BRUNO: A Deep Recurrent Model for Exchangeable Data

Iryna Korshunova\textsuperscript{1}  Jonas Degrave\textsuperscript{1}  Ferenc Huszár\textsuperscript{2}  Yarin Gal\textsuperscript{2}  Arthur Gretton\textsuperscript{4}  Joni Dambre\textsuperscript{1}
\textsuperscript{1}Ghent University  \textsuperscript{2}University of Oxford  \textsuperscript{3}Equal contribution  \textsuperscript{4}Now at DeepMind

Overview

BRUNO is a versatile meta-learning model that combines the expressiveness of deep neural networks with the data-efficiency of GPs to model exchangeable sequences of high-dimensional, complex observations like images.

BRUNO is exchangeable by construction, meaning that its joint distribution $p(x_1, \ldots, x_n)$ is permutation-invariant. As a consequence, BRUNO carries out an exact Bayesian inference, albeit implicitly.

BRUNO enjoys some properties that are desirable in practice:
- Predictive distribution $p(x_{n+1}|x_1, \ldots, x_n)$ is fast to evaluate
- $p(x_1|x_{n+1})$ is easy to sample from
- $p(x_{n+1}|x_1, \ldots, x_n)$ is differentiable with respect to the model parameters
- Can be trained efficiently in an RNN-like fashion

Exchangeability and Bayesian computations

A stochastic process $x_1, x_2, \ldots$ is exchangeable if for all $n$ and all permutations $\pi$:

$$p(x_1, \ldots, x_n) = p(x_{\pi(1)}, \ldots, x_{\pi(n)})$$

De Finetti’s theorem says that every exchangeable process is a mixture of i.i.d. processes:

$$p(x_1, \ldots, x_n) = \int p(\theta) \prod p(x_n|\theta) d\theta,$$

where $\theta$ is some parameter conditioned on which the data is i.i.d.

De Finetti’s theorem in terms of predictive distributions:

$$p(x_n|x_{n+1}) = \int p(x_n|\theta) p(f(x_n)|\theta) d\theta$$

This gives two ways for defining models of exchangeable sequences:
1) via explicit Bayesian modelling, e.g. like in the neural statistician [1]
2) via exchangeable processes, e.g. BRUNO

Exchangeability and meta-learning

Exchangeability is to meta-learning as convolutions are to vision.

Experiments

Fashion MNIST generation

OMNIGLOT few-shot generation

OMNIGLOT few-shot classification

Online set anomaly detection

Extra: conditional BRUNO

BRUNO can be easily extended to handle exchangeable sequences where every $x_i$ is associated with a vector of labels or tags $h_i$. Here, we model $p(x_1, h_1, x_2, h_2, \ldots)$.

ShapeNet 1-shot BRUNO samples conditioned on the camera angle

Bibliography