Reflex

Open Worlds, Reflective Statistics and Stochastic Modelling

Heinz Mühlenbein

Report number: 1994/14

Research Group for Adaptive Systems
SET.AS

Open Worlds, Reflective Statistics and Stochastic Modelling

Heinz Mühlenbein

German National Research Centre for Computer Science (GMD),
Schloß Birlinghoven, 53754 St. Augustin,
Germany.

1 The Problem

The real world is open and ambiguous. The problem of its openess has been neglected in science for a long time, especially in artificial intelligence. Most researchers in artificial intelligence still deal with closed worlds. A recent example is the CYC project from Lenat which started in 1984. Lenat believed that after entering about 10 million facts into CYC, that "CYC will grow by assimilating textbooks, literature, newpapers etc.". Now, in 1994 it turned out, that CYC has hardly enough knowledge for a small artificial domain in VLSI design.

Openess is a deep problem, it has to be taken seriously. The real world is not completely knowable. In fact, the domain of knowledge is very small compared to the huge unknown domain. Any system operating in the real world has to act with incomplete knowledge. This general observation has far reaching implications for probability theory and statistical inference. The theoretical discussion in these areas culminated in Popper's famous sentence:*"All knowledge is assumption knowledge"*. Every knowledge formulated as a hypothesis is preliminary, subject to rejection if new data is in contradiction to the hypothesis. In more technical terms this means for statistical inference: *Probabilistic hypotheses do not have a hypothesis probability.* It is possible to rate a number of hypotheses according to how good they explain the data, but it is not possible to compute a number which can be interpreted as the probability of a given hypothesis.

2 A Possible Solution

In the real world artificial systems face the same problem as living beings - they have to act in an open world with incomplete knowledge. Therefore it seems worthwhile to investigate how living beings cope with this situation. Research results are already available in biology, psychology and philosophy. I will only discuss some philosophical aspects. The importance of a philosophy for robot design has already been pointed out by the biologist Waddington. He wrote in 1974: *"The only way to make a robot anything more than an adding machine is to provide him with a philosophy."* For Waddington the essential function of a philosophy is to provide a mental machinery for dealing with a large variety of things.

I believe that reflection is a necessary element of such a philosophy. In an open world a system has to know what it knows and - still more important - what it does not know. The system can incrementally acquire the knowledge by selfassessment. The importance of knowing the boundary between knowledge and ignorance has been advocated by many philosophers both in east and west. I only want to mention the guideline from Lao Tsu: *Knowing ignorance is strength, ignoring knowledge is sickness.*

The scientific challenge for real world computing is to put this general philosophy into something similar to a calculus.

3 JANUS - a robot for open worlds

JANUS is a robot with two arms and two cameras designed to operate in open worlds. This term is applied to worlds that are not (yet) completely and unambigously definable from the information gathered about them so far. JANUS consists of a large numer of heterogeneous algorithms and sub-systems called *agents* that operate in parallel and independently. We have not yet been able to define a claculus in the statistical sense do deal in general with open worlds. Therefore a number of rudimentary principles, defining a general philosophy, has been developed for JANUS.

1. Adaptivity
   Agents should have the ability to adapt to the world and complement their basis heuristics with experience-based approximations of relevant functions.
2. Reflection
   The agents at every level of the system should be able to assess their own performance, and be able to say where they are expert and where they perform well. Likewise they should know what they do not know.
3. Exploration
   The system should not passively wait for supervision, but actively explore the problem domain.
4. Learn from nature
   When a non-trivial problem allows itself to be expressed in such a form that is similar to those seen in natural systems, borrow ideas from the way it has been solved here.
5. Reuse knowledge
   Try to reuse knowledge learnt in one task to solve similar looking problems in another task. Construct transformable hypotheses.
6. Open internal structure
   Allow the structure of the system to be of an open heterogeneous nature, to enable differing problem solving methods to work concurrently and complementarily, and allow the progressive addition of new methods.
7. Team work
   Solve complex problems by combining many simple solutions in an iterative manner, instead of attempting to construct a single global solution mechanism.
8. . Be optimistic
   Do not expect the worst case. Optimize for the average case.

We have started to formalize some of the above principles, namely reflection and exploration. The latter one is based on stochastic modelling. We will not describe this here, but one of the first problems solved by stochastic modelling.

4 Genetics - a case study in stochastic modelling

Stochastic modelling seems to be a promising calculus for dealing with ambiguity. A very important case study in stochastic modelling can be found in quantitative genetics. The case study deals with explaining the macroscopic observed evolution of livestock and plants by the microscopic chance model invented by Mendel.

The group of biometricians centered around Pearson succeeded in quantifying Darwin's evolution theory in purely macroscopic terms. They invented a variety of now standard statistical techniques, including those of correlation and regression. The theories required knowledge of the correlation between relatives for various characters (e.g. height). Empirical estimates of these correlations were obtained and used in the analysis. No genetics is involved.

After rediscovering Mendel's genetic chance model, researchers tried to derive the empirical laws from Mendel's chance model. Fisher solved the problem for a trait governed by one gene as early as 1922, but it took many years to solve the general case for an arbitrary number of genes. The general solution was proven by Kempthorne. He was able to predict the covariance between parent and offspring from Mendel's model. The prediction uses one of the most difficult statistical techniques in use today, namely the decomposition of the covariance. We have rediscovered this result and applied the method to the Breeder Genetic Algorithm. Unfortunately it turned out, that for complex fitness landscapes the predictive power of the covariance decomposition is very limited, less than of the purely macroscopic regression analysis. The reason for this seems to be that the assumptions needed for the decomposition are very severe and seldom fulfilled.

This result seems to indicate that Mendel's chance model leads to a macroscopic world which is very difficult to predict from the microscopic processes. We are currently investigating the implications of this result for theoretical biology.