Probabilistic Similarity Networks[*]

David Heckerman
Medical Computer Science Group
Knowledge Systems Laboratory
Departments of Computer Science and Medicine
Stanford, California 94305

October 5, 1990

Abstract

I address practical issues concerning the construction of normative expert systems, and examine the influence diagram as a potential framework for representing knowledge in such systems. I introduce an extension of the influence-diagram representation called a *similarity network*. A similarity network is a tool for constructing large and complex influence diagrams. The representation allows a user to construct independent influence diagrams for subsets of a given domain. A valid influence diagram for the entire domain can then be constructed from the individual diagrams. Similarity networks represent forms of conditional independence that are not represented conveniently in an ordinary influence diagram. I discuss in detail one such conditional independence, called *subset independence*, and examine how similarity networks exploit this form of independence to facilitate the construction of an influence diagram. Also, I describe the assessment of probability distributions for influence diagrams. I show that similarity networks exploit subset independence to simplify such probability assessments. I introduce a representation that is closely related to similarity networks, called a *partition*. This representation further exploits subset independence to simplify probability assessment. Finally, I examine a real-world normative expert system for the diagnosis of lymph-node pathology, called Pathfinder. The similarity-network and partition representations played a crucial role in the construction of this expert system.

1 Introduction

Over the last 2 decades, decision analysts have been using decision theory in conjunction with a collection of knowledge representations and heuristic techniques to provide clarity of action to individuals and groups who are confused about important decisions. Decision

---

[*] This work was supported by the National Library of Medicine under Grant RO1LM04529, and by the National Science Foundation under Grant IRI-8703710.

theory includes probability theory (sometimes referred to as subjective probability theory or Bayesian probability theory) and the maximum expected utility principle, which states that a decision maker should choose the alternative that maximizes his expected utility. Perhaps the most significant virtue of decision theory is that it is *normative*. That is, the theory provides a set of gold standards for how people wish they could behave when allocating scarce resources under uncertainty. It is well known that people often do not behave in accordance with these gold standards [5, 28, 16]. Thus, a decision analysis can save millions of dollars or even lives and suffering when the stakes are high.

For nearly the same period of time, knowledge engineers have been using a set of representations and heuristic techniques developed by researchers in artificial intelligence to build *expert systems*: computer programs that bring to bear the knowledge of an expert or group of experts on a class of decisions or *domain* (e.g., Mycin and Prospector are expert systems for the diagnosis and treatment of bacterial infections, and for site selection for mineral exploration, respectively [25, 4]). Decision theory and decision-analytic techniques have rarely been used in the construction of such systems. Although some artificial-intelligence researchers have avoided a decision-analytic approach on theoretical grounds [24, 29], most workers have turned to alternative approaches for decision making because they believe that the normative approach to constructing expert systems for large, real-world domains is impractical [26] [21, pages 184–199].

Over the last 6 to 7 years, several researchers working at the boundary of decision analysis and artificial intelligence have been attempting to build *normative expert systems*: expert systems that use a decision-theoretic model as the framework for knowledge representation and inference. From the perspective of decision analysis, such systems could provide decision assistance across a wide range of possible decisions in a given domain. By avoiding the expense of analysts and experts for every confusing and high-stakes decision to be made, a normative expert system could reduce significantly the costs of decision making. From the perspective of artificial intelligence, use of a normative theory as the framework for representing knowledge could improve dramatically the quality of expert knowledge that is delivered to the user of an expert system.

A major breakthrough on the path toward the creation of normative expert systems has been the development of the *influence diagram*, a representation that graphically represents the beliefs, alternatives, and preferences of a decision maker [15]. The influence diagram is a natural representation for the knowledge base of an expert system. The representation is mathematically precise, yet has a human-oriented qualitative structure that facilitates communication between the expert and a decision model. Moreover, influence diagrams can represent any decision problem.

The influence-diagram representation facilitates the three major facets of expert system development: *knowledge acquisition*, the process of capturing and encoding the knowledge of an expert or experts; *inference*, the generation of recommendations or relevant information based on user input and the expert knowledge; and *explanation*, the process of communicating such recommendations or relevant information to the user. Influence diagrams simplify knowledge acquisition, because we can use them to represent graphically assertions of conditional dependence and independence *before* we need to consider assessments of probabilities or utilities. We can use these assertions of conditional independence to decompose the assessment of a joint probability distribution into a collection of independent assessments of manageable size. Such decomposition helps us to focus attention during knowledge acquisition, and to decrease the size of the construction task. In addition, we can use the assertions of conditional independence in an influence diagram to increase the computational efficiency of decision-theoretic inference. Specifically, researchers have developed exact and approximate inference algorithms that exploit assertions of conditional independence in an influence diagram to avoid direct computations on the joint probability distribution associated with that diagram [22, 19, 17, 10, 3]. Furthermore, we can use the graphical representation of conditional independence to generate cogent explanations to the builders and users of normative expert systems [19, Chapters 5 and 10]. Given these features of the influence-diagram representation, it is not surprising that several normative expert systems have been constructed using the representation [27, 2, 20, 11, 18, 1].

In this paper, I address the pragmatic aspects of capturing and representing knowledge for normative expert systems. I introduce two graphical extensions to the influence-diagram representation called *similarity networks* and *partitions*. A similarity network is a tool for constructing an influence diagram, whereas a partition is a tool for assessing the probabilities associated with an influence diagram.[1] Both representations facilitate the development of large and complex models by exploiting forms of conditional independence that are not easily represented in an ordinary influence diagram. In this paper, I scrutinize these representations and the forms of conditional independence that they embody, and show how their use has made practical the construction of a normative expert system for medicine.

In Sections 2 and 3, we examine a large medical expert system called Pathfinder that motivated the development of similarity networks and partitions. We discuss the new representations, and investigate how these representations were used to simplify the construction of the Pathfinder influence diagram. In Sections 4 and 5, we examine the representations in detail by way of a simple medical example.

2 Pathfinder: A Normative Expert System

Pathfinder is a normative expert system that assists community pathologists with the diagnosis of lymph-node pathology [9]. The domain of Pathfinder is large by any standard of comparison for expert systems. Over 60 diseases can invade the lymph node (25 benign diseases, 9 Hodgkin's lymphomas, 18 non-Hodgkin's lymphomas, and 10 metastatic diseases). In addition, there are over 100 morphologic distinctions or *features* within lymph nodes that can be easily recognized on microscopic examination. Each feature is associated with two or more mutually exclusive and exhaustive *instances*. Also, Pathfinder contains over 30 features reflecting clinical, laboratory, immunological, and molecular biological information that are

---

[1] Some authors use the term influence diagram to refer to both the network representation and the probabilities that underlie the network. In this paper, however, I use the term to refer to only the network.

relevant to the diagnosis of lymph-node disease.

The goals of Pathfinder are (1) to provide a probability distribution over diseases—a *differential diagnosis*—given a set of set of observations made by the user, and (2) to suggest to the user additional features for observation that are likely to narrow the differential diagnosis. To accomplish these goals, the expert system uses the method of *sequential diagnosis* [7]. First, when evaluating sections of a lymph node, a user enters a set of salient morphologic observations, and Pathfinder uses Bayesian inference to compute the posterior probabilities of disease, given these observations. Next, the program computes the *value of clairvoyance* for each feature that has not yet been reported to the system [13]; and displays those features with the highest scores to the user. To avoid a combinatorial explosion, the value-of-clairvoyance computation employs the assumption that at most one additional feature will be reported by the user. Then, the user observes one or more of these features, and the process iterates. Case evaluation terminates when the user is satisfied with the differential diagnosis that has been determined. Because patient preferences vary, the utilities represented in Pathfinder are used only to suggest additional features for observation; Pathfinder does not recommend treatment.

The Pathfinder project began in 1983. Early versions of the program employed the decision-theoretic model represented by the influence diagram in Figure 1(a). Specifically, researchers on the Pathfinder team assumed that diseases were mutually exclusive and exhaustive, and that all features were conditionally independent, given disease. The assumption that diseases were mutually exclusive was appropriate, because co-occurring diseases almost always appear in different lymph nodes or in different regions of the same lymph node. Also, the large scope of Pathfinder made reasonable the assumption that the set of diseases was exhaustive. The assumption of global conditional independence, however, was inaccurate. For example, given certain diseases, finding that follicles are abundant in the tissue section increases greatly the chances that sinuses in the interfollicular areas will be partially or completely destroyed. Thus, in 1986, the expert on the project, Dr. Bharat Nathwani, and I attempted to represent explicitly the dependencies among features in the lymph-node domain, by constructing an influence diagram of the form shown in Figure 1(b).

The expert provided utilities that served as a surrogate for patient preference. The process of encoding utilities was tedious but manageable with available decision-analytic techniques. The expert, however, could not construct the probabilistic portion of the influence diagram. Because of the wide scope of the domain, the expert was uncomfortable assessing conditional dependencies among some features. Through many discussions with the expert, I isolated his difficulties with construction, and developed the similarity-network representation to overcome these difficulties. Given this representation, the expert was able to construct the influence diagram for Pathfinder shown in Figure 2. In addition, using some of the insights that motivated the creation of similarity networks, I developed the partition representation to facilitate the probability assessment for the influence diagram. This representation decreased the number of probability assessments required to construct a joint distribution for the lymph-node domain by more than a factor of five. Together, the similarity-network and partition representations produced a new version of Pathfinder whose

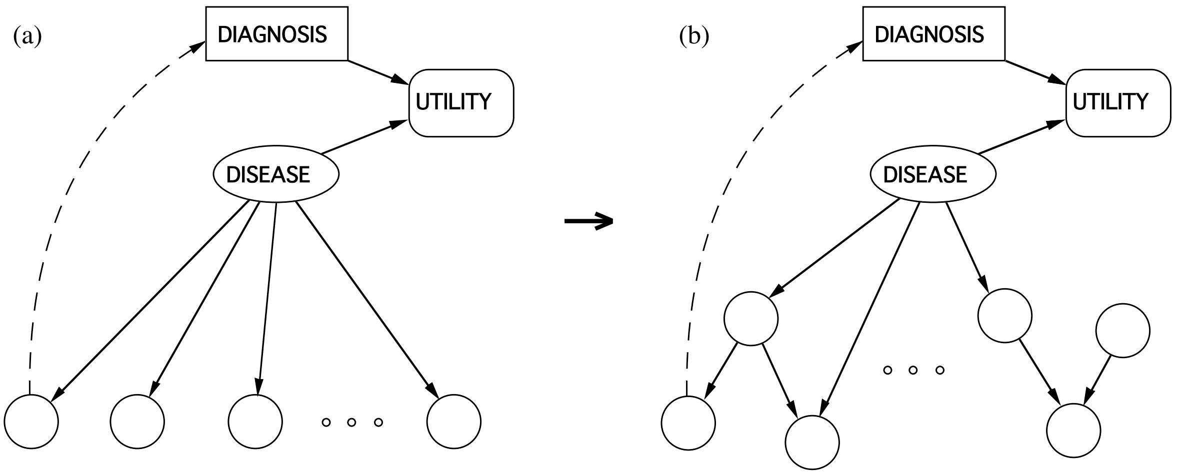

Figure 1: Two schematic influence diagrams for Pathfinder.

In both influence diagrams, the chance node DISEASE represents a set of mutually exclusive and exhaustive diseases. The decision node DIAGNOSIS represents all possible statements of the form, "the patient has disease x." The node UTILITY represents a patient's preferences for all disease–diagnosis combinations. The distinctions or features relevant to the diagnosis of disease are represented by the unlabeled chance nodes below DISEASE. The dotted arc represents the fact that some but not necessarily all features are observed at the time a diagnosis is rendered. In (a), all features are conditionally independent, given DISEASE. In (b), several features are conditionally dependent, given DISEASE. We required a similarity network to construct this more complex influence diagram.

diagnostic accuracy was superior to that of previous versions [8].

3 Extensions to the Influence-Diagram Representation

Influence diagrams that contain no decision nodes and no utility node are called *knowledge maps* [14], *belief networks* [19], or *probabilistic influence diagrams* [23]. Figure 3 illustrates the knowledge map corresponding to the schematic influence diagram for Pathfinder shown in Figure 1(b). In this knowledge map, the chance node DISEASE contains many mutually exclusive and exhaustive diseases. Furthermore, this node is not conditioned by any of the remaining nodes. Knowledge maps of this form are seen commonly in problems of diagnosis in which a single disease or *fault* is present.

A similarity network is a tool for building large and complex knowledge maps that have this special form. The disease node or—more generally—the *distinguished node* is the center of attention for the construction of a similarity network. The components of a similarity network include a *similarity graph* and a collection of *local knowledge maps*. Each node in a similarity graph represents an instance of the distinguished node, called a *hypothesis*. Edges in a similarity graph connect hypotheses that are similar or that are likely to be confused with one another by a user of the expert system. A local knowledge map is associated with each edge in the similarity graph. A local knowledge map for the edge between hypotheses h_i and h_j is a knowledge map constructed under the assumption that only h_i and h_j are possible. That is, a local knowledge map for hypotheses h_i and h_j is a knowledge map for discriminating only those two hypotheses. By constructing local knowledge maps, a person can concentrate on one manageable portion of the modeling task at a time.

Given a similarity network, we can construct a *global knowledge map* for the entire domain through simple graph manipulations on the local knowledge maps. The construction is *sound* in the sense that we can derive the assertions of conditional independence and dependence in the global knowledge map from the axioms of probability and the assertions of conditional independence and dependence in the local knowledge maps. Also, a simple algorithm exists for verifying that the assertions in the local knowledge maps are *consistent*. Thus, the global knowledge map constructed from a similarity network faithfully represents a person's assertions. We say that the global knowledge map is *valid*. In addition, the construction of the global knowledge map is *exhaustive* in the sense that any feature that is relevant to discrimination of the set of hypotheses as a whole will appear in the global knowledge map.

Similarity networks represent two forms of conditional independence called *subset independence* and *hypothesis-specific independence*, neither of which is represented conveniently in a knowledge map. In Section 4.6, we examine these forms of conditional independence, and discuss the problems posed by their representation in knowledge maps. Similarity networks take advantage of these forms of independence to decompose the construction of a knowledge map in much the same way that knowledge maps take advantage of ordinary conditional independence to decompose the construction of a joint probability distribution.

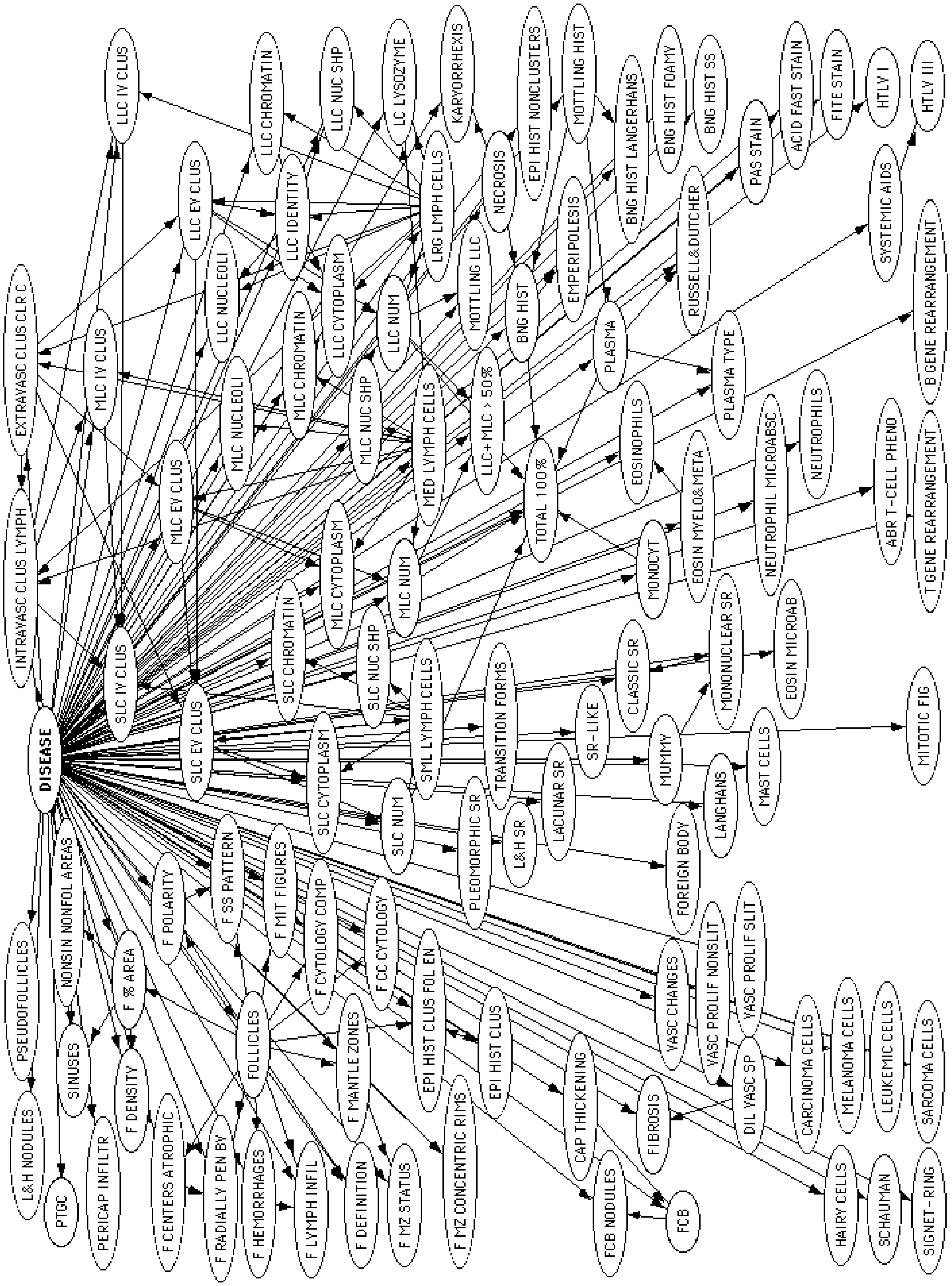

Figure 2: The Pathfinder influence diagram.

The influence diagram represents over 100 features that are relevant to diagnosis. The node DISEASE contains over 60 lymph-node diseases. The nodes DIAGNOSIS and UTILITY are not shown.

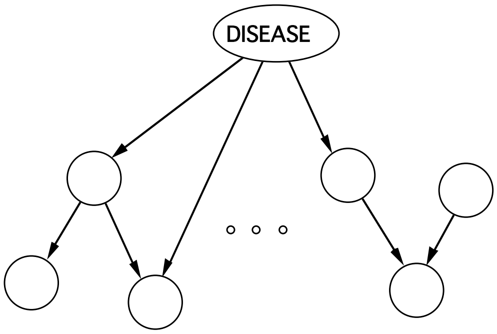

Figure 3: A schematic knowledge map for Pathfinder.
This influence diagram is the probabilistic part of the influence diagram shown in Figure 1(b). Influence diagrams containing only chance nodes and conditioning arcs are called *knowledge maps*. The chance node DISEASE contains many mutually exclusive and exhaustive instances, and is not conditioned by any of the remaining nodes. Similarity networks are suitable for building knowledge maps of this form.

In particular, similarity networks exploit people's ability to make judgments of subset independence and hypothesis-specific independence, without assessing the probabilities that underlie such judgments.

We can also use assertions of subset independence and hypothesis-specific independence to simplify the assessment of probabilities associated with a knowledge map. These assertions of conditional independence, as they are represented in a similarity network, simplify assessment somewhat. In Section 5, however, we examine the *partition*, a representation that exploits more fully these independencies for assessment. Figure 4 summarizes the roles of the partition, similarity network, and knowledge map in the construction of a joint probability distribution.

4 Similarity Networks: The Construction of a Knowledge Map

In this and the following section, we use the similarity-network and partition representations to construct and assess a knowledge map for a small medical expert system. The purpose of this exercise is to illustrate the basic concepts and techniques underlying these representations, and to demonstrate some of the advantages their use. Because the example is small, however, the full power of these representations for simplifying knowledge acquisition cannot be demonstrated. In [8], I describe highlights of this approach to knowledge acquisition applied to Pathfinder. There, the power of these representations is illustrated more fully.

The medical example that we examine is real, but it has been simplified for purposes of presentation. Dr. Harold Lehmann served as the expert for the domain. The figures in the chapter were generated by SimNet, an implementation on the Macintosh computer of the similarity-network and partition representations.

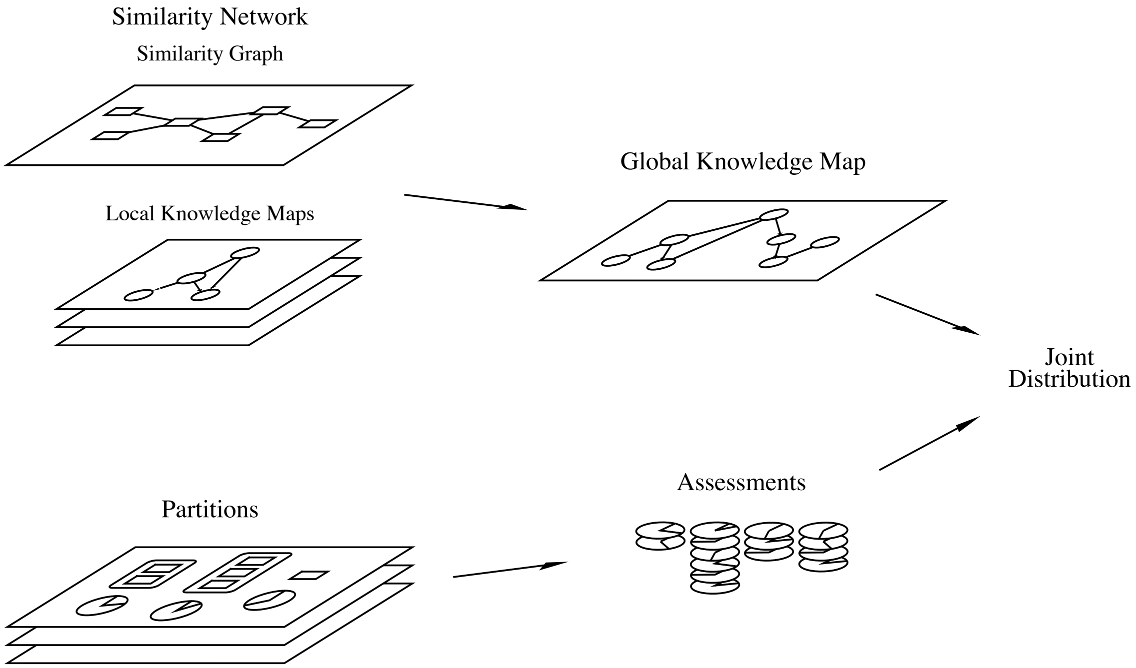

Figure 4: Decomposition of a joint probability distribution.

A joint distribution can be constructed from a global knowledge map and a set of assessments for each node in the map. The knowledge map itself can be constructed from a similarity network, consisting of a similarity graph and a collection of local knowledge maps. The assessments for each node in the knowledge map can be further decomposed using partitions.

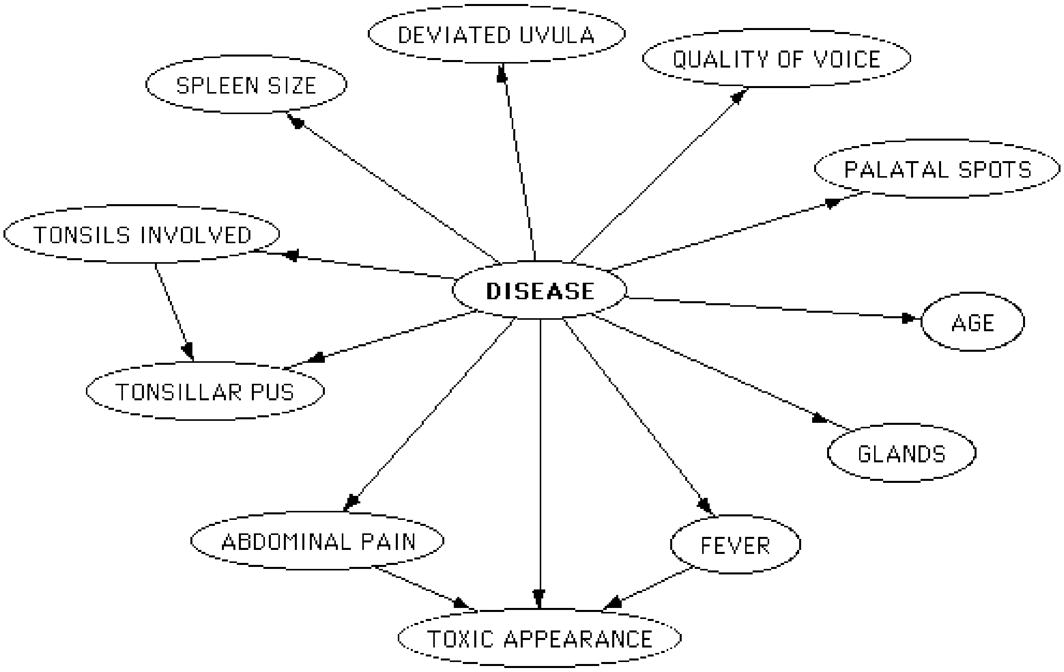

Figure 5: A knowledge map for sore throat.

This knowledge map describes the diagnostic dilemma for a patient between 5 and 18 years of age who comes to an emergency room with a severe sore throat. The node DISEASE represents the mutually exclusive and exhaustive causes of sore throat: VIRAL PHARYNGITIS, STREP THROAT, MONONUCLEOSIS, TONSILLAR CELLULITIS, and PERITONSILLAR ABSCESS. This node is the focus for the composition of a similarity graph. The remaining nodes represent evidence relevant to the diagnosis of the patient's disease.

Throughout the example, I will distinguish between the construction of a knowledge map, similarity network, or partition by a *person* and the construction of these representations by an *algorithm*. In particular, the terms *to compose* and *to construct* will refer to situations where a person and an algorithm generate a representation, respectively.

4.1 Composition of a Similarity Network

Suppose a patient between 5 and 18 years of age comes to an emergency room complaining of severe sore throat. A knowledge map for this situation is illustrated in Figure 5. The chance node DISEASE represents the causes of sore throat: VIRAL PHARYNGITIS, STREP THROAT, MONONUCLEOSIS, TONSILLAR CELLULITIS, and PERITONSILLAR ABSCESS. We assume that these diseases are mutually exclusive and exhaustive. The remaining nodes represent evidence relevant to the diagnosis of the patient's disease. We now discuss how to construct this knowledge map using a similarity network.

The focus for the composition of the similarity network is the *distinguished node* or *distinguished variable*. For medical domains, the distinguished variable represents a set of

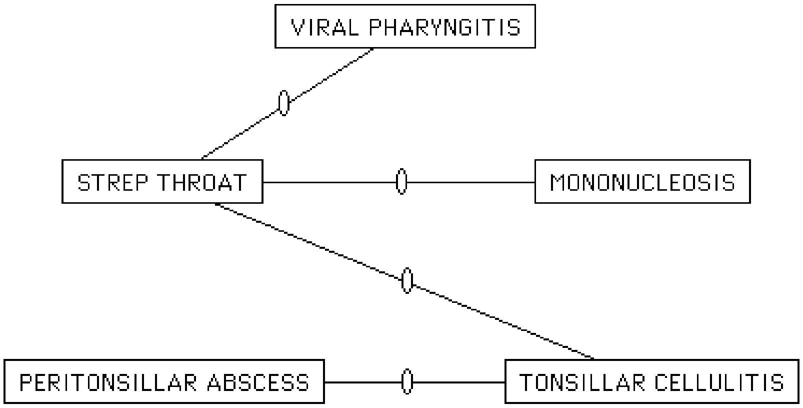

Figure 6: A similarity graph for sore throat.

The nodes in the graph represent the possible causes of sore throat. Undirected edges connect diseases that are similar. Although this graph is a tree (i.e., there is exactly one path between any two nodes in the graph), in general, similarity graphs can contain cycles.

mutually exclusive and exhaustive diseases. In general, we refer to the mutually exclusive exhaustive instances of this variable as *hypotheses.*

A similarity network consists of a *similarity graph* and a collection of *local knowledge maps.* To compose a similarity network, we first compose the similarity graph. The nodes in the similarity graph correspond to hypotheses of the distinguished node. Informally, the edges in the similarity graph connect hypotheses that are similar. We shall discuss soon the precise meaning of edges in a similarity graph. The similarity graph for sore throat is shown in Figure 6.

Next, we compose a local knowledge map for each pair of hypotheses that is connected in the similarity graph. To compose a local knowledge map for the hypothesis pair h_i and h_j, we imagine that one of these two hypotheses is true. Given this supposition, we compose a knowledge map consisting of the distinguished node—whose instances are restricted to h_i and h_j—and those *nondistinguished nodes* that are relevant to the discrimination of these hypotheses. Formally, we omit a node from the local knowledge map if and only if the node would be disconnected from the distinguished node (i.e., there would be no path between the node and the distinguished node) if we included it in the map.

Figure 7 shows the local knowledge map for the edge between TONSILLAR CELLULITIS and PERITONSILLAR ABSCESS in the similarity graph. The node at the top of the knowledge map represents the distinguished variable restricted to these two diseases. The nondistinguished nodes in the local map represent the features or disease findings that are relevant to the discrimination of the diseases TONSILLAR CELLULITIS and PERITONSILLAR ABSCESS. Notice that there are no arcs among these nondistinguished nodes. The missing arcs represent the assertion that, given that the patient has either TONSILLAR CELLULITIS or PERITONSILLAR ABSCESS, all findings in the map are independent. Also note that there are fewer findings in this local map than in the knowledge map for the entire domain (Fig-

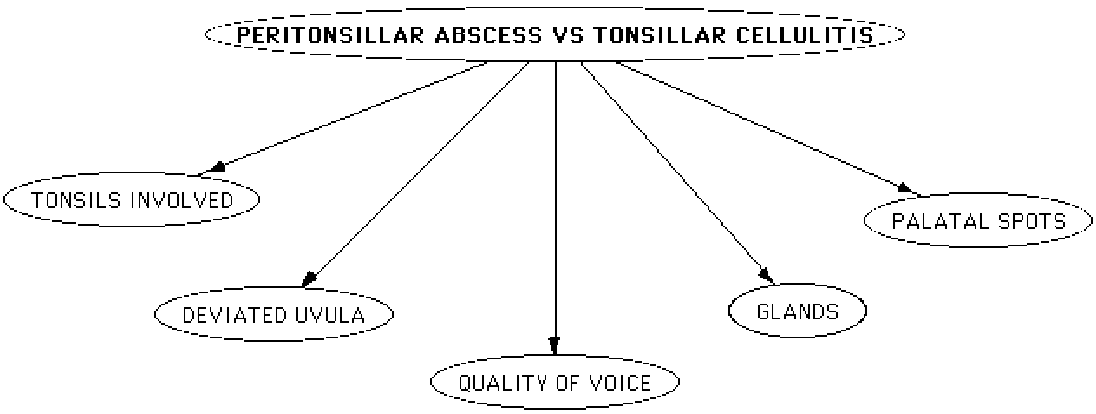

Figure 7: A local knowledge map.

In this knowledge map for the edge between PERITONSILLAR ABSCESS and TONSILLAR CELLULITIS in the similarity graph, the features exhibit mutual conditional independence. Only findings that are relevant to the discrimination of the two diseases are included in the map. Because the two diseases are similar, the knowledge map contains fewer nodes than does the knowledge map for the entire domain.

ure 5). This observation tends to be true, in general, because the diseases associated with local knowledge maps are similar.

Figure 8 shows the local knowledge map for the edge between STREP THROAT and VIRAL PHARYNGITIS in the similarity graph. Again, the map contains fewer features than does the knowledge map for the sore-throat domain as a whole. Now, however, some of the disease findings are conditionally dependent. The arc from TONSILS INVOLVED to TONSILLAR PUS reflects the expert's assertion that the probability of seeing pus on a patient's tonsils depends on whether the disease involves one tonsil, both tonsils, or neither tonsil. The arcs from FEVER and ABDOMINAL PAIN to TOXIC APPEARANCE reflect the observation that a patient is more likely to present with a toxic appearance if the patient has abdominal pain or a high fever. Although FEVER is relevant to the discrimination of STREP THROAT and VIRAL PHARYNGITIS indirectly through its effect on TOXIC APPEARANCE, the missing arc from the disease node to FEVER represents the assertion that temperature alone is not relevant to the discrimination of the two diseases.

The feature PALATAL SPOTS, among other features, appears in both of the local knowledge maps that we have examined. In the local knowledge map for TONSILLAR CELLULITIS and PERITONSILLAR ABSCESS, the instances of this feature are ABSENT and PRESENT. The same instances are associated with this feature in the local knowledge map for STREP THROAT and VIRAL PHARYNGITIS. In general, the instances associated with a feature in one local knowledge map must be identical to the instances associated with that feature is all other local knowledge maps.

Figure 9 shows the local knowledge maps for the remaining two edges in the similarity graph. Again, there are fewer features in each of these maps than there are in the knowledge

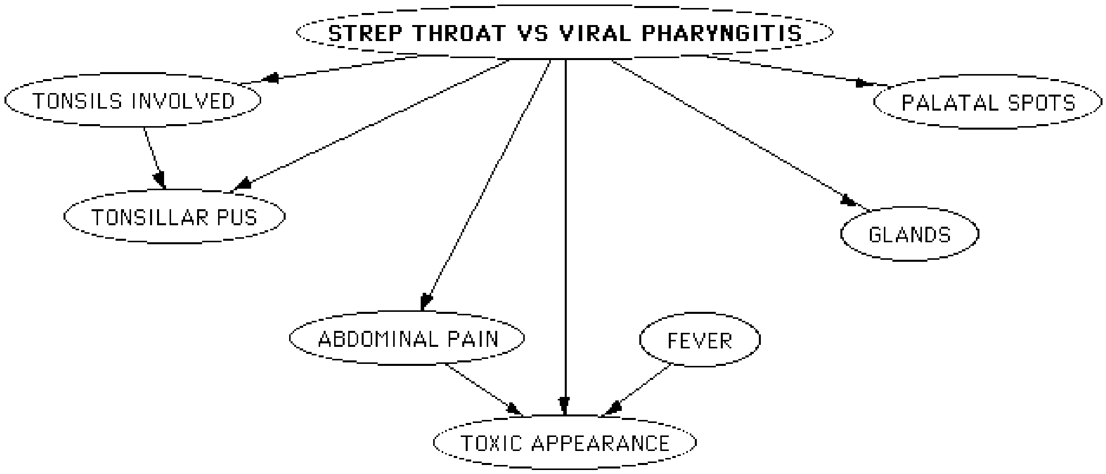

Figure 8: A local knowledge map with dependencies.
In this knowledge map for the edge between STREP THROAT and VIRAL PHARYNGITIS in the similarity graph, the feature TONSILLAR PUS is conditionally dependent on TONSILS INVOLVED, and the feature TOXIC APPEARANCE is conditionally dependent on ABDOMINAL PAIN and FEVER.

map for the entire domain.

In SimNet, the local knowledge map for each edge in a similarity graph is accessed via the oval on the edge (see Figure 6). Specifically, by clicking on an edge's oval, the user brings up a window in which the local knowledge map associated with that edge can be created or modified.

The formal criteria for drawing edges in a similarity graph are that (1) we connect two diseases only if we can compose a local knowledge map for the disease pair, and (2) the similarity graph must be connected—that is, there must be a path between any two nodes in the graph. There is no formal requirement that connected diseases be similar. As we have seen in this example, however, local knowledge maps for pairs of similar diseases tend to exclude many of the features that discriminate the set of diseases as a whole. Thus, an expert can simplify greatly his task of composing the local knowledge maps by connecting only similar diseases in the similarity graph (provided the graph remains connected). Indeed, in practice, I have found it useful to ask experts to draw a similarity graph using only considerations of similarity; I introduce the concept of a local knowledge map after the similarity graph is composed.

To simplify composition further, an expert may choose not to compose certain local knowledge maps, even if he believes that he can compose them. There is no need to build more knowledge maps than those that are required to create a connected similarity graph. Alternatively, an expert may choose to build a multiply connected similarity graph.[2] Cycles in the similarity graph provide additional opportunities to check the self-consistency of a user's knowledge.

---

[2] A multiply connected graph contains more than one path between some node pairs.

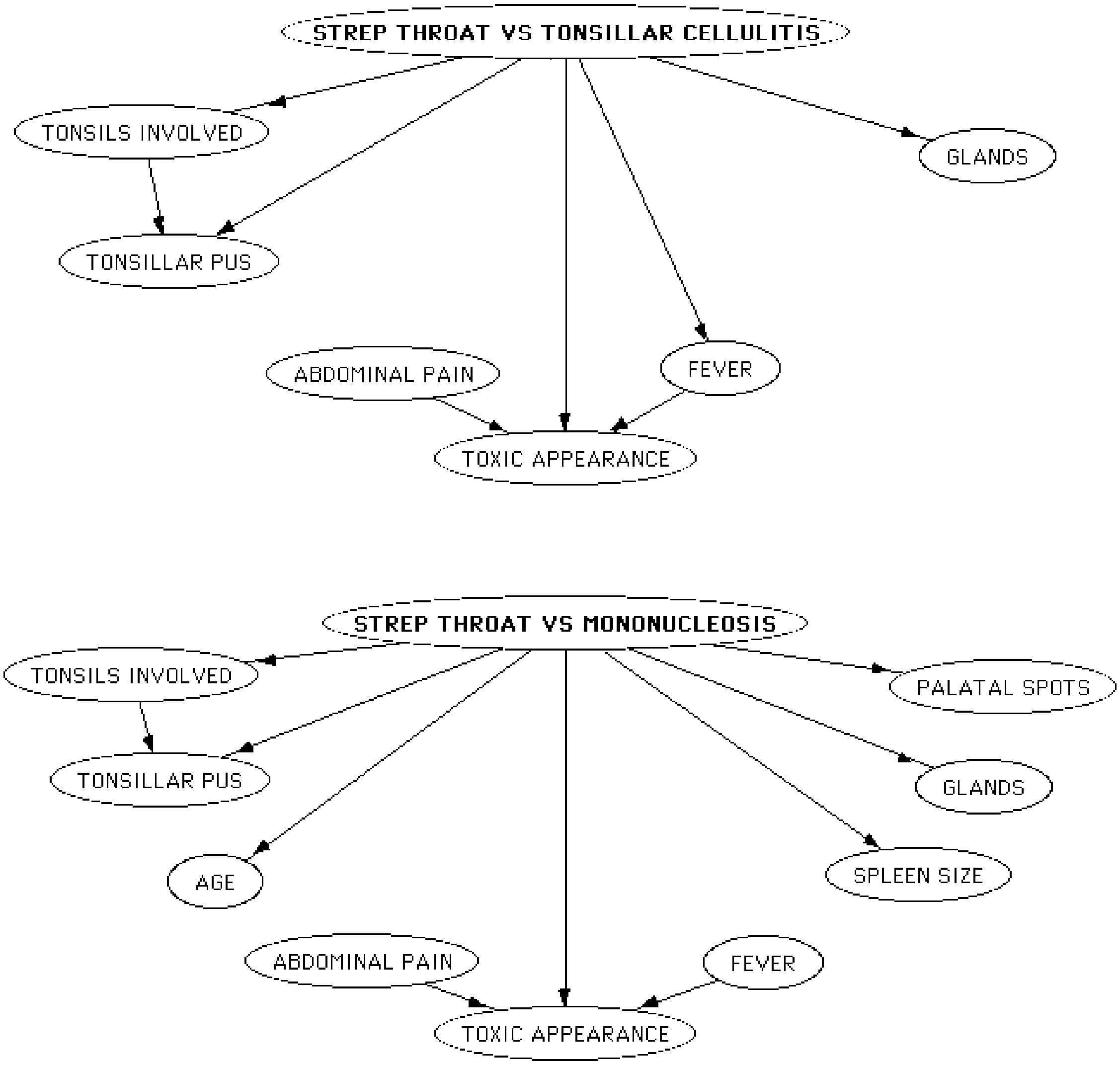

Figure 9: Local knowledge maps for the other two edges in the similarity graph. The knowledge map contains fewer nodes than does the knowledge map for the entire domain.

Given the similarity network that we have *composed*, we can now *construct* the knowledge map for the full sore-throat problem, called the *global knowledge map.* Specifically, we construct the global knowledge map by forming the *graph union* of the local knowledge maps in the similarity network. The operation of graph union is straightforward. The nodes in the graph union of a set of graphs is the simple union of the nodes in the individual graphs. Similarly, the arcs in the graph union of a set of graphs is the simple union of the arcs in the individual graphs. That is, a node (or arc) appears in the graph union, if and only if there is such a node (or arc) in at least one of the individual graphs. The set of instances associated with a feature in the global knowledge map is the same as the set of instances associated with that feature in each of the local knowledge maps. The set of instances associated with the disease or distinguished node in the global knowledge map is the union of all diseases or hypotheses in the similarity graph.

The global knowledge map for sore throat was shown in Figure 5. The node QUALITY OF VOICE, for example, appears in the global knowledge map because it appears in the local knowledge map for PERITONSILLAR ABSCESS and TONSILLAR CELLULITIS. The arc from DISEASE to ABDOMINAL PAIN appears in the global knowledge map because it is present in the local map for STREP THROAT and VIRAL PHARYNGITIS.

4.2 A Valid Knowledge Map

Under certain conditions, the construction of the global knowledge map from the similarity network is *sound.* That is, any joint distribution that satisfies the assertions of conditional independence implied by the local knowledge maps also satisfies the assertions of conditional independence implied by the global knowledge map. In addition, we can verify easily that the set of assertions implied by a collection of local knowledge maps is *consistent.* Given these two results, we know that the global knowledge map constructed from a similarity network accurately and coherently reflects the assertions of conditional independence of the person who composes that network. That is, the global knowledge map is *valid.*

These results apply to minimal knowledge maps as well. A *minimal* knowledge map is one in which no arc can be removed without contradicting one or more of the assertions of conditional independence made by the expert. Thus, a minimal knowledge map can represent both assertions of conditional independence and conditional dependence. We can check whether or not a collection of minimal local knowledge maps is consistent. In addition, the construction of a minimal global knowledge map is sound. To be more precise, suppose we have a similarity network in which all the local knowledge maps are minimal. Also, suppose we interpret the global knowledge map constructed from that similarity network to be a minimal knowledge map. Again, under certain conditions, any distribution that satisfies the assertions of conditional dependence implied by the local knowledge maps must also satisfy the assertions of conditional dependence implied by the global knowledge map. That is, if the local knowledge maps are minimal, then the global knowledge map is minimal as well.

The soundness result holds for those similarity networks in which the following constraints are satisfied:

1. The instances of the distinguished node (hypotheses) are mutually exclusive and exhaustive.
2. The similarity graph is connected.
3. Local knowledge maps are constructed for pairs of diseases.
4. The global knowledge map that is equal to the graph union of the local knowledge maps contains no directed cycles.
5. There are no arcs pointing to the distinguished node in any local knowledge map.
6. The joint distribution for the domain is *strictly positive* (i.e., there are no probabilities in the distribution that are equal to zero).

These constraints are sufficient conditions for soundness; they are not necessary conditions. In [8], I discuss how the constraints might be relaxed. For example, the soundness result applies as well to similarity networks where local knowledge maps are composed for hypothesis sets of arbitrary size. To build such a similarity network, we simply replace the similarity graph by a similarity hypergraph. A hypergraph consists of nodes and edges that connect sets of nodes. To ensure that the global knowledge map constructed from such a network is sound, we must replace only the constraint that the similarity graph be connected, using instead the constraint that the similarity hypergraph be connected. Also, we can use the representation in some situations where hypotheses are not mutually exclusive, and where the disease node is conditioned by other nodes.

4.3 An Exhaustive Construction

The construction of a global knowledge map from a similarity network is also *exhaustive*. That is, any feature that is relevant to the discrimination of the hypothesis set as a whole must appear in some local knowledge map, and hence in the global knowledge map. In [8, Chapter 3], I prove this result. Together, the soundness, consistency, and exhaustiveness results make the similarity-network representation extremely useful for knowledge acquisition.

4.4 Advantages of Using Similarity

The concept of similarity does not appear in the conditions for soundness, consistency, or exhaustiveness. Nonetheless, there are important advantages of composing knowledge maps for diseases that are similar. As we discussed in Section 4.1, one such advantage is that local knowledge maps for pairs of similar diseases tend to be small. If this fact were the only advantage of similarity networks for constructing a knowledge map, however, there would be no point in using them. If any feature appeared in more than one local knowledge map, we would be duplicating our efforts of composition, regardless of the size of the maps. In this section, we consider another advantage of composing local knowledge maps for pairs of similar diseases that makes the similarity-network representation a valuable tool for constructing knowledge maps.

As mentioned in Section 2, the Pathfinder expert could not compose directly the global knowledge map for the lymph-node domain. Specifically, he could not assess dependencies among certain features in the domain. When asked questions of the form

> Given any disease, does observing feature x change your belief that you will observe feature y?

the expert sometimes would reply

> I've never thought about these two features at the same time before. Feature x is relevant to the discrimination of a particular set of diseases. Feature y, on the other hand, is relevant to the discrimination of a different set of diseases. These two sets of diseases do not overlap, and I never confuse the first set of diseases with the second.

The expert had detailed knowledge about diagnosis in multiple *small worlds* or subsets of similar disease within the lymph-node domain. If two or more features were relevant to the same small world, the expert had no trouble assessing dependencies among those features. If two or more features were not relevant to a common small world, however, he could not evaluate the dependencies among them.

This observation is not that surprising. When an expert pathologist looks at a tissue section under the microscope, he immediately focuses on a relatively small set of diseases. He then expends the majority of his conscious effort looking for features that discriminate among these diseases. Almost by definition, these initial disease sets will consist of diseases that are similar to one another. In fact, many of the experts who experimented with the similarity-network representation independently adopted the following operational definition for similarity when composing the similarity graph:

> Two diseases are similar if and only if they are likely to be confused with each other in practice.

It is reasonable to expect that knowledge about dependencies among features relevant to these small worlds are more available to an expert, because he spends most of his time thinking explicitly about such features and the relationships among them. In extreme cases, such as the situation described in the previous paragraph, knowledge about the discrimination of highly dissimilar diseases may be unavailable to the expert.

The similarity-network representation is an ideal tool for combining knowledge about these small worlds into a coherent whole. Indeed, the representation was developed in direct response to the assessment predicament of the lymph-node expert described two paragraphs earlier. Using a similarity network, an expert can assess dependencies among features that are relevant only to pairs of similar diseases. Given the soundness, consistency, and exhaustiveness results, we can combine the knowledge about such dependencies, recorded in the local knowledge maps, to create a global knowledge map that faithfully represents the assertions of an expert for his domain. This knowledge map, in turn, endows an expert system with the ability to discriminate among any set of diseases, whether they are similar or dissimilar. Consider again, for example, the assessment predicament of the expert. Expressed in terms of a similarity network, the features x and y never appear in the same local knowledge map. Thus, given the procedure for constructing the global knowledge map and the soundness result, we know that x and y must be conditionally independent, given disease. This observation was apparent to neither the expert nor me before the similarity-network representation and its theory were developed.

Even in situations where an expert can compose a global knowledge map, a similarity network should prove useful for knowledge acquisition. Specifically, by composing local knowledge maps for pairs of similar diseases, the expert can use a similarity network to focus his attention on precisely those diagnostic subproblems with which he is familiar. The expert may thereby increase the quality of the knowledge he provides. In constructing the Pathfinder knowledge map, for example, local knowledge maps helped the expert avoid errors of omission in describing features relevant to lymph-node diagnosis.

4.5 Soundness and Consistency: Theoretical Considerations

At first glance, the soundness result may seem trivial. Consider, for example, the similarity network for three hypotheses h_1, h_2, and h_3, and three nondistinguished variables x, y, and z, shown in Figure 10(a). The network contains two local knowledge maps for the hypothesis pairs $\{h_1, h_2\}$ and $\{h_2, h_3\}$. In the local knowledge map for h_1 vs h_2, all nondistinguished variables are connected to the distinguished node, and, therefore, are included in the map. Similarly, all nondistinguished nodes are included in the local knowledge map for h_2 and h_3.

The global knowledge map constructed from the two local maps in the similarity network is shown in Figure 10(b). The global knowledge map asserts that z is independent of x and y, given h. This assertion is logically implied by the assertions of conditional independence in the local knowledge maps; hence, the construction of the global knowledge map is sound. To see this fact, let us consider the local knowledge map for h_1 and h_2. From the definition of missing arcs in a knowledge map, we know that z is independent of x and y, given h Formally,

$$p\,(z|x_i, y_j, h_k, \{h_1, h_2\}, \xi) = p\,(z|h_k, \{h_1, h_2\}, \xi) \qquad (1)$$

where x_i and y_j range over the instances of variables x and y, and where h_k is equal to h_1 or h_2. The first term refers to the probability distribution over z, given x_i, y_j, and h_k, and given the state of knowledge $\{h_1, h_2\}$ and ξ. The second term refers to a similar distribution. In both expressions, the symbol ξ denotes the state of knowledge that an expert brings to bear on the global problem. The set $\{h_1, h_2\}$, which conditions both probabilities, denotes the disjunction of h_1 and h_2. The disjunction appears in both expressions because, by definition, it is part of the state of knowledge of the local knowledge map. Now the hypothesis h_1 alone and the hypothesis h_2 alone, logically imply the disjunction of h_1 and h_2. Consequently, we

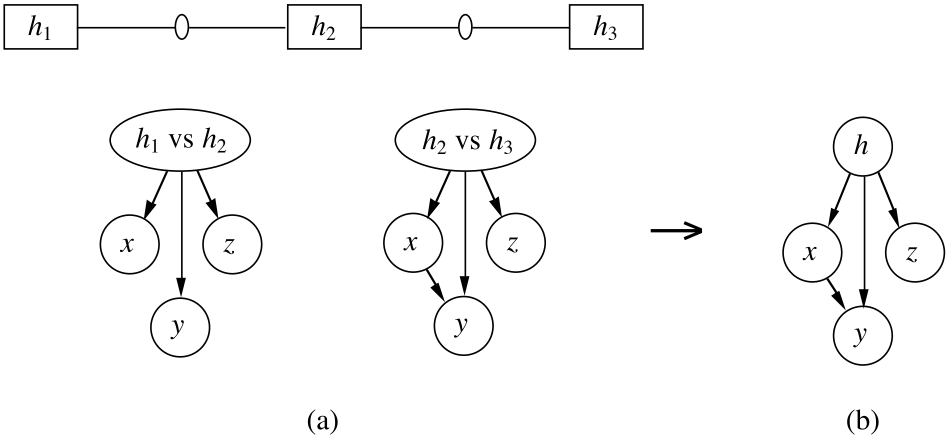

Figure 10: The construction of a valid global knowledge map.
(a) A similarity network for three hypotheses h_1, h_2, and h_3, and three nondistinguished variables x, y, and z. In the network, there are two local knowledge maps for the hypothesis pairs $\{h_1, h_2\}$ and $\{h_2, h_3\}$. Both local knowledge maps contain all the nondistinguished nodes. (b) The global knowledge map constructed from the similarity network. The assertions of conditional independence in the global knowledge map are logically implied by such assertions of conditional independence in the local knowledge maps. Hence, the construction of the global knowledge map is sound. The construction remains sound when the local knowledge maps and the global knowledge map are minimal. Furthermore, the similarity network is consistent, and thus the global knowledge map is valid.

can omit the disjunction from both sides of Equation 1 to obtain

$$p\,(z|x_i, y_j, h_k, \xi) = p\,(z|h_k, \xi) \tag{2}$$

where x_i, y_j, and h_k range over the same instances as in Equation 1. Thus, z is independent of x and y given both h_1 and h_2 alone. Similarly, from the local knowledge map for h_2 and h_3, we can show that z is independent of x and y, given h_2 and h_3 alone. Combining the observations for both local knowledge maps, we see that z is independent of x and y given any hypothesis of h.

Furthermore, suppose the local knowledge maps in Figure 10(a) are minimal. From the local knowledge map for h_2 and h_3, we know x and y are dependent, given that h_2 or h_3 is true. (If x and y were independent, given h_2 and given h_3, we could remove the arc from x to y, contradicting the minimality of the local knowledge map.) It follows that x and y must be dependent, given h, and there must be an arc between x and y in the global knowledge map. We can also show that each arc emanating from h must also be present in the global knowledge map. Thus, the soundness result also holds when the knowledge maps as minimal.

Finally, consider any probability distribution where (1) x, y, and z are conditionally independent, given h_1 and h_2, (2) only z is conditionally independent of x and y, given h_3, and (3) x, y, and z are dependent on h. This distribution satisfies all the assertions of conditional independence and dependence implied by the two local knowledge maps. Consequently, the similarity network in in Figure 10 is consistent, and the global knowledge map in the figure is valid.

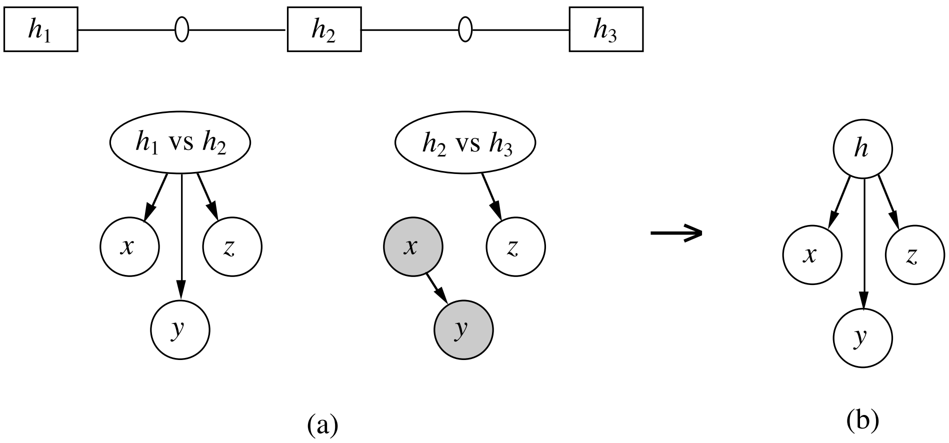

Figure 11: The construction of an invalid global knowledge map.
(a) A similarity network for three hypotheses and three nondistinguished variables. In the local knowledge map for the hypothesis set $\{h_2, h_3\}$, the nodes x and y (shaded) are not connected to h, and are omitted from the local knowledge map. (b) The global knowledge map constructed from the similarity network. If the local knowledge maps are minimal, we know that x and y are dependent given h_2 or h_3. Because these nodes have been omitted from the local knowledge map, however, this dependency is not recorded in the global knowledge map. This problem occurs because the similarity network is inconsistent.

In general, if each local knowledge map contains the same set of nondistinguished nodes, the proof of soundness is straightforward. When nodes are omitted from local knowledge maps, however, the proof is not so simple. Consider, for example, the similarity network for three hypotheses and three nondistinguished variables shown in Figure 11(a). The local knowledge map for h_1 and h_2 is identical to the corresponding map in Figure 10(a), whereas the the local knowledge map for h_2 and h_3 is different. In particular, the nodes x and y are disconnected from h in the local knowledge map for h_2 and h_3. Suppose the local knowledge map for h_2 and h_3 is minimal. In this case, we know that x and y are dependent, given h_2 or given h_3. Thus, x and y are dependent, given h. However, when we compose this local knowledge map using the procedure described previously, we omit the nodes x and y from the map (indicated by shading in the figure) because these nodes are disconnected from h. Consequently, the dependency does not get recorded in the global knowledge map, shown in Figure 11(b).

We could avoid this problem by including all nondistinguished variables in each map. Then, in this example, the dependency between x and y would be registered in the global knowledge map. This alternative, however, would destroy the benefits of the similarity-network representation discussed in Section 4.4.

Fortunately, we do not have to abandon the original procedure for composing local knowledge maps to guarantee soundness. As we shall see in the following paragraph, the difficulty in this example does not come from lack of soundness, but rather from the fact that the similarity network in Figure 11(a) is inconsistent. That is, there is no joint distribution over the variables h, x, y, and z that satisfies the conditional independence and dependence assertions implied by those local knowledge maps. Thus, the construction is sound, because we can derive any set of dependence and independence assertions from a contradiction, but the global knowledge map is invalid. When we make the similarity network in Figure 11 consistent, assuming there is a dependence between x and y in the local knowledge map for h_2 and h_3, we must add an arc from x to y in the local knowledge map for h_1 and h_2. Once this is done, the dependency between these nodes is registered appropriately in the global knowledge map. Consequently, the construction remains sound, and the global knowledge map becomes valid.

To see that the similarity network in Figure 11 is inconsistent, let us first examine the local knowledge map for h_2 and h_3, including in this map the nodes x and y. From the definition of missing arcs in a knowledge map, we get

$$p\,(y|x_k, h_l, \{h_2, h_3\}, \xi) = p\,(y|x_k, \{h_2, h_3\}, \xi) \tag{3}$$

where x_k ranges over the possible instances of the variable x, and where h_l is equal to h_2 or h_3. Now h_2 and h_3 alone logically imply the disjunction of h_2 and h_3. Therefore, we can remove this disjunction from the left-hand side of Equation 3 to obtain

$$p\,(y|x_k, h_l, \xi) = p\,(y|x_k, \{h_2, h_3\}, \xi) \tag{4}$$

where x_k and h_l have the same ranges as in Equation 3. Also, we have assumed that the local knowledge maps are minimal. Thus, x and y are dependent in the local knowledge map for h_2 and h_3, and we know that

$$p\,(y|x_i, \{h_2, h_3\}, \xi) \neq p\,(y|x_j, \{h_2, h_3\}, \xi) \tag{5}$$

for some instances $x_i \neq x_j$. From Equations 4 and 5, we get

$$p\,(y|x_i, h_l, \xi) \neq p\,(y|x_j, h_l, \xi) \tag{6}$$

for $h_l = h_2$ and $h_l = h_3$. Therefore, the local knowledge map for h_2 and h_3 dictates that x and y are dependent, given both h_2 and h_3 separately. The local knowledge map for h_1 and h_2, however, implies that x and y are independent, given h_1 and given, h_2. We thus obtain the contradiction that x and y are both conditionally independent and dependent, given h_2.

In general, if the members of a set of minimal local knowledge maps are consistent, and the other constraints discussed in Section 4.2 are satisfied, then arcs between nodes that are omitted from one local knowledge map must appear in other local knowledge maps. Thus, the arcs appear in the global knowledge map, and the construction of this map is sound [8]. We can extend this analysis easily to include nonminimal local knowledge maps. Also, there is a simple algorithm for testing the consistency of a similarity network [8]. We can use this algorithm to isolate and correct inconsistencies in a set of local knowledge maps. Consequently, we can omit nodes from local knowledge maps, and thereby retain the benefits of similarity networks for knowledge acquisition discussed in Section 4.4.

4.6 Assertions of Asymmetric Conditional Independence

A similarity network derives its power from its ability to represent assertions of conditional independence that are not conveniently represented in an ordinary knowledge map. In fact, a similarity network can represent two specific forms of such conditional independence.

To illustrate the first of these assertion types, suppose we have a variable h with mutually exclusive and exhaustive hypotheses $h_1, h_2, \ldots h_n$. Further, let $h_\subseteq$ denote a proper subset of the hypotheses of h. If x and h are independent, given that one of the elements of $h_\subseteq$ is true, we say that x is not relevant to $h_\subseteq$. Formally, a variable x *is not relevant to* the set $h_\subseteq$, given a state of knowledge ξ, if and only if

$$p\left(h_i | x_j, h_\subseteq, \xi\right) = p\left(h_i | h_\subseteq, \xi\right) \tag{7}$$

for all instances x_j of variable x, and for all hypotheses h_i in $h_\subseteq$. As in the previous section, the set $h_\subseteq$, which conditions both probabilities, denotes the disjunction of its elements. We call the form of conditional independence represented by Equation 7 *subset independence*. Using Bayes' theorem, we can derive an equivalent criterion for subset independence. In particular, we can show that a variable x is not relevant to the set of hypotheses $h_\subseteq$, given a state of knowledge ξ, if and only if

$$p(x|h_i, \xi) = p(x|h_j, \xi) \tag{8}$$

for all pairs $h_i, h_j \in h_\subseteq$. We shall return to Equation 8 when we discuss probability assessment in the following section.

Now suppose there is no arc from h to x in the local knowledge map for h_i and h_j. That is, suppose x and h are independent in the state of information $\{h_i, h_j\}$. By definition, we know that x is not relevant to $\{h_i, h_j\}$. Alternatively, suppose that x is omitted from this local knowledge map. From the definition of local knowledge map, we know that there would be no path from h to x, if x were included in the map. Consequently, x and h must be independent in the state of information $\{h_i, h_j\}$. Again, it follows that x is not relevant to $\{h_i, h_j\}$. In either case, using a similarity network, we can represent assertions of subset independence.

To illustrate the second form of conditional independence that we can encode in local knowledge maps, let us consider the similarity network in Figure 10. Using arguments similar to those in the previous section, we can show that x and y are dependent given h_3, but x and y are independent, given h_1 and h_2. Thus, a similarity network can represent assertions of conditional independence that are specific to individual hypotheses. We call this form of conditional independence *hypothesis-specific independence*.

Subset independence refers to relationships between the distinguished node and nondistinguished nodes. In contrast, hypothesis-specific independence refers to relationships among nondistinguished nodes. Nonetheless, both forms of independence are closely related in that they are *asymmetric*. In general, an assertion of conditional independence is asymmetric if it holds for only some instances of its variables. Assertions of subset independence and hypothesis-specific independence, in particular, hold for only proper subsets of the distinguished node.

As mentioned previously, we cannot easily encode assertions of asymmetric conditional independence in an ordinary knowledge map. In Section 5.4, we discuss this observation in detail.

5 Partitions: The Assessment of a Knowledge Map

Once the global knowledge map has been constructed, there are several alternative techniques for assessing the map. One approach is to assess directly conditional distributions for each variable in the global knowledge map. This approach is straightforward, but it does not take advantage of additional assertions of conditional independence represented in the similarity network. In this section, we examine how we can use assertions of subset independence to simplify assessment. In [8], I discuss how we can exploit both subset independence and hypothesis-specific independence to facilitate assessment.

5.1 Use of Similarity Networks for Assessment

There are two approaches available in SimNet for assessing a knowledge map that exploit assertions of subset independence. In one approach, only those assertions of subset independence represented by nodes missing from the local knowledge maps are employed. In another method, judgments of subset independence embodied by partitions are used. The first approach is illustrated in Figure 12 for the feature QUALITY OF VOICE. The rounded rectangle labeled with the feature name contains the mutually exclusive and exhaustive instances of the feature: NORMAL and MUFFLED. The two numbers under each disease are the probability distribution for the feature given that disease. For example, the probability that QUALITY OF VOICE is NORMAL, given STREP THROAT, is 0.9.

The black oval on the edge between PERITONSILLAR ABSCESS and TONSILLAR CELLULITIS reflects the fact that the feature QUALITY OF VOICE is present in the local knowledge map for the disease pair. Conversely, the white ovals on the remaining edges represent the fact that this feature is absent from the other local knowledge maps. As shown in the figure, when a feature is omitted from a local knowledge map, the conditional probability distributions on either side of an edge are equal. This observation follows from Equation 8 and the fact that any feature omitted from a local knowledge map cannot be relevant to the two diseases associated with that map. Consequently, for the feature QUALITY OF VOICE, we need to assess probability distributions given only PERITONSILLAR ABSCESS and TONSILLAR CELLULITIS. SimNet automatically propagates the probability distribution for TONSILLAR CELLULITIS throughout the remainder of the similarity graph.

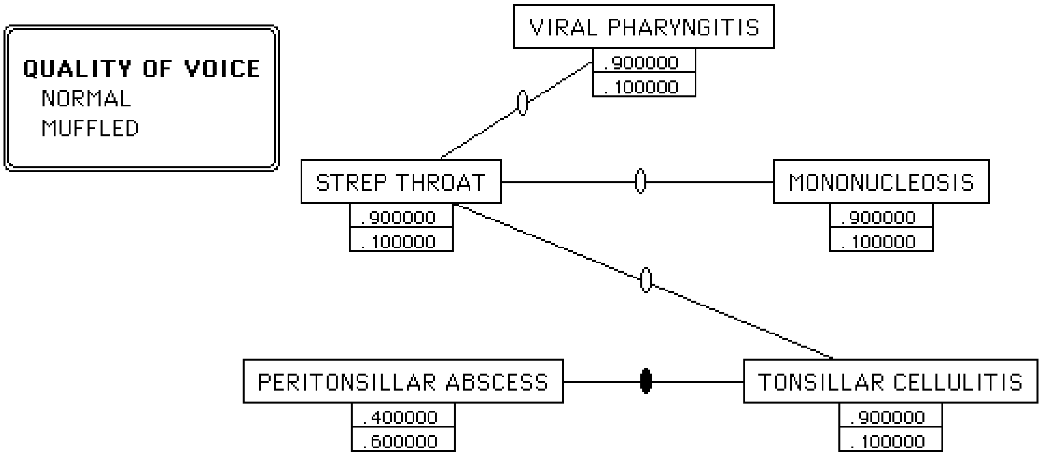

Figure 12: Probability assessment using a similarity network.

The probability distributions for the feature QUALITY OF VOICE given disease are shown. The rounded rectangle labeled with the feature name contains the mutually exclusive and exhaustive instances of the feature: NORMAL and MUFFLED. The numbers below each disease node are the probability distribution for QUALITY OF VOICE given that disease. The white ovals on the edges reflect the fact that the feature is absent in the corresponding local knowledge map. Conversely, the black oval signifies that the feature is present in the local knowledge map. Distributions bordering an edge with a white oval must be equal.

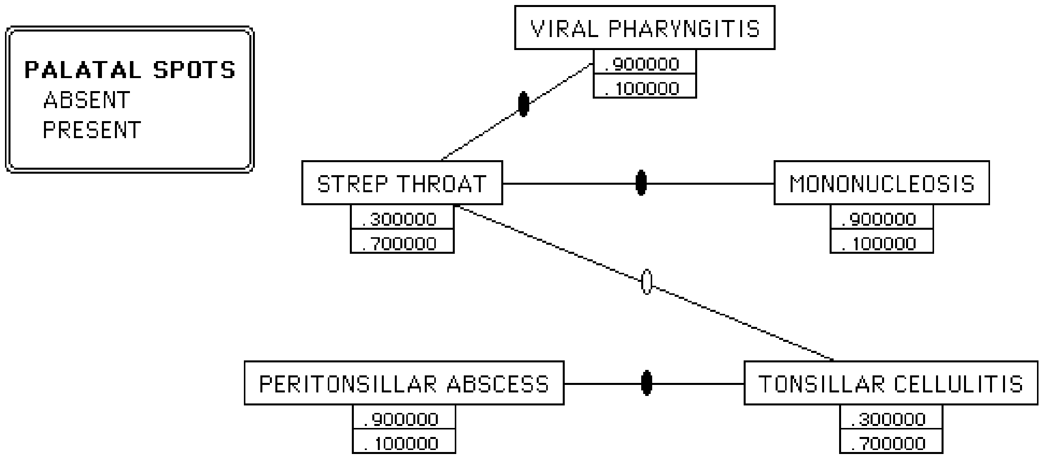

Figure 13: Hidden equivalence in a similarity network.

The assessment of the feature PALATAL SPOTS is shown. Although the distributions for VIRAL PHARYNGITIS, MONONUCLEOSIS, and PERITONSILLAR ABSCESS are equal, these equalities are hidden until the actual assessments are made, because there are no edges that directly connect any pair of these diseases in the similarity graph.

5.2 Use of Partitions for Assessment

A problem with this approach to assessment is illustrated in Figure 13. Specifically, the probability distributions for the feature PALATAL SPOTS given VIRAL PHARYNGITIS, MONONUCLEOSIS, and PERITONSILLAR ABSCESS are equal. Because we did not connect these diseases in the similarity graph, however, the equality of these distributions remains hidden until we assess the actual probabilities.

We can remedy this difficulty by composing a local knowledge map for every pair of diseases. For domains with more than just a few diseases or hypotheses, however, this alternative is impractical. Alternatively, we can compose a *partition* of the hypotheses for each nondistinguished variable to be assessed. In composing a partition, we place each hypothesis or instance of the distinguished variable into one and only one set. We place two or more hypotheses in the same set only if the nondistinguished variable associated with the partition *is not relevant to* those hypotheses in the set (see Equation 7). After composing the partition for a given nondistinguished variable, we assesses probability distributions for the variable given each hypothesis. Given Equation 8, however, we need to assess only one distribution for each set in the partition.

A partition for the feature PALATAL SPOTS is shown in Figure 14. In this partition, the possible causes of sore throat are divided into two groups: those diseases in which palatal spots are likely to be seen (STREP THROAT and TONSILLAR CELLULITIS), and those diseases for which palatal spots are not likely to be seen (MONONUCLEOSIS, PERITONSILLAR ABSCESS, and VIRAL PHARYNGITIS). The partition reflects the assertions that (1) the feature PALATAL SPOTS *is not relevant to* the disease pair STREP THROAT and TONSILLAR

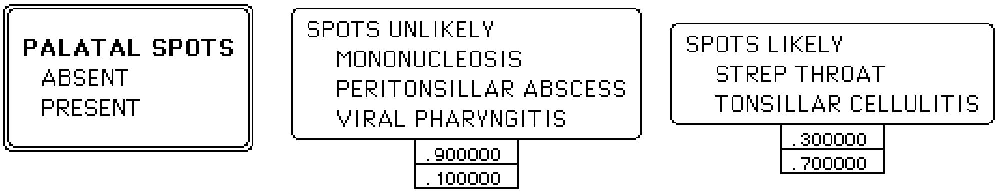

Figure 14: Assessment of probabilities using a partition.
The partition contains two sets of diseases labeled SPOTS UNLIKELY and SPOTS LIKELY. The partition reflects the assertion that PALATAL SPOTS is relevant to neither the diseases STREP THROAT and TONSILLAR CELLULITIS, nor the diseases MONONUCLEOSIS, PERITONSILLAR ABSCESS, and VIRAL PHARYNGITIS. Consequently, only two probability distributions are assessed.

CELLULITIS, and (2) the feature is *not relevant to* the disease triplet MONONUCLEOSIS, PERITONSILLAR ABSCESS, and VIRAL PHARYNGITIS. Consequently, we need to assess only two probability distributions. These distributions, shown below the hypothesis sets in Figure 14, are the same as those shown in Figure 13. By using this partition, however, we uncover equalities among the distributions for PALATAL SPOTS before we assess probabilities; we thereby avoid the assessment of three additional distributions.

When a feature is dependent on other features, we can compose a partition for each instance of the conditioning features.[3] Figure 15 illustrates this approach for the feature TOXIC APPEARANCE, which is conditioned by the features FEVER and ABDOMINAL PAIN. In the figure, partitions and assessments for three of the six instances of the feature's conditioning variables are shown. The three partitions in the figure correspond to the cases where a patient has no abdominal pain and a mild fever, no abdominal pain and a high fever, and abdominal pain and a high fever. In the first and third partition, the feature TOXIC APPEARANCE is relevant to neither PERITONSILLAR ABSCESS and TONSILLAR CELLULITIS, which are two localized diseases (diseases that tend to affect only a small area of the throat—usually one or more tonsils), nor MONONUCLEOSIS and STREP THROAT, which are two diseases that tend to affect a large area of the throat as well as other organs in the body. In the second partition, the feature is not relevant to the two localized diseases, but is relevant to the remaining diseases.

In general, partitions are better able to express assertions of subset independence for assessment than are similarity networks. To understand this point, consider all possible similarity networks that an expert might construct for a given domain. The similarity network that can represent the most assertions of subset independence is the one that has a completely connected similarity graph (i.e., a similarity graph in which all hypothesis pairs are connected directly). From the assertions of subset independence encoded in this similarity network, we can derive a single partition for each feature. Specifically, we place

---

[3] In general, an instance of a set of variables is an assignment of an instance to each variable in that set. Thus, if variable x has instances x_1 and x_2, and variable y has instances y_1 and y_2, then the pairs (x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2) comprise all instances of the set $\{x, y\}$.

hypotheses h_i and h_j in the same set for variable x if and only if the variable x is not present in the local knowledge map for h_i and h_j. We cannot, however, derive different partitions for different instances of the features that condition x.

5.3 Partitions and Classification Hierarchies

In assessing the joint probability distribution for Pathfinder, the use of partitions decreased the time to assess a knowledge map by more than a factor of five. At first, this observation may seem surprising, given that a partition must be composed for each conditioning instance of every feature. In the medical domains that I have investigated, however, two factors have contributed to the efficiency of the approach. First, the task of composing a single partition is straightforward. Apparently, as is the case with assertions of symmetric conditional independence, people find it easy to make judgments of subset independence without assessing the probabilities underlying such judgments.

Second, partitions often are identical or related from one feature to another, and more often are identical or related among the conditioning instances of the same features. For example, two of the partitions for TOXIC APPEARANCE in Figure 15 are identical, and these two partitions are closely related to the third. The partition for FEVER, shown in Figure 16, is almost identical to the partitions for TOXIC APPEARANCE.

The partitions for FEVER and those for TOXIC APPEARANCE are related in that each represents a slice through the same *classification hierarchy* of diseases. This classification hierarchy is shown in Figure 17. In the figure, diseases that cause sore throat are divided into two major categories: diseases that tend to affect a large area of the throat (DIFFUSE DISEASE) and diseases that tend to affect a small area of the throat (LOCALIZED DISEASE). The diseases in the former category are subdivided into those diseases that tend to affect multiple organs (MULTIORGAN DISEASE) and the disease VIRAL PHARYNGITIS, which does not affect multiple organs. The partition for FEVER represents a slice through this hierarchy at the level of abstraction DIFFUSE versus LOCALIZED DISEASE. The first and third partitions for TOXIC APPEARANCE in Figure 15 represent the slice through the hierarchy that is one level more specific for diffuse diseases. The second partition for TOXIC APPEARANCE reflects the slice that is the most specific for diffuse diseases.

In many domains, experts often fabricate one or more classification hierarchies to describe shared features among groups of hypotheses. In Pathfinder, for example, there are 13 distinct classification hierarchies across the approximately 100 features. In domains where such hierarchies have been explicated, it is likely that partitions will be identical or closely related. When the partitions for two or more features are related, we do not need to create each partition from scratch. Using SimNet, we can copy the structure of a partition for a conditioning instance of a feature, and assign this structure either to the partition of another conditioning instance of the same feature, or to the partition for an entirely different feature. Once the structure of the partition has been copied, we can modify the partition using simple Macintosh-style manipulations to reflect the judgments of subset independence for that feature and conditioning instance. Note that SimNet supports hierarchical set mem-

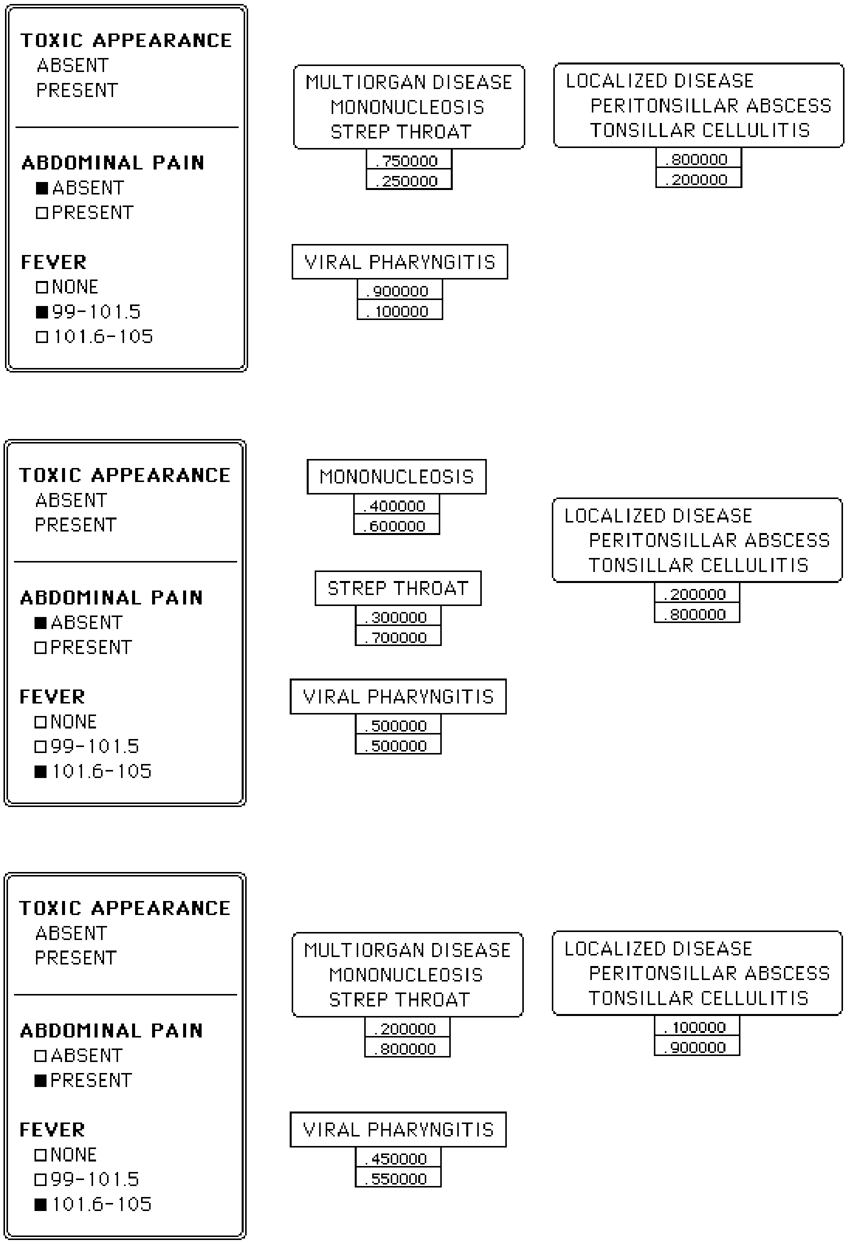

Figure 15: Assessing dependent features using partitions.

The figure contains partitions and associated assessments for three of the six conditioning instances of the feature TOXIC APPEARANCE. The three partitions correspond to cases where a patient has no abdominal pain and a mild fever, no abdominal pain and a high fever, and abdominal pain and a high fever, respectively. Only assessments for each set in a partition are required. Note that the first and third partitions are identical, and the second is closely related.

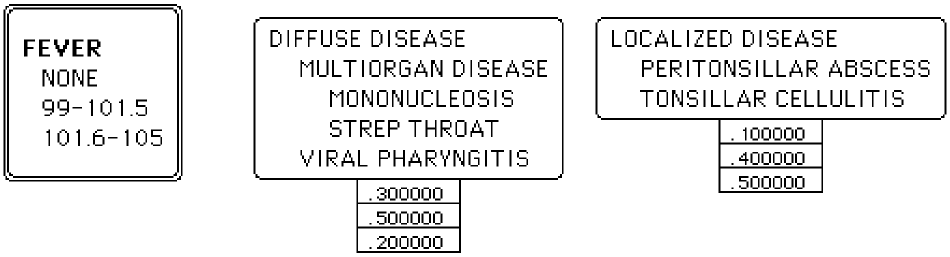

Figure 16: The partition for FEVER.

The partition is closely related to the partitions for TOXIC APPEARANCE. Note the hierarchical structure of the set DIFFUSE DISEASE. The set contains the disease VIRAL PHARYNGITIS and the set MULTIORGAN DISEASE, which, in turn, contains the diseases MONONUCLEOSIS and STREP THROAT. Hierarchical sets such as this one facilitate the composition of partitions.

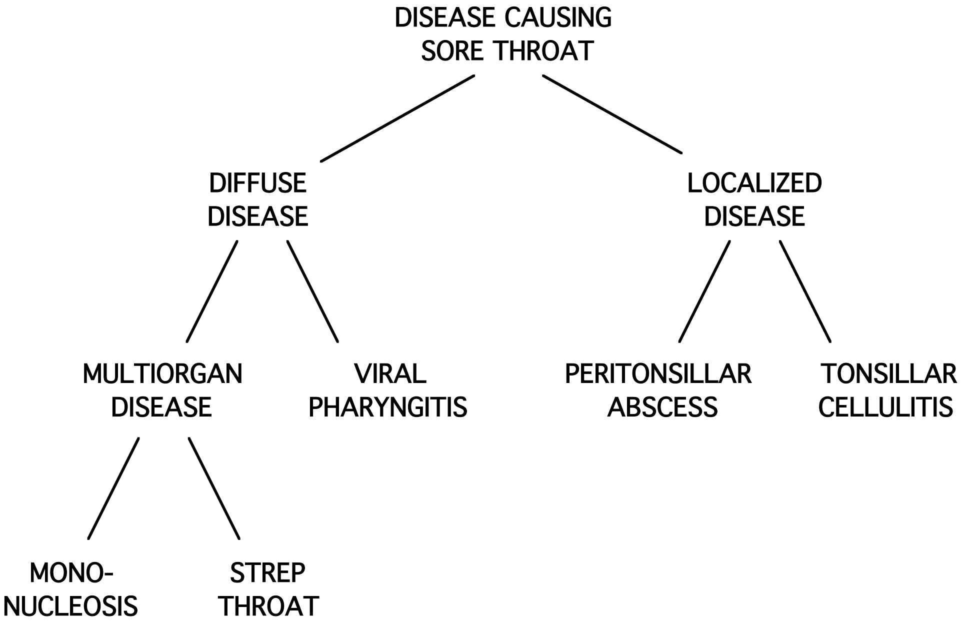

Figure 17: A classification hierarchy of the diseases causing sore throat.

In the hierarchy, diseases either are or are not localized to one portion of the throat. Two of the diseases that are not localized to one portion of the throat, MONONUCLEOSIS and STREP THROAT, affect multiple organs in the body. The partitions for TOXIC APPEARANCE and FEVER are slices of this hierarchy at different levels of abstraction.

bership within partitions. For example, in Figure 16, MONONUCLEOSIS is a member of the set MULTIORGAN DISEASE; this set, in turn, is a member of the set DIFFUSE DISEASE. This feature of the program makes it easy for a user to transform a partition that reflects a slice of a classification hierarchy at one level of abstraction to a partition that represents other levels of abstraction within the hierarchy.

5.4 Representation of Subset Independence in Ordinary Knowledge Maps

Asymmetrical assertions of conditional independence can be represented in an ordinary knowledge map with deterministic nodes. In this section, we examine the representation of subset independence in a knowledge map, and we discuss the merits of this representation relative to partitions.

Figure 18 contains the knowledge map that reflects the judgments of subset independence associated with the partitions for ABDOMINAL PAIN, FEVER, and TOXIC APPEARANCE in the sore-throat example. First, let us consider the feature ABDOMINAL PAIN. In the global knowledge map for the sore-throat problem (Figure 5), DISEASE is the only node that conditions ABDOMINAL PAIN. Consequently, there is only one partition for this feature. The sets within the partition for ABDOMINAL PAIN are represented by the instances of the deterministic node PARTITION FOR ABDOMINAL PAIN in the knowledge map. This node is deterministic because the sets in the partition are certain, given the set of possible diseases. The lack of an arc from DISEASE to ABDOMINAL PAIN represents the assertions of subset independence encoded by the partition for that feature. That is, given any set of diseases in the partition, knowing which disease in that set is the cause of the patient's sore throat does not change the probability that the patient will have abdominal pain. These same considerations apply to the feature FEVER.

Now consider the feature TOXIC APPEARANCE. In the global knowledge map, TOXIC APPEARANCE is dependent on both ABDOMINAL PAIN and FEVER. Therefore, the feature has a partition for each instance of these two features. The partitions and the sets within each of these partitions are represented by the deterministic node PARTITIONS FOR TOXIC APPEARANCE. Again, the node is deterministic because the partitions are known with certainty, given instances of DISEASE, ABDOMINAL PAIN, and FEVER. The lack of an arc from DISEASE to TOXIC APPEARANCE represents the assertions of subset independence associated with the partitions for the feature. In particular, given an instance of ABDOMINAL PAIN and FEVER, and a set in the appropriate partition, the probability that the patient's appearance will be toxic is independent of disease.

Although this knowledge map accurately encodes the assertions of subset independence for the three features, the representation is cumbersome. The partitions are hidden under the nodes in the knowledge map, and cannot be created or modified with simple graphic manipulations as they are in SimNet. Consequently, the representation of subset independence in ordinary knowledge maps does not facilitate assessment significantly. Similar remarks apply to the representation of hypothesis-specific independence in an ordinary knowledge

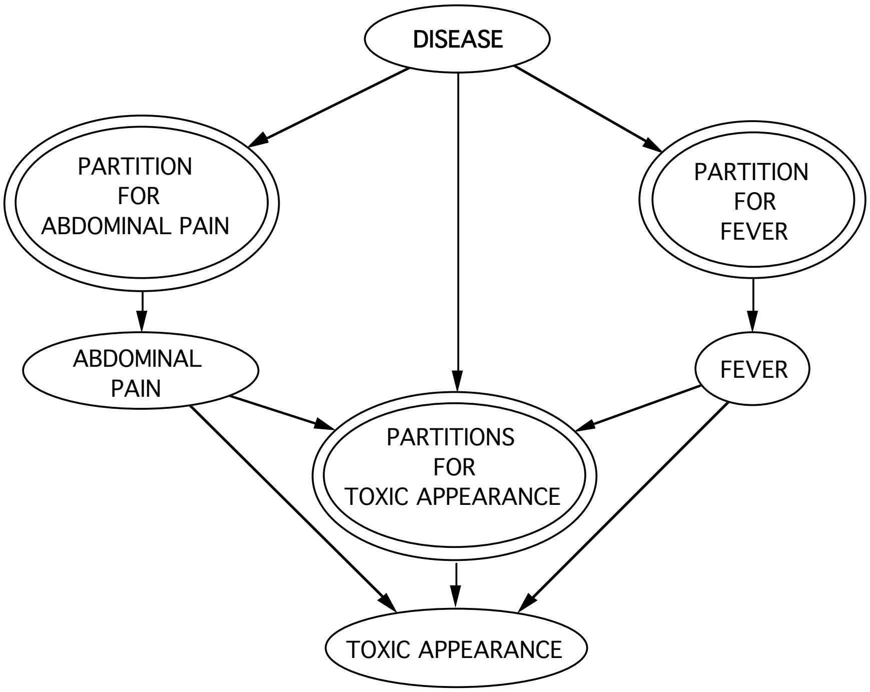

Figure 18: The representation of partitions in a knowledge map.

This knowledge map encodes the asymmetrical assertions of subset independence that are represented by partitions for the three features ABDOMINAL PAIN, FEVER, and TOXIC APPEARANCE. Because ABDOMINAL PAIN has no conditional predecessors except DISEASE, there is only one partition for the feature. The sets in the partition for ABDOMINAL PAIN are a deterministic function of DISEASE. Furthermore, the probability distribution for ABDOMINAL PAIN is independent of DISEASE, given the elements of the partition for the feature. Similar remarks apply to the feature FEVER. TOXIC APPEARANCE is conditioned both by ABDOMINAL PAIN and by FEVER. Consequently, TOXIC APPEARANCE has a partition for each combination of the instances of these two features. These partitions are a deterministic function of DISEASE, ABDOMINAL PAIN, and FEVER. The probability distributions for TOXIC APPEARANCE are independent of DISEASE, given ABDOMINAL PAIN, FEVER, and the elements of the partition for TOXIC APPEARANCE.

map.

5.5 Research Related to Partitions

There is an interesting connection between partitions and the Dempster–Shafer theory of belief. Several leaders in artificial-intelligence research have argued that probability theory is inadequate for reasoning under uncertainty. In providing motivation for the Dempster–Shafer theory, Gordon and Shortliffe argue that probability theory and other theories of uncertainty are inadequate, because they do not allow for the possibility that features or evidence bear on *sets* of diseases or hypotheses [6, page 324]:

> An advantage of the Dempster–Shafer theory over previous approaches is its ability to model the narrowing of the hypothesis set with accumulation of evidence, a process which characterizes diagnostic reasoning in medicine and expert reasoning in general. An expert uses evidence which may apply not only to single hypotheses but also to sets of hypotheses that together comprise a concept of interest.

Partitions, however, provide an alternative within probability theory for representing evidence that is relevant to sets of hypotheses. In particular, we can interpret the statement that "a piece of evidence applies to only a set of hypotheses" to mean that the evidence *is not relevant to* that set of hypotheses, by Equation 7. Furthermore, the probabilistic alternative to Dempster–Shafer theory can be more efficient computationally. Once the probabilities are assessed using partitions, the partitions can be ignored, and the effect of evidence on each hypothesis can be computed separately. Alternatively, the effect of evidence can be accumulated on set intersections of partitions, just as evidence is accumulated within the Dempster–Shafer theory. Here, however, we can accumulate evidence using Bayes' theorem, rather than Dempster's rule of combination. This alternative, when computationally feasible, might be useful for generating cogent explanations for the results of probabilistic inference.

Motivated by the same argument against the adequacy of probability theory, Pearl has developed a representation similar to partitions. In his book, Pearl discusses probability assessments in a domain in which a set of mutually exclusive and exhaustive hypotheses can be organized into what I have been calling a classification hierarchy [19, pages 333–344]. In this context, he examines the statement (my notation):

> Evidence x bears directly on $h_{\subseteq}$, but says nothing about the individual elements of $h_{\subseteq}$.

Pearl interprets this statement to mean

$$p\left(x|h_{\subseteq},\xi\right) = p\left(x|h_i,\xi\right), \quad h_i \in h_{\subseteq} \tag{9}$$

$$p\left(x|\bar{h}_{\subseteq},\xi\right) = p\left(x|h_i,\xi\right), \quad h_i \in \bar{h}_{\subseteq} \tag{10}$$

where $\bar{h}_{\subseteq}$ is the set of hypotheses in h that are not in $h_{\subseteq}$. Equations 9 and 10 imply that x is not relevant to the hypothesis set $h_{\subseteq}$ nor to the hypothesis set $\bar{h}_{\subseteq}$, respectively. Thus, Pearl's method is a special case of the partition approach to assessment in which we compose two-set partitions within a single classification hierarchy.

6 Conclusions

Researchers in the artificial-intelligence community have doubts that normative expert systems can be constructed for anything more complex than toy domains [26, 21]. For the most part, decision analysts share these doubts. They argue that, for a specific decision context, utilities reflecting the preferences of the decision maker are usually needed to prune an influence diagram to a manageable size. Hence, they find it almost inconceivable that a decision model designed to provide useful expertise over a wide range of decision problems can be constructed [12].

I hope that the research in this paper will convince workers in both the artificial-intelligence and decision-analysis communities that normative expert systems can be constructed and can provide valuable decision assistance in many domains. To the artificial-intelligence community I emphasize that we have not yet approached the limits to which decision theory can be applied to the representation of knowledge. By identifying forms of conditional independence that reflect the mechanisms that human experts use to manage the complexity of their domains, and by building representations that capture such forms of independence explicitly, researchers will be able to apply normative reasoning to domains that are much more complex than those being examined currently. To the decision-analytic community, I emphasize that a specific decision context is not always necessary to construct and use a decision model. Of course, the decision models that are built to handle large sets of decisions will themselves be large and complex. However, by identifying new forms of conditional independence and by extending the current influence-diagram representation to include explicit recognition of these independencies, decision analysts will be able to build expert systems that provide valuable assistance to people confronted with confusing, high-stakes decisions.

Acknowledgments

I thank Bharat Nathwani (the Pathfinder expert), Ross Shachter, and Eric Horvitz for inspiring several of the ideas in this paper; Mike Wellman and Jack Breese for providing technical feedback on the theory of similarity networks; Keung-Chi Ng, Mark Fischinger, Jaap Suermondt, and Marty Chavez for assisting with the programming of SimNet; Harold Lehmann for providing the knowledge for the sore-throat example; and Ron Howard, Peter Hart, Ted Shortliffe, David Matheson, and Jeff Bradshaw for thought-provoking discussions. Greg Cooper, Lyn Dupré, Dan Geiger, Peter Hart, Ross Shachter, and Ted Shortliffe provided useful comments on earlier drafts of this manuscript.

References

[1] I.A. Beinlich, H.J. Suermondt, R.M. Chavez, and G.F. Cooper. The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. In *Proceedings of the Second European Conference on Artificial Intelligence in Medicine,* London. Springer Verlag, Berlin, August 1989.

[2] G.F. Cooper. *NESTOR: A Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge*. PhD thesis, Medical Computer Science Group, Stanford University, Stanford, CA, November 1984. Report HPP-84-48.

[3] G.F. Cooper. Bayesian belief-network inference using nested dissection. Technical Report KSL-90-05, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, January 1990.

[4] R. Duda, J. Gaschnig, and P. Hart. Model design in the PROSPECTOR consultant system for mineral exploration. In D. Michie, editor, *Expert Systems in the Microelectronic Age*, pages 153–167. Edinburg University Press, Edinburgh, Scotland, 1979.

[5] W. Edwards, ed. Special issue on probabilistic inference. *IEEE Transactions on Human Factors in Electronics*, 7, 1956.

[6] J. Gordon and E.H. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26:323–357, 1985.

[7] G.A. Gorry and G.O. Barnett. Experience with a model of sequential diagnosis. *Computers and Biomedical Research*, 1:490–507, 1968.

[8] D.E. Heckerman. *Probabilistic Similarity Networks*. PhD thesis, Program in Medical Information Sciences, Stanford University, Stanford, CA, June 1990.

[9] D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani. Update on the Pathfinder project. In *Proceedings of the Thirteenth Symposium on Computer Applications in Medical Care,* Washington, DC, pages 203–207. IEEE Computer Society Press, Silver Spring, MD, November 1989.

[10] M. Henrion. Propagation of uncertainty by probabilistic logic sampling in Bayes' networks. In J.F. Lemmer and L.N. Kanal, editors, *Uncertainty in Artificial Intelligence 2*, pages 149–164. North-Holland, New York, 1988.

[11] M. Henrion and D.R. Cooley. An experimental comparison of knowedge engineering for expert systems and for decision analysis. In *Proceedings AAAI-87 Sixth National Conference on Artificial Intelligence,* Seattle, WA, pages 471–476. Morgan Kaufmann, San Mateo, CA, July 1987.

[12] S. Holtzman. *Intelligent Decision Systems*. Addison–Wesley, Menlo Park, CA, 1989.

[13] R.A. Howard. Value of information lotteries. *IEEE Transactions of Systems Science and Cybernetics*, SSC-3(1):54–60, 1967.

[14] R.A. Howard. Knowledge maps. *Management Science*, 35:903–922, 1989.

[15] R.A. Howard and J.E. Matheson. Influence diagrams. In R.A. Howard and J.E. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*, volume II, pages 721–762. Strategic Decisions Group, Menlo Park, CA, 1981.

[16] D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York, 1982.

[17] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistical Society B*, 50:157–224, 1988.

[18] K.G. Olesen, U. Kjaerulff, F. Jensen, F.V. Jensen, B. Flack, S. Andreassen, and S.K. Andersen. A MUNIN network for the median nerve: A case study on loops. *Applied Artificial Intelligence*, 3:385–404, 1989.

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

[20] J.A. Reggia and B.T. Perricone. Answer justification in medical decision support systems based on Bayesian classification. *Computers in Biology and Medicine*, 15:161–167, 1985.

[21] E. Rich. *Artificial Intelligence*. McGraw–Hill, New York, 1983.

[22] R.D. Shachter. Evaluating influence diagrams. *Operations Research*, 34:871–882, 1986.

[23] R.D. Shachter. An ordered examination of influence diagrams. *Networks*, 20, 1990.

[24] G. Shafer. Savage revisited. *Statistical Science*, 1:463–501, 1986.

[25] E.H. Shortliffe. *Computer-based Medical Consultations: MYCIN*. North-Holland, New York, 1976.

[26] E.H. Shortliffe and B.G. Buchanan. A model of inexact reasoning in medicine. *Mathematical Biosciences*, 23:351–379, 1975.

[27] D.J. Spiegelhalter and R.P. Knill-Jones. Statistical and knowledge-based approaches to clinical decision support systems, with an application in gastroenterology. *Journal of the Royal Statistical Society*, 147:35–77, 1984.

[28] A. Tversky and D. Kahneman. Judgment under uncertainty: Heruistics and biases. *Science*, 185:1124–1131, 1974.

[29] L. Zadeh. Is probability theory sufficient for dealing with uncertainty in AI: A negative view. In L.N Kanal. and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence*, pages 103–116. North-Holland, New York, 1986.