Electronic Transactions on Artificial Intelligence
Vol. 5 (2001), Section B: pp. *1-36*
http://www.ep.liu.se/ej/etai/2001/001/

Refereed Article

C.H. Bryant, S.H. Muggleton, S.G. Oliver, D.B. Kell, P. Reiser, and R.D. King:

Combining Inductive Logic Programming, Active Learning and Robotics to Discover the Function of Genes

Authors' affiliations:
CHB, SHM: University of York, U.K.
SGO: University of Manchester, U.K.
DBK, PR, RDK: University of Wales, U.K.

Abstract: *The paper is addressed to AI workers with an interest in biomolecular genetics and also to biomolecular geneticists interested in what AI tools may do for them. The authors are engaged in a collaborative enterprise aimed at partially automating some aspects of scientific work. These aspects include the processes of forming hypotheses, devising trials to discriminate between these competing hypotheses, physically performing these trials and then using the results of these trials to converge upon an accurate hypothesis. As a potential component of the reasoning carried out by an "artificial scientist" this paper describes ASE-Progol, an Active Learning system which uses Inductive Logic Programming to construct hypothesised first-order theories and uses a CART-like algorithm to select trials for eliminating ILP derived hypotheses. In simulated yeast growth tests ASE-Progol was used to rediscover how genes participate in the aromatic amino acid pathway of Saccharomyces cerevisiae. The cost of the chemicals consumed in converging upon a hypothesis with an accuracy of around* 88% *was reduced by five orders of magnitude when trials were selected by ASE-Progol rather than being sampled at random. While the naive strategy of always choosing the cheapest trial from the set of candidate trials led to lower cumulative costs than ASE-Progol, both the naive strategy and the random strategy took significantly longer to converge upon a final hypothesis than ASE-Progol. For example to reach an accuracy of* 80%, *ASE-Progol required 4 days while random sampling required 6 days and the naive strategy required 10 days.*

Publication and review history

This work was first presented at the Machine Intelligence 18 Workshop, held on September 19-21, 2001 at the University of York, UK. Attendance at this workshop was based on invitations. Speakers at the workshop were invited to submit their articles for consideration in the forthcoming workshop volume. The subsequent publication and review history for the present article was as follows.

1. First versions *published* by Linköping University Electronic Press on in August - October, 2001 and permanently available at
   http://www.ep.liu.se/ea/cis/2001/012/
2. First version *accepted* after due refereeing and according to scientific journal standards by the Electronic Transactions on Artificial Intelligence (ETAI) on 21.9.2001. Authors were invited to revise their articles, if they so desired, based on comments by the reviewers.
3. Revised version *accepted* on 4.12.2001.
4. Revised version *published* by Linköping University Electronic Press on 28.2.2002 and permanently available at the same URL as mentioned above.
5. Revised version appears in Electronic Transactions on Artificial Intellligence, Volume 5 (2001), section B, pages 1–36. The section and the annual volume are made permanently available at
   http://www.ep.liu.se/ej/etai/
6. Paper editions of the ETAI issue and the ETAI volume containing this artice *republished* by the Royal Swedish Academy of Sciences.

The review policy and the quality requirements for acceptance are documented at http://www.etaij.org/info/

Article maintenance

The Electronic Transactions on Artificial Intelligence maintains an Article Interaction Page for this article at
http://www.etaij.org/ra/mi/018/
Besides the publication and review history, it contains links for contacting the author(s), as well as to amendments and other related information for the article. It is intended to keep these links up-to-date in the future.

Copyright conditions