Title
Harmonia loosely praestabilitia: discovering adequate inductive strategies

Permalink
https://escholarship.org/uc/item/6nb9d3wq

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 22(22)

Authors
Bensusan, Hilan
Giraud-Carrier, Chrsitophe

Publication Date
2000

Peer reviewed

Harmonia loosely *praestabilita*: discovering adequate inductive strategies

Hilan Bensusan
Christophe Giraud-Carrier
Department of Computer Science, University of Bristol,
Bristol BS8 1UB, UK
hilanb,cgc@cs.bris.ac.uk

Abstract

Landmarking is a novel approach to inductive model selection in Machine Learning. It uses simple, barebone inductive strategies to describe tasks and induce correlations between tasks and strategies. The paper presents the technique and reports experiments showing that landmarking performs well in a number of different scenarios. It also discusses the implications of landmarking to our understanding of inductive refinement.

Introduction

One of the central goals of cognitive science is to uncover mechanisms that allow agents to produce and manage knowledge. Although informed by existing theories of living organisms, the chief contribution of artificial intelligence, is to investigate knowledge mechanisms in abstract, that is, independently of their psychological or neurological plausibility. Machine learning endeavours to study induction, one of the basis of knowledge production. It considers different inductive strategies, their performance in different scenarios.

Not surprisingly, different inductive strategies are adequate for different inductive tasks. Theoretical results show that there is no inductive strategy that can perform well in every conceivable task (Schaffer, 1994). Some practitioners of machine learning reacted to this predicament by insisting that not every conceivable inductive tasks is equally deserving of attention. If we concentrate on a subset of all conceivable tasks, some people claim that we should restrict ourselves to "real world problems", we can find a small number of strategies that can handle induction (Rao, Gordon & Spears 1995). The problem arises when we try to give a precise definition for the set of "real world problems". In any case, we face correlations between sets of tasks, or problems, and induction strategies. Strategies perform well only in a subset of the set of all tasks, this subset is often called the *area of expertise* of a strategy. Machine learning is then left to discover, by induction, correlations between inductive strategies and their area of expertise. One way of doing this is by automating this search for correlations between tasks and strategies. This process is often called *meta-learning* and a number of different approaches has been proposed (see Bensusan (1998,1999), Giraud-Carrier & Hilario (1998), Giraud-Carrier & Pfahringer (1999), Lindner & Studer (1999)). Meta-learning has a number of general consequences for the study of cognition.

This paper explores some of the general consequences of a new way of doing meta-learning, called *landmarking*. The technique has been introduced recently (Bensusan & Giraud-Carrier 2000; Pfahringer, Bensusan & Giraud-Carrier 2000) and some new results are reported here. Landmarking searches for correlations between tasks and inductive strategies by exploring the similarities between different strategies in order to locate the task in a map of areas of expertise. The discovery of similarities between strategies can prove to be a tool to refine inductive strategies and, ultimately, a way to sketch an explanation of human inductive success.

This paper is organised as follows. Next section introduces landmarking. The following section presents experiments that assess its performance. Then we consider some of its implication for the general study of induction and cognition. A last section concludes the paper.

Meta-learning through landmarking

Meta-learning is the endeavour to automatically discover correlations between tasks and inductive strategies. To simplify without loss of generality, let's concentrate on supervised learning tasks.[1] These tasks are composed by a set of examples described by attribute values and classified according to a target function. The induction of the difference in extension of the predicates "lemon" and "watermelon", for example, may include attributes such as COLOUR, SHAPE, SIZE. Something YELLOW, EGG-SHAPED, SMALL qualifies as lemon whereas something GREEN, ROUND, BIG is a watermelon. If the attributes that describe the example are not well-chosen, learning could be very difficult. Consider, as an example, the following worse set of attributes for the "lemon-watermelon" problem above: IS IT A VEGETABLE?, IS IT A FRUIT?, DOES IT FLY?. The two examples are now described as NO, NO, YES. The importance of the example description derives from the fact that inductive strategies rely on representations to generalise. Successful inductive hypotheses are the ones that can represent accurately the similarities and the differences relevant to the task.[2]

[1] Although there are different uses of the terms "induction" and "learning", in this paper we shall use the terms as interchangeable.

[2] Data representation is important because every learning strategy has what machine learning calls a *representational bias*, a preference for hypotheses with a specific representa-

Meta-learning tasks are inductive tasks. Here, the examples, instead of being lemons or watermelons, are inductive tasks classified according to the best inductive strategy to tackle them. Thus, we have: TASK1 -> NAIVE BAYES, TASK2 -> BACKPROPAGATION, TASK3 -> NEAREST NEIGHBOR etc where each of the inductive strategy mentioned after the arrow is the best strategy for the task before the arrow.[3] The meta-learning task is to use these examples to learn how to classify new tasks in terms of the most suitable inductive strategy. The crucial question for meta-learning is therefore how to describe tasks.

Different approaches to task description have been proposed. These include the use of statistical features of the dataset in the task (Michie et al. 1994) and the use of features of a decision tree representation of the task (Bensusan 1998; Bensusan 1999). In the latter, an inductive hypothesis, namely the one produced by a decision tree induction method, is used to describe the task. Landmarking also makes use of specific inductive methods to describe the task, but makes use of the method's performance rather than the method's induced hypothesis.

The basic idea of the landmarking approach is that the performance of an inductive strategy on a task uncovers information about the nature of the task. Tasks are described by a set of attributes corresponding to the performance of simple, efficient strategies on them. These strategies are expected to indicate which other, more refined strategy is the best to tackle the task. They act, therefore, as landmarkers, indicating where, in the space of all areas of expertise, the task belongs. It explores empirically the relationships between areas of expertise of different learners.

The kind of inference on which landmarking relies can be illustrated with the help of Figure 1. The rectangle represents a set of inductive tasks and the ellipses represent subsets of the set of tasks where a given inductive strategy performs well, that is, areas of expertise. Assume that $i1$, $i2$, and $i3$ are taken as landmarkers. In this case, landmarking concludes that problems on which both $i1$ and $i3$ perform well, but on which $i2$ performs poorly, are likely to be in $i4$'s area of expertise etc. Of course, the proximity of the areas of expertise of two strategies indicates some similarity between the inductive mechanisms behind them. For landmarking purposes, however, it is sufficient to concentrate on so-to-speak cartographic considerations. Tasks are described by how some landmarkers fare on them. Exploring the meta-learning potential of landmarking amounts to investigating how well a landmark learner's performance hints at the location of the respective learning tasks in the expertise map.

tion (Haussler, 1989; Russell & Grossof 1990). Thus, most Decision Tree induction algorithms prefer simpler decision trees, most rule induction algorithms prefer simpler rules. There is a trade-off between the need for good input representation and the strength of the strategy's preference (Craven & Shavlik, 1995).

[3]For a survey of the most used inductive strategies including all those to be mentioned in this paper consult Mitchell (1997).

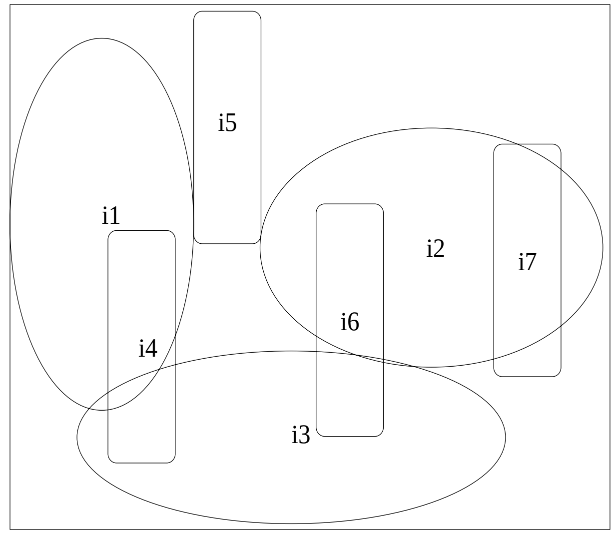

Figure 1: Example of map of areas of expertise

Landmarking relies on some simple and efficient inductive strategies to signpost the location of a task in a map of expertise areas. Landmarking discovers experimentally which inductive strategies are similar enough to have neighbouring areas of expertise. It therefore finds neighbourhoods of inductive strategies and, ultimately, draws a map of areas of expertise. While other approaches represent tasks in a way only indirectly related to their location in an expertise map, landmarking faces them as points in the map to be described in terms of their distance to some known milestones.

Landmarking is a tool to discover the areas of expertise of a learning device. In fact, this is the very goal of machine learning research: to establish the strength and the scope of different inductive strategies. In addition, it highlights which tasks fail to belong to the area of expertise of any of the existing inductive strategies. It can therefore direct the search of new strategies towards areas of the expertise map not currently covered by any learning method. If successful, it can guide the crafting of new inductive methods and work as the basis for a model of inductive refinement. Let's turn now to some experiments designed to measure its success.

Experiments with landmarking

A number of experiments to test landmarking are reported in (Bensusan & Giraud-Carrier 2000). The results show that landmarking successfully meta-learns in a number of different scenarios. Successful results mean that selection of inductive strategies can be done through the information contained in the performance of some landmark inductive strategies. In this section, we summarise some of these results and present new ones.

Experiments on landmarking can be done only through selecting a set of landmarkers. The landmarkers change according to what we call the *learners pool*, i.e., the set of target learners from which one must be selected. It remains to be investigated how close can we get from a universal set of landmarkers that can be used in any learners pool. In the experiments reported here, we used the following set of landmark learners. For each, we include the motivation for its inclusion in the set.

1. *Decision node:* A single decision node is chosen according to C5.0's information gain-ratio (Quinlan 1993, Mitchell 1997). The node is then used to classify test examples. This landmark learner aims to establish closeness to linear separability.

2. *Randomly chosen node:* A randomly chosen attribute is used to split the training set and classify the test examples. This landmark learner informs about irrelevant attributes.

3. *Worst node:* Here the gain-ratio information criterion is used to pick up the least informative attribute to make the single split. This landmarker, together with the first one, is supposed to tell us something else about linear separability: if neither the best nor the worst attribute produce a single well performing separation, it is likely that linear separation is not an adequate learning strategy.

4. *Naive Bayes:* The training set is used to estimate the probabilities required to use the Bayes theorem to classify test cases (Mitchell, 1997). This landmark learner intends to measure how conditionally independent the attributes are, given the class.

5. *1-Nearest Neighbor:* The test set is classified based on the classification of the closest training example (Mitchell 1997). This landmark learner measures how close instances that belong to the same class are.

6. *Elite 1-Nearest Neighbor:* This computes 1-Nearest Neighbor on a subset of all attributes. This elite subset is composed by the most informative attributes if the gain-ratio difference between them is smaller than 0.1[4]. Otherwise, the elite subset is a singleton and the learner acts like a decision node learner. This landmark learner intends to establish whether the task is a relational one, that is, if it involves parity-like relationships between the attributes (Clark & Thornton, 1997). In relational tasks, no single attribute is considerably more informative than others.

7. *Majority-class guesser:* The test set is classified according to the most common class in the training set. This landmark learner intends to inform about the frequency of the majority class.

8. *Linear Discriminant:* A linear approximation of the target function is sought (Gama, 1999). This landmark learner intends to measure closeness to simple linear separation.

[4]This threshold is based on the results reported in Bensusan (1999).

The performance of the different landmarkers are given by the average performance on all the existing examples of the induction problem, the so-called *instance space* of the induction made from 10 different subset of examples (*training sets*) of equal size. This approach, although different from the standard practice of *cross-validation* where the sets of examples are drawn without replacement, is an efficient way to estimate how the landmark learners perform in each task. Therefore, inductive tasks are described by landmarker's performance values. The task is then labelled by the learner with greater average accuracy on the task, using a 10-fold cross-validation approach. Each task can be labelled by a learner's name or as "tie" when the difference in performance between the best and the worst learner is less than 10%. A (meta-)dataset with 5 examples described by 4 landmarkers looks as follows:

```
0.42187,0.46875,0.46250,0.30781,Ripper
0.45312,0.42187,0.45000,0.26250,IB
0.54687,0.56250,0.45937,0.29844,C5.0tree
0.51562,0.59375,0.43750,0.28750,MLP
0.43750,0.51562,0.43125,0.27812,tie
```

Given the (meta-)dataset, the meta-learner aims at finding correlations between the performances of the learners in the pool and that of the landmarkers.

In the first experiment, we compared landmarking with an existing approach to task description for meta-learning. This approach uses a number of information-theoretical properties of the data to describe the task (Michie et al. 1994). We implemented this information-theoretical approach by considered the following 6 features defined on literature as meta-attributes: Entropy of the class, Average entropy of the attributes, Mutual information, Joint entropy, Equivalent number of attributes, Signal-to-noise ratio. The task was to select among the following 10 learning algorithms: C5.0, C5.0 with boosting, C5.0 rules, Multi-layer perceptron trained with backpropagation (MLP), Radial-based function networks (RBF), Linear discriminant, Ltree (see Gama, 1999), Naive Bayes (NB), Instance-Based inducer (IB) and Ripper. Landmarkers 1,2,3,4,5,6,8 were used. 320 Boolean tasks were considered. The 10 learning algorithms in the learner pool were also used for meta-learning in all experiments. Error rates were based on stratified 10-fold cross-validation. Results are given in Table 1. The first line reports the error rate of the default class that, in this case, was "tie".

The table shows that landmarking outperforms the information-based task description and therefore it is a suitable competitor. Notice that landmarking outperforms the information-based approach with all of the 10 meta-learners. Moreover, the difference in error is around 10% with the three C5.0 meta-learners. The table also shows that adding the information-based features to describe the task impairs landmarking performance.

Next, we considered a number of learners pools with two inductive strategies. Learners pools were composed by pairs of the following inductive strategies: C5.0(with boosting), C5.0(rules), Naive Bayes (NB), Instance-

Table 1: Comparison between different ways to describe tasks: performances of the landmarking approach (L), the information-based approach (Info) and the combined approach (Combined) using 10 different meta-learners.

Meta-learner	Land	Info	Combined
Default Class	0.460	0.460	0.460
C5.0boost	0.248	0.360	0.295
C5.0rules	0.239	0.333	0.301
C5.0tree	0.242	0.342	0.314
MLP	0.301	0.317	0.320
RBFN	0.289	0.323	0.304
LD	0.335	0.311	0.301
Ltree	0.270	0.317	0.286
IB	0.329	0.366	0.342
NB	0.429	0.407	0.363
Ripper	0.292	0.314	0.295
Average	0.298	0.339	0.312

Based induction (IB), Ripper and Multi-layer perceptron (MLP). Landmarkers 1,2,4,5,6,7,8 were used. Tasks were classified as a *tie* between the two strategies when the average error difference between the learners in the pool was less than 0.1. We used 927 artificially generated Boolean and MONK-like datasets (Thrun et al, 1991). Boolean instance spaces had between 5 and 12 attributes. The error rates given in table 2 are the average 10-fold cross-validation error of 5 inductive strategies used for meta-learning: IB, MLP, C5.0boost, Ripper and Radial Basis Function Network Induction (RBF).

Table 2: Landmarking to choose between pairs of learners

Learner pool	Error
NB-IB	0.383
NB-MLP	0.179
NB-Ripper	0.181
C5.0boost-MLP	0.246
C5.0boost-NB	0.359
C5.0rules-Ripper	0.204

In a different experiment, we looked at the suitability of inductive strategies and groups of similar inductive strategies. We considered that a task is suitable for a learner if it performs better than the average of 10 standard learners: C5.0, C5.0rules, C5.0boost, MLP, RBF, Linear Discriminant, Ltree, NB, IB and Ripper. For this experiment we used only landmarkers 1,2,3 and 6 as they are all decision node based and are arguably enough to diagnose at least whether decision tree induction is a good way to approach the task. We used 222 tasks from the set used in the previous experiment and the 10 standard learners mentioned above to perform the meta-learning induction. We looked at the suitability of IB, NB, C5.0boost, neural network inductive strategies (MLP and RBF) in general (NN), rule induction strategies (Ripper and C5.0rules) and decision tree strategies (C5.0, C5.0boost, Ltree). The error rates given in table 3 are the average 10-fold cross-validation error of the 10 inductive strategies used for meta-learning.

Table 3: Suitability of inductive approaches. Error rates for the default class prediction and for meta-learning with landmarking are given.

Approach	Default class	Landmarking
IB	0.420	0.297
NB	0.380	0.298
C5.0boost	0.510	0.449
NN	0.440	0.386
Rules	0.370	0.281
Trees	0.470	0.390

These results show that most meta-learners produce error levels smaller than the default error class and often the difference is substantial. Notice that error rate figures don't reflect the overall performance, that is the accuracy of the selected learning model. In another experiment, we tried to estimate this by using the 222 Boolean problems as tasks of a meta-learning training set and 18 other tasks to test the hypotheses and compare the selected approach with the best performing one. The 18 tasks of the test set were from the standard repository of benchmark induction problems maintained by the University of California at Irvine (UCI); these are commonly considered to be "real world problems". We chose the following problems: mushrooms, abalone, crx, sat, acetylation, titanic, waveform, yeast, car, chess(king-rook-vs-king), led7, led24, tic-tac-toe, MONK1, MONK2, MONK3, satimage, quisclas.

The results reported for this experiment are the average error difference between the best choice and the selected choice in the 18 UCI problems. If the average is in fact better than the chosen model, we consider the error difference between the chosen model and the average. Similarly if the meta-learner had chosen against the model that in fact is better than the average of the 10 learners. Here we used only C4.5 (Quinlan, 1993) as meta-learner. Average error difference appear in table 4.

Table 4: Average error difference between best and chosen option in the 18 UCI datasets

Approach	Error difference
IB	0.0356
NB	0.0165
C5.0boost	0.0443
NN	0.0314
Rules	0.0360
Trees	0.0211

The small average error difference shows that the chosen strategy, even when is not the best, performs well. It shows that landmarking seldom make choices that perform considerably worse than the best alternative. This is confirmed further by an experiment in the same scenario. Now we used only the 14 UCI tasks listed above as training set and tested the C4.5 hypothesis in the remaining four UCI tasks (MONK2, MONK3, satimage, quisclas). The results obtained have a greater variation than the previous one but shows that in some cases landmarking perform completely accurately. Table 5 summarises the new results.

Table 5: Average error difference between best and chosen option in 4 UCI tasks after training on 14 UCI tasks only

Approach	Error difference
IB	0.0675
NB	0.0605
C5.0boost	0.0000
NN	0.0000
Rules	0.0443
Trees	0.0172

These results, although still preliminary, show that landmarking is capable to select inductive approaches. They suggest that it pays off to run bare-bone, landmark inductive strategies on a number of tasks and learn how their performance relates to that of other, more fleshed-out strategies. This far, we have indicated how the performance of simple inducers in a task can be used for meta-learning. We move now to the significance of landmarking for a general theory of induction.

Discovering inductive strategies

For me [...] the problem of induction is a problem about the world: a problem of how we, as we are now [...], in a world we never made, should stand better than random or coin-tossing chances of coming out right when we predict by inductions that are based on our innate [...] similarity standard. Darwin's natural selection is a plausible partial explanation. W. V. O. Quine

One of the problems of explaining human (and animal) cognitive practices in general and inductive practices in particular is to account for success. Humans are remarkably good at inducing in familiar environments and seem to make heavy use of their background knowledge accumulated through inductions made in their lifetime history or received as cultural material. Studies on human induction on tasks similar to the MONK problems have established that prior knowledge influences the rate of concept learning, and the logical form of concepts formed during learning is a function of the logical form of the concepts previously acquired (Pazzani, 1991). In general, humans rely on previous acquisition of concepts and common-sense knowledge about the area to learn new concepts (Wisniewski & Medin, 1994; Heit, 1994). Background knowledge and the ability to meta-learn enable humans, when for instance engaged with scientific theory building, to perform successful inductions from one or few examples.

Human inductive trajectory from innate instincts to refined theories about the world is Quine's view of the problem of induction: a problem about the world. A plausible partial complement to Darwin's natural selection is to find a model of exploiting previous induction experience to boost performance. Such model, of course, has to accommodate the partial explanation role that natural selection plays. The inductive trajectory towards greater efficiency in familiar environments had its origins in evolutionary selection of relevant inductive mechanisms. Recent there have been attempts to characterise human innate inductive tendencies in terms of learning biases (Elman et al., 1996; Dessalles, 1998). Leaving aside the question of how our inductive practices are guided by our innate instincts, we can sketch a model of the human inductive trajectory according to which our similarity standards by means of which we generalise are partly product of evolution, partly a consequence of a gradual process of refinement. We claim that landmarking can be part of an account of inductive refinement.

Landmarking is a technique to select the most adequate inductive strategy for a task, but it can also be seen as an instrument for inductive refinement. It suggests ways in which better, increasingly appropriate inductive strategies, can be constructed from rudimentary ones. Landmarkers are simple inductive strategies that can characterise tasks. Thus, they can outline new inductive strategies to adequately cover areas of the expertise map; describing the area in terms of how different learning biases fare there is a step towards constructing more refined biases that can tackle it. As a way to describe tasks, landmarking has far-reaching consequences beyond strategy selection: to landmark a group of tasks could be the first step towards the development of an inductive strategy to tackle it. This is arguably what happens when a scientist applies various simple methods to a problem in order to get information about what more sophisticated method to develop. This could also be what happens when new problems had to be addressed by humans with only few, unrefined inductive tools. Landmarking is a way to discover relationships between different strategies and, as such, to establish what is needed to ease learning. In this sense, it not only bears similarities with other methods that exploit the nature of the task to decide which way to go (Clark & Thornton, 1997) but also can be seen as a general framework for those methods as it describes tasks only in terms of a portfolio of learning performances. The emerging picture is one where the records of failure and success of the current induction tools are used to inform how these tools need refinement. Successful learning, landmarking suggests, might require learning with previous mistakes and accomplishments.

Conclusions

Wär nicht das Auge sonnenhaft, die Sonne koennt'es nie erblicken. Goethe, Zahme Xenien, Werke, Weimar 1887-1918, bk 3, 1805.

Landmarking is a strategy to describe tasks so that no more than a small class of efficient learning algorithms is required. Tasks are described by their position in the expertise map. It can also be used to locate and explore expertise *terra incognita*. It can be seen as part of a model of inductive refinement whereby the description of a task in terms of landmarkers offers the raw material for the development of new induction tools. The picture offered by this model is one in which human inductive abilities are roughly tuned to their environment; no survival and no refinement could start from a completely alien inductive toolkit. Evolution gives part of the explanation. But the gradual refinement that sharpens the kit and assembles new instruments is what turns the original *harmonia* loosely *praestabilita* into an inductively adapted species.

Acknowledgements

This work is part of the METAL project supported by an ESPRIT Framework IV LTR Grant (Nr 26.257). We wish to thank the members of the Consortium for useful comments and discussion.

Bensusan, H. (1988). God doesn't always shave with Occam's Razor – learning when and how to prune. *Proceedings of the 10th European Conference on Machine Learning* (pp. 119-124). Berlin: Springer.

Bensusan, H. (1999). Automatic bias learning: an inquiry into the inductive basis of induction. *PhD Thesis* School of Cognitive and Computing Sciences, University of Sussex, UK.

Bensusan, H. & Giraud-Carrier. (2000) Discovering task neighbourhoods through landmark learning performances. *Submitted*

Clark, A., & Thornton, C. (1997). Trading spaces: computation, representation and the limits of uninformed learning. *Behaviour and Brain Sciences*, 20, 1, (pp. 57-90).

Craven, M. & Shavlik, J. (1995). Investigating the Value of a Good Input Representation. in: *Computational Learning Theory and Natural Learning Systems*, ed: Petsche, T. and Judd, S. and Hanson, S., Cambridge, MA, USA: MIT Press.

Dessales, J-L. (1998). Characterising Innateness in artificial and natural learning. *Proceedings of the Workshop on Human and Machine Learning, 10th European Conference on Machine Learning*, Technische Universität Chemnitz, Chemnitz, Germany.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking Innateness*. Cambridge, MA, USA: MIT Press.

Gama, J. (1999). Discriminant trees. *Proceeding of the Sixteeth International Conference on Machine Learning*, (pp. 134-142), San Mateo, CA: Morgan Kaufmann.

Giraud-Carrier, C. & Hilario, M., ed. (1998). *ECML'98 Workshop Notes - Upgrading Learning to the Meta-Level: Model Selection and Data Transformation*, Technische Universität Chemnitz, Chemnitz, Germany.

Giraud-Carrier, C. & Pfahringer, B., ed. (1999). *Proceedings of the ICML'99 Workshop on Recent Advances in Meta-Learning and Future Work*, J. Stefan Institute, Ljubljana, Slovenia.

Haussler, D. (1989). Quantifying Inductive bias: AI learning algorithms and Valiant's learning framework. *Artificial Intelligence*, 32, 2, (pp. 177-222).

Heit, E. (1994). Models of the Effects of Prior Knowledge on Category Learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 20, 6, (pp. 1264-1282).

Lindner, G. and Studer, R. (1999). AST: Support for Algorithm Selection with a CBR Approach. *Proceedings of the 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases.* (pp. 418-423). Berlin: Springer.

Michie, D., Spiegelhalter, J. & Taylor, C. C. (1994). *Machine Learning, Neural and Statistical Classification*. London: Ellis Horwood.

Mitchell, T. (1997). *Machine Learning*. New York, NY, USA: McGraw Hill.

Pazzani, M. (1991). Influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 17, 3, (pp. 416-432).

Phahringer, B., Bensusan, H. & Giraud-Carrier, C. (2000). Meta-learning by landmarking various learning algorithms. To appear in: *Proceedings of the Seventeenth International Conference on Machine Learning, ICML'2000*.

Quinlan, J. R. (1993) *C4.5: Programs for Machine Learning* San Mateo, CA, USA: Morgan Kaufmann.

Rao, R.B. and Gordon, D. and Spears, W. (1995). For every generalization action, is there really an equal and opposite reaction? *Proceedings of the Twelfth International Conference on Machine Learning, ICML'95* San Mateo, CA, USA: Morgan Kaufmann.

Russell, S. J. & Grossof, B. (1990). Declarative bias: an overview. in: Benjamim, P. (ed.) *Change of representation and inductive bias.* (pp. 267-308). Dordrecht, Netherlands: Kluwer.

Schaffer, C. (1994). A Conservation Law for Generalization Performance. *Proceedings of the Eleventh Conference on Machine Learning, ICML'94*. San Mateo, CA, USA: Morgan Kaufmann.

Thrun, S. et alii (1991). The MONK's problems - a performance comparison of different learning algorithms. Technical report CMU-CS-91-197. School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA, USA.

Wisniewski, E, J. & Medin, D. L. (1994). On the interaction of theory and data in concept learning. *Cognitive Science*, 18, (pp. 221-281).