

Guest editorial

Rui Camacho · Ross King · Ashwin Srinivasan

Published online: 24 July 2006
Springer Science + Business Media, LLC 2006

The papers in this issue are snapshots of ILP at the age of fourteen. Adolescence, as we all remember, was a time when change was normal and we knew everything. We are reminded here of Mark Twain's recollection: "When I was a boy of fourteen, my father was so ignorant I could hardly stand to have the old man around. But when I got to be twenty-one, I was astonished at how much he had learned in seven years". And so it is with ILP. Visiting friends and relatives would be surprised at the changes: some of which is captured in the papers here. In them we see a face of ILP in which systematic search for rules has been replaced by randomised ones; first-order trees have been replaced by random forests of first-order trees; observations no longer refer to the target concept being learnt; theories are constructed using very large ensembles of clauses, and agent programs are constructed from observing an expert. Compare this against the ILP that constructed a small number of definite clauses, each with about 4 literals, using a data set of no more than a few hundred examples of the target concept. And, as with Twain, we can expect to be astonished over the next few years.

As the character of ILP develops over the next few years, there are some traits that are intriguing, and some that require a pause for thought. Amongst the former is a trend towards the combination of statistical and relational information. We expect this will allow ILP to be applied across a much wider range of problems than previously attempted. Also interesting is the use of ILP as a general-purpose technique for constructing new features. This has repeatedly been shown to be a very good way to use an ILP system. There are two characteristics which may be viewed as a cause for concern by some. First, comprehensibility of theories constructed appears to be no longer a feature of all ILP-constructed theories. This

R. Camacho (✉)
LIACC/FEUP University of Porto, Portugal
e-mail: rcamacho@fe.up.pt

R. King
Department of Computer Science University of Wales, Aberystwyth Wales, United Kingdom
e-mail: rdk@aber.ac.uk

A. Srinivasan
IBM India Research Laboratory Indian Institute of Technology New Delhi, India
e-mail: ashwin.srinivasan@in.ibm.com

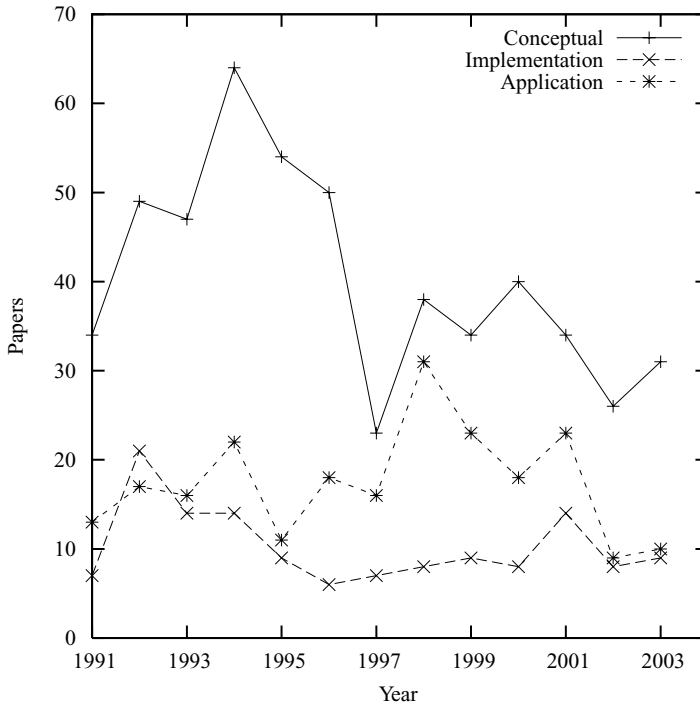


Fig. 1 A rough categorisation of ILP papers published from 1991. These are estimates obtained from a manual classification by one of us (A.S.) of the ILPNet2 bibliographic entries. The average values over the period are: 38 (Conceptual), 9 (Implementation) and 17 (Application)

is specially so with ILP systems capable of constructing and using very large numbers of clauses, or combining probabilistic and logical information in complex ways. Second, there appears to be a trend towards fewer conceptual and application-oriented papers (see Fig. 1). If these estimates are reflective of reality, then the problem needs addressing. We are pleased to see that the ILP application presented in this issue has some substantially different features to many in the recent past and it could take the place of the “mutagenesis” problem as the new *Drosophila* for ILP.

Now for the clerical details. This volume contains five papers based on original ones presented at the 14th International Conference on Inductive Logic Programming. All the papers were significantly re-worked and extended with respect to their original version that appear in the ILP 2004 Proceedings. ILP 2004 was held in Porto, Portugal, in September 2004. This annual meeting of ILP continues to be the premier forum for presenting the most recent and exciting work in the field. Six invited talks—three from fields outside ILP, but nevertheless highly relevant to it—and 20 full presentations formed the nucleus of the conference.

Van Assche et al. show how the concept of random forests can improve relational decision trees. They address and propose solutions for the problem of introducing aggregate functions. The resulting first order random forest induction algorithm was implemented and experimentally evaluated on a variety of datasets. The results indicate that first order random forests with complex aggregates are an efficient and effective approach towards learning relational classifiers that involve aggregates over complex selections.

Developing efficient methods to search the hypothesis space continues to be an important issue in ILP. Železný et al. demonstrate the benefits of randomising and restarting ILP search procedure. They show that significant speedups can be obtained using randomised restarts by comparing five search algorithms operating on two principally different classes of ILP problems.

In the paper by Tamaddoni-Nezhad et al. a problem that is increasingly important in biological applications is addressed. Observations do not necessarily refer to the target predicate and background knowledge cannot be assumed to be complete. Using a combination of abduction and induction an ILP program is used to construct a model by completing definitions for background predicates.

The paper by Goadrich et al. addresses the problem of dealing with large, highly-skewed datasets in ILP. They propose a randomized search method that collects good clauses from a broad spectrum of points along the recall dimension in recall-precision curves. The method is evaluated on information extraction tasks that typically involve many more negative examples than positive examples. Those at Porto in September 2004 may recall that the conference submission of this paper won the “Best Student Paper” award that year.

The paper by Könik et al. describes a method of relational learning by observation. In this, a learner automatically creates cognitive agent programs that model expert task performance in complex dynamic domains. Using observed behavior and goal annotations of an expert as the primary input, their method interprets them in the context of background knowledge, and returns an agent program that behaves similar to the expert.

ILP is still a young field, and one that will change much more before it achieves its true potential. But we are amongst those that believe that this potential is great: in the end, few other machine learning techniques have inherited the combination found in ILP of representational power, mathematical foundations and explicit use of background knowledge. The papers published in this special issue are all steps in the journey of growing up.