Extracting Job Title Hierarchy from Career Trajectories: A Bayesian Perspective

Huang Xu[1], Zhiwen Yu[1], Bin Guo[1], Mingfei Teng[2], Hui Xiong[2]
[1] Northwestern Polytechnical University, Xi'an, Shaanxi, 710129 China
[2] Rutgers University
{xuhuang@mail., zhiwenyu@, guob@}nwpu.edu.cn, {mingfei.teng, hxiong}@rutgers.edu

Abstract

A job title usually implies the responsibility and the rank of a job position. While traditional job title analysis has been focused on studying the responsibilities of job titles, this paper attempts to reveal the rank of job titles. Specifically, we propose to extract job title hierarchy from employees' career trajectories. Along this line, we first quantify the Difficulty of Promotion (DOP) from one job title to another by a monotonic transformation of the length of tenure based on the assumption that a longer tenure usually implies a greater difficulty to be promoted. Then, the difference of two job title ranks is defined as a mapping of the DOP observed from job transitions. A Gaussian Bayesian Network (GBN) is adopted to model the joint distribution of the job title ranks and the DOPs in a career trajectory. Furthermore, a stochastic algorithm is developed for inferring the posterior job title rank by a given collection of DOPs in the GBN. Finally, experiments on more than 20 million job trajectories show that the job title hierarchy can be extracted precisely by the proposed method.

1 Introduction

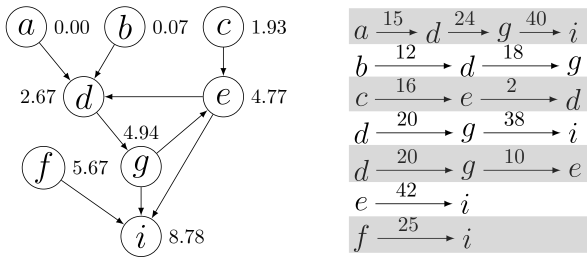
A job title is a short term in a few words which describes a job position held by an employee. It implies both the responsibility and the rank of the corresponding job position within an organization [Harper, 2015]. For instance, '*senior software engineer*' is a job title for a position with an engineering function in the software domain, and it has a higher rank than that of '*software engineer*'. Generally, the underlying rank of a job title denotes the credentials necessary to qualify for the corresponding job position, and a higher ranked job title is usually more difficult to get than a lower ranked one. In addition, the universe set of job titles in an organization naturally form a hierarchical directed graph, where nodes denote job titles and edges denote promotions, and the rank of a node denotes its level in the hierarchy. Figure 1a shows an example graph, where the alphabets represent job titles.

There have been thorough researches on modeling job titles with respect to the responsibilities [Bekkerman and Gavish, 2011; Javed *et al.*, 2016]. However, limited research attention has been paid to the hierarchy of job titles, whereas a properly defined hierarchy is desired in many applications, including career success evaluation [Heslin, 2005], career planning [Liu *et al.*, 2016], and job recommendation [Tang *et al.*, 2016; Li *et al.*, 2016]. For instance, during an interview, it is needed to assess the performance of candidates according to their working experience. Once the job title hierarchies of organizations that they worked for are presented, their performance can be evaluated by comparing their held job titles in resumes with the corresponding rank in the hierarchies to identify key indicators (e.g., early or late promotions).

In this paper, we attempt to explore a justifiable definition of job title hierarchy and hence design a robust method to establish the hierarchy. The core issue in the problem is determining the ranks of the job titles, but ranks are difficult to be detected from the name of job titles due to the ambiguity of phrases. For example, there are divergent titles for identical job positions or polysemous for distinctive job positions.

Fortunately, the universalized Online Professional Networks (OPNs) make a huge number of resumes publicly available. It provides a large collection of career trajectories, and each trajectory contains several job transitions, where a transition comprises two job titles and the tenure of the first job position till transiting to the second one. To this end, instead of analyzing the titles directly, we propose to infer the ranks of job titles by distilling their correlations with career trajectories. Particularly, from the perspective of career trajectories, a properly defined rank should have two essential properties. First, the relative ranks are consistent with the order of job titles in career trajectories, i.e., lower ranked job titles are usually followed by higher ranked ones. Second, the difference of two ranks should reflect the difficulty of being promoted from the former title to the latter one in a career.

However, finding the job title ranks possessing these two properties is nontrivial due to the wide variability of career moves, which commonly exist in workplaces, such as early promotion, lateral or even downward moves [Hess *et al.*, 2012]. This brings three challenges in revealing the ranks. First, it is hard to quantify the difficulty of being promoted to a job title, because the difficulty is an abstract concept that depends heavily on subjective cognition. Second, finding a common rank adaptive to individual diversity is difficult, because the difficulty of being promoted between identical job title pairs is different from person to person. Figure 1 illus-



(a) A job title hierarchy. (b) A set of job trajectories.

Figure 1: A job title hierarchy (a) detected from a set of job trajectories (b). The alphabets denote job titles. The numbers denote the rank of job titles and the length of tenures in (a) and (b), respectively.

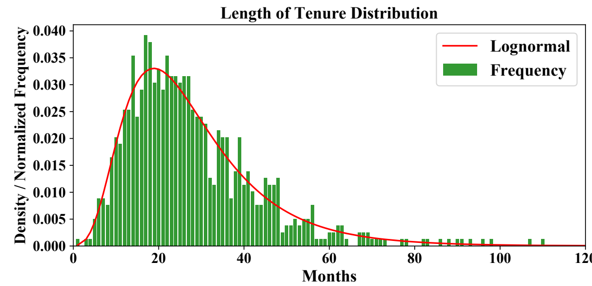


Figure 2: The distribution of the length of tenures between job titles *software engineer* and *senior software engineer* in an organization.

trates this phenomenon, where the numbers denote the Length of Tenure (LOT) (in Figure 1b) or ranks (in Figure 1a). To get promoted from d to g, people spend variate length of time (i.e., 18, 20 and 24 months), which indicates that the difficulty of a promotion is personal-dependent. Third, the solution should be adequately tolerant of conflicting trajectories since it will inevitably exist in reality. For instance, it is impossible to find ranks for d, g and e that are consistent with both trajectories ($d{\to}g{\to}e$) and ($c{\to}e{\to}d$), because it requires $d{<}e$ and $d{>}e$ simultaneously.

To address these challenges, we propose to adopt a Gaussian Bayesian Network (GBN) to extract the job title hierarchies. Specifically, the research contributes to the literature in a threefold way. First, we propose a novel measurement of the Difficulty of Promotion (DOP) based on the LOT in job trajectories. Our central premise is that the LOT reflects the DOP from one job position to another in a job transition. The DOP is quantified by a monotonic transformation of the LOT, based on the observation that the transformed LOT of a given pair of job titles can be well fitted by a normal distribution. Second, we represent collective and individual DOPs simultaneously by random variables to capture the macroscopic regularity and also the individual variations. Third, an efficient and conflict tolerant framework is presented to model the job title hierarchy. The joint distribution of job title ranks and DOPs in a career trajectory is modeled by a GBN, and an analytic form of the posterior distribution of ranks in the GBN is derived. Based on this, a computationally efficient algorithm is designed to detect the job title hierarchies.

The precision of the proposed method is evaluated on a dataset with 20 million job trajectories that are collected from a large OPN. The experimental results show that the proposed model could extract the job title hierarchy precisely comparing to the state-of-the-art benchmarks.

2 Job Title Hierarchy Model

In this section, we elaborate the insights observed from career trajectories, and then introduce the job title hierarchy model.

2.1 Measuring Difficulty of Promotion

The tenure of a job position is the time interval from the beginning timestamp of the position till transiting into a latter one in a career trajectory. A career trajectory contains one or more job transitions, where a transition comprises two job titles and the tenure of the former job position. We use t_{ij} to denote the LOT of a job position with job title i until transiting to a job position with job title j. In general, a longer t_{ij} usually implies a greater difficulty to get promoted from i to j. Thus the LOT provides an objective way to quantify the DOP between a pair of job titles.

However, the LOT of a given job title pair is hard to be described by a fixed number due to the wide variability of career performance. For instance, in a company, one employee may spend much shorter time than another to get promoted due to a better performance. The LOT usually has a different value in different career trajectories for an identical job title pair. Thus we treat t_{ij} as a random variable and describe it by its probability distribution to capture the underlying true value as well as its variability.

We have observed that, when measured by the number of months, the LOT follows a positively-skewed unimodal distribution. It means that most of the employees work on a job position for a short time period (e.g., around three years), and some of them work for an even shorter period (e.g., five months), while a minority of them keep a title for a longer period (e.g., ten years). Figure 2 shows the distribution of the LOT from job title '*software engineer*' to '*senior software engineer*' in our data sample. The LOT is concentrated in the interval from 10 to 30 months, with a less emphasis on five to ten months, while a long tail continues even after 100 months.

There are several well-known distributions that may fit the LOTs well, including Lognormal, Gamma, and Weibull [Wang *et al.*, 2013] distributions. Among these alternatives, we choose the Lognormal distribution for two reasons. First, it has optimal result in the goodness-of-fit test on the LOT data (see Section 4.2 for details). Second, it motivates an interpretable and computationally tractable job title hierarchy extraction model. In fact, because $log(t_{ij})$ follows a Gaussian distribution, the joint distribution of DOPs and job title ranks can be modeled by a GBN, in which the posterior distribution of job title rank has an analytic form.

Furthermore, we find out that, besides logarithm, other monotonic non-linear transformations, such as a n-th root transformation (i.e., $\sqrt[n]{t_{ij}}$,with $2 \leq n \leq 4$), can also make the transformed values of t_{ij} approximately normally distributed. In summary, if we use $g(\cdot)$ to denote one of these transformations, then

$$g(t_{ij}) \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2), \quad (1)$$

where parameters μ_{ij} and σ_{ij}^2 are the mean and variance, respectively. Because $g(\cdot)$ is a monotonic function of the LOT

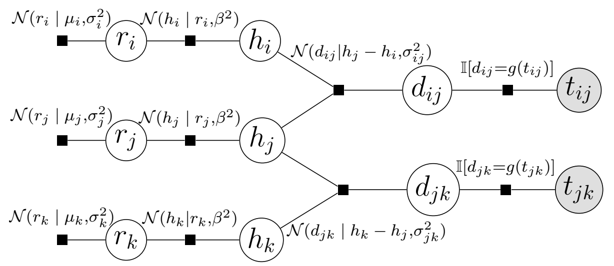


Figure 3: The factor graph of an example career trajectory. There are three job titles i, j and k in a queue with job title ranks r_i, r_j and r_k, correspondingly. Two job transitions (i, j) and (j, k) have length of tenures t_{ij} and t_{jk}, respectively.

and the LOT is an objective indicator of the DOP, we take $g(t_{ij})$ as the measure of DOP from job title i to j.

Definition 1. ***Difficulty of Promotion***. *The difficulty of promotion from job title i to j is $g(t_{ij})$, where t_{ij} is the LOT from i to j, and $g : \mathbf{t} \rightarrow \mathcal{R}$ is a monotonic function that maps a LOT to a random variable such that $g(t_{ij}) \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$.*

One interpretation of this definition is that the expected difficulty of being promoted from i to j is μ_{ij}, but it may vary around μ_{ij} with variance σ_{ij}^2 for a specific career. Besides, given job titles i, j and k, if μ_{ij} is larger than μ_{kj}, then the expected DOP from i to j is higher than that from k to j. Because the parameters are determined by the collection of tenures, we refer this difficulty to as *collective DOP*, which reflects the macroscopic regularity of the promotion process.

2.2 Bayesian Job Title Hierarchy Model

The primary goal of the model is to find an underlying rank r_i for a job title i such that the rank can reflect the difficulty to reach the title within a career. In fact, r_i cannot be observed directly, and only the relative ranks among job titles make sense since the ranks have no proper unit.

We reveal the relative ranks by distilling its correlations with career trajectories. Specifically, we assume that the difference of ranks essentially determine the DOP with some intermediate factors for two job titles in a career trajectory. For consistency, two requirements should be satisfied in the determination process. First, since DOP is normally distributed, the conditional distribution of the DOP given the ranks of two job titles should follow a Gaussian distribution, i.e.,

$$p(g(t_{ij}) \mid r_i, r_j) = \mathcal{N}(g(t_{ij}) \mid \mu_{ij}, \sigma_{ij}^2). \tag{2}$$

Second, given a pair of job titles r_i and r_j, the rank difference (i.e., $r_j - r_i$) should positively related to $g(t_{ij})$.

In addition, the universe set of job titles in an organization initiate a hierarchical directed graph, where nodes represent job titles, edges represent promotions, and the level of a node in the hierarchy is determined by the rank of the corresponding job title. Thus we define the job title hierarchy as follows:

Definition 2. ***Job Title Hierarchy***. *A job title hierarchy $G(V, E, \mathbf{r})$ is a directed graph with a hierarchy $\mathbf{r} := \{r_i | i \in V\}$, where $V = \{1, 2, \ldots, m\}$ is the set of job titles, and $E = \{(i, j)|i, j \in V\}$ contains (i, j) when there is at least one job transition from i to j. The rank $\mathbf{r}$ guarantees that $p(g(t_{ij}) \mid r_i, r_j) = \mathcal{N}(g(t_{ij}) \mid \mu_{ij}, \sigma_{ij}^2)$ for each $(i, j) \in E$.*

The job title hierarchy detection problem is revealing appropriate ranks $\mathbf{r}$ that satisfy the above requirements. To satisfy the first requirement, we assume that for a set of job titles, $\mathbf{r}$ has a factorizing Gaussian prior distribution, i.e.,

$$p(\mathbf{r}) = \prod_{i=1}^{m} \mathcal{N}(r_i \mid \mu_i, \sigma_i^2), \tag{3}$$

where μ_i denotes the expected rank of job title i, and σ_i^2 represents the variance of the rank. The priors could be interpreted as that the rank of a job title has an underlying true value, but it varies due to the wide variability of the promotion process. The value of σ_i^2 represents the confidence of our estimation, which will be described in Section 3.

For a specific employee, the actual difficulty of pursuing a job position usually depends on several personal-dependent factors, such as the work performance or the educational background. This makes the difficulty for an individual deviating from the collective DOP. Thus we use a Gaussian noise $a = \mathcal{N}(0, \beta^2)$ to introduce this variance into the model. Specifically, $h_i = (r_i + a)$ is the personalized rank of job title i for an individual.

$$p(h_i \mid r_i) = \mathcal{N}(h_i \mid r_i, \beta^2). \tag{4}$$

To satisfy the second requirement, the expectation of collective DOP (i.e., μ_{ij}) is set to as the difference of two personalized ranks, i.e.,

$$p(d_{ij} \mid h_i, h_j) = \mathcal{N}(d_{ij} \mid h_j - h_i, \sigma_{ij}^2), \tag{5}$$

where $d_{ij} = g(t_{ij})$ is the collective DOP from job title i to j. We refer $(h_j - h_i)$ to as the *individual DOP* because it reflects the difficulty of being promoted of an individual.

In summary, the model is a GBN [Kschischang *et al.*, 2001] specified by the conditional distributions in Equations 3, 4 and 5. An example of the model applied on a career trajectory is shown in Figure 3, where there are three job titles i, j and k in a queue in the career trajectory.

For clarity, we use symbols $\mathbf{h} := \{h_i \mid i \in V\}$ and $\mathbf{d} := \{d_{ij} \mid (i, j) \in E\}$ to represent the personal ranks and collective DOPs, correspondingly. The joint distribution of these variables in a career trajectory is denoted by $p(\mathbf{r}, \mathbf{h}, \mathbf{d})$. Without loss of generality, we arrange the random variables in a career trajectory by their subscripts in an natural order such that $(\mathbf{r}, \mathbf{h}, \mathbf{d}) = (r_1, \ldots, r_m, h_1, \ldots, h_m, d_{12}, \ldots, d_{m-1,m})$.

The joint distribution of $\mathbf{r}$, $\mathbf{h}$, and $\mathbf{d}$ can be derived by combining Equations 3, 4 and 5 according to Theorem 1 . Please refer to Section 2.3 in [Bishop, 2007] and Section 7.2 in [Koller and Friedman, 2009] for proof of the theorem.

Theorem 1. ***Conditional and joint distribution of linear Gaussian model***. *Let random variable y be a linear Gaussian of its parents $\mathbf{x}$, i.e., $p(y \mid \mathbf{x}) = \mathcal{N}\left(y \mid \theta^T\mathbf{x}, \sigma^2\right)$. Assume that $\mathbf{x} = (x_1, \ldots, x_k)$ are jointly Gaussian with distribution $\mathcal{N}(\mu, \Sigma)$. Then the distribution of y is a Gaussian distribution $p(y) = \mathcal{N}(y \mid \mu_y, \sigma_y^2)$, where:*

$$\mu_y = \theta^T\mu \quad \textit{and} \quad \sigma_y^2 = \sigma^2 + \theta^T\Sigma\theta. \tag{6}$$

The joint distribution $p(\mathbf{x}, y)$ is a normal distribution where

$$Cov[x_i; y] = \sum_{j=1}^{k} \theta_j \Sigma_{i,j}. \tag{7}$$

According to Theorem 1, the joint distribution is a multivariate Gaussian distribution and has the form of

$$p(\mathbf{r},\mathbf{h},\mathbf{d}) = \mathcal{N}(\mathbf{r},\mathbf{h},\mathbf{d} \mid \mu_J, \Sigma_J), \tag{8}$$

where the mean vector and the covariance matrix are

$$\mu_J = \begin{bmatrix}\mu^r \\ \mu^h \\ \mu^d\end{bmatrix} \quad \text{and} \quad \Sigma_J = \begin{bmatrix}\Sigma^{rr} & \Sigma^{rh} & \Sigma^{rd} \\ \Sigma^{hr} & \Sigma^{hh} & \Sigma^{hd} \\ \Sigma^{dr} & \Sigma^{dh} & \Sigma^{dd}\end{bmatrix}.$$

The components of the mean vector are $\mu^r = \mu^h = [\mu_1, \ldots, \mu_m]^T$ and $\mu^d = [(\mu_2 - \mu_1), \ldots, (\mu_m - \mu_{m-1})]^T$, where μ_i is defined in Equation 3. The components of the covariance matrix are $\Sigma^{hr} = \Sigma^{rh} = \Sigma^{rr} = diag(\sigma_1^2, \ldots, \sigma_m^2)$, $\Sigma^{hh} = diag(\sigma_1^2 + \beta^2, \ldots, \sigma_m^2 + \beta^2)$, and $\Sigma^{dd} = diag(\sigma_1^2 + \sigma_2^2 + 2\beta^2 + \sigma_{12}^2, \ldots, \sigma_{m-1}^2 + \sigma_m^2 + 2\beta^2 + \sigma_{m-1,m}^2)$. $(\Sigma^{dr})^T = \Sigma^{rd} \in \mathcal{R}^{m\times(m-1)}$, $(\Sigma^{dh})^T = \Sigma^{hd} \in \mathcal{R}^{m\times(m-1)}$ and

$$\Sigma_{i,j}^{dr} = \begin{cases}-\sigma_i^2 & j = i \\ \sigma_j^2 & j = i+1 \\ 0 & otherwise\end{cases}, \Sigma_{i,j}^{hd} = \begin{cases}-\sigma_i^2 - \beta^2 & j = i \\ \sigma_j^2 + \beta^2 & j = i+1 \\ 0 & otherwise\end{cases}.$$

3 Hierarchy Extraction

In this section, we first present the analytic form of the posterior distribution of $\mathbf{r}$ given corresponding DOPs in a career trajectory in the proposed model, and then present a stochastic job title rank extraction algorithm based on the posteriors.

3.1 Posterior Distribution Inference

We use $p(\mathbf{r},\mathbf{h} \mid \mathbf{d})$ to represent the posterior probability of $\mathbf{r}$ and $\mathbf{h}$ given $\mathbf{d}$ in a career trajectory. Based on the joint distribution in Equation 8, the posterior distribution can be derived by Theorem 2. Please refer to [Castillo and Kjærulff, 2003] for proof of the theorem.

Theorem 2. ***Conditionals of a Gaussian distribution***. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *have a joint multivariate Gaussian distribution with mean vector and variance matrix given by*

$$\mu = \begin{pmatrix}\mu^{\mathbf{x}} \\ \mu^{\mathbf{y}}\end{pmatrix} \quad and \quad \Sigma = \begin{pmatrix}\Sigma^{\mathbf{xx}} & \Sigma^{\mathbf{xy}} \\ \Sigma^{\mathbf{yx}} & \Sigma^{\mathbf{yy}}\end{pmatrix}$$

where $\mu^{\mathbf{x}}$ *and* $\Sigma^{\mathbf{xx}}$ *are the mean vector and covariance matrix of* $\mathbf{x}$. $\mu^{\mathbf{y}}$ *and* $\Sigma^{\mathbf{yy}}$ *are the mean vector and covariance matrix of* $\mathbf{y}$, *and* $\Sigma^{\mathbf{xy}} = (\Sigma^{\mathbf{yx}})^T$ *is the covariance of* $\mathbf{x}$ *and* $\mathbf{y}$. *Then the conditional probability of* $\mathbf{x}$ *given* $\mathbf{y}$ *is a multivariate Gaussian with mean vector* $\mu^{\mathbf{x}|\mathbf{y}=c}$ *and covariance matrix* $\Sigma^{\mathbf{x}|\mathbf{y}=c}$ *that are given by*

$$\mu^{\mathbf{x}|\mathbf{y}=c} = \mu^{\mathbf{x}} + \Sigma^{\mathbf{xy}}(\Sigma^{\mathbf{yy}})^{-1}(c - \mu^{\mathbf{y}}), \tag{9}$$

$$\Sigma^{\mathbf{x}|\mathbf{y}=c} = \Sigma^{\mathbf{xx}} - \Sigma^{\mathbf{xy}}(\Sigma^{\mathbf{yy}})^{-1}\Sigma^{\mathbf{yx}}. \tag{10}$$

According to Theorem 2, we have

$$p(\mathbf{r},\mathbf{h} \mid \mathbf{d}) = \mathcal{N}(\mathbf{r},\mathbf{h} \mid \mu_P, \Sigma_P), \tag{11}$$

where the mean vector and covariance matrix is calculated by substituting Equations 12, 13 and 14 into Equations 9 and 10.

$$\mu^{\mathbf{x}} = \begin{bmatrix}\mu^r \\ \mu^h\end{bmatrix} \quad , \quad \mu^{\mathbf{y}} = \mu^d. \tag{12}$$

$$\Sigma^{\mathbf{xx}} = \begin{bmatrix}\Sigma^{rr} & \Sigma^{rh} \\ \Sigma^{hr} & \Sigma^{hh}\end{bmatrix} \quad , \quad \Sigma^{\mathbf{xy}} = \begin{bmatrix}\Sigma^{rd} \\ \Sigma^{hd}\end{bmatrix}, \tag{13}$$

$$\Sigma^{\mathbf{yx}} = (\Sigma^{\mathbf{xy}})^T \quad , \quad \Sigma^{\mathbf{yy}} = \Sigma^{dd}. \tag{14}$$

Equation 10 shows that the posterior variance of $\mathbf{r}$ decreases when the number of observations on the job title increases. Thus the variance is the confidence of our estimation of $\mathbf{r}$, and a smaller variance indicates a higher confidence.

3.2 Hierarchy Extraction Algorithm

Based on the derived posterior distribution, we propose an algorithm to extract the hierarchy from job trajectories, as shown in Algorithm 1. Specifically, we use $\mathcal{T} = \{(i,j,t_{ij}) \mid i,j \in V, (i,j) \in E\}$ to denote all of the job transitions observed within an organization. The algorithm first calculates the variance of DOP (i.e., σ_{ij}^2) for each $(i,j) \in E$. It then iteratively updates the job title ranks by a *stochastic posterior updating* mechanism. In each iteration, the priors of ranks are set by an *adjacent prior assignment* strategy, and the posteriors are then calculated according to Equations 9 and 10.

Algorithm 1: Stochastic job title hierarchy extraction

Input : $\mathcal{T}, E, R = \emptyset, \beta^2, \mu_0, \sigma_0^2$
Output: R
for $(i,j) \in E$ **do**
  $\sigma_{ij}^2 \leftarrow variance[g(t_{ij})]$
for $(i,j,t_{ij}) \in \mathcal{T}$ **do**
  if $i \in R$ *and* $j \notin R$ **then**
    $\mu_j \leftarrow \mu_i; \sigma_j^2 \leftarrow \sigma_0^2$
  else if $i \notin R$ *and* $j \in R$ **then**
    $\mu_i \leftarrow \mu_j; \sigma_i^2 \leftarrow \sigma_0^2$
  else if $i \notin R$ *and* $j \notin R$ **then**
    $\mu_i \leftarrow \mu_0; \mu_j \leftarrow \mu_0; \sigma_i^2 \leftarrow \sigma_0^2; \sigma_j^2 \leftarrow \sigma_0^2$
  $\mu_i, \mu_j \leftarrow \mu^{\mathbf{r}|\mathbf{d}=g(t_{ij})}$
  $\sigma_i^2, \sigma_j^2 \leftarrow \Sigma^{\mathbf{r}|\mathbf{d}=g(t_{ij})}$
  $R \leftarrow R \cup \{i, j, \mu_i, \mu_j, \sigma_i^2, \sigma_j^2\}$

The *stochastic posterior updating* is an incremental strategy, which computes the posterior ranks for job titles in one job transition at a time. It has three advantages comparing to batched methods. First, it is conflict tolerant, because an unusual job transition only causes a small update of posterior and the impact is overwhelmed by principal transitions. Second, the ranks will be updated when new job transitions are available, which makes the algorithm adaptive when the organizational structure is adjusted. Third, it avoids calculating matrix inverse in Equation 9 and 10. In specific, $\mathbf{y}$ is a scalar in this strategy such that $(\Sigma^{\mathbf{yy}})^{-1}$ is implemented by $\frac{1}{\sigma_{\mathbf{y}}^2}$.

The *adjacent prior assignment* strategy is that when a job title i appears the first time, μ_i is assigned by the value of μ_j in current job transition. When both i and j are first appearing, the priors are set to as default values (i.e., μ_0, σ_0^2). This strategy is designed to rank a title when it is from an employee hired from outside (e.g, f in Figure 1a), since a title usually has a close rank with its adjacency title in job transitions.

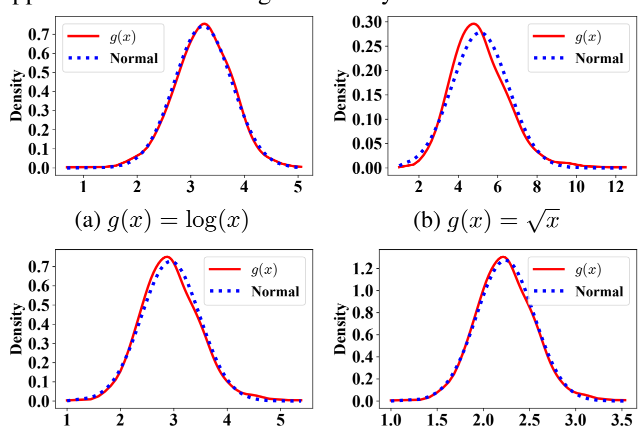
4 Experiments

4.1 Dataset

We have collected 20,243,120 digital resumes from a large OPN until October 2017. In each resume, there is a career trajectory that contains a list of job records, and each record consists of a job title, and a start and an end timestamp of the corresponding job position. After data cleaning, 10.2 million job transitions with complete information are extracted. Precisely, a job transition is a tuple (o, i, j, t_{ij}), where o, i, j and t_{ij} are the organization, the first job title, the second job title and the number of months (i.e., LOT), respectively. The hierarchies are extracted from 3.3 million job transitions among about 500,000 job titles, which are identified from top 10,000 organizations with the largest number of job transitions.

4.2 Normality Test of DOP

We elaborate the observation that the DOP (i.e., $g(t_{ij})$) can be well fitted by the Gaussian distribution, where $g(x)$ is one of the transitions in $log(x)$, $\sqrt{x}$, $\sqrt[3]{x}$ and $\sqrt[4]{x}$. Figure 4 shows an example of $g(t_{ij})$ and the fitted Gaussian distribution, where the data are the same with those in Figure 2. Specifically, the density of $g(t_{ij})$ is smoothed by the kernel density estimation method with the Gaussian kernel for clarity, and the Gaussian distribution is fitted by the maximum likelihood estimation method. It shows that the fitted Gaussian density is a close approximation to the original density.



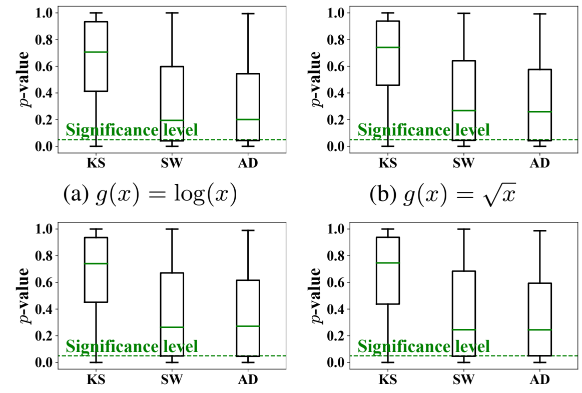
(a) $g(x) = \log(x)$ (b) $g(x) = \sqrt{x}$

(c) $g(x) = \sqrt[3]{x}$ (d) $g(x) = \sqrt[4]{x}$

Figure 4: An example $g(t_{ij})$ and the fitted Normal distribution.

To provide quantification evidence of the approximation, we test the distribution of the DOP by Shapiro-Wilk (SW), Kolmogorov-Smirnov (KS) and Anderson-Darling (AD) hypothesis tests [Razali and Wah, 2011]. These tests are designed to evaluate the goodness-of-fit of a distribution to a data sample, and they work well on evaluating whether a sample follows a Gaussian distribution (i.e., the normality) [Razali and Wah, 2011]. Specifically, these tests are conducted on top 1000 most frequently appeared job title pairs for data sufficiency, and the null hypothesis is that the sample comes from a normal distribution. Generally, the null hypothesis (i.e., normality) can not be rejected when the p-value is larger than a significance level.

Figure 5 shows the results of these normality tests on different $g(x)$, where the boxes denote the distribution of p-values and the dotted line shows the significance level of 0.05. It turns out that the p-values of KS test are far larger than the significance level, and the majority of p-values of SW and AD tests are larger than the significance level. The results suggest that, for a majority of tested job title pairs, the distribution of $g(t_{ij})$ can be well fitted by a Gaussian distribution.



(a) $g(x) = \log(x)$ (b) $g(x) = \sqrt{x}$

(c) $g(x) = \sqrt[3]{x}$ (d) $g(x) = \sqrt[4]{x}$

Figure 5: The results of three different normality tests on $g(t_{ij})$.

4.3 Validating Ordinal Concordance of Ranking

In the extracted job title hierarchy, the job titles in an organization are partially ordered regarding their ranks. A subset of the rank results is shown in Table 1, where the benchmarks include methods developed in TrueSkill™, Dot and Agony[1] (see Section 5). We use GBN to represent the proposed model, and only show the results of $g(x) = log(x)$ because four types of $g(x)$ have identical results regarding the orders. Hyperparameters are tuned on 5% of the dataset. The numbers in Table 1 are the mean (for GBN and Trueskill), the level (for Agony), or the normalized y-coordinate (for Dot).

We validate the ranks of job titles following typical paths in the graph by comparing them to a set of standard path templates. For example, (*consultant→senior consultant→manager→senior manager→partner*) is one of the standard templates for the job titles in Table 1. These templates are provided by human experts, and we collected 7,100 paths consists of 1.8 thousand job titles from online services[2].

The evaluating metrics are Kendall tau coefficient (τ) and Spearman's rank correlation coefficient (ρ). Specifically, both τ and ρ have high values when the compared two ranks are similar (or identical for a value of 1), and low when they are dissimilar (or fully different for a value of -1). The Normalized Discounted Cumulative Gain (NDCG) [Järvelin and Kekäläinen, 2002] is a widely used metric. We do not choose NDCG as a metric because it assumes that the items with higher ranks are more important to be right positioned, but there is no difference of importance in our situation.

Figure 6 shows the distribution of τ and ρ by comparing the results with the standard templates. Compared to the benchmarks, the proposed model achieves better performance regarding the majority values of τ and ρ. In addition, almost all of the paths have a τ and ρ larger than 0.85 and most of them are near 1. It indicates that the model generates nearly perfect job title orders on the majority of the template paths.

4.4 Validating DOP in Hierarchy

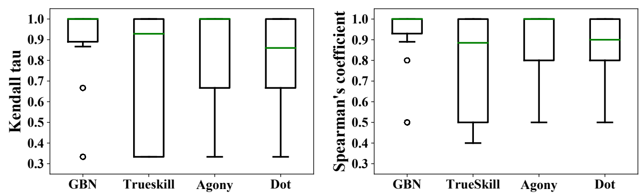
In addition to the orders, the difference of ranks should reflect the difficulty of being promoted. To validate this, we use the

[1] Acquired from http://trueskill.org/, https://www.graphviz.org/ and http://users.ics.aalto.fi/ntatti/software/agony.zip.

[2] For example, https://www.hierarchystructure.com

Comapny	Model	analyst	associate	auditor	sr. analyst	sr. auditor	sr. associate	consultant	sr. consultant	ass. manager	manager	sr. manager	director	partner
KPMG	**GBN**	**9.7**	**9.3**	**12.3**	**10.8**	**14.2**	**13.0**	**11.5**	**13.6**	**14.6**	**17.7**	**20.9**	**23.3**	**24.8**
	TrueSkill	8.6	11.1	12.4	13.6	20.3	19.2	16.6	20.0	23.9	26.7	27.1	27.5	28.6
	Agony	1	3	3	3	8	5	7	8	9	10	11	12	13
	Dot	4.0	3.0	6.0	6.0	9.0	6.0	8.0	9.0	10.0	11.0	12.0	13.0	14.0
EY	**GBN**	**13.4**	**12.5**	**13.2**	**16.1**	**16.2**	**16.2**	**13.8**	**17.0**	**18.4**	**20.4**	**23.9**	**26.1**	**29.6**
	TrueSkill	12.1	7.3	11.9	19.6	21.0	20.2	15.4	23.1	25.7	29.8	33.1	36.0	40.6
	Agony	11	3	3	12	13	12	13	14	15	16	17	19	20
	Dot	17.0	7.0	4.0	18.0	9.0	8.0	5.0	19.0	6.0	12.0	21.0	23.0	24.0
PwC	**GBN**	**15.1**	**15.1**	**15.8**	**19.0**	**18.8**	**18.9**	**16.1**	**19.3**	**21.4**	**23.1**	**26.3**	**29.3**	**32.6**
	TrueSkill	12.8	11.3	13.3	25.9	20.0	21.7	14.7	22.5	27.4	29.9	33.7	37.2	40.8
	Agony	4	5	9	5	10	11	8	10	12	13	17	18	19
	Dot	12.0	7.0	17.0	13.0	18.0	13.0	14.0	15.0	12.0	25.0	24.0	26.0	27.0
Deloitte	**GBN**	**5.4**	**7.7**	**8.0**	**9.0**	**11.4**	**10.1**	**9.1**	**12.7**	**12.8**	**16.5**	**20.2**	**23.8**	**25.8**
	TrueSkill	7.4	14.1	13.1	16.0	23.3	20.3	17.1	24.0	24.9	29.5	32.2	33.1	36.6
	Agony	1	7	6	2	8	8	8	10	9	12	13	15	16
	Dot	3.0	12.0	11.0	4.0	13.0	13.0	10.0	6.0	7.0	16.0	17.0	19.0	20.0

Table 1: The job title ranks of four models on several typical job titles from four companies in the accounting industry.



(a) Kendall τ. (b) Spearman's ρ.

Figure 6: The results of rank coefficients τ and ρ.

typical required years in job posting as an indicator for the real difficulty, and we take the difference of required years between two job titles as their DOP. We use the Root Mean Squared Error (RMSE) as a metric to evaluate the accuracy of our inference comparing to the real DOP. Specifically, $e = \frac{1}{|S|}||y_{ij} - g^{-1}(\mu_j - \mu_i)||_2$, where y_{ij} represents the required months from i to j for a set of job title pairs $(i, j) \in S$, and $g^{-1}(\cdot)$ in the benchmarks are identify functions.

	GBN $\log x$	GBN $\sqrt{x}$	GBN $\sqrt[3]{x}$	GBN $\sqrt[4]{x}$	TrueSkill	Agony	Dot
e	0.91	1.07	1.13	1.04	4.90	10.83	12.61

Table 2: The RMSE of the detected DOPs.

Table 2 shows the result of RMSE, where S contains 602 job title pairs which have direct promotional relations in 210 organizations. It shows that the proposed method has smaller RMSE in measuring the DOP compared to the benchmarks. While $\log x$ has smaller RMSE than other $g(\cdot)$s, different $g(\cdot)$s do not show significant impact on the result.

5 Related Work

There have been extensive studies focusing on job title classification tasks regarding job responsibilities [Bekkerman and Gavish, 2011; Javed *et al.*, 2016], but limited research efforts have been put on ranking job titles [Xu *et al.*, 2016; 2015]. From the application's perspective, organizational chart inference (OCI) [Zhang *et al.*, 2015] is similar to our problem, for it infers the reporting relations among employees according to their interactions in a social network. OCI differs from our work in that the difference of ranks is constant and the data source is social interaction. The method used in OCI is Agony, which will be introduced later.

There are several types of methods that are portable to our problem, including node ranking in graphs, hierarchy detection in social networks and player rating in competitions. Specifically, portable node ranking methods in graph include Topological Sorting (TS) [Kahn, 1962] and Layer Assigning (LA) in graph drawing [Ellson *et al.*, 2004]. However, they work only on a directed acyclic graph, but cycles exist in the job title hierarchy graph when there are downward career moves. To make TS and LA work, it is needed to remove a set of edges to make the graph acyclic, which is known as a Feedback Arc Set (FAS), and finding the minimum FAS is an NP-hard problem [Charbit *et al.*, 2007]. After the edge removal (usually by heuristics), TS and LA will assign a rank (or layer) to each node. LA (Dot) is a benchmark in our paper.

Social agony is a concept presented by Gupte et al. [Gupte *et al.*, 2011]. The authors assume that there is a rank for everyone in a social network, and it will cause a "social agony" when one person connects to a lower ranked person. The ranks are detected by integer programming. It differs from our situation in that the difference of ranks is constant. This method (i.e., Agony) is also applied to the OCI problem.

Some skill rating methods in competitions can be adapted to our problem by treating the job titles as competitors and the ranks as skills. A Bayesian skill rating system was proposed by [Herbrich *et al.*, 2007] and extended by [Dangauthier *et al.*, 2008] and [Guo *et al.*, 2012]. It differs from our problem in that a game has a binary output, which makes the posterior distribution of skills computationally intractable. Thus the posterior is inferred by the expectation propagation [Minka, 2001], which increases the computational complexity.

The above approaches treat the relative ranks as constants, and so spoil the natural difference of the DOPs, which makes it difficult to integrate LOTs into these methods.

6 Conclusion and Future Work

In this paper, we proposed a job title hierarchy extraction model. A novel measurement of the difficulty of promotion in careers was defined based on an observation about the length of tenure. A GBN was adopted to model both the job title ranks and the difficulty of promotion in career trajectories, and an efficient algorithm was designed to infer the posterior ranks in the model. The proposed method was validated by a real-world dataset. The model might be applicable to other hierarchy detection problems, where the difference of levels in the hierarchy can be quantitatively measured. We plan to map the job title ranks among organizations in the future.

Acknowledgments

This work was supported in part by the National Science Fund for Distinguished Young Scholars (No. 61725205), the National Basic Research Program of China (No.

2015CB352400), the National Natural Science Foundation of China (No. 91746301, 61332005, 61772428), and Baidu.

References

[Bekkerman and Gavish, 2011] Ron Bekkerman and Matan Gavish. High-precision phrase-based document classification on a modern scale. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 231–239. ACM, 2011.

[Bishop, 2007] C Bishop. Pattern recognition and machine learning, 1st edn. 2006. corr. 2nd printing edn. *Springer, New York*, 2007.

[Castillo and Kjærulff, 2003] Enrique Castillo and Uffe Kjærulff. Sensitivity analysis in gaussian bayesian networks using a symbolic-numerical technique. *Reliability Engineering & System Safety*, 79(2):139–148, 2003.

[Charbit *et al.*, 2007] Pierre Charbit, Stéphan Thomassé, and Anders Yeo. The minimum feedback arc set problem is np-hard for tournaments. *Combinatorics, Probability and Computing*, 16(1):1–4, 2007.

[Dangauthier *et al.*, 2008] Pierre Dangauthier, Ralf Herbrich, Tom Minka, and Thore Graepel. Trueskill through time: Revisiting the history of chess. In *Advances in Neural Information Processing Systems*, pages 337–344, 2008.

[Ellson *et al.*, 2004] John Ellson, Emden R Gansner, Eleftherios Koutsofios, Stephen C North, and Gordon Woodhull. Graphviz and dynagraph—static and dynamic graph drawing tools. *Graph drawing software*, pages 127–148, 2004.

[Guo *et al.*, 2012] Shengbo Guo, Scott Sanner, Thore Graepel, and Wray Buntine. Score-based bayesian skill learning. *Machine Learning and Knowledge Discovery in Databases*, pages 106–121, 2012.

[Gupte *et al.*, 2011] Mangesh Gupte, Pravin Shankar, Jing Li, Shanmugauelayut Muthukrishnan, and Liviu Iftode. Finding hierarchy in directed online social networks. In *Proceedings of the 20th international conference on World wide web*, pages 557–566. ACM, 2011.

[Harper, 2015] Charles Harper. *Organizations: Structures, processes and outcomes*. Routledge, 2015.

[Herbrich *et al.*, 2007] Ralf Herbrich, Tom Minka, and Thore Graepel. Trueskill$^{\mathrm{TM}}$: a bayesian skill rating system. In *Advances in neural information processing systems*, pages 569–576, 2007.

[Heslin, 2005] Peter A Heslin. Conceptualizing and evaluating career success. *Journal of Organizational behavior*, 26(2):113–136, 2005.

[Hess *et al.*, 2012] Narelle Hess, Denise M Jepsen, and Nicky Dries. Career and employer change in the age of the 'boundaryless' career. *Journal of Vocational Behavior*, 81(2):280–288, 2012.

[Järvelin and Kekäläinen, 2002] Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.

[Javed *et al.*, 2016] Faizan Javed, Matt McNair, Ferosh Jacob, and Meng Zhao. Towards a job title classification system. *arXiv preprint arXiv:1606.00917*, 2016.

[Kahn, 1962] Arthur B Kahn. Topological sorting of large networks. *Communications of the ACM*, 5(11):558–562, 1962.

[Koller and Friedman, 2009] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[Kschischang *et al.*, 2001] Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, 47(2):498–519, 2001.

[Li *et al.*, 2016] Jia Li, Dhruv Arya, Viet Ha-Thuc, and Shakti Sinha. How to get them a dream job?: Entity-aware features for personalized job search ranking. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 501–510, New York, NY, USA, 2016. ACM.

[Liu *et al.*, 2016] Ye Liu, Luming Zhang, Liqiang Nie, Yan Yan, and David S Rosenblum. Fortune teller: Predicting your career path. In *AAAI*, pages 201–207, 2016.

[Minka, 2001] Thomas P Minka. Expectation propagation for approximate bayesian inference. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 362–369. Morgan Kaufmann Publishers Inc., 2001.

[Razali and Wah, 2011] Nornadiah Mohd Razali and Yap Bee Wah. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1):21–33, 2011.

[Tang *et al.*, 2016] Liang Tang, Bo Long, Bee-Chung Chen, and Deepak Agarwal. An empirical study on recommendation with multiple types of feedback. In *KDD*, pages 283–292, 2016.

[Wang *et al.*, 2013] Jian Wang, Yi Zhang, Christian Posse, and Anmol Bhasin. Is it time for a career switch? In *Proceedings of the 22nd international conference on World Wide Web*, pages 1377–1388. ACM, 2013.

[Xu *et al.*, 2015] Huang Xu, Zhiwen Yu, Hui Xiong, Bin Guo, and Hengshu Zhu. Learning career mobility and human activity patterns for job change analysis. In *Data Mining (ICDM), 2015 IEEE International Conference on*, pages 1057–1062. IEEE, 2015.

[Xu *et al.*, 2016] Huang Xu, Zhiwen Yu, Jingyuan Yang, Hui Xiong, and Hengshu Zhu. Talent circle detection in job transition networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 655–664. ACM, 2016.

[Zhang *et al.*, 2015] Jiawei Zhang, Philip S Yu, and Yuanhua Lv. Organizational chart inference. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1435–1444. ACM, 2015.