

# Comparaison de l'expressivité des réseaux de neurones récurrents et des modèles de Markov cachés

Achille Salaün<sup>(1) (2)</sup>, Yohan Petetin<sup>(2)</sup> et François Desbouvries<sup>(2)</sup>

(1) Nokia Bell Labs (2) Samovar, CNRS, Telecom SudParis, Institut Polytechnique de Paris

Les modèles de Markov cachés (HMM = *Hidden Markov Models*) et réseaux de neurones récurrents (RNN = *recurrent neural network*) sont des modèles très utilisés dans le traitement de données séquentielles, par exemple en reconnaissance de la parole, prédiction de séries temporelles ou encore traduction automatique de texte. Bien que ces modèles aient été étendus et améliorés (les *Long Short Term Memory* et *Gated Recurrent Units* sont des architectures neuronales plus sophistiquées, les Modèles de Markov partiellement observés généralisent les chaînes de Markov cachées...), leur compréhension et comparaison théoriques demeurent un problème ouvert.

Le principe des RNN consiste à modéliser la loi jointe des observations par l'intermédiaire de variables latentes (les neurones) construites de manière déterministe à partir des observations. Les HMM reposent eux aussi sur des variables latentes (les états cachés), aléatoires cette fois-ci. Si dans de nombreuses applications les HMM considérés sont à états cachés discrets, les HMM à états continus ont trouvé de nombreuses applications dans les problèmes d'inférence bayésienne séquentielle et permettent également de modéliser une loi de probabilité sur les observations. En résumé, HMM (à états continus) et RNN sont des structures capables toutes deux de modéliser une loi jointe sur les observations par le biais de variables latentes continues. On remarque que la différence principale entre ces deux modèles provient de la structure de dépendance régissant les variables cachées (les neurones ou les états continus) et observées. Plus précisément, à un instant donné, la variable cachée d'un RNN est mise à jour de façon déterministe en fonction de la variable cachée et de l'observation précédentes ; tandis que celle d'un HMM dépend aléatoirement de la variable cachée précédente mais pas des observations. Ainsi les deux modèles montrent certaines différences et similitudes et peuvent finalement être décrits tous deux dans le cadre commun d'un modèle génératif unifié (GUM = *generative unified model*) que nous détaillons.

Nous étudions l'expressivité des modèles RNN et HMM dans le cadre unifié des GUM. En particulier, et afin de dresser une comparaison théorique du pouvoir modélisant de chacun de ces deux modèles, nous imposons la modélisation d'une loi jointe sur les observations telle que chaque marginale suit une loi normale centrée réduite. Sous cette contrainte, nous déterminons l'ensemble des lois jointes qu'il est possible d'obtenir avec chacun des trois modèles (HMM, RNN et GUM) : dans ce contexte, toute instance de GUM (dont HMM et RNN font partie) modélise n'importe quelle loi telle que la covariance entre deux observations dépend géométriquement de leur écart dans la séquence, et seulement de cet écart.

Nous concluons ainsi sur leurs expressivités respectives : ni les HMM, ni les RNN ne sont strictement inclus dans l'autre. Par ailleurs, si les HMM offrent un pouvoir modélisant supérieur à celui des RNN, c'est au prix d'un nombre plus important de paramètres à estimer. Nous ouvrons finalement sur certaines directions futures comme la généralisation de nos travaux au cas non gaussien ou l'utilisation pratique du modèle GUM dans un contexte d'apprentissage à partir d'une séquence d'observations.