TEACHING ARTIFICIAL INTELLIGENCE

Patrick H. Winston

I: SUBJECT MATTER

Rapporteur: Mr. M.J. Elphick

Abstract

This is a summary of a talk which involved a discussion of the Marr-Poggio stereo algorithm. Their word is described in detail in "A Theory of Human Stereo Vision", Proc. Royal Society of London, vol. 204, 1979.

What is Artificial Intelligence?

To teach any subject, one must start with a statement of what the subject is about and what it is that makes it worth doing.

> Artificial Intelligence is the enterprise of making computers exhibit substantial cognitive, sensory, and motor abilities, approaching or going beyond human levels. Artificial Intelligence has two general objectives: to make computers more useful and to understand intelligence for its own sake.

Note that this view of Artificial Intelligence is such that vision and manipulation are involved as well as reasoning and problem solving.

Note also that the objectives of Artificial Intelligence make for natural alliances with many other fields, particularly Computer Science, Education, Psychology, and Linguistics.

The objectives are so broad that people often have difficulty understanding that there is a subject at all. To address this difficulty, one must repeat that the field is defined by its objectives - there is no analogue to Maxwell's equations of Electromagnetic Theory, not yet at least.

How should Artificial Intelligence be Taught?

In teaching Artificial Intelligence, one must decide early what is to be on top: methods, principles, and issues or case studies. One approach is to deal with such topics as programming, search, control, constraint, and the criteria by which a representation can be judged. Another one explores particular programs in such areas as learning, expert problem solving, natural language understanding, vision, and manipulation.

The MIT subject titled Artificial Intelligence is taught to about 150 undergraduates annually. It takes a mixed approach, treating methods, principles, and issues, but spending most of the lecture time with illustrative case studies, emphasizing attention to representation throughout. The following material was used in the past year, filling about 30 one-hour lectures, supplemented by a number of one-hour guest lectures:

Introductory Lectures
- Subject overview
- Search techniques
- Constraint propagation: Waltz's drawing analysis program
- Games, minimax, alpha beta
- Goal trees
- GPS and production systems

Lectures on Expert Problem Solving:
- Expert systems overview
- Dendral
- Mycin
- Teiresias
- Internist
- Casnet
- AM

Lectures on Representation, Language and Learning
- Planner-like data bases
- Restrictions and filters
- Frame-oriented languages, inheritance, demons
- Augmented transition networks: LIFER
- The meaning of meaning: Winograd's SHRDLU
- Modern sentence parsing: the work of Marcus
- Learning parsing rules: the work of Berwick
- Analogy: the work of Winston

Lectures on Robotics
- Robotics overview
- Vision overview
- Shading: the work of Horn
- Stereo: the work of Marr and Poggio

Miscellaneous Lectures
- Harpy
- Hearsay
- Programming apprentices
- Parallel problem solving
- Locomotion
- Society theory
- Social issues

"Artificial Intelligence" (by P.H. Winston, Addison-Wesley, 1977) is used as the text, enabling rapid progress through much of the basic material, but still leaving much of the subject without supporting material other than reprinted technical papers.

An understanding of LISP is a prerequisite, enabling the occasional use of code in examples. Suitable LISP background is in "LISP" (by P.H. Winston and B.K.P. Horn, Addison-Wesley, 1981). "LISP" is an expansion and improvement of the second half of "Artificial Intelligence".

Other than an understanding of LISP, no prerequisites are presumed, making the subject quite accessible to lower level undergraduates. Enrollment is about evenly distributed among sophomores, juniors, and seniors.

Representation

Representation is a pervasive theme. In teaching the MIT subject, we dwell on what a representation is and how to identify good ones.

While it is possible to be quite philosophical about what a representation is, to get started, at least, it is more productive to think of a representation as a vocabulary of symbols together with some conventions for arranging them.

A good representation tends to have many characteristics, including the following:

> It makes the important things explicit.
>
> It exposes constraints or regularities.
>
> It is computable from a natural input.

An Example

The work of the late David Marr and Tomaso Poggio and their colleagues on stereo vision is a splendid example of the importance of good representation. The basic arguments they make are as follows:

> The problem in seeing in stereo is partly that of knowing where edges are and partly that of coping with too many possible matches.
>
> Since edges usually manifest themselves as intensity changes, they show up as bumps in the first derivative and zero crossings in the second.
>
> It is therefore convenient to match edges by matching the zero crossings they produce in the Laplacian (the Laplacian being the two-dimensional generalization of two differentiations). There is another argument, beyond convenience: the Laplacian is rotationally symmetric, eliminating the problem of pasting together several partial results using directional derivatives.

To deal with ambiguity in match, the thing to do seems to be to work first with blurred images, made by convolving the originals with Gaussian filters. Blurring allows operation at several scales, making it possible to get approximate distance by working at a coarse scale, with few lines. Approximate distance then limits what a given line can match at a finer scale. Working with three scales seems appropriate.

Gaussian filtering is the right way to do the blurring because it minimizes the space-frequency product, as we know from linear systems theory. Taking the Laplacian and doing the blurring can be done by one filter, which looks like a sombrero in two dimensions, which again is a result from linear systems theory. This filter can be approximated as the difference of two Gaussians.

The predictions made by the theory, in terms of the kind of performance in the face of various kinds of noise, closely mimics what is actually found in human experiments. Moreover, there is considerable support in the neurophysiological literature.

The theory is full of beautiful insights. The most critical, however, has to do with representation: transforming the original image into the zero-crossing map makes the edges explicit and exposes the constraints enforced by the physical world.

Interestingly, the zero-crossings seem to capture most, if not all, of the information in the image. Logan's theorem for one-dimensional signals states that zero crossings determine a signal uniquely, up to a scale factor, if the signal is band limited to one octave and obeys certain other obscure conditions. This seems true for two dimensions as well, but remains unproved.

Discussion

Professor Randell opened the discussion by asking if the speaker (in view of his opening remarks) had any comments on the recently published account of AI in "Machines Who Think", by Pamela MacCorduck (W.H. Freeman, 1979).

Dr. Winston had not yet read it, but **Professor van Rijsbergen** (as an outsider to AI) found its style, putting the work in this field into context through extensive quotations from the protagonists, provided an interesting overview.

Turning to the place of psychology in AI, **Professor van Rijsbergen** expressed the view that David Marr's work on vision had started from a psychological view of human vision, looking at the computation needed. **Dr. Winston** disagreed; there were several ways to use observations of biological systems, and he didn't think that Marr had taken the approach referred to.

Professor Michaelson suggested that such arguments could become more theological than scientific; a different understanding of the structure of a Teletype compared with that of a telephone exchange was clearly necessary.

Dr. Winston commented that one still needed to know something of the purpose of an exchange (to connect any two subscribers) in order to comprehend its structure.

TEACHING ARTIFICIAL INTELLIGENCE

II: SUCCESS AND METHOD

Abstract

This is a summary of a talk which involved a discussion of work by the author on analogy. The work is described in detail in "Learning and Reasoning by Analogy", CACM, vol. 23, no. 12, December 1980.

Pervasive Themes

People doing Artificial Intelligence should know about what it means to be successful and about what it means to use good methods. These themes are well worth concentrating on in an Artificial Intelligence subject.

Criteria for Success

In any scientific work, it is necessary to have some way of determining if the work is successful. For much of Artificial Intelligence, the following seems appropriate:

> There should be a task to be performed. Otherwise success will end up defined as being able to do what one is able to do.
>
> There should be an implemented program that performs the specified task. Otherwise it is too easy to ignore or overlook the hardest problems or to overstate what has been understood.
>
> The implemented program should perform by virtue of identifiable principles. Otherwise it is too easy to be seduced by something that is ad hoc, limited, and uninformative.

Methodology

To reach success, a sound methodology is imperative. Marr was extraordinarily articulate in these matters, arguing along the following lines:

> First, it is necessary to observe or define the competence to be understood, formulating some representative tasks to be performed.
>
> Second, a representation is selected or invented that exposes constraint or regularity.
>
> Third, a precisely defined computation problem is posed.
>
> Fourth, programs are implemented that perform the desired computation.

And fifth, the results are validated by implementation and experimentation.

All this seems obvious, but there are strong temptations that often throw research out of proper perspective. One such temptation results in being caught up with an attraction to a particular representation. Worse yet, there may be an attachment to some particular algorithm, with a corollary failure to understand that many algorithms usually can be devised once a computation problem is properly laid out.

An Example

The author has followed Marr's leadership enthusiastically, working in quite a different domain, that of solving problems by analogy:

The Competence to be Understood. On looking at a collection of plot outlines people can find corresponding people, describe similarities and differences, and talk about relative degree of match. The objective was to deal only with simple plot outlines with at most a few people and a few hundred facts. The following rendering of Macbeth, in a form easily read by a natural language interface, was to be representative:

> In Macbeth-plot there is Macbeth Lady-macbeth Duncan and Macduff. Macbeth is an evil noble. Lady-macbeth is a greedy ambitious woman. Duncan is a king. Macduff is a loyal noble. Macbeth is evil because Macbeth is weak and because Macbeth married Lady-macbeth and because Lady-macbeth is greedy. Lady-macbeth persuades Macbeth to want to be king. Lady-macbeth influenced Macbeth because Lady-macbeth is greedy and because Lady-macbeth married Macbeth. Macbeth murders Duncan with Lady-macbeth using a knife because Macbeth wants to be king and because Macbeth is evil. Lady-macbeth kills Lady-macbeth. Macduff is angry. Macduff kills Macbeth because Macbeth murdered Duncan and because Macduff loved Duncan and because Macduff is loyal.

The Representation. From the competence to be understood, it is plain that there must be a way to capture the relationships among objects and the acts and other relations they participate in. The representation that seemed best suited was a sort of case grammar, with the agents, objects, and acts or relations highlighted and the instruments, sources, destinations, and other cases also expressible.

The Regularities. Various examples indicated clearly that object classes, object properties, and acts and relations between objects all may be important, but there is variation from plot to plot. The usefulness of analogies, however, derives from the expectation that cause-effect relations in the past will be reflected in cause-effect relations in the future. Therefore it is to be expected that the CAUSE relations in the plots will link up the acts and relations that should be considered important.

The Computational Problem. The problem is to put the objects in two plots into the best possible correspondence, given that they are represented as networks of objects related by case structures, some of which are noted to be important because they cause something or are caused by something.

The Implementation. There are a number of decisions to make, given the matching problem. For example, one has to decide how to search the space of possible matches and how to combine individual pieces of evidence to form an overall score. Currently the match space is searched exhaustively and each matching case structure scores one point.

The Experiments. Once implemented, programs compared Macbeth, Hamlet, Othello, Julius Caesar, and The Taming of the Shrew. Satisfyingly they reported that each plot matched itself best, the tragedies matched each other better than any of them matched the comedy. The best nonself match was Macbeth with Hamlet, which made sense, inasmuch as both plots involved someone who wanted to be king, whose desire caused him to murder the king, which caused him to be killed in turn by a loyal person, thus ending up dead.

Of course it is not just a matter is going through this once. One iterates, trying for substantial progress on each step of each loop. Since the initial work was completed, ideas have been developed so that implemented programs can deal with exercises using other situations as precedents, generating production-like principles as a side effect.

Discussion

Dr. Grossman suggested that an extension to "funding by analogy" (comparing successful research proposals) might be profitable, or even to political advice (with analogies to Machiavelli). **Dr. Winston** agreed, remarking that much practice in medicine and law is based on analogies with previous situations; however, a continual accumulation of new knowledge is essential for performance at a high level.

Professor Dijkstra was worried about references to "causal chains"; the separation into causes and effect was often arbitrary, e.g. was an act called murder because the King was dead, or was the King dead because of an act of murder? **Dr. Winston** pointed out that the same duality occurs in physical analogies; in the water-pipe analogy for the flow of electrical current, does the voltage cause the current flow, or vice-versa? We understand analogies by projecting constraints from one situation to another analogical one.

Professor Wilson returned to the Shakespearean analogies, and objected that the approach was based on far too superficial a view of the plays. Such simplification would merely lead to false analogies. The speaker agreed that shallow analogies were possible; one should develop abstractions, or modify the questions posed, e.g. to "which two plays were most similar with respect to a king being killed?".

Professor Michaelson agreed that the descriptions and comparisons used depended very much on what one was trying to do. The models of ideal fluid flow and of ideal electrical current flow led to very similar equations; for a stylistic analysis of literary texts, one would want to look at quite different features than for plot outlines. **Dr. Winston** said that he felt rescued by these remarks - one should allow a cycle, in which we use what we know to form new bases for analogy.

Professor Katzenelson asked whether Dr. Winston's concept of analogy could be said to be based on homeomorphisms between graph structures. The speaker's response was that this view was too syntactic; the matching is based on what is worth consideration, and has much more to do with semantics than syntactic structures.

Professor Pyle said that he was worried that only additional supportive evidence was accepted in the matching process. Was no contradictory evidence considered? **Dr. Winston** replied that this was indeed the case, that at the level of finding correspondences very little attention was paid to contradictory evidence. He was not aware however of any psychological work on this aspect of analogy.