Discover Knowledge from Distribution Maps Using Bayesian Networks

Norazwin Buang
Research School of Information Sciences and Engineering (RSISE), Australian National University
Canberra, Australia
norazwin.buang@rsise.anu.edu.au

Nianjun Liu, Terry Caelli
National ICT Australia (NICTA)
Canberra Lab, ACT, Australia
nianjun.liu@nicta.com.au
terry.caelli@nicta.com.au

Rob Lesslie and Michael J. Hill
Bureau of Rural Sciences (BRS)
Canberra, Australia
rob.lesslie@brs.gov.au
hillmjdr@hotmail.com

Abstract

This paper applies a Bayesian network to model multi criteria distribution maps and to discover knowledge contained in spatial data. The procedure consists of three steps: pre processing map data, training the Bayesian Network model using distribution maps of Australia and testing the generalization and diagnosis of the model using individual states' maps. The Bayesian network that we used in this study is known as naïve Bayesian network. Results show that this environmental Bayesian network model can generalize the classification rules from training data for good prediction and diagnosis of a distribution map.

Keywords: Bayesian network, multi-criteria analysis, combining evidence, distribution maps.

1 Introduction

Analysis of spatial information in natural resource management is crucial to support a decision making process. However, with the advent of various technologies to acquire the data, analysis of multiple spatial data becomes a very challenging area. These technologies will produce the data with different accuracy and different resolution in the data. In spite of multi representations of spatial data, evidences of an area can be from different time intervals and different observer's views which make combination of those evidences complicated. Spatial data can have spatial attributes and non-spatial attributes. The former one is incorporated in spatial topological systems or relationships, and the latter is also called thematic data.

Environmental distribution map is crucial for decision support systems as it helps to monitor resource condition and also to identify the potential areas for investment. The information discovered from distribution maps is also contributing to applications such as land value determination, local and regional planning, pest and disease control, emergency response planning, agricultural productivity assessment and agricultural diversification (BRS, 2006). The complexity of a distribution map depends on the number of classes used to represent the data. It can be classified into as few as two classes or an infinite number of classes to represent the data. The more classes is used to quantise the data, the more precise the produced distribution map is, but also the more complexity of computation arises. In this study, we quantise the pixel values in a raster distribution map into five classes, where one colour is used to represent the associated class of a pixel.

In order to analyse the information contained in distribution maps, we need to discover as much knowledge as we can. Knowledge discovery on distribution maps includes combining multi criteria maps in order to extract general knowledge and interesting patterns from non-spatial attributes. The dependency between data that are usually uncertain make the analysis more complicated. Besides, different experts might have different opinions about the dependencies and factors involved in deriving any knowledge from spatial data. As a beginning, in this study, we focus on raster distribution maps as spatial data and discover knowledge from non-spatial attributes where we extract non-spatial attributes for each pixel in a distribution map. We apply Bayesian network model to do the generalization and diagnosis in predicting the class distribution of the data in a target map.

The aim of this study is to apply a Bayesian network in modelling multi criteria distribution maps. This paper provides a very basic introduction about Bayesian network. The technical details can be found in (Heckerman, 1995). The structure of this paper is as follows. We start in section 2, presenting several applications of Bayesian network in spatial data analysis. Section 3 describes the sources of the data used in this experiment. The process of the experiment is explained in section 4, including modelling distribution maps by naïve Bayesian network, data pre processing and model training and testing. The experimental results are discussed in section 5 and discussions are discussed in section 6. Finally, conclusions and future works are presented in section 7.

2 Literature Review

Spatial data can be derived from different type of sources. It has been used for discovering interesting knowledge and for making a decision. In order to do prediction and to produce a decision, the data needs to be represented in

Copyright (c)2006, Australian Computer Society, Inc. This paper appeared at the Australasian Data Mining Conference (AusDM 2006), Sydney, December 2006. Conferences in Research and Practice in Information Technology, Vol. 61. Peter Christen, Paul Kennedy, Jiuyong Li, Simeon Simoff and Graham Williams, Eds. Reproduction for academic, not-for profit purposes permitted provided this text is included.

a particular approach and one of them is Bayesian network. We will also describe several applications of Bayesian networks for prediction in this section.

A Bayesian network, also called belief net, is a directed acyclic graph (DAG) which consists of nodes to represent variables and arcs to represent dependencies between variables (Pearl 1986, Charniak 1991). Arcs or links also represent causal influences among the variables. The strength of an influence between variables is represented by the conditional probabilities which are summarized in a conditional probability table (CPT). Bayesian network is one of the graphical modelling techniques and it has been used widely in various applications including computer vision, medicine and spatial data analysis. There are two fundamentals idea in a Bayesian network according to De Vel et. al (2006). First, the notion of modularity where a complex system is decomposed into simpler parts and the second fundamental idea is their connections. The model also can deal with two main problems: uncertainty and complexity and therefore have an explanatory power for the modelling data.

In Bayesian networks, one can predicts the target values or the missing values given the model and other evidences. The class value for each pixel in a target node is obtained by finding the class associated with the highest posterior probability for that node. Bayesian networks work under the assumption of conditional independence where it estimates the posterior probabilities of the classification occurred in the training data. Bayesian networks have several advantages for data analysis. First, it can handle situations where some of the input data are missing. This is a great advantage, as having incomplete data is unavoidable in real world applications. Second, there are a few algorithms that have been developed for both structure and parameter training to learn the Bayesian networks from data. Third, Bayesian networks can be extended to model the structured data. This method is called probabilistic relational model (PRM) (Friedman 1999)

Representing spatial data using Bayesian networks for prediction has been applied successfully in many applications. Margaret et. al (2005) modelled satellite images using a Bayesian network for estimating leaf area index (LAI). The network was evaluated on a per pixel basis and the predicted results showed better classification than other classifiers such as neural networks and spectral vegetation indices, one of the pixel classifier approaches. On the other hand, Stassopoulou et al. (1998) used a Bayesian network to infer the risk of desertification of some burned forest in the Mediterranean region by combining several related evidences. The evidences used were from various sources with different resolution and accuracy. The network was also evaluated on a per pixel basis for modelling the data that came from different resolution.

Swayne (2004) used a Bayesian network and extended it into an influence diagram for multi objective modeling and decision support for a nonpoint source pollution model in Southern Ontario, Canada. They adjusted the conditional probability values of a Bayesian network to get a better decision based on the specific criteria preference. Besides prediction, the ability of Bayesian networks in detection applications has been successfully developed and the details can be found in Stassopoulou et al (2000) for building detection and Sebe et al. (2004) for skin detection.

3 Spatial Data Source

The data used in this study is acquired from the Multi Criteria Analysis Shell for Spatial decision support system (MCAS-S) provided by the Bureau Rural Sciences (BRS) of Australia (Michael et. al, 2005). Figure 1 and 2 show the distribution maps used to train the Bayesian network model. Figure 1(a) – (d) describe the distributions maps of population total, elevation, taxable income and accessible/remoteness index of Australia (ARIA) for 2001. The data has been quantized into five different classes and the scope for each class is also shown in the figure. One colour is used to represent a class while white pixel belongs to the background. These distribution maps are in a raster format where each pixel is associated with either a class or background.

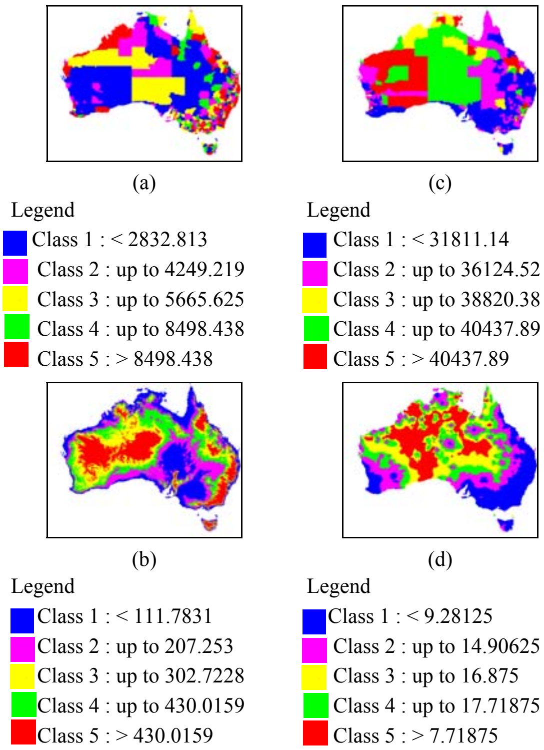
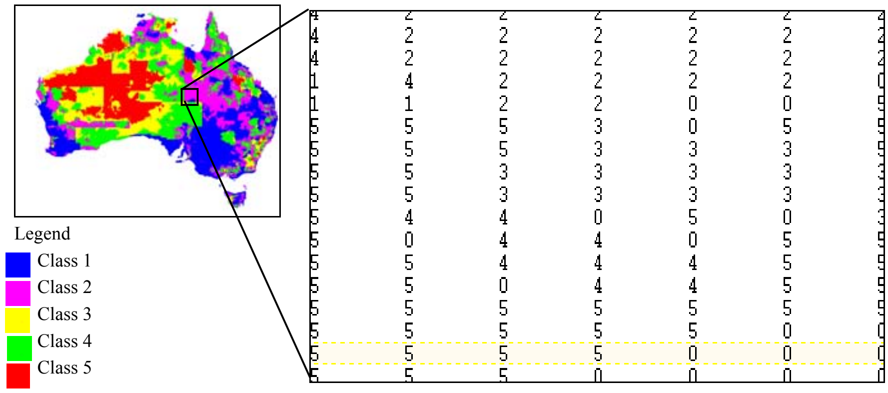


Fig. 1. Distribution maps of Australia (a) population total, (b) elevation, (c) taxable income, and (d) accessible remoteness index of Australia (ARIA).

In addition to the data shown in Figure 1, we are also given a distribution map of development potential as shown in Figure 2. Data of development potential distribution is also quantized into five different classes and the same colours are used to label the classes in this map. Figure 2(a) shows the distribution map of development potential of Australia while Figure 2(b) is a set of classes corresponding to the specified window in the map. All these five distribution maps are used in parameter training. The detail of the training process is discussed in section 4.



(a) (b)

Fig. 2. (a) Distribution map of development potential of Australia (b) pixels's classes in the specified region.

Besides the distribution map of the whole Australia, we are also provided with the distribution maps of the same criteria (population total, elevation, taxable income and ARIA) for the individual states in Australia (Queensland, Victoria, New South Wales, Australian Capital Territory, Western Australia and Adelaide). In this study, Victoria distribution maps are used for testing the Bayesian network model and the details are discussed in section 5.

To get the class value of each pixel, we pre processed each distribution map using Matlab and Java. The size of the distribution map image is 270 pixels height and 340 pixels width. Therefore, for each image, there are 91800 pixel values. However, only some parts of the image contain class values while others are just the background. Section 4.1 describes the process in details.

4 Bayesian Network (BN) for Modelling Spatial Data

4.1 General Framework

Figure 3 shows the flow chart of the experiment. The first step is data pre processing. The RGB values of each pixel are extracted from a raster distribution map and for each of them a class number is assigned based on its RGB values. A class number '0' is set to a background pixel, while class '1' to '5' are assigned to five different states or classes for each node. In the second step, the pixel values in the distribution map of Australia are used for training the Bayesian network model to obtain the model parameters- the Conditional Probability Table (CPT). In this study, we assume the structure of the Bayesian network is known and we focus only on parameter training. The next two steps are testing. We apply an individual state-Victoria distribution maps to test the performance of the proposed Bayesian Network Model. In the first experiment, we test the trained model to infer the distribution map of development potential of the whole state of Victoria, while in the second experiment we apply the model to infer the Victoria elevation distribution map. The proposed Bayesian network model will be discussed in details in the following subsections.

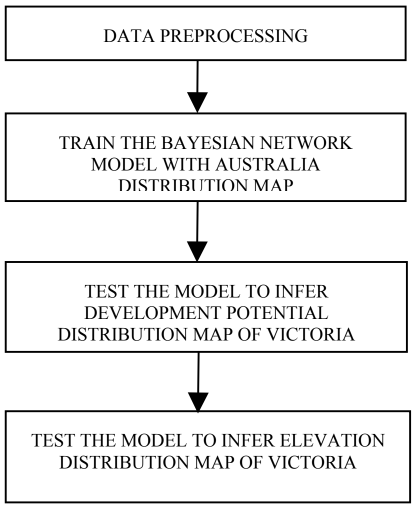


Fig. 3. General framework of the experiment

4.1.1 Modelling structure.

In the Bayesian network model used in this experiment, each distribution map is represented by a node. Figure 4 shows the structure of the Bayesian network. The *Development Potential* node is a child node for *population*, *elevation*, *tax-income* and *ARIA*. *Population* node, *elevation* node, *tax-income* node and *ARIA* node are parent nodes for *Development Potential* and also known as root nodes. There is no link between root nodes. Each node has five states that are 'vlow', 'low', 'med', 'high' and 'vhigh' associated with five classes in the distribution map, and the state's value corresponds to the probability of that class in the map. Class '1' corresponds to state '*vlow*'(very low), while class '2' corresponds to state '*low*'. Class '3' is the state '*med*' (medium) and class '4' equals to state '*high*'. Finally, class '5' corresponds to state '*vhigh*' (very high).

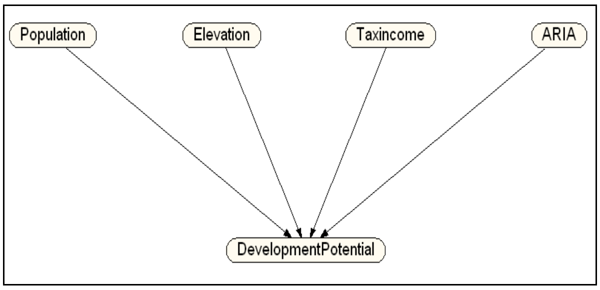


Fig. 4. The proposed Bayesian network model

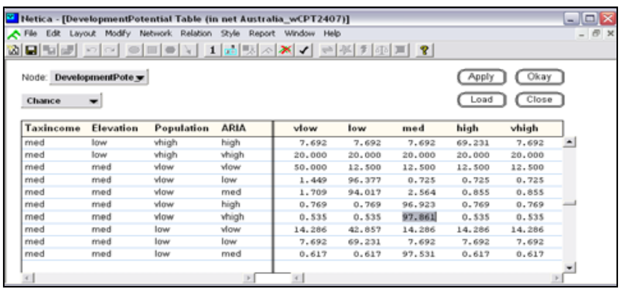
4.1.2 Training data and Trained Model

The non-spatial attributes from the distribution maps of Australia are used to train the model. We only select the value of each pixel that contains the available class for making up the training dataset as the format shown in Figure 5. Norsys's Netica Java software toolkit is used to train the Bayesian network. The data is incorporated into the network with *'.cas'* file format. There are six columns in the file including the IDnum. The .cas file is created by using Java codes.

Idnum	Population	Elevation	Taxincome	ARIA	DevelopPotential
1	vhigh	vlow	low	vhigh	med
2	med	low	med	low	med
3	vhigh	vlow	vhigh	med	low
:					
:					
:					
35881	low	high	vhigh	vlow	low
35882	vhigh	vlow	vlow	vhigh	vlow

Fig. 5. Format of the input data into Bayesian network (.cas file)

After training the model, a conditional probability table (CPT) is assigned to children nodes while prior probability is assigned to root nodes. According to the structure, *Development Potential* has a CPT while all other nodes have prior probabilities. As we are using the Netica learning algorithm, the prior distributions are Dirichlet functions. The *Development Potential* node consists of five states and four links directed into this node has $5^4 = 625$ rows in its CPT. Figure 6 shows parts of the CPT for *Development Potential* node derived from Netica (2006).



Taxincome	Elevation	Population	ARIA	vlow	low	med	high	vhigh
med	low	vhigh	high	7.692	7.692	7.692	69.231	7.692
med	low	vhigh	vhigh	20.000	20.000	20.000	20.000	20.000
med	med	vlow	vlow	50.000	12.500	12.500	12.500	12.500
med	med	vlow	low	1.449	96.377	0.725	0.725	0.725
med	med	vlow	med	1.709	94.017	2.564	0.855	0.855
med	med	vlow	high	0.769	0.769	96.923	0.769	0.769
med	med	vlow	vhigh	0.535	0.535	97.861	0.535	0.535
med	med	low	vlow	14.286	42.857	14.286	14.286	14.286
med	med	low	low	7.692	69.231	7.692	7.692	7.692
med	med	low	med	0.617	0.617	97.531	0.617	0.617

Fig. 6. Conditional Probability Table (CPT) for *Development Potential* node.

4.1.3 An example of proposed Bayesian Network Model

Here, an example of the mode in the pixel level is explained. Figure 7 shows the posterior probabilities for each class in *Development Potential* node when the state of population node is '*very low*', state of *Elevation* node is '*medium'*, state of *Taxable Income* node is '*medium'* and state of *AccessRemoteIndexAustralia (ARIA)* node is '*very high'*. From this sample, we can see that the posterior probabilities from class '1' to class '5' are 0.53, 0.53, 97.9, 0.53 and 0.53 respectively. As a state '*med'* holds the maximum probability (97.9%), the class of the pixel is inferred as class 3, where green colour is set to represent that pixel in the target distribution map.

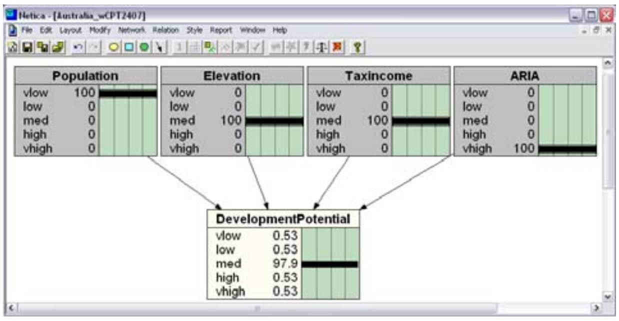


Fig. 7. Posterior probability inferred for each class of Development Potential node in Netica.

5 Experimental Results

This section presents the results of two testing experiments: generalization and diagnosis. The distribution maps of Victoria are used to test the model in both experiments. We compared on the values for each pixel. After the states' probabilities of each pixel are inferred, the pixel is set to the class/state with maximum probability and the pixel's colour is set accordingly. We only present the visual comparison in this paper for both testing experiments. We use the fastest known inference algorithm associated with Netica, a junction tree of clique algorithm for exact general probabilistic inference.

The first experiment is to test the model to infer class value of *Development Potential* node given other nodes (Population, Elevation, TaxableIncome and ARIA). Figure 8 shows the data and the result of the experiment. Figure 8(a)–(d) presents the data used as inputs to the model. Figure 8(e) is the development potential distribution map inferred by the Bayesian network model while Figure 8(f) is an empirical development potential distribution map, also as an expert knowledge. From the result, we can see that there are no big differences between the distribution map in Figure 8(e) and 8(f).

The second experiment is to infer the uncertainty—the class of the pixel with missing value of *Elevation* node given other nodes (Population, TaxableIncome, ARIA, and Development Potential). Figure 9 shows the data and the results of the experiment. Figure 9(a), (c), (d), (e) shows the data used as inputs to the model. Parts of the distribution map of elevation is removed and we used

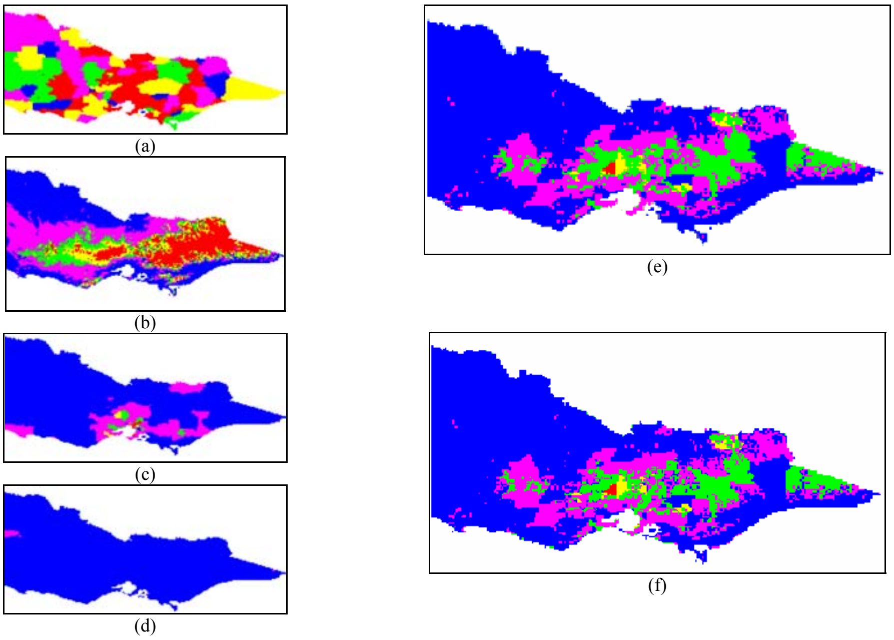


Fig. 8. Data and result of testing the generalization of the training data (a) Data Population (b) Data Elevation (c) Data Taxable Income (d) Data ARIA (e) Development potential distribution map inferred by the Bayesian network model (f) Empirical Development Potential

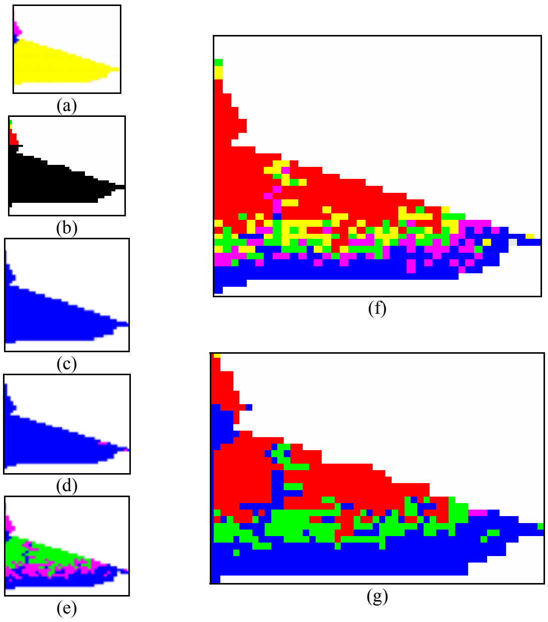


Fig. 9. Data and result of testing the diagnosis of the training data (a) Data Population (b) Data Elevation (missing data) (c) Data Taxable Income (d) Data ARIA (e) Data Development Potential (f) Data Elevation inferred by the model (g) Complete Data Elevation

black pixels to represent the missing values as shown in Figure 9(b). Figure 9(f) is the Elevation distribution map inferred by the Bayesian network model while Figure 9(g) is the elevation distribution map acquired from the MCAS-S application. More than 80% of pixel's colour matches quite well between the distribution maps in Figure 9(f) and 9(g).

6 Discussions

The use of Bayesian network in this context assumes that all data in distribution maps can be quantized into the same number of classes where each class has more or less the same number of pixels. In real world, this assumption might not be true. Moreover, we do not use a truncated function to limit the range in each class. For example, there might be a limit value for elevation that does not influence other nodes. This limitation can be solved if the expert opinions are included into the modelling process.

This problem is also related to the use of different modes between equal area distribution maps and equal interval distribution maps. In the future, it might be worth to understand how both modes affect the parameter estimation process in Bayesian networks.

The process to extract the class or state value for each pixel is not very accurate but still reliable. This is because of the difficulties to extract the values for each pixel especially at the boundaries between different states/classes. In this study, the class value for each pixel is only based on RGB values. It is quite hard to identify the exactly class value for each pixel that have similar colours or quite similar RGB values. As a result, some of the pixels are considered as missing if the program cannot distinguish the class values for that pixel.

This study only considered parameter training and the structure training is not included. There is a lot of research that focus on developing algorithms for structure training. In the future, we are planning to incorporate structure learning when constructing a Bayesian network. Since the dependencies between spatial environmental distributions are in diversity, the future work will include more complex model structure and model parameters' training. For the future, we are also planning to include learning algorithms when hidden variables are present. Despite all three problems discussed above, we are also facing difficulties to get more data for testing the model that we build.

7 Conclusions and Future Work

This paper describes how a Bayesian network can be used to model raster environmental distribution maps and their dependencies. It demonstrates that a Bayesian network is quite robust to discover knowledge from distribution maps, even though the present Bayesian network used in this study is only naïve Bayesian network. As their full potential has not yet been explored, we recommend it as one of the future works.

Acknowledgments The authors acknowledge Bureau Rural Sciences (BRS), Department of Agriculture, Fisheries, and Forestry for providing the data in the experiment. Financial support for this project from the Australian Research Council is gratefully acknowledged. National ICT Australia is funded by the Australian Government's Department of Communications, Information Technology, and the Arts and the Australian Research Council through Backing Australia's Ability and the ICT Research Centre of Excellence programs. The authors would also like to thank Mr. Nathan Brewer for his supports in reviewing and helping correcting this article.

8 References

BRS (Bureau of Rural Sciences) (2006): Guidelines for land use mapping in Australia: principles, procedures and definition.

Heckerman, D. (1995): A tutorial on learning with bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research (MSR).

Pearl, J (1986): Fusion, propagation and structuring in belief networks. Artificial Intelligence, 29, 241-288.

Charniak, E (1991): Bayesian networks without tears. AI Magazine. Vol. 12(4). 50-63.

N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. (1999): Learning probabilistic relational models. *In Proceedings of the Sixteenth International Joint Conferences on Artificial Intelligence* (IJCAI99).

De Vel, O, Liu, N., Caelli, T. and Caetano, T. (2006): An Embedded Bayesian Network Hidden Markov Model for Digital Forensics: *in IEEE Intelligence and Security Informatics Conference (ISI 2006)*, pp459-465, May 23-24, San Diego, USA

Kalacska, M., Sanchez-Azofeifa, G. A., Caelli, T., Rivard, B., Boerlage, B. (2005): Estimating leaf area index from satellite imagery using Bayesian Networks, *IEEE Transactions on Geoscience and Remote Sensing.*

Stassopoulou, A., Petrou, M., Kittler, J. (1998): Application of a Bayesian network in a GIS based decision making system. *International Journal of Geographical Information Science* 12 (1), 23±45.

Swayne, D., Jie Shi. (2004): Possible Courses: Multi-Objective Modelling and Decision Support Using a Bayesian Network Approximation to a Nonpoint Source Pollution Model. Complexity and Integrated Resources Management, *Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society*, iEMSs: Manno, Switzerland, :568-573.

Stassopoulou, A. and Caelli, T. (2000): Building Detection using Bayesian networks. *International J.Pattern Recognition Artificial Intelligence*, vol. 14, 715-733.

N. Sebe, I. Cohen, T. Gevers, T.S. Huang. (2004): Skin Detection: A Bayesian Network Approach. *International Conference on Pattern Recognition*, Cambridge, UK.

Michael J. Hill, Rob Lesslie, Andrew Barry, Simon Barry. (2005): A Simple, Portable, Spatial Multi-Criteria Analysis Shell – MCAS-S. In Zerger, A. and Argent, R.M. (eds) MODSIM 2005 *International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand* .

Netica Java API Library, http://www.norsys.com/. Accessed 1 August 2006.