Identifying Latent Variables in Dynamic Bayesian Networks with Bootstrapping Applied to Type 2 Diabetes Complication Prediction

CURRENT STATUS: UNDER REVIEW

BMC Medical Informatics and Decision Making BMC Series

Leila Yousefi
Brunel University

✉ Leila.yousefi@brunel.ac.uk *Corresponding Author*
ORCID: <https://orcid.org/0000-0003-1952-0674>

Mashael Al-Luhaybi
Brunel University

Lucia Sacchi
Universita degli Studi di Pavia Dipartimento di Studi Umanistici

Luca Chiovato
Fondazione Salvatore Maugeri Istituto Scientifico di Pavia Via Maugeri

Allan Tucker
Brunel University

DOI:

10.21203/rs.2.24145/v1

SUBJECT AREAS

Medical Informatics Endocrinology & Metabolism

KEYWORDS

Latent Variable; Diabetes; Dynamic Bayesian Networks; Time series Bootstrapping; Decision Making; Disease Complications

Abstract

Background: Type 2 Diabetes is a chronic disease with an onset that is commonly associated with multiple life-threatening comorbidities (complications). Early prediction of diabetic complications while discovering the behaviour of associated aggressive risk factors can reduce the patients' suffering time. Therefore, models of the time series diabetic data (which are often imbalanced, incomplete and involve complex interactions) are needed to better manage diabetic complications.

Aims: The aim of this work is to both deal with imbalanced clinical data using a bootstrapping approach, whilst determining the precise position of latent variables within probabilistic networks generated from the observations. The main motivation behind this paper is to stratify patient groups by means of latent variables to discover how complications in diabetes interact.

Methods: We propose a time series bootstrapping method for building Dynamic Bayesian Networks that includes hidden/latent variables, applied to a case for predicting T2DM complications. A combination of the IC* algorithm on time series bootstrapped data is utilised to identify the latent variables within a Bayesian model. Then, an exploration of inference methods assessed the influences of these latent variables.

Results: Our promising findings show how this targeted use of latent variables improves prediction accuracy, specificity, and sensitivity over standard approaches as well as aiding the understanding of relationships between these latent variables and disease complications/risk factors. The contribution of this paper compared to the previous papers in which time series bootstrapping is used for re-balancing the data and providing confidence in the prediction results.

Conclusion: Our results showed that our re-balancing approach by the use of Time Series bootstrapping method for an unequal number of time series visits demonstrated an improvement in the prediction performance. Additionally, the most highlighted contribution of this paper gained insight by interpreting the latent states (looking at the associated distributions of complications), which led to a better understanding of risk factors and patient-specific interventions: here the fact that the latent variable demonstrated that a patient falls into a sub-group that is hypertensive but not suffering from retinopathy.

Full Text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed.

However, the manuscript can be downloaded and accessed as a PDF.

Figures

Figure 1

IC*LS Diagram: The overall strategy of the proposed predictive model.

Figure 2

IC*LS graph: A Bayesian model of T2DM features. HV in this figure represents a hidden / latent variable.

Figure 3

Latent DBN Structure: Time series structures using the IC*LS approach and Fully Auto-Regressive dynamic links. The H, C, and O illustrate Hidden node, Comorbidity, and

Observed node, respectively.

Figure 4

Latent Variable Behaviour for predicting the onset of Liver disease: A latent prediction pattern of liver disease over time (a patient follow-ups).

Figure 5 Latent Variable Behaviour for predicting the onset of retinopathy: Latent variable prediction pattern of retinopathy over time (a patient follow-ups).

Figure 5

Latent Variable Behaviour for predicting the onset of retinopathy: Latent variable prediction pattern of retinopathy over time (a patient follow-ups).

Figure 6

Latent Variable Behaviour for predicting the onset of hypertension: Latent variable prediction pattern of hypertension over time (a patient follow-ups).

Figure 7 Bootstrap Confidence Interval: accuracy, sensitivity, specificity, and precision of liver disease prediction (Visit-based).

Figure 7

Bootstrap Confidence Interval: accuracy, sensitivity, specificity, and precision of liver disease prediction (Visit-based).

Figure 8

Prediction probabilities: The obtained posteriors for retinopathy, liver disease, and hypertension using Latent variable as the evidence.

Figure 9

A DBN Latent Model: From the left hand side, in the middle, and the right hand side demonstrate the patients history, the inferred latent variable probabilities, the prediction,

respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

[Supplementary_BMC_journal.pdf](#)