

Top Dishes @ Zomato

Team 17 - madNLP:

Amandeep Kaur amandeek@usc.edu USC ID: 5529301286

Dhruv Bhatia dhruvbha@usc.edu USC ID: 3567501596

Manish Dwibedy dwibedy@usc.edu USC ID: 7848467069

Minaxi Singla msingla@usc.edu USC ID: 4676995181

Introduction:

Zomato is restaurant search and discovery website, which features restaurant information, reviews, recommendations, star ratings that can help the user to make a decision. A customer can read a review that recommends a delicious dish, but it is difficult to map it to a particular menu item offered by the restaurant.

Furthermore, colloquial language may be used making it even more difficult to make a match. Overall rating of the restaurant is definitely useful, but it is also very useful to have per-item ratings.

Goal:

Our goal is to create a model to identify food mentions in the reviews and match them up with the respective menu items. Then, using sentiment analysis, we will find a “score” for each menu item mentioned in a review. We plan to enhance previous work done in this area by adding the following functionalities:

1. Discover location-based trends in the data.
2. Analyze as how does the restaurant category affects rating and price.
3. Analyze the effect of the user’s foodie-level on the score of the menu item.

Why is the idea interesting?

This idea is taking recommendation engine to the level of food search, rather than keeping it at the level of restaurant search. We foresee various applications of our work. To name a few – finding favorite menu items, recommendations of popular menu items, search by menu item, location based searches etc.

Related Work:

We draw our inspiration from the project – “**Product Features and Opinions**” [1] which focused on extracting product features (e.g. ‘size’, or ‘speed’) and opinions (e.g. ‘great’, or ‘horrible’) from the reviews. We plan to extend this work in various categories as mentioned in the **Goal** section.

Data:

We will be extracting the **English** data using Zomato APIs [4]. We will then be using **TextBlob** [9] python library to correct the misspelled words. Finally, **Microsoft Translation API** [5] will be used to translate data to **Hindi**. In order to perform successful sentiment analysis [2], data is to be annotated and this is the reason for choosing “Hindi” whose intricacies are very well known to us.

Implementation:

1. Extracting food mentions from the reviews: We plan to implement the concept of Named-Entity-Recognition [3] and Part of Speech Tagging to

- extract food mentions from the reviews. We will be using one of the techniques – **HMM** [8], **CRF** [7] or **Phonetic Matcher** to perform this task.
2. Matching food mentions with menu items: We plan to use various matching algorithms – **ExactMatch**, **PartialMatch**, **SubstringMatch**, **FuzzyMatch**, **SVMMatch** [6]
 3. Sentiment Analysis: After matching a food mention to a menu item, we plan to annotate the data and separate it into train, test, and development sets. We will develop an annotation tool to figure out the sentiment and determine if a dish is positively, neutrally, or negatively reviewed.

Evaluation:

We plan to evaluate each matcher mentioned above, examine the errors (ambiguous menu listings, incomplete Zomato data), and the resulting **Precision**, **Recall** and **F1** score. We will tune the models on development data. Training will be done on the training set. Testing data will be used to evaluate our score at the end. We will evaluate the whole system against a baseline system developed using well-established external tools and libraries like **TextBlob**, **CoreNLP** etc. running on data in 'English' language.

References:

- [1] Ana-Maria Popescu and Oren Etzioni, "Extracting Product Features and Opinions from Reviews", Proc. HLT/EMNLP 2005, 339-346.
- [2] Piyush Arora, "Sentiment Analysis for Hindi Language", M.S. thesis, IIIT, Hyderabad, India, 2013.
- [3] A. Nayan et al., "Entity Recognition for Indian Languages", Proc. IJCNLP-08 Workshop on NER for South and South East Asian Languages 2008, 97-104.
- [4] Zomato, [API], Available: <https://developers.zomato.com/documentation>
- [5] Microsoft, [Translator API], Available: <https://www.microsoft.com/en-us/translator/translatorapi.aspx>
- [6] Adam Cohen (2011, July 8), "Matching Algorithms" [Blog], Available: <http://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python>
- [7] A. Ekbal et al., "Language Independent Named Entity Recognition in Indian Languages", Proc. IJCNLP-08 Workshop on NER for South and South East Asian Languages 2008, 33-40.
- [8] Sudha Morwal and Deepti Chopra, "NERHMM: A Tool for Named Entity Recognition Based on Hidden Markov Model", IJNLC Vol. 2, No.2, April 2013, 43-49.
- [9] S. Loria, [TextBlob Library], Available: <https://textblob.readthedocs.org/en/dev/>

Division of Work:

Team-Member	Module
Dhruv & Minaxi	Documentation (Proposal/Report)
Manish	Data Collection
Amandeep, Dhruv, Manish, Minaxi	Data Annotation
Amandeep & Manish	Setting up the Baseline System
Dhruv & Minaxi	Named-Entity Recognition
Amandeep, Dhruv, Manish, Minaxi	Matching Algorithms
Amandeep	Sentiment Analysis

Word Count: 615