Advances in Computational Research
ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 4, Issue 1, 2012, pp.-91-94
Available online at http://www.bioinfo.in/contents.php?id=33

A LIBRARY RECOMMENDER SYSTEM USING COSINE SIMILARITY MEASURE AND ONTOLOGY BASED MEASURE

SHIRUDE S.B.* AND KOLHE S.R.

Department of Computer Science, North Maharashtra University, Jalgaon, MS, India.
*Corresponding Author: Email- snehalata.shirude@gmail.com

Received: February 21, 2012; Accepted: March 06, 2012

Abstract- We have proposed a recommender system UseLibrary, helping users of library to get proper journal articles, books and other library resources. Profile generation & maintenance and profile exploitation are two main tasks to develop the system. Currently in this system, dataset consists of 790 books and journal articles in PDF and TXT format. System learns the profiles of users and then provides recommendations. Information filtering is required for providing recommendations. Similarity between interest of user and available resources can be computed by different approaches. In this paper, we have compared two approaches, cosine similarity measure and ontology based measure. It is found that use of ontology based measure outperforms cosine similarity measure
Keywords- Recommender System, Cosine similarity measure, ontology, semantic concepts, digital library, filtering, similarity analysis, content based filtering

Citation: Shirude S.B. and Kolhe S.R. (2012) A Library Recommender system using cosine similarity measure and ontology based measure. Advances in Computational Research, ISSN: 0975-3273 & E-ISSN: 0975-9085, Volume 4, Issue 1, pp.-91-94.

Introduction

Recommender system is valuable software which is learning the interest of user, filtering resources according to the interest and then recommending resources to user. University library is a rich source of information for academicians. Recommender system can provide help to academicians to acquire proper resources from library. Development of such system has to generate profile of users, and then update profile whenever necessary. Then by learning profile, system can obtain interest of user. To provide recommendations to users system has to filter available resources. Content based, collaborative and hybrid are three approaches can be used. For recommendation information filtering is required. Similarity between interest of user and available resources can be computed by different approaches. Cosine similarity is generally used [18]. Use of ontology is also done in development of such systems [22]. In reference ontology there is explanation for the entities that exists in the world. Using this personal ontology is created and then recommendations are provided according to percentage of interest to different topics in ontology [21] [24]. Topics of research paper can be represented in the form ontology and then recommendations are provided by calculating interest [23]. In this paper, two approaches are experimented. In first experiment, the vector of interest is compared with all available resources in library using cosine similarity measure. In second experiment, we used ontology based approach. Domain ontology for the subject computer science is developed which consists of hierarchy of different concepts, terms related to concepts and keywords. Personal ontology is not created. This paper is organized in 8 sections. In section 2 & 3, system architecture and dataset used are described. Section 4 describes recommendations approach. Section 5 and 6 are giving comparison between approaches of computing similarity such as cosine similarity measure and ontology based measure while providing recommendations. Results are given in section 7.

System Architecture

Development of recommender system for efficient use of libraries is proposed. The objectives of the proposal are to help users of

library for finding the right book, relevant research paper/s, organizing sequence of lesson plan or book plan, developing relevant student group, filtering recommendations, etc. The block diagram shows the architecture for proposed recommender system.

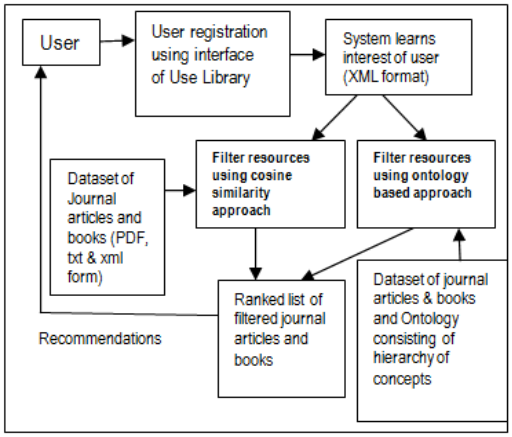

Fig. 1- System Architecture

Dataset Creation

The dataset of recommender system consists of user profile's data, journal paper's data, and various books' data. We have referred the structure of dataset from ACL Anthology Reference Corpus (ACL ARC). The ACL ARC is constructed from a significant subset of ACL anthology, a digital archive of conference and journal papers in natural language processing and computational linguistics [18]. Currently, in our dataset we have 575 journal papers and 215 books related to computer science. The journal papers and books are in PDF format. In TXT format we have abstracts of journal papers and index & keywords of books. User profiles are in XML format. The general syntax of XML file is as follows:

<User ID = " "> <FirstName> </FirstNAme> <LastName> </LastName> <MemberType> </MemberType> <Department> </Department> <Course> </Course> <Subject> </Subject> <EmailAddress> </EmailAddress> <ResearchTopic> </ResearchTopic> <PublishedWork> </PublishedWork> </User>.

In dataset, we also have storage of syllabi of various courses related to Computer Science subject [20].

Recommendation Approaches

System stores user profiles in XML form. Using XML form, a vector is created holding interest of user. Then system requires filtering from available resources according to interest of user. Currently, the system UseLibrary uses content based filtering. To find that, which resources are similar to the interest of user we use cosine similarity measure [18]. All resources of library are also stored in the form of vectors. In ontology based approach, we have used ontology representing semantic concepts and related terms [21-24].

Cosine Similarity Measure

We have vector I, in which interest of the user is stored. Each resource such as journal and book is represented in the form of vectors J and B respectively. Now system computes cosine similarity [18] as $COSSIM(I,J)=\frac{I \bullet J}{|I| \bullet |J|}$ and $COSSIM(I,B)=\frac{I \bullet B}{|I| \bullet |B|}$. UseLibrary then recommends the user the journal articles and books in decreasing values of cosine similarity. Interest of user is obtained by using published work of user if he/she is a research student. And if he/she is a postgraduate student then interest is obtained by referring syllabus of the course for which student have taken admission. By using TF-IDF scheme, we obtain journal articles and books which are containing more number of similar terms compared with interest vector. Due to this we are not able to get the resource which is semantically related to the interest of user.

Ontology based Measure

For this, ontology [22] is designed using Protégé editor which is representing semantic concepts and related terms. C1, C2, C3… Cm are all semantic concepts. Each semantic concept is annotated by set of related terms T1, T2, T3,.., Tn. Each term is the list of keywords K1, K2, K3,…, Kl. Now a specific list of keywords is annotated to one or more class of semantic concepts. We have vector I containing interest of user. System extracts related term/s Ti that matches with I using cosine similarity [18] such as

$$\mathrm{Re}latedTerm = similarity(I,Ti) = \frac{\sum_{i=1}^{k} Wi \bullet \sum_{j=1}^{n} Tj}{\sqrt{\sum_{i=1}^{k} Wi^2} \bullet \sqrt{\sum_{j=1}^{n} Tj^2}}$$

Here I is the interest of user and Wi are keywords consists in vector I. Then concepts which are annotated by RealtedTerm and also the subclasses in the hierarchy of ontology are retrieved [20]. Once we get list of concepts, system filter those resources which are close to all those concepts. This way system is able to filter the resources which are semantically related to the interest of user.

Results

To evaluate the performance of both approaches we used precision, recall, and F1-measure. Precision is the fraction of the documents retrieved that are relevant to the user's information need. Recall is the fraction of the documents that are relevant to query that are successfully retrieved. The weighted harmonic mean to precision and recall is F-measure [14]. This is known as F1-measure when recall and precision are evenly weighted.

$$\mathrm{Pr}ecision = \frac{Numberof\ \mathrm{Re}levent\ \mathrm{Re}trievedDocuments}{TotalNumberof\ \mathrm{Re}trievedDocuments},$$

$$\mathrm{Re}call = \frac{Numberof\ \mathrm{Re}levantNot\ \mathrm{Re}trievedDocuments}{TotalNumberof\ \mathrm{Re}levantDocuments}$$

and

$$F1 = \frac{2 \times \mathrm{Pr}ecision \times \mathrm{Re}call}{(\mathrm{Pr}ecision + \mathrm{Re}call)}.$$

To measure performance we considered sample data of 185 research journal papers and 25 books from the field of Artificial Intelligence in Computer Science. We have registered 10 users using UseLibrary. Table 1 shows data about documents retrieved and not retrieved for 10 users when only cosine similarity is used.

Table 1. Experimental results for pure Cosine Similarity

	No. of relevant retrieved doc. N_{rr}	No. of relevant not retrieved doc. N_{rn}	No. of irrelevant retrieved doc. N_{ir}	Total no. of re-trieved doc. N_{ret}	Total No. of relevant doc. N_{rel}
User1	24	5	2	26	29
User2	2	4	2	4	6
User3	12	2	15	27	14
User4	13	3	18	31	16
User5	9	1	1	10	10
User6	5	2	7	12	7
User7	15	4	9	24	19
User8	5	5	6	11	10
User9	30	6	18	48	36
User10	5	3	3	8	8

Table 2 shows data about documents when ontology based approach is used.

Table 2. Experimental results for ontology based approach

	No. of relevant retrieved doc. N_{rr}	No. of relevant not retrieved doc. N_{rn}	No. of irrelevant retrieved doc. N_{ir}	Total no. of re-trieved doc. N_{ret}	Total No. of relevant doc. N_{rel}
User1	21	8	2	23	29
User2	7	1	1	10	8
User3	19	2	7	26	21
User4	21	2	9	30	23
User5	9	2	2	11	11
User6	9	5	5	12	14
User7	12	3	6	18	15
User8	11	3	5	17	14
User9	27	7	16	42	34
User10	9	4	4	15	13

Table 3. Experimental results for both approaches.

	Precision (%)		Recall (%)		F1 (%)	
	Only Cosine Similarity approach	Ontology based approach	Only Cosine Similarity approach	Ontology based approach	Only Cosine Similarity approach	Ontology based approach
User1	92.31	91.30	82.76	72.41	87.27	80.77
User2	50.00	70.00	33.33	87.50	40.00	77.78
User3	44.44	73.08	85.71	90.48	58.57	80.85
User4	41.94	70.00	81.25	91.30	55.32	79.25
User5	90.00	81.82	90.00	81.82	90.00	81.82
User6	41.67	75.00	71.42	64.29	52.63	69.23
User7	62.50	66.67	78.95	80.00	69.77	72.73
User8	45.45	64.71	50.00	78.87	47.62	70.97
User9	62.50	64.29	83.33	79.41	71.49	71.05
User10	62.50	60.00	62.50	69.23	62.50	64.29
Average	59.34	71.69	71.92	79.50	63.51	74.87

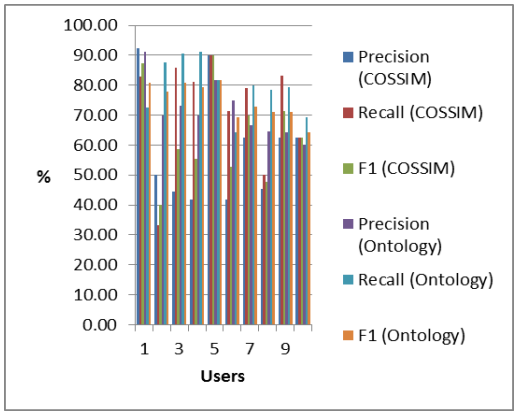

Fig. 2- shows a graph with these values for all 10 users

Conclusion

The experiment shows that recommender system can help academicians for acquiring library resources. We are developing such recommender system for efficient use of library, UseLibrary. Currently, system initially registers users and obtains interest of users and stores in XML format. Then to filter available resources two approaches are experimented, pure cosine similarity measure and ontology based approach. Cosine similarity computes how close the journal article and book is from interest of user. System ignores some relevant papers and books due to reason that they do not contain that many numbers of terms matching with the interest vector. So some approach in which concepts should be retrieved related to interest is necessary. Ontology allows defining such hierarchy of concepts. Using this approach when system filters journal articles and books to recommend the result is 71.69% which is much better than the cosine similarity approach.

References

[1] O'Connor M. and Herlocker J. (2001) *SIGIR*2001 *Workshop on Recommender Systems New Orleans LA*.
[2] Stuart E. Middleton, Harith Alani, David C. De Roure (2002) *Semantic Web Workshop*.
[3] Miquel Montaner, Beatriz Lopez and Josep Lluis DelaRosa (2003) *Artificial Intelligence Review* 19, 285-330.
[4] Kamal Yammine Mohammed A. Razek, Esma Aimeur and Claude Frasson (2004) *ITS* 2004, LNCS 3220, 720-729.
[5] Yan Yang, Jian Zhong Li. (2005) *Journal of Computer Science* 1(1), 40-46.
[6] Adomavicius G. (2005) *IEEE transactions on Knowledge and Data Engineering*, 17(6).
[7] Rosario Sotomayor, Joe Carthy and John Dunnion (2005) *American Association for Artificial Intelligence*.
[8] Nigel Shadbolt and Wendy Hall University of Southampton Tim Berners-Lee, Massachusetts Institute of Technology (2005) *IEEE Intelligent Systems*.
[9] Andre Vellino, David Zeber (2007) *IEEE/WIC/ACM International conference on Web Intelligence and Intelligent Agent Technology*.
[10] Cantador I., Fernandez M., Castells P. (2007) *First International Workshop on Social and Collaborative Construction of Structured Knowledge*.
[11] Ming Zhang, Weichum Wnag, Xiaoming Li (2008) *ICADL, LNCS* 5362.
[12] Jamil Itmazi, Miguel Megias (2008) *The International Arab Journal of Information Technology*, 5(3).
[13] Porcel C., Moreno J.M., Herrera-Viedma E. (2009) *Expert systems with Applications* 36, 12520-12528.
[14] Porcel C., Moreno J.M., Herrera-Viedma E. (2009) *EDULEARN*09 *Conference*.
[15] Porcel C., Herrera-Viedma E. (2009) *IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*.
[16] Juan A. Recio-Gracia, Belen Diaz-Agudo, Derek Bridge, Pedro A. Gonzalez-Calero (2010) *Expert Update*,10(2).
[17] Andre Vellino (2010) *ASSIST*.
[18] Kazunari Sugiyama, Min-Yen Kan (2010) *ACM*.
[19] Kuroiwa Takanori, Bhalla Subhash (2010) *Int. Journal of Computational Science and Engineering* 5(3-4) 207-213.

[20] Snehalata B. Shirude, Satish R. Kolhe (2011) *National Conference on Advances in Computing*, 177-179.
[21] I-En Liao, Shu-Chuan Liao, Kuo-Fong Kao and Ine-Fei Harn (2006) *ICADL*, 173-182.
[22] Nu´ ria Ferran, Enric Mor and Julia` Minguillo´n, *Library Management* 26, 206-217.
[23] Stuart E. Middleton, Harith Alani, David C. De Roure (2002) *Semantic Web Workshop.*
[24] Shu-Chuan Liao, Kuo-Fong Kao, I-En Liao, Hui-Lin Chen, Shu -O Huang (2009) *Electronic Library*, 27 (3), 96-508.