

Proposition de stage M2 2013/2014

Modèles graphiques probabilistes pour l'analyse de données issues de systèmes complexes

RESPONSABLE: Cyril Furtlehner

LABORATOIRE: Tao INRIA Saclay, LRI

ADRESSE: Université Paris-Sud, 91405 Orsay Cedex

MAIL: Cyril.Furtlehner@lri.fr,

Résumé :

Les modèles graphiques probabilistes tels les réseaux Bayésiens ou les champs Markoviens aléatoires (MRF) constituent un outil important pour l'inférence probabiliste. Le graphe permet de symboliser le jeu d'interactions entre variables, auxquelles sont associés des vertex, reliés entre eux par des liens dirigés dans le cas Bayésien ou non dans le cas des MRF, ceux-ci représentant alors des interactions probabilistes entre paires de variables.

Sans hypothèse particulière la construction de tels modèles à partir d'un jeu de données constitue un problème mal posé, et malgré des hypothèses restrictives il demeure souvent intractable. Cependant, sous certaines hypothèses simplificatrices (interactions de paires, variables Gaussiennes ou binaires...) des méthodes existent permettant de sélectionner de tels modèles avec différents compromis possibles à réaliser entre vraisemblance et simplicité.

La détermination d'un tel modèle peut non seulement être utile du point de vue de son exploitation pour des tâches d'inférence par exemple, mais encore être riche d'enseignements sur les données elles-mêmes, en particulier si le compromis entre vraisemblance et simplicité permet de ne conserver que l'information pertinente et d'éliminer le "bruit". On peut supposer en effet que dans certains cas le réseau d'interactions obtenu puisse être spécifique d'un individu ou d'une situation du système complexe sous-jacent génératif des données.

Par exemple si l'on considère les signaux EEG (ElectroEncephaloGraphy) recueillis par des casques BCI (BrainComputerInterface), une fois traités, ils se présentent en général sous la forme d'une matrice de covariance entre les signaux EEG provenant des différents capteurs. Trouver un MRF (Gaussien) représentatif de cette matrice de covariance, consiste alors à identifier le graphe des interactions les plus importantes et les couplages correspondants. L'hypothèse à explorer durant ce stage est que le modèle ainsi obtenu pourrait constituer une signature fiable et discriminante du sujet et/ou de la tâche effectuée pendant l'enregistrement.

Objectif du stage

Le stage s'intéresse à la détermination de modèles graphiques et à leur exploitation à partir de différents jeux de données possibles à analyser, données de trafic routier sur des réseaux de grande échelle, données BCI provenant de casques EGE, données FMRI.

Un certain nombre de questions se posent, concernant d'abord le bon compromis à trouver entre simplicité et vraisemblance. Puis, une fois obtenus ces modèles: comment les comparer, comment choisir une distance appropriée dans l'espace des modèles graphiques; peut-on rendre compte de la dynamique des systèmes complexes sous-jacents à l'aide de ce type de modèles?

Ce stage, outre les aspects pratiques d'analyse de données, peut être abordé sous l'angle algorithmique dans le but d'améliorer l'état de l'art dans ce domaine de détermination de modèles graphiques et également théorique d'un point de vue géométrique et de théorie de l'information.

Un pointeur

- J. Friedman, T. Hastie, R. Tibshirani, *Sparse inverse covariance estimation with the graphical lasso*, Biostatistics, 3 Vol 9 pp 432–441 (2008).