University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Library Philosophy and Practice (e-journal) | Libraries at University of Nebraska-Lincoln

Summer 4-1-2019

Recommender Systems for Digital Libraries: A review of concepts and concerns

VISHAL GUPTA
vg.vns3@gmail.com

Dr. Shriram Pandey Assistant Professor
Department of Library & Information Science, Banaras Hindu University, Varanasi, India, drshrirampandey@gmail.com

Recommender Systems for Digital Libraries: A review of concepts and concerns

Vishal Gupta
Research Scholar, Department of Library & Information Science, Banaras Hindu University, Varanasi, India. E-mail- vg.vns3@gmail.com

Dr. Shri Ram Pandey
Assistant Professor, Department of Library & Information Science, Banaras Hindu University, Varanasi, India. E-mail- drshrirampandey@gmail.com

Abstract

The study hopefully has given an understanding of Recommender System (RS) concept and trends of IR systems especially in the domain of digital libraries. It unfolded the concept of RS through the review of literature and presented an outline of the concepts. Paper discussed the importance of recommender systems in the digital library domain. Study further explain the concept of different kind of RS applied to different digital library software systems. This paper shows how recommender systems functions in different library systems and how these recommender system helps to the users to find and retrieve data or information from different databases. The basic aim of this paper is to know the future aspects of recommender systems in digital library systems and the implications according to its need. This paper contains about conceptual base of the recommender systems, their approaches and their usability in different field of information gathering systems.

Keywords: Digital Library, Recommender System, User Services, Filtering systems, Digital repositories.

1. Introduction

A Recommender System (RS) is a tool, which gives recommendations to the user according to their needs or query. Basically, the main feature or function of the Recommender System is to predict the user interest by relating the history, information, profile and queries used, searched, created and expressed by the user. From the beginning of Web 2.0, the internet began growing up and developing with tremendous speed. It provided the users with multiple opportunities, such as sharing knowledge, information, and the opinion of others. This did a favour to the development of social networks and commercial sites like Facebook, Amazon etc. Nowadays, authors can share their innovative ideas with millions of users around the world.

The aim of the Recommender System is to handle the information overload problem and improve the customer relationship management. This tool shows the result based on suitable items products and services and nowadays it is useful in getting recommendations for different fields like e-multimedia, e-governance, e-business, e-library etc. It can play a huge role in classifying the information for bibliographic control and information retrieval mechanism. Implementation of a recommender system provides lots of benefits to the users like identifying user interest, locating relevant information and helping in dynamic research activity [Isinkaye, Folajimi & Ojokoh 2015][1]. RS has emerged in the recent years as a remarkable tool meant for handling the information overload problem. Ultimately an RS addresses this phenomenon by pointing a user towards new, not-yet-experienced items that may be relevant to the user's current task [Ricci, 2015][2]. Recommender systems are helpful in searching contents since they help users to find information which can't be found anywhere else. However, recommender systems are often applied in many search engines to index and find non-conventional data. Recommender systems (RS) was introduced in the mid 90's and it's was first alluded by Jussi Karlgren [Karlgren, 1990][3] at Columbia University in a specialized report as a "Digital Bookshelf" in 1990. There has been significance contribution made by various experts since then. The paradigm shift of RS has gone through lots of technological advancement.

2. Literature Review

Recommender system has been so extensively used these days that it has become a preferable choice for researchers. Recommender systems evolved as an independent research area in the mid1970s in Duke University. The first recommender system that came into existence was Tapestry which was developed at the Xerox Palo Alto Research Centre [Sharma & Singh, 2016][4]. The motivation that led to its development was the growing number of incoming emails, mostly unnecessary ones that were too annoying sometimes and difficult to maintain. So what was done to overcome this problem was that users had their mailing lists made and only those people who were in the contact lists could send the mails to them or the ones the user might want to hear from while the others were sent to the spam list, exactly how it works in present in our email accounts. Numerous survey papers on Recommender systems have been published till date. For example, 4 in their paper reviewed more than 200 research articles about research-paper recommender systems and presented some descriptive statistics in the paper and a discussion about the major advancements and shortcomings and an overview of the most common recommendation concepts and approaches. First paper on recommender system was published in year 1998. Since then a significant number of papers had been published. Year 2005 John O'Donovan, Barry Smyth have taken trust as the percentage of correct predictions that a profile has made in general (profile-level trust) or with respect to a particular item (item-level trust) [O'Donovan & Smyth, 2005] [5]. Juan A. Recio-García presented a prototyping Recommender Systems in jCOLIBRI [Recio-García , 2008][6]. The goal of this recommender system is to support system developers in rapid prototyping recommender systems using Case-Based Reasoning (CBR) techniques. In year 2007 Paul Resnick proposed an idea of "influence limiter algorithm"[Resnik, 2007][7] in recommender system. This algorithm prevents any attack which shows the irrelevant result for our search. This algorithm limits the number of content that an attacker can modify. In 2008 Kleanthi Lakiotaki, Stelios Tsafarakis, and Nikolaos Matsatsinis proposed UTA-Rec [Lakiotaki, Tsafarakis & Matsatsinis, 2008][8]. UTARec is a Recommender System that incorporates Multiple Criteria Analysis methodologies.

4. Types of Recommender Systems

There are three major types of Recommender System:

4.1 Content-Based Recommender System (CB)

Content-Based Recommender System finds out about the client intrigue and serves separated recommendations for his needs. This sort of RS gives data to their client upon his prior quests, which put away in clients profile and gives different things and data sources.

Content-based RS, additionally acquainted with as exhaustive sifting, suggests things in light of a relationship between the substance of the things and a client's profile. The substance of everythingit appeared as an arrangement of inscription or terms, normally the words that come to pass in an archive. The client profile is appeared by similar terms and made up by breaking down the substance of data which has been seen by the client.

Numerous issues must be examined when executing a substance based separating System. Initially, terms can either be designated consequently or physically. At the point when terms are allocated naturally a technique must be picked that can separate these terms from things. Second, the terms must be spoken to with the end goal that both the client profile and the things can be thought about definitively. Third, an intellectual pecking order must be picked that can take in the client profile in view of seen questions and can make suggestions in light of this client information. The data source that substance based separating Systems are for the most part utilized with is content reports. A standard convention for term breaking down chooses single words from reports. The vector space display and inactive semantic ordering are two strategies that utilization these terms to speak to reports as vectors in a multidimensional space.

Importance criticism, hereditary calculations, neural systems, and the Bayesian classifier are among the learning strategies for taking in a client profile. The vector space display and idle semantic ordering can both be utilized by these learning strategies to speak to records.

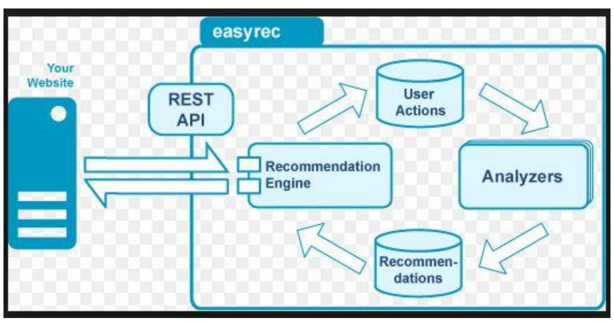

Figure 2: A basic representation how Content-Based Recommender System works (Das, 2019)[9].

4.2 Collaborative filtering Recommender Systems (CF)

Methods in this classification try to coordinate a client's normal nature over a thing as per what different clients have done previously. It begins by assessing an enormous number of client interchanges, evaluations, visits and different wellsprings of conduct and after that deliver a centralized server as indicated by these information. It at that point expects a client's temperament as per what other comparable clients and neighbour clients have done previously. The fundamental theory of CF is that a client may like concealed things on the off chance that it is loved by another watcher like him/her. In a generation System, the recommender yield would then be able to be characterized as, for point of reference 'individuals like you likewise preferred these things'. These systems are presently generally used to as a client and have legitimized as incredibly viable preliminary perusing and thus climb advertising. Be that as it may, keeping in mind the end goal to work adequately, they have to fabricate a rich ideal serving particular proposals and, that why, they require a lot of user generated info’s. One of the outputs of the decreasing amount of information is that CF can't find out things that no user has researched upon yet, the purported chilly things. Hence, the approach of different recommender Systems is to expose these things to users somehow, for instance, either by implementing them silently to a landing page or by implementing it in content-build moving with related to them diminishing in such way the number of frosty things in the database. While CF can accomplish brand new quality suggestions, it requires a type of a client profile to give proposals. It is, in this way, additionally difficult to apply it on sites that don't ask a client sign on, for example, CORE.

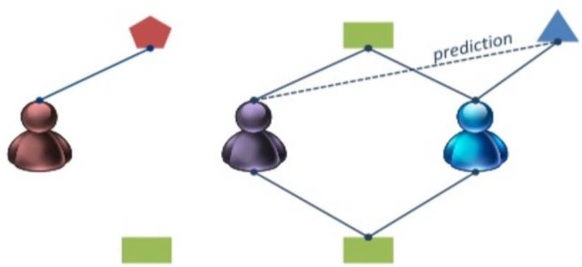

Figure 3: An illustration for how Collaborative Recommender system functions (Tjokro, 2019)[10]

4.3 Hybrid Recommendation System

Hybrid Recommendation Systems is a program wrapped with great highlights of various RS's to keep the confinements and the issues of a client’s inquiry or need. It additionally serves numeric and statistic procedures to demonstrate the best recommendations for clients request. Hybrid or mixed recommender Systems incorporated with at least two proposal procedures to climb better execution with least of the defect of an individual one. For the most part, community sifting is gathered with some other procedure in a shot to maintain a strategic distance from the increase issues.

4.3.1 **Weighted**

A weighted hybrid recommender System is one in which the score of a suggested thing is computed from the result of the greater part of the accessible proposal office exhibit in the System. For instance, the most straight forward joined hybrid would be a direct mix of proposal scores. The P-Tango System (Claypool et al. 1999) uses such a crossover. It first gives community oriented and content-based recommenders measure up to weight, however, bit by bit modifies the weighting as expectations about client appraisals are affirmed or disconfirmed. Pazzani's (Pazzani, 1999) mixed hybrid does not utilize numeric scores, yet rather treats the yield of each recommender (communitarian, content-based and statistic) as an arrangement of votes, which are then consolidated in an agreement plot.

The advantage of a weighted half and half is that the majority of the System's abilities are applied as a powerful influence for the suggestion procedure clearly and it is anything but difficult to perform post-hoc credit task and modify the cross breed as needs are. In any case, the verifiable suspicion in this system is that the relative estimation of the distinctive procedures is pretty much uniform over the space of conceivable things. From the discourse above, we realize this isn't generally so: a communitarian recommender will be weaker for those things with few raters.

4.3.2 **Switching**

A switching recommender system works in thing level affectability to the hybridization procedure: the System utilizes some rule to switch between suggestion methods. The DailyLearner System utilizes a substance/cooperative half and a half in which a substance based suggestion technique is utilized first. In the event that the substance based System can't make a suggestion with adequate certainty, at that point, a communitarian proposal is attempted. This exchanging hy does not totally maintain a strategic distance from the increasing issue; since both the cooperative and the substance based Systems have the "new client" issue. In any case, DailyLearner's substance based system is the closest neighbour, which does not require an extensive number of cases for precise grouping. What the communitarian strategy gives in an exchanging half and half is the capacity to cross types, to think of suggestions that are not shut semantically to the things past evaluated exceedingly, yet are as yet pertinent. For instance, on account of DailyLearner, a client who is occupied with the Microsoft against trust trial may likewise be keen on the AOL/Time Warner merger. Content coordinating would not probably prescribe the merger stories, but rather different clients with an enthusiasm for corporate power in the cutting edge industry might rate the two arrangements of stories profoundly, empowering the System to make the suggestion cooperatively. DailyLearner's hybrid has a "fallback" character – the transient model is constantly utilized first and the other procedure just becomes an integral factor when that strategy comes up short. Tran and Cohen (1999) proposed a more clear exchanging half and a half. In their System, the assertion between a client's past evaluations and the proposals of every method are utilized to choose the strategy to utilize for the following suggestion. Exchanging half and halves bring extra multifaceted nature into the suggestion procedure since the exchanging criteria must be resolved, and this presents another level of parameterization. In any case, the advantage is that the System can be touchy to the qualities and shortcomings of its constituent recommenders.

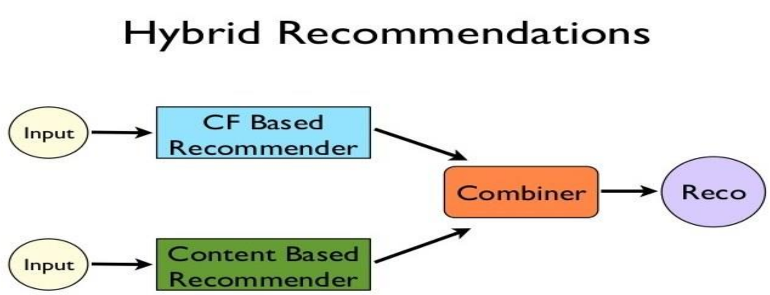

Figure 4: Hybrid Recommender System[11]

5. Application of Recommender Systems in Digital Libraries System

Recommender system is bow deployed in practice in various Information retrieval systems especially in digital library domain. Following table shows the Bird's eye view of the recommender system used in information retrieval systems.

Table: 1: Examples of Some Recommender Systems used in Digital Information Resource Systems

Digital Library Resources	Recommender systems Used	Type of Recommender systems
E-Prints	CoRe Recommender System	Hybrid Based RS
D-Space	Quambo Recommender Syestem	Content-based RS
OpenAIRE	Matchbook	Content-based RS
PubMedCentral	PURE, PubMedReco	Hybrid Based RS
CiteSeerX	Conceptual Impact-Based Recommender (CIBR)	Hybrid Based RS
JabRef	Mr.DLib (Scientific Recommender System)	Content-based RS

5.1 Recommender Systems for Digital Library System

Recommender systems are being used to develop the digital library system as well as other field of interest. The usability and functions of recommender systems are increasing day by day in different field of digital library systems. The RS are being applied according to the need of the user and the organization. Some of the recommender system are meant to work with repositories and databases. Here are some examples of the recommender systems which describes their need and functions.

5.1.1 The CORE (COnnecting REpositories) Recommendation System for EPrints

CORE, is a recommendation service that serves access to the colossal sum of inquiring about works, it helps to clients in finding articles concurring to what they examined. In the expansion, it moreover includes a recommender plug-in that can be introduced and bound together into a repository System, for the case, EPrints. When a store client sees an editorial page inside the store, the plugin sends to CORE information around the query. This may incorporate the identifier of the object and, when conceivable, its metadata, CORE at that point answers back to the store system and implants a list of recommended articles for perusing. These activities are produced by the CORE suggestion calculation.In light of the way that the CORE corpus is a vast database of records that mostly have content, it applies content-based sifting to deliver the rundown of recommended things. It changes the substance to an arrangement of term vectors and it finds comparative records by finding comparable vectors. The CORE Recommender does function in different areas, for example, on the CORE portal and in different institutional archives and diaries. From these spots, the recommender calculation gets data as information, for example, the identifier, title, creators, dynamic, year, source URL, and so forth. Furthermore, it tries to enhance these traits with extra accessible information, for example, reference tallies, number of downloads, regardless of whether the full-content accessible is accessible in CORE, and more related data. All these shape the arrangement of highlights that are utilized to locate the nearest archive in the CORE corpus.

5.1.2 Quambo for DSpace

Quambo is a recommender System tool which works as a plug-in for the DSpace open source database level. Quambo produces content based suggestions in light of a client chose set of cases, the way to deal with displaying content proposals to the client, and the encounters applying the System to an archive of specialized reports. It could be stuffed with the associate adjusted DSpace module to extend the thing space from which proposals can be made; a bigger thing space could enhance the assorted variety of the set from which to make suggestions

Late findings and overviews of the utilization of institutional databases in scholastic associations report that exploration researchers see little included an incentive infusing their organization's archive as a customary piece of their insightful exercises. It has built up an extra access to the DSpace open source store which creates and shows content proposals to make the institutional storehouse an all the more convincing asset for its clients.

Quambo's recommender System creates an arrangement of suggested things utilizing the Jaccard Similarity Index[12], modified to think about the weighting of where metadata fitting. Future research will look at the effectiveness of suggesting things from inside an alliance of stores and creating joint effort openings inside DSpace. Quambo deals with suggesting things inside a league; looking over a nearby reserve of metadata reaped from "selected" sources with a metadata gathering interface, and united inquiry against watchword seek prepared administrations.

5.1.3 Matchbook for OpenAIRE

Matchbook is a recommender service to enable keyword search for associations with most grounded records for financing and joint effort accomplishment, subsidized by a little grant (14.5k EUR) from OpenAIRE's Open Tender Calls. Matchbook can help forming consortia for scientific and logical foundations with distinguish potential accomplices with the correct disciplinary qualities and abilities they require, ranking them according their past achievement in anchoring financing determine their looks as per explicit national or worldwide funders or even individual subsidizing streams.

Matchbook functions as a fruitful evidence of-idea however would require greater improvement before having the capacity to be offered as an independent creation administration – specifically, the proposals are as of now made on a generally rare dataset (i.e., simply venture titles and watchwords, coordinated effort chronicles of organizations yet not constantly singular divisions) which implies that execution endures in contrast with comparative recommender administrations dependent on more extravagant data. A subsequent stage in this examination is include improved data - maybe including content mining of full undertaking depictions (likewise to the work-bundle level) and connecting the productions connected to ventures.

5.1.4 PURE for PubMed

PubMed article Recommender Systems (PURE), in light of content based filtering, has a web interface by which clients can include/erase their favored articles. When articles are enrolled, at that point PURE performs model-based grouping of the favored articles and prescribes the profoundly appraised articles by the forecast utilizing the prepared model. PURE updates the PubMed articles and reports the proposal by email on day by day base. This framework will be useful for scholars to lessen the time required for gathering data from PubMed.

5.1.5 PubMedReco for PubMed

PubMedReco is a recommender system for ongoing recommendations of medicinal articles from PubMed, a database of more than 23 million therapeutic references. PubMedReco can suggest medicinal article references while clients are chatting in a synchronous correspondence condition, for example, a talk room. Regularly, clients would need to leave their visit interface to open another internet browser window, and plan a fitting pursuit question to recover important outcomes. PubMedReco naturally produces the pursuit inquiry and shows pertinent references inside the equivalent incorporated UI. PubMedReco breaks down important catchphrases related with the discussion and utilizations them to scan for pertinent references utilizing the PubMed E-utilities programming interface. PubMedReco contributes by incorporate enhancements to the client experience for seeking PubMed from inside wellbeing discussions and visit rooms, and a machine learning model for distinguishing important watchwords

5.1.6 Conceptual Impact-Based Recommender (CIBR) for CiteSeerX

Conceptual Impact-Based Recommender (CIBR), is a hybrid recommender system which broadens the current conceptual recommender framework in CiteSeerx by including an express quality factor as a major aspect of the proposal criteria. To quantify quality, its framework considers the impact factor of each paper's creators as estimated by the creator's h-index. Investigations to assess the electiveness of hybrid frameworks how that the CIBR framework prescribes progressively applicable papers when contrasted with the conceptual recommender framework. Utilizing Conceptual Impact-Based Recommender (CIBR) in CiteSeerx Users can get applicable data from the database via looking by creator name and additionally keyword search. Clients may likewise get suggestions of papers they should need to peruse given by a current theoretical recommender framework. This framework suggests archives dependent on an automatically created client profile. Dissimilar to conventional substance based recommender system, the records and the client profile are spoken to as ideas vectors as opposed to catchphrase vectors and papers are prescribed dependent on theoretical matches instead of watchword coordinates between the profile and the data. In spite of the fact that the present framework gives suggestions that are on-subject, they are not really top notch papers.

5.1.7 Mr.DLib (Scientific Recommender System) for JabRef

Mr.DLib (Machine Readable Digital Library) is a free, open source, RESTful web based service that creates suggestions dependent on a solitary record and which offers different proposal approaches, for example, generalization based and content-based calculations with extra re-positioning utilizing bibliometric data. Also, Mr. DLib accumulates information about the suggestions made. These information are used to assess the different calculations, which over the long haul will advance scientific recommender frameworks.

Mr. DLib scientific recommender system is a plug-in into the JabRef reference manager which help clients to distinguish applicable papers out of tremendous measures of existing writing. They are especially valuable when utilized in mix with reference manager. Over 85% (Feyer, S., Siebert, S., Gipp, B., Aizawa, A., & Beel, J. 2017)[13] of JabRef clients expressed that they would value the coordination of a recommender system. In any case, the usage of writing recommender system requires involvement and assets that little organizations can't afford. Mr.DLib scientific recommender framework has been coordinated into JabRef with remembering the wants of the clients. Utilizing Mr.DLib's recommendations-as-an service, JabRef client's can find significant writing and keep themselves educated about the best in class in their particular fields.

6. Conclusions

Recommender Systems (RS) play a significant and catalytic role for any information retrieval system especially in digital library domain. Deployment of RS technologies in digital library brings many added advantages at users end. By using RS system in digital library system users can get more precise and relevant results as well as find a complete deferent searching experience powered by cognitive model of information retrieval.

References

[1] Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261-273.

[2] Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: introduction and challenges. In *Recommender systems handbook* (pp. 1-34). Springer, Boston, MA.

[3] Jussi Karlgren. (2019). A digital bookshelf: original work on recommender systems. [online] Available at: https://jussikarlgren.wordpress.com/2017/10/01/a-digital-bookshelf-original-work-on-recommender-systems [Accessed 11 Jan. 2019].

[4] Sharma, R., & Singh, R. (2016). Evolution of recommender systems from ancient times to modern era: a survey. Indian Journal of Science and Technology, 9(20), 1-12.

[5] O'Donovan, J., & Smyth, B. (2005, January). Trust in recommender systems. In Proceedings of the 10th international conference on Intelligent user interfaces (pp. 167-174). ACM.

[6] Recio-García, J. A., Díaz-Agudo, B., & González-Calero, P. A. (2008, October). Prototyping recommender systems in jcolibri. In Proceedings of the 2008 ACM conference on Recommender systems (pp. 243-250). ACM.

[7] Resnick, P., & Sami, R. (2007, October). The influence limiter: provably manipulation-resistant recommender systems. In Proceedings of the 2007 ACM conference on Recommender systems (pp. 25-32). ACM.

[8] Lakiotaki, K., Tsafarakis, S., & Matsatsinis, N. (2008, October). UTA-Rec: a recommender system based on multiple criteria analysis. In Proceedings of the 2008 ACM conference on Recommender systems (pp. 219-226). ACM.

[9] Das, S. (2019). *Beginners Guide to learn about Content Based Recommender Engine*. Retrieved from https://www.analyticsvidhya.com/blog/2015/08/beginners-guide-learn-content-based-recommender-systems/ [Accessed 11 Jan. 2019].

[10] Tjokro, M. (2019). Building Spotify's "Discover Weekly" with Spark – Towards Data Science. Retrieved from https://towardsdatascience.com/building-spotifys-discover-weekly-with-spark-4370d5d0df2f [Accessed 05 Jan. 2019].

[11] An Introduction to Recommendation Engines. (2015). Retrieved from https://dataconomy.com/2015/03/an-introduction-to-recommendation-engines/ [Accessed 05 Jan. 2019].

[12]En.wikipedia.org. (2019). *Jaccard index*. [online] Available at: https://en.wikipedia.org/wiki/Jaccard_index [Accessed 11 Jan. 2019].

[13] Feyer, S., Siebert, S., Gipp, B., Aizawa, A., & Beel, J. (2017, April). Integration of the scientific recommender system Mr. DLib into the reference manager JabRef. In *European Conference on Information Retrieval* (pp. 770-774). Springer, Cham.