Electronic Transactions on Artificial Intelligence
Vol. 5 (2001), Section B: pp. 245-274
http://www.ep.liu.se/ej/etai/2001/012/

Refereed Article

A.P. Cootes, S.H. Muggleton, R.B. Greaves, and M.J. Sternberg:

Automatic determination of protein fold signatures from structured superposition

Authors' affiliations:
AC, MS: Imperial Cancer Research Fund, U.K.
SM, RG: University of York, U.K.
AC, SM, MS: Imperial College of Science, Technology and Medicine

Abstract: *It remains unclear what principles underlie a protein sequence/ structure adopting a given fold. Local properties such as the arrangement of secondary structure elements adjacent in sequence or global properties such as the total number of secondary structure elements may act as a constraint on the type of fold that a protein can adopt. Such constraints might be considered "signatures" of a given fold and their identification would be useful for the classification of protein structure. Inductive Logic Programming (ILP) has been applied to the problem of automatic identification of structural signatures. The signatures generated by ILP can then be both readily interpreted by a protein structure expert and tested for their accuracy. A previous application of ILP to this problem indicated that large insertions/deletions in proteins are an obstacle to learning rules that effectively discriminate between positive and negative examples of a given fold. Here, we apply an ILP learning scheme that reduces this problem by employing the structural superposition of protein domains with similar folds. This was done in three basic steps. Firstly, a multiple alignment of domains was generated for each type of fold studied. Secondly, the alignment was used to determine the secondary structure elements in each of those domains that can be considered equivalent to one another (the "core" elements of that fold). Thirdly, an ILP learning experiment was conducted to learn rules defining a fold in terms of those core elements.*

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/2001/026/
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 245–274. 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/032/
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.