Learning Domain Semantics and Cross-Domain Correlations for Paper Recommendation

Yi Xie
Shandong University
Jinan, China
heilongjiangxieyi@163.com

Yuqing Sun*
Shandong University
MoE ERC of Digital Media Technology, Jinan, China
sun_yuqing@sdu.edu.cn

Elisa Bertino
Purdue University
West Lafayette, The United States
bertino@cs.purdue.edu

ABSTRACT

Understanding how knowledge is technically transferred across academic disciplines is very relevant for understanding and facilitating innovation. There are two challenges for this purpose, namely the semantic ambiguity and the asymmetric influence across disciplines. In this paper we investigate knowledge propagation and characterize semantic correlations for cross discipline paper recommendation. We adopt a generative model to represent a paper content as the probabilistic association with an existing hierarchically classified discipline to reduce the ambiguity of word semantics. The semantic correlation across disciplines is represented by an influence function, a correlation metric and a ranking mechanism. Then a user interest is represented as a probabilistic distribution over the target domain semantics and the correlated papers are recommended. Experimental results on real datasets show the effectiveness of our methods. We also discuss the intrinsic factors of results in an interpretable way. Compared with traditional word embedding based methods, our approach supports the evolution of domain semantics that accordingly lead to the update of semantic correlation. Another advantage of our approach is its flexibility and uniformity in supporting user interest specifications by either a list of papers or a query of key words, which is suited for practical scenarios.

CCS CONCEPTS

• **Information systems** *Information retrieval*.

KEYWORDS

semantic correlation; academic paper; recommendation

ACM Reference Format:
Yi Xie, Yuqing Sun, and Elisa Bertino. 2021. Learning Domain Semantics and Cross-Domain Correlations for Paper Recommendation. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3404835.3462975

*Corresponding Author

ACM ISBN 978-1-4503-8037-9/21/07…$15.00
https://doi.org/10.1145/3404835.3462975

1 INTRODUCTION

Knowledge propagation across academic disciplines is one of the most important factors for scientific innovation; many scientific discoveries are the result of multi-disciplinary research enabled by knowledge propagation. The impact of knowledge propagation in today research is very clearly shown by the fast increasing numbers of cross discipline publications and paper citations. For example, in *Biology* and *Information Science*, the number of cross-citation papers in *Scopus*[14] increased by 11.2% in the past 5 years. This trend is also reflected by research involving different fields within the same discipline – one example is the increased adoption of machine learning methods in the security, computer networks, and human-centered computing fields. To speed up scientific advances, it is thus critical for researchers to get well acquainted with state of the art research in related fields. To address such a need, an important mechanism is represented by systems for cross-discipline paper recommendation, which can recommend papers from different disciplines that may be relevant for scientists involved in some research task in a given discipline.

However, cross-domain recommendations have specific challenges compared to traditional recommendations, such as recommendation for items to buy or movies to see. The first challenge is the ambiguity of domain-dependent word semantics. Most computational linguistic methods tackle the word semantics [5, 24, 34] based on the word common usage. For example, the widely adopted $Word2Vec$ method learns the distributed word vectors from a corpus by embedding word usage features into a low-dimensional space. Although these methods have highly improved the performance of many natural language processing tasks, they are not able to take into account the semantic differences between scientific fields. For example, the phrase "neural network" in the *Computer Science* domain has different meanings with respect to meanings it has in the *Biology Science* domain. Besides, the semantics in each domain evolve over time as science advances. Take "AI" for instance, relevant researchers focused on machine learning in the 20th century, and switched their attention to genetic algorithms, neural networks and other domains in the 21st century. Such an evolution has specific domain characteristics, which are not taken into account by the current recommendation approaches.

Another challenge is how to model the academic influence across disciplines. For example, information theory was originally proposed to find fundamental limits for signal processing and communication systems and has since then fostered developments in other disciplines, such as behavioral science, neuroscience, biology, and computer science. Such influence is asymmetric. But most of the current researches focusing on knowledge propagation across recommendation platforms assume that transfer follows a symmetric pattern. For example, some researchers focus on product recommendations in e-commence [21, 42] and assume that the knowledge learned from one platform is similar to the knowledge learned from another. Such symmetric influence does not apply to cross-discipline scientific recommendations.

The third challenge is how to provide a flexible integrated approach by which users can express their interests. On most scientific publication websites, a user query is usually specified as a set of key words. However an approach, by which an academic service directly recommends papers to its registered users based on the publications or references of these users, seems a better fit for websites providing academic services, such as Google Scholar and Web of Science[14]. Although neural networks-based methods like Bert [7, 11] are designed for sequential text data and work well in many tasks, they are unable to seamlessly support these different forms of user interest specifications while sharing the learned semantics and correlation.

To tackle the above challenges, we adopt a generative model to learn word semantics on hierarchical domains; these domains reflect the hierarchical organizations of disciplines into research fields. In designing our approach, we take into account two human-centered factors concerning publications. One is that the scientific literature in the fields of a given discipline is classified with respect to hierarchical categories, which are widely used by authors and publishers, such as the *ACM Computing Classification System*[31] in *Computer Science*. Another is that the boundary between research fields is not strict. Researcher interests can be expressed either with respect to a general domain or a fine-grained domain, such as *Human-centered computing* or *Information retrieval*. Therefore the domain semantics should support flexible scopes and be probabilistic. To address these factors, we adopt a supervised generative method to learn the probabilistic associations of words with the existing discipline classification, called *domain semantics*. Compared to the traditional distributed embeddings as word semantics, such a method reveals the linguistic features with respect to the categories and the word usages for different scenarios. Besides, the generative method supports the character of open usage in linguistics, i.e., introducing new word and semantics evolution.

We then model the asymmetric influence between *domain semantics* with a neural network. Since in practice it is difficult to find hints about the correlations between un-cited papers, especially because of the large numbers of publications every year, a citation should be considered the deterministic evidence of a high correlation between papers. Thus a cross-domain citation reflects authors' consideration of the intrinsic influence from the referenced paper to another domain. To moderately highlight the difference between paper citations and papers without the citation relationship, we design a ranking metric to evaluate this correlation rather than using a definite score. This model can better represent author biases with respect to academic influence.

With the help of *domain semantics* and correlations, both paper content and user interest are modeled as probabilistic distributions over the target domain and the related papers are recommended, where an interest can be a query with key words or a list of publications. We use three publicly available datasets to verify our model and discuss the intrinsic factors of results in an interpretable way. The rest of this paper is organized as follows. Section 2 discusses related works. Then we present the technique details in Section 3 and 4. In Section 5, we present the evaluation results and discuss the interpretability, feasibility, robustness and reusability of our model. Finally, we conclude the paper in Section 6.

2 RELATED WORK

Paper Recommendation. The work most related to ours is work focusing on both the influence of a paper and the relatedness between a user interest and paper contents. For example, Yates et al. [40] adopt the PageRank method to quantify the quality of papers by the citation network. Sugiyama et al. [33] propose a comprehensive evaluation of paper recommendation by an adaptive neighbor selection method in a collaborative filtering framework. Some approaches consider other academic recommendation, such as expert ranking and paper recommendation [26, 39]. But these approaches do not take into account knowledge propagation for paper recommendation.

Considering user specific interest, Bollacker et al. [6] develop the CiteSeer digital library system on personalized paper recommendation, which consists of three-stages: feature extraction, personalized filtering of new publications, and discovery of interesting research and trends. For a general query, Chakraborty et al. [9] propose a diversified citation recommendation system that considers the semantically correlated papers, especially including those papers not using exactly the same keywords as the query. Although the above approaches consider the main aspects of academic recommendation, such as word semantics, paper influence, and user interests, they focus on a single discipline and thus cannot solve the problem of domain dependent semantics and cannot identify the asymmetric influence between domains.

Cross-domain Recommendation. Recently a few approaches have been proposed focusing on cross-domain recommendation by transferring the semantic knowledge from one domain to another homoplasmic domain, such as Epinions and Ciao [36, 38], EachMovie and MovieLens [12, 35]. Most of them are based on the scenario of a rating system and the learned knowledge about user interest is represented as some distributions over item categories [20, 23]. Zhang et al. [42] propose an active learning strategy to consider both specific and independent knowledge over all domains for recommendation in order to alleviate the data sparsity in multi-domain scenarios. Liu et al. [21] propose a domain-sensitive recommendation algorithm to predict a rating score by exploring user-item subgroups. Zhao et al. [43] extract users' preferences from their review documents, and transfer them based on user interactions from one domain to the other relevant domain. Krishnan et al. [18] propose a neural collaborative filtering method with domain-invariant components shared across the dense and sparse domains, to improve the user and item representations learned in the sparse domains. Bi et al. [3, 4] learn a feature mapping function between two domains by multi-layer perception based on the user representation similarity in source domain and target domain. However, these methods consider symmetric cross-domain recommendation scenarios, which is not appropriate for scientific papers. Another obvious difference is about the organization of categories. In the current approaches, the categories in each domain are organized in the form of collections rather than of hierarchies. The constraints on the inheritance on semantics make these methods not applicable to the problem of scientific papers recommendation.

Other related methods. Some neural network methods learn the embeddings of item tags and user attributes like gender [8, 32] and thus they can recommend a new item to a new user based on these embeddings, which however are not applicable to text. Although recurrent neural networks are specifically designed for sequential data, they cannot represent text, words and phrases in a uniform way. To understand recommendation results, some approaches enforce specific constraints on the model factors to express the semantics of latent dimensions, or provide feature-based explanations [2, 15, 26]. But they are based on the scenario of rating systems, and thus are not able to explain the intrinsic semantics for academic disciplines. Besides, they do not consider the different semantics and the asymmetric influence between domains.

3 DOMAIN SEMANTICS LEARNING

A discipline refers to a set of concepts and theories; each discipline is hierarchically organized into a set of categories, referred to as *domains*. These categories typically corresponds to the fields of the disciplines. Let $\mathbf{T}$ denote a scientific classification tree. Each node in $\mathbf{T}$ and its subtrees represent a specific domain. Each node $n \in \mathbf{T}$ except the root $r \in \mathbf{T}$ has a parent, denoted by $pa(n) \in \mathbf{T}$. For a given paper corpus $\mathbf{P}$, the word vocabulary involved in $\mathbf{P}$ is denoted by $\mathbf{W}$. The metadata of a paper include a title, authors, an abstract, keywords, and one or more tags corresponding to nodes in $\mathbf{T}$.

Formally, if $p \in \mathbf{P}$ is associated with tag $n \in \mathbf{T}$, then variable $y_{n,p}$ is set 1, otherwise -1. According to an intuitive notion of hierarchical categories, the inheritance property is often expressed as constraints on tags $y_{n,p} = 1$ implies $y_{pa(n),p} = 1, y_{pa(pa(n)),p} = 1, ..., y_{r,p} = 1$. Conversely, if p is not associated with n', then no descendant tag of $n' \in \mathbf{T}$ can be assigned to p.

3.1 Domain Dependent Semantics

As word usage is open and continuously evolves in scientific domains, we adopt a generative probabilistic model to learn word semantics, which is an extension of the topic model [5]. The traditional topic model is an unsupervised generative model, referred to as latent Dirichlet allocation method, which introduces latent variables to learn the intrinsic semantics from large document collections. Under such a model, a document is modeled as a finite mixture over a set of latent topics $1, \dots, K$, each topic k has a probabilistic distribution over a vocabulary of words, denoted by ϕ_k. Several modifications have been proposed supporting supervision [28, 30]. But they either cannot directly map the latent topics to a set of predefined tags, or overlook the hierarchical structures of tags. In this paper, we adopt the hierarchical supervised latent Dirichlet allocation (HSLDA) model [29] to extract domain semantics. It learns the word semantics as the associations with topics, while leveraging the hierarchical structure of the tags. The difference with previous approaches is the introduction of discipline classification and the restriction on choosing a tag path, which are specific to scientific semantics learning.

Algorithm 1 Domain semantics Extraction

Require: $\mathbf{P}, \mathbf{T}, \mathbf{W}$, topic number K, parameters α, α', γ.
Ensure: Z, η.
1: **for** each topic $k = 1, ..., K$ **do**
2: Draw a distribution over words $\phi_k \sim Dir_{|\mathbf{W}|}(\gamma \mathbf{1}_{|\mathbf{W}|})$
3: **end for**
4: **for** each node $n \in \mathbf{T}$ **do**
5: Draw topic proportions $\eta_n|\mu, \sigma \sim \mathcal{N}_K(\mu \mathbf{1}_K, \sigma \mathbf{I}_K)$
6: **end for**
7: Draw the global topic proportions $\beta|\alpha' \sim Dir_K(\alpha' \mathbf{1}_K)$
8: **for** each paper $p \in \mathbf{P}$ **do**
9: Draw topic proportions $\varepsilon_p|\beta, \alpha \sim Dir_K(\alpha\beta)$
10: **for** each word w_i in paper p **do**
11: Draw topic assignment $z_{i,p}|\varepsilon_p \sim Multinomial(\varepsilon_p)$
12: Draw word $w_{i,p}|\phi_{z_{i,p}} \sim Multinomial(\phi_{z_{i,p}})$
13: **end for**
14: Set the assignment of the root $y_{r,p} = 1$
15: **for** each node n in a breadth first traversal of $\mathbf{T}$ starting at the children of root r **do**
16: Draw node assignment of p, $\zeta_p|\bar{\mathbf{z}}_p, \eta_n, y_{pa(n),p} \sim \begin{cases} \mathcal{N}(\bar{\mathbf{z}}_\mathbf{p}^\top \eta_n, 1), & y_{pa(n),p} = 1 \\ \mathcal{N}(\bar{\mathbf{z}}_\mathbf{p}^\top \eta_n, 1) I(\zeta_p < 0), & y_{pa(n),p} = -1 \end{cases}$
17: Assign node n to paper p according to ζ_p
$$y_{n,d}|\zeta_p = \begin{cases} 1 & \zeta_p > 0 \\ -1 & otherwise \end{cases}$$
18: **end for**
19: **end for**

To make the paper self-contained, we present the adaptation details in Algorithm 1, where: α, α' and γ are external parameters; K is the topic number; $Dir_K(\cdot)$ is a K-dimensional Dirichlet distribution; $\mathcal{N}_K(\cdot)$ is a K-dimensional Normal distribution; $\mathbf{I}_K$ is the identity matrix; β is a global distribution over topics and it is generated by the Dirichlet distribution; $\mathbf{1}_K$ is the K-dimensional vector with each element equal to 1; $I(\cdot)$ is an indicator function that takes value 1 if its argument is true and 0 otherwise.

The hierarchically organized *domain semantics* embed word usage into the fine-grained categories, denoted by $\mathbf{DS}$. Formally, $\mathbf{DS}$ includes two parts: the word-specific probabilistic distributions $Z \in R^{|\mathbf{W}|*K}$ over latent topics, and the probabilistic correlations $\eta \in R^{|\mathbf{T}|*K}$ between the nodes and the topics. Let $w_{i,p} \in \mathbf{W}$ denote the i-th word in the sequential text of p. Its probabilistic distribution over topics is denoted by $z_{i,p} \in R^K$ mapping to the corresponding row of Z. The column vector $Z_k^T \in R^{|\mathbf{W}|}$ denotes the topic-specific distribution over words of topic k. For convenience, we use $\phi_{1:K}$ to denote Z^T, use $\phi_k \in R^{|\mathbf{W}|}$ to denote Z_k^T, and use $\eta_n \in R^K$ to denote the distribution over topics of node n.

The hierarchical tags are associated with p according to the the words in p. We first assign the root node to p, i.e., $y_{r,p} = 1$, indicating that all domain-dependent papers must be labeled by the root node. $\zeta_p \in R^{|T|}$ is a paper-specific distribution over nodes. Here $\bar{\mathbf{z}}_\mathbf{p}^\top = [\bar{z}_1, ..., \bar{z}_K]$ is the empirical topic distribution for p, in which each entry $\bar{z}_k$ is the proportion of the words in p assigned to topic k, i.e. $\bar{z}_k = |p|^{-1}\Sigma_{i=1}^{|p|}(z_{i,p} = k)$. In line 15, for every node $n \in \mathbf{T}$, two preconditions are used to determine whether it is applied to p: the topic distribution vector for p and whether the parent of n was applied. Note that n can be assigned to p only if its parent $pa(n)$ is also applied. Moreover, if $pa(n)$ is applied to p, it is possible that n is not applied to p.

Compared to other semantic models, there are three advantages in our domain semantic model: the **DS** supports flexible scopes of a domain, denoted as domain extents; our model considers the semantic differences between scientific fields to reduce the semantic ambiguity; the semantics in each domain can evolve over time as science advances, which is the property of topic model.

4 DOMAIN SEMANTIC CORRELATION

4.1 Correlation Metric and Mapping Function

Let $\mathbf{T}^\mathbf{S}$ and $\mathbf{T}^\mathbf{d}$ denote the node set of the hierarchical classification trees of domain **S** and **D**, respectively. $\mathbf{P}^\mathbf{S}$ and $\mathbf{P}^\mathbf{D}$ denote the two corresponding paper corpora. Since a paper citation reflects authors' consideration on the intrinsic influence by its references, the papers with cross-domain citations are more important and related to the source domain **S** than those without a citation. Although there may exist correlations between un-cited papers, it is time-consuming to find evidence or hints of such correlations while the knowledge extracted is relatively little, especially because of the large numbers of publications every year. As an alternative way, we introduce a ranking evaluation, referred to as *correlation metric*, instead than a definite score to increase the correlation for cited papers compared with papers without the citation relationship.

Let $E = \{(p,q)|p \in P^S, q \in P^D, p\ cites\ q\}$ denote the set of paper pairs with cross citation relationships from **S** to **D**. Similarly, $\bar{E} = \{(p,q)|p \in P^S, q \in P^D, p\ does\ not\ cite\ q\}$. Based on the learnt domain semantics **DS**, the probabilistic distributions between words and nodes is computed by $\varphi = Z \cdot \eta^\top$. Then for paper p, its **DS** representation is calculated as the expectation of word-specific probability distributions over nodes, i.e $\zeta_p = |p|^{-1}\Sigma_{i=1}^{|p|}\varphi_i$, where φ_i denotes the node probability distributions of word $w_{i,p}$. Our goal is to learn the asymmetric influence function π from **S** to **D** based on E and $\bar{E}$. Formally, $\pi(x : \mathbb{R}^{|\mathbf{T}^\mathbf{S}|}) \mapsto \mathbb{R}^{|\mathbf{T}^\mathbf{D}|}$ represents the semantic mapping between nodes $n_i \in \mathbf{T}^\mathbf{S}$ and $n_j \in \mathbf{T}^\mathbf{D}$.

Definition 1. ***Correlation Metric***. *Let* **S** *and* **D** *be domains and* π *be an asymmetric influence function defined between* **S** *and* **D**. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be the vectors of two papers in the same semantic space. A cross-domain* correlation metric $\mu(\mathbf{x}, \mathbf{y}) \mapsto R^+$ *should satisfy the condition that for any two pairs of papers* $(p,q) \in E$ *and* $(p,q') \in \bar{E}$, $\mu(\pi(p), q) \geqslant \mu(\pi(p), q')$.

In the above definition, the function $\mu(\mathbf{x}, \mathbf{y})$ measures the semantic correlation between two papers. $\pi(p)$ maps the contents of p from **S** to **D** such that papers in different domains can be mapped into the same space and their correlation calculated. So $\mu(\pi(p), q)$ highlights the semantic correlation between papers cited across-domains. The operator $\geqslant$ between correlations gives a preference, which is more flexible than a hard score as a firm distance. Theoretically, the combination of μ, π and the ranking mechanism $\geqslant$ allows one to compute the scientific influence. We introduce a neural network method to learn these functions.

In what follows, we adopt the distance function $\hat{\mu}$ instead of the correlation function μ just for convenience in the discussion as most objective functions are defined for minimization. The distance function satisfies three properties: non-negative, symmetric and triangular. Let θ denote all the parameters to be learned. The objective function is then regulated as follows:

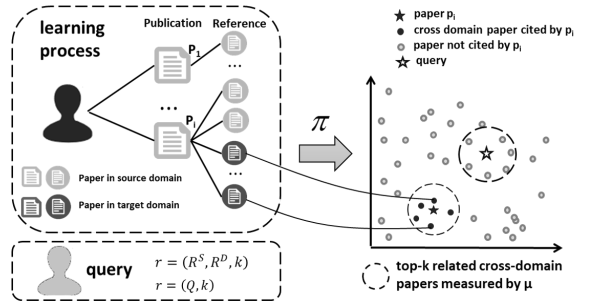


Figure 1: Semantic correlation and paper recommendation.

$$L = \sum_{(p,q)\in E, (p,q')\in \bar{E}} (\hat{\mu}(\pi(p), q) - \hat{\mu}(\pi(p), q')) \tag{1}$$

For the robustness purpose, we adopt the hinge loss function and add a regularization term $\gamma||\theta||_F^2$ to the objective function, $\gamma > 0$.

$$L = \sum_{\substack{(p,\,q)\,\in\,E \\ (p,\,q')\,\in\,\bar{E}}} (\max\{0, \hat{\mu}(\pi(p), q) - \hat{\mu}(\pi(p), q') + \epsilon\}) + \gamma||\theta||_F^2 \tag{2}$$

Now we discuss how to learn the mapping function $\pi(x : \mathbb{R}^{|\mathbf{T}^\mathbf{S}|}) \mapsto \mathbb{R}^{|\mathbf{T}^\mathbf{D}|}$. Theoretically, any neural network that satisfies the above metrics should work. It adopts the semantics based paper representation ζ_p^S as input $\mathbf{x}$ and outputs a vector $\mathbf{y} : \mathbb{R}^{|\mathbf{T}^\mathbf{D}|}$. We now define our multi-layer perceptron network and discuss several variants in the experiment section. The last layer representation is the desired embedding ζ_p^D of p in the semantic space of the target domain **D**.

$$\zeta_p^D = \pi(\zeta_p^S) = MLP(\zeta_p^S) \tag{3}$$

The difference with a direct representation ζ_q^D for paper q in domain D is that the cross-domain semantic representation ζ_p^D does not support the inheritance constraints. So we need to determine how to evaluate their correlations or distance $\hat{\mu}$. The neural network method is appropriate since it can theoretically simulate any function. It uses the concatenation of ζ_p^D and ζ_q^D and outputs their semantic distance. Formally, $\hat{\mu}(\mathbf{x}, \mathbf{y}) \mapsto R^+$. The actual formulas are similar to the neural function π and the details are omitted here. Thus the objective function is modeled as equation 4. Finally, the cross domain semantic correlation is the combination of the domain semantics mapping function, the correlation metric, and the ranking mechanism, denoted by $\rho = (\pi, \hat{\mu}, \geqslant)$.

$$L = \sum_{\substack{(p,\,q)\,\in\,E \\ (p,\,q')\,\in\,\bar{E}}} \max\{0, \hat{\mu}(\zeta_p^D, \zeta_q^D) - \hat{\mu}(\zeta_p^D, \zeta_{q'}^D) + \epsilon\} + \gamma||\theta||_F^2 \tag{4}$$

4.2 Cross-domain Recommendation

Since there are very large numbers of publications every year, we choose papers published in top conferences or high reputation journals to ensure paper quality. Then we focus on the most semantically related papers for recommendation. In this section, we discuss how to recommend cross-domain papers based on domain semantics and the influence function, as shown in Figure 1.

We consider two representative forms: a retrieval request in the form of a few keywords Q, and a list of papers as a user latent interest for an academic service provider to recommend papers based on one's historical publications R^S and references R^D in the target domain. It is possible in practice to obtain these papers since there are quite a few websites providing related academic services, such as CNKI, Google Scholar, and ResearchGate. We now formalize the cross-domain paper recommendation problem (**CPR** for short).

Definition 2. **CPR**. *Given the domain semantics* $\mathbf{DS^S}$ *and* $\mathbf{DS^D}$, *the correlation function* $\rho^{S\to D}$, *and a user request* $\mathbf{r} = (R^S, R^D, k)$ *or* $\mathbf{r} = (Q, k)$, *where* $R^S \subset \mathbf{P^S}, R^D \subset \mathbf{P^D}, Q \subset \mathbf{W}$ *and* $k \in N^+$ *is the user preference on the number of recommended papers, the cross-domain paper recommendation (***CPR***) problem is to find the top-k related papers from the candidate paper set* P^D *in domain* $\mathbf{D}$.

For $p \in R^S$, we compute the probabilistic representation ζ_p based on $\mathbf{DS^S}$. The publication based interest is computed as the expectation of the representations of papers in R^S, $I_p = |R^S|^{-1} \sum_{p \in R^S} \zeta_p$. Similarly, we have the reference based interest $I_r = |R^D|^{-1} \sum_{r \in R^D} \zeta_r$. After I_p is mapped to the target domain by the mapping function π, the overall cross-domain interest is computed as their combination, i.e $I = \gamma * \pi(I_p) + (1 - \gamma) * I_r$, where $\gamma \in [0..1]$ is a balance factor.

Considering the general case of a query, the keywords can be selected from the source domain or the target domain, denoted by Q^S and Q^D, respectively. Based on the word-specific probabilistic distribution φ over nodes, as discussed in Section 3.1, each keyword is computed as the probabilistic distribution over the corresponding domain semantics. Formally, $I_w^S = \frac{\sum_{w \in Q^S} \varphi_w^S}{|Q^S|}$, where φ_w^S is the word-specific probabilistic distribution on $\mathbf{DS^S}$. Similarly, we compute the interest for the target domain I_w^D against $\mathbf{DS^D}$. The combined query interest is computed as $I = \gamma * \pi(I_w^S) + (1 - \gamma) * I_w^D$.

Then, for each candidate paper $q \in P^D$, we compute the probabilistic representation ζ_q^D over $\mathbf{DS^D}$ and adopt $\hat{\mu}$ to compute the distance between a user interest and q. The top-k related papers are then recommended. In practice, many functions can be used to replace the functions in ρ, such as the *cosine similarity* for correlation metric μ, Euclidean distance for distance function $\hat{\mu}$, a linear mapping function for π.

5 EXPERIMENTAL EVALUATION

5.1 Datasets and Comparison Methods

We select two kinds of representative forms of categories and three real datasets. Considering the hierarchically organized categories, we adopt the 1998 ACM Computing Classification System (ACM CCS)[27] as the discipline classification, which is widely adopted in many *Computer Science* datasets. For example, $h.3.7$ refers to *Information Systems* → *Information Storage and Retrieval* → *Digital Libraries*. Each subdivision has its own concepts and theories and today there is an increase in knowledge influence between them. For example, advances in the domain "Machine Learning" highly influence the domain "Security" (e.g., "Deep Learning" methods are today used to detect Malicious Code). The crawled categories include 11 subdivisions and each contains 3 levels.

Table 1: Statistic on ACM Dataset.

Subdivision	Abbr.	#node	#paper
General Literature	A	5	117
Hardware	B	56	621
Computer Systems Organization	C	28	4179
Software	D	48	6300
Data	E	8	335
Theory of Computation	F	27	1614
Mathematics of Computing	G	21	1326
Information Systems	H	43	10428
Computing Methodologies	I	75	5908
Computer Applications	J	10	911
Computing Milieux	K	43	2096

We adopt two datasets related to CCS for experimental verification. One is about the publications in the ACM Digital Library (ACM for short), which contains 43380 conference and journal papers and was crawled by Tang et al. [37]. Each paper has a title, authors, an abstract, keywords, tags corresponding to CCS categories, and citation information. The statistics on the categories involved in each paper are shown in Table 1, which shows the number of nodes and papers in each subdivision against CCS level two. The second dataset is the patent database released by United States Patent and Trademark Office (PT for short)[17]. For our experiments we used a subset with a total of 70090 released patents from Jan 2017 to Dec 2017. Each patent contains the ownership, mark characteristics, classification, prosecution events, references, renewal, and maintenance history.

We also consider higher-level knowledge propagation between disciplines. The third dataset is crawled from *Scopus* and contains papers from 27 disciplines that are considered as categories, such as *Medicine*, *Social science*, *Computer science*, and etc. We select 328012 papers from discipline *Medicine* and 200222 papers from *Biochemistry, Genetics and Molecular Biology* within 2008 to 2017. Each paper has a title, authors, an abstract, keywords, references and discipline tags.

Since ours is the first approach for cross-domain paper recommendation, there is no previous approach that solves our exact problem. So, for comparison purposes we choose approaches developed for a problem most close to ours, such as the methods for reference recommendation and cross-platform recommendation methods for e-commerce. We then make some modifications to those approaches to adapt them to our scenario and thus be comparable to our approach. The selected baseline methods and different variants of our method are listed below. All the parameters of the baseline methods are empirically set to their optimal values.

TRRec[34]: It recommends cross-domain papers using the tf-idf vectors of papers to compute their relatedness. User interests are modeled by their publications and citations.

WNMF[41]: This is the weighted nonnegative matrix decomposition method, where each entry in the matrix is set to 1 if the researcher has cited the paper, and to 0 otherwise. The feature number is set to 10.

Table 2: Methods comparison on cross-domain recommendation.

Dataset	k	Method											
		TRRec	WNMF	TopicRec	MdRec	DsRec	CATN	GANRec	NCNRec	LCPR	CPR-pub	CPR-cite	CPR
ACM C-D	20	0.780	0.793	0.826	0.831	0.834	0.833	0.847	0.848	0.850	0.874	0.880	**0.901**
	30	0.769	0.786	0.820	0.826	0.835	0.829	0.832	0.838	0.843	0.851	0.877	**0.885**
	50	0.682	0.727	0.730	0.740	0.743	0.745	0.750	0.748	0.755	0.790	0.803	**0.825**
ACM H-I	20	0.791	0.833	0.839	0.838	0.850	0.841	0.852	0.859	0.867	0.879	0.886	**0.912**
	30	0.785	0.799	0.801	0.792	0.828	0.830	0.839	0.841	0.849	0.861	0.880	**0.893**
	50	0.726	0.732	0.765	0.761	0.780	0.780	0.781	0.787	0.793	0.825	0.848	**0.890**
Scopus	20	0.680	0.693	0.706	0.731	0.734	0.742	0.754	0.760	0.770	0.780	0.789	**0.793**
	30	0.659	0.676	0.680	0.706	0.715	0.719	0.721	0.747	0.750	0.769	0.773	**0.786**
	50	0.542	0.627	0.650	0.690	0.693	0.691	0.710	0.713	0.730	0.741	0.760	**0.773**

TopicRec[28]: It clusters the topics of researchers' publications, and computes the similarity of the topic proportions between researchers and papers.

MdRec[42]: An active learning strategy that considers both specific and independent knowledge over all domains so as to address the data sparsity in a multi-domain scenario.

DsRec[21]: A domain-sensitive recommendation algorithm to predict a rating while exploring the user-item subgroups.

CATN[43]: A cross-domain product recommendation system. It extracts multiple features for each user and item from the review documents, and learns the correlation between domains with an attention mechanism using the overlapping users in two domains.

GANRec[7]: A generative adversarial network based heterogeneous bibliographic network representation model, which incorporates citation relationship and paper contents to learn optimal representations of them, and recommend the top ranked papers as references by measuring the similarity scores.

NCNRec[13]: A neural citation network based paper recommendation model, which adopts a flexible encoder-decoder to represent the paper context, and uses a max time delay neural network with an attention mechanism to recommend the most related papers.

CPR: Our model, using both publications in source domain and references in target domain to learn the research interest.

LCPR: A variant of **CPR**, which learns the knowledge propagation based on the linear mapping method.

CPR-pub: A variant of **CPR**. Only the source domain publications are used to learn a researcher's interest. This is appropriate for users who are not familiar with the target domain.

CPR-cite: Our model with the researcher's references adopted as the target domain interest, which is appropriate for users who are interested in some specific cross-domain papers.

5.2 Cross-domain Recommendation

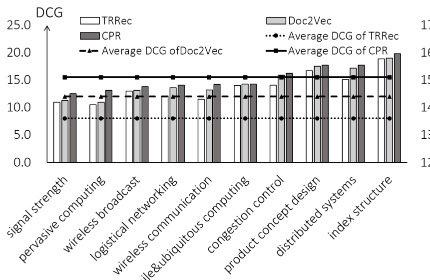
We verify the proposed method by considering two representative forms: personalized recommendation based on one's historical publications or references; and a retrieval request with keywords.

5.2.1 Cross Hierarchical Domain Recommendation. We first verify the performance of recommendation cross hierarchical domains. We use the CCS hierarchies and the ACM data set. Then we select several pairs of domains, where C and H are chosen as the source domains, D and I as the target domains. For each domain, the number of topics K highly affects the results, which is set according to the perplexity of our topic model, i.e. 15, 25, 22 and 40 for C, D, H, and I, respectively. The hyper-parameters α, α' and γ are sampled by Metropolis-Hastings[10], and the prior distributions of $p(\alpha)$, $p(\alpha')$ and $p(\gamma)$ are gamma distributed with a shape parameter of 1 and a scale parameter of 100. We choose 80 percent data in these domains to learn the domain semantics and correlation functions, and use the others for testing.

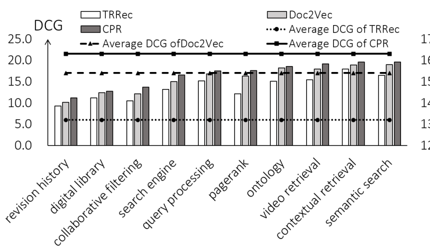
Because of the limitation of datasets, such as the shortage of experts whose publications and references are associated with tags at the same time, we select 300 researchers who have published at least 3 papers in source domain and cited at least 3 papers in target domain. For each user, m papers are randomly selected from her historical publications or references as her interest. In the experiments, for most users, we set $m = 3$. The ratio of positive samples E and negative samples $\overline{E}$ is set 1:7.

Since there are a large number of papers in a domain, for most researchers, the numbers of their publications are comparatively very small. To make the comparison results easily understood, we prepare k candidate papers for each user that contain at least one of her actually cited paper. We adopt the metric *nDCG@k* [16] to measure the recommendation results, since it is specifically designed for a recommendation approach that takes into account two important human-centered insights: (i) users are able to read only a limited number of recommended items and their interests follow a decay trend along with the ranking of items in the list; (ii) there is a limitation on the capacity that a device has for displaying results, such as screen space. Formally, $nDCG@k = \frac{DCG@k}{IDCG}$, in which $DCG@k = \sum_{i=1}^{k} \frac{rel_i}{log_2(i+1)}$, $rel_i = 5$ if the i-th paper is actually cited by the researcher, otherwise $rel_i = 0$. The comparison results with baseline methods are shown in Table 2, where we can see *CPR* outperforms the other methods on all datasets. Besides, the performance increases when incorporating both publications and references as user interest rather than only one of them.

5.2.2 Cross Discipline Recommendation. We use the *Scopus* dataset for this experiment. *Medicine* is selected as the source domain, while *Biochemistry, Genetics and Molecular Biology* are the target domains. 80% data in these domains are used for training. We randomly select 50 researchers in *Scopus* from the source domains to verify the performance. The settings of the hyper-parameters are the same as the settings for the experiment with the ACM dataset.



(a) Recommendation on C and D.



(b) Recommendation on H and I.

Figure 2: Cross-domain paper retrieval with keywords.

The topic numbers of domain *Medicine* and *Biochemistry, Genetics and Molecular Biology* are set to 15 and 7, respectively. The results, listed in the last 3 rows of Table 2, show that our model performs better than the baseline methods.

5.2.3 Evaluation on User Retrieval. To verify this general form, we choose the same domains of the previous experiments and adopt the keywords used by the highly cited papers in these domains. For each query, we recommend the most related 100 papers in the target domain D and I, respectively. Since many academic recommendation methods are not applicable for this retrieval form, the comparable baseline methods are listed below.

TRRec[34]: Papers are represented by their tf-idf vectors.

Doc2Vec[19]: An unsupervised algorithm that learns fixed-length feature representations of papers.

DCG@k is adopted to measure the cross-domain retrieval result. $DCG@k = \sum_{i=1}^{k} \frac{rel_i}{log_2(i+1)}$, where $k = 100$, $rel_i = 5$ if the i-th paper in the recommended list is cited by the papers in source domain, otherwise $rel_i = 0$. In Figure 2, we present the average performance of comparative models, as well as some query instances with key word. Besides the dominant results by our methods, there is an interesting phenomenon in that the performance on the queries involving human factors like "semantic search" is better than queries mentioning technologies like "collaborative filtering", which might be attributed to the fact that human factors has stronger domain characteristics, while the technologies applied to multiple domains have weak domain dependence.

Table 3: The performance of CPR on different settings.

Dataset	Model	Positive: negative			m		
		1:1	1:7	1:10	1	3	5
ACM C-D	CPR-pub	0.679	**0.854**	0.799	0.759	0.855	**0.871**
	CPR-cite	0.702	**0.841**	0.753	0.771	0.862	**0.862**
	CPR	0.791	**0.870**	0.780	0.761	0.869	**0.875**
Scopus	CPR-pub	0.720	**0.754**	0.730	0.664	0.754	**0.770**
	CPR-cite	0.742	**0.761**	0.753	0.681	0.761	**0.792**
	CPR	0.741	**0.769**	0.754	0.692	0.769	**0.820**

5.3 Evaluation of Parameter Influence

To quantify the influence of parameters, we perform the recommendation task on ACM and *Scopus* datasets with different settings. As the results in Table 3 show, *CPR* performs best when the ratio between positive and negative samples is 1 : 7 and increases with the increasing number of papers chosen as one's interest. In *Scopus*, *CPR-cite* performs better than *CPR-pub*, which shows that the cross domain influence is clearly recognized by authors' references. Comparatively, in *ACM*, there is no obvious difference between them. This might be due to two reasons: the flexible extent of hierarchical categories; the knowledge difference between fine grained domains is less distinct than cross disciplines.

5.4 Understanding the Results

Interpretable results are critical for the human centered paper recommendation task. Although our model is an extension of the topic model where the latent topics in the generative model cannot be directly mapped to an understandable space, the introduction of discipline categories, defined by experts, makes the domain semantics easily understood by researchers. Compared to previous approaches to interpretability, our model supports the discipline knowledge without sacrificing the accuracy of the learning results.

Understanding domain semantics. To check whether domain semantics *DS* reveal the difference of word usages between categories, we verify whether the *DS* representation of a paper can be used directly for predicating its tags. The ground truth is the tags that the authors have actually associated with the paper. For each paper p, we select the maximum probabilistic label path using our domain semantics extraction model. The selected comparison methods are as follows:

TRRec[34]: Papers are represented by their tf-idf vectors. A node vector is the expectation of its corresponding labeled paper vectors. The tags of a testing paper are predicted by the cosine similarity between node vectors and the paper vector.

Doc2Vec[19]: A neural network method which adopts Doc2Vec to train paper embeddings. The tags are predicted using the same approach as TRRec.

LLDA[30]: The LLDA model is trained without considering the hierarchies. The tags of a testing paper are predicted using the LLDA inference model.

Since the loss is worse when a parent node is correctly predicted than a child node, we introduce the average DCG as the measurement, $DCG = \sum_{i \in levels} \frac{rel_i}{i}$, where $rel_i = 2$ if the label on the i-th level is predicted correctly and 0 otherwise. The results in Figure 3

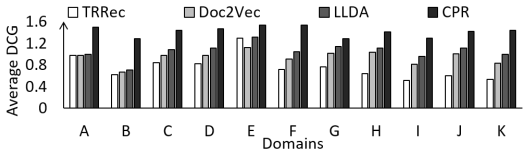
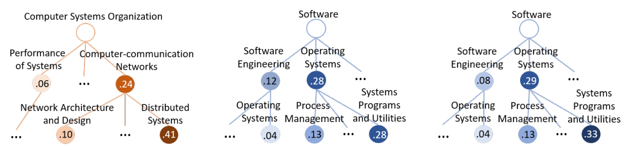


Figure 3: Methods comparison on tag predication.



(a) I_p in C (b) $\pi(I_p)$ in D (c) I_r in D

Figure 4: User interests over the classification trees.

show that our method is better than the neural network-based method and the usual topic model.

To illustrate the domain semantics in an intuitive way, we compare word semantics in a domain dependent context with a general semantic environment. We select domain *Information Systems* (H) and the probabilistic associations of word with domain semantics as the specific word representation, while the word embeddings generated by the $Word2Vec$[25] method on a large corpus *wikipedia* are regarded as the general vector. Then we reduce the word vectors into a 2-dimensional visual space by T-SNE [22], shown in Figure 5, where each node denotes a word. The closer the nodes, the more related the word semantics. As marked colored or cross nodes, the word "neural network" is highly correlated with "backpropagation" in the general space, but less related in a domain specific space. Comparatively, "neural network" is more related to "memory management unit" in domain H than in the general space. Such change reduces the ambiguity of words in different domains.

Understanding the semantic correlation cross domains. To understand how semantic correlation between domains change the word relatedness, we choose domain *Computer Systems Organization* (C) and *Software* (D) and several keywords in each domain. The keywords are represented by the probabilistic association with the corresponding domains and then are mapped to the 2-dimensional space for visualization by $T-SNE$, shown in Figure 6. The keyword *ubiquitous computing* in domain C (*uc* for short) is marked purple and its top-100 related words in C are marked pink. After mapped to domain D, their colors remain the same. The results show that their correlations are changed. To clearly illustrate this result, we select 4 keywords from the 100 words and mark them as purple crosses: *statistical analysis*, *energy efficiency*, *region management* and *portability*. In domain D, some words are still highly correlated with *uc*, such as *statistical analysis* and *energy efficiency*, because there are many papers mentioning them at the same time [1]. By comparison, *region management* and *portability* become less related to *uc*. Similarly, for the keyword *privacy*, colored orange, the top-100 related words are marked yellow, among which 4 keywords are marked orange crosses. There are similar conclusions as *uc*. That

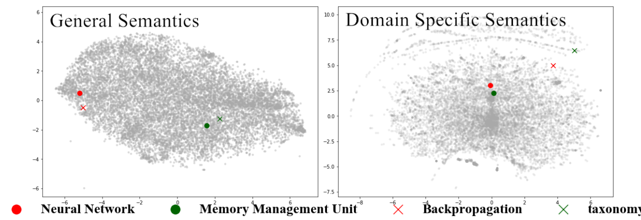


Figure 5: Word semantics in different semantic spaces.

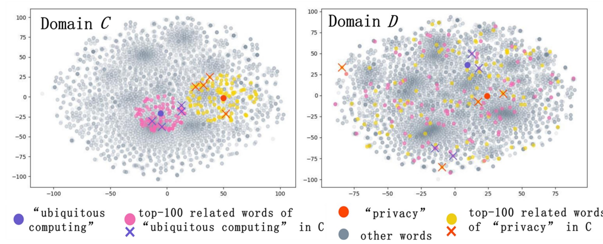


Figure 6: The word relatedness after semantic correlation.

is, *service-level agreements* and *resource discovery* keep close with *privacy* in D after knowledge propagation, while *social network* and *process algebra* become less related to *privacy*.

We now illustrate how a user interest is transferred cross domain in an intuitive way. We take researcher *K. Walsh* as an example. His interests in the source domain C are learned by his publications in C and then transferred to D by our model, shown as the shaded nodes over T in Figure 4(a) and 4(b), respectively. The darker the node, the higher the interest. The results in Figure 4(c) are obtained by using his actual references in D. By comparing Figures 4(b) and 4(c), we can see that the semantic correlation learnt by our method well match his cross-domain interest. The most interested keywords of *K. Walsh* are listed in Table 4. In the left part of the table, the frequently used keywords are the statistics on his publications and references in C and D, respectively. Then we select the most representative keywords which are associated with the distributions in Figure 4, and list them in the right part of the table. All the keywords are ranked by the frequency or the correlation against his interests. In the upper part, we highlight the same keywords on the left and right sides in red. From the red keywords, we can see that his research interest learnt by our model well matches his publications; the rest of the keywords and the exact recommendation results reported in Section 5.2.1 prove that our method can model his latent interest. The results in the bottom part show that the representative keywords against both $\pi(I_p)$ and I_r match his references, and our method can model his latent cross-domain interest.

5.5 Model Discussion

5.5.1 Model Practicality. Both the domain semantics model and correlation mapping function are pre-trained. Theoretically, the complexity of training the domain semantics model is linearly bounded by the size of samples, the height of classification tree and the paper length. But in fact, it is efficient since the method is a random process and the probability of a node to be visited is

Table 4: *K. Walsh*'s frequently used keywords and the modeled interests.

Top used keywords in pubications in C: filesharing, credence, p2p, reputation, gnutella, flow-based, decoy, network, client, sensor	**Representative keywords against I_p:** distributed, network, wireless, applications, mobile, sensor, node, p2p, reputation, grid
Top used keywords in references in D: software, engineering, visualization, propagation, history, single-view, compiler, taxonomy, stakeholder, task-specific	**Representative keywords against $\pi(I_p)$:** software, engineering, propagation, java, monitor, visualization, window, memory, management, i/o
	Representative keywords against I_r: software, engineering, propagation, java, visualization, monitor, window, compiler, memory, management

p^h, where h is the depth from the root to this node and p is the probability associated with this node. Once a node is not sampled, our sample operation will stop browsing its subtrees (see lines 15-18 in Algorithm 1). That is, for each paper, there is no need to do a full breadth-first search. In a real time recommendation process, the predictive process for a paper is similar to this random process. As for the semantic correlation between disciplines, our model supports the flexible scope of domain, that is to say a domain can be either a hierarchically organized discipline or fine grained categories. Their semantic correlations are learned in an efficient way, which are represented as the probabilistic mapping function between the whole categories. In a query, the user is allowed to specify the source and target domain in the form of either a subtree or a leaf. Then the recommendation algorithm chooses the papers from the scope of user defined domains.

The second aspect is the update of model. This process can be periodically invoked at the system free time, such as weekly or monthly update, and select a part of the corpus on newly published papers.

The third aspect is related to the efficiency in answering user queries, which are performed online. There are some strategies to accelerate this response, such as preparing a set of candidate papers for frequently used key words, which is also a widely adopted way for recommendation task or web search since in most cases user queries are in the form of keywords.

5.5.2 Model Robustness. As for the robustness of our model, we adopt a generative process which supports flexible scopes of domains. In practice, there is not a clear boundary for a scientific domain. Users may choose different levels of discipline categories, such as a paper may be associated with the CCS tags *Human-centered computing* → *Collaborative and social computing* or *Information systems* → *Information retrieval*. Besides, a discipline classification may be adjusted according to scientific progresses. Thus it is reasonable in our model to extract domain semantics on the basis of an existing discipline classification. Any new classification version can work for real usages through our model. Also the restriction on choosing a label path in the supervised learning process supports the hierarchical tags and the extensible domain semantics.

5.5.3 Reuse of Semantics. To check the reuse of learnt domain semantics and correlations, we verify our model on a new dataset containing documents with tags on the same categories but not used in the learning process. We use the knowledge learnt by *ACM* and the dataset *PT* to test. We randomly select the patents in domain C

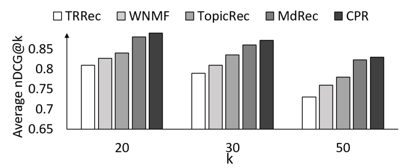


Figure 7: Recommendation between *C* and *D* on PT.

with references in domain D as interest query. For each patent, we prepare a set of patents in D that contains at least one patent being actually cited. Each time we map the interest to domain D, and rank the candidates based on the semantic correlations learnt by *ACM*. The comparison methods are reported in Section 5.1. The results in Figure 7 show that the learnt semantics can be applied to a new dataset without any further learning. Since scientific categories are widely adopted by the science community, the semantic and correlation knowledge can be acceptable in practice.

6 CONCLUSION

In this paper, we address the challenging problem of cross-domain paper recommendation. We introduce the notion of *domain semantics* to learn the word semantics against a hierarchical classification system, which reduces the discrepancies on the word usage in different domains. The intrinsic correlations between domains are represented as semantic correlations. A user interest is specified as either a set of papers or keywords. After mapping the interest to the target domain, papers are recommended based on the relatedness with them. We evaluate our method on real datasets and the results show that it outperforms other related methods. Domain semantics supported by our model can be easily understood by researchers, which is critical for the human-centered paper recommendation tasks. We discuss the feasibility, robustness and reusability of our model, and illustrate that such a recommendation system is applicable for real scenarios.

As future work, we plan to investigate the optimization techniques on academic recommendation cross multiple domains, such as real-time strategies. We also plan to use our model for other applications, such as cross-platform recommendation for e-commerce, especially for the hierarchically organized classification system.

ACKNOWLEDGMENTS

This work was supported by the Major Project of NSF Shandong Province (ZR2018ZB0420), the National Key R&D Program of China (2018YFC0831401), and the National NSF of China (91646119).

REFERENCES

[1] Eyhab Al-Masri and Qusay H Mahmoud. 2008. Investigating web services on the world wide web. In *Proceedings of the 17th international conference on World Wide Web*. ACM, 795–804.

[2] K. Balog, Filip Radlinski, and Shushan Arakelyan. 2019. Transparent, Scrutable and Explainable User Models for Personalized Recommendation. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 265–274.

[3] Ye Bi, Liqiang Song, M. Yao, Z. Wu, J. Wang, and J. Xiao. 2020. DCDIR: A Deep Cross-Domain Recommendation System for Cold Start Users in Insurance Domain. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1661–1664.

[4] Ye Bi, Liqiang Song, M. Yao, Z. Wu, J. Wang, and J. Xiao. 2020. A Heterogeneous Information Network based Cross Domain Insurance Recommendation System for Cold Start Users. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2211–2220.

[5] D. Blei, A. Ng, and M. Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[6] K. Bollacker, S. Lawrence, and C. Giles. 2000. Discovering relevant scientific literature on the web. *IEEE Intelligent Systems and their Applications* 15, 2 (2000), 42–47.

[7] Xiaoyan Cai, Junwei Han, and Libin Yang. 2018. Generative adversarial network based heterogeneous bibliographic network representation for personalized citation recommendation. In *Thirty-Second AAAI Conference on Artificial Intelligence*. 5747–5754.

[8] Yixin Cao, Xiang Wang, Xiangnan He, Zikun Hu, and Tatseng Chua. 2019. Unifying Knowledge Graph Learning and Recommendation: Towards a Better Understanding of User Preferences. In *The World Wide Web Conference*. 151–161.

[9] T. Chakraborty, N. Modani, R. Narayanam, and S. Nagar. 2015. Discern: a diversified citation recommendation system for scientific queries. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 555–566.

[10] S. Chib and E. Greenberg. 1995. Understanding the Metropolis-Hastings Algorithm. *The American Statistician* 49 (1995), 327–335.

[11] Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S. Weld. 2020. SPECTER: Document-level Representation Learning using Citation-informed Transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2270–2282.

[12] Paolo Cremonesi, Antonio Tripodi, and Roberto Turrin. 2011. Cross-domain recommender systems. In *2011 IEEE 11th International Conference on Data Mining Workshops*. Ieee, 496–503.

[13] Travis Ebesu and Yi Fang. 2017. Neural citation network for context-aware citation recommendation. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 1093–1096.

[14] M. Falagas, E. Pitsouni, G. Malietzis, and G. Pappas. 2008. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *The FASEB Journal* 22 (2008), 338 – 342.

[15] Jingyue Gao, Xiting Wang, Yasha Wang, and Xing Xie. 2019. Explainable Recommendation Through Attentive Multi-View Learning. In *The Thirty-Third AAAI Conference on Artificial Intelligence*. 3622–3629.

[16] K. Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* 20 (2002), 422–446.

[17] Jeeeun Kim and S. Lee. 2015. Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO. *Technological Forecasting and Social Change* 92 (2015), 332–345.

[18] Adit Krishnan, Mahashweta Das, Mangesh Bendre, Hao Yang, and Hari Sundaram. 2020. Transfer Learning via Contextual Invariants for One-to-Many Cross-Domain Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1081–1090.

[19] Quoc V Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning*. 1188–1196.

[20] Roy Ka-Wei Lee, Tuan-Anh Hoang, and Ee-Peng Lim. 2017. On analyzing user topic-specific platform preferences across multiple social media sites. In *Proceedings of the 26th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 1351–1359.

[21] J. Liu, Y. Jiang, Z. Li, X. Zhang, and H. Lu. 2016. Domain-sensitive recommendation with user-item subgroup analysis. *IEEE Transactions on Knowledge and Data Engineering* 28, 4 (2016), 939–950.

[22] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9, Nov (2008), 2579–2605.

[23] Tong Man, Huawei Shen, Xiaolong Jin, and Xueqi Cheng. 2017. Cross-Domain Recommendation: An Embedding and Mapping Approach.. In *IJCAI*. 2464–2470.

[24] J. Mcauliffe and D. Blei. 2008. Supervised topic models. In *Advances in neural information processing systems*. 121–128.

[25] Tomás Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*. 3111–3119.

[26] C. Moreira, P. Calado, and B. Martins. 2015. Learning to rank academic experts in the DBLP dataset. *Expert Systems* 32, 4 (2015), 477–493.

[27] E. Glinert T. Horton N. Mead R. Rada A. Ralston C. Rodkin B. Rous A. Tucker P. Wegner E. Weiss N. Coulter, J. French and C. Wierzbicki. 1998. Computing classification system 1998: Current status and future maintenance. *Computing Reviews* 39, 1 (1998), 1–5.

[28] M. Pazzani and D. Billsus. 2007. Content-based recommendation systems. In *The adaptive web*. Springer, 325–341.

[29] A. Perotte, F. Wood, N. Elhadad, and N. Bartlett. 2011. Hierarchically supervised latent Dirichlet allocation. In *Advances in Neural Information Processing Systems*. 2609–2617.

[30] D. Ramage, D. Hall, R. Nallapati, and C. Manning. 2009. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, 248–256.

[31] Bernard Rous. 2012. Major Update to ACM's Computing Classification System. *Commun. ACM* 55, 11 (Nov. 2012), 12.

[32] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and Philip S Yu. 2019. Heterogeneous Information Network Embedding for Recommendation. *IEEE Transactions on Knowledge and Data Engineering* 31, 2 (2019), 357–370.

[33] K. Sugiyama and M. Kan. 2015. A comprehensive evaluation of scholarly paper recommendation using potential citation papers. *International Journal on Digital Libraries* 16, 2 (2015), 91–109.

[34] Kazunari Sugiyama and Min-Yen Kan. 2010. Scholarly paper recommendation via user's recent research interests. In *Proceedings of the 10th annual joint conference on Digital libraries*. ACM, 29–38.

[35] Anu Taneja and Anuja Arora. 2018. Cross domain recommendation using multidimensional tensor factorization. *Expert Systems with Applications* 92 (2018), 304–316.

[36] Jiliang Tang, Suhang Wang, Xia Hu, Dawei Yin, Yingzhou Bi, Yi Chang, and Huan Liu. 2016. Recommendation with social dimensions. In *Thirtieth AAAI Conference on Artificial Intelligence*. 251–257.

[37] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 990–998.

[38] S Udayambihai and V Uma. 2018. Cross Domain Recommendation System Using Ontology and Sequential Pattern Mining. In *International Conference on ISMAC in Computational Vision and Bio-Engineering*. Springer, 1885–1893.

[39] Y. Xie, Y. Sun, and L. Shen. 2016. Predicating paper influence in academic network. In *Computer Supported Cooperative Work in Design (CSCWD), 2016 IEEE 20th International Conference on*. IEEE, 539–544.

[40] E. Yates and L. Dixon. 2015. PageRank as a method to rank biomedical literature by importance. *Source code for biology and medicine* 10, 1 (2015), 16.

[41] S. Zhang, W. Wang, J. Ford, and F. Makedon. 2006. Learning from incomplete ratings using non-negative matrix factorization. In *Proceedings of the 2006 SIAM International Conference on Data Mining*. SIAM, 549–553.

[42] Z. Zhang, X. Jin, L. Li, G. Ding, and Q. Yang. 2016. Multi-Domain Active Learning for Recommendation.. In *AAAI*. 2358–2364.

[43] Cheng Zhao, Chenliang Li, Rong Xiao, Hongbo Deng, and Aixin Sun. 2020. CATN: Cross-Domain Recommendation for Cold-Start Users via Aspect Transfer Network. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Virtual Event, China) *(SIGIR '20)*. New York, NY, USA, 229–238.