Deep learning for population size history inference: design, comparison and combination with approximate Bayesian computation

Flora Jay

Paris-Sud University/Paris-Saclay/CNRS/INRIA.

17 September 2020, 11.30am UK time

Abstract

In this talk, I will present the recent practical study conducted with my colleagues Théophile Sanchez, Jean Cury and Guillaume Charpiat on using and combining deep learning and ABC for demographic inference based on genomic data [1].

For the past decades, simulation-based likelihood-free inference methods have enabled researchers to address numerous population genetics problems. As the richness and amount of simulated and real genetic data keep increasing, the field has a strong opportunity to tackle tasks that current methods hardly solve. However, high data dimensionality forces most methods to summarize large genomic datasets into a relatively small number of hand-crafted features (summary statistics). Here we propose an alternative to summary statistics, based on the automatic extraction of relevant information using deep learning techniques. Specifically, we design artificial neural networks (ANNs) that take as input single nucleotide polymorphic sites (SNPs) found in individuals sampled from a single population and infer the past effective population size history. First, we provide guidelines to construct artificial neural networks that comply with the intrinsic properties of SNP data such as invariance to permutation of haplotypes, long scale interactions between SNPs and variable genomic length. Thanks to a Bayesian hyperparameter optimization procedure, we evaluate the performance of multiple networks and compare them to well established methods like Approximate Bayesian Computation (ABC). Even without the expert knowledge of summary statistics, our approach compares fairly well to an ABC approach based on handcrafted features. Furthermore, we show that combining deep learning and ABC can improve performance while taking advantage of both frameworks. Finally, we apply our approach to reconstruct the effective population size history of cattle breed populations.

References

[1] T. Sanchez, J. Cury, G. Charpiat, F. Jay. Deep learning for population size history inference: Design, comparison and combination with approximate Bayesian computation. Mol Ecol Resour. 2020; 00: 1–16. https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13224