Directory

[■ 서론 3](#_Toc529137966)

[Chapter 1 Data mining 5](#_Toc529137967)

[Chapter 2 Bayesian theory 6](#_Toc529137968)

[2.1Knowledge of probability 6](#_Toc529137969)

[2.2 Bayesian networks 7](#_Toc529137970)

[2.3 Bayesian network structure 8](#_Toc529137971)

[2.4 Bayesian inference 10](#_Toc529137972)

[2.4.1 Causality reasoning 10](#_Toc529137973)

[2.4.2 Diagnostic reasoning 11](#_Toc529137974)

[2.4.3 Explain 12](#_Toc529137975)

[2.4.4 D separation 12](#_Toc529137976)

[2.4.5 Fruit fructose problem Example 14](#_Toc529137977)

[2.5 Application: Spell check 15](#_Toc529137978)

[2.6 Bayesian network learning 16](#_Toc529137979)

[Chapter 3 Bayesian network application 18](#_Toc529137980)

[3.1 Application of BN in fault diagnosis 18](#_Toc529137981)

[3.2 Research on the application of BN in medical diagnosis system 19](#_Toc529137982)

[3.3 Bayesian filters 23](#_Toc529137985)

[3.4 Bayesian network data mining and knowledge discovery 2](#_Toc529137986)6

[Chapter 4 Summary 29](#_Toc529137987)

[■ 참고문헌 30](#_Toc529137988)

# ■ 서론

In our life there are a lot of uncertainty, people tend to make a common-sense reasoning, and this kind of reasoning is often uncertain. For example, if you see a person walking in with a wet hair, you might think it's raining outside, but you might be wrong; You see a man and a woman in a coffee shop with a kid, you might think they're one family, but you might be wrong. In the work, we also need to carry out scientific and reasonable reasoning, but the problems in the work are generally more complex, and there are many uncertainties, which brings great difficulties for accurate reasoning. Bayesian networks are proposed to address these uncertainties and incompleteness problems. Bayesian network is based on probabilistic inference (probability reasoning is through some process of variable information to obtain other probability information), for the uncertain relationship between these variables provides a way of expression, the combination of graphics and probability knowledge, more intuitive, compact, clear, effective, based on the analysis of uncertainty, and can effectively expression and multi-source information fusion. The bayesian network is able to analyze and make simple decisions based on limited, imperfect, and uncertain information, so the bayesian network is used in:

1. Fault diagnosis: find out the cause of the fault according to the fault diagnosis, carry out real-time monitoring and fault prevention according to the frequent fault or the existing state of the system. For example, fault diagnosis in Microsoft Windows software can help users solve hardware and software problems. Industrial fault Diagnosis (e.g. Auxillary Turbine Diagnosis by general electric), aerospace fault Diagnosis (e.g.Diagnosis of space shuttle propulsion systems jointly developed by NASA and Pockwell)
2. Expert system: provide expert level reasoning, simulate human intelligence, and solve practical problems in the professional field. For example, the application of bayesian networks in medicine.
3. Planning: predict the possibility of various events according to the causal probability reasoning, and get the planning of a project for a given goal.
4. Learning: provide help for learning and help beginners to quickly grasp the cause and effect of events.
5. Classification: using Bayesian network for cluster analysis and classification, it plays an important role in data mining and pattern recognition.

It is widely used in many fields. These successful applications fully demonstrate that Bayesian network technology is a powerful method of uncertainty reasoning. Meanwhile, the establishment of efficient and stable Bayesian network learning algorithm is the key to the application of Bayesian network. For many years, Bayesian network learning and its application have been a hot topic in domestic and foreign research. In order to have a more detailed and comprehensive understanding of Bayesian networks and their applications, here is to analyze the relevant knowledge of Bayesian networks and about the application of Bayesian networks.

Bayesian networks have the following characteristics:

( 1 ) It can show the causal relationship of the event very intuitively. Statistics, often use the regression algorithms usually takes a lot of historical data to establish a mathematical model of the expert's experience knowledge, this method can't fusion cannot explain the causal relationship between variables, life also often can't get enough data to establish model, Bayesian networks can be very good to overcome these defects, in the case of insufficient data, according to expert knowledge model is established.

( 2 ) You can do bidirectional reasoning. The reason can be inferred from the reason or from the result. When the data is provided to any variable, the Bayesian network can update the probability of all other variables in the model. So when you enter a data into a result variable, Bayesian network model will carry out reverse probability reasoning, reasoning the probability of cause variable, such reverse reasoning ability is not able to be achieved by other classical probability reasoning methods.

( 3 ) New data can be used to overturn previous reasoning.

( 4 ) Reasoning can be carried out in the case of incomplete data.

( 5 ) Different types of data can be combined.

( 6 ) All nodes in the Bayesian network model are visible.

■ 관련연구

# Chapter 1 Data mining

The data in all aspects of life are numerous and complex. People do not need to receive complex data comprehensively, but hope to extract useful information from the large amount of complex data. Data mining is a technique to find the law from large amounts of data by analyzing each data. Data mining tasks include correlation analysis, cluster analysis, classification analysis, anomaly analysis, specific group analysis and evolution analysis.

For example, stores have a large number of transaction data every day, from which you can see a lot of consumption records. For example, a consumer who buys a computer can buy relevant parts of a computer, and a consumer who buys baby milk powder can buy other baby products. These data will be reflected in the transaction data. If the data can be analyzed by computer, and then the data rules can be analyzed, it can be more convenient for consumers to choose, and also can provide the profits of the store.

Data mining is to analyze and learn some historical data, draw conclusions and use them for relevant prediction.

Steps of data mining include:

1. Understanding data and data sources

2. A acquire relevant knowledge and technology

3. Integration and inspection of data

4. Erasing error or different data

5. Make models and assumptions

6. A actual data mining

7. Test and verify mining results

8. Interpretation and application

# Chapter 2 Bayesian theory

Bayesian learning is a very effective way of learning, because its simple and practical has been studied by many people. The main research school was formed in the 1950s to 1960s, and after years of research, many applications were obtained, especially in the field of expert system. Bayes' reasoning is a probabilistic reasoning model, and the theoretical basis of Bayesian theory is related to probability theory, which provides good theoretical support for Bayesian theory.

## 2.1Knowledge of probability

The knowledge of probability is the basis of Bayesian theory. Conditional probability is an important concept in probability theory. Conditional probability refers to the probability that event A has occurred under the condition that another event B has occurred. The conditional probability is expressed as P (A|B), which is read as "the probability of A under the condition of B". If there are only two events A and B, then,

https://gss3.bdstatic.com/7Po3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D113/sign=2835f4587c1ed21b7dc92ae49e6fddae/b64543a98226cffcb1b6da18b8014a90f703eaf8.jpg

Edge probability is the probability that one event will occur, independent of the other. The edge probability is obtained by combining the unwanted events in the final result into the full probability of the event in the joint probability (the full probability is obtained by summation of the discrete random variable and the full probability by integration of the continuous random variable). This is called marginalization. The marginal probability of A is P(A), and the marginal probability of B is P(B). In these definitions there is no causal or chronological relationship between A and B. A could happen before B, or it could be the opposite, or both. A may cause B to happen, or it may be the opposite, or there is no causal relationship at all.

British mathematician Thomas Bayes developed to describe the relationship between two conditional probabilities, such as P(A|B) and P(B|A). According to the multiplication rule, we can immediately derive:

P(A exercise B) =P(A)\*P(B|A)=P(B)\*P(A|B).

The above formula can also be reduced to:

P(B|A) = P(A|B)\*P(B)/P(A).

Bayesian rule is about the conditional probability and the marginal probability of random events A and B

https://gss3.bdstatic.com/7Po3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D209/sign=3ea75d6f6659252da7171a040d9a032c/77094b36acaf2edd4492f69c8b1001e939019360.jpg

Where P(A|B) is the probability of A occurring in the case of B.

https://gss3.bdstatic.com/-Po3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D71/sign=addde126f0246b607f0eb075eaf8e24b/48540923dd54564e818c5b4cb5de9c82d0584fdd.jpg

Is a complete event group, i.e

https://gss1.bdstatic.com/9vo3dSag_xI4khGkpoWK1HF6hhy/baike/s%3D227/sign=7a8909cea01ea8d38e227306a00b30cf/dcc451da81cb39dbff8ed923d6160924ab18307c.jpg

In Bayesian rule, each noun has a conventional name:

Pr(A) is A prior probability or an edge probability of A. It's called a priori because it doesn't take into account any B aspect.

Pr(A|B) is known as the conditional probability of A after the occurrence of B, and is also known as the posterior probability of A due to the value of B.

Pr(B|A) is the conditional probability of B after the occurrence of A, and is also known as the posterior probability of B because of the value of A.

Pr(B) is the prior or edge probability of B, normalized constant.

## 2.2 Bayesian networks

Bayesian network consists of a directed acyclic graph (DAG) and a conditional probability table (CPT). College Bayesian networks is represented by a directed acyclic graph of a set of random variables and their conditional dependencies. It is parameterized by a conditional probability distribution. Each node is parameterized by P(node|Pa(node)), which represents the parent node in the network.

The following figure is a simple Bayesian network, whose corresponding full probability formula is:

P (a, b, c) = P (c | a, b) P (b | a) P (a)

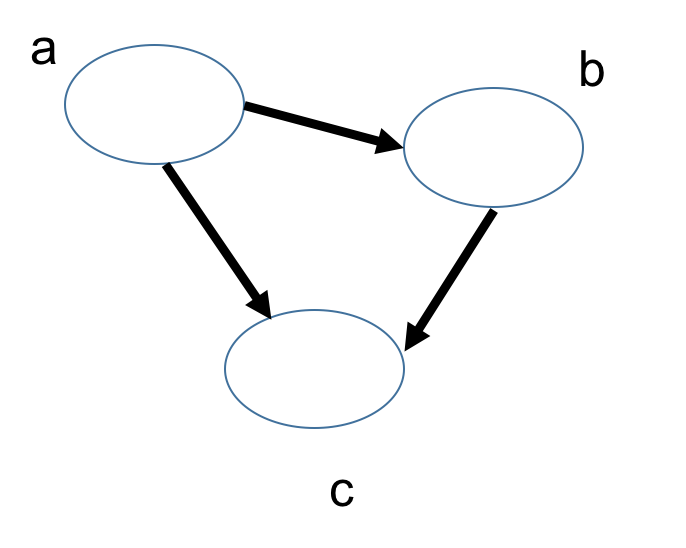


Fig .1 A simple BN

The total probability formula of the complex Bayesian network is:

P (x1, x2, x3, x4, x5, x6, x7) = P (x1) P (x2) P (x3) P (x4 | x1, x2, x3) P (x 5 | x1, x3) P (x6 | x4) P (x7 | x4, x5)

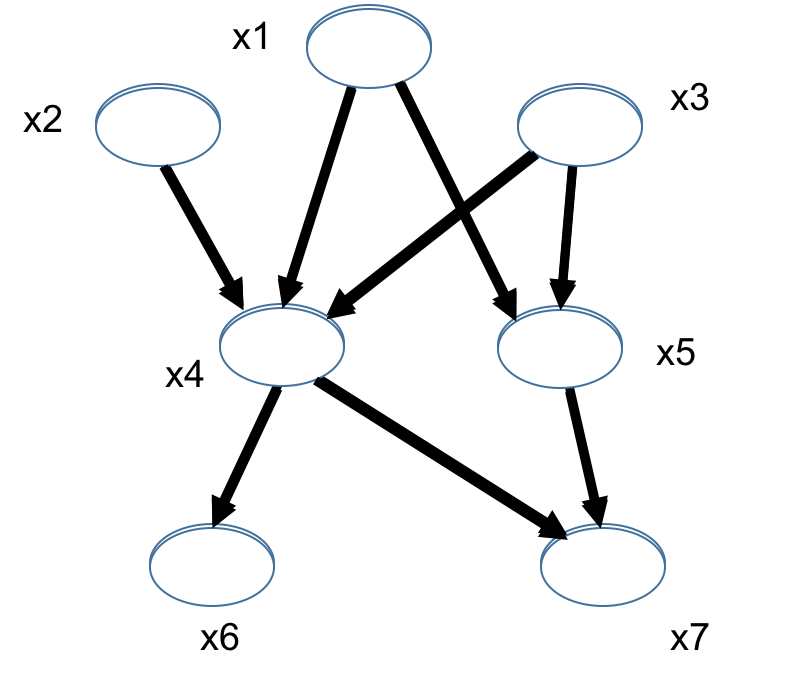


Fig .2 The complex BN

## 2.3 Bayesian network structure

Bayesian network structure form

Form 1: Head-To-Head

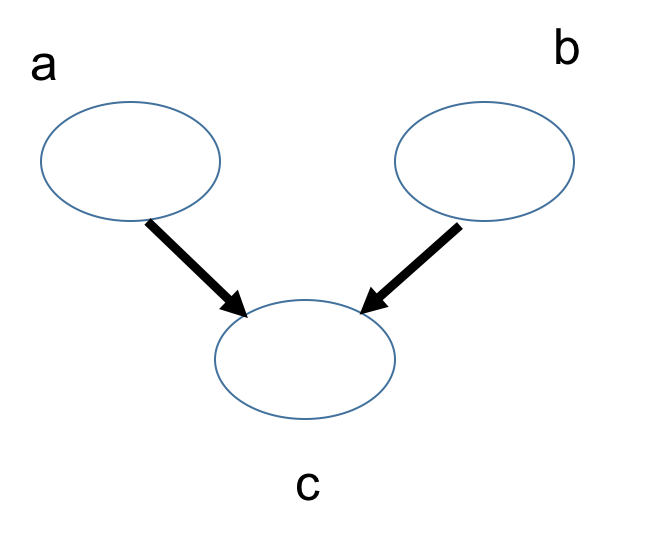


Fig .3 Head-To-Head

P(a,b,c) = P(a)\*P(b)\*P(c|a,b)

P(a,b)= P(a)\* P(b)

In other words, under the unknown condition of c, a and b are blocked and are independent, which is called head-to-head condition independence.

Form 2: Tail-To-Tail

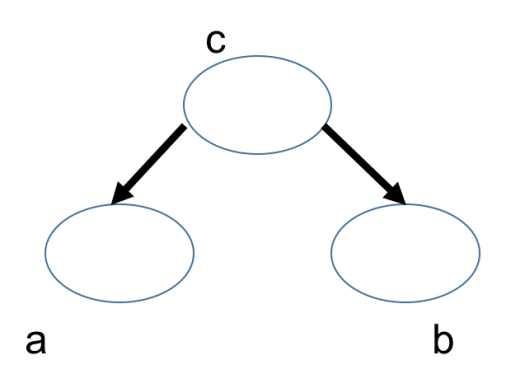


Fig .4 Tail-To-Tai

Consider the case of c unknown, and the case of c known:

When c is unknown, are: P (a, b, c) = P (c) \* P (a | c) \* P (b | c), at this point, can't it is concluded that P (a, b) = P (a) P (b), c is unknown, a and b is not independent.

When c known are: P (a, b | c) = P (a, b, c)/P (c), then P (a, b, c) = P (c) \* P (a | c) \* P (b | c) into the formula, get: P (a, b | c) = P (a, b, c)/P (c) = P (c) \* P (a | c) \* P (b | c)/P (c) = P (a | c) \* P (b | c), is known as c, a and b independent.

Therefore, under the given condition of c, a and b are blocked and are independent, which is called Tail-To-Tail condition independence

Form 3: Head-To-Tail

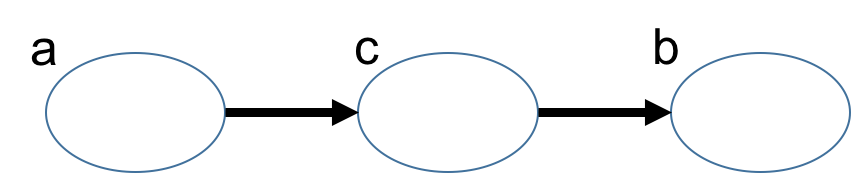
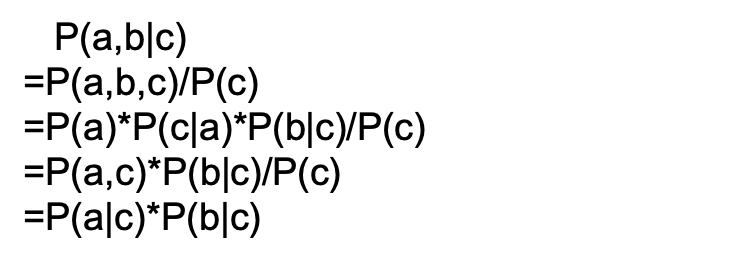


Fig .5 Head-To-Tai

The two cases are known in c and c:

C is unknown, there are: P (a, b, c) = P (a) \* P (c | a) \* P (b | c), but can't launch P (a, b) = P (a) P (b), c is unknown, a and b is not independent.

C is known, there are: P (a, b | c) = P (a, b, c)/P (c), and according to the P (a, c) = P (a) \* P (c | a) = P (c) \* P (a | c), can be obtained in:



## 2.4 Bayesian inference

Bayesian inference is a statistical method used to estimate certain properties of statistics.

It was the application of the Bayes' theorem. Thomas Bayes, a British mathematician, first proposed the theorem in a 1763 paper.

Bayesian inference is quite different from other statistical inference methods. It is based on subjective judgment, that is, you can estimate a value without the need for objective evidence, and then revise it according to the actual result.

Bayesian inference requires a lot of computation, so it has not been widely used for a long time in history. Only after the computer was born did it gain real attention. It has been found that many statistics are impossible to make objective judgments in advance, and the power of large data sets in the Internet age, coupled with high-speed computing power, has provided convenience to verify these statistics and created conditions for applying Bayesian inference.

There are three important reasoning modes in Bayesian networks: causal reasoning (top-down reasoning), diagnostic reasoning (bottom-up reasoning) and defense.

### 2.4.1 Causality reasoning

Let's illustrate the causal reasoning process through an overview. A given patient was a smoker (S), calculating his probability of emphysema (E) P (E|S). S is called reasoning evidence, E is called query node. First, we look for another parent node (C) of E and extend the probability.

P (E|S) =P (E, C|S) +P (E, ~C|S);

That is to say, the probability of emphysema in smokers is not the sum of the probability of emphysema in miners, that is, the probability of emphysema in smokers is the probability of emphysema in miners. Then we use the Bayes theorem:

P (E|S) =P (E|C, S) \*P (C|S) +P (E|~C, S) \*P (\*P);

Formula explanation:

P (E, C|S) = P (E, C, S) /P (S)

= P (E|C, S) \*P (C, S) /P (S) (Bias theorem)

= P (E|C, S) \*P (C|S) (reverse utilization of Bias theorem)

Similarly, the derivation of P (E, ~C|S) can be derived. It is necessary to find conditional probabilities of parental nodes of the expression and re-express joint probabilities (P (E, C | S), P (E, ~C | S). There is no parental relationship between C and S.

P (C|S) =P (C),

P (~C|S) = P (~C).

It can be obtained:

P (E|S) = P (E|S, C) \*P (C) +P (E|~C, S) \*P (S)

If we use the example data in the summary, we have P (E|S) = 0.9\*0.3+0.3\* (1-0.3) =0.48. From this example, it is not difficult to come up with the main operations of this reasoning.

1) According to the joint probability of V of the given evidence and all its parents, the conditional probability of the inquiry node of the given evidence is re-expressed.

2) Return to the probability that all parents are eligible to re express this joint probability.

3) Until all probability values can be obtained from the CPT table, the inference is completed.

### 2.4.2 Diagnostic reasoning

As an example, we calculate the probability P (~C |~E) of "it is not miners who can't emphysema". That is to say, in Bayesian networks, the conditional probability of a parent node is calculated from a child node. That is to say, a cause is deduced from the results, which is called diagnostic reasoning. Using Bayes formula can transform this reasoning into causal reasoning.

P (~C|~E) = P (~E|~C) \*P (~C) /P (~E).

From causal reasoning

P (~E|~C) = P (~E, S|~C) +P (~E, ~S|~C)

= P (~E|S, ~C) \*P (S) +P (~E|~S, ~C) \*P (~S)

= (1-0.3) \*0.4+ (1-0.10) \* (1-0.4) =0.82;

Therefore, we have to:

P (~C|~E) = P (~E|~C) \*P (~C) / P (~E) (Bias formula)

= 0.82\* (1-0.3) / P (~E)

= 0.574/ P (~E)

Alike,

P (C|~E) = P (~E|C) \* P (C) / P (~E)

= 0.34\*0.3/ P (~E)

= 0.102 /P (~E)

Due to the formula of total probability:

P (~C|~E) +P (C|~E) = 1

Generation can be obtained

P (~E) =0.676

Therefore,

P (~C|~E) = 0.849

This reasoning method mainly uses Bayes rules to be converted into causal reasoning.

### 2.4.3 Explain

If our evidence is only ~E (not emphysema), as mentioned above, we can calculate the probability that ~C patients are not coal miners. But if it is also given to S (the patient is not a smoker), then C should also become uncertain. In this case, we say ~ S explains to E, making C become uncertain. Such reasoning uses causal reasoning embedded in a diagnostic reasoning. In this process, the use of Bayes rules is an important step in the process of justification.

### 2.4.4 D separation

There are three such nodes in Bias network: S, L, E. Intuitively, L's knowledge (result) will affect S's knowledge (cause), and S will affect E's knowledge (another result). Therefore, there are many factors that must be considered in computational reasoning, which greatly affects the computational complexity of the algorithm, and may even affect the reliability of the algorithm. But for a given reason, S, L can not tell us more about E. That is to say, for S, L and E are relatively independent, so when calculating the relationship between S and L, we don't need to consider E too much, which will greatly reduce the computational complexity. In this case, we call S D to separate L and E. D separation is an effective way to find conditions independent.

For a given set of nodes, if for each undirected path between the nodes Vi and Vj in the Bayesian network, there is a node Vb on the path, if there are attributes:

1) Vb is in epsilon, and the two arcs on the path take Vb as the tail (that is, the arc starts at Vb).

2) Vb in epsilon, an arc on the path takes Vb as the head, and Vb as the tail.

3) Vb and any of its successors are not in epsilon. The two arcs on the path begin with Vb (that is, the arc ends at Vb).

It is said that Vi and Vj are blocked by Vb nodes.

CONCLUSION: If Vi and Vj are blocked by any node in the set of evidence, then Vi and Vj are separated by set D. The conditions of node Vi and Vj are independent of the given set of evidence, that is to say, E.

P (Vi|Vj, e) = P (Vi| E)

P (Vj|Vi, e) = P (Vj| E)

It is expressed as: I (Vi, Vj| E) or I (Vj, Vi| E).

Undirected path: DAG graph is directed graph, so the path should also be directed path. The undirected path referred to here is the path without considering the directivity in DAG graph.

Conditional independence: If one of the above three attributes exists, the node Vi and Vj conditions are independent of the given set of nodes.

Blocking: Given the set of evidence, when any of the above conditions is satisfied, Vb blocks the corresponding path.

D separation: if all paths between Vi and Vj are blocked, it is called evidence set E. D can separate Vi and Vj.

Taking the above medical diagnosis as an example, for the sake of simplicity, the set of evidence sets is chosen as a single node set. For a given node S, node E blocks the path between node C and node L, so C and L are conditionally independent and I (C, L | S) holds.

For a given node E, there is no blocking node between S and L. Therefore, S and L are not conditionally independent. Even if D separation is used, generally speaking, probabilistic inference is still a NP problem in Bayesian networks. However, some simplification can be used in an important network classification called Polytree. A Polytree network is a DAG. Between any two nodes of the DAG, there is only one path along each direction of the arc. The essence of D separation is to find the conditional independence semantics in Bayesian network, so as to simplify reasoning computation.

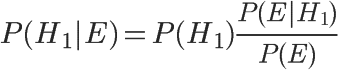
### 2.4.5 Fruit fructose problem Example

Two identical bowls, bowl one with 30 fruit drops and 10 chocolate drops, and bowl two with 20 each. Now select a bowl at random and pick out a sugar from it.

Let's say, H1 is bowl one, H2 is bowl two. Since these two bowls are identical, P(H1)=P(H2), that is, the probability that these two bowls are selected is equal before the fruit drop is taken out. So P(H1)=0.5, let's call this a prior probability, which means before we did the experiment, the probability from bowl one was 0.5.

And let's say that E is fructose, so the question becomes what is the probability of coming from bowl one given E, that is, P(H1|E). We call this probability a "posterior probability," a correction of P(H1) after the E event.

According to the conditional probability formula, we get:



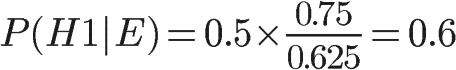
We know that P(H1) is equal to 0.5, and P(E|H1) is the probability of taking the fruit drop out of bowl 1, which is equal to 0.75, so we can get the answer by finding P(E). According to the full probability formula:

http://chart.googleapis.com/chart?cht=tx&chl=P(E)%3DP(E%7CH_%7B1%7D)P(H_%7B1%7D)%2BP(E%7CH_%7B2%7D)P(H_%7B2%7D)&chs=40

So:

http://chart.googleapis.com/chart?cht=tx&chl=P(E)%3D0.75%5Ctimes%200.5%2B0.5%5Ctimes%200.5%3D0.625&chs=40

Substitute the Numbers into the original equation, and get:



## 2.5 Application: Spell check

Usually, when you inadvertently enter the word in a nonexistent, the search engine will prompt you to enter a certain word. For example, when you type "Julw" into Google, the system will guess your intention: to search "July," which is called spell checking. Google's spell checker is based on Bayesian methods, according to an article written by an employee at Google.

When the user enters a word, it may be spelled correctly or incorrectly. If the correct spelling is remembered as c (for correct) and the incorrect spelling as w (for wrong), then the "spelling checker" has to do: in the case of w, try to infer c. In other words: you know w, and then in a number of alternatives, find the c that is most likely, which is the maximum.

According to Bayes’ theorem：

https://img-blog.csdn.net/20141112230845578

Since all the alternative c's correspond to the same w, their P(w) is the same, so as long as you maximize

https://img-blog.csdn.net/20141112231004421

P(c) denotes the occurrence "probability" of a certain correct word, which can be replaced by "frequency". If we have a large enough text library, the frequency of each word in the text library is equal to the probability of its occurrence. The higher the frequency of a word, the greater the P(c). For example, when typing the wrong word "Julw," the system tends to guess that the word you might want to type is "July," not "Jult," because "July" is more common.

P(w|c) is the probability of a misspelling w in the case of an attempt to spell c. To simplify the problem, assume that the closer the two words are in glyph, the more likely they are to be misspelled, and the larger the P(w|c). For example, a letter difference is more likely to occur than a two letter difference. You want to spell the word July, so the possibility of misspelling the word Julw (one letter difference) is higher than the word Jullw (two letters difference).

So, compare how often all words with similar spellings appear in the text library, and pick the one that appears most often, the one that the user most wants to type.

The main theoretical basis of Bayesian model is Bayesian theory. Its main advantages are:

1. The variation of the sample has little influence on the model, because the predicted result of each sample is the probability belonging to each class, which is then attributed to the maximum probability category. Therefore, the classification result will not change as long as the rank of the sequence of the probability belonging to the category remains unchanged.

2. A Bayesian learned quickly because it didn't have to repeat operations on large Numbers of samples.

3. In case of uncertainty, the Bayesian approach can make the assumption of uncertainty, which may be more realistic.

Disadvantages of Bayesian algorithm include:

1. In the Bayesian model, a lot of initial knowledge of probability is needed, and if the knowledge of probability cannot be mastered accurately, there will be many problems.
2. In some specific cases, the calculation is relatively cumbersome, especially for Bayesian networks.

## 2.6 Bayesian network learning

The Bayesian network constructed according to user's prior knowledge is called a prior Bayesian network. The Bayesian network obtained by combining the prior Bayesian network with data is called a posterior Bayesian network. The process of obtaining a posterior Bayesian network from a prior Bayesian network is called Bayesian network learning. Bayesian network can continue to learn, and the posterior Bayesian network obtained from the last learning can be transformed into the prior Bayesian network of the next learning. Users can adjust the prior Bayesian network before learning, so that the new Bayesian network can better reflect the knowledge contained in the data, as shown in the figure.



Fig .6 BN reflect the knowledge contained in the data

Bayesian network-based learning includes parameter learning and structure learning. At the same time, according to the different nature of sample data, each part includes two aspects: complete case data and incomplete case data. Parametric learning methods are mainly based on classical statistics and Bayesian statistics-based conditional probability table (CPT). Structural learning methods are mainly based on Bayesian statistical measures and coding theory based measurement methods. Structural learning is introduced below.

In Bayesian networks, a random variable Sh is defined to denote that database D is a random sample hypothesis from network structure S, and a prior probability distribution P (Sh) is given to denote the uncertainty of network structure. Then a posterior probability distribution P (Sh | D) is calculated. According to the Bayesian theorem, there are



Among them: P (D) is a normalized constant independent of structural learning, and P (D| Sh) is structural likelihood. Therefore, to determine the posterior distribution of network structure only needs to calculate the structural likelihood of data for each possible structure.

On the premise of unconstrained polynomial distribution, parameter independence, Dirichlet priori and data integrity, the structural likelihood of data is exactly the product of the structural likelihood of each (i, j) pair, i.e.



The formula was first given by Cooper and Herskovits in 1992. In general, the number of possible network structures with n variables is larger than that with n as an exponential function. It is difficult to exclude these assumptions one by one. There are two ways to deal with this problem: "model selection" and "selective model averaging". The former chooses a "good" model from all possible models (structural assumptions) and regards it as the correct model; the latter chooses a reasonable number of "good" models from all possible models and believes that these models represent all situations.

■ 제안 작품 소개

# Chapter 3 Bayesian network application

## 3.1 Application of BN in fault diagnosis

1995 years Microsoft company will BN applied to fault diagnosis for the printer [1-2], researchers put forward a certain inquiries - test methods - repair sequence framework. Through continuous calculation under different steps may be required from the effect of maintenance scheme to determine the optimal maintenance path. In here to use, BN used to update a given after the observational evidence of the probability of each fault source, at the same time, the costs will repair the fault diagnosis process source and source of trouble may occur the probability of comprehensive consideration is given to the diagnosis of optimization scheme.

In the figure, a simplified BN model is used for printer fault diagnosis. Then HP [3] improved the diagnosis method and extended the one-step forward analysis proposed by Microsoft to two-step forward analysis to solve the problem of multiple faults at the same time.

Commonly used automobile fault diagnosis based on fault tree analysis method. In 2008, Huang Y, McMurran R, Dhadyalla g. [4] successfully apply the BN in the off-line fault diagnosis of vehicle, proposed a three layer structure of BN diagnosis model to replace the traditional method of fault tree structure. This structure can solve the problem of the parent node too much, make each node's parent is not exceed four.

Figure is the structure of BN diagnosis model framework. In addition, the author adopts the object-oriented BN to fault warning diagnosis model is established to realize the multiple fault diagnosis. At the same time routine of BN diagnosis model usually only by experts in the field of forward conditional probability (from the source to the failure symptom), Scheiterer etc. [5] this paper presents a new method of BN diagnosis model is established, the method can make use of experts to provide two-way conditional probability to establish the diagnosis model, makes the model more complete and accurate.

BN is also used for fault diagnosis of engine rotor [6], power grid [7], vehicle power system [8], hydraulic pump [9], power transformer [10], solar power plant [11], mobile communication network [12] and manufacturing process [13].

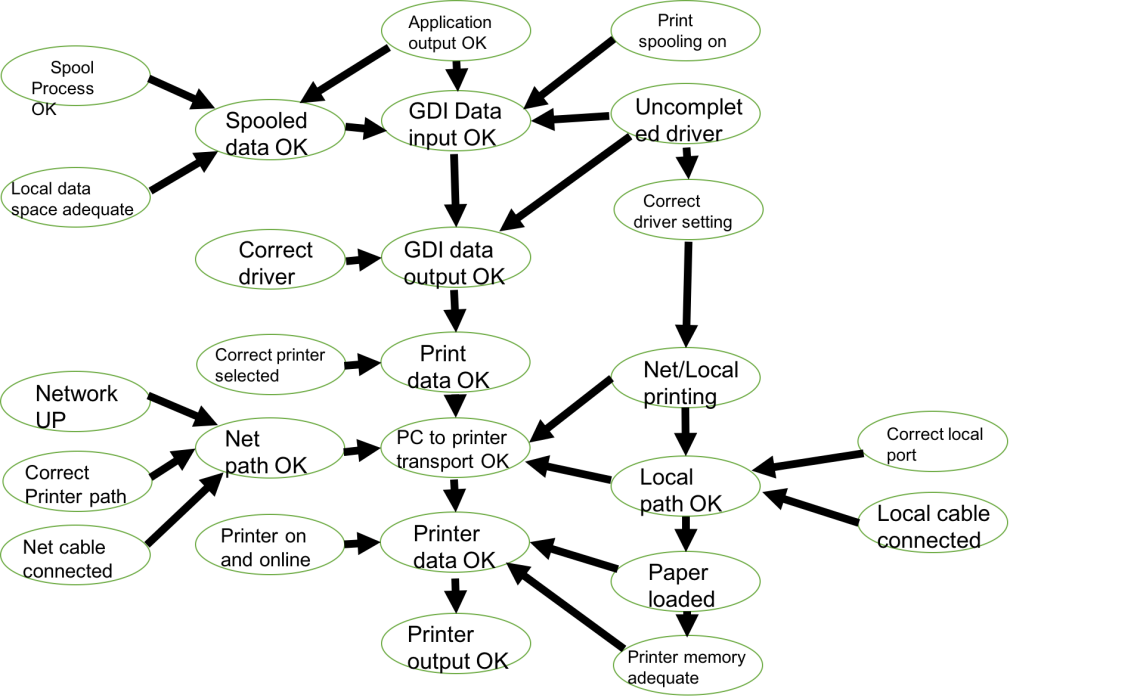


Fig .7 A simplified BN diagnosis model for printer fault

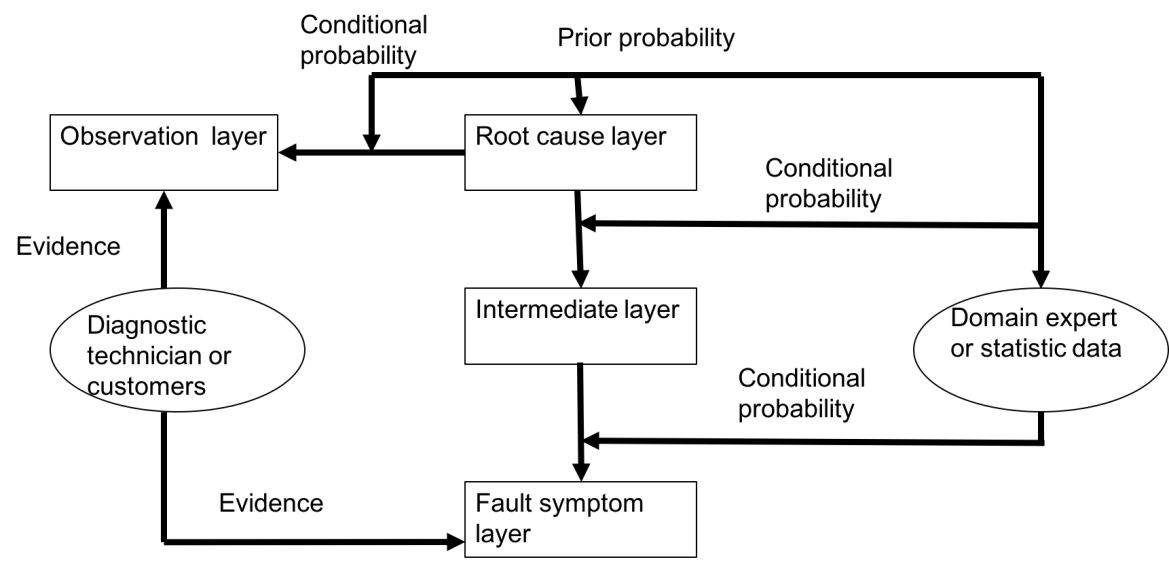


Fig .8 Framwork of BN model for automotive diagnosis

## 3.2 Research on the application of BN in medical diagnosis system

With the rapid development of The Times, various aspects of life are developing towards intelligentization. In the aspect of medical treatment, various expert systems gradually appear in people's vision. Because it has the advantages of improving the accuracy of diagnosis, simplifying the process of diagnosis and reducing the cost of diagnosis, it gradually becomes popular in life. With the improvement of people's living standards, the trend of diseases is more and more serious, such as cardiovascular and cerebrovascular diseases. Data show that cardiovascular and cerebrovascular diseases have become the number one threat to people's physical health, and there is a trend of youth. Therefore, the diagnosis of cardiovascular and cerebrovascular diseases has become a key issue. The diagnosis of cardiovascular and cerebrovascular diseases requires many examinations, such as blood pressure, blood lipids, blood sugar, etc.

How to use these tests to determine whether patients with cardio-cerebral vascular disease, or there are some ways to through these detected index to help judging experts is a realistic problem, or only examined a few of these patients, according to test data to determine whether the other patients sick, is also a realistic problem. Therefore, the diagnosis of such diseases has become an important research topic. The diagnosis of these diseases is a long process, and many detection technologies are needed to make the diagnosis. Therefore, the development of an intelligent system can help the experts to make the diagnosis, speed up the diagnosis and reduce the cost of diagnosis.

At present, there are the following problems in modern medical treatment of cardiovascular and cerebrovascular diseases:

1. As cardiovascular and cerebrovascular diseases are affected by many factors, and the influence among them is very complicated, the diagnosis of this disease has also become a problem of comprehensive assessment. People also hope to apply modern technology to medical diagnosis to help experts to diagnose diseases.

2. At the same time of cardio-cerebrovascular diseases diagnosis, patients need several checks and inspections by the indicators need to spend time and money is also a lot of, so how to use the less cost index of cardio-cerebrovascular disease diagnosis is a very valuable problem, so can reduce the cost of inspection need time and check.

3. During the diagnosis process, sometimes only one part of the indicators is checked, while no other indicators are checked. If the diagnosis is made without the results of these examinations, it is also a problem worth studying.

The main functional modules include: medical data import module, algorithm parameter setting module, training module, prediction module, warning module.

The most important thing in the whole system is:

1. Training system: be used to the training data to train, to generate the model used to predict

2. Prediction system: with training model to forecast the new instance, has forecast results

3. Warning system: if the forecast instance judged to be normal, but likely the person's health is not good, belong to the sub-health, if do not pay attention to, may enter the unhealthy condition, so it is necessary to carry out these examples suggest, it is necessary to pay attention to your health.

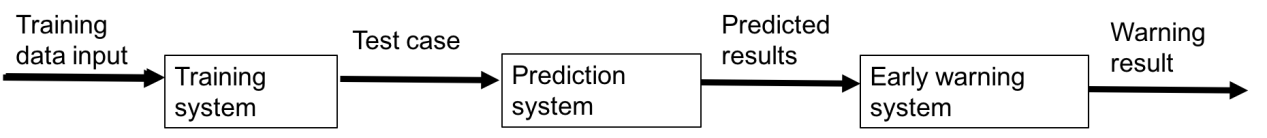
****

Fig .9 Medical diagnostic system process

The main modules are:

Data processing: the main function is to read the training data for later training, and can clear the data and reload the data

Graph analysis: it is used to display the classification results, to clearly show the classification accuracy in various situations, for the purpose of quantitative analysis

Warning display: it is used to analyze each result, analyze its risk degree, and give different warning information according to its analysis

System structure of medical diagnosis and warning system based on Bayesian,

Set algorithm parameters

(1) Cross-validation

Cross-validation is a key technology used in data mining to verify algorithm performance. Generally, it is explained as several fold cross validation, such as 10 fold cross validation. Existing data sets are divided into 10 equal parts, each part is taken as its detection data, the remaining 9 parts as training data for prediction, and then the average of each part's prediction results, that is, the prediction results of the model are obtained. Set the folded number across to test the performance of the algorithm under different folded Numbers, thus further analyzing. In addition, another parameter that affects the algorithm is the number of attributes. You can select different attributes to perform the operation of the algorithm.

(2) There are two ways to set the folding number. One is to set a single folding number for calculation, and the other is to set the folding number to increase. In each case, an algorithm performance test is conducted. In the same way, there are two ways to set the number of attributes. One is to set the single number of attributes, and the other is to use the increasing number of attributes.

Early warning mechanism

For an example, when you use Bayesian to make a prediction, you get two probabilities, one normal and one abnormal. Assuming that the two probabilities are p+, p-, if the probability that an instance is judged to be normal is much greater than the probability of being abnormal, then the person is largely determined to be normal. If a person is considered normal, but are considered normal probability not much larger than the probability of abnormal, it shows the person's health is not very obvious, that is to say it on the verge of health is not healthy, if you don't pay attention to your health, it is possible to be processed is not healthy, so an example is given to the difference of healthy and unhealthy can overall reaction to the people's health, the public can be expressed as

**△s= p+ - p-**

According to the formula, the larger the size of aspiration s indicates that people are healthier, and the smaller size of aspiration s indicates that people are in poorer health. For a person who is judged to be normal, it must be greater than 0. For someone who is abnormal, it must be less than 0. According to the size of allowance s, it is divided into four grades. All the cases considered here are those that are judged to be normal. That is to say, the cases that are judged to be normal are further divided into four grades, excellent, good, average and pass (basically normal, on the edge of normal and abnormal). In the early warning system, four different categories of people are displayed separately.

Performance testing under fixed folding

The experimental results show that the performance of the algorithm is affected by the combination of different attributes. Performance testing with fixed number of attributes. The experimental results show that the combination of different discount values affects the performance of the algorithm.

Bayesian network is applied in medical diagnosis to develop an expert diagnosis system based on Bayesian. Its functions include:

(1) Make a diagnosis of the new case to determine whether it is normal or abnormal. The judgment is based on the probability that the instance belongs to each class. The structural framework and system process of the system are given, and the parameter Settings of folding number and attributes are given (there are two factors influencing the algorithm, one is the influence of cross folding number, the other is the influence of the number of attributes. The performance of the algorithm is different under the influence of different folding number and different number of attributes. Such expert system can help experts make diagnosis and save time for diagnosis.

( 2 ) A warning mechanism is set up for each instance that needs to be diagnosed. In other words, although this instance is diagnosed as normal, the probability of belonging to each class is not large, which means that this instance is at the boundary between normal and abnormal, so it needs to be warned. The higher the probability of being normal, the better the health. On the contrary, the worse their health. Based on Bayesian predicted results, patients' health was divided and health warnings were given to some people whose health was hovering around health and unfitness to alert them to their health. So you have a clear sense of your health level, not just giving health and unfitness, doctors can also adopt different medical methods according to the status, and treat patients with targeted therapy.

## 3.3 Bayesian filters

Spam is a persistent headache that affects all Internet users. Correctly identifying spam is a tricky task. Traditional spam filtering methods mainly include "key words" and "check code". The former is filtered by specific words; the latter is a check code that computes the message text and compares it with known spam. They are not very effective and easy to avoid.

In 2002, Paul Graham proposed using "Bayesian inference" to filter spam. The effect, he said, was "unbelievably good." A thousand spam emails can filter out 995 and not one misjudgment.

In addition, this filter also has a self-learning function, will receive new emails, and constantly adjust. The more spam you receive, the more accurate it is.

Bayesian filter is a statistical filter, based on the existing statistical results. So, you must provide two pre-identified groups of mail, one normal, the other spam.

Use these two sets of emails to "train" the filters. The larger the two groups of emails, the better the training. Paul Graham used a mail-scale of 4,000 for normal and 4,000 for spam.

The "training" process is simple. First, parse all the messages and extract each word. Then, calculate how often each word appears in both normal and spam messages. For example, if we assume that the word advertising, of the 4,000 spam messages, contains 200 of them, then it occurs 5% of the time. Of the 4,000 normal emails, only two contained the word, and the frequency was 0.05%. (if a word only appears in spam, Paul Graham assumes that it appears in normal mail 1% of the time, and vice versa. I'm doing this to avoid the probability of 0. As the number of messages increases, the calculations are automatically adjusted. With this preliminary statistical result, the filter can be put into use. Suppose you receive a new email. Before the statistical analysis, we assumed that it was spam with a 50% probability. (Research has shown that 80% of E-mail received by users is spam. However, it is still assumed that the "prior probability" of spam is 50%.

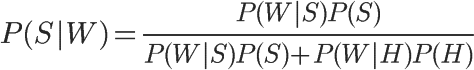
Use S for spam and H for healthy. So the prior probabilities of P(S) and P(H) are 50%.

../论文.png

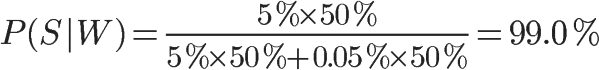
Then, analyze this email and found that it contained the word "advertisement". What is the probability that this email is spam?

We use the word W for "advertisement", and the question becomes how to calculate the value of P(S|W), which is the probability of spam (S) under the condition that a word (W) already exists.

And we can write it out immediately from our conditional probability formula：



In the formula, P(W|S) and P(W|H) mean the probability that the word occurs in spam and normal mail respectively. These two values can be obtained from the historical database. For the word "advertisement", it is assumed that they are equal to 5% and 0.05% respectively. In addition, the values of P(S) and P(H) are equal to 50%. So, immediately we can compute the value of P(S|W) :



So the probability that this new message is spam is 99%. This shows that the word "advertisement" has a strong ability of deduction, which increases the "prior probability" of 50% to 99% of "posterior probability".

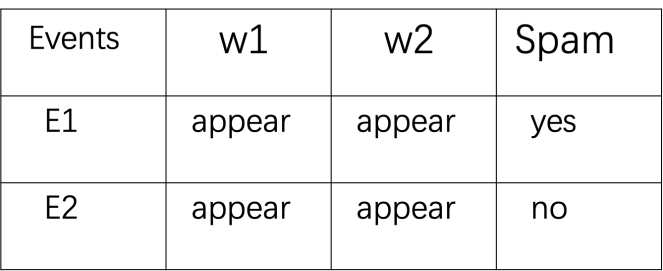
Can you conclude from the above step that this new email is spam?

And can't. Because an email contains many words, some words (like "AD") say it's spam, others say it's not. How do you know which word is right?

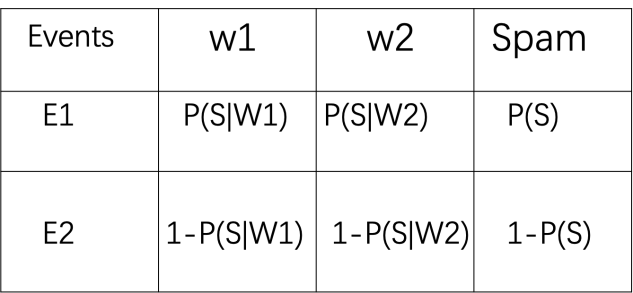
What Paul Graham did was to pick the top 15 words in the letter, the highest P(S|W), and calculate their combined probability. (if some words appear for the first time and you cannot calculate P(S|W), Graham assumes that this value is equal to 0.4. Because spam tends to be a fixed word, if you've never seen a word, it's probably a normal word.

Joint probability refers to the probability of another event in the case of multiple events. For example, if you know that W1 and W2 are two different words, and they both appear in an E-mail, then the probability that the E-mail is spam is the joint probability.

In cases where W1 and W2 are known, there are only two outcomes: spam (event E1) or normal mail (event E2).



The probability of W1, W2 and spam:

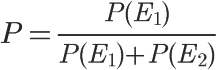


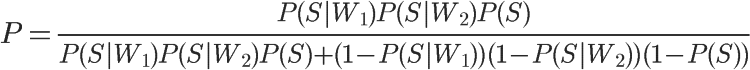
If all events are assumed to be independent events, then P(E1) and P(E2) can be calculated:

http://chart.googleapis.com/chart?cht=tx&chl=P(E_%7B1%7D)%3DP(S%7CW_%7B1%7D)P(S%7CW_%7B2%7D)P(S)&chs=40

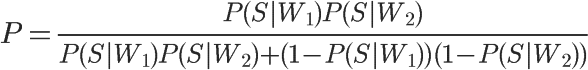
http://chart.googleapis.com/chart?cht=tx&chl=P(E_%7B2%7D)%3D(1-P(S%7CW_%7B1%7D))(1-P(S%7CW_%7B2%7D))(1-P(S))&chs=40

In the case that W1 and W2 have occurred, the probability of spam is equal to:

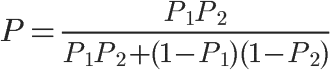




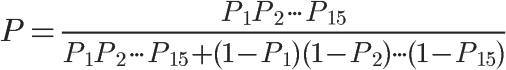
You plug in P of S equal to 0.5 and get:



If I write P(S|W1) as P1, and P(S|W2) as P2, the formula becomes:



Extending the formula to 15 words, the final probability calculation formula is obtained:



If an E-mail is spam, use this formula to calculate it. And then we need a threshold for comparison. The threshold value of Paul Graham is 0.9, and the probability is greater than 0.9, indicating that 15 words are jointly determined, and more than 90% of this email is likely to be spam; A probability less than 0.9 indicates normal mail.

With this formula, a normal letter will not be deemed spam even if the word “advertising “appears.

## 3.4 Bayesian network data mining and knowledge discovery

The following is an example of data mining and knowledge discovery using Bayesian networks, according to 10,318 senior students from Washington High School. Each student uses the following variables and their corresponding states to describe:

Gender (SEX): male and female;

Social economic status (SES): low, middle, lower, middle and upper.

Intelligence quotient (IQ): low, middle, lower, middle and upper.

Parents' encouragement (PE): low and high;

CP: Yes or no.

The goal is to find out the factors that affect high school students' intention to go to university from the data. The data has been compiled into the format shown in Table 1. Each data in Table 1 represents the number of combinations of statistics for 5 variables. For example, the first data indicates that the number of people in this combination is 4 (SEX = male, SES = low, IQ = low, PE = low, CP = yes), and the second data indicates that the number of people in this combination is 349 (SEX = male, SES = low, IQ = low, PE = low, CP = no). The subsequent data represent the number of people who successively rotate each variable for possible state statistics. Variables rotate from right to left, and states rotate according to the state order of the variables listed above.

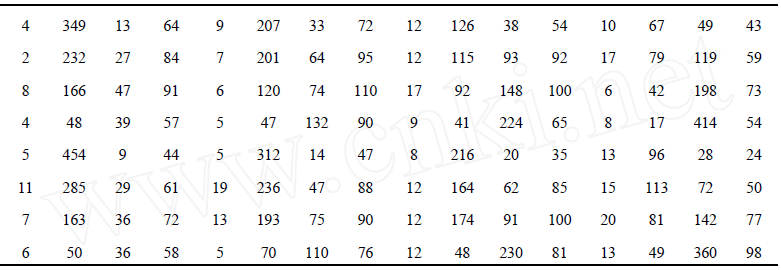


Fig .10 Statistics of persons in various situations (person)

Assuming that there are no hidden variables, we use the equivalent sample with capacity of 5 and a priori network with uniform distribution of P (x | Sh). After excluding the network structures of SEX and SES with parent nodes and CP with child nodes, it is assumed that all other network structures are equally possible. Because the data set is complete, the posterior probabilities of the network structure can be calculated by using the formula. Through exhaustive search of all network structures, we find one of the most possible network structures, such as graphs.

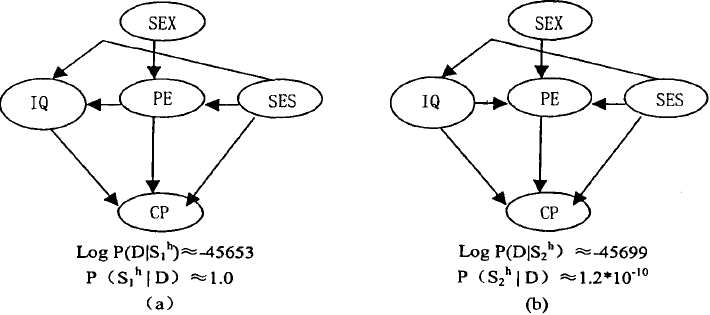


Fig .11 The possible network structures

From figure, we find that socioeconomic status (SES) has a direct impact on intelligence quotient (IQ). In order to verify this result, a new model is considered, that is, the direct impact of social and economic status (SES) on IQ in the model of figure is replaced by a hidden variable (H) pointing to SES and IQ. In addition, considering such a model, hidden variables point to SES, IQ and PE, and 2, 1 and 0 are removed from the SES-PE and PE-IQ links respectively. For each structure, the state number of hidden variables is changed from 2 to 6.

The posterior probabilities of these models are computed using the Laplace approximation Cheese and Stutz variant. In order to find the maximum posteriori to form Hs, EM algorithm is used and the maximum local maximum is obtained in 100 runs of HS with different random initialization. These models have Max posteriori probability graphs The probability of this model is 2 x 1010 times higher than that of the best model without hidden variables. Assuming that reasonable models are not neglected, there is strong evidence that there is a hidden variable affecting SES (socioeconomic status) and IQ (intelligence quotient).

■ 결론 및 소감

# Chapter 4 Summary

Bayesian network has been widely used since its emergence. It is necessary to recognize and learn Bayesian networks and extend them. This paper is only a preliminary understanding of Bayesian network, mainly about the application of Bayesian network in medical diagnosis. Among them, cardiovascular and cerebrovascular diseases have a great impact on people today. The diagnosis of cardiovascular and cerebrovascular diseases is also a complicated process, which requires a large number of examinations and costs. At the same time, it's a waste of time for patients and doctors. In addition, the general diagnosis does not provide detailed information about the patient's condition, such as the classification of the patient's health status, which can better inform the patient's physical condition.

In terms of medical diagnostic system, the following applications can be made according to relevant knowledge of Bayesian network:

1. Using Bayesian method in data mining to diagnose the health status of patients, fully mining useful information and helping experts to make diagnosis can improve the efficiency and accuracy of diagnosis. In the case of incomplete data, the cost of patients can also be judged by existing expert diagnostic information.

2. According to Bayes’ predicted results, the patient's health status level, which allows the patient to be more aware of their own health, and some healthy and unhealthy walking health warning effects, reminds the patient to take care of your health, at the same time that the doctor can receive treatment more accurately.

There are many other aspects of the Bayesian network that need to be discovered together. Bayesian network has many advantages, the use of Bayesian networks for events or attributes with uncertain relationship between modeling and reasoning in medical diagnosis, natural language understanding, fault diagnosis, heuristic search, image interpretation and target recognition and uncertainty reasoning and predict the many successful applications. These applications can be roughly divided into the establishment of system models to assist decision-making, feature fusion and classification of data analysis. Therefore, Bayesian networks have great research and application prospects in the fields of medicine, economy and network.

At the same time, Bayesian networks also have big problems in their wide application, and there are also many deficiencies in Bayesian networks. Only we keep learning and improving, the bayesian network will become more and more mature and widely used to bringing convenience to our life.

■ 참고문헌

[1].Breese.J,Heckerman D. Decision-theoretic trouble-shooting:a framework for repair and experiment[C]//Proceedings of the 12 Conference on Uncertainty in Artificial Intelligence. Porland, USA: Morgon Kaufmann Publisher Inc,1996:124-132

[2].Heckerman D, Breese J S, Rommelse K, Decision-the-oretic trouble shooting [J] , Communication of the ACM,1995,38(3):49-57

[3]..JensenF V, Kjerulff U, Kristiansen B,et al. The SACSO methodology for troubleshooting complex systems[J]. Journal of Aritificial Intelligence for Engineering Design, Analysis and Manufacturing,2000,15(4):321-333

[4]. Huang Y， McMurran R，Dhadyalla G. Probability based vehicle fautt diagnosis： Bayesian network method [J]. Journal of Intelligent Manufacture， 2008，19: 301-311.

[5]. Scheiterer R S，Obradovic D，Tresp V. Tailored-to-fit bayesian network modelling of expert diagnostic knowl­edge[J]. Journal of VLSI Signal Processing, 2007,49 ： 301-316.

[6]. Xu B G. Intelligent fault inference for rotating flexible rotors using Bayesian belief network[J]. ExpertSys- tems with Applications, 2012,39(1) ： 816-822.

[7]. Zhu Yongli，Wang Yan. Power system fault diagnosis based on Bayesian network[J]. Electric Power Auto­mation Equipment，2007,27(7) ；33-36.

[8] Cheng Yanwei，Xie Yongcheng，Li Guangsheng，et af Fault diagnosis method of vehicle power system based on Bayesian network [J]. Computer Engineering， 2011，37(23):251-253.

[9].Jiang Wanglu， Liu Siyuan. Fault diagnosis approach study of Bayesian networks based on multi-characteris­tic information fusion[J]. China Mechanical Engineer­ing， 2010,21(^8)：941-945. d)

[10]. Zhao Wengqing， Zhu Yongli， Wang Xiaohui. Combi­natorial Bayes network in fault diagnosis of power transformer [ J ] Electric Power Automation Equipment，2009，29(11) :6-9.

[11].Coleman A，ZalewskiJ. Intelligent fautt detection and diagnostics in solar plants[C] ^ Proceedings of the 2011 IEEE 6th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Tech­nology and Applications ( IDAACS 2011). [ S. l.]: IEEE，2011:948-953.

[12]. Chan A，McNaughtK R. Using Bayesian networks to improve fautt diagnosis during manufacturing tests of mobile telephone infrastructure[J]. Journal of the Op­erational Research Society，2008，59(4) :423-430.

[13]. Liu Y， Jin S. Application of Bayesian networks for diagnostics in the assembly process by considering small measurement data sets[J]. International Jour­nal of Advanced Manufacturing Technology， 2012， 2:1-9.

[14].Arroyo G, Sucar L，Villavicencio A. Probabilistic tem­poral reasoning and its application to fossil power plant operation [J ]. Expert Systems with Applications， 1998,5:317-324.

[15].KangCW，GolayMW.ABayesianbeliefnetwork- based advisory system for operational availability focused diagnosis of complex nuclear power systems [J]. Expert Systems with Applications，1 99 9，17: 21-32.

[16]. Torres-Toledano J G，Sucar L E. Bayesian networks for reliability analysis of complex systems lecture notes in computer science[C] ^Proceedings of the 6th Ibero- American Conference on AI: Progress in Artificial In­telligence. London: [s. n. ]，1998: 195-206.

[17]. Chen changjun, Huang bo, CAI zhihua, application of cost sensitive support vector machine in medical diagnosis, microcomputer information, 2008,15 (27) : 504-509

[18].Wang xiaoju, Li yonghua, application of rough set theory in medical diagnosis system, modern computer (professional edition), 2008,9 (6) : 106-110

[19].Cai qiong,Yu yi, Application of classifier model based on genetic algorithm in medical diagnosis, software guide, 2008,11 (03) : 179-184

[20].Zhang meng,Liang zhen,Zhu siqing，Application of data mining based on association rules in medical diagnosis, shandong science, 2008,3 (01) : 79-84

[21].Li deyun,Liu guipin,Liu siqiang，Design and implementation of a medical diagnostic expert system, journal of ningbo institute of technology, 2007,10 (05) : 225-230

[22].Chen bingcai,Zhao hao,Jiang shizhong，Medical diagnostic expert system based on bayesian nervous system, medical information, 2007,11 (10) : 31-35

[23].Yang bin,Wan shengchun, Application of data mining in fault diagnosis of large medical equipment, 2007,07 (10) : 107-201.

[24].Lin hepin,Guan renchu,Wang yan，Computer engineering, 2007,19 (16) : 621-626

[25].Xu ning，Medical diagnostic technology research based on case-based reasoning, computer knowledge and technology (academic exchange), 2007,9 (01) : 331-335

[26].Xu ning，Computer and information technology. 2006,10 (09) : 116-121

[27].Li xiaoyi,Xu zhaokang，Application of association rule mining in medical diagnosis, journal of liaodong normal university (natural science edition), 2006,11 (02) : 441-447

[28].Yang liping, Hu junjie，Application of fuzzy cognitive map in concomitant medical diagnostic system, computer engineering and application, 2005,22 (07) : 771-776

[29]. Fan Hong, Application of bayes in medical diagnosis system

[30]. HUANG Ying-ping, A review of the development and application of bayesian networks

[31]. Xu dingjie,Zhou yuandong,Sang yuli,Wei jinchen，Journal of Chinese medicine, 2005,07 (05) : 77-82

[32]. Zhang siqi,Zhou shuwen,Gong zhiguo,Dong mingchui，Data preprocessing, control engineering, 2005,09 (01) : 213-216

[33]. Chen yong,Chen weiguo, Medical diagnostic expert system research and application, medical device information, 2005,04 (12) : 211-214

[34]. Peter Norvig, director of research and development at Google, explains how spell checking works: http://norvig.com/spell-correct.html;