Your Profile Reveals Your Traits in Talent Market: An Enhanced Person-Job Fit Representation Learning

Ling Jian
bebetter@upc.edu.cn

China University of Petroleum, East China
Chongzhi Rao
China University of Petroleum, East China
Xiao Gu
China University of Petroleum, East China

Research Article

Keywords: Person-Job Fit, Multifaceted Feature Fusion, Representation Learning, Contrastive Learning

Posted Date: March 21st, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4112217/v1

Additional Declarations: No competing interests reported.

Your Profile Reveals Your Traits in Talent Market: An Enhanced Person-Job Fit Representation Learning

Ling Jian[1*], Chongzhi Rao[1†] and Xiao Gu[1†]

[1*]China University of Petroleum (East China), School of Economics and Management, No. 66 Changjiang West Road, Qingdao, 266580, Shandong, China.

*Corresponding author(s). E-mail(s): bebetter@upc.edu.cn;
Contributing authors: s22080030@s.upc.edu.cn; b20080011@s.upc.edu.cn;
†These authors contributed equally to this work.

Abstract

Person-job fit lies at the heart of online recruitment, measuring the compatibility between job seekers and vacancies. Current researchers mainly focus on job-resume matching between job requirements and work experiences, presenting two notable limitations. Free text representation is constrained by word-level polysemy and sentence-level anisotropy. Also, overreliance on final representations hampers the exploration of inter-feature relationships. Towards this end, we proposed a novel **A**ttentive **P**erson-**J**ob **F**it **M**ultifaceted feature **F**usion model (APJFMF), aiming at obtaining more precise and comprehensive interactive person-job fit feature representations. The main contributions are: (1) Introduction of a method for multifaceted feature extraction and fusion from multi-source heterogeneous person-job data; (2) Enhanced the person-job free text representations through unsupervised fine-tuning of BERT; (3) Investigated global feature interactions by integrating diverse attention mechanisms. Extensive experiments on a real-world recruitment dataset confirmed the effectiveness of APJFMF and its individual components. The code is released at https://github.com/raochongzhi/APJFMF.

Acknowledgements: Youth Innovation Team of Higher Education Institutions in Shandong Province-Data Intelligence Innovation Team at China University of Petroleum and commissioned project “Smart Employment Recommendation Algorithms” by Qingdao JiuYe Jie Big Data Technology Co., Ltd under Grant No. HX20220639.

Keywords: Person-Job Fit, Multifaceted Feature Fusion, Representation Learning, Contrastive Learning

1 Introduction

Online recruitment has become the primary method for talent acquisition, offering a wealth of potential candidates through various online platforms [1]. Platforms like Boss Zhipin, Zhilian Zhaopin, LinkedIn, and Jiuye Jie have transformed and streamlined the recruitment process. However, the demand for employment has created an information gap between candidates and recruiters, which results in low-quality and inefficient recruitment experiences [2]. Therefore, there's an urgent need for effective person-job fit technologies to accurately match candidates with suitable job opportunities and improve the overall efficiency of talent recruitment.

To enhance the efficacy of online recruitment platforms, scholars have endeavored to improve person-job fit using techniques such as Collaborative Filtering (CF) [3]. However, a notable obstacle known as the "cold start" challenge emerges, wherein the deficiency of user-job interaction data leads to diminished recommendation quality, particularly when relying exclusively on behavioral data. To tackle this challenge, researchers are now concentrating on person-job fit representation learning, primarily emphasizing two crucial elements: free text (typically work experiences and job requirements) and entity (typically structured or semi-structured data).

In the realm of free text representation learning, studies utilize machine learning and deep learning methods. Additionally, diverse attention mechanisms are incorporated to enhance feature representations. Despite the considerable success achieved by existing research in this field, certain limitations persist. Some unitized word embedding methods like Word2Vec, overlook the challenge of words having multiple meanings in different contexts despite using the same word embedding [4, 5]. Others employed sentence embedding methods such as BERT and ALBERT but encountered anisotropic and non-uniform distribution dilemmas. In addition, current research predominantly emphasizes unstructured free text feature representation, often neglecting (semi-) structured fields. Moreover, the interaction between free text features and entity features is frequently neglected.

Along this line, we proposed a novel **A**ttentive **P**erson-**J**ob **F**it **M**ultifaceted feature **F**usion approach (named **APJFMF**). First, we standardized resume formats and used PaddleOCR for parsing and content recognition. We then applied web scraping, regular expressions, and Named Entity Recognition (NER) [6] to comprehensively extract multifaceted entities from multi-source heterogeneous data [7]. Second, after using a pre-trained Chinese BERT model [8] for each sentence in work experiences and job requirements, we fine-tuned these embeddings using unsupervised SimCSE [9]. After employing Bi-LSTM to capture the contextual relationships between sentences, we applied attention and co-attention mechanisms to capture hierarchical and interaction information. Third, after classifying the extracted multifaceted entities and applying different embedding methods to each category, we concatenate and input them into the DeepFM model [10] to obtain entity feature representations. Last, attention mechanisms [11] are employed to integrate the acquired free text and entity feature representations for person-job fit prediction.

Experimental results affirmed the superior performance of our approach compared to state-of-the-art models. Ablation studies were also conducted to assess the isolated effectiveness of each component. Furthermore, we examined the interpretability of the model at the sentence, interaction, and matching levels. Therefore, the main contributions of this study are as follows:

- Presenting a multifaceted entity feature extraction and feature fusion method to enhance the comprehensiveness of person-job fit feature representations.
- Fine-tuning BERT sentence embeddings in an unsupervised contrastive learning manner to enhance the accuracy of person-job fit feature representations.
- Proposing a novel model, APJFMF, which explores global feature interactions to enhance the hierarchy and interactivity of feature representation through the integration of diverse attention mechanisms.

The remainder of this work is organized as follows. In Section 2, we review the relevant work on person-job fit and text representation learning. Section 3 shows the notations and definitions of this paper. In Section 4, we then elaborate on our proposed model APJFMF in detail and introduce the methodology of learning features from free text and entities we extracted. We run experiments and evaluations in Section 5 and draw conclusions in Section 6.

2 Related Work

2.1 Person-Job Fit

Person-job fit relies on exploring the latent feature representations of jobs and candidates, primarily from two fields: unstructured and (semi-) structured data. Researchers mainly focus on learning the unstructured text representation of resume work experience and job descriptions. For instance, Zhu et al [12] designed two parallel Convolutional Neural Networks (CNNs) to independently learn the latent text representations. Bian et al [13] employed a hierarchical attention-based Recurrent Neural Network (RNN) for person-job fit. Yan et al [14] utilized a Gated Recurrent Unit (GRU) to enhance text representations. Some studies also utilized Graph Neural Networks (GNNs) to learn latent relational representations [15–17]. Additionally, certain studies have employed bipartite heterogeneous graphs to facilitate the matching of individuals with job positions [18, 19].

Traditional neural networks face challenges with information and memory loss when handling long sequences. Current research enhances the representations using diverse attention mechanisms [11]. Qin et al [20, 21] designed four hierarchical capability-aware attention strategies to enhance representation learning. Hou et al [22] integrated job representations into user intent embeddings using self-attention mechanisms. He et al [23] applied a multi-head self-attention mechanism to enhance feature interaction modeling. Wang et al [16] employed a co-attentive CNN and an attentive

GNN to learn representations. In the same year, [24] developed a bidirectional interactive graph neural network utilizing Graph Convolutional Networks (GCNs) to jointly model interaction graphs and textual graphs, thereby acquiring feature representations of users and job positions.

The mentioned methods excel in unstructured text but overlook (semi-) structured data. Jiang et al [25] extracted entities features from recruitment data and learned from them. Huang et al [26] introduced a co-attention mechanism to capture interactions among non-textual features. Though delved into (semi-)structured data representation, their treatment of unstructured data was basic. Furthermore, they often used a simple concatenation approach to combine them, with limited exploration of interactions. This study advanced by employing attention and co-attention neural networks for text representations, multifaceted feature extraction and representation techniques for entity representations, and fusing them with attention networks, achieving a more comprehensive and interactive person-job representation learning.

2.2 Text Representation Learning

Natural language sentences consist of words and phrases that follow grammatical rules and convey complex semantic information, involving both sequential and hierarchical elements that are crucial for understanding them. Existing research on person-job fit predominantly employs the Word2Vec method for sentence representations [27–31] but overlooked word positional information within sentences. Furthermore, some studies employ Doc2Vec to capture paragraph semantics and syntax but encounter quality issues due to limited data [32]. To effectively capture the internal structure of sentences, including sequential and dependency information, researchers utilized BERT to extract semantic features of sentences [17, 26]. Due to the extensive parameter size of BERT subsequent BERT variants [33], researchers utilized ALBERT [23], which significantly reduces model parameters while maintaining performance.

Despite the significant strides of BERT and ALBERT, sentence representations exhibit anisotropic and non-uniform distribution. SimCSE addresses this challenge through contrastive learning and self-supervised training using pre-trained BERT outputs [9]. It effectively narrows the gap between similar samples and widens the separation between dissimilar ones, enhancing sentence representation capabilities. SimCSE outperforms BERT in text representation learning tasks, but its application to person-job fit representation learning remains unexplored in current research. Thus, our study employs unsupervised SimCSE for training text representations of resume work experiences and job requirements.

3 Notations and Definitions

In real-world recruitment scenarios, recruitment data usually consists of job postings, resumes, and job application records. A resume includes personal information like gender, education, institutions, and work experiences (projects, internships, history work, etc.). A job posting contains important details like location, salary, working hours, and job requirements. Job application records contain historical data on candidates' applications.

Table 1 Notations and Definitions

Notation	Definition
R	The resume set
J	The job posting set
$\mathbf{r}_i$	The i-th resume in the resume set R
$\mathbf{j}_k$	The k-th job posting in the job posting set J
$\mathbf{f}_i$	The work experience in the i-th resume $\mathbf{r}_i$
$\mathbf{f}_k$	The job requirements in the k-th job posting $\mathbf{j}_k$
$\mathbf{f}_{i,t}$	The t-th sentence in the work experience $\mathbf{f}_i$
$\mathbf{f}_{k,t}$	The t-th sentence in the job requirements $\mathbf{f}_k$
$\mathbf{e}_{i,t}$	The t-th entities in the i-th resume $\mathbf{r}_i$
$\mathbf{e}_{k,t}$	The t-th entities in the k-th job posting $\mathbf{j}_k$
$\mathbf{h}_{i,t}$	BERT embedding of the t-th work experience in the resume $\mathbf{r}_i$
$\mathbf{h}_{k,t}$	BERT embedding of the t-th job requirements in the job posting $\mathbf{j}_k$
$\mathbf{E}_{i,t}$	SimCSE embedding of the t-th work experience in the resume $\mathbf{r}_i$
$\mathbf{E}_{k,t}$	SimCSE embedding of the t-th job requirements in the job posting $\mathbf{j}_k$
n	The number of work experiences in the resume $\mathbf{r}_i$
N	The number of job requirements in the job posting $\mathbf{j}_k$
m	The number of entities in the resume $\mathbf{r}_i$
M	The number of entities in the job posting $\mathbf{j}_k$

Therefore, to define the problem of person-job fit, we use $\mathbf{r}_i = \{\mathbf{f}_i, \mathbf{e}_i\} \in R$ to denote the i-th resume in the resume set R, where $\mathbf{f}_i = \{\mathbf{f}_{i,1}, \mathbf{f}_{i,2}, ..., \mathbf{f}_{i,n}\}$ denotes the n pieces of work experiences in the resume, and $\mathbf{e}_i = \{\mathbf{e}_{i,1}, \mathbf{e}_{i,2}, ..., \mathbf{e}_{i,m}\}$ represents the m entities in the resume. Similarly, we use $\mathbf{j}_k = \{\mathbf{f}_k, \mathbf{e}_k\} \in J$ to denote the k-th job posting in the job posting set J, where $\mathbf{f}_k = \{\mathbf{f}_{k,1}, \mathbf{f}_{k,2}, ..., \mathbf{f}_{k,N}\}$ is denoted as the N pieces of job requirements in the job posting, while $\mathbf{e}_k = \{\mathbf{e}_{k,1}, \mathbf{e}_{k,2}, ..., \mathbf{e}_{k,M}\}$ is denoted as the M entities in the job posting. Finally, for each resume-job pair $s_l = (\mathbf{r}_i, \mathbf{j}_k, l_{i,k}) \in S$ in the historical application set S, we use $l_{i,k} \in \{0, 1\}$ to indicate whether the resume and the job is matching. To enhance clarity, important notations and their definitions are presented in Table 1.

4 Methodology

Our APJFMF model, as shown in Fig. 1, consists of four main components: feature extraction, free text feature representation, entity feature representation, and person-job fit prediction. Due to the multi-source heterogeneous nature of the person-job fit dataset, which includes multiple data sources (HTML, online documents, SQL database) and various data structures (structured, semi-structured, unstructured data), we utilize techniques such as web scraping, regular expressions, and Named Entity Recognition (NER) to comprehensively extract multifaceted entities from (semi-) structured data in the feature extraction phase (see Section 5.2). Then, free text features such as work experiences and job requirements undergo BERT-based processing and SimCSE fine-tuning to improve sentence-level feature representations. Entity features are categorized and processed, then input into the DeepFM module to obtain entity representation. Finally, An attention mechanism is applied to learn a joint feature representation of text and entities, which is further processed by a fully connected layer to predict the person-job matching probability.

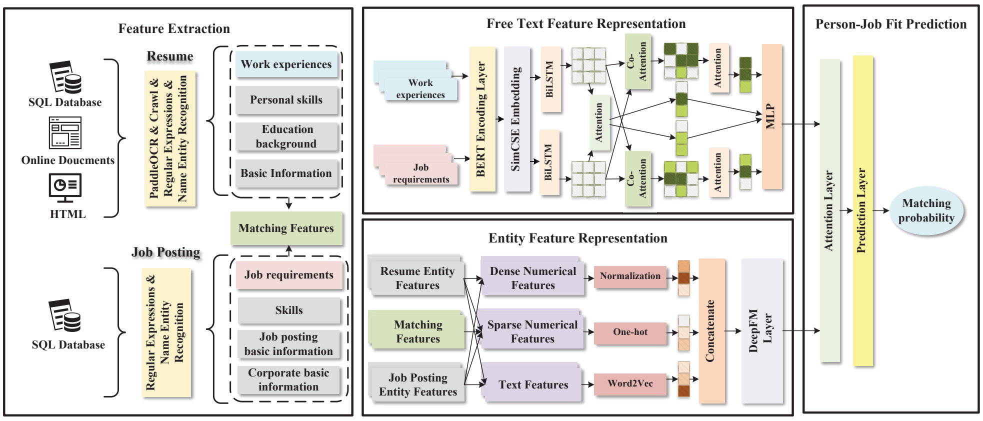

Fig. 1 A illustration of the architecture of the proposed model APJFMF, which can be separated into four parts: Feature Extraction, Free Text Feature Representation, Entity Feature Representation, and Person-Job Fit Prediction. Two different feature representation methods are used to learn free text feature representation and entity feature representation separately.

4.1 Learning Features from Free Text

For features from free text, we apply the BERT Chinese pre-trained model, as introduced by [34], to conduct sentence-level feature representation for work experience and job requirements. For each work experience sentence in the i-th resume $\mathbf{f}_i = \{\mathbf{f}_{i,1}, \mathbf{f}_{i,2}, ..., \mathbf{f}_{i,n}\}$ and each job requirement in the k-th job posting $\mathbf{f}_k = \{\mathbf{f}_{k,1}, \mathbf{f}_{k,2}, ..., \mathbf{f}_{k,N}\}$, we utilize the [CLS] vector of the BERT model as the sentence-level representation, as shown in Eq. 1:

$$\begin{aligned} \mathbf{h}_{i,t} &= BERT_{[cls]}(\mathbf{f}_{i,t}), \\ \mathbf{h}_{k,T} &= BERT_{[cls]}(\mathbf{f}_{k,T}). \end{aligned} \tag{1}$$

Then, we proceed with fine-tuning the BERT pre-trained sentence embeddings through the unsupervised SimCSE method [9]. The concept of unsupervised SimCSE involves feeding a set of sentences into Transformer models where dropout is applied to the fully-connected layers and attention probabilities. By inputting the same sentence $\mathbf{h}_i$ into the encoder twice, we obtain positive sample pairs $\mathbf{h}_i, \mathbf{h}_i^+$, while negative sample pairs $\mathbf{h}_i, \mathbf{h}_j^+$ are constructed by randomly selecting other sentences from the batch. The loss function is defined as shown in Eq. 2:

$$\ell_i = -\log \frac{e^{\sin(\mathbf{h}_i, \mathbf{h}_i^+)/\tau}}{\sum_{j=1}^{N} e^{\sin(\mathbf{h}_i, \mathbf{h}_j^+)/\tau}}, \tag{2}$$

by minimizing the loss ℓ_i, the aim is to bring similar sentence closer together while pushing dissimilar one apart, effectively learning data representations.

Following the fine-tuning of the BERT model via the SimCSE approach, we choose the last hidden state yielded by the model as the sentence representation. This refined sentence embedding can be calculated as $\mathbf{E}_i = \{\mathbf{E}_{i,1}, \mathbf{E}_{i,2}, ..., \mathbf{E}_{i,n}\}$ and $\mathbf{E}_k = \{\mathbf{E}_{k,1}, \mathbf{E}_{k,2}, ..., \mathbf{E}_{k,N}\}$ by Eq. 3:

$$\begin{aligned} \mathbf{E}_{i,t} &= SimCSE(\mathbf{h}_{i,t})[:, 0], \\ \mathbf{E}_{k,T} &= SimCSE(\mathbf{h}_{k,T})[:, 0]. \end{aligned} \tag{3}$$

Subsequently, since work experience and job requirements are both composed of paragraphs where sentences may have contextual dependencies, we incorporated a Bidirectional Long Short-Term Memory (BiLSTM) layer to learn the contextual information between sentences. This enables the model to capture the inter-sentence relationships effectively, enhancing the overall representation of the text data. We can calculated the final sentence representation $\mathbf{H}_i = \{\mathbf{H}_{i,1}, ..., \mathbf{H}_{i,n}\}$ and $\mathbf{H}_k = \{\mathbf{H}_{k,1}, ..., \mathbf{H}_{k,N}\}$ by Eq. 4:

$$\begin{aligned} \mathbf{H}_{i,t} &= BiLSTM\left(\mathbf{E}_{i,1:n}, t\right), \forall t \in [1, ..., n], \\ \mathbf{H}_{k,T} &= BiLSTM\left(\mathbf{E}_{k,1:N}, T\right), \forall T \in [1, ..., N], \end{aligned} \tag{4}$$

where $\mathbf{E}_{i,1:n}$ and $\mathbf{E}_{k,1:N}$ denote the sentence vectors input sequences of $\mathbf{E}_i$ and $\mathbf{E}_k$. And $\mathbf{H}_{i,t}$ and $\mathbf{H}_{k,T}$ are sentence representations of the t-th sentence in the work experience $\mathbf{f}_i$ and the t-th sentence in the job requirements $\mathbf{f}_k$.

Next, we developed an attention module to prioritize essential features in the model. Taking the resume part as an example: we begin by feeding the sentence representations into a fully connected layer. With the utilization of the tanh activation function and softmax normalization, attention scores $\alpha_{i,t}$ are calculated for each sentence representation, as depicted in Eq. 5:

$$\begin{aligned} \alpha_{i,t} &= \frac{exp(g_{i,t})}{\sum_{t=1}^{n} exp(g_{i,t})}, \\ g_{i,t} &= \mathbf{v}_\alpha^T tanh(\mathbf{w}_\alpha \mathbf{H}_{i,t} + \mathbf{b}_\alpha), \end{aligned} \tag{5}$$

where $\mathbf{v}_\alpha$, $\mathbf{w}_\alpha$, and $\mathbf{b}_\alpha$ are the parameters to be learned during the training process. $\alpha_{i,t}$ is the attention score of the sentence $\mathbf{H}_{i,t}$ among all the sentences in the work experience. As shown in Eq. 6, by taking the sum of the products of each sentence representation and its corresponding attention score, we obtain the final attention representation for the work experiences of the resume:

$$\mathbf{v}_i = \sum_{t=1}^{n} \alpha_{i,t} \mathbf{H}_{i,t}, \tag{6}$$

where $\mathbf{v}_i$ is the attention-based representation of the work experiences $\mathbf{f}_i$ in the resume $\mathbf{r}_i$. Similarly, $\mathbf{v}_k$ is the attention-based representation of the job requirements $\mathbf{f}_k$ in the job posting $\mathbf{j}_k$, which can be calculated in the same process. This attentiveness mechanism enhances the ability of the model to focus on critical information.

Moreover, we devised a co-attention module to investigate the hierarchical importance of feature interactions between resumes and job postings. Taking the resume part as an example, we begin by feeding the sentence representations of both work experiences and job requirements into a fully connected layer. With the utilization of the tanh activation function and softmax normalization, co-attention scores $\gamma_{i,T,t}$ are calculated for each sentence $\mathbf{H}_{k,T}$ in job requirements to each sentence $\mathbf{H}_{i,t}$ in work experiences, as depicted in Eq. 7:

$$\begin{aligned} \gamma_{i,T,t} &= \frac{exp(g_{i,T,t})}{\sum_{t=1}^{n} exp(g_{i,T,t})}, \\ g_{i,T,t} &= \mathbf{v}_{\beta}^{T} tanh(\mathbf{W}_{\beta} \mathbf{H}_{k,T} + \mathbf{U}_{\gamma} \mathbf{H}_{i,t}), \end{aligned} \tag{7}$$

where $\mathbf{v}_\beta$, $\mathbf{W}_\beta$ and $\mathbf{U}_\gamma$ are the parameters to be learned during the training process. As shown in Eq. 8, by taking the sum of each work experience sentence representation $\mathbf{H}_{i,t}$ and its corresponding co-attention score $\gamma_{i,T,t}$, we obtain the co-attention representation for the resume part.

$$\mathbf{x}_{i,T} = \sum_{t=1}^{n} \gamma_{i,T,t} \mathbf{H}_{i,t}, \tag{8}$$

where $\mathbf{x}_{i,T}$ is the representation of the work experiences in the resume towards the T-th job requirement $\mathbf{H}_{k,T}$ in the job posting based on the co-attention mechanism. Similarly, $\mathbf{x}_{k,t}$ is the representation of the job requirements in the job posting towards the t-th work experience $\mathbf{H}_{i,t}$ in the resume based on the co-attention mechanism, which can be calculated in the same process.

The obtained representation sequences $\{\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, ..., \mathbf{x}_{i,N}\}$ and $\{\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, ..., \mathbf{x}_{k,n}\}$ are fed into an attention module to obtain the final co-attention based feature representations. As shown in Eq. 9, we take the resume part for example

$$\begin{aligned} \zeta_{i,T} &= \frac{exp(g_{i,T})}{\sum_{T=1}^{N} exp(g_{i,T})}, \\ g_{i,T} &= \mathbf{v}_{\varepsilon}^{T} tanh(\mathbf{W}_{\varepsilon} \mathbf{x}_{i,T} + \mathbf{b}_{\varepsilon}), \\ \mathbf{x}_i &= \sum_{T=1}^{N} \zeta_{i,T} \mathbf{x}_{i,T}, \end{aligned} \tag{9}$$

where the parameters $\mathbf{v}_\varepsilon$, $\mathbf{W}_\varepsilon$ and $\mathbf{b}_\varepsilon$ are subject to learning during the training process. $\zeta_{i,T}$ is the attention score of the representation sequences $\{\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, ..., \mathbf{x}_{i,N}\}$. $\mathbf{x}_i$ is the final co-attention-based feature representation of the work experiences $\mathbf{f}_i$ in the resume $\mathbf{r}_i$. Similarly, $\mathbf{x}_k$ is the final co-attention-based feature representation of the job requirements $\mathbf{f}_k$ in the job posting $\mathbf{j}_k$, which can be calculated with the same process.

We concatenated the results of the attention and co-attention modules and fed them into an MLP layer for subsequent fusion with entity representations, depicted in the Eq. 10:

$$\mathbf{X}^{text} = MLP([\mathbf{v}_i; \mathbf{v}_k; \mathbf{v}_i - \mathbf{v}_k; \mathbf{v}_i \odot \mathbf{v}_k; \mathbf{x}_i; \mathbf{x}_k; \mathbf{x}_i - \mathbf{x}_k; \mathbf{x}_i \odot \mathbf{x}_k]). \tag{10}$$

4.2 Learning Features from Entities

Given the intricate interactions among the extracted multifaceted entity features, DeepFM stands out by amalgamating the linear and pairwise interaction capabilities of Factorization Machines (FM) with the intricate and nonlinear capabilities of Deep Neural Networks (DNN). Its proficiency lies in discerning complex patterns and dependencies within input features. Consequently, we utilized DeepFM to address the entity feature representation learning task.

For features extracted from Section 5.2, we classified them into numerical (dense and sparse) and textual groups. Dense numerical features undergo min-max normalization, resulting in the processed feature representation denoted as $\mathbf{v}_{dense}$. Sparse numerical features undergo one-hot encoding, producing the encoded feature representation referred to as $\mathbf{v}_{sparse}$. Textual features undergo Word2Vec embedding [35], followed by mean pooling to obtain the feature representation represented as $\mathbf{v}_{text}$. Feature fusion combines these representations as input to the DeepFM model, as depicted in the Eq. 11:

$$\mathbf{v}^{entity} = [\mathbf{v}_{dense}; \mathbf{v}_{sparse}; \mathbf{v}_{text}] \,. \tag{11}$$

The FM component, which was originally proposed by [36], is designed for learning feature interactions. We use the FM component to learn the first-order features of the entities, as Eq. 12:

$$\mathbf{X}_{FM} = \left\langle \mathbf{w} * \mathbf{v}^{entity} \right\rangle + \sum_{j_1=1}^{d} \sum_{j_2=j_1+1}^{d} \left\langle \mathbf{V}_i^{entity}, \mathbf{V}_j^{entity} \right\rangle \mathbf{v}_{j_1}^{entity} \cdot \mathbf{v}_{j_2}^{entity}, \tag{12}$$

where $\mathbf{w}$ is the weight vector, $\mathbf{V}_i^{entity}$ and $\mathbf{V}_j^{entity}$ are the output of the dense layer. $\mathbf{X}_{FM}$ is the output of the FM component.

The DNN component, which is a feed-forward neural network, is specifically employed to learn high-order feature interactions. We use the DNN component to learn the second-order features of the entities, as Eq. 13 shows:

$$\begin{aligned}\mathbf{a}^{(0)} &= [\mathbf{e}_1, \mathbf{e}_2, ...],\\ \mathbf{a}^{(l+1)} &= \sigma(\mathbf{W}^{(l)}\mathbf{a}^{(l)} + \mathbf{b}^{(l)}),\\ \mathbf{X}_{DNN} &= \sigma(\mathbf{W}^{|H|+1} \cdot \mathbf{a}^{H} + \mathbf{b}^{|H|+1}),\end{aligned} \tag{13}$$

where $\mathbf{a}^{(0)}$ is the output of embedding layer which is fed into deep neural network, $\mathbf{a}^{(l)}$, $\mathbf{W}^{(l)}$, $\mathbf{b}^{(l)}$ are the output, model weight and bias of the l-th layer, and $|H|$ is the number of hidden layers. $\mathbf{X}_{DNN}$ is the output of the DNN component.

Finally, we sum the FM component output $\mathbf{X}_{FM}$ and the DNN component output $\mathbf{X}_{DNN}$ to obtain the final representation of entity features $\mathbf{X}^{entity}$, as shown in Eq. 14:

$$\mathbf{X}^{entity} = \sum_i \mathbf{X}_i, i \in [FM, DNN]. \tag{14}$$

4.3 Person-Job Fit Prediction

To capture global interactions between free text and entity, we leveraged the feature representations acquired from both free text and entity $\left\{\mathbf{X}^{text}; \mathbf{X}^{entity}\right\}$ as the input to the attention layer, as depicted in Eq. 15:

$$\begin{aligned}&\mu_i = softmax\left(\mathbf{v}_o^T tanh\left(\mathbf{W}_o\mathbf{X}^i + \mathbf{b}_i\right)\right),\\ &D = \sum_i \mu_i \mathbf{X}^i, i \in [text, entity],\\ &\hat{Y} = \sigma(\mathbf{W}_\tau D + \mathbf{b}_\tau),\end{aligned} \tag{15}$$

where $\mathbf{v}_o^T$, $\mathbf{W}_o$, $\mathbf{b}_i$, $\mathbf{W}_\tau$ and $\mathbf{b}_\tau$ are the training parameters. μ_{text} and μ_{entity} are the attention score of $\mathbf{X}^{text}$ and $\mathbf{X}^{entity}$, respectively. And the sigmoid function $\sigma(\cdot)$ is employed to map the matching score into the range of [0,1]. Meanwhile, we adopted the binary cross-entropy loss to optimize our model.

5 Experiment

5.1 Data Description

In this paper, the large-scale dataset in our study was sourced from Qingdao Jiuye Jie Big Data Technology Co., Ltd up until April 24, 2023.

Considering the confidentiality of post-interview results within companies, undisclosed to the recruitment platform, we consider the assessment outcomes of resume submissions as implicit indicators of successful application. This approach is based on the understanding that a successful alignment leads to profile submission action from candidates and application approval action from recruiters, and rejections during interviews may involve factors beyond a mismatch, such as the applicant's interview performance.

Table 2 The Statistics of the Dataset

Statistics	Value
Number of resumes	16,722
Number of job postings	58,863
Number of successful applications	47,092
Average sentences per job requirement in job posting	9.77
Average sentences per work experience in resume	10.46
Average skill entities per job posting	3.74
Average skill entities per resume	9.76

We enhanced the negative sampling process in alignment with prior research on person-job fit, incorporating control over the category of positions alongside random 1:1 negative sampling. Specifically, we randomly selected an equal number of job postings from positions not previously submitted by the candidate and belonging to different second-level categories than those already submitted.

Table 2 shows the statistics of the dataset. To summarize, our dataset comprises 16,722 resumes and 58,863 job postings. After excluding 10,581 unsuccessful applications, we are left with a total of 47,092 successful applications. The mean count of sentences within job requirements and resume work experience is approximately 10. The average number of skill entities extracted from job postings is around 4, while for resumes, it is approximately 10.

5.2 Feature Extraction

For resumes, we started by standardizing resume formats, then used the Chinese OCR tool, PaddleOCR by Baidu PaddlePaddle [37], to extract resume text. We used regular expressions to divide resume text into four sections: work experience, personal skills, educational background, and basic information. Work experience was used for free text feature representation, while personal skills underwent Named Entity Recognition (NER) for skill extraction. Educational background and basic information were processed with regular expressions to extract gender, major, school, education, and location. To enhance information related to the "school" entity, we employed web crawlers to rank scores of universities from the 2023 *Best Chinese Universities Ranking* website [38]. Similarly, for job postings, we use regular expressions to extract location information and employ named entity recognition for skill extraction, following the same procedure as in the resume section.

To enhance entity representation, we use three matching features: degree match (1 if candidate's education meets job requirements, 0 otherwise), work location match (1 if candidate's location aligns with job location, 0 otherwise), and corporate location match (1 if candidate's location matches company headquarters, 0 otherwise).

5.3 Experiment Setting

Learning Features from Free Text Setting: In our experiments, we set a maximum sentence limit of 10 for both work experiences and job requirements based on statistical findings in Table 2. For sentences exceeding this limit, we applied truncation; for those falling short, we used padding. Then, we utilized a Chinese BERT pre-training model [8] for sentence embeddings, followed by unsupervised SimCSE training. Training parameters included a learning rate of 1e-5, a dropout rate of 0.3, and a batch size of 64, resulting in 768-dimensional embeddings for each sentence. Processed embeddings underwent a BiLSTM layer with an output dimension of 128, followed by the concatenation from attention and co-attention according to the Eq. 9. After obtaining a 1024-dimensional representation, we added an MLP layer with a dropout rate of 0.7, culminating in a singular 1-dimensional representation.

Learning Features from Entities Setting: For each skill entity extracted, we employed the skip-gram algorithm from Word2Vec with a vector dimension set at 8. Since the number of skill entities varies, unlike padding and truncation for sentences, we used the column-wise average matrix of word embeddings from Word2Vec as the embedding representation. The hidden layer dimensions in the FM module were set as [128, 64, 32]. In the DNN module, a dropout rate of 0.7 was added to prevent overfitting. Finally, by concatenating the 1-dimensional representations from both the FM and DNN modules, we obtained a singular 1-dimensional representation.

Person-Job Fit Prediction Setting: During training, we optimized the model by using the Adam algorithm. To address overfitting, a weight decay of 1e-5 and a dropout rate of 0.7 in the fully connected layer were implemented. The batch size was set at 128, and the learning rate was finely adjusted to 5e-3. The model was trained on a server with a 12-core CPU, 60 GB RAM, and a GeForce RTX 4090 GPU.

5.4 Performance Comparison

To assess the performance of our proposed APJFMF model, we chose several models as baselines, including classic classification models and state-of-the-art models in the field. For the traditional classification model, we included Logistic Regression (**LR**), Naive Bayes (**NB**), Decision Tree (**DT**), and AdaBoost (**AB**). For these methods, the average vector of all word vectors in a resume (or a job posting) was treated as the latent vector and then combined as the input. We also incorporate deep neural network baselines, considered state-of-the-art models for person-job fit.

- **PJFNN** [12]: PJFNN was a person-job fit model based on Word2Vec embedding and a two-layer Convolutional Neural Network (CNN).
- **APJFNN** [20]: APJFNN treated each resume experience and job requirement as separate sequences, employing four hierarchical ability-aware attention techniques to develop word-level representations for both.
- **BPJFNN** [20]: BPJFNN was a streamlined iteration of APJFNN, which employed Bidirectional Long Short-Term Memory (BiLSTM) networks to grasp the semantic essence of each word within both sequences.
- **JRMPM** [14]: JRMPM utilized a profiling memory module to investigate the preferences of both job providers and job seekers, subsequently enhancing their respective representations.

Table 3 Performance Comparison for Person-Job Fit

Model	AUC(%)	ACC(%)	Precision(%)	Recall(%)	F1(%)
NB	59.44	59.59	65.83	38.66	48.71
AB	65.89	65.91	66.39	63.46	64.89
LR	67.08	67.10	67.44	65.19	66.30
DT	72.27	72.26	71.30	73.85	72.55
JRMPM	57.92	57.88	60.33	60.34	52.73
PJFNN	74.39	74.40	74.28	74.28	74.16
BPJFNN	75.34	75.32	75.34	75.34	76.79
APJFNN	77.30	76.95	75.61	<u>79.07</u>	76.97
IPJF	78.38	78.37	77.70	77.70	78.53
INEXIT	<u>80.34</u>	<u>80.32</u>	<u>78.34</u>	78.34	<u>80.79</u>
APJFMF	**84.46**	**84.06**	**85.04**	**85.04**	**84.41**
Imp	5.13%	5.15%	8.55%	7.55%	4.48%

- **IPJF** [39]: IPJF was an interpretable person-job fit model that leveraged deep interactive representation learning to automatically learn the interdependence between a resume and job requirements.
- **INEXIT** [40]: INEXIT was a person-job fit model that explored the internal and external interactions for semi-structured multivariate attributes.

The results are presented in Table 3, demonstrating a clear superiority of our model over all the baselines by a substantial margin.

5.5 Ablation Study

To evaluate the impact of each component, we systematically removed individual core elements within APJFMF and compared their performance against the complete model.

- **APJFMF**$_{-SimCSE}$: APJFMF removed the SimCSE fine-tuning module, and directly employed the results from the pre-trained BERT model as the sentence representations.
- **APJFMF**$_{-DeepFM}$: APJFMF removed the entity representation module, meaning the DeepFM-based learning of entity representation was excluded, and only textual representations were employed for person-job fit.
- **APJFMF**$_{-freetext}$: APJFMF removed the free text representation module, meaning it relied solely on the learning of entity representations.
- **APJFMF**$_{-attention}$: APJFMF removed the attention mechanism module in the context of free text representation learning, and exclusively employed the representation derived from the co-attention mechanism module.
- **APJFMF**$_{-coAttention}$: APJFMF removed the co-attention mechanism module in the context of free text representation learning, and exclusively employed the representation derived from the attention mechanism module.

Table 4 presents the ablation study results for our model APJFMF. Removing the DeepFM module resulted in a 2.37% decrease in AUC, indicating that (semi-)structured data also carries valuable information. The 0.403% AUC decrease without

Table 4 Ablation Study of Our Model APJFMF

Model	AUC(%)	ACC(%)	Precision(%)	Recall(%)	F1(%)
APJFMF	**84.46**	**84.06**	**85.04**	**85.04**	**84.41**
APJFMF$_{-DeepFM}$	82.46	82.46	82.06	82.06	82.46
APJFMF$_{-attention}$	84.12	84.11	84.44	84.44	84.02
APJFMF$_{-coAttention}$	83.45	83.45	83.60	83.60	83.39
APJFMF$_{-SimCSE}$	82.27	82.27	81.81	81.81	82.32
APJFMF$_{-freetext}$	80.11	80.10	81.38	78.35	79.84

the attention module is because it focuses on important sentences, enhancing the grasp of the model in free text. The larger 1.20% AUC decrease without the co-attention module is due to its emphasis on the interaction between sentences, more directly enhancing the adaptability of the model in person-job fit tasks. Using BERT directly without fine-tuning resulted in a substantial 2.59% AUC decline. This is due to SimCSE, which enhances the accuracy of BERT representations by bringing similar samples closer and pushing dissimilar samples apart. Lastly, excluding the free text representation module showed a significant 5.15% AUC decline, highlighting the crucial role of free text in person-job fit.

5.6 Case Study

By utilizing the attention strategies we've proposed, our goal goes beyond improving matching performance. In this subsection, we'll clarify the matching results at three different levels to enhance interpretability.

Sentense-level: In our initial investigation, we aimed to evaluate the model's ability to highlight the most crucial sentences. The histogram in Fig. 2 illustrates the importance of each sentence. The length of the histogram corresponds to the sentences' importance within the entire paragraph, determined by attention scores α. The results indicate that the second sentence is the most significant, representing the job's unique characteristics directly. Conversely, the seventh and ninth sentences receive lower importance scores, offering less distinctive insights into job-specific features. Thus, our model effectively captures variations in sentence importance.

Interaction-level: We investigate the model's capability to capture feature interaction representations. Fig. 3 shows a matrix in which each cell visually presents the co-attention scores γ between each sentence in the resume and the job posting, with darker shades signifying higher scores. The results highlight the model's ability to capture complex interactions among features. Notably, the fifth sentence in the resume shows a strong alignment with the first sentence in the job posting requirements, mainly due to its semantic connection with "customer communication".

Matching-level: Our proposed solution's effectiveness in person-job fit is demonstrated in Fig. 4, where a resume achieves a matching probability of 0.8323 for job posting 1 and 0.2370 for job posting 2. Skill entities in the left panel are highlighted in purple. To understand the significant score difference, we conducted a detailed analysis from both textual and entity perspectives. Textual analysis revealed that the job seeker's background primarily revolves around the logistics sector, especially

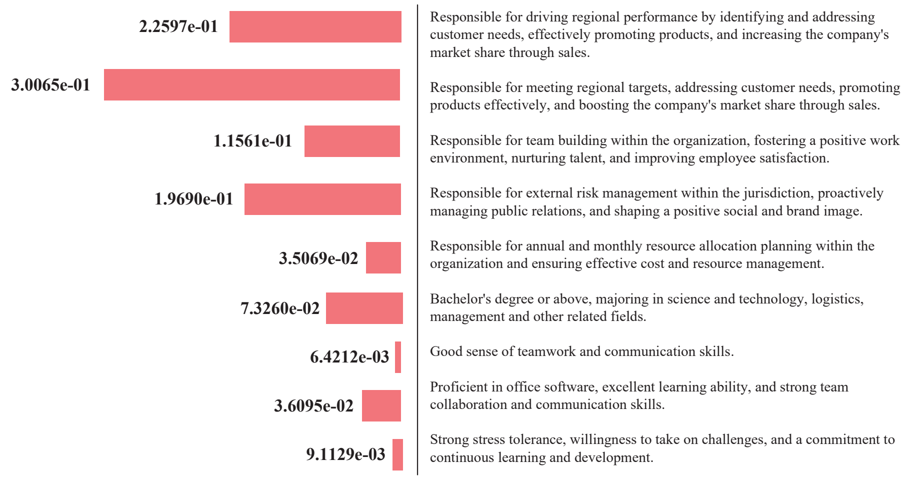

Fig. 2 An example showcasing the advantage of attention α in assessing the significance of each job requirement. The left bar charts represent the distribution of α values across all requirements.

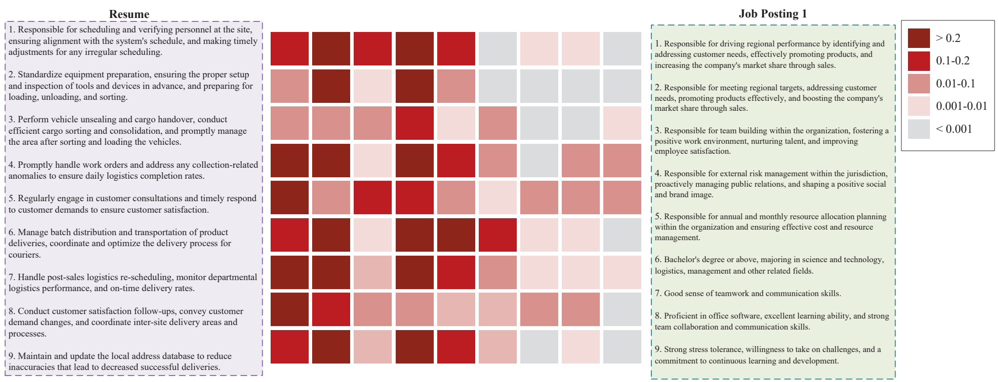

Fig. 3 An example showcasing the advantage of co-attention γ in assessing the significance between each resume and job requirement. The middle matrix represents the distribution of γ values across all requirements.

in transportation management and logistics planning. This aligns well with the job responsibilities and requirements in Job Posting 1, which focuses on logistics and management, while Job Posting 2 emphasizes blueberry cultivation with some management aspects. This empirical evidence confirms the precision of our model in predicting alignment between job postings and resumes.

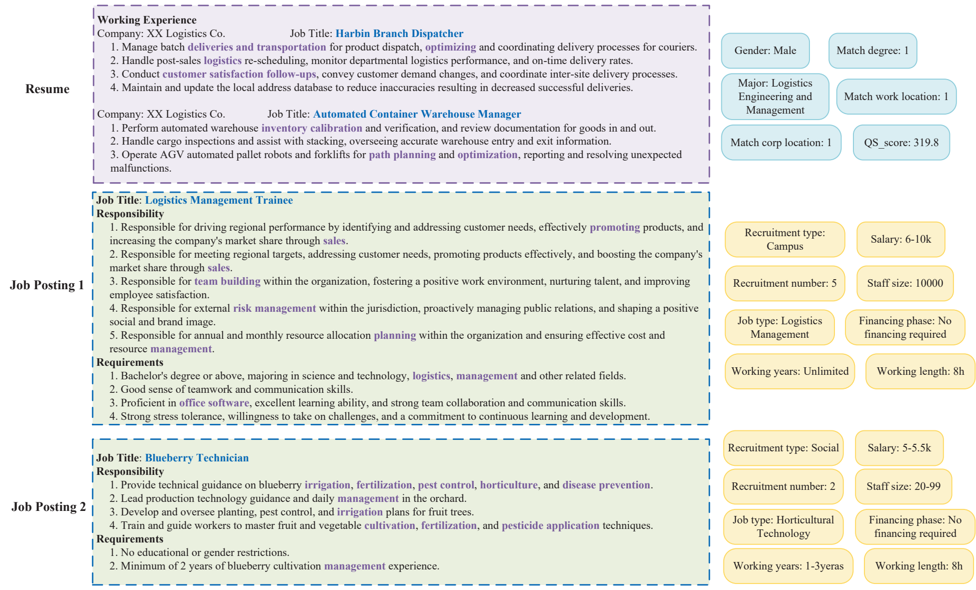

Fig. 4 An example demonstrating the matching effectiveness between a resume and job postings, where the resume has a matching probability of 0.8323 with the job posting 1 and a matching probability of 0.2370 with the job posting 2.

6 Conclusion

In this paper, we proposed a novel **A**ttentive **P**erson-**J**ob **F**it **M**ultifaceted feature **F**usion model (APJFMF). We hope the model will improve online recruitment performance by learning more precise and comprehensive interactive person-job fit representations. Specifically, to address the challenges of free text representation learning in person-job fit, we employed an unsupervised contrastive learning approach to fine-tune the representations of BERT. After utilizing BiLSTM to learn the contextual relationships between sentences, we seamlessly integrated attention and co-attention mechanisms. The attention mechanism captures sentence hierarchy, while the co-attention mechanism reveals resume-job interactions, enhancing the person-job fit representation learning. Additionally, we extracted multifaceted features from diverse data sources and applied the DeepFM method for entity feature fusion. Finally, we utilized attention mechanisms to fuse features between the free text and entity components, thereby exploring global feature interactions. We validated APJFMF using a large-scale real-world dataset from the online recruitment service platform. Our method outperformed several state-of-the-art models in extensive experiments, validating the contributions made by each module of APJFMF. In the future, we will incorporate user-historical application behaviors and job-historical acceptance and rejection behaviors into the model to explore the significance of user preferences and recruiter preferences in the person-job fit task.

Declarations

Data availability statements: The code pertinent to this study has been made publicly accessible in an open-source repository on GitHub, available at: https://github.com/raochongzhi/APJFMF. Due to the highly sensitive nature of personal resume information and the data confidentiality agreement signed between the researchers of this paper and the relevant enterprise, the data portion cannot be disclosed to the public. Nevertheless, this study will provide the format of the model's data input, allowing subsequent researchers to utilize their own acquired data to conduct research.

References

1. Wu L, Qiu Z, Zheng Z, et al (2023) Exploring large language model for graph data understanding in online job recommendations. arXiv preprint arXiv:230705722 https://doi.org/10.48550/arXiv.2307.05722

2. Ramanath R, Inan H, Polatkan G, et al (2018) Towards deep and representation learning for talent search at linkedin. In: Proceedings of the 27th ACM international conference on information and knowledge management, pp 2253–2261, https://doi.org/10.1145/3269206.3272030

3. Chen J, Yuan B, Jin C, et al (2022) Design and implementation of employee recommendation system based on neural graph collaborative filtering. In: International Conference on Image, Vision and Intelligent Systems, Springer, Singapore, pp 784–792, https://doi.org/10.1007/978-981-99-0923-0_78

4. Fu B, Liu H, Zhu Y, et al (2021) Beyond matching: Modeling two-sided multi-behavioral sequences for dynamic person-job fit. In: Database Systems for Advanced Applications. DASFAA 2021, Springer, Taipei, Taiwan, pp 359–375, https://doi.org/10.1007/978-3-030-73197-7_24

5. Xu A, Jian L (2023) A deep news headline generation model with reinforce filter. In: 2023 International Joint Conference on Neural Networks (IJCNN), IEEE, pp 1–7, https://doi.org/10.1109/IJCNN54540.2023.10192007

6. Fareri S, Melluso N, Chiarello F, et al (2021) Skillner: Mining and mapping soft skills from any text. Expert Systems with Applications 184:115544. https://doi.org/10.1016/j.eswa.2021.115544

7. De Mauro A, Greco M, Grimaldi M, et al (2018) Human resources for big data professions: A systematic classification of job roles and required skill sets. Information Processing & Management 54(5):807–817. https://doi.org/10.1016/j.ipm.2017.05.004

8. Cui Y, Che W, Liu T, et al (2020) Revisiting pre-trained models for chinese natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings. Association for Computational

Linguistics, Online, pp 657–668, https://doi.org/10.48550/arXiv.2004.13922

9. Gao T, Yao X, Chen D (2021) Simcse: Simple contrastive learning of sentence embeddings. In: 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Association for Computational Linguistics (ACL). Association for Computational Linguistics, pp 6894–6910, https://doi.org/10.18653/V1/2021.EMNLP-MAIN.552

10. Guo H, Tang R, Ye Y, et al (2017) Deepfm: A factorization-machine based neural network for ctr prediction. In: Sierra C (ed) Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017. ijcai.org, pp 1725–1731, https://doi.org/10.24963/IJCAI.2017/239

11. Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. Advances in neural information processing systems 30:6000–6010. https://doi.org/10.5555/3295222.3295349

12. Zhu C, Zhu H, Xiong H, et al (2018) Person-job fit: Adapting the right talent for the right job with joint representation learning. ACM Transactions on Management Information Systems (TMIS) 9(3):1–17. https://doi.org/0.1145/3234465

13. Bian S, Zhao WX, Song Y, et al (2019) Domain adaptation for person-job fit with transferable deep global match network. In: Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp 4810–4820, https://doi.org/10.18653/v1/D19-1487

14. Yan R, Le R, Song Y, et al (2019) Interview choice reveals your preference on the market: To improve job-resume matching through profiling memories. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, Boston, MA, USA, pp 914–922, https://doi.org/10.1145/3292500.3330963

15. Shalaby W, AlAila B, Korayem M, et al (2017) Help me find a job: A graph-based approach for job recommendation at scale. In: 2017 IEEE international conference on big data (big data), IEEE, pp 1544–1553, https://doi.org/10.1109/BigData.2017.8258088

16. Wang Z, Wei W, Xu C, et al (2022) Person-job fit estimation from candidate profile and related recruitment history with co-attention neural networks. Neurocomputing 501:14–24. https://doi.org/10.1016/j.neucom.2022.06.012

17. Yang C, Hou Y, Song Y, et al (2022) Modeling two-way selection preference for person-job fit. In: Proceedings of the 16th ACM Conference on Recommender Systems, pp 102–112, https://doi.org/10.1145/3523227.3546752

18. Chen H, Du L, Lu Y, et al (2024) Professional network matters: Connections empower person-job fit. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining, pp 96–105, https://doi.org/10.1145/3616855.3635852

19. Shi X, Wei Q, Chen G (2024) A bilateral heterogeneous graph model for interpretable job recommendation considering both reciprocity and competition. Frontiers of Engineering Management pp 1–15. https://doi.org/10.1007/s42524-023-0280-2

20. Qin C, Zhu H, Xu T, et al (2018) Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. In: Collins-Thompson K, Mei Q, Davison BD, et al (eds) The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018. ACM, pp 25–34, https://doi.org/10.1145/3209978.3210025

21. Qin C, Zhu H, Xu T, et al (2020) An enhanced neural network approach to person-job fit in talent recruitment. ACM Transactions on Information Systems (TOIS) 38(2):1–33. https://doi.org/10.1145/3376927

22. Hou Y, Pan X, Zhao WX, et al (2022) Leveraging search history for improving person-job fit. In: International Conference on Database Systems for Advanced Applications, Springer, pp 38–54, https://doi.org/10.1007/978-3-031-00123-9_3

23. He M, Shen D, Wang T, et al (2023) Self-attentional multi-field features representation and interaction learning for person–job fit. IEEE Transactions on Computational Social Systems 10(1):255–268. https://doi.org/10.1109/TCSS.2021.3134458

24. Yang ZR, He ZY, Wang CD, et al (2022) A bi-directional recommender system for online recruitment. In: 2022 IEEE International Conference on Data Mining (ICDM), IEEE, pp 628–637

25. Jiang J, Ye S, Wang W, et al (2020) Learning effective representations for person-job fit by feature fusion. In: CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020. ACM, pp 2549–2556, https://doi.org/10.1145/3340531.3412717

26. Huang Y, Liu DR, Lee SJ (2023) Talent recommendation based on attentive deep neural network and implicit relationships of resumes. Information Processing & Management 60(4):103357. https://doi.org/10.1016/j.ipm.2023.103357

27. Ma H, Xu Y, Ma W, et al (2020) A muti-field feature interaction convolutional neural network for resume recommendation. In: 2020 International Symposium on Autonomous Systems (ISAS), IEEE, pp 186–191, https://doi.org/10.1109/ISAS49493.2020.9378849

28. Zhang Y, Liu B, Qian J, et al (2021) An explainable person-job fit model incorporating structured information. In: 2021 IEEE International Conference on Big Data (Big Data), IEEE, pp 3571–3579, https://doi.org/10.1109/BigData52589.2021.9672057

29. Zhang Z, Luo Y, Wen Y, et al (2022) Cycleresume: A cycle learning framework with hybrid attention for fine-grained talent-job fit. In: CAAI International Conference on Artificial Intelligence, Springer, pp 260–271, https://doi.org/10.1007/978-3-031-20503-3_21

30. Fu B, Liu H, Zhao H, et al (2022) Market-aware dynamic person-job fit with hierarchical reinforcement learning. In: International Conference on Database Systems for Advanced Applications, Springer, pp 697–705, https://doi.org/10.1007/978-3-031-00126-0_54

31. Shi X, Song J, Wu J, et al (2023) Serialized knowledge enhanced multi-objective person-job matching recommendation in a high mobility job market. In: 56th Hawaii International Conference on System Sciences, HICSS 2023, Maui, Hawaii, USA, January 3-6, 2023, pp 980–989, URL https://hdl.handle.net/10125/102750

32. Benabderrahmane S, Mellouli N, Lamolle M (2018) On the predictive analysis of behavioral massive job data using embedded clustering and deep recurrent neural networks. Knowledge-Based Systems 151:95–113. https://doi.org/10.1016/j.knosys.2018.03.025

33. Lavi D, Medentsiy V, Graus D (2021) consultantbert: Fine-tuned siamese sentence-bert for matching jobs and job seekers. In: Proceedings of the Workshop on Recommender Systems for Human Resources (RecSys in HR 2021) co-located with the 15th ACM Conference on Recommender Systems (RecSys 2021), CEUR Workshop Proceedings, vol 2967. CEUR-WS.org, https://doi.org/arxiv.org/abs/2109.06501

34. Cui Y, Che W, Liu T, et al (2021) Pre-training with whole word masking for chinese bert. IEEE/ACM Transactions on Audio, Speech, and Language Processing 29:3504–3514. https://doi.org/10.1109/TASLP.2021.3124365

35. Grohe M (2020) word2vec, node2vec, graph2vec, x2vec: Towards a theory of vector embeddings of structured data. In: Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, pp 1–16, https://doi.org/10.1145/3375395.3387641

36. Rendle S (2010) Factorization machines. In: 2010 IEEE International conference on data mining, IEEE, pp 995–1000, https://doi.org/10.1109/ICDM.2010.127

37. Ma Y, Yu D, Wu T, et al (2019) Paddlepaddle: An open-source deep learning platform from industrial practice. Frontiers of Data and Domputing 1(1):105–115. https://doi.org/10.11871/jfdc.issn.2096.742X.2019.01.011

38. Ranking S (2023) Best chinese universities ranking, URL https://www.shanghairanking.cn/rankings/bcur/202311

39. Le R, Hu W, Song Y, et al (2019) Towards effective and interpretable person-job fitting. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp 1883–1892, https://doi.org/10.1145/3357384.3357949

40. Shao T, Song C, Zheng J, et al (2023) Exploring internal and external interactions for semi-structured multivariate attributes in job-resume matching. Int J Intell Syst 2023:1–16. https://doi.org/10.1155/2023/2994779