Exploiting Emotion on Reviews for Recommender Systems

Xuying Meng[1,2], Suhang Wang[3], Huan Liu[3] and Yujun Zhang[1]
[1]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Computer Science and Engineering, Arizona State University, Tempe, 85281, USA
{mengxuying, zhmj}@ict.ac.cn, {suhang.wang, huan.liu}@asu.edu

Abstract

Review history is widely used by recommender systems to infer users' preferences and help find the potential interests from the huge volumes of data, whereas it also brings in great concerns on the sparsity and cold-start problems due to its inadequacy. Psychology and sociology research has shown that emotion information is a strong indicator for users' preferences. Meanwhile, with the fast development of online services, users are willing to express their emotion on others' reviews, which makes the emotion information pervasively available. Besides, recent research shows that the number of emotion on reviews is always much larger than the number of reviews. Therefore incorporating emotion on reviews may help to alleviate the data sparsity and cold-start problems for recommender systems. In this paper, we provide a principled and mathematical way to exploit both positive and negative emotion on reviews, and propose a novel framework MIRROR, exploiting eMotIon on Reviews for RecOmmendeR systems from both global and local perspectives. Empirical results on real-world datasets demonstrate the effectiveness of our proposed framework and further experiments are conducted to understand how emotion on reviews works for the proposed framework.

Introduction

Recommender systems have become an imperative part to tackle the information overload problem by suggesting online users with products of potential interests (Wang et al. 2015; Shu et al. 2018). The majority of existing recommender systems mainly rely on users' historical ratings or reviews to conduct recommendation (Koren, Bell, and Volinsky 2009; Beutel et al. 2017). However, these recommender systems usually suffer from data sparsity and cold-start problems because compared to the great amount of items, a user only gives ratings to few items especially at the start. Therefore, online user-generated data, which can help to infer users' preferences, has become a key complementary source for recommender systems to alleviate the data sparsity and cold-start problems (Tang et al. 2016; Wang et al. 2017c). For example, social recommendation, which exploits user-user relationships guided by the homophily theory that two friends are more likely to have similar preferences, has shown to be effective for recommender systems (Wang et al. 2016).

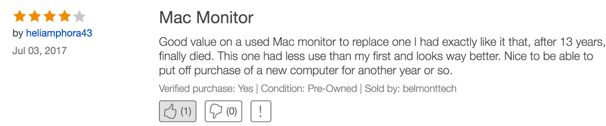


Figure 1: An example of emotion on reviews.

In real word, product review websites also provide various ways for users to express their emotion toward other users' reviews, which leaves abundant emotion information that has potential to improve recommendation performance. For example, Amazon users can indicate their emotion via commenting and replying posts under others' reviews; eBay users can give thumb up or down to express their emotion on reviews; and some commercial platforms like Ciao and Epinions provide rating mechanisms for emotion expression. In other words, emotion information is pervasively available. According to the sociologists and psychologists, users' emotion is a strong indicator of agreement or disagreement (Dunn and Schweitzer 2005; Bewsell 2012). For example, positive emotion such as appreciation and satisfaction indicates agreement, negative emotion such as antipathy and anger indicates disagreement (Wang et al. 2017b; 2017a). Therefore, emotion information can be used to infer a user's preference towards a product though the user has not given a rating to the product. For example, in Figure 1[1], *heliamphora43* submitted a review on Mac Monitor with a rating 4 and a user *Tracy* gave a thumb up to this review. From this, we can infer that *Tracy*'s rating on Mac Monitor will be close to 4 even if *Tracy* hasn't given her own rating. Furthermore, from real world data analysis, we find that the number of emotion on reviews is much larger than the number of reviews, which means that we have more abundant resource, i.e., emotion information, to help infer users' potential preferences. Therefore, exploiting emotion on reviews has great potential to help alleviate data sparsity and cold-start problems and improve the recommendation performance. The majority of existing work mainly focus on

[1]https://www.ebay.com/urw/Apple-Cinema-A1083-30-Widescreen-LCD-Monitor/product-reviews/82049072

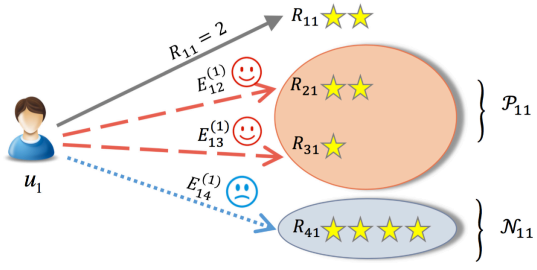


Figure 2: An example of user's positive and negative emotion reviews. The red dashed arrow with a smiling face denotes u_1 expresses positive emotion to the corresponding rating. The blue dotted arrow with a crying face denotes u_1 expresses negative emotion to the corresponding rating.

exploiting users' *emotion on products* (Costa and Macedo 2013; Zheng, Mobasher, and Burke 2016). However, little attention is paid to utilizing users' *emotion on reviews* for recommender systems.

Therefore, in this paper, we study the novel problem of exploiting emotion on reviews for recommender systems. In essence, we are faced with the following challenges: (1) Different from positive emotion, users seldom express their negative emotion (Beigi et al. 2016), although negative emotion is also a useful indicator of disagreement. The unbalanced data brings difficulties to take full advantage of positive and negative emotion; and (2) Users can give emotion on several reviews of the same item, among which these emotion may be all positive, all negative, or partial positive and partial negative. It's difficult to unify these cases in one model. Hence, modeling emotion information in recommendation is not a trivial problem. Thus, in this work, we investigate (1) how to model positive and negative emotion mathematically, and (2) how to incorporate emotion on reviews into a unified model for recommender systems. In an attempt to solve two questions, we propose a novel recommendation framework - MIRROR. The main contributions are summarized as follows:

- We study a new problem of exploiting emotion on reviews for recommender systems and statistically verify that positive and negative emotion can be used to infer users' preferences;
- We propose a novel framework MIRROR, which captures emotion in global and local perspectives to improve recommendation performance; and
- We conduct experiments on real-world datasets to evaluate the effectiveness of the proposed framework.

Problem Statement

We first introduce the notations used in this paper. Throughout the paper, matrices are written as boldface capital letters (e.g., $\mathbf{M}$), vectors are denoted as boldface lowercase letters (e.g., $\mathbf{m}$, $\mathbf{u}_1$, $\mathbf{x}_j^{(1)}$). For any matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$, $\mathbf{M}_{ij}$ is the (i,j)-th entry of $\mathbf{M}$, and the Frobenius norm of $\mathbf{M}$ is $||\mathbf{M}||_F = \sqrt{\sum_i \sum_j \mathbf{M}_{ij}^2}$. Also, we use calligraphic fonts for sets (e.g., $\mathcal{F}$). $|\mathcal{F}|$ represents the cardinality of set $\mathcal{F}$.

Let $\mathcal{U} = \{u_1, u_2, \dots, u_n\}$ be the set of n users, and $\mathcal{V} = \{v_1, v_2, \dots, v_m\}$ be the set of m items, and $\mathcal{R} = \{r_1, r_2, \dots, r_N\}$ be the set of N reviews. We use matrix $\mathbf{R} \in \mathbb{R}^{n\times m}$ to denote user-item rating matrix where $\mathbf{R}_{ij}$ represents the rating from user u_i to item v_j. $\mathbf{R}_{ij} = 0$ if a user u_i hasn't rated v_j. Generally, we have $N \ll n \times m$ as users only rate a small portion of items. In addition to giving ratings to products, users can also express their emotion on other users' reviews. We use $\mathbf{E} \in \mathbb{R}^{n\times N}$ to denote emotion-review relations where $\mathbf{E}_{ik}^{(j)} = 1, -1$ or 0 denotes user u_i has expressed positive, negative or no emotion, respectively, to user u_k's review on item v_j. We define positive emotion reviews $\mathcal{P}_{ij} = \{\mathbf{R}_{kj} | \mathbf{E}_{ik}^{(j)} = 1\}$ and negative emotion reviews $\mathcal{N}_{ij} = \{\mathbf{R}_{kj} | \mathbf{E}_{ik}^{(j)} = -1\}$ be the rating sets on v_j with u_i's positive and negative emotion. For example, in Figure 2, for an item v_1, u_1 have his own rating $\mathbf{R}_{11}$. Besides, towards other users' ratings $\mathbf{R}_{21}$, $\mathbf{R}_{31}$ and $\mathbf{R}_{41}$, u_1 expressed positive emotion on $\mathbf{R}_{21}$ and $\mathbf{R}_{31}$ and negative emotion on $\mathbf{R}_{41}$. In this case, u_1's positive and negative emotion reviews on item v_1 are $\mathcal{P}_{11} = \{\mathbf{R}_{21}, \mathbf{R}_{31}\}$ and $\mathcal{N}_{11} = \{\mathbf{R}_{41}\}$ separately. With these notations and definitions, the problem of recommendation with emotion information can be formally defined as: *Given the rating matrix* $\mathbf{R}$ *and emotion* $\mathbf{E}$ *towards others' reviews, we aim to find a recommender system to infer missing values in* $\mathbf{R}$.

Data Analysis on Emotion

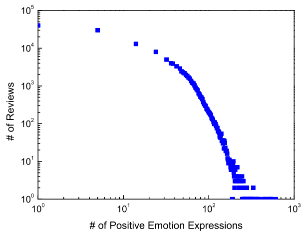
In this section, we will investigate the relations between users' preferences and users' emotion expressions, which lays the foundation for exploiting emotion on reviews for recommender systems. We first introduce the datasets used in this paper.

Datasets Description

We collected two real-world datasets Ciao and Epinions. For Ciao, users can rate *products* and *reviews* with scores from 1 to 5. For the evaluation purpose, we take the ratings on reviews as ground truth of emotion. In particular, we consider low review ratings from 1 to 3 as negative emotion, high review ratings 4 and 5 as positive emotion. For Epinions, the rating scope on reviews is from 1 to 6, where we regard ratings from 1 to 3 as negative emotion and 4 to 6 as positive emotion. To make it clear, all of the aftermentioned ratings are referred to product ratings. The detailed statistics of these datasets are illustrated in Table 1. It is evident from

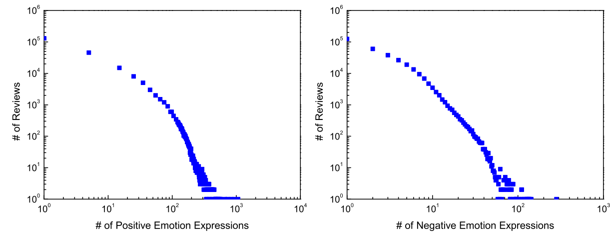
Table 1: Statistics of the datasets

Datasets	Ciao	Epinions
# of Users	44,311	194,872
# of Items	112,783	125,837
# of Ratings	7,345,419	706,200
# of Positive Emotion	7,795,044	6,319,020
# of Negative Emotion	176,933	138,593



(a) Positive emotion (b) Negative emotion

Figure 3: Power-law distribution of the Ciao dataset



(a) Positive emotion (b) Negative emotion

Figure 4: Power-law distribution of the Epinions dataset

these statistics that users are way more willing to express positive emotion than negative emotion.

In order to better understand the distributions of emotion, we further explore the degree distributions of positive emotion reviews and negative emotion reviews in Figure 3 and Figure 4. As can be observed from the distributions, users' emotion towards reviews present a power-law distribution in both datasets, which means that most reviews receive few emotion expressions and a few popular reviews receive extremely high numbers of emotion expressions. This is pretty common in recommendation systems (Beutel et al. 2017).

Analysis of Positive and Negative Emotion

Research findings from psychology and sociology suggest that emotion is a strong indicator of agreement or disagreement (Dunn and Schweitzer 2005; Bewsell 2012), which leads to two assumptions:

- Users are likely to share similar preferences with their positive emotion reviews; and
- Users are more similar to their positive emotion reviews than their negative emotion reviews.

To exploit these two assumptions for recommendation, we would like to first validate whether these assumptions hold in real-world scenario using the two datasets we introduced.

Let $\mathcal{F}_j$ be the set of user pairs from whom we observe ratings to the item v_j as $\mathcal{F}_j = \{(u_i, u_k) | R_{ij} \neq 0 \wedge R_{kj} \neq 0\}$. Among $\mathcal{F}_j$, if user u_i expresses positive and negative emotion on u_k's rating, we include pair (u_i, u_k) into $\mathcal{F}_j^p$ and $\mathcal{F}_j^n$ respectively. Besides that, we randomly select pairs (u_i, u_k) without any emotion interactions from each other, and include them into $\mathcal{F}_j^r$. With definitions of the pair division, we then define the difference between user pairs (u_i, u_k) on item v_j as $d_{ik,j} = (\mathbf{R}_{ij} - \mathbf{R}_{kj})^2$. For pairs in $\mathcal{F}_j^p$, $\mathcal{F}_j^n$ and $\mathcal{F}_j^r$, $j = \{1, 2, ..., m\}$, we construct difference vectors $\mathbf{d}^p$, $\mathbf{d}^n$ and $\mathbf{d}^r$ where $d_{ik,j}^p \in \mathbf{d}^p$, $d_{ik,j}^n \in \mathbf{d}^n$ and $d_{ik,j}^r \in \mathbf{d}^r$ denote rating differences for (u_i, u_k) on v_j with positive, negative and no emotion reviews, respectively.

For two vectors $\{\mathbf{x}, \mathbf{y}\}$, the null hypothesis is $H_0 : \mathbf{x} >= \mathbf{y}$ while the alternative hypothesis is $H_1 : \mathbf{x} < \mathbf{y}$, where the null hypothesis indicates that the mean of $\mathbf{x}$ is bigger than or equal to that of $\mathbf{y}$. To validate the assumptions, we conduct t-test on $\{\mathbf{d}^p, \mathbf{d}^r\}$ and $\{\mathbf{d}^p, \mathbf{d}^n\}$. For the t-test on $\{\mathbf{d}^p, \mathbf{d}^r\}$, the null hypothesis indicates users are less likely to share similar preferences with their positive emotion reviews; therefore if we reject the null hypothesis, then the assumption that users are more likely to share similar preferences with their positive emotion reviews is verified. For datasets Ciao and Epinions, the null hypothesis is rejected at the significant level $\alpha = 0.01$ with p-values of $1.8e^{-12}$ and $1.2e^{-164}$, respectively. For the t-test on $\{\mathbf{d}^p, \mathbf{d}^n\}$, the null hypothesis indicates users are less likely to be similar to their positive emotion reviews than their negative emotion reviews; therefore if we reject the null hypothesis, the assumption that users are more likely to be similar to their positive reviews than their negative reviews is proved. For datasets Ciao and Epinions, the null hypothesis is rejected at the significant level $\alpha = 0.01$ with p-values of $2.2e^{-68}$ and $9.9e^{-200}$, respectively. Thus we verified the relationships between emotion on reviews and user preferences, which pave a way to utilizing emotion on recommendation.

Proposed Framework– MIRROR

In this section, we discuss how to model emotion in global and local perspectives using the two assumptions we validated, in which we weight the importance of each rating based on global emotion information from all users, and predict each individual's preference based on local emotion information from the user himself.

Before modeling users' emotion, we introduce the basic recommendation algorithm we use in this paper. Matrix factorization is a popular model to build recommender systems (Takács et al. 2008; Jamali and Ester 2010; Meng et al. 2018). It decomposes the rating matrix $\mathbf{R}$ as follows:

$$\min_{\mathbf{U},\mathbf{V}} ||\mathbf{W} \odot (\mathbf{R} - \mathbf{U}^T\mathbf{V})||_F^2 + \alpha(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2), \quad (1)$$

where $\odot$ denotes Hadmard product. User latent feature matrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n] \in \mathbb{R}^{K \times n}$ includes the K-dimensional preference latent factors of each user u_i, and item latent feature matrix $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_m] \in \mathbb{R}^{K \times m}$ represents the K-dimensional characteristic latent factors of each item v_j. $\mathbf{W}$ is the indicator matrix that $\mathbf{W}_{ij}$ is equal to 1 if user u_i has rated item v_j and equal to 0 otherwise. The term $\alpha(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2)$ is introduced to avoid overfitting.

Global Emotion Modeling for Recommendation

Existing recommendation methods regard each rating equally weighted, which ignore that ratings are highly idiosyncratic. As shown in Figure 3 and Figure 4, the degree distributions of both positive and negative emotion of real-world datasets subject to power law, showing distinct influence with different numbers of emotion expressions. Ratings received many positive emotion expressions indicate higher importance and demand higher weights, whereas, ratings received plenty of negative emotion expressions may have bad impact on the item latent features and result in recommendation performance degradation. We then model the influence of emotion by the relative number of positive and negative emotion expressions, which gives additional powers and capabilities in recommender systems. We assume that $\mathbf{O}_{ij} = \mathbf{O}_{ij}^{pos} - \mathbf{O}_{ij}^{neg}$ is the difference between the positive emotion number $\mathbf{O}_{ij}^{pos}$ and the negative emotion number $\mathbf{O}_{ij}^{neg}$ over the rating $\mathbf{R}_{ij}$. Then we define emotion influence $\mathbf{W}_{ij}^{e}$ as a function f of relative number of positive and negative emotion expressions $\mathbf{O}_{ij}$. More specifically, we define:

$$\mathbf{W}_{ij}^{e} = \begin{cases} f(\mathbf{O}_{ij}) & \mathbf{O}_{ij} \neq 0 \\ 0 & otherwise \end{cases} \tag{2}$$

where the function f should limit the value of the emotion influence with $[0, 1]$ and be an increasing function of $\mathbf{O}_{ij}$, i.e., relatively more positive emotion expressions should have a higher value of emotion influence. In this work, we empirically find that $f(\mathbf{O}_{ij}) = sgn(\mathbf{O}_{ij})log(|\mathbf{O}_{ij}| + 1)$ works well for our proposed MIRROR, where $sgn(\mathbf{O}_{ij})$ denotes the sign of $\mathbf{O}_{ij}$.

The influence of emotion expressions plays an important role in recommendation. Users are willing to agree to ratings which have attracted many positive emotion expressions and few negative emotion expressions. Bewsell found that emotion greatly affect a user to formalize agreement and transact online (Bewsell 2012). To capture global information from emotion expressions, we can use the emotion influence to weight the importance of each ratings. Originally the importance of $\mathbf{R}_{ij}$ in Eq.(1) is controlled by $\mathbf{W}_{ij}$. Considering the influence of emotion, we should also include $\mathbf{W}_{ij}^{e}$, thus we have the new weight for $\mathbf{R}_{ij}$ as $\widetilde{\mathbf{W}}_{ij} = g(\mathbf{W}_{ij}, \mathbf{W}_{ij}^{e})$ where g is a function to combine two weights. With these new weights, the formulation to globally model users' emotion is as follows:

$$\min_{\mathbf{U},\mathbf{V}} ||\widetilde{\mathbf{W}} \odot (\mathbf{R} - \mathbf{U}^T\mathbf{V})||_F^2 + \alpha(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2), \tag{3}$$

where the importance of $\mathbf{R}_{ij}$ is controlled by $\widetilde{\mathbf{W}}_{ij}$ through a function g. When $\mathbf{W}_{ij} = 0$, i.e., user u_i hasn't rated v_j, it is obvious $\mathbf{W}_{ij}^{e} = 0$, too. Then we need $\mathbf{W}_{ij}$ equal to 0 in this condition. When $\mathbf{W}_{ij} = 1$, i.e., user u_i has rated v_j, then $\mathbf{W}_{ij}^{e}$ increase with $\mathbf{W}_{ij}^{e}$. Thus, $\widetilde{\mathbf{W}}_{ij}$ should also vary with $\mathbf{W}_{ij}^{e}$ in the same direction. In this way, we can construct $g(\mathbf{W}_{ij}, \mathbf{W}_{ij}^{e}) = \mathbf{W}_{ij} + \beta\mathbf{W}_{ij}^{e}$. The parameter β controls the relatively importance of $\mathbf{W}_{ij}$ and $\mathbf{W}_{ij}^{e}$ in the model.

Local Emotion Modeling for Recommendation

As we have proved, users are more likely to agree to their positive emotion reviews, especially compared to their negative emotion reviews. In other words, it indicates that the difference in ratings from user himself and from others with the user's positive emotion, is much lower than those with negative emotion. However, at the start, users only rated limited items, but their future ratings should also coincide with the observations. Thus we can also utilize them on our predicted ratings to face the cold start.

We first define the difference between a predicted rating $\mathbf{u}_i^T\mathbf{v}_j$ and an existing rating $\mathbf{R}_{kj}$ from u_i and u_k on the same item v_j as $d'_{ik,j} = (\mathbf{u}_i^T\mathbf{v}_j - \mathbf{R}_{kj})^2$. Let $d_{ij}^{p} \in \{d'_{ik,j} | \forall k \ s.t. \ \mathbf{R}_{kj} \in \mathcal{P}_{ij}\}$ and $d_{ij}^{n} \in \{d'_{ik,j} | \forall k \ s.t. \ \mathbf{R}_{kj} \in \mathcal{N}_{ij}\}$ denote the difference of the predicted rating $\mathbf{u}_i^T\mathbf{v}_j$ compared to one of u_i's positive and negative emotion reviews respectively. Then to model these emotion information for each (u_i, v_j) pair, there are five cases that we need to discuss.

- $Case\ 1 : \mathcal{P}_{ij} = \varnothing \ and \ \mathcal{N}_{ij} = \varnothing;$
- $Case\ 2 : \mathcal{P}_{ij} \neq \varnothing \ and \ \mathcal{N}_{ij} = \varnothing;$
- $Case\ 3 : \mathcal{P}_{ij} = \varnothing \ and \ \mathcal{N}_{ij} \neq \varnothing;$
- $Case\ 4 : \mathcal{P}_{ij} \neq \varnothing \ and \ \mathcal{N}_{ij} \neq \varnothing \ and \ d_{ij}^{p} - d_{ij}^{n} < 0;$
- $Case\ 5 : \mathcal{P}_{ij} \neq \varnothing \ and \ \mathcal{N}_{ij} \neq \varnothing \ and \ d_{ij}^{p} - d_{ij}^{n} \geq 0.$

Actually, $Case\ 1$ is the scenario for our basic model, and based on our studied datasets, negative opinions may be out of malicious purpose (Mobasher et al. 2007) and the proportion of $Case\ 3$ is pretty low. For example, there are only 0.93% and 0.82% reviews received only negative emotion in our previous datasets Ciao and Epinions, respectively. Thus we only need to consider $Case\ 2, 4$ and 5. For $Case\ 2$, according to the aforementioned observation, $\mathbf{u}_i^T\mathbf{v}_j$ should be close to ratings of positive emotion reviews by minimizing term d_{ij}^{p}. Since $Case\ 4$ satisfies our statistic assumption, we should not penalize this case. While for $Case\ 5$, we need to add a penalty to pull the predicted ratings closer to ratings in $\mathcal{P}_{ij}$. Therefore, it can be formulated by solving the following objective function:

$$\min \ \max(0, d_{ij}^{p} - d_{ij}^{n}), \tag{4}$$

where $d_{ij}^{n} = 0$ for $Case\ 2$ in order to be close to $\mathcal{P}_{ij}$. As users u_i may give more than one positive or negative emotion expressions on ratings of item v_j, we then use $(\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}_{*j}^{ip})^2$ and $(\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}_{*j}^{in})^2$ to respectively denote the average difference value of d_{ij}^{p} and d_{ij}^{n}. The above penalty term can be further reformulated as:

$$\min \ \max(0, (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}_{*j}^{ip})^2 - (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}_{*j}^{in})^2), \tag{5}$$

where $\overline{\mathbf{R}}_{*j}^{ip}$ and $\overline{\mathbf{R}}_{*j}^{in}$ are denoted as the average rating of positive and negative emotion reviews from u_i to v_j, respectively. However, when $\mathcal{P}_{ij}$ or $\mathcal{N}_{ij}$ is empty, we should ignore their corresponding impacts on predicted ratings. Then $\overline{\mathbf{R}}_{*j}^{ip}$ and $\overline{\mathbf{R}}_{*j}^{in}$ are defined as,

$$\overline{\mathbf{R}}_{*j}^{ip} = \begin{cases} \overline{\mathbf{P}}_{ij} & \mathcal{P}_{ij} \neq \varnothing \\ \mathbf{u}_i^T\mathbf{v}_j & otherwise \end{cases}$$

$$\overline{\mathbf{R}}_{*j}^{in} = \begin{cases} \overline{\mathbf{N}}_{ij} & \mathcal{N}_{ij} \neq \varnothing \\ \mathbf{u}_i^T\mathbf{v}_j & otherwise \end{cases} \tag{6}$$

where $\overline{\mathbf{P}}_{ij}$ and $\overline{\mathbf{N}}_{ij}$ represent the average rating of positive and negative emotion reviews, respectively. Actually, for any user-item pair (i, j) of $\overline{\mathbf{R}}^p$ and $\overline{\mathbf{R}}^n$, we just need to calculate $\overline{\mathbf{R}}^{ip}_{*j}$ and $\overline{\mathbf{R}}^{in}_{*j}$ once, since the corresponding term in Eq.(5) will always equal to 0 when $\mathcal{P}_{ij}$ or $\mathcal{N}_{ij}$ is empty.

Then we can find a unified term to locally model positive and negative emotion as:

$$\min \sum_{i=1}^{n}\sum_{j=1}^{m} \max(0, (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{ip}_{*j})^2 - (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{in}_{*j})^2). \tag{7}$$

Objective Function of MIRROR

With the introduction of emotion regularization to model both positive and negative emotion globally and locally, the final objective function of MIRROR is to minimize the following equation,

$$\begin{aligned}\min_{\mathbf{U},\mathbf{V}} &||\widetilde{\mathbf{W}} \odot (\mathbf{R} - \mathbf{U}^T\mathbf{V})||_F^2 + \alpha(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2) \\ &+ \gamma \sum_{i=1}^{n}\sum_{j=1}^{m} \max(0, (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{ip}_{*j})^2 - (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{in}_{*j})^2),\end{aligned} \tag{8}$$

where γ is introduced to control its local contribution of emotion regularization to model emotion on other users' reviews. Note that our proposed framework is a general model. Except dividing users' ratings on others' reviews into negative and positive emotion, there is another mechanism to express emotion, which submit "like" or "dislike", and "thumb up" or "thumb down". Those kinds of mechanisms can be naturally adapted to our framework.

Optimization Algorithm

The objective function in Eq.(8) is not convex if we update all the variables jointly. To optimize the objective function, following the common way, we update $\mathbf{U}$ and $\mathbf{V}$ alternatively by fixing one variable and update the other one using gradient descent. We use $\mathcal{J}$ to denote the objective function of Eq.(8) in the k-th iteration as follows:

$$\begin{aligned}\mathcal{J} = \min_{\mathbf{U},\mathbf{V}} &||\widetilde{\mathbf{W}} \odot (\mathbf{R} - \mathbf{U}^T\mathbf{V})||_F^2 + \alpha(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2) \\ &+ \gamma \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{M}^k_{ij}((\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{ip}_{*j})^2 - (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{in}_{*j})^2),\end{aligned} \tag{9}$$

where we define $\mathbf{M}^k \in \mathbb{R}^{n\times m}$ in the k-th iteration as,

$$\mathbf{M}^k_{ij} = \begin{cases} 1 & (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{ip}_{*j})^2 - (\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{in}_{*j})^2 > 0 \\ 0 & otherwise \end{cases} \tag{10}$$

The derivatives of $\mathcal{J}$ with respect to $\mathbf{u}_i$ and $\mathbf{v}_j$ are given as:

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \mathbf{u}_i} =& 2\sum_j \widetilde{\mathbf{W}}^2_{ij}(\mathbf{u}_i^T\mathbf{v}_j - \mathbf{R}_{ij})\mathbf{v}_j + 2\alpha\mathbf{u}_i \\ &+ 2\gamma\sum_j \mathbf{M}^k_{ij}(\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{ip}_{*j})\mathbf{v}_j \\ &- 2\gamma\sum_j \mathbf{M}^k_{ij}(\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{in}_{*j})\mathbf{v}_j\end{aligned} \tag{11}$$

Algorithm 1 The Proposed Framework MIRROR

Input: $\mathbf{R}$, $\mathbf{E}$, K, α, β, γ

Output: Predicted rating matrix $\widehat{\mathbf{R}}$

1: Compute $\widetilde{\mathbf{W}}$, $\mathcal{P}$ and $\mathcal{N}$ based on $\mathbf{R}$, $\mathbf{E}$ and β.
2: Initialize $\mathbf{U}$ and $\mathbf{V}$ randomly
3: Calculate $\overline{\mathbf{R}}^p$ and $\overline{\mathbf{R}}^n$ using Eq.(6)
4: **repeat**
5: Calculate $\mathbf{M}$ using Eq.(10)
6: Calculate $\frac{\partial \mathcal{J}}{\partial \mathbf{U}}$ using Eq.(11)
7: Calculate $\frac{\partial \mathcal{J}}{\partial \mathbf{V}}$ using Eq.(12)
8: Update $\mathbf{U}$ as $\mathbf{U} = \mathbf{U} - \tau_u \frac{\partial \mathcal{J}}{\partial \mathbf{U}}$
9: Update $\mathbf{V}$ as $\mathbf{V} = \mathbf{V} - \tau_v \frac{\partial \mathcal{L}}{\partial \mathbf{V}}$
10: **until** convergence
11: return $\widehat{\mathbf{R}} = \mathbf{U}^T\mathbf{V}$

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \mathbf{v}_j} =& 2\sum_i \widetilde{\mathbf{W}}^2_{ij}(\mathbf{u}_i^T\mathbf{v}_j - \mathbf{R}_{ij})\mathbf{u}_i + 2\alpha\mathbf{v}_j \\ &+ 2\gamma\sum_i \mathbf{M}^k_{ij}(\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{ip}_{*j})\mathbf{u}_i \\ &- 2\gamma\sum_i \mathbf{M}^k_{ij}(\mathbf{u}_i^T\mathbf{v}_j - \overline{\mathbf{R}}^{in}_{*j})\mathbf{u}_i\end{aligned} \tag{12}$$

With the gradients given above, we summarize the algorithm in Algorithm 1. Next, we briefly review Algorithm 1. We first randomly initialize user latent matrix $\mathbf{U}$ and item latent matrix $\mathbf{V}$. Then we calculate $\overline{\mathbf{R}}^p$ and $\overline{\mathbf{R}}^n$ in line 3. In each iteration, $\mathbf{M}^k_{ij}$ is updated in line 5. From line 6 to 9, we update $\mathbf{U}$ and $\mathbf{V}$ until achieving convergence, and τ_u, τ_v are the learning rates for updating $\mathbf{U}$ and $\mathbf{V}$. After convergence, the predicted rating is computed as $\widehat{\mathbf{R}} = \mathbf{U}^T\mathbf{V}$.

Experiments

In this section, we conduct experiments to demonstrate the effectiveness of our proposed framework MIRROR. Through the experiments, we aim to answer two questions:

- Does emotion on reviews improve the recommendation performance? and
- Does emotion on reviews help to solve the cold-start problem for recommendation?

Next, we first introduce the experiment settings followed by experiments to answer the two questions. Further experiments are conducted to understand the sensitivity of MIRROR to the hyper-parameters.

Experimental Settings

We choose two widely used evaluation metrics, i.e., Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), to evaluate the recommendation performance, which can be formally defined as $\sum_{(i,j)\in\mathcal{T}} |\widehat{\mathbf{R}}_{ij} - \mathbf{R}_{ij}|/|\mathcal{T}|$ and $\sqrt{\sum_{(i,j)\in\mathcal{T}}(\widehat{\mathbf{R}}_{ij} - \mathbf{R}_{ij})^2/|\mathcal{T}|}$, respectively. $\mathcal{T}$ denotes the set of ratings in the testing set, $|\mathcal{T}|$ is the size of $\mathcal{T}$. Note that previous work demonstrated that small improvement in

Table 2: Recommendation comparisons in terms of MAE and RMSE

Datasets	x%	MAE					RMSE				
		UCF	WNMF	MF	QMF	**MIRROR**	UCF	WNMF	MF	QMF	**MIRROR**
Ciao	10%	1.5701	1.9987	1.1483	1.1471	**1.0490**	2.1078	2.6155	1.4258	1.4249	**1.3449**
	20%	1.4407	1.1044	0.9499	0.9495	**0.8874**	1.9417	1.4782	1.2118	1.2116	**1.1659**
	40%	1.1287	0.9059	0.8473	0.8470	**0.8224**	1.5582	1.1752	1.1012	1.1009	**1.0839**
Epinions	10%	1.7971	1.2922	1.0849	1.0838	**1.0437**	2.2771	1.6600	1.3551	1.3536	**1.3279**
	20%	1.7365	1.0406	0.9545	0.9642	**0.9356**	2.2076	1.3280	1.2456	1.2223	**1.1953**
	40%	1.1889	0.9370	0.8867	0.8859	**0.8705**	1.6321	1.1999	1.1404	1.1392	**1.1265**

Table 3: Recommendation comparisons with 5% cold-start users in terms of MAE and RMSE.

Datasets	x%	MAE					RMSE				
		UCF	WNMF	MF	QMF	**MIRROR**	UCF	WNMF	MF	QMF	**MIRROR**
Ciao	10%	1.6301 -3.82%	2.1537 -7.76%	1.1761 -2.42%	1.1748 -2.41%	**1.0508** **-0.17%**	2.1607 -2.51%	2.7662 -5.76%	1.4543 -2.00%	1.4531 -1.98%	**1.3470** **-0.16%**
	20%	1.5206 -5.55%	1.2902 -16.82%	0.9795 -3.12%	0.9790 -3.11%	**0.8891** **-0.19%**	2.0296 -4.53%	1.7804 -20.44%	1.2440 -2.66%	1.2438 -2.66%	**1.1673** **-0.12%**
	40%	1.2682 -12.36%	1.1559 -27.60%	0.8851 -4.46%	0.8846 -4.44%	**0.8277** **-0.64%**	1.7402 -11.68%	1.6414 -39.67%	1.1430 -3.80%	1.1427 -3.80%	**1.0906** **-0.62%**
Epinions	10%	1.8344 -2.08%	1.4433 -11.69%	1.1006 -1.45%	1.0997 -1.47%	**1.0507** **-0.67%**	2.3107 -1.48%	1.8884 -13.76%	1.3710 -1.17%	1.3698 -1.20%	**1.3380** **-0.76%**
	20%	1.7749 -2.21%	1.2113 -16.40%	0.9820 -2.88%	0.9739 -1.01%	**0.9493** **-1.46%**	2.2450 -1.69%	1.6286 -22.64%	1.2742 -2.30%	1.2316 -0.76%	**1.2040** **-0.73%**
	40%	1.3183 -10.88%	1.1809 -26.03%	0.9357 -5.53%	0.9102 -2.74%	**0.8826** **-1.39%**	1.7920 -9.80%	1.6418 -36.83%	1.1638 -2.05%	1.1627 -2.06%	**1.1428** **-1.45%**

MAE or RMSE can have a significant impact on the quality of the top few recommendation (Koren 2008).

In datasets Ciao and Epinions, we randomly select $x\%$ of user ratings and corresponding emotion information related to the selected user ratings as the training set and the remaining $1-x\%$ as the testing set. To investigate the capability of the proposed framework in handling the data sparsity problem, we vary x as $\{10, 20, 40\}$ in this work. We then apply five fold cross validation for all the following experiments, and report the average MAE and RMSE.

Comparisons of Different Recommender Systems

To answer the first question, we compare MIRROR with the following representative methods,

- **UCF**: UCF predicts ratings by aggregating ratings from user u_i's K most similar users. We use the cosine similarity to calculate user-user similarity.
- **MF**: Matrix factorization based collaborative filtering (Koren, Bell, and Volinsky 2009) tries to decompose the user-item rating matrix into two latent matrices to predict ratings, which is the basic model of the proposed framework MIRROR.
- **WNMF**: Weighted nonnegative matrix factorization (Zhang et al. 2006) tries to decompose the user-item rating matrix into nonnegative matrices to predict ratings.
- **QMF**: Review quality aware collaborative filtering (Raghavan, Gunasekar, and Ghosh 2012) regards users' emotion as a measure to capture the quality of the rating and incorporate the quality scores into recommendation.

To fairly compare different methods, we set the parameters for all methods by a grid search strategy. For MIRROR, we empirically set $\alpha = 0.01$, $\beta = 0.1$, $\gamma = 0.5$, $K = 10$. More details about parameter selection for MIRROR will discussed in the following subsections. The comparison results on Ciao and Epinions datasets are shown in Table 2. We can make the following observations:

- All methods increase steadily with the increase of training set's size, and matrix factorization based methods outperform traditional user-orient collaborative factorization methods in general.
- MIRROR consistently outperforms other baseline methods on both datasets with significant performance gain. We perform t-test on results of MIRROR and QMF, it shows MIRROR is significantly better with a significant level of 0.05. The superiority of the proposed MIRROR can be attributed to the utilization of both global and local emotion.
- The performance superiority between MIRROR and other baseline methods increases when the training set become sparser, i.e., the proposed MIRROR performs much better when the training size is small. The proposed MIRROR is more robust to the data sparsity problem, which is due to the exploit of emotion information.

With these observations, we can draw an answer to the first question - with the exploit of rich emotion information, our proposed framework MIRROR not only outperforms the state-of-art recommender systems on recommendation performance but also can mitigate the data sparsity problem in recommender systems.

Recommendation for Cold-start Users

To answer the second question, we random select 5% users from the training set, and remove their ratings from the training set to the test set, while the emotion expressions for these 5% users are kept. In this way, these 5% users can be regarded as cold-start users. The results are shown in Table 3, where numbers inside parentheses denote the performance reductions compared to the performance without cold-start users in Table 2. From the tables, we draw the following observations:

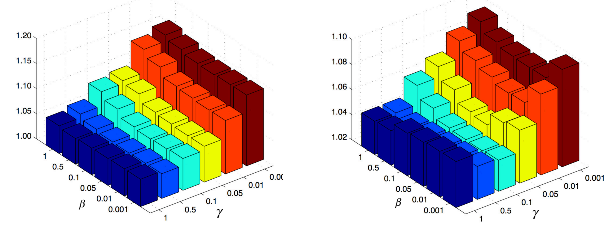
- With the involvement of cold-start users, the performance of every methods decreases. For example, the MAE performance of UCF with cold-start users decreases up to 12.39% compared to performance shown in Table 2.
- MIRROR decreases consistently slower than other baseline methods on both datasets, which means the proposed MIRROR is more robust to the cold-start user problem. That is because of the utilization of emotion information.

In summary, the proposed MIRROR outperforms the state-of-art recommender systems on handling cold-start users.

Parameter Analysis

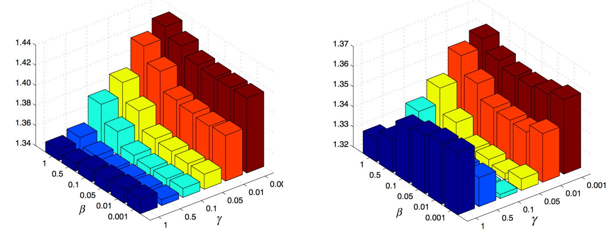
The proposed framework has two important parameters β and γ, which separately controls the contribution of global and local influence of emotion expressions. In this section, we investigate the impact of the parameters β and γ on the performance of MIRROR. We only show results on Ciao and Epinions with 10% without cold-start users since we have similar observations. We vary the values of β as $\{0.001, 0.01, 0.05, 0.1, 0.5, 1\}$ and γ as $\{0.001, 0.01, 0.05, 0.1, 0.5, 1\}$. From Figure 5 and Figure 6, we can have the following observations:

- With the increase of β (or γ), the performance of MIRROR first increases, suggesting that integrating global (or local) emotion can improve the recommendation performance. After the first increase, the performance then decreases in general, which is useful from a practical point of view to select parameters.
- Compared to β, MIRROR is more sensitive to γ. That is because users' local emotion varies while the global emotion information keeps stable. A large γ will lead to personal various emotion dominate the user preference learning process resulting a rapid and significant increase of recommendation performance.



(a) MAE on Ciao (b) MAE on Epinions

Figure 5: Parameter analysis in terms of MAE.



(a) RMSE on Ciao (b) RMSE on Epinions

Figure 6: Parameter analysis in terms of RMSE.

Related Work

To infer users' preferences on products and predict potential interests, product reviews are widely used by recommender systems (Yang et al. 2014; McAuley and Leskovec 2013). Among the great amounts of products, a user's ratings are extremely sparse especially at the very beginning, which leads to the sparsity and cold start problems due to the inadequacy of reviews. Recent years have witnessed great efforts in exploring external information to solve those problems. For example, Ma *et al.* exploit the characteristics of social relations and utilize friends' ratings by assuming that users should share the similar preferences in the rating space and the social relation space (Tang, Aggarwal, and Liu 2016). Tang *et al.* (Tang, Aggarwal, and Liu 2016) exploits signed social network for recommendation. Note that our work is inherently different from Tang *et al.* as we exploit users emotion on other users' reviews, which can directly infer a user's rating on items; while Tang *et al.* exploits trust and distrust network between users instead of user-review relations.

Emotion has been demonstrated to be effective for improving recommender systems (Zheng, Mobasher, and Burke 2016; Costa and Macedo 2013). For example, researchers utilize emotion with time and location to enrich a contextual situation (Zheng, Mobasher, and Burke 2016) or a user profile (Costa and Macedo 2013), but these emotion are not on reviews. Although users' emotion on reviews is a strong indicator to infer users' preferences on products, little attention is paid to exploit users' emotion on reviews to recommender systems. Most research just regards emotion on reviews as a signal to measure the reviews' helpfulness. For example, Wang et al. propose to predict reviews' helpfulness with products reviews, and in return, predict users' preference on products with review helpfulness (Wang, Tang, and Liu 2015). Raghavan et al. attach a weight based on emotion feedback to improve the importance of high quality reviews with probabilistic matrix factorization (Raghavan, Gunasekar, and Ghosh 2012). Our work exploits emotion on reviews for recommender system and proposed a novel unified model from global and local perspective to exploit both positive and negative emotion information to improve recommendation performance, which is significantly different from existing works.

Conclusion

In this paper, we study the novel problem of investigating emotion information for recommender systems. We validate two assumptions of users emotion on reviews and their agreement on ratings on real-world dataset. We propose a new framework guided by the two assumptions which integrates emotion information to matrix factorization based recommender systems. Experimental results show that the proposed MIRROR has significantly better performance than the state-of-the-art methods, and have better capacities to mitigate the data sparsity and cold-start problems. There are several interesting directions need further investigation as further work. For example, we intend to identify spam reviews and spam emotion expressions to obtain a more robust global model for recommender systems.

Acknowledgment

This work is supported by, or in part by, National Science Foundation of China (61672500, 61572474 and 61402446), and Program of International S&T Cooperation (2016YFE0121500). Suhang Wang and Huan Liu are supported by the National Science Foundation (NSF) under the grant #1614576 and Office of Naval Research (ONR) under the grant N00014-16-1-2257.

References

Beigi, G.; Tang, J.; Wang, S.; and Liu, H. 2016. Exploiting emotional information for trust/distrust prediction. In *SDM*, 81–89.

Beutel, A.; Cheng, Z.; Cheng, Z.; Pham, H.; and Anderson, J. 2017. Beyond globally optimal: Focused learning for improved recommendations. In *WWW*, 203–212.

Bewsell, G. R. 2012. Distrust, fear and emotional learning: an online auction perspective. *Journal of Theoretical & Applied Electronic Commerce Research* 7(7):1–12.

Costa, H., and Macedo, L. 2013. Emotion-based recommender system for overcoming the problem of information overload. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, 178–189.

Dunn, J. R., and Schweitzer, M. E. 2005. Feeling and believing: the influence of emotion on trust. *Journal of Personality & Social Psychology* 88(5):736.

Jamali, M., and Ester, M. 2010. A matrix factorization technique with trust propagation for recommendation in social networks. In *RecSys*, 135–142.

Koren, Y.; Bell, R. M.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *IEEE Computer* 42(8):30–37.

Koren, Y. 2008. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *SIGKDD*, 426–434.

McAuley, J., and Leskovec, J. 2013. Hidden factors and hidden topics: understanding rating dimensions with review text. In *RecSys*, 165–172.

Meng, X.; Wang, S.; Shu, K.; Li, J.; Chen, B.; Liu, H.; and Zhang, Y. 2018. Personalized privacy-preserving social recommendation. In *AAAI*.

Mobasher, B.; Burke, R.; Bhaumik, R.; and Williams, C. 2007. Toward trustworthy recommender systems:an analysis of attack models and algorithm robustness. *Acm Transactions on Internet Technology* 7(4):23.

Raghavan, S.; Gunasekar, S.; and Ghosh, J. 2012. Review quality aware collaborative filtering. In *RecSys*, 123–130.

Shu, K.; Wang, S.; Tang, J.; Wang, Y.; and Liu, H. 2018. Crossfire: Cross media joint friend and item recommendations. In *WSDM*.

Takács, G.; Pilászy, I.; Németh, B.; and Tikk, D. 2008. Investigation of various matrix factorization methods for large recommender systems. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, 6.

Tang, J.; Aggarwal, C.; and Liu, H. 2016. Recommendations in signed social networks. In *WWW*, 31–40.

Tang, J.; Wang, S.; Hu, X.; Yin, D.; Bi, Y.; Chang, Y.; and Liu, H. 2016. Recommendation with social dimensions. In *AAAI*, 251–257.

Wang, S.; Tang, J.; Wang, Y.; and Liu, H. 2015. Exploring implicit hierarchical structures for recommender systems. In *IJCAI*, 1813–1819.

Wang, X.; Lu, W.; Ester, M.; Wang, C.; and Chen, C. 2016. Social recommendation with strong and weak ties. In *CIKM*, 5–14.

Wang, S.; Aggarwal, C.; Tang, J.; and Liu, H. 2017a. Attributed signed network embedding. In *CIKM*, 137–146.

Wang, S.; Tang, J.; Aggarwal, C.; Chang, Y.; and Liu, H. 2017b. Signed network embedding in social media. In *SDM*, 327–335.

Wang, S.; Wang, Y.; Tang, J.; Shu, K.; Ranganath, S.; and Liu, H. 2017c. What your images reveal: Exploiting visual contents for point-of-interest recommendation. In *WWW*, 391–400.

Wang, S.; Tang, J.; and Liu, H. 2015. Toward dual roles of users in recommender systems. In *CIKM*, 1651–1660.

Yang, X.; Guo, Y.; Liu, Y.; and Steck, H. 2014. A survey of collaborative filtering based social recommender systems. *Computer Communications* 41:1–10.

Zhang, S.; Wang, W.; Ford, J.; and Makedon, F. 2006. Learning from incomplete ratings using non-negative matrix factorization. In *SDM*, 549–553.

Zheng, Y.; Mobasher, B.; and Burke, R. 2016. *Emotions in Context-Aware Recommender Systems*. Springer International Publishing.