# 레이블 전파에 기반한 커뮤니티 탐지를 이용한 영화추천시스템

신장 캄파폰*, 비라콘 폰싸이*, 이한형**, 송민혁**, 박두순***
*순천향대학교 컴퓨터학과
**순천향대학교 컴퓨터소프트웨어공학과
e-mail : khamphaphone12@gmail.com

# Movie recommendation system using community detection based on label propagation

Khamphaphone Xinchang*, Phonexay Vilakone*, Han-Hyung Lee**, Min-Hyuk Song**,
Doo-Soon Park***
*Dept. of Computer Science and Engineering, Soonchunhyang University
**,***Dept. of Computer Software Engineering, Soonchunhyang University

**Abstract**

There is a lot of information in our world, quick access to the most accurate information or finding the information we need is more difficult and complicated. The recommendation system has become important for users to quickly find the product according to user's preference. A social recommendation system using community detection based on label propagation is proposed. In this paper, we applied community detection based on label propagation and collaborative filtering in the movie recommendation system. We implement with MovieLens dataset, the users will be clustering to the community by using label propagation algorithm, Our proposed algorithm will be recommended movie with finding the most similar community to the new user according to the personal propensity of users. Mean Absolute Error (MAE) is used to shown efficient of our proposed method.

## 1. Introduction

Now a day, the recommendation system has become a part important for help people searching information in the social network, because it helps people easy to find the information they want, help people quickly to decide and can help people find information according to their need. The recommendation system is a type of filtering that relates to predictions ratings and user satisfaction, which helps users buy products according to their needs and interests [1]. At present, researchers interested in methods of community detection in large networks. Social networks can be represented by graphs which consist of a set of nodes and edges that connect these nodes. The nodes in social network represent individuals (person) or entity and the edges match the interaction between them. The person has a similar taste, characteristics, and preference will connect as ties (edges) in the social networks that leading to the formation of virtual groups or communities [2].

*** Corresponding Author: Doo-Soon Park
※ 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2017-2014-0-00720-002)

Detecting these communities can be useful for many applications such as finding for general research areas in the collaboration network, and finding for network of protein interactions in biological networks [3].

Label propagation algorithm is one of community detection in social network. Label propagation is a semi-supervisory learning algorithm that assigns labels to previously unlabeled data points. At the beginning of the algorithm, a generally small or subsets of data points has a label or classifications. These labels will be propagated to unmarked (unlabeled) of points throughout the course of the algorithm [4].

The purpose of this article is to obtain a method that is more proficient than ordinary recommendation methods. The proposed is used community detection based on label propagation and collaborative filtering. This concept is using label propagation algorithm to group users. Then finding the similar group for the new user by used cosine similarity based on personal information such as: age, gender and occupation of each user. After the similar group, we will take the movies that have the most average rating recommended to the new user. We assume that if users are similar, they may have the same taste and they may watch the same movie.

The work that we will present in this paper is as follows: The related works of this paper will be presented in section 2,

the algorithm of the proposed method of this paper will be presented in section 3 and the experimental result of this paper will be presented in section 4.

## 2. Related work

In this section we will briefly review the main works in the context. The main focus in this paper was in collaborative filtering and community detection based on label propagation algorithm. Recommendation system (RS) is one type of intelligent system that is working by storing the history of the users' ratings or support information to help provide product recommendations to users which is often seen in online shopping, e-commerce, and the program of social network.

### 2.1 Collaborative filtering

Collaborative filtering (CF) [5] is a popular technique used in the system. But there are often disadvantages for cold start items, both in the form of complete Cold Start Items and Incomplete Cold Start Items that do not have data collection rating before or have very little data collection that will be taken Used to give advice. CF can be easily divided into 2 types: User-Based Filtering is recommending the item by finding similarity between users. This algorithm predicts that if two users there are similar behavior they might be like the same item. Item-Based Filtering is recommended by finding products that are purchased with the same customers. In this paper, we used user-based collaborative filtering. We finding the similarity between users based on personal information such as age, gender and occupation.

### 2.2 Label propagation algorithm

Label propagation algorithm (LPA) is a heuristic solution that is proposed for detecting communities in the network [6]. The main rule of the LPA's heuristic is the continuous transfer of label information between nodes. Label propagation can be performed with a few programming codes and it can apply to networks with hundreds of millions of nodes and edges on standard computers, which are only a few truthful methods in the literature. Communities can be seen as densely connected groups (Or cluster) of nodes that are only loosely connected to the remaining networks. The LPA process can be described in the following steps in Algorithm1:

Algorithm1: LPA for community detection in network

**Input**: Network *G(N, E)*, maximum number of iterations *max* Iter.

**Output**: the result of community detection

(1) Initialize the unique label for each node in the network. For a given node $x$, $c_x(0)=x$.

(2) Iteration of label propagation:

a) Set iteration number t=1.

b) Arrange the nodes of the network in random order, and store the result in vector $X=\{x_1, x_2, \cdots, x_n\}$

c) For each node $x \in X$, let

$$C_x(t) = f(C_{x_{i1}}(t), \ldots, C_{x_{im}}(t), C_{x_{i(m+1)}}(t-1), \ldots, C_{x_{ik}}(t-1))$$

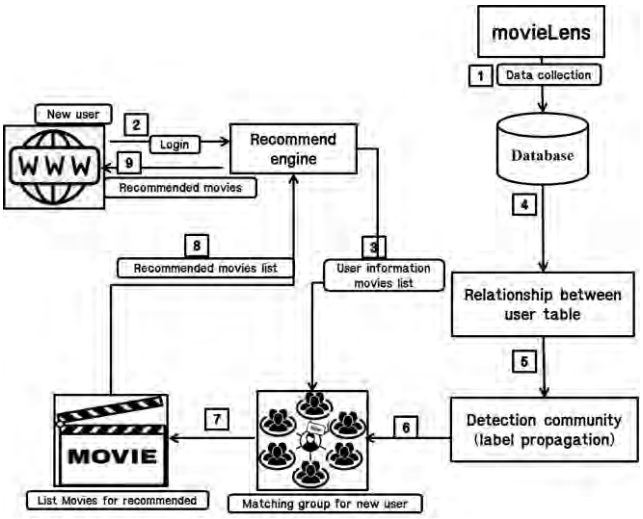
Where $x_{i1}, \ldots, x_{im}$ are the neighbor of $x_i$ those have already been update in the current iteration and $x_{i(m+1)} \ldots x_{ik}$ are neighbors those are not yet update in the current iteration. This function $f$ here returns the label that the maximum number of the neighbor has. If multiple simultaneously have the maximum number, then randomly select one of them to assign to the node.

c) If the label of every node does not change anymore, then stop the algorithm. Else, set $t=t+1$ and go to step (b).

(3) Community division: divide all node share the same label into a community, the type of label indicates the number of communities.

## 3. Proposed method

The process of our proposed method movie recommendation system by using community detection based on label propagation and collaborative filtering is presented in Fig.1.




**[Fig 1]** Flow chart of the process of proposed method

The operation of our movie recommendation method has been as follows:

**Step 1:** we collection data that use for processing including user information table (training set), movie table and rating table that is matrix between user-movie.

**Step 2:** new users should be participating in the system before recommended the movies.

**Step 3:** new users will provide their information such as name, age, gender and occupation to the system.

**Step 4:** in this step, the personal information of user is used to make relationship matrix between users, this matrix refer to user connection each other in the network.

**Step 5:** after we get the network graph from relationship matrix, users will clustering into many group by using community detect based on label propagation algorithm.

**Step 6:** after clustering users successfully, cosine similarity is used for finding which group is the most similar for new user. The cosine similarity value of new user and member in each group will be compared. The group will be selected as the new user group will be the group with the highest cosine similarity value.

**Step 7:** Once the most similar group to the new user is found, movies watched by members in the group will be ranked according to popularity and movies watched more often will be listed at the top. Otherwise, they are listed at the bottom.

**Step 8:** Top-N popular movie that we shorting previously will be select and prepared those movies for recommend engine.

**Step 9:** Top-N movies were recommended to the new user when they needs. However, the final decision about which movies the users will watch is depend on the new user.

## 4. Experimental Analysis

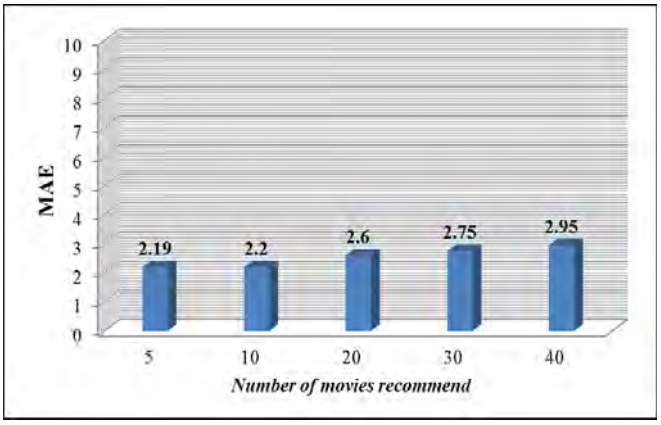
### 4.1 Dataset

Movies are one of the most popular and frequently watch in social network. People like to watch movies in order to relax but because the in the various websites there is an enormous amount. Therefore, the recommendation system becomes important. The dataset used for experiments in our paper is MovieLens dataset [7]. This dataset will be divided into 2 parts: 85% is used for modeling algorithm (training set) and 15% is used for precision prediction of our proposed methods (testing set). The dataset contains user table with 943 users, movies table with 1684 movies and rating table that each users gave to the movie 100.000 ratings.

### 4.2 Experimental result

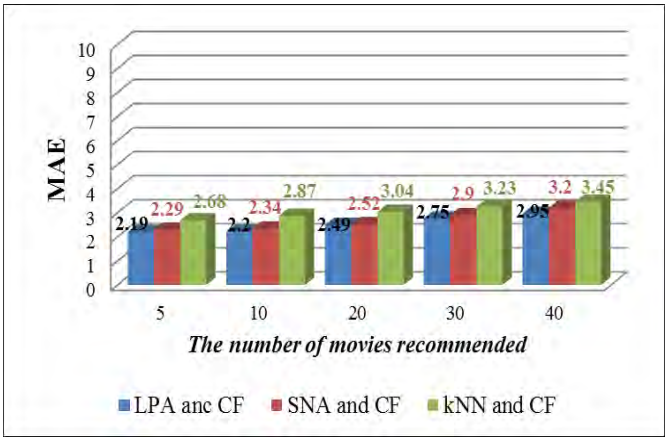
Accuracy is one of the important measures used to assess the validity of recommendations created. In general, User rating dataset that is included in the Movielens that we use is divided into two sets. One of these sets is called the $R_{Train}$, which is used to train algorithms and used to learn. The other set is called the $R_{Test}$ which is used to evaluate the accuracy of the methods we created. One of the important techniques used to analyze and measure the accuracy and precision of the generated recommendation system is Mean Absolute Error (MAE). MAE is defined as a measure of deviations or differences in the predicted ranking through the technique of filtering content based compatibility from the original ranking.

After we conducted the experiment of movie recommendation system, next we will measure the accuracy of the method. In here, Eq.2 is used. MAE can help measure the degree of satisfaction and estimate the accuracy of the recommender system. Commonly, the smallest of MAE is the highest accuracy of our recommendation system is. The result of movie recommendation system using community detection based on label propagation and collaborative filtering is shown in Fig 2.




**[Fig 2]** The result of movie recommendation system

In addition, to prove the accuracy the result of our proposed method will compare with other method. In here, we have conducted the movie recommendation using social network analysis and collaborative filtering which cluster user by using community detect based on edge betweennesse and one more method is movie recommendation system using *k*-nearest neighbor algorithm and collaborative filtering. The comparison results are shown on the Fig 3.




**[Fig 3]** The comparison result.

In Fig 3, the comparisons results have shown that our method s we offer are the most accurate. The best number of movies recommend was 5 and MAE=2.19.

## 5. Conclusion

Due to the social networking environment and the number of users increasing rapidly, it becomes difficult for users to find the information they need. The recommendation system

is used in a variety of fields and increases popularity. It's a way to help users easily find the information they need. Collaborative filtering is one of the most popular methods and is a successful method in this field. However, there are weaknesses such as cold start problems. The purpose of this study is to solve the problem of sharing filter using community detection based on label propagation in the recommendation system. The result has shown the effectiveness of movie recommendation system using label propagation and collaborative filtering more accuracy than other approaches that are using social network analysis and CF and using *k*-NN and CF.

**References**

[1] Suhasini Parvatikar, Dr. Bharti Joshi , "*Online Book Recommendation System by using Collaborative filtering and Association Mining",* IEEE International Conference on Computational Intelligence and Computing Research, 2015.

[2] Vilakone P, Xinchang K, Park D S and Hao F. "An efficient movie recommendation algorithm based on improved k-clique", Human-centric Computing and Information Sciences, 8:38, pp. 1-15, 2018.

[3] Crhisllane Rafaele dos Santos Vasconcelos, Túlio de Lima Campos,and Antonio Mauro Rezende, "*Building protein-protein interaction networks for Leishmania species through protein structural information*", BMC Bioinformatics, Published online 2018 Mar 6.

[4] Zhu, Xiaojin, Zoubin Grahramani *"Learning From Labeled and Unlabeled Data with Label Propagation"*, June 2002, CMU-CALD-02-107.

[5] Yongchang Wang, Ligu Zhu, "*Research on collaborative filtering recommendation algorithm based on mahout*", 2016 4th Intl Conf on Applied Computing and Information Technology/3rd Intl Conf on Computational Science/Intelligence and Applied Informatics/1st Intl Conf on Big Data, Cloud Computing, Data Science & Engineering, IEEE 2016.

[6] U. N. Raghavan, R. Albert, and S. Kumara, "*Near linear time algorithm to detect community structures in large-scale networks*", Physical Review E, 76(3):036106, 2007.

[7] Miller B N, Albert I, Lam S K, et al. MovieLens unplugged: "*experiences with an occasionally connected recommender system*". Proceedings of the 8th international conference on intelligent user interfaces. ACM, 263-266. 2003.