Discriminative Prediction of Adverse Events for Optimized Therapies Following Traumatic Brain Injury*

Priyantha Wijayatunga[1] and Nina Sundström[2]

[1] Department of Statistics, Umeå University, Umeå, Sweden
[2] Department of Radiation Sciences, Umeå University, Umeå, Sweden
{priyantha.wijayatunga,nina.sundstrom}@umu.se
https://www.umu.se/en/

Abstract. Traumatic brain injury (TBI) causes temporary or permanent alteration in brain functions. At intensive care units, TBI patients are usually multimodally monitored, thus rendering large volumes of data on many physiological variables. For the physician, these data are difficult to interpret due to their complexity, speed and volume. Thus, computational aids are recommended, e.g., for predicting patient's clinical status in near future. In this article, we describe a probabilistic model that can be used for aiding physician's decision making process in TBI patient care in real time. Our model tries to capture time varying patterns of patient's clinical information. The model is built by using a discriminative model learning framework so that it can predict adverse clinical events with a higher level of accuracy. That is, our model is built so that prediction of certain desired events are given more attention than that of the other less important ones. This can be achieved by estimating model parameters in such a way, for e.g. using a cost function, when a suitable model structure has been selected, that again can be done discriminatively. However, such estimation procedures have no closed form solutions, so numerical optimization methods are used.

Keywords: Dependence · Accuracy · Clinical · Real time.

1 Introduction

Traumatic brain injury is a type of head injury that causes temporary or permanent alteration in brain functions. It is one of the major health problems currently worldwide. Some of the main causes of TBI are automobile accidents, falls and sports accidents. TBI is a major cause of mortality and morbidity in younger people. In Sweden, according to media reports, from 15,000 to 20,000 TBI patients are hospitalized annually, rendering a huge number of hospital days and a great economic burden. A smaller portion of all TBI patients are emitted into intensive care units (ICU) due to the severity of their injuries.

* Supported by the County Council of Västerbotten and the Umeå University.

The clinical state of patients suffering severe TBI is often critical. These patients are multimodally monitored from the time of arrival to the ICU until their clinical state improves, or in the worst case, until the death occurs. An approximate mean treatment time in the ICU is two weeks. The setup for multimodal monitoring generally includes an extensive battery of high frequency physiological parameters such as intracranial pressure (ICP), arterial pressure (ABP), cerebral perfusion pressure (CPP), echocardiography (ECG), oxygenation, temperature and respiration. These clinical parameters are utilized to optimize the treatment of these critically ill patients. However, the information generated from such multimodal monitoring equipment is too complex and extensive to be handled by treating ICU physicians and neurosurgeons [2], thus computational aids are recommended. Currently there are no good means of effective use of such data for aiding the patient treatment process.

1.1 Aims and Objectives of Our Project

Our project is aimed at effective usage of these large datasets from TBI patients treated in the ICU. The goal is to offer a real-time tool to aid the patient treatment processes by generating individualized probabilistic predictions on patients' health conditions. In particular, we attempt to use advanced statistical and artificial intelligence (AI) techniques in our models to attain higher accuracies for predicting critical outcomes/events while overall prediction accuracies are also kept at high levels. Based on the vast amount of data collected in each single patient, we aim to develop an interactive tool to combine all physiological parameters and generate predictors of adverse events in a timely fashion as to be able to act upon them before they occur. Adverse events, associated with worse clinical outcome, that we set out to detect using an interactive tool are:

1. High ICP or low CPP over long coherent time periods, or in total over whole monitoring time.
2. Need for a high level of oxygenation ($> 50\%$) during treatment in the ICU
3. Need for treatment with a ventricular drainage (vdrain)
4. Pyrexia (core temperature $> 38.5°C$)

As the first step, we are developing simple probabilistic models for prediction of different clinical conditions of TBI patients in the ICU. Here we describe one such model including how it can be learnt to predict specific events, such as e.g. elevated ICP. First we give a brief overview of some of the current models described in the literature in the area of intracranial pressure monitoring.

1.2 Brief Overview on the Current Methods

Morphological features may provide insight to monitor and to understand ICP in an automatic fashion. Therefore, a probabilistic framework based on graphical models is used to track ICP peaks in real time exploiting temporal dependencies between successive peaks [6]. In a similar fashion, deep learning is used to model the relationship between intracranial hypertension and ICP waveform morphology to accurately detect presence of hypertension [5]. Furthermore, in one study computational methods have been developed to uncover causal structures of the brain physiological measures after subarachnoid hemorrhage [1].

2 A Simple Probabilistic Model

Here we present of simple probabilistic model that is a type of dynamic version of the so-called naïve Bayes model [8]. It can capture several time series in order to predict one time series into the future. If more than one series is needed, then multiples of such models can be used. Let X_t denote observation at time (discrete time points) t, for $t = 1, 2, ...$ and assume that it has no long-term trend. For simplicity we assume that the values are discrete. Then for p number of such time series $\{X_{1,t}\}, ..., \{X_{p,t}\}$ where the interest is to predict the $X_{1,t+1}, X_{1,t+2}, ...,$ we use the model

$$p(x_{1,t+k}|x_{1,[t:t-d_1]}, .., x_{p,[t:t-d_p]}) = \frac{p(x_{1,t+k}) \prod_{i=1}^{p} p(x_{i,[t:t-d_i]}|x_{1,t+k})}{\sum_{x'_{1,t+k}} p(x'_{1,t+k}) \prod_{i=1}^{p} p(x_{i,[t:t-d_i]}|x'_{1,t+k})}$$

where $k = 1, 2, 3, ...,$ integer $d_i \geq 0$ for $i = 1, ..., p$. Here $X_{i,[t:t-d_i]} = X_{i,t}, ..., X_{i,t-d_i}$ for $i = 1, ..., p$. Then the optimal prediction for the time $t + k$ is the posterior mode of conditional distribution of $X_{1,t+k}$ given $x_{1,[t:t-d_1]}, .., x_{p,[t:t-d_p]}$, i.e., $x^*_{1,t+k} = \text{argmax}_{x'_{1,t+k}}\, p(x'_{1,t+k}|x_{1,[t:t-d_1]}, .., x_{p,[t:t-d_p]})$. Note that d_i is moderately large enough to capture motifs for the time series, for each i. However when they are too large then, so is the conditional probability table $p(x_{i,[t:t-d_i]}|x_{1,t+k})$ for each i. Beside this simple model, we plan to implement other artificial intelligence tools such deep neural networks, decision trees, etc. in future.

It is obvious that the above model is just a version of the usual naïve Bayes model, where each feature variable is a time series of its most recent past few observations, including the class time series. Here the naïve Bayes assumption is that those time series are conditionally independent of each other given the present value of the class variable. So, it is clear that model structure is fixed since it is the nav̈e Bayes structure. But it is incomplete until when selection of time horizon into the past for each time series is done. When deciding the past time horizon the class time series, one simply need to find the Markov blanket of present value of the class time series with it. This can be implemented with general conditional independence tests and related algorithms such as $K2$ [3]. Similar tasks should be done with all other time series, separately. This is a simple approach and more complex way is to consider all the time series together. In addition we are also using the theory of vector autoregressive processes to build our network models [4].

After past time horizons of all time series have been decided, the parameters of the model have to be estimated. Common *generative* learning methods are maximum likelihood method, where the estimates are just empirical ratios of counts for respective sets of variables, and the Bayesian method, where the estimates are observed and hypothetical counts taken together. In *discriminative* learning methods, since the model is used to predict, say, X_0, accurately given data on a random variable vector, say, $X_{[n]}$, it is argued that, rather than maximizing the likelihood for $p(x_0, x_{[n]})$ (generative model) the condition likelihood for $p(x_0 \mid x_{[n]})$ should be maximized for the parameter estimation [7]. However this has no closed form solution (done only numerically). Sometimes, the data can be highly imbalance making it necessary to predict certain categories of the class variable more accurately, especially the minority categories than the majority categories. Such a case is presented in [8] where parameters are selected by optimizing a miss-classification cost function defined in such a way that the cost for an erroneous classification is higher for certain categories than the others. This is also a numerical optimization problem, often non-linear. However one major drawback of such a method is its tendency to over-fit the model to the training data. Furthermore, such discriminative learning method can result in model parameters that are not good for doing predictions with partial information on the feature variable set, that is one major advantage of probabilistic models such as Bayesian networks over regression models.

References

1. Claassen, J., Rahman, S. A., Huang, Y., Frey, H.- P., Schmidt, J. M., Albers, D., Falo, C. M., Park, S., Agarwal, S., Connolly, E. S. and Kleinberg, S.: Causal structure of brain physiology after brain injury from subarachnoid hemorrhage. PLoS ONE **11**(4): e0149878 (2016). https://doi.org/10.1371/journal.pone.0149878
2. Faltermeier, R., Proescholdt, M. A., Bele, S. and Brawanski, A.: Parameter optimization for selected correlation analysis of intracranial pathophysiology. Computational and Mathematical Methods in Medicine. **2015**, 652030 (2015). https://doi.org/10.1155/2015/652030
3. Murphy, K. P.: Machine learning: A probabilistic perspective. The MIT Press, Massachusetts (2012)
4. Nagarajan, R., Scutari, M. and Lèbre, S.: Bayesian networks in R: with applications in system biology. Springer, New York (2013)
5. Quachtran, B., Hamilton, R. and Scalzo, F.: Detection of intracranial hypertension using deep learning. In: Bayro-Corrochano, E., di Baja, G. S. and Medioni, G. (eds.) 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 2491–2496. IEEE Xplore (2016). https://doi.org/10.1109/ICPR.2016.7900010
6. Scalzo, F., Asgari, S., Kim, S., Bergsneider, M. and Hu, X.: Bayesian tracking of intracranial pressure signal morphology. Artificial Intelligence in Medicine **54**, 115–123 (2012)
7. Wijayatunga, P. and Mase, S.: Asymptotic properties of maximum collective conditional likelihood estimators for naïve Bayes classifiers. Technical Report B-432, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, Japan, (2006)
8. Wijayatunga, P., Mase, S. and Nakamura, M.: Appraisal of companies with Bayesian networks. International Journal of Business Intelligence and Data Mining **1**(3), 329–346 (2006)