

Recommender System for Journal Articles using Opinion Mining and Semantics

Anam Sardar, Javed Ferzund, Muhammad Asif Suryani, Muhammad Shoab

Department of Computer Science
COMSATS Institute of Information Technology
Sahiwal, Pakistan

Abstract—Till date, the dominant part of Recommender Systems (RS) work focusing on single domain, i.e. for films, books and shopping and so on. However, human inclinations may traverse over numerous areas. Thus, utilization practices on related things from various domains can be valuable for RS to make recommendations. Academic articles, such as research papers are the way to express ideas and thoughts for the research community. However, there have been a lot of journals available which recognize these technical writings. In addition, journal selection procedure should consider user experience about the journals in order to recommend users most relevant journal. In this work of journal recommendation system, the data about the user experience targeting various aspects of journals has been gathered which addresses user experience about any journal. In addition, data set of archive articles has been developed considering the user experience in this regard. Moreover, the user experience and gathered data of archives are analyzed using two different frameworks based on semantics in order to have better consolidated recommendations. Before submission, we offer services on behalf of the research community that exploit user reviews and relevant data to suggest suitable journal according to the needs of the author.

Keywords—*Recommendation system; journal recommendation system; user opinion; semantic similarity; text analysis*

I. INTRODUCTION

As the universe is getting digital, a large volume of structured, semi-structured and unstructured data is being generated very fast. This data is in terabytes, so it is referred as Big Data. Big data approaches are used to handle those types of datasets that are so big and complex that typically used applications software are not sufficient to exploit them fully. Because of the rapid increase in data volume, one is always flooded with Superfluity of choices in any domain [1]. A recommendation system uses the large volume of data in the form of text and sentiments available for summarization purpose to make serious and valid decisions. Recommender systems gather information from the users about their preferences for a particular item to make predictions for the product such as which bag I should buy or which paper I should read next [2]. Recommendations can be made based on user's interest which can be analyzed by the user's profile or considering their online or offline behavior e.g. RS is a subclass of information filtering system that tries to predict the "opinion" that a user would give to an item.

Recommender frameworks have turned out to be amazingly common in recent years, and are used in an

assortment of zones. Some prevalent applications incorporate music, books, movies, research papers; seek questions, social labels, and items in general. In any case, there are likewise recommended frameworks for specializations, partnership, jokes, eateries, life insurance, and Twitter pages [3]. Similarly, journal recommendation system has also become an important topic of discussion for research community which writes and publishes research articles, patents, and books. Because today we have numerous choices of journals that publish articles annually, quarterly, monthly and even bi-monthly, it becomes very difficult to choose an appropriate journal to submit your manuscript.

With an increase in the publication of different research papers in multiple journals of diversified fields, authors find it difficult to choose an appropriate journal for their research work. In submission of journal, article may result in rejection and the main reason for rejection is that the paper is not submitted to a relevant journal even when the paper itself is excellent. So there is a need to develop a Recommendation system that can suggest suitable journals to the authors. The journal recommendation system can provide services to authors on behalf of publishers of academic journals. The choice of journal directly influences the authorial decisions like impact on practitioners, CV value of publication and acceptance or rejection risk [4]. Now the core problems that arise while building a journal recommendation system are:

- which data set should be collected for a journal recommendation;
- where to store this amount of big data of journals;
- how to effectively perform data mining and sentiment analysis to make better journal recommendations;
- Providing accurate recommendations to the users with accuracy and exactness;
- which recommendation system technique would be best for journal recommendation system.

The solution to the above-mentioned issues is our proposed solution that is based upon user opinion to make suitable journal recommendations. Existing systems for journal recommendation works by matching title and abstract of the papers [5] and do not consider the user experience with journals. Previously most of the work has been done by just content similarity and didn't focus on other aspects e.g. low-level features. Our proposed system not only considers the

content similarity but also takes into account low-level features like subscription charges, access options etc. The main contribution of our paper is that our system collects user experiences also. For this purpose, we have conducted a survey that gathers user experience. Combining the content similarity with low-level features and user reviews for journal recommendation provide better recommendations.

In this work, the information about the user experience about an arrangement of journals focusing on different domains has been accumulated. This information incorporates journal domain, name and overview questions which address user experience with the journal.

Section 2 contains related work; Section 3 contains methodology and proposed framework followed by experimental setup and results.

II. LITERATURE REVIEW

Recommendation systems play a significant role in e-business and information sharing systems. Over two decades of research and different algorithms being implemented for recommendation engines it is declared that recommendation is not a one-size-fits-all problem. So, recommendation systems must need to be designed according to application-specific embedded tasks. Successful deployments must include user required tasks, for which different design choices are in practice. If authors are assumed to perform reasonable in the typical financial or economic sense, they should choose a journal for publication of their work according to where they can expect the highest average value adjusted for risk and expenses.

Journal recommendation systems have been studied by researchers in different backgrounds. For example, recommendation system is proposed to recommend the appropriate journals considering the other factors like price, openness, and subscription rather than just matching the content [5]. It is proposed that how author selects journal in development administration influenced by the quality and administration recognitions [4].

A hybrid research paper recommender system is introduced by researchers which improves the research paper recommendations by combining keyword-based search with implicit and explicit rating, citation analysis and source analysis [6]. The system uses 'Distance Similarity Index' (DSI) and the 'In-text Impact Factor' (ItIF) methods to improve the quality of recommendations.

A research paper recommender framework is proposed in view of the hypothesis that provides a clear indication of user interest by depending upon previously published articles of author. The system differentiates between senior and junior researchers and prunes the unnecessary citations and references [7]. Filtering these information sources result in the higher accuracy of recommendations. In [8], authors have discussed the online and offline evaluation of research paper recommender framework and conclude that offline evaluation in this domain does not provide promising results.

Docear's research paper recommender framework is proposed using content-based filtering in which user's data

(citations, references, and papers) is directed in mind maps and are then utilized for recommendations [9]. A research paper recommender framework is introduced using a Dynamic Normalized Tree of Concepts (DNTC) model and a complex ontology [10]. The system is evaluated offline using ACM digital library papers and the results show that this model performs better than the vector of space models.

Authors have discussed that Mendeley recommender system works by incorporating collaborative filtering and user feedback to produce recommendations in [11]. Results show that the proposed method provides better accuracy for new users. To serve the new researchers in getting a diagram of the research performed in a specific zone, authors have proposed keywords based retrieval procedure in [12] for giving an overview and a various arrangement of papers as a piece of the preliminary reading list. A literature review is presented on ontology-based recommender frameworks in the domain of e-learning [13]. This investigation demonstrates that intersection of information based proposal with other suggestion methods can upgrade the viability of e-learning recommender systems. Authors have discussed the performance of stereotype and most-popular recommendations in the domain of scholarly recommender frameworks in [14].

Researchers have discussed the new item problem and proposed a method of automatically analyzing the video and audio contents through low-level characteristics rather than just focusing on high-level features of the video content [15]. The Paper typically focused on the visual features. In [16] authors have proposed a real-time web service for providing recommendations for different items using opinion and ratings of people provided on Twitter, Facebook and other social media sites. Reviews about four products given by blippar site have analyzed using CF based approach.

A Latent Dirichlet Allocation approach that is used for sentiment mining and feature retrieval to improve the accuracy of recommendations is proposed in [17] and it was found that this technique provides the best results as compared to typical clustering techniques. An efficient user-modelling technique based on mind maps to recommend the Research papers is presented by researchers in [18].

In this paper, numerous variables concerning to mind-map-based user modelling were identified, and assess the variables' influence on user-modelling efficiency with an offline evaluation. Research work is done in which Authors have developed a hierarchical Poisson matrix factorization (HPF) for recommendation purpose. HPF model considers sparse user activities data, where every client has given criticism on just a little subset of things [19]. HPF handles both express appraisals, for example, various stars, and implicit ratings, for example, perspectives, snaps, or buys.

In [20], Apache Mahout is used to evaluating TF-IDF weighted technique of clustering. The dataset of tweets is used to evaluate the result of eliminating stop words from the dataset. The proposed system in [21] uses the slope-one recommendation algorithm to recommend micro-videos. The result shows that the strength of used algorithm provides better visualization interface and Hadoop framework provides high-level performance. The challenge of using Map Reduce

paradigm to parallelize CF technique is being addressed in [22]. The result shows that CF algorithms are not useful for Hadoop platform as it does not decrease the response time for an individual user. To overcome the issues like scalability, sparsity and imprecision etc. a CF method with Dimensionality technique is applied using Mahout in [23] to improve the recommendation accuracy of prediction and quality. Results show that approaches such as PCA and SVD can decrease the noise of high dimensional data, and provides an improvement in tackling the scalability and sparsity issues of prediction.

In [24] the authors have discussed that Recommendation systems are important platforms for users pursuing technical ways to find best choices available from a big amount of data. Directed edge recommendation problem is described in [25] where the user can recommend items to his connected user based on the algorithm that combines sharing preferences model and user preference model. Results demonstrate that joining the undertaking setting prompts more exact proposals as compared to group recommender system. The author provided an up-to-date and detailed survey of the recommended field, considering various kinds of interfaces, the range, and diversity of different recommendation system algorithms, the functionalities provided by these systems and their use of Artificial Intelligence methods [26].

III. PROPOSED STUDY AND DESIGN

The proposed approach comprises of two frameworks targeting user opinion and analysis of detailed content (i.e. journal archive data). Each of these frameworks is used to provide a consolidated recommendation. The theme of the proposed work is to explore user opinions and archive analysis which definitely results in better recommendations.

In preceding section introduction regarding recommender frameworks and related study have been provided. This section gives comprehensive insights about the proposed study considering journal recommender framework. Previously, it has been described that there are some studies that have been performed for journal recommendation considering various factors like matching the contents of the script, content matching combined with script charges, and access options, etc.

A. Framework for User Opinion Analysis

In Fig. 1, a conceptual framework is provided to analyze user opinion. First of all, the data is collected from the users by means of a survey paper in which user provide an opinion about his experience with the journals. Now the gathered data is unstructured and required some preprocessing before it can be analyzed. Preprocessing phase was a major challenge and a plenty of time is consumed during this phase. This textual preprocessing includes cleaning steps, such as removing duplicate characters, replacing special characters with spaces, removing stop words and word stemming.

From the cleaned data, attribute selection is made and separated into numerical/categorical and textual attributes. Then, by using different text analysis approaches it is analyzed that whether the user provides a positive opinion or negative opinion.

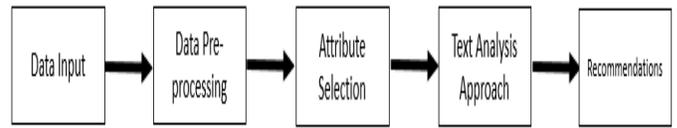


Fig. 1. A conceptual framework for user opinion analysis.

User opinion analyzed in this section is further used in the second framework to provide recommendations.

B. Framework for Semantic Similarity based Approach

In Fig. 2, a conceptual framework for semantic similarity based approach is provided. For recommendations, another data set is gathered based on the survey data collected in the above-mentioned step. This dataset includes archives about the journal.

Preprocessing phase is done in which TF-IDF approach is used. A Term Document Matrix is generated that describes the frequency of input words in the collection of documents based on term-term correlations. Then, by using KNN approach similarity is measured.

For checking semantic relationships, we used an approach based on the work which counts semantic connections in light of terms by utilizing semantic kernel. Semantic relationship implies which terms are co-related; in this manner it can enhance the clustering model. The work did in such manner additionally incorporate GVSM which is Generalized Vector Space Model (GVSM). GVSM accept that vectors are straightly autonomous so figure the term-term relationship. The similarity is measured using approach defined in [27].

If there is a matrix X which contains n archives and m terms, by applying GVSM we have semantic piece.

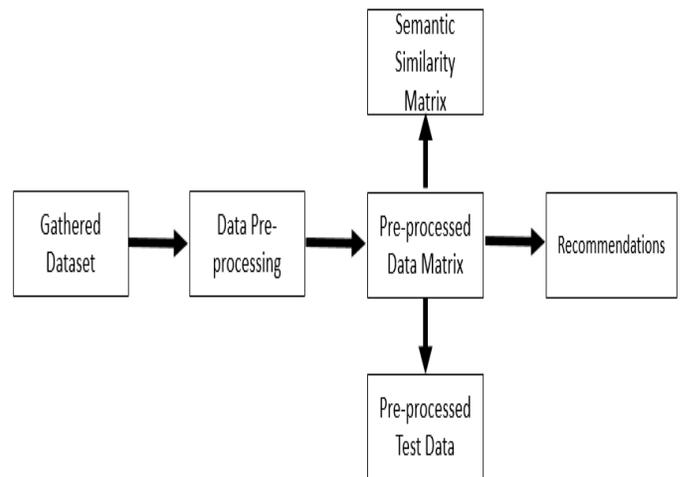


Fig. 2. Framework for semantic similarity based recommendation.

$$K = X^T G X \quad (1)$$

Here K is a gram matrix of lines; G is a gram matrix of Columns separately. In this way, cosine similarity can be figured by:

$$Sim = L^{-1/2}KL^{-1/2} \quad (2)$$

In the above conditions, G is vital and it must satisfy a portion of the properties, for example, G ought to be positive semi-distinct and represent the inner products of the term vectors. So there ought to be some estimation of G which is as under.

$$G_{ASSC} = QQ^T \quad (3)$$

$$G_{ASSC_N} = L_Q^{-1/2}QQ^TL_Q^{-1/2} \quad (4)$$

L_Q Is an $m \times n$ diagonal matrix whose components are the diagonal components of QQ^T . In this way the semantic kernel that relate to various estimates of K are:

$$K_{ASSC} = X^TQQ^TX \quad (5)$$

$$K_{ASSC_N} = X^TL_Q^{-1/2}QQ^TL_Q^{-1/2}X \quad (6)$$

Diverse measures of semantic kernel in view of term-term relationships, which is proportional to mapping archives to the higher semantic space where correlated terms are related with each other.

IV. EXPERIMENTAL DESIGN AND SETUP

In this section, comprehensive details are provided about the collection of data set. Detailed results are also shown in this section.

A. Data Collection and Analysis Process for User Opinion

This study involves the steps that need to be addressed in order to recommend journals based on user's opinion about the journal. In the first section, study design, information related to techniques existing for survey type research participants will be discussed. After that, the critical components to be considered for survey sort research will be explained. Further, the rules for setting up a questionnaire and choice of target population will be displayed. The fundamental reason of our examination is to explore the part of "user experience" in creating a positive or negative effect on the journal selection of researcher's community.

1) *Mode of Observation*: This study is based on a survey that is known as ex-post-facto design. Such type of study only reports what has happened or what is happening. It is a longitudinal study. Questionnaires were distributed to the faculty and data was collected face to face.

2) *Target Population*: The qualities of the target population, the intended interest group in each investigation is by all accounts seems to be critical as it will establish the framework of your research work. The directed application is by all accounts a pivotal point in this investigation; following are the different parameters that have been considered in such manner:

Age: 25-50 years of age

Education: Master's, M-Phil and PhD

Gender: both male and female

3) *Targeted Locations/Organization*: Researchers in this topographical region are chosen and features of the targeted audience have been given.

Following University with the named departments is selected for this study:

"COMSATS University"

Departments:

Bioinformatics

Computer Science

Math

4) *Observational Approach*: In this work, the survey was the fundamental wellsprings of gathering data from the specified group of audience. The polls of our study were utilized as a component for the gathering of data essential for this investigation. The Close community has been considered and ended questions are incorporated in the questionnaire. The survey comprises different questions about the user experience e.g. view about the journal response time, subscription charges, etc.

In addition, different factors were additionally considered with the goal that the investigation can have all the essential data and information that will lead toward successfully finishing of this study considering recommendation services.

5) *Data Collection*: Data is collected from the field that is the campus of COMSATS Institute of Information Technology from different departments. Prior to the rounding out the questionnaires, we led a session for the focused group of audience with the goal that every one of the surveyors must have important data that can help them in the correct filling of the survey. In addition, this action will help in getting the desired outcomes from this investigation.

Survey papers were given over to the researchers after a short portrayal and extension about the reason for this study. Essential information was recorded subsequent to getting back the filled questioners. Then this data was adjusted according to the need suited for recommendation purpose. For cleanness and simplicity surveys were provided in two different domains, offline and online.

In online, a survey was produced on Google Forms and was made accessible by giving the connections of this survey around, as this action will empower us to draw in the clients that incline toward the online medium. The substantial link of the survey appears below:

Showing Links of the Survey:

https://docs.google.com/forms/d/17fMH6u_6o_LxhTqhbYWYPxhbk2Sh8xANcgT0ZkBUYHw/edit

In addition, for the second kind of group of an audience, the survey was made accessible in hard shape with the goal that individuals can undoubtedly give the supposition in the

required arrangement and that movement will surely help us in pulling in the audience.

6) *Data analysis:* Analysis of the gathered information is very important and crucial task as it provides us with the information and results that we were looking for. In this study the simple information has been gathered and analyzed via different tools. This activity will help us in finding the relevant information. Moreover, the gathered textual information has been processed.

This section provides the detailed results which have been shown and the results depicts that incorporating user experience have impact on selected domain of study and can improve recommendation results. In previous sections complete descriptions have been provided.

The following are the outcomes which have been derived from the users in the form of a survey. All the gathered outcomes were plotted utilizing different tools and were appeared here one by one introducing the data in regards to each inquiry in this study. Here for the straightforwardness and brevity, the chosen results have been demonstrated which give noteworthy information in such manner.

Data in Fig. 3, apprised that 67% researchers find submission procedure helpful, 8% find it difficult, and 39% find this procedure fair.

Fig. 4 revealed that 60% researchers feel that archive papers do not help them in getting the idea about the journal, 5% researchers feel that archive papers help them partially while 35% feel archive papers inappropriate and do not provide any idea about the journal.

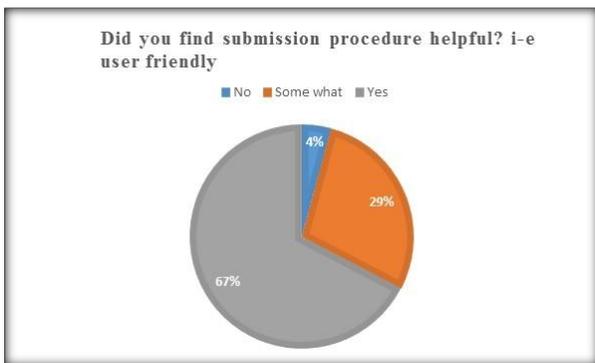


Fig. 3. Submission procedure.

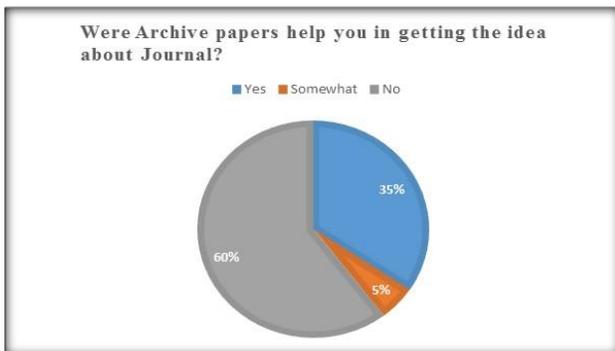


Fig. 4. Archive papers.

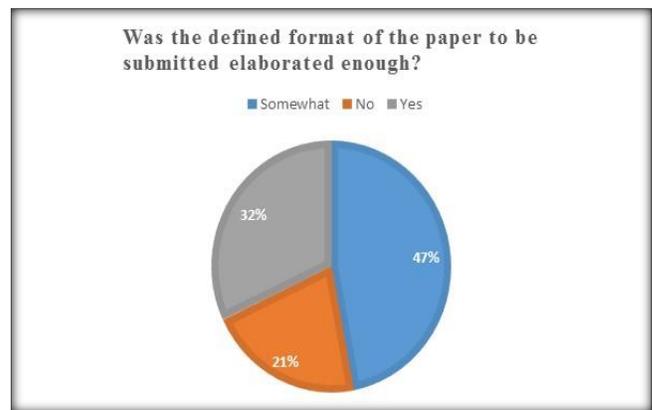


Fig. 5. Defined format.

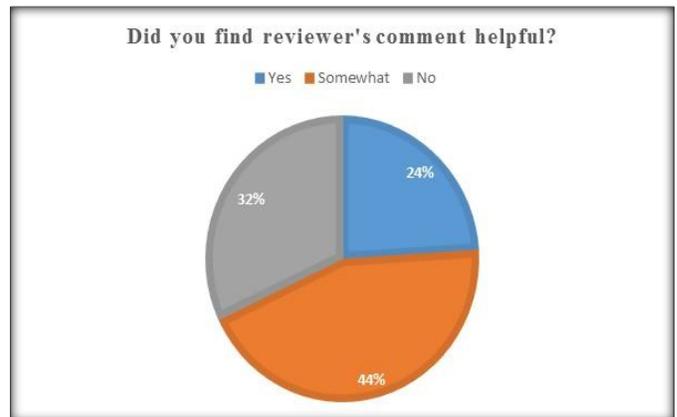


Fig. 6. Reviewer's comment.

Fig. 5 illustrated that 32% people were agreed with the statement that defined format was well elaborated, 47% people feel that it was ambiguous while 21% find it difficult.

Fig. 6 in the survey presented that 24% people have an opinion that reviewer's comment was helpful for improving their manuscripts, 32% were not satisfied with the comments and 44% people find the comments ambiguous.

Fig. 7 exhibited that 79% researchers' reveal that communication was supportive, 12% researchers take this communication as fair while 9% feel it was discouraging.

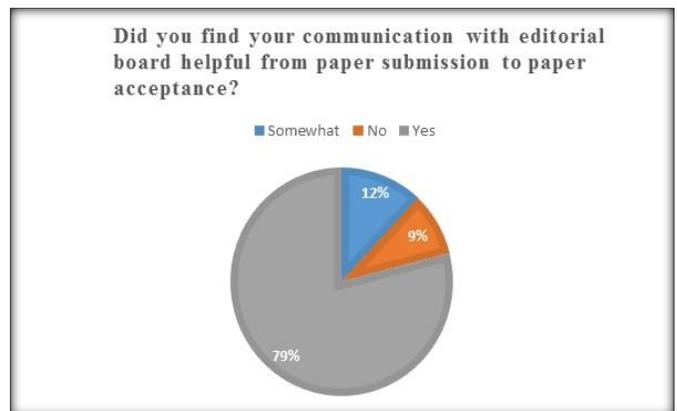


Fig. 7. Communication with editorial board.

All the gathered opinions from user over the questioners were processed and results were shown. For simplicity and better recommendations individual ratings of journal were found by considering positive and negative opinion from user. The results of some of the journals were shown respectively. For the pool of forty journals we tried to pick diverse journals. Following figures explain the experience of users with individual journals.

Fig. 8 reveals that researcher does not have good experience with “Acta Biomaterialia” journal. Average rating of this journal is 2.

Fig. 9 depicts that researchers have good experience with this journal named as “Biological Psychiatry”. Average rating of this journal is 3.

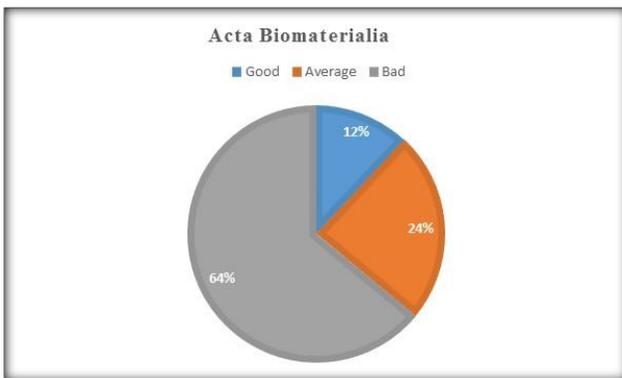


Fig. 8. Acta Biomaterialia.

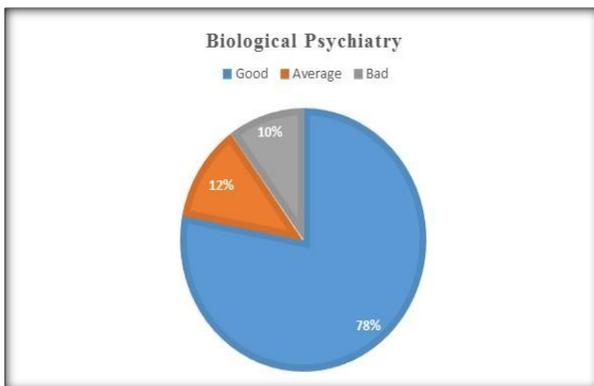


Fig. 9. Biological psychiatry.

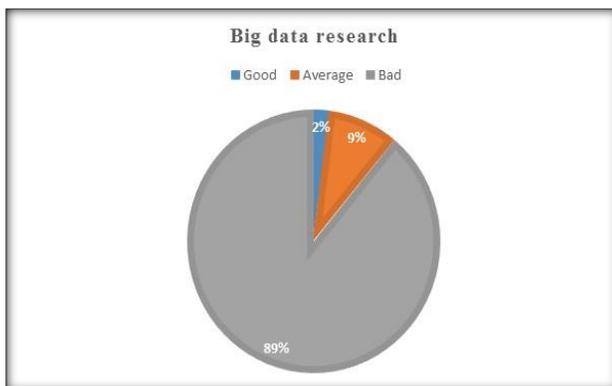


Fig. 10. Big data research.

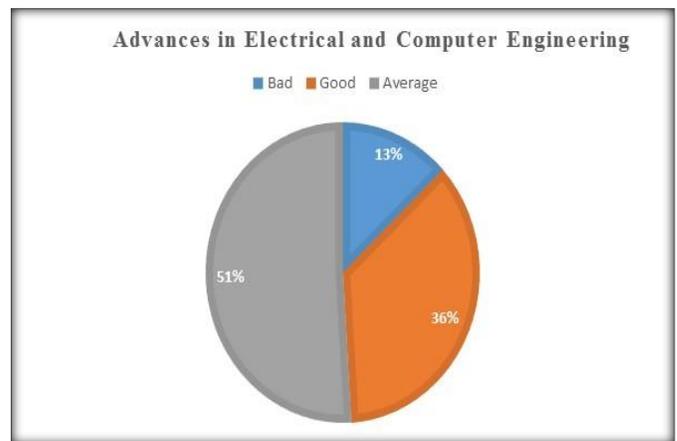


Fig. 11. Advances in electrical and computer engineering.

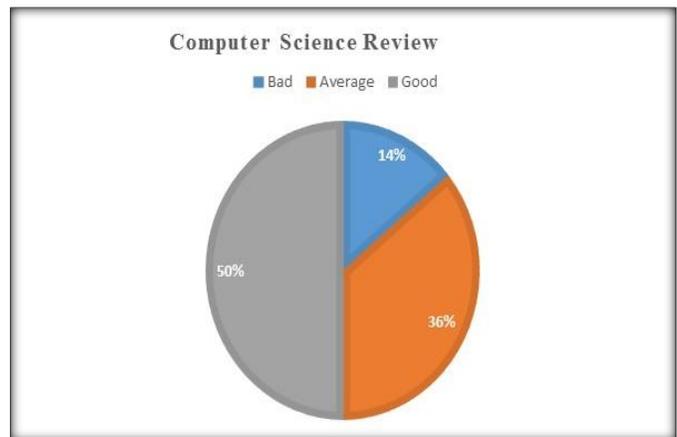


Fig. 12. Computer science review.

As per information, the experience of a client with “Big data research” journal is awful. Average rating of this journal is 1. It is described in Fig. 10.

It has been shown in the Fig. 11 that researchers have average experience with “Advances in Electrical and Computer Engineering” journal.

Average rating of this journal is 4.

Data indicated that most of the researchers have good experience with “Computer Science Review” journal. Average rating of this journal is 4. It is described in Fig. 12.

B. Data Set Description for Archive Data

For recommendations, we collected another dataset based on survey data collected in the above-mentioned step. At-least 40 research papers were collected along with their title, abstract and keywords for every journal against which user have provided the information in the survey. Journal attributes were also collected; which includes aim and scope of the journal, impact factor, and publication frequency and cite score. The user provides the title of the research paper, abstract and keywords in the form of text which is considered as input. Firstly, recommendations are generated within the dataset. Then the recommendations are proposed by combining the user opinion.

C. Journal Recommendations using Hybrid Approach

As we have processed the survey data and the results of user experience is available. Now, we are going to recommend the journals by combining simple journal recommendations with the user opinion. As defined above, term to term correlation is used to check the similarity.

To generate the journal recommendations, a query in the form of abstract is given which was related to computer science and big data. For checking the similarity value of a given query, it is added to the previously collected dataset of journal papers.

Results reveal in Fig. 13 that the given query has the most similarity with “Big data research” journal.

According to survey data, the average rating of “Big data research” journal is 3. So, it can be suggested the author submit the paper in this journal.

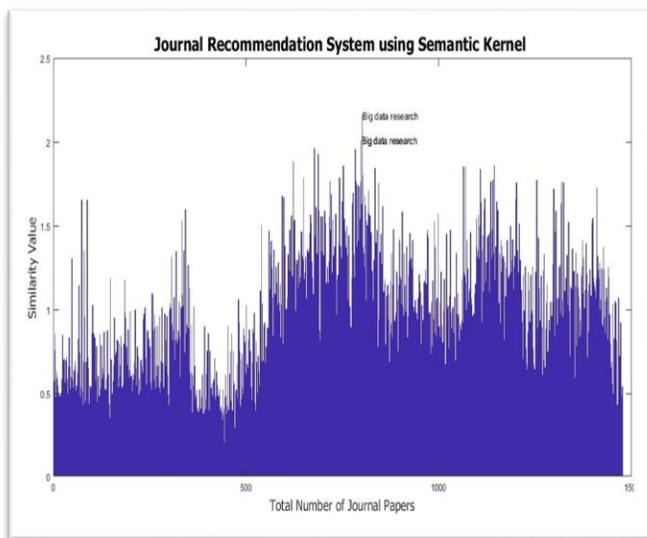


Fig. 13. Big data abstract.

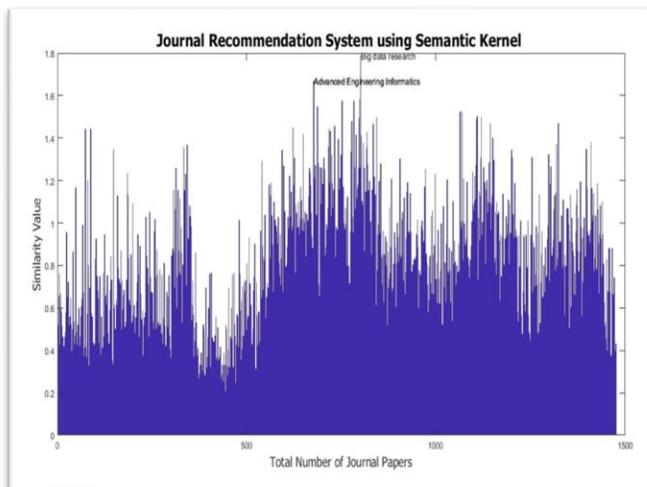


Fig. 14. Big data and information technology keyword.

A query in the form of a keyword is provided which relates to Information technology and big data combined with bioinformatics. Recommendation results show in Fig. 14 that the given keyword has the best match with “Advanced Engineering Informatics” and “Big data research” journal. As per overview information, the normal rating of “Big data research” journal and “Advanced Engineering Informatics” is 3 and 4, respectively. Along these lines, the author can choose among these two journals according to the priority.

A general keyword related to big data is used as a query to check the recommendations about the journal. Similarity value in Fig. 15 indicates that provided query has higher similarity with “Big data research” journal and is also similar to “Big data Analytics” journal.

As per survey result data, “Big data analytics” journal has an average rating of 2 and “Big data research” journal has an average rating of 3. Thus, the researcher can pick among these two journals as indicated by the need.

For journal recommendations, a keyword related to bio is added in the dataset. Similarity value in Fig. 16 indicates that it is suitable to choose “Biological Psychiatry” journal for the provided query.

Also, the survey results give 4 rating to this journal.

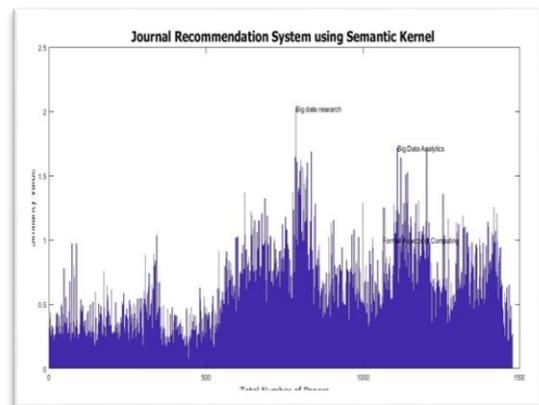


Fig. 15. General keyword for big data.

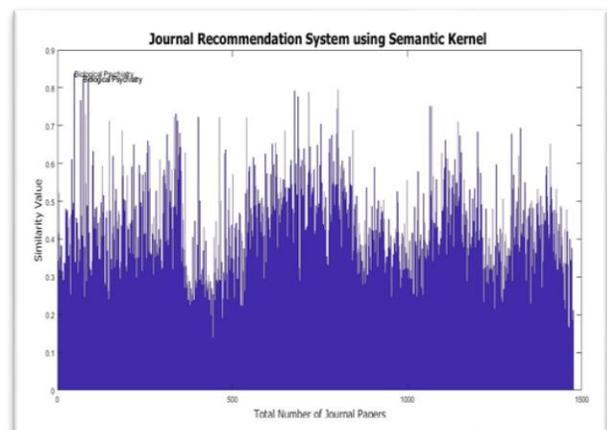


Fig. 16. Bio keyword.

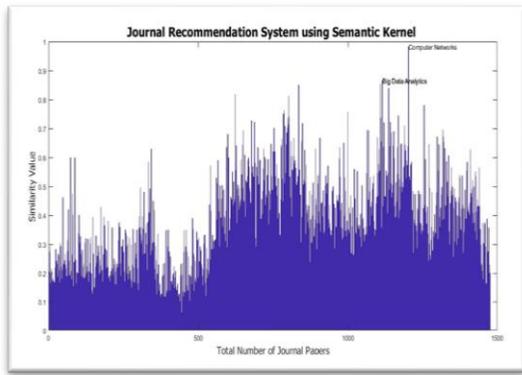


Fig. 17. Keyword for computer network.

Keyword related to the network is introduced in the dataset as a testing query which clearly has highest similarity with “Computer Networks” journal as show in Fig. 17.

Rating for “Computer Network” journal is 2.

V. CONCLUSION

In journal recommendation system better results were achieved using both user opinion and archives. The results show that our model will help researchers to fasten the paper submission procedure which enhance user experience.

Similarly, the selection of good similarity measure for semantic analysis is vital part of our proposed framework. In addition, the proposed work will be optimized for web-based application which helps us in making the user experience better.

In conclusion, this work may pave the way to other domains which certainly have impact on the life of the user.

In future, we aim to implement this work in different tools like Hadoop and spark in order to compare their relative recommendation accuracy.

REFERENCES

- [1] Kulkarni, S., *A Recommendation Engine Using Apache Spark*. 2015
- [2] Kim, M.C. and C. Chen, A scientometric review of emerging trends and new developments in recommendation systems. *Scientometrics*, 2015. 104(1): p. 239-263.
- [3] Lu, J., et al., *Recommender system application developments: a survey*. *Decision Support Systems*, 2015. 74: p. 12-32.)
- [4] Bröchner, J. and B.C. Björk, *Where to submit? Journal choice by construction management authors*. *Construction Management and Economics*, 2008. 26(7): p. 739-749.
- [5] Gutknecht, C., *Where to publish? Development of a recommender system for academic publishing*. 2014, University of Applied Sciences and Arts Northwestern Switzerland.
- [6] Gipp, B., J. Beel, and C. Hentschel. Scienstein: A research paper recommender system. In *Proceedings of the international conference on emerging trends in computing (icetic'09)*. 2009.
- [7] Sugiyama, K. and M.-Y. Kan. Scholarly paper recommendation via user's recent research interests. In *Proceedings of the 10th annual joint conference on Digital libraries*. 2010. ACM.
- [8] Beel, J., et al. A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation. In *Proceedings of the international workshop on reproducibility and replication in recommender systems evaluation*. 2013. ACM

- [9] Beel, J., et al. Introducing Docear's research paper recommender system. In *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*. 2013. ACM.
- [10] Al Alshaiikh, M., G. Uchyigit, and R. Evans, A research paper recommender system using a dynamic normalized tree of concepts model for user modelling. 2017.
- [11] Hristakeva, M., et al. Building recommender systems for scholarly information. In *Proceedings of the 1st Workshop on Scholarly Web Mining*. 2017. ACM.
- [12] Raamkumar, A.S., S. Foo, and N. Pang, Using author-specified keywords in building an initial reading list of research papers in scientific paper retrieval and recommender systems. *Information Processing & Management*, 2017. 53(3): p. 577-594.
- [13] Tarus, J.K., Z. Niu, and G. Mustafa, Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artificial Intelligence Review*, 2017: p. 1-28.
- [14] Beel, J., et al. Stereotype and Most-Popular Recommendations in the Digital Library Sowiport. In *Everything Changes, Everything Stays the Same? Understanding Information Spaces. Proceedings of the 15th International Symposium of Information Science (ISI 2017)-(isi2017)*. 2017. Humboldt-Universität zu Berlin
- [15] Deldjoo, Y., et al. Toward building a content-based video recommendation system based on low-level features. in *International Conference on Electronic Commerce and Web Technologies*. 2015. Springer
- [16] Esparza, S.G., M.P. O'Mahony, and B. Smyth, *Mining the real-time web: a novel approach to product recommendation*. *Knowledge-Based Systems*, 2012. 29: p. 3-11.
- [17] Jakob, N., et al. Beyond the stars: exploiting free-text user reviews to improve the accuracy of movie recommendations. in *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. 2009. ACM.
- [18] Beel, J., Towards effective research-paper recommender systems and user modeling based on mind maps. *arXiv preprint arXiv:1703.09109*, 2017.
- [19] Gopalan, P., J.M. Hofman, and D.M. Blei, *Scalable recommendation with poisson factorization*. *arXiv preprint arXiv:1311.1704*, 2013.
- [20] Moh, T.-S. and S. Bhagvat. Clustering of technology tweets and the impact of stop words on clusters. in *Proceedings of the 50th Annual Southeast Regional Conference*. 2012. ACM.
- [21] Shang, S., et al. A micro-video recommendation system based on big data. in *Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on*. 2016. IEEE.
- [22] Zhao, Z.-D. and M.-S. Shang. User-based collaborative-filtering recommendation algorithms on hadoop. in *Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on*. 2010. IEEE.
- [23] kumar Bokde, D., S. Girase, and D. Mukhopadhyay. An item-based collaborative filtering using dimensionality reduction techniques on Mahout framework. in *4th Post Graduate Conference for Information Technology (iPGCon-2015)*, Sangamner, Published in Spvryan's International Journal of Engineering Science and Technology (SEST). 2015.
- [24] Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A., &Riedl, J. (2002, January). Getting to know you: learning new user preferences in recommender systems. In *Proceedings of the 7th international conference on Intelligent user interfaces* (pp. 127-134). ACM.
- [25] Kotsogiannis, I., E. Zheleva, and A. Machanavajjhala. Directed Edge Recommender System. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. 2017. ACM.
- [26] Borràs, J., A. Moreno, and A. Valls, *Intelligent tourism recommender systems: A survey*. *Expert Systems with Applications*, 2014. 41(16): p. 7370-7389.
- [27] Farahat, A.K. and M.S. Kamel. *Document clustering using semantic kernels based on term-term correlations*. in *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on*. 2009. IEEE