Yoshua Bengio - Curriculum Learning (2009)

Tom Rochette <tom.rochette@coreteks.org>

December 21, 2025 — 77e1b28a

0.1 Context

0.2 Learned in this study

0.3 Things to explore

- Is it possible during training to determine that an example/sample is more "complicated" than others and to automatically partition samples in in some ordered complexity order?
- During training they added batches of 5000 words, would training have been more fluid (and resulted in even better results) if words were added one at a time?
- Is clustering (and other unsupervised approaches) a weak form of curriculum learning, where by grouping "similar" examples, it makes it easier to learn about them?

1 Overview

2 Notes

2.1 Abstract

- Curriculum learning can be seen as a particular form of continuation method

2.2 1. Introduction

- Called "shaping" when exploited in animal training
- The basic idea is to start small, learn easier aspects of the task or easier subtasks, and then gradually increase the difficulty level
- Elman makes the statement that this strategy could make it possible for humans to learn what might otherwise prove to be unlearnable

2.3 2. On the difficult optimization problem of training deep neural networks

- Deep learning methods attempt to learn feature hierarchies
- Features at higher levels are formed by the composition of lower level features
- A theoretical motivation for deep architectures comes from complexity theory: some functions can be represented compactly with an architecture of depth k, but require an exponential size architecture when the depth is restricted to be less than k
- Training deep architectures involves a potentially intractable non-convex optimization problem

2.4 3. A curriculum as a continuation method

- The basic idea is to first optimize a smoothed objective and gradually consider less smoothing, with the intuition that a smooth version of the problem reveals the global picture

2.5 4. Toy Experiments with a Convex Criterion

2.5.1 4.1. Cleaner Examples May Yield Better Generalization Faster

- One simple way in which easy examples could help is by being less "noisy"
- An example is considered noisy if it falls on the incorrect side of the decision surface of the Bayes classifier

2.5.2 4.2. Introducing Gradually More Difficult Examples Speeds-up Online Training

- A perceptron is trained on generated data, such that it has to determine the sign of the input $y = \text{sign}(w' x_{relevant})$ (w sampled from a Normal(0, 1))
- By controlling the amount of irrelevant inputs, the difficulty can be controlled (easy = no irrelevant input, difficult = many irrelevant inputs)
- The test error following curriculum learning is lower than through no curriculum (or anti-curriculum)

2.6 5. Experiments on shape recognition

- The task is to classify geometrical shapes into three classes: rectangle, ellipse and triangle
- The input is a 32x32 grey-scale image
- Two different datasets are generated
  - GeomShapes consists in images of rectangles, ellipses and triangles
  - BasicShapes only includes special cases: squares, circles and equilateral triangles
- Both sets contain other degrees of variability
  - Object position
  - Size
  - Orientation
  - Grey levels of the foreground and background
- The curriculum consists of a two-step schedule
  - Perform gradient descent on the BasicShapes training set, until "switch epoch" is reached
  - Then perform gradient descent on the GeomShapes training set

2.7 6. Experiments on language modeling

- A score is computed for the next word in a correct English sentence
- Collobert and Weston optimized the log-likelihood of the next word
- The ranking approach does not require computing the score over all the vocabulary words during training
- It is enough to sample a negative example

2.8 7. Discussion and Future Work

- The way we have defined curriculum strategies leaves a lot to be defined by the teacher
- It would be nice to understand general principles that make some curriculum strategies work better than others
- To reap the advantages of a curriculum strategy while minimizing the amount of human (teacher) effort involved, it is natural to consider a form of active selection of examples similar to what humans (and in particular children) do
- At any point during the "education" of a learner, some examples can be considered "too easy" (not helping much to improve the current model), while some examples can be considered "too difficult" (no small change in the model would allow to capture these examples)

3 See also

4 References

- Bengio, Yoshua, et al. “Curriculum learning.” Proceedings of the 26th annual international conference on machine learning. ACM, 2009.
- https://ronan.collobert.com/pub/matos/2009_curriculum_icml.pdf