National Bureau of Standards
Washington, DC
Electronic Computers Laboratory

Computer Seminar Notes

Russell A. Kirsch
Experiments With a Computer Learning Routine

Friday, July 30, 1954

COMPUTER SEMINAR NOTES

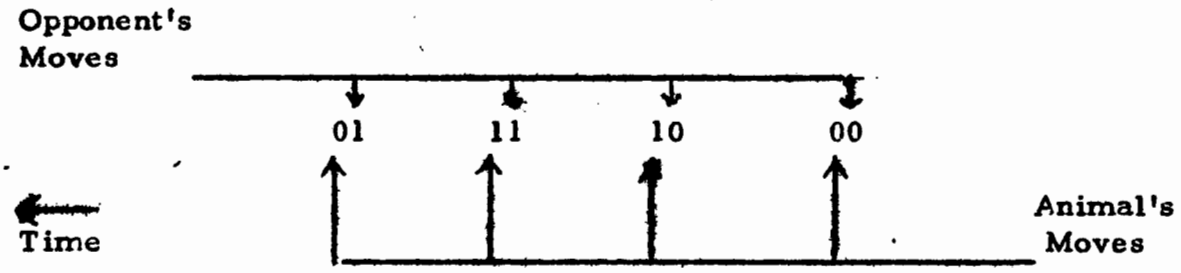
Friday, July 30, 1954

Speaker: R. A. Kirsch
Subject: Experiments With a Computer Learning Routine

In the October 1953 issue of the Proceedings of the IRE, Shannon described an experiment in which two computers played against each other in a coin matching game. One computer attempted to match a number generated by another computer which tried to unmatch. In this lab, the thought occurred to Dr. R. J. Slutz that we could attempt a similar experiment. Consequently, he prepared a routine to occupy SEAC's acoustic memory, and R. A. Kirsch prepared a routine to occupy the electrostatic memory. The nature of the two routines was entirely unknown to the opposing programmers. It was hoped that by playing the game at computer speeds and by allowing the computer to tabulate the score, a good indication of which routine was superior could be obtained since a great many trials could be made. In this report, a description will be given of the routine prepared by R. Kirsch and of later experiments with that routine playing against specially arranged opponent routines.

One approach to this problem could be to fit this situation to a model usually treated in the mathematical theory of games. This approach was not used. Instead, an attempt was made to devise what seemed like a model of the way in which an animal learns. Then this "animal model" was used to play the coin matching problem.

Evidently, the model has to have some learning mechanism which responds to stimulii of different types. The stimulii are defined in the following manner. At each "matching of the coin", two binary numbers are generated, one corresponding to the opponent's move and one corresponding to the animal model's attempted guess at what the opponent's move is. The game might then look something like this:

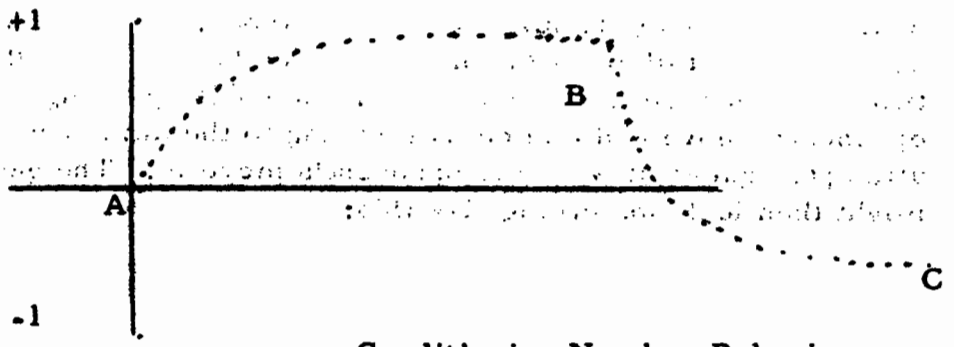


Win for: Opponent Animal Opponent Animal

The stimulus that the animal reacts to at any time is defined to be the group of 4 pairs of moves immediately past. Thus there are eight binary digits specifying the stimulus at any time. In the example above, the number 01111000 designates the stimulus, of which there could thus be 256 different types.

The learning mechanism consists of the animal becoming conditioned to a certain opponent's move following each stimulus. The animal model notes, for each stimulus, what move the opponent next makes. Then, the next time that same stimulus occurs, the animal duplicates the move of the opponent that followed the same stimulus previously. The more the opponent repeats the same move after any given stimulus, the more the animal model becomes "conditioned" to that move.

The conditioning mechanism is exponential in nature. In each of 256 cells corresponding to the 256 different possible stimulii, a number is stored which varies between +1 and -1. Whenever the opponent plays a 1, the number in the appropriate cell is increased by some fixed fraction of the difference between its previous value and +1. Thus if the opponent always follows a particular stimulus with a 1, the conditioning number will behave as in the figure in the region from A to B.



Conditioning Number Behavior

If the opponent should suddenly start playing zero's the number will go from B to C as shown in the figure. Part of the input data for the routine consists of a "time constant" for this exponential function, i.e., a parameter which determines how many identical repetitions of the same opponent move after a given stimulus will make the conditioning number reach 63% of its final value.

This exponential conditioning or learning function has characteristics similar to those of an animal. At the beginning, a single opponent move has more effect on the animal's learning than after many repetitions of that move. Also, when a stimulus is suddenly followed by a different opponent's response from that which the opponent had been making in the past, the conditioning number, like an animal's conditioning, becomes rapidly reversed at first, and more slowly later.

It is assumed that in the game situation, the opponent will, from time to time, change his strategy. This indicates that it would perhaps be of use to have a deconditioning or forgetting mechanism. Accordingly, provision is made, in the routine, for periodically deconditioning all conditioning numbers toward 0 by a fraction of their magnitudes. The analogy to an animal, here, is in the case when the animal has become conditioned to a certain second stimulus and suddenly the initiating stimulus is no longer applied. After a long enough period, he "forgets" the conditioning he previously displayed.

One final characteristic of this animal learning model must be mentioned, the mechanism whereby it finally determines the move to make. At each point in the game, the routine "knows" what the past 4 pairs of moves are and what the conditioning number corresponding to this stimulus is. A random number which ranges from +1 to -1 is then generated. This random number is compared with the conditioning number. If the random number is algebraically less than the conditioning number, the routine plays a 1, otherwise it plays a zero. This was originally inserted in the routine to prevent a simple analysis on the part of the opponent from discovering the "strategy". However, its importance became more evident in later experiments.

The first actual game played between the two coin-matching routines was run on SEAC on January 18, 1954. As a statistical test of significance of the results, it was agreed to stop the game after 10,000 trials had been run and to inspect the score. From information received from the statistical Engineering Laboratory at the Bureau, it was learned that the following statistical test could be applied to the results:

If 10,000 trials are made, and if at the end of that time one routine is ahead of the other by 196 wins, then there is at least a 2% bias in favor of the winning routine and the confidence level for that estimate is 98%.

After 10,000 trials were run, the difference in scores was 186, so it was considered not significant enough to make generalizations from this score.

In addition to the games played between the two coin-matching routines, several games were played between the machine and human opponents. In these cases, the stimulii were allowed to assume only from 4 to 32 different values depending upon the opponent rather than 256 as when playing against an automatic routine. The results were not always significant largely because it was not possible to play long enough games. Nevertheless, the machine's opponents did express surprise at the way the machine appeared to "learn" what their strategies were in many cases.

The most interesting games, however were those in which a special opponent routine was written to play against the animal model-learning routine.

In one game the opponent was a routine which played the inverse of the animal model's previous move 25% of the time and the remaining 75% of the time played randomly. The animal's conditioning exponential was adjusted to have a time constant of 4 moves, and the tallying routine was instructed to stop the game when there was a difference between scores of 250. The final score was 2960 to 2710 in favor of the animal model. In another identical game where the only

change was in the use of a different table or random numbers by the opponent, the animal model again won decisively, 3517 to 3267. In the next game two changes were made. The opponent routine again played the inverse of the animal's previous move, but only 10% of the time. The remaining 90% of the opponent's moves were random. The conditioning exponential time constant was also changed to 30 moves. This game was again won by the animal, but after a much longer time since its learning was slower and its opponent more capricious. The final score here was 23,131 to 22,881. A post-mortem analysis of the conditioning numbers showed that none of them was as large as 25% of their maximum values when the game finished. If this game had progressed farther, the animal's learning would have increased.

The main information gained from the above experiments was a measure of the relationship between randomness in the opponent, rate of learning of the animal, and length of game. In the experiments that followed, an attempt was made to determine the relationships between complexity on the part of the opponent and complexity in the animal's learning mechanism insofar as they affect the ability of the animal to "crack" the opponent's code. The physical situation that was simulated here was one of an animal adapting to an environment. It must be assumed that the environment is not capricious although it may be considerably more complex than the animal. To simulate this situation, the opponent must play a fixed strategy.

Since the opponent's strategy is fixed, the need for the forgetting mechanism described above is no longer present. This mechanism was therefore disconnected in the animal routine.

The final feature of the animal learning routine that is highly significant in playing against an environment that is complex but fixed is the combination of randomness and selectivity. When the animal finds that a particular stimulus implies no large conditioning number it plays some random move. If this move generally leads to a failure it remembers that; if it leads to neither failures nor successes it continues

to play randomly; and most important, when it finds a successful move, it plays that move in proportion to its success. The analogy to the evolutionary mechanism in species of random mutations and natural selection is fairly obvious.

A very interesting game was played on June 3, 1954, to test this new interpretation of the animal-learning model. An opponent routine was written which would always duplicate the animal's move of 10 plays back. Since the animal's stimulii were only defined by the past 4 plays the important question became: "Can the animal adapt to a strategy or environment which is in a sense more complex than the animal's learning or adaptation mechanism?" The conditioning time constant was set at 2 moves. The deconditioning or forgetting mechanism was disconnected, and the game was set to stop when one "side" was ahead by 1000 wins. During the first 45,500 moves, the animal was losing by amounts that were insignificant, the largest at any of the printout points being 276 (after about 30,000 moves). However, somewhere between the 45,500th move and the 46,000th move the animal suddenly seemed to "catch on" and thereafter won every single move. What actually happened was that in the absence of any significant learning the animal kept playing random moves until finally it played a string of more than 10 identical moves. Since the environment was duplicating the move of the animal 10 plays back, the environment began duplicating its own moves. The animal looking at the past 4 moves and detecting a successful strategy continued to play the same move and to become increasingly successful until it won all the remaining moves.

In this particular case there was a single strategy which the animal could not determine through analysis of the environment but which it did "discover" through playing randomly. Other environments against which the animal was matched did not have such clear cut "solutions", but in several of them the animal showed a decided ability to adapt successfully.

Other experiments were performed in which an attempt was made to insure that once a successful adaptation had been

discovered, it would not be lost through random variations, but no decisive results were obtained.

The last experiment that will be described here is one in which a fairly simple environment was created which did not possess the ability to analyze the animal's strategy, but which nevertheless decisively beat the animal despite (and in fact because of) the animal's attempts at adaptation. To understand the characteristics of this environment, consider the following repeating sequence of binary digits:

	01011100	01011100 ...
	12345678	12345678
	A	B

Note that the first 2-binary digit combination to occur is 01 and it is immediately followed by a 0 in position 3. The next time that same pair of digits occurs is in positions 3 and 4 of cycle A but this time it is followed by a 1 (in position 5). In fact this sequence of binary digits has the property than every combination of two binary digits occurs and is first followed first by one binary digit and the next time it occurs it is followed by the opposite binary digit. All such three binary digit combinations occur once before any one recurs. A method was devised for generating such a sequence for any length of binary digit combinations. Then using such a sequence consisting of all 5-binary digits combinations (a sequence of 2^5 or 32 bits length), the animal was set to adapt to this environment, but the animal was only capable of looking at the last 4 moves of the environment. Naturally, each time the animal attempted to duplicate the environment's response that last followed the given stimulus, the animal played the wrong move. The final score was 3508 to 3758 with the animal losing decisively!

A great deal of speculation can be indulged in here as to whether this last situation involves what may seem like a neurotic adaptation by the animal. Actually, most of the analogies that have been drawn in this discussion are highly

speculative in nature. Any rigor in the approach to a description of animal learning is, if anything, conspicuous by its absence. However, if this speculation can induce some further speculation which can be shown empirically to have validity, it will have accomplished its purpose.

One final word may be mentioned here in regard to a possible interesting demonstration using the model described here. Suppose that a manual opponent decides upon a simple strategy with only a small random element in it. Let this opponent play against the computer until each one of the stimulii to which the learning model responds will have occurred several times. There will then be stimulii of two types: those for which there is a conditioning number whose magnitude is large, and those for which it is small. However, the former can be considered to be elementary products which appear in the disjunctive normal canonical form of the logical function which describes the opponent's strategy. The latter type correspond to terms that do not appear. By techniques of the type described by Ledley in N. B. S. Report No. 3363, a simple logical expression can then be obtained which describes the opponent's strategy. In this manner, the computer cannot only be made to discover a way to beat a simple opponent, but it can also tell that opponent the logical formula for the strategy he is using!