

Abstract

The process of selecting an algorithm for use on a dataset is typically done by taking a pool of algorithms, optimizing the hyperparameters, and selecting a final algorithm based on its performance over an entire dataset. While the selected algorithm, when compared to its peers, has performed strongest over the entire dataset it typically has not performed strongest on every instance within the dataset. This dissertation found on the Kaggle Loan Dataset, a 58% reduction in mean absolute error (MAE) was possible using an ensemble of 8 algorithms, by selecting the best performing algorithm on a per-instance basis. This gives insight into the potential performance increase available for an optimal per-instance approach.

Per-instance algorithm selection often outperforms single algorithm approaches in many domains and has proven a useful avenue to improve performance. While per-instance regression models trained on instances and algorithm characteristics have proved useful, to date, they have been found to perform inconsistently across datasets leaving room for new per instance approaches.

In this dissertation, Siamese Algorithm Selection (SAS) is proposed as a new method of per-instance algorithm selection, utilizing a Siamese Neural Network (SNN) to learn Algorithm Performance Personas (APP), which are neighbourhoods of instances that map to similar performances. Trained on instance-performance pairs, Siamese Algorithm Selection uses the representations learned by the SNN in conjunction with k-nearest neighbours clustering algorithm for per instance algorithm selection.

It was found that the proposed method works and successfully outperformed the best single algorithm, reducing MAE by 15%. The selection accuracy, the number of times the best performing algorithm was selected on a per instance basis, showed that Siamese Algorithm Selection had a 8.5% higher selection accuracy compared to the best single algorithm.