N89-19837

Reducing Uncertainty by Using Explanatory Relationships

John R. Josephson, Ph.D., The Ohio State University

Abstract

Explanatory relationships can be used effectively to reduce the uncertainty that remains after diagnostic hypotheses have been scored using local matching.

1. Introduction

> The problem that the mind must solve is not that of reasoning with uncertainty, but reasoning DESPITE uncertainty -- how to come to robust conclusions despite uncertain data, inconclusive inference procedures, and incomplete knowledge.
>
> (-- B. Chandrasekaran)

Suppose that some black box **hypothesis source** delivers up a set of diagnostic hypotheses, each hypothesis given a confidence value on some scale. Suppose further that these confidence values can be taken to reflect "local match" or *prima facie* likelihood. That is, the confidence value associated with each hypothesis is a measure of its likelihood of being true, based only on consideration of the match between the hypothesis and the data with little or no consideration of interactions between potentially rival or otherwise related hypotheses. Thus we have a picture of a set of hypotheses where each has been somehow stimulated, evoked, and instantiated for the case, and at the current stage of processing each hypothesis has been scored in isolation from the others.

At this stage we have both a problem, and an opportunity to do something about it. The problem is that many hypotheses will probably have intermediate scores, representing hypotheses that can neither be taken as practically certain, nor as being of such a low confidence as to be ignorable. Some number of these hypotheses are presumably true, but how many and which ones?

The opportunity is that of bringing knowledge of interactions between the hypotheses to bear in order to reduce the degree of uncertainty associated with the hypotheses -- increasing confidence in some of them, and decreasing confidence in others.

Some types of interactions between hypotheses are:

- A and B are mutually incompatible.
- A is a more detailed refinement of B.
- A could be caused by B.
- A and B are mutually compatible, and are explanatory alternatives where their explanatory coverages overlap.

Besides hypothesis--hypothesis interactions of mutual incompatibility and support, which can arise by degrees as well as discretely, one especially interesting class of interactions concerns explanatory relations: what happens when two or more hypotheses represent alternative explanations for the same datum? The focus of this paper will be on explanatory relationships, and on how explanatory relationships impact on our estimates of confidence.

Knowledge of explanatory relationships gives us an opportunity to take advantage of some "best explanation" reasoning.

2. Best-Explanation Reasoning

Inference to the Best Explanation or *Abduction* is a form of inference that follows a pattern approximately like this:[1, 2, 3, 4]

> D is a collection of data (facts, observations, givens),
> H explains D (would, if true, explain D),
> No other hypothesis explains D as well as H does.
> ---
> Therefore, H is correct.

The strength of an abductive conclusion will in general depend on several factors, including:

- how good H is by itself, independently of considering the alternatives,
- how decisively H surpasses the alternatives,
- how thorough the search was for alternative explanations, and
- pragmatic considerations, including
  - the costs of being wrong and the benefits of being right,
  - how strong the need is to come to a conclusion at all, especially considering the possibility of seeking further evidence before deciding.

Abductions, as we have just characterized them, go from data describing something to an explanatory hypothesis that best accounts for that data.

3. Using Explanatory Relationships

Let us suppose that, besides a confidence value, each plausible diagnostic hypothesis (one which is not ruled-out) is associated with a description of which findings that hypothesis can explain.

One way to take advantage of these explanatory relationships is to set up a standard for when a diagnosis is complete. The diagnosis can be considered to be complete when all of the abnormal findings have been accounted for (explained). (This standard should be considered to be somewhat of an idealization, since for example unimportant findings need not be accounted for.)

Let us focus on the use of explanatory relationships to reduce the uncertainty that remains after the confidence scoring based on local matching. First we note that an overall abduction problem is set up - to account for all of the (abnormal) findings, and a series of small abduction problems is set up - to account for each particular (abnormal) finding. Our basic strategy will be to try to solve the overall abduction problem by solving some number of smaller and easier abduction problems.

First we solve the easiest little abduction problems, the ones in which we can have the most confidence. If a certain hypothesis is the only plausible explanation for some finding, then (supposing its local-match confidence value is not too low) it is entitled to as high confidence value, and entitled to be accepted into the overall composite hypothesis that represents the solution to the overall abductive problem. So first we form the set of**Essential** hypotheses consisting of those of the sort we have just mentioned.

If we are lucky the set of Essential hypotheses will together account for all of the (important abnormal) findings. If this occurs then the overall abduction problem is solved - the set of Essentials together constitutes the best explanation - and the diagnosis is complete. Hypotheses which are not part of this best explanation are lowered in confidence, since they are not needed as part of the final explanation, and everything that they explain can now be explained in some other (and in the context better) way.

If the Essentials do not explain everything, then next we form the set of **Clear Best** hypotheses consisting of those which explain findings for which there is no other explanation anywhere near as good. For example if some finding S can only be explained by A (moderate confidence), B (low confidence), and C (low confidence), then A is worthy of acceptance as being clearly the best way to explain S. Hopefully, the Essentials together with the Clear Bests will now explain everything, and the diagnosis can be considered to be complete (after perhaps removing some few hypotheses which are now explanatorily superfluous in the presence of the rest).

If the Essentials together with the Clear Bests do not explain everything we have done all we can do on the current evidence without resorting to guessing. Generally our best strategy under these circumstances would be to gather more data. In fact we are in a position to guide our data gathering by focusing on the problem of discriminating between alternative good explanations for important findings.

Yet sometime we have to decide quickly, and do not have enough time to gather further data. Also sometimes the cost of gathering further data is too high. Under these circumstances we still have the means available to do some clever guessing We can begin to include hypotheses which are best explanations for certain finding, but which are not far enough ahead of the alternatives, or not of high enough local-match confidence, to enable them to be accepted confidently. These Weakly Bests constitute the best guesses we can make under the circumstances.

Actually we can even do slightly better. Findings can be made to *vote* for the hypotheses which best explain them. The idea is that two different findings both pointing to the same hypotheses as the best explanation constitute (apparently) independent sources of evidence for the hypothesis, i.e. constitute converging lines of inference for the hypothesis. Hypotheses with more votes can be accepted more confidently than hypotheses with fewer votes, and enough can be accepted to complete the explanation. This phenomenon of converging lines of inference seems to be what the philosopher of science William Whewell (1794-1866) called "the concilience of inductions."

4. Conclusion

We have shown how a stage of diagnostic problem solving, where there are N viable plausible hypotheses, each with a confidence score based on local matching, can, by using explanatory relationships, be brought to a more advanced stage, where the number of hypotheses has been radically pruned to k hypotheses together representing a single compound hypothesis that explains a distinct portion of the data, and which is a "logically optimal" outcome in an abductive sense. I should point out that variations of this method have played an important role in several knowledge-based systems including the Red system for Red-cell antibody identification[5] and the Pathex system for diagnosing cholestatic liver diseases.

Note that, at the level of description we have been using, we might be describing the information processing of an "algorithmic computer", i.e. an instruction follower; or a "connectionist computer", i.e. one whose primitive processing elements work by propagating nudges and activation strengths. In either case what we are describing is the functional and semantic significance of various actions of the machine, not precisely how these actions are accomplished.

Note too that we might be describing a medical diagnosis engine, or a diagnoser of mechanical systems, a fragment of the processing of the vertebrate visual system, or the information processing that goes on when we recognize words in continuous speech. The strategy for reducing uncertainty that we have described is appropriate quite generally for a variety of abductive or interpretive information processing tasks.

5. Acknowledgments

Michael C. Tanner and Ashok Goel of Ohio State have contributed significantly to the development of these ideas, which have also benefited greatly from discussions with Jack Smith, Jr. and William Punch. This research has been supported in part by the Defense Advanced Research Projects Agency, RADC Contract F30602-85-C-0010 under the Strategic Computing Program, in part by the National Library of Medicine under grant LM-04298, and in part by the National Heart Lung and Blood Institute under grant HL-38776. Computer facilities have been enhanced through gifts from Xerox Corporation.

References

1. Josephson, John R., "A Mechanism for Forming Compositer Explanatory Hypotheses", *IEEE Transactions on Systems, Man and Cybernetics, Special Issue on Causal and Strategic Aspects of Diagnostic Reasoning*,Vol. SMC-17, No. 3, May, June 1987, pp. 445-54.

2. Peirce, C.S., *Abduction and Induction,* Dover, 1955, pp. 150ff., ch. 11.

3. Pople, H., "On the Mechanization of Abductive Logic", *Proceedings of the Third International Joint Conference on Artificial Intelligence*, 1973, pp. 147-152.

4. Eugene Charniak and Drew McDermott, *Introduction to Artificial Intelligence,* Addison Wesley, 1985.

5. Smith, Jack W.; Svirbely, John R.; Evans, Charles A.; Strohm, Pat; Josephson, John R.; Tanner, Mike, "RED: A Red-Cell Antibody Identification Expert Module", *Journal of Medical Systems*,Vol. 9, No. 3, 1985, pp. 121-138.