A Recommendation Algorithm For Collaborative Denoising Auto-Encoders Based On User Preference Diffusion

Xiu Wang, Xuejun Liu* and Xinyan Xu
College of Computer Science and Technology
Nanjing Tech University
Nanjing 211816, China
lxj_njgd@163.com

Abstract—**The main challenge of collaborative filtering is to use a relatively less effective score to get an accurate prediction. It is an effective method to incorporate user and item information with collaborative filtering. In this paper, we present a new method. Firstly, the heterogeneous information network is used to combine more information of users and items. Under the corresponding user interest semantic assumptions, the observed user implicit feedback extends the user's preference items along different meta-paths. And then combined with the Collaborative Denoising Auto-Encoder (CDAE) for top-N recommendation. Denoising Auto-Encoders learns latent representation of corrupted user-item preferences that can best reconstruct the full input. Experiments show that the proposed algorithm can effectively alleviate the sparse data and improve the performance in terms of recommendation accuracy.**

Keywords—Collaborative Filtering, heterogeneous information network, Denoising Auto-Encoders, meta-paths, sparse data.

I. INTRODUCTION

Collaborative filtering (CF) which make use of historical interactions or preferences has achieved significant success[1]. However, in practice, a user usually only interests in a small part of all the items, so the user-item interactions are rather sparse. In addition, the CF method cannot recommend new items because these items have never received any user feedback in the past. According to recent research[2], better recommendation results can be achieved by combination with user feedback data and additional information about users or items. The emergence of heterogeneous information network[3] provides a new idea and information integration framework for the integration of a variety of attributes and attributes of the recommended combination of technology.

In order to exploit relational heterogeneity of information network, we first diffuse the observed user's implicit feedback along the different meta-paths based on the corresponding user's interest semantic assumptions to produce possible user preferences. And then combined with the Collaborative Denoising Auto-Encoder (CDAE)[4] for top-N recommendation. The CDAE assumes that whatever user-item interactions observed are a corrupted version of the user preference. The model learns latent representations of corrupted user-item preferences, and can best reconstruct the complete input. During the training process, the user's item set is provided to the model and the training model to recover the entire item set. At prediction time, the existing preference set as input, the model will recommend new items to the user.

In this paper, A Recommendation Algorithm for Collaborative Denoising Auto-Encoders Based On User Preference Diffusion (UPD-CDAE) is proposed to solve the problem of top- N recommended problems, while avoiding the sparseness of data and the scalability of the algorithm. The rest of this paper is organized as follows. In the second section, we introduce the user preference diffusion based on the meta-path. The third section introduces the Collaborative Denoising Auto-encoder based on the user preference diffusion. The fourth section uses the experiment to verify the validity and efficiency of the new method.

II. THE USER PREFERENCE DIFFUSION BASED ON THE META-PATH

A. Basic definition

DEFINITION 1. An information network is defined as a directed graph $G=(V,E)$ with an object type mapping function $\phi : V \rightarrow \mathcal{A}$ and a link type mapping function $\varphi : E \rightarrow \mathcal{R}$, where each object $v \in V$ belongs to one particular object type $\phi(v) \in \mathcal{A}$, and each link $l \in E$ belongs to a relation type $\varphi(l) \in \mathcal{R}$ .When $|\mathcal{A}| > 1$ or $|\mathcal{R}| > 1$, the network is called heterogeneous information network; Otherwise known as homogeneous information network.

DEFINITION 2. The network schema is a meta template for a heterogeneous network $G=(V,E)$, denoted as $G_T = (\mathcal{A}, \mathcal{R})$. It is a directed graph defined over object and link types.

DEFINITION 3. A meta path $\mathcal{P}= A_0 \xrightarrow{R_1} A_1 \xrightarrow{R_2} \cdots \xrightarrow{R_k} A_k$ is a path defined in a network schema $G_T = (\mathcal{A}, \mathcal{R})$, which defines a composite relation $R_1 R_2 \cdots R_k$ between type A_0 and A_k. Where

978-1-5386-0497-7/17/$31.00 ©2017 IEEE

$A_0 = dom(R_1) = dom(\mathcal{P})$, $A_k = range(R_k) = range(\mathcal{P})$ and $A_i = range(R_i) = dom(R_{i+1})$. Where $dom(\cdot)$ denotes the domain of certain relationship, $range(\cdot)$ denotes the range.

B. Binary User Feedback

With m users $\mathcal{U} = \{u_1, u_2, \cdots, u_m\}$ and n items $\mathcal{I} = \{e_1, e_2, \cdots, e_n\}$, the user implicit feedback matrix $R \in \mathbb{R}^{m \times n}$ defined as follows:

$$R_{u_i,j} = \begin{cases} 1, & \text{if } (u_i, e_j) \text{ interaction is observed;} \\ 0, & \text{otherwise.} \end{cases}$$

The value 1 in the implicit feedback indicates that users are more interested in the items than the rest of the items. User preferences represent the user's interest in implicit feedback data. Intuitively, if we can understand the semantics meanings of user preferences and find similar items that the users were interested in, according to the semantics found, we can make recommendations to these users accordingly.

C. User Preference Diffusion

We want to diffuse the users preferences observed in the implicit feedback data through different meta-paths so that users can be associated with other items. By defining a user preference diffusion score between the target user and all possible items along different meta-paths, we can now measure the likelihood of unobserved user-item interactions in the information network under different semantic assumptions.

Given a math-path $\mathcal{P} = R_1 R_2 \cdots R_k$ with $dom(\mathcal{P}) = \text{user}$ and $range(\mathcal{P}) = \text{item}$, let $\mathcal{P}' = R_2 \cdots R_k$ with $dom(\mathcal{P}') = \text{item}$ and $range(\mathcal{P}') = range(\mathcal{P}) = \text{item}$. By extending the PathSim metric proposed in [5], the user preferences diffusion score for users u_i and items j along $\mathcal{P}$ are defined as follows:

$$s(u_i, e_j \mid \mathcal{P}) = \sum_{e \in I} \frac{2 \times R_{u_i,e} \times \left| \{p_{e \rightsquigarrow e_j} : p_{e \rightsquigarrow e_j} \in \mathcal{P}'\} \right|}{\left| \{p_{e \rightsquigarrow e} : p_{e \rightsquigarrow e} \in \mathcal{P}'\} \right| + \left| \{p_{e_j \rightsquigarrow e_j} : p_{e_j \rightsquigarrow e_j} \in \mathcal{P}'\} \right|} \tag{1}$$

Where $p_{e \rightsquigarrow e_j}$ is a path between e and e_j , $p_{e \rightsquigarrow e}$ is that between e and e, and $p_{e_j \rightsquigarrow e_j}$ is that between e_j and e_j .

By calculating the diffusion scores between all users and all items along the meta-path $\mathcal{P}$, we can get a diffused user preference matrix $\overline{R} \in \mathbb{R}^{m \times n}$, $\overline{R}_{u_i,j}$ represents the user u_j possible preference for the item j .

The process simulates the user information discovery by propagating user preferences along the different meta-paths in the heterogeneous information network. Diffusion scoring shows that the possibility of certain user variables interacting under certain semantic semantics can effectively alleviate sparse data and cold start problems.

III. Collaborative Denoising Auto-encoder Based On The User Preference Diffusion

A. Collaborative Denoising Auto-Encoder

Collaborative Denoising Auto-Encoder (CDAE) differs from traditional Denoising Auto-Encoder in that the input also encodes a latent vector for the user to make the recommended model more accurate. CDAE structure is shown in Figure 1.

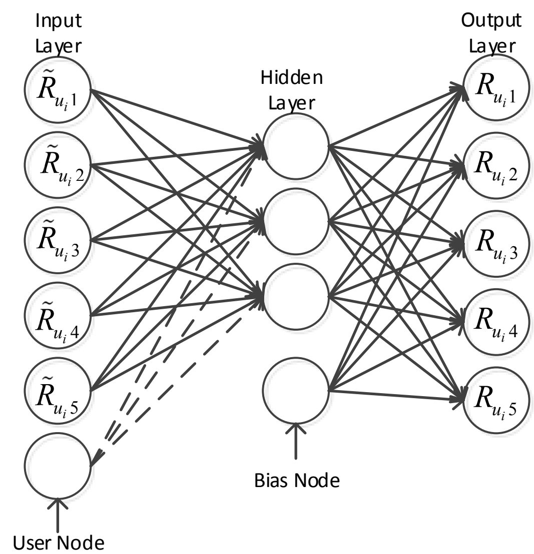

Figure 1. A sample structure of CDAE

In the input layer, there are $I+1$ nodes, where each of the I nodes corresponds to an item, and the last node is the user node, which means that the node and its associated weights are unique to each user u_i in the data. Given the user's historical feedback set for the item set, since a user usually only expresses a small interest in all the items, the user behavior vector is sparse. We obtain the user preference diffusion set $\mathcal{O}'$ by the formula (1), we can convert $\mathcal{O}'$ into a training set containing m instance $\{\overline{R}_{u_1}, \overline{R}_{u_2}, \cdots, \overline{R}_{u_m}\}$, where $\overline{R}_{u_i} = \{\overline{R}_{u_i1}, \overline{R}_{u_i2}, \cdots \overline{R}_{u_iI}\}$ is the I dimensionally extended feedback vector for user u_i on all the items in $\mathcal{I}$. $\mathcal{O}'_{u_i}$ will be expressed as the user's preference set.

There are K nodes in the hidden layer. We use $W \in \mathbb{R}^{I \times K}$ to represent the weight matrix between the item input nodes and the nodes in the hidden layer, and $V_{u_i} \in \mathbb{R}^K$ represent the weight vector of the user input node. V_{u_i} is a user-specific vector, that is, for each user, we have a unique vector. On the other hand, W_j and V_{u_i} can be seen as a distributed representation of the item j and the user u_i . Random noise is added to the user's preference diffusion data to get corrupted user preferences $\tilde{R}_{u_i}$. CDAE first maps the input to a latent representation z_{u_i} , calculated as follows:

$$z_{u_i} = h(W\tilde{R}_{u_i} + V_{u_i} + b) \quad (2)$$

Where $h(\cdot)$ is a mapping function(e.g. identity function or sigmoid function), $b \in \mathbb{R}^K$ is the offset vector.

At the output layer, there are I nodes representing the reconstructed input vector $\overline{R}_{u_i}$. The weight matrix is expressed as $W' \in \mathbb{R}^{I\times K}$. We use a bias node $b \in \mathcal{R}^I$ that represents the weight vector in the hidden layer. The latent representation is then mapped back to the original input space to reconstruct the input vector. The output value is:

$$\hat{R}_{u_i j} = f(W_j' z_{u_i} + b_j') \quad (3)$$

Where b' the offset vector of the output layer. $f(\cdot)$ is also a mapping function. We learn CDAE parameters by minimizing average reconstruction errors:

$$\underset{W,W',V,b,b'}{\arg\min} \frac{1}{U} \sum_{u=1}^{U} \ell(\overline{R}_{u_i}, \hat{R}_{u_i}) + \mathcal{R}(W, W', V, b, b') \quad (4)$$

$\mathcal{R}(\cdot)$ is a regularization term to control the complexity of the model. We use the square $L2$ Norm in this paper.

$$\mathcal{R}(\cdot) = \frac{\lambda}{2}(\|W\|_2^2 + \|W'\|_2^2 + \|V\|_2^2 + \|b\|_2^2 + \|b'\|_2^2) \quad (5)$$

We use Stochastic Gradient Descent (SGD) to lean the parameters. AdaGrad adaptive step is used during the learning process. The update formula is as follows:

$$\theta^{(t+1)} = \theta^{(t)} - \frac{\eta g_\theta^{(t)}}{\sqrt{\beta + \sum_{s=1}^{t} g_\theta^{(s)}}} \quad (6)$$

Here $\theta^{(t)}$ is the value of the parameter θ in the tSGD step, $g_\theta^{(t)}$ is the gradient at step t.

B. *Recommendation*

In the prediction phase, the predicted value of the item is not predicted by the correlation weight obtained by the training and the user's existing preference set (without corruption). And then construct the recommended list, select the top-N items of the prediction values as recommended items to the user.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, the experimental data using the Movie Lens 10M (ML) data set. For each data set, we will convert those ratings which are no less than 4 into 1, and set the other ratings as 0. We iterated to delete fewer than 5 ratings for users and items. For each user, we randomly selected 20% of the ratings as the test set, and the remaining 80% of the ratings as the training set.

We use Stochastic Gradient Descent (SGD) to learn the parameters for the proposed method. We set β=1 and try different steps $\eta \in \{1, 0.1, 0.01, 0.001\}$ and give the best results for each model. In addition, the hidden layer dimension K used to learn the potential feature is set to 50.

Given a top-N recommended list $C_{N,rec}$, the accuracy and recall are defined as follows:

$$\text{Presion@N} = \frac{|C_{N,rec} \cap C_{adopted}|}{N}$$

$$\text{Recall@N} = \frac{|C_{N,rec} \cap C_{adopted}|}{|C_{adopted}|}$$

Where $C_{adopted}$ means the items that used by the user in the test dataset.

A. *Comparisons with other recommendation methods*

In order to evaluate the performance of our algorithm, we compare it with FISM[6] and BPR[7]. For all comparison algorithms, the parameters are carefully selected by cross validation to ensure fair comparisons. The results are as follows:

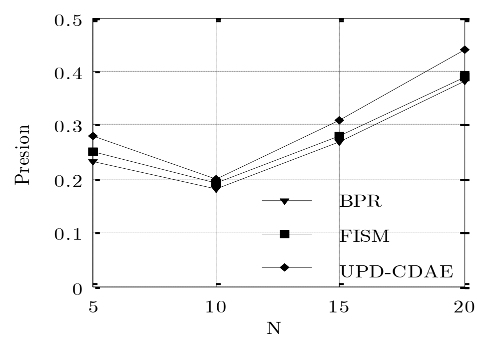

Figure 2. Precision comparison of different recommendation algorithms on ML Data

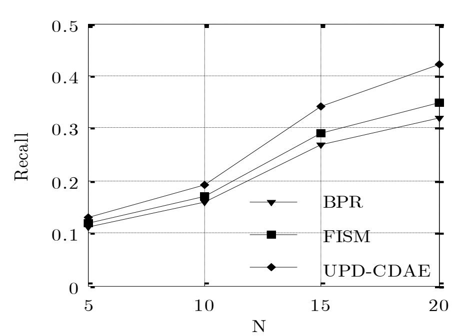

Figure3. Recall comparison of different recommendation algorithms on ML Data

Figure 2 shows the accuracy of the algorithms. The abscissa N represents the number of top-N and the ordinate indicates the precision. Figure 3 shows the recall for each algorithm, and the ordinate represents the recall. It can be seen from the figures, the UPD-CDAE algorithm we proposed always better than other methods.

B. The Effect of Data Sparse Degree on the Algorithm

In this experiment, 20%, 40%, 60% and 80% rating data is selected as the training set, and the remaining data as the test set. We take top-10 as an experimental test. The experimental results are shown in table I.

From the table can be found, with the training data set the proportion of the entire data set continues to improve, these algorithms recommended are constantly improved. The overall FISM method in the table outperform the BPR method. The UPD-CDAE method is more accurate than the FISM and BPE methods. This is because, in the case of sparse data, the UPD-CDAE method will observe the user's implicit feedback along different meta-paths to produce possible user preferences, mitigating sparse data problems, thereby increasing the recommended accuracy.

TABLE I. COMPARISON OF DIFFERENT SPARSENESS

Methd	Precision				Recall			
	20%	40%	60%	80%	20%	40%	60%	80%
BPR	0.1123	0.1336	0.1452	0.1683	0.1539	0.1627	0.1703	0.1779
FISM	0.1331	0.1412	0.1644	0.1756	0.1612	0.1715	0.1789	0.1837
UPD-CDAE	0.1458	0.1632	0.1749	0.1913	0.1831	0.1894	0.1936	0.1951

V. CONCLUSION

In this paper, we proposed a Recommendation Algorithm For Collaborative Denoising Auto-Encoders Based On User Preference Diffusion (UPD-CDAE). In the traditional collaborative filtering algorithm, there is data sparse motion problem. we first diffuse the observed user's implicit feedback along the different meta-paths based on the corresponding user's interest semantic assumptions to produce possible user preferences. Then we combine with the Collaborative Denoising Auto-Encoder for top-N recommendation. Through a series of experiments we show that compared with other recommended algorithms, the method further improve the recommended results, and can effectively alleviate the problem of sparse data. In the future work, we will consider combination with deep neural network structure- Such as convolution neural networks[8] and cyclic neural networks[9].

ACKNOWLEDGMENT

This work was supported by The National Natural Science Foundation of China(61203072); Primary Research & Developement Plan of Jiangsu Province(BE2015697).

REFERENCES

[1] Loepp B, Hussein T, Ziegler J, et al. Choice-based preference elicitation for collaborative filtering recommender systems[C]. ACM International Conference on Human Factors in Computing Systems. ACM, 2014:3085-3094.

[2] Shi Y, Larson M, Hanjalic A. Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges[M]. ACM, 2014 ,47 (1) :1-45.

[3] Yu X, Ren X, Sun Y, et al. Personalized entity recommendation: a heterogeneous information network approach[C]. ACM International Conference on Web Search and Data Mining. ACM, 2014:283-292.

[4] Wu Y, Dubois C, Zheng A X, et al. Collaborative Denoising Auto-Encoders for Top-N Recommender Systems[C]. ACM International Conference on Web Search and Data Mining. ACM, 2016:153-162.

[5] Sun Y, Han J, Yan X, et al. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks[J]. Proceedings of the Vldb Endowment, 2011, 4(11):992-1003.

[6] Kabbur S, Ning X, Karypis G. FISM: factored item similarity models for top-N recommender systems[C]. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2013:659-667.

[7] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[C]. Conference on Uncertainty in Artificial Intelligence. AUAI Press, 2009:452-461.

[8] Shen X, Yi B, Zhang Z, et al. Automatic Recommendation Technology for Learning Resources with Convolutional Neural Network[C]. International Symposium on Educational Technology. 2016:30-34.

[9] Wang H, Shi X, Yeung D Y. Collaborative Recurrent Autoencoder: Recommend while Learning to Fill in the Blanks[J]. 2016.