RECOMMENDATION SYSTEM USING DATA MINING AND CLUSTERING TECHNIQUES

[1]Dinesh Vijay Attarde, [2]Dr.Manmohan Singh,
[1]PG Student,[2]Associate Professor,
[1]Department of Computer Science & Engineering,
[1]RKDF School of Engineering, Indore,India

Abstract*—Today, there is a massive variety of various strategies and algorithms of data filtering and recommendations giving. In this paper we describe traditional techniques and give an explanation for what type of present day strategies have been evolved lately. All the paper long we can try and provide an explanation for strategies and their problems. In the end we are able to display proposed approach which assists to generate higher recommendations to user. Throughout this paper we design a new technique that overcomes the data-sparsity disadvantage and get better the performance accuracy.

***IndexTerms*—Recommendation System,Clustering,Recommendation methods.**

I. INTRODUCTION

The explosive increase in the quantity of to be had digital information and the wide variety of site visitors to the Internet have created a capacity project of records overload which hinders well timed get right of entry to items of interest at the Internet. Information retrieval systems, consisting of Altavista, Devil Finder and Google have fairly solved this difficulty other than prioritization and personalization (where a system maps available content material to person's interests and possibilities) of information were not present. This has expanded the insist for recommender structures greater than yet before.

Recommender structures are data filtering structures that cope with the difficulty of data excess [1] by filter very important data piece out of big quantity of dynamically generated information in line with user's preferences, interest, or experimental performance about item [2]. Recommender system has the ability to expecting whether or not a selected user might select an object or now not primarily based at the user's profile.

Recommender structures are useful to both to check provider and user [3]. They lessen transaction expenses of finding and selecting items in online shopping surroundings [4]. Recommendation structure has better decision making method and superiority [5]. In e-commerce setting, recommender methods increase revenues, for the truth that they are mighty method of promoting more products [3]. In scientific libraries, recommender methods help users by way of permitting them to maneuver past catalog searches. For this reason, the needs to use efficient and correct recommendation techniques inside a procedure with a view to provide vital and liable strategies for users cannot be there overemphasize.

II. TRADITIONAL RECOMMENDER APPROACHES

A. Content-based filtering

Content-based filter recommender systems work with profiles of users that are created at the beginning. A profile has information about a user and his taste. Taste is based on how the user rated items. Usually, when creating a profile, recommender systems make a survey, to get initial information about a user in order to avoid the new-user problem. [6]

Within the recommender method, the engine compares the items that had been through now absolutely rated with the aid of the user with the items he didn't rated and appears for similarities. Those items which are mainly similar to the absolutely rated ones might be advocated to the user.

B. Collaborative filtering

Collaborative filtering grew to become probably the most researched procedures of recommender systems on account that this procedure was recounted and described by way of Paul Resnick and Hal Varian in 1997[7]. The suggestion of collaborative filtering is to find users in a community that share appreciations [8]. If two users have identical or close to identical rated items in usual, then they've an identical tastes. Such users construct a group or a so called neighborhood. A user gets recommendations to these items that he/she hasn't rated earlier than, but that had been already absolutely rated by users in his/her nearby.

As opposed to simpler recommender systems where recommendations base on the most rated item and the most popular item methods, collaborative recommender systems care about the taste of user. The taste is considered to be constant or at least change slowly.

Going in details of methods of collaborative filtering we can distinguish most popular approaches: user-based, item-based and model-based approaches.

C. Hybrid recommendation approaches

For better results some recommender systems combine different techniques of collaborative approaches and content based approaches. Making use of hybrid strategies we are able to avoid some obstacles and issues of pure recommender systems, just like the cold-start problem. The blend of systems can proceed in one-of-a-kind ways [9]:

1) Separate implementation of algorithms and joining the outcome.
2) Utilize some rules of content-based filtering in collaborative approach.
3) Utilize some rules of collaborative filtering in content based approach.
4) Create a unified recommender system, which brings together both approaches

III. MODERN RECOMMENDATION APPROACHES

A. Context-aware approaches

Context is the information about the environment of a user and the details of situation he/she is in. Such details may play much more significant role in recommendations than ratings of items, as the ratings alone don't have detailed information about under which circumstances they were given by users. Some recommendations can be more suitable to a user in evening and doesn't match his preferences in the morning at all and he/she would like to do one thing when it's cold and completely another when its hot outside. The recommender systems that pay attention and utilize such information in giving recommendations are called context-aware recommender systems.

One of the biggest problems of context-aware recommender systems is obtaining context information. The information can be obtained explicitly by directly interacting with user asking him/her to fill out a form and making a survey. Although it is mostly desirable to obtain context information without making the whole rating and reviewing process complicated. Another way is gathering information implicitly using the sources like GPS, to get location, or a timestamp on transaction [10]. The last way of information extraction is analyzing users observing their behavior or using data mining techniques. For example: obtaining information from reviews. People usually like to write their reviews in free text form. The problem is to get important contextual information from such reviews. There are many text mining algorithms including artificial intelligence techniques [11].

B. Semantic based approaches

Most of the descriptions of items, users in recommender systems and the rest of the web are presented in the web in a textual form. Using tags and keywords without any semantic meanings doesn't improve the accuracy of recommendations in all cases, as some keywords may be homonyms. That's why understanding and structuring of text is a very significant part recommendation. Traditional text mining approaches that base on lexical and syntactical analysis show descriptions that can be understood by a user but not a computer or a recommender system. That was a reason of creating new text mining techniques that were based on semantic analysis. Recommender systems with such techniques are called semantic based recommender systems [12]. The performance of semantic recommender systems are based on knowledge base usually defined as a concept diagram (like taxonomy) or ontology [13].

C. Cross-domain based approaches

Finding similar users and building an accurate neighborhood is an important part of recommending process of collaborative recommender systems. Similarities of two users are discovered based on their appreciations of items. But similar appreciations in one domain don't surely mean that in another domain valuations are similar as well. Users sharing preferences in comedies are not by all means like the same type of horrors [14].

D. Peer-to-Peer approaches

The recommender systems with P2P approaches are decentralized. Each peer can relate itself to a group of other peers with same interests and get recommendations from the users of that group. Recommendations can also be given based on the history of a peer. Decentralization of recommender system can solve the scalability problem [15] [16].

E. Cross-lingual approaches

Cross-lingual recommender systems break the language barrier and gives opportunities to look for items, information, papers or books in other languages. Yang, Chen and Wu purposed an approach for a cross lingual news group recommendations. The main idea is to map both text and keywords in different languages into a single feature space, that is to say a probability distribution over latent topics. From the descriptions of items the system parses keywords than translates them in one defined language using dictionaries. After that, using collaborative or other filtering, the system gives recommendations to users [17].

With the help of semantic analysis it's possible to make a language-independent representation of text. Pascal Lops purposed an approach with a recommender system called MARS. Based on Word Sense Disambiguation algorithm that exploits online multilingual lexical database as sense repository the engine assigns right meaning to the words avoiding synonyms problems [18].

IV. PROPOSED METHOD

The major purpose of proposed system is to recommend excellent compatible item to the end user. With recognize to the second resolution, in order to experiment the performance of our system, we shall measure its capability to forecast a user's true ratings or preferences, i.e. system accuracy.

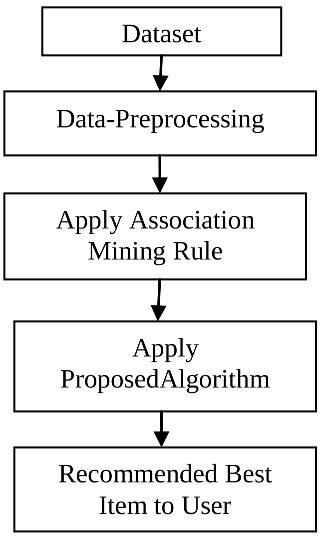


Fig.1. Recommended System Architecture

We performs following steps

1. Selection of the dataset.
2. Preprocessing of the data.

3. Applying the association mining rule on different clustering groups.
4. Applying Proposed algorithm.
5. To recommend the best items to the user.

Proposed Algorithm is an influential algorithm for mining frequent item sets for Boolean association rules. It works as shown in following Fig. 2

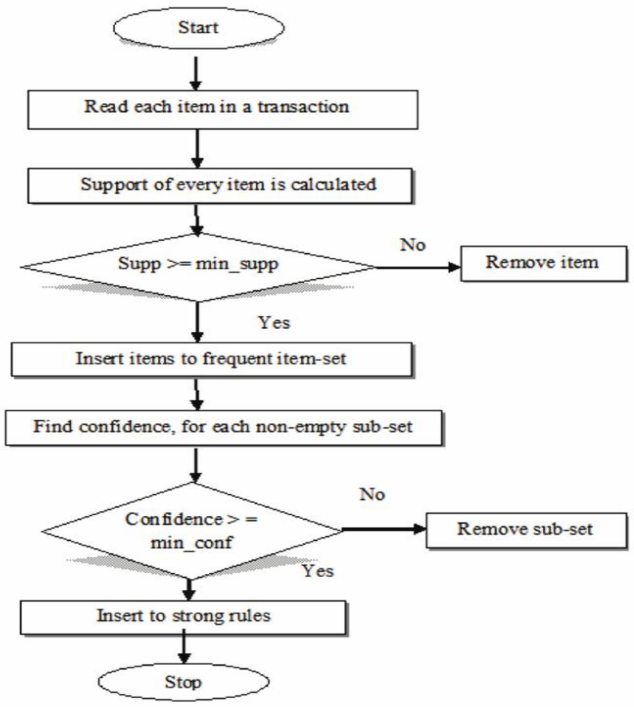


Fig. 2 Flowchart of algorithm

Where support and confidence values are calculated as below, in which let's consider X and Y are itemsets and N is total number of item sets in a database

$$\text{Support} = \frac{\text{Frequency(X, Y)}}{\text{N}}$$

and

$$Confidence = \frac{\text{Frequency(X, Y)}}{\text{Frequency(X)}}$$

V. EXPRIMENTAL SETUP

Mathematical Model Step 1:

User/Item	R1	R2	R3	R4	R5
U2	5	0	3	1	0
U4	0	4	4	4	4
U7	4	4	4	3	0
U10	5	0	0	3	3

Step 2: Converting dataset into Boolean data (where 0 means False and 1-5 means True)

User/Item	R1	R2	R3	R4	R5
U2	True	False	True	True	False
U4	False	True	True	True	True
U7	True	True	True	True	False
U10	True	False	False	True	True

Step 3: i. Frequent -1 Item set

R1=True	{ u2 , u7,u10}
R2 =True	{u4,u7}

R3 = True	{u2,u4,u7}
R4 =True	{u2,u4,u7,u10}
R5= True	{u4,u10}
R1=False	{ u4}
R2 =False	{u2,u10}
R3 =False	{u2,u10}
R4 =False	{ □ }
R5= False	{u2,u7}

ii. Frequent – 2 Item set

R1=True,R2 =True	{ u7}
R1 = True,R2 =False	{u2,u10}
R1 =True,R3 =True	{u2,u7}
R1=True,R5 =False	{u2,u7}
R1=False,R5 = True	{u4,u7}
R2=True,R3=True	{u4,u7}
R2=True,R4=True	{u4,u10}
R3=True,R4=True	{u2,u4,u7}
R3=True,R5=False	{u2,u7}
R4=True,R5=False	{u2,u7}
R4=True,R5=True	{u4,u10}

iii. Frequent 3-itemset

R1=True,R3,=True , R4 =True	{ u2,u7}
R1=True,R3=True,R5=False	{u2,u10}

VI. RESULTS AND ANALYSIS

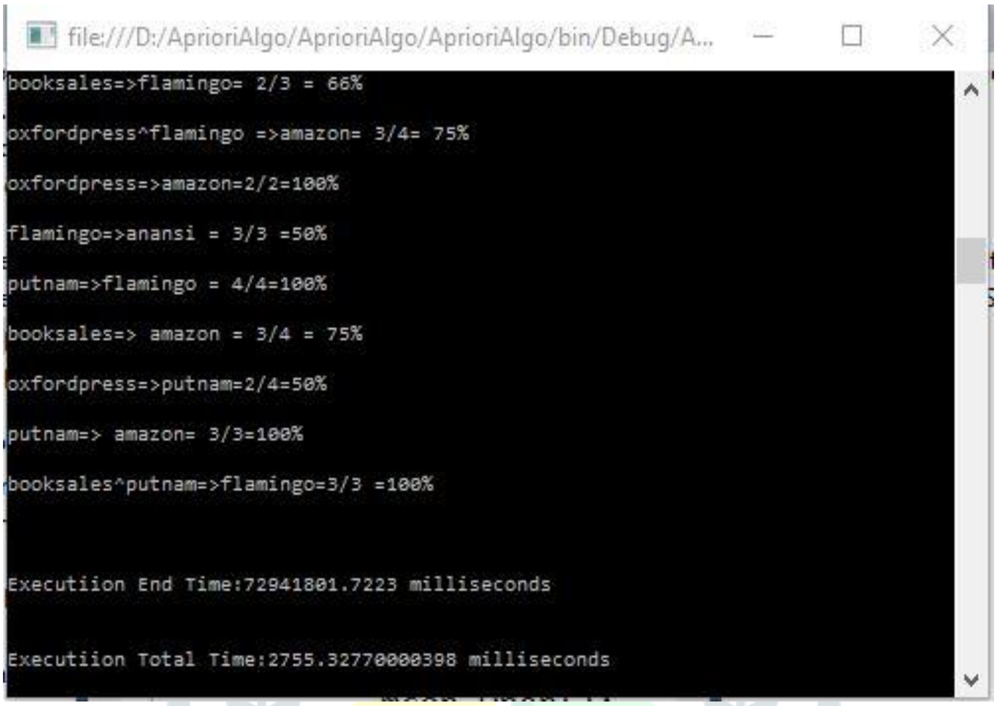
Here book dataset is used. It contain user id and item name field. On this dataset pre-processing is applied. So, it will generate user-item matrix. Here 50 x 50 size of data are used.

Table 1-Book Dataset

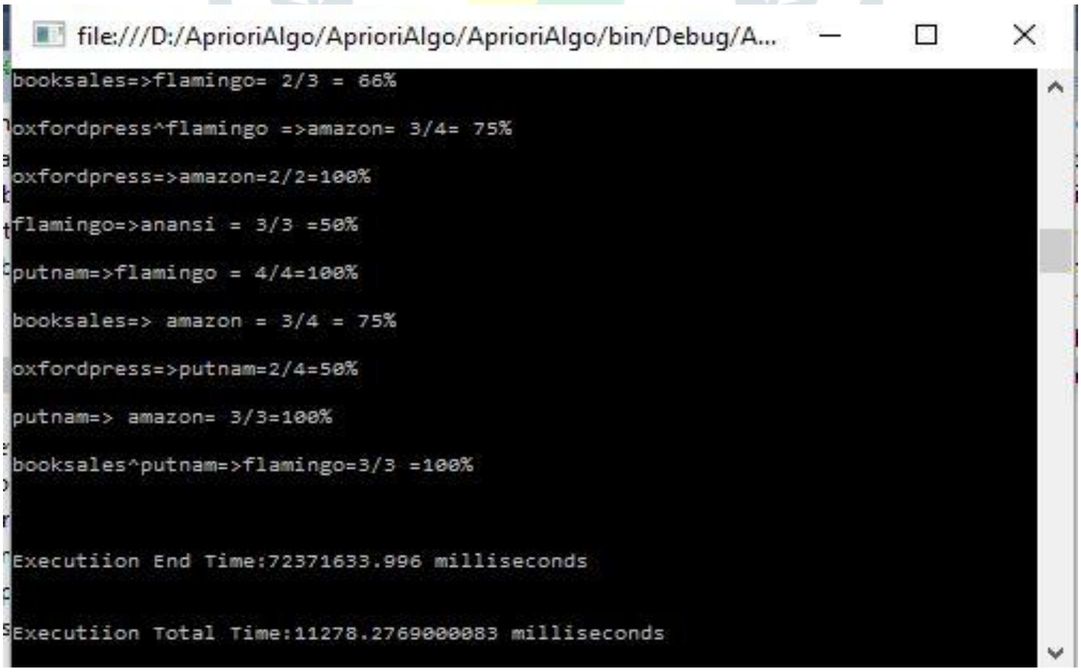
User ID	Item Name
276725	ConDes
276726	oxfordpress
276727	putnam
276730	anansi
276731	amazon
276732	flamingo
276733	booksales
276734	threeriverpress
276735	saturdayreviewpress
276725	ConDes
276736	rylandpeters
276738	acebook
276739	Berkleypublishing
276740	pearson
276741	signet

276748	anansipress
276755	signet acebook
276759	Thomoshardy
276760	st martin press
276761	avon trade
276762	penguin hall
276764	life's little
276765	instruction book
276766	HEART

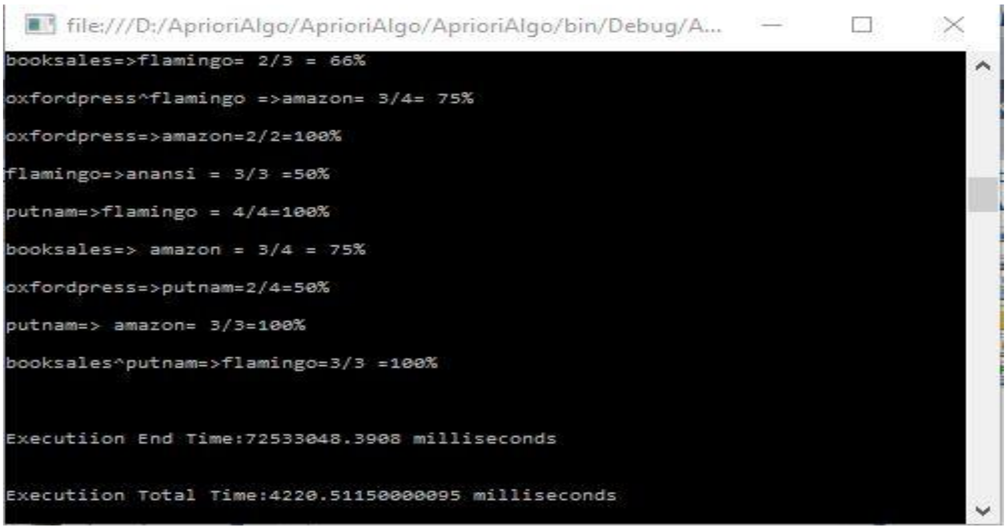
Ap-20



Ap-30



Ap-40



Where Minimum support = 2

Confidence	Execution time
20	2755
30	3277
40	4220
50	5023
75	6621

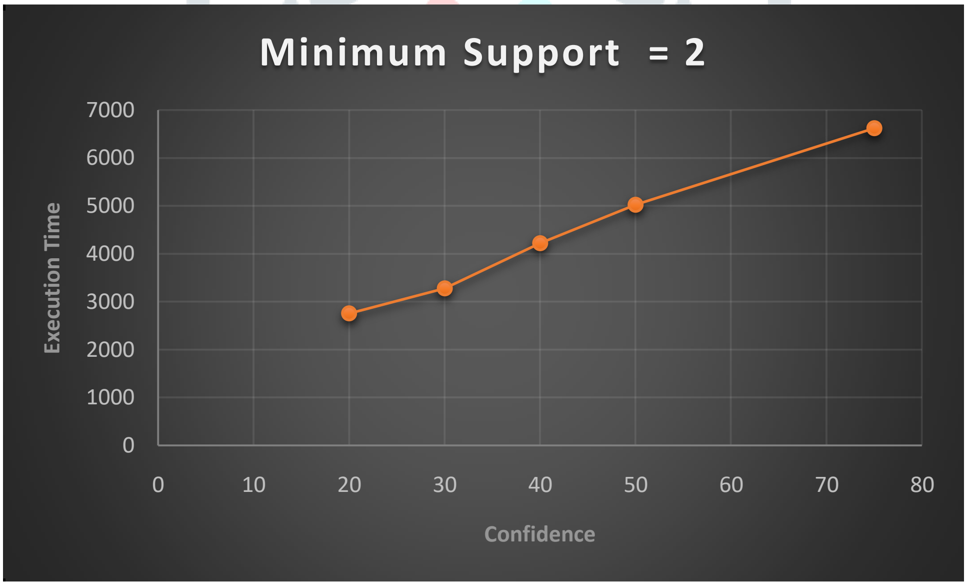


Chart-1 Confidence and Execution time graph

Where Minimum confidence = 20

Support	Execution Time
2	1004
3	870
4	565
5	400

Chart-2 Support and Execution time graph

VI. Conclusion

In this paper, we try to briefly describe the various types of recommendation techniques and its type. In proposed method, we have select book dataset, after that preprocessing the dataset so, it generate user-item matrix. The association mining rules apply on different clustering groups after that apply Proposed algorithm to recommend the best items to the user. This will result in a very efficient recommendation system with has its own intelligence to predict the best interest of the user and hence provide recommendations with high accuracy.

References

[1] J.A. Konstan, J. Riedl, "Recommender systems: from algorithms to user experience User Model User-Adapt Interact", 22 (2012), pp. 101-123 CrossRefView Record in Scopus

[2] C. Pan, W. Li, "Research paper recommendation with topic analysis", In Computer Design and Applications IEEE, 4 (2010) pp. V4-264

[3] Pu P, Chen L, Hu R., "A user-centric evaluation framework for recommender systems", In: Proceedings of the fifth ACM conference on Recommender Systems (RecSys'11), ACM, New York, NY, USA; 2011. p. 57–164.

[4] Hu R, Pu P., "Potential acceptance issues of personality-ASED recommender systems", In: Proceedings of ACM conference on recommender systems (RecSys'09), New York City, NY, USA; October 2009. p. 22–5.

[5] B. Pathak, R. Garfinkel, R. Gopal, R. Venkatesan, F. Yin, "Empirical analysis of the impact of recommender systems on sales", J Manage Inform Syst, 27 (2) (2010), pp. 159-188

[6] L. Candillier, F. Meyer, F. Fessant, and K. Jack, "State-of-the-art recommender systems," 2009.

[7] P. Resnick and H. R. Varian, "Recommender systems," Commun. ACM, vol. 40, pp. 56–58, March 1997. [Online]. Available: http://doi.acm.org/10.1145/245108.245121

[8] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: Item-to-item collaborative filtering," IEEE Internet Computing, vol. 7, pp. 76–80, January 2003. [Online]. Available: http://portal.acm.org/citation.cfm?id=642462.642471

[9] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Trans. on Knowl. and Data Eng., vol. 17, pp. 734–749, June 2005. [Online]. Available: http://dx.doi.org/10.1109/TKDE.2005.99

[10] G. Adomavicius and A. Tuzhilin, "Context-aware recommender systems," in Recommender Systems Handbook: A Complete Guide for Research Scientists and Practitioners. Springer, 2010.

[11] S. Aciar, "Mining context information from consumers reviews," 2010.

[12] A. Elgohary, H. Nomir, I. Sabek, M. Samir, M. Badawy, and N. A. Yousri, "Wiki-rec: A semantic-based recommendation system using wikipedia as an ontology," in Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on, 29 2010.

[13] E. Peis, J. M. Morales-del Castillo, and J. A. Delgado-Lpez, "Semantic recommender systems. analysis of the state of the topic." Seoul, South Korea: Department of Computer Science, YonseiUniverisity, 2008.

[14] S. Berkovsky, T. Kuflik, and F. Ricci, "Cross-domain mediation in collaborative filtering," in Proceedings of the 11th international conference on User Modeling, ser. UM '07. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 355–359. [Online]. Available: http : ==dx:doi:org=10:1007=978- 3-540-73078-144

[15] R. Kanawati and H. Karoui, "A p2p collaborative bibliography recommender system," in Proceedings of the 2009 Fourth International Conference on Internet and Web Applications and Services. Washington, DC, USA: IEEE Computer Society, 2009, pp. 90–96. [Online]. Available: http://portal.acm.org/citation.cfm?id=1585689.1586218

[16] R. Baraglia, M. Mordacchini, P. Dazzi, and L. Ricci, "A p2p recommender system based on gossip overlays (prego)," in Computer and Information Technology (CIT), 2010 IEEE 10th International Conference on, 2010, pp. 83 –90.

[17] C.-Z. Yang, I.-X. Chen, and P.-J. Wu, "Cross-lingual news group recommendation using cluster-based cross-training." Taiwan: Computational Linguistics and Chinese Language Processing, 2008, pp. 41–60.

[18] P. Lops, C. Musto, F. Narducci, M. De Gemmis, P. Basile, and G. Semeraro, "Mars: a multilanguage recommender system," in Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems, ser. HetRec '10. New York, NY, USA: ACM, 2010, pp. 24–31. [Online]. Available:http://doi.acm.org/10.1145/1869446.1869450