Theoretical Computer Science 558 (2014) 1–4

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Editorial

Guest Editors' foreword

This special issue of *Theoretical Computer Science* is dedicated to the 23rd International Conference on Algorithmic Learning Theory (ALT 2012) held in Lyon, France, October 29–31, 2012. It contains ten 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 revised, completed and polished versions of their work for this Special Issue. Once received, these papers underwent the usual refereeing process of *Theoretical Computer Science*. In the following, we shortly introduce each of the papers.

Algorithmic learning theory focuses on theoretical aspects of machine learning. It is dedicated to studies of learning from an algorithmic and mathematical perspective. Depending on the learning task considered, considerable interaction between various mathematical theories including statistics, probability theory, combinatorics, mathematical analysis, linguistics, and theory of computation is required. These studies comprise the investigation of various formal models of machine learning and statistical learning and the design and analysis of learning algorithms. This also leads to a fruitful interaction with the practical fields of machine learning, linguistics, psychology, and philosophy of science.

The first paper in this special issue belongs to the area of *inductive inference* of recursive functions which is one of the classical areas of algorithmic learning. In this setting the learner is usually fed augmenting finite sequences $f(0), f(1), f(2), \dots$ of the target function f. For each finite sequence the learner has to compute a hypothesis, i.e., a natural number. These numbers are interpreted with respect to a given enumeration of partial recursive functions comprising the target function. Then the number j output by the learner is interpreted as a program computing the jth function enumerated. The sequence of all hypotheses output by the learner has then to converge (to stabilize) on a program that, under the given correctness criterion, correctly computes the target function. This learning scenario is commonly called *explanatory inference* or *learning in the limit*. Since only finitely many values of the function have been seen by the learner up to the unknown point of convergence, some form of learning must have taken place. Usually, the goal is then to construct a learner that can infer all functions from a given target class U. Many variations of this model are possible. For finite learning one requires the point of convergence to be decidable. Another variation is to allow the learner to converge semantically, i.e., instead of stabilizing to a correct program, the learner is allowed to output infinitely many *different programs* which, however, beyond some point, all must correctly compute the target function. This model is commonly referred to as *behaviorally correct learning*. It should be noted that there is no single learner that can infer the whole class of recursive functions. Therefore, the study of further variations deserves interest. In their paper *Confident and Consistent Partial Learning of Recursive Functions* Ziyuan Gao and Frank Stephan considered partial learning originally introduced by Osherson, Stob, and Weinstein (1986), where the learner is required to output a correct program for the target function infinitely often and any other hypothesis only finitely often. Gao and Stephan refine this model by combining it with the confidence demand and with consistent learning. A confident learner is required to converge on every function, even it is not in the target class (but may stabilize to a special symbol "?"). The requirement to learn consistently in the context studied here means that the learner has to correctly reflect the information already obtained, and this demand is posed to all but finitely many of the hypotheses output. The resulting models are called confident partial learning and consistent partial learning, respectively. The paper contains many interesting results and masters several complicated proof techniques. In particular, it is shown that confident partial learning is more powerful than explanatory learning. On the other hand, the authors show that there are behaviorally correct learnable classes which are *not* confidently partially learnable. So, the learning model is also not trivial in the sense that it can infer every recursive function. Moreover, confident partial learning has another interesting property, i.e., it is closed under finite unions. The authors then study confident partial learning with respect to oracles, and

[1] The conference proceedings, including preliminary versions of these papers, appeared as "Algorithmic Learning Theory, 23rd International Conference, ALT 2012, Lyon, France, October 29–31, 2012. Proceedings," *Lecture Notes in Artificial Intelligence*, vol. 7568, Springer, 2012.

http://dx.doi.org/10.1016/j.tcs.2014.09.021

obtain some deep results. That is, in addition to the successively fed graph of the target function, the learner has access to an oracle. The second part of the paper combines partial learning with consistency. Since a consistent learner is preferable, these results deserve attention. On the positive site it is shown that every behaviorally correct learnable class is also is essentially class consistently partially learnable. On the other hand, the set of all recursive predicates is not essentially class consistently partially learnable. Finally, it is shown that PA-oracles are sufficient in order to partially learn every recursive function essentially class consistently.

Inductive inference is also widely studied in the context of formal language learning. In this setting the target is a formal language and the information given to the learner may be eventually all strings in the language (positive examples only), all strings over the underlying alphabet which are then marked with respect to their containment in the target language (complete data). Christophe Costa Florêncio and Sicco Verwer in *Regular Inference as Vertex Coloring* study the problem to learn the class of all regular languages from complete data in the limit. The hypothesis space chosen is the set of all deterministic finite automata (abbr. DFA). In this context it is known that it suffices to output in each learning step a minimal DFA that is consistent with all the data seen so far. This is, however, easier said than done, since the problem is known to be $\mathcal{NP}$-complete. Thus the idea is to reduce the learning problem to satisfiability and to exploit the enormous progress made for satisfiability solvers. The approach undertaken previously is to perform this in two steps, i.e., first the learning problem is translated into a graph coloring problem, and second the graph coloring problem obtained is translated into a satisfiability problem. Here the first step included some inequality constraints (requiring the constraint vertices to have a different color) as well as some equality constraints. So, these constraints had to be translated into the resulting satisfiability problem. The main contribution of the present paper is an improvement for the first step that allows for a direct translation of the inference problem into a graph coloring problem. In this way, one can also directly use sophisticated solvers for graph coloring. In this way the authors obtain new complexity bounds and a family of new learning algorithms.

Every learning model specifies the learner, the learning domain, the source of information, the hypothesis space, what background knowledge is available and how it can be used, and finally, the criterion of success. While the learner is always an algorithm, it may also be restricted in one way or another, e.g., by requiring it to be space and/or time efficient. A significant line of work over the past decade studies combinatorial measures of the complexity of learning and/or teaching. In learning problems the VC-dimension is often of central importance. In the framework of algorithmic teaching a helpful teacher chooses an informative sequence of labeled examples and provides them to the learner, with the goal of uniquely specifying the target concept from some a priori concept class of possible target functions. Several different combinatorial parameters related to this framework have been defined and studied, including the worst-case teaching dimension, the average teaching dimension, and the "recursive teaching dimension." In their paper *Algebraic Methods Proving Sauer's Bound for Teaching Complexity* Rahim Samei, Pavel Semukhin, Boting Yang, and Sandra Zilles show that Sauer's Bound can be adjusted to the recursive teaching dimension of the concept. That is, the well-known bound by Sauer on the cardinality of a finite concept class remains valid when the VC-dimension is replaced by the recursive teaching dimension. They further introduce and study classes whose size meets the upper bound and other properties of this measure. Moreover, the proof methods used allow to gain additional insight in the internal structure of recursive teaching plans.

Clustering (the partition of data into meaningful categories) is one of the most widely used techniques in statistical data analysis. A recent trend of research in this field is concerned with so-called perturbation resilience assumptions introduced by Bilu and Linial (2010). This means one asks for conditions under which the optimal clustering does not change when distances are perturbed.

Shalev Ben-David and Lev Reyzin define in *Data Stability in Clustering: A Closer Look* a new notion of stability that is implied by perturbation resilience. They discuss the implications of assuming resilience or stability in the data, where the strength of this resilience or stability is measured by a constant α. The authors show that it is $\mathcal{NP}$-hard to solve clustering instances that are α-perturbation resilient k-median or min-sum instances, for any constant $\alpha < 2$. Furthermore, it is shown that there is a fairly simple algorithm that solves any α-perturbation resilient k-median clustering instance for $\alpha > 5.7$. This algorithm is based on the observation that the data begin to show what is called strict separation, i.e., each point is closer to points in its own cluster than to points in other clusters provided that $\alpha > 5.7$. Moreover, they consider an analogous notion called β-additive perturbation resilience and show that it is $\mathcal{NP}$-hard to solve clustering instance that are β-additive perturbation resilient k-median or min-sum instances, for any constant $\beta < 1/2$.

The area of so-called bandit problems has attracted much attention in the ALT community for more than a decade. Bandit problems form a model of repeated interaction between a learner and a stochastic environment. In its simplest formulation the learner is given a finite number of arms, each associated with an unknown probability distribution with bounded support. Whenever the learner pulls an arm it gets some reward, drawn independently at random according to its associated distribution. The learner's objective is then to maximize the obtained cumulative reward. To do so, a trade-off between testing sufficiently often all the arms (exploration) and pulling more often the seemingly better arms (exploitation)