Healthcare Transformation with Informatics and Artificial Intelligence
J. Mantas et al. (Eds.)

doi:10.3233/SHTI230462

Provision of Decision Support Through Continuous Prediction of Recurring Clinical Actions

Michal WEISMAN RAYMOND[a,1] and Yuval SHAHAR[a]
[a] *Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel*

Abstract. We propose a framework for provision of decision support through the continuous prediction of recurring targets, in particular clinical actions, which can potentially occur more than once in the patient's longitudinal clinical record. We first perform an abstraction of the patient's raw time-stamped data into intervals. Then, we partition the patient's timeline into time windows, and perform frequent temporal patterns mining in the features' window. Finally, we use the discovered patterns as features for a prediction model. We demonstrate the framework on the task of treatment prediction in the Intensive Care Unit, in the domains of Hypoglycemia, Hypokalemia and Hypotension.

Keywords. Clinical Decision Support, Machine Learning, Temporal Data Mining.

1. Introduction

Deciding upon a patient's treatment can be a complicated task, as it entails dealing with continuously arriving streaming data of different sources, which must be constantly analyzed and combined with medical knowledge. Current solutions suggest the usage of known medical guidelines [1–3] or learning the optimal treatment anew, based on historical data, using reinforcement learning [4, 5]. In this study, we propose a different approach, which tries to continuously predict the next action that an experienced physician would be likely to perform, in a context similar to that of the current patient's, by learning from a large dataset of similar patients.

First, we convert the raw time-stamped clinical data into intervals which represent meaningful abstract concepts. Then, we partition the patient's timeline into feature and target time windows and mine frequent patterns within the feature windows. Next, we use these patterns as features to predict the dosage given in the target window, using a two-step machine learning approach which first determines the need for therapy, and if relevant, predicts the actual dose.

To implement this framework, we also integrate two temporal-analysis methods: knowledge-based temporal abstraction (KBTA)[6], which converts time intervals of raw time-stamped data into intervals which represent meaningful concepts; and frequent

[1] Corresponding Author: Michal Weisman Raymond, Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Beer Sheva, Israel; Email: michalweisman@gmail.com.

pattern detection using KarmaLego [7]. It discovers Time-Interval Relation Patterns, which have been previously used as features for machine learning models [8, 9].

To demonstrate the framework's ability to predict treatment, we applied it on the MIMIC-IV [10] dataset, to three different medical domains within the Intensive Care Unit (ICU) area: Hypoglycemia, Hypokalemia and Hypotension.

2. Methods

2.1. Computational Methods

To convert raw time-stamped data into time intervals, we first apply the KBTA method. Next, we partition the time intervals into the following time windows: (1) Features Time Window (FW) - The time window from which data is collected to be used in a prediction model. (2) Target Time Window (TW) The time window in which the outcome class or value should be predicted. (3) Prediction Gap (PG) - A time window starting at the end of FW and ending at the start of TW. These time windows form an instance, which is an entity that is constructed from the sequence of Time Windows: FW, PG and TW. Instance Gap (IG) is the gap between the start time of two consecutive instances. The sizes of FW, TW, PG and IG (denoted as $|FW|$, $|TW|$, $|PG|$ and $|IG|$, respectively), are supplied by the end user. **Figure 1** presents a visual illustration of the instances produced from a given patient's timeline.

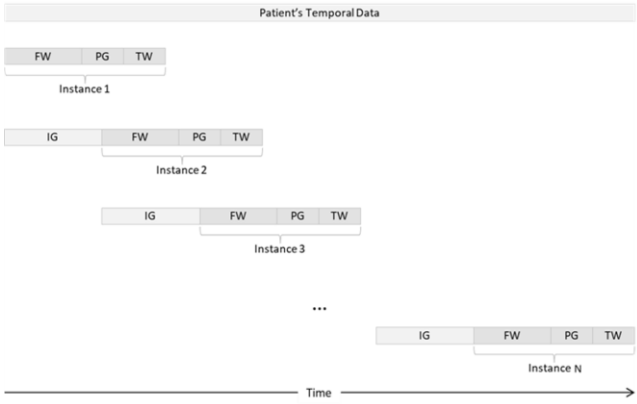

Figure 1. Instances extracted from a patient's time series.

After partitioning the patient's timelines into instances, a set of all instances (regardless of patient's association) is created. The instances are partitioned to train and test sets, and are grouped into two classes: a positive class for instances which received treatment in TW, and a negative class for those who did not. We then apply the KarmaLego algorithm for frequent pattern discovery, on TW of all instances in the train set, for each class separately (this division is important, as a pattern might meet the frequency condition only in a specific class, but not in the whole dataset). Next, we search for the set of mined patterns in the whole dataset, and compute the pattern's statistical characteristics in each instance, according to the statistics type provided by the user, (for example: pattern mean duration, horizontal support, etc.). We then use these characteristics as features in a machine learning model. First, we predict whether

treatment should be administered using Random Forest Classifier, and then we predict the dosage using Gradient Boosting Regressor.

2.2. Evaluation Methods

The dataset which was used for the evaluation of the proposed framework is the Medical Information Mart for Intensive Care (MIMIC) IV dataset [10] . It contains the clinical data of over 40,000 patients at the Beth Israel Deaconess Medical Center. Due to computational limitations, for each experiment, a subset of 50-100 patients was considered. To obtain enough positive samples, we used a subset of patients which received a relevant diagnosis **at some point** in their timeline. We performed evaluation on the following ICU clinical domains: Hypoglycemia - dextrose given IV or PO (mg); Hypokalemia - potassium chloride (mEq); Hypotension - fluids in the form of dextrose or sodium chloride solutions (mL), dopamine (mg) and norepinephrine (mg). We used the *Area Under the ROC Curve* (AUC) to estimate the framework's ability to detect non-treatment, and the *Mean Absolute Error* (MAE), *Root Mean Square Error* (RMSE), and their normalized versions (divided by the standard deviation) NMAE and NRMSE.

3. Results

We present the resulting performance metrics of the framework, with two sets of configurations, applied to the data of the different ICU clinical domains. Both sets included $|FW| = 24\ hours$ and $|TW| = 4\ hours$. Table 1 presents the results for no prediction gap, and mean pattern duration as the feature type. Table 2 presents the results for a prediction gap of two hours and horizontal support as the feature type. In both cases, the framework presented high AUC scores, implying that it was able to differentiate well between treatment and non-treatment cases. In addition, both configurations resulted in $NMAE<1$ and $NRMSE<1$, meaning that the errors were less than one standard deviation. In general, the results were surprisingly good even when a gap was used before the prediction window, suggesting the possibility of providing an advance alert to a therapy.

Table 1. Experiments results for the Hypoglycemia, Hypokalemia and Hypotension domains. Experimental configuration: |FW| = 24 hours, |TW| = 4 hours and |PG| = 0, Feature Type = mean duration.

Domain	Treatment	AUC	MAE	RMSE	NMAE	NRMSE
Hypoglycemia	Dextrose	0.947	2.389	4.671	0.355	0.693
Hypokalemia	Potassium chloride	0.8	0.27	1.976	0.041	0.298
	Fluids	0.928	32.15	60.359	0.392	0.736
Hypotension	Dopamine	0.944	10.08	45.951	0.076	0.346
	Norepinephrine	0.942	0.237	0.653	0.184	0.506

Table 2. Experiments results for the Hypoglycemia, Hypokalemia and Hypotension domains. Experimental configuration: |FW| = 24 hours, |TW| = 4 hours and |PG| = 2 hours, Feature Type = horizontal support.

Domain	Treatment	AUC	MAE	RMSE	NMAE	NRMSE
Hypoglycemia	Dextrose	0.937	2.42	4.66	0.363	0.699
Hypokalemia	Potassium chloride	0.772	0.916	3.548	0.193	0.707
	Fluids	0.933	37.14	71.347	0.38	0.73
Hypotension	Dopamine	0.95	13.61	51.126	0.102	0.383
	Norepinephrine	0.961	0.416	1.03	0.22	0.545

4. Discussion and conclusions

Medical treatment decisions are complex. Current solutions include using medical guidelines, which require a pre-determined diagnosis; and reinforcement learning approaches, which include learning optimal actions and require a large amount of data and a long time span to include long-term effects of actions. In this study, we proposed a different approach, which aims to predict the most probable action based on the collective experience of multiple physicians in similar situations.

First, we introduced a window partitioning algorithm, and provided a scripting language that supports the framework's flexibility. Next, we performed frequent pattern mining. Then, we used the patterns in a two-step prediction algorithm for treatment prediction, which first makes the binary prediction, of whether any treatment should be given, and if relevant, predicts the medication's dosage. We also introduced the usage of a prediction gap, which resulted in surprisingly good results. This can potentially allow the medical teams to receive an advance notification of necessary future actions.

The framework demonstrated its effectiveness in three medical domains: Hypoglycemia (dextrose treatment), Hypokalemia (potassium chloride treatment) and Hypotension (dopamine, norepinephrine, and fluids treatments). This study had several limitations, such as using the data of a single hospital, and including only patients which had a relevant diagnosis at some point during their stay. Thus, we aim to enhance our dataset in the future.

Acknowledgements

This work was partially supported by ONR award No. N629091912124.

References

[1] Shalom E, Goldstein A, Ariel E, Sheinberger M, Jones V, Van Schooten B, Shahar Y. Distributed application of guideline-based decision support through mobile devices: Implementation and evaluation. Artif Intell Med 2022; 129: 102324.

[2] Shalom E, Shahar Y, Parmet Y, Lunenfeld E. A multiple-scenario assessment of the effect of a continuous-care, guideline-based decision support system on clinicians' compliance to clinical guidelines. Int J Med Inform 2015; 84: 248–262.

[3] Shalom E, Shahar Y, Lunenfeld E. An architecture for a continuous, user-driven, and data-driven application of clinical guidelines and its evaluation. J Biomed Inform 2016; 59: 130–148.

[4] Zhao Y, Kosorok MR, Zeng D. Reinforcement learning design for cancer clinical trials. Stat Med 2009; 28: 3294–3315.

[5] Hassani A. Reinforcement learning based control of tumor growth with chemotherapy. In: 2010 International Conference on System Science and Engineering. 2010, pp. 185–189.

[6] Shahar Y. A framework for knowledge-based temporal abstraction. Artif Intell 1997; 90: 79–133.

[7] Moskovitch R, Shahar Y. Fast time intervals mining using the transitivity of temporal relations. Knowl Inf Syst 2015; 42: 21–48.

[8] Sheetrit E, Nissim N, Klimov D, Shahar Y. Temporal probabilistic profiles for sepsis prediction in the ICU. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019, pp. 2961–2969.

[9] Moskovitch R, Shahar Y. Classification of multivariate time series via temporal abstraction and time intervals mining. Knowl Inf Syst 2015; 45: 35–74.

[10] Johnson A, Bulgarelli L, Pollard T, et al. MIMIC-IV in PhysioNet.