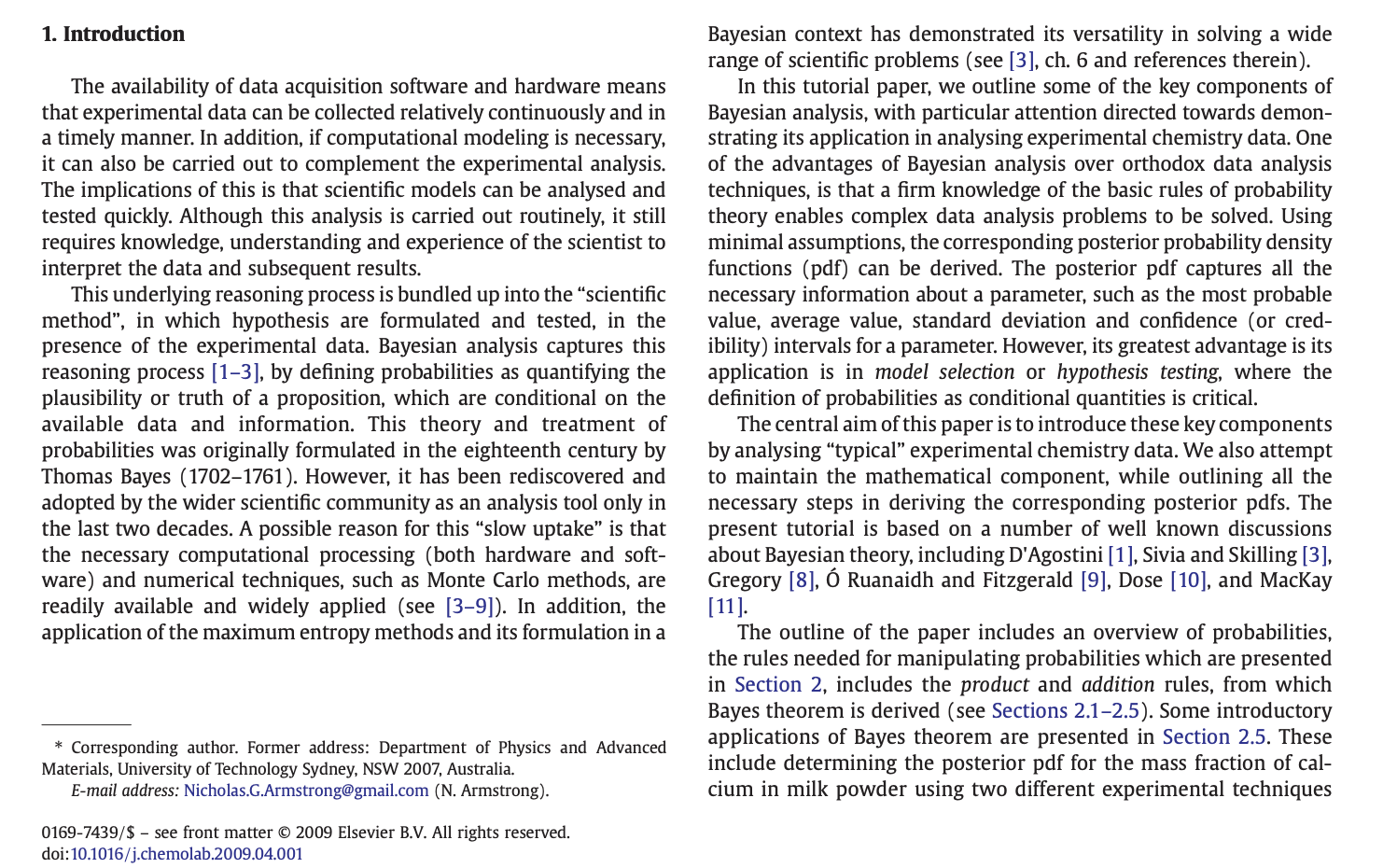
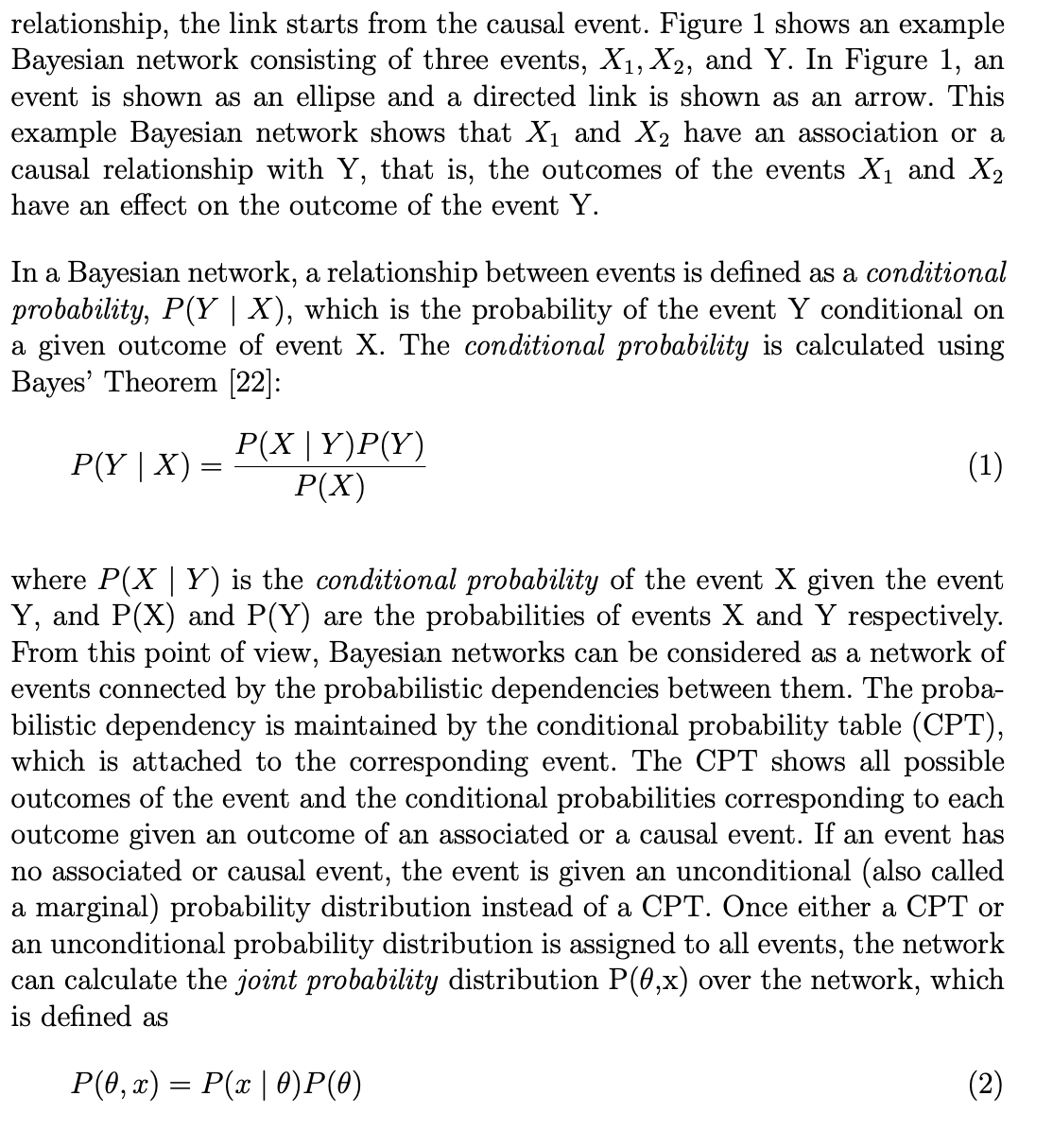
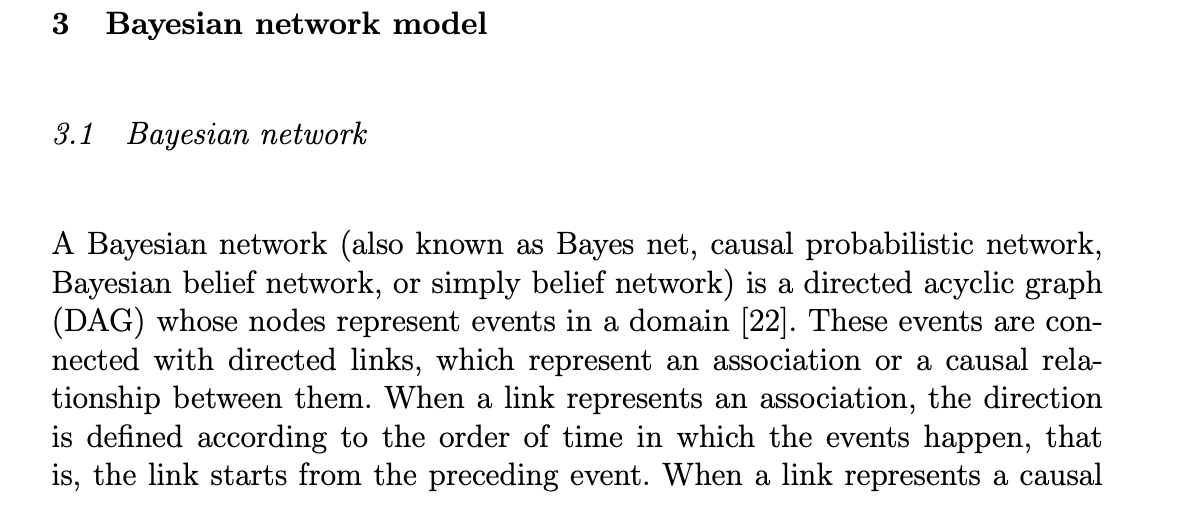
Bayes theorem

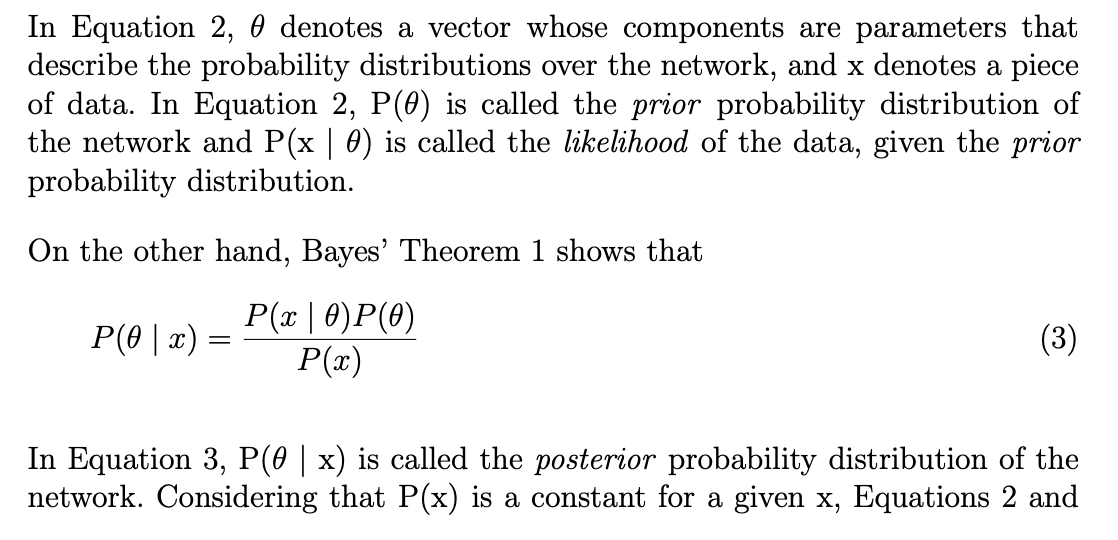
<https://ac.els-cdn.com/S0169743909000756/1-s2.0-S0169743909000756-main.pdf?_tid=a2b59e6b-16df-4e6f-b1dc-0051f0bc4978&acdnat=1542857421_c4009dbf975fe04551036ccd7f1de917>



<https://ac.els-cdn.com/S0169743909000756/1-s2.0-S0169743909000756-main.pdf?_tid=a2b59e6b-16df-4e6f-b1dc-0051f0bc4978&acdnat=1542857421_c4009dbf975fe04551036ccd7f1de917>

<https://ourarchive.otago.ac.nz/bitstream/handle/10523/919/dp2005-02.pdf>





There are a lot of uncertainty in our life, people tend to make a common-sense reasoning, but 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 it 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 it also 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 effencial 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, This article analyzes the relevant knowledge of Bayesian networks and 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 cannot explain the causal relationship between variables, It usually can't get enough data to establish model in daily life, Bayesian networks can overcome these defects well, 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.

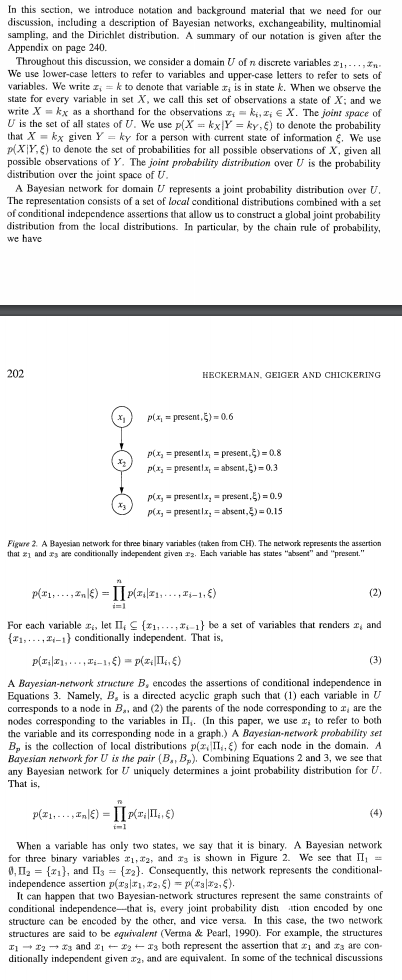
Bayesian algorithm research

With the rapid development of computer technology and network technology, people's ability to collect data by information technology has been greatly improved. People deeply realize, all kinds of data are stored in the computer system is a valuable information resource, which is likely to contain the many useful knowledge, the information or knowledge, will be provided or predict the unlimited business opportunities, key technology improvement, and even the important scientific discoveries, resulting in a remarkable economic and social benefits. However, due to the limitations of the tools people currently use, they cannot dig it out. Therefore, how to obtain useful and valuable information and knowledge in the practical field from various types of data and improve the efficiency of business management, production control, market analysis and scientific research has become a challenging task for computer researchers.

Knowledge discovery and data mining is made to adapt to the requirements, is the current database and artificial intelligence research hot topic, the goal is in the real world, with the amount, quality, variety of information sources in the form of complex, mining previously unknown, has potential application value, finally can be understand by user mode. In recent years, people have developed a variety of methods and technologies for knowledge discovery and data mining, which are mainly divided into the methods based on statistics, machine learning and mathematics. Among many methods of knowledge discovery and data mining, bayesian network, combined with graph theory and statistical knowledge, provides a method to express the causal relationship between variables. Based on probability theory and graph theory, nodes represent random variables, directed edges between nodes represent causal relationships between variables, and the degree of influence between variables is expressed by conditional probability attached to parent and child node pairs in the network. It is an ideal model for representing and dealing with uncertain knowledge.

Knowledge of probability

<https://link.springer.com/content/pdf/10.1007/BF00994016.pdf>



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,

(1)

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.

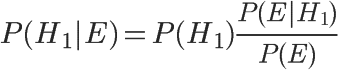
Fruit fructose problem Example

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 suppose that E is fructose, so the question becomes what is the probability of coming from bowl one given E, that is, P(H1|E). This probability is called a "posterior probability," a correction of P(H1) after the E event.

According to the conditional probability formula, we get:



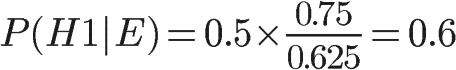
Let’s known that P(H1) is equal to 0.5, and P(E|H1) is the probability of taking the fruit drop out of bowl 1, which is equal to 0.75, so we can get the answer by finding P(E). According to the full probability formula:

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:

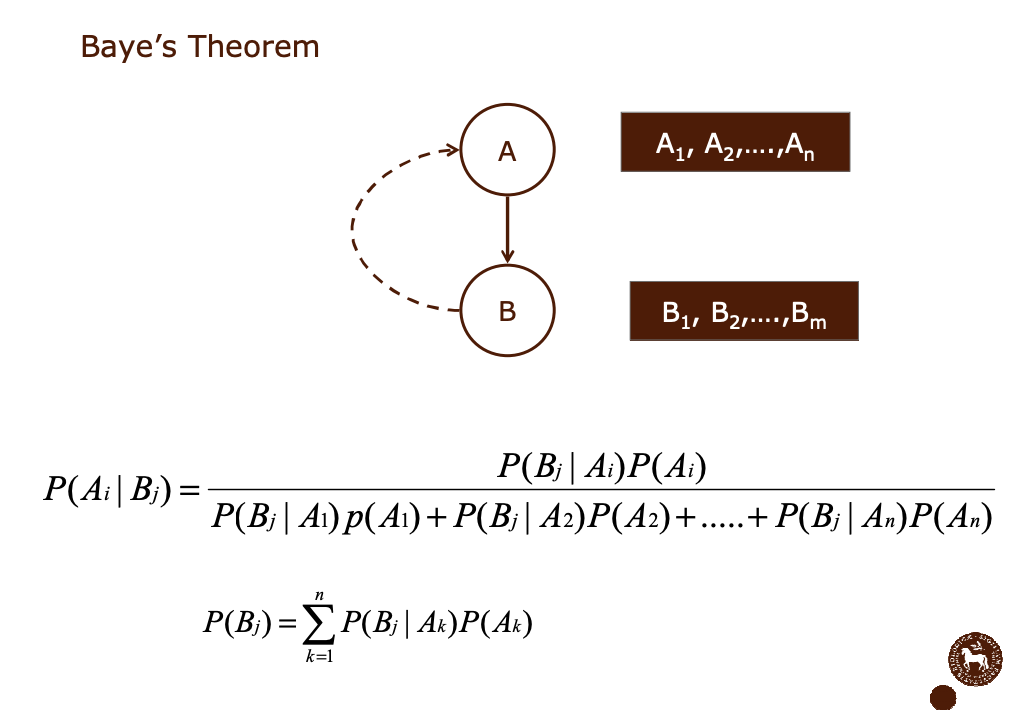
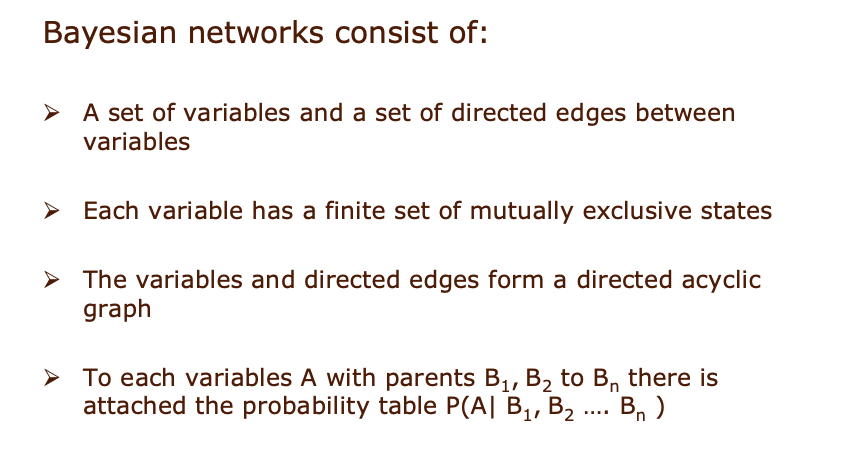
http://chart.googleapis.com/chart?cht=tx&chl=P(E)%3D0.75%5Ctimes%200.5%2B0.5%5Ctimes%200.5%3D0.625&chs=40

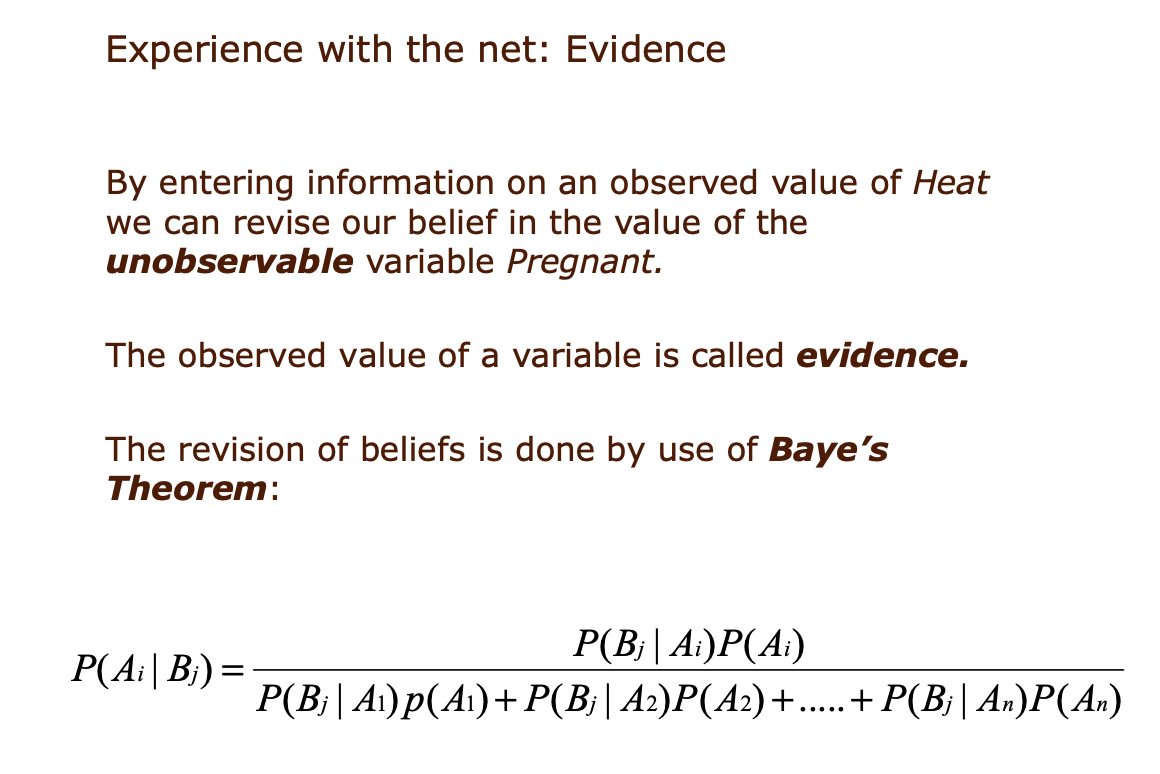
Substitute the Numbers into the original equation, and get:

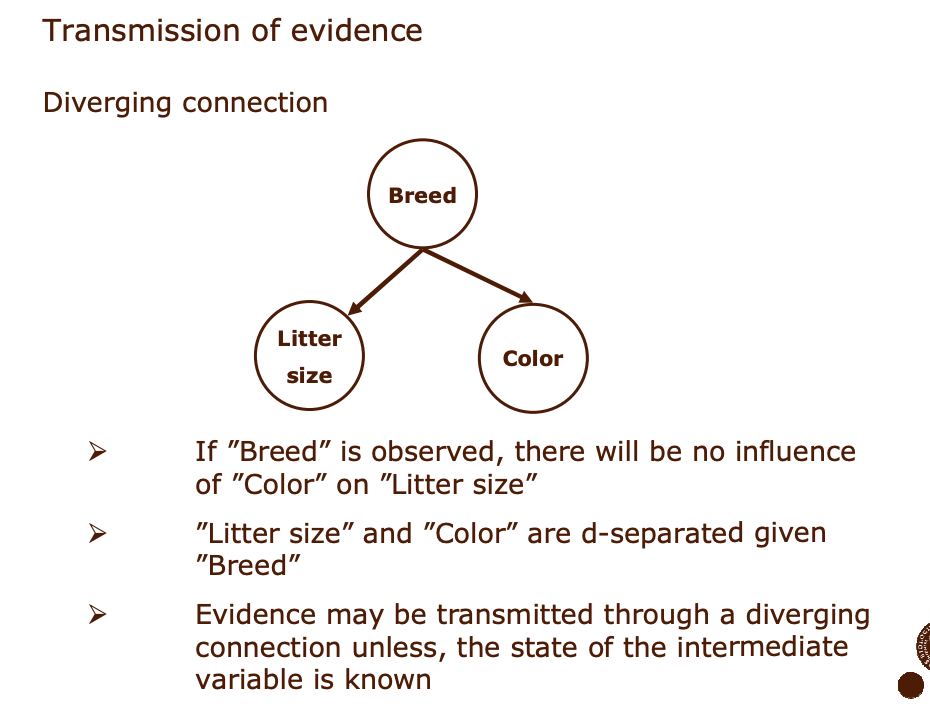
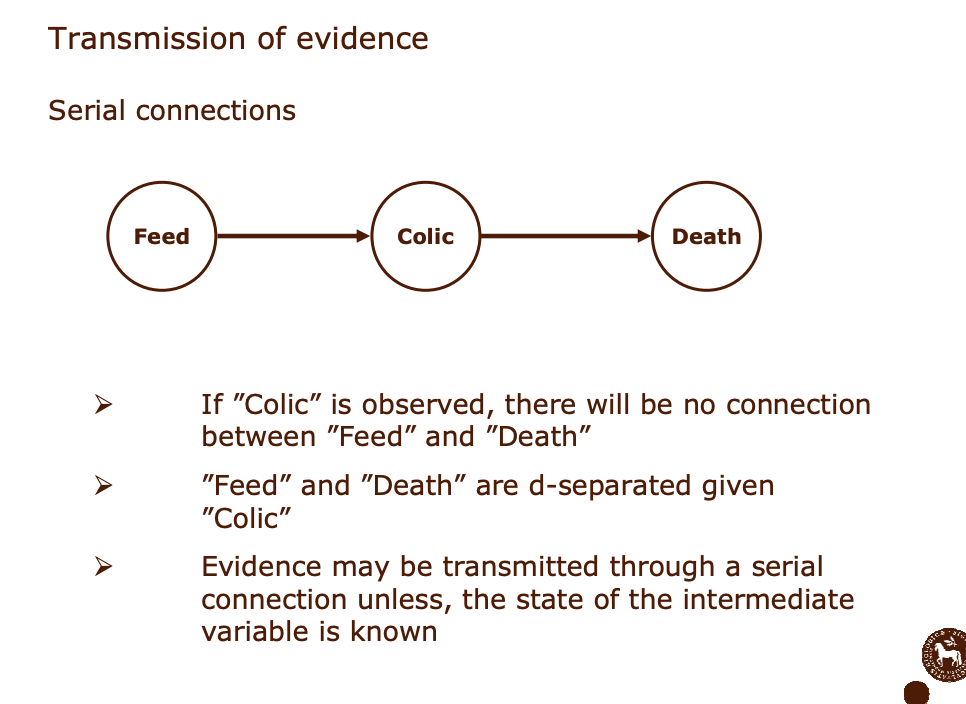


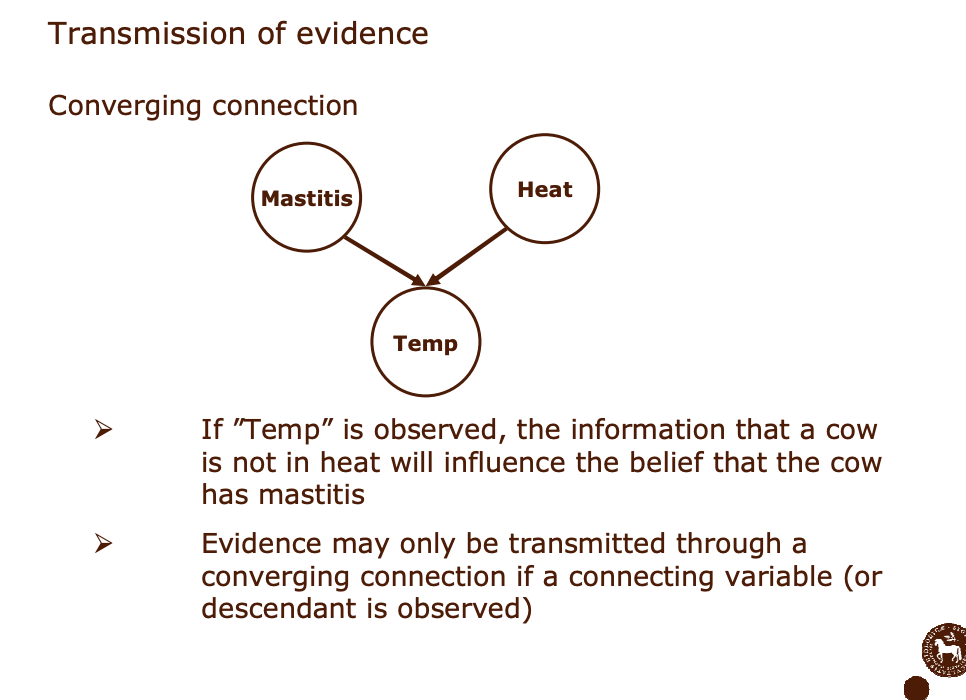
Bayesian network structure

<http://www.prodstyr.ihh.kvl.dk/vp/2009/slides/BayesNet-I-6.pdf>









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)

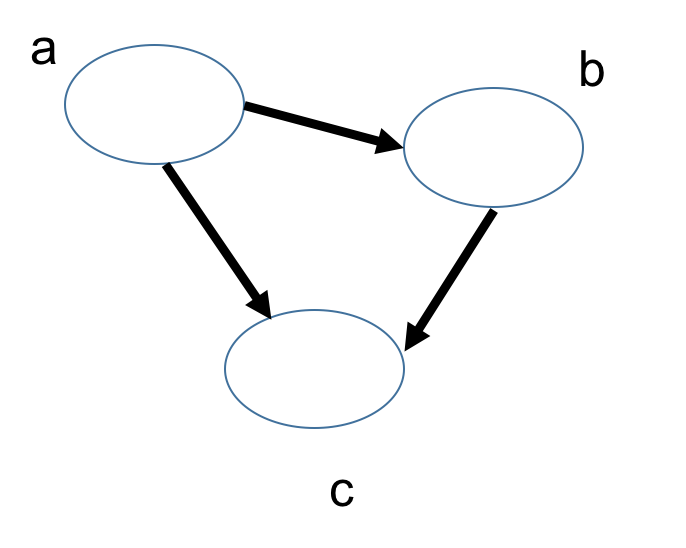


Fig . A simple BN

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)

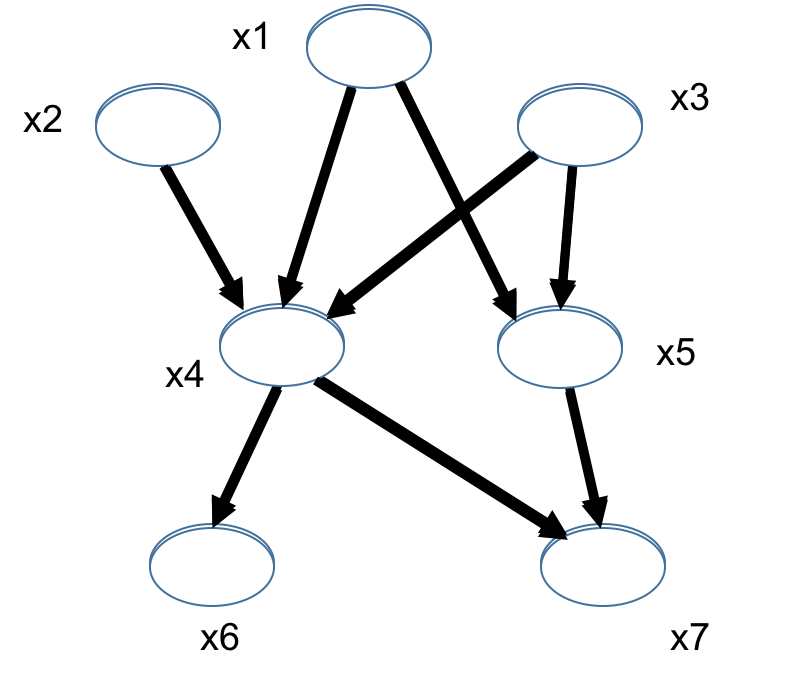


Fig .The complex BN

Bayesian network structure form

Form 1: Head-To-Head

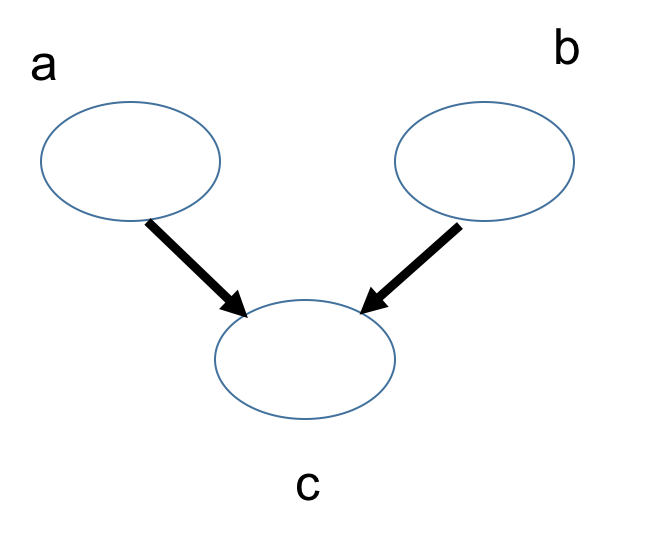


Fig . Head-To-Head

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

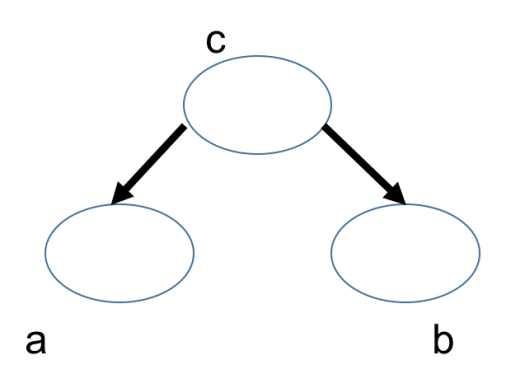


Fig .Tail-To-Tai

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

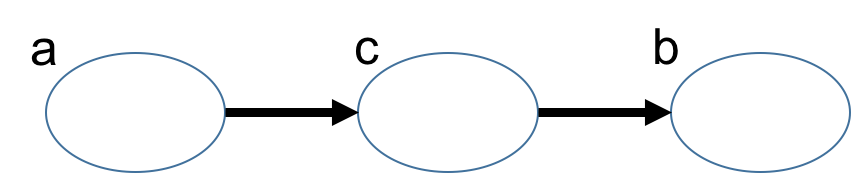
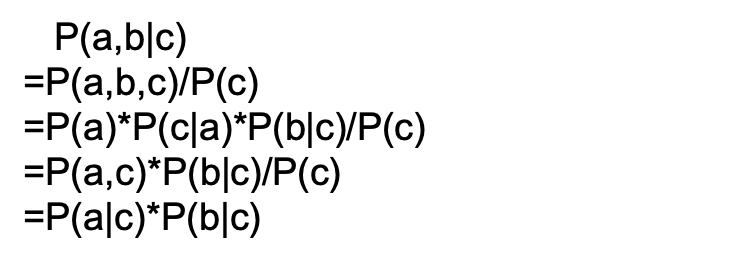


Fig . Head-To-Tai

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:



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.

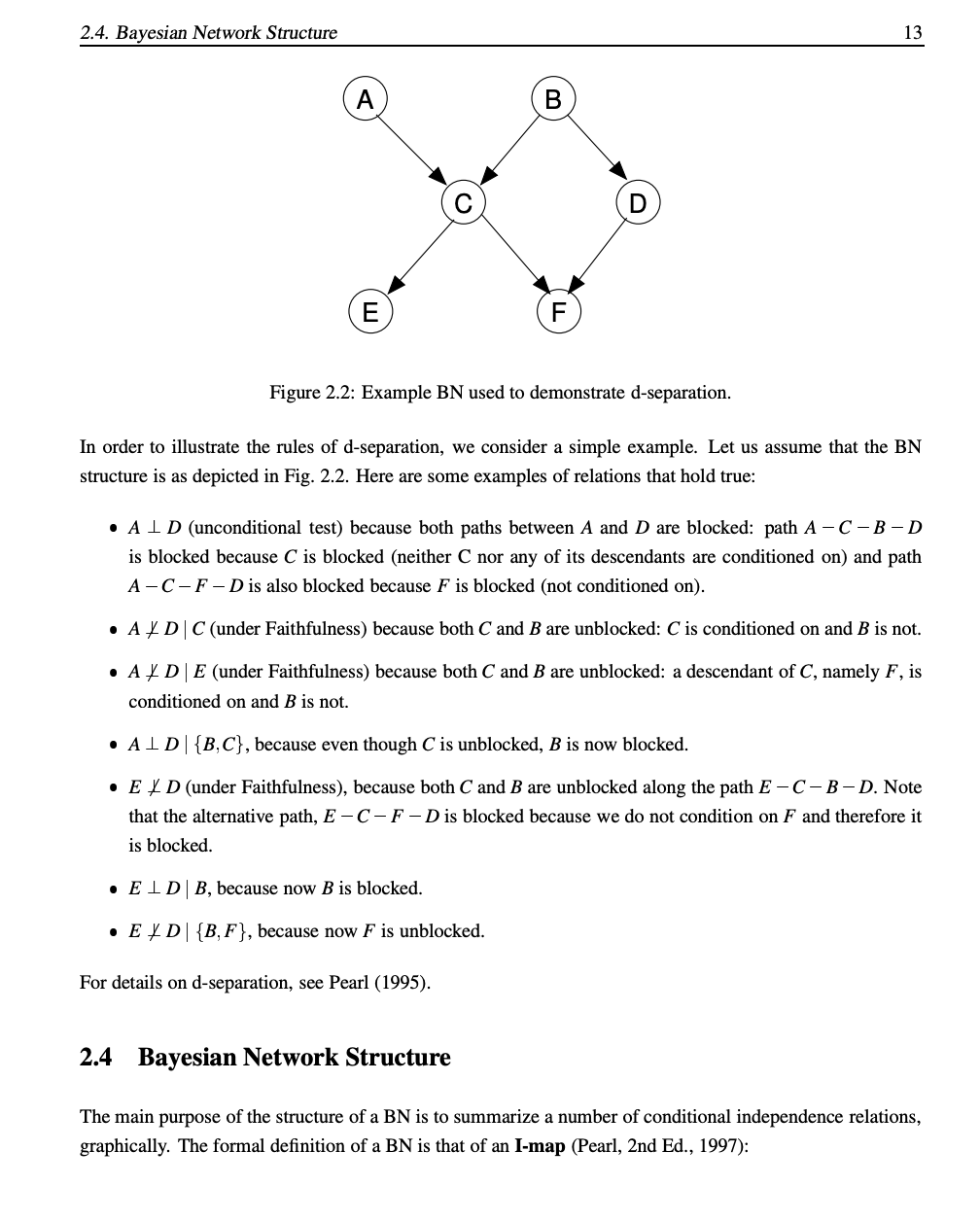
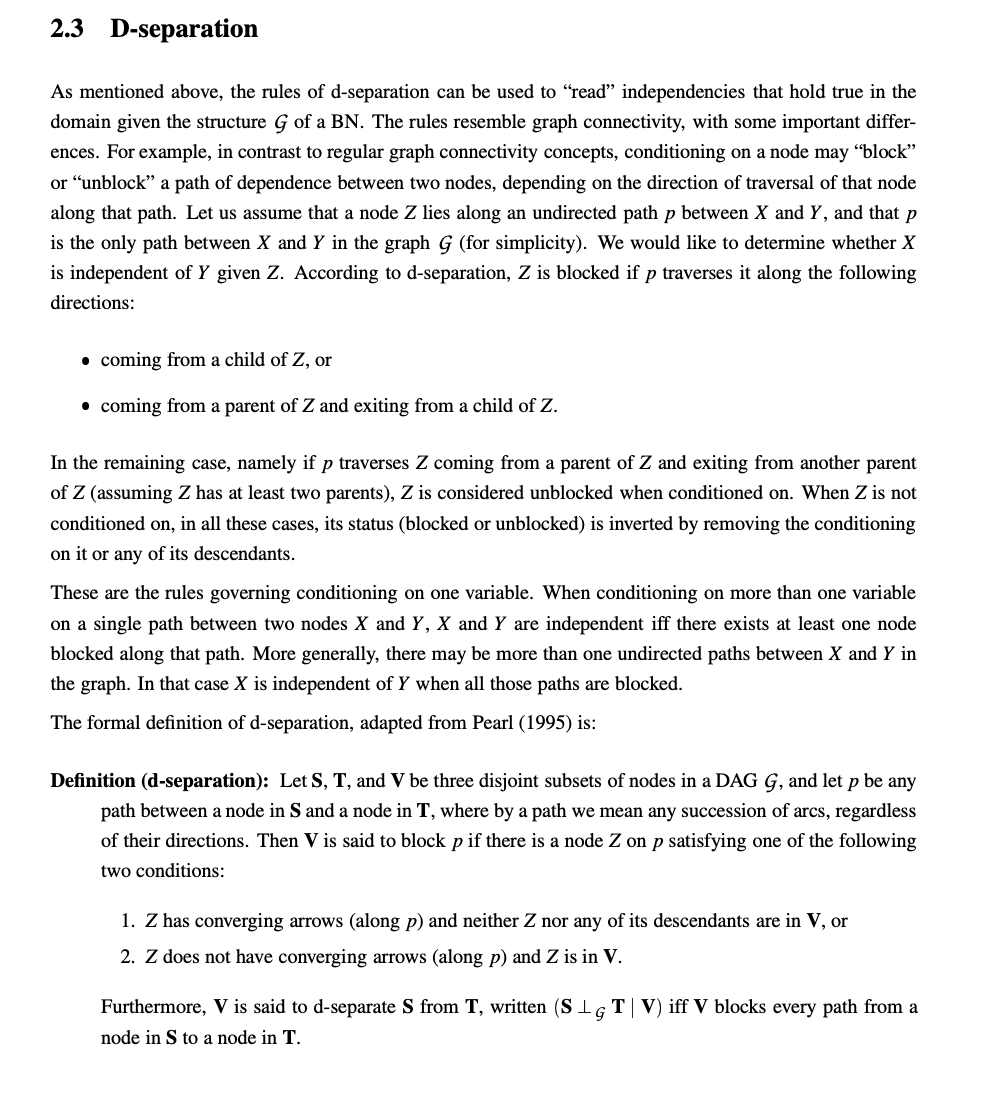
On the premise of unconstrained polynomial distribution, parameter independence, Dirichlet priori and data integrity, the structural likelihood of data is exactly the product of the structural likelihood of each (i, j) pair, i.e.



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.

D separation

<https://www.cs.cmu.edu/~dmarg/Papers/PhD-Thesis-Margaritis.pdf>



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, E don’t need to consider 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.

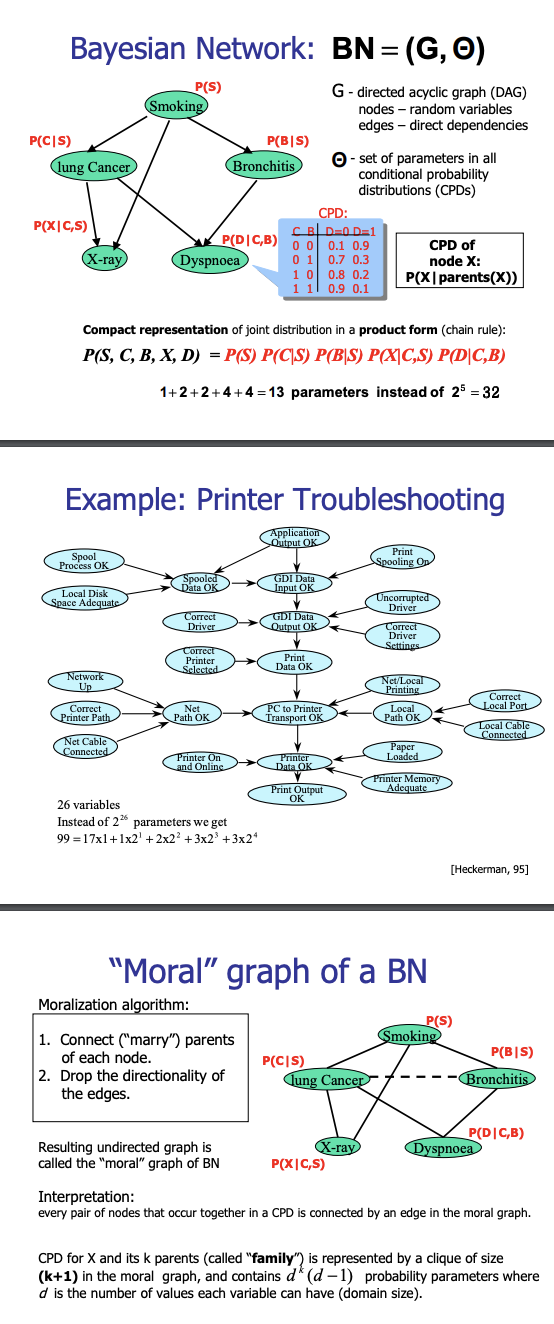
Blocking: Given the set of evidence, when any of the above conditions is satisfied, Vb blocks the corresponding path.

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.

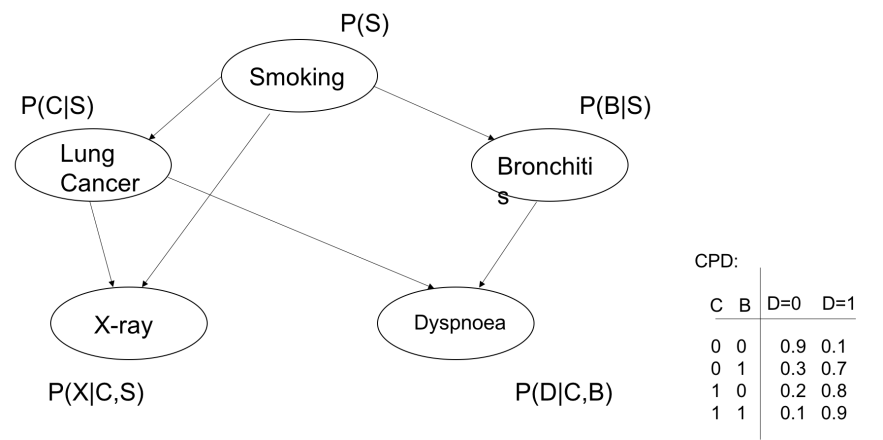
<http://www.ee.columbia.edu/~vittorio/Lecture12.pdf>

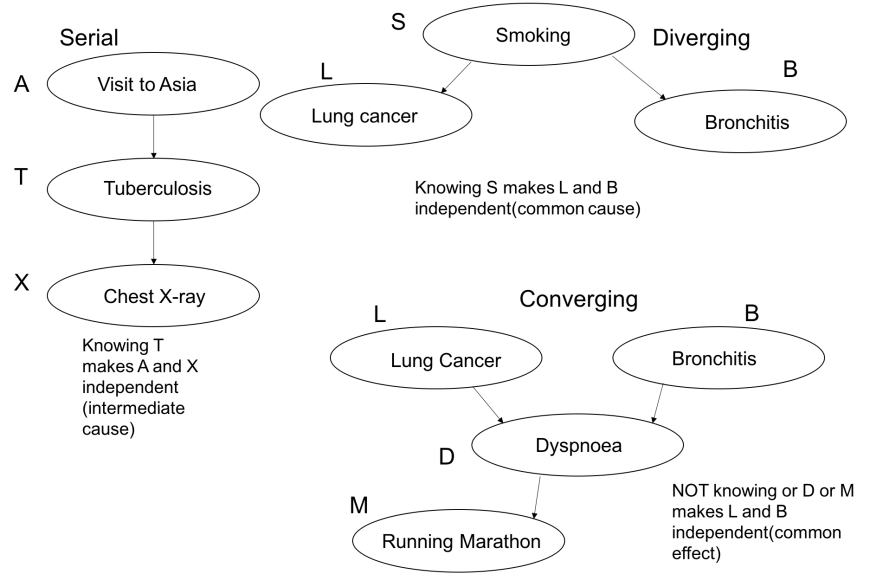


Smoking with the probability P (S), said one person in the case of smoking is lung cancer with the probability P (C | S) said, X - ray said need according to medical X-ray, lung cancer may lead to need according to the X-ray, smoking is likely to lead to need according to X light (so smoking is also a because of X - ray), therefore, smoking and lung cancer and need the probability of the X-ray of P (X | C, S) said.

Bronchitis, said a person in the case of smoking to the probability of Bronchitis with P (B | S), breathing difficulties, dyspnea said Bronchitis may cause difficulty in breathing, lung Cancer is likely to lead to difficulty breathing (so, lung Cancer and dyspnea with a), breathing difficulties caused by smoking and got Bronchitis with the probability P (D | C, B).

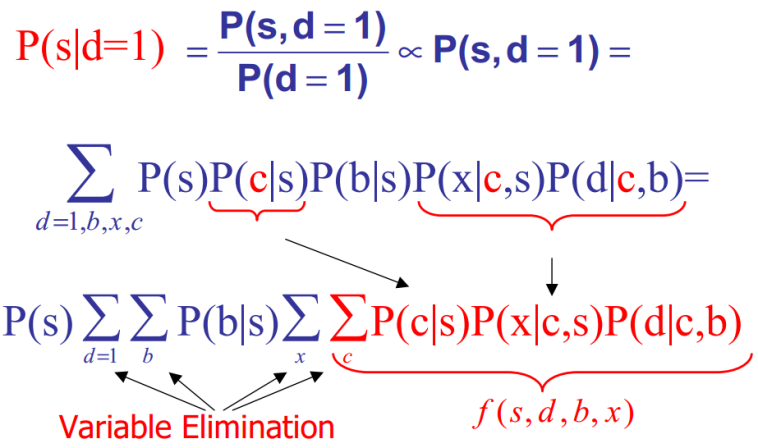
Lung Cancer shorthand for C, Bronchitis shorthand for B, dyspnea shorthand for D, and C = 0 means lung Cancer will not occur probability, C = 1 said lung Cancer occurrence probability, B is equal to 0 (B doesn't happen) or 1 (B) is similar to C, in the same way, the incidence of the D = 1 D, D = 0 D does not occur probability, dyspnea can be got a probability form.





What is the probability of sampling a person who has difficulty breathing?





At the beginning, all the variables are at the end of sigma(d=1,b,x,c), but since P(s) and "d=1,b,x,c" don't matter, you can go to the front. And P of b, b, b, b, 0, s doesn't matter with x and c, so you can also take it out and put it after sigma of b, so you're left with sigma of x and sigma of c on the right-hand side.

Application analysis of Bayesian network

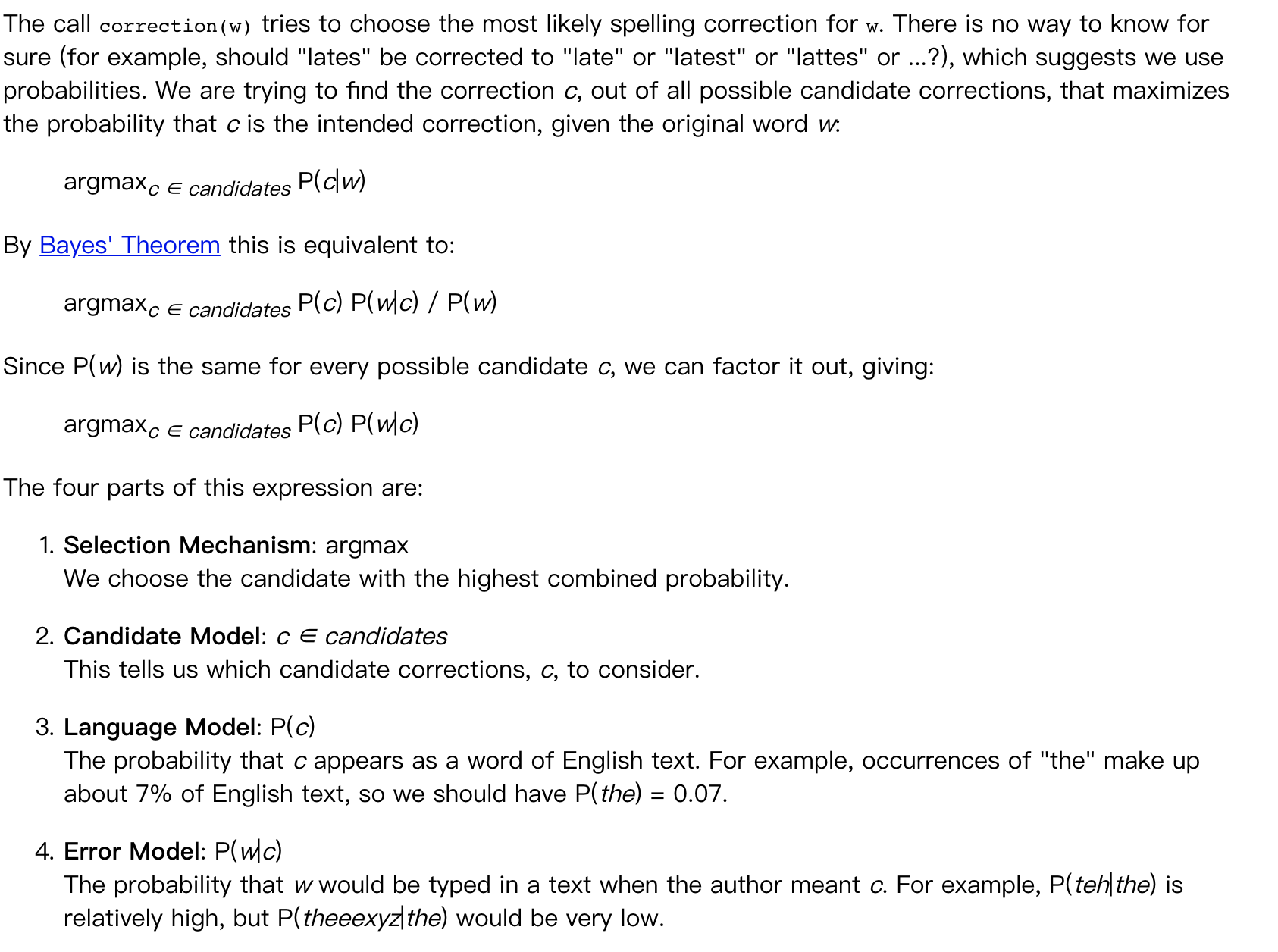
The human brain can't calculate precise probability numerically, so although group behavior can often be explained by Bayesian principle, and individual behavior may be deviated greatly by random adoption, our Bayesian computer may become a machine to study and think instead of human beings.

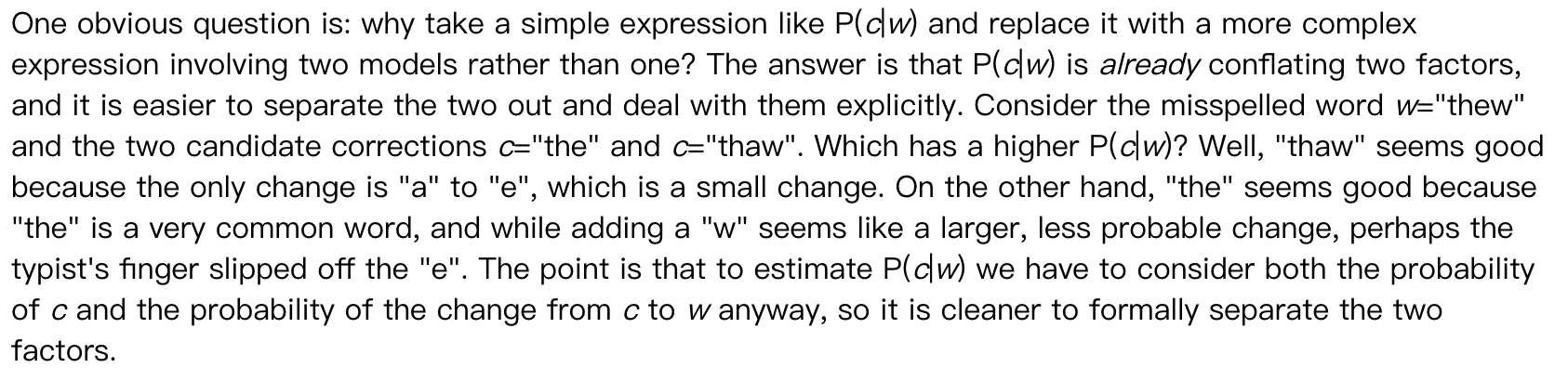
Bayesian theorem provides a new basis for probability theory. Traditional probability theory always assumes a prior probability based on prior knowledge, and then deduces probability on this basis. However, in many cases, the prior probability cannot be accurately obtained. There are always various uncertainties in our cognition. The Bayesian School regards the prior probability as a belief, and constantly revises the premise hypothesis by observing data, thus forming new beliefs, which continuously updates the posterior probability and makes the subjective probability approach the objective probability.

Bayesian Theorem provides logic of discovery. The Frequency School actually believes that the law behind phenomena exists objectively in the form of Plato's "idea". Bias theory requires us to understand the law with more and more strong faith. Several great philosophers, such as Einstein, set up a theoretical system based on assumptions in the narrow sense and the general theory of relativity, and then let others prove from experiments that their theory is completely consistent with the observed data. A lot of discoveries of electromagnetic theory and quantum theory are based on the observation data to find the best theoretical explanation. The discovery of Planck's theorem can explain the problem of blackbody radiation perfectly only by using quantum hypothesis, which opens the door of quantum theory.

Bayesian Theorem is highly consistent with the cognitive mechanism of human brain. The human brain does not really carry out accurate probability calculation, but uses some ingenious approximation mechanism (such as the number and shape of nerve impulses) to complete the reasoning process. At present, the cognitive theory of human brain based on Bayesian model has made considerable progress, which can effectively explain many cognitive behaviors of human brain, such as concept learning, object recognition, perception-motion integration and induction.

Application: Spell check

<https://norvig.com/spell-correct.html>



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

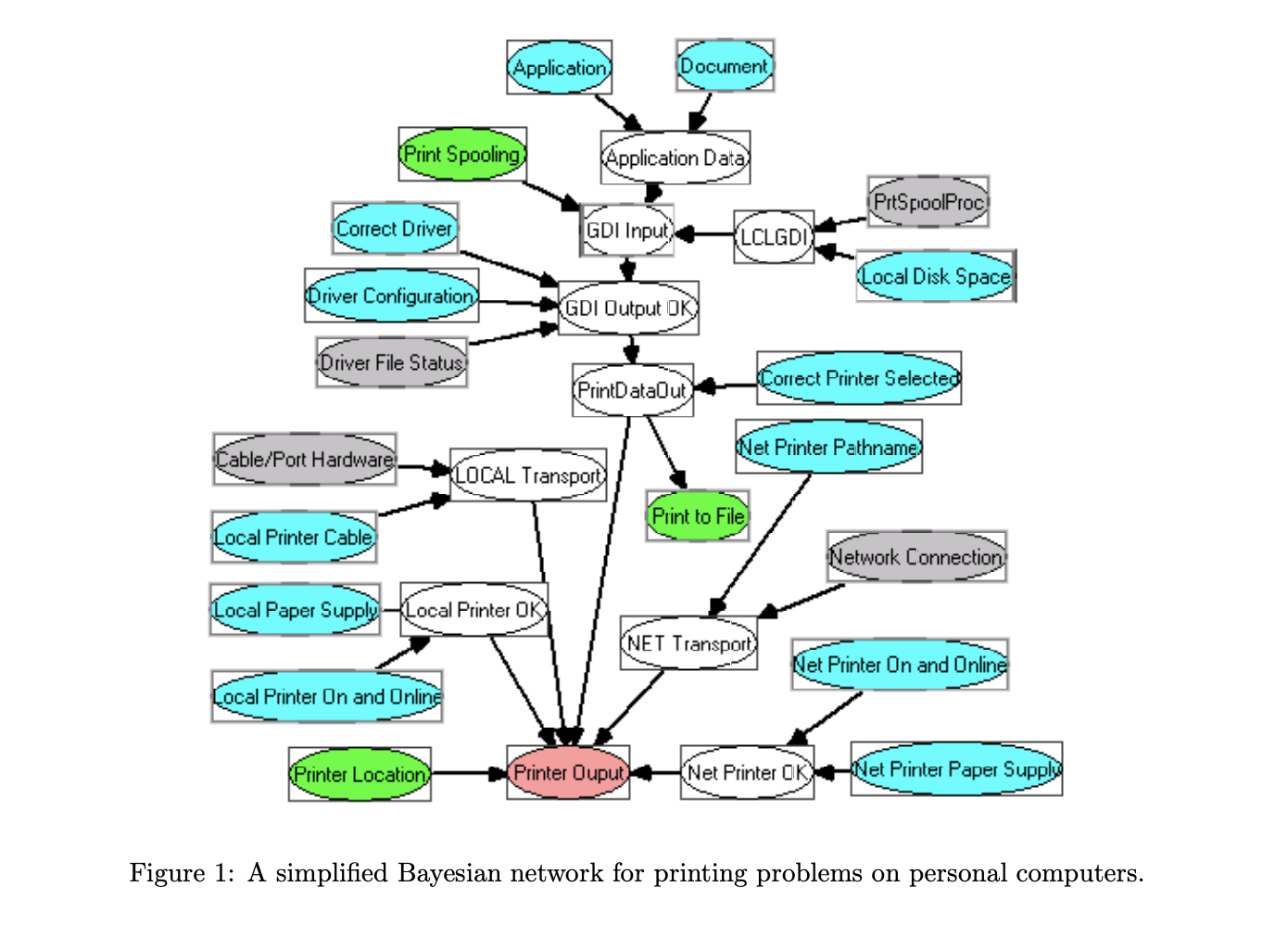
P(c) denotes the occurrence "probability" of a certain correct word, which can be replaced by "frequency". If we have a large enough text library, the frequency of each word in the text library is equal to the probability of its occurrence. The higher the frequency of a word, the greater the P(c). For 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).

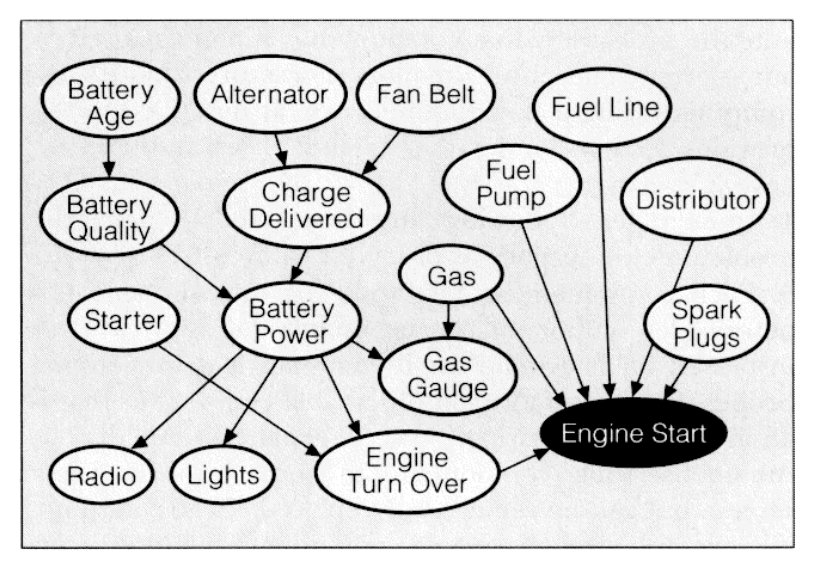
So, compare how often all words with similar spellings appear in the text library, and pick the one that appears most often, the one that the user most wants to type.

Application of Bayesian network in fault diagnosis

<https://pdfs.semanticscholar.org/fb78/bf93bedde7534755bab5f52b79682a439c7f.pdf>



[http://delivery.acm.org/10.1145/210000/203341/p49-heckerman.pdf?ip=115.145.243.116&id=203341&acc=ACTIVE SERVICE&key=0EC22F8658578FE1.B50D9BE1468BDDBD.4D4702B0C3E38B35.4D4702B0C3E38B35&\_\_acm\_\_=1542870368\_0981bbeb8c3fabefe115027acab38cf9](http://delivery.acm.org/10.1145/210000/203341/p49-heckerman.pdf?ip=115.145.243.116&id=203341&acc=ACTIVE%20SERVICE&key=0EC22F8658578FE1%2EB50D9BE1468BDDBD%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&__acm__=1542870368_0981bbeb8c3fabefe115027acab38cf9)



With the continuous growth of global demand for oil and gas resources, deep-sea oil exploration and development has attracted more and more attention. As the mainstream deep-sea oil and gas development mode, the safety and reliability of underwater production system is very important. M-span nozzles are widely used in underwater production systems, and they are responsible for transmitting the production fluid between different devices. Due to the frequent seabed activities and the harsh environment of high temperature and pressure, the cross-pipe is prone to failure. Therefore, the study on fault diagnosis of cross-pipe has become an urgent problem to be solved in offshore oil and gas development[3-6].

At present, researches on cross-pipe mainly focus on the simulation study of its failure mechanism, and few on the test and fault diagnosis methods. Yang wen et al. [3] firstly established the m-type cross-pipe fault diagnosis test system and studied how to obtain cross-pipe fault data. Then, the simulation verification test data are compared. Finally, a cross-pipe bayesian fault diagnosis network is developed based on experimental and simulation data.

The actual bayesian fault diagnosis network of type M cross-pipe is shown in figure 7. The network is composed of two layers, the upper layer is the fault layer, and the node is F1 ~ F11, which represents 11 possible fault barriers. The shape of the node is presented as the failure obstacle of the test point represented by the node, and the value is 1; Absent represents that no failure obstacle was found at the test points on the node representation, with a value of 0. According to the loading capacity of the test device, the following Settings were set. When the stress value monitored by any fault node exceeded 25 MPa, the status of the fault node was determined to be Present; otherwise, it was Absent. The lower layer is the fault symptom node, i.e. S1 to S12, which is distributed on both sides of the upper 11 fault nodes. It is necessary to analyze and process the collected data to determine the node status of fault symptom layer. Data preprocessing includes training sample selection and node interval division from two aspects. The determination of training samples should adhere to the principle of generalization and ensure an adequate number of samples. 10% of all simulation data and experimental data are randomly selected as verification data, which will not appear in the data of parameter learning. The remaining 90% of the data was then used as the parameter learning of bayesian network. The division of node interval refers to the division of the interval of bayesian network fault symptom nodes according to the characteristics of sample data. In particular, it is the process of data classification according to different types of fault obstacle in sample data. According to the training samples, the interval division results of symptom nodes are shown in Table1.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sensor01  /Mpa | Interval | Sensor02  /Mpa | Interval | Sensor03  /Mpa | Interval | Sensor04  /Mpa | Interval | Sensor05  /Mpa | Interval | Sensor06  /Mpa | Interval |
| 12.3 | 1 | 21.4 | 1 | 21.4 | 1 | 12.3 | 1 | 6.2 | 1 | 13.8 | 1 |
| 16.2 | 2 | 28.2 | 2 | 28.2 | 2 | 16.2 | 2 | 8.1 | 2 | 18.2 | 2 |
| 18.5 | 3 | 32.1 | 3 | 32.1 | 3 | 18.5 | 3 | 9.2 | 3 | 20.8 | 3 |
| 100 | 4 | 100 | 4 | 100 | 4 | 100 | 4 | 100 | 4 | 100 | 4 |
| Sensor07  /Mpa | Interval | Sensor08  /Mpa | Interval | Sensor09  /Mpa | Interval | Sensor10  /Mpa | Interval | Sensor11  /Mpa | Interval | Sensor12  /Mpa | Interval |
| 13.8 | 1 | 6.1 | 1 | 12.2 | 1 | 21.4 | 1 | 21.4 | 1 | 16 | 1 |
| 18.2 | 2 | 8.1 | 2 | 16.2 | 2 | 28..2 | 2 | 28.1 | 2 | 21 | 2 |
| 20.8 | 3 | 9.3 | 3 | 18.4 | 3 | 32.1 | 3 | 32.1 | 3 | 24 | 3 |
| 100 | 4 | 100 | 4 | 100 | 4 | 100 | 4 | 100 | 4 | 100 | 4 |

Fig . Fault symptom node interval division

Implementation of bayesian network algorithm

Bayesian network reasoning algorithm uses elimination algorithm. Take the posterior probability of solving the fault node F1 as an example. 

Fig . Cross-pipe fault diagnosis bayesian network

1) determine the elimination order as {F2, F3... F11}.

2) take the data collected by the sensor as a set of evidence E = {E1, E2... , E12} input into the network, and the set is {P (F1), P (F2)... , P (F11), P (E1 | F1, F7), P (E2 | F1, F7, F2)... , P (E12 | F11, F6)}.

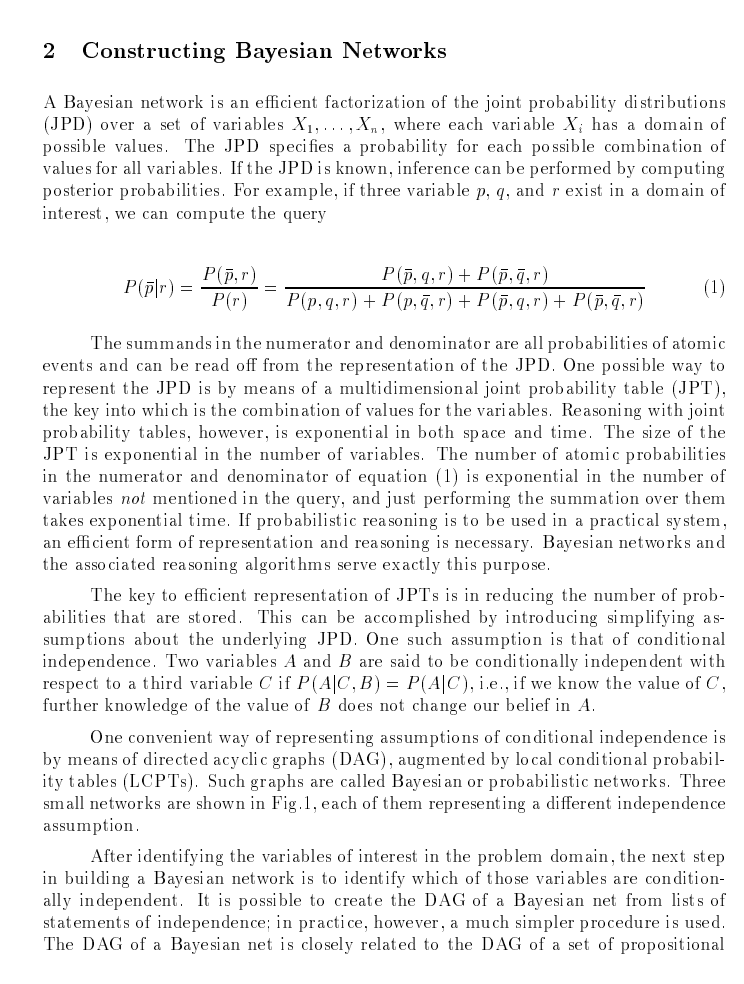
3) in accordance with the elimination order, the first variable F2 is eliminated. F2 is eliminated by finding the function related to F2 in the solution set, get {P (F1),P (F3)… , P (F11), P (E1 | F1, F7), P (E5 | F8, F3),... ", P (E12 | F11, F6), the number 1 (E2, E3, E4, F1, F7, F2),F8)}, in which 1 (E2, E3, E4, F1, F7, F2, F8) =∑ F2P (F2), P (E2, |, F1, F7, F2), P (E3, |, F7),F2) P (E4, |, F2, F8F3).

4) eliminate the remaining variables F3 ~ F11 in turn, and obtain {P (F1), 10 (F1)}

5) calculate h (F1) = P (F1), which means 10 (F1).

6) return P (F1, |, E) = h (F1)F sub 1h of f sub 1.Similarly, this algorithm can be used to query the posterior probability of other fault nodes.Analysis of bayesian network diagnosis results

The verification data shown in table 6 are input successively into s1-s12 of fault signs of bayesian fault diagnosis network of m-type cross-pipe bayesian fault diagnosis as shown in figure . According to this calculation, the posterior probability of F1-F11 of fault nodes is calculated by the mode.



1995 years Microsoft company will BN applied to fault diagnosis for the printer , 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 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. 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.



Fig .A simplified BN diagnosis model for printer fault

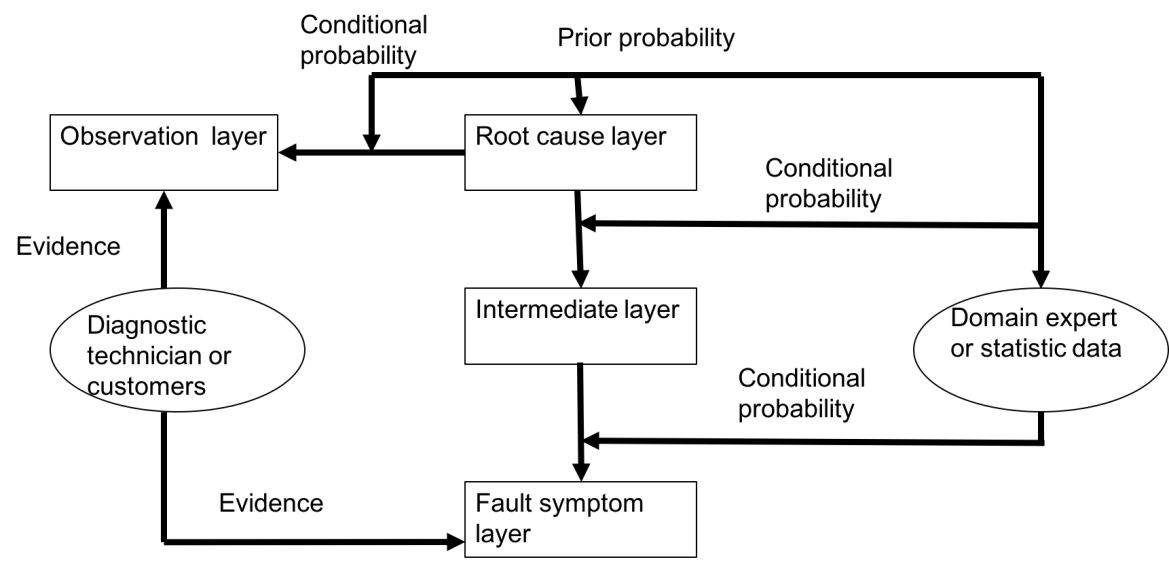


Fig . Framwork of BN model for automotive diagnosis

Bayesian network 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.

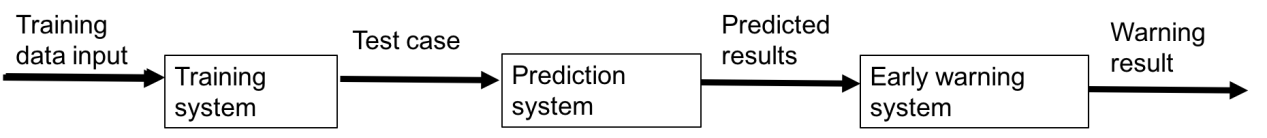
****

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.

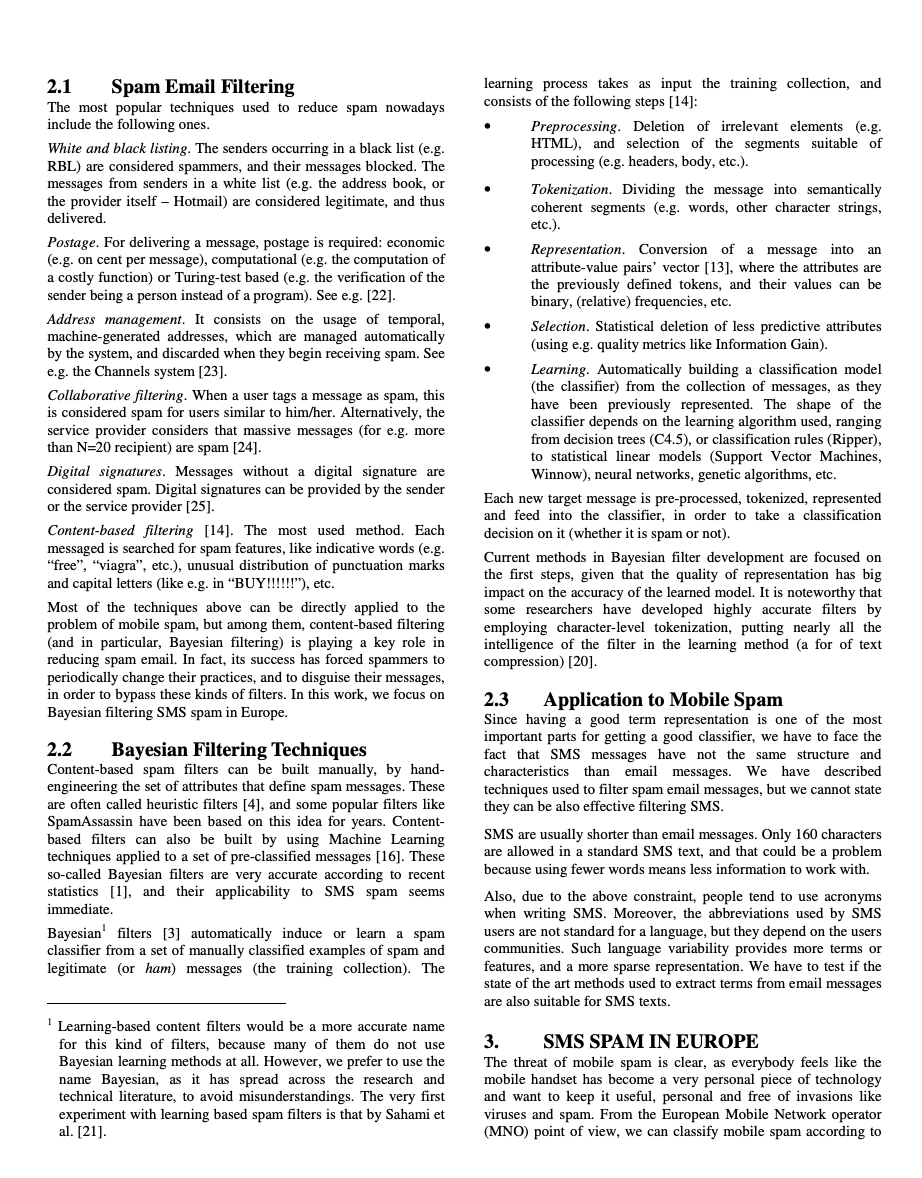
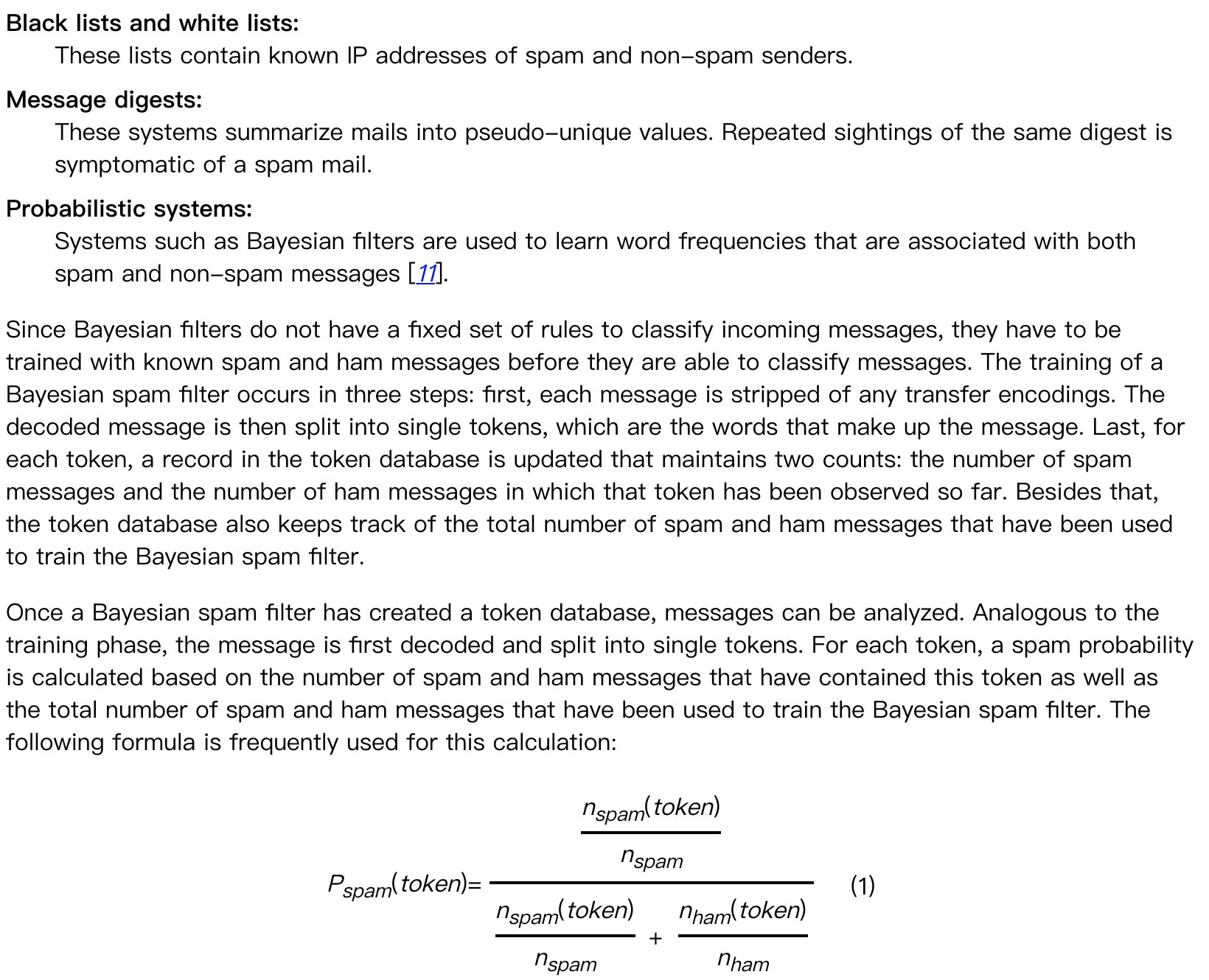
Bayesian network is applied in medical diagnosis to develop an expert diagnosis system based on Bayesian. Its functions include:

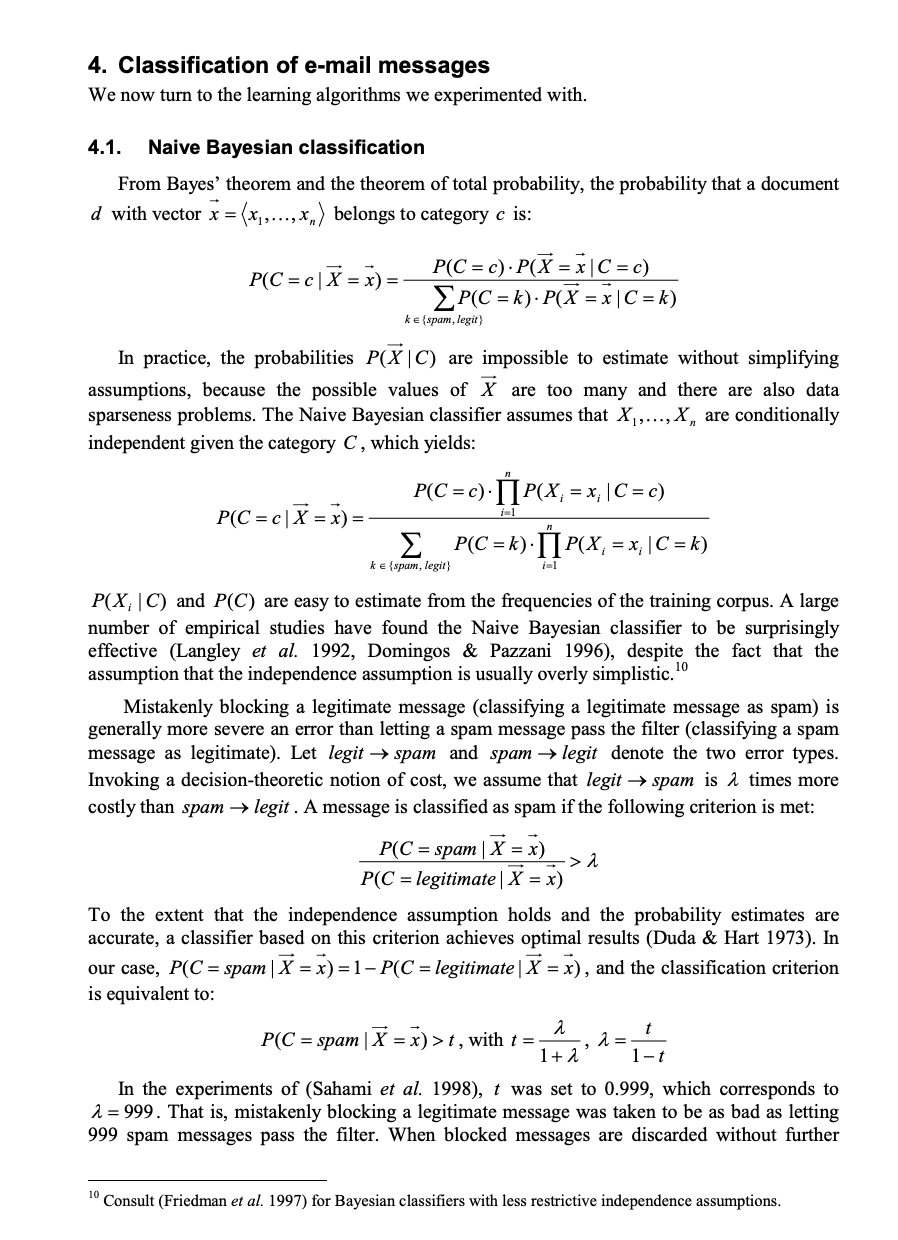
(1) Make a diagnosis of the new case to determine whether it is normal or abnormal. The judgment is based on the probability that the instance belongs to each class. The structural framework and system process of the system are given, and the parameter Settings of folding number and attributes are given (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.

Bayesian filters

<https://www.usenix.org/legacy/event/woot07/tech/full_papers/karlberger/karlberger_html/>





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%.

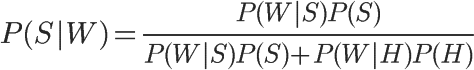
Use S for spam and H for healthy. So the prior probabilities of P(S) and P(H) are 50%.

../论文.png

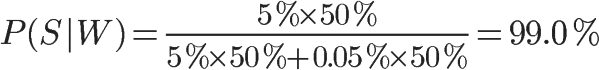
Then, I analyzed this email and found that it contained the word "advertisement". What is the probability that this email is spam?

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 here 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".

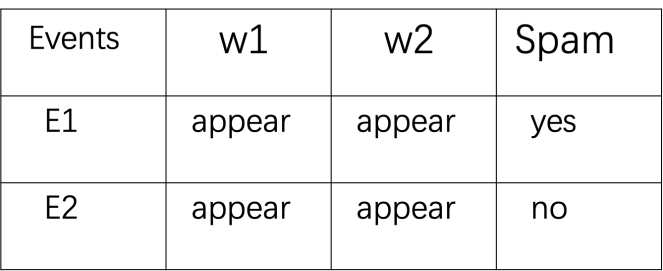
Can you conclude from the above step that this new email is spam?

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?

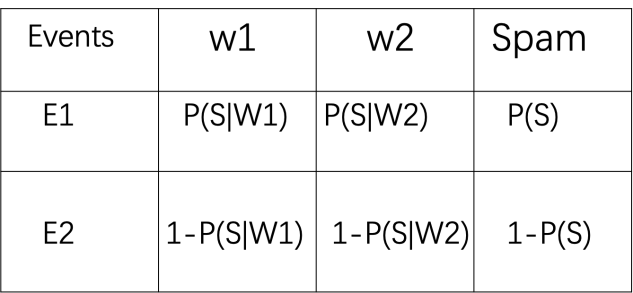
What Paul Graham did was to pick the top 15 words in the letter, the highest P(S|W), and calculate their combined probability. (if some words appear for the first time and you cannot calculate P(S|W), Graham assumes that this value is equal to 0.4. Because spam tends to be a fixed word, if you've never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

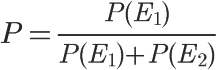


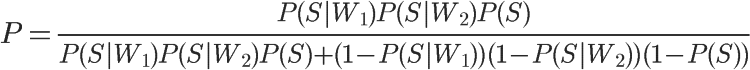
If all events are assumed to be independent events, then P(E1) and P(E2) can be calculated:

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

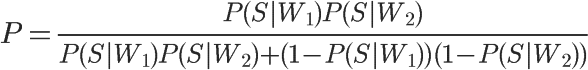
http://chart.googleapis.com/chart?cht=tx&chl=P(E_%7B2%7D)%3D(1-P(S%7CW_%7B1%7D))(1-P(S%7CW_%7B2%7D))(1-P(S))&chs=40

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

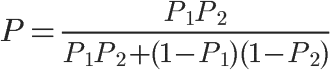




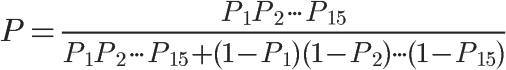
You plug in P of S equal to 0.5 and get:



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.

## 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.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4 | 349 | 13 | 64 | 9 | 207 | 33 | 72 | 12 | 126 | 38 | 54 | 10 | 67 | 49 | 43 |
| 2 | 232 | 27 | 84 | 7 | 201 | 64 | 95 | 12 | 115 | 93 | 92 | 17 | 79 | 119 | 59 |
| 8 | 166 | 47 | 91 | 6 | 120 | 74 | 110 | 17 | 92 | 148 | 100 | 6 | 42 | 198 | 73 |
| 4 | 48 | 39 | 57 | 5 | 47 | 132 | 90 | 9 | 41 | 224 | 65 | 8 | 17 | 414 | 54 |
| 5 | 454 | 9 | 44 | 5 | 312 | 14 | 47 | 8 | 216 | 20 | 35 | 13 | 96 | 28 | 24 |
| 11 | 285 | 29 | 61 | 19 | 236 | 47 | 88 | 12 | 164 | 62 | 85 | 15 | 113 | 72 | 50 |
| 7 | 163 | 36 | 72 | 13 | 193 | 75 | 90 | 12 | 174 | 91 | 100 | 20 | 81 | 142 | 77 |
| 6 | 50 | 36 | 58 | 5 | 70 | 110 | 76 | 12 | 48 | 230 | 81 | 13 | 49 | 360 | 98 |

Fig . 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.

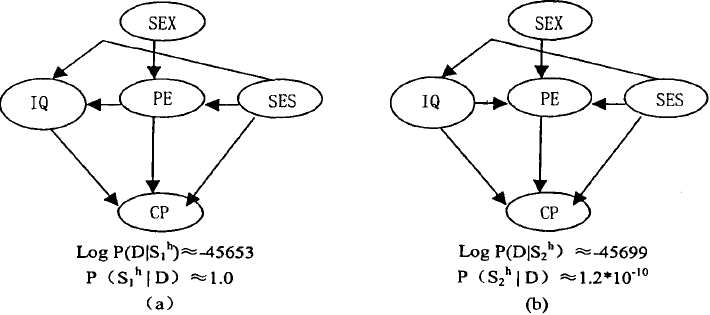


Fig . The possible network structures

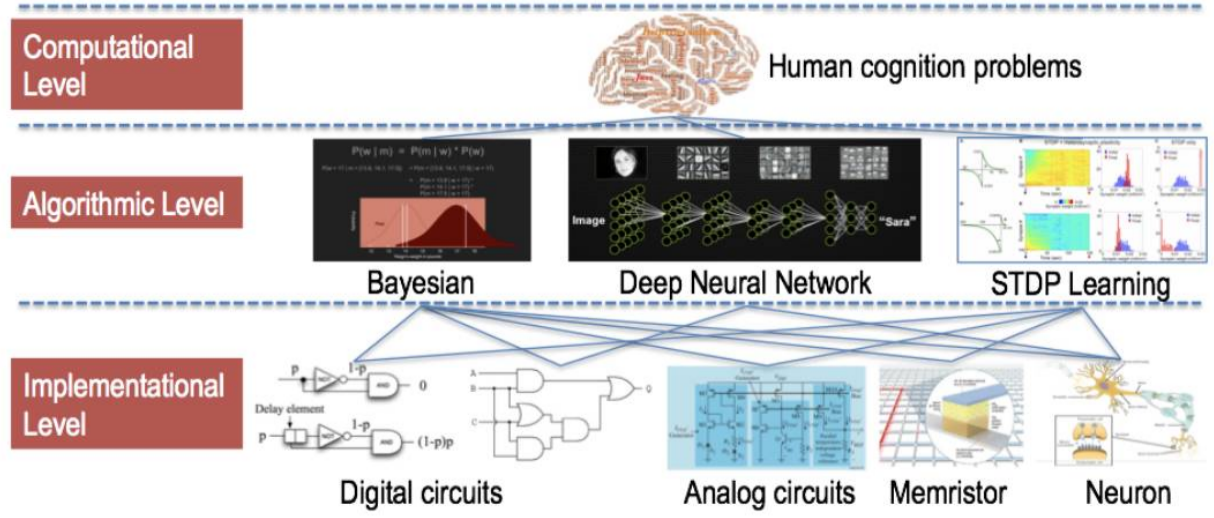
From figure, we find that socioeconomic status (SES) has a direct impact on intelligence quotient (IQ). In order to verify this result, a new model is considered, that is, the direct impact of social and economic status (SES) on IQ in the model of figure is replaced by a hidden variable (H) pointing 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).

Bayesian network and artificial intelligence

<https://www.taylorfrancis.com/books/9781439815922>

Learning and inference based on Bayesian theory is an important branch of artificial intelligence. From the perspective of human-like brain computing, Bayesian theory is closely related to artificial intelligence. Use the following figure to make a brief explanation.



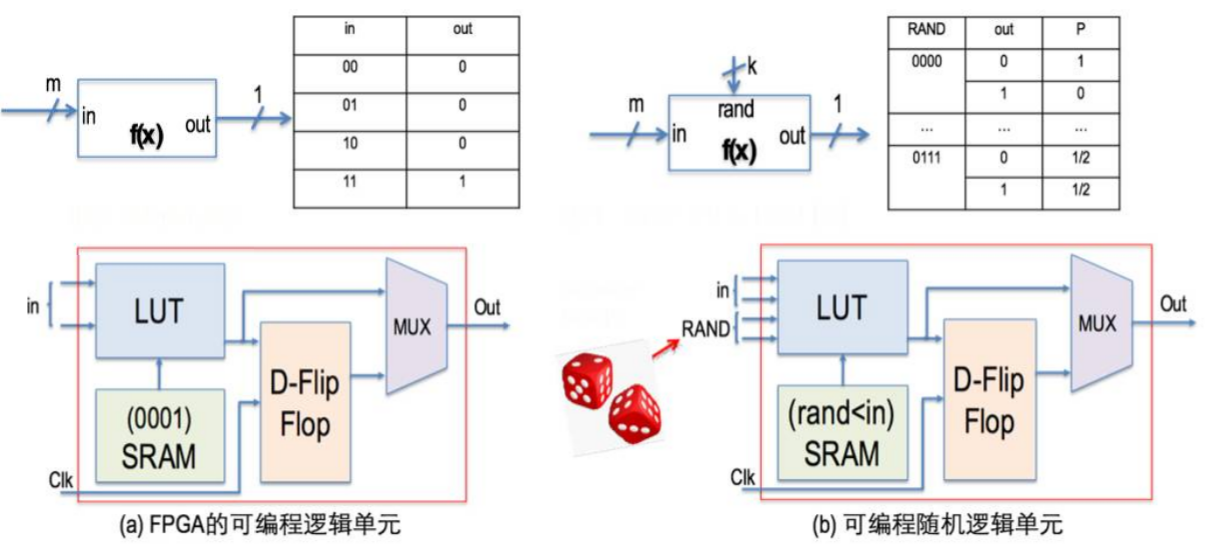
Understanding and analyzing human brain requires three levels of effort, namely, computation level, algorithm level and implementation level. Computing layer pays attention to functions and interfaces, that is, functions accomplished by certain human cognitive behaviors and corresponding inputs and outputs. Typical computer tasks include visual processing, memory, understanding, reasoning, language, learning and creativity. The computing layer is not concerned with the method and material foundation of the realization function.

The algorithm layer is concerned with the realization of human cognitive behavior, but it does not care about the specific material foundation. For example, the basic mechanism of human brain learning is synaptic plasticity, but the neurons here can be biological neurons, silicon-based circuits, or neuron programs implemented by software. Depth neural network can also be regarded as an algorithm for image classification, object detection and other cognitive activities. In fact, Bayesian theorem and its derivative methods also propose a large class of algorithms to realize human brain behavior.

The implementation layer solves the problem of which hardware is used to implement the corresponding algorithm. In the long evolutionary process, the earth creatures ultimately use neurons as computing units. For Bayesian algorithm, the digital circuit can be used as the implementation layer. Of course, the ability of random sampling and calculation needs to be enhanced by circuit design. Bayesian theory is an algorithmic framework for brain-like computing. According to the computational layer requirements of human brain cognitive behavior, corresponding algorithms can be constructed, then implemented by computer in software form, and special integrated circuits can be designed to accelerate the calculation.

The core of Bayesian computing is to solve the posterior probability. Its main method is Markov chain Monte Carlo (MCMC), because it is closer to the process of human brain perception. The idea of MCMC is to select random sampling according to likelihood probability, and make the Markov chain formed by sampling process converge to the posterior probability of the target. The form of the algorithm is not complicated, but the amount of computation can be enormous, especially in the high-dimensional data space, the convergence rate often becomes unacceptable. Among them, the main complexity lays in the need for a large number of samples, that is, a large number of iterations to achieve convergence.

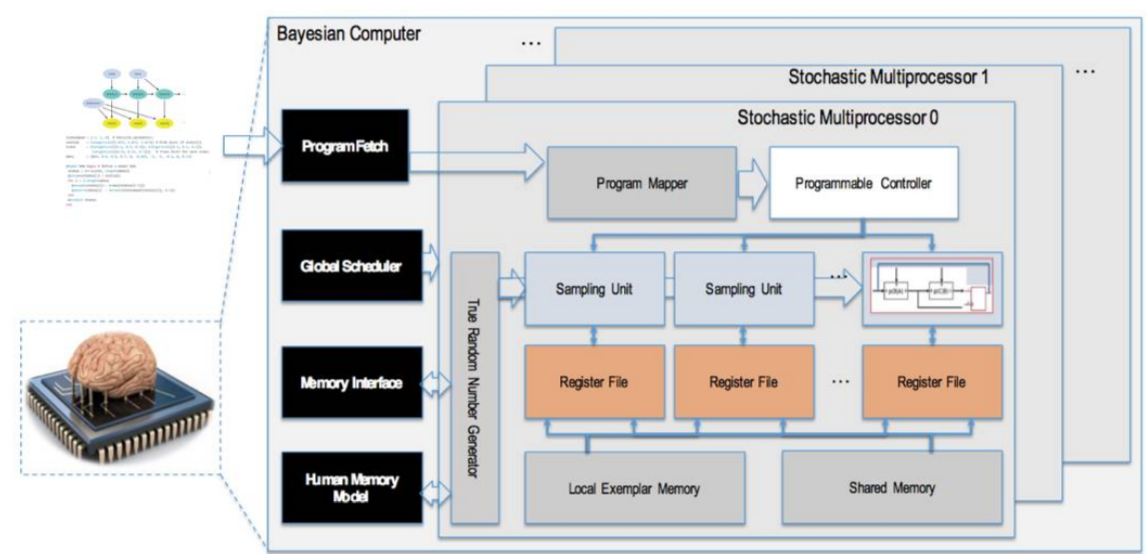
Bayesian computer. Bayesian computing is very important and computational complexity is very high, but the existing computers are not suitable for Bayesian computing. To solve this problem, only new hardware can be designed.



A circuit design method for random sampling using programmable unit of FPGA is presented. The figure is the original FPGA programmable cell, in which programmable logic is implemented with lookup tables. That is to say, the truth tables of logic functions to be implemented are stored in SRAM memory, and the combination of input variables corresponds to the address of finding the truth tables. In this way, only need to program the truth table, so that the programmable unit can realize any logic function, of course, at this time corresponding to the deterministic output of each input combination. In the figure, divide the input into two groups, one is the normal digital signal (in), the other is the random number generator (rand), and the corresponding truth table realizes the function Rand < in. In this way, the output is uncertain. Random number generators can only generate random numbers with specific probability distributions (such as uniform distributions). Through the above logic, can tailor such distributions (such as Rand < in) or combine them (such as adding two random gates that produce uniform distributions of 0 and 1 to form binomial distributions of random numbers). And a more varied probability variable is formed. The above description is similar for discrete distribution and continuous distribution of random variables, but floating-point operation units are needed as the basic circuits.

With the above stochastic combinational logic, can also introduce memory cells (such as D flip-flops) to form stochastic sequential circuits. By properly programming the circuit structure can construct the circuit structure for calculating MCMC. However, such a circuit can only be executed sequentially; that is to say, it only improves the performance of basic computing.

Of course, the above hardware can also be organized as processors. The following is a possible Bayesian computer architecture. The processor uses the random logic unit as the basic computing circuit, while the parallel MCMC is realized by programming. The basic element of probabilistic program is probabilistic variable, that is, the variable is no longer deterministic, but follows a certain probability.



Bayesian computers will serve as chassis for brain-like computing, allowing us to develop and experiment a series of cognitive methods, such as word learning, long-term and short-term memory models, induction, physical law learning, perception-motion integration, theoretical learning, and so on.

Summary

Bayesian network has been widely used since its emergence. It is necessary to identify, learn and extend bayesian networks. This paper first introduces bayesian theory, then reasoning and analysis of bayesian algorithm, and then the application of bayesian algorithm is analyzed.

According to the knowledge of bayesian network, it can be applied in the following fields:

1. Fault detection: bayesian algorithm has a high sensitivity in fault diagnosis and has a good application prospect in actual power system fault diagnosis.

2. Data processing field: as a common basic algorithm, bayesian statistics should not be underestimated. It plays an important role in machine learning. Especially in data processing, it has a good classification effect on the probability of event occurrence and event credibility analysis.

3. Medical field: bayesian method is used to diagnose the health status of patients, fully excavate useful information and help experts make diagnosis, which can improve the efficiency and accuracy of diagnosis. In the case of incomplete data, the cost of patients can also be judged by the existing expert diagnostic information. According to bayes' prediction results, patients' health status level makes patients more aware of their own health, and some healthy and unhealthy walk health warning effects, reminding patients to take care of your health, at the same time, doctors can receive treatment more accurately.

4. Data mining: with the development of database technology, the application of bayesian method in data mining is one of the hot issues in current research. Compared with other methods of data mining, bayesian method has the advantage of rich expression of probability information, synthesis of prior knowledge and solid mathematical foundation, which makes it more and more widely used.

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

At the same time, the bayesian network in its wide application also has a lot of problems, at the same time, there are many shortcomings of bayesian network. Only by continuous learning and improvement can bayesian network become more and more mature and widely used to bring convenience to our life.