Recommender System Utilizing User Review Text Corpora

Shenghan Liu, Catherine Hou, Vincent Le

Abstract

Over time, we rely more and more heavily on online platforms such as Netflix, Amazon, Spotify, which are embedded with the recommendation system in the applications. They know users' preferences by collecting their ratings, recording the clicks, combing the reviews and then recommending more items. In building the recommender system, review texts can hold the same importance as the numerical statistics because they contain key phrases that characterize how they felt about the review. For this project, we propose to build the recommender system with primary focus on the text reviews analysis through TF-IDF (term frequency-inverse document frequency) and AutoPhrase and to add targeted segmented analysis on phrases to attach sentiments to aspects of a restaurant to rank those recommendations. The ultimate goal is designing a website for deploying our recommender system and showing its functionality.

1 Introduction

Many recommender systems use ratings as an indicator of a user's sentiment towards a business, and remains the most popular and most used feature in terms of determining whether or not a business is 'good'. However, without utilizing the text of the reviews, most users will be left in the dark in regards to the features they may not realize would be important to them in data that simply cannot be represented in numbers. There are multiple factors that go into a rating: wait time, service, quality of food, cleanliness or even atmosphere - for example, a restaurant could have positive sentiment towards the food but negative sentiment towards the wait time. In order to solve this problem, our aim is to include such sentiments that can be found in the review text and turn that into data which can be used to further improve business recommendations to users. In order to do this, we extract the review text from a dataset of reviews and businesses and segment the text using a program called AutoPhrase. Afterwards, sentiment analysis is performed on the text segments which will reveal what parts of the text have positive or negative sentiment. This, coupled with the workings of a standard recommender system, should give users more interesting and personalized recommendations when it comes to businesses.

In investigations into prior recommender systems, many utilized a process known as Collaborative Filtering, which is a technique that collects data regarding user preferences and recommends items based upon the similarities of the users, and is used in many recommender systems. Other methods used in recommender systems include Alternating Least Squares, K-Nearest Neighbors and Stochastic Gradient Descent, among others. Such techniques are well regarded and highly used, but noticeably lack components that take advantage of text data, which has the potential to increase the accuracy and personalization of the recommendations given. Given that text data reveals much into why users give certain ratings, the text data can be used in order to create new features, or can be applied in different ways that can sort or rank businesses.

Much like previous works, the data used in this project is the Yelp Academic Data set, which is a dataset that was officially created and released to the public by Yelp and is used in many previous works in terms of recommender systems. This dataset contains 5 JSON files, in total containing data on roughly 5.2 million user reviews, 174,000 businesses, spanning 11 different metropolitan areas. Among this data, features such as business id, user id, review id, text, rating, and more are included in the data, which provides a more than adequate amount of data to work with in order to create a cohesive recommender system. For the sake of our project, we focused mainly on 2 JSON files: yelp_academic_dataset_business.json, which contains information about businesses, as well as yelp_academic_dataset_review.json, which contains information about reviews. Using these two files, we can merge data frames on business ids in order to track which businesses have good or bad reviews, and associate review text to certain businesses.

variable names	Description
business_id	Unique ID associated with a business
name	Name of the business
address	Address of the business
city	The city that the business resides in
state	The 2 character state code that the business resides in
postal_code	The postal code of the business
latitude	The latitude coordinates of the business
longitude	The longitude coordinates of the business
stars	The average number of stars received by the business
review_count	The number of reviews the business has received
is_open	0 or 1, for closed or open, respectively
attributes	Attributes of the business (may vary)
	Ex. "RestaurantsTakeout": True
categories	Categories that the business may fall under
	Ex. "Mexican", "Burgers"
hours	Business hours of the business

Table 1: Schema for yelp_academic_dataset_business.json

variable names	Description
review_id	Unique ID associated with a review
user_id	Unique ID associated with the user that wrote the review
business_id	Unique ID associated with a business
stars	The number of stars that the user gave the business
useful	The number of 'useful' votes received
funny	The number of 'funny' votes received
cool	The number of 'cool' votes received
text	The text of the review itself
date	The date the review was written, in YYYY-MM-DD format

Table 2: Schema for yelp_academic_dataset_review.json

2 Methods

Various methods were used in an attempt to experiment with our recommender system in order to determine which methods we would want to use in our system.

2.1 TF-IDF Recommendation Method

We experiment two approaches to start building our recommendation system – content-based and collaborative filtering. In preparation for the recommender system analysis, we join the business dataset and user dataset by the business id to get a more comprehensive dataset about the user and restaurants.

2.1.1 Restaurant-Restaurant Based

The first general approach we use to build a recommendation system is building a content-based recommender. The main idea is that: since each restaurant has the record of user review and rating, it is accessible to generate each restaurant review document from all its user reviews that can be used to recommend restaurants based on document similarity. Then, we decide to apply the method term frequency–inverse document frequency (TF-IDF), which assigns each word in a document a number that is proportional to its frequency in the document and inversely proportional to the number of documents in which it occurs.

For implementation, with the joined business and review dataset, we group the restaurant by its business_id, combine all user reviews on it within a document and calculate the average rating of all users' reviews on it. After getting each restaurant's document, we perform the data cleaning procedure to remove the unnecessary punctuation, extra space and lowercase all the words in the document to keep the consistency.

$$w_{i,j} = tf_{i,j} \cdot log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j

$df_{i,j}$ = number of documents

N = total number of documents

Then, for each restaurant, we run the term frequency–inverse document frequency (TF-IDF) Vectorizer to get the word distribution of each restaurant's review document. [4] The ngram parameter we choose for the TF-IDF vectorizer ranges from 1 to 3, which means we care about the single word phrase as well as possible two-word phrases and three-word phrases in our vectorizer process. The result of TF-IDF vectorizer throughout the whole grouped restaurant dataset has a significant number of entries (more than 100000), which means the vectorized phrases entries for the all restaurant are plentiful.

As for generating the recommended restaurants, we apply the cosine similarity to the TF-IDF vectorizer to get a similarity score matrix for all restaurants in the dataset.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|}$$

This similarity measures the cosine of the angle between the two vectors (A, B) being compared.The lower the angle between two vectors, the higher the cosine will be, hence yielding a higher similarity factor. The score of similarity ranges from -1 to 1, where 1 means exactly the same and -1 means opposite at all. In our case, if two restaurants' TF-IDF vectors preserve a high cosine score, it tells us that two restaurants have similar review word distribution, which is reasonable to recommend one while the user is querying another.

After generating the cosine matrix, we can easily generate a list of recommended restaurants given a sample restaurant business id we are interested in. We search through the cosine matrix and pop up the top N restaurants which have the closest similarity scores with our target restaurant. In case of the tie recommended restaurants, we decide to rank the restaurants based on the user's average rating. In the future, we might leave the option of filtering to the user.

2.1.2 User-User Based

The second general approach we experiment is to build a collaborative filtering recommender. The main idea is to generate the review document for each user and recommend similar users based on the similarity of their review word distribution. The experimenting procedure for this user-user based method matches with the restaurant-restaurant based method. We first combine the user review into one document and preprocess the review document for further TF-IDF vectorizer method. Then, we apply the TF-IDF vectorizer and use cosine similarity as evaluation criteria for recommending the similar users.

But we face quite a lot of challenges in the experiment. The most influential one is that our final website does not preserve a user property since we neither have actual active users as support of our database nor enable users to add review dynamically. Under such circumstances, without a unique user profile section, the user-user approach can not work at all because we can not link a user with another user in lack of their user information. Besides, in our exploration, we find a significant number of users from our dataset that only contain less than two review records. While searching for similar users of this kind of users, the results are quite biased because the user review document might only contain one sentence with very limited information and less representative word distribution. Without a good solution to handle these difficulties, we decide not to include this method in our implementation eventually.

2.2 Food Query Using Targeted Sentiment Analysis

Based on a food query like fried chicken, our recommender system will list the top five recommended restaurants using targeted sentiment analysis. Sentiment analysis determines whether a given review has positive, negative, or neutral sentiment. Polarity-based sentiment analysis is a binary output, while valence-based sentiment analysis is a continuous output of how positive or negative the document is [2]. A subset of sentiment analysis is called targeted sentiment analysis. Instead of finding the sentiment of a particular review, we want to find the sentiment towards particular aspects in a review: customer services, popular dishes, etc.

The first step is preprocessing the reviews text using AutoPhrase, an automatic phrase mining method that uses Wikipedia's database to find and annotate the highest quality single-word and multi word phrases. We trained the model on Yelp reviews text, then performed the phrasal segmentation step to annotate reviews text. Each phrase is annotated with "<phrase>...</phrase>" [3].

Original	"The fried chicken was delicious!"
AutoPhrase	"The <phrase>fried chicken</phrase> was delicious!"
VADER	'neg': 0.0, 'neu': 0.501, 'pos': 0.499, 'compound': 0.6114
Positive Mentions	fried chicken: 1

Table 3: An example of how targeted sentiment analysis works.

Next we run sentiment analysis on each sentence. The sentiment of the sentence is related to the sentiment towards the phrase. For example, if we had a sentence like "The <phrase>fried chicken</phrase> was delicious!", the phrase fried chicken has a positive sentiment.

To perform sentiment analysis we use a valence-based sentiment analysis tool called VADER (Valence Aware Dictionary and sEntiment Reasoner). VADER is a pre-trained model provided by NLTK (Natural Language ToolKit), a package that includes text processing functionalities. Training a model from scratch requires expensive resources (time, people), while the ratings in VADER have been previously trained by people's ratings [2].

For each word in a text it will assign it a value [-1, 1] of how positive or negative it is. For example, a word like "great" has a higher value than "good". VADER has four kinds of metrics: negative, neutral, positive, and compound. The first three describe the percentage of the sentence that falls under those categories. The compound is the sum of these values [2].

If the compound score for a sentence is greater than 0.3, then the phrase in that sentence is marked as positive. For each restaurant we aggregate these positive phrases by their total count. When a user queries for "fried chicken", we will sort our list based on the most positive mentions of "fried chicken". If there is a tie, the secondary ordering is based on the highest average star rating.

Info	**Description of Info**
Restaurant Name	Name of the Restaurant
Categories	Categories as listed in Yelp
Stars	Average Star Rating
Number of Mentions	Number of times the query had positive sentiment
Good Service	Number of times good/friendly service was mentioned
Website	URL for the restaurant's Yelp page

Table 4: The format that the recommender system displays its recommendations on our website. The objective of this table is to summarize a restaurant's aspects without having to read reviews.

3 Results

For the results section, we present a few case studies to evaluate how our recommendations perform by comparing our recommendation result with Yelp's search result. The cases below are generated with the location as Las Vegas because that is the city for which we have the most reviews.

For the restaurant query, the evaluation will focus on the relevance of results, quality of results and quantitative comparison. For the phrase query, the evaluation will focus on three aspects: the relevance of the results, the quality of the results, and other uses for our results.

3.1 Restaurant Query

Info	Description of Info
Restaurant Name	Name of the Restaurant
Restaurant Address	Address of Restaurant as listed in Yelp
Stars	Average Star Rating
Number of Reviews	Number of Reviews the Restaurant received in Yelp
Categories	Categories of Restaurant as listed in Yelp
Website	URL for the restaurant's Yelp page

Table 5: The interpretation of the restaurant query results

3.2 Case: Bellagio Patisserie

3.2.1 Relevance

Figures 1 and 2 below correspondingly represent the Bellagio Patisserie restaurant query results of our recommender website and Yelp official website. Note that for comparison purposes, we only compare the recommended restaurants, which does not involve our query restaurant Bellagio Patisserie.

Figure 1: This is the information of our query restaurant Bellagio Patisserie from Yelp website.


Bellagio Patisserie

Restaurant Address: 3600 S Las Vegas Blvd
Stars: 4.06
Number of Reviews: 139
Categories: Restaurants, Coffee & Tea, Cafes, Desserts, Food
Website: http://yelp.com/biz/55FnSahXfQoiueQjrKC86w/?show_platform_modal=True

Figure 2: This is the information of the top two recommended results from our recommender website. Jean Philippe Patisserie is a chained restaurant that occupies Top 1 and Top 2 while Cafe Belle Madeleine is the Top 3. For showing the uniqueness and making a better comparison, we just include one Jean Philippe Patisserie below.

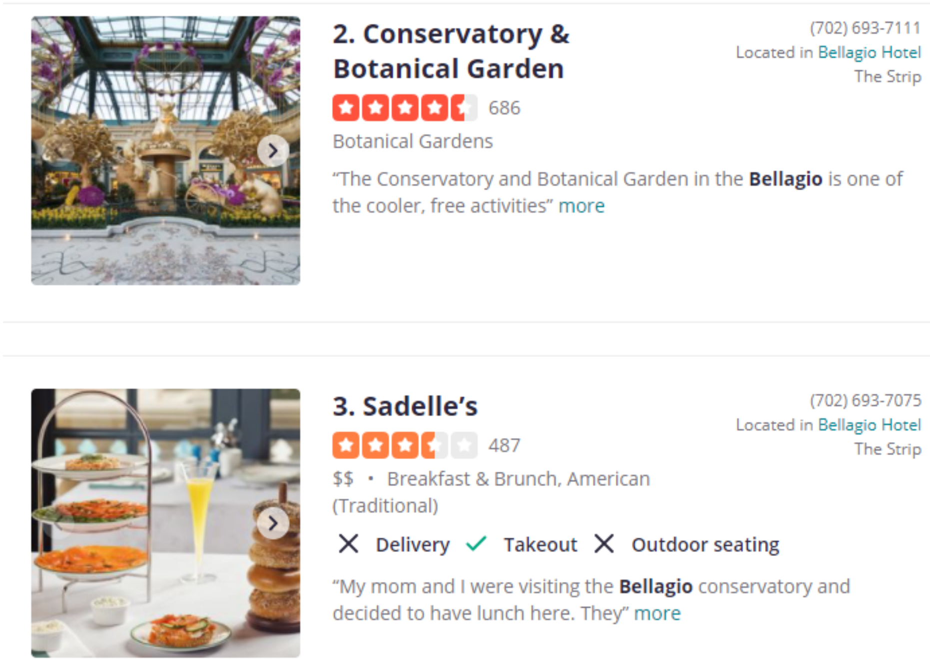

Jean Philippe Patisserie

Restaurant Address: 3600 S Las Vegas Blvd
Stars: 4.11
Number of Reviews: 1223
Categories: Cafes, Food, Desserts, Coffee & Tea, Restaurants
Website: http://yelp.com/biz/UUGoM4q4i8rK2CBRS0xDAw/?show_platform_modal=True

Cafe Belle Madeleine

Restaurant Address: 3655 Las Vegas Blvd S
Stars: 3.30
Number of Reviews: 265
Categories: Bakeries, Food, Cafes, Restaurants, Breakfast & Brunch, Coffee & Tea
Website: http://yelp.com/biz/jbbFSVdIrac2QBuSV-NBLw/?show_platform_modal=True

Figure 3: This is the information of the top two recommended results from Yelp Website. Conservatory & Botanical Garden is the Top 1 recommendation and Sadelle's is the Top 2 recommendation. Information not included in the Yelp results are categories, which in this case are [Botanical Gardens] and [Breakfast & Brunch, American (Traditional)] respectively. Screenshot taken on 3/7/2021, from Yelp. [1]



3.2.2 Quality

To evaluate the quality of results, we compare the consistency of the popular dishes of the query restaurant and in result restaurants.

Restaurant	Popular Dishes
Bellagio Patisserie (Query Rest.)	Nutella Crepe, Chocolate Almond Croissant, Tiramisu
Jean Philippe Patisserie	Tiramisu, Chocolate Cake
Cafe Belle Madeleine	Tiramisu, Chocolate Crossilet, Gelato
Conservatory & Botanical Garden	None (not a restaurant)
Sadelle's	Eggs benedict, Sticky bun, French toast

Table 6: The table contains the top 3 popular dishes of above four recommendation results and the query restaurant as listed in Yelp.

3.2.3 Quantity

Table 7 below shows our recommendation result in comparison with the result of Yelp Website. The 4 example restaurants are some random popular choices in Las Vegas.

3.3 Phrase Query

3.3.1 Website Example Output

Restaurant	Our Results Top 1	Yelp Results Top 1
KFC	KFC	KFC
Pho Kim Long	Pho Little Saigon	Pho King Vietnamese Kitchen
Scoop LV	Eis Cream Cafe	2 Scoops of Aloha LV
Sunrise Coffee	Sambalatte	Mothership Coffee Roasters

Table 7: Querying random popular choices in Las Vegas.

Figure 4: The top results of querying "fried chicken" in Las Vegas

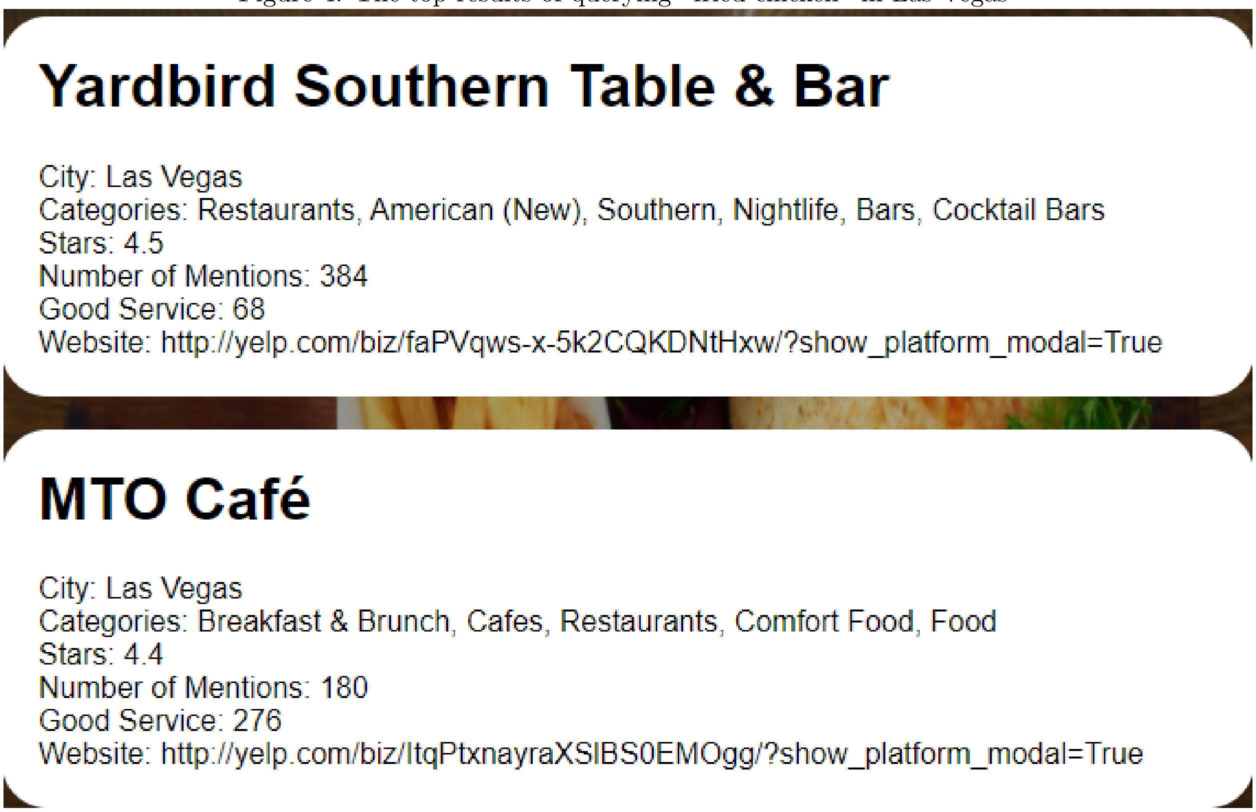


Figure 4 is a demonstration of the top two restaurants that we recommend using the query "fried chicken" in Las Vegas. "Number of Mentions" represents the number of times the query "fried chicken" was mentioned in a review with a positive sentiment. "Good Service" counts the instances where there was "friendly service" or "good service" in a review. We provided the businesses' Yelp links so that users could directly go to their Yelp page.

3.3.2 Relevance

In order to test our recommender system's relevance with Yelp's, we entered the queries into our system and Yelp's in order to observe how our results compared. These were the number of restaurants that had this dish mentioned in the reviews.

3.3.3 Quality

Query	Our Results	Yelp Results
Bacon Breakfast Sandwich	3	0
German Soft Pretzel	1	0
Macadamia Crusted Mahi Mahi	1	0
Grilled Asparagus	5	4

Table 8: Queries entered into both our recommender system and Yelp, and the number of mentions of the query in the review texts.

Info	Veggie House (Our Recommendation)	Gourmet China (Yelp's Recommendation)
Category	Vegan, Chinese	Chinese
Number of Reviews	1243	170
Star Rating	4.5	4.5
Number of Photos	49	12
Number of Reviews	185	70

Table 9: Top results for the "orange chicken" query.

In order to measure the quality of our recommendations in comparison to Yelp's recommendations, we used the query of "orange chicken" and compared the numbers between our recommendation and Yelp's.

3.3.4 Other Uses

To test the extensibility of our query, we tested our system with a "mother's day" query. This table contains the top two results from each system.

	Our Recommendation	Yelp's Recommendation
Top Result (categories)	Bouchon French, Cafes	Americana American (New)
Second Highest Result (categories)	Ohjah Japanese Steakhouse Japanese, Sushi Bars	Blume American (New), Cocktail Bars

Table 10: "Mother's Day" query on both our system and Yelp's.

Within the top results of each system, we compared the reviews that included "mother's day" of the top results. The number of reviews 4+ are how many reviews mentioning mother's day had at least a 4 star rating. The number of reviews <=3 are how many reviews mentioning mother's day are 3 and under.

Info	Bouchon	Americana
Number of Reviews	4260	644
Number of Stars	4	4.5
Number of Reviews 4+	14	6
Number of Reviews <=3	3	10

Table 11: Information regarding the "Mother's Day" top queries for our system and Yelp's.

4 Discussion

4.1 Interpreting Results

For the restaurant search query, when we searched for recommendations of some common and popular restaurants, the results of our recommender website and Yelp website mostly match with each other. As seen in Table 7, we tried a few popular restaurants such as KFC, Pho Kim Long, Sunrise Coffee, the results between our recommendations and Yelp's recommendations either had an overlap or were different restaurants with similar tags, which are hard to tell if our recommender website is better. But as we tried the restaurant that is not that popular, probably does not have a chain store at all, our recommender system did give a more comprehensive result.

Figure 1 shows the information of our outperformed query restaurant Bellagio Patisserie, which is a Coffee & Tea & Desserts restaurant inside a 4.5 star Hotel – Bellagio Hotel. Furthermore, in Las Vegas, there is only one such hotel and restaurant. Then, while searching for recommendation restaurants of such a special and unique restaurant, our top two recommended restaurants are completely different from the Yelp search results.

In Figures 2 and 3, we can see that our top two recommended restaurants both contain the Cafe & Tea tag while Yelp search results surprisingly gave botanical garden as the most similar result, which does not belong to a restaurant at all. It is easy to see that our recommender system popped up two more acceptable restaurants that highly matches with the tag of the search query restaurant.

However, we still need to compare if the cuisines or dishes are kind of similar between the query restaurant and results. In Table 6, we clearly observe that our query restaurant is popular for the desserts like Nutella Crepe, Chocolate Almond Croissant, Tiramisu, which perfectly matches with its Yelp category tag. And our two recommended restaurants both have Tiramisu and some kind of chocolate cakes as popular dishes, in comparison with the popular dishes (Eggs benedict, Sticky bun, French toast) of Yelp recommended restaurants. With such a high consistency between our recommended restaurants and the query restaurant, we are confident that our restaurant recommendation method gives a better result than Yelp website.

When we tried different queries we found that with general foods such as pizza, burgers, fried chicken, etc., the results between our recommendations and Yelp's recommendations had many overlaps, which is interesting considering we are only using text data in our rankings. Number of stars are only used to break ties. For very specific dishes, our recommender system found the restaurants that had the highest positive mentions of that dish, but the top 5 Yelp restaurants recommended did not have some of these dishes in their reviews.

Figure 4 shows an example of what our top two recommended restaurants would be in Las Vegas for fried chicken. From the summary we can understand that Yardbird has more positive sentiment towards their fried chicken, while MTO Cafe has more positive sentiment towards their service. Depending on how much our user considers the importance of food or service, they can decide which best fits their criteria, without having to read a single Yelp review.

As shown in Table 7, our ability to search for relevant restaurants for specific terms is stronger than Yelp's. Given our limited and older dataset of Las Vegas, we were able to find those dishes in the restaurants that Yelp could not find. For the german soft pretzel query, the top Yelp recommended restaurant did not sell any pretzels.

In the case study analyzing the quality of our search, initially it seems odd that we would recommend a vegan restaurant for a chicken dish, as shown in Table 8, but it was the most popular dish listed for Veggie House, with more reviews/photos about their orange chicken than for Gourmet China. Our recommender system was able to find a restaurant where more people enjoyed the orange chicken dish than Yelp's.

Tables 9 and 10 showcase how our query can be extended to beyond food. Our results offer a greater variety of categories (French, Japanese), while Yelp's results were both American (New). Americana has a slightly higher average star rating, while Bouchon has 6.6 times the number of reviews. There were 16-17 reviews relating to Mother's day in each restaurant. However, Bouchon had more 4+ stars Mother's day reviews and less <=3 stars Mother's day reviews than Americana. It is clear that a majority of people enjoyed Mother's day at our top recommended restaurant, while there was mixed sentiment at Yelp's top recommended restaurant. Our ability to perform targeted sentiment analysis for Mother's day exceeded Yelp's search result.

4.2 Impact and Applicability

While comparing our recommendations to the recommendations found on Yelp's own website, it can be difficult to determine, or even say at all, that our results are superior to Yelp's. However, one important factor to consider is that our recommender system uses only the text data found in reviews in order to make recommendations, with a small bit of sorting by number of stars afterward - Such comparable results using only text data showcases the hidden potential of text data in general. With this point in mind, projects like this will hopefully show that text data is immensely useful and inspire others to utilize text data in ways that consist of more than a simple string searching function. Unfortunately, text data, such as those in Yelp reviews, are more meant for human consumption, and is considered unstructured data. Therefore, unlike structured and even semi-structured data, which can be sorted and stored on tables for ease of use, text data is more likely to be difficult to organize and work with.

In terms of general applicability, we feel as though such techniques employed here would be beneficial for those with review posting functionality. For example, using Targeted Sentiment Analysis on review text could be done in order to more accurately discover the grievances of consumers and allow companies to address these grievances without reserving human labor to reading reviews or even having to survey their consumer base. On the other hand, companies should allow for users to search for products using this technique in order for the users to find better products that cater to their wants and needs. Overall, analysis text data in this way could benefit both companies and consumers alike, giving companies ways to tackle consumer concerns with more ease, and allowing users to have more personalized recommendations which may increase their overall satisfaction of the company and its products.

4.3 Limitations

Our recommendations based on food are strong for specific dishes that Yelp is unable to find (i.e. bacon breakfast sandwich). The biggest limitation that we had is the size of our dataset. The academic dataset that Yelp provides is a subset. So there are many times that a direct comparison between Yelp's and our results is not a good one because there are restaurants that will be recommended on Yelp that do not exist in the subset that we have.

Another limitation comes from the way we tokenize reviews for targeted sentiment analysis. We assumed that each sentence in a review is either positive or negative. A simple example is "The orange chicken was good. But the customer service was bad." However, we do not split compound sentences such as "The orange chicken was good, but the customer service was bad." More often than not, we can find the aspects sentence by sentence, which allows us to numerically summarize reviews in ways that Yelp is unable to (with our "Number of mentions" and "Good Service").

4.4 Future Work

In the future we hope to use a method to separate compound sentences in order to refine our targeted sentiment analysis. Then those specific cases like "we liked this, but not that" would be separated into two tokens: "we liked this", "but not that". Also, for the restaurant query, we would like to try more advanced models such as Alternating Least Square (ALS) Matrix Factorization and latent Dirichlet allocation (LDA). For our current exploration, the performance using TF-IDF is better than Yelp only under some circumstances. We want to make a better algorithm (like LDA, ALS) to outperform Yelp search results under more general circumstances.

Currently the aspects we can summarize are the number of times the query was present and the number of times reviews have positive sentiment towards service. There are many more we could incorporate into the website such as waittime, cleanliness, atmosphere, etc. This would provide an even better summary of a restaurant that users can use to evaluate whether they want to go there.

Our recommender system takes advantage of the phrases with the most positive sentiment. In future work we could also explore how to use phrases with negative sentiment, so that users can filter out less desirable aspects (i.e. long wait time, slow service) or so that restaurant owners can have an overview of what aspect they can improve on.

References

[1] No author. *Screenshot taken from Yelp.* https://www.yelp.com/search?find$_desc = Bellagio + Patisseriefind_loc = Las + Vegas\%2C + NVns = 1, Mar.7, 2021$.

[2] C.J. Hutto and Elic Gilbert. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.* Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, January 2015.

[3] Jingbo Shang, Jialu Liu, Meng Jiang, Xiang Ren, Clare R Voss, and Jiawei Han. *Automated Phrase Mining from Massive Text Corpora*. accepted by IEEE Transactions on Knowledge and Data Engineering, February 2018.

[4] Àlex Escolà Nixon. *Building a movie content based recommender using tf-idf*. https://towardsdatascience.com/content-based-recommender-systems-28a1dbd858f5, Aug 28 2020.