

Random forests – applications and developments

Several master thesis topics available.

The random forest algorithm [1] has become one of the most popular machine learning algorithms, thanks to its wide applicability, high predictive performance and low computational cost, e.g., in a recent investigation it turned out to be the winning approach out of 179 compared algorithms on over 120 datasets [2]. There are a number of aspects of the algorithm that need further study and which possibly can lead to improvements, including the handling of high-dimensional sparse data, missing values, parallelization, voting strategies, providing statistical guarantees and interpretability.

[1] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001

[2] M. Fernandez-Delgado, E. Cernadas, S. Barro, and D. Amorim. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? *The Journal of Machine Learning Research*, 15(1), pp. 3133–3181, 2014.

Please contact:

Henrik Boström
Professor in computer science - data science systems
KTH Royal Institute of Technology
School of Information and Communication Technology
Electrum 229, SE-164 40, Kista, Sweden
Phone: +46 790 43 06