LEARNING : THE CONSTRUCTION OF A POSTERIORI KNOWLEDGE STRUCTURES

Paul D. Scott,

Department of Computer and Communication Sciences, University of Michigan

Abstract

This paper is a critical examination of both the nature of learning and its value in artificial intelligence. After examining alternative definitions it is concluded that learning is in fact any process for the acquisition of synthetic a posteriori knowledge structures. The suggestion that learning will not prove useful in machines is examined and it is argued. that its main application in practical AI systems is in providing a means by which a system can acquire knowledge which is not readily formalizable. Finally some of the implications of these conclusions for future AI research are explored.

1: Introduction

In recent years machine learning seems to have undergone something of a renaissance. One manifestation of this is the recent publication of a book surveying the field (Michalski, Carbonell and Mitchell, 1983) Most of this book is devoted to reviewing what has been accomplished but, in what he clearly intended to be a provocative paper, Simon (1983) has raised a number of fundamental questions regarding the nature and value of machine learning.

In this paper I attempt to provide answers to these questions. In particular I shall try to define what learning is, why it is of great importance in artificial intelligence, how it relates to other branches of the subject and what these answers imply regarding future research in machine learning.

2. What Is Learning ?

When people use the term 'learning' in ordinary conversation they run little risk of being misunderstood. It is therefore somewhat surprising that A.I. researchers have had so much difficulty in arriving at a satisfactory definition. The usual explanation of this phenomena is the claim that the everyday use of 'learning' is very general and imprecise and actually refers to a heterogeneous collection of behaviors. There is much truth in this but the fact that people apply the same term to all these behaviors suggests that they have something fundamental in common.

The most widely accepted broad definition of learning within the A.I. community appears to be one relating it to improved performance. For example :-

> "Learning is any change in a system that allows it to perform better the second time on repetition of the same task or another task drawn from the same population"
>
> Simon (1983)

This is a functional definition. That is it defines 'learning' in terms of what it achieves rather than how it achieves it. Thus it is really a definition of the purpose of learning although it could be used as a definition of 'learning' if it is interpreted as meaning that any process which achieves that end is an example of learning.

Unfortunately if used in this way the definition is unsatisfactory in two respects. First it includes many behaviors which one would not want to classify as learning. For example, if I replace the old blade in my razor with a new one then I will perform the task of shaving better. It would be unreasonable to say the act of changing the blade constitutes an instance of learning every time I do it. Thus there are many changes which lead to improved performance which are not examples of learning.

The other problem with the definition is that it excludes many behaviors which would normally be classified as learning. For example, by the time the reader reaches the end of this sentence he or she will have learned that the author was born on a Tuesday. It is difficult to envisage a situation in which this knowledge could be used and hence it is clear that in this trivial example learning has nothing whatsoever to do with any performance.

It could be argued that common usage is at fault and that the involuntary acquisition of useless pieces of information should not be classed along with with the other behaviors which

*This work was supported by NSF grant # MCS-8203956

are regarded as examples of learning. However it it is not always clear when information is useless. In making a brief visit to a strange town I may happen to notice the location of the public library. Under normal circumstances this is the involuntary acquisition of useless information. If however a short time later someone stops me in the street and asks how to get to the library then the information suddenly becomes useful and will certainly improve my performance in answering the enquiry. Thus one cannot restrict the term 'learning' to the acquisition of useful knowledge or skills since the utility is not determined at the time learning occurs.

The functional definition of 'learning' in terms of improved performance does not therefore correspond to common usage of the term. One is therefore faced with a choice. One could decide that, when used in an A.I. context, 'learning' is a technical term whose meaning is determined by the functional definition rather than by common usage. Alternatively one could seek another definition.

The former course is possible but still presents difficulties. For example Samuel's checker player's (Samuel, 1963) ability to beat strong opponents was degraded by allowing the system to learn while playing weak opponents. At the same time it was certainly learning how to beat weak opponents. Thus by the functional definition it is learning with respect to one performance but not with respect to another drawn from the same population.

However, even if such difficulties can be overcome, there are strong arguments in favor of looking for alternative definitions. The first is that it is foolish to ignore common usage. For a concept to survive in everyday communications it must have some utility. Hence it is reasonable to suppose that there is some underlying unity about the collection of behaviors which are normally classified as 'learning'. The second reason is that the quest for such a definition, even if unsuccessful, should tell us something about the nature of learning processes.

3. The Organization Of Experience

We can identify certain common aspects of the systems and behaviors to which the term 'learning' is applied. First, any behavior described as learning seems to involve the notion that an event to which the system has been exposed influences the potential behavior of that system subsequent to that event. Furthermore, saying that a system learns appears to carry the implication that the system has the ability to store and retrieve information. The term 'learning' cannot be meaningfully applied to a system without such abilities. 'Learning' is associated with the storage rather than the retrieval of that information but there is a strong implication that the storage will be arranged in such a fashion that the information can be retrieved in appropriate circumstances.

These two attributes are closely connected since it is usually understood that the influence of an event on subsequent behaviour is mediated by retrieval of some information derived from the occurrence of that event.

Taking these attributes of learning together allows us to propose an alternative definition of learning. It is any process in which a system builds a retrievable representation of its past interactions with its environment. The term 'retrievable' should be understood to mean that the system itself can both access and interpret the representation. This definition may be more succinctly expressed :-

Learning is the organization of experience

Note that this definition, in contrast to the one discussed earlier, says nothing whatsoever about the purpose of learning.

What it does do is establish a strong link between learning and knowledge representation. This relationship can be made more explicit by the following equivalent definition :-

> Learning is any process through which a system acquires synthetic a posteriori knowledge.

In general a system will also have analytic knowledge such as rules of inference and synthetic a priori knowledge such as given facts about its environment. Both of these are supplied by the system designer.

4. Why Is Learning Important In Artificial Intelligence ?

The notion that knowledge representation is an essential part of any intelligent system is firmly established. Hence if we define learning as a process for building a representation of its environment then its potential utility is obvious. However learning is not the only way in which such a representation can be acquired. It can be explicitly supplied to the system as what is from the systems point of view synthetic a priori knowledge.

Many people seem to regard learning as an essential component of being intelligent. Hebb (1942,1949) distinguishes two meanings of the term 'intelligence'. 'Intelligence A' is the ability to acquire intelligent performance while 'Intelligence B' is the intelligent performance itself. Until recently, research in artificial intelligence has been very much more concerned with performance than with acquisition of that performance. That is in Hebb's terms it should perhaps be called 'artificial intelligence B'. The dominant research strategy for many of years was to try to discover ways in which an adequate representation of a specific problem domain can be constructed by the system designer and then utilized by the system to exhibit a desired type of performance. Thus the intelligent performance of a program is due to the combination of the designer and his program. The 'intelligence A' resides in the designer while the 'intelligence B' emanates from the program.

This suggests that a machine which is incapable of learning may be intellectually inferior to one which can. However, Simon (1983) notes that human learning appears to be an extremely tedious and inefficient process. It takes a long time to transfer expertise from one person to another. In contrast the knowledge structures of one computer program can be passed on to any number of other systems by means of a simple copy operation. Simon argues that this difference suggests there is not much point in trying to endow computers with human-like learning abilities. Why not just program the knowledge straight in?

This is an attractively simple argument. It does however carry the implicit assumption that 'just programming' will be a more effective way of endowing a system with knowledge than requiring it to go through some learning process. It is not at all obvious that this assumption is true.

Suppose one wished to construct an expert system for some domain. At present the process of developing an expert system requires a very large investment of effort by at least two people: one who already has the expertise that is to be incorporated into the expert system (the 'domain specialist') and one who can perform the programming necessary to incorporate it (the 'knowledge engineer'). The process involves an attempt to discover the rules and knowledge used by the domain specialist and then embody them in a form acceptable to the computer (usually as production rules). There are two obvious limitations to this approach. First of all it is very difficult and laborious. Secondly much of the domain specialist's expertise may not be introspectively available. He or she may be able to describe the general procedure he or she uses but is probably as incapable of explaining the reasons for exploring a particular possibility as a grandmaster is incapable of supplying an algorithm which replicates his chess playing ability.

Note that the creation of a human expert requires much less effort. The conventional education process requires that the teacher possess a sound knowledge of his subject but pedagogical skill appears to be much less important. Even a poor teacher usually succeeds in teaching his students a substantial fraction of what a very good teacher could have imparted. This is because the conventional education process is based on the premise that students are intelligent and therefore do almost all the work in building relevant mental representations themselves. Furthermore the conventional educational process normally involves the student in gaining experience by working on 'toy problems'. Much of the expertise he acquires is something that the teacher is incapable of formalizing. By repeated interaction with the problem domain the student will construct his own set of concepts which are useful for solving problems in that domain. If this were not true it would be just as easy to program a machine to write computer programs as it is to teach an introductory programming course.

The development of systems which could build their own representations would thus provide remedies to both the limitations of current expert system building techniques. First, since the system could build its own representation, the initial knowledge supplied need only be as complete and precise as that normally supplied in a classroom situation. Second, the system could extend this representation through experience to include concepts which its human teachers were unable to supply. Of course this process of learning by experience may be as time consuming for a machine as it is for a human. Certainly it makes sense to initialize the system with as much expertise as possible. However, as Simon notes, the machine expert has one enormous advantage over a human expert. The unteachable expertise which it acquires can be readily communicated to another machine. One simply makes a copy (or a thousand copies) of the final state of the original expert. In contrast, every human expert must individually go through the process of learning by experience. Viewed in this way it becomes clear that the copy process is complementary rather than an alternative to the learning process. Only the latter can create new knowledge structures.

Thus one situation in which learning is the only way a system can acquire a particular representation is when that knowledge cannot be readily formalized. Obviously it must be possible in principle to formalize the knowledge since otherwise a machine would not be able to acquire it. However it may be inordinately difficult or even impossible for a human to create such a formal representation explicitly.

Simon himself points out another situation in which learning would be necessary. This is the situation in which the structure of the system acquiring the representation is so complex that it is not practicable to modify it explicitly even if one knows what knowledge the system needs.

Thus it can be seen that learning is important in artificial intelligence because it provides a way in which a system can acquire knowledge that cannot be obtained by other means. What forms of knowledge have this characteristic is an important open question which deserves the serious attention of the AI research community.

5. The Growth Of Knowledge Structures

Defining learning as a process for acquiring a posteriori knowledge leads to some important conclusions regarding the place of machine learning within artificial intelligence. Computer scientists have long realized that there is an intimate relationship between data structures and the procedures which operate on them. We may apply this principle to knowledge structures.

There are two classes of operation which are applied to structures which represent knowledge. One class we call 'knowledge users'. These are the parts of the system which make use of the knowledge in the course of performing some task. The other class we call 'knowledge builders'. These are the parts of the system which construct and maintain knowledge structures.

The great effort which has gone into developing means of knowledge representation over the last decade or so has largely concentrated on the knowledge users. That is knowledge representation schemes have been developed with the aim of making them as usefully accessible as possible to the parts of the system responsible for its ultimate performance at its assigned task. This approach has yielded many valuable results. However it has only been possible because the system designers themselves played the role of knowledge builders. Since they were invariably much smarter than the knowledge user portions of their creations it made sense to construct representations for the latter's convenience alone.

However, if the AI community is going to pay serious attention to the problem of learning then new approaches to knowledge representation are necessary. In devising a representation scheme for a system which learns we must consider how it can be made easy to build and maintain as well as easy to use. Unlearnable representations, however versatile will have to be discarded.

Does this mean that we have to reject the powerful knowledge representation schemes that we already have? Not necessarily. The knowledge builders need to be able to manipulate such representations. If they are complex then the knowledge builders will need to be endowed with a fair bit of expertise about them. In other words they need metaknowledge. The simpler the knowledge structures the less metaknowledge is required.

Clearly this is a situation where important trade-offs must be made at the design stage. If you want a really simple learning component you have to be satisfied with a correspondingly simple knowledge representation. The art is going to be devising representation schemes which are sufficiently simple and uniform to make building and maintaining them easy yet sufficiently rich for the knowledge users to perform their task successfully.

It makes sense to avoid devising a knowledge structure such that the task of building and maintaining it is significantly harder than the task which the system is actually intended to perform. Interestingly it may be that this is not possible if the intended task is relatively simple. As the required repertoire of the system gets larger so the possibility of learning being worth the effort may increase. This may be why such versatile systems as human beings make so much use of learning.

If learning does prove most useful in very versatile systems it will dramatically change some widespread assumptions about how to make machines learn. With a few notable exceptions such as Lenat's AM (Lenat, 1976) most learning programs are based on the principle of improving their performance at some task by repeatedly attempting to perform that task. A learning system which forms part of a system with an enormous repertoire of potential tasks should acquire knowledge which is relevant to many of these tasks during the performance of other tasks. Ultimately it may have to acquire knowledge for knowledge's sake (as Lenat's AM does) rather than for its relevance to a specific task. This possibility is discussed in detail in Scott and Vogt (1983).

Another interesting consequence of this definition of learning is that some branches of artificial intelligence which have not traditionally been regarded as connected with learning are shown to be components of the learning process. For example truth maintenance systems (Doyle, 1979) can be viewed as systems which ensure the consistency of acquired knowledge structures.

I suspect that some readers may object that this type of process is not learning at all but reasoning or deduction or some similar concept. The point is that such processes cannot be meaningfully separated from 'pure learning'. Much human learning also makes extensive use of reasoning. Consider for example the well known section in Plato's Meno in which Socrates teaches a slave boy a special case of Pythagoras Theorem. Plato himself was so struck by the role of reasoning in human learning that he proposed that all learning is merely a form of recollection. It may be that in future artificial intelligence researchers will have to pay as much attention to inductive logic (Burks, 1977) as they have previously paid to deductive logic.

Most reviews of learning attempt to construct a taxonomy for the classification of the varied attempts which have been made to construct machine learning systems. Our definition of learning as a process for acquiring knowledge certainly suggests at least one way of categorizing such systems . They can be classified in terms of the knowledge structures built. Some authors (eg Michalski, Carbonell and Mitchell, 1983) have suggested what appear to be classification schemes of this nature but on closer examination they appear to be based not on the knowledge structure but rather the data structure used to represent the knowledge. I shall clarify this distinction by means of an example.

Samuel's checker player (Samuel, 1963) is often classed as a program which learns by adjusting coefficients in a polynomial expression. This is a perfectly correct characterization of the data structure that Samuel used. However classifying Samuel's program in this way leads to the paradoxical situation in which authors dismiss polynomial adjustment as too simplistic a view of learning and yet continue to find Samuel's program of interest. The reason is that what makes Samuel's program so interesting is not the fact that he uses a polynomial nor directly the way he adjusts it. Careful examination of Samuel's paper reveals that the polynomial is actually a model of the opponents behaviour based on the a priori assumptions that the opponent will use minimax and try to maximize piece advantage. Thus treating Samuel's program as a knowledge acquisition system we should say that the knowledge structure acquired is a representation of the opponent's checker playing behaviour.

My proposal is thus that we should classify learning programs in terms of what kind of knowledge is being acquired. This in no way detracts from the value of the traditional classifications in terms of either knowledge representation format or extent of assistance provided to the system by a teacher. It is orthogonal to these. Strangely it is often far from obvious exactly what knowledge is being acquired. For example I doubt if many people recognize what exactly Samuel's program is really learning on a single reading of his paper despite the strong hint given by its behaviour against weak opponents.

The overall conclusion to be drawn from this paper is that learning and knowledge representation are so closely connected that one cannot study the former without reference to the latter. Interest in machine learning seems to be growing rapidly. Hence I anticipate that over the next few years we are going to witness some radical rethinking of the way we view the problem of providing a system with knowledge it needs to solve problems in its allotted domain.

6. References

[1] Burks, A.W. "Chance, Cause and Reason" University of Chicago Press. 1977

[2] Doyle, J. "A Truth Maintenance System" Artificial Intelligence, Vol 12 , pp 231-272. (1979)

[3] Hebb, D.O. "The Effect of Early and Late Brain Injury Upon Test Scores and the Nature of Normal Adult Intelligence" Proc. Amer. Phil. Soc., Vol 85., pp 275-292. (1942)

[4] Hebb, D.O. "The Organisation of Behaviour" Wiley, New York. 1949

[5] Lenat, D. "AM : An Artificial Intelligence Approach to Discovery in Mathematics as Heuristic Search", Doctoral dissertation, Stanford University, July 1976.

[6] Michalski, R.S., Carbonell, J.G. and Mitchell, T.M. "Machine Learning", Tioga Publishing Co., Palo Alto. 1983

[7] Samuel, A.L. "Some Studies in Machine Learning using the Game of Checkers" In "Computers and Thought" Eds E. Feigenbaum and J. Feldman, McGraw-Hill, New York. 1963

[8] Scott, P.D. and Vogt, R.C. "Knowledge Oriented Learning" Proceedings IJCAI 1983

[9] Simon, H.A. "Why Should Machines Learn?" In Michalski, Carbonell and Mitchell (1983)