SpringerBriefs in Computer Science

More information about this series at http://www.springer.com/series/10028

Jinbo Xu · Sheng Wang
Jianzhu Ma

Protein Homology Detection Through Alignment of Markov Random Fields

Using MRFalign

Springer

Jinbo Xu
Toyota Technological Institute
Chicago, IL
USA

Jianzhu Ma
Toyota Technological Institute
Chicago, IL
USA

Sheng Wang
Toyota Technological Institute
Chicago, IL
USA

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-319-14913-4 ISBN 978-3-319-14914-1 (eBook)
DOI 10.1007/978-3-319-14914-1

Library of Congress Control Number: 2014960093

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

This short book is derived from our paper entitled "MRFalign: Protein Homology Detection Through Alignment of Markov Random Fields," which won the best paper award at a premier computational biology conference RECOMB2014 and also appeared at PLoS Computational Biology. The intended audience consists of students and researchers involved in developing computational methods for biological sequence analysis and protein structure and functional prediction, those who use sequence analysis tools to study biology problems, and those who would like to have a general idea about protein homology detection and fold recognition. We hope that the new Markov Random Fields (MRF) method described in this book will intrigue further study of protein homology detection and fold recognition. We also hope that the tool described in this book will be helpful for readers with biology backgrounds who need to quantify and analyze protein sequences to answer interesting questions.

This book covers sequence-based protein homology detection, a fundamental and challenging bioinformatics problem with a variety of real-world applications. The book first surveys a few popular homology detection methods such as the position-specific scoring matrix (PSSM) or Hidden Markov Model (HMM)-based methods and then is devoted to a novel MRF-based method which was recently developed by our group. Compared to HMM and PSSM, MRF can model long-range residue–residue interaction and thus, MRF-based methods are much more sensitive than HMM- and PSSM-based methods for remote homolog detection and fold recognition. This book also describes the installation, usage, and result interpretation of our programs implementing the MRF-based method.

The book is organized into four chapters. Chapter 1 describes the background and surveys the existing popular methods of homology detection and fold recognition. Chapter 2 describes a novel MRF-based method for homology detection and fold recognition. In particular, it covers how to build an MRF model for a protein sequence, how to score the similarity of two MRF models, and how to generate an MRF–MRF alignment optimizing the scoring function. Chapter 3 is devoted to the software implementing the ideas presented in Chap. 2, covering installation, usage, and result interpretation of the software. Chapter 4 describes the experimental

results of our MRF-based method for homology detection and fold recognition. Finally, conclusions are drawn in the last part of the book.

We are indebted to a few Ph.D. students in our group such as Dr. Jian Peng (now a faculty member at UIUC Computer Science Department), Dr. Feng Zhao, and Mr. Zhiyong Wang. We are also thankful to Dr. Söding, who developed the popular HHpred program for homology detection. It is their previous excellent work that leads to the MRF-based method described in this book.

Contents