Report Cover Page

ACERA Project
0905
Title
Combining GIS and Bayesian Networks in test-action strategies for risk assessment
Author(s) / Address (es)
Dr David Pullar, Brisbane, University of Qld Dr Carl Smith, Brisbane, University of Qld Dr Carmel Pollino, Canberra, ANU
Material Type and Status (Internal draft, Final Technical or Project report, Manuscript, Manual, Software)
Final Report
Summary

Environmental risk assessment involves making decisions that assess natural resource conditions relative to threats or disturbances. It is important to account for changing conditions over time due to natural processes, disturbances or remedial interventions. In this setting the spatial context and timing of decisions has to be represented and modelled. We develop models based upon Bayesian decision networks (Jensen 2001) to relate information and their uncertainties for doing environmental assessments and finding optimal decisions for management interventions.

The research is focused on methods and a software tool for practical environmental assessment problems. There are three distinct research outputs all linked by the use of Bayesian Networks for representing environmental problems with regard to information, its uncertainty, its assessment and decision processes. The outputs are:

i) Investigation of using Bayesian networks in environmental assessments that involve assessing natural resource conditions. This provides the rationale for linking Bayesian network models with spatial information and addresses aspects of how data is utilised in assessments. Two practical applications are presented for a land use risk model and vegetation condition assessment.

ii) A software tool was developed to link Bayesian network models to data in a GIS. It is used extensively in the previous investigation. The tool supports developing spatial models as Bayesian networks and the handling of data parameters and their uncertainties. Furthermore it identifies an archetypical way of combining mapping and monitoring data in environment assessment and natural resource applications.

iii) Investigation of Bayesian decision networks to find optimal solutions for management interventions in natural resource applications. We extend the use of BN's from a static model to a dynamic BN to account for changes over time. A practical application is presented to determine the best search and destroy strategy to manage invasive environmental weeds that disperse across a rural landscape over time.

ACERA Use only	Received By:	Date:
	ACERA / AMSI SAC Approval:	Date:
	ACERA / AMSI SAC Approval:	Date:

Executive Summary

Environmental risk assessment needs to confront uncertainties in the information used for natural resource assessment and uncertainties that arise in the assessment process itself; especially when the later involves judgements. Natural resource mapping information (for topography, soils, vegetation, etc.) is typically used in environmental risk assessment, but these information products have inaccuracies due to imprecision in mapping landscape features and in generalising data to map categories. Making judgements on the presence and impacts for environmental threats such as soil erosion, natural habitat destruction, etc., will introduce further uncertainties in relating information to risks. At the same time these criteria and decisions should be transparent and easily communicated to stakeholders. One approach for pursuing these aims is Bayesian networks (BN). BN's are capable of representing the causal assertions for information and their inherent uncertainty as belief probabilities. Structurally they are represented as a directed graph with nodes for random variables and arcs for relations between random variables. The BN graph structure can be saved, is easy to visualise and is a complete specification for a problem domain.

Many natural resource managers routinely use GIS in their work. GIS stores the states of objects and use spatial analysis to transform states into useful results. However it is difficult to store complex relationships and processes between objects in GIS. Hence we see an ideal match-up between GIS and BN's, one provides a model of the world and other provides data about the world

BN's are known for making probabilistic predictions, inferences, learning and decision making. A lesser known feature is they capture basic forms of causal reasoning for predicting and diagnosing situations. This project exploits this capability in two ways: i) to relate BN's to the data in a GIS in a way that encodes causal reasoning, and ii) to augment this same structure over time to capture state changes. The project has attempted to develop the best methodological basis for doing risk assessment with GIS and BN's.

The major outputs for this project are two methodological papers and a GIS software tool. Each of these outputs can stand on their own as research outcomes, therefore this report will not attempt to rehash their contents. Rather we describe the connections and gaps between these outputs as they relate to this project.

The contents for this report are brief. I will summarise the outputs, discuss events that arose over the course of the project that shaped the outputs, and briefly discuss future activities.

Contents

1. Summary of project outputs

The outputs for the projects are as follows:

i) Paper on "*Combining GIS and Bayesian Networks for environmental risk assessment and classification*".

This paper describes the rationale for applying Bayesian networks to risk assessment; in particular to risk characterisation and classification. There are many good texts available now describing the theory and practical use of Bayesian networks. However it does require some appreciation of probability theory and experience to build networks for environmental problems. The three investigators on this project ran an introductory 5 day course on the use of Bayesian networks for natural resource problems in 2008. One day was spent collecting field data for a vegetation survey and this was analysed on the final day to determine the ecosystem health of various vegetation communities sampled. Five groups worked independently to develop BN models of the problem but all groups used the same data. As expected each group produced a different model; but what was surprising was the huge variation in models with respect to their causal structure. This raises questions as to who was right, what model provides a more faithful account of the problem and its uncertainties, how to validate the models, and is there a more scientifically objective approach to structure model development? The answer to these questions rests in following a model structure that matches with the causal reasoning expressed in the graph language of a Bayesian network. In particular to appreciate (and master) the ways BN nodes are configured to capture predictive and diagnostic reasoning. A further appreciation of the roles nodes play as being root causes or final effects or cause-effect connections give greater insights on how to address problems with BN's. The roles for variables and their direction of causal reasoning are neatly categorised as: background, classification, mediator and observation nodes in the paper. Different texts use different terms but this structure is relatively universal for BN's. The paper attempts to relate this structure and associated reasoning paradigm to risk assessment problems. Two example applications are described; one on environmental risk characterisation and the other on vegetation condition assessment that incorporates impacts of anthropogenic disturbance to native vegetation. Both examples are frameworks within themselves but the paper demonstrates that their application needs to deal with inherent uncertainties in understanding real world phenomena. Importantly, the paper describes an archetype structure and practices to guide the way natural resource problems are modelled with Bayesian networks.

ii) Paper on "*Staying one step ahead of weed invasions"*

The paper applies a form of the archetype structure to a dynamic problem. It incorporates decisions in a BN and is aptly defined as a decision network. The paper is focused on spatial-temporal problems to find the optimal policy for surveillance-eradication to control a weed. We were very fortunate to collaborate with Queensland Biosecurity in getting data on an infestation with long term data surveys. In previous work we explored simulations models to map the extent and dispersal characteristics of Chilean Needle Grass (Fox et al. 2009). This new work explores the problem with a dynamic programming approach called a *partially observable Markov decision processes* (POMDP). The paper is very practical. It reviews the literature on characterising the impact of weed invasion and economic costs to control a weed. The paper builds BN models progressing from a static representation of the problem, to a temporal representation as a POMDP, and to a quasi-spatial POMDP. We focused on the interpretation of models and were pleased that the different progressions of BN's matched well with reported utility responses in the literature. The final model introduces a simplifying assumption to avoid doing a fully spatial POMDP, which would be hard to program and to interpret. This may be a finding in itself that using spatial pattern variables is sufficient to model the spatial interactions (when using probabilistic state transitions). We found the POMDP to be a very powerful paradigm for exploring spatial-temporal problems. It uses a particular form of optimisation that originates from dynamic programming where the solution to the problem is not an optimal state but rather a optimal series of decisions (or policy). We are enthusiastic about exploring this approach in future work, and in particular to try different optimisation strategies to address a range of policy approaches used in environmental management.

iii) Software tool and tutorial on combining GIS and Bayesian networks

This is a very tangible outcome following from the last. It develops a software tool in ArcGIS that implements the archetype structure in Bayesian networks for classification and assessment. In some respects, it can be thought of as a productivity tool to easily associate variables in a BN model with attributes in a GIS layer to do predictive reasoning. However it has been designed around the structure in the last paper and we believe will assist users to understand the semantics of problems in terms of the archetype categorises and their associated reasoning paradigms. We will attempt to assess if the tool and accompanying tutorial really assist users in understanding problems.

A basic feature of the tool is that it matches the different types of BN nodes to attributes in a GIS; this includes discrete/continuous types and probabilistic/deterministic kinds of nodes.

We will distribute the software from a University web site (http://www.gpem.uq.edu.au/CRSSIS/tools/) and set up procedures for getting feedback and bug reports. The software will be supported and we expect it to be reasonably popular as there is nothing equivalent that is available in the GIS marketplace or from research projects.

2. Discussion of activities

The project was affected by minor gaps and departures from the original proposal. Most the deliverables for the project are consistent with the proposal, but the emphasis did switch from decision networks to problems related to classification and impact assessment. This adjustment was a direct result of consulting with DAFF-BRS on a visit to Canberra in April. A brief summary of the outcomes of this visit are given in an appendix. We note here that upon visiting DAFF-BRS there was a much clearer direction of DAFF-BRS involvement by focusing on the problem of vegetation condition assessment and VAST. This was also an extremely interesting problem which we thought was well suited to our approach based on the archetype structure for BN's. We had hoped to explore how environmental impacts/transitions and ecosystem functions could be dealt with by a BN model. But time became too short to explore these possibilities. We will continue on this problem in the future and hope to have other outputs that apply the VAST framework in negotiating biodiversity offsets in a coal mining area.

We note that the work on mapping weeds has continued under a student project. Hence further results are expected on this aspect of the work.

We would also like to thank the DAFF-BRS collaborators for their helpful advice. In particular, Paul Pheoloung for his assistance in meeting with relevant people at DAFF in Biosecurity. We hope to continue our association with Paul on future Biosecurity projects that require a spatially explicit solution. We would also like to thank Richard Thackway and Rob Leslie for their time and providing direction for our work. We are keen to demonstrate our current and future findings to the Land and Forest Sciences Program in the BRS.

3. Future activities and directions

We greatly appreciate ACERA for funding this project. It allowed us to address important and practical problems facing management of natural resources. We plan to continue with these projects in various research settings. In particular we plan to support future development of the GIS-BN tool. The tool is an early version that will improve with time and hopefully an expanding user base. We are particularly excited about the possibility of improving the representation of uncertainties within GIS. As noted in the paper, different uncertainties can arise in a spatial context including uncertainties related to the purity of spatial attributes within land units. This relates to a classical problem in geography called the 'modifiable area unit problem" that we see would benefit from a fresh approach using probabilistic representations and reasoning.

Appendix: Outcomes of visit to DAFF-BRS Collaborators 6/4/2009

Meetings were arranged between DAFF-BRS collaborators to improve their involvement in the project. The intent of the meetings were to: i) describe the project and the progress to-date, ii) discuss potential for application of Bayesian networks for risk assessment of specific interest to DAFF. I gave presentations on the project and had informal meetings. The agenda for meetings clustered on two topics which I describe separately below.

Topic 1: *Progress to-date on test-action sequences to assess utility of surveillance and treatment decisions*.

Attended: Paul Pheulong, Debra Pedal, Michael Cole, Richard Thackway, Rob Leslie, Greg Hood.
DAFF Advice: There was general interest in approach. I was advised to use the modelling to develop take-home lessons on weed eradication. I was encouraged to continue interactions with Queensland Biosecurity and in particular to assess the minimum data requirements for surveillance.
Comments: There was not a strong sense that the project directly meets future objectives of DAFF. There was general interest but no clear directives, other than to continue to work with Queensland Biosecurity. This may also reflect changing priorities within DAFF in the area of Biosecurity.
Outcome: Emphasis in dealing with weed biosecurity was moved to a postgraduate thesis project. Bronwyn Kelly is collaborating with Biosecurity Queensland to predict the spread of Chilean Needle Grass along water courses in the Darling Downs. However, her timing for getting research results is outside this project so cannot be included in project report.

ii) **Topic 2**: Potential applications of Bayesian networks for risk assessment
Attended: Richard Thackway, Rob Leslie, Greg Hood
DAFF Advice: There was strong interest to apply BN's to the Vegetation, Assets, States and Transitions (VAST) framework developed by Thackway and Lesslie (2008). There were two specific suggestions:

a) Link existing land classifications (such as the ACLUMP, NVIS, etc.) with observation data (such as vegetation surveys, landcover mapping etc.) to improve VAST classification. The classification and its diagnostic criteria are explained in Thackway and Lesslie (2008).
   Thackway, R. and Lesslie, R. (2008) Describing and Mapping Human-Induced Vegetation Change in the Australian Landscape. *Environmental Management* 42:572–590
b) Action-utility modelling of how land management actions change a VAST class where the utility of the changes are assessed in terms of ecosystem services at a landscape scale (Yapp, Walker and Thackway, Unpublished manuscript).
   Yapp, G., Walker, J. And Thackway R. (Unpublished) Linking vegetation type and condition to ecosystem goods and services.

Comments: These discussions had a more definite direction for research. DAFF provided a number of research leads that they wanted me to explore. Their suggestions were well aligned with the project deliverable to develop an extension in ArcGIS that linked a Bayesian Network to environmental assessment.

Outcome: The first suggestion was realistic and consistent with the project deliverable, but it did require a change in application direction from earlier work. The second suggestion was also interesting, especially as it requires an action-utility decision network, but it was doubtful if anything could be completed in the project timeframe. Therefore we decided to focus on the first suggestion as a test case for software and methodological development, and as a precursor to developing a decision network for action-utility modelling.