Editorial

Flow of Information in Biological Systems

Information theory, as developed in 1948 by Claude Shannon, was primarily an instrument for understanding the transmission of electronic signals and modeling communication systems. In recent decades, information theory has branched out to take on topics from many diverse research areas, such as computer science, physics, economics, sociology, neuroscience, and most recently, bioinformatics and systems biology. An important type of signal processing and chemical information characterizes biological systems, and information theory can be applied to model biological processes and manage the intrinsic uncertainty involved in biological information. For modeling and handling cells, which are non-linear time-varying living systems and the complicated processes within and between them, classical information theory has to be extended to take into account the biological context in its diverse aspects. To cope with the degree of complexity involved, information theoretic concepts such as entropy and uncertainty measures, information divergence measures, mutual information, noisy-channel coding, etc. have been extended and/or generalized.

One could perhaps say that in practice the introduction of information theory into the life sciences began with contributions to medical imaging, made possible by the application of concepts such as mutual information (MI) used, for example, in the optimal alignment of CT and MRI images by maximizing their MI. Another field of application of Shannon's information theory has been in the quantification and analysis of variations in nucleic acid sequences. However, it has not been until recently that we have witnessed a second wave of applications of the theory, this time to network biology. Information theory can help quantify the expected and therefore missing and redundant information in biological data, and as such has the potential to assist researchers and data analysts, and this approach has been notably exploited in addressing the challenge of reverse engineering in understanding and analyzing biological networks.

Entropy has proven to be successful in various areas, and is of particular interest in biology, because on the one hand uncertainty is inherent to biology, and is part and parcel of the challenge of analyzing and understanding noisy and partial data, and on the other, the problem of causation versus correlation needs to be confronted. This latter has assumed central importance in genetics and molecular biology with the advent of huge amounts of data from high throughput sequencing, which requires methods to reconstruct underlying interactions from partial observations in order to unveil the causal mechanisms.

Despite the many possibilities opened up by information theory, the existing approaches have been limited in genomics and molecular biology. Although information theory is a theoretical framework that can be adapted to a wide range of problems, there are some undeniable challenges that should be considered before any analysis is undertaken. For example, the empirical estimation of probability mass functions for random variables and the construction of joint probability distributions for a group of random variables are tasks that are often challenging and not at all straightforward. Furthermore, the computations will become more complicated when a time delay is involved.

This special issue offers a review and overview of specific applications of information and algorithmic information theory, particularly to network biology. In this issue, Mousavian et al. [1,2] offer a comprehensive review that demonstrates the wide applicability of information theory in systems biology, emphasizing various types of biological networks. Their review is in two parts, each of which surveys applications of information theory to two types of biological networks. To this end, biological systems have been classified into 4 main classes based on the type of system components and the interactions between them: gene regulatory, metabolic, protein-protein interaction and signaling networks. In the first part [1], the emphasis is on gene regulatory and metabolic networks. In gene regulatory networks, information theoretic measures, especially mutual information, have been applied in network reconstruction. Furthermore, the dynamical behavior of metabolic systems and the robustness of such networks against perturbations have also been investigated using information theory. In the second part of this review [2], the information-theoretic studies reviewed focus on protein-protein interaction and signaling networks. In protein-protein interaction networks, various problems, including complex identification, complexity analysis, and the determination of subnetwork markers, among others, have been addressed using information theory. Moreover, the potentialities of information theoretic approaches in studying the rate of signal transmission in signaling networks and estimating the channel capacity in such networks, is explored.

Udo and Kuroda [3] review aspects of the theory recently applied to signal transduction, where mutual information regarding signal transduction has been determined to be a measure of information transmission. They provide an overview of how much information is transmitted in signaling pathways and how signal transduction transmits informational input and shapes biological output.

In Kiani et al. [4], tools (beyond statistical and graph-theoretical ones) for evaluating reconstruction methods are reviewed using information theory itself, both Shannon Entropy and algorithmic complexity. The field of algorithmic information theory (AIT) also

provides a set of complexity measures to gauge the information content of sequences and other objects, as shown by Zenil et al. [5]. It captures notions that were once considered to be only informal (e.g. simplicity and complexity), but in this issue Zenil et al. and Kiani et al. identify advances in the application of AIT to network biology, both in analyzing static and evolving networks, and in evaluating the results of other methods of reconstructing data. The paper by Zenil et al. addresses the suitability of mathematical measures of complexity, notably measures of algorithmic information content, to designing and applying new algorithms and tools in the area of genomics and bioinformatics, so as to enable us to address some of the most important challenges in genomics related to genome mapping and network profiling by means of the most powerful mathematical tools for pattern-detection (algorithmic information theory).

References

[1] Z. Mousavian, K. Kavousi, A. Masoudi-Nejad, Information theory in systems biology: Part I: Gene regulatory and metabolic networks, Semin. Cell Dev. Biol. 51 (2016) 3–13.

[2] Z. Mousavian, J. Díaz, A. Masoudi-Nejad, Information theory in systems biology. Part II: Protein–protein interaction and signaling networks, Semin. Cell Dev. Biol. 51 (2016) 14–23.

[3] S. Uda, S. Kuroda, Analysis of cellular signal transduction from an information theoretic approach, Semin. Cell Dev. Biol. 51 (2016) 24–31.

[4] N.A. Kiani, et al., Evaluating network inference methods in preserving the topology and complexity of reconstructed genetic networks, Semin. Cell Dev. Biol. 51 (2016) 44–52.

[5] H. Zenil, N.A. Kiani, J. Tegnér, Methods of information theory and algorithmic complexity for network biology, Semin. Cell Dev. Biol. 51 (2016) 32–43.

Ali Masoudi-Nejad
Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran

Hector Zenil
Unit of Computational Medicine, Science for Life Laboratory (SciLifeLab), Department of Medicine Solna, Center for Molecular Medicine, Karolinska Institute, Stockholm, Sweden
E-mail addresses: amasoudin@ibb.ut.ac.ir
(A. Masoudi-Nejad), hector.Zenil@ki.se (H. Zenil).