

EE 525
Recommendation Systems
Yazan Okasha and Souparni Agnihotri
Project Proposal

1. Summary

In this project, we are going to come up with a collaborative based filtering method in order to make better recommender system for our website - potluck.cafe. We have already collected the data from the users on this website. Using this data we can perform several data analytics techniques like k-means clustering to group similar users together and then make recommendations to the other users based on the same data.

2. Introduction

As a part of our research, we have been working towards building a fast, efficient and reliable recommender system. At the core of this research is our website - Potluck that obtains different food recipes from the Spoonacular API and queries the users on the recipes they may or may not like. Working with this data, we intend to come up with a collaborative filtering method, where we split the users into groups based on the recipes they have rated. To do this, we are going to try out different clustering measures (k-means clustering, spectral clustering, hierarchical clustering and top-k grouping). We will determine which algorithm is the most robust by computing cosine or jaccard similarity on a newly classified user. The higher the similarity measure, the better the algorithm on our given data.

In addition, we have already performed content based filtering on the recipe database and we are able to determine what the similar recipes are based on a given recipe. We can even perform an analysis on how different the recommendations are between content based and collaborative clustering

3. Project outline

3.1. Preprocessing

Once the data has been acquired, we need to do some preprocessing on the data. For example, if the data is too sparse, we will need to come up with a way to gather the data such that we will have a more densely filled user-recipe ratings matrix that would make our future computations more accurate.

3.2. Technical Specifications

The core of our project will involve building a collaborative based filtering model. That is, for each user, the recommender system will recommend items based on how similar users rated the item. We will be using a series of clustering measures:

- K-means clustering: We can consider an approach where we partition the users into k-clusters such that the users in each cluster share similar interests. Using these clusters, we will be able to effectively determine who the similar users are, and their ratings and make good recommendations.
- Spectral clustering: This is a different type of clustering which will work well on highly dense data where clusters are not immediately evident. We will perform the same clustering of users using this algorithm.
- Hierarchical clustering: Providing a broader way of clustering where we do not have to specify a value “k” for the number of clusters we want, this type of clustering seems to work well and gives us flexibility to determine how many clusters are ideal for our data.

4. Testing:

We will first determine which of the clustering measures were the most effective by computing the cosine and jaccard similarities on the recipes recommended by each of the clustering measures. Depending on the measure that gives us the highest similarity rating, we will choose which similarity measure to proceed with.

5. Analysis:

In this section we will evaluate the pros and cons of using each clustering metric and why one is better than the other. We may represent our analysis in terms of accuracy, confusion matrix and F-measure.