Negative results for the prediction of postprandial hypoglycemias from insulin intakes and carbohydrates: analysis and comparison with simulated data

Fabien Dubosson[1], Natalia Mordanyuk[2], Beatriz López[2], and Michael Schumacher[1]

[1] AISLab group, University of Applied Sciences Western Switzerland
[2] eXiT group, University of Girona

Abstract. Diabetic patients usually take insulin bolus right before eating a meal. A wrong dosage of insulin may lead to a hypoglycemia. Being able to anticipate such insulin-induced, postprandial hypoglycemias would enable warning of the patients about the risk associated with the quantity of insulin they are planning to take. In this work, we explore the feasibility of predicting these postprandial hypoglycemias by using information available at pre-meal time, such as glucose levels, planned insulin intakes and carbohydrates estimations. First, an experiment has been done on a dataset acquired on real patients, for which several classes of machine learning algorithms have been tried. The obtained results do not offer predictions that are useful enough to consider any usage in real-life applications. These kinds of datasets — acquired on real patients — suffer heavily from missing data and incorrect carbohydrates estimations though. In order to analyse the impact of these flaws on the obtained results, the same experiment has been run on a simulated dataset. Results support that even with the simulated dataset, which does not have missing data and which has precise carbohydrates intake, these features alone are not able to predict postprandial hypoglycemia. Therefore, improving the quality of patients annotations is not enough to solve the problem, and using these features without further features engineering does not offer good results.

Keywords: Hypoglycemia, Prediction, Insulin, Carbohydrates

1 Introduction

Type I diabetes patients usually take insulin bolus before starting a meal. The dosage of insulin depends on the quantity of carbohydrates they are going to ingest, and this quantity is usually estimated by the patients themselves. These estimations may be error prone, and taking too much insulin may lead to hypoglycemias which may be dangerous for patients lives in the long term.

In this work we explore the possibility of anticipating such postprandial hypoglycemias by using the information available juste before a meal: patient's glycemia levels, planned insulin intakes and carbohydrates estimations. Being able to do so would enable warning of the patients about the risk associated with the dosage of insulin they are planning to take.

A big part of the work going in the direction of hypoglycemias predictions focuses on live predictions from Continuous Glucose Monitoring (CGM) signals [1, 2]. Some are specially oriented toward patients with insulin pumps, as they try to detect when to stop basal insulin to prevent hypoglycemias [3, 4]. Our work, however, is targeting different audiences and objectives. The goal is to offer the possibility to anticipate hypoglycemias to patients not wearing any insulin pump or any continuous glucose monitoring device. This may assist patients to be confident about the planned insulin intake.

In [5], Reddy et al. presented a bolus calculator based on CGM device using Case-Based Reasoning (CBR) methods. They evaluated it while acquiring the dataset used in this study. Their work is closer to the objectives of this experiment than previously cited related works. However, their method is based on CGM devices as opposed to our work.

The two datasets are presented in Section 2. The Section 3 explains the methodology and presents the results, and finally results of this research are discussed in Section 4.

2 Datasets

This work is based on two datasets: a first one that have been acquired on type I diabetes patients, and a second one that have been generated by a simulator. These two datasets are presented in the following subsections.

2.1 Real-patients dataset

The database used in this study was provided by the Imperial College London [5]. The population consists of 10 patients (men and women), aged between 24 and 74 suffering from type I diabetes. The patients have been enrolled to a 6-weeks acquisition session, during which some of the patients' measures have been collected. In addition, the patients were provided a mobile application with an insulin bolus dose decision support system based on CBR [6].

The devices used for gathering the glucose levels of patients were Medtronic iPro2 Recorders[3]. The CGM took readings every 5 minutes, 24 hours a day over 7-10 consecutive days, and the units used to measure the glucose were *mmol/l*. The CGM data was supplemented with additional information provided by the patients, such as carbohydrates ingested (g), the insulin shots (U), alcohol consumption (*True* or *False*), or whether the patient performed some physical activities or not before the meal.

Initially, there were 2404 logbook entries (all the patients together). After a data cleaning process (due to missing values caused by the errors of the glucose

[3] `http://www.professional.medtronicdiabetes.com/ipro2-professional-cgm`

recorder, superfluous entries concerning sensor events, or due to small time interval meals), 1158 entries remained. Since the logbook entries are associated with CGM recordings, we also excluded all the sequences having too many missing data in the CGM, leaving a total of 891 entries.

2.2 Simulated dataset

The simulated dataset has been generated using the *UVA/PADOVA Type I Diabetes Simulator* [7, 8].

It consists of 10 virtual patients, for whom 500 days of data have been generated for each. The data consists of CGM measurements every 5 minutes with the associated quantity of carbohydrates ingested and insulin units taken. This represents 1500 entries per virtual patients for a total of 15 000. The necessary pre-processing have been done in order to convert units to the ones of the real dataset. Most where straightforward conversions, except for the insulin dosage. In the simulated dataset, the insulin dosage are reported for insulin pumps, with the following definition:

$$\mathrm{IIRt} = \mathrm{IU} \times \frac{6000}{\mathrm{BW}} + \frac{\mathrm{Basal}}{60} \times 5 \times \frac{6000}{\mathrm{BW}} \tag{1}$$

Where *IIRt* is the value provided by the simulator, *IU* is the insulin units we want to know, *Basal* in the basal insulin, and *BW* is the patients body weights. In order to use the same type of data as in the real-patients dataset to be able to compare results, the formulae has ben transformed as follow:

$$\mathrm{IU} = \frac{\mathrm{IIRt} \times \mathrm{BW} - \mathrm{Basal} \times 500}{6000} \tag{2}$$

The *IU* value has then been used as the insulin intake feature in the experiments.

3 Methodology and Results

Each log entry was labelled according to Zecchin [9]: the glucose below 70 mg/dL (3.889 mmol/l) was considered hypoglycemia, while glucose above 180 mg/dL (10 mmol/l) was considered hyperglycemia; other glucose levels correspond to the normoglycemia state. In this experiment we regrouped normoglycemia and hyperglycemia in the same class because we only want to predict hypoglycemias.

Since most nadirs occur around 2 hours after ingestion of a carbohydrate meal [10, 9], we looked for hypoglycemias between 1.5 hours and 2.5 hours after a meal for defining the insulin-induced hypoglycemia class labels. This results, as expected, in quite imbalanced datasets: 827 non-hypos for 64 hypos in the real-patients datasets, 14 923 non-hypos for 67 hypos in the simulated one.

Several families of machine learning algorithms have then been tried with the Python *scikit-learn* library[4]: linear classifier, nearest neighbors, random forest,

[4] `http://scikit-learn.org/`

extra trees and SVM. We used typical machine learning good practices: evaluations have been done with 10-folds cross-validation, within each cross-validation loop the models parameters have been fine-tuned with a cross-validated grid-search on the training set, and the classes weights have been set to compensate the class imbalance.

The selection of the scoring method is a more subjective task. The accuracy alone is not useful on imbalanced datasets because answering always the majority class gives high scores without being useful at all. Precision and Recall both have their utility. Precision relates to the number of false alarms, which is important to keep patients adherence to the system. Recall relates to the percentage of detected hypoglycemias, and we are trying to avoid hypoglycemias so it's important to detect the maximum number of them. The F1 score has been selected as it gives the harmonic means between precision and recall, but this choice over other scoring method is arbitrary.

Table 1. F1 scores

Algorithm	**Patients dataset** (Prior=13.4)	**Simulated dataset** (Prior=0.89)
Linear classifier	10.93	**1.15**
Nearest neighbors	5.76	0.55
Random forest	2.41	0.50
Extra trees	6.72	0.58
SVM	**13.70**	0.81

The prior F1-score of the real-patients dataset is 13.4% and the best result is obtained with the SVM classifier, which achieves 13.7%. The majority of non-hypos are classified correctly, as well as the majority of hypos that are also classified correctly. The number of false-alarm and the number of missed hypoglycemias prevent any application in real-word though. On the simulated dataset the results are not really better: the prior F1-score is 0.89% and the linear classifier achieves a score of 1.15%. This does not allow either any use in real-world scenarios.

4 Discussions

The best results obtained on the patients dataset are slightly better than the dataset priors, but are not good enough to be used in any real-world application. One point that have been noted while working on this dataset are unexplained glucose peaks, and the most likely cause of this should be missing carbohydrates information. Another related weak point of such real-life dataset is the fact that accurate carbohydrates estimations are difficult. In order to evaluate if these flaws may be explaining the difficulties of prediction on the real-patients dataset, we reproduced the same experiment on the simulated dataset.

The experiment on the simulated dataset, does not seem to offer significant improvements. Being generated by a simulator, the data should however behave

more predictively than a human body because the human body is much more complex and is sensible to the external environment. The simulator is also giving the exact and complete set of carbohydrates and insulin intakes, in contrast to human annotations.

This experiment supports first that the weakness in patients annotations alone is not enough to explain the difficulties of hypoglycemias prediction, otherwise good results would have been reached on the simulated dataset. Second, the absence of better results on the simulated dataset, despite having been tested with different families of machine learning algorithms, shows that the features used (glucose levels, insulin intakes and carbohydrates estimations) can not predict hypoglycemias as-is. This does not mean however that the problem may not be solved through further features engineering or by using more complex models. A more detailed study of the features and models would help to identify the factors involved in the difficulties of such predictions, and would permit to propose guidelines for improving the acquisition of new Type I diabetes datasets.

References

1. Palerm, C.C., and Bequette, B.W.: Hypoglycemia detection and prediction using continuous glucose monitoring—a study on hypoglycemic clamp data. (2007)
2. Eren-Oruklu, M., Cinar, A., and Quinn, L.: Hypoglycemia prediction with subject-specific recursive time-series models. (2010)
3. Hughes, C.S., Patek, S.D., Breton, M.D., and Kovatchev, B.P.: Hypoglycemia prevention via pump attenuation and red-yellow-green "traffic" lights using continuous glucose monitoring and insulin pump data. Journal of diabetes science and technology 4(5), 1146–1155 (2010)
4. Buckingham, B., Cobry, E., Clinton, P., Gage, V., Caswell, K., Kunselman, E., Cameron, F., and Chase, H.P.: Preventing hypoglycemia using predictive alarm algorithms and insulin pump suspension. Diabetes technology & therapeutics 11(2), 93–97 (2009)
5. Reddy, M., Pesl, P., Xenou, M., Toumazou, C., Johnston, D., Georgiou, P., Herrero, P., and Oliver, N.: Clinical Safety and Feasibility of the Advanced Bolus Calculator for Type 1 Diabetes Based on Case-Based Reasoning: A 6-Week Nonrandomized Single-Arm Pilot Study. Diabetes Technol Ther 18, 487–493 (2016)
6. Aamodt, A., and Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI communications 7(1), 39–59 (1994)
7. Man, C.D., Micheletto, F., Lv, D., Breton, M., Kovatchev, B., and Cobelli, C.: The UVA/PADOVA type 1 diabetes simulator: new features. Journal of diabetes science and technology 8(1), 26–34 (2014)
8. Visentin, R., Dalla Man, C., Kovatchev, B., and Cobelli, C.: The University of Virginia/Padova type 1 diabetes simulator matches the glucose traces of a clinical trial. Diabetes technology & therapeutics 16(7), 428–434 (2014)
9. Zecchin, C.: Online Glucose Prediction in Type 1 Diabetes by Neural Network Models. Università degli Studi di Padova (2014).
10. Brun, J.F., Fedou, C., and Mercier, J.: Postprandial reactive hypoglycemia. Diabetes & Metabolism 26(5), 337–351 (2000)