The Impact of Person-Organization Fit on Talent Management: A Structure-Aware Convolutional Neural Network Approach

Ying Sun[1,2,3], Fuzhen Zhuang[1,3*], Hengshu Zhu[2*], Xin Song[2], Qing He[1,3], Hui Xiong[2,4]
[1]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing 100190, China [2]Baidu Talent Intelligence Center, Baidu Inc.
[3]University of Chinese Academy of Sciences, Beijing 100049, China [4]Business Intelligence Lab, Baidu Research
{sunying17g, zhuangfuzhen, heqing}@ict.ac.cn, {zhuhengshu, songxin06}@baidu.com, xionghui@gmail.com

ABSTRACT

Person-Organization fit (P-O fit) refers to the compatibility between employees and their organizations. The study of P-O fit is important for enhancing proactive talent management. While considerable efforts have been made in this direction, it still lacks a quantitative and holistic way for measuring P-O fit and its impact on talent management. To this end, in this paper, we propose a novel data-driven neural network approach for dynamically modeling the compatibility in P-O fit and its meaningful relationships with two critical issues in talent management, namely talent turnover and job performance. Specifically, inspired by the practical management scenarios, we first creatively design an Organizational Structure-aware Convolutional Neural Network (OSCN) for hierarchically extracting organization-aware compatibility features for measuring P-O fit. Then, to capture the dynamic nature of P-O fit and its consequent impact, we further exploit an adapted Recurrent Neural Network with attention mechanism to model the temporal information of P-O fit. Finally, we compare our approach with a number of state-of-the-art baseline methods on real-world talent data. Experimental results clearly demonstrate the effectiveness in terms of turnover prediction and job performance prediction. Moreover, we also show some interesting indicators of talent management through the visualization of network layers.

ACM Reference Format: Ying Sun, Fuzhen Zhuang, Hengshu Zhu, Xin Song, Qing He, Hui Xiong. 2019. The Impact of Person-Organization Fit on Talent Management: A Structure-Aware Convolutional Neural Network Approach. In *The 25th ACM SIGKDD Conference on Knowledge Discovery & Data Mining (KDD'19), August 4–8, 2019, Anchorage, AK, USA*. ACM, NY, NY, USA, 9 pages. https://doi.org/10.1145/3292500.3330849

1 INTRODUCTION

Nowadays, in the competitive and fast-evolving business environments, how to effectively attract and select right talents to the right jobs becomes a critical challenge for modern organizations. As a result, the study of Person-Organization fit (P-O fit), which refers to the compatibility between employees and their organizations [19], becomes a major focus in the fields of organizational behavior and talent analytics. According to the attraction-selection-attrition (A-S-A) theory, organizations tend to hire individuals with similar values during the selection process, and individuals are also attracted to such organizations [33]. Indeed, P-O fit has been widely regarded as an effective indicator for proactive talent management, which has significant impact on outcomes such as work attitudes, turnover intention and job performance [21]. For example, a good P-O fit usually indicates a strong organizational commitment and lower talent turnover rate and vice versa [1].

In the literatures, while considerable research efforts have been made in understanding P-O fit, and some interesting findings were revealed (e.g. people with high agreeableness usually match up better with a supportive organizational climate), most of these studies rely heavily on the surveys and classic statistical models [4, 14]. Therefore, there are various limitations when applying these methods for talent management. First, the metrics used in the surveys are usually designed in a subjective way, which cannot capture the compatibility features between talents and organizations in an objective and automatic manner. Second, the compatibility always changes over time due to the dynamic nature of P-O fit, which is hard to be tracked in traditional surveys. Third, it is extremely difficult to quantitatively model P-O fit and its impact on talent management using survey data due to the complex real-world management scenarios [31].

Fortunately, the newly available big talent data provide unparalleled opportunities for researchers to understand talent and organizational behaviors and gain tangible knowledge about P-O fit in a dynamic, quantitative, and objective way. To this end, in this paper, we propose a novel data-driven solution, namely P-O Fit Neural Network (POFNN), for dynamically modeling the compatibility in P-O fit and its meaningful relationship with two critical issues in talent management, namely talent turnover and job performance. Specifically, inspired by the practical management scenario, we first creatively design a learning structure named Organizational Structure-aware Convolutional Neural Network (OSCN) for hierarchically extracting organization-aware compatibility features in P-O fit, rather than manually designing talent and organizational profiles (e.g. culture and values). In particular, with the help of OSCN, the influence of organizational environment (e.g., the influence from superior, peers and subordinates) can be holistically modeled. Then, to capture the dynamic nature of P-O fit and its consequent impact, we further exploit an adapted Recurrent Neural Network with attention mechanism to model the temporal information of P-O fit. Finally, we evaluate our approach with a number

*Fuzhen Zhuang 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/10.1145/3292500.3330849

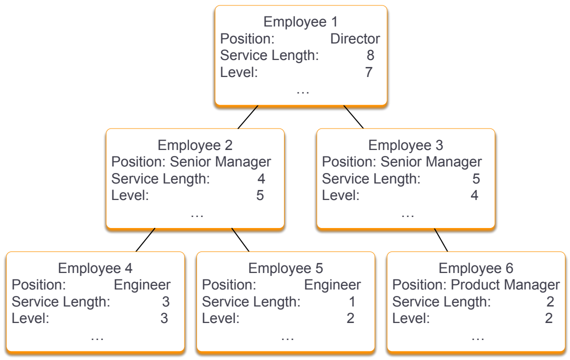


Figure 1: An example of the organization tree.

of state-of-the-art baseline methods on real-world talent data. Experimental results clearly demonstrate the effectiveness in terms of turnover prediction and job performance prediction, as well as some interesting indicators of talent management through the visualization of network layers. Moreover, our study reveals some management insights for enhancing proactive talent management. For instance, the results show that superiors usually have a great influence on the turnover intention of their subordinates.

Specifically, the major contributions of this paper can be summarized as follows:

- To the best of our knowledge, this paper is the first attempt of leveraging advanced data mining approaches for exploring the impact of P-O fit on talent management, which is a major focus of organizational behavior and talent analytics.
- We propose a novel data-driven neural network approach for dynamically modeling the compatibility in P-O fit and its meaningful relationship with two critical issues in talent management, namely talent turnover and job performance.
- We evaluate our approach with extensive experiments on real-world talent data. The results reveal some interesting insights for talent management and clearly validate the effectiveness of our approach in terms of helping turnover prediction and job performance prediction.

2 DATA DESCRIPTION

In this paper, a set of anonymized in-firm data of employees were collected from a high tech company in China, across a timespan of 45 months, ranging from 2015 to 2018.

Specifically, in the dataset, each employee has some static profiles (e.g., education background and previous employers), dynamic profiles that may change over days (e.g., reporting line, job position and level) and behavior records (e.g., communication network). In particular, from the reporting lines, the organization structure can be formulated as a tree structure, where each node represents an employee, as shown in Figure 1. Indeed, all the organizational relationships between employees can be found from the organization tree. For example, for each employee on the tree, her superior is the parent node, subordinates are the child nodes, and peers are the sibling nodes. Intuitively, the organization tree will change over time due the personnel changes in the company, such as the turnover, work transfer and promotion. In addition, the historical turnover and job performance records of employees are also contained in the dataset, which can help us to study the impact of P-O fit on consequent outcomes in talent management.

Table 1: Notifications used in this paper.

Notification	Explanation
N_t	Length of the sequential data
N_e	Number of employees
X^t	Feature matrix at the $t\text{-}th$ interval
A^t	Organization structure at the $t\text{-}th$ interval
I^t	Indicator for target employee at the $t\text{-}th$ interval
$\mathcal{X}$	Sequence of feature matrix
$\mathcal{A}$	Sequence of adjacency matrix
P_o^t	Environment profile matrix at the $t\text{-}th$ interval
P_e^t	Employee profile matrix at the $t\text{-}th$ interval
po^t	P-O fit representation vector at the $t\text{-}th$ interval
d_f	Dimension of feature vector
d_e	Dimension of employee profile
d_o	Dimension of environment profile
d_{po}	Dimension of P-O fit representation
d_l	Dimension of LSTM output
D_{per}/D_{sub}	Variance vector of peers/subordinates
M_{per}/M_{sub}	Average vector of peers/subordinates
F	Input feature vector of OSCN layer
U_{per}/U_{sub}	Aggregated vector of peers/subordinates
d_{in}	Dimension of OSCN layer input
d_{out}	Dimension of OSCN layer output

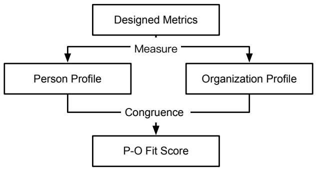
3 PROBLEM FORMULATION

In this paper, our goal is to temporally model P-O fit by anlayzing the latent relationship between organizational environment and employees, and to explore its applications on talent management. Therefore, given an organization tree, we define an adjacency matrix $A = [a^1, \cdots, a^{N_e}]$ to represent its structure, where $a^i \in \mathbb{R}^{N_e}$ indicates the $i\text{-}th$ node's parent node, a^i_j is equal to 1 if the $j\text{-}th$ node is the $i\text{-}th$ node's parent node, otherwise equal to 0. Meanwhile, we define the feature matrix of an organization as $X = [x^1, \cdots, x^{N_e}]$, where $x^i \in \mathbb{R}^{d_f}$ denotes the feature vector of the $i\text{-}th$ employee, d_f is the dimension of the feature vectors. As mentioned above, the structure of an organization and the profiles of the employees will change over time. So we define two sequential data of length N_t, written as $\mathcal{X} =< X^1, \cdots, X^{N_t} >$ and $\mathcal{A} =< A^1, \cdots, A^{N_t} >$, where X^t and A^t denote the feature matrix and adjacency matrix at the $t\text{-}th$ time interval, respectively.

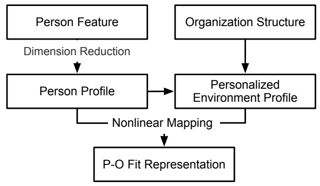
So far, we can formulate the problem of this study as learning a model M from historical data $\mathcal{A}$ and $\mathcal{X}$ to predict some future outcomes. Specifically, each sample at the $t\text{-}th$ time interval in the training set can be formulated as (X^t, A^t, I^t), where $I^t \in \mathbb{R}^{N_e}$ is the indicator for the target employee. The model is to predict y, which is a classification label indicating the outcomes, e.g., turnover or not. Without loss of generality, here we formulate the problem as a binary classification problem, i.e., $y \in \{0, 1\}$.

4 MODEL

In this section, we introduce the technical details of P-O fit Neural Network (POFNN). Specifically, we first describe the motivation and framework overview of POFNN. Then, we introduce the novel network structure Organizational Structure-aware Convolutional Neural Network (OSCN), which can extract environmental information from the organization's structure and the profiles of employees within. The notifications can be found in Table 1.



(a) A classic P-O fit modeling process.



(b) Our P-O fit modeling process.

Figure 2: Overview of P-O fit modeling.

4.1 P-O Fit Modeling

In this part, we describe the main idea of P-O fit modeling, which is also the motivation of our network. As mentioned in Section 1, P-O fit means the compatibility between employees and their organizations. Indeed, good P-O fit occurs when the characteristics of both employees and their organizations fit well. In the literature of organizational behavior, there are works on comparing person profile with organization profile to measure P-O fit [5]. The process is shown in Figure 2(a). In these works, metrics manually designed by experts are collected respectively from questionnaires for employees and organizations. Then statistical approaches are used to measure a P-O fit score as the congruence of these profiles. However, due to high cost of this process, it is difficult to apply it on real-world applications.

Therefore, it motivates us to automatically extract the profiles and model P-O fit with a neural-network based approach. The main process of our P-O fit modeling is shown in Figure 2(b). Instead of manually designing and measuring the metrics to form person profile, POFNN exploits features which can be automatically collected from employees' in-firm data, as is described in Section 3. Then a simple dimension reduction process is conducted for generating person profile from these features. And instead of directly measuring the organization profile with metrics, we suppose that the organizational environment is impacted by the employees within. Therefore, we obtain organization profile by combining the organization's structure with the profiles of the employees. Specially, instead of measuring an overall organization profile, we extract an unique environment profile for each employee, which varies according to the employees' positions. From the perspective of a certain employee, her organizational environment is usually influenced more by the closer colleagues, especially when the organization is big. Indeed, two employees in an organization may have little impact on each other if they are distant on the organization tree. And at last, instead of obtaining a P-O fit score with simple statistical approaches, we propose to use deep neural network to achieve a more complicated mapping from person and environment profiles to a P-O fit representation. We use this representation vector to predict outcomes in talent management. Since in this way more latent aspects of P-O fit can be modeled, we can achieve better prediction performance than simply using the P-O fit score.

4.2 Network Structure

Here we describe the structure of POFNN in detail, which is shown in Figure 3. The network has a recurrent structure for sequential data modeling. At each time interval, the feature matrix X^t and the organization structure matrix A^t are processed to form a P-O fit representation vector. First, a fully connected layer reduces the dimension of the input feature vectors and extracts the employee profiles $P_e^t \in \mathbb{R}^{d_e \times N_e}$, whose i-th column can be formulated as $\sigma(W_1^T X^t_{:,i} + b_1)$, where $W^1 \in \mathbb{R}^{d_f \times d_e}$ and $b_1 \in \mathbb{R}^{d_e}$ are the paramters, $\sigma(\cdot)$ represents the activation function. Then our proposed Organizational Structure-aware Convolutional Neural Network (OSCN) extracts the environment profiles from the organization structure as well as employee profiles, written as

$$P_o^t = OSCN(A^t, P_e^t).$$

As mentioned before, here each employee is attached with an unique environment profile representation, i.e., $P_o^t \in \mathbb{R}^{d_o \times N_e}$. Details of OSCN will be discussed in the next subsection. Then we select the target employee's environment profile and person profile with indicator vector I^t, written as

$$p_o^t = P_o^t I^t, \quad p_e^t = P_e^t I^t.$$

Then a deep neural network (DNN), which is composed of several fully connected layers, nolinearly maps them to form a P-O fit representation $po^t \in \mathbb{R}^{d_{po}}$, formulated as

$$po^t = DNN([p_o^t | p_e^t]),$$

where $[\cdot|\cdot]$ denotes concatenating two vectors of dimension d_1 and d_2 to form a vector of dimension $d_1 + d_2$. Then we have a sequence $PO = [po^1, \cdots, po^{N_t}]$, which contains P-O fit vectors of all the time intervals, we use Long Short-Term Memory (LSTM) to process them. LSTM outputs a status for each time interval with the historical information and the current P-O fit vector, written as

$$h^t = LSTM(po^t, h^{t-1}).$$

Considering the fact that some of the time intervals may have higher impact on the outcome (e.g., a sudden organizational change occuring in some month may cause employees' turnover intention), we use attention mechanism to catch those abnormal time intervals. The attention layer regard h^{N_t} as query and other LSTM outputs as keys. Then it adds a weight to each time interval (except the last one) indicating its importance, and then gets the weighted average of the time intervals, i.e.,

$$c_i = \mathrm{softmax}(\sigma(w_2^T [h^i | h^{N_t}] + b_2)),$$
$$o = \sum_{t=1}^{N_t - 1} c_t h^t,$$

where $w_2 \in \mathbb{R}^{2d_l}$ and $b_2 \in \mathbb{R}$ are the parameters. The output $o \in \mathbb{R}^{d_l}$ implicates the historical P-O fit information, we concatenate it with h^{N_t}, which contains the most recent P-O fit information. Then we map it with fully connected layer and sigmoid function to form a soft classification ranging from 0 to 1, written as

$$p = \mathrm{sigmoid}(w_3^T [o | h^{N_t}] + b_3).$$

4.3 Details for OSCN

Here we introduce the details for OSCN. As mentioned above, organizational environment may be different for employees in different positions even within the same organization. Therefore, we need a structure to extract organizational environment features for the employees according to their close colleagues on the organization tree.

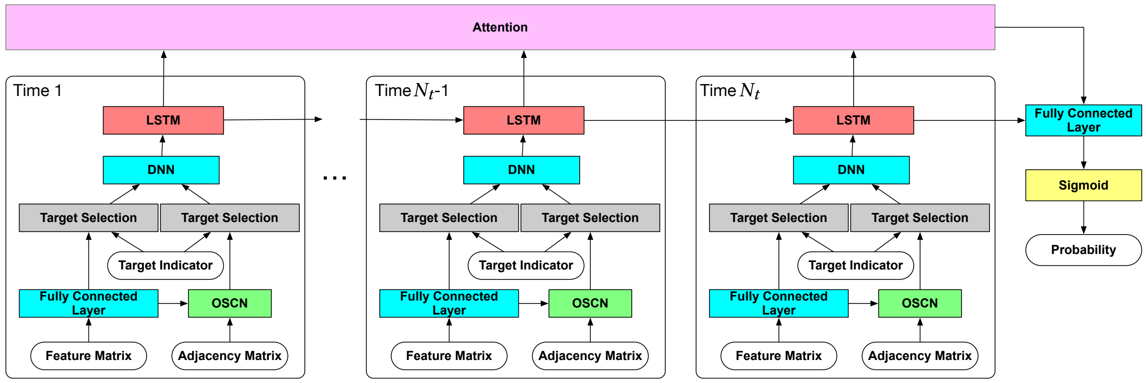


Figure 3: The overview of POFNN, which consists of both OSCN and LSTM structures with attention mechanism.

By simply regarding the organization tree as a graph, a straightforward idea is to apply Graph Convolutional Network (GCN) [18], which can pass message to the nodes from their neighboring nodes and output a vector containing the local information for each node. However, in a common GCN, the impact of different nodes depend mainly on their distance to the target position, thus GCN fails to distinguish different working relationships in an organization. For example, the influence on an employee's environment from her superior and her subordinates may be considered the same as they are all 1-step distant from her on the organization tree.

Based on the above consideration, we design OSCN to extract employees' organizational environment information, which can treat working relationships differently in convolution. Similar with the classic convolutional networks, the main idea of OSCN is weight sharing, and the weight of a node in the kernel lies on its relationship with the target. In our work, centered on one employee, we suppose that her superior, her subordinates and her peers give direct impact on her local organizational environment. So ideally, the convolution can be formulated as

$$P_o = W_{self}^T u_{self} + W_{sup}^T u_{sup} + W_{sub}^T u_{sub} + W_{per}^T u_{per},$$

where W_- represent the weights, and u_- represent the feature vectors of the colleagues. However, different from traditional convolutional networks that mainly focus on grid-like data, where the

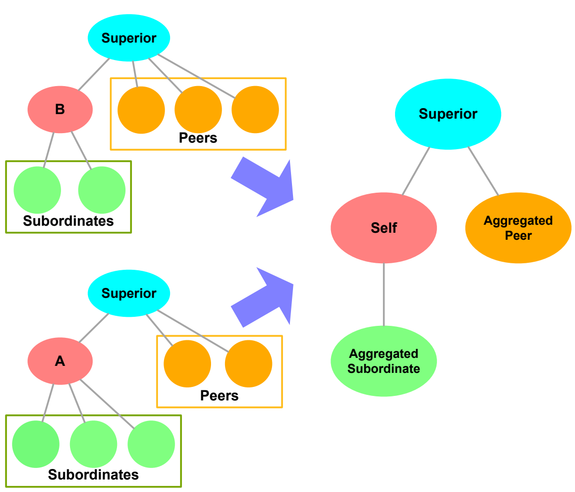


Figure 4: Schematic diagram of transformation to regular structure by node aggregation.

kernel can be simply formed as a matrix, the organization tree has a more irregular structure. The number of subordinates of each employee may be different, which brings difficulty to choosing a kernel that is applicable to all the positions on the tree.

To solve this problem, we propose to aggregate the employees with the same relationship. Specifically, supposing that all the subordinates of an employee have a similar impact on her, we aggregate them by calculating their average and variance vectors. And the same process is done on the peers. An example is given in Figure 4, where node A and node B have different local structures, but after the aggregating process, they are transformed into a same structure consisting of a superior node, an aggregated subordinate node and an aggregated peer node. Then we can apply a same kernel on them. For simplicity, here we use $F \in \mathbb{R}^{d_{in} \times N_e}$ to denote the input feature matrix of the OSCN layer, and use $A \in \mathbb{R}^{N_e \times N_e}$ to denote the adjacency matrix. Specifically, the average vectors of subordinates and peers, written as M_{sub} and M_{per}, can be formulated as

$$M_{sub} = \frac{FA^T}{e_1 e_2^T A^T}, \quad M_{per} = \frac{F(A^T A - \mathbb{1})}{e_1 e_2^T (A^T A - \mathbb{1})},$$

where each element in $e_1 \in \mathbb{R}^{d_{in}}$ and $e_2 \in \mathbb{R}^{N_e}$ is equal to one and $\mathbb{1}$ denotes the identity matrix. And the variance vectors, written as D_{sub} and D_{per} respectively, can be formulated as

$$D_{sub} = \sqrt{\frac{(F - M_{sub}A)^2 A^T}{e_1 e_2^T A^T}},$$

$$D_{per} = \sqrt{\frac{(F - M_{sub}A)^2 (A^T A - \mathbb{1})}{e_1 e_2^T (A^T A - \mathbb{1})}},$$

where $\sqrt{\cdot}$ and $\cdot^2$ are both element-wise operations. Then we get the aggregated feature vectors, written as $U_{sub}, U_{per} \in \mathbb{R}^{d_{out} \times N_e}$, with a fully connected layer, which can be formulated as

$$U_{sub} = \sigma(W_4^T [M_{sub}^T | D_{sub}^T]^T + b_4),$$

$$U_{per} = \sigma(W_4^T [M_{per}^T | D_{per}^T]^T + b_4),$$

Then we can apply convolution on the tree and get the output as

$$P_o = \sigma(W_{self}^T F + W_{per}^T U_{per} + W_{sub}^T U_{sub} + W_{sup}^T FA),$$

whose *i-th* column is the environment profile of the *i-th* employee.

The structure of an OSCN layer is shown in Figure 5. For simplicity, we only show the convolution of one node (yellow). In OSCN, the subordinates (green) and peers (orange) are aggregated to new

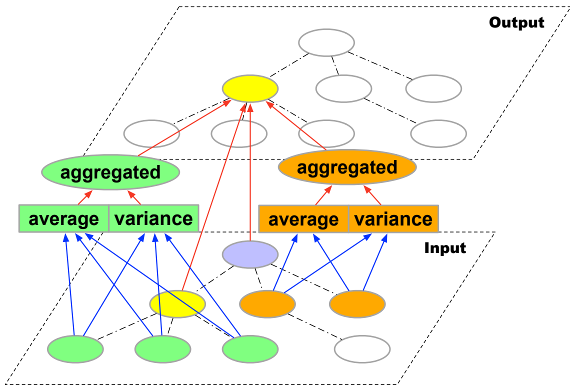


Figure 5: The schematic diagram of the OSCN.

feature vectors respectively. The convolution is done on each node and the output is a tree with the same structure as the input, where each node has a vector of local environment representation. It is notable that though we only use four kinds of relationships on the tree in this work (i.e., self, sibling, child and parent). The kernel can be easily extended according to the demand, e.g., cousin.

Finally, we stack several OSCN layers and use the output of the previous layer as the input of the next layer. The first output contains environmental information impacted from the peers, the subordinates and the superior. And the outputs of the deeper layers contain information from a broader scope of the organization, for example, from the superior of the superior, the peers of the superior, etc. We concatenate the output of all the layers as the final environment profile, which is shown in Figure 6.

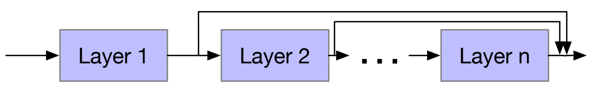


Figure 6: The schematic diagram of the stacked OSCN.

5 EXPERIMENTS

In this section, we evaluate the performance of POFNN on a real-world dataset. Specifically, we focus on two critical outcomes of talent management in our experiment: turnover and job performance. In turnover prediction, we predict if or not an employee will leave the company in the future. And in job performance prediction, we predict the employees' full-year performance rating.

5.1 Experimental Setup

5.1.1 Datasets. Details for the real-world dataset used in this paper can be found in Section 2. To avoid noise, we first filtered out the employees who stayed for less than one year in the company. Then, we regarded each month as a time interval and extracted features of each employee in each month. From the in-firm data, we extracted features like average frequency of communications with the superior in the month and job level at the end of the month. For simplicity, the organization trees are also extracted from the reporting lines at the end of the month. And to accelerate the training process, instead of using the whole organization tree for each sample, we extract a subtree consisting of her closest 100 employees. Then we conducted the following preprocessing for the two tasks respectively:

- **Turnover Prediction.** For each month, we chose some employees who will leave in the next 2 months as positive samples and some of the others as negative samples. Since the dataset is extremely imbalanced, we only sampled parts of instances with a ratio of 1:2. Then we formed sequential data consisting of organization trees and features within the 10 months untill the current month. Our training set contains samples with the last observed month between 2016.01 and 2017.12, and the test set contains samples with the last observed month between 2018.03 and 2018.06. To avoid information leaking, we removed training samples between 2018.01 and 2018.02, which may have overlap with test data.
- **Performance Prediction.** In each year from 2015 to 2017, we selected some employees who had annual performance appraisal. The performance can be clustered into 3 levels: excellent, normal and poor. We formed the sequence with features and organization structures in the first 10 months of the corresponding year. Specifically, we formulated two tasks of binary classification: 1) excellent performance prediction, where a sample is labeled as positive if the performance is excellent, and 2) poor performance prediction, where a sample is labeled as positive if the performance is poor. In each task, the training set contains samples for 2015 and 2016, and the test set contains samples for 2017.

Table 2: The statistics of datasets.

Dataset		#positive	#negative
Excellent Performance	Training	7,796	18,039
	Test	4,067	9,239
Poor Performance	Training	3,999	21,836
	Test	2,059	11,247
Turnover	Training	16,204	38,533
	Test	4,536	7,638

The statistics of datasets are shown in Table 2.

5.1.2 Baseline Methods. The baselines include four parts:

- Classic classification models including Decision Tree (DT), Logistic Regression (LR), Random Forest (RF) and Support Vector Machine (SVM). Since these methods are not designed for processing sequential data, we concatenated the sequential feature vectors into one feature vector. We also tried training with their averaged feature vectors, but the performance was far worse than concatenating them, due to page limitation, we did not display these results in this paper.
- Hidden Markov Model (HMM), which is a statistical model, it assumes the that the sequential data follows a Markov process with hidden states.
- Recurrent Neural Network (RNN), where we processed the feature vector of each time interval with DNN, then the sequential output was processed by LSTM with attention layer, which is the same as POFNN.
- Classic Graph Convolutional Network (GCN) [18], which is designed for extracting information from graphs. For the fairness of comparison, we replaced OSCN with GCN in POFNN and remained other settings the same.

5.1.3 Network Configuration. The network configuration can be found in Table 3. We tuned the parameters of the models with experiments and we found that slight changes on the parameters did not affect much on the performance, and we used the same configuration for both turnover prediction and performance prediction.

Table 3: The network configuration of NN based models.

Model	Name	Value
POFNN	DNN depth	3
	OSCN depth	2
	#Unit in each DNN layer	10
	#Unit in dimension reduction layer	12
	Ouput's dimension (d_{out}) of each OSCN layer	12
GCN	DNN depth	3
	GCN depth	2
	#Unit in each DNN layer	10
	#Unit in dimension reduction layer	12
	Ouput's dimension (d_{out}) of each GCN layer	20
RNN	DNN depth	5
	#Unit in each DNN layer	12

Since OSCN has more parameters than GCN, GCN needs its output vector of each layer to have higher dimension to get enough model complexity. And RNN needs lower model complexity than POFNN and GCN, because it only uses personnel features for prediction. Adam optimizer [17] is used for training.

5.1.4 Evaluation Metrics. We evaluated model performance by Cross Entropy (CE) and Area Under Curve (AUC). Cross Entropy can evaluate the distance between the predictions and the real labels, lower value of CE indicates higher ability on estimating the real distribution. AUC can measure classification performance under different thresholds, higher value of AUC indicates better overall performance. In particular, we used Area Under ROC Curve (ROC-AUC) and Area Under Precision-Recall Curve (PR-AUC) to evaluate the model performance.

5.2 Overall Performance

In this part, we introduce the overall evaluation of our approach in terms of turnover prediction and job performance prediction.

5.2.1 Evaluation on Turnover Prediction. The experimental results of turnover prediction are listed in Table 4. From the results, we can get the following observations. First, POFNN outperforms the baseline methods in terms of all the metrics. Second, the Neural Network (NN) based models outperform the classic models, because they can achieve more complicated modeling. Third, GCN and POFNN outperform the other models that only use personnel feature for prediction, indicating the effectiveness of P-O fit modeling on turnover prediction. Fourth, POFNN outperforms GCN in this task, indicating that OSCN is better at extracting environmental features from the organization tree. Last, though HMM can model sequential data, this generative model is not good at handling the complex observations in our work, thus performs even worse than most of the basic models.

5.2.2 Evaluation on Job Performance Prediction. We formulated job performance prediction as two binary classification tasks, namely excellent performance prediction and poor performance prediction. The results are demonstrated in Table 5 and Table 6 respectively. There are several observations. First, consistent with the result in turnover prediction, POFNN outperforms others in terms of all the metrics, which indicates that POFNN is an adapted model and can be applied on various applications in talent management. Second, while POFNN outperforms those methods, GCN performs almost the same as simple RNN. It indicates again that the environmental features extracted by GCN is not as good as those features extracted by OSCN in this task. Last, HMM still can not make good predictions, which is consistent with the result in turnover prediction.

Table 4: Performance on turnover prediction.

Model	CE	ROC-AUC	PR-AUC
Logistic Regression	0.830	0.584	0.522
Decision Tree	15.70	0.528	0.569
Random Forest	1.016	0.578	0.519
Support Vector Machine	0.738	0.569	0.495
Hidden Markov Model	14.53	0.538	0.524
Recurrent Neural Network	0.604	0.690	0.543
Graph Convolutional Network	0.592	0.708	0.575
P-O Fit Neural Network	**0.587**	**0.727**	**0.603**

Table 5: Performance on performance prediction (excellent).

Model	CE	ROC-AUC	PR-AUC
Logistic Regression	0.553	0.711	0.505
Decision Tree	13.40	0.556	0.484
Random Forest	0.815	0.659	0.450
Support Vector Machine	0.605	0.680	0.453
Hidden Markov Model	9.797	0.586	0.482
Recurrent Neural Network	0.544	0.726	0.522
Graph Convolutional Network	0.560	0.722	0.518
P-O Fit Neural Network	**0.533**	**0.745**	**0.543**

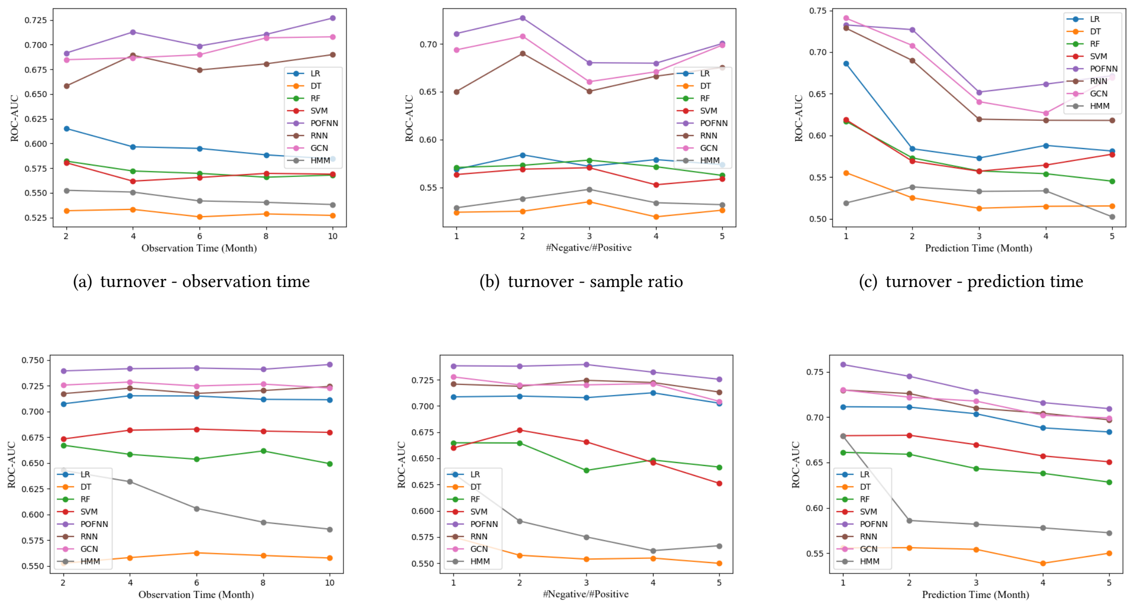
Table 6: Performance on performance prediction (poor).

Model	CE	ROC-AUC	PR-AUC
Logistic Regression	0.297	0.808	0.630
Decision Tree	7.606	0.685	0.489
Random Forest	0.758	0.767	0.609
Support Vector Machine	0.318	0.768	0.582
Hidden Markov Model	3.957	0.672	0.574
Recurrent Neural Network	0.286	0.827	0.658
Graph Convolutional Network	0.287	0.823	0.647
P-O Fit Neural Network	**0.280**	**0.834**	**0.669**

5.3 Robust Analysis

Here we introduce the evaluation of the robustness of our model. Without loss of generality, here we only evaluate the models on turnover prediction and excellent performance prediction in terms of ROC-AUC. Specifically, we generate the dataset with three kinds of different settings: 1) We adjusted the length of the sequential data, i.e., we used data of the x months before making the prediction ($x = 2, 4, 6, 8, 10$). 2) We predicted the outcome in the next y months with our model ($y = 1, 2, 3, 4, 5$). 3) We adjusted the number of samples in the dataset, so that the ratio of negative and positive samples is p ($p = 1, 2, 3, 4, 5$).

5.3.1 Evaluation on different observation period. Figure 7(a) and Figure 7(d) show the ROC-AUC performance of the models with different length of observation period. The x-axis represents the number of the observed months before making the prediction, i.e., the length of the sequential data N_t. It can be observed that the



(a) turnover - observation time (b) turnover - sample ratio (c) turnover - prediction time

(d) performance - observation time (e) performance - sample ratio (f) performance - prediction time

Figure 7: Robust analysis of our model under different experimental settings.

NN based models outperform the classic models, and POFNN performs the best. Meanwhile, in turnover prediction, NN based models perform better as the observation time getting longer, while the classic models perform worse. It indicates that NN based models are better at modeling the sequential working data. Also the results show that a short observation period is enough for predicting the performance, so the basic models also perform well in this case.

5.3.2 Evaluation on different ratios of positive and negative samples. Figure 7(b) and Figure 7(e) show the ROC-AUC performance of the models with different ratio of negative and positive samples. The x-axis represents the ratio. Since we only focus on the robustness of our model here, we did not set the total number of samples under each ratio to be the same, so the curve shows some fluctuation. It can be observed that POFNN outperforms the baselines under all the sample ratios.

5.3.3 Evaluation on different prediction period. Figure 7(c) and Figure 7(f) show the ROC-AUC performance of the models on predicting the outcomes in 1 to 5 months after the last observed month. It can be observed that POFNN performs the best. Though the models have a decrease on AUC performance as the prediction period getting longer, POFNN shows its efficiency on predicting the long-term outcomes. Indeed, predicting long-term outcomes is much tougher than predicting the short term outcomes. For example, when an employee has decided to leave the company, her turnover intention may reflect obviously on the behavior (e.g. little communication with colleagues). In this case, the models only considering employees' features also make good predictions. However, it may be too late for the human resource department to take actions with short-term prediction. Luckily, POFNN shows its ability to discover bad P-O fit that may potentially cause turnover in a long period, which is of great value in talent management.

5.4 Attention Layer Visualization

As described in Section 4, the attention layer attaches weights to the time intervals and outputs the weighted average of the LSTM outputs. Larger weight means greater impact on the outcome. Take turnover prediction as an example, a larger weight indicates abnormal P-O fit in some month, this may be caused by the changes of the organization (e.g., turnover of the superior). To study this influence, we drew the weights of 3000 randomly selected negative samples in turnover prediction on a heatmap. The result is shown in Figure 8. The x-axis represents the samples. And the y-axis represents the length between the month and the last observed month. For instance, if a sample's last observed month is 2017.10, then 1 on the y-axis denotes the month of 2017.09. Since our attention layer takes the last observed month as query, only the other 9 months are attached with weights and can be drawn on the heatmap. On the whole, we can find the months between 7 and 9 to be brighter and months less than 7 are darker, indicating that turnover in the future 2 months may be caused by bad P-O fit in the last 7 months. The result indicates that P-O fit may cause a quite long-term impact on turnover intention.

5.5 P-O fit Representation Visualization

We randomly selected 1000 samples and drew their P-O fit representation of the last observed month extracted by OSCN on a two-dimensional plane. Specifically, we applied t-Distributed Stochastic Neighbor Embedding (t-SNE) [28] for dimension reduction. The result is shown in Figure 9, where the red points represent the samples who are turnover and the blue points represent the samples who are not. It can be observed that these employees can be clustered into several groups with their P-O fit representation. Meanwhile, negative and positive samples have different distribution among the clusters. These observations indicates that there are

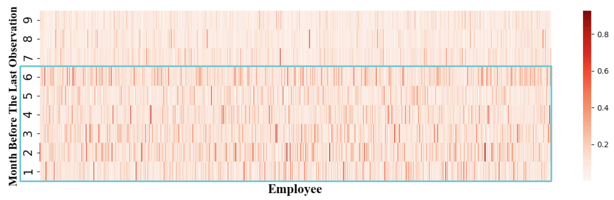


Figure 8: Weights in the attention layer (turnover).

several types of compatibility in the company and some of them can cause turnover behavior, showing the effectiveness of our extracted P-O fit representation.

5.6 Case study

The organizational environment of the employees may get impact from their colleagues in different ways. In each OSCN layer, environmental information is extracted from vectors of peers, superior and subordinates, we can measure the contribution of each relationship to the environment of the employees. Specifically, the contribution of relationship r to the i-th employees' current environmental vector can be calculated as $\|W_r^T u_r^i\|_1$, where $\|\cdot\|_1$ denotes $\mathcal{L}1$-norm, u_r^i denotes the vector of the relationshp and W_r denotes the weight. We randomly selected 100 negative samples in turnover prediction and calculated the contributions in the last observed month, then we drew them on a heatmap, which is shown in Figure 10, where the x-axis represents the employees, and y-axis represents working relationships. It can be observed that generally the employees' turnover intention is most affected by their superiors. Their subordinates have less influence and their peers have the least. Meanwhile, individual differences exist among these employees, for instance, sample 1 takes similar impact from all the three relationships, sample 2 gets dominated impact from the superior, and sample 3 is impacted most by the peers. It can also be observed that the impact from their subordinates are similar for most of the employees.

6 RELATED WORK

Generally, the related works of in this paper can be grouped into two categories, namely talent analytics and neural network based emerging applications.

6.1 Talent Analytics

Talent analytics [23, 36] focuses on applying data-driven technologies to understand large sets of people in order to make better

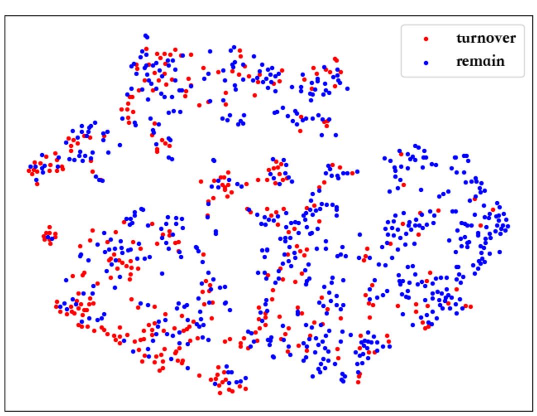


Figure 9: t-SNE for P-O fit representation.

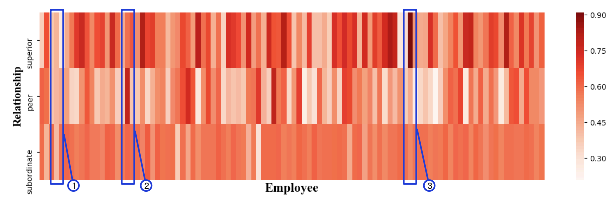


Figure 10: Weight of the colleagues (turnover).

organizational and operational decisions. Traditionally, studies of talent analytics are mainly based on surveys and straightforward statistical methods, e.g., statical hypothesis test and linear regression models. With these methods, researchers have designed a wide range of metrics to measure factors like job satisfaction and analyze their relationships with outcomes like turnover or revenue of operating unit [9, 30].

In recent years, the newly available big talent data provide unparalleled opportunity for researchers to understand talent behaviors and exploit advanced data mining approaches for talent analytics. For example, in the field of talent recruitment, many researchers studied the problem of person-job fit, job recommendation and market analysis [3, 16, 25, 27, 29]. Specifically, Qin *et al.* [32] proposed a novel ability-aware neural networks for enhancing the performance of person-job fit, which is based on the joint representation learning model PJFNN [40]. Zhu *et al.* [39] designed a sequential latent variable model for tracking the recruitment market trend. In the field of career development analysis, Li *et al.* [24] proposed a survival analysis approach to model the talent career paths with a focus on turnover and career progression and. Teng *et al.* [35] proposed a contagious neural network for turnover prediction. In addtion, there are also various emerging applications proposed in relevant domains, such as company profiling from a employee's perspective [26], and talent flow analysis [6, 37, 38].

Different from the above studies, in this paper, we focus on a new research topic, i.e., P-O fit in talent management, by proposing a novel data-driven neural network approach.

6.2 Neural Network Based Applications

Recent years have witnessed the rapid popularity of neural network technologies and applications [34]. Indeed, among various network structures, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are most widely used. Specifically, CNN is widely used for processing grid-like structured data [12], and made great success in processing image data with applications like image classification [20] and face recognition [22], as well as textual data [11]. Instead, RNN is a kind of neural networks widely used for modeling sequential data with applications like speech recognition [13], machine translation [2], and time series processing [8]. There are also works on combining CNN and RNN to process sequential grid-like structured data, such as applications of music tagging [7], visual recognition [10] and sentence modeling [15]. Inspired by previous studies, in this paper, our approach combines both RNN and CNN structures for hierarchically and dynamically extracting organization-aware compatibility features in P-O fit.

7 CONCLUSION

In this paper, we proposed a data-driven neural network approach for dynamically modeling the compatibility in Person-Organization fit (P-O fit) and its meaningful relationship with two critical issues in talent management, namely talent turnover and job performance. Specifically, we first designed an Organizational Structure-aware Convolutional Neural Network (OSCN) for hierarchically extracting organization-aware compatibility features in P-O fit. With the help of OSCN, the influence of organizational environment can be holistically modeled. Then, to capture the dynamic nature of P-O fit and its consequent impact, we further exploited an adapted Recurrent Neural Network with attention mechanism to model the temporal information of both talents and organizations. Finally, we conducted extensive experiments for evaluating our approach on real-world talent data, which clearly demonstrated the effectiveness in terms of turnover prediction and job performance prediction, as well as some interesting indicators of talent management through the visualization of network layers.

ACKNOWLEDGMENTS

The research work is supported by the National Key Research and Development Program of China under Grant No. 2018YFB1004300, the National Natural Science Foundation of China under Grant No. U1836206, U1811461, 61773361, 91746301, and the Project of Youth Innovation Promotion Association CAS under Grant No. 2017146.

REFERENCES

[1] Martha C Andrews, Thomas Baker, and Tammy G Hunt. 2011. Values and person-organization fit: Does moral intensity strengthen outcomes? *Leadership & Organization Development Journal* 32, 1 (2011), 5–19.

[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] Fedor Borisyuk, Liang Zhang, and Krishnaram Kenthapadi. 2017. LiJAR: A system for job application redistribution towards efficient career marketplace. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1397–1406.

[4] Daniel M Cable and D Scott DeRue. 2002. The convergent and discriminant validity of subjective fit perceptions. *Journal of applied psychology* 87, 5 (2002), 875.

[5] Jennifer A Chatman. 1989. Improving interactional organizational research: A model of person-organization fit. *Academy of management Review* 14, 3 (1989), 333–349.

[6] Yu Cheng, Yusheng Xie, Zhengzhang Chen, Ankit Agrawal, Alok Choudhary, and Songtao Guo. 2013. Jobminer: A real-time system for mining job-related patterns from social media. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1450–1453.

[7] Keunwoo Choi, György Fazekas, Mark Sandler, and Kyunghyun Cho. 2017. Convolutional recurrent neural networks for music classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2392–2396.

[8] Jerome T Connor, R Douglas Martin, and Les E Atlas. 1994. Recurrent neural networks and robust time series prediction. *IEEE transactions on neural networks* 5, 2 (1994), 240–254.

[9] Thomas H Davenport, Jeanne Harris, and Jeremy Shapiro. 2010. Competing on talent analytics. *Harvard business review* 88, 10 (2010), 52–58.

[10] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. 2015. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2625–2634.

[11] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N Dauphin. 2017. Convolutional sequence to sequence learning. *arXiv preprint arXiv:1705.03122* (2017).

[12] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*. Vol. 1. MIT press Cambridge.

[13] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. 2013. Speech recognition with deep recurrent neural networks. In *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE, 6645–6649.

[14] Timothy A Judge and Daniel M Cable. 1997. Applicant personality, organizational culture, and organization attraction. *Personnel psychology* 50, 2 (1997), 359–394.

[15] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188* (2014).

[16] Navneet Kapur, Nikita Lytkin, Bee-Chung Chen, Deepak Agarwal, and Igor Perisic. 2016. Ranking universities based on career outcomes of graduates. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 137–144.

[17] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[18] Thomas N Kipf and Max Welling. 2016. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907* (2016).

[19] Amy L Kristof. 1996. Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel psychology* 49, 1 (1996), 1–49.

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.

[21] Kristy J Lauver and Amy Kristof-Brown. 2001. Distinguishing between employees' perceptions of person–job and person–organization fit. *Journal of vocational behavior* 59, 3 (2001), 454–470.

[22] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. 1997. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks* 8, 1 (1997), 98–113.

[23] Alec Levenson. 2011. Using targeted analytics to improve talent decisions. *People and Strategy* 34, 2 (2011), 34.

[24] 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.

[25] Jia Li, Dhruv Arya, Viet Ha-Thuc, and Shakti Sinha. 2016. 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*. ACM, 501–510.

[26] Hao Lin, Hengshu Zhu, Yuan Zuo, Chen Zhu, Junjie Wu, and Hui Xiong. 2017. Collaborative Company Profiling: Insights from an Employee's Perspective.. In *AAAI*. 1417–1423.

[27] Qiaoling Liu, Faizan Javed, and Matt Mcnair. 2016. Companydepot: Employer name normalization in the online recruitment industry. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 521–530.

[28] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[29] 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.

[30] William H Mobley, Stanley O Horner, and At T Hollingsworth. 1978. An evaluation of precursors of hospital employee turnover. *Journal of Applied psychology* 63, 4 (1978), 408.

[31] Lyman W Porter, Richard M Steers, Richard T Mowday, and Paul V Boulian. 1974. Organizational commitment, job satisfaction, and turnover among psychiatric technicians. *Journal of applied psychology* 59, 5 (1974), 603.

[32] 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.

[33] Stephen P Robbins and Timothy A Judge. [n.d.]. Organizational behavior. 2001. *Google Scholar* ([n. d.]).

[34] Jinwen Sun, Keli Xiao, Chuanren Liu, Wenjun Zhou, and Hui Xiong. 2019. Exploiting Intra-day Patterns for Market Shock Prediction: A Machine Learning Approach. *Expert Systems with Applications* (2019).

[35] Mingfei Teng, Hengshu Zhu, Chuanren Liu, Chen Zhu, and Hui Xiong. 2019. Exploiting the Contagious Effect for Employee Turnover Prediction.. In *AAAI*.

[36] Vlad Vaiman, Hugh Scullion, and David Collings. 2012. Talent management decision making. *Management Decision* 50, 5 (2012), 925–941.

[37] 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.

[38] 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).

[39] 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.

[40] Chen Zhu, Hengshu Zhu, Hui Xiong, Chao Ma, Fang Xie, Pengliang Ding, and Pan Li. 2018. Person-Job Fit: Adapting the Right Talent for the Right Job with Joint Representation Learning. *ACM Transactions on Management Information Systems (TMIS)* 9, 3 (2018), 12.