HARVARD
School of Engineering and Applied Sciences

CS281: Advanced Machine Learning, Fall 2013

Forecasting Attrition in the USCG Workforce

Isaac M. Slavitt and Charles L. Hornbaker

Introduction

The United States Coast Guard has a highly structured enlisted workforce of over 33,000 members with set occupational specialties and a rigid path of advancement.

Each year, Coast Guard human resources managers must decide:

- How many promotions to carry out in a certain year within each paygrade of each enlisted specialty ("rating").
- How many new enlisted personnel to admit into boot camp at Training Center (TRACEN) Cape May, and how many newly accessed personnel to allow into a each rating.
- Whether or not special workforce management tools are needed, such as monetary incentives for reenlistments within a certain rating.

Problem Statement

How many members will leave the service next year?

Methods and Results

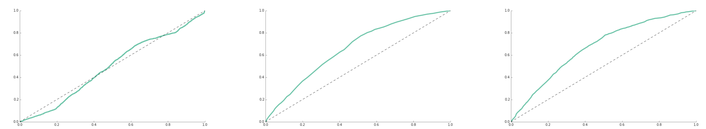
The data spanned FY 2005-2013. For each year, we had a snapshot of the entire enlisted workforce on the last day of the FY with each member's service and demographic information.

Feature selection

To achieve shrinkage, we compared fitting the data with ordinary ($\ell 2$) versus LASSO ($\ell 1$) logistic regression.[1]

	$\ell 2$	lasso
education	-0.009766	0.000000
female	-0.045346	0.000000
married	-0.102824	0.000000
minority	-0.062625	0.000000
tis	-0.031722	-0.054656

Table 1: Coefficient weights from different penalties.



(a) Logistic regression (AUC = 0.45) **(b)** k nearest neighbors (k = 101) (AUC = 0.66) **(c)** AdaBoost (AUC = 0.66)

Figure 1: ROC curves of various classification models.

As shown in Table 1, time in service alone was the most predictive variable; likely due to the extra dimensionality introduced by the other variables. Figure 1 shows various classifiers acting on individual records.

Modeling the service

Another way to model attrition is through a service-wide profile.

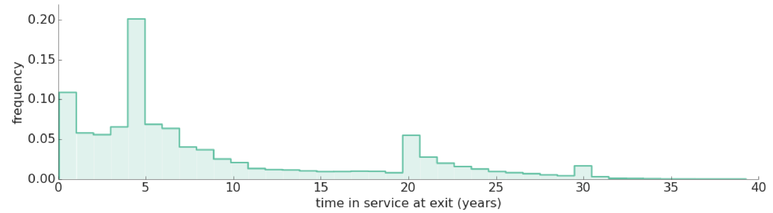


Figure 2: Histogram of time in service at exit.

We modeled the probability of leaving at a given time in service with a Gaussian Mixture Model (GMM).[2] Figure 3 shows the diagram for this model, while Figure 4 shows the resulting fit.

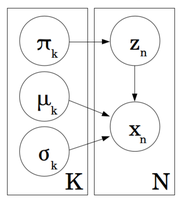


Figure 3: Plate diagram of Gaussian mixture model.

The generative story for this model can be interpreted as saying that employees have some kind of latent identity (e.g. "early wash-out" or "lifetime career") which informs their probability of leaving the service.

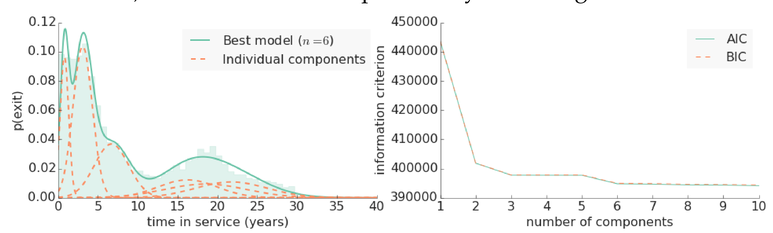


Figure 4: Results from fitting Gaussian mixture model.

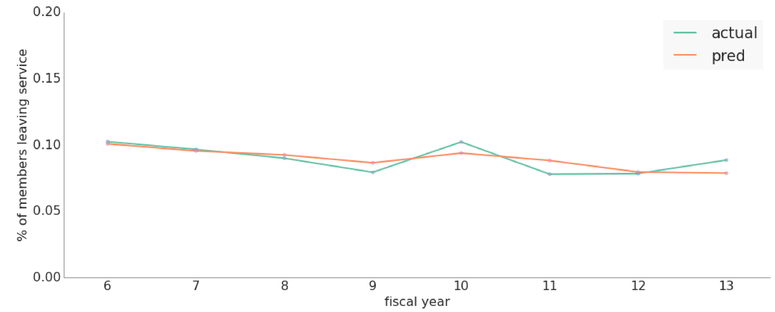


Figure 5: Predicted versus actual attrition, FY 2006-2013.

From our GMM, we obtain P(member will leave in the next FY). Selecting a sensible threshold based on historical data, we are able to treat our GMM as a classifier. Figure 5 shows the predicted attrition rates from the model.

Conclusions

The Coast Guard's current forecasting method is to review historic attrition rates and monitor current trends in order to guess the next three years' attrition rate. Beyond preliminary investigation, the true influences on attrition have never been quantified.[3] We present a model that can predict the attrition of USCG enlisted members with performance that constitutes a meaningful improvement over the current method of prediction, and traditional logistic regression, which would be the conventional approach to this type of problem. In addition, this is a model which can easily incorporate and fit new data, so that it can continue to be used in future years.

References

[1] K. P. Murphy, *Machine Learning: a Probabilistic Perspective*. Cambridge, MA: MIT Press, 2012.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[3] U.S. Coast Guard, *Fiscal Year 2012 Active Duty Enlisted Workforce Management Plan (Internal/Predicisional)*. Washington, DC: U.S. Government, 2011.

Acknowledgments

We are very grateful to the United States Coast Guard's Office of Workforce Forecasting and Analysis (CG-12A) for providing us with the data for this project. In particular, we would like to thank Mr. Carl Ellis for his help in getting the project approved and LT Ginny Gustafson for her continuing advice and feedback.