

Bayesian Networks for detecting relevant variable interactions for biochemical experiments

Reti Bayesiane per l'individuazione di interazioni rilevanti fra variabili in esperimenti biochimici

Debora Slanzi¹

Dipartimento di Statistica, Università Ca' Foscari, e-mail: debora@unive.it

Irene Poli

Dipartimento di Statistica, Università Ca' Foscari,
e European Center for Living Technology, e-mail: irenepoli@unive.it

Davide De March

European Center for Living Technology, e-mail: davidedemarch@unive.it

Michele Forlin

Dipartimento di Statistica, Università Ca' Foscari, e-mail: mforlin@unive.it

Riassunto: La scelta del disegno sperimentale per la costruzione di sistemi previsivi in ambito biochimico richiede spesso la selezione di fattori in insiemi ad alta dimensionalità. In questo lavoro si affronta il tema dell'identificazione dei fattori e delle loro interazioni adottando l'approccio delle reti bayesiane, che descrivono la struttura probabilistica fra i fattori, con la rappresentazione delle relazioni di dipendenza e indipendenza condizionata. I risultati evidenziano il ruolo dominante di una variabile sia nell'effetto principale, sia nelle interazioni con un sottogruppo dell'insieme iniziale dei fattori.

Keywords: Experiment designs, Bayesian networks, variable interaction.

In designing experiments we frequently encounter the problem of choosing a set of factors and their interactions to achieve a particular functionality for the system and formulate accurate predictions for unknown combinatorial compositions. Biochemical experiments are generally characterized by a large number of factors with a little prior knowledge on the factorial effects. Conventional fractional factorial experiments handle with difficulties the problem of the high dimensional search space, and factors are frequently canceled in the design without a deep understanding of their effects. For the same reason main effects analysis are only conducted ignoring the joint action of the factors, with the consequence of deriving unreliable predictive models. In this study we tackle the problem of identifying main effects and interactions among factors by adopting a class of probabilistic graphical models, namely the class of Bayesian Networks (Jensen, 2001). Bayesian networks provide a probabilistic description of the conditional dependence relation that can describe a set of variables proposed to derive an experimental design and a predictive model. A Bayesian Network, BN for short, specifies a multivariate probability distribution over a set of random variables through two components:

- the *network topology*, consisting of a set of nodes, which identify the random variables (experimental factors), and of arcs, which represent direct dependences between variables, and are described by direct arrows between nodes. This topology is generally named directed acyclic graph (DAG);
- the *network probability distributions*, consisting of a set of conditional probability distributions which describe the relations of dependence between nodes.

¹Address of correspondence: Dipartimento di Statistica, San Giobbe, Cannaregio 873, Venezia

