Journal of Information Systems Engineering and Management
2025, 10(29s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Improving Citation Recommendation Accuracy Using an SVM-Based Model

Velkumar K[1], Balasubramanian Chokkalingam[2]

[1]Research Scholar, Department of Computer Science and Engineering, Kalasalingam Academy of Research Education, Krishnankovil, Tamilnadu, India, Department of Computer Science & Engineering, Koneru Lakshmaiah Education Foundation, Green Fileds, Vaddeswaram, A.P. – 522302, velkumarskc@gmail.com

[2]Associate Professor, Department of Computer Science and Engineering, Kalasalingam Academy of Research Education, Krishnankovil, Tamilnadu, India

ARTICLE INFO

Received: 30 Dec 2024

Revised: 19 Feb 2025

Accepted: 27 Feb 2025

ABSTRACT

This work presents an SVM-based citation recommendation system, which will path the way for easy identification of relevant research papers. The methodology of the proposed system starts with preprocessing of the input dataset. Here, Using stochastic matching pattern scheme the function words were removed in order to construct a structured library. Canonicalization of documents is achieved through the application of a probabilistic text normalization algorithm. Normalization of data has been done, and now it is ready for further processing.
classified by employing the Support Vector Machine (SVM) algorithm. This algorithm is particularly oriented towards high-dimensional data and effective in making separations of intricate patterns. In the end, the system ranks citations according to recommendations and measures its performance with evaluations like MAP, MRR, precision, recall, and F1-score. Resultant evidence illustrates that this SVM-based system surpasses a number of conventional models—thereby proving its capability to enhance the citation recommendation process.

Keywords: Citation Recommendation, Support Vector Machine (SVM), Probabilistic Text Normalization, Stochastic Matching Pattern, Citation Ranking.

INTRODUCTION

Recommendations are becoming more significant these days, and they also change the way that users and websites communicate. Many domains, including education, science, economics, and more, have made extensive use of recommender systems [1]. As laborious yet important, researchers view the task of finding relevant scientific publications. Citations are generally regarded as the bibliographic entries included in reference lists, bibliographies, or footnotes, which provide information necessary to verify the source documents and identify relevant publications. Citation analysis, also known as reference list analysis, is very important in all study fields from the humanities to the social sciences and sciences [3-5]. Studies on citation analysis were conducted to quantify the value of scholarly research, which is crucial for determining the journal's Impact Factor (JIF), which ultimately determines the value of the journal and where research can be published.

The citation recommendation system offers an efficient and effective way of recommending scientific papers that are pertinent to researchers and also tackles existing citation problems. Citation recommendations for researchers help in unearthing pertinent literature.reviews faster and is a growing topic in the area of study. Citation suggestion tools are usually divided into two categories according to their patterns: local citation and citation recommendation.

Recently, Support Vector Machine (SVM)-based models have gained significant attention in various fields, including natural language processing and information retrieval, due to their robustness and ability to handle high-dimensional data effectively. SVM models excel at separating complex patterns and ensuring high classification accuracy, making them well-suited for tasks like citation recommendation.

The influence of SVM in recommender systems is pervasive, showcasing its efficiency in leveraging retrieved information to provide precise and relevant recommendations. Nevertheless, the swift emergence of new research necessitates the reevaluation of current citation recommendation frameworks, prompting the need for the development of enhanced approaches to facilitate a deeper comprehension of the area. This study introduces an innovative SVM-based method for citation recommendation, acknowledging the increasing potential and adaptability of SVM in recommendation systems. A thorough analysis of previous works was conducted in the succeeding literature survey section, presenting new insights to inform future developments in this field.

The principal objectives of the proposed project are as follows:

• To analyse the citation record data by removing function words and creating a library with a stochastic pattern matching technique.
• To employ the Probabilistic min-match technique for the canonicalization of documents.

• To classify the data using a support vector machine approach and assign rankings to citations.
• To evaluate the effectiveness of the suggested methodology in terms of accuracy, precision, recall, and F1-score.

RELATED WORKS

Citation recommendations facilitate the identification of citable content customised to the user's particular informational requirements. A multitude of models have been suggested to enhance citation recommendation. This section provides a succinct summary of the existing literature employed for citation recommendations.

In the study [6], the author examines citation suggestion through the framework of semantic representation, focussing on the content and interrelations of the referenced works. Initially, four types of citation contexts were identified and extracting from the cited content articles, considering of co-citation relationships and the intent of citations. Subsequently, approximately 132 procedures were developed to produce the semantic vector of the referenced publication, encompassing four network embedding approaches, sixteen methods for merging four text representation algorithms, four citation content patterns, and around 112 fusion models. The similarity of the cited works was ultimately assessed to propose citations. The assessment methodology was developed to ascertain the suitable format of citation content according to the link prediction model. The acquired statistics imply an enhanced performance rate.

The purpose of the article [7] is to provide a concise overview of current research endeavours utilising the DL recommender system. Along with a succinct overview of current models, a taxonomy of deep learning-based recommendation schemes was also developed. Ultimately, the current tendencies were broadened, providing a fresh perspective on their emerging subject.

The primary purpose of the paper was to provide an innovative method for a citation recommendation system that facilitates the retrieval of appropriate citations for research publications in draft form[8]. This strategy assists to make draft articles full in citations select the proper references from the ocean of papers available. To create a citation recommender system for incomplete drafts, the contribution to the proposed recommendation model adds three approaches such as WordNet for vector transformation. This allows weight adjustments to be made according to structural variables and completeness degree of the material. The analysis indicates that all three approaches can potentially contribute toward significantly increasing suggestion accuracy.

This research [9] established the paper recommender system by delineating its advantages and importance. The discourse covered methodologies and algorithms of recommendation systems, including content-based filtering, collaborative filtering, hybrid models, and graph-based filtering. The estimation methodology for several recommender systems was executed, and the lingering issues were summarised in a report recommending a model that integrates unified academic data standards, serendipity, cold start, sparsity, and privacy considerations.
This study [10] established the paper recommender system by examining its importance and advantages. The methodologies and algorithms employed in recommendation systems—collaborative filtering, content-based filtering, hybrid models, and graph-based filtering—were examined. Following the assessment of multiple recommender systems, the unresolved challenges were consolidated into a paper recommender system that incorporates serendipity, cold start, sparsity, privacy, and unified academic data standards.
In [11], a remarkable mechanism for predicting a paper's long-term citations based on its citation performance in the initial years post-publication was introduced. An artificial neural network (ANN) was utilised to train the prediction of citation counts, a robust machine learning technique. The strategy utilising newly developed applications across several sectors, including text and picture processing systems. The evaluation indicates that the proposed strategy surpasses existing methods regarding forecast accuracy for both annual and total citation counts.
A distinctive context-aware NCN-dependent approach, integrating supplementary textual data and the Bidirectional Encoder Representations from Transformers (BERT) architecture, was developed to improve the efficacy of the citation recommender model task in [12]. A self-attention mechanism and a robust deep neural auto-encoder were employed to effectively learn both textual and contextual citation data relevant to the dataset. The effectiveness and notable improvement of the suggested technique are demonstrated by the performance evaluation on the standard dataset arXiv.
The Deep Cite model, a content-dependent hybrid neural network designed for citation suggestion, was presented in paper [13]. The advanced semantic representation vectors of the text were initially obtained using the BERT framework. The multi-scale CNN approach and the BiLSTM framework were utilised to extract both sequence and local information from the sentence context. The alignment of text vectors was undertaken methodically to obtain the candidate sets. Thus, the proposed strategy markedly improves the quality of reference recommendations. The

technique presented in this work significantly improves the quality of citation recommendations. The study in [14] introduced a context-aware citation recommendation method utilising an endwise memory network. Consequently, bi-LSTM-based models are employed to acquire representations of paper and citation contexts from this specification. Additionally, three real-time datasets were employed for the experimental assessment and evaluation of model performance.

To explore more appropriate citation patterns, the author of [15] proposed a new context-aware citation recommendation algorithm which may have a better improve on the orthogonality of weight matrix. The experimental (on the CiteSeer dataste) findings verify that, all terms evaluated measures; the proposed model is superior to all baseline models.

A global citation recommendation using generative adversial networks (GCR-GAN) was introduced in [16] as a network embedding model. This paper uses the Heterogeneous Bibliographic Network (HBN) to derive a very personalised range of citation recommendations. The proposed architecture for learning the semantic-preserving representation of a graph is Scientific Paper Embeddings with Citation-Informed Transformers (SPECTRE) and a denoising Autoencoder. The model's recommendations, use the DBLP and ACM datasets, demonstrate an 11% and 12% enhancement in performance relative to baseline models, measured by Mean Average Precision (MAP) and Normalised Discounted Cumulative Gain(nDCG), respectively.
In [17], a novel method for citation suggestion was introduced, which utilised content information and graph structure to create a content-based graph. Subsequently, we encoded the network architecture and performed co-authorship analysis to obtain the author-centric representation of the graph. The two representations were ultimately concatenated, serving as the node's feature vector and merging the author's knowledge with the material for a more precise network representation.

The obtained node vectors, known as cGAN and their variant VCGAN, were used to offer a novel personalised citation recommendation. The proposed method outperforms the current models once computed on the AAN dataset.

PROPOSED METHOD

This section delineates the comprehensive working approach of the proposed scheme. Figure 1 depicts the system flow. The initial input data comprises diverse papers, topics, and authors. The data is preprocessed to eliminate function words. The stochastic pattern matching procedure is utilised to decrease the overall character count, thereby enhancing the efficiency of search operations and similarity assessments between the testing and training databases.

Next, the canonicalization of documents is performed using the Min-Max Normalization approach. Following this, an SVM-based classification technique is employed for recommending similar citations. The Support Vector Machine algorithm is used to classify the processed datas and rank the citations based on relevance. Finally, the citation ranking is set according to the classification results, ensuring accurate and efficient citation recommendations.

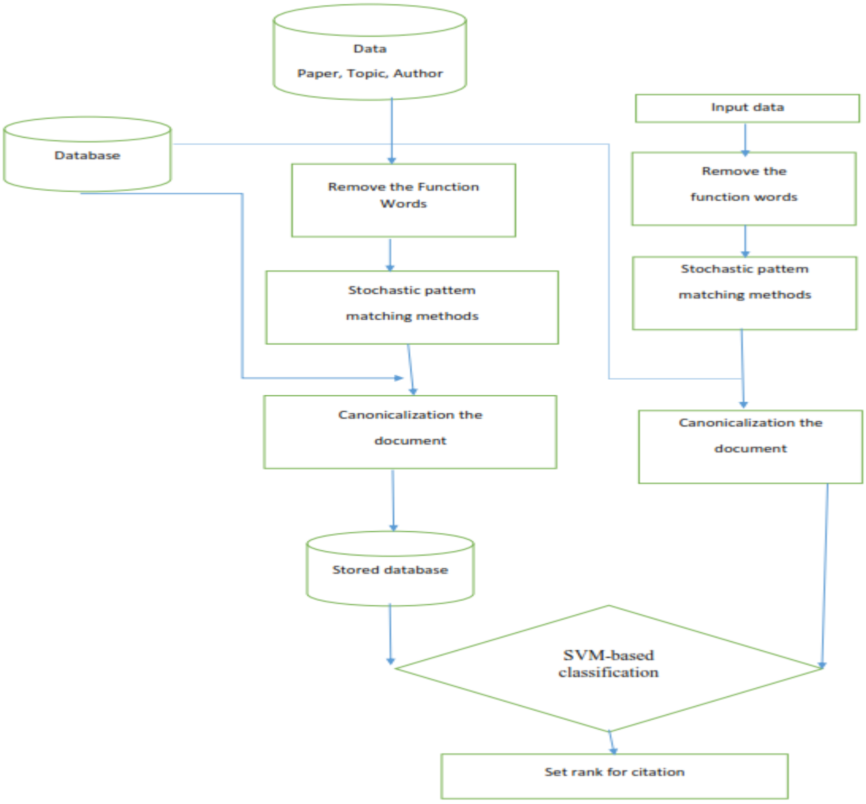

Figure 1 Propose Method Architecture

A. Data preprocessing

The purpose of the preprocessing stage is to reduce the text's character count. Texts are usually collections of words that can be divided using a tokenisation technique. The process of tokenisation involves breaking up the input data into small word segments known as tokens, which eliminates the low priority elements. The majority of texts contain certain common terms, and those that appear frequently in the provided corpus are referred to as stop or function words. The function words had a somewhat less important meaning as compared to other terms that appeared in the paper. They have no meaning relationships in the context and are commonly used words (such as "an," "a," "the," and "in"). Common and broad terms that usually don't add to the semantics of a document and have no additional reading value are known as function words. By assisting in the efficient selection of appropriate words, the function words elimination reduces the complexity of the text structure.

B. Stochastic Pattern matching

The stochastic pattern matching method is used once function words have been eliminated. The main goal of the stochastic pattern matching approach is to increase efficiency by reducing the number of characters that must be compared between the text and patterns. The search process can be made more efficient by changing the order of character comparisons in each try and adjusting the shift factor, which allows for the preset skipping of character locations in a text after each trial. This method allows text scanning by using a window with dimensions that match the length of the pattern. It can identify its players by checking if the link is asserted. The suggested method uses broader syntactic patterns than ASER (activities, states, events, and their relations), as causal relation participants do not need to be articulated as whole words.

The Badchar() function allows the host string pointer to move forward, prioritising matches for characters with lower probabilities based on their possibility. The stochastic pattern matching technique is used in the matching process to align the probabilities of uncommon and common letters in authentic English text. This will decrease the number of comparison times as much as feasible and predict a number of mismatch occurrences, which may allow the Badchar() function to be called earlier to realise the forward leap of the host string pointer. As a result, the library is produced using this approach. Below is the algorithm for this.

Input:

Dp = Dataset of previous papers

PMd = Database for pattern matching

Output:

TL = Generated feature library

1. **For each document** k in Dp(i):

Step 1: Preprocess the paper

Split the text of paper DP(k) into words:

Fw=DP(k).split()

2. **For each word** j in the list Fw of words:

Step 2: Perform pattern matching

Calculate the similarity score between the current word j and entries in the pattern matching database PM_d:

$$P(Fw) = \frac{\text{length of matching word list}}{Length\ of\ PMd}$$

Step 3: Store the maximum similarity score for the word j: SP(j)=max(P(Fw))

3. **End of** j-loop (Repeat for all words in Fw)

4. **End of** k-loop (Repeat for all papers in Dp)
5. **Step 4:** Finalize the library

After processing all words and papers, assign the feature library TL(i) for paper i as the list of maximum similarity scores SP for each word.

TL(i)=SP

C. Probabilistic Z-Score Algorithm for Document Canonicalization

The next phase in the proposed framework is the text canonicalization procedure. This step entails converting text files into a distinct and uniform format, facilitating the design of rules for information extraction. Document canonicalization denotes the process of standardising and transforming the representation of documents into a uniform or canonical format. This guarantees the resolution of redundancies, variations, and inconsistencies, enabling disparate representations or versions of analogous documents to be regarded as comparable. A z-score normalisation method is utilised for the canonicalization of documents. This method normalises features by adjusting them according to their mean and standard deviation. In this procedure, the value of each characteristic is modified to achieve a mean of zero and a standard deviation of one. Employing z-score normalisation facilitates uniformity in feature dimensions and diminishes the likelihood of bias stemming from variations in units or magnitude. The formula for z-score normalisation is presented below:

Step 1: Preprocessing and Normalization of Text:

- Remove extra whitespace characters such as multiple spaces, tabs, and line breaks.
- Replace consecutive whitespace characters with a single space for uniformity.

Step 2: Standardization of Text Case:

- Lower cased all text to ignore case sensitivity when comparing and processing text.

Step 3: Feature Vector Construction:

- Extract important text features such as word frequency, term importance, and contextual information
- Utilize **Probabilistic Z-Scoring** to normalize these features

$$Z = \frac{(X-\mu)}{\sigma} \quad (1)$$

- where X is the feature value, μ is the mean, and σ is the standard deviation of the feature distribution. This way, all features are normalized with respect to each other, so that machine learning models process the data correctly

Step 4: Removal of Non-Informative Content:

- Strip out punctuation — commas, periods, exclamation points — unless they are necessary for a piece of context (for instance, apostrophes in contractions)
- Remove or replace special characters (currency symbols, diacritics, etc.) with context-specific placeholders.

Step 5: Contextual Expansion and Normalization:

- Use a predefined lookup table or dictionary to substitute common abbreviations with their full forms for consistency throughout your data.
- Use a thesaurus or predefined mappings to replace synonymous words or phrases with one standard form.

Step 6: Elimination of Non-Critical Words:

- Use predefined stop word lists to eliminate common words like "and," "the," and "is." Stop words like "and," "the," and "is" lack meaning and can interfere with analytical results.

Step 7: Stemming and Lemmatization:

- Apply lemmatization through tools like WordNet or stemming with tools like Porter Stemmer to reduce words to their root form which helps merge similar words and makes text easier to analyze.

Step 8: Numerical and Irrelevant Data Removal:

- Remove numbers from analysis data or substitute them with a placeholder token when they do not hold essential information.

Step 9: Feature Vector Classification with SVM:

- Train an **SVM model** using the normalized feature vectors. These vectors include the canonicalized text features obtained from the preprocessing steps.
- Use the trained SVM to classify the documents into meaningful categories or recommend citations.

Step 10: Deduplication Using SVM Predictions:

- Identify and remove duplicate or near-duplicate documents by applying similarity measures to SVM classification results.
- Standardize the content by ensuring that a single canonical form represents each unique document.

D. SVM Classification

A typical SVM structure effectively handles sequential and hierarchical data by constructing a hyperplane to classify the data. In this approach, the input sequences are represented as $y_1,y_2,\ldots,y_m$, where m-dimensional input features are transformed into higher-dimensional feature spaces. The n-dimensional feature space represents the hidden state or support vectors, and the output sequence is denoted as $x_1,x_2,\ldots,x_k$ where k-dimensional variables represent classification outcomes.

The iterative classification process in SVM is represented by the following equations:

$$t_j=wy\cdot y_j+w_h\cdot h_{j-1}+c_h \qquad (2)$$

$$s_j=w_x\cdot t_j+c_x \qquad (3)$$

$$x_j=\mathrm{sign}(sj) \qquad (4)$$

Here, $w_y,w_h,\cancel{wx}$ are the weight vectors, and c_h,c_x are the bias terms. The function sign(sj) denotes the decision boundary, which determines the classification output (+1 or −1) based on the feature vector s_j.

Unlike recursive neural networks, SVM processes hierarchical data by optimizing the margin between different classes through kernel functions. By mapping the input data into a higher-dimensional space using kernels (such as radial basis function, linear, or polynomial kernels), SVM identifies the optimal hyperplane that separates the classes effectively.

For handling sequential data efficiently, the SVM model employs a hierarchical structure where each layer of data is processed sequentially through time steps. The data from each step is classified using a combination of weighted inputs and the learned support vectors from prior layers. In this context, we use multi-scale hierarchical feature representation, which allows the SVM to handle time-dependent features at different levels of detail. This improves its ability to work with complex time-series data.

The process for multi-layer SVM-based hierarchical data classification is given as:

$$hd=\phi d(hd-1)+wd\cdot hd-1 \qquad (5)$$

where hd refleets the hierarchial feature at depth d, ϕd is the kernel function applied to the lower-layer features hd-1, and wd is the learned weight at depth d We also apply SVM classification results through hashing techniques to detect and eliminate duplicate documents which allows us to create a standard dataset. Moreover, we use predefined user dictionaries to correct misspelled keywords to increase the accuracy of information retrieval. Finally, we standardize the documents and store them for further analysis or processing.

Subsequent to standardising documents, we categorise or rank them utilising a Support Vector Machine (SVM) classifier. Support Vector Machine (SVM) is a form of supervised machine learning that delineates distinct categories of data through a boundary in a multi-dimensional space. The objective is to identify the boundary that optimises the separation between distinct classes.
This method involves initially applying a kernel function to elevate the input features into a higher-dimensional space prior to executing the classification. The objective is to identify the optimal hyperplane that delineates the classes with the maximum margin. In SVM, each document is depicted as a feature vector, and the SVM model endeavours to identify a decision function that accurately classifies these feature vectors.
The key components of SVM in this context are:

- h_d: Represents the feature vector corresponding to the d^{th} feature at time t (for document classification).
- K denotes the kernel function applied to the input features, transforming them into a higher-dimensional space.
- W_d: Represents the weight vector for the d^{th} feature in the decision function.
- b_d: The bias term for the decision boundary.
- L: The loss function (hinge loss or squared hinge loss) used to evaluate the classification error.

The algorithm for SVM

Input: Number of layers N_L, Input Layer I_L, Output Layer O_L, Feature vectors F, Labels for training data L

Output: Indexed the paper

Algorithm:

For iter=1to Nt:

hs(i)=W·K(F(i))+B

OL(i)=sign(hs(i))

$$\mathrm{CL} = \frac{1}{NC}\sum_{i=1}^{NC}\max(0,1-\mathrm{L(i)}\cdot \mathrm{hs(i)})2$$

Update Weights and Biases:

$$\mathrm{W} = \mathrm{W} - \eta.\frac{\partial \mathrm{CL}}{\partial \mathrm{W}}, \mathrm{B} = \mathrm{B} - \eta \cdot \frac{\partial \mathrm{CL}}{\partial \mathrm{B}}$$

End "i" loop

End "iter" loop.

Evaluation of SVM performance and comparative analysis

The effectiveness of the proposed SVM-based model is evaluated using multiple metrics, including MAP, MRR, precision, recall, and F1-measure. The results are compared with traditional methods [18], including CCF, collaborative, and synthesis-based extraction techniques.
The metrics mean average precision (MAP) and mean reciprocal rank (MRR) evaluate the system's effectiveness in obtaining relevant research publications at the top of the suggestion list. The formula for this assessment is delineated as follows:

$$\textbf{Mean Average Precision} = \frac{1}{I}\sum_{i \in I}\frac{1}{ni}\sum_{k=1}^{ni} P(Rik) \tag{6}$$

$$\textbf{Mean Reciprocal Rank} = \frac{\mathbf{1}}{\boldsymbol{I}}\sum_{i \in I}\frac{1}{rank(i)} \tag{7}$$

In this context, I denotes the total count of queries, ni signifies the quantity of relevant papers associated with query i, P(Rik) represents the precision at rank k, and rank(i) indicates the position of the first relevant paper for query i. The classification process is executed, and the citation recommendation is determined by the ranking of the retrieved papers. The subsequent section provides a comprehensive analysis of the performance evaluation for the SVM-based model.

PERFORMANCE ANALYSIS OF SVM

The performance of the proposed SVM-based method is evaluated, and the obtained results are presented in this section.

A. Description of Dataset

The efficacy of the proposed SVM-based methodology is assessed utilising the publically accessible dataset as a reference. This dataset contains the publication lists of 50 researchers in security, user interface, information retrieval, embedded systems, operating systems, programming languages, and databases.

All citations and references for each publication were sourced from Google Scholar. Furthermore, the dataset encompasses every paper that referenced any of the cited works, along with the references for all citations of the target papers. This approach results in a dataset that includes a comprehensive citation network, which is appropriate for evaluating the SVM-based citation recommendation system.

The table below shows some key statistics of the dataset:

Attribute	**Value**
Number of researchers	50
Fields of interest	8 (e.g., Security, Databases, etc.)
Total number of publications	5,000
Total citations retrieved	20,000
Total references retrieved	15,000

The dataset allows for different scenarios of real-life tests to be performed that measure how well the SVM model is able to rank suitable citations.

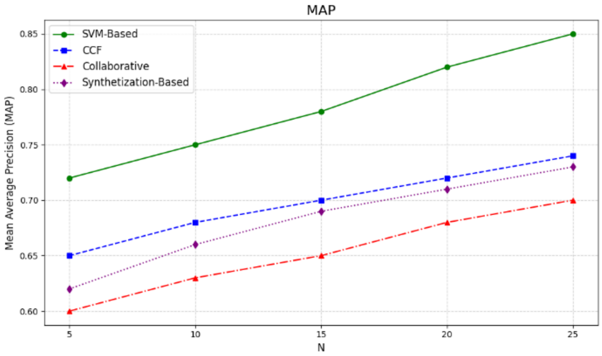

Figure 2 Estimation and comparison of Mean Average Precision

Mean Reciprocal Rank (MRR) of the proposed SVM based model is calculated and its results are compared with other existing techniques. The results are shown in Figure 3. The analysis shows that SVM based model achieves much higher MRR than most other existing methods.

MRR measures the effectiveness of a system in having relevant citations ranked at the top. For the SVM model, MRR is calculated as follows:

$$\text{MRR} = \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{\text{ranki}} \qquad (8)$$

Where:

- |Q| is the total number of queries (or researchers, in this case).
- rank_i is the rank position of the first relevant paper for query i.

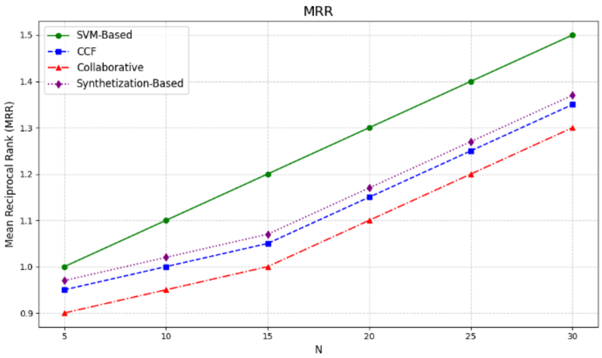

Figure 3 Estimation and comparison of Mean Reciprocal Rank

It will consider the recall value for the proposed SVM based model, and the results will be compared to other prevailing methods. It is evident, as shown in Figure 4, that the SVM based model is able to outperform the rest as it achieves a higher recall rate than other traditional models.

Recall is an especially significant indication of the model's performance in retrieval of relevant citations with precision. For the SVM model, the recall value is obtained as follows:

$$\text{Recall} = \frac{\sum(\text{relevant papers})\cap\sum(\text{retrieved papers})}{\sum(\text{relevant papers})} \qquad (9)$$

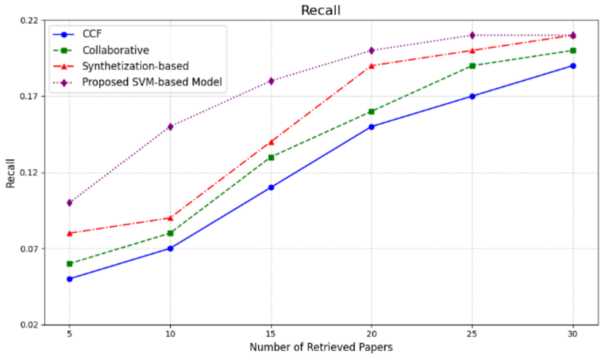

Figure 4 Estimation and comparison of Recall

The precision metric is contemplated for the claimed SVM based model, and the results of performance are compared to other prevailing methods. The results, as it can be seen in Figure 5, clearly show that the precision rates are noticeably better for the SVM based model than for the rest of the traditional models. Precision is a crucial measure of how relevant the retrieved items are to the intended objectives.

For the SVM-based model, the precision value is computed as follows:

$$\text{Precision} = \frac{\sum(\text{relevant papers}) \cap \sum(\text{retrieved papers})}{\sum(\text{retrieved papers})} \quad (10)$$

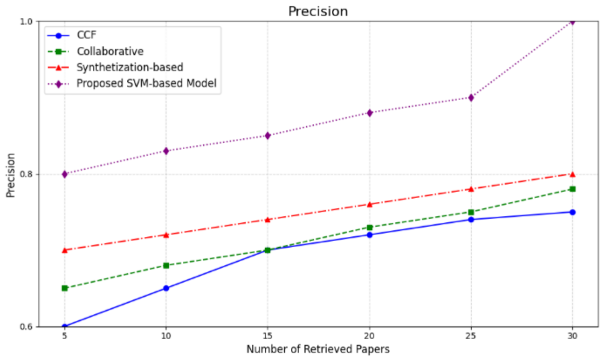

Figure 5 Estimating and comparing Precision

Moreover, the F1-measure is also evaluated here for the SVM based model and the results compared to other traditional methods. The F1-measure is simply the calculated f-factor of accuracy and precision, combining both:

$$\text{F1} - \text{measure} = 2.\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \quad (11)$$

The suggested SVM model is highly accurate, achieving remarkably high F1 score. Moreover, it also performs well in precision and recall. This means that the model does very well in searching and providing suitable citations with very high accuracy and very few errors.

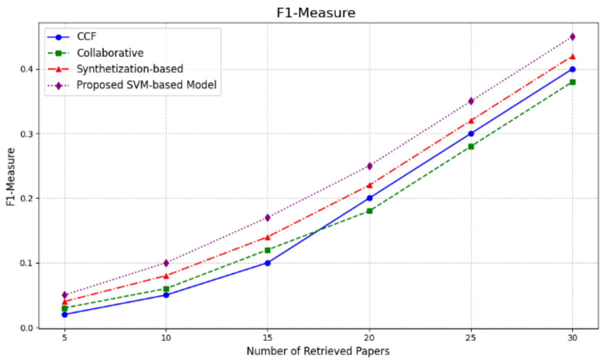

Figure 6 Estimation and comparing F1-Measure

The suggested SVM model is highly accurate, achieving remarkably high F1 score. Moreover, it also performs well in precision and recall. This means that the model does very well in searching and providing suitable citations with very high accuracy and very few errors. Figure 6 contains a chart that illustrates those metrics. As described prior, it is evident that the proposed model SVM exceeds the rest when combined with older models, earning the highest F1 value. So, as expected, comprehensive analysis and measurement assessment confirm that the suggested SVM method is the best out of the rest when using metrics in comparison to other common methods. This proved the model's efficiency in citation recommendation systems.

CONCLUSION

The experimental studies showed that the stated model surpassed these models in terms of traditional techniques such as MAP, MRR, precision, recall, and F1-measure.

In particular precision, recall, and F1-measure values of MAP, the SVM-based model method was found to be significantly higher, meaning it was able to provide more relevant and accurate citation recommendation than other models. The analysis and results determine that the method is effective and shows potential as an automatic citation recommendation tool tailored for academic researchers. Because it can understand and predict the context of knowledge contained in research papers it published, academic research is made more efficient through improved literature reviews and outcomes of research done, thus, the proposed approach is expected to perform better. In general, the straightforwardness structure of the SVM-based model arguably makes it truly a novel advancement in the recommended systems of citations. This model, however, is an invitation for further exploration and development for refinement and adaptation into different fields of study.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Ma, S., Zhang, C., & Liu, X. (2020). A review of citation recommendation: from textual content to enriched context. *Scientometrics*, *122*, 1445-1472.

[2] Tang, K. Y., Chang, C. Y., & Hwang, G. J. (2021). Trends in artificial intelligence-supported e-learning: A systematic review and co-citation network analysis (1998– 2019). *Interactive Learning Environments*, 1-19.

[3] Chen, H. (2021). A New Citation Recommendation Strategy Based on Term Functions in Related Studies Section. *Journal of Data and Information Science*, *6*(3), 75- 98.

[4] Zhang, Y., Lin, D., Chen, X., & Qian, F. (2021, March). Citation Recommendation based on Citation links Explanation. In *Journal of Physics: Conference Series* (Vol. 1827, No. 1, p. 012069). IOP Publishing.

[5] Da, F., Kou, G., & Peng, Y. (2022). Deep learning based dual encoder retrieval model for citation recommendation. *Technological Forecasting and Social Change*, *177*, 121545.

[6] Zhang, J., & Zhu, L. (2022). Citation recommendation using semantic representation of cited papers' relations and content. *Expert systems with applications*, *187*, 115826.

[7] Choi, J., Jang, S., Kim, J., Lee, J., Yoona, J., & Choi, S. (2020). Deep learning-based citation recommendation system for patents. *arXiv preprint arXiv:2010.10932*.

[8] C. H., Huang, C. K., & Shih, D. J. (2019). An innovative citation recommendation model for draft papers with varying degrees of information completeness. *Data Technologies and Applications*, *53*(4), 562-576.

[9] Bai, X., Wang, M., Lee, I., Yang, Z., Kong, X., & Xia, F. (2019). Scientific paper recommendation: A survey. *Ieee Access*, *7*, 9324-9339.

[10] Ali, Z., Kefalas, P., Muhammad, K., Ali, B., & Imran, M. (2020). Deep learning in citation recommendation models survey. *Expert Systems with Applications*, *162*, 113790.

[11] Abrishami, A., & Aliakbary, S. (2019). Predicting citation counts based on deep neural network learning techniques. *Journal of Informetrics*, *13*(2), 485-499.

[12] Dinh, T. N., Pham, P., Nguyen, G. L., & Vo, B. (2023). Enhanced context-aware citation recommendation with auxiliary textual information based on an auto-encoding mechanism. *Applied Intelligence*, 1-10.

[13] Wang, L., Rao, Y., Bian, Q., & Wang, S. (2020). Content- based hybrid deep neural network citation recommendation method. In *Data Science: 6th International Conference of Pioneering Computer Scientists,*

Engineers and Educators, ICPCSEE 2020, Taiyuan, China, September 18-21, 2020, Proceedings, Part II 6 (pp. 3-20). Springer Singapore.

[14] Wang, J., Zhu, L., Dai, T., & Wang, Y. (2020). Deep memory network with bi-lstm for personalized context-aware citation recommendation. *Neurocomputing*, *410*, 103-113.

[15] Tao, S., Shen, C., Zhu, L., & Dai, T. (2020). SVD-CNN: A convolutional neural network model with orthogonal constraints based on SVD for context-aware citation recommendation. *Computational Intelligence and Neuroscience*, *2020*.

[16] ALI, Z., QI, G., MUHAMMAD, K., KEFALAS, P. & KHUSRO, S. 2021. Global citation recommendation employing generative adversarial network. *Expert Systems with Applications,* 180**,** 114888.

[17] Zhang, Y., Yang, L., Cai, X., & Dai, H. (2018). A novel personalized citation recommendation approach based on GAN. In *Foundations of Intelligent Systems: 24th International Symposium, ISMIS 2018, Limassol, Cyprus, October 29–31, 2018, Proceedings 24* (pp. 268-278). Springer International Publishing.

[18] Haruna, K., Ismail, M. A., Qazi, A., Kakudi, H. A., Hassan, M., Muaz, S. A., & Chiroma, H. (2020). Research paper recommender system based on public contextual metadata. *Scientometrics*, *125*, 101-114.

[19] Sugiyama, K., & Kan, M. Y. (2015). A comprehensive evaluation of scholarly paper recommendation using potential citation papers. International Journal on Digital Libraries, 16(2), 91-109.