User Centred Networked Health Care
A. Moen et al. (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-806-9-140

Prognostic Data-Driven Clinical Decision Support - Formulation and Implications

Ruty RINOTT[a,1] Boaz CARMELI[a] Carmel KENT[a,] Daphna LANDAU[a] Yonatan MAMAN[a] Yoav RUBIN[a], Noam SLONIM[a]
[a] *IBM Haifa Research Labs, 165 Aba Hushi st., Haifa 31905, Israel*

Abstract. Existing Clinical Decision Support Systems (CDSSs) typically rely on rule-based algorithms and focus on tasks like guidelines adherence and drug prescribing and monitoring. However, the increasing dominance of Electronic Health Record technologies and personalized medicine suggest great potential for *prognostic data-driven CDSS*. A major goal for such systems would be to accurately predict the outcome of patients' candidate treatments by statistical analysis of the clinical data stored at a Health Care Organization. We formally define the concepts involved in the development of such a system, highlight an inherent difficulty arising from bias in treatment allocation, and propose a general strategy to address this difficulty. Experiments over hypertension clinical data demonstrate the validity of our approach.

Keywords. Clinical Decision Support, Data Driven, Machine Learning, Prognostic

1. Introduction

The need for Clinical Decision Support Systems (CDSSs) increases rapidly [1]. Most existing systems are rule-based systems focused at guidelines adherence, drug prescribing and monitoring, etc. [2]. The increasing pace by which Health Care Organizations (HCOs) adopt Electronic Health Record (EHR) technologies and the increasing recognition of personalized medicine importance suggest great potential for another type of CDSS, aiming to predict the outcomes of treatments considered for an individual patient via statistical and machine learning algorithms. We suggest a formal general description for such a *prognostic data-driven CDSS* (pdd-CDSS) and highlight an inherent difficulty associated with the development of such a system, related to the inherent bias in HCO's clinical data. We then propose a general strategy to address this difficulty and demonstrate our approach over clinical data of hypertension patients.

2. Methods

2.1. Defining Relevant Concepts

We consider a patient who is at stage *k* of disease *d*. The pdd-CDSS should assist the physician by predicting the expected outcome of relevant candidate treatments for this

[1] Corresponding Author.

individual patient, through mining the HCO's clinical data. Let T be a random variable with values $\{t_1,\ldots, t_{Nt}\}$, representing distinct candidate treatments. Let O be a random variable with values $\{o_1,\ldots, o_{No}\}$, representing distinct outcomes. We assume that the HCO maintains data about N_f clinical features, denoted by the random variables, $\{f_1,\ldots, f_{Nf}\}$. The sample population for the pdd-CDSS consists of N_p patients that have already been at stage k of disease d *and* their received treatment and resulting outcome are recorded in the HCO's database. These N_p patients can thus be divided into mutually exclusive and exhaustive *treatment groups*, according to the their treatment value, T, denoted $\{g_{t1},\ldots, g_{tNt}\}$.[2] The data mined by the pdd-CDSS can thus be represented by a matrix M, where $M(i,j)$ indicates the value of the i-th patient according to the j-th feature. The treatment and the outcome variables can be represented via two additional column vectors. Finally, we denote a *new* patient by the index i^*, and the data associated with her is represented via an additional row in M, while $T(i^*)$ and $O(i^*)$, are obviously unknown. All these notations are depicted in Fig. 1a.

2.2. Treatment Groups are Inherently Biased

Our first observation is that from a statistical perspective, different treatment groups often represent different populations, reminiscent to an *observational study* [3]. As an extreme example, let us assume that gender, denoted for example by f_j, affects treatment success. We further assume that in the HCO's data for all patients in g_{t1}, $f_j=M$, while for all patients in g_{t2}, $f_j=F$, e.g., due to the HCO's guidelines. Next, we consider a new female patient. Since there are no examples in the data for female patients who received treatment t_1, and assuming gender affects the treatment success, machine learning and statistical analysis algorithms will not be able to properly predict the outcome of applying t_1 to this new patient based on the HCO's records. In practice, we do not expect the distinction between the treatment groups to be that obvious. However any bias in baseline covariates between treatment groups will affect prediction ability and must be considered in the design of a pdd-CDSS. Next, we propose one strategy to address this issue.

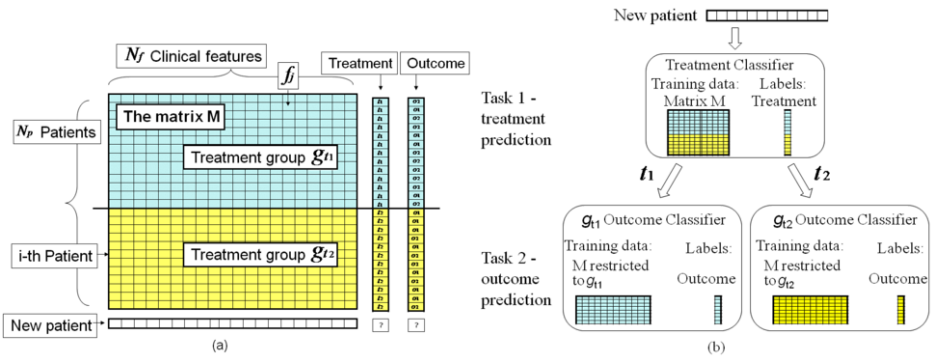

Figure 1. (a) Notations. (b) A flow chart for the proposed pdd-CDSS.

[2] For simplicity, if a patient received more than one treatment during the same stage of the disease, her assignment to a treatment group is done based on the most recent treatment she received.

2.3. A Valid Flow for pdd-CDSS

In the example above, while we could not predict the outcome of applying t_1 to the new patient, we could have predicted the outcome of applying t_2 to that patient. Thus, if the "customary"[3] treatment can be determined for a new patient, the outcome of that treatment may be reliably predicted. This suggests a strategy of limiting outcome prediction to "customary" treatments. However, identifying the "customary" treatment for a new patient might be far from trivial, involving complex considerations. Here, we propose to first exploit the bias in treatment allocation to predict the HCO's "customary" treatment. If a treatment group is clearly identified, it implies that the patients in that treatment-group are relatively similar to the new patient, in particular in the context of the covariates that distinguish the different treatment groups. Hence, outcome prediction can be reliably performed in that treatment group. Thus we propose to decompose outcome prediction for a new patient into two separate tasks (cf. Fig 1b):

- *Treatment prediction*: predict $T(i^*)$, i.e., the HCO's "customary" treatment for the new patient, using all N_p patients as training data.
- *Outcome prediction for the predicted treatment*: predict the outcome only for the predicted treatment; namely, predict $O(i^*)$ given that the treatment is $T(i^*)$, using only patients who underwent $T(i^*)$ as training data.

3. Results

We demonstrate our methodology over clinical data collected for hypertension patients as part of the Hypergenes project[4]. We identified three major possible treatments in the data: non-drug therapy (t_1); angiotensin II receptor blockers (t_2); and beta blockers (t_3). We focused on patients that suffer from Stage-1 hypertension and for which: (a) the treatment group is known and the date in which this treatment was assigned is known[5]; (b) Systolic and diastolic blood pressure (BP) were measured when treatment was selected and at an additional later time point. This led to a dataset of N_p=1771 patients with respect to 181 clinical features. Decrease in BP to below hypertension levels (diastolic < 90, systolic <140) was denoted as outcome o_1, while failing to do so was denoted by o_2. In treatment groups g_{t1}, g_{t2}, g_{t3}, we had 750, 475, and 63 patients, respectively, for which 39%, 51%, and 37%, had a resulting outcome o_1, respectively.

3.1. Prediction Algorithms

For both classification tasks (Section 2.3) we used a *k*-Nearest Neighbor (*k*NN) classifier [4]. Given a new patient, the algorithm finds her *k* NNs in the training data and predicts her label via a weighted majority of their labels. In the treatment-prediction task the training data consisted of all 1771 patients and the label was the given treatment. For the outcome-prediction task, we first predict the patients "customary" treatment, and then use the patients within this treatment-group as the training data, and their outcome as the label. An inherent challenge in k-NN

[3] Importantly, this "customary"' treatment is not necessarily optimal for the new patient. Rather, it solely reflects decisions made in the past in this HCO for patients with somewhat similar characteristics.

[4] For more details, See http://www.hypergenes.eu/.

[5] For simplicity we assume that this is also the date when treatment course started.

classification is to define the distance measure used to determine the NNs. Ideally, this measure should be adapted to the classification task, e.g., by assigning different weights to features based on their prediction power. In our context, while some features might contribute to treatment prediction, others might contribute to outcome prediction. Further, different features may affect the success of different treatments. For example, initial weight may significantly affect the success of a life-style change treatment while having a smaller affect on the success of drug therapy. This suggest that 4 different distance measures should be learned in our data; one for treatment prediction, and 3 for outcome prediction - one within each treatment group.

3.2. Information Based Distance

A natural way to quantify the dependency of a feature with a label is via the Mutual Information (MI) associated with their joint probability [5]. This measure is especially attractive in our context as it similarly applies for continuous and categorical random variables; allows capturing any type of dependency, including non-linear relations; and there is much literature on correcting MI estimates due to sample size effects, a dominant problem in real world clinical data. Here, we used the technique in [6] to estimate the MI between each feature and the relevant label. As expected, the MI value associated with a feature changes along with the task. For example, for the feature "Patient age" we observed high MI for the treatment-prediction task (0.17 bits), while nearly zero MI in all outcome-prediction tasks. In Table 1 we present the MI estimates in each prediction task for the three features with the highest observed MI. In all prediction tasks we used the obtained MI values to determine the distance measure [7], discarding features with MI<0.01 bits, and weighting the remaining features by their relative MI value. This led to 4 different similarity measures, where in each prediction task the most informative features contribute the most to the similarity estimates.

Table 1. Features MI (in bits) under different tasks for the three features with the highest MI per task.

Treatment Prediction	g_{t1} **Outcome Prediction**	g_{t2} **Outcome Prediction**	g_{t3} **Outcome Prediction**
Diastolic BP at decision (0.34)	Average systolic BP prior to decision (0.34)	Systolic BP at decision (0.08)	Systolic BP at decision (0.13)
Systolic BP at decision (0.23)	Average LDL cholesterol prior to decision (0.24)	Diastolic BP at decision (0.04)	Height (0.03)
Age (0.17)	Average systolic BP prior to decision (0.19)	LDL cholesterol at decision (0.02)	Alcohol consumption at decision (0.01)

3.3. Prediction Results

Using the distance measure learned for the treatment-prediction task and k=10 we predicted the treatment for all 1771 patients. The prediction accuracy was 85% suggesting a significant statistical bias between the treatment groups, exploited by the k-NN classifier. Next, we focused on patients for whom the predicted treatment was correct and relatively certain, i.e., there was a relatively clear majority for the correct label amongst the patient's k-NN. This resulted with 632, 361 and 10 patients in g_{t1}, g_{t2}, and g_{t3}, respectively. For each of these patients we predicted the outcome using the similarity measure learned within the relevant treatment-group and k=10, ending up with an average prediction accuracy of 66%.

4. Discussion

The increasing scale and complexity of recorded clinical features that affect treatment choice highlights the need for CDSSs [1]. Here we formally defined pdd-CDSS that utilize HCO's clinical data to predict patient outcome to candidate treatments. In recent years, several decision support tools have been developed that rely on mining clinical trials' results [8]. However, the increasing pace by which HCOs adopt EHR technologies suggests great potential in mining HCOs' clinical data, along with non obvious challenges. Here we discussed how treatment allocation bias hampers the ability to predict outcome for all candidate treatments. We suggested a framework to identify such biases and pinpoint for which treatments prediction can be made reliably. Treatment group bias has been discussed in papers that evaluate non-randomized clinical trials and observational studies [3] and various methods have been proposed to try and correct for this bias [9]. In contrast, here we do not aim to correct for treatment bias, but to narrow outcome prediction to cases where this bias is less harmful. Considering bias-correction tools within the pdd-CDSS framework suggested here is left for future research. The pdd-CDSS framework raises additional challenges. First, such systems cannot be detached from external knowledge sources such as published guidelines. Integrating guideline based CDSSs with pdd-CDSSs can add important information such as contraindications and sharpen the recommendations created by such systems. In parallel, much work remains in developing prediction algorithms that properly handle the heterogeneity in clinical data, consider dependencies between features, and more. Finally, remains the challenge of prediction evaluation. Obviously it is impossible to measure outcomes of treatments not delivered. Thus, alternative methods to evaluate the accuracy of prediction algorithms must be formulated.

References

[1] Kawamoto K, Lobach DF, Willard HF, Ginsburg GS. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine, *BMC Medical Informatics and Decision Making*, **9**:17 (2009), 1-14.

[2] Kuperman GJ, Bobb A, Payne TH, et al. Medication-related Clinical Decision Support in Computerized Provider Order Entry Systems: A Review, *Journal of the American Medical Informatics Association*, **14**:1, (2007), 29-40.

[3] Dreyer NA, Tunis SR, Berger M, Ollendorf D, Mattox P, Gliklich R. Why Observational Studies Should be Among the Tools Used in Comparative Effectiveness Research, *Health Affairs*, **29**, (2010), 1818-1825.

[4] Cover TM, Hart PE. Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, **13**:1, (1967), 21-27.

[5] Cover TM, Thomas JA. *Elements of Information Theory*, Wiley, New York, (2006).

[6] Slonim N, Atwal GS, Tkacik G, Bialek W. Information based clustering, *PNAS*, **102**:51 (2005), 8297-8302.

[7] Garcia-Laencina PJ, Sancho-Gomez J, Figueiras-Vidal A, et al. K nearest neighbours with mutual information for simultaneous classification and missing data imputation. *Neurocomputing*,**72** (2009), 83–93.

[8] Thome SD, Loprinzi CL, Heldebrant MP. Determination of Potential Adjuvant Systemic Therapy Benefits for Patients With Resected Cutaneous Melanomas, *Mayo Clinic Proceedings*, **77**:9 (2002) 913-917.

[9] Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Ed. Psychol.* 66 (1974), 688–701.