Expert Systems with Applications 41 (2014) 4388–4400

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Applying a validation framework to a working airport terminal model

Jegar Pitchforth *, Paul Wu, Kerrie Mengersen

Queensland University of Technology, Gardens Point Campus, Australia

ARTICLE INFO

Keywords:
Bayesian Belief Network
Validation
Application
Airport
Systems model

ABSTRACT

Validation is an important issue in the development and application of Bayesian Belief Network (BBN) models, especially when the outcome of the model cannot be directly observed. Despite this, few frameworks for validating BBNs have been proposed and fewer have been applied to substantive real-world problems. In this paper we adopt the approach by Pitchforth and Mengersen (2013), which includes nine validation tests that each focus on the structure, discretisation, parameterisation and behaviour of the BBNs included in the case study.

We describe the process and result of implementing a validation framework on a model of a real airport terminal system with particular reference to its effectiveness in producing a valid model that can be used and understood by operational decision makers. In applying the proposed validation framework we demonstrate the overall validity of the Inbound Passenger Facilitation Model as well as the effectiveness of the validity framework itself.

1. Introduction

Expert-informed Bayesian Networks, or Bayesian Belief Networks (BBN) (Pearl, 1987) are a popular systems modelling approach in cases where the behaviour of a system is not entirely known, or is difficult to observe. In such cases, validation presents a challenge that cannot be answered using the commonly used goodness-of-fit tests, such as AIC, BIC or DIC (Gelman & Hill, 2007; Riedwyl, 1967). In particular, such tests require an objective and directly observed output that can be used to train and then test the model parameters. Such an arrangement is useful in areas where the system output can be observed, but the range of interactions producing such results are too complex to be thought about all at once, such as the technical performance of an information system (Moullec et al., 2013) or the financial performance of some organisation (Guo, Yeh, Wang, & Lin, 2012).

However, in many cases BBNs are used precisely because no such output is possible to collect, such as in ecology, risk analysis and social behavioural studies (Grover, 2013). In other cases, such as in airport passenger flows, there is a theoretically observable output but gaining such data is expensive or difficult, making BBNs a much more practical and realistic alternative for describing and predicting the behaviour of the system. This is similar to identifying the effect of a latent variable in the model (see Yet, Perkins, Fenton, Tai, & Marsh (2013) for methods of achieving this), but in these cases the latent variable is the output of the model. In such domains there is no known method of determining overall validity as there is no ground truth data against which model outputs can be compared.

In these cases the question of validation is often only addressed in passing through expert self-checking or otherwise is answered with an incomplete view of validity constructs. For example, Scholten, Scheidegger, Reichert, and Maurer (2013) and Stark, Roth, and Farry (2013) both use very sophisticated approaches to using expert elicitation in their models, but are limited to validating the results using expert opinion either through direct interview or expert-created scenarios. In such cases it is not useful to provide accurate validity diagnostics, as the test data cannot be seen as ground truth making the diagnostic misleading and may lead to criteria contamination, or self confirmation (Brogden & Taylor, 1950). In their model of IT project success, Gingnell, Franke, Lagerström, Ericsson, and Lillieskõld (2014) focus validation attempts only on defending the assumptions of the Noisy OR-gates used in their elicitation to reduce expert workload, which provides an incomplete assessment of model validity. This method of validating the model lends weight to the approach of building confidence in validity incrementally rather than a binary judgment using a single diagnostic measure based on comparison to ground truth data, however it is incomplete from the perspective of Pitchforth and Mengersen (2013) framework. Another approach is to generate synthetic data and compare the model output against this (Aquaro et al., 2010), although this approach also focuses solely on the predictive validity of the model and is essentially equivalent to a Qualitative Features Analysis.

* Corresponding author. Tel.: +61 403 961 878.
E-mail address: jegar.pitchforth@qut.edu.au (J. Pitchforth).

http://dx.doi.org/10.1016/j.eswa.2014.01.013
0957-4174/

Almost no work has been conducted on frameworks to address validation in expert elicited models when there is no ground truth dataset available, despite validity issues being raised by critics as an issue with using experts for BBN modelling (Drescher et al., 2013). However, for a model to be successfully implemented into everyday decision scenarios the user must be confident that they are receiving an accurate representation of the system they are controlling. If model developers cannot reach some assessment of the validity of their work it is most likely that decision-makers will ignore model results and continue working using traditional methods. In some cases this simply results in non-optimal system operations, however in the case of critical infrastructure the result can be the development and maintenance of unreasonably high risk protocols.

In response to the lack of methods for validating BN models when no output data are available, a framework of validity tests was introduced by Pitchforth and Mengersen (2013) for practical application to expert-elicited and expert-informed BBNs. These tests were drawn from a variety of disciplines such as statistics, psychology and system dynamics to give helpful guidance on the strength of a model's validity where no ground truth data are available against which model outputs can be tested. However, the framework is still only theoretical and has not been established as practically effective through application to a working model of a real system.

Here we apply that framework to the Inbound Passenger Facilitation Model (IPFM), a model designed to represent the inbound passenger facilitation system at an Australian international airport. This is the first application of the framework by Pitchforth and Mengersen (2013) to a working BN model that cannot be trained and tested using directly observed output data, and has been used to demonstrate the validity of the model to airport stakeholders. In order to achieve this many tests described in this paper have been further developed from the original paper describing the framework.

In using the framework to validate a working airport terminal model, we aim to first demonstrate that the framework can be usefully applied to models with little or no observable output data and how this can be achieved. Our second aim is to demonstrate that by being subjected to the validity tests in the framework the IPFM has become a more valid representation of the real-world system than prior to their application.

1.1. Background

The IPFM is a model of airport terminal behaviour with a focus on inbound passenger facilitation times. Initial model development was in response to work by Hargreaves (2008) that took a deterministic approach to developing a measurement framework (as opposed to a model) for this system. While their work was comprehensive, the lack of a coherent holistic model of the system in question meant the results of the work were never applied in practical operations management. In addition, the sampling strategy required to quantify their metric framework proved infeasible. While limited samples are taken from the airport throughout the year, the rarity of such sampling along with the acknowledged error of the measurements means that such samples are unlikely to be useful for model validation purposes. In this case, using expert opinion to support observational data is an important step in creating reliable and valid systems models.

At the point of conception there were a range of theoretical goals set out for the model, such as integrating disparate datasets that were being maintained by numerous stakeholders, capturing the knowledge of experts and predicting the performance of the system in different scenarios. After a search of existing airport terminal performance models a BBN was identified as a suitable modelling tool for achieving these goals. These goals were gradually refined over the course of the project through interviews and workshops held at each iteration of model development.

Once initial reporting has been completed it became apparent that significant validation testing would be required before the model could be approved for use in managing critical infrastructure. However, much of the model was unable to be tested against objective data because it was too expensive to collect so had been quantified using expert elicitation. This led to an exploration of similar situations in which no directly observable outcome is possible, and ultimately to the development of the framework outlined in Pitchforth and Mengersen (2013) and applied in this paper. For approval to be used in daily operations, these tests of validity needed to be usefully applied and communicated to stakeholders in order to build their confidence that the model is a valid representation of inbound passenger processing.

1.1.1. Bayesian Belief Networks

BBNs are a member of the family of conditional joint probability models known as Bayesian Networks (BN) (Pearl, 1987). These models express systems in terms of the likelihood of each factor (or node) existing in a given state based on the direction and strength of influence from other nodes. There are three main features of a BBN before it is used for simulation; structure, discretisation and parameterisation. In some cases researchers may obviate discretisation by using continuous nodes (John & Langley, 1995), but this is rare in practical applications, and continuous nodes are usually used in conjunction with discrete nodes (Aguilera, Fernández, Fernández, Rumí, & Salmerón, 2011).

In the process of creating a BBN model, the researcher must first define the domain and scope of the model and arrive at some understanding of the structure of the network. If full data are available then these can be used directly to learn the network structure algorithmically. Alternatively, Principal Component Analysis (Joliffe, 2002) can be used to reduce the dimensionality of the data before running learning algorithms, which speeds the learning process in time-critical applications. For expert-elicited and expert-informed networks the number and subject of nodes is defined by the researcher through literature review and expert consultation, as is the number and direction of arcs between nodes.

If discrete nodes are created from continuous assessments, the node must be discretised before parameterisation. Deciding upon discretisation thresholds is a difficult process, as the resulting output of the network can be highly sensitive to this choice. There is a significant amount of research on discretising nodes from data sets in the case of Learning Bayesian Networks (Monti & Cooper, 2013), however very little has been explored in the case of expert-elicited or expert-informed BBNs (Uusitalo, 2007).

The final stage in model creation is to set prior parameters for each node through a conditional probability table (CPT) that specifies the likelihood of a node's state conditional upon the states of its parent nodes. It is this parameterisation through CPTs that provides the simulation capabilities for BNs generally.

From this process there are seen to be four areas affecting uncertainty in the validity of the BBN model:

1. Structure: The nodes included in the model, and the number and direction of links between nodes.
2. Discretisation: The way the state space has been divided within nodes.
3. Parameterisation: The conditional probabilities associated with node states.
4. Behaviour: The output of the model under interrogation.