The Silent Majority Speaks: Inferring Silent Users' Opinions in Online Social Networks

Lei Wang, Jianwei Niu, Xuefeng Liu, Kaili Mao*
State Key Laboratory of Virtual Reality Technology and System,
School of Computer Science and Engineering, Beihang University, Beijing 100191, China
{lei,niujianwei,liu_xuefeng,maokaili}@buaa.edu.cn

ABSTRACT

With the blossoming of social networking platforms like Twitter and Facebook, how to infer the opinions of online social network users on specific topics they had not directly given yet, has received much attention. Existing solutions mainly rely on one's previous posted messages. However, recent studies show that over 40% of users opt to be silent all or most of the time and post very few messages. Consequently, the performance of existing solutions will drop dramatically when they are applied to infer silent users' opinions, and how to infer the opinions of these silent users becomes a meaningful while challenging task. Inspired by the collaborative filtering techniques in cold-start recommendations, we infer the opinions of silent users by leveraging the text content posted by active users and their relationships between silent users. Specifically, we first consider both observed and pseudo relationships among users, and cluster users into communities in order to extract various kinds of features for opinion inference. We then design a coupled sparse matrix factorization (CSMF) model to capture the complex relations among these features. Extensive experiments on real-world data from Twitter show that our CSMF model achieves over 80% accuracy for the inference of silent users' opinions.

CCS CONCEPTS

• **Information systems** → **Social networks**; • **Computing methodologies** → *Factorization methods.*

KEYWORDS

Silent users; opinion inference; online social networks

ACM Reference Format:
Lei Wang, Jianwei Niu, Xuefeng Liu, Kaili Mao. 2019. The Silent Majority Speaks: Inferring Silent Users' Opinions in Online Social Networks. In *Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3308558.3313423

1 INTRODUCTION

Recent years have witnessed a booming development of online social networking platforms like Twitter and Facebook. In online social networks, even active users will not publish messages to directly show their opinions towards all topics. As a result, inferring users' opinions on specific topics they had not directly given yet, has attracted increasing attention [11, 13, 26]. When inferring a user's opinion on a specific topic, existing solutions generally are based on his/her previous posted messages on other topics, in order to profile himself/herself [7] or find his/her neighbors with similar user personal interests [9, 17, 30]. For example, [7] extracted two types of features from a user's posted content about other topics to tell if the user has certain opinion bias on all topics (i.e., some users are more likely to be positive or negative). Then the user's opinion on a specific topic can be correlated with or similar as his/her opinions on some other topics.

However, recent work on social science [6, 10, 22] has shown that a significant proportion (over 40%) of users choose to be *silent* all or most of the time, and seldom publish messages to express their opinions towards various topics. Having a better understanding of these silent users' opinions is important, as it greatly helps us to avoid misjudgement of overall population-level opinions, and advances a variety of real-world applications, such as targeted advertising [5, 14] and political election predictions [8]. Inferring the opinions of silent users is a quite challenging task. As silent users do not generate sufficient content, existing solutions which rely on one's historical posted messages often only focus on users who have historical messages, but ignore these silent users who do not post anything in a long period time. The performance of existing solutions will drop dramatically when they are used to predict the opinions of silent users [7, 17]. Different from existing work on opinion inference, we take a step towards the inference of silent users' opinions in this paper.

We find that inferring silent users' opinions is intrinsically similar to cold-start recommendation problem, whose objective is to predict rating scores for specific items rated by new users without any historical preference information. In recommendation systems, users tend to have similar tastes with their friends. As a result, rating scores of new users can be predicted by considering the preference information of their friends in social networks [16, 19]. Similarly, existing studies on social science theories have verified that users that are connected (become friends) are prone to exhibit similar opinions on a certain topic [2, 3, 18]. Therefore, we argue that we can infer silent users' opinions by leveraging the text content generated by active users and their relationships between silent users.

However, applying the methods designed for cold-start recommendations to infer silent users' opinions still entails many challenges due to the following reasons. 1) *Diversity of user-user relations*: Existing studies [4, 24] on social science have found that

*J. Niu is also with the Beijing Advanced Innovation Center for Big Data and Brain Computing (BDBC) and Hangzhou Innovation Institute of Beihang University.

WWW '19, May 13–17, 2019, San Francisco, CA, USA

ACM ISBN 978-1-4503-6674-8/19/05.
https://doi.org/10.1145/3308558.3313423

observed relationships (i.e., the follower/followee network structure) are only one type of user-user relations, while other types of user-user relations are also very important for shared opinions among users. 2) *Heterogeneous fusion*: Many features need to be considered when characterizing silent users and inferring their opinions, including structural features of the user-relationship network, opinion information of text content, users' interests, etc. Simultaneously considering all these features extracted from heterogeneous sources is quite difficult.

To address the aforementioned issues, the inference of silent users' opinions is modeled as a matrix factorization problem in this work, and a coupled sparse matrix factorization (CSMF) model is proposed to fuse text content data and user-relationship data for accurate inference. The main contributions of this paper are summarized as follows:

- To extract features of silent users, we first infer user-user relations and cluster users into communities. Then three sets of features are extracted for each user.
- A coupled sparse matrix factorization (CSMF) model is proposed to capture the complex relations among various features and infer silent users' opinions.
- We validate the effectiveness of our CSMF model based on real-world dataset from Twitter. Experimental results demonstrate that our model can be effectively used to infer silent users' opinions.

2 DATA DESCRIPTION

In this paper, Twitter is employed as the basis for our opinion inference experiments. The Twitter data we used consists of over $70,000$ messages on 20 hot political topics and the follower/followee network structure among over $6,000$ users. For each Twitter message, it is assigned an opinion label: -1 (negative), 0 (neutral) or $+1$ (positive). We employ 20 people to manually assign an opinion label for each message. In this paper, we restrict our attention mainly to negative, neutral and positive user-topic opinions. Then the final opinion of each user on a specific topic is computed by averaging the opinion labels of all his/her posted messages on this topic.

One factor we need to consider is whether we should require both users in a potential pair to connect with each other. On Twitter, mutual connections presumably indicate closer relationships, while one-way connections may correspond to a desire to pay attention (e.g., to a famous singer), rather than necessarily personal relationships. Existing studies [23] have shown that attention effects may be more important than homophily effects with respect to shared opinions in social networks. Therefore, we consider two possibilities when we define that a connection (edge) between a pair of users exists.

- Directed follow network: user u_i follows user u_j (u_j may or may not follow u_i in return).
- Mutual follow network: user u_i follows user u_j, and u_j follows u_i at the same time.

3 FEATURE EXTRACTION FOR USERS

In this section, we extract features for both silent and active users based on text content data and follower/followee network structure.

3.1 Inferring User-User Relations

Existing work [4] has found that there are two types of user-user relations in online social networks: observed relationships (or explicit relationships) and pseudo-relationships. Observed relationships can be reflected by the follower/followee network structure, and provide information on personal relationships among users. Pseudo-relationships are extracted from sentimental text content, where a connection between a pair of users is created if they hold consistent opinions on different topics. Since the inference of silent users' opinions is largely dependent on user-user relations, it is not a trivial matter to infer pseudo-relationships among users and consider both observed and pseudo relationships when inferring silent users' opinions.

Intuitively, a pseudo-friend relationship between a pair of users is likely to exist when they share the same opinions on many topics. In contrast, a pseudo-foe relationship between two users may exist if they have conflicting opinions. To infer the pseudo-relationship between two users, we first introduce the *meta path*, which is used to capture a sequence of relations defined between users [21]. For example, two users u_i and u_j are connected via the path "user-message-topic-message-user" (i.e., U-M-T-M-U) if both u_i and u_j have posted at least one message about a same topic T. Given a meta path (e.g., $\mathcal{P}$=U-M-T-M-U), the similarity between u_i and u_j is computed as follows:

$$sim(u_i, u_j) = \frac{2 \sum_{m_k \in M_{u_i}, m_l \in M_{u_j}} (1_{(p_{u_i \rightsquigarrow u_j} \in \mathcal{P})} \cdot Sc(m_k, m_l))}{|\{p_{u_i \rightsquigarrow u_i} \in \mathcal{P}\}| + |\{p_{u_j \rightsquigarrow u_j} \in \mathcal{P}\}|} \tag{1}$$

where M_{u_i} denotes the set of messages published by user u_i, and $p_{u_i \rightsquigarrow u_j}$ represents a path instance between u_i and u_j that follows the defined meta path $\mathcal{P}$. $Sc(m_k, m_l)$ is calculated as:

$$Sc(m_k, m_l) = \begin{cases} +1 & \text{if } sl(m_k) = sl(m_l) \\ -1 & \text{otherwise} \end{cases} \tag{2}$$

where $sl(m_k)$ is the opinion label of message m_k.

From the definition of $sim(u_i, u_j)$, we can find that it is a real value ranging between -1 and 1. A score $sim(u_i, u_j)$ close to 1 indicates that users u_i and u_j hold similar opinions with respect to different topics, and we denote them as *pseudo-friends*. On the contrary, two users with a score $sim(u_i, u_j)$ close to -1 indicates that they are *pseudo-foes* who hold conflicting opinions on many topics. Moreover, a score close to 0 indicates that these two users share both consistent and conflicting opinions. In this paper, we define a pseudo-connection between users u_i and u_j exists if their $sim(u_i, u_j)$ is larger than a threshold γ.

3.2 Clustering Users into Communities

We need to cluster users into communities, and extract a user's features based on the text content posted by the users in the same community and community structure information. Therefore, in this subsection, we shift our focus to how to cluster users into communities.

A lot of community detection algorithms have been proposed in order to group the users of a network into groups of users with denser connections internally and sparser connections between groups [15, 28, 29]. With user-user relations (including observed relationships and pseudo-relationships) in place, we employ the method proposed in [25], which can achieve the best performance of community detection on large social networks currently as far as we know, to cluster both silent and active users. This method unifies several commonly used clustering quality functions including modularity, cluster deletion, and sparsest cut, and places all these clustering quality functions within a unified optimization framework.

3.3 Features for Opinion Inference

In this subsection, we extract three sets of features for both silent and active users.

Content Features: Content features are used to describe the interested topics of users and the distribution of their opinions. As silent users publish very few messages, we cannot directly extract content features for them from their posted text content. To solve this problem, for each user, we consider the messages posted by both himself/herself and other users in the same community when extracting content features for him/her. For user u, his/her content feature vector is represented as:

$$f_u^C = (< tp_1, s_1 >, < tp_2, s_2 >, ..., < tp_N, s_N >) \tag{3}$$

where N is the total number of topics we consider, $tp_i (1 \leq i \leq N)$ equals to 1 if user u has posted messages about the i-th topic, and 0 otherwise. As mentioned before, we define the opinion of a user on a particular topic as the average opinion labels of all his/her published messages on this topic if he/she has posted messages on this topic. Here, we set s_i as user u's opinion on the i-th topic if u has posted messages on this topic, the average opinion of other users in the same community on the i-th topic if u has not but other users in the same community have posted messages on this topic, and the average opinion of all users in our dataset on this topic otherwise.

Intra-Community Structural Features: To extract intra-community structural features for each user, we consider the user-user relations among the whole users in the same community. For user u who resides in community C, we compute his/her degree, the average distance between u and all other users in C, the number of shortest paths between any two users in C that pass through u, and the Katz centrality [1], which is calculated as:

$$C_{Katz}(u) = \sum_{k=1}^{\infty} \sum_{v=1}^{|C|} \alpha^k (\mathbf{A}^k)_{u,v} \tag{4}$$

where α is an attenuation factor between 0 and 1, and matrix $\mathbf{A}$ is the adjacency matrix, i.e., $a_{u,v} = 1$ if user u is connected with user v, and $a_{u,v} = 0$ otherwise.

Inter-Community Structural Features: For each user, we employ inter-community structural features to indicate which community he/she is located in and the connections between the user and users in other communities. Given a set of users, if these users are clustered into K communities by the aforementioned community detection algorithm, the inter-community structural feature vector for user u can be represented as:

$$f_u^{Inter} = (s_1, s_2, ..., s_K) \tag{5}$$

where $s_i (1 \leq i \leq K) \in \{-1, 0, 1\}$. s_i equals to 0 if user u is located in the i-th community, 1 if u is connected with any user in the i-th community, and -1 otherwise.

4 INFERRING SILENT USERS' OPINIONS

In this section, we introduce how to utilize various kinds of features, and design a matrix factorization based framework to infer the opinions of silent users.

4.1 Data Modelling

We can extract various kinds of information from text content and user-user relation data. Here we build several matrices to capture features extracted from different sources and the relations among them.

Features of a user: For each user, we extract content features f^C, intra-community structural features f^{Intra} and inter-community structural features f^{Inter}. Then we place f^C, f^{Intra} and f^{Inter} of each user into a user feature matrix $\mathbf{Z}$, where each row denotes a user and each column stands for a kind of feature.

Opinion distribution patterns: To reflect opinion distribution patterns for individual users and communities, we extract user-topic opinion matrix $\mathbf{O}$ and community-topic opinion matrix $\mathbf{C}$. Specifically, each row of matrix $\mathbf{O}$ denotes a topic, and each column represents a user. Each element $o_{i,j}$ in $\mathbf{O}$ stands for the opinion of the j-th user on the i-th topic. $o_{i,j}$ equals to *null* if the j-th user had not posted messages on the i-th topic. Matrix $\mathbf{O}$ is very sparse because silent users seldom post messages and active users do not publish messages for all topics. To capture the community-topic opinion information, matrix $\mathbf{C}$ is built. Matrix $\mathbf{C}$ represents opinion distribution patterns for communities, where each row denotes a topic and each column stands for a community. Each community contains several users, and each element of $\mathbf{C}$ is the average opinion labels of users in a particular community on a particular topic. Compared with $\mathbf{O}$, $\mathbf{C}$ is a higher level but denser representation for opinion distribution patterns.

4.2 Coupled Sparse Matrix Factorization (CSMF) Model

In this subsection, we infer the opinions on different topics for each silent user, based on various matrices mentioned before. In the previous subsection, we build three matrices $\mathbf{O}$, $\mathbf{C}$ and $\mathbf{Z}$. Given the above settings, the goal of inferring the opinion of a specific user u_j on a particular topic t_i can be converted into filling the missing value in i-th row and j-th column of $\mathbf{O}$ with the help of $\mathbf{C}$ and $\mathbf{Z}$.

Basically, this goal can be achieved by solely factorizing matrix $\mathbf{O}$. Specifically, a more accurate but compact representation for users and topics in a low-rank space can be found through matrix factorization, and matrix $\mathbf{O}$ can be approximated by a multiplication of two low-rank matrices:

$$\mathbf{O} \approx \mathbf{T} \times \mathbf{U}^T \tag{6}$$

where $\mathbf{T} \in \mathbb{R}^{N \times D}$ and $\mathbf{U} \in \mathbb{R}^{M \times D}$ are low-rank matrices with $D \ll min(M, N)$, and M and N denote the total number of users and topics, respectively. The row vector $T(i, :)$ in $\mathbf{T}$ $(1 \leq i \leq N)$ and $U(j, :)$ in $\mathbf{U}$ $(1 \leq j \leq M)$ are the latent representations for the

i-th topic and j-th user, respectively. In traditional matrix factorization methods, matrix $\mathbf{O}$ can be approximated by minimizing the following objective function:

$$\min_{\mathbf{T},\mathbf{U}} \mathcal{L} = \frac{1}{2}||\mathbf{I} \circ (\mathbf{O} - \mathbf{T}\mathbf{U}^T)||_F^2 \tag{7}$$

where $||\cdot||_F$ represents the Frobenius norm of a matrix. $\mathbf{I}$ is an indicator matrix with its element $I_{i,j}$ equals to 0 if $o_{i,j}$ is *null*, 1 otherwise. The operator "$\circ$" denotes the entry-wise product (i.e., Hadamard product).

Note that in our case, the user-topic opinion matrix $\mathbf{O}$ is quite sparse since the observed user-topic opinions are only a small percent. As a result, the approximation of $\mathbf{O}$ will be not accurate enough if we solely factorize itself. To solve this problem, we can incorporate another two matrices $\mathbf{C}$ and $\mathbf{Z}$ to decompose these matrices collaboratively. To be more specific, we first factorize community-topic opinion matrix $\mathbf{C} \in \mathbb{R}^{N\times K}$ into two low-rank latent matrices $\mathbf{T} \in \mathbb{R}^{N\times D}$ and $\mathbf{G} \in \mathbb{R}^{K\times D}$, where K is the total number of communities. Since matrices $\mathbf{O}$ and $\mathbf{C}$ share the latent topic matrix $\mathbf{T}$, we can make $\mathbf{T}$ more accurate by jointly decomposing $\mathbf{O}$ and $\mathbf{C}$. Likewise, user feature matrix $\mathbf{Z}$ is also employed to help determine the latent user matrix $\mathbf{U}$ more accurately by factorizing $\mathbf{Z} \in \mathbb{R}^{M\times L}$ into $\mathbf{U} \in \mathbb{R}^{M\times D}$ and $\mathbf{F} \in \mathbb{R}^{L\times D}$, where L is the dimension of the user feature vector, and $\mathbf{O}$ and $\mathbf{Z}$ share the latent user matrix $\mathbf{U}$. Since matrices $\mathbf{C}$ and $\mathbf{Z}$ are much denser than $\mathbf{O}$, community-topic opinion matrix $\mathbf{C}$ and user feature matrix $\mathbf{Z}$ can make contributions to facilitating a more accurate factorization of $\mathbf{O}$ through shared matrices $\mathbf{T}$ and $\mathbf{U}$. That is to say, we can collaboratively decompose $\mathbf{O}$, $\mathbf{C}$ and $\mathbf{Z}$ as follows:

$$\mathbf{O} \approx \mathbf{T} \times \mathbf{U}^T; \mathbf{C} \approx \mathbf{T} \times \mathbf{G}^T; \mathbf{Z} \approx \mathbf{U} \times \mathbf{F}^T \tag{8}$$

where $\mathbf{O}$ and $\mathbf{C}$ share latent topic factor $\mathbf{T}$, and $\mathbf{O}$ and $\mathbf{Z}$ share latent user factor $\mathbf{U}$. Then we can achieve this goal by forming and minimizing the following objective function:

$$\begin{aligned}\min_{\mathbf{T},\mathbf{U},\mathbf{F},\mathbf{G}} \mathcal{L} = \frac{1}{2}||\mathbf{I} \circ (\mathbf{O} - \mathbf{T}\mathbf{U}^T)||_F^2 + \frac{\lambda_1}{2}||\mathbf{C} - \mathbf{T}\mathbf{G}^T||_F^2 \\ + \frac{\lambda_2}{2}||\mathbf{Z} - \mathbf{U}\mathbf{F}^T||_F^2\end{aligned} \tag{9}$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are used to control the loss in the matrix factorization. After the factorization, we can recover matrix $\mathbf{O}$ through the production of $\mathbf{T}$ and $\mathbf{U}^T$.

To further alleviate the data sparsity problem, we also take the correlations between users into account. Let matrix $\mathbf{S}$ be the correlation matrix of users, with each element $s_{i,j}$ denoting the correlation (similarity) between users u_i and u_j. $\mathbf{D}$ is a diagonal matrix with the i-th diagonal element as $d_{i,i} = \sum_{j=1}^{M} s_{i,j}$, and $\mathbf{L_S} = (\mathbf{D}-\mathbf{S}) \in \mathbb{R}^{M\times M}$ is the Laplacian matrix of the user correlation graph. Here, we use Pearson correlation analysis method [20] to obtain user correlation matrix $\mathbf{S}$ and the corresponding Laplacian matrix $\mathbf{L_S}$. Then the correlation information between users is considered as the trace of matrix $\mathbf{U}^T\mathbf{L_S}\mathbf{U}$, and can be obtained through the following deduction, which can guarantee that for two users u_i and u_j with higher correlation (i.e., the value of $s_{i,j}$ is large), they will be also closer in distance between the row vectors $U(i,:)$ and $U(j,:)$ in the matrix $\mathbf{U}$.

$$\begin{aligned}&\frac{1}{2}\sum_{i,j} s_{i,j}||U(i,:) - U(j,:)||_F^2 = \frac{1}{2}\sum_{i,j}\sum_{k=1}^{D} s_{i,j}(U(i,k) - U(j,k))^2 \\ &= \sum_{i,j}\sum_{k=1}^{D} s_{i,j}U^2(i,k) - \sum_{i,j}\sum_{k=1}^{D} s_{i,j}U(i,k)U(j,k) \\ &= \sum_{k=1}^{D} U^T(:,k)(\mathbf{D}-\mathbf{S})U(:,k) = tr(\mathbf{U}^T(\mathbf{D}-\mathbf{S})\mathbf{U}) \\ &= tr(\mathbf{U}^T\mathbf{L_S}\mathbf{U})\end{aligned} \tag{10}$$

where $tr(\cdot)$ represents the matrix trace. In this way, user correlation matrix $\mathbf{S}$ contributes to a more accurate $\mathbf{U}$.

After all components mentioned are combined, the final objective function of the coupled sparse matrix factorization model is formally defined as:

$$\begin{aligned}\min_{\mathbf{T},\mathbf{U},\mathbf{F},\mathbf{G}} \mathcal{L} = &\frac{1}{2}\underbrace{||\mathbf{I} \circ (\mathbf{O} - \mathbf{T}\mathbf{U}^T)||_F^2}_{\text{user-topic factorization}} + \frac{\lambda_1}{2}\underbrace{||\mathbf{C} - \mathbf{T}\mathbf{G}^T||_F^2}_{\text{community-topic factorization}} \\ &+ \frac{\lambda_2}{2}\underbrace{||\mathbf{Z} - \mathbf{U}\mathbf{F}^T||_F^2}_{\text{user feature factorization}} + \frac{\lambda_3}{2}\underbrace{tr(\mathbf{U}^T\mathbf{L_S}\mathbf{U})}_{\text{user correlation}} \\ &+ \frac{\lambda_4}{2}\underbrace{(||\mathbf{T}||_{2,1} + ||\mathbf{U}||_{2,1} + ||\mathbf{G}||_{2,1} + ||\mathbf{F}||_{2,1})}_{\text{row-wise sparsity regularization}}\end{aligned} \tag{11}$$

Finally, we can utilize the factorized matrices $\mathbf{T}$ and $\mathbf{U}$ to reconstruct $\mathbf{O}$ to $\widetilde{\mathbf{O}}$, namely $\widetilde{\mathbf{O}} = \mathbf{T} \times \mathbf{U}^T$. For a silent user u_j, his/her opinion on topic t_i is the value in i-th row and j-th column of $\widetilde{\mathbf{O}}$. In the objective function, the last term is a regularization of penalty to prevent overfitting, where the $\ell_{2,1}$-norm regularization is added on each latent matrix. The features we extract contain different types of knowledge, and we assume that only parts of them are useful for opinion inference. The $\ell_{2,1}$-norm can promote row-wise sparsity of the target matrix, making it suitable for feature selection. Hence, the last term can also help control group feature selection.

5 OPTIMIZATION ALGORITHM

The objective function we defined in the previous section is not jointly convex to all the variables. As a result, it is not easy for us to obtain a closed-form solution for minimizing this objective function. To solve this problem, we utilize an iterative algorithm, where we alternatively update one variable while fix other variables until convergence. Note that $\ell_{2,1}$-norm is not continuous on the origin. Hence, we cannot minimize the compound $\ell_{2,1}$ objective function directly. Motivated by recent studies on half-quadratic minimization approaches [27], we can first introduce an auxiliary variable, and then apply this auxiliary variable to transform the $\ell_{2,1}$-norm term of the objective function into an approximate form. For instance, $\mathbf{G}$ can be transformed as:

$$||\mathbf{G}||_{2,1} \approx tr(\mathbf{G}^T\mathbf{Q_G}\mathbf{G}) \tag{12}$$

where $\mathbf{Q_G}$ is a diagonal matrix with the i-th diagonal element as $q_{i,i}^{g} = \frac{1}{2||\mathbf{g_i}||_2}$. $\mathbf{T}$, $\mathbf{U}$ and $\mathbf{F}$ can be dealt with in the same way.

After having the above transformations, we then compute the derivative of the objective function with respect to $\mathbf{T}$, $\mathbf{U}$, $\mathbf{G}$ and $\mathbf{F}$ respectively:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \mathbf{T}} &= [\mathbf{I} \circ (\mathbf{T}\mathbf{U}^T - \mathbf{O})]\mathbf{U} + \lambda_1(\mathbf{T}\mathbf{G}^T - \mathbf{C})\mathbf{G} + \lambda_4 \mathbf{Q}_\mathbf{T}\mathbf{T} \\ \frac{\partial \mathcal{L}}{\partial \mathbf{U}} &= [\mathbf{I} \circ (\mathbf{T}\mathbf{U}^T - \mathbf{O})]^T\mathbf{T} + \lambda_2(\mathbf{U}\mathbf{F}^T - \mathbf{Z})\mathbf{F} \\ &\quad + \lambda_3 \mathbf{L}_\mathbf{S}\mathbf{U} + \lambda_4 \mathbf{Q}_\mathbf{U}\mathbf{U} \\ \frac{\partial \mathcal{L}}{\partial \mathbf{G}} &= \lambda_1[\mathbf{T}\mathbf{G}^T - \mathbf{C}]^T\mathbf{T} + \lambda_4 \mathbf{Q}_\mathbf{G}\mathbf{G} \\ \frac{\partial \mathcal{L}}{\partial \mathbf{F}} &= \lambda_2[\mathbf{U}\mathbf{F}^T - \mathbf{Z}]^T\mathbf{U} + \lambda_4 \mathbf{Q}_\mathbf{F}\mathbf{F}\end{aligned} \tag{13}$$

With the derivatives in place, a conjugate gradient method can be employed to iteratively minimize the objective function. We show the detailed process of the optimization algorithm in Algorithm 1.

Algorithm 1 Algorithm to minimize the objective function

Input: User-topic opinion matrix $\mathbf{O}$, community-topic opinion matrix $\mathbf{C}$, user feature matrix $\mathbf{Z}$, Laplacian matrix $\mathbf{L}_\mathbf{S}$ and regularization parameters.
Output: $\mathbf{T}$ and $\mathbf{U}$.

1: Initialize $\mathbf{T}$, $\mathbf{U}$, $\mathbf{G}$ and $\mathbf{F}$ randomly.
2: Compute the diagonal matrices $\mathbf{Q}_\mathbf{T}$, $\mathbf{Q}_\mathbf{U}$, $\mathbf{Q}_\mathbf{G}$ and $\mathbf{Q}_\mathbf{F}$.
3: **repeat**
4: Update $\mathbf{T}$, $\mathbf{U}$, $\mathbf{G}$ and $\mathbf{F}$ according to Eq. (13) by using the conjugate gradient method.
5: Update $\mathbf{Q}_\mathbf{T}$, $\mathbf{Q}_\mathbf{U}$, $\mathbf{Q}_\mathbf{G}$ and $\mathbf{Q}_\mathbf{F}$ with $q^t_{i,i} = \frac{1}{2||\mathbf{t}_i||_2}$, $q^u_{i,i} = \frac{1}{2||\mathbf{u}_i||_2}$, $q^g_{i,i} = \frac{1}{2||\mathbf{g}_i||_2}$, $q^f_{i,i} = \frac{1}{2||\mathbf{f}_i||_2}$.
6: **until** convergence

6 EXPERIMENTS

6.1 Evaluation Metrics

We employ two popular metrics, i.e., root mean square error (RMSE) and Accuracy, to measure the performance of opinion inference task. The metric RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N_T}\sum_{i,j}(o_{i,j} - \widetilde{o}_{i,j})^2} \tag{14}$$

where N_T stands for the number of opinions for testing. Since the value of $\widetilde{o}_{i,j}$ may be not an integer which exactly equals to -1, 0 or 1, we map the value of $\widetilde{o}_{i,j}$ with the *sign function* before computing the Accuracy. $\widetilde{o}_{i,j}$ is mapped to -1 if $\widetilde{o}_{i,j} \in (-\infty, -0.5]$, 0 if $\widetilde{o}_{i,j} \in (-0.5, 0.5)$, and 1 if $\widetilde{o}_{i,j} \in [0.5, +\infty)$. In experiments, a smaller RMSE value or a higher Accuracy value indicates a better inference performance.

6.2 Comparison with Baselines

In this subsection, we compare our CSMF model with several baseline methods. To analyze how the performance of these methods changes with the change of active user size, we select 30, 40, 60, 70 and 80 percent of the most active users as active users, and infer the opinions of the remaining portions of users (silent users). For users in the test set (silent users), all their posted messages are removed when extracting features for them, and their observed user-topic opinions are regarded as ground truth data. Then we predict their opinions using different methods. The baseline methods are listed as follows:

Table 1: RMSE comparisons using different methods and sizes of active users.

Active users	*RK*	*MF*	*STMF*	*DCMF*	*OP-SVM*	*LoCo*	*CSMF*
30%	0.92	0.80	0.75	0.71	0.68	0.70	**0.59**
40%	0.88	0.76	0.71	0.67	0.66	0.65	**0.56**
60%	0.87	0.72	0.69	0.64	0.65	0.64	**0.52**
70%	0.83	0.71	0.68	0.63	0.62	0.60	**0.50**
80%	0.81	0.71	0.65	0.62	0.61	0.58	**0.49**

- RK: Regression-Kriging is a liner combination estimator that combines a regression of the dependent variable on auxiliary variables with the regression residuals.
- MF: This method is the basic low-rank matrix factorization model, as shown in Eq. (7).
- STMF: It is proposed in [17], where a matrix factorization framework is built to predict user-topic opinions by considering users' previous posted text content and the similarities between different users.
- DCMF: This method [12] first learns hash codes for users and topics from content information, and then introduces an interaction regularization to tackle the data sparsity problem.
- OP-SVM: It is an opinion prediction model based on two types of features (sentiment features and opinion features) extracted from one's historical posted messages on other topics [7].
- LoCo: It is a matrix factorization framework designed for cold-start recommendation systems [19]. The inference of silent users' opinions is intrinsically similar to cold-start recommendation problem. Therefore, we adopt this method as a baseline method.

In this experiment, for the methods using regularization constraints, we adjust the regularization parameters of them in order to obtain their best performance. In our CSMF model, the best result is achieved when λ_1, λ_2, λ_3 and λ_4 are set as 1, 10^{-1}, 10^{-1} and 1 respectively. In addition, we only consider directed follow network when extracting features for users, and set γ, α and K as 0.7, 0.6 and 65 respectively based on the results of pre-performed testing for parameter tuning. We display the experimental results measured by RMSE and Accuracy in Table 1 and Figure 1.

From Table 1 and Figure 1, we can see that all these methods can achieve much better performance than randomly guessing (which is $\frac{1}{3}$ in our task). When we increase the size of active users, the performance of these methods gets improved gradually. RK method achieves the worst performance among all baseline methods, both by RMSE and Accuracy. The underlying reason may be that the relationships among different features are so complex that they cannot be captured by a liner model. Compared with basic MF model,

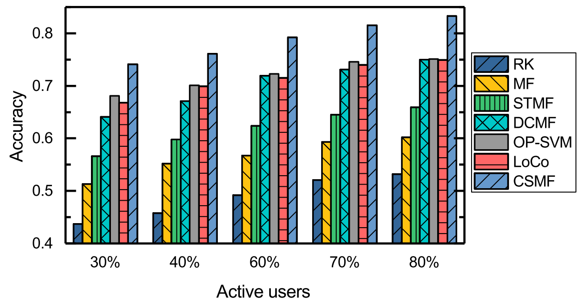

Figure 1: Accuracy comparisons using different methods and sizes of active users.

STMF, DCMF, LoCo and CSMF can obtain significantly improved results, indicating that the regularization constraints improve the inference of silent users' opinions. STMF performs much worse than DCMF, OP-SVM, LoCo and CSMF, which may be caused by that STMF relies on a large number of previous published messages per user to compute the similarities between users while few messages are available for each user in our dataset. For OP-SVM model, one type of features used in the model (i.e., opinion features) cannot be extracted when the user's all ground truth opinions on other topics are not available. As a result, OP-SVM model also obtains worse result than our CSMF model. Our proposed CSMF model always generates the best performance, and this observation demonstrates that incorporating text content and network structural features actually benefits the inference of silent users' opinions.

6.3 Effect of User-User Relations

When considering the explicit relationships (i.e., the follower/followee network structure) among users in the previous section, we define two types of connections between users, i.e., directed follow network and mutual follow network. In addition, we infer pseudo-relationships among users before extracting structural features for users. Here we analyze how the performance of inferring silent users' opinions is affected by the type of connection and pseudo-relationships. In this experiment, we select 60% of the whole users as active users. The experimental results are shown in Figure 2.

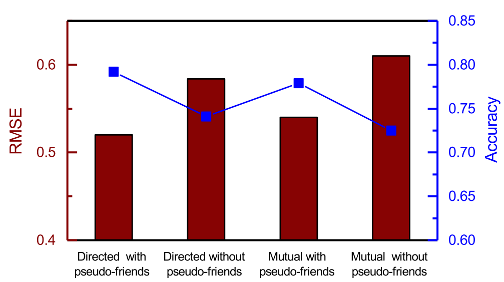

Figure 2: Effect of user-user relations on RMSE and Accuracy.

As we can see from Figure 2, the best performance is achieved when both directed follow network and pseudo-friend information are taken into consideration. For both two types of connections between users, much better results are generated after pseudo-relationships are combined with the explicit relationships. This result demonstrates that the pseudo-relationships are indeed highly correlated with shared opinions among users. Comparing the directed network with the mutual network, the directed network performs slightly better than the mutual network, indicating that attention effects are more important than homophily effects for shared opinions in social networks.

7 CONCLUSION

Existing studies on inferring users' opinions basically rely on one's previous posted messages. As silent users do not generate sufficient content, the performance of existing solutions will drop dramatically if they are applied to infer the opinions of silent users. To solve this problem, we model the opinion inference problem as a matrix factorization problem, and propose a coupled sparse matrix factorization (CSMF) model to infer the opinions of silent users. We propose several sets of features for opinion inference from different perspectives, and capture the complex relations among these features through the CSMF model. Experimental results based on real-world data from Twitter demonstrate that our model achieves much better result than the state-of-the-art methods.

ACKNOWLEDGMENTS

This work was supported by National Key R&D Program of China (2017YFB1301100), National Natural Science Foundation of China (61572060, 61772060, 61572218, 61728201) and CERNET Innovation Project (NGII20170315, NGII20160316).

REFERENCES

[1] Stephen P Borgatti. 2005. Centrality and network flow. *Social networks* 27, 1 (2005), 55–71.

[2] Hongxu Chen, Hongzhi Yin, Xue Li, Meng Wang, Weitong Chen, and Tong Chen. 2017. People opinion topic model: opinion based user clustering in social networks. In *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 1353–1359.

[3] Kewei Cheng, Jundong Li, Jiliang Tang, and Huan Liu. 2017. Unsupervised Sentiment Analysis with Signed Social Networks.. In *AAAI*. 3429–3435.

[4] Hongbo Deng, Jiawei Han, Hao Li, Heng Ji, Hongning Wang, and Yue Lu. 2013. Exploring and Inferring User-User Pseudo-Friendship for Sentiment Analysis with Heterogeneous Networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 378–386.

[5] Mauro Dragoni. 2017. A three-phase approach for exploiting opinion mining in computational advertising. *IEEE Intelligent Systems* 32, 3 (2017), 21–27.

[6] Wei Gong, Ee-Peng Lim, and Feida Zhu. 2015. Characterizing Silent Users in Social Media Communities.. In *ICWSM*. 140–149.

[7] Wei Gong, Ee-Peng Lim, Feida Zhu, and Pei Hua Cher. 2016. On Unravelling Opinions of Issue Specific-Silent Users in Social Media. ICWSM.

[8] Kristen Johnson, Di Jin, and Dan Goldwasser. 2017. Leveraging Behavioral and Social Information for Weakly Supervised Collective Classification of Political Discourse on Twitter. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vol. 1. 741–752.

[9] Jihie Kim, Jae-Bong Yoo, Ho Lim, Huida Qiu, Zornitsa Kozareva, and Aram Galstyan. 2013. Sentiment Prediction Using Collaborative Filtering.. In *ICWSM*.

[10] Yoree Koh. 2014. Report: 44% of Twitter Accounts Have Never Sent a Tweet. In *The Wall Street Journal*. News Corporation.

[11] Tao Li, Yi Zhang, and Vikas Sindhwani. 2009. A non-negative matrix tri-factorization approach to sentiment classification with lexical prior knowledge. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-Volume 1*. Association for Computational Linguistics, 244–252.

[12] Defu Lian, Rui Liu, Yong Ge, Kai Zheng, Xing Xie, and Longbing Cao. 2017. Discrete content-aware matrix factorization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 325–334.

[13] Shangsong Liang, Xiangliang Zhang, Zhaochun Ren, and Evangelos Kanoulas. 2018. Dynamic Embeddings for User Profiling in Twitter. In *Proceedings of the 24th*

ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 1764–1773.
[14] Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* 5, 1 (2012), 1–167.
[15] Zongqing Lu, Xiao Sun, Yonggang Wen, Guohong Cao, and Thomas La Porta. 2015. Algorithms and applications for community detection in weighted networks. *IEEE Transactions on Parallel and Distributed Systems* 26, 11 (2015), 2916–2926.
[16] Huifang Ma, Meihuizi Jia, Di Zhang, and Xianghong Lin. 2017. Combining tag correlation and user social relation for microblog recommendation. *Information Sciences* 385 (2017), 325–337.
[17] Fuji Ren and Ye Wu. 2013. Predicting user-topic opinions in twitter with social and topical context. *IEEE Transactions on Affective Computing* 4, 4 (2013), 412–424.
[18] Amin Salehi, Mert Ozer, and Hassan Davulcu. 2018. Sentiment-driven Community Profiling and Detection on Social Media. In *Proceedings of the 29th on Hypertext and Social Media (HT '18)*. ACM, New York, NY, USA, 229–237.
[19] Suvash Sedhain, Aditya Krishna Menon, Scott Sanner, Lexing Xie, and Darius Braziunas. 2017. Low-Rank Linear Cold-Start Recommendation from Social Data.. In *AAAI*. 1502–1508.
[20] Peter Sprent and Nigel C Smeeton. 2000. *Applied nonparametric statistical methods*. Chapman and Hall/CRC.
[21] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. 2011. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *Proceedings of the VLDB Endowment* 4, 11 (2011), 992–1003.
[22] Andrea Tagarelli and Roberto Interdonato. 2013. Who's out there?: identifying and ranking lurkers in social networks. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM, 215–222.
[23] Chenhao Tan, Lillian Lee, Jie Tang, Long Jiang, Ming Zhou, and Ping Li. 2011. User-level sentiment analysis incorporating social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1397–1405.
[24] Rodoula H Tsiotsou. 2015. The role of social and parasocial relationships on social networking sites loyalty. *Computers in Human Behavior* 48 (2015), 401–414.
[25] Nate Veldt, David F Gleich, and Anthony Wirth. 2018. A Correlation Clustering Framework for Community Detection. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 439–448.
[26] Svitlana Volkova, Glen Coppersmith, and Benjamin Van. 2014. Inferring user political preferences from streaming communications. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Vol. 1. 186–196.
[27] Yuqi Wang, Jiannong Cao, Lifang He, Wengen Li, Lichao Sun, and Philip S Yu. 2017. Coupled Sparse Matrix Factorization for Response Time Prediction in Logistics Services. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 939–947.
[28] Liang Yang, Xiaochun Cao, Di Jin, Xiao Wang, and Dan Meng. 2015. A unified semi-supervised community detection framework using latent space graph regularization. *IEEE transactions on cybernetics* 45, 11 (2015), 2585–2598.
[29] Lei Zhang, Hebin Pan, Yansen Su, Xingyi Zhang, and Yunyun Niu. 2017. A mixed representation-based multiobjective evolutionary algorithm for overlapping community detection. *IEEE Transactions on Cybernetics* 47, 9 (2017), 2703–2716.
[30] Guoshuai Zhao, Xueming Qian, and Xing Xie. 2017. User-service rating prediction by exploring social users' rating behaviors. *IEEE transactions on multimedia* 18, 3 (2017), 496–506.