Enhancing Job Recommendation through LLM-Based Generative Adversarial Networks

Yingpeng Du[1,4*†], Di Luo[2*], Rui Yan[2], Xiaopei Wang[3†], Hongzhi Liu[4†], Hengshu Zhu[5], Yang Song[6†], Jie Zhang[1]

[1]School of Computer Science and Engineering, Nanyang Technological University, Singapore
[2]Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China
[3]School of Languages and Communication Studies, Beijing Jiaotong University, Beijing, China
[4]School of Software and Microelectronics, Peking University, Beijing, China
[5]Career Science Lab, BOSS Zhipin, Beijing, China
[6]NLP Center, BOSS Zhipin, Beijing, China
dyp1993@pku.edu.cn,di_luo@ruc.edu.cn,ruiyan@ruc.edu.cn,wangxp@bjtu.edu.cn,liuhz@pku.edu.cn,
zhuhengshu@kanzhun.com,songyang@kanzhun.com,zhangj@ntu.edu.sg

Abstract

Recommending suitable jobs to users is a critical task in online recruitment platforms. Existing job recommendation methods often encounter challenges such as the low quality of users' resumes, which hampers their accuracy and practical effectiveness. With the rapid development of large language models (LLMs), utilizing the rich knowledge encapsulated within them, as well as their powerful reasoning capabilities, offers a promising avenue for enhancing resume completeness to achieve more accurate recommendations. However, directly leveraging LLMs is not a one-size-fits-all solution, as it may suffer from issues like fabricated generation and few-shot problem, both of which can degrade the quality of resume completion. In this paper, we propose a novel LLM-based GANs Interactive Recommendation (LGIR) approach for job recommendation. To alleviate the limitation of fabricated generation, we not only extract users' explicit properties (e.g., skills, interests) from their self-description but also infer users' implicit characteristics from their behaviors for more accurate and meaningful resume completion. Nevertheless, some users still suffer from the few-shot problem, which arises due to scarce interaction records, leading to limited guidance for high-quality resume generation. To address this issue, we propose aligning unpaired low-quality resumes with high-quality generated counterparts using Generative Adversarial Networks (GANs), which can refine resume representations for better recommendation results. Extensive experiments on three large real-world recruitment datasets demonstrate the effectiveness of our proposed method.

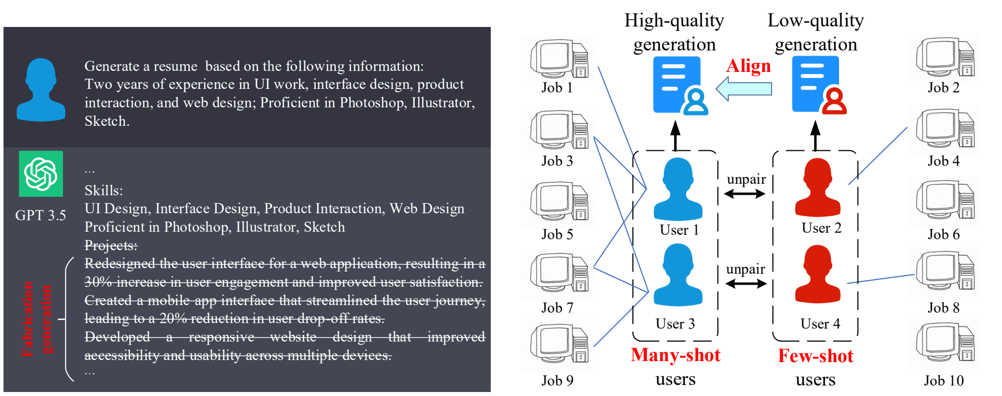
Introduction

Job recommendation is an essential task in today's online recruitment platforms, significantly improving recruitment efficiency by accurately matching job seekers (aka users) with suitable positions. Although existing job recommendation methods (Le et al. 2019; Jiang et al. 2020; Hou et al. 2022) have achieved considerable success in recent years, they still face significant challenges, such as the low quality of user resumes and interference from the few-shot problem (Gope and Jain 2017), hindering their practical accuracy and efficiency. For example, some users may not invest sufficient effort in crafting their resumes or lack comprehensive self-awareness, resulting in incomplete and low-quality descriptions of their skills and job preferences. Inspired by the recent remarkable capabilities and rapid development of large language models (LLMs), it is intuitive to utilize their extensive knowledge, powerful text comprehension, and reasoning abilities to improve and rectify low-quality resumes.

However, simply leveraging LLMs (Touvron et al. 2023; Brown et al. 2020) to enhance user resumes is not a one-size-fits-all solution for job recommendation. Due to the widespread fabrications and hallucinations within LLMs (Zhang et al. 2023), it is difficult to generate high-quality resumes without users' reliable interactive information. Fig.1 (A) illustrates the resume generation process for a user using simple completion with a well-known LLM, ChatGPT. It underscores that the generated results often contain excessive unrelated and fabricated information, rendering them unsuitable for recommendation. To alleviate this fabricated generation, we propose exploring users' interactive behaviors with recommender systems to mine their relevance to users' abilities and preferences, thereby assisting the LLMs in better profiling users for resume completion. Specifically, users generally possess particular job skills, residential addresses, and educational backgrounds, which make them interact with jobs that contain corresponding responsibilities, locations, and levels. As a result, we propose inferring users' implicit characteristics (e.g., skills, preferences) from their interaction behaviors to help LLMs profile users and generate high-quality resumes.

Although exploring users' interactive behaviors can help LLMs better profile users, they may still suffer from the few-shot problem, limiting the quality of resume completion for certain users. Specifically, users with few interaction records (aka the long-tail effect) still face challenges with

*These authors contributed equally.
†Corresponding authors.



(A) Extract valuable information beyond users' original resumes to allevaite fabrication of LLMs.

(B) Align the low-quality generation of few-shot users with high-quality generation.

Figure 1: The difficulty and motivation behind leveraging LLMs for job recommendation.

fabrications and hallucinations within LLMs, as they lack sufficient interactive guidance for high-quality resume completion. To alleviate this problem, we propose aligning the generated resumes of few-shot users with the high-quality resumes of users who have extensive interaction records as shown in Fig.1 (B). Due to the lack of paired high-quality and low-quality resumes for a specific user in real-world scenarios, we introduce a Generative Adversarial Networks (GANs) (Goodfellow et al. 2020) based method to align the *unpaired* resumes across different users, which can refine the generated resumes of few-shot users. Specifically, the generator aims to improve the representations of low-quality resumes by fooling the discriminator, while the discriminator strives to distinguish between the refined representations and the high-quality representations as effectively as possible. Through iterative training of GANs, the generator plays a crucial role in refining the representations of low-quality resumes, which can bridge the gap between few-shot users and many-shot users to enhance the quality of resume completion for all users.

To sum up, we propose an LLM-based GANs Interactive Recommendation (LGIR) method for job recommendation in this paper, which aims to address the limitations of fabricated generation in LLMs and the few-shot problem that degrades the quality of resume completion. To tackle the fabricated generation limitation, we extract valuable information beyond users' resumes. Specifically, we not only extract users' explicit properties from their self-descriptions but also infer their implicit characteristics from their behaviors, leading to accurate and meaningful resume completion. To mitigate the few-shot problem that restricts the quality of generated resumes, we propose a transfer representation learning strategy using GANs, which align low-quality resumes with unpaired high-quality resumes, enhancing the overall quality. We evaluate our model on three real-world datasets, demonstrating consistent superiority over state-of-the-art methods for job recommendation. Ablation experiments and a case study further substantiate the motivations and effectiveness behind our proposed method.

Related Work

Job Recommendation. Job recommendation has gained significant popularity in online recruitment platforms and can be primarily categorized into three groups: behavior-based methods, content-based methods, and hybrid methods. **Behavior-based methods** have been developed to leverage user-item interaction for job recommendation. Collaborative filtering based methods (Koren, Bell, and Volinsky 2009) have gained popularity among these approaches, which can be modified with deep neural networks (He and Chua 2017) and graph models (He et al. 2020) for more accurate recommendation results. **Content-based methods** utilize the rich semantic information present in resumes and job requirements using text-matching strategies or text enhancement techniques, such as CNN (Zhu et al. 2018), RNN (Qin et al. 2018), and memory networks (Yan et al. 2019). **Hybrid methods** combine the strengths of both behavior-based and content-based approaches. Specifically, they construct the embeddings of users and jobs based on their text content and leverage user-item interaction for job recommendation (Le et al. 2019; Jiang et al. 2020; Hou et al. 2022). However, these methods often suffer from the low quality of users' resumes. To address this challenge, we propose utilizing the rich knowledge and reasoning abilities encapsulated within LLMs to improve the resume quality for recommendation.

Large Language Models for Recommendation. Large Language Models (LLMs) (Touvron et al. 2023; Brown et al. 2020) are revolutionizing recommendation systems (Wu et al. 2023). Due to their extensive assimilation of knowledge (Liu, Zhang, and Gulla 2023), LLMs have the distinct advantage of comprehending contextual information (Geng et al. 2022), leading to improved recommendation accuracy and user satisfaction. They offer potential solutions to the cold-start problem with zero-shot recommendation capabilities (Sileo, Vossen, and Raymaekers 2022). Their capacity to generate language-based explanations also enhances recommendation interpretability (Gao et al. 2023). However, challenges arise in their direct application, including knowledge gaps and a tendency for unrealistic results

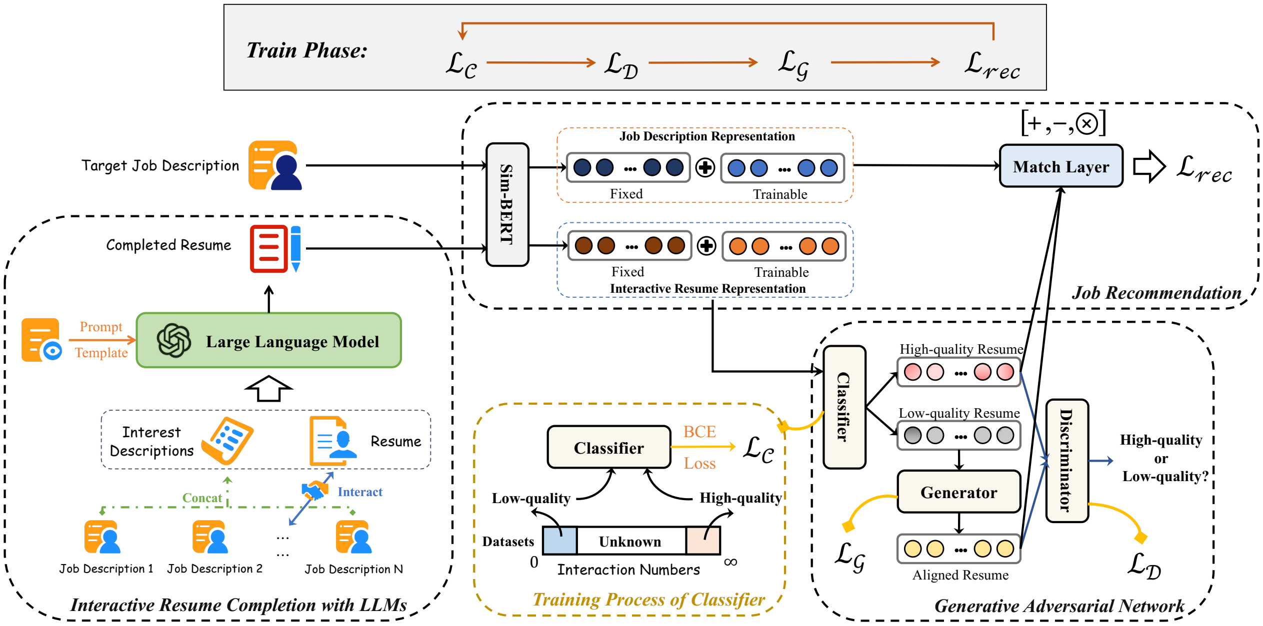


Figure 2: The architecture of the LLM-based GANs Interactive Recommendation (LGIR), mainly contains the interactive resume completion method for resume generation by LLMs and the GANs-based method for resume quality alignment.

(Liu et al. 2023). Recent studies utilize constructive prompts and in-context learning to control and direct LLM outputs, with methods such as (Hou et al. 2023)'s sequential recommendation prompts, (Gao et al. 2023)'s interactive recommendation framework, and (Wang et al. 2023)'s generative recommendation framework. Some also harness user behavior history for guidance (Chen 2023). Nonetheless, pervasive long-tail issues remain challenges, which can further exacerbate the hallucination problem of LLMs. To address these, our work uniquely employs Generative Adversarial Networks (GANs) to enhance representations of few-shot users, aiming to improve recommendation quality.

Problem Definition

Let $\mathcal{C} = \{c_1, \cdots, c_N\}$ and $\mathcal{J} = \{j_1, \cdots, j_M\}$ represent the sets of N users and M jobs, respectively. Each user or job is associated with a text document describing the resume or job requirement. Specifically, we denote the resume of user c as $T_c = [w_1, \cdots, w_{l_c}]$, where w_i is the i-th word in the resume and l_c denotes the the length of resume T_c. Similarly, the requirement description of job j with length l_j is denoted as $T_j = [w_1, \cdots, w_{l_j}]$. We suppose to know the interaction records between users and jobs, which can be represented as an interaction matrix $\mathcal{R} \in \mathbb{R}^{N\times M}$, where $\mathcal{R}_{ik} = 1$ if user c_i has interacted with the job j_k, and $\mathcal{R}_{ik} = 0$ otherwise.

In this paper, our goal is to recommend appropriate jobs to users. Formally, we propose learning a matching function $g(c_i, j_k)$ based on the interaction records $\mathcal{R}$ and the documents T. We then make the top-K recommendation based on this matching function.

The Proposed Method

The overall architecture of the proposed method is shown in Fig.2. Firstly, we propose an interactive resume completion method to alleviate the limitation of the fabricated generation in LLMs. Secondly, we propose a GANs-based aligning method to refine LLMs' representations of low-quality resumes. Finally, we propose a multi-objective learning framework for job recommendation.

A LLM-based Method for Resume Completion

To enhance the quality of users' resumes and thereby improve job recommendations, we propose leveraging the extensive knowledge and superior reasoning abilities of Large Language Models (LLMs). Specifically, we introduce two methods, named Simple Resume Completion (SRC) and Interactive Resume Completion (IRC), aimed at improving the quality of users' resumes for more accurate recommendations.

Simple Resume Completion with LLMs To improve the quality of users' resumes, we propose completing users' resumes using a prompting approach that directly leverages LLMs' knowledge and generation abilities. Specifically, we construct the prompt for LLMs based on the user's self-description as follows:

$$G_c = \text{LLMs}(prompt_{\text{SRC}}, T_c) \qquad (1)$$

where $prompt_{\text{SRC}}$ denotes the command that triggers the LLMs to complete the user u's resume based on his/her self-description T_c, the details of which are shown in the upper part of Fig.3. However, the SRC strategy may suffer from the fabricated and hallucinated generation of LLMs.

Interactive Resume Completion with LLMs To mitigate the limitation of fabricated generation in LLMs, we propose exploring users' interactive behaviors with recommender systems, thus assisting LLMs to better profile users for resume completion. For instance, users typically have specific job skills, residential addresses, and educational backgrounds, which influence their interactions with job positions containing corresponding responsibilities. Consequently, users' implicit characteristics (e.g., skills, preferences) can be inferred from their interaction behaviors for more accurate and meaningful resume completion. Specifically, we adopt a particular prompting approach for resume completion by LLMs, with consideration of both user's self-description and his/her interactive behaviors:

$$G_c = \text{LLMs}(prompt_{\text{IRC}}, T_c, R_c) \tag{2}$$

where $R_c = \{T_{j_k} | \mathcal{R}_{c,j_k} = 1\}$ denotes the requirements of jobs that the user c has interacted with. The details of the $prompt_{\text{IRC}}$ is shown in the lower part of Fig.3.

To utilize the user resumes and job requirements, we adopt the BERT to encode them into constant text embeddings $W_t \in \mathbb{R}^d$ (Yang et al. 2022). Specifically, we first maintain the text order and place a unique token $[CLS]$ before it, then we feed the combined sequence into the SIM-BERT model and use the output of the token $[CLS]$ as the semantic embeddings of the descriptive text (e.g., $W_{G_{c_i}} =$ SIM-BERT(G_{c_i})). Finally, we employ a multi-layer perceptron to encode these semantic embeddings:

$$x_{c_i} = \text{MLP}_{\text{user}}([P_i; W_{G_{c_i}}]), \tag{3}$$

$$x_{j_k} = \text{MLP}_{\text{job}}([Q_k; W_{T_{j_k}}]), \tag{4}$$

where G_{c_i} and T_{j_k} denote the user c_i's LLMs-generated resume and the job j_k's requirement description. $P_i, Q_k \in \mathbb{R}^d$ represent the ID embeddings for user c_i and job j_k, respectively. MLP_{user} and MLP_{job} denote the multi-layer perceptron with hidden layers $[2 \cdot d \to d_{e'} \to d_e]$ and the activation function $\text{Relu}(\cdot) = \max(\cdot, 0)$. d, d_e and $d_{e'}$ indicate the dimensions of hidden layers in the multi-layer perceptron.

A GAN-based Aligning Method for Resume Refine

While the exploration of users' interactive behaviors does enable LLMs to more effectively profile users, it may still encounter the few-shot problem. Specifically, users with limited interaction records might lead to difficulties in generating high-quality resumes. To address this challenge, we propose refining the low-quality resumes of few-shot users. The approach comprises two main components: a classifier designed to detect low-quality resumes, and Generative Adversarial Networks (GANs) employed for aligning resumes.

Classifier To detect the low-quality resumes for alignment, we propose a classifier $\mathcal{C}$ to distinguish between high-quality resumes and low-quality resumes, i.e.,

$$\mathcal{C}(x) = \sigma(W_2^c \cdot \text{Relu}(W_1^c \cdot x)) \tag{5}$$

where $W_1^c \in \mathbb{R}^{d_c \times d_e}$ and $W_2^c \in \mathbb{R}^{1 \times d_c}$ represent the parameters within the classifier $\mathcal{C}$ and we define them as $\Theta_\mathcal{C} = \{W_1^c, W_2^c\}$. We posit that users with either extremely

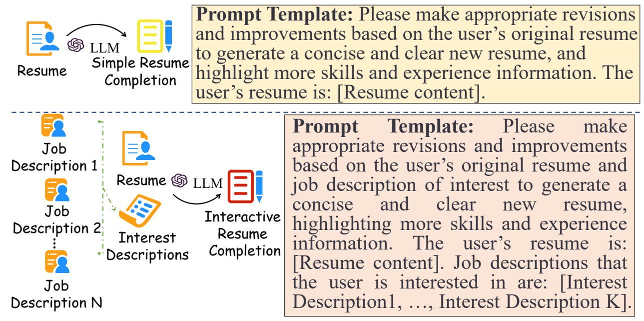


Figure 3: The difference between Simple Resume Completion and Interactive Resume Completion.

few or rich interaction records may respectively result in low-quality and high-quality resume generation by LLMs. To this end, we introduce the cross-entropy loss to train the classifier $\mathcal{C}$ on these partial users, i.e.,

$$\mathcal{L}_\mathcal{C} = \mathbb{E}_{(c_i, y_{c_i}) \sim T_\mathcal{C}}[y_{c_i} \cdot \log(\hat{y}_{c_i}) + (1 - y_{c_i}) \cdot \log(1 - \hat{y}_{c_i})] \tag{6}$$

where $\hat{y}_{c_i} = \mathcal{C}(x_{c_i})$ denotes the quality prediction for user c_i's generated resume, and $T_\mathcal{C} = T_\mathcal{C}^\uparrow \bigcup T_\mathcal{C}^\downarrow$ assembles the users for classifier learning ($T_\mathcal{C}^\uparrow = \{(c_i, 1) | \sum_k R_{ik} \geq \kappa_1\}$ and $T_\mathcal{C}^\downarrow = \{(c_i, 0) | \sum_k R_{ik} \leq \kappa_2\}$ represent the many-shot and few-shot users). y_{c_i} serves as the ground truth, where $y_{c_i} = 1$ if $c_i \in T_\mathcal{C}^\uparrow$ and $y_{c_i} = 0$ if $c_i \in T_\mathcal{C}^\downarrow$. The thresholds κ_1 and κ_2 are used to select the many-shot and few-shot users.

Generator To improve the resume quality, we introduce a generator $\mathcal{G}$ to refine the representations of low-quality resumes as identified by the aforementioned classifier $\mathcal{C}$. Specifically, the generator $\mathcal{G}$ aims to map the low-quality resume representations to their high-quality counterparts:

$$\mathcal{G}(x) = W_2^g \cdot \text{Relu}(W_1^g \cdot x) \tag{7}$$

where $W_1^g \in \mathbb{R}^{d_g \times d_e}$, $W_2^g \in \mathbb{R}^{d_e \times d_g}$ represent the parameters in the generator $\mathcal{G}$ and are defined as $\Theta_\mathcal{G} = \{W_1^g, W_2^g\}$.

Discriminator The principal function of the discriminator is to differentiate between samples originating from two distinct distributions. Specifically, we introduce a discriminator $\mathcal{D}$ to discern whether a given resume representation is a product of the generator's refinement process or a direct encoding of a high-quality resume:

$$\mathcal{D}(x) = \sigma(W_2^d \cdot \text{Relu}(W_1^d \cdot x)) \tag{8}$$

where $W_1^d \in \mathbb{R}^{d_s \times d_e}$, $W_2^d \in \mathbb{R}^{1 \times d_s}$ represent the parameters of $\mathcal{D}$, and are defined as $\Theta_\mathcal{D} = \{W_1^d, W_2^d\}$.

Adversarial Learning To align the representations of the low-quality and high-quality resumes, we propose engaging in a mini-max game between a generator and a discriminator (Goodfellow et al. 2020).

The discriminator $\mathcal{D}$ is responsible for distinguishing samples from distinct distributions. For the training of $\mathcal{D}$, we aim to maximize the following probability, which determines whether a representation stems from the generator's refinement or a high-quality generated resume:

$$\max_{\Theta_{\mathcal{D}}} \mathcal{L}_{\mathcal{D}} = \mathbb{E}_{c_{i_1} \sim \hat{T}^{\uparrow}_{\mathcal{C}}}[\log \mathcal{D}(x_{c_{i_1}})] + \mathbb{E}_{c_{i_2} \sim \hat{T}^{\downarrow}_{\mathcal{C}}}[1 - \log \mathcal{D}(\mathcal{G}(x_{c_{i_2}}))] \tag{9}$$

where $\hat{T}^{\uparrow}_{\mathcal{C}}$ and $\hat{T}^{\downarrow}_{\mathcal{C}}$ denote the high-quality and low-quality generated resumes detected by the classifier $\mathcal{C}$, respectively.

The generator $\mathcal{G}$ focuses on refining low-quality generated resume representations to resemble high-quality resume representations. For the training of $\mathcal{G}$, we minimize the generator loss by deceiving the discriminator $\mathcal{D}$:

$$\min_{\Theta_{\mathcal{G}}} \mathcal{L}_{\mathcal{G}} = \mathbb{E}_{c_i \sim \hat{T}^{\downarrow}_{\mathcal{C}}}[1 - \log \mathcal{D}(\mathcal{G}(x_{c_i}))] \tag{10}$$

Through iterative training of the generator and discriminator in a competitive manner, this adversarial training process drives both components to improve, ultimately leading to the creation of low-quality samples that increasingly resemble high-quality ones.

Multi-objective Learning for Recommendation

To explore the high-quality resume representations for improved recommendation, we utilize the Classifier $\mathcal{C}$ and Generator $\mathcal{G}$ to obtain aligned resume representations, denoted as z_{c_i}, for all users, regardless of whether they are few-shot users or many-shot users, i.e.,

$$z_{c_i} = \begin{cases} x_{c_i}, & \text{if } \mathcal{C}(x_{c_i}) \geq 0.5; \\ \mathcal{G}(x_{c_i}), & \text{if } \mathcal{C}(x_{c_i}) < 0.5 \end{cases} \tag{11}$$

To predict users' behaviors on jobs, we propose a deep model to capture the non-linear and complex relationship between the user c_i and the job j_k, i.e.,

$$\hat{R}_{i,k} = g(c_i, j_k) = W^p \cdot [z_{c_i} + x_{j_k}; z_{c_i} - x_{j_k}; z_{c_i} \odot x_{j_k}] \tag{12}$$

where $\odot$ denotes the element-wise product, $W^p \in \mathbb{R}^{1 \times 3 \cdot d_e}$ maps to a score or probability of j_k that user c_i will engage. For the recommendation target, we adopt the pairwise loss to define the recommendation objective function as follows,

$$\mathcal{L}_{rec} = \max_{\Theta} \sum_{(i,j_1,j_2) \in D} \log \sigma(\hat{R}_{i,j_1} - \hat{R}_{i,j_2}) - \lambda ||\Theta||^2 \tag{13}$$

where the train set $D = \{(c_i, j_1, j_2)\}$ means that user c_k gave positive feedback to job j_1 (i.e., $R_{i,j_1} = 1$) instead of job j_2 (i.e., $R_{i,j_2} = 0$). The Θ denotes all parameters that need to be learned in the proposed model and λ is the regularization coefficient of L2 norm $|| \cdot ||^2$.

Experiment

In this section, we aim to evaluate the performance and effectiveness of LGIR. Specifically, we conduct several experiments to study the following research questions:

- **RQ1**: Whether the proposed method LGIR outperforms state-of-the-art methods for job recommendation?
- **RQ2**: Whether LGIR benefits from inferring users' implicit characteristics from their behaviors for more accurate and meaningful resume generation?
- **RQ3**: Whether LGIR benefits from aligning the few-shot resumes with high-quality representations?
- **RQ4**: How LGIR achieves SOTA results in case level?

Dataset	# Users	# Items	# Interaction
Designs	12,290	9,143	166,270
Sales	15,854	12,772	145,066
Tech	56,634	48,090	925,193

Table 1: Statistics of the experimental datasets.

Experimental Setup

Datasets We evaluated the proposed method on three real-world data sets, which were provided by a popular online recruiting platform. These data sets were collected from 106 days of real online logs for job recommendation in the designer, sales, and technology industries, respectively. These data sets contained the rich interaction between users and employers. In addition, these data sets also contained text document information, which were the resumes of the users and the descriptions of job positions. The characteristics of these data sets are summarized in Table 1.

Evaluation Methodology and Metrics We spitted the interaction records into training, validation, and test sets equally. To evaluate the performance, we adopted three widely used evaluation metrics for top-n recommendation (Zhao et al. 2022): mean average precision ($MAP@n$), normalized discounted cumulative gain ($NDCG@n$) and mean reciprocal rank (MRR), where n was set as 5 empirically. We sampled 20 negative instances for each positive instance from users' interacted and non-interacted records. Experimental results were recorded as the average of five runs with different random initialization of model parameters.

Baselines We took the following state-of-the-art methods as the baselines, including content-based methods (i.e., BPJFNN (Qin et al. 2018)), collaborative filtering based methods (i.e., MF (Koren, Bell, and Volinsky 2009) and NCF (He et al. 2017)), hybrid methods (i.e., PJFFF (Jiang et al. 2020), SHPJF (Hou et al. 2022), SGL-text(Wu et al. 2021) , LightGCN-text(He et al. 2020), and LightGCN+SRC), and LLMs based method (i.e., SGPT-BE (Muennighoff 2022), SGPT-ST (Reimers and Gurevych 2019), SGPT-ST+SRC).

Implementation Details We adopted the ChatGLM-6B (Du et al. 2022) as the LLM model in this paper. For a fair comparison, all methods were optimized by the AdamW optimizer with the same latent space dimension (i.e., 64), batch size (i.e., 1024), learning rate (i.e., 5×10^{-5}), and regularization coefficient (i.e., 1×10^{-4}). We set $d = 768$, $d_{e'} = 128$, $d_e = 64$, and $d_c = d_s = d_g = 256$ for the proposed method. We carefully searched other special hyper-parameters for best performance, and early stopping was used with the patience of 50 epochs.

Model Comparison (RQ1)

Table 2 outlines the performance of various job recommendation methods, highlighting the top-2 results for each dataset. The conclusions drawn are as follows:

1. **Effectiveness of LGIR:** The proposed method LGIR consistently surpasses all baseline methods, improving

Models	Designs MAP@5	Designs NDCG@5	Designs MRR	Sales MAP@5	Sales NDCG@5	Sales MRR	Tech MAP@5	Tech NDCG@5	Tech MRR
SGPT-BE	0.0712	0.1140	0.2128	0.0526	0.0932	0.1726	0.1464	0.2092	0.3344
SGPT-ST	0.0694	0.1107	0.2077	0.0519	0.0926	0.1714	0.1422	0.2025	0.3289
SGPT-ST + SRC	0.0727	0.1177	0.2185	0.0511	0.0925	0.1719	0.1541	0.2194	0.3442
BPJFNN	0.1415	0.2156	0.3436	0.1138	0.2038	0.3030	0.2018	0.2948	0.4704
MF	0.1914	0.2913	0.4557	0.0887	0.1628	0.2789	0.4359	0.6054	0.7555
NCF	0.2071	0.3230	0.4944	0.1463	0.2670	0.3941	0.4105	0.5706	0.7414
PJFFF	0.1182	0.1855	0.3299	0.0690	0.1255	0.2199	0.2802	0.4040	0.6127
SHPJF	0.1862	0.2875	0.4531	0.1334	0.2436	0.3705	0.3710	0.5189	0.7016
SGL-text	0.2716	0.4309	0.5941	0.1508	0.2712	0.3945	0.4416	0.6230	0.7836
LightGCN-text	0.2664	0.4218	0.5955	0.1629	0.2980	0.4271	0.4676	0.6591	0.8093
LightGCN+SRC	0.2649	0.4189	0.5926	0.1611	0.2939	0.4204	0.4719	0.6661	0.8146
LGIR(ours)	**0.2887***	**0.4622***	**0.6319***	**0.1751***	**0.3225***	**0.4548***	**0.5086***	**0.7191***	**0.8434***
Imprvement	**6.28%**	**7.26%**	**6.11%**	**7.50%**	**8.22%**	**6.49%**	**7.78%**	**7.96%**	**3.54%**

Table 2: Performance of the proposed and baseline methods for job recommendation. $*$ indicates that the improvements are significant at the level of 0.01 with paired t-test.

Dataset	Method	MAP@5	NDCG@5	MRR
Designs	BASE	0.2627	0.4128	0.5829
	SRC	0.2601	0.4076	0.5781
	IRC	0.2859	0.4560	0.6220
	LGIR	**0.2887**	**0.4622**	**0.6319**
Sales	BASE	0.1617	0.2945	0.4250
	SRC	0.1652	0.3031	0.4331
	IRC	0.1671	0.3065	0.4359
	LGIR	**0.1751**	**0.3225**	**0.4548**
Tech	BASE	0.4994	0.7088	0.8374
	SRC	0.5048	0.7148	0.8435
	IRC	0.5056	0.7153	0.8400
	LGIR	**0.5086**	**0.7191**	**0.8434**

Table 3: Performance of the variants for ablation studies.

the best baseline by 6.65%, 7.40%, and 6.42% on designs, sales, and tech datasets, respectively.

2. **Limitations of LLM-only Methods:** LLMs methods (SGPT) perform poorly, indicating that relying solely on textual descriptions is ineffective due to inherent limitations such as meaningless information.
3. **Challenges with Hybrid Methods:** Hybrid methods like PJFFF and SHPJF, perform inadequately, likely due to the unstructured and varying organization habits of users.
4. **Success of GCN-based Methods:** GCN-based methods like LightGCN, which utilize preference encoding, achieve the best performance among baselines, signifying the importance of combining interactions and text.
5. **Simple Resume Completion's Limitations:** The strategy of simple resume completion (SRC) shows minimal improvement (e.g., LGCN vs. LGCN + SRC), revealing that merely leveraging LLMs isn't universally effective due to their tendency to generate fabricated content.

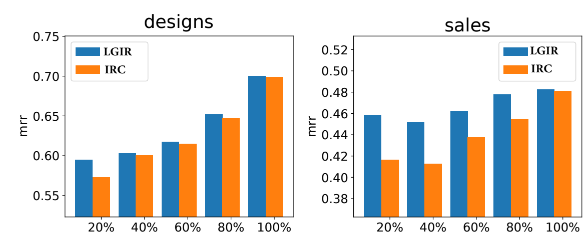


Figure 4: Performance comparison of LGIR and the variant IRC for few-shot analysis.

Ablation Study (RQ2&3)

To assess the effectiveness of the LGIR's module design, it's compared to several special cases:

- **BASE**: A two-tower text matching model that uses the original self-description from users for recommendation.
- **SRC**: Utilizes the generated resumes of users with a simple resume completion (SRC) strategy without GANs-based learning for job recommendation.
- **IRC**: Leverages the generated resumes with the interactive resume completion (IRC) strategy, but without GANs-based learning for aligning unpaired resumes.
- **LGIR**: The proposed method, including both the IRC strategy and GANs-based learning for recommendation.

Table 3 shows the performance of these methods, i.e. LGIR, BASE, SRC, and IRC. From the experimental results, we can get the following conclusions:

- RQ2: The SRC variant shows limited improvement over BASE, demonstrating that simply leveraging LLMs for job recommendation is not a one-size-fits-all solution. Issues with fabricated and hallucinated generation are addressed through the Interactive Resume Completion (IRC) strategy, which shows substantial improvement over both BASE and SRC. This highlights the necessity

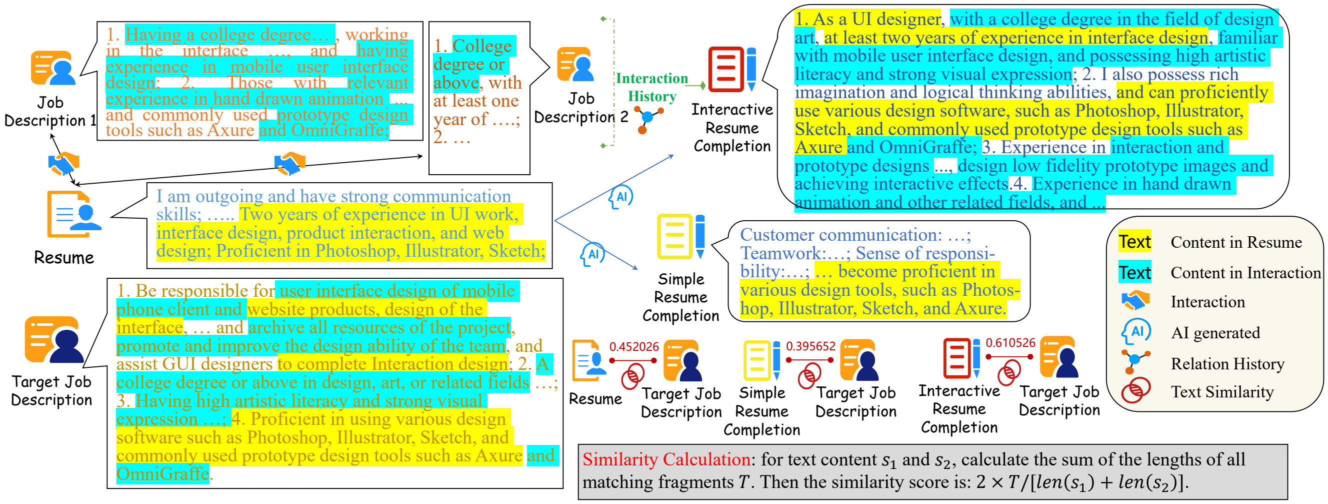


Figure 5: A real recruitment scenario where users have two historical interactions. The process explains how the model successfully integrates pertinent information from user resumes and interactive job descriptions that better reflect the user's abilities.

of inferring users' implicit characteristics based on their behaviors for more accurate resume generation.

- RQ3: The proposed method LGIR significantly outperforms the variants across all data sets, which benefits from the GANs-based learning to align the generated resumes of few-shot users with high-quality representations. Further in-depth analysis of the role of GANs is explored in the subsequent few-shot analysis.

Few-shot Analysis (RQ3)

The ablation study reveals the strengths of LGIR in aligning the generated resumes of few-shot users with high-quality representations. It is interesting to investigate how LGIR handles the challenges associated with few-shot scenarios, so a few-shot analysis was conducted, comparing LGIR with the **IRC** variant across different shot levels. Users were equally divided into five groups based on their interaction numbers (for example, the group 40% denotes the user set that falls within the $20\% - 40\%$ ranking range based on the number of interactions), and the recommendation performance of LGIR and IRC was compared across these groups.

The results in Fig.4 show LGIR consistently outperformed IRC in most cases, validating the effectiveness of the GANs-based learning scheme. Especially, LGIR showed a more pronounced improvement in groups with fewer interactions, confirming that GANs-based learning can align the resumes of few-shot users with those of users who have rich interaction records. This indicates that LGIR can effectively mitigate the problems associated with few-shot scenarios that often limit the quality of resume generation.

Case Study (RQ4)

In a real recruitment scenario depicted in Fig.5, we delve deeper into the outputs of LLMs and explore how they assist LGIR in achieving state-of-the-art results. The figure presents the user's resume, previous job interactions, target job description and two resume completion approaches: Simple Resume Completion (LLMs alone) and Interactive Resume Completion (LLMs guided by interactive history). We also highlight content relevant to a target job in the user's resume (in yellow) and interaction history (in blue).

The illustration reveals that the user's interaction history contains clues relevant to the target job, absent in the user's own resume. Using only the user's resume with LLMs results in nonsensical content, reducing the proportion of valuable information in the resume. Conversely, the interactive approach successfully integrates pertinent information and generates resumes that better express the user's abilities, even those they may not have articulated or recognized. Furthermore, we quantify this by calculating the pairwise similarity between texts, showing that interactive completion improved similarity from 0.45 to 0.61, a remarkable 35% enhancement. Therefore, exploiting the interactive behaviors of users helps LLMs accurately capture skills and preferences, contributing to better job recommendation results.

Conclusion

In this paper, we propose an LLM-based GANs Interactive Recommendation (LGIR) method for job recommendation. To alleviate the fabricated generation of LLMs, we infer users' implicit characteristics from their behaviors for more accurate and meaningful resume completion. To address the few-shot problem encountered during resume generation, we propose the GANs-based method to refine the low-quality resumes of users. The proposed method outperforms state-of-the-art baselines, which demonstrates the superiority of utilizing LLMs with interactive resume completion and alignment for job recommendation. The ablation study highlights the significance of each component within the LGIR framework, and the case study further illustrates its superiority in capturing users' skills and preferences.

Acknowledgements

This research is partially supported by the Agency for Science, Technology and Research (A*STAR) under its RIE 2025 – Industry Alignment Fund – Pre Positioning (IAF-PP) funding scheme (Project No: M23L4a0001). This work is also partially supported by the MOE AcRF Tier 1 funding (RG90/20) awarded to Dr. Zhang Jie. This research is partially supported by National Natural Science Foundation of China (NSFC Grant No. 62122089).

References

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901.

Chen, Z. 2023. PALR: Personalization Aware LLMs for Recommendation. *arXiv preprint arXiv:2305.07622*.

Du, Z.; Qian, Y.; Liu, X.; Ding, M.; Qiu, J.; Yang, Z.; and Tang, J. 2022. GLM: General Language Model Pretraining with Autoregressive Blank Infilling. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 320–335.

Gao, Y.; Sheng, T.; Xiang, Y.; Xiong, Y.; Wang, H.; and Zhang, J. 2023. Chat-rec: Towards interactive and explainable llms-augmented recommender system. *arXiv preprint arXiv:2303.14524*.

Geng, S.; Liu, S.; Fu, Z.; Ge, Y.; and Zhang, Y. 2022. Recommendation as language processing (rlp): A unified pretrain, personalized prompt & predict paradigm (p5). In *Proceedings of the 16th ACM Conference on Recommender Systems*, 299–315.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

Gope, J.; and Jain, S. K. 2017. A survey on solving cold start problem in recommender systems. In *2017 International Conference on Computing, Communication and Automation (ICCCA)*, 133–138. IEEE.

He, X.; and Chua, T.-S. 2017. Neural factorization machines for sparse predictive analytics. In *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*, 355–364.

He, X.; Deng, K.; Wang, X.; Li, Y.; Zhang, Y.; and Wang, M. 2020. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*, 639–648.

He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; and Chua, T.-S. 2017. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, 173–182.

Hou, Y.; Pan, X.; Zhao, W. X.; Bian, S.; Song, Y.; Zhang, T.; and Wen, J.-R. 2022. Leveraging Search History for Improving Person-Job Fit. In *Database Systems for Advanced Applications: 27th International Conference, DASFAA 2022, Virtual Event, April 11–14, 2022, Proceedings, Part I*, 38–54. Springer.

Hou, Y.; Zhang, J.; Lin, Z.; Lu, H.; Xie, R.; McAuley, J.; and Zhao, W. X. 2023. Large language models are zero-shot rankers for recommender systems. *arXiv preprint arXiv:2305.08845*.

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

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30–37.

Le, R.; Hu, W.; Song, Y.; Zhang, T.; Zhao, D.; and Yan, R. 2019. Towards effective and interpretable person-job fitting. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 1883–1892.

Liu, J.; Liu, C.; Lv, R.; Zhou, K.; and Zhang, Y. 2023. Is chatgpt a good recommender? a preliminary study. *arXiv preprint arXiv:2304.10149*.

Liu, P.; Zhang, L.; and Gulla, J. A. 2023. Pre-train, prompt and recommendation: A comprehensive survey of language modelling paradigm adaptations in recommender systems. *arXiv preprint arXiv:2302.03735*.

Muennighoff, N. 2022. SGPT: GPT Sentence Embeddings for Semantic Search. *arXiv preprint arXiv:2202.08904*.

Qin, C.; Zhu, H.; Xu, T.; Zhu, C.; Jiang, L.; Chen, E.; and Xiong, H. 2018. 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*, 25–34.

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Sileo, D.; Vossen, W.; and Raymaekers, R. 2022. Zero-shot recommendation as language modeling. In *European Conference on Information Retrieval*, 223–230. Springer.

Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Wang, W.; Lin, X.; Feng, F.; He, X.; and Chua, T.-S. 2023. Generative recommendation: Towards next-generation recommender paradigm. *arXiv preprint arXiv:2304.03516*.

Wu, J.; Wang, X.; Feng, F.; He, X.; Chen, L.; Lian, J.; and Xie, X. 2021. Self-supervised graph learning for recommendation. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 726–735.

Wu, L.; Zheng, Z.; Qiu, Z.; Wang, H.; Gu, H.; Shen, T.; Qin, C.; Zhu, C.; Zhu, H.; Liu, Q.; et al. 2023. A Survey on Large Language Models for Recommendation. *arXiv preprint arXiv:2305.19860*.

Yan, R.; Le, R.; Song, Y.; Zhang, T.; Zhang, X.; and Zhao, D. 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*, 914–922.

Yang, C.; Hou, Y.; Song, Y.; Zhang, T.; Wen, J.-R.; and Zhao, W. X. 2022. Modeling Two-Way Selection Preference for Person-Job Fit. In *Proceedings of the 16th ACM Conference on Recommender Systems*, 102–112.

Zhang, Y.; Li, Y.; Cui, L.; Cai, D.; Liu, L.; Fu, T.; Huang, X.; Zhao, E.; Zhang, Y.; Chen, Y.; et al. 2023. Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models. *arXiv preprint arXiv:2309.01219*.

Zhao, W. X.; Lin, Z.; Feng, Z.; Wang, P.; and Wen, J.-R. 2022. A revisiting study of appropriate offline evaluation for top-N recommendation algorithms. *ACM Transactions on Information Systems*, 41(2): 1–41.

Zhu, C.; Zhu, H.; Xiong, H.; Ma, C.; Xie, F.; Ding, P.; and Li, P. 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.