IDEA GROUP PUBLISHING

701 E Chocolate Avenue, Hershey PA17033, USA
Tel: 717/533-8845; Fax: 717/533-8661; URL: www.idea-group.com

ITB7253

Chapter IV

Heuristic Search-Based Stacking of Classifiers

Agapito Ledezma, Ricardo Aler and Daniel Borrajo
Universidad Carlos III de Madrid

Currently, the combination of several classifiers is one of the most active fields within inductive learning. Examples of such techniques are boosting, bagging and stacking. From these three techniques, stacking is perhaps the least used one. One of the main reasons for this relates to the difficulty to define and parameterize its components: selecting which combination of base classifiers to use and which classifiers to use as the meta-classifier. The approach we present in this chapter poses this problem as an optimization task and then uses optimization techniques based on heuristic search to solve it. In particular, we apply genetic algorithms to automatically obtain the ideal combination of learning methods for the stacking system.

INTRODUCTION

One of the most active and promising fields in inductive machine learning is the ensemble of classifiers approach. An ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way to classify new examples (Dietterich, 1997). The purpose of combining classifiers consists of improving the accuracy of a single classifier. Experimental results show that this is usually achieved.

There are several ways to construct such ensembles, but currently the most frequently used ones are bagging (Breiman, 1996), boosting (Freund & Schapire, 1995) and, less widely used, stacking (Wolpert, 1992). Bagging constructs a set of classifiers by subsampling the training examples to generate different hypotheses.

After the different hypotheses are generated, they are combined by a voting mechanism. Boosting also uses the voting system to combine the classifiers. But, instead of subsampling the training examples, it generates the hypotheses sequentially. In each repetition, a new classifier is generated whose focus are those instances that were handled incorrectly by the previous classifier. This is achieved by giving a weight to each instance in the training examples and adjusting these weights according to their importance after every iteration. Both, bagging and boosting use classifiers generated by the same base-learning algorithm and obtained from the same data. Finally, stacking can combine classifiers obtained from different learning algorithms using a high level classifier–the metaclassifier—to combine the lower level models. This is based on the fact that different classifiers are obtained from the same data and different learning algorithms use different biases to search the hypothesis space. This approach expects that the metaclassifier will be able to learn how to decide between the predictions provided by the base classifiers to improve their accuracy, much in the same way as a committee of experts.

One problem associated with stacked generalization is identifying which learning algorithm should be used to obtain the metaclassifier, and which ones should be the base classifiers. The approach we present in this chapter poses this problem as an optimization task, and then uses optimization techniques based on heuristic search to solve it. In particular, we apply genetic algorithms (Holland, 1975) to automatically obtain the ideal combination of learning methods for the stacking system.

BACKGROUND

The purpose of this section is to give enough background to understand the rest of the paper. Here, we will explain concepts related to ensembles of classifiers, bagging, boosting, stacking, and genetic algorithms.

Ensemble of Classifiers

The combination of multiple classifiers to improve the accuracy of a single classifier has had good results over several datasets that appear in recent papers about ensembles of classifiers (Bauer & Kohavi, 1999; Breiman, 1996; Freund & Schapire, 1996; Quinlan, 1996). According to Dietterich (1997), an ensemble of classifiers is a set of classifiers whose individual decisions are combined in some way to classify new examples. There are many ways to construct an ensemble of classifiers. Bauer and Kohavi (1999) have made a comparison of algorithms based on voting systems. Dietterich (2000) carried out a survey of the main methods to construct an ensemble of classifiers. One way to construct an ensemble of classifiers is based on subsampling the training set to generate a different set of hypotheses and then combine them. This is called bagging (Breiman, 1996). The second way is to create classifiers sequentially, giving more importance to examples that were

Related Content