Abstract

In type 1 diabetes management, mobile health applications are becoming a cornerstone to empower people to self-manage their disease. There are many applications addressed to calculate insulin doses based on the current information (e.g., carbohydrates intake) and a few of them are accompanied by modules able to supervise postprandial conditions and recommend corrective actions if the user falls in an abnormal state (i.e., hypoglycaemia or hyperglycaemia). On the other hand, mobile apps favour the gathering of historical data from which machine learning techniques can be used to predict if user conditions will worsen.

This work presents the application of k-nearest neighbour on the historical data gathered on patients, so that given the information related to a sequence of meals, the method is able to predict if the patient will fall in an abnormal condition. The experimentation has been carried out with the USA-Padova type 1 diabetes simulator over eleven adult profiles. Results corroborate that the use of sequential data significantly improve the prediction outcomes when forecasts distinguish the type of meal (breakfast, lunch and dinner).

Introduction

Type 1 Diabetes Mellitus (T1DM) is a chronic disease that demands a strict control of the Blood Glucose (BG) level of the patient. This BG control is required to avoid hypoglycaemia or hyperglycaemia events, which are associated to serious short-term and long-term complications, e.g., coma, blindness, severe kidney failure or even death. This paper studies how sequences of data (recorded by T1DM people) can be used to predict hypoglycaemia and hyperglycaemia events using the k-nearest neighbours (k-NN) method.

Methodology

In the problem faced in this paper, the following information is available: Time (T) in minutes; Carbohydrate (CH) intakes (mg); Bolus insulin dose (B); CGM readings (mg/dL). This information is used to create an event for each meal.

Table 1: Structure of a meal event

<table>
<thead>
<tr>
<th>T</th>
<th>CH</th>
<th>B</th>
<th>CGM</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1: 07:00 Breakfast data</td>
<td>Hypo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1: 12:00 Lunch data</td>
<td>Non Hypo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 1: 18:00 Dinner data</td>
<td>Non Hypo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 2: 07:00 Breakfast data</td>
<td>Hypo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 2: 12:00 Lunch data</td>
<td>Non Hypo</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Day 2: 18:00 Dinner data</td>
<td>Non Hypo</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Meal data (breakfast/lunch/dinner)

Given the set of meal events, these are sorted according to the time attribute and processed to create sequences of meals where each sequence contains all the ordered meals of a time window of 6 hours.

The class of each sequence of meals is labelled according to the postprandial status (PS) of the last meal of the sequence, following two procedures:

1. **PS is labelled as hypoglycaemia if a CGM reading between 2 and 6 hours after the bolus administration is below 70 mg/dL, otherwise it is labelled as non-hypoglycaemia.**
2. **PS is labelled as hyperglycaemia if CGM readings are above 180 mg/dL during at least 60 minutes between 2 and 6 hours after the bolus administration, otherwise it is labelled as non-hyperglycaemia.**

These sequences are then used to predict the class of a given sequence of meals using KNN, where \( k < \sqrt{n} \) (n is the number of examples).

Results

- **TPR:** True Positive Rate
- **FNR:** False Negative Rate
- **FPR:** False Positive Rate

<table>
<thead>
<tr>
<th>TPR</th>
<th>FNR</th>
<th>FPR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.88</td>
<td>0.12</td>
<td>0.25</td>
</tr>
</tbody>
</table>

The class of each sequence of meals is labelled according to the postprandial status (PS) of the last meal of the sequence, following two procedures:

1. **PS is labelled as hypoglycaemia if a CGM reading between 2 and 6 hours after the bolus administration is below 70 mg/dL, otherwise it is labelled as non-hypoglycaemia.**
2. **PS is labelled as hyperglycaemia if CGM readings are above 180 mg/dL during at least 60 minutes between 2 and 6 hours after the bolus administration, otherwise it is labelled as non-hyperglycaemia.**

The use of sequential data clearly outperforms the one-shot data.

Conclusions

1. Recommendations based on meals increases the overall performance (TPRs up to 0.88 and accuracies up to 83% have been achieved, while FNR and FPR are significantly lower, from 0.25 to 0.11). Furthermore, the use of sequential data clearly outperforms the one-shot data.
2. Recommendations should be based on type of meal (do not merge all the meals together).
3. Personalized databases with less amount of registers, does not favour nor the use of sequences, neither the distinction of different types of meals.

Forthcoming Research

- Eager mechanisms or its hybridization for sequence learning
- Fuzzy approach in the labelling

Acknowledgements

This work has received funding from the EU Horizon 2020 research and innovation programme under grant agreement No 689810 (PEPPER), and from the University of Girona under the grant MPCUdG2016 (Ajut per a la millora de la productivitat científica dels grups de recerca), and the Spanish MINECO under the grant number DPI2013-47450-C21-R. This work has been developed with the support of the research group SITES awarded with distinction by the Generalitat de Catalunya (SGR 2014-2016).