REZARTA ISLAMAJ DOĞAN

Computer Science Department
AV Williams, Rm 3228
University of Maryland
College Park, MD 20742

(301) 405-2717 (UMD)
(301) 594-6538 (NIH)
rezarta@cs.umd.edu
http://www.cs.umd.edu/∼rezarta

Research Interests

I am interested in computational biology problems. I work on novel algorithms to analyze genomic data using statistical methods and machine learning. I have worked on constructing, selecting and discovering appropriate motifs to model biological signals such as splice sites for accurate classification and prediction.

Education

UNIVERSITY OF MARYLAND — Ph.D. Computer Science
Ph.D. thesis: "New machine-learning techniques for sequence-data analysis with applications to splice-site prediction" — Fall 2007
Advisor: Lise Getoor

UNIVERSITY OF MARYLAND — M.S. Computer Science
M.S. scholarly paper: "Identification of protein-coding regions" — May 2003
Advisor: Lise Getoor

BOSPHORUS UNIVERSITY, ISTANBUL — B.Sc. Computer Engineering
B.S. thesis: "Three-dimensional representation of amino-acid characteristics" — June 2000
Advisors: Ethem Alpaydın and Uğur Sezerman
graduated with highest honors

Research and Professional Experience

DISSERTATION RESEARCH

The set of attributes or features selected to model an entity is very important for correct classification. For the problem of sequence classification, I have built an integrated process, which I refer to as *feature generation*. This algorithm allows the user to construct interesting features out of basic elements and to search effectively a large space of potential features. I have applied this approach to the problem of splice-site prediction. Predictive models for acceptor and donor site for two different organisms, human and Arabidopsis Thaliana, have achieved significant improvements in accuracy over existing, state-of-the-art approaches. In each case, the identified sets of features were used to discover biologically interesting motifs. An easy-to-use website, *SplicePort*, www.spliceport.org, can be used to predict new splice sites from user-input sequences, and to browse the whole collection of features for interesting signals. I have expanded the algorithm to construct more complex features, that also capture the three-dimensional characteristics of the genomic sequence. The new features improve the predictive power of the model, and also contribute to the discovery of new biological properties. I am interested in applying this approach for the discovery and prediction of other interesting signals, as well as expanding my research into other challenging computational problems.

PRE-DOCTORAL FELLOW — NCBI/NLM/NIH
June, 2002 - present — Bethesda, MD
I am working with Dr. W. John Wilbur (NCBI) and Dr. Lise Getoor (UMD). My research focuses on pre-mRNA sequence analysis and splice-site prediction improvement through machine-learning approaches.

TEACHING ASSISTANT — University of Maryland
September, 2000 - May, 2002 — College Park, MD
I assisted for Software Engineering and Database Design. I conducted discussion sessions and gave lectures, prepared assignments, prepared and graded term projects, graded exams and assisted students.

TEACHING ASSISTANT — Bosphorus University
September, 1999 - June, 2000 — Istanbul, Turkey
I prepared the course outline and presented material for two terms of Introduction to Pascal and C programming courses. I conducted discussion sections and prepared assignments. I graded assignments, term projects and exams. I helped students individually during office hours and directed them in group work during lab hours. I organized a competition challenge for term projects.

INTERN Software Engineer — SuperOnline
June 1999 - September 1999 — Istanbul, Turkey
I implemented several parts of an on-going project using data-driven dynamic HTML generation with Sybase Power Dynamo over a Sybase database.

INTERN Application Developer — IDB
August 1998 - September 1998 — Istanbul, Turkey

INTERN Programmer — GigaByte
June 1998 - July 1998 — Istanbul, Turkey

Publications

Book Chapters

[1] "A feature generation algorithm with applications to biological sequence classification" with Lise Getoor and W. John Wilbur, Chapter in *Computational Methods of Feature Selection,* Huan Liu and Hiroshi Motoda editors (2007).

Journal Papers

[1] "Features generated for computational splice-site prediction correspond to functional elements" with Lise Getoor, W. John Wilbur and Stephen M. Mount, (submitted 2007).

[2] "SplicePort: an interactive splice-site analysis tool" with Lise Getoor, W. John Wilbur and Stephen M. Mount, *Nucleic Acids Research,* (June 2007).

[3] "Structural footprinting in protein structure comparison: the impact of structural fragments." with Elena Zotenko, W. John Wilbur, Diane P. O'Leary and Teresa M. Przytycka, (submitted 2007).

Refereed Conferences

[1] "Feature generation algorithm: with application to splice-site prediction," with Lise Getoor and W. John Wilbur, *Proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases,* Berlin, Germany (September 2006).

[2] "A feature generation algorithm for sequences with application to splice-site prediction" with Lise Getoor and W. John Wilbur, *International Workshop on Feature Selection for Data Mining: Interfacing Machine Learning and Statistics,* Bethesda, Maryland (April 2006).

[3] "Three dimensional representation of amino acid characteristics" with Ugur Sezerman and Ethem Alpaydin, *23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Istanbul, Turkey (October 2001).

Talks and Presentations

[1] "Analysis of splicing motifs" (**invited talk**) *Splicing Regulator Motifs Workshop,* Erlangen, Germany (September 2006).

[2] "The proximity effect on feature construction for splice-site finding" *NIH Graduate Student Research Symposium,* Bethesda, Maryland (May 2006).

[3] "A machine learning solution for splice-site prediction" *TASSA Annual Conference,* Philadelphia, PA (March 2006).

[4] "Feature generation for sequences with application to splice-site prediction" *NIH Research Festival,* Bethesda, Maryland (October 2005).

[5] "Finding acceptor splice sites with AdaBoost" *NIH Second Annual Graduate Student Research Symposium,* Bethesda, Maryland (April 2005).

[6] "Classification of acceptor splice sites using AdaBoost with decision trees" *16th Annual Genomic Sequencing and Analysis Conference,* Washington DC (September 2004).

[7] "Identification of internal coding regions in mRNA sequences" *NIH First Annual Graduate Student Research Symposium,* Bethesda, Maryland (April 2004).

Honors and Awards

Faculty Horizons	June 2006
Graduated 5th, Computer Science, Bosphorus(Bogazici) University	June 2000
Dean's Honor List	1996-2000
Turkish Prime Ministry Fellowship	1997-2000
Rumeli Foundation Fellowship	1997-2000
Ranked 1st in National Mathematical Olimpiad, Albania	March 1996
Honorable Mention in 13th Balkan Mathematical Olympiad, Romania	April 1996
37th International Mathematical Olympiad, India	July 1996