SpringerBriefs in Computer Science

For further volumes:
http://www.springer.com/series/10028

Wyatt Travis Clark

Information-Theoretic Evaluation for Computational Biomedical Ontologies

Springer

Wyatt Travis Clark
Department of Molecular Biophysics
and Biochemistry
Yale University
New Haven, CT
USA

ISSN 2191-5768 ISSN 2191-5776 (electronic)
ISBN 978-3-319-04137-7 ISBN 978-3-319-04138-4 (eBook)
DOI 10.1007/978-3-319-04138-4
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013957360

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The development of effective methods for the prediction of ontological annotations is an important goal in computational biology, with protein function prediction and disease gene prioritization gaining wide recognition. While various algorithms have been proposed for these tasks, evaluating their performance is difficult due to problems caused both by the structure of biomedical ontologies and biased or incomplete experimental annotations of genes and gene products. In this work, we propose an information-theoretic framework to evaluate the performance of computational protein function prediction. We use a Bayesian network, structured according to the underlying ontology, to model the prior probability of a protein's function. We then define two concepts, misinformation and remaining uncertainty, that can be seen as information-theoretic analogs of precision and recall. Finally, we propose a single statistic, referred to as semantic distance, that can be used to rank classification models. We evaluate our approach by analyzing the performance of three protein function predictors of Gene Ontology terms and provide evidence that we address several weaknesses of currently used metrics. We believe this framework provides valuable and useful insights into the performance of protein function prediction tools.

Contents