



SPR-SMN: scientific paper recommendation employing SPECTER with memory network

Zafar Ali¹ · Guilin Qi¹ · Pavlos Kefalas² · Shah Khusro³ · Inayat Khan⁴ · Khan Muhammad⁵

Received: 8 August 2021 / Accepted: 30 May 2022 / Published online: 21 June 2022
© Akadémiai Kiadó, Budapest, Hungary 2022

Abstract

During the last decades, recommender systems are becoming quite popular since they provide great assistance to users on social networks and library websites. Unfortunately, the large volume of data combined with sparsity makes personalization a difficult task. In this regard, several models were introduced in the literature that suffers from the cold-start problem and the lack of personalization. In particular, the majority of these models ignore the relationship between the important factors and the semantic relations among the nodes (the authors, and the field of study) on the heterogeneous papers networks. Moreover, they fail to effectively capture researchers' preferences, which leads to inadequate recommendations. To overcome these problems, with this study we propose a scientific paper recommendation model called SPR-SMN, which employs the SPECTER document embedding model to learn context-preserving paper content representations. The model captures the long-range dependencies and researchers' preferences, by employing an end-to-end memory network and personalization module, respectively. We experimentally evaluate our method against baseline algorithms over two real-life datasets. The results indicate that the proposed method outperforms competing models.

Keywords Recommender systems · Paper recommendation · Memory network · SPECTER · Cold-start problem · Personalization

Introduction

The Web consists of millions of pages that contain information on almost every topic. This variety makes users strangled while searching for relevant information. To this end, numerous recommender systems that exploit the users' preferences have been proposed to support them with relevant information such as items, movies, locations, news, and other products (Amir et al., 2022; Christoforidis et al., 2018; Ma et al., 2020; Chen et al., 2021; Christoforidis et al., 2021a; Kefalas & Symeonidis, 2015). In that direction, a lot of work related to paper recommendations have been conducted (Cai et al., 2019; Bansal et al., 2016; Son &

✉ Zafar Ali
zafarali@seu.edu.cn; zafaralibagh6@gmail.com

Extended author information available on the last page of the article

Kim, 2017; Cai et al., 2018a) in the last five years. Based on the information filtering methods, researchers categorize these models into content-based (CB) (Bhagavatula et al., 2018; Zhao et al., 2016), collaborative filtering (CF) (Bansal et al., 2016; Wang & Li, 2015), and graph-based (GB) (Cai et al., 2019; Yang et al., 2018). The CB models employ the descriptions and features of papers and user profiles to produce recommendations. They generate useful recommendations when items' descriptions as well as user's past history and profile information are available, otherwise they are tackled by the cold-start problem (Christoforidis et al., 2018). The CF-based CR models exploit the past users' ratings along with the social network. The recommendations are robust if the users' ratings is available, otherwise they are tackled by sparsity which leads to imprecise predictions (Son & Kim, 2017). We can overcome this issue using GB models (Cai et al., 2019; Yang et al., 2018) that use additional relationships among nodes in the network. However, traditional GB models conceive recommendation like a link prediction task. Thus, these methods over-weight old and outdated nodes in the network (Son & Kim, 2017; Kefalas et al., 2018). To overcome the issues of traditional graph-based and CB methods, heterogeneous information networks for academic paper recommendation (HNPR) (Du et al., 2020) model exploited the author's collaboration, citation relations, and papers' research area to construct two types of heterogeneous information networks. The model employs a random walk-based strategy to conduct edge traversal in the constructed heterogeneous information networks and adopt natural language models to match word sequences for paper recommendation.

In recent years, different studies (Gupta & Varma, 2017; Cai et al., 2018a; Guo et al., 2019; Jiang et al., 2018) have employed homogeneous network representation learning (NRL) methods such as LINE (Tang et al., 2015; Christoforidis et al., 2021b) and Deepwalk (Perozzi et al., 2014) to overcome the issues of the traditional models. Nevertheless, these models cannot deal with the heterogeneity and multiplicity of citation networks. To overcome the heterogeneity issue several NRL-based models (Jiang et al., 2018; Cai et al., 2019; Kong et al., 2019) have been used in generating paper recommendations.

Albeit, the current NRL-based models (Jiang et al., 2018; Cai et al., 2018a, 2019; Gupta & Varma, 2017) resolved the heterogeneity problem in homogeneous embedding models, thus, they are unable to exploit the salient factors and relations corresponding to the heterogeneous information network objects. For instance, DBLP is a heterogeneous information network, where multiple relations exist among the network objects. These include relations among papers based on citation, authorship, the field of study, and so on. These relations establish a view that is indispensable to exploit by the NRL-based models. Nevertheless, existing NRL-based models are unable to consider such semantics and contextual information and therefore lack in capturing researchers' preferences and generating quality results (Chen et al., 2019). To learn content-based embeddings, existing studies employ Doc2vec (Le & Mikolov, 2014), BERT (Devlin et al., 2018), and SBERT (Reimers & Gurevych, 2019), language models. Yet, content-embedding approaches such as SBERT and BERT are trained on general English corpus, which in contrast to domain-specific embeddings models like SciBERT (Beltagy et al., 2019) and SPECTER (Cohan et al., 2020), cannot learn context-preserving node embeddings (Ali et al., 2021b). These models neglect to capture long-term dependencies and significance of salient factors and are therefore limited in making robust and justifiable recommendations.

To overcome these issues, we present a personalized paper recommendation model termed Scientific Paper Recommendation by employing SPECTER with Memory Network (SPR-SMN), which exploits semantic relations between the Heterogeneous Paper Networks (HPNs) nodes, such as papers, authors, the field of study, and content. The model employs the SPECTER language model to learn content-based node representations. Moreover, an

end-to-end memory network is used to exploit the long-range dependencies and robust semantics. The main **contributions** of this study are the following:

- We present a novel **Paper Recommendation** model that employs an **End-to-end Memory Network** with SPECTER, which effectively exploits useful semantic relations and contextual information to generate robust paper recommendations.
- The proposed model utilizes personalized information regarding a user, such as an authorship, the field of study, paper content, and citation relations, to capture researchers' preferences and make personalized recommendations.
- We conduct exhaustive experimentation on two real-world datasets to examine the performance of the proposed model related to other state-of-the-art counterparts in terms of standard evaluation metrics including nDCG, MAP, and recall.

Related work

The models providing paper recommendations can be classified into three main categories, namely: CB, CF, and GB. In the following sections, we explain state-of-the-art models that belong to these three classes.

CF and CB based citation recommendation models

CF-based models exploit users' friends' opinions (explicit or implicit) to make paper recommendations (Wang & Blei, 2011; Wang & Li, 2015). To this end, Collaborative topic regression (CTR) (Wang & Blei, 2011) recommends papers to users by seamlessly integrating both feedback matrix and paper content information into a unified model. However, the CTR model faces the cold-start problem when the user-item rating matrix is sparse. PCTR (Wang & Li, 2015) resolves this issue by extending the CTR model by integrating the network structure information along with user-item feedback information using a principled hierarchical Bayesian model, thereby overcoming the cold-start problem faced by CTR. Bansal et al. (2016), introduced a model that generates CF-based paper predictions by employing a gated recurrent unit (GRU) network. The model utilizes user friends' ratings along with the content of articles to make paper recommendations. Both these models utilize content and auxiliary information to overcome the sparsity and cold-start problems faced by traditional CF-based models. Khadka et al. (2020) generated a high-level representation of a paper employing its topic information to produce citation recommendations. The research also contributes a new dataset consisting of the publication history of researchers and the content of scientific publications. McNee et al. (2002) exploited citation network between authors and papers employing four types of CF-based methods to make recommendations for a query user.

In contrast to CF, the CB models represent research papers and users by exploiting the content, features, and descriptions of the corresponding papers and users to produce recommendations (Salloum & Rajamanthri, 2021; Ali et al., 2021c). To this end, Bollacker et al. (1998) proposed CiteSeer, which is the first CB-based academic paper recommendation system that exploits TFIDF vector and citation relations. Likewise, a CB model (Amami et al., 2016) makes use of Latent Dirichlet Allocation (LDA) to the textual content of research papers to generate their latent representations. In particular, the model builds the representations of the researcher's profile (based on author's written papers) and

candidate papers using LDA. Then, the model computes similarities between these representations to make final recommendations. Similarly, Science Concierge (Achakulvisut et al., 2016) employed Latent Semantic Analysis (LSA) to the content of manuscripts to provide recommendations.

The CB and CF-based models discussed above can assist users in tailoring personalized recommendations. Nevertheless, CF-based models encounter different problems, especially cold-start and sparsity. That is, when there is limited user rating and user profile information then CF models find it troublesome to produce justifiable results. This way, the recommendations delivered on such insufficient information can lead to inaccurate results (Son & Kim, 2017). In contrast, CB models require papers and user descriptions/features, and if such information is unavailable, they end up with the cold-start and overspecialization problems (Khusro et al., 2016). Besides, CB and CF models do not utilize auxiliary information sources and semantic relations in the HPNs. Therefore, these models fail to capture meaningful semantics and generate relevant recommendations.

Deep learning and graph-based citation recommendation models

During the last decade, several DL-based paper recommendation models employed multilayer perceptrons (MLPs) (Huang et al., 2015), convolutional neural networks (CNNs) (Yin & Li, 2017), recurrent neural networks (RNNs) (Bansal et al., 2016; Yang et al., 2018; Uddin et al., 2022), and generative adversarial networks (GANs) (Cai et al., 2018a) to produce quality recommendations (Ali et al., 2020a). For instance, Huang et al. (2015) generated recommendations for a citation context by using the semantic representations of citation contexts and relevant papers. This model utilized a multi-layer neural network to learn the probability of citing an article given the citation context. In the same direction, PCCR (Yang et al., 2018) adopted LSTM (Abro et al., 2020) to learn the representations of citation contexts and research manuscripts employing the context and paper encoders, respectively. Then, the model finds top-*k* citations for a given context using cosine similarity between the embeddings of corresponding context and candidate papers. Similarly, a personalized model (Wang et al., 2020a) utilized authors information and citations information using a BiGRU network to make context-aware recommendations. On the other hand, p-CNN (Yin & Li, 2017), a personalized citation recommendation model generates recommendations using CNN. In particular, it exploits the information about authors to compute relevance between citation context and relevant paper. Moreover, the model employs a discriminative training strategy to learn the parameters and generate relevant recommendations. Similarly, POLAR (Du et al., 2019) proposed an attention-based CNN model to produce paper recommendations. To capture salient factors and words, the model employs an attention matrix that captures both local and global weights.

Recently, sophisticated graph (Goyal & Ferrara, 2018) and network representation (Cui et al., 2019) techniques exploit semantic relations between the nodes of the graph or a network to learn their vector representations of corresponding nodes. Various paper recommendation models (Gupta & Varma, 2017; Cai et al., 2019; Kong et al., 2019; Ali et al., 2021a; Du et al., 2020) employed such embedding methods to make recommendations. To this end, Gupta and Varma (2017) employed Doc2vec (Le & Mikolov, 2014) and DeepWalk to learn the embeddings of papers' content and network structure, respectively. Then, the model exploits similarities between the learned representations to produce paper recommendations. Likewise, VOPRec (Kong et al., 2019) generated recommendations by integrating text-based vector representations and structured-based embedding employing

the Paper2vec (Ganguly and Pudi, 2017) and Struct2vec (Ribeiro et al., 2017) embedding methods, respectively. In contrast, BNR (Cai et al., 2019) and CR-HBNE (Ali et al., 2021a) employed Node2vec (Grover & Leskovec, 2016) to exploit the semantic relations between the objects of heterogeneous bibliographic network and learn the embeddings of participating nodes (i.e., papers, authors, content, venues, etc.). Finally, the learned node embeddings are utilized to make the final recommendations for a query manuscript. More recently, Dai et al. (2021) proposed the GRSLA model, which exploits author information by introducing a novel author embedding method. The model makes use of an encoder-decoder architecture using three neural networks to alleviate the extendibility issue of author embedding vector faced by existing global citation recommendation models (Son & Kim, 2017; Cai et al., 2018b; Dai et al., 2019). In the same direction, Zhang and Zhu (2021) studied citation recommendation problem from the perspective of semantic representation of cited papers’ relations and content. The study designs 132 methods by integrating different NRL-based methods with text representation learning methods and generates the vector representations of research papers and then cosine similarity between these representations is employed to deliver relevant recommendations.

Existing network representation learning-based recommendation models, such as VOPRec, GAN-HBNE (Cai et al., 2018b), GRSLA, and BNR gain superior results compared to the random-walk, and traditional CB and CF models. Nevertheless, these models are limited to exploiting the significance of semantic relations in HPNs and dealing with the “cold-start papers” problems. Furthermore, these models cannot effectively utilize contextual information and salient factors to capture the researcher’s preferences. A detailed survey related to the classification of paper recommender systems exploring information filtering methods, information sources exploited such as keywords, title, citation, user profile, etc., evaluation measures, and open challenges can be found in Kreutz and Schenkel (2022).

Preliminaries and problem definition

This section illustrates the preliminaries required for the proposed model. For simplicity, you may find all the symbols used in the rest of the paper at Table 1.

Definition 1 (*Heterogeneous Papers Network*). The network $G = (N, E)$ is a variant of Heterogeneous Information Network (HIN) (Ali et al., 2020c) with two mapping functions, viz., node type mapping $\varphi : N \rightarrow O$ and relation type mapping $\varphi : E \rightarrow R$. Here, each node $v \in N$ and edge $e \in E$ belong to a particular node type and relation type, respectively. Moreover, $N = A^s \cup P^s \cup F^s$, with $A = \{a_1, a_2, \dots, a_n\}$, $P = \{p_1, p_2, \dots, p_n\}$, $F = \{f_1, f_2, \dots, f_n\}$ representing sets of authors, papers, and field of study, respectively. Additionally, $E = \cup_{r \in R} E_r$ represents the edges, where E_r is the set of edges linked with relation $r \in R > 1$. In an HPN, we have $|O| + |R| > 2$.

Example 1 : The HPN in this work contains three objects, viz., papers, authors, and field of study (FOS) that establish relations with each other. In particular, the authorship relation network G_{ap} is established between authors and papers when an author writes a paper. For instance, if author a_i writes paper p_j , then the author-paper adjacency matrix $A^s \in \mathfrak{R}^{|A| \times |P|}$ will have value $e_{a_i, p_j} = 1$ to represent such relation. To capture researchers’ preferences, the proposed model employs paper-to-paper relation network G_{pp} and FOS of the paper

Table 1 Symbols and their details used in this research

Symbol	Details
G	The heterogeneous information network
N	The total number of nodes in the G
E	The total number of edges in the G
P	Set of papers $\{p_1, p_2, \dots, p_n\}$
A	Set of authors $\{a_1, a_2, \dots, a_n\}$
F	Set of papers field of study $\{f_1, f_2, \dots, f_n\}$
M	Set of memory slots $\{m_1, m_2, \dots, m_n\}$
$e_{i,j}$	An edge between two objects i and j
s_i	Relevance score between query and i -th paper
q	The query paper
e_q	Vector representation of query paper
c_i^p	Content-based embedding of paper p
ϕ_i^p	Context-aware personalized embedding of p
ϕ_i^q	Context-aware personalized embedding of q

$G_{pf} \in \mathfrak{R}^{|P| \times |F|}$ network. That is, if paper p_i establishes citation link with p_j , then the corresponding entry in the $P^s \in \mathfrak{R}^{|P| \times |P|}$ matrix has $e_{p_i, p_j} = 1$, otherwise 0. Likewise, we demonstrate relations between papers and their FOS, if there exists any relationship.

Problem Statement: Given a seed paper q along with the HPN $G = (V, E)$, the proposed model aims to exploit the semantics in the HPNs and recommend top@ k relevant papers for q .

Hypothesis: Using auxiliary information sources and structural modules incorporated in the HPNs will improve recommendation results and over-come the cold-start paper problem.

SPR-SMN model

Figure 1 presents the architecture of the proposed model that has a three-step process. First, it learns embeddings of the contents of papers by employing the pre-trained SPECTER (Cohan et al., 2020) model. Next, it exploits researchers' preferences by using authors' information, the FOS, and citation relations of the papers. Finally, it uses an end-to-end memory network with an attention mechanism to capture long-range dependencies and give weight to the significant information. The constituent modules responsible for these steps are discussed in detail in the following sections.

Content-based paper embedding

In this section, we present how the model employs the SPECTER (Cohan et al., 2020) document embedding model to learn content-based representations of scientific papers. To learn semantic-aware embedding, the SPECTER (Cohan et al., 2020) employs Citation-aware Transformers. General-purpose documents embedding models, i.e., BERT (Devlin et al., 2018) and SBERT (Reimers & Gurevych, 2019), can produce

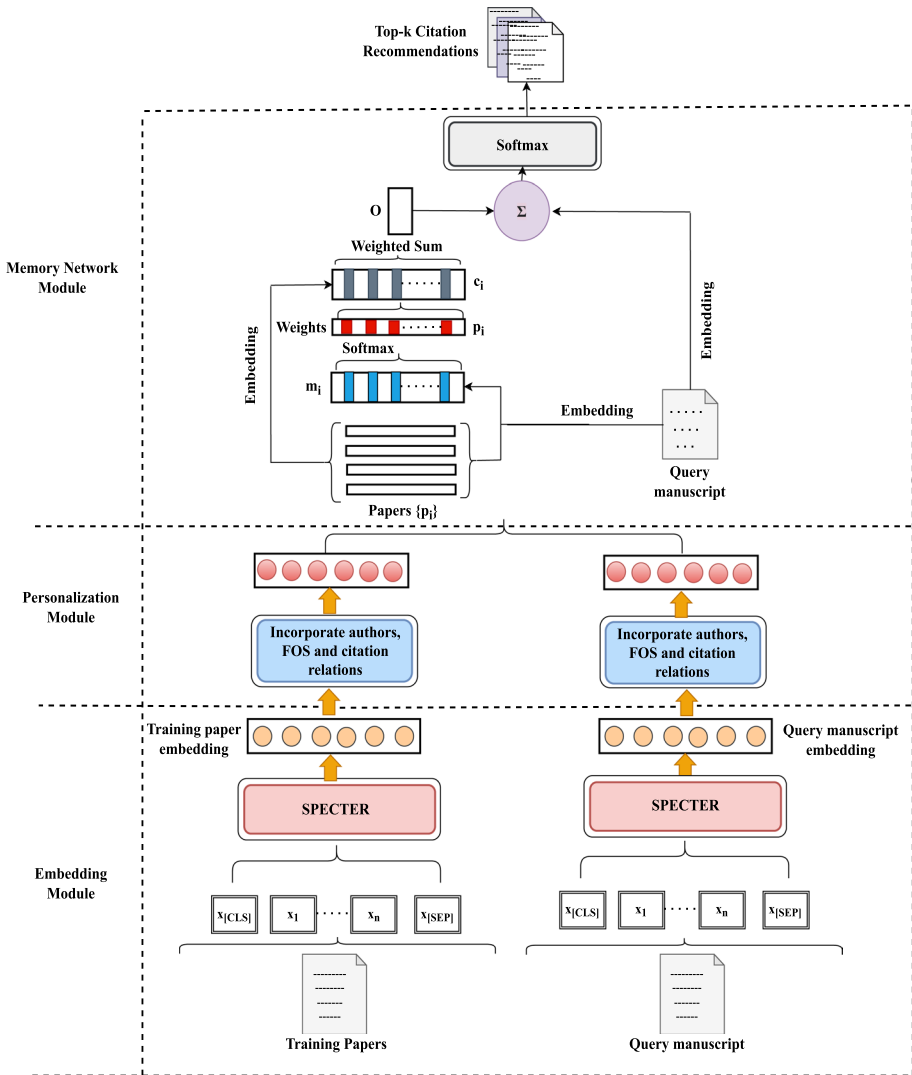


Fig. 1 The diagram of the SPR-SMN model

context-aware documents representations (Wang et al., 2020b; Gao et al., 2019). Nevertheless, traditional language models like Doc2vec and SBERT were trained on Wikipedia corpus, therefore they do not capture more relevant contextual information corresponding to scientific documents. Also, these models do not consider relatedness between documents established based on citation relations while generating document representations. To come up with a better solution, the SPECTER tunes the embeddings learned through SciBERT (Beltagy et al., 2019), which employs a corpus of scientific documents during training. To learn representations of a paper c_i^p , the SPECTER initially encodes the concatenated text (i.e., the abstract and title) of the paper utilizing Transformer LM (which is SciBERT), defined below.

$$c_i^p = \text{Transf}(\text{input})_{[CLS]}, \quad (1)$$

where Transf represents the forward function of Transformer. The model takes the concatenated WordPieces (of title and abstract) and $[CLS]$ token as an input, which is separated using the $[SEP]$ token. In order to enrich the embeddings learned using SciBERT, the SPECTER employs citation relations between documents as a relatedness signal to enrich the vector representations of documents learned through SciBERT. Besides, the SPECTER makes use of "hard negatives" along with "simple negatives" to learn more optimal and context-preserving embeddings. The SPECTER model learns nodes content-based embeddings by optimizing the following margin loss objective 2.

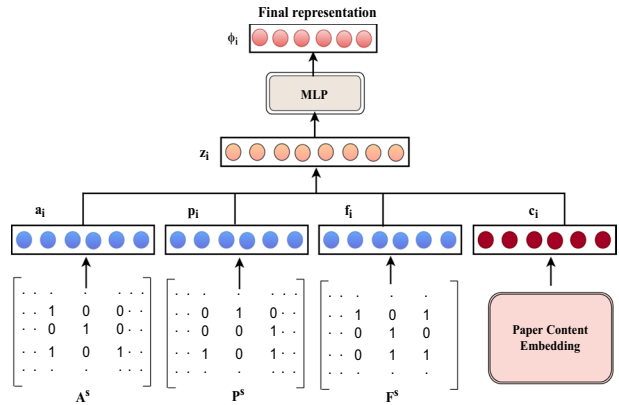
$$T_L = \max \{ (d(P^m - P^+) - d(P^q - P^-) + w), 0 \}, \quad (2)$$

where P^m is used for query paper, P^+ denotes the relevant paper, while P^- represents the irrelevant paper. Additionally, d denotes the Euclidean norm distance, and w represents the margin which is equal to 1. The model makes use of w to make sure that the value of P^+ is at least w closer to P^m as compared to the P^- . During training, the model aims to minimize the distance between query paper and related paper, while maximizing the distance between query paper and irrelevant paper. At inference time, for an input paper c_i , the model learn content-based paper embedding c_i^p by taking the SPECTER's Transformer pooled output activation. This way, the model captures the contextual information of the paper.

Personalization module

In paper recommendation models, researchers' preferences regarding authors, citations and paper field of study play a prominent role in producing individualized paper recommendations. For instance, the author(s) of a paper can have a great influence on the readers, and citations. Mostly, a researcher follows a particular researcher or a research group/s with similar research preferences. Likewise, an author who does collaboration with another researcher is more important, compared to the one who has no collaboration with different research interests (Ali et al., 2020c). Besides, those papers which are linked based on citation relation are considered to be more related. Also, researchers cite papers very carefully. Therefore, such relations has a great impact on personalizing the recommendations for a researcher. In particular, The probability of citing a paper already cited by an author in their previous paper is higher compared to any other random paper. Similarly, a researcher likes studying a research paper that targets the same field of study and authors with matching interests. Also, the study (Ali et al., 2020b) reveals that the most popular feature in citation recommendation models is field of study since it correlates papers based on similar keywords. Thus, each paper is marked with multiple tags that give a short description of its contents. To exploit such useful relations, the proposed model employs a personalization module, depicted in Fig. 2. This module exploits authors information, citation relations, and paper field of study. In this regard, the model first establishes relation matrix between papers, authors and FOS to capture researchers preferences. The author-paper relation matrix $L^a \in \mathfrak{R}^{|A| \times |D|}$ is maintained between authors and their papers. Similarly, the model establishes paper-citation relation matrix $P^s \in \mathfrak{R}^{|P| \times |P|}$. The paper-FOS relation matrix $F^s \in \mathfrak{R}^{|P| \times |F|}$ is created between papers and field-of-study. For instance, if a paper p_i belongs to a FOS f_j , the corresponding cell in the matrix gets $e_{p_i f_j} = 1$ and 0 otherwise.

Fig. 2 An illustration of personalized embedding module



Once we have these relation matrices, the model combines the i -th columns, i.e., $l_i^s \in L^s$, $p_i^s \in P^s$, and $f_i^s \in F^s$ of adjacency matrices with the content-based embedding c_i^p and passes through a non-linear layer as follows.

$$z_i^p = Rel(W_c^P c_i^p + W_u^P u_i^p + W_p^P p_i^p + W_f^P f_i^p + b^P) \tag{3}$$

Here, $W_c^P \in \mathfrak{R}^{p_p \times p_h}$, $W_u^P \in \mathfrak{R}^{p_p \times |A|}$, $W_p^P \in \mathfrak{R}^{p_p \times |P|}$, and $W_f^P \in \mathfrak{R}^{p_p \times |F|}$ denote the learnable weight matrices, and b^p represents the bias. Next, the model employs multi-layer perception to generate the final embedding vector ϕ_i^p of a paper, as defined below.

$$\phi_i^p = MLP(z_i^p), \tag{4}$$

where MLP represents a multi-layer perceptron which can be computed using the following equation.

$$MLP(y) = L_2[ReLU(W_i y + b_i)], \tag{5}$$

where $ReLU$ is used for non-linearity. To capture researchers preferences and produce personalized results, It is intuitive to exploit authors information, citation relations and paper’s field of study. Thus, to learn robust representations of training papers, SPR-SMN employs the aforementioned relations between heterogeneous papers networks (HPNs). Likewise, the model adopts this process to generate the context-aware representation of the query paper ϕ_i^q .

Memory network

The proposed model employs an end-to-end memory network (Sukhbaatar et al., 2015) module to exploit long-term contextual information and salient factors. The output of the personalization module is feed as an input to this module. The working of the memory network module is illustrated graphically as a final module in Fig. 1. To produce final recommendations for a query paper, the content-based paper representations ϕ_i^p are converted into memory slots $M_s = \mathfrak{R}^{p_m \times P}$, where P denotes the total number of papers. The model embeds query paper q into internal state vector in the controller and computes relevance between query paper and

memory slots by taking the inner product between them followed by a *tanh* activation, defined as given below.

$$s_i = \tan h(W_{am}[m_i : e_q] + b_{am}), \quad (6)$$

where s_i denotes the scoring function between query paper and relevant training papers and W_{am} represents weight metrics. The model then computes the attention weights for a query paper by giving s_i into the softmax function defined as $\alpha_i = \text{softmax}(s_i)$. The model employs an attention mechanism since it learns an adaptive weighting function, which assigns significance/weight to participating memory slots. The model computes the final output memory representations by obtaining a weighted sum over paper embedding vectors $C \in \mathfrak{R}^{p_m \times N}$, as computed below.

$$o = \sum_{i=1}^N \alpha_i c_i, \quad (7)$$

where o denotes weighted sum based on the relationships between training papers and query paper. If a single layer *MemN2N* is employed, then the output embedding o and the input vector e_q are summed and provided as an input into a final weight matrix followed by a *softmax* function to make predictions. Nevertheless, models that employ a single attention mechanism lack of capturing comprehensive semantics and context-aware representations of data (Sukhbaatar et al., 2015). Thus, the SPR-BMN uses multiple processing layers. In particular, to make predictions for a query paper, with the $k + 1 - th$ modeling layer/hop, the model adds the output o^k and the input e_q^k followed by *softmax* as follows.

$$\hat{a} = \text{softmax}[W(o^k + e_q^k)], \quad (8)$$

where $W \in \mathfrak{R}^{N \times p}$ denotes the weight matrix. The proposed model used three hops and selected top-10 recommendations for each query paper.

Model training

During training, the proposed model minimizes the cross-entropy loss computed between the prediction of the model, i.e., g , and the accurate result. Thus, the proposed model optimizes the following objective.

$$O = - \sum_{q \in C} \sum_{i \in P} P_q^{(i)} \log[P_q^{(i)}(g)], \quad (9)$$

where C , and $|P|$ denote training examples and set of research papers, respectively. If the proposed model recommends ground truth, i.e., p^i for the query paper q in top- k recommendations, then it is treated as an accurate result and we have $P_q^{(i)} = 1$, otherwise 0. In addition, the model utilizes stochastic gradient descent, which learns the parameters of the model based on the backpropagation method. To overcome possible overfitting, the model employs dropout. For better configuration, we experimented with different dropout rates.

Table 2 Datasets specifications

Dataset	Paper	Author	Venue	FOS	Citations
DBLP-V12	3,501,133	245,204	16,209	11,135	25,022,314
AAN	21,450	17,335	311	1755	113355

Experimental study

In this section, we present the evaluation protocol, the datasets, and the models used for comparison.

Datasets

To assess the results generated by models, we employed two datasets, viz., the DBLP-V12 and the ACL Anthology (AAN). Further details and statistics of the datasets are tabulated in Table 2.

DBLP The DBLP-Citation-network V12¹ has a relatively large size among the DBLP datasets with 3,501,133 research papers and 25,022,314 citations. The information it provides include papers’ titles, abstracts, authors, venues, keywords, and citation relations.

ACL Anthology Network (AAN) AAN dataset² is comparatively latest and holds papers related to computational linguistics and NLP. It holds research articles = 21,450, authors = 17,335, venues = 311, and citation relations = 113355.

Datasets train-test split

To conduct experiments, we split each dataset randomly into two sets called the training Y^t and test set Y^p . Training set consists of 80% of the papers, while test set Y^p possesses the remaining 20% papers. Additionally, $Y = Y^t \cup Y^p$ and $Y^t \cap Y^p = \emptyset$. For a seed manuscript, the model provides top@k paper recommendations using the Y^t . If the ground truth is recommended in the top@k, then the result is considered as relevant, otherwise not.

Metrics for evaluation

We employed recall, Mean Average Precision (MAP), and normalized Discounted Cumulative Gain (nDCG) as evaluation metrics, which are the most commonly used metrics in the relevant domain (Ali et al., 2020b; Kefalas & Manolopoulos, 2017; Kefalas et al., 2018).

Recall: examines the recommendations of models using the percentage of related results delivered in the top-k list of recommendations, where $k = \{20, 40, 60, 80, 100\}$.

$$Recall = \frac{1}{Q} \sum_{j=1}^Q \frac{R_p \cap T_p}{T_p}, \tag{10}$$

¹ <https://www.aminer.cn/citation>.

² <https://acl-arc.comp.nus.edu.sg/>.

where Q represents all target research papers, R_p denotes the list of *top-k* recommendations delivered for the seed paper p .

MAP: examines the significance of a model by examining whether the relevant articles are suggested in *top-k* or not. Additionally, it penalizes those errors that happen high up in the *top@k*.

$$AP@k = \frac{1}{GTP} \sum_{i=1}^k \frac{TP_{seen}}{i}, \quad (11)$$

where TP_{seen} denotes total true positives occurred till k . We select the cut-off value of AP as choose the cut-off value for the Average Precision (AP) as $AP@10$.

nDCG: evaluates the rank in the *top-k* of the true relevant papers suggested by the model (Ali et al., 2020c) and is computed as:

$$nDCG_g = \frac{DCG_g}{IDCG_g}, \quad (12)$$

where DCG is the weighted sum of the degree of relatedness of the ranked manuscript. $IDCG_g$ shows the DCG of ideal ordering, using which the DCG values are normalized.

Models used in the experiments

This section provides details about the models that are used as baselines for the proposed model. The following subsections explain these models.

- CCA (Gupta & Varma, 2017) learns the low-dimensional representations using the content and network proximity. To learn content and network embeddings, it uses Doc2Vec and DeepWalk embedding methods, respectively. Finally, the similarities between the learned vectors are computed to make relevant recommendations. We set the dimensions of DeepWalk to 64 and 300 for the Doc2Vec method.
- BNR (Cai et al., 2019) is an NRL method that explores network proximity and relevant papers' content to provide recommendations against a seed paper. To conduct fair experiments, the setting of the parameters is adjusted as follows: dimensions = 128, context size = 10, and walks per vertex = 80. Also, the tuning parameter β has great importance in the model performance and it gives the best results when it is set to $\beta = 0.7$. Also, we set the in-out parameter $p = 1$ and return parameter $q = 2$.
- SCR-NTR (Qiu et al., 2021) is a citation recommendation model that exploits network and textual information to generate recommendations. To learn text and network representations, the model utilizes BERT and HeGAN representation learning methods, respectively. The parameter setting for the text representation is as follows: Learning rate: 0.0001, Learning epoch: 5, the number of bidirectional transformer layers: 12, the size of the hidden layer: 768, the number of attention heads: 12, and the number of total parameters: 110 M. For network representation, the number of discriminator training per epoch and generator training per epoch are set to 15 each.
- NNRank (Bhagavatula et al., 2018) creates embeddings of nodes employing a neural network and provides *top-k* results for a query paper using the cosine between the embeddings learned for the corresponding nodes in the network. We utilize the learning rate of 0.001, batch-size = 512, abstract length = 500, nearest neighbors to 5, while the dense dimensions have a value of 75.

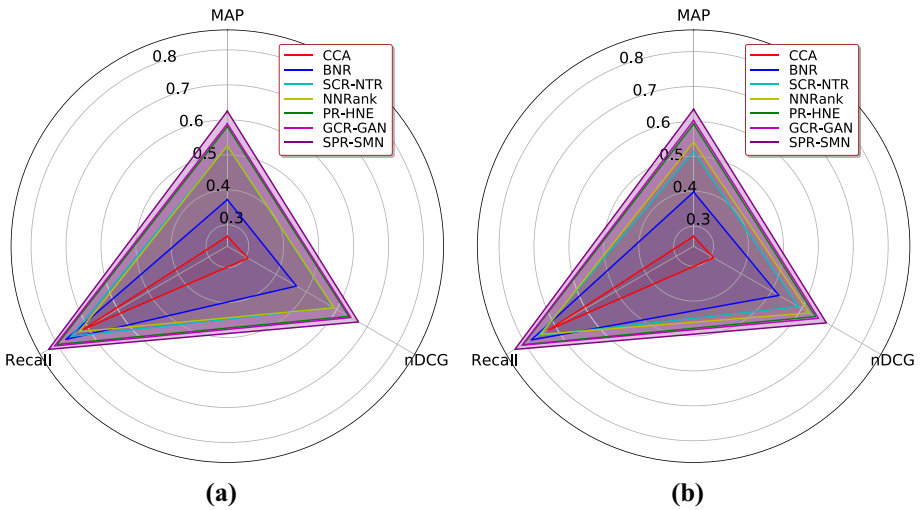


Fig. 3 Radar chart showing the MAP, nDCG, and Recall scores based on the **a** DBLP-V12, and **b** AAN datasets

- PR-HNE (Ali et al., 2020c) is a heterogeneous network embedding model that employs multiple relation networks to generate recommendations for a query manuscript. The model employs two proximity concepts to capture semantic relations between network objects and learn their representations. Finally, the dot product between these embedding is computed to rank top-k papers for a query manuscript. All the parameters are set according to the experimental setup of the PR-HNE model.
- GCR-GAN (Ali et al., 2021b) is a global citation recommendation model that exploits network structure and relevant content using generative adversarial network. Like the proposed model, GCR-GAN learns papers content-based embedding using the SPECTER document embedding model. However, the network-based representations are learned using the GAN network. On the other hand, the proposed model employs a personalization module and end-to-end memory network to capture user preferences and long-range contextual information, respectively. In addition, the proposed model exploits the paper field of study to exploit more semantics in the network. In this set of experiments, we used the default setting of parameters.

Comparative analysis with other baselines

This section analyzes the results generated by the proposed model compared to the baselines. In particular, we judge the recommendations produced by different models on the DBLP, and AAN datasets using the evaluation metrics, namely recall, MAP, and nDCG, as depicted in Fig. 3. Table 3 and Fig. 4 reveal that SPR-SMN outperform its counterparts on DBLP and AAN datasets, respectively. The results exhibit that CCA has generated the most insignificant results. The reason behind the poorer performance is its inability to exploit the structure of the heterogeneous bibliographic network, author information, and paper field of study. Moreover, the model utilizes doc2vec to learn textual representations, which in contrast to SPECTER cannot capture robust contextual information and therefore fails to learn semantic-preserving paper content embeddings.

Table 3 Results reported using the DBLP-V12, where **bold** indicates the best and—represents the second best performer

Model	MAP	nDCG	R@20	R@40	R@60	R@80	R@100
CCA	0.267	0.307	0.520	0.579	0.636	0.672	0.715
BNR	0.372	0.467	0.575	0.673	0.714	0.755	0.772
SCR-NTR	0.525	0.589	0.577	0.662	0.705	0.739	0.755
PR-HNE	0.581	0.640	0.589	0.656	0.721	0.764	0.801
NNRank	0.526	0.591	0.548	0.635	0.689	0.707	0.728
GCR-GAN	0.589 [−]	0.647 [−]	0.602 [−]	0.663 [−]	0.735 [−]	0.772 [−]	0.809 [−]
SPR-SMN	0.625	0.672	0.607	0.686	0.751	0.804	0.829

BNR on the other hand achieves better results compared to CCA as it exploits heterogeneous information sources to learn nodes embedding and produce citation recommendations. Nevertheless, the model generates ineffective vectors since it employs the DeepWalk method to exploit shallow node representation. SCR-NTR gains improvement in terms of nDCG and MAP scores as compared to the BNR model. The reason is that it employs more sophisticated network and text representation learning methods, viz., HeGAN and BERT, respectively to exploit semantic relations in the network. Yet, the model is limited in terms of using the BERT model, which in contrast to the domain-specific SPECTER model fails to produce context-aware content embedding since it is pretrained on general English Wikipedia corpus and does not consider hard negatives. Besides, the model is limited in terms of exploiting personalized information i.e., the paper field of study to generate individualized recommendations.

It is also notable that GCR-GAN returned the second-best results because of its ability to consider the network structure along with author and auxiliary information sources to make personalized results. However, the author vector in GCR-GAN is obtained using the author adjacency matrix, which contains only linked neighbors' information. Also, when a new author is introduced to the model, it requires to expand the dimension of the author's adjacency matrix and retrain the whole model, which leads to extendability problems. Therefore, GCR-GAN has been outperformed along with other competitors by the SPR-SMN in terms of nDCG, MAP, and Recall metrics, which demonstrates that the proposed model is comparatively efficient in producing better-ranked paper recommendations.

Also, we notice that the SPR-SMN has gained almost 4% and 3% improvement in terms of MAP and nDCG scores against GCR-GAN on the DBLP-V12 dataset. Considering the recall results (i.e., Rec@20, Rec@40, Rec@60, Rec@80, and Rec@100), SPR-SMN obtains better results over other baselines, which exhibits the stability and robustness in the results of the proposed model. Similarly, the proposed model has gained nearly 4% and 5% better MAP and nDCG scores related to the second-best model, i.e., GCR-GAN, on the AAN dataset. This is attributed to the fact that SPR-SMN exploits semantic relations and contextual information corresponding to research papers and authors, which helps the model exploit the researcher's preferences and produce quality recommendations. The reason for the SPR-SMN's superior results is its application of semantics employing the SPECTER model, which during training employs domain-specific corpus and utilizes citation-informed transformers to produce semantic-aware embedding of papers. Further, the personalization module exploits researchers' preferences, viz., author's information, relevant content, the paper field of study, and citation relations, which boosts the results.

Table 4 Results reported employing the AAN dataset, where **bold** indicates the best and—represents the second best performance

Model	MAP	nDCG	R@20	R@40	R@60	R@80	R@100
CCA	0.271	0.309	0.553	0.651	0.691	0.714	0.723
BNR	0.398	0.525	0.581	0.672	0.721	0.767	0.780
SCR-NTR	0.534	0.593	0.655	0.689	0.713	0.744	0.766
PR-HNE	0.593	0.646	0.604	0.681	0.740	0.781	0.807
NNRank	0.541	0.625	0.553	0.654	0.695	0.724	0.748
GCR-GAN	0.603 ⁻	0.656 ⁻	0.614 ⁻	0.681 ⁻	0.750 ⁻	0.784 ⁻	0.810 ⁻
SPR-SMN	0.647	0.694	0.621	0.692	0.756	0.813	0.841

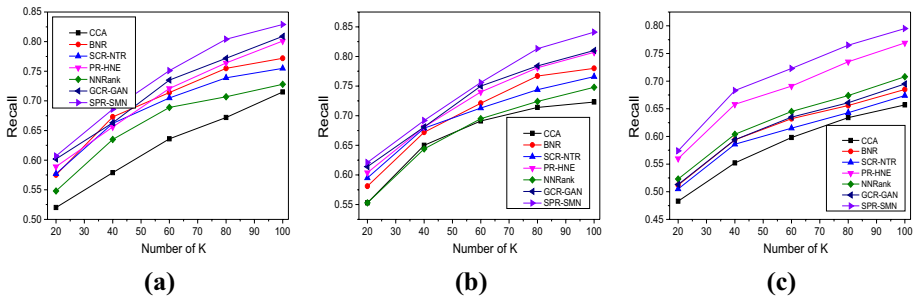


Fig. 4 Comparison of recommendations models regarding the recall scores **a** on the DBLP-V12 dataset, **b** on the ACL anthology dataset, and **c** with respect to the cold-start papers

Table 5 The influence of integrating different modules on the recommendation results, where **bold** results indicate the best model

Dataset	Model	MAP	nDCG	R@20	R@40	R@60	R@80	R@100
DBLP-V12	SPR-SMN_E	0.535	0.578	0.531	0.609	0.648	0.682	0.722
	SPR-SMN_{E+P}	0.593	0.645	0.585	0.657	0.701	0.732	0.763
	SPR-SMN_{E+P+M}	0.625	0.672	0.607	0.686	0.751	0.804	0.829
ACL Anthology	SPR-SMN_E	0.558	0.598	0.551	0.619	0.662	0.708	0.739
	SPR-SMN_{E+P}	0.598	0.651	0.592	0.669	0.704	0.741	0.768
	SPR-SMN_{E+P+M}	0.647	0.694	0.621	0.692	0.756	0.813	0.841

Additionally, the model stacks the memory layers to enhance the learning ability and grasp long-range dependencies of the significant factors (Table 4).

Ablation study

In this section, we analyze the impact of each module in the proposed model, in terms of MAP, nDCG, and recall scores. In particular, we judge the influence of integrating content embedding module **SPR-SMN_E**, personalization module **SPR-SMN_{E+P}** and memory network module **SPR-SMN_{E+P+M}**, over the two datasets. The results presented

Table 6 The influence of utilizing various relation networks on the recommendation results, where **bold** results indicate the best model

Dataset	Model	MAP	nDCG	R@20	R@40	R@60	R@80	R@100
DBLP-V12	SPR-SMN _{NP}	0.543	0.585	0.536	0.613	0.654	0.692	0.728
	SPR-SMN _A	0.579	0.602	0.597	0.663	0.704	0.731	0.753
	SPR-SMN _{AC}	0.596	0.645	0.601	0.677	0.724	0.752	0.772
	SPR-SMN _{ACF}	0.625	0.672	0.607	0.686	0.751	0.804	0.829
ACL Anthology	SPR-SMN _{NP}	0.564	0.603	0.558	0.626	0.669	0.715	0.747
	SPR-SMN _A	0.598	0.637	0.562	0.669	0.704	0.741	0.768
	SPR-SMN _{AC}	0.626	0.661	0.579	0.683	0.724	0.765	0.798
	SPR-SMN _{ACF}	0.647	0.694	0.621	0.692	0.756	0.813	0.841

in Table 5, clearly indicate that the SPR-SMN_E produced relatively insignificant results compared to other variants since this variant employs content information and does not capture personalized information and long-range dependencies. On the contrary, SPR-SMN_{E+P} gained a significant improvement over the previous variant since it exploits personalized information including author, the paper field of study, and citation relations. Finally, all components combined including the memory network denoted as SPR-SMN_{E+P+M}, further improve models' accuracy. This is due to the integration of an end-to-end memory network with an attention mechanism, which helps capture the long-range dependencies and exploit salient factors.

Impact of integrating information networks

We analyze the influence of the participating relation networks in the personalized embedding module. The results of this study are tabulated in Table 6. In particular, we examined the significance of incorporating authors' information, paper citation relations, and paper FOS. To do so, we compared the results of different variants of the proposed model. These include SPR-SMN_{NP}, which employs no relation networks, namely authors, citation, and FOS, except papers' content to generate recommendations. SPR-SMN_A extends previous version by adding author relation to enhance results. SPR-SMN_{AC} incorporates citation relations along with authors information. SPR-SMN_{ACF} integrates all relations networks including field of study.

These results demonstrate that SPR-SMN_{NP} produced relatively insignificant results compared to other variants since it employs none of the personalized author information, citation relationships, and field of study. Also, it can be noticed that exploiting author's relations in the SPR-SMN_A has a great impact on the relevance of the results. On the contrary, the field of study has comparatively less impact on the final recommendations. The results of final variant SPR-SMN_{ACF} exhibits that citation relation is the second influential relation exploited. To conclude, these findings suggest that when we integrate various information networks, the proposed model gains better results. Exploiting these relations help the model capture researchers' preferences to produce more individualized results.

Table 7 Analysis of results using cold-start papers, where **bold** results reveal the best and—results show the runner-up model

Model	MAP	nDCG	R@20	R@40	R@60	R@80	R@100
CCA	0.247	0.284	0.483	0.552	0.598	0.634	0.657
BNR	0.326	0.391	0.512	0.594	0.632	0.656	0.685
SCR-NTR	0.431	0.497	0.505	0.586	0.615	0.643	0.674
PR-HNE	0.489	0.557	0.560 [–]	0.658 [–]	0.691 [–]	0.735 [–]	0.769 [–]
NNiRank	0.465	0.551	0.523	0.604	0.645	0.674	0.708
GCR-GAN	0.503 [–]	0.563 [–]	0.513	0.594	0.635	0.661	0.695
SPR-SMN	0.574	0.657	0.574	0.683	0.724	0.765	0.795

Performance over cold-start papers

The problem of ‘cold-start papers’ occurs because of the unavailability of information about papers to recommend them to the users. The information that is unavailable regarding cold-start papers includes paper content, citation relations, the field of study, and the author’s information. To analyze the performance of the SPR-SMN, we selected 20,454 cold-start papers and used them in producing recommendations. Table 7 represents that even if we have missing information, the model can utilize auxiliary information sources to produce useful recommendations. In particular, if a paper has no citation information, the proposed model employs its field of study, content, and author’s information to make recommendations. We notice that our model gained 7% and 2% better MAP and Rec@100 results compared to the second best-performing model, i.e., GCR-GAN. This significant gain demonstrates that the context-preserving content-based embedding learned using the SPECTER model and exploiting personalized information helped the proposed model produce improved results. On the other hand, GCR-GAN and PR-HNE perform competitively with each other.

Impact of parameters

This section discusses the setting of different parameters employed by the SPR-SMN and their impact on the resulting recommendations. Besides, we use the SPECTER model to learn content-based representations of papers. SPECTER utilizes a 768-dimensional embedding for nodes. Also, we set the loss margin parameter to $w = 1$. For training, SPECTER employs five negative samples containing two hard negatives and three easy negatives. For the batch size, we choose value 16. To learn papers’ content embedding, we used the abstracts of research papers with a size of 517. To find the best learning rate, we used the grid search over a typical range of hyperparameters. To analyze the impact of the learning rate, we presented the MAP, nDCG, and Rec@20 results based on the DBLP and AAN datasets as depicted in Fig. 6. It is noticeable that the model functions poorly on a learning rate of 0.1, which shows that it is unable to converge well on a high learning rate. In contrast, a small value, i.e., 0.0001, requires more time for convergence. However, the results improve regularly on 0.001 and 0.005 values, and therefore, the model gains a significant change considering Recall, MAP, and nDCG on the DBLP dataset. Moreover, a similar kind of trend is observed in the AAN dataset. To come up with a good learning

Fig. 5 The influence of K value in the memory network module

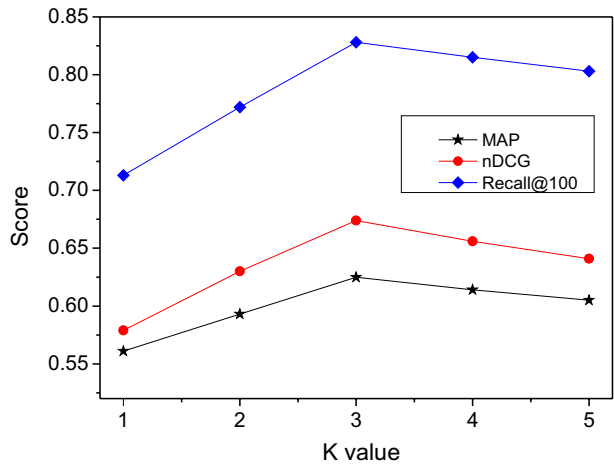
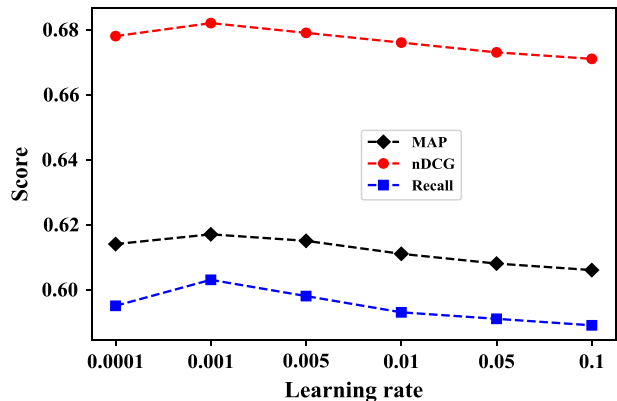


Fig. 6 Analyzing the influence of various learning rates on the final recommendations using the DBLP dataset



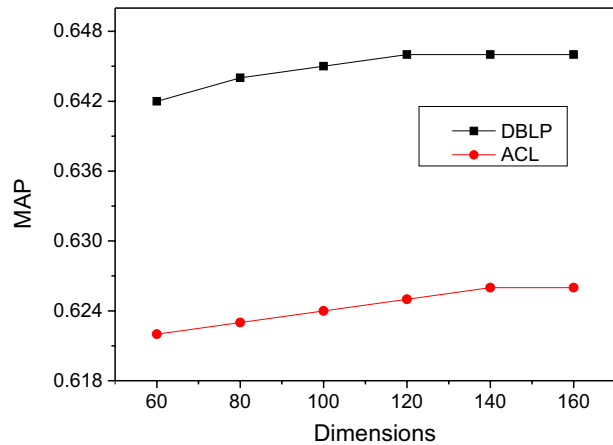
rate, we employ a fine-search mechanism and discover that 0.001 provides the best possible results on the DBLP and AAN datasets.

Finally, we analyze the model's performance for the layer number k of the memory network, its value varies from 1 to 5. Figure 5 depicts the impact of k on the model's results. We observe that the model gains the highest performance when k equals to 3. As the value of k increases from 1 to 3, the model achieves better results. But, when it goes from 3 to 5, then the performance of the model degrades. increasing k from 3 to 5, the performance is getting slightly lower. For simplicity, we set $k = 3$. On the other hand, the dimensionality of the MLP layer has a relatively less impact on the final results. In Fig. 7, we notice that the model achieves the highest performance when the values of the MLP layer are 120 and 140 on the DBLP-V12 and AAN datasets, respectively.

Error analysis

Our system achieves the MAP of 62.5 on the DBLP dataset. To further analyze the model and explore possible improvements, we conducted a manual error analysis. This is done by

Fig. 7 The dimensions of MLP layer in the personalization module using the DBLP-V12 and AAN datasets



randomly choosing 100 errors from our recommendation results. During the error analysis, we find that many errors occurred due to missing information regarding papers and authors. In particular, 33 out of 100 errors occurred due to the missing citation relations, and 27 happened due to the unavailability of paper contents (title and abstract). Further, 22 out of 100 inaccurate predictions are due to the unavailability of author information, the remaining 18 are caused by other factors like missing FOS, papers written in non-English languages, etc. Thus, with this in mind, we focused more effort on adding the missing citations, paper contents, and author information, which helped our model achieve an improved MAP score.

Conclusion and future work

Several models have been proposed in the literature to make personalized paper recommendations for researchers. However, these models are limited in exploiting several salient factors and semantic relations in the heterogeneous paper networks to capture researchers' preferences and generate relevant results. The existing models also encounter the 'cold-start papers' problem, which is addressed by the SPR-SMN model by employing contextual information and semantic relations corresponding to the network. The model uses SPECTER along with an end-to-end memory network to capture long-range contextual information. Its personalization module uses authors' information, the paper's FOS, and citations. The experimental results reveal the effectiveness of SPR-SMN against state-of-the-art baselines. The key findings of this research are the following.

- Employing SPECTER document embedding model can learn robust and semantic-preserving paper content embeddings.
- Exploiting personalized information, viz., paper citation relations, author's information, and field of study can better grasp the researcher's preference dynamics and produce justifiable results.
- Using an end-to-end memory network with an attention mechanism helps the model to exploit rich semantics and capture long-range dependencies.

- Integrating all these aspects into a unified model alleviates the cold-start paper and lack of personalization problems.

In the future, we will analyze the influence of temporal dynamics and the significance of contextual information using a hierarchical attention mechanism.

Declarations

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


References

- Abro, W. A., Qi, G., Ali, Z., Feng, Y., & Aamir, M. (2020). Multi-turn intent determination and slot filling with neural networks and regular expressions. *Knowledge-Based Systems*, 208, 106428.
- Achakulvisut, T., Acuna, D. E., Ruangrong, T., & Kording, K. (2016). Science concierge: A fast content-based recommendation system for scientific publications. *PLoS ONE*, 11(7), 1–11.
- Ali, Z., Kefalas, P., Muhammad, K., Ali, B., & Imran, M. (2020a). Deep learning in citation recommendation models survey. *Expert Systems with Applications*, 162, 113790.
- Ali, Z., Qi, G., Kefalas, P., Abro, W. A., & Ali, B. (2020b). A graph-based taxonomy of citation recommendation models. *Artificial Intelligence Review*, 53(7), 5217–5260.
- Ali, Z., Qi, G., Muhammad, K., Ali, B., & Abro, W. A. (2020c). Paper recommendation based on heterogeneous network embedding. *Knowledge-Based Systems*, 210, 106438.
- Ali, Z., Qi, G., Muhammad, K., Bhattacharyya, S., Ullah, I., & Abro, W. A. (2021a). Citation recommendation employing heterogeneous bibliographic network embedding. *Neural Computing and Applications*. <https://doi.org/10.1007/s00521-021-06135-y>
- Ali, Z., Qi, G., Muhammad, K., Kefalas, P., & Khusro, S. (2021b). Global citation recommendation employing generative adversarial network. *Expert Systems with Applications*, 180, 114888.
- Ali, Z., Ullah, I., Khan, A., Ullah Jan, A., & Muhammad, K. (2021c). An overview and evaluation of citation recommendation models. *Scientometrics*, 126(5), 4083–4119.
- Amami, M., Pasi, G., Stella, F., & Faiz, R. (2016). An LDA-based approach to scientific paper recommendation. In: *Natural Language Processing and Information Systems: Proceedings in 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24*, pp. 200–210
- Amir, N., Jabeen, F., Ali, Z., Ullah, I., Jan, A. U., & Kefalas, P. (2022). On the current state of deep learning for news recommendation. *Artificial Intelligence Review*. <https://doi.org/10.1007/s10462-022-10191-8>
- Bansal, T., Belanger, D., McCallum, A. (2016). Ask the gru: Multi-task learning for deep text recommendations. In: *Proceedings of the 10th ACM Conference on Recommender Systems (RecSys), RecSys '16*, pp. 107–114. New York, NY, USA
- Beltagy, I., Lo, K., & Cohan, A. (2019). *Scibert: A pretrained language model for scientific text*.
- Bhagavatula, C., Feldman, S., Power, R., & Ammar, W. (2018). Content-based citation recommendation. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 238–251). New Orleans: Human Language Technologies.
- Bollacker, K., Lawrence, S., & Giles, C. L. (1998). Citeseer: an autonomous web agent for automatic retrieval and identification of interesting publications. In: *Proceedings of the second international conference on Autonomous agents*, pp. 116–123
- Cai, X., Han, J., Li, W., Zhang, R., Pan, S., & Yang, L. (2018a). A three-layered mutually reinforced model for personalized citation recommendation. *Transactions on Neural Networks and Learning Systems*, 29(12), 6026–6037.
- Cai, X., Han, J., & Yang, L. (2018b). Generative adversarial network based heterogeneous bibliographic network representation for personalized citation recommendation. In: *Thirty-Second AAAI Conference on Artificial Intelligence*
- Cai, X., Zheng, Y., Yang, L., Dai, T., & Guo, L. (2019). Bibliographic network representation based personalized citation recommendation. *IEEE Access*, 7, 457–467.

- Chen, J., Liu, Y., Zhao, S., & Zhang, Y. (2019). Citation recommendation based on weighted heterogeneous information network containing semantic linking. *2019 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 31–36). Shanghai: IEEE.
- Chen, Y. L., Yeh, Y. H., & Ma, M. R. (2021). A movie recommendation method based on users' positive and negative profiles. *Information Processing and Management*, 58, 102531.
- Christoforidis, G., Kefalas, P., Papadopoulos, A., & Manolopoulos, Y. (2018). Recommendation of points-of-interest using graph embeddings. In: *5th IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2018, Turin, Italy, October 1-3, 2018*, pp. 31–40
- Christoforidis, G., Kefalas, P., Papadopoulos, A. N., & Manolopoulos, Y. (2021a). Reline: point-of-interest recommendations using multiple network embeddings. *Knowledge and Information Systems*, 63(4), 791–817.
- Christoforidis, G., Kefalas, P., Papadopoulos, A. N., & Manolopoulos, Y. (2021b). RELINE: point-of-interest recommendations using multiple network embeddings. *Knowledge and Information Systems*, 63(4), 791–817.
- Cohan, A., Feldman, S., Beltagy, I., Downey, D., & Weld, D. (2020). Specter: Document-level representation learning using citation-informed transformers. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 1–13). Association for Computational Linguistics.
- Cui, P., Wang, X., Pei, J., & Zhu, W. (2019). A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering*, 31(5), 833–852.
- Dai, T., Yan, W., Zhang, K., Qiu, C., Zhao, X., & Pan, S. (2021). Gated relational stacked denoising autoencoder with localized author embedding for global citation recommendation. *Expert Systems with Applications*, 184, 115359.
- Dai, T., Zhu, L., Wang, Y., & Carley, K. M. (2019). Attentive stacked denoising autoencoder with bi-lstm for personalized context-aware citation recommendation. *Transactions on Audio, Speech, and Language Processing (TACL)*, 28, 553–568.
- Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. [arXiv:1810.04805](https://arxiv.org/abs/1810.04805)
- Du, N., Guo, J., Wu, C. Q., Hou, A., Zhao, Z., & Gan, D. (2020). Recommendation of academic papers based on heterogeneous information networks. *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)* (pp. 1–6). IEEE.
- Du, Z., Tang, J., & Ding, Y. (2019). Polar: Attention-based cnn for one-shot personalized article recommendation. *Machine Learning and Knowledge Discovery in Databases* (pp. 675–690). Springer.
- Ganguly, S., & Pudi, V. (2017). Paper2vec: Combining graph and text information for scientific paper representation. *Advances in Information Retrieval* (pp. 383–395). Cham: Springer.
- Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-dependent sentiment classification with bert. *IEEE Access*, 7, 154290–154299.
- Goyal, P., & Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 151, 78–94.
- Grover, A., Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. In: *Proceedings of the 22nd ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 855–864
- Guo, L., Cai, X., Qin, H., Guo, Y., Li, F., Tian, G. (2019). Citation recommendation with a content-sensitive deepwalk based approach. In: *International Conference on Data Mining Workshops (ICDMW)*, pp. 538–543
- Gupta, S., Varma, V. (2017). Scientific article recommendation by using distributed representations of text and graph. In: *Proceedings of the 26th International Conference on World Wide Web Companion, WWW '17 Companion*, pp. 1267–1268.
- Huang, W., Wu, Z., Liang, C., Mitra, P., Giles, C. L. (2015). A neural probabilistic model for context based citation recommendation. In: *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence (AAAI)*, pp. 2404–2410
- Jiang, Z., Yin, Y., Gao, L., Lu, Y., Liu, X. (2018). Cross-language citation recommendation via hierarchical representation learning on heterogeneous graph. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR '18*, p. 635–644
- Kefalas, P., & Manolopoulos, Y. (2017). A time-aware spatio-textual recommender system. *Expert Systems with Applications*, 78, 396–406.
- Kefalas, P., & Symeonidis, P. (2015). Recommending friends and locations over a heterogeneous spatio-temporal graph. *MEDI*. Springer.
- Kefalas, P., Symeonidis, P., & Manolopoulos, Y. (2018). Recommendations based on a heterogeneous spatio-temporal social network. *World Wide Web*, 21(2), 345–371.
- Khadka, A., Cantador, I., Fernández, M.: Exploiting citation knowledge in personalised recommendation of recent scientific publications. In: *LREC (2020)*

- Khusro, S., Ali, Z., & Ullah, I. (2016). Recommender systems: Issues, challenges, and research opportunities. *Information Science and Applications (ICISA) 2016* (pp. 1179–1189). Springer.
- Kong, X., Mao, M., Wang, W., Liu, J., & Xu, B. (2019). Voprec: Vector representation learning of papers with text information and structural identity for recommendation. *Transactions on Emerging Topics in Computing (TETC)* (pp. 1–12). IEEE.
- Kreutz, C.K., Schenkel, R.: Scientific paper recommendation systems: a literature review of recent publications. ArXiv [abs/2201.00682](https://arxiv.org/abs/2201.00682) (2022)
- Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *Proceedings of the 31st International Conference on Machine Learning* (pp. 1188–1196). PLMR.
- Ma, Y., Mao, J., Ba, Z., & Li, G. (2020). Location recommendation by combining geographical, categorical, and social preferences with location popularity. *Information Processing and Management*, 57, 102251.
- McNee, S.M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S.K., Rashid, A.M., Konstan, J.A., Riedl, J. (2002). On the recommending of citations for research papers. In: *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pp. 116–125
- Perozzi, B., Al-Rfou, R., Skiena, S. (2014). Deepwalk: Online learning of social representations. In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, pp. 701–710
- Qiu, T., Yu, C., Zhong, Y., An, L., & Li, G. (2021). A scientific citation recommendation model integrating network and text representations. *Scientometrics*, 126(11), 9199–9221.
- Reimers, N., Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In: *EMNLP/IJCNLP*
- Ribeiro, L.F., Saverese, P.H., Figueiredo, D.R. (2017). struc2vec: Learning node representations from structural identity. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 385–394
- Salloum, S., & Rajamanthri, D. (2021). Implementation and evaluation of movie recommender systems using collaborative filtering. *Journal of Advances in Information Technology*. <https://doi.org/10.12720/jait.12.3.189-196>
- Son, J., & Kim, S. B. (2017). Academic paper recommender system using multilevel simultaneous citation networks. *Decision Support Systems*, 105, 24–33.
- Sukhbaatar, S., Szlam, A., Weston, J., Fergus, R.: End-to-end memory networks. arXiv preprint [arXiv:1503.08895](https://arxiv.org/abs/1503.08895) (2015)
- Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q. (2015). Line: Large-scale information network embedding. In: *Proceedings of the 24th International Conference on World Wide Web (WWW)*, pp. 1067–1077
- Uddin, M. N., Li, B., Ali, Z., Kefalas, P., Khan, I., & Zada, I. (2022). Software defect prediction employing bilstm and bert-based semantic feature. *Soft Computing*. <https://doi.org/10.1007/s00500-022-06830-5>
- Wang, C., Blei, D.M. (2011). Collaborative topic modeling for recommending scientific articles. In: *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 448–456
- Wang, H., & Li, W. (2015). Relational collaborative topic regression for recommender systems. *Transactions on Knowledge and Data Engineering (TKDE)*, 27(5), 1343–1355.
- Wang, J., Zhu, L., Dai, T., & Wang, Y. (2020a). Deep memory network with bi-lstm for personalized context-aware citation recommendation. *Neurocomputing*, 410, 103–113.
- Wang, L., Rao, Y., Bian, Q., & Wang, S. (2020b). Content-based hybrid deep neural network citation recommendation method. *International Conference of Pioneering Computer Scientists, Engineers and Educators* (pp. 3–20). Springer.
- Yang, L., Zheng, Y., Cai, X., Dai, H., Mu, D., Guo, L., & Dai, T. (2018). A lstm based model for personalized context-aware citation recommendation. *IEEE Access*, 6, 59618–59627.
- Yin, J., & Li, X. (2017). Personalized citation recommendation via convolutional neural networks. *Asia-Pacific web (APWeb) and web-age information management (WAIM) joint conference on web and big data* (pp. 285–293). Springer.
- Zhang, J., & Zhu, L. (2021). Citation recommendation using semantic representation of cited papers' relations and content. *Expert Systems with Applications*, 187, 115826.
- Zhao, W., Wu, R., & Liu, H. (2016). Paper recommendation based on the knowledge gap between a researcher's background knowledge and research target. *Information Processing and Management*, 52, 976–988.

Authors and Affiliations

Zafar Ali¹  · Guilin Qi¹ · Pavlos Kefalas² · Shah Khusro³ · Inayat Khan⁴ · Khan Muhammad⁵

Guilin Qi
gqi@seu.edu.cn

Pavlos Kefalas
kefalasp@csd.auth.gr

Shah Khusro
khusro@uop.edu.pk

Inayat Khan
inayat_khan@uop.edu.pk

Khan Muhammad
khan.muhammad@ieee.org

¹ School of Computer Science and Engineering, Southeast University, Nanjing, China

² Department of Informatics, Aristotle University, Thessaloniki, Greece

³ Department of Computer Science, University of Peshawar, Peshawar, Pakistan

⁴ Department of Computer Science, University of Buner, Buner, Pakistan

⁵ Visual Analytics for Knowledge Laboratory (VIS2KNOW Lab), Department of Applied Artificial Intelligence, School of Convergence, College of Computing and Informatics, Sungkyunkwan University, Seoul 03063, Republic of Korea