Abstract – Multi-objective optimization to train classifiers on feature subspaces

Joanna Klikowska

May 2024

Abstract

The thesis focused on the use of multi-objective optimization for employing feature selection to classify mainly imbalanced data. Feature selection determines a feature subspace for each model, and this mechanism ensures the diversity of models in the ensemble. Research in this area has shown that optimizing for feature selection gives good results, and the proposed methods sometimes outperform state-of-the-art methods. An additional advantage of the proposed algorithms using multi-objective optimization is the ability to select the most appropriate solution, which classical methods do not offer. The following research objectives confirmed the research hypothesis formulated at the beginning.

Using multi-objective optimization to train classifiers on feature subspaces produces models with no worse prediction quality than state-of-the-art methods and allows choosing a solution tailored to a user's needs.

Several objectives were formulated to prove the hypothesis.

- **Development of feature selection methods based on Multi-Objective Optimization for constructing single classifiers.**

  This objective was met by proposing methods using optimization to perform feature selection. The methods use optimization algorithms such as *GA* in the single-criteria version and *NSGA-II* in the multi-objective version. The advantage of using optimization is taking into account not only the quality of the built classifier but also the cost of features, which is particularly important in cost-sensitive learning. Simultaneous optimization of two criteria, maximizing performance and minimizing cost, in the case of multi-objective optimization, gives a set of solutions from which the user can choose the most suited to his needs. The proposed methods achieved comparable quality to classical methods, but the latter do not allow the possibility of choosing from a solution set and returning only one solution.

13.05.2024

Joanna Klikowska

- **Development of a multi-objective method for training classifier ensembles using learning Support Vector Machines base classifiers on subspaces of the feature space.**

  This goal was achieved by proposing the *SEMOOS* method. *SEMOOS* is an ensemble consisting of single *SVM* classifiers using multi-objective optimization and the *NSGA-II* algorithm to search the feature space and find two parameters of *SVM* classifiers. *NSGA-II* returns a set of such solutions, and each of them is used to train a model, which is then added to the pool, creating an ensemble classifier. Using the proposed center-based bootstrapping and model pruning in the method is optional. The method was tested on many imbalanced datasets and achieved satisfactory results compared to reference methods.

- **Development of a method for training classifier ensemble using learning decision tree base classifiers on subspaces of the feature space and aggregated criteria.**

  The objective was completed by presenting the *DE-Forest* method using the *Differential Evolution* optimization algorithm to find the best feature vector for the entire ensemble relative to various aggregated metrics. Such a vector is appropriately prepared, and based on it, individual decision tree models can be trained to create an ensemble. The proposed method often outperforms state-of-the-art methods.

- **Development of a multi-objective method for training classifier ensemble using learning decision tree base classifiers on subspaces of the feature space to form the non-random forest.**

  The goal was achieved by proposing the *MOOforest* method, an ensemble built from individual decision trees using the *MOEA/D* multi-objective optimization algorithm. *MOEA/D* searches the feature space for the entire ensemble based on two criteria simultaneously: *Precision* and *Recall*. Thanks to this, it returns a set of solutions from which one solution is selected using the *PROMETHEE II* function. The models are trained based on this solution that makes up the final ensemble. The proposed method, in many cases, outperforms the reference methods.

- **Development of a multi-objective method for training classifier ensemble using learning base classifiers on subspaces of the feature space and local optimization.**

  The goal was achieved by proposing the *MOLO* method. It is a novel method using multi-objective local optimization to build a diverse ensemble. Each base *DT* model is trained on one bootstrapped subset. The optimization searches through possible solutions in each step and adds one model to the ensemble. The restrictions prevent the search from spreading to a considerable extent, and thanks to it, the algorithm selects the paths that provide the best results at a given moment. The *MOLO* method has been tested for two sets of dual criteria and can also handle the three criteria case. Finally, *MOLO* returns several ensembles from which the user can choose the best one for his needs. Extensive tests for imbalanced data comparing the proposed and reference methods showed *MOLO* advantage.

- **Experimental evaluation of the obtained methods.**

  This objective was achieved by comparing all proposed methods: *SE-MOOS*, *DE-Forest*, *MOOforest*, and *MOLO*. Statistical tests and detailed results for each dataset showed each method's characteristics and competence areas.