

Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Computing and Information Systems

School of Computing and Information Systems

6-2013

Mining user relations from online discussions using sentiment analysis and probabilistic matrix factorization

Minghui QIU

Singapore Management University, minghui.qiu.2010@smu.edu.sg

Liu Yang

Singapore Management University, liuyang@smu.edu.sg

Jing JIANG

Singapore Management University, jingjiang@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/sis_research



Part of the [Databases and Information Systems Commons](#)

Citation

QIU, Minghui; Yang, Liu; and JIANG, Jing. Mining user relations from online discussions using sentiment analysis and probabilistic matrix factorization. (2013). *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL'13)*. 401-410.
Available at: https://ink.library.smu.edu.sg/sis_research/2053

This Conference Paper is brought to you for free and open access by the School of Computing and Information Systems at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Computing and Information Systems by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Mining User Relations from Online Discussions using Sentiment Analysis and Probabilistic Matrix Factorization

Minghui Qiu[†], Liu Yang^{†,‡}, Jing Jiang[†]

[†] School of Information Systems, Singapore Management University, Singapore

[‡] School of Software and Microelectronics, Peking University, China

{minghui.qiu.2010, jingjiang}@smu.edu.sg, yang.liu@pku.edu.cn

Abstract

Advances in sentiment analysis have enabled extraction of user relations implied in online textual exchanges such as forum posts. However, recent studies in this direction only consider direct relation extraction from text. As user interactions can be sparse in online discussions, we propose to apply collaborative filtering through probabilistic matrix factorization to generalize and improve the opinion matrices extracted from forum posts. Experiments with two tasks show that the learned latent factor representation can give good performance on a relation polarity prediction task and improve the performance of a subgroup detection task.

1 Introduction

The fast growth of the social Web has led to a large amount of interest in online social network analysis. Most existing work on social network analysis relies on explicit links among users such as undirected friendship relations (Liben-Nowell and Kleinberg, 2003), directed following relations (Hopcroft et al., 2011) and trust/distrust relations (Leskovec et al., 2010). However, besides these explicit social relations, the various kinds of interactions between online users often suggest other implicit relations. In particular, in online discussion forums, users interact through textual posts and these exchanged texts often reveal whether two users are friends or foes, or whether two users share the same viewpoint towards a given issue.

To uncover such implicit relations requires text analysis and particularly sentiment analysis. Re-

cently, Hassan et al. (2012) studied predicting the polarity of user interactions in online discussions based on textual exchanges. They found that the automatically predicted signed relations had an accuracy above 80%. The extracted signed network was further used to detect ideological subgroups. This is a piece of pioneering work that extracts online social relations based on text analysis.

In this paper, we further extend the idea of mining social relations from online forum posts by incorporating collaborative filtering. Our work is motivated by the observation that direct textual exchanges between users are sparse. For example, in the data set we use, only around 13% of user-user pairs have direct interactions. Collaborative filtering is a commonly used technique in recommender systems to predict missing ratings. The key assumption is that if two people have the same opinion on an item A , they are likely to also have the same opinion on a different item B . In online discussion forums, users express their opinions about each other as well as the various aspects of the topic under discussion, but not every user comments on every aspect or every other user. Collaborative filtering allows us to identify users with the same opinion even if they have not directly interacted with each other or commented on any common aspect.

Our method starts with extracting opinions on users and topic aspects from online posts using sentiment analysis. The results are two matrices indicating the sentiment polarity scores between pairs of users and pairs of a user and an aspect. To incorporate collaborative filtering, we choose probabilistic matrix factorization (PMF) (Salakhutdinov

and Mnih, 2008), a technique that has been successfully applied for collaborative filtering-based recommendation problems. PMF automatically discovers a low-rank representation for both users and items based on observed rating data. In our problem, the predicted sentiment polarity scores are treated as rating data, and the results of PMF are low-rank vectors representing each user in online discussions.

We evaluate our method on two tasks. The first is to predict the polarity of interactions between two users not from their own textual exchanges but from their interactions with other users or comments on topic aspects. The second is to use the latent vectors to group users based on viewpoints. We find that the latent factor representation can produce good prediction results for the first task and improve the clustering results of the second task compared with a number of baselines, showing the effectiveness of collaborative filtering for mining social relations from online discussions.

2 Related Work

Our work is closely related to recent studies on detecting subgroups from online discussions (Abu-Jbara et al., 2012; Dasigi et al., 2012; Hassan et al., 2012). Abu-Jbara et al. (2012) proposed to build discussant attitude profiles (DAP) from online posts and use these profiles to cluster users into subgroups. A DAP is a vector that contains the attitudes of a discussant towards other discussants and a set of opinion targets. We also extract opinions of users towards other users and opinion targets from posts, which are similar to DAPs. The difference is that we further apply probabilistic matrix factorization to derive a low-rank representation from the raw opinion scores. Our comparison with DAP-based clustering shows that probabilistic matrix factorization can improve subgroup detection. Hassan et al. (2012) proposed to predict the polarity of interactions between users based on their textual exchanges. They defined a set of interaction features using sentiment analysis and applied supervised learning for polarity prediction. In comparison, our work is unsupervised, that is, we do not use any ground truth of interaction polarity for training.

Probabilistic matrix factorization was proposed by Salakhutdinov and Mnih (2008) as a collabo-

rative filtering method for recommender systems. It has attracted much attention and been extended by Ma et al. (2008) and Wang and Blei (2011). In particular, Ma et al. (2008) proposed a SocRec model that combines social network information with rating data using the PMF framework to perform social recommendation. Our model bears similarity to SocRec in that we also consider two types of interactions, i.e. user-user interactions and user-aspect interactions. However, different from Ma et al. (2008), we predict both the user-user and user-aspect scores from textual posts using sentiment analysis, and the user-user opinion polarity scores are symmetric.

Part of our method uses sentiment analysis to extract opinions from text. This is built on top of a large body of existing work on opinion extraction, e.g. Choi et al. (2006) and Wu et al. (2009). As the sentiment analysis component is not our main contribution, we do not review existing work along this direction in detail here. Interested readers can refer to Pang and Lee (2008).

The idea of incorporating sentiment analysis into collaborative filtering algorithms has been explored by Kawamae (2011), Moshfeghi et al. (2011) and Leung et al. (2011). While their work also combines sentiment analysis with collaborative filtering, the purpose is to improve the accuracy of item recommendation. In contrast, we borrow the idea and technique of collaborative filtering to improve user relation mining from online text.

3 Method Overview

In this section, we provide an overview of our method. We first introduce some concepts.

User: We use *user* to refer to a discussant in an online discussion. Each user has an online ID, which can be used by other users to refer to him/her in a post. Users are both opinion holders and opinion targets. For example, User 1 below expresses a negative opinion towards another user in the following snippet.

User 1: Actually, I have to disagree with you.

Aspect: We use *topic aspect* or *aspect* to refer to an opinion target that is related to the topic under discussion. For example, when debating about whether one should vote for Obama, people may express

opinions on targets such as “President Obama” and “Republican party,” as shown in the following snippets. These aspects are all related to Obama’s presidential campaign. As we will explain later, the aspects we consider are named entities and frequent noun phrases.

*User 2: Americans should vote for **President Obama** because he picks good corporations as winners.*

*User 3: I simply point out how absolutely terrible the **Republican party** is.*

Polarity Score: A sentiment polarity score is a real number between 0 and 1, where 0 indicates a completely negative opinion and 1 indicates a completely positive opinion.

User-User Opinion Matrix: The opinions extracted from posts between users are represented by a user-user opinion matrix S , where entry $s_{i,j}$ is a polarity score between the i -th user and the j -th user. We assume that the polarity scores are symmetric.

User-Aspect Opinion Matrix: The opinions held by different users on the various topic aspects are represented by a user-aspect opinion matrix R , where entry $r_{i,k}$ is a polarity score indicating the i -th user’s opinion towards the k -th aspect.

Given the matrices S and R , we perform probabilistic matrix factorization to derive a low-rank vector representation for users and aspects such that if the polarity score between two users or a user and an aspect is high, the dot product between the corresponding two vectors is also high.

In Section 4, we will explain in detail how we identify topic aspects from a discussion thread and how we obtain polarity scores from posts. In Section 5, we will present the details of our probabilistic matrix factorization model.

4 Construction of Opinion Matrices

The opinion matrices are constructed from a single forum thread discussing some controversial topic.

4.1 Aspect Identification

As we have pointed out, there are two kinds of opinion targets, namely users and aspects. Users are clearly defined and can often be identified in posts by their IDs or second person pronouns. For aspects, however, there is not a pre-defined set. We observe that these topic aspects are usually named entities

or noun phrases frequently mentioned. We therefore use the OpenNLP toolkit¹ to perform chunking and obtain noun phrases and the Stanford NER tagger² to identify named entities from the posts.

Some of the candidate aspect phrases identified above actually refer to the same actual aspect, e.g. “Obama voter,” “Obama voters” and “the Obama voter.” We remove stop words from each candidate phrase and use the WordNet by Miller (1995) to obtain the lemma of each word such that we can normalize the candidate aspect phrases to some extent.

Finally, to select salient aspects for a given discussion topic, we count the number of times each candidate aspect has been expressed a positive or negative opinion on by all users, and select those candidate aspects which have opinion expressions from at least M users. We set M to 2 in our experiments. Figure 1 shows the top salient aspects for the thread on “Will you vote for Obama?” We acknowledge there are still duplicate aspects in the results like “Republican Party” and “GOP”. To normalize these aspects, some additional information such as Wikipedia entries and Google snippets may be considered. We will study this problem in our future work.

4.2 Opinion Expression Identification

Our next step is to identify candidate opinion expressions. This problem has been studied in Hu and Liu (2004), Popescu and Etzioni (2005), and Hassan and Radev (2010). Based on previous work, we do the following. We first combine three popular sentiment lexicons to form a single sentiment lexicon: the lexicon used in Hu and Liu (2004), MPQA Subjectivity Lexicon by Wilson et al. (2005) and Senti-WordNet by Baccianella et al. (2010). Our final sentiment lexicon contains 15,322 negative expressions and 10,144 positive expressions. We then identify candidate opinion expressions by searching for occurrences of words in this lexicon in the posts.

4.3 Opinion Relation Extraction

Given a post that contains an aspect and an opinion expression, we still need to determine whether the opinion expression is used to describe the aspect. This is a relation extraction problem. We use a supervised learning approach based on dependency

¹<http://opennlp.apache.org/>

²<http://nlp.stanford.edu/ner/index.shtml>

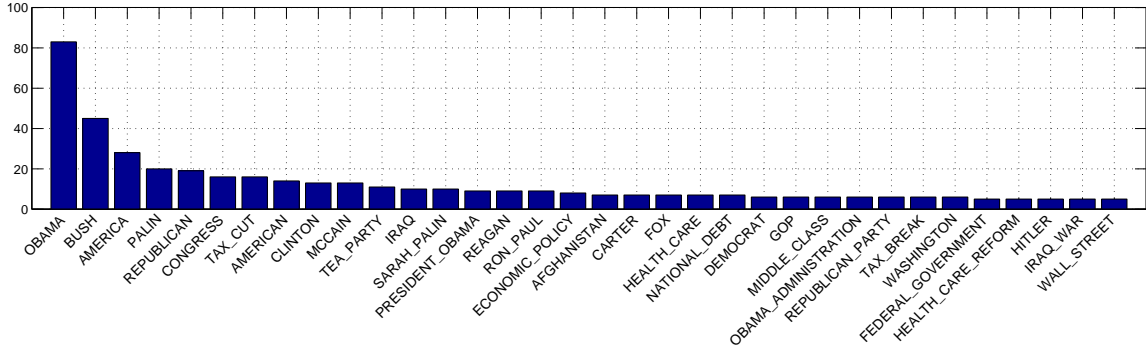


Figure 1: Salient aspects and number of users who express opinions on them in the thread “Will you vote for Obama?”

ID	Dependency path rule	Example
R1	$ADJ_{OP} \leftarrow amod \leftarrow N_{TR}$	I simply point out how <i>terrible</i> REPUBLICAN PARTY is.
R2	$ADJ_{OP} \rightarrow nsubj \rightarrow N_{TR}$	BUSH is even more <i>reasonable</i> for tax hike than Obama.
R3	$V_{OP} \rightarrow dobj \rightarrow N_{TR}$	I would never <i>support</i> OBAMA.
R4	$V_{OP} \rightarrow prep.* \rightarrow N_{TR}$	I'll <i>vote</i> for OBAMA.
R5	$V_{OP} \rightarrow nsubjpass \rightarrow N_{TR}$	DEMOCRATIC PARTY are ultimately <i>corrupted</i> by love of money.
R6	$N_{OP} \leftarrow dobj \leftarrow V \rightarrow nsubj \rightarrow N_{TR}$	PAKISTAN is increasing terrorist <i>threat</i> .
R7	$ADJ_{OP} \leftarrow amod \leftarrow N \rightarrow nsubj \rightarrow N_{TR}$	OBAMA was a <i>top</i> scorer for occidental college.
R8	$ADV_{OP} \leftarrow advmod \leftarrow V \rightarrow nsubj \rightarrow N_{TR}$	OBAMA is <i>smarter</i> than people.

Table 1: Examples of frequent dependency path rules in our training data. OP and TR refer to the opinion and the target. The opinion words are in italic and the aspect words are in uppercase.

paths. Previous work by Mintz et al. (2009), and Qiu et al. (2009) has shown that the shortest path between a candidate opinion aspect and a candidate opinion expression in the dependency parse tree can be effective in extracting opinion relations. We use the Stanford Parser from Klein and Manning (2003) to obtain the dependency parse trees for each sentence in the posts and then get the dependency paths between each pair of candidate aspect and opinion expression. We use dependency relations and POS tags of nodes along the path to represent a dependency path. Given a set of training sentences (we use the one from Wu et al. (2009)), we can get a set of dependency path rules based on their frequencies in the training data. Table 1 shows the frequent dependency path rules in our training data.

When a pair of aspect and opinion expression is identified to be related, we use the polarity of the opinion expression to label the relation. Finally, given a pair of users, we use the percentage of positive interactions between them over all subjective interactions (i.e. interactions with either positive or negative opinions) as extracted from their exchanged posts as the sentiment polarity score between the

two users, regardless of the reply-to direction of the posts. Similarly, given a user and an aspect, we also use the percentage of positive opinion relations extracted as the sentiment polarity score between them. Thus the user-user opinion matrix and the user-aspect opinion matrix are constructed. If there is no subjective interaction detected between two users or between a user and an aspect, the corresponding entry in the matrix is left empty. We will see later that empty entries in the matrices are not used in the probabilistic matrix factorization step.

5 Probabilistic Matrix Factorization

As we have pointed out earlier, a problem with the matrices extracted as described in Section 4 is that the matrices are sparse, i.e. many entries are empty. For the data set we use, we find that around 87% of entries in the user-user opinion matrix and around 90% of entries in the user-aspect opinion matrix are empty. In this section, we describe how we use Probabilistic Matrix Factorization (PMF) to represent users and aspects in a latent factor space and thus generalize the user preferences.

Our model is almost a direct application of proba-

bilistic matrix factorization from Salakhutdinov and Mnih (2008), originally proposed for recommender systems. The main difference is that the user-user opinion polarity scores are symmetric. Our model is also similar to the one used by Ma et al. (2008). We describe our model as follows.

We assume that there are K latent factors with which both users and aspects can be represented. Let $u_i \in \mathbb{R}^K$ denote the vector in the latent factor space for the i -th user, and a_k the vector for the k -th aspect.

Recall that the opinions extracted from posts between users are represented by a user-user opinion matrix S , and the opinions held by different users on the various topic aspects are represented by a user-aspect opinion matrix R . We assume that the polarity scores $s_{i,j}$ between the i -th and the j -th users and $r_{i,k}$ between the i -th user and the k -th aspect in the two matrices S and R are generated in the following way:

$$\begin{aligned} p(s_{i,j}|u_i, u_j, \sigma_1^2) &= \mathcal{N}(s_{i,j}|g(u_i^T u_j), \sigma_1^2), \\ p(r_{i,k}|u_i, a_k, \sigma_2^2) &= \mathcal{N}(r_{i,k}|g(u_i^T a_k), \sigma_2^2), \end{aligned}$$

where σ_1^2 and σ_2^2 are variance parameters, $g(\cdot)$ the logistic function, and $\mathcal{N}(\cdot|\mu, \sigma^2)$ is the normal distribution with mean μ and variance σ^2 .

We can see that with this generative assumption, if two users are similar in terms of their dot product in the latent factor space, then they are more likely to have positive interactions as extracted from their textual exchanges. Similarly, if a user and an aspect are similar, then the user is more likely to express a positive opinion on the aspect in his/her posts. The latent factors can therefore encode user preferences and similarity between two users in the latent factor space reflects whether they share similar viewpoints.

We also place the following prior over u_i and a_k :

$$\begin{aligned} p(u_i|\sigma_U^2) &= \mathcal{N}(u_i|\vec{0}, \sigma_U^2 \mathbf{I}), \\ p(a_k|\sigma_A^2) &= \mathcal{N}(a_k|\vec{0}, \sigma_A^2 \mathbf{I}), \end{aligned}$$

where σ_U^2 and σ_A^2 are two variance parameters for users and aspects, respectively, and \mathbf{I} is the identity matrix.

Figure 2 shows the plate notation for the generative model.

Let \mathcal{U} be a $K \times U$ matrix containing the vectors u_i for all U users, and \mathcal{A} be an $K \times A$ matrix containing

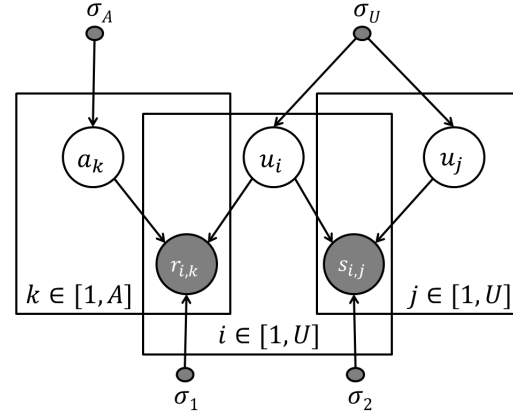


Figure 2: Probabilistic matrix factorization model on opinion matrices.

the vectors a_k for all A aspects. To automatically learn \mathcal{U} and \mathcal{A} , we minimize the following objective function:

$$\begin{aligned} \mathcal{L}(\mathcal{U}, \mathcal{A}, \mathcal{S}, \mathcal{R}) &= \frac{1}{2} \sum_{i=1}^U \sum_{k=1}^A \mathbb{I}(r_{i,k}) (r_{i,k} - g(u_i^T a_k))^2 \\ &+ \frac{\lambda_1}{2} \sum_{i=1}^U \sum_{j=1}^U \mathbb{I}(s_{i,j}) (s_{i,j} - g(u_i^T u_j))^2 \\ &+ \frac{\lambda_U}{2} \|\mathcal{U}\|_F^2 + \frac{\lambda_A}{2} \|\mathcal{A}\|_F^2, \end{aligned} \quad (1)$$

where $\lambda = \frac{\sigma_1^2}{\sigma_2^2}$, $\lambda_U = \frac{\sigma_1^2}{\sigma_U^2}$, and $\lambda_A = \frac{\sigma_1^2}{\sigma_A^2}$, $\mathbb{I}(s)$ is an indicator function which equals 1 when s is not empty and otherwise 0.

To optimize the objective function above, we can perform gradient descent on \mathcal{U} and \mathcal{A} to find a local optimum point. The derivation is similar to Ma et al. (2008).

Degenerate Versions of the Model

We refer to the complete model described above as PMF-UOM (PMF model based on User Opinion Matrices). PMF-UOM has the following two degenerate versions by considering either only the user-user opinion matrix or only the user-aspect opinion matrix.

PMF-UU: In this degenerate version of the model, we use only the user-user opinion matrix to learn the latent factor representation. Specifically, the objective function is modified such that we drop the sum

of the square errors involving R and the regularizer on \mathcal{A} .

PMF-UA: In this degenerate version of the model, we use only the user-aspect opinion matrix to learn the latent factor representation. Specifically, the objective function is modified such that we drop the sum of the square errors involving S .

6 Experiments

In this section, we present our experiments that evaluate our model.

6.1 Data Set and Experiment Settings

The data set we use comes from Abu-Jbara et al. (2012) and Hassan et al. (2012). The data set contains a set of discussion threads collected from two political forums (Createdebate³ and Politicalforum⁴) and one Wikipedia discussion session. We randomly select 6 threads from the original data set to evaluate our model. Some details of the data we use are listed in Table 2.

ID	topic	#sides	#sentences	#users
DS1	Vote for Obama	2	12492	197
DS2	Abortion Banned	6	3844	70
DS3	Profile Muslims	4	2167	69
DS4	England and USA	6	2030	62
DS5	Tax Cuts	2	1193	26
DS6	Political Spectrum	7	1130	50

Table 2: Some statistics of the data sets.

In our experiments, for the PMF-based methods, we set the number of latent factors to be 10 as we do not observe big difference when vary the latent factor size from 10 to 50. For the other parameters, we select the optimal setting for each thread based on the average of 50 runs. λ_U is chosen from $\{0.1, 0.01\}$, λ_A from $\{0.01, 0.001\}$ and λ from $\{1, 0.1\}$.

6.2 Relation Polarity Prediction

The first task we use to evaluate our model is to predict the polarity of interactions between two users. Different from Hassan et al. (2012), however, we are not using this task to evaluate the accuracy of sentiment analysis from text. Our experimental setting is completely different in that we do not make

use of the text exchanges between the two users but instead use their interactions with other users or aspects. The purpose is to test the effectiveness of collaborative filtering.

Experimental Setting: The experiments are set up in the following way. Given a pair of users i and j who have directly exchanged posts, i.e. $s_{i,j}$ is not empty, we first hide the value of $s_{i,j}$ in the matrix S . Let the altered matrix be $S_{-(i,j)}$. We then use $S_{-(i,j)}$ instead of S in the learning process as described in Section 5 to learn the latent factor representation. Let \hat{u}_i and \hat{u}_j denote the learned latent vectors for user i and user j . We predict the polarity of relation between i and j as follows:

$$\hat{s}_{i,j} = \begin{cases} 1 & \text{if } g(\hat{u}_i^T \hat{u}_j) > 0.5, \\ 0 & \text{otherwise,} \end{cases}$$

where $g(\cdot)$ is the logistic function to convert the dot product into a value between 0 and 1.

To judge the quality of the predicted polarity $\hat{s}_{i,j}$, we could compare it with $s_{i,j}$. But since $s_{i,j}$ itself is predicted from the textual exchanges between i and j , it is not the ground truth. Instead, we ask two human annotators to assign the true polarity label for user i and user j by reading the textual exchanges between them and judging whether they are friends or foes in the discussion thread. The annotators are asked to assign a score of 0 (indicating a negative relation), 0.5 (indicating a neutral relation) or 1 (indicating a positive relation). The lowest agreement score based on Cohen’s kappa coefficient among the 6 threads we use is 0.56, showing fair to good agreement. As ground truth, we set the final polarity score to 1 if the average score of the two annotators is larger than 0.5 and 0 otherwise.

We compare the PMF-based methods with two majority baselines: MBL-0 always predicts negative relations for all the user pairs (assuming most relations are negative) and MBL-1 always predicts positive relations (assuming most relations are positive).

We use *MAE* (mean absolute error) and *RMSE* (root mean square error) as defined below as performance metrics:

$$MAE = \frac{\sum_{i,j} |\hat{s}_{i,j} - l_{i,j}|}{N},$$

$$RMSE = \sqrt{\frac{\sum_{i,j} (\hat{s}_{i,j} - l_{i,j})^2}{N}},$$

³www.createdebate.com

⁴www.politicalforum.com

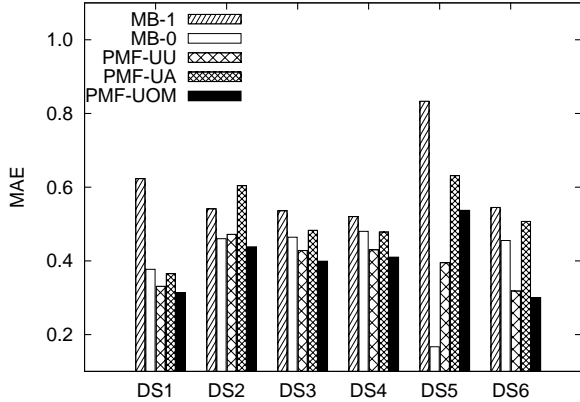


Figure 3: Comparing all the methods in terms of MAE.

where N is the total number of user pairs we test, and $l_{i,j}$ is the ground truth polarity score between user i and user j .

Results: We show the results of our model and of PMF-UU and PMF-UA in terms of MAE in Figure 3 and RMSE in Figure 4. The MAE values range between 0.31 and 0.44 except for DS5, which has a higher error rate of 0.53. The results show that even without knowing the textual exchanges between two users, from their interactions with other users and/or with topic aspects, we can still infer the polarity of their relation with decent accuracy most of the time.

The results also show the comparison between our model and the competing methods. We can see that overall the complete model (PMF-UOM) performs better than the two degenerate models (PMF-UU and PMF-UA). The differences are statistically significant at the 5% level without considering DS5, as indicated by a 2-tailed paired t-test. Comparing to the majority baselines, our model significantly outperforms MBL-1 at 1% significance level while outperforms MBL-0 on all the data sets except DS5. A close examination shows DS5 has very unbalanced relations (around 83% of relations are negative). Except for the unbalanced data set, our model has reasonably good performance.

6.3 Subgroup Detection

The second task we study is the problem of detecting ideological subgroups from discussion threads. The original data set has been labeled with the ground truth for this problem, that is, for each thread the

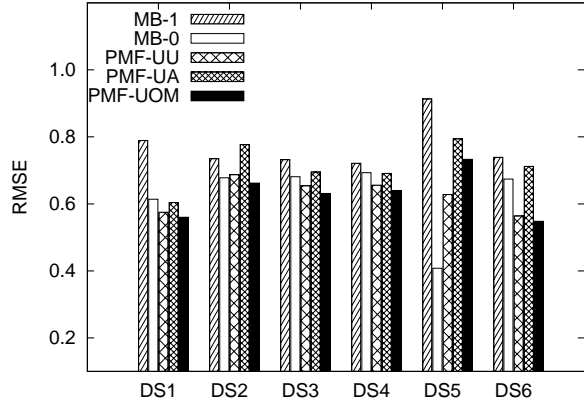


Figure 4: Comparing all the methods in terms of RMSE.

number of viewpoints is known and the viewpoint held by each user is labeled. A subgroup is defined as a set of users holding the same viewpoint.

Experimental Setting: Through this second experiment, we would like to verify the hypothesis that using the learned latent factor representation \mathcal{U} for users, we can better detect subgroups than directly using the opinion matrices S and R . For all the methods we compare, we first construct a feature vector representation for each user. We then apply K -means clustering to group users. The number of clusters is set to be the true number of viewpoints for each thread. The different methods are described below:

- **PMF-based methods:** We simply use the learned latent vectors \hat{u}_i after optimizing the objective function as the feature vectors to represent each user.
- **BL-1:** This is our own implementation to simulate the method by Abu-Jbara et al. (2012). Here each user is represented by a $(3 \times (U + A))$ -dimensional vector, where U is the number of users and A is the number of aspects, i.e. $(U + A)$ is the total number of opinion targets. For each opinion target, there are 3 dimensions in the feature vector, corresponding to the number of positive, neutral and negative opinion expressions towards the target from the online posts.
- **BL-2:** BL-2 is similar to BL-1 except that we only use a $(U + A)$ -dimensional vector to repre-

sent each user. Here for each opinion target, we directly use the corresponding sentiment polarity score $s_{i,j}$ or $r_{i,j}$ from the matrix S or R . For empty entries in S and R , we use a score of 0.5.

We use *Purity* (the higher the better), *Entropy* (the lower the better) and *Rand Index* (the higher the better) to evaluate the performance of subgroup detection (Manning et al., 2008). We further use *Accuracy* obtained by choosing the best alignment of clusters with the ground truth class labels and computing the percentage of users that are “classified” correctly.

Results: We first give an overview of the performance of all the methods on the task. We show the average performance of the methods on all the data sets in Figure 5. Overall, our model has a better performance than all the competing methods.

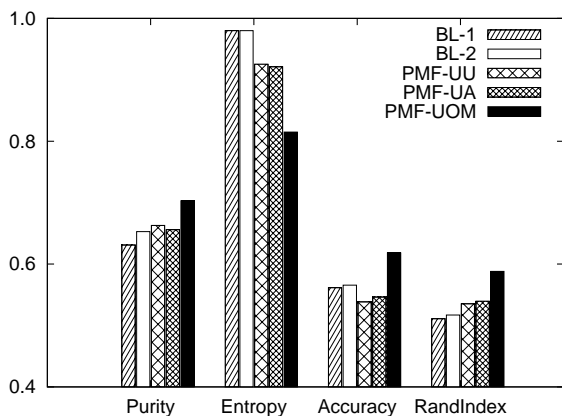


Figure 5: An overview of the average performance of all the methods on the 6 threads.

We present all the results in Table 3. We perform 2-tailed paired t-test on the results. We find that PMF-UOM outperforms all the other methods in terms of RandIndex at 5% significance level and outperforms other methods in terms of Purity and Entropy at 10% significance level. Furthermore, the PMF-UOM model outperforms its degenerative models PMF-UU and PMF-UA at 10% significance level in terms of all the measures.

We observe that PMF-UOM achieves the best performance in terms of all the measures for almost all threads. In particular, comparison with BL-1 and BL-2 shows that collaborative filtering can generalize the user preferences and help better group the users based on their viewpoints. The fact that

PMF-UOM outperforms both PMF-UU and PMF-UA shows that it is important to consider both user-user interactions and user-aspect interactions.

The Effects of Cluster Size: To test the effect of the number of clusters on the experiment result, we vary the number of clusters from 2 to 10 in all methods. We find that all methods tend to achieve better results when the number of clusters equals the ground truth cluster size. Overall, our method PMF-UOM shows a better performance than the other four methods when the number of clusters changes, which indicates the robustness of our method.

		BL-1	BL-2	PMF-UU	PMF-UA	PMF-UOM
DS1	<i>P</i>	0.61	0.61	0.61	0.61	0.62
	<i>E</i>	0.96	0.96	0.94	0.95	0.94
	<i>A</i>	0.59	0.59	0.55	0.57	0.60
	<i>R</i>	0.51	0.51	0.50	0.51	0.52
DS2	<i>P</i>	0.53	0.63	0.64	0.61	0.68
	<i>E</i>	1.17	1.22	1.14	1.09	0.99
	<i>A</i>	0.47	0.53	0.48	0.47	0.50
	<i>R</i>	0.50	0.50	0.56	0.56	0.58
DS3	<i>P</i>	0.66	0.68	0.62	0.60	0.68
	<i>E</i>	1.05	1.01	1.06	1.07	0.94
	<i>A</i>	0.61	0.63	0.48	0.47	0.58
	<i>R</i>	0.50	0.52	0.53	0.53	0.57
DS4	<i>P</i>	0.64	0.64	0.66	0.65	0.70
	<i>E</i>	0.92	0.94	0.90	0.91	0.85
	<i>A</i>	0.59	0.64	0.62	0.62	0.68
	<i>R</i>	0.49	0.52	0.52	0.51	0.56
DS5	<i>P</i>	0.86	0.86	0.86	0.86	0.86
	<i>E</i>	0.56	0.56	0.49	0.48	0.38
	<i>A</i>	0.70	0.70	0.57	0.60	0.71
	<i>R</i>	0.52	0.52	0.43	0.45	0.56
DS6	<i>P</i>	0.50	0.50	0.60	0.60	0.68
	<i>E</i>	1.35	1.35	1.03	1.04	0.79
	<i>A</i>	0.40	0.30	0.53	0.54	0.64
	<i>R</i>	0.53	0.53	0.68	0.68	0.74

Table 3: Results on subgroup detection on all the 6 threads. *P*, *E*, *A* and *R* refer to *Purity*, *Entropy*, *Accuracy* and *RandIndex*, respectively.

7 Conclusions

In this paper, we studied how to use probabilistic matrix factorization, a common technique for collaborative filtering, to improve relation mining from online discussion forums. We first applied sentiment analysis to extract user-user opinions and user-aspect opinions from forum posts. The extracted opinions form two opinion matrices. We then applied probabilistic matrix factorization using these

two matrices to discover a low-rank latent factor space which aims to better generalize the users' underlying preferences and indicate user similarities based on their viewpoints. Using a data set with 6 discussion threads, we showed that the learned latent vectors can be used to predict the polarity of user relations well without using the users' direct interaction data, demonstrating the effectiveness of collaborative filtering. We further found that for the task of subgroup detection, the latent vectors gave better performance than using the directly extracted opinion data, again showing that collaborative filtering through probabilistic matrix factorization can help address the sparseness problem in the extracted opinion matrices and help improve relation mining.

Our current work mainly focuses on the user opinion matrices. As future work, we would like to explore how to incorporate textual contents without opinionated expressions. One possible way is to consider the combination of matrix factorization and topic modeling as studied by Wang and Blei (2011) where we can use topic modeling to study textual contents.

Acknowledgments

We thank the reviewers for their valuable comments on this work.

References

- Amjad Abu-Jbara, Pradeep Dasigi, Mona Diab, and Dragomir R. Radev. 2012. Subgroup detection in ideological discussions. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 399–409.
- Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*.
- Yejin Choi, Eric Breck, and Claire Cardie. 2006. Joint extraction of entities and relations for opinion recognition. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, EMNLP '06*, pages 431–439, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Pradeep Dasigi, Weiwei Guo, and Mona T. Diab. 2012. Genre independent subgroup detection in online discussion threads: A study of implicit attitude using textual latent semantics. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 65–69.
- Ahmed Hassan and Dragomir Radev. 2010. Identifying text polarity using random walks. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, ACL '10*, pages 395–403, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012. Detecting subgroups in online discussions by modeling positive and negative relations among participants. In *Proceedings of the 2012 EMNLP*, pages 59–70.
- John Hopcroft, Tiancheng Lou, and Jie Tang. 2011. Who will follow you back?: reciprocal relationship prediction. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 1137–1146.
- Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.
- Noriaki Kawamae. 2011. Predicting future reviews: sentiment analysis models for collaborative filtering. In *Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11*, pages 605–614.
- Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1, ACL '03*, pages 423–430, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Predicting positive and negative links in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 641–650.
- Cane Wing-Ki Leung, Stephen Chi-Fai Chan, Fu-Lai Chung, and Grace Ngai. 2011. A probabilistic rating inference framework for mining user preferences from reviews. *World Wide Web*, 14(2):187–215.
- David Liben-Nowell and Jon Kleinberg. 2003. The link prediction problem for social networks. In *Proceedings of the twelfth international conference on Information and knowledge management*.
- Hao Ma, Haixuan Yang, Michael R. Lyu, and Irwin King. 2008. Sorec: Social recommendation using probabilistic matrix factorization. In *Proc. of ACM international conference on Information and knowledge management*.
- Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, July.

- George A. Miller. 1995. Wordnet: A lexical database for english. *Communications of the ACM*, Vol. 38, No. 11:39–41.
- Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. 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 2 - Volume 2*, ACL '09, pages 1003–1011, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Yashar Moshfeghi, Benjamin Piwowarski, and Jon M. Jose. 2011. Handling data sparsity in collaborative filtering using emotion and semantic based features. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '11, pages 625–634.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, January.
- Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 339–346, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009. Expanding domain sentiment lexicon through double propagation. In *Proceedings of the 21st international joint conference on Artificial intelligence*, IJCAI'09, pages 1199–1204, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Ruslan Salakhutdinov and Andriy Mnih. 2008. Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems*, volume 20.
- Chong Wang and David M. Blei. 2011. Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 448–456.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT/EMNLP*.
- Yuanbin Wu, Qi Zhang, Xuanjing Huang, and Lide Wu. 2009. Phrase dependency parsing for opinion mining. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3 - Volume 3*, EMNLP '09, pages 1533–1541, Stroudsburg, PA, USA. Association for Computational Linguistics.