Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)

A^{3}NCF: An Adaptive Aspect Attention Model for Rating Prediction

Zhiyong Cheng[1], Ying Ding[2], Xiangnan He[1], Lei Zhu[3*], Xuemeng Song[4], Mohan Kankanhalli[1]
[1] School of Computing, National University of Singapore, Singapore
[2] Vipshop US Inc., San Jose, CA, USA
[3] School of Information Science and Engineering, Shandong Normal University, China
[4] School of Computer Science and Technology, Shandong University, China
{jason.zy.cheng, leizhu0608}@gmail.com, mohan@comp.nus.edu.sg

Abstract

Current recommender systems consider the various aspects of items for making accurate recommendations. Different users place different importance to these aspects which can be thought of as a preference/attention weight vector. Most existing recommender systems assume that for an individual, this vector is the same for all items. However, this assumption is often invalid, especially when considering a user's interactions with items of diverse characteristics. To tackle this problem, in this paper, we develop a novel aspect-aware recommender model named A^{3}NCF, which can capture the varying aspect attentions that a user pays to different items. Specifically, we design a new topic model to extract user preferences and item characteristics from review texts. They are then used to 1) guide the representation learning of users and items, and 2) capture a user's special attention on each aspect of the targeted item with an attention network. Through extensive experiments on several large-scale datasets, we demonstrate that our model outperforms the state-of-the-art review-aware recommender systems in the rating prediction task.

1 Introduction

User ratings on E-commerce websites provide a good guidance for users to choose products. Therefore, *rating prediction* of products for users (who are new to those products) is a practical way to increase revenue for E-commerce companies, as it could guide the recommendation of products to potential customers. Matrix factorization (MF) [Koren *et al.*, 2009] has achieved great success in this task, as demonstrated by the Netflix Prize contest [Bell and Koren, 2007]. Relying on the user-item rating matrix, this method represents users' interests and items' features as latent factor vectors in a common latent space. However, a rating only reflects a user's overall satisfaction towards an item without explaining the underlying rationale. For example, a user may give a high rating to a phone because of its high-resolution camera or powerful battery, which cannot be told by the overall rating. As a result, MF methods cannot achieve fine-grained modeling of user preference on the various aspects of items, resulting in unexplained recommendations and the "cold-start" problem of users with few ratings [McAuley and Leskovec, 2013; Wang *et al.*, 2018; Cheng *et al.*, 2018].

To tackle the above limitations, researchers have paid extensive attentions to the textual reviews. As accompanying information of ratings, reviews contain rich information about users' preferences and items' characteristics, since users usually express their opinions on the aspects they care for. Different approaches have been developed to utilize the textual reviews for recommendations. Aspect-based models rely on external domain knowledge to identify aspect information in reviews for better user preference modeling [Zhang *et al.*, 2014; He *et al.*, 2015]. Alternatively, topic-based methods apply topic models (e.g., Latent Dirichlet Allocation [Blei *et al.*, 2003]) to extract aspect information automatically [McAuley and Leskovec, 2013; Ling *et al.*, 2014; Tan *et al.*, 2016; Cheng *et al.*, 2018]. Because of the powerful representation learning capability of deep learning techniques, researchers have also applied these techniques in reviews, aiming to extract better user preference and item characteristics [Catherine and Cohen, 2017; Zheng *et al.*, 2017].

Although those methods have achieved better performance than the MF methods using only ratings, they overlook the fact that *the attention or preference of a user on an aspect does not stay the same for different items*, which may result in misleading recommendations. It is well recognized that the preference of a user over the various aspects of items is different, which could be denoted as *an aspect preference/attention vector*. Existing methods assume that this vector keeps the same when facing different items. However, a user may pay the most attention to one aspect for an item but focus on another aspect for a different item, even when the two items belong to the same category. For example, users usually have different expectations on the same product with different prices. Taking *"cellphone"* as a typical instance: a user will expect an expensive phone to have high qualities on multiple aspects besides the basic functions, such as *high-resolution camera*, *long battery life*, or *fancy appearance*. On the other hand, for a cheap phone, users will not have high requirements on those aspects while paying more attention to the basic functions such as *the connection quality of making calls*.

Motivated by this observation, we propose an Adaptive

*Corresponding author.

Aspect Attention-based Neural Collaborative Filtering model (or $\mathbf{A^3NCF}$ for short), which can accurately capture the varying attentions that a user pays to each aspect of different items. In A^{3}NCF, a new topic model is developed to extract both users' preferences and items' characteristics simultaneously. It is different from previous topic-based methods [McAuley and Leskovec, 2013; Wang and Blei, 2011; Ling *et al.*, 2014; Tan *et al.*, 2016], which directly apply the LDA [Blei *et al.*, 2003] model in reviews and can only extract items' features. For each pair of user and item, their representations learned from the topic model are subsequently used in a neural collaborative filtering network for (1) guiding the learning of their final latent factors and (2) capturing the attention vector of the user with respect to the various aspects of this particular item with an attention network.[1] Finally, an unknown rating is predicted based on the attentive interaction of the user's and item's final latent factors. To this end, we expect that our model could achieve better rating prediction performance, due to (1) the aspect-aware representation learning of users and items via the powerful non-linear neural networks and (2) the delicately designed attention network of adaptive aspect attention modeling for each user-item pair.

To evaluate the effectiveness of our model, we conduct comprehensive experiments on both Amazon product datasets and the Yelp Dataset 2017 to compare our model with several state-of-the-art methods, which utilize both reviews and ratings with different strategies. Results show that our model outperforms those competitors by a large margin and also verify the effectiveness of the attention mechanism. To sum up, the main contributions of this work are as follows.

- We propose an aspect-aware rating prediction method based on a novel adaptive aspect attention modeling design. In particular, a new topic model is developed to extract both user and item features from reviews to guide the aspect-aware representation learning.
- We introduce an attention network to capture the varying attention vectors of each specific user-item pair.
- We conduct comprehensive experiments on publicly accessible datasets to comparatively evaluate and demonstrate the effectiveness of the proposed method.

2 Related Work

Many approaches have been developed to combine reviews and ratings for improving recommendation performance. In this section, we mainly review approaches falling into the *topic-based* and *deep learning-based* categories, which are closely related to our work.

Topic-based. A general approach of these methods is to extract latent topics from reviews using topic models [McAuley and Leskovec, 2013; Ling *et al.*, 2014; Zhang and Wang, 2016; Tan *et al.*, 2016; Cheng *et al.*, 2018] or non-negative matrix factorization (NMF) [Bao *et al.*, 2014; Qiu *et al.*, 2016] and learn latent factors from ratings using MF methods. Then, the latent topics and latent factors are combined in a way for final rating prediction. For example, HFT [McAuley and Leskovec, 2013] and TopicMF [Bao *et al.*, 2014] use a defined transform function to learn the latent topics and latent factors together. ITLFM [Zhang and Wang, 2016] and RBLT [Tan *et al.*, 2016] linearly combine them to form the final representations for users and items to model the ratings in MF. RMR [Ling *et al.*, 2014] also learns items' features using topic models on reviews, while it models ratings using a mixture of Gaussians rather than MF methods. Those works all apply the standard LDA [Blei *et al.*, 2003] on reviews to extract only items' features. In this paper, we develop a new topic model to directly extract users' preferences and items' features from the reviews.

Deep learning-based. With the success of deep learning in recommender systems [Covington *et al.*, 2016; He *et al.*, 2017], researchers also attempt to leverage textual reviews in deep learning recommendation models [Zhang *et al.*, 2016; Zheng *et al.*, 2017; Catherine and Cohen, 2017; Zhang *et al.*, 2017]. For example, [Zhang *et al.*, 2017] utilized the reviews as inputs of users and items into a neural network to learn their representations for top-n recommendation. UWRL [Tang *et al.*, 2015] and DeepCoNN [Zheng *et al.*, 2017] feed review texts into a feed-forward neural network or CNN for review rating prediction. A limitation of UWRL and DeepCoNN is that they need textual reviews even in the testing phase. TransNet [Catherine and Cohen, 2017] extends DeepCoNN by introducing an additional layer to simulate the review corresponding to the target user-item pair, which is then used for rating prediction.

All the above approaches assume that users' attention weights/vectors on various aspects stay the same for different items. However, this assumption is invalid, especially when considering the diverse characteristics of different items. Recently, [Cheng *et al.*, 2018] emphasized the importance of considering users' varying attentions on different aspects (w.r.t. different items) in recommender systems and proposed an aspect-aware rating prediction model. In their model, a heuristic method is used to estimate the aspect attention weights based on only reviews. In this paper, we propose an attentive neural network to capture users' varying attention vectors towards items using both reviews and ratings.

3 Our Model

3.1 Preliminaries

Problem Setting. Let $\mathcal{D}$ be a review collection for an item (i.e., product) set $\mathcal{I}$ from a specific category (e.g., *clothes*) written by a set of users $\mathcal{U}$, and each review $d_{u,i} \in \mathcal{D}$ comes with an overall rating $r_{u,i}$ to indicate the overall satisfaction of user u towards item i. The primary goal is to predict the unknown ratings of items that the users have not reviewed yet.

Intuition. A user may place *different importance to the various aspects of different items*, or *the attention weights on aspects are varied when facing different items*. Sometimes, an important aspect can dominate a user's attitude towards an item. For example, a fan of the famous NBA player *James Harden* is willing to purchase Adidas basketball shoes just because they are endorsed by this player. But when purchasing other basketball shoes, he would like to carefully consider

[1]A latent factor is regarded as an aspect in our context.

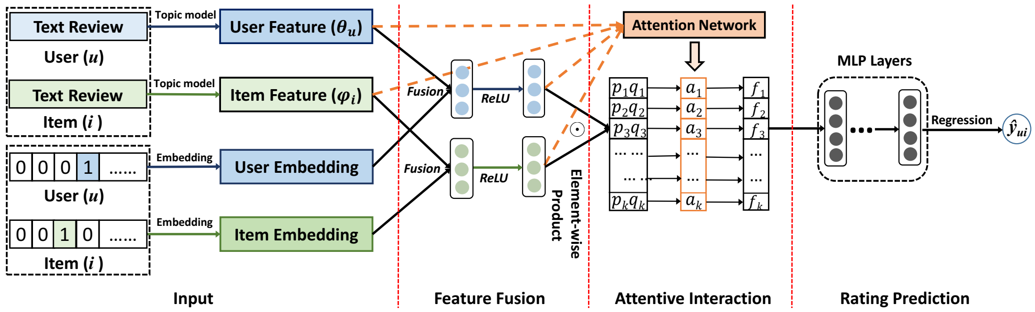


Figure 1: The structure of our A^{3}NCF model.

other factors, such as *whether the shoes are comfortable* and *how good is the cushioning*. Based on this observation, we argue that it is important to capture the attention that a user pays to each aspect of a targeted item when evaluating his attitude towards this item. The problem is how to accurately model users' attention weights towards different items.

When writing a review, users tend to write comments on either the aspects they care for or the most notable features of the targeted item. Therefore, we could extract a user u's preferences and an item i's characteristics of different aspects from reviews, denoted as $\boldsymbol{\theta_u}$ and $\boldsymbol{\varphi_i}$, respectively.[2] The interaction between $\boldsymbol{\theta_u}$ and $\boldsymbol{\varphi_i}$ will be used to estimate the attention vector $\boldsymbol{a_{u,i}}$ of user u for item i. As the mechanism behind the interaction could be very complex, we introduce an attention network to model it, since neural networks have shown strong ability in modeling the complex interactions [Nie *et al.*, 2015; He *et al.*, 2017].

3.2 A^3 Neural Collaborative Filtering

Figure 1 shows the structure of our Adaptive Aspect Attention-based Neural Collaborative Filtering (A^{3}NCF) model, which consists of four components.

The input part takes the reviews of users and items as well as the their identities as inputs. Based on the reviews, a new topic model is applied to extract users' preferences and items' characteristics, represented as $\boldsymbol{\theta_u}$ and $\boldsymbol{\varphi_i}$ for user u and item i, respectively (described in Sect. 3.5). The identity of a user u or an item i is transformed to a binarized sparse vector with an one-hot encoding, which is then projected to a dense vector via an embedding layer. The embedding layer is a fully connected layer. Identity-based embedded features and review-based features of users and items are then passed to the next layer.

The fusion part aims to fuse the embedded features and review-based features for better representation learning. The strategy of combining rating- and review-based features has been widely adopted for boosting recommendation performance in previous works. Different fusion methods can be applied, such as *concatenation*, *addition*, or *element-wise product*. Here, we adopt the *addition* fusion method, which has been applied in RBLT [Tan *et al.*, 2016] and ITLFM [Zhang and Wang, 2016] and achieves good performance. The difference is that we add a fully-connected neural layer directly after the fusion step for more sophisticated effects. This layer adopts the non-linear *ReLU* activation function [Maas *et al.*, 2013]. In experiments, we found that this additional layer can substantially improve the performance.

The attentive interaction part is the core of A^{3}NCF - capturing the targeted user's attention vector for different aspects of the targeted item. Let $\boldsymbol{p_u} \in \mathbb{R}^K$ and $\boldsymbol{q_i} \in \mathbb{R}^K$ be the representation vectors of user u and item i (learned from the fusion part), respectively. K is the dimension of latent vectors. The attentive interaction part outputs the representation of the user-item pair for the subsequent rating prediction. Let $\boldsymbol{F} = [f_1, f_2, ..., f_K]$ denote the output representation of the user-item pair. $\boldsymbol{F}$ is obtained by:

$$\boldsymbol{F} = \boldsymbol{a_{u,i}} \odot (\boldsymbol{p_u} \odot \boldsymbol{q_i}) \tag{1}$$

where $\odot$ denotes the element-wise product, and $\boldsymbol{a_{u,i}} \in \mathbb{R}^K$ is the attention vector of user u for item i. From the equation, we can see $f_k = a_{u,i,k} \cdot p_{u,k} \cdot q_{i,k}$, where f_k denotes the k-th factor in $\boldsymbol{F}$. It indicates that for the interaction of each factor between $\boldsymbol{p_u}$ and $\boldsymbol{q_i}$, there is an attention weight $a_{u,i,k}$ to capture the importance of this factor k of item i with respect to user u, namely, u's attention on the aspect k of item i. Therefore, $a_{u,i,k}$ is unique for each user-item pair. The attention mechanism for $\boldsymbol{a_{u,i}}$ will be introduced in Section 3.3.

The rating prediction part. The obtained interaction feature vector $\boldsymbol{F}$ is fed into fully connected layers as follows:

$$\boldsymbol{z_L} = \sigma_L(\boldsymbol{W}_L(\sigma_{L-1}(\boldsymbol{W}_{L-1} \cdots \sigma_1(\boldsymbol{W}_1\boldsymbol{F}+\boldsymbol{b}_1))+\boldsymbol{b}_{L-1})+\boldsymbol{b}_L)$$

where L denotes the number of hidden layers; $\boldsymbol{W}_l$, $\boldsymbol{b}_l$, and σ_l are the weight matrix, bias vector, and activation function for the l-th layer, respectively. We adopt the *ReLU* activation function [Maas *et al.*, 2013] for all the layers. The predicted rating $\hat{r}_{u,i}$ is obtained via a regression layer:

$$\hat{r}_{u,i} = \boldsymbol{W}\boldsymbol{z}_L + \boldsymbol{b}$$

where $\boldsymbol{W}$ and $\boldsymbol{b}$ are the weight matrix and bias vector, respectively.

3.3 Attention Mechanism

In this section, we introduce the attention mechanism in A^{3}NCF for capturing a user u's specific attention $a_{u,i,k}$ on an aspect/factor k of an item i. Since the user preference and item characteristics can be explicitly observed in reviews, we rely on the review-based features (i.e., $\boldsymbol{\theta_u}$ and $\boldsymbol{\varphi_i}$) to capture a user u's attention on the various aspects of an item i. Meanwhile, because the latent vectors learned in the *feature*

[2]Unless otherwise specified, notations in bold style denote matrices or vectors, and the ones in normal style denote scalars.

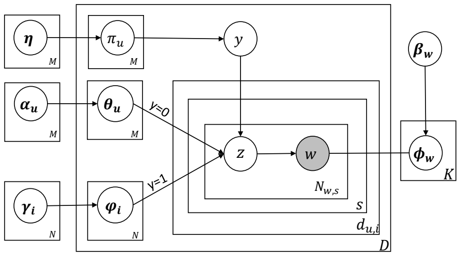


Figure 2: Graphical representation of the proposed topic model.

fusion part (i.e., $\boldsymbol{p_u}$ and $\boldsymbol{q_i}$) are used to learn the final interaction features $\boldsymbol{F}$ of the user-item pair, the attention vector is computed as:

$$\hat{\boldsymbol{a}}_{\boldsymbol{u,i}} = \boldsymbol{v}^T ReLU(\boldsymbol{W}_a[\boldsymbol{\theta_u}; \boldsymbol{\varphi}_i; \boldsymbol{p_u}; \boldsymbol{q_i}] + \boldsymbol{b}_a) \qquad (2)$$

where $\boldsymbol{W}_a$ and $\boldsymbol{b}_a$ are respectively the weight matrix and bias vector that project the input into a hidden layer, and $\boldsymbol{v}^T$ is the vector that projects the hidden layer into an output attention weight vector. $[\boldsymbol{\theta_u}; \boldsymbol{\varphi}_i; \boldsymbol{p_u}; \boldsymbol{q_i}]$ denotes the concatenation of the four feature vectors. *ReLU* is also used as the activation function here due to its effectiveness in neural attention networks [Cao *et al.*, 2018; Song *et al.*, 2018]. Following the standard setting of neural attention networks, $\boldsymbol{a_{u,i}}$ is obtained by a subsequent normalization with the *softmax* function.

$$a_{u,i,k} = \frac{exp(\hat{a}_{u,i,k})}{\sum_{j=1}^{K} exp(\hat{a}_{u,i,j})} \qquad (3)$$

It is worth mentioning that we also tried $[\boldsymbol{\theta_u}; \boldsymbol{\varphi}_i]$ or $[\boldsymbol{p_u}; \boldsymbol{q_i}]$ instead of $[\boldsymbol{\theta_u}; \boldsymbol{\varphi}_i; \boldsymbol{p_u}; \boldsymbol{q_i}]$ in Eq. 2 to calculate the attention weights, which leads to inferior performance. This also validates the assumption of combining them together is more effective.

3.4 Learning

Notice that user preferences and item features extracted from reviews are pre-processed by the proposed topic model (detailed in next subsection). Thus, the learning of A^{3}NCF is the same as general deep learning networks. As our task is rating prediction, we treat it as a regression task and the square loss is used for training. The objective function is

$$\mathcal{L} = \sum\nolimits_{(r_{u,i}) \in \mathcal{D}} (\hat{r}_{u,i} - r_{u,i})^2 \qquad (4)$$

where $r_{u,i}$ is the given rating in dataset $\mathcal{D}$, and $\hat{r}_{u,i}$ is the predicted rating. The stochastic gradient descent (SGD) algorithm is adopted for optimization.

3.5 Topical Feature Extraction

The features extracted from reviews for users and items are critical to our model, as they are not only related to the learning of the final representations for users and items, but also the core to compute the attention weights. As previous works have successfully applied topic models in extracting review features for rating prediction, such as [McAuley and Leskovec, 2013; Ling *et al.*, 2014; Tan *et al.*, 2016], we also use the topic modeling method. Different from those works which directly apply LDA [Blei *et al.*, 2003; Hong *et al.*, 2016b] in reviews to extract only items' characteristics, we develop a new topic model, which is capable of extracting both users' preferences and items' characteristics on different aspects simultaneously.

In our topic model, we assume that a set of latent topics (i.e., K topics) represent all the aspects that users discuss in the reviews. $\boldsymbol{\theta_u}$ is a probability distribution of latent topics, in which each value $\theta_{u,k}$ denotes the relative importance of an aspect to the user.[3] Similarly, $\boldsymbol{\varphi_i}$ is the probability distribution of aspects in item i's characteristics, in which each value $\varphi_{i,k}$ denotes the importance of an aspect k to the item i. $\boldsymbol{\theta_u}$ is determined based on all the reviews written by u; and $\boldsymbol{\varphi_i}$ is learned from the reviews of item i written by all users.

Algorithm 1: Generation Process of Our Topic Model.

for *each topic* $k = 1, ..., K$ **do**
  Draw $\boldsymbol{\phi_{k,w}} \sim Dir(\cdot|\boldsymbol{\beta_w})$;
for *each user* $u \in \mathcal{U}$ **do**
  Draw $\boldsymbol{\theta_u} \sim Dir(\cdot|\boldsymbol{\alpha_u})$;
for *each item* $i \in \mathcal{I}$ **do**
  Draw $\boldsymbol{\varphi_i} \sim Dir(\cdot|\boldsymbol{\gamma_i})$;
for *each review* $d_{u,i} \in \mathcal{D}, u \in \mathcal{U}, i \in \mathcal{I}$ **do**
  for *each sentence* $s \in d_{u,i}$ **do**
    Draw $y \sim Bernoulli(\cdot|\pi_u)$;
    if $y_s == 0$ **then**
      Draw $z_s \sim Multi(\boldsymbol{\theta_u})$;
    if $y_s == 1$ **then**
      Draw draw $z_s \sim Multi(\boldsymbol{\varphi_i})$;
    for *each word* $w \in s$ **do**
      Draw $w \sim Multi(\phi_{z_s,w})$

Based on these assumptions, we develop a new topic model, which is a generative probabilistic model as shown in Fig. 2. In the figure, the shaded circles indicate observed variables, while the unshaded ones represent the latent variables.M and N are the number of users and items, respectively. $N_{w,s}$ is the number of words in a sentence s. Given a corpus $\mathcal{D}$, which contains reviews of users towards items $d_{u,i} \in \mathcal{D}$. Our topic model mimics the processing of writing a review sentence by sentence as shown in Algorithm 1. A sentence usually focus on a single aspect (or topic z in our model), which could be from the user's preferences or from the item's characteristics. To decide the topic z_s for a sentence s, our model introduces an indicator variable $y \in \{0, 1\}$ based on a Bernoulli distribution, which is parameterized by π_u. Specifically, when $y = 0$, the sentence is generated from user u's preference; otherwise, it is generated according to item i's characteristics. Because for different items, users may comment on different aspects, which also reflects users' attentions on aspects of different items. Therefore, π_u is user-dependent, representing the inclinations of user u to comment from his own preferences or from the item i's characteristics.

In the topic model, $\boldsymbol{\alpha_u}$, $\boldsymbol{\gamma_i}$, $\boldsymbol{\beta_w}$, and $\boldsymbol{\eta}$ are pre-defined hyper-parameters and usually set to be symmetric for simplic-

[3]Here, a latent topic is regarded as an aspect as previous works.

Datasets	# users	# items	# ratings	Sparsity
Baby	17,177	7,047	158,311	0.9987
Grocery	13,979	8,711	149,434	0.9988
Home & Kitchen	58,901	28,231	544,239	0.9997
Garden	1,672	962	13,077	0.9919
Sports	31,176	18,355	293,306	0.9995
Yelp2017	169,257	63,300	1,659,678	0.9998

Table 1: Statistics of the evaluation datasets.

ity. Parameters need to be estimated including $\boldsymbol{\theta_u}$, $\boldsymbol{\varphi_i}$, $\boldsymbol{\phi_w}$ and π_u. Different inference methods for parameter learning in topic models have been developed, such as variation inference [Blei *et al.*, 2003] and collapsed Gibbs sampling [Cheng *et al.*, 2016; Cheng and Shen, 2016; Hong *et al.*, 2016a]. In this paper, we adopt the collapsed Gibbs sampling method for parameter inference.

4 Experiments

4.1 Experimental Setup

Datasets. We conducted experiments on two publicly accessible datasets: the Amazon Product Review dataset [McAuley and Leskovec, 2013][4] and the Yelp Dataset 2017.[5] The first dataset contains reviews and metadata of diverse products from Amazon. It contains 24 product categories. We adopted five categories and took the 5-core version for experiments, where each user or item has at least 5 interactions. The five categories are of different sizes and sparsity degrees (as shown in Table 1). The Yelp dataset contains reviews of local businesses in 12 metropolitan areas across 4 countries. We preprocessed all the datasets by removing duplicated reviews of the same user-item pairs and keeping only users and items with at least 5 reviews. For each review in the datasets, we extracted 'userID", "itemID", the corresponding rating score (1 to 5 rating stars) and textual review for experiments. In addition, we cleaned the reviews by removing punctuation, stopwords, and infrequent terms appearing less than 10 times in the datasets. The basic statistics of the datasets is shown in Table 1.

Experimental Settings. We randomly split each dataset into training, validation, and testing sets with ratio 8:1:1 for each user as in [McAuley and Leskovec, 2013; Ling *et al.*, 2014; Catherine and Cohen, 2017]. As each user has at least 5 reviews, we have at least 3 reviews per user for training, and at least 1 interaction per user for validation and testing. Note that we only used the review information in the training phrase, because reviews are unavailable in validation and testing phase in real-world scenarios.

Competitors. We compare the proposed model to several state-of-the-art methods which apply diverse strategies in exploiting both reviews and ratings for rating prediction. These methods are tuned on the validation dataset to obtain their optimal hyper-parameter settings for fair comparisons.

- **BMF.** It is a standard matrix factorization (MF) method with biased terms [Koren *et al.*, 2009].

[4]http://jmcauley.ucsd.edu/data/amazon/

[5]http://www.yelp.com/dataset_challenge/

Dataset	BMF	RMR	HFT	RBLT	TransNet	A^{3}NCF
Baby	1.176	1.152	1.117	1.119	1.098	**1.082***
Grocery	1.126	1.063	1.009	1.011	0.993	**0.985**
H & K	1.108	1.092	1.082	1.086	1.074	**1.051***
Garden	1.099	1.074	1.037	1.034	1.040	**1.021***
Sports	1.087	1.011	0.972	0.963	0.983	**0.940***
Yelp2017	1.284	1.263	1.185	1.204	1.183	**1.152***

Table 2: Comparisons of adopted methods in terms of RMSE with $K = 25$. The best performance is highlighted with bold face. The symbol * denotes a significance with $p - value < 0.05$ over the second best model based on a two-tailed paired t-test. "H & K" denotes "Home & Kitchen".

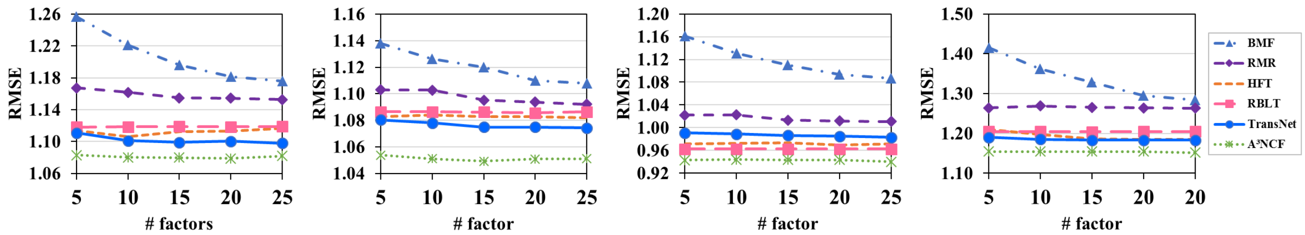
- **HFT.** It models ratings and review texts using MF and LDA [Blei *et al.*, 2003], respectively. Then an exponential transform function is used to associate the latent topics and latent factors for rating prediction [McAuley and Leskovec, 2013].
- **RMR.** Different from HFT and CTR, which use MF to model rating, it uses a mixture of Gaussian distributions to model the ratings [Ling *et al.*, 2014].
- **RBLT.** Similar to HFT, this method uses MF to model ratings and LDA to model review text. Different from HFT, it linearly combines the latent factors (learned from ratings) and latent topics (learned from reviews) to represent users and items [Tan *et al.*, 2016]. This strategy is also adopted by ITLFM [Zhang and Wang, 2016]. Here, we use RBLT as a representative method for this strategy.
- **TransNet.** This method adopts a neural network framework for rating prediction. Reviews of users and items are passed into two CNNs respectively to learn the latent representations of users and items [Catherine and Cohen, 2017]. The latent representations of a targeted user and a targeted item are concatenated and passed through a regression layer to estimate the rating.

Evaluation Metrics. The root-mean-square error (**RMSE**) is used in evaluation. A smaller RMSE indicates better performance.

Parameter settings. The number of MLP layers L in the rating prediction part of A^{3}NCF is set to 2 in our implementation. Besides, the dropout technique [Srivastava *et al.*, 2014] is used to prevent overfitting and the dropout ratio is 0.5. The learning rate is set to 0.001 for all the datasets and the Adam optimization method [Kingma and Ba, 2014] is used.

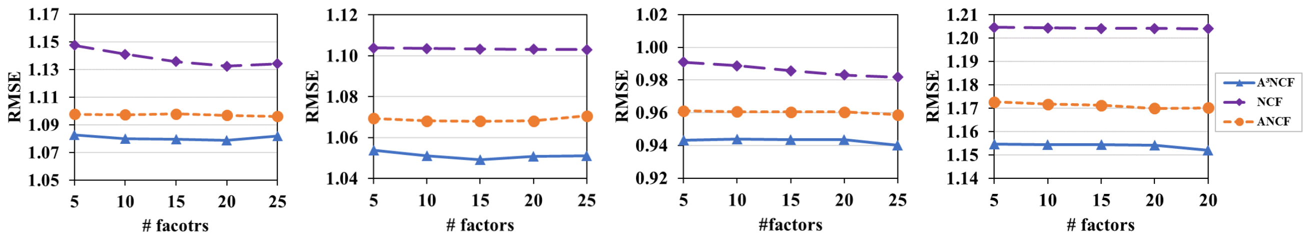
4.2 Experimental Results

Performance Comparison. Figure 3 shows the performance of all the adopted methods with respect to different numbers of predictive factors K. Due to the space limitation, we only show the results of four relatively larger datasets. We also report the concrete values of RMSE (based on $K = 25$) on all the six evaluation datasets in Table 2. Firstly, methods incorporating reviews all achieve better performance than BMF with a large margin, indicting the importance of utilizing reviews in preference modeling. Secondly, our method achieves the best performance over all datasets, significantly outperforming the state-of-the-art methods RBLT



(a) Baby (b) Home & Kitchen (c) Sports (d) Yelp 2017

Figure 3: Performance of all competitors *w.r.t.* the number of latent factors on four relatively larger datasets.



(a) Baby (b) Home & Kitchen (c) Sports (d) Yelp 2017

Figure 4: Performance of variants *w.r.t.* the number of latent factors on four larger datasets.

and TransNet. Based on the results in Table 2, the relative improvement over RBLT and TransNet is 2.9% and 2.2% respectively with significance testing. The performance of RBLT, TransNet, and HFT is comparable across all the datasets, followed by RMR. Lastly, with the increase of number of latent factors ($\#factors$), all the methods could obtain relatively better results.[6] Compared to BMF, review-based methods are relatively stable as they could already achieve relatively good performance with a small $\#factors$. It demonstrates the benefits of considering review information in preference modeling, especially when $\#factors$ is small. Compared to the baselines, our model achieves very good performance when $\#factors$ is only 5.

The substantial improvement of our model over the baselines could be credited to two reasons: (1) our model integrates features extracted from ratings and reviews via non-linear neural networks, and more importantly, applies more complicate interactions between users' and items' latent vectors instead of a simple concatenation; and (2) our model uses an attention mechanism to capture users' attention weights on different aspects of each item, and thus could achieve more accurate predictions.

Effects of Aspect Attention. In the design of A^3NCF, we assume that users place different importance on the same aspect of different items. To verify this assumption and validate the effectiveness of the attention mechanism in our model, we compare our model to the following two variants:

- **NCF.** It is a simplified version of our model without considering review information and attention mechanism. Thus, only the one-hot encodings of a user's and an item's identities are used as input in the networks.
- **Aspect-aware NCF (ANCF).** It is a variant of A^3NCF without the use of attention mechanism. Notice that it uses the review-based features extracted by our topic model.

The performance is shown in Figure 4. Similarly, we also only show the results on four relatively larger datasets. From the figures, we can observe: (1) ANCF greatly improves the prediction accuracy over NCF, which shows the effectiveness of using reviews for modeling users' preferences in recommendation. It also demonstrates the effectiveness of our networks on integrating the reviews and ratings. (2) A^3NCF outperforms ANCF with a large margin, which verifies our assumption that a user's attentions over aspects are varied for different items. This result also validates that our attention network in A^3NCF can effectively capture the attention weights for each user-item pair.

5 Conclusions

In this paper, we presented a novel adaptive aspect attention-based neural collaborative filtering (A^3NCF) model for rating prediction. In this model, a new topic model is developed to extract users' preferences and items' characteristics on different aspects from reviews. Besides, an attention network is introduced in A^3NCF to utilize the features extracted from reviews, aiming to capture the attention that a user pays to each aspect of the targeted item. This special design is due to the observation that a user may place different importance to the same aspect of different items. This is the first attempt to design a neural attention network based on reviews and ratings to tackle this problem. Experiments on

[6]Notice that if a too large $\#factors$ may cause overfitting, leading to performance deterioration. By increasing $\#factors$ till to 25 in our experiments, we have not observed obvious overfitting yet.

real-world datasets from Amazon show that our method significantly outperforms state-of-the-art methods.

Acknowledgments

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centre in Singapore Funding Initiative.

References

[Bao *et al.*, 2014] Yang Bao, Hui Fang, and Jie Zhang. Topicmf: Simultaneously exploiting ratings and reviews for recommendation. In *AAAI*, volume 14, pages 2–8, 2014.

[Bell and Koren, 2007] Robert M Bell and Yehuda Koren. Lessons from the netflix prize challenge. *ACM SIGKDD Explorations Newsletter*, 9(2):75–79, 2007.

[Blei *et al.*, 2003] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan):993–1022, 2003.

[Cao *et al.*, 2018] Da Cao, Xiangnan He, Lianhai Miao, Yahui An, Chao Yang, and Richang Hong. Attentive group recommendation. In *ACM SIGIR*, 2018.

[Catherine and Cohen, 2017] Rose Catherine and William Cohen. Transnets: Learning to transform for recommendation. In *ACM RecSys*, pages 288–296, 2017.

[Cheng and Shen, 2016] Zhiyong Cheng and Jialie Shen. On effective location-aware music recommendation. *ACM Transactions on Information Systems (TOIS)*, 34(2):13, 2016.

[Cheng *et al.*, 2016] Zhiyong Cheng, Shen Jialie, and Steven CH Hoi. On effective personalized music retrieval by exploring online user behaviors. In *ACM SIGIR*, pages 125–134, 2016.

[Cheng *et al.*, 2018] Zhiyong Cheng, Ying Ding, Lei Zhu, and Kankanhalli Mohan. Aspect-aware latent factor model: Rating prediction with ratings and reviews. In *WWW*, pages 639–648, 2018.

[Covington *et al.*, 2016] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *ACM RecSys*, pages 191–198, 2016.

[He *et al.*, 2015] Xiangnan He, Tao Chen, Min-Yen Kan, and Xiao Chen. Trirank: Review-aware explainable recommendation by modeling aspects. In *ACM CIKM*, pages 1661–1670, 2015.

[He *et al.*, 2017] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. Neural collaborative filtering. In *WWW*, pages 173–182, 2017.

[Hong *et al.*, 2016a] Richang Hong, Zhenzhen Hu, Ruxin Wang, Meng Wang, and Dacheng Tao. Multi-view object retrieval via multi-scale topic models. *IEEE Trans. Image Process.*, 25(12):5814–5827, 2016.

[Hong *et al.*, 2016b] Richang Hong, Luming Zhang, Chao Zhang, and Roger Zimmermann. Flickr circles: aesthetic tendency discovery by multi-view regularized topic modeling. *IEEE Trans. Multimed.*, 18(8):1555–1567, 2016.

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Koren *et al.*, 2009] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):42–49, 2009.

[Ling *et al.*, 2014] Guang Ling, Michael R Lyu, and Irwin King. Ratings meet reviews, a combined approach to recommend. In *ACM RecSys*, pages 105–112, 2014.

[Maas *et al.*, 2013] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. ICML*, 2013.

[McAuley and Leskovec, 2013] Julian McAuley and Jure Leskovec. Hidden factors and hidden topics: understanding rating dimensions with review text. In *ACM RecSys*, pages 165–172, 2013.

[Nie *et al.*, 2015] Liqiang Nie, Meng Wang, Luming Zhang, Shuicheng Yan, Bo Zhang, and Tat-Seng Chua. Disease inference from health-related questions via sparse deep learning. *IEEE Trans. Knowledge Data Eng.*, 27(8):2107–2119, 2015.

[Qiu *et al.*, 2016] Lin Qiu, Sheng Gao, Wenlong Cheng, and Jun Guo. Aspect-based latent factor model by integrating ratings and reviews for recommender system. *Knowledge-Based Systems*, 110:233–243, 2016.

[Song *et al.*, 2018] Xuemeng Song, Fuli Feng, Xianjing Han, Xin Yang, Wei Liu, and Liqiang Nie. Neural compatibility modeling with attentive knowledge distillation. In *ACM SIGIR*, 2018.

[Srivastava *et al.*, 2014] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[Tan *et al.*, 2016] Yunzhi Tan, Min Zhang, Yiqun Liu, and Shaoping Ma. Rating-boosted latent topics: Understanding users and items with ratings and reviews. In *IJCAI*, 2016.

[Tang *et al.*, 2015] Duyu Tang, Bing Qin, Ting Liu, and Yuekui Yang. User modeling with neural network for review rating prediction. In *IJCAI*, pages 1340–1346, 2015.

[Wang and Blei, 2011] Chong Wang and David M Blei. Collaborative topic modeling for recommending scientific articles. In *ACM KDD*, pages 448–456, 2011.

[Wang *et al.*, 2018] Xiang Wang, Xiangnan He, Fuli Feng, Liqiang Nie, and Tat-Seng Chua. TEM: Tree-enhanced embedding model for explainable recommendation. In *WWW*, pages 1543–1552, 2018.

[Zhang and Wang, 2016] Wei Zhang and Jianyong Wang. Integrating topic and latent factors for scalable personalized review-based rating prediction. *IEEE Transactions on Knowledge and Data Engineering*, 28(11):3013–3027, 2016.

[Zhang *et al.*, 2014] Yongfeng Zhang, Guokun Lai, Min Zhang, Yi Zhang, Yiqun Liu, and Shaoping Ma. Explicit factor models for explainable recommendation based on phrase-level sentiment analysis. In *ACM SIGIR*, pages 83–92, 2014.

[Zhang *et al.*, 2016] Wei Zhang, Quan Yuan, Jiawei Han, and Jianyong Wang. Collaborative multi-level embedding learning from reviews for rating prediction. In *IJCAI*, pages 2986–2992, 2016.

[Zhang *et al.*, 2017] Yongfeng Zhang, Qingyao Ai, Xu Chen, and W. Bruce Croft. Joint representation learning for top-n recommendation with heterogeneous information sources. In *ACM CIKM*, pages 1449–1458, 2017.

[Zheng *et al.*, 2017] Lei Zheng, Vahid Noroozi, and Philip S Yu. Joint deep modeling of users and items using reviews for recommendation. In *ACM WSDM*, pages 425–434, 2017.