

Implémentation d'algorithme de Viterbi pour le décodage en ligne d'une séquence d'états cachés

Un **modèle de Markov caché** (MMC) (*Hidden Markov Model* (HMM)) [1-2] est un modèle numérique d'un système ou un processus à temps discret ($t=1, 2, \dots, T$) doublement stochastique : une séquence mesurable (séquence d'observations continues ou discrets) $O = \{o_1, o_2, \dots, o_T\}$ générée par une séquence non mesurable ou cachée (séquence d'états discrets cachés) $X = \{x_1, x_2, \dots, x_T\}$. Le système modélisé est supposé être un processus Markovien de paramètres inconnus mais peuvent être estimés par un algorithme stochastique itératif d'apprentissage artificiel qui utilise uniquement la séquence d'observations, Fig. 1.

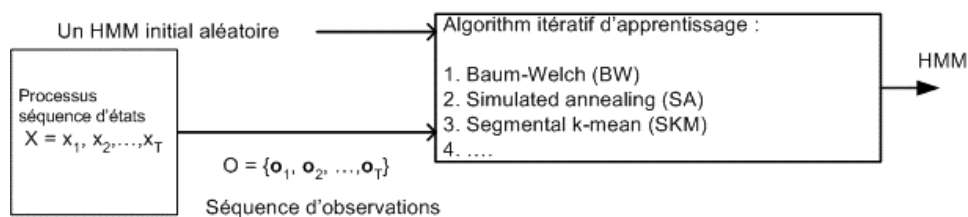


Fig. 1. Apprentissage d'un HMM à partir d'une séquence d'observations générée par le processus à modéliser.

Les HMMs sont largement utilisés notamment en reconnaissance de formes, diagnostic médical ou industriel, en intelligence artificielle ou encore en traitement automatique du langage naturel [1-3].

Une fois un HMM est construit, la séquence d'états cachés peut être décodée en utilisant un algorithme de programmation dynamique stochastique appelé algorithme de Viterbi [4]. Cet algorithme utilise la séquence d'observation et son HMM pour générer la séquence d'états cachés associée à la séquence d'observations, Fig. 2.

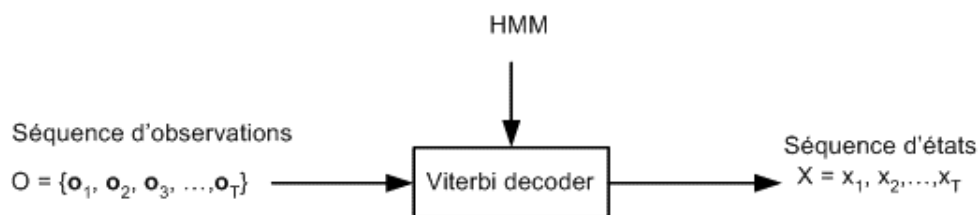


Fig. 2. Décodage de la séquence d'états cachés d'un processus en utilisant son modèle HMM et une séquence d'observations générée par ce processus.

Objectif de ce projet : L'algorithme de Viterbi est implémenté en Scilab [5]. Cet algorithme permet de de décoder une séquence d'état de $t=1, \dots, T$ en mais nécessite d'attendre la fin de la séquence d'observations ($t=T$). Cette implémentation est nécessaire pour certaines applications comme la reconnaissance d'un mot prononcé (il faut attendre la fin de la prononciation de ce mot pour pouvoir le décoder). Pour d'autres types d'applications comme, par exemple le diagnostic médical ou industriel, une détection en ligne ou en temps réel est nécessaire. Pour ce type d'applications, la longueur de la séquence

d'observation est assez importante ou même peut être considérée sans limite connue a priori. Dans ce cas il faut inférer ou décoder l'état caché à chaque instant courant du temps t où l'observation \mathbf{o}_t est disponible. L'objectif de ce projet est alors d'implémenter l'algorithme de Viterbi en ligne sous Scilab. Le code obtenu sera validé sur un problème de diagnostic médical.

Références

- [1] http://fr.wikipedia.org/wiki/Mod%C3%A8le_de_Markov_cach%C3%A9
- [2] http://en.wikipedia.org/wiki/Hidden_Markov_model
- [3] <http://www.esiee.fr/~alanit/Publications.html>
- [4] http://en.wikipedia.org/wiki/Viterbi_algorithm
- [5] <http://www.scilab.org>

Lieu du projet : ESIEE-Dépt. Informatique.

Moyens matériels : Machine PC sous Windows ou Linux,

Moyens logiciels fournis par le conseiller scientifique : logiciel Scilab, boîte à outils Scilab Hidden Markov Models

conseiller scientifique : T. AL-ANI
Bureau 5358, e-mail : t.alani@esiee.fr