Building blocks for large-scale learning based systems

Francis Maes (Montefiore Institute, ULg)

June 11, 2010

During the last decade, statistical machine learning has reached a high level of maturity to solve simple learning problems such as classification and regression, with efficient and theoretically well-founded algorithms. Although this is enough for many tasks, real-world applications such as natural language translation or protein folding involve learning problems that are much more complex than classification or regression.

In many fields of science, the classical approach to deal with complex problems is to decompose them into simpler sub-problems recursively until reaching atomic problems that we can solve easily. Surprisingly, this "divide-to-conquer" approach has nearly never been applied in the machine learning field yet. Two central questions must be solved to progress into this direction: what are the building blocks ? and what are the rules to assemble such building blocks to construct more and more complex systems in a principled way ?

This presentation focuses on these two questions. We propose the "Inference" framework where each building block is a learnable module that has an input, an output and an optional supervision. We propose building rules that make it possible to construct a complex inference by combining simpler inferences. We demonstrate the ability of our framework to capture various classical learning models, as well as its ability to construct highly complex system – and to learn them – in a principled way. In order to illustrate our approach, we rely on the protein folding problem and describe our work in progress on this problem.