

Chinese Restaurant Reviews Analysis

Yun Mo, Zifeng Wang, He Wang, Qintao Ying

STAT 628 Group 6

October 31, 2019

Outline

- 1 Data Preprocessing
- 2 All Restaurants Data Description
- 3 Chinese Restaurants Data Description
- 4 Word Frequency of Chinese Restaurant Reviews
- 5 Future work

Data Preprocessing

- ① Merge .json files based on "*business_id*"
- ② Extract the data which has "*Chinese*", "*Restaurant*" or "*Food*" in their catagories
- ③ Extract reviews of Chinese restaurants and clean the text

Text Cleaning

- Split text to words
- Remove punctuation
- Convert words to lower case
- Remove stop words(except "*not*", "*nor*", "*no*")

Example

Original Review:

'***No automatic doors, not baby friendly!*** I frequent this Chinese restaurant before and post partum. Dim sum is ok, dishes at dinner time are huge, service is great, BUT there are no automatic doors! How am I supposed to push my stroller inside?!!!'

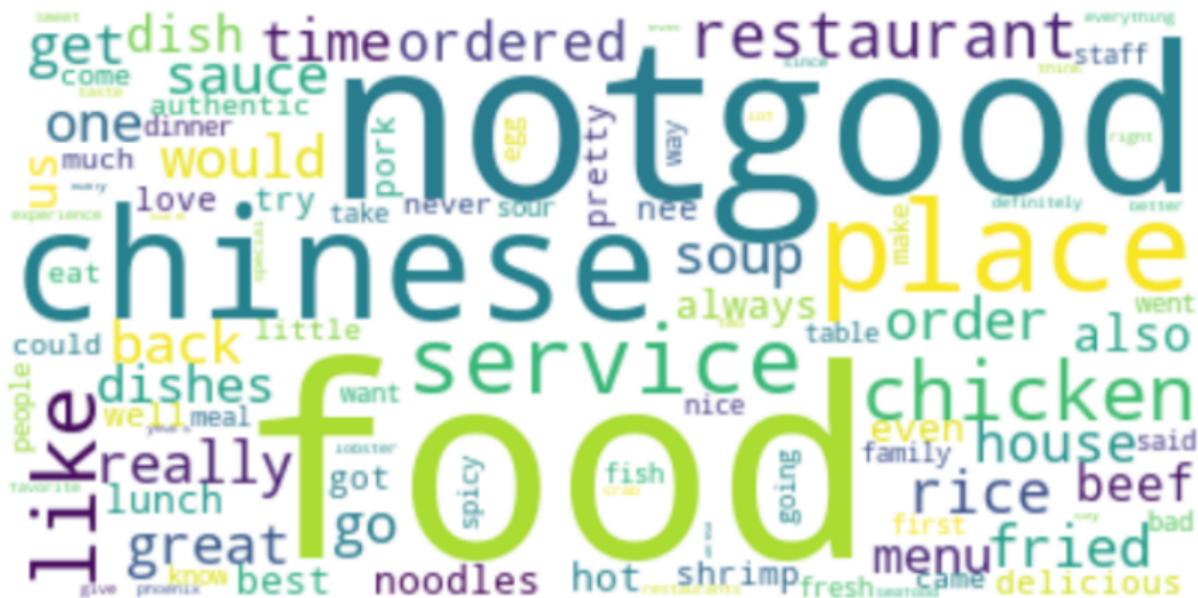


Cleaned Review:

'no automatic doors not baby friendly frequent chinese restaurant post partum dim sum ok dishes dinner time huge service great no automatic doors supposed push stroller inside'

- We didn't stem or lemmatize the words, because we want to easily extract words by tags(such as "*adjective*", "*nouns*")

WordCloud



Outline

- 1 Data Preprocessing
- 2 All Restaurants Data Description
- 3 Chinese Restaurants Data Description
- 4 Word Frequency of Chinese Restaurant Reviews
- 5 Future work

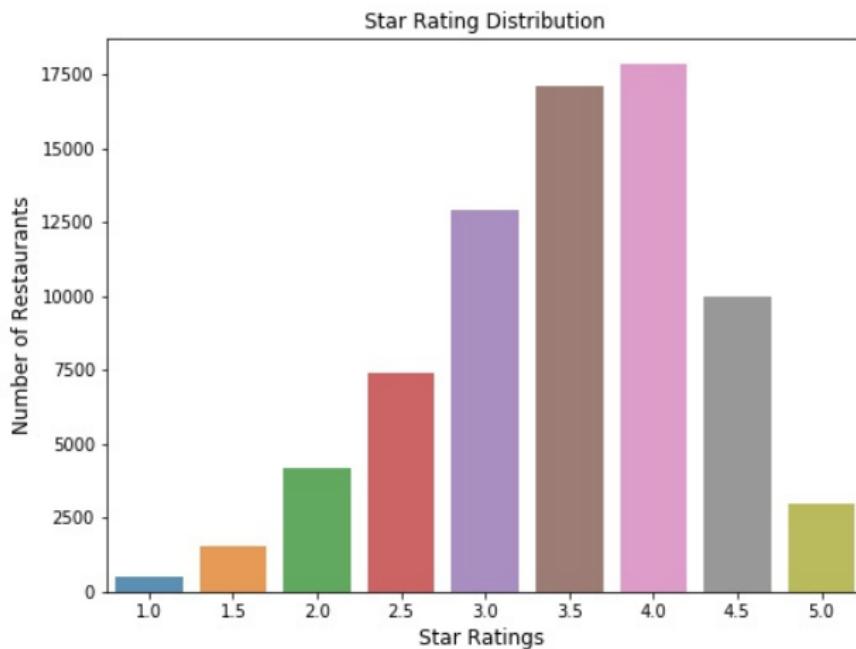
Restaurants Map

- Most of the restaurants are located in the United States and Canada.



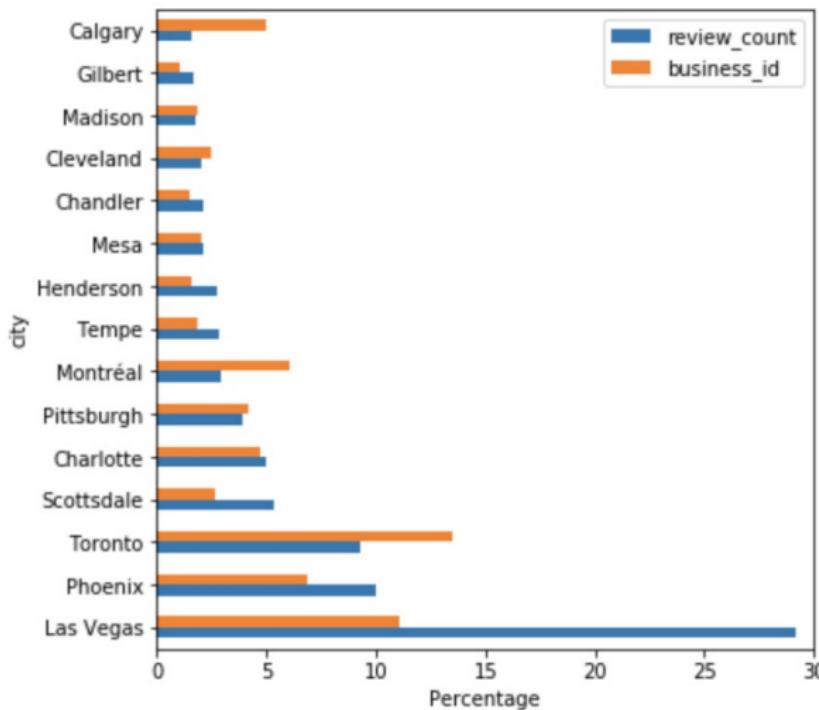
Rating Distribution

- 40 percent of restaurants' star ratings are ≥ 4 .



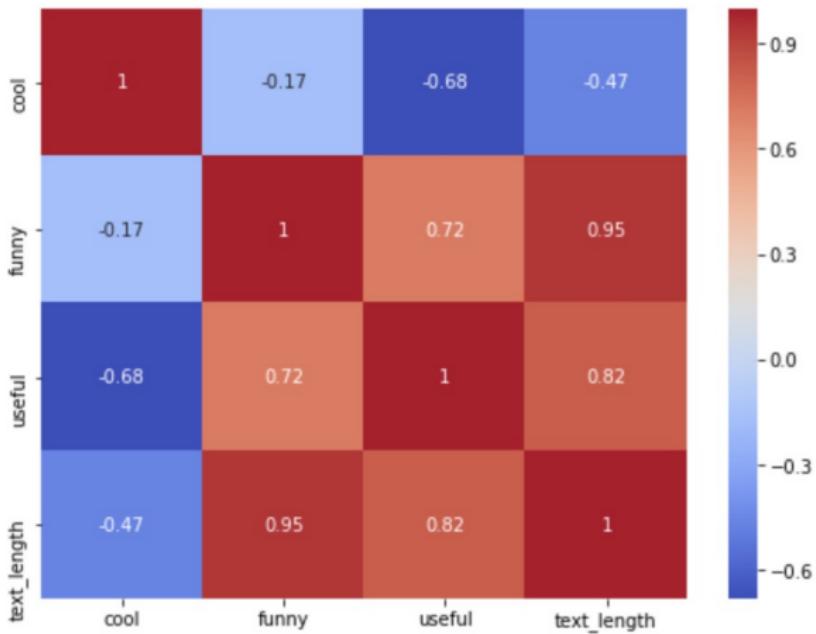
Top 15 Restaurants and Reviews Percentage

- Las Vegas has a large number of reviews.

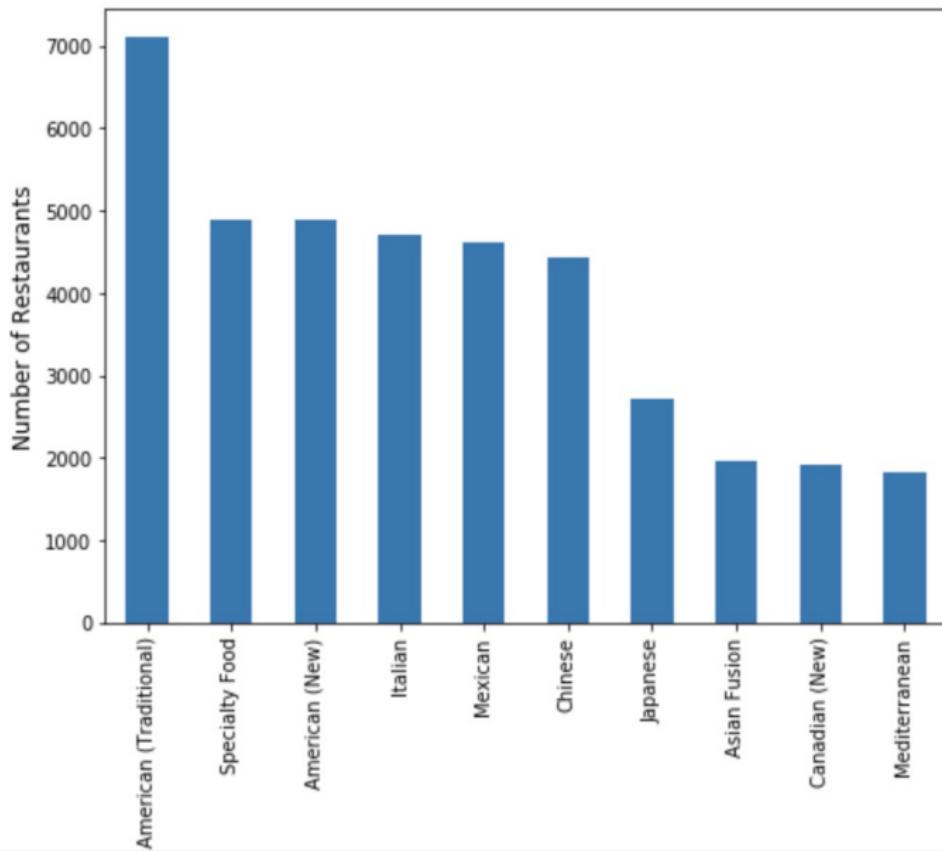


Text Correlation

- Funny and useful reviews have a positive correlation with text length.
- Cool reviews have a negative correlation with text length.



Cuisine Categories

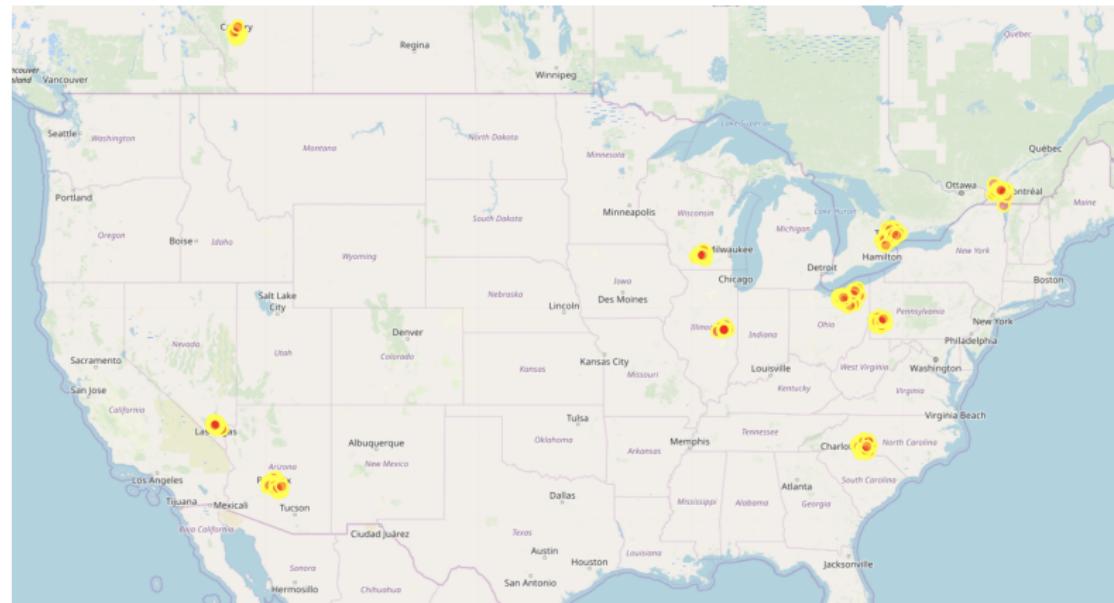


Outline

- 1 Data Preprocessing
- 2 All Restaurants Data Description
- 3 Chinese Restaurants Data Description
- 4 Word Frequency of Chinese Restaurant Reviews
- 5 Future work

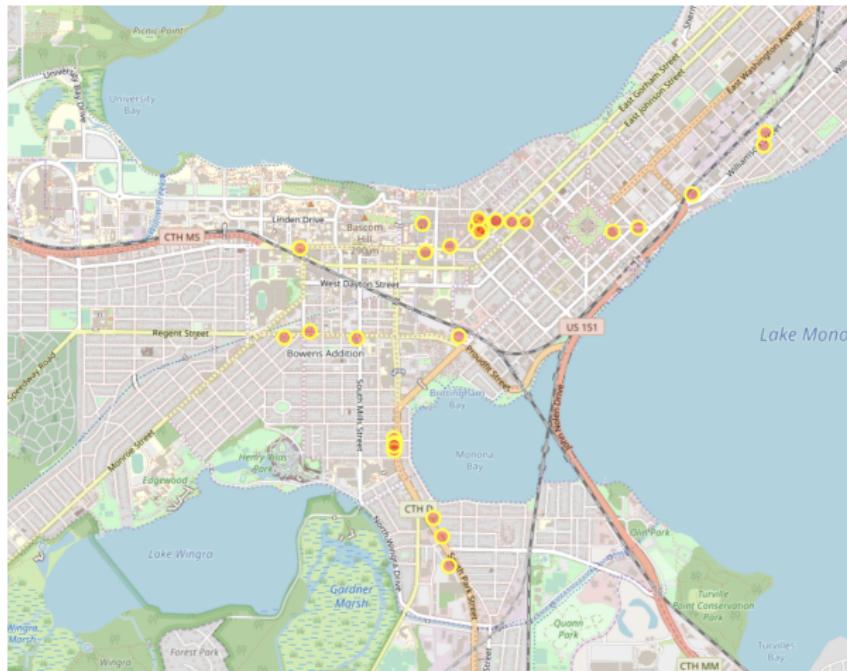
Chinese Restaurant Map

- Most of the Chinese restaurants are located around several big cities.



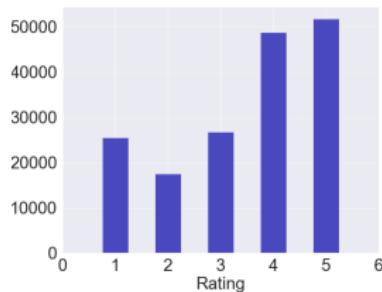
Chinese Restaurant around Madison

- You can take a closer look of the Chinese restaurants around Madison.

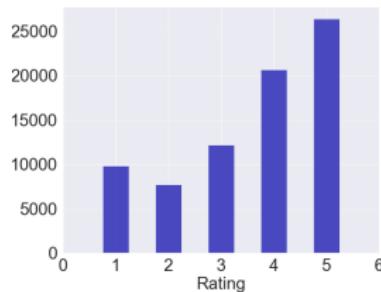


Restaurant with Different WiFi Type

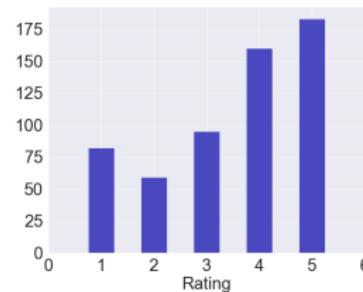
- These are three rating distributions of restaurants with different WiFi type
- Surprisingly, restaurants with no WiFi received most high-rating reviews.



(a) No WiFi



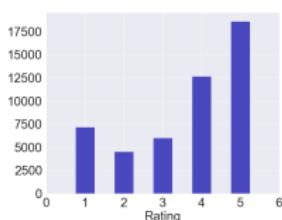
(b) Free WiFi



(c) Paid WiFi

Restaurant with Different Noise Level

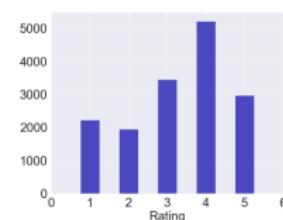
- These are four rating distributions of restaurants with different noise level.
- With the increasing of noise level, the more low-rating reviews that restaurants could receive.
- Here you can tell the Average noise level is the best one.



(d) Quiet



(e) Average



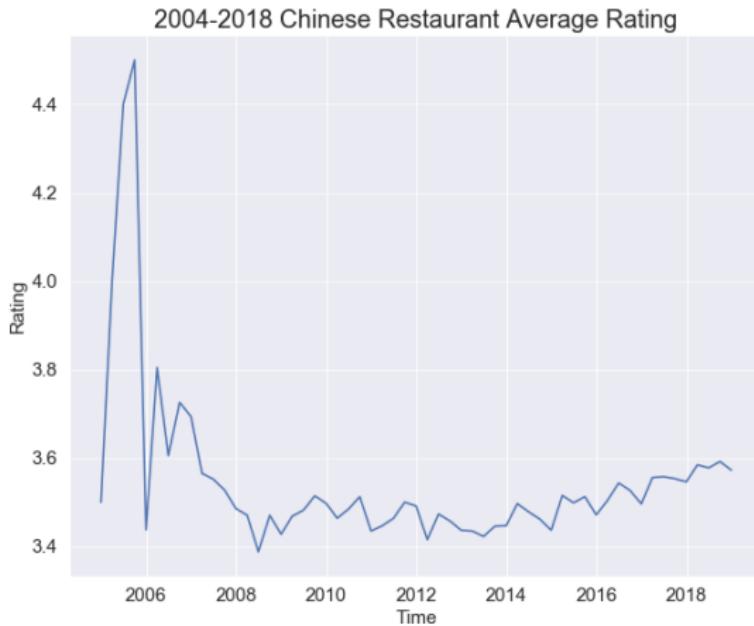
(f) Loud



(g) Very Loud

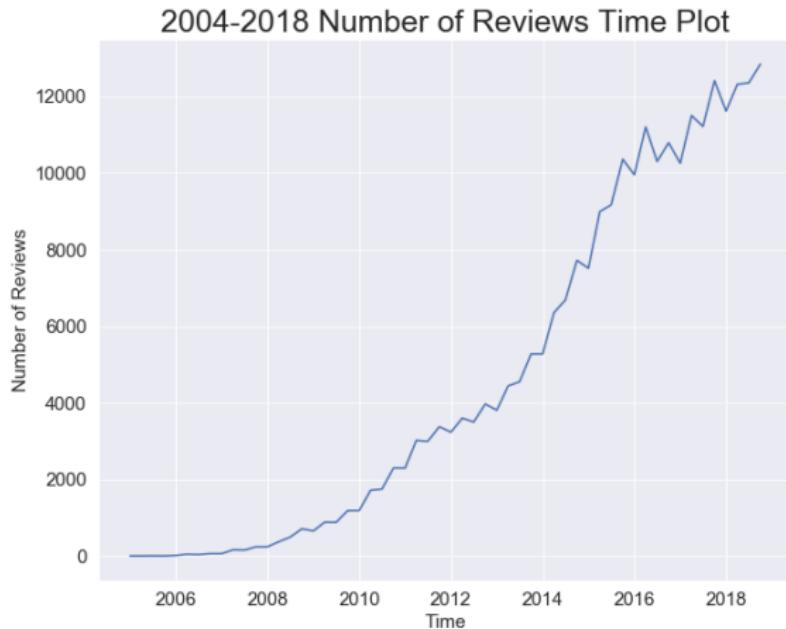
Rating Tendency

- The rating is getting more and more stable recent years.
- There is a small trend of increasing after 2015.



Number of Reviews

- The number of reviews is increasing greatly from approximately 0 to more than 12000.



Outline

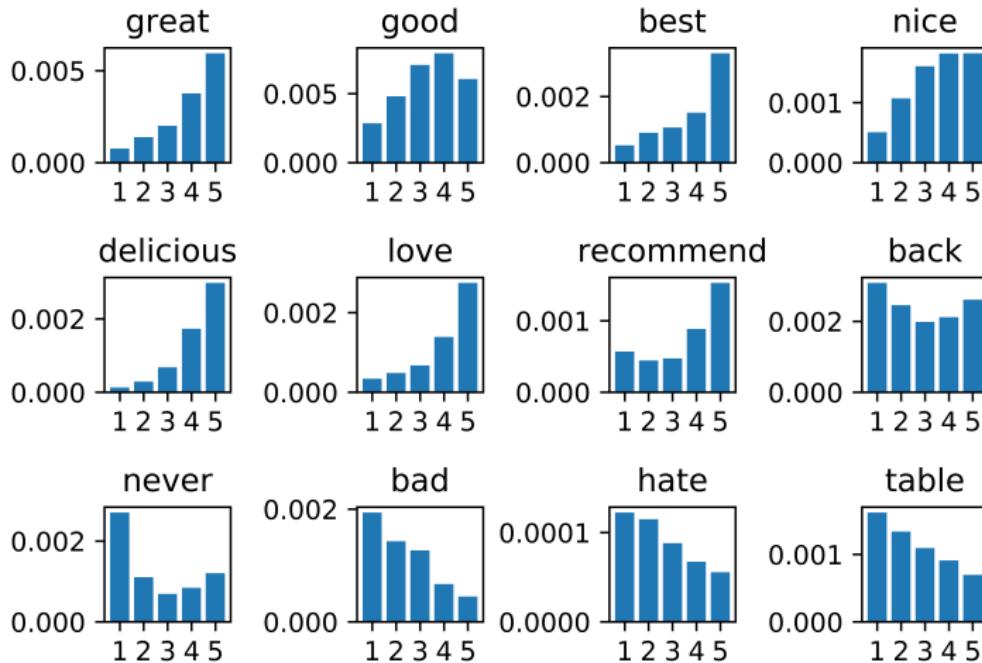
- 1 Data Preprocessing
- 2 All Restaurants Data Description
- 3 Chinese Restaurants Data Description
- 4 Word Frequency of Chinese Restaurant Reviews
- 5 Future work

Word Frequency – Food

- Positive words: beef, shrimp, egg, spicy, sushi, fresh

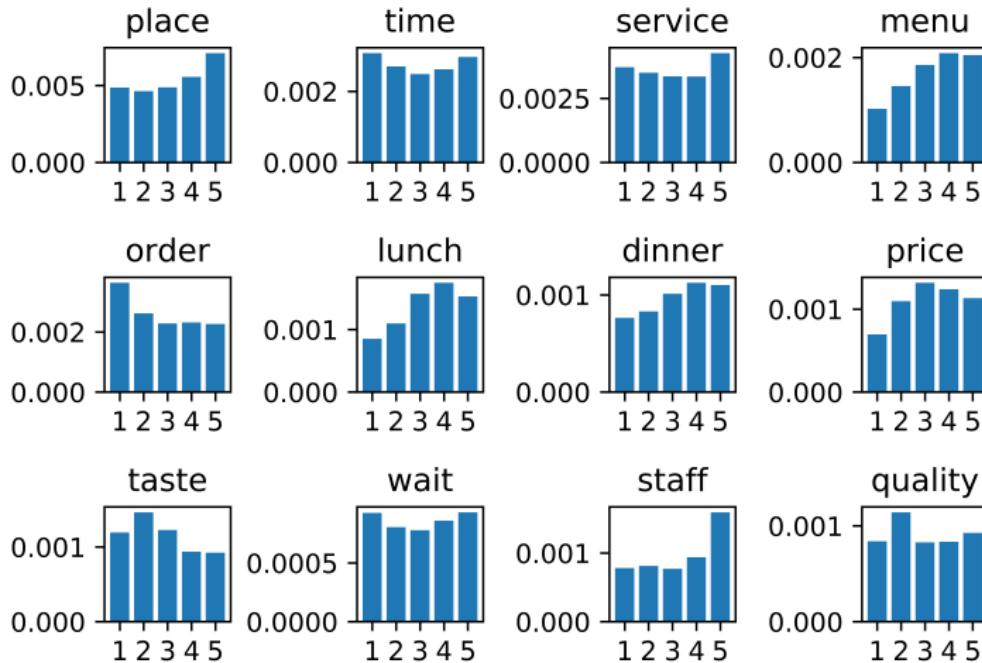


Word Frequency – Attitude

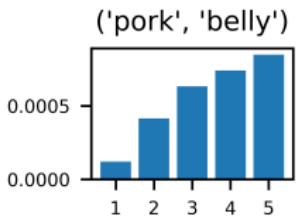
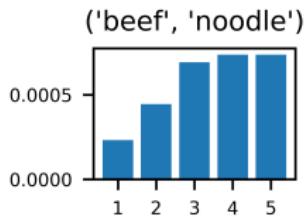
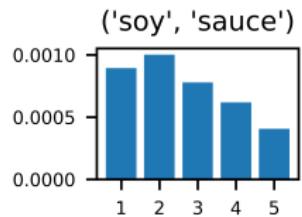
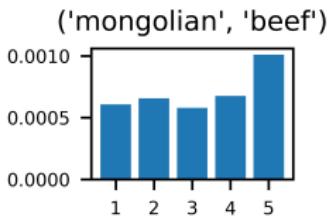
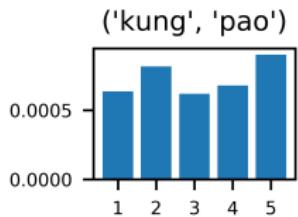
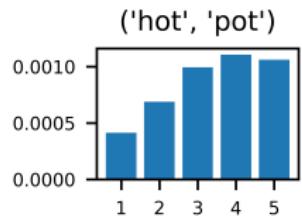
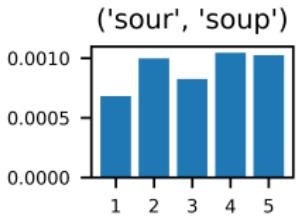
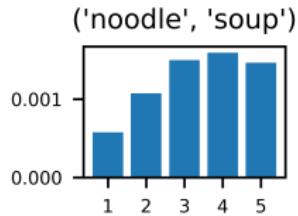
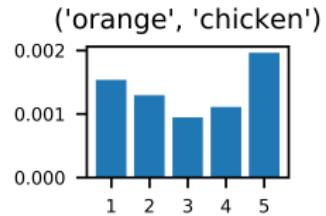


Word Frequency – Service

- Positive words: place, menu, staff
- Negative words: order, taste

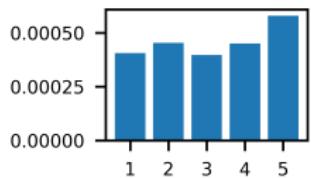


Phrase Frequency(composed of two words)

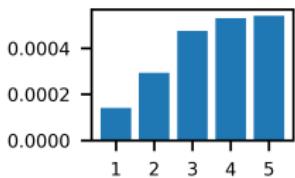


Phrase Frequency(composed of three words)

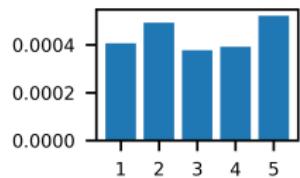
('egg', 'drop', 'soup')



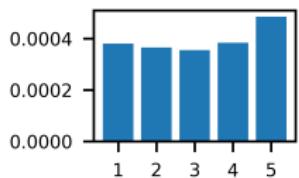
('beef', 'noodle', 'soup')



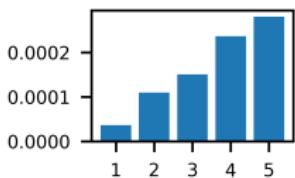
('kung', 'pao', 'chicken')



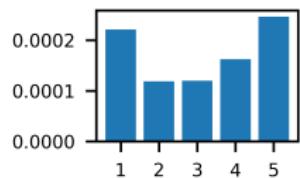
('chicken', 'fried', 'rice')



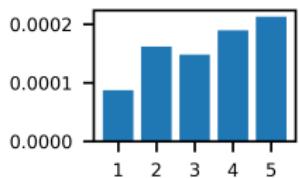
('hand', 'pulled', 'noodles')



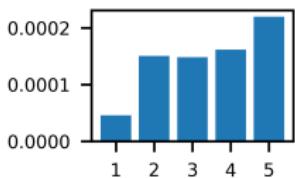
('egg', 'foo', 'young')



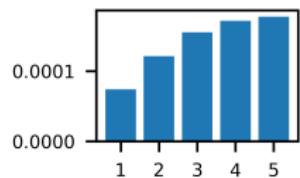
('hot', 'sour', 'soup')



('dan', 'dan', 'noodles')



('beef', 'chow', 'fun')



Outline

- 1 Data Preprocessing
- 2 All Restaurants Data Description
- 3 Chinese Restaurants Data Description
- 4 Word Frequency of Chinese Restaurant Reviews
- 5 Future work

Future work

- Extract useful words as features to overall describe different opinions of customers.
- Transform the selected words features to some numerical values by Word2Vec, GloVe and so on.
- Use various classifiers such as Logistic Regression and SVM to analyze which types of words affect rating.
- Find the connection between rating and restaurant categories.

Thanks!