Social Collaborative Viewpoint Regression with Explainable Recommendations

Zhaochun Ren‡ zhaochun.ren@ucl.ac.uk
Shangsong Liang‡ shangsong.liang@ucl.ac.uk
Piji Li§ pjli@se.cuhk.edu.hk
Shuaiqiang Wang† shuaiqiang.wang@jyu.fi
Maarten de Rijke^II derijke@uva.nl

‡University College London, London, United Kingdom
§The Chinese University of Hong Kong, Hong Kong, China
†University of Jyväskylä, Jyväskylä, Finland
^IIUniversity of Amsterdam, Amsterdam, The Netherlands

ABSTRACT

A recommendation is called explainable if it not only predicts a numerical rating for an item, but also generates explanations for users' preferences. Most existing methods for explainable recommendation apply topic models to analyze user reviews to provide descriptions along with the recommendations they produce. So far, such methods have neglected user opinions and influences from social relations as a source of information for recommendations, even though these are known to improve the rating prediction.

In this paper, we propose a latent variable model, called social collaborative viewpoint regression (sCVR), for predicting item ratings based on user opinions and social relations. To this end, we use so-called viewpoints, represented as tuples of a concept, topic, and a sentiment label from both user reviews and trusted social relations. In addition, such viewpoints can be used as explanations. We apply a Gibbs EM sampler to infer posterior distributions of sCVR. Experiments conducted on three large benchmark datasets show the effectiveness of our proposed method for predicting item ratings and for generating explanations.

Keywords

Recommender systems; User comment analysis; Topic modeling; Trusted social relations

1. INTRODUCTION

Recommender systems are playing an increasingly important role in e-commerce portals. With the development of social networks, many e-commerce sites have become popular social platforms that help users discuss and select items. Traditionally, a major strategy to predicting ratings in recommender systems is based on collaborative filtering (CF), which infers a user's preference using their previous interactions. Since CF-based methods only use numerical ratings as input, they suffer from a "cold-start" problem and unexplainable prediction results [12, 22].

Explainable recommendations have been proposed to address the "cold-start" problem and the poor interpretability of recommended results by not only predicting better rating results, but also generating explainable and understandable item aspects that attract a user's attention [54]. Most current solutions for explainable recommendations are based on content-based analysis methods [7, 22, 48]. Recent work on explainable recommender systems applies topic models to predict ratings and topical explanations [10, 22], where latent topics are inferred from user reviews. Each latent topic in a topic model is represented as a set of words, whereas each item is represented as a set of latent topics. These approaches face two important challenges: (1) Most existing methods neglect to explicitly analyze opinions for recommendation, thereby missing important opportunities to explain users' preferences. (2) Trusted social relations are known to improve the quality of CF recommendation [15, 50], however, current methods for explainable recommendations rarely use this information.

To improve the rating prediction for explainable recommendations, our focus is on developing methods to generate so-called viewpoints by jointly analyzing user reviews and trusted social relations. Centered around a concept, a *viewpoint* in the setting of recommender systems refers to a topic with a specific sentiment label. As an example, consider the concept "Chinese cuisine" and the topic "#kung pao chicken" with a positive sentiment. Compared to "topics" in previous explainable recommendation strategies [6, 48, 49], viewpoints contain richer information that can be used to understand and predict user ratings in recommendation tasks. We assume that each item and user in a recommender system can be represented as a finite mixture of viewpoints, and each user's viewpoints can be influenced by their trusted social relations. In Fig. 1 we show an example with multiple viewpoints, user reviews, trusted social relations, and ratings in a recommender system.

Three technical issues need to be addressed before viewpoints can successfully be used for explainable recommendations that make use of social relations: (1) the shortness and sparseness of reviews

WSDM 2017, February 06 - 10, 2017, Cambridge, United Kingdom

ISBN 978-1-4503-4675-7/17/02…$15.00

DOI: http://dx.doi.org/10.1145/3018661.3018686

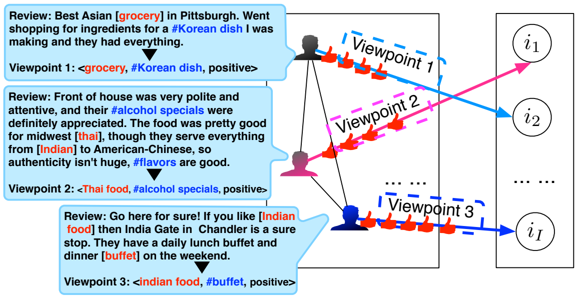


Figure 1: An example of trusted social relations, user reviews and ratings in a recommender system. Black arrows connect users with trusted social relations. "ThumpUp" symbols reflect the ratings of items. Concepts and topics have been highlighted in red and blue, respectively. Three viewpoints are represented in three different colors. A viewpoint is a mixture over a concept, a topic, and a sentiment (explained in Section 3).

make viewpoint extraction difficult; (2) because of the "bag of words" assumption, traditional topic models do not necessarily work very well in opinion analysis; (3) inferring explicit viewpoint statistics given trusted social relations among users and user reviews is not a solved problem.

In this paper, we address the above issues. According to our previous work on viewpoint modeling [31], we propose a latent variable model, called the *social collaborative viewpoint regression model* (abbreviated as **sCVR**), to predict user ratings by discovering viewpoints. Unlike previous collaborative topic regression methods [48], sCVR predicts ratings by detecting viewpoints from user reviews and social relations. sCVR discovers concepts, topics and sentiment priors from user reviews. It employs Markov chains to capture the sentiment dependency between two adjacent words; given trusted social relations, in sCVR we assign a viewpoint-bias to each user by considering the social influence of their trusted social relations. That is, given a user and an item, sCVR detects viewpoints and predicts ratings by jointly generating concepts, topics and sentiment labels in user reviews. Gibbs EM sampling is applied to approximate the posterior probability distributions.

We use three real-world benchmark datasets in our experiments: *Yelp 2013*, *Yelp 2014*, and *Epinions*. Extensive experiments on these datasets show that sCVR outperforms state-of-the-art baselines in terms of MAE and RMSE metrics.

To sum up, our contributions in this paper are: (1) To improve rating prediction for explainable recommendations, we generate viewpoints from user reviews and trusted social relations. (2) We propose a latent variable model, the social collaborative viewpoint regression model, to predict user ratings by jointly modeling concepts, topics, sentiment labels and social relations. (3) We prove the effectiveness of our proposed model on three benchmark datasets through extensive experiments, in which our proposed method outperforms state-of-the-art baselines.

2. RELATED WORK

We detail relevant related work on collaborative filtering, sentiment analysis, and solutions for explainable recommendations.

2.1 Collaborative filtering

In recent years, collaborative filtering (CF) based techniques have received considerable attention. Unlike content-based filtering strategies [23] that predict ratings using the analysis of user profiles, collaborative filtering [39] methods, either memory-based CF or model-based CF, predict ratings using user-item ratings matrices.

Early CF-based methods apply memory-based techniques. The most widely used memory-based CF methods include the nearest neighbor approach [1], user-based methods [32] and item-based methods [34]. Among the model-based CF methods, latent factor models [19] have become very popular as they show state-of-the-art performance on multiple datasets. Aimed at factorizing a rating matrix into products of a user-specific matrix and an item-specific matrix, matrix factorization based methods [19, 20, 28] are widely used. Zhang et al. [53] propose a localized matrix factorization approach to address data sparsity and scalability by factorizing block diagonal form matrices. Recently, ranking-oriented collaborative filtering algorithms have achieved good results: Shi et al. [37] use a list-wise learning to rank method, called CliMF. Following the memory-based CF framework, ListCF directly predicts a total order of items for each user based on similar users' probability distributions over permutations of commonly rated items [14].

Trusted social relations have been applied to enhance the performance of collaborative filtering [7, 18, 35]. Sharma and Cosley [35] analyze the influence of varying levels of social information on users' decisions. In social media recommendation tasks [15, 24, 50], social relations have been considered. Konstas et al. [18] adapt a random walk framework to take into account both the social annotation and friendships inherent in the social graph. Yang et al. [51] address recommendation and link prediction tasks based on a joint-propagation model between social relations and interests. Ye et al. [52] propose a generative model to describe users' behavior, given influences from social communities. Chen et al. [8] propose a collaborative filtering method to generate personalized recommendations on Twitter through a collaborative ranking procedure.

Unlike our proposed method, most previous work on collaborative filtering neglects the combination of content analysis for user reviews and social trust relationships. Furthermore, most previous work avoids mining the content of user reviews, i.e., viewpoint detection and analysis.

2.2 Sentiment analysis

In recent years, *sentiment analysis* has received a lot of attention. As a fundamental task in sentiment analysis, *sentiment classification* [45] is crucial to understand user generated content in product reviews. Lexicon-based methods [11, 43, 45] utilize a lexicon of sentiment words to predict sentiment labels, whereas corpus-based methods [16, 29, 55] classify sentences to sentiment polarities using corpora that are labeled with sentiment labels. With the recent success of deep neural networks [2, 13], more and more approaches to the sentiment classification task learn low-dimensional feature vectors, e.g., Tang et al. [40] propose a sentiment-specific word embedding method for short text sentiment classification in social media. For encoding relations between sentences in documents, a recurrent neural network has been proposed to learn representations of documents for sentiment classification [41]. By taking user information into account, Tang et al. [42] present a user-vector composition model to predict user ratings. Given a set of opinionated documents, Ren and de Rijke [30] address the task of summarizing contrastive themes by selecting meaningful sentences to represent contrastive themes.

In this paper, we focus on a combination of content-based recommendation and collaborative filtering, in which we not only predict sentiment labels in each review, but also consider topic aspects, user ratings and social trust communities in recommendation.

2.3 Explainable recommendations

The "cold-start" problem and poor interpretability are two serious issues for traditional collaborative filtering methods. To address these two issues, in recent years, more and more researchers have started to consider explainable recommendation. Explanations of recommendations often take the form of (labels derived from) a topic model [5, 33, 44, 54]. We follow this tradition but propose a richer notion of explanation: represented by tuples of a conceptual feature, a topic and a sentiment label, viewpoints are used to explain our results. Explainable recommendations are known to improve transparency, user trust, effectiveness and scrutability [44]. Vig et al. [46] propose an explainable recommendation method that uses community tags to generate explanations. Based on sentiment lexicon construction, the explicit factor models [54] and Tri-Rank [12] algorithms improve the ranking of items for review-aware recommendation. By combing content-based recommendation and collaborative filtering, Wang and Blei [48] apply topic models [6] to explainable recommendation problem to discover explainable latent factors in probabilistic matrix factorization. Chen et al. [7] take advantage of the social trust relations by proposing a hierarchical Bayesian model that considers social relationship by putting different priors on users.

Recent work on explainable recommendations focuses on user reviews. Diao et al. [10] propose a hybrid latent factor model integrating user reviews, topic aspects and user ratings for collaborative filtering. By using a multi-dimension tensor factorization strategy, Bhargava et al. [4] propose a recommendation approach by combining users, activities, timestamps and locations. The Hidden Factors as Topic model learns a topic model for items using the review text and a matrix factorization model to fit the ratings [25]. To tackle the sparsity in collaborative topic filtering, the Ratings Meet Reviews model has been proposed by adopting a mixture of Gaussians, which is assumed to have the same distribution as the topic distribution, to model ratings [22]. Ribeiro et al. [33] explain the predictions of classifiers in an interpretable manner and select representative predictions to users.

To the best of our knowledge, there is little previous work on explainable recommendation that jointly considers using user reviews and trusted social relations to improve the rating prediction: Represented by tuples of a conceptual feature, a topic and a sentiment label, viewpoints are used to explain our results.

3. PRELIMINARIES

Before introducing our social collaborative viewpoint regression model for explainable recommendations, we introduce our notation and key concepts. Table 1 lists the main notation we use.

Similar to the Ratings Meet Reviews model (RMR) [22], we assume that there are U users $\mathcal{U} = \{u_1, u_2, \ldots, u_U\}$; I items $\mathcal{I} = \{i_1, i_2, \ldots, i_I\}$; a set of observed indices $\mathcal{Q} = \{(u, i)\}$, where each pair $(u, i) \in \mathcal{U} \times \mathcal{I}$ indicates an observed rating $r_{u,i}$ with a user review $d_{u,i}$ from user u of item i. For user reviews $\mathcal{D} = \{d_1, d_2, \ldots, d_{|\mathcal{Q}|}\}$, we assume that each observed rating $r_{u,i}$ is associated with a user review $d_{u,i}$. Given an item i's reviews $\mathcal{D}_i$, each review $d \in \mathcal{D}_i$ is represented as a set of words, i.e., $d = \{w_1, w_2, \ldots, w_{|d|}\}$. If two users u_i and u_j trust each other, as evidenced in a user community, we define them to be a *trusted social relation* or simply *social relation* with trust value $\mathcal{T}_{u_i,u_j}$.

Next, we define the notions of topic, concept and sentiment. Following the standard definition [6], we define a *topic*, denoted as z, as a probabilistic distribution over words. Assuming K topics exist in the user reviews on which we focus, we set $z \in \{1, 2, \ldots, K\}$. We define a *concept*, denoted as e, as a rigid designator of a feature surrounded by a topic, e.g., the concept "French cuisine" with a topic about "tartare boeuf." Each user review is assumed to contain a concept $e_d \in \mathcal{E}$. *Sentiment* is defined as a probability distribution over the sentiment labels positive and negative. A *sentiment label* l_j is attached to each word w_j. Following [21], we assume that the sentiment label l_j for a word w_j depends on the topic z_j. Specifically, we set $l_j = -1$ when the word w_j is "negative," while $l_j = 1$ when w_j is "positive." Given a concept e, a topic z and sentiment label l, we define a *viewpoint* to be a finite mixture over triples $v = \langle e, z, l \rangle$ consisting of a concept, topic and sentiment.

Table 1: Glossary.

Symbol	Description
$\mathcal{I}$	candidate items
$\mathcal{U}$	candidate users
$\mathcal{D}$	user reviews
$\mathcal{N}$	vocabulary in review corpus $\mathcal{D}$
$\mathcal{T}$	trust values among users
$\mathcal{R}$	user ratings
$\mathcal{V}$	viewpoints set
$\mathcal{E}$	concepts set
$\mathcal{Z}$	topics set in $\mathcal{Z}$
$\mathcal{Q}$	observed indices
u	a user, $u \in \mathcal{U}$
i	an item, $i \in \mathcal{I}$
d	a review, $d \in \mathcal{D}$
v_d	a viewpoint in review d, $v_d \in \mathcal{V}$
e_d	a concept in review d, $e_d \in \mathcal{E}$
w_j	the j-th word present in a review, $w_j \in \mathcal{N}$
z_j	a topic present in word w_j, $z_j \in \mathcal{Z}$
l_j	a sentiment label present in word w_j
f_u	a viewpoint selected by user u
$r_{u,i}$	the rating value from user u to item i
π	distribution of viewpoints
θ_v^u	distribution of viewpoint v for user u
λ	distribution of concepts over viewpoints
μ	distribution of topics over viewpoints
$\phi_{v,z,l}$	distribution of words over v, z and l

Because user reviews are short, we assume that only one viewpoint v_d, represented as a combination of a concept e, a topic z and a sentiment label l, exists in each user review $d \in \mathcal{D}$. We assume that each item $i \in \mathcal{I}$ can be represented as a mixture over viewpoints, thus we set π_i to be a probability distribution of viewpoints in item i, μ to be a probability distribution of topics over viewpoints and λ to be a probability distribution of conceptual features over viewpoints. For words in user reviews, we set ϕ to be a probability distribution over viewpoints, topics and sentiment labels, which is derived from a Dirichlet distribution over hyper-parameter β.

It is common that rating scores are discrete [3, 49]. Importantly, unlike much previous work that predicts a decimal rating score given a user and an item, we apply a probabilistic rating distribution within the exponential family to provide more information to reflect users' rating habits, inspired by [3]. For each user $u \in \mathcal{U}$, we assume that u's ratings in a recommender system can be predicted by their viewpoint distribution over rating values, i.e., $\theta^u = \{\theta^u_{v_1}, \theta^u_{v_2}, \ldots, \theta^u_{v_V}\}$. Given a viewpoint $v \in \mathcal{V}$, $\theta^u_v \in \theta^u$ refers to a probabilistic distribution over each rating value $r \in [1, R]$, thus θ^u can be represented as an R-by-V matrix:

$$\theta^u = \begin{pmatrix} \theta^u_{1,v_1} & \cdots & \theta^u_{1,v_V} \\ \vdots & \ddots & \vdots \\ \theta^u_{R,v_1} & \cdots & \theta^u_{R,v_V} \end{pmatrix} \tag{1}$$

where each item $\theta^u_{r,v}$ denotes the probability of rating value r given user u and viewpoint v.

We assume that the viewpoint distribution θ_v^u is derived by a finite mixture over a personalized base distribution $\theta_{u,v}^0$ and viewpoint distributions of u's trusted relations. Given a user u and an item i, we set a multinomial distribution $f_{u,i}$, which derives from the viewpoints distribution π_i for item i, to reflect the viewpoint chosen by u for their rating to item i. If a user u writes a user review $d_{u,i}$ for item i, there is a corresponding rating $r_{u,i} \in [1, R]$ derived from a multinomial distribution over $\theta_{f_{u,i}}^u$.

Given observed indices $\mathcal{Q}$, observed data $\mathcal{R}$, $\mathcal{D}$ and $\mathcal{E}$, our target is to infer the user's viewpoint distribution θ and the item's viewpoint distribution π, which are then used to predict unknown ratings.

4. METHOD

In this section, we propose our *social collaborative viewpoint regression model*, abbreviated as sCVR. We start by detailing the model. We then describe our inference approach and explain our method to predict ratings using posterior distributions from sCVR.

4.1 Feature detection and sentiment analysis

We use descriptive keywords in an e-commerce platforms as concepts for items. Here we assume that E_i features exist in an item i's reviews. To discover the concept in a user review $d \in \mathcal{D}_i$, we employ word2vec [26] to calculate the similarity between a given concept $e \in \mathcal{E}_i$ and a user review d. Since the quality of the word vectors increases significantly with the amount of training data, we train a word2vec model using the latest Wikipedia data. Thereafter, we employ our trained model to predict the cosine similarity between a given concept e and each word w in a user review d. Given the cosine similarity $sim(e, w)$ between e and word w, $w \in d$, we calculate the similarity between e and review d according to Eq. 2:

$$sim(e, d) = \frac{1}{N_d} \sum_{w \in d} sim(e, w) \tag{2}$$

where N_d indicates the number of words in d. Given candidate concepts $\mathcal{E}_i$, the concept that is most similar to d will be considered as d's relevant concept. By ranking documents according to the similarity between candidate concepts and user reviews, we find the relevant concept for each user review.

We employ a state-of-the-art sentiment analysis method [38] to classify user reviews into positive and negative categories. The probability of a sentiment label is set as a prior value in our social collaborative viewpoint regression, which is detailed in §4.2.

4.2 Social collaborative viewpoint regression

Given observed indices $\mathcal{Q}$, users $\mathcal{U} = \{u_1, u_2, \ldots, u_U\}$, items $\mathcal{I} = \{i_1, i_2, \ldots, i_I\}$, ratings $\mathcal{R} = \{r_1, r_2, \ldots, r_Q\}$ and user reviews $\mathcal{D} = \{d_1, d_2, \ldots, d_Q\}$, our target is to infer distributions of viewpoints to predict unknown user ratings $\mathcal{Q}' = \{(u', i')\}$ from users to items, where $(u', i') \notin \mathcal{Q}$. To this end we propose *social collaborative viewpoint regression* (sCVR), a latent factor model. Unlike previous work, sCVR jointly models viewpoints, topics, concepts and sentiment labels in $\mathcal{D}$. In addition, sCVR explicitly models influences from a user's social relations on their own viewpoint distribution.

After preprocessing, for each user review $d \in \mathcal{D}$ we assume that there is a concept $e_d \in \mathcal{E}$, and for each word w in d there is a corresponding sentiment label l_w. We assume that there are, in total, V viewpoints and K topics in user reviews. Given an item $i \in \mathcal{I}$, we assume there is a probabilistic distribution π over viewpoints. Given a user review $d \in \mathcal{D}$, for each word $w_j \in d$, there is a topic z_j and a sentiment label l_j. We assume that a viewpoint v in d is derived via a multinomial distribution over a random variable π that indicates a probability distribution over viewpoints in each item; given viewpoint v, a concept e, a topic z and a sentiment label l are derived from probabilistic distributions over v. The probability distribution π is derived from a dirichlet mixture over a hyper parameter α.

Each user $u \in \mathcal{U}$ in sCVR is supposed to have F_u trusted social relations; each trusted relation u' shares a trust value $\mathcal{T}_{u,u'}$ with user u. For each user $u \in \mathcal{U}$, a probabilistic distribution over viewpoint v, θ_v^u is derived over viewpoint distributions of u's social relations and a base distribution of u, i.e., $\{\theta_v^{u_1}, \theta_v^{u_2}, \ldots, \theta_v^{u_{F_u}}\}$ and $\theta_{u,v}^0$. In sCVR we assume that u's rating $r_{u,i}$ for an item $i \in \mathcal{I}$ is derived from a multinomial distribution over θ_f^u, where f is a sampling viewpoint index derived from u's reviews, i.e., $f \in [1, V]$.

In sCVR, we consider the sentiment dependency between two adjacent words. A Markov chain is formed to represent the dependency relation between sentiment labels of two adjacent words. Given the j-th word w_j $(j > 0)$ in d, the sentiment label l_j is selected depending on the previous word. The transition probability distribution is derived from the sentiment label of l_{j-1} and a transition variable x_j. The transition variable x_j determines where the corresponding sentiment label comes from. If $x_j = 1$, then the sentiment label l_j of w_j is identical to the sentiment label l_{j-1} of word w_{j-1}; whereas if $x_j = -1$, the sentiment label l_j is opposite to l_{j-1}, which shows that the sentiment labels changes form one polarity to the other. Thus, we set the transition variable $x_j = 1$ when w_j and w_{j-1} are connected by a correlative conjunction, such as "and" and "both"; we set $x_j = -1$ when w_j and w_{j-1} are connected by an adversative conjunction, such as "but" and "whereas"; we set $x_j = 0$ for other kinds of conjunctions. The generative process of sCVR is shown in Fig. 2.

4.3 Inference

Because of the unknown relation among random variables, exact posterior inference for the sCVR model is intractable. We apply a Gibbs EM sampler [47] to conditionally approximate the posterior distribution of random variables in sCVR. Algorithm 1 summarizes the Gibbs EM sampling inference procedure that we will derive.

Specifically, Algorithm 1 is divided into two parts: an **E**-step and an **M**-step. Given item i and user u, for each user review d the target of our sampling in the **E**-step is to approximate the posterior distribution $p(\mathcal{V}, \mathcal{Z}, \mathcal{L} \mid \mathcal{W}, \mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{F})$. Conceptually, in this step we divide our sampling procedure into three parts.

- Firstly, given a user u and an item i, during the **E**-step, we sample the conditional probability of viewpoint $f_{u,i}$ given current state of viewpoints, i.e., $P(f_{(u,i)} \mid \mathbf{f}_{-(u,i)}, \mathcal{W}, \mathcal{V}, \mathcal{R})$.
- Secondly, given the values of inferred topics and sentiment labels, we sample the conditional probability of viewpoint v in each $d \in \mathcal{D}$, i.e., $P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z}, \mathcal{R})$.
- Lastly, given the current state of viewpoints, for word w_j we sample the conditional probability of topic z_j with sentiment label l_j transition label x_j, i.e., $P(z_j = k, l_j = l, x_j = x \mid \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{E}, \mathcal{R}, \mathcal{T}, \mathcal{F}, v)$.

During the **M**-step, given conditional probabilities derived during the **E**-step, we maximize each user u's viewpoint distribution θ_u, each viewpoint distribution π and the joint probability of viewpoints, concepts, topics, and sentiments over words, i.e., ϕ.

We now detail our sampling procedures. Given user u and item i, we first sample $f_{u,i}$ over $\mathbf{f}_{-(u,i)}$ without pair (u, i). So for user u's viewpoint over item i, we obtain $P(f_{(u,i)} \mid \mathbf{f}_{-(u,i)}, \mathcal{W}, \mathcal{V}, \mathcal{R})$

- For each viewpoint $v \in \mathcal{V}$:
  - Draw $\mu_v \sim Dir(\chi)$; $\lambda_v \sim Dir(\delta)$;
  - For each topic z:
    * Draw $\rho_{v,z} \sim Beta(\eta)$;
    * For each sentiment l:
      · Draw $\phi_{z,l,v} \sim Dir(\beta)$;
- For each user $u \in \mathcal{U}$:
  - Draw $\theta_v^u \sim Dir(\theta_{u,v}^0 + \frac{1}{F_u} \sum_{u' \in \mathcal{F}_u} \mathcal{T}_{u,u'} \theta_v^{u'})$;
- For each item $i \in \mathcal{I}$:
  - Draw $\pi_v \sim Dir(\alpha)$;
  - For each user review $d \in \mathcal{D}_{u,i}$ from user u:
    * Draw a viewpoint $v \sim Multi(\pi)$;
    * Draw a concept $e_d \sim Multi(\lambda_v)$;
    * Draw $\sigma \sim Dir(\tau)$;
    * For each word w_j in document d:
      · Draw a topic $z_j \sim Multi(\mu_v)$;
      · Draw $x_j \sim Multi(\sigma)$;
      · If $x_j = 1$, draw $l_j \sim l_{j-1}$
      · If $x_j = -1$, draw $l_j \sim (-1) \cdot l_{j-1}$;
      · If $x_j = 0$, draw $l_j \sim Bern(\rho_{v,z_j})$;
      · Draw word $w_j \sim Multi(\phi_{v,z_j,l_j})$:
  - For each ratings assigned by user u to i:
    * Draw viewpoint $f_{u,i} \sim Multi(\pi)$;
    * Draw rating $r_{u,i} \sim Multi(\theta_{f_{u,i}}^u)$;

Figure 2: Generative process of sCVR.

as:

$$P(f_{(u,i)} = y \mid \mathbf{f}_{-(u,i)}, \mathcal{W}, \mathcal{V}, \mathcal{R}) \propto \frac{n_{u,-i}^{r(u,i),y} + \theta_{r(u,i),y}^u}{n_u^y + R_u \cdot \theta_{r(u,i),y}^u} \cdot \frac{n_{f,-(u,i)}^{i,y} + n_v^{i,y} + \alpha}{n_{f,-(u,i)}^i + n_v^i + V\alpha}, \tag{3}$$

where R_u indicates how many times user u rates items, and $n_{f,-(u,i)}^{i,y}$ indicates the number of times that variable f has been assigned to y given item i, excluding user u; furthermore, $n_v^{i,y}$ indicates the number of times that viewpoint v in item i has been assigned to y; and $n_{u,-i}^{r(u,i),y}$ indicates the number of times that user u gives rating $r_{(u,i)}$ under $f = y$ for all items, excluding i. We calculate $\theta_{r(u,i),y}^u$ according to Eq. 4:

$$\theta_{r(u,i),y}^u = \theta_{u,y,r(u,i)}^0 + \frac{1}{F_u} \sum_{u' \in \mathcal{F}_u} \mathcal{T}_{u,u'} \cdot \theta_{r(u,i),y}^{u'}, \tag{4}$$

where $\mathcal{T}_{u,u'}$ indicates the trust value between user u and u', and $\mathcal{F}_u$ indicates the trusted social relations of user u. For review d written by user u for item i, we infer the conditional probability of viewpoint $v_d = v$ given all other random variables, i.e., $P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z}, \mathcal{R})$. So we have:

$$P(v_d = v \mid \mathcal{V}_{-d}, \mathcal{E}, \mathcal{W}, \mathcal{Z}, \mathcal{R}) \propto \tag{5}$$

Algorithm 1: Gibbs EM sampling for sCVR's inference

```
Input: α, β, η, τ, U, I, R, W
Output: θ, φ, μ, λ and π
ite = 0;
if ite<T then
    E-Step:
    for u = 1 to U do
        for i = 1 to I do
            Draw f_{u,i} = y from Eq. 3
            Update n_f^{i,y}, n_v^{i,y} and n_u^{r(u,i),y}
            Draw v_d = v from Eq. 6
            Update n^{i,v}, n_{v,e}, n_{v,z} and n_{z,l,v}^w for w ∈ d
            for j = 1 to N_d do
                Draw ⟨z_j, l_j, x_j⟩ from Eq. 6
                if x_j ≠ 0 then
                    Update n_{v,z_j}, n_{z_j,l_j,v}^{w_j} and n_{x_j}^{w_j}
                end
                if x_j = 0 then
                    Update n_{v,z_j}, n_{k_j,l_j,v}^{w_j}, n_{x_j}^{w_j} and n_{z_j,l_j,v}
                end
            end
        end
    end
    M-Step:
    Re-estimate θ_u, π, φ, μ and λ from Eq. 8;
    Maximize θ̂_{u,v}^0 from Eq. 9;
    ite = ite + 1 and go to E-Step;
end
```

$$\frac{n_{-d}^{i,v} + n_f^{i,v} + \alpha}{n_{-d}^i + n_f^i + V\alpha} \cdot \prod_{e \in \mathcal{E}} \frac{n_{v,e}^{-d} + \delta}{n_v^{-d} + E\delta} \cdot \prod_{z \in \mathcal{Z}} \frac{n_{v,z}^{-d} + \chi}{n_v^{-d} + K\chi} \cdot \prod_{l \in \mathcal{L}} \prod_{w \in \mathcal{N}_d} \frac{n_{z,l,v}^{w,-d} + \beta}{n_{z,l,v}^{-d} + N\beta},$$

where $n_{-d}^{i,v}$ indicates the number of times that viewpoint v has been assigned to user reviews, excluding d; $n_{v,e}^{-d}$ indicates the number of times that concept e has been assigned to viewpoint v in reviews, excluding d; $n_{v,z}^{-d}$ indicates the number of times that topic z has been assigned to viewpoint v excluding d; furthermore, $n_{z,l,v}^{w,-d}$ indicates how many words are assigned to topic z, viewpoint v and sentiment l, except for d. Given detected viewpoint $v_d = v$, for each word $w_j \in \mathcal{N}_d$ we sample the conditional probability of topic z_j with sentiment label l_j for word w_j, i.e., $P(z_j = k, l_j = l, x_j = x \mid v, \mathcal{X}_{-j}, \mathcal{L}_{-j}, \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{R}, \mathcal{F})$. Given the viewpoint v sampled at the document level, when $x_j \neq 0$ and $x_{j+1} \neq 0$ we can directly sample word w_j's topic z_j and sentiment label l_j using the probability in Eq. 6:

$$P(z_j = k, l_j = l, x_j = x \mid v, \mathcal{X}_{-j}, \mathcal{L}_{-j}, \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{R}, \mathcal{F}) \propto \frac{n_{v,k}^{-j} + \chi}{n_v^{-j} + K\chi} \cdot \frac{n_{k,l,v}^{w_j,-j} + \beta}{n_{k,l,v}^{-j} + N\beta} \cdot \frac{n_{-j,x}^{w_j} + \tau_x}{n_{-j}^{w_j} + \sum_{x \in \mathcal{X}} \tau_x} \cdot \frac{n_{-(j+1),x_{j+1}}^{w_{j+1}} + I(x_{j+1} = x_j) + \tau_{x_{j+1}}}{n_{-(j+1)}^{w_{j+1}} + 1 + \sum_{x \in \mathcal{X}} \tau_x}, \tag{6}$$

where $n_{v,k}^{-j}$ indicates the number of times that topic k has been assigned to viewpoint v, excluding the j-th word in d; n_v^{-j} indicates how many topics have been assigned to v, not including w_j; $n_{k,l,v}^{w_j,-j}$ indicates the number of times that word w_j has been assigned to topic z and sentiment l synchronously, excluding current one; $n_{-j,x}^{w_j}$ indicates the number of times that w_j assigned to x, excluding current word; and $I(x_{i+1} = x_i)$ gets value 1 if $x_{i+1} = x_i$, otherwise it gets 0. When $x_j = 0$, w_j's sentiment label l_j is derived from a Bernoulli distribution ρ_{v,z_j}; then the conditional probability $P(z_j = k, l_j = l, x_j = 0 \mid v, \mathcal{X}_{-j}, \mathcal{L}_{-j}, \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{R}, \mathcal{F})$ becomes:

$$P(z_j = k, l_j = l, x_j = 0 \mid v, \mathcal{X}_{-j}, \mathcal{L}_{-j}, \mathcal{Z}_{-j}, \mathcal{W}, \mathcal{R}, \mathcal{F}) \propto$$
$$\frac{n_{v,k}^{-j} + \chi}{n_v^{-j} + K\chi} \cdot \frac{n_{,k,l,v}^{w_j,-j} + \beta}{n_{k,l,v}^{-j} + N\beta} \cdot \frac{n_{-j,x}^{w_j} + \tau_x}{n_{-j}^{w_j} + \sum\limits_{x \in \mathcal{X}} \tau_x} \cdot \frac{n_{z,l,v}^{-j} + \eta_l}{n_{z,v}^{-j} + \sum\limits_{l \in \mathcal{L}} \eta_l}, \quad (7)$$

where $n_{z,l,v}^{-j}$ indicates how many words are assigned to viewpoint v, topic z and sentiment label l, excluding current w_j; whereas $n_{v,z}^{-j}$ indicates how many words are assigned to viewpoint v and topic z, excluding current w_j.

In the **M**-step, given conditional probabilities derived in the **E**-step, we estimate the parameters of user u's viewpoint distribution θ_u for each rating r, the viewpoint distribution π_i for each item i, the probability of topics, viewpoints and sentiment over words ϕ, viewpoint distributions over concepts λ and viewpoint distributions over topics μ, respectively, as follows:

$$\begin{aligned}
\theta_{r,v}^u &= \frac{n_u^{r,v} + \theta_{u,v,r}^0 + \frac{1}{F_u} \sum\limits_{u' \in \mathcal{F}_u} \mathcal{T}_{u,u'} \theta_{r,v}^{u'}}{n_{u,v} + R_u \cdot \left(\theta_{u,v,r}^0 + \frac{1}{F_u} \sum\limits_{u' \in \mathcal{F}_u} \mathcal{T}_{u,u'} \theta_{r,v}^{u'} \right)} \\
\pi_{i,v} &= \frac{n_{i,v} + \alpha}{n_i + V\alpha}; \qquad \phi_{v,z,l}^w = \frac{n_{v,z,l}^w + \beta}{n_{v,z,l} + N\beta} \qquad (8) \\
\mu_{v,e} &= \frac{n_{v,z} + \chi}{n_v + K\chi}; \qquad \lambda_{v,e} = \frac{n_{v,e} + \delta}{n_v + E\delta}.
\end{aligned}$$

Given posterior viewpoint distributions, we optimize the value of random variables θ_u^0 for each user u. Using two bounds defined in [27], we derive the following update rule for obtaining each user u's optimized viewpoint distribution in Eq. 8 via fixed-point iterations:

$$\hat{\theta}_{u,v}^0 \leftarrow \theta_{u,v}^0 \cdot \frac{\sum\limits_{v \in V} \Psi(n_{r,v}^u + \theta_{r,v}^u) - \Psi(\theta_{r,v}^u)}{\sum\limits_{v \in V} \Psi(n_v^u + R_u \cdot \theta_{r,v}^u) - \Psi(R_u \cdot \theta_{r,v}^u)}, \quad (9)$$

where $\Psi(x)$ is a digamma function defined by $\Psi(x) = \frac{\partial \log \Gamma(x)}{\partial x}$, and $\theta_{r,v}^u$ is defined in Eq. 4.

This concludes our derivations for Algorithm 1.

4.4 Prediction

After Gibbs EM sampling, for each user $u \in \mathcal{U}$, we have a matrix θ_u to describe the conditional probability of ratings given u's viewpoints, i.e., $P(r \mid v, u) = \theta_{r,v}^u$ over ratings. For each item $i \in \mathcal{I}$, we have a viewpoint distribution π_i, i.e., $P(v \mid i) = \pi_{v,i}$. Therefore, given user $u \in \mathcal{U}$ and item $i \in \mathcal{I}$, in order to predict an unknown rating between u and i, we calculate the probability of the rating $r_{u,i} = r$ by Eq. 10:

$$P(r_{u,i} = r \mid u, i) = \textstyle\sum_{v \in \mathcal{V}} \theta_{r,v}^u \cdot \pi_{i,v}. \quad (10)$$

By ranking $P(r_{u,i} = r \mid u, i)$ for each candidate rating r, we choose the rating r with the highest probability as the predicted rating for u and i.

5. EXPERIMENTAL SETUP

5.1 Research questions

We list the research questions that guide the remainder of the paper: (RQ1) What is the performance of sCVR in rating prediction and top-k item recommendation tasks? Does it outperform state-of-the-art baselines? (RQ2) Can sCVR generate explainable recommendation results? (RQ3) What is the effect of the number of viewpoints and the number of topics on the performance of sCVR? (RQ4) What is the effect of trusted social relations in collaborative filtering? Do they help to enhance the recommendation performance?

5.2 Datasets

We use three benchmark datasets in our experiments: the *Yelp dataset challenge 2013*, *Yelp dataset challenge 2014*[1] and *Epinions.com* dataset.[2] Each dataset has previously been used in research on recommendation algorithms [7, 22, 40]. In total, there are over 400,000 users, 80,000 items, 4,000,000 trusted social relations and 2,000,000 user reviews in our datasets. We show the statistics about our datasets in Table 2.

Table 2: Overview of the three datasets used in the paper.

	Yelp 2013	Yelp 2014	Epinions
items	15,584	61,184	26,850
reviews	335,021	1,569,264	77,267
users	70,816	366,715	3,474
relations	622,873	2,949,285	37,587

Yelp provides a business reviewing platform. Users are able to create a profile that they can use to rate and comment on services provided by local businesses. This service also provides users with the ability to incorporate a social aspect to their profiles by adding people as friends. Our first two datasets ("Yelp challenge 2013" and "Yelp challenge 2014" in Table 2) consist of data from the Yelp dataset challenge 2013 and 2014, respectively. The Yelp dataset challenge 2013 contains $15,584$ items, $70,816$ users and $335,021$ user reviews. Between the users, there are $622,873$ social relations. For the Yelp dataset challenge 2014, we find $366,715$ users, $61,184$ items, $1,569,264$ reviews and $2,949,285$ edges in the dataset. The two datasets are quite sparse, which may negatively most collaborative filtering methods based on ratings.

Epinions.com is a consumer opinion website on which people can share their reviews of products. Members of Epinions can review items, e.g., food, books, and electronics, and assign numeric ratings from 1 to 5. Epinions members can identify their own Web of Trust, a group of "reviewers whose reviews and ratings they have consistently found to be valuable." Released by [7], this dataset includes $3,474$ users with $77,267$ reviews for $26,850$ items; there are $37,587$ social edges in this dataset.

5.3 Evaluation metrics

We employ two offline evaluation metrics in our experiments: Mean Absolute Error (*MAE*) and Root Mean Square Error (*RMSE*). Both of them are widely used evaluation metrics for rating prediction in recommender systems. Given a predicted rating $\hat{r}_{u,i}$ and a ground-truth rating $r_{u,i}$ from user u to item i, the RMSE is calcu-

[1]Datasets are available at http://www.yelp.com/dataset_challenge.
[2]This dataset is available at http://epinions.com.

Table 3: Baselines and methods used for comparison.

Acronym	Gloss	Reference
CVR	Collaborative viewpoint regression	Section 4
sCVR	Social collaborative viewpoint regression	Section 4
Collaborative filtering methods		
CliMF	Maximize reciprocal rank method for item ranking	[37]
LRMF	List-wise learning to rank method for item ranking	[36]
NMF	Non-negative matrix factorization	[20]
PMF	Probabilistic matrix factorization	[28]
SoMF	Trust propagation matrix factorization	[15]
TrMF	Trust social matrix factorization	[50]
Explainable recommendation methods		
CTR	Collaborative topic regression model	[48]
EFM	Explicit factor model for item recommendation	[54]
HFT	Hidden factors as topics model	[25]
RMR	Ratings meet reviews model	[22]
SCTR	Social-aware collaborative topic regression	[7]

lated as:

$$RMSE = \sqrt{\frac{1}{R}\sum_{u,i}(r_{u,i} - \hat{r}_{u,i})^2}, \quad (11)$$

where R indicates the number of ratings between users and items. Similarly, MAE is calculated as follows:

$$MAE = \sqrt{\frac{1}{R}\sum_{u,i}|r_{u,i} - \hat{r}_{u,i}|}. \quad (12)$$

These two criteria measure the error between the true ratings and the predicted ratings.

Statistical significance of observed differences between the performance of two runs is tested using a two-tailed paired t-test and is denoted using $^{\blacktriangle}$ (or $^{\blacktriangledown}$) for strong significance for $\alpha = 0.01$; or $^{\vartriangle}$ (or $^{\triangledown}$) for weak significance for $\alpha = 0.05$.

5.4 Baselines and comparisons

We list the methods and baselines that we consider in Table 3. In this paper, we propose the social collaborative viewpoint regression model (sCVR); we write sCVR for the overall process as described in Section 4, which includes both viewpoint modeling and social relation modeling. We write CVR for the model that only considers viewpoint modeling in Section 4.

Our baselines include recent work on both collaborative filtering and explainable recommendation methods. To evaluate the performance of our viewpoint modeling methods in explainable recommendation, we contrast them against previous work on explainable recommendation: the hidden factors topic model (HFT) [25], the collaborative topic regression (CTR) [48], and the ratings meet reviews model (RMR) [22]. Using a sentiment lexicon analysis tool [54], we use EFM [54] as a baseline in our experiments for explainable recommendation. To evaluate the effect of social communities in explainable recommendation, we use social-aware collaborative topic regression (SCTR) [7] as another baseline. We also compare sCVR with recent collaborative filtering methods: probabilistic matrix factorization (PMF) [28], non-negative matrix factorization (NMF) [20], list-rank matrix factorization (LRMF) [36] and collaborative less-is-more filtering (CliMF) [37]. To compare sCVR with collaborative filtering using trusted social relations, we use trust matrix factorization (TrMF) [50] and social matrix factorization (SoMF) [15] as another two baselines.

Table 4: RQ1: MAE and RMSE values for rating prediction. Significant differences are with respect to SCTR (row with shaded background).

	Yelp 2013		Yelp 2014		Epinions	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Collaborative filtering						
CliMF	1.109	1.524	1.591	1.912	0.493	0.582
LRMF	1.653	1.944	1.897	2.042	0.517	0.626
NMF	1.130	1.591	1.284	1.763	0.595	0.691
PMF	1.427	1.853	1.424	1.902	0.526	0.688
SoMF	0.912	1.375	0.924	1.402	0.554	0.673
TrMF	1.109	1.524	1.134	1.564	0.542	0.667
Explainable recommendations						
CTR	0.915	1.169	0.971	1.294	0.525	0.612
EFM	0.912	1.182	1.124	1.452	0.532	0.644
HFT	0.844	1.072	1.094	1.336	0.517	0.604
LDA	1.232	1.622	1.294	1.677	0.526	0.612
RMR	0.812	1.013	0.937	1.283	0.514	0.602
SCTR	0.894	1.065	0.907	1.262	0.472	0.584
sCVR	**0.744**$^{\blacktriangle}$	**0.977**$^{\blacktriangle}$	**0.806**$^{\blacktriangle}$	**1.196**$^{\blacktriangle}$	**0.482**	**0.579**

6. RESULTS AND DISCUSSION

In Section 6.1, we compare sCVR to other baselines for rating prediction; we discuss the *explainability* of rating predictions in Section 6.2; in Section 6.3 we examine the performance of sCVR for varying numbers of viewpoints and topics; Section 6.4 examines the effect of social relations in sCVR.

6.1 Overall performance

To start, for research question **RQ1**, to evaluate the effectiveness of sCVR in personalized recommendation, we examine the performance of sCVR on the rating prediction task. Table 4 lists the performance of all methods in terms of MAE and RMSE. Because our baselines predict decimal rating values based on a Gaussian noise distribution, following Beutel et al. [3], we calculate the predictive probability, i.e., $P(r \mid \hat{r})$, for each predicted rating $\hat{r}$, and we use the discrete rating with highest predictive probability in our experiments. For all three datasets, sCVR outperforms other baselines, and significantly outperforms SCTR on the Yelp 2013 and 2014 datasets. The list-wise learning to rank methods (LRMF and CliMF) do not perform well in rating prediction, whereas methods considering social relations outperform other methods. To understand the benefits of viewpoint modeling (and in particular, the addition of concepts and sentiment), we compare sCVR with SCTR, which ignores concepts and sentiment during topic modeling. On the Yelp 2013 dataset, sCVR achieves a 16.7% and 8.2% decrease over SCTR in terms of MAE and RMSE, respectively, whereas on the Yelp 2014 dataset, it achieves decreases of 11.1% and 5.2%, respectively.

6.2 Explainability

Next, we turn to **RQ2**. Apart from being more accurate at rating prediction, another advantage of sCVR over collaborative filtering methods is that it provides *explainable* recommendation results. Explainability of recommendations is often assessed using examples [7, 48]; we follow a similar method. To illustrate the explainability of outcomes of sCVR, Table 5 shows 3 examples of detected viewpoints. In the example viewpoints in Table 5, we see concepts with relevant topics and corresponding sentiment labels. For each viewpoint, we find that relevant topics in the second col-

Table 5: **RQ2: Example viewpoints produced by sCVR on the Yelp 2013 dataset. Column 1 lists the concepts corresponding to the viewpoints; Column 2 list the topics in the viewpoints, Columns 3 and 4 list the probabilities of positive and negative labels for each topic, respectively, Column 5 lists the predicted rating scores with their probabilities for each viewpoint.**

Concept	Topic	Positive	Negative	Ratings
Italian	#topic 2: italian, pizza, well, pasta, menu, wine, favorite, eggplant, dinner, special	0.518	0.482	3/0.251
Fast food	#topic 12: burger, pizza, cheap, bad, drink, sausage, egg, lunch, garden, price	0.224	0.776	2/0.427
Steakhouses	#topic 7: potato, appetizer, good, place, pork, rib, bread, rib-eye, filet, beef	0.797	0.203	4/0.538

umn help to interpret the concept in the first column, and sentiment labels inform users on opinions in the viewpoint; we also find that the predicted ratings are consistent with the polarity of sentiment for each viewpoint.

In addition, to explicitly assess the quality of our explanations, we take a random sample of 100 recommendations and manually evaluate the corresponding explanations for accuracy of topic and sentiment label, and for agreement between sentiment label and actual rating.[3] Due to space limitations we only include sCVR, RMR (the best performing explainable baseline), and EFM (the only baseline that utilizes sentiment analysis during rating prediction); see Table 6. We find that sCVR outperforms RMR, for topic detection in terms of accuracy. sCVR outperforms EFM in terms of accuracy for sentiment label. As to the agreement between sentiment labels and predicted ratings, compared with EFM, sCVR offers an increase of up to 18.7%. This evaluation indicates that viewpoints make good sense; based on this outcome and the anecdotal evidence mentioned above, we believe that they are more informative explanations than mere topic labels, which were used in prior work; checking this through a user study is left as future work.

Table 6: RQ2: Explainability. Column 1 and 2 list the accuracy of topic detection and sentiment detection, respectively. Column 4 lists the agreement between sentiment labels and predicted ratings.

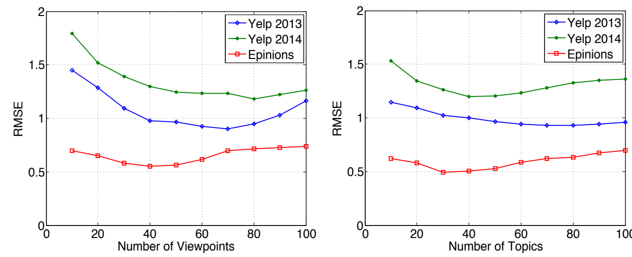
Method	Topics	Sentiment labels	Agreement
EFM	N/A	0.715	0.644
RMR	0.542	N/A	N/A
sCVR	**0.569**	**0.893**	**0.792**

6.3 Number of viewpoints and topics

Turning to **RQ3**, under the default value of the number of topics $K = 20$ in sCVR, in Fig. 3(a) we examine the RMSE performance of sCVR with varying numbers of viewpoints. We find that the performance of sCVR in terms of RMSE hits a minimum when the number of *viewpoints* equals 70 for the Yelp 2013 dataset; with fewer than 70, performance decreases but when the number exceeds 70, due to the redundancy of viewpoints in rating prediction, performance increases. Similar phenomena can be found for the Yelp 2014 and Epinions datasets. For Yelp 2014, sCVR achieves its best RMSE performance when the number of viewpoints equals 80, whereas for the Epinions dataset, it achieves its best RMSE performance when we set V to 40.

Under the default value of the number of viewpoints $V = 30$, we evaluate the RMSE performance of sCVR with varying numbers of topics in Fig. 3(b). We find that for the Yelp 2013 dataset, sCVR achieves its best RMSE performance when $K = 80$, whereas for

[3]We say that there is *agreement* if the rating $s \geq 3$ (< 3) and the sentiment label of the corresponding review is positive (negative).



(a) Different numbers of viewpoints (b) Different numbers of topics.

Figure 3: RQ3: RMSE of sCVR with different numbers of viewpoints and topics.

the Yelp 2014 dataset this value is 40. For the Epinions dataset, sCVR performs best when $K = 30$.

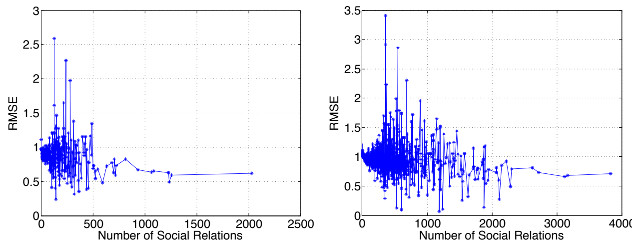
6.4 Effect of social relations

Finally, we address **RQ4**. To determine the contribution of social relations in the rating prediction task, we turn to Table 7, where columns 2–3 and 4–5 show the performance of CVR and sCVR, respectively, in terms of MAE and RMSE. Recall that CVR only detects viewpoints without considering social relations. We find that sCVR, which does consider social relations, outperforms CVR significantly on all three datasets. From Table 4, we also see that methods considering social relations perform quite well in terms of MAE and RMSE. For the Yelp 2013 dataset, sCVR achieves a 6.7% decrease over CVR in terms of RMSE. For the Yelp 2014 dataset, sCVR achieves a 7.4% decrease over CVR in terms of RMSE. In terms of RMSE, on the Epinions dataset, sCVR achieves a significant decrease over CVR of 18.7%. Thus, we conclude that social communities can successfully be applied to enhance the performance of rating prediction.

Table 7: RQ4: Effect of social communities on rating prediction in our three datasets.

	CVR		sCVR	
Dataset	MAE	RMSE	MAE	RMSE
Yelp 2013	0.862	1.049	**0.744**$^{\triangle}$	**0.977**$^{\triangle}$
Yelp 2014	0.953	1.291	**0.806**$^{\blacktriangle}$	**1.196**$^{\blacktriangle}$
Epinions	0.641	0.712	**0.482**$^{\blacktriangle}$	**0.579**$^{\blacktriangle}$

To evaluate the effect of the number of social relations, Fig. 4 shows the average RMSE performance for users with different numbers of social relations on the Yelp 2013 and Yelp 2014 datasets. From Fig. 4 we find that for both Yelp 2013 and Yelp 2014 datasets, RMSE performance shows a "wave-like" decrease as the number of social relations increases. Thus, users with more social relations, in most cases, will achieve better prediction results using sCVR.



(a) Yelp 2013 (b) Yelp 2014

Figure 4: RQ4: RMSE of sCVR with different numbers of social relations on the Yelp datasets.

7. CONCLUSION AND FUTURE WORK

We have considered the task of rating-based recommendations with explanations. To improve the rating prediction, we have identified two main problems: opinions in users' short comments, and complex trusted social relations. We have tackled these problems by proposing a novel latent variable model, the social collaborative viewpoint regression model, which detects viewpoints and uses social relations. The model has two key parts: viewpoint detection and rating prediction. Based on the probabilistic distribution of viewpoints, we predict users' ratings of items.

In our experiments, we have demonstrated the effectiveness of our proposed method and have found significant improvements over state-of-the-art baselines when tested with three benchmark datasets. Viewpoint modeling is helpful for rating prediction and item recommendation. We have also shown that the use of social relations can enhance the accuracy of rating predictions. The viewpoints used can be used as more informative explanations of items and of users' preferences than simple topic-based ones used previously.

Limitations of our work include the fact that it ignores topic drift over time [17]. Furthermore, as it is based on topic models, the conditional independence among topics may in principle lead to redundant viewpoints and topics. As to future work, we plan to explore whether ranking-based strategies that integrate our sCVR model can enhance the performance of item recommendation and whether a combination with interaction data [9] can help to improve performance even further. Also, the transfer of our approach to streaming corpora should give new insights. Finally, we would like to conduct user studies to verify the interpretability of the explanations that sCVR generates and to examine their usefulness across recommendation scenarios.

Acknowledgements. This work is supported by Ahold Delhaize, Amsterdam Data Science, Blendle, the Bloomberg Research Grant program, the Dutch national program COMMIT, Elsevier, the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement nr 312827 (VOX-Pol), the ESF Research Network Program ELIAS, the Royal Dutch Academy of Sciences (KNAW) under the Elite Network Shifts project, the Microsoft Research Ph.D. program, the Netherlands eScience Center under project number 027.012.105, the Netherlands Institute for Sound and Vision, the Netherlands Organisation for Scientific Research (NWO) under project nrs 612.001.116, HOR-11-10, 640.-006.013, CI-14-25, 652.002.001, 612.001.551, 652.001.003, 314-98-071, the Yahoo Faculty Research and Engagement Program, Yandex, and the EPSRC grant EP/K031953/1 ("Early-Warning Sensing Systems for Infectious Diseases"). All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

8. REFERENCES

[1] R. M. Bell and Y. Koren. Improved neighborhood-based collaborative filtering. In *KDD Cup and Workshop at the KDD*, 2007.
[2] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8): 1798–1828, 2013.
[3] A. Beutel, K. Murray, C. Faloutsos, and A. J. Smola. Cobafi: collaborative bayesian filtering. In *WWW*, 2014.
[4] P. Bhargava, T. Phan, J. Zhou, and J. Lee. Who, what, when, and where: Multi-dimensional collaborative recommendations using tensor factorization on sparse user-generated data. In *WWW*, 2015.
[5] M. Bilgic and R. J. Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Beyond Personalization Workshop, IUI*, 2005.
[6] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3:993–1022, 2003.
[7] C. Chen, X. Zheng, Y. Wang, F. Hong, and Z. Lin. Context-aware collaborative topic regression with social matrix factorization for recommender systems. In *AAAI*, 2014.
[8] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu. Collaborative personalized tweet recommendation. In *SIGIR*, 2012.
[9] A. Chuklin, I. Markov, and M. de Rijke. *Click Models for Web Search*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, August 2015.
[10] Q. Diao, M. Qiu, C.-Y. Wu, A. J. Smola, J. Jiang, and C. Wang. Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars). In *KDD*, 2014.
[11] X. Ding, B. Liu, and P. S. Yu. A holistic lexicon-based approach to opinion mining. In *WSDM*, 2008.
[12] X. He, T. Chen, M.-Y. Kan, and X. Chen. Trirank: Review-aware explainable recommendation by modeling aspects. In *CIKM*, 2015.
[13] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313 (5786):504–507, 2006.
[14] S. Huang, S. Wang, T.-Y. Liu, J. Ma, Z. Chen, and J. Veijalainen. Listwise collaborative filtering. In *SIGIR*, 2015.
[15] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *RecSys*, 2010.
[16] N. Kalchbrenner, E. Grefenstette, and P. Blunsom. A sentence model based on convolutional neural networks. In *ACL*, 2014.
[17] T. Kenter, P. Huijnen, M. Wevers, and M. de Rijke. Ad hoc monitoring of vocabulary shifts over time. In *CIKM 2015: 24th ACM Conference on Information and Knowledge Management*. ACM, October 2015.
[18] I. Konstas, V. Stathopoulos, and J. M. Jose. On social networks and collaborative recommendation. In *SIGIR*, 2009.
[19] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 6(8): 30–37, 2009.
[20] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, 2001.

[21] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, and H. Yu. Structure-aware review mining and summarization. In *Coling*, 2010.
[22] G. Ling, M. R. Lyu, and I. King. Ratings meet reviews, a combined approach to recommend. In *ACM RecSys*, 2014.
[23] P. Lops, M. De Gemmis, and G. Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender systems handbook*, pages 73–105. Springer, 2011.
[24] H. Ma, I. King, and M. Lyu. Learning to recommend with social trust ensemble. In *SIGIR*, pages 203–210, 2009.
[25] J. McAuley and J. Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *RecSys*, pages 165–172, 2013.
[26] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
[27] T. Minka. Estimating a dirichlet distribution, 2000.
[28] A. Mnih and R. Salakhutdinov. Probabilistic matrix factorization. In *NIPS*, 2007.
[29] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP*, 2002.
[30] Z. Ren and M. de Rijke. Summarizing contrastive themes via hierarchical non-parametric processes. In *SIGIR*, 2015.
[31] Z. Ren, O. Inel, L. Aroyo, and M. de Rijke. Time-aware multi-viewpoint summarization of multilingual social text streams. In *CIKM*, 2016.
[32] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *CSCW*, 1994.
[33] M. T. Ribeiro, S. Singh, and C. Guestrin. "Why should I trust you?" Explaining the predictions of any classifier. In *KDD*, 2016.
[34] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *WWW*, 2001.
[35] A. Sharma and D. Cosley. Do social explanations work?: studying and modeling the effects of social explanations in recommender systems. In *WWW*, 2013.
[36] Y. Shi, M. Larson, and A. Hanjalic. List-wise learning to rank with matrix factorization for collaborative filtering. In *RecSys*, 2010.
[37] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, N. Oliver, and A. Hanjalic. Climf: learning to maximize reciprocal rank with collaborative less-is-more filtering. In *RecSys*, 2012.
[38] R. Socher et al. Recursive deep models for semantic compositionality over a sentiment treebank. In *EMNLP*, 2013.
[39] X. Su and T. M. Khoshgoftaar. A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009: 4, 2009.
[40] D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin. Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL*, 2014.
[41] D. Tang, B. Qin, and T. Liu. Document modeling with gated recurrent neural network for sentiment classification. In *EMNLP*, 2015.
[42] D. Tang, B. Qin, T. Liu, and Y. Yang. User modeling with neural network for review rating prediction. In *IJCAI*, 2015.
[43] M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social web. *J. Am. Soc. Inf. Sci. Technol.*, 63(1):163–173, 2012.
[44] N. Tintarev and J. Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender Systems Handbook*, 2011.
[45] P. D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *ACL*, 2002.
[46] J. Vig, S. Sen, and J. Riedl. Tagsplanations: explaining recommendations using tags. In *IUI*, 2009.
[47] H. Wallach. Topic modeling: beyond bag-of-words. In *ICML 2006*, pages 977–984, 2006.
[48] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In *KDD*, 2011.
[49] Y. Xu, W. Lam, and T. Lin. Collaborative filtering incorporating review text and co-clusters of hidden user communities and item groups. In *CIKM*, 2014.
[50] B. Yang, Y. Lei, D. Liu, and J. Liu. Social collaborative filtering by trust. In *IJCAI*, 2013.
[51] S. Yang, B. Long, A. Smola, N. Sadagopan, Z. Zheng, and H. Zha. Like like alike: joint friendship and interest propagation in social networks. In *WWW 2011*, pages 537–546, 2011.
[52] M. Ye, X. Liu, and W. Lee. Exploring social influence for recommendation: a generative model approach. In *SIGIR*, 2012.
[53] Y. Zhang, M. Zhang, Y. Liu, S. Ma, and S. Feng. Localized matrix factorization for recommendation based on matrix block diagonal forms. In *WWW*, 2013.
[54] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *SIGIR*, 2014.
[55] J. Zhao, L. Dong, J. Wu, and K. Xu. Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In *KDD*, 2012.