Performance Comparison of Random Forests and Extremely Randomized Trees

Eftim Zdravevski

Performance Comparison of Random Forests and Extremely Randomized Trees

Eftim Zdravevski, Petre Lameski, Andrea Kulakov, Vladimir Trajkovik
Faculty of Computer Science and Engineering
Ss.Cyril and Methodius University, Skopje, Macedonia
Email: {eftim.zdravevski, petre.lameski, andrea.kulakov, vladimir.trajkovik}@finki.ukim.mk

***Abstract*—Random Forests (RF) recently have gained significant attention in the scientific community as simple, versatile and efficient machine learning algorithm. It has been used for variety of tasks due it its high predictive performance, ability to perform feature ranking, its simple parallelization, and due to its low sensitivity to parameter tuning. In recent years another tree-based ensemble method has been proposed, namely the Extremely Randomized Trees (ERT). These trees by definition have similar properties. However, there is no extensive empirical evaluation of both algorithms that would identify strengths and weaknesses of each of them. In this paper we evaluate both algorithms of several publicly available datasets. Our experiments show that ERT are faster as the dataset size increases and can provide at least the same level of predictive performance. As for feature ranking capabilities, we have statistically confirmed that both provide the same ranking, provided that the number of trees is large enough.**

Keywords*—*Random Forests, Extremely Randomized Trees, Decision Trees, Ensembles of Trees

I. INTRODUCTION

In recent years, ensemble methods for machine learning has been used extensively in the research community, as well as by industry practitioners. Classification predictive performance have been improved by growing an ensemble of trees and that vote for the most popular class. In particular, a very popular method for ensemble of decision trees is the Random Forests algorithm [1]. Another recently popularized method is the Extremely Randomized Trees [2].

The paper is organized as follows. The next subsection describes briefly these algorithms. Next is described the method for feature extraction used in the experiments. Section IV described the experimental setup and discusses the results. Finally, in section V we conclude the paper.

II. RANDOM FORESTS AND EXTREMELY RANDOMIZED TREES

Random forests are a combination of decision trees in which each tree depends on the values of a random feature space sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges a.s. to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them.

In [2] proposed the Extremely Randomized Trees tree-based ensemble method for supervised classification and regression problems. It randomizes both attribute and cut-point choice while splitting a tree node. In the extreme case, it builds totally randomized trees whose structures are independent of the output values of the learning sample. The strength of the randomization can be tuned to problem specifics by the appropriate choice of a parameter.

III. FEATURE EXTRACTION AND SELECTION

Extracting robust features from time series is a challenging task, but using a systematic approach, for our experiments we are generating a variety of features. A recent data mining competition for posture recognition of firefighters [3] was able to inspire different feature engineering approaches that are very effective [4], [5], [6]. Additionally, for feature selection we have used a ranking method, proposed in [7], the based that is also able to detect features that are subject to data drift, therefore that can potentially degrade performance over time. As a result of the evaluation of various feature subsets, we can analyze the performance of both classifiers.

IV. RESULTS

In order to evaluate the performance of both algorithms, we have selected a publicly available dataset for activity recognition [8]. This dataset, which is extensively described in [9], consists of raw readings of 1 chest-mounted accelerometer with 3 axes. The subjects were performing 7 different actions, but for experimenting the authors that published this dataset have been using only the 5 actions, so we did as well. The total number of instances in the dataset was 73899.

On Fig. 1 and Fig. 2 the average accuracy and execution time for 5 fold cross-validation depending on the number of features and classification algorithm are shown. It is evident that both algorithms are able to cope with redundant features fairly successfully, albeit the Extremely Randomized Trees are offering somewhat better accuracy. Moreover, the Extremely Randomized Trees are always performing faster and this becomes important as the number of features and instances increase.

Next we wanted to investigate the variance and stability of estimated feature scores by both algorithms, so we have repeated the experiments twice with the same feature sets. Table II lists some statistics based on the feature scores listed in Table II. Namely, Table II shows the feature scores of the 24 most common features that were used in the previous

This work was partially financed by the Faculty of Computer Science and Engineering at the Ss.Cyril and Methodius University, Skopje, Macedonia

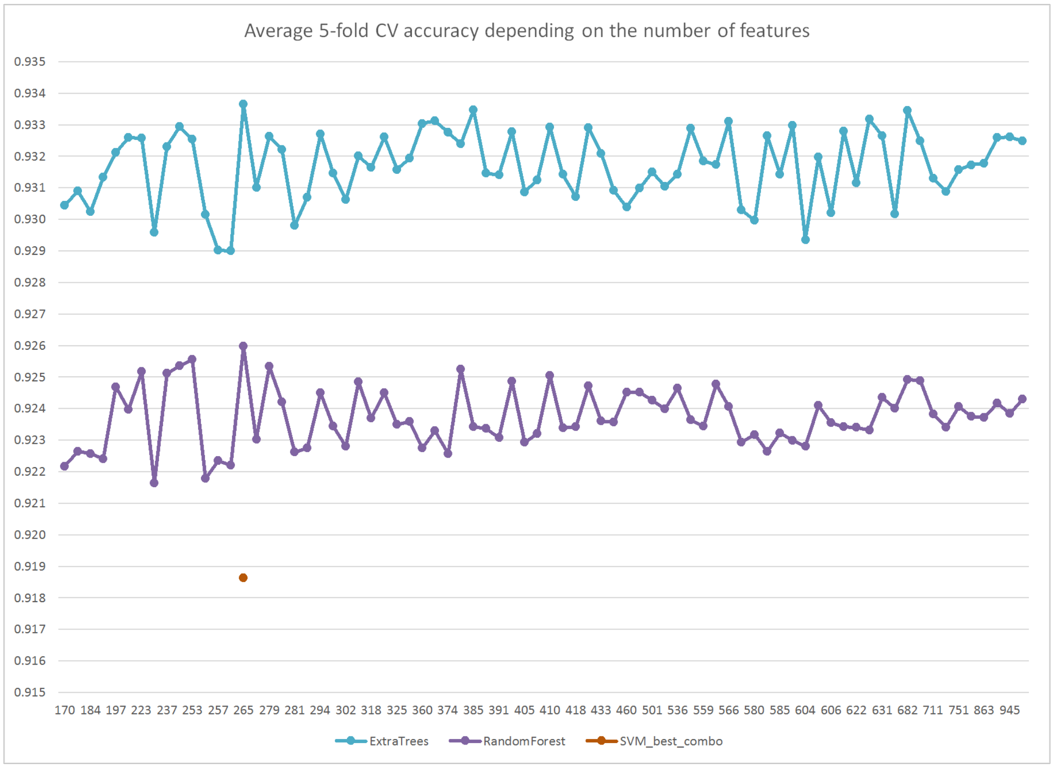

Fig. 1. Average 5-fold CV accuracy per personal model depending on the number of features and classification algorithm.

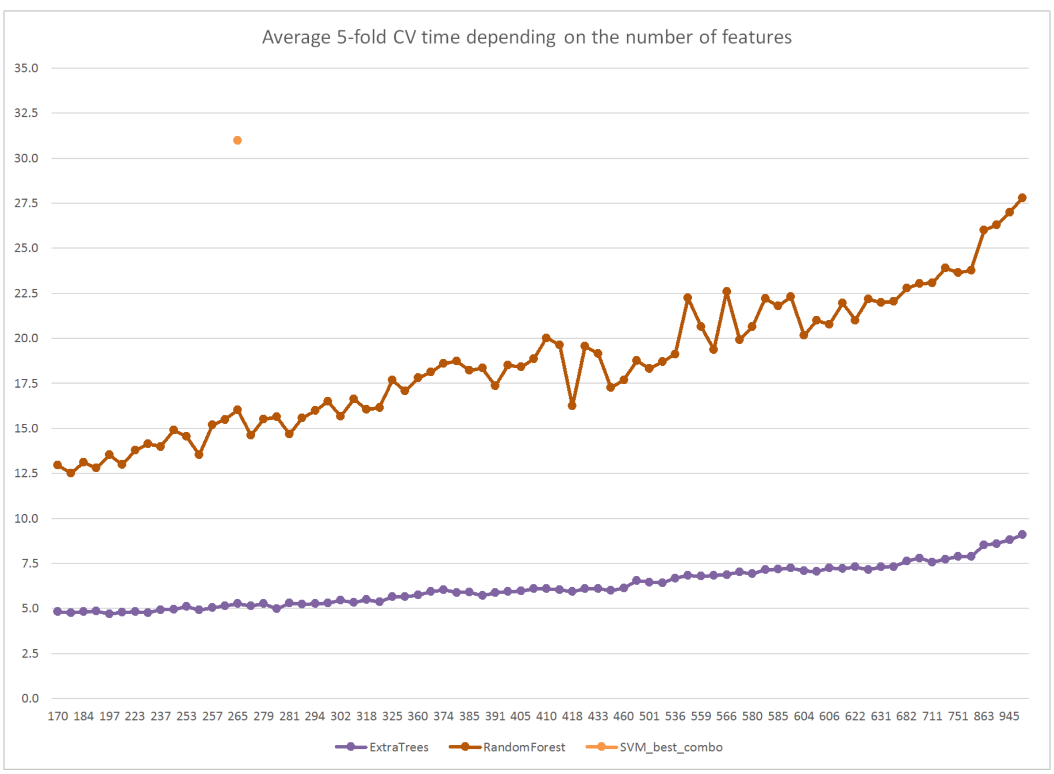

Fig. 2. Average 5-fold CV time in seconds per personal models depending on the number of features and classification algorithm.

evaluation. We performed 2-tailed T-Test of the feature scores obtained by both algorithms, as well as comparing the scores obtained by the 2 classifiers, as they are implemented in [10]. It is evident that the estimated feature scores are very stable due to the large number of trees. Then we calculated the standard deviation from the 2 scores of each feature for each classifier separately (columns ERT_std and RF_std in Table II). Calculating the mean of the standard deviations shows that both values are very low, albeit the Random Forest algorithm experiencing a lower value. Nonetheless, the statistical tests show that both algorithms can be safely used for feature importance estimation.

TABLE I. STATISTIC TESTS BASED ON DATA DESCRIBED IN TABLE II.

Metric	Result
ERT 2-tailed T-Test	0.99999813
RF 2-tailed T-Test	0.99999460
ERT vs RF T-Test	0.99999930
ERT mean of std per feature	0.00305626
RF mean of std per feature	0.00184345

V. CONCLUSION

We can conclude that both algorithms overcome redundant features very successfully, evident by the constant accuracy even if the number of features increases. It turns out that the Extremely Randomized Trees are resulting in somewhat better accuracy and are always performing faster, especially when the number of features and instances increase. Finally, based on the performed statistical tests we conclude that both algorithms provide similar feature importance estimates that are stable across multiple repetitions, regardless of the heavy randomization used intrinsically in the algorithms.

REFERENCES

[1] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32. doi: 10.1023/A:1010933404324. [Online]. Available: http://dx.doi.org/10.1023/A:1010933404324

[2] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006. doi: 10.1007/s10994-006-6226-1. [Online]. Available: http://dx.doi.org/10.1007/s10994-006-6226-1

[3] M. Meina, A. Janusz, K. Rykaczewski, D. Slezak, B. Celmer, and A. Krasuski, "Tagging firefighter activities at the emergency scene: Summary of aaia'15 data mining competition at knowledge pit," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*, Sept 2015. doi: 10.15439/2015F426 pp. 367–373.

[4] J. Lasek and M. Gagolewski, "The winning solution to the aaia'15 data mining competition: Tagging firefighter activities at a fire scene," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, 2015. doi: 10.15439/2015F418 pp. 375–380. [Online]. Available: http://dx.doi.org/10.15439/2015F418

[5] A. Zagorecki, "A versatile approach to classification of multivariate time series data," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, 2015. doi: 10.15439/2015F419 pp. 407–410. [Online]. Available: http://dx.doi.org/10.15439/2015F419

[6] E. Zdravevski, P. Lameski, R. Mingov, A. Kulakov, and D. Gjorgjevikj, "Robust histogram-based feature engineering of time series data," in *Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on*, ser. Annals of Computer Science and Information Systems, M. P. M. Ganzha, L. Maciaszek, Ed., vol. 5. IEEE, Sept 2015. doi: 10.15439/2015F420 pp. 381–388. [Online]. Available: http://dx.doi.org/10.15439/2015F420

[7] M. Boullé, "Tagging fireworkers activities from body sensors under distribution drift," in *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, and M. Paprzycki, Eds., vol. 5. IEEE, 2015. doi: 10.15439/2015F423 pp. 389–396.

[8] M. Lichman, "UCI machine learning repository," 2013. [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Activity+Recognition+from+Single+Chest-Mounted+Accelerometer

[9] P. Casale, O. Pujol, and P. Radeva, *Pattern Recognition and Image Analysis: 5th Iberian Conference, IbPRIA 2011, Las Palmas de Gran Canaria, Spain, June 8-10, 2011. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, ch. Human Activity Recognition from Accelerometer Data Using a Wearable Device, pp. 289–296. ISBN 978-3-642-21257-4. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-21257-4_36

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

TABLE II. Feature scores for the same dataset of 24 features with Random Forest (RF) and Extremely Randomized Trees (ERT) with 2 separate runs

FID	ERT_score1	ERT_score2	ERT_std	RF_score1	RF_score2	RF_std
1	0.0153947	0.0108758	0.0022595	0.0070526	0.0118251	0.0023863
2	0.0420794	0.0385484	0.0017655	0.0204782	0.0182116	0.0011333
3	0.0383765	0.0221985	0.0080890	0.0155013	0.0163048	0.0004018
4	0.0492405	0.0427291	0.0032557	0.0342975	0.0329070	0.0006953
5	0.0307088	0.0258354	0.0024367	0.0021842	0.0058967	0.0018563
6	0.0417594	0.0540148	0.0061277	0.0720881	0.0532936	0.0093973
7	0.1802803	0.2034607	0.0115902	0.1572349	0.1670721	0.0049186
8	0.0437639	0.0549879	0.0056120	0.0391848	0.0485794	0.0046973
9	0.0123589	0.0116197	0.0003696	0.0472597	0.0445050	0.0013774
10	0.0192220	0.0221837	0.0014809	0.0799508	0.0760616	0.0019446
11	0.0041322	0.0031622	0.0004850	0.0010633	0.0009209	0.0000712
12	0.0012066	0.0013322	0.0000628	0.0016406	0.0015975	0.0000216
13	0.0009440	0.0009317	0.0000061	0.0007464	0.0008139	0.0000338
14	0.0012444	0.0008918	0.0001763	0.0006912	0.0006698	0.0000107
15	0.0113268	0.0115136	0.0000934	0.0020207	0.0010306	0.0004951
16	0.0103095	0.0165596	0.0031251	0.0018599	0.0014173	0.0002213
17	0.0562460	0.0443660	0.0059400	0.0177128	0.0200320	0.0011596
18	0.0199697	0.0276082	0.0038193	0.0193860	0.0153704	0.0020078
19	0.0402050	0.0334358	0.0033846	0.0093499	0.0061775	0.0015862
20	0.0593671	0.0591298	0.0001186	0.0602517	0.0729347	0.0063415
21	0.0703908	0.0799185	0.0047639	0.1031800	0.1035752	0.0001976
22	0.0830415	0.0734411	0.0048002	0.1317565	0.1291026	0.0013269
23	0.1626861	0.1555547	0.0035657	0.1703122	0.1666465	0.0018329
24	0.0057459	0.0057007	0.0000226	0.0047964	0.0050541	0.0001289