≈ what was sent to McCarthy : Mon Jan 9, 1956.

Outline of Approach to T.M.

1) Large set of imp. problems that all seem to have certain basic sub-problems in common.

Some of them:

a) What is probability? - how to eval. in a gn. situation.

b) Genetic str., ~~[illegible]~~ organic evolution.

c) Understanding of psych. behavior as optimum behavior in an envt — ("learning behavior" in partic.)

d) Analysis of some random nets of pseudoneurons and deciding on optimum network strs.

e)

f) Info packing problem. → what is "original" or "creative" work?

2) These sub problems are:

a) ~~#~~ One has a bunch of ~~usual relatively successful~~ methods of dealing with probs that has been relatively successful in th. past. One is gn. a new prob. How can one use th. old methods to suggest new methods that have a hy probty. of being useful with th. new problem?

b) ~~Among~~ A decision is to be made betw. several hypotheses or course of action alternatives. Evidence from various relatively indip. sources, indicate different choices to be optimum. How should one combine this info in an optimum way to make a decision?

c) Suppose one has a computer characterized by a set of parameters. One ~~gives the comp~~ fixes th. parameters and gives the computer a set of problems to solve. One then gives th. params. a new set of values and gives th. computer th. same set of problems. ~~[illegible]~~ Can one devise a reasonable quantitative measure to determine how much "better" (or worse) th. 2nd set of solns. is than th. first?

d) Th. prob. of optimum construction methods. ~~[illegible]~~ Suppose one has a design for a computer ~~[illegible]~~ in terms of input-output behavior. What is a good ~~is the best~~ way to realize this computer physically? — i.e. how can one best break down its functions into "blocks"?

I have spent much time in analysing these 4 Q.'s in general and specific cases, untill

Outline of approach to T.M. (2

I feel at present, that I can usually solve these probs. in any partic. case.

Th. next ~~■~~ step was to find a partic. problem that ~~[illegible]~~ involved all of the above problems. Several possibilities that were investigated were:

1) optimum structure of net and of pseudoneuron, for "random" net model of human brain

2) Chess or checker playing machines.

3) Machine to learn to answer simple Q's in English, from a training sequ. of Q's and A's.

4) Machine to learn arith., alg. and math, in th. sense of being able to "fill in blanks" of incomplete equ. or proof.

?) 5) prediction of a time series in which both time, and th. elements of th. time series are discrete, discontinuous quantities.

For a preliminary analysis, 4) was chosen. Some reasons for this choice:

a) ~~■~~ problem 1) seemed to ~~be~~ depend for its soln, on just what kinds of problems were gn. to th. net. For this reason, analysis of this problem seemed at least as difficult as ~~■~~ 2, 3, 4, or 5, with the additional problem of ~~[illegible]~~ stochastic behavior ~~■~~ a random net

b) both 2 and 3 seemed to require much preliminary analysis peculiar to chess, checkers or language, and not too relevant to the ~~[illegible]~~ problems common to all such machines. ~~Each of these problems would require very much [illegible]~~ Very much time would be spent getting to the point where questions involving learning (which are the really interesting probleems) arise.

?) c) In 5) the nature of the time series must be decided upon. To have the elements of th. time series be examples of 2, 3 or 4 seems reasonable, so 5) is not essentially different from 2, 3 or 4.

≈ what was

Outline of approach to T.M.

d) The advantages of problem 4) are many: (3

1) Arithmetical problems of much simplicity can be devised, so that the amt. of preliminary analysis of math. that is nec. befor problems involving learning behavior arise, is minimized.

2) Problems involving any degree of complexity or any amt. of "creative" insight, can be easily devised — so we can always have problems of suff. difty to test the efficacy of our approach.

3) Often in math., the training sequ. is clear — i.e. it is often obvious as to what preliminary learning a machine must have had, befor it can approach a gn. problem with reasonable

.17 probty. of success in a reasonable amt. of time.

(to 4).22)

The work that I am doing ~~right~~ now, involves the application of some special statistical techniques to learning simple arithmetic. In particular, the problems involve the weighting of statistical conclusions drawn from different methods of analysis of the same problem. Also of much importance, is the devising of new methods of statistical analysis of reasonable probability of success, from older methods, known to be useful.

Specifically, let us consider a simple kind of arithmetic problem. Suppose we are presented with a series of ~~two row arrays~~ 4 x 7 rectangular arrays of the foll. type:

```
n n n n n n n
n = 0 1 1 0 n
n n 0 1 1 0 n
n n n n n n n
```

The n's (nulls) and the = sign always occur in the same place in each of the successive arrays but the first row of 0's and 1's is arbitrary, and the 2nd row is ~~the same as the first~~ always identical to the first row. After this preliminary "training sequence" of arrays, we are

Outline of Approach to T.M. (4

gn. an array with all but one square ~~incomplete~~ specified, and we are asked to complete that ~~unknown~~ square: As

For a human being, the problem is simple. He immediately realizes that the second row is the same as the first, and that 1 is the correct completion digit.

n n n n n n n
n = 1 0 1 1 n
n n 1 0 $\boxed{?}$ 1 n
n n n n n n n n

For a simple machine, the problem is more difficult. A first approach to th. problem, would be to make a statistical investigation of the frequency of 0's, 1's and n's in th. square in question. If the sample size is too small, it would be useful to count th. freq. of 0's, 1's and n's taken over the entire array. This would increase one's sample size by a factor of 28, but the

.21 information obtained would tend to be less relevant (.to 5).08)

.22 (from 3).17

~~Th. main disadvantage of this approach, is that the results are not particularly spectacular. It is not immediately apparent, that a machine that has learned, say, addition and subtraction, is inherently capable of much more than 7 year old child. Even if one were able to teach the machine to solve literal equations, and work out literal integrals, it would not be obvious that the machine was "thinking" in any useful sense, and much less obvious that it was doing any thing non-routine, requiring ability to formulate new concepts not specifically given to it.~~

not too good.

Th. main disadvantage to this approach is that it is not particularly spectacular. Learning ordinary addition and subtraction from examples, may seem to be a bit clever, but requires no obvious originality and inventiveness. It would seem to be a very long road, to the point

In chosing
~~Choice of~~ a suitable type of intelligent machine/ to begin work on ~~should be based~~ upon the following ~~criteria:~~ some of the following factors should be considered:

c.o. ~~The problems solved by the machine should be of sufficient difficulty, so that it is clear that~~

1) ~~As rather possib~~ ~~The machine should be as "[illegible] purpose" as possible~~. As Little study as possl. should be devoted to the peculiarities of the particular ~~pro~~ type of problem to be solved. Most study should be spent on those aspects that are common to all learning problems. Chess and Checker playing machines are rather poor in th~~e~~ respect~~s, in that the machines that would play these games, would be an extremely specialized ones — also~~, A great deal of work must be done in order to understand the games themselves in terms specific enough for machine programing. Most of this work ~~would be~~ is peculiar to chess or checkers, and has little bearing on the general problems of learning and intelligent behavior.

A machine built to work mathematical problems is ~~[illegible]~~ particularly good in this respect.
Simple mathematical problems — say arithmetic — are
Another such machine that has similar difficulties, is one that would learn to answer simple questions in English, after a training sequence of question-answer pairs. ~~Another is somewhat similar, but more useful, would be a machine to translate English [illegible] a language into a symbolic logic-type form, and the inverse of such a machine~~
rather well understood. The ~~principal~~ significant problems ~~difficulties~~ ~~[illegible]~~ would ~~[illegible]~~ immediately arrise in teaching the machine, rather than in ~~teaching~~ familiarizing the investigator with the intricacies of mathematics.

'Another important characteristic ~~of the~~ that the ~~machine~~ problem should have, ~~a graded complexity. It should be possible, at the beginning of the investigation, to~~ is versatility.
~~Specifically~~ If a machine has been designed

~~chose~~ with a particular type of problem in mind, it should be possible to give it a somewhat different kind of problem, to see that ~~[illegible]~~ the designer hasn't been too specific in building the solutions to the problems into the machine. This feature is particularly easy in mathematical machines, and somewhat less so in game playing machines.

Another desirable feature is the possibility of demonstrating inventiveness, or the creation of what appear to be new concepts. In a mathematical problem machine, this can be easilly demonstrated, but with a game playing machine this feature is less obviously demonstrable.

A ~~great~~ disadvantage of the mathematical machine is that it is not particularly spectacular. A chess machine can, at a rather low level of development < even without being able to learn > beat many human opponents. ~~It is unlikely, however, that any~~ reasonably simple ~~[illegible]~~ mathematical machine ~~can compete [illegible] with most mathematicians~~ successfully ~~in learning and invention, with even the~~ most low cost ~~[illegible] powered most [illegible]~~ mathematician. ~~There exists~~ This is a simple, natural way, to measure machine proficiency. With a mathematics machine, there appears no such obvious measures of ~~[illegible]~~ proficiency. Unfortunately, the "skill" of a chess machine is not necessarily an index as to its learning ability, tho it may, under certain circumstances, be legitimately used in this way.

The possibility of demonstrating inventiveness, or the creation of apparently new concepts, is also a desirable feature. A mathematical machine would be ideal in this respect. Suppose it had learned to solve quadratic equations with only real roots. It could then be presented with one that had ~~[illegible]~~ complex roots, and we would see if it could invent complex nos. ~~It is clear, however, that the training sequence leading to this problem must indicate that such novel concepts as complex numbers are admissable solutions to~~

It is clear, however, that in order for the machine to "invent" complex numbers, ~~[illegible]~~ a proper training sequence would have to be devised, in which newly invented concepts were given as permissable solutions to problems.

For a game playing machine, inventiveness would be far more difficult to demonstrate.

Note: In ~~[illegible]~~ Manuscript sent to McCarthy, "A Method of analysis" ≡ a partic "gp" and its associated parameters.

Outline of Approach to T.M.

Where the machine would invent complex numbers because it couldn't solve a certain quadratic equation that was presented to it with ordinary real nos. ~~ordinary real numbers~~. I feel fairly certain, however, that evidence of the machines "originality" will become apparent long before it is so blatantly demonstrated.

.08 (cont. from 4) .21

Th. first method, would give a probability of about $\frac{1}{2}$ for th. digit 1, or for 0. Th. second would give a proby. of $\frac{1}{7}$ for 1 or for 0, $\frac{19}{28}$ for "n", and $\frac{1}{28}$ for "=".

In weighting the probabilities given by the two methods, one might note that, in general, the first method would usually give more accurate results, and should be therefor given more weight. (this would be a simple statement if th. i.p. method is assumed).

A criterion of "accuracy" of statistical results, has been tentatively devised, as well as some methods for assigning wieghts to various prediction methods. Tho both analyses seem rather untractable in their exact form, it is felt that in many cases, even the most gross approximations would be adequate.

In the case of the simple arithmetic problem posed above, some much more accurate predictions involve "digram frequencies.". For instance one notes the frequency of the configurations

$$\genfrac{}{}{0pt}{}{1}{1},\ \genfrac{}{}{0pt}{}{0}{1},\ \genfrac{}{}{0pt}{}{1}{0},\ \genfrac{}{}{0pt}{}{0}{0},\ 1.1,\ 10,\ \text{etc.}$$

The digrams $\genfrac{}{}{0pt}{}{1}{1}$ and $\genfrac{}{}{0pt}{}{0}{0}$ would be found to be very reliable in predictions, and therefor predictions based on their use would ultimately be given much weight — altho at first they would be given little weight because of smaller sample size than some other

Out line of approach to T.M.

prediction methods.

A ~~?~~ kind of problem that seems ~~to come~~ next ~~on the list of~~ in complexity is ~~?~~ once in which a typical presentation to the machine is

n n n n n n n
n ~ 1 0 1 0 n
n n 0 1 0 1 n
n n n n n n n

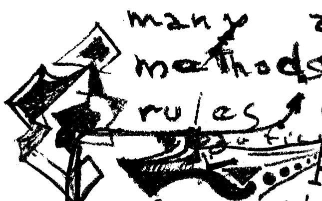
Here, ~ indicates complimentation. The first row of 0's and 1's is ~~?~~ varied randomly, and the second row is the compliment of the first.

After the machine has learned to work "=" type probems, it is necessary to use "trigram frequencies" to solve both "=" and "~" problems — to be able to distinguish between the two types.

Some hyer stages in the ~~Mana~~ complexity of arithmetic problems ~~?~~ are boolian multiplication and addition, than normal ~~arithmetic~~ addition and multiplication.

Each of these new types of problems seems to involve new methods of prediction, or learning. It has been found, however, that many apparently new and novel prediction methods can be generated from a few simple rules. ~~In particular, the use of analogy in prediction is particularly simple to use, via these rules.~~

In particular, these rules make the use of analogy in prediction, extremely natural.



~~At the present time, I am going thru a somewhat detailed analysis, to see just how well my general concepts of ? behavior solutions of problems in prediction, can be applied to a ? machine that is to learn arithmetic.~~

At th. present time I am working out a detailed analysis of the efficacy of my general prediction methods, in the problems of ~~simple arithmetic~~ learning basic mathematical processes.

Earlyer than
Jan 56

I ~~Each~~ Many approaches to problem list a few.

- nature of life.
- Self-reproducing
- Analy of Psych. behavior.
- Analysis of ~~the~~ neuroanatomy
- random nets
- optimum construction
- Evolution and ontogeny, gene coding
- Info. packing prob. (coding).
- Analysing some games
- statistical inference.

II I will use stat. inference.
problem [redacted] is instrumentation.
Why this may be easy
1) I believe only ~~a~~ very approx. soln. is by far good enuf for th. more complex parts of th. prob. That analog means will usually be nec.

e.g. f(x)'s need not be very good.

2) Th. soln. itself/ of ~~this~~ partic prob. will suggest methods of implimentation.

~~That~~ Both 1) and 2) are my own opinions based on my work with th. prob. and cannot be conclusively demonstrated at th. present time.

Note that usually 1 approach is used with many others.
Vary in ° of determinacy, randomness —
but any ~~good~~ stocastic approach $\leq$ a deterministic one.

T.M. : General description of the work and the progress as of now.

The general approach has been to ~~analysing~~ begin by analysing the methods by which human beings solve problems, and to try to describe the method in terms sufficiently specific for machine programming. Later, more attention was directed toward obtaining a method that would solve problems "reasonably well," rather than try to imitate slavishly the apparent workings of the human brain.

Much of the work has centered around determining just what "probability" is, and determining probabilities from what are apparently small samples.

First of all, in determining the "probability" of an expected event, one tries to categorize this even into a class of similar events, so that the event to be predicted has, in a sense, happend many times befor. On the basis of this classification, the relative frequency of occurance of the event ~~can be approximated~~, and the expected accuracy of this frequency, can be determined.

In general, this frequency ratio will depend markedly on just how one has decided to categorize the event. A rule that is usually used, is that the class should be as small as possible, yet lead to as sufficiently large as sample as possible. A rather ~~simple~~ elementary problem, involves the reconciliation of these usually conflicting criteria.

Somewhat more difficult, is the question of just what classifications to consider at all. Clearly, one can construct, ad-hock, a class of events, such that this class has occured many times in the past, yet, in the proper sense, the class has few members. An example of such ad-hock classification, would be the class of events consisting of the flip of a coin on Jan 3, 1958, and the rising of the sun every morning. The elements of the class, taken as a whole, occure often, yet there are only 2 essentially different kinds of events in the class.

Many criteria can be devised in attempts to recognize classes that are constructed ad-hoch. Usually it is desired that all of the members of the class be as similar as possible — yet in general the question of whether two events are similar or not, depends upon the use to which the information is to be put. Thus a spring and a capacitor may be similar from the point of view of ~~mathematics~~ circuit analysis but they usually differ significantly in physical and chemical structure.

Another criteria of un-ad-hochness is that the class be constructed in a "simple" way. The meaning of simplicity is rather ambiguous however. In general a description that is simple in terms of one set of words will be extremely complex in terms of another language.

A The general approach has been to analyse methods of making  statistical inferences.

Mention how "grouping" or "categorization" arizes in Selfridge's work — also in chess.

Many members in class, yet all as ~ as poss. to event to be predicted.

~~The for the most part, discussion~~

In the past, discussions of this problem have

B been on an abstract level, often by philosophers and philosophically minded statisticians. For the most part, the discussions hadn't been sufficiently concrete so that they were not directly verifyable, For Living beings, however, the problem of statistical inference is a ~~ ~~ very concrete one, and their lives depend upon the efficacy of their attempt to solve this problem.

use >2 sentences

My one attempts to solve this problem in terms sufficiently specific for a machine to understand, have been guided by observations on problem solving techniques used by myself and others. ~~and~~ ~~~~ I have attempted to check any conclusions by applying them to simple inductive problems.

As a first trial for a definition of the probability of an event about to occure, one usually catagorized the event in a set of similar events that have occored in the past. Then the probability of occurance is approximated by the ratio of the number of times events in the class did occure, to the number of times they "could have" occored.

This definition poses many problems. In order for the ratio to have much accuracy, we would like to have many events in the class of similar events. We also would like the class members to be as similar as possible. These two conditions are mutually incompatible. One ~~ ~~ suggests classes with many members, the other usually suggests classes with very few.

The problem of how to decide if events are "similar" is also difficult to solve. The classification of ~~ ~~ two events as "similar" ~~ ~~ can only be made if the manner in which this classification is to be used, is known. In statistical inference, there is often no obvious criterion of "similarity" of events.

Another problem is posed by the construction of ad-hock classes. For any ~~arbitrary~~ event to be predicted ~~and probability values~~, a class of events can be constructed, so that the event in question will have any arbitrary probability value. Altho ad-hock classes are usually recognizable by human beings, it is difficult to specify conditions by which a machine might recognize them,

Often their are several methods of classifying an event, and each yeilds a different probability value. Just how is one to chose from, or combine these values?

The approach that I have been using to solve these problems is by no means ~~meant~~ expected to be an optimum one. ~~It is meant only to arrive at a solution~~ good enuf ~~[illegible]~~ to ~~[illegible]~~ be able ~~to solve the simple inferential problems that arise in any thought~~ It is felt that the ~~ordinary~~ solution ordinarily used unconciously by human beings, is not an optimum solution, and, indeed, ~~that~~ it is questionable if such an ~~optimum~~ is definable in any useful way. ~~It is meant only to be a rough approximation good enuf to give interes~~ I feel than even a very approximate solution will be able to give rise to extremely clever machine behavior. I also feel that for the world in which modern man lives, ~~[illegible]~~ the solution to ~~this p~~ the inferance problem upon which he bases his behavior is not ~~a particularly~~ "good" solution.

The main problems I have focused my attention, are ~~[illegible]~~

1) The construction of classes of events and the assignment of parameters to them, to facilitate their use in prediction.

2) ~~[illegible]~~ The problem of probability evaluation of an event that is a member of several different classes, each suggesting a different probability.

3) Devising a criteque for determing just how satisfactory a run of predictions are. This criteque is to be used in ~~deciding whether~~ optimizing various prediction parameters.

Of these problems, I now have a tentative solution to 3). This particular problem has a solution that depends largely upon the application to which the predictions are to be put. However, ~~I mean that~~ probably the critique I have devised ~~may~~ will be "good enuf" for most applications.

For problem 2) I have a tentative result which I am now examining. It involves a rather unwealdy mathematical expression, and I am trying to determine some of the properties of the solution.

Problem 1) is most interesting and most difficult. I have devised several methods for obtaining new useful classes, from old classes which have been known to be useful. Some of these methods is a generalization ~~of~~ ~~analogy~~ ~~inversion~~ what might be regarded as "analogy". Others involve breakdown of classes ~~rather~~ of events into classes of sub entities, which are then transformed into new, predictionally useful, classes.

In all of these problems, I have tried to test the proposed solutions for reasonableness, by applying them to simple inferential problems occuring in science, and everyday life. Also, I have devised a sequence of mathematical examples, from which it is expected that a reasonable good machine should be able to infer what is desired in such arithmetic operations as addition, subtraction, multiplication, etc.

T.M. General Description of ~~[illegible]~~ Approach, and The Present state of The Work.

The particular problem I am most interested in, is that of constructing a machine that will be able to ~~accomplish tasks~~ solve problems that are ordinarily regarded as requiring creative skills. While most of my efforts have been ~~in the direction of duplicating~~ suggested by the methods by which man solves problems, I do not want to limit myself to this particular approach.

?

At the present time, it appears useful to study the very basic problem of statistical inference.

In solving any problem, it is necessary to draw inferences from similar problems that have been successfully solved in the past ~~[illegible]~~ as well as from past experiences ~~[illegible]~~ that are less directly relevant. ~~The problem of how much statistical weight to give various~~ [illegible]