Information Sciences 559 (2021) 212–235

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

# CLAVER: An integrated framework of convolutional layer, bidirectional LSTM with attention mechanism based scholarly venue recommendation

Tribikram Pradhan [a,b,*], Prashant Kumar [c], Sukomal Pal [a]

[a] Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India
[b] Department of Information and Communication Technology, Manipal Institute of Technology (MIT), Manipal Academy of Higher Education (MAHE), Manipal, India
[c] Amazon Development Centre Private Limited, Chennai, India

A R T I C L E I N F O

*Article history:*
Received 26 August 2019
Received in revised form 17 October 2020
Accepted 6 December 2020
Available online 15 December 2020

*Keywords:*
Recommendation system
Convolution neural network
Long short-term memory (LSTM)
Attention mechanism
Deep learning

A B S T R A C T

Scholarly venue recommendation is an emerging field due to a rapid surge in the number of scholarly venues concomitant with exponential growth in interdisciplinary research and cross collaboration among researchers. Finding appropriate publication venues is confronted as one of the most challenging aspects in paper publication as a larger proportion of manuscripts face rejection due to a disjunction between the scope of the venue and the field of research pursued by the research article. We present CLAVER--an integrated framework of Convolutional Layer, bi-directional LSTM with an Attention mechanism-based scholarly VEnue Recommender system. The system is the first of its kind to integrate multiple deep learning-based concepts, that requires only the abstract and the title of a manuscript to identify academic venues. An extensive and exhaustive set of experiments conducted on the DBLP dataset certify that the postulated model CLAVER performs better than most of the modern techniques as entrenched by standard metrics such as stability, accuracy, MRR, average venue quality, precision@k, nDCG@k and diversity.



## 1. Introduction

With rampant increase in data at hand from all walks of life it has become progressively difficult to extract accurate and specific information that users look for [33]. Recommender systems (RS) use abundance of data analysis techniques to furnish recommendations to the end users based on a set of components [3]. Recommender systems in academia gives out recommendations for citations [15], papers [43], academic venues [31], and collaborators [5]. These systems have justified their worth as they generate meaningful recommendations to users with personalised information services.

Researchers generally favour to publish in those academic venues (journals, conferences, or workshops) where they find acknowledgement of top notch papers applicable to their domain of research [37]. However, with swift evolution and expansion of domains of multidisciplinary areas and cross-collaboration, there has been an active change within the range of journals, henceforth making the choice of an appropriate venue an even more burdensome task [2]. For example, DBLP[1] dataset, a

* Corresponding author at: Department of Computer Science and Engineering, Indian Institute of Technology (BHU), Varanasi, Uttar Pradesh, India.
*E-mail addresses:* tpradhan.rs.cse16@itbhu.ac.in (T. Pradhan), kumrkcu@amazon.com (P. Kumar), spal.cse@itbhu.ac.in (S. Pal).

1 http://dblp.uni-trier.de/db/.

https://doi.org/10.1016/j.ins.2020.12.024
0020-0255/

repository of scholarly papers and their relationship within that repository exceeds 9585 conferences in the area of computer science,[2] and the number of journals is more than 4152[3] [42].

It is hence a problem for all the researchers to remain updated with the particulars of a new venue and sometimes approaching an inappropriate venues which result in rejection, hampers and/or delay the publication. A rise in challenges appears in interdisciplinary research areas to research institutions and their libraries to grasp the dynamic information needs of their users [40]. An overall and complete solution is provided to venue recommendation in terms of either top conferences and journals [19]. Given the limited scope of work done in the area of academic venue recommendation, over the last decade Collaborative Filtering (CF) has been mostly favoured [2]. The techniques used the similarity among researchers using their profile and citation patterns [49]. In a cold start situation where no sufficient ratings are available in the researcher-venue matrix, CF approaches seems to be inefficient. The recommendations are also not suitable for new researcher who has no publication history. Researcher-venue matrix which is one of the central component of the techniques is mainly sparse in nature due to the large number of researchers involved with less number of papers and scholarly venues.

On the other hand Content Based Filtering (CBF) approaches take a variant glance on the recommendation. CBF or content-based models consider individual profile of researcher's to extract only relevant contextual features from their papers and associated venue [17]. Latent Dirichlet Allocation (LDA) based topic modeling is considered to extract hidden patterns in order to compute similarity among venues [24]. Due to vagueness in juxtaposition of text CBF might not be able to furnish useful recommendation and also bear with new venues and new researchers (*cold start problem*). What considerably lowers quality of recommendations made by the CBF approaches is the limited content analysis it does [43].

Recently, network-based approaches (NB) are more chosen over other conventional methods [50]. We need to construct a social graph within available researchers and their co-authors where the connection between the co-authors forms an edge of the graph [22]. A paper written by identical set of authors will get the same recommendation results, disregarding the actual content. In NB based venue recommender system stability could be a problem. Keeping all the mentioned classes of recommendation techniques in mind, the results furnished in all of them are very low grade for a author who does not have any publication records. Inspite of showing content similarity, the set of venues that are less notoriety among the co-authors of a target researcher are very less frequently recommended. The majority of the existing methods have an unjustifiable predisposition against unhackneyed newly spawning venues and fledgling researchers with limited publication [22]. Moreover the transition of an author to an equivalent interdisciplinary field of research is not efficiently handled by these techniques.

A large portion of the methodologies examined above can not capture the semantics and, thusly, neglect to suggest appropriate publication venues. Because of its capacity to extract the complex structure and profound semantics in high dimensional data, deep-learning approaches have been successful in many areas such as visual object recognition, sentiment analysis and text-classification [18]. Moreover, deep learning models can learn multiple-abstract representations syntactic and semantic information, since more abstract concepts can be developed with different processing layers.

Over the last few years, deep learning has been regarded as an effective method for recommendation task in the field of academia [11]. However, in most of those approaches, we believe, the word sequence information and semantic information have only been partially utilized. Where convolution layers fail to capture the long-term temporal dependencies and also the positional relation of features together with global features, Long memory (LSTM) model proves to be efficient in capturing significant local features. However, LSTM only traps long run information within the forward direction ignoring the reverse flow of data during a backward direction. Bi-LSTM overcomes this shortcoming. Also, it's not feasible to supply selective weights to some significant features and ignoring irrelevant details using conventional deep learning techniques. the eye mechanism takes care of this issue in deep learning. Although a number of the work mentioned above have tried one or the opposite deep learning scheme, there has not been a comprehensive approach to handle all the problems.

We propose a deep learning-based decision support system - CLAVER (Convolutional layer, LSTM with Attention mechanism-based scholarly VEnue Recommender system) based on an integrated framework of convolution operation, LSTM, Bi-LSTM, and attention mechanism. Suitable steps have been incorporated to overcome the individual demerits of CNN and LSTM. Some of the major benefactions of the research work are enlisted below:

- We take a convolution layer to pull out the features from the content information (Abstract and Title), which are then straightforward fed into the LSTM layers (one-dimensional LSTM and Bi-LSTM model). These are to capture both preceding and succeeding contextual information from the features outputted by the convolution layer. Finally, an attention layer is engaged to get the weights of relevant words that contribute to the final venue recommendation. CLAVER works independently with respect to the experience and efficacy of an authors' former research contributions. Venues with less number of publications are also getting an equal chance for inclusion in the final recommendation. Thus it overcomes *cold-start problem* of new researchers and new venues.

[2] http://dblp.uni-trier.de/db/conf/.
[3] http://dblp.uni-trier.de/db/journals/.

- To address data sparsity[4] issues, we have to transform the high dimensional and sparse embedding matrix into a lower-dimensional and dense set. Several convolution filters are then applied to extract n-gram features that then decrease the dimensionality of the input data. As a result of this effectiveness of extracting concealed contextual features that are relevant makes convolutional deep learning approach highly preferred.
- The proposed system CLAVER learns from both top tier conferences and journals from multiple publishers like IEEE, Elsevier, ACM, Springer and others. The convolution layer taking into account the embedding vectors captures a surface level representation of the input attributes. Next in LSTM and Bi-LSTM another level of abstraction takes place that capture concepts from context. Ultimately, the distribution of weights is reinforced to more information-carrying terms from the input text. The entire scheme thus apprehends the diversity and encapsulates in abstraction leading to diversity[5] and stability[6] in recommendations.
- Based on the real-world DBLP dataset, we have conducted extensive experiments and reveal that our model significantly outperforms baseline methods by a large margin in terms of accuracy, MRR, nDCG@k, precision@k, average venue-quality (ave-quality), stability and diversity.

The remainder of this paper is structured as follows. In Section 2, we briefly summarize related work. We provide a more elaborate problem description in Section 3. Section 4, gives a detailed description of our attention-based model. Experiments and detailed statistics of dataset are presented in Section 5. Experimental results are reported in Section 6. Our work has been summarized and more insightful discussions have been reported in Section 7. Section 8 summarizes our work and the future direction.

## 2. Related work

A comprehensive review on recommender system was written and presented by Adomavicius et al. [1]. They highlighted predominantly three sorts of recommender frameworks dependent on their working standards. In addition, we have also included network-based and deep learning-based recommendations [50]. Some of them are elaborated in this section.

### 2.1. Recommendation based on collaborative filtering (CF) method

Yang et al. [46] introduced a model to answer the question, whether the set of papers published in one specific venue share common characteristics in writing styles or not exploiting lexical, syntactic and structural features. Yang et al. [47] proposed a novel framework to recommend suitable publication venues incorporating content similarity in terms of both topic and writing-style information. In order to recommend such venues, they adopted a strategy of differentiating the contributions of different types of neighboring papers. Yang et al. [48] exlored a model for various academic recommendations including scholarly venue recommendation for author-paper pairs considering the mutual effects of the latent correlation between relations. Hyunh et al. [16] introduced a novel framework to represent collaborative knowledge exploiting both graph theory and probability theory for recommendation in academia where three different components such as collaborative network, measures and rules are incorporated.

Yu et al. [49] presented a collaborative filtering based prediction model to extract continuity features from users browsing content to identify collaborative users based on their adjusted concept weights. Finally, concept wise information quantity is computed in terms of semantic relations between concepts to improve the efficiency of prediction of an academic recommendation. Liang et al. [21] explored a probabilistic model where user exposure is modeled as a latent variable and the model infers its value from data. Alhoori et al. [2] introduced a novel research analytical framework for academic venues recommendation considering a new weighting strategy for implicit ratings of venues that can provide better recommendations even for a new researcher.

### 2.2. Recommendation based on content-based filtering (CBF) method

Medvet et al. [24] proposed a topic matching procedure based scholarly venue recommendation system for scientific paper submission that only requires the title and abstract. Their approach can generate meaningful recommendations even at the early stages of the authoring process. Similarly, Wang et al. [42] proposed a contextual similarity based recommender system to recommend suitable journals in the domain of computer science which is designed based on the techniques of chi-square and softmax regression.

Errami et al. [10] introduced an approach eTBLAST in order to suggest journals considering only abstract as an input. Schurmie et al. [34] exlored the Journal/ Author Name Estimator (Jane) [7] that considered the biomedical database MEDLINE to suggest journals incorporating abstract as an input. This approach utilizes a Lucene tokenizer and Lucene MoreLikeThisAlgorithm to select a set of related literature in MEDLINE. Both weighted k-nearest neighbors and Lucene similarity scores are taken

[4] "Sparsity denotes average distance among pairs of related papers".
[5] "Diversity means how many different venues are recommended".
[6] "Stability denotes resiliency to change in ranked recommendations with the introduction of new papers".
[7] http://jane.biosemantics.org.

into consideration in order to rank articles. Choi et al. [7] designed an effective Evidence-Based Medicine (EBM) ranking algorithm that can automatically extract papers which are more relevant to clinical questions and are based on valid evidence. They mainly combined relevance and quality ranking using various fusion methods. Serrano et al. [35] proposed an Evidence-Based Medicine (EBM) ranking algorithm where both aspects related to the relevance and the quality of said documents are used in order to suggest the most relevant documents for clinicians.

Recently, a couple of online services began offering help for recommending journals based on keyword, title, and abstract matching. These services include Elsevier Journal Finder [8] [17], Springer Journal Suggester,[9] Edanz Journal Selector[10] and EndNote Manuscript Matcher[11] etc. Elsevier Journal Finder and Springer Journal Suggester considered journals only from their own publishers.

### 2.3. Recommendation based on network-based (NB) method

Klamma et al. [19] presented a community based recommendation model that can supports researchers in events finding. To achieve this objective, they employed event participation history to generate meaningful academic events to researchers. Silva et al. [36] presented a prototype system on the platform of ScholarMate research social network that holistically considers relevance, connectivity, and productivity parameters for generating effective journals. Pham et al. [26] presented a social network based model employing clustering techniques between researchers to select their neighborhoods and then they applied traditional CF algorithms for academic venue recommendation. Instead of using ratings data, Pham et al. [27] employed social relationship among researchers in order to identify their similar researchers that is used as a background data for academic venue recommendation and trust-based recommendation.

Chen et al. [6] implemented a prototype AVER to suggest academic venues to a target researcher. Their proposal mainly utlized a Random Walk with Restart (RWR) model on the co-publication network considering both author-author and author-venue relations. Later, Yu et al. [50] presented a random walk with restart model PAVE that employed scholarly features such as co-publication frequency, relation weight and researchers' academic level for scholarly venue recommendation.

Luong et al. [23] presented a social network based venue recommendation model that takes advantage of information analysed from an academic social network of researchers linked by their co-authorship relationship. Luong et al. [22] introduced a social network based venue recommendation model considering social network of researcher, linked by their publication history to recommend academic publication. Porcel et al. [28] introduced a hybrid recommender system that can generate personalized recommendations to improve teaching and learning processes incorporating both ontology and fuzzy linguistic modeling.

### 2.4. Recommendation based on hybrid method (HM)

Pradhan et al. [30] presented a novel unified architecture integrating social network analysis, content similarity, citation and co-citation analysis for academic venue recommendation. Their proposal also addresses cold start issues, diversity, and stability problems. Wang et al. [43] proposed a hybrid article recommendation approach incorporating CBF and CF approaches. Boukhris et al. [4] introduced a scholarly venue recommender system for the domain of computer science utilizing citation relationships between papers, co-authors, co-citers and co-affiliated researchers. Minkov et al. [25] presented a collaborative ranking predictions based hybrid recommendation model that recommends future events to users based on users past selections and feature descriptions of the events. Tang et al. [38] presented a cross-domain topic modeling approach based collaborator recommendation model that can rank and recommend potential cross-domain collaborators. They demonstrated the effectiveness and the efficiency of their model by experimenting in a coauthor network. Pradhan et al. [31] introduced an academic venue recommendation model which is an integrated framework employing paper-paper peer network model and venue-venue peer network model. Their proposal also addresses cold start, diversity and stability issues to some great extent.

### 2.5. Recommendation based on deep learning (DL) method

Article Recommendation based on dynamic attention deep learning-based model has been proposed by Wang et al. [44]. Models are trained to capture the underlying selection criteria for article selection. Ebesu et al. [9] proposed a novel probabilistic neural network based citation recommendation system which can jointly learn the semantic representations of citation contexts and cited papers. Hassan et al. [13] present a personalized research paper recommendation system. It facilitates this task by recommending papers based on users' explicit and implicit feedback. Yang et al. [45] introduced a context-aware citation recommendation that measures the overall relevance based on the learned distributed representation of citation contexts and the scientific papers.

[8] http://journalfinder.elsevier.com.
[9] http://journalfinder.com.
[10] https://www.edanzediting.com/journal-selector.
[11] http://endnote.com/product-details/manuscript-matcher.

Feng et al. [11] introduced a deep convolutional neural network based journal recommender system incorporating pre-trained word2vec to recommend suitable journals for biomedical literature based on the paper's abstract. Wang et al. [41] presented a recommender named CAMO that employs a multi-layer content encoder for simultaneously capturing the semantic information of multitopic and word order. Pradhan et al. [29] introduced a deep learning based journal recommendation model, employing bi-directional LSTM and hierarchical attention network for generating effective recommendations. Gao et al. [12] presented a comprehensive survey of the GNN-based knowledge-aware deep recommender system. They discussed about the graph embedding module, and how they address practical recommendation issues such as scalability, cold-start and so on. Pradhan et al. [32] presented a collaborator recommendation model employing both deep learning and biased random walk for potential collaborators that share similar research interests at the peer level.

## 3. Problem definition

**Definition 1. Venue Recommendation**. "Let us consider a set of papers which are unique $\{p_1, p_2, \ldots, p_n\}$ where $p_i$ is published in a venue $v_j \in V = \{v_1, v_2, \ldots, v_m\}, (m \leqslant n), V$ here is a collection of publication venues which is predefined [29]. It should be noted that, multiple paper may be published at any particular venue, and thus $m \leqslant n$"."When an input paper $p_0$ (seed paper) is given, the primary aim of the venue recommendation module is to supply an ordered list of suitable and effective publication venues $(v_{0_1}, v_{0_2}, \cdots, v_{0_k})$, $v_{0_1}$ being the most relevant among others and $v_{0_k}$ is k-th most relevant recommended venue. Similarly, the list follows a trend and recommends the venues in the decreasing order of relevance where each $v_{0_i} \in V$"."This can be visualized as a multiclass classification problem. Initially, it is required to identify unique publication venues and treat them as individual classes. Now we have, $C = \{c_1, c_2, \ldots, c_m\}$ (m⩽n) as the number of labels possible. A softmax function can then be used to estimate the probability of each of the individual classes $P(\mathrm{c_i}|\mathrm{p_0})$ $(i = \{1, 2, \ldots, m\})$, the sum of probabilities of all the individual classes here being 1"."For a given seed paper $p_0$, the venue recommendation task can be modelled as recommending an ordered list of classes $(c_{0_1}, c_{0_2}, \cdots, c_{0_k})$, such that $c_{0_1}$ is the most relevant and $c_{0_k}$ is k-th most relevant class (or venue) in the decreasing order of estimated probabilities $P(\mathrm{c_i}|\mathrm{p_0})$". Venue recommendations are provided if the title, and abstract of a seed paper are given to the system as input (Fig. 1).

## 4. Functional architecture of CLAVER

In the underlying section, an overview of CLAVER's functional architecture is expounded in elaborate details. (Fig. 1).

### 4.1. Overview

To address the task of recommending suitable venues, we attempt to propose CLAVER: a novel decision support system incorporating Convolutional layer, LSTM layer, along with Attention mechanism for scholarly VEnue Recommendation. The postulated solution comprises of multiple layers, each designed to deliver a particular task. CLAVER comprises of majorly five layers, namely embedding layer, convolution layer, LSTM layer, attention layer, and dense layer. These five primary layers can be described as follows:

(i) *Embedding Layer (Layer 1)*: The functionality of the Embedding layer is to shape, arrange, and organize the dataset rendering it appropriate for subsequent processing. This layer is synonymously known as the word representation layer as it is useful in getting a superficial representation of the abstract and the title for further use. The primary aim of this layer is to obtain a higher dimensionality representation of the input vector. This embedding layer was directly used from Keras, and the initial raw data was transformed into a higher dimensionality feature vector and fed into the subsequent convolution layer for further processing.
(ii) *Convolution Layer (Layer 2)*: The prime functionality of the Convolution layer is to extract n-gram features from both abstract and title texts. The CNN primarily consists of a regularised multilayer perceptron, which convolves the input using convolution layers, nonlinearity, and pooling layer (Subsampling). These layers, when stacked together, form a deep convolution neural network (CNN) and can be used to extract essential local features from the input feature vector. An additional operation, usually ReLU (Rectified Linear Unit), is used after every convolution operation. And pooling reduces the dimensionality of the input feature vector, which helps in decreasing the overall processing time.
(iii) *LSTM Layer (Layer 3)*: The LSTM layer can capture long term dependencies and also firmly tackle vanishing gradient problems. It uses a number gates in order to control the flow of data in the recurrent neural unit. Gates are nothing but layers which, when applied multiplicatively, can either keep the value from the gated layer or zero value depending on the 1 or 0 values of the gates, respectively. The three types of gates it uses comprises of output gate, input gate and a forget gate.

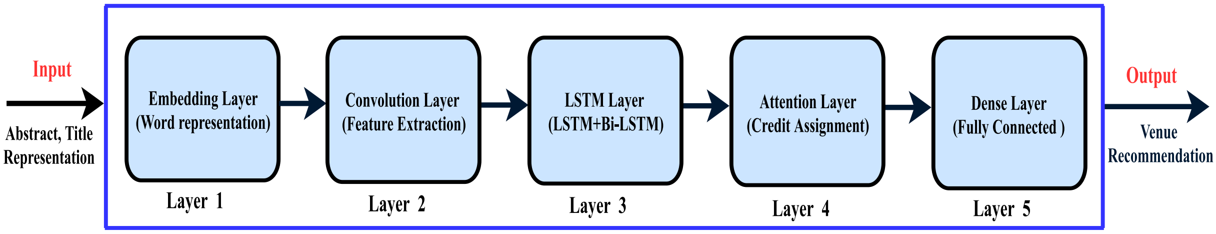


**Fig. 1.** Overview of proposed system CLAVER.

The two distinct model this layer comprises of are:

(i) *One-dimensional LSTM model*
(ii) *Bi-LSTM model*

The one-dimensional LSTM is used to extract features and information in the forward direction whereas BiLSTM accesses contextual information and feature in both forward and backward direction by using a combination of both a forward and a similar backward moving layer.

(iv) *Attention Layer (Layer 4)*: An attention mechanism is primarily devised and conciliated into CLAVER as it has an acquired capability to enhance the semantic feature extraction process of the model by identifying key-words and reducing the weight of non-keyword in the final prediction of an appropriate venue.
(v) *Dense Layer (Layer 5)*: The output of the preceding CNN, one-dimensional LSTM, and Bi-LSTM layers that represent the high-level features of abstract and title forms the input. Subsequently, the dense layer obtains a new high-level representation of input words from the previous layers by using a hidden layer with some dropout rate to map the input to the dimensionality of a desired output.

### *4.2. Embedding layer (Layer 1)*

The functionality of the Embedding layer is to shape, arrange, and organize the dataset rendering it appropriate for subsequent processing (Data preprocessing) as well as expanding the dimensionality of any input feature vector (Embedding). We acquired the data from DataBase systems and Logic Programming (DBLP), which originally contained 2,236,968 research papers from 4565 Journals and Conferences after all the missing value rows were removed. The Dataset in particular contains ‘abstract’, ‘authors’, ‘id’, ‘references’, ‘title’, ‘venue’, and ‘year’ columns. The columns used for our experiments were ‘abstract’, ‘title’, and ‘venue’. All the venues having less than 5 papers were removed from the Dataset in the initial filtering process, which finally left with 2,234,771 papers and 3216 venues to work on.

All the sentences from the title and the abstract were concatenated and then cleaned by converting all the words in the input text to lower case characters, then the stop-words and punctuations were removed, and finally, the verbs in the text was lemmatized. After cleaning the text, it was tokenized using Keras-Tokenizer with a maximum word limit of 5000. Padding was done to reach a maximum length of 300. Output (venue) was initially processed by encoding the name of the venues to labels by sklearn-LabelEncoder and then converting these labels to a one-hot encoding vector by sklearn-OneHotEncoder.

Some of the serious problems, such as losing word order and bloating up the dimensionality of the output space, encountered in traditional techniques such as one-hot encoding, were overcome by adapting a distributed word embedding technique. Our model’s input are the sentences of the abstract in $A_{text}$ and the title in $T_{text}$. The texts $A_{text}$ contains $n$ sentences, each of which is composed of multiple words. Similarly, the texts $T_{text}$ consists of multiple words. The training of the Dataset was done by using Keras word embedding method, which has the ability to convert each word $w_i \in R^d$ into a fixed-size vector, wherein $d$ denotes the dimensionality of the input word vector.

In our model, the size of the dimension $d$ was considered to be 300. Due to difference in the size of abstract and title for different papers, we set $L$ to be the maximum number of words appearing in both $A_{text}$ and $T_{text}$ among the input papers. Let $S$ denote the original representation of length $L$ (padded where necessary) appearing in the texts $A_{text}$, and the texts of $T_{text}$ of a given paper and can be represented as

$$S_{1:L} = w_1 \oplus w_2 \oplus \cdots \oplus w_L, S \in R^{L*d}, \quad (1)$$

where $\oplus$ is the concatenation operator, L being the maximum number of words, which is a scalar, and $S_{1:i+j}$ refers to the vector of the concatenation of the words $w_1, w_2 \cdots w_{i+j}$.

If the words are not in the pre-trained vocabulary, the corresponding word vectors are initialized to zero. Finally, a matrix S with a dimensionality of $L * d$, which represents the abstract, is obtained. Thus the representation of an abstract matrix S is used to represent that the text (words) appears in both $A_{text}$ and $T_{text}$, respectively. While LSTM comes with the capability to handle sequences of variable lengths; It certainly fails in utilizing the contextual information from the future tokens and is also a deficit in extracting the local contextual information. This problem can be solved by using a convolutional layer before LSTM.

### 4.3. Convolution layer (Layer 2)

The convolution layer is applied on the word embedding vectors obtained from the previous embedding layer to get a high-level representation of the input textual data in $S$. In the proposed model, the convolutional layer helps in capturing the sequence information and it also reduces the dimensionality of the given input data. We use CNN in this model to extract more semantic and abstract features.

There are various operations that the CNN model consists of, they are convolution operation, nonlinearity, and pooling, which come together to form a deep convolution neural network (CNN).

(i) *Convolution Operation*: The aim of this layer is to obtain high-level representations of the input textual data. Convolutions operation returns a feature map by applying a set of convolution filters.
(ii) *Pooling Operation*: This layer also known as the feature extraction layer. It mainly applies pooling operation to extracts word-wise essential and relevant features. Pooling operation primarily reduces the dimensions of the feature vector and, in turn, decreases the processing time.

#### 4.3.1. Convolution operation

The convolutional layer when coupled with max-pooling operation helps in extracting some rich feature vector from each of the convolved word window of length $l$ (i.e., 4 for the first convolutional layer) over the text, performing a convolution operation within each sliding window and the corresponding output of the $k$-th sliding window can be computed as

$$f_k = ReLU(W_c.W_{k-l+1:k} + b_c) \tag{2}$$

Where ReLU denotes a non-linear activation function, $W_{k-l+1:k}$ is the concatenation of $l$ word embeddings within the k-th window in word sequences in $S$ (text appear in both $A_{text}$ and $T_{text}$), $W_c$ is the convolution matrix and $b_c \in R$ is the bias.

In the proposed model, we decided to adopt ReLU (Rectified Linear Unit) as a nonlinear activation function as it supplies a significant reduction in the number of iterations required for convergence and also improves the learning dynamics of the deep neural networks.

Multiple filters were used and for the $q$th filter, it is applied to each possible window of words in $S$ $\{W_{1:l}, W_{2:l+1}, \cdots, W_{L-l+1:L}\}$ to produce a feature map

$$f_q = [f_{q1}, f_{q2}, \cdots, f_{q,L-l+1}] \tag{3}$$

with $f_q \in R^{L-l+1}$.

We can use $m$ number of filters which can extract multiple features maps $f_1, f_2, \cdots f_m$. We get a new feature representations $F \in R^{L-l+1 \, *m}$ as the column wise concatination of feature maps F=[$f_1, f_2, \cdots f_m$]. The i-th row $f^{(i)}$ of $F$ is the new feature representation generated at position i.

So, the result of the first convolution operation on $S$ will be

$$f^{(1)} = \left[f_1^{(1)}, f_2^{(1)}, \cdots, f_m^{(1)}\right] \tag{4}$$

As evident from Fig. 2, after the window size of 4 was applied in the first convolutional layer, 256 number of feature vectors (convolutional kernel) were obtained each having a dimension 297 $*$ 1. Consequently, after the second convolution, the dimensionality of the output vector will be 144 $*$ 1 having 128.

#### 4.3.2. Pooling operation (Feature Maps)

The primary purpose of the pooling operation is to reduce the spatial size of the feature vector and extract the key-features. The max-pooling operation is applied to obtain the most salient feature in every two-unit window for each $f_q^{(1)}$ of the input feature vector of $S$. The max-pooling operation gives $j_q^1$ result.

$$j_q^{(1)} = \left[j_{q1}^{(1)}, j_{q2}^{(1)}, \cdots, j_{q,(L-l+1)/2}^{(1)}\right] \tag{5}$$

where q is the qth filter of the convolution operation.

$$j_{qi}^{(1)} = max\left\{f_{q,2i-1}^{(1)}, f_{q,2i}^{(1)}\right\} \tag{6}$$

The convolutional layer proves to be effective in reducing the dimensionality of the input feature vector. As shown in Fig. 2, in the convolutional layer, 256 filters with window size 4 is used to capture the syntactic and semantic features of the input text. The dimensionality of input data is considered to be 300 $*$ 300, and the dimensionality of the first layer output data is 148 $*$ 1. Similarly, the dimensionality of the second layer is 72 $*$ 1, and hence the reduction in the dimensions of the input can be easily visualized.

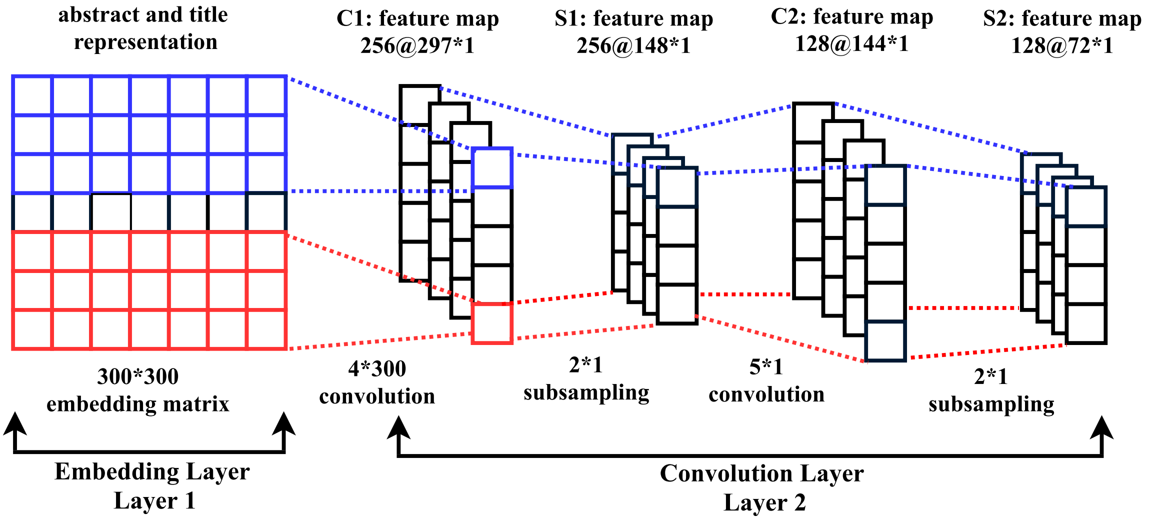


**Fig. 2.** Architecture of embedding layer and convolution layer.

### 4.4. LSTM layer (Layer 3)

We need to obtain the sequential information and determine long sequential patterns in the input feature vector, in order to predict a relevant journal for any given seed paper. The sequences of features obtained from the convolution operation in the preceding layer is limportant deficit in the supply of any sequential information. In the LSTM layer, one-dimensional LSTM is initially used for sequential modeling and sequential information extraction supplyingin the forward direction, exactly as illustrated in Fig. 3. As Bi-LSTM captures both the following and the previous sequential information, we use Bi-LSTM to get a finer feature vector.

#### 4.4.1. One-dimensional LSTM model

The traditional RNNs that works on the principle of backpropagation based on gradient descent posethe problem of vanishing gradient as well as exploding gradients. In order to resolve the problems of vanishing gradients and exploding gradients, Long short-term memory (LSTM) [14] can be used. In the recurrent hidden layer, LSTM consists of individual memory blocks units. In these memory blocks, which involve memory cells with self-connections, the temporal state of the network is stored. Additionally, there are special multiplicative units called gates that regulate the flow of information, as illustraded in Fig. 4.

In terms of three simple gates, including and relegation to the functionality of the input function (input gate), the output function (output gate) and the forget function (forget gate), LSTM can be formalised. These gates primarily make criticaltion, as exp ell. which short or long-term data is required and which inforamtion feature should be dropped.

(i) *The forget gate*: This is governed by $f_t$ in the figure, makes the decision to choose which part of the long term data should be dropped and which part of the long term data is important and is to be retained and passed on to the next term state $c_t$.
(ii) *The input gate*: This being controlled by $i_t$ in the figure, conditionally selects which part of the input $g_t$ and $h_{t-1}$ i.e. the former short term state should be filtered out while the others are relayed over to the next long term state i.e. $c_t$. This ensures that some part of the current state is relayed over to the and reflected in the long-term memory.
(iii) *The output gate*: This being constrained by $o_t$, establishes the outputs ($h_t$ and $y_t$) and also determines which portion of the long-term state $c_{t-1}$ suffices to be perused at any particular time step.

Mathematically each of the LSTM cell can be formulated as the following:

$$\mathbf{i}_{(t)} = \sigma\left(W_{Xi}^T \cdot \boldsymbol{X}_{(t)} + \boldsymbol{W}_{hi}^T \cdot \boldsymbol{h}_{t-1} + \boldsymbol{b}_i\right) \quad (7)$$

$$\boldsymbol{f}_{(t)} = \sigma\left(W_{xf}^T \cdot \mathbf{X}_{(t)} + \boldsymbol{W}_{hf}^T \cdot \boldsymbol{h}_{t-1} + \boldsymbol{b}_f\right) \quad (8)$$

$$\boldsymbol{o}_{(t)} = \sigma\left(\boldsymbol{W}_{xo}^T \cdot \boldsymbol{X}_{(t)} + \boldsymbol{W}_{ho}^T \cdot \boldsymbol{h}_{t-1} + \boldsymbol{b}_o\right) \quad (9)$$

$$\mathbf{g}_{(t)} = \tanh\left(\boldsymbol{W}_{xg}^T \cdot \boldsymbol{X}_{(t)} + \boldsymbol{W}_{hg}^T \cdot \boldsymbol{h}_{t-1} + \boldsymbol{b}_g\right) \quad (10)$$

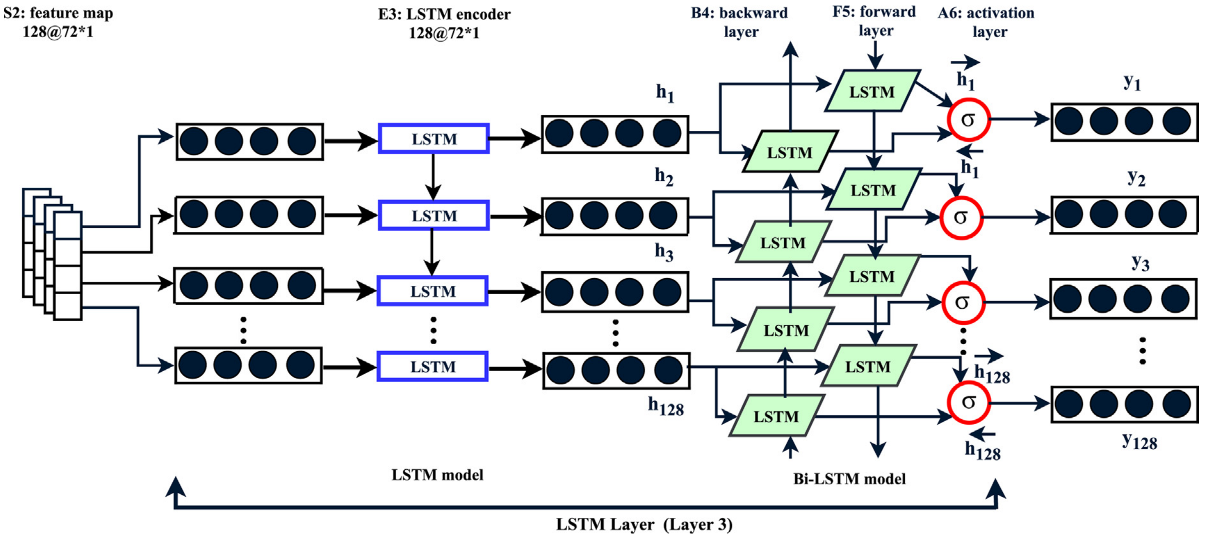


**Fig. 3.** Architecture of LSTM layer.

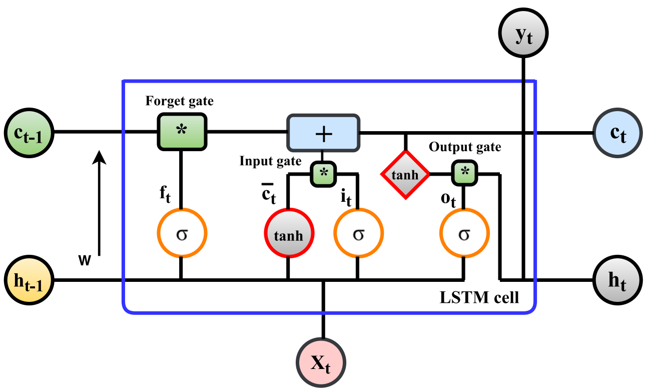


**Fig. 4.** Single LSTM cell.

$$\boldsymbol{c}_{(t)} = \boldsymbol{f}_{(t)} \otimes \boldsymbol{c}_{(t-1)} + \boldsymbol{i}_{(t)} \otimes \boldsymbol{g}_{(t)} \tag{11}$$

$$\boldsymbol{y}_{(t)} = \boldsymbol{h}_{(t)} = \boldsymbol{o}_{(t)} \otimes \tanh\left(\boldsymbol{c}_{(t)}\right) \tag{12}$$

where, $W_{xi}, W_{xf}, W_{xo}, W_{xg}$ are weight matrices of each of the four layers for their connections to the input vector $X_t$. $W_{hi}, W_{hf}, W_{ho}, W_{hg}$ are weight matrices of each of the four layers for their connections to the previous short term state $h_{t-1}$. $b_i, b_f, b_o, b_g$ are the bias terms for each four layers. LSTM has the ability to identify crucial and important input features, store it in long term state and finally extract all the necessary and important feature details.

### 4.4.2. Bi-LSTM model

We apply Bi-LSTM on the output feature vector $h_1, h_2, \cdots h_m$ obtained from the preceding LSTM operation, to perform even more comprehensive and thorough feature extraction. By concatenating the results obtained from a forward and a backward LSTM, the Bi-LSTM model generates an output. It contains the forward LSTM (represented by $\overrightarrow{LSTM}$) which peruses the feature successions from $h_1$ to $h_{128}$ and the backward LSTM (denoted by $\overleftarrow{LSTM}$) which peruses the feature successions from $h_{128}$ to $h_1$. Mathematically, the Bi-LSTM are defined as follows:

$$\overrightarrow{h_f} = \overrightarrow{LSTM}(h_i), \quad i \in [1, 128] \tag{13}$$

$$\overleftarrow{h_b} = \overleftarrow{LSTM}(h_i), \quad i \in [128, 1] \tag{14}$$

$$y_i = \overrightarrow{h_f} || \overleftarrow{h_b} \tag{15}$$

The Bi-LSTM gives a sequence of output vectors $y_1, \cdots, y_m$ for the corresponding output vectors $h_1, h_2, \cdots h_m$ of the preceding one dimensional LSTM model.

### 4.5. Attention layer (Layer 4)

We indeed observe that the relative contribution of various words in predicting an effective venue is different. Where some words prove to be crucial in the final prediction, some words do not hold much weightage. This predicament spawns the requirement of a mechanism to assign relative weights to different words. An attentive pooling is used to determine and formalise the representation of the input sequence by approximating their relative contribution, hence leading to the formulation of a surface level descriptor vector $J$ of the whole word sequence as depicted in Fig. 5.

Given a sequence of vectors $y_1, y_2, \cdots y_{128}$, we applied attention layer, in order to enhance the semantic feature extraction process of the model by identifying key-words and reducing the weight of non-keyword in the final prediction, to the context $J$ of the entire word sequence as characterized in Eqs. 16, 17, and 18. The attentive pooling layer can be mathematically formulated as following:

$$z_i = tanh(W_c * y_i + b_c) \tag{16}$$

$$\alpha_i = \frac{z_i^T u_w}{\sum_k exp(z_k^T u_w)} \tag{17}$$

$$J = \sum_i \alpha_i * z_i \tag{18}$$

where $y_i$ is the surface level feature representation (Bi-LSTM) for the $i$-th output vector $h_i$ (one dimensional LSTM) which is fed to a one-layer multi-layer perceptron (MLP) to obtain $z_i$ as hidden representation of $y_i$.

$W_c$ and $b_c$ represents the weights and the bias representation vectors for the Multi Layer Perceptron. $u_w$ represents a randomly seeded initial vector, these are basically the learnable parameters which are corrected and assimilated during the training phase, $\alpha_i$ are the relevant weight applied during the attention phase to the feature representation$y_i$. The relevance of any word in the feature vector is subject to a similarity score between the context vector $u_w$ and $z_i$. This perpetuates into a normalised weight $\alpha_i$ , delineated as the importance weight above, through a softmax function. Herein the feature representation vector $J$ is posited as a weighted sum over all the surface-level input feature representation vectors $y_1, y_2, \cdots y_{128}$.

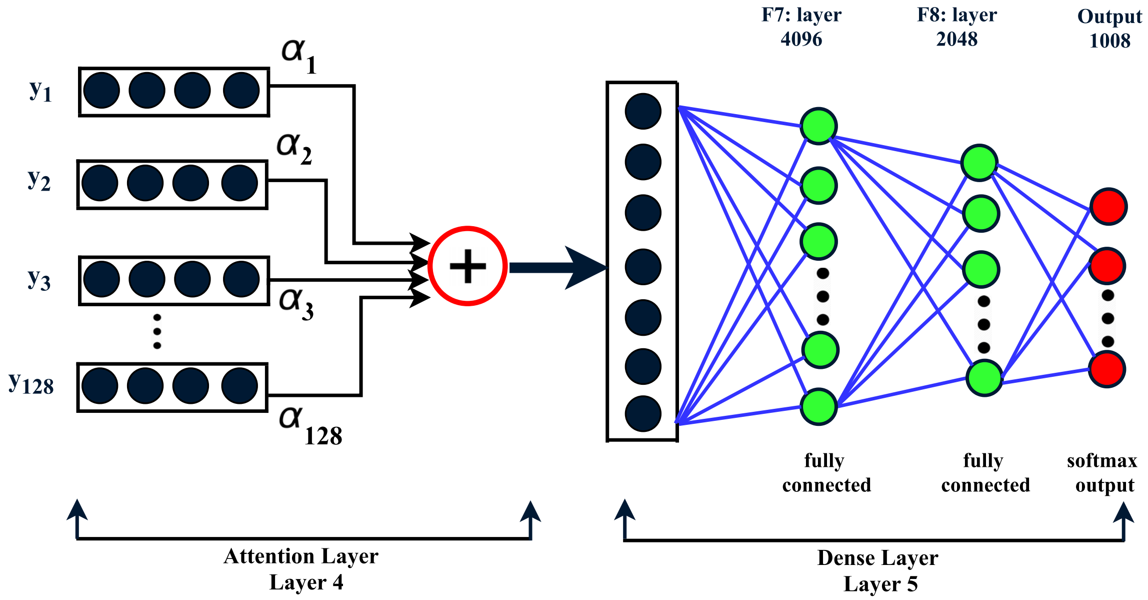


**Fig. 5.** Architecture of attention layer and dense layer.

### 4.6. Dense layer (Layer 5)

The Dense layer is used to create a high-level representation from the sequence of word vector $J$ from the text representation $J$ obtained from the preceding Attention mechanism. In order to map the text representation received from the previous layer to a desired dimensionality output vector, the dense layer uses a hidden layer with a certain drop out rate. The attention layer is followed by two dense layers, each of which has different sizes of neurons.

ReLu is used as the activation function of the first dense layer. We applied dropout between the two dense layers in order to avoid a potential overfitting issue. An array of dropout rates were experimented against to epiphanies an efficacious configuration. Using softmax as the activation function, the feature representation yields from the former layer are supplied to the second dense layer.

Given the training sample $R_1$ and $R_2$, where $T$ is the number of possible labels and the estimated probabilities $S_j \in [0,1]$ for each label j $\in$ [1, 2, $\cdots$, T], the softmax is defined as:

$$S_j = \frac{e_j^z}{\sum_{k=1}^{T} e_k^z} \tag{19}$$

Note that in the training dataset, we only have the gold-standard original venue of each source papers. In an attempt to minimise the error in prediction between the predicted venues and the gold-standard original venues we have utilized the categorical cross-entropy loss function:

$$L(\theta) = \sum_{j=1}^{T} Y_j log S_j \tag{20}$$

Here Y represents the gold-standard results. A trivial one-hot encoding method was employed, which represented all except the efficacious output as 0 and the correct output is labeled as 1. This encoding vector, of size L, is trained using Adam with mini-batch to correct and assimilate the trainable parameters $\theta$.

## 5. Experiments

The sequence of the experimentation section goes as follows; Firstly we outline the dataset followed by presenting an evaluation strategy and evaluation metrics employed for gauging the effectiveness of the proposed solution. We then perceptually expound the experimental settings used, parameter tuning employed along with the baseline methods in subsequent sub-sections. Majority of the experimentation were run on google colab XLA_CPU with a memory limit of 17 GB. All the relevant code was on Python programming language. The implementation of model was is done using Keras and other related libraries on a Tesla P100-PCIE modern GPU with 16 GB Stacked Memory Capacity.

### 5.1. Dataset description

The experimentation was done on DBLP-citation-network V10[12] which mirrors a real-world dataset. The dataset conflates a number of sources which includes DBLP, ACM, MAG (Microsoft academic graph) [39] from which we have extracted the citation relevant data. In the original DBLP dataset, there were four JSON files. First, we concatenated all the JSON files (dblp-ref-0.json, dblp-ref-1.json, dblp-ref-2.json, dblp-ref-3.json) and made a data frame. The dataset mentioned above consists of approximately 3,079,007 papers and about 25,166,994 citations.

All the paper in the dataset have a corresponding author, an abstract, a title, a list of references, a publication venue and the year of publication. All of the duplicate records, inconsistent entries in the database, papers with missing fields and journals having less than 5 number of papers were filtered out leaving us with an approximate of 2,234,771 papers. Non-textual content in the abstract of some of the papers were also excluded from the experiment. The dataset was then segregated into three incongruous parts consisting of a training set of 81%, validation set of 9%, and the remaining 10% contributed to the test set.

### 5.2. Evaluation strategy

Two kinds of evaluation is employed to access the performance of CLAVER juxtaposed various other frequently employed modern methods.

(a) *Coarse-level or offline evaluation*: This method provides a superficial understanding of the viability of CLAVER as compared to other methods. Inclusion of original venue of publication of a given paper and its corresponding rank among the top N recommendation is indicative of the accuracy of the result. The evaluation metrics employed here are MRR and

[12] https://aminer.org/citation.

Accuracy (detailed below). The above mentioned scenario is nomenclated as *offline* because such an evaluation of a system can be done only when the test data is annotated and the original venue of the given paper is known.

(b) *Fine-level or online evaluation*: This method of evaluation realistically craters to the needs of a researcher who requires an enumeration of multiple venues at her disposal where she could possibly communicate her work. This also explains why we have called this method *online*. Here we tend to dwell deeper into the intricacies of evaluation and incorporate relevance, quality and usefulness of the recommended results and present a list of venues which are ranked in terms of their graded relevance Eq. 31. The evaluation metrics used for this evaluation strategy are precision, average venue quality, nDCG and diversity.

### 5.3. Evaluation metrics

Multiple evaluation metrics that were used comprises of precision@k, MRR, accuracy, Average venue-quality (Ave-quality) and nDCG@k to access the vitality of CLAVER.

(a) Accuracy@N: In this scenario Accuracy is defined as the fraction of correct prediction where the actual venue of the publication is included in the top N recommendation made by our system to the total number of papers in the dataset [24]. N in the above mentioned accuracy metrics can vary and have multiple values between 3 and 15.

$$Accuracy@N = \frac{\text{correctly predict venues within top N}}{\text{Total number of test papers}} \tag{21}$$

(b) Mean Reciprocal Rank (MRR): As the name suggest in MRR the reciprocal of the position where the correct venue i.e the actual venue of the paper in question appears in the output result and their sum is then averaged out against the total number of papers.

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_{rel_i}} \tag{22}$$

here $rank_{rel_i}$ denotes the rank of the correct venue i.e the actual venue of the ith paper, in a set of papers Q, appears in the list of recommended papers.

(c) Precision: The ratio of the number of relevant items that are retrieved in a recommendation set to the total number of recommendations is Precision. In the aforementioned scenario in our paper, precision depends on the number of relevant venues recovered in the prediction set as denoted by the Eqn. 23.

$$Precision = \frac{|\text{relevant venues} \cap \text{ recommended venues}|}{\text{total recommended venues}} \tag{23}$$

Precision@k denotes that $k$ venues were recommended, i.e.,

$$Precision@k = \frac{|\text{relevant venues} \cap \text{recommended venues}|}{k} \tag{24}$$

(d) Normalized discounted cumulative gain (nDCG): nDCG refers to summation of the fraction of discounted system gain upon the discounted ideal gain for a rank $p$, where the summation of the relevance from rank 1 to rank $p$ is known as the gain at a rank $p$. Relevance value $(rel_{sj})$ refers to a score (0, 1 or 2) attributed to the $jth$ venue by a researcher. Ideal vector is a hypothetically formulated vector where the relevance scores $(rel_{ij})$ are arranged in descending order to ascertain maximum gain at any given rank.

$$DCG_{sp} = rel_{s1} + \sum_{j=2}^{p}\frac{rel_{sj}}{log_2(j)} \tag{25}$$

$$IDCG_p = rel_{i1} + \sum_{j=2}^{p}\frac{rel_{ij}}{log_2(j)} \tag{26}$$

$$nDCG_p = \frac{DCG_p}{IDCG_p} \tag{27}$$

(e) Diversity (D): Diversity refers to the mean dissimilarity, which is the converse of similarity, among the multiple pairing of items in the recommendation set [20].

$$D = 2 * \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(1 - Similarity(v_i, v_j))}{N(N-1)} \tag{28}$$

here $N$ denotes the size of the recommendation set, $v_i$ and $v_j$ refers to the venues in the recommendation list and Similarity$(v_i, v_j)$ refers to the similarity with respect to content(abstract, keywords, Title) between the venues $v_i$ and $v_j$.

(f) Average-Venue Quality (Ave-quality): Ave-quality measures the inherent quality of the recommendation made by our system CLAVER taking into account the Google's h5-index [50].

$$\text{Average-venue quality} = \frac{\sum_{v \in V} H5_v}{|V|} \tag{29}$$

where $V$ represents the list of the venues suggested by our model and $H5_v$ refers to the h5-index of any particular venue v. A high Ave-quality inherently signifies the quality of the recommendation.

(g) Stability: The Stability of a recommendation algorithm is determined by the consistency of the prediction made by the algorithm over an ephemeral span of time. It is analogous to Mean Absolute Shift (MAS), which expresses the robustness and the stability of a recommendation system in terms of the intrinsic regularity among the recommendations made by the system [3]. Stability is formally defined on the training set $R_1$ and the corresponding recommendation for a given seed paper, $P_1$. For any span of time, which is concomitant with the assimilation of new data points, lets assume that the new recommendation made by the model is, $P_2$. MAS can then be expressed as

$$Stability = MAS = \frac{1}{|P_2|} \sum_{(u,i) \in P_2} |P_2(u,i) - P_1(u,i)| \tag{30}$$

where $P_1, P_2$ are the suggestions entrusted by the model during the first and the second phases, respectively.

### 5.4. Experimental setting

In this section, we present the experimental dataset for offline and online evaluations. Then the procedure of online assessment is described in further sub-section. The colossal quantity of data is accompanied by an equivalent increase in the number of labels. Due to which it is difficult to be train and learn the proposed model.

To resolve these issues mentioned above, along with other operational constraints (resource and time), the experimentation is segregated into two stages to demonstrates the effectiveness of CLAVER.

(i) The first phase consists of preparing a dataset for offline strategy of evaluation
(ii) The second phase is associated with the dataset preparation for the online evaluation strategy.

#### 5.4.1. Preparing the dataset for offline assessment

Initially, we identified only those venues whose number of papers were less than 500 and removed such venues having several papers more than 500. Dataset contains total of $342,258$ research papers after preprocessing and is split into three parts training set = 81%, test set = 10%, and a validation set = 9%. The complete statistics of the overall dataset is depicted in Table 1.

Seed papers are chosen from the testing dataset. The choice of paper is influenced by the cold-start conundrum that may arise for new researchers and comparatively new venues. We have assimilated four categories of venues and a likewise four categories of researchers on the basis of the number of papers published at any particular venue ($v_c$) and the number of publication for any given researcher ($p_c$) [13,50] and the categories formed are enlisted as following.

(i) Category 1: $5 \leqslant v_c < 20$
(ii) Category 2: $20 \leqslant v_c < 50$
(iii) Category 3: $50 \leqslant v_c < 100$
(iv) Category 4: $100 \leqslant v_c$
(v) Category 5: $5 \leqslant p_c < 20$
(vi) Category 6: $20 \leqslant p_c < 50$
(vii) Category 7: $50 \leqslant p_c < 100$
(viii) Category 8: $100 \leqslant p_c$

A well rounded representation of all the papers is ascertained in each category.

**Table 1**
Dataset for offline assessment.

| Types | Training dataset | Validation dataset | Testing dataset |
|---|---|---|---|
| No. of papers | 277,229 | 30,803 | 34,226 |
| No. of venues | 2208 | 2208 | 2208 |

**Table 2**
Dataset for online assessment.

| Range of papers | Number of papers | Number of journals |
|---|---|---|
| $5 \leqslant X \leqslant 100$ | 240,667 | 1009 |
| $100 \leqslant X \leqslant 400$ | 229,547 | 1027 |
| $400 \leqslant X \leqslant 2000$ | 825,940 | 934 |
| $2000 \leqslant X \leqslant 10{,}000$ | 915,876 | 232 |
| $X \geqslant 10{,}000$ | 22,741 | 14 |
| Max. class size | 121,328 | 1 |
| Min. class size | 5 | 67 |
| Avg. class size | 694 | |
| All | 22,34,771 | 3216 |

**Table 3**
Dataset for online assessment.

| Types | Training dataset | Validation dataset | Testing dataset |
|---|---|---|---|
| No. of papers | 1,532,935 | 170,327 | 189,251 |
| No. of venues | 1008 | 1008 | 1008 |

*5.4.2. Preparing the dataset for online assessment*

After filtering out venues having papers less than a threshold of 500 we were left with 1008 venues and $1,892,513$ papers.

The dataset was segregated into three incongruous parts consisting of a training set of 81%, validation set of 9%, and the remaining 10% contributed to the test set. The venue-wise papers statistics and other dataset related stats are shown in Tables 2 and 3 respectively.

We restricted our experimentation to 20 sub-domains of computer science due to multiple operational constraints. An aggregate of 160 papers (8 from each sub-domains and perpetually 1 from each category) were selected from 20 sub-domains: compiler design (CD), human–computer interaction (HCI), information retrieval (IR), parallel and distributed systems (PDS), data mining (DM), wireless sensor network (WSN), security (SC), computer vision (CV), machine learning (ML), image processing (IP), artificial intelligence (AI), databases (DB), software engineering (SE), natural language processing (NLP), worldwide web (WWW), computer architecture (CO), web semantics (WS), algorithms and theory (AT), theory of computation (TC) and multimedia (MM).

*5.4.3. Procedure of online evaluation*

The evaluation demanded annotations signifying the relevance of the venues predicted which was assimilated from some volunteer researchers. A total of 78 researchers with relevant expertise in some of the aforementioned domains were entrusted with the suggestions of our model. The recommendation comprised of a list of 15 venues recommended for each paper.

The credentials of the experts were distinguished with a modicum of at-least three years of relevant research background. Most of the volunteer researchers were Ph.D with an exception of some who were either research assistants or students with a relevant masters' or bachelors' in the field of technology and science. The research or the experts who were chosen demonstrated avid interest and reconciliated felicitously with the topics of the papers in the dataset. Out of the 78 researchers, 6 were research assistants, 18 were senior research students, exactly 23 were assistant professor, there were 15 associate professors and the remaining 16 were professors.

All experts were from distinguished institution and universities like Indian Institute of Technology (IIT) Indore, IIT Roorkee, IIT Bhubaneswar, IIT (BHU) Varanasi, Jadavpur University, Central University Hyderabad, Banaras Hindu University (BHU) and Manipal University. The age of the volunteer researchers were in the following range: professors [22–25], associate professors [44], assistant professors [17,18,19,20], senior research students [29] and remaining research assistants were between [29,15] years old. There were 51 male researchers and 27 female researchers.

On the basis of the scope of any particular venue and its parity with the attributes of the paper such as the abstract and title as well as its domain, each expert consigns a relevance score $(r)$ to each venue recommended by the model for any particular seed paper.

$$\text{Relevance } (r) = \begin{cases} 2 & \text{perfectly matching} \\ 1 & \text{partial matching} \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

As we know that precision score is valid only for binary values, the recommendations whose relevance score is 2 are only taken into consideration omitting the recommendations with relevance scores of 1 and 0.

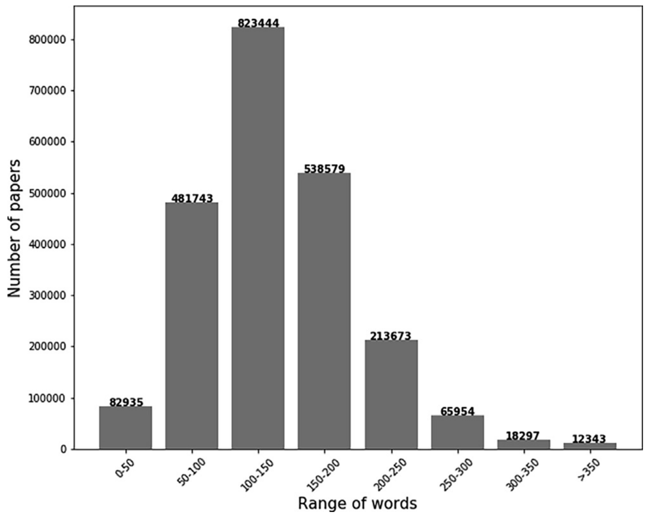


**Fig. 6.** Statistics of word count of abstracts.

The evaluation of our model is done in a more comprehensive manner and also the issues enlisted in Section 1 are taken into consideration. For the same we have examined the below mentioned research queries (RQs):

**RQ1:** Effectiveness of CLAVER juxtaposed to other modern methods?
**RQ2:** Quality and robustness of CLAVER in providing valid recommendation juxtaposed to other modern methods?
**RQ3:** Effectiveness of CLAVER in tackling issues like cold-start, data sparsity, stability and data diversity?

*5.5. Parameters and hyper-parameters tuning*

Title and abstract of paper are concatenated as (text) to pass through embedding layer of keras. Size of text is fixed at $L = 300$ (95% of texts have length below 300), each word is embedded to $d = 300$ length vectors. Text having a length less than 300 words are padded, and text has more than 300 words are truncated. We have taken the maximum size of the abstract as 300 because it is the maximum number of the words that most of the abstracts contain as depicted in Fig. 6.

We trained data using CNN model with two convolution layer having filter sizes of (4, and 5) respectively, two max-pooling layers with the filter size of 2 and stride of 2, dense layers of size 4,096 and 2,048 with the Dropout rate of 0.2. The dense layer is connected to the softmax layer of size 1,008 i.e., the number of total venues. For LSTM layer, we have used single LSTM layer followed by Bi-LSTM and four dense layers with Dropout (dropout rate = 0.3) and BatchNormalization. The final layer is the softmax layer having 1,008 labels representing a total number of Journals for training.

Model is trained on categorical cross-entropy loss function, Adam optimizer, and metrics as accuracy, top-3, top-5, top-10 and top-15 accuracy. Model is trained for 13 epochs, and it is validated after every epoch with a validation set of 10% of training data. Batch size of 1024 was used during training. Early stopping with a delta of 0.00001 and patience of 2 was applied on validation loss for a call back during training. Once the model is trained, we evaluate our model on test data.

*5.6. Baseline methods*

To determine the effectiveness of our model we have compared CLAVER with other modern methods which are elaborated below.

(a) Friend based model (FB): The cardinal characteristics of FB is that it provides recommendation based on the a graph where the immediate link (co-authors) and the second immediate like (co-authors' co-authors) for any given author in the graph are consider to provide valid venue recommendations. If a venue is frequented by any of the above mentioned neighbours then that venue is recommended. [8].
(b) Collaborative filtering model (CF): This method assimilates a memory-based CF approach. The fundamental assumption made here is that the probability of a work getting published at any particular venue commensurates with the domain of other articles published in that venue and their parity with the topic of the researchers work. [47].
(c) Co-authorship network-based model (CN): CN basically leverages a network of researchers where each researcher is connected to is co-author and so on. The prediction here is based on the reputation and credentials of the other's in the proximity of the author for whome the venues are to be suggested. Other factors like venues sub-domain, venue name and number of publications are considered [22].

(d) Content-based filtering model (CBF): CBF works on the idea whose foundation is laid on the basis of a similarity score, computed using LDA model, between the publications of the author and the domain of the publications at any particular venue [24].
(e) Personal venue rating-based collaborative filtering model (PVR): PVR is based on the idea of recommending venues which are rated on the basis of those publications in that venue which are also referred by an authors previous work [2].
(f) Random walk with restart model (RWR): Here the recommendations are provided by performing a random walk on a hybrid author-venue relationship graph along-with a restart technique. Here the probability of skipping to the neighbourhood nodes is identical for all nodes [6].
(g) Publication recommender system (PRS): PRS leverages a new form of CBF model integrated with softmax and chi-square regressing techniques. This method employs two segregated module, one for feature selection and the second is the regression module [42].
(h) Personalized academic venue recommendation models (PAVE): This method shows striking similarity with the RWR method with a miniature difference where a transfer matrix is used with a small bias. In contrast with RWR, here the probability of skipping over to the next node is based on factors like relation weight, co-publication frequency and researchers academic reputation [50].
(i) LSTM: LSTM (Long Short Term Memory) shows significant results in classifying text and therefore can be used to obtain a higher level rendition of the title and the abstract. This can subsequently be fed to a softmax function for journal classification.
(j) CNN: (Convolutional Neural Network) shows significant results in classifying text and therefore can be used to obtain a higher level rendition of the title and the abstract.

In addition to the above state-of-the-art techniques, we also incorporated validating CLAVER against some trivial deep-learning models like LSTM, CNN, Bi-LSTM, CNN + Bi-LSTM, RNN + CNN. Among the methods expounded above CF and PVR are based on Collaborative Filtering, PAVE and RWR are network or graph based approaches depending on co-authorship networks on which random walk algorithm is performed, CN and FB are network or graph based approaches depending on co-authorship networks, CBF is based on, as the very name suggests, content-based filtering technique, and LSTM, CNN, RNN, Bi-LSTM are all deep learning models.

## 6. Results and discussion

This section circumscribes a discussion on the effectiveness of CLAVER juxtaposed to other modern methods. For an effortless and comprehensive understanding of the results produced in this section we have segregated the discussion into two sections: Offline and Online evaluations in particular. A paired-sample t-test was also performed, taking into consideration all the evaluation metrics (precision, nDCG, MRR and Accuracy in particular), between CLAVER and another baseline method which performed the second-best in the experimentation. In the course of enhancing the clarity of this discussion we have marked the results which are statistically significant, having $p$ values below 0.05, with '*' and the baseline method performing second-best with '+' symbols.

### 6.1. Offline evaluation of CLAVER

The offline evaluation aims to evaluate the effectiveness of CLAVER taking into consideration the evaluation metrics of MRR and Accuracy. A comprehensive result can be seen in Table 4 while evaluating it for the journals having less than 500 papers. With these results it can be apparently justified that CLAVER, taking into account and compared to all the other approaches, gives a steadily robust performance and is concomitantly consistent, as it could assimilate the original venue of any given input paper nearly 47% of the time, that too in among the top 3 recommendations.

At position 3 the proposed model shows an initial accuracy of 0.4789 . The accuracy however steadily spirals upwards, showing a staggering accuracy of 0.7869 at the 15th position. According to the experimentation results PAVE can be considered as the second best of all the other baseline approached at position 3, 6 and 9 respectively. A comprehensive result for the two accuracy metrics can be seen in Table 5 while evaluating it for the journals having more than 500 papers, it could assimilate the original venue of any given input paper nearly 85% of the time in among the top 15 recommendations. FB certainly performs the worst among all the methods.

Likewise, while evaluating MRR, we can see that the proposed model CLAVER performs better than all the other modern baseline methods showing an MRR of 0.3038 in case of evaluation with journals having less than 500 papers. Similarly it shows an MRR of 0.3349 while evaluating with journals having more than 500 papers. PRS method performs the second best while FB again gives the worst results in terms of MRR metrics as it can be evidently understood from Table 5.

### 6.2. Online evaluation of CLAVER

The Online evaluation technique aims to evaluate the effectiveness of CLAVER juxtaposed to other modern methods while taking into consideration the evaluation metrics of nDCG, precision and Ave-quality.

**Table 4**
Overall results of accuracy and MRR (Journals < 500 papers).

| Approach | Acc@3 | Acc@6 | Acc@9 | Acc@12 | Acc@15 | MRR |
|---|---|---|---|---|---|---|
| FB | 0.0396 | 0.0784 | 0.1005 | 0.1397 | 0.1685 | 0.0279 |
| CF | 0.0743 | 0.0994 | 0.1317 | 0.1596 | 0.2016 | 0.0316 |
| CN | 0.0957 | 0.1256 | 0.1648 | 0.2006 | 0.2325 | 0.0412 |
| CBF | 0.1362 | 0.1673 | 0.1988 | 0.2005 | 0.2648 | 0.0414 |
| RWR | 0.1718 | 0.2005 | 0.2386 | 0.2719 | 0.2867 | 0.0579 |
| PVR | 0.1894 | 0.2037 | 0.2124 | 0.2756 | 0.2948 | 0.0671 |
| PRS | 0.2267 | 0.2652 | 0.2884 | $0.3659^{+}$ | $0.4732^{+}$ | 0.1059 |
| PAVE | $0.2274^{+}$ | $0.2695^{+}$ | $0.2997^{+}$ | 0.3244 | 0.3991 | $0.0878^{+}$ |
| CNN | 0.2593 | 0.3007 | 0.3906 | $0.4275^{+}$ | 0.4529 | 0.1287 |
| LSTM | 0.3932 | 0.4963 | 0.5865 | $0.6294^{+}$ | 0.6549 | 0.1593 |
| Bi-LSTM | 0.4076 | 0.5037 | 0.5479 | 0.6173 | 0.6482 | 0.1589 |
| CNN + Bi-LSTM | 0.4182 | 0.5096 | 0.5845 | 0.6362 | 0.6547 | 0.1873 |
| RNN + CNN | 0.3887 | 0.4779 | 0.4867 | 0.5343 | 0.5532 | 0.1562 |
| CLAVER | $0.4789^{*}$ | $0.5607^{*}$ | $0.6793^{*}$ | $0.7071^{*}$ | $0.7869^{*}$ | $0.3038^{*}$ |

Symbols '*' and '+' signifies statistically significant and the second best results in the table.

**Table 5**
Overall results of accuracy and MRR (Journals ⩾ 500 papers).

| Approach | Acc@3 | Acc@6 | Acc@9 | Acc@12 | Acc@15 | MRR |
|---|---|---|---|---|---|---|
| FB | 0.0555 | 0.0972 | 0.1250 | 0.1666 | 0.1944 | 0.0338 |
| CF | 0.0972 | 0.1111 | 0.1527 | 0.1805 | 0.2361 | 0.0451 |
| CN | 0.1111 | 0.1388 | 0.1805 | 0.2222 | 0.2500 | 0.0516 |
| CBF | 0.1527 | 0.1805 | 0.2083 | 0.2361 | 0.2916 | 0.0648 |
| RWR | 0.1944 | 0.2222 | 0.2500 | 0.2916 | 0.3194 | 0.0775 |
| PVR | 0.2083 | 0.2361 | 0.2368 | 0.3194 | 0.3472 | 0.0863 |
| PRS | 0.2497 | 0.2877 | $0.3265^{+}$ | $0.3987^{+}$ | $0.5467^{+}$ | $0.1356^{+}$ |
| PAVE | $0.2500^{+}$ | $0.2916^{+}$ | 0.3055 | 0.3611 | 0.4305 | 0.0906 |
| CNN | 0.4379 | 0.5264 | 0.7006 | 0.7208 | 0.7273 | 0.1287 |
| LSTM | 0.4563 | 0.5647 | 0.6752 | 0.7342 | 0.7501 | 0.1385 |
| Bi-LSTM | 0.4547 | 0.5659 | 0.6377 | 0.6984 | 0.7482 | 0.1294 |
| CNN + Bi-LSTM | 0.4874 | 0.5651 | 0.6659 | 0.7345 | 0.7842 | 0.2203 |
| RNN + CNN | 0.4231 | 0.4474 | 0.4645 | 0.4981 | 0.5042 | 0.5124 |
| CLAVER | $0.5382^{*}$ | $0.6428^{*}$ | $0.7317^{*}$ | $0.8124^{*}$ | $0.8532^{*}$ | $0.3349^{*}$ |

Symbols '*' and '+' signifies statistically significant and the second best results in the table.

#### 6.2.1. Precision@k

A comprehensive result for Precision metrics can be viewed in Table 6. From the experimentation results the superiority of our model CLAVER is evident. At position 3 the proposed model shows an initial precision of 0.9361, then showing an upward spike and reaching a precision of 0.9404 at position 7 before reaching a precision of 0.9356 till 15 position as displayed in Fig. 7.

The maxima and minima, for CLAVER model, is marked by a precision of 0.9510 at position 6, and a precision 0.9227 at position 2, respectively. Second best results were exhibited by PAVE with some exceptions at 5, 6, 7 and 8 positions where PRS model slightly outperform PAVE. FB again gives the worst results in terms of precision metrics.

#### 6.2.2. nDCG@k

A comprehensive result for nDCG@k metrics can be viewed in Table 7. It is clearly visible that the proposed CLAVER shows a significant improvement of nDCG over all other state-of-the-art methods. At position 2 the proposed model shows an initial nDCG@k of 0.9411, then showing a downward spike and reaching a nDCG@k score of 0.9394 at position 8 before getting an upward spiral again and reaching a nDCG@k score of 0.9435 till 15 position as displayed in Fig. 8 which clearly shows that the overall nDCG results of CLAVER is consistent and shows the highest nDCG of 0.9524 at position 3. PAVE model performs the second best while FB again gives the worst results in terms of nDCG metrics.

#### 6.2.3. Average venue quality (H5-Index) analysis

In Ave-quality and evaluation of the quality of the venues recommend by CLAVER is investigated juxtaposed to other approaches and the same is shown in Fig. 9. H5-Index of venues suggested by CLAVER has a mean of 94. The highest H5-Index of 99 is seen at position 5 which spirals down to an H5-Index of value 95 at position 15. The lowest quality of the venue suggested by CLAVER with an H5-index of 84 is seen at position 2.

**Table 6**
Precision results for proposed CLAVER and other compared approaches.

| Methods | P@3 | P@6 | P@9 | P@12 | P@15 |
|---|---|---|---|---|---|
| FB | 0.6052 | 0.5899 | 0.5753 | 0.5522 | 0.5798 |
| CF | 0.6125 | 0.6354 | 0.6357 | 0.6466 | 0.6354 |
| CN | 0.6801 | 0.6681 | 0.6715 | 0.6690 | 0.6592 |
| CBF | 0.7125 | 0.7121 | 0.7116 | 0.7016 | 0.7009 |
| RWR | 0.7460 | 0.7226 | 0.7159 | 0.7179 | 0.7208 |
| PVR | 0.8159 | 0.7848 | 0.7529 | 0.7618 | 0.7601 |
| PRS | 0.8079 | $0.8380^{+}$ | 0.7654 | 0.7609 | 0.7639 |
| PAVE | $0.8219^{+}$ | 0.8049 | $0.8395^{+}$ | $0.8393^{+}$ | $0.8412^{+}$ |
| CNN | 0.8574 | 0.8579 | 0.8564 | 0.8271 | 0.8376 |
| LSTM | 0.8803 | 0.8867 | 0.8786 | 0.8654 | 0.8537 |
| Bi-LSTM | 0.8748 | 0.8793 | 0.8694 | 0.8563 | 0.8469 |
| CNN + Bi-LSTM | 0.8792 | 0.8673 | 0.8685 | 0.8492 | 0.8338 |
| RNN + CNN | 0.8547 | 0.8495 | 0.8437 | 0.8347 | 0.8297 |
| CLAVER | $0.9361^{*}$ | $0.9510^{*}$ | $0.9382^{*}$ | $0.9461^{*}$ | $0.9356^{*}$ |

Symbols '*' and '+' signifies statistically significant and the second best results in the table.

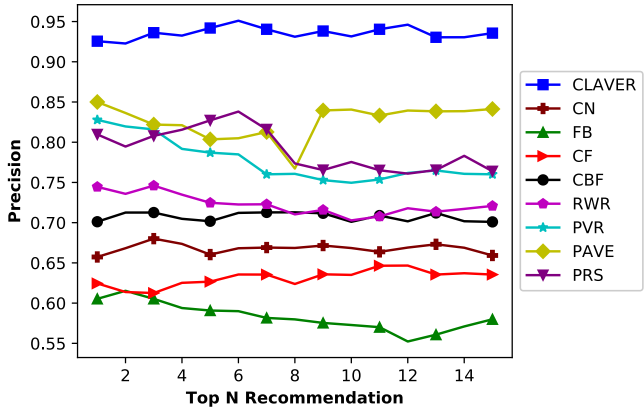


**Fig. 7.** Precision analysis for CLAVER and other approaches.

**Table 7**
nDCG results for proposed CLAVER and other approaches.

| Methods | nDCG@3 | nDCG@6 | nDCG@9 | nDCG@12 | nDCG@15 |
|---|---|---|---|---|---|
| FB | 0.6409 | 0.6229 | 0.6238 | 0.6244 | 0.6281 |
| CF | 0.6678 | 0.6767 | 0.6782 | 0.6800 | 0.6786 |
| CN | 0.6944 | 0.7028 | 0.6985 | 0.7009 | 0.7014 |
| CBF | 0.7562 | 0.7402 | 0.7408 | 0.7478 | 0.7530 |
| RWR | 0.7499 | 0.7494 | 0.7437 | 0.7502 | 0.7562 |
| PVR | 0.7847 | 0.7908 | 0.7867 | 0.7872 | 0.7778 |
| PRS | 0.7794 | 0.7830 | 0.7936 | 0.7849 | 0.7940 |
| PAVE | $0.8356^{+}$ | $0.8288^{+}$ | $0.8298^{+}$ | **0.8496** | **0.8368** |
| CNN | 0.8689 | 0.8657 | 0.8686 | 0.8581 | 0.8571 |
| LSTM | 0.8879 | 0.8705 | 0.8554 | 0.8492 | 0.8356 |
| Bi-LSTM | 0.8793 | 0.8691 | 0.8566 | 0.8398 | 0.8295 |
| CNN + Bi-LSTM | 0.8642 | 0.8632 | 0.8584 | 0.8473 | 0.8344 |
| RNN + CNN | 0.8547 | 0.8495 | 0.8437 | 0.8347 | 0.8297 |
| CLAVER | $0.9524^{*}$ | $0.9361^{*}$ | $0.9454^{*}$ | $0.9378^{*}$ | $0.9435^{*}$ |

Symbols '*' and '+' signifies statistically significant and the second best results in the table.

### *6.2.4. Computational complexity analysis*

The complexity of a deep learning model entails some computational limitations. The proposed model CLAVER consists of CNN, LSTM, Bi-LSTM layers, and Attention mechanism, which makes the model data-hungry and also bloats up the number of trainable parameters to the order of $10^6$. The training and testing were done using a Tesla P100-PCIE modern GPU with 16 GB Stacked Memory Capacity and a compute capability of 6.0 . It takes close to 2 min per epoch while training the dataset with a standard batch size of 64. During the training phase, the Tesla GPU was to throttled to 75% of its maximum capacity.

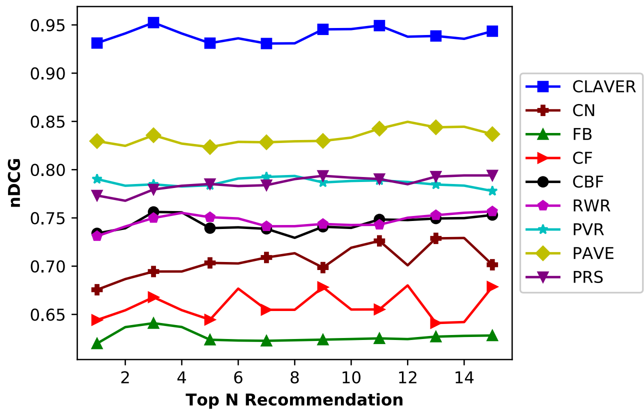


**Fig. 8.** nDCG analysis for CLAVER and other approaches.

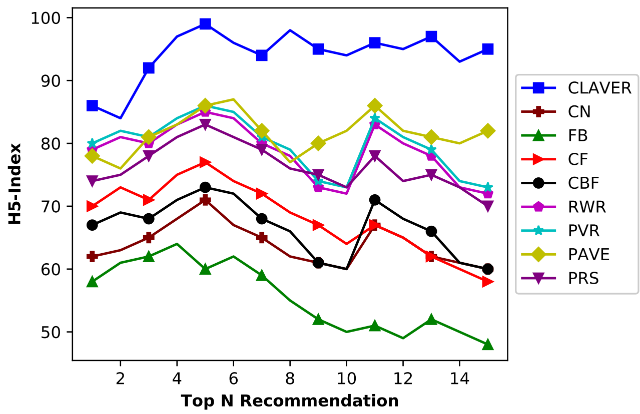


**Fig. 9.** Venue quality of CLAVER and other approaches.

The time to predict or verify a hypothesis took close to 15 s for 10 examples using the same hardware configurations. It can be safely established that a raw measurement of the amount of time taken in training and classification is heavily biased and depends on the hardware platforms, software packages and profiler quality used. The raw measurement can just help in providing a baseline and reflect more about the secondary requirements than the algorithm itself.The computational complexity of the model in terms of the number of trainable parameters or the overall time taken to train and predict a hypothesis exceeds other conventional methods. And therefore, the model requires a high-performance GPU to supply output in a realistic amount of time. However, the primary aim of the model was to outperform other conventional machine learning and simplistic deep learning architectures in terms of diversity and stability. And the model is more focused on overcoming the problems of cold-start and data sparsity; therefore, very less consideration is given to reduce the computational overload of the architecture.

### 6.2.5. Statistical significance test

A paired-samples t-test is a parametric test used with the aim of establishing the fact that the mean difference over a set of paired observation shows a considerable deviation from zero. This can be defined as following

$$t = \frac{m}{s/\sqrt{n}} \tag{32}$$

Here $m$ and $s$ are the average and standard deviation (SD) of the differences over a set of paired observation and n is the size of the sample space.

Mathematically the alternative and null hypothesis are defined below.

$H_0$: $\mu_d$ =0 (True mean difference is equal to zero).

$H_1$: $\mu_d \neq 0$ (True mean difference is not equal to zero).

In this paired-samples t-test, under the null hypothesis, it follows a t-distribution with $n - 1$ Degrees of Freedom (DF). Only $p$ values less than 0.05 were considered statistically significant at 5% level of significance ($\alpha$ =0.05). If the p-value is less than $\alpha$, we are rejecting the null hypothesis.

We performed such tests on overall precision, nDCG, accuracy between CLAVER, and other modern methods. The test was performed at multiple positions including 15, 12, 9, 6, and 3 and the results of the test was compared. 160 pairwise comparisons was required (DF = 159) for the multiple position during the evaluation. Similarly, we performed such a statistical test on the overall results of MRR, diversity, and stability to establish the effectiveness of our model CLAVER over others. The statistically significant results are annotated with a "*" symbol were CLAVER can easily be identified to out-perform other modern approaches with a statistically significant difference.

## 7. Study of the proposed approach

Some of the cardinal findings with respect to our Research Queries(RQs) as mentioned in Section 5.4.3 are summed up below:

### 7.1. ***RQ1***: *Effectiveness of CLAVER juxtaposed to other modern methods?*

A comprehensive study of the accuracy, nDCG@k, MRR and precision@k scores for CLAVER deep learning model were compared juxtaposed to other modern techniques. The precision and nDCG scores shows a significant superiority of CLAVER over other modern techniques. The performance measures with respect to MRR and accuracy also suggest that using CLAVER for a reliable venue prediction should be ideal as compared to other techniques. The results for the same are enumerated in Table 4, Table 5, Table 6, and Table 7 respectively.

### 7.2. ***RQ2***: *Quality and robustness of CLAVER in providing valid recommendation juxtaposed to other modern methods?*

The quality of the venues recommended by CLAVER is significantly high when it is compared to other modern niches in the field, and the same can be verified from Fig. 9. The mean H5-index of CLAVER demonstrates a most elevated average estimation of 94 after recommending 15 venues. The least H5-index for any venue recommended by CLAVER is demonstrably 84 which shoots up and culminates to a maximum of 99. The corresponding maxima and minima shown by the recommendations made by PAVE are demonstrably 87 and 76. PAVE model performs the second best while FB again gives the worst results in terms of H5 quality index of the venues.

### 7.3. ***RQ3***: *Effectiveness of CLAVER in tackling handle issues like cold-start, data sparsity, stability and data diversity?*

(i) **Cold-start issues**: In order to address prevalent problems like the "cold-start" issues, which comes into relevance when new venues spawn up or a new author wants to find venues to publish a paper, we assimilate certain high level features like the title and the abstract of the paper and train it with CNN model and LSTM model, respectively into the proposed model CLAVER. We applied CNN in order to extract latent factors from the content information, which then directly integrated into the process of recommendation seems to deal stinkingly well with the issues of cold-start. We investigated the performance of CLAVER while assessing for newly spawned venues and new researchers associated inputs (seed papers). The analysis of Table 8 depicts that, our model CLAVER could suggest relevant venues at a very embryonic stage irrespective of the novelty of the researcher or the newness of the venue. No co-authorship network or any past records of publications were required to provide an effective venue recommendation. The model, in order to predict an effective and relevant venue, requires only some trivial precocious informations like the title and the abstract along with the area of interest of the author who wants to publish their work. CLAVER does not show issues of cold start as discussed perceptually in the following section and also depicted in Table 9.

(ii) **Data sparsity**: In order to solve the issues of sparseness of data we have employed a very trivial convolution operation. A convolution operation transforms a high dimensionality of an input vector to a low dimensional output, capturing only relevant and essential features for the input feature vector. We have transformed the high dimensional and sparse embedding matrix into a lower-dimensional and dense set using deep learning-based convolutional operation. A convolution layer is adept in extracting contextual and temporal information from a mostly sparse, high dimensionality dataset. For example, in the convolutional layer, 256 filters with window size 4 move on the textual representation to extract the features. As the filters move on, many sequences that capture the syntactic and semantic features are generated. In this paper, we have considered that the dimension of input data is $300 * 300$, and the dimension of the first layer output data is $148 * 1$ and successively, the second layer output is of dimension $72 * 1$. Hence the convolutional layer is an effective way for dimensionality reduction. As specified in Table 9, CLAVER does not have data sparsity issues.

(iii) **Diversity**: To alleviate the problem of diversity, we integrated both convolution layer and LSTM layer to capture the contextual similarity-based relevant features and to extract contextual and temporal information from a mostly sparse, high dimensionality dataset.

**Table 8**
MRR results for CLAVER over new venue and new researcher.

| Methods | $5 <= v_c < 15$ | $15 <= v_c < 50$ | $50 <= v_c < 100$ | $100 <= v_c$ | $5 <= p_c < 15$ | $15 <= p_c < 50$ | $50 <= p_c < 100$ | $100 <= p_c$ |
|---|---|---|---|---|---|---|---|---|
| FB | 0.0437 | 0.0591 | 0.0683 | 0.0765 | 0.0677 | 0.0769 | 0.0797 | 0.0839 |
| CF | 0.0473 | 0.0637 | 0.0725 | 0.0753 | 0.0746 | 0.0828 | 0.0866 | 0.0887 |
| CN | 0.0526 | 0.0773 | 0.0866 | 0.0899 | 0.0877 | 0.0914 | 0.0957 | 0.0974 |
| CBF | 0.0668 | 0.0848 | 0.0932 | 0.1094 | 0.0817 | 0.0895 | 0.0917 | 0.1095 |
| RWR | 0.0793 | 0.0849 | 0.0853 | 0.0864 | 0.0861 | 0.0874 | 0.0891 | 0.1064 |
| PVR | 0.0798 | 0.0869 | 0.0851 | 0.0853 | 0.0893 | 0.0847 | 0.0893 | 0.0847 |
| PRS | 0.0972 | 0.0995 | 0.1049 | 0.1128 | 0.0903 | 0.0945 | 0.0963 | 0.0995 |
| PAVE | 0.0977+ | 0.1014+ | 0.1298+ | 0.1416+ | 0.1091+ | 0.1373+ | 0.1539+ | 0.1873+ |
| CLAVER | 0.2289* | 0.2431* | 0.2864* | 0.3079* | 0.2644* | 0.3059* | 0.3144* | 0.3459* |

Symbols '*' and '+' signifies statistically significant and the second best results in the table.

**Table 9**
Cold-start and other issues available.

| Methods | Cold-start | Sparsity | Diversity | Stability |
|---|---|---|---|---|
| FB | Yes (new researcher) | No | Yes | Yes |
| CF | Yes (researcher and venue) | Yes | No | Yes |
| CN | Yes (new venue) | No | Yes | Yes |
| CBF | Yes (new venue) | No | Yes | No |
| RWR | Yes (new researcher) | No | Yes | Yes |
| PVR | Yes (researcher and venue) | Yes | No | Yes |
| PRS | Yes (new venue) | No | Yes | No |
| PAVE | Yes (new researcher) | No | Yes | Yes |
| CLAVER | No | No | No | No |

In order to find the diversity, the dissimilarity among the recommended results, of the predicted venues by our model, we extracted the keywords from the papers by grouping together the publications at a particular venue, and computed the similarity score according to Eq. 28. As shown in Table 10, CLAVER shows the best diversity with a diversity score of 0.536. PVR approach comes as the near second, which shows a diversity score of 0.402. A mean D-score was considered as a benchmark to posit the diversity of an approach.

(iv) **Stability**: To maintain consistency of the recommendation results, we have prioritise the stability of the proposed system CLAVER. In order to solve for stability issues we have presented an integrated architecture comprising of LSTM and Bi-LSTM layers preceded by a convolution layer and followed by attention mechanism for capturing only relevant and essential features for the input feature vector and extract contextual and temporal information from a mostly sparse, high dimensionality dataset. The attention mechanism is used to acquire the sentimentality of the input texts to understand the semantics of the input texts and also to weigh more contributing words more relevant words that contribute to the final venue recommendation and also solves the stability issue.

A comprehensive study to investigate the stability of the recommendations provided by CLAVER is done as defined by Eq. 30. As evident from the stats depicted in Table 11, our proposed model CLAVER shows a MAS score that is comparatively lower than all other approaches. A MAS score of 3.046 is achieved on the DBLP dataset. The significance of this MAS score, 3.046 for CLAVER, can be described in terms of the mean shift in the prediction rating after addition of a new entry into the dataset. The threshold, in order to qualify any given approach as stable, was taken to be the mean of all the MAS score. CLAVER was proved to be the most stable and desirable model, through experimentation and investigation which involved adding approximately 3% of the data into the training dataset from the test set.

### 7.4. More insightful discussion on the results

The vitality of the results and the performance of the model as discussed in Section 6 portrays the ability of CLAVER in providing a desired result. A significantly high precision score depicts the efficacy of the model in predicting a relevant venue. There are some restraints of the model which may encumber its performance at multiple levels. A significantly high number of trainable parameters and extensive experimentation impedes the performance of the model due to hardware restrictions. The complexity of the model also commensurates with the increase in the number of output classes and the model becomes data hungry and perpetually contributes to hardware cost. Dependence of convolution layer on the initial parameter tuning (for a good point) to avoid local optima. Thus, a weakness of convolution operation is the considerable amount of work they require to initialize according to the problem at hand. This would require some expert knowledge in the domain. One of the limitations of our model is that it cannot recommend appropriate venues if the input paper is not related (new field or unique keywords) to any papers available in the training set. The potential of CLAVER in predicting a relevant venue may be hobbled by the scarcity of related papers published in any particular venue.

**Table 10**
Diversity of CLAVER and other approaches.

| Methods | Diversity (D) |
|---|---|
| FB | 0.238 |
| CF | 0.369 |
| CN | 0.281 |
| CBF | 0.215 |
| RWR | 0.317 |
| PVR | 0.402$^{+}$ |
| PRS | 0.252 |
| PAVE | 0.323 |
| CLAVER | 0.536* |

'*' denote statistically significant results over the second best ('+').

**Table 11**
Stability of CLAVER and other approaches.

| Methods | MAS |
|---|---|
| FB | 8.204 |
| CF | 7.865 |
| CN | 8.339 |
| CBF | 4.575 |
| RWR | 6.965 |
| PVR | 7.165 |
| PRS | 4.212$^{+}$ |
| PAVE | 6.594 |
| CLAVER | 3.046* |

'*' denote statistically significant results over the second-best ('+').

## 8. Conclusion

Academic venue recommendation has proven to be one of the most sought after field of research in the modern scenario. Although there is a heightened demand in this domain, most of the present techniques suffer from multiple issues like cold-start and data sparsity. They also fail to tackle the situation where new venue are spawned and added to the dataset. The solutions are deeply impeded by stability and diversity issues.

This article proposes a novel approach CLAVER which is a venue recommender system based on an aggregated deep learning approach. It is a coherent unified framework of a convolutional layer followed by Long short-term memory (LSTM) layer and a Bi-LSTM layer with attention mechanism to effectively predict a relevant venue for a paper using only the abstract and title of a any given paper. Additionally CLAVER deals reasonably well with issues like cold-start, data sparsity and others. A comprehensive array of experiments were conducted on a real world DBLP dataset, and we successfully demonstrated that CLAVER consistently outperforms various modern methods. It reveals significantly high scores of precision@k, nDCG@k, MRR, and accuracy than other state-of-the-art techniques. The venues recommended by CLAVER culminate great quality by suggesting venues which have a H5-index greater than the recommendations made by other best-in-class techniques without compromising on stability and diversity.

The evolution of digital technology presents us with vast opportunities to continuously improve the current model. Some of the low hanging fruits which would also make this model a wholistic solution include functionalities like updating the dataset over-the-air by employing a web crawler which could generate and re-compile the training dataset on-the-go. A feedback followed by a feedforward mechanism would be used to continually enhance the model. The feedback mechanism can be further enhanced by incorporating Information-Retrieval(IR) techniques like pseudo relevant feedback and relevant feedback to improve the efficacy of our system.

We also intend to explore, in the near future, the problem and issues posited by venue recommendation systems and provide an even more extensively monitored approach to further improve diversity, serendipity, accuracy, precision, novelty and coverage.

## 9. Compliance with ethical standards

The authors associated with this article proclaim no conflict of interest. A novel decision support system CLAVER is proposed by this article. It is an integrated framework of exploiting both convolution layer, LSTM layer along with attention mechanism. We assert no use of any creature or human subjects during the examination.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


[1] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Trans. Knowl. Data Eng. 17 (6) (2005) 734–749.
[2] H. Alhoori, R. Furuta, Recommendation of scholarly venues based on dynamic user interests, J. Inf. 11 (2) (2017) 553–563.
[3] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, Recommender systems survey, Knowl.-Based Syst. 46 (2013) 109–132.
[4] I. Boukhris, R. Ayachi, A novel personalized academic venue hybrid recommender, in: 2014 IEEE 15th International Symposium on Computational Intelligence and Informatics (CINTI), 2014, IEEE, pp. 465–470..
[5] P. Chaiwanarom, C. Lursinsap, Collaborator recommendation in interdisciplinary computer science using degrees of collaborative forces, temporal evolution of research interest, and comparative seniority status, Knowl.-Based Syst. 75 (2015) 161–172.
[6] Z. Chen, F. Xia, H. Jiang, H. Liu, J. Zhang, Aver: random walk based academic venue recommendation, in: Proceedings of the 24th International Conference on World Wide Web, 2015, ACM, pp. 579–584..
[7] S. Choi, B. Ryu, S. Yoo, J. Choi, Combining relevancy and methodological quality into a single ranking for evidence-based medicine, Inf. Sci. 214 (2012) 76–90.
[8] C. Desrosiers, G. Karypis, A comprehensive survey of neighborhood-based recommendation methods, in: Recommender Systems Handbook, 2011, Springer, pp. 107–144..
[9] T. Ebesu, Y. Fang, Neural citation network for context-aware citation recommendation, in: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2017, ACM, pp. 1093–1096..
[10] M. Errami, J.D. Wren, J.M. Hicks, H.R. Garner, etblast: a web server to identify expert reviewers, appropriate journals and similar publications, Nucleic Acids Res. 35 (Suppl_2) (2007) W12–W15..
[11] X. Feng, H. Zhang, Y. Ren, P. Shang, Y. Zhu, Y. Liang, R. Guan, D. Xu, The deep learning-based recommender system "pubmender" for choosing a biomedical publication venue: development and validation study, J. Med. Internet Res. 21 (5) (2019) e12957.
[12] Y. Gao, Y.-F. Li, Y. Lin, H. Gao, L. Khan, Deep learning on knowledge graph for recommender system: a survey, 2020. arXiv preprint arXiv:2004.00387..
[13] H.A.M. Hassan, Personalized research paper recommendation using deep learning, in: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization, 2017, ACM, pp. 327–330..
[14] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[15] W. Huang, Z. Wu, L. Chen, P. Mitra, C.L. Giles, A neural probabilistic model for context based citation recommendation, AAAI (2015) 2404–2410.
[16] T. Huynh, K. Hoang, Modeling collaborative knowledge of publishing activities for research recommendation, Comput. Collect. Intell. Technol. Appl. (2012) 41–50.
[17] N. Kang, M.A. Doornenbal, R.J. Schijvenaars, Elsevier journal finder: recommending journals for your paper, in: Proceedings of the 9th ACM Conference on Recommender Systems, 2015, ACM, pp. 261–264..
[18] D. Kim, D. Seo, S. Cho, P. Kang, Multi-co-training for document classification using various document representations: Tf–idf, lda, and doc2vec, Inf. Sci. 477 (2019) 15–29.
[19] R. Klamma, P.M. Cuong, Y. Cao, You never walk alone: recommending academic events based on social network analysis, in: International Conference on Complex Sciences, Springer, pp. 657–670..
[20] M. Kunaver, T. Požrl, Diversity in recommender systems–a survey, Knowl.-Based Syst. 123 (2017) 154–162.
[21] D. Liang, L. Charlin, J. McInerney, D.M. Blei, Modeling user exposure in recommendation, in: Proceedings of the 25th International Conference on World Wide Web, 2016, International World Wide Web Conferences Steering Committee, pp. 951–961..
[22] H. Luong, T. Huynh, S. Gauch, L. Do, K. Hoang, Publication venue recommendation using author network's publication history, Intell. Inf. Database Syst. (2012) 426–435.
[23] H.P. Luong, T. Huynh, S. Gauch, K. Hoang, Exploiting social networks for publication venue recommendations, KDIR (2012) 239–245.
[24] E. Medvet, A. Bartoli, G. Piccinin, Publication venue recommendation based on paper abstract, in: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI), 2014, IEEE, pp. 1004–1010..
[25] E. Minkov, B. Charrow, J. Ledlie, S. Teller, T. Jaakkola, Collaborative future event recommendation, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, 2010, ACM, pp. 819–828..
[26] M.C. Pham, Y. Cao, R. Klamma, Clustering technique for collaborative filtering and the application to venue recommendation, in Proc. of I-KNOW, Citeseer, 2010.
[27] M.C. Pham, Y. Cao, R. Klamma, M. Jarke, A clustering approach for collaborative filtering recommendation using social network analysis, J. UCS 17 (4) (2011) 583–604.
[28] C. Porcel, A. Ching-López, G. Lefranc, V. Loia, E. Herrera-Viedma, Sharing notes: an academic social network based on a personalized fuzzy linguistic recommender system, Eng. Appl. Artif. Intell. 75 (2018) 1–10.
[29] T. Pradhan, A. Gupta, S. Pal, Hasvrec: a modularized hierarchical attention-based scholarly venue recommender system, Knowl.-Based Syst. (2020) 106181..
[30] T. Pradhan, S. Pal, A hybrid personalized scholarly venue recommender system integrating social network analysis and contextual similarity, Fut. Gen. Comput. Syst. (2019).
[31] T. Pradhan, S. Pal, Cnaver: a content and network-based academic venue recommender system, Knowl.-Based Syst. 189 (2020) 105092.
[32] T. Pradhan, S. Pal, A multi-level fusion based decision support system for academic collaborator recommendation, Knowl.-Based Syst. (2020b) 105784..
[33] J. Protasiewicz, W. Pedrycz, M. Kozlowski, S. Dadas, T. Stanislawek, A. Kopacz, M. Glke.zewska, A recommender system of reviewers and experts in reviewing problems, Knowl.-Based Syst. 106 (2016) 164–178..
[34] M.J. Schuemie, J.A. Kors, Jane: suggesting journals, finding experts, Bioinformatics 24 (5) (2008) 727–728.
[35] J. Serrano-Guerrero, F.P. Romero, J.A. Olivas, A relevance and quality-based ranking algorithm applied to evidence-based medicine, Comput. Methods Programs Biomed. (2020) 105415..
[36] T. Silva, J. Ma, C. Yang, H. Liang, A profile-boosted research analytics framework to recommend journals for manuscripts, J. Assoc. Inf. Sci. Technol. 66 (1) (2015) 180–200.
[37] K. Sugiyama, M.-Y. Kan, Towards higher relevance and serendipity in scholarly paper recommendation by kazunari sugiyama and min-yen kan with martin vesely as coordinator, ACM SIGWEB Newsletter, (Winter), 4, 2015..
[38] J. Tang, S. Wu, J. Sun, H. Su, Cross-domain collaboration recommendation, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, ACM, pp. 1285–1293..
[39] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, Z. Su, Arnetminer: extraction and mining of academic social networks, in: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008, ACM, pp. 990–998..
[40] Á. Tejeda-Lorente, C. Porcel, E. Peis, R. Sanz, E. Herrera-Viedma, A quality based recommender system to disseminate information in a university digital library, Inf. Sci. 261 (2014) 52–69.

[41] C. Wang, T. Zhou, C. Chen, T. Hu, G. Chen, Camo: a collaborative ranking method for content based recommendation, Proceedings of the AAAI Conference on Artificial Intelligence 33 (2019) 5224–5231.
[42] D. Wang, Y. Liang, D. Xu, X. Feng, R. Guan, A content-based recommender system for computer science publications, Knowl.-Based Syst. 157 (2018) 1–9.
[43] G. Wang, X. He, C.I. Ishuga, Har-si: a novel hybrid article recommendation approach integrating with social information in scientific social network, Knowl.-Based Syst. 148 (2018) 85–99.
[44] X. Wang, L. Yu, K. Ren, G. Tao, W. Zhang, Y. Yu, J. Wang, Dynamic attention deep model for article recommendation by learning human editors' demonstration, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, ACM, pp. 2051–2059..
[45] L. Yang, Y. Zheng, X. Cai, H. Dai, D. Mu, L. Guo, T. Dai, A lstm based model for personalized context-aware citation recommendation, IEEE Access 6 (2018) 59618–59627.
[46] Z. Yang, B.D. Davison, Distinguishing venues by writing styles, in: Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries, 2012a, ACM, pp. 371–372..
[47] Z. Yang, B.D. Davison, Venue recommendation: Submitting your paper with style, in: 2012 11th International Conference on Machine Learning and Applications (ICMLA), vol. 1, IEEE, pp. 681–686..
[48] Z. Yang, D. Yin, B.D. Davison, Recommendation in academia: a joint multi-relational model, in: 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2014, IEEE, pp. 566–571..
[49] J. Yu, K. Xie, H. Zhao, F. Liu, Prediction of user interest based on collaborative filtering for personalized academic recommendation, in: 2012 2nd International Conference on Computer Science and Network Technology (ICCSNT), 2012, IEEE, pp. 584–588..
[50] S. Yu, J. Liu, Z. Yang, Z. Chen, H. Jiang, A. Tolba, F. Xia, Pave: personalized academic venue recommendation exploiting co-publication networks, J. Network Comput. Appl. 104 (2018) 38–47.