Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE @ DIRECT®

ELSEVIER

Theoretical Computer Science 298 (2003) 1–4

Theoretical
Computer Science

www.elsevier.com/locate/tcs

Foreword

This special issue of *Theoretical Computer Science* is dedicated to the 60th birthday of Professor Setsuo Arikawa. The authors of the papers included have been invited by the Special Issue Editors to submit completed versions of their work related to Professor Arikawa's research for this Special Issue.

Professor Setsuo Arikawa has had a long and distinguished career. He has personally made significant contributions in the areas of pattern matching, B-tree's, model inference, formal systems and inductive inference. All of these topics relate in some way to his life long commitment to understanding and automating the learning process. During this time he supervised more than 20 doctoral students.

Later in his career, Professor Arikawa took on several administrative tasks that enhanced the quality of the professional environment for those around him. He became the Director of the Computing Center and then the Library at Kyushu University. In 1990, he led the efforts to start a very successful series of meetings on Algorithmic Learning Theory. With extraordinary vision, he saw that learning was really part of the larger paradigm of discovery. Professor Arikawa was one of the principal organizers of a trans Japan research project in Discovery Science that spawned an annual international meeting starting in 1998.

These activities have earned Professor Arikawa the respect and admiration of colleagues around the world. He was fond of saying "if we recognize the importance or goodness in the works or ideas by other people, it is not enough for us to recognize it, we should state our recognition as early as possible." While we hope it is still premature to make comments about Professor Arikawa's entire career, we want to follow his lead and recognize his broad, outstanding contributions to the pursuit of understanding and automating the learning process.

Let us continue by shortly introducing the papers contained in this special issue. The first group consists of three articles all dealing with problems arising in the study of learning logic programs within the setting of discovery science.

Yamamoto takes a fresh look at different methods proposed to resolve the hypothesis finding problem. He puts all these methods on a common general ground by using upward refinement and residue hypotheses. This combination is then shown to be a complete method for any hypothesis finding problem in clausal logic.

In their paper, Sakamoto et al. consider the problem of learning elementary formal systems (EFS) in the query learning model initiated by Angluin. Originally introduced by Smyllan (Theory of Formal Systems, Princeton University Press, 1961)

PII: S0304-3975(02)00415-2

to study recursive function theory over strings, later EFS turned out to be similar to logic programs operating on strings and having a distinguished unary predicate. The true atoms for this predicate define a formal language. Arikawa and his co-workers studied EFS from many different points of view during a long period of time, thereby obtaining many deep and fundamental results. In particular, several important language classes describable by EFS have been characterized and their learnability has been shown. Sakamoto et al. extend this work by studying the various classes contained in the hierarchy of hereditary EFS with respect to their query complexity.

Lange et al. derive the motivation for their research from the design of a system for discovering knowledge in semi-structured documents. However, some of the information discovered may not be of interest to the user, and should be thus marked as such. Now, in order to improve the systems discovery capabilities, one has to perform further learning steps using this particular kind of negative information. Therefore, hypotheses cannot be expressed by using EFS. Instead, they had to incorporate negation into EFS resulting in advanced elementary formal systems. The extension is done along the lines of stratified logic programs. Learnability is studied and compared to the case before the extension. In Gold's classical paradigm of learning in the limit, some positive results do not transfer to the extended systems, but in Valiant's PAC model the main known positive result is shown to hold for the extended systems, too.

Lange and Nessel introduce the notion of decision lists over regular patterns. Regular pattern languages are a subclass of Angluin's pattern languages (every variable is allowed to occur at most once) and consist themselves a subclass of all regular languages. Moreover, regular pattern languages are also a subclass of hereditary EFS. On the other hand, decision lists have been introduced by Rivest for generalizing k-DNF and k-CNF. So, it is only natural to combine these approaches. And indeed, the new formalism provides a strict extension of even regular erasing pattern languages. Moreover, precise answers are obtained to the question which subclasses of the new decision lists are learnable and which are not.

The following three papers belong to the subfield of learning theory that shares the notion of refutation. This notion has been introduced by Arikawa and Mukouchi (Theoret. Comput. Sci. 137(1) (1995) 53–84) and was partly motivated by the design of automatic discovery systems. For such systems the choice of the hypothesis space is a critical parameter. The basic scenario can be described as follows. The learner is given a hypothesis space of uniformly recursive concepts in advance. Whenever the target concept can be correctly described by a member of this hypothesis class, then the learner has to identify it in the limit. If, however, the learner is fed data of a target concept having no correct description within the hypothesis class given, then the learner has to refute the whole hypothesis class after a finite amount of time by outputting a special refutation symbol and stopping the learning process. Thus, within the model of learning refutably, the learner either identifies a target concept or itself indicates its inability to do so.

This original approach is relaxed by Mukouchi and Sato. They consider a weakened correctness criterion and allow noisy examples. In their new model, the learner must succeed to learn a target concept if it has a (weak) k-neighbor in the hypothesis class;

otherwise it has again to refute the whole hypothesis class. Here a (weak) k-neighbor is defined in terms of a distance over strings.

Next, several variations of learning refutably are investigated by Jain et al. In this paper, the targets are drawn from the set of all recursive functions and an acceptable programming system is provided as hypothesis class. Hence, it is no longer appropriate to refute the whole hypothesis class, since it contains a correct description for every target. Nevertheless, the learner may not be able to successfully complete its learning task. This can be indicated by either outputting the refuting symbol and stopping the learning process, or by converging to the refutation symbol, or by outputting the refutation symbol infinitely often. All these models of learning refutably are studied, related to one another as well as their combination with other, previously studied learning models within the setting of inductive inference.

Merkle and Stephan introduce the notion of refutation in the limit. Moreover, they extend the notion of learning from positive data. Now, the data sequences are sequences of first-order sentences describing the target. Several new and interesting hierarchies are then established.

The next two papers deal with the complexity of learning problems. Yokomori considers a subclass of simple deterministic grammars the so-called very simple grammars. The resulting language class is incomparable to the class of regular languages. Then a learning algorithm is presented for identifying the class of simple grammars in the limit from positive data that achieves polynomial update time and makes at most polynomially many prediction errors. This result is important, since none of the well-known language classes in the Chomsky hierarchy is learnable from positive data at all.

Okamoto and Yugami investigate the average-case complexity of instance-based learning algorithms. A variant of k-nearest neighbor classifier is employed for the relevant algorithms and analyzed. Three types of noise are distinguished and the expected classification accuracy of the k-nearest neighbor classifier is expressed as a function of domain characteristics. This work completes earlier research undertaken by the same authors who have been supervised by Professor Arikawa as Ph.D. students.

Moreover, during his academic career Professor Arikawa always emphasized bridging theory and practice. He and his co-workers have been among the first applying results from learning theory to problems arising in molecular biology. In their paper, T. Akutsu et al. continue to study the identification of genetic networks.

Finally, the last paper included in this special issue deals with a current line of Professor Arikawa's research, i.e., the study of pattern matching algorithms. The basic problem here is to find all occurrences of a given pattern in a given text. In practice, documents are often stored in compressed form. So, the problem arises whether or not pattern matching can be performed without decompressing the documents. T. Kida et al. present the Collage System developed at Kyushu University which can be best described as a unifying framework for compressed pattern matching.

We would like to express our immense gratitude to the referees for their fine reports and their efficient work. Moreover, we are particularly thankful to Maurice Nivat and

Giorgio Aussiello for their support and for giving us the opportunity to compile this special issue.

Guest Editors
Takeshi Shinohara,
Carl H. Smith,
Thomas Zeugmann

Iizuka, Fukuoka, Japan
College Park, MD, USA
Lübeck, Germany