Chapter 33

META-LEARNING

Concepts and Techniques

Ricardo Vilalta
University of Houston

Christophe Giraud-Carrier
Brigham Young University

Pavel Brazdil
University of Porto

Abstract   The field of meta-learning has as one of its primary goals the understanding of
the interaction between the mechanism of learning and the concrete contexts in
which that mechanism is applicable. The field has seen a continuous growth in
the past years with interesting new developments in the construction of practical
model-selection assistants, task-adaptive learners, and a solid conceptual frame-
work. In this chapter we give an overview of different techniques necessary to
build meta-learning systems. We begin by describing an idealized meta-learning
architecture comprising a variety of relevant component techniques. We then
look at how each technique has been studied and implemented by previous re-
search. In addition we show how meta-learning has already been identified as an
important component in real-world applications.

Keywords: Meta-learning

1.   Introduction

We are used to thinking of a learning system as a rational agent capable
of adapting to a specific environment by exploiting knowledge gained through
experience; encountering multiple and diverse scenarios sharpens the ability of
the learning system to predict the effect produced from selecting a particular

course of action. In this case, learning is made manifest because the quality of the predictions normally improves with an increasing number of scenarios or examples. Nevertheless, if the predictive mechanism were to start afresh on different tasks, the learning system would find itself at a considerable disadvantage; learning systems capable of modifying their own predictive mechanism would soon outperform our base learner by being able to change their learning strategy according to the characteristics of the task under analysis.

Meta-learning differs from *base-learning* in the scope of the level of adaptation; whereas learning at the base-level is based 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 based on accumulating experience on the performance of multiple applications of a learning system. If a base-learner fails to perform efficiently, one would expect the learning mechanism itself to adapt in case the same task is presented again. Meta-learning is then important in understanding the interaction between the mechanism of learning and the concrete contexts in which that mechanism is applicable. Briefly stated, the field of meta-learning is focused on the relation between tasks or domains and learning strategies. In that sense, by learning or explaining what causes a learning system to be successful or not on a particular task or domain, we go beyond the goal of producing more accurate learners to the additional goal of understanding the conditions (e.g., types of example distributions) under which a learning strategy is most appropriate.

From a practical stance, meta-learning can solve important problems in the application of machine learning and Data Mining tools, particularly in the area of classification and regression. First, the successful use of these tools outside the boundaries of research (e.g., industry, commerce, government) is conditioned on the appropriate selection of a suitable predictive model (or combinations of models) according to the domain of application. Without any kind of assistance, model selection and combination can turn into stumbling blocks to the end-user who wishes to access the technology more directly and cost-effectively. End-users often lack not only the expertise necessary to select a suitable model, but also the availability of many models to proceed on a trial-and-error basis (e.g., by measuring accuracy via some re-sampling technique such as n-fold cross-validation). A solution to this problem is attainable through the construction of meta-learning systems. These systems can provide automatic and systematic user guidance by mapping a particular task to a suitable model (or combination of models).

Second, a problem commonly observed in the practical use of ML and DM tools is how to profit from the repetitive use of a predictive model over similar tasks. The successful application of models in real-world scenarios requires a continuous adaptation to new needs. Rather than starting afresh on new tasks, we expect the learning mechanism itself to re-learn, taking into account pre-

vious experience (Thrun, 1998; Pratt *et al.*, 1991; Caruana, 1997; Vilalta and Drissi, 2002). Again, meta-learning systems can help control the process of exploiting cumulative expertise by searching for patterns across tasks.

Our goal in this chapter is to give an overview of different techniques necessary to build meta-learning systems. To impose some structure, we begin by describing an idealized meta-learning architecture comprising a variety of relevant component techniques. We then look at how each technique has been studied and implemented by previous research. We hope that by proceeding in this way the reader can not only learn from past work, but in addition gain some insight on how to construct meta-learning systems.

We also hope to show how recent advances in meta-learning are increasingly filling the gaps in the construction of practical model-selection assistants and task-adaptive learners, as well as in the development of a solid conceptual framework (Baxter, 1998; Baxter, 2000; Giraud-Carrier *et al.*, 2004).

This chapter is organized as follows. In the next section we illustrate an idealized meta-learning architecture and detail on its constituent parts. In Section 3 we describe previous research in meta-learning and its relation to our architecture. Section 4 describes a meta-learning tool that has been instrumental as a decision support tool in real applications. Lastly, section 5 discusses future directions and provides our conclusions.

2. A Meta-Learning Architecture

In this section we provide a general view of a software architecture that will be used as a reference to describe many of the principles and current techniques in meta-learning. Though not every technique in meta-learning fits into this architecture, such a general view helps us understand the challenges we need to overcome before we can turn the technology into a set of useful and practical tools.

2.1 Knowledge-Acquisition Mode

To begin, we propose a meta-learning system that divides into two modes of operation. During the first mode, also known as the *knowledge-acquisition mode*, the main goal is to learn about the learning process itself. Figure 33.1 illustrates this mode of operation. We assume the input to the system is made of more than one dataset of examples (e.g., more than one set of pairs of feature vectors and classes; Figure 33.1A). Upon arrival of each dataset, the meta-learning system invokes a component responsible for extracting dataset characteristics or meta-features (Figure 33.1B). The goal of this component is to gather information that transcends the particular domain of application. We look for information that can be used to generalize to other example distributions. Section 3.1 details current research pointing in this direction.

During the knowledge acquisition mode, the learning technique (Figure 33.1C) does not exploit knowledge across different datasets or tasks. Each dataset is considered independently of the rest; the output to the system is a learning strategy (e.g., a classifier or combination of classifiers, Figure 33.1D). Statistics derived from the output model or its performance (Figure 33.1E) may also serve as a form of characterizing the task under analysis (Sections 3.1 and 3.1.1).

Information derived from the meta-feature generator and the performance evaluation module can be combined into a *meta-knowledge base* (Figure 33.1F). This knowledge base is the main result of the knowledge–acquisition phase; it reflects experience accumulated across different tasks. Meta-learning is tightly linked to the process of acquiring and exploiting meta-knowledge. One can even say that advances in the field of meta-learning hinge around one specific question: how can we acquire and exploit knowledge about learning systems (i.e., meta-knowledge) to understand and improve their performance? As we describe current research in meta-learning we will be pointing out to different forms of meta-knowledge.

2.2   Advisory Mode

The efficiency of the meta-learner increases as it accumulates meta-knowledge. We assume the lack of experience at the beginning of the learner's life compels the meta-learner to use one or more learning strategies without a clear preference for one of them; experimenting with many different strategies becomes time consuming. However, as more training sets have been examined, we expect the expertise of the meta-learner to dominate in deciding which learning strategy best suits the characteristics of the training set.

In the *advisory mode*, meta-knowledge acquired in the exploratory mode is used to configure the learning system in a manner that exploits the characteristics of the new data distribution. Meta-features extracted from the dataset (Figure 33.2B) are matched with the meta-knowledge base (Figure 33.2F) to produce a recommendation regarding the best available learning strategy. At this point we move away from the use of static base learners to the ability to do model selection or combining base learners (Figure 33.2C).

Two observations are worth considering at this point. First, the nature of the match between the set of meta-features and the meta-knowledge base can have several interpretations. The traditional view poses this problem as a learning problem itself where a meta-learner is invoked to output an approximating function mapping meta-features to learning strategies (e.g., learning model). This view is problematic as the meta-learner is now a learning system subject to improvement through meta-learning (Schmidhuber, 1995; Vilalta, 2001). Second, the matching process is not intended to modify our set of available
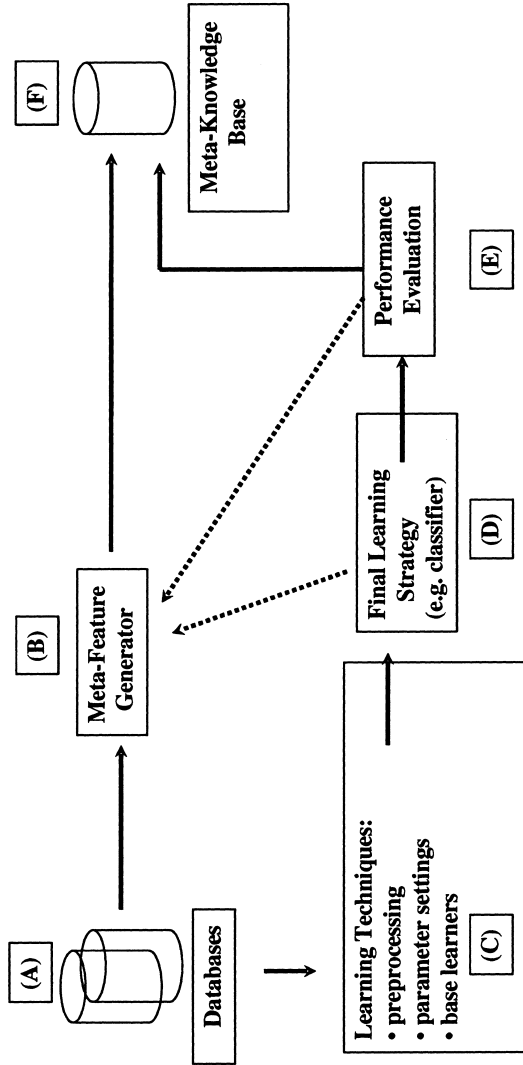
Figure 33.1. The Knowledge-Acquisition Mode

learning techniques, but simply enables us to select one or more strategies that
seem effective given the characteristics of the dataset under analysis.

The final classifier (or combination of classifiers; Figure 33.2D) is selected
based not only on its generalization performance over the current dataset, but
also on information derived from exploiting past experience. In this case, the
system has moved from using a single learning strategy to the ability of select-
ing one dynamically from among a variety of different strategies.
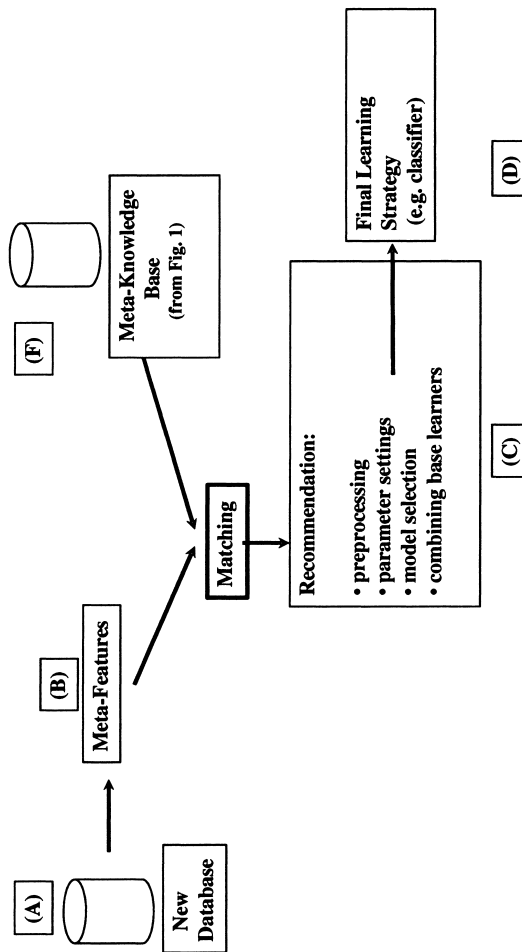

Figure 33.2.   The Advisory Mode

We will show how the constituent components conforming our two-mode meta-learning architecture can be studied and utilized through a variety of different methodologies:

1. The characterization of datasets can be performed under a variety of statistical, information-theoretic, and model-based approaches (Section 3.1).

2. Matching meta-features to predictive model(s) can be used for model selection or model ranking (Section 3.1.1).

3. Information collected from the performance of a set of learning algorithms at the base level can be combined through a meta-learner (Section 3.1.2).

4. Within the learning-to-learn paradigm, a continuous learner can extract knowledge across domains or tasks to accelerate the rate of learning convergence (Section 3.1.3).

5. The learning strategy can be modified in an attempt to shift this strategy dynamically (Section 3.2). A meta-learner in effect explores not only the space of hypotheses within a fixed family set, but the space of families of hypotheses.

3. Techniques in Meta-Learning

In this section we describe how previous research has tackled the implementation and application of various methodologies in meta-learning.

3.1 Dataset Characterization

First, a critical component of any meta-learning system is in charge of extracting relevant information about the task under analysis (Figure 33.1B). The central idea is that high-quality dataset characteristics or meta-features provide some information to differentiate the performance of a set of given learning strategies. We describe a representative set of techniques in this area.

3.1.1 Statistical and Information-Theoretic Characterization.
Much work in dataset characterization has concentrated on extracting statistical and information-theoretic parameters estimated from the training set (Aha, 1992; Michie *et al.*, 1994; Gama and Brazdil, 1995; Brazdil, 1998) (Engels and Theusinger, 1998; Sohn, 1999). Measures include number of classes, number of features, ratio of examples to features, degree of correlation between features and target concept, average class entropy and class-conditional entropy,

skewness, kurtosis, signal to noise ratio, etc. This work has produced a number of research projects with positive and tangible results (e.g., ESPRIT Statlog and METAL).

3.1.2   Model-Based Characterization.   In addition to statistical measures, a different form of dataset characterization exploits properties of the induced hypothesis as a form of representing the dataset itself. This has several advantages: 1) the dataset is summarized into a data structure that can embed the complexity and performance of the induced hypothesis (and thus is not limited to the example distribution); 2) the resulting representation can serve as a basis to explain the reasons behind the performance of the learning algorithm. As an example, one can build a decision tree from a dataset and collect properties of the tree (e.g., nodes per feature, maximum tree depth, shape, tree imbalance, etc.), as a means to characterize the dataset (Bensusan, 1998; Bensusan and Giraud-Carrier, 2000b; Hilario and Kalousis, 2000; Peng *et al.*, 1995).

3.1.3   Landmarking.   Another source of characterization falls within the concept of landmarking (Bensusan and Giraud-Carrier, 2000a; Pfahringer *et al.*, 2000). The idea is to exploit information obtained from the performance of a set of simple learners (i.e., learning systems with low capacity) that exhibit significant differences in their learning mechanism. The accuracy (or error rate) of these *landmarkers* is used to characterize a dataset. The goal is to identify areas in the input space where each of the simple learners can be regarded as an expert. This meta-knowledge can be subsequently exploited to produce more accurate learners.

Another idea related to landmarking is to exploit information obtained on simplified versions of the data (e.g. small samples). Accuracy results on these samples serve to characterise individual datasets and are referred to as *sampling landmarks*. This information is subsequently used to select a learning algorithm (Furnkranz, 1997; Soares *et al.*, 2001).

3.2   Mapping Datasets to Predictive Models

An important and practical use of meta-learning is the construction of an engine that maps an input space composed of datasets or applications to an output space composed of predictive models. Criteria such as accuracy, storage space, and running time can be used for performance assessment (Giraud-Carrier, 1998). Several approaches have been developed in this area.

3.2.1   Hand-Crafting Meta Rules.   First, using human expertise and empirical evidence, a number of meta-rules matching domain characteristics with learning techniques may be crafted manually (Brodley, 1993; Brodley,

1994). For example, in decision tree learning, a heuristic rule can be used to switch from univariate tests to linear tests if there is a need to construct non-orthogonal partitions over the input space. Crafting rules manually has the disadvantage of failing to identify many important rules. As a result most research has focused on learning these meta-rules automatically as explained next.

3.2.2   Learning at the Meta-Level.   The characterization of a dataset is a form of meta-knowledge (Figure 33.1F) that is commonly embedded in a meta-dataset as follows. After learning from several tasks, one can construct a meta-dataset where each element pair is made up of the characterization of a dataset (meta-feature vector) and a class label corresponding to the model with best performance on that dataset. A learning algorithm can then be applied to this well-defined learning task to induce a hypothesis mapping datasets to predictive models.

As in base-learning, the hand-crafting and the learning approach can be combined; in this case the hand-crafted rules can serve as background knowledge to the meta-learner.

3.2.3   Mapping Query Examples to Models.   Instead of mapping a task or dataset to a predictive model, a different approach consists of selecting a model for each individual query example. The idea is similar to the nearest-neighbour approach: select the model displaying best performance around the neighbourhood of the query example (Merz, 1995A; Merz, 1995B). Model selection is done according to best-accuracy performance using a re-sampling technique (e.g., cross-validation).

A variation to the approach above is to look at the neighbourhood of a query example in the space of meta-features. When a new training set arrives, the k-nearest neighbour instances (i.e., datasets) around the query example (i.e., query dataset) are gathered to select the model with best average performance (Keller *et al.*, 2000).

3.2.4   Ranking.   Rather than mapping a dataset to a single predictive model, one may also produce a ranking over a set of different models. One can argue that such rankings are more flexible and informative for users. In a practical scenario, users should not be limited to a single kind of advice; this is important if the suggested final model turns unsatisfactory. Rankings provide alternative solutions to users who may wish to incorporate their own expertise or any other criterion (e.g., financial constraints) on their decision-making process. Multiple approaches have been suggested attacking the problem of ranking predictive models (Gama and Brazdil, 1995; Nakhaeizadeh *et al.*,

2002; Berrer *et al.*, 2000; Brazdil and Soares, 2000; Keller *et al.*, 2000; Soares and Brazdil, 2000; Brazdil and Soares, 2003).

3.3   Learning from Base-Learners

Another approach to meta-learning consists of learning from base learners. The idea is to make explicit use of information collected from the performance a set of learning algorithms at the base level; such information is then incorporated into the meta-learning process.

3.3.1   Stacked Generalization.   Meta-knowledge (Figure 33.1F) can incorporate predictions of base learners, a process known as stacked generalization (Wolpert, 1997). The process works under a layered architecture as follows. Each of a set of base-classifiers is trained on a dataset; the original feature representation is then extended to include the predictions of these classifiers. Successive layers receive as input the predictions of the immediately preceding layer and the output is passed on to the next layer. A single classifier at the topmost level produces the final prediction. Most research in this area focuses on a two-layer architecture (Wolpert, 1997; Breiman, 1996; Chan and Stolfo, 1998; Ting, 1994).

Stacked generalization is considered a form of meta-learning because the transformation of the training set conveys information about the predictions of the base-learners (i.e., conveys meta-knowledge). Research in this area investigates what base-learners and meta-learners produce best empirical results (Chan and Stolfo, 1993; Chan and Stolfo, 1996; Gama and Brazdil, 2000); how to represent class predictions (class labels versus class-posterior probabilities; (Ting, 1994); what higher-level learners can be invoked (Gama and Brazdil, 2000; Dzeroski, 2002); and what are novel definitions of meta-features (Brodley, 1996; Ali and Pazzani, 1995).

3.3.2   Boosting.   A popular approach to combining base learners is called boosting (Freund and Schapire, 1995; Friedman, 1997; Hastie *et al.*, 2001). The basic idea is to generate a set of base learners by generating variants of the training set. Each variant is generated by sampling with replacement under a weighted distribution. This distribution is modified for every new variant by giving more attention to those examples incorrectly classified by the most recent hypothesis.

Boosting is considered a form of meta-learning because it takes into consideration the predictions of each hypothesis over the original training set to progressively improve the classification of those examples for which the last hypothesis failed.

3.3.3 Landmarking Meta-Learning. We mentioned before how land-marking can be used as a form of dataset characterization by exploiting the accuracy (or error rate) of a set of base (simple) learners called landmarkers. Meta-learning based on landmarking may be viewed as a form of learning from base learners; these base learners provide a new representation of the dataset that can be used in finding areas of learning expertise. Here we assume there is a second set of advanced learners (i.e., learning systems with high capac-ity), one of which must be selected for the current task under analysis. Under this framework, meta-learning is the process of correlating areas of expertise as dictated by simple learners, with the performance of other -more advanced-learners.

3.3.4 Meta-Decision Trees. Another approach in the field of learning from base learners consists of combining several inductive models by means of induction of meta-decision trees (Todorovski and Dzeroski, 1999; Todorovski and Dzeroski, 2000; Todorovski and Dzeroski, 2003). The general idea is to build a decision tree where each internal node is a meta-feature that measures a property of the class probability distributions predicted for a given example by a set of given models. Each leaf node corresponds to a predictive model. Given a new example, a meta-decision tree indicates the model that appears most suitable in predicting its class label.

3.4 Inductive Transfer and Learning to Learn

We have mentioned above how learning is not an isolated task that starts from scratch on every new task. As experience accumulates, a learning mech-anism is expected to perform increasingly better. One approach to simulate the accumulation of experience is by transferring meta-knowledge across domains or tasks; a process known as inductive transfer (Pratt *et al.*, 1991). The goal here is not to match meta-features with a meta-knowledge base (Figure 33.2), but simply to incorporate the meta-knowledge into the new learning task.

A review of how neural networks can learn from related tasks is provided by (Pratt *et al.*, 1991). Caruana (1997) shows the reasons explaining why learn-ing works well in the context of neural networks using backpropagation. In essence, training with many domains in parallel on a single neural network induces information that accumulates in the training signals; a new domain can then benefit from such past experience. Thrun (1998) proposes a learn-ing algorithm that groups similar tasks into clusters. A new task is assigned to the most related cluster; inductive transfer takes place when generalization exploits information about the selected cluster.

3.4.1 A Theoretical Framework of Learning-to-Learn. Several studies have provided a theoretical analysis of the learning-to-learn paradigm

within a Bayesian view (Baxter, 1998), and within a Probably Approximately Correct or PAC view (Baxter, 2000). In the PAC view, meta-learning takes place because the learner is not only looking for the right hypothesis in a hypothesis space, but in addition is searching for the right hypothesis space in a family of hypothesis spaces. Both the VC dimension and the size of the family of hypothesis spaces can be used to derive bounds on the number of tasks, and the number of examples on each task, required to ensure with high probability that we will find a solution having low error on new training tasks.

3.5   Dynamic-Bias Selection

A field related to the idea of learning-to-learn is that of dynamic-bias selection. This can be understood as the search for the right hypothesis space or concept representation as the learning system encounters new tasks. The idea, however, departs slightly from our architecture; meta-learning is not divided into two modes (i.e., knowledge-acquisition and advisory), but rather occurs on a single step. In essence, the performance of a base learner (Figure 33.1E) can trigger the need to explore additional hypothesis spaces, normally through small variations of the current hypothesis space.

As an example, DesJardins and Gordon (1995) develop a framework for the study of dynamic bias as a search in different tiers. Whereas the first tier refers to a search over a hypothesis space, additional tiers search over families of hypothesis spaces. Other approaches to dynamic-bias selection are based on changing the representation of the feature space by adding or removing features (Utgoff, 1986; Gordon, 1989; Gordon, 1990). Alternatively, Baltes (1992) describes a framework for dynamic selection of bias as a case-based meta-learning system; concepts displaying some similarity to the target concept are retrieved from memory and used to define the hypothesis space.

A slightly different approach is to look at dynamic-bias selection as a form of data variation, but as a time-dependent feature (Widmer, 1996A; Widmer, 1996B; Widmer, 1997). The idea is to perform online detection of concept drift with a single base-level classifier. The meta-learning task consists of identifying contextual clues, which are used to make the base-level classifier more selective with respect to training instances for prediction. Features that are characteristic of a specific context are identified and contextual features are used to focus on relevant examples (i.e., only those instances that match the context of the incoming training example are used as a basis for prediction).

4.   Tools and Applications

4.1   METAL DM Assistant

The METAL DM Assistant (DMA) is the result of an ambitious European Research and Development project broadly aimed at the development of methods and tools for providing support to users of machine learning and Data Mining technology. DMA is a web-enabled prototype assistant system that supports users with model selection and model combination. The project has as its main goal improving the utility of Data Mining tools and in particular to provide significant savings in experimentation time.

DMA follows a ranking strategy as the basis for its advice in model selection (Section 3.1.1). Instead of delivering a single model candidate, the software assistant produces an ordered list of models, sorted from best to worst, based on a weighted combination of parameters such as accuracy and training time. The task characterisation is based on statistical and information-theoretic measures (Section 3.1). DMA incorporates more than one ranking method. One of them exploits a ratio of accuracies and times (Brazdil and Soares, 2003). Another, referred to as DCRanker (Keller *et al.*, 1999), is based on a technique known as Data Envelopment Analysis (Andersen and Petersen, 1993; Paterson, 2000).

DMA is the result of a long and consistent effort in providing a practical and effective tool to users in need for assistance in model selection and guidance (Metal, 1998). In addition to a large number of controlled experiments on synthetic datasets and real-world datasets, DMA has been instrumental as a decision support tool within DaimlerChrysler and in the field of Computer-Aided Engineering Design (Keller *et al.*, 2000).

5.   Future Directions and Conclusions

One important research direction in meta-learning consists of searching for alternative meta-features in the characterization of datasets (Section 3.1). A proper characterization of datasets can elucidate the interaction between the learning mechanism and the task under analysis. Current work has only started to unveil relevant meta-features; clearly much work lies ahead. For example, many statistical and information-theoretic measures adopt a global view of the example distribution under analysis; meta-features are obtained by averaging results over the entire training set, implicitly smoothing the actual example distribution (e.g., class-conditional entropy is estimated by projecting all training examples over a single feature dimension.). There is a need for alternative -more detailed- descriptors of the example distribution in a form that can be related to learning performance.

Another interesting path for future work is to understand the difference between the nature of the meta-learner and that of the base-learners. In particular,

our general architecture assumes a meta-learner (i.e., a high-level generalization method) performing a form of model selection, mapping a training set into a learning strategy (Figure 33.2). Commonly we look at the problem as a learning problem itself where a meta-learner is invoked to output an approximating function mapping meta-features to learning strategies (e.g., learning model). This opens many questions, such as how can we improve the meta-learner which can now be regarded as a base learner? (Schmidhuber, 1995; Vilalta, 2001). Future research should investigate how the nature of the meta-learner can differ from the base-learners to improve the learning performance as we extract knowledge across domains or tasks.

We conclude this chapter by emphasizing the important role of meta-learning as an assistant tool in the tasks of model selection and combination (Section 4). Classification and regression tasks are common in daily business practice across a number of sectors. Hence, any form of decision support offered by a meta-learning assistant has the potential of bearing a strong impact for Data Mining practitioners. In particular, since prior expert knowledge is often expensive, not always readily available, and subject to bias and personal preferences, meta-learning can serve as a promising complement to this form of advice through the automatic accumulation of experience based on the performance of multiple applications of a learning system.

References

Aha D. W. Generalizing from Case Studies: A Case Study. Proceedings of the Ninth International Workshop on Machine Learning; 1-10, Morgan Kaufman, 1992.

Ali K., Pazzani M. J. Error Reduction Through Learning Model Descriptions. Machine Learning, 24, 173-202, 1996.

Andersen, P., Petersen, N.C. A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science*, **39**(10):1261-1264, 1993.

Baltes J. Case-Based Meta Learning: Sustained Learning Supported by a Dynamically Biased Version Space. Proceedings of the Machine Learning Workshop on Biases in Inductive Learning, 1992.

Baxter, J. Theoretical Models of Learning to Learn. In Learning to Learn, Chapter 4, 71-94, MA: Kluwer Academic Publishers, 1998.

Baxter, J. A Model of Inductive Learning Bias. Journal of Artificial Intelligence Research, 12: 149-198, 2000.

Bensusan, H. God Doesn't Always Shave with Occam's Razor – Learning When and How to Prune. In Proceedings of the Tenth European Conference on Machine Learning, 1998.

Bensusan, H., Giraud-Carrier, C. Discovering Task Neighbourhoods Through Landmark Learning Performances. In Proceedings of the Fourth Euro-

pean Conference on Principles and Practice of Knowledge Discovery in Databases, 2000.

Bensusan H., Giraud-Carrier C., Kennedy C. J. A Higher-Order Approach to Meta-Learning. Eleventh European Conference on Machine Learning, Workshop on Meta-Learning: Building Automatic Advice Strategies for Model Selection and Method Combination, Barcelona, Spain. 2000.

Berrer, H., Paterson, I., Keller, J. Evaluation of Machine-learning Algorithm Ranking Advisors. In Proceedings of the PKDD-2000 Workshop on Data-Mining, Decision Support, Meta-Learning and ILP: Forum for Practical Problem Presentation and Prospective Solutions, 2000.

Brazdil P. Data Transformation and Model Selection by Experimentation and Meta-Learning. Proceedings of the ECML-98 Workshop on Upgrading Learning to Meta-Level: Model Selection and Data Transformation, 11-17, Technical University of Chemnitz, 1998.

Brazdil, P., Soares, C. A Comparison of Ranking Methods for Classification Algorithm Selection. In Proceedings of the Twelfth European Conference on Machine Learning, 2000.

Brazdil, P., Soares, C., Pinto da Costa, J. Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results. Machine Learning, 50(3): 251-277, 2003.

Breiman, L. Stacked Regressions. *Machine Learning, 24:49-64,* 1996.

Brodley, C. Addressing the Selective Superiority Problem: Automatic Algorithm/Model Class Selection. Proceedings of the Tenth International Conference on Machine Learning, 17-24, San Mateo, CA, Morgan Kaufman, 1993.

Brodley, C. Recursive Automatic Bias Selection for Classifier Construction. Machine Learning, 20, 1994.

Brodley C., Lane T. Creating and Exploiting Coverage and Diversity. Proceedings of the AAAI-96 Workshop on Integrating Multiple Learned Models, 8-14, Portland, Oregon, 1996.

Caruana, R. Multitask Learning. Second Special Issue on Inductive Transfer. Machine Learning, 28: 41-75, 1997.

Chan P., Stolfo S. Experiments on Multistrategy Learning by Meta-Learning. Proceedings of the International Conference on Information Knowledge Management, 314-323, 1993.

Chan, P., Stolfo, S. On the Accuracy of Meta-Learning for Scalable Data Mining. Journal of Intelligent Information Systems, 8:3-28, 1996.

Chan P., Stolfo S. On the Accuracy of Meta-Learning for Scalable Data Mining. Journal of Intelligent Integration of Information, Ed. L. Kerschberg, 1998.

DesJardins M., Gordon D. F. Evaluation and Selection of Biases in Machine Learning. Machine Learning, 20, 5-22, 1995.

Dzeroski, Z. Is Combining Classifiers Better than Selecting the Best One? Proceedings of the Nineteenth International Conference on Machine Learning, pp 123-130, San Francisco, CA, Morgan Kaufmann, 2002.

Engels, R., Theusinger, C. Using a Data Metric for Offering Preprocessing Advice in Data-mining Applications. In Proceedings of the Thirteenth European Conference on Artificial Intelligence, 1998.

Freund, Y., Schapire, R. E. Experiments with a New Boosting Algorithm. In Proceedings of the 13th International Conference on Machine Learning, 148-156, Morgan Kaufmann, 1996.

Friedman, J., Hastie, T., Tibshirani, R. Additive Logistic Regression: A Statistical View of Boosting. Annals of Statistics 28: 337-387, 2000.

Fürnkranz, J., Petrak J. An Evaluation of Landmarking Variants, in C. Giraud-Carrier, N. Lavrac, Steve Moyle, and B. Kavsek, editors, Working Notes of the ECML/PKDD 2000 Workshop on Integrating Aspects of Data Mining, Decision Support and Meta-Learning, 2001.

Gama, J., Brazdil, P. A Characterization of Classification Algorithms. Proceedings of the Seventh Portuguese Conference on Artificial Intelligence, EPIA, 189-200, Funchal, Madeira Island, Portugal, 1995.

Gama, J., Brazdil P. Cascade Generalization, Machine Learning, 41(3), Kluwer, 2000.

Giraud-Carrier, C. Beyond Predictive Accuracy: What? Proceedings of the ECML-98 Workshop on Upgrading Learning to Meta-Level: Model Selection and Data Transformation, 78-85, Technical University of Chemnitz, 1998.

Giraud-Carrier, C., Vilalta, R., Brazdil, P. Introduction to the Special Issue on Meta-Learning. Machine Learning, 54: 187-193, 2004.

Gordon D. Perlis D. Explicitly Biased Generalization. Computational Intelligence, 5, 67-81, 1989.

Gordon D. F. Active Bias Adjustment for Incremental, Supervised Concept Learning. PhD Thesis, University of Maryland, 1990.

Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series, 2001.

Hilario, M., Kalousis, A. Building Algorithm Profiles for Prior Model Selection in Knowledge Discovery Systems. Engineering Intelligent Systems, 8(2), 2000.

Keller, J., Holzer, I., Silvery, S. Using Data Envelopment Analysis and Cased-based Reasoning Techniques for Knowledge-based Engine-intake Port Design. In Proceedings of the Twelfth International Conference on Engineering Design, 1999.

Keller, J., Paterson, I., Berrer, H. An Integrated Concept for Multi-Criteria-Ranking of Data-Mining Algorithms. Eleventh European Conference on Machine Learning, Workshop on Meta-Learning: Building Automatic Ad-

vice Strategies for Model Selection and Method Combination, Barcelona, Spain, 2000.

Merz C. Dynamic Learning Bias Selection. Preliminary papers of the Fifth International Workshop on Artificial Intelligence and Statistics, 386-395, Florida, 1995A.

Merz C. Dynamical Selection of Learning Algorithms. Learning from Data: Artificial Intelligence and Statistics, D. Fisher and H. J. Lenz (Eds.), Springer-Verlag, 1995B.

Metal. A Meta-Learning Assistant for Providing User Support in Machine Learning and Data Mining, 1998.

Michie, D., Spiegelhalter, D. J., Taylor, C.C. Machine Learning, Neural and Statistical Classification. England: Ellis Horwood, 1994.

Nakhaeizadeh, G., Schnabel, A. Development of Multi-criteria Metrics for Evaluation of Data-mining Algorithms. In Proceedings of the Third International Conference on Knowledge Discovery and Data-Mining, 1997.

Paterson, I. New Models for Data Envelopment Analysis, Measuring Efficiency with the VRS Frontier. Economics Series No. 84, Institute for Advanced Studies, Vienna, 2000.

Peng, Y., Flach, P., Brazdil, P., Soares, C. Decision Tree-Based Characterization for Meta-Learning. In: ECML/PKDD'02 Workshop on Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning, 111-122. University of Helsinki, 2002.

Pfahringer, B., Bensusan, H., Giraud-Carrier, C. Meta-learning by Landmarking Various Learning Algorithms. In Proceedings of the Seventeenth International Conference on Machine Learning, 2000.

Pratt, L., Thrun, S. Second Special Issue on Inductive Transfer. Machine Learning, 28, 1997.

Pratt S., Jennings B. A Survey of Connectionist Network Reuse Through Transfer. In Learning to Learn, Chapter 2, 19-43, Kluwer Academic Publishers, MA, 1998.

Schmidhuber J. Discovering Solutions with Low Kolmogorov Complexity and High Generalization Capability. Proceedings of the Twelve International Conference on Machine Learning, 488-49, Morgan Kaufman, 1995.

Skalak, D. Prototype Selection for Composite Nearest Neighbor Classifiers. PhD thesis, University of Massachusetts, Amherst, 1997.

Soares, C., Brazdil, P. Zoomed Ranking: Selection of Classification Algorithms Based on Relevant Performance Information. In Proceedings of the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases, 2000.

Soares, C., Petrak, J., Brazdil, P. Sampling-Based Relative Landmarks: Systematically Test-Driving Algorithms Before Choosing. Proceedings of the 10th Portuguese Conference on Artificial Intelligence, Springer, 2001.

Sohn, S.Y. Meta Analysis of Classification Algorithms for Pattern Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(11): 1137-1144, 1999.

Thrun, S. Lifelong Learning Algorithms. In Learning to Learn, Chapter 8, 181-209, MA: Kluwer Academic Publishers, 1998.

Ting, K. M., Witten I. H. Stacked generalization: When does it work?. In Proceedings of the 15th International Joint Conference on Artificial Intelligence, pp 866-873, Nagoya, Japan, Morgan Kaufmann, 1997.

Todorovski, L., Dzeroski, S. Experiments in Meta-level Learning with ILP. In Proceedings of the Third European Conference on Principles and Practice of Knowledge Discovery in Databases, 1999.

Todorovski, L., Dzeroski, S. Combining Multiple Models with Meta Decision Trees. In Proceedings of the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases, 2000.

Todorovski, L., Dzeroski, S. Combining Classifiers with Meta Decision Trees. Machine Learning 50 (3), 223-250, 2003.

Utgoff P. Shift of Bias for Inductive Concept Learning. In Michalski, R.S. et al (Ed), Machine Learning: An Artificial Intelligence Approach Vol. II, 107-148, Morgan Kaufman, California, 1986.

Vilalta, R. Research Directions in Meta-Learning: Building Self-Adaptive Learners. International Conference on Artificial Intelligence, Las Vegas, Nevada, 2001.

Vilalta, R., Drissi, Y. A Perspective View and Survey of Meta-Learning. Journal of Artificial Intelligence Review, 18 (2): 77-95, 2002.

Widmer, G. On-line Metalearning in Changing Contexts. MetaL(B) and MetaL(IB). In Proceedings of the Third International Workshop on Multistrategy Learning (MSL-96), 1996A.

Widmer, G. Recognition and Exploitation of Contextual Clues via Incremental Meta-Learning. In Proceedings of the Thirteenth International Conference on Machine Learning (ICML-96), 1996B.

Widmer, G. Tracking Context Changes through Meta-Learning. Machine Learning, **27**(3): 259-286, 1997.

Wolpert D. Stacked Generalization. Neural Networks, 5: 241-259, 1992.