QUANTITATIVE ARTIFICIAL INTELLIGENCE: HOW MUCH IS ENOUGH?

Philip C. Norem
214 Gilmore
Fort Erie, Ontario
Canada L2A 2M5

Abstract

It is shown that human beings have a modest capability for abstract thought ranging from 100 Kilobytes to 10 Megabytes. Accordingly, present artificial intelligence programs approximating 100 Kilobytes are roughly equal to the low end of human capabilities and should be capable of duplicating human reasoning abilities. Furthermore, it should be possible for a 1 Megabyte program to enable a computer to function in ordinary English with a substantial range of activities, making larger programs possible in English with the preface in a normal computer language.

All present systems for evaluating artificial intelligence programs suffer from an inability to compare their results with human processes for performing the same work. Accordingly, the estimated size for a human-like computer program has been best approached prior to the present by computing the number of cells and synapses in the human brain and then producing a rough ballpark estimate for the comparable computer program. The need for such clumsy approaches should be drastically reduced by the procedures outlined herewith. It will be shown that the brain capacity for abstract thought in the average person is the relatively modest value of 130 Kilobytes of RAM together with, of course, quite a bit of pattern recognition capability which is not measured in the same terms. However, by extrapolating through the measured memory of graduating students to those with doctorates in science or the equivalent expertise in other fields, the approximate figure is quite different: 3 Megabytes. Accordingly, it is very difficult for a small team of programmers to write programs of sufficient complexity to appear humanlike *if* the standard for that success is that the programs shall be as smart as they are and the method is the conventional one of writing the entire program in a conventional AI language such as LISP. However, if the standard is not that one but the variant of writing a program sufficiently smart to compete with a production line worker with a ninth grade education then the small team of programmers usually involved in the average project is quite sufficient. Indeed, there is a movement to make 130 Kilobyte programs medium sized ones especially in the artificial intelligence field. Accordingly, it may be that industrial robots with only modest minicomputers operating them may soon displace the large bulk of the average workforce. The problem in pattern recognition which seems the only barrier to this displacement may or may not prove a long term one. (The problem of pattern recognition among strings of data, series of pulses, has apparently been completely solved by "string comparators" which utilize high speed number crunching, and it may be that this procedure will prove adaptable to two-dimensional patterns.)[1].

Measuring Human Capabilities

The capacity of the human brain can be measured by the procedure of determining the maximum vocabulary for the test subject and then using the established relationship of Zipf to determine from the vocabulary the total number of words in the subject's memory bank. The procedure will be further detailed; for now let us take the example of a person who had a memory size of 80,000 words and a vocabulary of 8000 differently spelled words, the typical I.Q. 100 person, in which each word represents thirteen bits. In the corresponding case of an I.Q. 140 person with a doctorate or equivalent mental ability, the memory size would be a minimum of 1,500,000 words with a vocabulary of 150,000 differently spelled words. Each word would then represent eighteen bits.

The ratio of total vocabulary words to total mental storage capability is very close to ten as a consistent factor in English and other languages measured. Zipf [2,3] is responsible for this relationship, which derives from the application of information theory to the way languages are constructed and used. In particular he concludes that words in a sample of optimally written text occur with a frequency x/n, where n is the number of the word proceeding from the most frequently used word (n=1) to the least (n=8000) and x is a common multiplying factor. Thus, we have, for the total body of words in a given calculation:

$$\sum_{n=1}^{8000} x/n = 1.00 \text{ (solve for x)} \quad (1)$$

This calculation reflects the undoubted fact that the total of all words used in the text, whether unique or repeated, is one text the same size as itself. Since we are solving the equation for the case of a vocabulary of 8000 words we know that is the upper limit of the summation. When evaluated numerically on a computer the value of x is found to be x = .10455357

Having found x, if we wish to find now the total number of words in the text sample we set up the equation:

$$(x/x) \sum_{n=1}^{8000} 8000/n = \text{Number of words} \quad (2)$$

so that the least frequently appearing word will appear exactly once and the other words will appear as many times as their frequency dictates. Using the identity of Eq. (1) we have the new statement,

$$\frac{8000(1.00)}{x} = 8000/x = 8000/.10455 = 76515 \text{totalwords.}$$

[A correction so that each word appears an integral number of times reduces this figure to 73147].

Thus, the computation provides the total memory size under the assumption that English is a perfect language, which, in measurements reported by Zipf, it seems to approximate fairly closely. A more general measurement is possible, though with less accuracy, by determining the presence or absence of words which occur infrequently in the vocabulary and relating their presence to whether or not a chunk of memory large enough to include that word has been stored. The word "matchless" occurs roughly once per 140,000 words [4] in measured counts and thus, should not be found, on average, in the vocabulary of the average person. It is not [5].

For the I.Q. 140 person, the figure of 1,500,000 can only be used as a minimum because there are not many more than the 150,000 words of the measured vocabulary of college graduates [6]. The Unabridged Dictionary lists as entries about 450,000 but many are in fields which are not relevant to, for example, computer engineering and accordingly the simplifications which might otherwise be possible by using an ideal vocabulary simply cannot be achieved short of inventing this new vocabulary on the spot. Much progress has been made in that particular field by abbreviations such as HMOS, JFET, and EEPROM but the field still cannot be regarded as using the sort of finely tuned ideal language that basic English seems to be. Accordingly the fact that a person knows 150,000 words does not mean that he might not be able to learn and use even a million if the words were useful to him.

The question of memory is obviously one of the more crucial ones in determining human capabilities. Many cases are known of people with poor memories, enough so it will not be necessary to give an illustration; the converse case is more interesting. George Meany, former president of the AFL-CIO, claimed that he never forgot anything and his station in life would tend to give weight to his claim. If so, his memory bank would have contained the colossal total of 432 million words just counting his working hours at sixty words per minute of conversation. The statements of John Dean to the Watergate investigating committee verify that people, at least some people, are capable of dredging up substantially verbatim the entire texts of conversations years before and not considered important at the time. (Compare [7] to [8] for Mr. Dean.) Thus, it would not be extreme to feel that a ten million word total memory would reflect the approximate maximum of the human mind's capabilities, with a corresponding vocabulary of a million words.

The method of measuring the vocabulary of the average person, one with I.Q. of 100 deserves a special note because it was not done with the same procedures as Seashore and Eckerson [6] who determined what percentage of a random sample of the Unabridged Dictionary was known to the test subject and from there deduced his total vocabulary by multiplying by the ratio of size between the sample of the dictionary. The method used for average subjects was to measure the words missed by one-half of people taking an I.Q. test and then determine the number of words which are more frequently used in English and then assume that the person knows all of the more frequently used words and none of the less frequently used words. The Wechsler reference [5] gives the words in the test and their relative recognition factors, and Thorndike and Lorge [4] then give the number of words more frequently used than the chosen one. In view of the scarcity of published vocabulary measurements on normal adults (psychotics and the mentally impaired have been adequately studied) this figure of 8000 words will have to do.

As as aside, it will be important to clear up the otherwise puzzling question of how it happens that people seem to measure so differently in mental capabilities when Thomas Jefferson assured us that we were all born equal. The answer is twofold. The work of Zipf shows that a person with an 8000 word vocabulary has a fully complete basis on which to construct varied and grammatical sentences in an unending variety. Indeed, the structure of the language so produced is identical to the structures used by the person with ten times greater measured capabilities. As an example, the word "the" occurs at precisely the correct frequency in communication from an average person. In short, a person would not display his limitation in conversation or in writing, once he has passed the 8000 word vocabulary mark.

The second part of the answer is that, although people can determine things about vocabulary of others by asking these people direct questions about it, this is rather impolite; especially between strangers and only then does it count. It will be left to the reader as an exercise to determine the percentage of people who know the meaning of the word "reluctant."

Implications for Artificial Intelligence

It is clear from the conclusions of this paper that there is a very great chance for human-like computer programs in the next decade. It may well be that there are enough to them to make automation an overpowering fact of life for the average person in the workplace, resulting in drastic civil disruption. One hundred and thirty kilobyte programs are going to be a dime a dozen once written and put on ROMs. In fact, they fit neatly on single one-Megabit ROMs now nearing production and at a cost which will probably approach 89 cents or so in full scale production. Thus, the figure "dime a dozen" is indeed frighteningly accurate. Only the pattern recognition problem remains, and that only in two dimensions as of last Fall.

The conclusion for the leadership and scientific community of this paper seems less drastic. It seems unlikely that any major artificial intelligence effort will succeed in the production of more than a few ten-Megabyte programs if such efforts take traditional approaches; thus, for the next decade, the scientific community should remain in charge and at work. Indeed, the program DENDRAL [9], which analyzes mass spectrometer data to determine the chemical composition of test substances, seems to be the first case of an artificial intelligence program competitive with a person with a doctorate. In fact, it sometimes does better then such people and sometimes worse, in competitive analysis tests.

The work on meta-DENDRAL [10] (a program to produce DENDRAL type programs by automation) seems to be going slowly enough so that the next twenty years will still see human beings in charge, but the potential is clear for programs in the hundred-Megabyte range and nothing in the hardware development field would even cause a slowdown of such progress. (Optical disks can already hold programs in the Gigabyte size range if they are written.) It seems, however, that it takes a vast amount of intelligence to produce a relatively small increase in such values as status or ability to control others or salary. Accordingly, one can hope for a future in which the John Deans and the George Meanys of the world lose control to advanced robots only very slowly over many decades.

There remains only one uncertainty, but it is a large one. Now that is is clear that ordinary people make do with less than a Megabyte of memory for all functions it is clear that a one-Megabyte computer program designed to teach a computer to function in English is entirely reasonable. Such a program, and there seems no theoretical reason to require even difficult computer languages (perhaps BASIC might serve), would eliminate the requirement that the bulk of the program be in a computer language at all and thus it could be completed in about a year or so. Such a ten-Megabyte program, in such a short time, would be either a fantastic opportunity or a fantastic amount of trouble, probably both.

Bibliography

1. Yianilos, P.N., "A Dedicated Comparator Matches Symbol Strings Fast and Intelligently," Electronics, December 1, 1983, 56:113-117.

2. Pierce, J.R., Symbols, Signals, and Noise, Harper & Brothers, New York, 1961, pp. 238-249.

3. Zipf, F.K. Human Behavior and the Principle of Least Effort, Cambridge< Mass., Addison-Wesley, 1949.

4. Thorndike, E.L., Lorge, II, The Teachers Word Book of 30,000 Words, New York, Bureau of Publications, Teachers College, Columbia University, 1944.

5. Seashore, R.H., Eckerson, L.D., "The Measurement of Individual Differences in General English Vocabularies,", J. Educ. Psychol., 1940, 31, pp. 14-38.

6. The Washington Post, The Presidential Transcripts, Dell, 1974.

7. The New York Times, The Watergate Hearings, The Viking Press, New York, 1973.

8. Buchanan, B.G., Sutherland, G.L., & Feigenbaum, E.A., "Heuristic DENDRAL: A Program for Generating Explanatory Hypothesis in Organic Chemistry," Machine Intelligence, Vol. 4, pp. 209-254 (eds. Meltxer, B., & Michie, D.), Edinburgh: Edinburgh University Press, 1969.

9. McCorduck, Pamela, Machines Who Think, W.H. Freeman & Co., San Francisco, 1979.

A Note on Sorting Prolog Assertions

C.S. Kwok
Center of Computer Science and Applications
University of Hong Kong
Pokfulam Road
Hong Kong

In Prolog programs, arguments are often used to exchange information between relations. This kind of communication is tedious when a large amount of information has to be passed around. One way of solving the problem is to add assertions (unit clauses) to the knowledge base in execution time. These assertions are maintained as "global data structures." They are typically managed by the addition and deletion of clauses. However, most of the Prolog implementations offer little support for these operations. In C-Prolog [1], for example, a clause can only be added as the first or last clause of a relation. It would be useful if one could manipulate clauses in the same way as one manipulates lists. For example, sorting the assertions in the working memory of an expert system may lead the problem solver to follow a "better" path in the search space. In many applications, the working memory is build up incrementally. Thus, the insertion sort is a natural sorting method. The sorting may be combined with the addition of assertions using an assertcl(C) relation. A C-Prolog implementation of the articl(C) relation is given below.

```
/*************************************************************
 * assertcl(C) inserts a clause into the workspace.          *
 * The ordering of the clauses is defined by the             *
 * worse(C1, C2) relation, which holds when C1 is            *
 * less than C2 according to a certain ordering.             *
 *************************************************************

assertcl(Incl) :-
    make-rel(Name, Arity, Incl),
    sbc(Name, Arity, Incl, [], Stack),
    asserta(Incl).
    restore(Stack).

/*************************************************************
 * Clauses that are not "worse" than the clause to           *
 * be inserted are saved in a list :                         *
 *        r(a)                                               *
 *        r(b)                                               *
 *        r(c)                                               *
 * are saved as [r(a), r(b), r(c)]
 *************************************************************
sbc(Name, Arity, Incl, Instack, Outstack) :-
    make-rel(Name, Arity, WMc1),
    ( call(WMc1) ->
       ( worse(WMc1, Inc1)->

          Outstack = Instack;
          ( retract(WMc1),
            sbc(Name, Arity, Inc1, Instack, Interstack),
            Outstack = [WMc1 | Insterstack] ) );
       Outstack = Instack ).

/*************************************************************
 * restore a list backwards                                  *
 *************************************************************

restore([])
restore([X|Xs]) :-
```