[*Mechanical Translation*, Vol.7, no.2, August 1963]

Learning and Translating by Machines

by John F. Tinker, Research Laboratories, Eastman Kodak Company, Rochester, New York

To translate well, a machine must be furnished with rules that relate meaning to words. These rules may be expressed in terms of probabilities, if they cannot be expressed precisely. Less useful are descriptive rules, particularly those using concepts of psychology. That these rules can be satisfactorily formulated is strongly suggested by the fact that a child of four can adequately manipulate language.

To learn, a machine must be furnished with rules, besides those for performance, for critically evaluating its performance, and for modifying the performance rules. Learning is the process of successfully modifying the performance.

Creativity in humans is an example of this learning process. A human cannot perform better than his teacher if his rules of critical evaluation are identical with his teacher's. If he is to perform creatively, he must be able to modify all three elements of learning—performance, critique, improvement rules—not merely the first element. To teach a student to be creative, the teacher must specify the rules heuristically, not precisely. This is the same problem as programming a machine to learn. That the former can be done suggests that the latter is possible.

A good guide to the maximum amount to bet is the product of the probability of winning and the amount won. Spending on research is similar to wagering, and a sensible maximum to a research budget is the product of the probability of successful outcome during the budget time and the expected profit. If the probability of outcome is zero, the research budget should be zero, regardless of the profit.

Learning and translating by machine, it has been suggested, are fields in which the probability of successful outcome is zero. But is this so?

A computer can do anything that you can explain, carefully and patiently, to a child of four. A child of four can talk. This fact suggests that a machine can handle language.

Most machines existing today perform simple manipulations following precise, simple rules. More complicated rules can be followed, provided they are precisely expressed. For instance, the ability to follow different sets of steps, depending on the value of a particular number, is one possessed by a number of computers. Programs making use of random numbers are known. Rules stated in terms of probabilities, or heuristically stated, are satisfactory provided the statement is precise.

Translation by machine can be accomplished if the rules of language can be stated precisely. The rules of grammar and meaning have not been so stated, partly because of their complexity, and partly because descriptive rules are adequate for teaching language.

Rules of grammar exist and have been stated precisely enough for computer use in many instances. Rules of meaning pose more difficulties. Meaning is that attribute of a word which, by common agreement, refers to a defined concept. The agreement is reached by communicating in language. Definition results by comparison with related words. Can the meaning of words be specified by precise rules? Do these rules exist?

A child of four can construct grammatical sentences expressing an idea. He can extract meaning from sentences, and construct sentences from meaning. The rules he uses to do this have been given him in a descriptive fashion, and have been refined by trial and error. He cannot express the rules precisely, yet his use of them shows that he understands them precisely.

The child applies these rules to spoken language and, when he is older, to written language. Spoken and written language are not identical, but they stand in the same relation as do music and musical notation. In each case, the notation is adequate to determine many aspects of a performance but leaves the performer considerable freedom. All the discussions regarding interpretation, composition, and meaning can be transposed from music to language with considerable pertinency. Many terms are defined solely in terms of notation: a musical note, staccato, or presto are things in the notational scheme. They have consequences in the music but no specific counterparts. In the same way, a written word or sentence has consequence in the spoken language without necessarily having a specific counterpart.

A player piano is an example of mechanical translation to musical performance from musical notation. The roll of paper, which directs the motion of the keys, has on it most of the signals in the notation and a few more that have been supplied by a human programmer. A roll of paper is pulled across a sensing device and the notes are played when a hole crosses the sensor. The pitch is determined by the value of one coordinate, and the position in time is determined by the other. The tempo is related to the second coordinate so that, at a steady tempo, one bar of notation corresponds to a certain distance along the paper. When *ritardando* or *accelerando* appears in the notation, the programmer changes that distance. Note that terms "bar," "ritardando," and "accelerando" influence the music without appearing in the performance. The opera performers sing "Zitto, Zitto" ("softly") but not "piano" (softly), although both of these words appear in the notation. An ordinary player piano cannot vary its dynamics and is quite unable to play a loud note and a soft note at once, although this is not often required. Musical notation and the roll of a player piano are different repertories of signals designed to allow performance of music. The player-piano roll is more machine-oriented. The machine can play because a human programmer has formulated the rules precisely. It can play, but it can't learn.

"Learn" is an example of a concept that is defined more easily in a descriptive way than in a precise way. It is easy to rationalize when defining things, and so build conclusions into definitions. Rationalization of the opinion that learning is forever beyond machines is contained in the definition of learning as the transfer of control of a process from the conscious to the subconscious mind. That definition is a systematic relation between words of the sort that Samuel Johnson had in mind when he began the first dictionary. It is more useful for defining the psychological terms than for defining learning.

How does a student learn to play the piano? He is presented with, and perceives, three things: (1) a description of the process of converting the notation into a performance: (2) a body of rules by which to judge his performance; and (3) a series of remedies for common faults of performances. These three elements are the necessary and sufficient foundation of any learned skill: technique, critique, and hints for improvement. Once the student perceives the details of this program, he is able to increase the elegance of his performance on the piano to the level demanded by his critical judgment. He is able to begin learning. Learning is the process of refinement of these three elements—technique, critique, and improvement procedures.

Solving of an equation by successive approximation is an elementary learning process. The student is presented with the equations, with a criterion of successful solution, with directions for finding a bad solution, and with a procedure for improving a bad solution. He begins with a problem, a critique, a beginning, and a technique for improving. By going through the improvement procedure a step at a time, the student finds better and better solutions, finally reaching one that lies in the range specified by the critique as good enough. The success of the answer depends on the critique. If the solution is required to two significant figures, a few steps are enough; if a more exact solution is required, more steps must be taken. In general, refinement of the critique lengthens the learning process.

Learning a solution of an algebraic problem by successive approximation is elementary because each of the elements is precisely formulated. This is not true of learning to play. The equations specify precisely how the answer is related to the problem, but the musical instructions specify only incompletely how the performance is related to the notation. The directions for finding an approximate solution are simple, but the directions for beginning to play are complicated. The algebraic critique is exact, but the musical critique is approximate. Each element in algebra is precisely defined, and each element in music is heuristically defined. Learning to play requires the student to formulate and systematize the missing rules. He must learn not only the elements—technique, critique, and improvement—but how to refine those elements.

Refinement of the elements may be done automatically on an elementary level. The more adroit refinements that humans accomplish require insight and may involve creativity.

Creativity, insight, and humor are the more remarkable of outcomes of an instinct—the instinct to find semblance. The human mind continually searches lor similarity and is rewarded by the perception of similarity. Imitative behavior springs from this vigorous drive: a child imitates those around him to find similarity between his own actions and those of others. Cats imitate one another, monkeys mimic men. These examples are more obvious than the more sophisticated imitations of more mature people.

As a man matures, his search for similarity is carried to more and more abstract levels. A child imitates his father's movements; as he grows older, he begins to imitate behavior, then to imitate principles of behavior. Finally, he begins to see similarity among abstract propositions.

There is a French proverb, "Plus ça change, plus c'est la même chose." ["The more things change, the more they are the same."] It is a synonym of "There is nothing new under the sun." In both proverbs, beneath the commentary about slowness of change, we can detect the tendency of the mind to interpret the new in terms of the old, to seek similarity.

The search for similarity is clearly an advantage for a species. Often the solutions to yesterday's problems can be applied to today's problems, if they are similar. The catalogue of resemblances is a help in choosing a successful course of action with incomplete information.

The search for similarity is unceasing and does not always yield a useful result. Two things may be incongruous yet have similarity from some point of view. This condition can be built into a joke; humor depends on the logical relation of the incongruous.

If the logical relation is novel, it constitutes a creative insight. In this case, the rules for critical evaluation are important. An individual, taught a sufficient number of rules, is likely to assume that there are no more rules. Then the critique includes a rule that stabilizes itself, and vetoes any creative insight.

Creativity is the outcome of a sophisticated and knowledgeable search for similarity. To learn to play most successfully, the student must be taught in such a way that his insight and creativity can modify his critical evaluation of his performance. The critique must be specified, not precisely, but heuristically.

Teaching for creativity is the same problem as programming a computer for learning. In each case, the directions for judging the success of the task must be allowed flexibility, and further directions concerning this flexibility must be given.

A computer and a child show no internal resemblance, nor do they follow directions in the same way. The child's ability to deduce general rules of behavior from many examples, some of them inappropriate or wrong, has encouraged his teachers not to formulate the rules precisely but to rely on repetition and imitation as the mechanism for absorbing the rules.

The computer cannot absorb rules in this way. But, in talking, the child makes use of rules, even though they are not consciously or precisely formulated. The child makes use of these rules in such a way that, were the rules precisely formulated, a computer could follow them.

To be able to manipulate the language as successfully as he does, a child must have at his subconscious command a series of rules. How he deduces these rules, and the mechanics of his use of them, are not important to our argument. His use of them demonstrates that they exist. To put them into the form that a computer can use requires, not that they be invented or discovered, but that they be formulated.

To learn, an entity must have several choices of behavior; a means of judging the success of its choice, and a way of improving its judgment. It is difficult to design a computer of this sort and harder yet to program a present-day computer to behave this way, but it is possible in principle.

To produce high-quality translations, a computer must be able to learn to manipulate language and meaning. When the relations between language and meaning are specified, no matter in how complicated a way; when the criteria of high-quality translation are outlined, with suggestions about how to improve the criteria; and when the mode of improvement for each criterion is formulated, a computer can be built to produce high-quality translations. With technique, critique, and improvement rules specified heuristically, machine translation is at hand.

A child of four can do it—why not a machine?

Received November 5, 1962