A Cross-View Hierarchical Graph Learning Hypernetwork for Skill Demand-Supply Joint Prediction

Wenshuo Chao[1,2], Zhaopeng Qiu[2], Likang Wu[2,3], Zhuoning Guo[1],
Zhi Zheng[2,3], Hengshu Zhu[2,1]*, Hao Liu[1]*

[1] The Hong Kong University of Science and Technology (Guangzhou)
[2] Career Science Lab, BOSS Zhipin
[3] University of Science and Technology of China
{wchao829,zguo772}@connect.hkust-gz.edu.cn, {zhpengqiu, zhuhengshu}@gmail.com,
{wulk,zhengzhi97}@mail.ustc.edu.cn, liuh@ust.hk

Abstract

The rapidly changing landscape of technology and industries leads to dynamic skill requirements, making it crucial for employees and employers to anticipate such shifts to maintain a competitive edge in the labor market. Existing efforts in this area either rely on domain-expert knowledge or regarding skill evolution as a simplified time series forecasting problem. However, both approaches overlook the sophisticated relationships among different skills and the inner-connection between skill demand and supply variations. In this paper, we propose a Cross-view Hierarchical Graph learning Hypernetwork (CHGH) framework for joint skill demand-supply prediction. Specifically, CHGH is an encoder-decoder network consisting of i) a cross-view graph encoder to capture the interconnection between skill demand and supply, ii) a hierarchical graph encoder to model the co-evolution of skills from a cluster-wise perspective, and iii) a conditional hyperdecoder to jointly predict demand and supply variations by incorporating historical demand-supply gaps. Extensive experiments on three real-world datasets demonstrate the superiority of the proposed framework compared to seven baselines and the effectiveness of the three modules.

Introduction

The rapid advancement of information technology (*e.g.*, 5G, VR, and generative AI) has significantly reshaped the skill requirements. According to the released reports, the required skill set in the job market has changed by about 25% since 2015 (Communications 2022). Take machine-learning-related skills for example, there has been a booming demand for deep-learning-related skills since 2014 and Large Language Model (LLM) (Radford et al. 2018) related skills in the past year, but declining demand on classic statistical learning skills, as illustrated in Figure 1. Such shifts in skill requirements have led to a significant imbalance of skill demand and supply in various industries, commonly referred to as the skill gap (Larsen et al. 2018), which may result in unemployment and business failures as industries evolve (Kim, Hsu, and Stern 2006; McGuinness and Ortiz 2016; Donovan et al. 2022). It is crucial to forecast and analyze the skill supply and demand to help various stakehold-
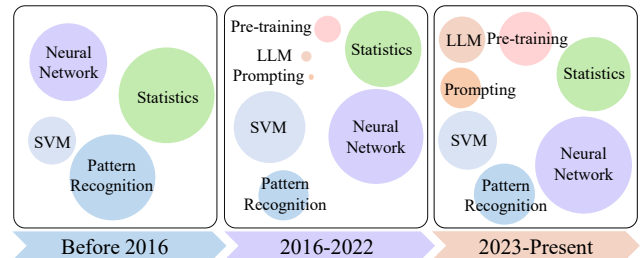

Figure 1: Real-world skill demand variation of machine-learning related jobs, data ranged from 2015 to 2023.

ers anticipate and proactively handle the skill gap. For instance, job seekers can plan relevant skill-learning paths in advance, companies can optimize their workforce planning strategies, and policymakers can develop policies to guide market development.

Existing studies on skill demand and supply prediction can be broadly categorized into expert-based and learning-based methods. Expert-based methods analyze skill shifts by exploring survey data (Belhaj and Tkiouat 2013). Such methods usually deliver coarse-grained forecasting results via expert domain knowledge (Bughin et al. 2018). Learning-based methods predict more fine-grained skill shifts by exploiting patterns and implications based on machine learning techniques. For example, Xu *et al.* (Xu et al. 2018) measure the popularity of skills via topic model and Macedo *et al.* (de Macedo et al. 2022) predict skill demands by adopting the recurrent neural network. However, both classes of approaches overlook the sophisticated demand-supply relationships among different skills grounded in structural and contextual domains. In this work, we investigate the joint prediction of skill demand and supply by explicitly modeling the collaborative relationships between different skills.

However, this is a non-trivial task due to the following technical challenges. First, the demand and supply variation of each skill in the labor market is correlated yet diversified. For instance, the breakthrough of LLM led to a great demand for prompt engineering, and the supply of this skill may slightly be lagged behind the demand to let talents familiar with it. Capturing the interconnection between

demand and supply variations perhaps can enhance one another predictive tasks. Second, the demand-supply variation of different skills is not standalone. Similar skills (*e.g.*, *C++* and *Java*) may share an evolution trend conditioned on the macro-level technology advancement and economic status. The challenge is identifying and preserving the synchronized evolution patterns of different skills to improve overall prediction accuracy. Third, despite the joint modeling paradigm sharing knowledge between skill demand and supply variations, the demand and supply predictions are still made separately. How to collectively output skill demand-supply predictions to further calibrate supply and demand predictions is another challenge.

In this paper, we propose a **C**ross-View **H**ierarchical **G**raph Learning **H**ypernetwork (CHGH) framework to predict skill supply and demand variations jointly. Specifically, the CHGH framework consists of a cross-view graph encoder module that enhances skill demand and supply representations by capturing their asymmetric relationship through an inter-view adaptive matrix. Besides, we devise a hierarchical graph encoder that models the high-level skill co-evolve trends by harnessing the cluster-wise skill variation correlation. Furthermore, we propose a hyper-decoder that outputs skill variations conditioned on the historical skill demand-supply gaps. By incorporating the paired relationship between the supply and demand of each skill, the hyper-decoder derives more accurate predictions to better support downstream analysis. The contributions of the paper are summarized as follows:

- To our knowledge, this is the first work that investigates the skill demand-supply joint prediction problem. By anticipating future skill demand and supply variations, employees and employers can effectively identify potential skill gaps and address them accordingly.

- We propose a cross-view hierarchical graph learning with hyper-decoder framework that simultaneously captures inter-dependencies between the skill supply and demand views, as well as the asymmetric cluster-wise correlation between skills.

- We evaluate the effectiveness of the proposed approach on three real-world datasets, and the experimental demonstrates the superiority of CHGH compared with seven baselines.

Preliminaries

In this section, we first introduce the real-world datasets used in this paper. Then, we formally define key concepts used in our task. Finally, we present the problem formulation of skill demand-supply joint prediction.

Data Description

The real-world dataset used in this work is extracted from an online recruitment platform, consisting of 2,254,733 job postings and 3,545,908 work experience descriptions. We divide the datasets into three main categories following the Global Industry Classification Standard (GICS) (Phillips and Ormsby 2016), including Information Technology (IT),

Datasets	IT	FIN	CONS
# of job description	902,442	645,899	692,163
# of work experience	363,376	223,330	455,120

Table 1: Statistics of datasets.

Finance (FIN), Consumer Discretionary and Consumer Staples (CONS), over the period from September 2017 to March 2019. The details of the dataset are illustrated in Table 1. In identifying job skills, we combined a list from Emsi Burning Glass (de Macedo et al. 2022) with our own additional skills, ensuring a comprehensive mix that addresses current industry trends. We preserve 7,446 skills that exist in the dataset by removing the rest that are too sparse in the job description. The occurrence relationship of skills in the job description is reported in Figure 2a.

Skill Supply and Demand Quantification

To quantify skill supply and demand shift, we first denote the skill set as $\mathcal{K}$ and discretize the investigated period into equal time steps $\mathcal{T}$. The job description jd and work experience we sets located at time step $t \in \mathcal{T}$ are denoted as $\mathcal{J}_t$ and $\mathcal{W}_t$, respectively. Then, we define the *Demand share* and *Supply share* to measure the relative importance of skills skill demand and supply, which indicates the proportion of a particular skill that appeared in the job description and work experience (de Macedo et al. 2022).

Definition 1 *Demand share. Demand share $\mathcal{D}_k^t$ is defined as the percentage of job descriptions that required the skill $k \in \mathcal{K}$ at the given time step $t \in \mathcal{T}$. The indicator function $\mathbb{1}_{k \in jd}$ is used to represent whether the skill k appears in the job description. $|\mathcal{J}_t|$ represents the total number of job descriptions at time step t.*

$$\mathcal{D}_k^t = \frac{\Sigma_{jd \in \mathcal{J}_t} \mathbb{1}_{k \in jd}}{|\mathcal{J}_t|}. \tag{1}$$

Definition 2 *Supply share. Supply share $\mathcal{S}_k^t$ is defined as the percentage of work experience that contained the skill $k \in \mathcal{K}$ at the time step $t \in \mathcal{T}$. $|\mathcal{W}_t|$ represents the total number of work experience at time step t.*

$$\mathcal{S}_k^t = \frac{\Sigma_{we \in \mathcal{W}_t} \mathbb{1}_{k \in we}}{|\mathcal{W}_t|}. \tag{2}$$

In this work, we use one month as the basic time step to derive demand share and supply share. Figure 2b depicts the distribution of skill demand-supply relationships. We can observe a strong positive correlation between skill demand and supply, which validate the necessity of joint demand-supply prediction. We further quantify the demand share and supply share difference.

Definition 3 *Skill gap. The skill gap SG_k^t for skill k at the time step t is defined as $SG_k^t = \mathcal{D}_k^t - \mathcal{S}_k^t$.*

The skill gap represents the demand and supply imbalance in the labor market, indicating the scarcity or abundance of skill k. Then we quantify the correlation between skills based on the co-occurrence information.
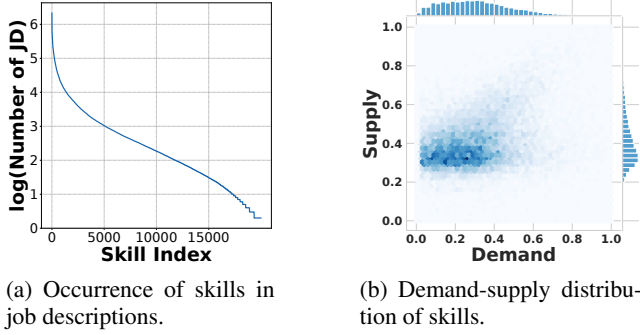
(a) Occurrence of skills in job descriptions.

(b) Demand-supply distribution of skills.

Figure 2: Skill distribution on IT dataset.

Definition 4 *Skill demand graph. Skill demand relation graph is defined as $\mathcal{G}_{\mathcal{D}} = (\mathcal{V}, \mathcal{E}_{\mathcal{D}}, \mathbf{A}^{\mathcal{D}})$. This graph represents the co-occurrence of skills in all job descriptions $\mathcal{J}$ of the training data. $\mathcal{V} = \mathcal{K}$ is the set of skills, and $\mathcal{E}_{\mathcal{D}}$ contains edges with weights that signify the normalized co-occurrence of skills in $\mathcal{J}$. $\mathbf{A}_{i,j}^{\mathcal{D}}$ is the corresponding weighted adjacency matrix*

$$\mathbf{A}_{i,j}^{\mathcal{D}} = \begin{cases} R_{i,j,\mathcal{J}}, & if \quad R_{i,j,\mathcal{J}} > \epsilon \\ 0, & otherwise \end{cases}, \quad (3)$$

$$R_{i,j,\mathcal{J}} = \frac{\Sigma_{jd \in \mathcal{J}} \mathbb{1}_{k_i, k_j \in jd}}{\Sigma_{jd \in \mathcal{J}} \mathbb{1}_{k_i \in jd}}, \quad (4)$$

where $R_{i,j,T}$ is the normalized co-occurrence ratio of skill k_i and k_j, and ϵ is a occurrence threshold.

Definition 5 *Skill supply graph. Skill supply relation graph is defined as $\mathcal{G}_{\mathcal{S}} = (\mathcal{V}, \mathcal{E}_{\mathcal{S}}, \mathbf{A}^{\mathcal{S}})$. This graph represents the co-occurrence of skills in all work experience descriptions $\mathcal{W}$ of the training data. It is constructed in a similar way as $\mathcal{G}_{\mathcal{D}}$, by utilizing $R_{i,j}^{\mathcal{W}}$ as normalized co-occurrence ratio in Eq. (4). $\mathbf{A}_{i,j}^{\mathcal{S}}$ is the weighted adjacency matrix*

$$\mathbf{A}_{i,j}^{\mathcal{S}} = \begin{cases} R_{i,j,\mathcal{W}}, & if \quad R_{i,j,\mathcal{W}} > \epsilon \\ 0, & otherwise \end{cases}. \quad (5)$$

Problem Statement

Given the demand sequence of all skills $\mathcal{D}_{\mathcal{K}}^{\mathcal{T}} = \{\mathcal{D}_{\mathcal{K}}^0, \mathcal{D}_{\mathcal{K}}^1, ... \mathcal{D}_{\mathcal{K}}^t\}$, where $\mathcal{D}_{\mathcal{K}}^t \in \mathbb{R}^{|\mathcal{K}| \times 1}$ denotes the demand share at time step t, the supply sequence $\mathcal{S}_{\mathcal{K}}^{\mathcal{T}}$, and skill gap historical sequence in time series $SG_{\mathcal{K}}^{\mathcal{T}} = \{SG_{\mathcal{K}}^0, SG_{\mathcal{K}}^1, ... SG_{\mathcal{K}}^t\}$ with elements $SG_{\mathcal{K}}^t \in \mathbb{R}^{|\mathcal{K}| \times 1}$. Our task is to utilize the skill co-occurrence graphs from the supply ($\mathcal{G}_{\mathcal{S}}$) and demand ($\mathcal{G}_{\mathcal{D}}$) views to jointly predict the demand and supply share of each skill k in the next time step,

$$\hat{Y}_{\mathcal{D}}^{t+1}, \hat{Y}_{\mathcal{S}}^{t+1} \leftarrow \mathcal{F}(\mathcal{S}_{\mathcal{K}}^{\mathcal{T}}, \mathcal{D}_{\mathcal{K}}^{\mathcal{T}}, SG_{\mathcal{K}}^{\mathcal{T}}, \mathcal{G}_{\mathcal{S}}, \mathcal{G}_{\mathcal{D}}), \quad (6)$$

where $\hat{Y}_{\mathcal{D}}^{t+1}, \hat{Y}_{\mathcal{S}}^{t+1}$ are the estimated supply and demand share of all skills at time step $t+1$.

Methodology

In this section, we present the CHGH framework in detail. As depicted in Figure 3, CHGH follows the encoder-decoder architecture comprising three major modules.

Cross-view Graph Encoder (CGE)

Existing studies on skill variation forecasting solely consider the dependencies in either demand or supply view. In fact, the demand and supply variation of skills are correlated yet diversified. To share the knowledge between demand and supply graph views (Khan and Blumenstock 2019; Liang et al. 2020), we propose the cross-view graph encoder, which consists of three components: skill temporal encoding, cross-view relation learning, and cross-view augmentation. Specifically, a skill temporal encoding block initially captures the temporal dependency within skill supply and demand sequences. After that, a cross-view relation learning block is invoked to identify the intricate asymmetric skill relationships between the supply and demand views. Finally, we employ a cross-view augmentation block to propagate knowledge from one view to another through the learned cross-view asymmetric skill relationships.

Skill temporal encoding. We first convert each skill $k \in \mathcal{K}$ to a d-dimensional vector $\mathbf{e}_k$ via an embedding matrix $\mathbf{E} \in \mathbb{R}^{|\mathcal{K}| \times d}$, where d is the dimension of each skill embedding. Take the supply sequence $\mathcal{S}_k^{\mathcal{T}} \in \mathbb{R}^{|\mathcal{T}| \times 1}$ of skill k for illustration, we further convert it to a matrix $\mathbf{S}_k^{\mathcal{T}} \in \mathbb{R}^{|\mathcal{T}| \times d}$ via a two-layer MLP. Then, we use two Long Short Term Memory (LSTM) layers to capture the temporal shifts,

$$\mathbf{h}_k^t = \text{LSTM}(\mathbf{S}_k^t, \mathbf{h}_k^{t-1}), \quad (7)$$

where $\mathbf{S}_k^t$ is the supply embedding derived in time step t and $\mathbf{h}_k^0$ is initialized by the skill embedding $\mathbf{e}_k$. After that, we adopt the attention mechanism $\text{Attn}(\cdot)$ (Wang et al. 2016) that adaptively reweighting the essential temporal steps for different skills to obtain the supply sequence-level representation,

$$\mathbf{H}_k^s = [\mathbf{h}_k^1; ...; \mathbf{h}_k^{\mathcal{T}}],$$
$$\text{Attn}(\mathbf{H}_k^s, \mathbf{e}_k) = \text{Softmax}\left(\mathbf{H}_k^s \mathbf{e}_k^\top / \sqrt{d}\right), \quad (8)$$
$$\mathbf{s}_k = \text{Attn}(\mathbf{H}_k^s, \mathbf{e}_k)^\top \mathbf{H}_k^s,$$

where $[;]$ denotes the row-wise concatenation. Finally, we use an MLP layer to fuse the skill embedding and supply sequence representation to obtain the supply representation of the skill,

$$\hat{\mathbf{s}}_k = \text{MLP}([\mathbf{e}_k || \mathbf{s}_k] \mathbf{W}_{\mathcal{S}}), \quad (9)$$

where $\mathbf{W}_{\mathcal{S}} \in \mathbb{R}^{2d \times d}$ is a learnable parameter, $||$ denotes the column-wise concatenation. Given the demand matrix $\mathbf{D}_k^{\mathcal{T}}$, we obtain skill demand representation $\hat{\mathbf{d}}_k$ in the same way.

Cross-view relation learning. The cross-view relation learning block derives cross-view skill relationships between the supply and demand views by calculating the similarity score given the skill supply and demand representations. The adaptive matrix $\mathbf{A}_p \in \mathbb{R}^{2|\mathcal{K}| \times 2|\mathcal{K}|}$ is defined as

$$\mathbf{E}_{\mathcal{S}} = [\hat{\mathbf{s}}_1; \cdots; \hat{\mathbf{s}}_{|\mathcal{K}|}], \quad \mathbf{E}_{\mathcal{D}} = [\hat{\mathbf{d}}_1; \cdots; \hat{\mathbf{d}}_{|\mathcal{K}|}],$$
$$\mathbf{X}_A = \tanh(\alpha[\mathbf{E}_{\mathcal{S}}; \mathbf{E}_{\mathcal{D}}]), \quad \mathbf{X}_B = \tanh(\beta[\mathbf{E}_{\mathcal{S}}; \mathbf{E}_{\mathcal{D}}]),$$
$$\mathbf{A}_p = \sigma(\text{Softmax}(\sigma(\mathbf{X}_A^\top \mathbf{X}_B - \mathbf{X}_B^\top \mathbf{X}_A)) - \delta),$$
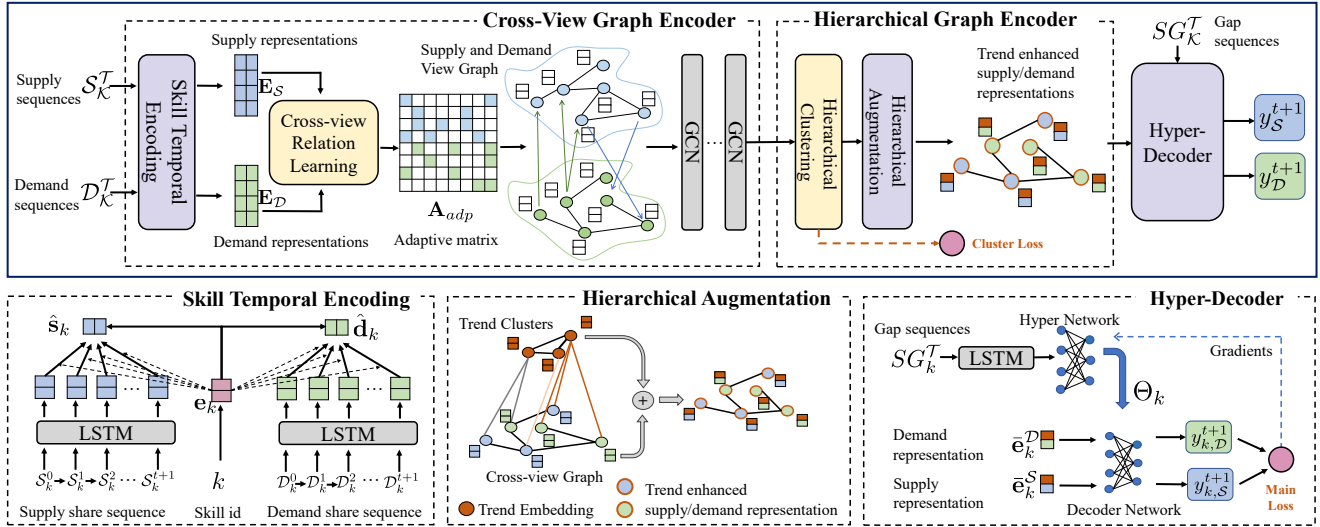$$\quad (10)$$

Figure 3: The CHGH framework mainly consists of three modules. (1) The cross-view graph encoder captures the asymmetric correlation for each skill between the supply and demand views. (2) The hierarchical graph encoder is responsible for preserving high-level co-evolution patterns between similar skills. (3) The conditional hyper-decoder jointly optimizes the supply and demand decoders based on historical skill gap tendencies.

where $\mathbf{E}_{\mathcal{S}}$ and $\mathbf{E}_{\mathcal{D}}$ are the embedding matrices derived from the skill temporal encoding block, $\alpha, \beta \in \mathbb{R}$ denote learnable scalars, σ stands for ReLU function, and δ is a hyperparameter to regulate the saturation of the adjacency matrix.

Cross-view augmentation. Leveraging the adaptive matrix $\mathbf{A}_p$, which encapsulates the cross-view asymmetric relationships, we further augment the cross-view skill demand and supply representations along with the intra-view skill relationships (*i.e.*, matrices $\mathbf{A}^{\mathcal{D}}$ and $\mathbf{A}^{\mathcal{S}}$). We devise a two-layer graph convolution operation (Kipf and Welling 2016) to derive skill supply and demand representations. The cross-view representation is enhanced in line with the approach proposed by (Wu et al. 2019b),

$$\mathbf{A}_{in} = \begin{bmatrix} \mathbf{A}^{\mathcal{S}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{\mathcal{D}} \end{bmatrix},$$

$$[\tilde{\mathbf{E}}_{\mathcal{S}}; \tilde{\mathbf{E}}_{\mathcal{D}}] = \mathbf{A}_p[\mathbf{E}_{\mathcal{S}}; \mathbf{E}_{\mathcal{D}}]\mathbf{W}_p + \mathbf{A}_{in}[\mathbf{E}_{\mathcal{S}}; \mathbf{E}_{\mathcal{D}}]\mathbf{W}_{in}, \quad (11)$$

where $[\tilde{\mathbf{E}}_{\mathcal{S}}; \tilde{\mathbf{E}}_{\mathcal{D}}]$ denotes the learned embedding from cross-view graph encoder. $\mathbf{W}_p$, $\mathbf{W}_{in} \in \mathbb{R}^{d \times d}$ are learnable parameters. We denote $\tilde{\mathbf{E}} = [\tilde{\mathbf{E}}_{\mathcal{S}}; \tilde{\mathbf{E}}_{\mathcal{D}}]$ in the following sections for the sake of simplicity.

Hierarchical Graph Encoder (HGE)

As aforementioned, the demand and supply of different skills may evolve synchronously. For example, the demand for *C++* and *Java* may boom under the policy of digital transformation. On the other hand, other skills, such as *accounting*, may not share such a co-evolve pattern under the policy guidance. In this part, we introduce the hierarchical encoder to capture the high-level skill co-evolve patterns to further enhance the skill relationship modeling.

Hierarchical clustering. To capture the cluster-wise trend among skills, we adopt Diffpool (Ying et al. 2018) to learn the cluster assignment in an end-to-end way. Specifically, we define an soft assignment matrix, $\mathbf{S} \in \mathbb{R}^{2|\mathcal{K}| \times c}$, where c is a hyper-parameter determining the number of trend clusters. The assignment matrix is derived by

$$\mathbf{S} = \text{Softmax}(\tilde{\mathbf{E}}\mathbf{W}_{\mathbf{S}}^{\top}), \quad (12)$$

where $\mathbf{S}$ describes the mapping relationship between bottom-layer skills and top-layer trend clusters, $\mathbf{W}_{\mathbf{S}} \in \mathbb{R}^{c \times d}$ is denoted as a mapping function that maps the node to a specific cluster. Then, we aggregate the connected skill representations to derive the cluster representations,

$$\mathbf{X}_h = \mathbf{S}^{\top}\tilde{\mathbf{E}}. \quad (13)$$

Hierarchical augmentation. Then, we integrate the learned cluster representations back into the skill representations to enhance the skill representations.

$$\hat{\mathbf{E}} = \text{Softmax}\left(\tilde{\mathbf{E}}\mathbf{X}_h^{\top}/\sqrt{d}\right)\mathbf{X}_h, \quad (14)$$

where $\hat{\mathbf{E}} \in \mathbb{R}^{2|\mathcal{K}| \times d}$ denotes the learned skill representations enhanced by the hierarchical encoder, which including both $\hat{\mathbf{E}}_{\mathcal{S}}$ and $\hat{\mathbf{E}}_{\mathcal{D}}$.

As clustering skills into appropriate trend clusters is a non-convex optimization problem, we introduce a clustering loss function (Ying et al. 2018),

$$\mathcal{L}_{\text{cluster}} = \frac{1}{2|\mathcal{K}|} \sum_{i=1}^{2|\mathcal{K}|} \text{H}(\mathbf{S}_i),$$

$$\text{H}(\mathbf{S}_i) = -\sum_{j=0}^{c} \mathbf{S}_{i,j} \ln \mathbf{S}_{i,j}, \quad (15)$$

where H denotes the entropy functions and $\mathbf{S}_{i,j}$ denotes the probability that skill supply and demand trend i belongs to the j-th trend cluster. The loss function forces clear distinctions between intra-cluster and inter-cluster skill variations.

Hyper-Decoder

Despite the cross-view encoder and the hierarchical graph encoder preserving cross-view multi-level skill correlations, separately output skill demand and supply predictions may derive inaccurate skill gaps and lead to biased conclusions. We further elaborate on this problem with the following motivation example.

Example 1 *Consider a machine learning engineer deciding on acquiring "Prompting" skill in the forthcoming months. If the predictive model underestimates the number of professionals who already have this skill, but correctly anticipates its rising demand, the engineer might face a saturated job market. Despite their investment in learning, they could struggle to find a competitive advantage for themselves to secure desired positions.*

In this work, we propose to incorporate historical skill gaps as auxiliary information to further calibrate supply and demand predictions. Specifically, inspired by the success of hypernetwork in generating generalizable parameter weights based on specific conditions (Pilault, Elhattami, and Pal 2021; Han et al. 2021), we propose a conditional hyper-decoder to force the framework collectively make predictions by considering the demand-supply gap. By encoding the prior skill gap sequence $SG_k^{\mathcal{T}}$ as condition $\mathbf{c}_k^{t+1}$, we generate the weight Θ_k for the hyper-decoder as follows,

$$
\begin{aligned}
\mathbf{c}_k^{t+1} &= \text{LSTM}(SG_k^t, \mathbf{c}_k^t), \\
\Theta_k &= \text{P}(\mathbf{c}_k^{t+1}, \Phi),
\end{aligned} \quad (16)
$$

where the LSTM layer transforms the skill gap sequence into conditions for supply and demand decoders, and $\mathbf{c}_k^0$ is initialized by skill embedding, $\mathbf{e}_k$. $\text{P}(\cdot)$ is the hyper-network that generates the weight for the MLP decoder given the condition $\mathbf{c}_k^{t+1}$, and the gradients are computed with respect to the weight of hypernet Φ. Θ_k is the generated decoder parameter for the skill k.

It's worth noting that updating the conditional hyper-decoder necessitates a carefully adjusted learning rate. An inappropriate learning rate may lead to slow convergence or exploding gradient issues. Thus, we normalize the condition $\mathbf{c}_k^{t+1}$ and introduce another three-layer MLP decoder as a supportive function to stabilize the training process,

$$
\begin{aligned}
\hat{y}_k^{\mathcal{S}} &= \text{Softmax}(\text{MLP}(\bar{\mathbf{e}}_k^{\mathcal{S}}, \Theta_k) + \text{MLP}(\bar{\mathbf{e}}_k^{\mathcal{S}})), \\
\hat{y}_k^{\mathcal{D}} &= \text{Softmax}(\text{MLP}(\bar{\mathbf{e}}_k^{\mathcal{D}}, \Theta_k) + \text{MLP}(\bar{\mathbf{e}}_k^{\mathcal{D}})),
\end{aligned} \quad (17)
$$

where $\hat{y}_k^{\mathcal{S}}, \hat{y}_k^{\mathcal{D}}$ represent our prediction goals for the next timestamp, and Θ_k is the replaced weight of MLP. $\bar{\mathbf{e}}_k^{\mathcal{S}}$ and $\bar{\mathbf{e}}_k^{\mathcal{D}}$ are the aggregated representations learned from cross-view graph encoder and hierarchical graph encoder, *i.e.*, $\bar{\mathbf{e}}_k^{\mathcal{S}} = \tilde{\mathbf{e}}_k^{\mathcal{S}} + \hat{\mathbf{e}}_k^{\mathcal{S}}, \bar{\mathbf{e}}_k^{\mathcal{D}} = \tilde{\mathbf{e}}_k^{\mathcal{D}} + \hat{\mathbf{e}}_k^{\mathcal{D}}$, where $\tilde{\mathbf{e}}_k^{\mathcal{S}} \in \tilde{\mathbf{E}}_{\mathcal{S}}, \tilde{\mathbf{e}}_k^{\mathcal{D}} \in \tilde{\mathbf{E}}_{\mathcal{D}}$, $\hat{\mathbf{e}}_k^{\mathcal{S}} \in \hat{\mathbf{E}}_{\mathcal{S}}, \hat{\mathbf{e}}_k^{\mathcal{D}} \in \hat{\mathbf{E}}_{\mathcal{D}}$.

Optimization

Following (Guo et al. 2022), we discretize the skill demand and supply into five trend categories, *i.e.*, *high, medium-high, medium, medium-low, and low*. Please refer to *Appendix* for a detailed processing procedure. The predictive loss is defined as

$$
\mathcal{L}_{\text{main}} = -\frac{1}{2|\mathcal{K}|} \sum_{k=1}^{|\mathcal{K}|} \sum_{j=1}^{m} y_{k,j}^{\mathcal{S}} \log \hat{y}_{k,j}^{\mathcal{S}} + y_{k,j}^{\mathcal{D}} \log \hat{y}_{k,j}^{\mathcal{D}}, \quad (18)
$$

where $y_{k,j}^{\mathcal{D}}$ and $y_{k,j}^{\mathcal{S}}$ are the ground truth labels of the skill k and transformed from the ground truth supply and demand shares in time period $t+1$. Take the demand label for example, $y_k^{\mathcal{D}}$ is a one-hot vector, and $y_{k,j}^{\mathcal{D}} = 1$ if the demand share $\mathcal{D}_k^{t+1}$ locates at the j-th trend class, otherwise $y_{k,j}^{\mathcal{D}} = 0$. m denotes the number of trend classes.

The overall training objective of CHGH is to minimize

$$
\mathcal{L} = \mathcal{L}_{\text{main}} + \lambda_1 \mathcal{L}_{\text{cluster}} + \lambda_2 ||\theta||_2^2, \quad (19)
$$

where $\mathcal{L}_{\text{main}}$ is predictive loss, $\mathcal{L}_{\text{cluster}}$ is the cluster assignment loss as defined in Eq. (15), and $||\theta||_2^2$ denotes the L2 regularization of learned parameters. λ_1 and λ_2 are the ratio of the clustering loss and the regularizer.

Experiments

In this section, we conduct extensive experiments to evaluate the effectiveness of our proposed framework. The source code of CHGH and all baselines are available online[1].

Experimental Setup

Data Processing and Evaluation metrics. For evaluation, we split each dataset described in Section into training, validation, and testing subsets with a ratio of 8:1:1. In our study, we formulate the demand and supply forecasting task as a classification problem. Our primary evaluation metrics include accuracy (ACC), weighted F1-score (F1), and the area under the receiver operating characteristic (AUC). We particularly underscore the importance of joint accuracy (J-ACC) as it represents the overall correctness of supply and demand predictions for a given skill. The definition of J-ACC is provided in *Appendix*. The hyper-parameters and implementation details are reported in *Appendix*.

Baselines. We compare CHGH with the following seven baselines and divide them into three classes:

- Statistical Time Series Model: **ARIMA** (Box and Pierce 1970) is a classical time series prediction model by taking differencing and moving the average into autoregression. **Vector Auto-Regressor** (Stock and Watson 2001) is a multivariate regressive model by capturing the relationship among variables.

- Recurrent-based Model: **Long-Short Term Memory (LSTM)** (Hochreiter and Schmidhuber 1997) introduces the forget gate and output gate to RNN to solve the vanishing gradient problem.

---

[1]https://github.com/vincent40416/Skill-Demand-Supply-Joint-Prediction

Datasets	IT				FIN				CONS			
Models	ACC	F1	AUC	J-ACC	ACC	F1	AUC	J-ACC	ACC	F1	AUC	J-ACC
ARIMA	0.3294	0.3314	0.5809	0.1161	0.3226	0.3256	0.5766	0.1072	0.3343	0.3375	0.5839	0.1140
VAR	0.3324	0.3264	0.5827	0.1170	0.3190	0.3216	0.5744	0.0997	0.3366	0.3424	0.5854	0.1049
LSTM	0.4864	0.4812	0.7878	0.2361	0.5058	0.5014	0.7960	0.2555	0.4810	0.4751	0.7852	0.2278
Transformer	0.4583	0.4544	0.7618	0.1893	0.4847	0.4818	0.7778	0.2340	0.3995	0.3927	0.7103	0.1675
Autoformer	0.3899	0.3700	0.7125	0.1540	0.3593	0.3314	0.6940	0.1312	0.3834	0.3614	0.7104	0.1476
Wavenet	<u>0.6170</u>	<u>0.6163</u>	<u>0.8867</u>	0.3816	0.5914	0.5919	0.8683	0.3567	<u>0.6702</u>	<u>0.6727</u>	<u>0.9157</u>	<u>0.4491</u>
MTGNN	0.6061	0.6040	0.8688	<u>0.3881</u>	<u>0.6621</u>	<u>0.7044</u>	<u>0.9209</u>	<u>0.4302</u>	0.6086	0.6085	0.8658	0.3522
Ours	**0.6777**	**0.6784**	**0.8968**	**0.4674**	**0.7227**	**0.7229**	**0.9234**	**0.5390**	**0.7704**	**0.7713**	**0.9398**	**0.6236**

Table 2: Overall performance comparison. The best results among all the models are highlighted in bold. The best baseline results are marked by underline.

Models	ACC	F1	AUC	J-ACC
Static Graph	0.5398	0.5378	0.8372	0.2944
+ Adaptive Graph	0.5724	0.5684	0.8638	0.3333
+ CGE	0.6179	0.6160	0.8814	0.3980
+ HGE	0.6596	0.6563	**0.9082**	0.4447
+ Hyper-Decoder	**0.6777**	**0.6784**	0.8968	**0.4674**

Table 3: Ablation study of each module on IT dataset.

- Transformer-based model: **Transformer** (Vaswani et al. 2017) adopts self-attention mechanism along with positional encoding. **AutoFormer** (Wu et al. 2021) introduces series decomposition blocks to capture series periodicity in long sequence prediction.
- Graph Neural Network based multivariate model: **Wavenet** (Wu et al. 2019b) uses graph convolutions and temporal 1D convolutions for spatio-temporal relation modeling. **MTGNN** (Wu et al. 2020) exploits unidirected dependencies among variables with mix-hop propagation to model the spatio-temporal dependencies.

Performance Comparison

Table 2 presents the overall performance comparison of various models across three different datasets: Information Technology (IT), Financial Industry (FIN), and Consumer Discretionary and Consumer Staples (CONS). The results lead us to the following observations. First, our proposed model outperforms all the baseline models across all datasets on all the evaluation metrics, demonstrating the effectiveness of our CHGH framework on skill supply demand joint prediction. Moreover, traditional time series forecasting models yield lower performance across all datasets and metrics, and the models based on neural networks demonstrate a significant improvement over traditional models. Furthermore, it is worth noting that the Recurrent-based method outperforms the Transformer-based methods in these datasets since the Transformer-based methods mainly focus on tackling long-horizon auto-correlation while overlooking the trend of variables. This outcome underscores the importance of the Recurrent-based model.

Besides, the results of Wavenet and MTGNN outperform statistical- and recurrent-based models. This is attributed to

their ability to propagate information through adaptive adjacency matrices grounded in skill embedding. Our proposed model further achieves noticeable improvement by preserving the demand-supply relationship and cluster-wise skill relationships, which were overlooked in previous studies.

Ablation Study

To delve deeper into the contributions of each component of our framework, we conducted an ablation study by gradually introducing designed modules into a basic model until forming the complete CHGH model. The variants of the model are as follows: **Static Graph** is the basic model which propagates the embedding learned from Skill Encoder based on fixed co-occurrence matrices and uses the plain MLP decoder. **Adaptive Graph** introduces the adaptive graph learner to generate an adjacency matrix on demand and supply views separately. **Cross-view Graph Encoder (CGE)** introduces asymmetric relations between demand and supply views to enhance skill representations (refer to Eq. (11)). **Hierarchical Graph Encoder (HGE)** further involves cluster-wise skill trend information (refer to Eq. (14)). **Hyper-Decoder** introduces the Hyper-decoder conditioned on historical skill gap sequences to enhance the joint prediction (refer to Eq. (17)). This represents the complete framework of CHGH.

As evident from Table 3, while introducing static and adaptive graphs offers performance improvements over naive sequential models, it is the cross-view learning framework that significantly boosts prediction results. The cross-view learning framework, which interchanges the relations between the demand and supply views, enhances the accuracy of the adaptive graph. The hierarchical graph encoder, designed to capture high-level trend information, further augments the accuracy. Lastly, introducing the conditional hyper-decoder showcases a marked improvement, especially in joint prediction accuracy, which indicates the effectiveness of joint prediction conditioned on historical skill gaps. The ablation study on other datasets is in *Appendix*.

Parameter Sensitivity

We present a sensitivity analysis of CHGH. We report ACC, F1, and AUC on the IT dataset.

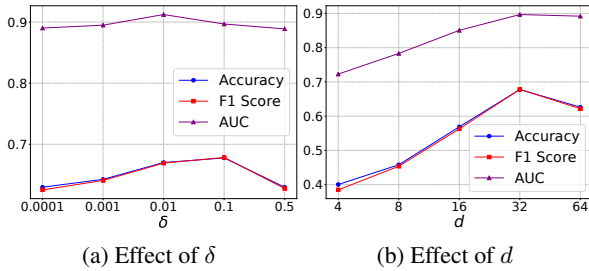First, we vary δ, which regulates the saturation of the edge

(a) Effect of δ (b) Effect of d

Figure 4: Parameter sensitivity on IT dataset.

Figure 5: Visualization of ground truth and predicted skill demand and supply variations.

generation, from 0.0001 to 0.5. A higher value implies that the generated edge becomes less. Figure 4a illustrates that when setting too large δ, the $\mathbf{A}_{adp}$ is likely to become so sparse that there exists no connection between views. Besides, when setting δ too small, the connection will reach to near fully connected graph, also degrading the performance.

Then, we vary d, the dimension of the skill embedding, from 4 to 64. As shown in Figure 4b, we observe that small dimension is insufficient to learn informative embedding, while high embedding dimension risks over-parameterization, causing the model to overfit to the dataset.

Case Study
As illustrated in Figure 5, there is a noticeable increase in both supply and demand in the *Natural Language Processing* (*NLP*) domain around the time of BERT model publication by Google in October 2018 (Devlin et al. 2018). Considering other concurrent factors that might have influenced the trend, many companies recognized the potential applications and advancements associated with BERT and other similar models, leading to increased interest in the field. Based on the prediction results of our model, there is an expected rise in both supply and demand for *Deep Learning* and *NLP* skills, which is consistent with observed real-world scenarios. Conversely, skills such as *HTML*, primarily associated with web development, do not appear to follow the same trend as *NLP* and *Deep Learning*, facing an oversupply after October 2018. Overall, the study illustrates how our model facilitates users in selecting suitable skills to learn.

Related Works
Skill supply and demand analysis. Traditional skill demand-supply analysis primarily involves statistical methods with expert knowledge to understand the pattern of transition of skills. Some researchers analyzed the skill demand shift caused by technology advancement (Bughin et al. 2018) and identified skill obsolescence (De Grip and Van Loo 2002) based on surveys. Due to the increasing usable data resources from online recruitment services, machine-learning-based models emerged as a powerful tool for substituting traditional statistic models. Xu et al. (2018) measures the popularity of skills with the topic model across various job criteria, such as company size and salary. de Macedo et al. (2022) adopted recurrent neural network in skill demand prediction. Despite the existence of research
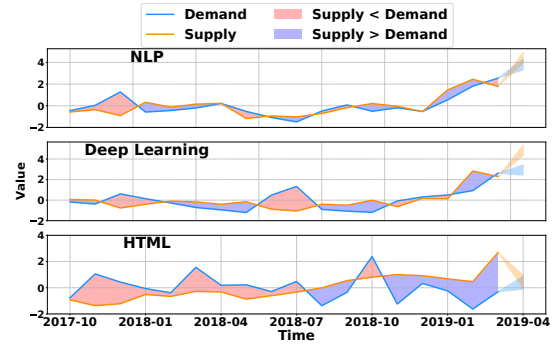
on skill-level supply (Belhaj and Tkiouat 2013) and demand prediction (de Macedo et al. 2022), they fail to address the connection between demand-supply in the predicting model. Besides, there are studies that delve into company-position-level analyses of supply and demand. These research efforts model the competitive nature of recruitment strategies employed by various companies (Wu et al. 2019a; Zhang et al. 2021; Guo et al. 2022; Qin et al. 2023).

Graph neural network. Graph neural networks received significant attention from academia and industry owing to their capability on learning non-Euclidean relationships (Shih, Sun, and Lee 2019; Wu et al. 2023; Zheng et al. 2021, 2023). The existing work could be divided into two categories, spectral-based methods (Defferrard, Bresson, and Vandergheynst 2016; Wu et al. 2019b) and spatial-based methods (Kipf and Welling 2016; Hamilton, Ying, and Leskovec 2017; Velickovic et al. 2017). The former adopted the message-filtering process and the latter follow the message-passing rule to aggregate information. In this work, we employ the GNN to facilitate information propagation across views and capture skill variation correlations.

Conclusion
In this work, we investigate the skill demand-supply joint prediction task, empowering employees and employers to anticipate future skill demand and supply variations and address potential skill gaps accordingly. To tackle the inherent challenges of this task, we proposed the CHGH framework. Specifically, a cross-view graph encoder is proposed to capture the inherent relationships between the skill supply and demand graph views. Furthermore, a hierarchical module was used to identify and preserve high-level skill trends that were caused by technological advancements or policy changes. We also constructed a conditioned hyper-decoder to calibrate supply and demand predictions by exploiting historical skill gaps as auxiliary signals. Extensive experiments demonstrated the superiority of CHGH against seven baselines, and the ablation study showed the effectiveness of each proposed module. By bridging the gap between employees' skills and industry requirements, this work facilitates a better alignment between individuals and the evolving labor market requirements.

Mathematical Notation

Symbol	Description
$\mathcal{K}$	Skill sets.
t	Timestamp.
$\mathcal{T}$	Time range $T = \{0, 1, ...t\}$.
$\mathcal{J}_t$	Job description at time t.
$\mathcal{W}_t$	Work experience at time t.
$\mathcal{D}_k^t$	Demand share of skill k at time t.
$\mathcal{S}_k^t$	Supply share of skill k at time t.
$\mathcal{SG}_k^t$	Skill gap of skill k at time t.
$\mathcal{G}_\mathcal{D}$	Skill demand relation graph.
$\mathbf{A}^\mathcal{D}$	Adjacency derived from $\mathcal{G}_\mathcal{D}$.
$\mathcal{G}_\mathcal{S}$	Skill supply relation graph.
$\mathbf{A}^\mathcal{S}$	Adjacency matrix derived from $\mathcal{G}_\mathcal{S}$.
$\hat{Y}_\mathcal{D}^{t+1}$	Estimated target of demand share
$\hat{Y}_\mathcal{S}^{t+1}$	Estimated target of supply share
$\mathbf{W}_*$	Model parameters
$\mathbf{E}$	Embedding derived from skill encoder.
$\mathbf{A}_p$	Learned adjacency matrix
$\mathbf{A}_{in}$	Combined intra-view adjacency matrix
$\bar{\mathbf{E}}$	Skill embedding derived from cross-view graph encoder.
c	Number of clusters.
$\mathbf{S}$	Assignment matrix.
$\mathbf{X}_h$	Trend embedding.
$\hat{\mathbf{E}}$	Skill embedding derived from hierarchical graph encoder.
$\mathbf{c}_k^{t+1}$	Conditions of hyper-decoder.
$\bar{\mathbf{e}}_k^\mathcal{D}$	Demand embedding as input of hyper-decoder.
$\bar{\mathbf{e}}_k^\mathcal{S}$	Supply embedding as input of hyper-decoder.
$y_k^\mathcal{D}$	Predicted demand share of skill k.
$y_k^\mathcal{S}$	Predicted supply share of skill k.

Discretizing Algorithm

The algorithm maps the demand share prediction goal $y_k^\mathcal{D}$ to discrete classes, the detail of the algorithm is formulated as:

$$std_k = STD(\mathcal{D}_k^\mathcal{T}),$$

$$\bar{\mathcal{D}}_k = \sum_{i=1}^{t} \mathcal{D}_k^i,$$

$$\hat{\mathcal{D}}_k^{t+1} = \frac{\mathcal{D}_k^{t+1} - \bar{\mathcal{D}}_k}{std_k}, \quad (20)$$

$$\hat{\mathcal{D}}_\mathcal{K}^{t+1} = \{\hat{\mathcal{D}}_0^{t+1}, \hat{\mathcal{D}}_1^{t+1}, \ldots, \hat{\mathcal{D}}_{|\mathcal{K}|}^{t+1}\},$$

where $STD(*)$ calculates the standard deviation of the sequence. After sorting $\hat{\mathcal{D}}_\mathcal{K}^{t+1}$ in ascending order, as $\tilde{\mathcal{D}}_\mathcal{K}^{t+1}$, we divide them into n classes with an equal number of elements in each class. The classes is represented as $C_1, C_2, \ldots, C_n$. which is formulated as:

$$\tilde{\mathcal{D}}_\mathcal{K}^{t+1} = \{\tilde{\mathcal{D}}_0^{t+1}, \tilde{\mathcal{D}}_1^{t+1}, \ldots, \tilde{\mathcal{D}}_{|\mathcal{K}|}^{t+1}\},$$

$$\tilde{\mathcal{D}}_0^{t+1} \leq \tilde{\mathcal{D}}_1^{t+1} \leq \cdots \leq \tilde{\mathcal{D}}_{|\mathcal{K}|}^{t+1}, \quad (21)$$

$$C_i = \{\tilde{\mathcal{D}}_{(i-1) \times \frac{n}{k}}^{t+1}, \tilde{\mathcal{D}}_{(i-1) \times \frac{n}{k}+1}^{t+1}, \ldots, \tilde{\mathcal{D}}_{i \times \frac{n}{k}-1}^{t+1}\},$$

where n is set to five trends in our experiment, representing *high, medium-high, medium, medium-low, and low*. The ground truth $y_{k,i}^\mathcal{D} = 1$ if $\hat{\mathcal{D}}_k^{t+1} \in C_i$, otherwise $y_{k,i}^\mathcal{D} = 0$.

Models	ACC	F1	AUC	J-ACC
Static Graph	0.4546	0.4348	0.7771	0.2230
+ Adaptive Graph	0.5826	0.5784	0.8627	0.3615
+ CGE	0.6839	0.6831	0.9135	0.4833
+ HGE	0.7218	0.7220	0.9170	0.5158
+ Hyper-Decoder	**0.7227**	**0.7229**	**0.9234**	**0.5390**

Table 4: Ablation study of each module on FIN dataset.

Models	ACC	F1	AUC	J-ACC
Static Graph	0.5183	0.5135	0.8177	0.2773
+ Adaptive Graph	0.5660	0.5628	0.8509	0.3242
+ CGE	0.7167	0.7172	0.9237	0.5414
+ HGE	0.7440	0.7441	0.9328	0.5800
+ Hyper-Decoder	**0.7704**	**0.7713**	**0.9398**	**0.6236**

Table 5: Ablation study of each module on CONS dataset.

$y_k^\mathcal{S}$ is discretized in a similar way based on the supply share sequence.

Joint Accuracy

To verify that both supply and demand are correctly predicted. We use $y_k^\mathcal{S}$ and $\hat{y}_k^\mathcal{S}$ as the true and predicted labels for the supply of the kth skill, $y_k^\mathcal{D}$ and $\hat{y}_k^\mathcal{D}$ be the true and predicted labels for the D goal of the kth instance. $|\mathcal{K}|$ is the total number of instances. Define the indicator function I as:

$$I(y_k^\mathcal{S}, \hat{y}_k^\mathcal{S}, y_k^\mathcal{D}, \hat{y}_k^\mathcal{D}) = \begin{cases} 1 & \text{if } y_k^\mathcal{S} = \hat{y}_k^\mathcal{S} \text{ and } y_k^\mathcal{D} = \hat{y}_k^\mathcal{D} \\ 0 & \text{otherwise} \end{cases}. \quad (22)$$

The joint accuracy (J-ACC) is then given by:

$$ACC = \frac{\sum_{k=1}^{|\mathcal{K}|} I(y_k^\mathcal{S}, \hat{y}_k^\mathcal{S}, y_k^\mathcal{D}, \hat{y}_k^\mathcal{D})}{|\mathcal{K}|}. \quad (23)$$

Implementation Details

For hyper-parameters, we choose the number of trend types as 5, the minimum length of sequences as 5, the embedding dim d as 32, the number of clusters c is 100, the head number of multi-head attention as 4, the number of LSTM layers as 3, and the output dimension of any other multi-layer perceptron as 32. We use Adam Optimization with a learning rate of 1e-3. The dropout is set to 0.3. The reducing rate of the learning rate scheduler is 0.1, step as 50. λ_1, λ_2 are set as 1e-5. We run each method 5 times and report the average results. The model is run on the machine with Intel(R) Xeon(R) Gold 5118 CPU @ 2.30GHz, NVIDIA GeForce RTX 3090 with 24G memory. The operating system is Centos Linux 7 (Core).

Ablation Study

The ablation study on the other dataset is shown in Table 4 and 5. The results show the same tendency as discussed in the study, indicating the effectiveness of the designed module across all the datasets.

Acknowledgements

References

Belhaj, R.; and Tkiouat, M. 2013. A Markov Model for Human Resources Supply Forecast Dividing the HR System into Subgroups. *Journal of Service Science and Management*, 6: 211–217.

Box, G. E.; and Pierce, D. A. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, 65(332): 1509–1526.

Bughin, J.; Hazan, E.; Lund, S.; Dahlström, P.; Wiesinger, A.; and Subramaniam, A. 2018. Skill shift: Automation and the future of the workforce. *McKinsey Global Institute*, 1: 3–84.

Communications, L. C. 2022. Our skills-first vision for the future. https://economicgraph.linkedin.com/blog/a-skills-first-blueprint-for-better-job-outcomes. Accessed: 2022-03-29.

De Grip, A.; and Van Loo, J. 2002. The economics of skills obsolescence: a review. *Research in Labor Economics*, 21: 1–26.

de Macedo, M. M. G.; Clarke, W.; Lucherini, E.; Baldwin, T.; Neto, D. Q.; de Paula, R. A.; and Das, S. 2022. Practical Skills Demand Forecasting via Representation Learning of Temporal Dynamics. In Conitzer, V.; Tasioulas, J.; Scheutz, M.; Calo, R.; Mara, M.; and Zimmermann, A., eds., *AIES '22: AAAI/ACM Conference on AI, Ethics, and Society, Oxford, United Kingdom, May 19 - 21, 2021*, 285–294. ACM.

Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In Lee, D. D.; Sugiyama, M.; von Luxburg, U.; Guyon, I.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, 3837–3845.

Devlin, J.; Chang, M.; Lee, K.; and Toutanova, K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR*, abs/1810.04805.

Donovan, S. A.; Stoll, A.; Bradley, D. H.; and Collins, B. 2022. Skills Gaps: A Review of Underlying Concepts and Evidence. CRS Report R47059, Version 3. *Congressional Research Service*.

Guo, Z.; Liu, H.; Zhang, L.; Zhang, Q.; Zhu, H.; and Xiong, H. 2022. Talent Demand-Supply Joint Prediction with Dynamic Heterogeneous Graph Enhanced Meta-Learning. In Zhang, A.; and Rangwala, H., eds., *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, 2957–2967. ACM.

Hamilton, W. L.; Ying, Z.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. In Guyon, I.; von Luxburg, U.; Bengio, S.; Wallach, H. M.; Fergus, R.; Vishwanathan, S. V. N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 1024–1034.

Han, J.; Liu, H.; Zhu, H.; Xiong, H.; and Dou, D. 2021. Joint Air Quality and Weather Prediction Based on Multi-Adversarial Spatiotemporal Networks. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, 4081–4089. AAAI Press.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.

Khan, M. R.; and Blumenstock, J. E. 2019. Multi-GCN: Graph Convolutional Networks for Multi-View Networks, with Applications to Global Poverty. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, 606–613. AAAI Press.

Kim, Y.; Hsu, J.; and Stern, M. 2006. An update on the IS/IT skills gap. *Journal of information systems education*, 17(4): 395.

Kipf, T. N.; and Welling, M. 2016. Semi-Supervised Classification with Graph Convolutional Networks. *CoRR*, abs/1609.02907.

Larsen, C.; Rand, S.; Schmid, A.; and Dean, A. 2018. *Developing Skills in a Changing World of Work: concepts, measurement and data applied in regional and local labour market monitoring across Europe*. Rainer Hampp Verlag Augsburg/München, Germany.

Liang, Y.; Huang, D.; Wang, C.; and Yu, P. S. 2020. Multi-view Graph Learning by Joint Modeling of Consistency and Inconsistency. *CoRR*, abs/2008.10208.

McGuinness, S.; and Ortiz, L. 2016. Skill gaps in the workplace: measurement, determinants and impacts. *Industrial relations journal*, 47(3): 253–278.

Phillips, R. L.; and Ormsby, R. 2016. Industry classification schemes: An analysis and review. *Journal of Business and Finance Librarianship*, 21(1): 1–25.

Pilault, J.; Elhattami, A.; and Pal, C. J. 2021. Conditionally Adaptive Multi-Task Learning: Improving Transfer Learning in NLP Using Fewer Parameters & Less Data. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Qin, C.; Zhang, L.; Zha, R.; Shen, D.; Zhang, Q.; Sun, Y.; Zhu, C.; Zhu, H.; and Xiong, H. 2023. A Comprehensive Survey of Artificial Intelligence Techniques for Talent Analytics. *CoRR*, abs/2307.03195.

Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I.; et al. 2018. Improving language understanding by generative pre-training.

Shih, S.; Sun, F.; and Lee, H. 2019. Temporal pattern attention for multivariate time series forecasting. *Machine Learning*, 108(8-9): 1421–1441.

Stock, J. H.; and Watson, M. W. 2001. Vector autoregressions. *Journal of Economic perspectives*, 15(4): 101–115.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. In Guyon, I.; von Luxburg, U.; Bengio, S.; Wallach, H. M.; Fergus, R.; Vishwanathan, S. V. N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, 5998–6008.

Velickovic, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; and Bengio, Y. 2017. Graph Attention Networks. *CoRR*, abs/1710.10903.

Wang, Y.; Huang, M.; Zhu, X.; and Zhao, L. 2016. Attention-based LSTM for Aspect-level Sentiment Classification. In Su, J.; Carreras, X.; and Duh, K., eds., *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, 606–615. The Association for Computational Linguistics.

Wu, H.; Xu, J.; Wang, J.; and Long, M. 2021. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y. N.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, 22419–22430.

Wu, L.; Zhao, H.; Li, Z.; Huang, Z.; Liu, Q.; and Chen, E. 2023. Learning the Explainable Semantic Relations via Unified Graph Topic-Disentangled Neural Networks. *ACM Transactions on Knowledge Discovery from Data*, 17(8): 1–23.

Wu, X.; Xu, T.; Zhu, H.; Zhang, L.; Chen, E.; and Xiong, H. 2019a. Trend-Aware Tensor Factorization for Job Skill Demand Analysis. In Kraus, S., ed., *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, 3891–3897. ijcai.org.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; Chang, X.; and Zhang, C. 2020. Connecting the Dots: Multivariate Time Series Forecasting with Graph Neural Networks. In Gupta, R.; Liu, Y.; Tang, J.; and Prakash, B. A., eds., *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, 753–763. ACM.

Wu, Z.; Pan, S.; Long, G.; Jiang, J.; and Zhang, C. 2019b. Graph WaveNet for Deep Spatial-Temporal Graph Modeling. In Kraus, S., ed., *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, 1907–1913. ijcai.org.

Xu, T.; Zhu, H.; Zhu, C.; Li, P.; and Xiong, H. 2018. Measuring the Popularity of Job Skills in Recruitment Market: A Multi-Criteria Approach. In McIlraith, S. A.; and Weinberger, K. Q., eds., *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, 2572–2579. AAAI Press.

Ying, Z.; You, J.; Morris, C.; Ren, X.; Hamilton, W. L.; and Leskovec, J. 2018. Hierarchical Graph Representation Learning with Differentiable Pooling. In Bengio, S.; Wallach, H. M.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, 4805–4815.

Zhang, Q.; Zhu, H.; Sun, Y.; Liu, H.; Zhuang, F.; and Xiong, H. 2021. Talent Demand Forecasting with Attentive Neural Sequential Model. In Zhu, F.; Ooi, B. C.; and Miao, C., eds., *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*, 3906–3916. ACM.

Zheng, Z.; Wang, C.; Xu, T.; Shen, D.; Qin, P.; Huai, B.; Liu, T.; and Chen, E. 2021. Drug Package Recommendation via Interaction-aware Graph Induction. In Leskovec, J.; Grobelnik, M.; Najork, M.; Tang, J.; and Zia, L., eds., *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, 1284–1295. ACM / IW3C2.

Zheng, Z.; Wang, C.; Xu, T.; Shen, D.; Qin, P.; Zhao, X.; Huai, B.; Wu, X.; and Chen, E. 2023. Interaction-aware Drug Package Recommendation via Policy Gradient. *ACM Trans. Inf. Syst.*, 41(1): 3:1–3:32.