Coupled Matrix Factorization and Topic Modeling for Aspect Mining

Ding Xiao[a], Yugang Ji[a], Yitong Li[b], Fuzhen Zhuang[c], Chuan Shi[a,*]

[a] *Beijing University of Posts and Telecommunications, Beijing, China*
[b] *Search Technology Center Asia, Beijing, China*
[c] *University of Chinese Academy of Sciences, Beijing, China*

Abstract

Aspect mining, which aims to extract ad hoc aspects from online reviews and predict rating or opinion on each aspect, can satisfy the personalized needs for evaluation of specific aspect on product quality. Recently, with the increase of related research, how to effectively integrate rating and review information has become the key issue for addressing this problem. Considering that matrix factorization is an effective tool for rating prediction and topic modeling is widely used for review processing, it is a natural idea to combine matrix factorization and topic modeling for aspect mining (or called aspect rating prediction). However, this idea faces several challenges on how to address suitable sharing factors, scale mismatch, and dependency relation of rating and review information. In this paper, we propose a novel model to effectively integrate *Ma*trix factorization and *To*pic modeling for *Asp*ect rating prediction (*MaToAsp*). To overcome the above challenges and ensure the performance, MaToAsp employs items as the sharing factors to combine matrix factorization and topic modeling, and introduces an interpretive preference probability to eliminate scale mismatch. In the hybrid model, we establish a dependency relation from ratings to sentiment terms in phrases. The experiments on two real datasets including Chinese Dianping and English Tripadvisor prove that MaToAsp not only obtains reasonable aspect identification but also achieves the best aspect rating prediction performance,

*Corresponding author

Email addresses: dxiao@bupt.edu.cn (Ding Xiao), jiyugang@bupt.edu.cn (Yugang Ji), liyitong0420@126.com (Yitong Li), zhuangfz@ics.ict.ac.cn (Fuzhen Zhuang), shichuan@bupt.edu.cn (Chuan Shi)

compared to recent representative baselines.

Keywords:
Topic modeling, Matrix factorization, Aspect mining, Rating prediction

1. Introduction

In order to tackle information overload problem, recommender systems are proposed to help users to find objects of interest through utilizing the user-item interaction information and/or content information associated with users and items. An important task in recommender system is to predict ratings of users to items. Ratings are very important, since users can directly judge whether the item is good or not from the rating. For example, restaurants with low rating are typically considered to provide worse service than those with high rating. Many methods have been proposed to predict the overall rating which reflects the quality of items from a general view (Salakhutdinov & Mnih, 2007; Koren et al., 2009; Balazs & Velásquez, 2016). However, in many situations, people may expect more subtle aspect ratings of items. For example, users not only care about the overall evaluation (i.e., overall rating) of a restaurant, but also take notice of the taste, environment and service of a restaurant. Some foodies may care more about the taste, while some business men may care more about the environment. This problem has inspired the research on aspect-level opinion mining (Wang et al., 2010; Lu et al., 2009; Li et al., 2015b; Xue et al., 2017). The goal of aspect-level opinion mining is to extract ad hoc aspects from online reviews and predict rating or opinion on each aspect.

Because of its great practical significance, there is a surge of research on aspect mining (or called aspect rating prediction) in recent years. Some earlier works (Wang et al., 2010; Lu et al., 2009) design two-stages methods to separately detect aspects of terms and predict aspect ratings. Taking advantage of observed overall ratings or external knowledge, recent researchers attempt to integrate the two tasks into a unified framework. Some works generate ratable aspects from reviews with whole overall ratings (Diao et al., 2014) or scarce overall ratings (Luo et al., 2014), and some works consider to integrate external knowledge (Wang & Ester, 2014). The aspect rating prediction usually utilizes the rating information, as well as reviews. How to effectively utilize the ratings and reviews information is a key issue for aspect rating prediction. Many approaches have been proposed to solve this issue.

For example, (Moghaddam & Ester, 2013; Luo et al., 2015; Wang & Ester, 2014) utilize latent variable models to combine rating and review. However, they fail to point out the intrinsic relationship between aspects and ratings. In other words, ratings and reviews haven't been combined tightly enough. Recently, in order to solve the problem of lacking intrinsic relationship between aspects and ratings, AIRS model (Li et al., 2015a) allows an aspect to directly affect the sampling of a latent rating on this aspect, but it is still short of the direct relationship between rating and words.

As we know, matrix factorization is an effective tool to handle rating information (Salakhutdinov & Mnih, 2007; Koren et al., 2009), and the topic modeling is widely used to process review information (Xu et al., 2015; Deerwester et al., 1990; Lu et al., 2009). So a direct way is to combine matrix factorization and topic modeling for aspect rating prediction. It is not trivial to effectively combine them to make full use of rating and review information. It will face the following challenges.

- It is desirable to carefully design the sharing factors of these two models. A basic idea of integrating two models is to share some common factors. However, what kind of factors to be shared can facilitate full use of rating and review information? It needs to be delicately designed.

- Scale mismatch exists in matrix factorization and topic modeling. As we know, in matrix factorization, the latent factors of users and items are usually arbitrary real numbers. However, the probabilities in topic modeling are usually from 0 to 1. As a consequence, the scales of the latent factor and the probability representation do not match.

- It is also difficult to design the dependency relation between rating and review information. Whether the ratings of users determine or depend on their reviews? We need to design proper dependency relation between rating and review information which complies with user behaviors.

Inspired by the above, our goal is to obtain more accurate aspect rating prediction by integrating both of matrix factorization and topic modeling. On one hand, we need to design the specific topic model and matrix factorization to detect the review information and rating information respectively. On the other hand, to integrate review and rating information, it is quite necessary to address the mentioned challenges.

In this paper, we propose a novel model to effectively integrate *Ma*trix factorization and *To*pic modeling for *Asp*ect rating prediction, called *MaToAsp*. The MaToAsp model respectively utilizes matrix factorization and topic modeling to handle rating and review information. Different from traditional matrix factorization, the latent factors of users and items in the proposed model are confined to [0, 1], and the multiplication of the latent factors of user and items is a preference probability from 0 to 1, not a rating score from 1 to 5. We design a transformation function to convert the preference probability to the same range of score. The proposed matrix factorization method not only handles the scale mismatch problem but also avoids the unexplainable latent factors of users and items in traditional matrix factorization. Different from traditional topic modeling, we design a topic model to predict aspect ratings at phrase level. We believe that the aspect rating prediction at phrase level can more subtly capture aspect sentiments, as recent works have validated (Lu et al., 2009; Luo et al., 2015).

After that, we integrate matrix factorization and topic modeling through sharing the latent factor of items, since it not only generates the aspect sentiments from reviews but also generates the preference probability with the latent factor of users. Moreover, we establish a dependency relation from rating to sentiment terms of phrases, which makes the distribution over sentiment term more reasonable. Finally, we design an objective function to concurrently optimize matrix factorization and topic modeling, and propose an iterative optimization framework to successively optimize MaToAsp with gradient descent and Gibbs sampling. Experiments are done on two real datasets including Chinese and English. We first qualitatively validate the effectiveness of aspect identification through a case study. Then we quantitatively verify the quality of aspect rating prediction on two criteria. Experiments show that the proposed MaToAsp always achieves the best performance on all conditions, compared to the state of the art methods.

The main contributions of this paper are summarized as follows:

- To the best of our knowledge, this work is the first to combine matrix factorization and topic modeling for aspect rating prediction, where the two parts can make full use of rating and review information respectively. Though some works (Bao et al., 2014; McAuley & Leskovec, 2013; Zhang & Wang, 2016) have attempted to utilize matrix factorization and topic modeling, it has not yet been exploited to combine matrix factorization and topic modeling for aspect rating prediction.

Our work is the first to explore it and the experiments can validate the effectiveness.

- Our work overcomes the mentioned challenges including suitable sharing factors, scale mismatch, and dependency relation of rating and review information. By setting the latent factors of items as the sharing part and designing the corresponding global optimization function, the rationality and effectiveness of MaToAsp is guaranteed.
- We compare our proposed MaToAsp with some representative methods on both of English and Chinese datasets. These experiments not only validate that MaToAsp is quite better in aspect rating prediction than others but also prove the convergence of MaToAsp.

2. Related Work

Since aspect rating prediction mainly utilizes rating and review information, here we briefly introduce the related work in rating prediction and review processing techniques, and then we present existing methods that integrate the rating and review information.

Rating prediction is a key task in recommender system. Many techniques have been proposed to solve this task. Among them, matrix factorization (Koren et al., 2009) based methods are a type of basic tools. Matrix factorization for collaborative filtering can be generalized as a probabilistic model (Salakhutdinov & Mnih, 2007). It factorizes user-item rating matrix into two low rank user-specific and item-specific matrices, then utilizes the factorized matrices to make further predictions (Srebro & Jaakkola, 2003). To be more specific, it allows us to learn two latent feature matrices corresponding to users' latent features and items' latent features, the dot product between users' features vector and items' features vector gives us the prediction of the rating users would assign them. Recently, some related works would like to integrate some extra information with matrix factorization. In (Guo et al., 2016), Guo et al. proposed a trust-based matrix factorization technique so as to avoid the problem of data sparsity and code start when predicting ratings. In (Taheri et al., 2017), Taheri et al. got rid of social relations and matrix factorization for rating prediction.

However, although matrix factorization models are effective on rating prediction, most of them have a shortcoming that the latent features have no obvious physical meanings. Thus, the recommended results are impossible to explain.

Topic modeling is widely used to handle the review information including aspect extraction and sentiment identification (Titov & McDonald, 2008; Yu et al., 2011). A model adopted the PLSA for aspect identification has been proposed by Lu et al. in (Lu et al., 2009) at the phrase level. A method to identify latent features in Chinese reviews was proposed in (Xu et al., 2015) using LDA and SVM. And AEP-LDA proposed by Zheng et al. in (Zheng et al., 2014) can be used to extract aspect words automatically from reviews. Recently, (Hu et al., 2017) proved that LDA can perform better for review summarization. Our model is designed based on the LDA framework at the phrase level combining matrix factorization method.

In recent years, combining both rating and review information becomes the key point to solve aspect rating prediction problem (Yu et al., 2017). Wang et al. (Wang et al., 2011) proposed a unified generative model LARAM without pre-specified key words. It aims at analyzing opinions expressed in each review at the level of topical aspects to discover each individual reviewer's latent rating on each aspect as well as the relative importance weight on different aspects when forming the overall judgment. Underlying the assumption that the aspects and corresponding ratings of reviews are influenced not only by the items but also by the reviews, Moghaddam et al. (Moghaddam & Ester, 2013) proposed a probabilistic model based on LDA and trained at the category level. Luo et al. (Luo et al., 2015) managed to generate fine-granularity aspects via head, modifier, rating and entity which is called quad-tuples. And Wang et al. (Wang & Ester, 2014) used external knowledge, product-level overall rating distribution and word-level sentiment lexicon, to extract aspects and aspects ratings simultaneously. Diao et al. (Diao et al., 2014) proposed a probabilistic model based on collaborative filtering and topic modeling without requiring knowledge of the aspect specific ratings or genres for inference. In (Bao et al., 2014), Bao et al. proposed a novel matrix factorization model (called TopicMF) which simultaneously considers the ratings and accompanied review texts by using a transform from item and user latent vectors to topic distribution parameters. By doing so, they combine latent factors in rating data with topics in user-review text, but still suffer the interpretation problem as traditional matrix factorization does. In (Zhang & Wang, 2016), Zhang et al proposed an effective model for rating prediction which integrated both of topic model and matrix factorization by sharing rating factor. This model assumed the latent factor of entities obtained by matrix factorization would indirectly affect the generation of words. In (Qiu et al., 2016), Qiu et al. proposed an aspect-based model on ratings and review texts for recommender system which took into account both of users' interested aspect information and items' property. In (Jin et al., 2016), Jin et al. attempted to learn review content vector, aspect representation and users' rating preference vector for rating prediction. Most of these works above achieve good performance in some conditions. However, they haven't successfully model the intrinsic connection between aspect and aspect rating, while, in our approach, we solve it by sharing the distribution of aspects with both rating and reviews.

Some works tried to build intrinsic connection between aspect and aspect rating. Work in (Li et al., 2015a) solved the problem of lacking intrinsic relationship between distribution of aspects and rating by allowing an aspect to directly affect the sampling of a latent rating on this aspect, but it ignored the direct relationship between words and ratings. Authors of (Wang & Blei, 2011) proposed Collaborative Topic Regression (CTR) to suggest scientific articles to potential readers. In (McAuley & Leskovec, 2013), the authors proposed the Hidden Factors and Hidden Topics (HFT) model, which learnt a Latent Dirichlet Allocation (LDA) model for items using the review text and a matrix factorization model to fit the ratings. Ling et al. (Ling et al., 2014) proposed a novel method to combine content-based filtering seamlessly with collaborative filtering, modeling the reviews and ratings simultaneously. These three models are all based on matrix factorization and topic modeling. They can solve the problem of building the intrinsic connection of rating and reviews in some terms, but still have some problems in establishing words and ratings' direct relationship and dealing with discrepancy of the distribution of topics in topic modeling and item feature vector in matrix factorization. In addition, they only focused on the problem of overall rating prediction rather than aspects rating problems, which is firstly studied in this paper.

3. Preliminary

In this section, we introduce the notations and concepts used in this paper.

User: A user i indicates a person who belongs to the user set. There are I users in total.

Item: An item j indicates a product which belongs to the product set (e.g., a restaurant in the Dianping dataset). J indicates the number of items.

Review: A review $d_{i,j}$ is a block of text that expresses user i's opinion about item j. An item j may have many reviews from different users, and all reviews of item j becomes a document $d_{(\cdot),j}$. Since there are J items, the total number of documents is also J.

Phrase: A phrase $l =< h, m >$ consists of a pair of words, which are extracted from the review. h denotes the head term, and m is the modifier term. A document $d_{(\cdot),j}$ contains L_j phrases.

Head term: The head term h is used to describe the aspect information. It decides which aspect the phrase l is expressing. For instance, "attitude" is a head term belonging to the aspect "Service". There are N_h head terms in total.

Modifier term: The modifier term m is used to describe the sentiment information. It is used to describe whether the aspect (decided by h) is good or bad. For instance, for the head term "attitude", "cold" or "passionate" may be used as the modifier term to express negative or positive sentiment. There are N_m modifier terms in total.

Overall rating: An overall rating $r_{i,j}$ along with a review $d_{i,j}$ is an integer rating from 1 to R, which is given by i^{th} user to j^{th} item.

Aspect: An aspect k is a specific side of item j, e.g., the environment of a restaurant. K indicates the number of aspects.

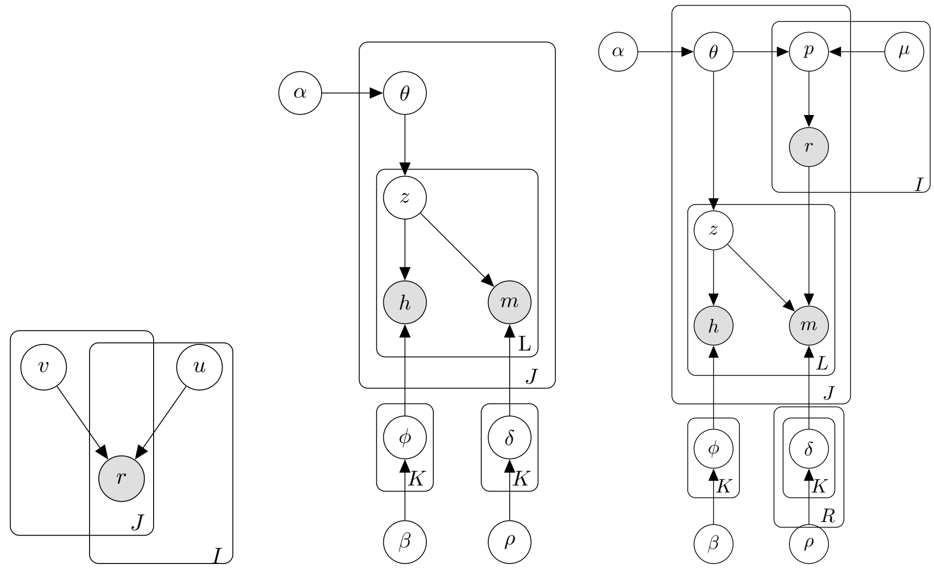
Aspect rating: An aspect rating $a_{i,j,k}$ is a numerical rating, which indicates the i^{th} user's sentiment tendency on the aspect k of the j^{th} item, and is a real number from score 1 to R. A review $d_{i,j}$ associates with K aspect ratings, which correspond to K aspects.

4. The Proposed MaToAsp Method

In this section, we will present our proposed MaToAsp model for solving these challenges mentioned in Section 1. Firstly, we briefly introduce matrix factorization for ratings and a topic model for reviews, respectively. Note that the matrix factorization and topic modeling do not simply employ existing methods, but are adopted to our problem setting. Thereafter, we propose the MaToAsp method based on these two models, in which the mechanism to solve those challenges is introduced. Finally, we give the learning algorithm.

4.1. Matrix Factorization for Rating

The low-rank matrix factorization is very popular for rating prediction (Salakhutdinov & Mnih, 2007; Koren et al., 2009). Its basic idea is to factorize



(a) Matrix Factorization Model

(b) Topic Model

(c) MaToAsp

Figure 1: The graphic models that utilize the rating and review information.

the user-item rating matrix into two matrices representing users' and items' distributions on latent features respectively. And Then, the rating prediction can be made through these two specific matrices. As the graphical model shown in Fig. 1(a), the predicted rating $\hat{r}_{i,j}$ for a user i and an item j can be calculated by

$$\hat{r}_{i,j} = g + b_i + b_j + \langle u_i \cdot v_j \rangle, \tag{1}$$

where g is the rating mean, b_i and b_j are rating biases of user i and item j. u_i and v_j are K-dimensional user and item factors, which represent preferences for user i and properties for item j, respectively. The parameters $\Phi = \{g, b_i, b_j, u_i, u_j\}$ are learned by solving the following problem

$$\min_{\Phi} \sum_{r_{i,j}} (r_{i,j} - \hat{r}_{i,j})^2 + \lambda(\sum_i \|u_i\|^2 + \sum_j \|v_j\|^2), \tag{2}$$

where the first term is the error of the predicted ratings, and the second term is the regularization to prevent poorly scaled solutions.

4.2. Topic Modeling for Review

Reviews contain users' opinions about items on various aspects. Topic models are widely used to uncover the latent aspects (topics) from reviews

(Lu et al., 2009; Xu et al., 2015). In this section, we propose a phrase-level LDA model to model the generative process of the reviews.

Suppose a document $d_{(\cdot),j}$ is the set of all reviews of an item $j \in \{1, 2, \cdots, J\}$, so there are J documents in total. Each document $d_{(\ldots),j}$ contains L_j phrases $l =< h, m >$ ($l \in \{1, 2, \ldots, L_j\}$). K denotes the number of aspects.

The generative process of a document is as follows:

- For each latent aspect $k \in [1, K]$:
  - Draw $\phi_k \sim Dirichlet(\beta)$
  - Draw $\delta_k \sim Dirichlet(\rho)$
- For each document $d_{(\cdot),j}$ with $j \in \{1, \ldots, J\}$:
  - Draw topic mixture proportion $\theta_j \sim Dirichlet(\alpha)$
  - For each phrase $l \in [1, L_j]$:
    - Draw topic assignment $z_{j,l} \sim Multinomial(\theta_j)$
    - Draw head term $h_{j,l} \sim Multinomial(\phi_{z_{j,l}})$
    - Draw modifier term $m_{j,l} \sim Multinomial(\delta_{z_{j,l}})$.

Given a set of documents D, phrase $l =< h, m >$ is the observed variable, α, β and ρ are the Dirichlet prior parameters, and $z, \Theta = \{\theta, \phi, \delta\}$ are latent variables need to learn. Each document $d_{(\cdot),j}$ has an associated K-dimensional topic distribution θ_j, which describes the probability that document $d_{(\cdot),j}$ belongs to each topic. ϕ_k is a N_h-dimensional word distribution which describes the probability that topic k belongs to each head term h. Furthermore, δ_k is a N_m-dimensional word distribution which describes the probability that topic k belongs to each modifier term m. The graphical model is shown in Fig. 1(b). From the graphical model, the joint probability is:

$$
\begin{aligned}
&p(h, m, z, \Theta|\alpha, \beta, \rho) \\
&\qquad = \prod_{j=1}^{J} p(\theta_j|\alpha) \prod_{l=1}^{L_j} \sum_{z=1}^{K} p(h_l|\phi_z) p(m_l|\delta_z) p(z|\theta_j) \prod_{k=1}^{K} p(\phi_k|\beta) p(\delta_k|\rho).
\end{aligned} \tag{3}
$$

The probability of observing the review text given the model parameters Θ (i.e., the likelihood) is

$$p(D) = \prod_{j=1}^{J} \prod_{l=1}^{L_j} \sum_{k=1}^{K} p(z_{j,l}|\theta_j) p(h_{j,l}|\phi_k) p(m_{j,l}|\delta_k), \tag{4}$$

where D is the corpus of all reviews for all items. We employ Gibbs sampling to estimate the posterior probability given the observed phrases. For each phrase l in each document $d_{(\cdot),j}$, the sample function is as follow:

$$\begin{aligned} & p(z_l = k | z_{\neg l}, h, m, \alpha, \beta, \rho) \\ & \quad \propto (n_{j,\neg l,k} + \alpha) \cdot \frac{n_{h,\neg l,k} + \beta}{\sum_{h'=1}^{N_h} n_{h',\neg l,k} + N_h \beta} \cdot \frac{n_{m,\neg l,k} + \rho}{\sum_{m'=1}^{N_m} n_{m',\neg l,k} + N_m \rho}, \end{aligned} \tag{5}$$

where $z_{\neg l}$ is the topic allocation of all phrases except the l^{th} phrase, $n_{j,\neg l,k}$ is the number of phrases assigned to topic k for document $d_{(\cdot),j}$, excluding current phrase l. Furthermore, $n_{h,\neg l,k}$ and $n_{m,\neg l,k}$ are the numbers of head terms and modifier terms assigned to topic k, excluding current phrase l respectively.

Note that unlike traditional Latent Dirichlet Allocation (LDA) model, the observed data in the proposed topic model are phrases rather than single words. We made this choice for two reasons. First, the bag-of-words representation of reviews is not so effective in classifying texts by opinion. Second, head terms and modifier terms convey different information. The former indicates the aspect, while the latter is used to express sentiment on this aspect. So they should naturally follow different distributions and priors.

4.3. A Unified Model

So far, we have introduced methods to model overall ratings and reviews, respectively. On one hand, matrix factorization employs the rating information to obtain user-specific and item-specific factors. On the other hand, topic modeling can discover aspects hidden in reviews. It is a natural thought to combine matrix factorization and topic modeling for aspect mining, but it faces some challenges mentioned above. In the following, we will discuss these challenges and our solutions in detail.

The first challenge is what kinds of latent factors can be shared. As we know, the item latent factor v_j and the item topic distribution θ_j have similar physical meanings. Both of them represent the property distribution for item j on K latent topics. So aligning these two parameters to link matrix factorization (v_j shown in Fig. 1(a)) and topic model (θ_j shown in Fig. 1(b)) is a reasonable thought. Sharing the same item property distribution combines rating and review more tightly than defining a transformation function between them as McAuley and Leskovec did in (McAuley & Leskovec, 2013).

However, sharing the same item property distribution will face the scale mismatch challenge. As we know, the probability $\theta_{j,k} \in [0, 1]$, while $v_{j,k}$ can be arbitrary real value. So they have different scales. In addition, θ_j in the topic model has an explicit physical meaning of topic distribution while v_j in matrix factorization is unexplainable. Therefore, we use θ_j to replace v_j as the sharing factor, and the range remains [0, 1]. In order to make the other parameter u_i in matrix factorization have the similar physical meaning and scale as v_j, we introduce a preference probability $p_{i,j}$ which is the multiplication of μ_i and θ_j(Since the meaning and scale of u have changed, we use symbol μ to replace u in the proposed model).

$$p_{i,j} = \sum_{k=1}^{K} \mu_{i,k} \cdot \theta_{j,k}, \tag{6}$$

where $\mu_{i,k} \in [0, 1]$ and $p_{i,j} \in [0, 1]$. Thus, the item topic distribution θ_j can be shared by two models. It also has another benefit that p has an explicit physical meaning. That is, $p_{i,j}$ is the preference or matching between users and items. Note that the physical meaning of $p_{i,j}$ can also be expressed by rating $r_{i,j}$. Therefore, the optimization objective of matrix factorization is the preference probability $p_{i,j}$, rather than the rating score $r_{i,j}$. However, $p_{i,j}$ is positively correlated with $r_{i,j}$, so we define a transformation function f to map from $p_{i,j}$ to $r_{i,j}$. That is,

$$r_{i,j} = f(p_{i,j}). \tag{7}$$

The third challenge is how to model the dependency relation between rating and review information. To solve the problem, we establish a dependency relation from modifier term m to rating r. The reason lies in that modifier term m is the qualitative assessment of a user's sentiment, while rating r is the quantitative assessment of a user's sentiment. High ratings correspond to positive modifier terms, and vice versa. So it is reasonable to establish a direct dependency relation between r and m. In addition, considering the generative process of a review, we believe that users form an intuitive impression (good or bad) as soon as they experienced an item, which is reflected by the appropriate rating r. Only after the impression (rating) is formed will the user select modifier terms to express his/her feelings. So in the proposed model, m is influenced not only by z but also by r directly. Although it increases the dimension of the parameter δ from $K \times N_m$ to $R \times K \times N_m$, the dependency relation makes the distribution of m more reasonable.

As a consequence, we propose a unified model MaToAsp to combine rating and review information tightly for aspect mining. The probabilistic graphical model is shown in Fig. 1(c). The final loss function of the unified model is:

$$\begin{aligned}\mathcal{L} = &\sum_{j=1}^{J}\sum_{i=1}^{I}(r_{i,j} - f(p_{i,j}))^2 + \lambda(\sum_{j=1}^{J}\|\theta_j\|^2 + \sum_{i=1}^{I}\|\mu_i\|^2) \\ &- \omega \log \prod_{j=1}^{J}\prod_{l=1}^{L_j}(\sum_{k=1}^{K}\theta_{j,k}\phi_{k,h}\delta_{r_j,k,m}).\end{aligned} \tag{8}$$

Here the preference/matching between user i and item j is $p_{i,j} = \mu_i \cdot \theta_j$. We propose the transformation function f is a linear function $f(p_{i,j}) = R \cdot p_{i,j}$ (R is the upper limit of ratings, and it is 5 in our experiments), since this choice has achieved good performance in our experiments. The first part of Eq. 8, including the first and second term, is the error of the predicted ratings with regularization term, while the second part is the log likelihood of the review corpus. ω is a tradeoff parameter which controls the contribution of these two parts. K is the number of aspects or topics.

4.4. The Learning Algorithm

Given the overall rating and review information, the optimization objective is defined as follows

$$\arg \min_{\Theta,\Phi,z} \mathcal{L}(\Theta, \Phi, z). \tag{9}$$

Parameters Θ are associated with review, which can be learned by Gibbs sampling. The parameters Φ are associated with rating, and we learn them with the gradient descent method. As shown in Fig. 1(c), our model is the combination of the two parts. And thus, under an iterative optimization framework, we sequently optimize MaToAsp with the following two steps.

$$\textit{update}\ \Theta^{new}, \Phi^{new} = \arg \min_{\Theta,\Phi} \mathcal{L}(\Theta, \Phi, z^{old}) \tag{10}$$

$$\textit{sample } z_{j,l}^{new} \textit{ with probability } p(z_{j,l}^{new}=k)=\phi_{k,h_{j,l}}^{new}\cdot\delta_{r,k,m_{j,l}}^{new} \tag{11}$$

In the first step, the topic assignments z for phrases are fixed, and Θ and Φ are updated by gradient descent. The gradients are as follows:

$$\frac{\partial\mathcal{L}}{\partial\theta_{j,k}}=-2\sum_{i=1}^{I}(r_{i,j}-f(p_{i,j}))\frac{\partial f(p_{i,j})}{\partial\theta_{j,k}}+2\lambda\theta_{j,k}-\omega\sum_{l=1}^{L_j}\frac{\phi_{k,h}\cdot\delta_{r,k,m}}{\sum_{k=1}^{K}\theta_{j,k}\cdot\phi_{k,h}\cdot\delta_{r,k,m}}. \tag{12}$$

$$\frac{\partial\mathcal{L}}{\partial\mu_{i,k}}=-2\sum_{j=1}^{J}(r_{i,j}-f(p_{i,j}))\frac{\partial f(p_{i,j})}{\partial\mu_{i,k}}+2\lambda\mu_{i,k}. \tag{13}$$

In the second step, i.e., Eq.11 parameters Θ and Φ are fixed, and topic assignments z are sampled by iterating all documents and all phrases. We sample the topic by the following function, which is the probability of the topic k being used for product j, multiplied by the probability of phrase $l =< h, m >$ being used for the topic k.

$$\begin{aligned} & p(z_l=k|z_{\neg l},h,m,r,\alpha,\beta,\rho) \\ & \propto\theta_{j,k}\cdot\frac{n_{h,\neg l,k}+\beta}{\sum_{h'=1}^{N_h}n_{h',\neg l,k}+N_h\beta}\cdot\frac{n_{m,\neg l,k,r}+\rho}{\sum_{m'=1}^{N_m}n_{m',\neg l,k,r}+N_m\rho}. \end{aligned} \tag{14}$$

Here the topic probability $\theta_{j,k}$ is determined in the first step instead of being sampled from a Dirichlet prior ($n_{j,\neg l,k}+\alpha$ in Eq. 5). Additionally, we add one dimension r into the third term $\frac{n_{m,\neg l,k,r}+\rho}{\sum_{m'=1}^{N_m}n_{m',\neg l,k,r}+N_m\rho}$, which influences the distribution of m together with topic z.

Finally, the two steps are repeated until the end of iterations.

4.5. Aspect Rating Prediction

Based on the results calculated above, we can use ϕ and δ to predict aspect ratings. According to the prior probability, we can obtain the aspect that each phrase belongs to by:

$$\mathcal{G}(l=<h,m>)=\arg\max_k\sum_r\phi_{k,h_{j,l}}\cdot\delta_{r,k,m_{j,l}}. \tag{15}$$

For each item, the predicted aspect rating for each aspect is as follows:

$$\hat{a}_{j,k}=\frac{\sum_{<h,m>\in d_{(\cdot),j}}\sum_r r\cdot\phi_{k,h_{j,l}}\cdot\delta_{r,k,m_{j,l}}}{\sum_{<h,m>\in d_{(\cdot),j}}\sum_r\phi_{k,h_{j,l}}\cdot\delta_{r,k,m_{j,l}}}. \tag{16}$$

Table 1: Statistics of the datasets

Datasets	#Users	#Items	#Reviews	#Phrases	Avg. Overall Rating
Dianping	14,519	1,097	216,291	696,608	3.97
TripAdvisor	192,108	5,579	437,088	4,562,247	4.10

5. Experiments

In this section, we conduct extensive experiments to evaluate the effectiveness of our method on two real datasets.

5.1. Data Preparation

We conduct experiments on two real datasets, Dianping and TripAdvisor. The first dataset is crawled from the Dianping website, a well-known social media platform in China, which provides a review platform for businesses and entertainments. In the Dianping website, a user can give a review to a restaurant after enjoying a service in this business. Besides an overall rating, the review information includes Chinese comments and three aspect ratings on Taste, Service, and Environment, respectively. We selected the restaurants located in Beijing and the time span of the reviews is from Jan. 1^{st}, 2011 to Jun. 31^{th}, 2015. In addition, we also employ TripAdvisor dataset. Accompanying with English comments, reviews in this dataset are not only associated with overall ratings, but also with ground truth aspect ratings on 3 aspects: Value, Service, and Food. In the TripAdvisor dataset, we select the hotels and restaurants which are in New York and the time span of this dataset is from Jan. 1^{st}, 2011 to Sep. 30^{th}, 2015. We also delete some sluggish restaurants, inactive users and abnormally active users, from the two original datasets. Note that, all the ratings in these two datasets are in the range from 1 to 5. The statistic information of these two datasets is shown in Table 1.

To fit the model, the data is preprocessed via the following procedures. The extracting process of TripAdvisor is similar to that in (Luo et al., 2014). The detailed steps are as follows: (1) Apply POS Tagging[1] to tag POS for each word in each review; (2) Select phrases according to the tagging and the rules as in Table 2; (3) Apply Porter Stemmer[2] to stem the phrases; (4) Combine the processed data into the phrase $l =< h, m >$. For Dianping dataset, the extracting process is partially different. Chinese does not need

[1]http://opennlp.sourceforge.net/
[2]http://www.tartarus.org/martin/PorterStemmer/

Table 2: Rules for TripAdvisor

No.	Rule
1.	$< noun, adjective >$
2.	$< adjective, adverb >$
3.	$< verb\ and\ past\ participle, adverb >$
4.	$< verb, verb\ and\ past\ tense >$
5.	$< verb\ and\ past\ tense, adverb >$

Table 3: Rules for Dianping

No.	Rule
1.	$amod(N, A) \rightarrow < N, A >$
2.	$acomp(V, A) + nsubj(V, N) \rightarrow < N, A >$
3.	$cop(A, V) + nsubj(A, N) \rightarrow < N, A >$
4.	$dobj(V, N) + nsubj(V, N') \rightarrow < N, V >$
5.	$< h_1, m > +conj\_and(h_1, h_2) \rightarrow < h_2, m >$
6.	$< h, m_1 > +conj\_and(m_1, m_2) \rightarrow < h, m_2 >$
7.	$< h, m > +neg(m, not) \rightarrow < h, not + m >$

Table 4: Prior words for aspect prior

Dataset	Category	Prior Words
Dianping	Taste	taste, flavor, dish, dishes
	Service	serving, attitude, waitress, service
	Environment	environment, location, room, decoration
TripAdvisor	Value	value, price, quality, worth
	Service	service, attitude, waiter, waitress
	Food	food, taste, dish, dinner

stemming, but need Word Segmenter[3] first. The rules for Chinese are from (Moghaddam & Ester, 2012), which are listed in Table 3.

To inject the prior knowledge of an aspect, we select some words as prior for each aspect; Table 4 lists some of these prior words. In order to facilitate understanding, we translate the Chinese words in Dianping into English.

The parameters used in the experiments are as follows. The Dirichlet priors α, β and ρ are set as 2, 0.5 and 0.5, respectively. The learning rate in gradient descent is 0.005, and the regularization coefficient λ in Eq. 8 is 0.01. The tradeoff parameter ω is set as 0.1 and 0.5 for Dianpings and TripAdvisors, respectively. Note that, since the topics should be mapped to the real-world aspects, we set K as 3 for Dianping datasets and TripAdvisor datasets, respectively.

5.2. Evaluation Metric

We select Root Mean Square Error (RMSE) and Person Correlation Coefficient (PCC) to evaluate the effectiveness of our model.

As is known to all, the lower the difference between real values and predicted values, the better the performance of aspect rating prediction is. RMSE (Gunawardana & Shani, 2009) is used to compare the predicted values with the real values. We have the real aspect rating vector $a_{j,k}$ for every

[3] http://nlp.stanford.edu/software/segmenter.shtml

item j as ground truth. Then we predict the aspect ratings $\hat{a}_{j,k}$ for every item j by our model. The function of RMSE is as follows:

$$RMSE = \sqrt{\frac{\sum_{j=1}^{J}\sum_{k=1}^{K}(\hat{a}_{j,k} - a_{j,k})^2}{J * K}}. \tag{17}$$

Smaller value of RMSE indicates a stronger predictor. Note that the RMSE criterion measures the average difference of predicted ratings and real ratings on all aspects.

Considering that aspect ratings are often used for recommendation, it is necessary to measure the relative ordering of products based on the predicted aspect rating and the real aspect. PCC (Gunawardana & Shani, 2009) is often a good choice. The corresponding function is defined as follows:

$$PCC = \frac{(J * K \sum \hat{a}_{j,k} a_{j,k} - \sum \hat{a}_{j,k} \sum a_{j,k})}{\sqrt{J * K \sum (\hat{a}_{j,k})^2 - (\sum \hat{a}_{j,k})^2} \cdot \sqrt{J * K \sum (a_{j,k})^2 - (\sum a_{j,k})^2}}. \tag{18}$$

As is shown in Eq. 18, the higher the PCC value, the better the performance of aspect rating prediction is. And the relevance is stronger when the absolute value of PCC is closer to 1, and weaker when the absolute value of PCC is closer to 0.

5.3. Comparison Methods

We compare the proposed model with three representative phrase-level methods.

- QPLSA (Luo et al., 2015) uses quad-tuples information to build a model based on PLSA framework. The model not only generates fine-granularity aspects of items, but also captures the relationship between phrases and ratings.

- GRAOS (Luo et al., 2014) is a semi-supervised model based on LDA framework. It also uses the quad-tuples information to capture the relationship between phrases and ratings.

- SATM (Wang & Ester, 2014) is a sentiment-aligned model based on LDA framework. The model uses two kinds of external knowledge: the item-level overall rating distribution and a word-level sentiment lexicon.

Table 5: Representative phrases for different aspects on two datasets

Datasets	Aspects	Representative Phrases(Ratings)
Dianping	Taste	amazing taste(4.55), delicious beef(4.45), standard dish(3.61), terrible food(2.18)
	Service	nice service(4.70), free parking(4.08), clumsy waiter(2.77), bad attitude(1.96)
	Environment	elegant environment(4.72), sumptuous decoration(4.41), remote location(3.10), old facility(2.29)
TripAdvisor	Value	superior value(4.63), great price(4.37), reasonable price(4.12), pricey fare(3.16)
	Service	best attitude(4.78), interesting waiter(4.27),unfriendly waitress(2.57), disgusting service(1.36)
	Food	incredible meat(4.57), delicious food(4.49), overcooked chicken(3.30), not fresh dish(2.33)

5.4. Aspect Identification

Firstly, we validate the effectiveness of aspect identification through a case study. Note that since these quite large phrases are not labeled and the time cost for manually annotating aspects of phrases is unbearable, it is hard to quantitatively evaluate the performance of aspect identification. In this experiment, we list the top 20 automatically mined phrases for each aspect on two datasets respectively, from which we select several meaningful phrases to be shown in Table 5. Specifically, we list 4 phrases and rank them by their scores in the descending order for each aspect. Moreover, Chinese phrases in Dianping are translated into English in order to facilitate understanding.

In general, the phrases are labeled with corresponding aspects, and the predicted ratings conform to the sentiments that the phrases express. On one hand, head terms represent the corresponding aspects, such as “price”, “attitude” and “meat” for aspects “Value”, “Service” and “Food” respectively. When other users see the head terms, they will understand the aspects which were talked about. On the other hand, the predicted ratings conform to the sentiments that the modifier terms express. That is, positive modifier terms tend to get high ratings, and vice versa. For example, for the “Food” aspect of TripAdvisor dataset, the phrase “incredible meat” gets a relatively high score (4.57) because “incredible” is a positive modifier term that expresses positive sentiment. While the phrase “notfresh dish” only gets a score of 2.33, since “not fresh” is a negative modifier term that expresses negative sentiment. The reasonable experimental results on both Chinese and English datasets validate the effectiveness of aspect identification qualitatively.

5.5. Accuracy Experiment

Then we validate the performance of the different methods in terms of RMSE PCC criteria. The number of aspects (topics) K is set as 3 for Dianping as well as TripAdvisor. The experiments are performed on different sizes of training datasets (i.e., 25%, 50%, 75%, and 100% of review data). The average results of ten runs are recorded.

Table 6: RMSE performance of different methods on two datasets. The right part shows the performance improvement of MaToAsp against other methods.

Dataset			QPLSA	GRAOS	SATM	MaToAsp	QPLSA	GRAOS	SATM
Dianping	25%	Mean	0.5812	0.4778	0.5796	**0.4520**	22.23%	5.40%	22.02%
		Dev.	0.0015	0.0032	0.0028	0.0023			
	50%	Mean	0.5793	0.4675	0.5766	**0.4457**	23.06%	4.66%	22.70%
		Dev.	0.0023	0.0029	0.0034	0.0013			
	75%	Mean	0.5708	0.4633	0.5639	**0.4439**	22.23%	4.19%	21.28%
		Dev.	0.0049	0.0015	0.0046	0.0035			
	100%	Mean	0.5642	0.4502	0.5590	**0.4381**	22.35%	2.69%	21.63%
		Dev.	0.0037	0.0017	0.0042	0.0029			
TripAdvisor	25%	Mean	0.6096	0.5720	0.5602	**0.5404**	11.35%	5.52%	3.53%
		Dev.	0.0034	0.0040	0.0039	0.0036			
	50%	Mean	0.5654	0.5587	0.4972	**0.4797**	15.16%	14.14%	3.52%
		Dev.	0.0060	0.0028	0.0065	0.0036			
	75%	Mean	0.5066	0.5510	0.4695	**0.4517**	10.84%	18.02%	3.79%
		Dev.	0.0026	0.0018	0.0069	0.0042			
	100%	Mean	0.4870	0.5415	0.4336	**0.4208**	13.59%	22.29%	2.95%
		Dev.	0.0011	0.0010	0.0045	0.0020			

5.5.1. RMSE Performance

Table 6 shows the aspect rating prediction performance of different methods on all aspects through RMSE criterion. The RMSE value is calculated by Eq. 17, and it measures the average difference between real scores and predicted scores of all items on all aspects. As shown in Table 6, we can find the following observations.

- MaToAsp achieves the best performance on all datasets. Moreover, its low standard deviations validate its stability. We think the good and stable performance attests to the appropriateness of the strategies employed in MaToAsp.

- We can also observe that GAROS performs well on the Dianping dataset, but badly on TripAdvisor dataset. In contrast, SATM performs well on the TripAdvisor dataset, but badly on Dianping dataset. The phenomenon shows that both GAROS and SATM are sensitive to the datasets, which implies that they may be not robust. However, the proposed MaToAsp always has good performance on these two datasets, which further reflects that MaToAsp is a stable and robust algorithm.

- In addition, with the increment of review data, the accuracy of all methods increases steadily because more information is available. However, we can observe that MaToAsp offers more significant performance improvements in smaller training data on both datasets. It shows that MaToAsp may be better suited for sparse data. We think the reason

Table 7: Pearson correlation coefficient of different methods on two datasets. The right part shows the performance improvement of MaToAsp against other methods.

Dataset			QPLSA	GRAOS	SATM	MaToAsp	QPLSA	GRAOS	SATM
Dianping	25%	**Mean**	0.5789	0.1290	0.3523	**0.5873**	1.45%	355.3%	66.70%
		Dev.	0.0121	0.0051	0.0097	0.0072			
	50%	**Mean**	0.5815	0.1287	0.3606	**0.5925**	1.89%	360.4%	64.31%
		Dev.	0.0098	0.0138	0.0116	0.0055			
	75%	**Mean**	0.5846	0.1328	0.3745	**0.6014**	2.87%	352.9%	60.59%
		Dev.	0.0057	0.0085	0.0119	0.0053			
	100%	**Mean**	0.5982	0.1388	0.3914	**0.6172**	3.18%	344.7%	57.70%
		Dev.	0.0183	0.0034	0.0087	0.0037			
TripAdvisor	25%	**Mean**	0.5190	0.2049	0.5109	**0.5254**	1.23%	156.4%	2.84%
		Dev.	0.0090	0.0071	0.0070	0.0070			
	50%	**Mean**	0.5484	0.2335	0.5350	**0.5548**	1.17%	137.6%	3.70%
		Dev.	0.0248	0.0094	0.0128	0.0091			
	75%	**Mean**	**0.5870**	0.2522	0.5766	0.5792	-1.33%	129.7%	0.45%
		Dev.	0.0043	0.0094	0.0124	0.0059			
	100%	**Mean**	0.5774	0.2764	0.5817	**0.5880**	1.84%	112.7%	1.08%
		Dev.	0.0150	0.0024	0.0073	0.0053			

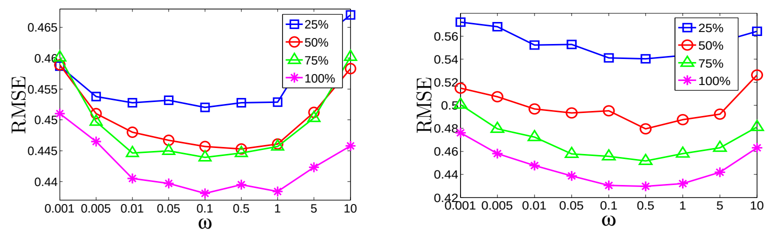
lies in that MaToAsp makes better use of review data, compared to the other methods we tested.

5.5.2. Relative Order Performance

In this section, we verify the ability of the different methods to maintain the relative order among items with PCC (see Eq. 18). The results are shown in Table 7. That is, MaToAsp obtains the highest PCC in almost all datasets. Moreover, all methods have worse performance for smaller dataset, while MaToAsp has comparatively better performance for sparse data. Note that RMSE and PCC evaluate the quality of solutions on different aspects. Some methods have significantly different performance on these two criteria (e.g., QPLSA and GRAOS). However, MaToAsp steadily performs best. Once again, this validates that MaToAsp is more effective and steady to model the correlations between aspects and ratings, and thus better maintains aspect ranking orders compared to other methods. The results also imply that the proposed method is very promising for aspect-level recommender systems, since it can generate very similar item order to the real order.

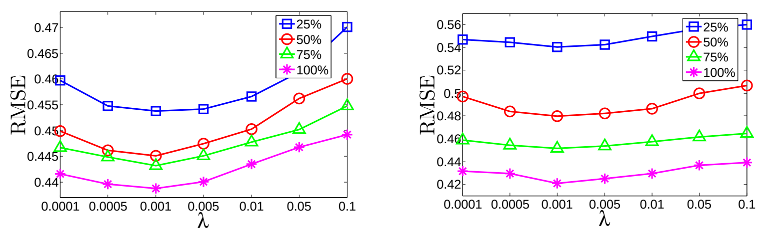
5.6. Parameter Experiment

Parameter ω is an important parameter that controls the contribution of ratings and reviews, as shown in Eq. 8. In this section, we investigate the sensitivity of MaToAsp to ω. We vary the value of ω in {0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10}, and calculate the RMSE results of different-size datasets on both Dianping and TripAdvisor. The results are shown in Fig.



(a) Dianping (b) TripAdvisor

Figure 2: RMSE performance of MaToAsp by varying the value of the tradeoff parameter ω on Dianping and TripAdvisor



(a) Dianping (b) TripAdvisor

Figure 3: RMSE performance of MaToAsp by varying the value of the tradeoff parameter λ on Dianping and TripAdvisor

2. From Fig. 2, we can observe that the performance of MaToAsp firstly increase and then decrease with the increment of ω. This is reasonable, since the proper balance of matrix factorization and topic modeling will have better performance. The MaToAsp performs best when $\omega = 0.1$ and $\omega = 0.5$ on Dianping and TripAdvisor, respectively. So we set the ω as 0.1 and 0.5 in the experiments.

Parameter λ is also an important parameter that controls the influence of regularization coefficient, as shown in Eq. 8. We conduct similar experiments on λ. The results are shown in Fig. 3. Similarly, the performance first increases, and then decreases with the increment of λ. The MaToAsp performs best when $\lambda = 0.001$ on both datasets. So we set the λ as 0.001 in the experiments.

5.7. Convergence Experiment

Finally, we conduct a convergence experiment. The X-axis shows the number of iterations, and the Y-axis show the loss of Dianping (left) and TripAdvisor (right), respectively. The results are shown in Fig. 4. MaToAsp converges after 4,000 iterations on Dianping, and 5,500 iterations on

TripAdvisor. The results show that MaToAsp converges steadily.

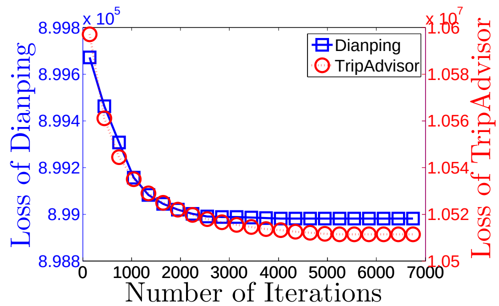


Figure 4: Performance of MaToAsp by varying the number of iterations

6. Conclusion

This paper proposes the novel MaToAsp model to effectively integrate matrix factorization and topic modeling for aspect rating prediction. Based on these two models, the MaToAsp shares the latent factor of items, introduces an interpretive preference probability to eliminate scale mismatch, and implements a dependency relation from the rating to the modifier terms to make the distribution of modifier term more reasonable. In addition, we design an iterative optimization framework to solve the designed optimization objective. The experiments on two real datasets including Chinese and English validate the effectiveness of the proposed MaToAsp. We consider this work is meaningful enough. On one hand, it is the first attempt to integrate matrix factorization and topic modeling for aspect mining, where the latent topics of matrix factorization can be explained as the aspects of shops. On the other hand, the accurate aspect rating prediction can not only help consumers make choices but also help shops improve their qualities, like service and environment.

7. Acknowledgments

This work is supported in part by the National Natural Science Foundation of China (No. 61375058), National Key Basic Research and Department (973) Program of China (No. 2013CB329606), and the Co-construction Project of Beijing Municipal Commission of Education.

References

Balazs, J. A., & Velásquez, J. D. (2016). Opinion mining and information fusion: a survey. *Information Fusion*, *27*, 95–110.

Bao, Y., Fang, H., & Zhang, J. (2014). Topicmf: Simultaneously exploiting ratings and reviews for recommendation. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada.* (pp. 2–8).

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *JASIS*, *41*, 391–407.

Diao, Q., Qiu, M., Wu, C., Smola, A. J., Jiang, J., & Wang, C. (2014). Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, New York, NY, USA - August 24 - 27, 2014* (pp. 193–202).

Gunawardana, A., & Shani, G. (2009). A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research*, *10*, 2935–2962.

Guo, G., Zhang, J., & Yorke-Smith, N. (2016). A novel recommendation model regularized with user trust and item ratings. *IEEE Trans. Knowl. Data Eng.*, *28*, 1607–1620.

Hu, Y., Chen, Y., & Chou, H. (2017). Opinion mining from online hotel reviews - A text summarization approach. *Inf. Process. Manage.*, *53*, 436–449.

Jin, Z., Li, Q., Zeng, D. D., Zhan, Y., Liu, R., Wang, L., & Ma, H. (2016). Jointly modeling review content and aspect ratings for review rating prediction. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016* (pp. 893–896).

Koren, Y., Bell, R. M., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *IEEE Computer*, *42*, 30–37.

Li, H., Lin, R., Hong, R., & Ge, Y. (2015a). Generative models for mining latent aspects and their ratings from short reviews. In *2015 IEEE International Conference on Data Mining, ICDM 2015, Atlantic City, NJ, USA, November 14-17, 2015* (pp. 241–250).

Li, S., Zhou, L., & Li, Y. (2015b). Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures. *Inf. Process. Manage.*, *51*, 58–67.

Ling, G., Lyu, M. R., & King, I. (2014). Ratings meet reviews, a combined approach to recommend. In *Eighth ACM Conference on Recommender Systems, RecSys '14, Foster City, Silicon Valley, CA, USA - October 06 - 10, 2014* (pp. 105–112).

Lu, Y., Zhai, C., & Sundaresan, N. (2009). Rated aspect summarization of short comments. In *Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009* (pp. 131–140).

Luo, W., Zhuang, F., Cheng, X., He, Q., & Shi, Z. (2014). Ratable aspects over sentiments: Predicting ratings for unrated reviews. In *2014 IEEE International Conference on Data Mining, ICDM 2014, Shenzhen, China, December 14-17, 2014* (pp. 380–389).

Luo, W., Zhuang, F., Zhao, W., He, Q., & Shi, Z. (2015). QPLSA: utilizing quad-tuples for aspect identification and rating. *Inf. Process. Manage.*, *51*, 25–41.

McAuley, J. J., & Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Seventh ACM Conference on Recommender Systems, RecSys '13, Hong Kong, China, October 12-16, 2013* (pp. 165–172).

Moghaddam, S., & Ester, M. (2012). On the design of LDA models for aspect-based opinion mining. In *21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012* (pp. 803–812).

Moghaddam, S., & Ester, M. (2013). The FLDA model for aspect-based opinion mining: addressing the cold start problem. In *22nd International*

World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013 (pp. 909–918).

Qiu, L., Gao, S., Cheng, W., & Guo, J. (2016). Aspect-based latent factor model by integrating ratings and reviews for recommender system. *Knowl.-Based Syst.*, *110*, 233–243.

Salakhutdinov, R., & Mnih, A. (2007). Probabilistic matrix factorization. In *Advances in Neural Information Processing Systems 20, Proceedings of the Twenty-First Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007* (pp. 1257–1264).

Srebro, N., & Jaakkola, T. S. (2003). Weighted low-rank approximations. In *Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA* (pp. 720–727).

Taheri, S. M., Mahyar, H., Firouzi, M., K., E. G., Grosu, R., & Movaghar, A. (2017). Extracting implicit social relation for social recommendation techniques in user rating prediction. In *Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, April 3-7, 2017* (pp. 1343–1351).

Titov, I., & McDonald, R. (2008). A joint model of text and aspect ratings for sentiment summarization. *proceedings of ACL-08: HLT*, (pp. 308–316).

Wang, C., & Blei, D. M. (2011). Collaborative topic modeling for recommending scientific articles. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011* (pp. 448–456).

Wang, H., & Ester, M. (2014). A sentiment-aligned topic model for product aspect rating prediction. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL* (pp. 1192–1202).

Wang, H., Lu, Y., & Zhai, C. (2010). Latent aspect rating analysis on review text data: a rating regression approach. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, July 25-28, 2010* (pp. 783–792).

Wang, H., Lu, Y., & Zhai, C. (2011). Latent aspect rating analysis without aspect keyword supervision. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011* (pp. 618–626).

Xu, H., Zhang, F., & Wang, W. (2015). Implicit feature identification in chinese reviews using explicit topic mining model. *Knowl.-Based Syst.*, *76*, 166–175.

Xue, W., Li, T., & Rishe, N. (2017). Aspect identification and ratings inference for hotel reviews. *World Wide Web*, *20*, 23–37.

Yu, D., Mu, Y., & Jin, Y. (2017). Rating prediction using review texts with underlying sentiments. *Information Processing Letters*, *117*, 10–18.

Yu, J., Zha, Z.-J., Wang, M., & Chua, T.-S. (2011). Aspect ranking: identifying important product aspects from online consumer reviews. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 1496–1505). Association for Computational Linguistics.

Zhang, W., & Wang, J. (2016). Integrating topic and latent factors for scalable personalized review-based rating prediction. *IEEE Transactions on Knowledge and Data Engineering*, *28*, 3013–3027.

Zheng, X., Lin, Z., Wang, X., Lin, K., & Song, M. (2014). Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification. *Knowl.-Based Syst.*, *61*, 29–47.