Exploiting the Contagious Effect for Employee Turnover Prediction

Mingfei Teng,[1,2] Hengshu Zhu,[2,*] Chuanren Liu,[4] Chen Zhu,[2] Hui Xiong[1,2,3,*]
[1]Rutgers University, [2]Baidu Talent Intelligence Center, [3]Baidu Business Intelligence Lab, [4]Drexel University
{mingfei.teng, hxiong}@rutgers.edu, {zhuhengshu, zhuchen02}@baidu.com, chuanren.liu@drexel.edu

Abstract

Talent turnover often costs a large amount of business time, money and performance. Therefore, employee turnover prediction is critical for proactive talent management. Existing approaches on turnover prediction are mainly based on profiling of employees and their working environments, while the important contagious effect of employee turnovers has been largely ignored. To this end, in this paper, we propose a contagious effect heterogeneous neural network (CEHNN) for turnover prediction by integrating the employee profiles, the environmental factors, and more importantly, the influence of turnover behaviors of co-workers. Moreover, a global attention mechanism is designed to evaluate the heterogeneous impact on potential turnover behaviors. This attention mechanism can improve the interpretability of turnover prediction and provide actionable insights for talent retention. Finally, we conduct extensive experiments and case studies on a real-world dataset from a large company to validate the effectiveness of the contagious effect for turnover prediction.

Introduction

Talent turnover will affect business performance of companies. When an unexpected turnover request is raised in a company, significant effort is required to search for a replacement, and there is a risk that the operation of the company will be disrupted if a suitable replacement is not found. The situation becomes even worse due to the contagious effect of external talent turnover when a group of employees influence each other and quit their jobs collectively (Felps et al. 2009). Indeed, talent plays an important role in daily business operation, not only because of its abilities and expertise, but also related to its collective influence on each other and the whole company. The quit of a talent often costs a large amount of business time, money, and performance, and the contagious effect of talent turnover will deconstruct the organizational structure and cause a dysfunction in the company. To alleviate the negative impact of talent turnover, it is critical for an employer to proactively predict potential turnovers, which in turn allows effective talent retention or successful talent replacement.

Previous approaches on turnover analysis focused on the ease and the desirability of job movement (March and Simon 1958). Specifically, the ease of movement is related to factors such as job market availability, unemployment rate, and personal skill level. Factors related to desirability of movement include job satisfaction, salary growth, promotions, organization's commitment, etc. A series of work (Mobley, Horner, and Hollingsworth 1978; Jackofsky and Peters 1983; Mcevoy and Cascio 1987; Trevor, Gerhart, and Boudreau 1997) analyzed the relationships of talent turnover and theses various factors. Later, different psychological paths that employees would follow when they quit their jobs were analyzed for modeling turnover behaviors (Lee and Mitchell 1994; Lee et al. 1996; Lee et al. 1999). These analyses were mainly based on linear statistical models for testing theoretical hypothesis. Recently, survival analysis has been extended to predict the timing of turnovers with multiple sources of information (Li et al. 2017).

All the above research efforts focus on the variables associated with the turnover employees and their working environments, while the social influence among them, specifically the contagious influence of prior turnovers on the following ones, as shown in Figure 1, is rarely explored and exploited. The contagious influence effect, or simply contagious effect, has a complex nature in several ways:

- The contagious effect varies from people to people and is shaped by factors such as the positions in the organization and the connection strength between employees;
- The contagious effect is time variant due to its broadcasting and decaying mechanisms on the social network of employees in a company;
- A series of prior turnovers may produce a cumulative effect in the time window for prediction. This effect is even stronger for larger teams.

Given these challenges, it is a non-trivial endeavor to incorporate dynamic contagious effect with comprehensive factors for effective talent turnover prediction. Indeed, the profiling of employees and their working environments is both heterogeneous and dynamic in nature. Therefore, the modeling framework should be able to process multiple sources of sequential information with different length, granularity, and format to effectively predict turnover behaviors and support proactive talent management.

*Corresponding Author.

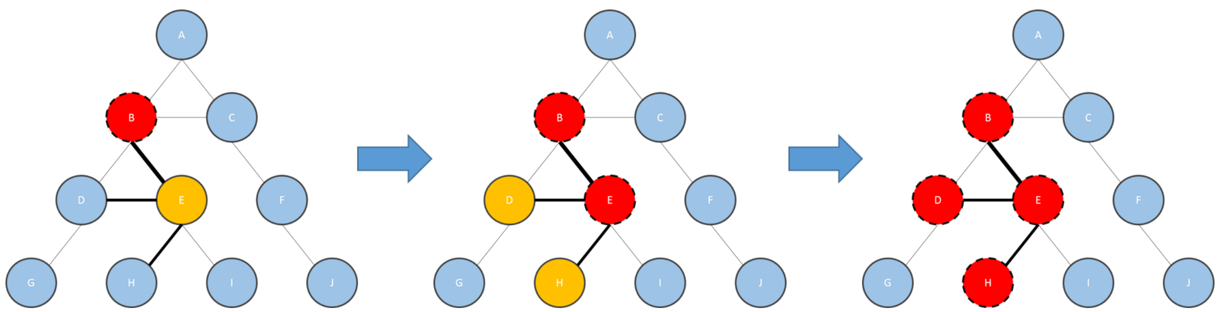


Figure 1: An example of contagious effect for employee turnover, where each node is an employee (i.e., blue is normal, red is a turnover, and yellow is an influenced one), the link between nodes represents the connection (i.e., stronger connection has bolder line weight).

In addition to providing accurate turnover predictions, it is also valuable to investigate the importance of various factors in each turnover's decision-making process. Quantifying and identifying important decision factors will reveal actionable insights such as motivations or reasons for the employees to quit their jobs. Such information is critical to facilitate employers in identifying unsteady points in their organizational structures and preventing talent turnovers for proactive talent retention and management.

To overcome the above difficulties, inspired by the application of RNN (recurrent neural network) on classification problem (Santos and Gattit 2014) and the attention mechanism (Bahdanau, Cho, and Bengio 2014; Yang et al. 2017), we build a contagious heterogeneous neural network (CEHNN) to integrate the peers' turnover sequence, the environmental change and static profile information, and add a global attention across the multiple chains. The attention mechanism helps to evaluate the contribution of each change (event) in all the sequences, so we can identify the most influential factor for the final turnover decision.

Our contribution can be summarized as below:

- We formulate the turnover prediction from a new angle, developing a new framework CEHNN to capture the contagious effect in the sequential employee turnovers.
- We use the framework CEHNN to process and integrate the sequential data from various sources with various formats, and design an attention mechanism across multiple sequences to evaluate the impact of different sequential factors on employee turnovers.
- We conduct extensive experiments on real-world data to validate the effectiveness of our framework and validate the contributing factors with case study.

Related Work

Recent years, there is a rising trend of applying advanced AI technologies to address talent related business problems (Zhu et al. 2018; Qin et al. 2018; Xu et al. 2018; Chen et al. 2017). Regarding the problem of turnover prediction, it is a hot topic that has been studied for years in human resource management. Previous research efforts generally fall into two categories. One focuses on analyzing the relationship of various factors and turnover, with tools of hypothesis test and linear models, while the other is trying to formulate the problem as survival analysis problems, and aims to predict the time to the occurrence of the turnover.

The early research of first category is mainly based on March and Simon's work (March and Simon 1958). Generally they proposed the turnover is determined by the desirability of movement and the ease of movement. The desirability of movement can be characterized by job satisfaction, salary growth, promotions, and organization's commitment, while the ease of movement can be characterized by the job market availability, unemployment rate and personal skill levels etc. The contributions of these factors have been extensively studied in the following work. Mobley (Mobley, Horner, and Hollingsworth 1978) quantitatively analyzed the correlations between job satisfaction, age-tenure, intention to quit and turnover, and made predictions based on regression analysis. Glenn (Mcevoy and Cascio 1987) found the turnover is lower among good performers, moderated by the turnover type, time span and level of unemployment, while Jackofsky (Jackofsky, Ferris, and Breckenridge 1986) and Trevor (Trevor, Gerhart, and Boudreau 1997) identified a curvilinear relationship between job performance and turnover, stating the turnover is higher for low and high performers. Trevor also identified the moderating influence of salary growth and promotions on the curvilinearity. Despite of the analysis of the interested factors and turnover, there also exist some research (Lee and Mitchell 1994; Lee et al. 1996; Lee et al. 1999) focusing on developing an unfolding model to describe and compare the psychological paths that employees take when they quit the jobs, the process of quitting from the initiation to the final decision is divided into different stages for discussion. One notable work is done by Felps, who studied the relation between job embeddedness and quitting, and mentioned there was a negative relationship between co-workers' job embeddedness and focal employee turnover (Felps et al. 2009). It indicates a contagious effect when talent turnover happens.

For the research on turnover survival analysis model, some existing approaches can be directly applied like the classic Cox proportional hazards model, which (Cox 1992) defined the hazard function of time and sample covariates. However it has a strong restriction that the time and covari-

Employee	Co-Worker	Role	Level	Intimacy	Start Date	End Date	Co-Worker Leave Date	Leave Date
Alice	Bob	Manager	1	0.9	2017-01-01	2018-02-01	2018-02-05	2018-04-15
Alice	Carl	Peer	0	0.6	2017-04-10	2018-03-01	2018-03-10	2018-04-15
Alice	David	Manager	2	0.2	2017-02-03	2018-04-9	N/A	2018-04-15
Ellen	Frank	Peer	0	0.4	2017-07-05	2018-01-07	2018-04-01	N/A

Table 1: A toy example of our processed pairwise turnover dataset.

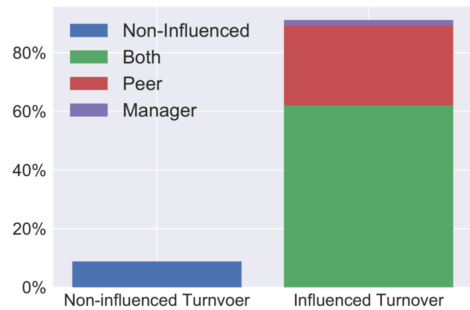


Figure 2: The comparison of the sizes of turnover grouped by whether they are under prior turnover influence.

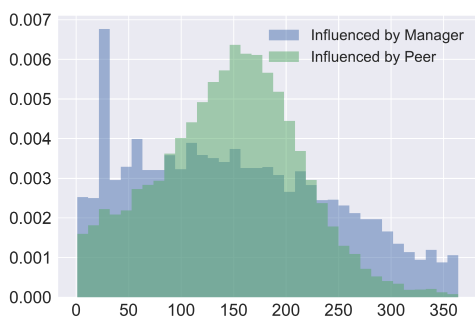


Figure 3: The density histogram of the mean of days between current turnover and prior turnover.

ates are independent and the weights of covariates are shared by all samples. In Li's work (Li et al. 2017), he adopted and extended the multi-task framework of Yan Li (Li et al. 2016), which treats the prediction of presence of employee at each time interval as a task, so the weights of the sample covariates varied from time to time.

Data Description

In this section, we explore the data to give a brief introduction, and more importantly to observe the contagious effect which inspires us of the way to formulate the research problem. The original dataset is provided by a high-tech company in China. It contains all of turnover records from 2016 to 2018, along with the profiles of employee. Both have been anonymized for privacy protection.

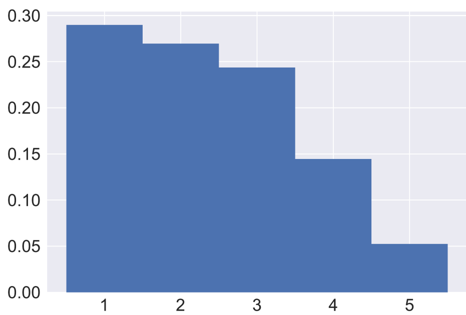
- *Profile Data*: The profile dataset includes the information such as anonymized employee ID, entry date, department, organization level and a year-based metric which characterizes the intimacy between each pair of employee in the company, generated according to the daily interactions.
- *Turnover Data*: This dataset includes the anonymized employee ID and leave date.

For our convenience, we integrated the profile data and turnover records in a pairwise way. Each integrated record would be a pair of co-workers (leader/subordinate or peers), for a period of time. Indirect managers and subordinates are also counted. For each record, the features include the anonymized employee's ID and co-worker's ID, the role of co-worker, their relative level, their relation start date and end date, and leave dates. Table 1 is an example dataset of the organized turnover records.

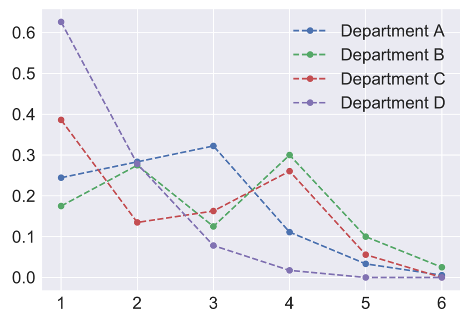
Based on the intermediate pairwise turnover dataset, we conduct further exploration to see how many turnovers happens under the influence of peers' or managers' turnover. Figure 2 shows the comparison of the size of turnover grouped by whether they are under a manager's or peer's turnovers influence within one year. We can see about 91% turnovers are under the influence of prior turnover (about 2% from managers only, 27% from peers only and 62% from both). Further we analyze the distribution of the mean of days between the pairwise turnovers. Figure 3 shows the distribution of the mean of days between employees' turnover and their closest prior manager/peer's turnover. We can see there is a decay effect, which indicates the influence would decrease with time as a trend. The interesting thing is in the density histogram related to peer, there is an apparent peak, indicating the influence of prior turnover may need to take certain amount of time to completely spread out and take effect, while in the density histogram related to manager, the distribution is almost decaying from the very beginning. Besides, in the group of turnover influenced by latest managers' turnover, the distribution of relative level of managers are plotted in Figure 4 (a). It can be found that more employees' turnovers are under the influence of their direct supervisors' turnovers. The number decays as the relative levels arise. However at a finer granularity, this is not the case, in Figure 4 (b), we can see the shape of distributions varies by department, in some of them, the prior managers' turnovers are followed by more subsequent turnovers in indirect subordinates than in direct ones.

Technical Details of CEHNN

In this section, we formulate a heterogeneous sequence classification problem, and propose the CEHNN (contagious effect heterogeneous neural network) framework as a solution,



(a) The density histogram of the relative level of turnover pair for the whole company



(b) The density histogram of the relative level of turnover pair for selected department

Figure 4: The density histogram of the relative levels of influential managers in our data.

followed by discussion of the global attention mechanism in the heterogeneous neural network.

Problem Formulation

Our primary goal is to predict talent turnovers based on the sequential or time-variant information from different sources. Specifically, for the i-th sample, we want to make binary classifications $\hat{y}_i$ at a decision time t_i based on static information A_i and a collection of M information series $\{\mathbf{s}_i^1, \mathbf{s}_i^2, \cdots, \mathbf{s}_i^M\}$. All the information $\mathbf{s}_i^m$ are collected during an observation period $[t_i - \Delta t_{\text{observe}}, t_i]$, and each information series $\mathbf{s}_i^m$ is a sequence $\{x_{i,1}^m, x_{i,2}^m, x_{i,3}^m, \cdots\}$. Since theses sequences are from different sources, they may vary in terms of lengths and dimensions.

In the context of turnover prediction, $\hat{y}_i = 1$ if the i-th employee will quit his/her job in the future prediction period $[t_i, t_i + \Delta t_{\text{predict}}]$. Otherwise $\hat{y}_i = 0$. Here we focus on $M = 2$ sequential information sources together with static employee profiles. One sequential information source is the co-workers' turnover events ordered by timestamp and denoted by $\mathbf{s}_i^{\text{turnover}}$. The co-workers includes prior managers and peers, etc. The other sequential information source is the dynamic environmental statistics of the employee and denoted by $\mathbf{s}_i^{\text{env}}$. The profile of employees is treated as the static information in A_i. Our task is to estimate the probability $P(y_i|\mathbf{s}_i^{\text{turnover}}, \mathbf{s}_i^{\text{env}}, A_i)$, so we can make a prediction based on:

Category	Name
Profile Input A	Department
	Organization level
	Key Staff
	Job Category
Sequence S^{turnover}	Relative level of employe i-j pair
	Days between the i-j relation end and observation end
	Prior turnover j's profile
	Relation type of employe i-j pair (peer/manager)
	Common work days of employe i-j pair
	Communication statistics of employe i-j pair
Sequence S^{env}	Monthly total employee in the department
	Monthly total turnover in the department
	Monthly total employee at the level
	Monthly total turnover at the level
	Monthly turnover rate at the level

Table 2: The list of input features.

$$\hat{y}_i = \begin{cases} 1, & \Pr(y_i = 1|\mathbf{s}_i^{\text{turnover}}, \mathbf{s}_i^{\text{env}}, A_i) > \phi, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

where ϕ is a discrimination threshold of employee turnover, and is set to 0.5 in our experiments.

Heterogeneous Neural Network

As discussed above, the three input sources are varied in terms of their length and format, so each needs to be taken care of before integration for the prediction. As a whole, a heterogeneous neural network is designed to address this.

For the prior two which are sequential data, we choose the long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997). As a variant of recurrent neural network, LSTM is powerful to process sequential data, with the ability to capture the long and short term dependency and overcome the exploding and vanishing gradients problems. A single LSTM cell is composed of an input gate i, a forget gate f, and an output gate o, which can be formulated as:

$$\begin{aligned} f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f), \\ i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i), \\ o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o), \\ c_t &= f_t \circ c_{t-1} + i_t \circ \tanh(W_c x_t + U_c h_{t-1} + b_c), \\ h_t &= o_t \circ \tanh(c_t), \end{aligned}$$

where $\{x_t\}_t$ is a series of input for the LSTM cell, $W_f, W_i, W_o, W_c, U_f, U_i, U_o, U_c, b_f, b_i, b_o, b_c$ are parameters to be trained, c_t is the cell state, h_t is the output of the cell, σ is the sigmoid function and $\circ$ is element-wise product.

We use two LSTM cells to process $\mathbf{s}_i^{\text{turnover}}$ and $\mathbf{s}_i^{\text{env}}$ respectively. Each element in $\mathbf{s}_i^{\text{turnover}}$ is a prior co-worker j's turnover, including features of employee j's profile and the intimacy of the pair i-j's relation; each element in $\mathbf{s}^{\text{env}}$ is

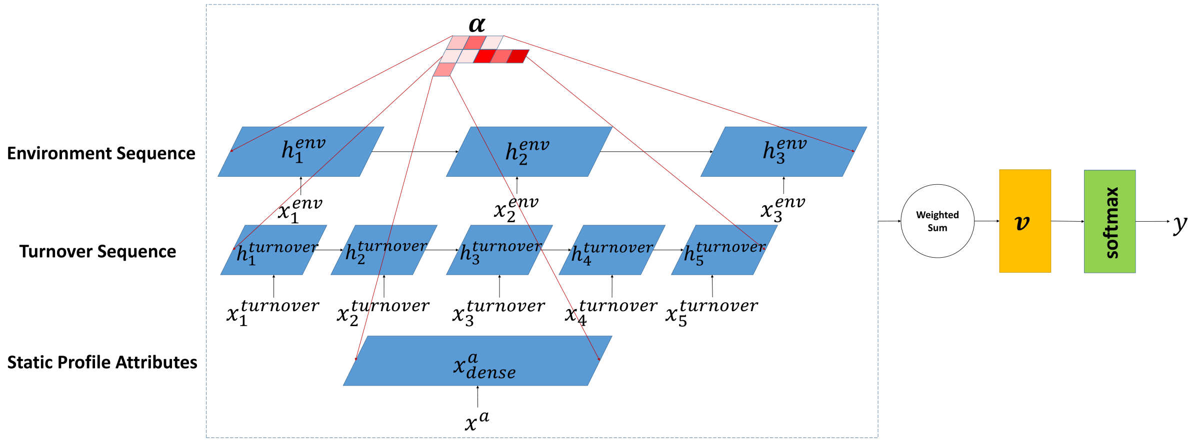


Figure 5: The overview of framework CEHNN for turnover prediction.

a monthly turnover statistics of specific department and organization level. Table 2 lists the detailed features for each part. Dropout layers are appended to both LSTM cells to avoid the overfitting.

For the static profile information A_i, we use a fully connected layer to process it.

Global Attention Mechanism

As mentioned before, we aim to evaluate the contributions of different factors into the final decision. At this point, we decide to introduce the attention mechanism. Attention mechanism for RNN (Bahdanau, Cho, and Bengio 2014) is proposed to be added into the sequence-to-sequence model, to achieve a word alignment effect, which reflects the importance of each word in source sentence for each word in the output sentence. By default, it is feasible to add one step attention in our framework, since we only have one step output, to each LSTM separately, however such implementation only helps to evaluate the importance of each event within a sequence. Alternatively, we decide to extend it to a global one across multiple sequences. Suppose for each sample we have M series of output from M LSTM cells, then the global attention can be formulated as:

$$u_t^m = \tanh(W^m h_t^m + b^m),$$

$$\alpha_t^m = \frac{(\exp(u_t^m)^\top u_c)}{\sum_{t\in\bigcup_m T_m} \exp(u_t^m)^\top u_c},$$

$$v = \sum_{t\in\bigcup_m T_m} \alpha_t^m (W^m h_t^m + b^m),$$

where h_t^m is the output of LSTM cell for m-th sequence at step t, u_t^m is the hidden representation of h_t^m, α_t^m is the normalized importance for t-th event in m-th sequence, v is the representation of the sequence as the aggregated weighted sum of hidden representation, T_m is the set of steps for m-th sequence, $\{W^m\}_m, \{b^m\}_m, u_c$ are parameters to be estimated through training. Based on these, the conditional probability to be estimated become:

$$\Pr(y_i = 1|\mathbf{s}_i^{\text{turnover}}, \mathbf{s}_i^{\text{env}}, c_i) = \text{softmax}(Wv + b). \quad (2)$$

Figure 5 shows the overall structure of our framework.

Experiment

In this section, we will evaluate the effectiveness of our framework. Specifically we trained and tested the framework CEHNN on a real-world dataset, comparing its performance with several state-of-the-art baselines. Meanwhile we will discuss the experiment result, as well as to analyze the attention weights generated through our framework by case study.

Name	Value
Total population	2,935
Positive samples	1,304
Negative samples	1,631
Dimension of Profile	400
Dimension of Turnover Sequence	419
Dimension of Environmental Sequence	4

Table 3: The statistics of experimental data.

Name	Value/Setting
Dimension of $\mathbf{S}^{\text{turnover}}$ LSTM	90
Dimension of $\mathbf{S}^{\text{env}}$ LSTM	20
Dimension of Profile Dense Layer	20
Dimension of Dense Layer after Attention	35
Dropout Probability	0.5
Attention Size	10

Table 4: The network configuration.

Method	Accuracy	Precision	Recall	F-measure	AUC-ROC
Logistic Regression	0.814±0.027	0.808±0.036	0.764±0.055	0.785±0.035	0.869±0.029
SVM	0.709±0.038	0.816±0.064	0.448±0.091	0.577±0.079	0.818±0.029
Random Forest	0.77±0.026	0.812±0.05	0.629±0.05	0.708±0.036	0.86±0.028
Gradient Boosting	0.833±0.03	0.861±0.038	0.744±0.05	0.798±0.039	0.910±0.026
Turnover sequence HMM	0.752±0.101	0.862±0.229	0.568±0.19	0.671±0.046	0.55±0.088
Environmental change HMM	0.656±0.115	0.619±0.178	0.684±0.206	0.638±0.077	0.577±0.032
Turnover sequence RNN	0.853±0.022	0.856±0.025	0.806±0.044	0.83±0.028	0.905±0.025
Environmental change RNN	0.761±0.034	0.79±0.061	0.633±0.07	0.702±0.049	0.804±0.045
CEHNN	**0.864±0.018**	**0.871±0.029**	**0.816±0.049**	**0.842±0.024**	**0.914±0.02**

Table 5: The overall performance of different approaches for turnover prediction.

Experimental Setup

Data Pre-processing. We set the observation time span $\Delta_{observe}$ for the sequence to be one year and the turnover prediction period $\Delta_{predict}$ to be three months. An employee who quits the job in the prediction period would be counted as a positive sample otherwise counted as a negative one. To isolate the influence in the prediction period, the decision time t for each sample was chosen with constraint to make sure there was no prior manager/peer turnover in the prediction period. The statistics of the data is listed in Table 3. We used 64% for training, 16% for validation, 20% for test.

Model Configuration. The dimension of our input and network configuration can be found in Table 4. We used the Adam optimizer(Kingma and Ba 2014) for training. The objective function is the cross entropy loss function.

Baseline Methods and Evaluation Metrics

We compared our framework with baselines to demonstrate its effectiveness comprehensively. The baselines selected fell into three categories, a) popular classification algorithms without ability to process sequential data, such as Logistic Regression, SVM, Random Forest and Gradient Boosting; b) algorithms able to deal with sequential data natively such as HMM c) LSTM for turnover sequence only and environmental change only. For algorithms not designed for sequential data we concatenated and padded the sequential data in preprocessing. HMM is unable to process multiple sequences with varied length, so we trained two HMM models, one was fed with the turnover sequence data concatenated with profile data, the other was fed with the environmental sequential data concatenated with profile data. The hyper-parameters of all above models were found by grid search within a predefined range based on experience and suggested best practice.

Since the problem is a binary classification problem, we chose accuracy, precision, recall, F-measure and the area under the curve of receiver operating characteristic (AUC-ROC) to measure the performance of the framework.

Performance Comparison

The baselines and the framework CEHNN were evaluated on the turnover dataset, the experimental result is listed in Table 5. We conducted each algorithm 10 times, each time we generated a random training/validation/test dataset with the same proportion. The values listed in the table are the means with 95% confidence interval. We use bold font to emphasize the top 1 for each metric.

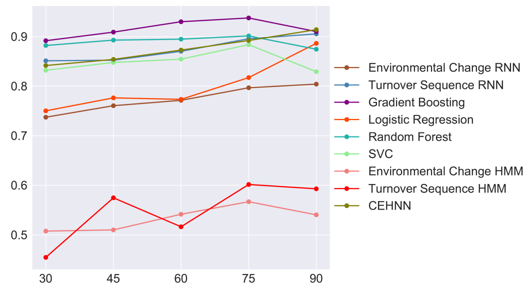
From the table, we have following observations: first, overall our framework performs better than others on all metrics; second, the turnover sequence RNN also achieves a competent result (the runner-up in accuracy, recall and F-measure), which demonstrates exploiting the prior turnover sequence will benefit the future turnover prediction, in contrast, the environmental change plays a less important role; third, the high dimension of concatenated data limits the performance of classical classifiers, but Gradient Boosting still achieves a competent performance; lastly, HMM performs poorly on the dataset, we think the reason may be there is a long dependency which cannot be captured.

Robustness Analysis

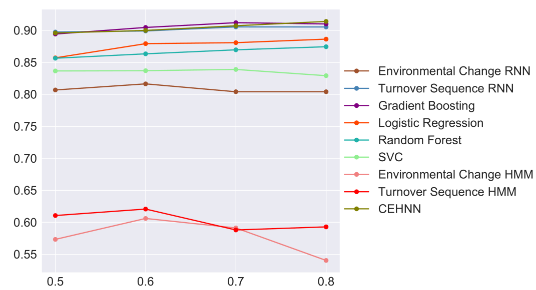
To validate the robustness of our framework, we conducted three experiments: a) we tested it under the values of $\Delta_{predict}$ set to $30, 45, 60, 75, 90$ days, as the prediction period is getting smaller, some positive (turnover) samples may turn to negative (non-turnover) ones; b) we tested it under the values of $\Delta_{observe}$ set to $3, 6, 9, 12$ months; c) we tested it under the values of ratio $= \frac{\text{\# of positive}}{\text{\# of negative}}$ set to $0.5, 0.6, 0.7, 0.8, 0.9$.

Figure 6 (a) compares the AUC-ROC values when we applied all the models to predict turnover for $30, 45, 60, 75, 90$ days. In general, the Gradient Boosting, Random forest, Turnover Sequence RNN and CEHNN outperforms the others. The Random Forest and Gradient Boosting performs better when the prediction period is short, while the Turnover Sequence RNN and CEHNN are good at predicting for a longer period, which demonstrates they can exploit the sequential data better and make a longer prediction than traditional classifiers.

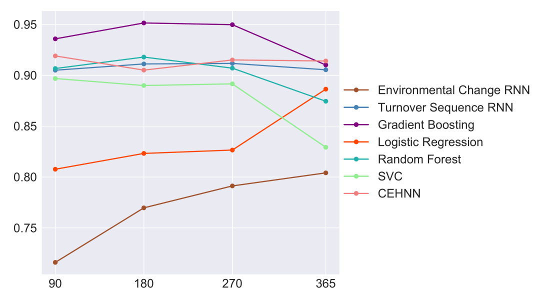
Figure 6 (b) compares the AUC-ROC values of all the models on the datasets with different ratios of positive and negative samples. Similar to Figure 6 (a), the Gradient Boosting, the Turnover Sequence RNN and CEHNN outperform the others. The Gradient Boosting and the Turnover Sequence RNN is slightly better than CEHNN. We consider this is due to the size of the dataset. In our original dataset, the ratio of positive and negative samples is close to 0.8, when we adjust the ratio to be $0.5, 0.6, 0.7$, we actually reduce the size of available data, which makes the framework



(a) AUC under different prediction period

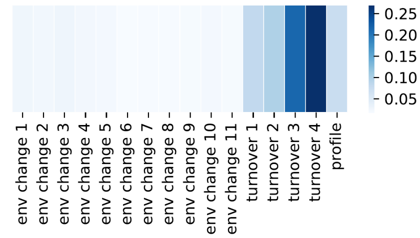


(b) AUC under different pos./neg. ratio

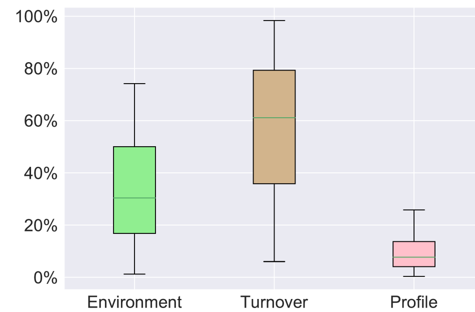


(c) AUC under different observation period

Figure 6: The robustness analysis of different methods.



(a) Weights for selected sample



(b) Boxplot of aggregated contribution

Figure 7: The visualization of weights from attention.

CEHNN lack of training.

Figure 6 (c) compares the AUC-ROC values of all the models on dataset with different lengths of observation period. In general, it can be observed that as the observation period increases, the framework CEHNN performs consistently well, which demonstrates its ability to process and exploit long sequential data. In contrast, the performances of traditional classifier Gradient Boosting and Random Forest drop explicitly as observation period increases, for which the reason might be the increasing high dimension of concatenated data make the prediction more challengeable.

Case Study

We conducted two case studies with our framework on the dataset. One is on the individual level, the other is on the organizational level. On the individual level, a sample (employee) was chosen in the dataset and the global weights for it was learned by the CEHNN and plotted in Figure 7 (a). The weights are composed of three parts, the leftmost is weights of the environmental change, the middle is the weights of turnover sequence, and the rightmost one is the weights of profile. It is found for this sample, the largest weight located in the turnover sequence, so the framework evaluate the 4th prior turnover as the largest contributing factor. On the organizational level, we evaluated the CEHNN on a test set of 587 samples, the weights are summed within environment change, turnover sequence and profile, as shown in Figure 7 (b). It is found that for the turnover prediction, the contributions from the prior turnover sequence, the environmental change and profile are approximately 61%, 30% and 9%.

Conclusion

We propose a contagious effect heterogeneous neural network (CEHNN) for turnover prediction within a period of future in this paper. The heterogeneous structure endows the framework the ability to process the sequential/non-sequential data from different sources together. Specifically it integrates the static profile information and the environmental change. Moreover, it exploits the contagious effect in the turnover sequences, formulating and solving the problem from a new angle. Further a global attention mechanism is implemented in the framework to detect a global importance of all involved factors, which gives more interpretability and actionable insight to our problem. Our experiment validate the effectiveness and robustness of the framework, and shows the value of global attention by case study.

Acknowledgements

This research was partially supported by grants from the National Natural Science Foundation of China (Grant No. 91746301 and 71531001).

References

Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; and Zheng, X. 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Bahdanau, D.; Cho, K.; and Bengio, Y. 2014. Neural machine translation by jointly learning to align and translate. *Computer Science*.

Chen, H.; Xiao, K.; Sun, J.; and Wu, S. 2017. A double-layer neural network framework for high-frequency forecasting. *ACM Trans. Manage. Inf. Syst.* 7(4):11:1–11:17.

Cox, D. R. 1992. *Regression Models and Life-Tables*. Springer New York.

Felps, W.; Mitchell, T. R.; Herman, D. R.; Lee, T. W.; Holtom, B. C.; and Harman, W. S. 2009. Turnover contagion: How coworkers' job embeddedness and job search behaviors influence quitting. *Academy of Management Journal* 52(3):545–561.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural Computation* 9(8):1735–1780.

Jackofsky, E. F., and Peters, L. H. 1983. The hypothesized effects of ability in the turnover process. *Academy of Management Review Academy of Management* 8(1):46.

Jackofsky, E. F.; Ferris, K. R.; and Breckenridge, B. G. 1986. Evidence for a curvilinear relationship between job performance and turnover. *Journal of Management* 12(1):105–111.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *CoRR* abs/1412.6980.

Kuang, K.; Cui, P.; Athey, S.; Xiong, R.; and Li, B. 2018. Stable prediction across unknown environments. In Guo, Y., and Farooq, F., eds., *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018*, 1617–1626. ACM.

Lee, T. W., and Mitchell, T. R. 1994. An alternative approach: The unfolding model of voluntary employee turnover. *Academy of Management Review* 19(1):51–89.

Lee, T. W.; Mitchell, T. R.; Wise, L.; and Fireman, S. 1996. An unfolding model of voluntary employee turnover. *Academy of Management Journal* 39(1):5–36.

Lee, T. W.; Mitchell, T. R.; Holtom, B. C.; Mcdaniel, L. S.; and Hill, J. W. 1999. The unfolding model of voluntary turnover: A replication and extension. *Academy of Management Journal* 42(4):450–462.

Li, Y.; Wang, J.; Ye, J.; and Reddy, C. K. 2016. A multi-task learning formulation for survival analysis. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1715–1724.

Li, H.; Ge, Y.; Zhu, H.; Xiong, H.; and Zhao, H. 2017. Prospecting the career development of talents: A survival analysis perspective. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 917–925.

March, J. G., and Simon, H. A. 1958. *Organizations*. Wiley-Blackwell.

Mcevoy, G. M., and Cascio, W. F. 1987. Do good or poor performers leave? a meta-analysis of the relationship between performance and turnover. *Academy of Management Journal* 30(4):744–762.

Mobley, W. H.; Horner, S. O.; and Hollingsworth, A. T. 1978. An evaluation of precursors of hospital employee turnover. *J Appl Psychol* 63(4):408–414.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; and Duchesnay, E. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830.

Qin, C.; Zhu, H.; Xu, T.; Zhu, C.; Jiang, L.; Chen, E.; and Xiong, H. 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*, SIGIR '18, 25–34. New York, NY, USA: ACM.

Rabiner, L. R. 1989. A tutorial on hidden markov models and selected applications in speech recognition. In *Proc. of the IEEE*, 257–286.

Santos, C. N. D., and Gattit, M. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *International Conference on Computational Linguistics*.

Trevor, C. O.; Gerhart, B.; and Boudreau, J. W. 1997. Voluntary turnover and job performance: Curvilinearity and the moderating influences of salary growth and promotions. *Journal of Applied Psychology* 82(1):44–61.

Xu, H.; Yu, Z.; Yang, J.; Xiong, H.; and Zhu, H. 2018. Dynamic talent flow analysis with deep sequence prediction modeling. *IEEE Transactions on Knowledge and Data Engineering* 1–1.

Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; and Hovy, E. 2017. Hierarchical attention networks for document classification. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1480–1489.

Zhu, C.; Zhu, H.; Xiong, H.; Ma, C.; Xie, F.; Ding, P.; and Li, P. 2018. Person-job fit: Adapting the right talent for the right job with joint representation learning. *ACM Trans. Manage. Inf. Syst.* 9(3):12:1–12:17.