Final report: M. Sznaier CNR sponsored visit to IEIIT-CNR, June 1-10, 2011

Description of the research and motivation.

Enhanced data collection and analysis capabilities can profoundly impact society, with benefits ranging from safer, self aware environments, to enhanced image-based therapies. A major impediment to realizing this vision stems from the curse of dimensionality. Simply put, existing techniques are ill-equipped to deal with the overwhelming volume of data.

(a) (b) (c) (d)

Fig. 1: Examples of sparsely encoded visual information. (a) Mamogram showing a cancerous lesion. (b) and (c) sample frames showing contextually abnormal events: onset of a tunnel fire and a person entering through an exit . (d) Tracking multiple targets. In all cases less than $O(10^{-6})$ of the data is relevant.

This situation is illustrated in Fig. 1, where, in all cases, decisions must be taken based on events discernible only in a small fraction of a very large data record: a short video sequence adds up to megabytes, yet actionable information (a change of behavior of a single target), may be encoded in just a few frames, e.g. less than 10^{-6} of the total data. Additional challenges arise from the quality of the data, often fragmented and corrupted by noise.

The objective of this short term visit was to start a new research line seeking to develop computationally tractable methods for compressive information extraction by exploiting a combination of methods at the confluence of 3 fields, each represented by one of the researchers: F. Dabbene (Convex Optimization), M. Sznaier (Identification and Machine Learning), and R. Tempo (Information Based Complexity). Briefly, in this approach the observed data is treated as the output of a hidden dynamical system, and actionable information is encapsulated in the relatively few parameters describing these systems. Thus, a key enabler is the ability to optimally identify these parameters from noisy, incomplete measurements.

Outcomes of the visit.

During this visit, we have taken steps towards opening a new research line, "*Probabilistic Information Based Complexity,*" that combines elements from classical Information Based Complexity with recent developments in Randomized Algorithms. We anticipate that this new line of research will lead to computationally tractable algorithms for optimal "risk-adjusted" estimation and identification that substantially outperform existing state of the art methods. These new results will be presented at a forthcoming workshop on robust control to be held in

Udine, Italy at the end of August. In addition, a journal paper summarizing these results is being prepared. Finally, during this visit, M. Sznaier presented a seminar entitled “Compressive Information Extraction” at the Politecnico di Torino.

Consiglio Nazionale delle Ricerche
Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni

Seminar Announcement

Mario Sznaier

Northeastern University, Boston, USA

Convex Relaxations of Identification Problems Arising in the Context of Information Extraction

Abstract -- This talk addresses the problem of extracting actionable information that is very sparsely encoded in high dimensional data streams. The central theme of our approach is the realization that actionable information can be often represented with a small number of invariants associated with an underlying dynamical system. Thus, in this context, the problem of actionable information extraction can be reformulated as identifying these invariants from (high dimensional) noisy data, and thought of as a generalization of sparse signal recovery problems to a dynamical systems framework. While in principle this approach leads to generically nonconvex, hard to solve problems, computationally tractable relaxations (and in some cases exact solutions) can be obtained by exploiting a combination of elements from convex analysis and the classical theory of moments.

In the second part of the talk we will illustrate the application of these results to some challenging problems such as video, image and genomic data segmentation, where the goal is to detect changes, for instance in scenes, activities, texture, or gene promoter expressions. We will conclude the talk by exploring the connection between information extraction, hybrid systems identification and machine learning, and discuss new research directions in optimization and systems theory motivated by these problems.

Friday, June, 9th 2011, 10:30

Sala Riunioni CNR IEIIT

For more info: Fabrizio Dabbene