

Web Usage Mining and Text Mining in the Environment of Web Personalization for Ontology Development of Recommender Systems

Tanya Bhattacharya¹, Arunima Jaiswal², Vaibhav Nagpal³

Department of Computer Science, Amity University Uttar Pradesh, Noida-201301, India

¹tanya.bhattacharya2@student.amity.edu

³vaibhav.nagpal@student.amity.edu

Asst. Professor, Amity University, Noida, India

²ajaiswal34@amity.edu

Abstract: Recommender system is something that every website or app that provides a reliable interface of user must have. It prevents user from wasting time in segregating what he needs and encourages efficient exploring. Various factors helps to acquire the data of user interest and utilize them in building a system which recommends them on the basis of profile of their activities and interest created. Semantic web helps to create protocols for ontology that defines the type and efficiency of the recommender system that we will use. Also various principles help to develop the same.

Keywords- Ontology, web usage mining, semantic web, web personalization, mining.

I. INTRODUCTION

Recommender system has been an unbounded area for research and innovation. The work on recommender system began in mid -1990s when the papers on collaborative filtering were published. Recommender systems have widely contributed in the fields of education, medicine, academics, movies and entertainment. A recommender system helps to tackle real life situations where we have to make choices and searches based on our needs. For example, which book to buy for purpose of engineering and which for leisure reading.

Since the beginning of early 90s researchers came up with many technologies to support and create a recommender system. All the technologies related to statistical sciences, image mining techniques and also text mining techniques had been used. In this paper we talk about various types of recommender systems present. Also various technologies will be discussed which were proposed and adapted by the users. The concept of web personalization which is an upcoming concept of this decade helps the user to get a whole new experience of exploring and solving their solutions. The basic idea of ontology helped to form an entire skeleton of recommender system and how it can be made better so as to come up with an efficient user profile will be discussed as well. This will comprise technical and non-technical basis which form the scope of defining such an efficient system. Since every organization or an individual is trying to improvise

their recommender system so that more people spend more time on their site which benefits them in one or the other way, few principles of designing a recommender system are also proposed. This is supported by web usage mining with semantic web forming its backbone.

II. RELATED WORKS

A lot of work has been done in these year. A lot of technologies were proposed by the researchers and these were adapted by the new upcoming websites and portals.

Ziegler N et al. (2004) proposed an efficient model of recommendation system was proposed in the field of e commerce. This model is based on the collaborative recommendation archetype adopting an item taxonomy with the help of which user profile is made. [1]

Celma O et al. (2005) proposed another such recommender system that was proposed was FOAFRealm system. This system is constructed o the concept of P2P or peer to peer network. This system has a library which enables the users to handle their owned defined profiles using FOAF which stands for Friend of a Friend Vocabulary. This is done by obtaining from taxonomic outlines of the ontologies. [7]

Nitin Agarwal et al. (2006) proposed a recommender systems that also utilized various statistical tools to enhance the process of recommending .This is done using weighted mean method which estimates the common features between user pairs. [3]

Punam Bedi et al. (2007) suggested that establishing web of trust is recommender system that uses the concept of creating agents. Agents accumulates data over other ontologies over the network. This helps to generate an invisible recommender system for the user. The agents are generated with the help of theory of Intuitionistic Fuzzy Sets (IFS). [4]

R.Q Wang et al. (2007) discussed a system based on relationship between user and the item according user's profile. This system uses OWL (Ontology Web Language) for modelling user and item. This algorithm analyses the

similarities between the users and item using OWL and helps formation of matrix, based on the information of the active user and the previous recommendations made to the user. [5]

Amancio Bouza et al. (2008) suggested a User Model Builder system that uses the concept of decision tree. It is achieved with the help of vectors of every product and use of system of ratings for their classification. This technique helps to select a characteristic gain with maximum level of information gain and helps to make node of decision tree which categorizes the characteristics on basis of relevance and irrelevance. [7]

Bela Gipp et al. (2009) discussed Scienstein's concept of text mining in proposing a recommender system. In this system collaborative classification and annotation provides the data required. These annotations are in form of comments that readers suggest and these comments are embedded in the text itself. [8]

Saman Shishehchi et al. (2010) proposed system for eradicating cold start problem. This system is divided into two sub-system. First sub system uses OWL to depict ontology. The second sub-system is further divided into sub-systems: viewer, user's profiles, recommendation storage & user interface. [12]

C.Ramesh et al. (2011) proposed a semantically enriched recommendation system. In this system web mining is done in order to obtain web personalisation and sequential pattern mining technique is done to discover the patterns of user's nature [13]

Sarunyoo Ngamsrithepparit et al. (2012) suggested a rule-based arrangement, by method for metaphysics ideas. Specialists can use this framework by means of web application, created by conveying JSP and servlet innovation. The web application helps managing philosophy, database and rule record by conveying the JAVA dialect for which a few APIs are accessible. [16]

K.sridevi et al. (2013) proposed a system for which personalize the information to fulfil the user requirements, called web personalization. This is done by monitoring the user's search behavior over web pages by monitoring user's activity patterns. [21]

Simon Philip et al. (2014) suggested using different algorithms and most accurate recommendation by the system using the algorithm is judged to be the best algorithm. Comparison of algorithm using for recommendations are made through CTR i.e., click-through rate. This method compares any two algorithms on the basis of the ratio of clicked recommendations. This comparison is also known as A/B test. [25]

Joonseok Lee et al. (2015) proposed a system, in this system gathering and processing of data is done by the bag-of-words model. In this system he applied methodology of K-Nearest

Neighbour (KNN). This model preference of target user is estimated and according to the behavior of the user the most suitable papers for that particular user is preferred. To perform the above task we apply clustering and KNN based recommendation algorithm. This method also reduces the chance of missing applicatory topics. [29]

Joeran Beel et al (2016) conducted experiments. The tests demonstrated that there are extensive disparities in the adequacy of indistinguishable proposal approaches in just somewhat diverse situations, and in addition expansive inconsistencies for somewhat diverse methodologies in indistinguishable situations. This helped to reproduce recommendations on the basis of user details and requirements. [30]

TABLE I: Author Wise Issues Addressed In Papers

S.no	Technology Proposed	Description	Year
1	Weighted mean method recommendation. [1]	Evaluate three similarities measures which helps to identify the needs of user and recommend the product efficiently.	2004
2	FOAFRealm System. [2]	Enables the users to handle profiles using FOAF which stands for Friend of a Friend Vocabulary	2005
3	Personalized recommendations in the context of e-commerce. [3]	Based on the collaborative recommendation and adapt item taxonomy with the help of which user profile is made.	2006
4	Establishing web of trust. [4]	Agents help to recommend based on Intuitionistic Fuzzy Set user on the basis of membership, non-membership and uncertainty of the data that user is looking for.	2007
5	User Build Model. [7]	Helps to select a characteristic gain with maximum level of information gain and helps to make node of decision tree.	2008
6	Scienstein's Model based on Text Mining. [9]	Collaborative classification and annotation are required. Commenting suggests readers and tagging is done in structuralized manner	2009
7	Recommender system in field of e- learning. [12]	Uses semantic based and rule based subsystems for recommendation technique	2010

S.no	Technology Proposed	Description	Year
8	Semantically enriched web usage. [13]	Presents a sequential algorithm which traces and recommends on the basis of common search pattern	2011
9	Rule Management System for Ontology based recommender system.[16]	Rules file enables the deployment with existing recommendation system. It also facilitate experts in real-time error checking.	2012
9	Web Personalization. [21]	Maximum user satisfaction is achieved by acquiring the facts of the knowledge from the users themselves.	2013
10	Content-Based approach for recommendation System for a Digital Library. [25]	Uses TF-IDF (Term Frequency Inverse Document Frequency) and cosine resemblance to user's query or profile of interest	2014
11	Neighborhood-based recommendation Systems. [29]	Works on the assumption that social relationship influences individuals' interest or preference.	2015
12	Reproducibility in recommender-systems Research.[30]	Explains large deviation in the effectiveness of similar recommendation approaches in only slightly different scenarios	2016

III. RESEARCH METHODOLOGY

1. Recommender system Designs

Recommender system helps to reduce the search time and prevents repetitive searching. This system helps to overcome the problem of management and manipulation of large amount of data that help the user to access the options available. Recommender systems are usually divided into 3 categories:

1.2 Content based recommendation

In content based method the profiling of each user is done explicitly through questionnaires, or implicitly-observed from the user's preferences. [10]

1.2 Collaborative based recommender system

This system recommends items to the particular user on the basis of the items rated by the other users. The first recommender system which was developed was Grundy system using stereotype which build separate user models and help them to recommend relevant books to each other. [18]

1.3 Hybrid based recommendation system

This system comprises of both content and collaborative based system for the purpose of avoiding limitations of both systems. These two systems can be combined to form hybrid by following methods:

i) Separately implementing both collaborative and content-based methods and then combining their results.

ii) Incorporating some content-based characteristics into a

Collaborative approach

iii) Incorporating some collaborative characteristics into a content-based approach

iv) Constructing a general unified model [19]

2. Ontology Based Model

Ontology basically signified the idea of representing domain in machine understandable form but also that can be comprehended by humans. This is done by developing mathematical relationship and using tools such as attributes, sets and axioms. Various purpose dependencies outline the concept of its extent of usage which severely relies on the base of knowledge being used. (2007) OWL is the common web ontology language that is being used for refining the purpose of definition and instantiation in the ontology structure being used. In OWL systems the arrangements take place in tree format, each child and sub child represents a concept. When the property of a class have been changed or altered in two forms, slot versioning aides to capture that alteration. In slot versioning, the altered form with differentiated property is created and super imposed over the present form in the existing OWL file. [4]

3. Web Personalization

Nowadays every activity we perform on web requires a user profile. These user profiles are either generated by taking information from the user or by making what user searches or any activity he performs. So we can say that either these are profiles when created by the information provided by user or pseudo profile when the system learns the information of preferences of the user After initial creation of the profile, the next time the user visits the same website or a portal his past activities help to pave the way for new searches and researches that user wishes to perform. [11]

3.1 Methods of generation of user profile

The benefactors that create it can be in form of technical and on technical basis of recommending-

3.3.1 Technical basis-

i) *Text Mining*- For searching anything user enters few keywords. These keyword could lead to coming up of millions of results which may or may not interest the user. However our

recommender system provides the user with the most accurate results right at the top. Basically there is setting up of priority of the content of his interest. This process is called text mining. Text Mining analyzes the keyword supplied and produces the best quality of information that the user is seeking for. This information is derived by creating pattern trends with the help of statistical pattern recognition and learning. It is not only valid for first time search, but also the next time the user visits the same location, more improvised information set will be provided.

ii) *QR code scanning*- Scanning of QR codes is now becoming a fast moving trend. It is currently being used for the purpose of facilitating data transfer. But if analyzed closely a QR code is nothing but a limited precise recommender system. It is facilitating users to look for the data directly, i.e., recommending the user to look for the exact data. QR code when scanned provides us the exact information user is seeking for. Now the user may individually search and get various results but QR code provides the exact set of information that the user is seeking for. It prevents the wastage of time and data and hence we can say it is though limited but precise recommender system.

iii) *Image mining and search*- The user provides an image and the recommender system generates the data from the database that matches the four required characteristics. This technique recognizes the three main characteristics- the pattern, color, resolution and perception of an image and gives the data that is related to the image we used as basis of our search. In these days for online shopping this method is being used. Apart from this, this technique plays a huge role in studying of topographical sciences. The image is provided and recommender system uses the information to generate and help the scientists, teachers and students to know about the geographical area and rest of the information regarding that place such as the climatic condition, physical condition about the inhabitants, flora and fauna of that place. Manual search of a place is literally impossible and it is not humanly possible to study the topography of an area. Hence making it all possible to extract information user is seeking for.

3.3.2 Non-technical basis-

i) *Rating scale*- This is easiest manner that a recommender system can be used. Making users rate any product or non-materialistic services like scholarly articles helps to decide their efficiency. Also the rating can be based on price also and hence the user can figure out a product on the basis of his needs. Not only this the recommender system helps the new user by providing them the insight of the items and they may choose what item will favor them. The product or service with higher prioritization will be visible to user first and saving time being wasted on exploring each and every item.

ii) *Social groupism*- This technique develops a recommender system based on the users sharing a common circle like common thoughts, preferences, ideas or even relations and

acquaintances. However in this case further exploring that may be necessary for the user gets omitted. Yet users tend to trust the choices of users belonging from common circle. [12] In the following pseudo code the system acquires the data. After this it takes the rating on the product by the user and then takes the help of those ratings and recommends the user the best suited information and helps in recommendation-class recsys

```
{void rating()
  For rating the products
  {if (Rating=x)
    Array_of_items= Product_category;}
void recommend()
  For recommending items based on rating provided by old users
  {if (Rating=X)
    return Product_category; };
```

3.4 Basic Principles for designing a recommender system

3.4.1 Principle 1

The system must provide some utility. Before beginning personalization, the system must serve few purposes. It must provide aggregation of popular items, products and services. There must be a forum for discussion of the same. Also portable storage must be provided in form of bookmarks and cookies.

3.4.2 Principle 2

The system must encourage participation of users. This is done by incorporating features such as comments, like/ dislike and tags. This way more efficient data accumulation is done on the basis of user response.

3.4.3 Principle 3

The participation of users must be social. Providing real time update to user and other users of the same circle make user feel to be connected and unified.

3.4.4 Principle 4

The system must provide instant fulfilment of need of the user. The personalized recommendation must come up as soon as some input is acquired from the user.

3.4.5 Principle 5

The user must be independent to provide the input and get the required recommendation. The “wisdom of crowd” factor must be optimized to enhance the quality of system being used.

3.4.6 Principle 6

The user must be provided “long tail” of access. This means the system must recommend varieties of information out of which user can choose.

3.4.7 Principle 7

The metadata must be properly exposed to all the users. Also it must be linkable to what the user is seeking. Every information must be linked to a unique and easy URL so that it gets accessed more often.

3.4.5 Principle 8

The data encapsulation and abstraction must exist. This helps the user to share only that information which they want to and also update the recommendation by choosing whether to see similar kind of data or not when they search further again.

3.5 Online and Offline Phase of Recommendation

Unlike semantically enriched recommendation system, conventional web usage based systems are not capable of using domain knowledge of web application. Web usage mining is the application of data mining technique to extract knowledge about the user's preferences. Web mining is done to understand the user's behaviour but it does not provide the relevant data log. Mapping of web information with the semantic space gives the intelligent and effective mining web usage. The process is distinguished into two phases—offline phase and online phase:

3.5.1 Offline Phase

First component of the architecture is the processing. The main source of input i.e., server log file generally contains noisy and unwanted data. Data cleaning is done in the processing phase. The first component of the offline phase is then followed by sequential pattern mining. [10]

3.5.2 Online Phase

In this phase the user navigational history is analogized with the sequential association rules for recommending new web pages to the active user in the real time. A very untimely web page that the user visited would not likely to affect the next page since the user would make a decision which web page to be clicked by the most recent pages. The maximum number of earlier page visits used for recommendation of the new page for the user is determined by the window count parameter. [10]

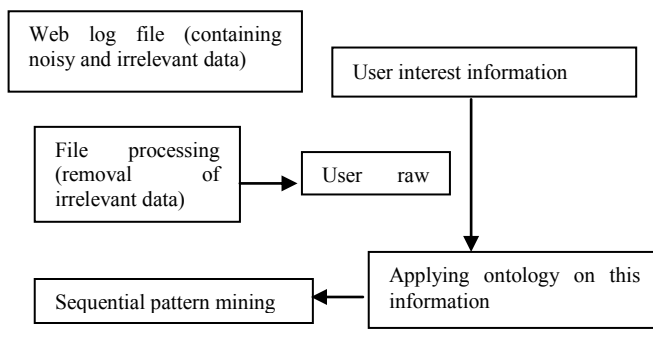


Fig. 1. Offline Phase Representation

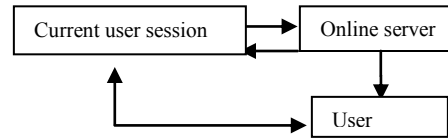
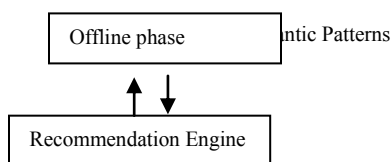


Fig. 2. Online Phase representation

3.6 Web usage mining

A recommender system quintessentially utilizes the application of web usage mining. It assists the recommender system to identify a particular pattern of usage of web which enables us to estimate and provide better utilization of data. Further it can be broken down into three parts—

3.6.1 Web Server data

Web server helps to accumulate data of user. These data involve information such as IP address, URL accessed and time spent on it. [23]

3.6.2 Application Data Server

It plays a vital role in case of e-commerce usages. It enables to keep a track of new and rising business related activities and incorporate in its all server logs.

3.6.3 Application Level Data

This enables to generate histories of all the activities and events every time the user logs in. It also helps in catching hold of criminal activities over web as all their activities will be monitored in one or the other manner. Many commercial and trade related websites are utilizing it as it helps to provide new information to users whenever they log in which essential links to their relevant topics. However this also questions the privacy of users. It is its biggest disadvantage. Since the information and data of all the users is combined and give a basis of profile for other users. Hence this questions the privacy as well as entire personalization. [26]

IV. SOLUTION TO CHALLENGES FACED

Recommender system design faces many issues. Firstly the input issue, which is regarding the user providing information that is fed to the collaborative filter algorithms. Secondly, the recommendations must be provided instantly but knowing when the recommendation isn't possible. Thirdly, generating a "trust based system" which provides the user information or data based on his choices. Lastly, sufficient information must be provided to quench the user requirements. [19]

Semantic web is used to solve these challenges faced during web services. The agents in this system help in gaining full potential web. Semantic web is used to enhance the searching mechanism combining with the ontology. The agents help in extraction of knowledge, meeting the user's queries, and

sequencing the web content according to user's preference. Semantic web provides various solutions to these problems by-

5.1 Using ontology languages OWL and RDF for designating the content in resource database.

5.2 Semantic web assures regular maintenance of structure of ontology as being dynamic in nature and produce over-time. Hence semantic web provides various technologies for maintaining ontology engineering task.

5.3 Combining ontology based system and personalized information's in various domains like e-learning, e-commerce helps in obtaining interesting results to individual's preferences.[12]

V. CONCLUSION AND FUTURE WORK

Various types of categorization formed the basis of a recommender system to exist. Further both the offline and online phases of a recommender system helps to segregate and collaborate the data at a differential level which enables the user access more accurate data as the pattern tracing is practiced. The feature of web personalisation enables generation of user profile on the basis of their search data, category and time. Despite of many advantages, these techniques face a serious issues related to trust and privacy. Since the activities of user are being monitored he may feel prone to malicious approaches. And he may also avoid giving away information, which may lead to development of false recommender system. Also further the data is clubbed and provides a particular set standard for a new user, overlooking his actual pattern of search and his individual approach. Furthermore recommender systems must be created addressing the issue of privacy and individuality. Also these can be then tested and applied on the websites related to acquiring financial and personal details of the user.

REFERENCES

- [1] Ziegler, N.; Lausen, G. y Schmidt-Thieme, L, "Taxonomy driven computation of product recommendations." Proc. of the 30th ACM Conference on Information and Knowledge Management, pp. 406-415, 2004.
- [2] Celma, O; Ramirez, M. y Herrera, P, "Foafing the music: A music recommendation system based on RSS feeds and user preferences", Proc. Of the 6th Int. Conference on Music Information Retrieval (ISMIR), pp. 464-457, 2005.
- [3] Nitin Agarwal, Ehtesham Haque, Huan Liu, and Lance Parsons, "Research Paper Recommender Systems: A Subspace Clustering Approach", Tempe AZ 85281, USA, 2006.
- [4] Punam Bedi, Harmeet Kaur, Sudeep Marwaha "Trust based Recommender System for the Semantic Web", IJCAI-07, pp. 2677-2682, 2007.
- [5] Díaz-Avilés, V.E. (2005). Semantic Peer-To-Peer recommender Systems (Tesis doctoral), <<http://citeseer.ist.psu.edu/semantic-peer-to-peer1.pdf>>[Retrieved 23/02/2008].
- [6] Wang, R.Q. y Kong, F.S, "Semantic-Enhanced Personalized Recommender System, Machine Learning and Cybernetics", Proc. of the Int. Conference on Machine Learning and Cybernetics, v. 7 (19-22), pp. 4069-4074, 2007.
- [7] Amancio Bouza, Gerald Reif, Abraham Bernstein, Harald Gall, "SemTree: Ontology-Based Decision Tree Algorithm for Recommender Systems", International Semantic Web Conference, Karlsruhe, Germany, Oct. 2008.
- [8] E. Peis; J. M. Morales-del-Castillo; J. A. Delgado-López, "Semantic Recommender Systems. Analysis of the state of the topic", Hipertext.net, Yearbook, issue 6, May 2008. <http://www.hipertext.net>, no. 6, 2008.
- [9] Bela Gipp, Jöran Beel, Christian Hentschel, "Scienstein: A Research Paper Recommender System", International Conference on Emerging Trends in Computing (ICETiC'09) Virudhunagar (India), pp. 309-315, Jan. 2009.
- [10] S. Salin and P. Senkul, "Using semantic information for web usage mining based recommendation", 24th ISCIS, Guzelyurt, Turkey, pp.236-241, 2009.
- [11] N. R. Mabroukeh and C. I. Ezeife, "Semantic-rich Markov models for Web prefetching", ICDMW, Miami, FL, USA, pp.465-470, 2009. ICDMW, Miami, FL, USA, pp.465-470, 2009.
- [12] Saman Shishehchi, Seyed Yashar Banihashem, Nor Azan Mat Zin, "A Proposed Semantic Recommendation System for E-Learning", Information Technology (ITSim), Symposium in, Vol. 1, 2010. International
- [13] C.Ramesh, Dr. K. V. Chalapati Rao, Dr. A.Goverdhan, "A semantically enriched web usage based recommendation model", International Journal of Computer Science & Information Technology (IJCSIT) Vol.3-No 5, pp. 193-202, Oct .2011.
- [14] F.Ricci, L. Rokach, B. Shapira, K.B Paul, "Recommender Systems handbook", Springer, NewYork, 2011.
- [15] Xiaohui Tao, Yuefeng Li, Ning Zhong, "A Personalized Ontology Model for Web Information Gathering", IEEE transactions on Knowledge and Data Engineering, Vol. 23, No. 4, April .2011.
- [16] S. Ngamsrithepparit, T. Supnithi, Y. Myat Thein, K. Runapongsa Saikaew, M. Buranarach, S. Poltree, "Rule Management System for Ontology-based Recommendation System", The Joint International Symposium on Natural Language Processing and Agricultural Ontology Service 2011, Bangkok, Thailand, Feb 9-10, 2012. <http://gear.kku.ac.th/~krunapon/research/pub/rule-management-snlp-2012.pdf>
- [17] K. Suneetha, and M. Usha Rani, "Web Page Recommendation Approach Using Weighted Sequential Patterns and Markov Model", Global Journal of Computer Science and Technology, 2012.
- [18] J. Lee, M. Sun, and G. Lebanon, "A comparative study of collaborative filtering algorithms", ArXiv Report 1205.3193, 2012.
- [19] Bhaganagare Ravishankar, Dharmadhikari Dipa, "Web Personalization Using Ontology: A Survey", IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661, Volume 1, Issue 3, pp. 37-45, May-June 2012.[20] R. Suguna, D. Sharmila, "An Efficient Web Recommendation System using Collaborative Filtering and Pattern Discovery Algorithms", International Journal of Computer Applications (0975 – 8887), Vol. 70– No.3, May 2013.
- [20] K. Sridevi, Dr. R. Umarani "Web Personalisation, "Web Personalisation Approaches: A Survey", International Journal of

Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 3, Mar. 2013

- [21] J. Lee, S. Kim, G. Lebanon, and Y. Singer, “*Local low-rank*”, International Conference on Machine Learning, 2013.
- [22] *Matrix approximation*”, International Conference on Machine Learning, 2013.
- [23] A.Said, “*Evaluating the Accuracy and Utility of Recommender Systems*”, PhD Thesis, Technische Universität, Berlin, 2013.
- [24] N .Manouselis, K. Verbert, “*Layered evaluation of multi-criteria collaborative filtering for scientific paper recommendation*”, Procedia Computer Science, Elsevier, New York, pp.1189–1197, 2013.
- [25] Simon Philip, P.B. Shola, Abari Ovy John, “*Application of Content-Based Approach in Research Paper Recommendation System for a Digital Library*”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5,-No. 10, 2014.
- [26] A.Said, A.Bellogin, “*Rival: a toolkit to foster reproducibility in recommender system evaluation*”, Proceedings of the 8th ACM Conference on Recommender Systems(ACM), New York, pp. 371–372, 2014.
- [27] Y.Shi, M. Larson, A. Hanjalic, “*Collaborative filtering beyond the user-item matrix: a survey of the state of the art and future challenges*”, ACM Comput. Surv, 47(1), 3:1–3:45, 2014.
- [28] J. Lee, S. Bengio, S. Kim, G. Lebanon, and Y. Singer, “*Local collaborative ranking*”, International Conference on World wide web, 2014.
- [29] Joonseok Lee, Kisung Lee, Jennifer G. Kim, Sookyoung Kim, “*Personalized Academic Paper Recommendation System*”, SRS’15, Sydney, NSW, Australia, 10 Aug. 2015.
- [30] Joeran Beel, Corinna Breiting, Stefan Langer, Andreas Lommatzsch, Bela Gipp, “*Towards reproducibility in recommender-systems Research*”, published online: Springer Science+Business Media Dordrecht, 12 Mar.2016.