

Inférence de réseaux biologiques à partir de données hétérogènes par apprentissage de modèles à noyaux multiples

Arnaud Fouchet*¹, Jean-Marc Delosme¹, et Florence d'Alché-Buc^{1,2}

¹IBISC, EA 4526, Université d'Evry-Val d'Essonne, 23 Bd de France, Evry, France

²INRIA-Saclay, TAO/AMIB, LRI umr CNRS 8623, Université Paris Sud, Orsay, France

30 mai 2013

Résumé

Les techniques de criblage du transcriptome à grande échelle ont ouvert la voie à l'inférence de réseaux de régulation à partir de données d'expression de gènes. Les niveaux d'expression des gènes, mesurés au cours du temps et dans des conditions expérimentales données, sont les variables d'état de ces systèmes dynamiques. Lorsqu'une de ces variables converge vers une valeur d'équilibre, cette valeur fournit une donnée statique, tandis que des valeurs prises lors du régime transitoire fournissent des données dynamiques. La plupart des méthodes d'inférence de réseaux de régulation utilisent soit l'un soit l'autre type de données.

Nous proposons un cadre permettant d'apprendre à partir de ces deux types de données et, pour inférer la régulation d'un gène par d'autres gènes, nous utilisons la régression à noyaux multiples avec sélection de noyaux. Nous construisons, grâce à une nouvelle méthode d'apprentissage, une régression non-linéaire dépendant de peu de variables et avec un faible temps de calcul.

L'approche a été testée sur des organismes réels et des organismes simulés de manière réaliste (données du challenge DREAM5). Les performances obtenues sont du niveau de l'état de l'art et les réseaux inférés présentent une certaine complémentarité par rapport à ceux obtenus par d'autres méthodes. Ainsi, en combinant les réseaux obtenus par un ensemble de méthodes incluant la nôtre, des performances supérieures sont obtenues pour l'inférence de réseaux de régulation.

*afouchet@ibisc.fr