

Classification probabiliste non supervisée et visualisation des données séquentielles

Rakia Jaziri ^{*,**}, Mustapha Lebbah^{*}, Younès Bennani^{*},

^{*}LIPN-UMR 7030 - CNRS, Université Paris 13,
99, av. J-B Clément F-93430 Villetaneuse
Prénom.Nom@lipn.univ-paris13.fr

^{**}Institut National de l'Audiovisuel,
4, av. de l'Europe 94 366 Cedex Bry-sur-Marne
rjaziri@ina.fr

Résumé

Nous proposons dans ce papier un nouvel algorithme de classification non supervisée à base de modèle de mélange topologique pour des données non i.i.d (non independently and identically distributed). Ce nouveau paradigme probabiliste, plonge les cartes topologiques probabilistes dans une formulation sous forme de chaînes de Markov cachées. Dans cette formulation, la génération d'une observation à un instant donné du temps est conditionnée par les états voisins au même instant du temps. Ainsi, une grande proximité impliquera une grande probabilité pour la contribution à la génération. L'approche proposée est évaluée en utilisant des données séquentielles réelles issues des bases de données de l'Institut Nationale de l'Audiovisuel (INA). Les résultats obtenus sont très encourageants et prometteurs.

Summary

We present a generative approach to learn a new probabilistic Self-Organizing Map (Pr-SOMS) for non independent and non identically distributed data sets. Our model defines a low dimensional manifold allowing friendly visualizations. To yield the topology preserving maps, our model has the SOM like learning behavior with the advantages of probabilistic models. This new paradigm uses HMM (Hidden Markov Models) formalism and introduces topological relationships between the states. This allows us to take advantage of all the known classical views associated to topographic map. We demonstrate our approach on the real-world data issued from French National Audiovisual Institute.