Monitoring and Predicting of Blood Sugar Level by Using an Artificial Intelligence Methodology

Introduction

HDCU, a hybrid data mining and case-based reasoning user modeling system, which is used to monitor and predict the blood sugar level in diabetics. The practical objective for this project is to reduce the cost of direct blood sugar self monitoring by minimizing the number of times that a diabetic needs to measure his or her sugar levels every day. From the technological point of view, the main aim is using the support vector machine as the classifier and implementing a case-based reasoning cycle as the problem-solved cycle in order to indirectly determine and predict blood sugar level in diabetics. The paper provides the implementation details along with the corresponding results for most conventional classifiers. Several comparative studies have been carried out concerning different sample size demonstrating the superiority of the proposed HDCU methodology in terms of reliability specificity and accuracy.

Inference Engine

Basically, the support vector machine is a learning method for the design of a feed forward network with a single hidden layer of nonlinear units. The hidden layer is central to the construction of the support vector learning algorithm which is the inner-product kernel between a "support vector" and the attribute drawn from the input vector.
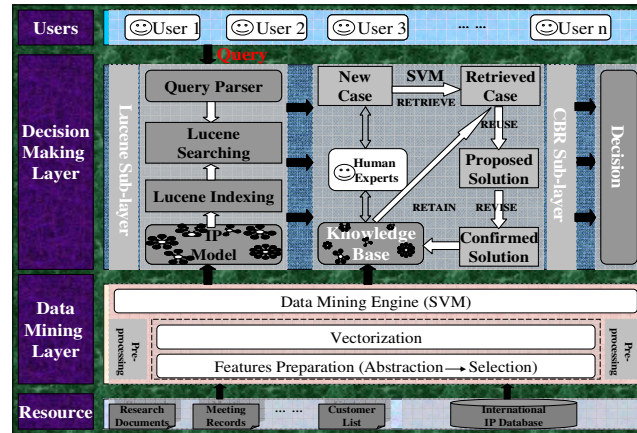
Given a training set of class-label pairs (x_i, y_i), $i = 1, ..., l$ where $x_i \in R^n$ and $y \in \{-1, 1\}^l$, the support vector machines require the solution of the following optimization problem:

$$\min_{W, b, \xi} \frac{1}{2} W^T W + C \sum_{i=1}^{l} \xi_I$$

Subject to

$$y_i (W^T \phi(x_i) + b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0.$$

Components & Levels

Seven main components of the system architecture:

- (1) A user class, turning a user into an appropriate blood sugar object instance.
- (2) A knowledge base, containing organized user model about diabetes.
- (3) A data mining engine SVM which classifies both patient information and diabetes domain information.
- (4) A set of pre-processing functions that provide the uniform data format in order to access the Data Mining engine.
- (5) A problem-solving life-cycle called the case-based reasoning cycle, assisting in the retrieve, reuse, revise and retain process in the knowledge base.
- (6) A case-based reasoning retrieval engine SVM which treats the new user as a testing case, and returns the most similar case.
- (7) A resource collection function which gathered all related information

The hybrid system was integrated with four levels:

- a) User Level   b) Data Mining Level
- c) Case-Based Reasoning Level
- d) Resource Level to achieve the diagnose functions.

Illustration of Experiment Results

Classifiers Measures	6 Dimension							
	Poly	NPoly	Puk	RBF	KNN	NB	RDT	ORBR
Correctly Classified	0.8511	0.8511	0.8511	0.8511	0.8298	0.3830	0.8191	0.1489
Seconds to build Model	0.5300	0.1900	0.1900	0.1900	0.0000	0.0000	0.0300	0.0000
Root Mean Squared Error	0.3859	0.3859	0.3859	0.3859	0.2114	0.5226	0.4006	0.9225
Root Relative Squared Error	1.0818	1.0818	1.0818	1.0818	1.1083	1.4650	1.1123	2.5860
Mean Absolute Error	0.1489	0.1489	0.1489	0.1489	0.2114	0.4999	0.2586	0.8511
Relative Absolute Error	0.5741	0.5741	0.5741	0.5741	0.8150	1.9271	0.9970	3.2805
Kappa statistic	0.0000	0.0000	0.0000	0.0000	0.2389	0.0815	0.0555	0.0000
Class	C1							
TP	1.0000	1.0000	1.0000	1.0000	0.9250	0.2880	0.963	0.0000
FP	1.0000	1.0000	1.0000	1.0000	0.7140	0.0710	1.0000	0.0000
Precision	0.8510	0.8510	0.8510	0.8510	0.8810	0.9580	0.8460	0.0000
Recall	1.0000	1.0000	1.0000	1.0000	0.9250	0.2880	0.9630	0.0000
F-measure	0.9200	0.9200	0.9200	0.9200	0.9020	0.4420	0.9010	0.0000
Roc Area	0.5000	0.5000	0.5000	0.5000	0.6940	0.6020	0.4560	0.5000
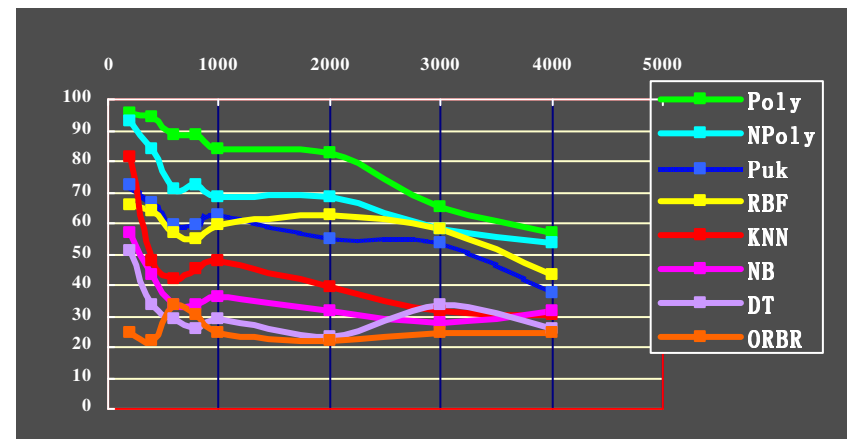
Experiment Design

- Do the weighting of four common inner-product kernels of the support vector machine: Poly Kernel, Normalized Poly Kernel, Puk and Radial Basis Function (RBF) Kernel; each with two groups of sample dataset, small size sample and large size sample.
- Do the weighting of four conventional classifiers the K-Nearest Neighbors (K-NN), the Naïve Bayes (NB), the Decision Tree (DT) and the Rule based reasoning function, each with the same two groups of sample dataset, small size sample and large size sample.

Researcher: Andrzej Bargiela
Andrzej.Bargiela@nottingham.edu.my

Researcher: Dino Isa
Dino.Isa@nottingham.edu.my

Researcher: Chen Zhi Yuan
eyx6czy@nottingham.edu.my

Researcher: Roselina Arelhi
roselina.arelhi@nottingham.edu.my

Researcher: Yong Su Ting
Su-Ting.Yong@nottingham.edu.my