Case base maintenance of a personalized insulin dose recommender system for Type 1 Diabetes Mellitus

Ferran Torrent, Joaquim Massana, and Beatriz López

Universitat de Girona, Politechnical School, Building EPS-4,
Campus Montilivi, Girona
ferran.torrent,joaquim.massana,beatriz.lopez@udg.edu
http://exit.udg.edu

Abstract. With the goal of supporting people suffering Type 1 Diabetes Mellitus (T1DM), some mobile applications are being developed based on artificial intelligence techniques. Some of these applications are based on Case-Based Reasoning methodologies (CBR) due to the advantage regarding a personal, adapted recommendation. However, the amount and quality of the cases in the CBR system will threat the system outcome. Most of the maintenance methods developed deals with classification tasks, while recommending an insulin dose (bolus) involves a regression task. In this paper, a new maintenance method presented, with the particularity of dealing with a regression tasks. The method is applied over the Pepper insulin dose recommender system, and tested using the UVA/Padova simulator, exhibiting the improvements of the proposal in terms of both, the person health and the case-base size.

1 Introduction.

The use of systems to continuously monitor blood glucose levels and the increased computation power of mobile devices such as smart-phones has permitted the use of artificial intelligence to aid people with T1DM. Mobile applications have the advantage of enabling connectivity with sensors able to measure glucose level as well as other physical conditions of the patient (e.g. physical activity), and actuators, and insulin pumps which inject insulin to the persons body.

In this context, CBR [9] has been proved as a useful methodology to develop adaptive and personalized insulin recommender systems [17] because it is capable to optimize insulin dosage using past experiences.

The performance of a CBR system depends on the experiences or cases stored in a case-based, so that they are key in providing recommendations. Moreover, in an insulin recommender system, the contents of the case-base should be maintained in order to capture the persons physiological changes evolution, and provide adapted recommendations over time. Insulin recommender systems based on CBR use to propose a simple technique to maintain the quality of the case base, while relying on a small set of attributes which are selected according to a given procedure, so case bases are kept small and efficient [12, 13, 17]. However, the metabolism of carbohydrates, glucose and insulin depends on many factors (e.g. stress, physical activity, etc.) which such simplistic approach could hide. Therefore, a case base editing technique is needed in order to keep the case base optimized to be efficient and dynamic to follow possible changes in the users physiology (concept drift [18]).

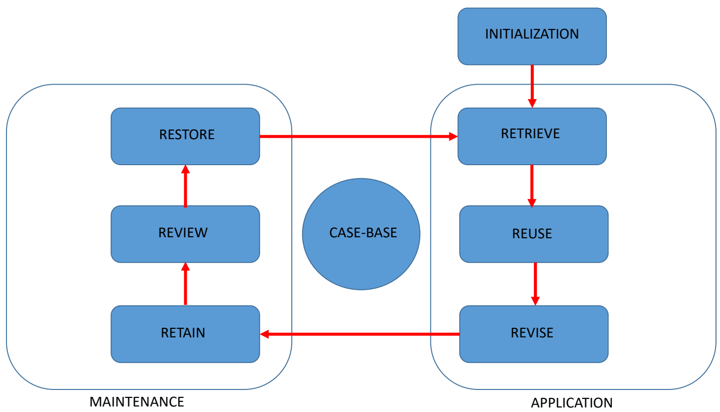

Fig. 1. CBR cycle extended with a maintenance phase.

This paper proposes a maintenance methodology for an insulin dose recommender system. The maintenance method consists of three steps: retain, review and restore, as it is proposed in [5] and depicted in Fig. 1. The particularity of the insulin recommendation system, which recommends a numeric value instead of a class, invites us to define new review-restore approaches to be used in a regression system instead of a classification system, which is the methodology found in the literature [10]. In addition, the maintenance method developed is proven to be capable of dealing with the concept drift.

2 Background.

The understanding of this paper requires some basics from case-based reasoning system and the Pepper recommender system from which the maintenance methodology is proposed.

2.1 Case Based Reasoning (CBR).

CBR is a lazy learning technique that uses past experiences in order to search a solution for a new problem. The basic CBR methodology was described in [1] in 4 steps, as shown in Fig. 2. First, given a query or problem situation or case, the retrieve step searches similar past experiences from the case base. Second, the reuse step proposes past solutions from the retrieved cases to the query case. Third, the revise step, the implemented solution is revised according to the outcome. In the last step, the retain one, the system decides about to store the query case or not according to some strategy, using several metrics. In so doing, more recent approaches as [5] proposes to unfold the CBR methodology into maintenance and application, moving the retain step into the maintenance part, and complementing it with a review and restore steps, as depicted in 1. This new approach highlights the importance of the quality of the cases and the size of the base of case regarding the CBR system performance.

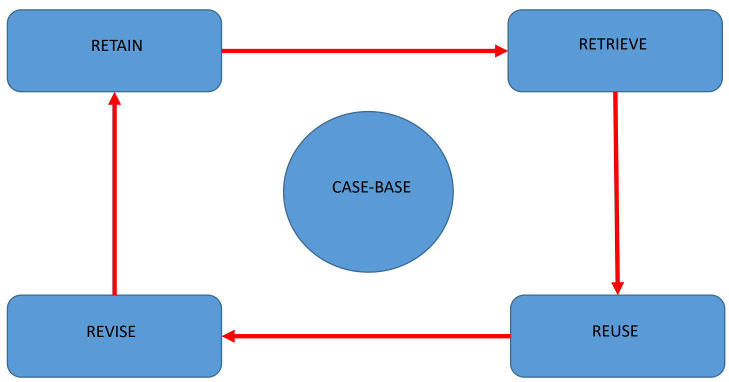

Fig. 2. Case Based Reasoning methodology.

2.2 Pepper insulin recommender.

The work presented in this paper is based on the Pepper recommender system proposed in [17], which follows a CBR methodology. The system takes advantage of mobile technology to gather information about different sensors which sends information to the mobile thorough blue-tooth low energy (physical activity band and continuous glucometer).
The goal of the system is to provide as insulin bolus for a T1DM person when is going to enjoy a food intake. Cases in the system consist of the attributes that characterise and contextualize the ingest (case description) and the solution. Attributes are the following: time of the day, amount of carbohydrates of the meal (g), post physical activity (quantified in a four-level graded scale) and future estimated physical activity (in the same scale). The solution consists of the ICR (insulin to carbohydrates ration) that enables the computation of the insulin dose (bolus).
The recommender system is linked to a continuous cyclometer which enable revising the case automatically. That is, the system analyses if the

postprandial blood glucose agrees some clinical constraints. If the bolus is not correct, the same postprandial measurements enable the system to calculate of the correct bolus according to the clinical knowledge as reported in [8,17]. Then, the system decides if the new case is introduced in the case base according to a "most recent" strategy: if two case description are equal, but not their solution, the most recent case is stored and the old one removed.
No other maintenance method was developed, and this paper focus on the full maintenance procedure of such system, including the review and restore stages.

3 Related work.

The authors in [10] propose a case base editing (or maintenance) methodology to keep the case base of a CBR binary classifier as small as possible but also capable to follow a concept drift. The methodology consists of applying a case base reduction technique, named Conservative Redundancy Reduction (CRR), but also, monitoring class instances variability to detect changes in the majority class in relevant regions of the solution space. Then, when a change is detected, old instances are removed (forgotten) because it is assumed that there is a concept drift.
The Instance-Based learning Algorithm3 (IB3) [2] was one of the first attempts to handle the concept drift monitoring of the cases accuracy and the retrieval frequency. In [14], the Locally Weighted Forgetting (LWF) algorithm was proposed to reduce the weights of the k-nearest neighbours (k-NN) of a new case, so a case is discarded if its weight falls below a threshold . In [15], Salganicoff achieved suitable results in time-varying and static tasks with a method called Prediction Error Context Switching (PECS). With the aim of controlling in an autonomous manner the size and the composition of the case-base, Beringer and Hllermeier [3] presented an Instance-Based Learning on Data Streams (IBL-DS) algorithm. In [7] Delany et al. Proposed a two-level learning technique with the aim of solving the concept drift issues.
Despite this research, there is still need of maintenance algorithms capable to deal with numerical solutions.

4 Pepper recommender maintenance.

The maintenance approach proposed in this paper consists of the three methods: recent strategy, time analysis and coverage and reachability analysis, which are applied in the retain, review and restore steps of the CBR methodology as described below.

4.1 Retain: Recent strategy.

The retain step is the same methodology of [17], which consists of before storing a new case, check if there is an equal case in the case base regarding the problem description, but that it has a different outcome.

That could happen because the information gathered in a case (time, carbohydrates and physical activity) is not complete regarding all possible conditions a patient could have (stress, body weight changes, etc.). So different solutions could arise from a the same problem description. Note, that extending the problem description has been studied by the authors, as well as in the literature, but the key factors as well as the complexity of the usability of the system developed has lead to follow a simple, but effective approach.
When two cases have identical description, the old case is removed and the new one is stored. Otherwise, the new case is stored but none is removed.
Note that this approach assumes that a pair of cases with equal attributes but different ICR is due to a concept drift.

4.2 Review and Restore: Time analysis.

The combination of values for the attributes of a case is large, and the "most recent" strategy is limited to the condition of finding identical case description; therefore, the capacity of following the concept drift (or physiological changes of a person) by this strategy is low.
Therefore, additional methods should be developed, as the ones proposed by [10]. However, such methods are based on a classification task, while the Pepper recommender system is dealing with a regression task. Therefore, solutions like [10] are not feasible.
Thus, we propose to statistically analyse the solutions of the case base throughout time, and if there are significant ICR differences in the timeline, we assume that there is a concept drift. Specifically, we sort the cases according their time-stamp (time when they were stored), and the mean and standard deviation of the ICR of the cases is calculated using a sliding window. When, significant differences are detected respect the first batch, the older cases are removed.

4.3 Review and Restore: Coverage and reachability analysis.

Redundancy of cases could happen because there are very similar cases in the case base. Traditional case base reduction techniques apply the review-restore methodology proposed in [5]. First, they review the case base and calculate metrics such as the coverage and the reachability of the cases that describe how useful are the cases. The coverage, is the set of target problems that it can be used to solve and the reachability, is the set of cases that can be used to provide a solution for the target. Second, the restore the case base, or delete those redundant cases in order to keep the minimum number of cases. However, these methods take advantage of a low number of classes and cannot be directly applied to regression tasks.
Therefore, we propose a method to decide whether an ICR of a particular case is similar enough to the ICR of another one. In particular, we propose to convert the minimum accuracy of insulin infusion to a maximum

allowed difference between ICRs to be considered equal. Thus, if the allowed error in the insulin infusion system is a half bolus then this value is converted to a maximum difference between ICRs, i.e. it is converted in the maximum error in terms of ICR that will derive in an insulin error equal to or lower than a half dose.
Once this difference between ICRs is calculated, reachability and coverage of each case are calculated. Next, given the reachability and the coverage of the cases, restore techniques such as CRR [6] or Iterative Case Filtering (ICF) [4] are applied. The CRR algorithm uses the coverage to remove the redundant cases. The ICR algorithm removes cases which have a smaller coverage than the reachability.

5 Results and discussion.

The experimentation has been carried out using the information of 11 adult virtual subjects on the UVA/Padova simulator [11]. Simulations have been carried out for 180 days and 20 repetitions. The experiments used a learning weights technique based on Salzberg [16]. The used glycaemic target range is [70,180] mg/dl for all the virtual subjects.
The following experimental scenarios are analysed:

- No maintenance.
- ICF.
- CRR.

Results are analysed first, in terms of the health implications of the maintenance operations, and second in terms of case-base size.
In terms of health, results are provided according to the time in target, that is, the time the person glucose level is between the minim and maximum levels prescribed. Fig. 3 shows how the proposed method improves the results regarding the absence of a maintenance step.

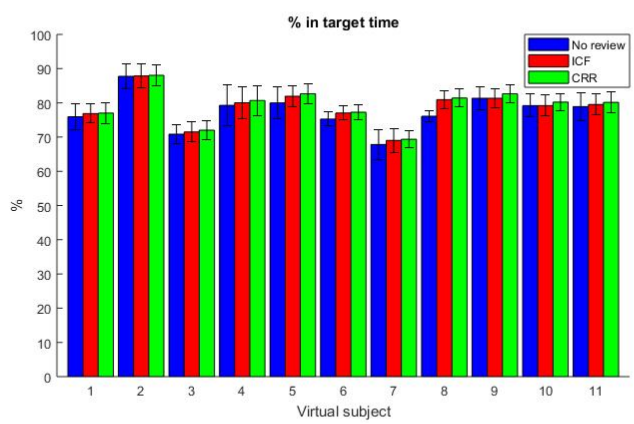

Fig. 3. In target time.

On the other hand, the health implications of a person being above or below the target are different. Fig. 4 shows the time above the target, while means the risk of suffering a hyperglycaemia event. The lower the better. As it can be observed in the Figure, the proposed method exhibits the better performance.

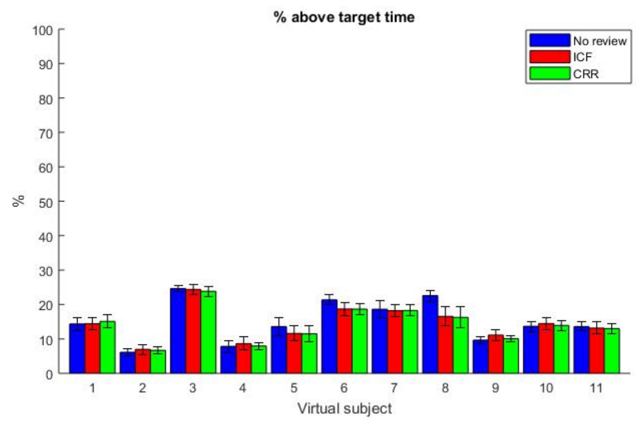

Fig. 4. Above target time.

Fig. 5 shows the time below the target, while means the risk of suffering a hypoglycaemia event. The lower, the better. As in the previous case, the proposed method exhibits the better performance.

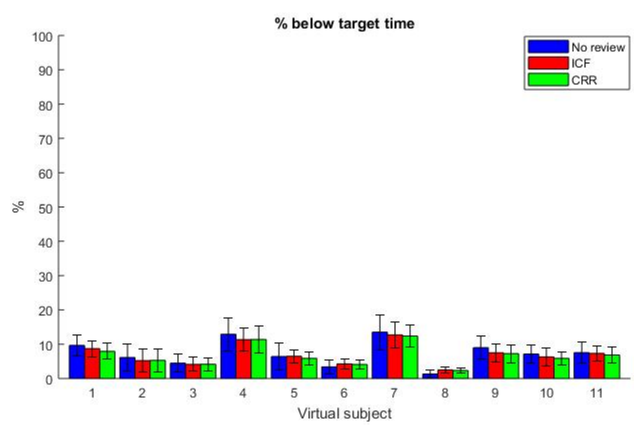

Fig. 5. Below target time.

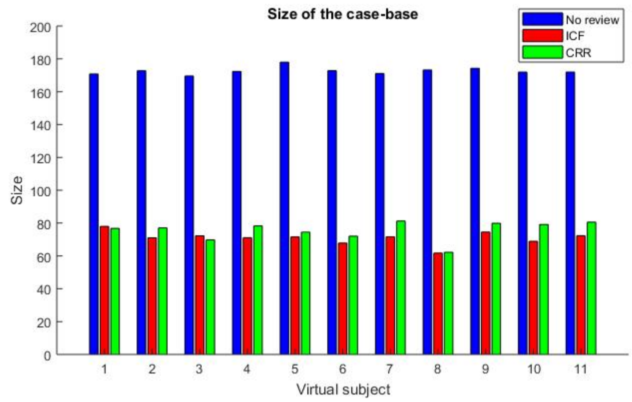

Fig. 6. Size of the case-base.

Finally, regarding the case-base size, Fig. 6 shows that there is a clear reduction in the size of the case-base using the proposed maintenance system.

6 Conclusions.

In order to deal with T1DM, mobile applications are being used. In particular, this paper deals with an insulin dose recommender system based on CBR. The CBR efficiency relies on the size and quality of the case-base. A new methodology of maintenance, that takes into consideration the physiological evolution of the person suffering T1DM (concept drift), is proposed in this paper. The results show clear improvements in case-base size and, what is more interesting, the health outcomes of the maintenance operations, measure by the time in target, that is the time in which the person is between the blood glucose level prescribed.

Acknowledgments.

This research project has been partially funded through BR-UdG Scholarship of the University of Girona granted to Joaquim Massana Raurich. Work developed with the support of the research group SITES awarded with distinction by the Generalitat de Catalunya (SGR 2014-2016), the MESC project funded by the Spanish MINECO (Ref. DPI2013-47450-C2-1-R) and the European Unions Horizon 2020 research and innovation programme under grant agreement No 680708.

References

1. Agnar Aamodt and Enric Plaza. Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Commun.*, 7(1):39–59, 1994.

2. David W Aha, Dennis Kibler, and Marc K Albert. Instance-based learning algorithms. *Machine learning*, 6(1):37–66, 1991.
3. Jürgen Beringer and Eyke Hüllermeier. Efficient instance-based learning on data streams. *Intelligent Data Analysis*, 11(6):627–650, 2007.
4. Henry Brighton and Chris Mellish. Advances in instance selection for instance-based learning algorithms. *Data mining and knowledge discovery*, 6(2):153–172, 2002.
5. Ramon Lopez De Mantaras, David McSherry, Derek Bridge, David Leake, Barry Smyth, Susan Craw, Boi Faltings, Mary Lou Maher, Michael T Cox, Kenneth Forbus, and Others. Retrieval, reuse, revision and retention in case-based reasoning. *The Knowledge Engineering Review*, 20(3):215–240, 2005.
6. Sarah Jane Delany and Pádraig Cunningham. An analysis of case-base editing in a spam filtering system. In *European Conference on Case-Based Reasoning*, pages 128–141. Springer, 2004.
7. Sarah Jane Delany, Pádraig Cunningham, Alexey Tsymbal, and Lorcan Coyle. A case-based technique for tracking concept drift in spam filtering. *Knowledge-based systems*, 18(4-5):187–195, 2005.
8. Pau Herrero, Peter Pesl, Jorge Bondia, Monika Reddy, Nick Oliver, Pantelis Georgiou, and Christofer Toumazou. Method for automatic adjustment of an insulin bolus calculator: In silico robustness evaluation under intra-day variability. *Computer Methods and Programs in Biomedicine*, 119(1):1–8, 2015.
9. Janet Kolodner. *Case-based reasoning*. Morgan Kaufmann, 2014.
10. Ning Lu, Jie Lu, Guangquan Zhang, and Ramon Lopez De Mantaras. A concept drift-tolerant case-base editing technique. *Artificial Intelligence*, 230:108–133, 2016.
11. Chiara Dalla Man, Francesco Micheletto, Dayu Lv, Marc Breton, Boris Kovatchev, and Claudio Cobelli. The UVA/PADOVA Type 1 Diabetes Simulator: New Features. *Journal of Diabetes Science and Technology*, 8(1):26–34, 2014.
12. Stewart Massie. *Complexity Modelling for Case Knowledge Maintenance in Case-Based Reasoning*. PhD thesis, Robert Gordon University, 2006.
13. Peter Pesl, Pau Herrero, Monika Reddy, Nick Oliver, Desmond G Johnston, Christofer Toumazou, and Pantelis Georgiou. Case-based reasoning for insulin Bolus Advice: evaluation of case parameters in a six-week pilot study. *Journal of diabetes science and technology*, 11(1):37–42, 2017.
14. Marcos Salganicoff. Density-adaptive learning and forgetting. In *Proceedings of the Tenth International Conference on Machine Learning*, volume 3, pages 276–283, 1993.
15. Marcos Salganicoff. Tolerating concept and sampling shift in lazy learning using prediction error context switching. In *Lazy learning*, pages 133–155. Springer, 1997.
16. Steven Salzberg. Distance metrics for instance-based learning. In *International Symposium on Methodologies for Intelligent Systems*, pages 399–408. Springer, 1991.

17. Ferran Torrent-Fontbona and Beatriz Lopez Ibanez. Personalised Adaptive CBR Bolus Recommender System for Type 1 Diabetes. *IEEE Journal of Biomedical and Health Informatics*, 2018.
18. Gerhard Widmer and Miroslav Kubat. Learning in the presence of concept drift and hidden contexts. *Machine learning*, 23(1):69–101, 1996.