Machine Learning, 50, 219–222, 2003

Introduction

Multistrategy Learning is concerned with developing learning methods and systems that integrate different inferential or computational strategies in solving a given *learning task*. In general, a learning task is defined as a composition of three components: the type of knowledge to be learned, the learner's prior knowledge, and the input information available to the learner (Michalski, 1994). Due to the complementary nature of different learning strategies, multistrategy learning systems have a potential for a much wider range of applications than monostrategy systems. Research in this field has also a close kinship to cognitive models of learning because human learning is multistrategy.

The field of Multistrategy Learning is now more than decade old, with its roots going back to the First International Workshop on Multistrategy Learning, organized by Ryszard Michalski and G. Tecuci in Harpers Ferry, WV in November 1991. Since that time, four more workshops have been organized, each one bringing together researchers from many countries and addressing a wide variety of issues (e.g., Esposito, Michalski, & Saitta, 1998). A motivating idea behind these efforts is that the development of full-fledged multistrategy learning systems could lead to a plethora of new applications of machine learning, and bring about a breakthrough in building intelligent systems. Efforts in this direction have demonstrated, however, that multistrategy learning represents an exciting but significant scientific challenge, and its major advancement will require sustained long-term efforts, and an incorporation of the most advanced ideas from such areas as knowledge representation, machine learning, automated inference, and others. So far, a relatively small number of researchers have embarked on research in this direction, and in most cases their attention has concentrated on classification learning, a relatively well-understood, but pervasive topic in practical applications of machine learning.

This special issue, the third issue on Multistrategy Learning, includes several papers constituting updated and extended versions of the papers selected from among those presented at the Fifth International Workshop on Multistrategy Learning, organized in Guimarães, Portugal in June 2000 by the editors of this issue (Michalski & Brazdil, 2000).

To give the reader a sense of organization of this issue, let us distinguish between two directions in multistrategy learning research, and show how the papers in this issue fit this framework. The papers in this issue can be classified into two major categories:

1. Combining different classification algorithms. Since any given learning goal (e.g., a classification learning) can be usually accomplished in several alternative ways, one multistrategy approach is to combine them in some simple way in order to achieve a better performance. Ensemble learning methods, such as bagging, voting, boosting, stacking, cascading, etc., can be viewed as different forms of this line of work. A combination of methods can be done in different ways. Some solutions opt for a pre-wired solution

(e.g., voting using uniform weights), and others opt for a more dynamic solution, e.g., by employing meta-learning.

Three papers in this volume belong to this broad category. The first one describes a method for combining classifiers using meta-decision trees (paper by Todorovski & Džeroski). The method can be seen as an extension of stacking or cascading that employs meta-learning. The second paper (by Brazdil, Soares, & Costa) describes a method for ordering (ranking) classification algorithms, and also relies on meta-learning. The top N classifiers recommended can be viewed as a kind of ensemble method. The third paper uses a multistrategy solution to solve the problem of matching schemas of different data sources (by Doan, Domingos, & Halevy). Predictions of different base learners are combined using a kind of stacking approach.

2. Multistage processing. In some applications, solving a given task requires several techniques employed sequentially. A typical example is the execution of data pre-processing steps before generating a model. This category of multistrategy learning methods resembles the approach to knowledge discovery in data (KDD), which typically involves several stages. Two papers fall into this broad category. The first one concerns clustering data before applying linear regression (by Torgo & Costa), and the second concerns determining patterns in graphs (by Inokuchi, Washio, & Motoda).

To give the reader a further guidance to this issue, in the following we summarize main ideas behind the individual papers included.

The first paper, "*Combining Classifiers with Meta Decision Trees*," by Todorovski and Džeroski, describes a technique extending the previous work on combining classifiers by using stacking and cascading. Meta Decision Trees (MDTs) are induced from data, and then used to select the most appropriate base-level classifier for a given example subset. Each leaf of the MDT characterizes a part of the dataset that has been identified as falling into the "area of expertise" of a given base-level classifier. The "area of expertise" is identified on the basis of meta-attributes characterizing classes of probability distributions of base-level classifiers. In experiments presented, MDTs outperformed several other methods, in particular, ordinary decision trees, typical voting schemes, boosting and bagging of decision trees, and were somewhat better than the SCANN method and the Select-Best method. MDTs can thus be viewed as representing the state-of–the-art in ensemble learning.

In "*Ranking Learning Algorithms*," Brazdil, Soares and Costa describe a method for supporting a selection of candidate algorithms. Although the paper is oriented towards classification algorithms, the method described could provide assistance in the selection of alternative strategies in a multistrategy system. The method uses a k-Nearest Neighbor algorithm to identify datasets that are most similar to the one at hand. The distance between datasets is assessed using a small set of characteristics that represent properties of algorithms correlated with the relative performance of the algorithms. The performance of algorithms on those datasets is used to generate a recommendation to the user in the form of a ranking. It is shown that if the top N (say top 3) recommendations are followed, the combination of the N algorithms represents a highly competitive option. As both accuracy and time are considered, the methodology will tend to suggest a solution that is appropriate for the task at hand. The authors demonstrated that the presented meta-learning method can lead to significantly better rankings than a baseline ranking method. The method presented provides a good basis for further applications in the area of ranking (and selection) of alternative solution strategies in a multistrategy system.

The problem of integrating data from multiple data sources, such as Internet or business enterprises, has recently drawn much attention in database and AI communities. A major bottleneck in building such systems is manual construction of semantic mappings between different data schemas that can be encountered. In "*Learning to Match the Schemas of Data Sources*," Doan, Domingos and Halevy show how this process can be speeded up, by employing multistrategy learning. Examples of mappings, prepared by the user, are used as training data for determining generalized mappings. For different tasks, different learning methods are employed, including Naïve Bayes Learner, hierarchical classifier (useful when dealing XML data), as well as some specific techniques, e.g., for recognizing names of regions (counties). Predictions of different learned classifiers are combined by a meta-learner using a stacking strategy. The meta-learner can be trained to adjust the weights assigned to different learners. The system has been evaluated on several real world domains. The results show a good overall accuracy for unseen data. This work contributes to the fast growing area of data extraction and integration.

In "*Clustered Partial Linear Regression*," Torgo and Costa describe a supervised multistrategy learning approach to performing regression. A clustering method is used to form subsets of the training data before the actual regression modeling takes place. This phase creates several partitions of the data, each containing cases relatively similar to each other. It is shown that this approach can lead to more accurate results. Among distinguishing features of the method is the way new cases are classified. For each case, all clusters are examined and for each one the appropriate membership probability is calculated. These probabilities are used as weights in the process of averaging different predictions. This method has some resemblance to bagging, but is obviously different. An early idea of clustering data before equation fitting has been implemented by Falkenheiner (1986), but the method proposed here is quite different.

The paper *Complete Mining of Frequent Patterns from Graph,* by Inokuchi, Washio and Motoda, addresses a problem akin to mining association rules. The methodology presented differs from more traditional ones in that it can deal with more complex data structures. Instead of using tables, the method uses graph-structured data to discover characteristic patterns. The authors propose novel principles and algorithms for deriving characteristic patterns frequently appearing in graph-structured data. Their methodology incorporates a multiplicity of techniques, such as canonical graph representation and efficient processing of matrices encoding adjacency information. Due to these techniques, many redundant paths in the search for graph-structured patterns can be eliminated. As a consequence, the system is able to perform very efficiently some difficult real-world tasks, such as an analysis of Web browsing patterns, analysis of chemical carcinogenesis, and an identification of graph-structured patterns. These results may be interesting to the Inductive Logic Programming (ILP) community, as they may be helpful for speeding-up existing ILP systems.

The papers included in this special issue represent but a very small sample of recent research on multistrategy learning; nevertheless, they give the reader some sense of the current research efforts in this area. It should be noted, however, that the field of multistrategy learning is at very early stage of development, and offers researchers a wide range of unexplored research topics.

In conclusion, we hope that this issue will bring this area to the attention of a wider range of researchers, and will stimulate further activities in this enthralling area that addresses some of the most challenging and fundamental problems in developing intelligent systems.

References

Esposito, F., Michalski, R. S., & Saitta, L. (Eds.). (1998). *Proceedings of the Fourth International Workshop on Multistrategy Learning* (MSL-98), Desenzano del Garda (Brescia), Italy.

Falkenheimer, B., & Michalski, R. S. (1986). Integrating qualitative and quantitative discovery: The ABACUS System. *Machine Learning, 1:4*, 367–401.

Michalski, R. S. (1994). Inferential theory of learning: Developing foundations for multistrategy learning. In *Machine learning: A multistrategy approach* (Vol. IV). San Francisco, CA: Morgan Kaufman.

Michalski, R. S., & Brazdil, P. B. (Eds.). (2000). *Proceedings of the Fifth International Workshop on Multistrategy Learning* (*MSL-2000*), Guimarães, Portugal, http://www.liacc.up.pt/ML/Events/MSLOO/.

Ryszard S. Michalski
Pavel Brazdil