KICS The Korean Institute of Communications and Information Sciences

Available online at www.sciencedirect.com

ScienceDirect

ICT Express 2 (2016) 150–154

www.elsevier.com/locate/icte

Smartphone-based personalized blood glucose prediction☆

Juan Li*, Chandima Fernando

Computer Science Department, North Dakota State University, Fargo, USA

Received 1 August 2016; received in revised form 6 October 2016; accepted 10 October 2016
Available online 25 October 2016

Abstract

Effective blood glucose control is essential for patients with diabetes. However, individual patients may not be able to monitor their blood glucose level regularly because of all manner of real-life interference. In this paper, we propose a personalized diabetes prediction mechanism that leverages smartphone-collected patient data and population data to drive personalized prediction. Unlike existing predictive models, this model utilizes pooled population data and captures patient similarities, and eventually produces a personalized blood glucose prediction for an individual. We have implemented the proposed model as a mobile application and have performed extensive experiments to evaluate its performance. The experimental results demonstrate that the proposed prediction mechanism can improve the prediction accuracy and remedy the problem of sparse data in the existing approaches.

Keywords: Diabetes; Smartphone; Blood glucose; Prediction; Personalized care

1. Introduction

Diabetes is becoming increasingly common around the world. In order to control blood sugar levels and prevent hypoglycemia, frequent blood glucose (BG) monitoring is needed by diabetes patients and their healthcare professionals. Diabetes control is aided by BG self-monitoring by facilitating the creation of an individualized BG profile. This profile can help healthcare professionals to draw up an individualized treatment plan for a particular patient. Moreover, it can also give diabetes patients and their families the ability to make appropriate day-to-day treatment choices about diet and physical activity, as well as about insulin or other agents [1].

However, frequent and regular monitoring of BG level is difficult and often impractical in a patient's daily life. The frequency with which diabetic patients should monitor their BG level varies from patient to patient. Most experts agree that insulin-treated patients should monitor BG at least four times a day, most commonly while fasting, before meals, and before bed. In some circumstances, however, patients are unable to maintain this frequent monitoring; for example, a patient may be in a meeting, may not have the testing devices with them, or may forget the test. For this reason, accurate prediction of BG level is very important for diabetes self-management.

There has been much research into automatic BG prediction using machine learning algorithms. For example, in the Artificial Pancreas Project [2], BG level is predicted so that the insulin flow can be continuously adjusted to meet patient needs. Despite the existing research, a challenge remains to predict BG level accurately and make personalized recommendations. Existing prediction models can be classified as either population-based prediction (e.g., [3–6]) or patient-based prediction (e.g., [7–10]). A population-based prediction of one patient's situation is based on independent population data. As shown in our experiments, the accuracy of this model in estimating individual risk is relatively low. Patient-based prediction normally requires a large amount of historical patient data to make a usable prediction. Sometimes, this requirement cannot be satisfied (e.g., for a new patient).

To address the aforementioned problems of existing approaches, our work shifts from using either population- or patient-based analysis and prediction to a synthesized population/patient-based analysis and prediction. We propose a personalized data-centric predictive model to predict patients'

* Corresponding author.
E-mail addresses: j.li@ndsu.edu (J. Li), chandima.fernando@ndsu.edu (C. Fernando).

Peer review under responsibility of The Korean Institute of Communications Information Sciences.

☆ This paper is part of a special issue titled Special Issue on Emerging Technologies for Medical Diagnostics guest edited by Ki H. Chon, Sangho Ha, Jinseok Lee, Yunyoung Nam, Jo Woon Chong and Marco DiRienzo.

http://dx.doi.org/10.1016/j.icte.2016.10.001

BG levels automatically based on their daily activity patterns collected from their smartphones, their historical information kept in their smartphones, as well as the historical population data. We leverage population data to drive personalized prediction. In particular, we propose a three-stage evolution model that includes a time-series regression model based on personal history, a pooled panel data (PPD) regression model, and a pre-clustered personalized regression model. A prediction system can choose an appropriate model from the three aforementioned models based on its data used for prediction.

2. Related work

There have been numerous studies of risk prediction related to diabetes management. In 1996, Shanker proposed the use of artificial neural networks to predict the onset of diabetes mellitus among the Pima Indian female population near Phoenix, Arizona [11]. Researchers have developed risk scores and predictive models for diabetes screening based on population studies [12]. More recently, Choi et al. developed two models to screen for prediabetes using an artificial neural network and support vector machine (SVM), and performed a systematic evaluation of the models using internal and external validation [13]. Zecchin et al. quantified the potential benefits of glucose prediction in terms of a reduction in the frequency/duration of hypoglycemia [14]. The Diabetes Support System (4DSS) designed by Ohio University tries to provide intelligent decision support systems for patients with type 1 diabetes (T1D) who are on insulin pump therapy [15].

Different machine learning algorithms have been proposed to predict BG level. Sandham et al. proposed the use of an artificial neural network and neuro-fuzzy systems to predict BG level for expert management of diabetes mellitus [8]. El-Jabali [16] used artificial neural networks to create a dynamic simulation model of T1D. Plis et al. [9] proposed a generic physiological model of BG dynamics to generate informative features for a support vector regression model to predict BG levels. Based on their experiments, this model outperforms diabetes experts at predicting BG levels, but its precision is still relatively low at about 42%. Decision trees (DTs) have been used for analysis and prediction in diabetes management [3,4]. For example, Pociot et al. proposed a novel DT-based analytical method to predict T1D mellitus [3]. Han et al. used DTs to build a model for diabetes prediction from the Pima Indians Diabetes dataset [4]. Sudharsan et al. applied random forest (RF), SVM, k-nearest neighbors, and naïve Bayes to predict hypoglycemia [5].

Based on the source of the training data, the aforementioned prediction models can be classified as either population-based (e.g., [3–6]) or patient-based (e.g., [7–10]). As already mentioned, they suffer from issues of low accuracy and/or sparse data.

3. Methodology

3.1. Data collection

Owing to the near-ubiquitous use of smartphones, we can use them to collect patient information such as BG measurements taken at regular intervals, and the corresponding daily events that impact the BG levels such as insulin, meals, exercise, and sleep. The ambient sensing features of a smartphone can help us collect some user information (e.g., exercise and sleep) automatically. Some other data (e.g., insulin dose) need to be entered manually. Because of its popularity, we mainly use touch-based gestures as the default input method. The collected data are preprocessed and automatically uploaded to the cloud.

3.2. Personalized prediction

We apply a three-stage evolution model to make more accurate and personalized BG predictions.

(1) Time-series prediction model based on patient data

Because smartphone-collected BG measurements have a natural temporal ordering, we can model the problem of predicting BG level as one of forecasting a time series. A time series is a sequence of data points, typically consisting of successive measurements made over a time interval. A time-series dataset differs from a regular one in that there is a natural ordering to the observations in the former. We confine our study to discrete time series. Our BG prediction problem is to use BG measurements up to time t to predict a future BG level at time $t+1$. This BG prediction problem can be modeled as follows:

$$\mathrm{BG}_{t+1} = f(\mathrm{BG}_t, \mathrm{BG}_{t-1}, \mathrm{BG}_{t-2}, \ldots, \mathrm{BG}_{t-n}). \quad (1)$$

For an observed BG series with n points, where t refers to the most recent observation and $t-n$ is the most distant observation, a future BG value at $t+1$ can be estimated with a function f. Function f is known as the model, and is used to obtain an estimated value BG_{t+1}.

(2) Pooled-panel-data regression model

The aforementioned time-series regression model makes predictions based on an individual patient's historical data. If the system has a large amount of historical data for the individual, this model can make better predictions. However, an individual patient may have very limited historical data. In the extreme case, the individual might not have any previous data at all. We understand that the inability to predict with time-series regression models is a small-sample issue. To address this issue, we propose to employ panel data to increase the sample size. A pooled panel data (PPD) regression model takes into account the fact that historical information of all patients enables cross-patient information sharing and thus overcomes issues related to data sparsity.

We choose linear regression to realize our PPD regression model for predicting BG level. Linear regression can model the relationship between the scalar dependent variable BG and the explanatory (or independent) variables. Let BG denote the dependent variable whose values we wish to predict, and let $X_1, \ldots, X_k$ denote the independent variables from which it is to be predicted; the value of variable X_i in period t (or in row t of the dataset) is denoted by X_{it}. For example, in our dataset, X_i could be insulin dose, hypoglycemic symptoms, meal ingestion,