Application of Random Forests and Deep Neural Networks to Survival Data

STAT-6494 Project Proposal

Wenjie Wang[*]

06 March 2018

Abstract

The classical survival models, such as Cox proportional hazard model, often require extensive efforts on variable selection or prior medical information to model interaction between patients' covariates and treatment covariates. While nonlinear models, such as neural networks and random forests, are able to model high-order interaction terms. It is of interest to apply these machine learning methods to survival data and compare their performance with classical statistical models.

Keywords: Cox Model, Machine Learning, Suicide Prevention

[*]wenjie.2.wang@uconn.edu; Ph.D. student at Department of Statistics, University of Connecticut.

1 Introduction and Objects

For survival data, medical researchers' interests often lie in discovery of significant treatment effects and important diagnosis covariates of patients. The classical survival models, such as Cox proportional hazard model, assume risk function in a simple linear form of covariates, which can be too simplistic to capture the underlying relationship between response and covariates. In addition, they often require extensive efforts on variable selection or prior medical information to model interaction between patients' covariates and treatment covariates. While nonlinear models, such as neural networks and random forests, are able to model high-order interaction terms. It is of interest to apply these machine learning methods to survival data and compare their performance with classical statistical models. It would be even more interesting to discover nonlinear relationship by machine learning methods and build a statistical model for better interpretation and capability for statistical inferences.

The specific objectives include:

- Explore and review existing machine methods for survival data including random forests and deep neural networks.
- Apply these methods for CT suicidal data.
- Compare the out-of-sample model fitting or prediction performance of these methods with classical survival models, such as Cox model.

2 Random Forests for Survival Data

Random forests (RF) proposed by Breiman (2001) is an ensemble tree method that introduces randomization to the base learning process. Breiman (2001) showed that RF may further improve the prediction performance of simple ensemble learning method. Ishwaran et al. (2008) extended RF method to random survival forests (RSF) method for analysis of right-censored survival data.

Other reference includes

- Strobl et al. (2007)
- Mogensen et al. (2012)

3 Deep Neural Networks for Survival Data

The regular Cox proportional hazards model has a linear relative risk function $r(\boldsymbol{x}, \boldsymbol{\beta}) = \boldsymbol{\beta}^\top \boldsymbol{x}$. In many applications, it is hard to assume a linear proportional hazards condition and thus high-level interaction terms are required. However, as the number of covariates and interactions increases, it becomes prohibitively expensive.

Katzman et al. (2016) proposed a Cox proportional hazards deep neural network method called DeepSurv for personlized treatment recommandations. DeepSurv is a multi-layer perceptron that predicts a patient's risk of death. The output of the network is a single node estimating the relative risk function $\hat{r}_\theta$ by the weights of the network θ.

Other reference includes

- Nair and Hinton (2010)
- Ioffe and Szegedy (2015)
- Klambauer et al. (2017)
- Srivastava et al. (2014)
- Kingma and Ba (2014)
- Nesterov (2013)
- Pascanu et al. (2012)

Reference

Breiman, L. (2001), "Random Forests," *Machine Learning*, 45, 5–32.

Ioffe, S. and Szegedy, C. (2015), "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in *International conference on machine learning*, pp. 448–456.

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., and Lauer, M. S. (2008), "Random Survival Forests," *The annals of applied statistics*, 841–860.

Katzman, J., Shaham, U., Bates, J., Cloninger, A., Jiang, T., and Kluger, Y. (2016), "DeepSurv: Personalized Treatment Recommender System Using a Cox Proportional Hazards Deep Neural Network," *ArXiv e-prints*.

Kingma, D. P. and Ba, J. (2014), "Adam: A Method for Stochastic Optimization," .

Klambauer, G., Unterthiner, T., Mayr, A., and Hochreiter, S. (2017), "Self-Normalizing Neural Networks," in *Advances in Neural Information Processing Systems*, pp. 972–981.

Mogensen, U. B., Ishwaran, H., and Gerds, T. A. (2012), "Evaluating Random Forests for Survival Analysis Using Prediction Error Curves," *Journal of Statistical Software*, 50, 1–23.

Nair, V. and Hinton, G. E. (2010), "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814.

Nesterov, Y. (2013), "Gradient Methods for Minimizing Composite Functions," *Mathematical Programming*, 140, 125–161.

Pascanu, R., Mikolov, T., and Bengio, Y. (2012), "Understanding the Exploding Gradient Problem," *CoRR*, abs/1211.5063.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014), "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *The Journal of Machine Learning Research*, 15, 1929–1958.

Strobl, C., Boulesteix, A.-L., Zeileis, A., and Hothorn, T. (2007), "Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution," *BMC Bioinformatics*, 8, 25.