Electronic Commerce Research and Applications 18 (2016) 83–94

Contents lists available at ScienceDirect

Electronic Commerce Research and Applications

journal homepage: www.elsevier.com/locate/ecra

A hybrid approach for movie recommendation via tags and ratings ☆

Shouxian Wei, Xiaolin Zheng *, Deren Chen, Chaochao Chen

College of Computer Science and Technology, Zhejiang University, 38 Yugu Road, Hangzhou, Zhejiang 310027, China

ARTICLE INFO

Article history:
Received 17 March 2015
Received in revised form 2 January 2016
Accepted 2 January 2016
Available online 8 February 2016

Keywords:
Hybrid movie recommendation
Tags and ratings
Personalized recommendation
Singular value decomposition (SVD)
Multiple correspondence analysis (MCA)

ABSTRACT

Selecting a movie often requires users to perform numerous operations when faced with vast resources from online movie platforms. Personalized recommendation services can effectively solve this problem by using annotating information from users. However, such current services are less accurate than expected because of their lack of comprehensive consideration for annotation. Thus, in this study, we propose a hybrid movie recommendation approach using tags and ratings. We built this model through the following processes. First, we constructed social movie networks and a preference-topic model. Then, we extracted, normalized, and reconditioned the social tags according to user preference based on social content annotation. Finally, we enhanced the recommendation model by using supplementary information based on user historical ratings. This model aims to improve fusion ability by applying the potential effect of two aspects generated by users. One aspect is the personalized scoring system and the singular value decomposition algorithm, the other aspect is the tag annotation system and topic model. Experimental results show that the proposed method significantly outperforms three categories of recommendation approaches, namely, user-based collaborative filtering (CF), model-based CF, and topic model based CF.

1. Introduction

Many online movie platforms provide abundant resources, which brings convenience for the general audience. However, given the rapid growth of network information resources, users have to spend plenty of time in searching movies that they are interested. Helping users to find resources that they want rapidly has become an important requirement. With the success of the movie market, the addition of new movies causes a problem of information overload. Recommender systems have been regarded as effective solutions to the information overload problem and have become an important research field.

As effective information resources, ratings may significantly affect the recommendation of unknown social media content. For example, MovieLens[1] predicts the unknown ratings of users on movies according to the existing known ratings. A recommendation is then made based on the similarities among movies and the common features that connect users. MovieLens can also provide an accurate suggestion list according to the social ratings. Tagging, which is an effective recommendation tool or extension of recommender systems, has also been extensively applied to major social media recommender systems recently (Movahedian and Khayyambashi 2015). Primarily, tags directly reflect the tastes and preferences of an individual consumer toward media content. In addition, tags do not have a strict organizational structure. Nonetheless, tags provide rich and clear topic information. As a result, information with long comments is omitted by users. Thus, tags can catalog user experience more flexibly. Therefore, tagging has increased in popularity and is extensively applied in personalized recommendations (Jung 2014). In addition, the main user activities involve posting text contents and annotating contents, as well as ratings or tags. Fig. 1 shows the typical online movie platform. All social content annotations (Chelmis and Prasanna 2013), (Hoi et al. 2011) are tagged and classified. As depicted in Fig. 2, movie recommendations are intended to improve user experience by providing a set of movies that are relevant to the tags and can be given a high score. The online movie platform is a typical representation of the collaboration and common interests of users (Colace et al. 2015). Therefore, we developed our movie recommendation strategy based on annotations.

With the development of online social networks, information overload has become a more severe problem. The Recommender system, as an effective tool to information filtering (Liu et al. 2013a), has recently been the focus of considerable attention. The recommender system suffers from poor accuracy problem because the user and the item have not interacted with the system

☆ Some content of this paper has been published on the 11th IEEE international conference on e-business engineering (ICEBE 2014).

* Corresponding author.

E-mail addresses: wsx@zju.edu.cn (S. Wei), xlzheng@zju.edu.cn (X. Zheng), drc@zju.edu.cn (D. Chen), zjuccc@zju.edu.cn (C. Chen).

[1] http://movielens.org.

http://dx.doi.org/10.1016/j.elerap.2016.01.003
1567-4223/© 2016 Elsevier B.V. All rights reserved.

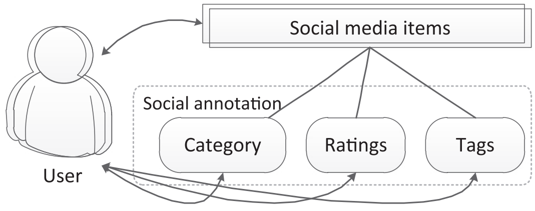

Fig. 1. A typical online movie platform.

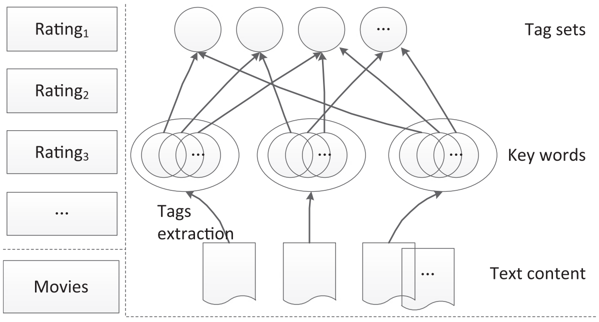

Fig. 2. Extraction of tags and movie ratings.

yet. Reducing the effects of the cold-start problem on personalized recommendation has become a research topic that has been extensively investigated to grantee premise accuracy. Moreover, the current recommender system, particular social media content recommendation (Lee and Phang 2015) must be improved. The following aspects are key to improving the recommender system: low recommendation precision and low automatic degree. Most of the recommendations made by users in the system are based on content filtering and on keywords used in searching. In addition, service is not persistent. Many recommendations are based on the login information of the user that is on record, browsing history of the user, and purchase information.

Related methods have been applied to different social media objects on the Internet, including micro-blogs (Mishne 2006), questions and answers, e-commerce, web bookmarks (Wetzker et al. 2010), blind dating systems, instant messaging applications, social games, social networks for business, music (Jschke et al. 2007), and photos (Sigurbjrnsson and van Zwol 2008). The hybrid method of recommendation is also employed in many applications, such as the temporal purchase patterns derived from sequential pattern analysis (SPA) (Choi et al. 2012). On one hand, these applications derived implicit ratings that can be used in online transaction data for collaborative filtering (CF). On the other hand, these applications used the temporal purchase patterns to eliminate the harmful effect on recommendation services through SPA. Eventually, CF and SPA are eventually integrated, improving recommendation quality. Each approach has its advantages. Meanwhile, the different approaches have their respective suitable application scenarios. Therefore, according to different scenarios, selecting different methods and making them work together can significantly improve recommendation performance. Another example is StereoTrust (Liu et al. 2013b), a trust model inspired by real-life stereotypes. StereoTrust analyzes the historical behavioral information to build the trust relationships to come up with recommendations. However, most of the aforementioned existing approaches, only generally highlight special scenarios. Moreover, the extensibility of these applications must be further improved.

In this study, we propose a hybrid movie recommendation approach via social tags and preferred ratings. First, we extract, normalize, and recondition social tags according to user preference. The user preference is based on social content annotation, which includes tags and ratings. Then, our model can benefit from unifying the potential capability of a personalized scoring system (e.g., singular value decomposition [SVD] of a matrix) and a tagging system (e.g., the preference-topic model, tag normalization and reconditioning). Finally, in terms of the recommendation results, our hybrid method has outperformed the existing user-based collaborative filtering (CF) algorithms including the user-based CF, the CF model, and the topic model based CF model.

The current paper is structured as follows: Section 2 reviews the related literature on the different recommendation approaches. Section 3 describes the hybrid model for movie recommendation. Section 4 describes the experiments and analyzes the experimental results. Section 5 concludes the paper and discusses future works.

2. Related works

In this section, we present related works on recommender systems, including recommendations based on social tagging, recommendations based on a topic model, and recommendations based on matrix factorization (MF) approaches.

2.1. Recommendations based on social tagging

Social tagging is extensively used in the industry. This process has generated many new applications, including Flickr, Delicious, and Last FM. Any user can make unconstrained annotations based on their own understanding and interests (Wen et al. 2014, 2012), and all annotations are visible to other users. In fact, this annotation mode is open and shared, and it reflects the actual view and understanding of the user. This concept revolutionizes information-resource organization, retrieval, and sharing. Moreover, annotation is generated based on swarm intelligence. The difference between tag annotation and previous recommender systems is how users select the keywords. This process reflects understanding toward resources. This understanding plays an important role in the links among users. Such tagging systems are dynamic. The problem that we face and need to overcome is the establishment and evolution of dynamic tags. The problem can be addressed by two methods (Yao et al. 2012), namely, taxonomy extraction and evolutionary taxonomy. A tag represents the main characteristics of information resources and simultaneously covers the relationships between users and resources and the relationships among users. A tag embeds the features derived from content and association. A recommender that displays effective content and CF may be developed by using tag-based datasets (Guy et al. 2010, Zhang et al. 2011). Given the aforementioned benefits of using social tagging, we not only take tags as the original data of our model, but also optimize tag data to make recommendations that are more accurate.

2.2. Recommendations based on a topic model

Topic models are used to discover topics by adopting a hierarchical Bayesian analysis of original texts (Blei et al. 2010) in a document. As the simplest topic model, latent dirichlet allocation (LDA) has been used in many applications, including recommender systems (Blei et al. 2003). Researchers have recently focused on the use of LDA to mine useful and rich-text content. Wang and Blei (2011) proposed collaborative topic regression (CTR), which combines the merits of MF and probabilistic topic modeling. Purushotham et al. (2012) combined CTR with SoRec (Ma et al. 2008) to generate a consistent and compact feature representation method called CTR-SMF to improve recommendation performance. However, all users are treated differently because CTR-SMF uses

SoRec to exploit social relationships. In another study, (Chen et al. 2014) point out that CTR-SMF ignores another important source of information: context. In the current study, we focus on overcoming this disadvantage of using CTR-SMF. Thornton and McDonald (2012) proposed a directory system that was constructed through a collaborative approach typically from a certain angle and with a specific value and expectation. This approach uses global tag co-occurrence to make recommendations for partially tagged photos. However, the approach does not consider the context of the user and his/her interaction with other users in the system. This co-occurrence approach is applied as a baseline in our experiment.

2.3. Recommendations based on matrix factorization (MF) approaches

The concept of MF is from the SVD in mathematics. An SVD-based recommendation approach (Ricci et al. 2011) is a combination of the baseline prediction method and the random grading prediction algorithm. Koren (2008) expresses the score prediction formula as follows:

$$\hat{r}_{ui} = \mu + b_i + b_u + q_i^T p_u \tag{1}$$

where μ is a training set within the arithmetic mean of all of the scores and b_u reflects the deviation degree of the ratings of user u from the global averages. Similarly, b_i reflects the item popularity in comparison with the average. Vector q_i indicates the popularity of the various features of a certain item for all users. Vector p_u reflects the popularity degree of all of the entries for a user. These components integrated the quantity features and the content features to enhance the effect of the recommended approaches.

MF (Hu et al. 2008, Mnih and Salakhutdinov 2007, Shan et al. 2010) is a popular approach for CF. This approach factorizes a user-item rating matrix into a user-specific matrix and an item-specific matrix. Then, the sum-of-squared-errors objective function is generally minimized with quadratic regularization terms. Social MF (Chen et al. 2013a, Jamali and Ester 2010) incorporates social relationships into the MF to improve the recommendation performance. Social regularization terms are typically incorporated to constrain the difference in taste between two users. For example, (Jamali and Ester 2010) used the average latent features of the direct neighbors of a user as the sole influence of the user in question based on the assumption that his/her taste is close to the average tastes of his/her friends or the members of his/her network. Implicit feedbacks can show the preference of a user by combining historical rating data (for example, the user can read a movie as a favorite positive feedback). In our approach, we reconsider the historical rating preferences. the rating preference is a vector that represents each movie score from a user. Tensor decomposition (Karatzoglou et al. 2010) is a CF algorithm based on multidimensional characteristics of users. Kilmer and Martin (2011) mainly introduced the HOSVD. HOSVD can obtain an approximate result of the original tensor after decomposition. We build the 3-dimensional tensor for HOSVD by users, movie items, and context. The context includes the ratings and tags. We use the ratings as an aspect of the *tag*-dimension of tensor. Therefore, the multi-dimensional features include user ratings features and tag annotations.

The recommender system has significantly progressed. Nonetheless, its effectiveness and adaptability in actual applications must be further improved. CF technology can be extended as the core of the recommended system. CF technology issues include cold start and high cost of hysteresis data and migration. These problems limit further application of the CF. Improving and solving the aforementioned problems will cause the CF technology to not only generate significant commercial interests but also enhance the loyalty of users. Our research is beneficial for exposing the defect in CF technology because our study expands the research scope and explores the changes in application (Adomavicius and Tuzhilin 2005).

3. Proposed method

In this section, we present our method, which is a hybrid approach for movie recommendation via tags and ratings. First, we formalize the social movie network (SMN) and define notations. Then, we describe the global framework and our motivation for creating our model. Next, we describe our preference-topic model with social tagging normalization and reconditioning. Finally, we provide our fusion recommendation method using an SVD-based algorithm.

3.1. Preliminaries

A few SMN employ the rating and tagging system. The best rating or tagging system should achieve the relative objective in which users are loyal moviegoers, particularly movie geeks. If an individual is not a professional critic, then he/she must select a movie that is worth watching. Thus, before watching a movie, most movie fans usually check the evaluations and ratings of movies made by others through searching the Internet or reading comments from critics on social movie websites. Such systems must respect the aesthetic and taste of a large number of fans and cannot insist a certain type of movie on users to prefer a certain type of movie. Instead, these systems should give priority to social factors as much as possible. Therefore, social media networks can describe the users, media items, tags, ratings, and assignments of tags to resources. The data structure of these networks can be represented by a four-tuple F, as follows:

$$F = \langle U, I, TA, \sigma, RA, \varsigma \rangle \tag{2}$$

where U is a set of users, I is a set of media items, TA is a set of tags, and σ is a ternary relation, TA represents tag assignments $\mathrm{TR}_{\mathrm{tag}}$, and RA is a set of ratings and ς is a ternary relation $\mathrm{TR}_{\mathrm{rating}}$, which are written as follows:

$$\mathrm{TR}_{\mathrm{tag}} \subseteq \{\langle u \times i \times ta \rangle : u \subseteq U, i \subseteq I, ta \subseteq TA\} \tag{3}$$

and

$$\mathrm{TR}_{\mathrm{rating}} \subseteq \{\langle u \times ra \times i \rangle : u \subseteq U, ra \subseteq RA, i \subseteq I\} \tag{4}$$

For example, the SMN describes the users, movies, topics, and movie annotations that correspond to the topics and personalized ratings. The data structure of the SMN is expressed as follows:

$$\mathrm{SMN} = \langle U, M, TO, \upsilon, RT, \tau \rangle \tag{5}$$

where U is a set of users, TO is a set of topics, M is a set of movies, and υ is a ternary relation. This relation denotes the annotation assignments as follows:

$$\upsilon \subseteq \{\langle u \times m \times to \rangle : u \subseteq U, m \subseteq M, to \subseteq TO\} \tag{6}$$

Furthermore, RT is a set of ratings and τ is a ternary relation that is written as follows:

$$\tau \subseteq \{\langle u \times rt \times m \rangle : u \subseteq U, rt \subseteq RT, m \subseteq M\} \tag{7}$$

These systems can enhance their value by utilizing the social annotations of users in recommendations. In summary, the SMN is a social media network. Thus, it can be represented by the aforementioned formalization description. The SMN is a special form of social media network, and both are the union sets of rating relations $\mathrm{TR}_{\mathrm{rating}}$ and tagging relations $\mathrm{TR}_{\mathrm{tag}}$, expressed as follows:

$$F \asymp \mathrm{SMN} = \{\mathrm{TR}_{\mathrm{tag}} \cup \mathrm{TR}_{\mathrm{rating}}\} \tag{8}$$

According to the aforementioned formalization description for the SMN, a score in a ternary relation TR_{rating} exists from a user to a movie item. The score derived from the data history of a user who provided a particular movie rating (“user ID”, “movie ID”, and “rating”) is used to predict the movie that the user will watch and the number of points that were not scored. Score r_{um} of the movie is applied along with the user-film score as a matrix. This matrix is a sparse matrix structure, and many of its elements are unknown. The usual corresponding fill rate is less than 1%. Each user is assumed to have multidimensional vector $\vec{u}$ that states his/her preference for a different movie style. Every movie m is also assigned multidimensional vectors of different styles according to user preference. Thus, the film-score matrix can be estimated using the following formula:

$$\hat{r}_{um} = \vec{u} * \vec{m}^T \tag{9}$$

The vectors of $\vec{u}, \vec{m}$ are the dimensions of the row vector of the user and movie item respectively. u represents the user and m represents the movie item. Gradient descent methods are often used to train the two variables. The results converge through several dozens of iterations. This traditional SVD method is prone to over-fitting. Thus, constraints must be implemented to reduce the noise in the training data. We reconstruct the basic factors as shown in Fig. 3 to establish the model of the recommender system and fuse the topic-based model with matrix decomposition optimization.

3.2. Motivation and framework

Our proposed method can refactor social tagging by using tag normalization and reconditioning. In addition, the proposed method can improve the ability of fusion by applying a preference-topic model and the SVD of social relationships. We build this model through the following processes: First, we construct the preference-topic model by using SMNs. Then, we extract, normalize, and recondition the social tags according to user preference based on social content annotation. Finally, we enhance the recommendation model using the supplement information of user historical ratings. Our proposed model aims to improve the fusion ability by applying the potential effect of two aspects generated by users. One aspect is the personalized scoring system and the SVD algorithm, and the other aspect is the tag annotation system and the topic model.

We conducted an experiment to differentiate our methods from the other methods. We call our proposed methods HR1, which is only built on social tagging and its preference-topic model, and HR, which is the global recommendation based on the development of the previous recommendation. For the personalized movie recommendation, we use the process shown in Fig. 4 to verify our method. User interests are the focus of the study of retrieving similar resources or a calculation method of resource similarity (Musto et al. 2009). This method includes attracting user interests (Chen et al. 2013b), denoting these interests, and updating.

We obtain the global processes of the proposed hybrid recommendation (HR) method shown in Fig. 4 through the aforementioned analysis. The thick line box illustrates the process of hybrid recommendation Algorithm 1 (HR1), which is considered only from the related operations for social tags, its topic model, and related similarity. On the basis of the aforementioned partial process, we use the tag as the implicit feedback and construct the global preferences and latent factor model based on the historical rating of the users.

3.3. Hybrid approach for movie recommendation

We build this model mainly through the following parts: First, we construct the preference-topic model by using the SMNs. Then, we use the normalization and reconditioning methods, including social influence, social features, tag co-occurrence, tag boosting, and tag removal, to reconstruct social tagging. Third, we improve the aforementioned preference-topic model using user historical ratings. Finally, we enhance the recommendation model through the fusion via tags and ratings.

3.3.1. Social movie networks and the preference-topic model

As shown in Fig. 3, users can upload any type of information on collaborative platforms and can express their opinions about the content that they enjoyed through textual feedback or reviews (Feng and Wang 2012, Sun et al. 2013, Guo et al. 2013).

We use M to denote the set of all movies in the SMN and m_i to denote the ith movie in M. Similarly, we use U to denote the set of all users who provide tags and ratings to the movies. We use TA to denote the set of all observed tags and TA_m to denote the set of tags annotated on movie $m \in M$. Then, we use $\vec{ta}$ to denote the vector variables of tag tokens in the corpus.

We introduce a latent variable to assign topics to tags, which is similar to that of tractional topic models. This variable takes on values in a finite set of $ntopics = \{1, 2, \cdots, K\}$, where K is the topic dimension of the hyper-parameter. $\vec{z}$ is the topic assignment vector corresponding to and has the same dimension as tag token vector $\vec{t}$. In addition, $z^i_m = k$, and $k \in$ ntopics refers to the assignment of topic k for the ith tag $t^i_m \in T_m$. In other words, T_m is drawn from the distribution of tags in the kth topic in the generative process, which will discussed later.

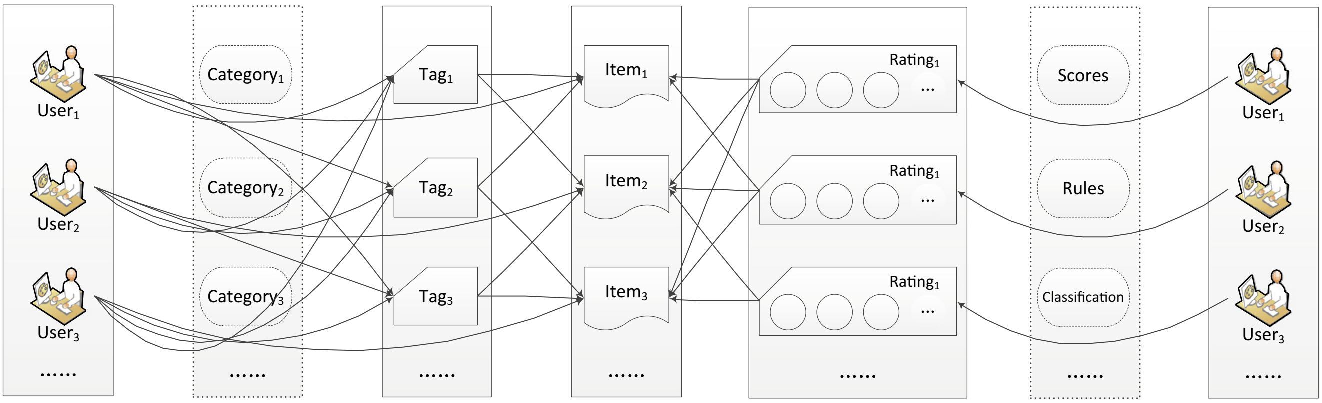

Fig. 3. Social media networks.

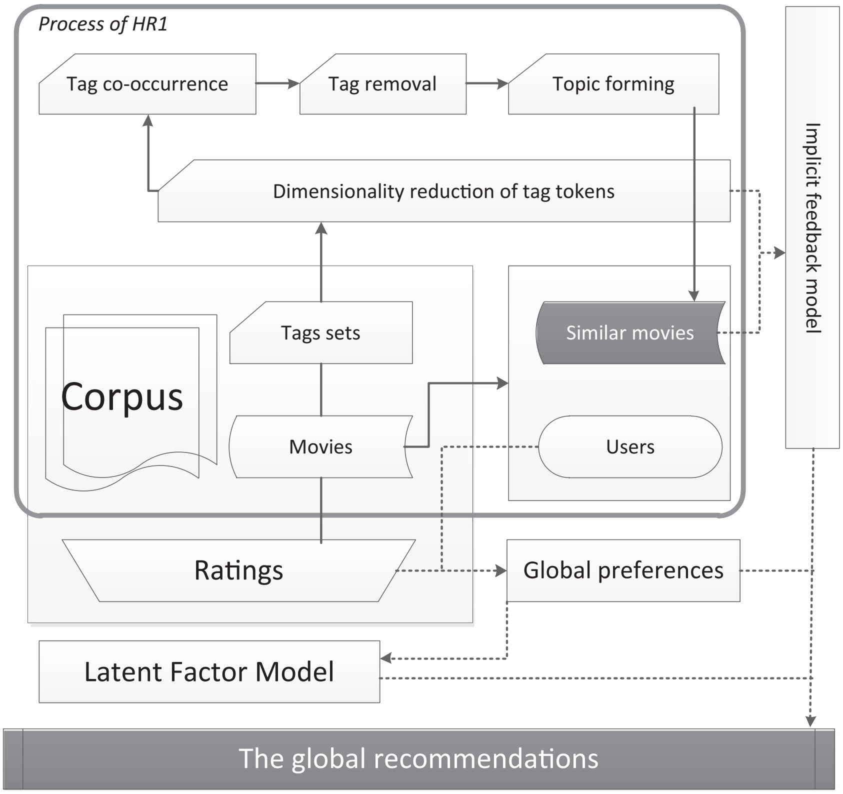

Fig. 4. Process of HR.

Table 1
Denotations of the graphical model.

M	Set of all movies in the SMN
U	Set of all users who provide tags and ratings to the movies
N	Count of topics of tag sets
α	Symmetric hyper-parameter of $\chi_{u,t}$
β	Multinomial probability $p(ta\|to)$ that merges with several topics $to \in TO$
K	Topic dimension of the hyper-parameter
$\vec{z}$	Topic assignment vector
$\chi_{u,t}$	Probability $p(to\|u)$, in which user u favors $to \in n$ topics
$\theta_{m,t}$	Probability $p(to\|u)$ by which movie m favors $to \in n$ topics
Φ	Distributions over the tags associated with each topic

We model the interest and preference of each user $u \in U$ probabilistically. We use $\chi_{u,t}$ to denote probability $p(to|u)$, in which user u favors a topic $to \in n$ topics. We collectively treat $\left\{\chi_{u,t} u \in U, t \in T\right\}$ as a random matrix denoted by X. We also assume that for each $u \in U, \chi_{u,t}(X_{u*}$, the uth row of $X)$ is drawn from the Dirichlet prior distribution $Dir(\alpha)$, which is the multinomial distribution of reference topics over the K latent topics. We should empirically set the symmetric hyper-parameter α at less than 1. Therefore, the majority of the probability mess shrinks to the corner of the $K-1$ simplex. Thus, our model should differentiate user-preferred topics effectively.

For each movie $m \in M$, we model user interest and preference probabilistically, where $\theta_{m,t}$ denotes probability $p(to|u)$ by which movie m favors a topic $to \in n$ topics. We collectively treat $\{\theta_{m,t} m \in M, to \in TO\}$ as a random matrix denoted by Θ. We also assume that for each $m \in M, \theta_{m,to}(\Theta_{m*}$, the mth row of $\Theta)$ is drawn from the Dirichlet prior distribution $Dir(\alpha)$, which is the multinomial distribution of reference topics over the K latent topics. We should empirically set the symmetric hyper-parameter α at less than 1. Therefore, the majority of the probability mess shrinks to the corner of the $K-1$ simplex. Thus, our model should differentiate movie topics effectively.

Furthermore, we assume that specific tag $ta \in TA$ has multinomial probability $p(ta|to)$ that merges with several topics $to \in TO$. A high probability indicates that ta is the expressive tag in the topic to. We similarly use matrix $\Phi = \left\{\varphi_{to,ta} : to \in TO, ta \in TA\right\}$ to denote the distributions over the tags associated with each topic. Moreover, for each row, $\varphi_{ta}(\Phi_{ta*}$, the tath row of Φ is the exact probability $p(to \mid ta)$ for each $to \in TO$ given a unique topic ta, which is also generated from a Dirichlet prior distribution $Dir(\beta)$. Symmetrically, β is the parameter used. The denotations used in this graphical model are shown in Table 1.

We use the social tag to train user preferences in this study. This tag can be represented by hybrid Algorithm 1, as shown in Fig. 5. Next, we fuse the user preferences by increasing the historical ratings to improve hybrid model 1 to construct the global hybrid algorithm.

We can obtain the topic distributions of each movie using the aforementioned model. In addition, the preference of each user for movies can be derived by inferring a topic model. We use $D_{KL}(P\|Q)$ to denote the Kullback–Leibler (KL) divergence[2] (information divergence), as shown in the following formula.

$$D_{KL}(P\|Q) = \sum_{x \in X} P(x) \ln \frac{P(x)}{Q(x)} \tag{10}$$

The relative entropy is zero when both probability distributions are identical $P(x) = Q(x)$. The probability distribution P(X) of information entropy is expressed as follows:

$$H(P) = -\sum_{x \in X} P(x) \ln P(x) \tag{11}$$

[2] http://en.wikipedia.org/wiki/Kullback-Leibler_divergence.

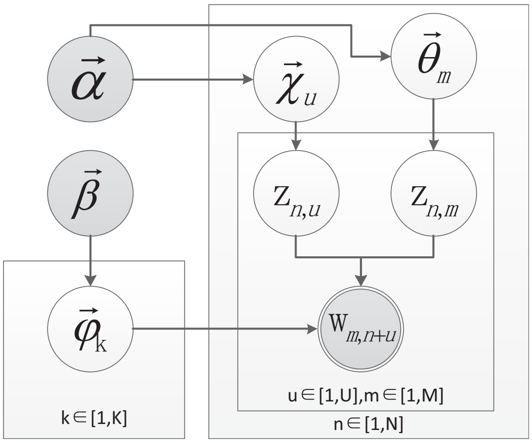

Fig. 5. Topic model based global tag annotation.

3.3.2. Social tagging and its normalization and reconditioning

In this section, we will refactor social tagging to improve the preference-topic model by using tag normalization and reconditioning. The more tag features are considered during recommendation, the more credible the recommendation results are (Zhang et al. 2011). The model does not eventually consume more computing resources. Moreover, given that not all of the tags would provide useful information in recommendation. Several tags are only considered noise. Noisy tags will reduce recommendation accuracy. Our goal is to improve recommendation performance by finding suitable approaches by reconstructing the tag set. However, automatically achieving such a goal is challenging. A number of tag preprocessing methods have been proposed to filter out irrelevant, non-informative, or redundant tags and enhance the effect of several tags as follows:

3.3.2.1. Social influence. The semantics of social tags are usually ambiguous because of the constant changes in user interests or informal definitions. Thus, directly applying social tags into the system is difficult (Chen et al. 2014b). The social influence (Hu et al. 2012) of a user is always decided by the overall situation, which involves the composition of behavior from the beginning to recent events in the social media network. If a user has a high social influence, then he/she is a positive participant. Conversely, if a user has a low social influence, then he/she is a noisy participant. Therefore, we exclude a few noisy participants (approximately to 5%) from the social tagging system to enhance global recommendations.

Topological potential is used to calculate social influence among users.

The Gaussian-type definition of topological potential is written as follows:

$$TP(v_i) = \frac{1}{n}\sum_{j=1}^{n} TP(j \to i) = \frac{1}{n}\sum_{j=1}^{n}\left(m_j \times e^{-\left(\frac{d_{j\to i}}{\alpha}\right)^2}\right) \tag{12}$$

The computing method (Hu et al. 2012) of topological distance calculates the behavior of a user. We can calculate user preferences and the corresponding degrees through personalized behaviors. We use the similarity of user preference as the determinant of topological distance. Calculating the superposition summation of the similarity degree between the current user and all other users can determine the social influence of a user.

L_k is defined as the steps of trust propagation between two users or the degree of similarity between two users. If two users have no similarity, then the propagative similarity should be considered. D_{ij} is the topological distance between the ith user and the jth user, which is expressed as follows:

$$f(x) = e^x - 1 \tag{13}$$

Therefore, the following formula can be derived:

$$D_{\mathrm{ij}} = \ln\left(\frac{1}{\sum_{k\in S_{ij}} f(L_k)} + 1\right) \tag{14}$$

3.3.2.2. Social features. The social feature (Chen and Shin 2013) can be viewed in the form of tags in social media networks. $\mathrm{SF}_{\mathrm{tag}_i}$ is assumed to indicate the popularity of tag$_i$. The influence of tag tokens can often be determined according to the number of occurrences. In this study, we use the number of users who gave the tag and the total number of this tag to determine its influence on the social tagging system. Based on the number of the tags used by all users, this tag can be labeled as common, general, or rare. We operate according to the different forms of tagging. If few users rarely used this tag, then the number is small. Such tags are unique because they lack semantics in social media and should be deleted. If a user uses a tag many times, then the type of content (in this situation, a movie) can be determined from this single tag. We give this type of content additional weight, as shown in the following formula:

$$\mathrm{SF}_{\mathrm{tag}_i} = \sum_{u_j\in U} \mathrm{Count}(\mathrm{tag}_i)_{u_j} \tag{15}$$

The frequency of the user's tag indicates the frequency with which users use a tag (total number), expressed as follows:

$$F_{u_j}(\mathrm{tag}_i) = \mathrm{Count}(\mathrm{tag}_i)_{u_j} \tag{16}$$

The user relative to the entire system represents the proportion of those using the tag, i.e.,

$$RF_{u_j}(\mathrm{tag}_i) = \frac{F_{u_j}(\mathrm{tag}_i)}{f_{\mathrm{tag}_i}} = \frac{\mathrm{count}(\mathrm{tag}_i)_{u_j}}{\sum_{u_j\in U}\mathrm{count}(\mathrm{tag}_i)_{u_j}} \tag{17}$$

We can determine the deductions using the previously presented definitions. First, we can obtain the probability of each tag in the entire corpora, as follows:

$$P(Tag_i) = \frac{\mathrm{Count}(Tag_i)}{\sum_{i\in items}\mathrm{Count}(Tag_i)} \tag{18}$$

Second, the probability of an item being generated can be calculated using all tags, as follows:

$$P(Item) = \prod_{\alpha=1}^{j} P(\mathrm{tag}_\alpha) \tag{19}$$

Similarly, we can obtain the probability of a corpora being generated, as follows:

$$P(corpus) = \prod_{\beta=1}^{k} P(Item_\beta) \tag{20}$$

3.3.2.3. Tag co-occurrence and tag boosting. At the time resources are recommended, the metadata (e.g., irrelevant tag and phrases lacking semantics) must be reduced (Yanagimoto and Yoshioka 2012).

The weight of co-occurrence tags must be addressed automatically. If multiple resources are marked twice with identical tags at the same time, then both tags co-occur, and multiple tags are similar. Tag co-occurrence and the calculation of social influence were examined by Hu et al. (2012). The relationship among tags was described by Jung (2014), Sun et al. (2013), and Pham et al. (2012). Tag graphs are constructed by mining these relationships. The recurrent co-occurrence of these tags has a high degree of similarity. Co-occurrence has been extensively used as a method of discovering the relationship between the two tags. The tag co-occurrence coefficient between two tags is the number of movies for which both tags are used in the same annotation. Two different normalization methods are essentially available, namely, symmetric and asymmetric. Based on the Jaccard coefficient (Naw and Hlaing 2013), the co-occurrence of two tags t_i and t_j can be normalized using the following equation. The coefficient takes the number of intersections between the two tags divided by the union of the two tags. In asymmetric measurements, tag co-occurrence can be normalized using the frequency of one of the tags computed in the following equation, which captures the frequency of one of the tags:

$$\mathrm{CO}_{\mathrm{sy}}(t_i, t_j) = \frac{|t_i \cap t_j|}{|t_i \cup t_j|}, \ \mathrm{CO}_{\mathrm{asy}}(t_j|t_i) = \frac{|t_i \cap t_j|}{|t_i|} \tag{21}$$

The Jaccard symmetric coefficient effectively identifies equivalent tags. By contrast, asymmetric tag co-occurrence diversifies candidate tags more effectively than its symmetric counterpart (Sigurbjrnsson and van Zwol). Tagging the resource with many similar tags is unnecessary because a user usually has a clear understanding of his uploaded resource. Rather, the user may tag the resource with tags for different reasons. In our work, we apply asymmetric tag co-occurrence.

Based on the aforementioned asymmetric tag co-occurrence method, candidate tags can be selected in the social comment network, which is composed of k similar neighbors. A person with high prestige usually has influence in a social network (Liu 2007). Candidate tags are promoted by adding boosting factor b as the weight of tag co-occurrence when a user with high prestige uses tag pairs.

3.3.2.4. Tag removal. In a SMN, users often annotate the movies according to their own description. These tags may include characteristics, categories, actors, styles, and even the feelings of the user. Tags are often used as the content of clustering and retrieval that are faced with semantic gap issues during the process of use. Therefore, a tag-based model must consider tag semantic features more than traditional content-based models. However, the tags that users define are often incomplete, inaccurate, and not sidewinded because these tags are derived from the subjective perspective of users based on their scope of knowledge and experience, as well as emotions. These tags are usually not considered comprehensive responses to the authenticity of content because the knowledge range, as well as experience and emotions, are sometimes impulsive. These tags are sometimes even regarded as completely unrelated content. These tags are occasionally inspirational (mental clarity). Nevertheless, this content cannot be captured. For example, a movie is in a foreign language, and the user knows little about the language that is used in the movie. Thus, the user tags this movie with another language. Sometimes, a few scenes from the movie that have not been presented before are shown, and unrelated tags are given.

In view of such phenomenon, we must apply noisy tag removal (Chen et al. 2013b) by deleting uncorrelated tags. Determining the correlation of a tag and dealing with tag co-occurrence is discussed in this work. We adopt an algorithm based on multiple correspondence analysis (MCA) (Abdi and Valentin 2007, Zhang et al. 2013) as our tag removal algorithm.

When users give tags to movies, a Boolean relation exists between a tag and a movie. In other words, a user has given a specific movie a specific tag. This action is called a status from the tag to the movie. The values of status are represented by T and F for existence and inexistence, respectively. We use movie m_i and tag t_j as examples. In this study, we build the relationship of all movies and tags to form a Boolean matrix (assumed MT). For each value of a mentioned matrix, Boolean relation, movie, and tag, we use the following formula:

$$MT_{ij} = \begin{cases} T, & \text{if the } j\text{th tag was given to the } i\text{th movie} \\ F, & \text{otherwise} \end{cases} \tag{22}$$

The reason for reducing the total tag set is to filter out the noisy (e.g., relevance is weak and the context semantic is scarce) tags or the tags that have weak correlations with concept topics. The remaining tags can predict the preference of users in topic distribution. In statistics, MCA is a data analysis technique for nominal categorical data and is used to detect and represent underlying structures in a data set. In this present work, the categorical data is the tag set. In the original data, each tag represents a dimension. A tag set has n tags, and the original dataset has n dimensions. The dimensionality of original tag sets is reduced, and the symmetric map is checked. This map is the graphical representation of MCA, which can be used to visualize the tag feature as points in a map with dimensions that depend on the amount of reduction in dimensionality. Thus, the correlation between the tag features can be measured using the cosine value of the angle between the two vectors representing tag features. The two dimensions could capture more than 95% of the total variance. In several cases, one dimension is sufficient, whereas more than two dimensions are preferred in other cases.

Fig. 6 shows a two-dimensional symmetric map in which many points represent the tag features, such as point P_1, which denotes the first tag feature and point P_2, which denotes the second tag feature. Point P_i denotes the ith tag feature. The acute angle in the figure denotes the correlation of both tags. As shown in Fig. 6, the smaller the angle between both tag features is, the smaller the correlation between both tags. We remove the tags that have a weak correlation, which will be discussed in the subsequent section.

3.3.3. Improvement of the aforementioned model with user historical ratings

In addition to label tags, a user expresses likes and dislikes for a movie. The rating often involves activities with regard to film score. Thus, the ratings are the more intuitive response from user interest in certain types of movies aside from tags. Ratings are computed using the following basic SVD++ equation in the recommendation system:

$$\hat{r}_{ui} = b_{ui} + q_i^T \left(p_u + \sum_{k \in N(u)} \beta_k v_k \right) \tag{23}$$

In the first part of Eq. (23), $b_{ui} = \mu + b_i + b_u$ is the preference model. The value of μ is the global average ratings. The values of b_i and b_u are the offset values of relative average ratings for the rating of the ith item and the rating of user u respectively.

In the second part of Eq. (23), $q_i^T p_u$ refers to the latent factor model. This part concludes that an item has the attributes and the preferences of a user for these attributes through the ratings from the users. The rating of the item from the user can be obtained using the dot product between d-dimensional user feature vector p_u and item feature vector q_i. The value of d is the preset experimental parameters, which is the dimensional count that the item has of the properties (as shown in the experiment).

In the third part, $q_i^T \sum_{k \in N(u)} \beta_k v_k$ is the implicit feedback model, where $N(u)$ is the implicit feedback set, v_k is the implicit feature vector, and β_k is the corresponding weight for the attributes, which is usually set at $\beta_k = |N(u)|^{-0.5}$. The performance of the recommendation system can be improved by introducing the implicit feedback information.

3.3.4. Fusion of recommendation via tags and ratings

Users can personalize their taste because of the particularity of social media and keywords, such as tags. The tag itself is a key

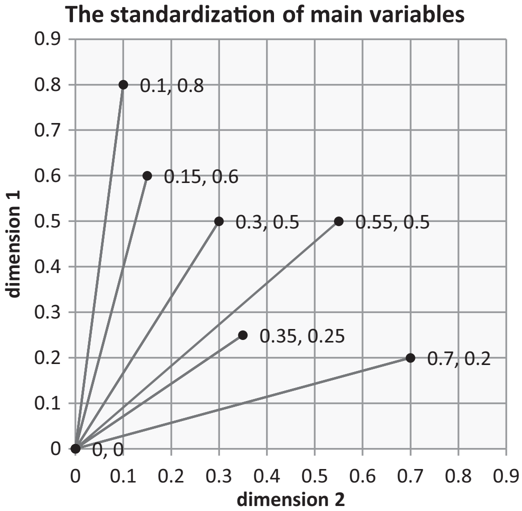

Fig. 6. Two-dimensional symmetric map (standardization of main variables).

word that provides project information from the user to social media entries. In addition to the convenience of the system of classification and indexing, tags can also be part of an active recommendation system. A number of tags tend to provide the same type of entries that belong to the same topic. The topic is usually obtained through information extraction, which is implicit distribution. Fig. 7 shows an extension of the graphical model that considers historical ratings. The gray shaded part is obtained from Fig. 5. The blue solid line box on the left denotes the feedback vector of user interest through the submission of user ratings $R_{u,i}$. By contrast, the V_i denotes the vector of topics from user u to the movie item i. The dotted box in the center denotes that the analysis of the records of the user and what the user left in the system to determine the points of user interests. The blue solid line box on the right denotes the vector of the point of user interests U_{V_j} based on the different tag weights.

We use the statistical method of term frequency-inverse document frequency (TF-IDF)[3] to obtain the importance evaluation of a tag or a word for the one in a file sets. In the preference-topic model, we have obtained the global topic distributions in the corpus and individual topic preference distributions for the movies that have been rated and tag annotated by a user. The importance of topics for a user increases proportionally according to the number of occurrences in the movie topic preference of the user, but, at the same time, decreases proportionally according to the number of occurrences in the corpus. Therefore, we adopt TF–IDF as the weighted method of a topic; we take the global probability as the term frequency (TF) and the preference probability as the inverse document frequency (IDF). In addition, a user creates a tag based on his/her own personality or interest, and the tag itself does not provide information of weight. We combined topic information and tag information as feedback, and added tag sets and topic distributions to obtain the following formula.

$$p_u^1 = p_u^0 + W(u,m)\sum_{m\in K(u)} v_{topic(m)} + \frac{1}{\sqrt{|T(u)|}}\sum_{n\in T(u)} v_{tag(n)} \tag{24}$$

The preference or interests of a user can be acquired by using explicit and implicit method. User interest vector p_u^0 is obtained based on the submission of the user. Whereas user interest vector p_u^1 is obtained by analyzing the records of the user and what the user had left in the system to determine the points of user interests. Moreover, $K(*)$ and $T(*)$ are for the topics and the tags of user u. The formula $W(*,m)$ represents the weight of the topic m. The representation of user interests can be denoted using the following tag sets:

$$q_i^1 = q_i^0 + W(u,m)\sum_{m\in K(i)} v_{topic(m)} + \frac{1}{\sqrt{|T(i)|}}\sum_{n\in T(i)} v_{tag(i)} \tag{25}$$

where q_i^0 is the former movie items feature vectors. q_i^1 is the latter movie items feature vectors. And $K(*)$ and $T(*)$ are for the topics and the tags of the ith item. The formula $W(*,m)$ represents the weight of topic m. The high value of the weight indicate a significant correlation degree.

In conclusion, we obtain a global formula of matrix decomposition based on the global ratings and the improved tags, as follows:

$$\hat{r}_{ui} = \mu + \mu_u + \sum_{j\in N(u,i)} S_{ij}\left[\frac{1}{\sqrt{|N(u,i)|}}(r_{uj} - \overline{r_u})\right] + (p_u^1)^T q_i^1 \tag{26}$$

where S_{ij} is the weighted average similarity from the tagging and rating of the two movie items for user m. The part of the tagging is from the aforementioned Eq. (10) (KL distance). The part of the ratings is from the cosine similarity of both items. In particular, when tags are inadequate, the similarity of both items entirely depends on rating.

4. Experimental evaluation

In this section, we conduct several experiments to compare the recommendation quality of HR1 and HR with other state-of-the-art CF recommendation methods. Our experiments aim to address the following questions: (1) How are tag sets filtered and reconstructed the tag sets? (2) How are the parameters for optimizing the topics model tuned? (3) How does the precision, recall and F-measure of our approach compare with that of existing contrastive algorithms?

4.1. Datasets

We validate our approach using the data from MovieLens[4]. MovieLens is a virtual community website that allows users to provide feedback on how much they liked the movie and annotate the movie based on their preferences. In addition, MovieLens is a recommender system that primarily applies CF technology. The dataset used in this experiment is shown in Table 2. We restrict the dataset to be validated to 7155 movies to avoid the negative influence from data with empty tags. Each movie is rated and tagged more than once.

4.2. Tag analysis

We divide the movie items into 11 groups with 927 movie items corresponding to the tag sets in each group. All movie items are considered integer variables. The corresponding values are obtained from the data column, tagWeight, which represents the marked times for a movie. The first group metric of tagged movies for MCA results is shown. Table 3 shows the iterative history of MCA. In addition, Table 4 shows the summary of the process. This table shows that 1000 iterations have been conducted.

We can obtain results through the iteration of each data group, as shown in Figs. 8 and 9. We select 95% of the tags that tend to have close semantics. For example, most of the objective points (Fig. 8) are the concentration of semantic correlations that are

[3] https://en.wikipedia.org/wiki/Tf-idf.

[4] http://www.grouplens.org
http://www.imdb.com
http://www.rottentomatoes.com.

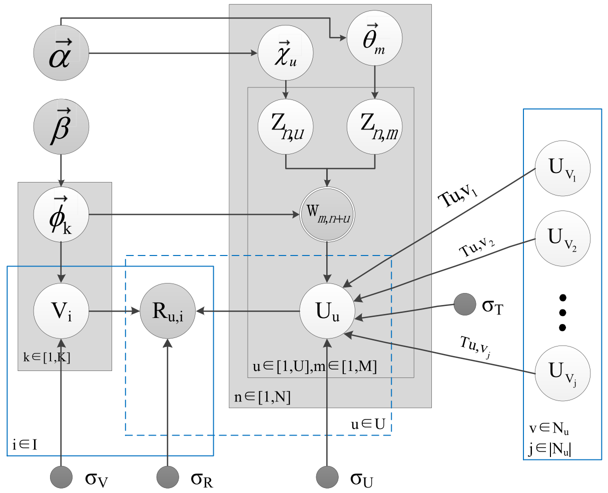

Fig. 7. Extension of graphical model with historical ratings.

Table 2
Experimental data.

Name	Users	Movies	Tags	Tag assignments	Ratings	Restricted movies
Count	2113	10,197	13,222	47,957	855,598	7155

Table 3
MCA iterative history.

MCA iterative history			
Number of iterations	The situation of considering variance		Loss
	Sum	Increment	
100[a]	994.915005	0.001770	0.084995

[a] Stop in the maximum number of iterations.

Table 4
Process summary

Process summary			
Dimensionality	Cronbach's Alpha	Remarks	
		Summary eigen value	iterance
1	1.000	994.927	1.000
1	1.000	994.927	1.000
Total		1989.830	2.000
average	100[a]	994.915	1.000

[a] Total Cronbach's Alpha Based on the average eigenvalue.

close. The results (Fig. 9) also seem similar in shape. The two figures display the multidimensional vectors that have been reduced into the two-dimensional vectors. The first identifying measurement is the mapping result of the two direct coordinates. We use the result of the second figure based on the description in the "Social tagging and its normalization and reconditioning" section. Several points must be removed, such as the tags of v799, v59, and v532. Normalizing and reconditioning the social tags provide a reasonable data set with considerable nominal categorical data. In general, approximately the same area of different variables on the same position classification points are linked to each other starting from (0,0). The scattering of variables that are associated with close distance between means is apparent in Fig. 8. The scattering of variables that are far from the origin is also apparent.

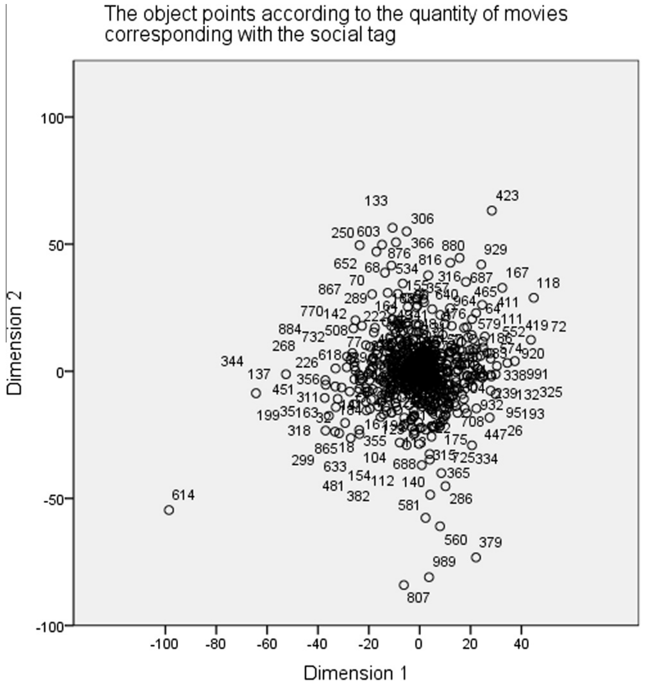

Fig. 8. Identification of measurement 1.

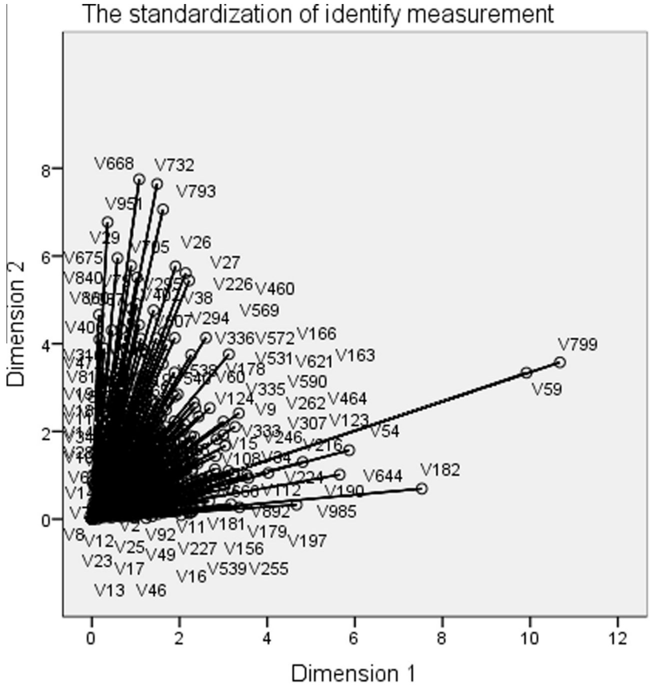

Fig. 9. Identification of measurement 2.

4.3. Tuning the parameters

To measure the quality of topics, the dataset is usually split into two parts: one for training, and the other for testing. For LDA, a test set is a collection of unseen documents w_d, and the model is described using topic matrix Φ and the hyper-parameter for the topic-distribution of documents. LDA parameters Θ are not considered because these parameters represent the topic-distributions for the documents of the training set. Thus, they can be ignored

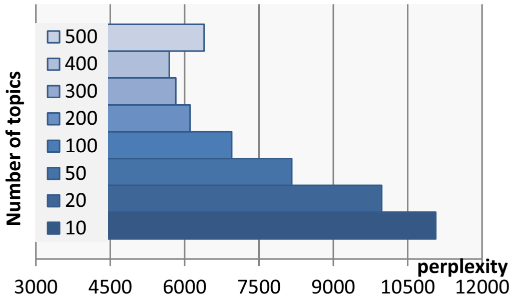

Fig. 10. Perplexity of the experiment corpus with different topic numbers.

in computing the likelihood of unseen documents. Therefore, we evaluate the log-likelihood as follows:

$$L(w) = \log p(w|\Phi, \alpha) = \sum_d \log p(w_d|\Phi, \alpha) \tag{27}$$

A set of unseen documents w_d are given topics Φ and hyperparameter α for the topic-distribution θ_d of documents. The likelihood of unseen documents can be used to compare models. The measure that is traditionally used for topic models is the perplexity of held-out documents w_d, which is defined as follows:

$$\text{perplexity (test set } w) = e^{-\frac{-L(w)}{\text{count of tokens}}} \tag{28}$$

High likelihood implies a good model. Minimum perplexity denotes the highest likelihood. When the number of topics are close to 400, the perplexity can derive the minimum value as shown in (Fig. 10). Therefore, we assume that the number of topics is equal to 350.

4.4. Experimental evaluation methodologies

We compare the proposed method and traditional CF methods that are based on the user and the item. The user-based CF contains the methods that calculate the similarity between Euclidean Distance and Pearson Correlation (abbreviated as UECF and UPCF, respectively). Similarity, the item-based CF can be classified into IECF and IPCF. In our experiments, we use the movie data from MovieLens. Each movie is rated using a discrete scale from 1 to 5. Each user can label each film with any tag, including words, phrases, or distinguished names. We use the table to support our evaluation metrics.

Precision and recall are extensively used in the recommendation system (Asabere et al. 2014) and statistical classification (Shaharanee and Hadzic 2014, 2013) fields. These ratios are used to evaluate the quality of the results. Precision is the ratio of the number of relevant items recommended and the total number of items. This ratio measures the precision of the recommendation system. Recall rate refers to the relevant items recommended and all related items in the item library. This rate measures the recall ratio of the recommendation system. With larger values for both ratios, the performance improved because both ratios are appropriate for the evaluation of whether the recommendation system provides good recommended content to users or not. The following equation presents precision and recall:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{29}$$

and

$$\text{Recall} = \frac{\text{TP}}{\text{Tp} + \text{FN}} \tag{30}$$

As previously mentioned, each movie has been rated using a discrete scale from 1 to 5. In our study, If the movie reaches a score 3, then that movie is what the user needs. We propose that higher precision and recall yield higher results. Nevertheless, both are contradictory in several cases. If we only recommend an item in extreme cases, which is accurate, then Precision is 100%, but recall is low. If we place all the returned items, then recall is 100%, but precision is low. Therefore, we should synthetically consider these components. The most common method is *F*-measure, as shown in the following equation.

$$F = \frac{(a^2 + 1)\ \text{Precision} * \text{Recall}}{a^2\ (\text{Precision} + \text{Recall})} \tag{31}$$

where we set the value of a to 1.

4.5. Experimental results and discussion

We perform comparative analysis to evaluate the recommendation performance of the proposed hybrid approach based on recall, precision, and *F*-measure (Hariri et al. 2012). In Figs. 11–13, UPCF and UECF represent the user-based CF with Pearson Correlation and Euclidean Distance. HR represents the proposed hybrid approach. HR1 represents the proposed hybrid approach without considering the rating. In addition, we also compare SVD++, CTR, PMF, and HOSVD. CTR is a topic-model-based recommendation model that incorporates the ratings and item contents. PMF is the basic MF method, which only considers ratings. HOSVD is a collaborative filtering algorithm that is a kind of tensor decomposition based on multidimensional characteristics of users. Kilmer and Martin (2011) mainly introduced HOSVD. HOSVD synthesis after the decomposition yield an approximate result of the original tensor. After the obtaining of approximate Tucker decomposition, we use the user-based CF to obtain the top-N recommend results on the aspect of ratings.

We used fivefold cross validation and determined that $K = 6, \lambda_u = \lambda_m = 0.01, a = 1$, and $b = 0.01$ result in the best performance for MF methods. a and b are tuning parameters $(a > b > 0)$ for the confidence parameter c_{ij}. More details are found in the work of Wang and Blei (2011). We also use fivefold cross validation to show that SVD and SVD++ deliver the best performance when $\alpha = 0.005$ and $\beta = 0.1$.

Through the analysis of the social tagging and related application, we trained the samples. Many negative samples exist in a large dataset. The existence of negative samples significantly reduces recommendation accuracy. Negative samples are the noises that affect the improvement of recommended indices. We improved recall by adopting the hybrid process in the model, as compared with the comparison the SVD++. This contribution is attributed partly to data preprocessing. CF based-on users has particular advantages in recall because of the oneness of using data set and the similarity between users. The recommendation is a comprehensive consideration. Therefore, a unilateral win does not yield good global results. Fig. 13 shows that our proposed method possessed fine stability and obtained more acceptable results.

The recommendation system presents a list of items for the user, but in general, most users are reluctant to peruse this list. Despite the increase in the size of the recommendation list, users only care about the recommendation results that at the beginning of the recommendation list. Therefore, with more target contents recommended from the whole dataset, better results are shown in a relatively few number of recommended candidates. For example, users can obtain more meaningful suggestion lists from 100 items with high precision than 200 items with low precision. Moreover, all kinds of effective indices of recommendation algorithms usually decline with the increase in the number of

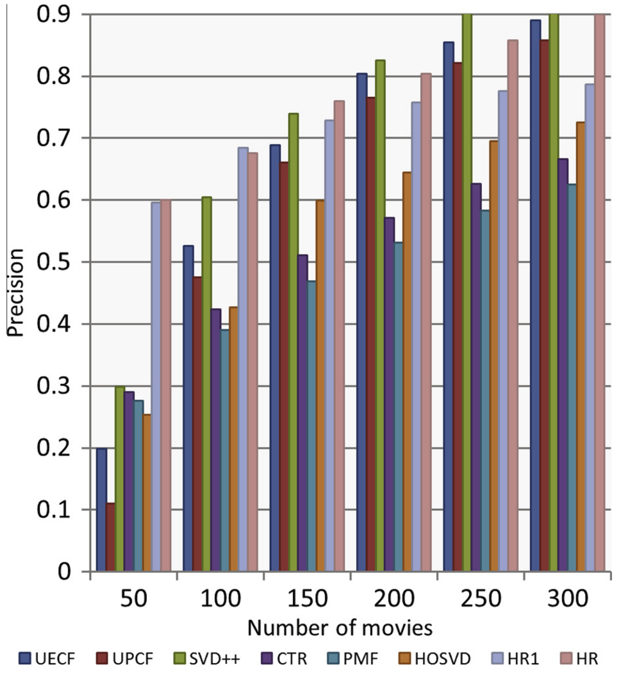

Fig. 11. Comparison of precision.

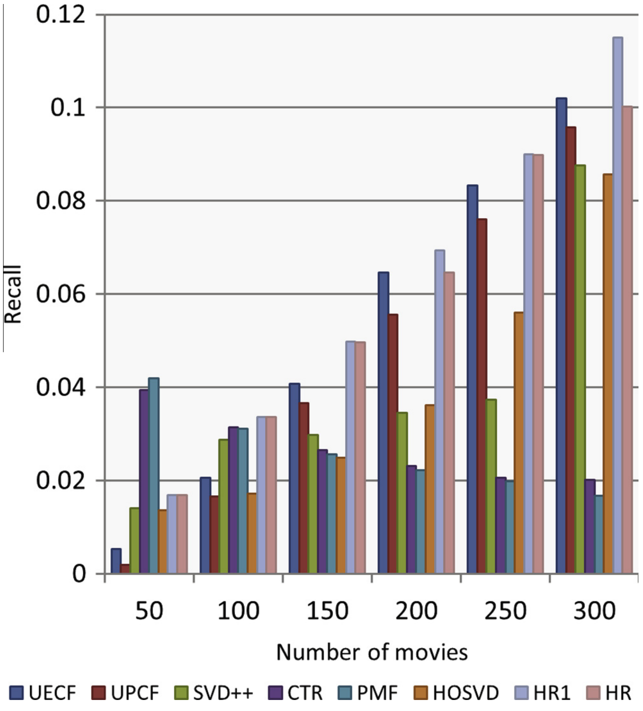

Fig. 12. Comparison of recall.

recommendations. Fig. 11 shows that, in the case of relatively few recommended items, our method is significantly better compared with the other method.

The aforementioned figure shows that the proposed method possesses high precision within 200 movies. SMNs often recommend excessive items to users. In addition, the proposed method is better than the user-based CF. The precision of the proposed method can be developed by combining user rating context information.

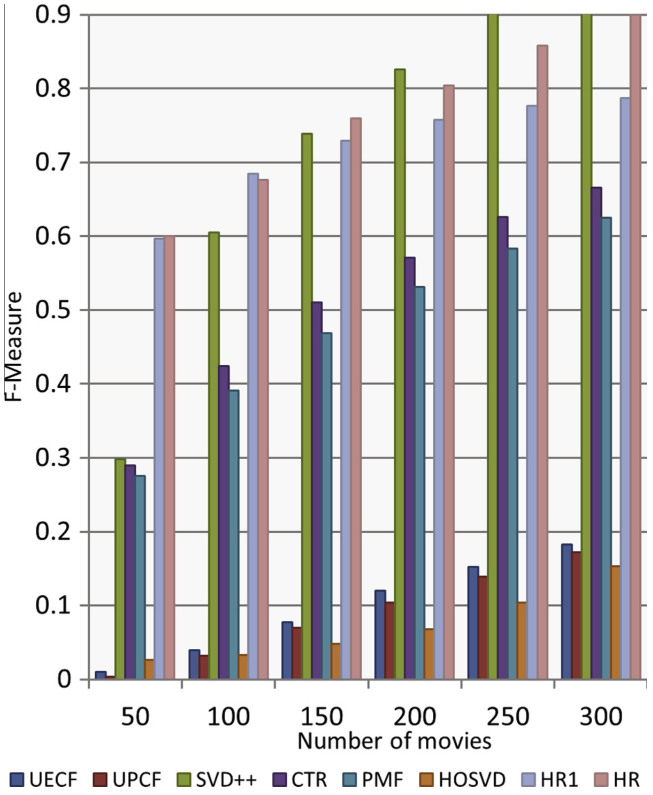

Fig. 13. Comparison of F-measure.

5. Conclusions and future works

Social media allow people to write, share, evaluate, comment, and communicate with each other through the web or mobile devices. Social media require a large number of viewers to contribute, extract, create consult, and spread information spontaneously. For example, in SMN, finding a preferred movie often requires frequent operations of users when they face vast resources, and personalized recommendation services can effectively solve the problem. However, the accuracy of a recommendation service is lower than expected. Thus, we propose a hybrid movie recommendation approach via social annotations and ratings. Based on user preference from social tagging and rating, we constructed SMNs and proposed the preference-topic model. Through a series of analyzes, including the extraction, normalization, and reconditioning of social tags, we combined the user historical rating to improve the hybrid model. The experimental results and the comparison of our model and existing some algorithms show that the proposed method has significantly improved in terms of recommendation accuracy.

In the future, we will enhance the performance of the algorithm and introduce the time factor into the recommendation to improve the responses to the changes in user focus. We will introduce the concept of social pulse (Pham et al. 2012, Jung 2014). For example, a user has long been interested in horror movies. Recently, he/she has focused considerable attention to comedy movies because of a certain actor. The system cannot recommend proper content to meet user requirements according to recent user preferences. In the personalized recommendation system, the types of user preferences lack dynamism, and user preferences cannot be acquired according to the change in time. Therefore, user preferences are constantly changing. In addition, future experiments will use different types of datasets (Feng and Wang 2012), including the Delicious and Last.fm.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 61379034), the National Key Technology R&D Program (No. 2014BAH28F05) and the Guangdong Province Science and Technology Program (No. 2013B040100004 and No. 2013B040403002).

References

Abdi, H., Valentin, D., 2007. Multiple correspondence analysis. Encyclopedia of Measurement and Statistics, 651–657.

Adomavicius, G., Tuzhilin, A., 2005. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17 (6), 734–749.

Asabere, N., Xia, F., Wang, W., Rodrigues, J., Basso, F., Ma, J., 2014. Improving smart conference participation through socially aware recommendation. IEEE Transactions on Human-Machine Systems 44 (5), 689–700.

Blei, D., Carin, L., Dunson, D., 2010. Probabilistic topic models. IEEE Signal Processing Magazine 27 (6), 55–65.

Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent dirichlet allocation. Journal of Machine Learning Research 3 (4–5), 993–1022.

Chelmis, C., Prasanna, V.K., 2013. Social link prediction in online social tagging systems. ACM Transactions on Information Systems (TOIS) 31 (4), 1–27.

Chen, C., Zeng, J., Zheng, X., Chen, D., 2013a. Recommender system based on social trust relationships. In: 2013 IEEE 10th International Conference on e-Business Engineering. IEEE, pp. 32–37.

Chen, C., Zheng, X., Wang, Y., Hong, F., Lin, Z., 2014a. Context-aware collaborative topic regression with social matrix factorization for recommender systems. In: Twenty-Eighth AAAI Conference on Artificial Intelligence. Twenty-Eighth AAAI Conference on Artificial Intelligence. pp. 9–15.

Chen, C., Zhu, Q., Lin, L., Shyu, M.-L., 2013b. Web media semantic concept retrieval via tag removal and model fusion. ACM Transactions on Intelligent Systems and Technology (TIST) 4 (4), 1–22.

Chen, J., Feng, S., Liu, J., 2014b. Topic sense induction from social tags based on non-negative matrix factorization. Information Sciences 280, 16–25.

Chen, X., Shin, H., 2013. Tag recommendation by machine learning with textual and social features. Journal of Intelligent Information Systems 40 (2), 261–282.

Choi, K., Yoo, D., Kim, G., Suh, Y., 2012. A hybrid online-product recommendation system: combining implicit rating-based collaborative filtering and sequential pattern analysis. Electronic Commerce Research and Applications 11 (4), 309–317.

Colace, F., De Santo, M., Greco, L., Moscato, V., Picariello, A., 2015. A collaborative user-centered framework for recommending items in online social networks. Computers in Human Behavior 51, 694–704.

Feng, W., Wang, J., 2012. Incorporating heterogeneous information for personalized tag recommendation in social tagging systems. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp. 1276–1284.

Guo, X., Zhang, R., Huai, J., Sun, H., Liu, X., 2013. Discovering user preference from folksonomy. In: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on. pp. 2114–2119.

Guy, I., Zwerdling, N., Ronen, I., Carmel, D., Uziel, E., 2010. Social media recommendation based on people and tags. In: Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval. ACM, pp. 194–201.

Hariri, N., Mobasher, B., Burke, R., 2012. Using social tags to infer context in hybrid music recommendation. In: Proceedings of the twelfth international workshop on web information and data management. ACM, pp. 41–48.

Hoi, S.C., Luo, J., Boll, S., Xu, D., Jin, R., King, I., 2011. Social Media Modeling and Computing. Springer Verlag, DE.

Hu, J., Wang, B., Liu, Y., Li, D.-Y., 2012. Personalized tag recommendation using social influence. Journal of Computer Science and Technology 27 (3), 527–540.

Hu, Y., Hu, Y., Koren, Y., Koren, Y., Volinsky, C., Volinsky, C., 2008. Collaborative filtering for implicit feedback datasets. In: 2008 Eighth IEEE International Conference on Data Mining. IEEE, pp. 263–272.

Jamali, M., Ester, M., 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In: Proceedings of the fourth ACM conference on recommender systems. ACM, pp. 135–142.

Jschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G., 2007. Tag recommendations in folksonomies. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). vol. 4702. Scopus, pp. 506–514.

Jung, J.J., 2014. Understanding information propagation on online social tagging systems: a case study on flickr. Quality & Quantity 48 (2), 745–754.

Karatzoglou, A., Amatriain, X., Baltrunas, L., Oliver, N., 2010. Multiverse recommendation: n-dimensional tensor factorization for context-aware collaborative filtering. In: Proceedings of the fourth ACM conference on Recommender systems. ACM, pp. 79–86.

Kilmer, M.E., Martin, C.D., 2011. Factorization strategies for third-order tensors. Linear Algebra and its Applications 435 (3), 641–658.

Koren, Y., 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In: Proceeding of the 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp. 426–434.

Lee, S.-Y.T., Phang, C.W.D., 2015. Leveraging social media for electronic commerce in asia: research areas and opportunities. Electronic Commerce Research and Applications 14 (3), 145–149.

Liu, B., 2007. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data. Springer, New York.

Liu, H., He, J., Wang, T., Song, W., Du, X., 2013a. Combining user preferences and user opinions for accurate recommendation. Electronic Commerce Research and Applications 12 (1), 14–23.

Liu, X., Datta, A., Rzadca, K., 2013b. Trust beyond reputation: a computational trust model based on stereotypes. Electronic Commerce Research and Applications 12 (1), 24–39.

Ma, H., Yang, H., Lyu, M., King, I., 2008. Sorec: social recommendation using probabilistic matrix factorization. In: Proceeding of the 17th ACM conference on information and knowledge management. ACM, pp. 931–940.

Mishne, G., 2006. Autotag: a collaborative approach to automated tag assignment for weblog posts. In: Proceedings of the 15th international conference on world wide web. ACM, pp. 953–954.

Mnih, A., Salakhutdinov, R., 2007. Probabilistic matrix factorization. In: Advances in neural information processing systems. Advances in neural information processing systems. pp. 1257–1264.

Movahedian, H., Khayyambashi, M., 2015. A semantic recommender system based on frequent tag pattern. Intelligent Data Analysis 19 (1), 109–126.

Musto, C., Narducci, F., De Gemmis, M., Lops, P., Semeraro, G., 2009. A tag recommender system exploiting user and community behavior. CEUR Workshop Proceedings 532, 25–32.

Naw, N., Hlaing, E.E., 2013. Relevant words extraction method for recommendation system. Bulletin of Electrical Engineering and Informatics 2 (3), 169–176.

Pham, X., Jung, J., Hwang, D., 2012. Beating social pulse: understanding information propagation via online social tagging systems. Journal of Universal Computer Science 18 (8), 1022–1031.

Purushotham, S., Liu, Y., Kuo, C.J.M., 2012. Collaborative topic regression with social matrix factorization for recommendation systems. Journal of Zhejiang University-Science 1, 759–766.

Ricci, F., Rokach, L., Shapira, B., Kantor, P.B., 2011. Recommender Systems Handbook. Springer Verlag, DE.

Shaharanee, I.N.M., Hadzic, F., 2014; 2013. Evaluation and optimization of frequent, closed and maximal association rule based classification. Statistics and Computing 24 (5), 821–843.

Shan, H., Banerjee, A., Banerjee, A., 2010. Generalized probabilistic matrix factorizations for collaborative filtering. In: Data Mining (ICDM), 2010 IEEE 10th International Conference on. IEEE, pp. 1025–1030.

Sigurbjrnsson, B., van Zwol, R., 2008. Flickr tag recommendation based on collective knowledge. In: Proceeding of the 17th international conference on world wide web. ACM, pp. 327–336.

Sun, F., Li, H., Zhao, Y., Wang, X., Wang, D., 2013. Towards tags ranking for social images. Neurocomputing 120, 434–440.

Thornton, K., McDonald, D., 2012. Tagging wikipedia: collaboratively creating a category system. In: Proceedings of the 17th ACM international conference on supporting group work. ACM, pp. 219–228.

Wang, C., Blei, D., 2011. Collaborative topic modeling for recommending scientific articles. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp. 448–456.

Wen, K., Li, R., Xia, J., Gu, X., 2014; 2012. Optimizing ranking method using social annotations based on language model. Artificial Intelligence Review 41 (1), 81–96.

Wetzker, R., Zimmermann, C., Bauckhage, C., Albayrak, S., 2010. I tag, you tag: translating tags for advanced user models. In: Proceedings of the third ACM international conference on web search and data mining. ACM, pp. 71–80.

Yanagimoto, H., Yoshioka, M., 2012. Relationship strength estimation for social media using folksonomy and network analysis. In: 2012 IEEE International Conference on Fuzzy Systems. IEEE, pp. 1–8.

Yao, J., Cui, B., Cong, G., Huang, Y., 2012. Evolutionary taxonomy construction from dynamic tag space. World Wide Web 15 (5), 581–602.

Zhang, C.-D., Wu, X., Shyu, M.-L., Peng, Q., 2013. A novel web video event mining framework with the integration of correlation and co-occurrence information. Journal of Computer Science and Technology 28 (5), 788–796.

Zhang, Z.-K., Zhou, T., Zhang, Y.-C., 2011. Tag-aware recommender systemsa state-of-the-art survey. Journal of computer science and technology 26 (5), 767–777.