International Journal of Advanced Computer Technology (IJACT)

ISSN:2319-7900

A Paper Recommender System Based on the Past Ratings of a User

Simon Philip[1]
Department of Computer Science
Federal University Kashere, Gombe, Nigeria
simonphilip102@yahoo.com

P.B. Shola (PhD)[2]
Department of Computer Science
University of Ilorin, Ilorin, Nigeria

E.P.Musa[3]
Department of Computer Science
Ramat polytechnic Maiduguri, Nigeria

ABSTRACT: Research paper recommender systems are software applications or systems that help individual users to find the most relevant research papers to their needs or tastes. These systems use filtering techniques to generate recommendations. These techniques are categorized majorly into collaborative-based filtering, content-based technique, and hybrid algorithm. Content-based filtering technique is adopted because of the availability of contents describing the research papers. The research papers were represented as vectors of weights using vector space model and TF-IDF weighing scheme. Cosine similarity was used as the function to compute the degree of similarity between research papers (represented as vectors of weights). In this paper, we proposed a recommendation algorithm based on the past ratings of an active user. The system does not provide recommendations to an active user based on the past ratings of other similar users. Also, this paper presents an experimental implementation of the proposed algorithm.

Keyword: Recommender Systems, Content-Based technique, TF-IDF, Cosine Similarity, Vector Space Model

1: Introduction

The capacity of computers to provide recommendations was recognized fairly early in the history of computing. Grundy [1], a computer-based librarian, was an early step towards automatic recommender systems.

Recommender systems have become extremely common in recent years. They are applied in a variety of applications. The most popular ones are probably movies, books, news, music, research papers, social tags, and products in general [2]. Most recommender systems typically produce a list of recommendations in one of the three ways- through collaborative-filtering technique, content-based technique, and hybrid algorithm [3].

Content-based recommender systems provide recommendations by comparing the content describing the items and the content describing the interest of the users. It also works by recommending items that are similar to the items a user has liked in the past [4]. The similarity of items is computed or calculated based on the features associated with the items being compared [5]. For example, if a user has liked a paper in the past, and the paper belongs to data mining research area, then the system can learn to recommend or suggest papers from this area.

According to [6, 7, 8], content-based recommender systems use the knowledge of the user's past experience (e.g. items bought in the past in the case of e-commerce) and the knowledge on items in general (items description). This type of recommender systems does not depend on the past ratings of other users to provide recommendations to an active user.

Content-based recommender systems use the past ratings of an active user and an appropriate filtering technique in finding relevant or desired items to the intended user. These systems assist users in overcoming the problem of information overload by suggesting or providing recommendations to the users based on their past ratings [9].

2- Related Work

GroupLens created MovieLens recommender system in 1997. MovieLens is a non-commercial recommender system and virtual community website that recommends films for its

ISSN:2319-7900

users to watch, based on their film preferences and using collaborative filtering. The website bases its recommendations on what the user provides to the website, including films the user likes, and what films other users with similar tastes prefer. When a user joins the website, they are given several randomly chosen movies and told to rate them from one to five stars, five being the best. The system then compares the user's ratings to those of other users with similar tastes, and then accordingly recommends films that the user has not yet seen [10]

Reference [11] applied content-based approach in research paper recommender system for a digital library. The authors used content-based technique as the recommendation technique. The system provides recommendations to users based on their taste or needs supplied as a query. The users' queries and research papers were represented as vectors of weights using vector space model and TF-IDF weighing scheme. Cosine similarity was used to determine similar papers to the users' queries (supplied as input). This system did not use the ratings of an active user to provide recommendations. The authors also proposed a recommendation algorithm based on users' query.

Reference [12] designed a paper recommendation system. The author used content-based filtering technique as the recommendation technique. The author used Jaccard similarity coefficient to compute similarity between users' query (users' attributes) and the attributes of the papers. The recommendations suggested by the system were sent via email to the intended users.

Reference [13] applied case-based recommendation technique in a restaurant recommender system. The adopted technique was used to select and rank restaurants. It was implemented to serve as a guide to attendees of the 1996 democratic national convention in Chicago and operated as a web utility.

Reference [14] designed a group recommender system for Facebook. The authors used hierarchical clustering and decision techniques to suggest or recommend the most suitable Facebook group (s) to Facebook users. He extracted profile information of the Facebook members at University of North Texas and used it as a test data.

3: Methodology

Content-based approach is adopted for the design and implementation of the research paper recommendation system based on the past ratings of an active user. This approach does not depend on the ratings of other users but uses the contents describing the items and the users' taste or needs. The researchers used vector space model (as an information retrieval model) and TF-IDF weighing scheme to represent research papers as vectors of weight. Cosine similarity was also used to determine the most relevant papers to the papers an active user has liked or rated in the past.

- Research paper Representation: The research papers are represented by a set of features. These features are: title of the paper, abstract, keywords, publisher, research area, and the authors
- Information Retrieval Model: Keyword-based vector-space model (with basic TF-IDF weighing scheme) was used to represent a research paper as a vector of weights, where each weight indicates the degree of association between a research paper and a term or keyword.
- TF-IDF: TF-IDF was used to determine the weight of a keyword or a term in a research paper i.e. how important a keyword or a term in a document. TF-IDF is given by:

$$tf - idf(t, d, D) = tf(t, d) \times idf(t, D) \quad (1)$$

Where:

t=term in the user's query, d= a document in the collection, D= a collection of documents

TF=Term Frequency given by:

$$tf(t, d) = \frac{N_{t,d}}{N_d} \quad (2)$$

IDF= Inverse Document Frequency which is given by:

$$idf(t, D) = log \frac{N}{|d \in D : t \in d|} \quad (3)$$

Where:

N = number of documents in the collection

N_d= Number of terms in the document d

$N_{t,d}$ = Number of times term t appears in document d

- Cosine similarity is a function that computes the degree of similarity between vectors. This method was used to determine how similar a research paper is to a paper that an active user has liked in the past. Given two research papers p_j, p_k represented as vectors of weights, their similarity is measured by:

$$\text{Sim}(p_j, p_k) = \frac{\overrightarrow{p_j}.\overrightarrow{p_k}}{|\overrightarrow{p_j}|.|\overrightarrow{p_k}|} = \frac{\sum_{i=1}^{n} w_{i,j}.w_{i,k}}{\sqrt{\sum_{i=1}^{n} w_{i,j}^2}\sqrt{\sum_{i=1}^{n} w_{i,k}^2}} \quad (4)$$

Where:

$W_{i,j}$ = Weight of term *i* in paper *j*

$W_{i,k}$ = Weight of term *i* in paper *k*

4: A Proposed Recommendation Algorithm Based On Papers an Active User Has Liked in The Past

The proposed algorithm for generating recommendations based on the papers an active user has liked in the past is shown below:

Note: According to reference [4], "Content-Based Recommender Systems also work by recommending items that are similar to the items a user has liked in the past"

i. START
ii. Get the paper (P_j) liked by an active user.
iii. Get the keywords describing P_j
iv. Determine the weight of each keyword using **TF-IDF** in paper P_j: The weight obtained represents the importance of a keyword in the document or paper P_j
v. Represent paper P_j as a vector of weights: The current paper P_j becomes a vector i.e. an array of floating points (weights)
vi. Retrieve research papers that are similar to the current research paper p_j to form a collection or corpus C
vii. FOR k=1 to N
    a. Determine the weight of each keyword in paper K(**using TF-IDF**), and represent the paper as a vector of weights
    b. Compute the similarity of paper K to paper P_j by using

$$\text{Sim}(P_k, P_j) = \frac{\sum_{i=1}^{n} w_{i,j}.w_{i,k}}{\sqrt{\sum_{i=1}^{n} w_{i,j}^2}\sqrt{\sum_{i=1}^{n} w_{i,k}^2}}$$

    c. Sim_Values[p_k]= $\text{Sim}(P, P_j)$

Note: *Sim_Values* is an associative array containing the similarity values of all relevant papers to P_j.

NEXT

viii. Sort the associative array *Sim_Values* in descending order with respect to similarity value
FOR p=1 TO N
    IF Similarity Value of "p" >=0.3 THEN
    Store and retrieve the details of paper 'p', and display it
    END
NEXT

ix. END

Note: P_j = the current documents or research paper liked by an active user.

N= number of documents or research papers that are similar to P_j.

$W_{i,j}$ = Weight of term i in paper J

$W_{i,k}$ = Weight of term i in paper K

Sim_Values = An associative array containing the similar papers in order of their similarity to P_j.

5: Implementation and Results

5.1 Dataset for the System

Sources of the research papers are the research papers published by academic staff of federal university kashere, and also the ones from open sources obtained on the internet.

5.2 Implementation

The system was implemented using PHP and MySQL technology. MySQl was used to design the database to store the details of the research papers. PHP was used to implement the above proposed recommendation algorithm.

ISSN:2319-7900

The choice of this Scripting language is based on the fact that digital libraries are developed as web applications.
The result of the implementation is discussed in section 5.3.

5.3: Results

The proposed recommendation algorithm was implemented in a web-based paper recommendation system. Section "a" presents the test result of the system while section "b" presents how the system works.

a. Test result. This section contains the test result of the system. Table 1 shows a list of paper recommendations and their similarity values computed using cosine similarity function. A paper titled "Development of software engineering ontology for Multi-site Software development" was rated by a user and the system provided or generated some recommendations (shown in table 1) based on their similarity values. The system displays the recommendations in descending order of their similarity values i.e. the paper that carries the highest similarity value would appear first in the list.

Table 1: Paper recommendations and their similarity values

S/N	Title	Similarity Value
1	Formalism and intuition in software engineering	0.54433
2	Aspects of abstraction in software development	0.53522
3	Engineering and software engineering	0.51344
4	Selecting empirical methods for software engineering	0.47140
5	The name and nature of software engineering	0.44721
6	Software engineering for self-adaptive systems: A research roadmap	0.33333
7	Some note on models and modeling	0.33333

b. How the recommender system works: Section I, II, and III describe how the recommender system works.

I. Users' Login Page: To have access to the system beyond the start page, users' authentication is necessary for security reasons and to deny unauthorized users from using the system. Hence, the login form requires a valid username and password to be able to access the system beyond the start page. The snapshot in figure 1 allows the users to login to the system.

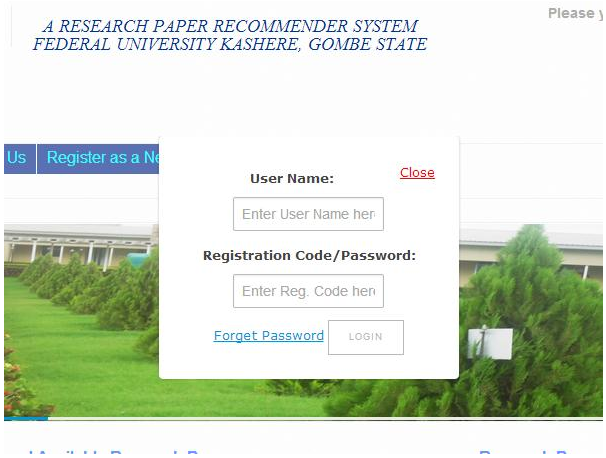

Fig. 1: Users' Login page

II. Users' Recommendation Page Based On The Papers An Active User Has Liked In The Past

After a successful logon to the system, the active user clicks on *my recommendations* link to view the recommendations generated by the system based on the papers the active user has liked in the past. The snapshot in figure 2 shows some recommendations displayed to an active user.

Fig. 2: Users' Recommendation page based on the papers liked in the past

III. Users' Recommendation Page Based On The Current Paper An Active User Likes

When a user searches for research papers in the system, the system displays papers based on the query supplied by the user. The user can like or dislike a paper from the ones displayed. When the user likes a paper, a confirmation message is displayed asking the user to confirm the action. If the user clicks on the OK button, the system suggests or generates paper recommendations based on the current paper being liked by the user. The snapshot in figure 3 shows some papers recommended to the user based on the current paper liked by the user.

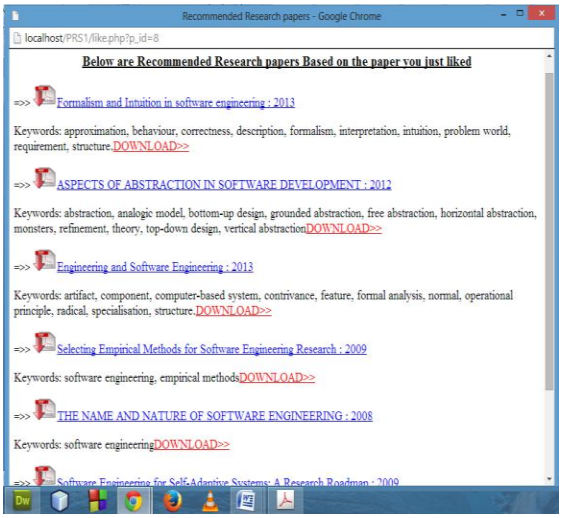

Fig. 3: Users' Recommendation page based on the current paper being liked

6: Conclusion and Future Work

In digital libraries, users are presented with too many options from which users choose the ones that are most relevant to their needs or tastes. This leads to information overload i.e. users have too many options. With the help of recommender systems, users are suggested or provided with most relevant recommendations that meet their tastes or needs.

This paper presents a paper recommender system that suggests or provides recommendations to the intended users based on the papers the users have liked in the past. This paper adopted content-based filtering technique to generate recommendations to the intended users. The system does not provide recommendations to an active user based on the past ratings of other similar users to the active user.

This paper also proposed a recommendation algorithm based on the papers a user has liked in the past. The result of the implementation of the proposed algorithm was compared with the result of the system in reference [33], and found to give better recommendations. This means that users get quality recommendations based on the papers the user have liked in the past compared to the one obtained when users present their tastes or needs in form of a query.

The next step of our future work is to adopt hybrid algorithm that combines both collaborative filtering technique and content-based approach. The researcher also recommends that the proposed recommendation algorithm (based on the past ratings of an active user) be evaluated to determine its effectiveness in providing recommendations.

References

[1] E.Rich, :User Modeling via stereotypes, Cognitive science, vol. 3, No.4, pp.329-354, October, 2010

[2] P. Gupta, A. Goel, J. Lin, A. Sharma, D.Wang, and R. B. Zadeh WTF; The who-to-follow system at Twitter, Proceedings of the 22nd international conference on World Wide Web

[3] H. Jafarkarimi, A.T.H. Sim and R. Saadatdoost A Naïve Recommendation Model for Large Databas

ISSN:2319-7900

es, International Journal of Information and Educa tion Technology, June 2012

[4] R. J. Mooney and L. Roy: Content-Based book recommendation using learning for text categoriza tion. In proceedings of the fifth ACM conference on digital libraries, pages 195-204, San Antonio, TX, June 2000.

[5] F. Ricci , L. Rokach, B. Shapira; Introduction to Recommender Systems Handbook,, Springer, 2011, pp. 1-35

[6] S.Rendle,K.Tso-Sutter,W.Huijsen, C.Freudenthaler, Z.Gantner, C.Wartena, R.brussee and M.Wibbels,: Report on State of the Art re commender Algorithms (Update), MyMedia pub lic Deliverable D4.1.2, March, 2011

[7] G.Salton,:Automatice text processing. The Trans formation, Analysis, and Retrieva of Information by Computer, Addison Wesley, 1989

[8] R.Baeza-yates and B.Ribeiro-Neto,: Modern Infor mation Retrieval. Addison Wesley, May, 1999

[9] Burke, R.: Hybrid Recommender Systems: Survey and experiment. User modeling and user-adaptive interaction pages 331-370 , November, 2002

[10] http://www.movielens.com

[11] S. Philip, P. B. Shola, and A. O. John; Application of Content-Based Approach in a Research Paper Recommendation System for a Digital Library, International Journal of Advanced Computer Science and Application, Volume 5, Issue 10, 2014

[12] B.S. Oladapo: A research paper recommender Sy tem, 2013 unpublished.

[13] J.L.Kolodner:A Restaurant Recommender System, 1993

[14] E.Baatarjav, J.Chartree, and T. Meesumrarni: Group Recommendation System for facebook, 2010.

Biographies

Simon Philip is a Graduate Assistant in Departmengt of Computer Science, Federal University Kashere. He obtained his B.Sc Computer Science at Gombe State University, Nigeria. He is currently undergoing his Master programme at University of Ilorin, Nigeria. He has finished his master's thesis and awaits his final examination.

Dr. P.B. Shola is an Associate Professor in Department of Computer Science, University of Ilorin, Nigeria. He obtained his PhD at University of Essex, Essex, England, United Kingdom. He is a member of Computer Society of Nigeria.

E. P. Musa, is a Senior Lecturer and Head of Computer Science Department, Ramat Polytechnic, Maiduguri, Borno State, Nigeria. He has MSc in IT and a Member of Nigeria Computer Society.