Applied Soft Computing 35 (2015) 310–332

Contents lists available at ScienceDirect

Applied Soft Computing

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

Controlling blood glucose variability under uncertainty using reinforcement learning and Gaussian processes

Mariano De Paula [a], Luis Omar Ávila [b], Ernesto C. Martínez [b,*]

[a] INTELYMEC – Centro de Investigaciones en Física e Ingeniería del Centro – CIFICEN – CONICET, Facultad de Ingeniería, Universidad Nacional del Centro de la Provincia de Buenos Aires – UNCPBA, Av. del Valle 5737, Olavarría B7400JWI, Argentina
[b] INGAR (CONICET—UTN), Avellaneda 3657, Santa Fe S3002 GJC, Argentina

ARTICLE INFO

Article history:
Received 17 April 2013
Received in revised form 24 June 2015
Accepted 25 June 2015
Available online 3 July 2015

Keywords:
Artificial pancreas
Diabetes
Gaussian processes
Policy iteration
Reinforcement learning
Stochastic optimal control

ABSTRACT

Automated control of blood glucose (BG) concentration with a fully automated artificial pancreas will certainly improve the quality of life for insulin-dependent patients. Closed-loop insulin delivery is challenging due to inter- and intra-patient variability, errors in glucose sensors and delays in insulin absorption. Responding to the varying activity levels seen in outpatients, with unpredictable and unreported food intake, and providing the necessary personalized control for individuals is a challenging task for existing control algorithms. A novel approach for controlling glycemic variability using simulation-based learning is presented. A policy iteration algorithm that combines reinforcement learning with Gaussian process approximation is proposed. To account for multiple sources of uncertainty, a control policy is learned off-line using an Ito's stochastic model of the glucose-insulin dynamics. For safety and performance, only relevant data are sampled through Bayesian active learning. Results obtained demonstrate that a generic policy is both safe and efficient for controlling subject-specific variability due to a patient's lifestyle and its distinctive metabolic response.

1. Introduction

Insulin-Dependent Diabetes Mellitus (IDDM) is a chronic disease characterized by the inability of the pancreas to produce sufficient amounts of insulin. A high level of BG concentration is known to cause serious health problems, including heart disease and stroke, hypertension, retinopathy, nephropathy, and neuropathy [1,2]. Poorly controlled diabetes mellitus is associated with multiple long-term complications that contribute to increased morbidity and mortality. Also, abnormal glycemic variability contributes to oxidative stress, which has been linked to the pathogenesis of diabetes [3,4]. Compensating for this deficiency in endogenous insulin production requires 4–6 insulin injections to be taken daily; the aim of this diabetes therapy is to maintain normoglycemia – i.e., a blood glucose level between 4 and 7 mmol/L. In defining the amount and timing of these injections, poor predictability of BG dynamics is a key issue that both patients and doctors must deal with [5]. Manual control of BG often results in high glycemic variability and the risk of a life-threatening hypoglycemic event is at stake. Hypoglycemia – i.e., low blood glucose levels – may lead to brain damage, coma and eventually death.

Closing the glucose control loop with a fully automated artificial pancreas will certainly improve the quality of life for insulin-dependent patients. Such a device is made up of a glucose sensor, an automated insulin infusion pump, and a feedback control strategy or control algorithm that calculates the insulin delivery based on a glucose signal. The major challenge for the development of a closed-loop control system is the glycemic variability between subjects and for the same subject over time. A reliable closed-loop system for blood glucose regulation should be able to adapt "on the fly" to each patient response in order to cope with daily variations in glucose metabolism. Another challenge in controlling variability of BG is sensor errors. These errors depend nonlinearly on the BG rate of change and are subjected to a delay due to its subcutaneous nature. Additionally, the sensor noise is non-white (non-Gaussian) and consecutive sensor errors are highly interdependent [6]. Despite significant advances, the available technology for continuous glucose monitoring is still ineffective to deal with many issues related to sensitivity, stability, calibration, and the physiological time lag between BG and interstitial glucose (IG) concentration. Significant delays in delivering insulin to the blood stream give rise to delayed effects of control actions which increase the risk of hypoglycaemia episodes. Hypoglycaemia has been identified as the

* Corresponding author. Tel.: +54 342 4534451; fax: +54 342 4553439.
E-mail address: ecmarti@santafe-conicet.gov.ar (E.C. Martínez).

http://dx.doi.org/10.1016/j.asoc.2015.06.041
1568-4946/

primary concern for a safe implementation of the artificial pancreas [7].

Pioneering works with the proportional-derivative (PD) and the proportional-integral-derivative (PID) types of controllers demonstrated the advantages of closed-loop control [8]. The resulting control strategies, however, generally do not prove sufficiently effective in maintaining euglycaemia after meals when the less invasive choice of the subcutaneous route is used. PID algorithms can be considered reactive, as they respond to observed glucose levels and are less equipped to take advantage of announced meals and patient-directed insulin boluses. In a recent experimental study with a PID controller using the subcutaneous–subcutaneous (SC–SC) route [9], the observed mean glucose levels are not sensibly different to the ones corresponding to the uncontrolled glucose dynamics. Also, the two-hour postprandial glucose levels were significantly higher than those observed in healthy subjects under similar conditions. Glucose variability has a highly detrimental effect on the performance of a PID controller since even for subjects with a strict lifestyle, gain scheduling [10] and adding a feedforward element based on manual entry of a meal disturbance (time and content) are required for satisfactory blood glucose control [41]. To prevent a hypoglycemic condition, PID controllers can be enhanced using a supervisory module that constrains insulin delivery by limiting the maximum infusion rate or by suspending altogether close-loop control when glucose levels are approaching a lower threshold or are decreasing too rapidly. However, the inability of PID controllers to accommodate system constraints in the computation of control actions further limits their potential for success with patients that have active life styles. Moreover, due to the myopic nature of its feedback law, a PID controller is unable to cope with delayed effects of control actions. Typically, in systems where the effects of control actions slowly unfold over time, a PID overreacts, which increases glycemic variability. The lag time associated with subcutaneous (SC) insulin infusion is an obstacle for any reactive control algorithm.

A more suitable control framework for systems with large lag times and constraints is model predictive control (MPC) which has been proposed as a promising architecture for insulin delivery in the artificial pancreas [11–14]. The predictive framework is a powerful tool not only to deal with time delays in the system response, but also to evaluate the future effects of a meal and thus achieving disturbance rejection. Constraint handling and penalizing input actions (which will avoid too aggressive control actions that may lead to hypoglycemia) are also advantages of model-based control methods. However, one serious drawback of model-based control systems is that the controller performance is strongly dependent on the accuracy of the model used to represent the glucose-insulin dynamics. Most of the glucose-insulin models proposed in the literature are physiological compartmental models that are generally representative of only an average subject under specific conditions [15–18]. The metabolic processes underlying insulin action involve complex interactions of hormones, which lead to significant variation in insulin sensitivity [19]. Another disadvantage is that to implement a MPC requires repeatedly obtaining an on-line solution of a mathematical program. This on-line optimization can be avoided with a single set of a priori optimizations via multiparametric programming; the on-line problem is thus reduced to the evaluation of an affine function obtained from a lookup table [14].

More recently, new approaches combining an Iterative Learning Control (ILC) scheme with MPC strategies have been proposed. These hybrid methods try to benefit from the repetitive nature of insulin therapy to improve iteratively the efficacy of insulin doses by using run-to-run control algorithms [20,21,42,43]. In a situation with frequent data sampling, iterative learning control (ILC) is the alternative of choice. ILC attempts to mimic human learning in order to take advantage of subject-specific variation patterns in the glucose-insulin dynamics. A key issue for a successful ILC implementation is the design of a feedback control law that can handle inter- and intra-patient glycemic variability. To address the latter, adaptive control of blood glucose is considered a worth exploring alternative [22,23]. Recently, an adaptive model-based control strategy has been proposed by Oruklu et al. [24] which can dynamically detect blood glucose variations, and on that basis reject glycemic disturbances. The adaptability of the controller is based on subject-specific recursive linear models developed using data from a continuous glucose monitoring (CGM) sensor along with a change detection algorithm. Metabolic variations in a subject's body are addressed by online model identification. At each step, model parameters are updated by using new glucose data, and the future time course of BG concentration is estimated. These parameters are then used in a model-based control algorithm for calculating the appropriate insulin infusion rate.

In recent years, clinical evaluations of different strategies for close-loop artificial pancreas systems have been reported. In Dassau et al. [25], a fully automated multi-parametric model predictive control algorithm with insulin *on-board* was experimentally tested with encouraging results. The first wearable AP outpatient study using a meal-informed MPC strategy was reported by Del Favero et al. [26] aiming to investigate the ability to control postprandial glucose. Despite promising results were obtained in short-term studies for a single meal (dinner), long-term randomized studies with numerous meals are needed to prove superiority of MPC over the commonly used bolus calculator. A bi-hormonal closed-loop artificial pancreas was experimentally assessed by El-Khatib et al. [27]. Even though results demonstrate the feasibility of safe BG control by a bi-hormonal artificial endocrine pancreas, inter- and intra-subject variability in metabolic responses to insulin and glucagon could hamper the effectiveness of the control algorithm. Reactive control algorithms need to integrate learning capabilities upon which they can promptly respond to the varying activity levels seen in outpatients, with unpredictable and unreported food intake and stress conditions, and may also provide the necessary personalized glucose control for individuals [28]. Controlling insulin delivery in a closed-loop using reinforcement learning algorithms revolves around obtaining robust, yet optimal control policies that are reactive to the immediate needs of the patient.

In this work, a novel approach for controlling variability in blood glucose concentration using simulation-based learning of a robust control policy is presented. To account for inter-patient variability, a policy iteration algorithm that combines reinforcement learning with Gaussian process approximation is proposed. A generic control policy is learned off-line using an Ito's stochastic process model of the glucose-insulin dynamics that simulates glycemic variability comprehensively. For safety and performance, only relevant data are sampled through Bayesian active learning.

2. Modeling glycemic variability

Glucose-insulin dynamics exhibits significant variability from patient to patient. Due to this uncertain behavior even the same insulin dose with the same meal routine and the same amount of physical exercise may result in different blood glucose responses to insulin injections on consecutive days. Furthermore, blood glucose levels vary among different patients according to carbohydrate contents, exercise levels, age and stress [29]. Natural inter- and intra-patient variability needs to be addressed in developing an optimal glucose control profile. Since the interactions between insulin, meals, exercise and other factors and their effect on blood glucose is an on-going phenomenon, there exists significant uncertainty in the actual response of a patient to control actions.