

S. Dua¹ and S.S. Iyengar²

Synopsis

¹Data Mining Research Laboratory
Computer Science Program
Louisiana Tech University
Ruston, Louisiana, USA

²Department of Computer Science
Louisiana State University
Baton Rouge, Louisiana, USA

Advances in Medical Signal Processing

In this section of the IMIA yearbook we present a unique assortment of research articles addressing some exigent and important problems in the area of medical signal processing. We present four research papers carefully selected from amongst the prestigious journals in this area, peer-reviewed and compiled for an amalgamated presentation.

Medical signal processing has manifested itself as a core component of research and development in medical informatics. Vast varieties of research originating from researchers around the world, coupled by an equal pace of product development in this area has lead this area to attract an unprecedented attention from both researchers and practitioners. This research has spanned from technological advancement in the area of analysis of signals originating in clinical domains, such as cardiocography and electrophysiology to signal processing practices in genomic engineering. Problems represented by increasing world health disease burden combined with escalating cost of healthcare will be greatly resolved by enhanced capability to computationally interpret signal data originating from these domains, as demonstrated in the following papers.

The first article in our compendium is “Visualization and evaluation of clusters for exploratory analysis of gene expression data” by Ju Han Kim *et al.* [1]. In the paper, authors present a unique computational framework for visualization and evaluation of clusters for exploratory analysis of gene expression data.

Biology has become an increasingly data rich discipline. Elucidation of molecular mechanisms in living cells is one of the major challenges of biology today. A recent breakthrough in methodology for experimental molecular biology is microarray technology, which has become an important tool for monitoring and analyzing gene expression profiles of thousands of genes simultaneously. The rate of collection of these data has far surpassed our ability to analyze the data for novel, non-trivial, and important knowledge. Data mining [2], commonly called knowledge discovery in large databases, has emerged as a set of powerful tools and methodologies for gene expression data analysis, by rapid, accurate, and practicable means. Among several data mining methodologies, unsupervised classification (clustering) has recently been applied to gene expression data analysis in

several computational and experimental settings [3], providing some exciting results. Clustering of gene expression data coupled by schemas for visualization of these clusters can provide valuable hypothesis development from the gene expression data by not limiting the analysis to individual genes already familiar in molecular physiology and virology.

In the paper, the authors follow up on their previously reported matrix incision tree (MITree) algorithm, a divisive hierarchical clustering algorithm that was able to reveal conceivable clusters in gene expression data. They emphasize that hierarchical classification can be viewed as a special nested sequence of partitions and proximity graph, which when converted to a proximity dendrogram, can present a unified view of the graphical representation of clustering structures having both hierarchical and partitional components. The proposed framework makes no assumption on data distribution to present a novel paradigm of *bi-incisional* hyperplanes [1] with same dimensionality whose union, as they demonstrate, can partition high-dimensional data such as gene expression data into more than two sub-spaces. It is asserted that in

contrast to the non-structure of K-means clustering algorithms[4] and the partial qualitative structure of SOM (SelfOrganizing maps)[5], the similarity measurements of MITrees's permit comprehensive multilevel quantitative visualization of similarity structures and substructures. They also present an efficient heuristic approximation algorithm by combining modified K-medoids algorithm[3] and an intervening incremental trimming-and-reassignment strategy. Conclusively, they tested their framework on four published and well-studied datasets, demonstrating the superiority of their results.

The second article in our compendium is "High Accuracy of Automatic Detection of Atrial Fibrillation using Wavelet Transform of Heart Rate Intervals" by Duverney *et al.*[6]. In this paper, the authors propose a wavelet transform based approach for automated detection of Atrial Fibrillation (AF).

Permanent and paroxysmal AF is a risk factor for the occurrence and the reoccurrence of stroke, which can occur as its first manifestation. Although initially, AF was believed to be a completely chaotic occurrence with ill-patterned electrical impulses within the heart's atria randomly, recent advances in this and related areas have eradicated this belief. Advances in signal processing techniques such as wavelet transforms have provided computational tools for analysis of signals that are expected to be associated with AF towards the development of a framework for computer-assisted automated identification of AF ([6] and references therein). Wavelet transform is a powerful tool for analysis of non-stationary signals, by representing them as linear combinations of atomic decompositions which allows extorting the simple constituents that make up a complicated structure or signal.

In the presented paper, the authors start their discussion with well laying

out the motivation for studying heart rate intervals for predicting and classifying AF. For example, the authors note that about one-fourth of Cerebrovascular strokes in USA and Europe have been associated with AF[7]. After discussing the recent advances in the related areas, the authors clearly describe their methods, discussing in details their population and recording methodology. The authors emphasize that due to its ability to precisely characterize at any time RR variability, novels methods based on Wavelet transforms associated with fractal classification could prove to be a powerful tool in AF detection. In the author's study, a quadratic spline of order three wavelet transform was employed which was successful in identifying periods of high HRV coefficients. A second step using fractal analysis was employed to classify these periods of high variability periods into physiological and pathological rhythms. When applied to the validation group, the author's methodology provided a sensitivity of 99.2% for the detection of AF QRS complexes, and a specificity of 99.9%. The authors also carefully study and present the limitations of their approach, pointing out that their approach does not guarantee null false negative rate.

The third article in our compendium is "Linear and Nonlinear Parameters for the Analysis of Fetal Heart Rate Signal from Cardiotocographic Recordings" by Maria Signrini *et al* [8]. The authors in this paper present a novel work in the automated detection of the status of fetal well-being.

The previous work in this area asserts that the fetal distress that can be preceded by alterations in heart interbeat intervals before any appreciable changes occur in the heart rate itself and both linear and nonlinear effects contribute to the signal generation pattern ([8] and references therein). These results also advocated that fetal heart rate regulation mecha-

nisms show an intrinsic nonlinear behavior. Although it was strongly believed that the heart rate properties could be particularly useful in monitoring the fetal wellbeing, the previous research in this area did not extract any quantitative indexes linking fetal heart rate signal patterns with patho-physiological fetal states. These revelations laid the motivation for the author's research.

In this paper, the authors introduce few parameters to improve the diagnostic ability of the traditional time domain cardiotocographic fetal heart rate analysis. These parameters allow obtaining a quantitative evaluation of both linear and nonlinear contributions to the heart rate generation. They assert that spectral parameters establish significant differences between normal and pathological subjects and are able to demonstrate that spectral parameters change under different fetal conditions such as activity or quiet sleep, corroborating their usefulness in the classification of fetal states. The authors claim that the ultra sound classical cardiotocography should receive a significant improvement by including the proposed parameters into a nonlinear multi-variate classifier. And also, by coupling physiology related indexes and nonlinear classifiers, it should be possible to enhance the reliability and robustness of fetal wellbeing nonstress monitoring in the antepartum period.

The final article in our compendium is "Signal processing techniques in genomic engineering" by Xin-Yu Zhang *et al.* [9]. In this paper, the authors have presented an overview of the applications of recent signal processing techniques on genomic signal analysis.

The authors have divided the analysis into two regimes of discussion: DNA sequence analysis and DNA microarray analysis. With an exponential generation of complete DNA sequences, it is particularly imperative to

decode the inherent features in these sequences. Many studies have been carried out to extract the characteristic segments, to reveal some hidden structures, to distinguish coding from non-coding regions in DNA sequences, and to explore structural similarity among DNA sequences ([9] and references therein). The authors have discussed the contributions in the area of application of wavelet transform to sequence structure prediction, DNA sequence comparison and classification, and in the exploration of the relationship between sequence structure and function.

In the subsequent part of the paper, the authors have discussed the research in the area of image signal processing in DNA microarray analysis. Multi-colored fluorescent techniques are often used in comparative hybridization detection in DNA microarray experiments. These techniques generate images (two-dimensional signals) which can be analyzed for gene expression data synthesis. The basic step is to transform an image of spots of varying intensities into gene expression matrix. The authors have surveyed the techniques for noise reduction, automation, and real-time analysis in microarray images. They have specifically discussed techniques of cellular neural networks in DNA

microarray parallel processing, and to obtain gene expression profile in real-time. It is followed by discussion on some potential and exciting avenues for future research, especially in the areas of biomodel-based technique in genomic data processing, usage of multidimensional signal processing technique, development of novel signal processing techniques for noise reduction, interpretation and application of CNN-based microprocessors in Microarray analysis.

The area of medical signal processing is important, elaborate and aggressively growing. The papers presented in this section of the 2004 IMIA yearbook present a succinct view of the spectrum of challenges in this exciting area.

References:

1. Kim JH, Kohane IS, Machado LO. Visualization and evaluation of clusters for exploratory analysis of gene expression data. *J Biomed Inform* 2002;35(1):25-36.
2. Han J, Kamber M. *Data Mining: Concepts and Techniques*. USA: Academic Press; 2001.
3. Sharan R, Shamir R. Algorithmic approaches to clustering gene expression data. In: Jiang T; Ying X, Zhang MQ, editors: *Current Topics in Computational Molecular Biology*. MIT Press; 2002. p. 269-300.
4. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. *Nat Genet* 1999;22:281-5.
5. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999; 96(6):2907-12.
6. Duverney D, Gaspoz JM, Pichot V, Roche F, Brion R, Antoniadis A, et al. High accuracy of automatic detection of atrial fibrillation using wavelet transform of heart rate intervals. *Pacing Clin Electrophysiol* 2002 Apr;25(4 Pt 1):457-62.
7. Camm AJ, Obel OA. Epidemiology and mechanism of atrial fibrillation and atrial flutter. *Am J Cardiol* 1996;78:3-11.
8. Signorini MG, Magenes G, Cerutti S, Arduini D. Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings. *IEEE Trans Biomed Eng* 2003 Mar;50(3):365-74.
9. Zhang XY, Chen F, Zhang YT, Agner SC, Akay M, et al. *Signal Processing Techniques in Genomic Engineering*. Proceedings of the IEEE 2002 Dec; 90(12):1822-33.

Address of the authors:

Sumeet Dua
Data Mining Research Laboratory
Computer Science Program
Louisiana Tech University
Ruston, LA – 71272 USA
E-mail: sdua@coes.latech.edu

S.S. Iyengar
Department of Computer Science
Louisiana State University
Baton Rouge, LA – 70803 USA
E-mail: iyengar@bit.csc.lsu.edu