

## Preface to the special issue

Robert H. Sloan

Published online: 17 August 2014  
© Springer International Publishing Switzerland 2014

The meeting series “International Symposium on Artificial Intelligence and Mathematics” (ISAIM) has been held every other year in Fort Lauderdale, Florida since 1990. The series is remarkable for the breadth of its scope, while still remaining within the meeting place of mathematics with AI. In recent years, the custom has been to have a few focused tracks as well as a general track. ISAIM 2012, the twelfth biennial ISAIM, featured three focused tracks:

- *Computational Social Choice*, organized by Judy Goldsmith and Jörg Rothe. Selected papers from that track made up an earlier special issue of this journal, *Algorithms, Approximation, and Empirical Studies in Behavioral and Computational Social Choice, Annals of Mathematics and Artificial Intelligence* 68/1-3 (2013).
- *Boolean and Pseudo-Boolean Functions*, organized by Endre Boros and Yves Crama.
- *Causal Learning from Complex Data Structures*, organized by David Danks.

Just under half of the papers at ISAIM 2012 were in the “regular” track, and the Boolean and pseudo-Boolean track was the second largest. This special issue is devoted to full length refereed versions of five of the best papers from those two tracks. The remainder of this introduction will attempt to place those papers in context.

We begin with a paper with strong ties to the beginning of the modern era of computational learning theory in the mid-1980s. In 1984, Leslie Valiant introduced the notion of Probably Approximately Correct (PAC) learning. The PAC model requires a classification learning algorithm to use polynomial resources (training data, and often also running time). The algorithm takes in a polynomial number of training examples drawn from a fixed but unknown probability distribution, and labeled according to their classification, typically into one of two classes. The algorithm must output a classifier that, with high probability, classifies most examples correctly, as long as those examples are drawn from that same probability distribution. In other words, the classifier must probably be approximately correct.

---

R. H. Sloan (✉)  
University of Illinois at Chicago, Chicago, IL, USA  
e-mail: sloan@uic.edu

As soon as Valiant published the PAC algorithm, a number of researchers noticed that the initial algorithms given for PAC learning depended crucially on each and every training example being classified correctly and being otherwise noise free. Of course, that is not what happens in practice! So, from the 1980s until today, there has been a line of research on what happens when there is one form or another of noise in the training examples for PAC learning. The paper *PAC-Learning in the Presence of One-sided Classification Noise* by Hans Simon continues this important line of research.

This paper gives us two related results that more or less completely settle the case of one-sided random classification noise. This means that training examples that are in fact negative are each mistakenly labeled positive with some constant probability and positive examples are always labeled correctly. Simon shows that the minimum one-sided disagreement algorithm, that is, finding the function that makes the minimum number of mistakes on the positive (labeled) training examples subject to the constraint of classifying all negative (labeled) training examples negative, uses within a logarithmic factor of the minimum possible amount of training data. Simon's second result is an analysis of how to implement the minimum one-sided disagreement algorithm efficiently for several interesting cases. The implementation will in fact work even in the case of adversarial one-sided noise.

The next two papers in this special issue concern the computational complexity of problems from areas of mathematics of particular importance to computer science in general, and AI in particular. The first concerns linear algebra, and the second concerns Horn formulas.

The second paper in this volume is a computational complexity paper about a linear algebra problem with a bit of the feeling of a predicate logic problem. *On Quantified Linear Implications* by Pavlos Eirinakis, Salvatore Ruggieri, K. Subramani, and Piotr Wojciechowski, is about the computational complexity of *quantified* linear programming, that is, linear systems that come with a quantifier string over the variables. Quantified linear programming can also be viewed as a quantified implication between two systems of linear equations; hence the title of the paper. The paper studies the computational complexity of the satisfiability decision problem. In particular, the paper provides a variety of new complexity-theoretic upper and lower bounds for these problems, considering zero, one, and two alternations of quantifiers. The paper should also be of interest to those looking at quantified linear algebra for the first time, because it gives a good overview of the subject.

The third paper in this volume is the paper drawn from the Boolean and pseudo-Boolean track of ISAIM 2012 and concerns Horn functions, which have long played a key role in AI. For most interesting classes of Boolean functions, questions concerning the computational complexity of minimizing the size of a Boolean formula to represent a particular function has been more or less settled. Generally the minimization problem is intractable. For Horn formulas, however, while we have many results, we also still have some important open problems. *Hardness Results for Approximate Pure Horn CNF Formulae Minimization* by Endre Boros and Aritanan Gruber takes on a couple of these. A *pure Horn* formula (called a *definite Horn* formula by many other authors) is a conjunction of clauses where every clause contains exactly one positive literal.

This paper continues the study of two different measures of formula size: minimizing the number of clauses in a formula, and minimizing the total number literals in a formula. For example, the formula  $(\bar{x}_1 \vee x_2) \wedge (\bar{x}_1 \vee \bar{x}_3 \vee x_4)$  has two clauses and five literals. The paper gives us new stronger non-approximation results for both measures.

Finally, the fourth and fifth papers in this special issue concern domain adaptation. Domain adaptation is a technique that has arisen from the applied side of machine learning. A frequent problem for real-world machine learning is that one receives labeled training data from, and thus one builds models for, some fixed domain with a particular probability dis-

tribution over that domain, but wants to deploy it on a different target domain. The learning algorithm also can use large quantities of *unlabeled* data from the target domain. Imagine, for example, an English-language speech recognition system trained on people with more or less standard American accents that needs to perform on a wide variety of different accents.

In *Robust Domain Adaptation* by Yishay Mansour and Mariano Schain, the authors introduce a measure of difference between the probability distributions of the labeled training data and the performance data that they call  $\lambda$ -shift, and prove some results about algorithmic robustness in terms of it, in particular for an SVM-inspired algorithm.

The intended fifth paper of the special issue, *Domain Adaptation—Can Quantity Compensate for Quality?* by Shai Ben-David and Ruth Urner, appeared in the previous volume of *Annals of Mathematics and Artificial Intelligence* 70/3 (2014), 185–202. It is a theoretical paper, giving conditions on the training and test probability distributions where one can prove that a Nearest Algorithm succeeds in domain adaptation.

Robert H. Sloan  
Chicago, USA  
July 2014