

## Knowledge Extraction by Dynamical Clustering of sea waves streaming data

Elvira Romano<sup>1</sup>, Antonio Balzanella<sup>1</sup>, Rosanna Verde<sup>2</sup>

<sup>1</sup> Dipartimento di Matematica e Statistica Università degli Studi di Napoli "Federico II",  
Via Cintia Complesso Monte Sant'Angelo I-80126, Napoli

<sup>2</sup> Dipartimento di Studi Europei e Mediterranei, Seconda Università di Napoli, San Leucio

**Abstract.** Data stream can be thought as a sequence of ordered data items, where the input arrives more or less continuously as time progress. There exist several applications producing data stream, e.g. telecommunication system, stock markets customer click streams etc.. In this paper we consider the problem of extracting knowledge by a dynamical clustering algorithm of sea waves streaming data, that is to say evolving streaming of data coming from a multisensor system. For this purpose we develop an updating version of Dynamical Clustering Algorithm [5]. This problem is very interesting from a practical point of view. It is based on the computation of a prototypal wave through a free-knot smoothing spline, optimizing a non linear problem. Thanks to this approach, it is possible to investigate in which way the incoming data change according to the various steps of process registration, and to have a summary description of the entire data thought prototypals flowing curves using a small amount of memory and time.

**Keywords:** data stream, data mining, clustering, sea waves propagation, free knots spline.

### 1 Introduction

In applications such as network monitoring, telecommunications data management, clickstream monitoring, manufacturing, sensor networks, and others, data takes the form of continuous data streams rather than finite stored data sets. These data are created by continuous activity over long periods of time and are therefore data which grow continuously over time. This has lead to the development of new strategies able to furnish an overview of the key characteristics in the data which change over time in a fast way. In the last years since it has been advances in hardware technology the analysis of such kind of data has gained more attention. An important area of interest is that of clustering. The clustering problem is especially interesting because of its application to data summarization and outlier detection, but is a difficult problem to solve because of large volumes of data arriving in a stream lead most traditional algorithms too inefficient. Regarding aspects of data storage, management and processing, the continuous arrival of data items in multiple, rapid, time-varying requires clustering adaptive strategies in the sense to up-to-date clusters, and able to taking new data items into consideration as soon as they arrive. The clustering problem has addressed many scientific area such as the database, data mining and statistics communities [3], [6],[10]. Despite these work presents interesting methods to classify datastream they have the aim to classify elements of one individual data stream, which is quite different from the problem that we consider here. In this paper we are going to analyze the streams themselves rather than single data items thereof. Our approach is different on the way of performing clustering. It share some ideas with the field of Dynamical Curve Clustering [14], [13], where the main aim is to classify a set of time stamped curves. We develop an updating version of Dynamical curve clustering where our focus is the analysis of *stream of curves* coming from an applicative study on sea waves streaming data. The idea is to classify this data using sliding windows of fixed size. The method incorporates good quality of clusters combined with efficient computational properties. It is incrementally updatable and is highly scalable on both the number of dimensions and the size of the data streams.