Multi-Domain Feature Representation and Multi-Dimensional Feature Interaction for Person-Job Fit

Bin Li, Jianwen Ding, Zhennan Chen, Hanpeng Jiang, Fan Lin*
School of informatics, Xiamen University, Xiamen, China
{30920211154159, 30920201153956, znchen, 30920221154272}@stu.xmu.edu.cn, *iamafan@xmu.edu.cn

***Abstract*—Person-job fit aims to use the algorithms to match jobseekers with job postings to overcome information overload on online recruitment platforms. Traditional matching algorithms are not ideal in the feature representation and interaction of resumes and job postings. To this end, we propose a person-job fit model PJFFRFI based on multi-domain feature representation and multi-dimensional feature interaction, which comprehensively considers the features of various domains and learns feature correlation vectors in different dimensions. Specifically, we first divide the features in resumes and job postings into seven domains, and design different representation methods according to the data type. Then we propose a feature enhancement module (FEM) based on multi-head self-attention to learn the feature correlation vectors in resumes and job postings. Moreover, we propose a feature interaction module (FIM) to facilitate feature interaction both inside and outside the domain. Extensive experiments on a real-world dataset demonstrate that the proposed method significantly surpasses the state-of-the-art methods.**

***Index Terms*—Feature representation, Multi-dimensional feature interaction, Feature enhancement, Person-job fit**

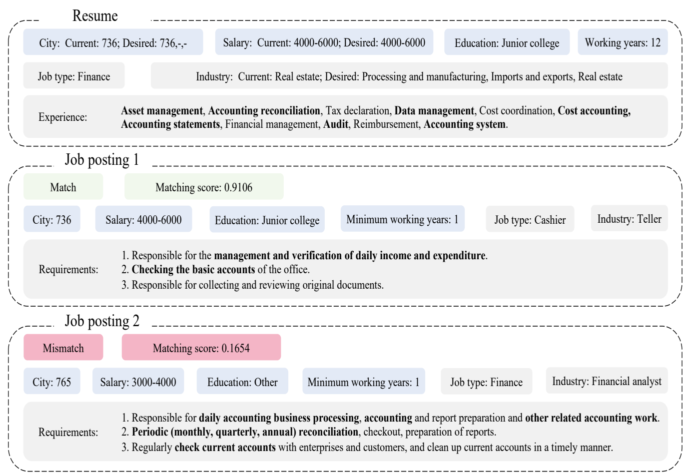

Fig. 1. Case study: resume and its candidate jobs.

I. Introduction

Due to the epidemic, online recruitment has recently taken over as the primary method [1] in the job market, making it simpler for both jobseekers and employers to find the right jobs and qualified candidates. By the end of 2022, LinkedIn had over 850 million users and 58 million companies registered in more than 200 countries. Due to the huge amount of data and fast-paced requirements of the job market, it is difficult to meet these requirements only through manual review. Therefore, it is extremely urgent to design an effective person-job fit algorithm to quantify the matching degree between jobseekers and job postings.

In the recent job market, a typical job matching idea is to match the requirements released by the employer with descriptive data such as experience of the jobseeker in Fig. 1. Thus, a series of deep learning algorithms dedicated to solving text matching came into being. Zhu et al. [2] proposed the PJFNN based on CNN, which adopted the hierarchical representation structure that can identify specific requirements the candidates meet in the job posting. Qin et al. [3] proposed a RNN model based on a hierarchical attention to learn word-level semantic representations of resumes and jobs and ability perception representations with different levels. However, these models only focus on text features, ignoring the possible impact of structured data (The blue part of Fig. 1). For example, when the job city posted by the employer is not in the jobseeker's plan, the jobseeker has a high probability of rejecting the job even if the text matching score is high. Furthermore, [4] introduced click stream and browse volume as structured data to assist text data, [5]–[7] also introduced different features such as numerical features identified as structured data, and considered the possible impact of interaction between structured data and textual data on matching results. Nevertheless, they did not fully mine the implicit feature interaction information from multiple dimensions. On the one hand, there is often a strong correlation between structural features. For instance, the more developed the city, the higher the salary level. Similarly, salary has a positive correlation with education and working years. On the other hand, there are also plentiful hidden information within the same domain features interaction and between different domain features interaction from resumes and job postings.

To sum up, we propose a person-job fit model based on multi-domain feature representation and multi-dimensional feature interaction, termed PJFFRFI. Our proposed model can capture the comprehensive interaction features based on learning different types of feature representations given a job posting and a resume. Specifically, we divide all the features in

DOI reference number: 10.18293/SEKE2023-148

resumes and job postings into three types as shown in Table 1, and adopt different embedding methods according to the data type. Meanwhile, considering the correlation between structure features, a feature enhancement module (FEM) is introduced to learn the feature correlation both in resume and job postings. Along this line, we propose a feature interaction module (FIM) to extract the hidden correlation features between resumes and job postings in the same domain and the hidden feature combinations between domains.

Our contributions can be summarized as follows:

- We propose a novel person-job fit model named PJFFRFI by learning embedding vectors of different domains and extracting feature correlations from different dimensions.
- We propose a feature enhancement module (FEM) to introduce correlation features among various domains in job postings. Furthermore, we design a novel feature interaction module (FIM), which leverages inner and outer feature interactions to learn the implicit feature correlations.
- Extensive experiments demonstrate our superior performance over the recent state-of-the-art methods by a large margin.

II. Related Work

In earlier study, person-job fit is regarded as a recommendation problem, relying on Boolean keyword matching which usually fails to give satisfactory recommendations. Malinowski et al. [8] proposed a bilateral selection model for the first time, which took the preference of both jobs and jobseekers into account and improved the recommendation performance. After that, the extensive application of collaborative filtering algorithm (CF) [9], [10] in job recommendation tasks has further promoted the progress of person-post matching tasks. However, the CF algorithm is based on historical interaction data and has the defect of cold start, which has attracted many experts to use different methods such as hybrid recommendation [11]–[13] to solve this problem.

Due to the advanced performance of deep learning technology in semantic mining, convolutional neural network (CNN) [2], recurrent neural network (RNN) [3], attention [14] and etc have been widely used in person-job fit. Luo et al. [15] integrated different types of information in a hierarchical representation and introduced adversarial learning to model job and resume representations. So as to make up for the semantic difference between job postings and resumes, Yao et al. [16] designed a knowledge-aware graph encoder and incorporated prior knowledge into graph representation learning to improve the performance.

In order to further enhance the semantic representation and learn more effective information in resumes and job postings, Bian et al. [4] proposed a multi-view co-teaching network from sparse interaction data, which introduced a relation-based module to complement the text-based matching module, realized the enhancement of semantic representation and data enhancement. FINN [5] divided features into three fields, and learned interaction signals of categorical features and textual features respectively. Jiang et al. [6] proposed a feature fusion method, which fused the expressive features of the job and candidate, so as to obtain a more comprehensive and effective representation. He et al. [7] proposed an end-to-end person-job fit model MUFFIN, which designed a module to learn the latent correlations between features in each field and a module based on multi-head self-attention with a residual connection to learn interactions.

Inspired by existing work, we propose a deep learning model to predict the matching scores between resumes and job postings. Compared with existing methods, our model divides all features into different domains and learns feature representations respectively. Moreover, we propose several feature interaction modules to extract feature correlations from multiple dimensions including within the job postings, within the domain, and between domains.

III. Methodology

A. Problem Definition

We denote the resume set as $R = \{r_1, r_2, \dots, r_m\}$ and job set as $J = \{j_1, j_2, \dots, j_n\}$, where m and n are total number of resumes and job postings respectively. Resume r_p and job posting j_q both have k domains, denoted as $r_p = \{r_{p,1}, r_{p,2}, \dots, r_{p,k}\}$ and $j_q = \{j_{q,1}, j_{q,2}, \dots, j_{q,k}\}$. Particularly, according to the characteristics of private dataset, $r_{p,i}$ may consists m_i features, denoted as $r_{p,i} = \{r_{p,i,1}, r_{p,i,2}, \dots, r_{p,i,m_i}\}$ while j_q has only one feature in one domain. For example, current salary and desired salary listed in the resume belongs to one domain named salary. The recruitment records are denoted as the set of $PJF = \{r_p, j_q, y_{p,q}\}$, where $y_{p,q} \in \{0, 1\}$. $y_{p,q} = 1$ means that the resume r_p successfully fits the job j_q while $y_{p,q} = 0$ means that the match fails. Our target is to design a deep learning model to predict $y_{p,q}$.

B. Overview

As shown in Fig. 2, the proposed model PJFFRFI groups all features from resumes and job postings into k domains as input. In the embedding module, there are two methods mapping numerical domain features (e.g., Salary), categorical domain features (e.g., City) and textual domain features (e.g., Requirements and Experience) into two hidden spaces, and get the embedding vectors of different domains respectively, denoted as $domain_i$ vector. Followed by a feature enhancement module (FEM), multi-head self-attention (MHSA) is used to introduce hidden correlation vectors between structured features, and output the feature enhancement vector in different domains, marked as $domain_i$ e_vector. Then, these vectors will be fed into the feature interaction module (FIM), including inner and outer interaction modules. Finally, we predict the matching score by prediction module.

C. Embedding Module

To begin with, we normalize numerical features with standard distributions and project categorical features into one-hot

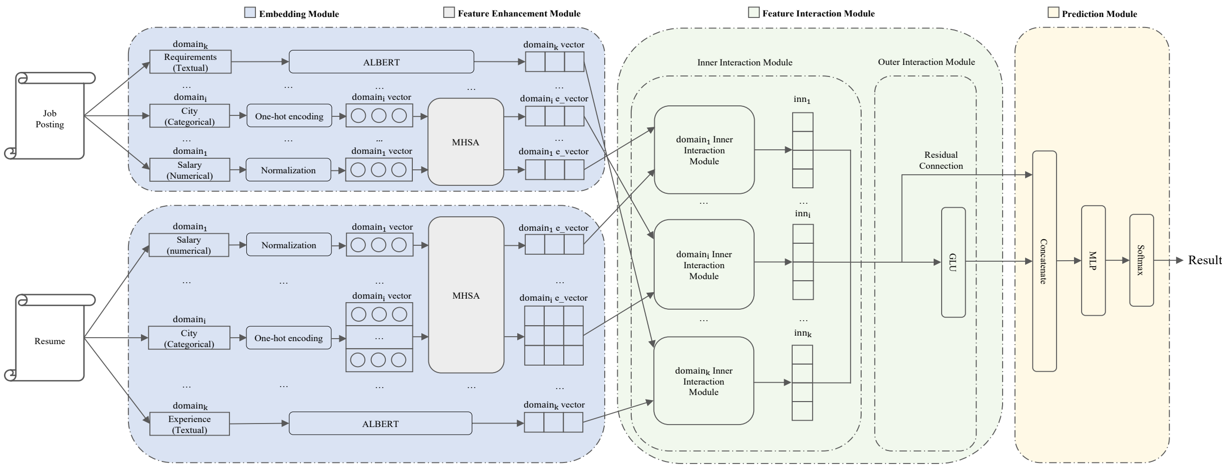

Fig. 2. Architecture of PJFFRFI.

vectors. To concatenate all types of features, each job and resume is represented as a vector. It is defined as:

$$\text{Input}^R = \left[x^R_{1,1}, x^R_{1,2}, \ldots, x^R_{1,m_1}, x^R_{2,1}, \ldots, x^R_{k,m_k}\right], \quad (1)$$

$$Input^J = \left[x^J_1, x^J_2, \ldots, x^J_k\right], \quad (2)$$

Where m_k represents the number of features of the kth domain in resume. $x^R_{i,t}$ and x^J_i represent the input of the tth feature in the ith domain in the resume and the ith domain feature in the job posting respectively. When the ith domain is of numeric type, the $x^R_{i,t}$ and x^J_i are scalar values. Otherwise, they are vectors.

Next, the numerical features and categorical features are mapped to a low-dimensional vector, which is defined as:

$$emb^R_{i,t} = V_i x^R_{i,t}, emb^J_i = V_i x^J_i, \quad (3)$$

where V_i is a shared embedding matrix. Meanwhile, textual features are embedded by a pretrained ALBERT [17]. The integration process can be denoted as:

$$emb^R_{i,t} = ALBERT(x^R_{i,t}), emb^J_i = ALBERT\left(x^J_i\right) \quad (4)$$

D. Feature Enhancement Module (FEM)

Since the basic information(education, living city etc.) of non-text features in resumes is more likely to be affected by factors such as the candidates' family and social relations, reflects personalized job seeking tendency. therefore, it is of significance to study the relevance between non-text features in resumes. In the same way, analysis of structural features in job postings is helpful in understanding recruiters' intentions, so we treat the embedding vectors of non-textual features as input and propose a FEM based on multi-head self-attention to introduce strong feature correlations both in resumes and job postings.

Multi-head self-attention has been successfully applied to semantic understanding [18], machine translation [19], etc. We apply it to capture the correlation between non-textual features in job postings and use the three most relevant features to achieve feature enhancement. The core of the self-attention is Query, Key, and Value. The multi-head self-attention is a combination of multiple self-attention modules, providing multiple representation subspaces to the attention layer. Taking the ith non-textual feature as an example, we calculate its similarity to other features and normalize with softmax. The correlation weight coefficient is calculated as:

$$\alpha^g_{i,j} = \frac{\left(Q^g emb^J_i\right) \cdot \left(K^g emb^J_j\right)^{\text{transpose}}}{\sqrt{d}}, \quad (5)$$

$$\tilde{\alpha}^g_{i,j} = \frac{\exp\left(\alpha^g_{i,j}\right)}{\sum_{j=1}^{k} \exp\left(\alpha^g_{i,j}\right)}, \quad (6)$$

$$\alpha_{i,j} = \frac{1}{|G|} \sum_{g=1}^{|G|} \tilde{\alpha}^g_{i,j}, \quad (7)$$

where Q^g and K^g are the Query and Key of the gth head. d represents the dimension of emb^J_i and $|G|$ represents the number of heads. We select the three most relevant features, multiply their embedding vectors by the corresponding weight coefficients and add them to the embedding vector of the ith feature to achieve feature enhancement. The final representation of the ith non-textual feature in the job posting can be defined as:

$$emb^J_i = emb^J_i + \sum_{j\epsilon Top\ 3\ related\ features} \alpha_{i,j} * emb^J_j \quad (8)$$

E. Feature Interaction Module (FIM)

1) Inner Interaction Module: Different from the common method of using average pooling to compress multiple features in each domain in the resume, we interact each of the features in the resume with the feature in the job posting in the corresponding domain, calculate the difference and the Hadamard product between two embedding vectors respectively, introduce the distance correlation vector and angle correlation vector between the features and then concatenate them with the two original embedding vectors to represent the interactions between resumes and job postings in the domain. Finally, the interaction vectors in the same domain are compressed by average pooling. The specific formula is as follows:

$$inn_i = \frac{1}{m_i}\sum_{t=0}^{m_i}\{W_i[emb_{i,t}^R \oplus emb_i^J \oplus emb_{i,t}^R - emb_i^J \oplus emb_{i,t}^R \odot emb_i^J] + b_{i,t}\}, \tag{9}$$

where $\oplus$ represents concatenation and $\odot$ represents the Hadamard product. m_i denotes the number of features contained in the resume in the ith domain. W_i and $b_{i,t}$ represent the parameters. In order to project all the domain features into same dimensional space, we use a MLP layer to map all the vectors.

2) Outer Interaction Module: One of the core problems in feature interaction is to extract the hidden feature combinations. Hence, in outer interaction module, we model the domain feature vectors with the same dimension to learn the meaningful combinations of domains. We propose a Gated Linear Unit (GLU) [20] to capture feature interactions between domains. First, we concatenate the interaction vectors between resumes and job postings in each domain:

$$Inn = inn_1 \oplus inn_2 \ldots \oplus inn_k, \tag{10}$$

where $\oplus$ represents concatenation and k represents the number of domains. Inn is fed to two convolutional layers in GLU. The output of the first convolutional layer with sigmoid function can help the model to learn the importance of features by backpropagation. The second convolutional layer has no activation function. The output of GLU is the Hadamard product of the output of the two convolutional layers, the formula is:

$$\mathrm{Conv}_1 = \frac{1}{1+\exp\left[-\left(W_1 * \mathrm{Inn} + b_1\right)\right]}, \tag{11}$$

$$Conv_2 = W_2 * Inn + b_2, \tag{12}$$

$$\widetilde{Out} = \mathrm{Conv}_1 \odot \mathrm{Conv}_2, \tag{13}$$

where W_1, b_1, W_2, b_2 are parameters, $\odot$ represents the Hadamard product. In order to avoid network degradation, we add a residual connection to reserve the interaction features between resumes and job postings learned by the inner interaction module. At last, the output of outer interaction module is:

$$Out = \widetilde{Out} + Inn \tag{14}$$

F. Prediction Module

In prediction module, MLP fed with the hidden vectors of interactions between jobs and resumes and the combinations of interaction features between domains is introduced to predict the matching scores. To simplify, we use d^0 to denote the input at 0th layer. Next, the output of each layer in the MLP can be formulated as:

$$d^{l+1} = ReLU\left(W_{l+1}d^l + b_{l+1}\right), \tag{15}$$

where d^{l+1} is the output of the lth layer, W_{l+1} and b_{l+1} are parameters. Then the predictive matching scores can be obtained through a two-dimensional vector which is the output of the last layer and are mapped into the matching probability through the softmax function:

$$\widehat{y_{p,q}} = softmax\left(W_{L+1}d^L + b_{L+1}\right), \tag{16}$$

where L indicates the depth of the MLP.

Finally, we use the cross-entropy loss function to optimize:

$$Loss = -\frac{1}{N}\sum\left[y_{p,q}\log\widehat{y_{p,q}} + (1-y_{p,q})\log\left(1-\widehat{y_{p,q}}\right)\right] \tag{17}$$

where $y_{p,q}$ and $\widehat{y_{p,q}}$ represent the matching label and predictive matching score of $< r_p, j_q >$ respectively. N indicates the total number of training samples.

IV. Experiments

A. Experimental Setup

1) Dataset: Our experiments use private data provided by recruitment platforms and companies. Overall, our dataset consists of 35,413 job postings, 1,648 resumes and 43,504 recruitment records. To meet the experimental conditions, we removed the job posting with null requirement and resume with null experience, and replaced the null value in the structured feature with "other". And we filter the dataset to make each resume have more than 6 suitable jobs and at least 3 unsuitable jobs to simulate the actual recruitment market. Finally, we get 43,504 recruitment records, among which 25,032 are positive and 18,472 are negative. Fig. 3 shows the distribution of person-job fit. It can be seen from (a) and (b) that most jobs only recruit one person, some recruit 2 to 4 people, and a few recruit more than 4 people. (c) shows that applicants generally send their resumes to multiple companies to increase employment opportunities, while (d) explains that a resume will usually receive offers from 4 to 21 companies. Through statistics, we found that almost all work experience resumes are less than 300 words, and most of the job descriptions are less than 500 words. In addition, Table 1 shows the 7 domains and 3 categories we designed.

2) Hyper-parameter Setting: In our experiments, we set the max number of words as 300 and 512 for experience of each resume and requirements of job posting respectively according to the statistical results. The dimension of numerical and categorical embedding vectors is 32 while the output dimension of ALBERT is 312. FEM uses a two-headed self-attention network with an embedding size of 32. For GLU in

TABLE I
DOMAINS AND CORRESPONDING FEATURES.

Domains	Type	Features in resumes	Features in job postings
Working years	Numerical	Working years	Minimum working years
Salary	Numerical	Desired salary; Current salary	Minimum Salary
Education	Categorical	Academic qualifications	Academic requirements
City	Categorical	Living city; Top 3 desired cities	Working location
Industry	Textual	Desired industries	Job industry
Job type	Textual	Current job title; Top 3 desired job titles	Job title
Description	Textual	Experience	Requirements

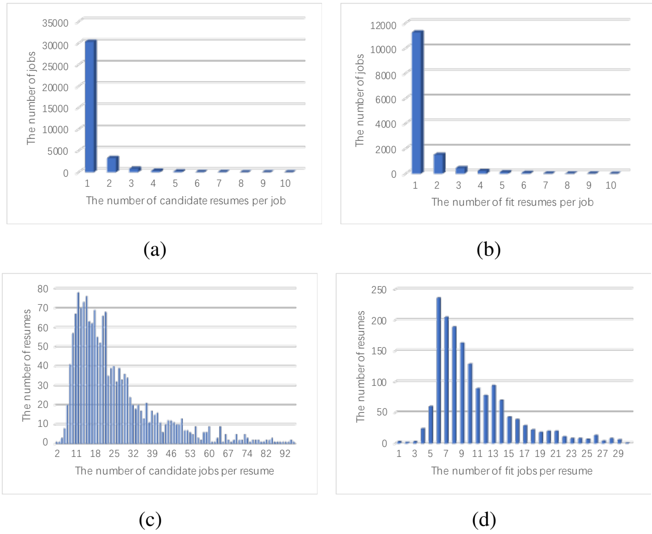

Fig. 3. Person-job fit distribution: (a) Candidate resumes per job. (b) Fit resumes per job. (c) Candidate jobs per resume. (d) Fit jobs per resume.

the outer interaction module, we set the size of embedding vectors to 32. The list of dimensions for the hidden layers in the MLP is $[224, 64, 32]$. Additionally, we set the batch size to 16 and use an early stop to prevent overfitting with a patience of 30 epochs.

3) Baselines: In order to verify the effectiveness of our proposed model PJFFRFI, we selected the following baselines, including 3 models with input features as text (t) and 6 models with input features as text and structure (s + t).

- PJFNN [2] uses a bipartite neural network to learn a joint representation of person-job fit from historical job applications, thereby mapping job posting and resume features to a shared latent representation.
- APJFNN [3] proposes a word-level semantic representation for both job requirements and jobseekers' experience based on RNN to exploit the rich information available at abundant historical job application data. BPJFNN could be treated as a simplified version of APJFNN model.
- Random Forests (RF), Gaussian Naive Bayes (GaussianNB) and Decision Trees (DT). We feed the embedding vectors for each resume and job posting from Section 3.3 as input to these models.
- MV-CoN [4] Utilizes a multi-view co-teaching network composed of text-based matching model and relation-based matching model to achieve representation enhancement and data enhancement.
- PJFFF [6] combines a CNN-based text learning representation module and a DeepFM-based entity semantic correlation analysis module for person-job fit.
- MUFFIN [7] groups all the features into several fields, and proposes two modules to learn the latent correlations in each fields and the field interactions.

4) Evaluation Metrics: We quantify the performance of each model using evaluation metrics commonly used in classification and recommendation tasks: accuracy, AUC, F1 score, precision, mean average precision (MAP).

TABLE II
COMPARISON OF THE EXPERIMENTAL RESULTS FOR PERSON-JOB FIT MODELS.

Model	Input Feature	Accuracy	AUC	F1 Score	Precision	MAP
PJFNN	t	0.523	0.5273	0.5064	0.6043	0.4844
BPJFNN	t	0.5842	0.6051	0.5338	0.6179	0.3349
APJFNN	t	0.6506	0.6942	0.5687	0.6991	0.5460
DT	t + s	0.512	0.5014	0.4287	0.5911	0.5301
RF	t + s	0.5383	0.4896	0.3061	0.5276	0.5900
Gaussian NB	t + s	0.5037	0.4953	0.4467	0.6248	0.6068
MV-CoN	t + s	0.6852	0.6638	0.5487	0.7487	0.6539
PJFFF	t + s	0.6543	0.7028	0.5391	0.7051	0.6217
MUFFIN	t + s	0.6725	0.6871	0.5634	0.7224	0.6638
PJFFRIL	**t + s**	**0.7066**	**0.7314**	**0.6056**	**0.7523**	**0.7017**

B. Experimental Results

The overall performance of the models is shown in Table 2, where the best results in terms of different evaluation metrics are marked in bold. Compared with other models, our proposed model PJFFRFI achieves the state-of-the-art results in all evaluation metrics. It indicates that our proposed feature interaction in three dimensions improves the performance of predicting matching scores between resumes and job postings. First, compared to models that only consider pure text features, the performance of models whose input features are text and structure features is generally higher than that of models with pure text features. Taking the best-performing APJFNN as an example, our proposed model improves 8.61%, 5.36%, 6.49%, 7.61%, 28.52% in terms of accuracy, AUC, F1 score, precision, MAP respectively, indicating that introduce of structure data does help to improve the performance of the model. Secondly, in models that considers structured data and text data, deep learning models far exceeds machine learning models such as DT, RF, Gaussian NB in all indicators, shows that deep learning is effective in semantic mining and structured data representation. Finally, compared with the models MUFFIN, PJFFF and MV-CoN whose model structure is similar to ours, our performance has also improved significantly. Compared to the muffin with the best overall performance in baselines, PJFFRFI improves 5.07%, 6.44%, 7.49%, 4.14%, 5.71% in five metrics, which explains that the model we propose is better at semantic representation and mining the implicit relationship in interaction features.

TABLE III
RESULTS OF ABLATION EXPERIMENTS.

Models	Accuracy	AUC	F1 Score	Precision	MAP
PJFFRFI (no_structure)	0.6546	0.6363	0.4791	0.6796	0.653
PJFFRFI (no_fem)	0.6673	0.6852	0.5169	0.6983	0.6751
PJFFRFI (no_inner)	0.6707	0.6961	0.5303	0.7249	0.6816
PJFFRFI (no_outer)	0.6562	0.6900	0.4517	0.6991	0.6775
PJFFRFI (no_residual)	0.6614	0.6852	0.5461	0.7025	0.6718
PJFFRFI	**0.7066**	**0.7314**	**0.6056**	**0.7523**	**0.7017**

C. Ablation Study

In this section, ablation experiments are conducted to study the contributions of some modules in PJFFRFI. Specially, we are interested in whether the structured data input and multi-dimensional feature interactions work and whether the residual connection network works. Therefore, we compare PJFFRFI with: 1) PJFFRFI without structured data input; 2) PJFFRFI without feature enhance module; 3) PJFFRFI without inner interaction module; 4) PJFFRFI without outer interaction module; 5) PJFFRFI without residual connection.

The results are shown in Table 3. It is apparent that removing the structured data input reduces the performance, indicating that structured data plays an important role in person-job fit. After removing PJFFRFI (no_fem), the overall performance of the model degrades 5.89%, 6.74%, 17.16%, 7.73%, 3.94% in five metrics respectively, which shows that FEM effectively identifies the implicit preference intention of recruiters and enhances it. We also observe that removing the FIM, performance decrease by 5.35%-7.68%, 5.07%-6%, 14.2%-34.07%, 3.78%-7.61%, 2.95%-3.57% respectively, verifying the effectiveness of extracting hidden features within and between domains. The comparison of results between PJFFRFI (no_residual) and PJFFRFI confirms that residual connection makes positive contributions.

D. Case Study

We further demonstrate the effectiveness of our model by case studies. Fig. 1 shows two candidate jobs for the same resume. The matching scores predicted for Job positing 1 and Job posting 2 are 0.9106 and 0.1654 respectively. It is shown that our model is able to predict the matching scores precisely. Referring to textual features marked by gray boxes, the bold matching words between Job posting 2 and Resume are more than Job posting 1. However, we can observe that in structured domains, Job posting 2 cannot meet the city expectations of the jobseeker and has ambiguous educational requirements, leading to its rejection. To sum up, our model is capable of adequately simulating the behavior of human resource managers, focusing not only on text matching but also on the interactions of informative structured features (e.g., Salary and City).

V. Conclusion

In this paper, we propose a novel model named PJFFRFI based on multi-domain feature representation and multi-dimensional feature interaction for person-job fit. PJFFRFI divides all features into multiple domains, maps numerical and categorical features to a low-dimensional space and learns semantic representation using pretrained ALBERT. Then, we design a feature enhancement module (FEM) and a feature interaction module (FIM) with inner and outer interactions to model the feature interactions in multiple dimensions, including the feature correlations in job postings, the feature interactions between resumes and job postings in the same domain and the hidden feature combinations between domains. Finally, we evaluate our model using a real recruitment dataset. Experimental studies confirm the superiority of PJFFRFI over existing models and verify the contributions of each module in PJFFRFI.

References

[1] Y. Chen and R. Pan, "Research on data analysis and visualization of recruitment positions based on text mining," *Advances in Multimedia*, vol. 2022, 2022.

[2] C. Zhu, H. Zhu, H. Xiong, C. Ma, F. Xie, P. Ding, and P. Li, "Person-job fit: Adapting the right talent for the right job with joint representation learning," *ACM Transactions on Management Information Systems (TMIS)*, vol. 9, no. 3, pp. 1–17, 2018.

[3] C. Qin, H. Zhu, T. Xu, C. Zhu, L. Jiang, E. Chen, and H. Xiong, "Enhancing person-job fit for talent recruitment: An ability-aware neural network approach," in *The 41st international ACM SIGIR conference on research & development in information retrieval*, 2018, pp. 25–34.

[4] S. Bian, X. Chen, W. X. Zhao, K. Zhou, Y. Hou, Y. Song, T. Zhang, and J.-R. Wen, "Learning to match jobs with resumes from sparse interaction data using multi-view co-teaching network," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 65–74.

[5] M. He, T. Wang, Y. Zhu, Y. Chen, F. Yao, and N. Wang, "Finn: Feature interaction neural network for person-job fit," in *2021 7th International Conference on Big Data and Information Analytics (BigDIA)*. IEEE, 2021, pp. 123–130.

[6] J. Jiang, S. Ye, W. Wang, J. Xu, and X. Luo, "Learning effective representations for person-job fit by feature fusion," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 2020, pp. 2549–2556.

[7] M. He, D. Shen, T. Wang, H. Zhao, Z. Zhang, and R. He, "Self-attentional multi-field features representation and interaction learning for person-job fit," *IEEE Transactions on Computational Social Systems*, 2021.

[8] J. Malinowski, T. Keim, O. Wendt, and T. Weitzel, "Matching people and jobs: A bilateral recommendation approach," in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, vol. 6. IEEE, 2006, pp. 137c–137c.

[9] M. Diaby, E. Viennet, and T. Launay, "Toward the next generation of recruitment tools: an online social network-based job recommender system," in *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*. IEEE, 2013, pp. 821–828.

[10] S. Bansal, A. Srivastava, and A. Arora, "Topic modeling driven content based jobs recommendation engine for recruitment industry," *Procedia computer science*, vol. 122, pp. 865–872, 2017.

[11] S. Ahmed, M. Hasan, M. N. Hoq, and M. A. Adnan, "User interaction analysis to recommend suitable jobs in career-oriented social networking sites," in *2016 International Conference on Data and Software Engineering (ICoDSE)*. IEEE, 2016, pp. 1–6.

[12] S. Yang, M. Korayem, K. AlJadda, T. Grainger, and S. Natarajan, "Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive statistical relational learning approach," *Knowledge-Based Systems*, vol. 136, pp. 37–45, 2017.

[13] S. Al-Otaibi and M. Ykhlef, "Hybrid immunizing solution for job recommender system," *Frontiers of Computer Science*, vol. 11, pp. 511–527, 2017.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[15] Y. Luo, H. Zhang, Y. Wen, and X. Zhang, "Resumegan: an optimized deep representation learning framework for talent-job fit via adversarial learning," in *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019, pp. 1101–1110.

[16] K. Yao, J. Zhang, C. Qin, P. Wang, H. Zhu, and H. Xiong, "Knowledge enhanced person-job fit for talent recruitment," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 2022, pp. 3467–3480.

[17] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942*, 2019.

[18] T. Mihaylov and A. Frank, "Discourse-aware semantic self-attention for narrative reading comprehension," *arXiv preprint arXiv:1908.10721*, 2019.

[19] H.-I. Liu and W.-L. Chen, "Re-transformer: a self-attention based model for machine translation," *Procedia Computer Science*, vol. 189, pp. 3–10, 2021.

[20] Y. N. Dauphin, A. Fan, M. Auli, and D. Grangier, "Language modeling with gated convolutional networks," in *International conference on machine learning*. PMLR, 2017, pp. 933–941.