EDMF: Efficient Deep Matrix Factorization With Review Feature Learning for Industrial Recommender System

Hai Liu, *Senior Member, IEEE*, Chao Zheng, Duantengchuan Li, Xiaoxuan Shen, Ke Lin, Jiazhang Wang, Zhen Zhang, Zhaoli Zhang, *Member, IEEE*, and Neal N. Xiong, *Fellow, IEEE*

***Abstract*—Recommendation accuracy is a fundamental problem in the quality of the recommendation system. In this article, we propose an efficient deep matrix factorization (EDMF) with review feature learning for the industrial recommender system. Two characteristics in user's review are revealed. First, interactivity between the user and the item, which can also be considered as the former's scoring behavior on the latter, is exploited in a review. Second, the review is only a partial description of the user's preferences for the item, which is revealed as the sparsity property. Specifically, in the first characteristic, EDMF extracts the interactive features of onefold review by convolutional neural networks with word-attention mechanism. Subsequently, L_0 norm is leveraged to constrain the review considering that the review information is a sparse feature, which is the second characteristic. Furthermore, the loss function is constructed by maximum *a posteriori* estimation theory, where the interactivity and sparsity property are converted as two prior probability functions. Finally, the alternative minimization algorithm is introduced to optimize the loss functions. Experimental results on several datasets demonstrate that the proposed methods, which show good industrial conversion application prospects, outperform the state-of-the-art methods in terms of effectiveness and efficiency.**

***Index Terms*—Deep matrix factorization, industrial recommender system, interactivity, L_0 norm, sparsity property.**

Manuscript received February 23, 2021; revised May 16, 2021, September 4, 2021, and October 18, 2021; accepted November 10, 2021. Date of publication November 16, 2021; date of current version April 13, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62177018, Grant 62177019, Grant 62107017, Grant 62011530436, Grant 62077020, Grant 62005092, and Grant 61875068, in part by China Postdoctoral Science Foundation under Grant 2020M682454, and in part by the Fundamental Research Funds for the Central Universities under Grant CCNU20ZT017 and Grant CCNU2020ZN008. Paper no. TII-21-0813.
(Corresponding author: Duantengchuan Li.)

Hai Liu and Zhaoli Zhang are with the National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079, China (e-mail: hailiu0204@ccnu.edu.cn; zl.zhang@ccnu.edu.cn).

Duantengchuan Li is with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: dtclee1222@whu.edu.cn).

Chao Zheng, Xiaoxuan Shen, and Neal N. Xiong are with the National Engineering Laboratory for Educational Big Data, Central China Normal University, Wuhan 430079, China (e-mail: zhengchao9528@gmail.com; shenxiaoxuan@mail.ccnu.edu.cn; 664519872@qq.com).

Ke Lin is with the Department of Control Science and Engineering, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: schris_lin@stu.hit.edu.cn).

Jiazhang Wang is with the Northwestern University, Evanston, IL 60208 USA (e-mail: jiazhang.wang@u.northwestern.edu).

Zhen Zhang is with the College of Sports Engineering and Information Technology, Wuhan Sports University, Hubei 430079, China (e-mail: zhangzhen@whsu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3128240.

Digital Object Identifier 10.1109/TII.2021.3128240

I. Introduction

Recommendation systems, which play an increasingly crucial role in overcoming information overload, have been widely applied in many industrial fields, such as in movies [1], music [2], news [3], e-commerce [4], online e-learning [5], and industrial applications [6]. The recommendation system matches the individual needs of users precisely with the real attributes of items in industrial applications, and it also can provide students personalized resource service in the online learning platforms.

Over the past decades, a large number of recommendation algorithms have been proposed. Matrix factorization (MF) [7]–[9] can be considered as a popular collaborative filtering-based approach that aims to represent the users and items with latent factors. Although these techniques have shown powerful performance, the sparsity problem of rating data is considered as a primary reason that affects the accuracy of these methods. Providing accurate recommendation service is difficult because some items or users have few ratings.

In order to improve the accuracy of recommender systems, numerous studies utilized extra information to construct an accurate latent factor of the user or item to enhance the recommendation performance, for instance, review [10]–[13], social networks [14], and side information [15], [16]. Experiments prove that the introduction of extra information can indeed improve the performance of recommendations. Review, which usually contains rich information about the preferences of an individual user and the characteristics of the items, plays an important role in mitigating the sparsity problem of recommendation. In most social media and e-commerce websites, such as Yelp and Amazon, users can be permitted to write reviews to express their personal opinions and give numerical star ratings. Thus, the review on the user's preferences for items (e.g., material, color, and quality) contains numerous information. This information will help us improve the recommendation accuracy.

1551-3203 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

Review and rating data have been utilized in existing works to enhance the rating prediction accuracy further. From the viewpoints of methodology, available methods can be mainly classified into two categories. Namely, the bags of words model and convolutional neural networks (CNNs) model [17]. In the bags of words model, several methods, such as the ratings meet reviews [11], collaborative topic regression [12], and hidden factors as topics [13] model, opt for topic modeling techniques to learn the topic factors of review, and integrate them into MF models to predict ratings. These approaches have shown significant improvements than conventional latent models, which merely rely on ratings. However, these works have some limitations. For instance, these works employ the bag of words representation for the latent topic. Thus, sentiment analysis and feature extraction of reviews via manual preprocessing are required here, whereas word order and local context information are ignored.

In recent years, CNNs have yielded tremendous success in industrial fields. Some studies also attempted to combine CNNs with MF frameworks to raise recommendation accuracy. In [18], the authors proposed a deep collaborative neural network (DeepCoNN) by jointly modeling the users and items in reviews to derive the latent factors of items and users from the reviews. To reveal the feature representation of reviews, Chen *et al.* [19] developed a novel CNNs-based model with two parallel neural attentional regression subnetworks. In [20], the authors observed that user-item pairs could represent the context-aware information in the recommended process. They proposed a representation learning model with a linear fusion mechanism, which could achieve high rating prediction accuracy. CNNs-based methods, which are better for feature extraction, do have a great advantage in accuracy when compared with the previous bags of words models. However, they exhibit a common trait of using aggregated review texts of users and items to extract latent factors, which cost colossal training time and high parameter complexity. Thus, the latent factors of user and item are extracted in a static and independent manner.

Because the CNNs-based methods train all the reviews related to the user/item every time, the scale of their parameter is very large. The time complexity and space complexity of the model are both high, which makes these models difficult to land in industry. At the same time, when extracting review features, they also ignore the characteristics of the review text itself. Considering these facts, we have done some explorations on the feature extraction of a review text. There is a relationship between reviews and rating data. A user's review is an interactive message in itself, which is called interactivity in this article, and the review text contains not only the user's attributes and preferences, but also the item's attributes, which is also regarded as an evaluation mechanism. When a user evaluates an item in an online platform, the rating data and the review data are actually two different ways of evaluating the users' shopping behavior for the item. Thus, a rating and a review should reach a match or fit to some extent. Namely, ratings and reviews are two different manifestations on the same event, and their feature representations should be as close as possible to some extent. Therefore, we align the features extracted from a onefold review text with the Hadamard product of user and item latent factors in the same space for feature alignment. Since our model is trained on a onefold review at a time, the training efficiency of the model is greatly improved, which can work well in practical applications. In addition, considering the characteristics of the review itself, the user's review could only partially reflect the user's rating behavior. As a result, a sparsity constraint on a onefold review feature is used in the proposed models.

To construct an efficient recommendation model, we fully utilize the reviews and propose a novel MF model with L_0-regularized review feature learning called an efficient deep MF (EDMF). Moreover, this article reveals the interactivity and sparsity of the review and constructs a novel recommendation model. The major contributions of this article can be summarized in the following three aspects.

1) A novel recommendation framework called EDMF, which introduces the onefold review feature representation vector with MF technique, is proposed to extract the review features between corresponding user and item. This framework can improve training efficiency and fully utilize the interactive characteristics of reviews.
2) A L_0 norm-based review modeling framework is proposed with sparseness constraint on review texts based on the finding that the review information could only manifest user preferences partially.
3) Given the interactive and sparsity characteristics of the review, four representative real-world datasets are taken for the comparative experiments to verify the performance of the proposed method. The results demonstrate that the EDMF achieves significant improvements for recommendation tasks in terms of prediction accuracy and training efficiency.

The rest of this article is organized as follows. Section II presents the feature analysis of the review information. Section III elaborates the proposed EDMF model in detail. Section IV introduces the optimization and parameter determination methods. Section V presents the experimental results on four public datasets using several state-of-the-art methods. Finally, Section VI concludes this article.

II. Analysis of Review Characteristics

A. Problem Formulation

In collaborative filtering approaches [21], [22], the recommendation task can be referred to the use of a group of like-preference users with common experience to recommend information that users are interested in. Given items M, users N, and the observed dataset $\mathbf{R} \in \mathbb{R}^{N\times M}$, the user-item rating matrix $\mathbf{Z}$ can be estimated from the data matrix $\mathbf{R}$. That is, the problem of the recommendation systems is to predict the unobserved elements in matrix $\mathbf{Z}$ from matrix $\mathbf{R}$, which needs to construct an estimator to remove the rating noise. The MF is a famous user-item matrix estimation algorithm based on collaborative filtering technology. Usually, the standard model of the rating data $\hat{\mathbf{R}}$ can be formulated as [23]:

$$\hat{\mathbf{R}} = \mathbf{U} \times \mathbf{V}^T + \mathbf{\Gamma}_U + \mathbf{\Gamma}_V + \bar{\mathbf{R}} \tag{1}$$

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

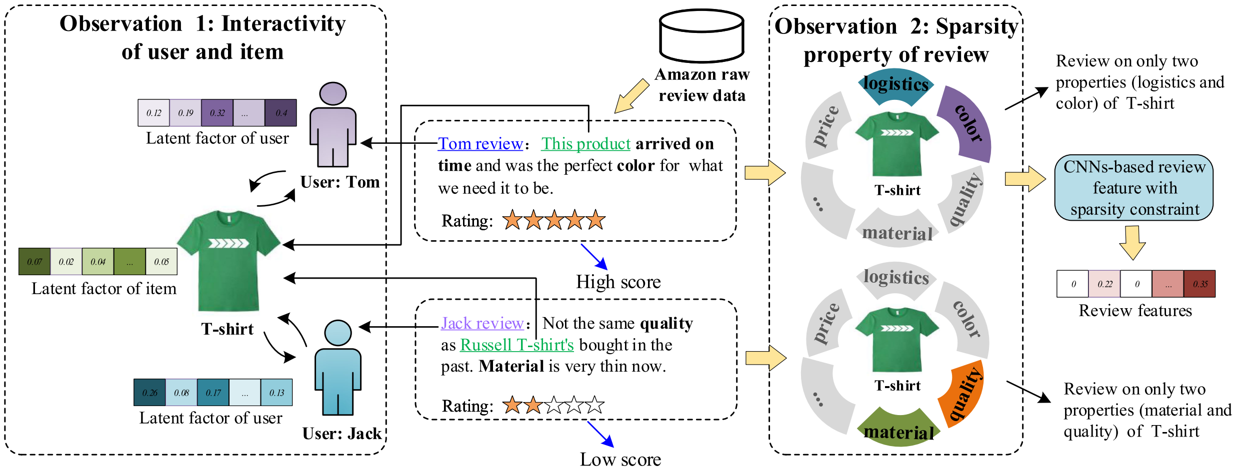


Fig. 1. Characteristic analysis of review from the Russell Athletic 'sShort sleeve t-shirt" on Amazon.

where $\mathbf{U} \in \mathbb{R}^{N\times k}$ and $\mathbf{V} \in \mathbb{R}^{M\times k}$ are the latent factors of users and items, respectively. k denotes the dimension of latent factors. Matrixes $\boldsymbol{\Gamma}_U$ and $\boldsymbol{\Gamma}_V$ are introduced to present the observed deviations of users and items, respectively. $\bar{\mathbf{R}}$ represents the overall average rating. Thus, the MF aims to utilize the rating and review information to build accurate users and items latent factor and bias.

B. Characteristics of Review Information

In fact, review text contains abundant semantic information of users and items, which is also a kind of interactive information between users and items. In this article, this characteristic of review is termed as interactivity. This finding motivates us to model a onefold review for the recommendation. In the first place, the model only needs to process one review text at a time instead of processing the whole text collection, thereby reducing the training time of the model greatly. Subsequently, the interactivity of the reviews will be revealed and considered in our proposed model.

Interactivity (see **Observation 1** in Fig. 1) and sparsity (see **Observation 2** in Fig. 1) are observed in the review information. The example in Fig. 1 shows that a review of the same item (short sleeve t-shirt) can be obtained high (by Tom) or low scores (by Jack). On the one hand, Tom decided to give five stars based on his various reasons for purchasing the t-shirt. However, his review only contains two aspects of user evaluation, namely, t-shirt "logistics" and "color". On the other hand, Jack provided several reasons for his low rating. However, his review may only reflect two other causes, namely, "material" and "quality". Therefore, the user's rating score behavior is a comprehensive evaluation of an item, which is a multifaceted evaluation result. But then again, the user's review could only partially reflect the user's rating behavior. In other words, multiple aspects of items, ranging from a few to dozens, are involved when a user scores an item. However, the user's review information often only contains the evaluation of several aspects of each item. Thus, this characteristic of review is identified as the sparsity property.

To achieve an efficient and robust recommendation performance, EDMF with L_0-regularized review feature learning has been proposed. Initially, review feature representation learning is proposed to extract the feature vector of onefold review through a convolution operation and word attention mechanism. Furthermore, the model applies the sparsity constraint on the obtained feature vector, which finally completes the rating prediction. In this way, the EDMF extracts the interactive characteristics of the review considering that the semantic information contained in the review text can merely reflect the user's rating behavior partially. Thus, our model utilized the review information effectively to predict users' interests and improve the prediction performance.

III. Our Proposed EDMF

A. Outline of EDMF

1) Perspective of the Module: EDMF can be summed as three main processes, namely, review feature representation learning, latent factor representation learning, and rating prediction based on the sparsity constraint (see Fig. 2). In the review feature representation learning, the review contextual feature vector is captured by inputting the onefold review text into the CNNs with a word attention mechanism. Then, the review text feature vector is used to constrain latent factor representation learning partially. The zero-mean spherical Gaussian priors [24] were assumed on the latent factor representation of users and items. Finally, the review text feature vector is used to partially constrain the Hadamard product of user and item latent factor vectors in rating prediction based on the sparsity constraint.

2) The View of Probability: In recent years, the maximum *a posteriori* (MAP) estimation method, which employs a prior probability density function (PDF) as a prior constraint, has been widely used for parameter estimation. It has played a key role in ill-posed inverse problems, which widely exist in practical applications. Inspired by these, introducing the MAP framework in the parameter optimization of recommendation is instinctive. The purpose of this introduction is to execute the MAP estimation of users and items latent factors given the

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

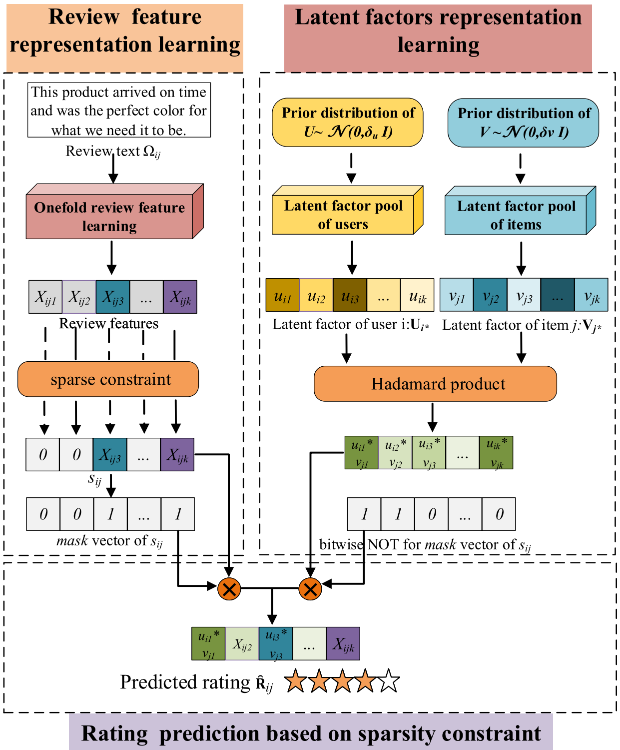


Fig. 2. Outline of EDMF framework.

ratings and reviews

$$\max_{\mathbf{U},\mathbf{V}} p\left(\mathbf{U},\mathbf{V}|\mathbf{R},\mathbf{\Omega},\mathbf{\Gamma}_U,\mathbf{\Gamma}_V,\mathbf{W},\mathbf{\Phi},\delta,\delta_U,\delta_V,\sigma\right). \quad (2)$$

Applying the Bayes' rule, (2) can be presented as

$$\begin{aligned}&\max_{\mathbf{U},\mathbf{V}} p\left(\mathbf{U},\mathbf{V}|\mathbf{R},\mathbf{\Omega},\mathbf{\Gamma}_U,\mathbf{\Gamma}_V,\mathbf{W},\mathbf{\Phi},\delta,\delta_U,\delta_V,\sigma\right)\\&=\max_{\mathbf{U},\mathbf{V}} \underbrace{p\left(\mathbf{R}|\mathbf{U},\mathbf{V},\mathbf{\Gamma}_U,\mathbf{\Gamma}_V,\delta\right)}_{\text{likelihood probability}}\times\underbrace{p\left(\mathbf{U}|\delta_U\right)p\left(\mathbf{V}|\delta_V\right)}_{\text{prior probability}}\\&\quad\times\underbrace{p(\mathbf{U},\mathbf{V}|\mathbf{\Omega},\mathbf{W},\mathbf{\Phi})p\left(\mathbf{W},\mathbf{\Phi}|\mathbf{\Omega},\delta_S\right)}_{\text{review with } L_0 \text{ norm prior probability}}.\end{aligned} \quad (3)$$

It can be seen that three PDFs' terms need to be defined. They will be constructed in detail as follows.

B. Latent Factors Representation Learning

1) Likelihood Probability: According to the rating data observation model (1), the measure errors in the rating score data are modeled as a set of independent and identically distributed random variables for all rating data. That is, each measure error obeys the Gaussian distribution $\epsilon\sim\mathcal{N}(0,\delta^2)$. There are two main reasons for assuming that the measurement error follows a Gaussian distribution. First, for many previous works [18]–[20], [24], they have assumed that the measurement error for scoring obeys a Gaussian distribution. They all constructed recommendation models based on this assumption and achieved relatively good results. Second, in probability theory, the central limit theorem establishes that, in many situations, when independent random variables are added, their properly normalized sum tends toward a normal distribution (Gaussian distribution) even if the original variables themselves are not normally distributed. Thus,

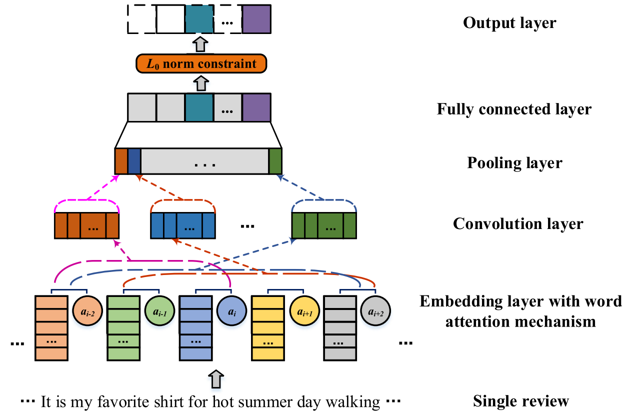


Fig. 3. Onefold review feature learning with L_0 norm constraint.

we adopt a probabilistic linear model with the Gaussian observation noise. Under this assumption, the probability density of the observed ratings can be formulated as

$$p\left(\mathbf{R}|\mathbf{U},\mathbf{V},\mathbf{\Gamma}_U,\mathbf{\Gamma}_V,\delta\right)=\prod_{j=1}^{\mathrm{M}}\prod_{i=1}^{\mathrm{N}}\left[\mathcal{N}\left(\mathbf{R}_{ij}|\hat{\mathbf{R}}_{ij},\delta^2\right)\right]^{\mathbf{1}_{ij}} \quad (4)$$

where $\mathbf{1}_{ij}$ denotes the indicator function, which is equal to 1 if the rating existed in matrix $\mathbf{R}$ and 0, if otherwise.

2) Prior Probability: For the second probability, the latent feature vectors of users and items are assumed as the Gaussian priors with zero mean, namely

$$p\left(\mathbf{U}|\delta_U\right)=\prod_{i=1}^{\mathrm{N}}\mathcal{N}\left(\mathbf{U}_{i^*}|0,\delta_U^2 I\right) \quad (5)$$

$$p\left(\mathbf{V}|\delta_V\right)=\prod_{j=1}^{\mathrm{M}}\mathcal{N}\left(\mathbf{V}_{j^*}|0,\delta_V^2 I\right) \quad (6)$$

where $\delta_V^2 I$ and $\delta_U^2 I$ are the covariance matrices of item latent matrix V and user latent matrix U, respectively.

C. Review Feature Representation Learning

1) Onefold Review Feature Learning: Convolution operation and attention mechanism are utilized for learning relevant semantic information features from a onefold user's review. Fig. 3 shows the network architecture of the review feature learning, which consists of five parts, namely, embedding layer with word attention mechanism, convolution layers, pooling layer, fully connected layer, and output layer.

In the first layer, a pretrained word vector model (i.e., word2vec) is utilized to convert the original text information into digital information and concatenate all the representation vectors of words to stack a text data into a review matrix $\mathbf{D}\in\mathbb{R}^{l\times d}$. Furthermore, a word attention matrix $\mathbf{\Phi}\in\mathbb{R}^{l\times 1}$ that represents the contribution of the word to the final review feature vector is adopted to derive the importance of each word in the review

$$\mathbf{D}\circ\mathbf{\Phi}=\begin{bmatrix}\cdots & \vert & \vert & \vert & \cdots\\ \cdots & \phi_{i-1}\mathbf{d}_{i-1} & \phi_i\mathbf{d}_i & \phi_{i+1}\mathbf{d}_{i+1} & \cdots\\ \cdots & \vert & \vert & \vert & \cdots\end{bmatrix}. \quad (7)$$

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

In the convolution layer, multiple shared weight convolution filters are used to extract the context features of each word with its local context. Specifically, *j*th convolution filter $\mathbf{W}^j$ with shared weights extract local contextual features c_i^j. The convolution operation is performed for every local contextual window when filter slides one step

$$c_i^j = \mathbf{W}^j * \left(\mathbf{D} \circ \mathbf{\Phi}\right)_{(:,i:(i+t-1))} + b^j \tag{8}$$

where *i* is the current index of the filter on the review word matrix ($i = 0, 1, \ldots, T - t$), $*$ denotes the convolution operator, $b_j \in \mathbb{R}$ is a bias for $\mathbf{W}^j$, and *t* signifies the length of the sliding window filter.

For the third layer, the features extracted by the convolution kernel are sampled by a down-sampling operation. The three previous layers have learned the local features of the review. To obtain global features, the fully connected network is leveraged to combine these local features. Thus, the review feature representation is regarded as a function

$$s_{ij} = \mathrm{C\,N\,N}_{\mathbf{W},\mathbf{\Phi}}\left(\Omega_{ij}\right) \tag{9}$$

where $\mathbf{W}$ denotes all the weight and bias variables in our network and $\mathbf{\Phi}$ represents the attention matrix of each word. Ω_{ij} is regarded as the user *i*'s raw review to item *j*.

Eventually, the output layer imposes a zero norm constraint on the results of the previous layer to obtain the final review feature representation.

2) Review Feature Representation With L_0 Norm: In our previous analysis, feature inequality has been observed between reviews and ratings. Rating score has comprehensive characteristic, which is represented by the Hadamard product of latent factors $\mathbf{U}_{i*}$ and $\mathbf{V}_{j*}$, namely $\mathbf{U}_{i*} \circ \mathbf{V}_{j*}$. In addition, a review has sparsity characteristic, in which the extracted vectors can only reflect the user preference partially. In other words, the vector extracted from a onefold review may contain only a few important values. Thus, the number of values in the review feature vector, which is related to L_0 norm mathematically for information nonzero element constraints, needs to be limited. L_0 norm is introduced to measure the number of nonzero elements in a vector, which can also be considered a sparsity constraint. Zero norm is the most direct and effective way to build a model to build the sparsity of review. Thus, it is natural for us to impose *a priori* constraint of the L_0 norm on the review features vector s_{ij}. $L_0(s_{ij})$ is utilized to show the result of the zero-norm constraint on the vector s_{ij}, which retains the sectional features learned in the review and constrains vectors of $\mathbf{U}_{i*} \circ \mathbf{V}_{j*}$.

Sparsity constraints have been performed by limiting the numbers of nonzero in the vectors of onefold review feature representation. The means for our sparsity constraint is to count the nonzero number of vectors, which is written as

$$C\left(s_{ij}\right) = \#\left\{(i,j,p) | s_{ijq} \neq 0\right\} \tag{10}$$

where q represents the index in the vector. $\#\{\bullet\}$ denotes the counting operator, which outputs the number of s_{ij} if the element value $s_{ijq} \neq 0$, that is, the L_0 norm of the review feature vector.

D. Rating Prediction Based on Sparsity Constraints

1) Review With L_0 Norm Prior Probability: Through the feature analysis of the L_0 norm in previous section, the review with L_0 norm prior probability is represented as $p(\mathbf{U}, \mathbf{V}|\mathbf{\Omega}, \mathbf{W}, \mathbf{\Phi})p(\mathbf{W}, \mathbf{\Phi}|\mathbf{\Omega}, \delta_S)$, where δ_s denotes the parameters of the sparsity constraint of L_0 norm.

To facilitate subsequent calculation process, we take negative logarithm on (3). Thus, MAP estimation is reformulated to minimize the loss function

$$\begin{aligned}\mathcal{L}_{\text{EDMF}} =& \sum_j^{\mathrm{M}} \sum_i^{\mathrm{N}} \frac{1_{ij}}{2}\left(\mathbf{R}_{ij} - \hat{\mathbf{R}}_{ij}\right)^2 + \frac{\lambda_U}{2}\|\mathbf{U}\|_F^2 + \frac{\lambda_V}{2}\|\mathbf{V}\|_F^2 \\ &+ \frac{\lambda_R}{2} \sum_j^{\mathrm{M}} \sum_i^{\mathrm{N}} 1_{ij} \left\|\left(\mathbf{U}_{i*} \circ \mathbf{V}_{j*} - s_{ij}\right) \circ mask\left(s_{ij}\right)\right\|_F^2 \\ &+ \lambda_S \sum_j^{\mathrm{M}} \sum_i^{\mathrm{N}} 1_{ij} L_0\left(s_{ij}\right)\end{aligned} \tag{11}$$

where $\|\bullet\|_F^2$ denotes the Frobenius norm square. The symbol s_{ij} is a *k*-dimension vector that represents the result after sparsity constraint with L_0 norm. The symbol $L_0(s_{ij})$ counts the number of nonzero in s_{ij}. Magic function $mask(s_{ij})$ can mark the positions of nonzero elements in s_{ij}. λ_S, λ_U, λ_V, and λ_R are all the weighted parameters that are the trade-offs between the first fidelity item and three regularization items.

2) Considering the Length of Review: Furthermore, the length of review is proportional to its confidence. Generally, longer review contains more information, which is more valuable for feature extraction. Thus, the model that considers the influence of the length of the review is called EDMF+, which can be written as

$$\begin{aligned}\mathcal{L}_{\text{EDMF+}} =& \sum_j^{\mathrm{M}} \sum_i^{\mathrm{N}} \frac{1_{ij}}{2}\left(\mathbf{R}_{ij} - \hat{\mathbf{R}}_{ij}\right)^2 + \frac{\lambda_U}{2}\|\mathbf{U}\|_F^2 + \frac{\lambda_V}{2}\|\mathbf{V}\|_F^2 \\ &+ \frac{\lambda_R}{2} \sum_j^{\mathrm{M}} \sum_i^{\mathrm{N}} 1_{ij} \left\|\left(\mathbf{U}_{i*} \circ \mathbf{V}_{j*} - s_{ij}\right) \circ mask\left(s_{ij}\right)\right\|_F^2 \\ &\times\ log(len(\Omega_{ij})) + \lambda_S \sum_j^{\mathrm{M}} \sum_i^{\mathrm{N}} 1_{ij} C\left(s_{ij}\right)\end{aligned} \tag{12}$$

where $len(\Omega_{ij})$ is the length of review. Intuitively, the longer the review text is, the more reliable the review information will be. The review information vectors extracted by the corresponding review feature representation learning can represent the interaction vectors between users and items better.

IV. Optimization and Parameters Determination

Considering that L_0 norm is nonconvex [25], solving (11) and (12) directly by using traditional optimization algorithms is difficult. Inspired by [26] and[27], an effective alternating minimization algorithm with semiquadratic splitting is introduced to optimize (11) and (12).

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

A. Optimization of L_0 Norm

According to the alternating minimization algorithm, (11) is solved by minimizing $\mathbf{U}$, $\mathbf{V}$, s_{ij}, and $\mathbf{M}_{ij}$ alternatively. In each step, one set of variables are fixed by using the values obtained from the previous iteration. In other words, four variables in (11) are updated in a cyclic manner.

1) Subproblem of $\mathbf{U}$*:* $\mathbf{U}$ estimation subproblem corresponds to minimize

$$\mathcal{L}_{\mathbf{U}} = \sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} \frac{1_{ij}}{2}\left(\mathbf{R}_{ij} - \hat{\mathbf{R}}_{ij}\right)^2 + \frac{\lambda_U}{2}\|\mathbf{U}\|_F^2 + \frac{\lambda_R}{2}\sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} 1_{ij}\left\|\left(\mathbf{U}_{i*}\circ\mathbf{V}_{j*} - s_{ij}\right)\circ \mathit{mask}\left(s_{ij}\right)\right\|_F^2 \tag{13}$$

which is a quadratic function, and thus, can be optimized directly by using gradient decent algorithm. Then, $\mathbf{U}_{i*}$ is updated on the basis of the iteration rule as follows

$$\begin{aligned}\mathbf{U}_{\mathbf{i}^*} &\leftarrow \mathbf{U}_{\mathbf{i}^*} - \alpha\tfrac{\mathcal{L}_U}{\partial \mathbf{U}_{\mathbf{i}*}}\\ &= \mathbf{U}_{i^*} - \alpha\left\{\sum_{j}^{\mathrm{M}} 1_{ij}\left(\mathbf{R}_{ij} - \widehat{\mathbf{R}}_{ij}\right) \times \mathbf{V}_{j^*} + \lambda_U \mathbf{U}_{i^*}\right.\\ &\left.+\sum_{j}^{\mathrm{M}} 1_{ij}\lambda_R\left(\left(\mathbf{U}_{i*}\circ\mathbf{V}_{j*} - s_{ij}\right)\circ \mathit{mask}\left(s_{ij}\right)\circ \mathbf{V}_{j^*}\right)\right\}.\end{aligned} \tag{14}$$

2) Subproblem of $\mathbf{V}$*:* Given that $\mathbf{V}$ and $\mathbf{U}$ are equivalent, the updated formulation with respect to $\mathbf{V}$ can be written as

$$\begin{aligned}\mathbf{V}_{\mathbf{j}*} &\leftarrow \mathbf{V}_{\mathbf{j}*} - \alpha\tfrac{\mathcal{L}_V}{\partial V_{\mathbf{j}*}}\\ &= \mathbf{V}_{j^*} - \alpha\left\{\sum_{j}^{\mathrm{M}} 1_{ij}\left(\mathbf{R}_{ij} - \hat{\mathbf{R}}_{ij}\right)\cdot \mathbf{U}_{i*} + \lambda_V \mathbf{V}_{j*}\right.\\ &\left.+\lambda_R\sum_{j}^{\mathrm{M}} 1_{ij}\left(\left(\mathbf{U}_{i*}\circ\mathbf{V}_{j*} - s_{ij}\right)\circ \mathit{mask}\left(s_{ij}\right)\circ \mathbf{U}_{i*}\right)\right\}.\end{aligned} \tag{15}$$

3) Subproblem of s_{ij}*:* Due to the L_0 norm term in (11), the minimization of this equation is a computationally intractable question. Thus, our method is only an approximation, thereby can make the problem easily addressed and uphold the algorithm property. We introduce a k-dimension auxiliary variable $\mathbf{M}_{ij}$ with the same dimension as the constraint target vector, which corresponds to s_{ij}, and (11) could be formulated as

$$\begin{aligned}\mathcal{L}_{\text{EDMF}} =& \sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} \frac{1_{ij}}{2}\left(\mathbf{R}_{ij} - \hat{\mathbf{R}}_{ij}\right)^2 + \frac{\lambda_U}{2}\|\mathbf{U}\|_F^2 + \frac{\lambda_V}{2}\|\mathbf{V}\|_F^2\\ &+ \frac{\lambda_R}{2}\sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} 1_{ij}\left\|\left(\mathbf{U}_{i*}\circ\mathbf{V}_{j*} - s_{ij}\right)\circ \mathit{mask}\left(s_{ij}\right)\right\|_F^2\\ &+ \lambda_S\sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} 1_{ij} C\left(\mathbf{M}_{ij}\right) + \beta\sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} 1_{ij}\left\|s_{ij} - \mathbf{M}_{ij}\right\|_F^2\end{aligned} \tag{16}$$

where $C(\mathbf{M}_{ij}) = \#\{(i,j,q)|\mathbf{M}_{ijq} \neq 0\}$ and is an automatically adaptive parameter to control the similarity between variables $\mathbf{M}_{ij}$ and their corresponding vector s_{ij}.

Thus, the recursion partial derivative equation for s_{ij} is given as

$$\begin{aligned}s_{ij} &\leftarrow s_{ij} - \alpha\tfrac{\mathcal{L}_{s_{ij}}}{\partial s_{ij}}\\ &= s_{ij} - \alpha\left\{\lambda_R\sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} 1_{ij}\tfrac{\partial\{(\mathbf{U}_{i*}\circ\mathbf{V}_{j*} - s_{ij})\circ\mathrm{mask}(s_{ij})\}}{\partial s_{ij}}\right.\\ &\left.+2\beta\sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} 1_{ij}\left(s_{ij} - \mathbf{M}_{ij}\right)\right\}.\end{aligned} \tag{17}$$

4) Subproblem of $\mathbf{M}_{ij}$*:* The objective function (16) with respect to $\mathbf{M}_{ij}$ is written as

$$\min_{\mathbf{M}_{ij}}\sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} 1_{ij}\left\{\frac{\lambda_S}{\beta}C\left(\mathbf{M}_{ij}\right) + \left\|s_{ij} - \mathbf{M}_{ij}\right\|_F^2\right\} \tag{18}$$

where $C(\mathbf{M}_{ij})$ denotes the number of nonzero elements in $\mathbf{M}_{ij}$. This apparently sophisticated subproblem can actually be solved fast because (18) can be spatially decomposed, and each element $\mathbf{M}_{ijq}$ can be estimated individually. This is the main benefit from our splitting scheme, which makes the altered problem empirically solvable. Thus, (18) is decomposed accordingly to

$$\sum_{q}^{k}\min_{\mathbf{M}_{ijq}}\sum_{j}^{\mathrm{M}}\sum_{i}^{\mathrm{N}} 1_{ij}\left\{\left\|s_{ij} - \mathbf{M}_{ij}\right\|_F^2 + \frac{\lambda_S}{\beta}\times H\left(\mathbf{M}_{ijq}\right)\right\} \tag{19}$$

where $H(\mathbf{M}_{ijq})$ is a binary function that returns to 1 if $\mathbf{M}_{ijq} \neq 0$ and 0, if otherwise. Each single term with respect to vector $\mathbf{M}_{ij}$ in (19) is

$$\mathcal{L}_{\mathbf{M}_{ijq}} = \left(s_{ijq} - \mathbf{M}_{ijq}\right)^2 + \frac{\lambda_S}{\beta}\times H\left(\mathbf{M}_{ijq}\right) \tag{20}$$

which reaches its minimum $\mathcal{L}_{\mathbf{M}_{ijq}}$ when

$$\mathbf{M}_{ijq} = \begin{cases} s_{ijq} & \frac{\lambda_S}{\beta} < \left(s_{ijq}\right)^2\\ 0 & \frac{\lambda_S}{\beta} \geq \left(s_{ijq}\right)^2\end{cases}. \tag{21}$$

In this step, we calculate the minimum loss $\mathcal{L}_{\mathbf{M}_{ijq}}$ of each dimension value of the vector based on the abovementioned derivation. In sum, that is, calculating $\sum_q^k \mathcal{L}_{\mathbf{M}_{ijq}}$ yields the global optimum for (19). Additionally, the alternating minimization of EDMF algorithm can be sketched in Algorithm 1.

Algorithm 1: EDMF With Review Feature Learning for Recommendation.

Input: $\mathbf{R}$: user-item rating matrix, $\mathbf{\Omega}$: review text of items
Set: dimensionality k, learning rate α, batch size b;
1: Matrix $\mathbf{U}_{i*}$ and $\mathbf{V}_{j*}$ initializing with random numbers, set $\mathbf{M}_{ij} = 0$;
2: While not $\mathcal{L}_{\text{EDMF}}$ is converged **do**:
  While not $\mathbf{U}, \mathbf{V}$ is converged **do**:
    with $\mathbf{V}_{j*}$, s_{ij}, and $\mathbf{M}_{ijq}$, solve for $\mathbf{U}_{i*}$ with (14);
    with $\mathbf{U}_{i*}$, s_{ij}, and $\mathbf{M}_{ijq}$, solve for $\mathbf{V}_{j*}$ with (15);
    with $\mathbf{U}_{i*}$, $\mathbf{V}_{j*}$, and $\mathbf{M}_{ijq}$, solve for s_{ij} with (17);
  end while
  with $\mathbf{U}_{i*}$, $\mathbf{V}_{j*}$ and s_{ij}, solve for $\mathbf{M}_{ijq}$ in (21).
end while
Output: EDMF model

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

TABLE I
STATISTICAL DETAILS OF THE DATASETS FOR RECOMMENDATION

Datasets	Users	Items	Ratings&Reviews	Sparsity
Automotive	2,928	1,835	20,473	99.619%
Movies_and_TV	14,169	17,795	673,342	99.733%
Video_Games	24,303	10,672	231,780	99.910%
Yelp_2018	36,989	48,813	1,578,463	99.913%

B. Parameter Determination

In the EDMF model, six parameters need to be determined. In review feature representation, k denotes the dimensionality of the review text and user/item latent factor vectors, which can control the representation ability of the proposed models. Meanwhile, the hyperparameter φ that affects the generalization ability refers to the parameter in dropout. The function of parameters λ_U, λ_V, λ_R, and λ_S are to adjust the constraint strength of regularization items in optimizing user and item representations.

To select the best values for the parameters, several algorithms, such as the L-curve method [28], generalized cross-validation (GCV) [29], and discrepancy principle [30], have been proposed. In this article, the GCV algorithm is leveraged to validate the parameter values in a large range and determine the best ones automatically. In other words, the four regularization parameters are determined heuristically. We suggest that promising performance can be achieved with the parameters φ=0.5, k=10, λ_U, λ_V=0.01, λ_R=0.1, λ_S=0.1, and β=10. More parameter details will be discussed in the next section.

V. EXPERIMENTS AND DISCUSSION

A. General Settings

1) Datasets: Four public datasets are selected from different fields to evaluate to the proposed model. Thereinto, movies and TV, video games, and automotive datasets are downloaded from Amazon 5-core. The remaining one is chosen from the famous Yelp Challenge 2018. In Table I, we compute the sparsity of the four public datasets. To satisfy the input requirements of all the comparison methods, we preprocess the raw-rating data to guarantee no less than 20 records are included for each user and item. Generally, each record in the datasets includes the review text and corresponding rating score, which ranges from one to five. In our experiments, each rating dataset is divided into 80% training and 20% testing, and are carried out using the fivefolds cross-validation technique.

2) Evaluation Metrics: To evaluate the performance of the comparison methods, we introduce two common metrics, namely, the root mean square error (RMSE) and the mean absolute error (MAE). The formulations of RMSE and MAE can be denoted as

$$\mathrm{RMSE} = \sqrt{\frac{1}{|\mathbf{R}_{\text{test}}|} \sum_{\mathbf{R}_{ij} \in \mathbf{R}_{\text{test}}} \left(\mathbf{R}_{ij} - \hat{\mathbf{R}}_{ij}\right)^2} \tag{22}$$

and

$$\mathrm{MAE} = \frac{1}{|\mathbf{R}_{\text{test}}|} \sum_{\mathbf{R}_{ij} \in \mathbf{R}_{\text{test}}} |\mathbf{R}_{ij} - \hat{\mathbf{R}}_{ij}|_{\text{abs}} \tag{23}$$

where $|\mathbf{R}_{\text{test}}|$ represents the cardinality of the test set and $|\bullet|_{\text{abs}}$ means the absolute value operation. The smaller the RMSE and MAE values are, the higher accuracy is achieved.

3) Tested Models: To compare with the proposed method, we selected six famous recommendation algorithms as follows:

a) *PMF* [24]: PMF adopts the Gaussian distribution, which is a standard MF, to model the latent factors of items and users.

b) *HFT* [13]: HFT is the first model to combine reviews and ratings and uses a latent Dirichlet allocation method.

c) *DeepCoNN* [18]: This method extracts the latent factors from item and user review documents by using two parallel deep neural networks.

d) *NARRE* [19]: Neural attentional regression model learns the latent feature vectors of items and users to explore the usefulness of reviews based on neural attention mechanism.

e) *CARL* [20]: CARL utilized CNNs and a dynamic linear fusion mechanism to construct model.

f) *EDMFWA:* A onefold review feature representation learning model without word attentional mechanism with sparsity constraint is proposed in this article.

g) *EDMF:* A onefold review feature representation learning model with sparsity constraint is proposed in this article.

h) *EDMF+:* EDMF+ is a variation of EDMF, which adds review length information to enhance the model performance.

B. Experimental Implementation

The parameters of the comparison methods are adjusted as suitable ones according to their articles. In EDMF, word latent vectors by the GoogleNews word vectors are initialized. In review feature representation learning, various window sizes (e.g., 3, 4, and 5) with shared weights are adopted to grasp the surrounding information of various length words. Subsequently, φ is set at 0.5, and k is equal to ten. EDMF+ has the same parameter settings as EDMF. The experiments are implemented using the software library TensorFlow [31].

C. Results and Discussion

1) Accuracy Analysis of EDMF+: All the recommendation methods are executed on the four public datasets. The comparisons of results are summarized in Table II. The best and second best values are denoted by the bold font and underline, respectively. Based on the rating prediction results, the comparative analysis can be obtained from three aspects.

First, the metric values of HFT, DeepCoNN, NARRE, and CARL methods are less than those of the PMF method in Table II. The four former methods consider the reviews for the rating prediction, whereas the latter methods only consider the rating information. The representation capacity of latent factor

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

TABLE II
COMPARISON OF RATING PREDICTION RESULTS BY USING THE PROPOSED METHOD AND SIX COMPARISON METHODS

Methods	Automotive RMSE	Automotive MAE	Movies_and_TV RMSE	Movies_and_TV MAE
PMF [24]	1.0768	0.8564	1.0428	0.7878
HFT [13]	1.0222	0.7277	1.0267	0.7579
DeepCoNN [18]	0.9305	0.6925	1.0096	0.7323
NARRE [19]	0.9187	0.6446	0.9947	0.7162
CARL [20]	0.9078	0.6207	0.9831	0.7048
EDMFWA	0.9089	0.6102	0.9859	0.7010
EDMF	0.9079	0.6057	0.9841	0.6900
EDMF+	**0.8930**	**0.5865**	**0.9762**	**0.6807**

Methods	Video_Games RMSE	Video_Games MAE	Yelp_2018 RMSE	Yelp_2018 MAE
PMF [24]	1.3965	1.0981	1.1938	0.9191
HFT [13]	1.1115	0.8435	1.1252	0.8702
DeepCoNN [18]	1.0706	0.8050	1.1218	0.8633
NARRE [19]	1.0607	0.7938	1.1106	0.8454
CARL [20]	1.0637	0.7942	1.1204	0.8494
EDMFWA	1.0598	0.7729	1.1043	0.8327
EDMF	1.0541	0.7690	1.0950	0.8244
EDMF+	**1.0521**	**0.7660**	**1.0924**	**0.8121**

Note: The best values are marked by bold font.

is improved substantially because of the rich users and items information in the review text.

Second, the deep learning-based methods (e.g., DeepCoNN, NARRE, and CARL) are usually superior to traditional methods. The reason is that deep learning techniques can nonlinearly extract the contextual features of review. Moreover, the dropout operation in deep learning-based methods can enhance recommender performance potentially and avoid the overfitting issue.

Third, the two proposed methods achieve the lowest values at the MAE and RMSE metrics in Table II. The major reason is that a review text can also reflect the parts of user rating behaviors, which is not considered in these comparison methods. Specifically, compared with the PMF method, EDMF achieves 6.2%–24.5% improvement in RMSE and 10.3%–30.44% in MAE. For the attention mechanism, we further conducted ablation experiments to remove the attention mechanism (called EDMFWA) while ensuring that the other conditions of the EDMF model remain unchanged. The experiment shows that the attention mechanism can more effectively extract features in the review. Moreover, EDMF+ obtains better prediction performance than EDMF at the aspects of MAE and RMSE. The values are improved by 0.2%–0.6% by the EDMF+ method because it utilizes the length information of reviews. Therefore, the proposed models outperform all the baseline methods consistently.

2) Training Efficiency Analysis of EDMF+: The training time of the four CNNs-based methods are compared. In Table III, EDMF achieves 70.8%/16.2%/16.5%/71.7% improvement of time in the training of the four public datasets than DeepCoNN, respectively. DeepCoNN, NARRE, and CARL use the aggregated review text to model user/item feature vectors. This way will cause the input data to be very large in the training process.

TABLE III
TRAINING TIME IN SECONDS FOR ONE EPOCH BY FOUR DEEP MODELS

	Automotive	Movies _and_TV	Video _Games	Yelp_2018
DeepCoNN [18]	7.2	82.3	28.4	572.6
NARRE [19]	15.6	181.0	60.8	1245.8
CARL [20]	9.7	110.2	36.6	751.0
EDMF	**2.1**	**69.0**	**23.7**	**161.9**
$\Delta\%$	70.8%	16.2%	16.5%	71.7%

Note: $\Delta\%$ denotes the improvement of EDMF over the best baseline performer.

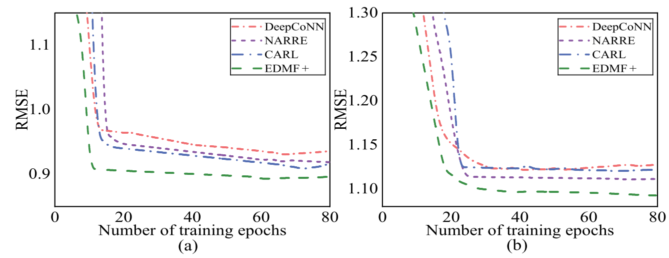


Fig. 4. Comparison of RMSE values by four different methods with the increasing number of training epochs. (a) Automotive. (b) Yelp_2018.

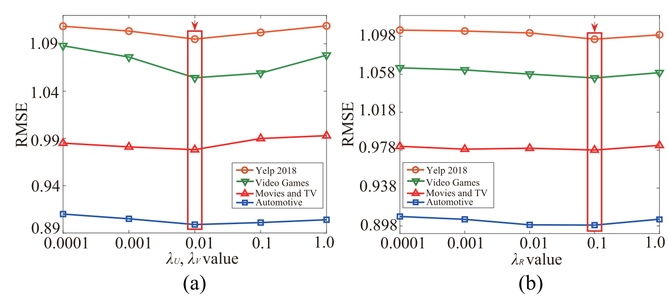


Fig. 5. Regularization parameter discussion in the proposed EDMF+ method. The robustness of the parameters can be verified on different datasets. (a) Parameters λ_U and λ_V. (b) Parameter λ_R.

The training time of NARRE lags behind that of DeepCoNN because the attention mechanism increases the training burden. Similarly, CARL leveraged factorization machines to construct the representations of item and user features, which enhances the training-time cost. Different from these methods, EDMF and EDMF+ models are trained on a onefold review at a time, thereby reducing the scale of parameters and enhancing training efficiency prominently. Experimental results prove that EDMF+ achieves better performance on training efficiency when compared with DeepCoNN, NARRE, and CARL methods. In Fig. 4, we have plotted the training process of these methods with increasing epoch number. The EDMF+ method (green dotted line in Fig. 4) achieves high converge speed and stability. Moreover, the lowest RMSE values illustrate that the review features can be extracted by the proposed network in Fig. 3.

3) Sensitivity Analysis of Parameters: To reveal the effect of rating and review information on model performance, we execute some comparison experiments, in which parameters λ_U, λ_V, and λ_R are set in [0.0001, 0.001, 0.01, 0.1, 1]. The related experimental results are demonstrated in Fig. 5, which

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

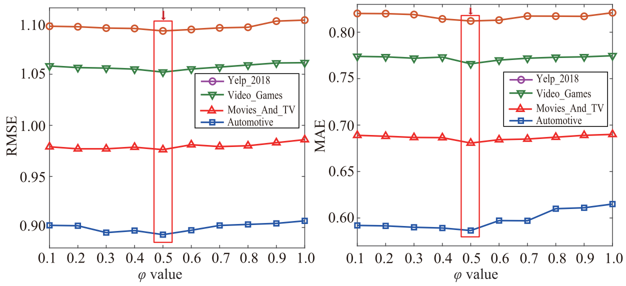


Fig. 6. Different datasets affect the dropout φ setting, which increases from 0.1 to 1. The parameter φ in proposed EDMF+ is robust (φ=0.5) to the changing datasets.

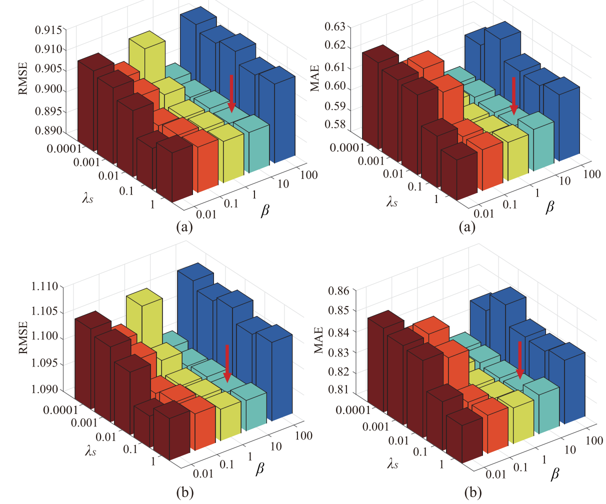


Fig. 7. Sparsity analysis of λ_S and β on (a) automotive, (b) Yelp_2018. The best parameters are indicated by red arrows.

shows the impact of parameters λ_U, λ_V, and λ_R on the four real-world datasets. According to the experience in [24], λ_U is set equal to λ_V. With the increase of parameters λ_U, λ_V, and λ_R, the best values are marked by the red rectangles in Fig. 5. The model obtains the lowest RMSE value when λ_R is equal to 0.1. Moreover, the models are robust on the different datasets because the models can converge with the same parameter setting. The suitable parameters λ_U, λ_V, and λ_R can balance the importance of ratings and review texts. The EDMF+ method achieves the highest prediction accuracy when λ_U and λ_V are equal to 0.01 and λ_R is equal to 0.1 on all four datasets. Thus, λ_U, λ_V, and λ_R can be set at abovementioned values in real scenarios. Moreover, k is robust on different datasets when k is equal to 10 on all four datasets.

The dropout ratio φ plays a vital role in improving the generalization ability of the proposed EDMF+. The experiments are executed on four public datasets with the changing dropout ratio φ. In Fig. 6, we draw the RMSE and MAE values by performing the EDMF+ method. The experimental results demonstrate that the optimal dropout ratio is 0.5 on four datasets. Therefore, dropout can improve the prediction performance effectively, and the dropout ratio φ is suggested to be 0.5 in real scenarios.

4) Sparsity Constraint Analysis: λ_S is a weight that directly controls the significance of $C(s_{ij})$, which is in fact a sparsity parameter. A large λ_S enables the result to have a strong sparsity constraint in the review feature vector. To demonstrate the influence of sparsity with L_0 norm on the model performance, an additional experiment, which reveals the relevance between model performance and the sparsity constraint, is designed. The parameter λ_S is tested in [0.0001, 0.001, 0.01, 0.1, 1] and β is tested in [0.01, 0.1, 1, 10, 100]. The different parameters of MAE and RMSE values of λ_S and β are drawn on histograms (see Fig. 7). The EDMF+ can obtain the best value when λ_S is equal to 0.1 and β is equal to 10, thereby validating that user review could only partially reflect the user rating behavior (i.e., the review is a sparsity constraint). The sparsity of review information is an important characteristic that can improve the performance of the EDMF+ method effectively.

D. Industrial Applications

In general, recommendation techniques can be utilized in many tasks. In route recommendations [32], the system can use the user's ratings and reviews of various attractions and stores to mine the user's preferences and plan possible points of interest for the user. It can also be combined with autonomous driving to provide precise route recommendations for users. The emergence of e-commerce platforms [4], [33] has given users a diversity of choices, and the review usually contains rich information about the preferences of the user and the characteristics of the items. Review-based recommendation techniques can be very good at mining user preferences to provide personalized recommendation services. Even in power systems [34], recommendations will play an important role. Electricity-based recommendation systems can learn how to use energy-efficient appliances from the experience of a large number of residential users (e.g., electricity plan report forms, usage times of household appliances), and recommend appropriate appliance usage plans to users while considering their lifestyle habits. Therefore, recommendation systems are being utilized widely in industrial applications and bring convenience to people's lives.

VI. Conclusion

In this article, we presented an EDMF with review feature learning for the industrial recommender system. First, two characteristics of the review were revealed. Through our observation, the reviews that contained interactive information between users and items could only reflect user rating behavior partially. L_0 norm was used to describe the review feature. The loss function was constructed by MAP. Then, the alternative minimization algorithm was introduced to solve the loss function. Furthermore, the length information of the review text, which was called EDMF+, was also considered to enhance recommendation ability. Experimental results on the four real-world datasets demonstrated that our methods have advantages in terms of prediction accuracy even when compared with state-of-the-art models. In the future, because the usefulness of the review is also an aspect that affects the accuracy of the recommendation system, we will introduce the latest generative adversarial networks to detect and neglect the abnormal reviews.

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

References

[1] H. Zhang, Y. Ji, J. Li, and Y. Ye, "A triple wing harmonium model for movie recommendation," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 231–239, Feb. 2016.

[2] D. Wang, X. Zhang, D. Yu, G. Xu, and S. Deng, "CAME: Content- and context-aware music embedding for recommendation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 1375–1388, Mar. 2021.

[3] A. F. Biten, L. Gomez, M. Rusiñol, and D. Karatzas, "Good news, everyone! Context driven entity-aware captioning for news images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12458–12467.

[4] W. Zhao, S. Li, Y. He, E. Y. Chang, J. Wen, and X. Li, "Connecting social media to e-commerce: Cold-start product recommendation using microblogging information," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 5, pp. 1147–1159, May 2016.

[5] X. Shen, B. Yi, Z. Zhang, J. Shu, and H. Liu, "Automatic recommendation technology for learning resources with convolutional neural network," in *Proc. Int. Symp. Educ. Technol.*, 2016, pp. 30–34.

[6] H. X. Hu, B. Tang, Y. Zhang, and W. Wang, "Vehicular ad hoc network representation learning for recommendations in Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2583–2591, Apr. 2020.

[7] Y. Koren, R. M. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Comput.*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[8] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.*, vol. 2009, 2009, Art. no. 421425.

[9] X. Luo, M. Zhou, S. Li, L. Hu, and M. Shang, "Non-negativity constrained missing data estimation for high-dimensional and sparse matrices from industrial applications," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1844–1855, May 2020.

[10] D. Li, H. Liu, Z. Zhang, K. Lin, and N. N. Xiong, "CARM: Confidence-aware recommender model via review representation learning and historical rating behavior," *Neurocomputing*, vol. 455, pp. 283–296, 2021.

[11] G. Ling, M. R. Lyu, and I. King, "Ratings meet reviews, a combined approach to recommend," in *Proc. 8th ACM Conf. Recommender Syst.*, 2014, pp. 105–112.

[12] C. Wang and D. M. Blei, "Collaborative topic modeling for recommending scientific articles," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2011, pp. 448–456.

[13] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: Understanding rating dimensions with review text," in *Proc. 7th ACM Conf. Recommender Syst.*, 2013, pp. 165–172.

[14] S. Deng, L. Huang, G. Xu, X. Wu, and Z. Wu, "On deep learning for trust-aware recommendations in social networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 5, pp. 1164–1177, May 2017.

[15] X. Shen *et al.*, "Deep variational matrix factorization with knowledge embedding for recommendation system," *IEEE Trans. Knowl. Data Eng.*, vol. 33, no. 5, pp. 1906–1918, May 2021.

[16] B. Yi *et al.*, "Deep matrix factorization with implicit feedback embedding for recommendation system," *IEEE Trans. Ind. Informat.*, vol. 15, no. 8, pp. 4591–4601, Aug. 2019.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. 25th Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[18] L. Zheng, V. Noroozi, and P. S. Yu, "Joint deep modeling of users and items using reviews for recommendation," in *Proc. 20th ACM Int. Conf. Web Search Data Mining*, 2017, pp. 425–434.

[19] C. Chen, M. Zhang, Y. Liu, and S. Ma, "Neural attentional rating regression with review-level explanations," in *Proc. World Wide Web Conf.*, 2018, pp. 1583–1592.

[20] L. Wu, C. Quan, C. Li, Q. Wang, B. Zheng, and X. Luo, "A context-aware user-item representation learning for item recommendation," *ACM Trans. Inf. Syst.*, vol. 37, no. 2, pp. 1–29, 2019.

[21] D. Lian *et al.*, "Scalable content-aware collaborative filtering for location recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 6, pp. 1122–1135, Jun. 2018.

[22] W. Wang, J. Chen, J. Wang, J. Chen, J. Liu, and Z. Gong, "Trust-enhanced collaborative filtering for personalized point of interests recommendation," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6124–6132, Sep. 2020.

[23] Y. Tan, M. Zhang, Y. Liu, and S. Ma, "Rating-boosted latent topics: Understanding users and items with ratings and reviews," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2640–2646.

[24] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1257–1264.

[25] J. Pan, Z. Hu, Z. Su, and M. Yang, "L_0-regularized intensity and gradient prior for deblurring text images and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 2, pp. 342–355, Feb. 2017.

[26] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," in *Proc. 2011 SIGGRAPH Asia Conf.*, 2011, pp. 1–12.

[27] S. Zhang, L. Jiao, F. Liu, and S. Wang, "Global low-rank image restoration with Gaussian mixture model," *IEEE Trans. Cybern.*, vol. 48, no. 6, pp. 1827–1838, Jun. 2018.

[28] P. C. Hansen, "Analysis of discrete ill-posed problems by means of the L-curve," *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.

[29] G. Golub, MichaelHeath, and GraceWahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[30] H. W. Engl, "Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates," *J. Optim. Theory Appl.*, vol. 52, no. 2, pp. 209–215, 1987.

[31] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operat. Syst. Des. Implementation*, 2016, pp. 265–283.

[32] W. Tu *et al.*, "Real-time route recommendations for E-taxies leveraging GPS trajectories," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3133–3142, May 2021.

[33] C. Wang, Z. Deng, J. Lai, and P. S. Yu, "Serendipitous recommendation in E-commerce using innovator-based collaborative filtering," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2678–2692, Jul. 2019.

[34] F. Luo, G. Ranzi, W. Kong, G. Liang, and Z. Y. Dong, "Personalized residential energy usage recommendation system based on load monitoring and collaborative filtering," *IEEE Trans. Ind. Informat.*, vol. 17, no. 2, pp. 1253–1262, Feb. 2021.

Hai Liu (Senior Member, IEEE) received the M.S. degree in applied mathematics and the Ph. D. degree in pattern recognition and artificial intelligence from the Huazhong University of Science and Technology, Wuhan, China, in 2010 and 2014, respectively.

Since June 2017, he has been an Assistant Professor with the National Engineering Research Center for E-Learning, Central China Normal University, Wuhan. From 2017 to 2019, he was a Postdoctoral Fellow with the Department of Mechanical Engineering, City University of Hong Kong, Kowloon, Hong Kong, where he was hosted by the Professor You-Fu Li. In 2020, he was selected as "China- European Commission Talent Programme" under National Natural Science Foundation of China (NSFC). He is a Senior Researcher with UCL Interaction Centre, University College London, London, U.K., where he will host by the Professor Sriram Subramanian; he will hold the position one year till February 2022. He has authored more than 70 peer-reviewed articles in international journals from multiple domains. More than six articles are selected as the ESI highly cited articles. His research interests include deep learning, artificial intelligence, recommendation system, head pose estimation, gaze estimation, educational technology, and pattern recognition.

Dr. Liu was the recipient of the first prize of Science and Technology Progress Award by the Hubei Province of China in 2020. He has frequently beena Reviewer for more than six international journals, including the IEEE Translations on Industrial Informatics, IEEE Translations on Cybernetics, IEEE/ASME Transactions on Mechatronics, and IEEE Translations on Instrumentation and Measurement. He is also a Communication Evaluation Expert for the NSFC from 2016 to present.

Chao Zheng received the bachelor's degree in computer science from the Huazhong University of Science and Technology, Wuhan, China, in 2017. He is currently working toward the master's degree in computer science and technology with the National Engineering Laboratory For Educational Big Data, Wuhan.

His research interests include deep learning, natural language processing, recommendation systems, and data mining techniques.

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.

Duantengchuan Li is currently working toward the Ph.D. degree in software engineering with the School of Computer Science, Wuhan University, Wuhan, China.

His research interests include recommendation system, representation learning, natural language processing, intention recognition, and their applications in services computing and software engineering.

Xiaoxuan Shen received the B.E. degree in mathematics and applied mathematics, in 2013.

He received the M.E. degree in education technology, 2017. He received the Ph.D. degree in education information technology from Central China Normal University (CCNU), Wuhan, China, in 2020.

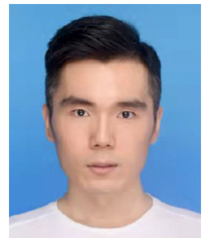
He is currently working as a Postdoctor with the National Engineering Laboratory for Educational Big Data, CCNU. His research interests include deep learning, representation learning, and their applications in recommendation system and intelligent e-learning environment.

Ke Lin is currently working toward the Ph.D. degree in control science and engineering with the Harbin Institute of Technology (Shenzhen), Shenzhen, China.

His research interests include deep reinforcement learning, machine learning, and their applications in recommendation system.

Jiazhang Wang is currently working toward the Ph.D. degree in computational photography with the ECE Department, Northwestern University, Evanston, IL, USA.

His research interests include pattern recognition, recommendation system, representation learning, computational photography, 3-D metrology, and machine vision.



Zhen Zhang received the M.S. degree in business administration and the Ph.D. degree in information management from Central China Normal University, Wuhan, China, in 2012 and 2018, respectively.

He is currently working with the School of Sports Engineering and Information Technology, Wuhan Sports University, Wuhan, as an Enrolled Lecturer. His research interests include recommendation system, representation learning, knowledge graph, and their applications in education and intelligent industry.

Zhaoli Zhang (Member, IEEE) received the M.S. degree from Central China Normal University, Wuhan, China, in 2004, and the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, in 2008, both in computer science.

He is currently a Professor with the National Engineering Research Center For E-Learning, Central China Normal University. His research interests include human–computer interaction, deep learning, signal processing, knowledge services, and software engineering.

Dr. Zhang is a Member of CCF (China Computer Federation).

Neal N. Xiong (Fellow, IEEE) received the Ph.D. degrees in sensor system engineering from Wuhan University, Wuhan, China, and in dependable sensor networks from the Japan Advanced Institute of Science and Technology, Nomi, Japan, in 2007.

He is currently a Professor with the Department of Mathematics and Computer Science, Northeastern State University, Tahlequah, OK, USA. His research interests include cloud computing, machine learning, representation learning, and optimization theory.

Dr. Xiong is currently the Editor-in-Chief, an Associate Editor, or an Editor member for more than ten international journals, including to be an Associate Editor for the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS AND INFORMATION SCIENCES.

Authorized licensed use limited to: Wuhan University. Downloaded on April 20,2022 at 01:03:33 UTC from IEEE Xplore. Restrictions apply.