**Movie recommendation system**

**based on genre preference**

Young Ha, Junseong Hong

Linear Algebra – final report

**December 18, 2018**

**1 Introduction**

Recommendation systems are those that provide information filtering that is appropriate for users of various items by guessing which items users might prefer. In this context, 'filtering' is an IT term that refers to the technology of selecting the appropriate item among the various items. Although many new technologies are used in recent developments, the underlying recommendation system is based on collaborative filtering and content-based filtering.

Collaborative filtering is a technique of analyzing large amounts of existing user behavior information and recommending items that users with similar preferences have enjoyed. The most common example is the 'Products Purchased by Other Users' service, which is commonly found on the online shopping page. For example, if a user who purchased a "diaper" often purchased a "beer," he or she would recommend a "beer" to the buyer who buys a "diaper."

This algorithm has the advantage of being intuitive and does not need to analyze the specific details of an item. In this case, it is not necessary to analyze whether diaper and beer are food or not, and whether they are related to each other that should be used together. However, it is recommended to new users based on records that users purchased both products together. In order to perform collaborative filtering, existing data must be used. However, they do not have to be asked directly by users. Collaborative filtering has the advantage of allowing users to search, view items, and use purchases while naturally using the site.

However, collaborative filtering has several disadvantages. First of all, there exists a problem called cold start. If existing data are needed for collaborative filtering, it is difficult to recommend new items that are added. For example, in the case of music services, when a new song is released, it becomes difficult to recommend it until the information builds up to recommend it. Cold start literally means 'difficult to start anew'. In addition to collaborative filtering, it is also used to address the problem of early information shortages in collaborative systems such as wiki.

The second problem is the long tail. Even if there are many system items, users are likely to be interested in only a few popular items. Therefore, many items of low interest to users often fail to provide sufficient information for recommendations. Chris Anderson[1] and Clay Shirky[2] said earlier that such asymmetrical drift is common. In other words, if there are many items managed by the recommendation system, collaboration filtering may be limited.

In the text, we present a movie recommendation algorithm based on new and intuitive genre preferences. We analyze the genre of a movie that users evaluated and make it a user's genre preference matrix. Then, we can predict how much they will like the movie.

**2 Data description**

We used the MovieLens data provided by Grouplens in this project.

Group lens Research has collected a set of racing data from the MovieLens website (<http://movielens.org>) for movie recommendation. Data sets were collected over a variety of periods depending on their size.

We used the small data set, which is recommended for education and development purposes. This data set includes 5-star ratings and text tags. There are total of 100,836 ratings, 3,683 tags, 9,742 movies and 610 users. These users were chosen arbitrarily, and all users rated at least 20 movies. Each user is represented by an id with an integer value. You can see the data in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Type | Ratings | Tag | Movies | Users |
| Size | 100,836 | 3,683 | 9,742 | 610 |

**3 Description of the selected method**

Step1. construct user-genre preference matrix U

First, based on the user's assessment, find a matrix showing preferences for a particular genre. When a user evaluates a particular movie, multiplies the movie's genre vector (or 1 or 0 in the genre) with a rating. Average this vector for each user. This results in a user by genre preference matrix, U.

Step2. construct genre-movie relevance matrix M

Find a genre by movie matrix M. We produced this from a given set of data. If a movie is applicable to that genre, it will be worth 1 or 0.

Step3. calculate UM for estimation

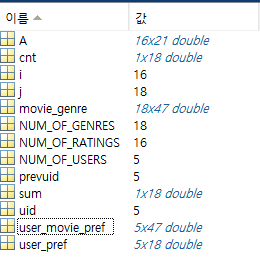
Matrix U and matrix M can now be multiplied to predict what user preferences will be for a particular movie.

**4 Results**

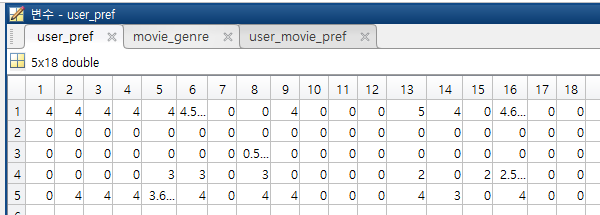
**4-1. result screen**

The result screen is shown below. If we normalize the matrices, we can get better results.

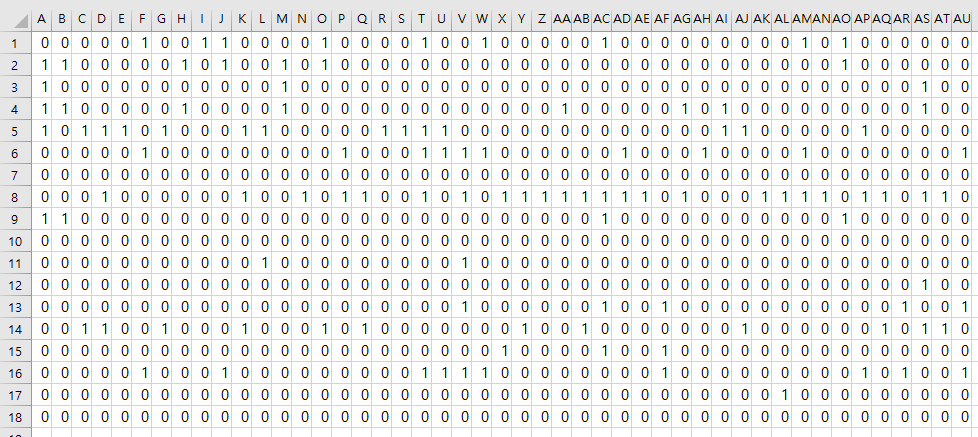
**4-1-1. variables**



**4-1-2. matrix U**

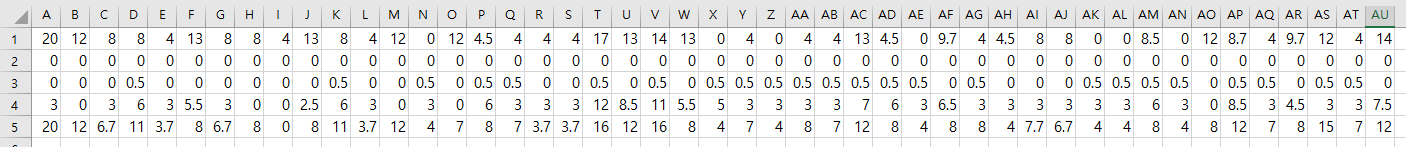


**4-1-3. matrix M**

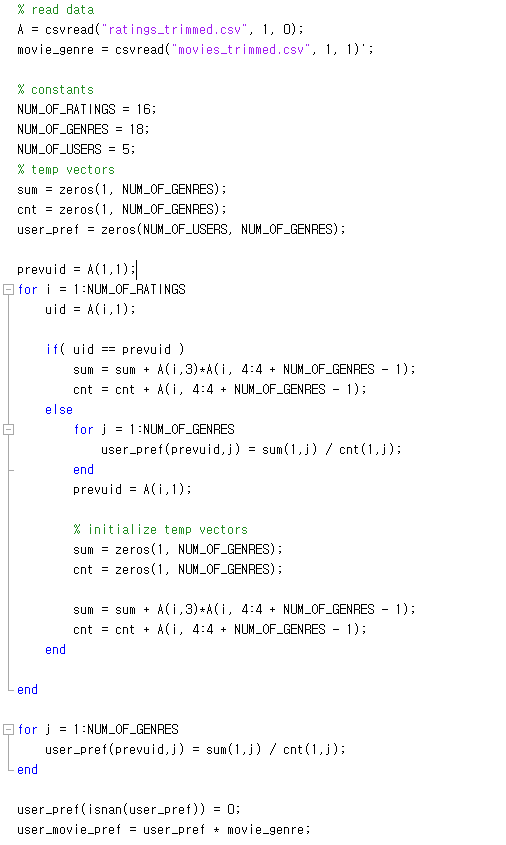


**4-1-4. User move preference matrix**

As shown below, row 2 are all zero. This is because they have not been evaluated before. In other words, problems such as cold start still remain.



**4-2. source code**



**5 References**

[1] The Long Tail: Why the Future of Business is Selling Less of More, Chris Anderson, July 11, 2006

[2] Power Laws, Weblogs and Inequality, Clay Shirky, February 8, 2003.