Differentiable Topics Guided New Paper Recommendation

Wen Li[1], Yi Xie[2,3], Hailan Jiang[4], and Yuqing Sun[1(✉)]

[1] Shandong University, Jinan, China
sun_yuqing@sdu.edu.cn
[2] National University of Defense Technology, Hefei, China
[3] Anhui Province Key Laboratory of Cyberspace Security Situation Awareness and Evaluation, Heifei, China
[4] Shandong Polytechnic, Jinan, China

Abstract. There are a large number of scientific papers published each year. Since the progresses on scientific theories and technologies are quite different, it is challenging to recommend valuable new papers to the interested researchers. In this paper, we investigate the new paper recommendation task from the point of involved topics and use the concept of subspace to distinguish the academic contributions. We model the papers as topic distributions over subspaces through the neural topic model. The academic influences between papers are modeled as the topic propagation, which are learned by the asymmetric graph convolution on the academic network, reflecting the asymmetry of academic knowledge propagation. The experimental results on real datasets show that our model is better than the baselines on new paper recommendation. Specially, the introduced subspace concept can help find the differences between high quality papers and others, which are related to their innovations. Besides, we conduct the experiments from multiple aspects to verify the robustness of our model.

Keywords: Paper recommendation · Topic model · GCN

1 Introduction

Currently, there are a large number of academic papers published every year. It's necessary to recommend researchers the valuable and interested papers. The number of citations is often regarded as an important indicator for the quality of papers. To describe the detailed contribution of a paper, a citation type can be further classified into three categories, *Background*, *Method* and *Result*. As an example, we show the papers concerning the technology *Transformer* [6], *GPT* [5], *BERT* [4], *GPT2* [3] and *BART* [2] in Fig. 1, which are labeled by

This work was supported by the National Nature Science Foundation of China, NSFC(62376138) and the Innovative Development Joint Fund Key Projects of Shandong NSF (ZR2022LZH007).
Wen Li and Yi Xie contributed equally to this work.

B. Luo et al. (Eds.): ICONIP 2023, LNCS 14452, pp. 44–56, 2024.
https://doi.org/10.1007/978-981-99-8076-5_4

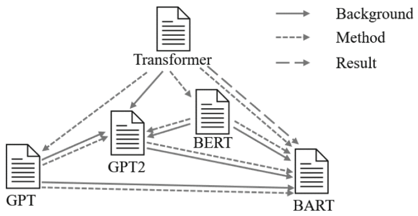


Fig. 1. An example of different citation types.

Semantic Scholar[1]. Arrows point to the citing papers, representing the direction of knowledge propagation. These points help the users more precisely find their interested topics, such as the inspiring theory, the technical methods, or the dataset and etc.

To recommend new paper, the existing methods typically leverage the academic network (AN for short) to model user interests and paper features [1,13]. However, they didn't consider the differentiable details on citations. Since the innovations in papers are various, the concept of subspace was used in this paper to describe the paper contents [21]. Besides, the citation-based recommendation methods are not applicable to new paper recommendation since it didn't have citation relationship.

To tackle the above challenges, we propose the differentiable topics based new paper recommendation model (DTNRec for short). Paper contents are classified into three subspaces according to the innovation forms as the usual way [21]: *Background*, *Method* and *Result*. We adopt the neural topic model (NTM for short) to get the topic distribution over subspaces as the paper embeddings, which are used to differentiate the innovation forms of paper. Considering the citations reflect the influence of cited papers and the author interests of citing paper, we adopt the asymmetric academic network to model this kind of knowledge propagation. The graph convolution network (GCN for short) operations are performed on this network to learn the user interests and paper influences, separately. For example, for the central paper p, its references are the neighbors during convolution to compute the interests for the authors of p, while its citations are used to compute its influences on the network. Then a new paper is recommended to the potentially interested users based on the paper content. Our contributions are as follows:

1. We label the paper content with subspace tags, then adopt the NTM to get the topic distribution over subspaces as paper embeddings.
2. We create the asymmetric academic network to model the academic propagation, where the directed edge points to the citing paper denoting the propagation. Based on this network and paper embeddings, we adopt the GCN operations to compute user interests and paper influences in a fine-grained way.

[1] https://www.semanticscholar.org/.

3. We conducted the experiments from multiple aspects to verify the effectiveness and robustness of our model.

2 Related Work

Collaborative filtering (CF for short) is a commonly used technique in recommendation systems. NeuMF [12] and BUIR [10] are both CF-based methods using user-item interaction data to get user and item representations. He et al. [11] proposed LightGCN to learn the user and item embeddings with neighborhood aggregation operation. Wang et al. [9] proposed alignment and uniformity as two properties that are important to CF-based methods, and optimized the two properties to get user and item representations. However, these methods only use interaction data of user and items, without considering other features.

The academic network consists of papers, authors, other related attributes and the relationships among them, which is important for paper recommendation task since it's rich in information. Existing works often used AN-based methods including KGCN [18], KGCN-LS [19], RippleNet [20], etc. to mine high-order information on the academic network, among which GCN is a widely used technique. However, these methods have cold-start problem and are not suitable for new paper recommendation since it lacks citation information.

Besides, paper contents are also considered to model user interests. JTIE [25] incorporated paper contents, authors and venues to learn user and paper representations. Xie et al. [26] proposed a cross-domain paper recommendation model using hierarchical LDA to learn semantic features of paper contents. Li et al. [13] proposed JMPR to jointly embed structural features from academic network and semantic features from paper contents. These methods alleviate the cold-start problem, but the diversity of paper innovations was not considered. Therefore, Xie et al. [21] proposed the subspace concept to label the paper content with *Background*, *Method* and *Result*. However, they didn't infer in subspace, that is they ignored the knowledge propagation among subspaces.

3 New Paper Recommendation Method

3.1 Problem Definition

Given a user set $\mathcal{U}$, a paper set $\mathcal{V}$, we aim to learn a prediction function $\mathcal{F}(u, q \mid \theta)$ that checks whether user $u \in \mathcal{U}$ has the potential interest of the new paper $q \in \mathcal{V}$, where θ denotes the parameters of function $\mathcal{F}$.

For an academic dataset, the academic network $\mathcal{G}$ is called the structural feature, where the nodes of $\mathcal{G}$ are papers, authors, and other related attributes, and the edges denote the relationships between them, including citation, etc. Each paper contains an abstract. The abstract describes the core content of a paper, which is called the semantic feature in this paper.

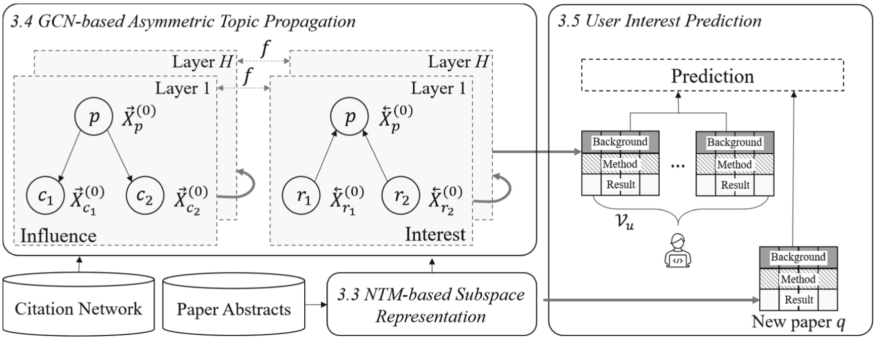


Fig. 2. Overall framework of DTNRec

3.2 Overall Framework

DTNRec include three modules, as shown in Fig. 2, i.e. the NTM-based subspace representation module, the GCN-based asymmetric topic propagation module, and the user interest prediction module. In the NTM-based subspace representation module, the paper abstract is labeled with subspace tags through the subspace tagging model. The resulting subspace text is fed into the NTM to obtain the topic distributions over subspaces as the paper content embeddings. In the GCN-based asymmetric topic propagation module, we adopt asymmetric GCN on the academic network $\mathcal{G}$ to model the asymmetric topic propagation among papers. The user interest prediction module predicts the probability on how much user u being interested in a new paper q.

3.3 NTM-Based Subspace Representation

Subspace Tagging. In order to differentiate the topics in papers, we inherit the subspace concept proposed in [21] and label the paper contents with three subspace tags, namely *Background*, *Method* and *Result*, respectively, denoted by the tag set $\mathcal{TS} = \{b, m, r\}$. We adopt the subspace tagging model in [22] to label the sentences of paper abstract with the subspace tags. The sentences for the same subspace represent the corresponding subspace text.

GSM-Based Paper Representation. The subspace texts are fed into the topic model to get the topic distributions over subspaces, which are regarded as the initial embeddings of paper content. The existing research results show that the topic model integrated with neural network has better performance than traditional topic model [23]. Therefore, we adopt the Gaussian Softmax distribution topic model (GSM for short) [23], which is based on variational autoencoder. Let $D \in \mathrm{N}^*$ denote the topic number. The output subspace topic distributions $\mathbf{x}_p^b \in \mathrm{R}^D, \mathbf{x}_p^m \in \mathrm{R}^D, \mathbf{x}_p^r \in \mathrm{R}^D$ for paper p are the corresponding embeddings, respectively. Paper p can be represented as matrix $X_p = \left(\mathbf{x}_p^b, \mathbf{x}_p^m, \mathbf{x}_p^r\right)^\top \in \mathrm{R}^{3\times D}$.

Different with the existing methods directly treat the paper content as a whole to obtain paper representation [13], our method label the paper content with subspace tags, which helps to distinguish paper innovations.

3.4 GCN-Based Asymmetric Topic Propagation

Each citation reflects the influence of the cited paper and the interest of the citing paper's authors. So we model the topic propagation between papers on the academic network as the asymmetric relations, denoted by $\mathcal{G}$. The academic influences and user interests are modeled, respectively, based on the citation relationships. For example, for a paper $p \in \mathcal{V}$ on $\mathcal{G}$, its references are the neighbors for convolution to compute the interests for the authors of p, while its citations are used to compute its influences on the network.

For any paper $p \in \mathcal{V}$ on $\mathcal{G}$, there are two matrix representations, denoted by the interest matrix $\overleftarrow{X}_p^{(h)}$ and the influence matrix $\overrightarrow{X}_p^{(h)}$, respectively, where $h \in \mathrm{N}^*$ denotes the depth of GCN, that is the number of GCN iterations. $\overleftarrow{X}_p^{(h)}$ and $\overrightarrow{X}_p^{(h)}$ both are initialized by the paper matrix X_p. The GCN kernel function is f, where $W \in \mathrm{R}^{\mathrm{D}\times\mathrm{D}}$, $U \in \mathrm{R}^{3\times\mathrm{D}}$, $V \in \mathrm{R}^{3\times\mathrm{D}}$ are all weights of f and $b \in \mathrm{R}^{3\times 3}$ is bias. Paper $p' \in \mathcal{V}$ cited paper p.

$$f(p,p',h) = \sigma\left(\overrightarrow{X}_p^{(h-1)} W \overleftarrow{X}_{p'}^{(h-1)\top} + U\overrightarrow{X}_p^{(h-1)\top} + V\overleftarrow{X}_{p'}^{(h-1)\top} + b\right) \tag{1}$$

To compute the influence of paper p, we choose citations of p as its neighbors. Since the number of paper neighbors may vary significantly over all papers, we uniformly sample a fixed-size set of neighbors for each paper instead of using all of them, denoted by $\mathcal{V}_p^{cit}$, to keep the computational pattern of each batch fixed and more efficient. We set $|\mathcal{V}_p^{cit}| = K \in \mathrm{N}^*$ as a hyper-parameter. Papers in $\mathcal{V}_p^{cit}$ are combined to characterize the influence of paper p, denoted by $\overrightarrow{X}_{\mathcal{V}_p^{cit}}^{(1)}$.

$$\overrightarrow{X}_{\mathcal{V}_p^{cit}}^{(1)} = \sum_{c\in\mathcal{V}_p^{cit}} f(p,c,1)\overrightarrow{X}_c^{(0)} \tag{2}$$

Then we aggregate $\overrightarrow{X}_p^{(0)}$ and $\overrightarrow{X}_{\mathcal{V}_p^{cit}}^{(1)}$ into one matrix $\overrightarrow{X}_p^{(1)}$ as p's first-order influence matrix, which is calculated as $\overrightarrow{X}_p^{(1)} = \sigma\left(\left(\overrightarrow{X}_p^{(0)} + \overrightarrow{X}_{\mathcal{V}_p^{cit}}^{(1)}\right) W^{(1)} + b^{(1)}\right)$.

In the same way, to compute the interest for the authors of paper p, we choose a fixed-size set of references of paper p as its neighbors, denoted by $\mathcal{V}_p^{ref}$. We set $|\mathcal{V}_p^{ref}| = K$, too. Then papers in $\mathcal{V}_p^{ref}$ are combined to characterize the interest for the authors of paper p, denoted by $\overleftarrow{X}_{\mathcal{V}_p^{ref}}^{(1)}$.

$$\overleftarrow{X}_{\mathcal{V}_p^{ref}}^{(1)} = \sum_{r\in\mathcal{V}_p^{ref}} f(r,p,1)\overleftarrow{X}_r^{(0)} \tag{3}$$

Then we aggregate $\overleftarrow{X}_p^{(0)}$ and $\overleftarrow{X}_{\mathcal{V}_p^{ref}}^{(1)}$ into one matrix $\overleftarrow{X}_p^{(1)}$ as p's first-order interest matrix, which is calculated as $\overleftarrow{X}_p^{(1)} = \sigma\left(\left(\overleftarrow{X}_p^{(0)} + \overleftarrow{X}_{\mathcal{V}_p^{ref}}^{(1)}\right) W^{(1)} + b^{(1)}\right)$.

We set the maximum depth of GCN as H. Through repeating the above process H times, we can get the H-order interest matrix $\overleftarrow{X}_p^{(H)}$ of paper p.

Given another paper q, to predict whether paper q will influence paper p or whether the author of paper p will be interested in paper q, we calculate the score $c(q,p)$. Since citation types are diverse, we adopt maximum pooling to find the largest topic association between different subspaces of paper p and paper q.

$$c(q,p) = MLP\left(maxpooling\left(X_q \overleftarrow{X}_p^{(H)\top}\right)\right) \tag{4}$$

We choose the cross entropy loss function. SP^+ and SP^- denote positive sample set and negative sample set, which are sampled according to the rule-based sample strategy [21]. Let $\hat{c}(q,p)$ denote gold label. Any paper pair (p,q) with citation relationship is sampled as positive, labeled as $\hat{c}(q,p) = 1$. The negative samples are selected from paper pairs without citation relationship according to the sample strategy in [21], labeled as $\hat{c}(q,p) = 0$.

$$L = \sum_{c(q,p) \in \mathrm{SP}^+ \cup \mathrm{SP}^-} c(q,p) \log \hat{c}(q,p) + \lambda \|\theta\|_2^2 \tag{5}$$

Existing recommendation methods typically initialize paper nodes randomly when using GCN. However, our method directly initializes paper nodes with paper's semantic-rich subspace embeddings. Besides, considering the asymmetry of topic propagation among papers, we conduct asymmetric GCN on the academic network to model user interest and academic influences in a fine-grained way.

3.5 User Interest Prediction

A new paper is recommended to the potentially interested users based on the content. Given a new paper q, we calculate the probability whether user u will be interested in paper q through the function $\mathcal{F}(u,q)$, where $\mathcal{V}_u$ denotes user u's history publications.

$$\mathcal{F}(u,q) = \max\left\{c(q,p)\mid, p \in \mathcal{V}_u\right\} \tag{6}$$

Since user interests change over time, we adopt the publications within a period as the user interests at different times. We calculate the probability on how much user u being interested in the paper q according to the user interests in different periods. In this way, the user interests are more accurately modeled.

4 Experiments

In this section, we verify the effectiveness of our model on real datasets for the new paper recommendation task. We select some baselines for comparative experiments and analyze the impact of hyper-parameter settings and model structure. Finally, we analyze the paper subspace embeddings.

4.1 Experimental Settings

Datasets. We use ACM[2] and Scopus[3] datasets. ACM dataset contains 43380 conference and journal papers in computer science. Scopus dataset is a multi-disciplinary dataset, and we use the papers within the area of computer science, with a total of 18842 papers. Every paper in the datasets contains the paper abstract, authors, publication year, citation relationship, etc.

Baselines and Metrics. We compare our model with several baselines. BUIR [10], LightGCN [11], NeuMF [12] and DirectAU [9] are CF-based methods using the user and item interaction data. KGCN [18], KGCN-LS [19], RippleNet [20] are AN-based methods, which introduce the side information such as keywords besides user-item interactions. NPRec [21] jointly embed the semantic features of paper content and structural features of academic network. DTNRec is our model.

In real recommendation scenarios, users usually pay attention to the first few items recommended. So we choose the $nDCG@k$ [8] as the metric to evaluate the ranking results. For each user, we prepare k candidates which contains at least one paper that is actually cited by the user. The candidate papers are ranked according to the value calculated by the function $\mathcal{F}$ (6). $DCG@k$ is calculated as $DCG@k = \sum_{i=1}^{k} \frac{rel_i}{\log_2(i+1)}$, where rel_i is a fixed value 5 if the i-th paper is actually cited by the user, otherwise 0. $IDCG = \sum_{i=1}^{|Ref|} \frac{5}{\log_2(i+1)}$ represents the DCG value corresponding to the best rank, where $|Ref|$ denotes the number of papers actually cited by the user in candidate papers.

4.2 Results

Performance Analysis. The evaluation results are shown in Table 1. It shows our model DTNRec outperforms the baselines on the new paper recommendation task. Because we introduce the concept of subspace, the paper innovations could be well differentiated. What's more, we fuse semantic features and structural features by performing asymmetric GCN on the academic network, whose nodes are initialized by paper content embeddings over subspaces. In this way, the user interests and paper influences are modeled in a fine-grained way. The CF-based models including BUIR, LightGCN, NeuMF and DirectAU performs worst since they only use interaction data of users and items, without considering other information such as paper content. The AN-based models including KGCN, KGCN-LS and RippleNet perform better than the CF-based methods, because the academic network contains rich high-order hidden information, which is beneficial for accurately modeling user preferences. Both CF-based methods and KG-based methods consider the structural features, without considering semantic features. NPRec considers both of them, so NPRec performs better than

[2] https://dl.acm.org/.
[3] https://www.scopus.com/.

AN-based models. However, the model structure of NPRec has limitations. It treats the paper content representation in a whole way rather than in subspaces, that is it ignored the knowledge propagation among subspaces.

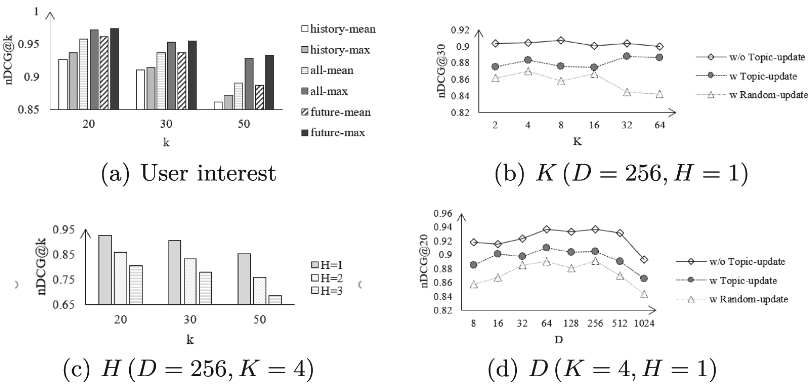
Table 1. New paper recommendation comparison.

nDCG@k	ACM			Scopus		
	k = 20	k = 30	k = 50	k = 20	k = 30	k = 50
BUIR	0.7734	0.7083	0.6681	0.7707	0.7156	0.6626
LightGCN	0.8266	0.7703	0.7314	0.8062	0.7639	0.7231
NeuMF	0.8234	0.7730	0.7419	0.8257	0.7808	0.7234
DirectAU	0.8357	0.7898	0.7423	0.8246	0.7819	0.7235
KGCN	0.8731	0.8592	0.8437	0.8507	0.8365	0.7592
KGCN-LS	0.9093	0.9010	0.8904	0.8660	0.8548	0.8063
RippleNet	0.9217	0.9088	0.8970	0.9040	0.8673	0.8465
NPRec	0.9736	0.9688	0.9645	0.9576	0.9349	0.9021
DTNRec	**0.9855**	**0.9844**	**0.9663**	**0.9735**	**0.9547**	**0.9329**

Impact of User Interest Calculation Method. Generally, the user interest will change over time. When we predict whether user u will be interested in a new paper q which is published after year Y, we should consider user u's interest after year Y, too. Therefore, we study the impact of using user u's interests at different times to make predictions. The experimental results are shown in Fig. 3(a). We calculate user interest in the following six ways.

- ***History-max*** denotes the user interest is computed as the function $\mathcal{F}$ (6), where $\mathcal{V}_u$ denotes the publications of user u before year Y.
- ***Future-max*** replaces $\mathcal{V}_u$ in *history-max* with the publications of user u after year Y.
- ***All-max*** replaces $\mathcal{V}_u$ in *history-max* with all the publications of user u.
- ***History-mean*** is the same as *history-max*, but replaces the operation of taking the maximum value in the function $\mathcal{F}$ (6) with taking the mean value.
- ***Future-mean*** replaces $\mathcal{V}_u$ in *history-mean* with the publications of user u after year Y.
- ***All-mean*** replaces $\mathcal{V}_u$ in *history-mean* with all the publications of user u.

In order to avoid information leakage, when computing user u's interest after year Y, we delete the citation relationship between papers published after year Y on the academic network, which means only u's publications after year Y and references before year Y are considered. The results in Fig. 3(a) show that *future* mode performs better than *history* mode and *all* mode. The *max* mode performs better than *mean* mode. When the user interest is calculated in the way of *future-max*, our model performs best, which also proves the user interest will change.



(a) User interest

(b) $K\,(D = 256, H = 1)$

(c) $H\,(D = 256, K = 4)$

(d) $D\,(K = 4, H = 1)$

Fig. 3. The results of the four figures are all carried out on Scopus dataset. (a) Comparison with different user interest computation methods. The hyper-parameter setting is $D = 256, H = 1, K = 4$. (b) (c) (d) are comparisons on model variants with different K, H, D, respectively. (b) (c) (d) all choose the *history-max* mode.

Ablation Study. To verify the impact of model structure on model performance, we conduct ablation experiments. The model variants are as follows.

- **w Random-update** randomly initializes $\overleftarrow{X}_p^{(0)}$ and $\overrightarrow{X}_p^{(0)}$. The parameters of $\overleftarrow{X}_p^{(0)}$ and $\overrightarrow{X}_p^{(0)}$ will be updated during the training process.
- **w Topic-update** initializes $\overleftarrow{X}_p^{(0)}$ and $\overrightarrow{X}_p^{(0)}$ with matrix X_p. And the parameters will be updated.
- **w/o Topic-update** initializes $\overleftarrow{X}_p^{(0)}$ and $\overrightarrow{X}_p^{(0)}$ in the same way as **w Topic-update**, but the parameters will not be updated.

The results are shown in Fig. 3(b) and 3(d). **w Random-update** performs worst. **w Topic-update** is better than **w Random-update**. And **w/o Topic-update**, which is also the final setting of our model, performs best. Because the paper subspace embeddings $\mathbf{x}_p^b, \mathbf{x}_p^m$ and $\mathbf{x}_p^r$, that are also the topic distributions output by NTM, are rich in semantic information. They do not need to be updated further. Instead, update brings information loss, resulting in model performance degradation.

Hyper-parameter Study. We analyzed the impact of hyper-parameter settings on model performance. Three hyperparameters are tested: the neighbor number K, the maximum depth of GCN H, the topic number D. The results are shown in Fig. 3(b), 3(c) and 3(d). Figure 3(b) shows that when K becomes larger, $nDCG@30$ of **w Random-update** will decrease due to the introduction of noise. **w Topic-update** and **w/o Topic-update** are not sensitive to the setting of K, which means the initialization by X_p weaken the influence of K on model performance. Figure 3(c) shows when H is set to 1, model performs better. As H increases, model performance decreases. Because there may be over-smooth problem with the increase of H. As shown in Fig. 3(d), we find as D increases, the value of $nDCG@k$ will first rise and then fall. Because a smaller D also means less semantic information. The topics in all papers may not be fully covered. The model performs best when D is set to 256. When D is too large, the model performance will decline, probably because the size of D can already cover all the topics, and continuing to increase will not obtain richer semantic information, but will introduce noise.

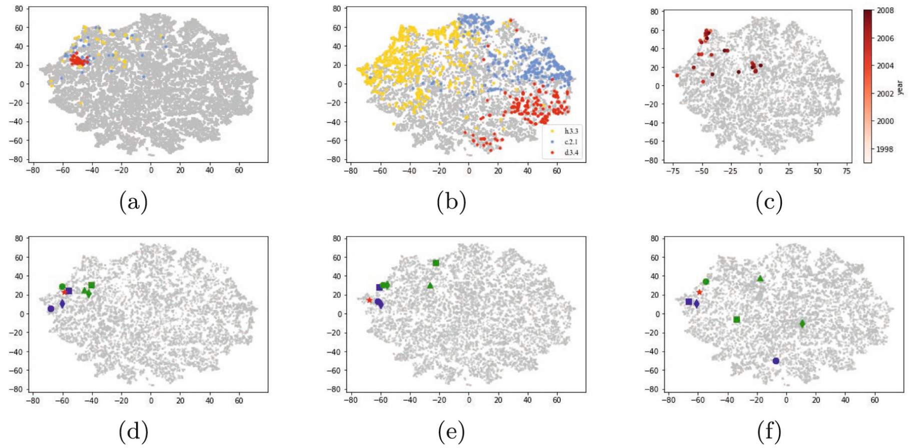


Fig. 4. Analyze the subspace embeddings. The results are based on ACM dataset. The topic number D is set as 16, so the subspace embeddings of papers are 16-dimensional. Then we reduce them into a 2-dimensional visual space by t-SNE [24]. Gray dots in each figure denote all papers. (a) Subspace embeddings of papers with similar background. (b) Background embeddings of papers with different CCS tags. (c) Background embeddings of Chengxiang Zhai's publications. (d) (e) (f) respectively analyze the *Background*, *Method* and *Result* embeddings of paper [17] and its references and citations.

Analyze the Subspace Embeddings. We analyze the subspace embeddings from different aspects, where the ACM Computing Classification System (ACM CCS) [7] is used as supplementary information.

In order to verify the necessity of subspace, we randomly selected a paper [14] with CCS tag h.3.3 (information search and retrieval). Then 50 papers with similar background to paper [14] are selected from paper set with the same CCS tag. The similarity is obtained by calculating the Euclidean distance of background embeddings. The smaller the distance, the more similar the background. As shown in Fig. 4(a), the red dots denote the background embeddings of the 50 papers. The yellow and blue dots represent method and result embeddings, respectively. We find that papers with similar background may have different methods and results. But the topics do not differ dramatically, but vary within a certain range of topics. So the consideration of subspace is necessary.

To verify whether the subspace embeddings of paper content could reflect CCS tag information, we randomly chose three CCS tags: h.3.3 (information search and retrieval), c.2.1 (network architecture and design) and d.3.4 (processors). As shown in Fig. 4(b), the papers with different CCS tag could be well differentiated.

Figure 4(c) shows some publications of researcher Chengxiang Zhai between 1998 and 2008. The red dots denote his publications, and the color shades correspond to publication years. It illustrates that the researcher's research interests will change within a field of study.

To study whether the subspace embeddings of paper content could reflect the relationship between a paper and its references and citations, we randomly selected a highly cited paper [17]. As shown in Fig. 4(d), 4(e) and 4(f), the red star denotes paper [17], and the green and blue shapes denote references and citations of [17], respectively. We can see that the topics between a paper and its references and citations are all close in different subspaces. But there are also differences, which reflect the topic propagation among different subspaces of papers. For example, the distance between the red star [17] and the green triangle [16] on Fig. 4(d) is closer than the distance on Fig. 4(e) and 4(f). Because the background of paper [17] and paper [16] are all related to the classification of web content, but they used different methods and thus got different results. Besides, the distance between the red star [17] and the blue circle [15] on Fig. 4(d) and 4(e) are closer than the distance on Fig. 4(f). Because the backgrounds of paper [17] and paper [15] are similar and both adopted user survey method. The results are different is due to the core issues of their research are different. It's worth mention that paper [17] and paper [15] have the same author Mika. It illustrates that the researchers tend to use similar methods in their publications.

5 Conclusion

We propose a differentiable topics based new paper recommendation model DTNRec. In DTNRec, we adopt the subspace tagging model and NTM to get embeddings of paper content. Then we model the user interest through the asymmetric GCN on the academic network. The experimental results show the effectiveness of our model.

References

1. Kreutz, C.K., Schenkel, R.: Scientific paper recommendation systems: a literature review of recent publications. Int. J. Digit. Libr. **23**(4), 335–369 (2022)
2. Lewis, M., Liu, Y., et al.: BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In: Annual Meeting of the Association for Computational Linguistics, pp. 7871–7880 (2020)
3. Radford, A., Wu, J., et al.: Language models are unsupervised multitask learners. OpenAI blog **1**(8), 9 (2019)

4. Devlin, J., et al.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT, pp. 4171–4186 (2019)
5. Radford, A., Narasimhan, K., et al.: Improving language understanding by generative pre-training (2018)
6. Vaswani, A., Shazeer, N., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
7. Coulter, N., et al.: Computing classification system 1998: current status and future maintenance report of the CCS update committee. Comput. Rev. **39**(1), 1–62 (1998)
8. Wang, Y., Wang, L., et al.: A theoretical analysis of NDCG type ranking measures. In: Conference on Learning Theory, pp. 25–54. PMLR (2013)
9. Wang, C., Yu, Y., et al.: Towards representation alignment and uniformity in collaborative filtering. In: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 1816–1825 (2022)
10. Lee, D., Kang, S., et al.: Bootstrapping user and item representations for one-class collaborative filtering. In: International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 317–326 (2021)
11. He, X., Deng, K., et al.: LightGCN: simplifying and powering graph convolution network for recommendation. In: International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 639–648 (2020)
12. He, X., Liao, L., et al.: Neural collaborative filtering. In: 26th International Conference on World Wide Web, pp. 173–182 (2017)
13. Li, W., et al.: Joint embedding multiple feature and rule for paper recommendation. In: Sun, Y., Liu, D., Liao, H., Fan, H., Gao, L. (eds.) ChineseCSCW. CCIS, vol. 1492, pp. 52–65. Springer, Singapore (2021). https://doi.org/10.1007/978-981-19-4549-6_5
14. Joachims, T.: Optimizing search engines using clickthrough data. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 133–142 (2002)
15. Aula, A., et al.: Information search and re-access strategies of experienced web users. In: International Conference on World Wide Web, pp. 583–592 (2005)
16. Dumais, S., et al.: Hierarchical classification of web content. In: ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 256–263 (2000)
17. Käki, M.: Findex: search result categories help users when document ranking fails. In: SIGCHI Conference on Human Factors in Computing Systems, pp. 131–140 (2005)
18. Wang, H., Zhao, M., et al.: Knowledge graph convolutional networks for recommender systems. In: The World Wide Web Conference, pp. 3307–3313 (2019)
19. Wang, H., Zhang, F., et al.: Knowledge-aware graph neural networks with label smoothness regularization for recommender systems. In: ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 968–977 (2019)
20. Wang, H., Zhang, F., et al.: RippleNet: propagating user preferences on the knowledge graph for recommender systems. In: ACM International Conference on Information and Knowledge Management, pp. 417–426 (2018)
21. Xie, Y., et al.: Subspace embedding based new paper recommendation. In: International Conference on Data Engineering (ICDE), pp. 1767–1780. IEEE (2022)
22. Jin, D., Szolovits, P.: Hierarchical neural networks for sequential sentence classification in medical scientific abstracts. In: Conference on Empirical Methods in Natural Language Processing, pp. 3100–3109 (2018)
23. Miao, Y., et al.: Discovering discrete latent topics with neural variational inference. In: International Conference on Machine Learning, pp. 2410–2419. PMLR (2017)

24. Van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(11) (2008)
25. Xie, Y., Wang, S., et al.: Embedding based personalized new paper recommendation. In: Sun, Y., Liu, D., Liao, H., Fan, H., Gao, L. (eds.) ChineseCSCW 2020. CCIS, vol. 1330, pp. 558–570. Springer, Singapore (2021). https://doi.org/10.1007/978-981-16-2540-4_40
26. Xie, Y., Sun, Y., et al.: Learning domain semantics and cross-domain correlations for paper recommendation. In: 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 706–715 (2021)