Collaborative filtering

In relation to getting data for the project, it is needed to received the users’ favorite genres, artists and songs (through Spotify) and get them to select, which concerts, they know that they want to attend at the festival.   
Based on this data from the festival attendants, the program will be able to recommend concerts for new users using collaborative filtering. Collaborative data thereby requires data to work and therefore the startup problem is present (1).

Collaborative filtering works by the principle that people with some characteristics in common, might also have other characteristics in common. Consider the following:   
Two people (Person A and B) like the same genre of music and both want to attend a certain concert. Person A want to attend another concert that Person B doesn’t know about. Then it is likely that Person B also want to attend the same concert (2). This can be done with several users and with several points of similarity to increase the quality of the recommendations.

## Pearson Correlation

One way to represent the correlation between two users is by using the Pearson correlation, which based on ratings of a set of items. The correlation is given by:

The correlation c between user X and Y is based on their rating of item k and the means of their ratings of all items in the set .

This correlation can then be used to predict a rating of an item by a user based on several other users’ ratings. This is given by the equation:

Where user X’s rating of item i is given by the ratings of the other users relative to their correlation with user X.  
(2)

## Matrix approach

Another way to use collaborative filtering is through constructing a matrix of the known users and the concerts.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Concert 1 | Concert 2 | … | Concert n |
| User 1 | Going | Going |  |  |
| User 2 | Going | Going |  |  |
| … | Going |  |  |  |
| User m | Going |  |  | Going |

The similarity between them are that all users are going to concert 1. It is then possible to count how many are going to another concert and on that basis, recommend whether a user should go to that concerts or not. This builds on the assumption that a larger number of similar users going to a concert equals a larger probability that the user in question also want to go to the concert. (1)

1(<https://www.youtube.com/watch?v=Eeg1DEeWUjA&feature=youtu.be>)

2(<http://recommender-systems.org/collaborative-filtering/>)

Hvis skal kunne bruge Pearson dimsen, så skal vi implementere et rating system. Det kunne være like/dislike eller hvor stor sandsynligheden er for at man tager til en koncert.