Linköping Electronic Articles in
Computer and Information Science
Vol. 5(2000): nr 39

Use of Inductive Logic Programming to Learn Principles of Protein Structure

Marcel Turcotte

Stephen H. Muggleton

Michael J. E. Sternberg

Linköping University Electronic Press
Linköping, Sweden

http://www.ep.liu.se/ea/cis/2000/039/

Published on December 21, 2000 by
Linköping University Electronic Press
581 83 Linköping, Sweden

Linköping Electronic Articles in
Computer and Information Science
ISSN 1401-9841
Series editor: Erik Sandewall

Recommended citation:
<Author>. <Title>. Linköping Electronic Articles in Computer and Information Science, Vol. 5(2000): nr 39. http://www.ep.liu.se/ea/cis/2000/039/. *December 21, 2000.*

This URL will also contain a link to the author's home page.

Abstract

Inductive logic programming (ILP) has been applied to learn rules which characterize protein folds. Several representations for the background set have been explored and the results have been interpreted in their biological context. In this paper, we present new results obtained with a background set containing information about protein topology. The new rules are more descriptive than the previous ones, *i.e.* where previous rules represented local motifs, often associated with functional regions, the new rules represent more complete descriptions, often similar to the descriptions found in SCOP. Cross-validation experiments were conducted for the 20 most populated folds. The overall cross-validated accuracy was found to be 75.1 ± 1.6 % for the more limited background knowledge, and 82.1 ± 1.4 % with additional information.

Authors' addresses

Marcel Turcotte and **Michael J. E. Sternberg**
Imperical Cancer Research Fund
Biomolecular Modelling Laboratory
P.O.Box 123
London, WC2A 3PX
United Kingdom

Stephen H. Muggleton
University of York
Department of Computer Science
Heslington, York YO1 5DD
United Kingdom

For technical reasons, this article is stored as two separate computer files, labelled "Part I" and "Part II", respectively. Please replace the present page by a printout of the "Part II" file.