Theoretical Computer Science 519 (2014) 1–3

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Preface

Guest Editors' introduction

This volume contains nine invited papers based on presentations given at the 22nd International Conference on Algorithmic Learning Theory (ALT 2011), which was held in Espoo, Finland, October 6–8, 2011. After discussions with the programme committee of the conference, the Guest Editors invited the authors of selected conference papers to submit extended and improved versions of their papers for this special issue. These submissions were then reviewed following the normal criteria of *Theoretical Computer Science*.

The ALT conference series is dedicated to the theoretical foundations of machine learning. It was established in Japan in 1990, and shares roots with the series of international workshops on Analogical and Inductive Inference (AII) established in 1986. After co-location in 1994, ALT and AII were merged, and the series has since continued to provide a forum for high quality contributions with a strong theoretical background and scientific interchange in a variety of areas, as can be seen from the breath of topics in this volume.

Online learning and bandit problems

In online learning the training data is presented sequentially and the learner updates its hypothesis after each data point. Sequential decision making tasks naturally require online learning, but it can also be used for computational reasons even when all the training data is available at once but manipulating the entire data set is computationally too expensive.

Of considerable interest is to learn a linear mapping from a d-dimensional Euclidean space to the set of real numbers, i.e., the task of linear prediction. In many practical problems one suspects that the weight vector w^* defining the linear mapping is sparse because not all inputs are relevant. This also implies that the weight vector will have a small 1-norm. How to design algorithms that can exploit this prior information has been the subject of intense investigation in recent years. Sébastien Gerchinovitz and Jia Yuan Yu study this problem in the online learning framework in their paper *Adaptive and optimal online linear regression on ℓ^1-balls*. Their main contribution is showing that the best achievable regret is subject to a phase transition depending on the value of the "intrinsic quantity" $\kappa = \sqrt{T}\|w^*\|_1 X/(2dY)$: For $\kappa < 1$, the best achievable regret scales as $d\kappa$, whereas for $\kappa > 1$ it behaves as $d \ln \kappa$. Here, T is the sample-size, X is the size of the ℓ^∞-ball that the inputs lie in, and Y is a bound on the size of the responses. They also give computationally efficient algorithms that essentially achieve this bound without knowing the values of $\|w^*\|_1$, X, Y or T.

The paper *Combining initial segments of lists* by Manfred K. Warmuth, Wouter M. Koolen, and David P. Helmbold falls broadly within the framework of predicting with expert advice. As an example, suppose you have K different policies for maintaining a memory cache of size N. Different policies work well for different access sequences, and you would like to combine the caches of the K policies dynamically into your own cache of size N such that no matter what the access sequence, you do not incur many more misses than the best of the K policies for that particular sequence. A naïve implementation of the well-known Hedge algorithm for predicting with expert advice is not computationally feasible, since there are roughly N^K ways of picking your combined cache. However, the paper comes up with efficient algorithms based on the special combinatorial structure of this set of N^K combinations. Also some lower bounds and hardness results are presented.

Bandit problems provide the simplest model to study learning in interactive, sequential scenarios with limited feedback: The learner takes actions, resulting in some reward that the learner observes. However, the learner gains no information about the rewards associated with the action not taken, hence the feedback about the environment is limited. The goal of the learner is to achieve as much reward as possible. Performance is measured in terms of the regret, i.e., the loss as compared to using the single best action from the beginning of time.

Most papers concerned with the stochastic version of bandit problem study the expected regret. However, a decision maker might also be interested in the risk, i.e., whether the regret is small not only in expectation, but also with high probability. In their paper *Robustness of stochastic bandit policies* Antoine Salomon and Jean-Yves Audibert show that in the classical setting of finite-armed stochastic bandit problems whether "small risk" policies exist hinges upon whether the total number of plays is known beforehand. That small risk is possible to achieve when this knowledge is available was known

0304-3975/$ – see front matter
http://dx.doi.org/10.1016/j.tcs.2013.09.018

beforehand. The new result is that without this knowledge, no algorithm can achieve small risk except when the class of distributions that can be assigned to the actions is restricted in some way.

Statistical learning theory and kernels

Probably approximately correct (PAC) learning is a computational learning model that has served as the framework for many fundamental results and also inspired a large number of other models. In the basic PAC setting, the unknown quantities are a target concept $f: X \to \{0, 1\}$ and a probability measure P over X. The learner receives a set of labeled examples $(x, f(x))$ and outputs a hypothesis $h: X \to \{0, 1\}$. For given ε and δ, the hypothesis must satisfy with probability at least $1 - \delta$ the property $P(f(x) \neq h(x)) \leqslant \varepsilon$. The analysis of a learning algorithm involves estimating the required number of examples and the computation time in terms of ε, δ and other relevant parameters of the problem.

Co-training under the conditional independence assumption is a model often used in PAC-style analysis of semisupervised learning. In this model, access to a large number of unlabeled examples can lead to a drastic reduction in the required number of labeled examples. The paper *Supervised learning and co-training* by Malte Darnstädt, Hans Ulrich Simon, and Balázs Szörényi poses the question of how much of this reduction is due to the unlabeled examples, and how much would result from the conditional independence assumption even without access to any unlabeled examples. It turns out that under this assumption, the number of labeled examples needed to co-train two concept classes, having VC-dimensions d_1 and d_2, is $O(\sqrt{d_1 d_2}/\varepsilon)$. For small ε this is significantly smaller than the lower bound $\Omega(d/\varepsilon)$ for learning a concept class of VC-dimension d without the conditional independence assumption.

Kernels are a powerful mathematical tool that have gained popularity in machine learning, among other reasons, because they sometimes allow computationally efficient implementation of algorithms that otherwise would require manipulating very high-dimensional feature vectors. Learning algorithms that operate in a high-dimensional feature space often employ some form of margin maximization as a means of avoiding overfitting.

The paper *Accelerated training of max-margin Markov networks with kernels* by Xinhua Zhang, Ankan Saha, and S.V.N. Vishwanathan considers structured output prediction, where in addition to the inputs, also the outputs to be predicted can have a complicated structure. Using the kernel paradigm this can be modeled assuming a joined feature map $\vec{\phi}$ that maps input–output pairs $(\vec{x}, \vec{y})$ into the feature space. One way of proceeding from there, and the one adopted in this paper, is max-margin Markov networks, which leads to a minimization problem where the objective function is convex but not smooth. Non-smoothness rules out some of the faster optimization methods. This paper shows how some known techniques for this kind of optimization can be modified so that they retain their convergence speed, getting to within ε of the optimum in $O(1/\sqrt{\varepsilon})$ iterations, and allow the iteration step to be implemented in an efficient manner that utilizes the structure of the outputs.

Corinna Cortes and Mehryar Mohri in their paper *Domain adaptation and sample bias correction theory and algorithm for regression* consider the situation when the training and test data come from different distributions. We assume there is little or no labeled data about the *target domain* where we actually wish to learn, but unlabeled data is available, as well as labeled data from a different but somehow related *source domain*. Previous work has introduced a notion of discrepancy such that a small discrepancy between the source and target domains allows learning in this scenario. This paper sharpens and simplifies the previous results for a large class of domains related to kernel regression. It then goes on to develop an algorithm for finding a source distribution that minimizes the discrepancy and shows empirically that the new algorithm allows domain adaptation on much larger data sets than previous methods.

Intelligent agents and inductive inference

Intelligent agents need to adapt to their environment to achieve their goals. The problem is made especially difficult by the fact that the actions taken may have long term effects.

Laurent Orseau studies the question of how to design agents which are "knowledge seeking" in the paper titled *Universal knowledge-seeking agents*. The knowledge-seeking agents are those who have a probabilistic world model. In a rather unorthodox manner, the immediate cost suffered by such an agent at some time step is defined as the conditional probability assigned to future outcomes based on the probabilistic world model that the agent chose to use. Arguably, an agent that uses an appropriate world model and that acts so as to minimize the long-term cost will choose actions that allow it to "discard" as many environments as quickly as possible. Performance is compared to the expected total cost suffered by the optimal agent that uses the probability distribution of the true environment as its world model. The main result, which is proved for certain "horizon functions," shows that the so-called AIXI agent's performance converges to the optimal performance provided that the environment is deterministic. A cost defined using the logarithm of the conditional probabilities, i.e., a Shannon-type cost, is also studied.

A discount matrix d is an $\infty \times \infty$ matrix: At time step t an agent using d would "discount" future rewards using the values in the tth column of d. A discount matrix leads to time-consistent behavior if for any environment the optimal policy given some history up to time t uses the same action as the optimal policy that is computed with a column of the discount matrix corresponding to some previous time instance (ties are assumed to be broken in an arbitrary, systematic manner). Tor Lattimore and Marcus Hutter prove a characterization of what discount matrices lead to time consistent discounting in their paper *General time consistent discounting*. They also study the sensitivity of behaviors to perturbations of a time-consistent

discount matrix. Finally, using a game theoretic approach, they show that there is a rational policy even if the discount matrix is time-inconsistent.

A formal language is just a set of strings over some fixed finite alphabet. Inductive inference of formal languages is the study of algorithms that map evidence on a language into hypotheses about it. In general, one studies scenarios in which the sequence of computed hypotheses stabilizes to an accurate and finite description (e.g., a grammar) of the target language.

The paper *Iterative learning from positive data and counters* by Timo Kötzing considers the variation that an iterative learner (a learner who has access only to the last hypothesis and datum) has additionally access to a counter. While it was known that this additional information yields a strictly more powerful learning model, it remained open why and how such a counter augments the learner's power. To answer this question, six different types of a counter are distinguished. In the previously studied case, the counter was incremented in each iteration, i.e., counting from zero to infinity (i.e., $c(i+1) = c(i)+1$). Further possibilities include *strictly increasing* counters (i.e., $c(i+1) > c(i)$), and *increasing and unbounded* (i.e., $c(i+1) \geqslant c(i)$ and the limit inferior of the sequence of counter values is infinity). The paper completely characterizes the relative learning power of iterative learners in dependence on the counter type allowed. It is shown that strict monotonicity and unboundedness are the only properties of the counters that augment the learner power in the iterative setting. The situation changes if other learning criteria are considered. For example, the learner may be required to never abandon a correct hypothesis, or its hypotheses should not depend on the order and the number of repetitions of the examples. It is then shown that for each choice of two different counter types there is a learning criterion that, when augmented with one of the counter types, yields different learnable classes than the same criterion when augmented with the other counter type.

Jyrki Kivinen *
Department of Computer Science, University of Helsinki, P.O. Box 68, FI-00014 University of Helsinki, Finland

Csaba Szepesvári
Department of Computing Science, University of Alberta, Edmonton, Alberta, T6G 2E8, Canada

Thomas Zeugmann
Division of Computer Science, Hokkaido University, N-14, W-9, Sapporo 060-0814, Japan

* Corresponding author.