Northeastern University, Khoury College of Computer Science

CS 6220 Data Mining - Final Project

Due: April 18, 2024(100 points)

Dharun Suryaa Nagarajan,Yichen Sun,Taiwei Cui

PROJECT GIT REPOSITORY

1 Introduction

Analyzing online passenger reviews provides valuable insights into the service quality and competitive positioning of airlines within the intensely competitive airline industry. These reviews serve as a comprehensive repository of firsthand experiences, reflecting passengers' opinions, expectations, and emotional responses. By utilizing tools such as sentiment analysis, airlines can identify areas for improvement and enhance service delivery to prioritize customer satisfaction. Surveys and feedback mechanisms play a crucial role in collecting consumer insights, aiding airlines in addressing key concerns.

2 Background

Online platforms such as Skytrax, Facebook, Twitter, and e-commerce websites serve as rich sources of public evaluations and textual remarks, offering insights into users' sentiments and attitudes. Sentiment analysis plays a vital role in automatically extracting and categorizing these sentiments across different levels, from aspect or feature level to document level, enabling the identification of positive, negative, or neutral expressions within user-generated content.
Concurrently, deep learning (DL), a subset of artificial intelligence (AI), has transformed the analysis of vast amounts of data, commonly referred to as big data. Employing hierarchical neural networks similar to the human brain, DL excels in various tasks such as fraud detection, voice recognition, and language translation.

References:

1. "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text".
2. "Sentiment analysis model for Airline customers' feedback using deep learning techniques" Heba Allah Samir, Laila Abd-Elmegid and Mohamed Marie.
3. "PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis" Ke Li, Chen Zang, Heng Yang.

3 Approach

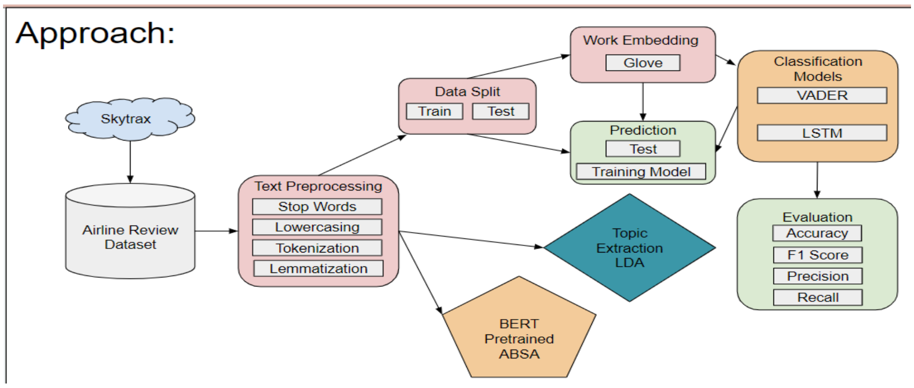
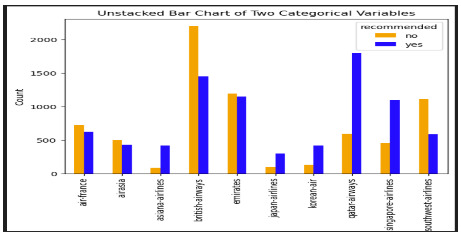


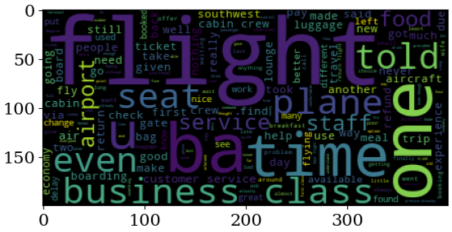
Figure 3.1: Approach

3.1 Data and Data Analysis

Amount of Data - We used the skytrax website which contains the detailed customer review of each airline represented. Using the Python libraries **beautifulsoup** and **requests**, we mined greater than 1000 pages of customer reviews. This gave us around 20k customer review for all the 10 airlines.

Data Analysis and Exploration - Since this is a sentiment analysis problem, we did some data





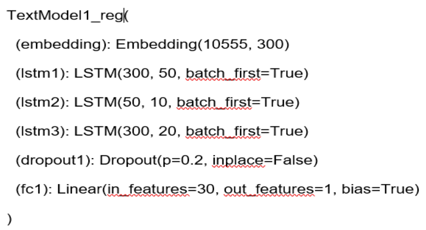
anaysis and pre-processing that helped us clean and understand the data. We had the highest amount of data for the BA airlines and the lowest for Japan Airlines. BA and Southwest airlines had the worst negative to positive sentiment ratio. Qatar and Singapore airlines had the best negative to positive ratio. We also noticed that before preprocessing the word cloud did not look good. It had a lot of stop words, non English words, lemmatization, tokenization. After pre-processing, the content became more aligned towards flight data.

3.2 Implementation

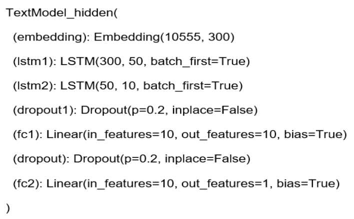
Sentiment Analysis Pre-processing: First use treebank-tag from NLTK to label the words with their POS to enhance performance of Lemmatizer. Then use NLTK.stem.WordNetLemmatizer to stem the words. We tokenized and sequentialized the words with package torch text. The network expects vectors as input. We used Standford GLoVe pretrained word2vec embedding as our embedding. These are words that we found no corresponding from GLoVe in our corpus:

$['waterish','unwelcomeness','unserviceability','unhelping','hamus','unpassionate','staggard']$ Most are misspell, some are failure of Lemmatizer. Model building up: We have tried 3 different models, all heavily based on LSTM.

Final Model : TextModel1 This model has 3 layers of LSTM units and 1 dense layers. Its structure is shown as below. The loss function is binary cross entropy. The input of the model is headers of reviews and contents of reviews. Layer lstm1 will take in the contents, and lstm3 will take in headers. Layer lstm2 take the output of lstm1 as input, and the droupout1 takes the last hidden status of lstm2 and last hidden status of lstm3 as input. The output of model is linear because we used BCEWithLogitsLoss, which combined sigmoid activation with binary cross entropy.
We then focus on the regression mission. This model uses the 2 LSTM trained in model2, and fine tuned 2 dense layer. This model takes mean square loss and try to predict the rating of each reviews.

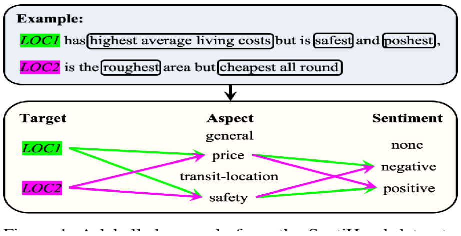


```
TextModel1_reg(
  (embedding): Embedding(10555, 300)
  (lstm1): LSTM(300, 50, batch_first=True)
  (lstm2): LSTM(50, 10, batch_first=True)
  (lstm3): LSTM(300, 20, batch_first=True)
  (dropout1): Dropout(p=0.2, inplace=False)
  (fc1): Linear(in_features=30, out_features=1, bias=True)
)
```

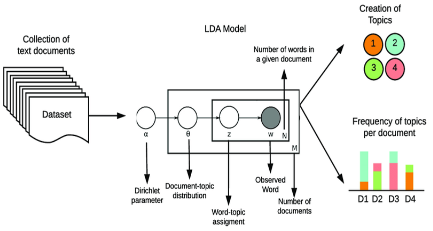


```
TextModel_hidden(
  (embedding): Embedding(10555, 300)
  (lstm1): LSTM(300, 50, batch_first=True)
  (lstm2): LSTM(50, 10, batch_first=True)
  (dropout1): Dropout(p=0.2, inplace=False)
  (fc1): Linear(in_features=10, out_features=10, bias=True)
  (dropout): Dropout(p=0.2, inplace=False)
  (fc2): Linear(in_features=10, out_features=1, bias=True)
)
```

Topic Extraction Objective here was to implement sentiment analysis but also to understand the features/characteristics that cause them. We used the preprocessed dataset which went through lowercasing, stopwords removal, tokenization and lemmatization.
1. The topic modeling failed to give us the characteristics cause of the abundance of non feature words like flight, British, Singapore etc.
3. We moved onto the better TF-IDF vectorization technique but the results failed there too cause of poly sentiment phrases in one review.
4. This led us to aspect based sentiment analysis. Using the tf-idf vectorized dataset, we used the pyABSA model based on the BERT language model to give out spect based sentiments.



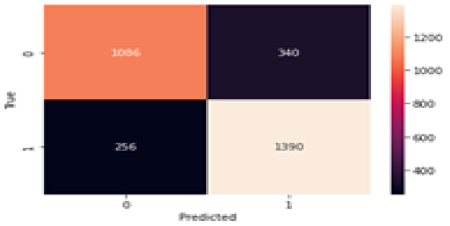
Example:
LOC1 has highest average living costs but is safest and poshest,
LOC2 is the roughest area but cheapest all round
Target
Aspect
Sentiment
LOC1
LOC2
general
price
transit-location
safety
none
negative
positive

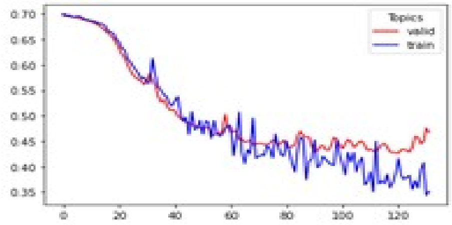


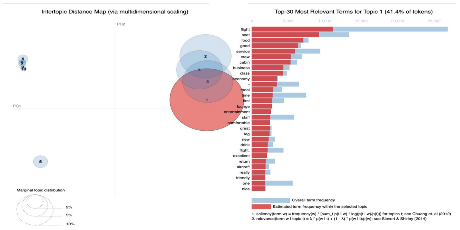
4 Results and Evaluation

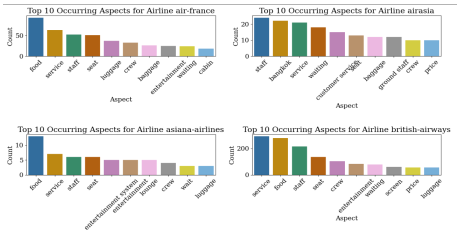
Well-Reasoned Evaluation
1. The accuracy of around 82% suggests that the model is performing reasonably well in classifying the testing data points.
2. The F1 score of approximately 0.82 indicates a balanced trade-off between precision and

recall, suggesting that the model is performing well in both minimizing false positives and false negatives.
3. The precision of approximately 0.82 indicates that the model has a high accuracy in predicting positive instances.
4. The recall of approximately 0.82 indicates that the model is capturing a significant portion of actual positive instances. 5. Further it can be noticed that, of the negative sentiment we see a majority of the complaints are about the Food and seats, some are about the entertainment unit. Using the ABSA we moved from flight as the topic to all the characterisitcs of the flight









5 Conclusions

By leveraging this sentiment analysis, airlines pinpoint areas for improvement and enhance service delivery to meet customer expectations. Additionally, sentiment analysis can aid in identifying trends and patterns in consumer feedback, allowing airlines to make data-driven decisions to stay competitive in the market.

Future Directions:
1. We haven't explore the header(title) of each review. We believe these texts are general idea. Of the commend, which worth exploring further.
2. The ABSA model is currently BERT pretrained model. In future to get better aspects from data we need to re-train/fine-tune the model to industrial data to update the parameters.
3. From industry point, we also need to capture airline type like Boeing, passenger class, source and destination, etc to deep dive on different segments of root cause analysis, this will require NER, POS tagging and other techniques.