



Dr Xiongcai (Peter) Cai

Research Fellow
Centre for Health Informatics,
The University of New South Wales

x.cai@unsw.edu.au

Dr Xiongcai (Peter) Cai is a researcher with a general background in AI with expertise in the fields of machine learning, data mining, social network analysis and health informatics. He has been spending the past decade researching and developing models to better understand behaviours and patterns that relate real world human activities.

Modeling of time series health data using Dynamic Bayesian Networks: An application to predictions of patient outcomes after multiple surgeries

Xiongcai Cai^a, Oscar Perez-Concha^a, Fernando Martin-Sanchez^b, Blanca Gallego^a

^aCentre for Health Informatics, Australian Institute of Health Innovation, University of New South Wales, NSW 2052

^bHealth and Biomedical Informatics Centre, University of Melbourne, Victoria 3010

SUMMARY

Objective: To develop dynamic predictive models for real-time outcome predictions of hospitalised patients.

Design: Dynamic Bayesian networks (DBNs) were built to model patient outcomes that dynamically depend on patient's clinical profiles, temporal patterns of ward transfers and surgery data. These models were applied to predict remaining days of hospitalisation (RDH) for patients undergoing multiple surgeries and their performance compared against a static model based on Bayesian networks (BNs).

Datasets: Hospital data from a Sydney metropolitan hospital.

Results: The basic model uses static information at time of prediction. The DBN model uses static and temporal information extracted from a series of surgeries; DBNs show a significant improvement in patient outcome predictions with respect to the static model.

Conclusion: Time series health data can be dynamically modelled by DBNs to improve predictions of outcomes for patients undergoing multiple surgeries.

INTRODUCTION

Healthcare systems are under increasing pressure to identify strategies to improve the current patterns of care. Although unstructured data analytics have been widely reported in big data, the importance of time series has not been fully explored yet. Real-time prediction of patient outcomes could benefit greatly from big data in health and time series analysis. In particular, prediction of RDH¹ is an important indicator to assess healthcare delivery and hospital management. Unexpectedly long length of stay may negatively impact patients and hospitals in a variety of ways, such as higher costs and increased exposure to adverse events. Current methods do not allow real-time automated stratification of risk. Rapidly identifying those patients at highest risk of extended RDH has a great potential to improve the quality of care, reduce avoidable harm and costs. DBNs² are specially suitable to tackle this problem, since they are probabilistic graphical models that allow temporal order, which can better capture the dynamical nature of the healthcare delivery processes: prognosis, treatment selection, surgery and recovery.

In this paper, we aim to develop a DBNs-based prediction model to investigate the roles of time series data for the prediction of RDH for patients undergoing multiple surgeries.

DESCRIPTION

Medical records of patients who underwent consecutive surgeries at a Sydney metropolitan hospital between 2008-2012 were analysed. There are 5733 records in the dataset. Each admission is characterised by a set of attributes, which include: patient's characteristics (such as age), surgery information (main procedure, number of procedures, length of surgery) and ward type. These attributes, together with days already in hospital, constitute the inputs to the static BN. The outcome to be predicted is RDH. BNs^{3,4} are static probabilistic graphical models that consists of nodes and arcs forming a directed acyclic graph, where nodes present domain variables (predictors), whereas arcs represent conditional probabilistic relationships among variables. DBNs² represent the evolution of a system over time, allowing a fixed structure network to present variables at multiple time points (slices), containing temporal dependencies between slices.

We learned both structures and model parameters: 1) In order to construct the static BNs (Figure 1), we learned intra-slice structures and reinforced them with domain knowledge. These BNs will be the baselines to compare with the DBNs; at the point of prediction, the BN will contain the information of the current



surgery, whereas the DBN will contain the temporal information of the consecutive surgeries. 2) We then fixed the intra-slice structure and learned the inter-slice dependencies. We computed the conditional probabilistic dependencies between any two-time slices by creating successive two-slice sequences and by learning the DBNs (Figure 2). For testing, we unrolled the DBN to the length of test sequences (Figure 3) and input test sequences as evidences to infer RDH of patients.

In our experiments, we performed 5-fold cross validation in both the learned BNs and DBNs. RDH is discretised into 12 bins. Compared to BNs, DBNs achieved a significant improvement in the prediction of RDH. Specifically, DBNs achieved 72.4% prediction accuracy, whereas BNs 25.8%. This implies a 180% improvement, which might be due to the ability of DBNs to dynamically update the model using temporal information from time series data. It requires about 30 minutes to construct the DBNs with 64-bit Windows 7 Enterprise, 2 cores of Intel® i7-3840QM CPU @ 2.80GHz and 8GB RAM.

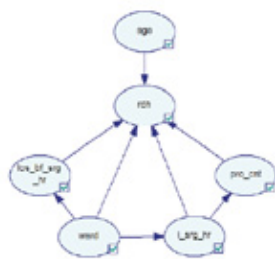


FIGURE 1. Static BN.



FIGURE 2. DBN.



FIGURE 3. Unrolled DBN.

CONCLUSION

We developed predictive DBNs models for predicting RDH of surgical patients using patient's trajectories and time series surgical information. Our experiments showed that DBNs significantly outperform BNs in RDH prediction after multiple surgeries. In the future, we plan to further apply DBNs in large-scale big data frameworks for health informatics.

Funding: This work was funded by National Health and Medical Research Council (NHMRC) Project grant 1045548 and Program Grant 568612.

Ethics approval: Ethics approval was obtained from the NSW Population and Health Services Research Ethics Committee and the NSW Human Research Ethics Committee. The corresponding author was responsible for the data analysis after the extraction. Its contents are the responsibility of the authors and do not reflect the views of NHMRC.

REFERENCES

1. V. Liu, P. Kipnis, M. K. Gould, and G. J. Escobar, "Length of stay predictions: improvements through the use of automated laboratory and comorbidity variables," *Medical care*, vol. 48, no. 8, pp. 739-744, 2010.
2. M. KP., *Dynamic Bayesian networks: representation inference and learning*, UC Berkeley, 2002.
3. L. P.J, v. d. G. L.C, and A.-H.A., "Bayesian networks in biomedicine and health-care," *Artificial Intelligence in Medicine*, vol. 30, pp. 201-214, 2004.
4. J. Pearl, *Causality: Models, Reasoning, and Inference*: Cambridge University Press, 2000.