Bayesian Network for Predicting Functional Disability in Critically Ill Patients

First Student, Second Student, Third Student
Master's Degree in Artificial Intelligence, University of Bologna
{ Mohammad.Pourtaheri}@studio.unibo.it

July 8, 2024

Abstract

The aim of this study is to develop a Bayesian Network model to predict the functional disability (sfdm2) in critically ill patients based on various physiological, demographic, and disease severity variables. The model is constructed using the pgmpy library in Python, fitting it with data after preprocessing and discretization. The key concepts explored include Conditional Probability Distributions (CPDs), Markov blanket, and Inference within the context of the Bayesian Network. Our experiments show that a simplified model achieved significant accuracy improvements and can be used to assist in early decision-making and planning for end-of-life care.

Introduction

Domain

This study focuses on developing a Bayesian Network model to predict the level of functional disability (sfdm2) in critically ill patients using a dataset comprising 9105 patients from five US medical centers (1989-1994). The model aims to improve end-of-life care by assisting in early decision-making. The work is inspired by the necessity to address the loss of control patients experience near the end of life, as discussed in relevant literature.

Aim

The purpose of this project is to construct and experiment with a Bayesian Network to predict the functional disability of critically ill patients. We aim to handle missing data, discretize continuous variables, and simplify the model to fit within computational constraints. Additionally, we compare different methods of constructing and fitting the Bayesian Network to evaluate their performance and accuracy.

Method

We utilized four models in this study:

- **Model 1:** Intuitive Bayesian Network using Bayesian Estimator with Dirichlet prior.
- **Model 2:** Intuitive Bayesian Network using MLE Estimator.
- **Model 3:** Intuitive Bayesian Network using Bayesian Estimator with Advanced Scored Dirichlet prior.
- **Model 4:** Hill Climbing to find the best Bayesian structure, followed by Bayesian Estimator with Dirichlet prior.

Bayesian Networks are graphical models that represent the probabilistic relationships among a set of variables. The Bayesian Estimator learns the network parameters from the data, with Dirichlet priors incorporating prior knowledge. Maximum Likelihood Estimation (MLE) is used to estimate parameters by maximizing the likelihood function. The Bayesian Dirichlet equivalent uniform (BDeu) prior is an advanced scoring method used to optimize the network structure.

Model

Data Preprocessing

Data preprocessing involved dropping non-contributory variables, handling missing values, and discretizing continuous variables. This step was crucial to ensure that the data fed into the Bayesian Network was clean and appropriately formatted.

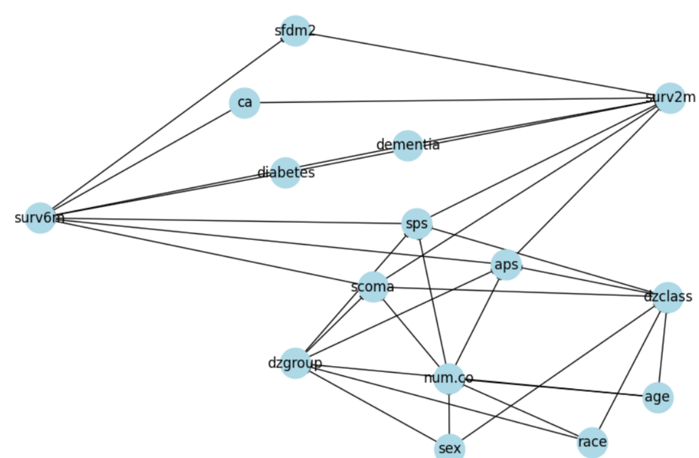

Figure 1: Initial Bayesian Network.

Initial Bayesian Network

The initial Bayesian Network structure was manually constructed based on domain knowledge and logical relationships between variables. The initial structure is shown in Figure 1.

Simplified Model

To address computational constraints, the initial model was simplified by selecting key variables while preserving critical relationships. This simplified model served as a baseline for further optimization.

Hill Climbing Optimization

The Hill Climbing algorithm was used to refine the Bayesian Network structure. This method iteratively adjusted the network to improve a scoring metric, resulting in a more accurate and realistic model. The optimized structure is shown in Figure 2.

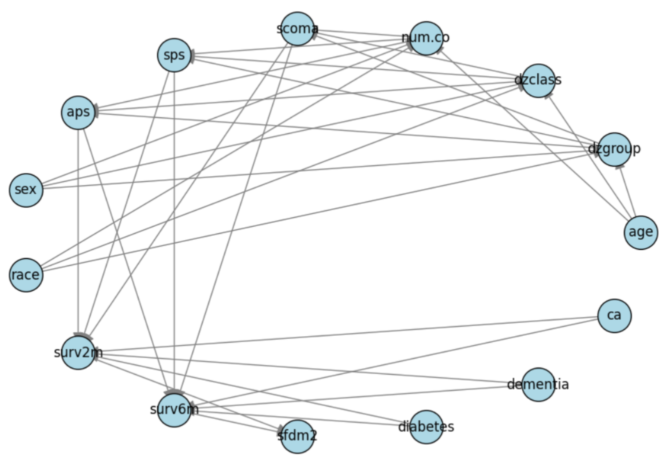

Figure 2: Bayesian Network optimized using Hill Climbing.

Results

Validation

The validation set was used to assess the model's performance. The initial model achieved a 50% accuracy on the validation set, while the Hill Climbing optimized model achieved 82% accuracy. This significant improvement demonstrates the effectiveness of the optimization process.

Method Comparison

Different methods were evaluated based on their ability to predict the target variable ‘sfdm2‘. The Bayesian Estimator with Dirichlet priors provided a good balance between complexity and accuracy, while the dirichlet prior with Hill Climbing optimization yielded the best results.

Analysis

Observations:

- **Dynamic Interdependencies:** The Hill Climbing model suggests more complex interdependencies between the variables.
- **Age and Dementia:** Age directly influencing dementia aligns with medical knowledge, showing that older age increases the risk of dementia.
- **Disease Group Influence:** ‘dzgroup‘ not only influences ‘dzclass‘ but also ‘ca‘ and ‘surv2m‘, reflecting the interrelated nature of disease severity and survival.
- **Survival Estimates Interplay:** ‘surv2m‘ influences ‘surv6m‘, ‘sps‘, ‘sfdm2‘, and ‘scoma‘, showing that short-term survival estimates impact both long-term survival and functional disability directly.
- **Physiological Measures:** Physiological measures (‘sps‘, ‘hrt‘, ‘aps‘, ‘crea‘, ‘resp‘, ‘meanbp‘, ‘temp‘) are interlinked, indicating that these factors collectively contribute to the survival estimates.
- **Health Conditions Interconnected:** Conditions like ‘diabetes‘ and ‘ca‘ have direct and indirect impacts on multiple nodes, highlighting their significant roles in influencing patient outcomes.
- **Complex Interdependencies:** The Hill Climbing model's structure reflects the complex, intertwined nature of medical variables. This complexity likely contributes to the higher accuracy as it better represents the underlying data patterns.
- **Direct and Indirect Influences:** By including both direct and indirect influences (e.g., ‘age‘ influencing ‘dementia‘, which in turn affects ‘num.co‘), the model aligns more closely with real-world medical scenarios.
- **Physiological Measure Interactions:** The inclusion of interactions between physiological measures (e.g., ‘sps‘, ‘aps‘, ‘hrt‘, ‘temp‘, ‘resp‘, ‘meanbp‘) allows the model to capture the collective impact of these variables on survival and disability outcomes.

Conclusion

This project demonstrated the feasibility of using Bayesian Networks to predict functional disability in critically ill patients. Despite initial computational challenges, the simplified and optimized models provided significant insights and reasonable accuracy. The Hill Climbing suggested structure provided a more accurate and realistic representation of the relationships between demographic, physiological, and disease-related variables. This improvement underscores the importance of dynamically capturing interdependencies and complex interactions in medical datasets to enhance predictive accuracy and utility in clinical decision-making. Future work will focus on further optimization, validation with larger datasets, and clinical integration.

Links to external resources

- Github: https://github.com/pm78p/SUPPORT2.DataSet.bayesian-network
- Dataset: https://archive.ics.uci.edu/dataset/880/support2

References