Machine Introspection for Machine Learning

Tristan Cazenave

LIP6, Tour 46-00 2ème étage
Université Pierre et Marie Curie, 4, place Jussieu
75252 PARIS CEDEX 05 FRANCE
e-mail: Tristan.Cazenave@lip6.fr

Primary subject classification : 03.12 Subsidiary subject classification : 03.04 03.09 03.16 01.05

Using consciousness, human beings can have some knowledge about their own behaviors ; this knowledge is very useful and consciousness is more developed in human beings than in any other animals. According to [Pitrat 1990], this is the reason why we overestimate our potential for being conscious. The mechanisms of consciousness could be much more powerful. [McCarthy 1996] gives some mechanisms that conscious machines should have. Introspect [Cazenave 1996] is a system that experiments some of these mechanisms for real. The system can observe its own behavior so as to detect its own inefficiencies and repair them. When it detects inefficiencies, it reasons on its reasoning process so as to understand why it has been inefficient. Then it modifies itself so as not to be inefficient again in similar situations. Introspect has mainly been used to discover knowledge in the game of Go. It has only been given the rules of the game, by playing the game it has observed, understood and modified itself and has discovered a lot of useful expert Go knowledge.

Go was developed three to four millennia ago in China; it is the oldest and one of the most popular board game in the world. Like chess, it is a deterministic, perfect information, zero-sum game of strategy between two players. In spite of the simplicity of its rules, playing the game of Go is a very complex task. On the contrary of Chess, it is impossible to make a brute force search of all the moves in the game of Go, the best Go playing systems rely on a knowledge intensive approach. A Go expert uses a large number of rules. Go programmers usually try to enter these rules by hand in a Go program. Creating this large number of rules requires a high level of expertise, a lot of time and a long process of trial and error. Moreover, even the people who are expert in Go and in programming find it difficult to design these rules. This phenomenon can be explained by the high level of specialization of these rules: once the expert has acquired them, they become subconscious and it is hard and painful for the expert to explain why he has chosen to consider a move rather than another one. He is not conscious of all the reasons why he has chosen the right move.

The difficulty of encoding Go knowledge is the consequence of a well known difficulty of expert system development: the knowledge engineering bottleneck. Writing a program that is able to observe itself, to detect its own inefficiencies, to create new knowledge so as to uncover them, and to use the new knowledge efficiently is a nice way to avoid this bottleneck by replacing the knowledge extraction process with an automated learning system based on introspection. Machine introspection and learning enable to get rid of the painful expert knowledge acquisition. Thus, computer Go is an ideal domain to test the efficiency of machine introspection when faced with expert human programmers associated to professional Go players (i.e. human introspection).

Introspect memorizes its mental actions so as to be able to observe itself after. It can transform itself into an unconscious program by compiling itself. The benefit of being unconscious is that it is faster because it does not have to interpret and memorize its behavior. The drawback is that it cannot introspect itself anymore. The unconscious program written by Introspect is named Gogol and consists of 1 000 000 lines of C++.

Gogol competed in the international computer Go tournament held during IJCAI97 together with 40 other participants. It finished as the best program based on academic research, playing better that the other programs directly written by Artificial Intelligence researchers and Go professionals. It has outperformed systems that have required more than 10 man*years of work. Introspect has also been used in other domains (games, pedestrian simulation and management) and has discovered, by introspection, knowledge that is more efficient than the knowledge given by experts.

References

[Cazenave 1996] - T. Cazenave. *Système d'Apprentissage par Auto-Observation. Application au Jeu de Go.* Ph.D. Thesis, Université Pierre et Marie Curie, Paris 6, 1996.

[Pitrat 1990] - J. Pitrat. *Métaconnaissance*, Editions Hermes, Paris 1990.

[McCarthy 1996] - J. McCarthy. *Making Robots Conscious of their Mental States*, in Muggleton S., editor, Machine Intelligence 15, Oxford University Press, 1996.