# How Much Can Machines Learn?

Many scientists are working on the problem of making digital machines act in the light of their own accumulated 'experience' instead of following detailed programs of instructions. Reporting on the progress they have made, Dr. Booth concludes that the surface of the problem has only just been scratched, though the modern high-speed computing machine opens up many possibilities.

Andrew Booth

IN the past two or three years much nonsense has been written concerning the supposed resemblances between automatic digital calculators and human beings, and this is particularly the case where authors seek to discuss the powers of machines to learn from experience. This article attempts to put the matter into perspective and to examine what has been achieved on digital computing machines and the way in which these achievements can find application in practice.

*PAVLOV'S DOGS*

Before discussing learning programmes for computers, it is worth while seeing the sort of learning which could easily be simulated on a machine. The psychologists tell us that the conditioning of a reflex, in the way first demonstrated by Pavlov on dogs, is in some sense a measure of the fact that a dog is an intelligent living organism. The way in which such a reflex is conditioned is probably too well known to need extensive description. Suffice it to say that to make a dog learn to respond in a given manner to an appropriate stimulus, it is either rewarded or punished a sufficient number of times after presentation of the initial stimulus and evaluation of its resulting action.

The classic example is that in which feeding was always accompanied by the ringing of a bell.



In due course it was found that the ringing of the bell on its own and without the presence of food was sufficient to produce salivation by the animal.

*TEACHING A WORM TO TURN*

In another series of experiments it was found possible to educate a worm always to turn either to the left or to the right. This was done by confining the worm in a T-shaped tube. If it made a turn at the junction in the wrong direction, it was punished by the application of an electric shock. It was then found that after about three hundred applications of the shock the worm was completely educated to turn in the appropriate direction.

*REWARDS AND PUNISHMENTS*

There has been considerable controversy as to whether training by reward or training by punishment or a mixture of both is most efficacious, although there seems no doubt in the case of human beings that training by punishment would be likely to produce more consistent results than the reverse.

An even more spectacular conditioning experiment has just been completed in the United States. The subjects were pigeons and they were trained to inspect a continuously moving set of transistors for sealing cracks. This they were able to do far more rapidly and reliably than were human beings, but, unfortunately, the labour unions have prevented this dilution of labour!

*EDUCATING COMPUTERS*

When the education of a computing machine is considered in the same light as that of the animal, a difficulty becomes immediately apparent. This is that a machine, from its very structure, has an almost perfect faculty for the learning and recitation of facts, always assuming that the correct programme is placed in its store.

This means, for example, that if a computing machine is required to 'learn' a poem, such learning can be achieved perfectly, supposing that an input programme is present in the machine which causes the poem to be recorded in the store, and that an output programme is also available so that the results of this recording can be typed out on a signal from the human operator.

*BEGGING THE QUESTION*

This, however, begs the whole question of machine learning, since the machine is simply doing what it has already been told to do by the human operator, and in this respect does not differ in principle from a typewriter upon which a human being records the poem which is then available for posterity to read as often as is desired.

*THE SHOPPING MACHINE*

The first published learning programmes for a computer were written by A G Oettinger[1] in 1952. They were on a slightly more sophisticated level than mere learning and repetition and probably the most interesting of his developments was that of a 'shopping' programme. In this there were recorded in the machine's store the names in code of a number of shops. Associated with each of these shops was a list, again in code, of the types of items which could be bought therein.

The education of the machine was as follows: the code number of a particular item was presented to it. The machine then proceeded to test each shop to see whether or not it sold the given item. If the desired item was sold by the shop, the next thing to be investigated was whether or not it was in stock.

*SIGNAL TO OPERATOR*

Now in modern computer usage in industrial and commercial organisations the question of stock would be decided by an internal stock valuation contained in the machine memory; in Oettinger's programme, however, the machine signalled when it arrived at a given shop to a human being who was operating the conditioning device.

If the item was in stock, an appropriate signal was given by the controller; if not, the reverse. In this way the machine was able to test a range of shops for a range of articles and to associate

Dr Booth first suggested the possibilities of machine translation of language and was responsible for most of the pioneer work in this field. The Computational Laboratory at Birkbeck College (University of London) gave the first demonstration of machine translation using punched card machinery in 1949, and thus forestalled the American demonstrations of the same type which were given in 1954.

Apart from the continuing interest in machine translation, Dr Booth's group at Birbeck College has been for some time occupied with investigations on computer simulation of learning processes, both in animal and human beings.

'A machine, from its very structure, has an almost perfect faculty for the learning and recitation of facts, always assuming that the correct programme is placed in its store'

availability of articles in its own store with the given shop code numbers. Thus, after a sufficient period of education, it was possible for the machine to go immediately to an appropriate shop when presented with the code number of the item desired.

### *PROGRAMME VARIANTS*

As in real shopping, it sometimes happened that a shop became out of stock of a given item, and Oettinger's machine programme allowed for this also, so that if, after playing the shopping game for some time, the human operator decided that a given shop was to be out of stock of a given object, the machine would then search around among the remaining shops until an alternative supplier was located, and then use this supply until it too was exhausted. The variations on programmes of this sort are of course very numerous, but all follow the same general logical pattern and all, in the opinion of the present writer at least, contribute very little to a true understanding of learning processes in the human animal.

It is interesting to note that a programme for performing a shopping operation of the type just mentioned can be written with as few as seventeen computer instructions together with a number of storage locations appropriate to the number of shops and articles for which searching is designed.

### *AN EARLY PROPOSAL*

The first published suggestion that the simulation of intelligent behaviour by a machine should make no assumptions regarding the machine's built-in structure appears to have been made by the present author in 1953.[2] The proposal was that the machine should be treated in roughly the same manner as a human infant. Signals for reward and punishment should be set up and in some manner these should cause changes in the contents of the machine's store, without in any way prejudging the results of the changes. Such alterations of the store contents could fairly simply be made either by means of a source of random numbers or by switching off the machine and assuming that when it is switched on again its storage device contains random numbers.

Although it was clear that such a training programme for a machine could fairly easily be tried in practice, a numerical investigation showed that significant practical trials were not feasible on the rather slow machines which were available in the early 1950's, and for this reason no experiments in the field were carried out.

### *CATEGORY COUNTS IN TRANSLATION*

In the field of machine translation it soon became obvious that some means of machine learning would be required if ambiguities were to be resolved, and the group at Birkbeck College worked out a method, known as the method of category counts, to enable such a learning process to form part of a machine translation programme. The ideas behind this process were extremely simple and have in fact been mentioned in a previous article in this journal.[3]

In essence, in order to decide between various ambiguous meanings of a given word contained in a passage to be translated, all that is necessary is for each of the technical words contained in the machine dictionary to have associated with it one or more category numbers. These numbers, which form a sort of universal decimal classification, are typical of the fields of human endeavour which the dictionary contains; thus, for example, 1 = medicine, 2 = mathematics, 3 = physics, 4 = engineering, 5 = sociology, 6 = psychology, 7 = chemistry, and so on.

As each word in the given text is received by the machine, unity is added to the storage location associated with the appropriate category. Thus, after a number of text words have been processed,



it will be found that a large number of counts occur for the category of the main subject under discussion in the given passage, whereas fewer counts occur for all others.

### *RESOLVING AMBIGUITIES*

When an ambiguous word is encountered, this will have several category numbers, each associated with one of the ambiguous meanings. To select the appropriate meaning, all that is necessary is to examine the category counts held in the machine store and to select among the ambiguous meanings that which is associated with the maximum category count.

Naturally a programme of this simplicity has many faults, particularly that if abrupt changes in subject occur ambiguities may be incorrectly resolved. Means are, however, readily available to take care of a situation of this sort; for example, two sets of category counts can be kept, the first a long-term one to cover the main subject, and the second a short-term one which is arranged so that at the end of each paragraph or even each sentence it is cleared to zero.

### *SUPERFICIAL RESEMBLANCE TO LEARNING*

The important things to notice about mechanisms of this sort are firstly that they are conceptually simple, and secondly that, although they bear a superficial resemblance to at least some types of human thinking, they reveal nothing about the way in which a human being differs from an unthinking animal or from a computing machine.

### *FIRST PRACTICAL ATTEMPT*

The first practical attempt to educate a machine without initial programming assumptions was made by R M Friedburg[4] in 1958, and an indication of the type of process involved will show how very far we still are from emulating the behaviour even of the simplest animal on a digital computer. Friedburg, using one of the most powerful IBM computers at present available, sought to make the machine produce a simple programme of its own without any assistance on the part of the human programmer. This was done by generating random numbers which could be interpreted by the machine as instructions.

A set of 64 such random numbers could be strung together and then tested to see whether the resulting programme performed the desired operation. Suitable criteria for reward and punishment were devised and with these the process of educating the machine was commenced.

The type of operation which the machine was to be taught was of the very simplest, for example: 'Transfer the number at the input of the machine to the output', and again: 'Generate the result of the logical operation "a and b" where a and b were the binary numbers zero and one.' In other trials the inclusive logical 'or' was the required operation, and again the logical operation 'not if, then.'

There is no need to discuss the underlying programming techniques which were used to produce the desired effect. The chief among them was a mathematical generator for pseudo-random numbers. The interesting thing is that a number of learning programmes were tried, each dignified by a pet code name. Two of the early ones were called 'Teddy' and 'Herman', and the results of the operation of these programmes are shown in the following table:

| Logical Operation | Teddy Trials | Teddy Successes | Herman Trials | Herman Successes |
|---|---|---|---|---|
| 'and' | 24,896 | 15 | 225,508 | 3 |
| 'inclusive or' | 44,539 | 9 | 97,978 | 6 |
| 'not if, then' | 18,633 | 8 | 141,306 | 3 |

### *VARYING EFFICIENCY*

It will be noticed that the programmes differ in their efficiency, but that in any case a stupendous number of trials was required to produce even the simplest education. For example, nearly a quarter of a million to produce three successful repetitions of the logical 'and' function.

In the simplest programme, 'Transfer input to output' similar large numbers of trials were required. In this case six different types of learning process were attempted and the numbers of trials required to produce correct education are given in the following table:

| Programme | Average number of trials for success |
|---|---|
| Samson | 4,603 |
| Herman | 2,890 |
| Teddy | 1,360 |
| Ramsey I | 416 |
| Homer | 321 |
| Ramsey II | 60 |

*Average number of trials for successful execution of the operation 'Transfer input to output'*

It will be noticed that there are two versions of a programme called 'Ramsey' and that the final one is the most efficient. It is important to notice that there has been no question of manipulation in producing this final and not too disappointing result of 60 trials. All that was done was that a different interpretation of random numbers was used in 'Ramsey' final. This in no way, as far at least as a human being can judge, influenced the way in which the programme produced its final result.

'Although many things are possible for the modern high-speed computing machine, none of these as yet approaches the miracle of even the new-born human brain'

### *IMPORTANCE OF EXPERIMENT*

The work of Friedburg is of great importance for our future understanding of possible educative processes for computing machines. It seems quite certain that vast numbers of repetitions will be required if machines are to learn on their own. Furthermore, our present understanding of the types of inbuilt unit logical operations which can lead most efficiently to learning on the part of the machine is at present very limited indeed.

It may be that the net of neurons which is present in the brain has logical properties which make learning a faster operation than was the case with the operations used by Friedburg and his collaborators, although a more likely explanation of the superiority of natural processes is to be found in the parallel operation which is made possible by the great number of neurons which are available.

### *IMITATING THE BRAIN*

At the present time a number of experiments are in process in different laboratories to simulate the behaviour of such neural nets, but the difficulty with all of these is that they make some fundamental assertion regarding the way in which neurons interact with one another; for example, a typical case is one in which a neuron produces an output if more than a certain number of its inputs are simultaneously stimulated. Furthermore, the definition of whether or not a given input is stimulated depends upon the number of times that input has been used previously.

Using simulation procedures of this kind progress has been made in such things, for example, as causing a network of only a few hundred simulated neurons to remember in some sense a pattern which is presented to it, and to reply with a given output when this pattern is re-presented with a slightly different orientation.

### *CHARACTER RECOGNITION*

One of the main incentives for work of the latter kind lies in the field of character recognition. Several *ad hoc* devices have been designed for this purpose and at least one is already in commercial production. These, however, appear to possess severe restrictions in that the recognition of characters which have quite small variations from the original type face may necessitate the complete redesigning of the system. Animals and human beings recognise type of different founts and in different orientations with ease and if such recognition is a function of their brain structure, it is to be hoped that by an investigation of idealised neuron nets the answer may be found to the problem of designing recognition machines of great flexibility.

These remarks have necessarily been of a somewhat superficial kind. Readers desiring more detailed information can examine the original papers which are mentioned at the end of this article. The general outcome, however, of all of this work is that the surface of the problem of machine learning has only just been scratched, and although many things are possible for the modern high-speed computing machine, none of these as yet approaches the miracle of even the new-born human brain.


### *REFERENCES*

[1] Oettinger, A G: Phil. Mag. *43* (1952) 1243.

[2] Booth, A D and Booth, K H V, *Automatic Digital Calculators* (1st edn) pp 211-216. Butterworths, London, 1953.

[3] Booth, A D, AUTOMATIC DATA PROCESSING, *1* (6) (1959) 18.

[4] Friedburg, R M, *et al*, IBM J of research and development 2 (Jan 1958) 2, and *3* (July 1959) 282.




**AMERICAN REPORT** from John Diebold and Associates, New York

# On The Track Of Lower Costs

In spite of extensive use of punched card and bookkeeping equipment, clerical costs at the Southern Railway Company (of America) continued to swell. Now with a machine to do a 'complete job' the tendency has been cut back.

THE South Carolina Canal and Railroad Company was founded in 1830 when only twenty-three miles of railroad track were in existence in the United States. Today, as the Southern Railway System, Southern Railway Company and its group of associated railway companies serve fourteen Southeastern states with a total of over 8,000 track miles. In keeping with its slogan of 'Serving the South,' Southern has invested some $450 millions in capital improvements since the second world war and by 1953 was the first major railroad to complete dieselisation. Southern also pioneered with electronic yard classification of freight cars, in the employment of mechanised methods for the maintenance of roadway and tracks, and in the utilisation of end-to-end radio communication systems for freight trains.

Southern was one of the pioneer users of electronic data processing equipment for commercial applications in America and was the first railroad to program and place freight revenue accounting on a computer. Visitors have come from many other railroads and from foreign countries to observe Southern's IBM 705 installation in Atlanta.

### *SOUTHERN'S APPROACH TO ELECTRONICS*

In spite of extensive utilisation of punched card and electro-mechanical bookkeeping equipment, Southern in 1954 was faced with an unmanageable growth rate in clerical costs.

The Accounting Department employed over 1,400 people and more than 900 were occupied with freight revenue accounting alone. Pre-war minimum clerical costs had been about four dollars per person per day, but after the war they began to go up, and have since risen 350 percent to $18 per day.

The need to accommodate a larger volume of data with improved accuracy and speed, together with zooming clerical costs, presented the railway with a dilemma which caused management to welcome the possibilities offered by electronic computers. As one accounting manager with 20 years' experience in freight document processing observed: 'We had for years been searching for a machine that would do the " complete job " and along came the computer—the first machine we could even hope might present an answer.'

### *KEY QUESTIONS*

The possibility of utilising an electronic computer for railway accounting was first suggested at Southern in June, 1954. The Controller and his planning staff addressed themselves directly to two key questions:

1. Are electronic computers really capable of processing business data?
2. Does Southern Railway have a paperwork