International Journal of Medical Informatics 112 (2018) 6–14

Contents lists available at ScienceDirect

International Journal of Medical Informatics

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

Bayesian averaging over decision tree models: An application for estimating uncertainty in trauma severity scoring

V. Schetinin[a,*], L. Jakaite[a], W. Krzanowski[b]

[a] University of Bedfordshire, United Kingdom
[b] University of Exeter, United Kingdom

ARTICLE INFO

Keywords:
Bayesian model averaging
Decision tree
Predictive posterior distribution
Trauma and injury severity scoring
TRISS

ABSTRACT

Introduction: For making reliable decisions, practitioners need to estimate uncertainties that exist in data and decision models. In this paper we analyse uncertainties of predicting survival probability for patients in trauma care. The existing prediction methodology employs logistic regression modelling of Trauma and Injury Severity Score (TRISS), which is based on theoretical assumptions. These assumptions limit the capability of TRISS methodology to provide accurate and reliable predictions.
Methods: We adopt the methodology of Bayesian model averaging and show how this methodology can be applied to decision trees in order to provide practitioners with new insights into the uncertainty. The proposed method has been validated on a large set of 447,176 cases registered in the US National Trauma Data Bank in terms of discrimination ability evaluated with receiver operating characteristic (ROC) and precision–recall (PRC) curves.
Results: Areas under curves were improved for ROC from 0.951 to 0.956 ($p = 3.89 \times 10^{-18}$) and for PRC from 0.564 to 0.605 ($p = 3.89 \times 10^{-18}$). The new model has significantly better calibration in terms of the Hosmer–Lemeshow $\hat{H}$ statistic, showing an improvement from 223.14 (the standard method) to 11.59 ($p = 2.31 \times 10^{-18}$).
Conclusion: The proposed Bayesian method is capable of improving the accuracy and reliability of survival prediction. The new method has been made available for evaluation purposes as a web application.

1. Introduction

Decision making in health care is subject to uncertainties that exist in data and decision models. In this regard machine learning (ML) methods have been intensively developed over the last decade to provide practitioners with reliable estimates of uncertainties in decisions and predictions, see e.g. [1]. Using ML methods, predictions of functional recovery and mortality after traumatic brain injury were considered in [2,3]. The combination of Glasgow Coma Scale scores with clinical and laboratory parameters of patients has provided a high prediction accuracy. Prediction of burn patient survival was undertaken in [4] using models that were developed on patient data. The data included information about the patient's age, sex, and percentage of burns in eight parts of the body.

In trauma care the evaluation of injury severity of patients has a long history of using logistic regression modelling known as the Trauma and Injury Severity Score (TRISS) [5–8]. The TRISS model allows practitioners to predict the probability of survival for a patient on arrival at a hospital. The predictions are based on screening tests which are recorded at an accident scene. Screening tests are evaluated by a trained scorer for injuries which a patient can obtain in the following six regions of the body: head, face, chest, abdomen, extremities, and external (skin, subcutaneous tissue and burns).

Estimates of survival probabilities of patients enable practitioners to identify cases for peer review and compare the survival rates of different patient groups. TRISS estimates are also used for benchmarking and monitoring of patient outcomes over time and between hospitals [9,10].

Uncertainties that exist in data as well as in the prediction model affect the results and might lead to misleading decisions. For this reason, practitioners have raised a concern about the ability of TRISS to provide reliable predictions and estimates of uncertainty [9,6,11].

The accuracy of predictions is compared against actual survival during development of prediction models. The relationship between predicted and actual probabilities can be visualised as a calibration curve [12,13]. Trauma care practitioners have found that the TRISS calibration curve is not ideal [14,7,11].

It has been found that the accuracy of TRISS predictions is

* Corresponding author.
E-mail address: vitaly.schetinin@beds.ac.uk (V. Schetinin).

https://doi.org/10.1016/j.ijmedinf.2018.01.009
Received 21 July 2017; Received in revised form 5 January 2018; Accepted 10 January 2018
1386-5056/ © 2018 Published by Elsevier B.V.

acceptable when the types and severities of patient injuries are typical [6]. However, for patients with four or more injuries as well as those with atypical combinations of injuries, the accuracy has to be improved. In practice, it is critically important to reliably estimate the uncertainty in a predicted survival probability. The uncertainty estimates are required in order to minimise risks of misleading decisions. Uncertainty can be represented by confidence intervals. These intervals are reliably estimated when the density of predicted probabilities is fully tractable, which is achievable only in trivial cases. Thus TRISS methodology that is based on theoretical assumptions cannot realistically estimate the uncertainty.

To tackle such problems, we adopt the methodology of Bayesian learning of models, which in theory provides reliable estimation of uncertainty intervals [15–17]. This approach, however, requires intensive computations, as discussed in [18,19].

The learning methodology is based on Bayesian Model Averaging (BMA) which defines a prediction model with parameters that determine the prediction ability. We use the Bayesian method for averaging over decision tree (DT) models which are known for their ability to select input variables that are relevant to the problem, as discussed in [20–22]. The given data are recursively split along input variables into reasonably small subsets. Splits made along variables are easy-to-interpret, and thus DT models built on the given data are able to assist practitioners with new insights [23].

In most practical cases, any given model is incapable of fully explaining the real-world data, which means that a single "true" model does not exist. The BMA methodology, adopted in our study, assumes that different models can be mixed together so that their average under certain conditions will approximate the true model of interest. The averaging strategy is often more efficient than model selection in real-world applications when the predictive ability (or fitness function) is not unimodal [24,25].

The trauma injury severity scoring problems have been considered in the Bayesian context. The study described in [26–28] has been undertaken with the main focus on the receiver operating characteristic (ROC) curve [29], following the standard practice in diagnostic test evaluation. This evaluation, however, is insufficient in the case of imbalanced patient data with a low rate of positive outcomes (e.g. mortality), see e.g. [30,31].

In this paper we propose a new approach to estimating the predictive posterior probability densities of injured patients and examine the accuracy and reliability of predictions on the patient data with low mortality rate. The proposed method is examined on the cases which were registered in the US National Trauma Data Bank (NTDB) [32] with multiple injuries. We also describe a web application [33] which has been developed and made available for evaluation by trauma care practitioners. The application assists practitioners to deliver reliable estimates of uncertainty intervals within which predicted survival probabilities are expected for a patient. Finally we discuss the main results, and draw conclusions.

2. Material and methods

2.1. Data

Our study is conducted on cases from the US NTDB, the major source of data about injured patients admitted to hospitals and emergency units [32]. The data include patient age, gender, type and regions of injuries along with some clinical and background information about patient state. The NTDB also includes the TRISS prediction and the outcome of care, alive or died, for each patient.

Table 1 shows the screening tests (or predictors) that are used by the TRISS method. The variables *Age*, *Blood pressure*, and *Respiration rate* are continuous, and the remaining variables are ordinal. The patient outcome is the *discharge status*: 0 is alive, and 1 is died. The table also shows the minimal and maximal values of each test.

Table 1
Screening tests.

#	Notation	Name	Min	Max
1	A	Age	0	100
2	G	Gender	0 Female	1 Male
3	T	Injury type	0 Penetrating	1 Blunt
4	B	Blood pressure	0	300
5	R	Respiration rate	0	200
6	G_E	GCS eye	1	4
7	G_V	GCS verbal	1	5
8	G_M	GCS motor	1	6
9	H_S	Head severity	0	6
10	F_S	Face severity	0	4
11	N_S	Neck severity	0	6
12	T_S	Thorax severity	0	6
13	A_S	Abdomen severity	0	6
14	S_S	Spine severity	0	6
15	U_S	Upper extremity severity	0	6
16	L_S	Lower extremity severity	0	6
17	E_S	External severity	0	6

For our study, we selected patient records which do not contain missing values in the above predictors. The maximal number of injuries in these records was 48. The number of these cases was 571,775, including 384,876 cases with 1–3 injuries and 186,899 with 4–48 injures. As we discussed in Section 1, the TRISS model has a limited ability to predict outcomes of patients with more than three injuries.

The injuries in NTDB are recorded using to the Abbreviated Injury Scale (AIS) codes, which assign a severity score to each injury. The AIS severity scores are based on mortality risk and range from 1, minor injuries including wounds to skin or subcutaneous tissue or closed fractures [34], to six, considered fatal [35]. However, it has been recently found that about 48.6% of patients with an injury severity of 6 survive [36].

Table 2 shows statistics of the screening tests A to E_S listed in Table 1 in the above groups of patients. The statistics are represented by values of the mean, standard deviation, median, and quartiles.

2.2. Logistic regression model

The use of logistic regression is a conventional way of predicting survival probabilities [5,13]. In trauma injury severity scoring, the logistic regression includes screening tests which are ordinal and continuous predictors. The ordinal variables represent severity scores of injuries obtained by a patient, the Glasgow Coma Scale (GCS), and the injury type. The continuous variables include age, systolic blood pressure, and respiratory rate.

The TRISS model is based on the above screening tests in Table 1, which are represented by the following two aggregated scores: Injury Severity Score (ISS), and Revised Trauma Score (RTS) [5,37]. A side effect of using the aggregated scores is unexplained fluctuations in the calculated survival probabilities, which affect the prediction accuracy, as discussed in [6,38].

The TRISS model determines the probability of survival, P, in the following logistic regression form:

$$P = \frac{1}{1 + e^{-b}}, \tag{1}$$

where $b = b_0 + b_1 \times RTS + b_2 \times ISS + b_3 \times A$.

Here b_0, …, b_3 are the regression parameters, and A is the dichotomised age: $A = 0$, if $age < 55$, and $A = 1$, otherwise. The parameters b were separately determined for blunt and penetrating types of injuries. As discussed in Section 1, the above TRISS model can consider only up to three of the most severe injuries which a patient can obtain.

The TRISS model assumes that the density of predicted values is Gaussian, $N(\mu, \sigma^2)$, where μ and σ^2 are the mean and standard deviation. This assumption, however, is often unrealistic, as discussed in [39–41].

Download English Version:

https://daneshyari.com/en/article/6926321

Download Persian Version:

https://daneshyari.com/article/6926321

Daneshyari.com