

Ratings Prediction on Yelp Data

Presentation I

Yilun Chen

Jiashun Cheng

Yishan Cai

Contents

- 1. Basic Sentiment Analysis
- 2. Features Exploration
- 3. Preliminary Analysis
- 4. Future Work

Basic Sentiment Analysis

1.Sentiment Analysis

2.Lexicons

- afinn: scores range from -5 to 5
- Bing: positive or negative
- NRC: categories by type of sentiment

3.How to Summarize

How to summarize

afinn

- **mean, median**

bing

- **(negative , positive)**

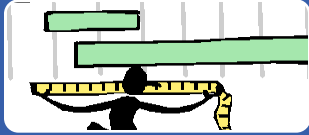
```
: num [1:2] 7 8
```

nrc

- **(anger, anticipation, disgust, fear, joy, negative, positive, sadness, surprise, trust)**

```
: num [1:10] 1 0 1 0 3 2 4 1 0 3
```

Features Exploration



Number of words



Number of third



Number of exclamation



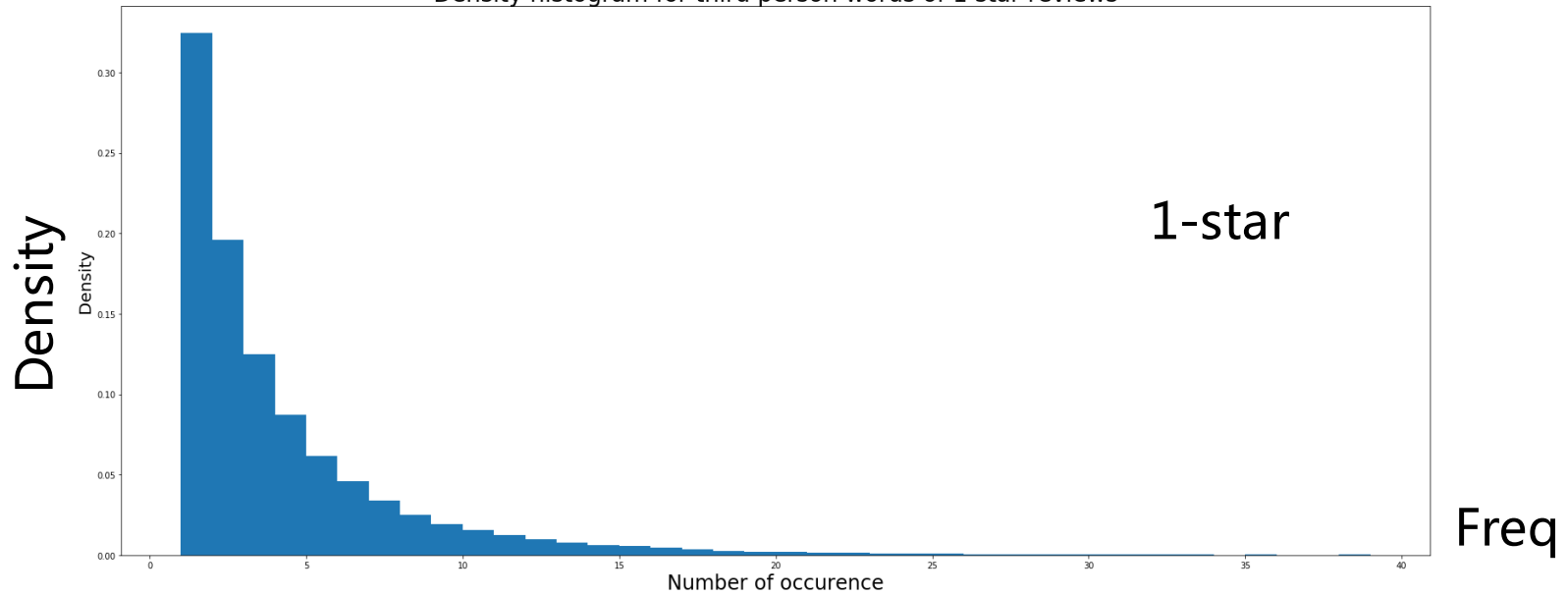
Number of question mark



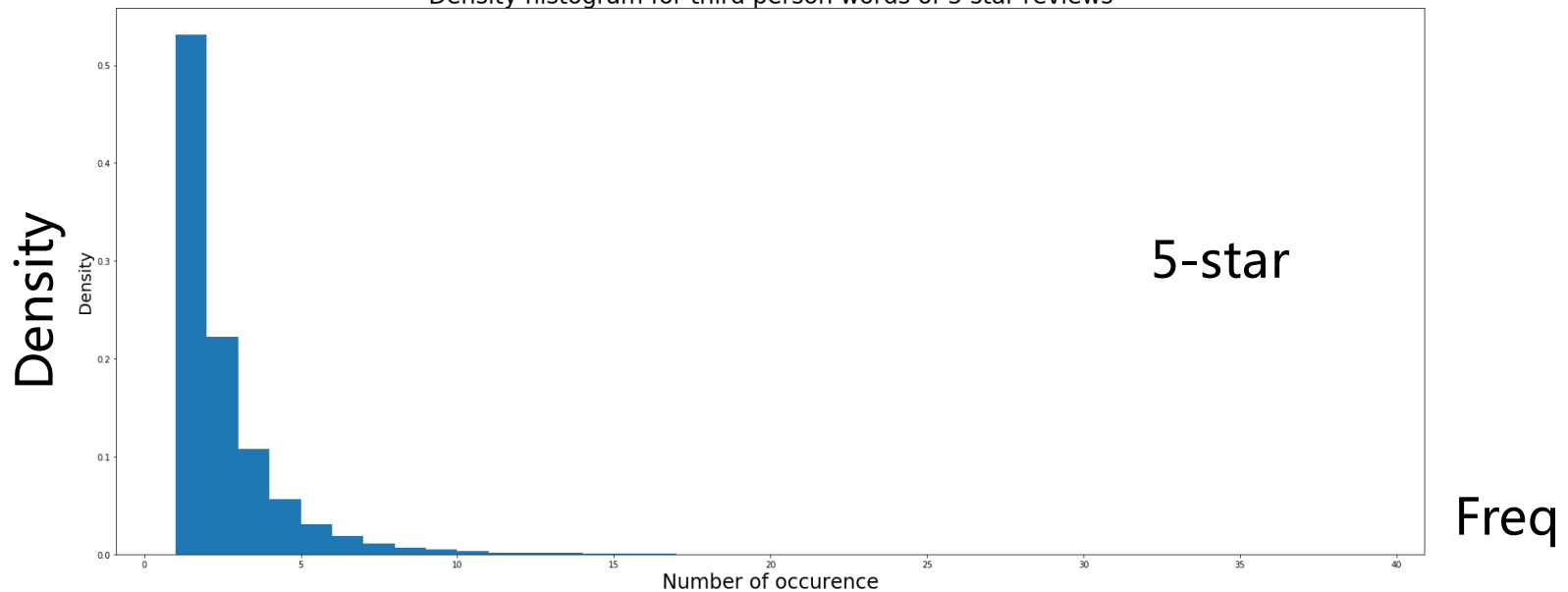
Average Score for restaurants/categories

Density Histogram-Third person words

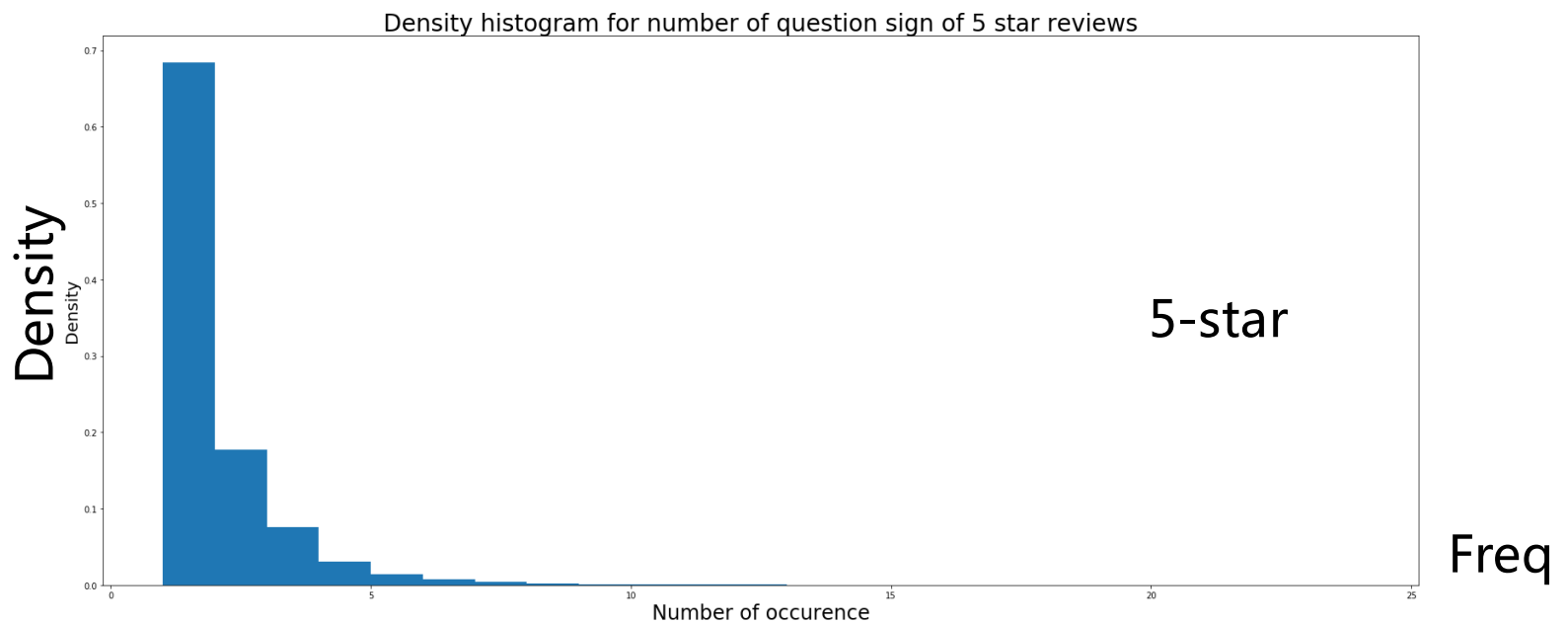
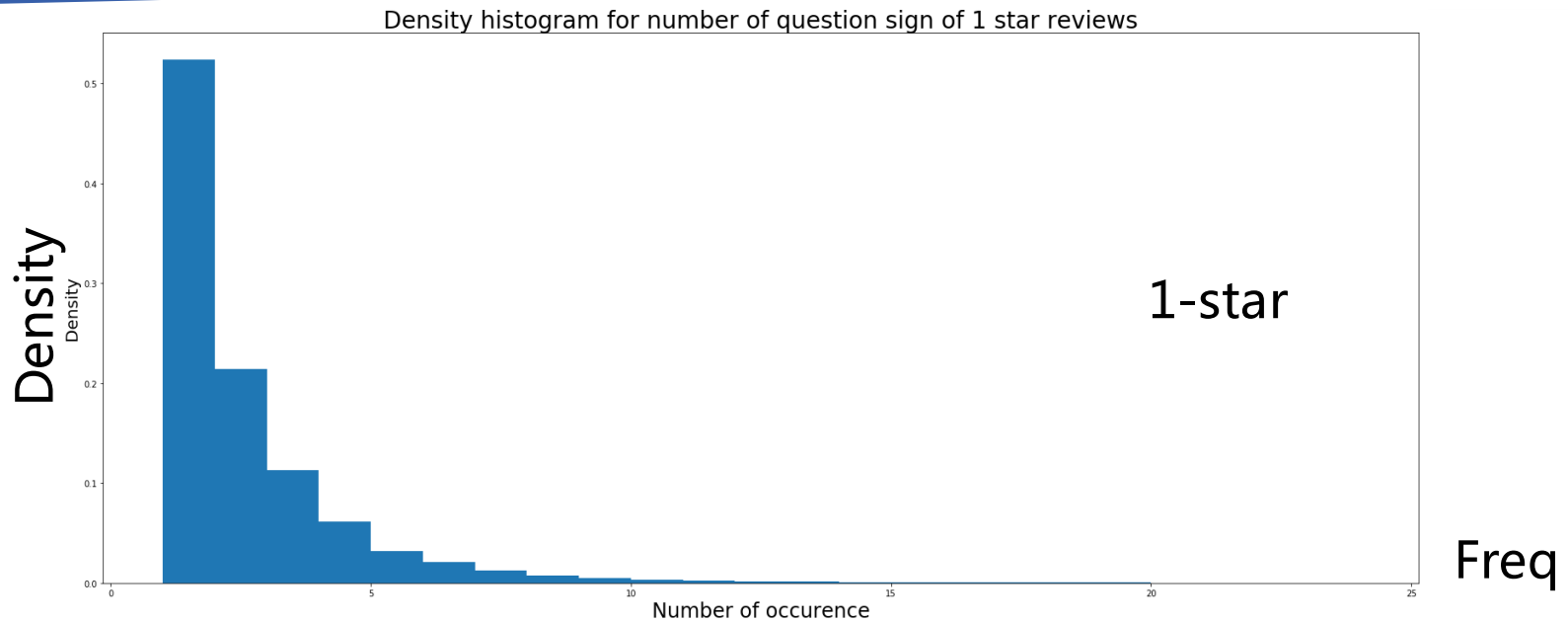
Density histogram for third person words of 1 star reviews



Density histogram for third person words of 5 star reviews



Density Histogram-Question mark



Preliminary Analysis

Clean

- Stop word, Punctuation
- Lemmatisation, Lowercase

Transfer

- Bag of word
- Wordvector

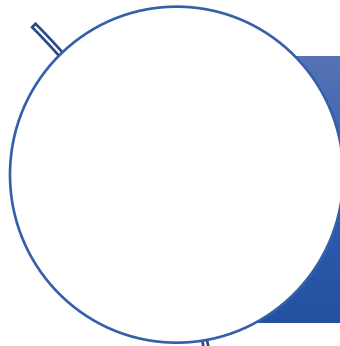
Fit

- SVM, LM, Logistic
- CNN, GBM

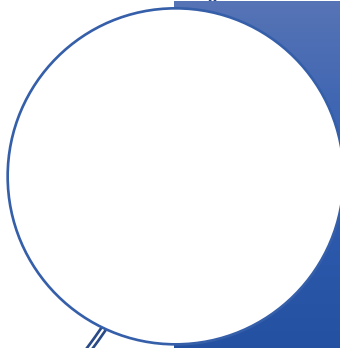
Preliminary Analysis



Future Work



Combine basic attributes information with text mining



Extract Features

- Prior knowledge
- Current discoveries

THANKS