Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Methods 129 (2017) 1–2

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Guest Editor's Introduction

Machine learning methods and systems for data-driven discovery in biomedical informatics

In the current era of big data, transforming biomedical big data into valuable knowledge has become one of the most important challenges in biomedical informatics. Simultaneously, various machine learning techniques have advanced rapidly and are now showing state-of-the-art performance in various fields. Consequently, applying machine learning to biomedical informatics is of great interest to both academia and industry. The objective of this special issue is to highlight exciting recent advances in applying machine learning methods and systems to biomedical informatics.

One of the most popular applications of machine learning in biomedical informatics may be computational genomics. Soylev et al. develop a method that combines diverse signatures such as read-pair, read-depth, and split-read into a single package to characterize most SV types, allowing easy and accurate SV discovery [1]. Jiao et al. propose a new guilt-by-approach to infer metagenomic data for predicting potentially novel pathways and functional association between proteins [2]. They demonstrate that microbial community profiling outperforms phylogenetic profiling, and reveals more functional associations. Lee et al. present a novel approach for detecting associations among genotypes, transcript, and phenotypes [3]. They applied the method to an Alzheimer's disease dataset and found genotype (shown by single nucleotide polymorphism)–transcript–phenotype (disease status) associations. Genome structural variations (SVs) are genomic alterations of >50 bp in size and play major roles in genome evolution and pathogenesis of diseases of genomic origin.

This special issue includes two articles on interesting applications of machine learning techniques to viral research. Deletions of hepatitis B virus (HBV) are associated with the development of progressive liver diseases leading to hepatocellular carcinoma (HCC). Accordingly, detecting the exact breakpoints of deletion with characteristics of HBV genome sequences from next-generation sequencing (NGS) outputs is critical for improving the prognosis and treatment of liver disease. Cheng et al. propose a novel analytical method named VirDelect (Virus Deletion Detect), which finds exact breakpoints of deletion effectively and efficiently, outperforming the latest state-of-the-art methods [4]. Kang et al. present a machine learning based method to predict candidate interactions between viral microRNA and endogenous human microRNA sponges, which bind to and inhibit microRNA [5]. Through computational prediction and experimental validation using luciferase reporter assay, western blot, and flow cytometry, a potential natural miRNA sponge that acts against microRNA derived from Kaposi's sarcoma-associated herpesvirus has been found.

Large-scale networks frequently occur in biomedical informatics as a means to represent complex interactions between multiple entities. Machine learning offers a set of effective tools to analyze large-scale biomedical data represented in networks. Ou-Yang et al. developed a node-based multi-view differential network analysis model to infer differential networks from multi-platform gene expression data [6]. They applied the model to real TCGA ovarian cancer samples and identified network rewiring associated with drug resistance. Using large-scale machine learning techniques, Choi et al. analyzed the emotional public response to a nationwide outbreak of Middle East respiratory syndrome (MERS) in Korea [7]. They collected mass media outlet data during the outbreak in 2015 and discovered an intriguing loop of information transfer between the media and the public. This method will be helpful for alleviating the unnecessary fear and overreaction of the public regarding infectious diseases.

Machine learning techniques can also be applied to data-driven pharmaceutical research. To find synergistic drug pairs, Chua et al. designed MASCOT, which leverages a machine learning based target prioritization method and the Loewe heuristic from pharmacology. MASCOT efficiently predicts synergistic target combinations with desired therapeutic effects and minimum off-target effects in a disease-related signaling network [8]. Ensemble learning algorithms and dimensionality reduction techniques were also used to predict drug-target interactions [9]. The authors applied three dimensionality reduction methods to find relevant features and applied ensemble models of decision trees and kernel ridge regression, resulting in significant improvement of drug–target interaction prediction.

Modern machine learning requires huge amounts of data to uncover underlying biological assumptions and principles that are otherwise difficult to find using conventional techniques. One of the most effective and realistic ways to generate and gather such a large volume of data relies on biological sensors, and the study by Sanzo et al. is one such example [10]. They propose a bimetallic biosensor composed of nanocoral Au decorated with Pt nanoflowers for H2O2 detection at low potentials, offering new perspectives for creating innovative glucose monitoring systems.

Various types of medical systems produce large volumes of time-series data, which can be effectively analyzed by machine learning techniques. Examples include electroencephalography (EEG), a standard non-invasive technique widely used in neural disease diagnosis and neuroscience. In particular, frequency-tagging (FT) is a technique used to measure EEG responses to stimuli. For automated analysis of FT responses in EEG, Montagna et al. propose a machine learning based pattern recognition

https://doi.org/10.1016/j.ymeth.2017.09.011
1046-2023/

technique, delivering performance with more than 90% accuracy [11].

In summary, this issue highlights recently proposed machine learning methods and systems that include innovations in computational genomics, viral research, network analysis, biosensors and monitoring systems, and biomedical measurement systems. Given the plethora of biomedical data that cannot be analyzed without computational methods and systems, we anticipate that numerous additional approaches based on machine learning will emerge to accelerate data-driven discoveries in biomedical informatics.

References

[1] Fereydoun Hormozdiari, Can Alkan, Arda Soylev, Can Kockan, Toolkit for automated and rapid discovery of structural variants, Methods 129 (2017) 3–7.
[2] Dazhi Jiao, Wontack Han, Yuzhen Ye, Functional association prediction by community profiling, Methods 129 (2017) 8–17.
[3] Seunghak Lee, Haohan Wang and Eric P. Xing, Backward Genotype-Transcript-Phenotype Association Mapping, Methods 129 (2017) 18–23.
[4] Ji-Hong Cheng, Wen-Chun Liu, Ting-Tsung Chang, Sun-Yuan Hsieh, Vincent Tseng, Detecting exact breakpoints of deletions with diversity in hepatitis B viral genomic DNA from next-generation sequencing data, Methods 129 (2017) 24–32.
[5] Soowon Kang, Seunghyun Park, Sungroh Yoon, Hyeyoung Min, Machine learning-based identification of endogenous cellular microRNA sponges against viral microRNAs, Methods 129 (2017) 33–40.
[6] Le Ou-Yang, Xiao-Fei Zhang, Min Wu, Xiaoli Li, Node-based learning of differential networks from multi-platform gene expression data, Methods 129 (2017) 41–49.
[7] Sungwoon Choi, Jangho Lee, Min-Gyu Kang, Hyeyoung Min, Yoon-Seok Chang, Sungroh Yoon, Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks, Methods 129 (2017) 50–59.
[8] Huey Eng Chua, Sourav Saha Bhowmick, Lisa Tucker-Kellogg, Synergistic target combination prediction from curated signaling networks: machine learning meets systems biology and pharmacology, Methods 129 (2017) 60–80.
[9] Ali Ezzat, Min Wu, Xiaoli Li, Chee Keong Kwoh, Drug-target interaction prediction using ensemble learning and dimensionality reduction, Methods 129 (2017) 81–88.
[10] Gabriella Sanzò, Irene Taurino, Francesca Puppo, Riccarda Antiochia, Lo Gorton, Gabriele Favero, Franco Mazzei, Sandro Carrara, Giovanni De Micheli, A bimetallic nanocoral Au decorated with Pt nanoflowers (bio) sensor for H2O2 detection at low potential, Methods 129 (2017) 89–95.
[11] Fabio Montagna, Marco Buiatti, Simone Benatti, Davide Rossi, Elisabetta, Farella, Luca Benini, A machine learning approach for automated wide-range frequency tagging analysis in embedded neuromonitoring systems, Methods 129 (2017) 96–107.

Guest editors
Sungroh Yoon
Seunghak Lee
Wei Wang