A Hierarchical Career-Path-Aware Neural Network for Job Mobility Prediction

Qingxin Meng[1], Hengshu Zhu[2*], Keli Xiao[4], Le Zhang[5], Hui Xiong[1,2,3*]
[1]Rutgers-the State University of New Jersey [2]Talent Intelligence Center, Baidu Inc.
[3]Baidu Business Intelligence Lab, Baidu Research [4]Stony Brook University
[5]School of Computer Science, University of Science and Technology of China
qm24@scarletmail.rutgers.edu, zhuhengshu@baidu.com,
keli.xiao@stonybrook.edu, laughing@mail.ustc.edu.cn, xionghui@gmail.com

ABSTRACT

The understanding of job mobility can benefit talent management operations in a number of ways, such as talent recruitment, talent development, and talent retention. While there is extensive literature showing the predictability of the organization-level job mobility patterns (*e.g.*, in terms of the employee turnover rate), there are no effective solutions for supporting the understanding of job mobility at an individual level. To this end, in this paper, we propose a *hierarchical career-path-aware neural network* for learning individual-level job mobility. Specifically, we aim at answering two questions related to individuals in their career paths: 1) who will be the next employer? 2) how long will the individual work in the new position? Specifically, our model exploits a hierarchical neural network structure with embedded attention mechanism for characterizing the internal and external job mobility. Also, it takes personal profile information into consideration in the learning process. Finally, the extensive results on real-world data show that the proposed model can lead to significant improvements in prediction accuracy for the two aforementioned prediction problems. Moreover, we show that the above two questions are well addressed by our model with a certain level of interpretability. For the case studies, we provide data-driven evidence showing interesting patterns associated with various factors (*e.g.*, job duration, firm type, *etc.*) in the job mobility prediction process.

ACM Reference Format:
Qingxin Meng, Hengshu Zhu, Keli Xiao, Le Zhang, Hui Xiong. 2019. A Hierarchical Career-Path-Aware Neural Network for Job Mobility Prediction. In *The 25th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDDâĂŹ19), August 4–8, 2019, Anchorage, AK, USA.* ACM, NY, NY, USA, 11 pages. https://doi.org/10.1145/3292500.3330969

1 INTRODUCTION

In this paper, we focus on providing data-driven solutions to the job mobility prediction problem. The importance of job mobility has been widely documented as a key element of human behaviors by researchers from different areas. For instance, Topel and Ward [24] claimed that work experience accumulation is mainly attributable to job changing activities for locating good job matches, especially for young employees. It has also been found that people have renewed interests in job movements by which job mobility occurs and results in different career paths [22]. In addition, evidence has been provided to support the significant relationship between individual's decision of migration and job mobility [3], the connection between social ties and job changes [26], the wage effect of cumulative job mobility [10], and the like.

From the perspective of human resource manager, it is important to understand the job mobility in the organization level as well as the individual level. The main purpose of related studies is to support the decision-making process regarding talent management. Understanding the potential career paths of an employee would help executives and department managers in internal promotion decisions to motivate key talents and reduce the turnover rate. Also, during the recruiting process, employers may be interested in knowing the probability for candidates to accept the job offers. Meanwhile, if there is a high chance of hiring, how long will they stay? On the other hand, from an employee's viewpoint, people also concern about their career development and growth for achieving professional success, and a question that may keep bothering them is: what is the best and fastest career path leading to the success in professional life?

However, job mobility prediction is not an easy task. Traditional studies of job mobility were largely based on limited survey data and focused on the empirical analysis of key factors affecting people's career paths [18, 25]. The rapid development of information technology and the emergence of professional social networks enable us to collect and analyze large-scale career path data from the real word. For example, as one of the earliest works in the individual-level job mobility prediction topics, Xu et al. [27] developed a framework to predict whether there is a large chance of job change in the next six months for individuals. Liu et al. [14] proposed a multi-view multi-task learning approach to predict the promotion in one's career path. These works considered work experience and daily activity data in their models, but the target problems were somehow general and had limited practical applications. Thus, in this paper, we propose to address the problem of job mobility prediction by answering two specific questions: (1) "Who is your next employer?" and (2)"How long will you work for your next employer?" The first question is regarding the position prediction, and the second one tells the eventual duration of your new job.

*Hui Xiong and Hengshu Zhu are corresponding authors.

KDD '19, August 4–8, 2019, Anchorage, AK, USA

ACM ISBN 978-1-4503-6201-6/19/08…$15.00
https://doi.org/10.1145/3292500.3330969

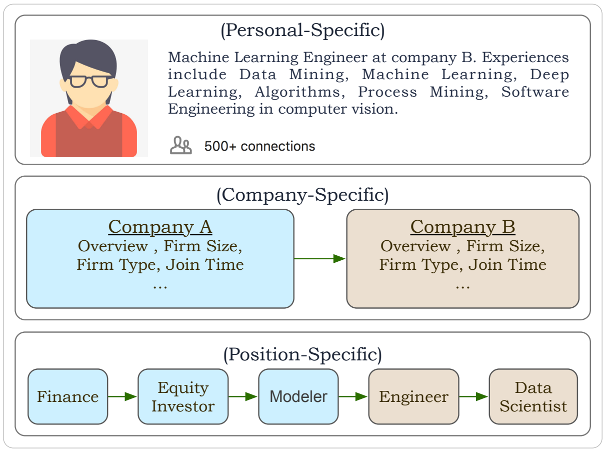


Figure 1: An example of the three-layer structure of our job mobility data.

The main challenges of the proposed prediction tasks are twofold. (1) We need to handle the dynamic hierarchical nature of career paths for employees, such as internal job mobilities and external job mobilities. For example, one person may experience several internal job transfers within a company before he/she hops to a new company. Both the internal transfers and external job hoppings may influence the direction of the future of the career path at different levels. Moreover, the data are complex with heterogeneous forms, including the personal-specific, company-specific, and position-specific data. For example, the personal self-introductions and company descriptions are freely structured, some features are categorized, while others are numerical. (2) The other challenge is to jointly consider the influence between environmental factors and individual historical patterns. The closeness between companies is one of the environmental factors. For example, one person working in a bank may have a high chance to hop to another bank. Meanwhile, tracing back the whole history of one's career path, which company or position takes the main role in the decision-making process, is another important factor we need to figure out.

We provide our solutions to the aforementioned issues and contribute to the literature in four ways as follows:

- To the best of our knowledge, among the existing work on individual-level job mobility prediction, we are the first to conduct dual highly specific tasks to predict people's next employer and the eventual duration.
- We propose a hierarchical career-path-aware neural network approach to integrate three levels of information, including personal-specific, company-specific and position-specific knowledge. The model embeds survival analysis and attention mechanism, which lead to a certain level of interpretability of results.
- For both proposed forecasting tasks, we demonstrate evidence showing the superiority of our model in comparison to several well-known benchmarks.
- Our model offers a new way to show data-driven evidence in support of the connection between specific factors and job mobility. As case studies, new evidence has been presented to show the impacts of various factors (*e.g.*, job duration, firm type, *etc.*) on the job mobility prediction performance.

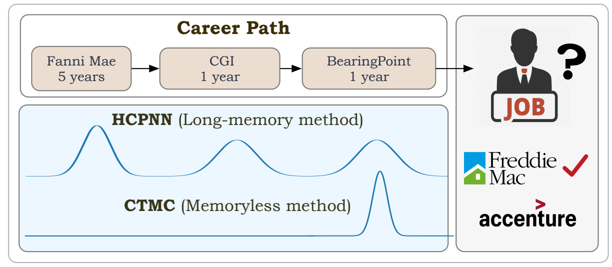


Figure 2: The prediction of a real case.

2 DATA DESCRIPTION

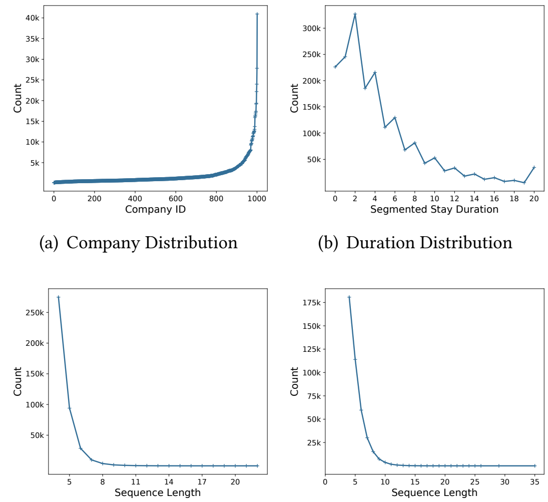
The data were collected from a famous online professional social platform, where users can build professional profiles introducing their education and work experience, like a public online curriculum vitae. We summarize the collected features into three categories, including personal-specific, company-specific, and position-specific data. Personal-specific information is static and includes freely structured self-description texts and the number of social connections. Company-specific features (*e.g.*, company name, type, size, etc.) and position-specific features (*e.g.*, position type, service duration, etc.) were collected as sequential data to describe the work timeline of the users. Our data contain both internal and external job transitions in their professional life. Figure 1 shows an example of the hierarchical structure of our data.

How to design an effective framework to model such a hierarchical structure becomes a key challenge in this job mobility prediction task. To handle the complex data structure, we formatted those three kinds of features into different levels. The static personal information were transformed into a vector as one level, while company-specific and position-specific features were transformed into a sequence of vectors respectively as the second and the third levels. Each level in the structure contains useful information that we do not want to mass them up in a simple machine learning model. Thus, we propose to construct a neural network model to handle the three level of inputs hierarchically. We will provide detailed discussion in section 3.

The second challenge and the motivation of our model are problem-specific. We believe the job mobility prediction is a sequential problem and even a long-distance dependent sequential problem. The decisions in people's career paths rely on two groups of factors. The first one refers to the work environmental factors, which describe the natural connections among company-specific characteristics, such as firm types. For instance, employees in a bank are highly prone to hop to another bank, rather than other manufactories. Although we do not have direct features to represent the similarities between companies, such information will be obtained by learning from people's job-hopping patterns. The second group of factors we should consider are the individual historical factors. In one's historical career path, she/he might have served several employers and been occupied in different positions. An effective model should be able to understand which experiences during the career path play the most important roles for future decisions. Such information should be captured during the model training process.

In particular, here we introduce a motivating example of the prediction problem in Figure 2, which is a real case in our sample. Specifically, the person has worked for three employers, namely “Fannie Mae”, “CGI”, and “BearingPoint”, one after another before he hopped to “Freddi Mac”. If we use the Markov Chain Model, which only considers the environmental factors to predict the next employer, the result is “Accenture”. The model considers the last employer “BearingPoint” as an important reference in the prediction process, given that “Accenture” and “BearingPoint” are both consulting companies. The result is reasonable as it only considers the environmental factors. On the other hand, our model intelligently gave higher attention for his first employer “Fanni Mae” than “CGI” and “BearingPoint”, which might be due to their associated duration. Therefore, it can successfully predict the right next employer “Freddi Mac”, which is closely tied up with the person’s first employer “Fannie Mae”. Indeed, “Fannie Mae” and “Freddi Mac” are two large house mortgage companies. To get the correct predictions, we need to jointly consider the environmental factors as well as the individual historical factors, which can be discovered from the detailed information of people’s career paths, such as the job duration of each position.

We also analyzed the characteristics regarding the distribution of our samples. Figure 3 (a) demonstrates the distribution of the occurrence number of the firms in our real-world dataset. As can be seen, 20% of the firms cover about 60% samples in the data. Figure 3 (b) shows the distribution of the job duration, which was split into 21 windows as 0.5 years, 1 year, 1.5 years, ..., 10 years, and more than 10 years. We can see that most people stay in one position for 1-3 years, and there is a decreasing pattern for the longer duration. Interestingly, there is also a “sawtooth” pattern in the job duration distribution, which may indicate that people try to avoid leaving a job in the odd number of half years. Also, both the length of company and position sequence have long tail shape too, as illustrated in Figure 3 (c) - (d). Our model needs to handle the imbalanced distribution for better predictions.



(a) Company Distribution (b) Duration Distribution

(c) Company Transition (d) Position Transition

Figure 3: The data distribution of different aspects.

3 PROBLEM AND METHODOLOGY

In this section, we formulate the job mobility prediction problem based on the data availability and then discuss the new method we proposed for addressing the problem.

3.1 Problem Statement

Let $u \in U$ denote a person, $c \in C$ denote a company, $p \in P$ denote a position, where U, C and P represent the full set of people, companies, and positions respectively. Given the company sequence $\overrightarrow{Q(u)}$, the position sequence $\overrightarrow{B(u)}$, and the personal-specific information $\Omega(u)$, we represent u’s three-layer career path as $S(u) = \left\{\overrightarrow{Q(u)}, \overrightarrow{B(u)}, \Omega(u)\right\}$. The company sequence $\overrightarrow{Q(u)}$ can be written as:

$$\overrightarrow{Q(u)} = \left\{(c_1, c_2, ..., c_g)|u\right\}, \quad (1)$$

and position sequence $\overrightarrow{B(u)}$ can be written as:

$$\overrightarrow{B(u)} = \left\{(p_{11}, p_{12}, ...), ..., (p_{g1}, ..., p_{gh})|u\right\}, \quad (2)$$

where p_{gh} describes the h-th position for his g-th employer c_g. For example, p_{24} represents the fourth position in the second company c_2 where a person worked.

Then, we formulate our problem as follows:

Problem 1. *Given a person’s three-layer career path,*

$$S(u) = \left\{\overrightarrow{Q(u)}, \overrightarrow{B(u)}, \Omega(u)\right\}, \quad (3)$$

where $\overrightarrow{Q(u)}$ stands for company sequence, $\overrightarrow{B(u)}$ stands for position sequence, and $\Omega(u)$ stands for personal information, we want to predict person u’s next employer c_{g+1} and the duration d_{g+1} at c_{g+1}.

3.2 An Overview of the Model

Now we introduce the methodology we proposed for addressing the job mobility prediction problem. The design of our model is rooted in the hierarchical data structure, and we name it as the *hierarchical career-path-aware neural network* (HCPNN). The model includes three main components, namely *Internal Job Mobility Representation*, *External Job Mobility Representation*, and *Prediction*.

Figure 4 illustrates the framework of our HCPNN. Specifically, for the component of *Internal Job Mobility Representation*, we embed the sequential position features as the inputs to a long short-term memory (LSTM) [7] layer. Then, we apply a local attention mechanism for obtaining the internal job mobility representation. For the component of *External Job Mobility Representation*, we first concatenate sequential company feature embeddings with the internal job mobility representation, then we feed them into another LSTM layer for training the external job mobility representation. Meanwhile, we conduct the embedding process for the personal-specific features, and then we apply the global attention mechanism to both external job mobility representation and static personal representation. Following that, we form a hierarchical job mobility neural network, which has the ability to learn the influences of internal and external job mobility on their next job decisions. Finally, for

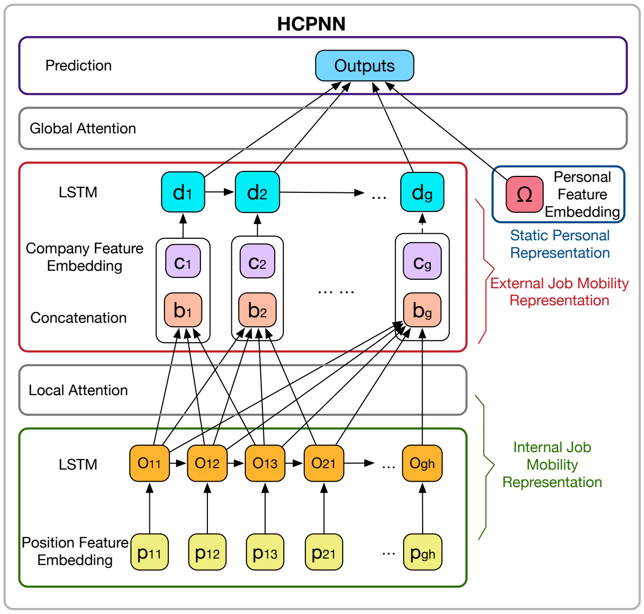


Figure 4: The graphical representation of the HCPNN model.

different learning tasks, the output from the HCPNN will be fed into different *prediction* widgets. As emphasized in the problem statement, we aim at predicting the next employer as well as the job duration with the next employer for every person.

3.3 Technical Details

Here, we introduce the details of *Internal Job Mobility Representation*, *External Job Mobility Representation*, and *Prediction* components mentioned above.

3.3.1 Internal Job Mobility Representation. The inputs of internal job mobility representation layer are position-specific features $\overrightarrow{B(u)}$. After the embedding, we feed them into an LSTM layer to learn the hidden representation of the position-specific sequential features. We choose LSTM to handle this task due to its predictive power, as well as the ability to alleviate the gradient vanishing problem in long-distance dependent sequential problems. In our framework, we refer the output of LSTM of this layer as $o_{11}, o_{12}, ..., o_{gh}$, and then we apply a local-attention mechanism with these outputs to get the final representation for internal job mobility. In particular, we propose to add an attention mechanism for obtaining the model interpretability, on which we rely for result analysis. Also, we use this mechanism to align the internal job mobility representation $b_1, b_2, ..., b_g$ and company sequence embeddings $c_1, c_2, ..., c_g$ with the same length.

The attention mechanism tries to capture the degree of attention for representing the importance of inputs in the learning process. Based on our proposed tasks and the data structure, we apply the attention technique as follows. For each attention output b_i, we assign attention based on the company sequence inputs before and include c_i. For instance, suppose position sequence p_{11}, p_{12}, p_{13} is associated with company c_1, and p_{21} is associated with c_2, then we assign the attention value on o_{11}, o_{12}, o_{13} to get the attention output b_1, and assign attention value on $o_{11}, o_{12}, o_{13}, o_{21}$ to obtain the output b_2. In this way, we prevent to use future information to predict the future. Also, the internal job mobility layer is aligned with the external job mobility layer. The local attention mechanism can be formulated as follows.

$$v_{ij} = tanh(W_a o_{ij} + b_a),$$
$$\alpha_{ij} = \frac{\exp(v_{ij}^T u_a)}{\sum_{i=1}^{g}\sum_{j=1}^{h}\exp(v_{ij}^T u_a)}, \quad (4)$$
$$b_g = \sum_{i=1}^{g}\sum_{j=1}^{h}\alpha_{ij}(W_a o_{ij}),$$

where W_a, b_a and u_a are training parameters, o_{ij} means the i-th company j-th position's hidden states learned from the first LSTM layer, and b_g is the output vector for g-th internal job mobility representation.

3.3.2 External Job Mobility Representation. Similar to the internal job mobility representation, we utilize an LSTM layer and attention mechanism to model the external job mobility information. First, we concatenate the aligned sequential company embedding data $c_1, c_2, ..., c_g$ with the internal mobility representation $b_1, b_2, ..., b_g$, then we feed them into another LSTM layer and obtain the output $d_1, d_2, ..., d_g$. Personal features Ω is further embedded. Then a global attention is computed based on both $d_1, d_2, ..., d_g$ and Ω for getting a final output. The attention technique implemented here not only integrates the personal information into our framework, but also improves the result interpretability of our model.

3.3.3 The Prediction Module. Our job prediction problem contains two major tasks: the next company and the duration at the next company. For the first one, we formulate it as a classification task as below. We first feed the output vector learned from the HCPNN model into a fully-connected layer where the output dimension matches our total company numbers. Then, we use a *softmax* activation function to normalize the probabilities $P(c)$ of each possible company. The process is demonstrated in figure 5.

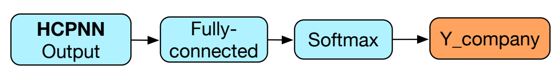


Figure 5: The process of predicting next employer.

Based on the maximum likelihood estimation, we optimize the loss function for predicting the next employer, which is formulated as follows. Given a person u, we maximize

$$L^u_{company} = \sum_{i=2}^{g} \log\left(P(c = c_i)|S(u)\right). \quad (5)$$

In the optimization process, we can not predict the first employer c_1, as $\overrightarrow{B(u)}$ and $\overrightarrow{Q(u)}$ are empty before c_1. So we summarize the log-likelihood of company sequence, start from the second index.

For the second task of predicting job duration, we integrate survival analysis into our framework. We regard the event a person joins a company as her start life in this company. And the event of leaving the company as a death event.

Table 1: The statistics of experimental data.

Name	value
Number of samples	414,266
Number of companies	1,002
Number of normalized position types	26
Max/min/mean company sequence length	22/4/4.52
Max/min/mean position sequence length	35/4/5.14
Observed time periods	1988.1-2018.11

Survival analysis has been widely used for estimating the occurrence time of an event with censored observations. We denote the probability of an event does not happen before time t as $P(T_{survival} \geq t)$, and the instantaneous rate of the occurrence of the target event at time t as $\lambda(t)$, so we have

$$\frac{d}{dt} P(T_{survival} \geq t) = -\lambda(t) P(T_{survival} \geq t). \tag{6}$$

To solve the Equation 6, we can get

$$P(T_{survival} \geq t) = exp\left(-\int_0^t \lambda(\tau)\, d\tau\right). \tag{7}$$

Sometimes, we can only observe the survival event within a period time t, after time t we cannot continue the observations. This is called right-censored data. And the probability can be computed by Equation 7. And if a target event occurred at the exact time t', the probability is computed as

$$P(T_{survival} = t') = \lambda(t') exp\left(-\int_0^{t'} \lambda(t)\, dt\right), \tag{8}$$

where the meaning of the function can be explained as the joint probability of the event happening at the exact time t' and the event does not happen before time t'.

In our problem, we first use a fully-connected layer to transform the output learned from the HCPNN into $k+1$ dimension, where the first k dimensions can represent the individual turnover probability for the segmented time period $\left[(0, \frac{1}{k}T), [\frac{1}{k}T, \frac{2}{k}T), ..., [\frac{k-1}{k}T, T)\right]$. The last dimension denotes the turnover probability after T. Note that T is the longest observation time in our problem. The larger of k, the higher precision of the simulation. In this way, we can transform the task of predicting the next duration to a survival analysis problem. Figure 6 illustrates the process of job duration prediction.

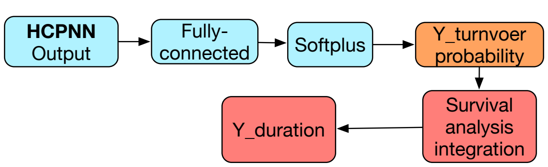


Figure 6: The process of predicting job duration.

The technical details can be summarized as follows. Let $\lambda(\tau)|S(u)$ denote the individual turnover probability for the next employer under the condition of previous career path $S(u)$, where $\tau \in (0, +\infty)$. We will use $\lambda(\tau)$ for short in the following. We have $\lambda(\tau) > 0$ by definition, so we use the *softplus* function to constrain the value to be positive. So the log-likelihood of predicting the duration at the next company can be computed as:

$$\begin{aligned} L^u_{duration} &= \log\left(\prod_{i=2}^{g} P\big(d = d_i | S(u)\big)\right) \\ &= \sum_{i=2}^{g} log\left(\lambda(d_i) \exp\left(-\int_0^{d_i} \lambda(\tau)\, d\tau\right)\right) \\ &= \sum_{i=2}^{g} log\left(\lambda(d_i)\right) - \sum_{i=2}^{g} \int_0^{d_i} \lambda(\tau)\, d\tau. \end{aligned} \tag{9}$$

Table 2: The features used in HCPNN.

Personal Specific Feature	
Number of social connections	Numerical
Self-introduction	Free text
Company Specific Feature	
Job duration at the company	Numerical
Company personnel flow (in/out/transfer)	Numerical
Company description	Free text
Company id, Size ,Type, Location, Age	Categorical
Position Specific Feature	
Duration in the position	Numerical
Position type	Categorical

The reason why the summation starts from index 2 is the same with that we explained for computing $L^u_{company}$. If we split the observation time into two parts, $(0, T)$ and $(T, +\infty)$, the job-hopping events occurred after time T will be treated as right-censored data points. The Equation 9 can be rewritten as:

$$\begin{aligned} L^u_{duration} = \sum_{i=2, d_i<T}^{g} log\left(\lambda(d_i)\right) &- \sum_{i=2, d_i<T}^{g} \int_0^{d_i} \lambda(\tau)\, d\tau \\ &- \sum_{i=2, d_i \geq T}^{g} \int_0^{T} \lambda(\tau)\, d\tau. \end{aligned} \tag{10}$$

By summarizing the loss function for predicting the next employer and job duration, we get our final loss function as below:

$$Loss = -\sum_{u \in U} \left(\alpha L^u_{company} + (1-\alpha) L^u_{duration}\right), \tag{11}$$

where the α is the tuning parameter of these two types of loss functions. Given the probabilities of $P(c_{g+1}|S(u))$ and $\lambda(\tau)|S(u)$, it is easy to deduce the most possible next hopping company to be $argmax\{P(c_{g+1}|S(u)\}$. As for the next job duration, we need to calculate the integration of $P(d = \tau|S(u))$ over time $\tau \in (0, +\infty)$, the formulation will be

$$\begin{aligned} duration &= \int_0^{+\infty} \tau \cdot P(d = \tau | S(u))\, d\tau \\ &= \int_0^{+\infty} \tau \cdot \lambda(\tau) \exp\left(-\int_0^{\tau} \lambda(s)\, ds\right) d\tau. \end{aligned} \tag{12}$$

Since the Equation 12 is non-linear, and there is no analytic solution, we can use simulation to solve the integration problem. We segment the time window $t \in (0, +\infty)$ into z intervals, then use the function above to calculate the integration.

4 EXPERIMENTS

In this section, we introduce the details of experiments conducted on a real-world dataset for validating our HCPNN.

4.1 Experimental Setup

The data were collected from a well-known online professional social platform. We filtered out the samples with the number of the external job transitions less than four. And we selected those companies having the highest occurrence frequency as our research targets. The major statistics of the data are summarized in Table 1.

As described before, our data have three levels, personal-specific features, company-specific features, and position-specific features. To handle the rich forms of data (free text, numerical and categorical features), we preprocessed the data with the following methods. For the free text feature, such as company description and personal self-introduction, we used the *wold2vec* [20] embedding method to transform a word into a vector. Then we computed the mean value of the embedding for every dimension respectively, in this way we got a fixed length of the vector for the free text of varying length. For the categorical features, the number of types of which less than ten, we used one hot encoding; for those with the number of types more than ten, such as company ID, we used a fully-connected layer for the embedding process. To process the job duration at companies and positions, we first segmented the time less than ten years by every half year into 20 small windows, and the time larger than 10 years was set into one category. In this way, the job duration was segmented into 21 categories, namely 0.5 years, 1 year, 1.5 years,..., 10 years, and more than 10 years. We segmented job duration in this way because it is hard to predict the exact leaving time when an employee stays service for more than 10 years. We counted and normalized the personnel flow in/out/transfer number of every company for every three years. Thus, given a specific company c and a timestamp m, we can draw the corresponding flow in/out/transfer value of the company c at the time period $m-1$. We treated them as company-specific features. The features used in our model are summarized in Table 2. After preprocessing of the data, we set up the configuration of our HCPNN based on our preliminary experiments. The key dimensions and value settings of the model are reported in Appendix A.3 Table 9.

4.2 Baselines

We compared our model with state-of-the-art techniques, which are listed as follows: non-sequential models (*e.g.*, Logistic Regression (LR), Random Forest (RF), and Decision Tree (DT)), sequential models (*e.g.*, Conditional Random Field (CRF) [11], Continuous Time Markov Chain (CTMC) [1] [1]) , and the stochastic time series models (*e.g.*, Poisson Process (PP) [9], Multi-variable Hawkes Process (MHP) [16]). Also, we tested two modified versions of our method HCPNN named HCPOP and HCPOS. HCPOP model does not contain internal transition representation layer, while HCPOS does not contain the survival analysis technique, the job duration prediction was treated as a classification problem. We modified the CRF and MHP to fit our problems, the technique details will be introduced in Appendix A.4.

[1] https://github.com/kmedian/ctmc

4.3 Evaluation Metrics

For predicting the next employer, we use *Accuracy@k* (*Acc@k*) and *mean reciprocal rank* (*MRR*) to evaluate the results, where $Acc@k = \frac{1}{N}\sum_{i=1}^{N} I(rank(i) \le k)$, and $MRR = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{rank(i)}$, where the N is the total number of predictions, and $rank(i)$ represents the real label rank in the predicting ranking list. If $rank(i) \le k$, then $I(rank(i) \le k)$ equals one, else equals zero. In this paper, we set $k = 1, 15, 30$ respectively. The higher value of *Acc@k* and *MRR*, the better performance. For predicting the job duration, we use *mean absolute error* $MAE = \frac{1}{N}\sum_{i=1}^{N}|p_i - r_i|$ and *Root Mean Square Error* $RMSE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(p_i - r_i)^2}$ to evaluate the performances. p_i and r_i are the predicted job duration and the real job duration.

4.4 The Overall Performance

To validate the effectiveness of our model, we first randomly split the samples by (0.8/0.1/0.1) as the training/validation/test datasets. And the overall performance including predicting next employer and job duration respectively. The results of predicting next employer are reported in Table 3. We calculated the improvements of our model against all the other baselines. We can observe that the tree-based model are not able to effectively handle this predicting task, and the performances of the sequential models are better than that of non-sequential models. Our model has the best performances with significant improvements. For example, we achieved improvements of 231.8%, 160.3%, 121.4%, and 609.1%, in terms of *Acc@1*, *Acc@15*, *Acc@30*, and *MRR*, against the DT. Comparing to the best baseline, CTMC, our model also resulted in a consistent superior. To validate the improvement of HCPNN over HCPOP is statistically significant, we randomly split the data by (0.8/0.2) ten times, and conducted a standard student t-test. As the results, the p-value is very small for both employer and duration predictions, demonstrating a statistically significant improvement, and validating the importance of internal job mobility representation layer in our model. More detailed results about t-test are reported in Appendix A.5 Table 11.

The results of predicting job duration are summarized in Table 4. Similar to Table 3, we computed the performance improvement of HCPNN against all the other baselines. We can observe that the stochastic time series models and the variant of our model HCPOS achieved relative better performances, indicating that the task should be considered as a time series problem. Our model achieved the best performance with obvious advantages, while HCPOS, which uses the same structure but without the survival analysis, resulted in worse performance, even comparing it to MHP and PP. These results confirm the importance of our framework as well as the survival analysis on the duration prediction task.

4.5 Robustness Analysis

We also conducted additional experiments to confirm the robustness of our method. We first randomly split the dataset by samples with different training proportion settings (i.e., 90%, 80%, 70%, 60%, and 50%), the results of which are reported in Table 5. We can observe that with the training proportion increasing, the performance is improving as well. Furthermore, we split the dataset by years as well, for instance, if we set the splitting year as 2005, the whole sample sequences will be truncated by the year 2005, the points

Table 3: The overall performance (next employer prediction).

Model	Acc@1	Improvement	Acc@15	Improvement	Acc@30	Improvement	MRR	Improvement
DT	0.022	231.8%	0.156	160.3%	0.243	121.4%	0.022	609.1%
RF	0.021	247.6%	0.157	158.6%	0.250	115.2%	0.021	642.9%
LR	0.054	35.2%	0.313	29.7%	0.420	28.1%	0.120	30.0%
CRF	0.053	36.5%	0.320	26.9%	0.433	24.2%	0.120	30.0%
CTMC	0.060	21.7%	0.336	20.7%	0.457	17.6%	0.089	75.4%
HCPOP	0.071	2.8%	0.402	1.0%	0.534	0.7%	0.154	1.3%
HCPNN	**0.073**	-	**0.406**	-	**0.538**	-	**0.156**	-

[1] The improvement of our HCPNN over HCPOP is statistically significant with a p-value consistently less than 0.01.

Table 4: The overall performance (duration prediction).

Model	MAE	Improvement	RMSE	Improvement
DT	3.839	28.8%	5.608	31.3%
RF	4.070	32.8%	5.782	33.4%
LR	3.096	11.7%	4.857	20.7%
CTMC	4.128	33.8%	5.872	34.4%
PP	3.143	13.0%	4.228	8.9%
MHP	3.029	9.7%	4.214	8.6%
HCPOS	3.095	11.7%	4.898	21.4%
HCPOP	2.739	0.2%	3.880	0.7%
HCPNN	**2.734**	-	**3.852**	-

[1] The improvement of our HCPNN over HCPOP is statistically significant with a p-value consistently less than 0.01.

Table 5: The performance on randomly split samples.

Ratio	Acc@1	Acc@15	Acc@30	MRR	MAE	RMSE
0.9	0.074	0.405	0.538	0.157	2.729	3.855
0.8	0.072	0.403	0.534	0.155	2.732	3.892
0.7	0.071	0.401	0.532	0.154	2.722	3.912
0.6	0.070	0.398	0.528	0.152	2.746	3.884
0.5	0.068	0.393	0.524	0.149	2.724	3.919

Table 6: The performance on splitting data by years.

Year	Acc@1	Acc@15	Acc@30	MRR	MAE	RMSE
2005	0.042	0.297	0.419	0.106	2.692	3.517
2006	0.041	0.313	0.440	0.109	2.556	3.366
2007	0.043	0.313	0.437	0.109	2.566	3.271
2008	0.045	0.328	0.455	0.115	2.651	3.241
2009	0.046	0.331	0.460	0.116	2.466	2.999
2010	0.048	0.340	0.470	0.120	2.277	2.796

in a sequence before the year 2005 will be used for training, and the points in the sequence after 2005 will be used for predicting. The results are shown in table 6. We can observe that with the splitting year approaching recent, the performance improves. The results of two different splitting settings are stable, demonstrating the robustness of our model HCPNN.

4.6 Attention Analysis

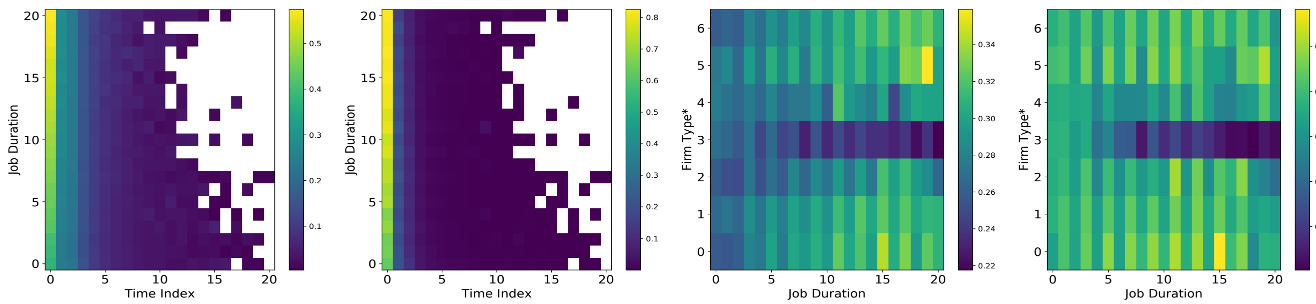
With the attention mechanism, our HCPNN model offers new opportunities to investigate the importance of considered factors and related patterns in the job-mobility prediction tasks. Here, we show some examples in which we study the characteristics of three job-mobility factors, including the job duration, the firm type, the time index of career paths.

In Figure 7 (a) - (b), each column represents a time index which is set to be the position distance prior to the last job. For example, the last job has a time index of zero; the one before the last job has a time index of 1, and so forth. Each row represents the duration of a job, and the color of each grid shows the mean value of attention. The brighter of the color, the higher attention. The grids in white are missing values (no observation). Two interesting patterns can be found: (1) The longer stay with an employer, the higher attention (importance) it has; (2) A job appearing in a later position in one's career path has higher attention. Specifically, we find that 76.8% of the people in our sample have the highest attention weights for their last jobs.

On the other hand, the firm type also matters. As demonstrated in Figure 7 (c) - (d), an interesting pattern can be found. In general, with the job duration increases, the importance of an employer increases as well. However, this pattern is reversed for government-based organizations. That is, the longer people stay in the government, the less attention it has for the job mobility.

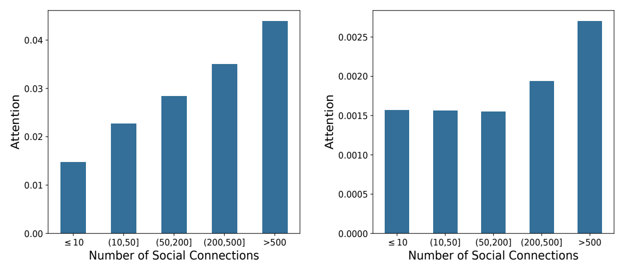
4.7 Individual Effect and Firm Effect

We also find the evidence of the existence of individual effect and organization effect in the predictions. We showcase the importance of number of social connections in the job mobility prediction in Figure 8. As can be seen, with the number of social connections increases, the attention increases as well. Moreover, the HCPNN will pay more attention to personal information when predicting next employer than predicting job duration. These findings are consistent with [26] regarding the relationship between social ties and job mobility. We also evaluate the mean attention grouped by companies and plot the sorted attention in Figure 9, which appears as a sinh curve. We report the top-10 companies with the highest attention and compare them to the top-10 companies with highest occurrence number. The result is listed in Table 7. None of those companies overlapped. As can be seen, most of the top-10 companies with highest attention are emerging high-tech



(a) Predicting Next Employer (b) Predicting job duration (c) Predicting Next Employer (d) Predicting job duration

Figure 7: The attention analyses of job mobility patterns.
(* 0: Sole Proprietorship, 1: Privately Held, 2: Joint Venture, 3: Government, 4: Educational Institution, 5: Non-Profit Organization, 6: Public.)



(a) Predicting Next Employer (b) Predicting job duration

Figure 8: The attention on number of social connections.

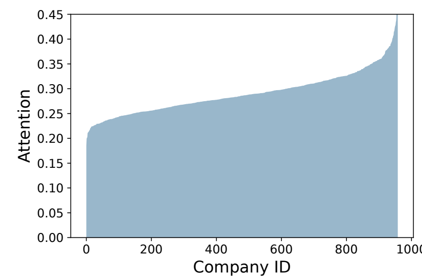


Figure 9: The sorted attention value on companies.

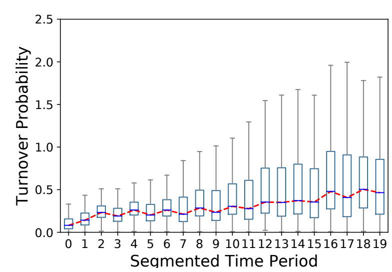


Figure 10: The turnover probability over time before 10 years.

companies, while the most of the top-10 companies with highest occurrence frequency are relative old famous companies.

4.8 Individual-level Turnover Analysis

To analyze the patterns of turnover probability for individuals, we gathered the individual-level turnover probability for all samples and plot them in Figure 10. We can observe that with the working years increasing, the instantaneous turnover probability steady increases too. We also found an interesting phenomenon, which shows the individual turnover probability follows a "sawtooth" shape. This is consistent with our findings regarding the job duration distribution, as shown in Figure 3 (b). The pattern indicates that people tend to stay with an employer for integer years rather than odd number of half years. Our model learned this pattern without any pre-defined constraints.

Table 7: Attention on companies.

Top 10 companies with highest attention
Facebook, LinkedIn, SapientNitro, GE Oil & Gas
Amazon Web Services, BBVA Compass bank, inVentiv Health
IndusInd Bank, Societe Generale Corporate and Investment Banking
Everything Everywhere (EE)
Top 10 companies with highest occurrence number
PricewaterhouseCoopers, Deloitte, Microsoft
Oracle, JPMorgan Chase, Bank of America, Citibank
Accenture, Hewlett Packard Enterprise, IBM

5 RELATED WORK

Career path analysis is a hot topic in management and psychology fields due to its significant values for guiding the decision-making process of organizations as well as individuals. Those works were largely based on limited survey data and gave qualitative analyses of key factors that would influence one's career path [18, 25]. Recent years, AI technology has enhanced the development and re-designed the paradigm of human resource management in many aspects [17, 19, 23, 28, 32], of the area, career path analysis is one hot target problem. For example, Li et al. [13] designed a neural network framework to predict the next employer and positions together. Li et al. [12] proposed a survival analysis to model the promotion and turnover within one company, which is different from our trans-company analysis. Xu et al. [29] analyzed the talent flow into and out of the target organizations, regions, or industries.

The technologies used in our model are associated with recurrent neural networks, as well as sequential event data analysis. Various recurrent neural network approaches have been developed to address the time series problem, such as LSTM [7], and Gated Recurrent Unite (GRU)[5]. These techniques have been widely used due to their strong performance as well as the ability to capture long-term temporal dependencies, especially in the text mining and image recognition areas [4, 31]. After that, attention-based model are introduced to improve the prediction power of RNNs further [2, 15]. Recent years, sequential event data and survival analysis models have been developed to solve various problems [6, 16, 30]. Jing and Smola [8] applied RNN to model the user return pattern of a musician application. Ren et al. [21] proposed a deep learning model to analyze both censored and uncensored data. Our research is different from the above works in two aspects. First, we use a hierarchical LSTM and attention mechanism to model a hierarchical sequence data. Second, we do not suppose any preliminary assumptions on the form of hazard rate, as the preliminary assumptions may be against the true nature of the real values.

6 CONCLUDING REMARKS

In this paper, we focused on understanding job mobility at an individual level. Specifically, the goal is to predict the next potential employer of an individual and how long he/she will stay in the new position. Along with this line, we proposed a *hierarchical career-path-aware neural network* for answering these two questions. Our approach was designed to provide a certain level of interpretability by embedding the attention mechanism. As shown in our experimental results, our method provided much better accuracy for both prediction tasks. Finally, based on the assigned attention, we also provided data-driven evidence to show the importance of various factors (*e.g.*, job duration, firm type, *etc.*) for job mobility prediction.

ACKNOWLEDGMENTS

The research work was partially supported by the National Natural Science Foundation of China (No.91746301) and (No.71531001).

REFERENCES

[1] William J Anderson. 2012. *Continuous-time Markov chains: An applications-oriented approach.* Springer Science & Business Media.
[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
[3] Ann P Bartel. 1979. The migration decision: What role does job mobility play? *The American Economic Review* 69, 5 (1979), 775–786.
[4] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259* (2014).
[5] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).
[6] Nan Du, Hanjun Dai, Rakshit Trivedi, Utkarsh Upadhyay, Manuel Gomez-Rodriguez, and Le Song. 2016. Recurrent marked temporal point processes: Embedding event history to vector. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 1555–1564.
[7] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
[8] How Jing and Alexander J Smola. 2017. Neural survival recommender. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining.* ACM, 515–524.
[9] Alan Karr. 2017. *Point processes and their statistical inference.* Routledge.
[10] Kristen Keith and Abagail McWilliams. 1995. The wage effects of cumulative job mobility. *ILR Review* 49, 1 (1995), 121–137.
[11] John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. (2001).
[12] Huayu Li, Yong Ge, Hengshu Zhu, Hui Xiong, and Hongke Zhao. 2017. Prospecting the career development of talents: A survival analysis perspective. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 917–925.
[13] Liangyue Li, How Jing, Hanghang Tong, Jaewon Yang, Qi He, and Bee-Chung Chen. 2017. Nemo: Next career move prediction with contextual embedding. In *Proceedings of the 26th International Conference on World Wide Web Companion.* International World Wide Web Conferences Steering Committee, 505–513.
[14] Ye Liu, Luming Zhang, Liqiang Nie, Yan Yan, and David S Rosenblum. 2016. Fortune Teller: Predicting Your Career Path.. In *AAAI*, Vol. 2016. 201–207.
[15] Minh-Thang Luong, Ilya Sutskever, Quoc V Le, Oriol Vinyals, and Wojciech Zaremba. 2015. Addressing the rare word problem in neural machine translation. *ACL* (2015), 11–19.
[16] Hongyuan Mei and Jason M Eisner. 2017. The neural hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems.* 6754–6764.
[17] Qingxin Meng, Hengshu Zhu, Keli Xiao, and Hui Xiong. 2018. Intelligent Salary Benchmarking for Talent Recruitment: A Holistic Matrix Factorization Approach. In *2018 IEEE International Conference on Data Mining (ICDM).* IEEE, 337–346.
[18] Amalia R Miller. 2011. The effects of motherhood timing on career path. *Journal of population economics* 24, 3 (2011), 1071–1100.
[19] Chuan Qin, Hengshu Zhu, Tong Xu, Chen Zhu, Liang Jiang, Enhong Chen, and Hui Xiong. 2018. Enhancing Person-Job Fit for Talent Recruitment: An Ability-aware Neural Network Approach. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval.* ACM, 25–34.
[20] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks.* ELRA, Valletta, Malta, 45–50. http://is.muni.cz/publication/884893/en.
[21] Kan Ren, Jiarui Qin, Lei Zheng, Zhengyu Yang, Weinan Zhang, Lin Qiu, and Yong Yu. 2019. Deep Recurrent Survival Analysis. *AAAI* (2019).
[22] Rachel A Rosenfeld. 1992. Job mobility and career processes. *Annual review of Sociology* 18, 1 (1992), 39–61.
[23] Dazhong Shen, Hengshu Zhu, Chen Zhu, Tong Xu, Chao Ma, and Hui Xiong. 2018. A Joint Learning Approach to Intelligent Job Interview Assessment.. In *IJCAI.* 3542–3548.
[24] Robert H Topel and Michael P Ward. 1992. Job mobility and the careers of young men. *The Quarterly Journal of Economics* 107, 2 (1992), 439–479.
[25] Charles M Vance. 2005. The personal quest for building global competence: A taxonomy of self-initiating career path strategies for gaining business experience abroad. *Journal of World Business* 40, 4 (2005), 374–385.
[26] Bernd Wegener. 1991. Job mobility and social ties: Social resources, prior job, and status attainment. *American Sociological Review* (1991), 60–71.
[27] Huang Xu, Zhiwen Yu, Hui Xiong, Bin Guo, and Hengshu Zhu. 2015. Learning career mobility and human activity patterns for job change analysis. In *2015 IEEE International Conference on Data Mining (ICDM).* IEEE, 1057–1062.
[28] Huang Xu, Zhiwen Yu, Jingyuan Yang, Hui Xiong, and Hengshu Zhu. 2016. Talent circle detection in job transition networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 655–664.
[29] Huang Xu, Zhiwen Yu, Jingyuan Yang, Hui Xiong, and Hengshu Zhu. 2018. Dynamic Talent Flow Analysis with Deep Sequence Prediction Modeling. *IEEE Transactions on Knowledge and Data Engineering* (2018).
[30] Zeyang Ye, Lihao Zhang, Keli Xiao, Wenjun Zhou, Yong Ge, and Yuefan Deng. 2018. Multi-User Mobile Sequential Recommendation: An Efficient Parallel Computing Paradigm. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* ACM, 2624–2633.
[31] Liang Zhang, Keli Xiao, Hengshu Zhu, Chuanren Liu, Jingyuan Yang, and Bo Jin. 2018. CADEN: A Context-Aware Deep Embedding Network for Financial Opinions Mining. In *2018 IEEE International Conference on Data Mining (ICDM).* IEEE, 757–766.
[32] Chen Zhu, Hengshu Zhu, Hui Xiong, Pengliang Ding, and Fang Xie. 2016. Recruitment market trend analysis with sequential latent variable models. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* ACM, 383–392.

Appendix A

A.1 Position Normalization

The position names in the dataset are not standardized, they are not even in one kind of language. So we used multiple keywords matching method to normalize those names into 26 categories. The normalized types of positions are listed in Table 8.

A.2 Data Preprocessing

Our dataset contains sequential data with different length. In order to fit to the non-sequential models (*i.e.* Logistic regression, Decision Tree and Random Forest), we have to transform the input features into a vector with a fixed length. To deal with the problem, we used a *Bag-of-Companies* model, which is similar to the concept of *Bag-of-words*. We ignored the sequential information among companies, and only counted the occurrence number of each company, and calculated the cumulative duration in that company. Except for companies and durations, we only recorded the last values of the sequential features. The non-sequential features remained the same with what we used in HCPNN. At last, we concatenated all the features into one vector with a fixed length. In this way, we fit our sequential data to the non-sequential models.

A.3 HCPNN network configuration

We summarize the key dimensions and values of HCPNN in Table 9.

A.4 Baseline Setting

We summarize the details of baseline methods, especially the modified CRF and MHP methods as follows:

- **CTMC** [1]: It is a stochastic model to describe a series of events which the state spaces are discrete, yet the time is continuous. It also called the memoryless process, because the future states are solely dependent on the present state. This is a model that can predict the next state and the duration of the next state simultaneously. In our experiment, we set the state to be working in a specific company.
- **PP** [9]: It defines the occurrence probability of an event over a real-time line. When the instantaneous occurrence probability λ is a constant, we call it stationary or homogeneous Poisson Process, which we deployed in this paper.
- **CRF**[11]: It is an undirected graphical and discriminative model allowing long-distance dependencies and integration of rich features. The nodes in the graph denote the random variables, while the edges denote the direct influence or dependency relations between the variables. We used the linear conditional field in our experiment. We have

$$P(Y|X) = exp\left(\sum_{i,k} \lambda_k t_k(Y_{i-1}, Y_i, X, i) + \sum_{i,l} u_l s_l(Y_i, X, i)\right), \tag{13}$$

where $t_k(Y_{i-1}, Y_i, X, i)$ denotes the transition probability transferred from Y_{i-1} to Y_i at the sequence position (X, i), which corresponding the transfer probabilities from one company to another company. $s_l(Y_i, X, i)$ represents the probability of Y_i at the sequence position (X, i). λ_k and u_l are two weight coefficients for these two functions. In our problem, Y_i means a specific company in the position i. The train process is the same with what is broadly applied in NLP tasks [2], but the predicting process is different, since the viterbi algorithm will use the future information to deduce the historical sequence. To handle this problem, we used the original definition as described in Equation 13 to calculate the probabilities of next employer. More specifically, we used the parameters $s_l(Y_i, X, i)$, $t_k(Y_{i-1}, Y_i, X, i)$ learned from training process, combining with the known historical company sequence $\overrightarrow{Q(u)}$ to calculate the probabilities. We set the tunning parameters λ_k and u_l to be 1.
- **MHP**[16]: We used a multi-variable Hawkes process defined in [16] in this paper with modifications. It assumes the event intensity rate is not only caused by a self-excited rate μ, but also influenced by the events happened before, and the influence degree is proportional to the time span between the events and event types. Suppose a company $c \in C$, where C denotes the whole company set, and we have N companies in total. We want to simulate the instantaneous occurrence probability $\lambda(\tau)$ over time $(0, \infty)$ by training three parameters, namely self-excited intensity rate μ, the event influence parameter σ and the time decay parameter δ. To facilitate the understanding of MHP, we first summarize the notation descriptions and their dimensions in Table 10.
The key process of the algorithm can be described in equation:

$$\lambda_{c_{g+1}}(\tau) = \mu_{c_{g+1}}(\tau) + \sum_{i=1}^{g} \sigma_{c_i, c_{g+1}} \exp\left(-\delta_{c_i, c_{g+1}}((g+1) - i)\right), \tag{14}$$

where $\lambda_{c_{g+1}}(\tau)$ is the individual turnover probability when she works for her $(g+1)$-th company c_{g+1}. For every person, the turnover probability is influenced by two factors, one is the self-excited factor of $\mu_{c_{g+1}}(\tau)$, the other is all the employers he/she worked for before. The influence degree is controlled by the company type and the time index distance. The longer of the time distance, the weaker of the influence. The objective function is the maximize the log-likelihood of predicting duration as describe in Equation 10. When we get the $\lambda_{c_{g+1}}(\tau)$ for every person for the next company, we can use the survival analysis integration function described in Equation 12 to compute the expectation of the duration.

A.5 T-test for HCPNN and HCPNO

The results of standard student t-test on comparing HCPNN and HCPNO are summarized in Table 11.

[2]https://github.com/tensorflow/tensorflow/tree/master/tensorflow/contrib/crf

Table 8: Position name normalization.

Position Types				
Accounting	Sales	Administrative	Supporter	Consulting
Social Service	Engineering	Education	Entrepreneurship	Finance
Health Care	Human Resources	Information Technology	Law	Military
Marketing	Media	Operation	Real Estate	Purchaser
Product Management	Quality Assurance	Researcher	Program Management	Arts and Design
		Business Development		

Table 9: The network configuration of HCPNN.

Name	Dimension/value
duration embedding	10
company id embedding	50
position id embedding	5
company description words embedding	50
personal description words embedding	50
hidden states of company LSTM layer	150
hidden states of position LSTM layer	20
output dimension of local attention layer	20
output dimension of global attention layer	80
dropout probability	0.9
number of samples in a batch	64
the loss tuning parameter α	0.5
segmentation to compute the integration of survival analysis z	21

Table 10: Notation description in MHP.

Notation	Description
$i, g \in N^+$	the time index in the company sequence.
$c_i \in C$	a person's i-th employer in his/her career path.
$\lambda(\tau)$	the individual turnover probability.
$\mu \in \mathbb{R}^N$	the self-excited turnover probability
$\sigma \in \mathbb{R}^{NN}$	the intensity influence rate between pair-wised companies
$\delta \in \mathbb{R}+^{NN}$	the time decay parameter between pair-wised companies

Table 11: The results of standard student t-test with 95% confidence interval.

Model	HCPNN	HCPOP	p-value
Acc@1	0.0726 ± 0.0004	0.0712 ± 0.0003	1.3e-5
Acc@15	0.4039 ± 0.0009	0.3995 ± 0.0010	4.9e-8
Acc@30	0.5353 ± 0.0010	0.5308 ± 0.0009	4.3e-8
MRR	0.1555 ± 0.0004	0.1534 ± 0.0004	9.6e-7
MAE	2.7288 ± 0.0056	2.7357 ± 0.0043	5.7e-4
RMSE	3.8846 ± 0.0084	3.8925 ± 0.0073	1.0e-2