Report

Group 50
(Hsu, Ya Cheng Zhao, Shu jing Yang, Mingyu)

Task

1. Classification: Given a list of ingredients of a recipe, output type of cuisine;
2. Prediction: Given a partial recipe, predicting remaining ingredients may included;
3. Metrics: Explore relevant metrics to evaluate performance;
4. Experiments: Compare possible methods to appropriate baselines.

Roles of the Teammates

Hsu, Ya Cheng (s2139419) - coding & literature searches
Mingyu Yang (s2117294) - coding & algorithm study
Shujing Zhao(s2037249) -coding & report

Data Preprocessing:

The raw data listed out ingredients of cuisines. Intuitively, we can treat each ingredient as a boolean feature for a cuisine. That is, if a given cuisine has a particular ingredient i, the feature i would have a value of 1, and otherwise 0. By converting the original dataset, now we have a

Methods:

Clustering/ Classification

We plan to try different non-linear clustering models for this problem, like SVM (with Gaussian kernel), Naive Beyesian and Neural Network.And because of the large and sparse feature matrix we had, we may need some dimensionality reduction method such as PCA and MDS.

Collaborative Filtering

We plan to use model-based Collaborative Filtering for this problem. And we would like to research and try different similarity methods here, like Jaccard Similarity, Spearman rank correlation and Cosine Similarity.

Evaluation:

For the classification part, we will compare our model with linear classification models, like logistic regression and SVM (with linear kernel).

For the CF part, we will compare our model with user-based and item-based CF models.

Metrics:

For classification- precision, recall, accuracy, F1 score;

For prediction-accuracy.