



ARTICLE

Joint Modeling of Citation Networks and User Preferences for Academic Tagging Recommender System

Weiming Huang^{1,2}, Baisong Liu^{1,*} and Zhaoliang Wang¹

¹Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo, 315211, China

²Inner Mongolia Metal Material Research Institute, Baotou, 014000, China

*Corresponding Author: Baisong Liu. Email: lbs@nbu.edu.cn

Received: 05 February 2024 Accepted: 17 April 2024 Published: 20 June 2024

ABSTRACT

In the tag recommendation task on academic platforms, existing methods disregard users' customized preferences in favor of extracting tags based just on the content of the articles. Besides, it uses co-occurrence techniques and tries to combine nodes' textual content for modelling. They still do not, however, directly simulate many interactions in network learning. In order to address these issues, we present a novel system that more thoroughly integrates user preferences and citation networks into article labelling recommendations. Specifically, we first employ path similarity to quantify the degree of similarity between user labelling preferences and articles in the citation network. Then, the Commuting Matrix for massive node pair paths is used to improve computational performance. Finally, the two commonalities mentioned above are combined with the interaction paper labels based on the additivity of Poisson distribution. In addition, we also consider solving the model's parameters by applying variational inference. Experimental results demonstrate that our suggested framework agrees and significantly outperforms the state-of-the-art baseline on two real datasets by efficiently merging the three relational data. Based on the Area Under Curve (AUC) and Mean Average Precision (MAP) analysis, the performance of the suggested task is evaluated, and it is demonstrated to have a greater solving efficiency than current techniques.

KEYWORDS

Collaborative filtering; citation networks; variational inference; poisson factorization; tag recommendation

1 Introduction

Tag recommendations help to limit the tagging vocabulary, thus improving annotation efficiency. Users can complete the tag attachment process with just a few clicks using the tag recommendation system. A tagging system is a crucial tool that helps users organize large quantities of academic information efficiently. Users can efficiently manage information and locate relevant resources using a tagging system. Users can tag articles individually in RefWorks (www.refworks.com), Connotea (www.connotea.org), CiteULike (www.citeulike.org) and Zotero (www.zotero.org). Zotero provides a free tagging system, as illustrated in Fig. 1. Tags selected by various individuals may be idiosyncratic and unpredictable. Moreover, different user actions can cause issues with entity recognition [1]. Tag



recommendations enhance annotation efficiency by restricting the tagging vocabulary. Users can finish the tag attachment procedure efficiently using the tag recommendation system with minimal effort.

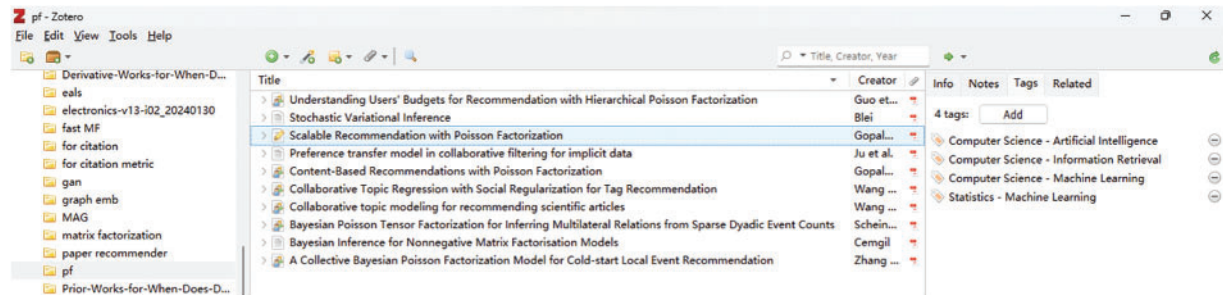


Figure 1: Zotero's personalized tagging system allows users to customize article tags. Personalized labelling service is provided on the right side to facilitate literature management

Based on the realized methods, the current label recommendation can be broadly classified into three categories: Content-based methods, co-occurrence-based methods, and user preference-based methods. Content-based methods directly utilize item content for label recommendation. Content-based label recommendation generally does not consider user information but only the association between content information and item labels, such as term frequency–inverse document frequency (TF-IDF) [2]. Later on, the industry has developed methods that consider the order of words, e.g., reference [3] uses a capsule network to encode the intrinsic spatial relationships between the parts and the whole that make up the viewpoint-invariant knowledge, and another example in [4] builds a convolutional neural network considering the document topic and global semantics and inputs textual and word-vector features to predict the most relevant labels for unlabeled or few-labeled assigned items. The content-based approach is mainly used in the cold-start phase. Co-occurrence-based methods primarily utilize the co-occurrence of tags between items (i.e., the item-pagination matrix) for tagging. The fundamentals of co-occurrence-based and user preference-based approaches are similar to Collaborative Filtering (CF) approaches. As in [5], deep latent embeddings are learned from different item auxiliary information using a Variational Autoencoder (VAE), and generative distributions can be formed on each auxiliary information by introducing latent variables parameterized by a deep neural network. As in [6], correlations between user attribute labels and concepts extracted from usable parts of scientific papers are defined based on statistical, structural and semantic aspects.

However, the object of study of tag recommendation is the three subjects of users, items and tags, as well as the interaction between the subjects. Users may pay attention to items in various fields and have different paging habits. Therefore, ignoring user preferences will simplify the model, but at the same time, it will inevitably cause some critical information to be missing, which affects the recommendation ability of the model, so some studies have begun to consider the interaction between the three subjects. For example, Huang et al. [7] used a Tag-aware Attentional Graph Neural Network (TA-GNN) to extract user-label interaction and item-label interaction from the user-label-item graph structure. Another example is [8], which uses an autoencoder network of learned hidden representations to encode multiple types of relationships between entities, i.e., between users and labels, between items and labels, and between labels, and uses a decoder component to reconstruct the original inputs based on the learned latent representations. Literature [9] uses a metric learning approach to explore the distance relationships between user \times item \times label triples and applies existing

metric learning-based approaches (i.e., latent relational metric learning (LRML), collaborative metric learning (CML), symmetric metric learning (SML)) to label recommendation.

However, users are accustomed to choosing their preferred labels for academic labelling recommendations, so papers read by similar users may be given the same label. For example, a mathematician may label the paper “Stochastic Variational Reasoning” [10] as “Statistics”, while a computer scientist may label the paper as “Machine Learning”. In addition, there is a citation relationship between articles, and usually, two articles with a citation relationship may be about the same topic, so that they may have the same label. Therefore, finding an effective way to combine citation networks, paper labelling and user paper matrices, remains challenging. This paper proposes new approaches to solve the multi-label recommendation problem for systems that provide closed-label recommendations. The main contributions of this paper are as follows:

To the best of our knowledge, this is the very first work to study multifactor Poisson factorization for tag-oriented recommender systems.

By extending the Poisson Factorization model, we propose to exploit the additivity of the gamma distribution to seamlessly integrate the three networks, namely, the paper labelling matrix, the citation network, and the user paper interaction matrix, into the same principle model.

Extensive experiments on real-world datasets have shown that the model significantly outperforms baseline models based on the paper label or user paper matrix. Furthermore, the computational complexity of the model is low.

[Section 2](#) of this paper describes the related work on tagging recommendation, Poisson factorization recommendation, etc. [Section 3](#) describes our Poisson Decomposition recommendation model with social regularization. [Section 4](#) describes using the coordinate ascent algorithm to infer the model parameters. Finally, [Section 5](#) validates the experimental performance of the model on the CiteULike dataset.

2 Related Works

This section provides an overview of the current research on tag recommendation and the Poisson factorization-based recommender system closely associated with this work.

2.1 Tag Recommendation

The article emphasizes the importance of tag recommendation systems in academic resources to alleviate information overload and improve user tagging experience [11]. It outlines three primary tag recommendation methods: Content-based, co-occurrence-based, and user preference-based approaches [1].

2.1.1 Content-Based Tag Recommendation

Content-based tag recommendation methods primarily focus on the association between an item’s content and its labels, often ignoring user information and behaviors patterns [12–14]. Initially, these methods treated the text as a “Bag of Words” (BoW) [15], using word frequency [16] for feature extraction and feeding these features into multi-label classifiers. However, this approach did not accurately capture the importance of each word, leading to the adoption of Term Frequency–Inverse Document Frequency (TF-IDF) [2] for more effective text representation. Despite advancements, these methods still overlooked the sequential structure of text, neglecting sentence-level information. To address this, deep learning techniques, such as capsule networks [3] and convolutional neural networks [4], have

been employed to consider both word-level and sentence-level information, enhancing the model's ability to predict relevant tags [13]. Additionally, external knowledge bases like knowledge graphs [17] and pre-trained models [18] have been integrated to improve the recognition of new concepts and to capture richer semantic information. However, these additions can introduce complexity and noise. Despite these efforts, content-based methods still struggle to reflect personalized user preferences and group intelligence.

2.1.2 Co-Occurrence-Based Tag Recommendation

Co-occurrence-based methods enhance tag recommendation by exploring correlations between tags, leveraging conditional probabilities and deep learning techniques like Variational AutoEncoders (VAE) [5] to learn latent embeddings and predict label recommendations [6]. These methods also consider tag structures, as seen in [14], which uses graph diffusion mechanisms and hierarchical tag structures [19] for better performance. Furthermore, label semantic-based approaches, such as those using Graph Convolutional Networks (GCNs) [20] and pre-trained word vectors like Global Vectors for Word Representation (GloVe), enrich the semantic information of labels [21], guiding the content representation for improved classification. To capture the complex interactions in tag recommendation, Neural Graph for Personalized Tag Recommendation (NGTR) [22] employs graph neural networks within a tensor decomposition model, integrating neighborhood representations to enhance the encoding of entity relationships.

2.1.3 Tag Recommendation Based on User Preferences

User preference-based methods in tag recommendation systems address the multi-label classification problem by considering the interactions among users, items, and tags. These methods often combine user and item representations, from user-to-user, user-to-item, and item-to-item interactions, to capture individual preferences and enhance model accuracy [23]. Techniques such as hierarchical attention models and graph neural networks integrate collaboration signals and generate richer entity representations [24]. Additionally, some works explore different vector spaces to learn entity representations and measure semantic correlations [25], such as using hyperbolic spaces and tangent space optimization [26]. To directly capture the interactions between users, items, and tags, methods like the Tag-aware Attentional Graph Neural Network (TA-GNN) [7] and variational self-encoders [8] have been proposed, which utilize attention mechanisms and metric learning to understand better and leverage these interactions [9]. This paper contributes to the field by proposing a collaborative filtering model that integrates tag co-occurrence and user preference, aiming to optimize the utilization of interaction data in academic tagging.

2.2 Poisson Factorization for Recommendation System

Poisson factorization is a matrix decomposition technique applied in recommender systems, notably for its effectiveness in handling sparse datasets. It has been utilized in models like Collaborative Topic Poisson Factorization (CTPF) [27] and Collaborative Topic Regression (CTR) [28] to integrate topic distributions with item representations and alleviate data sparsity. The multilayer Poisson decomposition model [29] is introduced to describe the combination of users' purchase budgets and preferences for items. The method reduces computational complexity and capitalizes on the value of long-tail items. Poisson factorization excels in sparsity modelling, diminishing the impact of unvisited items on predictions, and demonstrates flexibility, scalability, and robustness in the presence of noisy or missing data. These attributes make Poisson factorization a powerful tool for various applications, including recommender systems, text mining, and image processing. This paper proposes a novel

approach that regularizes the Poisson factorization of the paper-tag matrix using citation networks and user-paper interactions, offering improvements over existing methods that do not consider citation networks. It can fuse multiple contexts using the additivity property of the Poisson distribution.

3 Problem Description and Preparatory Knowledge

The two tasks closely related to tag recommendation are item recommendation and multi-tag classification, and part of the tag recommendation also draws on the idea of these two. As shown in Fig. 2a, the general item recommendation recommends a collection of items $I = \{i_1, i_2, \dots, i_n\}$ for the current user u_i based on the user's behavior and other information.

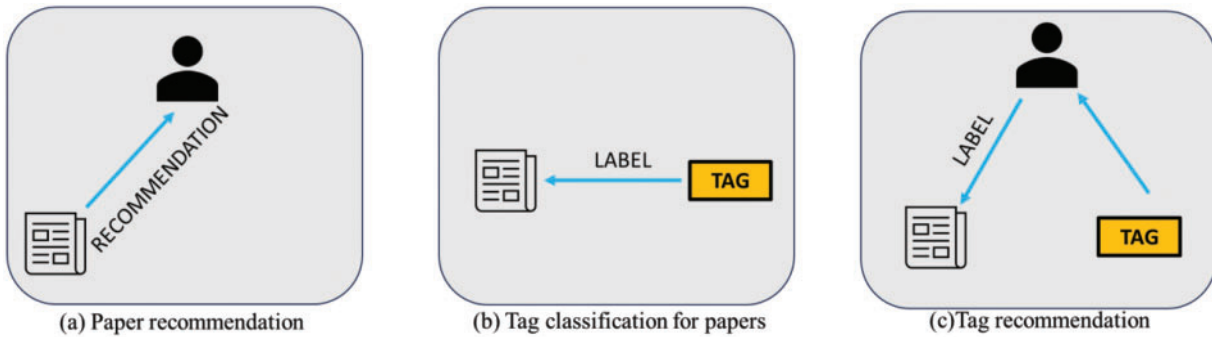


Figure 2: Paper recommendation, tag classification for papers and tag recommendation

The multi-tag classification problem generally refers to tagging a set of tags $V = \{v_1, v_2, \dots, v_m\}$ for the current item i_j based on its content information, as shown in Fig. 2b.

Tag recommendation refers to the process that when the current user u_i wants to tag the target item i_j (which often contains content information such as text, images or videos), the system automatically recommends a tag set $V = \{v_1, v_2, \dots, v_m\}$ to the user for paper tagging. This tagging process is affected by the habit of user and project information. The relationship between user u_i and project i_j , tag v_m is shown in Fig. 2c.

We base our study on the CiteULike dataset, which is collected from the CiteULike website. CiteULike provides three entities: Users, papers, tags, and their multidimensional relationships. To describe the formal model of CiteULike, we first give the relevant definitions:

Definition 1. User paper interaction network, the network formed by the user's clicks, favorites and evaluation records of papers, can be expressed as an undirected graph structure, denoted as $R = (U, I, S)$, where U denotes the set of users, I denotes the set of papers, and S denotes the set of user behaviors such as clicks, favorites, evaluations, etc. The CiteULike does not differentiate between multiple behaviors, and thus can be taken as $R = (U, I)$.

Definition 2. Citation network, a directed graph network formed by paper-to-paper citation relationships, denoted as $G = (I, E)$, where I denotes the set of papers and E denotes the set of edges of citation relationships between papers, i.e., $E = \{(i_k, i_j) | i_k, i_j \in I\}$.

Definition 3. Paper-tag network, an undirected graph network formed by the relations between papers and tags, it denoted as $C = (I, V)$, where I denotes the set of papers and V denotes the set of paper tags.

This paper further gives a heterogeneous data utilization model that fuses the three relationships of definitions 1~3. This model can also be generalized to similar industry data.

Gopalan et al. [27] proposed the Bayesian Poisson Factorization (BPF) model. The BPF model assumes that the user's responses to items (e.g., clicks, favorites, purchases, follows, ratings, comments, retweets, etc.), y_{ui} , follow a Poisson distribution:

$$y_{ui} \sim \text{Poisson}(\eta_u^T \theta_i) \quad (1)$$

where η_u and θ_i are K-dimensional non-negative vectors following a *Gamma* distribution, i.e.,

$$\eta_u \sim \text{Gamma}(\lambda_{ua}, \lambda_{ub}) \quad (2)$$

$$\theta_i \sim \text{Gamma}(\lambda_{ia}, \lambda_{ib}) \quad (3)$$

where $\lambda_{ua}, \lambda_{ub}$ is the shape parameter of the Gamma distribution and $\lambda_{ia}, \lambda_{ib}$ is the scale parameter of the Gamma distribution. We can infer the values of η_u and θ_i by Gibbs sampling or variational inference.

4 Joint Modeling of Citation Networks and User Preferences

4.1 Context-Enhanced Poisson Factorization Models

We take advantage of the additivity of the gamma distribution and use the Poisson factorization model as a general framework, as modelled in the following equation:

$$y_{vi}|y_{-vi} \sim \text{Poisson}\left(\eta_v^T \theta_i + \sum_{j=1}^{|J|} \delta_{vj} s_{ij} y_{ji} + \sum_{w=1}^{|W|} \tau_{vw} s'_{iw} y_{wi}\right) \quad (4)$$

where η_v is the hidden feature vector of the tag, θ_i is the hidden feature vector of the paper, s_{ij} is the user preference-based similarity between paper j and paper i , δ_{vj} is the attenuation coefficient of the user preference-based similarity, s'_{iw} is the citation network-based similarity between paper w and paper i , and τ_{vw} is the decay coefficient of citation network similarity. y_{ji} is the set of papers that have been labelled i , and y_{wi} is the set of papers that have been labelled i as well. In fact, the two sets are consistent. However, due to the different semantics of the two summation terms $\sum_{j=1}^{|J|} \delta_{vj} s_{ij} y_{ji}$ and $\sum_{w=1}^{|W|} \tau_{vw} s'_{iw} y_{wi}$, to semantically distinguish these two terms, we take the middle variable of these two terms to be different. The probability plot of the model is shown in Fig. 3. From the above equation, we can see that when the value of y_{ji} or y_{wi} in the matrix is 0, there is no need to calculate $\sum_{j=1}^{|J|} \delta_{vj} s_{ij} y_{ji}$ or $\sum_{w=1}^{|W|} \tau_{vw} s'_{iw} y_{wi}$. For sparse matrices, this feature of Poisson factorization can save a large part of the computation. In the massive paper queue, the model call data is often only a few of them, so for the academic platform-oriented tagging recommendation, the Poisson factorization algorithm is a very suitable method.

4.2 Article Similarity Metrics Based on User Preferences

This section introduces symmetric metapaths and computes the path similarity of the head and tail nodes (PathSim). The task can be described as, for example, finding the path similarity in the meta-path 'paper \rightarrow reader \rightarrow paper'. The formula for path similarity is as follows:

$$s(x, y) = \frac{2 \times |p_{x \rightarrow y}|}{|p_{x \rightarrow x}| + |p_{y \rightarrow y}|} \quad (5)$$

where $s(x, y)$ is the path similarity between node x and node y . $|p_{x \rightarrow y}|$ represents the number of paths between node x and node y in the case of a meta-path P . Intuitively, when there are more paths between nodes, the two nodes are more similar. The denominator in the formula is the number of total paths of the two nodes back to themselves as a normalization term.

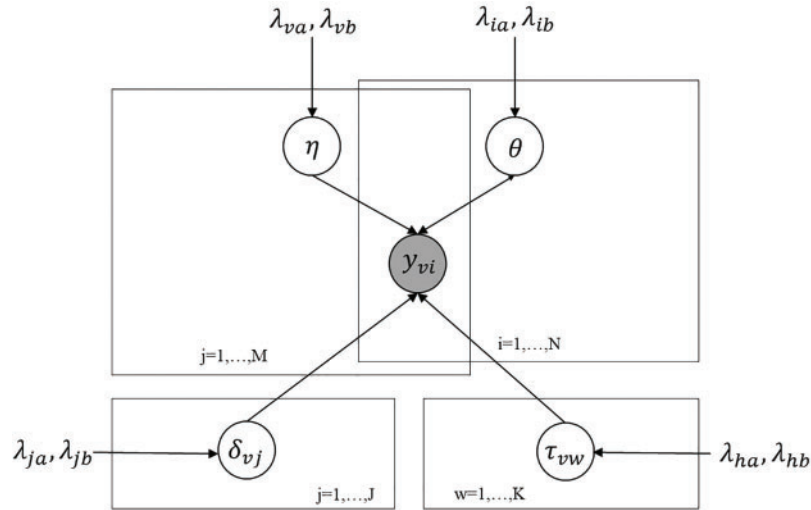


Figure 3: Probabilistic graphical models

Take the meta-path ‘paper → reader → paper’ as an example. The goal is to find the similarity of the paper. Suppose there are readers $A1, A2, A3, A4$, and papers $B1, B2, B3$, and their interaction matrices are shown in the following [Table 1](#):

Table 1: Interaction matrices

	$B1$	$B2$	$B3$
$A1$	1	1	0
$A2$	1	1	1
$A3$	0	0	1
$A4$	1	1	0

The data in the table can be represented as a bipartite graph as shown in [Fig. 4](#). From [Fig. 4](#), it can be easily noticed that $B1$ has 3 paths to $B2$, $B1 \rightarrow A1 \rightarrow B2$ and $B1 \rightarrow A2 \rightarrow B2$ and $B1 \rightarrow A4 \rightarrow B2$. It can be obtained that $|p_{B1 \rightarrow B2}| = 3$.

$B1$ has three paths back to $B1$, $B1 \rightarrow A1 \rightarrow B1$ and $B1 \rightarrow A2 \rightarrow B1$ and $B1 \rightarrow A4 \rightarrow B1$. Then $|p_{B1 \rightarrow B1}| = 3$. $B2$ has 3 paths back to $B2$ as $B2 \rightarrow A1 \rightarrow B2$, $B2 \rightarrow A2 \rightarrow B2$ and $B2 \rightarrow A4 \rightarrow B2$. Then $|p_{B2 \rightarrow B2}| = 3$. Then we can get:

$$s(B1, B2) = \frac{2 \times 3}{3 + 3} = 1 \tag{6}$$

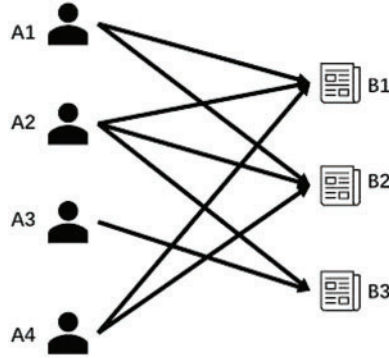


Figure 4: Bipartite graph for interaction matrices

We use matrices cash skill to speed up the computation. The user-paper matrix is R . We use the Commuting Matrix (CM) to simplify the calculation process. Commuting Matrix (CM) $M = R^T R$. Then, the meta-path similarity of the papers in the interaction matrix is:

$$s_{ij} = \frac{2M_{ij}}{M_{ii} + M_{jj}} \quad (7)$$

where M_{ij} is the element of the corner scale i, j in the commuting matrix. Then the procedure of adopting matrix for $s(B1, B2)$ is as follows:

$$M = R^T R = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 3 & 1 \\ 3 & 3 & 1 \\ 1 & 1 & 2 \end{bmatrix} \quad (8)$$

$$s_{12} = \frac{2M_{12}}{M_{11} + M_{22}} = \frac{2 \times 3}{3 + 3} = 1 \quad (9)$$

In the specific implementation, the Gamma distribution can fit the decay process well and can be introduced as a decay factor to learn the representation of user preferences. Take $\delta_{vj} \sim \text{Gamma}(\gamma_{vj}^{shp}, \gamma_{vj}^{rte})$ to denote the user's interest to thesis j decays according to the *Gamma* distribution. Then the likelihood of a tag v reaching a new paper i through paper j is:

$$z_{vji} = \delta_{vj} s_{ij} \quad (10)$$

Then, the likelihood of a tag v reaching a new paper i through a collection of papers J can be calculated as follows:

$$z_{vji} = \sum_{j=1}^{|J|} \delta_{vj} s_{ij} \quad (11)$$

4.3 Spatial Similarity Measures Based on Citation Networks

Even if the papers have significant textual gaps, they may still have intrinsic theoretical similarities that cannot be obtained by computing textual representation vectors. Given a citation network, it is necessary to utilize as much of the citation network as possible to obtain a low-dimensional vector representation of the papers. Utilizing only neighboring nodes may prevent the model from losing much information about the structure. Once too many layers of network node information are utilized,

too much noise may be introduced into the model. Therefore, we utilize a node’s neighboring edges and adjacent nodes as the contextual information of the current node. Intuitively, two nodes are more similar if two nodes have more neighboring nodes.

Let the meta-path similarity of the paper i and j in the interaction matrix is s'_{ij} . In the specific implementation, Gamma is similarly introduced as a decay factor in the representation learning of citation networks. Take $\tau_{vw} \sim \text{Gamma}(\gamma_{vw}^{shp}, \gamma_{vw}^{rte})$. Then, the likelihood of a tagged v pair reaching the citation network of a new paper i through paper w can be expressed by the following equation:

$$z_{vwi} = \tau_{vw} s'_{iw} \tag{12}$$

Then, the spatial similarity of the citation network for a user to reach a new paper i node on the citation network through the cluster of paper set J is described as:

$$z_{vwi} = \sum_{j=1}^{|J|} \tau_{vw} s'_{iw} \tag{13}$$

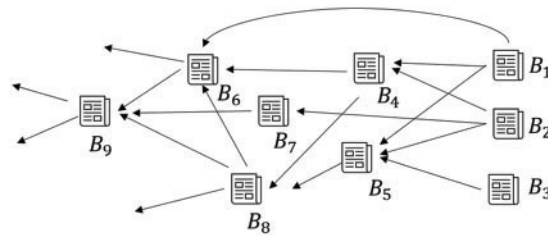


Figure 5: Citation networks

Table 2: The adjacency matrices G of citation networks

	B1	B2	B3	B4	B5	B6	B7	B8	B9
B1	0	0	0	0	0	0	0	0	0
B2	0	0	0	0	0	0	0	0	0
B3	0	0	0	0	0	0	0	0	0
B4	1	1	0	0	0	0	0	0	0
B5	1	1	1	0	0	0	0	0	0
B6	1	0	0	1	0	0	0	1	0
B7	0	1	0	0	0	0	0	0	0
B8	0	0	0	1	1	0	0	0	0
B9	0	0	0	0	0	1	1	1	0

Suppose there are nine papers $B1, B2, B3, B4, B5, B6, B7, B8, B9$. The citation network is shown in Fig. 5 and its adjacency matrices is shown in Table 2. As we can see, the adjacency matrix G of the directed graph is an asymmetric matrix. We take its Commuting Matrix (CM) as:

$$M' = (G + G^T) (G + G^T)^T \quad (14)$$

Let us take the meta-path ‘paper → paper → paper’ as an example. The goal is to find the paper’s similarity in citation. We simplify the citations network to an undirected graph so their adjacency matrices became symmetric. We use matrices cash skill to speed up the computation. The adjacency matrices are G , we take the Commuting Matrix (CM) $M' = (G + G^T) (G + G^T)^T$, then the meta-path similarity of the papers in the interaction matrix is:

$$s'_{ij} = \frac{2M'_{ij}}{M'_{ii} + M'_{jj}} \quad (15)$$

where M'_{ij} is the element of the corner scale i, j in the commuting matrix.

$$M' = (G + G^T) (G + G^T)^T = \begin{bmatrix} 3 & 2 & 1 & 1 & 0 & 1 & 0 & 3 & 1 \\ 2 & 3 & 1 & 0 & 0 & 1 & 0 & 2 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 4 & 3 & 2 & 1 & 1 & 2 \\ 0 & 0 & 0 & 3 & 4 & 2 & 1 & 0 & 1 \\ 1 & 1 & 0 & 2 & 2 & 4 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 2 & 1 & 0 \\ 3 & 2 & 1 & 1 & 0 & 2 & 1 & 4 & 1 \\ 1 & 1 & 0 & 2 & 1 & 1 & 0 & 1 & 3 \end{bmatrix} \quad (16)$$

$$s'_{56} = \frac{2M'_{56}}{M'_{55} + M'_{66}} = \frac{2 \times 3}{4 + 4} = 0.75 \quad (17)$$

$$s'_{57} = \frac{2M'_{57}}{M'_{55} + M'_{77}} = \frac{2 \times 1}{4 + 2} = 0.33 \quad (18)$$

So paper 5 is more similar to paper 6 than to paper 7.

4.4 Approximate Posterior Inference

We use variational inference to solve the model, and the details of the solution process are described below. δ_{vj} and τ_{vw} obey a gamma distribution, $\delta_{vj} \sim \text{Gamma}(\gamma_{vj}^{shp}, \gamma_{vj}^{rte})$, $\tau_{vw} \sim \text{Gamma}(\gamma_{vw}^{shp}, \gamma_{vw}^{rte})$, let $z_{vik}^M \sim \text{Poisson}(\eta_{vk}^T \theta_{ik})$, $z_{vji} \sim \text{Poisson}(\delta_{vj} s_{ij} y_{ji})$, $z_{vwi} \sim \text{Poisson}(\tau_{vw} s'_{vw} y_{wi})$, then Eq. (13) can be reduced to:

$$y_{vi} | y_{\neg vi} = \sum_{k=1}^K z_{vik}^M + \sum_{j=1}^{|J|} z_{vji} + \sum_{w=1}^{|W|} z_{vwi} \quad (19)$$

Due to the additivity of *Poisson* distribution, the sum of multiple *Poisson* distributions y_{vi} still obeys *Poisson* distribution. We can obtain the Variational family of the above model as follows:

$$q(z_{vik}^M, z_{vji}, z_{vwi}, \delta_{vj}, \tau_{vw}, \eta_v, \theta_i) = \prod_{v,i,k} z_{vik}^M \prod_{v,j,i} z_{vji} \prod_{v,w,i} z_{vwi} \prod_{v,j} \delta_{vj} \prod_{v,w} \tau_{vw} \prod_v \eta_v \prod_i \theta_i \quad (20)$$

According to the conditional conjugate model, each variational parameter equals the expectation of the other corresponding parameters in the complete condition. Detailed proofs are shown in Appendix (including Table A1). In summary, the parameters are updated as follows:

$$(\gamma_{vk}^{shp}, \gamma_{vk}^{rte}) \leftarrow \left(\lambda_{va} + \sum_i y_{vi} \phi_{vik}^M, \lambda_{vb} + \sum_i \frac{\gamma_{ik}^{shp}}{\gamma_{ik}^{rte}} \right) \quad (21)$$

$$(\gamma_{ik}^{shp}, \gamma_{ik}^{rte}) \leftarrow \left(\lambda_{ia} + \sum_v y_{vi} \phi_{vik}^M, \lambda_{ib} + \sum_v \frac{\gamma_{vk}^{shp}}{\gamma_{vk}^{rte}} \right) \quad (22)$$

$$(\gamma_{vj}^{shp}, \gamma_{vj}^{rte}) \leftarrow (\lambda_{ja} + \sum_i y_{vi} \phi_{vji}, \lambda_{jb} + \sum_i s_{ij} y_{ji}) \quad (23)$$

$$(\gamma_{vw}^{shp}, \gamma_{vw}^{rte}) \leftarrow (\lambda_{ha} + \sum_i y_{vi} \phi_{vwi}, \lambda_{hb} + \sum_i s'_{iw} y_{wi}) \quad (24)$$

$$\phi_{vi}^M \propto \langle \phi_{vi1}^M, \dots, \phi_{vik}^M, \dots, \phi_{viK}^M \rangle \quad (25)$$

We expect the latent variable as following:

$$\phi_{vik}^M = G_q [\eta_{vk}^T \theta_{ik}] = \exp \{ \mathbb{E}_q [\log \eta_{vk}^T + \log \theta_{ik}] \} = \exp \{ \Psi(\gamma_{vk}^{shp}) - \log \gamma_{vk}^{rte} + \Psi(\gamma_{ik}^{shp}) - \log \gamma_{ik}^{rte} \} \quad (26)$$

By the same reasoning, it is easy to obtain the following equation:

$$\phi_{vji} = \exp \{ \Psi(\gamma_{vj}^{shp}) - \log \gamma_{vj}^{rte} + \log s_{ij} y_{ji} \} \quad (27)$$

$$\phi_{vwi} = \exp \{ \Psi(\gamma_{vw}^{shp}) - \log \gamma_{vw}^{rte} + \log s'_{iw} y_{wi} \} \quad (28)$$

4.5 Optimization and Complexity Analysis

Algorithm 1 shows the entire optimization of our proposed method. The coordinate ascent method is more efficient than the batch algorithm. The computational complexity per iteration of the batch algorithm is $O((W + R)K)$, where R and W are the total number of non-zero observations in the document-user and document-word matrices, respectively. For the coordinate ascent method, this is $O((w_d + r_d)K)$, where r_d is the number of users who rated the sampled document d and w_d is the number of unique words in it. We assume that only one document is sampled in each iteration. In Fig. 3, sums involving polynomial parameters can be tracked for efficient memory usage. The constraint on memory usage is $O((D + V + U)K)$. As in Table 3, the update algorithm is of the same order of magnitude of complexity as the Bayesian Personalized Ranking (BPR), which does not introduce relational contexts. Therefore, on the premise of pre-processing contextual content and not increasing the arithmetic burden of the online model, the computational complexity of the model proposed in this paper remains the same as that of the BPR algorithm without fusing content information. Therefore, it is currently a more advanced algorithm regarding computational efficiency.

Table 3: Time complexity of methods with implicit feedback

Method	Time complexity
BPR	$O(R K)$
PMF	$O(MNK)$
ALS	$O((M + N)K^3 + R K^2)$
eALS	$O((M + N)K^2 + R K)$
Context-aware-eALS	$O((M + N)K^2 + R K)$
OUR METHOD	$O((w_d + r_d)K)$

Algorithm 1: Optimal coordinate update**Input:** C , paper-tag matrix G , the adjacency matrix of the citation network R , user-paper matrix $\lambda = \{\lambda_{va}, \lambda_{vb}, \lambda_{ia}, \lambda_{ib}, \lambda_{ja}, \lambda_{jb}, \lambda_{ha}, \lambda_{hb}\}$, starting value of model parameters**Output:** $\gamma = \{\gamma_{vk}^{shp}, \gamma_{vk}^{rte}, \gamma_{ik}^{shp}, \gamma_{ik}^{rte}, \gamma_{vj}^{shp}, \gamma_{vj}^{rte}, \gamma_{uv}^{shp}, \gamma_{uv}^{rte}\}$, model parameters

1. Initialize parameters
2. while not converge do
3. For each rating y_{ui} do
4. Update $\phi_{vik}^M, \phi_{vji}, \phi_{uvi}$ with Eqs. (26)–(28), respectively,
5. Normalize ϕ_{ui} with Eq. (25),
6. End For
7. For each article v do
8. Update $\gamma_{vk}^{shp}, \gamma_{vk}^{rte}$ with Eq. (21), respectively,
9. Update $\gamma_{vj}^{shp}, \gamma_{vj}^{rte}$ with Eq. (23), respectively,
10. Update $\gamma_{uv}^{shp}, \gamma_{uv}^{rte}$ with Eq. (24), respectively,
11. End For
12. For each tag i do
13. Update $\gamma_{ik}^{shp}, \gamma_{ik}^{rte}$ with Eq. (22), respectively
14. End For
15. While $\|\gamma^{(\theta+1)} - \gamma^{(\theta)}\| \leq 0.001$, break
16. Return

5 Experimental Results and Analysis

We examine the algorithm's efficacy using the citeulike-a and citeulike-t datasets [30]. The two datasets both contain the tags corresponding to the papers, records of user interactions with the articles, citation networks, and the bag of words for the articles. The statistical information of the dataset is shown in Table 4.

Table 4: Datasets

Statistics	citeulike-a	citeulike-t
Users	5551	7947
Articles	16980	25975
Tags	46391	52946
Citations	44709	32565
User-article pairs	204987	134860
Article-tag pairs	239253	290830

5.1 Evaluation Metric

We use implicit feedback for the computation, where we recommend K tags to papers based on the ranking of the predicted values and evaluate them based on the articles that users have checked

in. We use two metrics to evaluate the quality of the sorted list of tags: The Area Under Curve (AUC) and Mean Average Precision (MAP), which are defined as follows:

$$AUC = \frac{\sum_{i=1}^P rank_i - \frac{P(P+1)}{2}}{P \times N} \quad (29)$$

where $rank_i$ denotes the predicted ranking of the i_{th} positive sample, P denotes the number of positive samples, and N denotes the number of negative samples.

In the actual calculation process, the samples are usually ranked according to the predicted probability value. Then the ranked samples are taken as positive samples one by one, and the corresponding $rank_i$ values are calculated. Then these values are brought into the AUC calculation formula to obtain the final AUC value.

$$MAP = \frac{1}{I} \sum_{i \in I} \frac{1}{ni} \sum_{k=1}^N P(R_{ik}) \quad (30)$$

Average precision (AP) is the average of precision values at all ranks where relevant research papers are found and Mean Average Precision (MAP) given by Eq. (30), is the average of all APs. Where $P(R_{ik})$ denotes the precision of returned papers from the top until paper k is reached, N represents the length of the recommendation list, ni is the number of relevant papers in the recommendation list and I is the set of papers.

5.2 Baselines

In order to validate the effectiveness of introducing contextual information and optimization methods, we have chosen the following recommended techniques to compare with our approach:

PF [27]. Poisson Factorization (PF) is a method that estimates the parameters of a variable by variational inference, using the Poisson distribution as a prior for the variable.

CTR [30]. Collaborative Topic Regression (CTR). Collaborative Topic Regression method that combines Poisson decomposition and topic distribution can mitigate data sparsity and cold-start problems of articles.

mostPop [31]. mostPop is a tag-popularity-based ranking method.

userKNN [32]. User-based K-Nearest Neighbors (userKNN) is the user-based near-neighbor method which calculates the similarity between users based on their reading history using similarity metrics. Here, the paper is considered a user and the tag is an item.

itemKNN [32]. Item-based K-Nearest Neighbors (itemKNN) is the item-based nearest neighbor method, based on the reading records of papers, using similarity metrics to calculate the similarity between items. Here, the paper is considered a user and the tag is an item.

BPR [33]. Bayesian Personalized Ranking (BPR) is the classical ranking learning algorithm based on the principle that positive samples should be ranked before negative samples, and the recommendation model is trained by learning the ranking loss function.

CDR [34]. Collaborative Deep Ranking (CDR) is a hierarchical Bayesian deep learning framework combining a deep feature representation of the paper's content and implicit user preferences to reduce sparsity.

ALS [35]. Alternating Least Squares (ALS) is an implicit feedback-based matrix decomposition method that employs negative sample complete sampling.

We randomly draw 80% of the data from the thesis-tag matrix as training data and the remaining 20% as test data. For a fair comparison, we set the parameters of different algorithms concerning the corresponding literature or experimental results of the compared algorithms.

5.3 Result

5.3.1 Performance Comparison

The experimental results obtained by baselines on the citeulike-a and citeulike-t datasets are shown in Fig. 6. We can find following:

1) The algorithm proposed in this paper has greatly improved over the Poisson decomposition algorithm that does not introduce context, proving that it is feasible to introduce context through the additivity of the gamma distribution. This finding is consistent with the conclusion of the paper [1].

2) Our method shows more than 3% improvement over the alternating element least square method, we believe it is because our model uses a more rational approach to compute the trust strength, and our model successfully combines two contexts to mitigate the data sparsity problem. The eALS models do not model users' implicit feedback well due to the extremely sparse data in tag recommendations.

3) The recommendation performance of the model introducing contextual information is similar to that of the collaborative filtering-based model, suggesting that introducing contextual information requires an appropriate approach.

4) The performance of BPR based on negative sample sampling is worse than the ALS method based on complete negative sample sampling, which suggests that random negative sample sampling affects the model.

5) the recommendation accuracy of userKNN is higher than itemKNN in memory-based tagging recommendation algorithms. The reason is that many papers share fewer tagged, which leads to the fact that item-based paper similarity calculation is less accurate than user-based user similarity calculation. This observation is consistent with that of literature [36]. ItemKNN has the lowest accuracy and MAP, and collaborative filtering based solely on paper similarity computation cannot adequately characterize the reading preference of users.

5.3.2 The Effect of the Dimension of the Latent Vector

For the tag recommendation algorithm proposed in this paper, the dimension K_d of the latent feature vector is also an important parameter that affects the algorithm's performance. In incremental steps, we vary the value of K_d , from 10 to 150 and observe the variation of AUC and MAP on the citeulike-a dataset.

The experimental results are shown in Fig. 7. It can be observed that the values of AUC and MAP increase first with the increment of K_d . After reaching the optimal value, AUC and MAP decrease with the increment of K_d . On the citeulike-t dataset, the effect of parameter K_d on AUC and MAP shows a similar trend. This observation suggests that over-increasing the value of K_d may introduce some noise to reduce the accuracy of the recommendation algorithm.

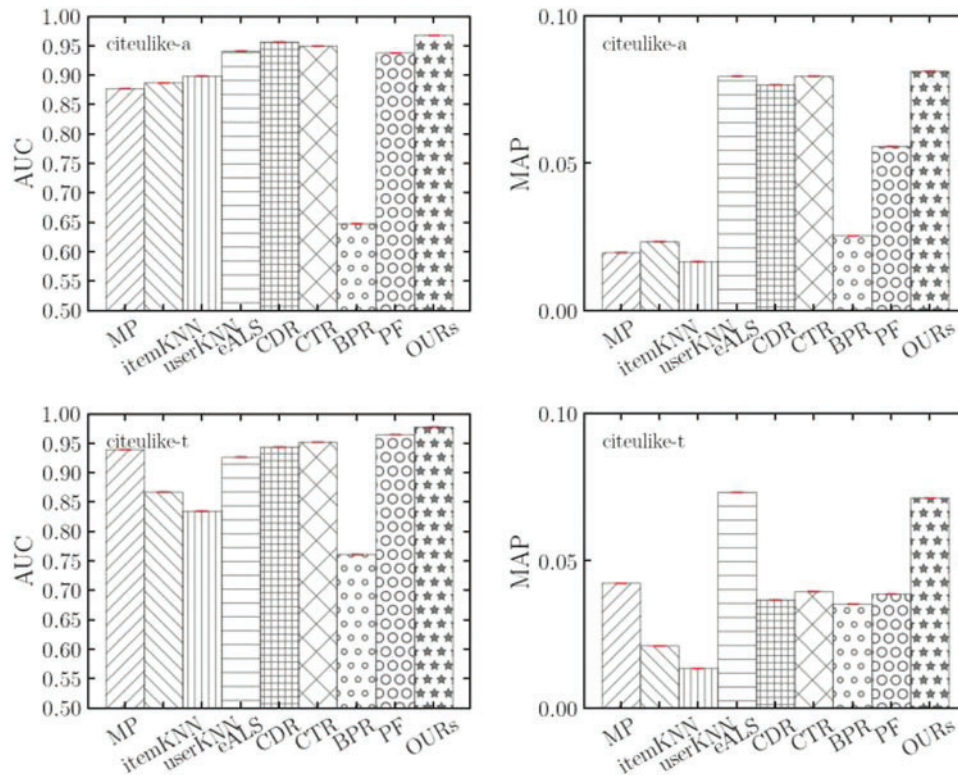


Figure 6: Performance comparison ($K_d = 50$)

5.3.3 Ablation Experiments

In this paper, we propose a tag recommendation algorithm that fuses context regularization through the additivity of the gamma distribution. We implement regularization of the article-tag interaction model with the citation and user-article networks. In this section, we exclude the citation and user-article network, respectively, and compare the changes of AUC and MAP with one missing term and without regularization.

- PF. This model removes items representing user-profiles and citation networks in Eq. (4).
- PF+ citation. This model removes items representing user profiles.
- PF+ profile. This model removes items representing citation networks.
- This is the full context-augmented Poisson factorization model with user profiles and citation network as context.

Fig. 8 shows the results evaluated on two datasets regarding AUC and MAP. There are several interesting observations:

- Jointly training mode is crucial. Compared with the Poisson decomposition without context, our model improves the AUC by 3.1% and 1.2% and the MAP by 46.1% and 84.0% on the two datasets, respectively. This demonstrates the effectiveness of gamma distribution additivity for feature extraction. This mechanism enables our model to capture various features in context for each tag by dynamically adjusting the feature weights for better tag recommendation.
- Effectiveness of context-sensitive modelling. By comparing PF with citation and PF with profile, it can be found that the model using citation as context performs relatively better (AUC improves by about 2.0% on citeulike-a dataset and 0.4% on citeulike-t dataset, MAP improves by about 19.1% on citeulike-a dataset and 50.4% on citeulike-t dataset). This suggests that citation network have a larger impact than user preference in labelling recommendations. Furthermore, the model is slightly better in citeulike-a than citeulike-t datasets. This is reasonable because of the different data sparsity.

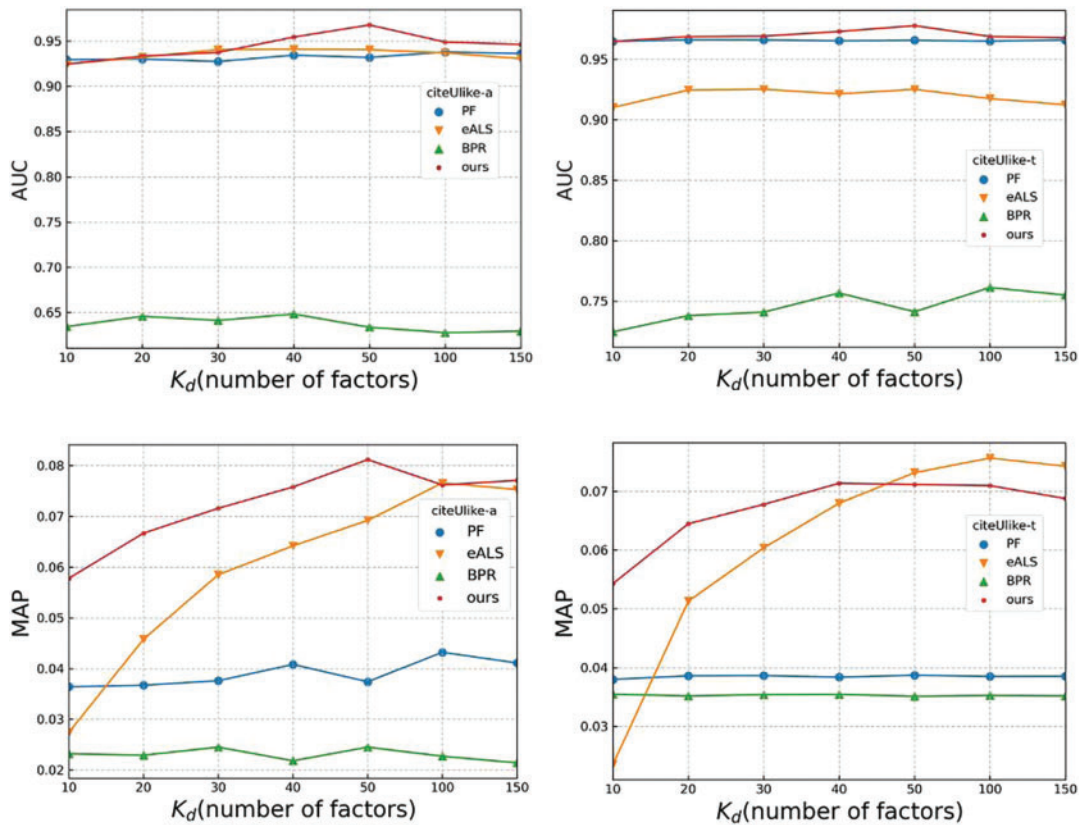


Figure 7: Prediction accuracy of four implicit feedback methods vs. K_d

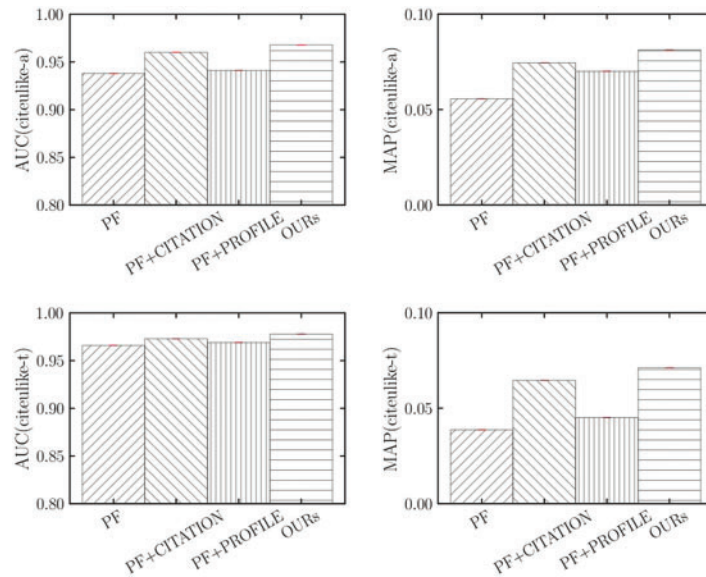


Figure 8: Ablation tests ($K_d = 50$)

6 Conclusion and Future Work

With the rapid growth of academic resources, the reading and organizing academic resources become more and more time-consuming. Tag recommendation is an integral part of the academic resource platform, which can help users personalize the classification of papers, facilitate users' finding papers, and reduce information overload. In order to fully utilize the three kinds of relational data in tag recommendation, this paper proposes a context-aware tag recommendation model based on Poisson decomposition. We explicitly integrate user preferences and citation networks into the model and solve it with gradient climbing method, which significantly reduces the complexity. Experimental results on two real tagging recommendation datasets show that the method in this paper outperforms current methods in terms of tag recommendation accuracy and computational efficiency.

However, academic tag recommendation systems may exacerbate the phenomenon of hot and cold areas or topics of academic research, leading to over-attention to certain areas of research and neglect of others. This is not conducive to the diversity and balanced development of academic research. We will look into this issue in our follow-up work.

Acknowledgement: Not applicable.

Funding Statement: This work was supported by the National Natural Science Foundation of China (No. 62271274).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Weiming Huang; data collection: Weiming Huang, Zhaoliang Wang; analysis and interpretation of results: Weiming Huang, Baisong Liu; draft manuscript preparation: Weiming Huang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data supporting this study's findings are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] P. Xu, H. Liu, B. Liu, L. Jing, and J. Yu, "Survey of tag recommendation methods," *J. Softw.*, vol. 33, no. 4, pp. 1244–1266, 2022. doi: [10.13328/j.cnki.jos.006481](https://doi.org/10.13328/j.cnki.jos.006481).
- [2] R. Ashvanth, G. Deepak, J. Sheeba Priyadarshini, and A. Santhanavijayan, "KMetaTagger: A knowledge centric metadata driven hybrid tag recommendation model encompassing machine intelligence," in *Intelligent Systems Design and Applications*, Springer, Cham, pp. 1–11, 2023.
- [3] K. Lei, Q. Fu, M. Yang, and Y. Liang, "Tag recommendation by text classification with attention-based capsule network," *Neurocomputing*, vol. 391, pp. 65–73, 2020. doi: [10.1016/j.neucom.2020.01.091](https://doi.org/10.1016/j.neucom.2020.01.091).
- [4] L. Yang, R. Yang, T. Chen, H. Fei, and J. Tang, "MF-TagRec: Multi-feature fused tag recommendation for GitHub," in *Proc. PRICAI*, Shanghai, China, 2022, pp. 17–31. doi: [10.1007/978-3-031-20868-3_2](https://doi.org/10.1007/978-3-031-20868-3_2).
- [5] J. Yi, X. Ren, and Z. Chen, "Multi-auxiliary augmented collaborative variational auto-encoder for tag recommendation," *ACM Trans. Inf. Syst.*, vol. 41, no. 4, pp. 1–25, 2023. doi: [10.1145/3578932](https://doi.org/10.1145/3578932).
- [6] D. Boughareb, A. Khobizi, R. Boughareb, N. Farah, and H. Seridi, "A graph-based tag recommendation for just abstracted scientific articles tagging," *Int. J. Coop. Inf. Syst.*, vol. 29, no. 3, pp. 2050004, 2020. doi: [10.1142/S0218843020500045](https://doi.org/10.1142/S0218843020500045).
- [7] R. Huang, C. Han, and L. Cui, "Tag-aware attentional graph neural networks for personalized tag recommendation," in *Proc. IJCNN*, Shenzhen, China, 2021, pp. 1–8. doi: [10.1109/IJCNN52387.2021.9533380](https://doi.org/10.1109/IJCNN52387.2021.9533380).
- [8] W. Zhao, L. Shang, Y. Yu, L. Zhang, C. Wang and J. Chen, "Personalized tag recommendation via denoising auto-encoder," in *Proc. WWW*, vol. 26, no. 1, pp. 95–114, 2023. doi: [10.1007/s11280-021-00967-3](https://doi.org/10.1007/s11280-021-00967-3).
- [9] Z. Fei, J. Wang, B. Huang, and X. Xiang, "Pairwise metric learning with angular margin for tag recommendation," *IEEE Access*, vol. 11, pp. 27020–27033, 2023. doi: [10.1109/ACCESS.2023.3246090](https://doi.org/10.1109/ACCESS.2023.3246090).
- [10] M. D. Hoffman, D. M. Blei, C. Wang, and J. W. Paisley, "Stochastic variational inference," *J. Mach. Learn. Res.*, vol. 14, pp. 1303–1347, 2012.
- [11] Y. Wu, Y. Yao, F. Xu, H. Tong, and J. Lu, "Tag2Word: Using tags to generate words for content based tag recommendation," in *Proc. CIKM*, Indianapolis, Indiana, USA, 2016, pp. 2287–2292. Accessed: Jan. 01, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:14840323>.
- [12] R. Tang, C. Yang, and Y. Wang, "A cross-domain multimodal supervised latent topic model for item tagging and cold-start recommendation," *IEEE Multimed.*, vol. 30, no. 3, pp. 48–62, 2023. doi: [10.1109/MMUL.2023.3242455](https://doi.org/10.1109/MMUL.2023.3242455).
- [13] L. Wang, Y. Li, and W. Jing, "KEIC: A tag recommendation framework with knowledge enhancement and interclass correlation," *Inf. Sci.*, vol. 645, pp. 119330, 2023. doi: [10.1016/j.ins.2023.119330](https://doi.org/10.1016/j.ins.2023.119330).
- [14] S. Yu, Q. Li, M. Liu, and Z. Wang, "Observation is reality? A graph diffusion-based approach for service tags recommendation," in *Proc. ICSSOC*, vol. 14420, pp. 100–114, 2023. doi: [10.1007/978-3-031-48424-7](https://doi.org/10.1007/978-3-031-48424-7).
- [15] L. Li, P. Wang, X. Zheng, Q. Xie, X. Tao and J. D. Velásquez, "Dual-interactive fusion for code-mixed deep representation learning in tag recommendation," *Inf. Fusion*, vol. 99, pp. 101862, 2023. doi: [10.1016/j.inffus.2023.101862](https://doi.org/10.1016/j.inffus.2023.101862).
- [16] R. Gharibi, A. Safdel, S. M. Fakhrahmad, and M. H. Sadreddini, "A content-based model for tag recommendation in software information sites," *Comput. J.*, vol. 64, no. 11, pp. 1680–1691, 2019. doi: [10.1093/comjnl/bxz144](https://doi.org/10.1093/comjnl/bxz144).
- [17] N. Engleitner, W. Kreiner, N. Schwarz, T. Kopetzky, and L. Ehrlinger, "Knowledge graph embeddings for news article tag recommendation," in *Proc. ICSS*, Amsterdam, Netherlands, 2021. Accessed: Jan. 01, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:238207994>.

- [18] X. Wang, P. Zhou, Y. Wang, X. Liu, J. Liu and H. Wu, “ServiceBERT: A pre-trained model for web service tagging and recommendation,” in *Proc. ICSOC*, Dubai, UAE, 2021, pp. 464–478. doi: [10.1007/978-3-030-91431-8_29](https://doi.org/10.1007/978-3-030-91431-8_29).
- [19] J. Gao *et al.*, “STAR: Spatio-temporal taxonomy-aware tag recommendation for citizen complaints,” in *Proc. CIKM*, Beijing, China, 2019. Accessed: Jan. 01, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:207757333>.
- [20] L. Xiao, X. Huang, B. Chen, and L. Jing, “Label-specific document representation for multi-label text classification,” in *Proc. EMNLP-IJCNLP*, Hong Kong, China, 2019. Accessed: Jan. 01, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:202771535>.
- [21] Y. Wei, Z. Cheng, X. Yu, Z. Zhao, L. Zhu and L. Nie, “Personalized hashtag recommendation for micro-videos,” in *Proc. MM*, Nice, France, 2019, pp. 1446–1454. doi: [10.1145/3343031.3350858](https://doi.org/10.1145/3343031.3350858).
- [22] Y. Yu, X. Chen, L. Zhang, R. Gao, and H. Gao, “Neural graph for personalized tag recommendation,” *IEEE Intell. Syst.*, vol. 37, no. 1, pp. 51–59, 2022. doi: [10.1109/MIS.2020.3040046](https://doi.org/10.1109/MIS.2020.3040046).
- [23] J. Sun, M. Zhu, Y. Jiang, Y. Liu, and L. Wu, “Hierarchical attention model for personalized tag recommendation,” *J. Assoc. Inf. Sci. Technol.*, vol. 72, no. 2, pp. 173–189, 2021. doi: [10.1002/asi.24400](https://doi.org/10.1002/asi.24400).
- [24] X. Chen, Y. Yu, F. Jiang, L. Zhang, R. Gao and H. Gao, “Graph neural networks boosted personalized tag recommendation algorithm,” in *Proc. IJCNN*, Glasgow, UK, 2020, pp. 1–8. doi: [10.1109/IJCNN48605.2020.9207610](https://doi.org/10.1109/IJCNN48605.2020.9207610).
- [25] J. Zhao, Q. Zhang, Q. Sun, H. Huo, Y. Xiao, and M. G., “FolkRank++: An optimization of FolkRank Tag recommendation algorithm integrating user and item information,” *KSII Trans. Internet Inf. Syst.*, vol. 15, no. 1, pp. 1–19, 2021. doi: [10.3837/tiis.2021.01.001](https://doi.org/10.3837/tiis.2021.01.001).
- [26] W. Zhao, A. Zhang, L. Shang, Y. Yu, and L. Zhang, “Hyperbolic personalized tag recommendation,” in *Proc. DASFAA*, vol. 13246, pp. 216–231, 2022. doi: [10.1007/978-3-031-00126-0](https://doi.org/10.1007/978-3-031-00126-0).
- [27] P. Gopalan, J. M. Hofman, and D. M. Blei, “Scalable recommendation with hierarchical poisson factorization,” in *Proc. UAI*, Amsterdam, Netherlands, 2015, pp. 326–335. Accessed: Jan. 01, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:1723648>.
- [28] C. Wang and D. M. Blei, “Collaborative topic modeling for recommending scientific articles,” in *Proc. SIGKDD*, San Diego, California, USA, 2011, pp. 448. doi: [10.1145/2020408.2020480](https://doi.org/10.1145/2020408.2020480).
- [29] Y. Guo, C. Xu, H. Song, and X. Wang, “Understanding users’ budgets for recommendation with hierarchical poisson factorization,” in *Proc. IJCNN*, Melbourne, Australia, 2017, pp. 1781–1787. doi: [10.24963/ijcai.2017/247](https://doi.org/10.24963/ijcai.2017/247).
- [30] H. Wang, B. Chen, and W. J. Li, “Collaborative topic regression with social regularization for tag recommendation,” in *Proc. IJCAI 2013*, Beijing, China, 2013, pp. 2719–2725. Accessed: Jan. 01, 2024. [Online]. Available: <http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/7006>.
- [31] Y. Zhang, J. Hu, S. Sano, T. Yamasaki, and K. Aizawa, “A tag recommendation system for popularity boosting,” in *Proc. MM*, California, USA, 2017, pp. 1227–1228. doi: [10.1145/3123266.3127913](https://doi.org/10.1145/3123266.3127913).
- [32] J. Wang, A. P. de Vries, and M. J. T. Reinders, “Unifying user-based and item-based collaborative filtering approaches by similarity fusion,” in *Proc. SIGIR*, Seattle, Washington, USA, 2006, pp. 501. doi: [10.1145/1148170.1148257](https://doi.org/10.1145/1148170.1148257).
- [33] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, “BPR: Bayesian personalized ranking from implicit feedback,” in *Proc. UAI’09*, Arlington, Virginia, USA, 2009, pp. 452–461.
- [34] H. Ying, L. Chen, Y. Xiong, and J. Wu, “Collaborative deep ranking: A hybrid pair-wise recommendation algorithm with implicit feedback,” in *Proc. PAKDD 2016*, Auckland, New Zealand, 2016, pp. 555–567. doi: [10.1007/978-3-319-31750-2_44](https://doi.org/10.1007/978-3-319-31750-2_44).
- [35] Y. Hu, Y. Koren, and C. Volinsky, “Collaborative filtering for implicit feedback datasets,” in *Proc. ICDM*, Pisa, Italy, 2008, pp. 263–272. doi: [10.1109/ICDM.2008.22](https://doi.org/10.1109/ICDM.2008.22).
- [36] T. Bogers and A. van den Bosch, “Recommending scientific articles using citeulike,” in *Proc. RecSys 2008*, Lausanne, Switzerland, 2008, pp. 287–290. doi: [10.1145/1454008.1454053](https://doi.org/10.1145/1454008.1454053).

Appendix

Details of Variational Inference

The Poisson decomposition employs variational inference as an optimization objective function to solve for the parameters. A hidden variable factor k is added to the model, and for the observation y_{vi} , We add k intermediate variables y_{vik} , such that $y_{vik} \sim \text{Poisson}(\eta_{vk}^T \theta_{vi})$ and $y_{vi} = \sum_k y_{vik}$, y_{ui} is converted to a polynomial by the transformation $\sum_k y_{vik}$, while y_{vi} still satisfies the Poisson distribution, due to the principle of additivity, the sum of multiple Poisson distributions obeys a Poisson distribution.

We use variational inference to solve the model, and the details of the solution process are described below. δ_{uj} and τ_{vw} obey a gamma distribution, $\eta_{vk} \sim \text{Gamma}(\gamma_{vk}^{shp}, \gamma_{vk}^{rte})$, $\theta_{ik} \sim \text{Gamma}(\gamma_{ik}^{shp}, \gamma_{ik}^{rte})$, $\delta_{vj} \sim \text{Gamma}(\gamma_{vj}^{shp}, \gamma_{vj}^{rte})$, $\tau_{vw} \sim \text{Gamma}(\gamma_{vw}^{shp}, \gamma_{vw}^{rte})$, let $z_{vik}^M \sim \text{Poisson}(\eta_{vk}^T \theta_{ik})$, $z_{vij} \sim \text{Poisson}(\delta_{vj} s_{ij} y_{ji})$, $z_{vwi} \sim \text{Poisson}(\tau_{vw} s'_{iw} y_{wi})$, then Eq. (19) can be reduced to:

$$y_{vi} | y_{-vi} = \sum_{k=1}^K z_{vik}^M + \sum_{j=1}^{|J|} z_{vij} + \sum_{w=1}^{|W|} z_{vwi} \quad (31)$$

Due to the additivity of Poisson distribution, the sum of multiple Poisson distributions y_{vi} still obeys Poisson distribution. We can obtain the Variational family of the above model as follows:

$$q(z_{vik}^M, z_{vij}, z_{vwi}, \delta_{vj}, \tau_{vw}, \eta_v, \theta_i) = \prod_{v,i,k} z_{vik}^M \prod_{v,j,i} z_{vij} \prod_{v,w,i} z_{vwi} \prod_{v,j} \delta_{vj} \prod_{v,w} \tau_{vw} \prod_v \eta_v \prod_i \theta_i \quad (32)$$

For Eq. (31), we extracting the portion associated with η_{vk} yields

$$p(\eta_{vk} | y_{vi}, \lambda_{va}, \lambda_{vb}) \propto \eta_{vk}^{\lambda_{va}-1} \exp(-\lambda_{vb} \eta_{vk}) \prod_i [(\eta_{vk}^T \theta_{ik})^{y_{vik}} \exp(-\eta_{vk}^T \theta_{ik})] \propto \eta_{vk}^{\lambda_{va} + \sum_i y_{vik}} \exp[-(\lambda_{vb} + \sum_i \theta_{ik}) \eta_{vk}] = \text{Gamma}(\lambda_{va} + \sum_i y_{vik}, \lambda_{vb} + \sum_i \theta_{ik}) \quad (33)$$

For y_{vik} , according to Lagrange's median theorem, we can get there must exists $y_{vi} = \sum_k y_{vik}$ and $\sum_k \phi_{vik}^M = 1$, make $y_{vik} = y_{vi} \phi_{vik}^M$. Since θ_{ik} obeys the Gamma distribution, its expectation is $\frac{\gamma_{ik}^{shp}}{\gamma_{ik}^{rte}}$. So the iterative formula for $(\gamma_{vk}^{shp}, \gamma_{vk}^{rte})$ can be rewritten as $(\gamma_{vk}^{shp}, \gamma_{vk}^{rte}) \leftarrow \left(\lambda_{va} + \sum_i y_{vi} \phi_{vik}^M, \lambda_{vb} + \sum_i \frac{\gamma_{ik}^{shp}}{\gamma_{ik}^{rte}} \right)$.

By the same reasoning, it is easy to obtain the following equation:

$$(\gamma_{ik}^{shp}, \gamma_{ik}^{rte}) \leftarrow \left(\lambda_{ia} + \sum_v y_{vi} \phi_{vik}^M, \lambda_{ib} + \sum_v \frac{\gamma_{vk}^{shp}}{\gamma_{vk}^{rte}} \right) \quad (34)$$

$$(\gamma_{vj}^{shp}, \gamma_{vj}^{rte}) \leftarrow (\lambda_{ja} + \sum_i y_{vi} \phi_{vji}, \lambda_{jb} + \sum_i s_{ij} y_{ji}) \quad (35)$$

$$(\gamma_{vw}^{shp}, \gamma_{vw}^{rte}) \leftarrow (\lambda_{ha} + \sum_i y_{vi} \phi_{vwi}, \lambda_{hb} + \sum_i s'_{iw} y_{wi}) \quad (36)$$

$$\phi_{vi}^M \propto \{\phi_{vi1}^M, \dots, \phi_{vik}^M, \dots, \phi_{viK}^M\} \quad (37)$$

The variational inference method assumes at this point that there exists a Poisson distribution q . The iterative method makes the distribution q move closer and closer to the target distribution p , which we need to solve for, and thus makes the following transformations:

$$\phi_{vik}^M = \frac{\eta_{vk}^T \theta_{ik}}{y_{vi}} \propto \exp(\log(\mathbb{E}_q[\eta_{vk}^T \theta_{ik}])) = \exp(\mathbb{E}_q[\log(\eta_{vk}^T \theta_{ik})]) = \exp\{\mathbb{E}_q[\log \eta_{vk}^T + \log \theta_{ik}]\} \quad (38)$$

where $\mathbb{E}_q(\cdot)$ represents the expectation under the distribution q . Since there exists $\log \eta_{vk}^T = \Psi(\gamma_{vk}^{shp}) - \log \gamma_{vk}^{rte}$, the final iterative formula for ϕ_{vik}^M is given as:

$$\phi_{vik}^M = G_q[\eta_{vk}^T \theta_{ik}] = \exp\{\mathbb{E}_q[\log \eta_{vk}^T + \log \theta_{ik}]\} = \exp\{\Psi(\gamma_{vk}^{shp}) - \log \gamma_{vk}^{rte} + \Psi(\gamma_{ik}^{shp}) - \log \gamma_{ik}^{rte}\} \quad (39)$$

By the same reasoning, it is easy to obtain the following equation:

$$\phi_{vji} = \exp\{\Psi(\gamma_{vj}^{shp}) - \log \gamma_{vj}^{rte} + \log s_{ij} y_{ji}\} \quad (40)$$

$$\phi_{vwi} = \exp\{\Psi(\gamma_{vw}^{shp}) - \log \gamma_{vw}^{rte} + \log s'_{iw} y_{wi}\} \quad (41)$$

Then we can derive and obtain the update equations are summarize in [Table A1](#):

Table A1: Hidden variables, complement conditionals, variational parameters and relevant expectations

Latent variable	Type	Complete conditional	Para.	Latent variable
η_v	Gamma	$\lambda_{va} + \sum_i z_{vik}^M, \lambda_{vb} + \sum_i \theta_{ik}$	$\gamma_{vk}^{shp}, \gamma_{vk}^{rte}$	$\lambda_{va} + \sum_i y_{vi} \phi_{vik}^M, \lambda_{vb} + \sum_i \frac{\gamma_{ik}^{shp}}{\gamma_{ik}^{rte}}$
θ_i	Gamma	$\lambda_{ia} + \sum_v z_{vik}^M, \lambda_{ib} + \sum_v \eta_{vik}$	$\gamma_{ik}^{shp}, \gamma_{ik}^{rte}$	$\lambda_{ia} + \sum_v y_{vi} \phi_{vik}^M, \lambda_{ib} + \sum_v \frac{\gamma_{vk}^{shp}}{\gamma_{vk}^{rte}}$
δ_{vj}	Gamma	$\lambda_{ja} + \sum_i z_{vji}, \lambda_{jb} + \sum_i s_{ij} y_{ji}$	$\gamma_{vj}^{shp}, \gamma_{vj}^{rte}$	$\lambda_{ja} + \sum_i y_{vi} \phi_{vji}, \lambda_{jb} + \sum_i s_{ij} y_{ji}$
τ_{vw}	Gamma	$\lambda_{ha} + \sum_i z_{vwi}, \lambda_{hb} + \sum_i s'_{iw} y_{wi}$	$\gamma_{vw}^{shp}, \gamma_{vw}^{rte}$	$\lambda_{ha} + \sum_i y_{vi} \phi_{vwi}, \lambda_{hb} + \sum_i s'_{iw} y_{wi}$
z_{vik}^M	Mult	$\log \eta_{vk} + \log \theta_{ik}$	$\widetilde{z_{vik}^M}$	$\exp\left\{\Psi(\gamma_{vk}^{shp}) - \log \gamma_{vk}^{rte} + \Psi(\gamma_{ik}^{shp}) - \log \gamma_{ik}^{rte}\right\}$
z_{vji}	Mult	$\log \delta_{vj} + \log s_{ij} y_{ji}$	$\widetilde{z_{vji}}$	$\exp\left\{\Psi(\gamma_{vj}^{shp}) - \log \gamma_{vj}^{rte} + \log s_{ij} y_{ji}\right\}$
z_{vwi}	Mult	$\log \tau_{vw} + \log s'_{iw} y_{wi}$	$\widetilde{z_{vwi}}$	$\exp\left\{\Psi(\gamma_{vw}^{shp}) - \log \gamma_{vw}^{rte} + \log s'_{iw} y_{wi}\right\}$