Discussion

The time scale of artificial intelligence: Reflections on social effects

R.J. SOLOMONOFF
Oxbridge Research, P.O. Box 559, Cambridge, MA 02238, USA

R. Solomonoff was graduated from the University of Chicago in 1951 with a degree in Physics. Since that time he has mainly been working on the mechanization of inductive inference – the most successful approach being algorithmic complexity theory. He has extended this theory to include the optimization of both hardware and software for general problem solving. He is now a principal scientist at Oxbridge Research, Cambridge, MA.

Six future milestones in AI are discussed. These range from the development of a very general theory of problem solving to the creation of machines with capacities well beyond those of a single human. Estimates are made for when these milestones will occur, followed by some suggestions for the more effective utilization of the extremely rapid technological growth that is expected.

Keywords: Artificial intelligence, social effects, future developments.

I will first give a brief discussion of recent developments in AI, and then a description of the expected future milestones with estimates of when they will occur and some expected social effects.

Milestone A. The 'modern' phase of AI can be regarded as beginning in 1956 at the Dartmouth Summer Study Group on Artificial Intelligence. At that time many people in this field came from all over to talk about what they were doing and what they expected to do. It marked the beginning of the much accelerated work in this area.

One of the earliest developments was the 'General Problem Solver' of Newell and Simon – a first attempt at a general theory of AI. From this beginning, they moved on to study human problem solving – which developed into what is now called 'cognitive psychology'. Parallel with this work, was the development of 'expert systems', which depend not so much on general principles, but on knowledge of many facts in a particular field. Public awareness of these expert systems has grown rapidly in recent years, as has the fraction of AI manpower devoted to them.

Another very relevant development has been the study of large parallel computers with novel architectures. These studies are important because, first, the human brain is a very large parallel computer and the design and study of computers of this type can give the needed insight on how the human brain works. Another point is that the information processing capacity of present-day serial machines seems to be less than that of the human brain. If we are to emulate its behavior, we must have at *least* its computing capacity, and large parallel machines are most certainly the least expensive way to do this.

Milestone B. The next milestone in the development of AI might be a general theory of problem solving. Here 'problem solving' is to be understood in a very general sense, and includes processes which, if they were performed by a human, would be regarded as 'creative' or 'insightful'.

Some areas that would have to be covered by such a theory are:

(1) Learning: based both on input data, and the machine's own experience in problem solving;
(2) Devising and testing new concepts to be used in solving problems;

North-Holland
Human Systems Management 5 (1985) 149–153

(3) Taking in information and storing it in a manner useful for problem solving; and
(4) Methods of implementation on existing computers and/or the design of new kinds of computers that would be needed.

Milestone C. A critical point in AI development would be a machine that could usefully work on the problem of self-improvement. Newell and Simon were not successful in their attempts to get their 'General Problem Solver' to improve it's own methods of operation. While Lenat's 'Eurisko' has been successful in several problem areas, he has not been able to get it to devise good heuristics for itself. He is, however, optimistic about the progress that has been made and is continuing this work.

Milestone D. Another milestone will be a computer that can read almost any English text and incorporate most of the material into its data base – just as a human does. It would have to store the information in a form that is useful for solving whatever kinds of problems it is normally given.

Since there is an enormous amount of information available in electronic data bases all over the world, a machine with useful access to this information could grow very rapidly in its ability to solve problems and in a real sense in its understanding of the world.

Milestone E will be a machine that has a general problem solving capacity near that of a human, in the areas for which it has been designed – presumably in mathematics, science and industrial applications.

Milestone F will be a machine with a capacity near that of the computer science community.

Milestone G will be a machine with a capacity many times that of the computer science community.

Can we estimate when these milestones will occur?

For Milestone B – a general theory for AI – I feel that anything between one and 50 years is possible, with 2 to 25 years being much more likely. At present, there are too few people in AI working on theories of this sort. To aggravate the problem, recent commercial success of AI – mainly expert systems – has lured many bright graduate students away from general theory, to work on industrial applications.

Some promising work on general theory at the present time are: Lenat's work, which I've mentioned [1]; and Bradshaw, Langley and Simon on how scientists discover scientific laws [2].

Neither of these are explicit attempts at a general theory of intelligence, but they work on problems in ways that are readily generalized.

Some more direct work on general theory are: Minsky's work 'The Society of Mind' is an attempt to describe the operation of the human brain in terms of a large number of small problem solvers working parallel with relatively infrequent intercommunication [3]; and my own work on training sequences, problem solving and learning [4].

It may be possible to get something that is superficially like Milestone E without a general theory. The current Japanese '5th generation computer' project attempts to program a large number of 'expert systems' and put them all in a very large, very fast computer. Though expert systems all try to simulate parts of the human conscious mind, many of the more interesting human activities are mainly performed by the unconscious mind. If the unconscious mind works very much like the conscious mind (but we are merely less aware of its workings), then there is no difficulty here. However, if as is widely suspected, the unconscious mind is significantly different from the conscious – then the present expansion of expert systems will have serious limitations.

It is not necessary to know just how the unconscious mind works in order to emulate it – but slavishly imitating the workings of human consciousness would seem to be a poor approach.

Milestone D – understanding English – is being approached from several directions.

One is the study of ethnic languages, their grammars and semantics.

A somewhat different approach has been developed within the AI-community, in which machines are programmed to respond to commands or questions in English. The emphasis is on whether the program responds in the desired way, not on whether it 'understands' the input in terms of traditional grammatical and semantic concepts.

A third approach is through learning. The machine is taught English starting with very simple sentences. After it has learned to respond to them properly, it is given somewhat more complex sentences – just as a child learns language.

Most likely these three methods should be combined to obtain a system that acquires most rapidly, an understanding of English. The learning component is, I think, essential. The meaning of words and phrases vary considerably with context – sometimes grossly, other times subtly. Programming all of these nuances into a machine would seem to be too arduous a task to be done well by a human. It would be far better and less subject to error, if the machine learned as humans do, how the larger context of a phrase controls its meaning.

When can we achieve Milestone E? Milestone B seems to be the most critical bottleneck. From that point to achieving both C and D might be as little as five or ten years, and from there to milestone E, only a few years more.

Let us examine the significance of Milestone E. At such a time we would have a machine with the problem-solving capability of a human, in several fields. For reasons that will become clear later, we will at first want to emphasize mathematics and science – computer science in particular.

Twenty years from now, the hardware cost of such a machine might be as little as several hundred thousand dollars, and it will be halved every four years or so.

At this rate, artificial intelligence will eventually cost less then human intelligence. Note that while the cost of training a very intelligent machine is very large, the cost of training the next one is very small, since the information in memory can be rapidly transferred from one machine to another.

The most important features of very intelligent machines are not related to their cost however. Machines of this sort are able to do things far beyond the capabilities of humans or groups of humans. For example, they can be designed to process information from many modalities very rapidly – optical, radar, sound, radio, telephone, etc. As our machines become faster, such processing would become invaluable in weather prediction or the administration and control of very large projects, such as space programs, the construction of ever-larger computers, and providing food and shelter for billions of people.

Some of the most critical capabilities of very intelligent machines depend on their being *much* more intelligent than humans. How long will it take to go to Milestone F and then to G?

The number of creative scientists and engineers that are responsible for the advancement of computer science, are at most several thousand. After we have reached Milestone E, it shouldn't take much more than ten years to construct ten thousand duplicates of our original 'Milestone E' machine, and have a total computing capability close to that of the computer science community. The ten year figure seems reasonable when one notes that the cost of these machines will keep halving every four years or so, and also that the new 'artificial' computer scientists will help speed the construction of the new machines.

While there is normally an exponential decrease in computing cost with time (halving every four years or so), when the artificial intelligence community is as large as the human scientific community, the halving time itself will halve, so we get halving in two years instead of four.

Suppose c is the size of our computer science community at time t. We define this to be 1 at time zero. R is the rate at which we expend money on our AI computers to effectively increase the size of our computer science community. t is the time in years, from our origin point. x is the amount of computing power we get for a dollar at a particular time. We will set $x = 1$ at $t = 0$. First,

$$\mathrm{d}c/\mathrm{d}t = Rx. \tag{1}$$

The rate of increase of our (partly artificial) computer science community is the product of our rate of expenditure times the efficiency of that expenditure. Next

$$\mathrm{d}\ln x/\mathrm{d}t = Ac. \tag{2}$$

This says that the rate of change of the log of our efficiency is proportional to the size of our computer science community. If c were to be kept constant at 1, then we would want eq. (2) to give a doubling of x every four years. This gives $A = (\ln 2)/4 = 0.1733$. With conditions $c = 1$ and $x = 1$ at $t = 0$, we obtain from (1) and (2)

$$\mathrm{d}c/\mathrm{d}t = A(c^2 - 1)/2 + R \tag{3}$$

This equation has the property that for any positive value of R, the value of c will at some finite time $t = T$, approach infinity.

For $R = 1$, $T = 4.62$ years,
if $R = 0.1$, $T = 11.11$ years,
if $R = 0.01$, $T = 21.51$ years.

A value of $R = 1$ means that if we kept x constant at 1, at the end of one year we would have invested enough in our AI computer to equal its capacity to that of the human computer science community.

Usually, when infinities like this one occur in science, they indicate a breakdown of the validity of the equations as we approach the infinity point'. The critical part of the equations appears to be continued exponential decrease in computation cost. So far, this rate of improvement has been possible only because of radically new technologies that were introduced – i.e. first vacuum tubes, then transistors, then integrated circuits, and then large scale integrated circuits. There appear to be several new technologies on the horizon that are adequate for maintaining the progress for several more orders of magnitude – as for the technologies over the horizon that have not yet been discovered, we only have a faith based on performance of the past.

A decrease in computation cost by a factor of 1000 would, at the present rate of progress, take about 40 years. At the present time, a reasonable guess at the cost of hardware with the computing power of a human brain might be ten million dollars. Ten thousand of such machines would cost 100 billion dollars now, and 100 million dollars forty years from now. This 100 million would put us at $t = 0$ for eq. (3). At a continued expenditure of ten million dollars a year, it would take about 11 more years to get to the 'infinity point'. Though infinity is a bit high, it seems very likely that we could achieve a growth factor of at least 100 in these 11 years – and so we reach Milestone G. What would be the effect of a scientific community equivalent that is 100 times as large as what we have now?

The last 100 years have seen the introduction of special and general relatively, automobiles, airplanes, quantum mechanics, large rockets and space travel, fission power, fusion bombs, lasers, and large digital computers. Any one of these might take a person years to appreciate and understand. Suppose that they had all been presented to mankind in a single year! This is the magnitude of 'future shock' that we can expect from our AI-expanded scientific community.

In the past, introduction of a new technology into the culture has usually been rather slow, so we had time to develop some understanding of its effect on us, to adjust the technology and culture for an optimum 'coming together'. Even with a slow introduction, our use of a new technology has sometimes been very poor.

The use of nuclear energy for military purposes has been expensive, difficult to control and has obtained us neither military goals nor security of any sort. Nuclear energy for power generation in the United States, has cost much more than expected. In both cases we have had many years to consider how to use this technology best – yet, perhaps because of the difficulties of the problems involved, we have not done very well. We have spent enormous amounts of money and manpower and have attained relatively little of value.

Can we use very intelligent machines to help us solve the problems associated with the surfeit of new technologies of the future?

There appear to be at least two ways to do this.

First, attainment of Milestone B is likely to give us a much better understanding of the human mind than we have ever had. We should be able to get our intelligent machines to explain each new technology in a way that is intelligible to man. If this can't be done, and the new technology is essentially un-understandable to man, then man would be foolish indeed to use it in any way!

However, understanding does not always assure success in dealing with very complex problems. Mankind will continue to have to make decisions under conditions of uncertainty. In the past he has usually chosen his courses of action relatively blindly – controlled more by his own perceived wants and needs than by considerations of the likelihoods of alternative possible futures and their effects upon him.

In this area, very intelligent machines can help us in one very important way – they can predict the results of social action.

Normally, there are several limitations, both theoretical and practical on our ability to predict the future accurately.

These limitations are:

(1) The models we use for prediction are not the best possible, and we are unable to find better ones;

(2) We have a limited computing capacity and have already used all of it;

(3) Predictions can be self-modifying: we can make the prediction, but as soon as we make it

public, this brings about conditions that invalidate it;

(4) Quantum mechanical limitation on prediction. This is to some extent similar to (3); and

(5) Selectively feeding data to our predictor so that it obtains the result we think we want, or by otherwise biasing the prediction process.

Difficulties 1 and 2 are both inherent in the nature of all real-world predictions. No matter how long we search for good models of our system, there is always the possibility that if we looked a little longer we would find a much better model. At Milestone G we will have much better models than we have now, as well as a much greater computing capacity for applying them.

Difficulty 3 is very important in predicting social action. If the prediction is self-denying, then there may exist no public prediction that is correct. If it is self-conforming, there may be several different predictions that can be made – any of which would be correct if made public. For example, one prediction might be that many people would be hurt and another might be that no-one was hurt. Under these conditions we would want to give the predictor 'ethical guidelines' upon which to make a choice, or have a human intermediary decide what prediction to make public.

Difficulty 5 is less of a problem if the machine has independent access to all available information. However, it is often possible for a human to inadvertently define a question so that the reply must be badly biased.

Within these necessary limitations we will be able to obtain much better predictions than ever before. And, as before, it is not certain that even this capability will be used wisely.

What seems most certain is that the future of man – both scientific and social – will be far more exciting than the wildest eras of the past.

References

[1] Lenat, D. and J. Brown, Why AM and Eurisko appear to work, in: Proceedings Nat. Conf. on AI, 22–26 August (1983) 236–240.

[2] Bradshaw, G., P. Langley and H. Simon, Studying scientific discovery by computer simulation, Science 222, 4627 (December 1983) 971–974.

[3] Minsky, M., The society of mind, forthcoming.

[4] Solomonoff, R., Perfect training sequences (Oxbridge Research, Cambridge, MA, 1982).