



The International Association for
Pattern Recognition
Technical Committee 20

Pattern Recognition for Bioinformatics

Interim Report

December 2004 – May 2005

Raj Acharya, Chair, IAPR-TC20
Professor and Head, Computer Science and Engineering
The Pennsylvania State University
220 Pond Laboratory
University Park, PA 16802
U.S.A.
Tel: +1 814 8659505, Fax: +1 814 8653176
Email: acharya@cse.psu.edu
URL: www.cse.psu.edu/~acharya

1 TC-20 Aims and Scope

1.1 Introduction

The past decade has witnessed an explosion in the amount and complexity of biological data such as DNA and protein sequences, gene and protein expressions, structures, pathways, genetic information, biomedical text data, and molecular images. Although the analyses of these data involve pattern recognition and data mining, the novel and efficient data analyses techniques have not realized their true potential. Bioinformatics can be viewed as a field of discovering knowledge from life sciences data with the aid of Information Technology, to find answers to unresolved problems in biology. The pattern recognition methods have various applications and the potential in discovering biological knowledge from life sciences data.

The information stored in DNA, a chain of four nucleotides (A, T, G, and C), is first converted to mRNA through the process of *transcription* and then converted to the functional form of life, proteins, through the process of *translation*. The initiation of translation or transcription process depends on the presence of specific patterns of DNA, RNA, and motifs. Research on detecting specific patterns of DNA sequences such as genes, protein coding regions, and promoters, leads to the better understanding of molecular level function of the cell. Comparative genomics focuses on comparisons across the genomes to find conserved patterns over the evolution, which should possess some functional significance. Construction of evolutionary trees is useful to know how genome and proteome are evolved over all species by ways of a complete library of motifs and genes.

A protein's functionality or its interaction with another protein is mainly determined by its 3-D structure. Prediction of protein's 3-D structure from its 1-D amino-acid sequence remains an open problem in structural genomics; protein-protein interactions determine essential functions in living cells. Computational modeling and visualization tools of 3-D structures of proteins help biologists to infer cellular activities.

The challenge in functional genomics is to analyze gene expression data accumulated by microarray techniques to discover the clusters of co-regulated genes and thereby gene regulatory networks, leading to the understanding of regulatory mechanisms of genes and pathways. Molecular imaging provides techniques for in vivo sensing and imaging of molecular events, which measure biological processes in living organism at the molecular and cellular level. The techniques to fuse and integrate different kinds of information derived from different life science data are yet to realize its full potential.

The ever expanding knowledge of biomedical and phenotype data, combined with genotypes, is becoming difficult to be analyzed by traditional text-based methods. Advanced data mining techniques, where the use of ontologies for constructing precise descriptors of medical concepts and procedures, are required in the field of medical informatics. The vast amount of biological literature is posing new challenges in the field of text mining. These text mining techniques along with the aid of information fusion methods could help find pathways and interaction networks.

With compared to the recent emergence of bioinformatics problems, several pattern recognition tools have evolved over the last three decades to discover specific patterns embedded in data. These techniques include statistical, structural, syntactic, neural networks, computational intelligence, graphical models, data mining techniques, and their hybrids. TC20 has interests in using these techniques in pattern recognition problems in the following areas:

- Computational and comparative genomics
- Gene expression analysis and functional genomics
- Alignment of sequences: DNA, protein, structures, etc.
- Phylogenic analysis of species, sequences, structures
- Structural genomics and proteomics
- Functional, molecular, and cellular imaging
- Data mining, text mining, data integration, and visualization
- Information fusion such as combining sequences, expressions, texts, and images, etc.
- Pathway analysis, gene regulatory networks, etc.
- Disease modeling
- Medical informatics

1.2 Goal

The goal of TC-20 is to bring together pattern recognition scientists and life scientists to find solutions to problems in bioinformatics and to foster multidisciplinary research in the pattern recognition community.

One of the goals is to facilitate the pattern recognition and the life science communities to interact through its memberships and thereby to match-make pattern recognition techniques and bioinformatics applications. An example of the benefits of proliferation of pattern recognition techniques in bioinformatics applications could be the discovery of new drugs.

The committee has a membership list and maintains a website for the exchange and dissemination of ideas among its members. It maintains a website having links to other websites of similar interests, the information on conferences and likely events, and useful data sources. The TC-20 will organize and support related workshops and seminars to bring its members and interested researchers together and disseminate information on latest research in the field.

3. TC-20 Structure and Organization

3.1 Chairmanship

In November 2004 during the last ICPR in Cambridge, a proposal to form the TC-20, put forward by the present Chairs, was accepted by the IAPR Executive Committee. The current chairmanships will hold until November 2006.

- TC20 Chair: Prof. Raj Acharya, The Pennsylvania State University, University Park, USA
- TC20 Vice-Chair: A/Prof. Jagath C. Rajapakse, Nanyang Technological University, Singapore

3.2 Membership

The founding membership was formed by the invitation of the TC-20 Chairs. TC-20 has 23 founding members worldwide.

3.3 Founding members

- Raj Acharya, The Pennsylvania State University, USA, (acharya@cse.psu.edu) (Chair)
- Francisco Azuaje, University of Ulster, Northern Ireland (fj.azuaje@ulster.ac.uk)
- Vladimir Brusic, Institute of Infocomm Research, Singapore, (vladimir@i2r.a-star.edu.sg)
- Phoebe Chen, Deakin University, Australia, (phoebe.chen@deakin.edu.au)
- Marchiori Elena, Vrije Universiteit Amsterdam, The Netherlands, (elena@cs.vu.nl)
- Mariofanna Milanova, University of Arkansas at Little Rock, USA, (mgmilanova@ualr.edu)
- Gary Fogel, Natural Selection, Inc., USA, (gfogel@natural-selection.com)
- Saman Halgamuge, University of Melbourne, Australia, (saman@unimelb.edu.au)
- Visakan Kadirkamanathan, The University of Sheffield, United Kingdom, (visakan@sheffield.ac.uk)
- Nik Kasabov, Auckland University of Technology, New Zealand, (nkasabov@aut.ac.nz)
- Irwin King, Chinese University of Hong Kong, Hong Kong, (king@cse.cuhk.edu.hk)
- Alexey V. Kochetov, Russian Academy of Sciences, Russia, (ak@bionet.nsc.ru)
- Graham Leedam, Nanyang Technological University, Singapore (asgledam@ntu.edu.sg)
- Ajit Narayanan, University of Exeter, United Kingdom, (A.Narayanan@ex.ac.uk)
- Nikhil R. Pal, Indian Statistical Institute, India, (nikhil@isical.ac.in)
- Muthu Palaniswami, The University of Melbourne, Australia, (swami@ee.mu.oz.au)
- Jagath C. Rajapakse, Nanyang Technological University, Singapore (asjagath@ntu.edu.sg) (Vice Chair)
- Gwenn Volkert, Kent State University, USA, (volkert@cs.kent.edu)
- Roy E. Welsch, Massachusetts Institute of Technology, USA, (rwelsch@mit.edu)
- Kay C. Wiese, Simon Fraser University, Canada, (wiese@sfu.ca)
- Limsoon Wong, Institute of Infocomm Research, Singapore, (limsoon@i2r.a-star.edu.sg)
- Qiang Yang, Hong Kong University of Science and Technology, Hong Kong, (qyang@cs.ust.hk)
- Yanqing Zhang, Georgia State University, USA, (yzhang@cs.gsu.edu)

3.4 Website

The IAPR website is currently has the following URL:

<http://www.cse.psu.edu/~acharya/IAPR/iapr.htm>

It is currently hosted by the Computer Science and Engineering, The Pennsylvania State University, University Park, USA and will soon have a IAPR specific domain.

The website is still in the infancy state and will be enhanced with links to major information sources, data sources, software tools, literature, etc.

3.5 Mailing List

The current mailing list consists of the founding members and is being used to discuss future activities of TC-20. Once the membership is open to public, the email list will be used as a moderated discussion forum.

4. Events

TC-20 will support and organize events pertaining to bioinformatics and pattern recognition. As a regular event, TC-20 will organize the International Workshop on Pattern Recognition in Bioinformatics (PRIB) in conjunction with the biennial International Conference on Pattern Recognition (ICPR), the major event of IAPR.

Although ICPR takes place once in two year if the TC members or local IAPR Chapters are interested, PRIB could be held in the interlude, annually.

4.1 First International Workshop on Pattern Recognition in Bioinformatics (PRIB'06)

First International Workshop on PRIB is planned to be held in conjunction with ICPR 2006, Hong Kong, on August 20, 2006.

Workshop Chair:

Raj Acharya, Pennsylvania State University, USA

Program Co-Chairs:

Jagath C. Rajapakse, Nanyang Technological University, Singapore
Limsoon Wong, Institute of Infocomm Research, Singapore

Publicity Chairs:

(to be announced)

Local Organization Chair:

(to be announced)

Program Committee:

(to be announced)

The papers presented in the workshop will be published in the Proceedings of the First International Workshop on Pattern Recognition in Bioinformatics (PRIB'06).

It is also planned to publish the extended version of papers selected from the workshop in an edited book.

5. Other Activities

5.1 Conferences

Several TC-20 members are actively engaged as Chairs of upcoming related conferences:

- Limsoon Wong serves as a Program Co-Chair of *2005 International Joint Conference of InCoB, AASBi, and KSBI (BIOINFO2005)* to be held in Busan, Korea, from September 22 – 24.
- Gary Fogel serves as the General Chair and Gwenn Volkert serves as a Technical Co-Chair at *2005 IEEE Symposium on Bioinformatics and Computational Biology (CIBCB'05)* to be held in San Diego, CA, from November 14 – 15, 2005;
- Mariofanna Milanova serves as a General Chair of *2005 International Conference on Machine Learning and Applications (ICMLA'05)* to be held in Los Angeles, CA, from December 15 – 17, 2005.

- Phoebe Chen and Limsoon Wong serve as General Co-Chairs of *the Fourth Asia-Pacific Bioinformatics Conference (APBC2006)*, to be held in Taipei, Taiwan, from February 13 – 16, 2006.
- Marchiori Elena serves as a Technical Co-Chair of the *Fourth European Workshop on Evolutionary Computation and Machine Learning in Bioinformatics (EvoBIO2006)* to be held in Budapest, Hungary, from April 10-12, 2006;

5.2 Editorial activities

Several TC-20 members serve as either Editors or Associate Editors of related journals:

- Nikhil Pal is the Editor-in-Chief of *IEEE Transactions on Fuzzy Systems*
- Limsoon Wong is a Managing Editor of *Journal of Bioinformatics and Computational Biology*
- Raj Acharya and Gary Fogel serve as Associate Editors of *IEEE Transactions on Bioinformatics and Computational Biology*.

6. 2005 – 2006 Work-plan

The main objectives for the next year are:

- Improve the webpage, mainly focusing on: the useful links to major bioinformatics data, information sources, tools, and major events in the field, and the development of a section linking pattern recognition methods and bioinformatics applications
- Collect benchmark datasets and performance evaluation algorithms to test pattern recognition applications in bioinformatics, and create a database
- Organization of First International Workshop on PRIB.
- Increase of current membership: after the formation of the sub-committees: the membership of TC-20 will be open to interested researchers via the website.

6.1 TC-20 Sub-committees

The objectives of the TC-20 will be achieved through the formation of subcommittees:

- Publicity and Publication Sub-committee (PPS): members work together for improving and keeping the website up-to-date, maintaining the emailing list and discussion forum, and preparing biennial interim report and quarterly newsletter
- Methods and Applications Sub-committee (MAS): members work together to build a database of various applications in bioinformatics, various techniques in pattern recognition, and how the pattern recognition techniques and applications are met through links to and survey of the recent literature
- Events and Contests Sub-committee (ECS): members find the opportunities to support and organize events and contests to increase the awareness of pattern recognition applications in bioinformatics

TC-20 founding members are invited to serve in the sub-committees. A TC20 Sub-committee will be led by a TC-20 Sub-committee Chair.