Sentiment Analysis and Rating Prediction on Hotel Reviews

Milestone - 1
Raj Gandhi
Advisor: Dr. Rajendra Raj
Topics to cover

- Summary of Project
- Literature study - 1
- Literature study - 2
- Literature study - 3
- Data collection, cleaning and Preparation
- Next steps (Milestone 2)
Summary of the Project

- Project focuses on doing sentiment analysis on the hotel reviews.
- Getting the customer’s interest
- Implementing Random Forest classifier for sentiment analysis
- Creating a web based UI for visualizing recommendation
Literature Study - 1

Paper: Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naive Bayes

- This paper focuses on sentiment analysis of trip advisor
- Data collection: Web crawler
- Classifier: Naive Bayes
- Accuracy: 72%

Relevance to the project: Understanding the sentiment analysis methodology and comparing the classifier accuracy with proposed one.

Literature Study - 2

Paper: Sentiment analysis using product review data

- Goal of the paper: Sentiment polarity categorization
- Data collection: Reviews data from Amazon.com
- Classifier used: Naive Bayes, SVM and Random Forest
- Better performance by Random Forest

Paper: Product Recommendation using Sentiment Analysis of Reviews: A Random Forest Approach

- Goal of the paper: Sentiment polarity categorization
- Data Collection: Kaggle (product reviews)
- Classifier: Multinomial Naive Bayes, LinearSVC, Random Forest
- Best Accuracy by Random Forest

Data Collection, Cleaning and Preparation

- Data has been collected from Kaggle
- Dataset having 1000 hotels and their reviews
- Important attributes: location, name, review data, rating, title
- Data Cleaning and Preparation:
  - Removing empty/missing values
  - Detecting languages other than English
  - Cleaning unwanted characters
Next steps (Milestone - 2)

- Using nltk python library, train and test the Random forest classifier
- Document results
- Design database schema
- Web UI for data visualization
Thank you