Music Recommendation System

Ceng History X - Iteration Report 1

Team Members:

1) Asena Ok, 1746296, asenaok@gmail.com
2) Ayşe Aybüke Taşdirek, 1746353, aybuketasdirek@gmail.com
3) Birant Altınel, 1745744, birantaltinel@gmail.com
4) Hacer Nihal Tarkan, 1744291, tarkan.nihal@gmail.com

Tasks for Iteration 1:

Planning Tasks	Responsible Person	Document vs. Attachments	Implementation	Is completed?
Improving GUI	Hacer Nihal Tarkan	Added to SVN	Yes	Under Construction
Accelerating Current Recommendation Algorithm	Aybüke Taşdirek	Documented	Yes	Yes. Committed in SVN
Improving Current Collaborating Filtering Algorithm	Asena Ok			
Separating and Storing Data for Evaluation Module	Birant Altınel	Added to SVN	Yes	Yes.
Implementing Evaluation Module	Aybüke Taşdirek - Asena Ok	Documented	Yes	Partially completed. Committed in SVN
Research on Tag Based Recommendation	Hacer Nihal Tarkan	A survey given to Instructor.	No	Yes.
Extracting the Algorithm of Evaluation Module	Asena Ok		No	No. Committed in SVN

1. Accelerating Current Recommendation Algorithm

After improving the current collaborative filtering algorithm we can get recommendation faster. The improving one is committed in SVN. Please see details below:

- Old algorithm:
  - Searching in database every time
  - Responds in very long time (about 2 minute)
- New algorithm:
  - Usage of existing indexing
  - Responds in 3-5 seconds

2. Separating and Storing Data for Evaluation Module

In order to make an Evaluation on the current recommender system, we decided to split the data into two different parts, so that we could use one of the parts to train the recommender and use the second part to evaluate the outputs of the recommender. In order to achieve this, we have created a second database which is a much more reduced version of the original data structure that we used in the primary database. It only holds the essential information that is required to generate information to be used in evaluation. There are two reasons, in our opinion, that a secondary more basic database was useful. Firstly, using a simpler data structure dramatically reduces the capacity of the additional data from roughly 30 GB to a size of about 10 GB. Secondly, using less data makes is possible to traverse and mine the database faster, and produce the information quicker.

The reduced data structure includes only two different type of nodes: User and Song. The album and performer information is only stored as id numbers as an attribute to the Song nodes.

3. Implementation Evaluation Module

We evaluate the recommendations on training database. It includes %90 of total data that we get from ARGE-DOR. We split the data according to time. The recommendations are compared with the songs that the users've listened after a specific time. To indicate accuracy of the current algorithm we just check the name of the songs. We've achieved a consistency about ⅙ percent. After

improving the current algorithm by using other approaches like content-based recommendation this ratio will be improved. Please see the project 'Data_Import_End of Iteration_1' implementation of evaluation module

4. Research on Tag Based Recommendation

We are trying different approaches to make a recommendation in terms of speed and accuracy. We accelerated the current collaborative filtering algorithm. That is, we did something for the speed of our system. We also wanted to improve accuracy of our system. Then, we think that we can use semantic approach to tag songs. Tag is a short description of a song. We searched for tag based recommendation methods. There are two main kind of tag based recommendation technique. These are social tag based ones and auto taggers. The first one is done by user generated social tags. And, the other one is done by applying signal processing on audios. However, we do not have tag data and audios. Therefore, the common approaches are not appropriate for our system. We also thought that we can tag songs according to their names. But, the name data is over misspelled. Moreover, song names are not in one language. Therefore, tag based recommendation is not appropriate approach for our system under these conditions. You can see more information from the paper that I provided in the attachment.

5. Improving GUI

The GUI we created was poor in terms of showing results. Also, we need to add some tables to show evaluation results and make it easy to compare our recommendations and evaluations. We also beautified our GUI.

Iteration 2 Planning

During iteration 2 our only focus will be implementing a content-based approach to the current algorithm. We are planning to complete implementation of content-based algorithm and achieve a hybrid recommendation system.