

# Analysis and Forecast of Employee Turnover Based on Decision Tree Method



Haining Yang and Xuedong Gao

**Abstract** The decision tree is established by the employee turnover as a target variable, and gender, marital status, household registration, registered permanent residence, age, driving duration, education background, post, and category of employees as input variables using C5.0 decision tree algorithm, and recommendations about human resources management are proposed based on the mined rules. The research results show that this approach has more accurate forecast of employee turnover, and the knowledge is found easily to be understood. It can help companies to master the law of employee turnover and reduce the brain drain level.

**Keywords** Employee turnover · Decision tree · Data mining

## 1 Presentation of the Problem and Review of Literature

In recent years, the economic development of our nation has gradually move from the industrial age to the information age, and human resources, as the medium of knowledge and creators, have become the core driving force behind the development of enterprises. At the same time, the increasingly open nature of employment information and increasing specialization of recruitment services have led to more frequent personnel turnover, with the turnover rate of mid-level to senior workers becoming as high as 50–60% in some domestic enterprises, with average employment periods of less than 3 years [1]. Mass departures not only brings about loss in enterprise recruitment and training costs, but also has negative effects on the operation of the enterprise such as loss of important client, leakage of key technology, decrease the enterprise competitive power, even cause the big business

---

H. Yang (✉) · X. Gao  
Donlinks School of Economics and Management, University of Science and Technology  
Beijing, Beijing, China  
e-mail: [yanghaining@apiins.com](mailto:yanghaining@apiins.com)

X. Gao  
e-mail: [gaoxuedong@manage.ustb.edu.cn](mailto:gaoxuedong@manage.ustb.edu.cn)

failure. Therefore, effectively predicting employee turnover patterns and reducing losses incurred in these turnovers has become an important question in the practice and theoretical research of human resource management.

So far, research on employee turnover mainly focuses in the fields of economics and organizational behavior, with research in economics utilizing methods of econometric analysis and game theory analysis to analyze supply and demand patterns in the labour market behind employee turnover. Xu and Zhao used the Probit model to study employee turnover, demonstrating that regional economical development level will affect employee turnover [2]. Based on data from 676 listed companies from 1997 to 2007, Yang's research on CEO turnover demonstrated that CEO turnover in listed companies of our nation is not only an financial process, but also a social-political process [3]. Zou and Dong constructed game models from the perspective of a library and its employees, and used evolutionary game theory to analyze stability strategies from both sides [4].

Research in organizational behavior mainly employs questionnaires and statistical analysis methods to analyze influencing factors and its mechanisms behind employee turnover. Wang and others conducted a statistical analysis on 932 questionnaires on the basis of stress interaction theory and job requirement resource models, and demonstrated that employees' job requirement had a positive impact on intention of leaving employment, impulsive personality has a mediative effect on the impact of job requirement on intention of leaving employment, and social support has an intermediary effect for the relationship between impulsive personality, and job requirements and intention of leaving employment [5]. Through statistical analysis of 433 questionnaires, Dan and others found that a negative correlation exists between employees' perception of status and employees' perception of union status, with intention of leaving employment [6]. After analyzing 267 employees and 92 supervisors' paired samples, Ye and others demonstrated that employees' sense of workplace exclusion and marginalization have a positive impact on turnover intention [7].

Research results in economics demonstrated market trends of employee turnover from a macro perspective, but these research abstracts their research subject, ignoring the influence of gender, age, needs, perception, emotion, and other individual factors in a worker's decision to leave a job. Research in organizational behavior reveals the intrinsic motivation of employee turnover from the aspects of employees' psychological factors, organizational factors and incentive factors, but the research data used in these studies are mostly based on the "intention of leaving employment" questionnaire, rather than the actual turnover behavior of employees, and are therefore unable to make direct predictions on turnover behavior. And so, this essay uses actual turnover data from enterprises along with decision tree algorithms to analyze and predict staff turnover, and to provide relevant strategical suggestions.

Decision tree is a data mining method that uses inductive learning. It is a tree-like decision graph with additional probability results and an intuitive graphical method using statistical probability analysis. In machine learning, decision tree is a prediction model. It represents a mapping between object attributes and object values. Each node in the tree represents the judgment conditions of object attributes,

and its branches represent the objects that meet the node conditions. The leaf nodes of the tree represent the predicted results of the object's ownership. The reasons why decision tree classification model is widely used are as follows:

- (1) *Compared with neural network or Bayesian classification, the classification principle of decision tree is simple and easy to understand and accept by users.*
- (2) *In the process of decision tree classification, there is no need for human activities. Setting any parameters is more suitable for the requirement of knowledge discovery.*
- (3) *Decision Tree Classification does not require any data sets other than training data sets. And test additional information beyond the data set to ensure that the decision tree and others compared with other classification methods, this method has higher classification speed.*
- (4) *Compared with other classification models, the decision tree classification method has the following advantages.*

It has very good classification accuracy.

To sum up, this method summarizes corresponding classification rules from a set of unordered and irregular cases, and can also test the classification rules using actual classification results. Widely used in a variety of research fields [8]. Gu and Xu used decision tree classification method to model the user loyalty of professional virtual communities [9]. Peng first used a metrology method to analyze user characteristics, and then used a decision tree to classify the user characteristics [10]. Sun and other, using 16009 traffic accident data to analyze the factors affecting traffic accidents, using a method based on the C5.0 decision tree [11]. C5.0 is an algorithm in decision tree model. It was developed by J. R. Quinlan in 79 years and ID3 algorithm was proposed. It mainly aims at discrete attribute data, and then continuously improves to form C4.5. It adds the discretization of contiguous attributes on the basis of ID3. C5.0 is a classification algorithm of C4.5 applied to large data sets. It mainly improves the execution efficiency and memory usage. C4.5 algorithm is a revised version of ID3 algorithm. Gain Ratio is used to improve the method. The segmentation variable with the largest Gain Ratio is selected as the criterion to avoid the problem of over-matching of ID3 algorithm. C5.0 algorithm is a revised version of C4.5 algorithm, which is suitable for processing large data sets. Boosting method is used to improve the accuracy of the model, also known as Boosting Trees. It is faster in software calculation and occupies less memory resources. Application results in various fields show that the data mining method using decision tree method has better capacity for interference and interpretability, and that this method can process multiple types of data, and has low requirements for the data used.

## 2 Indicator Selection

In the research of factors influencing employee intention in leaving employment, some scholars use statistical analysis on questionnaires to discover the impact of factors such as employee gender, age, education, marital status, registered permanent residence, job position, continuous service age, number of resignations, probability of promotion, number of promotions, etc. on employee-organizational commitment and intention of leaving employment [12, 13]. Based on research results and data of existing literature, this paper chose 9 indicators: gender, age, time in current company, education, marriage, household registration, registered permanent residence, job category, and job level, to analyze the status of employee turnover for an electronic material manufacturing enterprise. Mobley believes that leaving a job means that the employee has worked in a position in the organization for a period of time [14]. After consideration, the individual is deliberately (Deliberate Willfulness) to leave the original job, thus losing his position and related interests, and the original company/Organization is no longer relevant. Based on Mobley's previous separation factors, Rodger, Peter & Stefan also performed post-analysis to update the effects of various factors on turnover behavior in a number of studies, in terms of demographic variables: on-the-job cognitive ability [15]. There is a low correlation between the test and the resignation. The high-education, family responsibilities (marital status, number of children, minimum child age, age) are low, women, young people, and younger people tend to leave, and the recruitment test has helping to stay in the job; in terms of job satisfaction, organizational and work environment factors: the distribution of performance compensation justice helps to improve organizational commitment, low overall satisfaction, does not meet employee expectations, salary dissatisfaction expectant, dislikes supervisor, no satisfied colleagues, poor cooperation with colleagues, lower social status, lower understanding of company organization and procedures, lower self-awareness of work-related abilities, increased conflicts and burdens at work, less opportunities for promotion. The low autonomy of work tends to leave the company, and the salary increase and the positive effect of Role-orientation can help employees to stay; At the level of work content compared with the external environment: the scope of work is positively related to the high-growth needs of employees and job satisfaction, high work repetition, high unemployment rate, high job opportunities, high intentions to find work, tend to leave, Compared with the outside world, the job satisfaction is high, and those who have high investment in work are helpful to staying in the job; in other projects: absentee, latecomer and poor performers tend to leave; in the sense of organizational retreat and behavior: high organizational commitment. Those who help to retain their jobs, have high intentions to leave their jobs, those who have low willingness to work, and the effectiveness of the evaluation work are more likely to leave than the current high.

Xu was using the “China Employer-Employee Matching Data Tracking Survey” data, using the Probit model system to analyze the factors influencing the departure of employees of state-owned enterprises [16]. Assuming that there is no off-the-job search, employees leave the company generally through three decision-making processes: The first is to measure the cost of job conversion, the second is to get a new job, and finally the desire and ability to accept a new job. We found the special characteristics of the departure of Chinese employees, for example, the turnover rate of agricultural permanent residence employees is relatively high; the factors affecting compensation are more complicated, and the benefits are due to the influence is weak; the opportunity to obtain new jobs depends on the macro environment, so the turnover rate is also affected by the regional economic level and wage level. In addition, there is a clear gender difference between men and women and a lower turnover rate for female employees.

Feng summarizes and analyzes the research situation of foreign work family conflict [17]. It is found that in recent years the number of papers and quotations in working family conflict show exponential growth. From the published journals and disciplinary attributes, mainly concentrated in high-level journals of the psychology and organizational behavior field, and disciplines of cross-fusion properties become increasingly evident, showing a diversified development trend. From the issue of countries and research institutions, mainly concentrated in the United States, Europe and other developed countries and regions, the influence of Chinese scholars in this academic field continues to rise, the number of published in the world ranks second. From the high citation and co-citation literature, the classical literature is still the main source of knowledge in this field, the literature review, the construction of different theories and variables, and the empirical research on the conflict of work family from different levels. The literature constitutes the main content of the classical literature research, and the theoretical and empirical research system of work family conflict has been formed. From the core author group, the number of papers published in the field become more, but has not yet formed a stable core of the author group. From the keywords collinear and strategic coordinates, pressure, satisfaction, gender, social support, resources, role conflict, health are high-frequency keywords, representing hot spots from 30 years. Strategic coordinates show that the role of conflict and social support research despite the frequency of the composition of the hot spots, but in the study of the external links and internal development has not yet formed a high degree of recognition of the research field, “work family conflict type”, “stress conflict and control”, “work life satisfaction” form the hotspot of work-family conflict research, “gender and emotional exhaustion”, “job performance”, “work family promotion”, “work family balance”, stress and “workplace needs”, “psychological distress”, “mental health and quality of work”, “working time flexibility and health” constitute the work of the family conflict in the hot areas, “character and emotions”, “shift work”, “social support and self-assessment”, “role conflict”, “time-based conflict and balance”, “organizational support and work attitude”, “workplace flexibility”, “female research” constitute a work family conflict relative Popular areas.

Ma adopting questionnaire survey to 29 manufacturing enterprises and 772 industrial workers as the research object, this study investigates the influencing mechanism of employee-oriented human resource practices and career growth on turnover intention by using multi-level linear model [18]. The empirical study results show that the employee-oriented human resource practices can significantly improve employee career growth and reduce turnover intention, and employee career growth is the cross-level mediating variable between employee-oriented human resource practice and turnover intention. The employee-oriented human resources practice has no significant moderating effect on the relationship between employee's career growth and employee turnover intention. But there are significant differences in the influence of employee career growth on turnover intention among different enterprises.

### **3 Data Processing and Descriptive Statistics**

To analyze the characters of turnover staff more precisely, this paper classifies and discretizes some indicators. The details are as follows:

#### **3.1 Age**

Continuous age data was discretized, with the 25–39-year-old employees were converted into “25–29 years old”, “30–34 years old” and “35–39 years old”. Due to the small number of employees under the age of 24 and over 40, they were not classified in detail, and are collectively referred to as “under 24 years old” and “40 years old or older”.

#### **3.2 Job Category**

Job categories are divided into five categories based on the work content of employees: “management”, “production”, “technical”, “distribution” and “support”. “Management” refers to the general manager and deputy general manager of the enterprise and the strategic management positions responsible for production planning and market development; “production” refers to the operators and on-site management personnel and equipment maintenance personnel of the production workshop of the enterprise; “technical” refers to technical positions such as quality inspection, technology development, and information system maintenance; “distribution” includes procurement positions and sales positions of the enterprise; “support” includes auxiliary positions such as administrative, financial, and personnel affairs of the enterprise.

### **3.3 Job Level**

Job level is divided based on employee's responsibility and power within their position: "senior manager", "mid-level manager", "junior manager" and "general employee", among which "senior manager" refers to the general manager of the company and deputy general manager; "mid-level manager" refers to the shop supervisor, department manager, etc.; "junior-level manager" refers to team leaders in the company's workshop.

Using 237 employee data of the company from 2015, and using SPSS 20.0 software to produce descriptive statistics of each indicator, the statistical results are shown in Table 1.

The results of descriptive statistical analysis show that the company's turnover rate in 2015 was 32.50%, which is relatively high. The employees of the company are mainly male, aged between 25–34 years old; non-local employees and agricultural household registry holders are relatively higher in percentage, and the education level overall is low, with the number of employees with vocational college education or above being less than 40%. Job positions are concentrated in quality inspection and technology research and development, and the job level distribution system present typical linear hierarchy characteristics.

### **3.4 Decision Tree Construction and Results Analysis**

70% of the 237 data were randomly selected as training samples. The SPSS modeler software was used to model the decision tree using C5.0 algorithm. The 8 variables of employee gender, marital status, household registration, registered permanent residence, age, time in current company, education, and job category were entered (only one of all departing employees belongs to the level of middle manager, with the rest being general employees, which holds no analytical significance, and so post level was not used as an input variable), a decision tree is constructed for the classification of employee turnover (on-the-job/departure) status as shown in Fig. 1.

Using the remaining 30% of the data as test samples, the prediction results of the decision tree are verified as shown in Table 2. The verification results show that the prediction accuracy rate is 88.16%, and the prediction result has relatively high accuracy.

**Table 1** Descriptive statistics of each indicator

<b>Turnover Status</b>	<i>Departed</i>		<i>Currently employed</i>		
	<b>32.50%</b>		<b>67.50%</b>		
Gender	<i>Male</i>		<i>Female</i>		
	86.10%		13.90%		
Marital Status	<i>Unmarried</i>		<i>Married</i>		
	51.50%		48.50%		
Household Registration	<i>Local</i>		<i>Non-Local</i>		
	6.30%		93.70%		
Registered permanent residence	<i>Urban</i>		<i>Rural</i>		
	20.30%		79.70%		
Age	<i>24 &amp; Below</i>	<i>25-29</i>	<i>30-34</i>	<i>35-39</i>	<i>40 &amp; Above</i>
	13.10%	44.70%	26.60%	8.00%	7.60%
Time in Current Company	<i>1 Year</i>		<i>2 Years</i>		
	20.30%		11.80%		
	<i>3 Years</i>		<i>4 Years</i>		
	12.70%		20.30%		
	<i>5 Years</i>		<i>6 Years</i>		
	15.60%		3.00%		
	<i>Less than a year</i>		16.50%		
Education	<i>Elementary School</i>		<i>Middle School</i>		
	0.42%		6.75%		
	<i>Vocational High School</i>		<i>High School</i>		
	21.94%		36.71%		
	<i>Vocational College</i>		<i>Undergraduate</i>		
	20.68%		11.39%		
	<i>Masters</i>		<i>Doctorate</i>		
1.27%		0.84%			
Job Type	<i>Management</i>		<i>Technical</i>		
	9.30%		28.30%		
	<i>Distribution</i>		<i>Production</i>	<i>Support</i>	
	5.50%		44.70%	12.20%	
Job Level	<i>Senior Manager</i>		<i>Junior manager</i>		
	1.27%		7.17%		
	<i>Mid-Level Manager</i>		<i>Regular Worker</i>		
	5.91%		85.65%		



**Table 2** Test results

Partition	Training sample		Testing sample	
	Quantity	Proportion (%)	Quantity	Proportion (%)
Accurate	150	93.17	67	88.16
	11	6.83	9	11.84
Accurate	161	100	76	100

Using rules set of the decision tree mining and the actual situation of the enterprise, meaningful rules found are as shown in Table 3.

Using the above rules, it is found that the employees leaving employment have the following three characteristics, and that the enterprise should adjust its human resource management method accordingly.

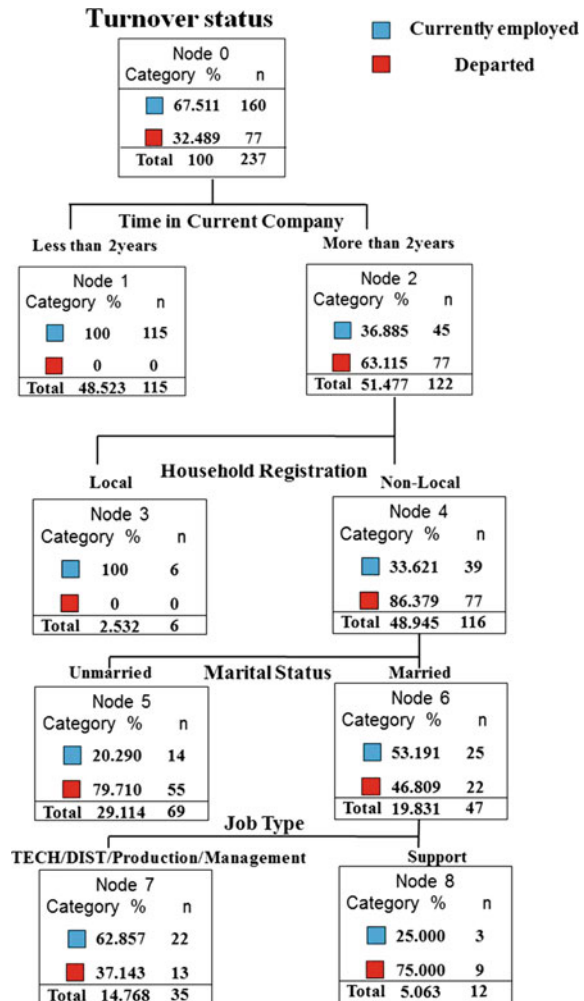
- (1) *Time in company is an important factor in staff turnover.* Employees employed for more than 3 years have a turnover rate higher than 60%. The reason behind this phenomenon might be due to problems in this company’s salary adjustment model, which may not be giving continuous stimulation to employees. The company should adjust their compensation policy in a timely fashion, to reflect the effect of work experience on salary level, and lower loss rate of older employees.
- (2) *Marital status has a relatively large influence on staff turnover, with unmarried staff having a significantly higher departure rate than married staff.* Most of the employees in this enterprise are composed of non-local males from rural areas. Unmarried males have no familial burdens, and therefore change their employment locations frequently. Whereas married male workers often come to cities to work as a family unit, with their spouse and children working and living in the same city, therefore having lower mobility rates. When employing workers, the enterprise should notice the influence of marital status, and increase the hiring proportion of married personnel.

**Table 3** Meaningful rules

No.	Condition IF	Result THEN
1	Time in Company in [1 year/2 years/Less than a Year]	Currently employed
2	Time in Company in [3 years/4 years/5 years/6 years] + Household Registration = Local	Currently employed
3	Time in Company in [3 years/4 years/5 years/6 years] + Household Registration = Non-Local + Marital Status = Married + Job Type in [Technical/Distribution/Production/Management]	Currently employed
4	Time in Company in [3 years/4 years/5 years/6 years] + Household Registration = Non-Local + Marital Status = Unmarried	Departed
5	Time in Company in [3 years/4 years/5 years/6 years] + Household Registration = Non-Local + Marital Status = Married + Job Type in [Support]	Departed

- (3) *Household registration is an important factor affecting staff turnover.* Staff with local or foreign household registration, married and highly skilled jobs is relatively stable, while those with non-local household registration, unmarried and engaged in simple labor, have a higher turnover rate. This shows that staff with non-local household registration is highly mobile. Staff with local household registration should be recruited as much as possible during the recruitment process. It also shows that household registration plays an important role in the human resource management.

**Fig. 1** The structure of the decision tree



## 4 Conclusion

This paper uses employee's gender, age, time in current company, education, marriage, household registration, registered permanent residence, and job category as input variables, constructs a decision tree with employee turnover status as the target variable, mining information related to employee turnover, and uses the results of decision tree analysis to make recommendations for the enterprise's human resource management. The research results show that when using the decision tree method to predict and analyze the actual turnover of employees, difficulty of data collection is relatively low, and the results have clear testable standards. Information produced by this method is more easily accepted and adopted by enterprises, and using this information, enterprises can grasp the rules of employee turnover more fully, and adjust their human resources management activities accordingly, moreover, the data-driven employee turnover prediction method is mainly based on objective experiments and is not affected by subjective factors. Therefore, the proposed method can be integrated into the support decision-making system to help improve the employee turnover behavior prediction ability of human resources decision makers; furthermore, the analysis of the main factors affecting employee turnover can help enterprise decision makers to target employee turnover. They are inclined to adopt corresponding coping plans, or formulate policies to try to retain excellent employees, or take measures to avoid the enterprise losses caused by staff turnover to the greatest extent.

## References

1. Yang, C., & Dong, X. (2016). Research on warning and management of enterprise employee turnover crisis—based on a non-language communication perspective. *Economic Problems*, (8), 83–86.
2. Xu, F., & Zhao, Z. (2015). Analysis of enterprise employee turnover behavior—an economic review of China's experience. *Economic Perspectives*, (11), 55–67.
3. Yang, D. (2012). Efficiency logic or power logic—corporate politics and compulsory resignation of CEOs of listed companies. *Chinese Journal of Sociology*, 32(5), 151–178.
4. Zou, P., & Dong, Y. (2015). Analysis on the countermeasures of employee turnover crisis in university libraries based on game theory. *Information Science*, (1) 43–48.
5. Wang, X., Li, A., Xiong, G., & Sun, H. (2016). Research on the impact of job requirements of frontline employees on turnover intention and chain adjustment mechanism. *Journal of Management and Business Research*, 13(8), 1191–1198.
6. Shan, H., Hu, E., Bao, J., & Zhang, M. (2015). Research on the influence of level of employee status perception on intention of leaving employment in non-state-owned enterprises. *Journal of Management and Business Research*, 12(8), 1144–1153.
7. Ye, R., Ni, C., & Huang, S. (2015). The impact of workplace exclusion and workplace marginalization on employee intention of leaving employment: the adjustment of employee performance. *Management Review*, 27(8), 127–140.
8. Zhang, Y., & Zhang, Z. (2015). A customer loss prediction model based on C5.0 decision tree. *Statistics & Information Forum*, (1), 89–94.

9. Gu, B., & Xu, J. (2015). User loyalty mining based on knowledge sharing in professional virtual communities. *Information Science*, (1), 105–110.
10. Peng, X., Zhu, Q., & Liu, W. (2015). Analysis of microblog user features and classification research—using Sina Weibo as a case study. *Information Science*, (1), 69–75.
11. Sun, Y., Shao, C., Zhao, D., & Ouyang, S. (2014). C5.0 decision tree prediction model of traffic accident severity. *Journal of Chang'an University (Natural Science Edition)*, 34(5), 111–116.
12. Cui, X. (2003). Research on the influence of personal characteristics of employees on organizational commitment and intention of leaving employment. *Nankai Business Review*, 6(4), 4–11.
13. Wang, G. (2015). Comparative study on commitment differences of enterprise information workers organizations from the perspective of individual characteristic variables. *Journal of Guangzhou University (Social Science Edition)*, 14(5), 47–53.
14. Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. (1979). Review and conceptual analysis of the employee turnover process. *Psychological Bulletin*, 86(3), 493–522.
15. Rodger, W. G., Peter, W. H., & Stefan, G. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26(3), 463–488.
16. Xu, F., & Zhao, Z. (2015). Analysis of the turnover behavior of enterprise employees—economic review of chinese experience. *Economic Perspectives*, 55–66.
17. Feng, L. (2018). Research situation of work family conflict in foreign countries: based on bibliometric analysis. *Business Management Journal*, 4, 187–208.
18. Ma, W., Yan, K., Chen, X., & Weng, Y. (2019). The influencing mechanism employee-oriented human resource practice and career growth on turnover intention: a multi-level model based on industrial workers in the new era. *Soft Science*, 1–11.