Serialized Knowledge Enhanced Multi-objective Person-job Matching Recommendation in a High Mobility Job Market

xiaowei shi
School of Economics and Management, Tsinghua University, Beijing 100084, China
shixw.17@sem.tsinghua.edu.cn

Jiarun Song
Meituan, Beijing, 100012, China
songjiarun@meituan.com

Junchao Wu
Meituan, Beijing, 100012, China
wujunchao@meituan.com

Qiang Wei
School of Economics and Management, Tsinghua University, Beijing 100084, China
qiangwei@tsinghua.edu.cn

Abstract

In a high mobility job market, accumulated historical sequences information from persons and jobs bring opportunities and challenges to person-job matching recommendation, where the latent preferences may significantly determine the success of person-job matching. Moreover, the sparse labels further limit the learning performance of recommendation methods. To this end, we propose a novel serialized knowledge enhancement multi-objective person-job matching recommendation method, namely SMP-JM. The key idea is to design a serialized multi-objective method from "intention-delivery-review", which effectively solves the problem of sparsity through the transmission of information and the serialization constraints between objectives. Specifically, we design various attention modules, such as self-attention, cross-attention and an orthogonal multi-head attention, to identify correlations between diversified features. Furthermore, a multi-granularity convolutional filtering module is design to extract personal latent preference from the historical sequential behaviors. Finally, the experimental results on a real-world dataset validate the performance of SMP-JM over the baseline methods.

Keywords: person-job matching, multi-objective learning, sequential features

1. Introduction

According to a job market survey, 35.8% of white-collar workers are considering changing jobs, and over 70% of persons have already started resumes preparation and deliver[1]. Moreover, the high mobility of the job market can be observed, i.e., the rapid generation of new jobs and high frequency of persons switching jobs, brings new characteristics to person and job matching recommendation, especially for online recruitment platforms.

The high mobility job market has accumulated affluent historical sequences for the person-job matching, such as person's historical work experience sequence, resume update sequence, job delivery sequence and corresponding job review sequence. These historical sequences contain substantial and latent preference and attribute information of persons and jobs as well as the interactions between them, which may contribute in-depth insights of person-job matching. Thus, it is promising to design an appropriate machine learning method to discover knowledge from historical sequential features to facilitate effective person-job matching recommendation.

Nevertheless, the high mobility market also amplifies the personalization and dynamics of the bilateral preferences from both persons and jobs, since both of them may have more diverse choices in the person-job matching process, which might increase the complexity in designing a recommendation method.

In addition to generating more diversified and complex bilateral preferences, the rapid change of persons and jobs exacerbates the sparsity of labels, i.e., more new or updating persons and jobs enter and leave the platforms frequently, e.g., leading to more cold start cases. Meanwhile, due to a large number of failed resume delivery and non-delivery behaviors, i.e., resulting into null labelling, the sparsity problem

[1]http://www.ce.cn/xwzx/gnsz/gdxw/201809/27/t20180927_30395952.shtml

URI: https://hdl.handle.net/10125/102750
978-0-9981331-6-4

caused by one-class label is quite prominent in this context and significantly weakens the recommendation performance.

The above characteristics put forward higher requirements for designing an intelligent person-job matching recommendation method, which is expected to possess three merits. First, it can comprehensively analyze the persons' and jobs' bilateral and diversified preferences from historical sequences. Second, it can well cope with reciprocity nature of person-job matching problem and output satisfactory recommendation for both sides. Third, the recommendation model is expected to well cope with the sparse labels.

Related researches have utilized collaborative filtering technology (Peng, 2016), content-based recommendation technology (Guo, 2014) and pattern recognition technology to implement the matching of persons and jobs. Except the structured features of persons and jobs, some researches started to incorporate two types of important unstructured texts, i.e., the person resumes and job descriptions (JD), to learn more detailed attributes and preferences (Qin, 2020).

However, existing person-job matching methods have been facing some challenges. First, the textual features contain and reflect little persons or jobs' personalized and dynamic preferences, which nevertheless can be detected from historical interactions. Few existing works take the abundant sequential features in historical interactions into consideration, which obviously limits the recommendation performance. Second, most of them only focus on the explicit and manifested information (Le, 2019), e.g., listed skills and requirements, while lacking the detection of certain important latent preferences, such as the preferences reflected from mutual clicks between persons and jobs, and the long-term and short-term preferences changing from sequential features, especially in the dynamic and diversified context. These latent preferences are also the key factors in determining whether the person and the job are successfully matched with reciprocity (Belavina, 2020). Third, the sparse labels, which is aggravated in the high mobility context, cannot be well addressed with most of the existing methods.

Therefore, this paper proposes a serialized knowledge enhanced multi-objective person-job matching recommendation method, called SMP-JM. This method can take the advantage of the historical sequences and establish a reciprocal matching recommendation in a sequential and multi-step multi-objective manner, which can not only capture the latent and diversified preferences from persons and jobs, but also tolerant the labelling sparsity.

First of all, the profile and auxiliary historical behaviors of persons and jobs i.e., the work experience, the resume update, the job delivery, and the job review, are comprehensively incorporated into the method, from which, the latent and diversified preferences and attributes of persons and jobs can be better learned to facilitate in-depth analysis in a dynamic manner.

Furthermore, since the person-job matching is essentially a multi-objective reciprocal process, we leverage the historical delivery and review information to design a deep learning module with both attention mechanism and gating mechanism so as to implement the learning of the bilateral latent preferences of persons and jobs. In order to further fit the dynamic characteristics of personalized preferences in the context of high mobility, a multi-granularity convolutional filtering module is designed to identify long-term and short-term personalized preferences.

Finally, to solve the problem of labelling sparsity, a serialized multi-objective recommendation is proposed, showing its advantage, because the sharing and transmission of information between objectives can effectively alleviate the problem of sparse information in single objective.

An effective person-job matching can be formulated as a multi-objective learning with multi-step sequential relationship. The first step is to detect whether a person have intention to apply for a job, the next is to justify whether the person delivers his/her resume to the preferred jobs, and the final one is to determine whether (the employer of) a job feedbacks a positive review to a delivery. Thus, we construct the person-job matching into a three-objective learning, i.e., the prediction of person's intention within a time window, the prediction of person's delivery and the prediction of job's review. Among them, the third objective reflects the degree of matching of person and job, and achieves the goal with reciprocity. Furthermore, it's naturally in alignment with a typical multi-step job seeking sequence (intention-delivery-review), i.e., the delivery will start after the person has an intention to apply for a job, and the review will be feedbacked after the person has completed the delivery. This serialization constraint can be integrated to help formulate a serialized multi-objective framework to optimize the search space and solve the problem with sparsity.

Along this line, the information and target dependencies between the three objectives need to be established. For target dependency, we design a sequential relationship between multiple objectives, which can guide and optimize the complex search space, as well as improving the effectiveness of multi-objective learning. For information dependency, we propose a comprehensive attention mechanism to implement the information transfer between inter and intra of multiple features. Compared with CNN and RNN, the attention mechanism can flexibly capture long-distance information dependencies (Tian, 2021). Therefore, this paper optimizes and designs an orthogonal multi-head attention mechanism to achieve efficient integration of the information in each sequence. Finally, extensive experiments on a large-scale real-world dataset validate the effectiveness of our SMP-JM method compared with baselines. The contributions of the method are summarized as follows:

- Establish a sequential multi-objective framework to achieve satisfactory person-job matching recommendation, based on which the transferring of information and objective dependencies can well alleviate the problem of one-class label and sparsity.
- Model the domain characteristics of bilateral matching and latent preferences, with which the multi-granularity convolutional filtering modules are designed to subtly extract dynamic personal latent preference from the historical resume deliveries and review behaviors.
- Design comprehensive attention mechanisms, such as the orthogonal multi-head attention mechanism, which can better support the effective integration and learning of interaction relationships and important latent preferences.

The rest of this paper is organized as follows. In Section 2, we briefly introduce some related work. Section 3 introduces the preliminaries and formally defines the problem of person-job matching. Then, technical details of SMP-JM framework are presented in Section 4. Extensive experiments are discussed in Section 5. Section 6 concludes the paper.

2. Related Research

2.1. Job Recommendation

Many existing job recommendation methods are the direct migration of traditional recommendation algorithms, neglecting domain characteristics (Zhang and Cheng, 2015). In recent years, some researches have gradually taken into account the specialty and characteristics of job recruitment. Almalis (2015) integrated structured and unstructured job descriptions to calculate the "fitness" between persons and jobs. Sato (2017) proposed six kinds of interactions between persons and jobs, and assigned different scores to different interactions. Yang (2017) found that the precision is more important than the recall in the job recommendation.

The reciprocal matching feature has also been considered at the algorithm level in recent years. Through the matching algorithm, if the person's skills can meet the job requirements without excessive waste, the reciprocal recommendation will be achieved (Lian, 2017; Dave, 2018). Qin (2018) leveraged a Recurrent Neural Network to project words into latent representations. Zhu (2018) proposed a bipartite Convolutional Neural Network that can estimate whether a person fits a job, as well as identifying which specific requirements in the job have been satisfied by the person. Jiang (2020) fused the representations for the explicit and implicit preference of persons and recruiters by LSTM to get a more comprehensive representation for person-job fit. Qin (2020) incorporated two hierarchical topic-based ability-aware structures to guide the learning of semantic representation and incorporate the global meaning for job requirements and person experiences. He (2021) grouped the features of resumes and jobs into several fields and explored attention as well as residual connection. Recurring to gorgeous features, it is an effective way to learn the latent correlations between each field of features from both resumes and jobs.

Although the characteristics of person-job matching have been widely studied, the effect of recommendation model still suffered from the sparse and one-class label. This paper establishes a sequential multi-objective framework, which effectively solves the above problems through the interaction of information between multiple objectives.

2.2. Multi-objective Learning

Multi-objective learning is to train data with multiple objectives simultaneously, using shared representations to learn the common or specific knowledge contained in multiple objectives, as the learned features from one objective may be useful in learning another related objective (Caruana,1997; Vandenhende, 2020). Multi-objective learning has been extensively considered in the researches, e.g., natural language processing (NLP), speech recognition to vision (Krishna, 2018; Chen, 2018).

One of the widely used multi-objective learning models was proposed by Caruana (1997), which has a shared-bottom structure. Although this structure substantially reduces the risk of overfitting, it easily suffers from optimization conflicts caused by the differences between multi-objectives. To solve the above problem, instead of sharing bottom hidden layers, some recent approaches added different types of constraints on objective specific parameters (Misra, 2016). Cao (2018) designed a multi-objective convolutional neural network, in which, four Task Specific Networks (TSNets) and one Shared Network (SNet) are connected with partially shared structures. Inspired by the above idea, expert-bottom pattern was proposed to control how expert modules are shared across all objectives at the bottom of the multi-objectives model (Ma, 2018; Tang, 2020). Moreover, to solve the problem of negative transferring when objectives are not very relevant (Wang, 2021), Google proposed MMOE model (Ma, 2018) to construct gate control mechanism for each objective. Tencent (Tang, 2020) split the experts into shared experts and private experts, and divided the sample space by loss function to resolve the heterogeneous relationship between objectives. Therefore, the crucial challenges in multi-objectives learning are how to model relationships among objectives appropriately and how to exploit them to improve the performance of each objective (Zhao, 2019).

One direction to model relationships between objectives is to transfer probabilities in the output layers of different objectives (Wen, 2020). However, this treatment can only allocate simple probability information via the scalar product, ignoring richer and more useful representations, which might result in a great loss. Based on domain knowledge, this paper relies on the sequence characteristics between multi-objectives in the job seeking to build information transfer relationship and constraints between objectives, which can help optimize the search space and improve the efficiency and effectiveness of multi-objective learning.

3. A Serialized Multi-objective Person-Job Matching Recommendation

3.1. Problem Formulation

Generally speaking, the person-job matching recommendation is to, measure the person's intention, preference and the degree of matching between person and job, in a multi-step sequential manner, so as to achieve effective person and job recommendation with reciprocal benefit.

Two types of features will be used to predict the matching of person and job: sequential features and non-sequential features. The sequential features include person’s work experience sequence, resume update sequence, delivery sequence, and job’s review sequence. The non-sequential features include basic attributes of persons: age, school, city, etc., and basic attributes of jobs: name, business type, job description, etc. Formally, we denote the person set as $C = {c_i}_{i=1}^{numc}$ and the job set as $P = {p_j}_{j=1}^{nump}$,where $numc$ and $nump$ are the numbers of persons and jobs, respectively. Further, we denote the non-sequential features corresponding to person c_i as F^{c_i}, and the non-sequential features corresponding to job p_j as F^{p_j}. The job descriptions in F^{p_j} are unstructured text, and we denote them as $F_{des}^{p^j}$. Historical work experience sequence, resume update sequence, the delivery sequence and the job’s review sequence of person c_i are denoted as E^{c_i} , U^{c_i} , D^{c_i} and R^{c_i} , respectively, where, the description feature of the work experience in E^{c_i} is denoted as $E_{des}^{c_i}$

Therefore, a historical delivery, denoted as H={c_i, y_1, p_j, y_2 , y_3 }, means person c_i generates intention y_1, and then delivers resume to job p_j, i.e., generating label y_2, as well as gets the review feedback from the job, i.e., getting label y_3, where $y_1 \in \{0,1\}$, $y_2 \in \{0,1\}$ and $y_3 \in \{0,1\}$. The multiple objectives of person-job matching are to predict the person's intention y_1, the person's preference for the job y_2, that is, whether to deliver, and job preference for person, i.e., y_3. For one delivery, the generation of three labels are naturally endogenous with the multi-step sequential relationship, i.e., first generating intention, next delivery and then review. Therefore, a partial dependency relationship of three label, $y_1 \geq y_2 \geq y_3$, can be implanted to improve the search efficiency and prediction accuracy. In summary, given sequential and non-sequential features, the comprehensive matching function M can be learnt, with the mapping of M $(F, E, U, D, R) \rightarrow \{y_1, y_2, y_3\}$.

3.2. A Multi-objective Learning Framework

Figure 1 illustrates the overall framework of the multi-objective person-job matching recommendation method. Concretely, SMP-JM takes two types of features, i.e., sequential and non-sequential features, from persons and jobs as input. An embedding layer is designed to map different types of features into vectorization representation. Next, the embedding features are fed into expert module. Specifically, all of sequential features are input into BiGRU to get comprehensive encoding of update, delivery, review and work experience sequential features, based on which, the potential dependency within the same type of features can be learnt.

For more complex unstructured text features, an Orthogonal Multi-head Attention (OMA) mechanism is proposed to achieve multi-spatial semantic capture and improve the parameters learning efficiency, which can fully tap the semantic interaction of each potential dimension and realize the mining of global and local semantic relations. Moreover, for four types of sequential features, this paper uses a sequential features integration layer and convolution networks with different kernel granularity to filter and integrate the sequence information with different time lengths.

Further, the outputs of the above series of expert modules are used as the input of three multi-objectives towers. For each tower, it uses the attention gate to filter the overall diversified signs. The gated mechanism can automatically realize the selection of information under each tower. Specifically, this paper establishes the information dependency between output of each tower, which can further optimize the search space and improve the efficiency and accuracy of prediction. Finally, we generate the intention score, preference score and matching score based on the above designs. The objective function utilizes label loss, reciprocity loss and sequence constraints based on three scores to achieve end-to-end multi-objective learning.

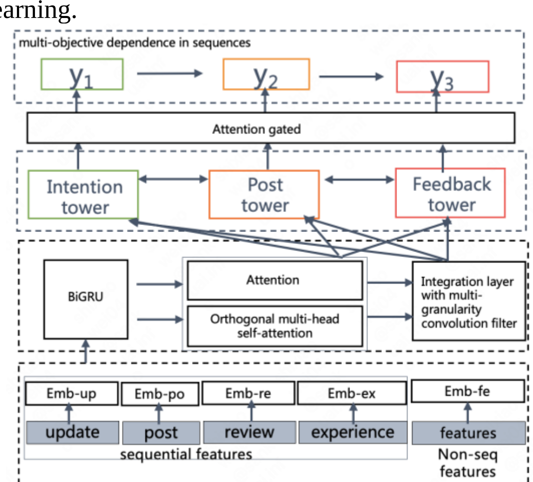

Figure 1. the framework of SMP-JM.

4. Method Implementation

4.1. BiGRU Layer

Firstly, the numerical features are normalized with standard distribution and the categorical features are projected into one-hot vectors. Therefore, the vectorized representation matrices of four kinds of sequential features and non-sequential features can be obtained, namely E, U, D, R and F. Among them, each item in E contains the basic attribute and description text vector of the job. U is the time series matrix of resume updates, where the daily update time is represented as a row vector of the matrix. The sequential basic attribute and description text vector of the delivered job are included in D. Each item in F contains the basic attribute vector of the candidate and the label of the review.

Encoding layer is composed of a double BiGRU, where one is for learning the semantic information of text feature, and the other is for encoding sequential feature. BiGRU can capture semantics and dependencies of sequential information, as well as possessing less training parameters and higher training efficiency than BiLSTM (Han, 2020). Specifically, with the word vectors of input $x_j \in \mathbb{R}^{n*d}$ and a pair of forward and backward GRU, two sequences of hidden states $\overleftarrow{h_j} \in \mathbb{R}^{n_j*d}, \overrightarrow{h_j} \in \mathbb{R}^{n_j*d}$ can be generated, and the final output $h_j \in \mathbb{R}^{n_j*d}$ is generated by summing $\overleftarrow{h_j}$ and $\overrightarrow{h_j}$, where n_j is the max length of input and d is the dimension of hidden states.

Therefore, the text features of job description F_{des}, and the text features of work experience E_{des}, can be integrated into BiGRU, deriving h_{jdes}, h_{edes}. The sequential features, which include the text features, i.e., experience, delivery and review features, E, D and R, are also brought into the isomorphic BiGRU for information integration, deriving h_E, h_D and h_R. Furthermore, the update sequential feature U is encoded by a heterogeneous BiGRU to get h_U.

$$\overleftarrow{h_j} = [\overleftarrow{h_j^1}, \overleftarrow{h_j^2}, \overleftarrow{h_j^3} \dots \overleftarrow{h_j^n}] = \overleftarrow{GRU_j}(x_j) \quad (1)$$

$$\overrightarrow{h_j} = [\overrightarrow{h_j^1}, \overrightarrow{h_j^2}, \overrightarrow{h_j^3} \dots \overrightarrow{h_j^n}] = \overrightarrow{GRU_j}(x_j) \quad (2)$$

$$h_j = [h_j^1, h_j^2, h_j^3 \dots h_j^n] = [\overrightarrow{h_j^1} + \overleftarrow{h_j^1}, \dots, \overrightarrow{h_j^n} + \overleftarrow{h_j^n}] \quad (3)$$

4.2. Attention Layer

With ample features, the comprehensive and valuable relationships among them need to be detected to learning the potential information supporting multi-objective decision. Thus, attention layer is designed to understand potential dependencies within and between each type of features. A biased general alignment is adopted for non-sequential features F (Sordoni, 2016) to derive ATT_F, and the formulas are shown in (4) - (6), where W_h and W_o are projection matrices and b_r is the biased term.

$$s_m^i = tanh(W_h \cdot X_m + W_o \cdot X_i + b_r) \quad (4)$$

$$a_n^i = exp(s_n^i) \,/\, \textstyle\sum_{m=1}^{2k-1} exp\,(s_m^i) \quad (5)$$

$$attention\,(X_i, X_m) = \textstyle\sum_{n=1}^{2k-1} a_n^i X_n \quad (6)$$

However, single attention inhibits the jointly learning from different representation subspaces (Vaswani, 2017). Therefore, for text with more complex information, the Multi-head Attention (MA) s is utilized to learn the comprehensive information dependencies. Although multi-head can capture effective and robust relevance and improve the performance of the model, some researches argued that many of the head projected dimensions are redundant and some heads might attend on similar feature in input space (Voita, 2019; Cordonnier, 2020). Therefore, instead of learning projection matrices W_Q^i , W_K^i and W_V^i of each head i, separately, QW_Q , KW_K and VW_V are set as three entries to learn, and a matrix M is used to transform the original matrix into orthogonal multi-head space. Then, the multi representations are connected to obtain the final orthogonal multi-head attention, as shown in formulas (7-8), where M is the eigenvector matrix corresponding to the first k eigenvalues of $XW（XW）^T = XWW^TX^T$, $X \in \{K, Q, V\}$, and f is the attention function. The two types of text features, h_{jdes} and h_{edes}, are input into the orthogonal multi-head attention mechanism, and the corresponding attention feature representation $OMATT_{F_{des}}$ and $OMATT_{E_{des}}$ can be obtained.

$$OMATT^i = f(M_Q^i QW_Q, M_K^i KW_K) \cdot M_V^i VW_V \quad (7)$$

$$OMATT = OMATT^1 \oplus \ldots \oplus OMATT^8 \quad (8)$$

4.3. Sequential Features Integration

We take the review sequence as an example to depict the integration of sequential features. Based on the review sequential features, the latent preferences of jobs in sequential form can be obtained, denoted as P, where $P = [p_0,\ p_1, \ldots, p_{maxr}]$ and $maxr$ denotes the maximum length of the review sequential feature. The initial preference is set as p_0=$h_{R,0}$, and the latent preference P can be calculated by formulas (9)-(12) with encoded review sequential feature h_R. $\odot$ is the element-wise multiplication, and W_r, W_z, U_z, U_r, U_u and v_z, v_r, b are related parameters. The tanh() is configured as the activation function.

$$r = softmax({v_r}^T \tanh\,(W_r h_{R,i+1} + U_r * p_i) \quad (9)$$

$$upd = \tanh\,(U_u\big(r \odot h_{R,i+1}\big) + b) \quad (10)$$

$$z = softmax({v_z}^T \tanh\,(W_z h_{R,i+1} + U_z * upd) \quad (11)$$

$$p_{i+1} = z \odot upd + (1 - z) \odot p_i \quad (12)$$

Further, to capture the dynamics of latent preferences, the review sequence is needed to be processed granularly. Although the timestamp is considered as one of the key features, the dependency on the length of time varies differently with different features. Therefore, the inception strategy in image processing is adopted to deploy a Multi-granularity Convolution Filter (MCF) layer to integrate different time-dependent characteristics. Specifically, the multi-granularity convolution filter layer takes the inherent attributes of time into account and characterizes the commonly used time periods 2, 3, 7 and 12, as kernel sizes to filter. Therefore, a 1D convolutional filter with four kernel sizes can be designed, and the output $conv_{1*\mathrm{n}}$ can be constructed in formula (13) with latent sequential preferences of jobs, where f is the convolution kernel function.

Similarly, the encoded sequential features can be learnt by a similar sequential feature integration layer, getting $\boldsymbol{conv}^E$, $\boldsymbol{conv}^U$, $\boldsymbol{conv}^P$, $\boldsymbol{conv}^R$ (formula (14)), which indicates the comprehensive preference of persons and jobs.

$$\boldsymbol{conv_{1*num}} = \sum_{\boldsymbol{e=0}}^{\boldsymbol{num-1}} \boldsymbol{f_{1*num}(e)\, P * e} \quad (13)$$

$$\boldsymbol{conv} = \boldsymbol{chan}(\boldsymbol{conv_{1*2}}, \boldsymbol{conv_{1*3}}\ , \boldsymbol{conv_{1*7}}\ , \boldsymbol{conv_{1*12}}) \quad (14)$$

4.4. Tower Layer and Loss Function

Given a set of features handled by the expert layers, ATT_F $\boldsymbol{conv}^E$, $\boldsymbol{conv}^U$, $\boldsymbol{conv}^P$, $\boldsymbol{conv}^R$, the tower layer for each objective is defined as formula (15), and the output of each tower can be obtained as $Tower_1$, $Tower_2$ and $Tower_3$, respectively. The three outputs have an endogenous multi-step sequential relationship as indicated previously. The step n is affected by the output of this tower and the output of the previous n-1 tower, as shown in formula (16). The latent semantic similarity is proposed to measure the current information transmission relationship, as shown in formula (17), where d is the dimension of the output of tower. Based on the construction of sequence dependence, a MLP is used to predict the label for each objective, as shown in formula (18).

$$Tower_{objective} = MLP(gate(\ ATT_F, \boldsymbol{conv}^E, \boldsymbol{conv}^U, \boldsymbol{conv}^P, \boldsymbol{conv}^R)) \quad (15)$$

$$o_n = \frac{\exp\,(w_n)}{\exp\,(w_{n-1}) + \exp\,(w_n)} Tower_n + \frac{\exp\,(w_{n-1})}{\exp\,(w_{n-1}) + \exp\,(w_n)} g(o_{n-1}) \quad (16)$$

$$w_n = \frac{g(o_{n-1}) \cdot Tower_n}{\sqrt{d}} \quad (17)$$

$$\widehat{y_i} = sigmoid(MLP(o_i)) \quad (18)$$

Finally, the loss function of in this method includes three parts: label loss, reciprocity loss and constraint loss (formula (19)). The label loss under each objective, ${L^{label}}_i$, is the cross-entropy loss function of the predicted label $\widehat{y_i}$ and real label y_i (formula (20)), and is integrated for three objectives by weights (formulas (21-22)) to balance the difference of learning rate and loss scale between objectives. Reciprocity loss comes from the reciprocity particularity of person-job matching, i.e., the consistency of job preference and person preference in a successful interaction (Xia, 2019). The output of the corresponding tower layer is used as the latent preference of person and job, and the goal of the loss function is to minimize the distance between both parties for each delivery with positive review (formula (23)). Constraint loss (formula (24)) is used to restrict the dependency of target values between multi-step sequential multiple objectives, i.e., $y_1 \geq y_2 \geq y_3$.

$$L^{total}(t) = L^{label}(t) + L^{reciprocity}(t) + L^{constraint}(t) \quad (19)$$

$$L^{label}{}_i = y_i \, log(\widehat{y_i}) + (1 - y_i) log\,(1 - \widehat{y_i}) \quad (20)$$

$$L^{label}(t) = \sum_{i \in [[1,2,3]]} \alpha_i(t) L^{label}{}_i(t) \quad (21)$$

$$\alpha_i(t) = \frac{exp\,((L_i(t) - L_i(t-1)/L_i(t-1)))}{\sum_{n \in [1,2,3]} exp\,((L_n(t) - L_n(t-1)/L_n(t-1)))} \quad (22)$$

$$L_{reciprocity}(t) = y_3 * dis(Tower_2, Tower_3) \quad (23)$$

$$L_{constraint} = cons(\widehat{y_1} \geq \widehat{y_2} \geq \widehat{y_3}) \quad (24)$$

5. Experiments

5.1. Data Description

The dataset is from a leading IT enterprise's talent and job database. In order to protect the privacy, the resumes have been anonymized and the text information in the job experience is desensitized by word segmentation tools and word2vector. We extracted the delivery records of persons who have more than three delivery actions in history and corresponding job review information from January to March, 2022. In total, the dataset includes 28,558 persons and 20,231 jobs, with 267,728 pairs of delivery records.

5.2. Evaluation Methods and Baselines

For comparison, the widely used metrics, i.e., accuracy (Accu), recall, precision (Prec), F1 and AUC are utilized, i.e., the greater the better. The proposed SMP-JM method is compared with several mainstream methods, i.e., DeepFM, PJFNN, PJFFF and MUFFIN. DeepFM combines the factorization machines and deep learning for feature learning in a neural network architecture, which emphasizes both low- and high- order feature interactions (Guo, 2017). PJFNN utilizes a bipartite neutral network to learn the joint representation of person-job fitness from historical job applications (Zhu, 2018). PJFFF extracts features from the resume and job description and learns the representation of implicit preference by the historical applications (Jiang, 2020). MUFFIN is a multi-filed feature based person-job fit method. Recurring to gorgeous features, MUFFIN designs a solution to learn the latent correlations between each field of features from both resumes and jobs (He, 2021). The features and key modules used by the baseline methods and the SMP-JM are summarized in Table 1. It can be found that only PJFFF uses historical interaction data to learn the latent preferences of persons and jobs. However, the impact of the sequential nature in the interactions is not considered. Moreover, baseline methods are mostly single-objective works of job recommendation or person-job bilateral matching. Therefore, we compare results on each objective, and performances are as shown in Table 2.

Table 1. A Summary of the features and key module.

	Unstructured	text	Historical interaction	Key module	Obje.
DeepFM	√	×	×	FM DNN	single
PJFNN	×	√	×	CNN	single
PJFFF	√	√	√	DeepFM CNN LSTM	single
MUFFIN	√	√	×	pre-trained embedding MA MLP	single
SMP-JM	√	√	√	BiGRU OMA MCF	multi

Table 2. The performance of SMP-JM and baselines.

	AUC	Recall	F1	Accu	Prec
SMP-JM obj 1	**0.7769**	**0.7235**	**0.7841**	**0.7100**	**0.8559**
SMP-JM obj 2	**0.9282**	**0.8226**	**0.8450**	**0.8371**	**0.8685**
SMP-JM obj 3	**0.8072**	**0.7699**	**0.7890**	**0.8870**	**0.8091**
DeepFM obj 2	0.6797	0.6270	0.6250	0.6423	0.6233
DeepFM obj 3	0.6194	0.5711	0.5678	0.6631	0.5645
PJFNN obj 3	0.7071	0.6600	0.7039	0.7562	0.7540
PJFFF obj 3	0.7734	0.6923	0.7355	0.8033	0.7845
MUFFIN obj 3	0.7354	0.6788	0.7057	0.8007	0.7349

The results show that, the SMP-JM method outperforms all the baseline methods on every objective, which verifies that the proposed method can subtly distinguish preferences of persons and jobs, i.e., essentially leading to high-accuracy predictions. Especially, DeepFM performs the worst in the objective 3, due to the lack of design for bilateral matching, which asserts the importance of modeling the bilateral information. PJFFF performs better than other baselines, since it to some extent captures the explicit and implicit features for bilateral matching. On this basis, the proposed SMP-JM method well deals with the sequential features in a reciprocal manner and exquisitely utilizes the multi-step sequential relationship between all objectives, which further enriches the mutual labelling abundancy. In doing so, i.e., by serializing the multi-objectives, SMP-JM can better deal with the sparse labels, and achieve the prediction outperformance.

5.3. Ablation Experiments

The ablation experiments from four angles are conducted with results shown in Table 3. First, we explore the influence of orthogonal multi-head attention mechanism, i.e., SMP-JM - w/o head. It can be seen that the performances of objective 1, 2, and 3 are slightly reduced (e.g., about 0.5% on accuracy in each objective), as the design of the orthogonal multi-head mechanism is mainly to improve the efficiency. With the addition of the orthogonal multi-head mechanism, the parameter scale is reducing from 38230030 to 38035825, i.e., about 230000. Next, by removing multi-granularity temporal convolutional layer, SMP-JM- w/o multik is compared. It is found that the multi-granularity temporal convolutional layer contributes an improvement of roughly 3% to accuracy, as it is able to better capture diverse length-dependent sequential features. Then, we verify whether the design of attention is effective by excluding all of the self and cross attention, that is SMP-JM- w/o atten, and the reduced accuracy of the objectives 1, 2 and 3 are about 7%, 4% and 12%, respectively. Specifically, it is crucial for objective 3, as the attention mechanism shows significant advantage in capturing the long distances and potential dependencies. These results confirm the necessity and value of attention modeling. Finally, we verify the importance of historical sequential features, i.e., SMP-JM- w/o seq. It is certain that the accuracy of objective 1 have dropped significantly to about 54%, because sequential features are important source used to learn intention of persons. The accuracy of objectives 2 and 3 are also reduced by 11% and 20% respectively, indicating the important role of the learnt latent preferences of persons and jobs based on sequential features in improving the accuracy of person-job matching.

Table 3. The performance of ablation experiments.

	AUC	Recall	F1	Accu	Prec
SMP-JM obj 1	**0.7769**	**0.7235**	**0.7841**	**0.7100**	**0.8559**
SMP-JM obj 2	**0.9282**	**0.8226**	**0.8450**	**0.8371**	**0.8685**
SMP-JM obj 3	**0.8072**	**0.7699**	**0.7890**	**0.8870**	**0.8091**
SMP-JM - w/o head obj 1	0.7747	0.7097	0.7735	0.7086	0.8499
SMP-JM- w/o head obj 2	0.9097	0.8224	0.8378	0.8296	0.8538
SMP-JM- w/o head obj 3	0.8048	0.6991	0.7477	0.8739	0.8035
SMP-JM- w/o multik obj 1	0.7036	0.6977	0.7411	0.6830	0.7902
SMP-JM- w/o multik obj 2	0.8863	0.8071	0.8141	0.8021	0.8212
SMP-JM- w/o multik obj 3	0.7822	0.7222	0.7504	0.8506	0.7810
SMP-JM- w/o atten obj 1	0.6165	0.6039	0.6501	0.6198	0.7040
SMP-JM- w/o atten obj 2	0.8529	0.7843	0.7864	0.7593	0.7887
SMP-JM- w/o atten obj 3	0.7143	0.5987	0.6489	0.7402	0.7084
SMP-JM- w/o seq obj 1	0.4289	0.5037	0.5117	0.5459	0.5199
SMP-JM- w/o seq obj 2	0.7264	0.7723	0.7496	0.7235	0.7282
SMP-JM- w/o seq obj 3	0.6788	0.5206	0.5277	0.6888	0.5351

5.4. Hyper-parameter Sensitivity Experiments

In this section, the sensitivity of four hyper-parameters of SMP-JM is analyzed. Figure 2 shows the AUC when varying the size of hyper-parameters. it is obvious that except the number of dimensions of hidden-state is too small (less than 32), the AUC remains stable with the changes of the dimension. Then, it can be observed that the performance is almost on horizontal lines when the head move from 4 to 10, indicating that SMP-JM is insensitive to the change of the number of head.

Further, we adjust the length of the sequential features within a certain range. It can be found that with the decrease of the length of the deliver/review sequence, the AUC decreases slightly, but the overall performance is stable when it varies within a certain range, i.e., from five to nine. It shows that a certain amount of historical interaction information can bring stable effect to the model and a similar conclusion can be verified in the update sequence.

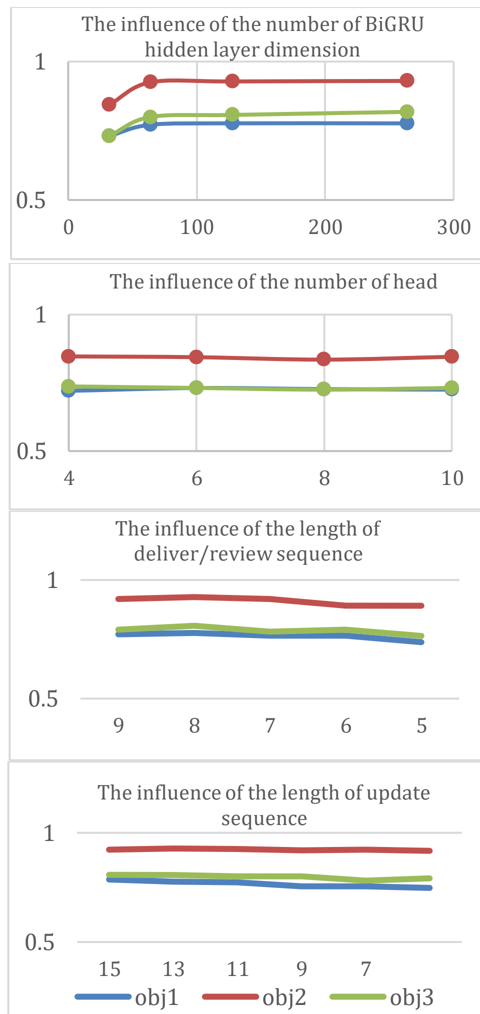

Figure 2. The AUC with the size varying.

5.5. Robustness Check

The serialized multi-objective learning possesses a merit on alleviating the labelling sparsity, which is quite prominent in the high mobility person-job matching context. Specifically, many jobs have few person-review records, and most of persons' deliveries are somewhat of cold-start nature. The transferring of information between multiple objectives and the constraints based on the sequence relationship can be well captured in the proposed method, which can effectively deal with these problems. Therefore, the robustness check is conducted on the serialized multi-objectives SMP-JM method and the single-objective MUFFIN method under different degrees of data sparsity in objective 3, that is, the prediction of bilateral preferences matching. Referring to Hu's cold-start validation method (2018), we divided the dataset into 5 equal datasets, and gradually increased the training data from one to four, corresponding to 20%, 40%, 60%, and 80% of the training data. The vertical axis of Figure 3 represents AUC, and the horizontal axis represents different sizes of training set. The results show that the proposed SMP-JM method shows stable performance under different sizes of training set. However, with the increase of data sparsity, the proportion of the training set decreases, the AUC of MUFFIN method decreases significantly. It shows that the serialized multi-objective learning has a satisfactory capability to improve the stability on the person-job matching recommendation, effectively solving the problems caused by the sparse data and labels.

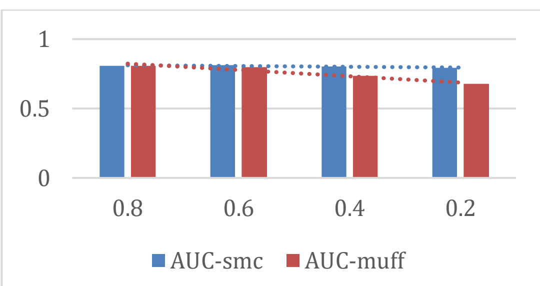

Figure 3. The AUC with the training set varying.

6. Conclusion

In order to cope with the opportunities and challenges brought by high mobility person-job matching context, we propose a novel person-job matching recommendation method SMP-JM. SMP-JM is an end-to-end multi-objective learning method, which integrates the historical sequences, including work experience, resume update and delivery sequence of persons, and review sequence information of jobs, etc. Through modeling the transferring of information between multiple objectives, it realizes effective prediction of persons' intention, preferences and bilateral matching. Then, the learning of latent preferences is well developed by designing various attention mechanisms, self-attention and orthogonal multi-head attention. The sequence information can be well processed by the design of integration layer with a multi-granularity temporal convolution, and the problem caused by sparse labels can be carefully alleviated by capturing the dependencies and information transfer relationships between objectives. Extensive data experimental results show the outperformance of the proposed SMP-JM method. Future studies will investigate the interpretability and efficiency of the method.

Acknowledgements

This work is supported by National Natural Science Foundation of China (72172070) and the MOE Project of Key Research Institute of Humanities and Social Sciences at Universities (17JJD630006).

7. References

Almalis, N. D., Tsihrintzis, G. A., Karagiannis, N., & Strati, A. D. (2015, July). FoDRA—A new content-based job recommendation algorithm for job seeking and recruiting. In 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA). IEEE. (pp. 1-7).

Belavina, E., Girotra, K., Moon, K., & Zhang, J. (2020). Matching in Labor Marketplaces: The Role of Experiential Information. Available at SSRN 3543906.

Cao, J., Li, Y., & Zhang, Z. (2018). Partially shared multi-task convolutional neural network with local constraint for face attribute learning. In Proceedings of the IEEE Conference on computer vision and pattern recognition. pp. 4290-4299.

Caruana, R. (1997). Multitask learning. Machine learning, 28(1), 41-75.

Chaudhari, S., Mithal, V., Polatkan, G., & Ramanath, R. (2021). An attentive survey of attention models. ACM Transactions on Intelligent Systems and Technology (TIST), 12(5), 1-32.

Chen, Y., Zhao, D., Lv, L., & Zhang, Q. (2018). Multi-task learning for dangerous object detection in autonomous driving. Information Sciences, 432, 559-571.

Cordonnier, J. B., Loukas, A., & Jaggi, M. (2020). Multi-head attention: Collaborate instead of concatenate. arXiv preprint arXiv:2006.16362.

Dave, V. S., Zhang, B., Al Hasan, M., AlJadda, K., & Korayem, M. (2018, October). A combined representation learning approach for better job and skill recommendation. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management. pp. 1997-2005.

Guo, H., Tang, R., Ye, Y., Li, Z., & He, X. (2017). DeepFM: a factorization-machine based neural network for CTR prediction. arXiv preprint arXiv:1703.04247.

Guo, X., Jerbi, H., & O'Mahony, M. P. (2014, September). An analysis framework for content-based job recommendation. In 22nd International Conference on Case-Based Reasoning (ICCBR), Cork, Ireland, 29 September-01 October 2014.

Han, Y., Liu, M., & Jing, W. (2020). Aspect-level drug reviews sentiment analysis based on double BiGRU and knowledge transfer. IEEE Access, 8, 21314-21325.

He, M., Shen, D., Wang, T., Zhao, H., Zhang, Z., & He, R. (2021). Self-Attentional Multi-Field Features Representation and Interaction Learning for Person-Job Fit. IEEE Transactions on Computational Social Systems.

Hu, B., Shi, C., Zhao, W. X., & Yu, P. S. (2018, July). Leveraging meta-path based context for top-n recommendation with a neural co-attention model. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1531-1540.

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

Krishna, K., Toshniwal, S., & Livescu, K. (2018). Hierarchical multitask learning for ctc-based speech recognition. arXiv preprint arXiv:1807.06234.

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

Lian, J., Zhang, F., Hou, M., Wang, H., Xie, X., & Sun, G. (2017). Practical lessons for job recommendations in the cold-start scenario. In Proceedings of the Recommender Systems Challenge 2017.pp. 1-6.

Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L., & Chi, E. H. (2018, July). Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 1930-1939.

Misra, I., Shrivastava, A., Gupta, A., & Hebert, M. (2016). Cross-stitch networks for multi-task learning. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3994-4003.

Qin, C., Zhu, H., Xu, T., Zhu, C., Jiang, L., Chen, E., & Xiong, H. (2018, June). 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. pp. 25-34.

Qin, C., Zhu, H., Xu, T., Zhu, C., Ma, C., Chen, E., & Xiong, H. (2020). An enhanced neural network approach to person-job fit in talent recruitment. ACM Transactions on Information Systems (TOIS), 38(2), 1-33.

Sato, M., Nagatani, K., & Tahara, T. (2017). Exploring an optimal online model for new job recommendation: Solution for recsys challenge 2017. In Proceedings of the Recommender Systems Challenge 2017.pp. 1-5.

Sordoni, A., Bachman, P., Trischler, A., & Bengio, Y. (2016). Iterative alternating neural attention for machine reading. arXiv preprint arXiv:1606.02245.

Tang, H., Liu, J., Zhao, M., & Gong, X. (2020, September). Progressive layered extraction (ple): A novel multi-task learning (mtl) model for personalized recommendations. In Fourteenth ACM Conference on Recommender Systems. pp. 269-278.

Tian, Z., Wang, C., Xiao, Y., & Lin, Y. (2021). Flexible scene text recognition based on dual attention mechanism. Concurrency and Computation: Practice and Experience, 33(22), e5863.

Vandenhende, S., Georgoulis, S., & Gool, L. V. (2020, August). Mti-net: Multi-scale task interaction networks for multi-task learning. In European Conference on Computer Vision. pp. 527-543).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Voita, E., Talbot, D., Moiseev, F., Sennrich, R., & Titov, I. (2019). Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned. arXiv preprint arXiv:1905.09418.

Wang, Y., Zhang, D., & Wulamu, A. (2021). A Multitask Learning Model with Multiperspective Attention and Its Application in Recommendation. Computational Intelligence and Neuroscience, 2021.

Wen, H., Zhang, J., Wang, Y., Lv, F., Bao, W., Lin, Q., & Yang, K. (2020, July). Entire space multi-task modeling via post-click behavior decomposition for conversion rate prediction. In Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. pp. 2377-2386.

Xia, B., Yin, J., Xu, J., & Li, Y. (2019). WE-Rec: A fairness-aware reciprocal recommendation based on Walrasian equilibrium. Knowledge-Based Systems, 182, 104857.

Yang, S., Korayem, M., AlJadda, K., Grainger, T., & Natarajan, S. (2017). Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach. Knowledge-Based Systems, 136, 37-45.

Yi, P., Yang, C., Li, C., & Zhang, Y. (2016, October). A job recommendation method optimized by position descriptions and resume information. In 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC).IEEE. pp. 761-764.

Zhao, J., Du, B., Sun, L., Zhuang, F., Lv, W., & Xiong, H. (2019, July). Multiple relational attention network for multi-task learning. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & Data Mining. pp. 1123-1131.

Zhu, C., Zhu, H., Xiong, H., Ma, C., Xie, F., Ding, P., & 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.