The Power of Machine Learning in the Biological Context

Johannes Stübinger

Department of Statistics and Econometrics, University of Erlangen-Nürnberg, Germany

Submission: April 23, 2019; **Published:** June 06, 2019

*Corresponding author: Johannes Stübinger, Department of Statistics and Econometrics, University of Erlangen-Nürnberg, Germany

Abstract

In the recent past, both the rapid growth of big data and the exponential increase in computing power have enabled the use of Machine Learning. In biology, too, this type of artificial intelligence finds very great accusation, as it opens new fields of research. Therefore, this paper provides a comprehensive overview of Machine Learning in biology by consolidating and organizing the extensive literature available in this field of research.

Keywords: Machine learning; Biology; Big data; Supervised learning; Unsupervised learning; Reinforcement learning; Artificial intelligence

Introduction

Machine Learning describes the process of creating artificial intelligence from experience without being explicitly programmed [1]. Therefore, Machine Learning enables IT systems to recognize patterns and structures based on existing data. The acquired know-how of complex relations is generalized and used for new challenges or for the analysis of previously unknown data. Within the recent past, both the rapid growth of large amounts of data and the exponential increase in computing power permit the application of Machine Learning [2]. Therefore, academic interest in studying state-of-the-art algorithms and statistical models has surged over the past years. In general, Machine Learning is divided in three substreams. Supervised Learning aims at a structural mapping of the data by determining a function based on training data marked by labels. In contrast, Unsupervised Learning recognizes patterns in data that is not marked by labels. Reinforcement Learning describes the process of learning an autonomous strategy in order to maximize the expected future reward. The algorithms are deployed successfully in various areas of our lives, e.g., self-driving cars [3], speech and text recognition [4], identification of capital market anomalies [5], and the prediction of soccer match results [6]. This manuscript supplies a comprehensive overview of Machine Learning in biology by consolidating and organizing the extensive literature available in this research field.

Supervised learning

Supervised Learning aims at finding a function that perfectly maps an input object to a desired output object based on exemplary input-output pairs. In the optimal case the algorithm is able to always correctly determine the output of unseen inputs. If the output is a discrete (continuous) feature, we apply a classification (regression).

In the classification field, Neumann et al. [7] establish a phenotypic high-throughput screening platform combining potent gene silencing by ribonucleic acid interference, time-lapse microscopy and computational image processing. In addition, Bleakley et al. [8] use classification algorithms to derive and reconstruct biological networks from heterogeneous data. In the same research area, Sionov et al. [9] employ decision trees to find out which genes in a network affect others.

In the regression field, Frontzek et al. [10] show the power of support vector machines for predicting the nonlinear dynamics of individual biological neurons. The prediction of protein accessible surface areas from their primary structures is conducted by Yuan et al. [11].

Unsupervised learning

In contrast, Unsupervised Learning handles unlabeled data, i.e., we only possess inputs and while cluster analysis groups the available data according to similar inputs, dimensionality reduction reduces the complexity in the data set by identifying and removing variables without information content.

An important component of cluster analysis is the reaction of cells as a result of drug consumption. Loo et al. [12] explain a heterogeneous cell response by a dynamic mixture of subpopulations with different physiological characteristics. Similarly, Singh et al. [13] show that patterns of signal heterogeneity may reveal functional differences between cell populations.

Horn et al. [14] use dimensionality reduction to detect interactions between signaling factors affecting different quantitative phenotypes of Drosophila Melanogaster cells. Consequently, they were able to reconstruct signaling pathways and identify a conserved Ras-MAPK signaling regulator.

Reinforcement learning

Finally, Reinforcement Learning describes a process in which the algorithm autonomously learns a strategy to maximize expected future rewards. Specifically, the agent, i.e., computer program capable of independent and selfdynamic behavior, continuously acquires knowledge by interacting with his environment which contains all information about possible states and rewards. Problems are modeled by statistical approaches (Markov decision processes, Monte Carlo simulations, dynamic programming) and solved with the aid of trial and error.

Reinforcement Learning is successfully utilized within the Brain-Machine Interfaces. DiGiovanna et al. [15] propose an algorithm that learns complete routines from interaction with its environment instead of an explicit training signal. Mahmoudi et al. [16] utilize both a continuous perception-action-reward cycle and the motor commands and information from the brain to decipher the intended actions of the users. Wang et al. [17] develop a reinforcement learning model based on a quantized attention-gated kernel - the high success rates indicate a powerful decoding capability for more demanding BMI tasks.

Analyzing medical images is a very important field of application for reinforcement learning. Sahba et al. [18] present an algorithm that estimates the location and volume of the prostate in transrectal ultrasound images. Nixon et al. [19] consolidate the implementation of medical imaging and computer vision techniques.

The operating mode of nervous systems can be modeled via reinforcement learning. The results of Frank et al. [20] demonstrate a neurocomputational dissociation between striatal and prefrontal dopaminergic mechanisms in reinforcement learning. In addition, genetic analyses show that they contribute to rewarding and avoiding human learning.

Conclusion

This manuscript provides a comprehensive literature review of Machine Learning in the context of biology. The subsections Supervised Learning, Unsupervised Learning and Reinforcement Learning are structured, and the corresponding literature is explained. In the future, Machine Learning will certainly become even more important in biology. Not least because it could lead to other vital solutions, such as the cultivation of high yielding, drought-tolerant plants or the modification of cancer cells.

References

1. Michie D, Spiegelhalter DJ, Taylor CC (1994) Machine Learning, Neural and Statistical Classification, Ellis Horwood Ltd.

2. Al-Jarrah OY, Yoo PD, Muhaidat S, Karagiannidis GK, Taha K (2015) Efficient machine learning for big data: A review. Big Data Research 2(3): 87-93.

3. Stilgoe J (2018) Machine learning, social learning and the governance of self-driving cars. Social studies of science 48(1): 25-56

4. Alm CO, Roth D, Sproat R (2005) Emotions from text: machine learning for text-based emotion prediction. In Proceedings of the conference on human language technology and empirical methods in natural language processing. Association for Computational Linguistics Pp. 579-586.

5. Knoll J, Stübinger J, Grottke M (2019) Exploiting social media with higher-order factorization machines: Statistical arbitrage on high- frequency data of the S&P 500. Quantitative Finance 19(4): 571-585.

6. Stübinger J, Knoll J (2018) Beat the Bookmaker–Winning Football Bets with Machine Learning (Best Application Paper). In International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer, Cham, Pp: 219-233.

7. Neumann B, Walter T, Hériché JK, Bulkescher J, Erfle H, et al. (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 464(7289): 721-727.

8. Bleakley K, Biau G, Vert JP (2007) Supervised reconstruction of biological networks with local models. Bioinformatics 23(13): 57-65.

9. Soinov LA, Krestyaninova MA, Brazma A (2003) Towards reconstruction of gene networks from expression data by supervised learning. Genome biology 4(1): R6.

10. Frontzek T, Lal TN, Eckmiller R (2001) Predicting the nonlinear dynamics of biological neurons using support vector machines with different kernels. In IJCNN'01. International Joint Conference on Neural Networks Pp. 1492-1497.

11. Yuan Z, Huang B (2004) Prediction of protein accessible surface areas by support vector regression. Proteins 57(3): 558-564.

12. Loo LH, Lin HJ, Singh DK, Lyons KM, Altschuler SJ, et al. (2009) Heterogeneity in the physiological states and pharmacological responses of differentiating 3T3-L1 preadipocytes. J Cell Biol 187(3): 375-384.

13. Singh DK, Ku CJ, Wichaidit C, Steininger RJ, Wu LF, et al. (2010) Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. Mol Syst Biol 6(1): 369.

14. Horn T, Sandmann T, Fischer B, Axelsson E, Huber W, et al. (2011) Mapping of signaling networks through synthetic genetic interaction analysis by RNAi. Nat Methods 8(4): 341-346.

15. DiGiovanna J, Mahmoudi B, Fortes J, Principe JC, Sanchez JC (2009) Coadaptive brain–machine interface via reinforcement learning. IEEE transactions on biomedical engineering 56(1): 54-64.

16. Mahmoudi B, Sanchez JC (2011) A symbiotic brain-machine interface through value-based decision making. PloS one 6(3): 1-14.

17. Wang F, Wang Y, Xu K, Li H, Liao Y, et al. (2017) Quantized attention-gated kernel reinforcement learning for brain – machine interface decoding. IEEE transactions on neural networks and learning systems 28(4): 873-886.

18. Sahba F, Tizhoosh HR, Salama MM (2008) Application of reinforcement learning for segmentation of transrectal ultrasound images. BMC Med Imaging 8(8): 1-10.

19. Nixon M, Aguado AS (2012) Feature extraction and image processing for computer vision. Academic Press.

20. Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A 104(41): 16311-16316.

**Your next submission with Juniper Publishers
will reach you the below assets**

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats
  (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission

https://juniperpublishers.com/online-submission.php