

CS 584-B Natural Language Processing

TripAdvisor reviews Analysis and Prediction of Attribute values

Krina Shah (10459200)
Varun Reddy (10474196)

Introduction and Background



- 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
- Enables customers to filter hotels based on different amenities as well as reviews provided by other customers
- Overall rating provided to a hotel is not a reliable measure of individual service quality such as service rating, cleanliness rating, value rating, location rating and sleep rating
- Understand the different services offered and features for specific Hotels is essential

Objective and Research Question



Objective is to analyze customer reviews to determine the following:

- What are the different categories/services in hotels that customers generally mention in their review for a Hotel?
- Based on the customer reviews can we point to the positive and negative features in the Hotel?
- Given the review, predicting rating for individual categories i.e. service rating, cleanliness rating, value rating, location rating and sleep quality rating.

Business Impact



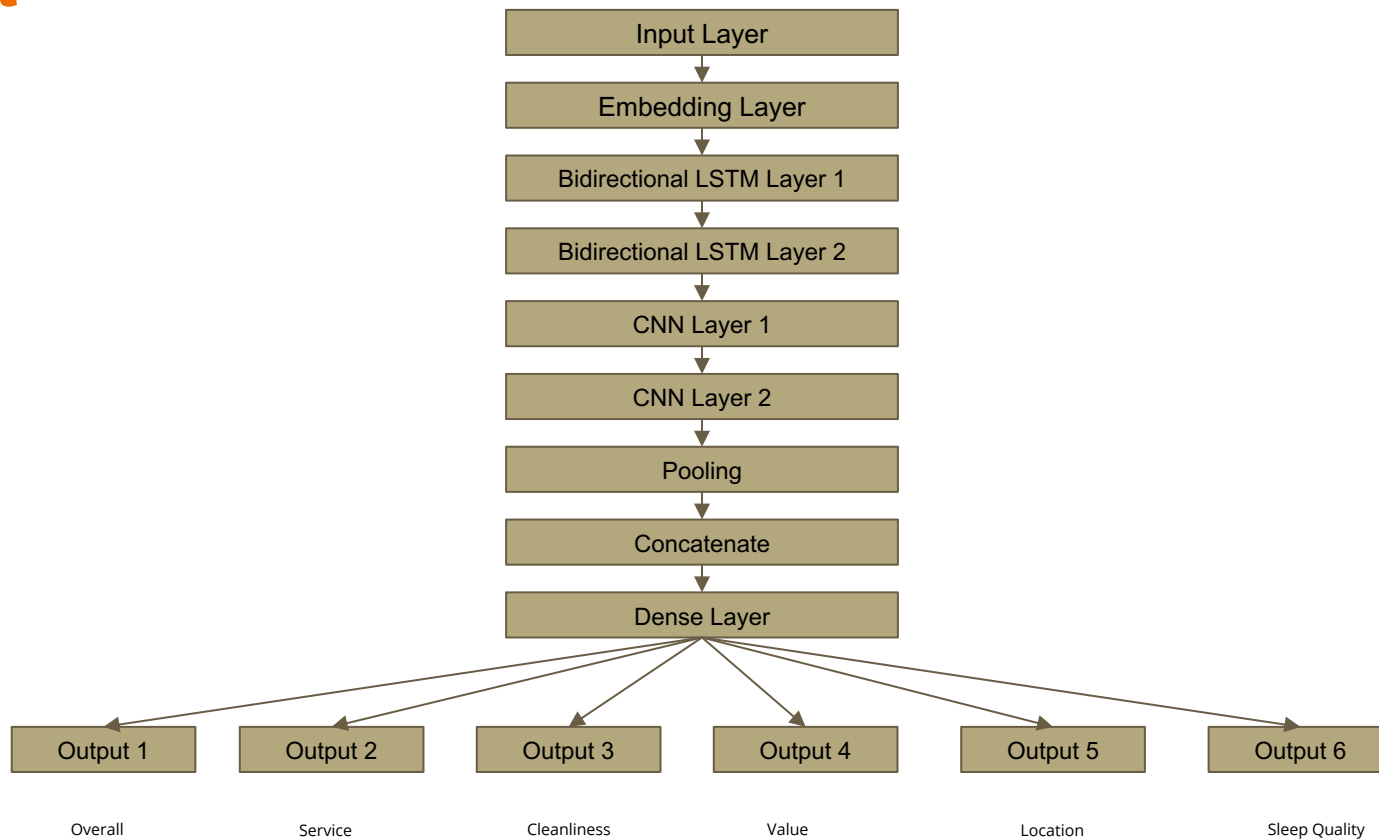
- This analysis can help Hotels to determine specific areas where they need improvements, especially the areas which are negatively mentioned in the reviews by customers and thereby improve the guest experience
- This in turn, can also help the business to retain existing customer base, gain new ones and get more attraction online through the positive reviews they can potentially gain after making the said improvements
- It can help bridge the gap between reviews left by customers and the overall ratings they have provided as sentiment analysis of text reviews will explain whether overall Hotel services have had a positive or negative impact on their stay

Data Set and Preprocessing

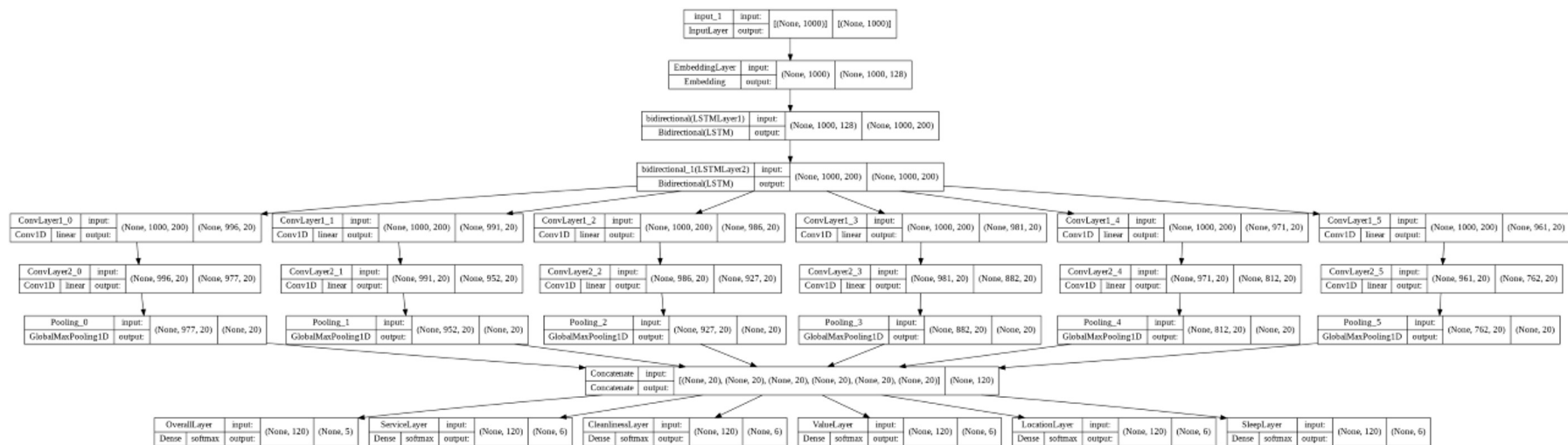
- Source of dataset : <http://www.cs.cmu.edu/~jiweil/html/hotel-review.html>
- Data was in JSON format, we converted in CSV format with necessary attributes use for our analysis
- Our source variable was the **review text** given by the user and our target variable was the overall rating, service rating, cleanliness rating, value rating, location rating and sleep rating
- The text was pass through series of preprocessing such as removal of punctuation, numbers, url, stopwords and then text was tokenize and lemmetized

	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

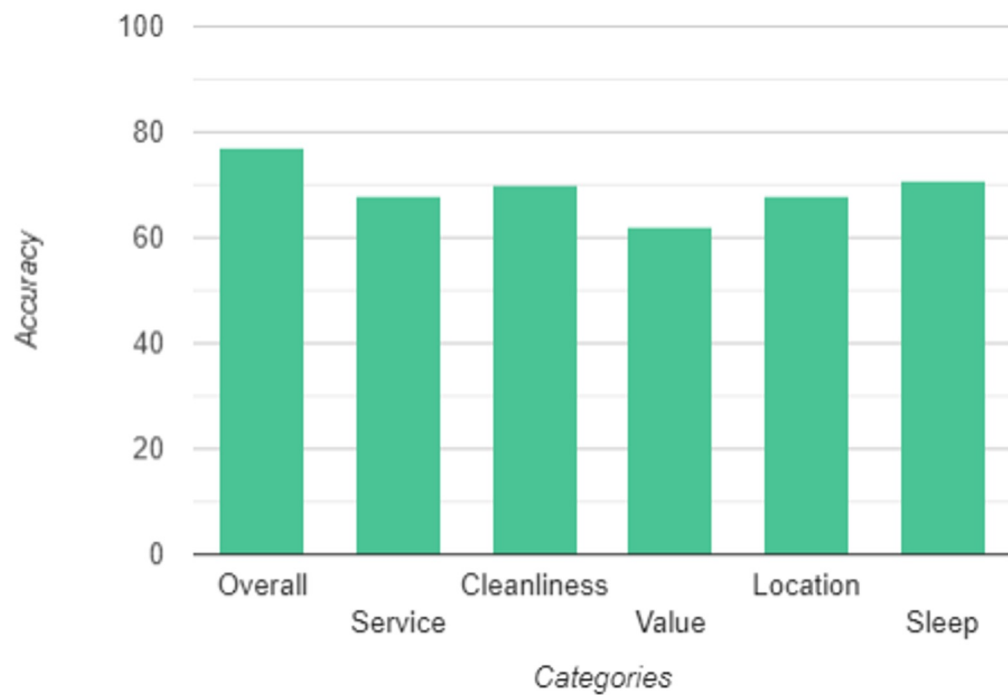
Model



Model



Results

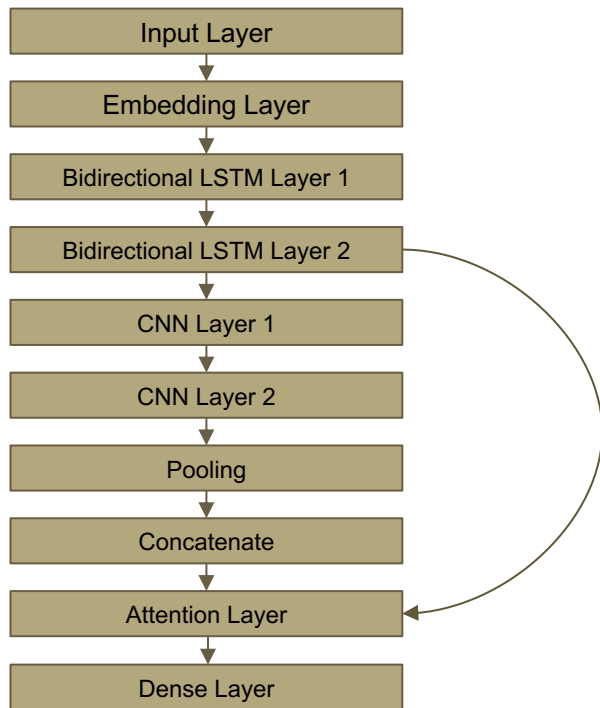


Future Work

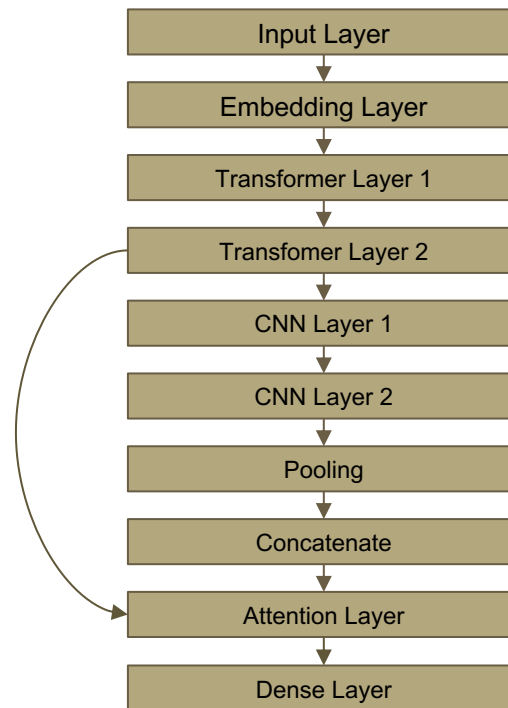
- Adding attention layer from LSTM and CNN output
- Making model using Transformers or BERT
- Change in sentiment over the time



Future Work



RCNN with Attention layer



Transformer with Attention Layer

References

- J. Zheng and L. Zheng, "A Hybrid Bidirectional Recurrent Convolutional Neural Network Attention-Based Model for Text Classification," in IEEE Access, vol. 7, pp. 106673-106685, 2019, doi: 10.1109/ACCESS.2019.2932619.
- The asymmetric effect of review valence on numerical rating: Tsao, Hsiu-Yuan & Chen, Ming-Yi & Lin, Hao-Chiang & Ma, Yu-Chun (2018)
- Apply word vectors for sentiment analysis of APP reviews Xian Fan; Xiaoge Li; Feihong Du; Xin Li; Mian Wei (2016) 3rd International Conference on Systems and Informatics (ICSAI) Year: 2016 | Conference Paper | Publisher: IEEE
- Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers Victoria Ikoro; Maria Sharmina; Khaleel Malik; Riza Batista-Navarro (2018) Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS) Year: 2018 | Conference Paper | Publisher: IEEE
- Comparative study of Twitter Sentiment On COVID - 19 Tweets Anu J Nair; Veena G; Aadithya Vinayak (2021) 5th International Conference on Computing Methodologies and Communication (ICCMC) Year: 2021 | Conference Paper | Publisher: IEEE
- A framework for sentiment analysis with opinion mining of hotel reviews Kudakwashe Zvarevashe; Oludayo Olugbara (2018) Conference on Information Communications Technology and Society (ICTAS) Year: 2018 | Conference Paper | Publisher: IEEE

Thank You