IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DATA-ENABLED INTELLIGENCE FOR DIGITAL HEALTH

Received August 6, 2019, accepted August 30, 2019, date of publication September 11, 2019, date of current version September 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940644

A Medical-History-Based Potential Disease Prediction Algorithm

WENXING HONG[1], ZIANG XIONG[1], NANNAN ZHENG[1], AND YANG WENG[2]
[1]School of Aerospace Engineering, Xiamen University, Xiamen 361005, China
[2]College of Mathematics, Sichuan University, Chengdu 610064, China

Corresponding author: Wenxing Hong (hwx@xmu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0830300, in part by the Science and Technology Program of Fujian, China, under Grant 2018H0035, and in part by the Science and Technology Program of Xiamen, China, under Grant 3502Z20183011.

ABSTRACT As an important application of medical informatization, healthcare big data analysis has been extensively researched in the fields of intelligent consultation, disease diagnosis, intelligent question-answering doctors, and medical assistant decision support, and have made many achievements. In order to improve the comprehensiveness and pertinence of the medical examination, this paper intends to use healthcare big data analysis combined with deep learning technology to provide patients with potential diseases which is usually neglected for lacking of professional knowledge, so that patients can do targeted medical examinations to prevent health condition from getting worse. Inspired by the existing recommendation methods, this paper proposes a novel deep-learning-based hybrid recommendation algorithm, which is called medical-history-based potential disease prediction algorithm. The algorithm predicts the patient's possible disease based on the patient's medical history, providing a reference to patients and doctors to reduce the problem of delaying treatment due to unclear description of the symptom or limited professional knowledge. The experimental results show that our approach improves the accuracy of the potential diseases prediction.

INDEX TERMS Healthcare big data, deep learning, recommendation, disease predicting.

I. INTRODUCTION

With the rapid development of the internet, electronic medical information has become popular with many cities around the world, such as electronic medical records(EMR) to replace traditional paper medical records, online appointments, and online reports, thus accumulated large-scale of healthcare data. Healthcare big data covers a wide range of areas, and any data that is directly or indirectly related to healthcare falls into this category. The most common categories include: healthcare services data [1], [2], biomedical data [3]–[6], health insurance data, medical research and management data [7], public health data [8], [9], behavioral and emotional data, health statistics, population management data, and environmental data. Many approaches of these fields are developed upon machine learning [10], it is usually used to deal with classification, regression and feature extraction [11] problems. If healthcare big data can be timely acquired, properly stored and effectively analyzed, it is of great significance for improving lifestyle, diagnosis and treatment efficiency, medical service quality, accelerating drug research and development, and optimizing medical system. Potential disease prediction can help patient do targeted medical examinations by pointing the possible diseases, otherwise patients may miss significant medical examinations for lacking of professional knowledge, which will lead to severe health problems.

Inspired by the achievement of recommender system research, we argue that giving medical advice can be viewed as a kind of recommendation. We could dig the latent relations of diseases as the e-commercial product recommendation did, thus those the most possible diseases that the patient may have could be listed.

Traditional recommendation methods include collaborative filtering [12], content-based recommendations [13], hybrid recommendations [14]. Some of them have been adopted to industry. Among these methods above, CF is used mostly in candidate selection problem. The system predicts a score for each item, and items are ranked by the score, which can be viewed as the relevance between user and item (or patient and disease in this paper). With the development

The associate editor coordinating the review of this manuscript and approving it for publication was Yonghong Peng.

of deep neural networks (DNN), it provides a powerful tool which can be used to handle big data problems. DNN provides a novel way to model user-item relevance. Based on large amounts of data accumulated by the internet, we can extract features of users and items automatically through DNN. Some of these features are hard to find artificially and unexplainable.

Many previous collaborative filtering based methods only model second-order relations of item pairs or user pairs. Actually, take shopping on the internet as an example, multiple items with same or complementary usage are likely to be consumed together [15]. High-order relations exist in the real world scenario which have been verified in [16]. Second-order model does not indicate these high-order relations properly. Therefore, high-order relations have better performance than low-order in estimating user preference. And DNN is one feasible way to model high-order feature relations. Besides, DNN can handle large amounts of data with acceptable time and space consuming, and prevent overfitting with some special tricks.

In this paper, we construct our model with DNN to calculate the high-order relations among diseases. Like [17] and [15] indicated before, users' historical interactions are not weighted equally, as well to medical scenario, different symptom or disease has different relevance, so we adopt attention network to map the weight of each disease in the patient's medical history. With the weighted relevance among diseases, we first conduct a weighted-sum operation to model the patient's latent feature, which is viewed as the health condition of the patient, and then multiply with target disease embedding. After that we use DNN to model high-order relations among features. Inspired by DeepFM [18], we argue that low-order disease relations are not neglectable, considering both high-order and low-order relations seems to be more reasonable in the real world data. So we conduct Factorization Machines to model low-order disease relations. By combining both high-order and low-order relations, the model gives a score which means the possibility of the disease that the user may have. Ranked by the possibilities, we select top-N diseases as the final predicting list. We term this algorithm as medical-history-based potential disease prediction algorithm.

II. RELATED WORK

The key to get optimal performance of recommender system is modeling user's preference accurately, we regard it as the patient's health condition features in our research. In this section, we briefly review some state-of-the-art methods that related to our approach.

A. HYBRID RECOMMENDATION

As mentioned in section I, there are several traditional base models using different data sources and structures in recommender system research. Although these methods could recommend items based on user interests, the advantages of them vary from fields. CF utilizes the given relations among users or items to generate the recommendation; content-based method depends on the features of item and context, knowledge-based model reasons according to explicit knowledge domain models. These methods have advantages and shortcomings respectively. To achieve better performance, one feasible method is to construct a model that combines the advantages of these methods.

B. BASE MODEL

Deep Item-based Collaborative Filtering is introduced in [15] (short for DeepICF), it models high-order feature relations between user and item. It provides an end-to-end way to generate user and item representation, given the close relationship to our proposed model. We briefly review DeepICF in this subsection.

The architectures of the model that DeepICF based called Neural Attentive Item Similarity Model [17] (short for NAIS) and DeepICF [15] are shown in Figure 1. There are two parts of input. One is the user's historical interactions, which is multi-hot representation formed with items that the user has interacted (e.g., browsed or bought) before. The other is the target item that the user has not interacted yet, formed as one-hot representation. The items are mapped to dense vectors respectively through the embedding matrix, an element-wise product operation is conducted to extract relations of each pair of item embeddings. The items that user has interacted are not equally weighted in the real world scenario [17], so they use the attention network to calculate the weight of the items. To make a fixed size vector for further calculation in DNN, a weighted sum-pooling is demonstrated as follows,

$$f_{att}(V_{u,i}) = \frac{1}{|V_{u,i}|^{\alpha}} \sum_{v \in V_{u,i}} a(\mathbf{v}) \cdot \mathbf{v} = \frac{1}{\left(|R_u^+| - 1\right)^{\alpha}} \sum_{j \in R_u^+ \setminus i} a(q_j \odot p_i) \cdot (q_j \odot p_i), \quad (1)$$

where a($\mathbf{v}$) is the attention function that takes vector $\mathbf{v}$, which is the element-wise product of user interacted item j and target item i, as input, and outputs the relevance of $\mathbf{v}$ in the weighted average pooling. R_u^+ is the item set that the user u has interacted with. $R_u^+ \setminus i$ is to exclude the influence of the target item i in constructing u's profle to predict $y_{u,i}$. The structure of the attention network is illustrated in Figure 2. Both methods use a multi-layer perceptron with a hidden layer to parameterize the attention function,

$$a(\mathbf{v}) = softmax'(\mathbf{h}^T ReLU(\mathbf{W}\mathbf{v} + \mathbf{b})), \quad (2)$$

where $\mathbf{W}$ and $\mathbf{b}$ denote the weight matrix and bias of the attention network. $softmax'$ is a variant of the softmax function to normalize the weights, which is defined as,

$$softmax'(a(\mathbf{v})) = \frac{\exp a(\mathbf{v})}{[\sum_{\mathbf{v} \in V_{u,i}} \exp a(\mathbf{v})]^{\beta}}, \quad (3)$$

β is defined as a hyper-parameter to smooth the output of softmax function, we set β to 0.5 in our experiments. When accomplished those steps above, we get a

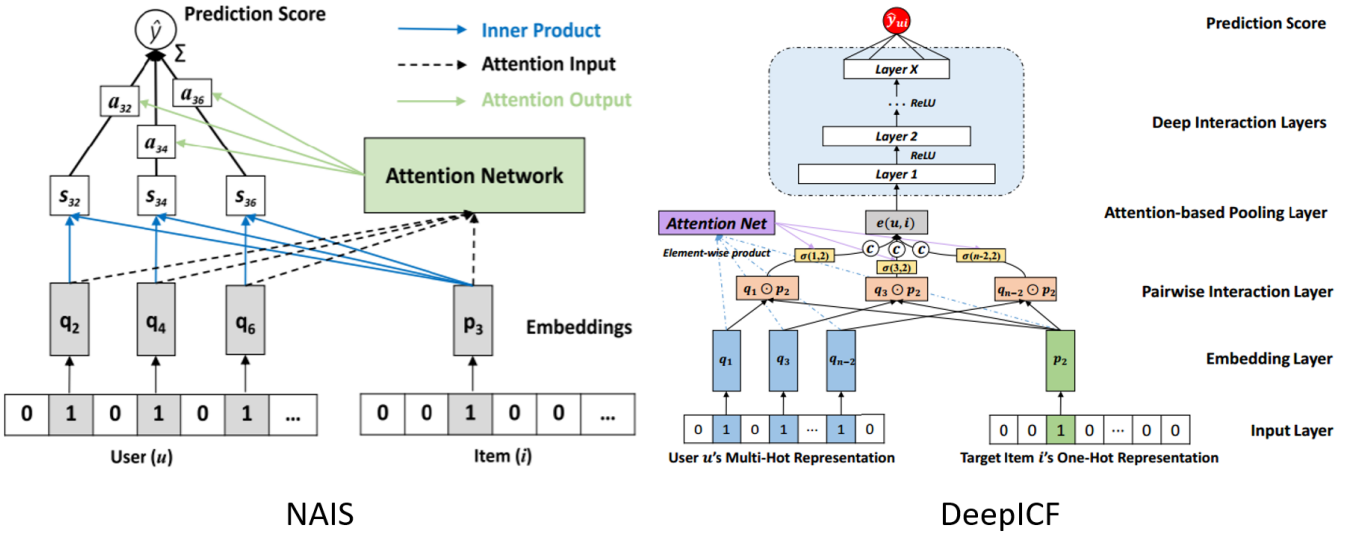

FIGURE 1. The architecture of NAIS & DeepICF, they the share same input and output format. NAIS uses the sum of weighted inner product of historical interactions and the target item embedding vectors to represent user's preference, based on it, DeepICF constructs a DNN to capture high-order relations among dimensions of embedding.

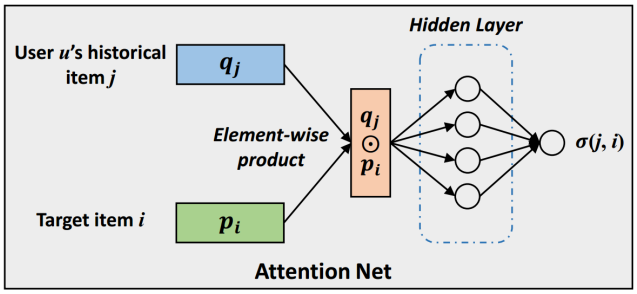

FIGURE 2. The architecture of the Attention network. We conduct a neural network to calculate the relevance between q_j and p_i that denotes the weight value.

k-dimensional vector, let it represented as $e_{u,i}$, set $e_{u,i} = f_{att}(\cdot)$. $e_{u,i}$ is input to a stacked multi-layer perceptron to achieve the high-order relations modeling. The deep interaction layers are defined as follows,

$$\begin{cases} e_1 = ReLU(W_1 e_{u,i} + b_1) \\ e_2 = ReLU(W_2 e_1 + b_2) \\ \dots \\ e_L = ReLU(W_L e_{L-1} + b_L), \end{cases} \tag{4}$$

$$\hat{y}_{u,i} = sigmoid(z^T e_L) + b_u + b_i, \tag{5}$$

where W_l and z denote the weight matrices, b_l, b_u, b_i denotes the bias vectors. $\hat{y}_{u,i}$ is the output of the whole model, ranges from 0 to 1 which represents the preferance of the user on the item.

C. FACTORIZATION MACHINE

In many recommending scenarios, features are represented in large amounts of categorical fields. When transformed into one-hot format, features are represented by extremely high dimensional vector, which leads to severe sparsity [19]–[21] problem in modeling, so we conduct a Factorization Machine [22] (short for FM) model to handle this problem. Besides, sharing same embedding matrix with DNN, FM is able to model low-order relations in feature pairs.

General linear model considers features individually, which ignores relations among features. But quantity of datasets from the real world indicate that there exists relations among features. From the perspective of the e-commerce website, the purchase of cosmetics is mostly for women, while the purchase of sporting goods is mostly for men. Such feature relations contain quantity of information, which leads to great influence on the performance of recommender system. Factorization Machine extracts the second-order relations which are defined as follows,

$$y = \omega_0 + \sum_{i=1}^{n} \omega_i x_i + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \omega_{ij} x_i x_j, \tag{6}$$

the equation is divided into two parts. The first part forms by the first and the second factors, which denote the linear model, ω_0 denotes the intercept, x_i and ω_i denote the i-th feature and the weight corresponding to it respectively. The second part is the second-order combinations of features. ω_{ij} denotes the weight of the combination of the i-th and j-th feature.

Due to the sparsity of features, there are few feature combinations meet the condition that both i-th feature and j-th feature are not "0", w_{ij} is hard to obtain by training. Thus we exploit accessorial vectors for each feature, set $V_i = (v_{i1}, vi2, \dots, vik)$, $\omega_{ij} = V_i V_j^T$. We have,

$$V = \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1k} \\ v_{21} & v_{22} & \dots & v_{2k} \\ \vdots & \vdots & & \vdots \\ v_{n1} & v_{n2} & \dots & v_{nk} \end{pmatrix}_{n \times k} = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix}, \tag{7}$$

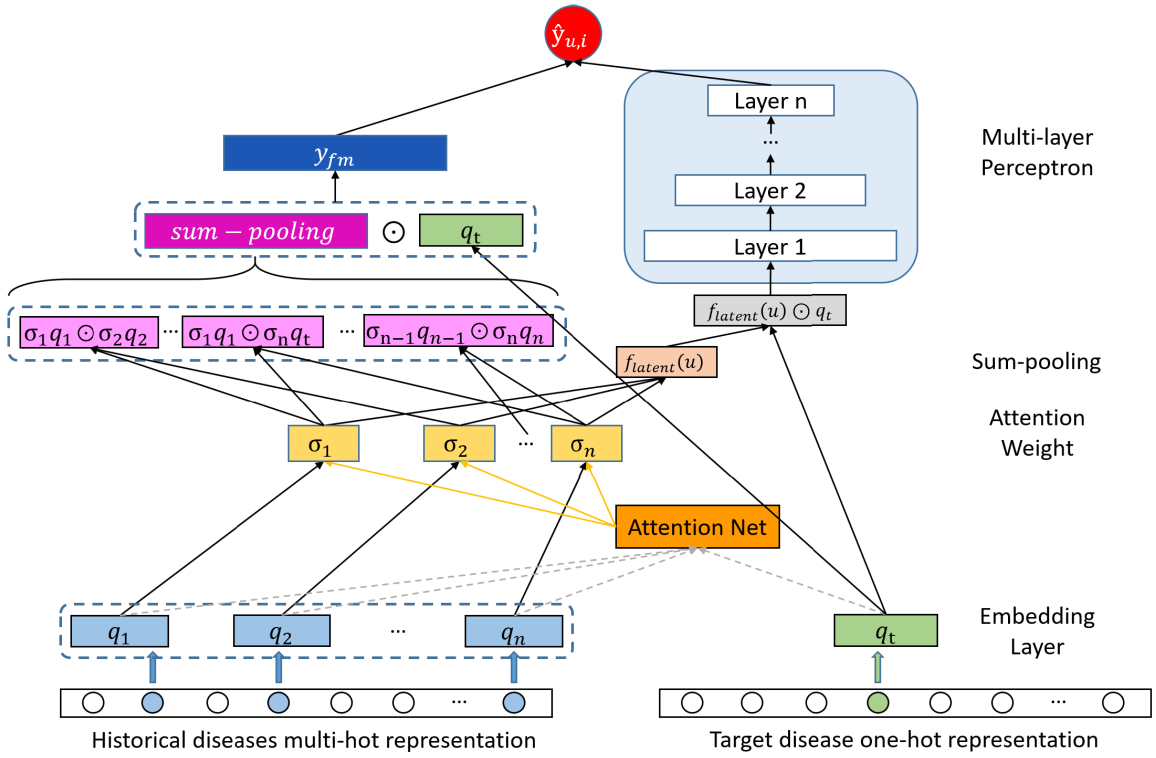

FIGURE 3. The architecture of our proposed model, each historical disease is weighted by it's relevance with the target disease, the FM part extract the low-order diseases combination features, and the deep part extract the high-order nonlinear feature relations with multi-layer perceptron. The FM part and deep part share the same embedding vectors.

so the weight matrix W is defined as:

$$\hat{W} = VV^T = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_n \end{pmatrix} (V_1^T, V_2^T, \ldots, V_n^T), \tag{8}$$

III. POTENTIAL DISEASE PREDICTION ALGORITHM BASED ON MEDICAL HISTORY

As reviewed in section II, NAIS method introduced the attention network to weight the items of user historical interactions, confirmed that user interacted items do not make equal contribution to the recommending of the target item. Based on NAIS, DeepICF introduced an end-to-end method to model user preference to the given items by considering high-order feature relations through deep neural networks. Inspired by DeepFM [18], we argue that low-order relations contain a small number of informations are not neglectable and should be considered simultaneously. We construct our model as follows.

A. INPUT DEFINITION

The architecture of our model is shown in Figure 3, the input contains two parts of data with the same dimensions. The blue circle denotes the disease that the patient once had in the whole disease set. While the green circle denotes the target disease is going to be predicted from the same disease set. The disperse and sparse history vector is mapped into several low dimension and dense vectors through an embedding matrix, each dimension in these vectors denotes a latent feature of the disease, which is randomly initialized and will be trained through the historical medical data.

B. PATIENT LATENT FEATURE REPRESENTATION

The number of historical diseases may vary among patients, to represent the latent feature of patients, we adopt the weighted sum-pooling on historical diseases as the latent features of patients, which is defined as follows,

$$f_{latent}(u) = \sum_{i \in I_u \setminus t} a(q_i, q_t) \cdot q_i, \tag{9}$$

$f_{latent}(u)$ denotes the latent feature of the patient u, it can be viewed as health condition of patient u, q_i and q_t denote the embedding vectors of i-th historical disease and target disease t respectively. The relevance of diseases are different, so the historical diseases are weighted by their relevance. The weight of each disease $a(\cdot)$ (we set $\boldsymbol{\sigma_i} = a(q_i, q_t)$ in Figure3.) is calculated in an Attention Network function defined in Equation (2).

C. MULTI-LAYER PERCEPTRON

Supposing the embedding dimension is set to k, we have k-dimensional vector $f_{latent}(u)$ that denotes the latent feature of the patient u from equation (9). We adopt an element-wise

product operation between $f_{latent}(u)$ and q_t to represent their relations (eg. similarity or causality). The elements in the vector denote different indicators of health condition, to exploit the interplay among these indicators, we stack a multi-layer perceptron to extract nonlinear and high-order relations among these latent features, for it provides valuable signal, which is hard to find manually, to estimate the relevance between medical history and the target disease, which is defined as follows,

$$\begin{cases} L_0 = f_{latent}(u) \odot q_t \\ L_1 = BN(ReLU(W_1 L_0 + b_1)) \\ L_2 = BN(ReLU(W_2 L_1 + b_2)) \\ \dots \\ L_n = BN(ReLU(W_n L_{n-1} + b_n)), \end{cases} \tag{10}$$

BN [23] denotes the batch normal operation, which reduces the model dependency on variables initialization, replacing the traditional dropout [24] layer. W_i and b_i denote the weight matrix and bias of each layer.

D. LOW-ORDER RELATIONS

Inspired by DeepFM [18], we argue that the deep part missed some low-order relations which contains information that is not neglectable. The deep part extracted pairwise relation between historical disease and the target disease, and uses DNN to mine the high-order relations among latent features. But the latent features are weighted overlay of these pairwise relations, in real world scenario, the relation among two or three diseases combination in historical diseases also contributes to the target disease, we define it as low-order relations. Combining both low-order and high-order relations make more contribution to the performance of predicting. Different from DeepFM, we abandon the linear part of FM, since the number of historical disease are vary among patients, unlike those fixed dimension features in click-through rate (CTR) prediction. With the weights of historical diseases calculated in section III.B, we map the interacted items with the weights and calculate the element-wise product with the target disease embedding vector. The FM part is defined as:

$$f_{fm}(u) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} < (a_i q_i), (a_j q_j) > x_i x_j, \tag{11}$$

simplified as:

$$f_{fm}(u) = \frac{1}{2} \sum_{f=1}^{k} ((\sum_{i=1}^{n} q_{i,f} x_i)^2 - \sum_{i=1}^{n} q_{i,f}^2 x_i^2), \tag{12}$$

$q_{i,f}$ denotes the f-th value of the i-th historical disease embedding vector, x_i is binary, if the i-th disease in the set is one historical disease of the user, $x_i = 1$, otherwise $x_i = 0$, k and n denote the dimension of embedding vectors and the number of the whole disease set, respectively. We set

$$y_{fm}(u) = \frac{1}{2}((\sum_{i=1}^{n} q_i x_i)^2 - \sum_{i=1}^{n} q_i^2 x_i^2) \odot q_t, \tag{13}$$

The output of the low-order part is a k-dimensional vector, which contains second-order relations among historical diseases.

E. OUTPUT AND LEARNING

For the last part of our approach, we concatenate the low-order and high-order outputs, and adopt an output layer with sigmoid function to calculate the final score which ranges from 0 to 1. The closer to 1 the score is, the higher possibility of the disease the patient may have.

$$\delta(\hat{y_{ui}}) = Sigmoid(z^T concat(y_{fm}(u), L_n)) + b_u + b_i, \tag{14}$$

where z is the weight vector, b_u and b_i denote the patient bias and disease bias of output layer. Our model can be treated as a bipartition problem, the loss function is defined as:

$$loss = \frac{-1}{|R^+| + |R^-|}[\sum_{(u,i) \in R^+} \log \delta(\hat{y_{ui}}) + \sum_{(u,j) \in R^-} \log(1 - \delta(\hat{y_{uj}}))] + \lambda \|\Theta\|^2, \tag{15}$$

Θ is the parameters of the L_2 regularization for preventing overfitting. R^+ denotes the positive examples, and R^- denotes the negative examples which sample from the disease set that the user didn't have (introduced in Section IV.A).

IV. EXPERIMENTS

In this section, to verify the effectiveness of our proposed disease predicting model, we conduct some experiments and compare our proposed model with the other state-of-the-art models on a public dataset. The evaluation result indicates that our proposed model achieved better performance than any other state-of-the-art models.

A. EXPERIMENT SETUP

Datasets: We conduct the experiment on the public dataset provided by [25], it contains 103,886 pairs of sample of 1355 genes and 6563 phenotypes. This dataset provides a link between genes and Human Phenotype Ontology (HPO) terms. All phenotype terms associated with any disease that is associated with variants in a gene are assigned to that gene in this dataset. We adopt the widely used *leave-one-out* [26], [27] evaluation protocol to evaluate our model in this section. The last phenotype is selected to testing data for each gene and the rest for training. We sample 99 random phenotypes which are not linked to the gene as the negative testing example. In other words, the gene in the testing dataset is linked to 1 positive phenotype and 99 negative phenotypes, our work is to ensure the positive phenotype ranks higher than any other negative phenotypes on the output score. To achieve better performance on these supervised learning models, we sample

TABLE 1. Performance of HR@10 and NDCG@10 of our model and competitors. Embedding size is set to 32.

Metrics	HR@10	NDCG@10
FISM	0.7534	0.6943
NAIS	0.7664	0.7180
DeepICF	0.7848	0.7344
Our model	**0.7941**	**0.7695**

a number of negative examples for training. Consistent with these researches [15], [17], we set the negative sampling ratio to 4, which leads to the dataset nearly 5 times bigger.

Evaluation Metrics: We adopt two metrics to evaluate the performance of models in our experiment: *Hit Ratio* at rank k (HR@k) [26] and *Normalized Discounted Cumulative Gain* at rank k (NDCG@k) [19], [28], set $k = 10$. HR@10 is the hit ratio that the positive phenotypes of the gene is present at the top-10 ranked list, and NDCG@10 evaluate the position of the positive phenotypes in the top-10 list, the higher the ranking, the higher the score.

B. COMPETITORS

FISM [29]: It stands for the mainstream in learning-based item-based collaborative filtering, it provides a method that learns the item-item similarity matrix as the product of two low dimensional latent factor matrices, dealing with the problem of sparsity.

NAIS [17]: FISM's modeling fidelity can be limited by its assumption that all historical items of a user contribute equally in estimating the similarity between the user profile and the target item. The relevance between interacted items and the target item varies among pairs. So they designed an attention network to weight the interacted items of users, representing the score as the assumption of the weighted and paired inner product.

Base Model (DeepICF) [15]: It is built upon NAIS, using DNN to model high-order relations that NAIS missed. The DNN shows effective improvement on modeling high-order feature relations, and they verified that the hyper-parameters including neural network depth and embedding size have a certain effect on the performance of the model.

C. PERFORMANCE EVALUATION

Considering all models of competitors are embedding-based, we set the embedding size of all models to 32, the hidden layers of the attention network are set to $32 \times 16 \times 1$ and the smoothing parameter β is set to 0.5. The hidden layers of DeepICF and the deep part of our model are constructed with 3 layers of fully connected neural networks (or MLP), the layers of MLP are set to $32 \times 16 \times 16 \times 1$.

The evaluation performance is presented in Table 1, we can see that our proposed model achieved the best performance (both HR@10 and NDCG@10) among all competitors.

First of all, we can see from the Table 1 that the performance of NAIS is better on FISM on both HR@10 and NDCG@10, we owe this achievement to the attention network, it weighted the observed phenotypes with the relevance between each observed phenotype and the target phenotype, reduced the interference of those irrelevant phenotypes. Secondly, we can see that DeepICF significantly outperforms FISM and NAIS, it shows that high-order relations of the latent features of phenotypes contain more information that is helpful for modeling relevance between gene and the target phenotype. Thirdly, our proposed model achieves the best performance among all competitors, which means the low-order relations of phenotype combination are also unneglectable, considering both high-order and low-order relations achieves better performance than any single part of them.

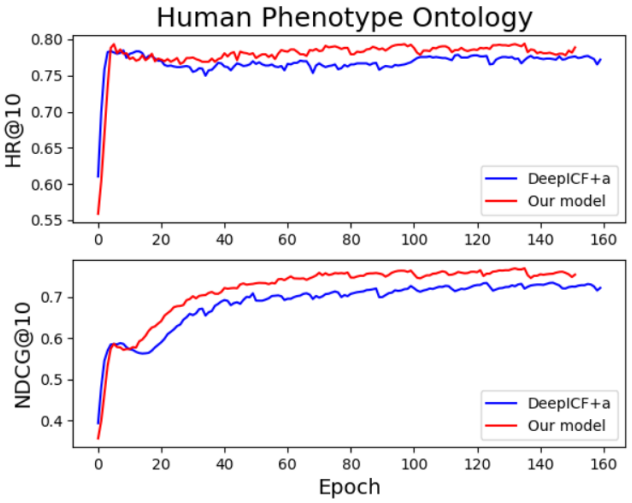

FIGURE 4. Testing performance of DeepICF and our proposed model on the public HPO dataset.

Figure 4 shows the growth of the HR@10 and NDCG@10 with the training step grows. The HR@10 achieved the best performance and remained stable early, while the NDCG@10 grew a few steps before stable, that means those positive phenotypes ranked in top-10 early and ranked higher when training step grew. Besides, the closer to HR@10 the NDCG@10 is, the closer to top-1 the positive phenotypes are, it seems our work also performs the best from this perspective. Since the hyper-parameters of the deep part of our model are the same to DeepICF [15], and they did detailed hyper-parameters study in their work, thus we do no more further research on the hyper-parameters comparison in this paper.

V. CONCLUSION

In this paper, we proposed a medical-history-based potential disease prediction algorithm, and overcame the shortcomings of the state-of-the-art models listed in section IV.B with combining the advantages of FM and DNN. Our proposed model can not only handle high-order relations among disease features, but also consider the low-order combination of diseases that leads to a certain disease, thus the model's comprehensiveness is improved. We conduct some significant experiments on the real-world datasets that predict the potential phenotype based on the observed phenotypes of the gene. We owe the improvement of our proposed model to the

following advantages: 1) we weighted the historical diseases with an attention network, the noise of irrelevant-disease is reduced; 2) by learning the deep part and the FM part jointly, the low-order relations and high-order relations are merged to predict the final score, that performs better than any single part of it.

Accurate potential disease prediction plays an important role in medical examination assisting, we believe that there are certain relations among diseases, as the genes, to be studied, it reminds doctors or patients those diseases which are latently relevant, provides medical references from the view of statistics, and improves the comprehensiveness and pertinence of the medical examination. However, since the prediction of the potential disease is pure statistic based, the prediction result and the latent features of diseases are unexplainable, we need to add some professional knowledge with some methods to make the prediction result more convincing. Thus for future study, there are two feasible directions. One is using the symptom of patients and other side information to improve the prediction accuracy. The other is considering the explicit features and other professional knowledge of diseases to make prediction more explainable with some strategies.

REFERENCES

[1] P. Jacquemart and P. Zweigenbaum, "Towards a medical question-answering system: A feasibility study," *Med. Inform. Eur.*, vol. 95, pp. 463–468, 2003.
[2] A. B. Abacha and P. Zweigenbaum, "MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies," *Inf. Process. Manage.*, vol. 51, no. 5, pp. 570–594, 2015.
[3] Y. Peng and X. Zhang, "Integrative data mining in systems biology: From text to network mining," *Artif. Intell. Med.*, vol. 41, no. 2, pp. 83–86, Oct. 2007.
[4] N. Zeng, Z. Wang, B. Zineddin, Y. Li, M. Du, L. Xiao, X. Liu, and T. Young, "Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach," *IEEE Trans. Med. Imag.*, vol. 33, no. 5, pp. 1129–1136, May 2014.
[5] N. Zeng, Z. Wang, Y. Li, M. Du, and X. Liu, "A hybrid EKF and switching PSO algorithm for joint state and parameter estimation of lateral flow immunoassay models," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 9, no. 2, pp. 321–329, Mar./Apr. 2012.
[6] N. Zeng, Z. Wang, Y. Li, M. Du, and X. Liu, "Inference of nonlinear state-space models for sandwich-type lateral flow immunoassay using extended Kalman filtering," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 7, pp. 1959–1966, Jul. 2011.
[7] P. E. Johansson, G. I. Petersson, and G. C. Nilsson, "Personal digital assistant with a barcode reader—A medical decision support system for nurses in home care," *Int. J. Med. Inform.*, vol. 79, no. 4, pp. 232–242, 2010.
[8] J. R. Slagle, M. W. Gaynor, and E. J. Halpern, "An intelligent control strategy for computer consultation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 2, pp. 129–136, Mar. 1984.
[9] E.-Y. Lin, D.-L. Yang, and M.-C. Hung, "System design of an intelligent nutrition consultation and recommendation model," in *Proc. 9th Int. Conf. Ubiquitous Intell. Comput., 9th Int. Conf. Autonomic Trusted Comput.*, Sep. 2012, pp. 740–745.
[10] Y. Peng, "A novel ensemble machine learning for robust microarray data classification," *Comput. Biol. Med.*, vol. 36, no. 6, pp. 553–573, 2006.
[11] Y. Peng, Z. Wu, and J. Jiang, "A novel feature selection approach for biomedical data classification," *J. Biomed. Inform.*, vol. 43, no. 1, pp. 15–23, 2010.
[12] C. Li, "Research on the bottleneck problems of collaborative filtering in E-commerce recommender systems," Ph.D. dissertation, Hefei Univ. Technol., Hefei, China, 2009.
[13] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, vol. 463. New York, NY, USA: ACM Press, 1999.
[14] H.-L. Xu, X. Wu, X.-D. Li, and B. P. Yan, "Comparison study of Internet recommendation system," *J. Softw.*, vol. 20, no. 2, pp. 350–362, 2009.
[15] F. Xue, X. He, X. Wang, J. Xu, K. Liu, and R. Hong, "Deep item-based collaborative filtering for top-N recommendation," *ACM Trans. Inf. Syst.*, vol. 37, no. 3, 2019, Art. no. 33.
[16] E. Christakopoulou and G. Karypis, "HOSLIM: Higher-order sparse linear method for top-N recommender systems," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2014, pp. 38–49.
[17] X. He, Z. He, J. Song, Z. Liu, Y.-G. Jiang, and T.-S. Chua, "NAIS: Neural attentive item similarity model for recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 12, pp. 2354–2366, Dec. 2018.
[18] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, "DeepFM: A factorization-machine based neural network for CTR prediction," 2017, *arXiv:1703.04247*. [Online]. Available: https://arxiv.org/abs/1703.04247
[19] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proc. WWW*, vol. 1, 2001, pp. 285–295.
[20] M. Papagelis, D. Plexousakis, and T. Kutsuras, "Alleviating the sparsity problem of collaborative filtering using trust inferences," in *Proc. Int. Conf. Trust Manage.* Berlin, Germany: Springer, 2005, pp. 224–239.
[21] A. B. Barragáns-Martínez, E. Costa-Montenegro, J. C. Burguillo, M. Rey-López, F. A. Mikic-Fonte, and A. Peleteiro, "A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition," *Inf. Sci.*, vol. 180, no. 22, pp. 4290–4311, 2010.
[22] S. Rendle, "Factorization machines," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 995–1000.
[23] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: https://arxiv.org/abs/1502.03167
[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[25] S. Köhler, L. Carmody, N. Vasilevsky, J. O. B. Jacobsen, D. Danis, J.-P. Gourdine, M. Gargano, N. L. Harris, N. Matentzoglu, and J. A. McMurry, "Expansion of the human phenotype ontology (HPO) knowledge base and resources," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1018–D1027, 2018.
[26] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua, "Neural collaborative filtering," in *Proc. 26th Int. Conf. World Wide Web*, 2017, pp. 173–182.
[27] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proc. 25th Conf. Uncertainty Artif. Intell.*, 2009, pp. 452–461.
[28] D. Cao, X. He, L. Nie, X. Wei, X. Hu, S. Wu, and T.-S. Chua, "Cross-platform app recommendation by jointly modeling ratings and texts," *ACM Trans. Inf. Syst.*, vol. 35, no. 4, p. 37, 2017.
[29] S. Kabbur, X. Ning, and G. Karypis, "FISM: Factored item similarity models for top-n recommender systems," in *Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2013, pp. 659–667.

WENXING HONG received the Ph.D. degree in system engineering from Xiamen University, China, in 2010. Since 2010, he has been with the Department of Automation, Xiamen University, where he is currently an Associate Professor. He is the author or coauthor of more than 25 papers. He has led and participated in more than 15 research projects, including the National Natural Science Foundation of China project. His current research interests include the areas of data mining, big data, artificial intelligence, recommendation systems, and FinTech. He is the Dean of the Research Center for Systems and Control, Xiamen University. He is a member of the CCF. He has been served as the General Secretary of the International Conference on Computer Science and Education (ICCSE) and the Fujian Systems Engineering Society, from 2006 and 2010.

ZIANG XIONG received the B.E. degree from Xiamen University, China, in 2018, where he is currently pursuing the master's degree with the Department of Automation. He is currently with the System and Control Center Laboratory, Xiamen University, where he is currently involved in deep learning and recommendation systems.

NANNAN ZHENG received the B.E. degree from Xiamen University, China, where she is currently pursuing the master's degree with the Department of Automation. She is also with the System and Control Center Laboratory, Xiamen University, where she is currently involved in deep learning and recommendation systems.

YANG WENG received the B.S. and Ph.D. degrees from the Department of Mathematics, Sichuan University, Chengdu, China, in 2001 and 2006, respectively. He was a Postdoctoral Fellow of Nanyang Technological University, Singapore, from August 2008 to July 2010. Since 2006, he has been with the College of Mathematics, Sichuan University, where he is currently an Associate Professor. His current research interests include statistic machine learning and nonparametric Bayesian inference.