Yongmiao Hong, Oliver Linton, Jiajing Sun, and Meiting Zhu's contribution to the Discussion of "the Discussion Meeting on Probabilistic and statistical aspects of machine learning"

Hong, Yongmiao; Linton, Oliver; Sun, Jiajing; Zhu, Meiting

DOI:
10.1093/jrsssb/qkad152

Yongmiao Hong, Oliver Linton, Jiajing Sun, and Meiting Zhu's contribution to the Discussion of 'the Discussion Meeting on Probabilistic and statistical aspects of machine learning'

Yongmiao Hong†, Oliver Linton‡, Jiajing Sun§, & Meiting Zhu¶
Academy of Mathematics and Systems Science, Chinese Academy of Sciences†
Faculty of Economics, University of Cambridge‡
School of Mathematics, University of Birmingham§
School of Economics, Xiamen University¶

This paper proposes a neural network-based approach for automating offline change-point detection. The authors show that CUSUM and generalized CUSUM are a special case of their neural network class. They emphasize misclassification error rates and their theoretical contribution is to establish some elegant results for these under i.i.d. unit variance Gaussian data with a possible change in mean. Their theoretical results outline the conditions on the ERM neural network that allow it to achieve comparable performance to the classic CUSUM test for this very specific setting. The framework relies on N training data samples that are independent and identical copies where $N >> n^2 \log n$, which seems like a lot of training is needed! The CUSUM test only needs the sample of size n. In many financial applications we have a single time series $\{X_1, \ldots, X_n\}$ and we are interested in when and how change-points occur throughout the whole observation period. It is not clear how or why we should divide the data into training and testing samples and some guidance on this would be appreciated. In those applications considerable care needs to be taken in how to account for time series dependence and where this is treated nonparametrically issues arise with bandwidth selection. The self-normalization method proposed by Shao (2010) does not require the choice of tuning parameter and so is also automatic, and is quite widely used for identifying change-points within a given dataset, Shao and Zhang (2010). However, this method is known to have poor power properties. Recently, Hong et al. (2022) proposed the adjusted-range based self-normalization, in-

stead of the usual long run variance normalization, and this appears to work better under some long memory alternatives.

The current paper avoids the normalization issue altogether by making use of training data, and perhaps where that exists their method can have advantages, but it would be interesting to know how competitive their method is on financial time series and on realistic sampling schemes with longer range dependence.

References

[1] Hong, Y., O. B. Linton, B. McCabe, J. Sun, & S. Wang (2023). Kolmogorov-smirnov type testing for structural breaks: A new adjusted-range based self-normalization approach. To appear in *Journal of Econometrics*. Available at SSRN 3850894.

[2] Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3), 343–366.

[3] Shao, X., & Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491), 1228–1240.