

MACHINE LEARNING

► Dès aujourd'hui, l'Institut de Martigny accueille près de 150 spécialistes du

L'IDIAP joue ses atouts dans la bio-

Les algorithmes d'apprentissage, utilisés dans la reconnaissance de la parole, s'appliquent



Sylvie Gardel

Les compétences suisses en matière de bio-informatique ne sont plus à démontrer. L'Institut suisse de bio-informatique, celui de l'ISREC, celui en commun entre l'EPFZ et l'Université de Zurich, et surtout Swiss-Prot, une des bases de données les plus importantes pour l'analyse des protéines, placent en effet notre pays parmi ceux qui excellent dans ce secteur. Il faut dire que la bio-informatique, mariage entre l'analyse de l'information biologique (en majorité sous forme de séquences génétiques et de structures de protéines) et les outils informatiques sur lesquels elle est fondée, a pris un essor fulgurant depuis le décryptage de la carte du génome humain.

domaine public). Deux voies d'approche sont généralement utilisées: la combinaison de méthodes bio-informatiques avec de nouvelles stratégies expérimentales (génomique comparative, analyse des profils d'expression, etc.) et la recherche de méthodes prédictives basées sur une analyse de plus en plus sensible des séquences (similarité, motifs fonctionnels ou structuraux).

Mais qui dit étude des séquences, dit aussi traitement de signaux de plus en plus subtils (signal processing, en anglais) et apprentissage statistique (Machine Learning) avec des outils comme des réseaux de neurones artificiels (neural networks). Dès aujourd'hui et jusqu'à vendredi, ces deux domaines de recherches, basés sur le développement d'algorithmes mathématiques, sont à l'honneur du séminaire organisé par l'Institut Dalle Molle d'intelligence artificielle perceptive (IDIAP) de Martigny, en collaboration avec l'École polytechnique fédérale de Lausanne (EPFL), où près de cent

traitements des signaux avec le laboratoire du professeur Murat Kunt de l'EPFL». Les deux institutions étudieraient par ailleurs un rapprochement plus formel dans le but d'optimiser leurs connaissances en bio-informatique, notamment.

Le Machine Learning extrait de grandes bases de données des concepts généraux

Autant d'atouts que le petit Institut valaisan entend désormais mettre en avant dans les recherches de génomique fonctionnelle et structurale. Et plus particulièrement encore, les techniques de Machine Learning. Cette expression anglo-saxonne désigne un ensemble de techniques souvent basées sur les statistiques qui ont pour but d'extraire de grandes bases de données des concepts généraux et unificateurs. Ce sont les algorithmes dits d'apprentissage. Ainsi, de la reconnaissance vocale à l'étude des séquences de gènes ou de protéines, le processus demeure le même. «Il existe, par exemple, une multitude de manières d'é-

gnés, classés, etc.) par comparaison directe, mais à travers la construction d'un modèle qui tente, dans une première étape, d'en capturer les propriétés communes.

Une des grandes familles de problèmes sur lesquels travaille l'IDIAP relève des séquences d'information, notamment présentes dans la reconnaissance de la parole ou des images vidéo. Dans ces cas-là, la taille des séquences est en effet très variable, ce qui nécessite des modèles particuliers. «Pour résoudre ces problèmes, on utilise des modèles statistiques comme les chaînes de Markov (un thème qui sera abordé par le professeur Anders Krogh de l'Université de Copenhague, ce mercredi 4 septembre, dans le cadre de l'étude des protéines et de l'ADN, ndlr).» Pour l'analyse de données essentiellement séquentielles, les modèles probabilistes fondés sur les chaînes de Markov (simples ou cachées) se sont révélés particulièrement fructueux. L'ADN, d'abord représenté par les lettres A, C, T, G - les quatre molécules primordiales de l'ADN - l'est aujourd'hui