Hindawi
Scientific Programming
Volume 2018, Article ID 8503452, 14 pages
https://doi.org/10.1155/2018/8503452

Research Article

Personalized Recommendations Based on Sentimental Interest Community Detection

Jianxing Zheng and Yanjie Wang

School of Computer and Information Technology, Shanxi University, Taiyuan, Shanxi 030006, China

Correspondence should be addressed to Jianxing Zheng; jxzheng@sxu.edu.cn

Received 4 April 2018; Accepted 19 June 2018; Published 5 August 2018

Academic Editor: Emiliano Tramontana

Communities have become a popular platform of mining interests for recommender systems. The semantics of topics reflect users' implicit interests. Sentiments on topics imply users' sentimental tendency. People with common sentiments can form resonant communities of interest. In this paper, a resonant sentimental interest community-based recommendation model is proposed to improve the accuracy performance of recommender systems. First, we learn the weighted semantics vector and sentiment vector to model semantic and sentimental user profiles. Then, by combining semantic and sentimental factors, resonance relationship is computed to evaluate the resonance relationship of users. Finally, based on resonance relationships, resonant community is detected to discover a resonance group to make personalized recommendations. Experimental results show that the proposed model is more effective in finding semantics-related sentimental interests than traditional methods.

1. Introduction

Recently, socialized recommendation has become one of the most popular means of recommendations in various recommender systems that have been applied in the fields of E-commerce, social media platforms, web search engines, and so on [1]. In socialized recommender systems, mining socialized relationships is critical for pushing or sharing interesting people and things with target users. Additionally, discovering similar users with common interests is important for determining recommendations. To address these problems, user profile techniques are used to reflect users' interests by describing socialized relationships and representing the history of browsing contents [2]. Accurate user profiles can depict the preference of users, which is helpful for accurate socialized recommendations.

Semantics analysis involves identifying relevant information from a serious of views, products and sources [3]. As a vector representation method, word embeddings are useful for identifying semantic relations from the context of documents or subjects [4]. Word embeddings use neural network architectures to implement distributed representations of words [4, 5]. The distributed representations in vector allow computation of the semantic similarity of words, which can efficiently discover high quality semantics-relevant subjects [6]. Therefore, we can model user profiles and analyze their semantic relationships regarding word embeddings to improve the quality of socialized recommendations.

Sentiment analysis, or opinion mining, has been applied to many fields, such as marketing, opinion monitoring, and information retrieval [7]. Sentiment analysis includes positive, negative, or neutral attitudes. Through analyzing semantic information of strong emotional and subjective texts, sentiment analysis captures users' emotional behavior characteristics and sentimental attitudes [8]. Sentiment user profiles can be searched for target users according to the presence and frequency of terms and opinionated words in documents [9]. A sentiment user profile depicts a user's sentiment degree for a subject or a topic [10]. Thus, users with similar sentiments or opinions may have a resonant interest tendency. Therefore, detecting a group of users with common sentiments provides a beneficial method for personalized resonant recommendations.

Considering the semantic and sentimental relationships between user profiles, we can divide users with similar sentimental interests into a cluster, forming a community. The target of community detection is to identify a series of clusters so that the users in the same cluster have high similarity but are very dissimilar with users in other clusters [11]. Based on this assumption, we can capture the interests of similar users in the same community for target users to improve accuracy of recommendation systems and user's satisfaction. However, very little research has been conducted merging semantics and sentiments for community detection. In fact, regarding content resources, although users browse the same topics, view the same products, and comment on the same topics, they often have different opinions and emotions for these topics. Some users support the topic while some users are against it. The results of community detection considering content resource do not reflect detailed interest opinions. Users in the same community are not emotionally congruent [12]. Therefore, identifying similar emotional interest users based on a community for socialized recommendations is a problem. By combining the semantic interests and sentimental interests, we can use resonant community to discover a group of similar users with common interest and sentiments, which can efficiently supplement a user's related interests to promote the accuracy of recommender systems.

In this paper, we present a resonant sentimental interest community- (RSIC-) based recommendation model to improve the accuracy performance of recommender systems. The proposed method considers the integration of semantics and sentiment. As a popular tool, word2vec proposed by Mikolov is used to learn distributed word representation to model the semantics vector and sentiment vector [10]. Then, considering semantics and sentiment factors, resonance relationship is computed to evaluate the correlation of users. Additionally, based on resonance relationships among users, a RSIC is detected to discover resonance group, which includes users with common interests and sentiments. Finally, based on resonance users, a collaborative strategy is adopted to select semantics-related subjects for updating the interest subjects of target users.

The remainder of this paper is organized as follows. Section 2 discusses state of the art about RSIC-based recommendation model. In Section 3, we give a RSIC-based recommendation overview. Section 4 introduces the notations and methods to build a semantic user profile and sentiment user profile. In Section 5, we apply the RSIC detection and collaborative filtering recommendation. Section 6 shows the experimental results and discussions. Section 7 contains the conclusions of the paper.

2. Related Works

Research on socialized recommendations and sentiment user profiles is relevant to the proposed model. In this section, we discuss related works involved in user profiles, community detection and recommendation approaches.

2.1. User Profiles. For most recommender systems, user profile modeling mines users' preferences for personalized search from users' histories or similar users' contents [1, 2, 13]. By extracting a vector representation of the words, Boratto [14] proposed a novel method to model user profile and detected segments of users. Through integrating annotated tags and ratings, Du [15] proposed a multilevel user profiling model to make personalized search. By analyzing the feedback interactions between users, Liu [16] presented an unsupervised approach to automatically update the profiles. Kumar [17] constructed clustered user-interest profile in terms of use singular value decomposition (SVD), which includes clusters of semantically or syntactically related tags, to identify topics of users' interests. By considering the entities and aspects in the user's comments, Meguebli [18] focused on building a method of user profiles and article profiles, and then matching these profiles for the purpose of personalized recommendation. By constructing user profiles from folksonomy systems, Xie [19] computed user similarities within a random walking distance and identified user communities. Based on the behaviors of community neighbors, enriching user profiles was proposed to solve the data sparsity problems of conventional single user profiling. Considering the contextual information sources, White [20] evaluated the utility of social, historic, task, collection, and user interaction sources to model user profiles for predicting users' future interests. Accounting for the sentiment factor, Xie [21] incorporated sentiment information and proposed a SenticRank framework to create a personalized search. In the framework, the authors utilized the content-based method and collaborative strategy to obtain personalized ranking recommendations.

As emotions are important in daily life, some studies have made emotion prediction for individuals. Considering a probabilistic graphical model, Cui [22] combined user-interest and social influence factors to make an emotion prediction framework for individuals. As bursty sentiment-aware topics can reveal sentiment-aware events, Qi [23] proposed a Time-User Sentiment/Topic Latent Dirichlet Allocation (TUS-LDA) to detect bursty sentiment-aware topics, which solved the problem of context sparsity problem in conventional LDA-based models. Although existing studies have got the remarkable achievements for personalized user profile modeling and individual emotion prediction, these works have not fully integrated the interest user profile and sentiment user profile to describe users' interests.

2.2. Community Detection. The existing community detection methods identify communities by utilizing the network topology information. Some researches focused on the graph clustering approaches to detect nonoverlapping or disjointed communities, such as block model approximation, label propagation, and modularity maximization. Considering the local importance of a node in a community and the representability of the community, Bai [24] proposed a fast graph clustering description model to discover communities for a large-scale network. By constructing a sparse graph regarding the similarity matrix, Xiao [25] aggregated multiple clustering results from different clustering algorithms to obtain the final clustering of different shapes and sizes. Chen [26] designed a Bayesian mixture network (BMN) model to make overlapping communities detection for weighted networks, which presented soft partition and soft memberships solutions to solve the problems of detecting weighted networks and measuring the membership degree of a node belonging to a community. As users' interests are changing over time, Feng [27] developed a time-weighted overlapping community detection method in terms of association rule mining in order to model dynamic user interests for personalized recommendations. Lancichinetti et al. presented a local fitness maximization method to make overlapping and hierarchical community detection [28]. Newman proposed the Fast-Newman algorithm to improve the quality of community detection [29]. Considering community detection as a matrix blocking problem, Chen [30] recognized matrix column similarities and computed a partial clustering of the vertices in a dense subgraph to analyze graph structures and complex networks.

To our knowledge, social emotions are contagious and resonant. Some researches have focused on exploiting social emotion mining from the latent semantics and sentiments of individual words. Rao [31–33] detected social emotion and investigated social emotion classification for short texts in terms of topic models, such as affective topic model and topic-level maximum entropy (TME) models. Rao et al. [33] also concentrated on generating sentiment topic model by merging latent topics with social emotions, which can be used in the social emotion classification and social emotion lexicons modeling. Lee [34] identified individual user sentiments embedded in the messages, task-oriented content, and proactiveness to analyze collective sentiments, which can affect collective cocreation thinking, especially for the innovation process of cocreation communities. Zou [35] utilized community detection methods to investigate how to exploit weak dependency connections in communities as an aspect of social contexts for microblog sentiment analysis, including sentiment consistency and emotional contagion. In our paper, we detect a resonant sentimental community and identify the most similar users with concordant sentiments for personalized recommendations.

2.3. Recommendation Approaches. Most recommender systems use three approaches to make recommendations, including content-based, collaborative filtering (CF), and hybrid approaches. Based on a personal ontology user profile, Cantador [36] investigated the tastes and preferences of users and computed their social relationships to identify communities of interest. Because short-form messages often express users' interests and opinions, Esparza [37] investigated users and products profiles from associated reviews and computed their relevance to make product recommendations. Taking into account the similarities among user profiles, cooccurrence of user names, and interaction behaviors, Xiong [38] presented a probabilistic graphical model to accurately measure the social relationships in online social networks for recommendations. By exploiting the high-order relational information of tag data and customizing different types of relations' influences, Zhu [39] developed a heterogeneous hypergraph embedding framework for document recommendation. Based on users' historical web search behaviors, Bai [40] utilized external usage information to the news service for news personalization. By considering structured and unstructured data with different semantics, Zhang [41] proposed an integrated framework, called collaborative knowledge base embedding (CKE), to learn the implicit representations and semantic representations in terms of collaborative filtering strategy. Based on the semantics of tags, Li [42] categorized tags into emotional types and developed an emotion ontology called UniEmotion for music recommendation. Although the content information has been successfully used for various recommender systems, sentimental information is also worthy of attention for capturing a personal user profile and improving the accuracy of document recommendations.

3. Recommendation Framework Based on the RSIC

We design a recommendation framework based on the resonant community. The resonant community is comprised of users with similar interests and common sentiments. Figure 1 shows the process of RSIC detection and resonant interest selection.

In Figure 1, the recommendation framework involves three steps: weighted user profile modeling, resonance relationship calculation, and RSIC-based recommendation. First, by applying the term frequency-inverse document frequency (TF-IDF) mechanism and ontology structure, we compute the interest degree of subject for each user. Considering word embeddings and sentiment dictionary, a weighted semantic user profile and weighted sentiment user profile are modeled. Then, based on two kinds of weighted user profiles, we compute the resonance relationship between users, which differentiates the interests and emotions among users. Finally, we conduct RSIC detection by considering resonance relationships between users. According to the belonged community, resonance group is selected for target users to implement an interest update. By ranking updated interests, we push top-k subjects and their relevant microblogs to target users.

4. Resonance Relationship

4.1. Content Interest. Messages posted or reposted by users contain many noun entities which reflect preferences of users. In our paper, we adopt the TF-IDF mechanism to measure the weight of a subject in messages. First, given a corpus, by removing stop words and splitting words, we identify significant noun entities in messages and calculate their TF-IDF weight. For a message m, it can be represented as $m = \{(w_1, weit_{1m}), (w_2, weit_{2m}), \ldots, (w_p, weit_{pm})\}$.

Here, $weit_{tm}$ is the relative importance of term t in m, which can be computed using the TF-IDF scheme as follows:

$$weit_{pm} = \frac{freq_{pm}}{\max_l (freq_{lm})} \times \log \frac{N_m}{n_p}, \quad (1)$$

where $freq_{pm}$ is the frequency number of term p in microblog m and $\max_l(freq_{lm})$ is the frequency number of term l which has the maximum frequency in m. N_m is the total amount of microblogs and n_p is the quantity of microblogs that contain term p. The weight $weit_{tm}$ depicts

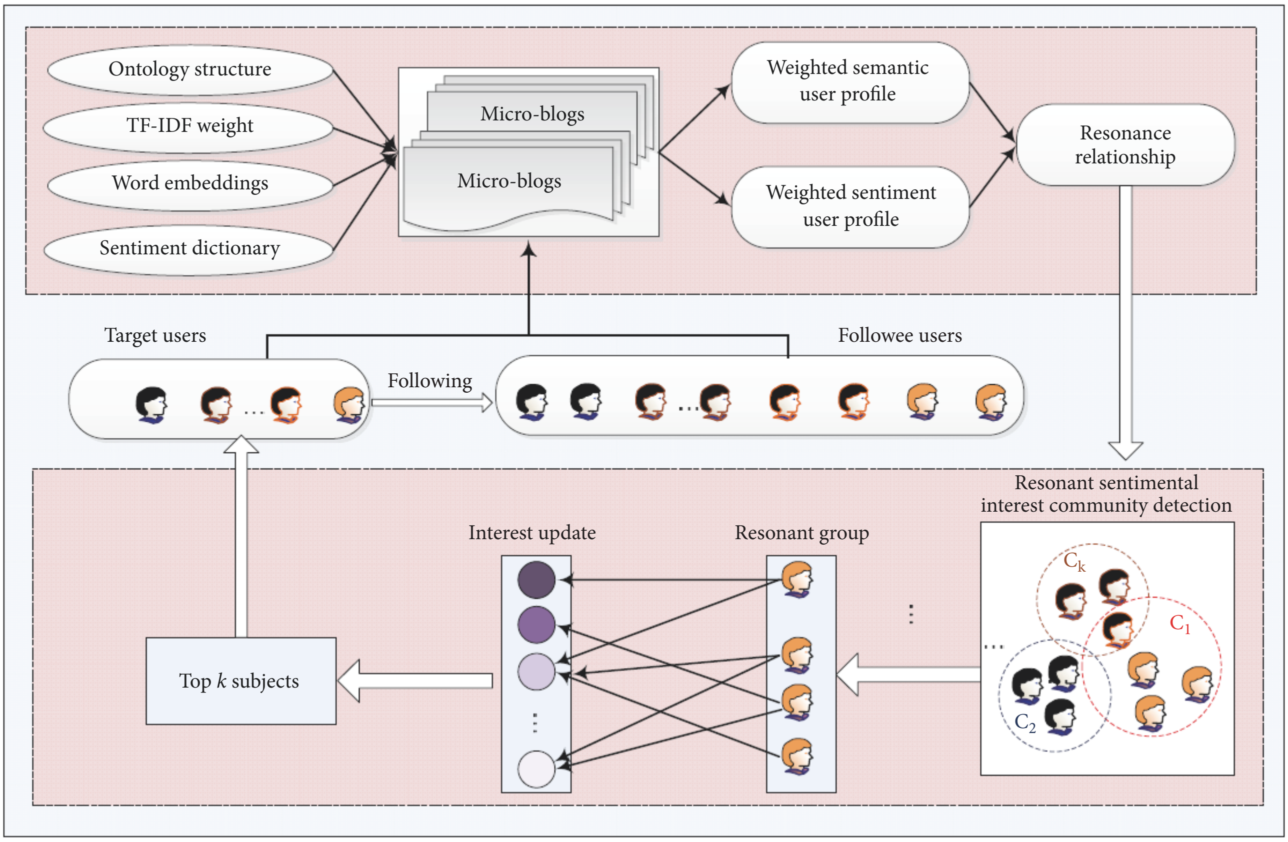

FIGURE 1: RSIC-based recommendation framework.

the importance contribution of term t on the representation of the microblog m.

Then, for a given user u, considering all the user's microblogs, we infer content interest degree of subject s as

$$Cid_u(s) = \frac{\sum_{m \in M_u} weit_{sm} \times \eta(s,m)}{\sum_{s_i \in S} \sum_{m \in M_u} weit_{s_i m} \times \eta(s_i,m)}, \quad (2)$$

where M_u is the set of microblogs for user u. S is a subject set over knowledge base C. The knowledge base C is involved in society, sports, economics, culture, and IT topics, which are from the category classifications of the Baidu Wikipedia. Figure 2 shows an example of a classification structure for the five topics. $\eta(s,m) = 1$ if $s \in m$; otherwise $\eta(s,m) = 0$.

4.2. Weighted Semantic User Profile. However, the interest degree does not reflect close relationships between subjects from semantics, which is unsuitable for discovering similar users with latent semantic interests. Therefore, we introduce word embeddings to represent the characteristic of subjects and improve descriptions of user profiles.

The neural word embeddings proposed by Google's word2vec include the CBOW and n-gram models [6]. Word2vec adopts $|k|$ dimensional vector autoencoders to train a large quantity of text for representing the characteristic of words or contexts of words [9]. Figures 3-4 show an illustration of the CBOW and n-gram models, respectively. The CBOW model is a three-layer neural network to predict a word as the output of a vector regarding the context as input, while the n-gram model learns the vector representation of the context through the center word [9].

In our work, we try to learn the word vector using the n-gram model, which extracts the multidimensional vector representation of a word to represent the characteristic of a subject. Given a sequence of training noun entities $w_1, w_2, \ldots, w_T$, for a word w_j, its context includes the previous c words and following c words as $C(w_j) = \{w_{j-c}, w_{j-c+1}, \ldots, w_{j+c-1}, w_{j+c}\}$. Then, the n-gram model maximizes the conditional probability $P(C(w_j) \mid w_j)$ to learn the word vector representation of w_j. By maximizing the words' average log probability, the n-gram model learns the objective function as follows [10]:

$$L = \frac{1}{T}\sum_{j=1}^{T} \sum_{-c \le i \le c, i \ne 0} \log P\left(w_{j+i} \mid w_j\right), \quad (3)$$

where c is the size of the training context. For the n-gram model, we use a hierarchical softmax function to speed up training. Finally, we adopt a weighted path from the root to a leaf node to represent the vector of word as $= \{d_1 : \tau_1, d_2 : \tau_2, \ldots, d_k : \tau_k\}$.

Then, for an interest subject w_p, assigning its content interest weight to the vector representation, we utilize an integrated weighted vector to differentiate semantics and

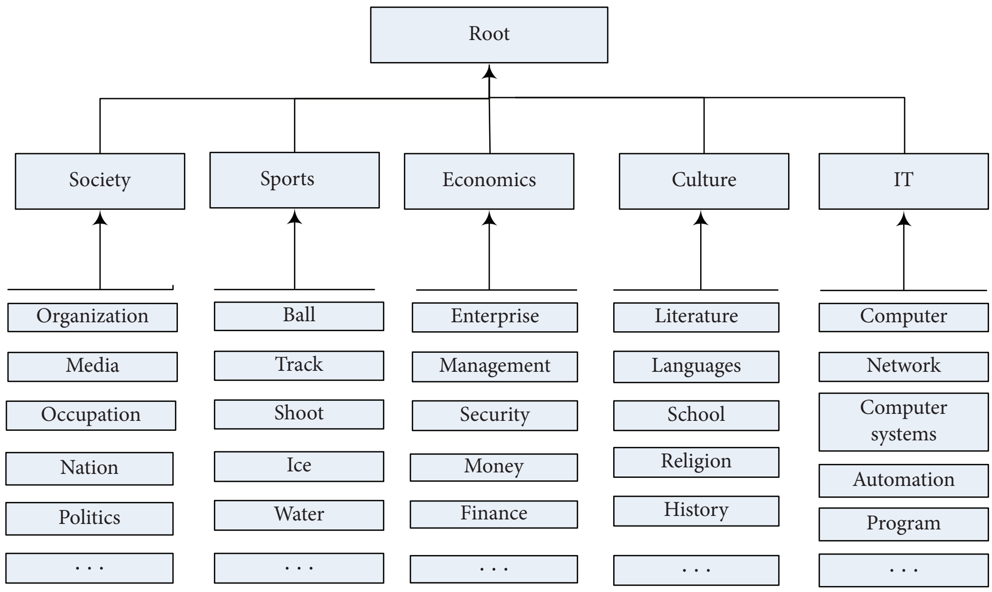

FIGURE 2: A fragment of a knowledge base for the five topics.

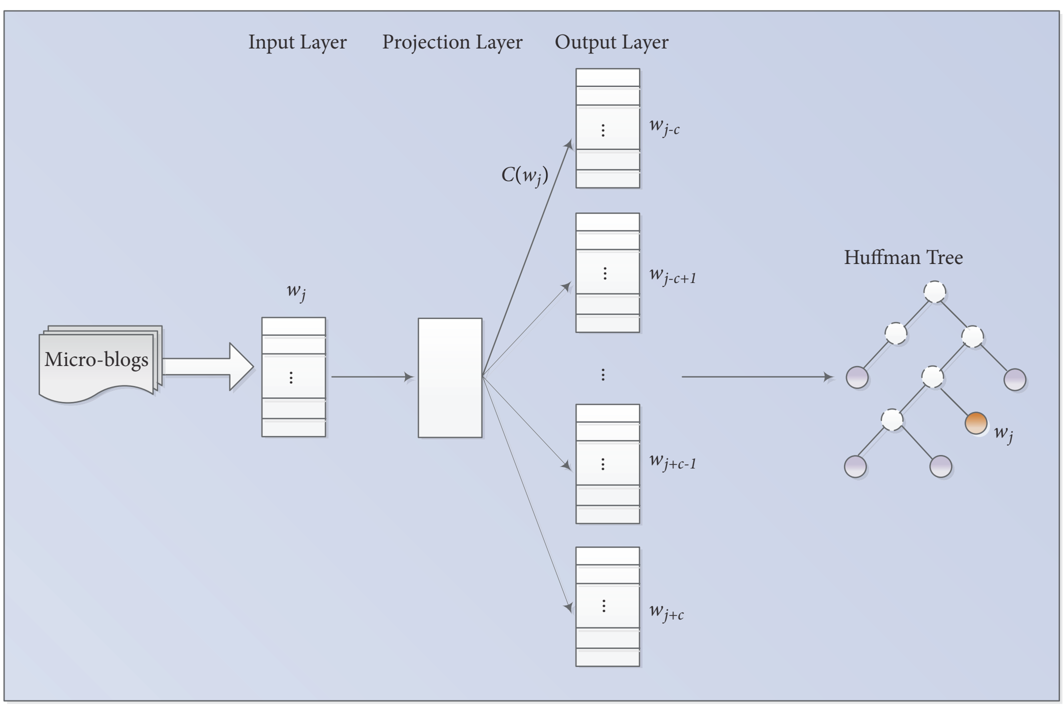

FIGURE 3: An illustration of CBOW.

model the semantic user profile for a user, which is formalized below:

$$\vec{V}(u) = \frac{\sum_{p=0}^{n} Cid_p(u)\,\vec{v}_p}{\sum_{p=0}^{n} Cid_p(u)}. \tag{4}$$

The weighted semantic user profile can be efficiently used to depict the closeness of users from aspects of latent integrated semantics. For two users, u_i, u_j, we can measure their semantic similarity by cosine metric as

$$sim_{ij}^{1} = \frac{\vec{V}(u_i) \bullet \vec{V}(u_j)}{\left\|\vec{V}(u_i)\right\| \cdot \left\|\vec{V}(u_j)\right\|}. \tag{5}$$

4.3. Weighted Sentimental User Profile. In the microblog scenario, documents are short and discrete, which provides

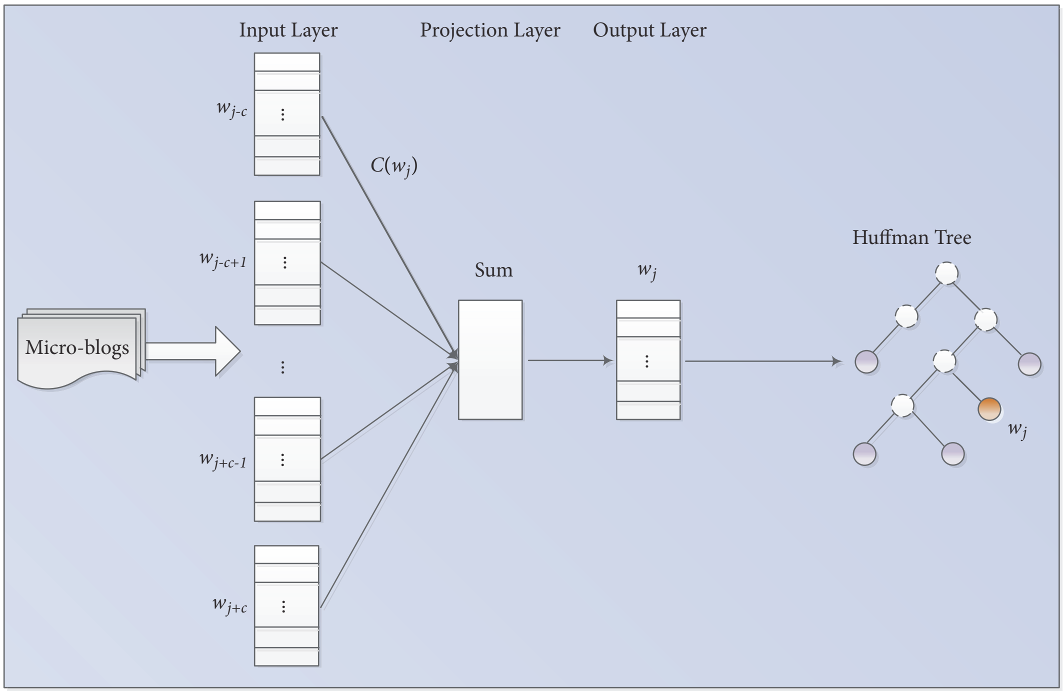

FIGURE 4: An illustration of n-gram.

users with fragmented information. One topic can trigger multiple messages, which includes positive comments, negative judgements, and neutral points. The subjects in each message express a kind of sentiment or emotion. Different subjects imply different sentiment tendencies. In our paper, we use the emotional vocabulary ontology [43] to detect the sentiment degree of each subject. Considering occurrence frequency of a subject in one's microblogs, we define the sentiment degree weight as

$$W_u(p) = \frac{\sum_{m \in M_u} sed(p) \times \lambda(p, m)}{\sum_{p_i \in O} \sum_{m \in M_u} sed(p_i) \times \lambda(p_i, m)}. \quad (6)$$

Here, *sed* is a sentiment degree of a subject, which involves several grades, such as scores 1, 3, 5, 7, and 9. $\lambda(p, m) = 1$ if $p \in m$; otherwise $\lambda(p, m) = 0$. The sentiment degree weight describes the sentimental importance of a subject for a user's sentiments.

Additionally, for a sentiment subject p, we learn its sentiment vector representation as v_p^s. Considering all the sentimental entities of user u, we utilize a weighted sentiment vector to model sentiment user profile as

$$\overrightarrow{V^s}(u) = \frac{\sum_{p=0}^{n'} W_u(p) \times \overrightarrow{v_p^s}}{\sum_{p=0}^{n'} W_u(p)}. \quad (7)$$

The weighted sentiment user profile describes a user's sentiment preference from aspects of sentimental semantics, which shows the emotional subjects the user prefers. Based on weighted sentimental user profiles, we infer users' sentimental similarity as

$$sim_{ij}^2 = \frac{\overrightarrow{V^s}(u_i) \bullet \overrightarrow{V^s}(u_j)}{\left\| \overrightarrow{V^s}(u_i) \right\| \cdot \left\| \overrightarrow{V^s}(u_j) \right\|}. \quad (8)$$

For users u_i and u_j, based on semantic similarity and sentimental similarity, we compute their weighted sum to model users' resonance relationship, shown in

$$sim_{ij} = \alpha sim_{ij}^1 + (1 - \alpha) sim_{ij}^2. \quad (9)$$

The coefficient weight α evaluates the relative importance of the semantic similarity and sentimental similarity on the measurement of the resonance relationship.

5. RSIC-Based Recommendation

Based on the resonance relationship between users, we determine close connected edges and implement community detection. The generated communities include users with common interests and similar emotions, named as resonant sentimental interest community. Considering the RSIC, we implement more accurate personalized recommendations. For all users, we first construct the resonance graph $G(V, E)$ based on their resonance relationships, where V is the set of user vertices, E is the set of resonance edges which represent two users having a higher similarity. Equation

(10) defines the resonance edges using the corresponding resonance relationship between users:

$$e\left(u_i,u_j\right)=\begin{cases}1 & if\, sim_{ij}>\varepsilon\\ 0 & else\end{cases} \tag{10}$$

where cutoff $\varepsilon \geq 0$ controls the number of resonance edges in the graph. Different thresholds generate different numbers of connection edges for a resonance relationship graph. In (10), we ensure that those connective nodes using the resonance edges in the graph are users with similar interests and common sentiments. The resonance edges are used to cluster resonant users into a community.

Based on the resonance graph $G(V,E)$, we implement RSIC detection. As stated in [28], the fitness function measures the contribution of internal edges of nodes in the graph and external edges with other nodes in the remainder of the graph, which is shown as follows:

$$f(G)=\frac{d_{in}^{G}}{\left(d_{in}^{G}+d_{out}^{G}\right)^{\beta}}, \tag{11}$$

where d_{in}^{G} and d_{out}^{G} are the total internal and external degrees of the nodes of graph G. Parameter $\beta > 0$ determines the scale of communities. In our paper, we compute the weighted internal and external degrees of graph G as $d_{in}^{G}=\sum_{i\in G}\sum_{j\in G} sim_{ij}\cdot e(u_i,u_j)$ and $d_{out}^{G}=\sum_{i\in G}\sum_{j\in \overline{G}} sim_{ij}\cdot e(u_i,u_j)$. Then, we set $\beta = 1$ to have iterative operations to detect the overlapping communities. The weighted internal and external degrees reflect the close relationship between users from aspects of semantics and sentiments. For a resonance graph G and a node p, we utilize the fitness contribution $f_G(p)=f_{G+\{p\}}-f_{G-\{p\}}$ to determine whether the node belongs to the community G. $f_{G+\{p\}}^{\rho}$ and $f_{G-\{p\}}^{\rho}$ define the fitness of the new graph G with p inside and outside.

Based on the fitness, the detailed steps of RSIC detection are given in Algorithm 1.

Algorithm 1 describes the detection process of a community for a user node u. For all the vertex nodes in the graph, we detect communities until each node is contained in at least one community. Algorithm 2 presents the process of community detection.

Algorithm 2 shows the process of RSIC detection for all vertexes in the initialized resonance graph G. At each iteration, steps (3)-(4) perform the community detection for a new node, which is selected from the last subgraph G. By implementing Algorithms 1 and 2, the RSIC of each vertex is discovered, and each vertex belongs to one community.

In each RSIC, the nodes in the community are similar regarding the subject semantics and sentiment attitudes. Based on this view, for each user, we utilize the RSIC to discover one's resonance group, as $crg(u)=\{u_j \mid e(u,u_j)=1, u_j \in G_i\}$. The resonance group includes users who have close linkages with the target user. Then, considering the subjects deriving from the resonance group, we collaboratively predict the interest degree for the target users. Given a subject p, the interest degree of the subject based on the resonance group is computed as

$$Cid_p^R(u)=\frac{\sum_{u_j\in crg(u)} Cid_p(u)\times sim\left(u,u_j\right)}{\sum_{u_j\in crg(u)} sim\left(u,u_j\right)}. \tag{12}$$

By ranking the interest degree, the top-k subjects are selected for helping to provide related microblogs to target users.

6. Experiments and Discussions

6.1. Experiment Strategies. In this section, we evaluate performance of the proposed RSIC-based method by introducing some other recommendation strategies to make comparisons, including LDA, CF, and TF-IDF methods.

For a given subject s, we can get content interest degree of a user by TF-IDF mechanism in (2). Considering all the subjects, we can rank their content interest degree and provide top-k subjects for target users to make recommendations.

Based on the semantic similarity in (5), we can compute the similarity of two users to measure their interest closeness. For a given user u_i, considering one's similar users set N_{u_i}, the collaborative interest degree is defined as in

$$Cfid_{u_i}(s)=\frac{\sum_{u_k\in N_{u_i}} Cid_{u_k}(s)\times sim_{ik}^{1}}{\sum_{u_k\in N_{u_i}} \left|sim_{ik}^{1}\right|}. \tag{13}$$

By ranking the collaborative interest degree of subjects, we can select top-k subjects and push to target users.

LDA topic method is a generative probabilistic graphical model for personalized topic models [23]. The model generates documents of latent topics in terms of two assumptions. Each document can be represented as a multinomial distribution over a set of T topics, and the topic is a multinomial distribution related to the set of vocabulary words, which are, respectively, defined as $p(w \mid z)$, $p(z \mid d)$, where z, w, and d denote the latent topic, the word, and the document, respectively. The topic $z_i = j$ can be denoted as $P(z_i = j) = Multinomial(\theta_j^{(d_i)})$. A multinomial distribution related to the set of vocabulary words can be denoted as $P(w_i \mid z_i = j) = Multinomial(\varphi_{w_i}^{(j)})$, which depicts the meaning of the topic. Then, the document distribution and word distribution are Dirichlet distributions, which can be defined as $\theta_j = Dirichlet(\alpha)$ and $\varphi_i = Dirichlet(\beta)$.

In experiments, we adopt Gibbs sampling and set the hyperparameters $\alpha = \beta = 0.01$ to model latent topic distributions. According to top-k topic distribution, we select their maximal related subjects for each topic and give recommendation to the user.

6.2. Experiment Datasets. To examine the quality of the proposed RSIC-based recommendation method, we used the *NLPIR* dataset and *Application* dataset to verify the efficacy of the system. The two datasets contained subjective emotional Weibo microblogs or users' emotional comments, involved in lots of words in the emotional vocabulary ontology proposed in [43]. For both of the datasets, we rely on

Input:
node u.
Output:
community G.
(1) A loop is performed over all adjacent nodes of u;
(2) Select the adjacent vertex u', where $f_G^{u'} = \max\{f_G^{u_k} \mid e(u, u_k) = 1\}$, generating a subgraph G;
(3) Calculate the fitness of each vertex of G;
(4) **if** $\exists p \in G$,satisfy $f_G(p) < 0$ **then**
(5) Delete p, yielding a new subgraph G';
(6) **end if**
(7) **if** 4 occurs **then**
(8) Repeat from (3).
(9) **else**
(10) Repeat from (1) for subgraph G'.
(11) **end if**

Algorithm 1: Community detection.

Input:
graph $G(V, E, w)$.
Output:
communities C.
(1) **while** $V \neq \varnothing$ **do**
(2) Select the node u_i from V having the maximal resonance edges with other nodes in G.
(3) Detect the community G_i of u_i by Algorithm 1;
(4) $S = S \cup G_i$;
(5) $V = V/S$;
(6) Generate the subgraph G by the remaining nodes in V;
(7) Add the set G_i into the C;
(8) **end while**

Algorithm 2: RSIC discovery.

the timestamps to split users' microblogs set into two parts. The data in the earlier period was explored to model semantic and sentimental user profiles. The latter period was used for recommendation tests.

The *NLPIR* dataset was derived from the NLPIR website (http://www.nlpir.org/), which was collected from Sina Weibo. The dataset was from December 4, 2011, to December 23, 2011, and 114 users with more than 4,337 training microblogs were used to learn their user profiles. In addition, for 114 users, we selected their 1,228 followees and followees' 1,873 microblogs to model followees' user profiles. Considering all user profiles, we compute their resonance relationships and made a RSIC-based recommendation. Finally, we adopted 114 users' 5,065 testing microblogs to test the accuracy of the RSIC-based method. In the dataset, the number of followees' microblogs is small.

For the *Application* dataset, we selected 3,449 users to crawl the Sina Weibo (http://open.weibo.com) and get their microblogs, which were from April 10, 2013, to April 29, 2013. By deleting the sentences with fewer than two characters, we preserved 26,293 microblogs to conduct the experiments. In particular, 7,279 training microblogs and 9,986 followees' microblogs contributed to user profile modeling; meanwhile, 9,028 testing microblogs were used for verifying performance. In the *Application* dataset, the number of 9,986 followees' microblogs is large enough to get abundant subjects to model semantic user profiles. The details of the two datasets are shown in Table 1.

Table 1: Statistics results for two datasets.

	NLPIR dataset	***Application dataset***
No. of users	114	435
No. of followees	1,228	3,014
No. of users' training micro-blogs	4,337	7,279
No. of followees' micro-blogs	1,873	9,986
No. of users' testing microblogs	5,065	9,028

6.3. Experiment Metrics. In our experiments, we used precision, recall, and F1 measure to evaluate the performance of various recommendation methods, which were calculated as follows:

$$Precision = \frac{|S_T \cap S_R|}{|S_R|}, \quad (14)$$

$$Recall = \frac{|S_T \cap S_R|}{|S_T|}, \quad (15)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}, \quad (16)$$

where $S_T = \{s \mid I(s) > 0\}$ is the set of real subjects which were involved in the test microblogs and S_R is the set of subjects from the recommendation list.

6.4. Experiment Results. In the experiments, we first model weighted semantic user profile and sentimental user profile. Then, by computing the resonance relationship of users, we detect the RSIC community and select the resonance group of users from the community. Based on the interest degree in (12), we can rank the subjects in community and make top-k recommendations.

In the process of calculating the resonance relationships, we set the dimension of the word vector as 100 to depict the semantic user profile and sentimental user profile. After computing the interest degree of the subject in terms of resonance group, we changed the value of recommendation list size k to make personalized recommendations. As the resonance relationship between users was determined by both semantic similarity and sentiment similarity, the results of the RSIC were affected by the relative weight of the semantic similarity and sentiment similarity. By changing the value of the relative weight coefficient α, we considered the resonance relationship to observe the variations in the recommendation results. Meanwhile, the cutoff ε also controlled the scale of resonance edges among users, which formed a different resonance relationship graph and affected the results of the communities. Figures 5-6 show the results of precision, recall, and F1 under different recommendation list sizes by setting the fixed coefficient α and cutoff ε. As shown in the figures, when k increases, the performance of the RSIC recommendation method is clearly superior to LDA, CF, and TF-IDF methods. For example, in the *NLPIR* dataset, we can see that the precision of the RSIC method with a recommendation list size of 5 can get an approximately value of 0.67 while the values of LDA, CF, and TF-IDF are 0.53, 0.58, and 0.51. That is, in the RSIC method, more than 381 of 570 recommended subjects are matched with the target users' interests. In the microblog scenario, for a topic, different users often describe different semantics regarding different diverse text contents. However, their similar topic sentiment can assist in identifying the same interests for users. By considering the sentiment effects, more accurate subjects can be selected for the target users.

Additionally, for all methods in Figures 5-6, we can also observe that their precision values decrease and recall values increase with the recommendation list size increasing from 5 to 25. Especially for a larger recommendation list size, the recall of the RSIC method is superior to other methods, which shows that the introduction of sentiment can efficiently enhance the relevance of subjects.

By adjusting the values of coefficient α, we can analyze the influence of semantic similarity and sentiment similarity factors for the recommendation results, which are shown in Figures 7-8. By setting the values of the cutoff as $\varepsilon = 0.5$ and $\varepsilon = 0.6$ for two datasets, Figures 7-8 show the trend of precision, recall, and F1 results at different recommendation list sizes under different coefficient α. In the figures, we can see that both the precision and recall curves first increase and then decrease for different recommendation lists. On two datasets, the performance can get a better result at $\alpha = 0.7$ and $\alpha = 0.5$ for different values of k. For the *NLPIR* dataset, the few microblogs can trigger a small semantic similarity between users; and the sentiment factor can obviously affect the resonance relationship of users, which effectively improve the performance of recommendation results. For the *Application* dataset, the adequate microblog contents can reflect the users' semantic similarity, which is beneficial for selecting interest subjects. Thus, in the second dataset, the values of precision and recall are maximally at weighted coefficient $\alpha = 0.5$, which shows that both the semantics factor and sentiment factor have the same significance in the process of discovering resonance users for resonant community detection. The phenomenon shows that an appropriate weight coefficient can achieve good results for RSIC-based recommendations.

As the cutoff ε can affect the scale of the detected community, we set different cutoffs to identify different resonant communities. Based on different resonant communities, we obtain different resonance groups to make the RSIC recommendations. For two datasets, Figures 9-10 show the precision, recall, and F1 results at different recommendation lists under different cutoff ε values with $\alpha = 0.7$ and $\alpha = 0.5$. In the figures, we can see that the best performances are achieved at $\varepsilon = 0.5$ and $\varepsilon = 0.6$, respectively. For example, in the *NLPIR* dataset, the values of precision and recall first rise and then decline with the cutoff ε changing. With the recommendation list size increasing from 5 to 25, the maximal precision and recall become 0.67, 0.48, 0.37, 0.31, and 0.26 and 0.44, 0.62, 0.73, 0.79, and 0.84 at $\varepsilon = 0.5$, respectively. As we expected, in (10), the large cutoff is inappropriate for acquiring resonance graph with similar users, which creates a small community and few resonance users. A small number of resonance users cannot provide rich interest subjects for the target user, which affects the performance of the RSIC method. However, a small cutoff helps to model many similar resonance relationships; and many users in the RSIC generate a large resonance group. Although the sentiment of the users in the resonance group is close and consistent, their semantic similarity is small. The subjects from the resonance group are not accurate and relevant, which leads to a poor precision and recall. Therefore, it is suitable to set an appropriate threshold for detecting community and filtering out the best number of resonance users to make recommendations.

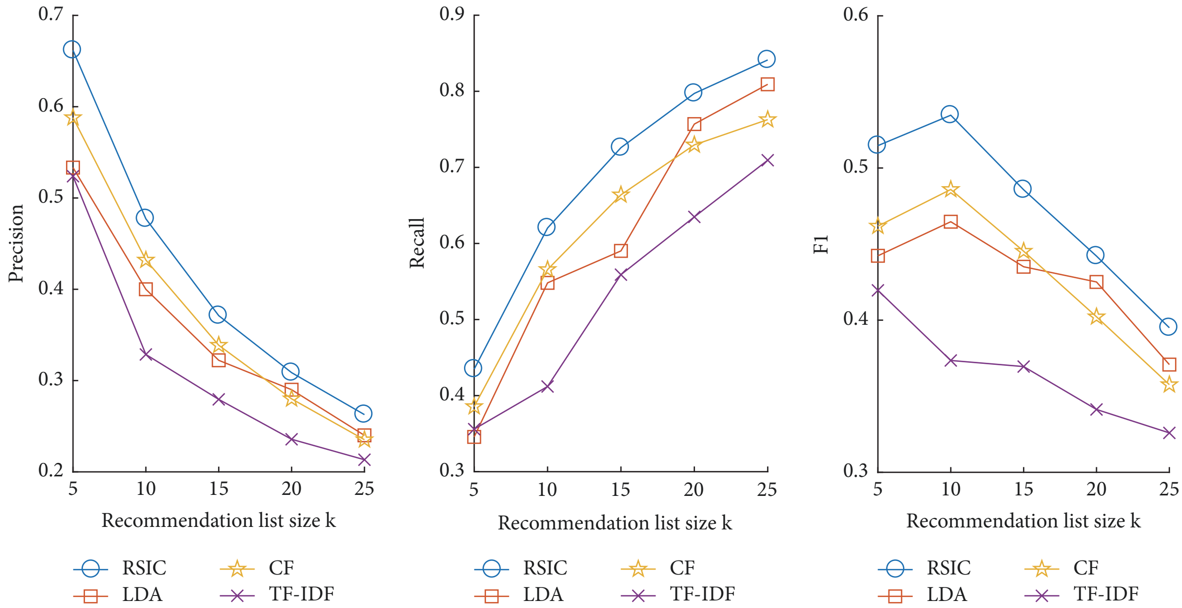

FIGURE 5: *NLPIR* dataset: precision, recall, and F1 values for various recommendation methods under different k ($\alpha = 0.7$; $\varepsilon = 0.5$).

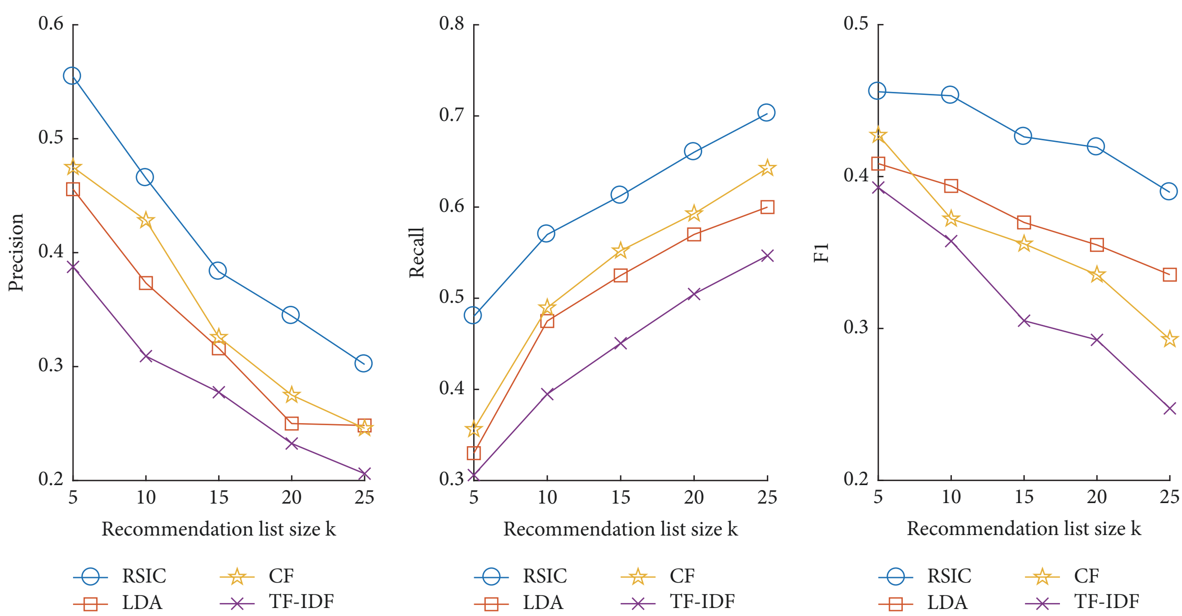

FIGURE 6: *Applications* dataset: precision, recall, and F1 values for various recommendation methods under different k ($\alpha = 0.5$; $\varepsilon = 0.6$).

7. Conclusions

This paper proposed a method to merge the sentiment factor into user's semantic interests for computing resonance relationships between users. Then, considering the resonance relationship, a resonant sentimental interest community was detected for personalized recommendations. From evaluation of the method, some insights were found, as described below.

In our experiment, the RSIC recommendation method outperformed semantics-based LDA, CF, and TF-IDF methods, both in indexes of accuracy and recall. We contributed to designing a weighted RSIC-based recommendation method by taking into account the sentiment and semantics factors for community detection. In addition, both sentiment and semantics were beneficial for mining resonance users with common interests in a community. Especially for uses with high implicit semantic similarity, the sentiment can efficiently select their accurate and relevant subjects.

Interestingly, users' interests are diverse and multigranular. How to model the sentiment user profile in different grain subjects and discover multigrain RSIC is a promising problem. In most cases, the resonance similarity of users is large for the coarse-grain subject, while, in the fine-grain

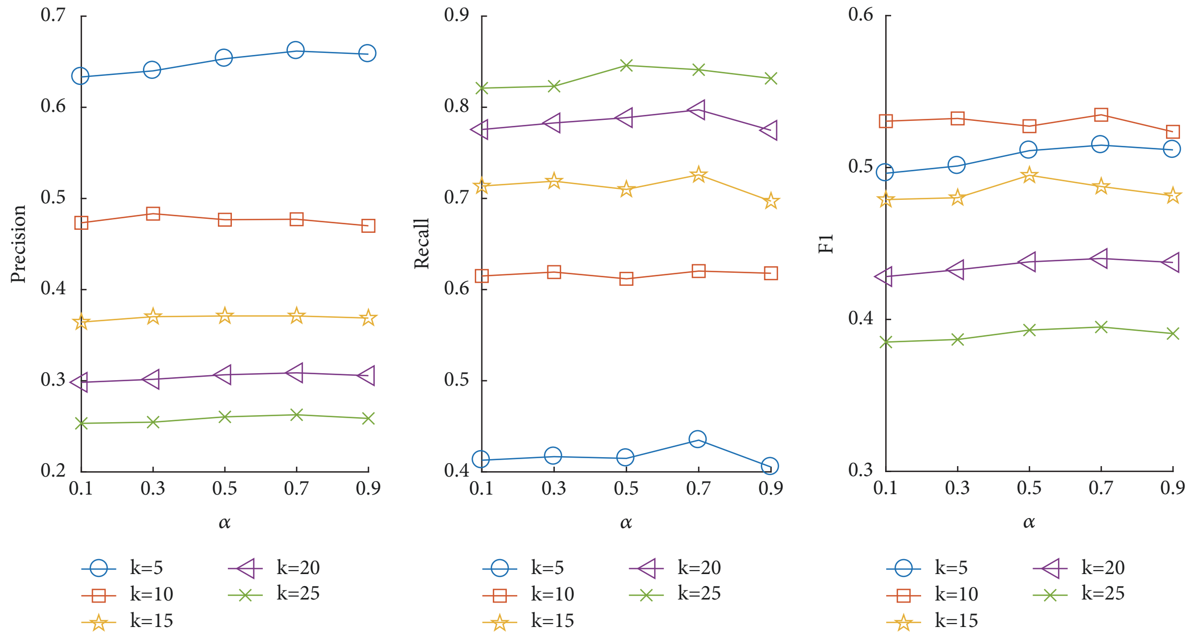

Figure 7: *NLPIR* dataset: precision, recall, and F1 values for the RSIC recommendation method with different k at different coefficient α (ε =0.5).

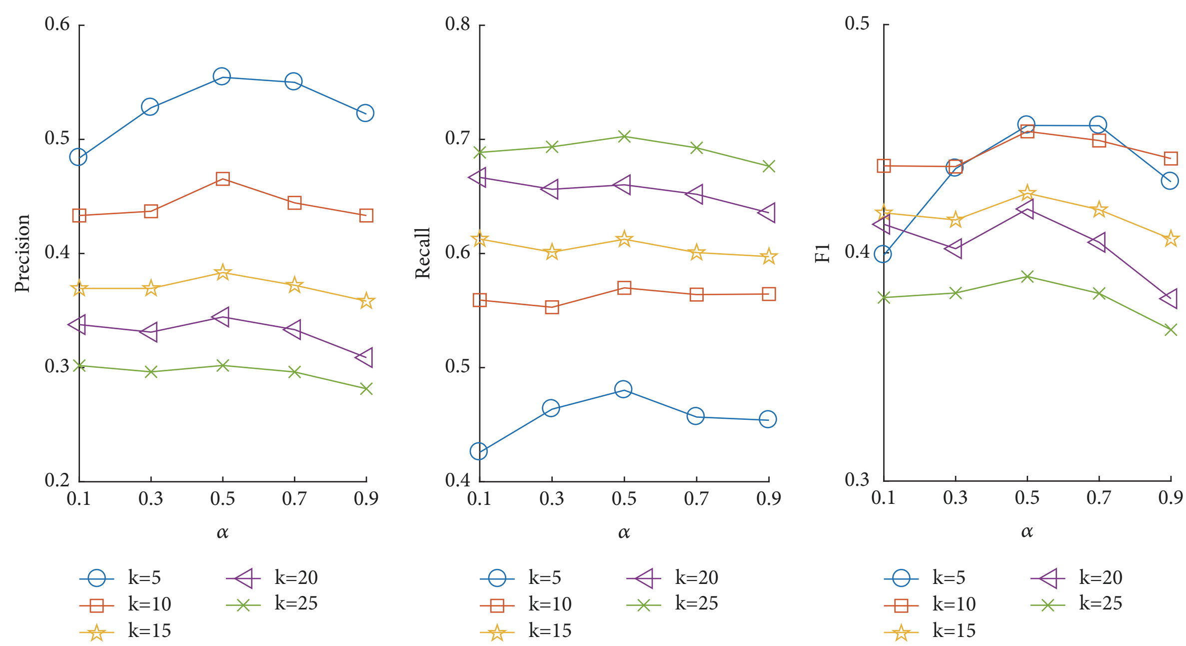

Figure 8: *Applications* dataset: precision, recall, and F1 values for the RSIC recommendation method with different k at different coefficient α (ε =0.6).

subject, the resonance similarity between users is generally small. In different grain subjects, we want to investigate the mechanism of community detection in terms of resonance similarity of users, which can generate different granular communities. According to communities of different granularity, we can consider subjects in coarse-grain community and in fine-grain community to get diverse combination results. Hence, how to design optimized combination recommendation results in communities of different granularity is important for improving users' satisfaction. In the future, it is expected that multigranularity RSIC can make more accurate recommendation services.

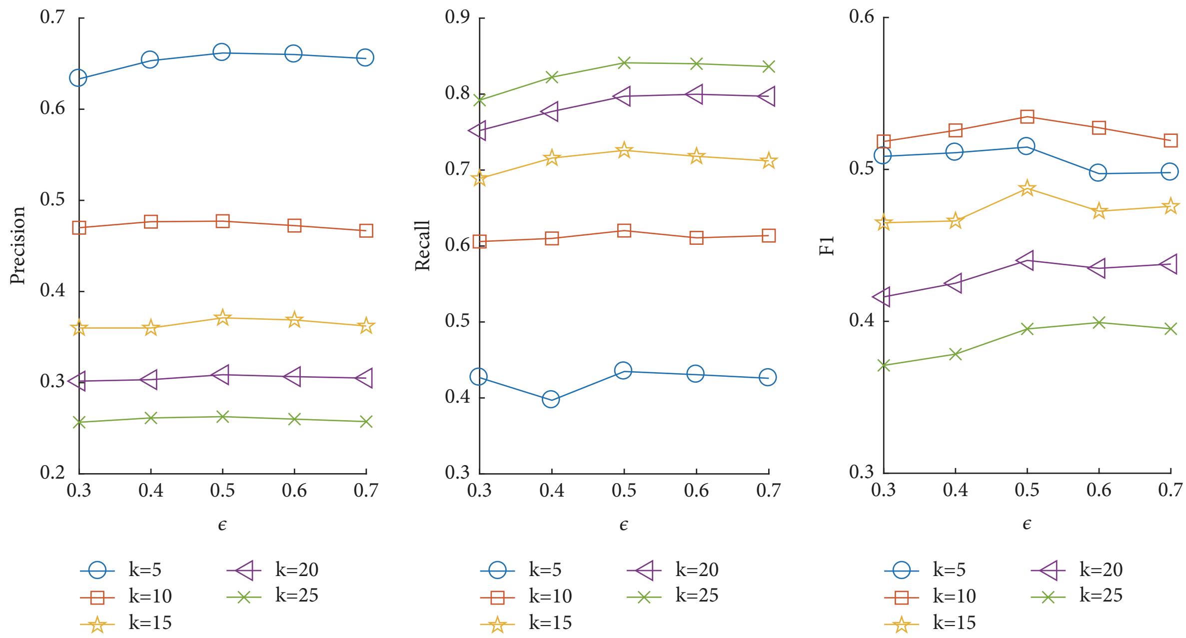

Figure 9: *NLPIR* dataset: precision, recall, and F1 values for the RSIC recommendation method with different k at different cutoff ε (α =0.7).

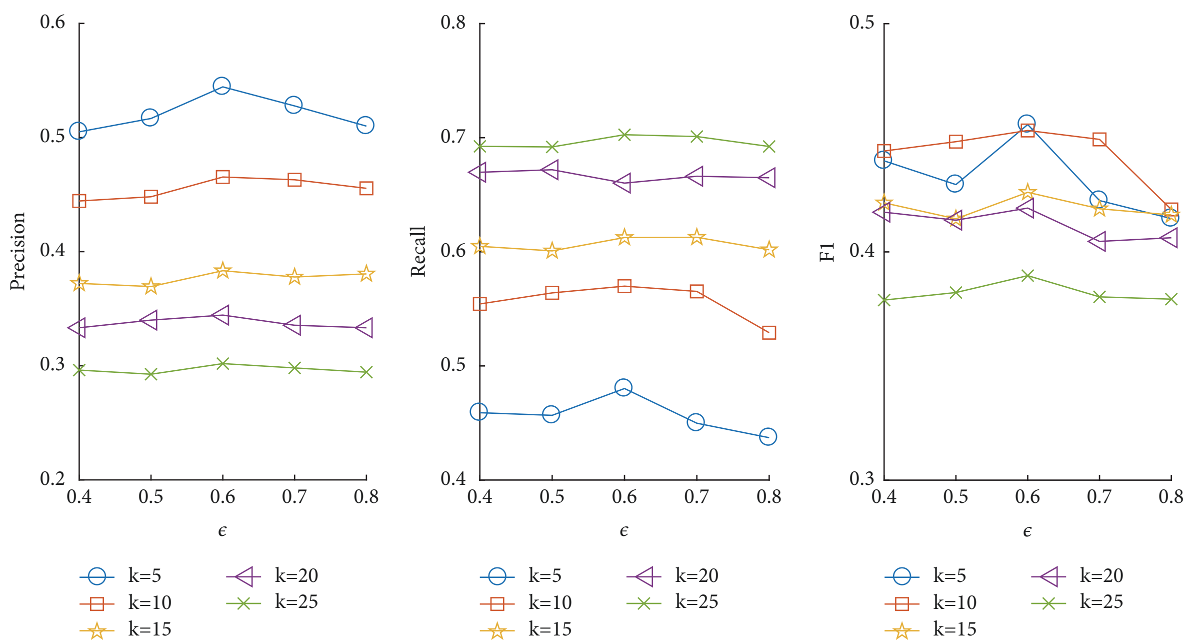

Figure 10: *Applications* dataset: precision, recall, and F1 values for the RSIC recommendation method with different k at different cutoff ε (α =0.5).

Data Availability

To examine the quality of the proposed RSIC-based recommendation method, we used the corpus of NLPIR dataset and Application dataset to verify the efficacy of the system. The NLPIR dataset was derived from the NLPIR website (http://www.nlpir.org/). The Application dataset was collected from Sina Weibo (http://open.weibo.com). The data used to support the findings of this study have not been made available according to Sina's personal information protection policy.

Conflicts of Interest

The authors declare that there are no conflicts of interest of this paper.

Acknowledgments

This work was partially supported by Youth Science Fund Project of the National Natural Science Foundation of China (no. 61603229) and the project of the Natural Science Foundation of Shanxi (201601D202041).

References

[1] K. Ikeda, G. Hattori, C. Ono, H. Asoh, and T. Higashino, "Twitter user profiling based on text and community mining for market analysis," *Knowledge-Based Systems*, vol. 51, pp. 35–47, 2013.

[2] H. Xie, Q. Li, and X. Mao, "Context-Aware Personalized Search Based on User and Resource Profiles in Folksonomies," in *Web Technologies and Applications*, vol. 7235 of *Lecture Notes in Computer Science*, pp. 97–108, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[3] G. A. León-Paredes, L. I. Barbosa-Santillán, and J. J. Sánchez-Escobar, "A Heterogeneous System Based on Latent Semantic Analysis Using GPU and Multi-CPU," *Scientific Programming*, vol. 2017, 19 pages, 2017.

[4] J. Li, J. Li, X. Fu, M. A. Masud, and J. Z. Huang, "Learning distributed word representation with multi-contextual mixed embedding," *Knowledge-Based Systems*, vol. 106, pp. 220–230, 2016.

[5] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of Machine Learning Research*, vol. 3, pp. 1137–1155, 2003.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean, *Efficient estimation of word representations in vector space*, vol. 3, 2013.

[7] M. A. Paredes-Valverde, R. Colomo-Palacios, M. d. Salas-Zárate, and R. Valencia-García, "Sentiment Analysis in Spanish for Improvement of Products and Services: A Deep Learning Approach," *Scientific Programming*, vol. 2017, 6 pages, 2017.

[8] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Systems with Applications*, vol. 69, pp. 214–224, 2017.

[9] F. Enríquez, J. A. Troyano, and T. López-Solaz, "An approach to the use of word embeddings in an opinion classification task," *Expert Systems with Applications*, vol. 66, pp. 1–6, 2016.

[10] R. Xu, T. Chen, Y. Xia, Q. Lu, B. Liu, and X. Wang, "Word embedding composition for data imbalances in sentiment and emotion classification," *Cognitive Computation*, vol. 7, no. 2, pp. 226–240, 2015.

[11] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," *Data Mining: Concepts and Techniques*, 2012.

[12] S. F. Mousavi, M. Safayani, A. Mirzaei, and H. Bahonar, "Hierarchical graph embedding in vector space by graph pyramid," *Pattern Recognition*, vol. 61, pp. 245–254, 2017.

[13] B. Kyoungsoo and S.-G. Cheongju, "Social group recommendation based on dynamic profiles and collaborative filtering," *Neurocomputing*, vol. 209, pp. 3–13, 2016.

[14] L. Boratto, S. Carta, G. Fenu, and R. Saia, "Using neural word embeddings to model user behavior and detect user segments," *Knowledge-Based Systems*, vol. 108, pp. 5–14, 2016.

[15] Q. Du, H. Xie, Y. Cai et al., "Folksonomy-based personalized search by hybrid user profiles in multiple levels," *Neurocomputing*, vol. 204, pp. 142–152, 2016.

[16] Z. Zhang, Y. Liu, G. Xu, and H. Chen, "A weighted adaptation method on learning user preference profile," *Knowledge-Based Systems*, vol. 112, pp. 114–126, 2016.

[17] H. Kumar, S. Lee, and H.-G. Kim, "Exploiting social bookmarking services to build clustered user interest profile for personalized search," *Information Sciences*, vol. 281, pp. 399–417, 2014.

[18] Y. Meguebli, M. Kacimi, B. L. Doan, and F. Popineau, "Building rich user profiles for personalized news recommendation," in *Proceedings of the In Proceedings of the 22nd Conference on User Modelling, Adaptation and Personalization Workshops*, p. 11, Aalborg, Denmark, 2014.

[19] H. Xie, Q. Li, X. Mao, X. Li, Y. Cai, and Y. Rao, "Community-aware user profile enrichment in folksonomy," *Neural Networks*, vol. 58, pp. 111–121, 2014.

[20] R. W. White, P. Bailey, and L. Chen, "Predicting user interests from contextual information," in *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '09)*, pp. 363–370, Boston, Mass, USA, July 2009.

[21] H. Xie, X. Li, T. Wang et al., "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, 2016.

[22] T. Rui, P. Cui, and W. Zhu, "Joint user-interest and social-influence emotion prediction for individuals," *Neurocomputing*, vol. 230, pp. 66–76, 2017.

[23] K. Xu, G. Qi, J. Huang, T. Wu, and X. Fu, "Detecting bursts in sentiment-aware topics from social media," *Knowledge-Based Systems*, vol. 141, pp. 44–54, 2017.

[24] L. Bai, X. Cheng, J. Liang, and Y. Guo, "Fast graph clustering with a new description model for community detection," *Information Sciences*, vol. 388-389, pp. 37–47, 2017.

[25] W. Xiao, Y. Yang, H. Wang, T. Li, and H. Xing, "Semi-supervised hierarchical clustering ensemble and its application," *Neurocomputing*, vol. 173, pp. 1362–1376, 2016.

[26] Y. Chen, X. Wang, X. Xiang et al., "Overlapping community detection in weighted networks via a Bayesian approach," *Physica A: Statistical Mechanics and its Applications*, vol. 468, pp. 790–801, 2017.

[27] H. Feng, J. Tian, H. J. Wang, and M. Li, "Personalized recommendations based on time-weighted overlapping community detection," *Information and Management*, vol. 52, no. 7, pp. 789–800, 2015.

[28] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," *New Journal of Physics* , vol. 11, Article ID 033015, 20 pages, 2009.

[29] M. E. Newman, "Mixing patterns in networks," *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics*, vol. 67, no. 2, 2003.

[30] J. Chen and Y. Saad, "Dense subgraph extraction with application to community detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 7, pp. 1216–1230, 2012.

[31] Y. Rao, Q. Li, L. Wenyin, Q. Wu, and X. Quan, "Affective topic model for social emotion detection," *Neural Networks*, vol. 58, pp. 29–37, 2014.

[32] Y. Rao, H. Xie, J. Li, F. Jin, F. L. Wang, and Q. Li, "Social emotion classification of short text via topic-level maximum entropy model," *Information and Management*, vol. 53, no. 8, pp. 978–986, 2016.

[33] Y. Rao, Q. Li, X. Mao, and L. Wenyin, “Sentiment topic models for social emotion mining,” *Information Sciences*, vol. 266, pp. 90–100, 2014.

[34] H.-H. M. Lee and W. Van Dolen, “Creative participation: Collective sentiment in online co-creation communities,” *Information and Management*, vol. 52, no. 8, pp. 951–964, 2015.

[35] Z. Xiaomei, Y. Jing, Z. Jianpei, and H. Hongyu, “Microblog sentiment analysis with weak dependency connections,” *Knowledge-Based Systems*, vol. 142, pp. 170–180, 2018.

[36] I. Cantador and P. Castells, “Extracting multilayered Communities of Interest from semantic user profiles: Application to group modeling and hybrid recommendations,” *Computers in Human Behavior*, vol. 27, no. 4, pp. 1321–1336, 2011.

[37] S. Garcia Esparza, M. P. O’Mahony, and B. Smyth, “Mining the real-time web: A novel approach to product recommendation,” *Knowledge-Based Systems*, vol. 29, pp. 3–11, 2012.

[38] L. Xiong, Y. Lei, W. Huang, X. Huang, and M. Zhong, “An estimation model for social relationship strength based on users’ profiles, co-occurrence and interaction activities,” *Neurocomputing*, vol. 214, pp. 927–934, 2016.

[39] Y. Zhu, Z. Guan, S. Tan, H. Liu, D. Cai, and X. He, “Heterogeneous hypergraph embedding for document recommendation,” *Neurocomputing*, vol. 216, pp. 150–162, 2016.

[40] X. Bai, B. Barla Cambazoglu, F. Gullo, A. Mantrach, and F. Silvestri, “Exploiting search history of users for news personalization,” *Information Sciences*, vol. 385, pp. 125–137, 2017.

[41] F. Zhang, N. J. Yuan, D. Lian, X. Xie, and W. Ma, “Collaborative Knowledge Base Embedding for Recommender Systems,” in *Proceedings of the the 22nd ACM SIGKDD International Conference*, pp. 353–362, San Francisco, California, USA, August 2016.

[42] Q. Li, S. Shah, A. Nourbakhsh, X. Liu, and R. Fang, “Hashtag recommendation based on topic enhanced embedding, tweet entity data and learning to rank,” in *Proceedings of the 25th ACM International Conference on Information and Knowledge Management, CIKM 2016*, pp. 2085–2088, usa, October 2016.

[43] L. H. Xu, H. F. Lin, Y. Pan, H. Ren, and J. M. Chen, “Constructing the affective lexicon ontology,” *Journal of the China society for scientific and technical information*, vol. 2008, no. 27, pp. 180–185, 2008.