Domain Structure and the Complexity of Diagnostic Problem Solving

Thomas D. Wu
MIT Laboratory for Computer Science
545 Technology Square
Cambridge, Massachusetts 02139

Abstract

This paper provides a quantitative analysis of domain structure and its effects on the complexity of diagnostic problem solving. It introduces a hypothesis about the modular structure of domains and proposes a measured called explanatory power. The distribution of explanatory power reveals the inherent structure of domains.

We conjecture that such structure might facilitate problem solving, even when the problem solver does not exploit it explicitly. To test this hypothesis, we create a domain without structure by randomizing the distribution of explanatory power. We use the structured and randomized knowledge bases to study the effect of domain structure on two diagnostic algorithms, candidate generation and symptom clustering. The results indicate that inherent domain structure, even when not encoded explicitly, can facilitate problem solving. Such facilitation occurs for both the candidate generation and symptom clustering algorithms. Moreover, domain structure appears to benefit symptom clustering more than candidate generation, suggesting that the efficiency of symptom clustering derives in part from exploiting domain structure.

Introduction

Artificial intelligence has long touted the importance of the domain in problem solving. A recurring theme is that regularities in the domain and knowledge about them helps reduce the complexity of search (Feigenbaum 1977). As the oft-quoted saying goes, "in the knowledge lies the power." But this power has never been characterized or quantified in detail. Moreover, the relationship between domain knowledge and domain structure remains somewhat unclear. In this paper, we begin to formalize the study of domain structure, by quantitatively studying domain structure and its effects on complexity.

This study is not only quantitative but also comparative. Comparative studies are useful in isolating key design issues in building intelligent systems. In particular, we compare two approaches to diagnostic problem solving: candidate generation and symptom clustering. In (Wu 1990), we showed that a symptom clustering approach offers increased efficiency for the diagnosis of multiple disorders. In this paper, we begin to explain why. It appears that some of the efficiency is gained by exploiting inherent domain structure.

This study is one of the first to separate the confounding concepts of domain knowledge and domain structure. Efficiency in search can be gained either by knowledge about domain structure or by the inherent domain structure itself. Most artificial intelligence work has studied the former topic, by explicitly encoding knowledge about structure in the form of abstraction hierarchies, component diagrams, and so on. But what happens if domain structure exists and knowledge about it is not explicitly encoded? Can such inherent domain structure still facilitate problem solving, even if the program does not "know" about it? In this paper, we untangle these issues by isolating the issue of domain structure from that of domain knowledge.

This paper proposes a hypothesis about the nature of diagnostic domains. This hypothesis states that:

> *Diagnostic domains are characterized by clusters of causes (disorders) that cause similar clusters of effects (symptoms)*

Because these sets of disorders and symptoms partition a domain into modules, we call this the *modularity hypothesis of domain structure.* To put it more loosely, the knowledge of a domain is not uniform but "lumpy", with modules of related disorders and symptoms. This is exemplified by medicine, which can be divided by organ systems into cardiovascular disorders and symptoms, gastrointestinal disorders and symptoms, and so on. Further modularity exists within each of these subspecialties as well. The modularity hypothesis can be tested experimentally by defining the explanatory power of a disorder. Then domain structure can be measured as the distribution of explanatory power. The experimental results obtained in this paper constitute evidence for the existence and importance of domain structure.

Background

In set-covering diagnosis, knowledge bases have a bipartite structure, with disorders and symptoms that

are linked whenever a disorder can cause a symptom. Each disorder has a set of possible effects, called its disorder profile. Each symptom has a set of possible causes, called its symptom profile (called a conflict set in the literature). These are the disorders in the knowledge base that have can cause that symptom. A diagnostic problem, or case, consists of a set of symptoms to be explained. In the general case, links in the knowledge base can have conditional probabilities and disorders can have prior probabilities, but in this study we restrict ourselves to the task of categorical diagnosis.

Candidate generation (de Kleer & Williams 1987, Reggia, Nau, & Wang 1983, Reiter 1987) is a method for converting a set of symptom profiles to a set of minimal candidates. A candidate is a set of disorders that constitutes a proposed solution to the diagnostic problem. A candidate is valid when it explains all of the given symptoms. A candidate is minimal when it is valid and no proper subset of the candidate is also valid. For instance, consider the simple knowledge base shown below:

Symptom	Possible causes
Fever	Flu, TB, Hep, Mal
Cough	Flu, TB, Asth, Bron

Here, the two symptom profiles have Flu and TB in common. Suppose a patient has both fever and cough. Candidate generation yields the following set of minimal candidates:

[Flu], [TB],
[Hep,Asth], [Hep,Bron], [Mal,Asth], [Mal,Bron]

Symptom clustering (Wu 1990, 1991) is another method for converting a set of symptom profiles to a set of minimal candidates. It works by constructing partitions of symptoms called symptom clusterings. Each symptom clustering consists of a set of clusters, and each cluster has a differential diagnosis that explains that cluster. For instance, the two symptom clusterings for the problem above are:

Clustering 1:	(Fever,Cough)	⟵	{Flu, TB}
Clustering 2:	(Fever)	⟵	{Hep, Mal}
	(Cough)	⟵	{Asth, Bron}

Each differential contains a disjunction of disorders, any one of which explains its cluster. The minimal candidates can be obtained by taking the cartesian product of the differentials. Thus, we obtain the following candidates:

{Flu, TB} = [Flu], [TB]

{Hep, Mal} × {Asth, Bron} =
[Hep,Asth], [Hep,Bron], [Mal,Asth], [Mal,Bron]

These are all of the minimal candidates for the two symptoms of fever and cough. In general, symptom clustering computes an approximation to the complete set of minimal candidates produced by candidate generation. The computation is approximate mainly because nonminimal candidates may be included in order to achieve a compact factored representation.

Thus, a symptom clustering represents an explanatory structure, postulating how a given problem is decomposed into subproblems. In the process, it factors sets of candidates into a compact form. Clusterings compactly represent a set of candidates because the candidates are stored as an implicit cartesian product without having to compute the cartesian product explicitly. This results in efficient search of the candidate space, as suggested by experiments in (Wu 1990). Further details on symptom clustering are provided in (Wu 1991).

Explanatory Power

The study of complexity requires large real-world knowledge bases. For this study, we use the QMR knowledge base (Miller, McNeil, et al. 1986), formerly the INTERNIST program (Miller, Pople, & Myers 1982). This knowledge base contains over 4000 symptoms and 600 diseases, covering over 80 percent of the topics in general internal medicine. In particular, we generated test cases by selecting prerenal azotemia as a target disorder. Prerenal azotemia has 14 possible causes in the knowledge base. We generated 10 cases, labeled A through J, by stochastically picking symptoms from this pool of 14. We used the frequency values in the QMR knowledge base as the probability of a symptom being included in the case. Stochastic selection therefore simulated the causation of symptoms in a natural case. For each of the 10 cases, we generated 10 runs by randomly permuting the symptoms in each case, for a total of 100 runs. Random ordering is necessary because the performance of diagnostic algorithms depends greatly on the order in which symptoms are presented. The resulting runs are the same used in (Wu 1990), and they are described in more detail there. Although each case contains one single-disorder candidate, namely [Prerenal Azotemia], the vast majority of minimal candidates contain multiple disorders, making these cases a test of multidisorder diagnosis.

Hence, the prerenal azotemia subdomain is the universe seen by a diagnostic program. This subdomain contains a set of observable symptoms, which are the 14 possible effects of the target disorder, prerenal azotemia. The subdomain also contains a set of complete and partial explanations for these symptoms. These disorders are competitors of the target disorder. The definition of a subdomain is shown in figure 1.

The relevant part of each disorder profile is that which intersects the observable symptoms. We call the size of this relevant disorder profile its explanatory power. It indicates the fraction of observable symptoms a disorder is able to explain. We define explana-

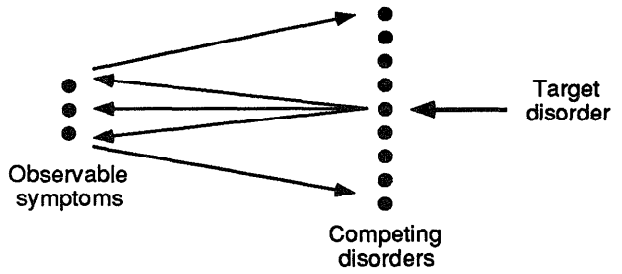

Figure 1: Definition of a subdomain. A subdomain is the subset of disorders and symptoms seen by a diagnostic program for a given problem. A target disorder causes a set of observable symptoms, which then implicate a set of competing disorders.

tory power to be

$$\frac{|\text{Effects}(d) \cap \text{Effects}(d_0)|}{|\text{Effects}(d_0)|}$$

where Effects(d) is the disorder profile for d, d_0 is the target disorder, and Effects(d_0) is the set of observable symptoms. We can examine explanatory power graphically. The prerenal azotemia subdomain is shown in figure 2(a). In this figure, disorders are arranged with the highest explanatory power in the center. The distribution of explanatory power can also be quantified, as shown in figure 3(a). Note that the distribution is non-Gaussian, with most disorders having a low explanatory power, but a significant few with a high explanatory power. This distribution is typical of the domain; for comparison, an average of other explanatory power distributions is shown in figure 3(b). This summary distribution shows that explanatory power is generally shifted to the left; however, the individual features of each distribution are lost by averaging.

The form of the distribution suggests that the subdomain has a high degree of structure. The distribution shows that most disorders have low explanatory power, while a significant number of disorders have high explanatory power. We call this phenomenon *explanatory structure*. It means that similarity among disorders is not random. Rather, a given problem triggers many disorders with low explanatory power and several disorders with high explanatory power. Explanatory structure occurs because the links in a knowledge base are not distributed randomly. Rather, there are clusters of symptoms and corresponding clusters of disorders. Links between clusters are relatively dense, while links between clusters are relatively sparse. Most disorders are therefore unrelated, resulting in low explanatory power. But each disorder belongs to a cluster of related disorders that have high explanatory power. In the rest of this paper, we explore explanatory structure experimentally to study its effect on complexity.

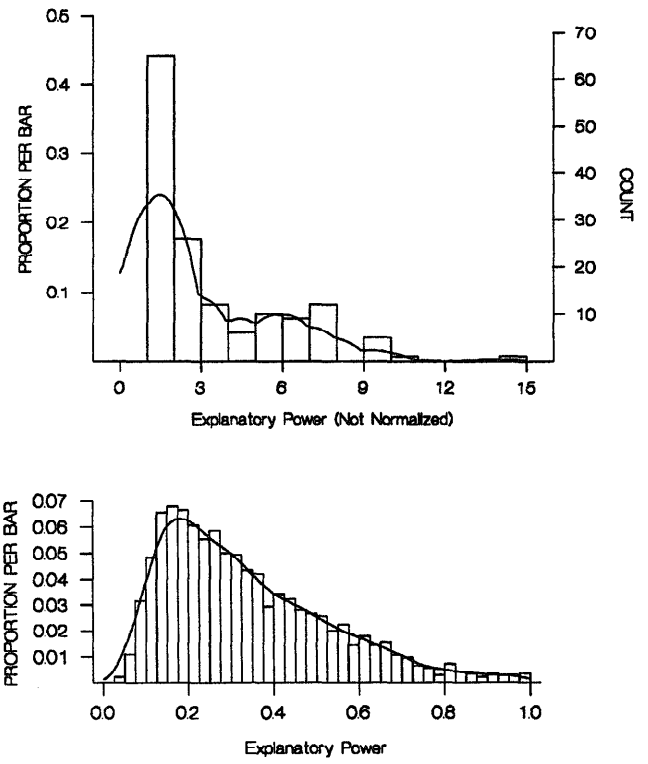

Figure 3: The distribution of explanatory power. The top graph shows the distribution for the prerenal azotemia subdomain. The bottom graph shows an average of distributions for several randomly selected target disorders in the QMR knowledge base. The continuous line provides kernel smoothing of the distribution.

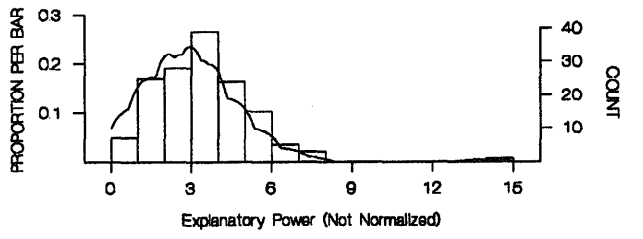

Figure 4: Distribution of explanatory power for a randomized knowledge base. Compare this figure with the original distribution in figure 3(a).

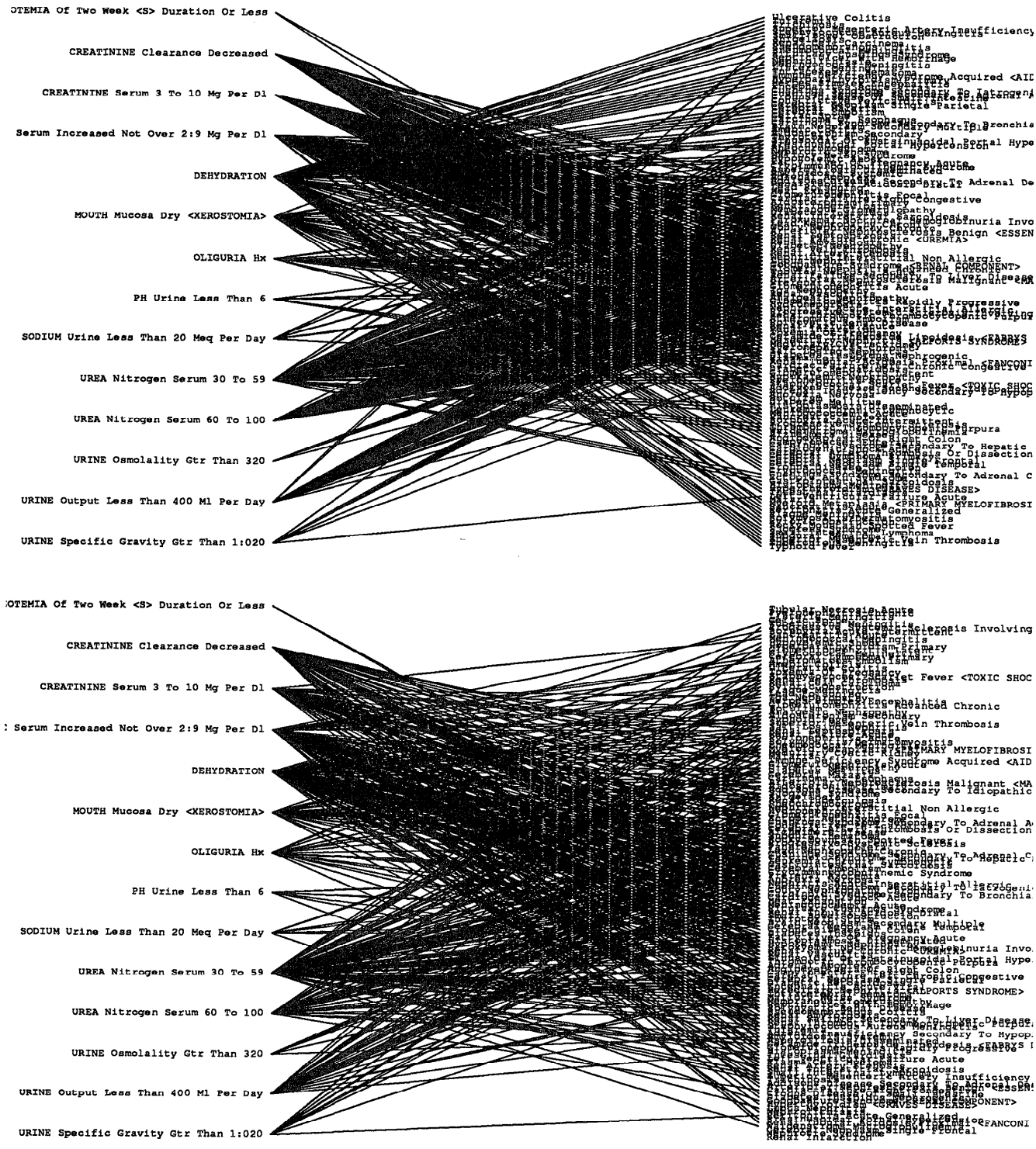

Figure 2: Structured (top) and randomized (bottom) subdomains, based on the target disorder of prerenal azotemia. Observable symptoms are shown on the left, competing disorders on the right. A link between a disorder and symptom indicates that the disorder is a possible cause for the symptom. The disorders in each graph are arranged so that those with the most links are located in the center.

Experiment

In this experiment, we experimentally remove structure in the knowledge base and observe the result. We remove structure by randomizing the distribution of disorders among symptom profiles. Each symptom profile keeps the same size as before, but the contents of each profile are redistributed among the original set of competing diseases. The only restriction is that each symptom profile must contain the target disorder. This is done to maintain the semantics of the subdomain, so that all symptoms are possible effects of the target disease.

If we apply this procedure to the original prerenal azotemia subdomain, we get the randomized knowledge base shown in figure 2(b), where disorders again are arranged with the highest explanatory power in the center. In this figure, the distribution of explanatory power shows some variation, which would be expected in any random process. Nevertheless, it can be seen that the links are spread out more evenly over the competing disorders than in the structured knowledge base. To show this effect more clearly, the distribution of explanatory power for the randomized knowledge base is plotted in figure 4. This shows that the distribution of explanatory power has been changed from a skewed distribution to a normal one.

The symptom clustering and candidate generation algorithms were implemented. Each algorithm was executed on the 100 stochastically generated and permuted runs for the prerenal azotemia subdomain. However, the candidate generation algorithm was only executed on 35 runs in four of the simplest cases, since more difficult runs would not terminate in a reasonable amount of time. The total number of nodes kept during search for each run was recorded.

The results are shown in figure 5. Figure 5(a) shows a comparison between symptom clustering and candidate generation on the structured knowledge base. Figures 5(b) and 5(c) compare space complexities between the structured (QMR) and randomized knowledge bases for each algorithm. Finally, figure 5(d) compares the two algorithms on the randomized knowledge base.

Discussion

Figure 5(a) indicates that symptom clustering is more efficient than candidate generation when domain structure is present. The slope of 1/5 suggests that the space complexity of symptom clustering is reduced by a power of 5 compared that for candidate generation; that is, the complexity of symptom clustering is the fifth-root of that of candidate generation. Thus, at least for this subdomain, symptom clustering represents the search space more compactly. On a log-log scale, a straight line represents only a polynomial reduction in complexity. Still, the savings are substantial, allowing the symptom clustering algorithm to solve real-world diagnostic problems with reasonable complexity.

Figure 5(b) compares the computational behavior of candidate generation on the original, structured knowledge base versus the randomized knowledge base. The figure shows that domain structure leads to a much lower space complexity, but with little consistent pattern. This suggests that candidate generation does not exploit the inherent domain structure systematically. Nevertheless, candidate generation does appear to benefit from inherent domain structure substantially. The slope of 1/9 suggests that domain structure reduces space complexity of minimal candidates by a power of 9.

On the other hand, figure 5(c) measures the effect of domain structure on symptom clustering. Again, the space complexity is decreased, but there appears to be a better correlation in complexity between the structured and randomized domains. The correlation suggests that the evidential structures represented by symptom clustering correspond to the structure of the domain. Thus, its computational behavior is more responsive to structure present between disorder profiles. This might be expected, since disorders with similar profiles are usually placed in the same differential diagnosis, while disorders with different profiles are usually placed in different differentials. Domain structure thereby gives a sharper contrast between plausible and implausible clusterings. The zero slope indicates that symptom clustering uses domain structure very effectively. This slope, which is smaller than that of figure 5(b), suggests that symptom clustering exploits domain structure more effectively than candidate generation.

Finally, figure 5(d) compares the two algorithms on the randomized knowledge base. Surprisingly, even in the absence of domain structure, symptom clustering performs better than candidate generation. The slope of 1/4 means that clustering is more efficient even in randomized domains. Hence symptom clustering may be simply represent candidates more efficiently, regardless of the domain.

Admittedly, this experiment does not control all aspects of domain structure. But by isolating one attribute, explanatory power, and studying one subdomain in depth, we obtain a more controlled and informative experiment. Still, other attributes potentially affect diagnostic complexity. For instance, the medical domain is characterized by large symptom profiles, as evidenced by the prerenal azotemia subdomain, where symptoms had between 2 and 76 possible causes each. Other domains might not trigger so many competing disorders. In addition, while we changed the distribution of explanatory power, we did not change its average value. Explanatory power, even after randomization, averaged only 4 out of the 14 observable symptoms. This is significant, since it is combinations of these partial explanations that cause much of the complexity of candidate generation. Symptom clustering gains efficiency by representing such combinations

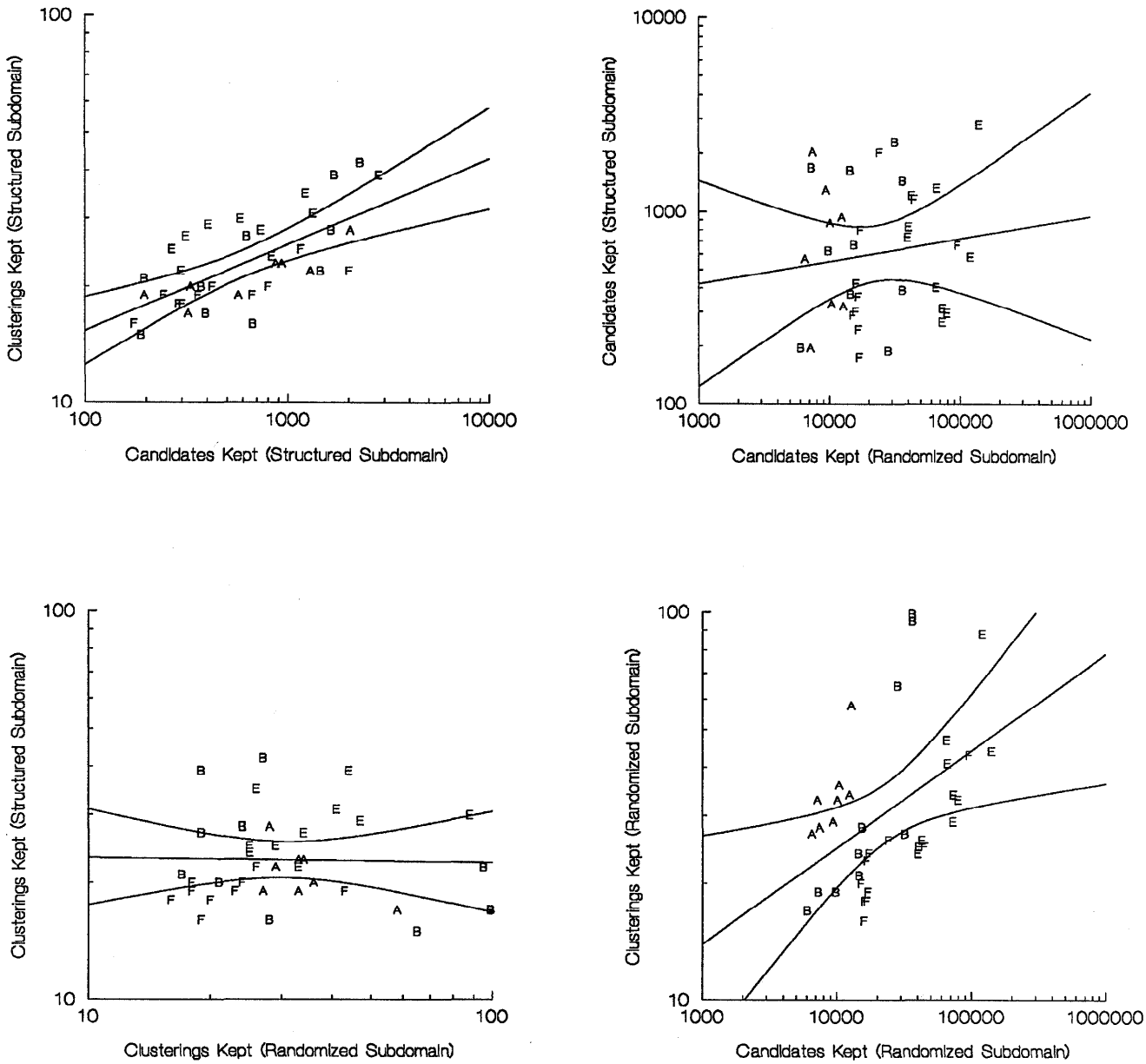

Figure 5: Complexity results: (a, top left) Comparison of symptom clustering and candidate generation on a structured domain. (b, top right) Comparison of structured and randomized domains for candidate generation. (c, bottom left) Comparison of structured and randomized domains for symptom clustering. (d, bottom right) Comparison of symptom clustering and candidate generation on a randomized domain. Graphs show linear fits of data, along with their 95th percentile confidence bands. Each letter represents a different set of symptoms in a random permutation.

compactly in a factored form, while candidate generation must represent each combination explicitly.

Finally, it is interesting that domain structure appears in such a specific example. We would expect large-scale structure at the level of organ systems. For instance, cardiac diseases cause a characteristic cluster of heart symptoms, while gastrointestinal diseases cause a different cluster of digestive symptoms. But despite the fact that prerenal azotemia subdomain lies within a single organ system, it apparently still has structure. Even at this level, disorders and symptoms exhibit domain structure. We can see this structure in the symptoms for prerenal azotemia, where some symptoms deal with urinary chemistry while others deal with manifestations of dehydration. We conjecture that domain structure exists at different levels because the underlying object or device being diagnosed has structure at different levels. In any case, the existence of structure at different levels means that symptom clustering may offer gains in efficiency for a broad range of diagnostic situations.

Conclusions

Artificial intelligence has developed a distinctive set of concepts, theories, and principles. One of the most fundamental of these principles has been the use of domain structure to simplify problem solving. This paper provides a first step towards characterizing and quantifying domain structure, by identifying an attribute of domain structure called explanatory power. By altering the distribution of explanatory power, we can remove structure from a domain. Randomized domains give us an experimental tool to study the effect of domain structure on the complexity of problem solving. The experimental results here suggest that inherent structure of a domain facilitates problem solving, even when knowledge about such structure is not encoded explicitly. Such facilitation occurs for both the candidate generation and symptom clustering algorithms. Moreover, the results suggest that inherent structure may benefit the symptom clustering algorithm more than the candidate generation algorithm. This may help to explain why symptom clustering provides an efficient means of diagnosing multiple disorders and why structure is such a critical tool for dealing with complexity (Simon 1973).

Acknowledgements

This research was supported by National Institutes of Health grant R01 LM04493 from the National Library of Medicine and by National Research Service Award T32 GM07753. I am grateful to Ramesh Patil, Peter Szolovits, and Randy Davis for their helpful discussions and to Randolph Miller for allowing use of the QMR knowledge base for testing purposes.

References

de Kleer, J. and Williams, B. C. Diagnosing multiple faults. *Artificial Intelligence*, 32:97–130, 1987.

Feigenbaum, E. A. The art of artificial intelligence: 1. Themes and case studies in knowledge engineering. In *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, pages 1014, 1029, 1977.

Miller, R. A., McNeil, M. A., et al. The Internist-1/ Quick Medical Reference project—status report. *Western Journal of Medicine*, 145:816–822, 1986.

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

Reggia, J. A., Nau, D. S, and Wang, P. Y. Diagnostic expert systems based on a set covering model. *Intl. Journal of Man-Machine Studies*, 19:437–460, 1983.

Reiter, R. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–96, 1987.

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

Wu, T. D. Efficient diagnosis of multiple disorders based on a symptom clustering approach. In *Proceedings, Eighth National Conference on Artificial Intelligence*, pages 357–364, 1990.

Wu, T. D. Efficient diagnosis of multiple disorders: A symptom clustering approach. Doctoral dissertation, Massachusetts Institute of Technology, 1991 (in preparation).