*Journal of Nonlinear Systems and Applications (2010) 63–71*
*Copyright © 2010 Watam Press* *http://www.watam.org/JNSA/*

# LEARNING SELF-ORGANIZING MIXTURE MARKOV MODELS

Nicoleta Rogovschi, Mustapha Lebbah and Younès Bennani *

**Abstract.** This paper describes a new algorithm to learn Self-Organizing map as Markov Mixture Models. Our model realizes an unsupervised learning using unlabelled evolutionary data sets, namely those that describe sequential data. The new formalism that we present is valid for all structure of graphical models. We use E-M (Expectation-Maximisation) standard algorithm to maximize the likelihood. The graph structure is integrated in the parameter estimation of Markov model using a neighborhood function to learn a topographic clustering of not i.i.d data set. The new approach provides a self-organizing Markov model using an original learning algorithm. We provide three applications of our model: (1) dealing with continuous data using Gaussian distribution; (2) dealing with categorical data without any coding to encode variables using probability tables; (3) dealing with binary data using Bernoulli distribution.

## 1 Introduction

Since many years, temporal and spatial sequences have been the subject of investigation in many fields, such as statistics, pattern recognition, web mining, and bioinformatics. The easiest way to treat sequential data would be simply to ignore the sequential aspects and treat the observations as independent and identically distribution (i.i.d) in the first stage. For many applications, the i.i.d assumption will be a poor one. Often in many application the treatment is decomposed in two steps; the first one is clustering task. In second stage the result of clustering is used to learn a probabilistic model by relaxing the i.i.d. assumption, and one of the simplest ways to do this is to consider a Markov model.

Hidden Markov Models (HMMs) are the most well-known and practically used extension of Markov models. They offer a solution to this problem introducing, for each state, an underlying stochastic process that is not known (hidden) but could be inferred through the observations it generates. In fact the probabilistic graphical modelling motivates different graphical structures based on the HMM [1, 2]. Another variant of the HMM worthy of mention is the factorial hidden Markov model [3], in which there are multiple independent Markov chains of latent variables, and the distribution of the observed variable at a given time step is conditional on the states of all of the corresponding latent variables at that same time step. Many related models, such as hybrids of HMMs with artificial neural networks [4, 5, 6].

Clearly, there are many possible probabilistic structures that can be constructed according to the needs of particular applications. Graphical models provide a general formalism for motivating, describing and analysing such structures. Therefore, it will be very important to have algorithms able to infer from a data set of sequences not only the probability distributions but also the topological structure of the model, i.e., the number of states and the transitions interconnecting them. Unfortunately, this task is very difficult and only partial solutions are today available [7, 8, 9]. In order to overcome the limitations of HMMs, in [9] the author proposes a novel and an original machine learning paradigm, which is titled topological HMM, that embeds the nodes of an HMM state transition graph in an Euclidian space. This approach models the local structure of HMM and extract their shape by defining a unit of information as a shape formed by a group of symbols of a sequence.

Other attempts have been made for combining HMMs and SOM (Self-Organizing Map of Kohonen) to form hybrid models that contain the clustering power of SOM with the sequential time series aspect of HMMs [6]. In many of these hybrid architectures, SOM models are used as front-end processors for vector quantization, and HMMs are then used in higher processing stages. In [10, 11], a vector sequence is associated with a node of SOM using DTW (dynamic time warping) model. Other works exist and differ in the manner of combination [12, 13, 14]. In [14] the authors propose an original combined model which is the offspring of a crossover between the SOM algorithm and the HMM theory. The model's core consists in a novel unified/hybrid SOM-HMM algorithm where each cell of SOM map presents an HMM. The model is coupled with a sequence data training method, that blends together the SOM unsupervised learning and the HMM dynamic programming algorithms. Of course, there is a lot more litterature on HMMs and their applications than can be covered here, but this survey wants to be representative of the

*Nicoleta Rogovschi, Mustapha Lebbah and Younès Bennani are with LIPN-UMR 7030 - CNRS, Université Paris 13. 99, av. J-B Clément F-93430 Villetaneuse e-mail: firstname.secondname@lipn.univ-paris13.fr.