HYDI: A Hybrid System with Feedback for Diagnosing Multiple Disorders

by

Yeona Jang

August, 1993

HYDI: A Hybrid System with Feedback for Diagnosing Multiple Disorders

by

Yeona Jang

Submitted to the
Department of Electrical Engineering and Computer Science
on August, 1993
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Abstract

This thesis investigates the advantages of introducing feedback between the processes of automated diagnosis and automated knowledge acquisition. The introduction of such feedback results in an "hybrid" system that can learn from its own problem-solving experience by analyzing the results of previous diagnoses and incorporating their key features into an associative knowledge base which, in turn, assists future diagnosis. Experimental results show that such a system is capable of an efficiency/accuracy trade-off when applied to the problem of diagnosing multiple disorders.

A primary feature of this work is a new mechanism, called the "diagnostic-unit" representation, for remembering results of previous diagnoses in a "decomposed-and-merged" form. The *diagnostic-unit representation* is explicitly designed to capture the most likely causal relationships between disorders and clusters of findings. Unlike typical bipartite "If-Then" representations, the diagnostic-unit representation uses a general graph representation to capture more complex causal relationships between disorders and findings. Diagnostic units provide guides for decomposing a set of findings into smaller subsets for which the most likely explanations can be immediately inferred. Diagnostic units thus allow decompositional abductive diagnosis to be done efficiently and effectively. They also can facilitate one's understanding of the structure inherent in the diagnosis domain.

In addition to the basic diagnostic-unit concept, this thesis develops

experience-based strategies for incrementally deriving and updating diagnostic units and the various relationships between them. Techniques for selecting diagnostic units relevant to a given problem and then combining them to generate overall solutions are also described. These strategies and techniques are implemented in a computer system called HYDI, and have been tested in the domain of diagnosing heart failure.

Thesis Supervisor: Peter Szolovits
Title: Professor, Electrical Engineering and Computer Science

Acknowledgements

My deepest thanks go to Peter Szolovits. He has been my advisor in and out of academics. His solid guidance helped keep me on course. I am especially thankful for his unreserved encouragement, patience, and support from day one.

Special thanks are due also to William Long. His knowledge and patience inspired me, and discussion with him led me to many insights. I was extremely fortunate to have his Heart Failure Program as a resource for my own research.

I thank Randall Davis, who did not seem to mind my interruptions despite his busy schedule, for providing useful feedback on earlier drafts of this work.

I am thankful for comments supplied by the members of the Clinical Decision Making Group at MIT. They have been a source of friendship as well as colleagueship. Thanks to all those who made it such a pleasurable and productive working environment. I am particularly thankful to Ira Haimowitz and Tze-Yun Leong for their ever-willingness to cheer me up.

I want to thank my parents for their lifetime of love, care, and support.

And, I thank my husband Alex for his heroic draft-reading efforts and innumerable suggestions. Most importantly, without his love, support, and sacrifices, I could never have finished this thesis.

Finally, I would like to acknowledge the research support from National Institutes of Health under Grant R01 LM04493 from the National Library of Medicine and under Grant R01 HL33041 from National Heart, Lung, Blood Institute. This research was also supported, in part, by the Harvard, MIT, and Tufts/New England Medical Center Combined Research Training Program in Medical Informatics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

This thesis addresses three problems. The first is the diagnosis problem, in particular, the problem of efficiently identifying the most likely causal events for a given body of evidence. The second is the knowledge acquisition problem, particularly, the problem of acquiring knowledge about the context sensitivity of the conclusions that can be drawn from an individual piece of evidence. This type of knowledge can guide the grouping of a given body of evidence into easier subproblems, and thus can be called knowledge about domain structure. The third is the problem of representing such domain-structure knowledge. The three problems are not independent of each other: The choice of knowledge to acquire depends on the goal of problem solving, and mechanisms chosen to represent and use the knowledge affect overall diagnostic performance. Though reasonable methods have been proposed for solving each of these problems individually, the slow advancement in the development of a system capable of competent and efficient medical diagnosis gives evidence that the direct combination of such independent solutions does not necessarily yield acceptable results. The methods described in this thesis seek to achieve efficient diagnosis, by simultaneously considering the

importance and interrelationship of all three problems.

1.1.1 Basic Approach

The basic approach explored in this thesis is the introduction of feedback between the processes of problem solving and knowledge acquisition. The introduction of such feedback results in a hybrid system that generates hypotheses to account for a given body of evidence, analyzes the results of the diagnosis, and incorporates their key features into an experiential knowledge base which, in turn, assists future diagnosis. This hybrid architecture for a diagnostic system embodies a learning by doing paradigm.

Basic Approach to Diagnosis Problem

The basic approach to the diagnostic problem is hybrid reasoning that makes use of association-based reasoning in conjunction with causal-model-based reasoning. The goal is to make the solution of complex diagnostic problems more robust and efficient. While reusing an existing system for causal-model-based reasoning, this thesis develops and analyzes a decompositional abductive technique for association-based reasoning.

For an intuitive understanding of decompositional abductive diagnosis, suppose that findings f_1, f_2, ..., f_n are the evidence upon which a diagnosis is based. One reasonable question to ask is "Can we group these findings into relatively independent subsets of findings for which the most likely hypotheses can be immediately inferred?" In other words, can we solve the problem by decomposition and abduction? If so, a solution to the original problem can be generated quickly, by combining solutions to the subsets of findings where each partial solution represents a disorder that explains part of the overall malfunction.

Problem decomposition allows a complex problem to be solved efficiently, by simplifying it into subproblems [24, 32, 70, 95]. Abduction also allows efficient problem solving, by avoiding step-by-step reasoning from first prin-

ciples. Abductive inference makes "backward" inferences based on known causal relations, to explain or justify a conclusion [18, 38, 45, 76, 89]. In other words, given the truth of proposition Q and the implication $P \rightarrow Q$, abduction is the immediate conclusion of P as an explanation for Q (even though this may be an incorrect explanation). For example,

> Flu can cause headache.
> Headache is present.
> ---
> $\Rightarrow$ Conclude flu as an explanation for the presence of headache.

It is the potential efficiency that motivates the use of decomposition and abduction for multidisorder diagnosis.

Decompositional techniques are efficient, however, only when a problem is decomposed correctly [60]. Unfortunately, the task of finding correct decompositions for diagnosis is a difficult task, for there are exponentially many ways of decomposing a given set of findings. Similarly, abduction is an effective technique for solving the diagnosis problem only when all known causal relations are most likely. Otherwise, poor performance with respect to accuracy will result (the actual, or best, explanation for headache may be emotional stress, fatigue, or a serious intracranial disease such as brain tumor). In light of these observations, this thesis attempts to explicitly represent *knowledge about the context sensitivity of the conclusions that can be drawn from findings*. Such knowledge provides guides for decomposing a given set of findings and also can be formulated to capture only the most likely causal relations.

A part of this thesis describes and analyzes techniques designed to address the issue of how to efficiently use such knowledge in decompositional abductive diagnosis, more specifically 1) a technique, called "deep matching adaptation," for identifying relevant knowledge pieces based not only on similarity on the surface but also on similarity in underlying causality; 2) a similarity metric, called "specificity-reflected similarity," for determining levels of relevance of such knowledge pieces; and 3) a technique, called

"dependency-guided picking," for combining chosen knowledge pieces to generate a solution to a given diagnostic problem.

Basic Approach to Knowledge Acquisition and Representation Problems

An important issue is how to acquire and represent knowledge that facilitates decompositional abductive diagnosis. This research investigates the role of experience in knowledge acquisition, by viewing problem-solving experiences as guidance for abstracting the context sensitivity of findings associated with particular disorders. Knowledge about the context sensitivity of findings is acquired from experience, and stored in a structured form that allows decompositional abductive diagnosis to be reduced to the retrieval of relevant knowledge pieces.

A new mechanism, called the *diagnostic-unit representation*, is designed to organize knowledge about the context sensitivity of findings. In the diagnostic-unit representation, disorders and sets of findings that are in the most likely causal relation are explicitly grouped into diagnostic units. Disorders and sets of findings in diagnostic units are linked by causal relationships. Each diagnostic unit can be conceptualized, at the most abstract level, as follows:

$$\text{Disorder } d \longrightarrow \text{Set } F \text{ of findings.}$$

In this relation, a link between a disorder and a set of findings means that the disorder is believed to be the most likely cause for the findings in the set (not just some plausible cause). Unlike in typical bipartite "If-Then" representations, each association between a disorder and a set of findings is supported by a highly likely causal mechanism that underlies the association. To efficiently represent such complex causal relationships between disorders and sets of findings, the diagnostic-unit representation uses a general graph representation. Diagnostic units provide guides for decomposing a set of findings into smaller subsets for which the most likely explanations can be

immediately inferred, thereby allowing decompositional abductive diagnosis to be done efficiently and effectively. Once constructed, diagnostic units also have the potential to facilitate one's understanding of the structure inherent in the diagnosis domain. In addition to the basic diagnostic-unit concept, this thesis develops methods for transforming problem-solving experience (observations about diagnosis) into diagnostic units (more general problem-solving rules for decompositional abductive diagnosis).

The efficiency and effectiveness of the techniques and methods developed in this thesis are tested in the domain of heart failure diagnosis,[1] by implementing a computer system called HYDI. Heart failure diagnosis is chosen as a test domain because 1) the domain is large, 2) causal relationships between findings and disorders are generally many-to-many, uncertain, and indirect, 3) multiple coexisting diseases are common, and 4) a relatively robust causal-model-based system for heart failure diagnosis is available.

1.1.2 Contributions

The contribution of this work is two-fold. One contribution is the development of new methods for solving problems in complex diagnostic domains. Classical techniques for diagnosis include association-based and model-based reasoning. In general, an association-based reasoning system [11, 19, 26, 42, 65, 66, 79, 90] can solve problems efficiently, but is fragile in the sense that it is only good at solving prespecified, familiar problems. Conversely, a causal-model-based reasoning system [21, 23, 35, 37, 56, 62, 72, 83, 98] can solve not only familiar but also unfamiliar problems from first principles, but is slow. This work is motivated by the desire to develop a diagnostic technique whose efficiency is comparable to that of association-based reasoning but with robustness which approaches that of causal-model-based reasoning. This work

[1] Heart failure is "a condition in which the pumping action of the ventricle of the heart is inadequate" [59].

shows that such a system *is* in fact possible, by combining association-based and causal-model-based reasoning.

Another contribution would be in automating the acquisition of knowledge about domain structure. Patterns of findings can be useful indicators of the existence of disorders and corresponding underlying causal mechanisms. Expert physicians seem to examine findings for such indications. Unfortunately, the ill-structuredness of the medical diagnosis domain [92] makes the task of finding the correct decomposition of findings difficult. A computer program that can automatically identify such patterns for diagnosis thus has great utility to physicians. This research explores methods for automatically discovering such patterns by accumulating a problem-solver's own experience in a decomposed and merged form: The results of the accumulation are domain-structure knowledge that identifies how a given body of evidence can be structured by grouping it into subproblems.

1.2 Diagnosis of Multiple Disorders

This section describes the assumptions under which the diagnosis problem is addressed in this thesis. It also describes how diagnostic problems and their solutions are represented.

1.2.1 Assumptions made for Diagnosis

Multifault assumption: A common diagnosis assumption is the single-fault assumption, where only a single disorder, or a *fault*, is assumed to produce all findings [11, 20, 74, 90]. In medical domains like heart failure or Acquired Immune Deficiency Syndrome (AIDS), however, patients suffering from multiple coexisting disorders are not uncommon [63]. In light of the prevalence of multiple disorders, this thesis addresses the diagnosis problem under the multifault assumption that more than one disorder may be present simultaneously. While expanding the power and scope of diagnos-

tic techniques, handling of the multifault assumption generally implies an increase in computational complexity. In multidisorder diagnosis, multiple disorders may plausibly explain a set of findings. Unfortunately, it is often unknown *a priori* how many disorders coexist. For example, when a patient shows increased Aspartate Aminotransferase (AST) and complains of pain in the upper quadrant of his or her abdomen, the findings could be caused by hepatic graft-versus-host disease (GVHD), by cardiac cirrhosis,[2] by both, or by any combination of other disorders that can cause these findings. As a consequence, identification of the most likely combinations of disorders potentially requires a search through a hypothesis space that grows exponentially with the number of disorders [13, 43].

Disorders are not always independent: Most existing systems for diagnosing multiple faults perform diagnostic inferences based on the restrictive disorder independence assumption [42, 65, 66]. This assumption specifies that given findings are grouped into subsets such that all the findings in each set are caused by the same disorder, and each subset must be explained by a unique disorder that does not explain findings in any other subset. The disorder independence assumption allows for substantial reduction in the computation involved in decompositional diagnosis – especially the process of combining partial solutions to an entire solution. While computationally attractive, however, the disorder independence assumption makes it difficult to deal appropriately with situations in which one disorder can cause other disorders. In most medical domains, disorders are not always independent of each other. For example, congestive heart failure can cause cardiac cirrhosis. This thesis addresses multidisorder diagnosis without assuming that disorders are independent. Relaxing the disorder independence assumption complicates decompositional diagnosis, especially the process of synthesizing a solution from partial solutions. This thesis develops and analyzes tech-

[2]Cardiac cirrhosis is a liver disease characterized by the formation of fibrous tissue in the liver as a result of passive congestion of the liver due to congestive heart failure.

niques for dealing with such increased computational complexity.

1.2.2 Representation of Diagnostic Problems

A diagnostic problem, or *case*, is to identify, for a given set of findings, the most likely causes producing the findings in the set.

Input: Diagnostic problems are represented in the form of a set of findings. Findings include the history of a patient such as known diagnoses and therapies received, subjective symptoms such as a patient's complaints, physical examination, and objective signs revealed either by observations or by various special laboratory tests.

Output: A *diagnosis* consists of not only the disorders primarily suspected of causing a given set of findings, but also an underlying pathophysiologic mechanism that explains how these disorders are producing the findings in the set. Providing an underlying causal mechanism as an integral part of a diagnosis is important for two reasons. First, while a single listing of the suspected disorders is computationally less burdensome to generate, oversimplified accounts of underlying causality provide few insights about how these disorders are producing the findings, and thus give little guidance in therapy planning and management. It is important to provide an underlying pathophysiologic mechanism, particularly when intermediate links within the causal chain are important determinants in the appropriate therapy for the patient. Moreover, because therapies may change existing symptoms or produce their own symptoms, it is important that one be able to understand the causes of observed findings. Secondly, without a pathophysiologic mechanism that explains how findings are related to their primary suspected disorders, it is difficult to evaluate whether the results of a diagnosis "make sense" or are little more than random guesses.

In general, it is important to find not only some explanations for a set of findings but also the most likely explanations, particularly when high risk

and costs are associated with treatments or tests. The computational cost of finding the globally most likely explanations may, however, be unacceptably high. The methods presented in this thesis seek to efficiently identify explanations that while not necessarily globally optimal, are reasonable approximations to the desired optimal solutions.

This thesis uses a general graph notation to represent causal explanations for findings. An example of this graph representation is shown in Figure 1.1.

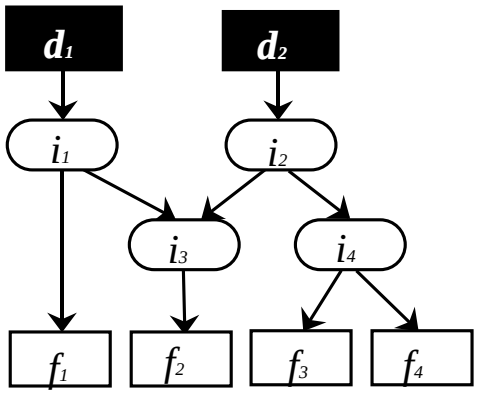


Figure 1.1: An example of a causal explanation for given findings

Black rectangular and oval nodes represent pathophysiologic states. Pathophysiologic states are states of living organisms, and their components, that arise from bodily abnormality or the failure of some organ or mechanism to function properly. Pathophysiologic states are divided into elemental disorders and intermediate states. Elemental disorders are either pathophysiologic states that are defined at a level needed for differential diagnosis, such as myocardial infarction,[3] or pathophysiologic states that do not require any

[3]Myocardial infarction, commonly known as a heart attack, is the formation of an infarct (the morphological changes indicative of cell death) as a result of interruption of the blood supply in cardiac muscle.

further causes, such as constrictive pericarditis.[4] Intermediate states are the remaining pathophysiologic states, such as salt and water retention. In Figure 1.1, each black rectangular node represents an elemental disorder, and each oval node an intermediate state. Each rectangular node represents a finding such as pedal edema. Each link represents a direct causal relation between the two clinical entities represented by the corresponding nodes. The causal graph shown in Figure 1.1 represents a causal explanation with the elemental disorders producing the findings via the intermediate states identified in the graph.

The same set of findings can be explained in many ways. The main driving force of this research is the question "How can we efficiently find causal graphs that represent the most likely causal explanations for a set of findings?"

1.3 Research Hypotheses

This section describes the hypotheses tested in this thesis.

1.3.1 Hybrid Reasoning Approach to Diagnosis

In general, association-based reasoning is faster than model-based reasoning, while the former is less robust than the latter. This thesis explores a hybrid reasoning approach which seeks to take advantage of the complementary strengths of association-based and causal-model-based reasoning. The goal is to develop a diagnostic system that is both robust and efficient.

Hypothesis 1 (Hybrid reasoning for diagnosis): The hybrid use of association-based and causal-model-based reasoning can enhance the overall performance of a diagnostic system.

[4] Constrictive pericarditis is inflammation of the membrane surrounding the heart which leads to thickening.

The hybrid reasoning architecture investigated in this thesis consists of a relatively robust causal-model-based component capable of first principles diagnosis and an association-based component for diagnosing disorders that "have been seen before." The central idea of the hybrid reasoning architecture for combining association-based and causal-model-based reasoning is to solve diagnostic problems, whenever possible, using association-based reasoning; but if association-based reasoning fails to solve a problem, then causal-model-based reasoning is performed to solve the problem step-by-step from first principles. While reusing an existing causal-model-based reasoning system, this thesis develops a decompositional abductive technique for the association-based component.

It is assumed that the associative knowledge base on which the association-based problem-solving component operates is initially empty, and consequently the component is not initially capable of reliable diagnosis. It is worth noting that this assumption is analogous to physicians who first start medical practice: In general, while possessing sufficient amounts of domain knowledge, such physicians lack the experience needed to make efficient association-based diagnosis. As various types of diagnostic problems are solved, the results of diagnoses are incorporated as an integral part of the associative knowledge base. Familiar kinds of problems then can be solved more efficiently, without search, as experienced physicians appear able to do [25, 27, 46, 47, 48].

1.3.2 Abstraction of Causal Knowledge

In general, "knowledge inundation" makes diagnosis from first principles computationally intensive. For instance, consider the causal relationships between disorders and findings. Such relationships are often many-to-many, indirect, and uncertain. As knowledge is added, entities in a knowledge base are increasingly likely to interact in uncertain ways with each other. In light of this observation, this thesis organizes complex pathophysiologic knowledge in a structured form so that knowledge relevant to a diagnostic problem can

be efficiently identified and retrieved (by the association-based component).

To provide an intuitive understanding of the organization structure, let us consider the causal knowledge base shown in Figure 1.2(a).

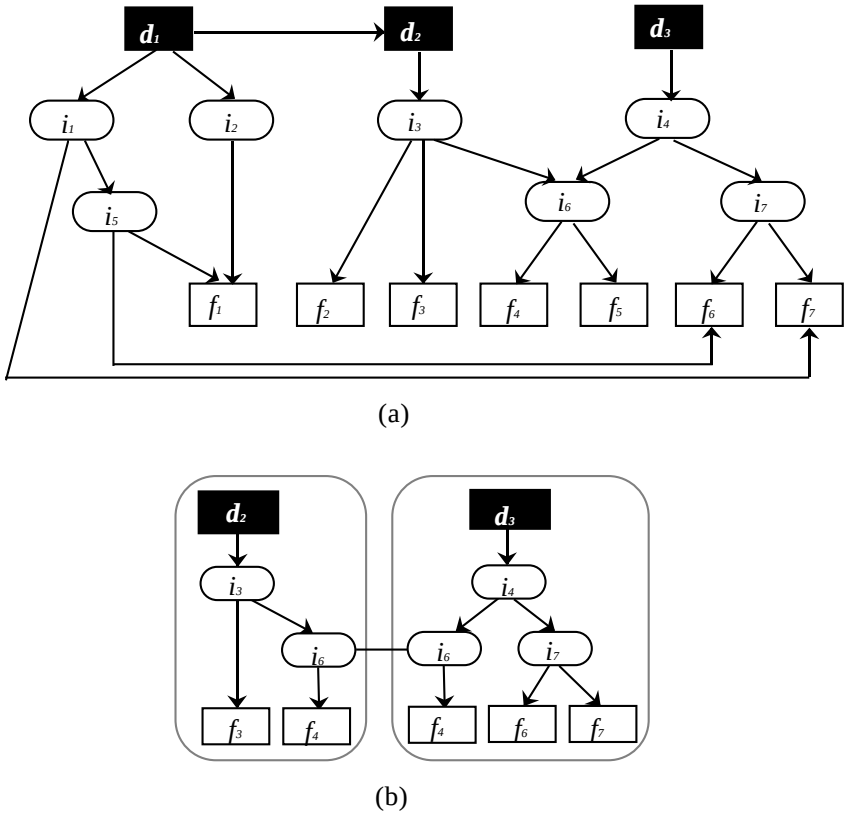


Figure 1.2: Organizing a knowledge base such that domain decomposability is captured; (a) Original causal knowledge base; (b) Abstract knowledge base that captures domain decomposability

The causal knowledge base exhibits little structure, and provides little guidance in detecting patterns of findings that may exist in a patient. Now, suppose that the causal knowledge base is abstracted to the knowledge base shown in Figure 1.2(b). Underlying this transformation is the premise that ill-structured causal knowledge can be grouped into relatively independent

modules that capture the decomposability of the diagnosis domain. In the abstracted knowledge base, the structure is more apparent, though some of the nodes and causal links have been removed. Nodes and links are removed if they are deemed to be "diagnostically insignificant." This thesis defines diagnostically insignificant causal relations as causal relations that are plausible in principle, but are not likely to appear as parts of diagnoses in practice. Diagnostically significant causal relations, causal relations that are not only plausible in principle but also likely to be parts of diagnoses of patients, remain in the abstracted knowledge base. Remaining nodes and causal links are organized into two modules where each module consists of a single disorder, a set of intermediate states, a set of findings, and a set of causal links. While the two modules are not completely decomposable, they are "nearly decomposable" in the sense that links within each module are relatively dense and strong, but links between two modules are relatively sparse and weak. Each module then can be used as a whole independently of the other. Such "structuring" or abstraction provides insights about how given findings can be grouped into subproblems. This thesis develops techniques for using such knowledge about domain structure to perform decompositional abductive diagnosis efficiently and effectively.

Case-based reasoning provides another example of a modular representation of knowledge [36, 44, 53, 55, 84, 94]. In general, cases are stored as atoms. Each stored case represents a chunk of (relatively low-level) knowledge that is used independently of other cases, to solve problems. The concept of knowledge modules is also explored in the area of learning [12, 28, 39, 57, 85, 86].

What to Abstract: Context Sensitivity of Findings

The next issue is what knowledge to abstract (in modules shown in Figure 1.2(b)). What to abstract is not independent of the goal of problem solving, since it depends heavily on the use to which we intend to put the knowledge. Abstracting knowledge in an intentional vacuum may result in an abstraction that is too general to efficiently and effectively lead any particular

problem-solving process to a solution.

This thesis abstracts knowledge which allows search-intensive causal-model-based diagnosis to be reduced to faster decompositional abductive diagnosis. To do so, it observes that the context sensitivity of findings can greatly affect overall performance. In most medical domains, findings in isolation may have more than one cause, but findings as a whole can constrain each other's cause, consequently reducing the number of causes to consider. In addition, the significance of a finding depends on other findings that occur together with it. As a consequence, changing some findings may even require findings that remain the same to be explained differently. For example, high cardiac output is often the most likely explanation for systolic ejection murmur.[5] If findings that strongly suggest low cardiac output are presented, however, aortic stenosis[6] might be a better explanation for systolic ejection murmur than high cardiac output. This example illustrates that the most likely cause of a finding cannot be determined by looking at its immediate cause in isolation, since it generally depends on the relative strength of causes of other findings that occur simultaneously.

The context sensitivity of findings implies that findings can be grouped into sets such that the findings in each set, taken together, immediately suggest the most likely disorder and pathophysiologic mechanism that best explains how this disorder is causing the findings in the set. This thesis explicitly represents, as a unit, such a set of findings and its most likely disorder and underlying pathophysiologic mechanism. Such knowledge about the context sensitivity of findings allows diagnosis of multiple disorders to be reduced to reasoning about how to decompose a set of findings into smaller subsets which can be solved immediately and relatively independently of each other.

[5] Systolic ejection murmur is an adventitious sound heard on auscultation of the heart that is most intense at the time of maximum flow of blood from the heart.

[6] Aortic stenosis is narrowing of aortic valve or its orifice due to lesions of the wall with scar formation.

How to acquire: Role of Experience

The key question that arises is how to acquire rules that capture the context sensitivity of findings. A typical method, often used in expert systems development, is an interview with a domain expert. Cognitive and AI researchers have found, however, that it is generally hard to directly draw, even from expert physicians, such experiential knowledge [75, 97]. In an attempt to overcome the difficulties associated with manual compilation, this thesis views experience accumulation as the processes through which knowledge is structured into a ready-to-use coherently simplified whole. This view allows the results of previous diagnoses to be used to acquire general problem-solving rules for diagnosis, *i.e.,* experiential knowledge. In the results of diagnosis, causal relationships that are critical to diagnosis are highlighted, while insignificant ones are suppressed. The results of previous diagnosis, therefore, provide a useful level of abstraction that allows diagnostic problems to be solved efficiently.

Observe that the automation of the experiential knowledge acquisition process has in effect introduced feedback between the reasoning and knowledge-acquisition processes. The assessment of a hypothesis computed during problem solving becomes a key link in a feedback loop. Whenever problem solving is completed, the associative knowledge base is adjusted. Such dynamic incorporation of revisions, *i.e.,* new experience, into the associative knowledge base allows the incorporation of changes in knowledge, prior to beginning a new problem-solving cycle. This adjustment thus feeds forward to future diagnosis. The following hypotheses summarize the notion of feedback between the reasoning and knowledge-acquisition processes.

Hypothesis 2 (Role of experience in the acquisition of general rules for diagnosis): The results of previous diagnoses can be used to incrementally derive, and update, knowledge about the context sensitivity of findings and thus to guide decompositional abductive diagnosis.

Hypothesis 3 (Role of experience in domain understanding and problem solving): As knowledge about the context sensitivity of findings is abstracted, understanding of the decomposability of the diagnosis domain is enhanced, and subsequent problem solving can be done more efficiently.

The associative knowledge base investigated in this thesis differs from most existing knowledge bases in the way knowledge is acquired. In most existing knowledge-based systems, including the causal-model-based system used in this thesis, knowledge is acquired from domain experts, and then encoded manually into a knowledge base. Once encoded, knowledge generally remains fixed in the knowledge base, and used repeatedly for each problem. In this regard, this type of knowledge bases can be called *static.* In contrast, contents of the associative knowledge base investigated in this thesis are derived from the problem-solver's own experience, and change with experience. The associative knowledge base thus can be called a *dynamic* knowledge base.

How to Represent: Representation of Experiences in a Decomposed-and-Merged Form

The choice of problem-solving algorithms and representational mechanisms often has a significant impact on overall problem-solving performance. The issue of how to remember the results of previous diagnoses must therefore be addressed. More specifically, one needs to decide whether to treat each solved case as "atomic" or decomposable. This choice of a "grain size" can affect overall reasoning performance and domain understanding. A common approach, investigated in most existing case-based reasoning systems, is to store every solved case as an independent atom [36, 53, 55, 84, 94]. While easy to implement, storing each case as an atom can limit the reusability of previous cases in future problem solving, by reducing the possibility of finding matches particularly when only parts of cases match. It may also result in inefficient use of memory space, because even very similar cases

are stored separately. In addition, such "redundancy" can adversely affect overall problem-solving efficiency. As solved cases are added to a system, we expect the system to be able to solve a problem without search (by directly retrieving a solution to a similar previous case). In deliberating about what case to use, however, all the stored cases need to be considered. If there are many cases, then simply searching them to find the best match(es) can be very time-consuming.

In an attempt to deal with these difficulties, this thesis stores the results of previous diagnoses in a "decomposed-and-merged" form. To provide an intuitive picture, consider two hypothetical solved cases, Case_1 and Case_2, shown in Figure 1.3(a). Case_1 is decomposable into a, b, c, and d, while Case_2 can be decomposed into a, b', d', and e.

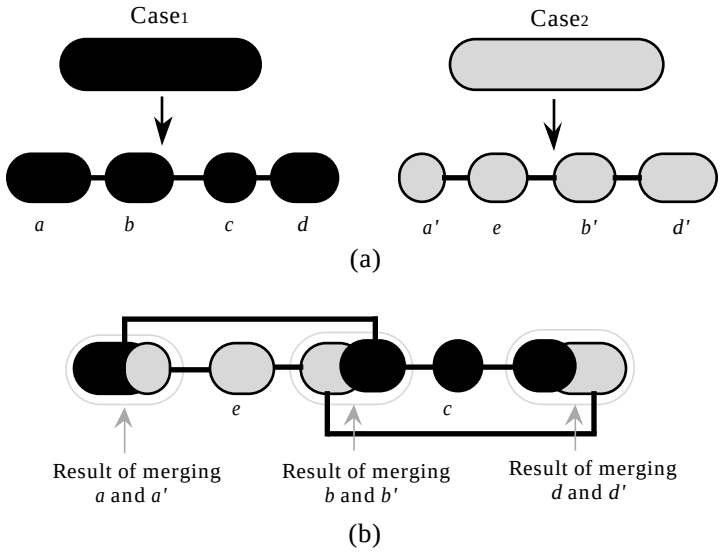


Figure 1.3: An example of storing solved cases in a decomposed-and-merged form; a) Case_1 is decomposable into a, b, c, and d, while Case_2 can decomposed into a', e, b', and d'; b) A form in which Case_1 and Case_2 are stored, supposing that a and a', b and b', and d and d', respectively, can be merged to produce a combined whole

This thesis attempts to merge components from different solved cases when

possible. For expository purposes, suppose that a in Case_1 and a' in Case_2 can be merged, as can b and b', and d and d', respectively. Once merged, Case_1 and Case_2 can be stored in the decomposed-and-merged form shown in Figure 1.3(b).

Storing solved cases in the decomposed-and-merged form raises the issues of how to decompose and how to merge. The issue of how to decompose is addressed by decomposing a solved case, more specifically a causal graph representing the most likely causal explanation, such that each component of the decomposed case captures the context sensitivity of findings. The basic approach taken to address the merging issue is as follows: If components are instantiations of the same underlying pathophysiologic mechanism of a disorder, then these components, from different solved cases, can be merged to produce a coherent combined whole.

By remembering solved cases in a decomposed form, parts of cases can be accessed and used more easily. By remembering similar cases in a merged form, resources such as memory space and processing time can be used more efficiently.

The primary mechanism investigated in this thesis, for remembering components of decomposed diagnoses, is the diagnostic-unit representation. The diagnostic-unit representation attempts to capture the context sensitivity of findings by explicitly storing, in separation from other plausible causal relations, the most likely causal relations between disorders and sets of findings. Thus, in the diagnostic-unit representation all known causal relations are most likely. The diagnostic-unit representation uses a general graph representation to efficiently specify the often complex underlying causal relationships that link disorders and sets of findings. Each diagnostic unit is a general graph such that the set of findings identified in the graph, as a whole, strongly suggest the identified disorder and underlying causal mechanism.

This thesis investigates experience-based strategies for incrementally deriving and updating diagnostic units and the various relationships between

them.

How to Use: Diagnostic Units as Constraint for Evidence Grouping

The knowledge base that consists of diagnostic units is a *categorical* knowledge base that contains only the most likely causal associations between disorders and sets of findings. Diagnostic units represent *grouping constraints* stating how a given set of findings can be grouped into subproblems for which the most likely explanations can be concluded immediately. In the diagnostic-unit representation paradigm, therefore, the grouping of given evidence into subproblems can be reduced to a search for relevant diagnostic units. This thesis addresses the issue of how to determine relevant diagnostic units, by matching diagnostic units against the given evidence.

Unfortunately, in such medical domains as heart failure, patients with the exact same set of findings rarely occur. In addition, different patients can manifest different findings, even when they are suffering from the same disease and underlying pathophysiologic mechanism. Since diagnostic units used in this thesis are acquired from diagnostic experience, knowledge captured in diagnostic units is generally imperfect. As a consequence, it is generally necessary to translate, or "adapt," existing diagnostic units so that they match a new problem. This thesis develops an approximate technique for performing this adaptation. The technique is based on a notion of *deep matching* which matches diagnostic units against a problem at the causal level as well as at the finding level. The goal is to increase the usability of diagnostic units in later diagnosis.

In domains where diagnosis is based on categorical knowledge, *plausibility criteria* are required to select the best hypotheses [77]: A hypothesis is a set of disorders, or faulty components, which can explain all the given findings when considered together. A common plausibility criterion used in most recent diagnostic algorithms is *minimality* [23, 82, 83, 101, 102]. A hypothesis is minimal when none of its subsets can account for all of the findings. In

general, plausibility criteria such as minimality identify some plausible solutions under the disorder independence assumption or the assumption of set additivity of observations [73]. Such assumptions reduce computational complexity. Accuracy is traded for efficiency, however, by not taking into consideration the possibility that disorders and findings can interact so as to mask or support each other. This thesis attempts to improve accuracy without unduly degrading efficiency. Toward this end, a similarity metric which takes the specificity of findings into account is developed for gauging the similarity between a problem and a diagnostic unit. In addition, dependency among disorders is used to guide the process of combining highly similar diagnostic units into an overall solution. By doing so, interactions between disorders and findings are taken into consideration.

Experimental results indicate that the diagnostic-unit representation effectively captures domain decomposability. They also support that the techniques developed in this thesis, for using diagnostic units, are able to effectively exploit this domain decomposability when performing decompositional abductive diagnosis.

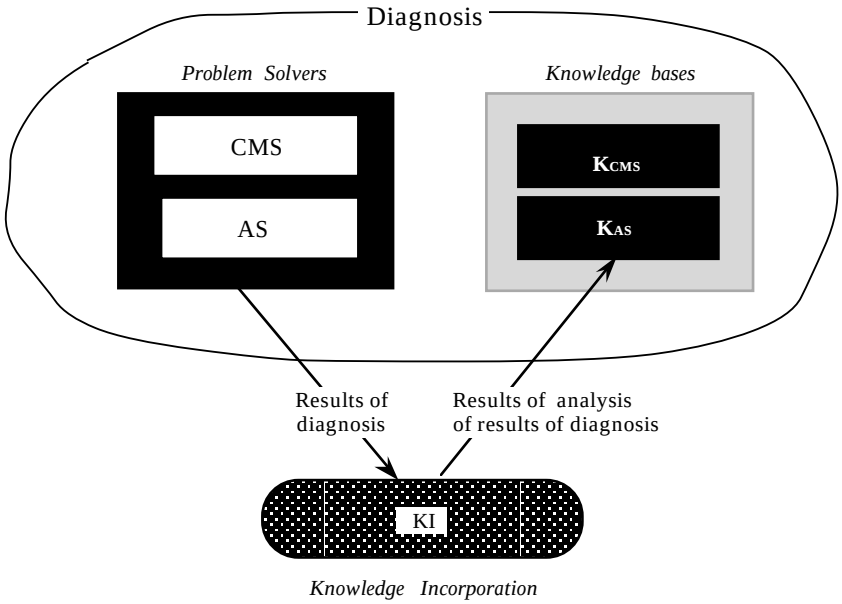
1.4 Hybrid Diagnostic System with Feedback

To facilitate further discussion, this section describes the logical structure of a hybrid diagnostic architecture designed explicitly to support feedback between the problem-solving and knowledge-acquisition processes.

A diagnostic system is said to be a hybrid diagnostic system if it is capable of performing a diagnostic task both associatively and from first principles. Furthermore, a diagnostic system is said to contain feedback if it is capable of incorporating its own diagnostic experience into its knowledge base.

1.4.1 HYDI

HYDI is a hybrid diagnostic system with feedback. It diagnoses multiple disorders without assuming that disorders are independent, and automatically acquire knowledge about the context sensitivity of findings from its own problem-solving experience. Figure 1.4 summarizes the basic architecture of HYDI.



CMS: Causal-model-based reasoner of HYDI (HF)
AS: Association-based reasoner of HYDI (Decompositional abductive diagnosis)
KCMS: Causal knowledge base for use by CMS
KAS: Associative knowledge base for use by AS
KI: Knowledge incorporator of HYDI

Figure 1.4: Architecture of HYDI

Problem-Solving Component of HYDI

The problem-solving component of HYDI consists of the causal-model-based problem solver, CMS, and the association-based problem solver, AS. The

intention is to take advantage of the robustness of causal-model-based reasoning and the efficiency of association-based reasoning. Since the focus of this thesis is not on the specifics of causal-model-based reasoning, HYDI uses an existing probabilistic causal-model-based system, HF, for CMS. The HF program is a probabilistic causal-model-based diagnostic system which automates, from first principles, heart failure diagnosis in the human cardiovascular system [62, 61]. For AS, this thesis develops a decompositional abductive diagnosis technique that exploits knowledge structures in the diagnostic-unit representation.

Knowledge Bases of HYDI

HYDI has two knowledge bases. One is the causal knowledge base, K_{CMS}, that contains domain causal knowledge. It is used mainly by CMS. The other is the associative knowledge base, K_{AS}, in which associative knowledge, specifically knowledge about the context sensitivity of findings, is stored for use by AS.

Initially, HYDI has no associative knowledge, but is able to solve diagnostic problems from first principles. As HYDI solves various types of diagnostic problems, however, its associative knowledge base evolves away from its initial empty state. HYDI automatically acquires knowledge about the context sensitivity of findings, by analyzing and accumulating its own problem-solving experience in the decomposed-and-merged form. As the associative knowledge base grows with experience, HYDI can solve familiar kinds of problems more efficiently.

Hybrid Reasoning for Diagnosis

HYDI performs hybrid reasoning to solve a diagnostic problem. HYDI's basic hybrid-reasoning flow is as follows. Given a diagnostic problem, AS first tries to solve the problem. If it fails to generate an acceptable solution to the problem, then the more robust CMS is called to solve the problem

step-by-step from first principles. As the associative knowledge base grows with experience, CMS is expected to be called less to solve a problem from first principles. From the cognitive perspective, it is worth noting that this problem-solving behavior is analogous to physicians. When students graduate from medical school, they presumably possess considerable amounts of medical domain knowledge. Their wealth of knowledge does not seem to be effectively activated, however, when they are confronted by real-world problems. In general, many years of additional training are required before students are considered "ready to practice." In contrast to "fresh" students who generally use step-by-step reasoning strategies, those experts in the areas of the patient's diseases appear able to recognize relevant pathophysiologic details and zero in on the target immediately [1, 3, 10, 25, 27, 47, 48]. So much so that an observer might feel that no detailed reasoning is involved in their thought processes at all. From a pragmatic standpoint, the use of association-based diagnosis supported by a causal model of knowledge reduces the possibility of failing to make a reasonable diagnosis (even though there is the potential problem that AS incorrectly thinks it can solve.).

Knowledge Incorporator

A goal of this dissertation research is to develop an experience-guided mechanism for abstracting "expertise" that allows a diagnostic system to get to believed-to-be good solutions without search. Such general problem-solving rules for diagnosis are acquired from experience and stored in K_{AS} by the knowledge incorporator. Whenever a diagnostic problem is solved, the knowledge incorporator analyzes and decomposes the results of the diagnosis (specifically, the most likely causal explanations for the given set of findings). The results of the analysis are potential diagnostic units. The potential diagnostic units are then used to refine existing knowledge in K_{AS}. Two diagnostic units are merged to produce a combined whole if they represent the same underlying pathophysiologic mechanism of a disorder. It is the result of merging that is stored in K_{AS}. Such dynamic incorporation of

new experience into K_{AS}, prior to beginning a new problem-solving cycle, allows the adjustment to feed forward to future diagnosis. The diagnostic-unit representation allows the perturbation caused by any change in K_{AS} to be localized only to affected diagnostic units.

In general, the knowledge incorporation process attempts to abstract diagnostic units whenever possible. The cost of abstraction is that omitted causal details are sometimes important for problem solving. In HYDI which is supported by a causal knowledge base, however, this cost is relatively insignificant: Omitted details can be made available from the causal knowledge.

1.5 Guide to the Thesis

Chapter 2 describes previous work related to the work described in this thesis. The hybrid reasoning investigated in this research uses two types of knowledge representation: a causal representation and a diagnostic-unit representation. Chapter 3 describes a causal representation to model medical domain principles for diagnosis. It also examines features of diagnostic experience that suggest a possibility of gradual improvement in overall performance. Chapter 4 discusses the diagnostic-unit representation for remembering the results of previous diagnoses. This thesis uses diagnostic experience to acquire and incrementally update diagnostic units and the relationships between them. Chapter 5 addresses issues that arise in incorporating new diagnostic experience into an existing body of experiential knowledge. Chapter 6 describes a decomposition-based abductive diagnosis method that exploits the domain structure identified through diagnostic experience. Chapter 7 presents the results of experiments, conducted to test the effectiveness of the diagnostic-unit representation and the techniques developed for using diagnostic units in decompositional abductive diagnosis. Chapter 8 summarizes this research, and suggests future directions which this research could extend to.

Chapter 2

Related Work

2.1 Diagnosis of Multiple Disorders

Gorry addressed the problem of multiple disorders based on "pattern-sorting function [34]." The pattern-sorting function groups findings into clusters. A finding cluster is considered to be valid if there exists a disorder that can explain all the findings in the cluster. The program decides sequentially which clusters to pursue based on probability and utility.

Another system for diagnosing multiple disorders is INTERNIST [42, 65, 66]. INTERNIST is intended to deal with the entire scope of general internal medicine. Symptoms and diseases are matched based on forward and backward conditional probabilities. INTERNIST identifies the disease with the highest score based on these probabilities, and then builds a differential diagnosis around this disease. The differential diagnosis is the list of diseases that are competing explanations for the same findings. The top-ranked disease in the differential diagnosis is chosen. If the chosen disease cannot explain all the given findings, another differential diagnosis is formed based on the unexplained findings. This process continues until all the given findings are explained.

One problem with INTERNIST's sequential approach is that it does not handle interacting disorders and findings appropriately. After concluding a disease, INTERNIST removes all the findings that can be explained by the disease from further consideration. This can be particularly problematic, if removed findings may actually be caused by other disease. In general, INTERNIST's sequential approach is effective, in domains where each finding is caused by a single disorder, and disorders are independent.

The CADUCEUS program [79] addresses the deficiencies in INTERNIST by using a combined hierarchical-causal network. The hierarchies contain pre-specified differential diagnoses. Causal links connect physiologically related disease categories to different hierarchies. Differential diagnoses are triggered by causal links. CADUCEUS derives a hypothesis for a problem, by triggering multiple differential diagnoses in the hierarchies and then following various subsumption and causal relationships. A main problem with CADUCEUS is that it depends on differential diagnoses that are hierarchically structured in advance. In ill-structure domains like medical diagnosis, however, the task of organizing differentials cleanly is a difficult problem in itself.

SYNOPSIS is a system for diagnosing multiple disorders in the domain of internal medicine [102]. It generates a plausible hypothesis, by finding a plausible candidate. A candidate is a set of clusters of findings such that the union of all the clusters in the set is equal to the set of findings presented for diagnosis, and for any two clusters in the set their intersection is an empty set. A candidate is plausible if all findings in a cluster must be explainable by the same single disorder, and each cluster must be explained by a unique disorder that does not explain another cluster. A plausible candidate is found by decompositional search. Decompositional search is done on the search space which is computed dynamically for each problem, based on known causal relations between diseases and findings. Each causal relation is represented in the bipartite form of "Disorder $d \rightarrow$ Finding f," where a link represents

that disorder d can cause finding f.

The current implementation of SYNOPSIS does not make an attempt to discern the difference in strengths of associations between diseases and findings. While this greatly simplifies computational requirements, there is no guarantee that a hypothesis found by SYNOPSIS is the most likely diagnosis. In addition, SYNOPSIS groups findings into clusters, implicitly assuming the context insensitivity of findings. It treats findings as atoms, and examines them one at a time independently of each other. As a result, findings f and f', each of which can be caused by d, can be grouped together around d, even though the simultaneous occurrence of f and f' could lower the likelihood d as a cause. SYNOPSIS can be extended by adding probability to take the difference in the strength of a disorder in producing a particular finding. It then must compute a probability for each plausible candidate. The computation of a probability of a candidate is proven to be expensive [102], however, requiring time that is exponential in the number of findings in a case.

More rigorous approaches to multidisorder diagnosis appear in ABEL and HF. ABEL is a program for acid-base and electrolyte disorders [72]. It divides physiological knowledge along different levels of detail, including clinical, physiological, and biochemical levels. It reasons about interactions between diseases at different levels of abstraction. HF [61, 62] is another system that reasons with detailed pathophysiological knowledge. HF makes a diagnosis, using a probabilistic causal reasoning technique. Both ABEL and HF require detailed physiologic knowledge to solve problems, and thus are generally suited to domains with well understood physiology.

2.2 Combining Association-based and Causal-model-based Reasoning

Several attempts have been made to combine association-based and causal-model-based reasoning, to take advantage of their complementary strengths.

For example, GORDIUS combines association-based and causal reasoning techniques under the Generate, Test, and Debug (GTD) paradigm [91]. Association-based reasoning is used to generate a plausible hypothesis. Causal reasoning is used to test and debug the plausible hypothesis.

Goel's computational model uses association-based and model-based reasoning to design a physical device [33]. Association-based reasoning, more specifically case-based reasoning, is used to identify candidate components of the device. Then, model-based reasoning is then used to modify the candidate components so that the components together produce desired behavior of the device.

CASEY is another system which combines association-based and causal-model-based reasoning for diagnosing heart failure [55]. Association-based reasoning, more specifically comparison-based case-based reasoning, is performed to generate a hypothesis. If a diagnosis cannot be made by association-based reasoning, then a causal-model-based reasoning system, specifically HF, is called to solve the problem.

2.3 Knowledge Acquisition

Much of research concern so far has been centered around the development of problem-solving mechanisms for dealing with computational complexity in a manageable way. As a result, while the power of a knowledge-based system derives from domain-specific knowledge, the concern about how to fill out knowledge entries in a knowledge base has been relatively inactive. In most of existing systems for medical diagnosis, knowledge was compiled, typically by knowledge engineers, through expensive lengthy interviews with domain

experts. In turn, compiled knowledge was encoded into a system, which is a time-consuming and painstaking task. Not only the construction cost, but also costs due to human errors make the labor-intensive way of knowledge acquisition unattractive. It is very unlikely that all of knowledge required in problem-solving can be known *a priori* and entered without errors for use by a program [46, 47]. This might explain why there seem to be many people who have more confidence in the medical diagnosis made by "expert physicians" combining facts in a heuristic manner than in the diagnosis emanating from a large system which operates on knowledge compiled from a person by another person.

The automatic acquisition of knowledge becomes increasingly important, as knowledge-intensity of a domain increases. Problem-solving experience offers a valuable source for knowledge that allows efficient problem solving. Effective utilization of experiences in any organized fashion has emerged as a pressing concern to overcome the drawbacks of one-time labor-intensive knowledge acquisition. Early attempts at learning can be traced back to Arthur Samuel's checkers-playing program. His program not only played games with its opponents but also exploited experience at games to improve its later performance [87]. The process of bringing learning into a more central position in Artificial Intelligence was accelerated by Winston's work on blocks-world learning [100]. Such issues as when, where, and how to remember for effective knowledge acquisition have also received researchers' attention [10, 31]. Attempts have been made to acquire rules inductively from particular examples. Inductive learning algorithms enable a system to automatically extract general rules [4, 80], expertise [9], concepts [6, 39, 40], or structures [8, 41] from externally supplied examples. Another area of learning is the learning of control knowledge to augment the efficiency of problem solving [3, 68, 69]. A system attempts to learn a sequence of operators that can be grouped as a macro operator. A macro operator is treated as a single operator by a problem-solving algorithm. By reducing a search space, chunking of knowledge can improve efficiency of problem

solving [12, 29].

2.4 Explanation-Based Learning

It is possible to use explanation-based learning to learn disease concepts. Explanation-based learning is a technique for learning a concept without the use of many training examples [22, 68]. It consists of two steps: generation of an explanation for a given example, followed by generalization of the explanation. A domain theory is used to generalize the explanation, and it is the domain theory that makes it possible to learn a concept from only a few training examples. The direct application of explanation-based learning to disease concept learning, however, would be problematic without appropriate extensions to handle the following difficulties.

First, explanation-based learning is generally only effective in domains where pruning is simple. If a domain theory consists of causations where each effect has a single certain corresponding cause, the generation of an explanation is a relatively simple task: The cause of an effect can be identified with certainty. If there are many "levels" of uncertain multiple causations, however, exponentially many potential explanations need to be considered to generate a good explanation. The issue of how to deal with the potential combinatorial explosion that arises in generating an explanation emerges as an overriding concern.

Second, explanation-based learning generally assumes that a complete domain theory is available. This assumption is motivated by the desire to generalize explanations into provably correct ones. It is hard, if even possible, to come by a complete domain theory in ill-understood domains like medicine. In such domains, learning empirically justifiable experience is likely to be more sensible.

Third, knowledge learned using explanation-based learning can generally only be applied when exact matches occur. Within the context of medical diagnosis, this restriction implies that a learned disease concept cannot be

used in diagnosis, unless a diagnostic problem matches the concept exactly. In most medical domains, patients with the exact same findings rarely occur. This characteristic of medical diagnosis suggests that partial matching may have greater utility.

Finally, much of explanation-based learning focuses primarily on learning concepts one at a time, independently of each other. In most medical domains where multiple disorders are not uncommon, however, a disease can have different descriptions, depending on what other diseases occur with it. Dependencies among concepts have to be handled appropriately if useful disease concepts are to be learned.

2.5 Integration of Problem solving and Learning

What to learn is not independent of the goals of problem solving. In the light of the interrelationship between problem solving and learning, systems such as SOAR coupled learning to problem solving. SOAR is a rule-based general-purpose problem-solving system which is integrated with explanation-based learning [57]. An explanation-based learning component of SOAR analyzes explanations and chunks macro rules that summarize the explanations [22, 57, 49, 85].

PRODIGY [68] is another general purpose problem-solving system which is integrated with learning mechanisms. Much of learning in PRODIGY is directed at automatically acquiring control rules from experience. The goal is to improve efficiency of a search process. PRODIGY differs from SOAR in that it attempts to learn from its failure as well as success. If it pursues an unsuccessful path, PRODIGY tries to come up with an explanation of reasons for the failure. This explanation is then used to construct control rules that will help PRODIGY avoid pursuing unpromising search paths in later problem solving.

CASEY [55] is a diagnostic system that acquires associative knowledge from experience to improve later diagnostic performance. CASEY is a case-based, more specifically comparison-based, reasoning system grounded on HF. CASEY remembers, individually as independent atoms, cases that have been successfully solved for its own use in future problem solving. Each stored case represents an associative rule that links a problem situation to its solution. Solved cases are stored in a MOP (Memory Organization Packet) structure [86]. The number of memory structures grow exponentially in the number of cases. Although a hashing scheme could be used to locate previous cases relevant to a new case, storing exponentially many generalizations and cases poses serious difficulties [2]. CASEY solves a problem, by directly inspecting old cases to find best matches against the problem. Because all heart failure cases involve multiple, interacting diseases, diseases and findings can combine to form innumerable problem situations. It is difficult to find two patients with exactly same set of findings. This can adversely affect problem-solving performance of CASEY which relies on the routine recurrence of similar cases.

2.6 Case-based Reasoning

Case-based reasoning is an approach to perform problem-solving assignments based on a collection of stored previous episodes [14, 30, 36, 50, 51, 52, 54, 81, 84, 94]. The use of previous experience in future problem solving raises the issue of how to store previous cases. A common approach, investigated in case-based reasoning, is typically to store previous cases individually as independent atoms. Stored cases represent relatively low-level knowledge for solving problems which have been encountered previously.

Indexing schemes are required to access stored episodes during future problem solving. A common indexing scheme, investigated in the literature, often uses pre-determined features for indexing. In complex domains, like design or medical diagnosis, however, it is often hard to identify primary fea-

tures *a priori*, because cases which are dissimilar on the surface may be more relevant to a problem than ones which are similar in appearance. To address the problems of appearance-based indexing, attempts have been made to retrieve cases based on generalized indices – for instance, relevant indices determined by explanation-based learning [7], underlying common causes [55], and abstract indexing vocabulary specified in a hierarchy of normative causal interactions [96].

Once best matches are located, they can be adapted, if necessary, to fit the current problem. For example consider CHEF, a case-based planner applied in the domain of cooking [36]. It adapts the retrieved recipe to meet current requirements, by applying a series of modification strategies. Adaptation strategies are generally built around domain-specific heuristics. In complex domains, however, the adaptation of retrieved cases based on heuristic associations and search can be computationally expensive. Computational burdens can be relieved with an reductionistic approach, by decomposing an adaptation space into smaller adaptation spaces. In Goel's computational model for designing physical devices, for example, "modification-generation plans" are used for decomposing an adaptation space to fix behavioral difference between a new problem and a retrieved one [33]. Each plan, indexed by a specific type of behavioral differences, identifies candidate components which need structural modifications. Qualitative behavior-structure models are used for adaptation. Each behavior-structure model specifies how output behaviors are produced by the structure of a design.

A case-based reasoning technique can be used in conjunction with a problem-decomposition mechanism. When presented with a new problem, the issue that must be dealt with is how to decompose the given problem. Problem-decomposition techniques, explored in the case-based reasoning literature, are generally examined in relatively highly structured domains where how to decompose a problem can be specified *a priori* with relative ease [44, 58, 81, 96]. In ill-structured domains like medical diagnosis, however, pre-specification of how to decompose a problem is often difficult. Such

difficulty appears reflected in previous applications of case-based reasoning to medical diagnosis: Comparison-based reasoning that transfers an entire solution is usually used [52, 55]. Comparison-based reasoning transfers an entire solution to a previous case for a new case, by modifying it based on differences between the new and the previous.

Chapter 3

Causal Knowledge Representation

This chapter formally describes a causal representation of pathophysiologic knowledge. It also discusses the complexity of diagnosis based on the causal representation. While causal representation is not new [62, 75, 78], the formal investigation of a causal representation of medical knowledge helps understand what diagnostic experiences are, and thus facilitates the discussion of representational and acquisitional approaches explored in this thesis.

3.1 Pathophysiologic Knowledge in Causal Representation

A causal knowledge base for medical diagnosis contains pathophysiologic knowledge that describes the malfunctioning of a human body. This section describes a way of modeling pathophysiologic knowledge in terms of causal relationships between clinical entities. Clinical entities include elemental disorders, intermediate states, and findings, which were described in Section 1.2.2. To facilitate discussion, let $\mathcal{D}$ denote a set of variables that represent elemental disorders, $\mathcal{I}$ a set of variables that represent intermediate

states, and $\mathcal{F}$ a set of variables that represent findings. Let $\mathcal{U}$ denote the set of variables that represent all clinical entities: In other words, $\mathcal{U} = \mathcal{D} \cup \mathcal{I} \cup \mathcal{F}$.

An additional notation is introduced to describe value assignments for variables. For any variable x in $\mathcal{U}$, let V_x denote a set of values that can be assigned to x. For any v in V_x, value assignment $x := v$ (for example, aortic stenosis := *present*) represents the instantiation of the value of x as v. An instantiated variable is a variable with a value assigned to it. A *finding* is an instantiated finding variable in $\mathcal{F}$, an *elemental disorder* an instantiated elemental disorder variable in $\mathcal{D}$, and an *intermediate state* an instantiated intermediate state in $\mathcal{I}$. This thesis assumes that for any variable x in $\mathcal{U}$, its values can be classified into two classes: normal and abnormal.

Causal mechanisms in a human body are often uncertain. Causal relationships between clinical entities are thus uncertain. Such uncertainty can be modeled in probabilistic terms with a particular probability being used to represent the degree of belief in a causal dependency between clinical entities. An uncertain direct causal relation can be represented as follows.

Notation (Direct causal relation): For any $a \in \mathcal{D} \cup \mathcal{I}$ and any $b \in \mathcal{U}$ such that a is a direct cause for b, $a \xrightarrow{Pr(b|a)}_c b$ denotes a direct causal relation from a to b, where given that a (and nothing else) occurred, a can cause b with probability of $Pr(b|a)$.

Conditional probabilities associated with each causal link represents the strength of a cause in producing a particular effect. According to probability combining rules based on a noisy-or assumption, it can be computed, from the evidence collected, how likely it is that various disorders are present. (See [62] for details.)

Note that findings only occur as effects in direct causal relations. Findings, however, can cause states or other findings. Like HF, this thesis handles such findings as if they are also pathophysiological states by creating corresponding pathophysiologic states. More specifically, consider a finding f with n effects a_1, a_2, ..., and a_n: Each a_i is either a finding or a pathophysiological

state. These n causal relations are depicted in Figure 3.1(a). For finding f, a variable i representing a corresponding pathophysiologic state is created (as an element of either $\mathcal{D}$ or $\mathcal{I}$ accordingly). Then, i is defined as a clinical entity with $n+1$ effects – more specifically, finding f and its n effects, as shown in Figure 3.1(b). A direct causal link from i to f is established. In addition, a direct causal relation from finding f to its effect a_i is represented as a direct causal relation from i to a'_i, where a'_i is a_i if a_i is a state, or is a corresponding pathophysiologic state of a_i if a_i is a finding.

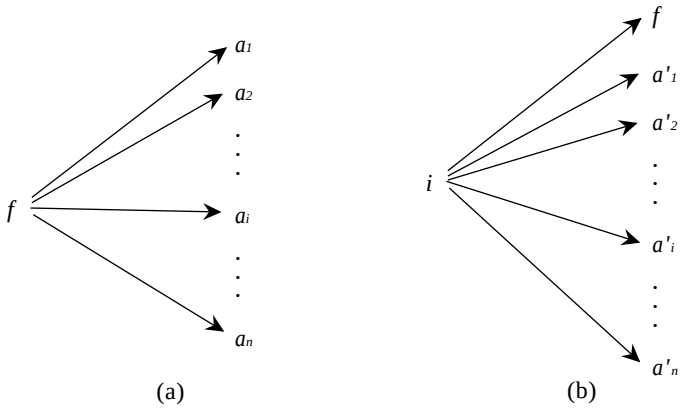


Figure 3.1: Representation of a finding f with n effects: For all $i = 1, 2, \ldots, n$, a'_i is a_i if a_i is a state, or is a corresponding pathophysiologic state of a_i if a_i is a finding.

Finding f may have causes as well. Suppose that f has m causes $b_1, b_2, \ldots,$ and b_m, as shown in Figure 3.2(a). Each cause b_j of f is handled as a state causing i, rather than f. More specifically, the causal relation from b_j to f is represented as the causal relation from b'_j to i, where b'_j is b_j if b_j is a state, or is a corresponding pathophysiologic state of b_j if b_j is a finding. These causal relations are shown in Figure 3.2(b).

To provide an intuitive understanding, consider a finding on-aortic-valve-replacement which can cause aortic stenosis (an elemental disorder). Figure 3.3(a) depicts a causal dependency between aortic stenosis and on-aortic-

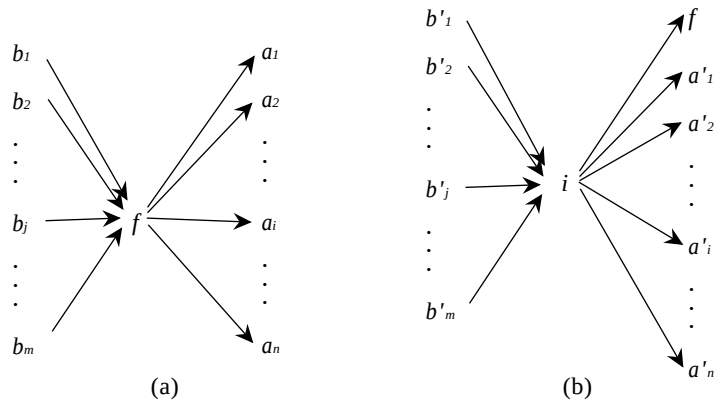


Figure 3.2: Representation of a finding f with m causes: For all $j = 1, 2, \ldots, m$, b'_j is b_j if b_j is a state, or is a corresponding pathophysiologic state of b_j if b_j is a finding.

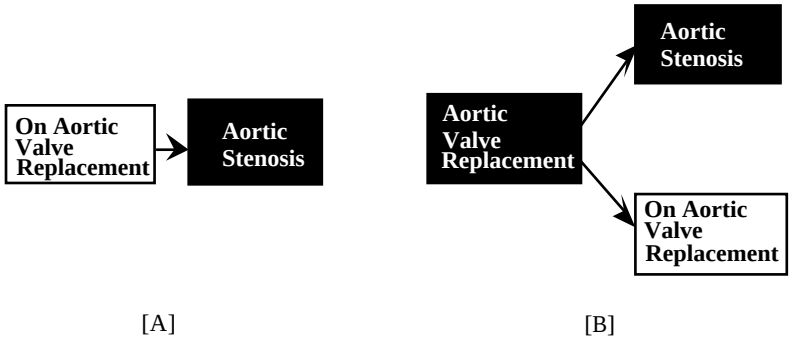


Figure 3.3: Representation of a causal relation between the finding on-aortic-valve-replacement and its cause aortic stenosis. A black rectangular node represents an elemental disorder variable, and a unfilled rectangular node represents a finding variable.

value-replacement. Figure 3.3(b) shows how this causal relation is modeled in this research. Instead of a direct causal link from the node representing on-aortic-value-replacement to the node representing aortic stenosis, a pathophysiologic state that represents aortic valve replacement is created, and two direct causal links are established from the node representing this

pathophysiologic state to the node representing on-aortic-value-replacement and to the node representing aortic stenosis, respectively. The modeling of findings that can cause states or other findings in this way guarantees that findings always appear as effects.

Let $\mathcal{L}$ be the set of all possible direct causal relations, *i.e.*, $\mathcal{L} = \{a \xrightarrow{Pr(b|a)}_c b \mid a \in \mathcal{D} \cup \mathcal{I}, b \in \mathcal{U}$ and $\exists$ a direct causal relation from a to $b\}$. Pathophysiologic knowledge modeled in terms of direct causal relations can be conceptualized as a causal graph of nodes and links, where each node represents a clinical entity variable,[1] and each link represents a direct causal relationship between the two clinical entities represented by the nodes it connects. To facilitate further discussion, this thesis adapts a standard graph notation [16] to define a causal graph.

Notation ($\mathcal{L}_{|D,I,F}$): For any $D \subseteq \mathcal{D}, I \subseteq \mathcal{I}$, and $F \subseteq \mathcal{F}$, let $\mathcal{L}_{|D,I,F}$ denote the set of all possible direct causal links between elements in D, I, and F. In other words, $\mathcal{L}_{|D,I,F} = \{a \xrightarrow{Pr(b|a)}_c b \mid a \in D \cup I, b \in D \cup I \cup F$ and $\exists$ a direct causal relation from a to $b\}$.

Definition 1 (Causal graph): A causal graph G is a list (D, I, F, L), where $D \subseteq \mathcal{D}, I \subseteq \mathcal{I}, F \subseteq \mathcal{F}$, and $L \subseteq \mathcal{L}_{|D,I,F}$. D, I, F, and L are called the disorder set, intermediate state set, finding set, and causal link set, respectively, of G.

A causal graph is a collection of direct causal relations between clinical entities. Let $\mathcal{CG}$ be the universe set of all causal graphs: In other words, $\mathcal{CG} = \{(D, I, F, L) \mid D \subseteq \mathcal{D}, I \subseteq \mathcal{I}, F \subseteq \mathcal{F}$, and $L \subseteq \mathcal{L}_{|D,I,F})\}$.

Throughout this proposal, the symbol $\mathcal{C}$ is used to denote a causal graph that represents pathophysiologic knowledge defined in a medical domain of concern. In other words, $\mathcal{C}$ is the causal graph $(\mathcal{D}, \mathcal{I}, \mathcal{F}, \mathcal{L})$. Causal graph $\mathcal{C}$ is referred to as a domain causal network, or shortly a causal network.

[1] Except when noted otherwise, a node and a variable are used interchangeably.

The causal graph representation of pathophysiologic knowledge can be made more concrete with an example from the heart failure domain. As an example, consider the causal graph shown in Figure 3.4. The causal graph is a part of a domain causal network that models malfunctioning of the human cardiovascular system. Each black rectangular node represents an elemental disorder variable in $\mathcal{D}$, an oval node an intermediate state variable in $\mathcal{I}$, and a rectangular node a finding variable in $\mathcal{F}$. Even though for convenience causal links have not been labeled with conditional probabilities, corresponding conditional probabilities are associated with each causal link.

As shown in the example causal graph, a finding can be caused by more than one disorder. For example, nocturnal dyspnea[2] can be caused by any combination of aortic valve replacement, aortic stenosis, aortic valve disease, acute as well as chronic mitral regurgitation,[3] and mitral stenosis.[4] Figure 3.4 also illustrates that a disorder can produce a finding through more than one causal path – with different degrees of likelihood. For example, there is more than one causal path from aortic stenosis to nocturnal dyspnea.

3.2 Diagnosis of Multiple Disorders based on Causal Knowledge

This subsection discusses diagnosis from first principles modeled in terms of causal relations between clinical entities. It also attempts to touch upon some related computational issues.

Additional definitions are introduced below, to facilitate further discussion.

Definition 2 (Instantiated causal graph): For any causal graph $G \in \mathcal{CG}$,

[2]Dyspnea is a shortness of breath.

[3]Mitral regurgitation is backflow of blood from the left ventricle into the left atrium due to failure of the value to close completely.

[4]Mitral stenosis is narrowing orifice of the mitral valve obstructing free flow from atrium to ventricle.

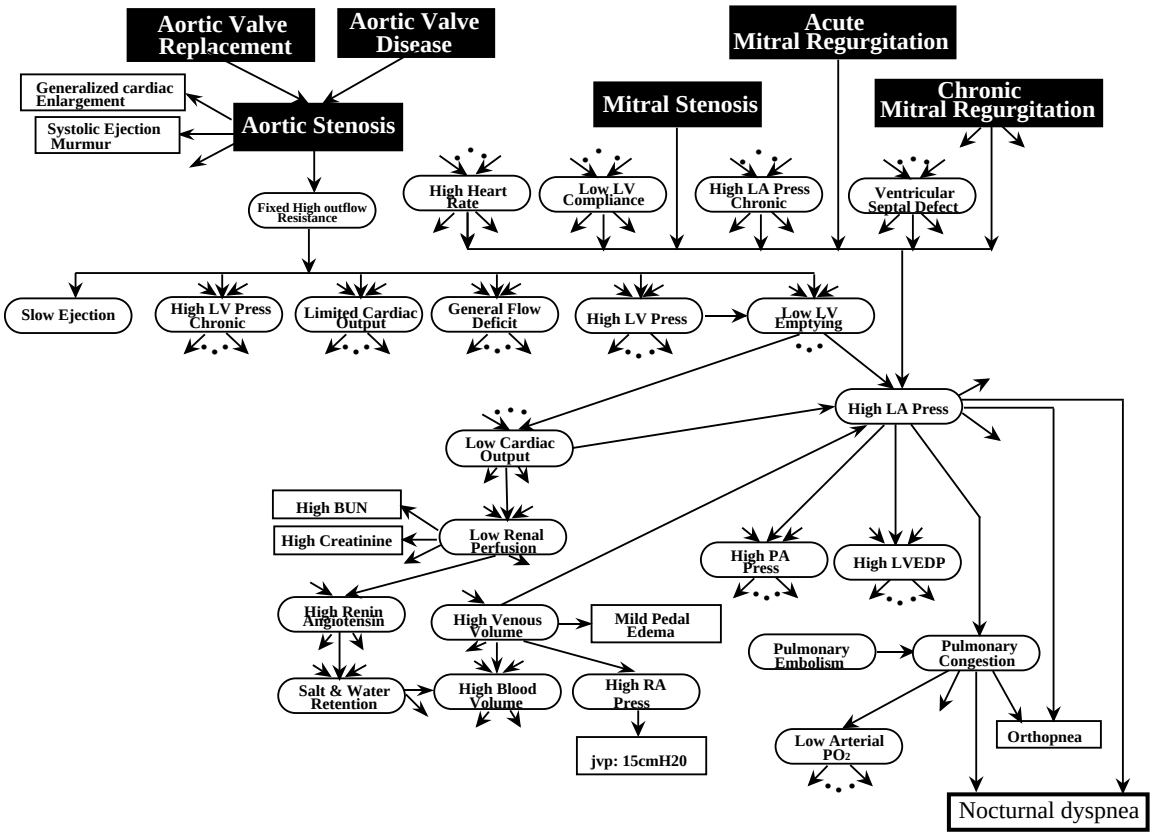


Figure 3.4: A part of a domain causal network that models malfunctioning of the human cardiovascular system

> G is said to be instantiated if each and every node in G is instantiated to a certain value.

Similarly, an instantiated set is defined as follows:

Definition 3 (Instantiated set): For any variable set X, X is said to be instantiated if each and every variable in X is instantiated to a certain value.

Definition 4 (Equality of instantiated sets): For any two instantiated variable sets X and Y, X and Y are said to be equal if

1. for any variable $z \in X \cap Y$, the value assignment of z in X is equal to that of z in Y,
2. for any variable $x \in X$, $x \in Y$, and
3. for any variable $y \in Y$, $y \in X$.

This thesis uses the symbol ψ to denote a diagnostic problem, *i.e.*, a set of findings collected for diagnosis: ψ is an instantiated subset of $\mathcal{F}$. Diagnostic problem solving is the task of identifying the most likely body state producing the findings in ψ. Each instantiation of the domain causal network $\mathcal{C}$ corresponds to a particular body state. Diagnostic problem solving thus becomes the task of identifying the most likely instantiation of $\mathcal{C}'$ that explains the findings in ψ, where $\mathcal{C}'$ is a causal graph such that

1. $\mathcal{C}'$ is a subgraph of $\mathcal{C}$ that is obtained by removing from $\mathcal{C}$ nodes that represent findings not in ψ, along with any direct causal links to these nodes, and
2. The finding set of $\mathcal{C}'$ is an instantiated set equal to ψ.

$\mathcal{C}'$ is a domain causal network which is tailored to diagnostic problem ψ. For ease of exposition, we may call $\mathcal{C}'$ the tailored causal network for ψ. Note that the modeling, as pathophysiologic states, of findings that can cause other states or findings guarantees that all findings in $\mathcal{F}$ always appear as leaves in $\mathcal{C}$. As a result, the finding removal process to produce a tailored causal network can be performed without considering issues of how to remove nodes with children.

To provide an intuition on a tailored causal network, consider the fictional domain causal network $\mathcal{C}$ shown in Figure 3.5. Formally, $\mathcal{C}$ is the causal graph $(\mathcal{D}, \mathcal{I}, \mathcal{F}, \mathcal{L})$, where

$\mathcal{D} = \{d_1, d_2, d_3\}$

$\mathcal{I} = \{i_1, i_2, i_3, i_4, i_5, i_6\}$

$\mathcal{F} = \{f_1, f_2, f_3, f_4, f_5, f_6, f_7\}$

$\mathcal{L}$ = a set of direct causal links shown in Figure 3.5.

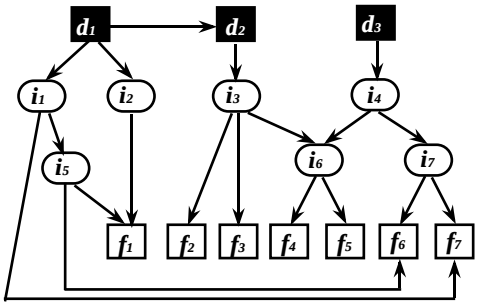


Figure 3.5: A fictional domain causal network $\mathcal{C}$ used for expository purposes

For expository purposes, it is assumed that each node in the example $\mathcal{C}$ is a binary variable that can take on the value of either *present* or *absent*, and that for any variable x in $\mathcal{C}$ *present* is an abnormal value of x, and *absent* is a normal value of x.

Now, let us consider an example diagnostic problem which consists of findings $f_3 :=$ *present*, $f_4 :=$ *present*, $f_6 :=$ *present*, and $f_7 :=$ *present*. In other words,

$$\psi = \{f_3 := \textit{present},\ f_4 := \textit{present},\ f_6 := \textit{present},\ f_7 := \textit{present}\}$$

The tailored causal network $\mathcal{C}'$ for ψ is a causal graph obtained by removing nodes f_1, f_2, and f_3 from $\mathcal{C}$. Figure 3.6 shows the tailored causal network $\mathcal{C}'$.

The following additional notations are introduced to facilitate the discussion of a diagnostic solution.

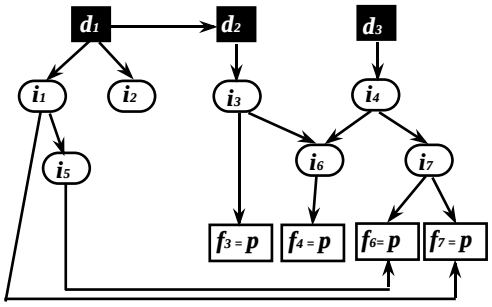


Figure 3.6: A tailored causal network for the diagnostic problem ψ that consists of $f_3 := present$, $f_4 := present$, $f_6 := present$, and $f_7 := present$. p's in the tailored causal network stand for $present$.

Notation: For any causal graph G in $\mathcal{CG}$, let

$d\langle G\rangle$ denote the disorder set of G,

$f\langle G\rangle$ denote the finding set of G,

$i\langle G\rangle$ denote the intermediate state set of G,

$l\langle G\rangle$ denote the causal link set of G,

$di\langle G\rangle$ denote $d\langle G\rangle \cup i\langle G\rangle$, and

$dif\langle G\rangle$ denote $d\langle G\rangle \cup i\langle G\rangle \cup f\langle G\rangle$.

The symbol $\cup$ used in the above notation represents a union of variable sets that denotes the *conjunction*, not logical disjunction, of events asserted by instantiating the variable set union. For example, consider a variable set A which consists of a variable a which is instantiated as v_a, and a variable set B which consists of a variable b with v_b assigned to it. $Pr(A \cup B)$ stands for the probability that a is instantiated as v_a, and b as v_b – in other words, $Pr(a := v_a, b := v_b)$.

Any instantiation of a tailored causal network for a diagnostic problem can be an explanation for the findings presented in the problem.

Definition 5 (Causal explanation): For some instantiated subset F of $\mathcal{F}$, let $\mathcal{C}'$ be a tailored causal network for F. A causal explanation for the findings in F is an instantiation of $\mathcal{C}'$.

A causal explanation for the findings in ψ is an instantiated causal graph, more specifically, an instantiated tailored causal network for ψ.

Due to uncertainty in the domain knowledge itself, it is often difficult to find a solution to a diagnostic problem with absolute certainty. For such problems, probabilities that summarize which disorders are more likely than others can be used to support diagnostic judgments. As mentioned earlier, any instantiation of a tailored causal network for ψ can be an explanation for the findings in ψ. Each of these causal explanations tells us that the pathophysiologic states in the causal explanation can cause the findings accordingly, but not necessarily do so.

The diagnostic task tackled in this thesis is to find the most likely causal explanations for the findings in ψ, in other words, to find the most likely instantiation of the tailored causal network for ψ. A *diagnostic solution* to ψ is defined as a causal explanation $S^{\star}_{\psi}$, for ψ, that satisfies the following qualitative relationship:

> For any causal explanation G for ψ, the findings in ψ are more, or equally, likely to be caused by the disorders in $d\langle S^{\star}_{\psi}\rangle$ via the pathophysiologic mechanism identified by $S^{\star}_{\psi}$ than by those in $d\langle G\rangle$ via the pathophysiologic mechanism identified by G.

This relationship is articulated in the following definition of a diagnostic solution.

Definition 6.1 (Optimal diagnostic solution): For any diagnostic problem ψ, the optimal diagnostic solution to ψ is a causal explanation $S^{\star}_{\psi}$ for ψ such that for any causal explanation G for ψ, $Pr(di\langle S^{\star}_{\psi}\rangle \mid \psi) \geq Pr(di\langle G\rangle \mid \psi)$.

Despite of the use of tailoring, $\mathcal{C}$ and its instantiations are often too large to convey useful information to a user. A user of a diagnostic system is usually more interested in abnormal states that produce given findings. To make a diagnostic solution more informative and insightful, this thesis trims the optimal diagnostic solution in Definition 6.1, by pruning away nodes with normal values. The following defines a reduced optimal diagnostic solution.

Definition 6.2 (Reduced optimal diagnostic solution): For any diagnostic problem ψ, the reduced optimal diagnostic solution to ψ is a maximal subgraph G of the optimal diagnostic solution $S^{\star}_{\psi}$ such that

1. $dif\langle G\rangle$ is a subset of $dif\langle S^{\star}_{\psi}\rangle$ such that every node in G is instantiated to an abnormal value,
2. every node in $dif\langle S^{\star}_{\psi}\rangle - dif\langle G\rangle$ is instantiated to a normal value, and
3. $l\langle G\rangle = \{a \rightarrow b \mid a \in di\langle G\rangle, b \in dif\langle G\rangle$, and $\exists$ a direct causal relation from a to b in $S^{\star}_{\psi}\}$.

Suppose, for example, that the instantiated tailored causal network shown in Figure 3.7 is the most likely causal explanation for $f_3 := present$, $f_4 := present$, $f_6 := present$, and $f_7 := present$.

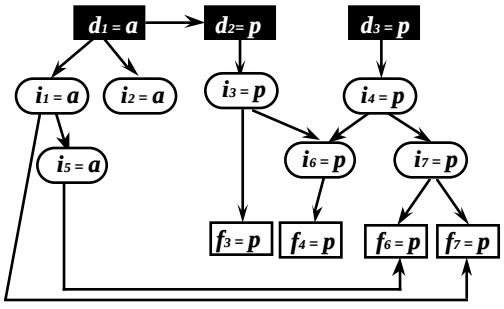


Figure 3.7: The most likely instantiation of the tailored causal network shown in Figure 3.6. p's and a's shown in the instantiated causal network denote $present$ and $absent$, respectively.

Then, the reduced optimal diagnostic solution to our diagnostic problem is the trimmed best instantiation shown in Figure 3.8. For convenience, reduced optimal diagnostic solutions are generally referred to as "optimal diagnostic solutions."

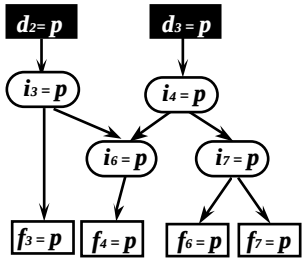


Figure 3.8: The optimal diagnostic solution to ψ obtained by trimming the best instantiation shown in Figure 3.7

Note that an underlying pathophysiologic mechanism, as well as the disorders primarily suspected of causing a given set of findings, is returned as an integral part of a diagnostic solution. The pathophysiologic mechanism in a diagnostic solution can provide additional assurance by explaining how the primary disorders and the findings are related to each other. In addition, the underlying pathophysiologic mechanism in a diagnostic solution makes it easier to reason about the solution, and determine if the solution found by a program "makes sense." Finally, the underlying pathophysiologic mechanism in a diagnostic solution makes it easier to detect possible sources of flaws in a diagnostic solution. This will, in turn, help confirm or refine the correctness of the knowledge base used by a program.

Unfortunately, a causal network can be instantiated in exponentially many ways, and thus there are exponentially many causal explanations to consider, in order to find the most likely one. The exhaustive enumeration-and-evaluation approach is likely to be unwieldy. Fortunately, the search can be reduced by using heuristics. For example, in order to relieve computa-

tional complexity, HF uses heuristics such as pruning [61]. It still, however, has to search exponentially many relevant pathways.[5]

Heuristic-based diagnostic systems apply heuristics to find diagnostic solutions that approximate the desired optimal ones.

Definition 7 (S-generated diagnostic solution): For some heuristic-based diagnostic system S and some diagnostic problem ψ, a diagnostic solution to ψ which is computed by S is called an S-generated diagnostic solution to ψ.

Diagnostic solutions generated by heuristic-based systems, thus, are generally "satisficing" solutions [93], in other words, causal explanations that are judged to be the most likely ones based on the heuristics employed.

3.3 Nature of Diagnostic Experience

This section attempts to describe, at a broad level, the implication of previous diagnostic solutions for an understanding of the diagnosis domain and for future diagnosis. Observe that diagnostic solutions identify diagnostically useful knowledge: In a diagnostic solution, the most likely causal relationships between clinical entities are selected, while insignificant ones are pruned away. The resulting pruned version of a detailed causal network thus can provide diagnostically valuable abstraction of disease models. Such abstraction raises the level of an understanding of vital causal dependencies among disorders and findings. In addition, diagnostic episodes provide a valuable source for acquiring "expertise" which allows subsequent problem solving to be done more efficiently. As mentioned in Chapter 1, this thesis develops methods for automatically acquiring such expertise from a problem-solver's own experience.

[5] HF has to consider at least an average of 10^{30} possible causal explanations for a diagnostic problem [2].

This thesis remembers solved cases (specifically, instantiated causal graphs representing the most likely causal explanations for findings) in a decomposed form by analyzing them with respect to elemental disorders. Each diagnostic solution is decomposed into smaller instantiated causal graphs each of which is obtained by collecting, for each elemental disorder in the solution, all nodes and causal links in the solution that are reachable from the elemental disorder. For example, consider again the diagnostic solution shown in Figure 3.8. Since the diagnostic solution shown in Figure 3.8 consists of two elemental disorders, it is decomposed into the two components shown in Figure 3.9. Each component is a subgraph of the diagnostic solution rooted at a particular elemental disorder in the diagnostic solution.

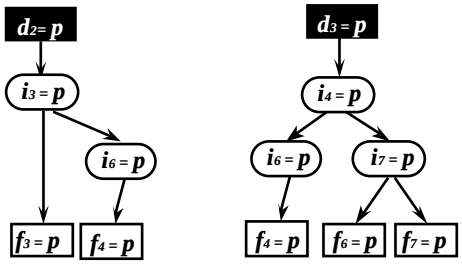


Figure 3.9: Components of the diagnostic solution shown in Figure 3.7 as a result of the decomposition

By remembering solved cases in a decomposed form, parts of cases can be accessed and used more easily. Such flexibility increases the possibility of finding matches particularly when parts of cases match.

The components of a solved case are then merged with components of other solved cases to produce a coherent whole. Components of solved cases can be merged if they are instantiations of the same underlying pathophysiologic mechanism of a particular elemental disorder. It is the combined whole, not individual similar components, that is remembered for use in future diagnosis. By remembering similar cases in a merged form, resources such as memory space and processing time can be used more efficiently.

Chapter 4 describes the diagnostic-unit representation which provides a basis for storing components of solved cases. Chapter 5 addresses the issues of decomposing solved cases and merging components. Chapter 6 describes new techniques for using knowledge, gained from previous diagnoses and represented in the diagnostic-unit representation, to perform decompositional abductive diagnosis.

Chapter 4

Diagnostic-Unit Representation

The diagnostic-unit representation is a new mechanism for representing knowledge about the context sensitivity of findings. It forms a basis for storing components of solved cases for use in future diagnosis. Primary constructs of the diagnostic-unit representation include diagnostic units and links that represent relationships between diagnostic units.

4.1 Diagnostic Units

Diagnostic units are designed to capture the context sensitivity of findings associated with particular disorders. They are clusters of evidence and hypotheses where all of the findings in each cluster are more, or equally, likely to be explained by the hypothesis identified in the cluster than by any other hypothesis for the same findings. To facilitate the formal discussion of diagnostic units, the following definitions are made.

Definition 8 (Source set of a causal graph): For any causal graph $G \in \mathcal{CG}$, the source set of G, denoted by $s\langle G \rangle$, is the set of nodes in G which do not have incoming links in G.

Definition 9 (Optimal diagnostic unit): For some instantiated subset F of $\mathcal{F}$ and some $d \in \mathcal{D}$, an optimal diagnostic unit of d, denoted by $du^\star(d, F)$, is an instantiated subgraph of $\mathcal{C}$ such that

1. $s\langle du^\star(d, F)\rangle = \{d\}$,
2. $f\langle du^\star(d, F)\rangle = F$, and
3. for any instantiated subgraph G of $\mathcal{C}$ such that $f\langle G\rangle = F$, $Pr(di\langle du^\star(d, F)\rangle \,|\, F) \geq Pr(di\langle G\rangle \,|\, F)$.

Less formally, an optimal diagnostic unit is an instantiated causal graph, with a single elemental disorder root, such that the causal explanation identified by the graph can be immediately inferred to be the most likely causal explanation for all the findings in the graph.

Optimal diagnostic units are designed to identify *diagnostic contexts* in which *global optimality* is implied. A diagnostic context specifies that when all of the findings identified in a diagnostic unit occur together, the disorder and underlying causal mechanism identified in the diagnostic unit can be concluded as a diagnosis for the findings. Global optimality means that the disorder and underlying mechanism identified in a diagnostic unit are the most likely diagnosis for the findings.

In principle, for any combination of an elemental disorder and a subset of $\mathcal{F}$, the corresponding optimal diagnostic unit can be computed. The issue is one of computational resources. The number of subsets of $\mathcal{F}$ to consider grows exponentially in the number of variables in $\mathcal{F}$. This thesis deals with such combinatorial explosion by acquiring diagnostic units that appear empirically useful. In effect, it views diagnostic problem solving as a process by which diagnostic units and relationships between them are gradually recovered.

Definition 10 (Empirical approximation of an optimal diagnostic unit): For some instantiated subset F of $\mathcal{F}$ and some $d \in \mathcal{D}$, an empirical approximation to the optimal diagnostic unit $du^\star(d, F)$, denoted by $du(d, F)$, is an instantiated subgraph of $\mathcal{C}$ such that

1. $s\langle du(d,F)\rangle = \{d\}$,
2. $f\langle du(d,F)\rangle = F$, and
3. for any previously encountered, instantiated subgraph G of $\mathcal{C}$ such that $f\langle G\rangle = F$, $Pr(di\langle du(d,F)\rangle \,|\, F) \geq Pr(di\langle G\rangle \,|\, F)$.

An empirical approximation of an optimal diagnostic unit is an instantiated causal graph, with a single elemental disorder root, such that previous experiences indicate that the causal explanation identified by the graph can be immediately inferred to be the most likely causal explanation for the findings in the graph. To provide an intuitive understanding of a diagnostic unit, an example of a diagnostic unit is shown in Figure 4.1.

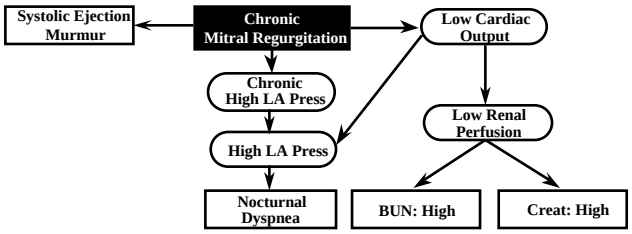


Figure 4.1: An example of a diagnostic unit

The diagnostic unit shown in Figure 4.1 is an instantiated causal graph with a single root representing the elemental disorder chronic mitral regurgitation. The following qualitative relationship is embedded in the diagnostic unit:

> Experiences so far indicate that the findings in the graph are more, or equally, likely to be caused by chronic mitral regurgitation via the underlying mechanism identified in the graph than by any other causal explanation for the same findings.

Empirical approximations of optimal diagnostic units are designed to identify diagnostic contexts in which *approximate optimality* is implied. A

diagnostic unit is approximately optimal if, according to past experience, the disorder and underlying mechanism identified in the diagnostic unit are believed to be the most likely diagnosis for the findings. Empirical approximations of optimal diagnostic units are generally referred to simply as diagnostic units.

Different sets of findings can suggest the same elemental disorder via different underlying pathophysiologic mechanisms. For example, myocardial infarction frequently produces ventricular systolic dysfunction with relatively high likelihood. Myocardial Infarction can also be a highly likely cause of ventricular diastolic dysfunction – though with less frequency. Each of the dysfunctions has a different underlying pathophysiologic mechanism, and thus corresponds to a different diagnostic unit of myocardial infarction. In light of the possibility that an elemental disorder can be associated with more than one diagnostic unit, a set of diagnostic units for an elemental disorder is defined as follows:

Definition 11 (Optimal diagnostic-unit set): For some d in $\mathcal{D}$, the optimal diagnostic-unit set of d, denoted by $DU^{\star}(d)$, is the set of optimal diagnostic units of d.

This thesis acquires approximations of optimal diagnostic-unit sets from experience.

Definition 12 (Empirical approximation of optimal diagnostic-unit set): For some d in $\mathcal{D}$, an empirical approximation to the optimal diagnostic-unit set $DU^{\star}(d)$, denoted by $DU(d)$, is a set of diagnostic units of d that are acquired from experience.

Empirical approximations of optimal diagnostic-unit sets are generally referred to simply as diagnostic-unit sets.

4.2 Links between Diagnostic Units

The definition of a diagnostic unit does not mention how a diagnostic unit is related to other diagnostic units. This section defines two types of links, to represent relationships between diagnostic units. One type of link is called a causal relation link. A causal relation link represents a causal dependency between two diagnostic units.

Definition 13 (Causal relation link): For any two diagnostic units du_1 and du_2 rooted at d_1 ($\in \mathcal{D}$) and d_2 ($\in \mathcal{D}$), respectively, du_1 is said to be causally related to du_2 if there exists a diagnostic solution S_ψ such that

1. du_1 and du_2 are subgraphs of S_ψ, and
2. there exists a causal path W in S_ψ from d_1 to d_2 such that d_1 and d_2 are the only elemental disorders in W.

The other type of link, called a non-causal relation link, represents a dependency between diagnostic units which are not causally related but still share common nodes.

Definition 14 (Non-causal relation link): For any two diagnostic units du_1 and du_2 rooted at d_1 ($\in \mathcal{D}$) and d_2 ($\in \mathcal{D}$), respectively, du_1 is said to be non-causally related to du_2 if there exists a diagnostic solution S_ψ such that

1. both du_1 and du_2 are subgraphs of S_ψ,
2. there exists no causal path in S_ψ either from d_1 to d_2 or from d_2 to d_1, and
3. there exists a node n such that there exist in S_ψ a causal path W_1 from d_1 to n and a causal path W_2 from d_2 to n such that d_1 and d_2 are the only elemental disorders in W_1 and W_2, respectively.

Knowledge represented in the diagnostic-unit representation can be conceptualized as a graph where each node represents a diagnostic-unit set, and each link represents a relationship between diagnostic units in different diagnostic-unit sets. Such a graph is called a *diagnostically-operative causal graph,* or simply a DOC graph.

Definition 15 (Diagnostically-operative causal graph): A diagnostically-operative causal graph G is a pair (CU, L), where CU is a set of diagnostic-unit sets and L is a set of links between diagnostic units in the diagnostic-unit sets in CU.

A pictorial notation is used to represent the primary constructs of the diagnostic-unit representation. A filled oval node denotes a diagnostic-unit set, while an oval node surrounded by a filled oval node denotes an element of the "surrounding" diagnostic-unit set. A causal relation between diagnostic units is denoted by an arrow with a circle on it, while a non-causal relation between diagnostic units is denoted by a solid line with a square on it. To make the pictorial notation more concrete, an example of a DOC graph is shown in Figure 4.2.

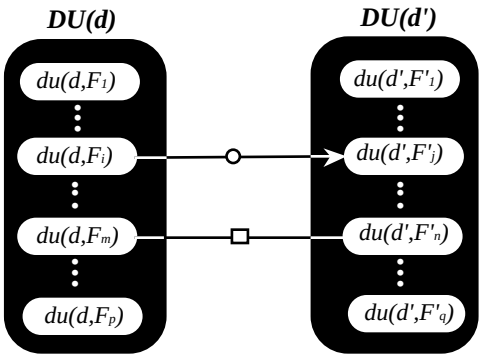


Figure 4.2: An example of a diagnostically-operative causal graph

The example DOC graph consists of two diagnostic-unit sets: $DU(d)$ and $DU(d')$ for some elemental disorders d and d', respectively. Diagnostic-unit set

$DU(d)$ has p diagnostic units, $du(d, F_1), du(d, F_2), \ldots,$ and $du(d, F_p)$, where all F_i's are instantiated subsets of $\mathcal{F}$. Diagnostic-unit set $DU(d')$ consists of q diagnostic units, $du(d', F'_1), du(d', F'_2), \ldots,$ and $du(d', F'_q)$, where each F'_j is an instantiated subset of $\mathcal{F}$. The example DOC graph shows that diagnostic unit $du(d, F_i)$ is causally related to $du(d', F'_j)$, and that $du(d, F_m)$ and $du(d', F'_n)$ are connected by a non-causal relation link.

4.3 Implications of Diagnostic Units for Diagnosis

Each optimal diagnostic unit groups as a unit a disorder, a underlying pathophysiologic mechanism, and a set of findings that are in the most likely causal relation. Diagnostic contexts and global optimality captured in diagnostic units facilitate the identification of how to group a diagnostic problem into subsets of findings for which the most likely causal explanations can be immediately inferred. Diagnostic units thus provide "good" abstraction of causal knowledge for diagnostic tasks. Chapter 5 describes an experience-based knowledge incorporation technique for acquiring diagnostic units that approximate optimal ones.

Chapter 5

Experience-based Acquisition of Diagnostic Knowledge

Once diagnostic units are available, a solution to a diagnostic problem can be proposed by selecting and combining the diagnostic units relevant to the problem. In addition, diagnostic units can enhance an understanding of the vital relationships between disorders and sets of findings. The issue that must be dealt with is where and how to acquire diagnostic units and relationships between them. One approach is manual compilation by interviewing expert physicians. Manual compilation is, however, often expensive and time-consuming. In an attempt to deal with difficulties associated with manual compilation, this research seeks to automatically acquire diagnostic units and relationships between them, by analyzing results of diagnoses. As discussed in Section 3.2, in the results of diagnosis highly likely causal relations are highlighted, while insignificant ones are pruned away. This feature suggests the possibility of using the results as a source of gradually recovering optimal diagnostic units and relationships between them.

5.1 Knowledge Incorporation Process

This section investigates an experience-based method for acquiring diagnostic units and the relationships between them. It is assumed in this research that the associative knowledge base K_{AS}, in which knowledge acquired from experience is stored, is initially empty. In other words, no diagnostic units are known *a priori*. As diagnostic experience grows, however, empirically useful diagnostic units and relationships between them are gradually recovered. New diagnostic experience is assimilated into the existing K_{AS} by the process called *knowledge incorporation process*. The knowledge incorporation process is a two-stage process that consists of the "diagnostically-operative causal graph transformation process" and the succeeding "joining-up process."

5.1.1 Diagnostically-Operative Causal Graph Transformation Process

The *diagnostically-operative causal graph transformation process*, or in short the DOC transformation process, constructs DOC graphs for new diagnostic solutions.

The following additional definitions are introduced to facilitate further discussion.

Definition 16 (Restricted intermediate state set $i\langle d \,|\, G\rangle$): For any causal graph $G \in \mathcal{CG}$ and any $d \in d\langle G\rangle$, the intermediate state set of d restricted to G, denoted by $i\langle d \,|\, G\rangle$, is the set of intermediate states in G that are reachable from d. Thus, $i\langle d \,|\, G\rangle \subseteq i\langle G\rangle$.

Definition 17 (Restricted causal link set $l\langle d \,|\, G\rangle$): For any causal graph $G \in \mathcal{CG}$ and any $d \in d\langle G\rangle$, the causal link set of d restricted to G, denoted by $l\langle d \,|\, G\rangle$, is the set of causal links in G that are reachable from d. Thus, $l\langle d \,|\, G\rangle \subseteq l\langle G\rangle$.

Definition 18 (Restricted finding set $f\langle d \,|\, G\rangle$): For any causal graph $G \in \mathcal{CG}$ and any $d \in d\langle G\rangle$, the finding set of d restricted to G, denoted by $f\langle d \mid G\rangle$, is the set of findings in G that are reachable from d. Thus, $f\langle d \,|\, G\rangle \subseteq f\langle G\rangle$.

Given a new diagnostic solution, the DOC transformation process first compiles diagnostic units by collecting, for each elemental disorder in the diagnostic solution, all nodes and causal links in the diagnostic solution that are reachable from the elemental disorder. This procedure for the DOC transformation process is summarized in the algorithm shown in Figure 5.1.

DOC Transformation Algorithm:

Input: A diagnostic solution S_ψ
Output: The corresponding DOC graph for S_ψ

Step I: For each elemental disorder d in $d\langle S_\psi\rangle$, compile a diagnostic unit du by collecting all nodes and causal links in S_ψ that are reachable from d: In other words, du is an approximate diagnostic unit such that $s\langle du\rangle = d, i\langle du\rangle = i\langle d \,|\, du\rangle, l\langle du\rangle = l\langle d \,|\, du\rangle$, and $f\langle du\rangle = f\langle d \,|\, du\rangle$.

Step II: For any two approximate diagnostic units compiled in Step I, establish either a causal or a non-causal link between them accordingly (See Definitions 13 and 14).

Figure 5.1: An algorithm for the DOC transformation process

The transformation of a new diagnostic experience to a DOC graph can be made concrete with an example. Consider a diagnostic problem described below:

History:	61 year old male with normal weight and orthopnea.
	Having known diagnosis of old MI, and aortic valve replacement.
Vital signs:	Blood pressure = 90/50 $mmHg$
	Heart rate = 72 bpm
	Respiration rate = 16 bpm
Physical exams:	Chest revealed basilar rales
	JVP (Jugular Veneous Pressure) = $15cmH_20$
	Normal jugular pulse was observed.
	Auscultation revealed a systolic ejection murmur in I-II/VI at apex and mild pedal edema.
Lab findings:	Chest X ray: generalized cardiac enlargement and vascular redistribution
	Na = 140 meg/l
	K = 4.0 meg/l
	BUN (Blood Urea Nitrogen) = 15 $mg/100ml$
	Creatinine = 1.0 $mg/100ml$
	Normal acid base status

A diagnostic solution to the example diagnostic problem is shown in Figure 5.2.

The three diagnostic units rooted at aortic stenosis, aortic value replacement, and old MI, respectively, can be compiled from the diagnostic solution shown in Figure 5.2. Each diagnostic unit is obtained by collecting all nodes and causal links in the diagnostic solution that are reachable from the corresponding elemental disorder. For example, the diagnostic unit du(`Aortic Stenosis`) shown in Figure 5.3 is the diagnostic unit complied from the diagnostic solution, by collecting all nodes and causal links that are reachable from aortic stenosis.

The diagnostic solution shown in Figure 5.2 shows that aortic valve replacement is causing aortic stenosis. A causal relation link is thus established from the diagnostic unit rooted at aortic value replacement (du(`Aortic valve replacement`)) to the diagnostic unit rooted at aortic stenosis (du(`Aortic stenosis`)). In addition, the diagnostic solution shows that old MI and aortic stenosis are non-causally related to each other. Thus, the diagnostic unit rooted at old MI (du(`Old MI`)) and the diagnostic unit rooted at aortic

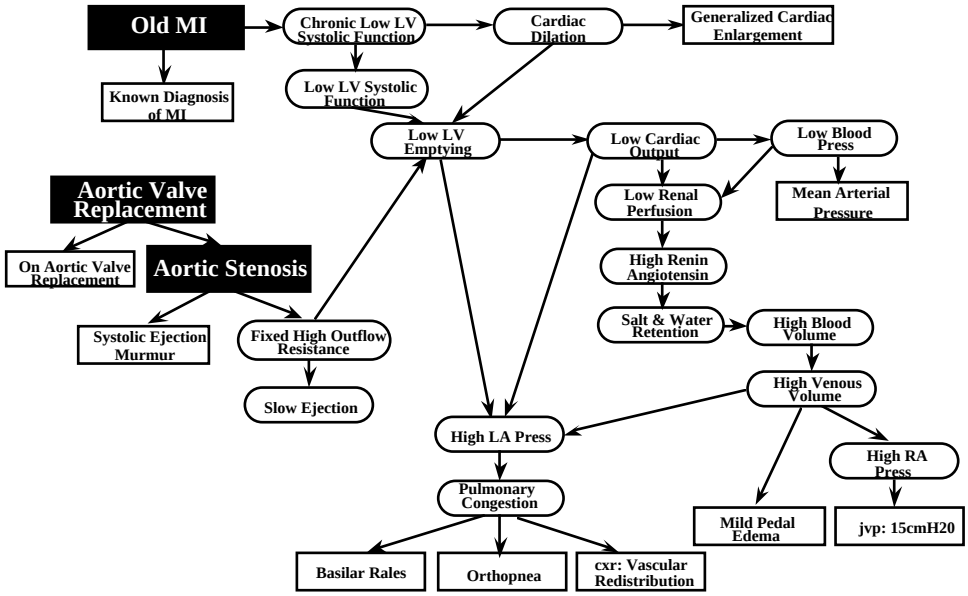


Figure 5.2: A diagnostic solution to the example diagnostic problem

stenosis (du(`Aortic stenosis`)) are connected by a non-causal relation link. Figure 5.4 shows the corresponding DOC graph for the diagnostic solution in Figure 5.2.

5.1.2 Joining-Up Process

The *joining-up process* incorporates newly acquired DOC graphs into the existing associative knowledge base K_{AS}. In particular, the joining-up process incorporates the following types of experiential knowledge into K_{AS} by the joining-up process.

- Newly compiled diagnostic units and relationships between them.
- Statistical information such as the frequency with which a diagnostic unit occurs.

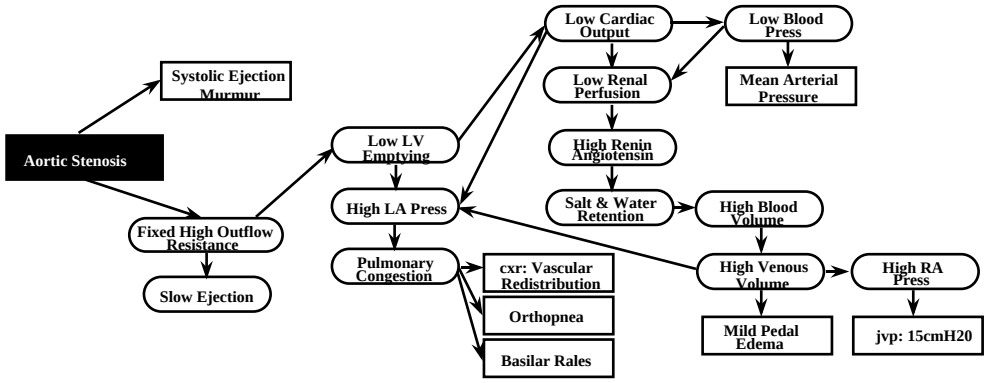


Figure 5.3: Diagnostic unit du(`Aortic stenosis`), rooted at aortic stenosis, compiled from the diagnostic solution shown in Figure 5.2

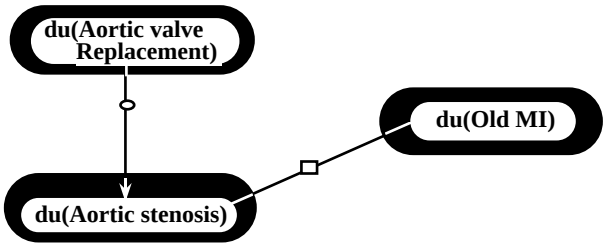


Figure 5.4: The corresponding DOC graph for the diagnostic solution shown in Figure 5.2

To facilitate further discussion, the following additional definition is introduced.

Definition 19 (Union of causal graphs): For any set of causal graphs G_1, G_2, ..., and G_n, where $G_i = (D_i, P_i, F_i, L_i)$, the union of the G_i's is defined as follows: $\bigcup_{i=1}^{n} G_i = (\bigcup_{i=1}^{n} D_i, \bigcup_{i=1}^{n} P_i, \bigcup_{i=1}^{n} F_i, \bigcup_{i=1}^{n} L_i)$.

Let DOC_ψ denote the DOC graph, for a diagnostic solution S_ψ, produced by the preceding DOC transformation process. Each diagnostic unit in DOC_ψ

is incorporated into K_{AS}. An algorithm for incorporating a diagnostic unit in DOC_ψ into K_{AS} is described in Figure 5.5.

Diagnostic Unit (DU) Incorporation Algorithm:

Input: 1. A diagnostic unit du, rooted at an elemental disorder d, in DOC_ψ
2. K_{AS}

Output: Updated K_{AS} as a result of incorporating du

Let $\widehat{DU}(d)$ be a diagnostic-unit set of d in K_{AS}.

- If $\widehat{DU}(d)$ exists in K_{AS}, then
  - if there exists a diagnostic unit $\widehat{du}$ in $\widehat{DU}(d)$ which du can be incorporated into, then modify the structure of $\widehat{du}$ by unioning $\widehat{du}$ and du. The union of causal graphs is defined in Definition 19. Increase the frequency with which $\widehat{du}$ has occurred by 1.
  - Otherwise, update $\widehat{DU}(d)$ by adding du to the set, setting the frequency associated with du to 1.
- If no $\widehat{DU}(d)$ exists in K_{AS}, then create $\widehat{DU}(d)$ with du as its element, setting the frequency associated with du to 1.

Figure 5.5: An algorithm for incorporating a newly compiled diagnostic unit into K_{AS}

The incorporation of new diagnostic units raises the issue of determining whether a new diagnostic unit du should be used to update an existing diagnostic unit $\widehat{du}$ in K_{AS} or be considered to be a new element of the corresponding diagnostic unit set. In other words, the question that needs to be answered is whether or not du and $\widehat{du}$ should be considered to be both a partial recovery of the same optimal diagnostic unit. The merging issue

is addressed with a *merging threshold*. If two diagnostic units are a partial recovery of the same optimal diagnostic unit, then they are likely to have a similar underlying structure. Two diagnostic units are merged to produce a combined diagnostic unit if they have the roots representing the same elemental disorder and the underlying structure of the two diagnostic units match by more than a certain percentage, *i.e.,* the merging threshold.

The merging threshold can affect the sizes of diagnostic-unit sets and of individual diagnostic units. The size of a diagnostic-unit set is defined as the number of diagnostic units in the diagnostic-unit set. The size of a diagnostic unit is defined as the number of nodes in the diagnostic unit. The sizes of diagnostic unit sets can affect reasoning efficiency, accuracy of solutions, and an understanding of domain structure. An experiment conducted to examine the effect of merging threshold on the sizes of diagnostic-unit sets is presented in Chapter 7.

Dependencies between diagnostic units are also updated in K_{AS}. Dependencies between diagnostic units in a DOC graph are by no means certain. Instead, a dependency between diagnostic units represents that the diagnostic units are related to each other with relatively high likelihood. An experiment, which is to be reported in Chapter 7, shows that a diagnostic unit can appear in a solution without the diagnostic units that are connected to it in the associative knowledge base. Nonetheless, links between diagnostic units provide useful information for guiding the construction of an overall solution to a diagnostic problem. Chapter 6 develops an algorithm which uses dependency among diagnostic units to construct an overall diagnostic solution.

In addition, the joining-up process also performs bookkeeping for use in later diagnosis, updating statistics such as how frequently a diagnostic unit has occurred so far. Such statistical information provides a useful guide for recognizing diagnostic units that are relevant to a diagnostic problem.

5.2 Knowledge in the Associative Knowledge Base

As diagnostic solutions are incorporated, the chunks of knowledge about the context sensitivity of findings are stored in the associative memory K_{AS}, ready for use in future diagnosis. Note that the associative knowledge K_{AS} is a DOC graph. Unlike a causal knowledge base (as shown in Figure 3.4) which exhibits little structure, K_{AS} shows more apparent structure in which patterns of findings that may exist in a diagnostic problem can be detected with relative ease. Such patterns of nodes and causal links are grouped into diagnostic units. Each diagnostic unit identifies a diagnostic context. While diagnostic units are not completely decomposable, they are "nearly decomposable" in the sense that links within each diagnostic unit are relatively strong, and hence each diagnostic unit is generally used as an atom. On the other hand, links between diagnostic units are relatively weak, as empirically verified by the experiment described in Chapter 7: Diagnostic units appear in solutions without the diagnostic units that are connected to them in the associative knowledge base. The abstraction of causal knowledge in the form of a DOC graph effectively provides insights about how given findings can be grouped into immediately solvable subproblems.

The knowledge incorporation process generalizes diagnostic experience (low-level rules for problems that have actually been encountered) to more general problem-solving rules for diagnosis (diagnostic units). The generalization is done gradually by accumulating diagnostic episodes that the program has encountered. Diagnostic units represent what the program knows from its own past experience. Unfortunately, in such medical domains as heart failure, different patients can manifest different findings, even when they suffer from the same disorder via the same underlying mechanism. As a result, a diagnostic unit compiled from a particular diagnostic solution is generally an imperfect disease description. A more general disease descrip-

tion is acquired by merging such imperfect disease descriptions, based on similarity in underlying structure. The degree to which a disease description is generalized depends on previous experiences. Even though imperfect, generalized descriptions are then used to solve subsequent problems presented to the program.

The knowledge incorporation process differs from explanation-based learning [22, 49, 68, 69] or example-based, inductive learning [4, 6, 9, 50, 80] in that it does not attempt to learn perfect units before they can be used in problem solving: Diagnostic units can be used in problem solving even though they are imperfect. Explanation-based learning learns rules from few training examples, using domain theories. Example-based learning learns a rule from many training examples which are believed to be covered by the rule. In general, both explanation-based and example-based learning techniques assume a unidirectional sequential relationship between the learning and problem-solving processes: a "learning first, and then problem solving" paradigm. Under this paradigm, these techniques attempt to learn rules that are generalized to the point where the rules can be directly applied to problems on an exact matching basis.

The issue here is the tradeoff between problem-solving efficiency and the cost of learning general rules. If rules are not general enough to cover most of the cases that they could ever cover, then they need to be modified, case by case, to solve problems. On the other hand, if rules are general enough to cover the full set of cases, then problem solving can be done by directly applying the rules to a problem, without any modification. While such straightforward application of rules on an exact matching basis reduces problem-solving time, learning such completely general rules can be expensive, if possible at all.

This thesis explores a strategy for solving diagnostic problems with whatever imperfect rules are available. The available rules are locally modified, case by case, so that they are applied to a given diagnostic problem.

Chapter 6

Decompositional Abductive Diagnosis

The previous chapters described the diagnostic-unit representation for organizing knowledge about the context sensitivity of findings, and the knowledge incorporation process for transforming experience into such general problem-solving rules. This chapter develops and analyzes a decompositional abductive diagnosis technique which exploits knowledge structures in the diagnostic-unit representation.

This thesis views decompositional abductive diagnosis as a two-stage process. The first stage is the grouping of a given body of evidence into subproblems. The basic approach is to find relevant diagnostic units based on an approximate technique called "deep matching adaptation." The second stage is to construct a diagnostic solution to the problem, from the relevant diagnostic units selected by the preceding evidence-grouping process.

6.1 Evidence-Grouping Process

The macro-finding captured in each diagnostic unit, as a whole, represents a clinical indicator that the specified set of findings strongly supports the

existence of the disorder and underlying pathophysiologic mechanism identified in the diagnostic unit. In other words, diagnostic units provide guides for decomposing a set of findings into smaller, immediately solvable subsets. In the diagnostic-unit representation paradigm, therefore, the grouping of given evidence becomes a search for relevant diagnostic units. This thesis addresses the issue of how to select relevant diagnostic units, by matching existing diagnostic units against the given evidence.

Depending on control strategies used to select relevant diagnostic units, the use of diagnostic units may degrade or improve overall problem-solving performance. Not all diagnostic unit selected are ones that are part of a correct diagnosis. Some diagnostic units are falsely chosen as relevant ones.

Definition 20 (True- and false- positive diagnostic units): A diagnostic unit that is part of a correct diagnosis is called a *true-positive diagnostic unit,* or shortly a true positive. A diagnostic unit that is not part of a correct diagnosis is called a *false-positive diagnostic unit,* or in short a false positive.

A common matching method is to simply count the number of findings that match on the surface. While easy to implement, this matching method, called "simple matching" in this thesis, is only effective when cases that are similar on the surface occur frequently. Unfortunately, in most medical domains, patients with the exact same findings rarely recur. Since diagnostic units used in this thesis are acquired from experience, knowledge captured in diagnostic units is generally incomplete. As a consequent, a diagnostic unit with syntactically dissimilar findings could be a better choice than that with more findings that match on the surface [2].

Type I matching error: If a true-positive diagnostic unit is overlooked due to dissimilarity in appearance, then a Type I matching error has occurred.

The straightforward application of simple matching to incomplete diagnostic units is prone to Type I matching errors. This type of errors may be reduced by adapting existing diagnostic units so that they match a new problem. This thesis explores this possibility by developing an approximate method, called "deep matching adaptation." A key feature of deep matching adaptation is that it considers not only similarities in appearance, but also similarities in underlying causality.

The issue that arises in matching a diagnostic unit against a diagnostic problem is what to do with those findings that are unmatched on the surface. There are two kinds of unmatched findings. One is unmatched findings in a diagnostic unit. The other is unmatched findings in a diagnostic problem. Since unmatched findings in a diagnostic unit are not known to a patient, they have no effect on the patient's state. In light of this observation, this thesis handles unmatched findings in a diagnostic unit by removing them from the diagnostic unit. Such removal can invalidate the diagnostic unit, however. The issue regarding the validity of a diagnostic unit is addressed in the succeeding hypothesis-construction process.

For unmatched findings in a diagnostic problem, this thesis checks to see if they can be explained by the diagnostic unit. To this end, a technique called "causal accounting" is investigated.

6.1.1 Causal Accounting

Causal accounting is a simple method for tailoring diagnostic units based on underlying causality of findings. Causal accounting allows an unmatched finding in a diagnostic problem to be added to a diagnostic unit (and treated as a matching finding), if there exists in the diagnostic unit a pathophysiologic state which can directly cause the unmatched finding. Causal accounting can be made more concrete with an example. Suppose that the diagnostic unit in Figure 6.1 is being matched against a diagnostic problem which consists of

the findings of systolic ejection murmur,[1] high BUN (Blood Urea Nitrogen), high creatinine,[2] and orthopnea.[3]

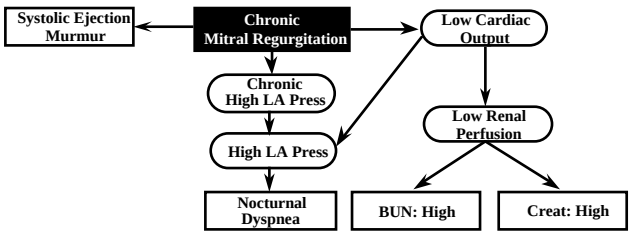


Figure 6.1: A diagnostic unit rooted at chronic mitral regurgitation

Orthopnea is an unmatched finding in the problem. According to HF's causal model, high LA (Left Atrial) pressure can directly cause orthopnea as well as nocturnal dyspnea. Thus, causal accounting allows orthopnea to be added to the diagnostic unit as an effect of high LA pressure.

Definition 21 (Causally matching finding and accounting state): Let ψ be a diagnostic problem, and du be a diagnostic unit being matched against ψ. For any unmatched finding f in ψ, f is said to be *explainable by* du if there exists a pathophysiologic state i such that

1. $i \in di\langle du \rangle$, and
2. $\exists$ a causal relation $i \longrightarrow f \in l\langle \mathcal{C} \rangle$, where $\mathcal{C}$ is the domain causal network.

f is called a *causally matching finding*, and i an *accounting state* of f.

[1]Systolic ejection murmur is an adventitious sound heard on auscultation of the heart that is most intense at the time of maximum flow of blood from the heart.

[2]Creatinine is one of the nonprotein constituents of blood, and increased quantities of it are found in advanced stage of renal disease.

[3]Orthopnea is a shortness of breath in any but erect sitting or standing position.

In our example, orthopnea is a causally matching finding, and high LA pressure is an accounting state of orthopnea. For expository simplicity, a finding that matches on the surface is called a syntactically matching finding.

Definition 22 (Syntactically matching finding): Let ψ be a diagnostic problem, and du be a diagnostic unit being matched against ψ. A *syntactically matching finding* is a finding that matches on the surface, *i.e.,* is an element of both ψ and $f\langle du\rangle$.

Systolic ejection murmur, high BUN, and high creatinine are syntactically matching findings in our example.

Causal accounting offers the potential to reduce the possibility of making Type I matching errors, and can efficiently increase the usability of diagnostic units in problem solving. In addition, it does not require any particular domain-specific heuristics to adapt diagnostic units. Only knowledge about direct causal dependencies, available in the causal knowledge base K_{CMS} and used primarily for first principles diagnosis by CMS, is needed for adaptation.

The depth of causality examined by causal accounting is limited to direct causal dependencies, however, in order to avoid a costly complete propagation of new evidence impact throughout the entire network. In this regard, causal accounting can be viewed as a one-step lookahead version of a more general accounting principle. From this viewpoint, it becomes apparent that causal accounting trades accuracy for computational efficiency, and hence can degrade the extent to which end results represent reasonable approximations to the desired optimal hypotheses.

Type II matching error: If a diagnostic unit selected by a matching strategy is false positive, a Type II matching error has occurred.

Causal accounting can increase the possibility of making Type II matching errors. Techniques for dealing with Type II matching errors are investigated in the succeeding hypothesis-construction process.

CASEY [55], a case-based reasoning system for heart failure diagnosis, explored an adaptation method similar to causal accounting. To solve a new diagnostic problem, CASEY directly inspects and adapts previous cases, treating each case as an independent atom. It then transfers the entire diagnostic solution of the best matching previous case to the new problem.

CASEY matches a previous case as follows: Let ψ and ψ' denote a new diagnostic problem and a previous diagnostic problem, respectively. Let $S_{\psi'}$ be the stored solution to the previous case ψ' (specifically, a causal explanation for ψ'). CASEY takes a set of the pathophysiological states in $S_{\psi'}$ which have direct causal links to at least one of the findings in ψ', and matches this set against a set of the pathophysiological states which can directly cause at least one finding in ψ. For expository convenience, let I_1 be the former set and I_2 the latter. Then,

$$I_1 = \{i | i \in di\langle S_{\psi'}\rangle \text{ and } \exists \text{ a causal relation } i \longrightarrow f \in l\langle S_{\psi'}\rangle \text{ for some } f \in f\langle S_{\psi'}\rangle\}$$
$$I_2 = \{i | i \in di\langle \mathcal{C}\rangle \text{ and } \exists \text{ a causal relation } i \longrightarrow f \in l\langle \mathcal{C}\rangle \text{ for some } f \in \psi\},$$

where $\mathcal{C}$ is the domain causal network. As shown in I_1, in order for a pathophysiologic state in $S_{\psi'}$ to be used in matching, it has to be a direct cause for at least one finding in $S_{\psi'}$. CASEY determines the similarity between ψ' and ψ, by counting the number of pathophysiological states that are both in I_1 and in I_2, *i.e.*, $|I_1 \cap I_2|$. A previous case with the largest number of matching pathophysiologic states is then chosen as the best match.

CASEY's matching method is essentially the same as determining matching degrees based on the number of underlying causes, as opposed to findings, in common. In principle, therefore, a previous case can be chosen as the best match in CASEYeven when a new case and the chosen one do not have any

common finding. Such generous adaptation could adversely affect the suitability of a solution, transferred from the best match, as the most likely causal explanation.

6.1.2 Deep Matching Adaptation

The evidence-grouping process finds relevant diagnostic units by a technique called "deep matching adaptation." In an attempt to take into account similarity in underlying causality as well as similarity in appearance, deep matching adaptation uses causal accounting in determining relevant diagnostic units.

Deep matching adaptation adapts diagnostic units to fit a given diagnostic problem, as follows: For a diagnostic unit, 1) make a copy of the diagnostic unit, 2) for each unmatched finding f in the diagnostic problem, if f can be explained by causal accounting, then add to the copy a node n that represents f and a direct causal link to n from the corresponding accounting state, 3) remove all unmatched findings in the diagnostic unit from the copy, and then 4) return the modified copy. This procedure for deep matching adaptation of a diagnostic unit is summarized in Algorithm DMAT shown in Figure 6.2.

Algorithm DMAT takes as input a diagnostic unit and a diagnostic problem. It adapts the diagnostic unit to fit the problem. To provide an intuitive understanding of how Algorithm DMAT adapts a diagnostic unit, let us apply Algorithm DMAT to the diagnostic unit shown in Figure 6.3(a) and a diagnostic problem ψ which consists of findings f_1, f_5, f_7, f_8, and f_9. For expository purposes, it is assumed that intermediate states i_1 and i_2 can directly cause findings f_7 and f_8, respectively – in other words, $i_1 \longrightarrow f_7 \in l\langle\mathcal{C}\rangle$ and $i_2 \longrightarrow f_8 \in l\langle\mathcal{C}\rangle$. Let du' be a copy of the diagnostic unit shown in Figure 6.3(a). Algorithm DMAT modifies the set M of matching findings and du' as follows:

Algorithm Deep Matching Adaptation (DMAT):

Input: 1. A diagnostic unit du in K_{AS}
2. A diagnostic problem ψ

Output: An adapted diagnostic unit du' of du

Step I: Copy du to du'. Let M be the set of matching findings. It is initialized to $f\langle du\rangle \cap \psi$.

Step II (Adaptation by adding causally matching findings): For any unmatched finding f in ψ, if f can be explained by causal accounting,

- add to du' a node n representing f and a direct causal link from i to n, where i is a node in du' representing an accounting state of f, and
- add f to M.

Step III (Adaptation by pruning away unmatched findings in a diagnostic unit): Remove from du' modified by Step II all nodes with no paths to any findings in M modified by Step II.

Step IV: Return the adapted diagnostic unit du' produced by Step III.

Figure 6.2: Algorithm DMAT for adapting a diagnostic unit based on deep matching adaptation

1. $M \longleftarrow \{f_1, f_5\}$, the intersection of $f\langle du\rangle$ and ψ.
2. $M \longleftarrow \{f_1, f_5, f_7\}$ since i_1 can directly cause f_7. Add to du' a node representing f_7 and a causal link from the node representing i_1 to the newly added finding node.

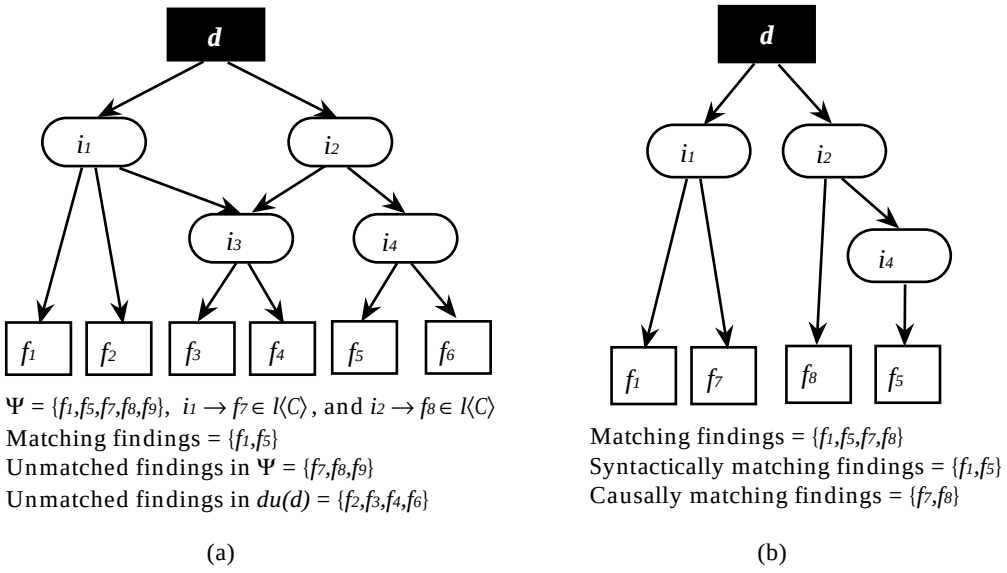


Figure 6.3: An example of deep matching adaptation; (a) A diagnostic unit to be matched against $\psi = \{f_1, f_5, f_7, f_8, f_9\}$; (b) An adapted diagnostic unit returned by Algorithm DMAT

3. $M \longleftarrow \{f_1, f_5, f_7, f_8\}$ since i_2 can directly cause f_8. Add to du' a node representing f_8 and a causal link from the node representing i_2 to the newly added finding node.

All unmatched findings in the diagnostic unit – f_2, f_3, f_4, and f_6 – are then removed from the modified du'. Figure 6.3(b) shows the adapted diagnostic unit returned by Algorithm DMAT. All findings in the adapted diagnostic unit shown in Figure 6.3(b) are matching findings. The matching findings include two syntactically matching findings f_1 and f_5 and two causally matching findings f_7 and f_8.

6.1.3 Algorithm for Evidence-Grouping Process

The evidence-grouping process applies deep matching adaptation to each diagnostic unit that exists in K_{AS}. Algorithm Evidence-Grouping shown in

Figure 6.4 summarizes the evidence-grouping process. Adapted diagnostic units returned by Algorithm Evidence-Grouping are passed to the succeeding hypothesis-construction process. The hypothesis-construction process then constructs an overall solution to the given problem from these adapted diagnostic units.

Algorithm Evidence-Grouping:

Input: 1. A diagnostic problem ψ
2. K_{AS}

Output: A set $\mathcal{Q}$ of adapted diagnostic units

1. $\mathcal{Q} \longleftarrow$ `Nil`;
2. For each diagnostic unit du in K_{AS}
   $\mathcal{Q} \longleftarrow \text{DMAT}(du, \psi)$;
3. Return $\mathcal{Q}$;

Figure 6.4: Algorithm Evidence-Grouping for selecting relevant diagnostic units

Indexing refers to choosing the features of an object that will be used as pointers to it in memory. As shown in Algorithm Evidence-Grouping, the evidence-grouping process does not use any pre-specified features as an index to access diagnostic units. Instead, it simply examines all diagnostic units in K_{AS}. This method appears reasonable in medical domains like heart failure. In most medical domains, all of the findings shown in a case are not equally important. Moreover, the relative importance of findings can change from patient to patient, depending on the relative strength of causes for other findings that occur together. Such patient-specificity of primary clinical features makes it difficult to specify the *a priori* significance of findings in a case. In addition, diagnostic units in K_{AS} are modified by experience, and their find-

ings are likely to change dynamically with experience. As a consequence, the accessing of diagnostic units based on pre-determined fixed pointers is not desirable in medical domains. CASEY is another diagnostic system which takes the patient-specificity of primary features into account for indexing. Instead of indexing by pre-specified features, CASEY uses indexing by every feature. It uses all the findings of a case for indexing. The price that has to be paid for not using particular pre-specified indices is computational efficiency. The amount of computation involved in the evidence-grouping process grows proportionally to the number of diagnostic units to examine.

6.1.4 Related Algorithms for Adaptation

This subsection describes two adaptation techniques based on existing matching methods: CASEY-style matching adaptation and simple matching adaptation. These methods will be compared to the deep matching adaptation method, by empirical study reported in Chapter 7.

CASEY-Style Matching Adaptation

CASEY-*style matching adaptation* is an adaptation method based on CASEY's matching method, which was described in Section 6.1.1. The procedure for CASEY-style matching adaptation of a diagnostic unit is summarized in Algorithm CMAT shown in Figure 6.5.

To provide an intuitive understanding of CASEY-style matching adaptation, let us apply Algorithm CMAT to the diagnostic unit shown in Figure 6.3(a) and the diagnostic problem which consists of findings f_1, f_5, f_7, f_8, and f_9. Given the input, Algorithm CMAT modifies the set M of matching findings and a copy du' of the diagnostic unit as follows:

1. $M \longleftarrow \{f_1, f_5\}$, the intersection of $f\langle du\rangle$ and ψ.
2. $M \longleftarrow \{f_1, f_5, f_7\}$ since i_1 not only can directly cause f_7 but also has f_1 and f_2, as its direct effects, in the diagnostic unit. Add to du' a

---

Algorithm CASEY-style Matching Adaptation (CMAT):

Input: 1. A diagnostic unit du in K_{AS}
2. A diagnostic problem ψ

Output: An adapted diagnostic unit du' of du

Step I: Copy du to du'. Let M be the set of matching findings. It is initialized to $f\langle du\rangle \cap \psi$.

Step II: For any an unmatched finding f in ψ, if there exists a pathophysiologic state i in du' such that i

1. has at least one finding in du' as its direct effect, and
2. can directly cause f, in other words $\exists$ a causal relation $i \longrightarrow f \in l\langle \mathcal{C}\rangle$,

then

- add to du' a node n representing f and a direct causal link from i to f, where i is an accounting state in du for f, and
- add f to M.

Step III (Adaptation by pruning away unmatched findings in a diagnostic unit): Remove from du' modified by Step II all nodes with no paths to any findings in M modified by Step II.

Step IV: Return the adapted diagnostic unit du' produced by Step III.

Figure 6.5: Algorithm CMAT for adapting a diagnostic unit based on the CASEY-style matching method.

node representing f_7 and a causal link from the node representing i_1 to the newly added finding node. Even though i_2 can directly cause f_8, i_2 does not have any finding as its direct effect in the diagnostic unit. Thus, unlike deep matching adaptation, CASEY-style matching adaptation does not add f_8 to du'.

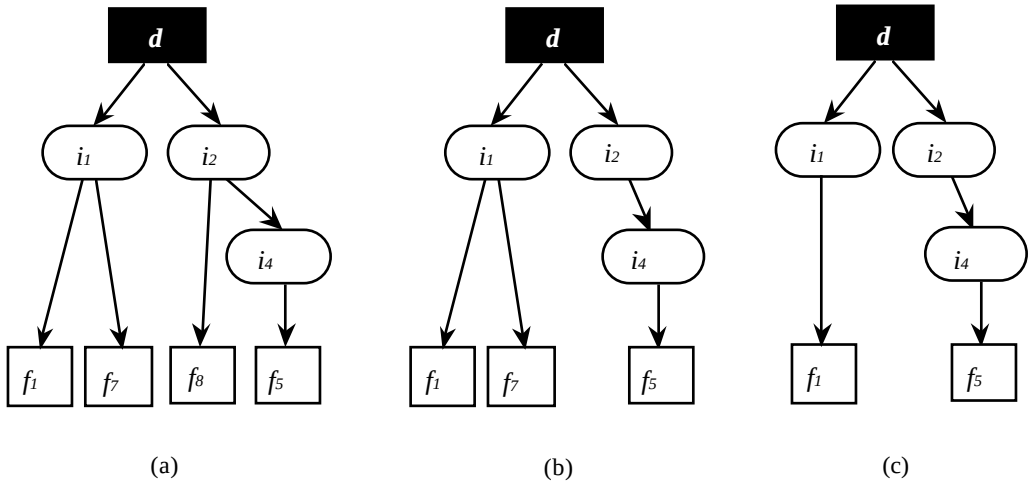


Figure 6.6: Results of various adaptation techniques applied to the diagnostic unit shown in Figure 6.3(a) and the diagnostic problem which consists of f_1, f_5, f_7, f_8 and f_9; (a) An adapted diagnostic unit returned as output by Algorithm DMAT; (b) An adapted diagnostic unit returned by Algorithm CMAT; (c) An adapted diagnostic unit returned by Algorithm SMAT

All unmatched findings in diagnostic unit – f_2, f_3, f_4, and f_6 – are then removed from the modified du'. Figure 6.6(b) shows the adapted diagnostic unit returned as output by Algorithm CMAT. The matching findings, *i.e.*, the findings in the adapted diagnostic unit shown in Figure 6.6(b), include two syntactically matching findings f_1 and f_5 and a finding f_7 which do not match on the surface but is accounted for by CASEY-style matching.

CASEY-style matching adaptation differs from deep matching adaptation in the way that an unmatched finding in a diagnostic problem is considered to be explainable by a diagnostic unit (as well as in that in CASEY, a whole

case corresponds to a diagnostic unit). CASEY-style matching adaptation considers an unmatched finding f in a diagnostic problem to be explainable by a diagnostic unit du, if there exists a pathophysiologic state i in the diagnostic unit that both can directly cause f and has at least one finding in the diagnostic unit as its direct effect – *i.e.*, $\exists$ a causal relation $i \longrightarrow f \in l\langle \mathcal{C} \rangle$, and $\exists f' \in f\langle du \rangle$ such that $\exists$ a causal relation $i \longrightarrow f' \in l\langle du \rangle$.

Simple Matching Adaptation

Simple matching adaptation is an adaptation technique for selecting relevant diagnostic units based on appearance. All the findings in a diagnostic unit, except the syntactically matchings, are removed. This procedure for simple matching adaptation is summarized in Algorithm SMAT shown in Figure 6.7.

Algorithm Simple Matching Adaptation (SMAT):

Input: 1. A diagnostic unit du in K_{AS}
2. A diagnostic problem ψ

Output: An adapted diagnostic unit du' of du

Step I: Copy du to du'. Let M be the set of matching findings. It is initialized to $f\langle du \rangle \cap \psi$.

Step II (Adaptation by pruning away unmatched findings in a diagnostic unit): Remove from du' all nodes with no paths to any findings in M.

Step III: Return the adapted diagnostic unit du' produced by Step II.

Figure 6.7: Algorithm SMAT for adapting a diagnostic unit based on syntactic similarity

Figure 6.6(c) shows the adapted diagnostic unit as a result of applying Algorithm SMAT to the diagnostic unit shown in Figure 6.3(a) and the diagnostic problem which consists of findings f_1, f_5, f_7, f_8, and f_9. All unmatched findings in diagnostic unit – f_2, f_3, f_4, and f_6 – are removed from a copy of the diagnostic unit. The matching findings, *i.e.,* the findings in the adapted diagnostic unit shown in Figure 6.6(c), include only two syntactically matching findings f_1 and f_5.

Simple matching adaptation differs from deep matching adaptation in that causal accounting is not performed. Simple matching only takes syntactic similarity into account to select relevant diagnostic units.

6.2 Hypothesis-Construction Process

The hypothesis-construction process generates a solution to the original problem by combining the adapted diagnostic units chosen by the evidence-grouping process. Note that each diagnostic unit returned by the evidence-grouping process is adapted to explain some subset of the given finding set. As a consequence, any combination of adapted diagnostic units can be an explanation for all of the given findings if the union of adapted diagnostic units in the combination is equal to the given finding set. The issue is how to find a combination which results in the most likely causal explanation.

The problem is that not all adapted diagnostic units selected by the evidence-grouping process are ones that are part of a correct diagnosis. Some of the adapted diagnostic units are, in fact, falsely chosen as relevant ones. Unfortunately, during testing, the evidence-grouping process using Algorithm Evidence-Grouping returned many diagnostic units, and most of them were false positives: More specifically, the evidence-grouping process returned an average of 66 adapted diagnostic units, and over 90% of them were false positives. The issue is how to pick "correctly" true positives from the output of the evidence-grouping process. Before addressing this issue, Section 6.2.1 analyzes deep matching adaptation. The analysis shows that there is no guar-

antee that a causal explanation identified in an adapted diagnostic unit is the most likely causal explanation for the findings in the adapted diagnostic unit.

6.2.1 Analysis of Deep Matching Adaptation

Although adaptation can efficiently increase the usability of diagnostic units, it could make it difficult to isolate true-positive diagnostic units. Two types of adaptation are performed in deep matching adaptation: causal accounting and removal of findings from a diagnostic unit. This subsection analyzes the impacts of causal accounting and removal of findings on the approximate optimality that is implied in diagnostic units.

Observe that deep matching adaptation does not propagate the impacts of the addition of new findings to, and the removal of findings from, a diagnostic unit throughout the entire network. Instead, it only examines direct causal dependencies. In this regard, it can be considered as a one-step lookahead technique. This narrowing of perspective gives valuable insights and analytical simplicity. By doing so, however, deep matching adaptation trades accuracy for computational efficiency. While it maintains local consistency in causation, addition or removal of findings in such a one-step lookahead way could affect global optimality unfavorably.

More specifically, consider a conditional probability statement $Pr(B \mid A) = p$. The statement means that if A is known to be true and A is the *only* thing known, a probability p can be attached to B. It does not convey any information about the probability of B under any other condition. What is identified in a diagnostic unit du is that given that all the finding in the diagnostic unit occur, then the causal explanation identified in the diagnostic can be concluded as a highly likely causal explanation for the identified findings, in relation to other causal explanations for the same findings. This qualitative relationship of du to other causal explanations holds when all findings in du occur. Thus, if all of the findings in the diagnostic unit are present in a

patient, the diagnostic unit can be used, with relatively high confidence, as (a part of) a diagnosis for the patient. On the other hand, if any of the findings in the diagnostic unit are absent or new findings are present, then the use of the diagnostic unit to explain the patient' findings could be accompanied by significant errors.

Vulnerable Causal Accounting

The depth of causality examined by causal accounting is restricted to direct causal dependencies. By avoiding propagating impacts of new evidence, causal accounting may not preserve the approximate optimality that is implied in diagnostic units.

For a more formal discussion of the vulnerability of causal accounting, suppose that for a set F of findings, $H, H_1, H_2, \ldots$, and H_p are hypotheses for F, and that H is the most likely hypothesis for F. Now suppose that a new finding f ($\notin F$), which can also be caused by $H, H_1, H_2, \ldots$, and H_p, is observed. Figure 6.8 graphically summarizes the causal relations defined in this example.

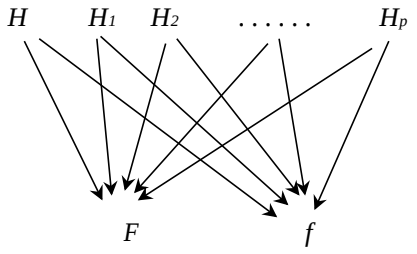


Figure 6.8: Examples of causal relations

The question is what is the most likely hypothesis for all of f and the findings in F. H could be a reasonable candidate which can be suggested without computation. Unfortunately, however, the fact that H is the most likely hypothesis for F does not guarantee that H is also the most likely one

for $\{f\} \cup F$. It can be proven as follows. Conditional independence between findings allows the impact of the new datum f to be computed incrementally. By Bayes' theorem [75],

$$Pr(H \mid F, f) = Pr(H \mid F) \frac{Pr(f \mid H)}{Pr(f \mid F)} \tag{6.1}$$

$$Pr(H_i \mid F, f) = Pr(H_i \mid F) \frac{Pr(f \mid H_i)}{Pr(f \mid F)}, \forall i = 1, 2, \ldots, p \tag{6.2}$$

Consider the righthand side of each of equations (6.1) and (6.2). Because H is the most likely hypothesis for F, for all $i = 1, \ldots, p$ $Pr(H \mid F) \geq Pr(H_i \mid F)$. This, however, does not necessarily imply that for all $i = 1, \ldots, p$

$$Pr(H \mid F, f) \geq Pr(H_i \mid F, f)$$

because there is no guarantee that it is always true that for all $i = 1, 2, \ldots, p$

$$Pr(f \mid H) \geq Pr(f \mid H_i).$$

For example, suppose that the diagnostic unit shown in Figure 6.9(a) is being matched against the problem which consists of systolic ejection murmur, orthopnea, and high BUN. The diagnostic unit shown in Figure 6.9(b) is the adapted diagnostic unit returned by Algorithm DMAT. What is captured in the diagnostic unit shown in Figure 6.9(a) is that nocturnal dyspnea is computed to be evidence in strong favor of high LA pressure. This does not necessarily imply, however, that orthopnea can also be strong evidence for high LA pressure. In order to determine that, the impact of new evidence must be fully propagated.

Vulnerability of Removal of Unmatched Findings

Similarly, the removal of unmatched findings in a diagnostic unit from the diagnostic unit can invalidate the diagnostic unit. For formal discussion,

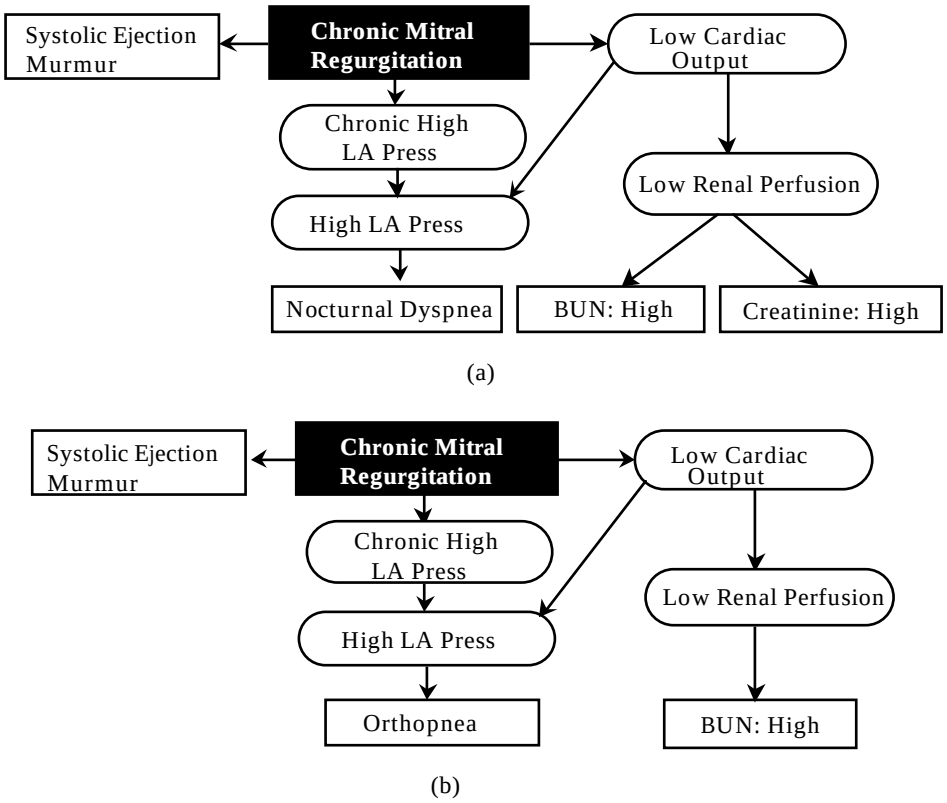


Figure 6.9: An example of deep matching adaptation, for a diagnostic problem which consists of systolic ejection murmur, orthopnea, and high BUN

reconsider the set F of findings for which hypotheses include $H, H_1, H_2, \ldots,$ and H_p, and for which the most likely hypothesis is H. Suppose that a finding f' in F is absent (see Figure 6.10 for causal relations defined in this example).

It can be proven that there is no guarantee that H is the most likely hypothesis for $F - \{f'\}$, in a similar way that new evidence impacts were proven. The impact of the absent datum f' can be computed incrementally. By Bayes' theorem [75],

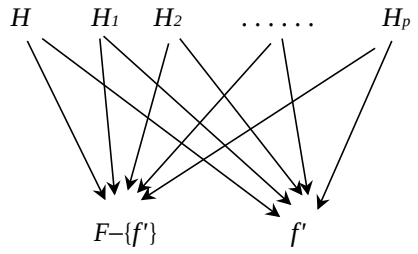


Figure 6.10: Examples of causal relations

$$\begin{aligned} Pr(H \,|\, F) &= Pr(H \,|\, F - \{f'\}, f') \\ &= Pr(H \,|\, F - \{f'\}) \frac{Pr(f' \,|\, H)}{Pr(f' \,|\, F - \{f'\})} \end{aligned} \tag{6.3}$$

$$\begin{aligned} Pr(H_i \,|\, F) &= Pr(H_i \,|\, F - \{f'\}, f') \\ &= Pr(H_i \,|\, F - \{f'\}) \frac{Pr(f' \,|\, H_i)}{Pr(f' \,|\, F - \{f'\})} \\ &\quad \text{for all} \quad i = 1, 2, \ldots, p \end{aligned} \tag{6.4}$$

Consider the lefthand side of each of equations (6.3) and (6.4). Because H is the most likely hypothesis for F, for all $i = 1, \ldots, p$

$$Pr(H \,|\, F) \geq Pr(H_i \,|\, F),$$

and thus for all $i = 1, \ldots, p$

$$Pr(H \,|\, F - \{f'\}) Pr(f' \,|\, H) \geq Pr(H_i \,|\, F - \{f'\}) Pr(f' \,|\, H_i).$$

This does not, however, necessarily imply that for all $i = 1, \ldots, p$

$$Pr(H \,|\, F - \{f'\}) \quad \geq \quad Pr(H_i \,|\, F - \{f'\}) \tag{6.5}$$

The inequality in equation (6.5) is true only when for all $i = 1, 2, \ldots, p$

$$Pr(f' \mid H_i) \geq Pr(f' \mid H).$$

For example, the diagnostic unit shown in Figure 6.9(a) specifies that high BUN and high creatinine together are evidence in strong support of low renal perfusion. It is not implied in the diagnostic unit, however, that high BUN alone can be strong evidence for low renal perfusion. The impact of absent evidence is uncertain until it is fully propagated.

Implication of One-Step Lookahead Adaptation

Unfortunately, the impact of new, or absent, evidence is computationally more intensive to propagate if causes can influence findings indirectly via several causal links [62, 75, 78]. Causal accounting and the removal of findings avoid the costly complete propagation of new/absent-evidence impacts. The sum of local consistency ensured in adaptation, however, does not guarantee that the end result of adaptation is globally optimal. Indeed, the application of adaptation to a diagnostic unit that does not share enough points of similarity with a given case could overlook context sensitivity of findings captured in the diagnostic unit, and consequently result in an adapted diagnostic unit which is essentially dysfunctional.

The issue is how to select, from the output of the evidence-grouping process, adapted diagnostic units that approximate the desired global optimum. One approach is to check each adapted diagnostic unit du, to see if the causal explanation identified in du is the most likely causal explanation for the findings in du. Such a check can be performed by computing $Pr(G \mid f\langle du \rangle)$, for each other causal explanation G for the same findings, and then comparing this probability with $Pr(di\langle du \rangle \mid f\langle du \rangle)$. Unfortunately, this approach is computationally expensive, for there are exponentially many causal explanations to consider (as discussed in Section 3.2). The rest of this chapter describes techniques developed to address the issue of how to efficiently pick adapted diagnostic units that approximately meet global optimality.

6.2.2 Specificity-Reflected Similarity Metric

A common picking method is to count the number of matching findings, and then pick the units with the largest number, as used in CASEY [55]. This method, called a *simple similarity metric* in this thesis, is easy to implement. Not all matching findings in an adapted diagnostic unit are, however, of the same kind. Some findings are included because they match on the surface (syntactically matching findings). Some findings are included because they are considered to be matching findings by causal accounting (causally matching findings). Causally matching findings are added to an adapted diagnostic unit, but their impact is not fully propagated. As a result, it is unknown how strongly these causally matching findings support the existence of the diagnostic unit, while for syntactically matching findings, it is indicated by previous experience that they strongly do. This thesis develops a similarity metric, called *specificity-reflected similarity metric*, to take this difference into account.

The specificity-reflected similarity metric is motivated by the observation that some findings do better than others in identifying the existence of a disorder. Findings in a diagnostic unit are divided into two groups: specific findings and non-specific findings.

Definition 23 (Specific and non-specific findings): For some diagnostic unit du in K_{AS}, *specific findings* of du are findings that play a significant role in identifying the existence of du. Findings in du that are not specific findings are called *non-specific findings* of du.

"Playing a significant role" clause in Definition 23 is implemented via comparison of the specificity attached to each finding in a diagnostic unit with a threshold. The specificity of a finding represents a level of the significance of a finding in identifying a particular disorder and the underlying pathophysiologic mechanism. HF is capable of providing specificities that range between 0 and 1, which the knowledge incorporation process of HYDI remembers for each finding in a diagnostic unit. The remembered specificity of a finding in

a diagnostic unit is compared with a threshold, to determine if the finding can be considered to be specific. During testing, a finding in a diagnostic unit with the specificity higher than 0.8 was considered as a specific finding of the diagnostic unit. In order to determine a threshold value, a test was conducted to learn the threshold which gives the best outcome in identifying diagnostic units. The test showed that the threshold of 0.8 produced the best outcome.

Unlike the simple similarity metric, the specificity-reflected similarity metric uses the following three measures for gauging the similarity between the problem and an adapted diagnostic unit:

- The number of matching specific findings
- The total number of matching findings
- The frequency with which a diagnostic unit has occurred so far, specifically the number of merging operations that have been applied to the diagnostic unit.

A matching specific finding is a finding that is both a specific finding of a diagnostic unit and a matching finding in the adapted version of the diagnostic unit.

Definition 24 (Matching specific finding): For some diagnostic unit du in K_{AS}, let du' be an adapted diagnostic unit of du returned by Algorithm DMAT. A *matching specific finding* is a finding that is both a specific finding of du and a finding in du'.

For some adapted diagnostic unit du, let $s(du)$, $t(du)$, and $f(du)$ denote the number of matching specific findings, the total number of matching findings, and the frequency with which du has occurred so far, respectively. To provide an intuitive understanding of $s(du)$ and $t(du)$, let us reconsider the

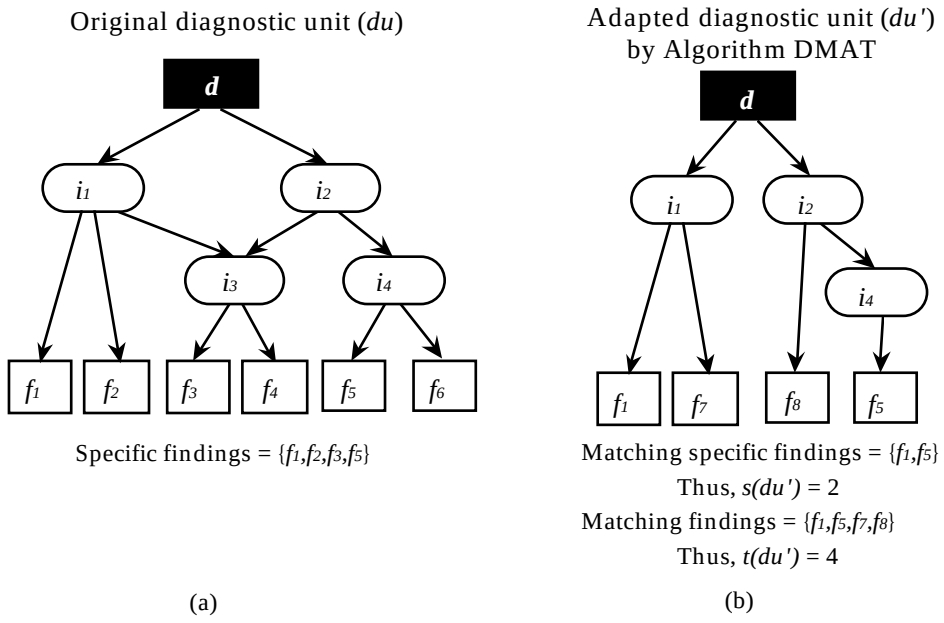


Figure 6.11: An example of measures used in the specificity-reflected similarity metric

deep matching adaptation example used in Section 6.1.2. For expository convenience, the original diagnostic unit du is shown again in Figure 6.11(a). Figure 6.11(b) reshows the adapted version du' of du. The original diagnostic unit shown in Figure 6.11(a) has four specific findings – f_1, f_2, f_3, and f_5. Of them, two specific findings f_1 and f_5 remain in the adapted diagnostic unit: Thus $s(du') = 2$. There are four matching findings in the adapted diagnostic unit. In other words, $t(du') = 4$.

The specificity-reflected similarity metric picks an adapted diagnostic unit with the highest rank, where ranks of adapted diagnostic units are determined as follows: For any two adapted diagnostic units, the one with the larger number of matching specific findings gets higher rank. In case of a tie, the one with the larger total number of matching findings gets higher rank. In case of a tie, the one that has occurred more commonly gets higher rank.

Underlying this ranking procedure is the following heuristics.

Heuristic 1: Two cases with the same underlying causes will share more findings than two cases with different underlying causes.

Heuristic 1, which is a straightforward observation on the property of the causal regularity of pathophysiologic functioning of a human body, provides guidance for selecting more likely causes for findings. It allows adapted diagnostic units which share more findings with a given problem to get higher ranks.

The algorithm shown in Figure 6.12(a) summarizes the specificity-reflected similarity metric. Note that matching specific findings are all syntactically matching findings. In this regard, the specificity-reflected similarity metric gives more weight to syntactically matching findings than to causally matching findings. For comparison purposes, an algorithm for the simple similarity metric is presented in Figure 6.12(b). The simple similarity metric, used in CASEY, determines similarity by counting the total number of findings in adapted diagnostic units. Syntactically and causally matching findings are thus given the same weight.

6.2.3 Dependency-Guided Picking Method

If the adapted diagnostic unit picked by the specificity-reflected similarity metric does not explain all of the given findings, other adapted diagnostic units need to be chosen to account for the unexplained findings. A common approach is to choose these units based on the disorder independence assumption. In most medical domains, however, disorders are not always independent of each other. In light of such dependency between disorders, this thesis uses dependencies between diagnostic units to guide the picking process.

Basically, additional units are chosen by applying the specificity-reflected similarity metric to adapted diagnostic units which have causal or non-causal

---

Similarity Metrics

Input: A set $Unsel$ of adapted diagnostic units
Output: An adapted diagnostic unit in $Unsel$ with the highest rank

(a) Algorithm for Specificity-Reflected Similarity Metric (SRSM):

Return an adapted diagnostic unit in $Unsel$ with the highest rank, where the rank is determined as follows:
For any two adapted diagnostic units, du and du', in $Unsel$ du gets the higher rank,

1. if $s(du) > s(du')$,
2. if $t(du) > t(du')$, when $s(du) = s(du')$, or
3. if $f(du) \geq f(du')$,
   when $t(du) = t(du')$ and $s(du) = s(du')$.

(b) Algorithm for Simple Similarity Metric (SSM):

Return an adapted diagnostic unit in $Unsel$ with the highest rank, where the rank is determined as follows:
For any two adapted diagnostic units, du and du', in $Unsel$ du gets the higher rank,

1. if $t(du) \geq t(du')$.

Figure 6.12: (a) Algorithm for Specificity-Reflected Similarity Metric; (b) Algorithm for Simple Similarity Metric

relation links to the adapted diagnostic units which have already been picked. If there are no such adapted diagnostic units, all remaining adapted diagnostic units are considered. This picking process is repeated either until all of the given findings are explained or until there is no unselected adapted diagnostic unit. The dependency-guided picking process is summarized in the algorithm shown in Figure 6.13.

Algorithm for Dependency-Guided Picking (DGP):

Input: 1. A set $Unsel$ of adapted diagnostic units that have not been selected
2. A set Sel of adapted diagnostic units that have been selected
3. A set E of findings that have been explained
4. A diagnostic problem ψ

Output: Modified Sel and E

1. While $E \neq \psi$ and $Unsel \neq$ `Nil`
2. $Candidates \longleftarrow$ The set of adapted diagnostic units in $Unsel$ that are causally or non-causally related to at least one adapted diagnostic unit in Sel;
3. $du \longleftarrow$ IF $Candidates \neq$ `Nil`
4. THEN SRSM($Candidates$)
5. ELSE SRSM($Unsel$);
6. $Sel \longleftarrow Sel \cup \{du\}$;
7. $Unsel \longleftarrow Unsel - \{du\}$;
8. $E \longleftarrow E \cup f\langle du \rangle$;
9. Return Sel and E;

Figure 6.13: Dependency-Guided Picking Algorithm for picking next units

Algorithm DGP returns a set of adapted diagnostic units that are selected to be parts of a diagnostic solution, and a set of findings that are explained by these selected units.

6.2.4 Algorithm for Hypothesis-Grouping Process

The hypothesis-construction process generates a diagnostic solution to the given problem from the output of the evidence-grouping process. The specificity-reflected similarity metric (Algorithm SRSM) and the dependency-guided picking method (Algorithm DGP) are used to select, from the output, diagnostic units that are judged to be parts of a diagnostic solution. If adapted diagnostic units chosen by Algorithm DGP can explain all the findings in the problem, then the hypothesis-construction process constructs a diagnostic solution by unioning the adapted diagnostic units: The union of causal graphs was defined in Definition 19. Otherwise, it returns a "Failed" signal to notify that the generation of a diagnostic solution to the problem is failed. This hypothesis-construction process is summarized in the algorithm shown in Figure 6.14.

6.3 Algorithm for Decompositional Abductive Diagnosis

Decompositional abductive diagnosis is performed by the evidence-grouping process followed by the hypothesis-construction process. The grouping of evidence is done by applying Algorithm Evidence-Grouping to a given diagnostic problem. The construction of a diagnostic solution to the problem is done by applying Algorithm Hypothesis-Construction to the output of Algorithm Evidence-Grouping. The procedure for decompositional abductive diagnosis is summarized in Algorithm DAD shown in Figure 6.15.

Algorithm DAD takes as input a diagnostic problem and the existing K_{AS}. It returns as output either a diagnostic solution to the problem or a "Failed"

Algorithm Hypothesis-Construction:

Input: 1. A diagnostic problem ψ
2. The set $\mathcal{Q}$ of adapted diagnostic units returned by the evidence-grouping process

Output: A diagnostic solution to ψ

1. $du \longleftarrow \text{SRSM}(\mathcal{Q})$;
2. $Sel \longleftarrow \{du\}$;
3. $E \longleftarrow f\langle du \rangle$;
4. Sel and $E \longleftarrow \text{DGP}(\mathcal{Q} - Sel, Sel, E, \psi)$;
5. IF $E = \psi$
6. THEN Return a diagnostic solution obtained by unioning all the adapted diagnostic units in Sel
7. ELSE Return "Failed" signal;

Figure 6.14: Algorithm Hypothesis-Construction

signal. Algorithm DAD first adapts each diagnostic unit in K_{AS} to fit the problem, by applying Algorithm Evidence-Grouping. Each adapted diagnostic unit returned by Algorithm Evidence-Grouping represents a disorder and underlying mechanism that can explain part of the overall malfunction. Algorithm DAD then applies Algorithm Hypothesis-Construction to the output of Algorithm Evidence-Grouping, to select adapted diagnostic units that are considered to be parts of a diagnostic solution. Such adapted diagnostic units are selected by taking into account differences in syntactically and causally matching findings and dependency among disorders.

Algorithm Decompositional Abductive Diagnosis (DAD):

Input: 1. A diagnostic problem ψ
2. K_{AS}

Output: A diagnostic solution to ψ

1. $\mathcal{Q}$ ⟵ Evidence-Grouping(ψ, K_{AS});
2. S_ψ ⟵ Hypothesis-Construction($\psi, \mathcal{Q}$);
3. Return S_ψ;

Figure 6.15: Algorithm Decompositional Abductive Diagnosis

6.4 Problem-Solving Algorithm for HYDI

HYDI performs hybrid reasoning to solve a diagnostic problem. HYDI's hybrid problem-solving flow is summarized in the algorithm shown in Figure 6.16.

Given a set of findings, the associative problem solver AS first tries to generate a diagnostic solution, by applying Algorithm DAD to the problem. AS uses decompositional abductive diagnosis which finds relevant diagnostic units based on deep matching adaptation, and constructs a diagnostic solution based on the specificity-reflected similarity metric and the dependency-guided picking method. If a diagnostic solution generated by AS is not acceptable, then the more robust CMS is called to solve the problem step-by-step from first principles. HYDI uses HF as its causal-model-based problem solver CMS. Acceptability of diagnostic solutions generated by AS is discussed in Chapter 7.

Algorithm HYDI-Diagnosis:

Input: A diagnostic problem ψ
Output: A diagnostic solution to ψ

1. $S_\psi \longleftarrow \text{DAD}(\psi, K_{AS})$;
2. $S_\psi \longleftarrow$ IF S_ψ is acceptable
3. THEN S_ψ
4. ELSE Call CMS;

Figure 6.16: Problem-Solving Algorithm for HYDI

Chapter 7

Empirical Analysis

An empirical study was conducted to test the efficiency and effectiveness of the techniques developed in this thesis. The data set used for the empirical analysis consisted of 300 cardiac patients from The New England Medical Center Hospital. Graphic results reported here were smoothed by the SYSTAT program [99].

In order to reduce bias due to case ordering, 50 independent trials were conducted on 50 different random case orderings, and their results were averaged together. In each trial, the associative knowledge base, K_{AS}, of HYDI was initially empty, and all 300 patients were run. Each diagnostic problem was solved with knowledge that was available in the knowledge bases at the time of problem solving. Each time a diagnostic problem was solved, a diagnostic solution was incorporated into K_{AS}, and hence used in the next diagnosis. For evaluation purposes, the most likely causal explanations generated by HF were assumed to be "correct" diagnoses.[1]

Section 7.1 presents an experiment conducted to examine the effect of merging threshold (used to determine if two diagnostic units can be merged) on the number of diagnostic units in K_{AS}. The remainder of the chapter

[1]Experience with physicians shows that the performance of HF is at a level acceptable by experts [63, 64].

presents experiments conducted to evaluate the problem-solving performance of the techniques developed in this thesis. While absolute performance is interesting, of more interest is the relative performance of previously studied alternative algorithms. Section 7.2 compares deep matching adaptation with simple and CASEY-style matching adaptation. Section 7.3 compares the specificity-reflected similarity metric and the dependency-guided picking method with the simple similarity metric and the simple picking method. Finally, Section 7.4 compares hybrid reasoning with each of causal-model-based and association-based reasoning.

7.1 Final Size of the Associative Knowledge Base of HYDI

The amount of computation involved in decompositional abductive diagnosis depends on the number of diagnostic units that are available in K_{AS}, which in turn is affected by the setting of the merging threshold. Because computational efficiency is an important aspect of automated diagnosis, an experiment was conducted to evaluate the effect of the merging threshold on the number of diagnostic units in K_{AS}. Experimental runs were made for five different merging threshold values – 75%, 80%, 85%, 90%, and 95%. For each merging threshold value, independent trials were conducted on the 50 random case orderings. In each trial, as diagnostic problems were solved, diagnostic units and relationships between them were discovered and incorporated into K_{AS}, which was initially empty. The procedure for each trial is summarized in the algorithm shown in Figure 7.1. A merging threshold value ρ, given as input, was used for knowledge incorporation. At the end of each trial, the size of K_{AS} was measured. The size of K_{AS} is defined as the total number of diagnostic units available in K_{AS}.

At the end of the experiment, of the total 83 elemental disorders defined in HF, diagnostic units for 77 disorders were identified in K_{AS}. Table 7.1

Procedure used in each trial to examine the effect of merging threshold on the size of K_{AS}:

Input:
1. A set Ψ of the 300 patients in random order
2. Empty K_{AS}
3. Merging threshold value ρ

1. For each patient ψ in Ψ
2. Solve the diagnostic problem;
3. Incorporate a verified diagnostic solution into K_{AS} by the knowledge incorporation process, using the merging threshold value ρ;
4. Measure the final size of K_{AS};

Figure 7.1: The procedure used by each trial to examine the sensitivity of the final size of K_{AS} to merging threshold

summarizes the average final size of K_{AS} for each of the various merging threshold values.

Merging threshold	Avg. size of a diagnostic-unit set	Avg. size of K_{AS}	Avg. # of links of a diagnostic unit
75%	1.26	97	5.4
80%	1.33	102	5.1
85%	1.56	120	5.0
90%	1.99	153	4.6
95%	3.35	258	4.5

Table 7.1: Average size of K_{AS} at the end of the experiment

The size of the diagnostic-unit set of a disorder is defined as the number of elements in the diagnostic-unit set, *i.e.,* the number of diagnostic units

identified for the disorder. Average final sizes of diagnostic-unit sets are shown in the second column of Table 7.1. The average size of K_{AS} was measured by multiplying the average size of a diagnostic-unit set by the number of disorders identified in K_{AS}. The experiment shows that the average size of K_{AS} grows asymptotically with experience. For example, Figure 7.2 shows how the size of K_{AS} changed as experience grew for a 95% merging threshold.

Figure 7.2: Change in the size of K_{AS} (*i.e.*, the total number of diagnostic units available in K_{AS}) as a function of experience for a 95% merging threshold

The average final sizes of K_{AS} are shown in the third column of Table 7.1. As shown in the table, an average of approximately 3 diagnostic units were identified for each disorder, even at the "tightest" merging threshold of 95%. Recall that, as shown in Algorithm Evidence-Grouping, the evidence-grouping process examines each diagnostic unit in K_{AS}. The experimental results im-

ply that an average of only 258 diagnostic units[2] need to be examined by the evidence-grouping process. Each diagnostic unit had an average of approximately 5 causal and/or non-causal relation links to other diagnostic units.

7.2 Empirical Analysis of Adaptation Techniques

This thesis developed deep matching adaptation for modifying diagnostic units, which are generally imperfect, so that they are applicable to a given diagnostic problem. An experiment was conducted to compare deep matching adaptation with two other existing adaptation techniques: CASEY-style and simple matching adaptation.

7.2.1 Adaptation Techniques Compared

The following three different evidence-grouping strategies were compared:

- EG-DMAT: The evidence-grouping strategy that applies deep matching adaptation (*i.e.,* Algorithm DMAT described in Section 6.1.2) to each diagnostic unit in K_{AS}.
- EG-CMAT: The evidence-grouping method that applies CASEY-style matching adaptation (*i.e.,* Algorithm CMAT described in Section 6.1.4) to each diagnostic unit in K_{AS}.
- EG-SMAT: The evidence-grouping method that applies simple matching adaptation (*i.e.,* Algorithm SMAT described in Section 6.1.4) to each diagnostic unit in K_{AS}.

[2]The total number of diagnostic units available in K_{AS} can be obtained by multiplying the average size of a diagnostic-unit set by the number of disorders identified. At the end of the experiment, the average size of a diagnostic-unit set was 3.35, and a total of 77 disorders were identified. Thus, the total number of diagnostic units available in K_{AS} was on average about 258 ($\approx 3.35 \times 77$).

7.2.2 Performance Dimensions Used

Each of the evidence-grouping strategies returned diagnostic units which had been adapted to explain subsets of findings in a diagnostic problem. Some of the adapted diagnostic units were true positives, and some were false positives. Performance was measured along the dimensions of false-positive rate and true-positive accountability. The *false-positive rate* is defined as the ratio of the number of false positives to the total number of diagnostic units selected by an evidence-grouping strategy. The *true-positive accountability* is defined as the percentage of findings in a diagnostic problem that the true positives selected by an evidence-grouping strategy can explain. For clarity, let ψ be a diagnostic problem, and $\mathcal{Q}$ be the set of adapted diagnostic units returned by an evidence-grouping strategy. Then,

$$\text{False-positive rate} = \frac{|\mathcal{Q}_F|}{|\mathcal{Q}|}$$

$$\text{True-positive accountability} = \frac{|\cup_{du \in \mathcal{Q}_T} f\langle du \rangle|}{|\psi|} \times 100$$

where $\mathcal{Q}_T$ and $\mathcal{Q}_F$ are the sets of true and false positives, respectively, in $\mathcal{Q}$: $\mathcal{Q} = \mathcal{Q}_T \cup \mathcal{Q}_F$.

7.2.3 Test Procedure

Independent trials were conducted on 50 different random orderings of the 300 patients. In each trial, K_{AS} was initially empty, and all 300 patients were run. For each diagnostic problem, an adaptation technique was applied to diagnostic units that were available in K_{AS} at the time of problem solving. Each time a diagnostic problem was solved, a diagnostic solution was incorporated into K_{AS} which was in turn used in the next diagnosis. A 95% merging threshold was used for knowledge incorporation. This procedure for each trial is summarized by the algorithm shown in Figure 7.3.

Procedure used by a trial for comparing adaptation techniques

Input: 1. A set Ψ of the 300 patients in random order
2. Empty K_{AS}

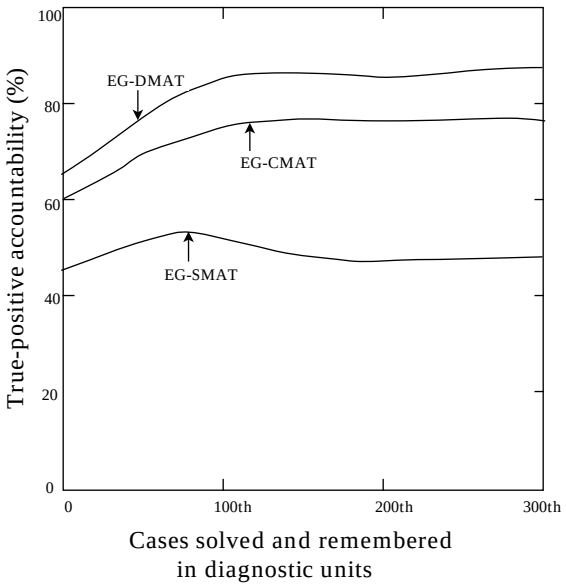
1. For each patient ψ in Ψ
2. For each diagnostic unit du in K_{AS}
3. $Q_D \longleftarrow \text{DMAT}(du, \psi)$; for analysis of EG-DMAT
4. $Q_C \longleftarrow \text{CMAT}(du, \psi)$; for analysis of EG-CMAT
5. $Q_S \longleftarrow \text{SMAT}(du, \psi)$; for analysis of EG-SMAT
6. Measure false positive rate and true-positive accountability for Q_D, Q_C, and Q_S;
7. Incorporate a verified diagnostic solution into K_{AS} by the knowledge incorporation process, using a 95% merging threshold;

Figure 7.3: The procedure used by each trial for comparing adaptation techniques

7.2.4 Analysis

Figure 7.4 plots true-positive accountability as a function of experience. It also summarizes the average true-positive accountability of each evidence-grouping strategy. The results demonstrate that deep and CASEY-style matching adaptation is empirically able to do better than simple matching adaptation, in translating patterns from one case to another.

One thing to note in Figure 7.4 is that after a certain number of cases are solved, changes in the true-positive accountability appear to remain fairly small. This could either be because all possible cases have already been seen or because similar problems occur infrequently. In this study, the latter



Evidence-grouping method	EG-DMAT	EG-CMAT	EG-SMAT
Avg. true-positive Accountability	82%	73%	49%

Figure 7.4: Comparison of evidence-grouping strategies with respect to true-positive accountability

seemed to be a more likely reason for this phenomenon, because if all possible cases had already been seen, then higher accountability than observed (closer to 100%) would be expected. In order to verify this speculation, the following statistic was gathered: For each diagnostic problem, it was measured, retrospectively, what percentage of the diagnostic units in a diagnostic solution had been seen before the problem was solved. One would expect that at least these patterns should be recognized as relevant ones. Such a percentage is referred to as the expected true-positive accountability, or for short expected accountability. According to the experiment, the most likely causal explanation for a problem consisted of an average of 8 diagnostic units (*i.e.*,

8 true positives), and 40% of them were ones that had been seen before the problem was solved. In other words, the average expected accountability was 40%. Figure 7.5 graphically summarizes the distribution of cases in terms of expected accountability. This result empirically verifies that cases with the same particular combinations of diagnoses do not recur routinely, even though precedents could account for different parts of the solution.

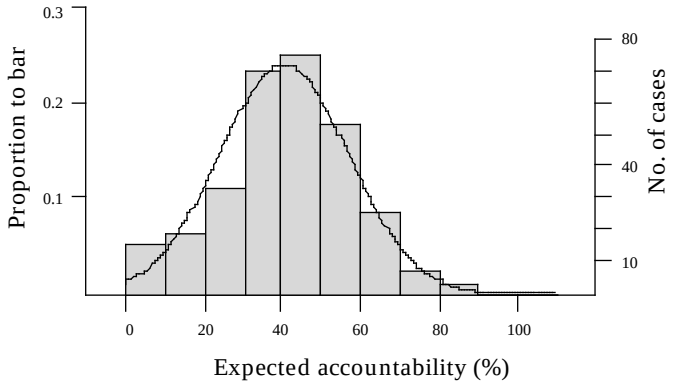


Figure 7.5: The distribution of cases, during the study, in terms of expected accountability

Figure 7.6 attempts to compare evidence-grouping strategies in terms of expected accountability and the true-positive accountability that was actually achieved. For expository convenience, the latter is referred to as actual accountability. In Figure 7.6(a), the x-coordinate represents expected accountability, and the y-coordinate represents actual accountability. The evidence-grouping strategies based on deep and CASEY-style matching adaptation achieved actual accountability higher than expected accountability for most of cases. On the other hand, simple matching adaptation produced relatively poor performance. In terms of average actual accountability, all of the three evidence-grouping strategies achieved an average actual accountability higher than the average expected accountability (40%). Average accountabilities are plotted, for ease of comparison, in Figure 7.6(b). These results empirically demonstrate the advantage of techniques that adapt rules to fit

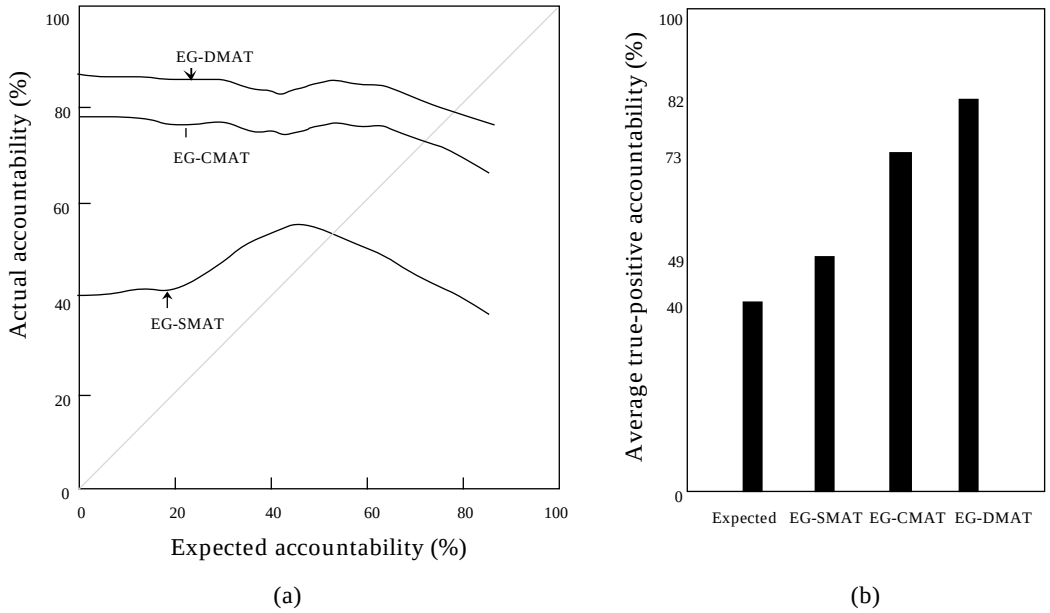
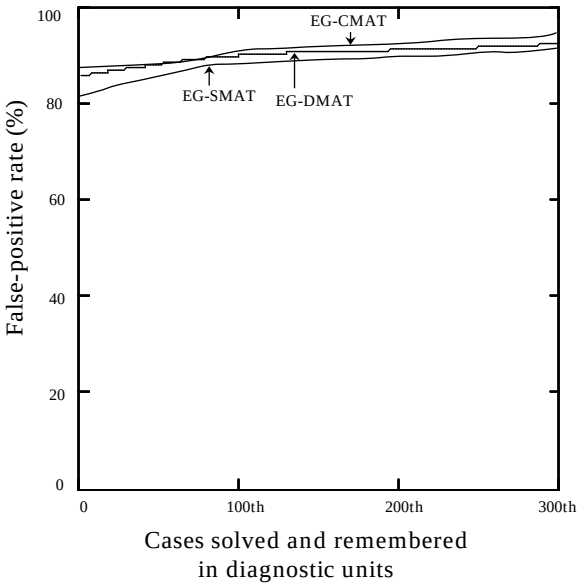


Figure 7.6: Actual accountability *vs.* expected accountability

a problem, over techniques that apply rules on an exact match basis, particularly when rules are imperfect.

Figure 7.7 shows the change in false-positive rates as experience grew. It also shows a table of average false-positive rates for each evidence-grouping strategy. The results shown in the figure empirically verify that adaptation of diagnostic units results in high false-positive rates as well as high true-positive accountability. While false-positives are not desirable, Section 7.3 presents an experiment which demonstrates that to construct a diagnostic solution, such a high false-positive rate is less problematic than one may think, and what is more important is the number of true positives selected.

For ease of comparison, Table 7.2 summarizes the average true-positive accountability and false-positive rate results. The table suggests that deep matching adaptation enjoys an advantage over simple matching adaptation in transferring knowledge from case to case. Causal accounting used in deep



Evidence-grouping method	EG-DMAT	EG-CMAT	EG-SMAT
Avg. false-positive rate	90%	92%	89%

Figure 7.7: False-positive rates as a function of experience and the comparison of matching strategies with respect to average accountability and average false-positive rates

Evidence-grouping method	EG-DMAT	EG-CMAT	EG-SMAT
Avg. true-positive Accountability	82%	73%	49%
Avg. false-positive rate	90%	92%	89%

Table 7.2: Summary of average true-positive accountability and false-positive rate for each evidence-grouping strategy

matching adaptation appears to degrade false-positive rates only slightly, while substantially enhancing true-positive performance. This result empirically supports the importance of causal similarity in recognizing relevant diagnostic units, particularly when matches between cases are only par-

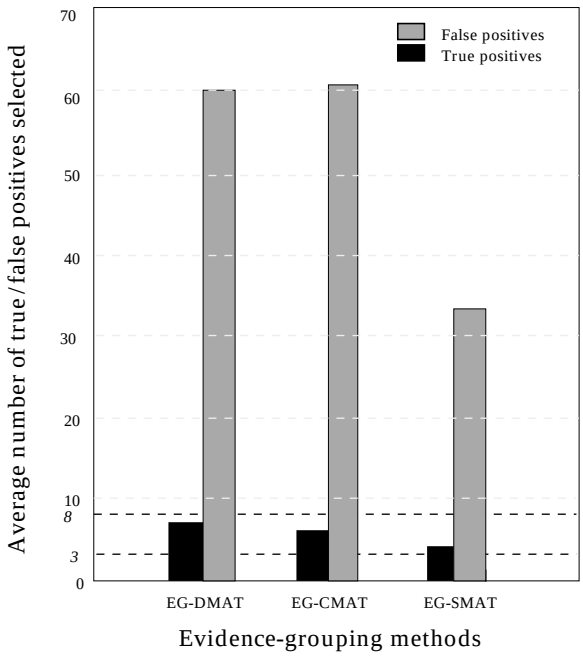
tial. Deep matching adaptation also appears to do better than CASEY-style matching adaptation. The former achieved higher true-positive accountability but lower false-positive rates than the latter (even though the difference in false-positive rates is marginal). From this experiment, we may conclude that deep matching adaptation is an effective technique for taking advantage of similarity between cases, in domains where cases similar on the surface rarely occur.

Finally, Figure 7.8 summarizes the average number of true and false positives chosen by each evidence-grouping strategy. The second column shows the average total number of adapted diagnostic diagnostic units returned by each evidence-grouping strategy. Diagnostic units returned are either true positive or false positive. The third column lists the average number of true positives selected, while the fourth column lists the average number of false positives selected.

As mentioned previously, the most likely explanation for a diagnostic problem consisted of an average of 8 diagnostic units, 3 of them being ones that had been seen before the problem was solved. Note that EG-DMAT was able to find 82% of them *i.e.,* about 6.55 true positives. This demonstrates the capability of deep matching adaptation to identify relevant diagnostic units, particularly when experience is limited, and consequently diagnostic units are imperfect.

Table 7.8 also shows that adaptation substantially reduced the number of diagnostic units that need to be considered to construct a solution (from 258 diagnostic units that were available in K_{AS}[3] to 66 diagnostic units (26% of 258), 67 diagnostic units (26% of 258), and 38 diagnostic units (15% of 258), for EG-DMAT, EG-CMAT, and EG-SMAT, respectively). The outputs of all three evidence-grouping strategies were, however, still heavily populated with false positives. More specifically, 90%, 92%, and 89% of the adapted diagnostic units returned by EG-DMAT, EG-CMAT, and EG-SMAT, respectively,

[3]As discussed in Section 7.1, the average number of diagnostic units in K_{AS}, for a 95% merging threshold, was 258.



Evidence-grouping strategy	Avg. # of diagnostic units selected	Avg. # of true positives selected	Avg. # of false positives selected
EG-DMAT	66.46	6.55	59.91
EG-CMAT	67.48	5.67	61.81
EG-SMAT	37.61	4.02	33.59

Figure 7.8: Performance of evidence-grouping strategies in selecting true positives and false positives for a diagnostic problem

were false positives. The next section reports on an evaluation of the capability of the hypothesis-construction process to deal with such high false-positive rates.

7.3 Empirical Analysis of Hypothesis Construction Strategies

An experiment was conducted to test the effectiveness of the specificity-reflected ranking method and the dependency-guided picking method. In particular, these techniques were compared with the simple similarity metric and the simple picking method. Toward this end, four hypothesis-construction strategies were implemented based on different combinations of similarity metrics and picking methods. The experiment was run on the output of the evidence-grouping process that selects relevant diagnostic units by applying deep matching adaptation (*i.e.,* Algorithm DMAT).

7.3.1 Hypothesis-Construction Strategies Compared

The following four hypothesis-construction strategies were compared.

- HC-STRATEGY$_1$: The hypothesis-construction strategy based on the specificity-reflected similarity metric and the dependency-guided picking method. Algorithm Hypothesis-Construction shown in Section 6.2.4 is used for HC-STRATEGY$_1$.

- HC-STRATEGY$_2$: The hypothesis-construction strategy based on the specificity-reflected similarity metric and the simple picking method. An algorithm for the picking diagnostic units based on these techniques is given in Figure 7.9.

- HC-STRATEGY$_3$: The hypothesis-construction strategy based on the simple similarity metric and the dependency-guided picking method. An algorithm for picking diagnostic units based on these techniques is given in Figure 7.10.

- HC-STRATEGY$_4$: The hypothesis-construction strategy based on the simple similarity metric and the simple picking picking method. An algorithm for this picking method is given in Figure 7.11.

Simple Picking Algorithm for HC-STRATEGY$_2$ **based on the specificity-reflected similarity metric (SPSRSM):**

Input:
1. A set $Unsel$ of adapted diagnostic units that have not been selected
2. A set Sel of adapted diagnostic units that have been selected
3. A set E of findings that have been explained
4. A diagnostic problem ψ

Output: Modified Sel and E

1. While $E \neq \psi$ and $Unsel \neq$ `Nil`
2. $du \longleftarrow \text{SRSM}(Unsel)$;
3. $Sel \longleftarrow Sel \cup \{du\}$;
4. $Unsel \longleftarrow Unsel - \{du\}$;
5. $E \longleftarrow E \cup f\langle du \rangle$;
6. Return Sel and E;

Figure 7.9: Simple Picking Algorithm, based on the specificity-reflected similarity metric, for picking next adapted diagnostic units

For clarity, the four hypothesis-construction strategies are also summarized in Figure 7.12, and algorithms for them are shown in Figure 7.13.

7.3.2 Performance Dimensions used

Each of the four hypothesis-construction strategies returns a set Sel of adapted diagnostic units that are judged to be parts of a diagnostic solution, and a set E of findings that can be explained by the selected adapted diagnostic units. Not all adapted diagnostic units in Sel are true positives. Performance was measured along the dimensions of true-positive accountability and false-positive rate. Because true-positive accountability is the percent-

---

Dependency-Guided Picking Algorithm for HC-STRATEGY$_3$ **based on Simple Similarity Metric (DGPSM):**

Input:
1. A set $Unsel$ of adapted diagnostic units that have not been selected
2. A set Sel of adapted diagnostic units that have been selected
3. A set E of findings that have been explained
4. A diagnostic problem ψ

Output: Modified Sel and E

1. While $E \neq \psi$ and $Unsel \neq$ `Nil`
2. $Candidates \longleftarrow$ The set of adapted diagnostic units in $Unsel$ that are causally or non-causally related to at least one adapted diagnostic unit in Sel;
3. $du \longleftarrow$ IF $Candidates \neq$ `Nil`
4. THEN SSM($Candidates$)
5. ELSE SSM($Unsel$);
6. $Sel \longleftarrow Sel \cup \{du\}$;
7. $Unsel \longleftarrow Unsel - \{du\}$;
8. $E \longleftarrow E \cup f\langle du \rangle$;
9. Return Sel and E;

Figure 7.10: Dependency-Guided Picking Algorithm, based on the simple similarity metric, for picking next adapted diagnostic units

---

Simple Picking Algorithm for HC-STRATEGY$_4$ based on the simple similarity metric (SPSSM):

Input: 1. A set $Unsel$ of adapted diagnostic units that have not been selected
2. A set Sel of adapted diagnostic units that have been selected
3. A set E of findings that have been explained
4. A diagnostic problem ψ

Output: Modified Sel and E

1. While $E \neq \psi$ and $Unsel \neq$ `Nil`
2. $du \longleftarrow \text{SSM}(Unsel)$;
3. $Sel \longleftarrow Sel \cup \{du\}$;
4. $Unsel \longleftarrow Unsel - \{du\}$;
5. $E \longleftarrow E \cup f\langle du \rangle$;
6. Return Sel and E;

Figure 7.11: Simple Picking Algorithm, based on the simple similarity metric, for picking next adapted diagnostic units

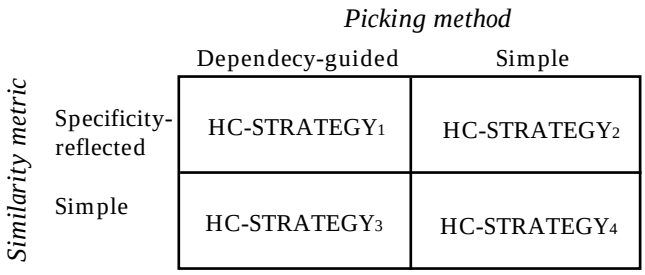


Figure 7.12: Hypothesis-construction strategies based on various combinations of similarity metrics and picking methods

age of findings that are explained by the true positives in Sel, one minus true-positive accountability is the percentage of findings that are either explained by false positives or not explained by any of the diagnostic units in Sel. False-positive rate is the percentage of diagnostic units in Sel that are false positives. Formally,

$$\text{True-positive accountability} = \frac{|\cup_{du \in Sel_T} f\langle du \rangle|}{|\psi|} \times 100$$

$$\text{False-positive rate} = \frac{|Sel_F|}{|Sel|} \times 100$$

where ψ denotes a diagnostic problem, and Sel_T and Sel_F are the sets of true and false positives, respectively, in Sel: $Sel = Sel_T \cup Sel_F$.

7.3.3 Test Procedure

Once again, independent trials were conducted on 50 different random orderings of the 300 patients. Each hypothesis-construction strategy constructed diagnostic solutions from the diagnostic units chosen by applying deep matching adaptation. A 95% merging threshold was used for knowledge incorpo-

Strategies for hypothesis construction

Input: 1. A diagnostic problem ψ
2. A set $\mathcal{Q}$ of adapted diagnostic units returned by Algorithm Evidence-Grouping

Output: Sets of adapted diagnostic units

(a) HC-STRATEGY$_1$;

;; Select adapted diagnostic units by the specificity-reflected
;; similarity metric and the dependency-guided picking method.
1. $du \longleftarrow \text{SRSM}(\mathcal{Q})$;
2. $Sel \longleftarrow \{du\}$;
3. $E \longleftarrow f\langle du\rangle$;
4. Sel and $E \longleftarrow \text{DGP}(\mathcal{Q} - Sel, Sel, E, \psi)$;

(b) HC-STRATEGY$_2$;

;; Select adapted diagnostic units by the specificity-reflected
;; similarity metric and the simple picking method
1. $du \longleftarrow \text{SRSM}(\mathcal{Q})$;
2. $Sel \longleftarrow \{du\}$;
3. $E \longleftarrow f\langle du\rangle$;
4. Sel and $E \longleftarrow \text{SPSRSM}(\mathcal{Q} - Sel, Sel, E, \psi)$;

(c) HC-STRATEGY$_3$;

;; Select adapted diagnostic units by the simple similarity
;; metric and the dependency-guided picking method.
1. $du \longleftarrow \text{SSM}(\mathcal{Q})$;
2. $Sel \longleftarrow \{du\}$;
3. $E \longleftarrow f\langle du\rangle$;
4. Sel and $E \longleftarrow \text{DGPSM}(\mathcal{Q} - Sel, Sel, E, \psi)$;

(b) HC-STRATEGY$_4$;

;; Select adapted diagnostic units by the simple similarity
;; metric and the simple picking method.
1. $du \longleftarrow \text{SRSM}(\mathcal{Q})$;
2. $Sel \longleftarrow \{du\}$;
3. $E \longleftarrow f\langle du\rangle$;
4. Sel and $E \longleftarrow \text{SPSSM}(\mathcal{Q} - Sel, Sel, E, \psi)$;

Figure 7.13: Algorithms for various hypothesis-construction strategies

ration. This test procedure is summarized in Figure 7.14.

Procedure used by a trial for Comparing Hypothesis-Construction Strategies

Input: 1. A set Ψ of the 300 patients in random order
2. Empty K_{AS}

1. For each patient ψ in Ψ
2. For each diagnostic unit du in K_{AS}
3. $\mathcal{Q} \longleftarrow$ Evidence-grouping(du, ψ);
4. $Q_1 \longleftarrow$ HC-STRATEGY$_1(\psi, \mathcal{Q})$;
5. $Q_2 \longleftarrow$ HC-STRATEGY$_2(\psi, \mathcal{Q})$;
6. $Q_3 \longleftarrow$ HC-STRATEGY$_3(\psi, \mathcal{Q})$;
7. $Q_4 \longleftarrow$ HC-STRATEGY$_4(\psi, \mathcal{Q})$;
8. Measure false positive rate and true-positive accountability for Q_1, Q_2, Q_3, and Q_4;
9. Incorporate a verified diagnostic solution to ψ into K_{AS} by the knowledge incorporation process using a 95% merging threshold;

Figure 7.14: A test procedure used for comparing various hypothesis-construction strategies

7.3.4 Analysis

Figure 7.15 summarizes the average performance of each hypothesis-construction strategy, with respect to average true-positive accountability and false-positive rate. For ease of comparison, the results are shown in a matrix format as well as in a tabular form. For each picking method, the specificity-reflected similarity metric did better than the simple similarity metric, with respect

Strategy	Avg. True-Positive Accountability (σ)	Avg. False-Positive Rate (σ)
HC-STRATEGY$_1$	81%	13%
HC-STRATEGY$_2$	71%	12%
HC-STRATEGY$_3$	77%	19%
HC-STRATEGY$_4$	67%	31%

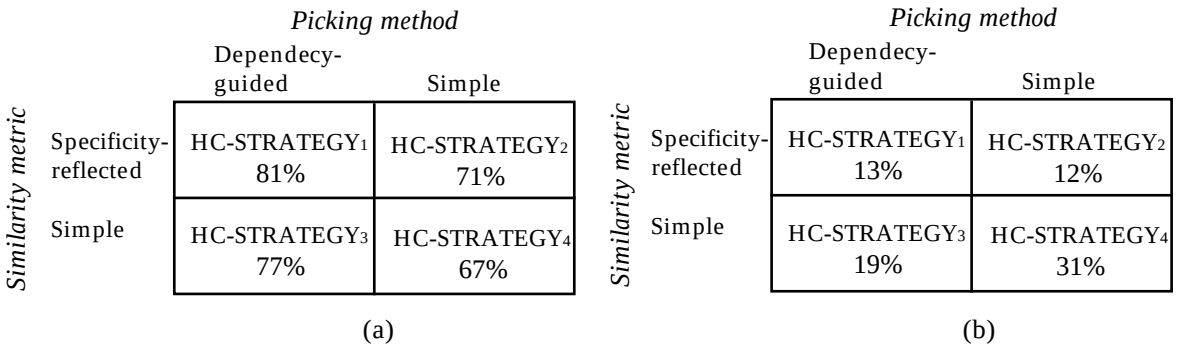


Figure 7.15: Performance of hypothesis-construction strategies based on various combinations of similarity metrics and picking methods; (a) Average true-positive accountability; (b) Average false-positive rate

to both average true-positive accountability and average false-positive rate. Similarly, for each similarity metric, the dependency-guided picking method did better than the simple picking method.

In order to determine the likelihood that these results could be achieved by random chance, the t-test [88] was conducted for each pair of hypothesis-construction strategies. In particular, we are interested in the differences in performance, μ_d and μ'_d, of two hypothesis-construction strategies x and y, where μ_d is the difference between the true-positive accountability of strategy x and that of strategy y, and μ'_d is the difference between the false-positive rate of strategy x and that of strategy y. With respect to true-positive accountability, the data in the samples are the pairs (tx_i, ty_i), $1 \leq i \leq 50$, where tx_i and ty_i are the average true-positive accountabilities of strategies

x and y, respectively, for the ith trial. One hypothesis tested is that strategy x does better, with respect to true-positive accountability, than strategy y, in other words,

$$\mu_d > 0.$$

Similarly, with respect to false-positive rate, the data in the samples are the pairs (fx_i, fy_i), $1 \leq i \leq 50$, where fx_i and fy_i are the average false-positive rates of strategies x and y, respectively. Another hypothesis tested is that

$$\mu'_d < 0$$

i.e, strategy x has lower false-positive rate than strategy y.

The results of the t-tests of differences are summarized in Table 7.3.[4] The results statistically verify that, with respect to true-positive accountability, the hypothesis that $\mu_d > 0$ is true with probability higher than 99% for each pair of hypothesis-construction strategies shown in the table, except the pair HC-STRATEGY$_2$ and HC-STRATEGY$_3$. In other words, the probability that $\mu_d \leq 0$ is true is less than 1%. For the pair HC-STRATEGY$_2$ and HC-STRATEGY$_3$, the t-test shows that with probability higher than 99% HC-STRATEGY$_3$ does better than HC-STRATEGY$_2$ with respect to true-positive accountability. It may be concluded from these results that with probability higher than 99%, HC-STRATEGY$_1$ does better than HC-STRATEGY$_3$, which does in turn better than HC-STRATEGY$_2$, which in turn does better than HC-STRATEGY$_4$.

[4] $\widehat{t_{tp}}$ and $\widehat{t_{fp}}$, in Table 7.3, are computed by the following formula:

$$\hat{t} = \frac{\Sigma d_i / n}{\sqrt{\frac{\Sigma {d_i}^2 - (\Sigma d_i)^2/n}{n(n-1)}}}$$

where n is the number of observations, specifically 50 in this experiment. Each d_i is the difference of true positive accountabilities (*i.e.,* $tx_i - ty_i$) for the hypothesis that $\mu_d > 0$, and of false-positive rates (*i.e.,* $fx_i - fy_i$) for the hypothesis that $\mu'_d < 0$. The corresponding t value for the significance level of 1% for the one sided test is 2.405 when the degree of freedom is 49 [88].

Strategies Compared		w.r.t TPA	w.r.t. FPR
Strategy x	Strategy y	$\widehat{t_{tp}}$	$\widehat{t_{fp}}$
HC-STRATEGY$_1$	HC-STRATEGY$_2$	14.48	1.25
HC-STRATEGY$_1$	HC-STRATEGY$_3$	6.49	-8.08
HC-STRATEGY$_1$	HC-STRATEGY$_4$	16.43	-23.26
HC-STRATEGY$_2$	HC-STRATEGY$_3$	-6.88	-10.59
HC-STRATEGY$_2$	HC-STRATEGY$_4$	4.82	-22.39
HC-STRATEGY$_3$	HC-STRATEGY$_4$	11.21	-16.40

Table 7.3: Results of t-test of difference with respect to true-positive accountability and false-positive rate; TPA and FPR stand for true-positive accountability and false-positive rate, respectively; $\widehat{t_{tp}}$ is significant at the 1% level provided $\widehat{t_{tp}} \geq 2.405$, while $\widehat{t_{fp}}$ is significant at the 1% level provided $\widehat{t_{fp}} \leq 2.405$.

With respect to false-positive rate, the results statistically verify that for each pair of hypothesis-construction strategies shown in the table, the hypothesis that $\mu'_d < 0$ is true with probability higher than 99%. Overall, we may conclude that with the probability higher than 99%, HC-STRATEGY$_1$ does better than HC-STRATEGY$_2$, which does in turn better than HC-STRATEGY$_3$, which does in turn better than HC-STRATEGY$_4$.

The results of the t-tests statistically illustrate the effectiveness of combining the specificity-reflected similarity metric and the dependency-guided picking method, in selecting true positives to construct diagnostic solutions.

7.4 Diagnostic Performance of HYDI

An experiment was conducted to test the effectiveness and efficiency of hybrid reasoning, by comparing it with both of causal-model and association-based reasoning.

Hybrid reasoning was implemented in HYDI. HYDI used HF as its causal-model-based solver. For HYDI's association-based solver, decompositional abductive diagnosis (for expository convenience, referred to as DAD) was implemented that selects relevant diagnostic units based on deep matching

adaptation, and constructs a diagnostic solution by applying the specificity-reflected similarity metric and the dependency-guided picking method. The experiment also included a comparison of HYDI with both of HF and DAD.

HYDI performs hybrid reasoning to solve diagnostic problems. The central idea of the hybrid reasoning architecture is that for a diagnostic problem, DAD first tries to solve the problem; if a diagnostic solution generated by DAD is not acceptable, then HF is called to solve the problem. Two strategies for determining whether or not a diagnostic solution generated by DAD is acceptable were considered. One strategy is to accept a causal explanation generated by DAD if the explanation can account for all the findings in the problem. This strategy was implemented in HYDI_1. The other strategy is to accept a causal explanation if the explanation not only can explain all the findings in the problem, but also is "close to a correct diagnosis." The "close to a correct diagnosis" clause was implemented as a comparison of true-positive accountability with a threshold. During testing, 88%, which is the average true-positive accountability achieved by HYDI_1, was used as the threshold value. This strategy was implemented in HYDI.

Performance was measured along the dimensions of accuracy and running time. Accuracy was measured in terms of true-positive accountability and false-positive rate. Running time was measured on a SUN SPARC Station 2.

Figure 7.16 plots the change in true-positive accountability and false-positive rate for HYDI_1 and HYDI, as experience grew. HYDI did better than HYDI_1, with respect to both true-positive accountability and false-positive rate.

Table 7.4 summarizes average percentage of findings in a diagnostic problem that are explained by a diagnostic solution. Unlike the three other systems, DAD was able to explain only an average of 89% of findings in a diagnostic problem. This empirically demonstrates that the association-based DAD is less robust than the other systems.

Figure 7.17 summarizes the results of the experiment, in terms of aver-

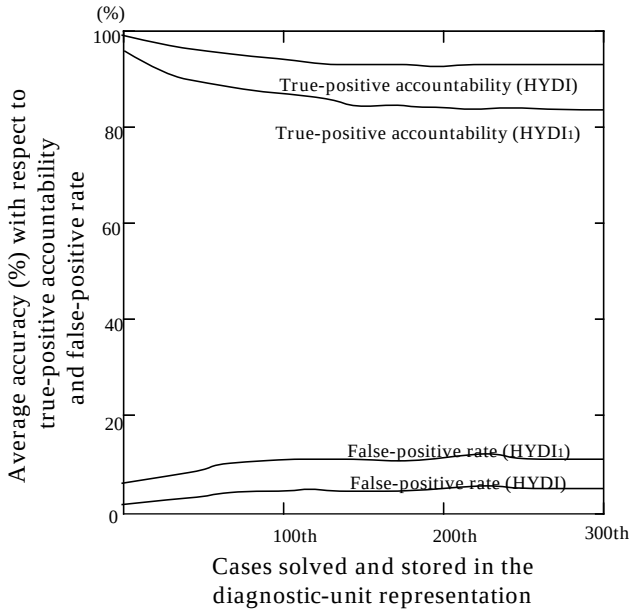


Figure 7.16: Changes in true-positive accountability and false-positive rate as experience grew

A diagnostic solution S to a problem	Average % of findings explained by S
DAD-generated diagnostic solution	88.9%
HF-generated diagnostic solution	100%
$HYDI_1$-generated diagnostic solution	100%
HYDI-generated diagnostic solution	100%

Table 7.4: Average percentage of findings that were explained by a diagnostic solution generated by each system

age accuracy and running time. During evaluation, HF-generated diagnostic solutions were assumed to be "correct" diagnoses. In other words, HF has 0 false-positive rate and 100% true-positive accountability. As summarized in the table shown in Figure 7.17, the experiment indicates that HYDI achieved an average of 96.8% true-positive accountability and 4.3% false-positive rate.

System	Accuracy		Avg. Running Time (σ)
	Avg. TPA	Avg. FPR	
HF	100%	0	52.3 sec (138.2 sec)
DAD	81.0%	13.1%	3.7 sec (12.2 sec)
$HYDI_1$	88.1%	10.9%	4.0 sec (13.6 sec)
HYDI	96.8%	4.3%	17.7 sec (43.1 sec)

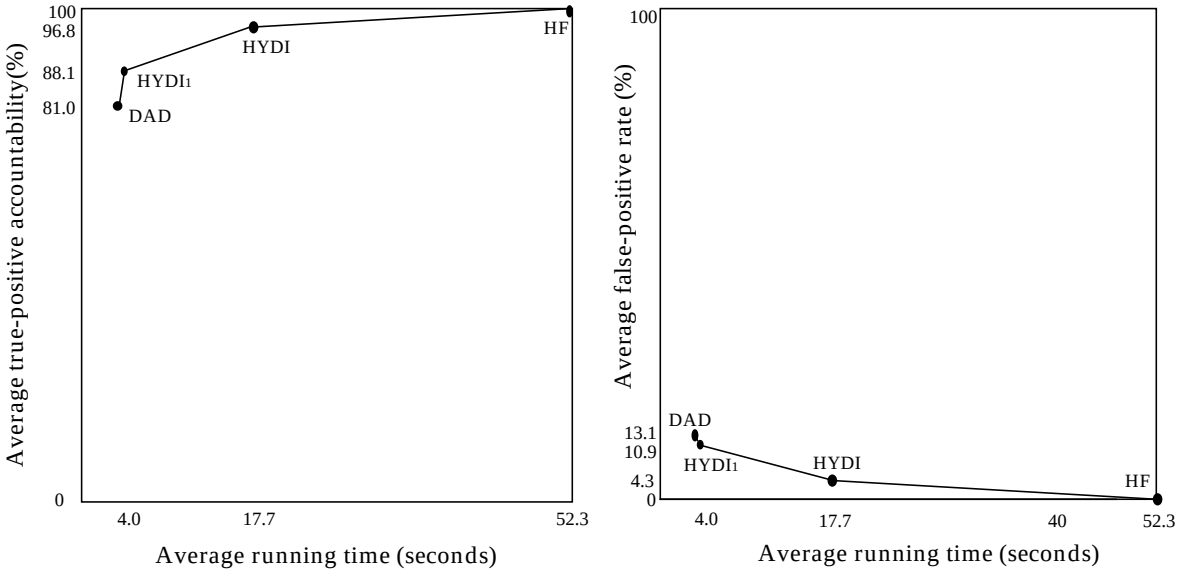


Figure 7.17: Average accuracy and running time

This result may appear disappointing, but in fact is relatively good, given the fact that most of the problem solving in HYDI was done by AS (which based its problem solving on previous experience), and the fact that similar cases occurred infrequently (only an average of 40% of the diagnostic units in a solution were seen previously).

For ease of comparison, the performance figures shown in the table are also plotted in Figure 7.17. This result empirically suggests that attractive tradeoffs between accuracy and efficiency can be achieved. In particular,

compared with HF, HYDI was able to achieve a 300% increase in speed, at only a 3% decrease in accuracy. HYDI also shows a smaller standard deviation in running time than HF.

Finally, the experiment shows that as experience was accumulated, CMS was called less frequently for diagnosis. In HYDI_1, after an average of 27 cases, DAD was generally able to find some diagnostic solutions that explain all the findings. In HYDI, even though CMS was called more frequently than in HYDI_1, it was still called less as experience was gathered (See Figure 7.18).

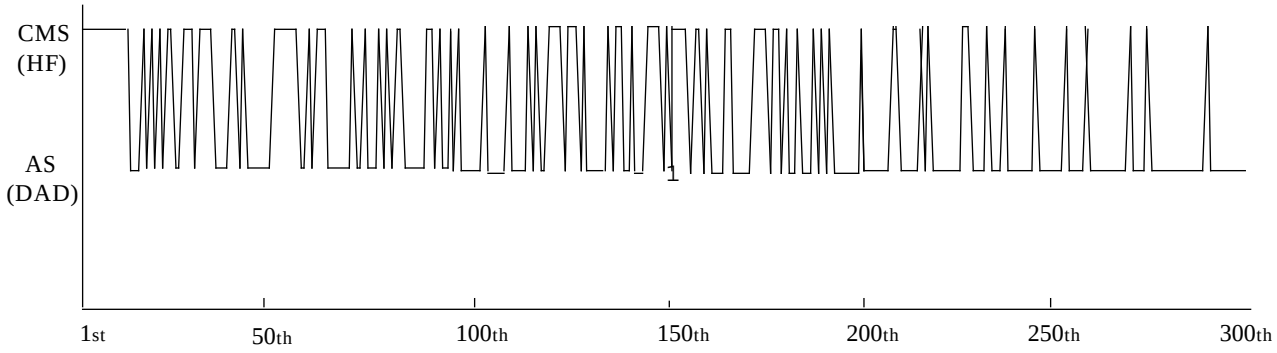


Figure 7.18: Changes in problem solvers within HYDI

Chapter 8

Conclusion

8.1 Summary

This thesis has developed a new approach to diagnosis, based on integrating association-based reasoning, causal-model-based reasoning, and learning techniques into a single system. The result is a hybrid system that can solve familiar problems efficiently by learning from its own experience, while maintaining the ability to reason from first principles when necessary. Algorithms for the hybrid system have been developed, and the efficiency and effectiveness of the algorithms have been tested against existing techniques. Major results of this thesis are summarized in this chapter.

8.1.1 Hybrid Reasoning Architecture

An hybrid reasoning architecture was designed for combining association-based reasoning and reasoning from a causal model. The central idea is to solve diagnostic problems, whenever possible, using only association-based reasoning; if association-based reasoning fails to solve a problem, then the problem is solved step-by-step from first principles. While an existing causal-model-based system was reused, novel techniques were developed for performing decompositional abductive diagnosis with rules which are generally

imperfect. The work on decomposition and abduction was motivated by their potential efficiency.

The effectiveness and efficiency of the hybrid reasoning architecture was tested, in the domain of heart failure diagnosis, against each of causal-model-based and association-based reasoning. The empirical results indicate that the hybrid system achieves efficiency comparable to that of association-based reasoning but with robustness which approaches that of causal-model-based reasoning.

8.1.2 Modular Representation of Knowledge about Context Sensitivity of Findings

This thesis developed a new formal representation for domain-structure knowledge. This representation accounts, to a large extent, for the efficiency and effectiveness of decompositional abductive diagnosis.

A common existing representational scheme is the bipartite "If-Then" representation in which the unit of rules is an entire problem-solving situation and its solution, *i.e.,* "If a problem, then a solution." In the diagnosis domain, a problem is a set of findings upon which a diagnosis is based, and a solution is a hypothesis which can account for all the findings in the problem. In the multifault diagnosis domain, a hypothesis is typically a set of disorders. The direct mapping of problem situations to their solutions allows a problem to be solved quickly, by simply applying applicable rules. The method is fragile, however, unless all possible problem-solving situations are specified in advance. Unfortunately, because domains like medical diagnosis often involve multiple, interacting diseases, findings and disease states can be combined in a prohibitively large number of ways. As a consequence, these domains are not amenable to the If-Then representation, because they require that too many rules be specified. Another problem with this representation is that it generates "rote memory" solutions which do not specify which disorders cause which findings and how.

This thesis has dealt with these problems by using smaller, modular abstractions that encode the context sensitivity of findings. The approach was motivated by the following observation about diagnoses. While the domain at the level of problems and solutions appears to have little regularity, collections of findings, intermediate states, and diseases often recur individually, thus suggesting a fine-grain regularity in the domain. Diagnostic units are modular abstractions that capture such regularity in the domain. Each diagnostic unit includes a single disease and its pathophysiologic effects. It also records the co-occurring diseases that can directly influence this pathophysiology, to account for disease interactions. By identifying the various finding contexts associated with each disease, diagnostic units, unlike If-Then rules, provide guides for structuring evidence into subproblems for which the most likely explanations can be immediately inferred.

8.1.3 Transformation of Experiences into General Problem-Solving Knowledge

This thesis has described possible experience-based strategies for acquiring and refining diagnostic units. Instead of storing a problem and its solution as an atom, HYDI analyzes them, and incorporates the derived knowledge into abstract, general principles. The basic approach is to transform experience into such knowledge by remembering solved cases in a decomposed-and-merged form. Such knowledge is then used in solving familiar problems in the future.

8.1.4 Decompositional Abductive Diagnosis

Once knowledge is available in the diagnostic-unit representation, diagnosis for a problem can be made by activating the diagnostic units supported by the findings, and then combining selected diagnostic units. The following techniques were developed for performing such decompositional abductive diagnosis.

Deep Matching Adaptation: The selection of the diagnostic units relevant to a problem raises the issue of how to match. One approach would be to find exact partial matches, in other words, to select a diagnostic unit if the findings specified in it are a subset of the given findings. Unfortunately, diagnostic units that are acquired from experience are generally imperfect, and exact partial matches are not likely to be found. This thesis developed deep matching adaptation which locally adjusts diagnostic units so that they can match a problem.

Deep matching adaptation tailors a diagnostic unit to a particular case in two ways – by adding unmatched problem findings to the diagnostic unit and by removing unmatched diagnostic-unit findings from the unit. For adding unmatched problem findings to a diagnostic unit, causal accounting was investigated. Experiments empirically support the advantage of deep matching adaptation over the existing simple and CASEY-style matching adaptation techniques, particularly when diagnostic units are imperfect.

Specificity-Reflected Similarity Metric: The specificity-reflected similarity metric was designed for gauging the level of similarity between adapted diagnostic units and a problem. Experiments indicate that taking into account the ability of a finding to identify the existence of a diagnostic unit is useful heuristics when selecting relevant diagnostic units.

Dependency-Guided Picking Method: In addition, a disorder-dependency-based technique was developed for constructing overall solutions from adapted diagnostic units. The empirical results indicate that dependency among disorders is a good guide for directing the hypothesis-construction process, handling interacting disorders and findings appropriately.

8.1.5 Summary of Analysis

HYDI starts out with a causal model and an empty associative knowledge base. It then develops its associative knowledge by learning, from its own problem-solving experience, "essential" disease descriptions that identify which causal relations in a causal model are important for diagnosis and which can be ignored. These descriptions are associations of sets of findings and diseases that regularly appear in diagnoses, and are represented as diagnostic units. In diagnostic units, detailed reasoning structures are compiled into simple associations between sets of findings and their most highly likely hypotheses. HYDI can save time by retrieving relevant descriptions rather than dynamically generating new ones each time a problem is solved.

- As empirically demonstrated, the diagnostic-unit representation appears to be an effective way of capturing domain decomposability. The capture of domain decomposability in turn facilitates the recognition of problem structure, and allows an expanded capability to solve complex diagnostic problems.

- As it solves more problems, HYDI automatically acquires and refines, from its own experience, diagnostic units. HYDI then gains efficiency by retrieving and combining diagnostic units, which represent partial solutions, rather than recomputing them from scratch. Thus, HYDI's efficiency improves with experience.

HYDI also demonstrates that combining association-based reasoning with reasoning from first principles can have significant advantage over each method used alone.

- HYDI overcomes the efficiency limitations of model-based reasoning and the inability of association-based reasoning to deal with unfamiliar problem situations. It efficiently solves familiar problems by retrieving diagnostic units, making small local changes to them, and then combining them. When HYDI recognizes that it does not know how to solve a

particular problem, however, it solves the problem from first principles by using HF.

HYDI also overcomes some of the major weaknesses of case-based reasoning.

- First, parts of past cases can be accessed and used easily, by transforming the coarse-grain cases into finer-grain diagnostic units and relationships between them.
- Second, even though dissimilar on the surface, HYDI can discern that retrieved diagnostic units are applicable to a new problem by analyzing differences and tailoring them to the problem. Empirical results suggest that the techniques developed for using diagnostic units to perform decompositional abductive diagnosis effectively exploit the structure inherent in the diagnosis domain.

8.2 Future Work

There are many ways in which the work described here could be extended This section describes some of them.

Hypothesize-test-refine model: HYDI can be extended to a hypothesize-test-refine paradigm. In many diagnostic problems, determining what to do next is as important as determining what disorders are present. Diagnostic units could help determine the next set of tests to perform, and make predictions outlining the expected results. According to a comparison of the expected results to the observed results, a system could select amongst competing diagnostic units and refine its diagnoses.

Multiple-level models and reasoning on multiple levels of detail: Diagnostic units could be generalized to multiple-level models, by not-

ing similarities between them. For example, diagnostic units representing aortic valve disease and aortic stenosis could be generalized to aortic disease. The similarity between aortic valve disease and aortic stenosis and the ways of differentiating them can be specified within aortic disease unit. The issue that must be dealt with is how to determine suitable levels of abstraction. One approach would be a generalization method used in CASEY. CASEY generalizes two diagnostic solutions if they share findings, disorders, or intermediate states. The result of generalization is a concept which consists of findings, disorders, and intermediate states that are common in both of the solutions. Unfortunately, this method exhibits substantial combinatorial complexity problems. More importantly, not all generalizations are meaningful abstractions. It is worth noting that a mechanism which can automatically discover appropriate levels of abstraction would be useful for not only diagnosis but also medical training.

Reasoning on multiple levels of description also raises the issue of how to access generalized descriptions. An indexing mechanism is required that gives the problem solver control over the level of detail being used at a given time.

Reasoning about noise: Like HF, the current implementation of HYDI assumes that input findings have been interpreted and filtered by a user. In medical diagnosis, like many other real-world domains, some findings can be "noisy" (including findings outside of the domain) or insignificant, and need not be explained. HYDI could be extended to reason about noisy, insignificant findings, using a detailed causal model and knowledge represented as diagnostic units.

Learn from failure: Failure could be used to reduce the chances of making faulty diagnoses in the future. When a solution produced by DAD is judged to be unacceptable, HYDI simply stores the solution generated by HF. It does not, however, make use of the information to determine

what mistake it made in arriving at the faulty solution. By examining the faulty solution generated by DAD and the correct solution generated by HF, HYDI could find the differences between the correct solution and the knowledge used to make a faulty substitution, and adjust knowledge in K_{AS} to reflect that the substitution is not acceptable.

Similarly, HYDI can be extended to incorporate a user's opinion, by allowing a user to correct a solution generated by HYDI. By examining the solution corrected by the user and the original faulty solution, HYDI could once again find the differences, and adjust knowledge in K_{AS} accordingly.

Refining HF's knowledge base: This thesis uses HF as a standard, assuming that HF produces correct solutions. While HF contains many basic principles of physiology, it also contains uncertain knowledge, and employes heuristics to generate solutions. As a result, it is not always guaranteed that the solutions generated by HF are correct diagnoses; occasionally HF gives incorrect answers. HYDI could be extended to identify the knowledge in HF that might have led to the faulty conclusions. This would also require HYDI to examine the solution generated by HF and the solution corrected by a user, and determine what knowledge in HF was responsible.

Evaluation: Currently, the techniques developed in this thesis are tested on only 300 patients. It would be interesting to see how accuracy and efficiency change with larger numbers of patients. In addition, evaluation of these techniques in other domains would help to further develop them.

Bibliography

[1] B. Adelson. "When novices surpass experts: The difficulty of a task may increase with expertise." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:3.

[2] D.S. Aghassi. "Evaluating Case-based Reasoning for Heart Failure Diagnosis." TR 478, Massachusetts Institute of Technology, Laboratory for Computer Science, Cambridge, MA, 1990.

[3] J.R. Anderson. *Cognitive Skills and Their Acquisition*, Hillsdale, NJ: Erlbaum, 1981.

[4] B. Arbab and D. Michie. "Generating Rules from examples." *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 1985.

[5] W.M. Bain. "A Case-based Reasoning System for Subjective Assessment." *Proceedings of the National Conference on Artificial Intelligence*, American Association for Artificial Intelligence, 1986.

[6] E.R. Bareiss, B.W. Porter, and C.C. Weir. "PROTOS: an examplar-based learning apprentice." Special issue on 2nd Knowledge acquisition for knowledge-based workshop, *International Journal of Man-Machine Studies*, 1988.

[7] R. Barletta and W. Mark. "Explanation-Based Indexing of Cases." In *Proceedings of a Workshop on Case-Based Reasoning,* Morgan Kaufmann, 1988.

[8] J.S. Bennett. "A Knowledge-based system for acquiring the conceptual structure of a diagnostic expert system." *Journal of Automated Reasoning*, 1, pages 49-74, 1985.

[9] B.P. Bloomfield. "Capturing expertise by rule induction." *Knowledge Engineering Review*, 1(4): 30-36, 1986.

[10] A.L. Brown. "Knowing when, where, and how to remember: A problem of metacognition." In R. Glaser *Advances in Instructional Psychology* 1, pages 77-165, NJ: Lawrence Erlbaum Assoc., 1978.

[11] B.G. Buchanan and E.H. Shortliffe. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project.* Addison-Wesley, MA, 1984.

[12] H. Bushke. "Learning is organized by chunking." *Journal of Verbal Learning and Verbal Behavior*, 15, pages 313-324.

[13] T. Bylander, D. Allemang, M. Tanner, and J.R. Josephson. "The computational Complexity of abduction." *Artificial Intelligence,* 49:25-60, 1991.

[14] J.G. Carbonell. "Derivational Analogy: A theory of reconstructive problem solving and expertise acquisition." In *Machine Learning*, volume 2. Morgan Kaufmann, Los Altos, CA, 1986.

[15] B. Chandrasekaren and S. Mittal. "Deep versus compiled Knowledge approaches to diagnostic problem-solving." *Proceedings of the National Conference on Artificial Intelligence*, 1982.

[16] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms.* The MIT Press, Cambridge, MA, 1990.

[17] W.J. Clancey. "Knowledge Acquisition for classification expert systems." In *Proceedings of ACM Annual Conference*, pages 11-14, 1984.

[18] P.T. Cox and T. Pietrzykowski. "General diagnosis by abductive inference." In *Proceedings of the 1987 Symposium on Logic Programming,* pages 183-189, 1987.

[19] R. Davis, B.G. Buchanan, and E.H. Shortliffe. "Production Rules as a Representation for a Knowledge-Based Consultation Program." In W.J. Clancey and E.H. Shortliffe, editors, *Readings in Medical Artificial Intelligence*, pages 98–130, Addison-Wesley, Reading, MA, 1984.

[20] R. Davis. "Diagnostic reasoning based on structure and behavior." *Artificial Intelligence*, 24:347-410, 1984.

[21] R. Davis and W. Hamscher. "Model-based reasoning: Troubleshooting." In H. Shrobe, editor, *Exploring Artificial Intelligence: Survey talks from the National Conference on Artificial Intelligence,* pages 297-346, Morgan Kaufman, 1988.

[22] G. Dejong and R. Mooney. "Explanation-based learning: An alternative view." *Machine Learning*, 1(2), 1986.

[23] J. de Kleer and B.C. Williams. "Diagnosing multiple faults." *Artificial Intelligence,* 32:97-130, 1987.

[24] R. Descartes. *Discours de la Methode,* 1637.

[25] D.A. Evans and V.L. Patel. *Cognitive Science in Medicine*, MIT Press, Cambridge, 1989.

[26] L.M. Fagan, E.H. Shortliffe, and B.G. Buchanan. "Computer-Based Medical Decision Making: From MYCIN to VM." In W.J. Clancey and E.H Shortliffe, editors, *Readings in Medical Artificial Intelligence: The First Decade*, Addison Wesley, 1984.

[27] P.J. Feltovich, P.E. Johnson, J.H. Moller, and D.B. Swanson. "LCS: the role and development of medical knowledge in diagnostic expertise." In W.J. Clancey and E.H. Shortliffe, editors, *Readings in Medical Artificial Intelligence*, pages 275-319, Addison-Wesley, Reading, MA, 1984.

[28] Richard E. Fikes, Peter E. Hart, and Nils J. Nilsson. "Learning and executing generalized robot plans." *Artificial Intelligence*, 3:251–288, 1972.

[29] D. Fisher. "Improving inference through conceptual clustering." *Proceedings of American Association of Artificial Intelligence 1987*, pages 461-465, Morgan-Kaufmann, 1987.

[30] J.H. Friedman, J.L. Bentley, and R.A. Finkel. "An algorithm for finding best matches in logarithmic expected time." *ACM Transactions on Mathematical Software*, 3(3):209-226, 1977.

[31] B.R. Gains. "An overview of knowledge acquisition and transfer." *International Journal of Man-Machine Studies*, 26(4):453-472, 1987.

[32] H. Gelernter. "Realization of a geometry-theorem proving machine." In E.A. Feigenbaum and J. Feldman, editors, *Computers and Thought*, pages 134-152, McGraw-Hill, New York, 1963.

[33] A. Goel. "A Model-Based Approach to Case Adaptation." In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* pages 143-148, 1991.

[34] G.A. Gorry. "Strategies for computer-aided diagnosis." *Mathematical Biosciences,* 2:283-318, 1968.

[35] R. Greiner, B.A. Smith, and R.W. Wilkerson. "A correction to the algorithm in Reiter's theory of diagnosis" (research note). *Artificial Intelligence,* 41:79-88, 1989.

[36] K.J. Hammond. "CHEF: A model of case-based planning." In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 267-271, 1986.

[37] W.C. Hamscher. *Model-based Troubleshooting of Digital Systems.* PhD thesis, Massachusetts Institute of Technology, August 1988.

[38] G. Harman. "The inference to the best explanation." *Philosophical Review,* LXXIV:88-95, 1965.

[39] F. Hayes-Roth. "Learning by example." In *Cognitive Psychology and Instruction*, A.M. Lesgold, *et al.*, (eds.), 1978.

[40] B. Hayes-Roth and F. hayes-Roth. "Concept learning and the recognition and classification of examplars." *Journal of Verbal Learning and Verbal Behavior*, 16:119-136, 1977.

[41] F. Hayes-Roth and J. McDermott. "Knowledge acquisition from structural descriptions." In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, 1977.

[42] D. Heckerman and R.A. Miller. "Towards a better understanding of INTERNIST-1 knowledge base." In R. Salamom, B. Blum, and M.

Jørgensen, editors, *MEDINFO-86: Proceedings of the Fifth Conference on Medical Informatics,* pages 22-26, Washington, North-Holland, 1986.

[43] D.E. Heckerman. "A tractable inference algorithm for diagnosing multiple disease." In M. Henrion, R.D. Shachter, L.N. Kanal, and J.F. Lemmer, editors, *Uncertainty in Artificial Intelligence 5, Machine Intelligence and Pattern Recognition Series,* 10:263-171. North-Holland, 1990.

[44] T.R. Hinrichs. "Toward an architecture for open world problem solving." In *Proceedings of a Workshop on Case-Based Reasoning,* Morgan Kaufmann, 1988.

[45] J.R. Hobbs, M. Stickel, P. Martin, and D. Edwards. "Interpretation as abduction." In *Proceedings of the* 26^{th} *Annual Meeting of the Association of Computational Linguistics,* pages 95-103, 1988.

[46] P. E. Johnson, A.S. Duran, F.A. Hassebrock, J. Moller, and M. Prietula. "Expertise and Error in Diagnostic Reasoning." *Cognitive Science* 5, 1981.

[47] P.E. Johnson. "What kind of Expert Should a System be?" *Journal of Medicine and Philosophy* 8, 1983.

[48] J.P. Kassirer and G.A. Gorry. "Clinical problem solving: A behavioral analysis." *Ann. Intern. Med.,* 89:245-255, 1978.

[49] R.M. Keller and S.T. Kedar-Cabelli. "Explanation-based generalization: A unifying view." *Machine Learning*, 1(1), 1986.

[50] D. Kibler and D.W. Aha. "Learning Representative Examples of Concepts: An Initial Case Study." In *Proceedings of Fourth International Workshop on Machine Learning*, pages 24-30, Irvine, CA: Morgan Kaufmann, 1987.

[51] J.L. Kolodner, R.L. Simpson, Jr., and K. Sycara, "A process model of case-based reasoning in problem solving." In *Proceedings of the National Conference on Artificial Intelligence*, pages 284-290, American Association for Artificial Intelligence, 1985.

[52] J.L. Kolodner and R. Kolodner. "Using Experience in Clinical Problem Solving: Introduction and Framework." In *IEEE Transactions on Systems, Man, and Cybernetics,* 17, pages 420-431, 1987.

[53] J.L. Kolodner. "Extending problem solving capabilities through case-based inference." In *Proceedings of the* 4^{th} *International Machine Learning Workshop*, Morgan Kaufman, 1987.

[54] J.L. Kolodner. "Retrieving the Best Case for a Case-Based Reasoner." In *Proceedings of the Eleventh Annual Conference of the Cognitive Science Society,* pages 155-162, Lawrence Erlbaum, 1989.

[55] P.A. Koton. "Using experience in learning and problem solving." PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[56] C.A. Kulikowski and S.M. Weiss. "Representation of expert knowledge for consultation: The CASNET and EXPERT projects." In P. Szolovits, editor, *Artificial Intelligence in Medicine*, Westview Press, 1982.

[57] J.E. Laird, P.S. Rosenbloom, and A. Newell. "Chunking in SOAR: The anatomy of a general learning mechanism." *Machine Learning*, 1:11-46, 1986.

[58] J. Lancaster and J.L. Kolodner. "Varieties of learning from problem solving experience." In *Proceedings of the Tenth Annual Conference of the Cognitive Science Society,* Lawrence Erlbaum, 1988.

[59] Laurence Urdang Associates Ltd. *The Bantam Medical Dictionary*, Bantam Books, 1982.

[60] G. Leibniz. *Philosophische Schriften,* 1690.

[61] W.J. Long, S. Naimi, M.G. Criscitiello, and R. Jayes. "The development and use of a causal model for reasoning about heart failure." In *Symposium on Computer Applications in Medical Care*, pages 30–36, IEEE, 1987.

[62] W.J. Long. "Medical diagnosis using a probabilistic causal network." *Applied Artificial Intelligence*, 3:367-383, 1989.

[63] W.J. Long, S. Naimi, and M.G. Criscitiello. "Development of a Knowledge Base for Diagnostic Reasoning in Cardiology." *Computers and Biomedical Research,* 25:292-311, Academic Press, 1992.

[64] W.J. Long, S. Naimi, M.G. Criscitiello, and R. K. Adusumilli.

"Validation of a causal probabilistic medical knowledge base for diagnostic reasoning." *Deep Models for Medical Knowledge Engineering,* ed. E. Keravnou, pages 147-160, Elsevier Science Publishers B.V., 1992.

[65] R.A. Miller, H.E. Pople, Jr., and J.D. Myers. "Internist-1, an experimental computer-based diagnostic consultant for general internal medicine." *New England Journal of Medicine,* 307:468-476, 1982.

[66] R.A. Miller, M.A. McNeil, S.M. Challinor, F.E. Masari, Jr., and J.D. Myers. "The Internist-1/Quick Medical Reference project - Status report." *Western Journal of Medicine,* 145:816-822, 1986.

[67] Miller and Starr. *The structure of human decision*, England Cliffs, NJ, Prentics-Hall, 1967.

[68] S. Minton, J.G. Carbonell, C.A. Knoblock, D.R. Kuokka,O. Etzioni, and Y. Gil. "Explanation-based Learning: Optimizing Problem Solving Performance through Experience." In *Paradigms for Maching Learning*, 1989.

[69] R.M. Mitchell. "Generalization as Search." *Artificial Intelligence*, 18:203-226,1980.

[70] A. Newell and H.A. Simon. *Human Problem Solving.* Prentice-Hall, 1972.

[71] V.L. Patel and G.J. Groen. "Knowledge based solution strategies in medical reasoning." *Cognitive Science*, 10:91-116, 1986.

[72] R.S. Patil. "Causal representation of patient illness for electrolyte and acid-base diagnosis." PhD thesis, TR-267, Massachusetts Institute of Technology, 1981.

[73] R.S. Patil. "Artificial intelligence techniques for diagnostic reasoning in medicine." Survey talk presented at the 6^{th} National Conference on Artificial Intelligence (AAAI-87), 1987.

[74] S.G. Pauker, G.A. Gorry, J.P. Kassirer, and W.B. Schwartz. "Towards the Simulation of Clinical Cognition: Taking a Present Illness by Computer." In W.J. Clancey and E.H. Shortliffe, editors, *Readings in Medical Artificial Intelligence*, pages 131–159, Addison-Wesley, Reading, MA, 1984.

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

[76] C.S. Peirce. *Collected Papers.* Harvard University Press, 1960.

[77] Y. Peng and J.A. Reggia. "Plausibility of diagnostic hypotheses: The nature of simplicity." In *Proceedings of the National Conference on Artificial Intelligence,* pages 140-145, American Association for Artificial Intelligence, 1986.

[78] Y. Peng and J.A. Reggia. *Abductive Inference Methods for Diagnostic Problem Solving,* Springer Verlag, 1990.

[79] H.E. Pople Jr. "Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics." In *Artificial Intelligence in Medicine,* ed. P. Szolovits, volume 51 of AAAS SELECTED SYMPOSIUM SERIES, pages 119-190, Westview Press, Boulder, Colorado, 1982.

[80] J.R. Quinlan. " Generating production rules from decision trees." *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, pages 304-307, Milan, Italy: Morgan Kaufmann, 1987.

[81] M. Redmond. "Distributed Cases for Case-Based Reasoning; Facilitating Use of Multiple Cases." In *Proceedings of the Eighth National Conference on Artificial Intelligence,* pages 304-309, 1990.

[82] J.A. Reggia, D.S. Nau, and P.Y. Wang. "Diagnostic expert systems based on a set covering model." *International Journal of Man-Machine,* 32:57-96, 1987.

[83] R. Reiter. "A theory of diagnosis from first principles." *Artificial Intelligence,* 32:57-96, 1987.

[84] E. Rissland and K. Ashley. "Credit Assignment and the Problem of Competing Factors in Case-Base Reasoning." In *Proceedings of DARPA Workshop on Case-Based Reasoning,* pages 327-344, 1988.

[85] P.S. Rosenbloom and A. Newell. "The Chunking of goal hierarchies: A generalized model of practice." *Universal subgoaling and Chunking: The Automatic Generation and Learning of Goal Hierarchies.* Kluwer Academic Publishers, Boston, MA, 1986.

[86] R. C. Schank. *Dynamic memory: A theory of Reminding and Learning in Computers and People*, Cambridge University Press, Cambridge, 1982.

[87] A.L. Samuel. "Some studies in Machine Learning Using the Game of Checkers." In E.A. Feigenbaum and J. Feldman, editors, *Computers and Thought.* NY: MaGraw-Hill, 1963.

[88] L. Sachs. *Applied Statistics: A Handbook of Techniques.* Springer series in statistics, Springer-Verlag, 1984.

[89] B. Selman and H.J. Levesque. "Abductive and default reasoning: A computational core." In *Proceedings of the National Conference on Artificial Intelligence,* pages 343-348. American Association for Artificial Intelligence, 1990.

[90] E.H. Shortliffe. *MYCIN: Computer-based Medical Consultations*, American Elsevier, 1976.

[91] R.G. Simmons. "Combining Associational and Causal Reasoning to Solve Interpretation and Planning Problems." PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1988.

[92] H.A. Simon. "The structure of ill-structured problems." *Artificial Intelligence,* 4:181-201, 1973.

[93] H.A. Simon. *A study of Decision-making Processes in Administrative Organization*, NY, Free Press, 1976.

[94] R.L. Simpson. "A Computer Model of Case-Based Reasoning in Problem Solving: An Investigation in the Domain of Dispute Mediation." GIT-ICS-85/18, Georgia Institute of Technology, 1985.

[95] J.R. Slagle. "A heuristic program that solves symbolic integration problems in freshman calculus." *Journal of the Association for Computing machinery,* 10:507-520, 1963.

[96] K.P. Sycara and D. Navinchandra. "Influences: A Thematic Abstraction for Creative Use of Multiple Cases." In *Proceedings of a Workshop on Case-Based Reasoning,* Morgan Kaufmann, 1991.

[97] A. Tveretzky and D. Kahnemann. "Causal Schemata in judgments under uncertainty." In *Progress in Social Psychology*, ed. M. Fishbein. Hillsdale, N.J., Lawrence Erlbaum, 1977.

[98] S.M. Weiss, C.A. Kulikowski, S. Amarel, and A. Safir. "A Model-Based Method for Computer-Aided Medical Decision Making." *Artificial Intelligence,* 11:145-172, 1978.

[99] L. Wilkinson, M. Hill, and E. Varg. *SYSTAT Graphics,* SYSTAT, Inc. 1989.

[100] P.H. Winston. "Learning structural descriptions from examples." In P.H. Winston, editor, *The Psychology of Computer Vision*, McGraw-Hill, 1975.

[101] T.D. Wu. "Efficient diagnosis of multiple disorders based on a symptom clustering approach." In *Proceedings of the National Conference on Artificial Intelligence,* pages 357-364. American Association for Artificial Intelligence, 1990.

[102] T.D. Wu. "A Decompositional Search Algorithm for Efficient Diagnosis of Multiple Disorders." PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1992.