M.N.M. van Lieshout and C. Lu's contribution to the Discussion of 'the Discussion Meeting on Probabilistic and statistical aspects of machine learning'

M.N.M. van Lieshout[1,2] and C. Lu[2]

[1] CWI, Amsterdam, The Netherlands
[2] University of Twente, Enschede, The Netherlands

Correspondence: M.N.M. van Lieshout, CWI, Amsterdam, The Netherlands. Email: Marie-Colette.van.Lieshout@cwi.nl.[1]

We congratulate Professors Li, Fearnhead, Fryzlewics and Wang on their fine work that represents classic test statistics for change-point detection as a neural network based classifier and develops improved offline detection algorithms for historical, labelled data.

The paper focuses on real-valued observation vectors in a temporal regression model with training data being either labelled historical data or obtained by simulation from a model. In many applications though, the data consist of both spatial and temporal components. For instance, the observations may take the form of a series of point patterns (e.g., mapped tree locations at different census times) or a single observation from a spatio-temporal point pattern (e.g., occurrence locations and times of fire incidents). For the latter, detection of changes in intensity in a model-based Bayesian test setting was investigated by Altieri et al. (2015). Do the authors believe that an adapted neural network approach could be competitive in this context? A complication would be that the change-point is due to complex changes in inter-point interaction for which neither a known model nor labelled historical data is available. Do the authors see a way forward here?

Machine learning ideas could benefit spatio-temporal statistical practice more widely. Specifically for point pattern analysis, Lu et al. (2023) employed random forest importance scores for variable selection, whilst Jalilian et al. (2023) trained neural networks on simulated data to distinguish spatial structural differences. A similar motivation as that of Li et al. is seen in point process intensity estimation. Usually, the intensity function is assumed to be log-linear in spatial and temporal covariates. Lu et al. (2023), proposed a tree-based model, XGBoostPP, which forms the intensity function based on a covariate vector $\boldsymbol{z}(\boldsymbol{s})$ as

$$\log\{\lambda(\boldsymbol{s})\} = \sum_{k=1}^{K} f_k\{\boldsymbol{z}(\boldsymbol{s})\}.$$

[1]This research was funded by the Dutch Research Council (NWO) for the project 'Data driven risk management for fire services' (18004).

Here, $f_k\left\{\boldsymbol{z}(\boldsymbol{s})\right\}$ are tree predictors that output the response on the leaf where a covariate value $\boldsymbol{z}(\boldsymbol{s})$ lies in. For model fitting, we customized a penalized weighted Poisson log-likelihood loss function

$$\sum_{k=1}^{K}\Omega(f_k) - \sum_{\boldsymbol{x}\in X} w(\boldsymbol{x})\log\left\{\lambda(\boldsymbol{x})\right\} + \int_S w(\boldsymbol{s})\lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s},$$

where X denotes the point process on S and $\Omega(f_k)$ is proportional to the L_1-norm of leaf responses. The tree structures and corresponding leaf responses are optimized iteratively; the weights w are calculated based on the estimated inhomogeneous K-function. The classic log-linear intensity function can be represented as a reparameterized XGBoostPP; neural networks may offer alternatives.

In the reverse direction, the STIT tessellation (Nagel and Weiss, 2005) from spatial statistics can be used for partitioning responses (cf., Ge et al., 2019).

REFERENCES

Altieri, L., Scott, E. M., Cocchi, D. and Illian, J. B. (2015). A changepoint analysis of spatio-temporal point processes. *Spat. Stat.*, 14, 197–207.

Jalilian, A. and Mateu, J. (2023). Assessing similarities between spatial point patterns with a Siamese neural network discriminant model. *Adv. Data. Anal. Classi.*, 17, 21–42.

Lu, C., Van Lieshout, M. N. M., De Graaf, M. and Visscher, P. (2023). Data-driven chimney fire risk prediction using machine learning and point process tools. *Ann. Appl. Stat.*, to appear.

Lu, C., Guan, Y., Van Lieshout, M. N. M. and Xu, G. (2023). XGBoostPP: Tree-based estimation of point process intensity functions, working paper.

Nagel, W. and Weiss, V. (2005). Crack STIT tessellations: Characterization of stationary random tessellations stable with respect to iteration. *Adv. Appl. Probab.*, 37, 859–883.

Ge, S., Wang, S., Teh, Y. W., Wang, L. and Elliott, L. (2019). Random tessellation forests. In *NeurIPS 32*.