Deep Learning

AlejandroMllo

This document serves as a very brief summary of the topics covered in each chapter of the book Deep Learning [1].

Disclaimer: This document is completely extracted from [1], the author does not attribute any ownership over the material.

12 Applications

- Deep learning is based on the idea that a large population of features acting together can exhibit intelligent behavior.
- Nowadays NN are trained mostly using GPUs or the CPUs of many machines networked together.
- Obtaining good performance with GPUs:
  - Most writable memory locations are not cached, so it can actually be faster to compute the same value twice, rather than compute it once and read it back from memory.
  - Multithreaded code.
  - Memory operations can be faster if they can be coalesced. Coalesced reads or writes occur when several threads can each read or write a value that they need simultaneously, as part of a single memory transaction.
  - Make sure that each thread in a group executes the same instructions simultaneously.
- Is better to distribute the workload of training and inference across many machines.
- Asynchronous SGD causes the learning process to be faster overall.
- Model compression: large models learn some function $f(\boldsymbol{x})$, but do so using many more parameters than are necessary for the task. As soon as is possible to fit this function, a training set containing infinitely many examples can be generated, simply by applying f to randomly sampled points $\boldsymbol{x}$. Then the new, smaller model is trained to match $f(\boldsymbol{x})$ on these points.
- Use dynamic structure to accelerate data-processing systems. NN use conditional computation.
- To accelerate inference in a classifier use a cascade of classifiers. This strategy may be applied when the goal is to detect the presence of a rare object (or event).
- Different kinds of preprocessing are applied to both the training and the test set with the goal of putting each example into a more canonical form to reduce the amount of variation that the model needs to account for.
- Dataset augmentation: improve the generalization of a classifier by increasing the size of the training set by adding extra copies of the training examples that have been modified with transformations that do not change the class.
- The task of speech recognition is to map an acoustic signal containing a spoken natural language utterance into the corresponding sequence of words intended by the speaker.
  - Unsupervised pretraining phase considered unnecessary (it did not bring significant improvements).
  - CNNs that replicate weights across time and frequency are commonly used.
- Natural Language Processing (NLP): use of human languages by a computer.
  - Techniques that are specialized for processing sequential data are commonly used.
  - Word-based language models operate on an extremely high-dimensional and sparse discrete space, they define a PD over sequences of tokens in a natural language.
  - Neural Language Models (NLM): class of language model designed to overcome the curse of dimensionality problem for modeling natural language sequences by using a distributed representation of words.

- Machine translation is the task of reading a sentence in one natural language and emitting a sentence with the equivalent meaning in another language. This systems often involve many components.
  - They used attention-based systems that are composed of:
    1. A process that reads raw data.
    2. A list of feature vectors storing the output of the reader.
    3. A process that exploits the content of the memory to sequentially perform a task, at each time step having the ability to put attention on the content of one memory element (or a few, with a different weight).
- Other application involve:
  - Recommender systems: collaborative filtering algorithms, reinforcement learning, etc.
  - Knowledge representation, Reasoning and Question answering.
    - Determine how distributed representations can be trained to capture the relations between two entities.
- Reinforcement learning requires choosing a trade-off between exploration and exploitation. Exploitation refers to taking actions that come from the current, best version of the learned policy. Exploration refers to taking actions specifically to obtain more training data.

References

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.