SUMMER: Bias-aware Prediction of Graduate Employment Based on Educational Big Data

FENG XIA, Federation University Australia, Australia and Dalian University of Technology, China

TENG GUO, Dalian University of Technology, China

XIAOMEI BAI, Anshan Normal University, China

ADRIAN SHATTE, Federation University Australia, Australia

ZITAO LIU, TAL Education Group, China

JILIANG TANG, Michigan State University, USA

The failure of obtaining employment could lead to serious psychosocial outcomes such as depression and substance abuse, especially for college students who may be less cognitively and emotionally mature. In addition to academic performance, employers' unconscious biases are a potential obstacle to graduating students in becoming employed. Thus, it is necessary to understand the nature of such unconscious biases to assist students at an early stage with personalized intervention. In this paper, we analyze the existing bias in college graduate employment through a large-scale education dataset and develop a framework called SUMMER (bias-aware gradUate eMployMENT pRediction) to predict students' employment status and employment preference while considering biases. The framework consists of four major components. Firstly, we resolve the heterogeneity of student courses by embedding academic performance into a unified space. Next, we apply a **Wasserstein generative adversarial network with gradient penalty (WGAN-GP)** to overcome the label imbalance problem of employment data. Thirdly, we adopt a temporal convolutional network to comprehensively capture sequential information of academic performance across semesters. Finally, we design a bias-based regularization to smooth the job market biases. We conduct extensive experiments on a large-scale educational dataset and the results demonstrate the effectiveness of our prediction framework.

CCS Concepts: • **Applied computing** → **Education**; • **Computing methodologies** → **Machine learning**;

Additional Key Words and Phrases: Graduate employment, prediction, bias, educational big data, data analysis

ACM Reference format:

Feng Xia, Teng Guo, Xiaomei Bai, Adrian Shatte, Zitao Liu, and Jiliang Tang. 2022. SUMMER: Bias-aware Prediction of Graduate Employment Based on Educational Big Data. *ACM/IMS Trans. Data Sci.* 2, 4, Article 39 (March 2022), 24 pages.

<https://doi.org/10.1145/3510361>

Authors' addresses: F. Xia, School of Engineering, IT and Physical Sciences, Federation University Australia, Ballarat, 3353, Australia; School of Software, Dalian University of Technology, Dalian, 116620, China; email: f.xia@acm.org; T. Guo, School of Software, Dalian University of Technology, Dalian, 116620, China; email: teng.guo@outlook.com; X. Bai, Computing Center, Anshan Normal University, Anshan, 114007, China; email: xiaomeibai@outlook.com; A. Shatte, School of Engineering, IT and Physical Sciences, Federation University Australia, Ballarat, 3353, Australia; email: a.shatte@federation.edu.au; Z. Liu, TAL Education Group, Beijing, 100080, China; email: liuzitao@100tal.com; J. Tang, Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA; email: tangjili@msu.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

2577-3224/2022/03-ART39 \$15.00

<https://doi.org/10.1145/3510361>

1 INTRODUCTION

Education, as the basic means of improving individual abilities, makes students competitive in the job market. However, not every graduate can succeed in gaining employment. According to an analysis from the **National Center for Education Statistics (NCES)**, the employment rate of full-time undergraduate students and part-time undergraduates was 43 percent and 81 percent, respectively, in 2018 [17], with these rates having decreased since the year 2000 (53 percent and 85 percent). The data from the statistical office of the **European Union (EU)** shows that the employment rate in the EU in 2018 of 20–34 years old was 83.4% for tertiary education and 65.8% for upper secondary general education [39]. The current prevalence of the COVID-19 pandemic also contributes to the poor state of the labor market [2] and is further impacting the employment situation of young people (From Figure 1, it can be seen that the unemployment rate of the youth rises in the first quarter of 2020). The failure of obtaining employment can lead to serious negative psychosocial outcomes including depression [5, 14, 30], particularly for students who may have poorer coping mechanisms due to being cognitively and emotionally less mature [40, 45]. Therefore, the timely detection of students facing challenges in obtaining appropriate employment and providing those students with personalized intervention and guidance is of great importance. However, it can be challenging to detect which students are at risk of not obtaining employment because recruitment can be impacted by a range of factors [37, 54].

While every recruiter aims to hire the best employees, the process of hiring can be impacted by both objective and subjective factors. In addition to academic performance [29], recruitment decisions can be affected by unconscious biases, such as gender and institutional prestige [10, 12, 13, 51]. These biases not only lead to an imbalance in the hiring process, but also result in an inequality of employment that favours uniformity rather than diversity in the workplace [18, 21], particularly for fresh graduates who may have very limited or no work experience. Therefore, it is necessary to understand biases in recruitment, which can further be exploited for the prediction of graduates' employment. While previous research has focused on understanding such biases in employment, these studies have relied mainly on questionnaires and surveys, which are time- and cost-consuming to implement, and their limited scope may not be representative of all students [19].

Campus life is information-intensive [4, 23, 34], and modern intelligent education management systems allow for rich data about students to be derived from digitized records. This enables data-driven development [33, 43, 47, 48] and provides new opportunities to explore and understand the nature of graduate employment and its associated barriers. However, there are still several key challenges to overcome. First, such data is much more complex than that based on questionnaires, thus advanced analytic techniques are needed. Second, the number of graduates that cannot land a job is much smaller compared to those who can successfully obtain jobs, thus employment analysis and prediction are highly imbalanced. Third, there may exist biases in employment that can vary by study major and discipline, whereas the majority of existing algorithms seldom consider all possible biases.

In this paper, we aim to explore the employment biases present in different majors, using demographic and academic performance data to make two employment-related predictions:

- **Employment status:** predict students with trouble in obtaining employment at an early stage.
- **Employment preference:** predict whether a student's employment choice will be a government office or an enterprise at an early stage.

The experimental process of this research is shown in Figure 2. First, we analyze the employment biases for each major from four domains including gender, ethnic group, hometown, and enrolment status. Second, based on possible employment biases, we propose the **SUMMER**

Fig. 1. The youth (15–24 years old) unemployment rate of countries in **Group of Seven (G7)** in the last 10 years (data is from **Organization for Economic Co-operation and Development (OECD)**) [16]. The G7 is an international intergovernmental economic organization consisting of seven major developed countries: Canada (CAN), France (FRA), Germany (DEU), Italy (ITA), Japan (JPN), the United Kingdom (GBR) and the United States (USA), which are the largest IMF-advanced economies in the world.

(**bias-aware gradUate eMployment pRediction**) prediction framework, which is comprised of four key components. In the first component, we resolve the heterogeneity of students' courses through embedding academic performance into a space of unified dimension by autoencoder. Next, to overcome the label imbalance problem, a **Wasserstein generative adversarial network with gradient penalty (WGAN-GP)** is applied to generate data of the minority class. Further, considering the sequential information between semesters, a **temporal convolutional network (TCN)** is utilized. Finally, we design a regularization to smooth the weight caused employment biases of different majors.

This work is based on our previous work [24], which aimed to explore the bias of graduate employment and make an effective prediction for graduate employment. In previous work [24], we focused on predicting employment status only. In the current work, we make predictions for employment preference in addition to employment status so as to improve the understanding of and guidance for students into employment. For the prediction tasks, we devise the new prediction framework named SUMMER. Compared with the MAYA prediction framework proposed in our previous work [24], we further improve the prediction performance by introducing WGAN-GP and TCN technology in this work.

Our contributions can be summarized as follows:

- We provide a comprehensive analysis on employment bias from the perspectives of employment status and employment preference.
- We propose a bias-aware framework for graduate employment prediction based on students' academic performance and demographic characteristics.
- We conduct extensive experiments upon a large-scale educational dataset, with results demonstrating the effectiveness of the proposed framework.

The rest of this paper is organized as follows. In Section 2, related work is reviewed. The problem formulation is presented in Section 3. In Section 4, we analyze the employment biases by majors.

Fig. 2. The flowchart of this research.

In Section 5, the SUMMER prediction framework is introduced in detail. In Section 6, we analyze the results of our experiment. We present the discussion and conclusion of our work in Section 7.

2 RELATED WORK

2.1 Employment and Recruitment

The employment and recruitment process has attracted significant attention in recent decades. Researchers have studied patterns behind the recruitment process using a variety of computational and statistical methods, which are summarised in this section. For instance, Naim et al. [38] presented a computational framework for automatically quantifying verbal and nonverbal behaviors of college students in job interviews and demonstrates the effectiveness of their framework based on a real-world dataset. Qin et al. [41] proposed a personalized question recommender system (Skill-Graph) for enhancing job interview assessment based on a knowledge graph of job skills. Yan et al. [50] designed a preference-based matching network to profile the latent preference from the histories of all interviewed candidates and all job applications, with experimental results showing that the proposed model could improve the performance of job-CV matching. Shen et al. [42] developed a latent variable model to jointly model job description, candidate resume and interview assessment and demonstrate its performance using a large-scale, real-world interview dataset. Kong et al. [29] carried out a series of experiments to explore the relationship between students' academic performance and their graduate employment outcomes. Hora [26] explored employability based on cultural capital theory, with results demonstrating that cultural fit dominates the recruitment process and that employers prefer to match diverse applicant personalities and competencies to the personalities of existing employees.

In addition to conducting research at a theoretical level, some scholars have developed a range of tools related to student employment. For example, Liu et al. [35] developed a tool for career exploration based on the intuitiveness of node-link diagrams and the scalability of aggregation-based techniques to help students in understanding the process of employment. Uosaki et al. [44]

developed a career support system to help international students find a job in Japan using a learning log system and eBook reading. Liu et al. [36] designed a job recommendation service framework for university students, which utilises a student profiling based re-ranking rule to recommend a list of potential jobs. Currently, COVID-19 is the focus of worldwide attention and has had a huge impact on the labor market, including employment of college graduates [2].

In summary, previous research has explored graduate employment from a range of perspectives and techniques. A popular method of data collection in previous research has been questionnaires. While this approach offers greater flexibility in obtaining data, it can be time- and cost-consuming and difficult to scale questionnaires to large-scale student populations [19].

2.2 Temporal Convolutional Network

Temporal dynamic data is an important data type in real life. Recurrent neural networks, and their variants, are one of the most effective machine learning models for processing this type of data. In 2018, Bai et al. [3] proposed a sequence model, named **temporal convolutional network (TCN)**, which outperforms current popular models on many mainstream datasets. Some scholars have enhanced the original TCN model and proposed several TCN variant models. For instance, Farha et al. [15] developed a multi-stage architecture for the temporal action segmentation task based TCN model, with experimental results showing that this model achieves state-of-the-art performance on some challenging datasets. Lin et al. [32] proposed a hierarchical attention-based TCN based on the attention mechanism and achieved good performance in diagnosing myotonic dystrophy using the model. Gao et al. [20] designed an optimized fault diagnosis model for power converters, based on TCN, with experiments demonstrating that this approach is efficient and can be adaptively applied to different real-world applications. You et al. [52] proposed hierarchical TCNs for making accurate dynamic recommendations based on users' sequential multi-session interactions with items. The results of experiments demonstrate that this model outperforms state-of-the-art dynamic recommendation methods.

The TCN and its variants have been successfully applied to various fields, including finance, meteorology, architecture, traffic, and medicine. Deng et al. [11] proposed a knowledge-based prediction framework, named KDTCN, for stock trend prediction. They combine two technologies, knowledge graph and TCN, and improve the performance of stock prediction tasks. Yan et al. [49] used TCN to predict El Niño-Southern Oscillation and achieved outstanding performance. Pau et al. [46] applied a deep TCN model as a building energy surrogate model for processing annual multivariate weather data. Kok et al. [28] developed a sepsis detection tool for hospitals and this tool can achieve 98.8% on AUC and 98% on ROC. Zhang et al. [53] applied the TCN model on short-term prediction of traffic passenger demand, with experiments showing the proposed MTL-TCNN with the ST-DTW algorithm is a promising method for short-term passenger demand prediction in a multi-zone level.

3 PROBLEM STATEMENT

In this section, we will introduce some notations and then formally define the problem to be solved in this work. In a university, let $M = \{1, 2, \dots, m\}$ denote the set of majors (i.e., programs) and the set of students in every major is defined as $Q = \{\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_m\}$. In other words, $\mathcal{N}_m$ represents the set of students in major m . For student i in major m , we define two types of features: **academic feature** as $\mathbf{a}_i^m \in \mathbb{R}^n$, where n represents the embedding dimension, and **demographic feature** as $\mathbf{d}_i^m \in \mathbb{R}^p$, where p is equal to 4 because we use four demographic features, including gender, ethnic group, hometown, and enrolment status. Note that academic feature is obtained through a special representation method proposed in this paper, which will be introduced in Section 5. The **final employment status** and **final employment preference** are denoted as $y_i^m \in \{0, 1\}$, where

Fig. 3. The hometown geographical distribution of students. The depth of color represents the percentage of students.

0 and 1 respectively represent students who have not found a job and those who have found a job, and $o_i^m \in \{0, 1\}$, where 0 and 1 respectively represent whether the type of work belongs to government or enterprise. Let $\mathbf{D}^m = [\mathbf{d}_1^m, \mathbf{d}_2^m, \dots, \mathbf{d}_{|N_m|}^m] \in \mathbb{R}^{|N_m| \times p}$, $\mathbf{A}^m = [\mathbf{a}_1^m, \mathbf{a}_2^m, \dots, \mathbf{a}_{|N_m|}^m] \in \mathbb{R}^{|N_m| \times n}$, $\mathbf{y}^m = [y_1^m, y_2^m, \dots, y_{|N_m|}^m] \in \mathbb{R}^{|N_m|}$ and $\mathbf{o}^m = [o_1^m, o_2^m, \dots, o_{|N_m|}^m] \in \mathbb{R}^{|N_m|}$ represent the demographic feature matrix, the academic performance matrix, the employment status vector and the employment preference vector of major m . The details of features used in this research are described in the following section.

There are two prediction problems in our research:

Employment Status Prediction: given the feature vector $\mathbf{d}_i^m$ and the corresponding academic performance vector $\mathbf{a}_i^m$, we predict the final employment status y_i^m .

Employment Preference Prediction: given the feature vector $\mathbf{d}_i^m$ and the corresponding academic performance vector $\mathbf{a}_i^m$, we predict the final employment preference o_i^m .

4 BIAS ANALYSIS

In this section, we first introduce the dataset used in this research. In addition, we explore bias in employment status and employment preference from both college and major perspectives, respectively.

4.1 Dataset

The dataset used in this work involves 2,133 Chinese university students from 31 provinces of China. A detailed geographical distribution is shown in Figure 3. All students enrolled in 2013 and graduated in 2017. The students are from 64 majors across 12 colleges. The data is securely stored and managed by the university the students attended. We were granted access to a dataset which had been well preprocessed by means of data de-identification, anonymisation and pseudonymisation. This dataset consists of three types of information, which are described below.

4.1.1 Demographic Data. Students are required to submit personal information at the time of admission to college, including hometown, gender, and ethnic group. The demographic data includes 8,532 records.

Table 1. Employment Bias Analysis of Employment Status from College-view

College	Employment Status			
	Hometown	Gender	Enroll Status	Ethnic Group
College of Physical Sciences and Technology	0.2072	0.1201	0.0025**	0.9639
Literature College	0.9543	0.9595	0.9097	0.8876
Music College	0.7144	0.9123	0.0507	0.1594*
Higher Vocational and Technical College	0.7662	0.5452	0.1064	0.5648
Education College	0.0469*	0.0351*	0.1303	0.1348
College of Chemistry and Life Sciences	0.5078	0.7611	0.1778	0.4046
Business College	0.9648	0.5745	0.9304	0.1616
Art College	0.0947	0.0216*	0.9621	0.8973
College of Mathematical and Information Sciences	0.4229	0.1255	0.8849	0.0043**
College of Social Development	0.9658	0.6917	0.8470	0.7782
College of Foreign Languages	0.5724	0.2382	0.1495	0.9367
Sport College	0.2196	0.3299	0.7453	0.2460

(P-value of Chi-square Test; *Represents $p < 0.05$ and **Represents $p < 0.01$).

4.1.2 Academic Performance Data. Students' academic performance data contains scores and credits of courses. There are in total 195,234 academic records.

4.1.3 Employment Data. When finding a job, students need to sign tripartite agreements to guarantee their legal rights, thus universities own records of students' employment status information about related companies and government agencies. This dataset consists of 2,133 employment data records.

4.2 Bias in Employment

In this section, we define employment bias from two perspectives: employment status and employment preference. Employment status represents whether or not the student has found a job at graduation. In this case, the bias here means that recruiters have stereotypes of college students from certain majors due to certain traits that are not related to personal ability [27, 31]. For example, in some majors, male students are more likely to receive job opportunities than female students. Moreover, all employment units in our dataset can be divided into two types according to organization code: government office and enterprises. Here the bias refers to the preferences of various students in the type of work. For example, in some majors, male students tend to work in enterprises, while female students are more willing to work in government roles.

We analyze the bias in employment from two views: college-view and major-view. We check the bias from four aspects: **gender** [18], **ethnic group** (the minority or the majority) [1], **administrative level of hometown** (city or county) [25], and **enrolment status** (whether passing the college entrance examination at one time). The Chi-square test is used here to examine the impact of these features on 12 colleges and 64 majors involved in our dataset (note that Fisher's exact test is used for several majors with small sample sizes). The details of bias analysis are given below.

4.2.1 Employment Bias of College-view. All results of employment bias on college-view are shown in Tables 1 and 2. It can be seen that bias exists on both employment status and employment preference. For employment status, the colleges with bias are shown as follows:

- **Administrative Level of Hometown:** Education College.
- **Ethnic Group:** College of Mathematical and Information Sciences.

Table 2. Employment Bias Analysis of Employment Performance from College-view

College	Employment Preference			
	Hometown	Gender	Enroll Status	Ethnic Group
College of Physical Sciences and Technology	0.9686	0.8316	0.5967	0.7309
Literature College	0.9489	0.9434	0.6438	0.8931
Music College	0.9965	0.6988	0.8759	0.0438*
Higher Vocational and Technical College	0.1177	0.9120	0.3104	0.3274
Education College	0.2646	0.8289	0.4951	0.9946
College of Chemistry and Life Sciences	0.0137*	0.4199	0.0303*	0.8107
Business College	0.6881	0.0109*	0.0625	0.8319
Art College	0.8869	0.7146	0.8441	0.9197
College of Mathematical and Information Sciences	0.9706	0.5757	0.8974	0.8720
College of Social Development	0.8875	0.8985	0.225	0.2946
College of Foreign Languages	0.3879	0.8217	0.025*	0.5767
Sport College	0.2924	0.7632	0.9291	0.8848

(P-value of Chi-square Test; *Represents $p < 0.05$ and **Represents $p < 0.01$).

- **Enrolment Status:** College of Physical Sciences and Technology.
- **Gender:** Education College, Art College.

For employment preference, the colleges with bias are shown as follows:

- **Administrative Level of Hometown:** College of Chemistry and Life Sciences.
- **Ethnic Group:** Music College.
- **Enrolment Status:** College of Chemistry and Life Sciences, College of Foreign Languages.
- **Gender:** Business College.

4.2.2 Employment Bias of Major-view. Since there are too many majors to appear in one table, we chose to represent the results of analysis in a scatter plot of the P value. Bias analysis of major-view is shown in Figures 4 and 5, respectively. For employment status, the majors with bias are shown as follows:

- **Administrative Level of Hometown:** Physical Education.
- **Ethnic Group:** Information and Computing Science, Computer Science and Technology.
- **Enrolment Status:** Preschool education, English, Electronic Information Science and Technology, Food Science and Engineering.
- **Gender:** English, Applied Psychology, Electronic Information Science and Technology.

For employment preference, the majors with bias are shown as follows:

- **Administrative Level of Hometown:** Physical Education, Applied Chemistry.
- **Ethnic Group:** Musical Performance.
- **Enrolment Status:** Preschool Education, Japanese.
- **Gender:** Marketing, Tourism Management, Fine Arts.

These observations suggest that employment bias does exist in some majors and the existence of bias does affect graduates' employment. Note that since the purpose of this paper is to detect the presence of employment bias and design algorithms to consider its influence in employment prediction, rather than analyze its causes, the mechanisms of bias formation are not further analyzed here.

Fig. 4. The distribution of p -value with respect to employment status. Subfigures denote the results of Chi-square test in terms of hometown, ethnic group, enrolment status and gender, respectively. Each black dot represents the p -value of a certain major. When p -value is less than the threshold (i.e., 0.05 depicted as a red star), the hypothesis is acceptable, that is, bias exists.

5 DESIGN OF SUMMER

In this section, we provide a detailed description of the proposed framework, SUMMER, which is targeted at overcoming common challenges in the prediction of employment status and the prediction of employment preference. Figure 6 shows an illustration of the SUMMER framework. The framework has four components, including representation learning of academic performance, data augmentation for label imbalance, the prediction model, and bias-aware optimization. In the next section, we explain each component in detail.

5.1 Academic Performance Representation

When taking academic performance as a feature, the heterogeneity of curriculum is always a challenge due to the differences in students' selection of courses. A popular method is to calculate the summarizing statistics (e.g., the mean or **Grade-Point Average (GPA)**) as the agent to represent academic performance of students. This method can effectively overcome the heterogeneity of the curriculum, but the information loss may be significant when the data distribution varies widely. In this case, we propose a C matrix, and based on C , we use an auto-encoder to get the embedding representation to tackle the heterogeneity issue.

5.1.1 C Matrix. To solve the problem caused by the heterogeneity in the data, we embedded student exam scores through a variant of one-hot encoding, which replaces the 1 in one-hot encoding with the corresponding exam grade, and details are shown as follows. We create the matrix $C_s \in \mathbb{R}^{n_s \times m_s}$ where n_s and m_s represent the number of students and the number of courses for a

Fig. 5. The distribution of p -value with respect to employment preference. Each subfigure is the corresponding feature's result of Chi-square test. Each black dot represents the p -value of a certain major. When p -value is less than the threshold (i.e., 0.05 depicted as an orange star), the hypothesis is acceptable, that is, bias exists.

Fig. 6. The illustration of SUMMER framework.

certain major, respectively. In our dataset, $s = 1, 2, \dots, 6$ since students have valid grades across six semesters except for the two-semester graduation project and social practice. c_{ij} is the grade of student i on course j . If a student does not attend a particular course, the corresponding element remains 0. The size of this matrix is different for each semester as shown in the following matrix. For example, if there are 300 students and 500 courses in the first semester, then the size of C_1 matrix is 300×500 .

$$\begin{Bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{Bmatrix}$$

5.1.2 Representation Learning. Due to the diversity of courses from various majors and students' autonomy in choosing courses, C matrix is quite sparse. Thus, we use the autoencoder to get the embedding representation that is the academic performance matrix A . The hidden layers of the autoencoder are divided into two parts: the encoder part and the decoder part. The layers consistently encode and decode the input data. The input of the i th layer is considered as the output of $(i-1)$ th layer. The hidden layers can automatically capture the characteristics of input data and keep them unchanged. To capture the temporality among semesters, we use the autoencoder to embed the matrix of each semester, respectively. In each hidden layer, we adopt the following nonlinear transformation function:

$$\begin{aligned} \mathbf{h}_{(2)} &= f(\mathbf{W}_{(2)}\mathbf{h}_{(1)} + \mathbf{b}_{(2)}) \\ \mathbf{h}_{(3)} &= f(\mathbf{W}_{(3)}\mathbf{h}_{(2)} + \mathbf{b}_{(3)}) \\ \mathbf{h}_{(i)} &= f(\mathbf{W}_{(i)}\mathbf{h}_{(i-1)} + \mathbf{b}_{(i)}), i = 3, 4, \dots, k \end{aligned} \quad (1)$$

where f is the activation function and $\mathbf{W}_{(i)}$, $\mathbf{b}_{(i)}$ are the transformation matrix and the bias vector of the layer i . The corresponding Loss function is as follows (Equation (2)):

$$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \quad (2)$$

where $\mathbf{x}$ represents the input and $\hat{\mathbf{x}}$ represents the output (i.e., $\hat{\mathbf{x}} = f(\mathbf{W}\mathbf{x} + \mathbf{b})$, where $\mathbf{W}$ and $\mathbf{b}$ represent the transformation matrix and bias vector of the model, respectively). We use C as the input and minimize the reconstruction error between the output and the original input. Then, we take the output of the encoder as the academic performance matrix A' .

5.2 Data Augmentation for Label Imbalance

In general, the number of students who fail to secure employment is smaller. Moreover, for employment preference, there is also a huge difference between the number of people who choose commercial corporations and the number choosing government agencies (shown in Figure 7). Thus, the label imbalance problem exists in both of these prediction tasks.

The main idea here is that we use the **GAN (generative adversarial networks)**-related model to learn the distribution behind the data and then generate new data samples to balance the dataset. GAN is a generative model based on neural network structure, which is widely used in various fields [8, 55]. The GAN training strategy is a game between two competing networks: the generator and the discriminator. The generator maps a noise source to the input space. The discriminator receives a true data sample or a generated sample and is required to distinguish between the two. The loss function is shown as follows:

$$\min_G \max_D V(D, G) = \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim p_z(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (3)$$

Fig. 7. The proportion of students in different categories.

where G and D represent the generator and the discriminator, respectively. $\mathbf{x} \sim p_{\text{data}}(\mathbf{x})$ represents that $\mathbf{x}$ is taken from the real data distribution. $\mathbf{z} \sim p_z(\mathbf{z})$ represents that $\mathbf{z}$ is taken from the noise distribution.

However, in the traditional GAN model, the divergences are potentially not continuous with respect to the generator's parameters, leading to training difficulty. Scholars have proposed a series of GAN variants to improve its training efficiency and performance. Recently, Gulrajani et al. [22] proposed WGAN-GP and demonstrated its strong performance in various fields. Thus, we employ WGAN-GP to augment data in order to improve generalization performance. The objective of WGAN-GP is:

$$L = \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_g} [D(\tilde{\mathbf{x}})] - \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_r} [D(\mathbf{x})] + \lambda \mathbb{E}_{\tilde{\mathbf{x}} \sim \mathbb{P}_{\tilde{\mathbf{x}}}} \left[(\|\nabla_{\tilde{\mathbf{x}}} D(\tilde{\mathbf{x}})\|_2 - 1)^2 \right] \quad (4)$$

where $\mathbb{P}_{\tilde{\mathbf{x}}}$ is defined as a sample uniformly along straight lines between pairs of points sampled from $\mathbb{P}_g$ and $\mathbb{P}_r$.

The process of using the WGAN-GP model to solve the problem of label imbalance is as follows: The generator G shown in Figure 6 takes a random vector from a uniform distribution as input. It outputs a vector including all features of the corresponding class (i.e., the students who failed in securing employment). Next, the generated data and the real data are entered into discriminator D for classification. Through repeated training, the D cannot identify the generated data from the real data. Then we use G to generate data of students who failed in securing employment until the two categories are balanced. In other words, we aim to implicitly learn the distribution of data of students who failed to secure employment to further generate new samples.

5.3 Prediction Model

Compared with the previous work [24] that uses the LSTM variant as the main prediction model, we use the basic TCN model proposed in [3] as the main prediction model for graduate prediction, for the following reasons:

- (1) Compared with the LSTM model, the TCM model provides the flexibility to adjust the receptive field through choosing larger filter sizes and increasing the dilation factor.
- (2) While the data involved in this paper is relatively large in the field of employment prediction in terms of both data volume and data dimension, it is relatively small compared with typical datasets (billions of data) used in **NLP (Natural Language Processing)** experiments or **CV (Computer Vision)** experiments for which TCN models are popular. In this case, we abandon TCN variants like [32] that make the model more complex, in order to overcome the issue of overfitting.

The basic TCN model used in this research is based on a generic architecture proposed in [3], which involves three key components: Causal Convolutions, Dilated Convolutions, and Residual Connections. These components are described in the next sections.

5.3.1 Causal Convolutions. In order to comply with the basic principle that there can be no leakage from the future into the past, causal convolutions are used to process the input information, which means that output at time t is convolved only with elements from time t and earlier in the previous layer. Moreover, 1-D **fully-convolutional network (FCN)** architecture and zero paddings of length (kernel size - 1) are used to keep the length of each layer constant. In a word, TCN is basically equivalent to the combination of 1-D FCN and causal convolutions.

5.3.2 Dilated Convolutions. Dilated convolutions are used to achieve a larger receptive field with fewer parameters.

For a sequence input vector $X \in \mathbb{R}^n$ and a filter $\mathcal{F} : \{0, \dots, k-1\} \rightarrow \mathbb{R}$, the dilated convolution operation F on elements of the sequence is shown as follows:

$$O(X_j) = (X *_d \mathcal{F}) = \sum_{i=0}^{k-1} \mathcal{F}(i) * X_{X_j-i \cdot d} \quad (5)$$

where d is the dilation factor, k is the filter size, and the subscript $X_j - i \cdot d$ denotes the direction of the past. The dilated convolution become a regular convolution when $d = 1$. Dilated convolutions exist in each layer and include parameter d ($d_t = 2^l$ [3]), an activation function $f(\cdot)$ and a residual connection that combines the layer's input and the convolution signal. This part can be represented by matrix multiplication [11] and the details are shown as follows:

$$\tilde{Z}_t^{(j,l)} = f(W_0 \tilde{Z}_{t-d}^{(j,l-1)} + W_1 \tilde{Z}_t^{(j,l-1)}) \quad (6)$$

$$Z_t^{(j,l)} = \tilde{Z}_t^{(j,l-1)} + V \tilde{Z}_t^{(j,l)} + e \quad (7)$$

where $\tilde{Z}_t^{(j,l)}$ and $Z_t^{(j,l)}$ represent results after dilated convolution and adding the residual connection at timestamp t , respectively. $W = [W_0, W_1]$, $W_i \in \mathbb{R}^{F_w \times F_w}$, is the weight matrices of filter. $V \in \mathbb{R}^{F_w \times F_w}$ denotes the weight matrix and F_w denotes the number of filters. $e \in \mathbb{R}^{F_w}$ represents the bias vector for the residual block.

5.3.3 Residual Connections. According to the dilated convolutions mentioned above, the depth of the network increases exponentially if we want to increase the receptive field of the TCN. The whole structure of residual blocks is shown in Figure 8. Residual blocks are involved in the problem caused by the depth of the network like the gradient vanishing problem. Within residual blocks, two layers of dilated causal convolution are involved with ReLU activation. Moreover, Weight normalization is applied to convolutional filters to prevent uncontrolled changes in the gradient. Finally, the spatial dropout mechanism is involved after each dilated convolution for regularization.

5.4 Bias-aware Optimization

5.4.1 Bias-aware Regularization. As mentioned above, employment bias varies by major. This finding motivates us to introduce regularization to eliminate the influence of bias in various majors based on the assumption that it will reduce fluctuations in weighting due to different biases across professions. To dynamically adjust the strength of regularization according to the degree of bias, we design a variable weight decay factor, which represents the difference between bias in a certain major and other majors. The cross-major smoothed regularization with a variable weight decay factor is defined below.

Fig. 8. The diagram of TCN model.

For students in major m , the corresponding regularization term in the loss function is as follows:

$$\Omega_m = \frac{1}{2} \sum_{n \neq m}^M \|W * (\mathbf{u}_m - \mathbf{u}_n)\|_F^2 \quad (8)$$

where M represents the collection of all majors. $*$ represents the Hadamard product of two matrices (if one of them is a vector, then the vector is expanded into a matrix through the broadcast mechanism). W is the weight matrix of TCN mentioned above. $\|\cdot\|_F^2$ denotes the Frobenius norm. $\mathbf{u}_m$ is the bias vector for major m . In other words, it is the feature importance vector of students of major m in employment prediction. Each element of $\mathbf{u}_m$ is represented by the Chi-square test p value corresponding to the feature calculated in the Section 4 through a transformation function. In other words, the lower the p value, the greater the weight of the bias (note that theoretically, the importance of the academic performance could be chosen as an arbitrary value, since the subtraction would be 0). The transformation function is defined as follows:

$$f(x) = \frac{e^{1-x} - e^{1+x}}{e^{1-x} + e^{1+x}} \quad (9)$$

5.4.2 Optimization. Based on the discussion above, we formulate the whole loss function of our SUMMER prediction framework below.

For students in major m , the corresponding loss function is as follows:

$$\begin{aligned} \mathcal{L}_m &= \frac{1}{2} \sum_m^M ((W \mathbf{x}_i^m - y_i^m)^2 + \Omega_m) \\ &= \frac{1}{2} \sum_m^M \left((W \mathbf{x}_i^m - y_i^m)^2 + \frac{1}{2} \sum_{n \neq m}^M \|W * (\mathbf{u}_m - \mathbf{u}_n)\|_F^2 \right) \end{aligned} \quad (10)$$

Fig. 9. The results of representation learning for academic performance. S1-S6 denote the six semesters, respectively.

where x_i^m and y_i^m represent the input features and labels of the corresponding student i from major m , respectively. M represents the set of all majors. W is the weight matrix of TCN mentioned above. $\|\cdot\|_F^2$ denotes the Frobenius norm. u_m indicates represents the degree of bias in each of the input dimensions in major m .

6 EXPERIMENTS AND RESULTS

In this section, we present detailed experimental results to demonstrate the effectiveness of our proposed SUMMER framework. We first introduce the representation results of academic performance. Then, we introduce the experimental settings of employment status and its prediction results. Finally, we introduce the experimental settings of employment preference and its prediction results.

6.1 Representation of Academic Performance

To deal with the heterogeneity of courses enrolled in by students each semester, we design a C matrix to denote students' academic performance. The four-year university life involves 8 semesters. Campus recruitment takes place densely at the beginning of the final year of study. Hence, only the academic performance of the previous three years (or six semesters S1 to S6) would affect students' employment. Autoencoder is applied to embed academic performance data to overcome the heterogeneity of course selection. We test different dimensions including 3, 6, 12, 24, 32, 64, 80, 96 and the performance is shown in Figure 9. The value of the loss function fluctuates slightly. That is, even vectors with low dimensions can still effectively represent the academic performance of each student. Thus, we choose 3 as the dimension of representation for computational efficiency.

6.2 Prediction Results of Employment Status

We first predict the employment status with features including academic performance, gender, ethnic group, enrolment status, hometown, and student major. To verify the effectiveness of our SUMMER framework, we design prediction experiments including two settings: comparison with TCN-based SUMMER's variants and comparison with representative baselines.

6.2.1 Comparison with TCM-based SUMMER's Variants. Table 3 displays the prediction performance of SUMMER and its variants. We design a three-step experiment to test the performance with metrics, i.e., accuracy, recall, and F1-score, to understand the results collectively.

Table 3. Performance of SUMMER Variants on Employment Status Prediction

Variants	Accuracy	Recall	F1-score
TCN+Raw Data	0.869	0.500	0.475
TCN+WGAN-GP	0.875	0.690	0.747
TCN+WGAN-GP+New Loss	0.890	0.786	0.830

TCN+Raw Data represents the prediction result of TCN model training on the original imbalanced data. *TCN+WGAN-GP* represents the prediction result of TCN model on the balanced dataset generated through data augmentation process. *TCN+WGAN-GP+New Loss* represents the prediction result of the TCN model with the proposed loss function on the balanced dataset generated through the data augmentation process.

In the first step, we use raw data to fit the original TCN model. The imbalanced label issue exists in raw data, leading to the occurrence of unexpected results on recall and F1 score. The computational logic of precision and recall is to calculate the corresponding metric for each category and then take the mean value. Based on the significant label imbalance of our employment dataset, the algorithm inevitably ignores the minority class and predicts all the samples into the majority class, which results in one type of precision and recall approaching 1 and the other approaching 0. In this case, the mean value of final indicators recall and F1 score is around 0.5, which is consistent with Table 3.

In the second step, WGAN-GP is used to solve the label imbalance problem. In the case of a fixed test set, we separately test the prediction performance based on the raw training set and the prediction performance based on the training set with data augmentation. According to the principle of control variables, such an experimental design can effectively detect the contribution of data augmentation. The data generation process as follows:

- First, stratified sampling is used to divide the raw data into two categories: training set $\mathbf{a}$ and testing set $\mathbf{b}$.
- Second, we use WGAN-GP on the training set $\mathbf{a}$ to generate samples of the minority class. In the resulting training set $\mathbf{a}'$, the number of students in the two classes is equal.

Next, we use the training set $\mathbf{a}'$ to fit the model and test it on the original testing set $\mathbf{b}$. The performance shown in Table 3 verifies its effectiveness. In the final step, we add the bias-based regularization into the optimization loss based on the last step and the results suggest its importance.

6.2.2 Comparison with Baseline Methods. In addition to the comparison with TCN-based variants, we compare the SUMMER framework with several popular algorithms:

- **SVM** [7]: SVM is a classic algorithm and is widely used in the field of data mining.
- **Random Forest** [6]: is a classic ensemble algorithm that achieves good performance in various applications.
- **XGBoost** [9]: XGBoost is a boosting-tree-based method and is widely used in various data mining scenarios with good performance.
- **MAYA**: MAYA is a targeted framework for graduate employment proposed in our previous work [24]. The proposed framework is improved based on the MAYA framework. Compared with the proposed framework, the MAYA framework has the following differences: (1) In the MAYA framework, an LSTM model with a temporal dropout mechanism is used as the prediction model; and (2) In the MAYA framework, a normal GAN model is used for data augmentation.

Fig. 10. Performance of employment status prediction on raw training dataset.

Fig. 11. Performance of employment status prediction on balanced training dataset.

Note that we test the performance of these algorithms from two aspects. On the one hand, we fit algorithms based on the raw training set a and test them on the testing set b . The results are shown in Figure 10. It's shown that the prediction is not accurate and the fluctuation is quite large due to the label imbalance of raw data. To overcome this problem, we fit algorithms based on the balanced training set a' and test them on b . The results are shown in Figure 11. The performance is improved significantly.

Next we analyze the input features and the parameters involved in the experiment of employment status prediction.

6.2.3 Input Features. It is highly important to identify the students who might encounter difficulties in employment at an early stage so that teachers and other support services can intervene in a timely manner. Therefore, we conduct a test on the number of semesters involved in academic performance data. As shown in Table 4, the more semesters of academic performance included as input, the better the performance of the algorithm. This implies that educational administrators should assess students' future employment status each semester rather than just once. In addition, the results show that academic performance in the later stage of college life contributes more to employment projections. As job interviews generally occur in the late stage of college life, academic performance during this period may also reflect the student's focus on performing well in job interviews.

Moreover, we test the contribution of demographic features in prediction. A two-step experiment is designed: First, we use the academic performance of six semesters only to predict employment status. Second, we use the academic performance of six semesters and demographic features to do the prediction with loss function Equation (10). The results are shown in Table 4. As expected, the introduction of demographic features improve the prediction performance significantly.

6.2.4 Learning Rate. Learning rate that controls the update speed of the model is an important parameter in the SUMMER framework. In Figure 12, we analyze the performance of various

Table 4. Performance of Employment Status Prediction with Different Inputs

Inputs	Accuracy	Recall	F1
1 semester + demographic feature	0.75463	0.64678	0.63418
2 semesters + demographic feature	0.80287	0.65231	0.69371
3 semesters + demographic feature	0.86224	0.66746	0.70124
4 semesters + demographic feature	0.87731	0.68842	0.72121
5 semesters + demographic feature	0.88612	0.71248	0.77452
6 semesters + demographic feature	0.89000	0.78646	0.83025
Academic performance	0.87313	0.68911	0.73875
Academic performance + demographic feature	0.89000	0.78646	0.83025

' n semesters' represents using academic performance of n semesters as input.

Table 5. Effect of Different Optimization Strategies on Prediction of Employment Status

Optimization Function	Accuracy	Recall	F1-score
Equation (11)	0.874	0.691	0.746
Equation (10)	0.890	0.786	0.830

Fig. 12. Analysis of learning rates in prediction of employment status.

learning rates and find that the model can achieve the best prediction performance when the learning rate is set to 0.01.

6.2.5 Bias-based Regularization. We design an experiment to test the effectiveness of bias based regularization. We use Equations (10) and (11) as loss function, separately and the prediction performance is shown in Table 5. The bias-based regularization can improve the performance remarkably.

$$\mathcal{L}_{normal} = \frac{1}{2} \sum_{m=1}^M (\mathbf{W}\mathbf{x} - y)^2 + \|\mathbf{W}\|_F^2 \quad (11)$$

where $\mathbf{W}$ represents the weight matrix of the model. $\mathbf{x}$ and y represent the input features and labels of the corresponding data, respectively.

6.2.6 Transformation Function. As mentioned in Section 5.4.1, we need a monotonically decreasing function between 0 and 1 as the transformation function to process the P value. Here,

Table 6. Effect of Different Transformation Functions on Prediction of Employment Status

Transformation Function	Accuracy	Recall	F1-score
$y = 1 - x$	0.864	0.751	0.806
$y = 1 - x^2$	0.873	0.768	0.801
$y = \frac{1}{1+e^x}$	0.870	0.778	0.813
Equation (9)	0.890	0.786	0.830

Table 7. Performance of SUMMER Variants on Employment Preference Prediction

Variants	Accuracy	Recall	F1-score
TCN+Raw Data	0.877	0.687	0.751
TCN+WGAN-GP	0.895	0.711	0.776
TCN+WGAN-GP+New Loss	0.920	0.817	0.861

TCN+Raw Data represents the prediction result of TCN model training on the original imbalanced data. *TCN+WGAN-GP* represents the prediction result of TCN model on the balanced dataset generated through data augmentation process. *TCN+WGAN-GP+New Loss* represents the prediction result of the TCN model with the proposed loss function on the balanced dataset generated through the data augmentation process.

three functions are selected that meet the requirements for comparison experiments and results are shown in Table 6. It can be seen from the above results that the effect of Equation (9) is better than that of other transformation functions.

6.3 Prediction Results of Employment Preference

We predict the employment preference of successfully employed students through the same experiment setting mentioned in Section 6.2. The features used in this part include academic performance, gender, ethnic group, enrolment status, hometown, and their major. Similarly, comparison with TCN-based SUMMER's variants and comparison with representative baselines are used to verify the effectiveness of our SUMMER framework.

First, we evaluate the performance of the SUMMER framework and its variants, and the results are shown in Table 7, which demonstrates the validity of each part involved in our proposed framework. Compared with the results of employment status shown in Table 3, there are two obvious differences. Firstly, the performance of *TCN+RawData* is much improved, because the label imbalance in the employment preference prediction is much less severe than in the employment status prediction. Second, in some cases students' employment preferences are closely related to their major, causing the framework to perform better on employment preference prediction than on employment status prediction.

Second, the SUMMER framework is compared with several popular algorithms described in Section 6.2 including SVM, Random Forest, XGBoost, and MAYA framework. The prediction results on the raw dataset are shown in Figure 13. First, the experimental results demonstrate the advantages of our proposed prediction framework SUMMER. Moreover, compared with the prediction on employment status, all algorithms make significant improvements in performance on the prediction of employment preference and the performance of all algorithms is relatively stable. The reason for this is that the label imbalance in the employment preference prediction is much less severe than in the employment status prediction, as already mentioned above.

Fig. 13. Performance of employment preference prediction on raw training dataset.

Fig. 14. Performance of employment preference prediction on balanced training dataset.

Moreover, we test all algorithms on the balanced dataset generated by WGAN-GP and show the performance in Figure 14. All results demonstrate the validity of our proposed framework.

The input features and parameters involved in this experiment are described below.

6.3.1 Input Features. Identifying a student's employment preferences at an early stage can provide targeted guidance to help the student find a satisfying job. In this case, we first conduct a test on the number of semesters involved in academic performance data. As shown in Table 8, prediction performance grows slowly from the fourth semester. With the addition of more data on academic performance, the predicted improvement in effect was not as great as in the experiment of employment status prediction. The reason for this is that employment preference is also very relevant to a student's major.

Moreover, we test the contribution of demographic features in prediction. In the same experimental setting with the prediction experiment of employment status, we use a two-step experiment. First, we use the academic performance of six semesters only to predict employment status. Second, we use the academic performance of six semesters and demographic features to perform the prediction with loss function Equation (10). The results are shown in Table 8. As shown, the demographic features improve the prediction performance significantly.

6.3.2 Learning Rate. Similarly, we test the influence of learning rate on prediction performance. In Figure 15, we analyze the performance of various learning rates and find that the model can achieve the best prediction performance when the learning rate is set to 0.01.

6.3.3 Bias-based Regularization. We design an experiment to test the effectiveness of bias based regularization on prediction of employment preference. We use Equations (10) and (11) as loss function separately and the prediction performance is shown in Table 9. The bias-based regularization can improve the performance remarkably.

Table 8. Performance of Employment Preference Prediction with Different Inputs

Input	Accuracy	Recall	F1
1 semester + demographic feature	0.82436	0.73648	0.74313
2 semesters + demographic feature	0.83215	0.75371	0.76371
3 semesters + demographic feature	0.88244	0.77377	0.79332
4 semesters + demographic feature	0.89123	0.78121	0.81233
5 semesters + demographic feature	0.90377	0.79314	0.83445
6 semesters + demographic feature	0.92000	0.81701	0.86134
Academic performance	0.89233	0.71091	0.76008
Academic performance + demographic feature	0.92000	0.81701	0.86134

' n semesters' represents using academic performance of n semesters as input.

Table 9. Effect of Different Optimization Strategies on Prediction of Employment Preference

Optimization Function	Accuracy	Recall	F1-score
Equation (11)	0.894	0.731	0.773
Equation (10)	0.920	0.817	0.861

Table 10. Effect of Different Transformation Functions on Prediction of Employment Preference

Transformation Function	Accuracy	Recall	F1-score
$y = 1 - x$	0.890	0.800	0.821
$y = 1 - x^2$	0.901	0.803	0.833
$y = \frac{1}{1+e^x}$	0.912	0.811	0.842
Equation (9)	0.920	0.817	0.861

Fig. 15. Analysis of learning rates in prediction of employment preference.

6.3.4 Transformation Function. We test the effect of different transformation functions on the prediction of employment preference. As above, here we select three transformation functions that meet the requirements for comparison experiments and results are shown in Table 10. It can be seen from the displayed results that the effect of Equation (9) is better than that of other transformation functions.

7 CONCLUSION

In this paper, we analyzed a large-scale educational dataset for predicting graduates' employment status and employment preference. Since the impact of unconscious biases on graduate employment are significant, we first analyze the employment bias of different majors from two aspects: employment status and employment preference, and verify the existence of employment bias. Then, based on the homogeneity problems that exist in the prediction of employment status and employment preference, SUMMER, a prediction framework, is proposed to predict graduates' employment status and employment preference. We incorporate autoencoder to ease the data-sparsity issue and deal with the label imbalance problem using WGAN-GP. TCN is used to capture the sequentiality between semesters and a bias-based regularization is introduced to weaken the impact of biases.

Our extensive experiments based on a real-world education dataset demonstrate the proposed framework can improve the prediction performance significantly, and SUMMER outperforms other baseline models significantly, including MAYA, LSTM and XGBoost.

There are multiple directions for future work:

- (1) We plan to expand our dataset and explore graduate employment from more perspectives, including psychology, sociology and pedagogy.
- (2) We will examine intervention strategies for students with employment problems at an educational as well as a psychological level.
- (3) While this study focused on student-centric data, future work would include data acquisition from various companies to further study this issue from the perspective of employers.
- (4) Last but not the least, we also intend to integrate the SUMMER framework into a modern educational management system and apply it to detect the employment status of graduating students.

ACKNOWLEDGMENTS

The authors would like to thank Shihao Zhen and Dongyu Zhang from Dalian University of Technology for help with experiments. Part of this work was done when the first author worked at Dalian University of Technology.

REFERENCES

- [1] Omar Al-Ubaydli and John A. List. 2019. How natural field experiments have enhanced our understanding of unemployment. *Nature Human Behaviour* 3, 1 (2019), 33–39.
- [2] Stefania Albanesi and Jiyeon Kim. 2021. Effects of the COVID-19 recession on the US labor market: Occupation, family, and gender. *Journal of Economic Perspectives* 35, 3 (2021), 3–24.
- [3] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *ArXiv Preprint ArXiv:1803.01271* (2018).
- [4] Xiaomei Bai, Fuli Zhang, Jinzhou Li, Teng Guo, Abdul Aziz, Aijing Jin, and Feng Xia. 2021. Educational big data: Predictions, applications and challenges. *Big Data Research* 26 (2021), 100270.
- [5] Abigail Barr, Luis Miller, and Paloma Ubeda. 2016. Moral consequences of becoming unemployed. *Proceedings of the National Academy of Sciences* 113, 17 (2016), 4676–4681.
- [6] Leo Breiman. 2001. Random forests. *Machine Learning* 45, 1 (2001), 5–32.
- [7] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 3 (2011), 1–27.
- [8] Jiawei Chen, Yuexiang Li, Kai Ma, and Yefeng Zheng. 2020. Generative adversarial networks for video-to-video domain adaptation. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. AAAI Press, 3462–3469.
- [9] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 785–794.
- [10] Aaron Clauset, Samuel Arbesman, and Daniel B. Larremore. 2015. Systematic inequality and hierarchy in faculty hiring networks. *Science Advances* 1, 1 (2015), e1400005.

- [11] Shumin Deng, Ningyu Zhang, Wen Zhang, Jiaoyan Chen, Jeff Z. Pan, and Huajun Chen. 2019. Knowledge-driven stock trend prediction and explanation via temporal convolutional network. In *Companion Proceedings of The 2019 World Wide Web Conference*. ACM, 678–685.
- [12] Shady Elbassouni, Sihem Amer-Yahia, and Ahmad Ghizzawi. 2020. Fairness of scoring in online job marketplaces. *ACM Transactions on Data Science* 1, 4 (2020), 1–30.
- [13] Paula England, Andrew Levine, and Emma Mishel. 2020. Progress toward gender equality in the United States has slowed or stalled. *Proceedings of the National Academy of Sciences* 117, 13 (2020), 6990–6997.
- [14] Teresa M. Evans, Lindsay Bira, Jazmin Beltran Gastelum, L. Todd Weiss, and Nathan L. Vanderford. 2018. Evidence for a mental health crisis in graduate education. *Nature Biotechnology* 36, 3 (2018), 282.
- [15] Yazan Abu Farha and Jurgen Gall. 2019. MS-TCN: Multi-stage temporal convolutional network for action segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 3575–3584.
- [16] The Organisation for Economic Co-operation and Development. 2020. Youth unemployment rate (indicator). <https://data.oecd.org/unemp/youth-unemployment-rate.htm>.
- [17] National Center for Education Statistics. 2020. College student employment. https://nces.ed.gov/programs/coe/pdf/coe_ssa.pdf.
- [18] Heather L. Ford, Cameron Brick, Karine Blaufuss, and Petra S. Dekens. 2018. Gender inequity in speaking opportunities at the American Geophysical Union fall meeting. *Nature Communications* 9, 1 (2018), 1358.
- [19] Daniel Fuerstman and Stephan Lavertu. 2005. The academic hiring process: A survey of department chairs. *PS: Political Science and Politics* 38, 4 (2005), 731–736.
- [20] Yating Gao, Wu Wang, Qiongbai Lin, Fenghuang Cai, and Qinqin Chai. 2020. Fault diagnosis for power converters based on optimized temporal convolutional network. *IEEE Transactions on Instrumentation and Measurement* 70 (2020), 1–10.
- [21] Konstantinos Giannakas, Murray Fulton, and Tala Awada. 2017. Hiring leaders: Inference and disagreement about the best person for the job. *Palgrave Communications* 3, 1 (2017), 17.
- [22] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. 2017. Improved training of Wasserstein GANs. *arXiv preprint arXiv:1704.00028* (2017).
- [23] Teng Guo, Xiaomei Bai, Tian Xue, Selena Firmin, and Feng Xia. 2021. Educational anomaly analytics: Features, methods, and challenges. *Frontiers in Big Data* 4 (2021), 811840.
- [24] Teng Guo, Feng Xia, Shihao Zhen, Xiaomei Bai, Dongyu Zhang, Zitao Liu, and Jiliang Tang. 2020. Graduate employment prediction with bias. In *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, 670–677.
- [25] Finn Hedefalk and Martin Dribe. 2020. The social context of nearest neighbors shapes educational attainment regardless of class origin. *Proceedings of the National Academy of Sciences* 117, 26 (2020), 14918–14925.
- [26] Matthew T. Hora. 2020. Hiring as cultural gatekeeping into occupational communities: Implications for higher education and student employability. *Higher Education* 79, 2 (2020), 307–324.
- [27] Amanda J. Koch, Susan D. D’Mello, and Paul R. Sackett. 2015. A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology* 100, 1 (2015), 128.
- [28] Christopher Kok, V. Jahmunah, Shu Lih Oh, Xujuan Zhou, Raj Gururajan, Xiaohui Tao, Kang Hao Cheong, Rashmi Gururajan, Filippo Molinari, and U. Rajendra Acharya. 2020. Automated prediction of sepsis using temporal convolutional network. *Computers in Biology and Medicine* 127 (2020), 103957.
- [29] Jie Kong, Meng Ren, Ting Lu, and Congying Wang. 2018. Analysis of college students’ employment, unemployment and enrollment with self-organizing maps. In *International Conference on E-Learning and Games*. Springer, 318–321.
- [30] Augustine J. Kposowa, Dina Aly Ezzat, and Kevin Breault. 2019. New findings on gender: The effects of employment status on suicide. *International Journal of Women’s Health* 11 (2019), 596–575.
- [31] Linda Hamilton Krieger. 1995. The content of our categories: A cognitive bias approach to discrimination and equal employment opportunity. *Stanford Law Review* 47 (1995), 1161–1248.
- [32] Lei Lin, Beilei Xu, Wencheng Wu, Trevor W. Richardson, and Edgar A. Bernal. 2019. Medical time series classification with hierarchical attention-based temporal convolutional networks: A case study of myotonic dystrophy diagnosis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 83–86.
- [33] Jiaying Liu, Xiangjie Kong, Feng Xia, Xiaomei Bai, Lei Wang, Qing Qing, and Ivan Lee. 2018. Artificial Intelligence in the 21st century. *IEEE Access* 6 (2018), 34403–34421.
- [34] Jiaying Liu, Feng Xia, Lei Wang, Bo Xu, Xiangjie Kong, Hanghang Tong, and Irwin King. 2021. Shifu2: A network representation learning based model for advisor-advisee relationship mining. *IEEE Transactions on Knowledge and Data Engineering* 33, 4 (2021), 1763–1777.
- [35] Li Liu, Deborah Silver, and Karen Bemis. 2018. Application-driven design: Help students understand employment and see the “big picture”. *IEEE Computer Graphics and Applications* 38, 3 (2018), 90–105.
- [36] Rui Liu, Wenge Rong, Yuanxin Ouyang, and Zhang Xiong. 2017. A hierarchical similarity based job recommendation service framework for university students. *Frontiers of Computer Science* 11, 5 (2017), 912–922.

- [37] Yuetian Luo and Zachary A. Pardos. 2018. Diagnosing university student subject proficiency and predicting degree completion in vector space. In *Thirty-Second AAAI Conference on Artificial Intelligence*. AAAI Press, 7920–7927.
- [38] Iftekhar Naim, Md Iftekhar Tanveer, Daniel Gildea, and Mohammed Ehsan Hoque. 2016. Automated analysis and prediction of job interview performance. *IEEE Transactions on Affective Computing* 9, 2 (2016), 191–204.
- [39] European Statistical Office. 2019. Employment rates of recent graduates. https://ec.europa.eu/eurostat/statistics-explained/index.php/Employment_rates_of_recent_graduates.
- [40] Carolyn Parkinson, Adam M. Kleinbaum, and Thalia Wheatley. 2018. Similar neural responses predict friendship. *Nature Communications* 9, 1 (2018), 332.
- [41] Chuan Qin, Hengshu Zhu, Chen Zhu, Tong Xu, Fuzhen Zhuang, Chao Ma, Jingshuai Zhang, and Hui Xiong. 2019. DuerQuiz: A personalized question recommender system for intelligent job interview. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2165–2173.
- [42] Dazhong Shen, Hengshu Zhu, Chen Zhu, Tong Xu, Chao Ma, and Hui Xiong. 2018. A joint learning approach to intelligent job interview assessment. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*. 3542–3548.
- [43] Ke Sun, Lei Wang, Bo Xu, Wenhong Zhao, Shyh Wei Teng, and Feng Xia. 2020. Network representation learning: from traditional feature learning to deep learning. *IEEE Access* 8 (2020), 205600–205617.
- [44] Noriko Uosaki, Kousuke Mouri, Chengjiu Yin, and Hiroaki Ogata. 2018. Seamless support for international students' job hunting in Japan using learning log system and ebook. In *2018 7th International Congress on Advanced Applied Informatics*. IEEE, 374–377.
- [45] Marijtje A. J. Van Duijn, Evelien P. H. Zeggelink, Mark Huisman, Frans N. Stokman, and Frans W. Wasseur. 2003. Evolution of sociology freshmen into a friendship network. *Journal of Mathematical Sociology* 27, 2–3 (2003), 153–191.
- [46] Paul Westermann, Matthias Welzel, and Ralph Evins. 2020. Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones. *Applied Energy* 278 (2020), 115563.
- [47] Xindong Wu, Xingquan Zhu, Gong Qing Wu, and Wei Ding. 2013. Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering* 26, 1 (2013), 97–107.
- [48] Feng Xia, Jiaying Liu, Hansong Nie, Yonghao Fu, Liangtian Wan, and Xiangjie Kong. 2019. Random walks: A review of algorithms and applications. *IEEE Transactions on Emerging Topics in Computational Intelligence* 4, 2 (2019), 95–107.
- [49] Jining Yan, Lin Mu, Lizhe Wang, Rajiv Ranjan, and Albert Y. Zomaya. 2020. Temporal convolutional networks for the advance prediction of ENSO. *Scientific Reports* 10, 1 (2020), 1–15.
- [50] Rui Yan, Ran Le, Yang Song, Tao Zhang, Xiangliang Zhang, and Dongyan Zhao. 2019. Interview choice reveals your preference on the market: To improve job-resume matching through profiling memories. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 914–922.
- [51] Luoying Yang, Zhou Xu, and Jiebo Luo. 2020. Measuring female representation and impact in films over time. *ACM Transactions on Data Science* 1, 4 (2020), 1–10.
- [52] Jiaxuan You, Yichen Wang, Aditya Pal, Pong Eksombatchai, Chuck Rosenberg, and Jure Leskovec. 2019. Hierarchical temporal convolutional networks for dynamic recommender systems. In *The World Wide Web Conference*. ACM, 2236–2246.
- [53] Kunpeng Zhang, Zijian Liu, and Liang Zheng. 2019. Short-term prediction of passenger demand in multi-zone level: Temporal convolutional neural network with multi-task learning. *IEEE Transactions on Intelligent Transportation Systems* 21, 4 (2019), 1480–1490.
- [54] Yang Zhang and Tao Cheng. 2019. A deep learning approach to infer employment status of passengers by using smart card data. *IEEE Transactions on Intelligent Transportation Systems* 21, 2 (2019), 617–629.
- [55] Mu Zhou, Yixin Lin, Nan Zhao, Qing Jiang, and Zengshan Tian. 2020. Indoor WLAN intelligent target intrusion sensing using ray-aided generative adversarial network. *IEEE Transactions on Emerging Topics in Computational Intelligence* 4, 1 (2020), 61–73.

Received January 2021; revised November 2021; accepted January 2022