Enhancing Job Recommendation through LLM-Based Generative Adversarial Networks

Yingpeng Du$^{1,4*}$, Di Luo$^2*$, Rui Yan$^3$, Xiaopei Wang$^3*$, Hongzhi Liu$^{4*}$, Hengshu Zhu$^5$, Yang Song$^5*$, 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, dliu@ruc.edu.cn, ruiyan@ruc.edu.cn, wangxp@bjtu.edu.cn, liuhz@pku.edu.cn, zhuangshu@kanzhu.com, songyang@kanzhu.com, zhangji@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
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.
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-term 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 \( C = \{c_1, \ldots, c_N\} \) and \( J = \{j_1, \ldots, j_M\} \) 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 \( T_c = [w_1, \ldots, w_{l_c}] \), where \( w_i \) is the \( i \)-th word in the resume and \( l_c \) denotes the length of resume \( T_c \). Similarly, the requirement description of job \( j \) with length \( l_j \) is denoted as \( T_j = [w_1, \ldots, w_{l_j}] \). We suppose to know the interaction records between users and jobs, which can be represented as an interaction matrix \( R \in \mathbb{R}^{N \times M} \), where \( R_{ik} = 1 \) if user \( c_i \) has interacted with the job \( j_k \), and \( R_{ik} = 0 \) 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 \( R \) 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:

\[
G_c = \text{LLMs}(\text{prompt}_{\text{SRC}}, T_c)
\]  

where \( \text{prompt}_{\text{SRC}} \) denotes the command that triggers the LLMs to complete the user’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:

\[ G_c = \text{LLMs}(\text{prompt}_{\text{IRC}}, T_c, R_c) \]

where \( R_c = \{T_{jk}|R_{c,jk} = 1\} \) denotes the requirements of jobs that the user \( c \) has interacted with. The details of the \( \text{prompt}_{\text{IRC}} \) 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., \( W_{G_{c,i}} = \text{SIM-BERT}(G_{c,i}) \)). Finally, we employ a multi-layer perception to encode these semantic embeddings:

\[ x_{ci} = \text{MLP}_{\text{user}}([P_i; W_{G_{c,i}}]), \]
\[ x_{jk} = \text{MLP}_{\text{job}}([Q_{jk}; W_{T_{jk}}]), \]

where \( G_{ci} \) and \( T_{jk} \) denote the user \( c_i \)'s LLMs-generated resume and the job \( j_k \)'s requirement description. \( P_i, Q_{jk} \in \mathbb{R}^d \) represent the ID embeddings for user \( c_i \) and job \( j_k \), respectively. MLP\(_{\text{user}} \) and MLP\(_{\text{job}} \) denote the multi-layer perceptron with hidden layers \([2 \cdot d \rightarrow d\prime \rightarrow d]\) and the activation function \( \text{Relu}(\cdot) = \max(\cdot, 0) \), \( d \), \( d\prime \) 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. For instance, users typically have more accurate and meaningful resume completion. Specifically, we introduce the cross-entropy loss to train the classifier \( C \) on these partial users, i.e.,

\[ L_c = \mathbb{E}_{(c, y_c) \sim D_c} [y_c \cdot \log(y_c) + (1 - y_c) \cdot \log(1 - y_c)] \]

where \( y_{ci} = C(x_{ci}) \) denotes the quality prediction for user \( c_i \)'s generated resume, and \( T_c = T_1^c \cup T_2^c \) assembles the users for classifier learning \( T_1^c = \{(c_i, 1)| \sum_k R_{c,ik} \geq \kappa_1\} \) and \( T_2^c = \{(c_i, 0)| \sum_k R_{c,ik} \leq \kappa_2\} \) represent the many-shot and few-shot users. \( y_{ci} \) serves as the ground truth, where \( y_{ci} = 1 \) if \( c_i \in T_1^c \) and \( y_{ci} = 0 \) if \( c_i \in T_2^c \). The thresholds \( \kappa_1 \) and \( \kappa_2 \) are used to select the many-shot and few-shot users.

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

\[ G(x) = W_g^d \cdot \text{Relu}(W_g^{d-1} \cdot x) \]

where \( W_g^d \in \mathbb{R}^{d \times d_e}, W_g^{d-1} \in \mathbb{R}^{d_e \times d} \) represent the parameters in the generator \( G \) and are defined as \( \Theta_G = \{W_g^d, W_g^{d-1}\} \).

Discriminator The principal function of the discriminator is to differentiate between samples originating from two distinct distributions. Specifically, we introduce a discriminator \( D \) 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:

\[ D(x) = \sigma(W_d^{d} \cdot \text{Relu}(W_1^{d-1} \cdot x)) \]

where \( W_d^d \in \mathbb{R}^{d \times d_e}, W_1^d \in \mathbb{R}^{1 \times d_e} \) represent the parameters of \( D \), and are defined as \( \Theta_D = \{W_d^d, W_1^d\} \).

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 \( D \) is responsible for distinguishing samples from distinct distributions. For the training of \( D \), we aim to maximize the following probability, which determines whether a representation stems from the generator’s
The proposed method LGIR achieves SOTA results in case level?

**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., BPJFFNN (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-SE (Muennighoff 2022), SGPT-ST (Reimers and Gurevych 2019), SGPT-ST+SRC).
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.

<table>
<thead>
<tr>
<th>Models</th>
<th>Designs</th>
<th>Sales</th>
<th>Tech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAP@5</td>
<td>NDCG@5</td>
<td>MRR</td>
</tr>
<tr>
<td>SGPT-BE</td>
<td>0.0712</td>
<td>0.1140</td>
<td>0.2128</td>
</tr>
<tr>
<td>SGPT-ST</td>
<td>0.0694</td>
<td>0.1107</td>
<td>0.2077</td>
</tr>
<tr>
<td>SGPT-ST + SRC</td>
<td>0.0727</td>
<td>0.1177</td>
<td>0.2185</td>
</tr>
<tr>
<td>BPJFNN</td>
<td>0.1415</td>
<td>0.2156</td>
<td>0.3436</td>
</tr>
<tr>
<td>SGPT-ST + SRC</td>
<td>0.0727</td>
<td>0.1177</td>
<td>0.2185</td>
</tr>
<tr>
<td>BJFF</td>
<td>0.1182</td>
<td>0.1855</td>
<td>0.3299</td>
</tr>
<tr>
<td>SITF</td>
<td>0.0782</td>
<td>0.1382</td>
<td>0.2629</td>
</tr>
<tr>
<td>LightGCN</td>
<td>0.2071</td>
<td>0.3218</td>
<td>0.5955</td>
</tr>
<tr>
<td>LightGCN + SRC</td>
<td>0.2649</td>
<td>0.4189</td>
<td>0.5926</td>
</tr>
<tr>
<td>LGIR (ours)</td>
<td>0.2887*</td>
<td>0.4622*</td>
<td>0.6319*</td>
</tr>
</tbody>
</table>

Table 3: Performance of the variants for ablation studies.

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

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...
I am outgoing and have strong communication skills; … Two years of experience in UI work, interface design, product interaction, and web design. Proficient in using various design tools such as Axure and OmniGraffe; 3. Experience in interaction and prototype design. Design low-fidelity prototype images and achieving interactive effects. 4. Experience in hand drawn creation and other related fields.

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


