Sentiment Analysis and Rating Prediction on Hotel Reviews

Raj Gandhi
Advisor: Dr. Rajendra Raj
Motivation

• Reviews play an important role whenever someone wants to search a hotel. Users greatly rely on these reviews and these reviews may be helpful if they are more relevant.

• Today recommendation systems on Hotels rely largely on content-based approach while proposed system rely on mood-based approach which require sentiment analysis.

• Consider a scenario where the user searches for a hotel and gets recommendations based on ratings. However, these reviews could be based on one of the factors, like location while the user wants a hotel who provides great food.

• Users would be able to take more informed decision if they get recommendation based on their interests.
Background and Related Work

• In the current systems like TripAdvisor, Hotels.com, etc. we have content-based recommendation systems and there is no way to get recommendations based on mood of the customers.

• Getting the Sentiment Score from the reviews and then recommending the Hotels based on customer’s criteria could help the end user with better decision.

• As per the two researches[1] [2] done on Hotel Reviews, Sentiment Analysis uses Naïve Bayes and SVM approaches which produces upto 75% of accuracy.

• However, another two researches[3][4] done on another dataset (movies & product reviews) for sentiment analysis uses SVM, Naïve Bayes and Random Forest training model, of which Random Forest provides more than 91% of accuracy.

References:
Proposed Work

• Do the data exploration to learn more about the data and its important attributes.

• Data cleaning & preparation and load the data to the database.

• Define popular categories like food, business, and try to extract these categories from the reviews using NLP. Then the reviews will be fed to a function of Sentiment Analysis which will help understand the mood of the review. If the review leans towards positivity this hotel will get a reward in that category (else penalty in that category for negative reviews). For e.g. There is a positive review of food on Hotel California, then it will receive higher points on food and may climb up in search results in food category. Geo-Location of hotel would also impact this search results.

• Create a web-based UI for users to see recommendations based on interest.
Results and Evaluations

• Hotel recommendations are based on Sentiment Analysis which will use Random Forest for training the model.

• Results will be analyzed on the accuracy of the model and metrics used will be:
  • Accuracy
  • Precision
  • Recall
  • F-score

Results from the above metrics will be compared to the existing model of Naïve Bayes and SVM.
Milestones

• Milestone – 1: Data exploration, Data cleaning and preparation and define popular categories like Food, Landscape, Business and extract this features from reviews

• Milestone – 2: Implement Sentiment Analysis using Random Forest Model and create a web UI for the application.

• Milestone – 3: Results evaluation, Documentation and Future work.
Thank You