Activités en deep learning de l'équipe SC

Cédric Févotte

DeepLearning@IRIT
1er juillet 2020

Content

- Optimisation for the training of neural networks
  PhD Camille Castera
- Deep generative models for Monte Carlo sampling
  PhD Florentin Cœurdoux
- Deep learning for music recommendation
  Postdoc Paul Magron
- Deep learning for embedded satellite image compression
  PhD Vinicius Alves de Oliveira

Algorithms for Training Neural Networks 1/2

Motivation

- Deep Neural Networks usually trained with noisy first-order information (stochastic gradients)
- Hard technical limitations, in particular expensive computational cost and limited storage
- Lack of guarantees (convergence of algorithms, etc.)

Goals

- Build new algorithms for this task, efficiently exploiting second-order information.
- Prove the convergence and other satisfying properties of these algorithms.

Collaborators

IRIT: Camille Castera (SC) - Cédric Févotte (SC) - Edouard Pauwels (ADRIA)

TSE: Jérôme Bolte

Algorithms for Training Neural Networks 2/2

Problem to solve: $\min_{\theta \in \mathbb{R}^P} \mathcal{J}(\theta)$

Idea: Taking inspiration from Newton's second law of dynamics

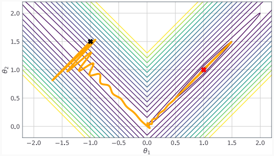
$$\underbrace{\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}(t)}_{\text{Accel.}} + \underbrace{\alpha\,\frac{\mathrm{d}\theta}{\mathrm{d}t}(t)}_{\text{Friction}} + \underbrace{\beta\,\nabla^2\mathcal{J}(\theta(t))\frac{\mathrm{d}\theta}{\mathrm{d}t}(t)}_{\text{Transvers. damping}} + \underbrace{\nabla\mathcal{J}(\theta(t))}_{\text{Gravity}} = 0$$

Equivalent formulation where the Hessian is implicit!

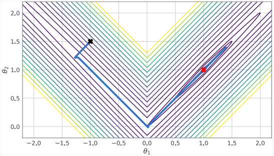
$$\begin{cases} \frac{\mathrm{d}\theta}{\mathrm{d}t}(t) + \beta\nabla\mathcal{J}(\theta(t)) & +(\alpha - \frac{1}{\beta})\theta(t) + \frac{1}{\beta}\psi(t) = 0 \\ \frac{\mathrm{d}\psi}{\mathrm{d}t}(t) & +(\alpha - \frac{1}{\beta})\theta(t) + \frac{1}{\beta}\psi(t) = 0 \end{cases}$$

Discretize to obtain **INNA**, a new 2nd order algorithm with two hyper-parameters α and β driven by Newton's laws.

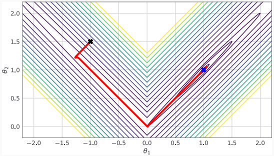
α small, β small



α small, larger β



larger α, larger β



C. Castera, J. Bolte, C. Févotte, E. Pauwels, *An Inertial Newton Algorithm for Deep Learning.* Journal of Machine Learning Research (To appear)

Deep generative models for Monte Carlo sampling

F. Coeurdoux, N. Dobigeon, P. Chainais

- Goal: Estimate a set of parameters θ from empirical data $\mathcal{X}$.
- ABC algorithm estimates the posterior of a parameter by simulating the model to produce artificial data sets $\mathcal{X}$ from distribution $p(\theta)$.

Bayes formula for posterior inference:

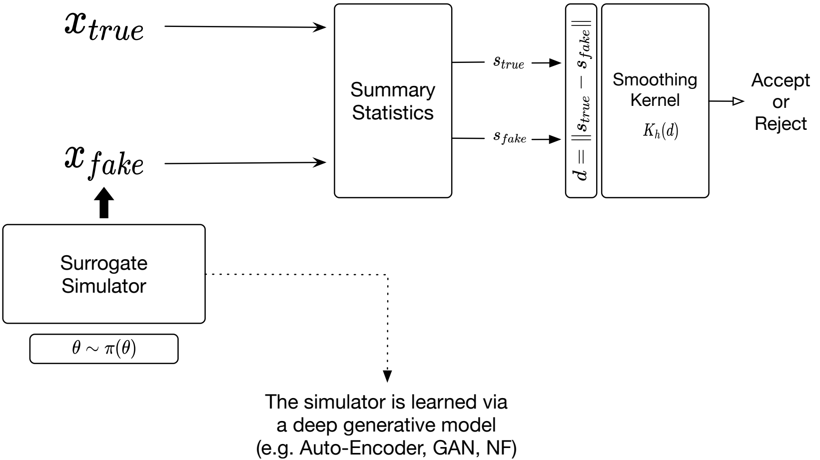
$$\text{Posterior} \propto \text{Prior} \times \text{Likelihood}$$
$$p\left(\theta \mid \mathcal{X}\right) \propto p\left(\theta\right) p\left(\mathcal{X} \mid \theta\right)$$

- Problem: A data simulator can be extremely computationally expensive or even not always available

Proposal: Leverage deep generative models (e.g Auto-Encoder, GAN, Normalizing Flows) for simulating data conditionally to the parameters of interest : $\mathcal{X} \sim p(\mathcal{X} \mid \theta)$.

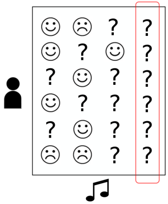
Deep generative models for Monte Carlo sampling

F. Coeurdoux, N. Dobigeon, P. Chainais



Deep learning for music recommendation (P. Magron, C. Févotte)

- Goal: predict songs that a user might enjoy.
- Main idea: exploiting *similarities* between users (and/or songs) from the observed preferences $\mathbf{P}$.
- The cold-start problem: how to handle new songs with no preference data?



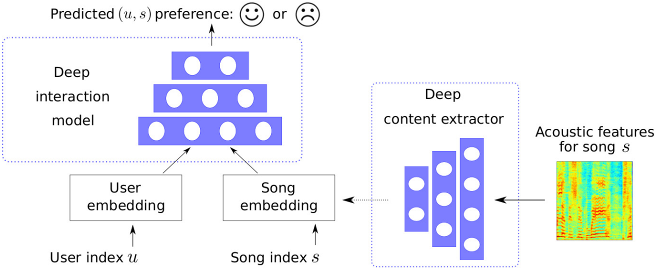
Content-aware collaborative filtering:

$$p_{u,s} = \psi(\mathbf{w}_u, \mathbf{h}_s) \quad \text{with} \quad \mathbf{h}_s \approx \phi(\mathbf{x}_s)$$

- Decompose the observed data with latent factors that characterize *user preferences* $\mathbf{w}_u$ and *songs attributes* $\mathbf{h}_s$.
- Regularize the song factor $\mathbf{h}_s$ with side content information (=acoustic features $\mathbf{x}_s$) to enable cold-start recommendation.

Proposal: Leverage deep learning for modeling both ψ (refined user/song interactions) and ϕ (extract relevant content features).

Deep learning for music recommendation (P. Magron, C. Févotte)

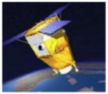


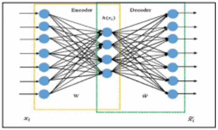
State-of-the-art performance (NDCG in %, higher is better):

Interaction	Content extractor	Warm-start	Cold-start
Linear	None	27.9	–
	Deep	30.4	23.8
Deep	None	38.8	–
	Deep	**40.9**	**25.9**

P. Magron, C. Févotte, "Neural content-aware collaborative filtering for cold-start music recommendation", *submitted to the ACM Transactions on Information Systems*, 2021 (`https://arxiv.org/abs/2102.12369`).

Deep learning for embedded satellite image compression (1/2)





- Motivations
  - Impressive performance, in terms of rate-distortion trade-off, of learned image compression methods based on convolutional neural networks (CNN).
  - Strong computational constraints on board satellites.
- Goals
  - Design a reduced-complexity CNN-based image compression framework that outperforms current embedded compression algorithms.
  - Design a CNN-based framework that could integrate other functionalities of the satellite image processing chain e.g. denoising.

Collaborators: Vinicius Alves de Oliveira[1,2], Marie Chabert[1], Thomas Oberlin[3], Charly Poulliat[1], Mickael Bruno[4], Christophe Latry[4], Mikael Carlavan[5], Simon Henrot[5], Frederic Falzon[5], Roberto Camarero[6]

[1]University of Toulouse IRIT/INP-ENSEEIHT, [2]TéSA Toulouse, [3]University of Toulouse ISAE-SUPAERO, [4]CNES Toulouse, [5]Thales Alenia Space Cannes, [6]ESA Noordwijk

Deep learning for embedded satellite image compression (2/2)

- Two contributions:
  - Reduced-complexity end-to-end variational autoencoder for on board satellite image compression.
    - Starting from a state-of-the-art framework, drastic dimension reduction targeting the number of layers, of filters, the size of the filters and the activation functions.
    - Proposition of a simplified entropy model adapted to the representation learned from satellite images.
  - Learned-based satellite image denoising and compression. Two proposals for different types of Earth observation missions:
    - Joint compression and denoising using a single on board framework.
    - Sequential learned compression (on board) and denoising (on ground).