Personalized Paper Recommender System for Researchers

Prepared By

Pradip Makwana

12MCEC41

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2014

Personalized Paper Recommender System for Researchers

Major Project

Submitted in partial fulfillment of the requirements

For the degree of

Master of Technology in Computer Science and Engineering

Prepared By

Pradip Makwana

(12MCEC41)

Guided By

Prof. Sapan H. Mankad

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2014

Certificate

This is to certify that the Major Project Report entitled “**”Personalized Paper Recommender System for Researchers”** submitted by **Pradip Makwana (Roll No: 12MCEC41)**, towards the partial fulfillment of the requirements for the degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project part-I, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Sapan H. Mankad
Guide & Assistant Professor,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Prof. Vijay Ukani
Associate Professor
Coordinator M.Tech - CSE
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Sanjay Garg
Professor and Head,
CSE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr K Kotecha
Director,
Institute of Technology,
Nirma University, Ahmedabad

Undertaking for Originality of the Work

I, **Pradip Makwana**, Roll. No. **12MCEC41**, give undertaking that the Major Project entitled "**Personalized Paper Recommender System for Researchers**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Nirma University, Ahmedabad, is the original work carried out by me and I give assurance that no attempt of plagiarism has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student
Date:
Place:

Endorsed by
Prof. Sapan H. Mankad
(Signature of Guide)

Acknowledgements

I owe this moment of satisfaction with deep sense of gratitude to **Prof. Sapan H. Mankad**, CSE Department, Institute of Technology, Nirma University, Ahmedabad for his guidance and whole hearted inspiration has been of greatest help to me in bringing out this work in its present shape. The direction, advice, discussion and constant encouragement by him has been so helpful that it enabled me to complete the work successfully.

My deepest thanks is extended to **Prof. Vijay Ukani**, PG CSE - Coordinator, Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad for an exceptional support and continual encouragement throughout the Major Project.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr K Kotecha**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also like to thanks all faculty members for their support cooperation in this project. I express my sincere gratitude to my consulted faculties for giving me the different aspects the topic and helped in every steps performed during the project work.

At last but not least, I thank to all those have contributed during the preparation of the topic. We thank our friends and guides for their support throughout the project work.

- Pradip Makwana

12MCEC41

Abstract

Today, there is a huge research work going on personalized recommendation. Social network being more popular all over the word and emerged to fulfill the need of users in the different fields. Here we discuss on personalized recommendation of researchers who are working in different domain and related papers. We also show that how social network is useful to meet users need. From the perspective of the researchers, we can conclude that having more similar the research topics of researchers are, the stronger is their similarity in the preferences. In our procedure, we firstly extract keywords which are representing the researchers then after using tf values of the keyword measure similarity between researcher and documents. And base on this matter we defiantly define that if the researchers have similar research topics, the stronger is their context similarity in the preferences.

Contents

Certificate **iii**

Undertaking **iv**

Acknowledgements **v**

Abstract **vi**

List of Tables **viii**

List of Figures **ix**

1 Introduction **1**

1.1 Recommendation System 1

1.2 Personalized recommendation system 2

1.3 Personalized Paper Recommendation System 3

1.4 Organization of Thesis 3

2 Related Work **5**

3 Problem Statement **9**

4 Proposed Work **10**

5 Result Observation **18**

6 Conclusion **22**

List of Tables

4.1 Top 10 similar results of Researcher 1 11

4.2 Top 10 similar results of Researcher 2 11

4.3 Top 10 similar results of Researcher 3 12

4.4 Top 10 similar results of Researcher 1 14

4.5 Top 10 similar results of Researcher 2 14

4.6 Top 10 similar results of Researcher 3 15

4.7 Top 10 similar results of Researcher 1 16

4.8 Top 10 similar results of Researcher 2 16

4.9 Top 10 similar results of Researcher 3 17

List of Figures

5.1 Results of Similarity Score of User1 and Documents 19

5.2 Results of Similarity Score of User2 and Documents 19

5.3 Results of Similarity Score of User3 and Documents 19

5.4 Results of Similarity Score of User1 and Documents 20

5.5 Results of Similarity Score of User2 and Documents 20

5.6 Results of Similarity Score of User3 and Documents 21

Chapter 1

Introduction

1.1 Recommendation System

Recommendation systems are a subclass of information filtering system that seek to predict the 'rating' or 'preference' that user would give to an item. Contrasted with hunt frameworks, recommender frameworks give the likelihood to clients to find new things or thing classifications that they may not at first consider when defining the inquiry question. Explore on recommender frameworks has increased considerably over the past a few years, since the function and quality of recommendation becomes more intensely popular in an extraordinary mixed bag of online administrations. Furthermore, various genuine information sets that are made accessible in the group, and an arrangement of challenges that accentuate different proposal purposes, have further helped the advancement of exploration on recommender systems[5].

Recommender systems (RS) automatically select the most appropriate items to each user, thus shortening his product searching time and adapting the selection as his particular preferences evolve over time. Besides, recommender frameworks are characterized into the accompanying classes, in light of how suggestions are made:

Content-based recommendations: The user is recommended items like the ones the user favored in the past.

Collaborative recommendations: The client is prescribed things that individuals with comparative tastes and inclination loved in the past.

Hybrid approaches: These methods combine collaborative and content-based methods

1.2 Personalized recommendation system

E-commerce has been developing quickly keeping the pace with the web. Its quick development has made both organizations and clients confront another circumstance. Although organizations are harder to get by because of more rivalries, the open door for clients to pick among more items has expanded the load of data preparing before they select which items help. One answer for understand these methodologies is personalized recommendation[3] that helps clients to discover the items they might want to buy by handling an arrangement of proposed items for every given client. It is key to perceive every client's one of a kind and specific needs and prescribe a customized rundown.

The goal of such recommendations is to adapt the content of the site to the specific needs of the individual user, presenting to him/her the most attractive and relevant items.

Traditionally, there are two main methods to fulfill the task of personalized recommendation. They are typically based on (1) content, i.e., recommend items with content that is similar to the content of the items already consumed by the target users; (2) collaborative filtering(CF), i.e., providing items related to people who are related to the target user by some kind of similarity. Collaborative Filtering (CF) has become one of the most popular techniques of personalized recommendation. It is based on the assumption that similar users share the same interest. However, for the new users who hardly have interacted with others, there isnt enough information to use collaborative method to find similar users.

1.3 Personalized Paper Recommendation System

Personalized recommendation can be applied to outside of commercial applications. These days, many academic papers are coming out from a lot of conferences and journals. Academic researchers should go through all the conferences and journals which are related to their field of research and find out if there is any new articles that may relate to their current works. Sometimes they search the articles from Google scholars or Citeseer with the key words that might show interesting articles to them. However, these two methods require users to commit their time to search articles, which is labor-intensive, and also do not guarantee that they will find the exact articles related to their field of research.

In order to reduce their workload, we suggest developing the scholarly paper recommendation system[2] for academic researchers, which will automatically detect their research topics they are interested in and recommend the related articles they may be interested in based on similarity of the works. We believe this system will save the researchers time to search the articles and increase the accuracy of finding the articles they are interested in.

1.4 Organization of Thesis

Chapter 2 Presents a related work of the work to date in the domains of personalized recommendation for researchers that is relevant to the work presented in the remainder of the thesis.

Chapter 3 Defines the problem statement of the domain.

Chapter 4 Describes detail work done in this domain and steps and formulas of our personalized paper recommender system.

Chapter 5 Defines the how are results are different with the other and improvements in methods and accuracy.

The thesis ends with discussing the conclusions derived from the work and explores some future enhancements that can be made to get more accurate output and enhance the accuracy of the result.

Chapter 2

Related Work

To help create significant recommendations for specialists, suggestion frameworks have begun to influence the dormant hobbies in the production profiles of the scientists themselves. To alleviate sparsity, identify potential citation papers through the use of collaborative filtering. On a scholarly paper recommendation dataset, we show that recommendation accuracy significantly outperforms state-of-the-art recommendation baselines as measured by nDCG and MRR.[1]

A author's published works constitute a clean indicator of the inert hobbies of an analyst. A key some piece of our model is to upgrade the profile inferred straightforwardly from past works with data originating from the past works' referenced papers and additionally papers that refer to the work. In our investigations, we separate between lesser analysts that have just distributed one paper and senior scientists that have numerous distributions. We demonstrate that sifting these wellsprings of data is worthwhile when we moreover prune uproarious references, referenced papers and production history, we accomplish measurably critical larger amounts of proposal exactnes.[2]

In this paper discussion on the importance of social network of researchers for personalized recommendation of researchers and papers. In begin by briefly describing collaborative filtering method for personalized recommendation and its cold start problem of the new uses. [3]

Collaborative Filtering (CF) technique. It is become one of the most popular techniques of personalized recommendation. It is based on the assumption that similar users share the same interest. [4]

Collaborative Filtering Recommender Systems :

Collaborative filtering approaches are based on the following assumption: users that have similar tastes on some items may also have similar preferences on other items. Thus the main idea is to utilize behavior history from other like-minded users, in order to find good recommendations for the current active user. The typical user behavior history is presented as a m x n user-item matrix R. Accordingly, there are m users and n items in the dataset. Every entry is the user s rating for itemj . The rating can be in the form of binary action (view a product or not), or numerical rating (integer between 1 and 5). The rating can be provided by the user in an explicit way. Or it implicitly reflects users preference to this product, from his/her transaction history . The user-item matrix is usually incomplete and sparse.

Content-based Recommender Systems:

In the content-based approach, the idea is to detect items that are most similar to the users existing profile. A user profile is composed of his/her previous transaction history, such as what he/she viewed or purchased before. After the user profile is set up, how to determine similarity between an item and the profile is the key challenge. Various approaches have been developed, such as cosine similarity with TF/IDF term weight, Bayesian classifiers, clustering, etc.

This paper introduce MovieGEN, an expert system for movie recommendation using machine learning and cluster analysis based on a hybrid recommendation approach.[5]

Session-aware recommender system has all user behaviors (including search, click, purchase, etc.) in record. Such system makes recommendations based on the users short-term behavior within a session, as well as the users long-term behavior across multiple sessions. It focuses on modeling a users purchase intention in each session. By recommending the right product at the right time, the session-aware recommender system could enhance the users conversion rate and their satisfaction/utility.[6]

Social network extraction framework from the Web, called Referral Web [7]. The system focuses on co-occurrence of names on Web pages using a search engine. It estimates the strength of relevance of two persons X and Y by putting a query X and Y to a search engine: If X and Y impart a solid connection, we can discover much proof that may incorporate their individual homepages, arrangements of co-creators in specialized papers, references of papers, and organizatinal graphs. Interestingly, a path from a person to a person (e.g., from Henry Kautz to Marvin Minsky) is obtained automatically using the system. Later, with improvement of the WWW and Semantic Web innovation, more data on our every day exercises has gotten accessible on the web. Programmed extraction of social relations has much more terrific potential and request now contrasted with when Referral Web is initially created.

Social network mining system called POLYPHONET [8] an advanced social network extraction system from social web.

A system called Fink [9], for extraction, aggregation and visualization of online social networks for a Semantic Web community. Social networks are obtained using analyses of Web pages, e-mail messages, and publications and selfcreated profiles (FOAF files). TheWeb mining component of Flink, similarly to that in Kautzs[7] work, employs a co occurrence analysis.Given a set of names as input, the segment utilizes a web search tool to acquire hit considers for distinctive names well as the co-occurrence of those two names. The framework focuses on the Semantic Web group. Along these lines, the expression "Semantic Web OR Ontology" is added to the query for disambiguation.

An end-to-end framework that concentrates a users social network. That framework recognizes one of a kind individuals in email messages, finds their homepages, and fills the fields of a contact location book and the other individual's name. Connections are set in social network between the holder of the page and persons found on that page. A newer version of the system targets co-occurrence information on the entire Web, integrated with name disambiguation probability models.[10]

A probabilistic models for the Web appearance disambiguation issue [11]: the set of Web pages is part into groups, then one group could be acknowledged as holding just important pages: all different bunches are immaterial.

A calculation for the issue of cross-report distinguishing proof and following of names of diverse sorts [12]. Here form a generative model of how names are sprinkled into archives.

Discussion about a name-disambiguation module [13]. Its idea is this: for an individual whose name is not basic, for example, Yutaka Matsuo, we have to include no words; for an individual whose name is regular, we ought to include a few words that best recognize that individual from others. In an amazing case, for an individual whose name is extremely basic, for example, John Smith, numerous words must be added.

Chapter 3

Problem Statement

Problem Statement: Today too many research work are going on in different research area. There are so many new researcher debut in research field and senior researcher extends their area long way. There are so many research paper published today in different journals and conferences. So this system is used to new user/researcher or junior /senior researchers to help for finding similar research work which should help in their work, because there are so many research paper are available today so it is difficult to find exact similar research paper to user/researchers domain area. So this system recommend most similar paper for researcher based on their published papers/documents/journals/interest through similarity measure, Which gives most similar paper to help in their research work.

Chapter 4

Proposed Work

Method 1: simple similarity measure

I had used scholarly dataset1 for implementation purpose, which contain 13 users and 597 number of documents. Scholarly have also dataset2 which contain 50 users and 100,531 number of documents.

Below are the steps for the implementation purpose:

1. Concate feature vector text file of researchers published papers of every user in one text file.
2. Preprocessing
3. Fetch TF values for every researchers keyword from every researcher had published document files
4. Find median value of similar words
5. Fetch TF values for every researchers keywords from every document files and put in document-researcher keyword matrix
6. Find similar word in researchers published papers and total published documents
7. Measure similarity between researchers published papers and published documents
8. Select top n number of papers which is related to research area

Researcher 1	
Similarity Score	Similar Paper Index
0.687794037379863	'P04-1080$_{r}ecfv.txt'$
0.633943256823162	'P02-1054$_{r}ecfv.txt'$
0.620216440816256	'P03-1070$_{r}ecfv.txt'$
0.602255823206644	'P01-1051$_{r}ecfv.txt'$
0.583462802738823	'P06-1074$_{r}ecfv.txt'$
0.555877553301710	'P06-1073$_{r}ecfv.txt'$
0.552986408925387	'P01-1026$_{r}ecfv.txt'$
0.548795375454533	'P01-1034$_{r}ecfv.txt'$
0.542897813447576	'P00-1026$_{r}ecfv.txt'$
0.534447757305736	'P03-1062$_{r}ecfv.txt'$

Table 4.1: Top 10 similar results of Researcher 1

Researcher 2	
Similarity Score	Similar Paper Index
0.693084729888297	'P01-1037$_{r}ecfv.txt'$
0.672905271315925	'P02-1006$_{r}ecfv.txt'$
0.665051090531532	'P03-1003$_{r}ecfv.txt'$
0.663607760454874	'P01-1014$_{r}ecfv.txt'$
0.660899509142885	'P05-1026$_{r}ecfv.txt'$
0.655729570957063	'P02-1035$_{r}ecfv.txt'$
0.650582363955386	'P05-1029$_{r}ecfv.txt'$
0.644546558776135	'P06-1112$_{r}ecfv.txt'$
0.640912541160463	'P05-1076$_{r}ecfv.txt'$
0.630553739580646	'P01-1031$_{r}ecfv.txt'$

Table 4.2: Top 10 similar results of Researcher 2

Researcher 3	
Similarity Score	Similar Paper Index
0.737658373441029	'P02-1035$_recfv.txt'$
0.733213741703121	'P06-1037$_recfv.txt'$
0.712458723306390	'P01-1010$_recfv.txt'$
0.712162726335095	'P05-1011$_recfv.txt'$
0.702976445027025	'P04-1088$_recfv.txt'$
0.694945435412705	'P06-1128$_recfv.txt'$
0.694248356423435	'P04-1042$_recfv.txt'$
0.691604688728979	'P00-1061$_recfv.txt'$
0.690767584299282	'P04-1081$_recfv.txt'$
0.687770254366291	'P03-1038$_recfv.txt'$

Table 4.3: Top 10 similar results of Researcher 3

Method 2 : User Profile Construction

We propose recommending papers based on an individuals recent research interests as modeled by a profile derived from their publication list. We hypothesize that this will result in high recommendation accuracy as we believe that a users research interests are reflected in their prior publications.

We first construct each researchers profile using their list of previous publications, and then recommend papers by comparing the profiles with the contents of candidate papers.

We first divide researchers into (i) junior researchers, and (ii) senior researchers. This is because the two types of researchers publication lists exhibit different properties. We define junior researchers as having only one recently published paper, which has yet to attract any citations (i.e., no citation papers). Senior researchers differ in having multiple past publications, where their past publications may have attracted citations.

Our method starts with our former scholarly paper recommendation system [number], and as such it is instructive to first describe our system and its basis. It consists of three steps:

Step 1: Construct a user profile P_{user} from a researchers list of published papers;
Step 2:: Compute feature vectors f^{P_j} (j = 1, , n) for each of the papers;

Step 3: Compute the cosine similarity Sim(P_{user}, f^{P_j}) between P_{user} and f^{P_j} (j = 1, , t), and recommend papers with high similarity to the target user.

Now we show the formula for step1 to construct user profile for junior researcher and senior researchers.

For junior researchers,

$$P_{user} = f^P + \sum_{y=1}^{l} W^{P \to Pref_y} f^{Pref_y} \tag{4.1}$$

For senior researchers,

P$_{user} = f^P + \sum_{x=1}^{k} W^{Pcit_x \to P} f^{Pcit_x} + \sum_{y=1}^{l} W^{P \to Pref_y} f^{Pref_y}$(4.2)

Where $Pcit_x$ (x = 1, , k) and $Pref_y$ (y = 1, , l) denote papers that cite P and papers that P refers to, respectively. In addition, $W^{Pcit_x \to P}$ and $W^{P \to Pref_y}$ are weights for the citation papers and weights for the reference papers, respectively. We can define these weights in a more general form as follows: Let $W^{u \to v}$ be the coefficient used to compute the weight for between target v and its source u. In addition, let f^u and f^v be the feature vectors of the source u and target v papers, respectively. Then cosine similarity sim(f^u,f^v) between the two vectors is used as $W^{u \to v}$.

Researcher 1	
Similarity Score	Similar Paper Index
0.719228759370801	'P04-1080$_{r}ecfv.txt'$
0.663309601185123	'P02-1054$_{r}ecfv.txt'$
0.643126760525625	'P03-1070$_{r}ecfv.txt'$
0.636369117657868	'P06-1074$_{r}ecfv.txt'$
0.596611313023832	'P01-1034$_{r}ecfv.txt'$
0.590890281176923	'P06-1073$_{r}ecfv.txt'$
0.588559590005847	'P00-1026$_{r}ecfv.txt'$
0.583438359076361	'P01-1051$_{r}ecfv.txt'$
0.578883476855913	'P01-1026$_{r}ecfv.txt'$
0.575914475031329	'P04-1012$_{r}ecfv.txt'$

Table 4.4: Top 10 similar results of Researcher 1

Researcher 2	
Similarity Score	Similar Paper Index
0.785438800639890	'P01-1037$_{r}ecfv.txt'$
0.762942849574503	'P02-1006$_{r}ecfv.txt'$
0.758709261797852	'P03-1003$_{r}ecfv.txt'$
0.736264845129093	'P04-1073$_{r}ecfv.txt'$
0.733209899587872	'P02-1054$_{r}ecfv.txt'$
0.730582994833413	'P06-1135$_{r}ecfv.txt'$
0.730327180098239	'P05-1027$_{r}ecfv.txt'$
0.729154427847985	'P05-1026$_{r}ecfv.txt'$
0.709249806494022	'P06-1112$_{r}ecfv.txt'$
0.704019572623653	'P05-1029$_{r}ecfv.txt'$

Table 4.5: Top 10 similar results of Researcher 2

Researcher 3	
Similarity Score	Similar Paper Index
0.754869639214806	'P06-1037$_recfv.txt'$
0.741759452782148	'P05-1011$_recfv.txt'$
0.739294697770865	'P02-1025$_recfv.txt'$
0.726237566786361	'P01-1010$_recfv.txt'$
0.723239062030783	'P04-1088$_recfv.txt'$
0.720985923253721	'P02-1035$_recfv.txt'$
0.718370705835827	'P03-1038$_recfv.txt'$
0.715469480002374	'P05-1012$_recfv.txt'$
0.714554129429456	'P02-1043$_recfv.txt'$
0.714269461838079	'P06-1097$_recfv.txt'$

Table 4.6: Top 10 similar results of Researcher 3

Method 3 : Euclidean Distance Measure Similarity

The basis of many measures of similarity and dissimilarity is Euclidean distance. The distance between vectors X and Y is defined as follows:

$$d\left(x,y\right) = \sqrt{\sum_{i=1}^{n} left(x_i - y_i^2} \tag{4.3}$$

Euclidean distance is the square root of the sum of squared differences between corresponding elements of the two vectors. Euclidean distance is most often used to compare profiles of respondents across variables. Here we using x as feature vector of number of documents and y define feature vector of user profile. Below shows results for the similarity by simple and user profile base Euclidean distance similarity.

Researcher 1			
Simple similarity		User profile base similarity	
Similarity Score	Similar Paper Index	Similarity Score	Similar Paper Index
0.921779417942383	'P04-1080$_recfv.txt'$	0.943578902227339	'P03-1070$_recfv.txt'$
0.917977516530854	'P01-1034$_recfv.txt'$	0.941340670012709	'P01-1034$_recfv.txt'$
0.917733234634991	'P03-1070$_recfv.txt'$	0.937252618381823	'P04-1080$_recfv.txt'$
0.914883249325158	'P00-1026$_recfv.txt'$	0.935277645769011	'P01-1016$_recfv.txt'$
0.910285482322611	'P04-1012$_recfv.txt'$	0.934144656475369	'P05-1071$_recfv.txt'$
0.908800553569143	'P05-1071$_recfv.txt'$	0.933270949248077	'P01-1051$_recfv.txt'$
0.907163621042714	'P01-1016$_recfv.txt'$	0.932874645165048	'P00-1026$_recfv.txt'$
0.905051723762410	'P00-1002$_recfv.txt'$	0.931751194529405	'P00-1002$_recfv.txt'$
0.904510446527962	'P04-1062$_recfv.txt'$	0.930863900243086	'P06-1073$_recfv.txt'$
0.903064492343940	'P04-1009$_recfv.txt'$	0.930840874161614	'P03-1071$_recfv.txt'$

Table 4.7: Top 10 similar results of Researcher 1

Researcher 2			
Simple similarity		User profile base similarity	
Similarity Score	Similar Paper Index	Similarity Score	Similar Paper Index
0.935895840948468	'P02-1065$_recfv.txt$	'0.948315882348694	'P02-1065$_recfv.txt'$
0.933587404713376	'P04-1077$_recfv.txt$	'0.945237529387075	'P05-1043$_recfv.txt'$
0.932526442879500	'P05-1043$_recfv.txt$	'0.943685373620587	'P04-1077$_recfv.txt'$
0.931292692873408	'P00-1053$_recfv.txt$	'0.942820353378009	'P06-1008$_recfv.txt'$
0.931218991987617	'P04-1075$_recfv.txt$	'0.942491728178304	'P02-1039$_recfv.txt'$
0.930861539695490	'P06-1019$_recfv.txt$	'0.942490830317586	'P01-1014$_recfv.txt'$
0.930128895880718	'P03-1064$_recfv.txt$	'0.942421685667535	'P05-1010$_recfv.txt'$
0.930108351597521	'P02-1039$_recfv.txt$	'0.942041580382302	'P02-1036$_recfv.txt'$
0.929219549268518	'P04-1076$_recfv.txt$	'0.941587470726869	'P06-1131$_recfv.txt'$
0.929185611440611	'P02-1058$_recfv.txt$	'0.941379581185800	'P01-1019$_recfv.txt'$

Table 4.8: Top 10 similar results of Researcher 2

Researcher 3			
Simple similarity		User profile base similarity	
Similarity Score	Similar Paper Index	Similarity Score	Similar Paper Index
0.931273204658357	'P02-1023$_recfv.txt'$	0.949507278522075	'P04-1081$_recfv.txt'$
0.927308591223292	'P06-1097$_recfv.txt'$	0.946387514900068	'P04-1088$_recfv.txt'$
0.923415792783361	'P04-1081$_recfv.txt'$	0.945243007906533	'P03-1038$_recfv.txt'$
0.922996716867842	'P03-1038$_recfv.txt'$	0.945017085306724	'P06-1097$_recfv.txt'$
0.921240388627427	'P04-1077$_recfv.txt'$	0.944956188279298	'P01-1004$_recfv.txt'$
0.920172152394308	'P01-1004$_recfv.txt'$	0.943775983494387	'P02-1010$_recfv.txt'$
0.920013565719454	'P02-1038$_recfv.txt'$	0.943208511446480	'P04-1064$_recfv.txt'$
0.919226543258773	'P06-1001$_recfv.txt'$	0.942882303727965	'P01-1011$_recfv.txt'$
0.917913591020113	'P03-1037$_recfv.txt'$	0.942710826209850	'P06-1103$_recfv.txt'$
0.917705723134810	'P00-1073$_recfv.txt'$	0.942639396055412	'P02-1023$_recfv.txt'$

Table 4.9: Top 10 similar results of Researcher 3

Chapter 5

Result Observation

Here shown output of method1, method2 and method3 for researcher 1,2 and 3.

For Researcher 1:

In simple similarity measure method1 output paper 'P04-1080_recfv.txt' has similarity score is about 0.6877 as shown in table 4.1 while in method2 user profile construction measure score is raised up to 0.706 as shown in table 4.4. In both measure rank is same but in most of the paper the order of other paper is also changed as change in similarity score like paper 'P06-1073_recfv.txt' in table 4.1 rank is 6th in method 1 while it is on 10th in method 2 in table 4.4. While in table 4.1 paper 'P03-1062_recfv.txt' is on 10th position in method1 while it is not present in table 4.4 in output of method2. Same like in Euclidean distance similarity measure similarity score is raised for every paper and rank is slightly changed seeing in table 4.7. In short on an average similarity score for every paper is raised in user profile construction measure.

For Researcher 2:

As shown in table 4.2 and 4.5 similarity score is raised very high in method 2. First 3 paper are contain same rank in both method. Rank of the other paper is changed all most but the similarity score is raised in method 2 of every paper. Same like in Euclidean method similarity score of paper is raised and order is slightly changed moreover.

For Researcher 3:

Here shown in table and score of 'P02-1035_recfv.txt' is decreased and it fall down from 1st in method1 to rank 6th in method2 while score of the other papers is increased all over. Paper 'P06-1037_recfv.txt' becomes 1st in method2 while it is on 2nd in method1.

In Euclidean method 'P02-1023_recfv.txt' paper lost its first position and not secure top 10 position in user profile construction method.

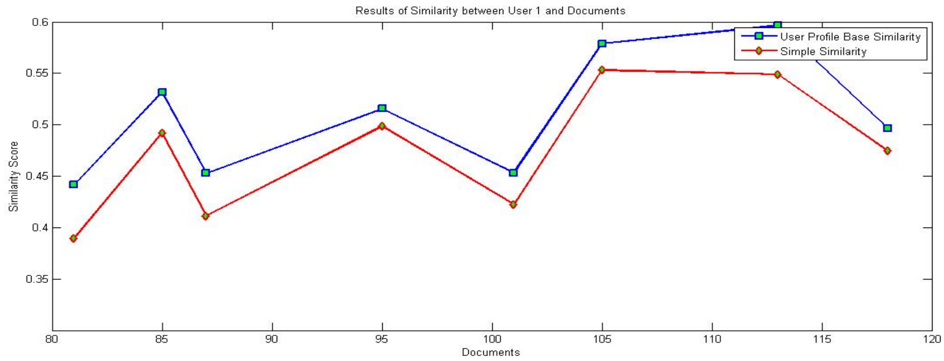

Figure 5.1: Results of Similarity Score of User1 and Documents

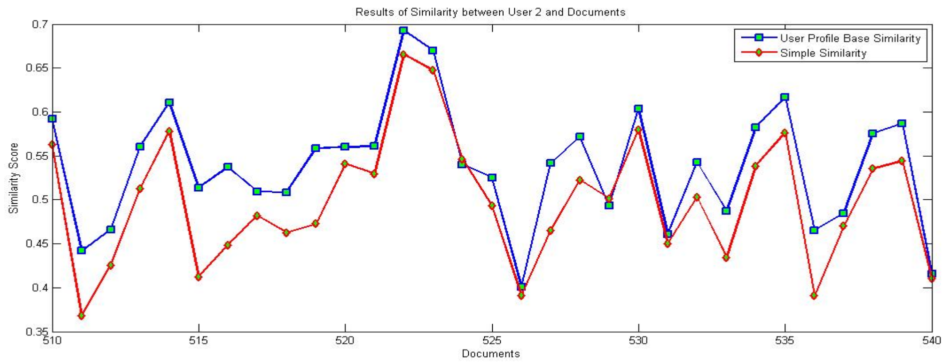

Figure 5.2: Results of Similarity Score of User2 and Documents

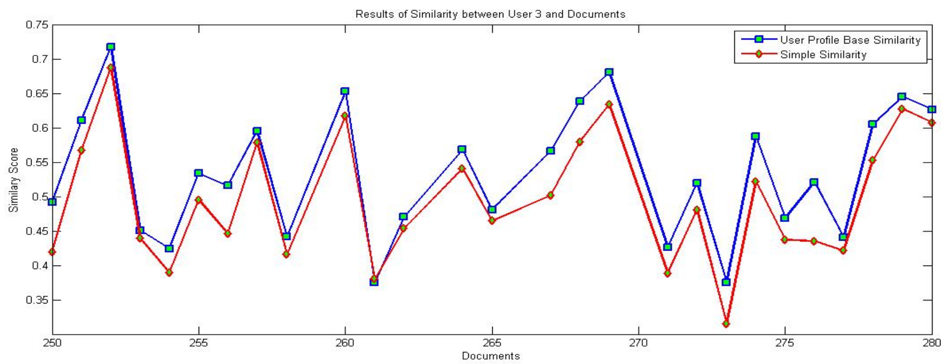

Figure 5.3: Results of Similarity Score of User3 and Documents

Here above 3 figure shows that for every documents method2 has better similarity score than method1.

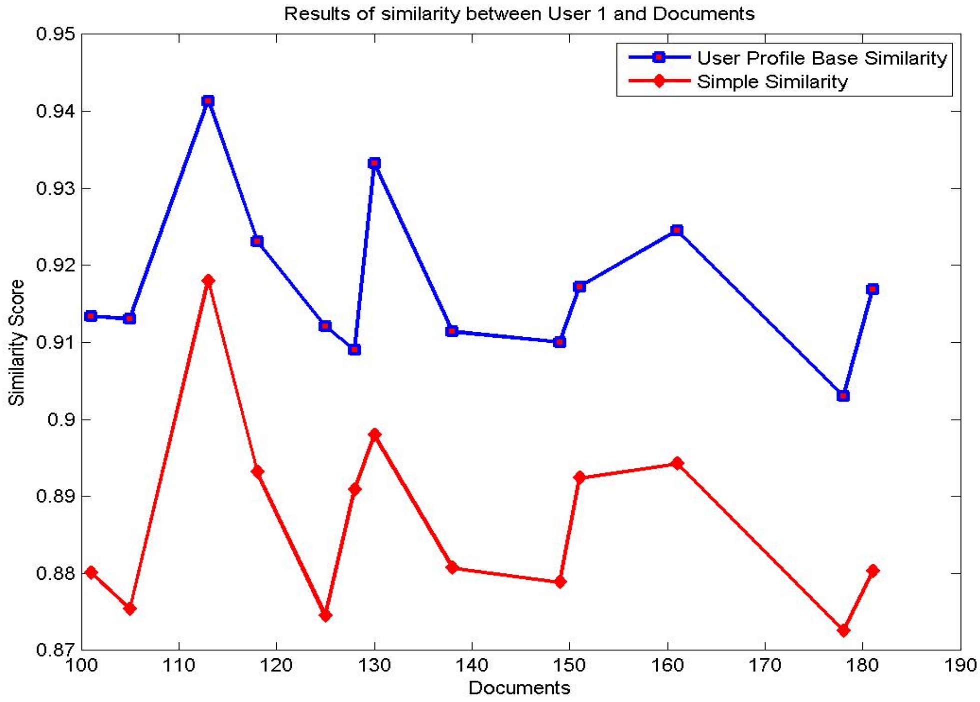

Figure 5.4: Results of Similarity Score of User1 and Documents

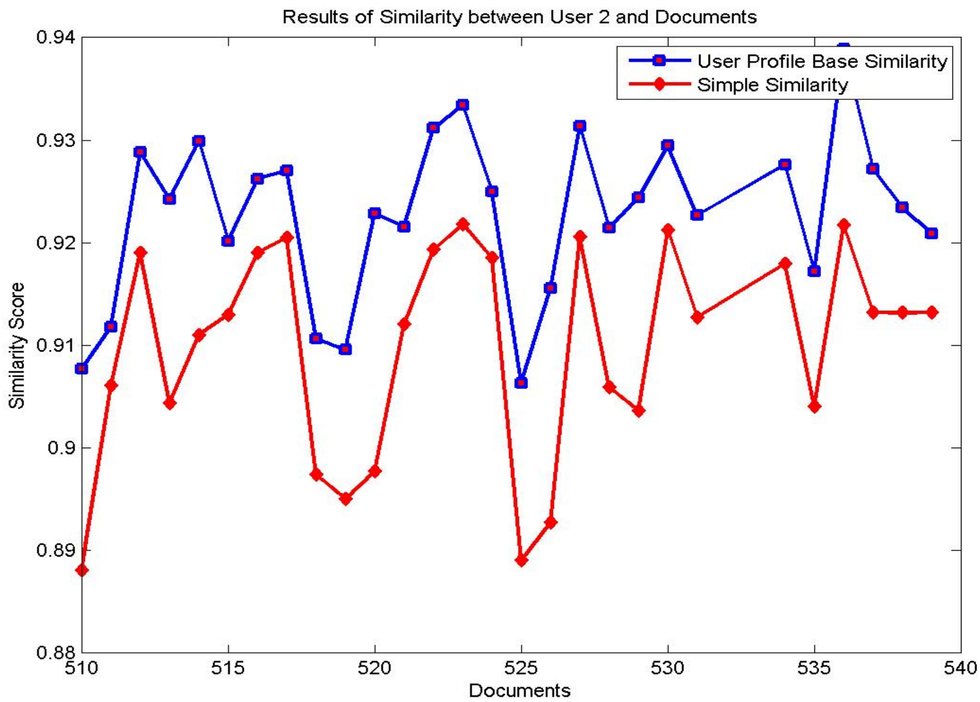

Figure 5.5: Results of Similarity Score of User2 and Documents

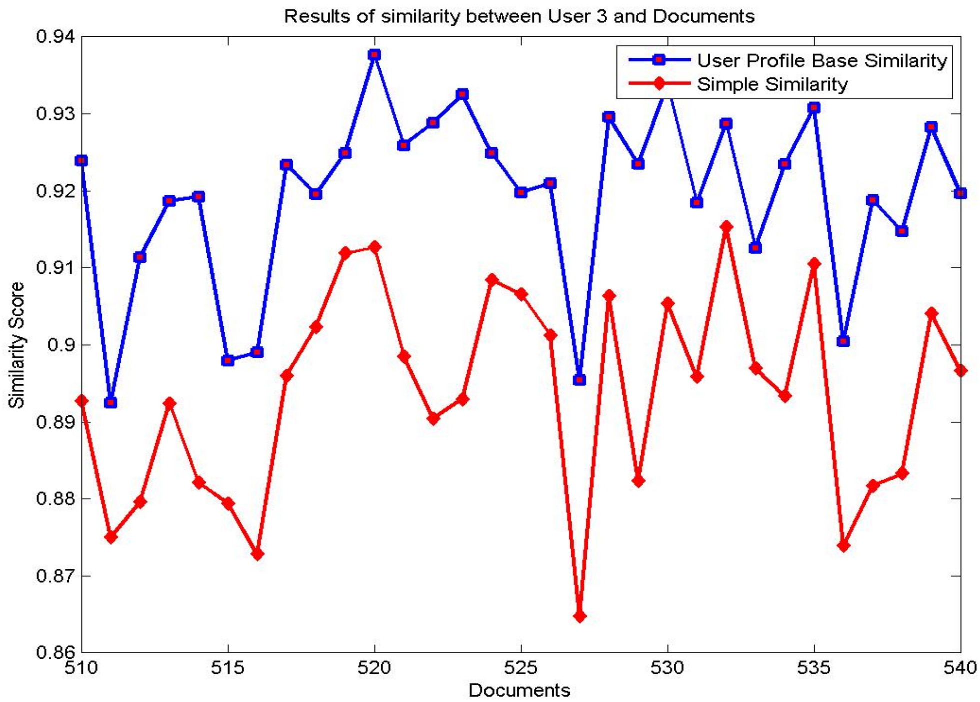

Figure 5.6: Results of Similarity Score of User3 and Documents

Chapter 6

Conclusion

In this paper we propose a method1 for user to find the documents which are related to his/her research area. User and paper relevance score defines the user has interest in that paper. Higher the score mean high interest. And in method 2 we have separate junior researchers and senior researcher and then by using their citation and reference papers we have mad user profile for them. By method2 we had also increase similarity score and get more accurate result. This method is very useful for the new user and also who are working in academics and different institutes.

Future scope by weighting more to the authors, newly published papers, publication, journal, conferences. it will give more accurate output and also it may create new way of research.

References

[1] Kazunari Sugiyama Min-Yen Kan Exploiting Potential Citation Papers in Scholarly Paper Recommendation JCDL2013

[2] K. Sugiyama and M.-Y. Kan. Scholarly Paper Recommendation via Users Recent Research Interests. In Proc. of the 10th ACM/IEEE Joint Conference on Digital Libraries (JCDL 10), pages 2938, 2010.

[3] Cheng Chen, Chengjie Mao, Yong Tang, Guohua Chen, and Jinjia Zheng1: Personalized Recommendation Based on Implicit Social Network of Researchers School of Computer Science, South China Normal University, Guangzhou, China.

[4] Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using collaborative filtering to weave an information tapestry. Commun. ACM 35, 2002.

[5] Eyrun A. Eyjolfsdottir, Gaurangi Tilak, Nan Li : A Movie Recommendation System. University of California Santa Barbara

[6] Jian Wang : Session Aware Recommender System in E-commerce. University of California Santa Cruz, 2013.

[7] H. Kautz, B. Selman, and M. Shah. The hidden Web. AI magazine, 1997.

[8] Matsuo, Y., Mori, J., Hamasaki, M.: POLYPHONET: An advanced social network extraction system from the Web. Journal of Web Semantics,2007.

[9] P. Mika. Flink: Semantic web technology for the extraction and analysis of social networks. Journal of Web Semantics, 2005.

[10] I. Davis and E. V. Jr. RELATIONSHIP: A vocabulary for describing relationships between people.

[11] R. Bekkerman and A. McCallum. Disambiguating web appearances of people in a social network, 2005.

[12] X. Li, P. Morie, and D. Roth. Semantic integration in text: From ambiguous names to identifiable entities, 2005.

[13] D. Bollegara, Y. Matsuo, and M. Ishizuka. Extracting key phrases to disambiguate personal names on the web, 2006.