ENTREERECC

An Automatic Entree Recommending Technique Based on Yelp Review
YELP REVIEWS

- Restaurants, Places of Interest etc.
- Star Rating
- Reviews
• However, reading reviews may be cumbersome
• Too many reviews, too little time
• How can we know what item is popular?
• Yelp single out similar reviews and highlight the word
• Crispy Tofu is a nice entree
• What about Mandarin Work and Mahomet?
• Are these reviews positive?
• Yelp does not differentiate entree names from others

• Yelp probably does not use sentiment analysis

• We need a better entree recommending system
INTRODUCING ENTREEREC

• An automatic technique to recommend popular entrees based on existing Yelp reviews

• Define popular entrees

• Need a classifier and sentiment analysis
SYSTEM OVERVIEW

• Extract noun phrases from the review
• Train the Naive Bayes Classifier
• Identify the reviews that are talking about entrees
• Sentiment analysis on those reviews
NOUN PHRASE EXTRACTION

- Illinois Chunker

- Sample output:

- Use python to process the output data and extract the noun phrases (NP’s)
• Incorrect use of grammar:

"The crispy tofu is a must even if you do like tofu it’s so good the pan fried noodles yes yes yes and don’t get me started about the duck."

[NP (DT the) (VB pan) (JJ fried) (NNS noodles) (UH yes) (UH yes) (UH yes) (CC and) ]

• Irregular use of punctuation marks:

// Sesame ball - these are stuffed with red bean paste. I can’t tell if they are a dessert or an appetizer. Either way it is delicious

// Hot and Sour soup and Egg drop Soup - the MSG is to die for... haha I think i like the Egg drop soup a little more. The consistency is soo perfect. "

[NP (NN //) (NNP Crispy) (NN tofu) ]
• Stop words

Crispy Tofu and Garlic Beef

the Sesame Balls

Crispy Beef with Garlic; Sweet and Sour Chicken

• Might not want to remove all stop words

• Removal of ‘the’ and ‘their’
CLASSIFICATION

- Open source Node.js classifier on npm
  https://www.npmjs.org/package/classifier
- Two labels: *food* and *not*
- Training data
SENTIMENT ANALYSIS

- Stanford CoreNLP
- Deep Learning for Sentiment Analysis
  
  http://nlp.stanford.edu/sentiment/

  The website includes a online demo of sentiment analysis

- Negative, neutral or positive
• Read output from Illinois Chunker

  In JSON format

  Each sentence is associated with noun phrases

• Use npm Classifier for noun phrases

• Sentiment analysis for sentence with food noun phrases

• If sentiment positive, increase counter for the noun phrases (entree name)

• Output highest ranked entree names
EVALUATION

• Recommended popular entrees for Golden Harbor
• Performance is relatively slow
• Will be performed offline
FUTURE WORK

• Integrate EntreeRec into an IR system
• Crawl reviews instead of using downloaded data
• Shallow parsing informal text data
THANK YOU!