Feb 11, 74, Politics (not CMI, hvr). TM61½

Discn. w. Pat. Winston re: my staying at MIT past March 31, 74!
He discussed it w. Marv.
My impressn.: W. doesn't feel that Ind. Inf. research is terribly relevant to A.I. — so, as long as I'm working on that, he'd
05 → rather not have me here. If t. nature of t. project "changed" he would perhaps, reconsider.

Hvr. Marvin likes to support AI on a broader front — so he'd like me to stay, even if I work on induction theory.

I think M. & W tentatively decided that M. & I mite go out and hunt for funds for me alone, pretty much. As W. puts it

[~~W. doesn't like to get funding from much > 1 source, because of additional work involvd.~~]

"As long as they have room for me on t. 3rd floor — it's O.k. w. him" — T. implication being that as far as W. is concernd, he has no great desire to keep me here, but as long as I don't do any harm, e.g. sop up any funds or room, it's o.k. w. him. At any rate — W. seemed ~~[illegible]~~ hostile!

So, anyway — discuss this w. Marv. — os as well as how to go about shopping for funds. Mention Licklider as good person to talk to as far as suggesting Gvt. agencys. RLE [Brooks]

Discn. w. Marv.: Apparently there is a big money shortage now. Also much disagreement on ratio of theoretical to practical A.I. work — w. Marvin & Seymor for more theory & Pat for more practical.

Anyway M. suggests I write proposal. He will discuss with Seymore, sources of funds. Size of project to be presented (i.e. no. of people on it) will depend on how much money agency wants to spend. Preliminary proposal draft: What I have done in induction, what needs to be done, what I intend to do, and why it is important.

67.01 spec.
62.01

21274 TM 67

Proposal:

.01.66.90 (61½.90) spec

Training Sequences for Mechanized Learning. (Perhaps better title)

Project for research in Mechanized ~~[illegible]~~ Learning.

Introduction: The background of the present ~~For many years, the pr~~ project, is the theoretic analysis of the induction process — the process by wich a person ~~with perception~~ is able to discover regularities in the world about him, and use ~~[illegible]~~ these discoveries ~~[illegible]~~ in the successful prediction of new phenomena he has not yet seen.

~~In [illegible]~~

The project itself will consist of the construction of ~~training sequences from a variety of fields,~~ ~~and~~ sequences of problems of increasing difficulty, from various fields, and the ~~constr~~ parallel construction of machines that are able to solve these problems ~~in the~~ as they are presented.

The present theory of inductive inference makes it possible to devise such problem sequences. It tells us how much of a "heuristic jump" is needed to solve ~~contained in~~ a new problem, if we understand just how the machine has solved problems up to that point. The size of the "heuristic jump" tells us just how much

68.01

hard ware and/or ~~how much~~ time, the machine is expected to need to find a solution to the new problem.

At first the problems will be chosen from ~~[illegible]~~ ~~very~~ well understood areas in mathematics and formal linguistics. As the total training sequence grows larger and the machine has incorporated many heuristic ~~[illegible]~~ devices into ~~[illegible]~~ ~~[illegible]~~ its discovery mechanisms, it will be possible to give the machine problems that are not so completely understood, from a greater variety of fields.

Why this work is important (depends on agency)

possib applications (Military, non-military, depending on agency)

Applications, importance: (1) Devising trng seqs for ~~[illegible]~~ People is aided by analysis of machines' learning of seqs.
(2) Steps toward very hy power machines that could learn to understand English text & other data that has not been very carefully designd for it. (Perhaps go into this at some length — how, ordinarily, it is very dificult to ~~to~~ teach machines things — one must be very careful in devising trng seqs.

22274 Pro

TM73

.01;68.40 : (1) Previous work in induction theory : what it can do

— It can tell us (1) if a discovery procedure is adequate. (2) It suggests ways to devise discovery procedures.

(2) ~~Most work on getting ma [illegible]~~ Much of the V.G. work ~~in [illegible]~~ on induction in AI, has been based upon introspection & protocol analysis. Pgms written were attempts to work the way humans do.

Induction theory gives us a uniform set of constructs for expressing such induction systems — for criticizing them, for improving them. It makes it possible to select features from several different induction programs and combine them to make a program more powerful than any of the original programs.

(3) The goal of the research is not simply to find good ways to do mechanized induction. — Rather it is to find such methods that are themselves discoverable by a machine that has had an acceptable training sequence.

We want a machine that learns how to do induction from a training sequence that is ~~relatively~~ not specificly oriented toward teaching specific induction techniques. It is only in this way that we can obtain a machine that will learn by its experience with problems that have not be selected with the goal of training the machine in any way. It is such a machine that would be most able to deal with the greatest variety of problem appearing in the real world.

71.01

22274 Pro

TM73

01;68.40 : (1) Previous work in induction theory: what it can do — It can tell us (1) If a discovery procedure is adequate. (2) It suggests ways to devise discovery procedures.

(2) Much of the V.G. work on induction in A I, has been based upon introspection & protocol analysis. Pgms written were attempts to work t. way humans do.

Induction theory gives us a uniform set of constructs for expressing such induction systems — for criticizing them, for improving them. It makes it possible to select features from several different induction programs and combine them to make a program more powerful than any of the original programs.

(3) The goal of the research is <u>not</u> simply to find good ways to do mechanized induction. — Rather it is to find such methods that are themselves <u>discoverable</u> by a machine that has had an acceptable training sequence. We want a machine that <u>learns</u> how to do induction from a training sequence that is not specificly oriented toward teaching specific induction techniques. It is only in this way that we can obtain a machine that will learn by its experience with problems that have <u>not</u> be selected with the goal of training the machine in any way. It is such a machine that would be most able, to deal with the greatest variety of problem appearing in the real world.

12274

Pro

.01! 73.40 : (9) (a) The Nature of t. (Ind. Int.) Problem : ~~Under~~ Induction, learning. How it is done by ~~Animals~~ living Creatures — How to do it mechanically.

(b) Why it is an impt. problem! — Understanding human learning for better teaching — Understanding t. / creative discovery process / in humans so we can devise envts., trg., to encourage it — Devising ~~learning~~ machines for / learning under conds in which it is not feasible for humans to ~~[illegible]~~ / employ : e.g. ~~[illegible]~~

NSF? Ford Foundn

NASA extraterrestrial envts for Nasa (hy accel., hy radiation, hy pressure

NIH "hy tempt. / low temperature, etc) : Medical Learning / MD over very large population over very long period of time. Well beyond capability of any human. Present statistical techniques employ simple models for induction — are very limited in their capability. We want to get much more clever induction for very large amts of data. A human Md. dies with his experience, for t. most part not passed on — or taught. We want to retain t. learning of ~~all~~ a potentially infinite lifetimes.

For making intelligent decisions more rapidly than Humans can

ARPA DOD Army, Air Force, (Nuclear Attack / Attack in a way that was not anticipated, not too brite — Explain how / machine could be used / How Machine would have to be tested under various simulated conds. to get confidence in it. Compensates for lack of brilliance, by extreme speed),

Learning to recognize signatures of Missiles of various types.

Learning to recognize Nuclear test v.s. earthquake.

Integration of Data from many modalities (IR, Radar, Sonar, intelligence reports) very rapidly, to infer position of Nuc. submarines, missiles, missile bases,

Navy Army Slackout Battlefield

NIH A very fast machine could ~~[illegible]~~ reach thru the skin of a patient and remove an appendix in less than a second. — On the way in, / ~~noticing~~ making exploratory observations of various conditions that may or may not make an appendectomy the best thing to do, perhaps deciding not to do it, after 2 or 3 milliseconds penetration.

75.01

.01 : 74.90 : NASA ; Air Force : ([redacted] } Very rapid emergencies occurring in the piloting of aircraft or spacecraft.

Human is too slow to respond, Situation is to complex for ~~rbalt~~ routine preprogramming for ~~[illegible]~~ forseeable emergencies.

- Industrial production
- Detecting Epidemics at early level
- Infor Retrieval
- General Assistant & Getting from Moon.

NIH : / As part of A R central medical intelligence system, an induction machine could ~~dett~~ discover an incipient epidemic by ~~the~~ noticing various apparently ~~[illegible]~~ isolated symptoms in a great number of people over a large population. The ~~interrelation~~ of such symptoms with weather conditions, ~~ambient~~ environment level of various environmental pollutants, ~~[illegible]~~ disease conds in other countries would not be detected by ~~normally use~~ ordinarily available statistical methods. The induction techniques ~~[illegible]~~ to be used will obtain the best possible predictions from the available information, using the ~~[illegible]~~ computing resources available. Suspicion of an epidemic could be confirmed by ~~[illegible]~~ further medical tests.
Such a central medical system could ~~to~~ then recommend various measures to ~~[illegible]~~ cut off the epidemic well befor it ever got started.

~~Industrial pr~~ Use of robots for industrial production would be much facilitated if they could be taught a very complex task in the way a human is taught — by watching someone else perform the task. At the present time, robots are very limited in what they can do. Tasks involving complex interaction of effectors and sensors are usually well beyond their capabilitys. In cases where such interactive behavior is possible, ~~the~~ each ~~specialized~~ ~~tasks~~ job must be specifically programmed — a very difficult and time consuming task. An inductive machine could learn such tasks very quickly after observing a human perform the task a few times.

224 14 Pro

01, 77.40
75.40 spec.

TM78

The Generalized informational assistant (Marvin's idea)

This is an info retrieval device, in which one descbs one's problem (or what one *thinks* one's problem is) & t. machine asks Q's of one & makes suggestions of idea areas or of things to read or of relevant discoverys, activities of people, inventions, conferences, places to go, kinds of expts. to try (both gedanken & actual)

This assistant would also help as a "memory aid." One would descb to it t. thing one wanted to think of, along w. various probable constraints & it would ask Q's then try to guess what was wanted.

Council on Library Resources (Ford Found)

To Mar 5, 74
630-730P
Advisor?
195 Albany St
Camb

Demo of 52'x64" proj color TV.

T. Sugg are *all* general A.I. goals. Ind inf. research is one area of AI in which any success can usually be translated into useful progress in most any other area of A.I.

POD
ARPA

Learning to integrate various environmental & political & economic symptoms to decide whether an enemy country was preparing for an attack on some other country.

General emphasis of abilitys of hy level learning machines to deal w. *unforseen* events, for which they were not specifically pgmd, & give list of situations in which this would be vital.

For Motivation, art: New Scientist. 10 Mar 66: "New Learning Machines for Future Aerospace Systems. (D.R. Moore, A.C. Spanke, — USAF Avionics lab. W. Patt [redacted] AFB, Ohio. pp 626-628

→ ① Present state of Inf. research; How long will it be befor we get progress? Just why we are *more* likely to make breakthru rite now.

② On Gorry's work: How t. big problem was to obtain *probys* to be usd in decision theoretic Model. How *now* Induction Theory is in some position to give *real* Help with this problem — perhaps solved completely.

22774 — Pro

TM79

(Govry wrote 3 proposals to NIH — all rejected — perhaps because of Nixon's impounding of funds. — Now funds have been restored.)

1) ⓐ Perhaps work w. Govry to get Money from NIH

→ ⓑ " Phone Peggy to ask her who to talk to.

— Tho she may not know specifically, she can give info on the hierarchy.

2) Discuss [redacted] my early work in induction (~1956 IRE)

[redacted] What its problems were (Utility evaln. — no basic understanding of induction Theory).

3) Discuss ~~more~~ more modern methods of dealing w. This problem — its successful applicn. to some aspects of grammar induction, where other methods have faild. Its "adequacy" in t. ~~MS~~ Σ sq error sense.

4) The proposed 3 pronged attack on ~~~~ induction: — a) Basic Tng. seq. starting w. very simple concepts. use of [redacted] arith, alg, math, & formal linguistics in tng. seq.

— Where material is very well understood.

b) ~~Use~~ Tng. Seq. ~~of~~ of an ongoing complex problem. Attempt to find / inadequate set of "factors" for the concepts used. That we have a very clear defn. of adequacy in this context.

Use of Med diagnosis, induction of concepts from "block world", [redacted] from Chess, from Electronic ckt analysis, from simple computer Pgm debugging.

---

c) Continued work on basic problems in induction Theory that are most relevant to applications. e.g. ① Better methods to combine pgms. ② Optimum (least cost) solu. of theoretical problem of finding best changes in "Prod'n body" in response to a gn. "Δ corpus". ③ Work on IPG Theory.

→ (80.01)

5m × 360 = 1.8×10³

Bonny (Ihordu 530 p)

flux = Mxwl.

$\mu^{-\frac{1}{2}} m^{\frac{1}{2}} l^{\frac{3}{2}} t^{-1}$

hν = en.

$h t^{-1} = m l^2 t^{-2}$

$h = m l^2 t^{-1}$

e =

~~~~ = H flux

$Q \cdot L^{-1} T^{-1}$

$\frac{m l^2 t}{Q}$

$Q = \mu^{-\frac{1}{2}} m^{\frac{1}{2}} L^{\frac{1}{2}}$

$\frac{h}{e} = \frac{m l^2 t^3}{\mu^{-\frac{1}{2}} m^{\frac{1}{2}} l^{\frac{1}{2}}}$

$= \mu^{\frac{1}{2}} m^{\frac{1}{2}} l^{\frac{3}{2}} t^3$

weber flux $m l^2 t^{-1} Q^{-1}$

el charge = Q

flux × e = $m L^2 t^{-1}$

$= \frac{h}{t\,t} = $ ~~~~
~~~~

3374 Solomonoff Mar 6, 74 Pro

Proposal for the Application of Recent Progress in the theory of Inductive Inference to Practical Problems in Machine Learning, Prediction and Training Sequence Generation.

The last decade has witnessed important progress in understanding the theoretical mechanisms of inductive inference. Research by Chaitin, Kolmogorov, Solomonoff and Willis, has clarified much in an area that had up to this time been dominated by philosophical speculation, intuition and inspired guesswork.

Although these new results have been applied to some practical problems - such as linear regression and curve fitting, the goal of these studies has been not to obtain new techniques, but to show that the new formulations of induction yield correct solutions for problems that are already well understood.

The proposed applications will be of several kinds. In previous artificial intelligence work progeams have been written to solve problems in mathematics, linguistics, and electronics, and to play games such as chess and checkers. These programs have been very specific, in the sense of being devised for very narrow task environments, and being of little value for slightly different kinds of tasks.

The proposed work in induction will be toward machines that are able to work over a much broader spectrum of problem environments - machines which learn to work problems by being shown an adequate set of mostly correct examples.

Much of the previous work of this sort has met with very limited success. In some cases - i.e. perceptron-like machines - Minsky and Papert have made the nature of the limitations quite clear - the machines are linguistically incapable of expressing the needed concepts.

Induction theory makes it very easy to construct machines that do not have this linguistic limitation.

Another difficulty with previous induction machines (e.g. those of Uttley and Rosenblatt) is that the number of examples needed to learn a particular concept was inordinately large. In some cases, the machine would have to be shown all cases of the concept in order to be able to learn it - so that for such problems the machine performs no inductive inference at all.

Again induction theory permits us to deal with this difficulty in a most general way. It tells us the minimum number of examples that are needed to discover a particular concept (Schubert 1973, Willis 1970) as well as what sort of training environment is needed, and suggest algorithms for most rapid learning.

These proposed investigations involve programming a machine so that it can learn to work problems in a particular field.

A somewhat more general approach that will be persued would train a maahine just as a child is trained - by starting out with a very simple set of problems and by a suitable sequence of training examples, bring the machine to any needed level of competence in a particular field. The initial machine need have fewer induction techniques built into it than in the previously described machines. Again, induction theory tells us which sets of built-in techniques are adequate for this kind of learning. It is felt that this sort of machine would probably be able to work problems better than those previously described, and from a much broader field of problem environments. Its training sequences would, however, be much longer and as in the case of a human being most of the training would not be in fields directly related to the problems the machine would ultimately be solving.

A third area of research will be continued work on fundamental problems in induction theory that are most relevant to applications. One of these problems involves optimum methods of combining various successful induction techniques. Antoher concerns optimum methods of computation to obtain the best possible results in the least amount of time, using any specific hardware configuration, or ways of finding optimum hardware to solve the particular kinds of problems that occur in induction. These hardware-software optimization problems occur

3374 p10

in all areas of practical computation, but in induction theory, it appears to be easier to state the problem in a general, unambiguous way that suggests promising approaches.

3574 Pro

TM84

01.83.90: My present idea of this proposal:

1) Abstract (short summary)

2) Introdn (longer summary — also tells what t. various sections are about

3) Main body: various sections that were summarized in 2)

4) Appendices: Papers & Reports & "working papers" relevant to t. proposal. [Eric willis: Tug says (1962); IE³ review: winston: perhaps my working paper on ind Inf results.

The main body desbs t. problem, tells a little why its impt. & gives some applicns. (The main thrust of why its impt. & applicns, will be in an auxilliary letter — geared to each agency.)

Next, what previous work has been done by others on this problem, along with criticisms of it.

Next, Give summary of main deficiencies of previous approaches. ~~Mainly~~

Give outline of progress in induction theory of last ~10 yrs. (Evans: winston) Give t. main points that are relevant to overcoming the diftys of the "previous approaches".

(In desbg the work of others, do not mention deficiencys in it unless they are extremely clearly there. Also, when mentioning deficiencies of an approach, mention the essential progress in A.I. that was obtaind via that approach.

These ~~errors~~ remarks are made w.t. understanding that the people whose work I refer to will probly be reviewing this proposal!)

32174 Pro

.01; 84.40 : Abstract: (use title of 80.01) See 107.01–.21 for revision TM 105

The proposed investigations will ~~continue the previous work on induction and generalize it~~ apply the recent results on induction theory to the most successful induction approaches to programs available at the present time. It should be possible for programs to arrive at their present proficiency by a suitable training sequence given to an initially less ~~able induc~~ able induction program. By continued training, it would go on to greater proficiency in more general problem universes.

A parallel investigation will devise a training sequence starting from the most elementary problems and continuing on up to the most difficult ones. A machine will be devised that is able to follow along in learning to work this training sequence.

Along with these two [illegible] program developments, the theoretical analysis of induction will continue, with emphasis on its application to practical problems.

.23

The "Introduction" starts with 80.01 – 83.40. – perhaps use that as a long "abstract".

.29 A "true" introduction! More like an extended table of contents.

→ Description of just what the inductive inference problem is — give a formalization & some examples of problems that can be ~~redo~~ expressed in that form. Mention a few areas in which a machine that could do this would be of much value (e.g. weather prediction in essentially different way from usual methods), — Med diagnoses – decision theory.)

106.01

32174 Pro

.01: 105.40: Previous work in this field: Uttley, Rosenblatt, Kochan, Solomonoff, et al, KGS, IE3 (see my 1964 paper for list), Evans, Winston. TM106

Discuss difficulties, inadequacies of this work. (perhaps refer to ..., but mention Winston in more detail, since it is post 1964.

[Note lack of work on probability — explain why its impt — orig. in decision theory — to operate or not — shall we treat it as an enemy attack? Probability of snow.

How recent induction research will overcome these diffys & inadequacies. Mention specific inadequacies that are common to several of past attempts. Show very specifically — say in 1 or 2 particular cases, how induction research will solve these problems.

A more detailed description of the approaches that will be pursued: ① give example of application to Winston's work ② Give example of simple training in Arith, alg, calc, analysis, — that all of "pgmg" will be done by giving examples — & examples will be gn. w/ less & less care as we continue. That the tng. sequence will be very similar to that given to a child.

Discuss theoretical problems that need be solved — give specific problems & expected solns a/o approaches

Discuss Hardware! Methods of use of hair memy, of many ALU's in ll of cache memys, of theorem on mean flip time vs. min possl. flip time. Perhaps insert a little math on hair memy utilizn. — explain why its impt. — that TM feasibility depends on hardware, that speed x2 = 2 yrs.

107.01

322 74 Pro

short Abstract: (use title of 80.01)
from 105.01–.23

…01; 106.40: ↓↓ The proposed investigations will utilize [redacted] recent results in inductive inference theory to overcome the limitations of the best inductive inference programs that [redacted] are presently available. These [redacted] programs will be revised so that by presenting to them suitable training sequences of usually correct examples, their internal structures will be modified so that they will be able to deal with more dificult problems or problems in different fields.

TM107

A parallel investigation will devise a machine and an associated training sequence that will start from very elementary problems and work up to more difficult ones. Such a machine is expected to be more varsitile than the machine previously described, but it will require a much longer training sequence.

Along with these program developements, the theoretical analysis of induction will continue, with emphasis on its application to practical problems.

…z1

See 80.01 — 83.40 for a longer ≈ introduction ≈ abstract.

<u>Extended introduction</u> & Giving contents & summary of f. following sections. (See 105.29 – 106.40 for prelimy. approach to introdn)

tac

32474 Pro

107.40 ; Introdn: (1) The nature of the induction problem and The value of a good mechanized soln. — This will be a brief statement of what each section is about. TM108

Weather prediction, medical diagnosis, discovery of patterns in data, extrapolation of economic trends, and all of human learning behavior are examples of inductive inference.

All of these problems can be usefully expressed in the form of extrapolation of a sequence of symbols.

Mention that all learning processes are examples of induction

Ordinarily induction is performed by humans, sometimes aided by computers.

There are several reasons why we want to mechanize this procedure. A major goal is the purely scientific one of understanding the induction process in humans. Such understanding would be invaluable in teaching or otherwise enhancing the learning processes in humans. Our models of human learning of complex tasks are extremely rudimentary. A working induction machine would be a very good preliminary model for human learning.

In weather prediction, we have only recently begun to get completely mechanized predictions that are better than human predictions. Humans seem to make these predictions significantly differently from the ways now used by machines, and at present, machine weather prediction is severely limited by the size and speed of available computers. One aspect of the proposed induction research is in the direction of combining the methods used by humans with those presently used by machines. Another aspect of the proposed work is the optimization of induction of hardware efficiency — How can we best design hardware or utilize existing hardware, so as to get the best possible prediction for whatever fixed amount we spent on computing?

In the case of medical diagnosis, relatively little has been done by machine. perhaps Mention Gorry & others in his Biblog. 109.01

3-24-70 (TM 109)

01 108.90 ... a computer program, once written, need never be rewritten — only maintained and updated with new data. Very good human diagnosticians are available, but they are limited in number and are very expensive. Effective computer diagnosis could reduce this cost considerably. While the costs of computation may, at present, be prohibitive, this will not continue to be so, since the costs of computation continue to halve about every 2 yrs. Again, however, the study of methods to obtain the best possible diagnosis for a given computation cost is of much importance.

In most cases the use of humans for induction problems is adequate, but where the work occurs in hostile environments such as space, the ocean floor, or involves very rapid decisions — such as in aeronautics, astronautics or in decisions involving nuclear attack, mechanization can be invaluable.

.15

O.K. ! instead of a long introduction telling what the various sections are about, have a contents, w. an abstract of each section. This will make the thing a lot easier to read & understand. — So just write the various sections now, & write the abstracts later.

The section "on what is induction & why do we need to mechanize it" 106.01ff Could be much stronger. One should finish reading it w. the idea that This is one of the most impt. things That has to be done in the world — Now!

Anyway! Next! What has been done by others! Give brief mention of most impt. work, but refer to my IE3 paper! Then mention a few studies that have been relevant to the proposed work!

my SOS '62; KGS; Evans, Winston, (Sussman?); (Fogel?)

Mention specific areas where work could be continued! lack of probty calcns. in most, & why probty is important for decisions (Nuc. attack, Med decisions)

Deficiencies of these works! (1) Took too much cost (KGS) (2) didn't give probys (3) didn't have adequate lang. & had no way to extend it. (4) ... Had fixd set of concepts, methods to get new ones were absent or inadequate. (5) Device did not have 112.01

01:109.~~10~~ facilitys to review its past ~~behavior, to decide~~ successes & failures & decide *why* it had succeeded or failed & modify future behavior accordingly. TM112

---

so: A repeat of 105.29ff as a contents:

1) What is the problem of Inductive Inference and what is the value of a mechanized solution? 105.29; 108.01–109.15

2) Previous work on Mechanized induction. ~~and the~~ Unsolved ~~The~~ ~~inadequacies of this work — what needs to be done.~~ ~~Unsolved problems~~ Problems that remain to be solved.

3) ~~A~~ The approach to mechanized induction suggested by recent theoretical research. ~~How the inadequacies of previous work can be dealt with.~~

perhaps July 1 section 116

4) A description of the propsed work. ~~How various~~ How it will deal with the difficulties encountered in past induction research → 118.10
~~inadequacies of difficulties occurring in previous work will be dealt with~~ ~~How it will deal with the problems of the past previous~~ difficulties encountered in past induction research.

5) ~~Theoretical Problems that need to be~~

Problems in the optimization of hardware for inductive inference. The concept of "information processing capacity" of an element or complex of elements.
~~Some methods of highly efficient usage of m[illegible] Optimum Methods to utilize memories of different speeds. Optimum Use of many small quick central processing units with sharing of most of main memory. How these structures differ from other large parallel processors (e.g. Illiac IV).~~

~~6) Some important unsolved problems in induction theory.~~
Optimum utilization of a set of memory banks, of differing access times. Optimum utilization of a set of fast cpu's with large common memory banks.

6) Some important unsolved problems and unexplored areas in induction theory. 113.13 IPC 114.01

01:112.40: 2) Previous work on Mechanized Induction. Problems that remain to be solved.

TM114

At this point, I should familiarize myself w. T. Evans & P. Winston's stuff:

Re: my summary of T. Evans in my 66IE³ paper: P1690: 124

I think his main problem was to devise a lang. to descr. each object. These descrs. should be s.t. it is easy to devise good "relations", so A : B = C : X can be solved.

So t. problem is devising a lang., gn. t. set of pictures.

10:112.18 4) The general approach: Start with a model of human induction either obtained from introspection, or from protocol analysis. Such models have been studied extensively (Newell, Simon, T. Evans to start, Winston, e.g.). Define each of the concepts used and assign it a symbol. Obtain a "bit cost" for each symbol. This is, loosely speaking, −log₂ of the frequency of use of that concept in the past. It will also depend on the constraints that are necessary in using that concept.

Using these symbols and their costs, and variously computed punctuation costs, write out the prediction algorithm of the model. Ordinarily, it will be possible to do this in more than one way.

From these algorithm descriptions we can obtain 2 new kinds of information.

.37 First, we are able to get probability values, where, in the original algorithm, we were only able to get single "best" predictions.

115.01

01 : 114.40 : Second; After using the algorithm for prediction, the bit costs of the concepts used are changed somewhat. The new bitcost values can be used to ~~[illegible]~~ tell how to combine the old concepts to obtain new/trial concepts. The new trial concepts will have bitcosts ~~[illegible]~~ roughly equal to the sum of the bitcosts of their component concepts. For trail runs of new predictions, we should use new concepts of minimum bit cost. — These will correspond to concepts that are formed of only a few concepts that have been very useful in the past.

While the forgoing method of obtaining new concepts would be fine if we had an arbitrarily large/amount of computing power available, it is not particularly good in practical cases of moderate complexity. The reason is, we have not taken the computation cost of the old and new concepts into account.

~~To do this, we must~~

This can be done ~~a~~ any of several ways: or(? name 3!). A fairly general way is to have a second machine observe the behavior of the first. It observes the concepts used by the first machine, and it notes the computation cost associated with various combinations of these concepts. From these observations it makes a model from which it is able to estimate the expected computation cost of any new concepts that are obtained by combining ~~[illegible]~~ old ones. 116.01.

32774 Pro

01: 115.40 : ~~These~~ / computation cost estimates need not be very good for ~~as to be able~~ the first machine to be able to ~~grow~~ discover fairly ~~good~~ useful concepts ~~for the first machine,~~ but it will take longer to find those concepts if the second machines ~~estimates are poor.~~ TM116

As the ~~[word]~~ body of data used by the machines increases, the second machine will be able to make better computation cost estimates on the basis of larger sample size — even if its models are poor.

In a [redacted] machine with much experience in ~~[word]~~ probability estimates of this kind, it should be possible to use the first machine to devise [redacted] improved models for the second machine's computation cost estimates.

~~If~~ when ~~costs~~ expected computation cost estimates are available, they are used, in combination with low ~~UBK~~ total bit costs to select out new trial concepts that are ~~[word]~~ most likely to be useful, and which will require a reasonable amount of computation to test.

Re: 114.37 ! We must have several / BW prodn. methods in || to get proby values. We always have t. standard Bern. seq. method, but usually we can do far better than that — be devising Bern seqns. using various constraints — or / variously derived subsequences of t. corpus.

We are interested in going beyond previous induction programs in, 2 ways (1) Obtaining good values of proby. This is ~~[word]~~ well beyond the use of event counting to obtain frequencies. It involves construction of the most general / possible stochastic models possl. 117.01

32874 Pro Discovery of ~~important~~ new concepts to be used in induction: ~~Successful dis~~

01:116.40: 2) ~~Devising new concepts to be used in induction:~~ Previous induction systems have been very limited in this respect. ~~Most~~ ~~of them were able to [illegible]~~ Some of the older ones used little more than the optimization of a fixed set of parameters ~~to obtain better induction~~. ~~Some~~ More complex systems were able to ~~devise~~ ~~[illegible]~~ ~~new concepts that were~~ go beyond parameter adjustment, but were still severly limited in the range of ~~[illegible]~~ environmental regularities that they could ~~were capable of~~ discover. TM117

There have been a few systems that were not limited in this way, that could, in theory, discover any regularity in the environment that ~~was describable in a fin~~ had a finite description. These systems have all, however, been heuristically inadequate, so they had no way to achieve these discoveries ~~discover complex new concepts~~ in any reasonable length of time. ~~[illegible]~~ Discovering important ~~interesting~~ new concepts was theoretically possible, but not practically achievable. → 119.35–.37

~~The set of all prediction algorithms can be listed, and, conceptually, in some thought at least, they can each be tested on the applied to the data of the past. As function of both the past performance and information (complexity part, as an index), each of these algorithms can be weighted suitably, and the [illegible] set of them can be used to make predictions as predictions that are as good as possible with the given data.~~

138: 119.37 In practice this procedure is not entirely possible. Many of the prediction algorithms take an enormous amount of time to evaluate. Many of them never ~~actually~~ converge at all

32974 Pro.

TM118

01:117.90 and it is impossible to discover / this in a finite amount of time.

Instead of testing all possible algorithms, it is necessary to restrict the set tested to those that are likely to converge in a reasonable time.

If an estimate is obtainable for the convergence time of any prediction algorithm, an optimum order in which the algorithms should be tested / can be computed — optimum in the sense of best expected prediction accuracy for a given amount of time spent testing algorithms. At any particular point the testing program, can stop and the best algorithms found up to that time can used for prediction.

This is, in essence, the induction system that we propose.

(1) Direct param derv.
(2) FoK " "
(3) Mt. Carlo param Derv.
→ 120.01

.20 The method of generation of prediction algorithms is fairly well understood. It is also possible to list them in a priori likelyhood of effectiveness. The part of the system that needs most work is the calculation of expected testing times for the algorithms. For many large classes of algorithms (Give a list of such), it is fairly easy to obtain such estimates, but for the truly original algorithms that will be generated, there is no general estimation method. Various approximation methods will have to be devised and analysed.

It should be noted that

119.01

01:118.40 : ~~estimation method is not very good, the system as a whole~~ TM 119

the effectiveness of the whole induction system is not critically dependent upon the accuracy of the subsystem that estimates algorithm ~~[illegible]~~ testing times. [335P] → 335P

Poor estimates will mean that the machine wastes ~~[illegible]~~ time testing algorithms that are ~~[illegible]~~ impractical/ or impossible. ~~Some of which do not at all converge. This will slow down the discovery of algorithms that are essential/ new, but will certainly permit the discover~~

Because of this inefficiency such a system would be able to discover essentially new algorithms, but at a much slower rate. ~~than a system with a better testing time estimator.~~

Eventually, if the entire system has worked up to ~~[illegible]~~ a capability of dealing with fairly complex induction problems, it will be able to work on the problem of improving its own algorithm testing time estimator. ~~This being a normal inductive inference problem~~

At first, the system will use a fixed program for estimating testing times of algorithms. This program will be improved from time to time by its (human) operator.

---

35 The set of prediction algorithms can be listed and, conceptually at least, each can be tested on the known data. [illegible] Each of these algorithms can be weighted as a function of its informational complexity and the goodness of its performance on known data. This weighted sum will give predictions that are as good as possible with respect to the ~~[illegible]~~ available data. → 117.38

37

120.01

329 79 Pro

.01: 119.40: | SN : 118.20 spac

On 1. Method of generating new Pams.

TM120