International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

A Novel Automatic Journal Recommender System

S. Prasanna Priya, M. Karthikeyan

Abstract: Recommender Systems (RS) have developed into an important application in several user domains. RS may aid users to discover appropriate items in vast data. The selection of journal publication is generally based on the research domain or topics of document. The traditional method for journal recommendation is carried out by analyzing the document and matching its topics with relevant journal utilizing content-based examination. Though, this approach might create errors because of disparities in manuscript comparisons. In this paper, a novel Jaccard based Journal Finder Neural Network is proposed with pearson correlation coefficient (JJFNN-PC) for journal recommendation. The proposed recommender system allows the researchers to automatically find appropriate publication with journal title and abstract. Similarity coefficient is computed among the journal database and journal title and abstract of user distinctly through Jaccard similarity. The obtained outcome is used for training the novel JFNN that automatically find appropriate publication for user research article. The pearson correlation coefficient is established to validate the correlation between title and abstract of the recommended journal. The experimental result of automatic journal selection process provides the exact journal list and obtained better performance with accuracy of 98.41%.

Keywords: Recommender system, journal, JFNN, Jaccard's similarity, Pearson Coefficient.

I. INTRODUCTION

The advancement in the web technologies in the recent past has attracted both common people and the e-commerce society to have their trade-off with mutual benefits. The e-commerce organization has the ultimate goal for profit with enhanced customer satisfaction and retention [1]. For accomplishing their objectives, the e-commerce society has established a recommender systems (RS) that suggest the products based on the preferences and the behavior of the user [2]. The RS has solved the problem of plenty for the customers/users. Each sites has their own RS that are probable to fit their customers requirement [3].

Many researchers had developed different RS for various domains, however the attention in RS grows with increased user demands on products. Henceforth, there is a requirement for the high quality RS with the capability to provide recommendation at fine-tuned level for complex applications [4]. One such complex domain is the research paper recommendation, which is valuable resource for the researchers over the vast amount of journals in the repository [26]. Many researchers find it difficult to publish their work in the appropriate journal and experience delay in publication. Hence in the present work, a novel Journal Finder Neural Network (JFNN) was proposed to aid the researchers. The proposed system is fundamentally developed with the novel SIGMOID2 function with Jaccard Similarity (JS) for automatic journal recommendation that provide an improved search performance with better accuracy.

The following contributions are accomplished to propose the intended RS as:

1. Developing JS that retrieves all similar journals with journal title and abstract.
2. Improving the automatic journal recommendation process and provide the exact journal selection of search with JFNN–Sigmoid2 based on JS.
3. Estimating the fitness of recommendation with Pearson correlation coefficient and estimate the proposed RS performance.

The proposed RS is described in the subsequent sections as: Section 2 includes the works related on RS; Section 3 clarifies the proposed RS approach; Section 4 presents all the established results with inferences; the final section provides a conclusion.

II. REVIEW ON RELATED WORKS

The Collabrative filtering technique was implemented over the citation-web for generating the rating matrix among the scholarly papers. This work was intended for recommend certain additional features using the relationship of paper and its citation with the given paper. Six diverse algorithms are investigated for citation selection and evaluated offline [5].

The work distinguished the relations among the research papers to be direct and indirect with three approaches viz., content-based, citation analysis, and citation context to establish those relations. from comparison the citation analysis is found to be the appropriate method for recommending the research paper with additional features [6].

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
S. Prasanna Priya*, Assistant Professor, Thiru A. Govindasamy Govt Arts College, Tindivanam, Tamilnadu, India.
Email-id: prasannapriyatdm@gmail.com.
Dr. M. Karthikeyan, Assistant Professor, Department of Computer and Information Science, Faculty of Science, Annamalai University, Annamalainagar, Tamilnadu, India. Email-id: karthiaucse@gmail.com

A novel Scholarly paper recommender system was proposed based on the networking and social learning among the conference participants. The recommender algorithm provided the list of research papers based on the established social ties of participants with other conference [7,8]

Lu et al. 2014 generated the recommendations of news with multi-dimensional similarity that was integrated with Jaccard-Kmeans clustering.

The result provided has enhanced the scalability and the accuracy of the recommendation system despites of data sparsity [9].

The survey of text similarity approaches presented to measure a resemblance concerning sentences, words, text and paragraphs is a significant component in numerous challenges like data retrieval, word-sense disambiguation, text clustering, automatic essay scoring, machine conversion and document summarization [10]. The granularity among character N-gram and word N-gram in managing multilingual automated arrangement was furthermore examined the granularity among character N-gram and word N-gram in tackling multilingual automated arrangement. The experimental results displayed the word N-gram are much effectual than word granularity in resolving multilingual arrangement. Utilization of Jaccard index measures in two collections of patterns are same, and it tries to measure the resemblance concerning trusted patterns and recently discovered patterns, but it intended for particular problematic, lacking [11].

The Cluster-based similarity aggregation (CSA) was presented with programmed resemblance aggregating scheme for analogous ontology. The scheme initially computed five dissimilar fundamental measures to generate five resemblance matrixes. Pre-alignment is attained from the matrix to provide higher suggested scheme accuracy [12].

The usage of machine learning along with techniques for classifying the document automatically for managing large numbers of news articles, or web page descriptions, lightening the load on domain expert was extensively described. It developed to allow network-based utilization or workflow schemes to manage data much more effectually [13].

The automatic text classification was presented with the storage and retrieval of text. It offered the simplicity of utilizing and capability for dynamically acclimatizing to novel text forms. The simulated result showed the outcomes of document classification that fulfill the requirements of high-volume application [14]. The model was presented for automatically categorizes the text. Programmed text categorizations are usually demarcated as a text-based assignment of more predefined classifications to documents that was deliberated as text-based RS [15].

The Content-based Journals RS was presented with a hybridization of softmax regression and chi-square feature selection. It recommends the realted journals by order of priority with the document abstract, and the experimental outcome provided 61.37% accuracy [16].

Several works in the literature highlights supporting access and navigation problem in large scientific publications libraries, mostly from a perspective of Information Retrieval. Yet there are numerous works who have taken into justification personalization-based methods to difficult, significant to improve the RS before search engines. In this paper, novel JJFNN-PC with sigmoid2 was proposed to enhance the results provided by the similarity measure designed. And also improve the automatic journal recommendation to search and select the appropriate journal with better accuracy

III. METHODOLOGY

The main goal of JJFNN-PC is providing personalized access to documents retrieved. The overall design flow of JFNN with Sigmoid2 is shown in fig 1. Web User Interface devoted to the user for profile creation and management, once the user uploaded the title and abstract of the paper. The given informations are extracted to discover the similarity of journals by using the JS. A JFNN with sigmoid2 function devoted to build and maintain individual user profiles; retrieve query results from the scientific journal collection and then recommend appropriate journals. The outcomes from JFNN is evaluated with the pearson correlation coefficient to enhance the accuracy on journal recommendation.

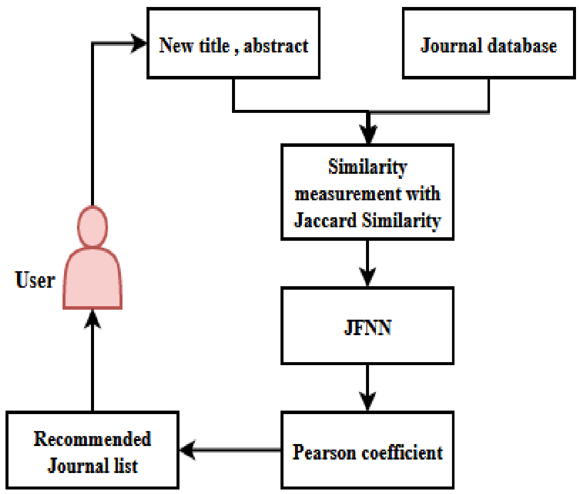

Fig. 1.Overall design flow of JJFNN-PC with Sigmoid2

Jaccard similarity

The Jaccard similarity between any two objects or data, and b are estimated with the equation (1) as

$$J(A,B)=|A\cap B|/|A\cup B| \quad (1)$$

Similarity measure runs faster and also improves the quality of results. In the current work, two different features are employed for the journal recommendataion i.e. the journal title and journal abstract. Let J[1], J[2] … J[P] be the sets of journal articles JA with journal title (TX) and journal abstract (AX). If the journal paper of user has the title (TY) and abstract (AY) the similarity measures among the two title and abstract are estimated as in equation (2) and (3).

For title: $J(TX,TY)=|TX\cap TY|/|TX\cup TY|$ (2)

For abstract: $J(AX,AY)=|AX \cap AY|/|AX \cup AY|$ (3)

The cumulative similarity for the journal n the journalist is given as

Jaccard_Sim= J(TX,TY)+ J(AX,AY) (4)

When the estimated Jaccard_Sim has the value higher than 0.75, the journal is found to be recommended.

Pseudo code for Jaccard similarity

Algorithm 1 Jaccard Similarity Measure

1: Procedure [JS_Title, JS_Abstract, Jaccard_Sim, Recom, Not_- Recom] = JSMAbstract, Title, Journals

Each Journal contains Research Articles

2: Journal← {J{1}, J{2},…,J{p}}

3: $J\{1\} \leftarrow \{J_1A_1, J_1A_2, \dots, J_1A_{u1}\}$;

4: $J\{2\} \leftarrow \{J_2A_1, J_2A_2, \dots, J_2A_{u2}\}$;

5: $J\{p\} \leftarrow \{J_pA_1, J_pA_2, \dots, J_pA_{up}\}$

6: $AX \leftarrow Abstract$; $TX \leftarrow Title$;

7: $for\ i \leftarrow 1 to |Journals|$ do

8: $for\ j \leftarrow 1 to |Journals\{i\}|$ do

9: AY← abstract of J_iA_j; TY←Title of J_iA_j

10: $JS_{Title}(i,j) \leftarrow |TX \cap TY|/|TX \cup TY|$

11: $JS_{Abstract}(i,j) \leftarrow |AX \cap AY|/|AX \cup AY|$

12: $Jaccard_{sim}(i,j) \leftarrow JS_{Title}(i,j) + JS_{Abstract}(i,j)$

13: end

14: end

15: $Row \leftarrow$ Number of rows in $Jaccard_{sim}$

16: $Column \leftarrow$ Number of columns in $Jaccard_{sim}$

17: $Recom \leftarrow \emptyset$

18: for $i \leftarrow 1$ to Row do

19: for $j \leftarrow 1$ to $Column$ do

20: if $Jaccard_{sim}(i,j) > Threshold$ then

21: $Recom \leftarrow [Recom, i]$

22: break

23: end

24:end

25:end

26: $Not\_Recom \leftarrow \{1,2,\dots,Row\} - Recom$ Set difference

Journal Finder Neural Network (JFNN)

It contains three sequential layers as shown in fig 2. Each scheme fundamentally has a three successive layered scheme as namely input, output and hidden layer.

Input layer: The input values are provided.for the journal recommendation through the Proposed JFNN, three inputs are provided viz., Jouranal ID, journal title (JT) and journal abstract (JA). Hence there are three nodes at the input layer.

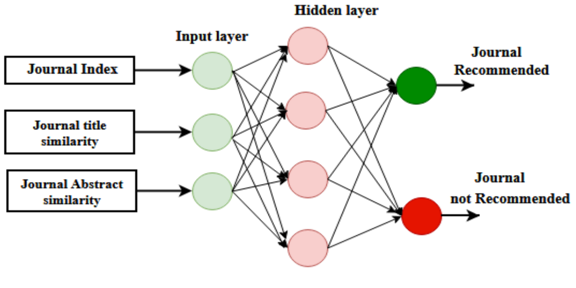

Fig. 2.The general structure of JFNN

*Hidden layer(s):*In general hidden layers are a collection of neurons among in and out layers, it might be either single or multiple layers. In the proposed JFNN there are four nodes at the hidden layer.

Output layer: Generally, it has two nodes and the outcome values might be, 0 (not recommended) and 1 (recommended)

Hidden units are added to the network one by one. Every fresh hidden element accepts a weighted association from each and every net's actual inputs and likewise from each earlier hidden unit. Each new unit, therefore, adds a new single-unit layer to the network. The processing capabilities are stowed in internal unit association strengths, called weights. Input strength is contingent on weight ranges. Weight range may be negative-positive, or zero. Negative weight denotes the signal is decreased or inhibited. Zero weight denotes that there is no any association between two neurons. Weights are adjusted to attain the necessary outcome. This process of adjusting weights of nodes is performed during the training and the obtained JFNN may either recommend a journal or not recommend it.

The most normally used sigmoid activation function is a given by

$$F(x) = \frac{1}{1+e^{-sum}} \quad (3)$$

Where, $sum = \sum_{i=1}^{n} x_i W_i$ (4)

The sum denotes a weighted inputs and weights product summation between the layers. For better learning and accuracy, a novel sigmoid2 activation function is introduced as in equation (5) for the layers.

$$\sigma(x) = \frac{4 \times ex}{(1+ex)^2} \quad (5)$$

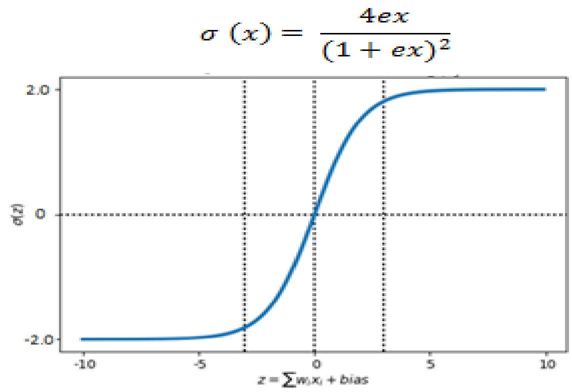

Fig. 3.Sigmoid2 function

Pseudocode for sigmoid2 function

Algorithm 2 Sigmoid2 function

1: $procedure\ sig = SIGMOID2\ (value)$

2: $ex \leftarrow exp(value)$

3: $sig \leftarrow \frac{4.0 \times ex}{(1+ex)^2}$

Pseudocode for JFNN with sigmoid2

Algorithm 3 Journal Finder Neural Network with SIGMOID2

1: $Procedure\ h = JFNN\ (JS_{Title}, JS_{Abstract}, Jaccard_{sim}, Recom, Not\_Recom)$

2: $Rows \leftarrow$ Number of rows in $Jaccard_{sim}$

3: for $i \leftarrow 1$ to $Rows$ do

4:

$x(i,:) \leftarrow [JS_{Title}(i,:), JS_{Abstract}(i,:), Jaccard_{Sim}(i,:)]$

5:**end**

6:$op\_units \leftarrow zeros(1, Row)$;$op\_units\,(Recom) \leftarrow 1$

7: $W_f, U_f, b_f \leftarrow$ Set of learnable parameters for the forget gate.

8: $W_{\tilde{c}}, U_{\tilde{c}}, b_{\tilde{c}} \leftarrow$ Another set of learnable parameters.

9: $W_i, U_i, b_i \leftarrow$ Set of learnable parameters, defined for the input gate.

10:$W_o, U_o, b_o \leftarrow$ Set of learnable parameters, defined for output gate.

11: $h_0 \leftarrow zeros\,(1,p)$; $C_0 \leftarrow h_0$

12: for $t \leftarrow 1$ to n do

13: $f(t) \leftarrow \sigma(W_f x(t) + U_f h(t-1) + b_f)$

14: $\tilde{C}(t) \leftarrow SIGMOID2(W_{\tilde{c}} x(t) + U_{\tilde{c}} h(t-1) + b_{\tilde{c}})$

15: $i(t) \leftarrow \sigma(W_i x(t) + U_i h(t-1) + b_i)$

16: $C(t) \leftarrow f(t) \odot C(t-1) + i(t) \odot \tilde{C}(t)$

17: $O(t) \leftarrow \sigma(W_o x(t) + U_o h(t-1) + b_o)$

18: $h(t) \leftarrow SIGMOID2(C(t)) \odot O(t)$

19: **end**

20: Set ip_units, $jfnn\_units$ and optimizer to define $JFNN$ Network (L)

21: Normalize the dataset (Di) into values from 0 to 1.

22: Select training window size (tw) and organize Di accordingly

23: for $n \leftarrow 1$ to $Batch\_size$ do

24: Train the Network (L)

25: **end**

26: Run Predictions using L

27: Calculate the loss function

Pearson Coefficient

Pearson Coefficient computes similarity by drawing a line between attributes of two objects. Correlation between two objects results in a positive slope line. Pearson Coefficient is more robust to un-normalized data. For the given JS title and JS abstract, the aggregate column and rows of JS title as well as abstract are considered as the matrix by sum together the matrix; the coefficient value is achieved. Pearson Coefficient is calculated by the formula below

$$\rho X,Y = \frac{cov(JT,JA)}{\sigma JT \sigma JA} \quad (5)$$

Pseudocode for the Pearson coefficient

Algorithm 4 Pearson Coefficient

1: $[P\_C] = Perron\_Coefficient\,(JS_{Title}, JS_{Abstract})$

2: $Row \leftarrow$ Number of rows in JS_{Title};

3: $Columns \leftarrow$ Number of columns in JS_{Title};

4: $N \leftarrow Rows \times Columns$

5: $SumX \leftarrow \sum_{i=1}^{p} \sum_{j=1}^{up} JS_{Title}(i,j)$;

6: $Squar\,e_{sum}X \leftarrow sumX + sum\,X^2$

7: $SumY \leftarrow \sum_{i=1}^{p} \sum_{j=1}^{up} JS_{Abstract}(i,j)$

8:$Squar\,e_{sum}Y \leftarrow sumY + sum\,Y^2$

9: $SumXY \leftarrow \sum_{i=1}^{p} \sum_{j=1}^{up} (JS_{Title}(i,j) \times JS_{Abstract}(i,j))$;

10: $P\_C \leftarrow \frac{N(sumXY - SumX)sumY}{\sqrt{N^2(Squar\,e_{sum}X - SumX^2)(Squar\,e_{sum}Y - sumY^2)}}$

IV. RESULTS AND DISCUSSION

In the present work, the computer science based journal papers are collected for its title and abstract over 10 journals with 50 research articles. For training the JJFNN-PC, 60% of data is used for training and testing is performed with the remaining data. The experimental outcomes of JJFNN-PC provide the automatic journal selection process for better search performance are provided in this section.

Accuracy

Accuracy might be computed with the percentage of properly classified instances as given below

$$Accuracy = \frac{no.of\ correctly\ classified\ relavent\ journals}{Total\ classification\ of\ journals} \quad (6)$$

The experimental result of proposed JJFNN-PC model provides the better accuracy of 94.81%.

Precision

The precision metric exhibits the positive class predicted exactness to be positive among the true class in the proposed RS.

$$precision = \frac{no.correct\ relavent\ Journals\ recommended}{total\ no.of\ journal\ recommended} \quad (7)$$

The proposed JJFNN-PC obtained the better precision of 100%

Recall

The recall metric exhibits the exactness with which the positive class is predicted to be positive among all relavent journals through the proposed RS.

$$Recall = \frac{no.correct\ relavent\ Journals\ recommended}{total\ no.of\ relavent\ journal\ in\ dataset} \quad (8)$$

The proposed JJFNN-PC model obtained the better recall of 93.78 %.

F-Measures

The *F measure* is to measure the accuracy-test and is *defined* as the weighted harmonic mean of the test recall and precision.

$$F = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

The experimental results of proposed JJFNN-PC model obtained the better F-measure of 96.76 %

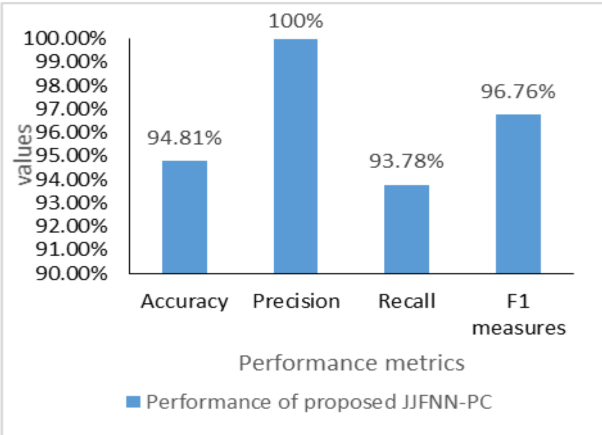

Fig. 4.Performance of Proposed JJFNN-PC

V. CONCLUSION

The JJFNN-PC can greatly facilitate the task of searching for a scientific journal. The users have to spend their efforts and time for accessing the knowledge contained in scientific publications. Errors in matching function are seen in the traditional journal recommendation system.

The proposed JJFNN-PC is used for the automatic journal selection process to provide a better search performance. The JJFNN-PC showed that employing JS and pearson correlation coefficient has improved the accuracy performance for finding the exact journal.

Jaccard consumed less time for similarity value prediction than the existing methods. In conclusion, the iterations in similarity checking algorithm determined the operating time. The accuracy, precision, sensitivity, and F-score of the proposed RS are measured as 94.81%, 100 %, 93.78 % and 96.76 % respectively. The proposed RS can be used for other domain based journal selctions and also upgraded with the facility for effective journal recommendation system with other inputs like citation score, publishing term etc.

Conflict of Interest: We have no conflicts of interest to disclose.

REFERENCES

1. Ramzan B, Bajwa IS, Jamil N, Amin RU, Ramzan S, Mirza F, Sarwar N. An Intelligent Data Analysis for Recommendation Systems Using Machine Learning. Scientific Programming. 2019;2019.
2. Pu P, Chen L, Hu R. Evaluating recommender systems from the user's perspective: survey of the state of the art. User Modeling and User-Adapted Interaction. 2012 Oct 1;22(4-5):317-55.
3. Sarwar BM, Karypis G, Konstan J, Riedl J. Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering. InProceedings of the fifth international conference on computer and information technology 2002 Dec 27 (Vol. 1, pp. 291-324).
4. Khusro S, Ali Z, Ullah I. Recommender systems: issues, challenges, and research opportunities. InInformation Science and Applications (ICISA) 2016 2016 (pp. 1179-1189). Springer, Singapore.
5. Sibaroni Y, Widyantoro DH, Khodra ML. Survey on research paper's relations. In2015 International Conference on Information Technology Systems and Innovation (ICITSI) 2015 Nov 16 (pp. 1-6). IEEE.
6. Asabere NY, Xia F, Meng Q, Li F, Liu H. Scholarly paper recommendation based on social awareness and folksonomy. International Journal of Parallel, Emergent and Distributed Systems. 2015 May 4;30(3):211-32.
7. Xia F, Asabere NY, Liu H, Deonauth N, Li F. Folksonomy based socially-aware recommendation of scholarly papers for conference participants. InProceedings of the 23rd International Conference on World Wide Web 2014 Apr 7 (pp. 781-786).
8. Lu M, Qin Z, Cao Y, Liu Z, Wang M. Scalable news recommendation using multi-dimensional similarity and Jaccard–Kmeans clustering. Journal of Systems and Software. 2014 Sep 1;95:242-51.
9. Gomaa WH, Fahmy AA. A survey of text similarity approaches. International Journal of Computer Applications. 2013 Apr 13;68(13):13-8.
10. Lecluze C, Rigouste L, Giguet E, Lucas N. Which granularity to bootstrap a multilingual method of document alignment: character N-grams or word N-grams?. Procedia-Social and Behavioral Sciences. 2013 Oct 25;95:473-81.
11. Tran QV, Ichise R, Ho BQ. Cluster-based similarity aggregation for ontology matching. Ontology Matching. 2011 Oct 24;814.
12. Calvo RA, Lee JM, Li X. Managing content with automatic document classification. Journal of Digital Information. 2004 Jun;5(2).
13. Appiani E, Cesarini F, Colla AM, Diligenti M, Gori M, Marinai S, Soda G. Automatic document classification and indexing in high-volume applications. International Journal on Document Analysis and Recognition. 2001 Dec 1;4(2):69-83.
14. Goller C, Löning J, Will T, Wolff W. Automatic Document Classification-A thorough Evaluation of various Methods. ISI. 2000 Nov;2000:145-62.
15. Wang D, Liang Y, Xu D, Feng X, Guan R. A content-based recommender system for computer science publications. Knowledge-Based Systems. 2018 Oct 1;157:1-9.
16. Fan Z, Chen S, Zha L, Yang J. A text clustering approach of Chinese news based on neural network language model. International Journal of Parallel Programming. 2016 Feb 1;44(1):198-206.
17. Elmongui HG, Mansour R, Morsy H, Khater S, El-Sharkasy A, Ibrahim R. TRUPI: Twitter recommendation based on users' personal interests. InInternational Conference on Intelligent Text Processing and Computational Linguistics 2015 Apr 14 (pp. 272-284). Springer, Cham.
18. Costa H, Furtado B, Pires D, Macedo L, Cardoso A. Recommending POIs based on the User's Context and Intentions. InInternational Conference on Practical Applications of Agents and Multi-Agent Systems 2013 May 22 (pp. 166-177). Springer, Berlin, Heidelberg.
19. Caraballo AA, Arruda NM, Nunes BP, Lopes GR, Casanova MA. TRTML-a tripleset recommendation tool based on supervised learning algorithms. InEuropean Semantic Web Conference 2014 May 25 (pp. 413-417). Springer, Cham.
20. Baldominos A, Albacete E, Saez Y, Isasi P. A scalable machine learning online service for big data real-time analysis. In2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD) 2014 Dec 9 (pp. 1-8). IEEE.
21. Aouay S, Jamoussi S, Gargouri F. Feature based link prediction. In2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA) 2014 Nov 10 (pp. 523-527). IEEE.
22. Pecli A, Giovanini B, Pacheco CC, Moreira C, Ferreira F, Tosta F, Tesolin J, Dias MV, Lima Filho SP, Cavalcanti MC, Goldschmidt RR. Dimensionality Reduction for Supervised Learning in Link Prediction Problems. InICEIS (1) 2015 Apr 27 (pp. 295-302).
23. Vialardi C, Chue J, Peche JP, Alvarado G, Vinatea B, Estrella J, Ortigosa Á. A data mining approach to guide students through the enrollment process based on academic performance. User modeling and user-adapted interaction. 2011 Apr 1;21(1-2):217-48.
24. Zeng L, Li Z. Text classification based on paragraph distributed representation and extreme learning machine. InInternational Conference in Swarm Intelligence 2015 Jun 25 (pp. 81-88). Springer, Cham.
25. Mitra V, Wang CJ, Banerjee S. Text classification: A least square support vector machine approach. Applied Soft Computing. 2007 Jun 1;7(3):908-14.
26. Champiri ZD, Asemi A, Binti SS. Meta-analysis of evaluation methods and metrics used in context-aware scholarly recommender systems. Knowledge and Information Systems. 2019 Nov 1;61(2):1147-78. doi:10.1007/s10115-018-1324-5