Topic Modeling and Aspect Based Sentiment Analysis on Yelp reviews

MICHIGAN ENGINEERING
UNIVERSITY OF MICHIGAN

Aparna Gopalakrishnan, Mexi Liang, Xue Xu

Introduction

Sentiment Analysis is a complex subject given the nuances of how varied humans write and the short-form nature of some of these texts. Aspect-Based Sentiment Analysis (ABSA) helps to understand the sentiment or opinions from text.

In this project we applied general topic model and Large Language Model techniques to Yelp reviews of restaurants, identifying key aspects ('food quality', 'tase', etc.) and sentiments

Objective

- For customers searching for a specific restaurant we want to provide a list of top aspects along with their sentiment score for that restaurant
- For business owners, we want to provide a comprehensive solution for improving aspects of reviews written for their restaurants

Data Resource and Cleaning

- Utilized Yelp dataset for reviews and business info.
- Convert JSON to CSV for efficiency.
- Filtered non-restaurant reviews, focusing on California.
- Identified restaurants using business categories.
- Resulted in refined California restaurant review dataset.

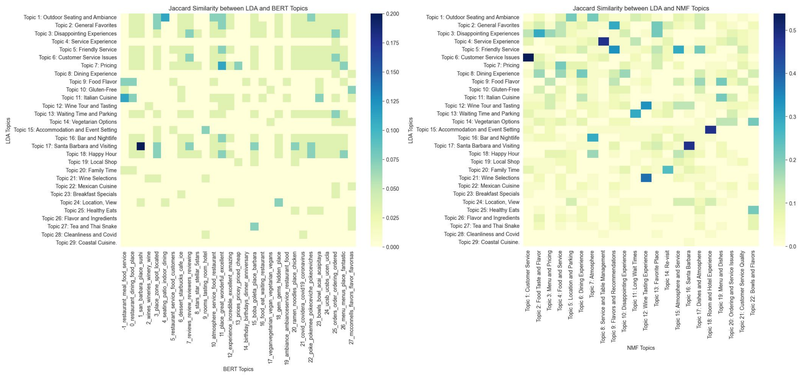
Process and Methodology

Aspect Extraction	Aspect Sentiment Polarity Detection	Review Recommendation
Extract Aspects from Reviews	Calculate Sentiment Polarity with Confidence Score	Ranking Reviews Based on Aspect Ratings for Similar type of Restaurants
Methods: **Topic model**: LDA, NMF, BERTopic **LLM**: SetFit Pre-trained Deberta	**Methods**: Pre-trained Deberta	**Methods:** Cosine Similarity

Result and Evaluation

LDA, NMF, BERTopic Topic Models

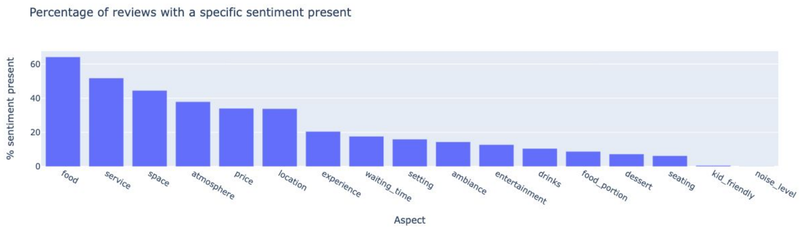
- 21 aspects was extracted. #: LDA 18, NMF 11, BERTopic 17
- Three topic models are focusing on aspects like "Setting," "Service," and "Ambiance."
- LDA captured broader themes, NMF provided detailed insights with overlap, and BERTopic extracted specific cuisines and geographical references.



- **Jaccard similarity** shows moderate similarity between LDA and NMF (50%) and lower with BERTopic (20%), indicating consistency in some topics.

Pre-trained Deberta Model

- **Sentiment distribution analysis**: "price" is often associated with negative sentiment (35%), while "atmosphere" and "food" are predominantly linked with positive sentiment.

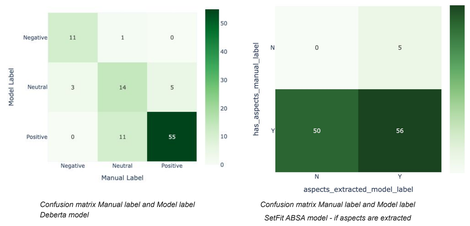


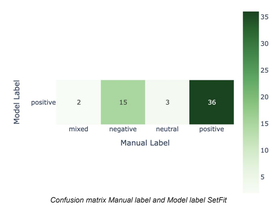
- **Comparison between Owner-provided and model-inferred aspects:** price positive labels from owners were less frequent but inferred positively by the model, and "kid friendliness" owner-provided labels didn't align well with model inference.

- **Evaluation:** Proxy accuracy calculated at 80.0%.

Setfit Model

- **Evaluation Result:** Aspect Extraction Accuracy: 50.5%; Sentiment Prediction Accuracy: 64.3%
- **Observations**: Model relies on exact word matches, lacking contextual understanding (e.g., cannot recognize "burrito" as a type of food).
- **Limited sentiment mapping**: Only positive sentiment detected, neglecting negative, mixed, and neutral sentiments identified during training.





Review Recommendation

- Designed a method for users to explore similar restaurants with positive reviews. Requires input of restaurant business ID and user-selected aspect.
- **Threshold Selection:** we establish a threshold value of 0.7, make sure provide enough number of similar restaurants.

Discussion and Conclusion

1. **Topic Models** show **consistent** in aspect “Setting”, “Service”, and “Ambiance”. **Surprising** emergence of **topics** like COVID and vegetarian related issues.
2. **Topic Model Limitation**: Aspect extraction relying on human inspection; **Sensitivity of topic models** to parameters.
3. **Fine-tuning the Deberta model** significantly **improves performance**, revealing positive sentiments from users about aspects such as atmosphere and food.
4. **Poor accuracy** of SetFit ABSA model in identifying aspect presence and sentiment.
5. **Ethical Considerations**: Potential challenges from dishonest and inaccurate reviews, including those from paid Yelp users.
6. **Acknowledgment** of sentiment biases introduced by Large Language Models (LLMs) and human evaluation to mitigate