COMP7703 Machine Learning
Article Review 2

Thanat Chokwijitkul 44522328

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems?. *J. Mach. Learn. Res*, 15(1), 3133-3181.

The paper, "Do we need hundreds of classifiers to solve real world classification problems?" by Fernández-Delgado et al. analyses the performance of 179 classifiers belonging to 17 families against the entire UCI database along with 4 real-world datasets, summing up to 121 datasets. The authors' primary objective was to identify the classifier with the potential to achieve the best performance for any dataset.

The authors applied four comparison methodologies, including the Friedman ranking (a nonparametric statistical test), the probability of achieving maximum accuracy (PAMA), the percentage of maximum accuracy (PMA) and the probability of achieving more than 95% baseline (P95).

The experiment yielded that the most efficient classifier was parallel random forest (parRF_t) implemented in R using the caret and RandomForest packages. parRF_t achieved Friedman ranking of 32.9 with average accuracy of 82.0% and maximum accuracy of 94.1%. It is followed by random forest (rf_t) and support vector machine (svm_C) with maximum accuracies of 93.6% and 92.3% respectively. Fernández-Delgado et al. also concluded that random forests (RF) and support vector machines (SVM) are the best families since the resulted top 20 classifiers included 6 RFs and 5 SVMs.

Fernández-Delgado et al.'s paper is an intriguing empirical research that evaluated the performance of a reasonably large number of classifiers on a massive dataset. It is possible that the paper was written for machine learning researchers and practitioners who are interested in the field of classification, and may be used as a resource for the development of classification algorithms and future study in the field. The research involved a huge volume of work as the formats and patterns of the UCI datasets used at the time were not all the same. Experimenting with a classifier with a single dataset is uncomplicated, but trying to use it as an adapter for such massive amount of different datasets without the concept of transfer learning (applying knowledge gained from one problem to solve other different but related problems) in mind may be quite unattractive.

Some may regard that there is no such thing as one best classifier, only classifiers with better performance on particular tasks. Consequently, using a combination of learning algorithms for better performance sounds more promising in practice. The objective of this approach is not to discover the best classifier but to find an ensemble of learning algorithms which works well with a particular set of data.

Although Fernández-Delgado et al.'s conclusion is not utterly surprising, the intention of this research is more than just trying to identify one best classifier. It is considered beneficial to have a study on the comparisons of such significant amount of classifiers with quantitative data as the supporting evidence because the assumption about the non-existence of one best classifier is usually based on the "no free lunch" theorem (on average, all algorithms are equivalent against the set of all problems), without prior knowledge or evidence. Another key lesson learned from this study is that practical problems confronted among machine learning researchers and practitioners are usually a subset of all possible problems and no matter how large the set of the possible learning algorithms is, it is still manageable.

Even though there was an indication of bias against the real-world datasets, Fernández-Delgado et al.'s research successfully identified the globally best classifier for any dataset with insightful supporting evidence. As machine learning is now a mainstream presence and no longer some esoteric practice, the data produced from this research can serve as a considerably informative resource for further study in the field of classification.