World Wide Web (2019) 22:221–240
https://doi.org/10.1007/s11280-018-0558-1

Explainable recommendation with fusion of aspect information

Yunfeng Hou[1] · Ning Yang[1] · Yi Wu[1] · Philip S. Yu[2,3]

Received: 28 September 2017 / Revised: 21 March 2018 / Accepted: 1 April 2018 /
Published online: 13 April 2018
© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract Explainable recommendation has attracted increasing attention from researchers. The existing methods, however, often suffer from two defects. One is the lack of quantitative fine-grained explanations why a user chooses an item, which likely makes recommendations unconvincing. The other one is that the fine-grained information such as aspects of item is not effectively utilized for making recommendations. In this paper, we investigate the problem of making quantitatively explainable recommendation at aspect level. It is a nontrivial task due to the challenges on quantitative evaluation of aspect and fusing aspect information into recommendation. To address these challenges, we propose an Aspect-based Matrix Factorization model (AMF), which is able to improve the accuracy of rating prediction by collaboratively decomposing the rating matrix with the auxiliary information extracted from aspects. To quantitatively evaluate aspects, we propose two metrics: User Aspect Preference (UAP) and Item Aspect Quality (IAQ), which quantify user preference to a specific aspect and the review sentiment of item on an aspect, respectively. By UAP and IAQ, we can quantitatively explain why a user chooses an item. To achieve information incorporation, we assemble UAPs and IAQs into two matrices UAP Matrix (UAPM) and IAQ Matrix (IAQM),

✉ Ning Yang
yangning@scu.edu.cn

Yunfeng Hou
YunFengHarry@163.com

Yi Wu
wuyihyper.scu@gmail.com

Philip S. Yu
psyu@uic.edu

[1] College of Computer Science, Sichuan University, Chengdu, China

[2] Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA

[3] Institute for Data Science, Tsinghua University, Beijing, China

respectively, and fuse UAPM and IAQM as constraints into the collaborative decomposition of item rating matrix. The extensive experiments conducted on real datasets verify the recommendation performance and explanatory ability of our approach.

Keywords Explainable recommendation · Recommender system · Matrix factorization

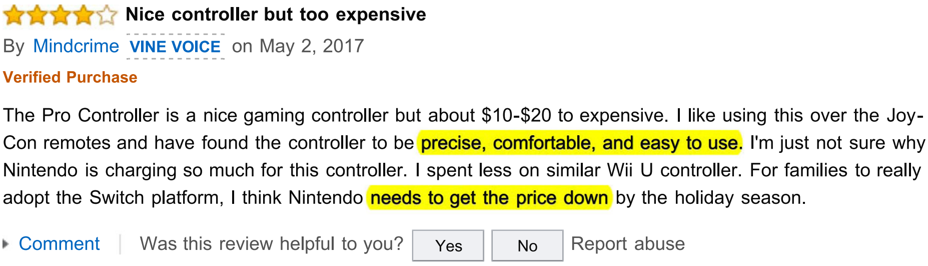
1 Introduction

Explainable recommendation has been attracting increasing attention in recent years [42]. The existing methods for explainable recommendation [1, 21, 42] often suffer from two defects. One is the lack of quantitative fine-grained explanations why a user chooses an item, which likely leads to less convincing recommendations. The other one is that fine-grained information is not effectively fused into the process of recommendation, and there is still room to improve the accuracy of rating prediction.

Recently, some methods utilizing aspect information extracted from review text have been proposed [10, 22, 39, 46], where two types of aspects are defined. One is defined as a noun word or phrase representing an item feature [10, 46]. The other type of aspect is defined as a set of words that characterize a topic of item in the reviews [22, 39]. Existing methods generally exploit the first type of aspect for explainable recommendation, which makes it hard to contrastively analyze recommendatory reasons because aspects may be different among different items. Figure 1 gives an illustration of the second type of aspect, from which we can observe that the user focuses two aspects of the item: the operation aspect and the price aspect. When we recommend this item to user with corresponding aspect evaluation, it will strengthen the recommendatory persuasion.

In this paper, we investigate the problem of making quantitatively explainable recommendation at aspect level, where we take the second definition of aspect since item topic is able to offer fine-grained aspect information for more convincing explainable recommendations. However, it is a nontrivial task due to the following challenges.

- **Quantitative Evaluation of Aspect** In order to make recommendatory explanation more convincing, we need to quantitatively evaluate aspects. Even if two users assign the same overall rating score for an item, they possibly have different preferences to aspects. For example, suppose two users book rooms of the same hotel and they assign the same rating score for the item. One of them may care more about the location aspect, and the other may mainly consider the service aspect. Similarly, for two items that have



★★★★☆ **Nice controller but too expensive**

By Mindcrime VINE VOICE on May 2, 2017

Verified Purchase

The Pro Controller is a nice gaming controller but about $10-$20 to expensive. I like using this over the Joy-Con remotes and have found the controller to be precise, comfortable, and easy to use. I'm just not sure why Nintendo is charging so much for this controller. I spent less on similar Wii U controller. For families to really adopt the Switch platform, I think Nintendo needs to get the price down by the holiday season.

Figure 1 Aspect example

the same overall rating score, their review sentiment on each aspect may be different. We need to capture the differences at aspect level.

- **Fusing Aspect Information into Recommendation** Rating prediction is the important task in recommendation. Review text provides a wealth of relevant information beyond ratings. How to combine aspect information learned from review text with current recommendation technique for improving the accuracy of rating prediction is a challenge. We need a suitable approach to incorporate these aspect evaluation information.

To address these challenges, in this paper, we propose an Aspect-based Matrix Factorization model, called AMF. The main idea of AMF is fusing auxiliary aspect information into matrix factorization to improve the accuracy of rating prediction. Figure 2 shows the work flow of AMF, which contains two parts. In the first part, to quantitatively evaluate aspects, we propose two metrics: User Aspect Preference (UAP) and Item Aspect Quality (IAQ). UAP reveals user preference to a specific aspect and IAQ assesses the review sentiment of item on a specific aspect. By UAP and IAQ, we can quantitatively explain why a user chooses an item. We assemble UAPs and IAQs into two matrices UAP Matrix (UAPM) and IAQ Matrix (IAQM), respectively. UAPM and IAQM code the auxiliary aspect information which are used as constraints fused into the factorization of item rating matrix, as shown in the second part in Figure 2. By collaboratively decomposing original rating matrix with the auxiliary information extracted from aspects, AMF is able to improve the accuracy of rating prediction.

Our main contributions can be summarized as follows:

1. We propose an Aspect-based Matrix Factorization model (AMF) for explainable recommendation, which can improve the accuracy of rating prediction by fusing auxiliary information extracted from aspects into the factorization of item rating matrix.
2. We propose two metrics, User Aspect Preference (UAP) and Item Aspect Quality (IAQ), by which we can quantify the user preference to an aspect and assess the review

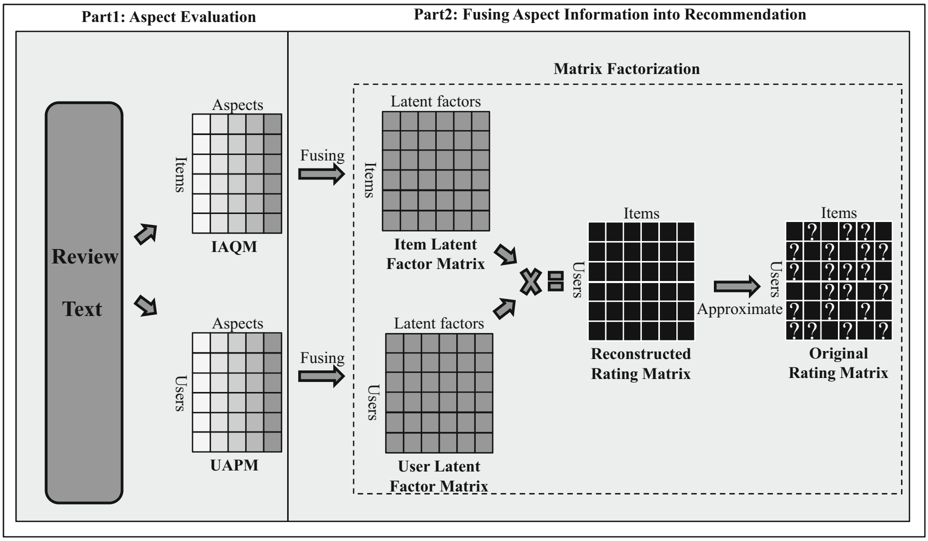


Figure 2 The work flow of AMF

sentiment of item on an aspect, so as to make the recommendations explainable and convincing.

3. The extensive experiments conducted on real datasets verify the recommendation performance and explanatory ability of our approach.

In the rest of the paper, we describe the details of aspect evaluation in Section 2. The details of AMF are presented in Section 3. We analyze the experimental results in Section 4. We review the related works in Section 5 and conclude in Section 6. Table 1 summarizes the notations used in this paper.

2 Aspect evaluation

In this section, we present how to quantitatively evaluate aspects based on review dataset. At first we formally define the aspect as follows:

Definition 1 (Aspect): An aspect $\mathcal{A}$ is a set of related words that characterize a topic of item in reviews,

$$\mathcal{A} = \{w_1, ..., w_{J_{\mathcal{A}}}\}\ (w_i \in V, 1 \le i \le J_{\mathcal{A}}), \tag{1}$$

Table 1 Notations

Symbol	Description
$\mathcal{A}$	aspect
M	the number of users
N	the number of items
L	the number of aspects
K	the number of latent factors
D	the highest rating score in dataset
$p_m^{\mathcal{A}}$	the preference of user m to aspect $\mathcal{A}$
$c_m^{\mathcal{A}}$	the mentioned number of aspect $\mathcal{A}$ in all reviews of user m
c_m	the review number of user m
$q_n^{\mathcal{A}}$	the review sentiment of item n on aspect $\mathcal{A}$
$a_n^{\mathcal{A}}$	the attention factor of item n on aspect $\mathcal{A}$
$s_n^{\mathcal{A}}$	the sentiment factor of item n on aspect $\mathcal{A}$
$c_n^{\mathcal{A}}$	the mentioned number of aspect $\mathcal{A}$ in all reviews of item n
c_n	the review number of item n
r_n	the average rating of item n
$\boldsymbol{R}$	the rating matrix of user-item, $\boldsymbol{R} \in \mathbb{R}^{M \times N}$
$\boldsymbol{P}$	User Aspect Preference Matrix (UAPM), $\boldsymbol{P} \in \mathbb{R}^{M \times L}$
$\boldsymbol{Q}$	Item Aspect Quality Matrix (IAQM), $\boldsymbol{Q} \in \mathbb{R}^{N \times L}$
$\boldsymbol{U}$	user latent factor matrix, $\boldsymbol{U} \in \mathbb{R}^{M \times K}$
$\boldsymbol{V}$	item latent factor matrix, $\boldsymbol{V} \in \mathbb{R}^{N \times K}$
$\boldsymbol{X}$	latent factor matrix between $\boldsymbol{P}$ and $\boldsymbol{U}$, $\boldsymbol{X} \in \mathbb{R}^{L \times K}$
$\boldsymbol{Y}$	latent factor matrix between $\boldsymbol{Q}$ and $\boldsymbol{V}$, $\boldsymbol{Y} \in \mathbb{R}^{L \times K}$

where w_i is a word in vocabulary V of review dataset and $J_{\mathcal{A}}$ is the number of related words of aspect $\mathcal{A}$. For example, "music", "sound", and "singing" are related words on music aspect.

According to the concept of aspect, we extract aspects (denoted by $\mathcal{A}_1$, ..., $\mathcal{A}_L$) from review dataset and map the words of each review onto corresponding aspects. For the first purpose, we leverage Latent Dirichlet Allocation (LDA) [41] to model the topics of review dataset. By LDA, we obtain the topics and their corresponding topic words, and we regard each topic in LDA as an aspect. We will further discuss the influence of aspect number in Section 4.2. To map words of each review into corresponding aspects, we utilize Aspect Segmentation Algorithm (ASA) proposed by Wang et al. [39] and use the topic words as keywords of corresponding aspect in the input of ASA.

Intuitively, users have different preferences to different aspects. Likewise, for items, different reviews have different sentimental orientation for different aspects. For example, some people may comment the color aspect of an item with positive words like "beautiful", while some other people may comment it with negative words like "ugly". Inspired by these observations, we propose two metrics: User Aspect Preference (UAP) and Item Aspect Quality (IAQ), to measure the user preference to an aspect and the review sentiment of one item on an aspect, respectively. The formal definitions of UAP and IAQ are given as follows:

Definition 2 (User Aspect Preference (UAP)): The preference of user m to aspect $\mathcal{A}$ is defined as:

$$p_m^{\mathcal{A}} = \frac{c_m^{\mathcal{A}}}{c_m} \times \frac{(D-1)}{1+e^{-c_m}} + 1, \tag{2}$$

where $c_m^{\mathcal{A}}$ is the mentioned times of aspect $\mathcal{A}$ in all reviews of user m, and c_m is the review number of user m.

$\frac{c_m^{\mathcal{A}}}{c_m}$ reflects the preference degree of user m to aspect $\mathcal{A}$, i.e., the importance of aspect $\mathcal{A}$ to user m. The more the user cares about an aspect, the more that aspect is mentioned. To depict user preference more accurately, we exploit a logistic function to regularize c_m. We rescale $\frac{c_m^{\mathcal{A}}}{c_m} \times \frac{1}{1+e^{-c_m}}$ into value range $[1, D)$, where D is assigned 5 in correspondence with the highest rating score in many real recommender systems.

Definition 3 (Item Aspect Quality (IAQ)): The review sentiment of item n on aspect $\mathcal{A}$ is defined as:

$$q_n^{\mathcal{A}} = a_n^{\mathcal{A}} \times s_n^{\mathcal{A}}, \tag{3}$$

where $a_n^{\mathcal{A}}$ is the attention factor and $s_n^{\mathcal{A}}$ is the sentiment factor.

Equation (4) and (5) below give the computational formulas of $a_n^{\mathcal{A}}$ and $s_n^{\mathcal{A}}$. $a_n^{\mathcal{A}}$ and $s_n^{\mathcal{A}}$ reveal collective attention (i.e., concerned degree) and sentiment (i.e., satisfied degree) of users who review item n on aspect $\mathcal{A}$ respectively. The $a_n^{\mathcal{A}}$ is defined as:

$$a_n^{\mathcal{A}} = \frac{c_n^{\mathcal{A}}}{c_n} \times \frac{1}{1+e^{-c_n}}, \tag{4}$$

where $c_n^{\mathcal{A}}$ is the mentioned times of aspect $\mathcal{A}$ in all reviews of item n and c_n is the review number of item n.

$\frac{c_n^{\mathcal{A}}}{c_n}$ reflects concerned degree of aspect $\mathcal{A}$ for item n. The more the aspect is mentioned, the larger the fraction is. To depict concerned degree more accurately, we exploit a logistic function to regularize c_n. The value range of $a_n^{\mathcal{A}}$ is [0, 1].

The $s_n^{\mathcal{A}}$ is defined as:

$$s_n^{\mathcal{A}} = \sum_{j=1}^{J_{\mathcal{A}}} \beta_n^{\mathcal{A},w_j} \times f_n^{\mathcal{A},w_j}, \tag{5}$$

where $\beta_n^{\mathcal{A},w_j}$ is the sentiment score of word w_j related to aspect $\mathcal{A}$ of item n and $f_n^{\mathcal{A},w_j}$ is the frequency of word w_j related to aspect $\mathcal{A}$ of item n.

In order to compute $\beta_n^{\mathcal{A},w_j}$, we use the context-aware model Latent Rating Regression (LRR) [39], which explores the sentiment score of the word by considering contextual factors instead of directly assigning the constant sentiment score from lexicon in previous work [2]. In LRR, r_n (the average rating of item n) is assumed to be a sample drawn from a Gaussian distribution:

$$r_n \sim N(\sum_{i=1}^{L} \alpha_n^{\mathcal{A}_i} \sum_{j=1}^{J_{\mathcal{A}_i}} \beta_n^{\mathcal{A}_i,w_j} f_n^{\mathcal{A}_i,w_j}, \delta^2), \tag{6}$$

where $\alpha_n^{\mathcal{A}_i}$ and δ^2 are weight of aspect $\mathcal{A}_i$ for item n and variance, respectively. By LRR, we get $s_n^{\mathcal{A}}$ valued in the range of $[1, D]$, where D is the highest rating score in real recommender system. With $a_n^{\mathcal{A}}$ and $s_n^{\mathcal{A}}$, we can calculate $q_n^{\mathcal{A}}$.

We assemble UAPs and IAQs into two matrices UAP Matrix (UAPM) and IAQ Matrix (IAQM), which are formally defined as follows:

Definition 4 (User Aspect Preference Matrix (UAPM)): The UAPM, denoted by $\boldsymbol{P}$, is a matrix of $\mathbb{R}^{M\times L}$, where M is the number of users and L is the number of aspects. The cell at ith row and jth column of $\boldsymbol{P}$ is denoted by $\boldsymbol{P}_{i,j}$, $\boldsymbol{P}_{i,j} = p_i^{\mathcal{A}_j}$, $1 \le i \le M$, $1 \le j \le L$.

Definition 5 (Item Aspect Quality Matrix (IAQM)): The IAQM, denoted by $\boldsymbol{Q}$, is a matrix of $\mathbb{R}^{N\times L}$, where N is the number of items and L is the number of aspects. The cell at ith row and jth column of $\boldsymbol{Q}$ is denoted by $\boldsymbol{Q}_{i,j}$, $\boldsymbol{Q}_{i,j} = q_i^{\mathcal{A}_j}$, $1 \le i \le N$, $1 \le j \le L$.

3 Aspect-based matrix factorization

In this section, we present the details of Aspect-based Matrix Factorization model (AMF) by which a rating prediction can be made. As we have mentioned, our straightforward idea is to fuse auxiliary information extracted from aspects into the factorization of the original rating matrix. In other words, AMF collaboratively decomposes original rating matrix with UAPM and IAQM. Following the general idea of matrix factorization [17], AMF assumes that the original rating matrix can be reconstructed by two low rank latent factor matrices: $\boldsymbol{R} \approx \hat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}}$, where $\boldsymbol{R} \in \mathbb{R}^{M\times N}$ is the original rating matrix and $\hat{\boldsymbol{R}} \in \mathbb{R}^{M\times N}$ is the reconstructed rating matrix. $\boldsymbol{U} \in \mathbb{R}^{M\times K}$ is the user latent factor matrix, and a row vector of $\boldsymbol{U}$ implicitly represents the preferences of one user, where K is the number of latent factors. $\boldsymbol{V} \in \mathbb{R}^{N\times K}$ is the item latent factor matrix, and a row vector of $\boldsymbol{V}$ represents the inherent properties of one item. Naively, AMF generates the matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ by minimizing the following optimization objective:

$$J(\boldsymbol{U},\boldsymbol{V}) = \frac{1}{2}\|\boldsymbol{R} - \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}}\|_{\mathrm{F}}^2 + \frac{\lambda}{2}(\|\boldsymbol{U}\|_{\mathrm{F}}^2 + \|\boldsymbol{V}\|_{\mathrm{F}}^2). \tag{7}$$

In Equation (7), there is only the observed rating information of items (represented by $\boldsymbol{R}$) used for approximating $\boldsymbol{U}$ and $\boldsymbol{V}$. In order to approximate $\boldsymbol{U}$ and $\boldsymbol{V}$ more accurately, we fuse UAPM (represented by $\boldsymbol{P}$) and IAQM (represented by $\boldsymbol{Q}$) as constraints into the collaborative decomposition of $\boldsymbol{R}$. $\boldsymbol{P}$ can be factorized as $\boldsymbol{P} = \boldsymbol{U}\boldsymbol{X}^{\boldsymbol{T}}$, where $\boldsymbol{X} \in \mathbb{R}^{L\times K}$ is a latent factor matrix. Note that $\boldsymbol{P}$ shares $\boldsymbol{U}$ with $\boldsymbol{R}$. $\boldsymbol{Q}$ can be factorized as $\boldsymbol{Q} = \boldsymbol{V}\boldsymbol{Y}^{\boldsymbol{T}}$, where $\boldsymbol{Y} \in \mathbb{R}^{L\times K}$ is another latent factor matrix. Similarly, $\boldsymbol{Q}$ shares $\boldsymbol{V}$ with $\boldsymbol{R}$. The idea here is that the rating information of item, which is represented by matrix $\boldsymbol{R}$, through $\boldsymbol{U}$ and $\boldsymbol{V}$, can fuse with the user preference information on aspects and item sentiment information on aspects, which are represented by $\boldsymbol{P}$ and $\boldsymbol{Q}$. By fusing auxiliary information extracted from aspects, AMF is able to reconstruct the original rating matrix $\boldsymbol{R}$ more accurately so as to improve the accuracy of rating prediction.

Algorithm 1 *Aspect-based Matrix Factorization Algorithm*

```
Input:
Matrices $\boldsymbol{R}$, $\boldsymbol{P}$, $\boldsymbol{Q}$, parameters $\gamma_1, \gamma_2, \lambda_1$,
the number of latent factors $K$, the number of aspects $L$.
Output:
Reconstructed rating matrix $\hat{\boldsymbol{R}}$.
Set $\theta$ as step size, $i$ as iteration number,
$I$ as maximum iteration number, $e$ as error threshold;
Initialize $i = 0$;
Initialize matrix $\boldsymbol{U}$, $\boldsymbol{V}$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ with the uniform distribution in the range of $[-0.1, 0.1]$;
Calculate the initial value of the objective function $J^0$ according to (8);
Initialize $\Delta J^i = J^0$;
while $(\Delta J^i > e)$ and $(i < I)$ do
    for each $\boldsymbol{R}_{m,n} \neq 0$ do
        $\boldsymbol{U} = \boldsymbol{U} - \theta\partial_{\boldsymbol{U}} J$;

        $\boldsymbol{V} = \boldsymbol{V} - \theta\partial_{\boldsymbol{V}} J$;

        $\boldsymbol{X} = \boldsymbol{X} - \theta\partial_{\boldsymbol{X}} J$;

        $\boldsymbol{Y} = \boldsymbol{Y} - \theta\partial_{\boldsymbol{Y}} J$;

    end for
    $i = i + 1$;
    Calculate the $i$-th iteration value of objective function $J^i$ according to (8);
    $\Delta J^i = J^i - J^{i-1}$;
end while

$\hat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}}$
```

By fusing the aspect information, we can redefine the optimization objective of AMF as to minimize the following function:

$$
\begin{aligned}
J(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{X}, \boldsymbol{Y}) = {} & \tfrac{1}{2}\|\boldsymbol{R} - \boldsymbol{U}\boldsymbol{V}^{\boldsymbol{T}}\|_{\mathrm{F}}^2 + \tfrac{\gamma_1}{2}\|\boldsymbol{P} - \boldsymbol{U}\boldsymbol{X}^{\boldsymbol{T}}\|_{\mathrm{F}}^2 + \tfrac{\gamma_2}{2}\|\boldsymbol{Q} - \boldsymbol{V}\boldsymbol{Y}^{\boldsymbol{T}}\|_{\mathrm{F}}^2 \\
& + \tfrac{\lambda_1}{2}(\|\boldsymbol{U}\|_{\mathrm{F}}^2 + \|\boldsymbol{V}\|_{\mathrm{F}}^2 + \|\boldsymbol{X}\|_{\mathrm{F}}^2 + \|\boldsymbol{Y}\|_{\mathrm{F}}^2),
\end{aligned}
\tag{8}
$$

where $\|\cdot\|_{\mathrm{F}}$ represents Frobenius norm. The first term of the right side of Equation (8) controls the factorization error of $\boldsymbol{R}$; the second term regulates the error of factorization of $\boldsymbol{P}$; the third term controls the error of factorization of $\boldsymbol{Q}$; and the last term is the regularizing term used to avoid overfitting. γ_1, γ_2 and λ_1 are parameters controlling the contributions of different parts.

Algorithm 1 gives the procedures of AMF, where a local minima of J is obtained through an iterative process of gradient descent, and the gradients of objective function are given by the following equations:

$$
\begin{aligned}
\partial_{\boldsymbol{U}} J &= (\boldsymbol{U}\boldsymbol{V}^T - \boldsymbol{R})\boldsymbol{V} + \gamma_1(\boldsymbol{U}\boldsymbol{X}^T - \boldsymbol{P})\boldsymbol{X} + \lambda_1\boldsymbol{U}, \\
\partial_{\boldsymbol{V}} J &= (\boldsymbol{U}\boldsymbol{V}^T - \boldsymbol{R})^T\boldsymbol{U} + \gamma_2(\boldsymbol{V}\boldsymbol{Y}^T - \boldsymbol{Q})\boldsymbol{Y} + \lambda_1\boldsymbol{V}, \\
\partial_{\boldsymbol{X}} J &= \gamma_1(\boldsymbol{U}\boldsymbol{X}^T - \boldsymbol{P})^T\boldsymbol{U} + \lambda_1\boldsymbol{X}, \\
\partial_{\boldsymbol{Y}} J &= \gamma_2(\boldsymbol{V}\boldsymbol{Y}^T - \boldsymbol{Q})^T\boldsymbol{V} + \lambda_1\boldsymbol{Y}.
\end{aligned}
$$

4 Experiment

In this section, we present the details of the experiments conducted on real datasets. We first determine the number L of aspects and the number K of latent factors, and then verify the performance of AMF. At last, we perform several case studies to show the explainability of AMF and the ability of AMF to capture differences of users and items at aspect level. The experiments are executed on a Windows 7 PC with an Intel Core CPU of 3.3 GHz and 16 GB RAM, and all algorithms are implemented in Python.

4.1 Settings

4.1.1 Dataset

Our experiments are conducted on two real datasets. The first dataset is TripAdvisor which contains over 178K hotel reviews [39]. The second dataset is an Amazon review dataset containing over 231K video game reviews [9, 23]. The statistics of the two datasets are presented in Table 2. The density is computed by formula: $\frac{\#Reviews}{\#Users \times \#Items}$, where #Reviews, #Users and #Items are the review number, the user number and the item number, respectively. From Table 2, one can observe that data is extremely sparse.

Table 2 The statistics of datasets

Dataset	#Users	#Items	#Reviews	Density
TripAdvisor	145318	1759	178616	0.07%
Amazon	24303	10672	231780	0.09%

4.1.2 Baselines

In order to demonstrate the effectiveness of AMF, we compare it with five baseline methods: (1) Probabilistic Matrix Factorization (PMF) [25] PMF is a matrix factorization that outperforms traditional memory-based algorithms on the large sparse rating matrix. (2) Nonnegative Matrix Factorization (NMF) [18] NMF approximates an original rating matrix by factorizing it as a product of two nonnegative matrices, which makes rating prediction without utilizing review data. (3) LDAMF [19] LDAMF leverages the information in reviews with LDA model and regards topic distribution of each item as latent factors of corresponding item in a matrix factorization. (4) Hidden Factors as Topics (HFT) [21] HFT jointly models reviews and ratings, which combines the topic distribution over reviews and latent factor vectors obtained by a matrix factorization by using an exponential transformation function. (5) Explicit Factor Model (EFM) [43] EFM integrates explicit and implicit features into a factor model based on phrase-level sentiment analysis.

4.1.3 Model evaluation

We use Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) to evaluate the prediction accuracy of AMF, which are defined as follows:

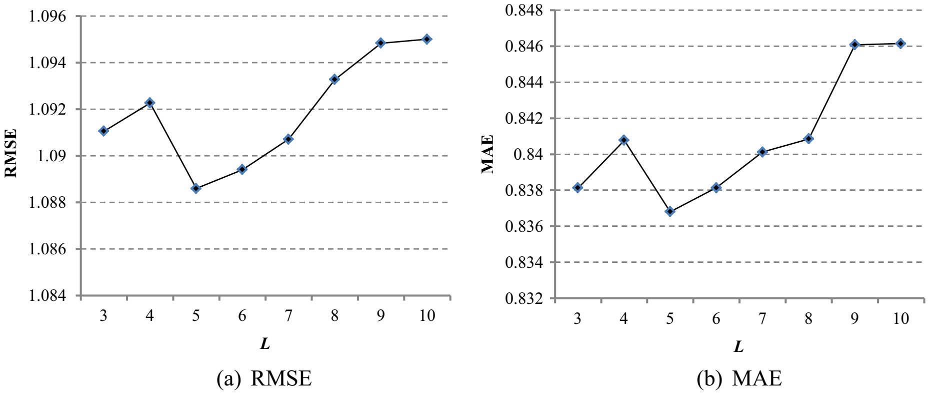
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}, \tag{9}$$

$$MAE = \frac{\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right|}{n}, \tag{10}$$

where y_i and $\hat{y}_i$ are the true and estimated rating scores, respectively, and n is the number of instances in the test set.

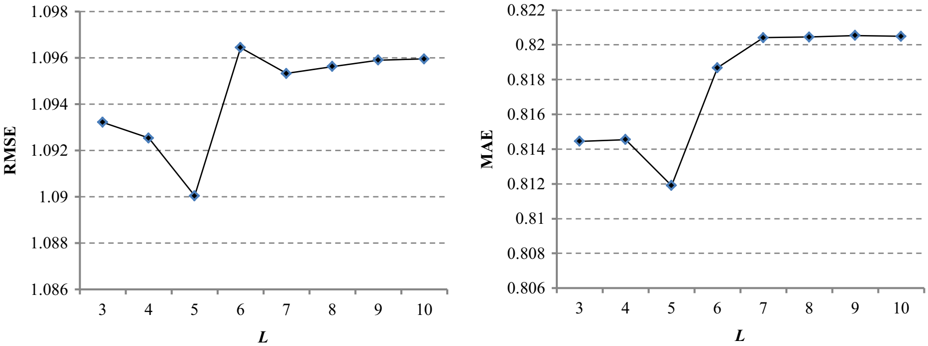
4.2 Sensitiveness of parameter

We first determine the hyper-parameters of (8) by a grid search in the range of (0, 1] with a step size of 0.05. As a result, the hyper-parameters γ_1, γ_2 and λ_1 are set to 0.6, 0.1, 0.05, respectively. Then we investigate the sensitiveness of the number L of aspects and the number K of latent factors.



(a) RMSE (b) MAE

Figure 3 Tuning the number L of aspects on TripAdvisor



(a) RMSE (b) MAE

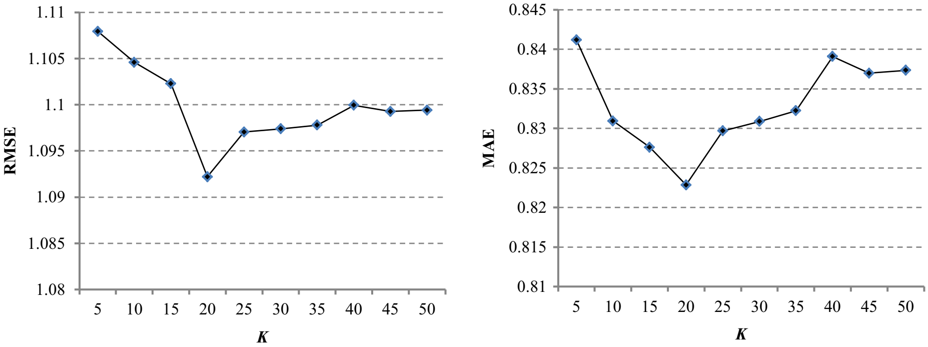
Figure 4 Tuning the number L of aspects on Amazon

Figures 3 and 4 show the RMSEs and MAEs of AMF over L = 3, 4, 5, 6, 7, 8, 9, 10. We can see that the RMSE and MAE over the both two real datasets achieve the minimum when $L = 5$, so we choose the value 5 as the aspect number for the two datasets in the following experiments.

We tune K from 5 to 50, and Figures 5 and 6 show the effects of different values of K. The RMSE and MAE of AMF, on TripAdvisor dataset and Amazon dataset, achieve the minimum at 20 and 35, respectively. Thus we choose the values 20 and 35 as the number of latent factors on TripAdvisor dataset and Amazon dataset in the following experiments, respectively.

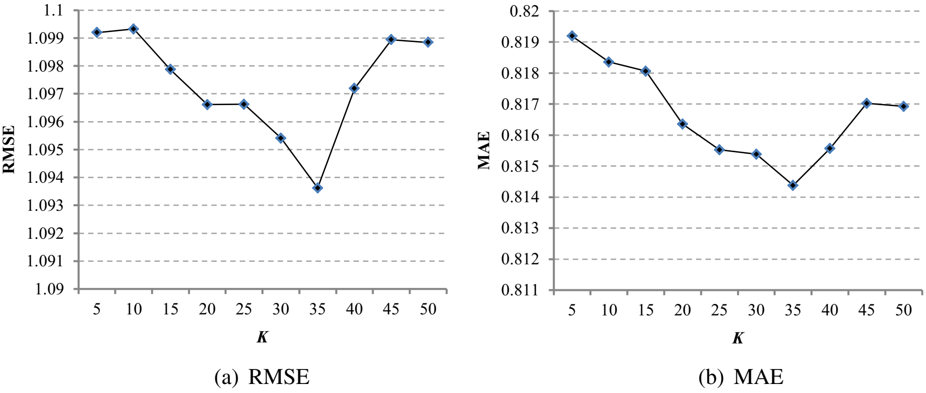
4.3 Prediction accuracy

We split each dataset into training set and testing set with a ratio of 8:2. We repeat experiments ten times and use the average as the final prediction result. AMF outperforms baseline methods significantly at the 0.01 level in terms of paired t-test. The prediction results on the two datasets are shown in Figures 7 and 8, from which we can see that the RMSE and MAE



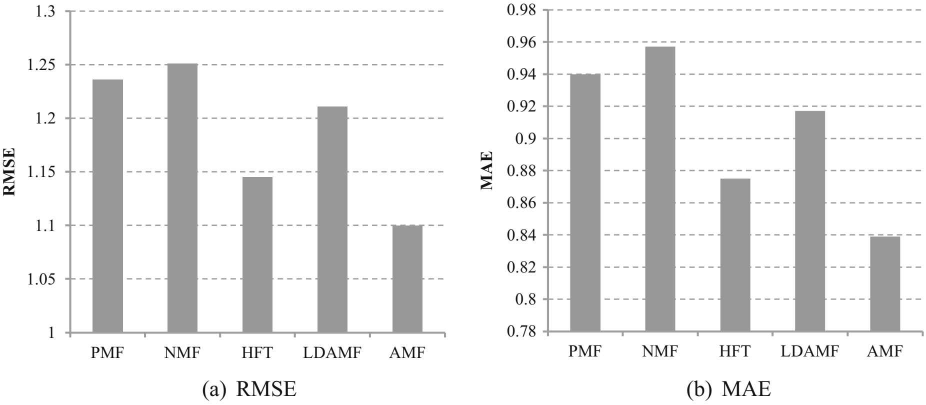
(a) RMSE (b) MAE

Figure 5 Tuning the number K of latent factors on TripAdvisor



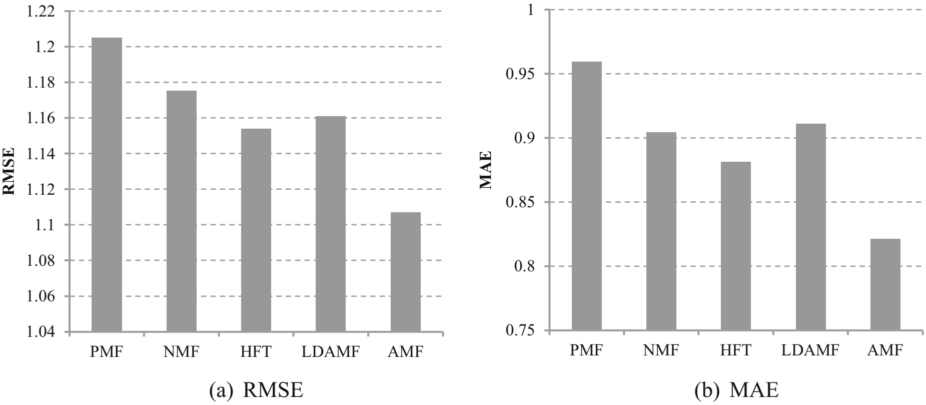
(a) RMSE (b) MAE

Figure 6 Tuning the number K of latent factors on Amazon



(a) RMSE (b) MAE

Figure 7 Performance comparison on TripAdvisor



(a) RMSE (b) MAE

Figure 8 Performance comparison on Amazon

Table 3 RMSE and MAE comparisons of different percentages of test set on Amazon and TripAdvisor datasets

Metric	Method	Amazon					TripAdvisor				
		10%	20%	30%	40%	50%	10%	20%	30%	40%	50%
RMSE	NMF	1.156	1.175	1.186	1.208	1.264	1.236	1.246	1.249	1.251	1.253
	PMF	1.192	1.205	1.206	1.217	1.243	1.227	1.232	1.235	1.236	1.238
	LDAMF	1.158	1.161	1.168	1.169	1.172	1.206	1.211	1.213	1.214	1.219
	HFT	1.138	1.154	1.193	1.223	1.230	1.144	1.145	1.147	1.185	1.187
	EFM	1.119	1.129	1.152	1.183	1.224	1.108	1.113	1.121	1.148	1.172
	AMF	1.102	1.107	1.112	1.121	1.148	1.096	1.101	1.104	1.114	1.119
MAE	NMF	0.886	0.904	0.914	0.933	0.983	0.942	0.952	0.955	0.957	0.960
	PMF	0.952	0.959	0.961	0.965	0.975	0.931	0.938	0.939	0.940	0.941
	LDAMF	0.910	0.911	0.913	0.918	0.919	0.913	0.917	0.918	0.919	0.924
	HFT	0.869	0.881	0.892	0.921	0.947	0.874	0.875	0.876	0.912	0.916
	EFM	0.829	0.835	0.847	0.863	0.891	0.855	0.857	0.861	0.887	0.904
	AMF	0.817	0.821	0.823	0.826	0.841	0.838	0.839	0.840	0.843	0.846

of AMF on both datasets are significantly lower than the RMSE and MAE of the baseline methods. The reasons can be analyzed as follows:

1. In contrast with PMF and NMF, AMF makes rating prediction by fusing auxiliary information extracted from reviews, which is significant for improving the accuracy of rating prediction.
2. Although LDAMF and HFT also make rating prediction based on the models incorporating auxiliary information extracted from reviews, unlike AMF, they do not take into consideration sentiment information which is yet helpful to supplement user profile.
3. Compared to AMF, EFM may be suffering from the lack of explicit features of users or items, which leads to information missing for the construction of profile. AMF models aspects rather than features, which can alleviate this problem because each aspect in AMF contains a certain amount of related words, i.e., explicit features.

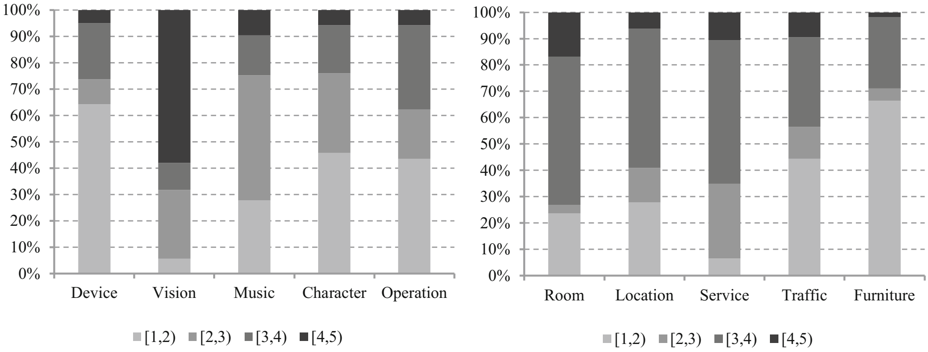
We also conduct experiments with different percentages of test set for each compared method. As shown in Table 3, AMF outperforms all baseline methods.

4.4 Case studies

We first verify the explainability of AMF, and then we show the ability of AMF to capture differences of users and items at aspect level.

4.4.1 Explainability

In this section, we first propose a new metric Satisfaction Degree on Aspects (SDA) to measure user satisfaction on aspects. Then we give two case studies to show how AMF explains the recommendation results by SDA. Finally, we make statistics to investigate the explainability on the two datasets.



(a) Amazon

(b) TripAdvisor

Figure 9 The distribution of $p_m^{\mathcal{A}_i}$ over all aspects on Amazon and TripAdvisor datasets

To evaluate the overall satisfaction degree of one user on aspects toward the specified item, we propose a new metric, Satisfaction Degree on Aspects (SDA), which is defined as

$$SDA = \frac{\sum_{i=1}^{L} w_m^{\mathcal{A}_i} \times q_n^{\mathcal{A}_i}}{\sum_{i=1}^{L} w_m^{\mathcal{A}_i} \times M_{q^{\mathcal{A}_i}}}, \tag{11}$$

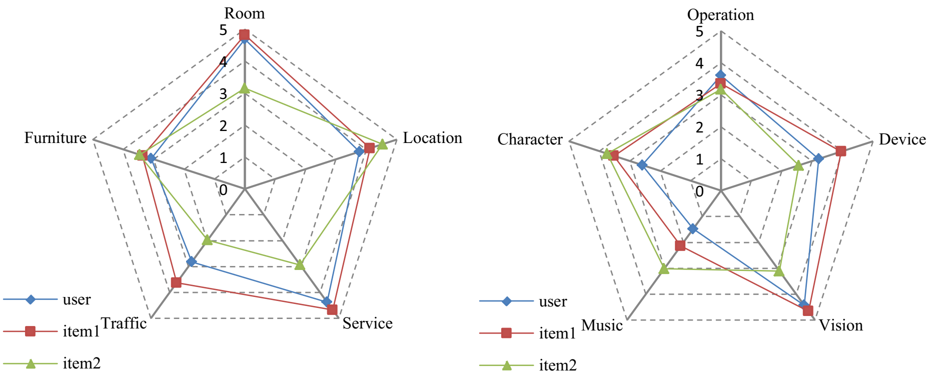
where $w_m^{\mathcal{A}_i} = \frac{p_m^{\mathcal{A}_i} - N_{pm}}{M_{pm} - N_{pm}}$ if $M_{p_m} \neq N_{p_m}$ $(1 \leq i \leq L, w_m^{\mathcal{A}_i} \in [0, 1])$, otherwise, $w_m^{\mathcal{A}_i} = 1$, and M_{p_m} and N_{p_m} are the maximum value of $p_m^{\mathcal{A}_i}$ and the minimum value of $p_m^{\mathcal{A}_i}$, respectively. Figure 9 shows the distribution of $p_m^{\mathcal{A}_i}$ over all aspects on the two datasets where the horizontal axis represents the aspects, and the vertical axis represents the ratio of the data points that fall into each interval of $p_m^{\mathcal{A}_i}$. From Figure 9, we can see that the ratio of each interval is nonzero, which indicates that SDA is a reliable metric for the user satisfaction as there are no outlier data points that can dominate the value of $p_m^{\mathcal{A}_i}$. In SDA, $w_m^{\mathcal{A}_i}$ measures the relative importance of aspect $\mathcal{A}_i$ for user m. $M_{q^{\mathcal{A}_i}}$ is the upper bound of $q_n^{\mathcal{A}_i}$. Here the idea of SDA is to measure the gap between the aspect sentiment of the recommended item and the aspect sentiment of user expected item.

We follow the idea that the explainability of recommendation is generally evaluated by examples [10, 38]. We give two case studies described as follows:

1. The first case study is conducted on TripAdvisor dataset. Table 4 shows the UAPs of a sampled user (user name: Jerri_Blank) and the IAQs of two items reviewed by

Table 4 The UAPs of the user and the IAQs of the two items on TripAdvisor dataset

Object	Room	Location	Service	Traffic	Furniture
User: Jerri_Blank	4.69	3.77	4.38	2.82	3.08
Item1: Hotel_208453	4.82	4.12	4.68	3.63	3.36
Item2: Hotel_230114	3.15	4.53	2.94	1.98	3.46



(a) The visualization of Table 4

(b) The visualization of Table 5

Figure 10 The visualization of the data in Tables 4 and 5

the user (item1 name: Hotel_208453 and item2 name: Hotel_230114) on TripAdvisor dataset. In Table 4, "Room", "Location", "Service", "Traffic", "Furniture" are the five aspects extracted from TripAdvisor dataset. To quantitatively analyze the user satisfaction degree toward the two items, according to (11), we compute SDA toward the two items: $SDA_{item1} = 0.913$, and $SDA_{item2} = 0.676$. In result, $SDA_{item1} > SDA_{item2}$, which shows that the user is more satisfied with the aspects of item1. In addition, Figure 10a gives the visualization of the data in Table 4. Intuitively, from Figure 10a, it can be seen that the user cares most about "Room" aspect and "Service" aspect, where item1 is much stronger than item2. Although item2 is a little bit stronger on "Location" aspect than item1, the user cares a bit less about this aspect. Thus, the user is more likely satisfied with item1 because it has superiority of the IAQs, which is coincident with the comparison result of SDA. Actually, compared to item2, the predicted overall rating of the user toward item1 is higher, which is consistent with the ground truth.

2. The second case study is conducted on Amazon dataset. The sampled data is exhibited in Table 5. According to (11), we compute SDA of the user toward the two items in Table 5: $SDA_{item1} = 0.799$, and $SDA_{item2} = 0.619$. In result, $SDA_{item1} > SDA_{item2}$, which reveals that the user is more satisfied with the aspects of item1. Additionally, Figure 10b gives the visualization of the data in Table 5. From Figure 10b, it can be seen that "Vision" is the most important aspect to the user, where item1 is much stronger than item2. Although item2 is much stronger on "Music" aspect than item1, this is the aspect that the user cares the least. Thus, the user is more likely satisfied with item1 because it has superiority of the IAQs, which is coincident with the comparison result of SDA. In comparison to item2, the predicted overall rating of the user toward item1 is higher, which is consistent with the ground truth.

Table 5 The UAPs of the user and the IAQs of the two items on Amazon dataset

Object	Device	Vision	Music	Character	Operation
User: R. Garrelts	3.22	4.43	1.48	2.59	3.61
Item1: B000038IFX	3.96	4.66	2.14	3.52	3.35
Item2: B000BKHYM6	2.56	3.11	3.02	3.74	3.17

The two case studies demonstrate that AMF can help users to contrastively analyze the fine-grained differences between different recommended items.

Finally, we make statistics to investigate the consistence of predicted overall rating and SDA, namely, whether or not one item of higher predicted overall rating has higher SDA. Without loss of generalization, we assume that one user assigns item A higher overall rating than item B. If $SDA_{itemA} > SDA_{itemB}$, it is Positive Satisfaction; otherwise, Negative Satisfaction. We take two steps to fulfill the test. First, for each user in test sets of the two datasets, we randomly select two items. One of them is assigned an overall rating less than 4 and the other is assigned an overall rating equal to or greater than 4. Then we calculate the ratio of Positive Satisfaction for the two datasets by formula: $\frac{\#Positive\ Satisfaction}{\#user}$, where $\#Positive\ Satisfaction$ and $\#user$ are the number of Positive Satisfaction and the number of users, respectively. We repeat the test five times and use the average as the final statistical result. The ratios of Positive Satisfaction on Amazon dataset and TripAdvisor dataset are 84.31% and 79.12%, respectively. It indicates that most items of higher overall rating have higher satisfaction degree on aspects. In other words, these items suit the taste of users. At the same time, a few items of lower overall rating might also have higher satisfaction degree on aspects, which is because that some users may be satisfied with part of aspects of one item even if they assign a lower overall rating to that item.

4.4.2 Capturing user preference to aspects

As we have mentioned before, even if two users assign the same rating score to the same item, their preferences to aspects may be different. Figure 11 illustrates the situation by the comparison of the UAPs of two sampled users (user1 name: “blackaciddevil” and user2 name: “Michael Kerner”) on Amazon dataset. “Device”, “Vision”, “Music”, “Character”, “Operation” are the five aspects extracted from Amazon dataset. For the sampled item (item name: B002I094AC), the two users all assign the same rating score 4. Although their rating scores are the same for this item, from Figure 11, we can observe that their preferences to aspects are different. User1 shows great interests in “Vision”, “Operation” and

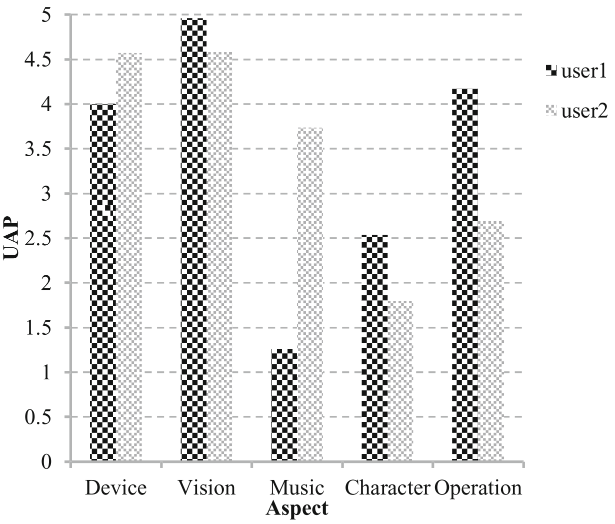
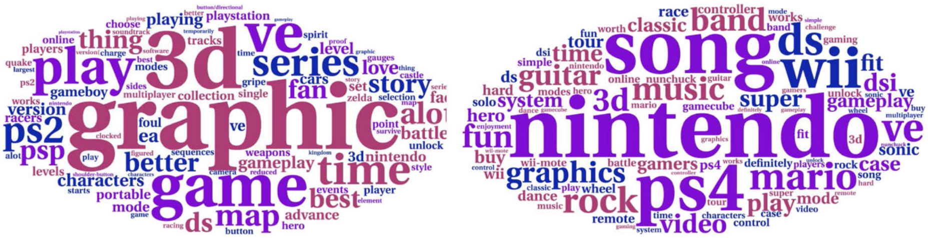


Figure 11 The comparisons of the UAPs of the two users



(a) Profile of user1

(b) Profile of user2

Figure 12 The preference profiles of the two users

"Device" aspects. User2 prefers to "Device", "Vision" and "Music" aspects. On "Music" and "Operation" aspects, their preferences are obviously different.

Figure 12 shows the preference profiles of the two users on aspects, which illustrates the related words frequently mentioned by the two users. The more frequently a related word is mentioned, the bigger its corresponding font size is. Figure 12 tells us that user1 focuses on "graphic", "3d", "game" and "time", while user2 cares more about "nintendo", "song", "ps4" and "wii", which indicates that the vital preferences (represented by the related words) are different even if the users assign the same rating score to an item.

4.4.3 Capturing sentiment difference on item aspects

Even if two items have the same overall rating score, their review sentiment on aspects may be different. Table 6 shows two sampled items (item1 name: Hotel_86984 and item2 name: Hotel_148789) that have the same overall rating score 4.1 and their IAQs on TripAdvisor dataset. Figure 13 shows the result of the comparison of their IAQs. We can observe that the review sentiment of item1 on "Room" and "Furniture" aspects is better than that of item2. The review sentiment of item2 on "Traffic" aspect outperforms that of item1.

In conclusion, for two users who assign the same rating score to the same item, they possibly have different preferences to aspects. Likewise, for two items that have the same rating score, their review sentiment on each aspect may be different. AMF can effectively capture the aspect-level differences, which is significant to make the recommendations more convincing.

5 Related work

In this section, we briefly review the related work with our research, including collaborative filtering, sentiment-based recommendation, and topic-based recommendation.

Table 6 The IAQs of the two items

Item	Room	Location	Service	Traffic	Furniture
Item1	4.38	3.57	4.66	1.36	3.72
Item2	3.41	4.09	4.77	3.69	2.59

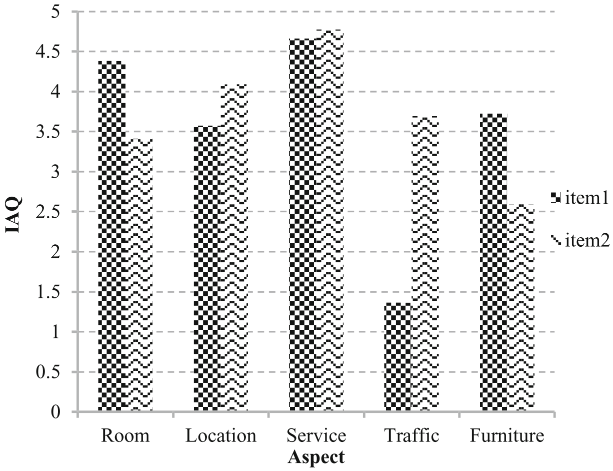


Figure 13 The comparisons of the IAQs of the two items

Collaborative filtering Collaborative Filtering (CF) based approaches [4, 15, 34, 38] make rating prediction using a decomposition of a user-item rating matrix, and fall into two categories, the memory based CF [6, 11, 12] and the model based CF [16, 17, 31]. The memory-based CF approaches generally recommend items based on user similarity or item similarity evaluated based on transaction records. Matrix factorization (MF) is one of the most popular model-based CF methods in recent years which achieves the state-of-the-art performance on multiple practical applications. There are various MF algorithms, including Singular Value Decomposition (SVD) [17, 33], Non-negative Matrix Factorization (NMF) [18], Probabilistic Matrix Factorization (PMF) [25], and Max-Margin Matrix Factorization (MMMF) [31]. However, these approaches only rely on the rating information of items for recommendation, which limits their recommendation performance.

Topic-based explainable recommendation Topic-based approaches utilize topic models, e.g. LDA [24, 45], to uncover topics from reviews for recommendation. McAuley et al. come up with a Hidden Factors as Topics model that jointly models reviews and ratings, which combines the topic distribution over reviews and latent factor vectors obtained by a matrix factorization [21]. Ling et al. propose a unified model that combines topic models with CF to improve prediction accuracy [19]. Bao et al. propose a novel matrix factorization model TopicMF which simultaneously considers the ratings and reviews [3]. Almahairi et al. propose two models to learn distributed representations from reviews for collaborative filtering [1]. Wang et al. leverage topic models to discover explainable latent factors in matrix factorization for explainable recommendation [38]. Musat et al. propose a topic profile collaborative filtering to predict ratings at topic level [27]. Chen et al. propose a context-aware collaborative topic regression method with social matrix factorization for recommendation [4]. However, these methods cannot quantify user preferences and the sentiment of item reviews at topic level, and most of them ignore sentiment information which is able to help to explain recommendation results.

Sentiment-based explainable recommendation Sentiment-based methods utilize the sentiment analysis [2, 13, 20, 26] for recommendation. They explore user opinion [32, 35, 39, 40] from reviews by means of a sentiment lexicon [8, 36, 37] or corpus [14, 28, 44] and integrate these opinion information into recommendations. Pappas et al. propose a sentiment-aware nearest neighbor model for multimedia recommendations [29]. Pero et al. propose a rating prediction framework utilizing both ratings provided by users and opinions inferred from their reviews [30]. Zhang et al. propose an Explicit Factor Models to integrate explicit sentiment information extracted from reviews into matrix factorization for recommendations [43]. Diao et al. propose a probabilistic model to jointly model ratings and sentiments for recommendation based on collaborative filtering and topic modeling [7]. He et al. devise a generic algorithm TriRank for recommendation ranking on tripartite graphs of user-item-aspect [10]. Chen et al. propose a tensor decomposition algorithm to learn to rank user preferences based on phrase-level sentiment analysis across multiple categories [5]. However, these methods cannot serve our goal because they cannot contrastively analyze fine-grained sentiment differences of user reviews on items.

6 Conclusions

In this paper, we propose an Aspect-based Matrix Factorization model (AMF) for explainable recommendation. AMF makes recommendation by fusing auxiliary aspect information extracted from reviews into matrix factorization. To quantitatively evaluate aspects, we propose two metrics: User Aspect Preference (UAP) and Item Aspect Quality (IAQ). UAP reveals user preference to a specific aspect and IAQ assesses the review sentiment of item on an aspect, by which we can make the recommendations explainable and convincing. To improve the accuracy of rating prediction, we assemble UAPs and IAQs into two matrices UAP Matrix (UAPM) and IAQ Matrix (IAQM), respectively, which are used as constraints fused into collaborative decomposition of the item rating matrix. The results of the extensive experiments conducted on real datasets verify the recommendation performance and explanatory ability of AMF.

Acknowledgments This work is supported by NSFC through grant 61173099, Ministry of Education of China through 6141A02033304, and in part by NSF through grants IIS-1526499 and CNS-1626432.

References

1. Almahairi, A., Kastner, K., Cho, K., Courville, A.: Learning distributed representations from reviews for collaborative filtering. In: Proceedings of the 9th ACM Conference on Recommender Systems, pp. 147–154. ACM (2015)
2. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC, vol. 10, pp. 2200–2204 (2010)
3. Bao, Y., Fang, H., Zhang, J.: Topicmf: simultaneously exploiting ratings and reviews for recommendation. In: AAAI, vol. 14, pp. 2–8 (2014)
4. Chen, C., Zheng, X., Wang, Y., Hong, F., Lin, Z., et al.: Context-aware collaborative topic regression with social matrix factorization for recommender systems. In: AAAI, vol. 14, pp. 9–15 (2014)
5. Chen, X., Qin, Z., Zhang, Y., Xu, T.: Learning to rank features for recommendation over multiple categories. In: Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 305–314. ACM (2016)
6. Deshpande, M., Karypis, G.: Item-based top-n recommendation algorithms. ACM Transactions on Information Systems (TOIS) **22**(1), 143–177 (2004)
7. Diao, Q., Qiu, M., Wu, C.Y., Smola, A.J., Jiang, J., Wang, C.: Jointly modeling aspects, ratings and sentiments for movie recommendation (jmars). In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 193–202. ACM (2014)

8. Ding, X., Liu, B., Yu, P.S.: A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 International Conference on Web Search and Data Mining, pp. 231–240. ACM (2008)
9. He, R., McAuley, J.: Ups and Downs: modeling the visual evolution of fashion trends with one-class collaborative filtering. In: Proceedings of the 25th International Conference on World Wide Web, pp. 507–517. International World Wide Web Conferences Steering Committee (2016)
10. He, X., Chen, T., Kan, M.Y., Chen, X.: Trirank: review-aware explainable recommendation by modeling aspects. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, pp. 1661–1670. ACM (2015)
11. Huang, S., Wang, S., Liu, T.Y., Ma, J., Chen, Z., Veijalainen, J.: Listwise collaborative filtering. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 343–352. ACM (2015)
12. Jin, R., Chai, J.Y., Si, L.: An automatic weighting scheme for collaborative filtering. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 337–344. ACM (2004)
13. Jo, Y., Oh, A.H.: Aspect and sentiment unification model for online review analysis. In: Proceedings of the 4th ACM International Conference on Web Search and Data Mining, pp. 815–824. ACM (2011)
14. Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A sentence model based on convolutional neural networks. In: Procedding of the 52th Annual Meeting of Association for Computational Linguistics (2014)
15. Konstas, I., Stathopoulos, V., Jose, J.M.: On social networks and collaborative recommendation. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 195–202. ACM (2009)
16. Koren, Y.: Factor in the neighbors: scalable and accurate collaborative filtering. ACM Transactions on Knowledge Discovery from Data (TKDD) **4**(1), 1 (2010)
17. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. Computer **42**(8), 30–37 (2009)
18. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems, pp. 556–562 (2001)
19. Ling, G., Lyu, M.R., King, I.: Ratings meet reviews, a combined approach to recommend. In: Proceedings of the 8th ACM Conference on Recommender Systems, pp. 105–112. ACM (2014)
20. Lu, Y., Castellanos, M., Dayal, U., Zhai, C.: Automatic construction of a context-aware sentiment lexicon: an optimization approach. In: Proceedings of the 20Th International Conference on World Wide Web, pp. 347–356. ACM (2011)
21. McAuley, J., Leskovec, J.: Hidden factors and hidden topics: understanding rating dimensions with review text. In: Proceedings of the 7th ACM Conference on Recommender Systems, pp. 165–172. ACM (2013)
22. McAuley, J., Leskovec, J., Jurafsky, D.: Learning attitudes and attributes from multi-aspect reviews. In: 2012 IEEE International Conference on Data Mining (ICDM), pp. 1020–1025. IEEE (2012)
23. McAuley, J., Targett, C., Shi, Q., Van Den Hengel, A.: Image-based recommendations on styles and substitutes. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 43–52. ACM (2015)
24. Mei, Q., Ling, X., Wondra, M., Su, H., Zhai, C.: Topic sentiment mixture: modeling facets and opinions in weblogs. In: Proceedings of the 16th International Conference on World Wide Web, pp. 171–180. ACM (2007)
25. Mnih, A., Salakhutdinov, R.R.: Probabilistic matrix factorization. In: Advances in Neural Information Processing Systems, pp. 1257–1264 (2008)
26. Moghaddam, S., Ester, M.: Ilda: interdependent Lda model for learning latent aspects and their ratings from online product reviews. In: Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 665–674. ACM (2011)
27. Musat, C.C., Liang, Y., Faltings, B.: Recommendation using textual opinions. In: IJCAI International Joint Conference on Artificial Intelligence, EPFL-CONF-197487, pp. 2684–2690 (2013)
28. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing, vol. 10, pp. 79–86. Association for Computational Linguistics (2002)
29. Pappas, N., Popescu-Belis, A.: Sentiment analysis of user comments for one-class collaborative filtering over ted talks. In: Proceedings of the 36Th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 773–776. ACM (2013)
30. Pero, Š., Horváth, T.: Opinion-Driven Matrix Factorization for Rating Prediction. In: International Conference on User Modeling, Adaptation, and Personalization, pp. 1–13. Springer (2013)

31. Rennie, J.D., Srebro, N.: Fast maximum margin matrix factorization for collaborative prediction. In: Proceedings of the 22nd International Conference on Machine Learning, pp. 713–719. ACM (2005)
32. Schouten, K., Frasincar, F.: Survey on aspect-level sentiment analysis. IEEE Trans. Knowl. Data Eng. **28**(3), 813–830 (2016)
33. Srebro, N., Jaakkola, T.: Weighted low-rank approximations. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 720–727 (2003)
34. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. Advan. Artif. Intell. **2009**, 4 (2009)
35. Tan, Y., Zhang, Y., Zhang, M., Liu, Y., Ma, S.: A unified framework for emotional elements extraction based on finite state matching machine. In: Natural Language Processing and Chinese Computing, pp. 60–71. Springer (2013)
36. Thelwall, M., Buckley, K., Paltoglou, G.: Sentiment strength detection for the social Web. J. Am. Soc. Inf. Sci. Technol. **63**(1), 163–173 (2012)
37. Turney, P.D.: Thumbs up Or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp. 417–424. Association for Computational Linguistics (2002)
38. Wang, C., Blei, D.M.: Collaborative topic modeling for recommending scientific articles. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 448–456. ACM (2011)
39. Wang, H., Lu, Y., Zhai, C.: 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, pp. 783–792. ACM (2010)
40. Wu, Y., Ester, M.: Flame: a probabilistic model combining aspect based opinion mining and collaborative filtering. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pp. 199–208. ACM (2015)
41. Yang, L., Qiu, M., Gottipati, S., Zhu, F., Jiang, J., Sun, H., Chen, Z.: Cqarank: jointly model topics and expertise in community question answering. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, pp. 99–108. ACM (2013)
42. Zhang, Y.: Incorporating phrase-level sentiment analysis on textual reviews for personalized recommendation. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pp. 435–440. ACM (2015)
43. Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., Ma, S.: Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In: Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval, pp. 83–92. ACM (2014)
44. Zhao, J., Dong, L., Wu, J., Xu, K.: Moodlens: an emoticon-based sentiment analysis system for chinese tweets. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1528–1531. ACM (2012)
45. Zhao, W.X., Jiang, J., Yan, H., Li, X.: Jointly modeling aspects and opinions with a Maxent-Lda hybrid. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pp. 56–65. Association for Computational Linguistics (2010)
46. Zuo, Y., Wu, J., Zhang, H., Wang, D., Lin, H., Wang, F., Xu, K.: Complementary aspect-based opinion mining across asymmetric collections. In: 2015 IEEE International Conference on Data Mining (ICDM), pp. 669–678. IEEE (2015)