



Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

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



Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modeling with neural networks

Seokho Kang

Department of Systems Management Engineering, Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon 16419, Republic of Korea

ARTICLE INFO

Article history:

Received 25 September 2017
Received in revised form 15 January 2018
Accepted 15 February 2018

Keywords:

Type 2 diabetes mellitus
Drug failure prediction
Sequential modeling
Personalized prediction
Neural network

ABSTRACT

Patients with type 2 diabetes mellitus are generally under continuous long-term medical treatment based on anti-diabetic drugs to achieve the desired glucose level. Thus, each patient is associated with a sequence of multiple records for prescriptions and their efficacies. Sequential dependencies are embedded in these records as personal factors so that previous records affect the efficacy of the current prescription for each patient. In this study, we present a patient-level sequential modeling approach utilizing the sequential dependencies to render a personalized prediction of the prescription efficacy. The prediction models are implemented using recurrent neural networks that use the sequence of all the previous records as inputs to predict the prescription efficacy at the time the current prescription is provided for each patient. Through this approach, each patient's historical records are effectively incorporated into the prediction. The experimental results of both the regression and classification analyses on real-world data demonstrate improved prediction accuracy, particularly for those patients having multiple previous records.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Presently, diabetes mellitus is one of the most prevalent diseases with a continually growing number of diabetic patients worldwide. Type 2 diabetes mellitus (T2DM), in which the body of an individual is not able to use insulin properly owing to insulin resistance along with a relative insulin deficiency, is the most common form, estimated to account for 90–95% of the patients with diabetes [1]. Though the specific causes of T2DM have not been identified yet, it is known to be highly associated with genetic, environmental, and lifestyle factors [2]. Owing to the complexity and chronicity, T2DM patients typically require continuous long-term medical treatment, mainly based on oral hypoglycemic agents to achieve the desired glucose level [3,1]. During the treatment, the patients regularly visit a hospital for diagnosing their current medical conditions and accordingly receiving prescriptions of anti-diabetic drugs.

Abundant data for diabetic patients including electronic medical records (EMR) have been collected in the recent decades [4], to which machine learning can be applied to effectively address the complex relationships between the multiple factors in data [5,6]. Most studies have focused on applying machine learning methodologies to the diagnosis and risk prediction of diabetes [7–18].

Regarding T2DM, some studies have presented a data-driven modeling approach that builds prediction models to estimate the subsequent prescription efficacy at the time of a prescription is provided [19–21]. Some related studies in the literature that applied machine learning to diabetes data are summarized in Table 1. An accurate prediction of the efficacy would contribute to the data-driven clinical decision support for the patients. However, achieving the desired accuracy is difficult and still challenging because the efficacy of each prescription for a patient is significantly affected by various personal factors such as personal characteristics, medical condition, prescription details, and presence of diabetic complications.

As personal factors, we focus on the sequential dependencies in the records for the prescriptions and their consequences for each T2DM patient, in which the current record is affected by the previous records [7,22,23]. In this regard, here we propose a patient-level sequential modeling approach to enable a personalized prediction of the efficacy of a prescription for a patient based on his/her previous records. Though the conventional approach is to perform a prescription-level modeling that does not directly exploit the previous records to make a prediction, this approach predicts the efficacy of the current prescription of a patient as a function of the sequence of all the previous records for the patient. This allows improving the prediction accuracy for patients with multiple previous records. We investigate the effectiveness of incorporating the sequential depen-

E-mail address: s.kang@skku.edu

<https://doi.org/10.1016/j.artmed.2018.02.004>

0933-3657/© 2018 Elsevier B.V. All rights reserved.

Table 1
Summary of literature applying machine learning to diabetes data.

Objective	Machine learning task	Prediction model	Literature
Disease diagnosis (whether diabetes or not)	Binary classification	Decision tree	[18]
		Ensemble	[10,11,17]
		Feed-forward neural network Support vector machine	[7,8,12–15] [9,16]
Treatment efficacy prediction (whether patient is in control)	Binary classification	Decision tree, <i>k</i> -nearest neighbors, Naive Bayes	[19]
		Support vector machine	[20]
	Multi-class classification Regression	Recurrent neural network Ensemble	Present study [21]
		Recurrent neural network	Present study

dences through experimental regression and classification analyses using the EMR data collected from Seoul National University Hospital (SNUH).

The remainder of this paper is organized as follows. In Section 2, we overview the data used in this study. In Section 3, we describe the sequential modeling approach for predicting the efficacy of a prescription for T2DM. Section 4 reports the experimental results on the EMR data to demonstrate the effectiveness of the approach. Finally, the conclusion and future work are provided in Section 5.

2. Data

2.1. Data collection

In this study, we used the EMR data of 3169 T2DM patients collected from SNUH during 2003 to 2013. The data collection was conducted after the protocol had been approved by the institutional review board of SNUH. In the data collection environment, a patient who undertook a test for the medical condition received a prescription within a week according to the test results, and undertook the test again to check the glucose level to confirm the efficacy after 2–6 months of the prescription, as illustrated in Fig. 1. A T2DM patient generally went through this process several times over a long duration. Prescription-level instances were constituted to fulfill this requirement, and accordingly, we obtained 45,384 instances from the data.

The available information at the time of the prescription and the prescription efficacy were used to derive the input and output variables, respectively. As the input variables, the personal characteristics and medical conditions examined by the test and the

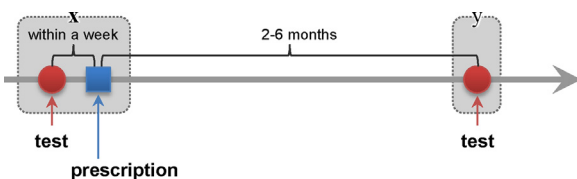


Fig. 1. Constitution of the input and output variables for each instance.

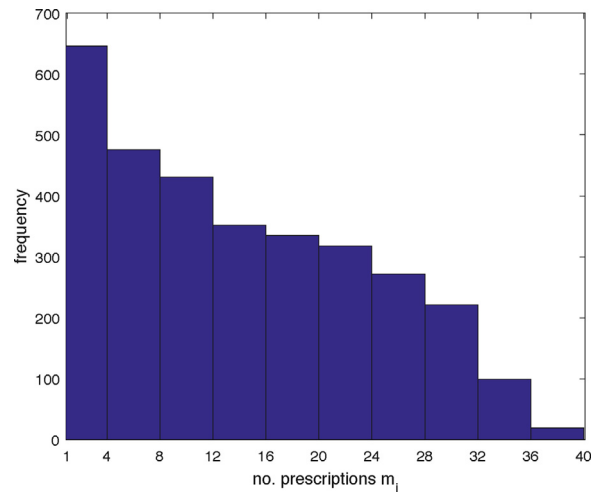


Fig. 2. Number of prescriptions for the patients.

prescription details of anti-diabetic and other drugs were used. For the prescription of anti-diabetic drugs, the input variables were continuous-valued, determined as the amount of daily dose of the hypoglycemic agents, except for the case of insulin for which the variable type was binary. For the prescription of other drugs, the input variables were binary-valued, defined based on whether each drug category was prescribed or not. The output variable was defined based on the prescription efficacy 2–6 months after the prescription. Glycated hemoglobin (HbA1c), which is a widely used measure to evaluate the glycemic level for the diagnosis of T2DM, was employed as an indicator of the efficacy [24–26]. HbA1c itself was used as the continuous-valued output variable for the regression analysis, while the output variable for the classification analysis was binary-valued, defined based on whether HbA1c increases by 0.5 or is greater than 7 according to the glycemic recommendation of the American Diabetes Association [1]. All the variables used in this study are listed in Table 2.

The following notations are used herein. Let $i \in \{1, \dots, 3169\}$ denotes the patient id, and a prescription for the i -th patient is represented by a prescription id $t \in \{1, \dots, m_i\}$ assigned in order of occurrence, where m_i is the total number of prescriptions for the patient. It should be noted that for each patient the value of m_i is different. The instance of the t -th prescription for the i th patient is defined as $(\mathbf{x}_i^{(t)}, y_i^{(t)})$, where $\mathbf{x}_i^{(t)}$ and $y_i^{(t)}$ are the t -th input vector and output value for the i th patient, respectively. Accordingly, the i th patient corresponds to the set of instances $\{(\mathbf{x}_i^{(t)}, y_i^{(t)})\}_{t=1}^{m_i}$.

2.2. Data exploration

Typically, a T2DM patient regularly visits a hospital multiple times to receive prescriptions for long-term treatment. However, from the data perspective, there are numerous patients corresponding to only a few instances, e.g. those who have started to visit the hospital near the end of the data collection period and those have been transferred from or to another hospital. The number of prescriptions for a patient also depends on the relative seriousness of his/her medical condition. Fig. 2 presents the histogram of the distribution of number of prescriptions m_i over the patients in the data. The patients received an average of 14.32 prescriptions, with minimum, maximum, and mode values of 1, 40, and 1, respectively. The patients with less than 5 prescriptions accounted for 20.38% of the total. The number of patients with a large number of prescriptions was relatively small.

In this study, the entire instances were divided into six intervals based on t : (1) $t = 1-2$, (2) $3-5$, (3) $6-10$, (4) $11-15$, (5) $16-20$,

Download English Version:

<https://daneshyari.com/en/article/6853322>

Download Persian Version:

<https://daneshyari.com/article/6853322>

[Daneshyari.com](https://daneshyari.com)