Diversity of Recommendation with Considering Data Similarity among Different Types of Contents

Jiyeon Kim
Institute of General Education, Daejin University, Pocheon, Republic of Korea
Email: jykim629@daejin.ac.kr

Youngchang Kim, Hyesun Suh, and Jongjin Jung
Department of Computer Engineering, Daejin University, Pocheon, Republic of Korea
Email: kshkyc@naver.com, {jako403, jjjung}@daejin.ac.kr

***Abstract*—Recommendation methods have important objectives of accuracy and diversity but the traditional researches have been mainly focused on the accuracy of recommendation in terms of quality. At present, the diversity of recommendation is also important to people in terms of quantity in addition to quality since people's desire for content consumption have been stronger rapidly than past. In this paper, we pay attention to similarity of data gathered simultaneously among different types of contents. With this motivation, we propose an enhanced recommendation method using correlation analysis with considering data similarity between two types of contents which are movie and music. Specifically, we regard folksonomy tags for music as correlated data of genres for movie even though they are different attributes depend on their contents. That is, we make result of new recommendation movie items through mapping music folksonomy tags to movie genres in addition to the recommendation items from the typical collaborative filtering. We evaluate effectiveness of our method by experiments with real data set. As the result of experimentation, we found that the diversity of recommendation could be extended by considering data similarity between music contents and movie contents.**

***Index Terms*—data mining, collaborative filtering, folksonomy tag, recommendation, personalization**

I. Introduction

Information technologies have changed communication society to knowledge society. People have confused by big data which have been produced massively from information system, web, social network service and smart devices. They have difficulty in selecting the reasonable information necessary for their situation. Therefore it is necessary to filter or refine or analyze the information through big data process, especially data mining [1]. Recently, many researchers have focused on personalization with diversity in addition to accuracy of recommendation in data mining areas [2]-[4]. Traditional collaborative filtering researches have been processed in terms of accuracy of recommendation. They have concentrated on the recommendation using user's profiling and personal tendency in conjunction with social network services [5]-[7]. Currently, diversity of recommendation is very important. In terms of viewpoint, collaborative filtering approaches with user experience have been studied more and more. These researches are based on folksonomy tag [8]. Folksonomy means collaborative tagging or social tagging which are classified by users directly to retrieve their necessary information. Folksonomy tag has advantage of flexibility in terms of expansion of analysis because they are not static hierarchy of taxonomy in typical classification system but dynamic horizontal structure. However, folksonomy has two problems in analyzing process, synonyms and neologism. That is, folksonolomy needs thesaurus to process synomyms and stemming work to remove stopwords and to distinguish terms of similar means [9]. To overcome these problems, [8] and [10] established tag cloud architecture with dynamic navigation link and [11] constructed an ontology-based dynamic data catalog using tag cloud. However, these researches have still got limitation of diversity problem because they used folksonomy tags only within one type of contents.

In this paper, we pay attention to similarity of data gathered simultaneously among different types of contents. For example, we note that folksonomy tags for music can be correlated to genres for movie and their values can affect each other in terms of association rule. That is, we regard folksonomy tags for music as correlated data of genres for movie even though they are different attributes depend on their contents. With this motivation, we propose an enhanced recommendation method using correlation analysis with considering data similarity between two types of contents which are movie and music. In this paper, we propose a collaborative filtering algorithm for diversity of recommendation and

Manuscript received October 15, 2015; revised December 28, 2015.

doi: 10.12720/jait.7.2.76-80

evaluate effectiveness of our algorithm by experiments with real data set.

II. The Proposed Recpmmendation Method Using Folksonomy

A. Overview of the Proposed Method

We propose an enhanced recommendation method using correlation analysis with considering data similarity between two types of contents. We typically deal with the movie contents and music contents. In our method, the core concept is focused on the inter-relationship between movie genres and folksonomy tags of music for diversity of recommend. Flow of the proposed method is shown in Fig. 1.

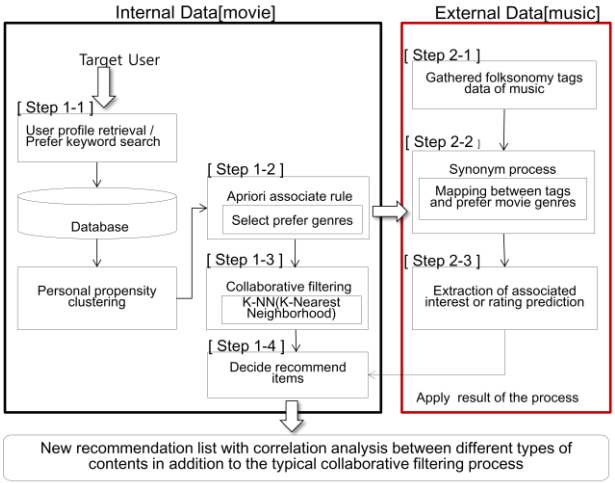


Figure 1. Flow of the proposed method.

As shown in Fig. 1, Flow of the proposed method is divided into internal data process for movie contents and external data process for music contents. In the internal data process, the method decides recommend items for movie contents by typical collaborative filtering algorithms. In the external process, the method extracts associated interest and rating prediction through the synonym process of mapping between folksonomy tags for music and prefer movie genres generated from the internal process. As the result, our method can generate new recommendation list by analyzing inter-relationship between movie contents and music contents. It will contribute to improve accuracy of recommendation by calculating with doubled data amount of both movie and music. Above all, it will also contribute to make diversity of recommendation by analyzing with different types of contents.

B. Procedure of the Proposed Method

Our proposed method processes correlation analysis with data similarity between two types of contents which consist of movies and music. Specifically, our method recommends the contents to the users based on the result of mapping folksonomy tags of music to movie genres information which are preferred by user group. The proposed method processes the following steps of procedure.

Step 1-1: User profile retrieval / Prefer keyword search

- Process clustering work using user profile data which contains age, occupation, sex and etc.
- Make first level similar user groups for their consumption data of movie contents gathered from MovieLens by GroupLens Research Project [12] team.

Step 1-2: Data matrix generation

- Examine the preferred movie genres for user groups selected from step 1-1.
- Generate data matrix among the user groups by Apriori algorithm.

Step 1-3: Collaborative filtering

- Make second level similar user groups by the similarity calculation algorithm of K-nearest neighborhood model.

Step 1-4: Recommendation list generation

- Generate recommend movie items according to the preference rating scores from step 1-3 algorithm.

Step 2-1: Music folksonomy tags extraction

- Using OpenAPI of Last.fm [13], gather the music data and their folksonomy tags which are related to movie genre information for the recommend items from step 1-4.

Step 2-2: Mapping process using synonyms

- Examine synonyms for movie genre corresponding to the music tags through semantic network of synonym dictionary
- Process mapping between movie genres and music tags.
- Decide new recommendation items of movie genres.

Step 2-3: New recommendation list creation

- Extract associated interest of movie contents correspond to music contents by calculation of rating prediction.

Create a new recommendation list which consists of the result of user's preferred genres from the above step and the recommendation list from the typical collaborative filtering from step 1-4.

III. Experimental Senario Process

A. Movie Data Set Gathering for Experiment

We use MovieLens as experimental data set. The movie genres are shown in Table I.

TABLE I. Movie Genres in Movielens

1	2	3	4
Animation	Adventure	Comedy	Action
5	**6**	**7**	**8**
Drama	Thriller	Crime	Romance
9	**10**	**11**	**12**
Children's	Documentary	Sci-Fi	Horror
13	**14**	**15**	**16**
Western	Mystery	Film-Noir	War
17	**18**	**15**	
Fantasy	Musical		

The used data set consists of approximately 100 million of data which include preference behavior rating scales. The rating scales are priced by 5 points for 3,884 movies by 6,040 users.

B. Synonyms Gathering for Movie Genre to Music Tag

We use synonym dictionary of Thesaurus [14] as the gathering method of synonyms to connect the inter-relationship between movie genres and music folksonomy tags for the movie. We collected total 473 synonyms for 18 movie genres in Table II. For example, Fig. 2 shows the result of retrieval synonyms corresponding to 'animation' of movie genre.

TABLE II. SYNONYMS FOR 18 MOVIE GENRES

Genre	Synonyms
animation	ardo, bounce, brio et al 37
adventure	experience, elan, trip et al 19
comedy	fun, humor, camp et al 43
action	deal, life, activity, power et al 51
drama	comedy, farce, melo et al 25
thriller	shocker, squeaker et al 7
crime	atrocity, breach, felony et al 48
romance	love, fling, affair et al 15
children's	baby, kid, youth et al 44
documentary	broadcast, film, narrative et al 7
sci-fi	SF, futurism, sci-fi movie et al 6
horror	disgust, dread, panic et al 23
western	westerly, westbound et al 6
mystery	enigma, problem, thriller et al 37
film-noir(darkness)	black, blackout et al 26
war	bloodshed, combat et al 16
fantasy	fancy, illusion et al 38
musical	sweet, melodic, orchestral et al 25

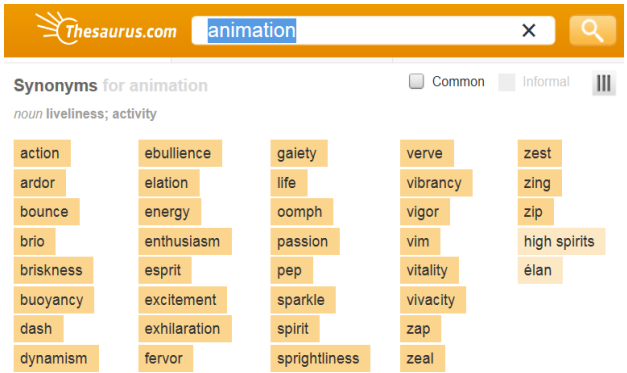


Figure 2. The synonyms corresponding to 'animation'.

We found the synonyms of 'children' instead of movie genre of 'children's' and the synonyms of 'darkness' instead of movie genre of 'film-noir'. As the result of retrieval, the movie genre of 'action' had the most number of synonyms and the movie genre of 'sci-fi' had the least number of synonyms. We counted the same frequency for each genre if there was a duplicate of the synonym words for the 18 genres.

We collected total 473 synonyms of 18 genres. Table II shows the result of collection. To extract folksonomy tags for 18 genres in Table I, we also gathered music data set from Last.fm. To acquire meaningful analysis, we collected most of top 100 music items among music data which had tags corresponding to each movie genre. At this point, we used OpenAPI of Last.fm to gather music items. As the result, we gathered total 1,750 music items as shown in Table III.

TABLE III. MUSIC ITEMS WHICH HAVE TAGS FOR MOVIE GENRES

Genre	Number	Genre	Number
Animation	100	Documentary	50
Adventure	100	Sci-Fi	100
Comedy	100	Horror	100
Action	100	Western	100
Drama	100	Mystery	100
Thriller	100	Film-Noir	100
Crime	100	War	100
Romance	100	Fantasy	100
Children's	100	Musical	100

For gathered music items, every item has one or more folksonomy tags corresponding to movie genre. The internal tag structure of music is shown in Fig. 3. This data set can be used to extract inter-relationship of movie genres.

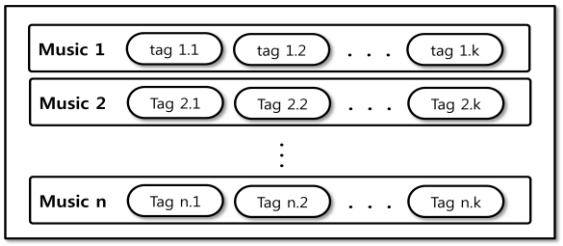


Figure 3. The internal tag structure of music item.

Table IV shows the number of internal folksonomy tags for each music item in the gathered music data set.

TABLE IV. INTERNAL TAGS OF MUSIC FOR MOVIE GENRES

Genre	Tags	Genre	Tags
Animation	645	Documentary	425
Adventure	973	Sci-Fi	2,859
Comedy	6,092	Horror	2,517
Action	1,445	Western	2,975
Drama	1,691	Mystery	1,632
Thriller	1,235	Film-Noir	1,865
Crime	965	War	2,737
Romance	2,455	Fantasy	4,921
Children's	1,642	Musical	4,520

As shown in Table IV, we can verify 645 tags for 100 music items which include synonyms corresponding to movie genre of 'animation'. We also verify that the music items for 'comedy' include the most amounts of tags and the music items for 'documentary' include the least

amounts of tags. It means that the genre of 'documentary' has relatively weak relationship with music compared with the other genres and the genre of 'comedy' has strong relationship with music.

C. Mapping Folksonomy Tags with Movie Genres

We use folksonomy tags of music to make additional recommendation for movie through mapping process folksonomy tags to movie genres. Fig. 4 shows the mapping process.

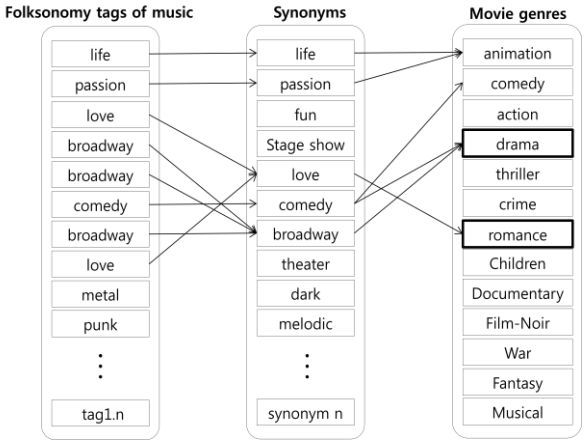


Figure 4. Mapping internal tags of music to movie genres.

As the result of experimentation, the tag of 'broadway' is synonym of relationship with 'drama' and it is found 54 times in music items corresponding to the genre of 'horror'.

D. Recommendation Based on Collaborative Filtering with K-Nearest Neborighhood

We implemented similarity calculation algorithm on Eclipse RUNA. Usually K-nearest neighborhood algorithm is used to find similar user group and Pearson correlation coefficient algorithm is used to calculate the similar distance from the user to the other user for neighborhood. However Pearson correlation coefficient algorithm has problem of inaccuracy because it doesn't consider difference of rank scaling among users. To improve this problem, we employ adjusted cosine-based similarity calculation method as the following equation.

$$Similarity(i,j)=\frac{\sum_{u\in U}(R_{u,i}-\overline{R_u})(R_{u,j}-\overline{R_u})}{\sqrt{\sum_{u\in U}(R_{u,i}-\overline{R_u})^2(R_{u,j}-\overline{R_u})^2}} \quad (1)$$

In the equation (1), $R_{u,i}$ means rank score of item i which is gotten from user u, $\overline{R_u}$ means average rank score from user u. In this paper, we make cluster of neighborhood for a user using the above similarity calculation method. Among the items which the user didn't select yet, we find n items which are given high preference from the neighborhood for the user.

E. Extended Recommendation through Mapping Process Between Movie Genres and Music Tags

We make a recommendation list by collaborative filtering with K-nearest neighborhood algorithm as shown in Table V.

In the Table V, if neighborhood for a user preferred to 'horror' of movie genre, the proposed method would recommend the movies Century (1993), Reckless (1995) and 8 Seconds (1994) to the user by collaborative filtering method. And if the method generated the genre of 'drama' as the additional recommendation genre by the mapping process in addition to the existing genre of 'horror', it would extend the recommendation items as the movies Richard III (1995), Jefferson in Paris (1995) and Sonic Outlaws (1995) in addition to the existing recommendation list. As the result of experimentation, we found that the proposed method extended the recommendation of movie using inter-relationship between movies and music. Specifically, the inter-relationship meant the correlation analysis with data similarity between movie genres and music folksonomy tags.

TABLE V. RECOMMENDATION LIST BY COLLABORATIVE FILTERING

movie	type
Balto (1995)	Comedy|Horror
Wyatt Earp (1994)	Drama|Horror
8 Seconds (1994)	Horror|Thriller
Boxing Helena (1993)	Horror|Sci-Fi|Thriller
Sonic Outlaws (1995)	**Comedy|Drama**
White Squall (1996)	Action|Horror|Sci-Fi|Thriller
Little Voice (1998)	Drama|Romance
Jefferson in Paris(1995)	**Drama|Horror**
Century (1993)	Horror
Reckless (1995)	Horror
Richard III (1995)	**Drama**

IV. CONCLUSION

In this paper, we proposed two dimensional recommendation method using inter-relationship between two different types of contents. Specifically, we tried to make diversity of recommendation through mapping between movie genres and music folksonomy tags. That is, we made result of new recommendation movie items through mapping music folksonomy tags to movie genres in addition to the recommendation items from the typical collaborative filtering. As the result of experimentation, we found that the diversity of recommendation could be extended by considering data similarity between two types of contents. We call our concept two dimensional recommendation because the proposed recommendation processes not data analysis within only a certain type of content but data correlation analysis with data similarity between two types of contents.

Currently we are extending the proposed two-dimensional recommendation method to high dimensional recommendation among two or more types of contents in view of the diversity of recommendation. We also consider the accuracy of our high dimensional recommendation through verification work for actual customers in real contents distribution market.

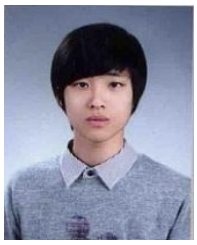
ACKNOWLEDGMENT

This research project (2014-individual9500) was supported by Ministry of Culture, Sports and Tourism (MCST) and from Korea Copyright Commission in 2015.

REFERENCES

[1] C. F. Surprenant and M. R. Solomon, "Predictability and personalization in the service encounter," *Journal of Marketing*, vol. 51, no. 2, pp. 86-96, April 1987.
[2] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce," in *Proc. 2nd ACM Conference on Electronic Commerce*, 2000, pp. 158-167.
[3] X. Su and T. M. Khoshgoftaar, "Collaborative filtering for multi-class data using Bayesian networks," *International Journal on Artificial Intelligence Tools*, vol. 17, no. 1, pp. 71-85, 2008.
[4] M. A. Ghazanfar and A. Prugel-Bennett, "Leveraging clustering approaches to solve the gray-sheep users problem in recommender system," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3261-3275, 2014.
[5] D. Park, "Improved movie recommendation system based-on personal propensity and collaborative filtering," *Journal of Information Processing Society of Korea*, vol. 2, no. 11, pp. 475-482, 2012.
[6] J. Park, Y. Jo, and J. Kim, "Social network: A novel approach to new customer recommendations," *Journal of Intelligence and Information Systems*, vol. 15, no. 1, pp. 123-140, Mar. 2009.
[7] M. Kim and K. Kim, "Recommender systems using structural hole and collaborative filtering," *Journal of Intelligence and Information System*, vol. 20, no. 4, pp. 107-120, December 2014.
[8] D. Kim, K. Lee, and H. Kim, "Improved tag selection for tag-cloud using the dynamic characteristics of tag co-occurrence," *Journal of KIISE*, vol. 15, no. 6, pp. 405-413, June 2009.
[9] G. Özbal, H. Karaman, and F. N. Alpaslan, "A content-boosted collaborative filtering approach for movie recommendation based on local and global similarity and missing data prediction," *Computer Journal*, vol. 54, no. 9, pp. 1535-1546, 2011.
[10] D. Kim, B. Lee, and C. Kim, "A study on tag cloud architecture as a dynamic navigation link," *Journal of Advanced Information Technology and Convergence*, vol. 9, no. 8, pp. 203-211, August 2011.
[11] H. Lee and M. Sohn, "Ontology-based dynamic data-catalog construction using tag cloud," in *Proc. Spring Conference on Korea Intelligent Information System Society*, May 2012, pp. 149-155.
[12] GroupLens Research Project. [Online]. Available: www.grouplens.org
[13] Last.fm. [Online]. Available: www.last.fm
[14] Thesaurus. [Online]. Available: www.thesaurus.com

Jiyeon Kim received B.S. degree (1992), M.S. degree (1997) and Ph.D. degree(2008) in department of Computer Engineering from Inha University, Incheon, Korea. From 2001 to 2005, she worked as an assistant professor in department of Internet Business Information at Chungkang College, Icheon, Korea. From 2009 to 2010, she worked as a visiting researcher in Center for Decision and Information Technology, Dallas, U.S.A. Currently she is a professor in Institute of General Education at Daejin University, Pocheon, Korea. His research interests include DRM (Digital Right Management), big data, data mining.



Youngchang Kim is currently undergraduate student in department of Computer Engineering at Daejin University, Pocheon, Korea. His research interests include DRM (Digital Right Management), big data, data mining.

Hyesun Suh received a B.S. degree in department of Statistics at Sunshin Woman's University. She received a M.S. degree(1992) and Ph.D. degree(1997) in Statistics at Korea university. From 1997 to 2012, she worked as a project manager/senior researcher in Samsung Card Cooperation, Seoul, Korea. Currently she is a professor in department of Computer Engineering at Daejin University, Pocheon, Korea. His research interests include DRM (Digital Right Management), data mining, business modeling, Analysis of Consumer's Adoption Intentions.

Jongjin Jung received B.S. degree (1992), M.S. degree (1995) and Ph.D. degree (2000) from the Inha University, Incheon, Korea. From 1998 to 2002, he worked as an assistant professor in department of Internet Information from Kyungmoon College, Pyungtaek, Korea. From 2009 to 2010, he worked as a visiting researcher in Center for Decision and Information Technology, Dallas, U.S.A. Currently he is a professor in department of Computer Engineering at Daejin University, Pocheon, Korea. His research interests include DRM (Digital Right Management), data mining, pattern recognition, knowledge-based system.