Richa Meherwal

Netid: meherwa2

CS 410 Fall 2020

Technology Review - Latest Applications of Text Mining in evaluating customer satisfaction in Travel industry sector

INTRODUCTION

In todays world, extracting value out of online reviews to analyse various aspects of customer satisfaction is deemed necessary. With the advent of online shopping and rise fo e-commerce sectors, it is now a customer right to know and understand the service/product before investing in it. Although overall rating is a quick way to help a customer figure out whether a service or product is something the customer would like to hire or buy, it is through the reviews and the detailed description in the online comments that the customer is influenced to buy or not buy a product. For example, when shopping on an e-commerce website, a customer clicks on the product displayed on the initial page by looking at the overall rating. On the next page, where the product description, reviews are provided in details, the customer quickly scrolls through the comments to see if the rating is actually a reflection of the reviews that are seen. It is like a second and final screening before buying a service/product. This is why text mining has diversified its applications to various sectors in the industry including airlines, hotels, healthcare etc. In this technology review, I will introduce a few of the application of text mining on analysing reviews in airlines and hotel industry. These two industries are a major portion of the travel industry, so we will actually be reviewing the latest research in harvesting value from the online reviews on travel industry.

Application of Text Mining on Hotel Reviews

One of the latest research [1] has been identifying the importance of determinants of customer satisfaction in hotels using text mining. Many hotel managers are now looking forward to incorporating feedback from social media. This has enabled a lot of research in measuring or quantifying customer satisfaction levels using text mining techniques such as sentiment analysis and topic extraction. Sentiment analysis is performed on document-level, sentiment-level and aspect-level. Aspect based Sentiment Analysis requires a lot of topics/aspects to be extracted and them perform sentiment analysis on it. Most of the research that has already been done, have always linked review ratings to hotel selection. However , in this research, the goal is to identify important topics from the UGC reviews and then perform a sentiment analysis on the most meaningful topics.

DATASET

This dataset consists of 410 hotels in Ho Chi Minh City, Vietnam up to Feb 2017. 58,381 travelers with purchased evidence were giving 75,933 reviews.A manually annotated dataset is created by the researchers to assign aspect-level annotation for 3500 reviews.

PRE-PROCESSING

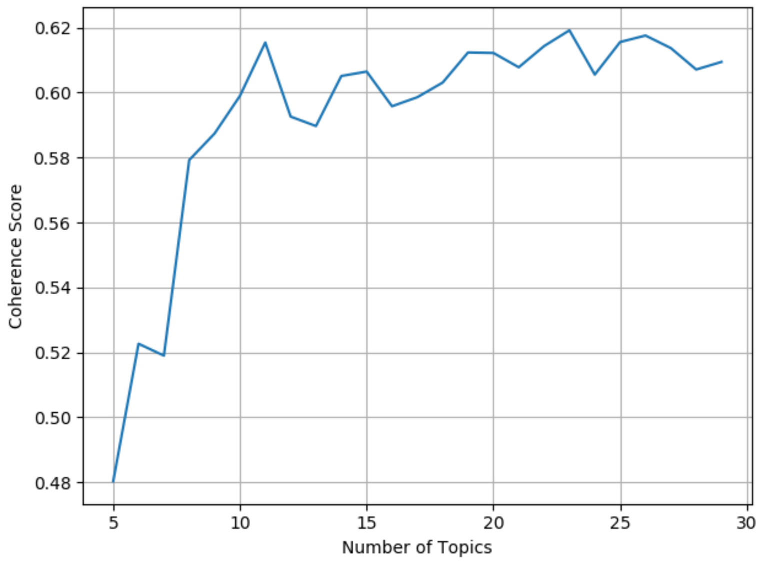
Language detection, noise removal and Lemmatization

PROCESS

A deep learning algorithm BiLSTM-CRF is used to extract aspect terms and associate a sentiment with every topic. It is fed with a new annotated dataset. The interesting thing to note here, is the researchers now pair it with IOB which basically determined the position of each word in noun phrase with inside-outside tagging before feeding it back to the model. In the below table, B denotes beginning of aspect, I denotes inside of an aspect. POS- for positive and NEG- for negative.

Words:	**Staff**	are	very	friendly	,	but	the	**rooftop**	**bar**	is	quite	noisy	.
Labels:	**B-POS**	O	O	O	O	O	O	**B-NEG**	**I-NEG**	O	O	O	O

The model is trained with annotated dataset. And then aspect terms and polarities are extracted for the non-annotated dataset. A coherence score shown below determines the optimal number of topics to extract. Here the number is 11 as score starts to flatten from there onwards.



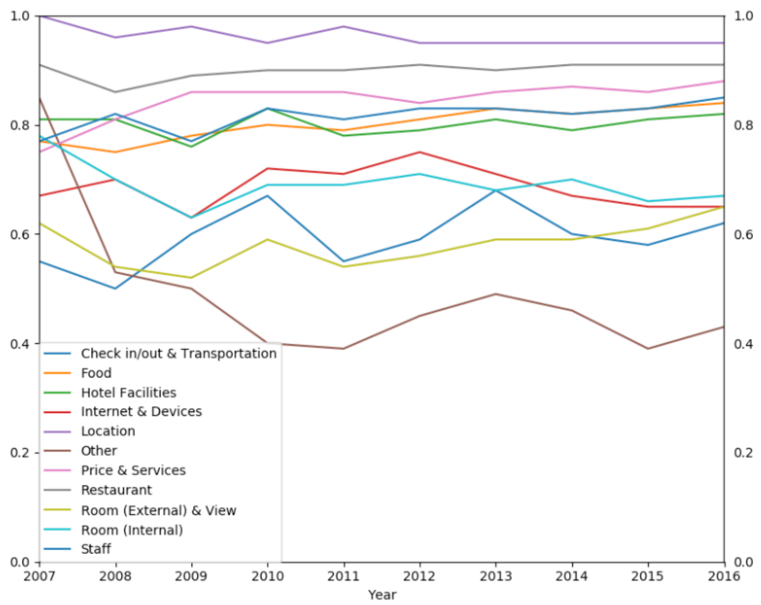
Aspect terms are then fed into LDA to extract the topics.

METHOD OF ANALYSIS

It is finally inferred as shown in the table here that each topic such as Food is associated with more positive (21.7%) than any other sentiment. Similarly, other observation was that Food , Staff , Internal Room Facilities, Internal Room Facilities & View occupied approximately 69.60% of all aspect terms. This means that customers most often talk about only these topics at large. Below is a graph of trend along the years for every topic and the positive polarity score on the y-axis. We can observe some aspects show a positive trend such as Food.

Inferred Topic	#Positive	#Neutral	#Negative	%Aspects
Food	**86,698**	**14,210**	9,100	18.90
Hotel Facilities	38,183	5,970	5,844	8.59
Price & Service	43,397	3,289	5,367	8.94
Check in/out & Transportation	6,414	3,323	1,831	1.99
Staff	55,672	7,632	7,120	12.10
Internet & Devices	7,859	365	3,326	1.98
Other	4,144	543	5,714	1.79
Room (Internal)	50,117	8,515	22,753	13.98
Restaurant	8,692	617	468	1.68
Room (External) & View	76,127	3,8997	**28,170**	24.62
Location	30,314	970	343	5.43

Table 9: The result for interring topics and Summarization



Application of Text Mining in Airline Industry

A recent research article [2] on airline customer satisfaction text mining emphasises that relevant features can be accurately extracted from OCRs as the key to identify the competitive edge that is needed for airline companies to excel. The analysis is done using LDA, a popular topic extraction technique in Text Mining on a UGC dataset. The dimensions discovered not only give an overall insight into what areas customer satisfaction is seen negatively and positive but it gives a layered analysis on what topics carry a sentiment in certain types of passenger based on location, time etc.

DATASET

This dataset comprises of 55,000 OCRs, covering over 400 airlines and passengers from 170 countries. It is collected from Air Travel Review (ATR), a widely used website for review collection for airlines industry. It has over 681 airlines and 725 airports worldwide. The final dataset had a publication date, review and answer in Yes/No to a question - “Would you recommend this airline”. Some reviews consisted of contextual information such as nationality , cabin flown.

PRE-PROCESSING

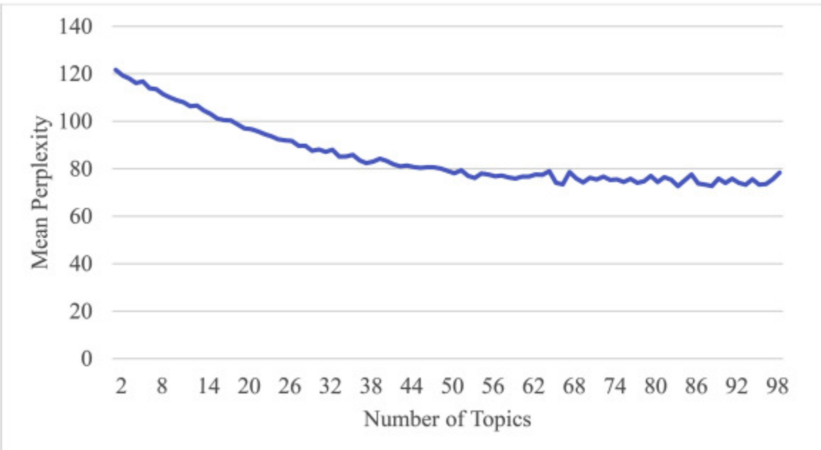
A web crawler was developed to crawl the ATR website and extract all the reviews in the html code. Challenges that were faced included 1) dealing with casual language and grammar 2) pulling out only those dimensions that give valuable information 3) to map text to number using encoding techniques 4) to deal with different languages. Using entity recognition airport names, prices, dates, flight codes, airline names were identified and replaced with a standard term starting with '_' , so as to separate satisfaction keywords from context.

TOOLS USED

Python 3.6, NLTK, sklearn

TOPIC EXTRACTION PROCESS

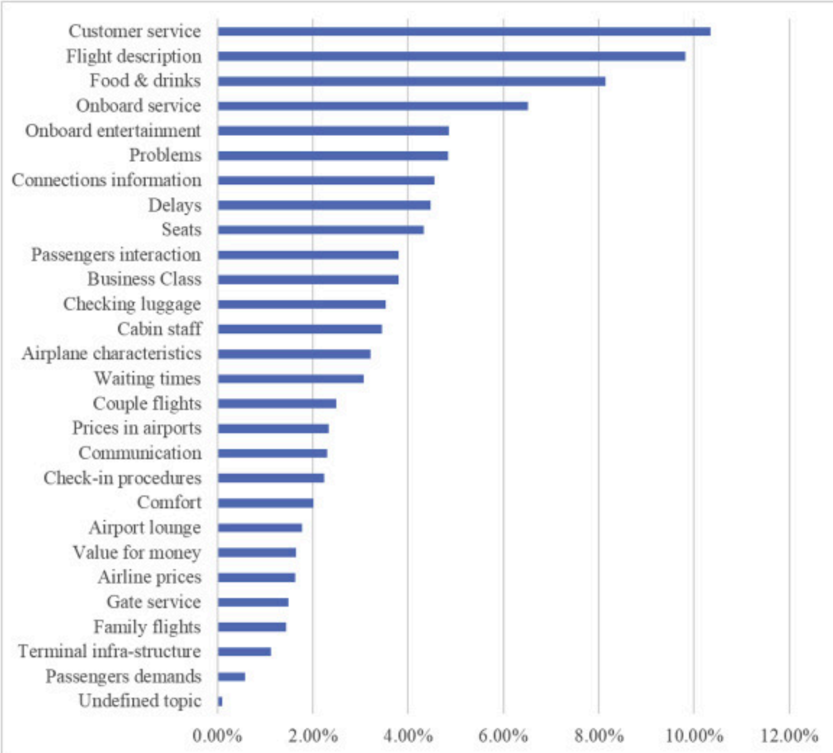
POS tagging was applied and only nouns and adjectives were kept. Nouns were used to help with identifying the topics for customer satisfaction whereas adjectives were used for sentiment analysis. A matrix was created where rows were docs/reviews and columns were TF-IDF for each term. A Latent Dirichlet Allocation model was used. In LDA, each topic can be shown through a word distribution. The challenge here was that number of topics was not known, setting less number of topics could miss out on information and more number of topics could make the model more complex. To combat this, a perplexity measure was used to identify the appropriate number of topics where lower the perplexity measure, better is the model.



Once LDA was fitted and number of topics was determined, then the word to topic likelihood was used to determine the top ranked words that are relevant to the topic. A consensus process is followed till researchers establish which words correctly describe a given topic.

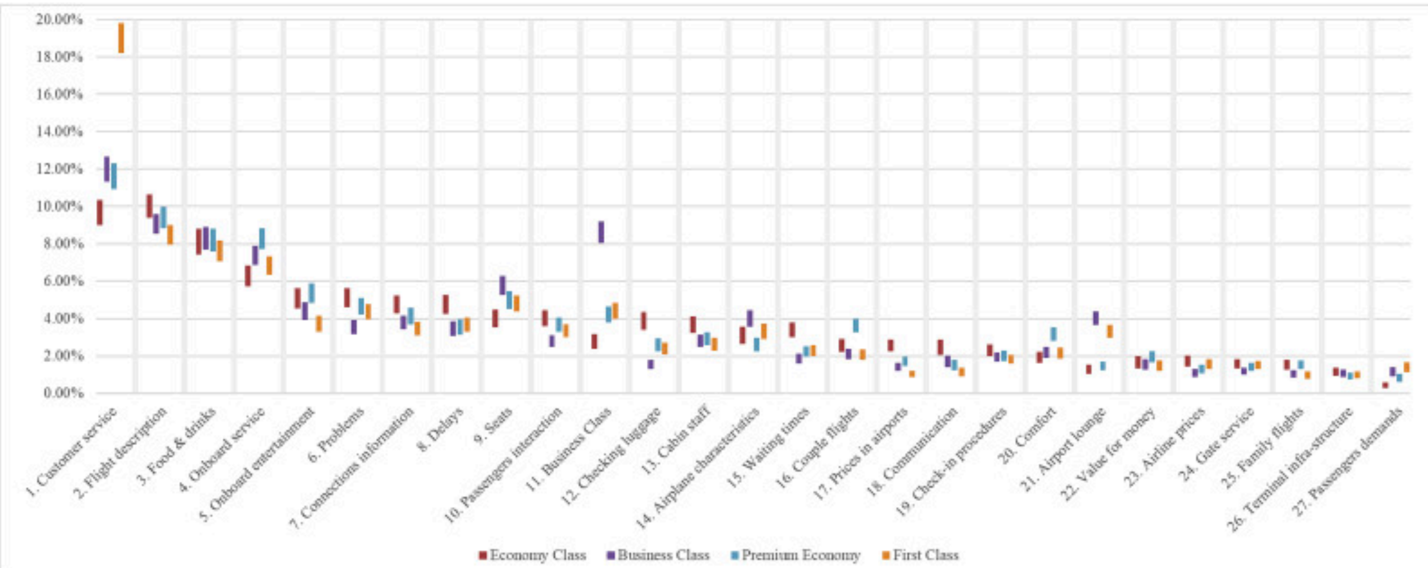
METHOD OF ANALYSIS AND ACCURACY

Distribution of topics across all reviews is noted. Then, reviews are grouped by criteria such as airline, time, passenger nationality, type of passenger, and type of cabin flown and then distribution of topics is noted for each group. For sentiment analysis, Naive Bayes Classifier is used. Labeled dataset of negative reviews, those that were rated 0 and positive reviews , those that were rated 10 are fed into the classifier. The dataset is then fitted and sentiment strength coefficients was assigned to each adjective. This provides the list of important adjectives and the sentiment scores associated with them. Similar meaning words (882 adjectives) were then grouped together and the final topic word list was now with 27 dimensions of satisfaction.



Another way the data was analysed was using confidence intervals. Dimensions are laid down like shown against the percentage CI.

Each of these CI graphs were built for each group and dimensions. It was easy to infer what group of people cared about which dimensions and also compare them with other groups. For example First class care a lot more about customer service then other cabin flown types. Similarly, business class care a lot more about airport lounge than economy and premium economy.



Dimensions and adjectives were then modelled with logistic regression to predict the recommendation with an accuracy of 79.95%.

CONCLUSION

TEXT MINING ON HOTEL REVIEWS

In my opinion, the graph i.e positive polarity score against aspect terms could have been visualised better in sense that line graphs are not smoothed and it is difficult to interpret what kind of trend is observed over the years. Topics such as “Hotel Facilities” and “Others” could be broken down into more specifics. The scope for improvement I see here, is to segregate these topics reviews and analyze them at a more granular level to find out the exact features affecting the quality of each of their services. For example, Staff features could be more hospitality, more accommodating with on-demand services , polite/rude and so on. Although sentiment numbers do indicate that the overall customer satisfaction is positive , it fails to find out for what exact topics under Staff is it negative and does not offer any suggestive help to the managers as to where the areas of improvement are. If ratings were available, they too could have considered as a prior in providing final polarity score.

TEXT MINING ON AIRLINE REVIEWS

In my opinion, there was certainly more refining to be done with extracting relevant adjectives/ dimensions. Dimensions like “Business Class” or “Problems” are not very descriptive and do not indicate what the issue or criteria for satisfaction is. As I described earlier for hotel reviews, these areas could be broken down at a granular level. Furthermore, Accuracy for LDA could be improved by applying a similar approach like Aspect Based Extraction using a deep learning method.

Note that both models use different measures to extract the optimal number of topics to be assigned for LDA. One uses perplexity measure , the other a coherence score. Both are evaluation techniques to check how accurate the prediction is and could be used interchangeably. Also, the annotated dataset was very small and accuracy could have been improved further if more human annotations were made available.

REFERENCES

[1] Tran Xuan, Thang & Ba, Hung & Huynh, Van-Nam. (2019). Measuring Hotel Review Sentiment: An Aspect-Based Sentiment Analysis Approach. 10.1007/978-3-030-14815-7_33.

[2] Lucini, Filipe & Tonetto, Leandro & Fogliatto, Flavio & Anzanello, Michel. (2019). Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. Journal of Air Transport Management. 83. 10.1016/j.jairtraman.2019.101760.