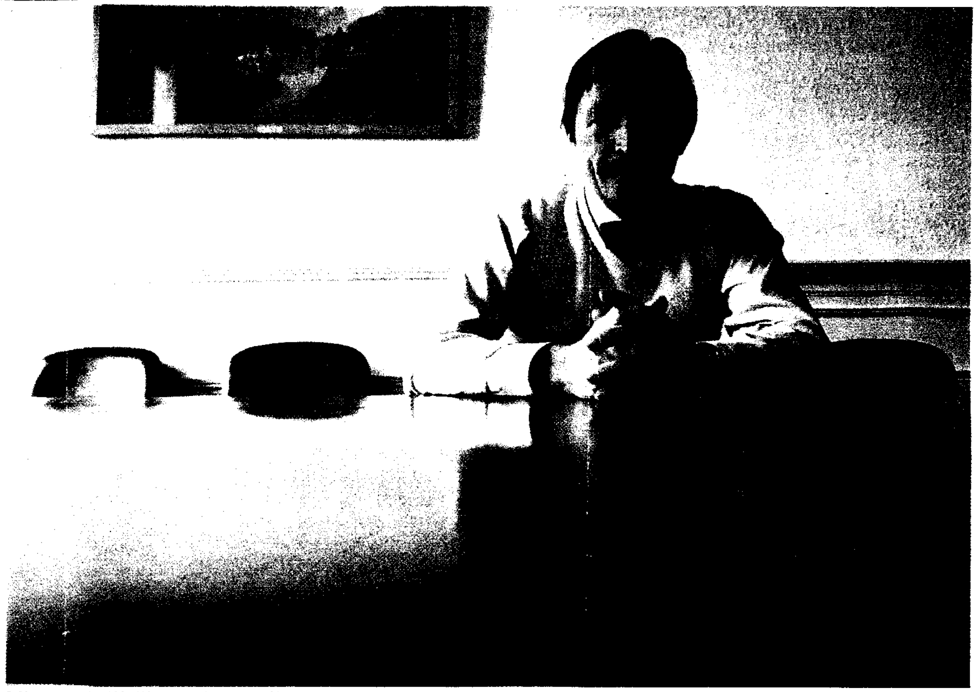
TRYING TO MIMIC THE MIND

by

R. S. Michalski
T. Durham

In Computing Magazine, April 18, 1985.

For machines, as for human beings, knowledge without understanding can be a dangerous thing. Tony Durham reports on moves at Illinois University to enable computers to check that their conclusions match up with reality



TOM KIDD

Michalski: 'We'll be able to build complex structures, but still make them consistent with all known facts'

TRYING TO MIMIC THE MIND

Professor Ryszard Michalski's intelligent systems group, part of the intelligence laboratory at the University of Illinois, is one of the world's leading centres of research on machine learning.

Michalski says that his group's research has three branches, which develop in parallel. There is the theoretical work, which seeks an improved understanding of the process of learning, and there is the work on programs and algorithms which actually demonstrate learning. 'The third direction,' says Michalski, 'is applying those programs to very specific tasks, to see whether those programs indeed do something useful.'

The applied work has been mainly in agriculture and machine, with subsidiary projects on chess and other subjects. One result which has achieved almost mythical status in the expert systems world was a set of rules for diagnosing soybean diseases. The rules were formulated by Michalski's AQ11 learning program, using 290 training examples. Each example was a description of various features of a diseased plant, together with an expert diagnosis of the plant's disease.

The machine-generated rules gave the correct disease top ranking 97.6% of the time. Another set of rules, developed by human 'knowledge engineers' interviewing an expert plant pathologist, put the correct disease first only 71.8% of the time.

'Learning has a special place in artificial intelligence(ai), in my opinion,' says Michalski. 'We probably all agree that learning is central to intelligent behaviour. It is hard to imagine an intelligent person or a truly intelligent system that could not learn, that could for example repeat the same error again and again, never able to improve no matter how many times the procedure is repeated.

'We are now discovering after many years of trying ai without learning, that further progress in almost all areas of ai very much demands progress in machine learning.'

As in the soybean example, machine learning may help to solve the knowledge bottleneck problem in building expert systems. Michalski extends this to other areas of ai: 'in order to improve our vision systems, we have to be able to teach them visual concepts by example, rather than define everything by hand in the form of a program,' he says. 'The same in natural language understanding, the same in intelligent tutoring systens, or in robotics. All the major areas of artificial intelligence are now at the stage where further progress would greatly benefit by having learning capabilities implanted in the systems.'

Humans, and machine learning programs, often form general hypotheses from specific examples. Michalski points out that this process of inductive inference introduces an element of uncertainty. 'It leads to hypotheses or theories which may not be true, though the initial starting set of facts was true,' he says. 'Learning is a very dangerous type of activity because it may introduce things which are uncertain. So therefore the results of learning processes have to be tested. Now in order to test those results we have to understand those results.'

Michalski calls this the principle of comprehensibility. He argues that machine learning methods should produce knowledge in a form which corresponds closely to the way the same knowledge would be modelled in the human mind.

'The reason for strongly insisting on the principle of comprehensibility is that since you generate knowledge which may not be true, you are introducing a potential danger,' says Michalski. If the knowledge is used, without understanding, in some new situation, 'it may lead to really very dangerous situations'. Michalski is serious about this. If we use a knowledge-based system to recognise intercontinental missiles, he points out, we'd better understand how it works.

The solution he proposes is a knowledge ratification process. 'Knowledge is generated by machines,' he explains, 'but a panel of experts looks at the knowledge, and ratifies it, and says, "this is the kind of knowledge which we have tested on some real data, and we are sufficiently confident that the knowledge can now be accepted by human society."'

Michalski's learning systems are intended to be useful in their own right, and not merely as models of human learning.

But he argues that learning research must stay close to the human model. If the aim is to protect the principle of comprehensibility, we must know what kinds of things people can comprehend. Michalski believes machine learning researchers should work hand in hand with cognitive psychologists, as he has recently done himself.

Of course, a machine might learn a rule which is comprehensible to humans, without necessarily learning the same rule as a human would. Nor need the machine learn its rule in the same way as a human does. 'They are separate issues,' Michalski agrees. 'But you need a good model of the human mental processes which are involved in understanding and evaluating the result. So the studies which we do concentrate on the evaluation by humans of those results.'

But there is another motive, too, for studying human learning. 'We also try to understand how people arrive at those results,' Michalski adds, 'because those studies may give us some clues to better machine algorithms.'

He believes that it may be possible to formulate a computational theory of what is capable of being learned. But he points out that in practice, knowledge may be required in different forms for different purposes. His group has tried to build algorithms which produce knowledge in relatively simple notations, including one called annotated predicate calculus, which can be translated directly into natural language. The idea is that it may then be simple to translate the knowledge into other forms for specific purposes. ▶

There is usually a trade-off. Learning algorithms are easier to write and quicker to run, if they use a relatively restricted language to express the rules that they learn. For example, the decision trees that are generated by Expert-Ease and similar programs are efficient, but can only learn a limited class of concepts. At the opposite extreme, no one knows how to write a program capable of learning every concept that can be expressed in English.

'So therefore you have to narrow down your language,' says Michalski. 'But then if you narrow down too much, you would be learning very simple, trivial concepts which are not that interesting.

'For example, by restricting your language only to decision trees, as is done in Expert-Ease, you get solutions that sometimes are incomprehensible, because you have a very limited language. But you do it very efficiently. We have been working with a more general language. Therefore the knowledge which we generate by our algorithm is in many cases much more easy to comprehend, and simpler in many ways. But it takes more computation.'

But Michalski argues that the trade-off is not always linear. 'By cleverly choosing the operators of your language,' he says, 'you may suddenly gain more mileage by doing that, than you lose by increasing the complexity of your algorithm.'

Most ai researchers choose to represent knowledge in a form which allows some deductive inference mechanism to operate efficiently. Michalski's work, however, focuses on knowledge representations which can be learned – representations which allow efficient inductive inference. 'I would not say there is a conflict,' explains Michalski. 'There are simply two different criteria.

'Most of the work in expert systems, knowledge representation and inference systems went into building representations which are good for deduction. We are saying that's not enough, because you not only want to use knowledge, you also want to update it, and even replace it by better knowledge.' Michalski therefore argues that you should not only consider the power of the language for representing things, and how easy it is to use the knowledge for drawing conclusions. 'You should take one more criterion into consideration, namely how easy is to modify knowledge, or construct new knowledge. That is exactly what we are trying to go, to somehow balance and take into consideration all three criteria.'

In March, Michalski presented a paper at the Turing Institute's Machine Intelligence 11 workshop, on what he calls 'incremental learning' – the step by step acquisiton of knowledge. 'Incremental learning,' he explains, 'is a form of learning that we humans perform all the time. When we talk about very large knowledge structures, large knowledge bases which we will be building in the future for expert systems, then we will be facing incremental learning, because we cannot assume that each time we get more facts we have to completely redesign our knowledge base.'

The Illinois research has thrown attention on an important and potentially useful difference between human and machine learning. You know that nettles sting, but do you recall all the occasions that helped you to learn this? 'People have a hard time remembering facts,' says Michalski, 'but they remember, relatively well, simple and powerful theories. Machines have a hard time deriving both powerful and simple theories, but can remember facts very well. So why not take advantage of the machine's ability to store and retrieve many facts?'

This led to the idea of 'full-memory incremental learning'. This is something machines can do and humans cannot. Michalski points out that humans 'store some facts but not all of them. They just store representative facts'. The danger here is that as humans extend their theories, these may become inconsistent with facts which were once known and now are forgotten. (Though scientists, Michalski says, can be fairly sure of maintaining consistency because they have a vast body of facts recorded in books.)

Machines with their vast memories need not fall into the trap of inconsistency. 'We'll be able to build very complex structures, modify them, but still make them consistent with all known facts,' Michalski says.

Intriguingly, philosophers have had difficulty with the idea of inductive inference. 'Not only Popper, but also Russell, Polya, and many other philosophers over many centuries, starting with Aristotle, had certain problems with induction.'

One mistake the philosophers made, according to Michalski, was to imagine that induction was meant to produce true theories. This, he says, is impossible.

His answer is to separate theory generation from validation, as in the knowledge ratification process which checks that knowledge is not dangerous to human society.

'We cannot possibly do complete verification by machines,' Michalski says, 'because we cannot put on the machine a complete model of the human mind, the human body, human goals, human expectations about the future and human relations. It is simply impossible to build a perfect model of a human being on a machine.'

Tony Durham is a freelance journalist. ■