Ensemble of Classifiers based on Decision Trees and Random Vector Functional Link Neural Network

Abstract

Ensemble of classifiers, also known as multiple classifier system, is a widely researched and frequently applied approach in machine learning. A multitude of studies corroborate that the combination of many unstable classifiers into one aggregated classifier leads to an improved performance compared to a single instance of such unstable classifier. Decision trees and neural networks are popular examples of such unstable classifiers.

There are basically three fundamental reasons behind the success of ensemble methods: statistical, computational and representational. An ensemble of learning algorithm reduces the risk of choosing a wrong hypothesis by taking an average of the different hypotheses given by the constituent learning algorithms (statistical). An ensemble that employ local searches from various starting points may achieve a better approximation of the unknown function than a single learning algorithm (computational). By combining the hypothesis of individual learning instances, an ensemble expands the space of the representable functions (representational). In addition to these, bias-variance decomposition, margin-theory and strength-correlation all corroborate the improved generalization capability of ensemble methods compared to a single learning algorithm. Moreover, most ensemble methods naturally fit distributed or parallel computing environment. Thus, ensemble methods are widely popular with machine learning researchers and enthusiasts.

This thesis is largely motivated by the aforementioned advantages of ensemble methods. More specifically, we work on decision tree and neural network ensembles. Random Forest (RF), an ensemble of decision trees, is among the most widely used machine learning algorithms while the popularity of random vector functional link (RVFL) neural network is also escalating because of its impressive performance in several diverse domains and its faster training time. The first part of this thesis is based on random forest (including a hybrid model based on RF and RVFL) while the second part is based on random vector functional link neural network ensembles.

Decision trees in conventional random forest usually employ binary splits at each node. The binary splits result in very deep trees. We utilize RVFL to create muti-way splits based decision trees. The proposed method provides a rich insight into the data by grouping the confusing or hard to classify samples for each class and thus, provides an opportunity to employ fine-grained classification rule over the data. Also, by using multi-way splits, we obtain shallow trees that agree with Ockham's razor principle of building smaller trees. Similarly, we also present our work in oblique decision trees in the first part of this thesis. In oblique decision trees, an oblique hyperplane is employed instead of an axis-parallel hyperplane at each node to partition the data. Trees with such hyperplanes can better exploit the geometric structure to increase the accuracy of the trees and reduce the depth. The present realizations of oblique decision trees do not evaluate many promising oblique splits to

select the best. We propose a random forest of heterogeneous oblique decision trees that employ several linear classifiers at each non-leaf node on some top ranked partitions.

In the second part of this thesis, inspired by the success of representational learning, we propose multi-layer (deep) random vector functional link neural networks. Specifically, we propose two variants of multi-layer (deep) random vector functional link neural networks. The first variant is built with RVFL based autoencoders while the second variant is built hierarchically by stacking several hidden layers on top of each other. We then, propose an ensemble of such multi-layer neural networks.

Publications

Journal

1. **R. Katuwal**, P.N. Suganthan, and L. Zhang, "An ensemble of decision trees with random vector functional link networks for multi-class classification," Applied Soft Computing, 2017.
2. **R. Katuwal**, P.N. Suganthan, and L. Zhang, "Heterogeneous Oblique Random Forest," Pattern Recognition (submitted).
3. **R. Katuwal**, and P.N. Suganthan, "Multi-layer Random Vector Functional Link for Classification" Applied Soft Computing (submitted).

Conference

1. **R. Katuwal** and P.N. Suganthan, "An ensemble of kernel ridge regression for multi-class classification," Procedia Computer Science, vol. 108, pp. 375–383, 2017, International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland
2. W. X. Cheng, **R. Katuwal**, P. N. Suganthan and X. Qiu, "A heterogeneous ensemble of trees," 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, 2017, pp. 1-6.
3. **R. Katuwal** and P.N. Suganthan, "Enhancing multi-class classification of random forest using random vector functional neural network and oblique decision surfaces," in 2018 International Joint Conference on Neural Networks (IJCNN), July 2018, pp. 1–8.
4. **R. Katuwal** and P.N. Suganthan, "Dropout and DropConnect based Ensemble of Random Vector Functional Link Neural Network", 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bengaluru, India, 2018.
5. **R. Katuwal** and P.N. Suganthan, "Random Vector Functional Link based Deep Neural Network", in 2019 International Joint Conference on Neural Networks (IJCNN) (submitted).