

Champs de Markov conditionnels pour le traitement de séquences¹

Trinh Minh Tri Do*, Thierry Artières*

*LIP6, Université Paris 6

8 rue du capitaine Scott

75015 Paris France

Do@poleia.lip6.fr, Thierry.Artieres@lip6.fr

Résumé. Les modèles conditionnels du type modèles de Markov d'entropie maximale et champs de Markov conditionnels apportent des réponses aux lacunes des modèles de Markov cachés traditionnellement employés pour la classification et la segmentation de séquences. Ces modèles conditionnels ont été essentiellement utilisés jusqu'à présent dans des tâches d'extraction d'information ou d'étiquetage morphosyntaxique. Cette contribution explore l'emploi de ces modèles pour des données de nature différente, de type « signal », telles que la parole ou l'écriture en ligne. Nous proposons des architectures de modèles adaptées à ces tâches pour lesquelles nous avons dérivé les algorithmes d'inférence et d'apprentissage correspondant. Nous fournissons des résultats expérimentaux pour deux tâches de classification et d'étiquetage de séquences.

1 Introduction

La classification, la segmentation et l'étiquetage de données séquentielles sont des problématiques au cœur de nombreux domaines comme la bioinformatique, la reconnaissance de l'écriture, l'extraction d'information. Une des problématiques principales dans ce type de domaine consiste en effet à transformer une séquence observée (un signal écrit par exemple) en une séquence d'étiquettes (on utilise également le terme de labels). Cette tâche peut être réalisée à différents niveaux. On cherche à segmenter le signal écrit d'une phrase en une séquence de mots, de même que le signal écrit de chaque mot doit être segmenté en une séquence de caractères, etc.

Les modèles Markoviens cachés (MMC) constituent l'approche la plus utilisée pour résoudre ce type de tâches bien qu'ils reposent sur des hypothèses d'indépendance fortes sur les données et qu'ils soient appris de façon non discriminante. Ce dernier point vient du fait que ce sont des modèles génératifs et qu'ils définissent une loi de probabilité conjointe $P(X, Y)$ sur la séquence d'observations X et la séquence d'étiquettes associée Y . Diverses

¹ Ce travail est en partie financé par le programme IST de la communauté européenne, à travers le réseau d'Excellence PASCAL IST-2002-506778.