International Journal of Computer Theory and Engineering, Vol. 2, No. 2 April, 2010
1793-8201

Modeling Clinical Database using Time Series based Temporal Mining

Akash Rajak, and Kanak Saxena

***Abstract*—Clinical database contains huge amount of information about patients, medical tests and doctor's prescription. As the result of laboratory test, findings of physical examination and diagnosis are noted during time points or periods, by adding time dimension in clinical database we can retrieve vital information from the database. This paper presents the study of diabetes mellitus type 1 patients on the basis of time series based representation of clinical temporal data.**

***Index Terms*—temporal database, blood glucose level, insulin plasma level, type 1 diabetes mellitus, time series.**

I. Introduction

It is almost inconceivable to represent clinical data and reason about them without a temporal dimension. Time is important when time-oriented clinical data are considered as part of various decision support applications, such as determining a diagnosis, prescribing therapy and browsing electronic patient records for management or research purpose.

Experienced physicians are able to combine several significant contemporaneous findings, to abstract such findings into clinically meaningful higher-level concepts in a context-sensitive manner, and to detect significant trends in both low-level data and abstract concepts. Thus, it is desirable to provide short, informative, of time-oriented clinical data stored on electronic media, and to be able to answer queries about abstract concepts that summarize the data. Providing these abilities would benefit both a human physician and an automated decision-support tool that recommends therapeutic and diagnostic measures based on the patient's clinical history up to the present. Such concise, meaningful summaries, apart from their immediate value to a physician, would support the automated system's further recommendations for diagnostic or therapeutic interventions, provide a justification for the system's or for the human user's actions, and monitor plans suggested by the physician or by the decision-support system [1]-[5].

A. Diabetes mellitus Type 1

Diabetes mellitus is a metabolic disorder characterized by the decreased ability or completely inability of the tissues to utilize carbohydrates, accompanied by changes in the metabolism of fat, protein, water and electrolytes. The disorder is due to a deficiency or diminished effectiveness of the hormone insulin. This hormone is required to convert sugar, starches, and other food into energy. Diabetes mellitus is characterized by constant high levels of blood glucose (sugar) [6]. There are many types of diabetes: (i) Type 1 diabetes, (ii) Type 2 diabetes, (iii) Pre-diabetes, and (iv) Gestational diabetes. Type 1 diabetes is an autoimmune disease that results in destruction pegof insulin-producing beta cells of the pancreas. Lack of insulin causes an increase of fasting blood glucose (around 70-120 mg/dL in nondiabetic people) that begins to appear in the urine above the renal threshold (about 190-200 mg/dl in most people), thus connecting to the symptom by which the disease was identified in antiquity, sweet urine [7]. With Type 1 diabetes, your pancreas does not make insulin. Without insulin, too much glucose stays in your blood. Over time, high blood glucose can lead to serious problems with your heart, eyes, kidneys, nerves, and gums and teeth [8]. This paper simulates the effects of changes in insulin and diet on the blood glucose profiles of Type 1 diabetes mellitus patients. The results and observations shown are based on AIDA [9] simulator.

B. Structure of Paper

In section 2, we will discuss the functioning of AIDA simulator. In section 3, we focus on our proposed methodology. Section 4, provides a case study. Section 5 deals with limitations of simulator. Finally section 6, concludes the paper.

II. Simulator

AIDA is an interactive PC-based freeware computer program, which contains a simple model of glucose-insulin interaction in the human body. It is intended for simulating the effects on the blood glucose profile of changes in insulin and diet for a typical type 1diabetic patient. It also contains a very simple knowledge based system, which can identify problems in the displayed case. A list of suggestions, which might correct some of these problems, can also be generated. This list is provided solely as a prompt to the sort of insulin dosage adjustments that users might like to try simulating with AIDA. The system uses the concepts from mathematical techniques such as linear programming for calculating optimum insulin-dosage and is different from the way in which clinician's reason while trying to make insulin-dosage adjustments to improve a patient's glycaemic control [10]. The clinicians rather adopt step by step approach.

The simulator is initially provided with patient data regarding the carbohydrate intake, insulin dosage and blood glucose level. The simulator then projects the graph showing the variation in blood glucose level and plasma insulin level corresponding to carbohydrate intake and insulin dosage.

Manuscript received October 22, 2009. This work is supported by Uncertain Reasoning Framework for Data Mining by AICTE F. No.: 8023/BOR/RPS-97/2006-07.

Akash Rajak is with the Krishna Institute of Engineering and Technology, Ghaziabad, India (phone: +91-9873718407, e-mail: akashrajak@rediffmail.com, aakash@kiet.edu).

Kanak Saxena is with the Samrat Ashok Technological Institute, Vidisha, India (e-mail: ksv1909@yahoo.com).

The system also suggests alternative therapy plan for improving glycaemic control.

When simulated, two graphs will be shown on the graphical simulator display. The upper graph will show the "observed" blood glucose readings recorded via the data entry screen, while the lower graph will provide a composite display of information regarding insulin and carbohydrate intake. Superimposed on these graphs, as shown below, are the predicted steady state blood glucose and plasma insulin profiles as calculated by the AIDA model. The purpose of AIDA is to demonstrate the glycaemic effect of changes in either the dietary intake or insulin regimen–as either a self learning or demonstration exercise [10].

III. Time Series Based Methodology

Clinical data can contain temporal information, but it is often treated as static. In order to extract important information from temporal clinical data, we need to find an appropriate representation of our temporal sequences before applying mining techniques [11]. In our representation we apply time series for clinical data by using the concept of bands. The various steps are as follows:

A. Defining the Safety, Lower and Upper bands

The band ranges can be defined as:

1) Safety band

If the person is having blood glucose level between 72-180 mg/dL it would fall in safety band. A point (x_i, y_i) would fall in safety band if, $\alpha \le y_i \le \beta$, for $0 \le i \le n$. where α and β are the minimum and maximum blood glucose values. In our case, we are assuming value for α is 72 mg/dL and for β is 180 mg/dL.

2) Lower band

A severely low blood sugar level may lead to unconsciousness. A point (x_i, y_i) would fall in lower band if, $y_i < \alpha$, for $0 \le i \le n$. The simulator can be used to simulate the situation where patient takes his insulin, but rushes off to work without having breakfast. In such a situation he may be running a significant risk of hypoglycemia in the mid-morning. Fig. 1 shows the state of hypoglycemia.

3) Upper band

A severely high blood sugar may results in various symptoms like breathlessness. A point (x_i, y_i) would fall upper band if, $y_i > \beta$, for $0 \le i \le n$. For example one can simulate what would happen if patient 'x' forgot to take his morning insulin injection and can predict effect on his blood glucose profile. It can be seen that omitting morning insulin injection would send him markedly hyperglycemic in the afternoon and leave her at significant risk of developing diabetic ketoacidosis. Fig. 2 shows the state of hyperglycemic.

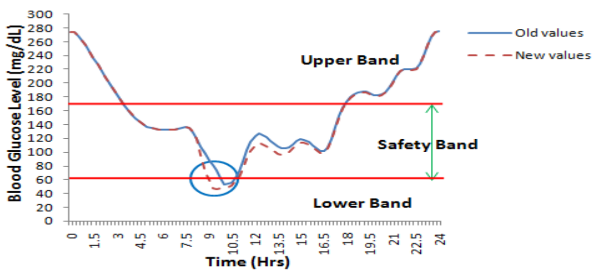

Figure 1. Hypoglycemia State

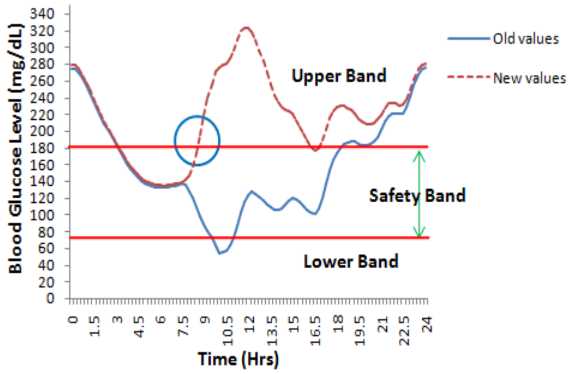

Figure 2. Hyperglycemia State

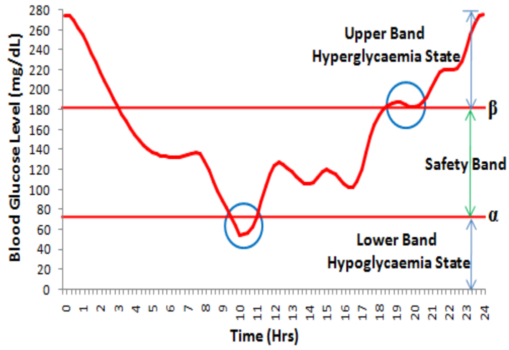

Figure 3. Blood Glucose Level

Fig. 3, if the blood glucose level crosses the safety bands it would enter either in lower or upper bands and thus results in unconscious state for patients. This is demonstrated by using a circle. The main advantage of using band ranges are that we can estimate the weight of each segment and predict the closeness of each other in order to find any early sign that can warn of a crucial event (such as hypoglycemia or hyperglycemia) [11].

B. Finding the Function of a Subsequence

In time series representation we can subdivide sequences into meaningful subsequences. The subsequences can be obtained using windowing and by finding a piecewise linear function able to approximately describe the entire initial sequence [12]-[13]. The subsequence can be defines as a continuous function defined on some closed interval [11].

C. Segmenting a Subsequence

All the subsequences of sequence are segmented and weights are assigned to them [14].

D. Approximating the Area under a Subsequence.

E. Estimating the Weight of a Segment.

IV. Case Study

A 45 year old man, and was diagnosed as having diabetes at the age of 14. He is currently on a regimen of combined short and (or) intermediate acting insulin preparations four times per day. He tends to higher blood glucose values overnight but has low blood glucose in the mid-morning. The blood sugar levels are various time intervals are given in Table I. We will try using the simulator to see how one could redistribute his insulin doses to improve his overall control [10].

A. Case

1) Treatment: As per his blood glucose level the meal and insulin dose suggested are given in Table II and Table III. The diet consists of 6 meals. The patient was on a 4 times insulin injection regimen. The simulated blood glucose level and plasma insulin level are demonstrated in figure 3 and figure 4 respectively.

2) Results: Fig. 3 shows the patient had low blood glucose (lower band) after breakfast and very high blood glucose values at bed time and mid night (upper band).

B. Case

1) Treatment: On the basis of above observations, the treatment can be modified as:

- *decrease dose of morning Actrapid*
- *increase dose of evening Actrapid*

There is no alteration in the diet plan has been considered. Only 20% changes in the insulin doses have been made. The modified insulin doses are shown in Table IV.

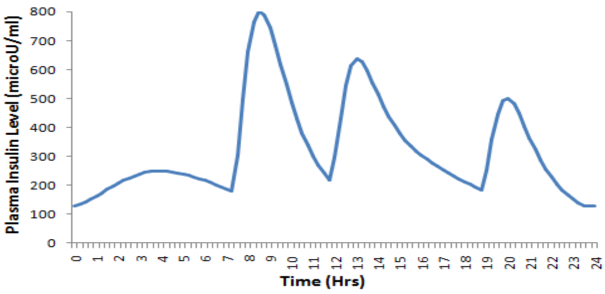

Figure 4. Plasma Insulin Level

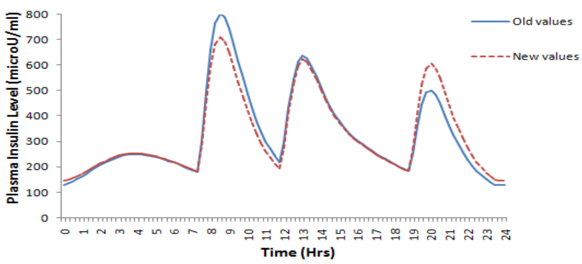

Figure 5. Plasma Insulin Level

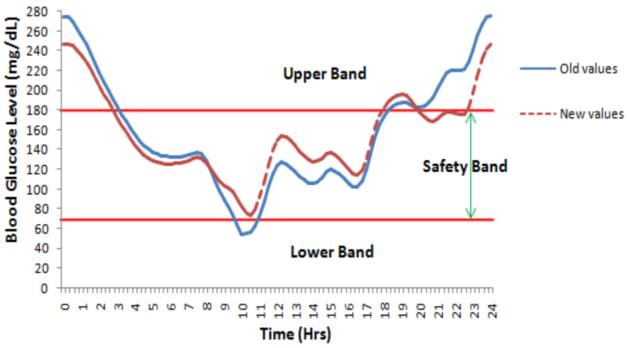

Figure 6. Blood Glucose Level Based on Insulin

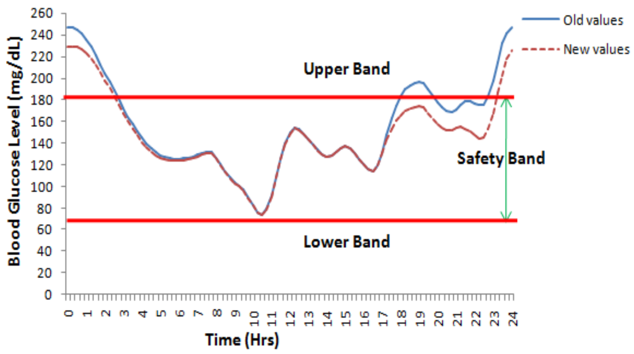

Figure 7. Blood Glucose Level Based on Carbohydrates

TABLE I. BLOOD GLUCOSE VALUES

Time (hh:mm)	06:00	08:45	13:30	17:00	20:00	24:00
Blood Glucose (mg/dL)	144	61.2	81	126	216	252

TABLE II. MEALS

Meals:	Break-fast	Snack	Lunch	Snack	Supper	Snack
Time (hh:mm):	07:30	10:00	12:00	16:00	19:00	22:00
Carbo. (gm):	30	20	50	20	40	20

TABLE III. INSULIN DOSES

Insulin Injection Dose (Time):	No. 1 (07:00)	No. 2 (11:30)	No. 3 (18:30)	No. 4 (23:00)
Actrapid :	6	4	3	0
Lentard MC:	0	4	0	6

TABLE IV. MODIFIED DOSES

Insulin Injection Dose (Time):	No. 1 (07:00)	No. 2 (11:30)	No. 3 (18:30)	No. 4 (23:00)
Actrapid :	5	4	4	0
Lentard MC:	0	4	0	6

TABLE V. MODIFIED MEAL

Meals:	Break-fast	Snack	Lunch	Snack	Supper	Snack
Time	07:30	10:00	12:00	16:00	19:00	22:00
Carbo. (gm):	30	20	50	15	35	20

2) Results: Fig. 6 shows there is a remarkable change in patient blood glucose level, the curve is not falling in lower band, so there is no chances of hypoglycemia. But after supper the patient is recorded with high blood glucose level. So, again we have to modify the treatment.

IACSIT
International Association of Computer Science and Information Technology
WWW.IACSIT.ORG

C. Case

1) Treatment: In this treatment we are modifying the diet of the patient. As the patient is having high glucose level at supper we reducing his carbohydrate intake, which are given in Table V.

2) Results: Fig. 7 shows the resultant blood glucose level. We can see very clearly that it has been reduced surprisingly during supper.

V. LIMITATIONS OF SIMULATOR

Following are the limitations of simulator: (i) the results are based on carbohydrate intake, no other foods are considered such as protein, fats etc., (ii) the simulation results are based on patients 24 hours blood glucose level and insulin dosages, estimated for one set of patient data on one day may not necessarily be accurate several days later, (iii) the system does not deals with other complications of diabetes like a person already having heart or kidney problem along with diabetes, or a women with gestational diabetes, (iv) other factors are also not considered such as exercise of stress which greatly affects the lives of many diabetic patients and (v) the simulator is only for educational purpose, so it cannot be intended for patient use.

VI. CONCLUSION AND FURTHER WORK

In this paper we have discussed how we can apply the concept of time series for planning the therapy of diabetes mellitus patient for type 1. We simulated a case on AIDA simulator and also suggested the possible treatment for the patient having insulin-dependent diabetes. Similarly the simulator can be used in planning the therapy and monitoring the blood-insulin level of diabetic patient.

A further work includes finding the function of a subsequence, segmenting a subsequence, approximating the area under a subsequence and estimating the weight of a segment and this work will be based on the concepts of safety, lower and upper band ranges.

REFERENCES

[1] A. Rajak, and K. Saxena, “Temporal-abstraction database mediators based on clinical guidelines,” in Proceedings of IEEE International Advance Computing Conference, Patiala, India, pp. 2274-2279, March 6-7, 2009.
[2] A. Rajak, and K. Saxena, “Managing temporal databases,” in proceedings of International Conference on Managing Next Generation Software Applications, Coimbatore, India, pp. 107-114, December 5-6, 2008.
[3] A. Rajak, and K. Saxena, “Modeling temporal database: a survey,” International Journal of Computer and Electronics Engineering, India, vol. 1, no. , pp. 77-82, Jan-June 2009.
[4] A. Rajak, and M. K. Gupta, “Association rule mining-applications in various areas,” in Proceedings of International Conference on Data Management, Ghaziabad, India, pp. 3-7, February 25-26, 2008.
[5] A. Rajak, and K. Saxena, “Research issues related to temporal mediators related to clinical domain,” International Journal of Recent Trends in Engineering, Finland, vol. 1, no. 1, pp. 535-540, May 2009.
[6] 2009, Aug.). Diabetes. [Online]. Available: http://diabetesinformationhub.com
[7] (2009, Aug.). Diabetes mellitus. [Online]. Available: http://en.wikipedia.org/wiki/diabetes_melitus
[8] (2009, Aug.). Diabetes Type 1. [Online]. Available: http://www.nlm.nih.gov/medlineplus/diabetestype1.html
[9] E. D. Lehmann, T. Deutsch, and D. Broad, “The AIDA v4.3a diabetes simulator,” http://www.2aida.org.
[10] E. D. Lehmann, and T. Deutsch, “A physiological model of glucose-insulin interaction,” in Proceedings of Thirteenth Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Orlando, Florida, USA, pp. 2274-2275, 1991.
[11] M. Samia, “A representation of time series for temporal web mining,” in Proceedings of 16th GI-Workshop on the Foundations of Databases, Moheim, Germany, pp. 103-107, June 2004.
[12] M. Datar, and S. Muthukrishnan, “Estimating rarity and similarity over data stream windows,” in Proceedings of the 10th European Symposium on Algorithms, Rome, Italy, September 2002.
[13] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, “Fast subsequence matching in time series databases,” in Proceedings of ACM SIGMOD International Conference on Management of Data, Minneapolis, USA, pp. 419–429, 1994.
[14] E. Keogh, and M. Pazzani, “Relevance feedback retrieval of time series data,” in Proceedings of SIGIR ’99, Berkley, CA USA, August 1999.

Akash Rajak was born in Guna (M.P), India in 1978. He has completed M.C.A from U.T.D campus, Dr. H. S. University, Sagar (M.P) (Central University) in 2002. He is pursuing his Ph. D in computer science from Barkatullah University, Bhopal. His research interests are in the areas of temporal data mining, intelligent systems and time series analysis. He is working as Associate Professor in Krishna Institute of Engineering and Technology, Ghaziabad, India. He has 7 publications in various international journals and conferences. He has more than 7 years of teaching and research experience. Mr. Rajak is a member of various professional societies like ACEEE, IACSIT, and IAENG. He is also a reviewer of various international conferences and international journals. He is program committee chair of various ACEEE conferences. He has also reviewed the books of Tata McGraw-Hill Education, India.

Kanak Saxena received her Ph. D. in computer science from the Devi Ahilya University, Indore, India. She is professor in the Computer Applications Department at the Samrat Ashok Technological Institute affiliated to Rajiv Gandhi Technical University, Bhopal. Her current research focuses on database systems, parallel computing, data uncertainty and design and other interests include network security and performance and software engineering. She is the member of the editorial board of international journals and member of various international committees of international conferences.