Applied Soft Computing 20 (2014) 40–53

Contents lists available at ScienceDirect

Applied Soft Computing

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

Modeling glycemia in humans by means of Grammatical Evolution

J. Ignacio Hidalgo [a,*], J. Manuel Colmenar [b], José L. Risco-Martin [a], Alfredo Cuesta-Infante [b], Esther Maqueda [c], Marta Botella [d], José Antonio Rubio [d]

[a] Dept. Arquitectura de Computadores y Automática, UCM, Spain
[b] CES Felipe II, Univ. Complutense de Madrid (UCM) at Aranjuez, Spain
[c] Servicio de Endocrinología y Nutrición (SEN), Hosp. Virgen de la Salud, Toledo, Spain
[d] SEN, Hosp. Univ. Principe de Asturias, Alcalá de Henares, Spain

ARTICLE INFO

Article history:
Received 31 March 2013
Received in revised form 2 October 2013
Available online 4 December 2013

ABSTRACT

Diabetes mellitus is a disease that affects to hundreds of millions of people worldwide. Maintaining a good control of the disease is critical to avoid severe long-term complications. In recent years, several artificial pancreas systems have been proposed and developed, which are increasingly advanced. However there is still a lot of research to do. One of the main problems that arises in the (semi) automatic control of diabetes, is to get a model explaining how glycemia (glucose levels in blood) varies with insulin, food intakes and other factors, fitting the characteristics of each individual or patient. This paper proposes the application of evolutionary computation techniques to obtain customized models of patients, unlike most of previous approaches which obtain averaged models. The proposal is based on a kind of genetic programming based on grammars known as Grammatical Evolution (GE). The proposal has been tested with in silico patient data and results are clearly positive. We present also a study of four different grammars and five objective functions. In the test phase the models characterized the glucose with a mean percentage average error of 13.69%, modeling well also both hyper and hypoglycemic situations.

1. Introduction

Diabetes mellitus is a disease caused by a defect in either the secretion or in the action of insulin, which is essential for the control of blood glucose levels. Both of them cause in cells not to assimilate the sugar and, as a consequence, there is a rise in blood glucose levels, or hyperglycemia. Several types of diabetes differ in origin. According to the ADA (American Diabetes Association) we can distinguish four types of diabetes:

- Type 1 Diabetes (T1DM): cells do not produce insulin because of an autoimmune process. Currently, requires the person to inject insulin or wear an insulin pump.
- Type 2 Diabetes (T2DM): results from insulin resistance, where cells fail to use insulin properly, sometimes combined with an absolute insulin deficiency.
- Gestational diabetes: appears in the gestation period in one out of ten pregnant women. Pregnancy is a change in the body's metabolism, since the fetus uses the mother's energy for food, oxygen and others. This causes a decrease in the secretion of insulin from the mother.
- Other types: such as problems on β-cells, genetic defects affecting insulin action, induced by drugs, and genetic syndromes.

In most cases, diabetic patients with long time evolution need exogenous insulin either injected into various injection doses, or introduced by an insulin pump. It is important to maintain good glycemic control to prevent not only from the acute complications specific to diabetes (diabetic ketoacidosis and hypoglycemia, defined as blood glucose value less than 70 mg/dl), but also from a set of multi-chronic complications associated with diabetic patients: nephropathy, retinopathy, microangiopathy and macroangiopathy.

In recent years, it has been shown that a strict glycemic control in critically ill patients improves performance and reduces medical costs [1,2]. Glucose levels control is a demanding and difficult task for both patients and their families. To keep good levels of blood glucose, the patient must have some capacity of prediction to know what level of glucose would have if ingested a certain amount of food or injected with a quantity of a insulin of a certain kind. In fact, the objective is to avoid not only long periods of hyperglycemia (glucose levels $\geq$120 mg/dl) but also episodes of severe hypoglycemia (glucose levels $\leq$40 mg/dl) that can lead to patient death.

* Corresponding author. Tel.: +34 679562712; fax: +34 913947527.
E-mail addresses: joseignacio.hidalgoperez@gmail.com, hidalgo@dacya.ucm.es (J.I. Hidalgo).

1568-4946/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2013.11.006

One of the aspects that make it difficult to control blood glucose level is the lack of a general model of response to both insulin and the various factors mentioned above, due to the particularities of each patient [3]. Models in the literature apply classical modeling techniques, resulting in linear equations, defined profiles, or models with a limited set of inputs. Here we propose a novel technique that involves obtaining the patient model using genetic programming (GP). GP eliminates barriers in building the model, such as linearity or limitations on the input parameters.

Evolutionary techniques such as GP, have certain characteristics that make them particularly suited to address optimization problems and complex modeling. First, they are conceptually simple in its application but have a theoretical basis defined and widely studied. GP has demonstrated its applicability to many real problems, and is intrinsically parallelizable to work with a set of solutions. Furthermore, EAs have great potential to incorporate knowledge about the domain and to incorporate other search mechanisms (not necessarily evolutionary).

One of the best known applications of GP is symbolic regression and the application of one of its variants, Grammatical Evolution (GE), allows to obtain solutions that incorporate non-linear terms. GE is an evolutionary computation technique established in 1998 by Conor Ryan's group at the University of Limerick (Ireland) [4]. GP aims to find an executable program or function that respond to the reference data. The key advantage is that GE applies genetic operators to a whole chain, which simplifies the search application in different programming languages. In addition, there are no memory problems, unlike with GP where the tree representation could have the well know problem of bloating (an excessive growing of the computer structures in memory). Hence, we propose to apply GE to find a custom model that describes and predicts the blood glucose level in a patient. Our method takes the historic data of a patient consisting in previous glucose levels, ingested carbohydrates and injected insulin, and obtains an expression that can be used to predict near future glucose values. The contributions of this work are:

- We propose a method based on GE to obtain individualized and customized glycemia (glucose level in blood) models in humans.
- We have tested this proposal with five in silico patients taken from AIDA simulator [5].
- We present a study of four different grammars and five objective functions.
- We have selected the best models for each patient and run a test phase with a new dataset. In the test phase the models characterized the glucose with a mean percentage average error of 13.69%, reflecting also a good representation of both hyper and hypoglycemic situations.

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 details how Grammatical Evolution can be applied to this problem. Section 4 shows the general model we propose, as well as the grammars, particular models and objective functions we have studied for the glucose estimation problem. Section 5 is devoted to the experimental setup, while Section 6 presents the results obtained in both training and test phases. Finally, Section 7 explains the conclusions and the future work.

2. Related work

Glucose level control is a very demanding and difficult task for both patients and their families. Trying to keep a good control of blood glucose involves to perform blood glucose regular measurements (which involves at least one puncture in each measure or using a continuous monitoring system during some periods), insulin dose estimation, carbohydrates estimation, analyze that information somehow and to have some capacity of prediction that allows the patient to know what level of glucose would have if ingested a certain amount of food or injected with a quantity of a insulin of a certain kind.

As we have already mentioned, one of the main problems in controlling and predicting blood glucose levels is the lack of reliable models of response to both insulin and the various factors involved. Although there are some general approximations, there are hardly few adapted to the particularities of each patient [3,6]. The models in the literature apply classical modeling techniques, resulting in linear equations defined profiles, or models with a limited set of inputs. There are other factors that make a good control hard to achieve [7]. For instance, we can mention that there is a significant delay between insulin administration and the appearance of insulin in the blood stream with the use of subcutaneous (SC) insulin. This delay time limits the achievable control performance on subcutaneous administration of insulin.

In [6] authors propose the use of models to maintain margins of robustness when there is a mismatch between the model and the patient. The approach used there is personalized using information a priori known (i.e., easy access) of patients to limit conservatism. However, this model only applies to linear models and cannot incorporate other factors such as exercise or stress that clearly affect the expected levels of glucose. Models based on data for individual subjects are often inaccurate, since clinical data in T1DM are not extensive enough to identify the exact models [8,9]. To obtain continuous series of data, glucose levels should be measured using a subcutaneous continuous glucose monitoring (CGM) system. To calculate the dose of insulin the patient or the physician may use different mechanisms and control algorithms. Hence, we can also find some personalized control approaches [10–13] corresponding to clinical practice. Current treatment for subjects with T1DM uses rates of basal insulin delivery, insulin to carbohydrate ratios (CHO) and individual correction factors, typically from observations of the specialist.

There are also some models used in artificial pancreas systems or models of closed loop control [14,15]. The main risk is hypoglycemia as a result of excessive insulin administration. However we know that it is possible to reach a good control with approximate models, provided that the model is related to the control objective [16,17]. Again, the most important factor for the focus of this paper is the lack of accurate individualized models. If there is an accurate model of the subject's response to insulin, the design of the controller is relatively simple using classical control techniques. Autoregressive models (AR) may be applied to overcome problems of identifiability [18,19], although those are not useful for controlling since they have not an exogenous input. Some protocols have also been proposed to improve the reliability of the models [9,14,20] but the possibilities for the design of experiments are limited due to the strict security requirements and limitations in clinical protocols.

There have been also different approaches to facilitate the diabetes control from commercial companies. However, most of them have been designed only for specific glucometers and when providing insulin recommendations, the model is not available. *Glucofacts Deluxe* by *Bayern* [21], *CoPilot Health Management System* by *Abbot* [22], and *MenaDiab* [23] by *Menarini* are some of them.

Although there are many works that use control models, up to the date the modeling problem has not been addressed by evolutionary computation techniques that, as mentioned, have a high potential to incorporate to the model factors which are difficult to quantify, in other words to collect system dynamics. The main new aspect is the use of individualized models, i.e. we obtain a solution of the problem for each set of data on a single patient or individual.