Explanation, Irrelevance and Statistical Independence *

Solomon E. Shimony
Computer Science Department
Box 1910, Brown University
Providence, RI 02912
ses@cs.brown.edu

Abstract

We evaluate current explanation schemes. These are either insufficiently general, or suffer from other serious drawbacks. We propose a domain-independent explanation system that is based on ignoring irrelevant variables in a probabilistic setting. We then prove important properties of some specific irrelevance-based schemes and discuss how to implement them.

Introduction

Explanation, finding causes for observed facts (or evidence), is frequently encountered within Artificial Intelligence. For example, some researchers (see [Hobbs and Stickel, 1988], [Charniak and Goldman, 1988], [Stickel, 1988]) view understanding of natural language text as finding the facts (in internal representation format) that would explain the existence of the given text. In automated medical diagnosis (for example the work of [Cooper, 1984], [Shachter, 1986], and [Peng and Reggia, 1987]), one wants to find the disease or set of diseases that explain the observed symptoms. In vision processing, recent research formulates the problem in terms of finding some set of objects that would explain the given image.

Following the method of many researchers (such as cited above), we characterize finding an explanation, as follows: given world knowledge in the form of (usually causal) rules, and observed facts (a formula), determine what needs to be assumed in order to *predict* the evidence. Additionally, we would like to select an explanation that is "optimal" in some sense.

There are various schemes for constructing explanations, among them the pure proof theoretic "theory of explanation" (see [McDermott, 1987]), set minimal abduction [Genesereth, 1984], and others; these are usually insufficiently discriminating (i.e. they would not be able to choose between many candidate explanations), because they only supply a partial ordering of explanations, that may result in an mutual incomparability of the best candidates (see [Charniak and Shimony, 1990]). A better explanation construction method is Hobbs and Stickel's weighted abduction [Hobbs and Stickel, 1988], and our variant of it, cost-based abduction [Charniak and Shimony, 1990]. However, the latter schemes do not handle negation correctly, because of the independence assumptions inherent to them. In fact, cost-based abduction may prefer an *inconsistent* (i.e. 0 probability) explanation to a reasonably probable explanation, as we show in [Shimony, 1990].

Another method suggested recently is the *coherence* metric, presented in [Ng and Mooney, 1990]. However, coherence suffers from some anomalies, as shown by [Norvig, 1991]. One anomaly is that if the proof subgraph of an explanation happens to contain several intermediate nodes, then that explanation may be spuriously preferred. Another anomaly occurs in cases where we may not want things to be explained by the same fact, and coherence will fail there. Coherence also fails to deal with uncertainty, or with cases where rules or predicates have priorities.

Probabilistic schemes for explanation are sufficiently discriminating, as they provide a total ordering of candidate explanations. These schemes also have a natural semantics, the probabilities of things occurring in the world. For these reasons, we focus exclusively on explanation in a probabilistic setting. Unfortunately, while there are numerous probabilistic explanation schemes, all of them are either insufficiently general or have other deficiencies, as shown by Poole in [Poole and Provan, 1990]. One of the schemes, Maximum A-Posteriori (MAP) model explanations (called MPEs in [Pearl, 1988]) is used here as a starting point, because it maximizes both internal consistency of the explanation (the probability of the model) and its "predictiveness" of the evidence. Formally, the MAP is the assignment $\mathcal{A}$ to all the variables that maximizes $P(\mathcal{A}|\mathcal{E})$, or the "most probable scenario given the evidence, $\mathcal{E}$". We make the simplifying assumption that the world knowl-

*This work has been supported in part by the National Science Foundation under grants IST 8416034 and IST 8515005 and Office of Naval Research under grant N00014-79-C-0529. The author is funded by a Corinna Borden Keen Fellowship. Special thanks to Eugene Charniak for helpful suggestions and for reviewing drafts of the paper.

edge is represented as (or at least is representable as) a probability distribution in the form of a belief network, as is done by many researchers in the field, such as in [Charniak and Goldman, 1988], [Cooper, 1984], and others.

MAP explanation has its proponents in the research community. Derthick, in his thesis [Derthick, 1988], talks about "reasoning by best model", which is essentially finding the most probable model given the evidence, and performing all reasoning relative to that model. We adopt that idea as a motivation for finding a single "best" explanation.

A serious drawback of MAP explanations and scenarios is that they exhibit anomalies, in the form of the *overspecification* problem, as demonstrated by Pearl in [Pearl, 1988]. He presents an instance of the problem (the vacation planning problem), a variant of which (where the belief network representation is made explicit) is presented below.

Suppose that I am planning to take some time off on vacation, and take a medical test beforehand. Now, the medical test gives me a 0.2 probability of having cancer[1] (i.e. probability 0.8 of being healthy). Now, if I am alive and go on vacation whenever I am healthy, then the most-probable state is: I am healthy, alive, and on vacation (ignoring temporal problems). Suppose, however, that I now plan my vacation, and am considering 10 different vacation spots (including one of resting at home), and make them all equally likely, given that I'm healthy (i.e. no preference). Also, assume that if I'm not well, then I will (with high probability) stay at home and die, but I still have a small probability of surviving and going on vacation. One way of representing the data is in the form of a belief network, as shown in figure 1. The network has three nodes: alive, healthy, and vacation spot[2]. The evidence is a conjunction, where we also allow assignments of U (unassigned) to appear for evidence nodes. The latter signifies that we are interested in an explanation to whatever value the node has, without actual evidence being introduced. Thus, in the example, the "evidence" can be stated as "alive = U".

In this network, since we have 10 vacation spots, the probability of any scenario where I am alive is 0.08; but the scenario where I die of cancer is the most likely (probability of 0.18)! This property of most-probable scenarios is undesirable, because it is not reasonable to expect to die just because of planning ahead one step.

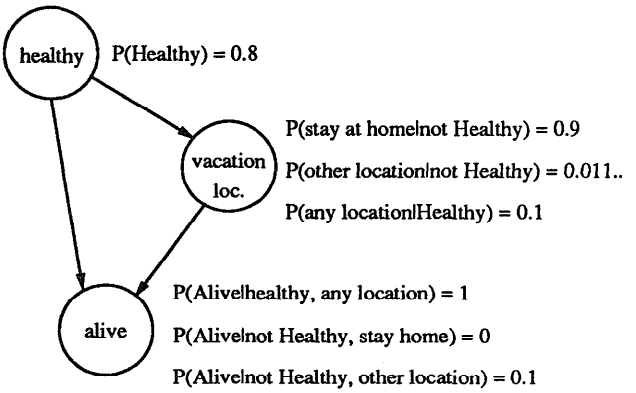

Figure 1: Belief network for the vacation-planning problem

Pearl suggests that we do not assign nodes having no "evidential support", i.e. those that have no nodes below them in the belief network. This is a good idea, but it is not sufficient, as it clearly does not help in the above example, because the "vacation location" node does have evidential support.

Despite its shortcomings, the MAP scheme does not suffer from potential inconsistencies. We use it as a starting point, and argue that by using a *partial* Maximum A-Posteriori model as an explanation, we can solve the overspecification problem. We use the intuition that we are *not interested* in the facts that are *irrelevant* to our observed facts, and consider models (explanations) where irrelevant variables are *unassigned*.

We propose two ways to define the class of partial models (or assignments) that we are interested in, i.e. to decide what is irrelevant. The first attempt, independence-based partial assignments, uses statistical independence as a criterion for irrelevance. We then define independence-based partial MAP as the highest probability independence-based assignment. We show that it alleviates the overspecification problem in some cases (it solves the vacation-planning problem correctly). We then outline a method of adapting our algorithm for finding complete MAPs, introduced in [Shimony and Charniak, 1990], to compute irrelevance-based partial MAPs. We do that by proving some important properties of independence-based assignments.

Using independence-based MAPs still causes assigning values to variables that we would think of as irrelevant. We propose δ-independence, an improved criterion for deciding irrelevance that is more liberal in recognizing facts as irrelevant. It specifies that a fact is irrelevant if the given facts are independent of it within a tolerance of δ.

[1] We assume here that (cancer = not healthy), i.e. assume that there are no other diseases, for the purposes of this example.

[2] The terms *nodes* and *variables* are used interchangeably. We use lower case names and letters for variable names, and their respective capitalized words and letters to refer to either a particular state of the variable, or as a shorthand for a particular assignment to the variable. In addition, a variable name appearing in an equation without explicit assignment, means that we actually have a set of equations, one for each possible assignment to the variable. E.g. $P(x) = P(y)$ where x and y are boolean valued nodes stands for $P(x = T) = P(y = T) \wedge P(x = F) = P(y = T) \wedge P(x = T) = P(y = F) \wedge P(x = F) = P(y = F)$. Assignments are treated as sample-space events when applicable.

Irrelevance-based MAPs

We now define our notion of best probabilistic explanation for the observed facts as the most probable partial model that ignores irrelevant variables. The criteria under which we decide which variables are irrelevant will be defined formally in the following sections. For the moment, we will leave that part of the definition open-ended and rely on the intuitive understanding of irrelevance. Suffice it to say that our definitions of irrelevance will attempt to capture the intuitive meaning of the term.

Definition 1 *For a set of variables B, an assignment[3] $\mathcal{A}^S$ (where $S \subseteq B$), is an* irrelevance-based assignment *iff the nodes $B - S$ are irrelevant to the assignment.*

In the vacation planning example, we would say that the vacation-location is irrelevant to the assignment $\{Alive, Healthy\}$.

Definition 2 *For a distribution over the set of variables B with evidence $\mathcal{E}$, an assignment $\mathcal{A}^S$ is an* irrelevance-based MAP *if it is the most probable irrelevance-based assignment that is complete w.r.t. the evidence nodes, such that $P(\mathcal{E}|\mathcal{A}^S) = 1$.*

This is a meta-definition. We will use different definitions of irrelevance-based assignments to generate different versions of irrelevance-based MAPs. With the "intuitive" definition, in our vacation-planning example, the irrelevance based MAP is $\{Alive, Healthy\}$, which is the desired scenario.

We say that the irrelevance-based MAP w.r.t. the evidence $\mathcal{E}$ is the best explanation for it. Note that the definition above is not restricted to belief networks. Our formal definitions of irrelevance, however, will be restricted to belief networks, and will rely on the directionality of the networks, the "cause and effect" directionality. In belief networks, an arc from u to v states that u is a possible cause for v. Thus, the only possible causes of a node v are its ancestors, and thus (as in Pearl's evidential support), all nodes that are not ancestors of evidence nodes are unassigned. Additionally, we do not assign (i.e. are not "interested" in) nodes that are irrelevant to the evidence given the causes. The ancestors are only *potentially* relevant, because some other criterion may cause us to decide that they are *still* irrelevant, as shown in the next section.

Independence-based MAPs

Probabilistic irrelevance is traditionally viewed as statistical independence, or even independence given that we know the value of certain variables (the independence model that is due to Pearl). The latter is known, in the case of belief networks, as d-separation. However, using d-separation as a criterion for deciding which nodes are irrelevant does not suffice for our example, because clearly the "vacation spot" and "alive" nodes are not d-separated by the "healthy" node. As a starting point for our notion of probabilistic irrelevance, we use Subramanian's strong irrelevance ([Subramanian, 1989]). In that paper, $SI(f, g, M)$ is used to signify that f is irrelevant to g in theory M if f is not necessary to prove g in M and vice versa (see [Subramanian, 1989] for the precise definition). We use the syntax of that form of irrelevance, but change the semantics. That is because we are interested in irrelevance of f to g even if g is not true. We define probabilistic irrelevance relative to sets of models, rather than theories (as in [Subramanian, 1989]). This is necessary because the (more general) probabilistic representation does not have implications, just conditional probabilities.

Partial assignments induce a set of models. For example, for the set of variables $\{x, y, z\}$, each with a binary domain, the assignment $\{x = T, y = F\}$ with z unassigned induces the set of models $\{(x = T, y = F, z = F), (x = T, y = F, z = T)\}$. We will limit ourselves to the sets of models induced by partial assignments, and use the terms interchangeably. We say that $In(f, g|\mathcal{A})$ if f is independent of g given $\mathcal{A}$ (where $\mathcal{A}$ is a partial assignment), i.e. if $P(f|g, \mathcal{A}) = P(f|\mathcal{A})$. We allow f and g to be either sets of variables or assignments (either partial or complete) to sets of variables. This is similar to Pearl's independence notation, $I(X, Y, Z)$, where variable set X is independent of variable set Z given variable set Y. The difference is that Pearl's notation does not require a certain assignment to Y, just that the assignment be known; whereas our notation does require it. For any disjoint sets of variables X, Y, Z, we have that $I(X, Y, Z)$ implies $In(X, Z|\mathcal{A}^Y)$, but *not* vice-versa.

We now define our first notion of an irrelevance-based assignment formally (we call it *independence-based assignment*):

Definition 3 *An assignment $\mathcal{A}^S$ is an independence-based assignment iff for every node $v \in S$, $\mathcal{A}^{\{v\}}$ is independent of all its ancestors that are* not *in S, given $\mathcal{A}^{S\uparrow^+(v)}$.*[4]

The idea behind this definition is that the unassigned nodes above each assigned node v should remain unassigned if they cannot affect v (and thus cannot be used to explain v). Nodes that are not ancestors of v are never used as an explanation of v anyway, because they are not potential causes of v.

Definition 4 *An independence-based MAP is an irrelevance-based MAP where independence-based as-*

[3] $\mathcal{A}$ denotes assignments. The superscript denotes the set of assigned nodes. Thus, $\mathcal{A}^S$ denotes an assignment that is *complete* w.r.t. S, i.e. assigns some value (but not U) to each node in S.

[4] $\uparrow$ is shorthand for "immediate predecessors of". $\uparrow^+$ is the non-reflexive, transitive closure of $\uparrow$. Thus, $\uparrow^+(v)$ is the set of all the ancestors of v. We omit the set-intersection operator between sets whenever unambiguous, thus $S\uparrow^+(v)$ is the intersection of S with the set of ancestors of v.

signments are substituted for irrelevance-based assignments.

In our example, using independence-based MAPs, we have a best scenario of {Alive, Healthy, vacation location undetermined} with a probability of 0.8 as desired. We do not assign a value to vacation location because the only node v with unassigned ancestors is v=alive, and the conditional independence In(alive, vacation spot | Healthy) holds.

Properties of Independence-based Assignments

The independence constraints in the definition of independence-based assignments leads to several interesting properties, that are desirable from a computational point of view.

We make the following observation: if, for each assigned variable v, v is independent of all of its unassigned *parents* given the assignment to the rest of its parents, then the entire assignment is independent of the unassigned ancestors. Thus, to test whether an assignment is independence-based, we only need to test the relation between each node and its parents, and can ignore all the other ancestors. Formally:

Theorem 1 *For all assignments $\mathcal{A}^S$ that are complete w.r.t. S, the nodes of some subset of belief network B, if for every node $v \in S$, $In(\mathcal{A}^{\{v\}}, \uparrow(v) - S | \mathcal{A}^{S\uparrow(v)})$, then $\mathcal{A}^S$ is an independence-based assignment.*

Proof outline: We construct a belief network B', that is the same as B except for intermediate nodes inserted between nodes and their parents. The extra nodes map out all possible assignments to each node v and its parents, where nodes are collapsed together whenever v is independent of some subset of its parents given the assignment to the rest of its parents. Then we show that the marginal distribution of B' is the same as B. We use the d-separation of nodes in B' to show independence of nodes and their ancestors in the constructed network, and carry these independence results back to the original network, B.

Theorem 1 allows us to verify that an assignment is independence-based in time linear in the size of the network, and is thus an important theorem to use when we are considering the development of an algorithm to compute independence-based MAPs. Additionally, if B has a strictly positive distribution, then theorem 1 also holds in the reverse direction. This allows for a linear-time verification that an assignment is *not* independence-based.

The following theorem allows for efficient computation of $P(\mathcal{A}^S)$:

Theorem 2 *If $In(v, \uparrow(v) - S, \mathcal{A}^{S\uparrow(v)})$ for every node $v \in S$, then the probability of $\mathcal{A}^S$ is:*

$$P(\mathcal{A}^S) = \prod_{v \in S} P(\mathcal{A}^{\{v\}} | \mathcal{A}^{S\uparrow(v)})$$

Proof outline: Let O be the set of belief network nodes not in S. We argue that, because of the independence constraints, we can write the joint probability of the entire network, $P(nodes(B))$ as the product $P(\mathcal{A}^S)P(\mathcal{A}^O)$, for any possible assignment to the nodes of O. The joint probability of a belief network can always be written as a product of probabilities of nodes given their parents. To calculate $P(\mathcal{A}^S)$, we marginalize $P(\mathcal{A}^O)$ out, by summing over all the possible values of the unassigned nodes. Thus, we can write:

$$P(\mathcal{A}^S) = \mathcal{S} \prod_{v \in S} P(\mathcal{A}^{\{v\}} | \mathcal{A}^{S\uparrow(v)})$$

with

$$\mathcal{S} = \sum_{\mathcal{A}^O_c} \prod_{v \in O} P(\mathcal{A}^{\{v\}} | \mathcal{A}^{\uparrow(v)})$$

where the c subscript denotes all possible *complete* assignments, in this particular case all complete assignments to the set of nodes O. We then argue that $\mathcal{S}$ is the sum of the probabilities of a complete sample space, and thus is equal to 1.

The theorem allows us to find $P(\mathcal{A}^S)$ in linear time for independence-based assignments, as the terms of the product are simply conditional probabilities that can be read off from the conditional distribution array (or other representation) of nodes given their parents.

Algorithmic Issues

In order to be able to adapt standard algorithms for MAP computation to compute independence-based MAPs, we need to be able to do two things efficiently: a) test whether an assignment is independence-based, and b) evaluate its probability. This is usually a minimal requirement[5], whether the form of our algorithm is simulation, best-first search or some other method. In the case of independence-based MAPs, theorems 1 and 2 indeed provide us with linear-time procedures to meet these conditions, which allows us to take a complete-MAP algorithm and convert it to an independence-based MAP algorithm.

We presented a best-first search algorithm for finding complete (rather than partial) MAP assignments to belief networks in [Shimony and Charniak, 1990]. The algorithm finds MAP assignments in linear time for belief networks that are polytrees (i.e. the underlying graph, with all edges replaced by undirected edges, is a set of trees), but is potentially exponential time in the general case, as the problem is provably NP-hard.

The algorithm was modified to compute independence based MAPs. We describe the algorithm and modifications more fully in [Shimony, 1991], but review it here. An agenda of states is kept, sorted by current probability, which is a product of all conditional probabilities seen in the current expansion. The operation of the algorithm is summarized in the following table.

[5] We can survive without requirement a) if we have a scheme that enumerates independence-based assignments, but even in that case theorem 1 will help us out.

1	queue evidence "state" into agenda
2	de-queue agenda into current
3	if current state is not complete go to step 5
4	if need only one result, halt
5	extend current state, and go to step 2

The modifications are in checking for completeness in step 3, and in extending the current agenda item. In both cases, the extension consists of picking a node, and assigning values to its neighbors. Each such assignment generates a new state. The states are evaluated and queued onto the agenda. The difference is that with the modified algorithm, when extending a node, we never assign values to its children, and some of the parents need not be assigned, which actually saves some work w.r.t. the complete-MAP algorithm. Completeness in the modified algorithm is different in that an agenda item may be complete even if not all variables are assigned, and in fact we use the results of theorem 1 directly to check for completion. We will not pursue further details of the algorithm here, as it is discussed in [Shimony, 1991].

Evaluating Independence-based MAPs

We can see that independence-based assignments solve the vacation-planning problem, in that "vacation spot" is irrelevant to "alive" given "Healthy", using the conditional independence criterion of definition 3. However, this definition of irrelevance is still insufficient because slightly changing conditional probabilities may cause assignment to variables that are still intuitively irrelevant, which may in turn cause the wrong explanation to be preferred (the *instability problem*). The latter problem manifests if we modify the probabilities in our vacation-planning problem slightly, as in the following paragraph.

Change the probability of being alive given the location so that probability of "Alive" given "Healthy" and staying at home is still 1, but only 0.99 given "Healthy" and some other location (say an accident is possible during travel). We no longer have independence, and thus are forced into the bad case of finding the "not alive" scenario as the best explanation. This is counter-intuitive, and we need to find a scheme that can handle "almost" independent cases.

δ-independence and Explanation

In order to improve the relevance performance of independence based explanation, we will attempt to relax the independence constraint that stands at the heart of the scheme. This will allow us to assign fewer variables, hopefully ones that are not independent but still intuitively irrelevant. We relax the exact independence constraint by requiring that the equality hold only within a factor of δ, for some small δ.

Definition 5 *We say that* **a** *is δ-independent of* **b** *given $\mathcal{A}^S$, where* **a**, **b** *and S are sets of variables (in our notation: δ-$In(\mathbf{a}, \mathbf{b}|\mathcal{A}^S)$), iff*

$$\min_{\mathcal{A}^{\mathbf{b}}} P(\mathcal{A}^{\mathbf{a}}|\mathcal{A}^S, \mathcal{A}^{\mathbf{b}}) \geq (1-\delta) \max_{\mathcal{A}^{\mathbf{b}}} P(\mathcal{A}^{\mathbf{a}}|\mathcal{A}^S, \mathcal{A}^{\mathbf{b}})$$

This definition is naturally expanded for the case of **a** and **b** being (possibly partial) *assignments* rather than sets of variables (in which case read **a** for $\mathcal{A}^{\mathbf{a}}$, and **b** for $\mathcal{A}^{\mathbf{b}}$). This definition is parametric, i.e. δ can vary between 0 and 1.

Definition 6 *An assignment $\mathcal{A}^S$ is δ-independence based iff $\delta - In(\mathcal{A}^{\{v\}}, O\uparrow^+(v)|\mathcal{A}^{S\uparrow^+(v)})$, for every $v \in S$ (where O is the set of variables not in S).*

Note that the case of $\delta = 0$ reduces to the independence-based assignment criterion. In the case of our modified vacation-planning problem, and $\delta = 0.1$, we get the desired δ-independence based MAP of $\{Alive, Healthy\}$, alleviating the instability problem. Using δ-independence as a measure of irrelevance solves the vacation-planning problem and its modified counterpart, but we need to show that finding δ-independence based MAPs is practical. To do that, we need to prove locality theorems similar to theorems 1 and 2. In the former case, it works:

Theorem 3 *If $\mathcal{A}^S$ is a complete assignment w.r.t. subset S of belief network B, and for every node $v \in S$, $\delta - In(\mathcal{A}^{\{v\}}, \uparrow(v) - S|\mathcal{A}^{S\uparrow(v)})$ (for $0 \leq \delta \leq 1$), then $\mathcal{A}^S$ is a δ-independence-based assignment to B.*

Proof outline: Expand the probability of node v given its parents as a marginal probability, summing over states of the unassigned indirect ancestors of v. The sum of probabilities over all states of the ancestor nodes equals 1. Using a convexity argument, we show that the minimum probability of v, given any states of its ancestors, occur when all the parents of v are assigned. Thus, the minimum probability of v given some assignment to its direct parents is smaller than the probability of v given any assignment to the indirect ancestors of v. A similar argument can be made for the respective maxima. Thus, given that the minimum and maximum probabilities (of v given *parents*) are within a factor of $1-\delta$ of each other, the minimum and maximum over the states of all unassigned ancestors are also within factor $1-\delta$.

Computing the exact probability of δ-independence based assignments is hard, but the following bound inequalities are always true:

$$P(\mathcal{A}^S) \leq \prod_{v \in S} \max_{\mathcal{A}^{U\uparrow(v)}} P(\mathcal{A}^{\{v\}}|\mathcal{A}^{S\uparrow(v)}, \mathcal{A}^{U\uparrow(v)})$$

$$P(\mathcal{A}^S) \geq \prod_{v \in S} \min_{\mathcal{A}^{U\uparrow(v)}} P(\mathcal{A}^{\{v\}}|\mathcal{A}^{S\uparrow(v)}, \mathcal{A}^{U\uparrow(v)})$$

Proof outline: Apply the argument of theorem 3 to all $v \in S$, and the multiplicative property of belief networks (as in the proof of theorem 2) to get the inequalities above.

The bounds get better as δ approaches 0, as their ratio is at least $(1-\delta)^{|S|}$. They can be computed using only *local* information, the conditional independence arrays of v given its parents, for each v in the assignment. In the worst case, we need to scan all the possible value assignments to the (currently unassigned) parents of each v, but the computation time is still linear in $|S|$. In practice, it is possible (and feasible) to precompute the bounds for each v and each possible δ-independence based assignment to v and its parents (*not* all possible δ-independence based assignments to the *entire network*). In [Shimony, 1991], we show how to adapt our basic MAP algorithm to compute δ-independence based MAPs.

There is, however, a problem with δ-independence, that of determining the correct value of δ. In the above example, where did the value $\delta = 0.1$ come from? We could state that this is some tolerance beyond which we judge variables to be sufficiently near to independence, and that we can indeed pick some value and use it for all explanation problems successfully. That does not appear to be the case in certain examples we can cook up, especially when nodes can have many possible values (say, more than 10 values). A good solution to the problem is using a variable δ. We are looking into a method of making δ dependent on prior probabilities of nodes, but are also considering basing its value on the number of values in a node's domain.

Summary

Previous research ([Poole and Provan, 1990], [Charniak and Shimony, 1990], and [Shimony, 1990]) has shown that existing explanation systems have major drawbacks. We have looked at probabilistic systems for a solution. We started off with MAP explanations, and observed that the system suffers from overspecifying irrelevant variables. We defined the concept of partial MAPs, and a particular kind of partial MAP called "irrelevance-based MAP", in which "intuitively irrelevant" nodes are left unassigned.

We then defined irrelevance as statistical independence, showed how it helps in certain cases, and proved important properties of independence-based assignments that facilitate designing an algorithm to compute them. Independence-based MAPs still suffer from irrelevant assignments, and we discussed the relaxation of the independence based assignment criterion, by using δ-independence, to solve the problem.

References

Charniak, Eugene and Goldman, Robert 1988. A logic for semantic interpretation. In *Proceedings of the ACL Conference.*

Charniak, Eugene and Shimony, Solomon E. 1990. Probabilistic semantics for cost-based abduction. In *Proceedings of the 8th National Conference on AI.*

Cooper, Gregory Floyd 1984. *NESTOR: A Computer-Based Medical Diagnosis Aid that Integrates Causal and Probabilistic Knowledge.* Ph.D. Dissertation, Stanford University.

Derthick, Mark 1988. *Mundane Reasoning by Parallel Constraint Satisfaction.* Ph.D. Dissertation, Carnegie Mellon University. Technical report CMU-CS-88-182.

Genesereth, Michael R. 1984. The use of design descriptions in automated diagnosis. *Artificial Intelligence* 411–436.

Hobbs, Jerry R. and Stickel, Mark 1988. Interpretation as abduction. In *Proceedings of the 26th Conference of the ACL.*

McDermott, Drew V. 1987. Critique of pure reason. *Computational Intelligence* 3:151–60.

Ng, Hwee Tou and Mooney, Raymond J. 1990. On the coherence in abductive explanation. In *Proceedings of the 8th National Conference on AI.* 337–342.

Norvig, Peter 1991. Personal communication.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann, San Mateo, CA.

Peng, Y. and Reggia, J. A. 1987. A probabilistic causal model for diagnostic problem solving (parts 1 and 2). In *IEEE Transactions on Systems, Man and Cybernetics.* 146–162 and 395–406.

Poole, David and Provan, Gregory M. 1990. What is an optimal diagnosis? In *Proceedings of the 6th Conference on Uncertainty in AI.* 46–53.

Shachter, R. D. 1986. Evaluating influence diagrams. *Operations Research* 34:871–872.

Shimony, Solomon E. and Charniak, Eugene 1990. A new algorithm for finding map assignments to belief networks. In *Proceedings of the 6th Conference on Uncertainty in AI.*

Shimony, Solomon E. 1990. On irrelevance and partial assignments to belief networks. Technical Report CS-90-14, Computer Science Department, Brown University.

Shimony, Solomon E. 1991. Algorithms for irrelevance-based partial maps. Submitted to the Conference on Uncertainty in AI.

Stickel, Mark E. 1988. A prolog-like inference system for computing minimum-cost abductive explanations in natural-language interpretation. Technical Report 451, Artificial Intelligence Center, SRI.

Subramanian, Devika 1989. *A Theory of Justified Reformulations.* Ph.D. Dissertation, Stanford University. Technical report STAN-CS-89-1260.