Ten Simple Rules for Deep Learning in Biology

This manuscript (permalink) was automatically generated from Benjamin-Lee/deep-rules@c87caf4 on December 10, 2018.

Authors

- **Benjamin D. Lee**
  0000-0002-7133-8397 · Benjamin-Lee
  School of Engineering and Applied Sciences, Harvard University; Department of Genetics, Harvard Medical School; Lab41, In-Q-Tel

Introduction

Deep learning (DL), a subfield of machine learning (ML) that focuses on deep artificial neural networks, is exploding in popularity and is increasingly being used for biological data analysis [1]. However, DL itself remains an active area of research, and the complexity of this field poses a large barrier of entry to those who would like to correctly utilize state-of-the-art DL technology in biological research applications. While much has been written about the impressive results of DL in biology, there is a comparative dearth of literature articulating best practices for its use, with most instructional literature focusing on ML, rather than DL [2].

To fix this problem, we solicited input from a diverse community of researchers, who wrote this manuscript collaboratively using the GitHub version control platform [3] and Manubot [4]. In the course of our discussions, several themes became clear: the importance of understanding and applying ML fundamentals as a baseline for utilizing DL, the necessity for extensive model comparisons and careful evaluation, and the need for critical thought in interpreting results generated by means of DL, among others. The rules we established range from high-level guidance to the implementation of best practices, and it is our hope that these rules will provide actionable, DL-specific advice for both new and experienced DL practitioners alike who would like to employ DL in biological research. By increasing the accessibility of DL techniques to biology, we aim to improve the overall quality and reproducibility of DL in the literature, enabling these powerful methods to be properly utilized to generate new scientific insights.

Rule 1: Concepts that apply to machine learning also apply to deep learning

Rule 2: Use traditional methods to establish performance baselines

Rule 3: Understand the complexities of training deep neural networks

Rule 4: Know your data and your question

Rule 5: Choose an appropriate neural network architecture and data representation

Rule 6: Tune your hyperparameters extensively and systematically

Rule 7: Address deep neural networks' increased tendency to overfit the dataset

Rule 8: Do not necessarily consider a DL model as a black box

Rule 9: Interpret predictions in the correct manner

Rule 10: Don't share models trained on sensitive data

Conclusion

References

1. **Opportunities and obstacles for deep learning in biology and medicine**
Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M. Hoffman, … Casey S. Greene
Journal of The Royal Society Interface (2018-04) https://doi.org/gddkhn
DOI: 10.1098/rsif.2017.0387 · PMID: 29618526 · PMCID: PMC5938574

2. **Ten quick tips for machine learning in computational biology**
Davide Chicco
BioData Mining (2017-12) https://doi.org/gdb9wr
DOI: 10.1186/s13040-017-0155-3 · PMID: 29234465 · PMCID: PMC5721660

3. https://github.com/Benjamin-Lee/deep-review

4. https://greenelab.github.io/meta-review/