Project Increment 1

Related Work

This is a Natural Language Processing project of which focus is on sentiment analysis. We would like to extract meaningful phrases from reviews and classify them into most important features (and possibly their ratio) that hotels need to distribute their resources accordingly.

Dataset

The dataset was taken from Kaggle: https://www.kaggle.com/andrewmvd/trip-advisor-hotel-reviews which includes 20491 rows x 2 columns where rows represent reviews of customers who have experienced hotel services and columns are of the review content and their ratings.

Detail Design of Features and Analysis

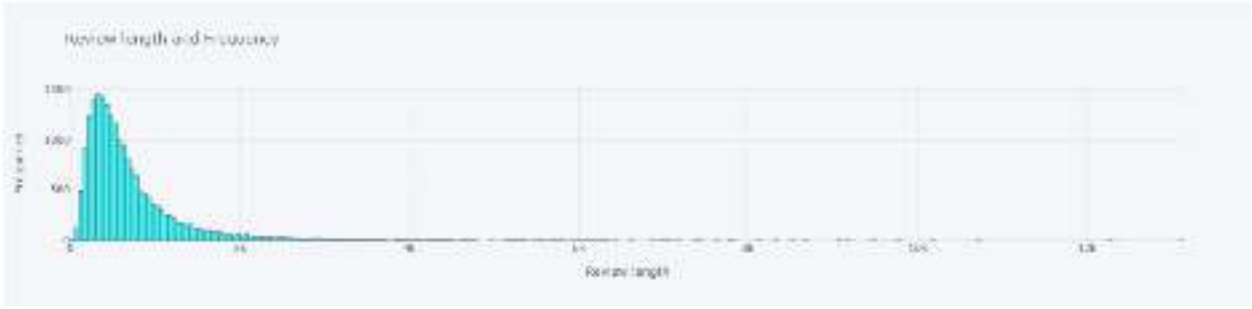
- Punctuation removal
- Exploratory analysis, such as review length vs frequency, or review length vs ratings
- Word cloud generation
- Spacy library's rule-based matching
- Review phrase extraction
- Clustering phrases / classification with the support of word vectors

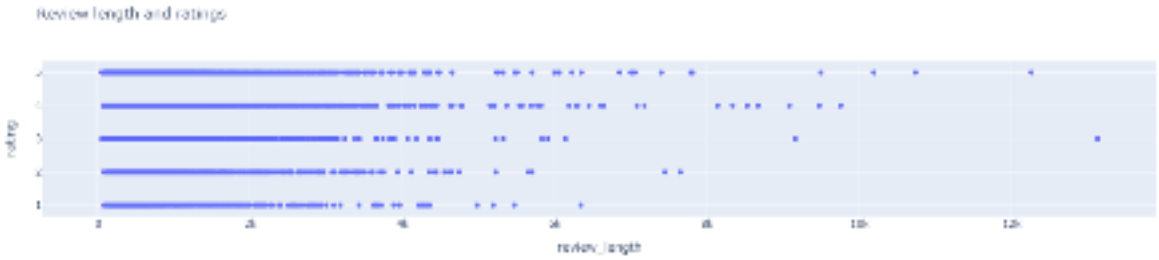
Implementation

Please find the code attached.

Preliminary Results

- We can see the pattern that those who are more satisfied are more likely to give reviews. Also, the data set is not balanced among ratings.





Project Management

Implementation status report

Work completed:

- Punctuation removal
- Exploratory analysis, such as review length vs frequency, or review length vs ratings
- Word cloud generation
- Spacy library's rule-based matching
- Review phrase extraction

Work to be completed

- Clustering phrases / classification with the support of word vectors

Issues / Concerns

- Should we use word vector to find the similarity of the extracted phrases with hotel's common features?

- Should we use a pretrained model?
- Should we cluster them and find the similar content of the phrases?

Github link: https://github.com/teohangxanh/5290/tree/master/Project

Works Cited

Agarwal, Basant, et al. “Sentiment Analysis Using Common-Sense and Context Information.” Computational Intelligence and Neuroscience, vol. 2015, 2015, pp. 1–9, www.ncbi.nlm.nih.gov/pmc/articles/PMC4381572/, 10.1155/2015/715730. Accessed 1 Apr. 2019.

Bhadane, Chetashri, et al. “Sentiment Analysis: Measuring Opinions.” Procedia Computer Science, vol. 45, 2015, pp. 808–814, www.sciencedirect.com/science/article/pii/S1877050915003956, 10.1016/j.procs.2015.03.159. Accessed 24 Nov. 2019.

Cyril, C Pretty Diana, et al. “An Automated Learning Model for Sentiment Analysis and Data Classification of Twitter Data Using Balanced CA-SVM.” Concurrent Engineering, 20 July 2021, p. 1063293X2110314, 10.1177/1063293x211031485. Accessed 26 Oct. 2021.

Devika, M.D., et al. “Sentiment Analysis: A Comparative Study on Different Approaches.” Procedia Computer Science, vol. 87, 2016, pp. 44–49, 10.1016/j.procs.2016.05.124. Accessed 14 June 2019.

Le, Bac, and Huy Nguyen. “Twitter Sentiment Analysis Using Machine Learning Techniques.” Advanced Computational Methods for Knowledge Engineering, 2015, pp. 279–289, 10.1007/978-3-319-17996-4_25. Accessed 15 Oct. 2020.

“Optimization of Sentiment Analysis Using Machine Learning Classifiers | Human-Centric Computing and Information Sciences.” Rdcu.be, rdcu.be/cAaGY. Accessed 26 Oct. 2021.

Puschmann, Cornelius, and Alison Powell. “Turning Words into Consumer Preferences: How Sentiment Analysis Is Framed in Research and the News Media.” Social Media + Society, vol. 4, no. 3, July 2018, p. 205630511879772, 10.1177/2056305118797724.