Nakamura, Masatoshi; Ochi, Yoshimichi; Motogaito, Hiroki; Goto, Masashi
Garrote trees as tree structured regression analysis. (English) Zbl 1407.62261
J. Jpn. Soc. Comput. Stat. 30, No. 1, 65-80 (2017).

Summary: In regression analysis, stochastic models are often constructed to model relationships between outcomes and explanatory variables. We derive statistical interpretation about the underlying structure of data based on these models. When we use a linear regression model and the model provides good fitting to the data, it is straightforward to interpret the relation. However, there are cases where it may be difficult to formulate a linear model reflecting actual characteristics in detail. In such cases, a tree-structured approach is recommended, such as classification and regression trees (CART), which develops a tree and provides an interpretation of the data based on the fundamental model derived from the tree. Random forest (RF) involves an ensemble learning method based on the trees and can predict outcomes more precisely. However, RF cannot provide a tree-structured model for interpreting the data. We examine a nonnegative garrote (NNG), a shrinkage estimator, and propose garrote trees (GT) as an adjustment of RF based on NNG. In addition, GT can lead making trees that are useful for interpretation of data. Two case studies of diabetes and prostate cancer data illustrate predictive accuracy and descriptive features of GT. Finally, our simulation studies show that the proposed method is highly accurate predictively and provides a potential ability to interpret the data from new meaningful standpoints.

MSC:
62J05 Linear regression; mixed models
62J07 Ridge regression; shrinkage estimators (Lasso)
62H30 Classification and discrimination; cluster analysis (statistical aspects)
68T05 Learning and adaptive systems in artificial intelligence

Keywords:
tree-structured approaches; ensemble learning; CART; random forest; nonnegative garrote; classification

Full Text: DOI

References:
[1] Banerjee, M., Ding, Y. and Noone, A. M. (2012). Identifying representative trees from ensembles. Statistics in Medicine, 31, 1601-1616.
[2] Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). Classification and Regression Trees. Belmont, CA: Wadsworth.
[3] Breiman, L. (1995). Better subset regression using the nonnegative garrote. Technometrics, Zbl 0862.62059
[4] Breiman, L. (2001). Random forests. Machine Learning, 45, 5-32. Zbl 1007.68152
[5] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. The Annals of Statistics, 32, 407-499. Zbl 1091.62054
[6] Friedman, J. H. (1991). Multivariate adaptive regression splines. The Annals of Statistics, 19, 1-67. Zbl 0765.62064
[7] Friedman, J. H. and Popescu, B. E. (2008). Predictive learning via rule ensembles. Annals of Applied Statistics, 2, 916-954. Zbl 1149.62051
[8] Hoerl, E. and Kennard, R. W. (1970) Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12, 55-67. Zbl 0202.17205
[9] Meinshausen, N. (2009). Forest garrote. Electronic Journal of Statistics, 3, 1288-1304. Zbl 1326.62093
[10] Nakamura, M., Shimokawa, T., Sakamoto, W. and Goto, M. (2013). Regression analysis using lasso random forest, Bulletin of the Computational Statistics of Japan, 26, 1-15.
[11] Rosa, J. C., Veiga, A. and Medeiros, M. C., (2008). Tree-structured smooth transition regression models based on CART algorithm. Computational Statistics and Data Analysis, 52, 2469-2488. Zbl 1452.62044
[12] Sugimoto, T., Simokawa, T. and Goto, M. (2005). Tree-structured approaches and recent advances. Bulletin of the Computational Statistics of Japan, 18, 123-164.
[13] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, 58, 267-288. Zbl 0850.62538
[14] Zou, H. and Trevor H. (2005). Regularization and variable selection via the Elastic Net. Journal of the Royal Statistical Society, 67, 301-320. - Zbl 1069.62054

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.