Artificial Intelligence in Medicine 69 (2016) 22–32

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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

Evolving classification of intensive care patients from event data

Mark Last[a,*], Olga Tosas[b], Tiziano Gallo Cassarino[c], Zisis Kozlakidis[c], Jonathan Edgeworth[b]

[a] Department of Information Systems Engineering, Ben-Gurion University of the Negev, Marcus Family Campus, Rager St., Beer-Sheva 84105, Israel
[b] Department of Infectious Diseases, Guy's and St. Thomas' NHS foundation Trust, Westminster Bridge Road, London SE1 7EH, United Kingdom
[c] The Farr Institute of Health Informatics Research, University College London, 222 Euston Road, London NW1 2DA, United Kingdom

ARTICLE INFO

Article history:
Received 18 November 2015
Received in revised form 16 April 2016
Accepted 19 April 2016

Keywords:
Evolving classification
Decision trees
Logistic regression
Event data streams
Intensive care

ABSTRACT

Objective: This work aims at predicting the patient discharge outcome on each hospitalization day by introducing a new paradigm—evolving classification of event data streams. Most classification algorithms implicitly assume the values of all predictive features to be available at the time of making the prediction. This assumption does not necessarily hold in the evolving classification setting (such as intensive care patient monitoring), where we may be interested in classifying the monitored entities as early as possible, based on the attributes initially available to the classifier, and then keep refining our classification model at each time step (e.g., on daily basis) with the arrival of additional attributes.
Materials and methods: An oblivious read-once decision-tree algorithm, called information network (IN), is extended to deal with evolving classification. The new algorithm, named incremental information network (IIN), restricts the order of selected features by the temporal order of feature arrival. The IIN algorithm is compared to six other evolving classification approaches on an 8-year dataset of adult patients admitted to two Intensive Care Units (ICUs) in the United Kingdom.
Results: Retrospective study of 3452 episodes of adult patients ($\geq$ 16 years of age) admitted to the ICUs of Guy's and St. Thomas' hospitals in London between 2002 and 2009. Random partition (66:34) into a development (training) set $n = 2287$ and validation set $n = 1165$. Episode-related time steps: Day 0—time of ICU admission, Day x—end of the x-th day at ICU. The most accurate decision-tree models, based on the area under curve (AUC): Day 0: IN (AUC = 0.652), Day 1: IIN (AUC = 0.660), Day 2: J48 decision-tree algorithm (AUC = 0.678), Days 3–7: regenerative IN (AUC = 0.717–0.772). Logistic regression AUC: 0.582 (Day 0)—0.827 (Day 7).
Conclusions: Our experimental results have not identified a single optimal approach for evolving classification of ICU episodes. On Days 0 and 1, the IIN algorithm has produced the simplest and the most accurate models, which incorporate the temporal order of feature arrival. However, starting with Day 2, regenerative approaches have reached better performance in terms of predictive accuracy.

1. Introduction

Over the recent years, the nature, the scale and the speed of data collected within healthcare has changed dramatically, creating new challenges and opportunities. For example, we may be interested to utilize data mining techniques for estimating the probabilities of various discharge outcomes on each day of a given hospital episode. This can be considered as an *evolving classification* problem, where each patient is repeatedly assigned a probability distribution over the optional classes, such as A (discharged alive) vs. D (discharged dead) as more clinical data becomes available. The evolving classification problem considered in this paper is different from the well-known problem of *incremental learning from evolving data streams* [1,2], where the model should be adapted to changing system dynamics in response to new data samples that are continuously arriving over time.

In this paper, we introduce a new paradigm for evolving classification of event data streams. We extend an oblivious read-once decision-tree algorithm, called information network (IN), to deal with evolving classification. The new algorithm, named incremental information network (IIN), restricts the order of selected features by the temporal order of feature arrival. The IIN algorithm

* Corresponding author.
E-mail addresses: mlast@bgu.ac.il (M. Last), Olga.Tosas@gstt.nhs.uk (O. Tosas), t.cassarino@ucl.ac.uk (T.G. Cassarino), z.kozlakidis@ucl.ac.uk (Z. Kozlakidis), jonathan.edgeworth@gstt.nhs.uk (J. Edgeworth).

http://dx.doi.org/10.1016/j.artmed.2016.04.001
0933-3657/

is evaluated on the outcome prediction task in an 8-year dataset of adult patients admitted to two Intensive Care Units (ICUs) in the United Kingdom.

The rest of this paper is organized as follows. Section 2 covers related work on evolving classification algorithms and risk prediction in intensive care units. Section 3 describes the analyzed dataset and the evaluated classification algorithms. The results of the data analysis are presented in Section 4. Section 5 concludes the paper with some insights and directions for future research.

2. Related work

Most classification and regression algorithms, such as logistic regression [3], decision trees [4,5], and support vector machines [6], are not designed for the "evolving classification" task as they consider all predictive features at the same time while ignoring the potentially temporal nature of various features and feature sets. Millan-Giraldo et al. [7] deal with a streaming data scenario, where one or several attributes of incoming instances arrive only after some delay. They suggest the following three straightforward strategies for an early classification of streaming data with delayed attributes: *Do-nothing* (ignore the values of delayed attributes when they become available), *Put-and-reclassify* (reclassify an instance after all attributes become available), and *Wait-and-classify* (classify an instance *only* after all attributes become available). According to the experimental evaluation of Ref. [7], *Wait-and-classify* proves to provide the most accurate results out of the above three strategies, especially when the delayed attributes are the most relevant ones. In case of hospital episodes, *Do-nothing* means classifying a given patient at a single time point (e.g., 24 h after admission) and then ignoring all data arriving afterwards, *Put-and-reclassify* can be interpreted as repeatedly classifying a patient episode on arrival of new attributes, and *Wait-and-classify* implies that patients are classified only on the discharge day when all episode attributes become known. Of course, predicting the episode outcome on the discharge day is nearly useless in the clinical setting.

The delayed attributes scenario of Ref. [7] is related to the novel paradigm of entity stream mining introduced by Krempl et al. in Ref. [8]. This paradigm assumes monitoring a set of entities, such as hospital patients, in the course of their lifetime (e.g., during a given hospital episode). At various time points, each entity is linked to structured or unstructured records ("instances") generated by entity-related events such as medical tests. While many different learning tasks may be defined over such an entity stream, we focus here on the evolving classification task, where an entity classification is required at multiple time steps based on partial sequences of entity-related events.

Stratification of patients into risk groups is important for comparing quality-of-care across different hospitals and units, evaluating the results of clinical trials, and other purposes [9]. Back in 1985, Knaus et al. [10] presented APACHE II, a point score system for estimating the risk of ICU death outcome from 12 physiologic measurements, age, and previous health status. Based on its worst value measured within 24 h after ICU admission, each physiological parameter is assigned a severity weight on a scale of 0–4. The sum of all points is called APACHE II score. Its maximum possible value is 71 though in practice it usually does not exceed 55. The predictive capability of APACHE II was evaluated on 5815 ICU admissions in 13 US hospitals during 1982. The data collection process was actively controlled by the authors of Ref. [10]. A statistically significant increase in the death rate was shown for each 5-point increase in the APACHE score. The predictive capability of APACHE II was evaluated using logistic regression analysis with the outcome as the dependent variable. The area under receiver operating characteristic (ROC) curve (AUC) reported in Ref. [10] for the logistic regression model based on APACHE II score is 0.863. Since the paper [10] does not specify any cross-validation procedure, the reported predictive performance may be based on the developmental (training) data only and thus it may be higher than the true (validation) accuracy. Though the paper [10] emphasizes the importance of the early patient classification at the time of ICU admission (rather than after 24 h), no alternative models for such early prediction are proposed.

Contrary to Ref. [10], Lemeshow et al. [11] propose two mortality probability models (MPM), named MPM0 and MPM24, for use at ICU admission and 24 h after admission, respectively. This is a relatively large study, collected data on 19,124 adult ICU patients at 139 hospitals in 12 European and North-American countries. In case of multiple ICU admissions, only data from the first ICU episode was used. The quality of the collected data was monitored by the physician coordinators at each hospital. The collected records were randomly partitioned into developmental (training) and validation samples with the training/validation ratio of 0.65:0.35. The variables to be used in each multiple logistic regression model were chosen based on statistical tests and clinical plausibility. The data for inducing the MPM0 model included patients who stayed in the ICU for less than 24 h and the validation area under the ROC curve of that model reached 0.824. However, such short ICU stays were excluded from the training set of the MPM24 model, which has shown a slightly higher AUC of 0.836. Out of 13 variables included in the MPM24 model, 5 variables were available at admission and 8 additional variables were assessed at the 24-h mark.

Both the authors of Refs. [12] and [9] indicate the need of accurate risk prediction models for patients who stay in ICU beyond 72 h. Hence, in Ref. [12], the MPM24 model has been adapted to 48-h and 72-h prediction by adjusting the constant coefficient of the logistic regression equation induced for the 24-h model while keeping the coefficients of all independent variables unchanged. This approach has resulted in a decrease in the validation AUC of the MPM48 model to 0.796 (vs. 0.836 of MPM24) and a further decrease to 0.752 for the MPM72 model. In their Discussion section, the authors of Ref. [12] try to find an explanation of this counter-intuitive result, since generally we would expect a later classification model, based on more accumulated information about the patient, to be more accurate.

Trujillano et al. [13] calculate the probability of hospital mortality with three decision-tree classification algorithms: CART [4], CHAID [14], and C4.5 [5]. All evaluated models are aimed at severity estimation for patients within the first 24 h of their admission only. The authors of Ref. [13] indicate that the main benefits of decision trees include the high interpretability of the resulting decision rules along with the relative homogeneity of patient groups associated with each terminal node ("leaf") of the tree. A retrospective dataset of 2864 patients was randomly partitioned in a 70:30 ratio, to form the development and the validation sets. On the validation set, all decision-tree models have reached in Ref. [13] a reasonable AUC level of 0.75–0.76, which was very close to the APACHE II AUC (0.77), but lower than the AUC of logistic regression (0.81).

Portela et al. [15] present the INTCare, a Pervasive Intelligent Decision Support System, which supports intensive care medical activities. The system was used for predicting Organ Failure (Cardiovascular, Coagulation, Respiratory, Hepatic, and Renal) and the Outcome (live or death) of 129 patients in a Portuguese ICU, based on the first five days of their stay. The attributes were collected from bedside monitors, lab results, drugs system, and hospital records. The predictive accuracy of an ensemble of classification models varied across targets between 43% and 83% (64% for predicting the patient outcome).

The overall conclusion is that outcome prediction models for ICU patients are mainly focused on risk assessment after 24 h in the intensive care and, in the case of MPM0, at the time of ICU admis-