Decision Making in Uncertainty: A Bayesian Network for Plant Disease Diagnoses

4 authors, including:

S.M.Aqil Burney

Institute of Business Management

89 PUBLICATIONS **543** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project packages for radioactive materials View project

Project Welcome View project

Decision Making in Uncertainty:
A Bayesian Network for Plant Disease Diagnoses

S. M. Aqil Burney[1]
College of Computer Science
& Information System
IoBM University, Karachi, Pakistan
aqil.burney@iobm.edu.pk .l

Jawed Naseem[2]
SSITU, SARC
Pakistan Agriculture Research Council
Karachi, Pakistan

Abstract: *Decision making in uncertain situation with missing and partial truth is challenging in crop cultivation management like any other domain of application. In crop production attack of disease is a significant risk factor affecting yield and quality of crop. Rusts is one of the economically important fungal diseases of wheat difficult to diagnose in field condition due to ambiguity in classifying factors. Computer based soft computing methods can provide several intelligent solution for disease diagnoses in plant more precisely. In this paper a model for probabilistic decision making system in uncertain situation is discussed. The model is utilized to develop Bayesian Network for diagnoses of leaf, stem and strip rust disease of wheat. Proposed Bayesian network efficiently capture interdependence of classifying factor like color, shape and distribution of disease in different parts of plant along with removing uncertainty by employing conditional dependence. The proposed BN achieve upto 81% accuracy in wheat disease diagnoses.*

Key Words: Decision making, Uncertainty Wheat, Rust disease, Bayesian network,

I. INTRODUCTION

Decision making in uncertain situation with missing and partial truth is challenging in every field of science and agriculture has no exception. Disease attack during cultivation of crop is one of the major risk in crop cultivation management. Timely decision subject to prevailing environmental condition is required to control the disease and reduce risk. Rusts are economically important disease of wheat. Three distinct types of rusts, leaf rust, stripe rust and stem rust occur on wheat. The potential yield loss caused by these diseases depends on host susceptibility and weather conditions, but the loss also is influenced by the timing and severity of disease outbreaks relative to crop growth stage. The greatest yield losses occur when one or more of these diseases occur before the heading stage of development. Early detection and proper identification of disease is critical to disease management and control. Symptoms of various wheat diseases are so common that it is difficult to identify the disease appropriately without detail knowledge. Even the expert with sufficient knowledge can do mistake due to ambiguity in classification.

Computer based technologies can be utilized for decision making with ambiguous information. Artificial Intelligence (AI) is the area of computer science which focuses on developing machines and computer systems requiring intelligence like humans being[Iowa, 2006]. Using AI techniques and methods researchers are creating systems which can mimic human expertise in any field of science. Application of AI ranges from creating robots to soft computing models (softbot) that can reason like human expert and suggest solutions to real life problems. AI can be used for reasoning on the basis of incomplete and uncertain information and delivering predictive knowledge.

In AI several machine learning techniques and method can be employed to perform automated task which are difficult to perform manually. Bayesian networks are one of classifying technique which effectively employed in uncertain and ambiguous situation.

In this paper a mechanism is discussed to develop a probabilistic reasoning system for decision making in uncertain situation. The model is used to develop Bayesian Network for diagnoses of rust disease in wheat crop. Section II discuss challenges and issues in wheat disease diagnoses, Section III describes development of Bayesian network. Section VI discuss outcome of the experiment along with efficacy of the proposed system, Section V highlight future work.

II DYNAMICS OF WHEAT DISEASE DIAGNOSES & CONTROL

Wheat cultivation is associated with several risk posed by environment, economic stability and management of crop. One of the economically significant risks is disease attack during cultivation. Wheat is attacked by several diseases during cultivation including rust disease. The wheat rust

fungi are obligate parasite as they can grow and multiply in nature only on living plant tissue. Rust disease affect crop yield significantly due to their wide distribution, capacity to form new races that can attack previously resistant cultivars, ability to move long distances, and potential to develop rapidly under optimal environmental conditions [Wegulo, 2012]. Stem rust is capable of destroying entire wheat fields over a large area within a period of just a few weeks. The following are the important parameter in wheat rust disease [Table-1]

- Parts of plant infected:
- Shape and distribution of lesions
- Lesion color
- Degree of damage
- Tearing of Tissue

In different rust disease of wheat one or many of the above factors can be used to diagnose the disease.

	Leaf rust	Stem rust	Stripe rust
Pustule location	Leaf, mainly on the upper surface	Stem and leaf, upper and lower surfaces of leaf; occasionally on head and seeds	Leaf, upper surface; occasionally on head and seeds
Pustule arrangement	Single and random	Single and random	Stripes
Pustule shape and size	Round or slightly elongated; small to medium	Oval shaped or elongated; small to large	Round, blister-like; small
Tearing of host epidermis	Rare, visible with magnification	Conspicuous	None
Optimum temperature for infection	59-68 F	59-84 F	45-54 F
Optimum temperature for disease development	68-77 F	79-86 F	50-59 F

Table-1 Comparison of Wheat Rust disease [Wegulo, 2012]

Leaf and stripe rust can be distinguished by the color and shape of pustules and the location of the infection. However the symptom of these three types of disease has very slight variation which makes it difficult to distinguish one from another[Table-1]. Leaf rust pustules are orange brown in color, circular to oval in shape and chiefly found scattered on the upper surface of leaves. Stripe rust pustules are yellow-orange. Initially, the pustules are small and circular, but develop into yellowish stripes on the upper leaf surfaces, leaf sheaths and inside glumes.

III BAYESIAN BELIEF NETWORK

Bayesian network is a probabilistic graphical model used to represent knowledge system about a uncertain domain[Ben-Gal, 2007]. Any system having inherent uncertainty can be represented by Bayesian network. The simplest example of BN is a estimating probability of rain on a given day which is dependent on certain factor like temperature humidity and weather condition on last few days. In BN each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics. Bayesian Network is based on Baye's theorem which explains conditional dependence of one variable on other. The prior probability of event used to estimate posterior probability.

Formally, Bayesian network B is an annotated acyclic graph that represents a JPD over a set of random variables V. The network is defined by a pair

$$B = (G, \Theta)$$

where G is the DAG(directed Acyclic graph) whose nodes X_1, X_2, . . ., X_n represents random variables, and whose edges represent the direct dependencies between these variables. The graph G encodes independence assumptions, by which each variable X_i is independent of its non-descendants given its parents in G. The second component Θ denotes the set of parameters of the network. This set contains the parameter $\theta_{xi} \mid \pi_i = PB\ (xi \mid \pi_i)$ for each realization x_i of X_i conditioned on π_i, the set of parents of X_i in G. Accordingly, B defines a unique JPD over V, namely:

$$P_B(X_1, X_2, \ldots \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid \pi_i) = \prod_{i=1}^{n} \theta_{Xi} \mid \pi_i$$

If Xi has no parents, its local probability distribution is said to be unconditional, otherwise it is conditional. If the variable represented by a node is observed, then the node is said to be an evidence node, otherwise the node is said to be hidden or latent. The conditional independence statement of the BN provides a compact factorization of the JPDs. Instead of factorizing the joint distribution of all the variables by the chain rule is applied. The reduction provides an efficient way to compute the posterior probabilities given the evidence

Learning Bayesian Network

Bayesian network explicitly define the interdependence among variable of interest. In practical application learning Bayesian network is one of the crucial steps. The process involves learning topology or structure of network to depict causal relationship among variable and secondly estimating the parameter. Different approaches are used for learning the BN. The most common approach is learning through data and using expert knowledge. In this paper a hybrid approach is adopted. Expert knowledge is helpful in defining the structure while learning through data is effective for estimating the parameter. In learning through data a prior probability density function is assigned to each parameter vector and training data is used to compute the posterior parameter distribution and the Bayes estimates.

Probabilistic Reasoning through BN

The ultimate objective of developing BN is to inference the most probable outcome based on available evidence. BN is mathematically represented through JPD in a factored form which can be used to evaluate all possible inference by marginalization, i.e. summing out over "irrelevant" variables. Two types of inference support are often considered: *predictive support* for node Xi , based on evidence nodes connected to Xi through its parent nodes called *top-down reasoning*, and *diagnostic support* for node Xi , based on evidence nodes connected to Xi through its children nodes known as *bottom-up reasoning*.

The complexity of JPD increases with increasing number of nodes. Even if the variable have binary outcome JPD has size $O(2n)$, where n is the number of nodes. Hence, summing over the JPD takes exponential time. In general, the full summation (or integration) over discrete (continuous) variables is called exact inference and known to be an NP-hard problem. However, some efficient algorithms exist to solve the exact inference problem in restricted classes of networks. One of the most popular algorithms is the message passing using Junction Tree algorithm.

The junction tree algorithm [Kahle] is a method to extract marginalization in general graphs. In essence, it entails performing belief propagation on a modified graph called a junction tree. The basic premise is to eliminate cycles by clustering them into single nodes. The general problem here is to calculate the conditional probability of a node or a set of nodes, given the observed values of another set of nodes.

The basic concept in junction tree is clustering of predicted attributes. In belief updating instead of approximating joint probability distribution of all targeted variable (cliques) cluster attributes are formed and potential of clusters are used to approximate probability. So basically junction tree is the graphical representation of potential cluster nodes or cliques and a suitable algorithm to update this potential. Junction tree algorithm involve several steps as moralizing the graph, triangulation junction tree formulation, assigning probabilities to cliques, message passing and reading cliques marginal potentials from junction tree. Consistency in junction tree is a requirement which ensure that potential of a particular node with in two different cliques marginal probability of the node of interest is same .

IV   BN RUST DISEASE DIAGNOSES

The development of Bayesian Network of rust disease diagnoses is carried out through six tire processes as below;

i.   Identification of parameter/ variable of interest
ii. Identifying relationship, interdependence among variable
iii. Representing structure/topology of network through Directed Acyclic Graph(DAG)
iv. Estimating Conditional probabilities and joint probability distribution(JPD)
v.   Belief updating using junction tree algorithm by marginalizing/ factoring JPD
vi. Inference BN through message passing algorithm

We have used a hybrid approach for learning the network. In the first step expert knowledge together with technical detail of occurrence of disease is used to identify the variable of interest define the interdependence of various factors [Fig. 2] and their expected probability. The following factors are identified significant in diagnoses of disease [Table1]

- Parts of plant infected,
- Shape and distribution of lesions,
- Lesion color,
- Degree of damage of tissue
- Visibility of damage
- Occurrence of Disease(Common, Occasional, Rare)

- In the second steps parameter learning of conditional probability dependence of variable is determined using data. The collected data divided in two parts as learning and test data set. Individual

record selected randomly in two data set. However data contain replicate of all possible outcomes of identified variable. Open source tool BNsoft used for structure learning through data. In third step model generated reviewed by the expert.
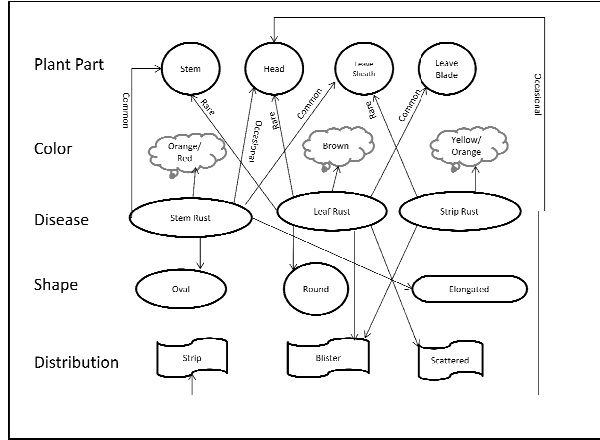

Fig. 2 Dynamics of Rust Disease Diagnoses

The developed Bayesian Belief network of rust (BBNRust) disease diagnoses depicted in Fig 2. The network efficiently estimates probability of occurrence of respective disease subject to instantiation of dependent variable. The network is capable of diagnosing the disease in case of missing instant of particular variable. The system can update the probability as soon as more information is available about variable.
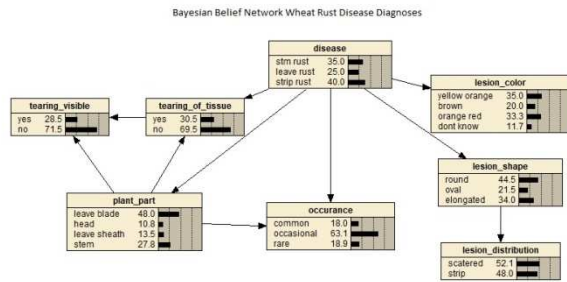
The JPD of the BN is given as under

P (D) = P(LC) x P(Occ) x P(LS/B) x P(ToT/TV)
        x P(PP/TV,TOT)
where
D = Disease, LC = Lesion Color, Occ = Occurrence
LS = Lesion Shape, LD = Lesion distribution
ToT = Tearing of Tissue, TV = Tearing visible
PP = Plant Part

Simplifying joint probability distribution to marginalize the require probability is carried out Using junction tree algorithm. The following sets of clique are formed;

Clique   Member nodes (* means home)
[Joined To]
 0 [1]   (disease, plant_part, *occurance)
 1 [0 2 3]   (*tearing_of_tissue, disease,
                  *plant_part)
 2 [1]   (*tearing_visible, tearing_of_tissue,
plant   plant_part
 3 [1 4 5]   (*lesion_shape, *disease)
 4 [3]   (*lesion_distribution, lesion_shape)
 5 [3]   (*lesion_color, disease)

II   RESULT AND DISCUSSION

Decision making in uncertain situation is a challenge particularly in plant disease diagnoses. Bayesian belief network proved to be an effective method for diagnoses of rust disease in wheat. The BN efficiently estimated conditional dependence of diagnostic parameters by capturing causative relationship between variable. Expert knowledge along with learning through data successfully identified underlying structure of the system.
We proposed a model (Fig 2) for developing a system for decision making in uncertain situation
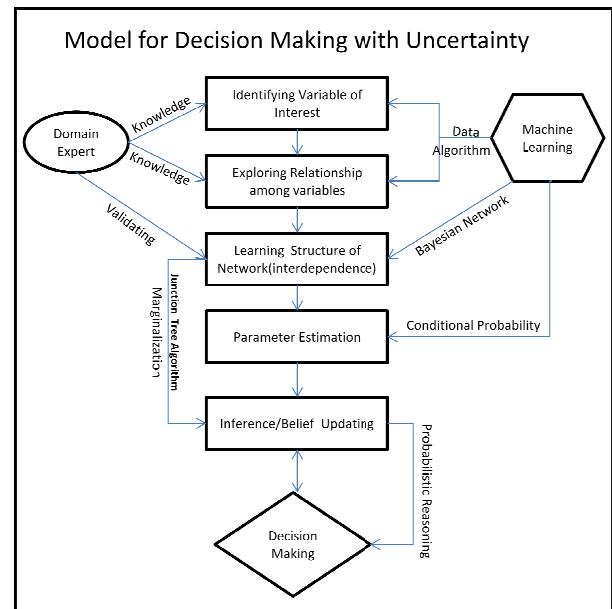
The mechanism is multi facet process involving domain expert as well as state of art computer based machine learning methods to develop the system.

We have proposed a hybrid system as in many situations it is difficult use expert knowledge alone or purely learning the structure through data. The hybrid approach ensures to capture the relationship which can not be distinguished with data. The model involves identifying variable of interest, exploring relationship and estimating parameter.

Employing the model BN for rust disease diagnoses is developed (Fig. 1). The BN diagnose the disease up to 81% accuracy. However variation exists among different kind of disease. The diagnoses of Stem Rust disease is more accurate as compare to other disease (Table 1)

Disease Diagnosed	Accuracy Rate (%)
Stem Rust	87.5
Leave Rust	78.3
Strip Rust	76
Over All	**81.3**

The proposed system is flexible as well as scalable. Bayesian network ensure inclusion of more variable of interest in rust diagnoses over period of time. Further the network can be extended for diagnoses of other plant diseases.

The overall accuracy of 81 % is not optimum main reason is the fact that shape and distribution of lesion is still posing confusion as human inspection may contribute to inaccuracy. The possible option is to use images recognition for distribution of lesion.

VI   FUTURE WORK

The limiting factor of the proposed network, as mentioned, is the more precise recognition of shape of lesion which can be achieved by image processing. Authors have plan to undertake research for incorporation of automated image recognition component in the proposed system.

REFRENCES

1. Ben-Gal, I , 2007, "Bayesian Networks", Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons
2. Burney, S. M Aqil, Nadeem Mehmood, 2012 "Generic Temporal and Fuzzy Ontological Framework, (GTFOF) for Developing Temporal-Fuzzy Database Model for Managing Patient's Data, Journal of Universal Computer Science, vol. 18, no. 2 (2012), 177-193
3. Burney, S. M Aqil,, Jawed Naseem, 2010, "Efficient Probabilistic Classification Methods for NIDS", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No. 8, pp 168-172
4. Iowa, Ames, 2006,"Arterial Intelligence: An Overview" Iowa State University, https:// pdfs.semanticscholar.org
5. Kahle, David et al, 2008, Junction Tree Algorithm Rice University", https://www.cs.helsinki.fi/u/bmmalone/probabili stic-models-spring- 2014/JunctionTreeKahle.pdf
6. Kolhe Savita, Raj Kamal, Harvinder S. Saini, G.K. Gupta, 2011,"A web-based intelligent disease-diagnosis system using a new fuzzy-logic based approach for drawing the inferences in crops", Computers and Electronics in Agriculture Volume 76, Issue 1, Pages 16-27
7. Magarey, R.D.; Travis, J.W.; Russo, J.M.; Seem, R.C. & Magarey, P.A. 2002. Decision Support Systems: Quenching the Thirst. Plant Disease, Vol. 86, No. 1, pp. 4-14,
8. Murphy, K. (1998). A brief introduction to graphical models and Bayesian networks. http://www.cs.ubc.ca/~murphyk/Bayes/ Inintro. html.
9. Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Francisco.
10. Wegulo, Stephen N. et al, 2012, "Rust Diseases of Wheat" Institute of Agriculture and Natural Resources, University of Nebraska–Lincoln, http://extensionpublications.unl.edu/assets/pdf/g 2180.pdf
11. Zhi Ping Ding, 2011, "The Development of Ontology Information System Based on Bayesian Network and Learning," Advances in Intelligent and Soft Computing", Volume 129, , Pages 401-406

AUTHOR'S PROFILE

Dr. S. M. Aqil Burney is Professor at College of Computer Science and Information Systems (CCSIS) at Institute of Business Management (IoBM) Karachi, one of the leading Business School in Pakistan. Dr. Burney was a Meritorious Professor (R.E.) and approved supervisor in Computer Science and Statistics by the HEC, Govt. of Pakistan. He was also the founder Project Director (UBIT) & Chairman of the Department of Computer Science, University of Karachi. He is also member of various higher academic boards of different universities of Pakistan. His research interest includes artificial intelligence, soft computing, neural networks, fuzzy logic, data science, statistics, simulation and stochastic modeling of mobile communication system and networks and network security, currently heading the detp. of actuarial science and risk management at CSCIS - IoBM. Teaching mostly MS(CS) Ph.D(CS))courses such as Data Warehousing, data mining & ML and information retrieval systems, fuzzy systems ,advanced theory of statistics, Markov chains and Financial Time Series.

Jawed Naseem is Principal Scientific Officer (RE) in Pakistan Agricultural Research Council. He has MCS & M.Sc (Statistics) and currently a Ph D. scholar department of Computer Science, University of Karachi, Pakistan. His research interest includes data modeling, machine learning, probabilistic reasoning, Information Management & Security and Decision Support System particularly in health care and agricultural research. He has experience in research & education at national regional (SAARC) and international level.