Meta-Learning

Christophe Giraud-Carrier
Brigham Young University, USA

Pavel Brazdil
University of Porto, Portugal

Carlos Soares
University of Porto, Portugal

Ricardo Vilalta
University of Houston, USA

INTRODUCTION

The application of Machine Learning (ML) and Data Mining (DM) tools to classification and regression tasks has become a standard, not only in research but also in administrative agencies, commerce and industry (e.g., finance, medicine, engineering). Unfortunately, due in part to the number of available techniques and the overall complexity of the process, users facing a new data mining task must generally either resort to trial-and-error or consultation of experts. Clearly, neither solution is completely satisfactory for the non-expert end-users who wish to access the technology more directly and cost-effectively.

What is needed is an informed search process to reduce the amount of experimentation with different techniques while avoiding the pitfalls of local optima that may result from low quality models. Informed search requires meta-knowledge, that is, knowledge about the performance of those techniques. Meta-learning provides a robust, automatic mechanism for building such meta-knowledge. One of the underlying goals of meta-learning is to understand the interaction between the mechanism of learning and the concrete contexts in which that mechanism is applicable. Meta-learning differs from base-level learning in the scope of adaptation. Whereas learning at the base-level focuses on accumulating experience on a specific learning task (e.g., credit rating, medical diagnosis, mine-rock discrimination, fraud detection, etc.), learning at the meta-level is concerned with accumulating experience on the performance of multiple applications of a learning system.

The meta-knowledge induced by meta-learning provides the means to inform decisions about the precise conditions under which a given algorithm, or sequence of algorithms, is better than others for a given task. While Data Mining software packages (e.g., SAS Enterprise Miner, SPSS Clementine, Insightful Miner, PolyAnalyst, KnowledgeStudio, Weka, Yale, Xelopes) provide user-friendly access to rich collections of algorithms, they generally offer no real decision support to non-expert end-users. Similarly, tools with emphasis on advanced visualization help users understand the data (e.g., to select adequate transformations) and the models (e.g., to tweak parameters, compare results, and focus on specific parts of the model), but treat algorithm selection as a post-processing activity driven by the users rather than the system. Data mining practitioners need systems that guide them by producing explicit advice automatically. This chapter shows how meta-learning can be leveraged to provide such advice in the context of algorithm selection.

BACKGROUND

STABB is an early precursor of meta-learning systems in the sense that it was the first to show that a learner's bias can be adjusted dynamically (Utgoff, 1986). VBMS may be viewed as the first simple meta-learning system (Rendell et al., 1989). It learns to choose one of three symbolic learning algorithms as a function of the number of training instances and the number of features. The StatLog project extended VBMS significantly by considering a larger number of dataset characteristics, together with a broad class of candidate models and algorithms for selection (Brazdil & Henery, 1994). The aim was to characterize the space in which a given algorithm achieves positive generalization performance.

The MLT project focused on the practice of machine learning and produced a toolbox consisting of a number of learning algorithms, datasets, standards and know-how (Kodratoff et al., 1992; Craw et al., 1992). Considerable insight into many important machine learning issues was gained during the project, much of which was translated into meta-rules that formed the basis of a kind of user-guidance expert system called Consultant-2.

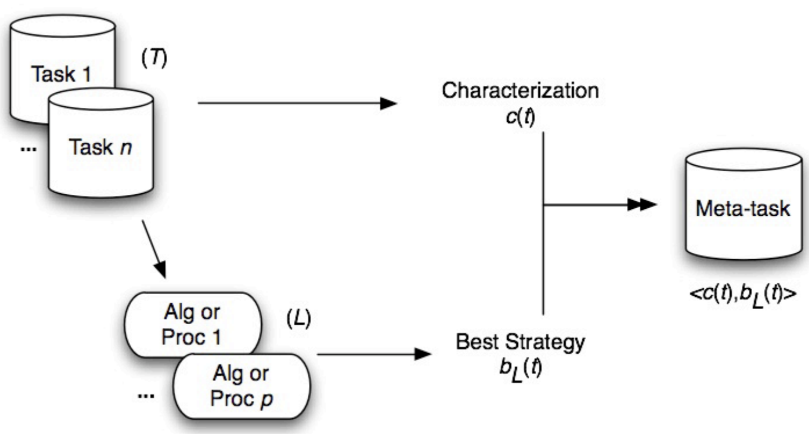
Born out of practical challenges faced by researchers at Daimler Benz AG (now), CITRUS is perhaps the first implemented system to offer user guidance for the complete data mining process, rather than for a single phase of the process (Engels, 1996; Wirth et al., 1997). Algorithm selection takes place in two stages, consisting of: 1) mapping tasks to classes of algorithms, and 2) selecting an algorithm from the selected class. The mapping stage is driven by decomposition and guided by high-level pre/post-conditions (e.g., interpretability). The selection stage consists of using data characteristics (inspired by the Statlog project) together with a process of elimination (called "strike-through"), where algorithms that do not work for the task at hand are successively eliminated until the system finds one applicable algorithm. Although there is no meta-learning in the traditional sense in CITRUS, there is still automatic guidance beyond the user's own input.

Finally, theoretical results, such as the NFL theorems and their consequences have helped in identifying limitations and opportunities for meta-learning (Schaffer, 1994; Wolpert & Macready, 1995; Wolpert, 2001). Additionally, extensive empirical studies have confirmed the theory, and provided additional insight into learning that may serve both as a source of direct meta-knowledge and as input to meta-learning (Aha, 1992; Holte, 1993; Lim et al., 2000).[1]

MAIN FOCUS

Meta-learning, in the context of model selection, consists of applying learning mechanisms to the problem of mapping learning tasks to algorithms. Let L be a set of learning algorithms and T be a set of learning tasks such that for each t in T, $b_L(t)$ represents the algorithm in L that performs best on t for some user-defined performance criterion (e.g., predictive accuracy, execution time).[2] Since learning tasks may be unwieldy to handle

Figure 1. Meta-dataset construction



Related Content