Master Thesis - Project Specification

HbA1c Impact Estimation of Detected Events from Continuous Glucose Monitoring Data

THONY PRICE

Master in Computer Science
Date: February 4, 2019
Supervisor: Pawel Herman
Examiner: Erik Fransen
Swedish title: Estimerad HbA1c Inverkan Av Detekterade Event ur Kontinuerlig Glukos Matardata
School of Electrical Engineering and Computer Science

Abstract

Diabetes affect millions of people today and health care emphasize on self management by the patients. Continuous glucose monitoring sensors have enabled close monitoring of the current and historical blood glucose concentration for both patients and clinicians. This data may also include valuable insight about patterns that have an impact on the long term progression of blood glucose development in patients.

This thesis aims to investigate a new approach to detect events in blood glucose data and estimate their impact on the long term blood glucose concentration progression. The method includes filtering the signal, derive a qualitative representation which is fed to a hidden Markov model. The model detects events which by intervention analysis is labeled with an estimated impact.

Contents

Chapter 1

Introduction

This chapter intends to provide the reader an overview of the current state of diabetes healthcare as well as explain the outline and aim of this thesis project.

1.1 Project Introduction

In all countries the disease burden related to diabetes is already high, and it is steadily increasing [8]. In 2017 the estimated prevalence of diabetes was 451 million people globally and approximately 5 million deaths were attributed to diabetes [3]. Aside from reduced life expectancy, diabetes increase the risk of multiple other serious conditions such as heart disease and stroke [8]. With a projected prevalence of 693 million diabetes patients in 2045 and given the seriousness of diabetes, proper medical care for patients are of utmost importance [3].

Diabetes is a group of metabolic diseases that is characterized by hyperglycemia. It is a consequence of defects in insulin secretion, insulin action, or both [1]. Insulin is necessary to maintain normal blood glucose concentration (BGC) internally by facilitating cellular glucose uptake and regulating carbohydrate metabolism [13]. The vast majority of cases of diabetes fall into two broad categories, type 1 diabetes (T1D) and type 2 diabetes. T1D is caused by an absolute deficiency of insulin secretion, thus patients need to induce exogenous insulin on a

regular basis to maintain a balanced BGC [1].

Maintaining balanced BGC is an every day challenge of T1D patients. Treatment guidelines put heavy emphasis on self management activities that benefit a balanced regulation. Guidelines include recommendations of eating patterns, exercise, carbohydrate consumption etc. Patients traditionally consult a clinician regularly to get evaluate these activities and plan ahead for the continued self management process [4].

Continuous glucose monitoring (CGM) sensors are wearable devices that measures the blood glucose frequently, usually every 1-5 minutes. CGM sensors have revolutionized the ability for clinicians to review a patients data and deliver care driven by extensive data [6]. Care based CGM data have proven effective in lowering patients long term average BGC, known as HbA1c [5]. The data enable precise monitoring of immediate processes, such as accuracy of a specific insulin dosing or patient response to a certain meal.

Start slimming down to more specifics here! However, there is a lack of research in deriving insights for long term medical advice autonomously from CGM data.

1.2 Project Aim

Given CGM data of a patient, autonomously generate insight of which events and patterns in the data indicates a certain progression of long term HbA1c. For example, if provided 8 week's worth of data some segments of the data may infer a healthy development, such as healthy meals and exercise. The aim is to detect these events in the CGM data in order to classify which activities have a positive impact on a patient's HbA1C.

1.2.1 Objective

The objective of this thesis is to investigate how an autonomous system, which fulfills the criterions below, can be implemented and in which configuration it performs optimally. The system should:

- Detect events in from a batch CGM time series data (such as meals, exercises or sleep).
- Estimate intervention caused by one or more events (what impact did the event(s) have on the continued time series).

1.2.2 Research Question

1. Is it possible to detect events from a sequence of qualitative representations generated from a batch CGM data using a hidden Markov model?
2. Given one or more events in (1.), can intervention analysis estimate the long term impact of a single, or a set of events?

Chapter 2

Background

This chapter sets out define diabetes and the relevant aspects of T1D care enough to understand the content in latter chapters. It further describes theory of the techniques in the suggested method and reviews related work that should assist putting this thesis into context of the current field.

[This chapter is under construction. It should be continued when the method is decided upon.]

Research Suggestions:

1. CGM and connection to diabetes
2. Filtering/smoothing CGM data, what impact does it have?
3. Meal/Event Detection
4. Charbohydrate estimation (Samadi et. al.)
5. Time series analysis
6. Time series clustering
7. Time series labelling
8. Event evaluation
9. Supervised/unsupervised problem?
10. Podova - Logged T1D Trial to test against
11. Pretrainning?

Chapter 3

Method

3.1 Data

The data is collected during a 4,5 month pilot study which includes 20 patients. All patients are at least 18 years old and have been injecting insulin at least 1 year prior to the study. Each patient are equipped with a CGM sensor for the entirety of the study. For each patient, the CGM sensor measures the blood glucose concentration (BGC) at intervals of 1-5 minutes. The (time, BGC) tuples constitutes the structure of each patient's data set. Additionally, other events such as meals and exercise are logged manually by patients. A summary of collected data points are presented in Table 3.1.

Notation	**Field**	**Format**
t	Time	Date
t_{BG}	Glucose value at t	Float
t_{ACT}	Physical activity at t	Integer
t_{INS}	Injected insulin at t	Float
t_{IMG}	Food image at t	Float
t_{EVENT}	Manually reported by patient at t	Text

Table 3.1: The data for each patient include continuous measurements at time steps t of intervals between 1-5 minutes. Each measurement at t *always* include BG value and *may* contain other field presented in the table.

3.2 Implementation

The objective of the proposed system is to analyze the data for a closed time interval, identify events and classify them accordingly with respect to their influence over future measurements. The analysis is not performed in real time, all data is available immediately to the algorithms. The proposed steps of implementation can be overviewed in Figure 3.1. Each step is described in more detail in its corresponding section below.

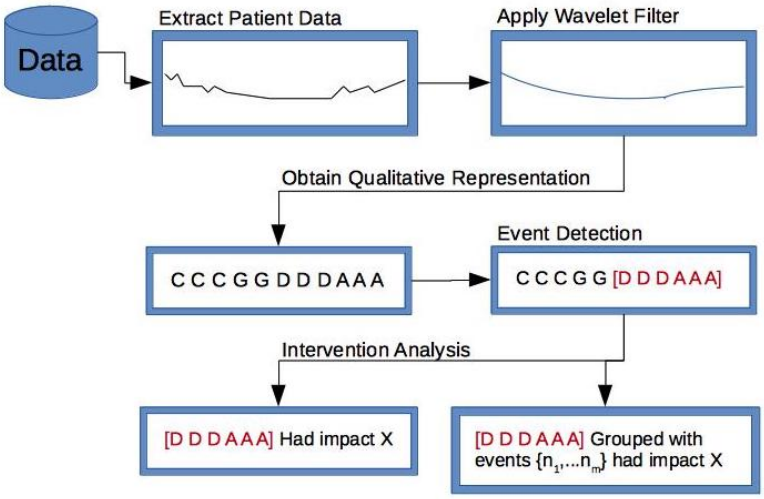


Figure 3.1: Schematics of implementation.

3.2.1 Wavelet Filter

Studies have shown data from CGM sensors is subject to distortion. This is caused by diffusion processes and by time-varying systematic under/overestimations due to calibrations and sensor drifts [7]. Noise can trigger false positives in event detection because abrupt fluctuations overrides the true underlying derivatives of the curve [6]. Wavelet filters have been used repeatedly with CGM data and proved successful in reducing noise while retaining events such as spikes [9], [6], [11]. Figure 3.2 displays an example of wavelet filtering applied to CGM data.

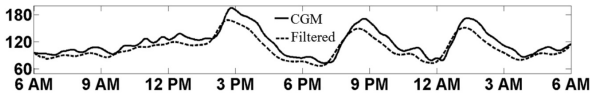


Figure 3.2: Wavelet filter applied on CGM data. Vertical axis represents glocose concentration [mg/dl]. Image courtesy of Samadi et al. [11].

3.2.2 Qualitative Representation

To identify events in the de-noised CGM data, feature extraction is used. Feature extraction can be achieved by either a qualitative or quantitative method. The qualitative method offer benefits such as more transparent reasoning and ability to provide explanations for for the solutions it provides [12].

In qualitative representation by triangular shapes, a CGM data segment can take seven shape variables. Figure 3.3 shows the different shapes. Each is a unique combination of the first and second order derivatives on the curve of the current segment. The derivates can be read from segment of adjacent points, allowing the CGM data series to be presented as a sequence of shapes describing fluctuations in BGC.

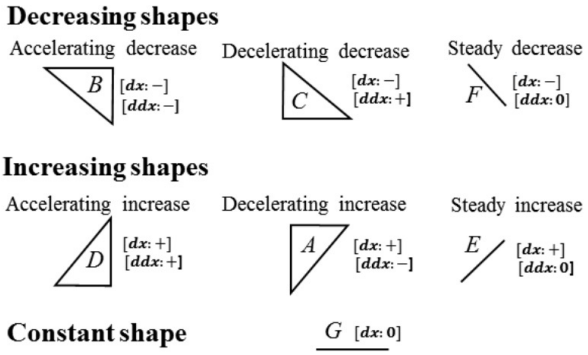
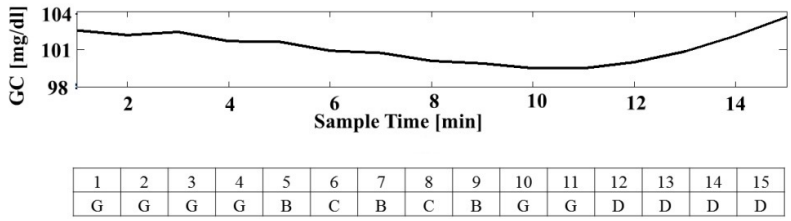


Figure 3.3: Scheme of the qualitative variables A-G. Image courtesy of Samadi et al. [11].

3.2.3 Event Detection

With the qualitative representation, event detection can be performed by analyzing the sequence of shapes. In figure 3.4 an event could be

triggered by identifying a continuously accelerating increase (for example the four sequential D's from time-step 12).



1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
G	G	G	G	B	C	B	C	B	G	G	D	D	D	D

Figure 3.4: Shape sequence representation if CGM curve. Image courtesy of Samadi et al. [11].

Given a sequence of observations (i.e. shapes) with some underlying state (i.e. eating, sleeping, exercising) a hidden Markov model can provide an estimate of what state a current time step most probably corresponds to. The Markov approach includes two steps:

The evaluation problem: Calculating probability distribution by training a model on sequences with labeled states.

The decoding problem: Determining the hidden states by observing a unlabelled sequence and finding the optimal state sequence associated with the given observation sequence [10].

3.2.4 Intervention Analysis

Intervention analysis provides a tool to asses how much a given event has changes the series (if at all) [2]. The analysis is able to detect 4 patterns:

1. Permanent constant change to the mean level.
2. Brief constant change to the mean level.
3. Gradual increase or decrease to a new mean level.
4. Initial change followed by gradual return to previous mean level.

The mean level in our case is the mean BGC. Intervention analysis compare the mean average before the event and its progression afterwards. The characteristics of the progression curve defines which of the four patterns an event has triggered. Because changes in mean BG

concentration are subtle and changes takes place over a longer timespan, an alternative to the approach is suggested.

An event by itself may not cause a observable change in the mean average but combined events might. The intervention analysis is also evaluated on sets of adjacent events. The output can hint on when the aggregated data implies a positive development of the mean. In other words, the analysis can suggest when combined activities by a patient results in a change of his or her long term HbA1c value.

3.3 Evaluation

The method is evaluated in two steps:

Detection of events: For each patient, a 80/20% data split is made into training and test data respectively. The training data is utilized to solve the Markov evaluation problem and find model parameters. The labels are removed from the test data the events are produced from the Markov decoding step of the algorithm. A comparison of the removed (true) labels and the proposed ones are made and accuracy and precision measurements serves the basis of this evaluation.

Intervention analysis: Given that the algorithm has produced a set of events with suggested interventions. A clinician is presented the same data and suggested events. The clinician performs a independent assessment of the probable intervention caused by these events. The algorithmic output is then compared to the clinicians and an accuracy and precision score is studied to evaluate the performance of the intervention analysis.

Chapter 4

Results

4.1 Event Detection

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

4.2 Events Clustering

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

4.3 Scores against Doctors

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

4.4 Podova Scores

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 5

Discussion

5.1 Header 1

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Chapter 6

Conclusion

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Bibliography

[1] American Diabetes Association. "Diagnosis and Classification of Diabetes Mellitus". In: *Diabetes Care* 33.Supplement 1 (2010), S62–S69. ISSN: 0149-5992. DOI: 10.2337/dc10-S062.

[2] George EP Box et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[3] N. H. Cho et al. "IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045". In: *Diabetes Research and Clinical Practice* 138 (Apr. 2018), pp. 271–281. ISSN: 0168-8227. DOI: 10.1016/j.diabres.2018.02.023.

[4] Debbie Cooke et al. "Structured Type 1 Diabetes Education Delivered Within Routine Care". In: *Diabetes Care* 36.2 (2013), pp. 270–272. ISSN: 0149-5992. DOI: 10.2337/dc12-0080.

[5] Kathleen Dungan and Neha Verma. "Monitoring technologies–continuous glucose monitoring, mobile technology, biomarkers of glycemic control". In: *Endotext [Internet]*. MDText. com, Inc., 2018.

[6] Andrea Facchinetti. "Continuous Glucose Monitoring Sensors: Past, Present and Future Algorithmic Challenges". In: *Sensors (Basel)* 16.12 (Dec. 2016). PMC5191073[pmcid], p. 2093. ISSN: 1424-8220. DOI: 10.3390/s16122093.

[7] Andrea Facchinetti et al. "Modeling the glucose sensor error". In: *IEEE Transactions on Biomedical Engineering* 61.3 (2014), pp. 620–629.

[8] Nita Gandhi Forouhi and Nicholas J. Wareham. "Epidemiology of diabetes". In: *Medicine (Abingdon)* 42.12 (Dec. 2014). 25568613[pmid], pp. 698–702. ISSN: 1357-3039. DOI: 10.1016/j.mpmed.2014.09.007.

[9] Nicolas Magdelaine et al. “Wavelets for CGM off-line denoising”. In: (Feb. 2016). DOI: 10.13140/RG.2.1.2048.2326.

[10] Lawrence R Rabiner. “A tutorial on hidden Markov models and selected applications in speech recognition”. In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286.

[11] Sediqeh Samadi et al. “Meal detection and carbohydrate estimation using continuous glucose sensor data”. In: *IEEE journal of biomedical and health informatics* 21.3 (2017), pp. 619–627.

[12] Venkat Venkatasubramanian et al. “A review of process fault detection and diagnosis: Part III: Process history based methods”. In: *Computers and Chemical Engineering* 27.3 (2003), pp. 327–346. ISSN: 0098-1354. DOI: https://doi.org/10.1016/S0098-1354(02)00162-X.

[13] Gisela Wilcox. “Insulin and insulin resistance”. In: *Clin Biochem Rev* 26.2 (May 2005). PMC1204764[pmcid], pp. 19–39.