

Received April 20, 2020, accepted May 3, 2020, date of publication May 8, 2020, date of current version May 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2993289

Dynamic Collaborative Filtering Based on User Preference Drift and Topic Evolution

CHARINYA WANGWATCHARAKUL AND SARTRA WONGTHANAVASU[✉], (Member, IEEE)

Department of Computer Science, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand

Corresponding author: Sartra Wongthanavas (wongsar@kku.ac.th)

This work was supported in part by the Thailand Research Fund (TRF) under Grant RTA6080013 and in part by the Computer Science Department of Khon Kaen University, Thailand.

ABSTRACT Recommender systems are efficient tools for online applications; these systems exploit historical user ratings on items to make recommendations of items to users. This paper aims to enhance dynamic collaborative filtering on recommender systems under volatile conditions in which both users' preferences and item properties dynamically change over time. Moreover, existing collaborative filtering models mainly rely on solving data sparsity by adding side information to improve performance. We propose a model to capture the user preference dynamics in the rating matrix by using a joint decomposition method to extract user latent transition patterns and combine latent factors together with the associated topic evolution of review texts by using topic modeling based on the dynamic environment. We evaluate the accuracy on real datasets, and the experimental results show that the model leads to a significant improvement compared with the state-of-the-art dynamic CF models.

INDEX TERMS Recommender systems, user preference drift, topic evolution, dynamic collaborative filtering, data sparsity.

I. INTRODUCTION

The use of recommendation systems (RSs) has become increasingly widespread in the field of online applications. While browsing, users can be provided with suggestions based on the types of movies, books or musics that they search for. Well-known companies, including Amazon, LinkedIn, and Netflix, employ systems that make product recommendations that match the buyer's interests to induce customers to make further purchases [1]. Using previous ratings, recommendation systems wish to find and understand users' preferences and tastes by analyzing online information. Most recommendation models for predicting ratings are based on collaborative filtering (CF), where the user's preferences are predicted based on the relationship between users and items [6], [7], [45], [46].

In fact, users' preferences may drift over time, which tends to frequently redefine their tastes [4], [63]. The changes in users' preferences can originate from substantial reasons, like personality shift, or transient and circumstantial ones, like seasonal changes in item popularities. Disregarding these

temporal drifts in modeling users' preferences can result in unhelpful recommendations. For example, in the movies domain, the user can change his rating behavior, her preference for genre over language, or start favoring drama over comedy. Additionally, besides rating information provided by users, users can provide feedback in the form of tags or written product reviews. Such user-generated content can serve as a useful source for deriving explanatory information that may increase the user's understanding of the underlying criteria and mechanisms that led to the results. Many e-commerce leaders such as Amazon and Netflix have made recommender systems a crucial part of their services assisting enhance user satisfaction and loyalty. The user preference dynamics have a significant impact on RSs. Therefore, accurately estimating user's interests dynamically is challenge in this research domain, and it has attracted considerable attentions by users from systems.

Many of the researches have been carried out to consider temporal dynamics, in which users' preferences drift over time. For example, [25] used the drift and the decay factors for tracking the changes of the global behaviour of users and items to perform rating predictions but did not use the personal behaviour. [21] provided a novel algorithm to

The associate editor coordinating the review of this manuscript and approving it for publication was Pasquale De Meo.

compute the time weights for different items by decreasing weights to historical data. This approach deemphasizes the importance of historical data; however, the historical data might be helpful to facilitate other users' recommendation. [24] proposed a model for capturing a customer's purchasing interest on products recommendation. The model focused on the current stage and propose a clustering-based method and a multi-modal graph ranking method to obtain the recommended item list. Additionally, [37] proposed a method for detecting a user's time-variant pattern in order to improve the performance of collaborative filtering. Further, social media systems, considered by [59] for dynamic user preference and recommendation, used intrinsic interests and temporal context-aware mixture models. However, most of them only consider the temporal dynamics between users and items within a relatively short time period, which cannot explicitly capture the exact interest of users.

Another issue with CF approaches is data sparsity. The operation of missing rating prediction in the user-item interaction will be drastically reduced if the data are limited. The sparse data leads to generate potentially untrustworthy predictions [2], [35]. Although CF models can effectively predict temporal rating data exist, their capability of tracking the dynamics of user preferences is limited on sparse data, and there are few dynamic CF models that both consider drift on the basis of users' preferences and handle data sparsity [53], [60].

In recent years, review texts, which provide additional useful information that can be interpreted as user interests, have contributed to improving the prediction accuracy [13]. Reviews are created by users to provide their opinions to support the reason why they give a rating to appearance items and which feature of the item applies most to its ratings. To solve the rating sparsity issue, the review topics can be helpful for enriching the available ratings by using topic modeling to mine the underlying structure hidden in the unstructured texts automatically. For example, [34] considered the temporal changes in user and item latent features together with the associated review texts in a single learning stage. Additionally, [36] proposed hidden factors as topics (HFT) to exploit review topics and latent factors to enhance the rating, and [38] showed that review texts and topic profiles indeed correlate with ratings. They used review topics to weight ratings, but they did not address the review topic evolution. To consider the dynamics of user preferences, we assume that the review topics dynamically change over time, and we focus on discovering hidden review topics based on the underlying temporal patterns guiding the topic evolution.

There are a few approaches that have handled user preference drift and other subproblems, such as data sparsity [29], [30], [42], [55]. Recently, OCF-DR [29] use neighborhood factor to track the drift of users' preference. TCMC [42] was proposed by combining multimodal information with scores to reduce data sparsity, but this model also considered multimodal data that did not uncover the hidden topic factor from reviews to explain the score on each time step.

Motivated by this idea, it is interesting to learn how RSs can analyze text reviews and ratings to solve the data sparsity problem based on user preference dynamics and achieve higher accuracy. The contributions of this work are summarized as follows:

- We propose a novel model to capture changes in users' preferences in the rating matrix and adapt the influence of review texts to improve the accuracy for data sparsity over the time evolution.
- We exploit users' opinions as review texts to find hidden topic evolutions that can explain why a user rates the score by using a topic modeling technique at different time steps. We believe that the review texts play an important role in users' decisions.
- We evaluate the accuracy on seven widely used datasets to demonstrate the effectiveness of the proposed model, and the experimental results indicate that the model leads to a significant improvement compared with the state-of-the-art dynamic CF models.

The remainder of this paper is organized as follows. Section II provides a brief discussion of related work. Section III presents the proposed recommendation model and how to address an optimization problem regarding the models. Then, Section IV provides the empirical experimental results of the model compared with the state-of-the-art approaches using well-known datasets. Finally, conclusions, discussions and future work are given in Section V.

II. RELATED WORK

In this section, we review several popular RS approaches, including (1) dynamic CF and (2) topic modeling based on contents.

A. DYNAMIC COLLABORATIVE FILTERING

The existing models of CF RSs are classified into memory-based (neighbor-based) and model-based approaches [46]. Memory-based CF approaches are based on the fact that similar users have similar patterns of rating behavior and users provide similar ratings to similar items. Recently, [6] present a novel method for RS known as IPWR (improved PCC weighted with rating preference behavior) similarity measure. It takes into account the user RPB (rating preference behavior) towards an item rating to improve standard Pearson correlation coefficient similarity measure method. The model result performs better than state-of-the-art similarity measure methods. Conversely, model-based CF approaches use the collection of ratings to learn a model, which is then used to make rating predictions. Newly, [7] proposed the model to consider users' preference similarity with social trust (i.e., both explicit and implicit trust) and create a more complete unified rating profile. They define function significance by considering users' positive and negative opinions to enhance the RS performance of CF-based when only a limited set of ratings is available. However, the models ignore the user preference dynamics domain.

Dynamic RSs have become a new trend in the research field [1], [10], [44], and traditional CF models do not concentrate on changing users' preferences over time; they assume that users' preferences are stationary [4]. Hence, the challenging problem with the approach taken by most researchers is that they mainly ignore the drifting preference. To consider the temporal dynamics, time information has been introduced into CF approaches to model users' dynamically evolving preferences and interests [4], [21], [25].

Since data generally change dynamically over time, there are memory-based models that use time-aware approaches, such as the time-weighted CF [21] proposed to improve the quality by using the weights of rating scores. DTDM [47] exploits clustering and a time impact factor matrix to monitor the degree of user interest drift in the class based on the nearest neighbors' time factor coefficient matrix. Another approach, [15] considers the user's response and the dynamic interests of social user and proposes a similarity measurement. However, the main drawback of techniques is the requirement of loading a large amount of in-line memory.

In contrast, the dynamic CF of model-based approaches can instantly respond to user requests [4], [18], [25], [37], [56]. For example, modeling approaches based on probabilistic matrix factorization, such as time singular value decomposition (TimeSVD++) [25], add the time change into the neighborhood model to be combined with the latent factor for capturing users' preferences. Moreover, there are models based on tensor factorization [43], [57], such as Bayesian probabilistic tensor factorization (BPTF) [57] where proposed ratings are represented as triples (user, item, time); then, these triples are organized into a three-dimensional tensor. The models learn the global evolution of latent features, and ratings are predicted using the inner product of the latent factor vector, but the disadvantage of the model is that it is not sensitive enough to capture the local changes in preferences.

However, it is widely known that CF approaches suffer from a data sparsity problem [39], [41], [50]. Recently, a few temporal dynamic models have considered this problem. Temporal collective matrix factorization (TCMF) [42] was proposed considering both temporal dynamics and multimodal information for addressing data sparsity, but they exploit only implicit feedback of user comments and ignore the hidden meaning of additional information. Moreover, TMRvCo [55] considers temporal dynamics and side information. This approach uses a more appropriate item correlation measure in the cofactor and associates the item factors of the cofactor with that of MF, but it focuses on item correlation dimensions, which is a different method than our work.

B. TOPIC MODELING BASED ON CONTENTS

Topic modeling has been applied to interpret contents in unstructured texts. It can be used to automatically extract valuable hidden elements. Conventional topic modeling techniques such as probabilistic latent semantic analysis (PLSA) and latent Dirichlet allocation (LDA) are widely used to infer latent topical structures from documents [9], [14], [20], [23].

In contrast, nonnegative matrix factorization [27], a distinct topic mining method compared to the above probabilistic perspective, models the underlying components as coordinate axes, and each document corresponds to a unique point in the latent topic.

As mentioned in the previous section, a sparse dataset leads to a lack of common rated items, and items cannot be selected to make effective predictions. Hence, the contents are useful information for improving performance. Recently, topic-model-driven recommendation domains are demonstrated in many existing studies [3], [8], [11], [52]. Collaborative topic regression (CTR) [52] utilizes LDA to discover the hidden latent structure of scientific articles. It incorporates the latent topic spaces with traditional collaborative filtering to recommend scientific articles to users of an online community. However, their main assumption is that each item only corresponds to one stationary text, regardless of different users who interact with the item. This does not satisfy that each item has multiple reviews. Moreover, Shivam Bansal *et al.* [8] proposed a latent semantic-based approach by using topic modeling, which automatically provides the best suggestions to determine the most relevant user-job connection mapping according to the skills and preferences of a user.

By incorporating users' review information, some works exploited the meanings of reviews to help the system grasp what the user is most interested in and which aspect of the item contributes most to the rating to overcome the data sparsity problem [13], [32], [51], [62]. For example, HFT model [36] combines ratings with review texts for product recommendations. They focused on the hidden factors in product ratings with hidden topics in product reviews, and exponential transformation was adopted to guarantee that the topic probability distribution of each review was associated with the user or item topic factor. The aspect-aware latent factor model (ALFM) [16] exploits reviews to analyze each user's attention to different aspects of the target item and then integrates the attention weights into the matrix factorization. By associating latent factors with aspects, the aspect weights are integrated with latent factors for rating prediction.

These methods discover static topics. They ignore explaining that the topics that the user prefers can change between past and present, while models that utilize topic modeling to identify the interest of users change over time [19]. In recent years, topic evolution has referred to model topics changing over time [5]. This research helps to overcome information overload and represent complex and dynamic phenomena [58]. The earliest work on topic evolution proposed topic detection and tracking (TDT) [5]. This approach mainly focused on the topic evolution of text streams, which includes topic content evolution and topic strength evolution. Moreover, [49] presented a time-based collective factorization model for topic discovery and monitoring in news to discover and track hidden topics, and this model relies on the underlying temporal patterns guiding the topic evolution. Additionally, this model is able to connect topics between different time slots.

On the other hand, some of the works [12], [17], [54], [64] have explored deep neural networks to perform an in-depth understanding of textual item content and achieved impressive effectiveness by generating more accurate item latent models. In dual-regularized matrix factorization (DRMF) [54], a multilayered neural network model that stacks a convolutional neural network and a gated recurrent neural network is applied to generate independent distributed representations of contents of users and items. Moreover, in deep cooperative neural networks (DeepCoNN) [64], reviews are first processed by two CNNs to learn representations of users and items, which are then concatenated and passed into a regression layer for rating prediction. A limitation of the model is that the performance decreases greatly when reviews are unavailable in the testing phase.

For review information, both topic modeling and deep neural network are used to extract features to represent items, and regularization strategy is mostly used to combine with matrix decomposition. However, deep learning is well known to operate similar to a black box. It requires a massive amount of data to fully support its rich parameterization and provides explainable predictions seem to be a substantially challenging task [33], [40]. Meanwhile, topic modeling can provide the word co-occurrence relation to supplement for information loss. Hence, in this paper we focus on discovering hidden review topics by using topic modeling to guide the topic evolution.

In summary, the state-of-the-art strategies either do not capture users' preference shifts or analyze the topic evolution of user feedback reviews; consequently, they do not perform both of these tasks simultaneously.

III. PRELIMINARIES

In this section, we will present some preliminaries and the formal definition of a CF recommendation, and then we will present the motivation for our work.

We establish some terminologies for the task of dynamic CF to fill in an incomplete rating matrix whose ratings are associated with timestamps. For each user and item, we are given a set of data, where each data point is represented by the following five-tuple:

(user, item, rating, reviews, timestamp)

Specifically, we denote the set of users as $\mathcal{U} = \{u_1, u_2, \dots, u_i\}$ and the set of items as $\mathcal{V} = \{v_1, v_2, \dots, v_j\}$ in each time period $\mathcal{T} = \{t_1, t_{l-1}, t_{l-2}, \dots, t_n\}$. The preference rating is represented by matrix R_{ij} , and we exploit the additional information from the collection of review texts as $\mathcal{D} = \{d_1, d_2, \dots, d_n\}$ to find the hidden factors of items by using the topic distribution and link them together.

Nonnegative matrix factorization (NMF) is a powerful matrix decomposition technique in which the domain of the data is inherently nonnegative and parts-based decompositions are desired [28]. In general, NMF seeks a nonnegative matrix $W \in \mathbb{R}^{m \times d}$ and a nonnegative matrix $H \in \mathbb{R}^{d \times n}$, and d is the number of latent factors such that the factorized matrix $D \approx WH$. For squared loss, NMF finds a low-rank approxi-

mation to a data matrix D by minimizing the Frobenius norm of $\|D - WH\|_{fro}$ under nonnegativity and scaling constraints on the factors W and H .

Capturing users' preferences is a main task in the prediction process. In recent years, the nonnegative constraint of NMF in latent factor space has been used to model a user's preferences [18], [31]. The model simplifies the rating by characterizing the features of both users and items to extract latent factors from a rating matrix. Specifically, given matrices $U \in \mathbb{R}^{i \times k}$ and $V \in \mathbb{R}^{k \times j}$, corresponding to the number of k latent factors, the predicted rating is then $\hat{R}_{ij} \approx UV^T$. The goal is to find entrywise nonnegative matrices U and V by minimizing the following objective function:

$$(U, V) = \|R - UV\|_F^2, \quad U \geq 0, V \geq 0 \quad (1)$$

where $\|\cdot\|_F$ is the Frobenius norm. Our model is formally defined as follows:

Definition 1 (User-Item Rating): The user-item interaction matrix $R \in \mathbb{R}^{i \times j}$ is factorized into the user factor matrix $U \in \mathbb{R}^{i \times k}$ and the item factor matrix $V \in \mathbb{R}^{k \times j}$ (k is the dimension of the latent factor) at each time t . The predicted ratings $\hat{R}_{ij}^t \approx U^t V^t$ are estimated at the present time period t and the previous time period $t-1$ by minimizing the following objective function:

$$(U, V, t) = \|R^t - U^t V^t\|_F^2 \quad (2)$$

$$(U, V, t-1) = \|R^{t-1} - U^{t-1} V^{t-1}\|_F^2 \quad (3)$$

where $\|\cdot\|_F$ is the Frobenius norm.

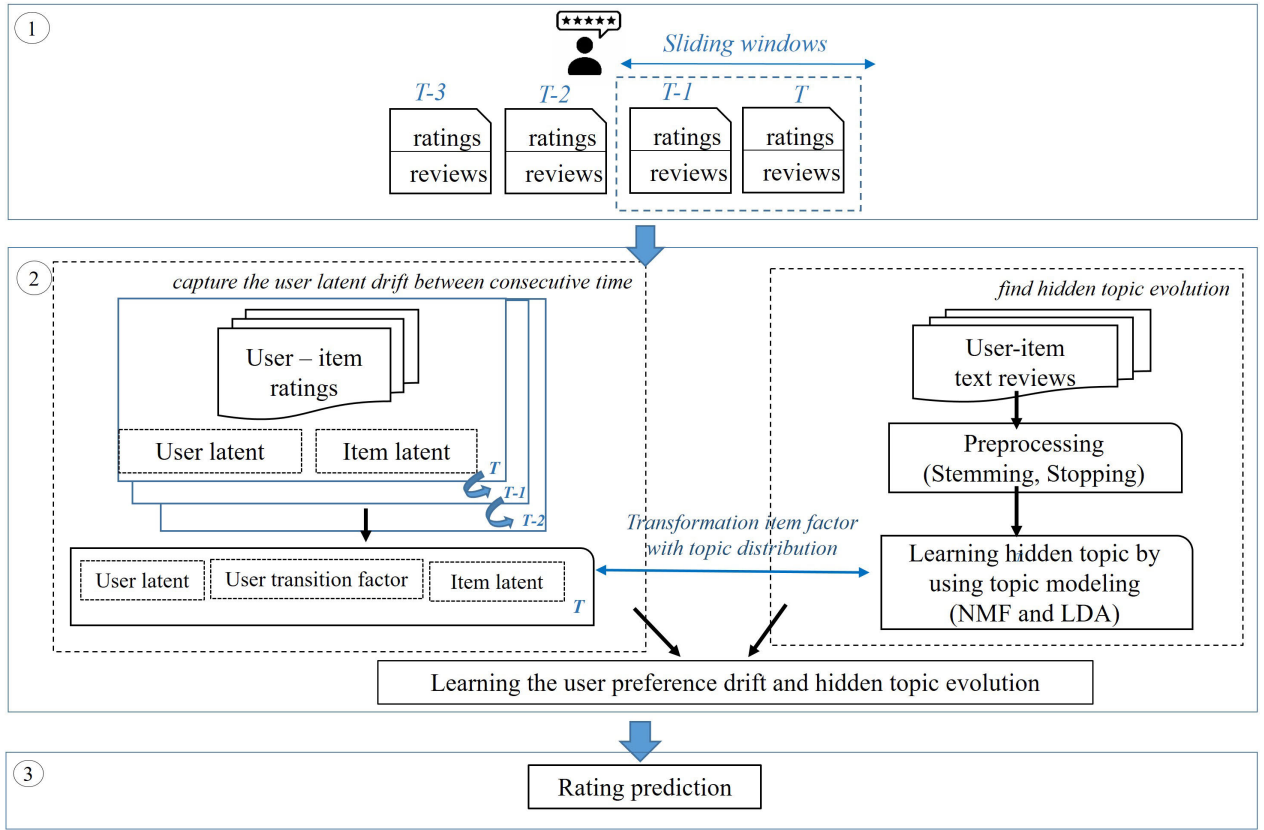
Definition 2 (Users' Preferences Over Time): The users' preferences change over time. The relationship of user factor matrix U_i^t at present time t and U_i^{t-1} at previous time $t-1$ in consecutive time are considered for each user i . For user i , the user factor matrix on the time series is $\mathcal{U} = \{U_i^t, U_i^{t-1}, U_i^{t-2}, \dots, U_i^{t-T}\}$ to learn the dynamics of users' preferences from time period $t-1$ to t .

Definition 3 (User-Item Reviews): The review comes with an overall rating R_{ij} to indicate the overall satisfaction of user i with item j . Given text reviews, \mathcal{D}_{ij} is a collection of reviews of item set \mathcal{V} written by a set of users \mathcal{U} . Review topics are the aspects of an item that a user discusses in a review. To learn the topic, the item factors $V_{j,k}$ are linked with topic distributions $\theta_{d,k}^{i,j}$ by using topic modeling and softmax transformation [36]

$$\theta_{d,k}^{i,j} = \frac{\exp(\kappa V_{j,k})}{\sum_{k'=1}^K \exp(\kappa V_{j,k'})} \quad (4)$$

where $\theta_{d,k}^{i,j}$ implies the k th dimension of the topic distribution of the review d written by user i for item j . Let κ be a parameter to control the 'peakness' of the transformation.

Definition 4 (Review Topic Evolution): As time passes, the topic of review collection evolves. The topic models find the topic distribution at different time steps to capture the topic evolution over the time span.



① Acquiring Data; ② Learning user preference each time-step based on rating and reviews; ③ Recommendation

FIGURE 1. The proposed method.

IV. PROPOSED METHOD

In this section, we describe the proposed method, shown in Figure 1. Given a set of historical ratings and reviews, the first task is to capture the latent user factor drift between consecutive time points to track the user transition factor. To consider both the user preference dynamics and topic evolution, we also learn the hidden topic evolution on reviews via topic modeling and transform the topic distributions of reviews and latent factors to improve the prediction performance.

A. CAPTURING THE USER FACTOR DRIFT

Motivated by the fact that there is a change in user latent vector U_i in each time step, we consider a model for the dynamics of the user latent factors in a time series. We assume that user interests evolve smoothly during one time period. Let us learn the relationship between U_i^t and U_i^{t-1} from the previous time $t-1$ to the present time t .

The conventional NMF assumes that the data and the latent factors are static. Clearly, this assumption is not appropriate for temporally drifting data. For capturing the user factor drift, we apply state transition matrix $Z_i^{(t)}$, to represent the mapping of the user latent factor between time steps with the linear dynamical systems (LDS) [22], [48], which is an effective way to reveal the relationship between user latent factors at each time step, where $U_i^{(t)} = Z_i^{(t)} U_i^{(t-1)}$. The above discussion motivates the following objective function that is

optimized every time step for the following minimization problems:

$$(U_i, Z_i, t, t-1) = \|U_i^t - Z_i^t U_i^{t-1}\|_F^2 \quad (5)$$

where the transition hypermatrix $Z_i \in \mathbb{R}^{D \times D}$ for user i models the transition of user i 's preferences in the latent space from the previous time to the next.

Accordingly, for each time step t , given $R^{(t)}$ and $U^{(t-1)}$, joining the above two decompositions, we derive:

$$\begin{cases} R^{(t)} \approx U^{(t)} V^{(t)} \\ R^{(t)} \approx Z^{(t)} U^{(t-1)} V^{(t)} \end{cases} \quad (6)$$

We propose a model to extend NMF for tracking the drift of the user latent factor, as in (2) and the state transition concept, as in (5), as a unified model over the user factor matrix and item factor matrix to avoid overfitting and to capture the rating and the preference drift in the period, as shown by the following joint minimization problem:

$$\begin{aligned} \arg \min_{U, V, Z} & \frac{1}{2} \sum_{t=1}^T \|R^{(t)} - U^{(t)} V^{(t)}\|_F^2 \\ & + \sum_{t=2}^T \|U^{(t)} - Z^{(t)} U^{(t-1)}\|_F^2 \\ \text{s.t. } & U^{(t)} \geq 0, \quad V^{(t)} \geq 0, \quad Z^{(t)} \geq 0 \end{aligned} \quad (7)$$

TABLE 1. Notations.

Notation	Description
i, j	numbers of users and items
k	latent factors and hidden topic factor
$R^{(t)} \in \mathbb{R}^{i \times j}$	the rating matrix at present time t
$R^{(t-1)} \in \mathbb{R}^{i \times j}$	the rating matrix at previous time $t - 1$
$\hat{R}_{ij}^{(t)}$	the predicted rating of item j by user i at time t
T	set of time period
$U_i^{(t)} \in \mathbb{R}^{i \times k}$	the user latent matrices at time t
$U_i^{(t-1)} \in \mathbb{R}^{i \times k}$	the user latent matrices at time $t - 1$
$V_j^{(t)} \in \mathbb{R}^{k \times j}$	the item latent matrices at time t
$Z^{(t)} \in \mathbb{R}^{k \times k}$	the user factor transition matrix of $U^{(t)}, U^{(t-1)}$
λ_1	temporal regularization parameter
$\alpha_1, \alpha_2, \alpha_3$	l_1 -norm regularization parameter for user factor, item factor and transition factor
$D^{(t)} \in \mathbb{R}^{i \times j}$	the review matrices at present time t
$d_{ij}^{(t)}$	the reviews of item j by user i at time t
n	the number of words in reviews D
$\theta_{jk}^{(t)}$	topic distribution for item j on topic k at time t
$\beta_{kn}^{(t)}$	word n 's distribution for topic k at time t

where $\|\cdot\|_F^2$ denotes the Frobenius norm. The first term in the cost function represents the minimization of the errors between the observed data and the recovered data. The second term shows the minimization of the errors of estimating the transition matrix over $t - 1$ transitions for capturing the user factor drift.

For a clearer understandings, we explain user factor drift based on temporal dependence between the latent user vectors U_i^t and U_i^{t-1} from the previous time $t - 1$ to the present time t . We proposed the transition hypermatrix $Z_i \in \mathbb{R}^{K \times K}$ for user i models the transition of user i 's preferences in the latent space from the previous time to the next. Z_i captures the time-invariant aspect of the user i 's evolution in this intrinsic feature space. For example, consider the $K = 2$ latent space. The identity transition matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ represents a stable latent user vector that does not change much over time; the transition matrix $\begin{bmatrix} 1.1 & 0 \\ 0 & 1 \end{bmatrix}$ represents a gradual shift pattern toward the first factor; the transition matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ represents the alternating changing pattern between two latent vectors. In a higher-dimensional latent space, several patterns could occur at the same time. At the current time t , we learn the latent user vector U_i^t and latent item vector V_j^t for each user i and each item j , as well as the transition matrix $Z_i \in \mathbb{R}^{K \times K}$ from all the ratings collected up to time t .

B. LEARNING THE TOPIC EVOLUTION OF REVIEWS

The reviews can address the rating sparsity problem because the discovered review topics can be used to enhance the ratings in CF. Given a corpus D , which contains reviews of users toward items $\{d_{i,j} | d_{i,j} \in \mathcal{D}, i \in \mathcal{U}, j \in \mathcal{V}\}$, we assume

that a set of latent topics (i.e., K topics) covers all the topics that users discuss in the reviews. We consider the review topic evolution in each time period $\mathcal{T} = \{t, t-1, t-2, \dots, t-n\}$.

1) LEARNING BY LATENT DIRICHLET ALLOCATION

LDA is a generative probabilistic model for a given text collection. The basic idea of LDA-based topic models is that documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words [9]. Topic models are applied in many models to incorporate explicit/implicit data [8], [20]. Essentially, there is an assumption that K latent topics are hidden in the given N document corpus, and each topic is represented as a multinomial distribution over the M words in the vocabulary extracted from the above data collection. Each document is generated by sampling a mixture of these topics and then sampling words from that mixture. More precisely, the generative process for each document in the text archive is illustrated as follows:

1. For the n th ($n = 1, 2, \dots, N$) document d in the whole N -document corpus, choose $\theta_n \sim \text{Dirichlet}(\alpha)$;
2. For each word $w_{n,m}$ in document d ,
 - (a) Choose topic assignment $z_{n,m} \sim \text{Multinomial}(\theta_n)$
 - (b) Find the corresponding topic distribution $\varphi_{z_{n,m}} \sim \text{Dirichlet}(\beta)$;
 - (c) Sample a word $w_{n,m} \sim \text{Multinomial}(\varphi_{z_{n,m}})$.

Finally, the hidden unobserved random variables, topic-word distributions φ and document-topic distributions θ could be learned through Gibbs sampling and variational expectation-maximization (EM) algorithm via maximizing the probability $P(D|\alpha, \beta)$.

Two parameters need to be inferred in this model: one is the distribution θ of document-topic, and the other is the distribution φ of topic-word.

2) LEARNING BY NONNEGATIVE MATRIX FACTORIZATION

NMF is a linear algebra method that embeds the original high-dimensional data into a low semantic space from nonnegative hidden structures, which are viewed as coordinate axes in the transformed space with geometric perspectives [28]. The goodness of this formalization is evaluated by the square loss between the original term-document vector d_{ij} and the linear combinations $\sum_{k=1}^K \theta_k \beta_{kn}$. Therefore, the following exists:

$$\mathcal{J}_{review} = \arg \min_{\theta, \beta} \sum_{n=1}^N \left\| d_n - \sum_{k=1}^K \theta_k \beta_{kn} \right\|_F^2 + \lambda \sum_{n=1}^N \|v_n\|_1 \quad (8)$$

$$\text{s.t.} \begin{cases} \|\beta^T \beta - I_K\|_F^2 \leq v, & (v \geq 0) \\ \theta_{mk} \geq 0, & m = 1, 2, \dots, M \\ \beta_{kn} \geq 0, & n = 1, 2, \dots, N; \\ & k = 1, 2, \dots, K \end{cases} \quad (9)$$

where each d_n of N reviews in a given corpus is supposed to be linearly combined by K factors, $\|\cdot\|_F^2$ marks the square of a matrix's Frobenius norm, v and λ are two hyperparameters

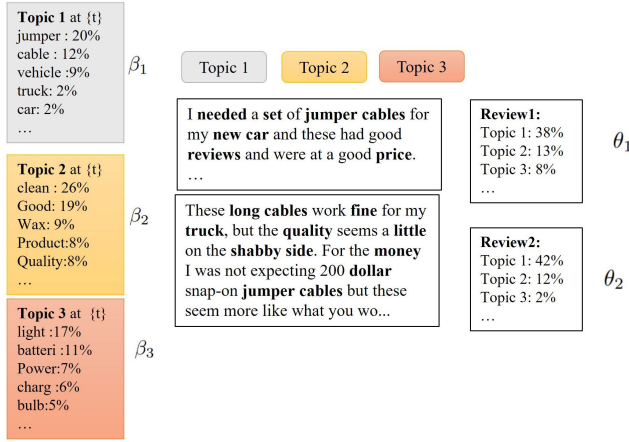


FIGURE 2. Illustration of topic model on Automotive reviews output.

with nonnegative values, and I_K is a $K \times K$ diagonal matrix. The smaller ν is, the closer it is to orthogonal θ , while the larger λ is, the sparser ν_n will be.

Topic evolution indicates that the same topic shows dynamism and differences over time. The evolution of the topic is reflected in two aspects. First, the topic intensity changes over time. We use the distribution of topic k with time t to define the intensity of the topic. We analyze the topics from Mar. 2001 to Aug. 2012. The different changes in the intensities of the topics will be calculated at different times, which are divided into seven time slices each year. The intensity of the topic, is shown by the following:

$$\delta'_k = \frac{1}{D'} \sum_{d=D'} \theta_{d,k} \quad (10)$$

where δ'_k is the intensity of a topic, D' is the number of documents, and $\theta_{d,k}$ is the distribution of topic k .

Through this model, we can obtain the topic distributions of users and reviews in time slice t . We apply the items' topic factors fully influence topics of their related reviews. Exponential transformation is adopted to guarantee that topic probability distribution of each review is associated with item topic factor as shown in Equation 4. In addition, the output of topic model is the topics distribution as K -dimensional vector for each review. The proposed model utilizes integrated factor formed by hidden topic and latent factors to model users' rating behaviors. While topic factor is jointly learned from both review texts and rating information, latent factor is mainly from rating information. After training model, each word in this model will be displayed as K dimensional hidden topic vector with element as probability for appearance of this word in corresponding topic. We sum up the values of all words in review for each dimension to construct representation of each review.

The example of output and our feature vector is illustrated in Figure 2 on Automotive dataset. The figure show top 5 words associated with 3 random topics. Note that LDA-based is a probabilistic mixture of mixtures model

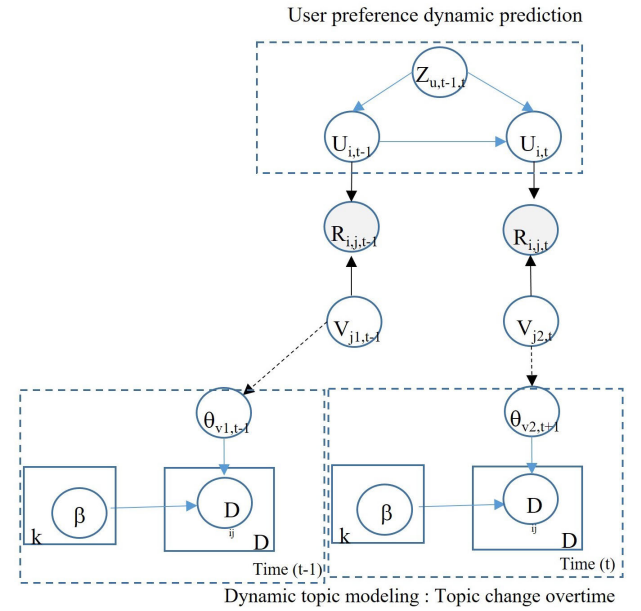


FIGURE 3. The graphical illustration of the proposed model based on NMF.

for grouped data. The observed words within the reviews are the result of probabilistically choosing words from a specific topic, where the topic is itself drawn from a document-specific multinomial that has a global Dirichlet prior. This means that words can belong to various topics in various degrees. In topic models, each reviews is a probability distribution of a series of topics, where each topic is a conditional probability of words. The topics in text reviews evolve over time and it is of interest to model the dynamics of underlying topics.

C. JOINT USER FACTOR TRANSITION AND TOPIC EVOLUTION

Our model attempts to combine the idea of user factor transition for rating prediction and topic modeling to uncover latent topic factors in review texts. In particular, we correlate the topic factors with the corresponding latent factors of both users and items. Specifically, as demonstrated in the previous section, we learn a topic distribution $\theta_{d,j}$ for each review d_{ij} . This actually records the extent to which each of K topics is discussed by user i for item j .

$$\begin{aligned} \mathcal{J}_{(all)} = \arg \min_{U,V,Z} & \frac{1}{2} \sum_{t=1}^T \|R^{(t)} - U^{(t)}V^{(t)}\|_F^2 \\ & + \sum_{t=2}^T \|U^{(t)} - Z^{(t)}U^{(t-1)}\|_F^2 \\ & + \sum_{n=1}^N \left\| d_n^{(t)} - \sum_{k=1}^K \theta_k^{(t)} \beta_{kn}^{(t)} \right\|_2^2 \\ & + \lambda_2 \sum_{n=1}^N \|v_n\|_1 + \xi \|U^T U - I_K\|_F^2 \end{aligned}$$

$$\begin{aligned}
& + \alpha_1 \|U^{(t)}\|_1 + \alpha_2 \|V^{(t)}\|_1 + \alpha_3 \|Z^{(t)}\|_1 \\
& + \lambda_1 \|Z^{(t)} - I\|_F^2 \\
& \text{s.t. } \beta^{(t)} \geq 0, V^{(t)} \geq 0, Z^{(t)} \geq 0
\end{aligned} \quad (11)$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm. The first term in the cost function represents the minimization of the errors between the observed data and the recovered data using the NMF t times. The second term shows the minimization of the errors of estimating the transition matrix over $t - 1$ transitions to capture the user factor drift, and $\|\cdot\|_1$ stands for the L1 norm. The temporal regularization $\lambda \|Z^{(t)} - I\|_F^2$ controls the extent to which we bias the decomposition toward $U^{(t-1)}$. Thus, the λ parameter $\in (0, \infty)$ balances present and past information; it quantifies the extent to which the model is past (i.e., $\lambda \rightarrow \infty$) or present oriented (i.e., $\lambda \rightarrow 0$).

The objective of our model is to learn the optimal U and V for accurately modeling users' ratings while simultaneously obtaining the most likely topics according to reviews with the constraint of the transformation function. Thus, we reach the following objective function for our model using NMF for uncovering hidden topics. However, we can find a local minimum for the objective function using multiplicative updates as introduced by [14]:

Considering the Karush-Kuhn-Tucker (KKT) [26] first-order conditions applied to our problem, we derive the following:

$$U^{(t)} \geq 0, V^{(t)} \geq 0, Z^{(t)} \geq 0, \theta^{(t)} \geq 0, \beta^{(t)} \geq 0 \quad (12)$$

$$\begin{aligned}
\nabla_{U^{(t)}} \mathcal{L} & \geq 0, \quad \nabla_{V^{(t)}} \mathcal{L} \geq 0, \quad \nabla_{Z^{(t)}} \mathcal{L} \geq 0, \\
\nabla_{\theta^{(t)}} \mathcal{L} & \geq 0, \quad \nabla_{\beta^{(t)}} \mathcal{L} \geq 0
\end{aligned} \quad (13)$$

$$\begin{aligned}
U^{(t)} \odot \nabla_{U^{(t)}} \mathcal{L} & = 0, \quad V^{(t)} \odot \nabla_{V^{(t)}} \mathcal{L} = 0 \\
Z^{(t)} \odot \nabla_{Z^{(t)}} \mathcal{L} & = 0 \\
\theta^{(t)} \odot \nabla_{\theta^{(t)}} \mathcal{L} & = 0, \quad \beta^{(t)} \odot \nabla_{\beta^{(t)}} \mathcal{L} = 0
\end{aligned} \quad (14)$$

where \odot is the elementwise product.

From the loss function \mathcal{L} in Equation (11), we derive the gradients according to each parameter:

$$\nabla_{V^{(t)}} \mathcal{L} = U^{(t)T} U^{(t)} V^{(t)} - (U^{(t)T} R^{(t)} - \alpha_2) \quad (15)$$

$$\nabla_{Z^{(t)}} \mathcal{L} = \sum_{t=2}^T (Z U^{(t-1)} U^{(t-1)T} - U^{(t)} U^{(t-1)T}) - \alpha_3 \quad (16)$$

$$\begin{aligned}
\nabla_{U^{(t)}} \mathcal{L} & = V^{(t)T} (V^{(t)} U^{(t)} - R^{(t)}) \\
& + \lambda Z^{(t)T} (Z U^{(t)} - U^{(t-1)})
\end{aligned} \quad (17)$$

By substituting the corresponding gradients into Equation (14), we derive the following update equations:

For example, substituting Eq. (15) into Eq. (14), we have

$$V \odot \sum_{t=1}^T (V^{(t)} U^{(t)} U^{(t)T}) = V \odot \sum_{t=1}^T (R^{(t)} U^{(t)T}) \quad (18)$$

Then, the MU rule for V is derived, i.e.,

$$V \leftarrow V \odot \left(\sum_{t=1}^T (R^{(t)} U^{(t)T}) \right) \oslash \left(\sum_{t=1}^T (V^{(t)} U^{(t)} U^{(t)T}) \right) \quad (19)$$

$$\begin{aligned}
Z & \leftarrow Z \odot \left(\sum_{t=2}^T (U^{(t)} U^{(t-1)T}) \right) \\
& \oslash \left(\sum_{t=2}^T (Z U^{(t-1)} U^{(t-1)T}) \right)
\end{aligned} \quad (20)$$

$$\begin{aligned}
U^{(t)} & \leftarrow U^{(t)} \odot (V^T R^{(t)} + \lambda Z^T U^{(t-1)}) \\
& \oslash (V^T V U^{(t)} + \lambda Z^T Z U^{(t)})
\end{aligned} \quad (21)$$

Recall that our goal is to simultaneously optimize the parameters associated with ratings (U and V) and the parameters associated with topics (θ and β). The topic distribution θ is fit by V . As presented above, U and V are fit by MU in Equations (19) - (21), while β is updated through Equation (8). Therefore, we design a procedure that alternates between the following two steps following algorithm 1.

Algorithm 1 The Proposed Algorithm

Input : Matrices $R^{(t)}, R^{(t-1)}$ for rating interaction, TF-IDF

matrix $D^{(t)}$ for review collection,
regularization parameter $\{\lambda, \alpha_1, \alpha_2, \alpha_3, \epsilon\}$,
factor size k , iterMax, subIterMax

Output: Predicted data $\hat{R}^{(t)}$

```

1 Randomly nonnegative initialize  $U^{(t)} \in \mathbb{R}^{i \times k}$ ,
 $V^{(t)} \in \mathbb{R}^{k \times j}$ ,  $Z^{(t)} \in \mathbb{R}^{k \times k}$ ;
2 while ( $iter < ltermax$ ) or ( $\mathcal{J}_{(all)}$  not converged) do
3   Update  $V^{(t)}$  according to iterative Eq. (19);
4   Update  $Z^{(t)}$  according to iterative Eq. (20);
5   Update  $U^{(t)}$  according to iterative Eq. (21);
6   if  $iter = subIterMax$  then
7     Compute  $topicdistribution \theta_{i,k}^{(t)}$  by transformation
      Eq. (4);
8     while ( $\mathcal{J}_{review}^{(t)}$  not converged) do
9       Update  $\beta_{k,n}^{(t)}$  for NMF based on Eq. (8);
10      Or
11      Update  $\beta_{k,n}^{(t)}$  for LDA based on Eq. (8);
12    end
13  end
14  Compute  $\mathcal{J}_{(all)}^{(t)}$  based on Eq. (11);
15   $iter = iter + 1$ ;
16 end
17 Predicted data  $\hat{R}^{(t)} = U^{(t)} V^{(t)}$ ;

```

Accordingly, the generative process for finding $\mathcal{J}_{review}^{(t)}$ from reviews in the corpus D by using NMF-based and LDA-based methods are different. The generative process of the LDA-based method for review can be simply described

TABLE 2. Rating prediction results on Automotive dataset.

user id	item id	actual rating	predicted rating (TimeSVD++)	predicted rating (Proposed model)
550	1206	3	2.74	2.97
	1256	4	3.65	3.92
	1369	3	2.58	2.95
	1456	3	1.26	2.89
	1567	2	2.20	1.99
	1777	5	3.17	4.51
	1852	4	2.63	3.52
	1859	4	2.75	3.58
	1987	3	2.46	2.79
	2805	1	2.62	1.08

by the following conditional distribution, $P(D|U, V) = \prod_{n \in N_d^{u,v}} \sum_k \theta_{d,k}^{u,v} \beta_{kn}$. We multiply overall reviews in the dataset and all words in each review. The two terms in the product are the likelihood of these particular topics ($\theta_{d,k}^{u,v}$) and the likelihood of particular words for this topic (β_{kn}). The sub-process iterates through all reviews d and all word positions n and updates their topic assignments. We assign each word to a topic (an integer between 1 and k) randomly, with probability proportional to the likelihood of that topic occurring with that word and set k as topic factor with probability proportional $\theta_{d,k}^{u,v} \beta_{kn}$. Alternatively, the NMF-based process reconstructs words' occurrence counts in each review. The reconstruction loss can be formulated by the squared loss between the original term-document vector d_n and the linear combinations $\sum_{k=1}^K \theta_{d,k}^{u,v} \beta_{kn}$. However, both the proposed NMF-based and LDA-based require the topic proportions ($\theta_{d,k}$) that are not sampled from a Dirichlet distribution but are instead determined based on the value of $V_{j,k}$ taken during the previous step.

For example, we examine the rating prediction results following by $\hat{R}^{(t)} = U^{(t)}V^{(t)}$ for each user in Table 2 with a user (user id = 550). In summary, our temporal model is good at capturing the preference dynamics and thus yields improvement on the performance.

V. EXPERIMENTS

In this section, the recommender system of the proposed model is evaluated. The experimental data preparation and results are presented and discussed.

A. DATASETS

We use seven datasets from Amazon¹ which are collected by Stanford University [36]. Statistics of the datasets are shown in Table 3. These datasets are a collection of user ratings with a corpus of text reviews and timestamps corresponding to various types of items that are available on public websites. The numerical rating scores of all reviews lie in the range of 1 to 5 and can only be integers. For example, the baby dataset consists of 122,150 ratings and 98,630 reviews from 19,526 users on 7,060 items. We are interested in all the interaction difference datasets from 1998 to 2014, even for users or items

¹Datasets: <http://jmcauley.ucsd.edu/data/amazon>

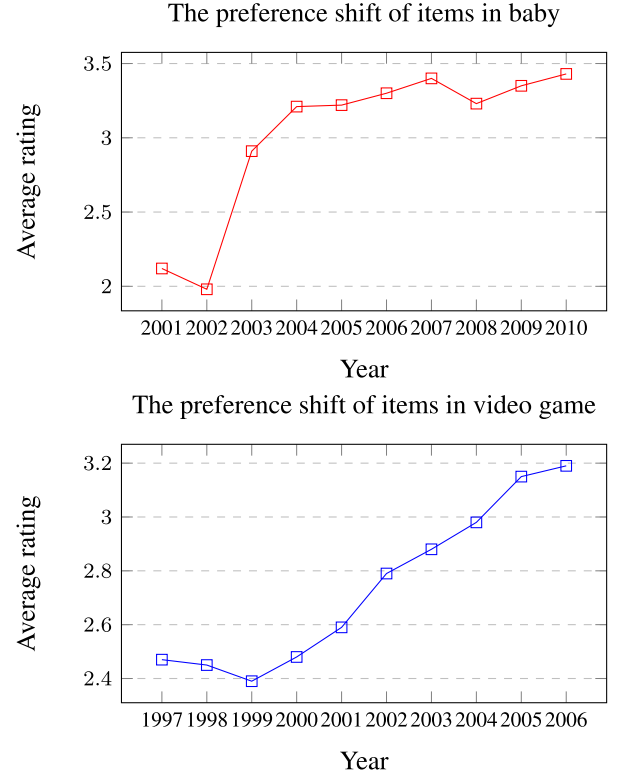


FIGURE 4. The preference dynamics of ratings in the (a) baby and (b) video game datasets.

that have only one review. Figure 4 shows the ratings drift based on the average ratings per sliding time window over time. An abrupt shift in ratings occurred for the Baby dataset in 2003, with a sudden jump from approximately 1.98 to greater than 2.91. The second significant temporal effect is that the ratings in the Video game dataset tend to increase over time.

Additionally, we discuss how the training sets and test sets are constructed and how to measure the performances of the model before reporting the results. First, we sort the data chronologically and construct training and testing subsets by shifting a time period. Then, we split the dataset into \mathcal{T} time periods corresponding to 10 months. Given a set of training months, we predict the preference for a user at a test month. We declare a time period equal to 10 months, and we consider all the past months (1^{st} - 9^{th}) of each time period as the training set and the last month (10^{th}) as the test set. For example, Automotive is selected for 90 months, which split the dataset into $\mathcal{T} = 9$ time periods, corresponding to 10 months. The 1^{st} - 9^{th} month of each period is used as the training set and the 10^{th} as testing set. Therefore, we have nine different test sets of the 10^{th} , 20^{th} , 30^{th} , 40^{th} , 50^{th} , 60^{th} , 70^{th} , 80^{th} and 90^{th} test months. For each test month, we average the results overall users and report the mean values. The size of the training and testing set are shown in Table 4.

B. REVIEW DATA PREPROCESSING

In preprocessing the review data, we attempt to clean our review data as much as possible using the natural language

TABLE 3. Description of datasets.

Datasets	users	items	ratings	reviews	sparsity	timestamps
Automotive	1,836	2,930	38,595	20,474	99.282%	Jul.2005-May.2014
Baby	19,526	7,060	122,150	98,630	99.911%	Mar.2001-Aug.2012
Cell phone& Accessories	100,532	20,469	436,520	205,250	99.978%	Dec.2000-Aug.2012
Fine Food	243,059	65,858	1,256,965	535,721	99.992%	Dec.1999-Aug.2012
Video game	421,256	35,880	2,535,441	656,112	99.983%	Nov.1997-Aug.2012
Toys & Game	875,021	225,896	1,989,554	1,325,256	99.998%	Nov.2000-Aug.2012
Beauty	998,195	212,485	208,936	1,572,549	99.999%	Jan.1998-Aug.2012

TABLE 4. Size of training and testing set.

Datasets	#datasets (months)	#training (months)	#testing (months)
Automotive	90	81	9
Baby	140	126	14
Cell phone& Accessories	140	126	14
Fine Food	150	135	15
Video game	180	162	18
Toys & Game	140	126	14
Beauty	180	162	18

toolkit (NLTK)² in Python. The idea is to remove the punctuation, numbers and special characters in one step using regex replace, which will replace everything, except letters, with a space. Then, we will remove shorter words because they generally do not contain useful information. Finally, we will make all the text lowercase to nullify case sensitivity. After that, we remove the stop words from the review data because they are mostly cluttered and hardly carry any information. Stop words are terms such as ‘it’, ‘they’, ‘am’, ‘a’, ‘by’, ‘doing’, ‘it’, and ‘how’.

To remove stop words from the documents, we have to tokenize the text, i.e., split the string of text into individual tokens or words.

C. EVALUATION METHODS

To evaluate the performance of the RS based on user preference drift and topic evolution, we employed the common evaluation metric, as follows:

RMSE : the root mean square error is widely used for evaluating the performance of RS models. In this paper, we present the prediction accuracy of the recommender in terms of rating predictions, which is given by the following:

$$RMSE = \sqrt{\frac{\sum_{i,j} (R_{i,j} - \hat{R}_{i,j})^2}{N}} \quad (22)$$

where N ratings of a test month, the ground-truth ratings in R and the predicted ratings in \hat{R} , where lower values of $RMSE$ indicate higher performance in the rating prediction.

²NLTK: <https://www.nltk.org/>

D. COMPARED METHODS

We conduct experiments on each dataset to assess the performance of the proposed method in comparison with baseline recommendation models: TimeSVD++, Bayesian temporal matrix factorization (BTMF), hidden factor topic (HFT), and temporal collective matrix factorization (TCMF).

1) TimeSVD++

This approach is a baseline temporal recommender model that extends SVD++ by introducing a time-variant bias for each user and item in every individual time step to address the dynamics of user preferences, but it does not exploit auxiliary information to enhance the accuracy due to data sparsity [25].

2) BAYESIAN PROBABILISTIC TENSOR FACTORIZATION

The ratings proposed by the BPTF model are represented as triples (user, item, time). These triples are organized into a three-dimensional tensor. Finally, the tensor is decomposed, and ratings are predicted using the inner product of the latent factor vector [57].

3) HIDDEN FACTOR AS TOPICS

In HFT, topics in review texts are associated with item parameters using LDA topic modeling, which has been shown to perform well in terms of rating predictive accuracy, to overcome data sparsity, but temporal data are not considered [36].

4) BAYESIAN TEMPORAL MATRIX FACTORIZATION

BTMF is a temporal recommender model that extends SVD++ by introducing a time-variant bias for each user and item in every individual time step. It achieves the best performance among the matrix factorization schemes by introducing priors for the hyperparameters of TMF to capture the conditional distributions of users, items and user feedback on items [61].

5) TEMPORAL COLLECTIVE MATRIX FACTORIZATION

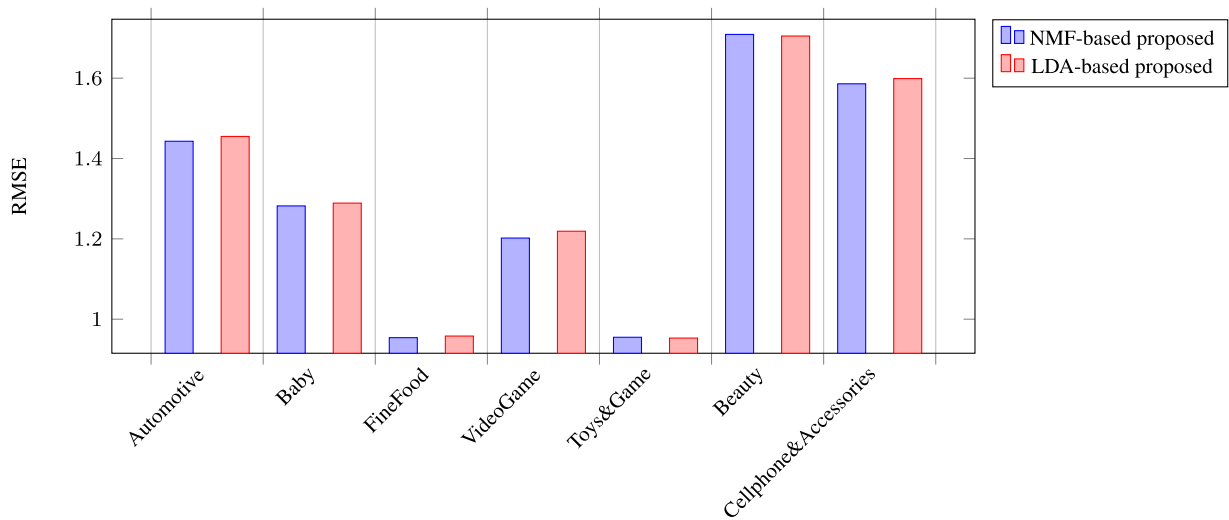
The TCMF model captures preference dynamics through a joint decomposition model that extracts the user temporal patterns from the rating and multimodal information between consecutive time points. However, the model does not track the hidden topic factors from review information to discover the evolution of users’ preference topics [42].

6) TMRevCo

The model is based on matrix factorization model that factorizes rating matrix into latent user and item factors for

TABLE 5. Comparison of RMSE on different models with latent topic $K = 10$.

Datasets	Time SVD++ (a)	BPTF (b)	HFT (c)	BTMF (d)	TCMF (e)	TM RevCo (f)	Proposed (NMF) (g)	Proposed (LDA) (h)	improvement (%)					
									(a&g)	(b&g)	(c&g)	(d&g)	(e&g)	(f&g)
Automotive	1.546	1.564	1.539	1.499	1.499	1.492	1.443	1.455	7.15	8.37	6.67	3.92	3.89	2.54
Baby	1.356	1.338	1.330	1.315	1.312	1.311	1.282	1.289	5.73	4.39	3.71	2.54	2.36	1.72
Fine Food	1.130	1.128	1.113	1.008	0.985	1.007	0.954	0.958	18.41	18.22	16.65	5.64	3.19	5.15
Video game	1.379	1.322	1.290	1.286	1.224	1.275	1.202	1.219	14.74	10.01	7.39	7.03	1.86	4.66
Toys&Game	1.043	1.030	0.999	0.989	0.985	0.987	0.955	0.953	9.26	7.86	4.65	3.56	3.14	3.56
Beauty	1.910	1.822	1.815	1.790	1.779	1.781	1.709	1.705	11.76	6.60	6.20	4.74	4.10	4.45
Cell phone& Accessories	1.641	1.635	1.637	1.679	1.623	1.652	1.586	1.599	3.48	3.14	3.21	5.89	2.35	3.37

**FIGURE 5.** Comparison of NMF-based proposed method and LDA-based proposed method in terms of RMSE with topic factor = 10.

rating prediction. The model focus on dynamic user factor and item correlation measure in CoFactor and associate the item factors of CoFactor with that of matrix factorization [55].

All these approaches are tested in the same experimental environment. In the compared approaches, regularization parameters, learning rate, size of latent factors are optimized experimentally. The most precise values are either determined by empirical results or suggested by the original papers. For temporal recommendation models including TimeSVD++, we train the model by fitting the data in the training set and learn model parameters by the stochastic gradient descent (SGD) algorithm as suggested in [25], which ignores the review texts when generating the prediction in the test month and the factor dimension(f) set to 10. The settings of BPTF given by [57] are adopted in our experiments; consequently, we set $\mu_0 = 0$, $v_0 = K$, $\beta_0 = 1$, $\mathbf{W}_0 = \mathbf{I}$, $\tilde{v}_0 = 1$, $\tilde{\Lambda} = 1$, and $\tilde{W}_0 = 1$, where L is 200, K is set from 5 to 50, with additional $Z_0 = \mathbf{I}$ in BTMF as suggested in [61] and we chose the best one, which has the latent space dimension of 10 in the hyper-priors. Meanwhile, the compared models are considered regarding coping with data sparsity, including HFT and TCMC. For HFT, the offset (B) and bias terms (B_i) are initialized by average ratings and

global bias item, parameters are fit using L-BFGS and we set $n = 0.1$ as suggested in [36]. TCMC is set with regularization parameters: $\lambda = \beta = \gamma = 1$, as suggested in [42], and we fix $\epsilon = 10^{-4}$, $\text{maxIter} = 10^3$ and obtain the best results on the temporal regularizer parameter $\lambda = 1$. While, TMRevCo, we set $\lambda_c = 0.001$ and $\lambda_r = 0.05$ as suggested by [55].

E. EXPERIMENTAL RESULTS

The results are evaluated in terms of RMSE and are shown in Table 5. We compare the proposed method with the temporal models TimeSVD++,³ BPMF,⁴ and BPTF.⁵ Additionally, we also compare our method with the HFT model, which incorporates the rating and hidden topics factor to enhance ratings and address data sparsity but ignores the preference dynamics. While, TMRevco model combines the dynamic user factor of TimeSVD++ with the hidden topic of each review texts mined by the topic model and applies item correlation measure in CoFactor and associate the item factors of CoFactor. Meanwhile, the TCMF model considers the

³TimeSVD++: <http://www.librec.net>

⁴BPMF: <https://github.com/xinychen/transdim>

⁵BPTF: <http://www.cs.cmu.edu/~lxiong/bptf/bptf.html>

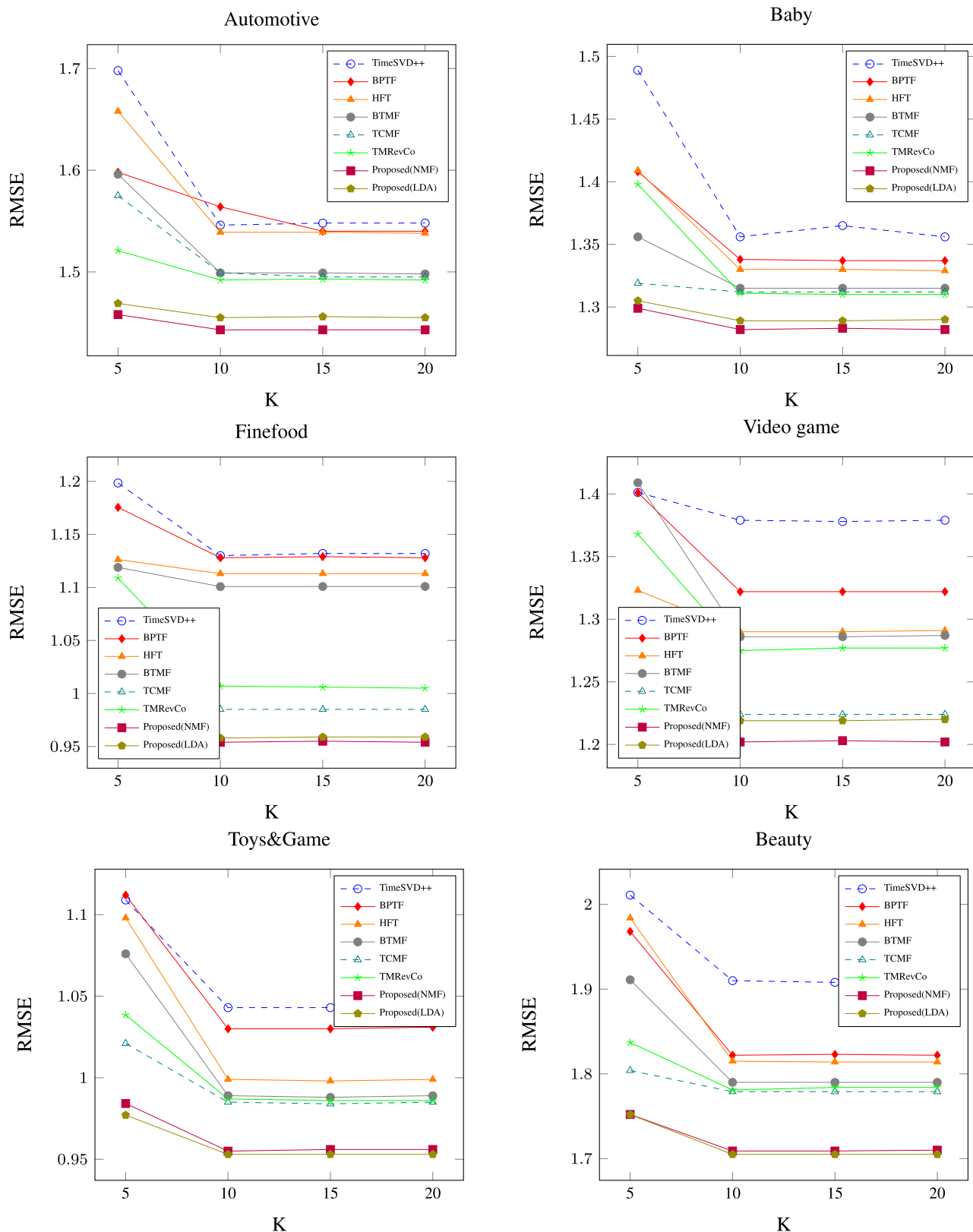


FIGURE 6. Comparison of the proposed method with baselines in terms of RMSE on public datasets.

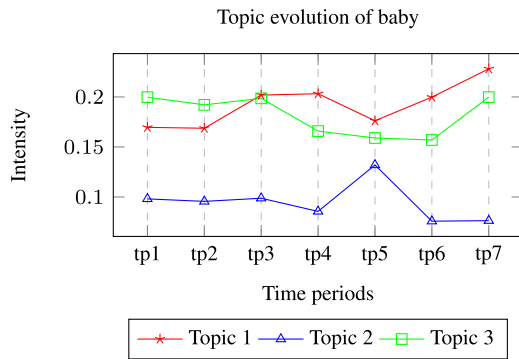


FIGURE 7. Intensity topic evolution of the top3 topics on baby.

user preference drift and solves the data sparsity problem based on the accuracy of the recommendation, as it does not track the hidden topic evolution of reviews that may dynamically change over time. The proposed model outperforms the compared models, and the best performance is indicated in bold font. This occurs because the proposed model observes the dynamics of user interactions on ratings and reviews to capture user preference drifts and address the data sparsity problem with hidden topic review evolution. Figure 6 shows a comparison of the proposed method with state-of-the-art approaches in terms of RMSE. The results also show the effect of performance accuracy for the examined models over the number of latent factors(k). We observe that all models improve their learning ability and the best average result is on $k = 10$.

From the results of our experiment, we can see the changes of topic interests. For example, the intensity of each topic is shown in Figure 7. The ordinate represents the intensity of the top 3 topics in the Baby dataset in each time slice, which is calculated with equation (10). Figure 7 shows us that the intensity of topic 1 is higher than that of the other topics during the time slice of 4-7. This result demonstrates that the influence of topic 1 is strongest. In addition, the intensity of each topic changes with time. This result shows that each topic will go through an evolutionary process that is consistent with reality. Next, we also track topics from reviews automatically using dynamic LDA-based and NMF-based methods in each time period. These methods take different approaches to model texts, i.e., corresponding to probabilistic generative processes and geometrically linear combinations, respectively. Therefore, in Figure 5 we compare the result to determine which is better for different topic numbers (5, 10, 15, 20, 50) for both methods corresponding to the datasets, and the hyper parameters α and β in LDA are set to 1 and 0.01, respectively. The iterative parameters of NMF are set to 100 iterations.

Obviously, the proposed NMF-based method tends to outperform the LDA-based method for short review topic mining and is positively affected in the case of dynamics, which leads to draw some analysis. However, LDA and NMF can both automatically learn topics from short texts, but they do so in different manners to model texts, i.e., corresponding to

a probabilistic generative process and geometrically linear combinations, respectively. Meanwhile, the general review texts are short and sparse, and due to this manner, ordinarily exhibit an absence of data regarding the word-word co-occurrences, which is undoubtedly not gainful for statistical topic models such as LDA. Furthermore, our model adopted in the LDA-based method is the stochastic Gibbs sampling for each word in texts, which would introduce significant variances in learning and inference, especially for sparse and short texts. However, the NMF-based approach first encoded the whole corpus with term frequency-inverse document frequency (TF-IDF) weights, which not only considering the frequency but also induces discriminative information (IDF) for each word. In conclusion, the greater amount of information encoded and definite algorithms likely enable the tendency of NMF to produce higher-quality topics than does LDA from short texts [14] and leading to improve accuracy.

VI. DISCUSSION

Summarizing our results, our model can be used as a novel form of temporal dynamics for tracking the evolution of users' preferences over time to capture the transition of users' preferences in latent factor space and exploit users' opinions as review texts to find hidden topic evolutions that can explain why a user rates the score by using a topic modeling technique at different time steps. However, such temporal models e.g., time SVD++, BPTF, BTMF, focus only rating interactions. They ignore data sparsity problem. The sparse data leads to low prediction accuracy. On the other hand, HFT is a model that can exploit review topics and latent factors to enhance the rating and solve the rating sparsity issue, but does not capture preference drifts. While, TCMC combines multimodal information with scores to reduce data sparsity, but this model also considered multimodal data that did not uncover the hidden topic factor from reviews to explain the score on each time step. For TMRevCo, the model considers temporal dynamics and item correlation dimensions on reviews but does not capture the transition of users' preference each time step. Our results clearly show that the proposed model outperforms the compared models. This occurs because the proposed model observes the dynamics of user interactions on ratings and reviews to capture user preference drifts and address the data sparsity problem with hidden topic review evolution.

VII. CONCLUSIONS

This paper aims to enhance dynamic CF on recommender systems under volatile conditions in which both users' preferences and item properties change dynamically over time. Moreover, the existing CF models mainly rely on solving data sparsity by adding side information to improve the performance. We proposed a model to capture the dynamics of users' preferences in the rating matrix by using a joint decomposition method to extract user latent transition patterns and combine latent factors, together with the associated topic evolution of review texts by using dynamic topic modeling. We evaluated the accuracy on seven real datasets for solving

the data sparsity problem, and the experimental results show that the model leads to a significant improvement compared with the baseline and state-of-the-art method.

In the future, we plan to extend this work to incorporate external users' interests, such as social relationships and fashion trends that can change over time. Another interesting extension to our model would be the ability to capture evolving, emerging and fading topic interests of users to improve the prediction performance of the dynamic recommendation system.

REFERENCES

- [1] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.
- [2] B. A. Hammou, A. A. Lahcen, and S. Mouline, "An effective distributed predictive model with matrix factorization and random forest for big data recommendation systems," *Expert Syst. Appl.*, vol. 137, pp. 253–265, Dec. 2019.
- [3] M. Al-Ghossein, P.-A. Murena, T. Abdesslem, A. Barré, and A. Cornuéjols, "Adaptive collaborative topic modeling for online recommendation," in *Proc. 12th ACM Conf. Recommender Syst.* New York, NY, USA: Association for Computing Machinery, 2018, pp. 338–346, doi: [10.1145/3240323.3240363](https://doi.org/10.1145/3240323.3240363).
- [4] I. Al-Hadi, N. M. Sharef, M. N. Sulaiman, and N. Mustapha, "Review of the temporal recommendation system with matrix factorization," *Int. J. Innov. Comput. Inf. Control*, vol. 13, no. 5, pp. 1579–1594, 2017.
- [5] J. Allan, *Topic Detection and Tracking: Event-Based Information Organization*. New York, NY, USA: Springer, 2012.
- [6] M. Ayub, M. A. Ghazanfar, Z. Mehmood, T. Saba, R. Alharbey, A. M. Munshi, and M. A. Alrige, "Modeling user rating preference behavior to improve the performance of the collaborative filtering based recommender systems," *PLoS ONE*, vol. 14, no. 8, 2019, Art. no. e0220129, doi: [10.1371/journal.pone.0220129](https://doi.org/10.1371/journal.pone.0220129).
- [7] R. Ayub, M. A. Ghazanfar, Z. Mehmood, K. Alyoubi, and A. Fakeeh, "Unifying user similarity and social trust to generate powerful recommendations for smart cities using collaborating filtering-based recommender systems," *Soft Comput.*, vol. 23, Nov. 2019, doi: [10.1007/s00500-019-04588-x](https://doi.org/10.1007/s00500-019-04588-x).
- [8] S. Bansal, A. Srivastava, and A. Arora, "Topic modeling driven content based jobs recommendation engine for recruitment industry," *Procedia Comput. Sci.*, vol. 122, pp. 865–872, Jan. 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877050917326960>, doi: [10.1016/j.procs.2017.11.448](https://doi.org/10.1016/j.procs.2017.11.448).
- [9] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
- [10] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey," *Knowl.-Based Syst.*, vol. 46, pp. 109–132, Jul. 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950705113001044>
- [11] C. Cai and H. Xu, "A topic sentiment based method for friend recommendation in online social networks via matrix factorization," *J. Vis. Commun. Image Represent.*, vol. 65, Dec. 2019, Art. no. 102657. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1047320319302780>
- [12] R. Catherine and W. Cohen, "Transnets: Learning to transform for recommendation," in *Proc. 11th ACM Conf. Recommender Syst.* New York, NY, USA: Association Computing Machinery, 2017, pp. 288–296, doi: [10.1145/3109859.3109878](https://doi.org/10.1145/3109859.3109878).
- [13] L. Chen, G. Chen, and F. Wang, "Recommender systems based on user reviews: The state of the art," *User Model. User-Adapted Interact.*, vol. 25, no. 2, pp. 99–154, Jun. 2015, doi: [10.1007/s11257-015-9155-5](https://doi.org/10.1007/s11257-015-9155-5).
- [14] Y. Chen, H. Zhang, R. Liu, Z. Ye, and J. Lin, "Experimental explorations on short text topic mining between LDA and NMF based schemes," *Knowl.-Based Syst.*, vol. 163, pp. 1–13, Jan. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0950705118304076>
- [15] J. Cheng, Y. Liu, H. Zhang, X. Wu, and F. Chen, "A new recommendation algorithm based on user's dynamic information in complex social network," *Math. Problems Eng.*, vol. 2015, pp. 1–6, Mar. 2015.
- [16] Z. Cheng, Y. Ding, L. Zhu, and M. Kankanhalli, "Aspect-aware latent factor model: Rating prediction with ratings and reviews," in *Proc. World Wide Web Conf.*, 2018, pp. 639–648, doi: [10.1145/3178876.3186145](https://doi.org/10.1145/3178876.3186145).
- [17] J. Y. Chin, K. Zhao, S. Joty, and G. Cong, "ANR: Aspect-based neural recommender," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*. New York, NY, USA: Association Computing Machinery, 2018, pp. 147–156, doi: [10.1145/3269206.3271810](https://doi.org/10.1145/3269206.3271810).
- [18] F. C. T. Chua, R. J. Oentaryo, and E.-P. Lim, "Modeling temporal adoptions using dynamic matrix factorization," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 91–100.
- [19] A. Daud, "Using time topic modeling for semantics-based dynamic research interest finding," *Knowl.-Based Syst.*, vol. 26, pp. 154–163, Feb. 2012, doi: [10.1016/j.knsys.2011.07.015](https://doi.org/10.1016/j.knsys.2011.07.015).
- [20] E. Delpisheh, A. An, H. Davoudi, and E. Boroujerdi, "Time aware topic based recommender system," *Big Data Inf. Anal.*, vol. 1, p. 8, Nov. 2016.
- [21] Y. Ding and X. Li, "Time weight collaborative filtering," in *Proc. 14th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*. New York, NY, USA: ACM, 2005, pp. 485–492, doi: [10.1145/1099554.1099689](https://doi.org/10.1145/1099554.1099689).
- [22] Z. Ghahramani and G. E. Hinton, "Parameter estimation for linear dynamical systems," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep. CRG-TR-96-2, 1996.
- [23] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*. New York, NY, USA: ACM, 1999, pp. 50–57, doi: [10.1145/312624.312649](https://doi.org/10.1145/312624.312649).
- [24] W. Hong, L. Li, and T. Li, "Product recommendation with temporal dynamics," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12398–12406, Nov. 2012.
- [25] Y. Koren, "Collaborative filtering with temporal dynamics," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. New York, NY, USA: ACM, 2009, pp. 447–456, doi: [10.1145/1557019.1557072](https://doi.org/10.1145/1557019.1557072).
- [26] H. W. Kuhn and A. W. Tucker, "Nonlinear programming," in *Proc. 2nd Berkeley Symp. Math. Statist. Probab.* Berkeley, CA, USA: Univ. of California Press, 1951, pp. 481–492. [Online]. Available: <https://projecteuclid.org/euclid.bmsmp/1200500249>
- [27] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.
- [28] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization" in *Proc. 13th Int. Conf. Neural Inf. Process. Syst.* Cambridge, MA, USA: MIT Press, 2000, pp. 535–541. [Online]. Available: <http://dl.acm.org/citation.cfm?id=3008751.3008829>
- [29] K. Li, X. Zhou, F. Lin, W. Zeng, B. Wang, and G. Alterovitz, "Sparse online collaborative filtering with dynamic regularization," *Inf. Sci.*, vol. 505, pp. 535–548, Dec. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0020025519307157>
- [30] Z. Li, J. Liu, X. Zhu, and H. Lu, "Multi-modal multi-correlation person-centric news retrieval," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*. New York, NY, USA: ACM, 2010, pp. 179–188, doi: [10.1145/1871437.1871464](https://doi.org/10.1145/1871437.1871464).
- [31] Z. Li, H. Zhang, S. Wang, F. Huang, Z. Li, and J. Zhou, "Exploit latent Dirichlet allocation for collaborative filtering," *Frontiers Comput. Sci.*, vol. 12, no. 3, pp. 571–581, Jun. 2018.
- [32] G. Ling, M. R. Lyu, and I. King, "Ratings meet reviews, a combined approach to recommend," in *Proc. 8th ACM Conf. Recommender Syst. (RecSys)*. New York, NY, USA: ACM, 2014, pp. 105–112, doi: [10.1145/2645710.2645728](https://doi.org/10.1145/2645710.2645728).
- [33] Z. Lipton, Y.-X. Wang, and A. Smola, "Detecting and correcting for label shift with black box predictors," in *Proc. 35th Int. Conf. Mach. Learn.*, vol. 8, J. Dy and A. Krause, Eds. Stockholm, Sweden: Stockholmsmässan, Jul. 2018, pp. 3122–3130. [Online]. Available: <http://proceedings.mlr.press/v80/lipton18a.html>
- [34] Y. Liu, Y. Liu, Y. Shen, and K. Li, "Recommendation in a changing world: Exploiting temporal dynamics in ratings and reviews," *ACM Trans. Web*, vol. 12, no. 1, pp. 1–20, Feb. 2018, doi: [10.1145/3108238](https://doi.org/10.1145/3108238).
- [35] L. Luo, H. Xie, Y. Rao, and F. L. Wang, "Personalized recommendation by matrix co-factorization with tags and time information," *Expert Syst. Appl.*, vol. 119, pp. 311–321, Apr. 2019.
- [36] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: Understanding rating dimensions with review text," in *Proc. 7th ACM Conf. Recommender Syst. (RecSys)*. New York, NY, USA: ACM, 2013, pp. 165–172, doi: [10.1145/2507157.2507163](https://doi.org/10.1145/2507157.2507163).
- [37] S.-H. Min and I. Han, "Detection of the customer time-variant pattern for improving recommender systems," *Expert Syst. Appl.*, vol. 28, no. 2, pp. 189–199, Feb. 2005. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417404001101>
- [38] C.-C. Musat, Y. Liang, and B. Faltings, "Recommendation using textual opinions," in *Proc. 23rd Int. Joint Conf. Artif. Intell. (IJCAI)*. Palo Alto, CA, USA: AAAI Press, 2013, pp. 2684–2690. [Online]. Available: <http://dl.acm.org/citation.cfm?id=2540128.2540515>

- [39] S. Natarajan, S. Vairavasundaram, S. Natarajan, and A. H. Gandomi, "Resolving data sparsity and cold start problem in collaborative filtering recommender system using linked open data," *Expert Syst. Appl.*, vol. 149, Jul. 2020, Art. no. 113248. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417420300737>
- [40] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, "Deep learning vs. traditional computer vision," in *Proc. Sci. Inf. Conf. Las Vegas, NV, USA: Springer*, 2019, pp. 128–144.
- [41] E. Poirson and C. D. Cunha, "A recommender approach based on customer emotions," *Expert Syst. Appl.*, vol. 122, pp. 281–288, May 2019.
- [42] D. Rafailidis, P. Kefalas, and Y. Manolopoulos, "Preference dynamics with multimodal user-item interactions in social media recommendation," *Expert Syst. Appl.*, vol. 74, pp. 11–18, May 2017. [Online]. Available: <https://app.dimensions.ai/details/publication/pub.1023393133>
- [43] D. Rafailidis and A. Nanopoulos, "Modeling the dynamics of user preferences in coupled tensor factorization," in *Proc. 8th ACM Conf. Recommender Syst. (RecSys)*. New York, NY, USA: ACM, 2014, pp. 321–324, doi: [10.1145/2645710.2645758](https://doi.org/10.1145/2645710.2645758).
- [44] C. Rana and S. K. Jain, "A study of the dynamic features of recommender systems," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 141–153, Jan. 2015. [Online]. Available: <https://app.dimensions.ai/details/publication/pub.1007079586> and <https://www.springerlink.com/content/165422u737864003/fulltext.pdf>
- [45] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.*, vol. 2009, pp. 1–19, Jan. 2009.
- [46] G. Suganeshwari and S. P. Syed Ibrahim, "A survey on collaborative filtering based recommendation system," in *Proc. 3rd Int. Symp. Big Data Cloud Comput. Challenges*, V. Vijayakumar and V. Neelanarayanan, Eds. Cham, Switzerland: Springer, 2016, pp. 503–518.
- [47] B. Sun and L. Dong, "Dynamic model adaptive to user interest drift based on cluster and nearest neighbors," *IEEE Access*, vol. 5, pp. 1682–1691, 2017.
- [48] J. Sun, R. Kush, and K. Subbian, "Dynamic matrix factorization: A state space approach," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 1897–1900, doi: [10.1109/ICASSP.2012.6288274](https://doi.org/10.1109/ICASSP.2012.6288274).
- [49] C. K. Vaca, A. Mantrach, A. Jaimes, and M. Saerens, "A time-based collective factorization for topic discovery and monitoring in news," in *Proc. 23rd Int. Conf. World Wide Web*. New York, NY, USA: ACM, 2014, pp. 527–538, doi: [10.1145/2566486.2568041](https://doi.org/10.1145/2566486.2568041).
- [50] D. V́eras, R. Prudêncio, and C. Ferraz, "CD-CARS: Cross-domain context-aware recommender systems," *Expert Syst. Appl.*, vol. 135, pp. 388–409, Nov. 2019.
- [51] B. Wang, B. Chen, L. Ma, and G. Zhou, "User-personalized review rating prediction method based on review text content and user-item rating matrix," *Information*, vol. 10, no. 1, p. 1, 2018.
- [52] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*. New York, NY, USA: Association Computing Machinery, 2011, pp. 448–456, doi: [10.1145/2020408.2020480](https://doi.org/10.1145/2020408.2020480).
- [53] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative filtering and deep learning based recommendation system for cold start items," *Expert Syst. Appl.*, vol. 69, pp. 29–39, Mar. 2017.
- [54] H. Wu, Z. Zhang, K. Yue, B. Zhang, J. He, and L. Sun, "Dual-regularized matrix factorization with deep neural networks for recommender systems," *Knowl.-Based Syst.*, vol. 145, pp. 46–58, Apr. 2018.
- [55] T. Wu, Y. Feng, J. Sang, B. Qiang, and Y. Wang, "A novel recommendation algorithm incorporating temporal dynamics, reviews and item correlation," *IEICE Trans. Inf. Syst.*, vol. 101, no. 8, pp. 2027–2034, Aug. 2018.
- [56] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, and J. Sun, "Temporal recommendation on graphs via long- and short-term preference fusion," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*. New York, NY, USA: ACM, 2010, pp. 723–732, doi: [10.1145/1835804.1835896](https://doi.org/10.1145/1835804.1835896).
- [57] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell, "Temporal collaborative filtering with Bayesian probabilistic tensor factorization," in *Proc. SIAM International Conf. Data Mining*. Philadelphia, PA, USA: SIAM, 2010, pp. 211–222.
- [58] M. Yang, Q. Qu, X. Chen, W. Tu, Y. Shen, and J. Zhu, "Discovering author interest evolution in order-sensitive and semantic-aware topic modeling," *Inf. Sci.*, vol. 486, pp. 271–286, Jun. 2019.
- [59] H. Yin, B. Cui, L. Chen, Z. Hu, and X. Zhou, "Dynamic user modeling in social media systems," *ACM Trans. Inf. Syst.*, vol. 33, no. 3, pp. 1–44, Mar. 2015.
- [60] F. Zafari, I. Moser, and T. Baarslag, "Modelling and analysis of temporal preference drifts using a component-based factorised latent approach," *Expert Syst. Appl.*, vol. 116, pp. 186–208, Feb. 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417418305815>
- [61] C. Zhang, K. Wang, H. Yu, J. Sun, and E.-P. Lim, "Latent factor transition for dynamic collaborative filtering," in *Proc. SIAM Int. Conf. Data Mining*, 2014, pp. 452–460, doi: [10.1137/1.9781611973440.52](https://doi.org/10.1137/1.9781611973440.52).
- [62] W. Zhang and J. Wang, "Integrating topic and latent factors for scalable personalized review-based rating prediction," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 11, pp. 3013–3027, Nov. 2016.
- [63] E. Zheng, G. Y. Kondo, S. Zilora, and Q. Yu, "Tag-aware dynamic music recommendation," *Expert Syst. Appl.*, vol. 106, pp. 244–251, Sep. 2018.
- [64] L. Zheng, V. Noroozi, and P. S. Yu, "Joint deep modeling of users and items using reviews for recommendation," in *Proc. 10th ACM Int. Conf. Web Search Data Mining*. New York, NY, USA: Association Computing Machinery, 2017, pp. 425–434, doi: [10.1145/3018661.3018665](https://doi.org/10.1145/3018661.3018665).



CHARINYA WANGWATCHARAKUL received the B.S. degree in computer science from Burapha University, Thailand, in 2006, and the M.Sc. degree in computer science from Thammasat University, Thailand. She is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science, Faculty of Science, Khon Kaen University, Thailand. Her main research interests include machine learning and recommender systems.



SARTRA WONGTHANAVASU (Member, IEEE) received the B.S. degree in mathematics from Khon Kaen University, Thailand, in 1982, the M.S. degree in computer science from the Illinois Institute of Technology, USA, in 1996, and the Ph.D. degree in computer science from the Asian Institute of Technology, Thailand, in 2001. He has authored a variety of articles and served as a Reviewer for dozens of journals, such as the IEEE INTELLIGENT SYSTEMS, the IEEE TRANSACTIONS ON CYBERNETICS, *Expert Systems with Applications*, and *Neural Computing and Applications*. His research interests include machine learning, computer vision, cellular automata, and knowledge engineering.

...