BAYESIAN NETWORK AND ITS APPLICATION IN MAIZE DISEASES DIAGNOSIS

Guifen Chen[1,2] *, Helong Yu*[1,2,*]
[1] *Computer Science and Technology Institute, Jilin University, Changchun 130021, China*
[2] *Information Technology Institute, Jilin Agricultural University, ChangChun 130118, China*
[*] *Corresponding author, Address: Information Technology Institute, Jilin Agricultural University, 2888 XinCheng Street, ChangChun, 130118, P. R. China, Tel:+86-431-84532775, Fax:+86-431-84542775, Email: yuhelong@yahoo.com.cn*

Abstract: Bayesian network is a powerful tool to represent and deal with uncertain knowledge. This paper mainly introduces some technologies and methods of modeling Bayesian network, which are used in the building Maize Diseases Diagnosis system. In the construction of Bayesian network, noisy-or model and transformation from certainty factor to probability are used. Then maize disease diagnosis system based on BN is built by Netica (a BN software package). The practice proves that BN is an effective tool for maize disease diagnosis.

Key words: Bayesian Network; maize; disease; diagnosis

1. INTRODUCTION

There exists a lot of uncertainty phenomenon and problem. Uncertainty in agriculture is more extensive and complex. So, in order to create an effective intelligent system, uncertain knowledge must be dealt with.

From representation of uncertainty knowledge, there are two methods of dealing with uncertainty. One is rule-based method, and the other is model-based method. The advantage of rule-based method is that its computation is convenient, and its disadvantage is that its syntax is not systemic. The advantage and disadvantage of model-based method are contrary to the rule-based method.

From measurement of uncertainty, the methods of dealing with uncertainty are fuzzy theory and probability theory. Fuzzy theory mainly deals with vagueness, and probability theory mainly deals with randomness.

Therefore, Bayesian network in the article is a model-based probability method.

For the uncertainty in agriculture, there are some good model and application, but mainly rule-based. This method is adapted to knowledge represented by rule. However, uncertainty in agriculture is various. It is not adequate for rule to represent this uncertainty.

In order to solve it, Bayesian network is introduced. It widens knowledge representation that increases the reliability of expert system.

Bayesian network is a combination of probability theory and graph theory. Study of Bayesian network originates from the 1980's. Since 1990's, its study and application have stirred great concern. Compared with rule based method, the syntax of Bayesian network is clearer, which can reason in dual direction and can be constructed and debugged rapidly. The disadvantage of Bayesian network is that the computation complexity is high.

This paper mainly introduces the application of Bayesian network in maize disease diagnosis system.

As far as the computing complexity is concerned, in the Bayesian network construction, noisy or technology are adapted to simplify network structure and condition probability table.

On running the system, we find that the result is conformed to domain expert. It proves that it is effective to use Bayesian network to represent and deal with uncertain knowledge in agriculture.

2.   BAYESIAN NETWORK

2.1   Bayesian Network Syntax

BN= (Structure, CPT)
(1) Structure contain nodes and arcs
Nodes: random variable.
- Nodes can be continuous or discrete.
- Nodes can have two states or more.
- Nodes can be deterministic or nondeterministic.
Arcs: relationships between nodes.
- Arcs represent causal relationships of nodes.
- Arc between x and y represents that x has direct causal influence only.
(2) CPT: Condition Probability Table
- Each node has condition probability which is stored in a table(CPT).

- Value in table is *P(X_i|parents(X_i))*, *parents(X_i)* is the set of parent nodes of X_i.
- Root node is particular, as it has no parent node and has only prior probability: *parents (X_i) = Φ*, so *P(X_i|parents(X_i))= P(X_i)*.

2.2   BN Semantics

- Local semantic: represent conditional independence in the net
- Global semantic: represent global probability distribution

$$P(X_1...X_n) = \prod_{i=1}^{n} P(X_i \mid Parents(X_i))$$

We can conclude that Bayesian Network is combination of network structure and CPT, or global probability distribution is combination of conditional independence and local probability.

3.   BN BUILDING

Before being deduced, Bayesian network must be constructed. As we know, Bayesian network has two parts: structure and CPT, so the process of constructing Bayesian network is to construct structure and CPT(David J.Spiegelhalter,1993).

There are three methods to construct Bayesian network: manual construction, machine learning and combination of them. This article mainly introduces manual method, which constructs Bayesian network by domain expert elicitation(E.Charles, J.Kahn, etc,1997).

3.1   Elication of BN Structure

In this process, variables and relationships between them should be determined.

First, select variable set. It is important to limit the number of variables. So, it is necessary to choose important variables which are

- Query variables: or object variables, they are outputs of net and what we want to know.
- Evidence variables: or observation variables, they are inputs of net and used to reason states of query variables.
- Context variables: or middle variables, they are used to connect query variables and evidence variables.
- Controllable variables: or adjustable variables, they are used to control and adjust net.

If objects are mutex, they can be states of a variable, else being various

variables. Arc cause$_i$->effect represent cause$_i$ is one of cause for effect. There are two cases:

- Multi-causes, one effect.
- One cause, multi-effects.

According to the formula

$P(X_1...X_n) = P(X_n|X_{n-1},...,X_i)P(X_{n-1} | X_{n-2},...,X_1)....P(X_2 | X_1)P(X_1)$, we can find that the

$$= \prod_{i=1}^{n}(X_i | Parents(X_i))$$

right sequence of adding nodes is:

- Add root nodes.
- Add nodes that be influenced directly by root nodes.
- Repeat the above two steps until leaf nodes are added.

3.2   Elicitation of Condition Probability Table

There are three kinds of probability, namely objective probability, frequent probability and subjective probability, which originate from data, domain experts and literature.

In this process, the state of each variable and qualitative probability should be determined. This can be obtained by domain expert and literature.

In order to decrease the size of network, state number should be limited. In the meanwhile, states should be mutex.

Generally, probability given by domain expert is qualitative, so the transformation from qualitative probability to quantitative probability is necessary (Table 1).

Table 1. The transformation from qualitative probability to quantitative probability

Qualitative Probability	Quantitative Probability
always	0.99
generally	0.85
ofen	0.78
usually	0.73
Not ofen	0.50
sometime	0.20
occasionally	0.15
Usually not	0.10
seldom	0.30

3.3   Two methods used in building maize disease diagnosis system

3.3.1 'Noisy-or' Technology

This model has three assumptions:
- Parents and child are Boolean variables.
- Inhibition of one parent is independent of the inhibitions of any other parents.
- All possible causes are listed. In practice this constraint is not an issue because a leak node can be added (a leak node is an additional parent of a Noisy-or node).

Now, we can have a definition of noisy-or:
- A child node is false only if its true parents are inhibited.
- The probability of such inhibition is the product of the inhibition probabilities for each parent.
- So the probability that the child node is true is 1 minus the product of the inhibition probabilities for the true parents.

For Fig. 1, we can get this formula:

$$P(F \mid H_1, H_2, ... H_n) = 1 - \prod_{i=1}^{n}(1 - p_i)$$
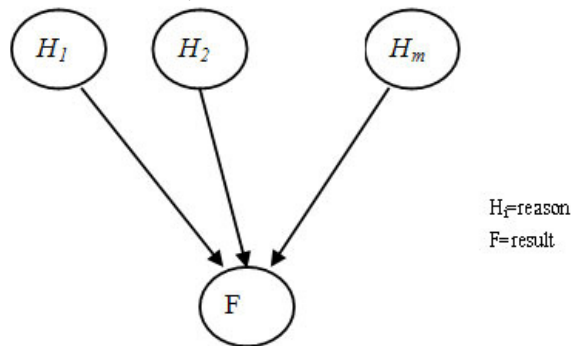

H_i=reason
F=result

Fig. 1: Noisy-or model

Generally, for node havii   parent nodes, if use 'Noisy-or', it needs O (k) parameters, if not, it needs $O(2^k)$ parameters. Obviously, BN is simplified.

3.3.2 Transformation from Certainty Factor to Probability of BN

H_i=reason
F=result

In the maize disease diagnosis, the knowledge given by domain expert is rule-based, and measurement for the belief is certainty factor, that is:

IF *A* THEN *B* *CF(B|A)*
Definition of certainty factor (CF):

$$CF(B|A) = \begin{cases} \dfrac{P(B|A)-P(B)}{1-P(B)} & ifP(B|A) > P(B) \\ 0 & ifP(B|A) = P(B) \\ \dfrac{P(B|A)-P(B)}{P(B)} & ifP(B|A) < p(B) \end{cases}$$

However, in the Bayesian network, uncertainty is measured by probability. So, in order to construct Bayesian network, it needs to transform CF to probability(F.trai. 1996; Kevin B.Korb, Ann E.Nicholson, 2006;Nevin Lianwen Zhang, 1996). From above formula, we can obtain:

$$P(B|A) = \begin{cases} CF(B|A)(1-P(B))+P(B) & ifCF(B|A) \ge 0 \\ (CF(B|A)+1)P(B) & ifCF(B|A) < 0 \end{cases}.$$

So, in order to get probability, it needs to know *P(B)*, which is prior probability of node *B*.

P(B) can be obtained from domain expert, literature, or assume *P(B)*=0.5.

4. IMPLEMENTATION OF MAIZE DISEASE DIGNOSIS SYSTEM BASED ON BAYESIAN NETWORK

Construcion of a Bayesian network for a domain problem needs communication and cooperation of Bayesian network expert, domain expert and BN software tool(P.J.F Lucas, 2005).

There are two types of nodes in the Maize Disease Diagnosis System，which are disease nodes and symptom nodes.

The disease nodes are Boolean variables, which contain states: 'happen' and 'unhappen'[P.J.F Lucas,2001; Radim Jirousck,1997], while Symptom nodes may contain multiple states.

This BN is a two-layer network, in which the upper layer is composed of disease nodes and the lower layer is composed of symptom nodes.

Obviously，the arc direction is from disease nodes to symptom nodes.

Fig.2 is a part of BN structure for maize disease diagnosis. In this structure there are four disease nodes and four symptom nodes, which corresponds to four diseases and four symptoms. The four diseases are maize dwarf mosaic, maize sheath blight, maize northern blight and bipolarismaydis.
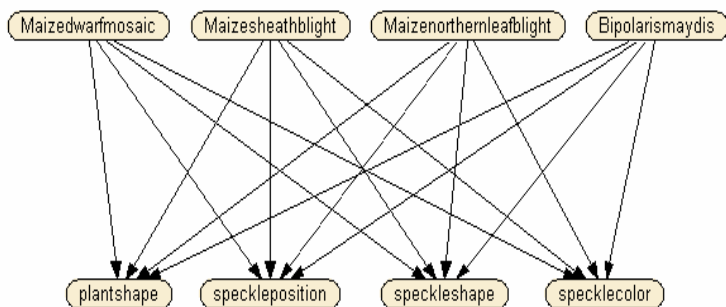
Fig.2: A part of BN structure

With the 'Noisy-or' technology and probability transforming from CF to probability, node's CPT is achieved. We can find the inference results from Fig.3, which are the posterior probability of disease when plant shape is normal, speckle position is lamina and speckle shape is others.
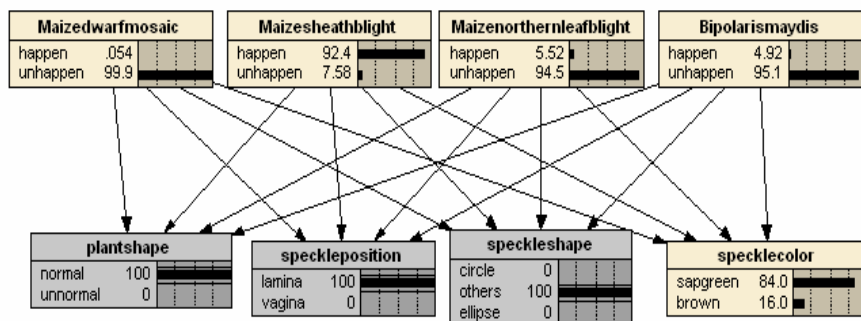

Fig.3: inference results of the BN

5.   CONCLUSIONS

BN is a strong tool for representing and dealing with uncertain knowledge. There exists a lot of uncertainty knowledge in maize disease diagnosis. So it is natural to use BN to build maize disease diagnosis system.

While building BN, Noisy-or model and transformation from CF to probability are used to decreasing network scale and simplify the network structure.

In running the maize disease diagnosis system, we find that the reasoning result is conformed with the solution given by domain expert, as proves that it is effective to use Bayesian network to represent and deal with uncertain knowledge in disease diagnosis.

Obviously, BN can be used in the diagnosis of maize pesticide, which will be done in the near future.

ACKNOWLEDGEMENTS

REFERENCES

David J.Spiegelhalter.1993.Bayesian Analysis in Expert Systems, Statistical Science, Volume 8, Issue 3: 219-247.

E.Charles,J.Kahn, etc.1997.Construction of a Bayesian network for mammographic diagnosis of breast cancer, Comut.Biol.Med: 19-29.

F.trai. 1996.A bayesian network for predicting yield response of winter wheat to fungicide programs, Computers and electronics in agriculture: 111-121.

Kevin B.Korb, Ann E.Nicholson. 2006.Bayesian Artificial Intelligence, CRC Press:.225-260

Nevin Lianwen Zhang, 1996.Exploiting causal independence in Bayesian network inference, Journal of artificial intelligence: 301-328.

P.J.F Lucas, 2005.Bayesian network modeling through qualitative patterns. Artificial Intelligence: 233-263.

P.J.F Lucas.2001.Certainty-Factor-Like structures in Bayesian belief networks, Knowledge-based systems: 327-335.

Radim Jirousck.1997.constructing probabilistic models, International journal of medical informatics 45: 9-18.