in the review text. Looking at the zero count figure, we can see that most zero are given to sleep quality rating, and the least zero count were given to cleanliness rating. This suggests that People have hardly mentioned their sleep quality, this tells us that while looking for a hotel people never look for how good or bad the sleep quality is but give more attention to other categories such as its service, its value, its location and especially how clean the hotel is. As the least zero count is given to cleanliness rating, cleanliness of hotel is most common thing people mention in their review.

7 Conclusion and future work

So we can conclude from the the above study that there are various parameters that are not mentioned while the user provides reviews and using this model we were able to predict the Sentimental value for each of the said category such as Sleep, Cleanliness, etc based of the given review. This would not only help the users filter out the hotels based on the attribute, but also give a hotel owner a detailed analysis as in where to improve. Here in our scenario we have not included the attention value in our final demonstration. In the future, we can highlight the words that would contribute to output accuracy of each word. The accuracy might further be increased by using Transformers architecture models such as BERT but given the huge parameter size of the model, we were constrained by the system resources.

References

[1] Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics. *Tourism Management*, 52:498–506, 2016. ISSN 0261-5177. doi: https://doi.org/10.1016/j.tourman.2015.07.018. URL `https://www.sciencedirect.com/science/article/pii/S0261517715001715`.

[2] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

[3] Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.

[4] Nikolaos Korfiatis, Elena García-Bariocanal, and Salvador Sánchez-Alonso. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3):205–217, 2012. ISSN 1567-4223. doi: https://doi.org/10.1016/j.elerap.2011.10.003. URL `https://www.sciencedirect.com/science/article/pii/S1567422311000639`.

[5] Hsiu-Yuan Tsao, Ming-Yi Chen, Hao-Chiang Lin, and Yu-Chun Ma. The asymmetric effect of review valence on numerical rating: A viewpoint from a sentiment analysis of users of tripadvisor. *Online Information Review*, 43, 09 2018. doi: 10.1108/OIR-11-2017-0307.

[6] Jin Zheng and Limin Zheng. A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification. *IEEE Access*, 7:106673–106685, 2019. doi: 10.1109/ACCESS.2019.2932619.

[7] Kudakwashe Zvarevashe and Oludayo O. Olugbara. A framework for sentiment analysis with opinion mining of hotel reviews. In *2018 Conference on Information Communications Technology and Society (ICTAS)*, pages 1–4, 2018. doi: 10.1109/ICTAS.2018.8368746.