Available online at www.sciencedirect.com

ScienceDirect

Theoretical Computer Science 382 (2007) 167–169

Theoretical Computer Science

www.elsevier.com/locate/tcs

Foreword

This special issue of *Theoretical Computer Science* is dedicated to the Fifteenth International Conference on Algorithmic Learning Theory (ALT 2004) held at Padova University, Padova, Italy, October 2–5, 2004. It contains eight articles that were among the best in the conference.[1] The authors of these papers have been invited by the Special Issue Editors to submit completed versions of their work for this special issue. Once received, these papers underwent the usual refereeing process of *Theoretical Computer Science*.

Algorithmic learning theory is mathematics concerning computer programs which learn from experience. This involves considerable interaction between various mathematical disciplines including the theory of computation, statistics, and combinatorics. There is also considerable interaction with the practical, empirical fields of machine and statistical learning in which a principal aim is to predict, from past data for phenomena, useful features of future data for the same phenomena.

The papers in this special issue cover a broad range of topics of current research in the field of algorithmic learning theory. The categories featured are Inductive Inference, Query Learning, Logic Based Learning, Pattern Recognition, Statistical Supervised Learning, Online Sequence Prediction, and Approximate Optimization Algorithms.

Below we briefly introduce each of these subjects, relate the paper to it and point out the paper's specific research. Note that each of the formal methods of automated learning reflects various facets of the wide range of activities that can be viewed as *learning*.

Learning can be viewed as an indefinite process and as a finite activity with a defined termination. Models studied in Inductive Inference usually focus on indefinite learning processes, requiring only eventual success of the learner to converge to a satisfactory conclusion. This line of research was pioneered by Gold who introduced his model of learning in the limit. In this model the learner receives a stream of examples of a certain target concept and is required to output a sequence of hypotheses converging to this target. When two consecutive hypotheses output differ, then a *mind change* occurred. Since in general convergence is not decidable, there is always some uncertainty and one can interpret the number of mind changes as a measure of uncertainty.

The paper by Martin, Sharma and Stephan studies the problem why uncertainty is sometimes desirable. For answering this question the authors introduce and explore the concept of *frugal* learnability reflecting the data consumption complexity. Frugality is then explored in relation to mind change complexity and interesting tradeoffs are observed.

While in the classical inductive inference paradigm the learner passively receives the data, there are also models in which the learner is allowed to ask queries, e.g., Angluin's query learning model. We can imagine the learning via queries scenario as the interaction between a teacher and a learner that can communicate using some prespecified query language. For example, when learning the concept of a table, the learner may ask "Is a sofa an example of a table?", or when learning a language, a query may be of the form "Is w a string from the target language?". This type of query is referred to as a *membership query*. Alternatively, the learner may be allowed to ask "Is G a grammar for the target language?". This type of question is called an *equivalence query*. It should be obvious how to generalize equivalence queries to any learning domain.

Angluin also considered a combination of the modes of information presentation, i.e., the learner initially receives a finite set of data belonging to the target and is then allowed to ask membership queries. After having asked finitely

[1] The conference proceedings, including preliminary versions of these papers, appeared as Lecture Notes in Artificial Intelligence, vol. 3244, Springer, 2004.

0304-3975/$ - see front matter © 2007 Published by Elsevier B.V.
doi:10.1016/j.tcs.2007.03.036

many membership queries the learner has to output a hypothesis about the target and this hypothesis must be correct. So, here learning is a finite process with a defined termination.

Besombes and Marion study the learnability of regular tree languages in this model and present an efficient algorithm. From the viewpoint of computer science, tree languages are interesting because they model XML documents.

Another modification of the information presentation is considered in the paper by Sloan, Szörényi and Turán. The goal is to learn an unknown threshold function over a Boolean domain $\{0, 1\}^n$. Threshold functions are quite important, since they are widely used. The learner is allowed to ask membership and equivalence queries. But additionally, the learner is also given another function as input which should be considered as the initial guess of the target function. So, the main task is then to revise the initial guess until it correctly describes the target function. Note that an affirmative answer to an equivalence query directly yields the information that the target has been reached, and the learner can and has to stop. Solving such a revision problem in principle is trivial. Thus, the main problem is to design an efficient learner. The efficiency is measured in the number of queries asked. This number is measured depending on the number n of variables and, since an initial guess is given, the syntactic distance d between the target and the initial guess.

While this task is still easy to achieve if one allows polynomially many queries in the number of variables, the authors present a *super-efficient* algorithm. The number of queries that a super-efficient algorithm is allowed to ask must be uniformly bounded *polylogarithmically* in the number n of variables and polynomially in the syntactic distance d. As a matter of fact, the algorithm presented is polylogarithmic in the number of variables and linear in the syntactic distance.

The paper by Bulatov, Chen and Dalmau belongs also to the field of query learning. In their paper, only equivalence queries are allowed. The authors study the learnability of concept classes that consist of those concepts that are invariant under a “generalized majority–minority operation”. This generalizes previous work on the learnability of various intersection-closed concept classes including certain classes of quantified Boolean formulae such as 2-CNFs. The algorithm works using a new hypothesis representation called “signatures”.

Learning classifiers is another important topic which has been intensively studied within the learning theory community. An important area where classifiers are needed is pattern recognition. Here the learner observes a sequence (x_t, y_t) of training examples, where the x_t are the objects to be classified and the y_t are the labels. As an example one may think of the x_t as handwritten digits and the y_t are then from $\{0, \dots, 9\}$. The goal is then to learn a classifier that predicts as accurately as possible the label of any further instance.

Cesa-Bianchi surveys a number of results obtained by him and his co-authors concerning the behavior of algorithms for learning classifiers based on the solution of a regularized least-squares problem.

The next paper belongs to the field of Statistical Supervised Learning. That is, one aims again at learning a classifier. The data are generated by an underlying distribution P typically not given to the learner. Then the learner is fed a training set of labeled data which are generated identically and independently with respect to the distribution P. The classifier has to be chosen from a given class of functions such that it best imitates the underlying distribution. Such a learning strategy is successful if the difference between the sample and true performance is small for every function in the class provided the sample size is sufficiently large. This property is usually referred to as *uniform convergence* over the class of functions.

What can be achieved depends on the function class. If the class is too rich, then it may contain for every randomly generated training set a function precisely fitting it. Thus, the learner may overfit. Therefore, the notion of the *capacity* of a class has been introduced. Intuitively, the higher the capacity of the class the greater the risk of overfitting. Consequently, the problem is how to measure the capacity of a class. In the literature, two measures have been successfully used, i.e., the Rademacher and the Gaussian complexity.

Ambroladze, Parrado-Hernández, and Shawe-Taylor generalize and simplify some Rademacher and Gaussian complexity inequalities which are fundamental to application of these quantities in learning theory bounds. In particular, a new and general proof for the “contraction principle” for Rademacher averages is presented and interesting applications to Lipschitz functions are outlined.

The next paper deals with the problem of sequence prediction. Here the sequence is randomly generated by μ and one has to predict the next symbol x_n from an observed sequence $x_1, \dots, x_{n-1}$. Solomonoff proved in 1978 a very fundamental result concerning this problem. He showed that the prediction of a universal semimeasure M converges quite fast and with probability 1 to the true sequence generator μ provided that μ is computable. But Solomonoff’s result does not say anything concerning the problem of whether or not the convergence holds for any individual sequence. In particular, it remained open whether Solomonoff’s result also holds for all Martin-Löf random sequences.

Hutter and Muchnik answer this question negatively. They also construct a non-universal enumerable semimeasure M for which convergence always holds, whenever the sequence $x_1, x_2, \ldots$ is Martin-Löf with respect to a computable μ.

Support Vector Machines (abbreviated as SVM) have recently attracted considerable attention. They provide a successful technique for classification, function approximation and other key problems in statistical learning. The basic idea of SVM is to map the points to be classified into a high dimensional space. The dimension of the space must be so large that the points can be separated by a linear hyperplane. From all possible separating linear hyperplanes one has to choose one which has good generalization abilities. Statistical learning theory ensures that a good generalization error is achieved whenever the sample can be embedded in a space where some linear hypothesis can be found that separates positive examples from negative ones with large margin. So, one is interested in finding the maximum-margin hyperplane and this means one has to solve a convex quadratic optimization problem.

However, the density of the matrix representing the "quadratic part" of the cost function causes major problems, since storing the whole matrix is prohibitive. Thus, traditional techniques usually cannot be applied. Several methods have been proposed in the literature to overcome this difficulty. One method is decomposition. This method keeps track of a current feasible solution and improves it iteratively. In each iteration, a so-called working set is used (a subset of the variables) and the optimization problem is solved for the variables in the working set, while the remaining variables remain unchanged. Since the success of this method is quite sensitive to the selection of the working set, finding good policies for a working set selection is an important problem.

Simon studies the complexity of working set selection. The policy considered is the one selecting the subset with the largest first-order reduction. The author then shows that this selection is NP-complete, thus providing a theoretically founded explanation for why this selection is rarely used in practice.

In the second part of his paper, he shows that an approximation suffices and an efficient approximation algorithm is presented.

We would like to thank all the referees for their fine reports and their efficient work. Special thanks go to the members of the program committee of ALT 2004 for selecting the papers. We are very grateful to all authors for submitting their papers and for all their efforts to improve and to polish their articles.

Finally, we are particularly thankful to Giorgio Ausiello for the opportunity to compile this special issue.

Waterloo, Newark and Sapporo, March 3, 2007.

Special Issue Editors
Shai Ben-David
University of Waterloo,
David R. Cheriton School of Computer Science,
200 University Avenue West,
N2L 3G1 Waterloo, Ontario,
Canada
E-mail address: shai@cs.uwaterloo.ca.

John Case
Department of Computer and Information Sciences,
101A Smith Hall,
University of Delaware,
Newark, DE 19716-2586, USA
E-mail address: case@cis.udel.edu.

Thomas Zeugmann*
Division of Computer Science,
Graduate School of Information Science and Technology, Hokkaido University,
N-14, W-9 Sapporo 060-0814, Japan
E-mail address: thomas@ist.hokudai.ac.jp.

* Corresponding editor.