Employee Attrition:A Classic Problem revisited via Machine Learning

Ponnur Srujan, Sourav Poddar, Charan Simha Reddy, Poonam Joshi
Project Guide: Dr.Snehanshu Saha

October 10, 2017

1 Abstract

Employee turnover (attrition) is a major cost to an organization, and predicting turnover is at the forefront of needs of Human Resources (HR) in many organizations. Until now the mainstream approach has been to use logistic regression or survival curves to model employee attrition. However, with advancements in machine learning (ML), we can now get both better predictive performance and better explanations of what critical features are linked to employee attrition. We intend to develop a predictive model that is in the same ballpark as commercial products in terms of ML accuracy. Then we'll use the variable importance plots (feature ranking) to identify key parameters for training and testing.
Key Terms: Predictive Analysis, Workforce Analysis, Machine Learning, Ensemble Methods

2 Introduction

Organizations face huge costs resulting from employee turnover. Some costs are tangible such as training expenses and the time it

takes from when an employee starts to when they become a productive member. However, the most important costs are intangible. Consider what's lost when a productive employee quits: new product ideas, great project management, or customer relationships. With advances in machine learning and data science, it's possible to not only predict employee attrition but to understand the key variables that influence turnover.

Understanding people and why they decide to stay at or leave a job is arguably one of the most important questions for HR to answer. Identifying attrition risk calls for advanced pattern recognition in surveying an array of variables. If a human were to try and detect attrition risk among engineers in Palo Alto with less than 2 years of tenure, the variance analyses to reach that conclusion are innumerable, like finding a needle in haystack, but machine learning allows us to connect these dots in seconds, freeing HR representatives to spend time supporting teams instead of analyzing data.

3 Literature Survey

Field literature review shows that these general causes affecting voluntary turnover intention could be grouped in the following broad categories: 1) organizational causes, related to: Human Resource Management (HRM), organizational culture, job structure and content, leadership style and internal marketing strategies, policies and practices; 2) extra organizational causes, related to: individual characteristics and labor market aspects.

Organizational culture was found to have an affect on organizational behavior and may be proposed as a potential moderator of voluntary turnover. Tziner et al., 2012, in their study hypothesized the moderating effects of the dimensions of organizational culture on the relationship between Organizational Citizenship Behavior (OCB) and turnover intentions.

In another study organizational culture dimensions only partially met the initial requirement for a variable to be defined as moderator, where the variable was linearly uncorrelated to the predictor variable and was also uncorrelated to the criterion variable

(Ibid and in Zedeck, 1971; Keppel and Zedeck,1989). In addition results from a sample of 102 employees showed solid support for the interactive effects of the dimensions of organizational culture and OCB in relation to turnover intentions (Cohen and Zaidi, 2002; Tang and Wang, 2005). In addition, there is evidence that perceptions of organizational politics, which can be highly salient aspects of organizational culture, were positively related to turnover intentions (Chang et al.,2009).

4 Methodology

4.1 Features and data:

Features

- Total prior jobs, Position changes

- Average Tenure Before, Total Experience

- Working Hours, Salary, et cetera.

Data

- IBM Watson HR Employee Attrition and Performance

- Kaggle HR Data Set

4.2 Ensemble Clustering:

Attrition prediction is a scenario that takes historic employee data as input to then identify individuals that are inclined to leave. The basic procedure is to extract features from the available data that might have previously been manually analyzed and to build predictive models based on a training set with labels relating to the employment status. Normally it can be formalized as a supervised classification problem, while the uniqueness is that population of employees with different employment status may not be equal. Training such an imbalanced data set requires re-sampling or cost-sensitive learning techniques. For sentiment analysis on unstructured data such as text, preprocessing techniques that extract

analysis-friendly quantitative features should be applied. Commonly used feature extraction methods for text analysis include word-to-vector, term frequency, or term frequency and inverse document frequency, etc. Algorithms for building the model depend on the data characteristics. In case a specific algorithm does not yield the desired results, we have found that ensemble techniques can be deployed to effectively boost model performance.

5 Outcome/Conclusion

- Impact analysis

- Causality between two different time series data- Attrition and company turnover

- Identify key features

6 References

- Ladelsky Limor Kessler ,THE EFFECT OF ORGANIZATIONAL CULTURE ON IT EMPLOYEES TURNOVER INTENTION IN ISRAEL

- Zheng WeiBo1*, Sharan Kaur 2and Tao Zhi 3,A critical review of employee turnover model (1938-2009) and development in perspective of performance, African Journal of Business Management Vol. 4(19), pp. 4146-4158, December Special Review, 2010

- W. Stanley Siebert,Nikolay Zubanov,Arnaud Chevalier,Tarja Viitanen, Labour Turnover and Labour Productivity in a Retail Organization, IZA Discussion Paper No. 2322 September 2006

- Jacob Rubæk Holm, The effects on performance of voluntary and involuntary labour turnover in an evolutionary signalling model