A BAYESIAN NETWORK FOR MEDICAL DIAGNOSIS

By

HAOBO LIU

Bachelor of Medical Science

Hunan Medical University

Changsha, Hunan

People Republic of China

1993

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 1998

A BAYESIAN NETWORK FOR

MEDICAL DIAGNOSIS

Thesis Approved:

Blayne E. Mayfield

Thesis Adviser

J. Chandler

Jacques E. LaFrance

Wayne B. Powell

Dean of the Graduate College

ACKNOWLEDGMENTS

I have been blessed with a grand opportunity to further my education in America. Many people have made the pursuit of my Master's Degree a positive experience. First and foremost, I would like to express my sincere appreciation to my advisor, Dr. Nick Street, for his intelligent guidance, invaluable expertise, and friendship during this study, and for his patience in correcting this thesis. Without him, I am sure I would not have finished my degree in a timely manner. I also wish to express the gratitude to my other two committee members: Dr. John Chandler and Dr. Jacques LaFrance for their kind suggestions and assistance during this study, which made the process of earning my degree enjoyable as well as educational. I feel very fortunate to have had such intelligent and kind individuals on my committee. Furthermore, I would like to thank the dean of Computer Science department, Dr. Blayne Mayfield, agreed to be my advisor alternative in my final defense. And I will surely cherish the valuable instruction provided by other faculty and staff during these past two years forever.

I would also like to express my special thanks to my family and friends, without whose support and encouragement I would not have been able to complete the requirements for my degree in the manner that I have. In particular, I am deeply indebted to my dear wife, Haihui Huang whose understanding, unconditional love and support have upheld me during this study and throughout my life. I will never be able to thank her enough.

TABLE OF CONTENTS

Chapter Page

I. INTRODUCTION 1

Probability in Expert System 2
Bayesian Network 3
Applications 4
Suitability 5
Intent of Study 6

II. LITERATURE REVIEW 9

Expert System 9
Expert System in Medicine 11
Functions 12
Methodologies 13
History 15

III. PROBABILISTIC REASONING 19

Probability 19
Kolmogorov's Axioms 20
Prior Probability 20
Conditional Probability 21
Bayes' Rule 21
Normalization 22
Properties of Bayesian Networks 24
Joint Probability Distribution 24
Conditional Probability Table 25
Conditional Independence Relations 26

IV. MESSAGE-PASSING ALGORITHM 29

Computation with Single Parent 30
Computation with Multiparents 34
Propagation 37

V. DIAGNOSTIC AGENTS DESIGN AND CONSIDERATION 40

Steps of Building an Expert Agent 40
Knowledge Collection 41
Building Network 43

VI. TESTING RESULT 47

List of Results 47
Analysis 49

VII. DISCUSSION 51

Summary 51
Future Work 52

REFERENCES 54

LIST OF FIGURES

Figure Page

1. Bayesian network....25
2. X and Y are blocked by E and z in E (Tail-to-Tail)....27
3. X and Y are blocked by E and z in E (Head-to-Tail)....27
4. X and Y are blocked by E and Z not in E (Head-to-Head)....28
5. Fragment of a singly-connected network showing the partition of parents and children to the computing variable X....31
6. The message passing in local network during updating of variable X (only single parent)....35
7. A part of a belief network....36
8. Belief propagation by message passing process....38
9. The Bayesian network of ectopic pregnancy and acute salpingitis....44
10. The Bayesian network of ectopic pregnancy....45
11. The Bayesian network of acute salpingitis....46

LIST OF TABLES

Table Page

1. Conditional probability table for X_4 25

2. Conditional probability table for X_6 26

3. Relations among F, C and ectopic pregnancy 42

4. Probability of symptoms, signs and lab tests for ectopic pregnancy and acute salpingitis 42

5. Information of ten made-up patients 48

6. The comparing of diagnosis by doctor and agent 48

CHAPTER I

INTRODUCTION

Artificial intelligence (AI) is a branch of computer science that deals with symbolic representation of knowledge and its use in problem solving. It has been regarded as a revolution in software by some experts. As an important subfield of AI, expert systems are knowledge-based systems that symbolically encode concepts derived from experts in a certain field. The resulting system performs problem analysis based on that knowledge and provides advise and solutions. What's more, an expert system can extend the analysis and decision-making ability of an expert to general users and even provide valuable advise for the experts themselves. Obviously, the expert system has a better "memory" than the experts. It can collect and keep all the knowledge in its "head". Therefore, it is not a surprise when the expert system makes a better judgement than human experts.

Expert systems often implement a rule-based approach that uses boolean logic to process input from the user, using its knowledge base to generate a prediction or suggestion. But the problem with this approach is that it can not deal effectively with uncertainty. In the real world, not everything is crystal clear; not everything can be answered with "yes" or "no." Actually, in many situations, people use probability to express their decisions or judgements.

Probability in Expert System

Probability is a powerful tool to express uncertainty in an expert system. The problem of dealing with uncertainty is crucial in the entire expert system field, because most everyday reasoning and decision making is based on uncertain premises. Most of our actions in daily life are based on guesses, often requiring explicit weighting of conflicting evidence. Probability theory was created by Thomas Bayes[4], who was a British minister, in 18th-century. It was seldom used in the computer industry before the 1980s. Previous work in knowledge representation had focused on symbolic logic programs. The modularity of if-then rules contributed to the success of the expert systems during that period. From the 1980s on, things changed as scholars realized the importance of dealing with uncertainty.

The certainty factor (CF) has been created as a basis for the system MYCIN to deal with uncertainty[34]. It is a relatively informal mechanism for quantifying the degree to which, based on the presence of a given set of evidence. Certainty factors have been most widely applied to domains that use incrementally acquired evidence.

The Dempster-Shafer theory[33] was designed to deal with the difference between uncertainty and ignorance. Instead of computing the probability of a proposition, it computes the probability that the evidence supports the proposition. But most current scholars regard it as merely an alternative way to use probability.

Fuzzy logic[44] does not use a precise probability figure to represent uncertainty. Instead, it uses ranges of values to represent input variables in a fuzzy system, coupled with rules that produce output ranges based upon the input values. Fuzzy set theory is very controversial. On the one hand, it is an extremely popular area of research. It is successfully used in all sorts of consumer and industrial products. The Japanese have used it to control passenger trains, digital cameras, washing machines, air conditioners, antilock brakes etc. On the other hand, there is a lot of opposition to fuzzy set theory in the AI community because it is unable to describe uncertainty. Some scholars even question the fundamental rules of fuzzy set theory.

Many expert systems that used probability paid little attention to the theory of probability. They just used probabilities to express the strength of the evidence. But they ignored the rules of probability to evaluate and organize the information. When some important pieces of evidence are changed, expert systems such as MYCIN[34] are not sensitive enough to correctly modify their conclusions. That's why they are unreliable in practical usage. Most expert systems were used just for research purposes. However things changed after the establishment of probabilistic structure. Probability has flourished both on the practical side and on the scholarly side. Many articles on this structure have been published in the recent years. This structure is called the Bayesian network.

Bayesian Network

A Bayesian network (or belief network) is used to model uncertainty in a domain. Both quantitative and qualitative techniques are used. The basic idea of the Bayesian network is that the problem domain is modeled as a set of nodes interconnected by directed lines and arcs. Each node in the network represents a particular occurrence or condition, called a variable. The lines indicate the causal effect of the variables on each other. Actually a Bayesian network is an acyclic directed graph. The lines represents causal relationships between nodes, and the natural flow of the graph does not enable conditions to cycle back to prior conditions. This prevents the algorithm, which we will introduce, from getting into infinite loops or becoming deadlock.

In addition to storing the relationships among nodes, a Bayesian network contains the probabilities associated with each relationship and the distinct possible output states of each node in the network. These states are mutually exclusive and comprehensive. A Bayesian network represents the entire joint probability distribution (see chapter 3) over the domain variables. After the Bayesian network is established, it must be seeded with an initial estimation for all of the probabilities involved. Prior probability is also called unconditional probability. It means the probability of an event occurring in the condition of having no other information. For the root nodes, they mean the random probability of a certain state occurring. When new evidence is given, the network automatically updates the probabilities for the parent and child nodes. The recalculation continues to propagate across the network, fine-tuning the accuracy of all probabilities. The ability to revise the probability of an event by considering the new states of events, that are caused by the first event is the main advantage of the bayesian formalism.

Applications

Bayesian networks provide a practical use of AI. A Bayesian network is generally applied to problems when there is uncertainty in the data or in the knowledge about the domain. It has been applied particularly to problems which require diagnosis of problems from a variety of input data. Some of the most well-known examples of Bayesian networks are medical diagnostic tools such as PATHFINDER, BNG[13]. Recently more products have been created in other areas. Following are some general areas in which Bayesian networks are used:

- Medical diagnostic systems
- Analysis in the natural, biological and social sciences
- Real-time weapons scheduling
- Intel processor fault diagnosis (Intel)
- Generator monitoring expert system (General Electric)
- Troubleshooting (Microsoft)

Both scholars and businessman are fascinated with Bayesian networks. That means more research and more money will be directed into this area. So we can predict more applications be found in the future.

Suitability

Bayesian network technology is similar to two of the most creative computational technologies available today: fuzzy logic and neural networks. Although the fundamental concepts behind each of the three areas are quite different, they are all used to make systems more intelligent and practical. A fuzzy system can learn from experiences. It can develop its fuzzy rules based on its own experiences. A neural network is a system that emulates the cognitive abilities of the brain by establishing recognition of particular inputs and producing the appropriate output. A neural network is trained using presented inputs to establish their own internal weights and relationships guided by feedback. The Bayesian networks resemble neural networks in their variable dependency. The probabilities in a Bayesian network are roughly analogous to the internal weights in a neural network. Neural networks, however, are free to form their own internal workings and adapt on their own. Compared to neural networks, Bayesian networks have the following advantages:

- The expert can provide knowledge in the form of causal structure.
- The network is understandable and extensible.
- They can be used easily with missing data.

Bayesian networks can be used whenever classical knowledge-based systems might be used. Compared with the classical knowledge-based systems, a Bayesian network has the following advantages:

1. a more modular representation of uncertain knowledge, which makes them easier to maintain and to adapt to different contexts,
2. a more intuitive knowledge representation (polytree diagrams) for domain experts, and
3. making it easier for them to be involved in maintaining a system.

Intent of Study

Computer programs to assist with medical decision making have long been anticipated by physicians with both curiosity and concern. Scientists have worked on this field for almost forty years. They have succeed in some areas. But the progress in patient-specific consultation systems has been slow. A wide variety of techniques have been used in the experimental design and implementation of such systems, which include simple logic, mathematical modeling, pattern recognition, and the analysis of large data bases[35]. Also some systems have been deployed, such as ATTENDING[18], ONCOCIN[12, 36,37], and MYCIN[34]. But the results were not satisfactory. All of them were used only by the research unit that created them. In 1959, some researchers recognized the relevance of Bayes's theorem[4] to the task of diagnosis. Because computers were traditionally viewed as numerical calculating machines, it was clear that they could be used to compute the pertinent probabilities based on observations of patient-specific values. Many Bayesian diagnosis programs have been developed by using the Bayesian network. Some of them have been shown to be accurate in selection among competing explanations of a patient's disease state[26]. In England, De Dombal and associates[8] made a Bayesian system for the diagnosis of acute abdominal pain. It works well and is used extensively in British emergency departments.

The goal of this research is to build a diagnostic agent with a Bayesian network. This agent can help doctors in emergency departments to make a differential diagnosis between ectopic pregnancy and acute salpingitis. Acute salpingitis is one of most common diseases in obstetric and gynecologic areas. Ectopic pregnancy is a very dangerous disease though it is not as common as acute salpingitis. They have a lot of similar symptoms which could easily confuse the physician. Acute salpingitis is the most common misdiagnosis in cases of ectopic pregnancy. Its misdiagnosis could even cause the death of a patient. In my thesis, I will do the following work:

1. Compare these two diseases and collect the medical data for building the diagnostic agent.
2. Construct the Baysian network with those data .
3. Implement the message-passing algorithm introduced by Judea Pearl in 1986[22] to estimate the probability of the two diseases.
4. Evaluate the result and discuss the work further.

The reasons why I adopt this algorithm are: First, it is sensitive and accurate. When new evidence appears, the probability is modified accordingly. This feature fits the situations found in medical diagnosis. Secondly, this algorithm is efficient. Third, the algorithm is understandable. Fourth, it is one of the most popular algorithms used in the real medical world.

CHAPTER II

LITERATURE REVIEW

Expert System

The study of expert systems is an important subfield of artificial intelligence (AI)[29]. It is a knowledge-based system and is regarded as an "intelligent" system. What is generally considered to be "intelligence" can be divided into a collection of observations or facts and a means for utilizing these facts to reach goals. For example, a goal might be to determine why a car will not start. The expert system prunes these facts to eliminate from consideration any facts and rules that won't lead the user to a specified goal. The portion of intelligence that generates new facts from existing ones and arrives at the goal is the "inference mechanism."

Expert systems can be applied to problems that are solved primarily using formal reasoning. The problem is solved through a dialog, or "consultation," with the expert system. In a simple expert system, each question is answered with "yes" or "no". After each question, either the program may request an answer to another question or it makes an inference based on the facts it already has accumulated.

Knowledge engineers are used to develop expert systems. They are skilled at observing and analyzing the methods used by human experts to solve problems in a particular discipline. These methods, or heuristics, are stored as part of the data.

There are three basic components of an expert system. The first component, the rule-base, is a static database that contains all the knowledge about the domain. The second component, the working memory, houses the dynamic database to store the new facts obtained from the user or inferred from known facts. The inference engine is the third component which contains the general problem-solving logic.

One of the most common types of expert systems is the ruled-based system. In a rule-based system, knowledge is represented as IF-THEN statements (rules). When the IF portion of a rule is true in the current situation, the action specified by the THEN portion is executed or said to fire.

The working memory contains facts that describe what is known about a particular problem. When a program is started, the working memory is empty. As the consultation progresses and the system learns more about the problem, the new knowledge is put into working memory. The knowledge in working memory is used to fire additional rules. As each rule fires, the conclusion is added to working memory with the facts already known.

The inference engine has two tasks: one is inference, and the second is control. The inference component uses the facts in working memory to try to create new rules. After all conditions of a rule are triggered, the rule fires and the conclusions are added to working memory. The control component determines the order in which the rules are scanned.

Expert Systems in Medicine

Expert systems in medicine are the computer programs used to support clinical decision making. They are also called medical decision-support systems. They deal with medical data about patients and the relative medical knowledge that is necessary to interpret such data. Generally, these systems are divided into three types:

1. Systems for information management. These systems provide environments for storing and retrieving clinical data and knowledge. For example, Hospital Information Systems provide access to patients' data needed for clinical decision making. Bibliographic Retrieval Systems allow rapid access to pertinent information from current literature. These systems are similar to other commercial information systems.
2. Systems for focusing attention. Examples are Clinical Laboratory Systems that flag abnormal values and possible explanations for those abnormalities and Pharmacy Systems that provide information about effects and side-effects of drugs and possible interactions[39] among the drugs. Such programs are designed to remind the physician of diagnosis or problems that might otherwise have been overlooked. They typically use simple logic, displaying fixed lists or paragraphs as a standard response to a definite or potential abnormality. These systems need professional medical knowledge for correct performance. They help ensure that physicians don't ignore the potential damage caused by routine treatments.
3. Systems for patient-specific consultation. Such systems provide diagnosis and advice based on sets of patient-specific data. They may follow simple logic, such as

algorithms, and may be based on statistical theory and cost-benefit analysis. Some of the diagnostic assistants suggest differential diagnoses. With additional evidence, they can narrow the range of etiologic possibilities.

The boundaries among these three categories are not sharp, but the distinctions can help us to understand the different functions of the systems.

Functions

The goals of developing expert systems for medicine are as follows[35]:

1. To improve the accuracy of clinical diagnosis through approaches that are systematic, complete, and able to integrate data from diverse sources.
2. To improve the reliability of clinical decisions by avoiding unwarranted influences of similar but not identical cases.
3. To improve the cost efficiency of tests and therapies by balancing the expenses of time and inconvenience against the benefits and risks of definitive actions.
4. To improve our understanding of the structure of medical knowledge, with the associated development of techniques for identifying inconsistencies and inadequacies in that knowledge.
5. To improve our understanding of clinical decision-making, in order to improve medical teaching and to make the system more effective and easier to understand.

The third type of medical decision-support system mentioned above generally falls into two categories: those that assist physicians with determining what is true about a patient (usually the correct diagnosis) and what to do for the patient (such as, what test to order, whether to treat, what therapy plan to institute). Many systems assist clinicians with both activities (for example, diagnostic programs often help physicians to decide what additional information would be most useful in narrowing the differential diagnosis for a given case), but the distinction is important, because advice about what to do for a patient cannot be formulated without balancing the costs and benefits of possible actions. Assessments of what is true about a patient, based on a fixed set of data that are already available, can theoretically be made without consideration of cost and risk. That means we need to pay more attention to the first question before we make the decision. Thus, a "pure" diagnostic program leaves to the user the task of determining what data to gather, or it requires a fixed set of data for all patients. From this point, it is easier for engineers to build an expert system to answer the second question rather than the first one. However, it is unrealistic to view diagnosis as a process separable from considerations of the available options for data collection and therapy.

Methodologies

Since the beginning of expert systems technology, knowledge acquisition and representation have been considered the major constraint in the development of expert systems in the medical field. Knowledge acquisition and representation is a kind of knowledge model that can be used to predict or explain behavior in the world. Thus, diagnosis is based on a causal explanation of what is happening to the patient, and therapy is based on predictions about how the disease process can be modified. The knowledge models people have used in medicine are the following[13]:

1. Bayesian Networks.

   As introduced in Chapter One, a Bayesian network is a mechanism[2] to calculate the probability of a disease, in light of specified evidence, from the a priori probability of the disease and the conditional probabilities relating the observations to the diseases in which they may occur. We will discuss it in detail later.

2. Rule-Based Reasoning

   Rule-based reasoning is the most general structure. It uses knowledge encoded in generation rules (IF…THEN). Rules usually have a conditional part and an action part. Each rule represents one of the knowledge units related to an expert field.. Many related rules may correspond to an inference chain, which deduces a useful conclusion from several known facts. Rule-based reasoning has been the most popular choice of knowledge engineers for building expert systems in medicine.

3. Neural Network

   A neural network is essentially a type of information processing technology inspired by studies of the designs in the brain and nervous system. These systems are made up of many simple, highly interconnected processing elements that dynamically interact with each other to "learn" or "respond to" information rather than carrying out algorithmic steps or programmed instructions.

4. Case-Based Reasoning

   Medical doctors solve new problems by analogy with old cases and explain reasons in terms of prior experience. Computer systems that solve by analogy with old ones are called case-based reasoning (CBR) systems[27]. CBR systems solve problems by searching a collection of stored cases to find and retrieve the cases that most closely resemble a newly presented case using some similarity criteria.

5. Object-Oriented Programming

Object-oriented programming refers to all data structures as objects. Each object contains two basic types of information: information that describes the object and information that specifies what the object can do. It provides a natural way of representing real-world objects.

History

Since the earliest days of computers, health professionals have anticipated the day when machines would assist in the diagnostic process. The first articles dealing with this possibility appeared in the late 1950s (by Ledley and Lusted[16]) and experimental prototypes were shown to be accurate within a few years thereafter[41]. Several problems prevented the clinical introduction of such systems, however, ranging from the limitations of the scientific underpinnings to the logistical difficulties developers encountered when encouraging clinicians to use and accept the systems. But diagnostic systems received enhanced opportunities for progress from several sources, including the rapid development of the technological base (the hardware, software and the methods for interacting with them), the rapid growth of awareness of and interest in computers and information-management systems, and the growth of medical information systems for helping professionals with other biomedical research. A wide variety of techniques have been used in the design and implementation of decision-support systems. The simplest logic has been problem-specific algorithms designed by clinicians and then encoded for use by a computer. Although such algorithms have been useful for triage purposes and as a didactic technique used in journals and books where an overview for a problem's management has been appropriate, they have been largely rejected by physicians as too simplistic for routine use[11]. In addition, the advantage of their implementation on computers are not clear.

In the 1960s, medical expert systems with the implementation of programs that performed well-known statistical analysis appeared. These programs focused on the diagnosis part of the consultation. Some of the programs also used simple logic and mathematical modeling. They took as input a set of findings and selected the appropriate disease from a fixed set, using methods such as pattern recognition through discriminant functions, Bayesian decision theory, and decision-tree techniques.

Since the early 1970s, a growing body of researchers have been applying the techniques of AI to the development of diagnostic and therapy management consultation programs[15,21,36]. The AI field is closely tied to psychology and to the modeling of logical processes by the computer. Psychological studies of problem solving by medical experts have accordingly been influential in medical AI research. Medical expert systems became a hot topic in AI research and several applications of expert systems were developed. Internist-1 for example, was a large system designed to assist with diagnosis in general internal medicine[19]. MYCIN was a program designed to assist with therapy selection for patients with bacteremia or meningitis[38]. It explored the power of inferential rules as a mechanism for storing knowledge in a computer and was among the first system to emphasize the importance of explanatory capabilities in medical decision-support tools[43]. CASNET demonstrated the utility of detailed models of causal or pathophysiological relationships as the basis for pursuing diagnoses or proposing management strategies[24]. It was designed to assist physicians with the management of patients with glaucoma. But none of these systems was in routine clinical use because of limitation of knowledge representation techniques and physician resistance.

In the 1980s, medical expert systems developed very fast. They had a great impact on many areas of the medical field where knowledge provides the power for solving important medical problems. According to a survey conducted in 1992[13], the total number of expert systems uncovered in all fields was approximately 2500 in the 1980s. It shows that this field was very attractive to expert system developers in the 1980s. Comparing the growth rate of the expert systems between 1970s and 1980s, the trend is encouraging. In the 1970s, researchers were focusing on developing intelligent programming techniques. Only a handful of systems were built. During the 1980s, the number of developed expert systems increased from 50 in 1985 to 2200 in 1988[13]. The impressive growth rate of expert systems is an indicator of the acceptance of the technology by industry. These surveys also showed that expert systems were merging with the mainstream of information processing that was previously dominated by conventional data processors.

Again, we can attribute the large growth rate in developed systems in part to the new hardware and software technologies. In the 1970s, most expert systems were developed on powerful and expensive workstations, using languages like LISP, PROLOG, and OPS. Only the few people who could afford the platforms, and had the patience to learn the complexities of the available languages, had the chance to develop an expert system. During the 1980s, PCs became prominent in the computer world. Engineers developed easy-to-use expert system software development tools called "shells". A shell is a programming environment that contains all of the necessary utilities for both developing and running an expert system. These well-known shells like ProMD, HUGIN, NEXPERT, KAPPA, and ClassicaD3 were created for use on PCs. Therefore, the opportunity to develop an expert system was placed in the hands of many individuals.

In the 1990s, the complexity and volume of medical knowledge have increased continuously. A total of 233 medical expert systems were found between 1992 and 1996[13]. Researchers are trying to develop these expert systems with medical knowledge at all levels of medical care in order to achieve high-quality medical care and to reduce costs. Some scholars said the basic methodological problems like knowledge representation and inference mechanisms were no longer holding the spotlight and problems of introducing the systems in the clinical environment and questions of application-oriented research were receiving the attention.

The applications for expert systems in medicine appear to be increasing at an almost exponential rate. However, among the expert systems that have been implemented, there are questions concerning the actual success of at least some of these implementations. Most of the systems published in papers have not been successful in practice, especially in the clinical environment. Why? The answer is complex. We do have lots of logistic and scientific challenges that lie ahead. But people's enthusiasm for making diagnostic machines has never cooled down. More money and more people are focusing on it. This makes the future bright for medical expert systems.

CHAPTER III

PROBABILISTIC REASONING

Probability

The concept of probability has been debated for hundreds of years. The scholars are divided into two main camps — the subjectivists and the frequentists. The subjectivists think the probability of a certain event is the degree to which someone believes it, as indicated by their willingness to bet or take other actions. Meanwhile, the frequentists contend that the probability of certain event is the frequency with which it occurs. From the history, both of them were partially right. The result seems to be a kind of combination of the two definitions. Most Bayesian statisticians compromise on the meaning of probability. They agree that their goal is to calculate objective probabilities from frequency data, but they advocate the use of subjective prior probability to improve the calculations. Bayesian and frequentist statisticians tend to agree on the objective. The Bayesian prefers to assess prior subjective probabilities for the different possible statistical models and uses the data to update these prior probabilities to posterior probabilities, while the frequentists prefer to rely on the data alone to estimate the model.

Kolmogorov's Axioms

In order to express the probability clearly, we will use some logic symbols in the formulas. A. N. Kolmogorov[29] introduced the following three axioms:

Rule 1 : (non-negative rule). Any probability P (A) is a number between 0 and 1.

$0 \le P(A) \le 1$

Rule 2 : The outcome of an event which is true has probability 1, the outcome of an event which is false has probability is 0.

$P(\text{True}) = 1 \qquad P(\text{False}) = 0$

Rule 3: (additive rule) The union of two probability events is given by

$P(A \vee B) = P(A) + P(B) - P(A \wedge B).$

Note: If A and B are mutually exclusive (disjoint sets), then

$P(A \vee B) = P(A) + P(B).$

Prior Probability

Prior probability is also called unconditional probability. It is denoted as P (A). It means the probability of an event occurring in the condition of having no other information. Usually it is an assigned value or a value from statistics. In a belief network, it is the initialized data stored in each root node before the estimation. A is called a random variable. It can be multi-valued. For example, if there are five balls with different colors (red, blue, black, green, white) in a box, and we regard the color as our concerned variable, the probabilities of getting each color are:

P (Color = red) = 0.20

P (Color = green) = 0.20

P (Color = black) = 0.20

P (Color = blue) = 0.20

P (Color = white) = 0.20.

In this example, the variable A has a domain of five values.

Conditional Probability

When we know some evidence before we estimate the probability of certain events, prior probability is no longer appropriate. For instance, consider the question "Will Lee go to school? It is said he is ill". We assume A = "Lee won't (or will) go to school", B = "Lee is ill". If we want to estimate A, we use the following notation: $P(A \mid B)$. It means the probability of A given condition B.

Product rule: $P(A \wedge B) = P(A \mid B)P(B)$ where $P(B) > 0$; or

$P(A \wedge B) = P(B \mid A)P(A)$ where $P(A) > 0$;

Note: Most of the time, we use $P(A, B)$ instead of $P(A \wedge B)$ for convenience.

Independence rule: $P(A \mid B) = P(A)$ if and only if $P(A \wedge B) = P(A)\,P(B)$.

Bayes' Rule

If we rearrange the two forms of the product rule, we can get the following Bayes' rule[4]:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{1}$$

This rule is very useful. In many cases, we know three members and need to calculate the fourth member. The diagnostic processes in medicine are often such a situation. For example, the doctor may know that patients having hepatitis have a 70% chance to be jaundiced. The prior probability of hepatitis in the society is 1/5,000, and the frequency of people to be jaundiced is 1/10,000. According to Bayes' rule, we get:

$P(H) = 1/50000$

$P(J) = 1/10000$

$P(J \mid H) = 0.7$

$P(H \mid J) = P(J \mid H)\, P(H)/P(J) = (0.7 * 1/50000) / 0.0001 = 14\%$

In our daily life, we often have conditional probabilities on causal relationships and want to derive a diagnosis. It is also easier for people to estimate causal probabilities than diagnostic ones.

Normalization

This is a powerful technique we can use in the calculation of probability. It makes the estimation much easier. From Bayes' rule, we have (assume that A has n possible values, and A_j is any one of them):

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{P(B)}$$

We can change it to:

$$P(A_1 \mid B) + P(A_2 \mid B) + \ldots + P(A_n \mid B) =$$

$$\frac{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \ldots + P(B \mid A_n)P(A_n)}{P(B)} \qquad (2)$$

Since the left side of formula (2) is equal to 1, we get:

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \ldots + P(B \mid A_n)P(A_n) \qquad (3)$$

then we have:

$$P(A_j \mid B) = \frac{P(B \mid A_j)P(A_j)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \ldots + P(B \mid A_n)P(A_n)}$$

This process is called Normalization. So we treat 1 / P (B) as a normalizing constant, and obtain the following form for Bayes' rule:

$$P(Y \mid X) = \alpha P(X \mid Y)P(Y) \qquad (4)$$

We can extend this single evidence rule to a multiple evidence rule as:

$$P(Z \mid Y, X) = \alpha P(X \mid Z)P(Y \mid Z)P(Z) \qquad (5)$$

where α is a normalization constant.

For equation (5), if variables Z and X are conditionally independent, then

$$P(Z \mid Y, X) = P(Z \mid Y) \qquad (6)$$

Although the above simplified form of Bayesian updating is useful, note that it works only when the conditional independence among the variables holds. We will use it a lot in the following chapter.

Properties of Bayesian Network

The most distinctive feature of Bayesian networks, stemming largely from their causal organization, is their ability to represent and respond to changed configurations. How does it accomplish this task? To answer this question, we need to know the properties of this network. A Bayesian network is a graph with the following properties:

1. The nodes of the network present the variables (propositions).
2. The relation between two variables is denoted by a direct link between nodes. The expression X → Y means that X is the parent of Y and it directly influences Y.
3. Each node has a conditional probability table to keep the specific influence ration that all parents pass to this node.
4. The graph has no directional cycles.

With this data structure, we can compute any necessary probabilities in the domain[3,7].

Joint Probability Distribution

Figure 1 illustrates a simple but typical Bayesian network. It describes the causal relationships among the variables. The network provides a complete description of the domain. For a belief network representing variables, a joint probability is given by the following formula:

$$P(X_1, X_2, \ldots, X_n) = \prod_{j=1}^{n} P(X_j \mid F(X_j))$$

where F (X_j) is the set of parents of X. In the example from Figure 1,

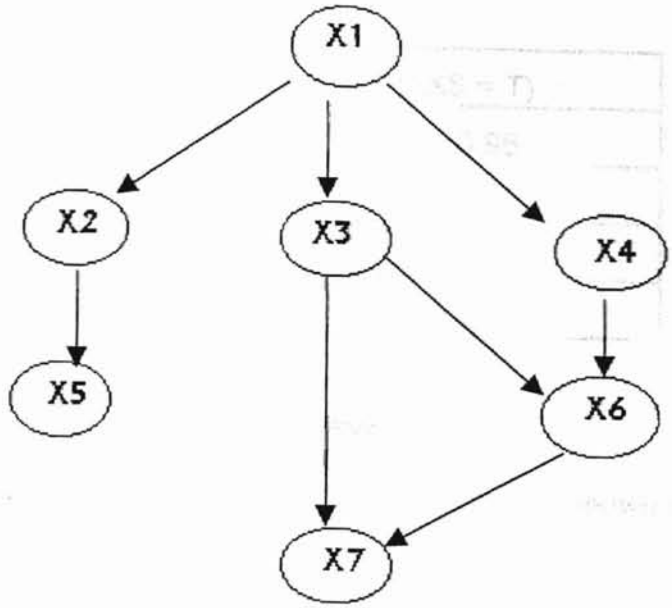

Figure 1. Bayesian Network

P $(X_1, X_2, X_3, X_4, X_5, X_6, X_7)$

$= P(X_7|X_6, X_3)\, P(X_6|X_4, X_3)\, P(X_3|X_1)\, P(X_4|X_1)\, P(X_5|X_2)\, P(X_2|X_1)\, P(X_1).$

Conditional Probability Table

Table 1 and Table 2 are examples of conditional probability tables (CPT) for X_6 and X_4 in Figure 1. Assume that there are only two kinds of values for X_4 and X_6: true and false.

X1	P (X4=T)
T	0.85
F	0.1

Table 1. Conditional probability table for X_4.

X3	X4	P (X6 = T)
T	T	0.95
T	F	0.85
F	T	0.3
F	F	0.02

Table 2. Conditional probability table for X_6

For each node in the network, we need to specify a CPT. As shown above, each row in the table contains a conditional probability of each node value for a conditioning case. A conditioning case is just a possible combination of values for the parent nodes. For each row in the table, the sum of values is equal to 1.

The disadvantage of using tables to store these values is the large space requirement. For a node with n parents that have m kinds of values, the space we need is m^n. However, in most real-world domains, each variable is directly influenced by only a few other variables[6]. Therefore, the storage requirement remains manageable.

Conditional Independence Relations

We have known that conditional independence could simplify the computation of probability. But how can we know that the conditional independence relation holds in the Bayesian network? Here we will introduce a method called d-separation or direction-dependent separation[30, 31]. For a triple set of nodes x, y, and z, two links are involved. We define a set of nodes E as a subset of the network variables for which we have direct evidence. We say E d-separates two nodes x and y if every undirected path from the node x to the node y is blocked given E. Conditional independence is obtained in the following three cases:

(1) Tail-to-Tail (Figure 2): x and y are conditionally independent if z is in set E.

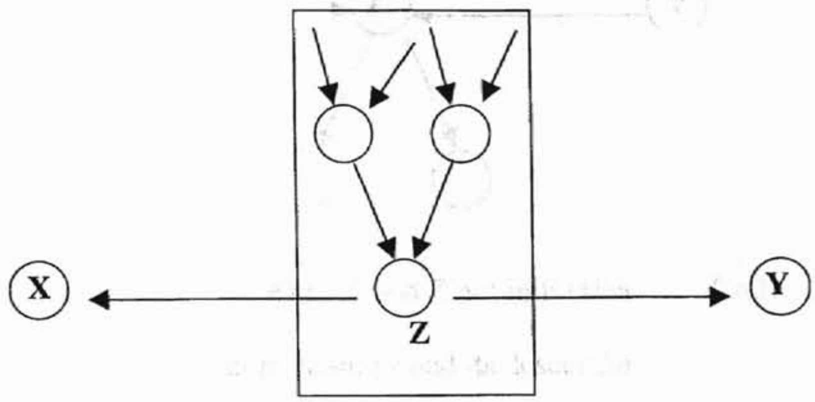

Figure 2. x and y are blocked by E and $z \in E$ (Tail-to-Tail).

(2) Head-to-tail (Figure 3): x and y are conditionally independent if z is in E.

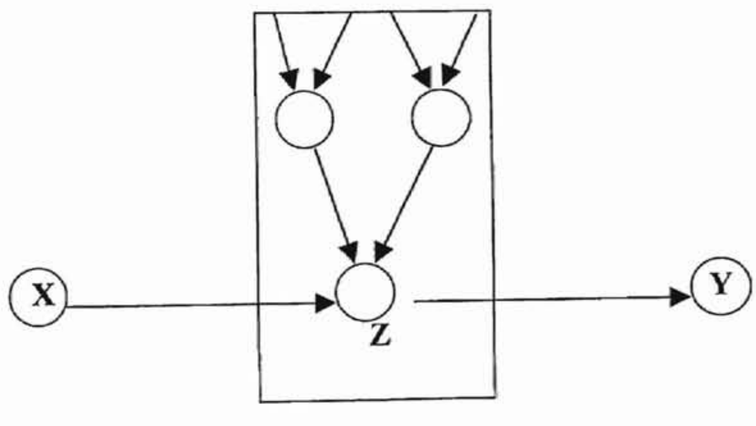

Figure 3. x and y are blocked by E and $z \in E$ (Head-to-Tail)

(3) Head-to-head (Figure 4): x and y are conditionally independent if z and its descendants are not in E.

In Figure 1, X_1 and X_5 are d-separated by X_2; X_4 and X_3 are d-separated by X_1, but not separated by $\{X_1, X_6\}$.

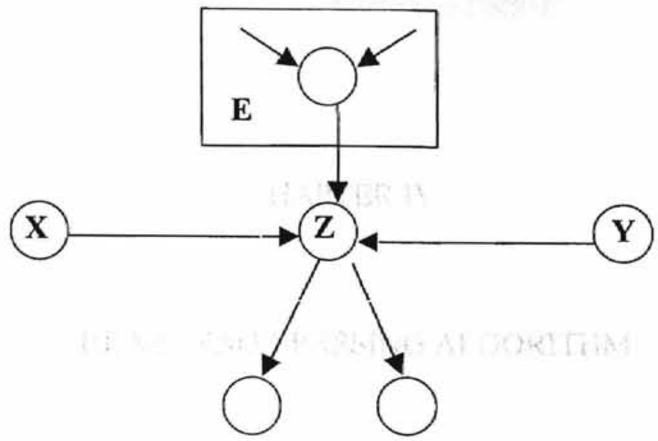

Figure 4. x and y blocked by E and Z not in E (Head-to-Head)

(Z is the set including z and its descendants)

The above concepts are very important to help us understand the calculation in a Bayesian network. The derivation of a message-passing algorithm is based on these concepts. In the following chapter, we will introduce the message-passing algorithm in detail.

CHAPTER IV

THE MESSAGE-PASSING ALGORITHM

This algorithm was created by Judea Pearl in 1986[22]. It only works on singly-connected networks. In Pearl's approach, "the network is not only a passive parsimonious code for storing factual knowledge but also a computational architecture for reasoning about that knowledge."[22] The links in the network are treated as the channels that direct and propel the flow of data in the process of querying and updating beliefs. The nodes are treated as processors whose functions are not only maintaining the parameters of belief for the host variables but also managing the communication links which connect with its neighbors (parents and children). The computation can be activated by a change of evidence or a clock or at random. When a certain processor is activated, it interrogates the belief parameters associated with its neighbors and compares them to its own parameters. If the parameters have no changes, no action is taken. Otherwise, it needs to update its parameters. This will activate similar revisions at the neighboring nodes and will begin a multidirectional propagation process[20, 32]. This chain reaction will stop when a new equilibrium is reached.

Computation with Single Parent

After establishing the Bayesian network and the CPT for each node, we can estimate the probability of any of the nodes. Since the multi-connected network is extremely complicated, and the algorithm is still in discussion, we will introduce an algorithm that works only in a singly-connected network, which is also called polytree. A polytree has one more restriction compared with other networks. That is, there is only one path between any two nodes in the network. Figure 1 is not a polytree, because there are two paths between X_1 and X_6 ($X_1 \rightarrow X_4 \rightarrow X_6$ and $X_1 \rightarrow X_3 \rightarrow X_6$) and three paths between X_1 and X_7. But if we omit the links between X_3 and X_6, X_6 and X_7, then it becomes a polytree. Figure 5 is a local part of a singly-connected network. It describes the relationship of nodes in a network. Suppose we want to compute the variable X given the set of evidence E. The evidence E is divided into E_X^+ and E_X^-.

- E_X^- stands for the descendants of X.
- E_X^+ stands for the ancestors of X.

We can begin with following equation:

$$P(X \mid E) = P(X \mid E_X^+, E_X^-).$$

Using Bayes' rule we get:

$$P(X \mid E_X^-, E_X^+) = \frac{P(E_X^- \mid X, E_X^+) P(X \mid E_X^+)}{P(E_X^- \mid E_X^+)}.$$

From Figure 5, we know that E_X^- and E_X^+ are conditionally independent. We can treat

$1 / P\,(E_X^- \mid E_X^+)$ as a normalizing constant. Therefore, the above equation becomes:

$$P(X \mid E) = \alpha\ P(X \mid E_X^+)P(E_X^- \mid X) \qquad (7)$$

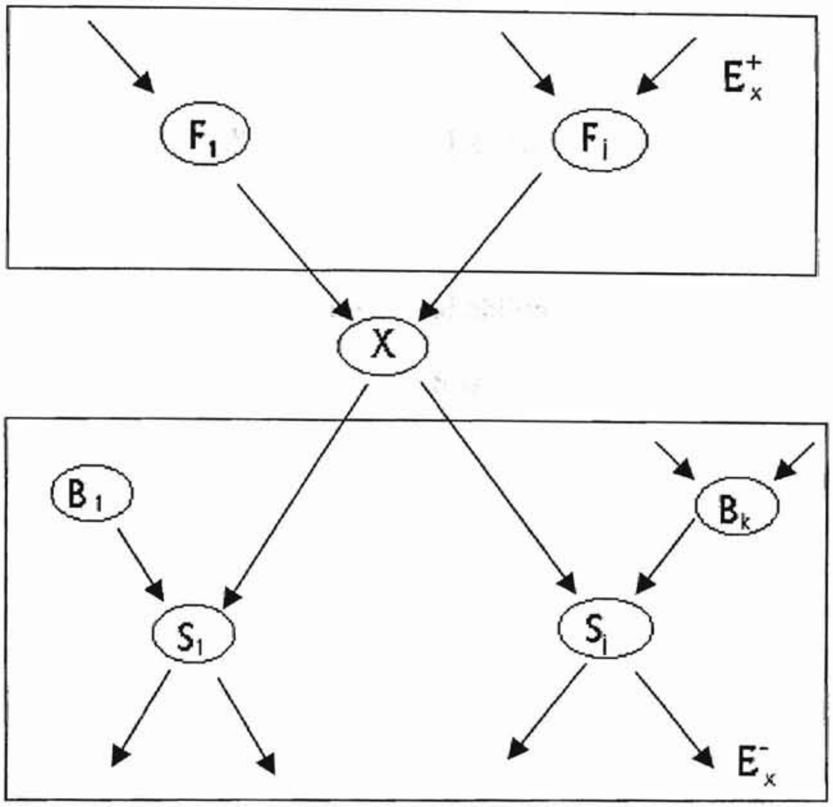

Figure 5. Fragment of a singly-connected network showing the partition of parents and children to the computing variable X.(From E. Rich 1991)

For the convenience of illustration, we will use the following symbols which are introduced by Pearl.

- $Bel(X_j) = P(X_j \mid E)$, X_j means the jth possible value of X. $Bel(X_j)$ stands for belief in X_j and denotes the dynamic actual value of the updated node probability.
- $\pi(X_j) = P(X_j \mid E_X^+)$, which represents the causal support attributed to X.
- $\lambda(X_j) = P(E_X^- \mid X_j)$, which means the diagnostic support from X's descendants.

So

$$Bel(X_j) = \alpha\, \lambda(X_j) \pi(X_j). \qquad (7)$$

As we observed, E^-_x can be partitioned into E^{1-}, E^{2-}, ... , E^{n-}, one for each subtree emanating from X. Since X d-separate these subtrees, conditional independence holds. We get:

$$\lambda(X_j) = \prod_K P(E^{K-} \mid X_j). \qquad (8)$$

The values of λ and π are stored with each node of the network. Then λ and π of query variable X are determined by its parents and children. For instance we want to compute the gth multiplicand in the product of (8). S is the gth child of X and S_k has K possible values. We use:

$$P(E^{g-} \mid X_j) = \sum_K P(E^-_g \mid X_j, S_K) P(S_K \mid X_j) \qquad (9)$$

We can replace (9) with the following:

$$P(E^{g-} \mid X_j) = \sum_K P(S_K \mid X_j) \lambda(S_K) \qquad (10)$$

This equation means that the gth multiplicand is obtained by the λ stored in the gth child of X times the entries in its conditional independent table. To make the step meaningful, we treat each multiplicand as a message sent by gth child of X. If the child is called S, the message will be denoted by $\lambda_S(X)$. Therefore, equation (10) is changed to

$$\lambda_S(X_j) = \sum_K P(S_K \mid X_j) \lambda(S_K) \qquad (11)$$

For the second part in (7), we have

$$\pi(X_j) = \sum_e P(X_j \mid F_e) P(F_e \mid E^+_X)$$

where F is a parent of X and e is the number of values of F. We can rewrite the above equation as follows:

$$\pi(X_j) = \sum_e P(X_j \mid F_e)\left[\alpha\pi(F_e)\prod_m \lambda_m(F_e)\right], \tag{12}$$

where m varies over the siblings of X. We call the expression in the brackets the message $\pi_X(F)$ which F transmits to X:

$$\pi_X(F_j) = \alpha\pi(F_j)\prod_m \lambda_m(F_j). \tag{13}$$

A more useful expression of (13) is

$$\pi_X(F_j) = \frac{Bel(F_j)}{\lambda_X(F_j)}. \tag{14}$$

So now we rewrite (7) as

$$Bel(X_j) = \alpha\prod_K \lambda_K(X_j)\sum_e P(X_j \mid F_e)\pi_X(F_e) \tag{15}$$

Figure 6 illustrates the message passing among the node X and its parents and children. It also shows the processes of propagation:

1. Processor X is activated, and updates its parameters by using (7).
2. Belief updating and belief revision involve updating and transmitting two types of messages. First, the strength of the evidential support that X obtains from its descendants is updated using the following equation:

$$\lambda(X_j) = \lambda_1(X_j)\lambda_2(X_j)\ldots\lambda_K(X_j) = \prod_K \lambda_K(X_j). \tag{16}$$

3. The second message refers to the strength of a causal support that X obtains from its ancestors, which is computed by:

$$\pi(X_j) = \beta\sum_K P(X_j \mid F_K)\pi_X(F_K). \tag{17}$$

4. Bottom-up propagation. The message which X sends to its parent F computed by:

$$\lambda_X(F_e) = \sum_j P(X_j \mid F_e)\lambda(X_j).$$

5. Top-down propagation. The message which X sends to the gth child S is computed by:

$$\pi_S(X_j) = \alpha\pi(X_j)\prod_{m \neq g}\lambda_m(X_j), \text{ or}$$

$$\pi_S(X_j) = \alpha\,\frac{Bel(X_j)}{\lambda_S(X_j)}.$$

Computation with Multiple Parents

Since we understand the computation with single parent, the multiple parents computation is the extension of the previous one. Figure 6 is a classical diagram from Pearl's article for illustration of the propagation in the network. Although it is only a fragment of the network, the rest of the computation is a recursive repetition of the same process.

In Figure 7, the node A in the network is the query variable. The possible values of A are denoted A_1, A_2, ..., A_n. Incoming evidence to node A through instantiated variables, will be denoted D.

The arc B→A from Figure 7 partitions the graph into two parts: an upper subgraph G^+_{BA} and a lower subgraph G^-_{BA}. Data contained in G^+_{BA} and G^-_{BA} will be denoted D^+_{BA} and D^-_{BA}, respectively. Similarly, each of arcs C→A, A→X, and A→Y partitions the graph into two subgraphs, containing corresponding data. As we see from the Figure 7, node A separates the network into four groups: G^{+}_{BA}, G^{+}_{CA}, G^{-}_{AX} and G^{-}_{AY}.

Using Bayes' rule,

$$Bel(A_j) = P(A_j \mid D^{+}_{BA}, D^{+}_{CA}, D^{-}_{AX}, D^{-}_{AY})$$

$$= \alpha P(A_j \mid D^{+}_{BA}, D^{+}_{CA}) P(D^{-}_{AX} \mid A_j) P(D^{-}_{AY} \mid A_j)$$

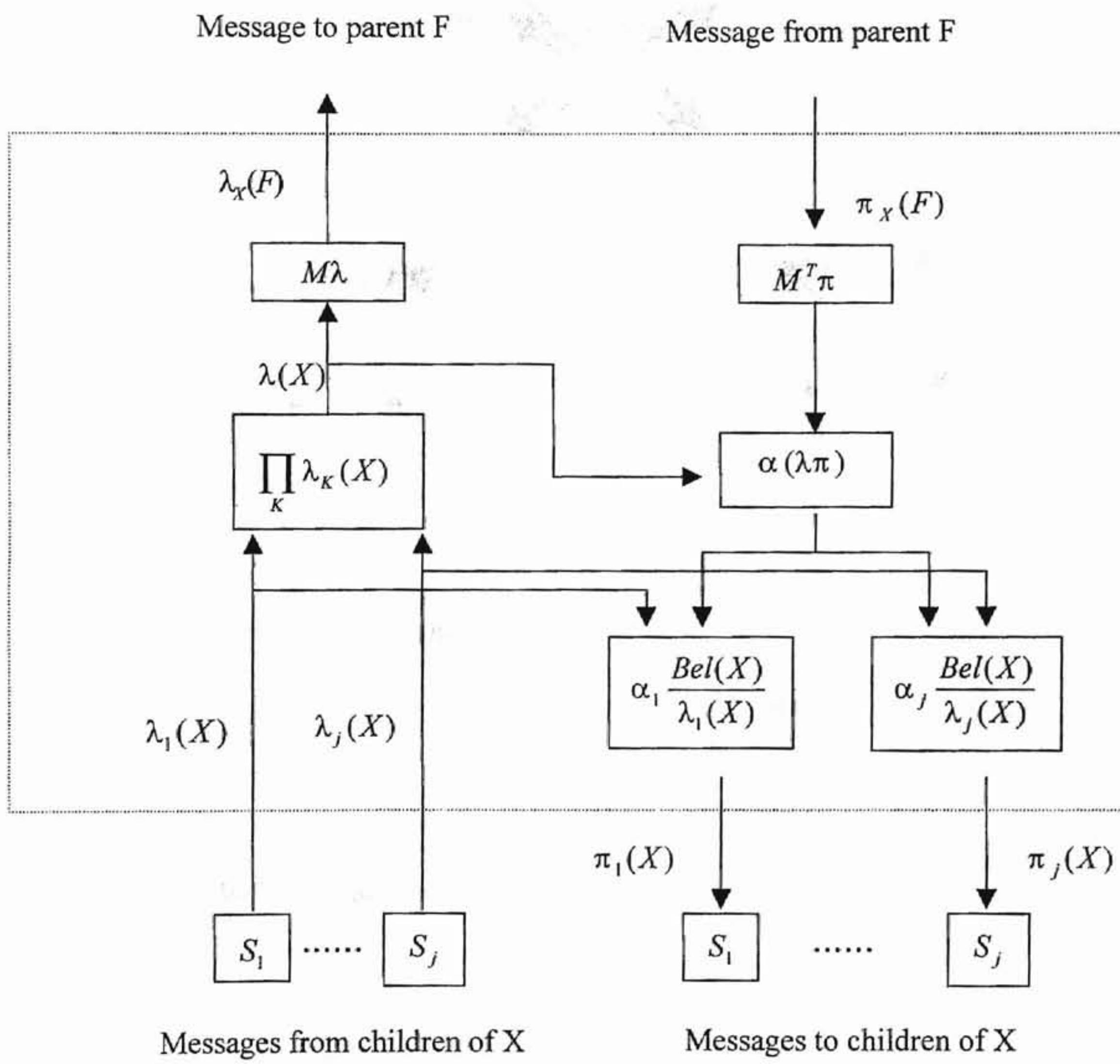

Figure 6. The message passing in local network during updating of variable X (only single parent) (From Pearl 1986)

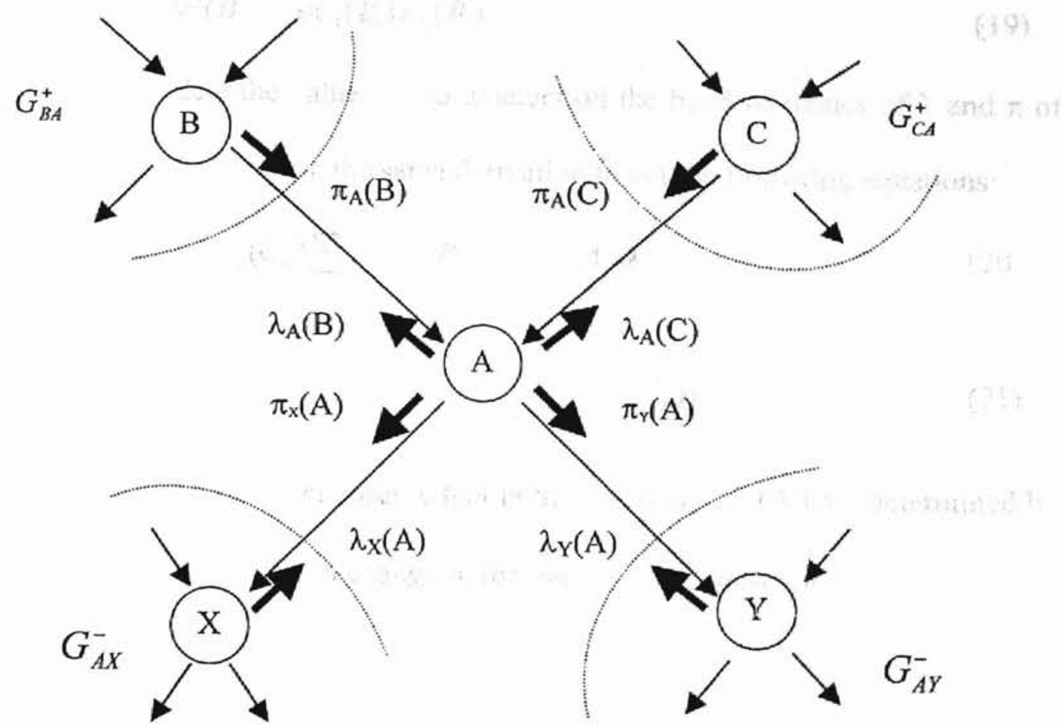

Figure 7. A part of a belief network (from Pearl 1986).

where α is a normalizing constant. Furthermore,

$$Bel(A_j) = \alpha P(D^{-}_{AX} \mid A_j) P(D^{-}_{AY} \mid A_j) (\sum_{i,k} P(A_j \mid B_i, C_k) P(B_i \mid D^{+}_{BA}) P(C_k \mid D^{+}_{CA}))$$

In the last equation, the current strengths of incoming arcs to A will be denoted as $\pi_A(B_i)$ and $\pi_A(C_k)$, which are called casual supports. The current strengths of outgoing arcs from A will be denoted as $\lambda_X(A_j)$ and $\lambda_Y(A_j)$, which are called diagnostic supports. Now, we rewrite the above equation as following:

$$Bel(A_j) = \alpha \lambda_X(A_j) \lambda_Y(A_j) \sum_{i,k} P(A_j \mid B_i, C_k) \pi_A(B_i) \pi_A(C_k) \qquad (18)$$

We can compute the parent B's belief distribution by:

$$Bel(B_i) = \alpha\pi_A(B_i)\lambda_A(B_i) \tag{19}$$

The next step is to update the values of parameters on the basis of values of λ and π of the neighboring arcs. We can use the same derivation to get the following equations:

$$\lambda_A(B_i) = \alpha\sum_k(\pi_A(C_k)\sum_j P(A_j \mid B_i, C_k)\lambda_X(A_j)\lambda_Y(A_j)) \tag{20}$$

$$\pi_X(A_j) = \alpha\lambda_Y(A_j)(\sum_{i,k} P(A_j \mid B_i, C_k)\pi_A(B_i)\pi_A(C_k)) \tag{21}$$

From equations (20) and (21), we observe that both $\lambda_A(B_i)$ and $\pi_X(A_j)$ are determined by their neighboring parameters and a change in the value of the causal parameter π will not affect the value of the diagnostic parameter λ at the same arc, and vice versa. Therefore, no circular reasoning will take place.

Propagation

Figure 8 depicts five successive stages of belief propagation through a binary tree. Our example is similar to that of Pearl[22]. White tokens represent values $\lambda_A(B_i)$ that A sends to its parent B, while black tokens represent value $\pi_X(A_j)$ that A sends to its child X.

State (a): Initially, the tree is in a state of equilibrium.

State (b): A datum activates a node, so a white token is sent to its parents.

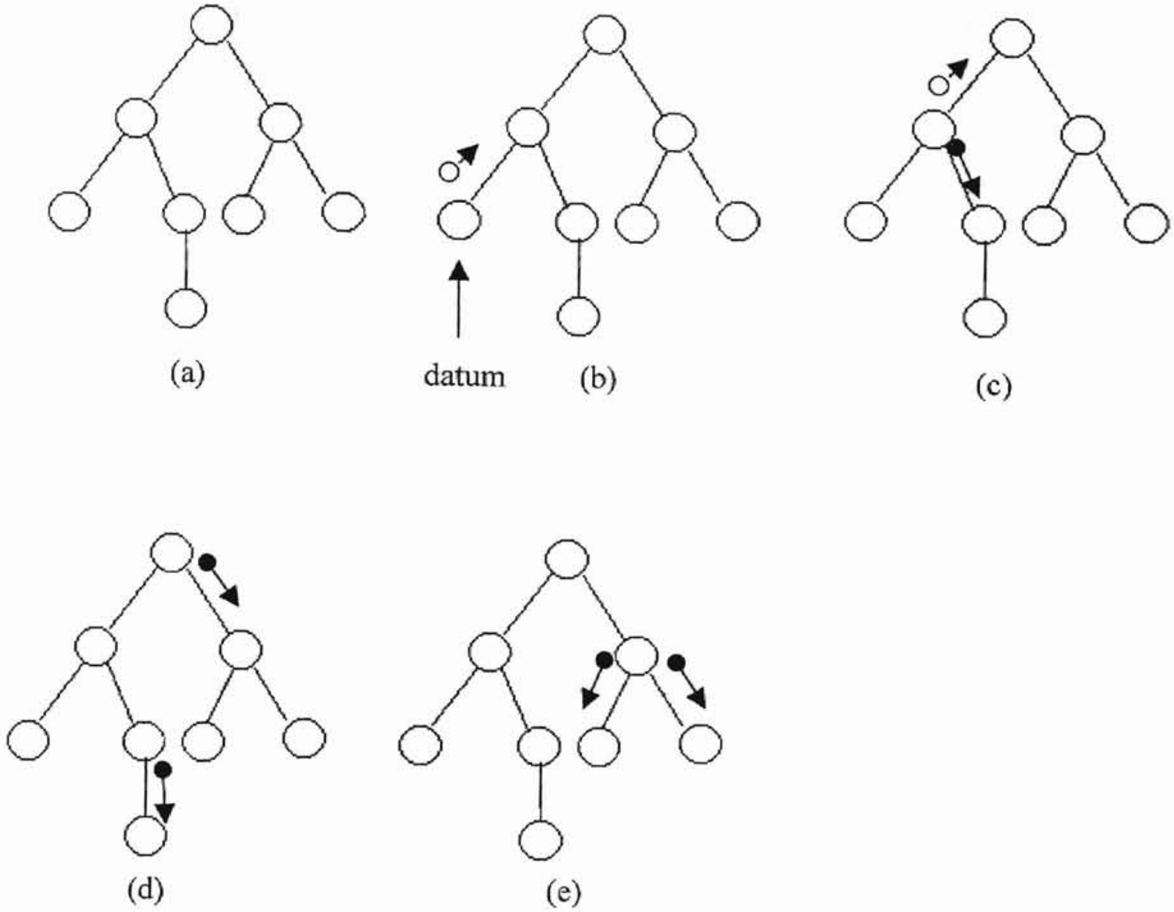

Figure 8. Belief propagation by message passing process

State (c): The parent sends a black token to its children as response and sends a white token to its parents.

State (d) to (e): The process continues until all the tokens are absorbed and the network reaches a new equilibrium.

The *message-passing algorithm* has the following advantages:

1. It makes each step understandable and meaningful.
2. It is a natural mechanism for exploiting the independence embodied in a sparsely constrained system.

3. Translating messages by subtask decomposition results in a substantial reduction in complexity.

CHAPTER V

DIAGNOSTIC AGENT'S DESIGN AND CONSIDERATION

Steps of Building an Expert Agent

As pointed out by Pepper[24], "All human diagnosticians, whether they work in automotive repair or medicine, have certain characteristics in common. Both groups have an internal mental model of the task domain. This model is a body of knowledge about the parts of the mechanism or organism they are trying to fix and about how those parts fit together. This model is closely tied to two additional knowledge sources: the expert's formal understanding of the laws of the domain and a large loosely structured body of knowledge consisting of common sense and experience gained simply by living in the world. Taken together, these three knowledge sources are very powerful and enable human beings to solve new problems…" So, regardless of the type of diagnostics performed, either industrial or medical, the same method can be used to build a diagnostic agent. The design steps to be taken are as follows:

1. Collect the knowledge about the domain.
2. Analyze the acquired knowledge and choose the set of relevant variables that describe the domain.
3. Choose an ordering for the variables and record their attributes and associated values in conditional tables.

4. Structure and model the knowledge.

5. Implement the model.

Knowledge collection

As mentioned in Chapter One, the purpose of this research is to build a diagnostic agent using a Bayesian network. Our domain is medical diseases and symptoms. This agent will make a differential diagnosis between ectopic pregnancy and acute salpingitis. This agent will be used to test if the message-passing algorithm is sensitive and accurate enough to make a differential diagnosis. I chose fifteen common symptoms, two predisposing factors and the above two diseases as variables. The attribution and relation among them are listed in Table 3 and Table 4. Since accurate data are based on tremendous statistical research, I collected the statistical data which was already published and also estimated some other data by myself after discussions with some doctors. We can adjust those data if statistical data become available. The computable data of ectopic pregnancy and salpingitis are listed in Table 4[1, 5, 9, 10, 14, 17, 23, 25, 40, 42]. All the data related to ectopic pregnancy were obtained from statistical research of over 1000 patients with ectopic pregnancy. Some of the data related to acute salpingitis were from the estimation.

We use F to stand for fallopian tube infection or surgery history and C to stand for Congenital abnormalities of the fallopian tube. Data in Table 3 are attributions which are estimated by individual experience.

Table 3. Relations among F, C and ectopic pregnancy

F	C	Ectopic Pregnancy (true)
T	T	0.65
T	F	0.40
F	T	0.50
F	F	0.004

Table 4. Probability of symptoms, signs and lab tests for

ectopic pregnancy and acute salpingitis

Symptoms, Signs And Lab test	Incidence (Percent)			
	Ectopic Pregnancy		Acute Salpingitis	
Symptoms:	T	F	T	F
1. Pregnancy symptoms	20	0.5	--	--
2. Abdominal pain	90	1	90	1
3. Amenorrhea	85	2	--	--
4. Vaginal bleeding	70	6	10	5
5. Dizziness and syncope	10	2	8	2
6. Fever	8	3	52	3
Signs:				
7. Abdominal tenderness	70	2	80	2
8. Adnexal tenderness	80	1	70	1
9. Adnexal mass	45	0.3	20	0.2
10. Uterine enlargement	25	2	--	--
11. Cul-de-sac fullness	60	1	--	--
12. Orthostatic hypotension	10	1	--	--
Lab tests:				
13. White cell count > 15000/μl	20	6	70	5
14. β-hCG (+)	92	2	10	1
15. Pelvic Ultrasound	60	0.3	--	--

Prior probability of F = 0.03.

Prior probability of C = 0.001.

Prior probability of acute salpingitis = 0.03.

Building Network

Usually, in medical diagnostic processes, doctors collect the inducing factors that cause certain disease and the symptoms that are caused by the disease. That means only three layers of variables are needed for a medical diagnostic expert system. We divided the variables into three layers: predisposing factors layer, disease layer and symptom layer. With the variables above, we can build the network as in Figure 9. Since we use the message-passing algorithm, the main design problem is already solved by Pearl. We will separate the network in Figure 9 into two parts: the ectopic pregnancy part (Figure 10) and salpingitis part (Figure 11). This will simplify the calculation. What's more important is that we can get more published data from medical references. It is obvious that it is much easier to find the probability of fever given respiratory infection in published material than find the probability of fever given diseases respiratory infection and hepatitis. Since the properties of the network are not changed, Pearl's algorithm still works, but is based on data that is easier to collect. As you will see, our results support this point. The content stored in nodes in the network will list as follows:

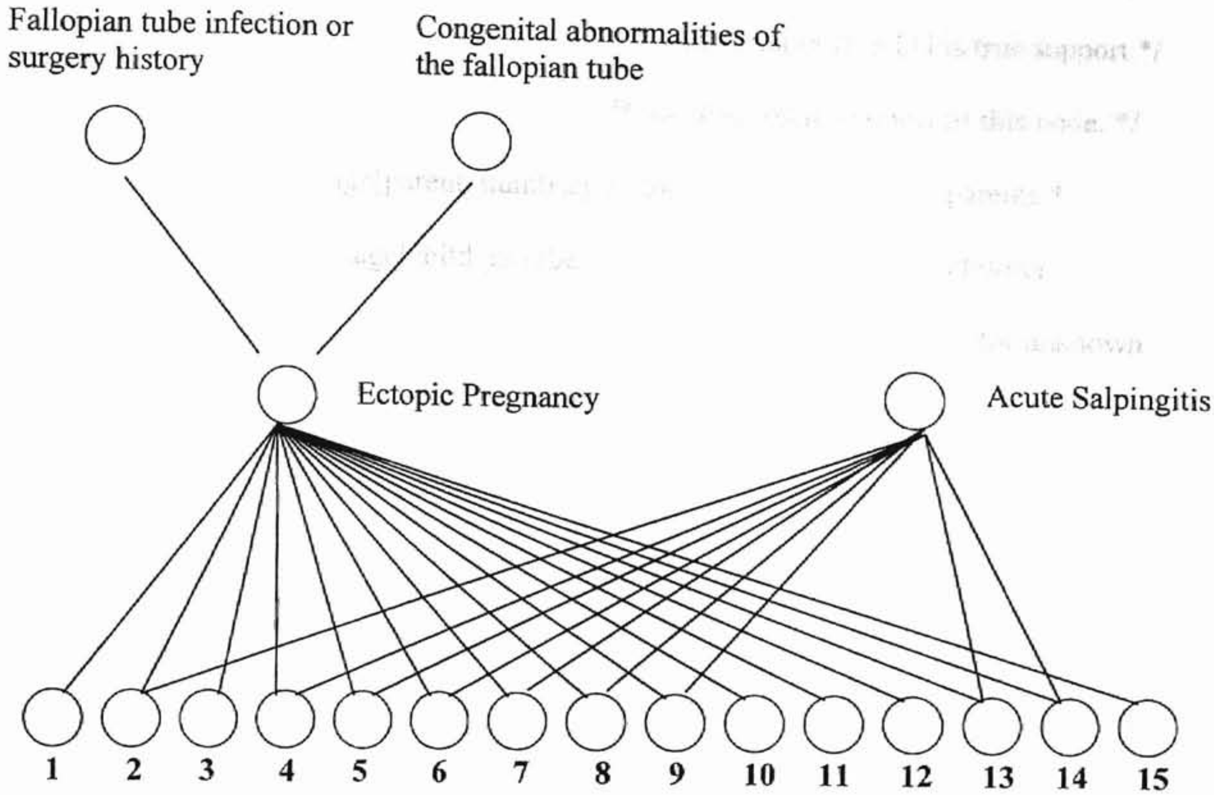

Figure 9. The bayesian network of ectopic pregnancy and acute salpingitis.

Structure of node

```
{
    char name[max_length]              /* the name of this node */
    int node_ID                        /*  The ID number of this node in network */
    int parents_number                 /* the number of parents of this node */
    int parent[parent_number]          /* the array of the parents of the node */
    int child_number                   /* the number of children of the node */
    int child[child_number]                /* the array of the children of the node */
    Table c_table[MAX_ATTRIBUTION] /* Conditional table */
```

```
    float pi[2]                              /* the causal support of this node. pi[0]
                                                is false support, pi[1] is true support */
    float lamda[2]                           /* the diagnostic support of this node. */
    float parent_message[parent_number]  /* the messages from the parents */
    float children_message[child_number] /* the messages from the children */
    int evidence                             /* 0 for false, 1 for true and 2 for unknown */
}
```

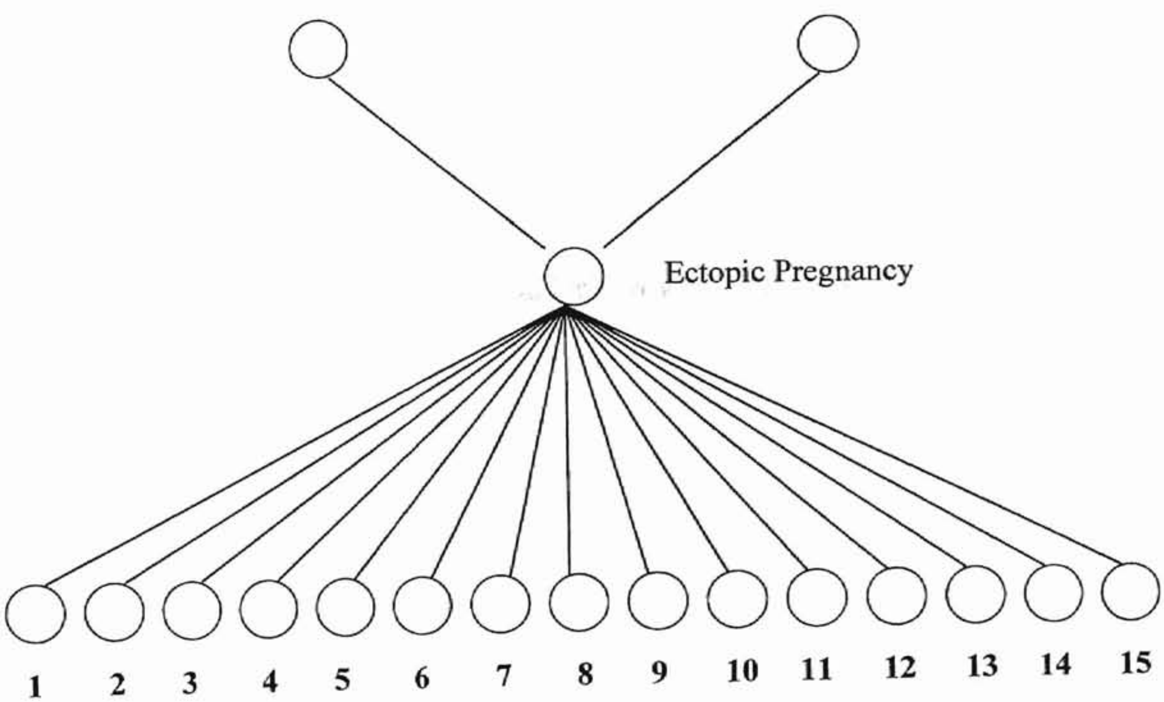

Figure 10. The Bayesian network of ectopic pregnancy.

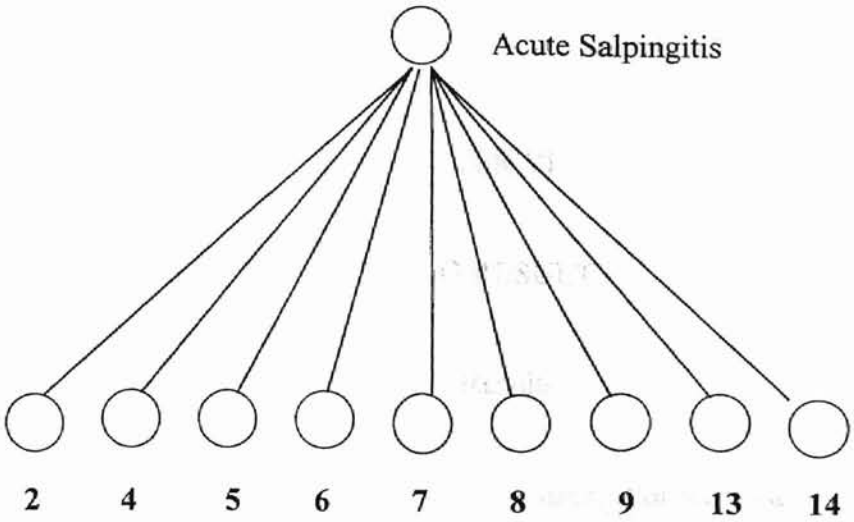

Figure 11. The Bayesian network of acute salpingitis

Since the network is established and the necessary knowledge is ready, the next step is to implement and test this agent which is designed by using the message-passing algorithm. The following chapter shows experimental results and the analysis of my agent. There, the capabilities of this algorithm will be examined for its use in medicine diagnosis.

CHAPTER VI

TESTING RESULT

List of Results

In order to test this work, I made up ten cases. For each case, I created several predisposing factors and symptoms, then I asked a doctor of gynecology and obstetric to make a diagnosis. After comparing the results given by the doctor and the results produced by this agent, I believe that this algorithm matches the situations in the real medical world. I realize that this algorithm has its limitations also and more work is needed to be done in order to make the system appropriate for clinical medicine. Table 5 contains the ten made-up patients' symptoms and Table 6 lists the comparing results.

In Table 5:

"-" means this evidence is negative.

"+" means this evidence is positive.

"~" means this evidence is unknown.

In Table 6:

"AS" means Acute salpingitis. The probability of AS larger than 50%.

"EP" means Ectopic pregnancy. The probability of EP is larger than 50%.

"UN" means unknown (need more evidence before making a diagnosis). Both probability of acute salpingitis and ectopic pregnancy are less than 50%.

"Both" means the patient has both of the diseases. The probability of both diseases are higher than 50%.

"H" means healthy. All the abnormal symptoms are negative.

Table 5 Information of ten made-up patients

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
F. tube inf./surg.history	-	+	-	+	+	-	-	-	~	-
Cong. Abnor. Of F. tube	-	-	-	+	~	-	-	+	~	-
Pregnancy symptoms	-	-	-	+	+	-	-	-	+	-
Abdominal pain	+	+	-	-	+	-	+	+	-	+
Amenorrhea	-	-	~	+	-	~	+	+	+	-
Vaginal bleeding	-	-	+	-	-	+	-	-	-	-
Dizziness and syncope	+	-	+	+	-	+	-	-	-	-
Fever	+	-	-	-	-	-	+	-	-	-
Abdominal tenderness	-	-	-	+	-	-	+	-	-	+
Adnexal tenderness	~	+	-	-	-	+	~	-	-	-
Adnexal mass	-	-	-	-	-	-	-	-	-	+
Uterine enlargement	-	-	-	-	-	-	-	-	+	-
Cul-de-sac fullness	-	-	-	-	-	-	-	-	+	-
Orthostatic hypotension	+	~	~	-	-	-	-	-	-	-
WBC count > 15000/ul	+	~	~	-	+	+	+	+	-	+
Beta-HCG	-	~	~	+	-	-	~	+	+	-
Pelvic Ultrasound	-	~	~	+	~	~	~	~	~	~

Table 6 The comparing of diagnosis by doctor and agent

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
Probability of E. P.	1	89	0.03	100	20	2.7	99	99	0	48.6
Probability of A. S.	99	93	0.07	15	44	69.6	99	89.5	0	99
Diagnosis by agent	AS	Both	UN	EP	UN	AS	Both	Both	H	AS
Diagnosis by doctor	*	**	UN	EP	UN	AS	***	Both	H	AS

"*" (For P1) The evidence strongly supports any infection in the body, but the probability of getting EP is very low.

"**" (For P2) Both of these diseases are possible, but doctor would not make a diagnosis before getting the lab result.

"***" (For P7) The same suggestion as P2.

As we see in table 6, most of the diagnoses made by my agent match the doctor's diagnoses. P1, P2 and P7 are different. The doctor's opinions are listed above.

Analysis

Although the diagnostic matching rate is higher than 80%, as we noticed, there are also several cases that are not matched. The reasons of these errors may lay in the following areas:

1. The network is incomplete. That means the symptoms in my network are not caused only by acute salpingitis and ectopic pregnancy, but can also be caused by other diseases. For example, fever can be caused by any infection. That's the reason the doctor didn't think the first patient (P1) has acute salpingitis. She thought other infection diseases were possible.
2. The knowledge database is not accurate enough. In cases of P2 and P7, the probability of both diseases are about 90%. That means the agent made a sure diagnosis while the lab tests were unknown. But in the real world, most of the doctors would not make a decision before some important lab examinations are done. So, my agent over-estimated the probability of these diseases. On the other hand, if all the knowledge data are accurate, this may not be a shortcoming of this agent

because if the symptoms obviously suggest certain disease, it is not always a mistake to make a diagnosis.

3. The assumption of conditional independence of symptoms usually does not apply and can lead to substantial errors in certain settings.
4. In many domains, it may be inaccurate to assume that relevant conditional probabilities are stable over time. Furthermore, diagnostic categories and definitions are constantly changing, as are physicians' observational techniques, thus invalidating data previously accumulated.

The above four problems may lead to the deviation of my agent's diagnosis from that of the physician. Frankly, they are also the limitations of Pearl's algorithm. In order to solve the first and second problems, we need the cooperation of knowledge engineers and experts in medicine. With help from them, we can build a complete network and get the most accurate knowledge database. Obviously, to solve the fourth problem, we need to update the knowledge database frequently according to the changing situation. As for the third problem, I think we still can't find an efficient way to deal with it today. The best recommendation is don't believe the diagnostic agent with 100% confidence and keep an eye on it. It may make mistakes like the human being.

CHAPTER VII

CONCLUSION

Summary

Computer programs to assist with medical decision making have long been anticipated by physicians with both curiosity and concern. In the past forty years, a lot of research has been done in this area and many medical expert systems have been developed. Motivation for the development of expert systems in medicine has been abundant. A physician may have knowledge of most diseases, but, due to the extensive number of diseases, a physician could benefit from the support provided by an expert system to quickly isolate the disease. This is also the task of my agent. In my research, I built a medical diagnostic agent using the theory of Bayesian networks. I think the results of this agent are encouraging. It demonstrated that this agent was sensitive enough to handle a medical diagnosis, even though the Bayesian approach has its limitations. But I do think we can narrow the chance of making errors. As we cannot expect a physician to be 100% correct, neither can we expect the expert system to always be right. I think such agent can be a powerful tool for physicians like other technical equipment. I don't think it is realistic that the diagnostic machines will totally replace human physicians. Compared with other methodologies like rule-based reasoning, neural networks and case-based reasoning, the Bayesian approach has more potential for development. Although, today there are only a few applications that were built with the Bayesian approachs in medicine expert systems, I believe it is just beginning. With the development of the software technologies, more related tools and shells for the Bayesian approach will be marketed. Then, researchers will feel at ease about developing the practical medical agents with a Bayesian approach.

Future Work

The development of medical expert systems brings with it many formidable technical, behavioral, legal, and ethical problems that must be addressed by the researchers in this field. These include acquiring and representing medical knowledge, validating the systems, getting physicians and patients to accept them, and deciding who will be responsible for clinical decisions made with the help of these systems. In assessing applications, it is pertinent to examine the following research issues that affect the success of expert system in medicine[13].

- What is the appropriate domain in medicine?
- How is the clinical knowledge to be acquired and represented? How does it facilitate the performance goals of the system described?
- Is the system accepted by users for whom it is intended?
- Is the interface with the user adequate?
- Is it suitable for dissemination?
- What are legal and ethical problems?

In recent years, the applications of expert systems in medicine appear to be growing at an almost exponential rate. However, among the expert systems that have been implemented, few of them have been successful in practice, especially in the clinical environment. In some cases, failures have definitely occurred, and many of these failures have been due to an improper selection of domains or a neglect of the critical factor of expert system maintenance. In others, failure may be traced to the choice of the wrong knowledge acquisition and representation methods. However, the most problems encountered in the implementation of expert systems have not been a fault of the methodology. My research results proved that the Bayesian network is a good approach to building medical diagnostic systems. To make the real world realize the usage of this theory and accept the expert systems built with this theory, the issues I listed above will require more attention in the future.

REFERENCES

1. Alsuleiman S.A., Grimes E.M., Ectopic pregnancy: A review of 147 cases, *Journal of Reproductive Medicine*, 27: 101, 1982.
2. Aliferis C.F., Cooper G.F., A new formalism for temporal modeling in medical decision-Support systems, *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*, 213-217, 1995.
3. Alag, S., A bayesian decision-theoretic framework for real-time monitoring and diagnosis of complex systems: theory and application, Doctoral Dissertation, Department of Mechanical Engineering, University of California at Berkeley. 1996.
4. Bayes T., An essay towards solving a problem in the doctrine of chances, 1763. *Philosophical Transactions of the Royal Sociaty of London*, 53:370-418. Reprinted with commentary by Barnard G.A. in Biometrika, 45:293-315, 1965.
5. Brenner P.F., Ray S., Mishell D.R., Ectopic pregnancy a study of 300 consecutive surgically treated cases, *Journal of American Medical Womens Association*, 243: 673, 1980.
6. Chavez R.M., Cooper G.F., A randomized approximation algorithm for probabilistic inference on bayesian belief networks, *Networks*, 20:661-685, 1990.
7. Cooper G.F., The computational complexity of probabilistic inference in bayesian networks in NP-hard. *Artificial Intelligence*, 42(2-3), 393-405, 1993.
8. De Dombal F.T., Leaper D.J., Staniland J.R., Computer-aided diagnosis of acute abdominal pain, *British Medical Bulletin* , 376-380, 1972.
9. De Cherney, A.H., Minkin, M.J., Spangler, S., Contemporary management of ectopic pregnancy, *Journal of Reproductive Medicine*, 26: 519, 1981.
10. Greiss F.C., Think ectopic pregnancy, *American Family Physician* 17:126, 1978.
11. Grimm R.H., Shimoni K., Harlan W.R., Evaluation of patient-care protocol use by

various providers, *New England Journal of Medicine*, 292:507-511, 1975.

12. Hickam D.H., Shortliffe E.H., Bischoff M.B., et al. The treatment advice of a computer-based cancer chemotherapy protocol advisor, *Annuals of Internal Medicine*, 103:928-936, 1985.
13. Jay L., *The Handbook of Applied Expert Systems*, CRC Press, 1998.
14. Kilchin J.D., Wein R.M., Nunley W.C., Ectopic pregnancy: Current clinical trends, *American Journal of Obstetrics and Gynecology*, 134:870, 1979.
15. Langlotz C.P., Shortliffe E.H., Adapting a consultation system to critique user plans, *International Journal of Man-Machine Study* 19:479-496, 1983.
16. Ledley R.S., Lusted L.B., Reasoning foundations of medical diagnosis, *Science*, 130:9-12,1959.
17. McBride J.R., Wait R.B., Ectopic pregnancy: A review of 212 cases, *Texas Medicine*, 76:121, 1980.
18. Miller P.L., *Expert Critiquing Systems: Practice-Based Medical Consultation by Computer*, New York , Springer-Verlag NY, 1986.
19. Miller R.A., Pople H.E., Myers J.D., Internist-1: An evolving computer-based diagnostic consultant for general internal medicine, *New England Journal of Medicine* 307:468-476,1982.
20. Musick R., Minimal assumption distribution propagation in belief networks, *In 9th Conference Uncertainty in Artificial Intelligence*. 232-241, 1993.
21. Pauker S.G., Gorry G.A., Kassirer J.P., Toward the stimulation of clinical cognition: Taking a present illness by computer, *American Journal of Medicine*, 60:981-996,1976.
22. Pearl J., Bayesian Decision Methods, *Encyclopedia of Artificial Intilligence*, Wiley Interscience, 1986.
23. Pelosi M.A., Apuzzio J., Dwyer J.W., Early diagnosis of ectopic pregnancy, *Female Patient*, 8: 42, 1983.
24. Pepper J., An Expert System for Automative Diagnosis, *The MIT Press*, Cambridge, MA, 1990.
25. Portuondo J.A., Remacha M.J., Llaguno M.R., Ectopic pregnancy early diagnosis

limitations, *International Journal of Gynaecology and Obstetrics*, 20:371, 1982.

26. Reggia J.A., Tuhrim S., *Computer-Assisted Medical Decision Making*, New Nork, Springer-Verlag NY, 1985.
27. Rich, E., Knight, K., *Artificial Intelligence*, second edition, McGraw-Hill, 1991.
28. Robert L.W., Real time estimation of bayesian networks, *In 12th Conference Uncertainty in Artificial Intelligence*, 533-544, 1996.
29. Russell S. and Norvig P., *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.
30. Silvia A., Luis M. de Campos, An algorithm for finding Minimum d-separating sets in belief networks, *In 12th Conference Uncertainty in Artificial Intelligence*, 3-10, 1996.
31. Smith J.E., Holtzman S., Matheson, Structuring conditional relationships in influence diagrams, Operations Research, 41(2):280-297, 1993.
32. Satnam A., Alice M.A., Inference using message propagation and topology transformation in vector gaussian continuous networks, *In 12th Conference Uncertainty in Artificial Intelligence,* 20-28, 1996.
33. Shafer G., A mathematical theory of evidence, *Princeton University Press*, 1976.
34. Shortliffe B., *Rule-based Expert System*, Addison-Wesley, 1984.
35. Shortliffe E.H., Buchanan B.G., Feigenbaum E.A., Knowledge engineering for medical decision making: A review of computer-based clinical decision aids, *Proceedings of the IEEE*, 67:1207-1224, 1979.
36. Shortliffe E.H., Medical expert systems: Knowledge tools for physicians, *Western Journal of Medicine*, 145:830-839, 1986.
37. Shortliffe E.H., Computer programs to support clinical decision making, *Journal of the American Medical Association*, 258:1, 1987.
38. Shortliffe E.H., *Computer-Based Medical Consultations: MYCIN*, New York, Elsevier North Holland, 1976.
39. Tatro D.S., Briggs R.L., Chavez-Pardo R., et. al: Online drug interaction surveillance, *American Journal of Hospital Pharmacy*, 32: 417-420, 1975.
40. Thorneycroft I.H., When you suspect ectopic pregnancy. *Contemporary OB/GYN*,

22:91, 1983.

41. Warner H.R., Toronto A.F., Veasy L.G., Experience with Bayes' theorem for computer diagnosis of congenital heart disease. *Annuals of the New York Academy of Sciences*, 115:2-16, 1964.
42. Weckstein L.N., Boucher A.R., Tucher H., Accurate diagnosis of early ectopic pregnancy, *Obstetrica Si Gynecologia*, 65:393, 1985.
43. Weinstein M.C., Fineberg H.V., *Clinical Decision Analysis*, Philadelphia, WB Saunders,1980.
44. Zadeth L.A., The role of fuzzy logic in the management of uncertainty in expert systems, *Fuzzy Sets and Systems*, 11, 199:228, 1983.

2

VITA

Haobo Liu

Candidate for the Degree of

Master of Sciences

Thesis: A BAYESIAN NETWORK FOR INTERNAL MEDICAL DIAGNOSIS

Major Field: Computer Science

Biographical:

Personal Data: Born in Hunan, P. R. China, on April 13, 1970, the son of Shuzhen Chen and Zaiquan Liu; married to Haihui Huang in 1995.

Education: Graduated from NO.1 Middle School, Yiyan, Hunan, P. R. China in June, 1987. Received Bachelor of Medical Science degree from Hunan Medical University, Changsha, Hunan, and P. R. China in July, 1993. Completed the requirements for the Master of Sciences degree with a major in Computer Science at Oklahoma State University in December, 1998.

Experience: Employed by the Baoan People's Hospital as a neurosurgeon, Shenzhen, Guangdong, P. R. China, 1993 - 1995. Employed by the Department of Computer Sciences at Oklahoma State University in Stillwater, Oklahoma as a Teaching Assistant, 1998.