The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2514-9288.htm

# Topic optimization–incorporated collaborative recommendation for social tagging



Xuwei Pan, Xuemei Zeng and Ling Ding
*Department of Management Science and Engineering, Zhejiang Sci-Tech University, Hangzhou, China*

Received 20 November 2021
Revised 20 March 2022
1 May 2022
23 May 2022
Accepted 22 June 2022

## Abstract

**Purpose** – With the continuous increase of users, resources and tags, social tagging systems gradually present the characteristics of "big data" such as large number, fast growth, complexity and unreliable quality, which greatly increases the complexity of recommendation. The contradiction between the efficiency and effectiveness of recommendation service in social tagging is increasingly becoming prominent. The purpose of this study is to incorporate topic optimization into collaborative filtering to enhance both the effectiveness and the efficiency of personalized recommendations for social tagging.

**Design/methodology/approach** – Combining the idea of optimization before service, this paper presents an approach that incorporates topic optimization into collaborative recommendations for social tagging. In the proposed approach, the recommendation process is divided into two phases of offline topic optimization and online recommendation service to achieve high-quality and efficient personalized recommendation services. In the offline phase, the tags' topic model is constructed and then used to optimize the latent preference of users and the latent affiliation of resources on topics.

**Findings** – Experimental evaluation shows that the proposed approach improves both precision and recall of recommendations, as well as enhances the efficiency of online recommendations compared with the three baseline approaches. The proposed topic optimization–incorporated collaborative recommendation approach can achieve the improvement of both effectiveness and efficiency for the recommendation in social tagging.

**Originality/value** – With the support of the proposed approach, personalized recommendation in social tagging with high quality and efficiency can be achieved.

**Keywords** Social tagging, Topic model, Collaborative filtering, Recommender systems, Personality, Logistic function

**Paper type** Technical paper

## 1. Introduction

As a popular component of Web 2.0 technologies, social tagging systems have grown rapidly since its emergence. In social tagging systems, users can completely freely add one or more descriptions according to their liking for a series of resources such as books, photos, music and videos (Golder and Huberman, 2005). Consequently, social tagging systems have become an effective tool that integrates functions of organizing, sharing, retrieving and discovering information resources (Zhou *et al.*, 2010) and can filter the "noise" that is constantly generated on the Internet (Chi and Mytkowicz, 2007).

The user tagging behavior in social tagging systems changes the binary relationship between users and resources in traditional recommender systems and constructs a "user–resource–tag" ternary relationship, which provides a new solution for the recommendation of information resources (Kubatz *et al.*, 2011). Tags can not only describe the characteristics of resources but also express user preferences and interests in resources (Xie *et al.*, 2016). Social tagging fully exerts the wisdom of group users and provides an important data


*Funding:* This is study is supported by Zhejiang Provincial Natural Science Foundation of China (LZ18G010001).

Data Technologies and Applications
Emerald Publishing Limited
2514-9288
DOI 10.1108/DTA-11-2021-0332

source for achieving accurate recommendation (Marinho *et al.*, 2011). Integrating tags into recommendation services has become an important direction in the field of personalized recommendation. Previous social tagging recommendation methods make full use of relationships among users, resources and tags to improve the recommendation effect from different perspectives (Ifada and Nayak, 2014; Shokeen and Rana, 2020; Wang, 2017; Zhen *et al.*, 2009). However, most existed methods mainly focus on recommendation effectiveness, especially on improving recommendation accuracy, while recommendation efficiency has not been paid enough attention.

With the continuous increase of users, resources and tags, social tagging systems gradually present the characteristics of "big data" such as large number, fast growth, complexity and unreliable quality. In a "big data" environment, social tagging recommendation has encountered new challenges: (1) In social tagging systems, users, resources and tags have grown rapidly, so that a large number of tagging behaviors have caused a large number of junk tags to appear, and the consistency and reliability of tags have been reduced (Shepitsen *et al.*, 2008). It is becoming hardly possible to accurately obtain user interest preferences directly from tags (Indra and Thangaraj, 2019). (2) Recommendation service is a user-oriented service that starts with user needs and ends with satisfying user needs, and inefficient recommendation services will affect user satisfaction (Parkhomenko *et al.*, 2019). However, most existing social tagging recommendations confuse the offline optimization process and online service process, which results in too long online time for users and affects the efficiency of online recommendation. In other words, the continuous increase in users, resources and tags in social tagging systems leads to increases in the complexity of recommendations, which not only makes recommendations become inefficient but also cannot guarantee that the recommendation results can better meet user needs.

In order to solve the problems of tags redundancy, inconsistency and uncertainty, tags should be optimized first. The topic model of tags is a choice to solve this problem (Li *et al.*, 2011; Ramage *et al.*, 2009; Zhong *et al.*, 2017). In the topic modeling of tags, not only the relationships among users, resources and tags are considered, but also tags are clustered into some consistent topics according to the tag's characteristics. Therefore, the redundancy, uncertainty and inconsistency of tags are reduced, and the user's preferences and features of resources can be presented more effectively. Meanwhile, based on the topic model of tags, the corresponding users' latent preference model and resources' latent affiliation model on topics are constructed from the history tagging data, which can be separated from the online recommendation service process for a target user. Consequently, combing the idea of optimization before service, we present a different collaborative recommendation approach for social tagging that soundly incorporates the offline topic optimization into the online recommendation service. In our proposed approach, the recommendation process is explicitly divided into the offline topic optimization phase and the online recommendation service phase. In the offline phase, we first construct a topic model of tags with the "user–resource–tag" ternary relationships; then the topic model is used to optimize the latent preference model of users and the latent affiliation model of resources on topics. In the online recommendation service process, the latent preference model of users and the latent affiliation model of resources created in the offline phase are incorporated into obtaining the target user's interesting topics and generating the corresponding recommendation list, respectively.

The proposed topic optimization–incorporated collaborative recommendation for social tagging brings two advantages. One is that the constructed topic model alleviates encountered problems on redundancy, uncertainty and inconsistency of tags in a "big data" social tagging environment and conduces to obtain user's preferences more accurately. Another advantage is that incorporating offline topic optimization into online recommendation service not only reduces the pressure (e.g. time-spending and computing complexity) of the user-visible online service process by strengthening the user-invisible offline optimization process but also ultimately guarantees the quality and efficiency of the recommendation service. The main contributions of this paper are concluded as follows:

(1) A new idea is put forward to solve the contradiction between quality and efficiency of recommendation in social tagging. Combining the idea of optimization before service, the process of recommendation implementation in social tagging is divided into two explicit phases: offline topic optimization and online recommendation service. By integrating the two processes, personalized recommendation service with high quality and efficiency can be achieved.

(2) Logistic function is used to convert the tagging frequency. According to the characteristics of the user's tagging behavior, we use Logistic function to depict the frequency relationship between users and tags, resources and tags more accurately.

(3) The approach we proposed calculates the user's preference from the perspective of the topic and combines the topic and user's interest by establishing relationship matrices. The constructed user–topic preference matrix and resource–topic affiliation matrix reflect the latent preference of users and the latent affiliation of resources on topics.

(4) Experimental exploration is carried out on MovieLens 20M and CiteULike dataset. Experimental results show that our approach can achieve improvement in both effectiveness and efficiency for the recommendation in social tagging.

This paper is organized as follows: Section 2 discusses the related work. Section 3 describes the proposed approach in detail. Section 4 shows the experimental evaluation and results. Section 5 concludes the whole paper.

## 2. Related works

At present, the tag-based recommendation method has been widely used in the recommendation field. According to different research perspectives, social tagging recommendation methods can be divided into three categories: graph-based methods, tensor-based methods and topic-based methods.

In social tagging systems, users, resources and tags constitute a complex relationship network, which can be studied using graph theory–related theories such as bipartite graphs, tripartite graphs and hypergraphs (Landia *et al.*, 2013). Guan *et al.* (2010) proposed to represent users, tags and documents in the same semantic space. The distance between two documents is measured by their relevance. And the documents that are close enough, that is, the more relevant, will be recommended to users. Zhang *et al.* (2011) proposed an algorithm based on hybrid mass diffusion, which uses both user–resource graph and resource–tag graph for personalized recommendation. To solve the problem of sparsity in social tags, Zhang *et al.* (2013) built a ternary interaction graph and then applied random walk to explore the transfer relationship between users and resources. Liu *et al.* (2017) propose a hybrid method that combines the collaborative filtering (CF) method and graph-based interest propagation for movie recommendation.

Graph-based methods often use two-dimensional vectors to represent the relationship between two entities and cannot dig into the user's behavior and explore the internal relationship of multiple tags described by the user on the same resource. Tensors-based approach proposed by Symeonidis (2009) describes the relationship between the three entities (e.g. user–resource, resource–tag and user–tag). They developed a unified framework to model the three entities that exist on the social tagging system, namely users, products and tags (Symeonidis *et al.*, 2010). In their proposed model, these data are modeled by a third-order tensor, on which the high-order singular value decomposition (SVD) method and the kernel-SVD smoothing technology are used to perform multichannel latent semantic analysis and dimensionality reduction. Rafailidis and Daras (2013) proposed a tensor factorization and tag clustering model for resource recommendation in social tagging systems. The method they proposed has contributed to solving cold start and sparsity issues. A common problem with tensor modeling when generating quality recommendations for large datasets is scalability. Ifada and Nayak (2014) proposed a tensor-based recommendation method using a probabilistic ranking method. This method uses block stripe parallel matrix multiplication to generate a reconstruction tensor and then probabilistically calculates the user's preference for ranking recommended resources.

Tensor-based recommendation methods are also suitable for generating user or tag recommendation lists. It greatly reduces the difficulty of recommending multidimensional data and supports multimodal recommendations in a simple way (Hong *et al.*, 2019). However, this way of analyzing only the relationship between objects often ignores the meaning of the objects themselves, such as the semantics of tags and the characteristics of resources. Mining the semantic features of tags can more accurately grasp user interests and better describe resource characteristics (Li *et al.*, 2011), so topic modeling methods are used to improve recommendation performance. In order to alleviate the inherent sparsity of the data and the vocabulary problems introduced by having a completely unrestricted lexicon, Harvey *et al.* (2010) proposed a method based on Latent Dirichlet Allocation (LDA) topic modeling. This method reduces the dimensionality of the data to provide more accurate resource rankings with higher recall. Yao *et al.* (2018) proposed an algorithm to model the generation of tags based on both users and resources, thereby solving the coupling relationship between social tags. Further, Liu (2019) combines tag frequency, time and ordinal position to compute the user's interest degree. Considering the semantic relationship, Wang and Blei (2011) proposed an approach combining the merits of traditional CF and probabilistic topic modeling, which provides an interpretable latent structure for users and resources. However, these approaches cannot apply to the recommendation of unstructured resources because it topicalizes the content of the recommended resources. Chen *et al.* (2016) paid more attention to the semantic information of tags and links between tags and users and resources and proposed a tag and rating–based CF model for resource recommendation. Topic modeling is used to separately mine the semantic information of tags of each user and each resource, and then the semantic information is merged into matrix decomposition to factorize rating information and capture the bridging characteristics of tags and hierarchies between users and resources. Similarly, Liu *et al.* (2020) propose a CF algorithm using a topic model called user-item-tag LDA. Similar methods are applied to the cross-domain recommendation. Wang and Lv (2020) propose a Tag-informed Cross-Domain Collaborative Topic Regression model, which exploits shared tags as bridges to link related domains through an extended collaborative topic modeling framework. To sum up, compared with the method based on graph theory and the method based on tensor, the topic-based method not only considers the relationship between users, resources and tags but also integrates the characteristics of tags and resources for a comprehensive recommendation, which can better meet the personalized needs of users for resources (Belém *et al.*, 2017). In addition, the topic-based method can divide tags into multiple clusters with consistent topics, which reduces the redundancy, uncertainty and inconsistency of tags (Duan *et al.*, 2015; Ifada, 2014; Xu *et al.*, 2020).

However, the above studies mainly focus on the recommendation quality, especially on improving recommendation accuracy, while the conflict between recommendation efficiency and recommendation effectiveness has not been paid enough attention. Additionally, they commonly use the direct number relations in the topic modeling, while the fact that the closeness of the relationship between tags and users as well as tags and resources is not straightforward linear relation has not been considered. Therefore, in this paper, we put forward a framework that divides the recommendation process into the offline topic optimization phase and the online recommendation service phase, and these two phases are incorporated naturally. Besides, we exploit Logistic function to better express closeness between tags and users as well as tags and resources in the topic modeling.

## 3. Proposed approach

The goal of this study is to incorporate topic optimization into CF to enhance both the effectiveness and the efficiency of personalized recommendation for social tagging. Figure 1 illustrates the overall framework of our proposed approach, which is divided into two stages: the offline topic optimization phase and the online recommendation service phase. To alleviate problems of redundancy, uncertainty and inconsistency of the tags in "big data" social tagging environment, the topic model is exploited to optimize tag data and explore the potential relationships among users, tags and resources. To reduce the time complexity of online computation, the offline optimization phase models, stores and preprocesses the offline data. The online recommendation service phase makes the personalized resources recommendation based on the usage data of previous retrievals in the offline phase. Consequently, our proposed approach can be beneficial to not only improve the quality of recommendation through topic optimization of "big data" tags but also enhance the efficiency of the user-visible online recommendation phase by incorporating offline topic optimization into online recommendation service.

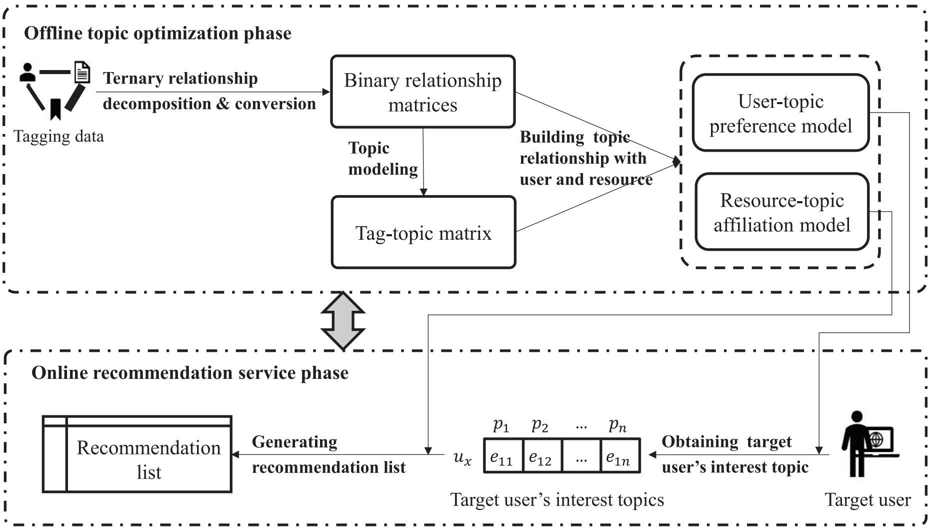


**Figure 1.** The overall framework of our proposed approach

*3.1 Offline topic optimization*

The offline topic optimization phase exploits the topic model to optimize tags and explore the potential relationships among users, tags and resources by preprocessing the history tagging data. This phase consists of three interactive tasks: ternary relationship decomposition and conversion, topic modeling and building topic relationships with users and resources. Below are the logically interactive steps of the three tasks:

(1) Ternary relationship decomposition and conversion. The "user–resource–tag" ternary relationship is decomposed into three two-dimensional matrices: user–tag matrix, resource–tag matrix and user–resource matrix. In order to better characterize the closeness of relationships between users and tags as well as resources and tags, Logistic function is used to converse the value of the unit in the user–tag matrix and the resource–tag matrix.

(2) Topic modeling. Based on the conversed user–tag matrix and resource–tag matrix, LDA model is exploited to transform tags into some clusters, which represent tag–topics.

(3) Building topic relationships between users and resources. The above three two-dimensional matrices and constructed topic models are combined to create the user–topic preference model and the resource–topic affiliation model, which present the latent preference of users on topics and the latent affiliation of topics for resources, respectively. The collaborative recommendation idea is utilized by incorporating topic–topic similarity and resource–resource similarity into the process of constructing models.

*3.1.1 Ternary relationship decomposition and conversion.* In social tagging systems, users, resources and tags constitute a ternary relationship. We decompose the ternary relationship into three binary relationships.

The ternary relationships among "user–resource–tag" can be represented by a three-dimensional matrix $M = [m_{u,r,t}]_{|U|\times|R|\times|T|}$, where $U$ is the user set, $U = \{u_1, u_2, \ldots, u_n\}$, $R$ is the resource set, $R = \{r_1, r_2, \ldots, r_m\}$, $T$ is the tag set, $T = \{t_1, t_2, \ldots, t_l\}$, $|U|$, $|R|$ and $|T|$ represent the number of users, resources and tags, respectively, and $m_{u,r,t}$ represents a tagging action of a user on a resource. The three-dimensional matrix is decomposed into three primary two-dimensional matrices: $UT_0 = [a_{u,t}]_{|U|\times|T|}$, $RT_0 = [b_{r,t}]_{|R|\times|T|}$ and $UR_0 = [c_{u,r}]_{|U|\times|R|}$. The elements in matrix $UT_0$ and $RT_0$ are the number of times that a tag is tagged by a user and a resource, respectively. The elements in matrix $UR_0$ are 0 or 1, where 1 means that a user has tagged a resource and 0 means that a user has not tagged a resource. The reduction process of the ternary relationship is illustrated in Figure 2.

In the primary decomposed $UT_0$ matrix and $RT_0$ matrix, the number of times of tagging represents the closeness of the relationship. However, the closeness of the relationship between tags and users and between tags and resources generally does not have a linear growth with frequency. For example, the closeness of the relationship between users and tags $Rev_{u,t}$ usually does not grow at the same rate. At the initial stage, the user $u$ only performs a small amount of tagging using tag $t$, $Rev_{u,t}$ will get a rapid growth. With more use of tag $t$ by user $u$, the growth rate of $Rev_{u,t}$ will be slower and ultimately tends to 0. Therefore, it is not applicable to directly use frequency to describe the closeness of the relationship between users and tags from the perspective of actual user tagging behavior (Pan *et al.*, 2017). The closeness of the relationship between resources and tags $Rev_{r,t}$ is similar to $Rev_{u,t}$.

Logistic function is derived from the population growth model and is used to describe the population growth trend (Richards, 1959). The population at the initial stage has approximately

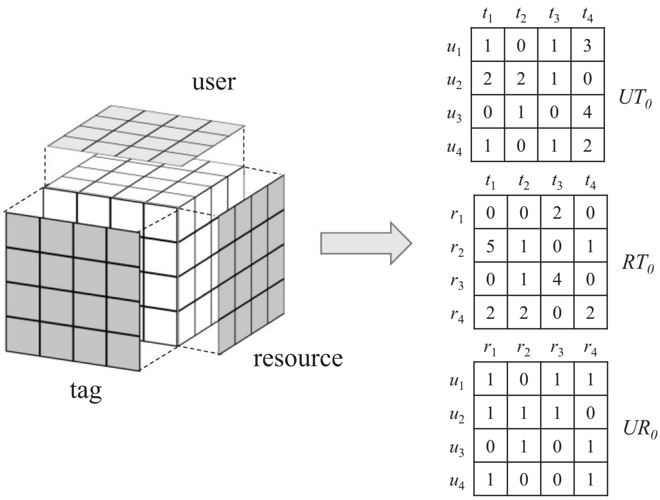


**Figure 2.**
An example of social tagging data decomposing

exponential growth. Then, as the population gradually saturates, the growth rate slows down to linear and finally the growth rate tends to 0. So Logistic function is suitable to describe the closeness of the relationship between tags and users as well as tags and resources, in that the growth is from fast to slow and eventually tends to be stable. Consequently, we define the Logistic function to describe the correlation of tags with users and resources, as shown in Formula 1.

$$Rev_{i,j} = \begin{cases} 0 & , \quad n_{i,j} = 0 \\ \dfrac{1}{1 + e^{-k\left(n_{i,j} - n_0\right)}} & , \quad n_{i,j} > 0 \,\text{and}\, n_{i,j} \in N \end{cases}, \tag{1}$$

where $Rev_{i,j}$ represents the correlation between two objects, here objects refer to users, resources or tags; $n_{i,j}$ represents the frequency between two objects; $n_0$ represents the intermediate value of all frequencies, which normally is the average or median of $n_{i,j}$; $k$ is the growth rate of the curve, usually set $k = 1$ according to the characteristic of user tagging behavior (Pan and Ding, 2018). From the definition, we can see $Rev_{i,j} \in [0, 1]$. Especially, when $n_{i,j} = 0$, $Rev_{i,j} = 0$, indicating that there is no association between the two objects; when $n_{i,j} = n_0$, $Rev_{i,j} = 0.5$. Figure 3 shows the graph of $Rev_{i,j}$ when $n_0 = 3$ and $k = 1$.

Based on the primary users–tags matrix $UT_0$ and resources–tags matrix $RT_0$ obtained previously, we use the Logistic function to convert the values in the units of two matrices from frequency to their corresponding results of Logistic function $Rev_{i,j}$. The converted matrices with values of Logistic function $Rev_{i,j}$ are named $UT$ and $RT$, respectively.

*3.1.2 Tag–topic modeling.* Compared with a single tag, a tag cluster is often composed of multiple tags, showing distinct topic information, which conduces to alleviate problems of redundancy, uncertainty and inconsistency of tags. With the help of clustering ideas and methods, the topic model can transform tags into tag clusters with distinct topics.

LDA is a topic model that can give the topic of each document in the document set in the form of a probability distribution (Blei *et al.*, 2003). The input of LDA is a corpus including documents and their corresponding words, and the output is a potential topic distribution. In social tagging systems, if a user or a resource with its related tags is regarded as

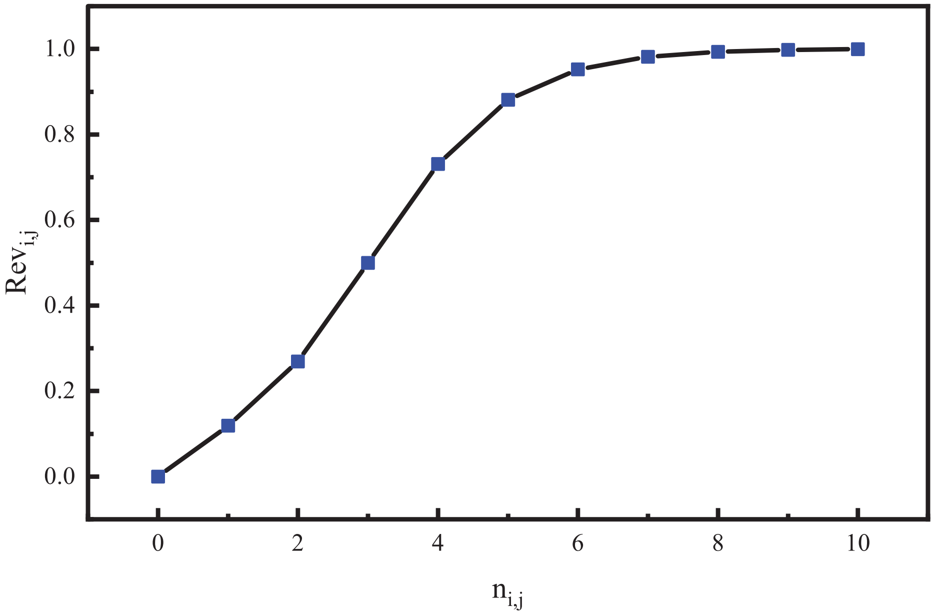


**Figure 3.**
Graph of $Rev_{i,j}$ when $n_0 = 3$ and $k = 1$

a document and a tag is regarded as a word, the corpus constituted by users (or resources) and tags can be trained by the LDA model to obtain the distribution of the tag–topic (Newman *et al.*, 2011). Therefore, using LDA model for tag–topic modeling not only fully considers the relationships among multiple tags to interpret related tags' semantics features of resources for user-personalized demand mining but also greatly improves the efficiency of tag data processing (Das *et al.*, 2015).

This paper applies LDA to model tag–topics. The input of LDA is the corpus constituted by users and tags, i.e. the *UT* matrix we have processed with the Logistic function earlier. The output is the "tag–topic" distribution matrix $TP = [d_{t,p}]_{|T|\times|P|}$, reflecting the probability of each tag appearing under each topic, where $|P|$ represents the number of topics and $d_{t,p}$ is the probability of the tag $t$ appearing under the topic $p$.

When applying LDA to model tag–topics, the number of topics, denoted as $K_{\text{topic}}$, is a user-specified parameter, which needs to be manually set. The perplexity is a valid evaluation index to determine the value of $K_{\text{topic}}$ (Jacobi *et al.*, 2016). $|P|$ is equal to the final optimal value of $K_{\text{topic}}$ obtained.

*3.1.3 Building topic relationships with user and resource.* Based on the constructed tag–topic matrix, we can build relationships between users and topics and between resources and topics. To further discover the latent relationships, the collaborative similarities are computed and then used in the relationships building process. Therefore, there are three subtasks in this process: building the direct relationships between users and topics as well as resources and topics, computing the similarity of resources and topics and finding users' latent preference on topics and the latent affiliation of the resource on topics.

3.1.3.1 Building topic direct relationships with users and resources. Tag–topic model divides the cluttered and inconsistent tags into some clusters with distinct topics, which helps to better describe the characteristics of users and resources. Tags are not only in the tag–topic distribution matrix *TP* but also in the user–tag relationship matrix *UT* and the resource–tag relationship matrix *RT*. We can use the bridge of tags to directly construct "user–topic" relationship and "resource–topic" relationship. Thereby, on the one hand, characteristics of user and resource are more accurately described by topics, and on the other hand, the data size is reduced, which is beneficial to improve next computing performance. We use the principle of matrix multiplication to achieve this conversion, converting the *UT* matrix into the user–topic matrix *UP*, $UP = UT \times TP$. Similarly, the *RT* matrix can be converted into the resource–topic matrix *RP*, $RP = RT \times TP$. An example of the *UP* construction process is illustrated in Figure 4.

3.1.3.2 Similarities computation. According to the idea of CF, computing the similarity between two objects can help to find a latent relationship. In our proposed approach, resource–resource similarity and topic–topic similarity are used to obtain the user's latent preference for similar topics and the association between topics and similar resources.

*Resource–resource similarity.* The calculation of resource–resource similarity can use either the user–resource matrix *UR* or the resource–topic matrix *RP*. We define the resources similarity calculated based on *UR* as *RS_user* and the similarity calculated based on *RP* as *RS_topic*. Cosine similarity is used to calculate them. In order to fully consider the impact of users and topics on resource similarity, resource–resource similarity *sim_res* is combined by *RS_user* and *RS_topic* in a linear way, which is shown as Formula 2.

$$sim\_res(r_x, r_y) = \lambda RS\_user(r_x, r_y) + (1 - \lambda) RS\_topic(r_x, r_y), \tag{2}$$

where $r_x$ and $r_y$ are two different resources, $\lambda$ is the adjusting parameter, $\lambda \in (0, 1)$. Based on this calculation, we can obtain the resource–resource similarity matrix *RS*.

*Topic–topic similarity.* When calculating the topic–topic similarity *sim_topic*, we follow the same steps as above for calculating the resource–resource similarity *sim_res*. We define the similarity calculated based on *UP* as *PS_user* and the similarity calculated based on *RP* as *PS_res*. The topic–topic similarity *sim_topic* is defined as shown in Formula 3.

$$sim\_topic(p_x, p_y) = \gamma PS\_user(p_x, p_y) + (1 - \gamma) PS\_topic(p_x, p_y), \tag{3}$$

where $p_x$ and $p_y$ represent two different topics, $\gamma \in (0, 1)$. Based on this calculation, we can obtain the topic–topic similarity matrix *PS*.

3.1.3.3 Constructing user–topic preference model and resource–topic affiliation model. To discover the user's preference on topics and the closeness between resources and topics, direct topics' relationship with users and resources are further transformed by combining resource–resource similarity or topic–topic similarity. The user–topic preference model is denoted as $\mathrm{Pref}_{\text{user-topic}}$, which is deduced by the user–topic matrix *UP* and topic–topic similarity matrix *PS* as shown in Formula 4.

$$\begin{array}{cc} & \text{tag} \\ \text{user} & \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1l} \\ a_{21} & a_{22} & \cdots & a_{2l} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nl} \end{bmatrix} \end{array} \times \begin{array}{cc} & \text{topic} \\ \text{tag} & \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1k} \\ d_{21} & d_{22} & \cdots & d_{2k} \\ \vdots & \vdots & & \vdots \\ d_{l1} & d_{l2} & \cdots & d_{lk} \end{bmatrix} \end{array} = \begin{array}{cc} & \text{topic} \\ \text{user} & \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1k} \\ e_{21} & e_{22} & \cdots & e_{2k} \\ \vdots & \vdots & & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nk} \end{bmatrix} \end{array}$$

**Figure 4.** An illustration of the process of computing the user–topic matrix

$$\mathrm{Pref}_{\mathrm{user-topic}} = UP \times PS. \tag{4}$$

The topic–topic similarity matrix combining user preferences on topics is helpful to discover potential topics with which users are not directly associated. So $\mathrm{Pref}_{\text{user-topic}}$ can obtain a potential interest in the topic from the user. Similarly, to find the potential association between topics and resources, we construct the resource–topic affiliation model by combining the resource–topic matrix *RP* and the resource–resource similarity matrix *RS*. The difference from the construction of $\mathrm{Pref}_{\text{user-topic}}$ is that the resource–topic matrix *RP* should be transposed. So, the resource–topic affiliation model is computed as shown in Formula 5.

$$\mathrm{Aff}_{\mathrm{topic-res}} = RP^{T} \times RS. \tag{5}$$

*3.2 Online recommendation service*

Based on the user–topic preference model and resource–topic affiliation model obtained in the offline optimization stage, the pressure of online recommendation service is reduced to a great extent. The online recommendation service phase mainly focuses on two subtasks: obtaining the target user's interest topics and generating a final recommendation list.

In the offline phase, the user–topic preference library and the resource–topic affiliation library are generated and stored. The target user's interest topics $\mathrm{Pref}'_{\mathrm{user}_x-\mathrm{topic}}$ can be selected from the user–topic preference library, which is a vector of user's interest in each topic. Then, the user's preference for each resource can be obtained according to Formula 6.

$$\mathrm{Score}_{u_x}(r) = \mathrm{Pref}'_{\mathrm{user}_x-\mathrm{topic}} \times \mathrm{Aff}_{\mathrm{topic-res}}. \tag{6}$$

*3.3 Time complexity analysis*

The proposed approach divided the social tagging recommendation into two stages: the offline topic optimization phase and the online recommendation service phase, so we need to, respectively, analyze time complexity of them.

Offline computing time is mainly spent on building the user–topic preference model and resource–topic affiliation model. Considering the matrix sparsity, the time complexity of building those two models is approximately $O(|P||U| + |P||R|)$ and $O(|R||U| + |P||R|)$, respectively. Considering the computing overlap between two models, the total time complexity of offline phase is $O(|P||U| + |P||R + |R||U||)$.

In order to improve the real-time performance of the recommendation, reducing the time complexity of the online phase is more important than the time complexity of the offline phase. In the online phase, the interest topics of the target user can quickly be matched from the user–topic preference model, and the corresponding recommendation list can be generated by combing the topic resource affiliation model. Therefore, the time complexity of calculating the target user's preference score for a certain resource in the online recommendation process is $O(|X|)$, where $|X|$ is the number of interest topics of the target user.

## 4. Experimental evaluation

*4.1 Dataset*

To evaluate our proposed approach, we conduct experiments in two real-world datasets MovieLens 20M (https://grouplens.org/datasets/movielens/20m/) and CiteULike (www.citeulike.org/). The dataset MovieLens 20M records the tagging of each movie by each user.

Users are randomly selected, and each user has rated at least 20 movies. In order to reduce the sparsity, the users whose tagging times were less than 4 times were deleted. After data preprocessing, it contained 18,211 tags, 19,441 movies and 3,538 users. Similarly, the dataset CiteULike is a paper bookmarking site that allows users to submit and tag papers to help users discover papers relevant to their field of study, and it contains 90,291 tags, 440,132 resources and 4,226 users.

### *4.2 Evaluation approach*

To examine the performance of our proposed approach comprehensively, the evaluation is conducted from two aspects: the quality of recommendation and the efficiency of recommendation. We adopted precision, recall and F-measure to evaluate the quality of recommendations. If $T(u)$ is the user's actual feedback list on the test set, $R(u)$ is the recommended resources list; the indexes of the quality of recommendation are defined as follows.

$$\text{Precision} = \frac{1}{|U|}\sum_{u}\frac{|R(u)\cap T(u)|}{|R(u)|}, \tag{7}$$

$$\text{Recall} = \frac{1}{|U|}\sum_{u}\frac{|R(u)\cap T(u)|}{|T(u)|}. \tag{8}$$

We use the time complexity and the actual running time to evaluate the recommendation efficiency. The total recommended time includes offline time and online time. The offline time is the time taken from the process of starting to process the data to generate the user–topic preference model and the resource–topic affiliation model. The online time is defined as the time taken between the start of acquiring the interest preferences of the target user and the end of generating the recommendation list. Moreover, our operating environment is that the processor is Intel® Core™ i5-8265U, the RAM is 8 GB and the system type is 64-bit.

### *4.3 Determination of parameters and verification of effectiveness*

In our proposed approach, there are some parameters that should be determined first. These parameters are the number of topics $K_{\text{topic}}$ and the value of optimal similarity combination parameters $\lambda$ and $\gamma$. In order to prevent redundancy, we take the dataset MovieLens 20M as an example to introduce the parameter determination process of the experimental dataset.

*4.3.1 Optimal number of topics.* In Section 3.1.2, we mentioned that perplexity is used to determine the optimal number of topics $K_{\text{topic}}$. The default $K_{\text{topic}}$ ranges from 20 to 300. Then we calculate the perplexity under different $K_{\text{topic}}$. The experiment result is shown in Figure 5. When the perplexity value is the smallest, the tag–topic is more clearly divided. We got the optimal number of topics $K_{\text{topic}} = 50$. By substituting the number of excellent topics $K_{\text{topic}} = 50$ into the LDA model, the "tag–topic" distribution matrix $TP$ can be obtained. The matrix describes the probability distribution of 18,211 tags under 50 topics. Some example topics extracted from tags using LDA are shown in Table I. In addition, it is worth mentioning that we used the LDA model on genism to achieve automatic adjustment of two hyperparameters alpha and eta.

*4.3.2 Optimal similarity combination parameters.* After the tag–topics are identified, we vary the value of the resource–resource similarity parameter $\lambda$ and the topic–topic similarity parameter $\gamma$ to find their optimal values, which will affect the performance of our proposed approach. We take 10-fold cross-validation to calculate the evaluation value

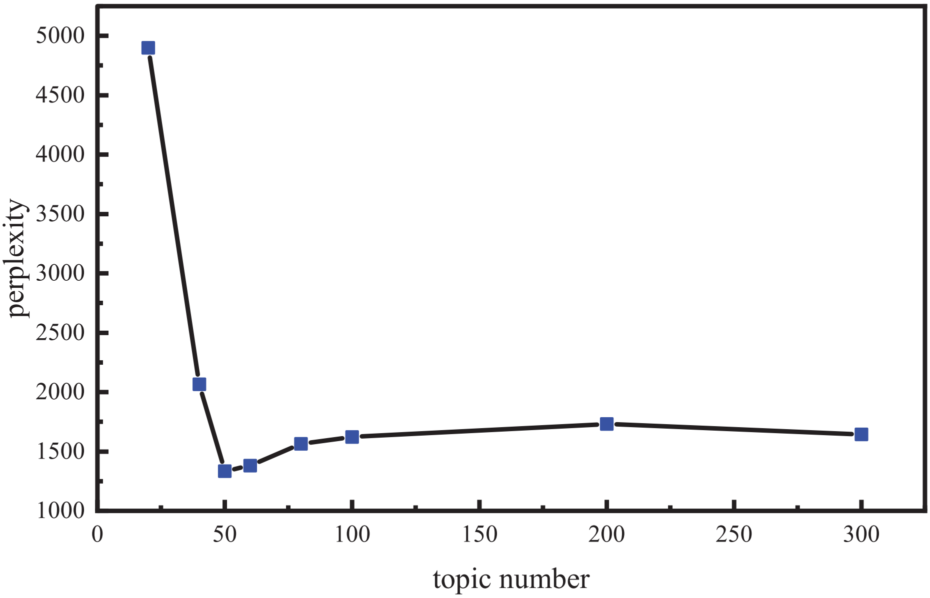


**Figure 5.**
Perplexity under different number of topics on MovieLens 20M

| Topic | Tags | | | | | | |
|---|---|---|---|---|---|---|---|
| Topic1 | Gay (0.014) | French (0.013) | French film (0.013) | France (0.010) | Ensemble cast (0.009) | Animation (0.008) | – |
| Topic2 | Sci-fi (0.011) | Superhero (0.007) | Nudity topless (0.007) | Funny (0.007) | Robert de Niro (0.007) | Mark Wahlberg (0.007) | – |
| Topic3 | bd (0.028) | DVD ram (0.024) | Criterion (0.020) | DVD video (0.016) | Betamax (0.015) | bd video (0.011) | – |
| Topic4 | Long (0.040) | History (0.010) | Dance (0.009) | British new wave (0.009) | National film registry (0.006) | Biography (0.006) | – |
| Topic5 | Sci-fi (0.018) | Atmospheric (0.009) | Classic (0.008) | Atylized (0.006) | Robots (0.005) | Aliens (0.005) | – |

**Table I.**
Example topics extracted from tags using LDA

under different similarity parameters. The top-N ($N = 10$) recommendation experimental results are shown in Figure 6. We can see that when $\lambda = 0.2$, $\gamma = 0.4$, the effect of the recommendation model is the best. From the optimal values of $\lambda$ and $\gamma$, it can be seen that whether computing resource similarity or topic similarity, the similarity based on the resource–topic matrix has a relatively greater contribution to the recommendation accuracy.

*4.3.3 Verification of effectiveness of frequency conversion with Logistic function.* In Section 3.1.1, we convert tagging frequency by Logistic function. Then we verify the validity of frequency conversion by Logistic functions. We design an experiment to compare the effects of our approach with Logistic function and without Logistic function.

Figure 7 shows the F-measure comparison between our approach with Logistic function and without Logistic function. It can be seen that the recommendation quality of our approach (i.e. the approach with frequency conversion using Logistic function) is better, and the average improvement rate is about 1.4 per cent. This shows that our conjecture is correct. The relationship between users, tags and resources can be described more accurately after frequency conversion through Logistic function.

*4.4 Comparison*

In order to verify whether our proposed approach can achieve the balance between effect and efficiency of recommendation, a comparative experiment is conducted. The CF, a general LDA-based approach and a hybrid recommendation approach named HR_Wei (Wei *et al.*, 2016) are chosen in the comparison study. A comparative experiment evaluates from the aspects of recommendation quality and recommendation efficiency. The CF recommends resources to users according to the similarity value between user ($u_x$) and user ($u_y$) who both have similar preferences. The general LDA-based recommendation takes the $UT_0$ as the input matrix in LDA modeling to obtain the tag–topic distribution matrix $W$, and the user profile and resource profile are formed by multiplying $W$ with $UT_0$ and $RT_0$,

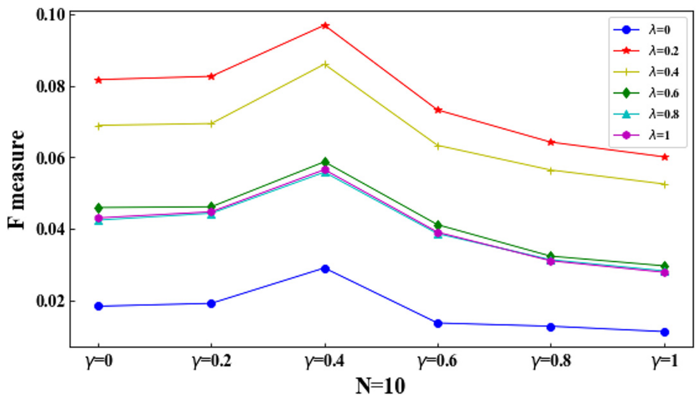


**Figure 6.**
Comparison of F-measure of different similarity combination parameters for top-N recommendation on MovieLens 20M

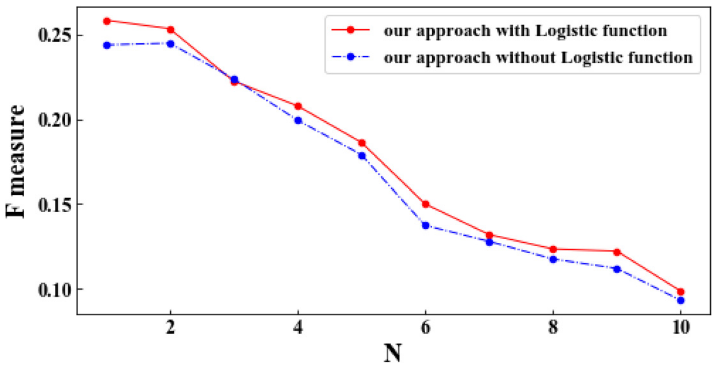


**Figure 7.**
The F-measure comparison of our approach with and without Logistic function on MovieLens 20M

respectively. Then the recommendation list is generated based on the calculation result of the similarity between the two profiles. HR_Wei constructs social networks and a preference–topic model, extracts and reconditions the social tags according to user preference based on social content annotation and enhances the recommendation model by using supplementary information based on user historical ratings (Wei *et al.*, 2016).

*4.4.1 Comparison of recommendation quality.* Figure 8 shows the evaluation results of the four approaches on MovieLens 20M. It can be seen that with the values of *N* increasing, the precision of several approaches gradually decreases. However, no matter what value *N* takes, the precision of our proposed approach is the highest. Among them, CF has the worst performance, followed by the LDA-based approach. The precision of HR_Wei and our approach is relatively close at the start, but our approach is still the best. The improvement rates of precision of our proposed approach are 49.8, 14.8 and 11.9 per cent. The recall rises sharply at the start and gradually stabilizes after reaching a certain level. Similarly, our proposed approach also has the highest recall rate. Among them, the performance of CF is the worst, the recall of LDA-based approach is close to that of HR_Wei and our approach is still the best. The improvement rates of recall of our proposed approach are 65.9, 5.7 and 3 per cent. Figure 9 shows the evaluation results of the four approaches on CiteULike. On this dataset, our approach is still the best performing.

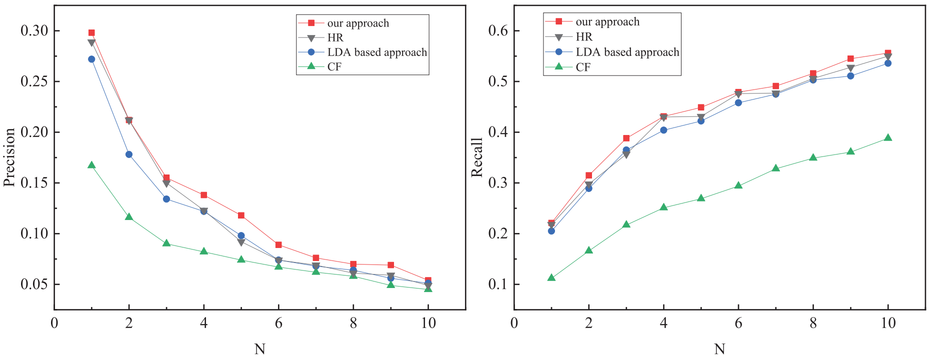


**Figure 8.**
Precision and recall comparison at top-N recommendation for each approach on MovieLens 20M

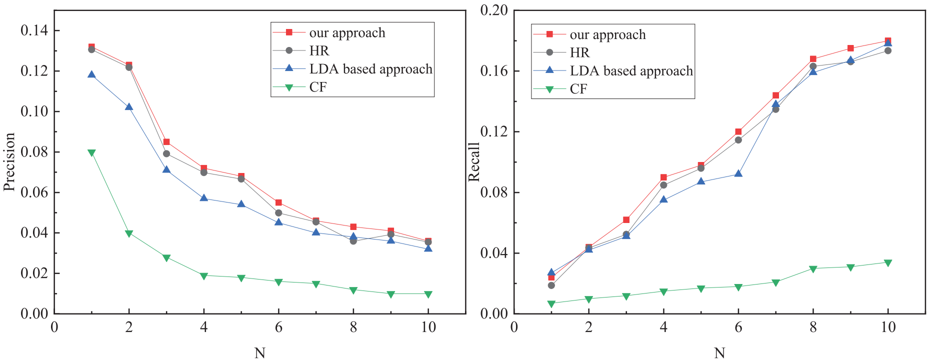


**Figure 9.**
Precision and recall comparison at top-N recommendation for each approach on CiteULike

We use paired $t$-tests to judge the significance of the approach's results for each dataset. The test results are shown in Table II. The results demonstrate that our approach significantly outperforms the baseline approaches.

*4.4.2 Comparison of recommendation efficiency.* Table III is the result of the time complexity comparison of the four approaches. The total time complexity of the four approaches is of the same magnitude. However, from the perspective of online recommendation list generation, the time complexity of our proposed approach is relatively low.

Table IV is the total time required for the four approaches on MovieLens 20M. It can be seen from Table IV that our approach takes a lot of time in offline phases, but the time spent

| Test instance | Our approach | HR_Wei approach | LDA-based approach | CF approach |
|---|---|---|---|---|
| *MovieLens 20M – Precison* | | | | |
| Mean | 0.128 | 0.118 | 0.112 | 0.081 |
| Std. | 0.077 | 0.079 | 0.069 | 0.037 |
| $t$ | | 4.434 | 5.408 | 3.635 |
| $p$ value | | 0.002 | 0.000 | 0.005 |
| *Movielens 20 M – Recall* | | | | |
| Mean | 0.439 | 0.427 | 0.417 | 0.273 |
| Std. | 0.106 | 0.108 | 0.105 | 0.089 |
| $T$ | | 4.092 | 11.027 | 23.033 |
| $p$ value | | 0.000 | 0.000 | 0.006 |
| *CiteULike – Precison* | | | | |
| Mean | 0.071 | 0.067 | 0.059 | 0.025 |
| Std. | 0.034 | 0.031 | 0.029 | 0.021 |
| $t$ | | 3.608 | 6.021 | 8.255 |
| $p$ value | | 0.000 | 0.000 | 0.000 |
| *CiteULike – Recall* | | | | |
| Mean | 0.111 | 0.105 | 0.102 | 0.02 |
| Std. | 0.056 | 0.052 | 0.055 | 0.009 |
| $t$ | | 6.163 | 3.294 | 6.112 |
| $p$ value | | 0.000 | 0.009 | 0.000 |

**Table II.** The paired $t$-tests results for each datasets

| | CF approach | LDA-based approach | HR_Wei approach | Our approach |
|---|---|---|---|---|
| Order of magnitude of total time complexity | $n^2$ | $n^2$ | $n^2$ | $n^2$ |
| Orders of magnitude of online recommendation service time complexity | $n^2$ | $n^2$ | $n^2$ | $n$ |

**Table III.** Time complexity comparison of four approaches

| | Offline time (s) | Online time (s) |
|---|---|---|
| CF approach | 0 | 1.01 |
| LDA-based approach | 235 | 0.92 |
| HR_Wei | 845 | 0.63 |
| Our approach | 638 | 0.12 |

**Table IV.** The time required for the four approaches on MovieLens 20M in online phase is short; CF approach takes less time offline, but the time spent online is longer. The general LDA-based approach is somewhere in between.

This result shows that although a certain amount of time is sacrificed in the offline optimization process proposed in our approach, it greatly improves the efficiency of online recommendation process. In addition, the implementation of our approach also illustrates the feasibility of the recommended service model that integrates the optimization process.

### *4.5 Summary and discussion*

From the experimental results on MovieLens 20M obtained above, we can conclude that when $K_{\text{topic}} = 50$, $\boldsymbol{\lambda} = 0.2$, $\gamma = 0.4$, the recommendation approach proposed in the paper is optimal. $\lambda$ and $\gamma$ are parameters for calculating the similarity of resources. The smaller the value of $\lambda$ and $\gamma$, the greater the contribution of the similarity calculation based on the resource–topic matrix to the recommendation accuracy. The experiment reflects that whether it is computing resource–resource similarity or topic–topic similarity, the contribution of similarity calculation based on the "resource–topic" matrix to the recommendation accuracy is relatively greater. It implies that the way of using the tag–topic to describe the characteristics of the resource more truly reflects the similarity of resources and topics.

It is worth noting that the recommendation quality of our approach is better than the approach removed conversion. This suggests that frequency converting by Logistic function has a positive effect on the quality of recommendations.

Although the total time complexity of the four approaches is in the same order of magnitude, from the point of view of the time to generate the recommendation list in online service phase, our approach requires a short time. This shows that the incorporated topic optimization approach proposed in this paper greatly improves the efficiency of online recommendation by strengthening the offline optimization process, and it also further proves the feasibility of the recommendation service model of the incorporated optimization process proposed in this paper. It is acceptable to sacrifice offline time to achieve better online services.

We can call the transition point from the system's self-processing optimization process (user-invisible online process) to personalized recommendation service process (user-visible online process) as "User Decoupling Point" (UDP), as shown in Figure 10. The recommendation service in social tagging is user-centric. By moving the UDP back to the right, the efficiency of online recommendation services visible to users can be improved, and user satisfaction can be improved. Therefore, from a systematic perspective, this article strengthens the invisible self-processing optimization process of the system and weakens the user-visible service process, improving the efficiency of personalized recommendation services.

In the era of big data, our approach provides a solution to the contradiction between large-scale data processing and timely response to the individual needs of users. The idea of

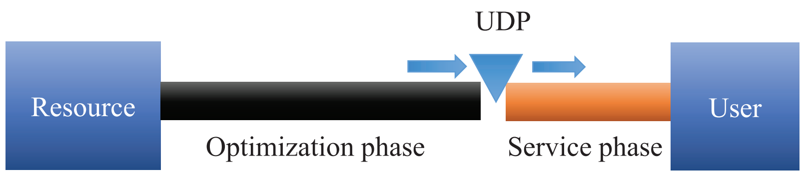


**Figure 10.**
Conceptual diagram of "User Decoupling Point" (UDP)

integrating offline processing and online service can also be applied to other service areas by strengthening the offline stage to improve both quality and efficiency of online services.

## 5. Conclusion and future work

Combining the idea of optimization before service, this paper proposes a collaborative recommendation approach that incorporates the topic optimization in social tagging. The proposed approach optimizes the tag data through topic modeling and integrates offline tag optimization phase with online recommendation phase. The experimental results show that our approach has improved the recommendation quality and efficiency compared to other approaches. Our approach effectively solves the contradiction between large-scale data processing and timely response to users' personalized needs. One important property of our approach is that we use the Logistic function to convert the tagging frequency in offline optimization phase. The experiment results show that it improves the quality of recommendations.

This paper makes an attempt to solve the contradiction between the efficiency and effectiveness of recommendation service in social tagging by incorporating offline topic optimization into online recommendation service. However, there are some problems that should be further explored. (1) This paper examines user's interest topics from a static perspective. The user's implicit preferences can be mined through tags, but the user's tagging behavior is affected by time, and the feedback of the user's interest also changes with time, that is, there is a "user's interest drift" phenomenon. Future research will consider the influence of time to improve the proposed approach. The user's annotation timing can be considered in the research to improve user interest mining. The time-forgetting curve can be used to describe user interests. (2) We use LDA for topic optimization, and there are actually many methods that can be used for topic optimization, including enhanced LDA methods, clustering methods, etc. These different topic optimization methods are worth exploring. (3) The integration of deep learning methods and the solution of the cold start problem are also future work that should be paid attention to. Methods based on deep learning can be used for topic modeling, such as topic models based on the Generative Adversarial Network, and can also be used to characterize massive data of users, resources and tags to learn the essential characteristics of datasets from samples. We can try to alleviate the cold start problem by mining users' social relationships.

## ORCID iDs

Xuwei Pan http://orcid.org/0000-0001-5583-4000
Xuemei Zeng http://orcid.org/0000-0001-6234-4683
Ling Ding http://orcid.org/0000-0002-7016-1390

## References

Belém, F.M., Almeida, J.M. and Gonçalves, M.A. (2017), "A survey on tag recommendation methods", *Journal of the Association for Information Science and Technology*, Vol. 68 No. 4, pp. 830-844.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent Dirichlet Allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.

Chen, C., Zheng, X., Wang, Y., Hong, F. and Chen, D. (2016), "Capturing semantic correlation for item recommendation in tagging systems", *Thirtieth AAAI Conference on Artificial Intelligence*, AAAI Press, Phoenix, AZ, available at: www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11994 (accessed 2 June 2020).

Chi, E.H. and Mytkowicz, T. (2007), "Understanding navigability of social tagging systems", *Presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'07)*, ACM Press, San Jose, p. 11.

Das, R., Zaheer, M. and Dyer, C. (2015), "Gaussian LDA for topic models with word embeddings", *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Vol. 1: Long Papers)*, Association for Computational Linguistics, Beijing, pp. 795-804.

Duan, J., Ai, Y. and Ii, X. (2015), "LDA topic model for microblog recommendation", in Chen, W.L., Ma, B., Zhang, M., Lu, Y.F. and Dong, M.H. (Eds), *Proceedings of 2015 International Conference on Asian Language Processing*, IEEE, New York, NY, pp. 185-188.

Golder, S.A. and Huberman, B.A. (2005), "The structure of collaborative tagging systems", *ArXiv Preprint Cs/0508082*. doi: 10.48550/arXiv.cs/0508082

Guan, Z., Wang, C., Bu, J., Chen, C., Yang, K., Cai, D. and He, X. (2010), "Document recommendation in social tagging services", *Proceedings of the 19th International Conference on World Wide Web*, Association for Computing Machinery, Raleigh, NC, pp. 391-400.

Harvey, M., Ruthven, I. and Carman, M. (2010), "Ranking social bookmarks using topic models", *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, Association for Computing Machinery, Toronto, ON, pp. 1401-1404.

Hong, M., Akerkar, R. and Jung, J.J. (2019), "Improving explainability of recommendation system by multi-sided tensor factorization", *Cybernetics and Systems*, Vol. 50 No. 2, pp. 97-117.

Ifada, N. (2014), "A tag-based personalized item recommendation system using tensor modeling and topic model approaches", *Sigir'14: Proceedings of the 37th International Acm Sigir Conference on Research and Development in Information Retrieval*, Assoc Computing Machinery, New York, NY, p. 1280.

Ifada, N. and Nayak, R. (2014), "Tensor-based item recommendation using probabilistic ranking in social tagging systems", *Proceedings of the 23rd International Conference on World Wide Web*, Association for Computing Machinery, New York, NY, pp. 805-810.

Indra, R. and Thangaraj, M. (2019), "An integrated recommender system using semantic web with social tagging system", *International Journal on Semantic Web and Information Systems (IJSWIS)*, Vol. 15 No. 2, pp. 47-67. doi: 10.4018/IJSWIS.2019040103

Jacobi, C., van Atteveldt, W. and Welbers, K. (2016), "Quantitative analysis of large amounts of journalistic texts using topic modelling", *Digital Journalism*, Vol. 4 No. 1, pp. 89-106. doi: 10.1080/21670811.2015.1093271

Kubatz, M., Gedikli, F. and Jannach, D. (2011), "LocalRank – neighborhood-based, fast computation of tag recommendations", in Huemer, C. and Setzer, T. (Eds), *E-Commerce and Web Technologies*, Springer, Berlin, Heidelberg, pp. 258-269.

Landia, N., Doerfel, S., Jäschke, R., Anand, S.S., Hotho, A. and Griffiths, N. (2013), "Deeper into the Folksonomy Graph: FolkRank adaptations and extensions for improved tag recommendations", *Computer Science*, available at: https://arxiv.org/abs/1310.1498v1 (accessed 30 July 2020).

Li, F., Shen, H. and He, T. (2011), "Tag-topic model for semantic knowledge acquisition from blogs", *2011 7th International Conference on Natural Language Processing and Knowledge Engineering (NLPKE)*, IEEE, Tokushima, pp. 221-226.

Liu, H. (2019), "Resource recommendation via user tagging behavior analysis", *Cluster Computing*, Vol. 22 No. 1, pp. 885-894.

Liu, H., Feng, S. and Yu, G. (2017), "An interest propagation based movie recommendation method for social tagging system", *2017 International Conference on Machine Learning and Cybernetics (ICMLC)*, IEEE, New York, NY, Vol. 1, pp. 130-135.

Na, L., Ying, L., Xiao-Jun, T., Ming-Xia, L. and Wang, C. (2020), "Improved user-based collaborative filtering algorithm with topic model and time tag", *International Journal of Computational Science and Engineering*, Vol. 22 No. 2-3, pp. 181-189. doi: 10.1504/IJCSE.2020.107340

Marinho, L.B., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G. and Symeonidis, P. (2011), "Social tagging recommender systems", in Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B. (Eds), *Recommender Systems Handbook*, Springer US, Boston, MA, pp. 615-644.

Newman, D., Bonilla, E.V. and Buntine, W. (2011), "Improving topic coherence with regularized topic models", in Shawe-Taylor, J., Zemel, R.S., Bartlett, P.L., Pereira, F. and Weinberger, K. Q. (Eds), *Advances in Neural Information Processing Systems*, Vol. 24, Curran Associates, Granada, Spain, pp. 496-504.

Pan, X.W. and Ding, L. (2018), "Considering correlation retarded growth for personalized recommendation in social tagging", *Presented at the WHICEB 2018 Proceedings*, AIS Electronic Library (AISeL), Wuhan, p. 65.

Pan, X.W., Ding, L., Zhu, X.Y. and Yang, Z.X. (2017), "A social approach to high-level context generation for supporting context-aware m-learning", *Eurasia Journal of Mathematics, Science and Technology Education*, Vol. 13 No. 7, pp. 3675-3686.

Parkhomenko, A., Gladkova, O. and Parkhomenko, A. (2019), "Recommendation system as a user-oriented service for the remote and virtual labs selecting", in Auer, M.E. and Tsiatsos, T. (Eds), *The Challenges of the Digital Transformation in Education*, Springer International Publishing, Cham, pp. 600-610.

Rafailidis, D. and Daras, P. (2013), "The TFC model: tensor factorization and tag clustering for item recommendation in social tagging systems", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 43 No. 3, pp. 673-688.

Ramage, D., Hall, D., Nallapati, R. and Manning, C.D. (2009), "Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora", *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Singapore, pp. 248-256.

Richards, F.J. (1959), "A flexible growth function for empirical use", *Journal of Experimental Botany*, Vol. 10 No. 2, pp. 290-301.

Shepitsen, A., Gemmell, J., Mobasher, B. and Burke, R. (2008), "Personalized recommendation in social tagging systems using hierarchical clustering", *Proceedings of the 2008 ACM Conference on Recommender Systems – RecSys '08*, ACM Press, Lausanne, pp. 259-266.

Shokeen, J. and Rana, C. (2020), "A study on features of social recommender systems", *Artificial Intelligence Review*, Vol. 53 No. 2, pp. 965-988.

Symeonidis, P. (2009), "User recommendations based on tensor dimensionality reduction", in Iliadis, M., Tsoumakasis, V. and Bramer (Eds), *Artificial Intelligence Applications and Innovations III*, Springer US, Boston, MA, pp. 331-340.

Symeonidis, P., Nanopoulos, A. and Manolopoulos, Y. (2010), "A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22 No. 2, pp. 179-192.

Wang, C. and Blei, D.M. (2011), "Collaborative topic modeling for recommending scientific articles", *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, San Diego, CA, pp. 448-456.

Wang, J. and Lv, J. (2020), "Tag-informed collaborative topic modeling for cross domain recommendations", *Knowledge-Based Systems*, Vol. 203, p. 106119. doi: 10.1016/j.knosys.2020.106119

Wang, L. (2017), "Personalized movie recommendation based on social tagging systems", *Presented at the 2017 7th International Conference on Advanced Design and Manufacturing Engineering (ICADME 2017)*, Atlantis Press, Paris, France, pp. 412-416.

Wei, S., Zheng, X., Chen, D. and Chen, C. (2016), "A hybrid approach for movie recommendation via tags and ratings", *Electronic Commerce Research and Applications*, Vol. 18, pp. 83-94.

Xie, H., Li, X., Wang, T., Lau, R.Y.K., Wong, T.-L., Chen, L., Wang, F.L. and Li, Q. (2016), "Incorporating sentiment into tag-based user profiles and resource profiles for personalized search in folksonomy", *Information Processing & Management*, Vol. 52 No. 1, pp. 61-72.

Xu, B., Lin, H., Lin, Y. and Guan, Y. (2020), "Integrating social annotations into topic models for personalized document retrieval", *Soft Computing*, Vol. 24 No. 3, pp. 1707-1716.

Yao, J., Wang, Y., Zhang, Y., Sun, J. and Zhou, J. (2018), "Joint Latent Dirichlet Allocation for social tags", *IEEE Transactions on Multimedia*, Vol. 20 No. 1, pp. 224-237.

Zhang, Z., Zeng, D.D., Abbasi, A., Peng, J. and Zheng, X. (2013), "A random walk model for item recommendation in social tagging systems", *ACM Transactions on Management Information Systems*, Vol. 4 No. 2, pp. 1-24. doi: 10.1145/2490860

Zhang, Z.K., Zhou, T. and Zhang, Y.C. (2011), "Tag-aware recommender systems: a state-of-the-art survey", *Journal of Computer Science and Technology*, Vol. 26 No. 5, p. 767.

Zhen, Y., Li, W.J. and Yeung, D.Y. (2009), "TagiCoFi: tag informed collaborative filtering", *Proceedings of the Third ACM Conference on Recommender Systems*, ACM Press, New York, NY, pp. 69-76.

Zhong, S., Lei, K., Huang, X. and Wu, J. (2017), "Topic representation: a novel method of tag recommendation for text", *2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA)*, IEEE, NY, New York, pp. 671-676.

Zhou, T.C., Ma, H., Lyu, M.R. and King, I. (2010), "Userrec: a user recommendation framework in social tagging systems", *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, Atlanta, GA*, 11–15 July 2010.


## Further Reading


Agarwal, D. and Chen, B.C. (2010), "fLDA: matrix factorization through Latent Dirichlet Allocation", *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, Association for Computing Machinery, New York, NY, pp. 91-100.



**Corresponding author**
Xuwei Pan can be contacted at: panxw@zstu.edu.cn