Quantile Regression for Workforce Analytics

Karthikeyan Natesan Ramamurthy, Kush R. Varshney, and Moninder Singh
Business Analytics and Mathematical Sciences Department
IBM Thomas J. Watson Research Center, Yortktown Heights, NY USA 10598
{knatesa,krvarshn,moninder}@us.ibm.com

***Abstract*—Understanding the behavior of a constantly changing workforce is key to making business decisions in modern organizations. In this paper, we develop frameworks based on quantile regression to estimate the productivity and attrition profiles of employees from revenue, headcount, and incentive data. Results show the advantages of quantile-specific profiles compared to those obtained with other regression schemes.**

***Index Terms*—productivity profile, attrition profile, quantile regression, workforce behavior**

I. Background

Workforce analytics is a broad area comprising many scientific techniques that help in understanding and predicting the behavior of the workforce in a business using available data. It has a huge potential for improving the profitability of business. Two key determiners of the financial performance of an organization at the workforce level are the productivity, and the constant influx and outflux of its employees. New employees have to be sufficiently trained before they can reach the maximum level of productivity, and attrition causes a temporary reduction in the overall productivity. Hence, it will be beneficial for any business to complement its existing information technology tools for improving customer satisfaction with workforce analytics methods that focus on its employees.

Many existing methods for modeling productivity and attrition of the workforce do not consider the temporal information. However, the actual workforce behavior is event-driven, and time-dependent. One example is that providing incentives to employees may result in improved retention [1], whereas lack of incentives for a long time may lead to attrition. Furthermore, in order to help with robust business planning and decision making, it necessary to understand the behavior of the workforce at various levels: from best-case to worst-case. This can be achieved using quantile-specific profiles. In this paper, we propose a workforce analytics framework based on quantile regression [2] to estimate time-varying attrition and productivity profiles from incentive and revenue data respectively, and demonstrate their predictive power.

II. Proposed Framework and Results

The proposed framework will be described for estimating profiles based on attrition and promotion data. The same framework can also be used with other types of incentives such as salary raise. The basic intuition behind this model is that for any employee, the propensity to quit the company, $h[t]$, depends on the number of months, t, since the last promotion. Given that the number of attrits in a month n is $y[n]$, and the number of actives in month n who were promoted t months ago is $x_n[t]$, the attrition model is given as $y[n] = \sum_{t=0}^{T-1} x_n[t]h[t]$, where T is the length of the profile. When this data is available for several months, the profile h can be estimated using a regression model. We used the quantile regression approach proposed in [2], and solved a linear program to estimate profile for a given quantile $0 < q < 1$. The high (low) quantile-profile models the behavior of the workforce who have a higher (lower) inclination to attrit.

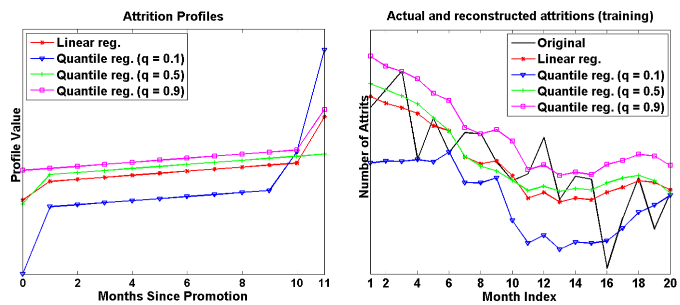

Fig. 1. Attrition profiles of employees and attrition fit using the profiles.

The proposed framework was used to obtain year-long profiles using 26 months of attrition and 38 months of promotion data in an organization. The first 20 months of attrition data were used to train the proposed model and the rest of the data were used to test its predictive power. The profiles were constrained to be non-negative and also non-decreasing with minimum and maximum slopes. For comparison, we used linear regression with the same constraints. The profiles and the fitted attritions are given in Figure 1. Comparing the figures, it can be seen that the low-quantile and high-quantile profiles model the extreme scenarios, whereas the medium quantile profile is close to that of the average case. Note that the ordinates in the plots are hidden due to confidentiality of the data. The normalized RMSE with test data for quantile regression ($q = 0.5$) was 0.136, whereas for linear regression it was 0.154. Using this framework to model productivity profiles of newly hired salespeople with sales revenue and headcount data [3] results in similar informative profiles.

References

[1] M. Singh, K. R. Varshney, J. Wang, A. Mojsilović, A. R. Gill, P. I. Faur, and R. Ezry, "An analytics approach for proactively combating voluntary attrition of employees," in *Proc.IEEE ICDMW*, 2012.

[2] R. Koenker and G. Bassett, Jr., "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, Jan. 1978.

[3] K. R. Varshney, M. Singh, M. Sharma, and A. Mojsilović, "Estimating post-event seller productivity profiles in dynamic sales organizations," in *Proc.IEEE ICDMW*, 2011.