BioMedical Engineering OnLine

Differential sequential patterns supporting insulin therapy of new-onset type 1 diabetes

Rafał Deja[1*], Wojciech Froelich[2] and Grażyna Deja[3]

*Correspondence:
rdeja@wsb.edu.pl
[1] Department of Computer Science, Academy of Business in Dabrowa Gornicza, Cieplaka 1c, Dabrowa Gornicza, Poland
Full list of author information is available at the end of the article

Abstract

Background: In spite of numerous research efforts on supporting the therapy of diabetes mellitus, the subject still involves challenges and creates active interest among researchers. In this paper, a decision support tool is presented for setting insulin therapy in new-onset type 1 diabetes.

Methods: The concept of differential sequential patterns (DSPs) is introduced with the aim of representing deviations in the patient's blood glucose level (BGL) and the amount of insulin injections administered. The decision support tool is created using data mining algorithms for discovering sequential patterns.

Results: By using the DSPs, it is possible to support the physician's decisionmaking concerning changing the treatment (i.e., whether to increase or decrease the insulin dosage). The other contributions of the paper are an algorithm for generating DSPs and a new method for evaluating nocturnal glycaemia. The proposed qualitative evaluation of nocturnal glycaemia improves the generalization capabilities of the DSPs.

Conclusions: The usefulness of the proposed approach was evident in the results of experiments in which juvenile diabetic patients actual data were used. It was confirmed that the proposed DSPs can be used to guide the therapy of numerous juvenile patients with type 1 diabetes.

Keywords: Data mining, Medical patterns, Diabetes mellitus

Introduction

In recent years, the number of cases of diabetes mellitus has increased significantly [1]. Therefore, controlling the concentration of glucose in the patient's blood is still an important problem that is being investigated by many researchers. The problem is a challenge since the strict medical procedure for controlling BGL cannot be defined [2]. The main reason is the high variability between people and the numerous different initial clinical states of patients after they are admitted to the hospital. This especially concerns the onset of diabetes, when the patient's state is unstable. For that reason, the existing medical approach to determining the appropriate therapy is based almost exclusively on the physician's experience [3]. Researches has been focused mainly on the so called issue of 'closing the loop' [4]. The idea is to create a system that can simulate the pancreas by automatic adjustment of insulin injections or, in some cases, by delivering glucagon to achieve the appropriate glucose level. The problem often requires building a decision-support system based on the prediction of the level of blood glucose.

In this paper, we propose a decision support tool for setting basal insulin therapy at the onset of type 1 diabetes. The proposed method is based on the concept of sequential patterns [5]. After determining the sequential patterns that underlie a given sequence of terms, it is possible to predict the plausible continuation of the sequence. From the medical perspective, sequential patterns can be treated as medical guidelines [3,6,7]. The advantage of using medical guidelines based on sequential patterns is that they are easy for physicians to interpret [8].

The concept of sequential patterns has been used extensively in different application domains [9], especially in medicine [10-12]. Sequential patterns have been used to identify diagnoses that diverse hospitals have in common [13] or to determine the subsets of examinations that were done together and that were followed by patients [14]. Several algorithms to discover sequential patterns have been proposed [5,15]. The concept of mining diabetes data using sequential patterns first was proposed in [16]. The patterns that were discovered were used to construct a decision tree that expressed the possible flow of medical events. Recently, [17] proposed the use of sequential patterns for planning diabetic therapy. The template-based patterns discovered from historical data were used to recommend the insulin dosage on the basis of the known value of the level of blood glucose [17].

One of the limitations of the approach proposed in [17] is the low generalization capability of the patterns that were discovered over the population of all patients. For that reason, the patterns were mined and used separately for each individual patient. In this paper, we describe our effort to extend the application of sequential patterns for supporting diabetic therapy by increasing their generalization capabilities. It is necessary to mention that such generalization is quite difficult for pre-meal insulin dosage, because each patient starts a meal with different BGL, and the measurement of the energy value of the meal is highly imprecise. For that reason, we left that challenge for future researches and focused our investigation in this research on mining patterns concerning only long-lasting basal insulin.

The contributions of this paper are as follows:

- A new approach for the qualitative evaluation of nocturnal glycaemia is introduced.
- The concept of differential sequential patterns and an algorithm for mining them are proposed.
- A decision support tool is developed based on differential sequential patterns. The decision tool's output serves as a medical guideline for physicians when, the proper course of therapy must be determined quickly, e.g., for patients admitted to the hospital - new-onset diabetes.

The remainder of the paper is organized in the following way. We start with the Medical background section where the therapy of diabetes is presented. The basics of sequential patterns are given in Methods section, where later on the contribution of this paper is presented. First, the procedure of qualitative evaluation of nocturnal glycaemia is formalized. Second, the concept of differential sequences is introduced, and an algorithm for the generation of sequential patterns based on differential sequences is proposed. The validation of the proposed approach and the discussion of the results that were obtained are described in Results of experiments section.

Medical background

The main goal of effective diabetes treatment is to maintain the state of near normo-glyceamia. It is said that the normal glucose level should range between 70-140 mg/dl throughout the day, although the recommended range before a meal is 70-100 mg/dl [3]. Therefore repeated measurements of blood glucose during the day are performed [6]. Typically, the onset of type 1 diabetes among children is rapid, and such patients should be hospitalized. The initial daily insulin dosage is prescribed based on several basic data (e.g. the patient's weight, gender and age), some test results (initial glicaemia level, acid-base balance, CRP, C-peptyd and HbA1c) and the proposed diet (patient's energy demand). Currently, there is no strict medical procedure that allows immediately setting the proper insulin dosage based on the data listed above. Thus, in the following days, the daily glucose profile is taken, and the insulin dosage is adaptively modified. The daily glucose profile consists of glicaemia measurements every 2-3 hours during the night, fasting, before each meal, and 2 hours after each meal. The patient is released from the hospital when treatment is established, so that during the day, the blood glucose level is close to normal.

The overall daily insulin dosage is divided with 30% of the dosage administered as the basal dosage and 70% administered as the pre-meal insulin dosage. (Every patient with type 1 diabetes mellitus should take two types of insulin.) For the basal insulin the long-lasting (e.g., glargina insulin) is used. It is a biosynthetic recombinant analog insulin, that provides 24 hours of continuous activity, becomes fully acting as soon as 2 hours after administration, and is peakless. The basal insulin is usually injected in the evening (one injection per day), and it ensures the correct level of glucose during the night and before meals. In the following days, the results of current measurements of glucose are used to determine the dosage. The dose for the next day can be reduced when hypoglycemia occurred in the previous day, or it can be increased id hyperglycemia is observed in the glucose profile. In the case of increasing the glucose level (more than 200 mg/dl) during the night, the short acting insulin is administered immediately as the so-called 'basal correction.'

The second part of the therapy involves administering pre-meal insulin which, is short acting. This insulin is served just before each meal - usually 4-5 times a day.

Below, we provide example data concerning dosages of basal insulin and measurements of the levels of glucose in the blood. These data were collected during two exemplary days of a patient's stay in the pediatric, endocrinology, and diabetology clinical hospital in Katowice, Poland. The data (Table 1) consist of the basal insulin dosage administered at 10 P.M. and four measurements of blood glucose level during the night (at 0:00, 3:00, 5:00, and 7:00 A.M.).

The nocturnal blood glucose level was interpreted in accordance with accepted medical standards [3] in the following way: when BGL is less than 70 mg/dl, it is treated as below normal; the normal level is between 70 and 140 mg/dl; the higher level (called mild-hyperglycemia) is recognized in the range of 140-200 mg/dl; and hyperglycemia is observed above 200 mg/dl. Thus, the corresponding symbolic values for the discretized glycaemia attribute are presented in Table 2.

When planning therapy, the physician considers the patient's clinical status and some test results [3], but the most important factor when the insulin dosage is initially specified is the patient's weight. We used this indication such that the given basal insulin dosage is

Table 1 Exemplary clinical data

Patient	Date	Time[h]	Value	Description
425	2008-11-26	22	10	Basal insulin
425	2008-11-27	0	109	Glycaemia
425	2008-11-27	3	62	Glycaemia
425	2008-11-27	5	60	Glycaemia
425	2008-11-27	7	71	Glycaemia
425	2008-11-27	22	9	Basal insulin
425	2008-11-28	0	114	Glycaemia
425	2008-11-28	3	112	Glycaemia
425	2008-11-28	5	105	Glycaemia
425	2008-11-28	7	99	Glycaemia
425	2008-11-29	22	8	Basal insulin

recalculated into units per 10 kg of the patient's weight. For example, when considering the first day of therapy for patient 425, the basal insulin dosage is 10 units, and the patient's weight is 32 kg. Thus, in our research the considered/normalized basal insulin dosage was rounded to 3 units per 10 kg.

Methods

In this section, step-by-step we describe the way the decision support tool for therapy setting was created. We start with preliminaries about the notion of sequential patterns.

The basics of sequential patterns

Let $E = \{e_1, e_2, \ldots, e_n\}$ denote the set of items (also called elements or events). Any subset of items $a_i \subseteq E$ is called an itemset. The sequence $a = \langle a_1, a_2, \ldots, a_k \rangle$ is defined as an ordered list of itemsets a_i. The term k-sequence denotes any sequence of cardinality $k = card(a)$, where k is the number of itemsets it contains. The particular item $e_j \in E$ may recur in the k-sequence by being included in several itemsets a_i.

The sequence $a = \langle a_1, a_2, \ldots, a_n \rangle$ is contained [5] in the other sequence $b = \langle b_1, b_2, \ldots, b_m \rangle$, i.e. $a \subseteq b$ if there exist integer numbers $i_1, i_2, \ldots, i_n$ such that $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \ldots, a_n \subseteq b_{i_n}$ [5]. The sequence a is called a subsequence of b, and the sequence b is called a supersequence of a.

Suppose the collection of sequences $S = \{s^1, s^2, \ldots, s^m\}$ is available, where $s^j = \langle s_1^j, s_2^j, \ldots, s_n^j \rangle$ denotes the jth sequence consisting of s_i^j itemsets. The support (1) of sequence a is a fraction of sequences in S that contain a.

$$sup(a) = \frac{card\left(s^j \in S | a \subseteq s^j\right)}{card(S)} \tag{1}$$

Table 2 Glycaemia discretisation

Blood glucose level [mg/dl]	Interpretation	Discrete value
<70	Hypoglycemia	1
[70, 140)	Normoglycemia	2
[140, 200]	Mild-hyperglycemia	3
>200	Hyperglycemia	4

The objective is to determine such subsequences $a \subseteq s^j$ that are frequently contained in any sequences from S, i.e., with the support $sup(a) \geq sup_{min}$, where sup_{min} is a threshold given by experts. Such sequences are in fact a kind of patterns in data, therefore they are called frequent patterns.

Several algorithms for mining frequent patterns have been proposed, the first and best known among them being AprioriAll and AprioriSome [5]. The goal of the other algorithms was to decrease the computational cost of mining [18-20]. An overview of sequential patterns mining algorithms is available in [21].

Qualitative estimation of nocturnal glycaemia

In the case of the hospital for diabetic children that was considered, the measurements of BGL are made at 0:00, 3:00, 5:00 and 7:00 A.M. After numerous experiments, it became clear that the application of the arithmetic mean of four numerical measurements [17] was not fully correct and led to decreased support for the patterns that were discovered from the sequences stored as historical data. For that reason, we proposed a new approach.

In fact, qualitative evaluation of nocturnal glycaemia is usually done by doctors to estimate the level of glycamia during the night, and, further, it is used to determine the basal insulin dosage that is administered on the same day at 22:00 P.M. Formally, let us denote the sequence of BGL measurements at 0:00, 3:00, 5:00 and 7:00 as $bgl = \langle bgl_0, bgl_3, bgl_5, bgl_7 \rangle$, where bgl_k is a discrete value of the blood glucose level measured at time $k \in \{0:00, 3:00, 5:00, 7:00\}$. Using that sequence, it is proposed that the level of nocturnal glycaemia be estimated by the following calculation:

$$d = \begin{cases} normoglycemia & \Leftrightarrow \forall_k (bgl_k = normoglycemia) \\ hypoglycemia & \Leftrightarrow \exists_k (bgl_k = hypoglycemia) \\ hyperglycemia & \Leftrightarrow \exists_k (bgl_k = hyperglycemia) \wedge \forall_k (bgl_k \neq hypoglycemia) \\ mild-hyperglycemia & otherwise \end{cases}$$

Table 3 illustrates the exemplary data given previously in Table 1 after the qualitative evaluation of nocturnal glycaemia. The qualitative evaluation is made in the column 'Evaluation', without any changes in the numerical values of the insulin dosage.

Differential sequential patterns

Let us define an ordered pair $s_i = \langle z_i, c_i \rangle$, where z_i denotes the value of basal insulin dosage, and c_i is the value of nocturnal BGL. In this way every sequence

Table 3 Exemplary clinical data after the qualitative evaluation of nocturnal glycaemia

Patient	Date	Time	Evaluation	Description
425	2008-11-26	22	3	Basal insulin
425	2008-11-27	Night	Hypoglycemia	Glycaemia
425	2008-11-27	22	3	Basal insulin
425	2008-11-28	Night	Normoglycemia	Glycaemia
425	2008-11-28	22	3	Basal insulin

$s^j = \langle s_1, s_2, \ldots, s_n \rangle$ is related to the basal-insulin therapy of the jth patient. Every itemset within such sequence is, in fact, an ordered pair of events related to a single day of therapy. The goal of the therapy is to discover frequent patterns $p \subset s^j, j \in [1, n]$ that with high probability will recur for a newly-admitted patient, i.e., $p \subset s^{n+1}$.

As stated in Medical background, the basal-insuline therapy is often considered by physicians in terms of deviations of glycaemia and corresponding insulin dosages. To reflect this procedure, let us assume that the insulin dosage and the glycaemia level are denoted as a and b, respectively, at the day of patient's admission; in this way, $s_1 = \langle a, b \rangle$, where a and b are constants. The following insulin dosage (the next day) can be raised (denoted as '+'), reduced (denoted as '-'), or it can be kept at same level (denoted as '0'). A similar operation is performed with the glycaemia level. For $i \geq 2$ we have:

$$
d_i = \begin{cases} - & , \text{for: } z_i < z_{i-1} \\ 0 & , \text{for: } z_i = z_{i-1} \\ + & , \text{for: } z_i > z_{i-1} \end{cases}
$$

$$
g_i = \begin{cases} - & , \text{for: } c_i < c_{i-1} \\ 0 & , \text{for: } c_i = c_{i-1} \\ + & , \text{for: } c_i > c_{i-1} \end{cases}
$$

To represent the therapy of a patient, we redefine a differential sequence as $S = \langle s_1, s_2, \ldots, s_n \rangle$, where $s_i = \langle d_i, g_i \rangle$. For example, the sequence $S = \langle \langle 3, 3 \rangle, \langle 3, 2 \rangle, \langle 3, 1 \rangle, \langle 2, 2 \rangle \rangle$ is converted to the differential sequence $D = \langle \langle a, b \rangle, \langle 0, - \rangle, \langle 0, - \rangle, \langle -, + \rangle \rangle$ or shorter when omitting brackets $D = \langle a, b, d0, g-, d0, g-, d-, g+ \rangle$ - note that d stands for insulin dosage and g for glycaemia.

The standard support measure given by formula (1) is used for the evaluation of DSP.

Discovering DSP from historical data

In this section an algorithm for mining DSP is proposed. The goal of the mining algorithm is to determine the set of differential patterns with the given support level. The input data for the algorithm are the set of differential sequences S prepared by pre-processing the clinical data gathered and interpreted according to description given in Qualitative estimation of nocturnal glycaemia.

First, the initial pattern candidate a_k is initiated with the itemset $< d_0, g_0 >$ from the first sequence from the set S. The support of the candidate pattern for each sequence in the set S is evaluated. Then, the pattern is enhanced with the following itemset $< d_j, g_j >$ of the considered sequence, creating a new candidate pattern. The support of the new candidate pattern is evaluated, and the algorithm repeats. The loop is ended if the support of the candidate pattern is less than the given level $psup_{min}$ or if there are no more itemsets in the sequence. The patterns with the support above $psup_{min}$ and having the required length (here ≥ 6) are stored. The steps described above are repeated for each unique sequence from S.

Algorithm 1: Discovery of differential sequential patterns

input : $S = \{s^1, s^2, \ldots, s^n\}$ the set of patients' source differential sequences, $psup_{min}$ -
      the acceptable support level

output: $A = \{a_1, a_2, \ldots, a_m\}$ the set of patterns

$A \leftarrow \emptyset$;

foreach $s^i \in S$ **do**

    $a_k \leftarrow \emptyset$;

    $count(a_k) \leftarrow 0$;

    foreach $< d_j, g_j > \in s^i$ **do**

        $a_k \leftarrow a_k + < d_j, g_j >$;

        if $a_k \notin A$ **then**

            foreach $s^l \in S$ **do**

                if $a_k \subseteq s^l$ **then**

                    $count(a_k) \leftarrow count(a_k) + 1$;

                end

            end

            if $count(a_k)/n > psup_{min}$ *and* $card(a_k) \geq 6$ **then**

                $A \leftarrow A + a_k$;

            end

            else

                break;

            end

        end

    end

end

The algorithm presented above (Algorithm 1) is based on the AprioriAll [5] sequential patterns discovery algorithm. There are many different variations of this algorithm implementation for different purposes with the task to reduce the time and memory computation complexity [22-24]. The algorithm above has been implemented by authors with worst time complexity not greater than $O(n \cdot k^2)$ where n is the number of sequences and k is the average length of the sequences.

Decision support

The main application of the mined DSPs is to support setting the therapy for a newly-admitted diabetic patient. The main idea of that support is to help the physician with the decision of changing the current treatment (increase or decrease the insulin dosage). That decision can be made with greater certainty by considering the knowledge gathered from similar cases presented in the form of differential patterns. In this way, the proposed supporting tool relies on the set of the mined differential patterns. The following steps lead the physician in determining the appropriate treatment for the patient:

1.   The patient is admitted to the hospital cared for by the physician during a certain initial period (1-3 days)

2.   The clinical data of the patient gathered during that initial period are converted into the form of differential sequence a^j (as described in this Section).

3.   The computer program retrieves a subset of the patterns $S \subseteq P$ from the set of all available DSPs such that a^j is supported by each sequence from S, i.e., $S = \{s : a^j \subseteq s \land s \in P\}$. This set is called a set of treatment support patterns (TSP).

4.   The selected patterns serve as the medical guidelines for the physician, supporting her or his decision regarding the following therapy.

After the patient's hospitalization ends, all of the gathered data are added in the form of differential sequences to the set of all sequences and the algorithm for mining DSPs is run again.

Results of experiments

The objective of the following experiments was to validate the theoretical approach proposed in this paper. We used the data of 102 children with onset of type 1 diabetes collected by the Department of Pediatrics, Endocrinology and Diabetology at the Silesian Medical University in Katowice, Poland. First, the data were processed, i.e., the discretization of BGL and insulin values was performed as described in Medical background. Afterwards the qualitative evaluation of nocturnal glycaemia was made.

The goal of the first experiment was to mine DSPs from the entire available dataset. By measuring the support, it was possible to check the generalization capabilities of the proposed DSP-based model over the set of all patients. The results are given in Table 4.

Table 4 shows that the support of individual DSPs reached quite high values. After adding the support of the first three patterns that refer to alternative versions of therapy for the initial three steps, the cumulative support $(0.39 + 0.24 + 0.23) = 0.86$ was evaluated as high. This means that applying only three of the most general DSPs, it is possible to propose the therapy for 86% of the patients.

The objective of the second experiment was the evaluation of the usefulness of the mined DSPs for newly-admitted patients, and 5-fold cross validation with random sampling was applied. The set of patients' sequences was partitioned into five parts. In each trial, the selected part served as the testing set, and the rest of sequences constituted the learning set. Differential patterns were mined from the learning set, and, then, their support in the testing set was calculated. The results of the experiment are presented in Tables 5 and 6.

Table 4 The set of DSPs mined from the entire set of data

DSP	Supp
$\langle a, b, d0, g0, d0 \rangle$	0.39
$\langle a, b, d0, g-, d- \rangle$	0.24
$\langle a, b, d0, g+, d0 \rangle$	0.23
$\langle a, b, d0, g0, d0, g0 \rangle$	0.16
$\langle a, b, d0, g0, d0, g- \rangle$	0.15
$\langle a, b, d0, g0, d0, g0, d0 \rangle$	0.13
$\langle a, b, d0, g0, d0, g-, d0 \rangle$	0.13
$\langle a, b, d0, g-, d0, g0 \rangle$	0.13
$\langle a, b, d0, g+, d0, g0 \rangle$	0.13
$\langle a, b, d0, g-, d-, g+ \rangle$	0.11
$\langle a, b, d0, g-, d0, g+ \rangle$	0.11

Table 5 Differential patterns, the results of 5-folded cross validation

1st trial	Supp	2nd trial	Supp
$\langle a, b, d0, g0, d0 \rangle$	0.4	$\langle a, b, d0, g0, d0 \rangle$	0.5
$\langle a, b, d0, g-, d- \rangle$	0.3	$\langle a, b, d0, g+, d0 \rangle$	0.3
$\langle a, b, d0, g0, d0, g0, d0 \rangle$	0.2	$\langle a, b, d0, g+, d0, g0 \rangle$	0.25
$\langle a, b, d0, g0, d0, g- \rangle$	0.2	$\langle a, b, d0, g-, d- \rangle$	0.2
$\langle a, b, d0, g0, d0, g-, d0 \rangle$	0.2	$\langle a, b, d0, g0, d0, g- \rangle$	0.15
$\langle a, b, d0, g0, d0, g0 \rangle$	0.2	$\langle a, b, d0, g-, d-, g0 \rangle$	0.15
$\langle a, b, d0, g+, d0 \rangle$	0.15		
$\langle a, b, d-, g-, d+ \rangle$	0.15		

Similarly, as in previous experiment, the support that was obtained was evaluated as high. Taking into account the patterns with a minimal support level greater or equal to 0.2 and with the length of 2.5 days or more, the non-empty set of treatment support patterns was deduced in 89% of test examples in each trial. The mined patterns can be used effectively to support the therapy of newly-admitted patients.

We draw the following medical observations regarding the results that were obtained:

- The correction of the initially-administered insulin dosage is often not needed, e.g., the DSP $= \langle d0, g0, d0, g0, d0 \rangle$ represents that fact.
- The physician is not changing the dose hastily just after one day of observation, e.g., DSP $= \langle g+, d0 \rangle$, $\langle g-, d0 \rangle$ is used, where g+,g- corresponds to the glycaemia level.
- The physician is trying to reduce the insulin dosage and maintain it at that level even if the body's response is not clear, e.g, that is represented by DSP $= \langle d0, g-, d- \rangle$.
- Considering the whole set of DSPs, it was apparent that the last dosage is usually the same as the previous dosage, meaning that the patient is leaving the hospital with the treatment settled and verified with the previous days.
- Another interesting observation is that the body's response can change even when the insulin dosage is not changed, see, e.g. DSP $= \langle d0, g-, d0, g+ \rangle$.

Conclusions

In this paper, the treatment of onset type 1 diabetes patients was analyzed. It was shown that the proposed differential sequential patterns for setting the basal insulin dosage can serve as guidelines for physicians during the day-after-day therapy. Also, by having these patterns, the physician has the opportunity to foresee the possible consequences of the prescribed therapy. Differential sequences can help the physician to decide the proper therapy faster. The usability of the proposed differential sequential patterns and the proposed mining algorithm was verified experimentally using real medical data. The results that were obtained provide evidence of the usefulness of the proposed approach.

Table 6 Differential sequential patterns mined from every learning trial

3rd trial	Supp	4th trial	Supp	5th trial	Supp
$\langle a, b, d0, g0, d0 \rangle$	0.35	$\langle a, b, d0, g0, d0 \rangle$	0.35	$\langle a, b, d0, g0, d0 \rangle$	0.4
$\langle a, b, d0, g+, d0 \rangle$	0.25	$\langle a, b, d0, g+, d0 \rangle$	0.15	$\langle a, b, d0, g+, d0 \rangle$	0.3
$\langle a, b, d0, g0, d0, g- \rangle$	0.15	$\langle a, b, d0, g-, d-, g+, d0 \rangle$	0.15	$\langle a, b, d0, g+, d0, g0 \rangle$	0.2
$\langle a, b, d0, g+, d0, g0 \rangle$	0.15	$\langle a, b, d0, g-, d-, g+ \rangle$	0.15	$\langle a, b, d0, g0, d0, g-, d0 \rangle$	0.15
$\langle a, b, d0, g+, d0, g0, d0 \rangle$	0.15	$\langle a, b, d0, g0, d0, g0 \rangle$	0.15	$\langle a, b, d0, g0, d0, g- \rangle$	0.15

Deja *et al. BioMedical Engineering OnLine* (2015) 14:13

Page 10 of 11

Author details
[1]Department of Computer Science, Academy of Business in Dabrowa Gornicza, Cieplaka 1c, Dabrowa Gornicza, Poland. [2]Institute of Computer Science, University of Silesia, Bedzinska 39, Sosnowiec, Poland. [3]School of Medicine in Katowice, Department of Pediatrics, Pediatric Endocrinology and Diabetes, Medykow 16, Katowice, Poland.

References
1. WHO. Fact Sheet No. 312. http://www.who.int/mediacentre/factsheets/fs312/en/.
2. Toussi M, Lamy J-B, Toumelin PL, Venot A. Using data mining techniques to explore physicians' therapeutic decisions when clinical guidelines do not provide recommendations: methods and example for type 2 diabetes. BMC Med Inform Decis Mak. 2009;9(1):28.
3. ADA. American diabetes association. Standards Med Care Diabetes-2012 Diabetes Care. 2012;35:11–63. doi:10.2337/dc12-s011.
4. Shalitin S, Phillip M. Closing the loop: combining insulin pumps and glucose sensors in children with type 1 diabetes mellitus. Pediatr Diabetes. 2006;7:45–9.
5. Agrawal R, Srikant R. Mining sequential patterns In: Yu PS, Chen ALP, editors. ICDE, Proceedings of the Eleventh International Conference on Data Engineering, March 6-10. Taipei, Taiwan: IEEE Computer Society; 1995. p. 3–14.
6. Bangstad HJ, Danne T, Deeb L, Jarosz-Chobot P, Urakami T, Hanas R. Ispad clinical practice consensus guidelines. insulin treatment in children and adolescents with diabetes. Pediatr Diabetes. 2009;12:92–9.
7. Rewers M, Pihoker C, Donaghue K, Hanas R, Swift P, Klingensmith G. Ispad clinical practice consensus guidelines. assessment and monitoring of glycemic control in children and adolescents with diabetes. Pediatr Diabetes. 2009;12:71–81.
8. Groot P. Experiences in Quality Checking Medical Guidelines using Formal Methods. In: Proceedings verification and validation of software systems (VVSS 2007). The Netherlands: Eindhoven: Technische Universiteit Eindhoven; 2007. p. 164–78.
9. Koper A, Nguyen H. Sequential pattern mining from stream data In: Tang J, King I, Chen L, Wang J, editors. Advanced Data Mining and Applications. Lecture Notes in Computer Science, vol. 7121. Berlin Heidelberg: Springer; 2011. p. 278–91. doi:10.1007/978-3-642-25856-5_21.
10. Yuliana OY, Rostianingsih S, Budhi GS. Discovering sequential disease patterns in medical databases using freespan mining approach. In: International conference on advance computer science and information system. Universitas Indonesia: Faculty of Computer Science; 2009.
11. Alonso F, López-Illescas Á, Martínez L, Montes C, Valente JP. Knowledge discovery using medical data mining. In: Medical data analysis. Springer; 2002. p. 1–12.
12. Kléma J, Nováková L, Karel F, Stepankova O, Zelezny F. Sequential data mining: A comparative case study in development of atherosclerosis risk factors. Syst Man Cybernet Part C: Appl Rev IEEE Trans. 2008;38(1):3–15.
13. Concaro S, Sacchi L, Bellazzi R. Temporal data mining methods for the analysis of the ahrq archives. Proc Am Med Inform Assoc 2007 Annu Symp. 2007.
14. Baralis E, Bruno G, Chiusano S, Domenici VC, Mahoto NA, Petrigni C. Analysis of medical pathways by means of frequent closed sequences. In: Setchi R, Jordanov I, Howlett RJ, Jain LC, editors. KES (3), Lecture notes in computer science. vol. 6278. Berlin Heidelberg: Springer; 2010. p. 418–25.
15. Srikant R, Agrawal R. Mining sequential patterns: Generalizations and performance improvements. In: Apers PMG, Bouzeghoub M, Gardarin G, editors. EDBT, Lecture notes in computer science. vol. 1057. Berlin Heidelberg: Springer; 1996. p. 3–17.
16. Rahaman SB, Shashi M. Sequential mining equips e-health with knowledge for managing diabetes. Int J Inf Process Manag. 2011;2(3):65–71.
17. Froelich W, Deja R, Deja G. Mining therapeutic patterns from clinical data for juvenile diabetes. Fundamenta Informaticae. 2013;127(1):513–28.
18. Zaki MJ. Efficient enumeration of frequent sequences. In: 7th Intl. conf. on information and knowledge management. New York, NY, USA: ACM; 1998. p. 68–75.

Deja *et al. BioMedical Engineering OnLine* (2015) 14:13

Page 11 of 11

19. Han J, Pei J, Mortazavi-Asl B, Chen Q, Dayal U, Hsu MC. FreeSpan: frequent pattern-projected sequential pattern mining. New York, NY, USA: ACM; 2000, pp. 355–59.

20. Cheng H, Yan X, Han J. IncSpan: incremental mining of sequential patterns in large database. In: KDD '04: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: ACM; 2004. p. 527–32.

21. Han J, Cheng H, Xin D, Yan X. Frequent pattern mining: current status and future directions. Data Min Knowl Discov. 2007;15(1):55–86.

22. Cavique L. A network algorithm to discover sequential patterns. In: Progress in artificial intelligence. Berlin Heidelberg: Springer; 2007. p. 406–14.

23. He D. Using suffix tree to discover complex repetitive patterns in dna sequences. In: Conf Proc IEEE Eng Med Biol Soc. vol. 1. IEEE; 2006. p. 3474–7.

24. Shakya S, Singh A, Singh D. A time efficient algorithm for web log analysis. Int J Comput Appl. 2013;75(9):23–30.