OPEN ACCESS Freely available online

ISSN: 2153-0602
Journal of
Data Mining in
Genomics & Proteomics

Perspective

Machine Learning Methods for Investigating Protein Functionalities

Eitan Mrozek*

Department of Bioinformatics, Silesian University of Technology, Gliwice, Poland

DESCRIPTION

Machine learning techniques are powerful tools for predicting the function of genes, especially when experimental validation is difficult or costly. Gene function prediction can help us understand the roles of coding and non-coding genes in various biological processes, such as gene regulation, protein interaction, disease development and differentiation. Machine learning techniques can leverage various types of data, such as sequence, structure, expression, interaction and annotation, to infer gene function from patterns and relationships. The function of genes, both coding and non-coding, is often hard to determine by experimental methods in molecular biology. Hence, computational methods, which frequently use machine learning, can be helpful for guiding and predicting function. Machine learning has been seen as a "black box" before, but it can be more precise than simple statistical testing methods. Lately, deep learning and big data machine learning techniques have grown quickly and reached an impressive level of performance in many areas, such as image classification and speech recognition. This Research Topic examines the possibility of machine learning for gene function prediction.

There are many machine learning techniques that have been applied to gene function prediction, such as Support Vector Machines (SVM), Artificial Neural Networks (ANN), decision trees (DT), Random Forests (RF), K-Nearest Neighbors (KNN), Bayesian Networks (BN) and Deep Learning (DL). These techniques can be classified into supervised, unsupervised and semi-supervised methods, depending on whether they use labelled or unlabelled data for training and testing. Supervised methods require a large amount of labelled data to learn a function that maps input features to output labels. Unsupervised methods do not use any labels, but instead cluster or group similar data points based on their features. Semi-supervised methods use a combination of labelled and unlabelled data to improve the performance of supervised or unsupervised methods.

Challenges and opportunities for machine learning techniques on gene function prediction

Data quality and quantity is the availability and reliability of data sources for gene function prediction varies widely across different organisms, conditions and domains. Machine learning techniques need to deal with issues such as noise, missing values, imbalance, redundancy and heterogeneity of data. Moreover, machine learning techniques need to cope with the high dimensionality and sparsity of data, as well as the complexity and diversity of gene functions.

Data integration and fusion is the function prediction can benefit from integrating and fusing multiple types of data, such as sequence, structure, expression, interaction and annotation. Machine learning techniques need to develop effective methods for combining different data sources in a coherent and consistent way, while avoiding information loss or conflict. Moreover, machine learning techniques need to exploit the complementarity and synergy of different data sources to enhance the accuracy and coverage of gene function prediction.

Interpretability and explainability: Machine learning techniques often produce black-box models that are difficult to interpret and explain. Gene function prediction requires machine learning techniques that can provide meaningful and understandable results that can be validated by biological knowledge and experiments. Moreover, machine learning techniques need to provide confidence measures and uncertainty estimates for their predictions, as well as identify novel or unknown gene functions that require further investigation.

Transferability and generalizability: Machine learning techniques often suffer from over fitting or under fitting problems when applied to new or unseen data. Gene function prediction requires machine learning techniques that can transfer and generalize their models across different organisms, conditions and domains. Moreover, machine learning techniques need to adapt and update their models in response to

Correspondence to: Eitan Mrozek, Department of Bioinformatics, Silesian University of Technology, Gliwice, Poland, E-mail: eitanzek007@edu.pl

Received: 03-Jan-2023, Manuscript No. JDMGP-23-20676; **Editor assigned:** 06-Jan-2023, JDMGP-23-20676 (PQ); **Reviewed:** 20-Jan-2023, QC No. JDMGP-23-20676; **Revised:** 27-Jan-2023, Manuscript No. JDMGP-23-20676 (R); **Published:** 03-Feb-2023, DOI: 10.4172/2153-0602.23.14.281

Citation: Mrozek E (2023) Machine Learning Methods for Investigating Protein Functionalities. J Data Mining Genomics Proteomics. 14:281.

new or changing data. Machine learning techniques on gene function prediction are an active and promising research area that can advance our understanding of the molecular mechanisms of life. By developing novel methods and applications for gene function prediction, machine learning techniques can contribute to various fields of biology, medicine and biotechnology.