Engineering Applications of Artificial Intelligence 127 (2024) 107306

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Assessing growth potential of careers with occupational mobility network and ensemble framework

Jiamin Liu [a,1], Tao Wang [a,1], Feng Yao [a], Witold Pedrycz [b,c,d], Yanjie Song [a,*], Renjie He [a,*]

[a] College of Systems Engineering, National University of Defense Technology, Changsha, 410073, Hunan, China
[b] Department of Electrical & Computer Engineering, University of Alberta, Edmonton AB, T6G 2R3, Canada
[c] Systems Research Institute, Polish Academy of Sciences, Warsaw, 00-901, Poland
[d] Department of Computer Engineering, Faculty of Engineering and Natural Sciences, Istinye University, Sariyer/Istanbul, Turkiye

ARTICLE INFO

Keywords:
Human capital
Occupational mobility network
Growth potential of career
Machine learning

ABSTRACT

The growth potential of a career reflects its future prospects and is an important consideration for individuals and organizations when career planning. There is still a lack of quantitative assessment tools for growth potential of careers. In this study, considering the key role of human capital in human resource management, as well as the excellent performance of complex network and machine learning in big data analysis and prediction, a career growth potential assessment model with human capital ensemble is proposed through human capital-based occupational mobility network and ensemble learning. First, an occupational mobility network is constructed based on online professional dataset to associate occupations with each other. Then, five dimensions of human capital measurements are designed to quantify human capital in terms of education, experience, social capital, occupational size, and concentration. These are then combined with the occupational mobility network to create a new network that depicts human capital flows among occupations. Finally, an ensemble framework for assessing career growth potential is constructed to integrate multidimensional human capital information in the network and obtain quantitative scores of growth potential. This study is the original attempt to adopt a data-driven idea and an intelligent approach to understand career growth potential. The experimental results show that it also makes a useful exploration for modeling human capital flows and intelligent assessment of career prospects.

1. Introduction

In the era of Industry 4.0, the production and operation modes of enterprises are becoming increasingly intelligent. Accordingly, the management mode has to be improved, especially human resource management (HRM) (Shamim et al., 2016; Trotta and Garengo, 2018). Leveraging big data and artificial intelligence technologies, many HRM tasks are heading towards intelligence, such as recruitment (Shen et al., 2021), personnel allocation (Bocewicz et al., 2023; Fischer et al., 2020), training (Wang et al., 2021) and talent retention (Teng et al., 2021). In addition, career planning is also one of the topics of interest to HR managers and individual talents. With the development of industry and economy, new knowledge and technology are continually emerging and converging to create novel working methods and career paths. People are presented with a plethora of career options, necessitating that they consider not only the workplace and salary, but also the potential for development. Therefore, it is worthwhile to assess the growth potential of careers to assist decision-making for career planning.

In the past, many researches have focused on individual potential when exploring career development (Wicks, 2001; Sibunruang and Kawai, 2023; Healy et al., 2022). However, an individual's career development is often limited by the development of the particular occupation he or she chooses. The growth potential of careers often reflect the technical and economic characteristics of the times. For example, with the evolution of information technology, programmer has become one of the most promising careers. At present, the rise of new energy technology puts electric vehicle engineer at the perfect opportunity. Also, the decline in infrastructure demand has triggered the decline of civil engineers. Thus, it is of great value for individuals and enterprises to understand the growth potential of careers.

In this study, growth potential of career is defined as careers' capacity to have desirable prospects for development in the future.

* Corresponding authors.
E-mail addresses: liujm_1020@163.com (J. Liu), wangtao@nudt.edu.cn (T. Wang), yaofeng@nudt.edu.cn (F. Yao), wpedrycz@ualberta.ca (W. Pedrycz), songyj_2017@163.com (Y. Song), renjiehe@nudt.edu.cn (R. He).
[1] Co-first authors.

https://doi.org/10.1016/j.engappai.2023.107306
Received 27 April 2023; Received in revised form 13 August 2023; Accepted 12 October 2023
Available online 17 October 2023
0952-1976/

Currently, some universities,[2,3] governments,[4] and consulting platforms[5] have considered career prospect or potential to be an important factor for students and job seekers when career planning. But there is still missing a systematic and intelligent way to predict or evaluate career growth potential. As mentioned earlier, the emerging information technology has achieved good results in the intelligence of HRM, which is an innovation of management engineering. Such a new approach is also needed for the assessment of career growth potential. Realizing the assessment based on artificial intelligence can objectively and quickly provide judgment on career prospects, which not only helps individuals to make career decisions but also helps organizations to plan talent development.

Simultaneously, as a large number of industries have shifted from labor-intensive to knowledge-intensive, HRM has been upgraded from strategic HRM to human capital management. The knowledge and skills of talents are the core competitiveness of companies. Human capital has been found to be the strong predictor of growth potential in firms and individuals (Bonaventura et al., 2020; Ge et al., 2023). It is all the more important, then, to take human capital into account when evaluating career growth potential.

Overall, this study first needs to quantify human capital and relate it to careers, and then assess the growth potential of careers based on intelligent techniques. Therefore, in this study, we start by constructing an occupational mobility network (OMN) describing occupations and their interrelationships based on data from online professional networking sites. We then design five dimensions of human capital measurements (HC^5) to quantify human capital in terms of education, experience, social capital, occupational size, and concentration. We combine HC^5 measurements with the OMN to create HC^5 measurements in which human capital flows between occupations are depicted. Next, we draw on the idea of ensemble learning to construct a predictive model called CPACE based on ensemble framework. The model can integrate multidimensional human capital information from the HC^5 network and effectively assess growth potential of careers.

The contributions of our study are as follows:

(1) **The combination of human capital, occupational mobility, and machine learning enables an intelligent quantitative assessment of career growth potential.** To the best of our knowledge, this is the first time to use data-driven and intelligent methods to understand career growth potential.

(2) **The design of multidimensional human capital measurement effectively quantify human capital and improve occupational mobility network.** Single dimensional measurement is one-sided. For example, in terms of education, a doctor with one year's service is more valuable than an undergraduate with ten years' service. But from the perspective of experience, the undergraduate is more expensive. Our method can provide a more comprehensive and objective approach to model human capital flows by reasonable quantification and complex network.

(3) **The prediction model based on ensemble learning integrates multidimensional human capital information and significantly enhances the accuracy of career growth potential assessment.** Simple arithmetic operations lack complete theoretical support. Direct concatenation of features may lead to redundancy and interference. The proposed ensemble framework can process multidimensional information coming from HC^5 network and extract valuable knowledge before integration. It also helps to identify the factors influencing career prospects and the mechanisms of career mobility.

This paper is organized as follows: Section 2 describes the work in related areas. Section 3 introduces the data sources and construction of the HC^5 network. Section 4 elaborates the framework and components of CPACE model. Then experimental results are presented in Section 5. The final section summarizes the study and discusses application scenarios, limitations and future directions.

2. Related work

Occupational mobility analysis. Occupational mobility is a key area of study within HRM. Scholars have investigated the factors that influence occupational mobility. Ref. Ng et al. (2007) proposed a theoretical framework comprising constructual factors, individual difference, and decisional factors to explain occupational mobility. Some studies demonstrated that education level, salary, etc., can motivate or discourage workers to change jobs (Sicherman, 1990; Parrado et al., 2007). Quantitative methods are increasingly being used in this field of research, such as Cortes et al.'s use of the gravity equation to measure the costs of occupational mobility (Cortes and Gallipoli, 2017), and Xu et al.'s application of data mining to learn occupational mobility patterns (Xu et al., 2015). Complex networks are also widely used to depict occupational mobility (Schmutte, 2014). At the micro level, Ref. Vaccario et al. (2021) and Xu et al. (2016) analyzed and predicted individuals' career decisions based on mobility networks. Besides, at the macro level, Ref. Nimczik (2017) and Park et al. (2019) classified labor markets and explored employment trends by analyzing labor flows. However, few studies have explored the relationship between occupational mobility and the growth potential of occupation itself.

Human capital measurement. In 1776, Smith proposed that human capital is the acquired and useful ability of society members (Smith, 2010). Generally speaking, human capital embraces the knowledge, skills, and experience of individuals and their contribution in terms of productivity (Baron, 2011). As a large number of industries shift from labor-intensive to technology-intensive, calculating human capital by headcount is no longer applicable. For the quantification of human capital, most studies use one of the ideas of educational background, age heaping, contribution to the economy (i.e. income), and consumed resources (i.e. costs) (A'Hearn et al., 2009; Folloni and Vittadini, 2010; Wossmann, 2003; Le et al., 2005). Specific indicators include the percentage of highly educated managers, average salary, the average age of employees, employee turnover, annual staff training costs, the scale of the project team (Lu et al., 2014). At present, human capital measurement involves different resources and different dimensions, and there is no unified and recognized method for its quantification. Additionally, there is a lack of means to integrate existing methods.

Growth potential prediction. Some HRM-related studies have sought to investigate the growth potential of both organizations and individuals. Sedláček et al. analyzed the growth potential of startups based on the business cycle (Sedláček and Sterk, 2017). Using a worldwide start-up network, Bonaventura et al. used the closeness of enterprise nodes to predict the success of the companies in the future (Bonaventura et al., 2020). Human capital has also been identified as an essential reference factor in the estimation and forecasting of growth potential. Safavi et al. modeled career mobility by combining experience, corporate expansion and employee retention to uncover key organizations in the history of computing (Safavi et al., 2018). Ge et al. constructed mobility networks by quantifying talent flows as human capital from the educational dimension, and used PageRank and Reverse PageRank to estimate the ranking of companies (Ge et al., 2023). Our team has also successfully predicted the growth potential of employees based on working expertise, interpersonal environment, and competency (Liu et al., 2021a, 2019; Li et al., 2020). However, few studies have explored growth potential at the career level.

[2] www.waldenu.edu/programs/resource/careers-with-growth-potential-and-the-degree-youll-need.
[3] clas.wayne.edu/philosophy/programs/career-outlook.
[4] www.careers.govt.nz/articles/7-careers-with-a-promising-future.
[5] www.shrm.org/resourcesandtools/hr-topics/organizational-and-employee-development/career-advice/pages/find-jobs-with-career-potential-in-your-region.aspx.

This literature review shows that current researches lack attention to growth potential of careers. Occupational mobility network is an effective method to analyze labors and human resources, and it is valuable to combine the network with intelligent technology for assessing career growth potential. Meanwhile, at a time when human capital is becoming increasingly important, it is more appropriate to use human capital to improve occupational mobility network. But the quantification and multidimensional integration of human capital is still a challenge. This study also makes some explorations in quantifying human capital with multiple dimensions.

3. HC^5 Occupational mobility network

3.1. Data preparation

The data come from an open dataset provided by Zhang and his colleagues on profiles of LinkedIn users (Zhang et al., 2015) [6] . The profiles record the occupations where users have worked. Occupations discussed in occupational mobility network are the jobs listed in LinkedIn profiles. However, these occupations are filled in by users without uniform standards. Besides, some occupations are too obscure and lack representativeness due to their limited quantity. Therefore, we focus on common occupations that appear more frequently in the dataset and standardize occupation names through appropriate matching rules. Additionally, we map standardized LinkedIn occupations to official occupational classification for more information.

Profile Dataset. LinkedIn is one of the world's largest online professional networks, with millions of users sharing their learning and working experiences. The profile dataset includes 2.98 million individuals across 186 countries, 154 industries, and thousands of occupations. Each professional profile contains items such as industry, locality, education, experience, etc. Moreover, the dataset records frequently visited profiles of every profile's visitors, which are referred to as 'co-visit' profiles. Also, given that the records from 1980 to 2012 are the most complete, we only choose records from these years. After preprocessing, we select 612,463 profiles that are in English and include at least three periods of work experience. All user information has been anonymized.

Occupation standardization. Raw occupations in the dataset are expressed in various ways. For example, *Business Development Manager - WW Services Sales* and *Manager: Business Development* both mean *business development manger*. To unify occupation names, we first select some occupations as preliminary references. Occupations with a total frequency greater than 100 are designated as common occupations. Common occupations are then manually modified as the preliminary standardization. Two steps of automatic matching are then carried out for unstandardized occupations. The first step is the complete matching of occupation names. If an occupation text contains all words of a preliminary-standardized occupation name, it will be converted to the corresponding standardized occupation. If the occupation text matches multiple standardized names, it will be identified as the occupation with the highest number of matching words, i.e., the maximum complete matching. The second step is the complete matching of occupation name stems, matching occupation texts with standardized occupation names after stemming. The standardized name is the maximum complete matching occupation with more than two matching words. This process results in a total of 1720 standardized occupations.

Mapping to O*NET. The O*NET program[7] is an online platform supported by the United States Department of Labor to collect occupation-specific information to understand the nature of the workforce and provide guidance for employees. O*NET has its own taxonomy that provides a uniform classification of occupations. We map LinkedIn occupations with O*NET classifications to utilize the information in O*NET. O*NET classifications are more coarse-grained than LinkedIn occupations, and multiple LinkedIn occupations may correspond to a single O*NET classification. For example, LinkedIn occupations *software engineer* and *iOS developer* are both mapped to *software developer* in the O*NET taxonomy. We finally map the 1720 LinkedIn occupations to 301 O*NET classifications.

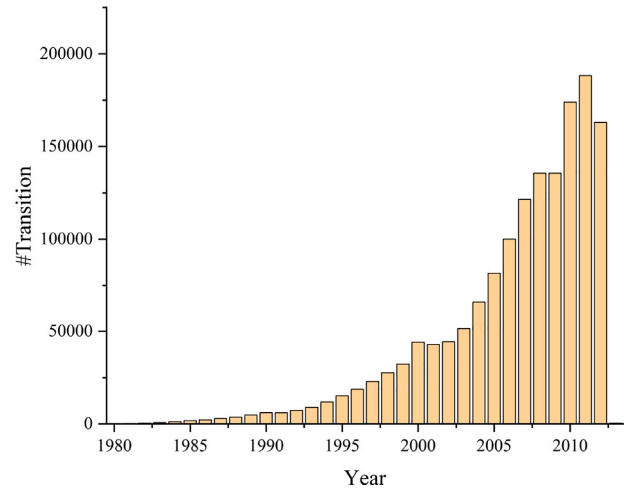


Fig. 1. Number of job transitions per year.

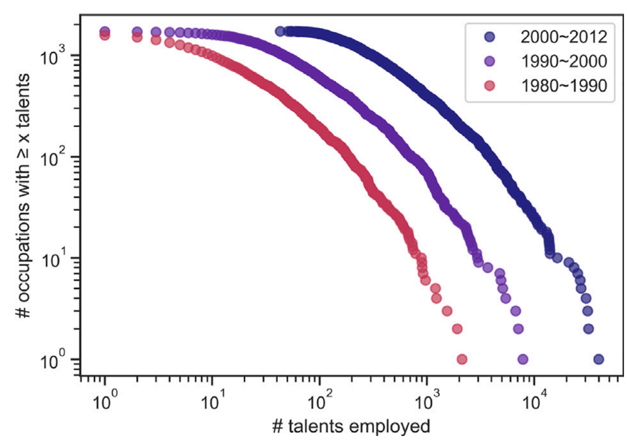


Fig. 2. Cumulative distribution of talents employment.

Description. From 1980 to 2012, there has been an upward trend in the number of career transitions per year (see Fig. 1). Each talent experiences an average of 5.36 occupations. Initially, these transitions were minimal in the 1980s, however, they sharply increased in 2004. Following this, there was a plateau in 2008 and 2009, but the growth rate returned to a high level in 2010. The number of career transitions peaked in 2011 at 188,228. Additionally, there is evidence of a Matthew effect in the number of talents, with more talents concentrated in certain occupations (see Fig. 2), such as *manager* and *consultant*. It also shows that the total employment has increased over several decades.

3.2. Architecture of network

The HC^5 occupational mobility network is a complex network model describing inter-occupational human capital flows.

To begin with, the concepts of human capital and occupational mobility should be clarified. In this study, we recognize human capital as the set of skills, knowledge, and abilities embodied in people and the

[6] www.linkedin.com.
[7] www.onetonline.org.

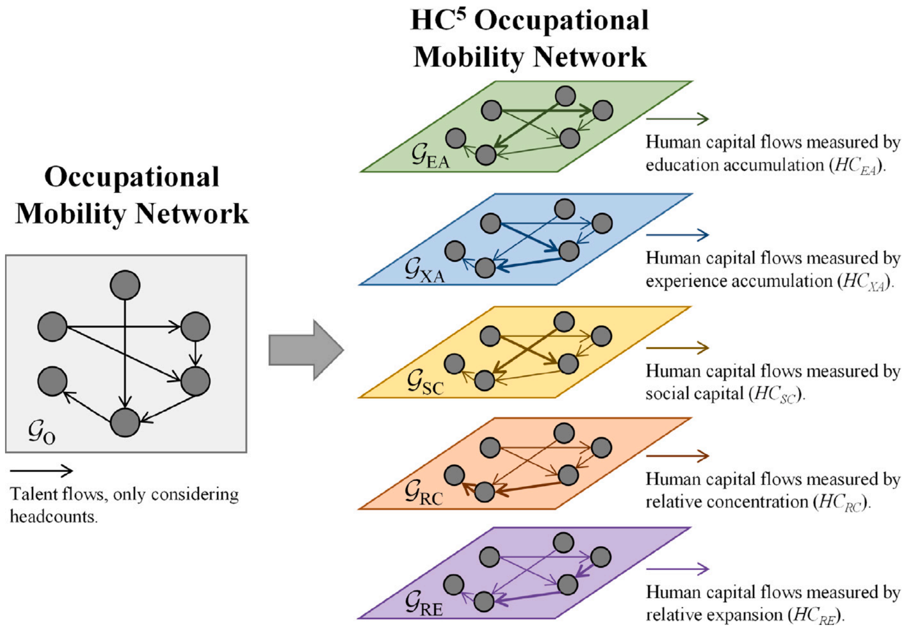


Fig. 3. Architecture of HC^5 occupational mobility network.

contribution of these attributes to productivity and creativity (Lenihan et al., 2019; Chowdhury et al., 2014). For occupational mobility, it is essentially a change in an employee's working status (Price, 1979). An employee's working status is not static, and his or her occupation changes with promotions, demotions, switching careers, and other actions. The shift of an employee from occupation u to occupation v is known as occupational mobility.

We can construct OMN $\mathcal{G}_{\mathcal{O}}$ based on the talent flow between occupations. The network's nodes represent different occupations, and edges correspond to the number of talents transition between them. OMN reveals how talent moves between occupations; however, it can only reflect the change in the headcounts. Nowadays, the value of talents is not only reflected in quantity, but more importantly in the human capital they represent.

Therefore, this study improves OMN from five dimensions of human capital measurements (HC^5), including education accumulation (HC_{EA}), experience accumulation (HC_{XA}), social capital (HC_{SC}), relative concentration (HC_{RC}), and relative expansion (HC_{RE}). These measurements can model occupational mobility based on multiple human capital flows. Specifically, we quantify human capital based on each of these five dimensions and use them to modify the edge weights to construct the HC^5 occupational mobility network. Accordingly, the HC^5 network contains five sub-networks: $\mathcal{G}_{EA}$, capturing education accumulation; $\mathcal{G}_{XA}$, capturing experience accumulation; $\mathcal{G}_{SC}$, capturing social capital; $\mathcal{G}_{RC}$, capturing relative concentration; and $\mathcal{G}_{RE}$, capturing relative expansion. The overall architecture is shown in Fig. 3.

Among the above five measurements of human capital, HC_{EA}, HC_{XA}, and HC_{SC} are based on the capabilities and resources carried by individuals. Meanwhile, HC_{RC} and HC_{RE} are more macro-level measurements to quantify human capital according to the overall trend of talent flows. We will detail the calculation of each HC^5 measurement in Section 3.4.

3.3. Occupational mobility network

The study constructs an occupational mobility network (OMN) to model talent flows between occupations. In OMN, $\mathcal{G}_{\mathcal{O}}(\mathcal{V}, \mathcal{E}, \mathcal{W})$, each node $v \in \mathcal{V}$ represents an occupation, and each edge $e(u, v) \in \mathcal{E}$ describes the transition from occupation $u \in \mathcal{V}$ to occupation $v \in \mathcal{V}$ in time t with weight $w^t(u, v) \in \mathcal{W}$. In the general case, weight $w^t(u, v)$ is the number of talents changing their jobs from u to v at year t.

Fig. 4 illustrates the evolution of OMN from 1980 to 2012, with node colors representing the main types of occupations in the O*NET taxonomy. Over the past 30 years, the number of common occupations has increased from 1453 in the 1980s to 1720 in the 2000s, indicating the emergence of new occupations. Meanwhile, talent exchanges between occupations have become more frequent as the number of edges increases from nearly 10,000 to above 200,000. In the 2010s, the number of career transitions in the first three years alone reached 180,000. The network has become closer with a larger density and a lower average shortest path length, indicating that the career paths are more diversified. The change in assortativity also suggests that occupation mobility of the same O*NET type are becoming weaker. We also use Sankey diagram to examine changes in occupational mobility between different types (see Fig. 5). Occupational mobility was low between 1980 and 1990, but each decade witnessed a large influx of newcomers to the labor market. Moreover, the high acceptance of job hopping by the new generation of workers led to an increasingly fierce trend of occupational mobility (Lyons et al., 2012). Meanwhile, although the largest destinations for workers when changing jobs are still the same types of occupations, the choices and routes vary. There is a trend of talent integration from different professions.

However, OMN only reflects changes in the number of personnel. To further reveal the human capital flows behind the talent flows, OMN needs to be improved using the HC^5 measurements.

3.4. HC^5 measurements

3.4.1. HC_{EA}: Modeling education accumulation

Human capital is closely related to the education level of talents. We classify a person p's highest education level $edu(p)$ into four levels and assign different scores:

$$edu(p) = \begin{cases} 1, & p\text{'s highest education level is } other \\ 2, & p\text{'s highest education level is } bachelor \\ 3, & p\text{'s highest education level is } master \\ 4, & p\text{'s highest education level is } doctor \end{cases} \tag{1}$$

Then, regarding the logistic skill gain model (Yelle, 1979), the sigmoid function is used to describe the accumulation of human capital

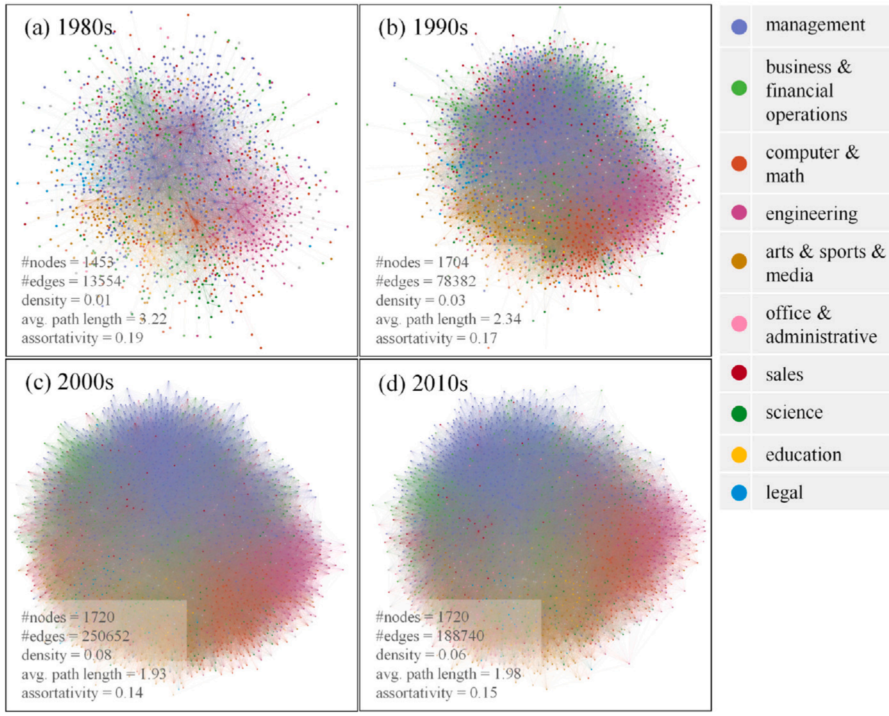


Fig. 4. Evolution of OMN over four decades.

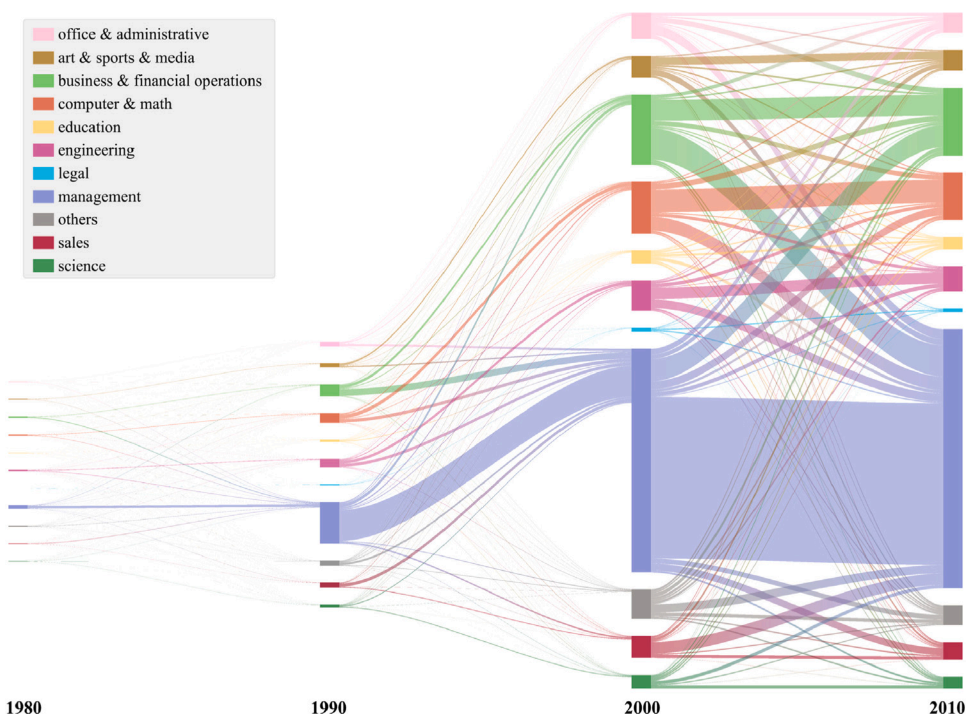


Fig. 5. Sankey diagram of occupational mobility.

obtained through education with the improvement of education level:

$$HC_{EA}(p) = \left(1 + \exp\left[-\frac{edu(p)}{\alpha}\right]\right)^{-1} \tag{2}$$

where α is the control parameter of the curve's steepness.

We then use HC_{EA} to improve the weights of edges and transform $\mathcal{G}_{\Theta}$ into $\mathcal{G}_{EA}(\mathcal{V}, \mathcal{E}, \mathcal{W}_{EA})$ describing education accumulation:

$$w^t_{EA}(u,v) = \sum_{p:u\rightarrow v|t} HC_{EA}(p) \tag{3}$$

where $w^t_{EA}(u,v) \in \mathcal{W}_{EA}$ represents the human capital based on education accumulation in the flow of talents from occupation u to occupation v in year t.

3.4.2. HC_{XA}: Modeling experience accumulation

The accumulation of work experience of talents is also the accumulation of knowledge and skills, which are important components of human capital. Longer careers lead to more accumulation of experience and thus enhance human capital value. We also refer to logistic skill gain model to quantify the experience of each talent (Safavi et al., 2018; Yelle, 1979). In detail, experience-related human capital of employee p at year t is modeled as a sigmoid function of his or her career length up to that year $l(p,t)$:

$$HC^t_{XA}(p) = \left(1 + \exp\left[-\frac{l(p,t) - \overline{l(t)}}{\gamma}\right]\right)^{-1} \tag{4}$$

where $\overline{l(t)}$ is the average career length of employees at year t. γ is the parameter controlling the steepness of curve.

The directed edges of $\mathcal{G}_{\mathcal{O}}$ can be transformed into concentrated flows of experienced people with corresponding human capital:

$$w^t_{XA}(u,v) = \sum_{p:u\to v|t} HC^t_{XA}(p) \tag{5}$$

We then obtain $\mathcal{G}_{XA}(\mathcal{V},\mathcal{E},\mathcal{W}_{XA})$ characterizing human capital flows for the experience accumulation dimension, and $w^t_{XA}(u,v) \in \mathcal{W}_{XA}$.

3.4.3. HC_{SC}: Modeling social capital

Social capital can be regarded as an essential resource that talent possess, and can be considered as a soft skill. Personal social networks and communities have been confirmed as determinants of outcomes ranging from education to work (Chetty et al., 2022). This research uses co-visit relations referred in the profile dataset to construct a social network. Each node represents a user p, while each edge is a co-visit tie meaning that the two users are more likely to be browsed together. More details can be found in Ref. Liu et al. (2021b). The degree of node p, $degree_{SN}(p)$, is used to measure human capital from the perspective of social capital:

$$HC_{SC}(p) = degree_{SN}(p) \tag{6}$$

To capture the influence of social capital, we transform $\mathcal{G}_{\mathcal{O}}$ to $\mathcal{G}_{SC}(\mathcal{V},\mathcal{E},\mathcal{W}_{SC})$ by resetting the weights with HC_{SC}:

$$w^t_{SC}(u,v) = \sum_{p:u\to v|t} HC_{SC}(p) \tag{7}$$

where $w^t_{SC}(u,v) \in \mathcal{W}_{SC}$.

3.4.4. HC_{RC}: Modeling relative concentration

Relative concentration replaces the number of people moving with the proportion of people moving. Relatively concentrated talent flows tend to occur between occupations with similar duties or at adjacent ranks in occupation series (Liu et al., 2021a), like *junior software developer* and *software developer*. It can exclude the influence of the absolute number of personnel. To a certain extent, it indicates the aggregation of highly specialized knowledge, skills, and resources. It is likely that a concentration of talent will be seen in highly specialized occupations. Thus, relative concentration can capture the flow of highly specialized human capital. It is defined as:

$$HC^t_{RC}(u,v) = \frac{mov^t_{(u,v)}}{\sum_{z\in\mathcal{V}} mov^t_{(u,z)}} \tag{8}$$

where $mov^t_{(u,v)}$ is the number of people moving from occupation u to occupation v in year t.

The weights of $\mathcal{G}_{\mathcal{O}}$ are changed according to HC_{RC}:

$$w^t_{RC}(u,v) = HC^t_{RC}(u,v) \tag{9}$$

The OMN describing relative concentration $\mathcal{G}_{\mathcal{O}}$ to $\mathcal{G}_{RC}(\mathcal{V},\mathcal{E},\mathcal{W}_{RC})$ is thus constructed, and $w^t_{RC}(u,v) \in \mathcal{W}_{RC}$.

3.4.5. HC_{RE}: Modeling relative expansion

Relative expansion depicts the increase in occupational size, assuming that high-value human capital can help occupations develop rapidly. As human capital contributes to incomes, it measures human capital through the final promotion effect of talent on occupational size:

$$HC^t_{RE}(v) = \frac{\log(\sum_{u\in\mathcal{V}} mov^t_{(u,v)}) - \log(\sum_{z\in\mathcal{V}} mov^t_{(v,z)})}{\log(M^t_v) + 1} \tag{10}$$

where M^t_v is the number of people in occupation v in year t.

Drawing on models of growth rates in ecology and economics (Safavi et al., 2018; Hoffmann and Poorter, 2002; Marti et al., 2021), the relative growth of occupation v in year t is modeled as the difference between the logarithm of inflow and outflow of personnel. The difference is standardized by the number of talents working in v in year t before transitions.

We apply normalized HC_{RE} to improve the weights and transform $\mathcal{G}_{\mathcal{O}}$ into $\mathcal{G}_{RE}(\mathcal{V},\mathcal{E},\mathcal{W}_{RE})$ describing relative expansion:

$$w^t_{RE}(u,v) = w^t(u,v) \cdot \frac{\exp\left[\beta \cdot HC^t_{RE}(v)\right]}{\max_{v,t}\left(\exp\left[\beta \cdot HC^t_{RE}(v)\right]\right)} \tag{11}$$

where β is a parameter to weight the curve.

3.5. Human capital flows in HC^5 network

We use HC^5 measurements to reconstruct the edges in OMN and form the HC^5 occupational mobility network. Fig. 6 shows the typical HC^5 edges. We extract the most important 100 edges and their associated nodes under each HC^5 measurement and highlight the dominant dimension.

HC_{EA} is mainly notable in IT-related occupations, including the flows from *software engineer* to *senior software engineer*, and from *project leader* to *project manager*, suggesting the mobility of highly educated talents in high-tech careers. HC_{XA} is especially pertinent to managerial occupations, such as the progression from *director* to *senior director* to *vice president*, which is in line with the fact that senior executives are usually filled by employees with greater experience and seniority. HC_{SC} flows are conspicuous in sales-related occupations and occupations at the outset of a student's career, demonstrating the necessity of social resources for these groups. Also, it is found that HC_{SC} is the main dimension connecting various types of occupations ranging from *software engineer* to *brand manager*. Besides, HC_{RC} mainly connects leaf nodes in the network. It tightly links specialized occupational clusters and emphasizes the promotion channels for small-scale occupations. Edges dominated by HC_{RE} are observed around *sales manager* and *account manager*, reflecting the expansion of the sales industry and the rapid circulation of related human capital.

4. Career growth potential assessment model with human capital ensemble

4.1. Model framework

The study aims to assess the growth potential of careers using the characteristics of occupational mobility and human capital flows. It can be achieved by solving the following problem: Given an occupation v, forecast the outlook of v based on features extracted from HC^5 network.

The assessment of career growth potential is then formalized as a binary classification problem with multi-view data. Suppose we have five dimensional feature set $X = [X_1, X_2, X_3, X_4, X_5]$, where $X_1 = [x^1_1, \dots, x^1_i \dots, x^1_N]$, $\dots$, $X_5 = [x^5_1, \dots, x^5_i \dots, x^5_N]$. We denote the labels as $Y = [y_1, \dots, y_i \dots y_N]$. A classification model aims to learn a mapping function f: $x^{(1,2,3,4,5)}_i \mapsto \hat{y}_i$ for all i and minimize the difference between $\hat{y}_i$ and y_i. Specifically, labels are provided by O*NET in the *Bright Outlook Occupations* list. Occupations mentioned in the list are labeled as '*bright*', while the rest are labeled as '*non-bright*'. The proposed classification model, CPACE, predicts occupation outlooks and can also give the probability of each occupation being classified as either *bright* or *non-bright*. Ultimately, the probability of an occupation being

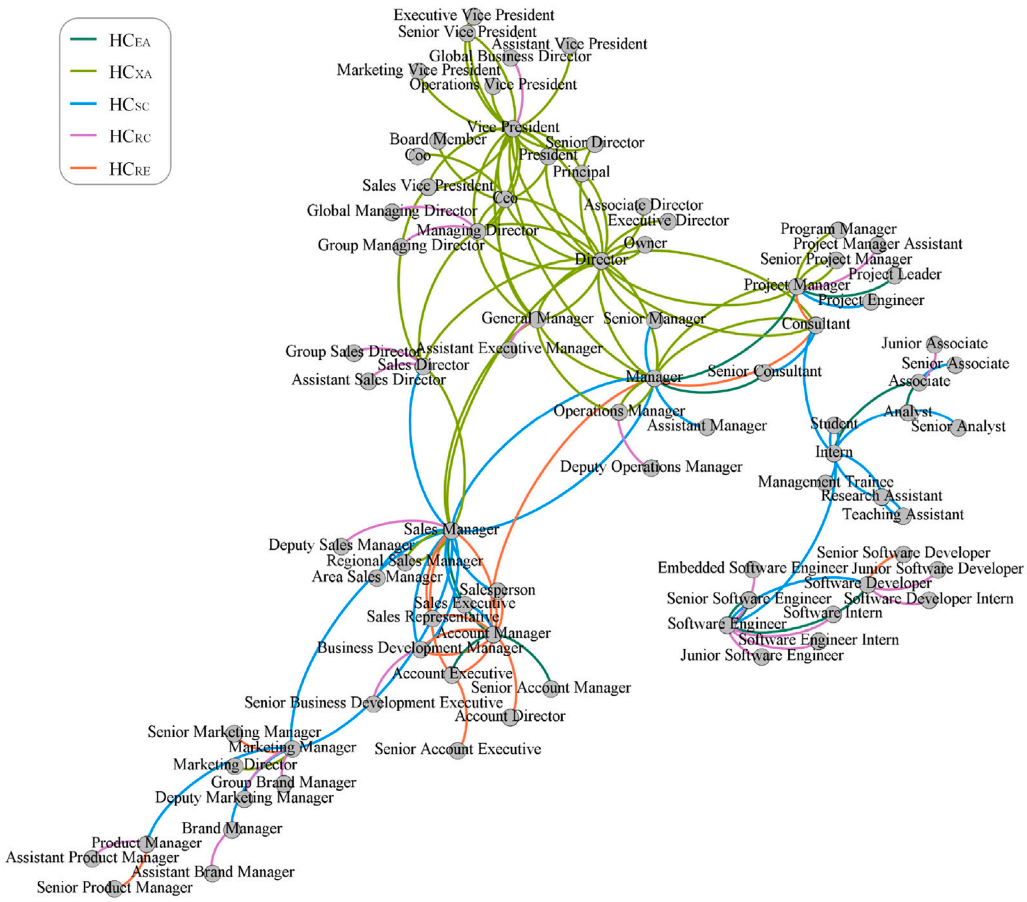


Fig. 6. Typical HC^5 Edges.

classified as *bright* is defined as the quantitative value of its growth potential.

Meanwhile, we integrate multidimensional human capital information by means of ensemble learning. It should be noted that the HC^5 network contains five sub-networks, each of which reflects unique inter-occupational human capital flows. Therefore, an appropriate method is required to integrate human capital across variables. Simple combination of addition or subtraction lacks scientific basis and may lead to the loss of important information. The direct concatenation of features of each human capital variable may cause interference. The ensemble method can enhance the prediction of the model by fusing valuable knowledge while preserving the uniqueness of each sub-network.

Thus, based on HC^5 network, the study proposes a model for assessing career growth potential that integrates multiple human capital measurements through ensemble learning, i.e., the **c**areer growth **p**otential **a**ssessment model with human **c**apital **e**nsemble (CPACE).

The ensemble framework of the CPACE model is shown in Fig. 7. It mainly consists of six modules, including five prediction modules and one ensemble module. The five prediction modules are the five independent classifiers of **c**areer growth **p**otential **a**ssessment (CPA). The ensemble module combines the results of each prediction module and finally achieves the integration of human capital in all dimensions.

4.2. Modules

Prediction module. Prediction modules are used to obtain the preliminary prediction results of career growth potential based on the features from HC^5 network. The five classifiers CPA_{EA}, CPA_{XA}, CPA_{SC}, CPA_{RC}, and CPA_{RE} respectively classify occupation outlook according to features from the five sub-networks in HC^5 network. They use two different base-classifiers to learn HC^5 features, get the prediction probability (PP) of each occupation being estimated as $bright$, and then input the probability to a voter. The output of voter is the prediction probability of the label $bright$ calculated by prediction modules. For the base-classifiers, we select a generalized linear classifier, support vector machine (SVM) (Cortes and Vapnik, 1995), and an ensemble learning classifier, random forest (RF) (Breiman, 2001), due to their superior performance in this problem. This also avoids over-fitting and enhances the generalization ability of the model. Voter uses the mechanism of soft voting, which calculates the average prediction probability of each label for the two base-classifiers as the predictive results.

Ensemble module. The ensemble module combines and further learns the predictive results obtained by prediction modules. It adopts the idea of stacking which is a hierarchical structure. Taking two-level stacking as an example, multiple classifiers are applied to prediction in the first level, and the predicted results are fed as input to the second-level classifier. The output of the second-level classifier is then taken as the final result of the model. In CPACE, first-level classifiers are the five sub-classifiers listed above. Next, in the second layer, we choose logistic regression (LR) as the algorithm for fusing sub-classifiers (Cox, 1958). In detail, the predicted probabilities of CPAs for the label *bright* are used as inputs to LR. Thus, the model integrates HC^5 measurements to learn the final classification probability of occupation outlook, i.e., career growth potential.

4.3. Feature set

The feature set is the input of CPACE model. Each CPA classifier uses different HC^5 features. A total of 13 features are constructed to describe every occupation. The features consist of two parts: general features and HC^5 features. General features describe the basic situation of occupations, including STEM category, the proportion of the tertiary

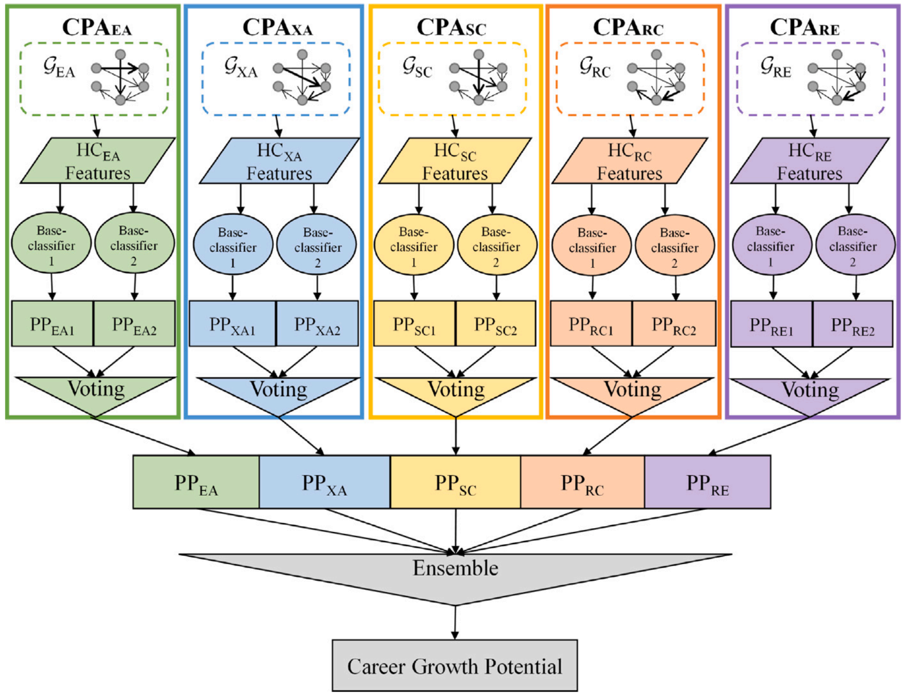


Fig. 7. Framework of CPACE.

Table 1
Features for CPACE.

Feature	Notation	Type
General features		
STEM category	$STEM$	Binary
Proportion of the tertiary industries	IND	Numeric
Proportion of developed localities	LOC	Numeric
Education level	ELV	Numeric
Education requirements	ERQ	Categorical
HC^5 features		
Average neighborhood degree	NDG	Numeric
Constraint	CON	Numeric
PageRank	PR	Numeric
Reverse PageRank	RPR	Numeric
PR-RPR	$PRPR$	Numeric
Community	CM	Categorical
Similar influx	SIN	Numeric
Similar outflux	SOT	Numeric

industries, the proportion of developed localities, education level, and education requirements. HC^5 features are designed based on the HC^5 network, which reflects the characteristics of occupational mobility and human capital flows. Average neighborhood degree, PageRank (PR), Reverse PageRank (RPR), PR-RPR, constraint, community, similar influx, and similar outflux are the HC^5 features used in CPACE. HC^5 features vary with different human capital dimensions. Their details are listed in Table 1.

STEM category. STEM is an acronym for the four subjects of science, technology, engineering, and mathematics. An occupation is classified as a STEM occupation based on whether it is included in O*NET's list of STEM-related occupations.

Proportion of the tertiary industry. The profile dataset includes the industry of users. According to *the clarification of three industries*,[8] these industries are divided into the tertiary industry and other industries. The feature is the proportion of tertiary industry workers in each occupation, which indicates the type of work for the occupation. See Ref. Liu et al. (2021b) for more details on these industries.

Proportion of developed localities. The profile dataset includes the locality of users. According to *world economic outlook database*,[9] these countries are divided into developed localities and other localities. The feature is the proportion of talents who have worked in the occupation belongs to developed localities, indicating the economic state of the occupation. See Ref. Liu et al. (2021b) for more details on these localities.

Education level. The education level of an occupation is the average education level of talents who have worked in the occupation. The levels are calculated according to Eq. (1).

Education requirements. The information comes from the Occupational Employment Statistics (OES) survey administered by the United States Bureau of Labor Statistics[10] which records the typical education level required to enter each occupation.

Average neighborhood degree. It is the average degree of neighborhoods of each node, reflecting the reachable resources of occupations. The average neighborhood degree of an occupation v in HC^5 network is:

$$NDG_v^t = \frac{1}{wdegree_v} \sum_{u \in Nb(v)} w_*^t(u,v) \times degree_u \tag{12}$$

where $Nb(v)$ are the neighbors of v, $wdegree_v$ is the weighted degree of v (without direction), $degree_u$ is the degree of node u which belongs to $Nb(v)$, and $w_*^t(u,v)$ is the weight of the edge that links u and v.

PageRank. It is the importance of occupation in occupational mobility estimated by the amount and quality of inflow of human capital. The PageRank algorithm is an iterative process that iterates until it converges (Page et al., 1999). Firstly, the algorithm assigns the initial

[8] www.stats.gov.cn/tjsj/tjbz/201301/t20130114_8675.html.

[9] www.imf.org/en/Publications/WEO/weo-database/2021/April/select-country-group.

[10] www.bls.gov/oes/2019/may/oes_nat.htm.

Table 2
Paramater settings.

Model	Parameter	Range	Retained Value
RF	criterion, θ_1 n_estimators, θ_2 min_samples_split, θ_3 max_depth, θ_4	entropy, gini [50,400] [2,16] [10,50]	$RF_{EA}(\theta_1$ = entropy, $\theta_2 = 400$, $\theta_3 = 8$, $\theta_4 = 40)$; $RF_{XA}(\theta_1$ = entropy, $\theta_2 = 300$, $\theta_3 = 2$, $\theta_4 = 40)$; $RF_{SC}(\theta_1$ = entropy, $\theta_2 = 400$, $\theta_3 = 4$, $\theta_4 = 30)$; $RF_{RC}(\theta_1$ = entropy, $\theta_2 = 300$, $\theta_3 = 2$, $\theta_4 = 20)$; $RF_{RE}(\theta_1$ = entropy, $\theta_2 = 400$, $\theta_3 = 2$, $\theta_4 = 50)$.
SVM	kernel, ϑ_1 gamma, ϑ_2 C, ϑ_3	linear, rbf, poly, sigmoid [0.001,1000] [0.001,1000]	$SVM_{EA}(\vartheta_1$ = poly, $\vartheta_2 = 3.2$, $\vartheta_3 = 0.01)$; $SVM_{XA}(\vartheta_1$ = rbf, $\vartheta_2 = 10$, $\vartheta_3 = 1)$; $SVM_{SC}(\vartheta_1$ = rbf, $\vartheta_2 = 0.8$, $\vartheta_3 = 10)$; $SVM_{RC}(\vartheta_1$ = rbf, $\vartheta_2 = 0.8$, $\vartheta_3 = 10)$; $SVM_{RE}(\vartheta_1$ = rbf, $\vartheta_2 = 0.4$, $\vartheta_3 = 100)$.

value $PR_v(0)$ to all occupations, and $\sum_{v=1}^{N} PR_v(0) = 1$. Then, in iteration k, the PageRank value of occupation is defined as:

$$PR_v^t(k) = \frac{1-d}{N} + d \times \sum_{u \in \mathcal{V}} \frac{w_*^t(u,v) \times PR_u^t(k-1)}{Out_u^t} \tag{13}$$

where d is the damping probability with a default value of 0.85, N is the number of occupations in the HC^5 network, and Out_u^t is the weight of all outgoing edges of occupation u.

Reverse PageRank. Contrary to PR, RPR calculates the importance of occupation by the outflow of human capital. The calculation processes of reverse PageRank are similar to that of PageRank, but all the direction of edges in the network are reversed (Bar-Yossef and Mashiach, 2008).

PR-RPR. It is the ratio of PR to RPR. It is the difference in the importance of occupation between human capital inflow and outflow, reflecting the retention ability of occupation to human capital. It is defined as:

$$PRPR_v^t = \frac{PR_v^t}{RPR_v^t + \xi} \tag{14}$$

where ξ is the smoothing factor with a small value. If $PRPR > 1$, the occupation has a strong ability to retain human capital. Otherwise, it may face serious challenges in human capital loss.

Constraint. It demonstrates the strength of an occupation's role as a connection, measuring the ability to control and obtain resources. Constraint is associated with structural holes (Burt, 1995). In OMN, if an occupation is connected to two occupations that have no talent flow between them, the occupation then occupies a structural hole. Constraint is one of the classic indicators to measure structural holes. It evaluates the restrictions of an occupation caused by its lack of access to structural holes. The lower the constraint of an occupation, the more likely it is to occupy a structural hole.

Community. Communities are the subsets of densely interconnected nodes in the network, showing the occupational clusters formed by human capital flows. In the study, we utilized Louvain algorithm to detect occupation communities (Blondel et al., 2008). Louvain algorithm is a community discovery algorithm based on modularity. It has been applied in the network analysis of labor mobility and achieved desirable performance (Park et al., 2019).

Similar influx. It refers to the proportion of human capital from occupations with the same O*NET type as the target occupation in the total influx of the target occupation.

Similar outflux. It is the proportion of human capital leaving for occupations with the same O*NET type as the source occupation in the total outflux of the source occupation.

The correlation between features and occupation outlook are analyzed in Appendix. IND, SOT, SIN, NDG, PR, $PRPR$ are positively correlated to occupation outlook, while $STEM$, ELV, CON are negatively correlated to occupation outlook. Also, negative correlations are found in GC_RC for both *SIN* and *RPR*. Different ERQ and CM categories have different correlations with outlook. In summary, occupations with higher node importance are more likely to be assessed as *bright*. Education affects career prospects, but neither too high nor too low is good. The type of an occupation is highly correlated with its future development.

5. Experiments

This section evaluates the effectiveness of the HC^5 network and CPACE model through experiments. By using the CPACE model to assess career growth potential, we acquire satisfactory results. CPACE significantly improves the prediction result. Ablation experiments verify the importance of HC^5 features. Moreover, the quantification of career growth potential enables us to analyze the mechanisms of career mobility, providing valuable insights into career development.

5.1. Model performance

The experiment is carried out with the use of Python environment. We use RF and SVM for a single HC^5 dimension as baselines. Grid search and four-fold cross-validation are applied to optimize parameters. Specific parameter settings are shown in Table 2. CPAs and CPACE follow these parameters to train base-classifiers. Then, five-fold cross-validation is used for assessing growth potential of careers. Accuracy and F1 score are used to evaluate model performance. Accuracy measures a model's ability to correctly predict occupation outlooks, while F1 score measures the ability to completely and precisely find *bright* occupations.

Table 3 shows the performances of baselines, CPA classifiers, and CPACE model. They reflect the predictive ability of models to occupation outlooks under single and multiple human capital dimensions. SVM is superior to RF in HC_{XA}, HC_{SC}, HC_{RC} features. In terms of HC_{EA} and HC_{RE} features, there is little difference between RF and SVM. CPA classifiers perform better than baselines and can compensate for the bias caused by the single model. Specifically, the accuracy of $\mathrm{CPA_{EA}}$ and $\mathrm{CPA_{RC}}$ is below 0.79. $\mathrm{CPA_{EA}}$ also shows the lowest F1 score. $\mathrm{CPA_{XA}}$ and $\mathrm{CPA_{SC}}$ perform better, achieving the accuracy of more than 0.79 and the F1 score of about 0.81. $\mathrm{CPA_{RE}}$ has the highest accuracy of 0.801 and the highest F1 score of 0.818. It can be concluded that HC_{RE} has the highest predictive ability for career growth potential, with HC_{SC} and HC_{XA} following closely behind. The CPACE model, which comprehensively considers five human capital measurements, achieves the highest accuracy of 0.821 and the highest F1 score of 0.833. Compared to the lowest baseline, $\mathrm{RF_{EA}}$, its accuracy improves by 4.8% and its F1 score improves by 4.3%.

To further verify the effectiveness of the ensemble framework, we also design a comparison model $\mathrm{CPA_{(w/o)\ EM}}$. $CPA_{(w/o)EM}$ does not use the ensemble module to fuse HC^5 features of each dimension. It directly uses all HC^5 features as input for training two base-classifiers, and finally gets the result by a voter. In addition, given that the feature set comes from multiple HC^5 dimensions, it is a multi-view dataset. We can use multi-view learning approach to integrate variables for classification (Zhao et al., 2017; Sun, 2013). We select the classical multi-view learning models MCCA (Tenenhaus and Tenenhaus, 2011; Kettenring, 1971) and SVM-2K (Farquhar et al., 2005). The parameter adjustment for $\mathrm{CPA_{(w/o)\ EM}}$ is the same as other CPAs. Kernel functions chosen for SVM-2K correspond to the SVM models under each human capital dimension. In MCCA, the number of canonical components is 25, and regularization adopts the Ledoit-Wolf covariance estimate. The

Table 3
Model performance.

Feature	Model	Accuracy	F1 score
HC_{EA}	RF_{EA}	0.773	0.790
	SVM_{EA}	0.772	0.790
	CPA_{EA}	**0.780**	**0.801**
HC_{XA}	RF_{XA}	0.765	0.784
	SVM_{XA}	0.782	0.796
	CPA_{XA}	**0.793**	**0.809**
HC_{SC}	RF_{SC}	0.779	0.793
	SVM_{SC}	0.790	0.811
	CPA_{SC}	**0.795**	**0.811**
HC_{RC}	RF_{RC}	0.758	0.773
	SVM_{RC}	0.779	0.801
	CPA_{RC}	**0.785**	**0.807**
HC_{RE}	RF_{RE}	0.786	0.802
	SVM_{RE}	0.789	0.803
	CPA_{RE}	**0.801**	**0.818**
HC^5	SVM-2K	0.794	0.812
	MCCA	0.808	0.814
	$CPA_{(w/o)\ EM}$	0.802	0.820
	CPACE	**0.821**	**0.833**

evidence in Table 3 suggests that models based on multiple HC^5 dimensions all outperform baselines. However, their ability to integrate and utilize features varies. SVM-2K has the worst performance, followed by $CPA_{(w/o)\ EM}$ and MCCA. The CPACE proposed in this paper achieves the best result, which demonstrates that the ensemble framework is more appropriate for the integration of multidimensional human capital features.

Next, we run an ablation experiment to explore which components are the most important when ensemble. The result is shown in Table 4. CPA_{RC} and CPA_{RE}, which predict occupation outlook by measuring human capital at the macro level of talent flows, demonstrate more important roles when combining with other classifiers. In particular, CPA_{RC} is not ideal enough when performing prediction tasks alone, but its removal causes a significant loss in the accuracy and F1 score, indicating that the component is highly effective in integration. These two classifiers are followed by CPA_{SC}, CPA_{XA}, and CPA_{EA}. Although the predictive power of each CPA varies, eliminating any of them leads to a decline in the performance. Therefore, the five dimensions of human capital measurements are effective in addressing this problem.

5.2. Feature importance

Although CPACE achieves high prediction accuracy, the interpretability of the results is also important in machine learning (Li et al., 2022). Thus, feature importance analysis is used to further understand the model.

Random forest can calculate the importance of individual features, which we use to evaluate the efficacy of the features. Fig. 8 shows the importance of some features in the RF classifier of CPACE. *NDG*, *ELV*, and *IND* are the three most critical features. For communities, computer-related and HRM-related communities have greater impact on the prediction. This is also consistent with the correlation analysis in Section 4.3.

Since random forest is merely one of the components of CPACE, this study then explores the importance of each feature through ablation experiments. The results are shown in Table 5. All features have positive effects on predicting occupation outlook, as the accuracies decrease after elimination. The greatest decrease is observed after eliminating *CM* and *ERQ*. In addition to the high correlation with career growth potential, it may also be because categorical features create multiple sub-features after One-Hot encoding. Besides, although the feature importance of *STEM* in RF is not substantial, the feature's contribution to accuracy in ablation experiments is obvious. *STEM* may

Table 4
Ablation experiments of sub-classifiers.

Method	Accuracy	F1 score
(w/o) CPA_{EA}	0.815	0.828
(w/o) CPA_{XA}	0.814	0.830
(w/o) CPA_{SC}	0.813	0.827
(w/o) CPA_{RC}	0.808	0.823
(w/o) CPA_{RE}	0.806	0.823

Table 5
Ablation experiments of features.

Feature	Accuracy	Feature set	Accuracy
(w/o) $STEM$	0.796	(w/o) PR	0.814
(w/o) IND	0.804	(w/o) RPR	0.816
(w/o) LOC	0.812	(w/o) $PRPR$	0.815
(w/o) ELV	0.797	(w/o) CM	0.740
(w/o) ERQ	0.786	(w/o) SOT	0.808
(w/o) NDG	0.805	(w/o) SIN	0.814
(w/o) CON	0.815	(w/o) HC^5	0.657

play a significant role in SVM. It has been noted that features such as *ELV*, *NDG*, and *SOT* have strong influences on predictive power. Furthermore, the exclusion of all HC^5 features diminishes the accuracy to only 0.657, which validates that HC^5 features are appropriate and effective.

With the findings in ablation experiments and Section 4.3, we summarize the key factors that influence career prospects. The correlation between educational factors and career growth potential is evident, but a higher level of education does not necessarily equate to better outcomes. Careers dominated by undergraduates have better prospects, while the prospects of careers with either excessive or insufficient demands for educated people are impeded (see Appendix). Secondly, the type or industry of a career heavily influences career prospects. Additionally, the dynamics of human capital flow also plays a significant role. For an occupation, the resources available surrounding it are more critical than its importance in the network. The neighborhoods of $bright$ occupations bring together a large amount of human capital. Moreover, the 'way out' of a career is more crucial than its 'way in', and employees in high-potential careers are more likely to pursue their career paths in similar fields.

5.3. Career growth potential analysis

Using CPACE's predictions, we conduct an in-depth analysis of career growth potential.

We discover that senior occupations of management, as well as typical occupations in popular and emerging industries tend to have better prospects. Fig. 9 shows career growth potential in the form of word cloud. *Bright* occupations are represented in orange, while *non-bright* occupations are depicted in purple. The darker the color, the greater the level of *bright/non-bright*. The font size is proportional to the occupational size. Career growth potential is higher for *general management*, *sales*, and *computer* occupations, and lower for *human resource management* and *research* occupations.

We then list the 15 most promising occupations assessed by CPACE in Table 6. Most of them are related to *information technology* and *business consulting*. In 2012, along with the impact of big data and internet technology, many enterprises carried out digital transformations to reshape their business and management modes (CMSWIRE, 2011). IBM's 2012 annual report said that IBM is reinventing the enterprise with intelligent technology in the new era of smarter computing. At the same time, a large number of customers seek IBM's involvement and support for information technologies. This transformation had the potential to fuel the biggest wave of business technology (IBM, 2011). Some talent demand reports also prove the reasonableness of CPACE's predictions. McKinsey noted that a new era of work had

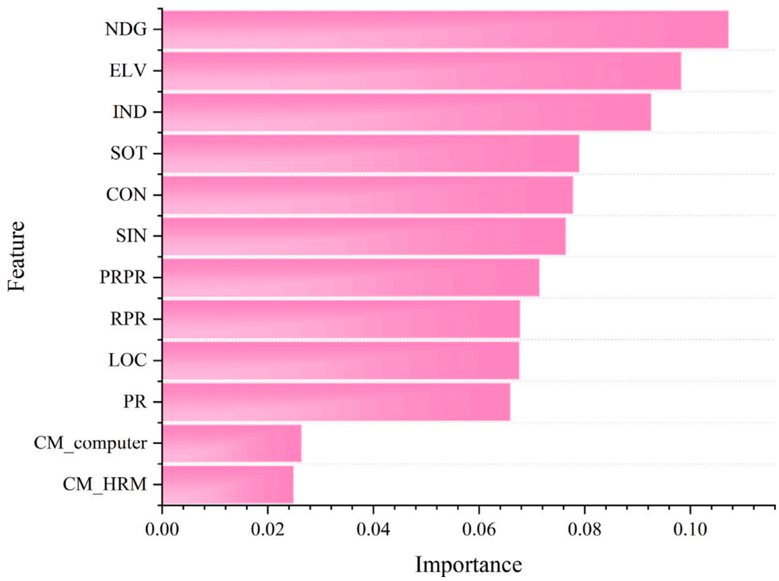


Fig. 8. Feature importance in random forest.

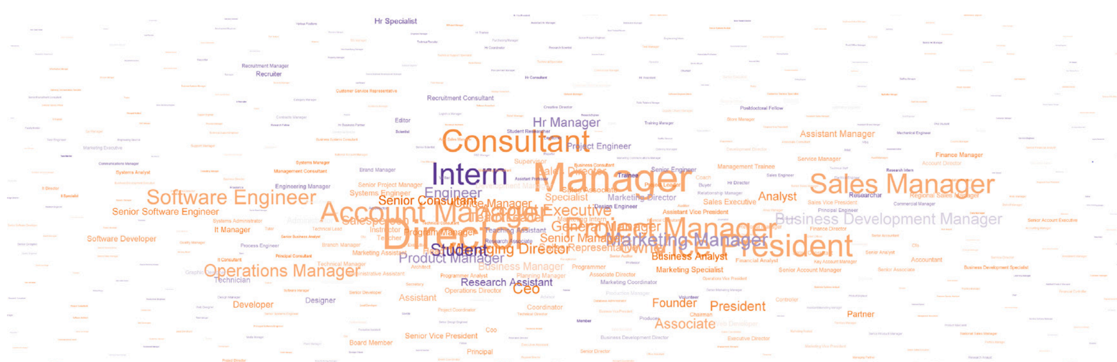


Fig. 9. Word cloud of career growth potential.

dawned (Company, 2011). Knowledge workers and interactive workers were indispensable for companies and countries seeking to prevail in the competition. Many lower-skilled workers (e.g. *construction*, *administrative support*) were laid off, but jobs for higher-skilled professionals (e.g. *business*, *computer*) continued to grow. Moreover, on websites like Computerworld[11] and LinkedIn,[12] experts believed that skills such as data mining, software development, business intelligence, online marketing, etc., were among the most in-demand. These are precisely the kinds of skills that high-potential occupations evaluated by CPACE ought to possess.

Subsequently, we further analyzed the pattern of talent mobility based on career growth potential. Fig. 10(a) shows the potential of talents' first, second, and last occupations. Fig. 10(b) shows the potential of talents' previous, current, and next occupations. The points in the box represent the average. The previous/first occupation has the least potential, while the current/second occupation has moderate potential. The next/last occupation has the highest potential. The gap of potential between the first and last occupation is the most distinct. As a whole, the potential of talents' occupations increases gradually as their career steps forward. This implies that when changing careers, talents tend to choose occupations that may have better prospects.

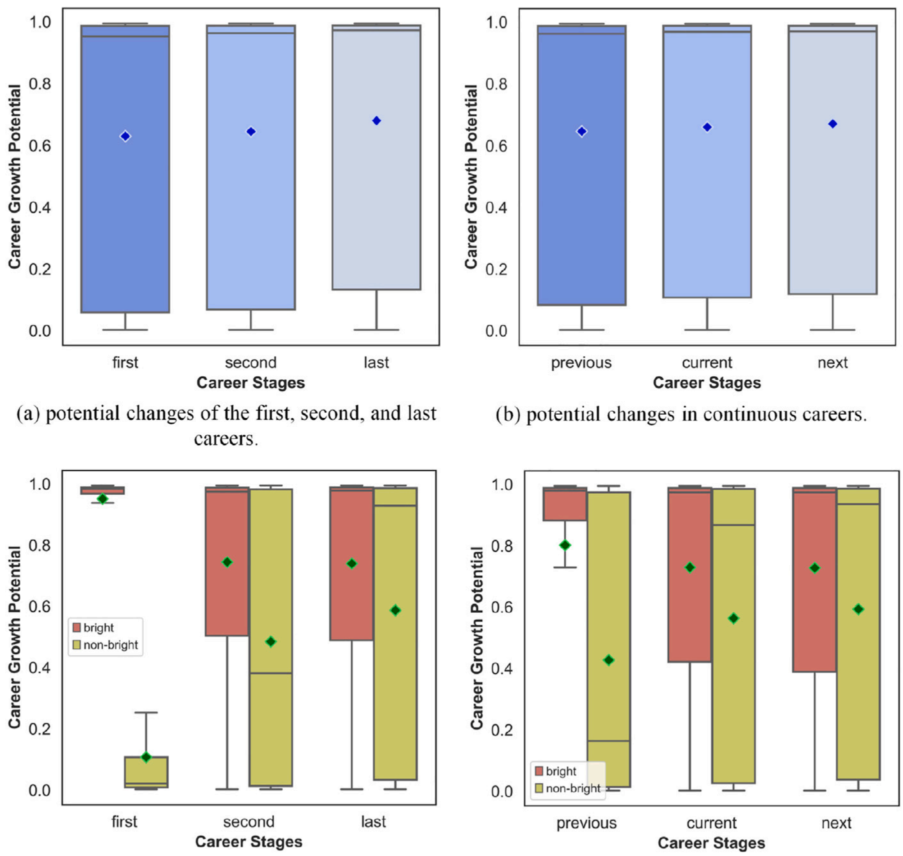
Table 6
The 15 most promising occupations.

Rank	Occupation	Potential
1	Software Manager	0.99498
2	Management Consultant	0.99491
3	Principal Consultant	0.99480
4	Systems Architect	0.99467
5	Solution Designer	0.99449
6	Business Consultant	0.99425
7	Systems Manager	0.99419
8	Lead Developer	0.99415
9	Application Manager	0.99406
10	Project Management Consultant	0.99390
11	Technical Architect	0.99381
12	Software Design Engineer	0.99377
13	Business Analyst	0.99371
14	Solution Architect	0.99363
15	Enterprise Architect	0.99358

We then compare the changes in career growth potential across career stages for groups with *bright* and *non-bright* initial jobs, respectively. In Fig. 10(c), the potential of first job for *bright* group is concentrated in the range of 0.95 to 1.0, while that of *non-bright* group is concentrated in the range of 0 to 0.1. However, after career

[11] www.computerworld.com/article/2732426/career-watch--the-most-in-demand-skills-of-2012.html.

[12] blog.linkedin.com/2014/12/17/the-25-hottest-skills-that-got-people-hired-in-2014.



(a) potential changes of the first, second, and last careers.

(b) potential changes in continuous careers.

(c) potential changes of the first, second, and last careers within different groups.

(d) potential changes in continuous careers within different groups.

Fig. 10. Changes in career growth potential at different career stages.

transitions, there is a decline in career growth potential of the *bright* group. In contrast, the potential of the *non-bright* group increases substantially, reaching an average of approximately 0.5. Eventually, the career growth potential of the *bright* group stabilized at an average of about 0.7. Although half of the talents still engage in high-potential occupations, a large number of individuals have seen decreases in their career growth potential. But basically, they are able to maintain work in *bright* occupations. Conversely, the *non-bright* group continues to progress in terms of career growth potential, culminating in an average of 0.6. A considerable proportion of them are employed in high-potential occupations. Besides, the potential distributions and trends in the continuous career stages of both groups (see Fig. 10(d)) are similar to those in Fig. 10(c), with changes being more moderate. The career growth potential of the *bright* group is consistently superior to *non-bright* group. The increase of the potential in *non-bright* group surpasses the decrease in *bright* group. What is the same is that most of the talents in both groups end up working in $bright$ occupations ($potential > 0.5$).

As a result, we conclude that: (1) Talents gravitate towards careers with higher growth potential. (2) It is challenging for talents to stay employed in careers with the brightest prospects all the time, but engaging in careers with relatively high potential can be an achievable endeavor. (3) Talents who begin their careers in low-potential occupations will exert greater effort to find more promising occupations.

6. Discussion and conclusion

In this study, with the use of occupational mobility network and machine learning, we propose an innovative approach based on multi-dimensional human capital quantification and integration to evaluate growth potential of career. Specifically, we exploit profile and occupation data available on the internet to construct HC^5 occupational mobility network, capturing human capital flows between occupations. In the HC^5 network, human capital is measured by five dimensions: education, experience, social capital, expansion of occupational size, and concentration of talent flows. Based on this, we design an ensemble framework for HC^5 features of to predict career prospects and quantify career growth potential. The experimental results validate the effectiveness of our method to measure human capital and assess growth potential of careers. Analysis of feature importance reveals that educational level of practitioners, resources surrounding the career, and the work type significantly influence the potential. Furthermore, we find that growth potential increases with career changes, suggesting that individuals gradually gravitate towards careers with higher potential. People in *non-bright* occupations are more eager to move up the ladder, while those in *bright* occupations desire to maintain their advantages.

The assessment of career growth potential based on occupational mobility can assist job seekers better plan their careers. Our research proves that talents tend to gravitate towards high-potential occupations, and the choice of education and industry plays a crucial role in career development. And our model quantifies the outlooks of different occupations within the occupational mobility network, which is valuable for individual career choice and career transition decisions. In terms of career choice, job seekers can select bright occupations with more available resources according to the assessment results. This can contribute to personal development. As for career transition, job seekers can make more informed decisions about changing occupations by using the model to understand their position and opportunities in the job market. When deciding to switch jobs, they can utilize the assessment to compare career prospects in various occupations and choose a more progressive career path. However, job seekers may also require information on core skills, popular training areas, and other relevant details when planning their careers (Kasman and Ali, 2022). Since only occupational mobility is considered in the proposed model, they need to obtain relevant information through other channels. In other words, this study only provides recommendations for job seekers' goals but does not address how to reach these goals. Future research could integrate data on skills, training, job responsibilities to better support individual career planning.

For organizations, CPACE is capable of scientifically depicting the flow of human capital and accurately assessing career growth potential, thereby providing reliable decision-making basis for HRM. The key application scenarios are employee career planning and human resource planning. In employee career planning, managers can use the model to understand the opportunities for growth in specific careers. They can share the assessing results with employees, providing personalized career development advice and formulating individual career plans that are more beneficial for individual development. This may enhance employees' job satisfaction and loyalty (Adekola, 2011; Khuong et al., 2020). Moreover, when an organization develops its human resource strategy, managers can be more targeted in talent acquisition and reserve based on the career prospects provided by the model. Meanwhile, managers can also adjust the organizational structure and post setting according to the prospects to adapt to future directions in the job market. The outlook for each occupation in the job market is not static. For example, according to analysis by Eloundou et al. (2023), occupations such as software engineer are susceptible to being replaced by artificial intelligence due to the impact of large language models. Being able to timely or even proactively grasp human capital dynamics would be advantageous for organizational development. However, the model cannot replace human resource managers. It aids managers in understanding the job market and supporting decision-making. But such assistance is not direct enough. Considering that humanism is an important value in HRM (Latemore et al., 2020), future research should focus on building a human-centered overall solution to achieve high-quality user interaction.

This paper still has some limitations in both data and method. Restricted by the data set, we cannot cover the talent flows of all occupations in reality. Also, occupational mobility may present different patterns under different times, spaces, and industries, but these elaborate conditions are not considered in this study. In future study, we want to expand the volume of data and strive to obtain a sufficient amount of profiles from representative industries and regions. Furthermore, we can refine regions and industries to construct time-evolving occupational mobility models to analyze and compare the dynamics of human capital flows under different industries and spatio-temporal conditions. Besides, discrimination is a persistent subject in HRM studies. Gender and race have a significant impact on career development (Huang et al., 2020; Evans and Herr, 1991). The data may contain such intrinsic biases that could potentially influence assessment outcomes. However, we currently lack data on gender, race, etc., thus making it unable to measure the biases. As identifying and quantifying discrimination with data-driven methods is a worthwhile problem, we intend to collect relevant information and explore it in the future. In terms of the method, CPACE adopts classical machine learning models. Future study can apply state-of-art computing methods like graph neural networks to improve the assessment model to better understand OMN, and incorporate the idea of multi-view representation learning to better utilize multi-dimensional human capital information (Li et al., 2019).

In summary, we introduce a data-driven and intelligent method that enables a scientific and accurate assessment of career growth potential, helping individuals and organizations perceive job market situations in time. The assessment results can assist career planning as well as human resource planning. Future research will focus on improving the assessment model in terms of data, algorithm, and overall solution to provide more comprehensive, direct, and powerful support for HRM-related decisions.

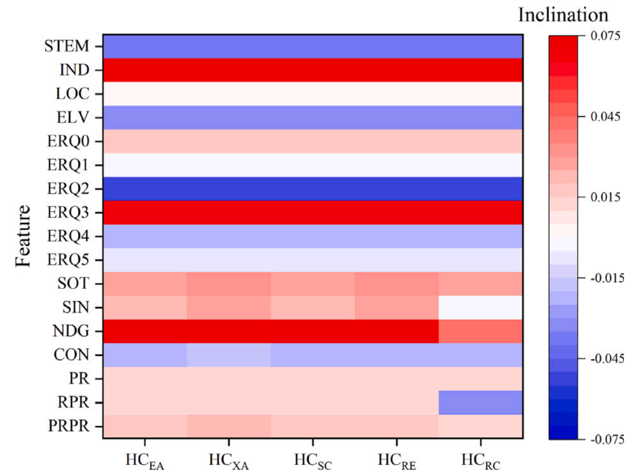


Fig. A.1. Inclination of features.

CRediT authorship contribution statement

Jiamin Liu: Methodology, Software, Writing – original draft. **Tao Wang:** Methodology, Visualization, Writing – review & editing. **Feng Yao:** Project administration, Resources. **Witold Pedrycz:** Supervision. **Yanjie Song:** Formal analysis, Writing – review & editing. **Renjie He:** Conceptualization, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We are grateful to Li Cheng for precious insights for the development of this paper. Especially thanks to Prof. Pedrycz for his valuable suggestions.

Appendix. Inclination analysis of features

We analyze the correlation between each feature and occupation outlook. An inclination I indicator is designed to identify the relationship between features and occupation outlook:

$$I = \frac{1}{N_1} \sum_{v \in V_1} x_v - \frac{1}{N_0} \sum_{v \in V_0} x_v \tag{A.1}$$

where V_1, V_0 are the sets of bright occupations and non-bright occupations respectively; N_1 is size of V_1, while N_0 is the size of V_0; x_v is the value of the feature corresponding to occupation v. If $I > 0$, the feature is positively correlated with occupation outlook. Otherwise, they are negatively correlated.

Fig. A.1 shows the inclination of general feature. Categorical features are processed by One-Hot encoding. *IND* shows a strong positive correlation, while *STEM* and *ELV* are negatively correlated with occupation outlook. For *ERQ*, the positive correlation is pronounced for the education requirement of bachelor's degree (*ERQ3*). Occupations with either lower (associate's degree, *ERQ2*) or higher (master's and doctor's degrees, *ERQ4* and *ERQ5*) are more likely to be evaluated as *non-bright* occupations. Most HC[5] features are positively related to

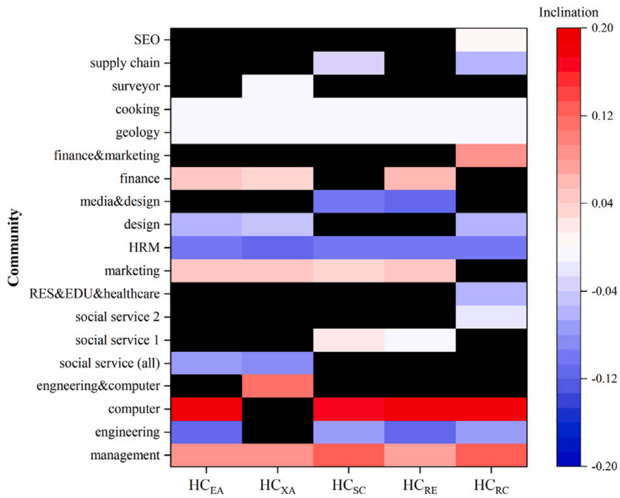


Fig. A.2. Inclination of communities.

occupation outlook, with *NDG* being the most relevant. Only *CON* has a negative correlation. However, a smaller constraint means a stronger ability to control resources. This may suggest that occupations with high importance are enriched with human capital and have promising futures. Nevertheless, the correlation pattern of HC_{RC} differs from the other HC^5 measurements, with negative correlations for both *SIN* and *RPR*. This may imply that the concentrated inflow from a single occupation harms career prospect.

The inclination of communities is shown in Fig. A.2. The communities varied with HC^5 measurements. We name these communities according to their main types of occupations in the community, which include: *management*, *engineering*, *computer*, *marketing*, *HRM*, *design*, *social service*, *finance*, *geology*, *cooking*, *surveyor*, *supply chain*, *search engine optimization* (SEO). *Social service* include *research* (RES), *education* (EDU), *healthcare*, *administrative support*, *life service*, *legal profession*, *media*, etc. *Management* and *computer* are the occupational communities most positively correlated with occupation outlook, followed by *finance* and *marketing*. *Engineering*, *HRM*, *design*, and *media* are the communities most negatively associated with occupation outlook. There is also a significant negative correlation between *social service* and occupation outlook, but the effect decreases after excluding *research*, *education*, and *healthcare* occupations (*social service 1*). What is more, it even becomes a weakly positive correlation in HC_{SC} after excluding *media* occupations (social service 2).

References

Adekola, B., 2011. Career planning and career management as correlates for career development and job satisfaction a case study of Nigerian bank employees. Aust. J. Bus. Manage. Res. 1 (2), 100–112.

A'Hearn, B., Baten, J., Crayen, D., 2009. Quantifying quantitative literacy: Age heaping and the history of human capital. J. Econ. Hist. 69 (3), 783–808. http://dx.doi.org/10.1017/S0022050709001120.

Bar-Yossef, Z., Mashiach, L.-T., 2008. Local approximation of pagerank and reverse pagerank. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management. pp. 279–288.

Baron, A., 2011. Measuring human capital. Strateg. HR Rev. 10 (2), 30–35. http://dx.doi.org/10.1108/14754391111108338.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Statist. Mech.: Theory Exp. 2008 (10), P10008.

Bocewicz, G., Golińska-Dawson, P., Szwarc, E., Banaszak, Z., 2023. Preventive maintenance scheduling of a multi-skilled human resource-constrained project's portfolio. Eng. Appl. Artif. Intell. 119, 105725. http://dx.doi.org/10.1016/j.engappai.2022.105725.

Bonaventura, M., Ciotti, V., Panzarasa, P., Liverani, S., Lacasa, L., Latora, V., 2020. Predicting success in the worldwide start-up network. Sci. Rep. 10 (1), 345.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Burt, R.S., 1995. Structural Holes: The Social Structure of Competition. Harvard University Press.

Chetty, R., Jackson, M.O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R.B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., et al., 2022. Social capital I: measurement and associations with economic mobility. Nature 608 (7921), 108–121.

Chowdhury, S., Schulz, E., Milner, M., Van De Voort, D., 2014. Core employee based human capital and revenue productivity in small firms: An empirical investigation. J. Bus. Res. 67 (11), 2473–2479.

CMSWIRE, 2011. How digital business will evolve in 2012: 6 big ideas. URL https://www.cmswire.com/cms/social-business/how-digital-business-will-evolve-in/2012-6-big-ideas-013938.php.

Company, M., 2011. Preparing for a new era of work. URL https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/preparing-for-a-new-era-of-work.

Cortes, G.M., Gallipoli, G., 2017. The costs of occupational mobility: An aggregate analysis. J. Eur. Econom. Assoc. 16 (2), 275–315. http://dx.doi.org/10.1093/jeea/jvx006.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Cox, D.R., 1958. The regression analysis of binary sequences. J. R. Stat. Soc. Ser. B Stat. Methodol. 20 (2), 215–232.

Eloundou, T., Manning, S., Mishkin, P., Rock, D., 2023. GPTs are GPTs: An early look at the labor market impact potential of large language models. arXiv e-prints arXiv:2303.10130.

Evans, K.M., Herr, E.L., 1991. The influence of racism and sexism in the career development of African American women. J. Multicult. Counsel. Dev. 19 (3), 130–135.

Farquhar, J., Hardoon, D., Meng, H., Shawe-taylor, J., Szedmák, S., 2005. Two view learning: SVM-2k, theory and practice. In: Advances in Neural Information Processing Systems, Vol. 18. MIT Press.

Fischer, G.S., da Rosa Righi, R., de Oliveira Ramos, G., da Costa, C.A., Rodrigues, J.J., 2020. Elhealth: Using internet of things and data prediction for elastic management of human resources in smart hospitals. Eng. Appl. Artif. Intell. 87, 103285. http://dx.doi.org/10.1016/j.engappai.2019.103285.

Folloni, G., Vittadini, G., 2010. Human capital measurement: A survey. J. Econ. Surv. 24 (2), 248–279. http://dx.doi.org/10.1111/j.1467-6419.2009.00614.x.

Ge, C., Shi, H., Xu, X., Jiang, J., 2023. Assessing firm human capital based on labor mobility network. IEEE Trans. Eng. Manage. 70 (3), 806–818. http://dx.doi.org/10.1109/TEM.2020.2996644.

Healy, M., Hammer, S., McIlveen, P., 2022. Mapping graduate employability and career development in higher education research: a citation network analysis. Stud. Higher Educ. 47 (4), 799–811. http://dx.doi.org/10.1080/03075079.2020.1804851.

Hoffmann, W.A., Poorter, H., 2002. Avoiding bias in calculations of relative growth rate. Ann. Bot. 90 (1), 37–42.

Huang, J., Gates, A.J., Sinatra, R., Barabási, A.-L., 2020. Historical comparison of gender inequality in scientific careers across countries and disciplines. Proc. Natl. Acad. Sci. 117 (9), 4609–4616.

IBM, 2011. 2012 IBM annual report. URL https://www.ibm.com/annualreport/2012/bin/assets/2012_ibm_annual.pdf.

Kasman, P.S.P., Ali, H., 2022. Literature review factors affecting decision making and career planning: environment, experience and skill. Dinasti Int. J. Digit. Bus. Manage. 3 (2), 219–231.

Kettenring, J.R., 1971. Canonical analysis of several sets of variables. Biometrika 58 (3), 433–451.

Khuong, M.N., Mai, T.P., Phuong, N.T.M., 2020. The impacts of human resource management practices on employees' motivation and loyalty. Manage. Sci. Lett. 10 (11), 2673–2682.

Latemore, G., Steane, P., Kramar, R., 2020. From utility to dignity: Humanism in human resource management. In: Virtuous Cycles in Humanistic Management: From the Classroom to the Corporation. Springer International Publishing, Cham, pp. 91–118.

Le, T., Gibson, J., Oxley, L., 2005. Measures of Human Capital: A Review of the Literature. New Zealand Treasury Working Paper 05/10, New Zealand Government, The Treasury, URL http://hdl.handle.net/10419/168147.

Lenihan, H., McGuirk, H., Murphy, K.R., 2019. Driving innovation: Public policy and human capital. Res. Policy 48 (9), 103791. http://dx.doi.org/10.1016/j.respol.2019.04.015.

Li, J., Long, Y., Wang, T., Shen, D., Zhang, Z., 2020. A data-driven method for competency evaluation of personnel. In: Proceedings of the 3rd International Conference on Data Science and Information Technology. ACM, pp. 93–104. http://dx.doi.org/10.1145/3414274.3414489.

Li, Y., Yang, M., Zhang, Z., 2019. A survey of multi-view representation learning. IEEE Trans. Knowl. Data Eng. 31 (10), 1863–1883.

Li, X., Ye, P., Li, J., Liu, Z., Cao, L., Wang, F.-Y., 2022. From features engineering to scenarios engineering for trustworthy AI: I&I, C&C, and V&V. IEEE Intell. Syst. 37 (4), 18–26.

Liu, J., Huang, J., Wang, T., Xing, L., He, R., 2021a. A data-driven analysis of employee development based on working expertise. IEEE Trans. Comput. Soc. Syst. 8 (2), 410–422. http://dx.doi.org/10.1109/TCSS.2020.3046726.

Liu, J., Li, J., Wang, T., He, R., 2019. Will your classmates and colleagues affect your development in the workplace: Predicting employees' growth based on interpersonal environment. In: 2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService). IEEE, pp. 71–78. http://dx.doi.org/10.1109/BigDataService.2019.00016.

Liu, J., Wang, T., Shen, D., Zhang, Z., Yao, F., He, R., 2021b. Discovering industrial and geographical separation in online professional networks: information cocoons-like effect on professional information exchange. In: 2021 7th International Conference on Big Data and Information Analytics (BigDIA). IEEE, pp. 197–204.

Lu, W.-M., Kweh, Q.L., Huang, C.-L., 2014. Intellectual capital and national innovation systems performance. Knowl.-Based Syst. 71, 201–210. http://dx.doi.org/10.1016/j.knosys.2014.08.001.

Lyons, S.T., Schweitzer, L., Ng, E.S., Kuron, L.K., 2012. Comparing apples to apples: A qualitative investigation of career mobility patterns across four generations. Career Dev. Int. 17 (4), 333–357.

Marti, G., Nielsen, F., Bińkowski, M., Donnat, P., 2021. A review of two decades of correlations, hierarchies, networks and clustering in financial markets. Prog. Inf. Geom.: Theory Appl. 245–274.

Ng, T.W., Sorensen, K.L., Eby, L.T., Feldman, D.C., 2007. Determinants of job mobility: A theoretical integration and extension. J. Occup. Organ. Psychol. 80 (3), 363–386.

Nimczik, J.S., 2017. Job mobility networks and endogenous labor markets. In: Beiträge Zur Jahrestagung des Vereins Für Socialpolitik 2017: Alternative Geld- Und Finanzarchitekturen - Session: Labor Contracts and Markets I. ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Issue: E05–V3.

Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999–66, Stanford InfoLab, URL http://ilpubs.stanford.edu:8090/422/.

Park, J., Wood, I.B., Jing, E., Nematzadeh, A., Ghosh, S., Conover, M.D., Ahn, Y.-Y., 2019. Global labor flow network reveals the hierarchical organization and dynamics of geo-industrial clusters. Nature Commun. 10 (1), 3449. http://dx.doi.org/10.1038/s41467-019-11380-w.

Parrado, E., Caner, A., Wolff, E.N., 2007. Occupational and industrial mobility in the United States. Labour Econom. 14 (3), 435–455. http://dx.doi.org/10.1016/j.labeco.2006.01.005.

Price, J.L., 1979. The study of turnover. Am. J. Sociol. 85 (1), 239–240. http://dx.doi.org/10.1086/227004.

Safavi, T., Davoodi, M., Koutra, D., 2018. Career transitions and trajectories: A case study in computing. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, pp. 675–684. http://dx.doi.org/10.1145/3219819.3219863.

Schmutte, I.M., 2014. Free to move? A network analytic approach for learning the limits to job mobility. Labour Econom. 29, 49–61. http://dx.doi.org/10.1016/j.labeco.2014.05.003.

Sedláček, P., Sterk, V., 2017. The growth potential of startups over the business cycle. Amer. Econ. Rev. 107 (10), 3182–3210. http://dx.doi.org/10.1257/aer.20141280.

Shamim, S., Cang, S., Yu, H., Li, Y., 2016. Management approaches for industry 4.0: A human resource management perspective. In: 2016 IEEE Congress on Evolutionary Computation (CEC). pp. 5309–5316. http://dx.doi.org/10.1109/CEC.2016.7748365.

Shen, D., Qin, C., Zhu, H., Xu, T., Chen, E., Xiong, H., 2021. Joint representation learning with relation-enhanced topic models for intelligent job interview assessment. ACM Trans. Inf. Syst. 40 (1), http://dx.doi.org/10.1145/3469654.

Sibunruang, H., Kawai, N., 2023. Promoting career growth potential: political skill, the acquisition of social resources and ingratiation. J. Manage. Organ. 29 (2), 227–246. http://dx.doi.org/10.1017/jmo.2021.25.

Sicherman, N., 1990. Education and occupational mobility. Econ. Educ. Rev. 9 (2), 163–179. http://dx.doi.org/10.1016/0272-7757(90)90044-6.

Smith, A., 2010. The Wealth of Nations: An Inquiry Into the Nature and Causes of the Wealth of Nations. Harriman House Limited.

Sun, S., 2013. A survey of multi-view machine learning. Neural Comput. Appl. 23 (7).

Tenenhaus, A., Tenenhaus, M., 2011. Regularized generalized canonical correlation analysis. Psychometrika 76 (2), 257–284.

Teng, M., Zhu, H., Liu, C., Xiong, H., 2021. Exploiting network fusion for organizational turnover prediction. ACM Trans. Manage. Inf. Syst. 12 (2), http://dx.doi.org/10.1145/3439770.

Trotta, D., Garengo, P., 2018. Industry 4.0 key research topics: A bibliometric review. In: 2018 7th International Conference on Industrial Technology and Management (ICITM). pp. 113–117. http://dx.doi.org/10.1109/ICITM.2018.8333930.

Vaccario, G., Verginer, L., Schweitzer, F., 2021. Reproducing scientists' mobility: A data-driven model. Sci. Rep. 11 (1), 10733.

Wang, C., Zhu, H., Wang, P., Zhu, C., Zhang, X., Chen, E., Xiong, H., 2021. Personalized and explainable employee training course recommendations: A Bayesian variational approach. ACM Trans. Inf. Syst. 40 (4), http://dx.doi.org/10.1145/3490476.

Wicks, A., 2001. Occupational potential: A topic worthy of exploration. J. Occup. Sci. 8 (2), 32–35. http://dx.doi.org/10.1080/14427591.2001.9686487.

Wossmann, L., 2003. Specifying human capital. J. Econ. Surv. 17 (3), 239–270. http://dx.doi.org/10.1111/1467-6419.00195.

Xu, H., Yu, Z., Xiong, H., Guo, B., Zhu, H., 2015. Learning career mobility and human activity patterns for job change analysis. In: 2015 IEEE International Conference on Data Mining. pp. 1057–1062. http://dx.doi.org/10.1109/ICDM.2015.122.

Xu, H., Yu, Z., Yang, J., Xiong, H., Zhu, H., 2016. Talent circle detection in job transition networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, Association for Computing Machinery, New York, NY, USA, pp. 655–664. http://dx.doi.org/10.1145/2939672.2939732.

Yelle, L.E., 1979. The learning curve: Historical review and comprehensive survey. Decis. Sci. 10 (2), 302–328.

Zhang, Y., Tang, J., Yang, Z., Pei, J., Yu, P.S., 2015. Cosnet: Connecting heterogeneous social networks with local and global consistency. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1485–1494.

Zhao, J., Xie, X., Xu, X., Sun, S., 2017. Multi-view learning overview: Recent progress and new challenges. Inf. Fusion 38, 43–54.