Special feature: probabilistic graphical models and its applications to biomedical informatics—part 1

Joe Suzuki[1] · Brandon Malone[2]

Recent advances in low-cost, high-throughput measurement technologies, such as RNA-seq, have brought Big Data to the world of health care. Machine learning and artificial intelligence approaches are among the most promising techniques for extracting useful biological signals from the noise inherent in these technologies. Integrating such data with traditional medical data poses even more challenges. Nevertheless, this field of research is poised to offer effective, personalized medical treatment.

This special feature focuses on probabilistic graphical modeling approaches to biomedical informatics. A wide variety of problems remain open in this domain. For example, regulatory and interaction networks can naturally be expressed as Bayesian or Markov networks; however, integrating heterogeneous data types, such as single nucleotide polymorphism information and RNA-seq, is often approached in an ad hoc, problem-dependent manner. Biological signals are known to exhibit complex, time-dependent and often non-acyclic dependencies; rigorously modeling such dynamic processes remains an open challenge. Furthermore, many biological problems suffer from "large p, small n," in which many more variables (such as genes) are present than the number of samples. These settings remain challenging for typical machine learning approaches.

Many difficulties remain in medical informatics, as well. For example, many hospitals maintain databases of patient information; however, the information for individual patients is typically sparse, and sometimes even incorrect. Extracting useful knowledge from such unstructured data sources often requires probabilistic processing techniques. Similarly, many health care professionals do not have experience interpreting machine learning results. Thus, information retrieval and

✉ Joe Suzuki
 suzuki@math.sci.osaka-u.ac.jp

[1] Graduate School of Science, Osaka University, Osaka, Japan

[2] Heidelberg University Hospital, Heidelberg, Germany

visualization techniques are also relevant, open questions for biomedical informatics.

This special feature will be published in two parts. Part 1 in this issue includes four invited papers. Professors Michio Yamamoto and Kei Hirose deal with factor analysis using their R package fanc and explain why the notion of sparseness is essential in their article "Graphical tool of sparse factor analysis"; Professors Shizue Izumi and Masataka Taguri address structural mean models (SMMs) for estimating causal treatment effects in the presence of non-ignorable non-compliance in clinical trials and propose the goodness-of-fit (GOF) test for the SMMs in their article "A global goodness-of-fit test for linear structural mean models"; Dr. Song Liu, and Professors Kenji Fukumizu and Taiji Suzuki consider detecting changes in a Markov network using a density ratio approach in their article "Learning Sparse Structural Changes in High-dimensional Markov Networks: A Review on Methodologies and Theories"; and Professors Yangbo He, Jinzhu Jia, and Zhi Geng consider causal network models, in particular, the model space of causal networks, decomposition learning of structures from observational data, local structural learning approaches, and the active learning for optimal designs of intervention in their article "Structural Learning of Causal Networks."

We hope those inspiring papers contribute the development of probabilistic graphical models and its applications to biomedical informatics.