The 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
21-23 September 2017, Bucharest, Romania

Application of Regularized Ranking and Collaborative Filtering in Predictive Alarm Algorithm for Nocturnal Hypoglycemia Prevention

Galyna Kriukova[1] , Nadiya Shvai[1] , Sergei V. Pereverzyev[2]
[1] Department of Mathematics, Faculty of Computer Sciences, National University of "Kyiv-Mohyla Academy", Skovorody Str., 2, Kyiv, Ukraine, kriukovagv@ukma.kiev.ua, http://www.ukma.edu.ua
[2] Johann Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenberger Str., 69, Linz, Austria, sergei.pereverzyev@oeaw.ac.at, http://www.ricam.oeaw.ac.at

Abstract – **In this paper we propose a predictive alarm ranking algorithm for prevention of nocturnal hypoglycemia (NH) in patients with insulin-treated diabetes. We adapt collaborative filtering method for prognosis of risk of NH occurrence in diabetic patients. We illustrate the proposed algorithm and test it to quantify its accuracy with medical records obtained within the framework of the European FP7-funded project DIAdvisor. Results show that the algorithm with proper ranking and preprocessed data outperforms well-known NH-prediction approaches in terms of sensitivity, specificity and F1-score.**

Keywords – **predictive algorithm; regularization; ranking; collaborative filtering**

I. INTRODUCTION

Diabetes mellitus is a common and serious autoimmune disease characterized by the inability of the pancreas to produce insulin, an essential hormone needed to convert food into energy, and thus to regulate blood glucose concentration. The management goal is to delay or prevent serious long-term diabetic complications, including blindness, kidney failure, strokes, heart attacks [1]. According to Diabetes Controls and Complication Trial [1] the risk of diabetes can be prevented by proper blood glucose monitoring and regulation. Hence, in order to control blood glucose evolution, self-monitoring of blood glucose with use of glucometers (from capillary blood obtained through finger prick) or continuous glucose monitoring devices is essential. However, blood glucose levels abnormally fluctuate, and too little insulin injection results in chronic high blood glucose levels, too much can cause hypoglycemia. For both patients and doctors diabetes management can pose a rather complicated task: patients are required to track their blood glucose levels and daily activities, and doctors should make appropriate therapeutic adjustments based on the monitoring data. A clinically important task in Diabetes treatment is prevention of nocturnal hypoglycemic events, as blood glucose level falls below 70 mg/dl [2]. Naturally, would be beneficial if the problem of blood glucose regulation could be treated proactively, i.e. the alerts would be given not at the moments of blood glucose excursions, but beforehand in a predictive way based on previous blood glucose measurements. This task can be addressed as a learning-to-rank problem. However, this task is ill-posed, so application of regularization methods is required.

There are several techniques that could be used to target this problem described in the literature [3]–[8]. Nevertheless, wide variability among individual patients in terms of biologic and behavioral factors such as awareness of hypoglycemia, treatment decisions and strategies, sensitivity to insulin, response to lifestyle factors, and response to treatment result in high complexity of the task. We develop a prediction algorithm using "collaborative filtering" (CF) technique based on assumption that an unknown behaviour, preference or characteristics of a person can be estimated using the similarity of this person to other person [9]. CF may be used to process all the available information and to be able to adjust our ranking function for a particular individual. It was shown [10], that this approach in diabetology is a well-approved method for predicting the level of methabolic control. CF predictions are specific to the individual, but use information gleaned from many other persons. The fundamental assumption of CF is that if persons A and B rate some inputs identically, hence will rate other inputs similarly.

The paper is organized as follows. In section II, the problem statement is presented and our approach is con-

The authors gratefully acknowledge the support of the project AMMODIT funded within EU H2020-MSCA-RICE. S.V. Pereverzyev is partially supported by the Austrian Science Foundation (FWF), project I1669.

sidered. In section III we discuss a theoretical background for regularized ranking algorithm. In section IV, we illustrate the proposed algorithm with clinical data and discuss its accuracy.

II. Problem Statement

The ranking problem can be stated as follows: we are given a sample set $z = \{(x_i, y_i)\}_{i=1}^m$, where $x_i \in X$ is input vector, and $y_i \in Y(Y = [0, M], M > 0)$ denotes output (in our case this is the risk of nocturnal hypoglycemia). The goal is to find a scoring function $f_z : X \to Y$ that ranks future instances x and x' with labels $f(x) < f(x')$ if $x \prec x'$ (x' is more alarming than x). To account for individual patient's differences, function f is trained for each patient separately.

For vector of blood glucose measurements during the day $g = (g^1, g^2, g^3, \dots, g^l) \in \mathbb{R}^l$ ranking function f should predict risk of NH. To use information from other individuals we want to use outputs from other persons with the same or close daily bloog glucose measurements. Accordingly, we apply k-mean clustering on space of vectors $\mathbb{R}^l$ from our dataset (we used $k = 20$), hereby for each vector $g = (g^1, g^2, g^3, \dots, g^l)$ is marked with cluster label g_{lk}. From the dataset we find those records, for which input vector falls in the same cluster and find the risk associated with the cluster (i.e. risk of NH occurrence for other individuals when appear in the cluster) $x^1, \dots, x^n$ (some of them may be undefined). At this point PCA is used to reduce complexity and sparsity of data. These ranking are then used as input $x = (x^1, \dots, x^n)$ and risk of NH occurrence as output y.

The general setting of the ranking problem has been well studied in [11] on the basis of algorithmic stability analysis. In our ranking problem we wish not just to accurately predict pairwise ordering but also preserve the magnitude of the scores or the difference between ratings, therefore we estimate quality of a ranking function by its expected ranking error corresponding to least squares ranking loss [12]. To overcome problem of overfitting, function f is selected from some hypothesis space $\mathcal{H}$ using Tikhonov regularization scheme. A natural choice of a hypothesis space $\mathcal{H}$ is a Reproducing Kernel Hilbert Space (RKHS). Regularization in a RKHS has a number of attractive features, including the effective error bounds and stability analysis relative to perturbations of the data. For further detailed information, please refer to [11], [12].

III. REGULARIZED RANKING ALGORITHM

Let X be a compact metric space and $Y = [0, M]$, for some $M > 0$. An input $x \in X$ is related to an output $y \in Y$ through an unknown probability distribution $\rho(x, y) = \rho(y|x)\rho_X(x)$ on $Z = X \times Y$, where $\rho(y|x)$ is the conditional probability of Y given x and $\rho_X(x)$ is the marginal probability of x. The distribution ρ is given only through a set of samples $z = \{(x_i, y_i)\}_{i=1}^m$. The ranking problem aims to automatically learn a function $f_z : X \to \mathbb{R}$ that ranks future instances.

A ranking loss function $l = l\big(f, (x, y), (x', y')\big)$ is utilized to evaluate the prediction result of $f(\cdot)$. The loss reflects the coherence between scores and ranking: if feature vectors with higher scores also have higher rank, then the loss is small. Otherwise, the loss is large. We further define the risk function as the expected loss function with respect to the distribution $\rho(x, y)$:

$$\mathcal{E}_l(f) = E_{(x,y),(x',y')\sim\rho}[l\big(f, (x, y), (x', y')\big)].$$

Empirical l-error of f with respect to $z = \{(x_i, y_i)\}_{i=1}^m$ will be defined as

$$\widehat{\mathcal{E}}_l(f, z) = \frac{\sum_{i=1}^{m-1}\sum_{j=i+1}^{m} l\big(f, (x_i, y_i), (x_j, y_j)\big)}{\sum_{i,j=1}^{m} 1_{y_i > y_j}}.$$

The learning task may be considered as a minimization problem with empirical risk function as target function. In this case the choice of the loss function corresponds to the choice of a ranking model. For magnitude-preserving ranking the most natural loss function is square-loss function

$$l = \big(f(x') - f(x) - (y' - y)\big)^2.$$

The quality of ranking function f is expressed with the expected error

$$\mathcal{E}(f) = \int_Z \int_Z (y - \bar{y} - (f(x) - f(\bar{x})))^2 d\rho(x, y) d\rho(\bar{x}, \bar{y}). \tag{1}$$

Note, that although $\mathcal{E}(f)$ is a convex function of f, it does not imply the uniqueness of function f^* that would minimize (1). In the space of square integrable functions with respect to the marginal probability measure ρ_X the risk $\mathcal{E}(f)$ can be minimized by a family of functions $f_\rho(x) + c$, where

$$f_\rho(x) = \int_Y y d\rho(y|x)$$

is the so-called target function or Bayes estimator, and c is a generic constant, which may take different values at different occurrences. The function $f_\rho(x)$ is also called regression function, as it is a solution for the regularized least-square method [13].

However, the target function $f_\rho(x)$ can not be found in practice, because the conditional probability $\rho(y|x)$ is unknown, and only a sample of it, the training set $z = \{(x_i, y_i)\}_{i=1}^m$, is available. Therefore, it is convenient to look for a function f from some hypothesis space $\mathcal{H}$ minimizing the approximation error $\|f - f_\rho\|_{\mathcal{H}}$.

A natural choice of a hypothesis space $\mathcal{H}$ is a Reproducing Kernel Hilbert Space (RKHS). From [14] it is known that every RKHS is generated by a unique symmetric and positive definite continuous function $K : X \times X \to \mathbb{R}$, called the reproducing kernel of $\mathcal{H}_K$, or Mercer kernel. The RKHS $\mathcal{H}_K$ is defined to be a closure of the linear span of the set of functions $\{K_x := K(x,\cdot) : x \in X\}$ with the inner product $\langle\cdot,\cdot\rangle_K$ defined as $\langle K_x, K_{\bar{x}}\rangle_K = K(x,\bar{x})$. The reproducing property takes the form $f(x) = \langle f, K_x\rangle_K$.

The regularization-based algorithm can be defined by the following regularization scheme:

$$f_z^\lambda = \arg\min_{f\in\mathcal{H}_K} T(f),$$

where functional $T(f) = T_{K,\lambda,z}(f)$ is defined as follows

$$T(f) = \frac{1}{m^2}\sum_{i,j=1}^{m}\big(y_i - y_j - (f(x_i) - f(x_j))\big)^2 + \lambda\|f\|^2_{\mathcal{H}_K}$$

Data-free regularization scheme is the following

$$f^\lambda = \arg\min_{f\in\mathcal{H}_K}\{\mathcal{E}(f) + \lambda\|f\|^2_{\mathcal{H}_K}\},$$

where $\|f\|^2_{\mathcal{H}_K}$ is the norm in RKHS $\mathcal{H}_K$, and λ is a regularization parameter. We define the integral operator $L : \mathcal{H}_K \to \mathcal{H}_K$ as

$$Lf = \int_X\int_X f(x)(K_x - K_{\bar{x}})d\rho_X(x)d\rho_X(\bar{x}). \qquad (2)$$

One can see that the operator L is self-adjoint and positive linear operator on $\mathcal{H}_K$, and that the function f^λ is a Tikhonov-regularized solution of the equation

$$Lf = Lf_\rho.$$

Finally,

$$\langle Lf, f\rangle_K = \int_X f^2(x)\rho_X(x) - \Big(\int_X f(x)\rho_X(x)\Big)^2 \geq 0,$$

and the operator L is self-adjoint. Using standard representer theorem [13], [14] the minimizer of the optimization problem has the form

$$f_x^\lambda = \sum_{i=1}^{m} c_{i,z}K_{x_i},$$

with $c_z = (c_{1,z},\ldots,c_{m,z})^T \in \mathbb{R}^m$. Therefore under the assumption that $f_\rho \in \mathcal{H}_K$ the regularized solution of the problem can be obtained in the following form:

$$f^\lambda = (L + \lambda I)^{-1} L\mathbf{y},$$

or in the discrete form

$$f_z^\lambda = \left(\frac{1}{m^2}S_x^*DS_x + \lambda I\right)^{-1}\frac{1}{m^2}S_x^*D\mathbf{y}, \qquad (3)$$

where $S_x : \mathcal{H}_K \to \mathbb{R}^m$ is the so-called sampling operator, i.e.

$$S_x(f) = (f(x_1), f(x_2), \ldots, f(x_m))^T,$$

and its adjoint $S_x^* : \mathbb{R}^m \to \mathcal{H}_K$ can be written as

$$S_x^*c = \sum_{i=1}^{m} c_iK_{x_i},\ c = (c_1,\ldots,c_m)^T,$$

$D = m\mathbf{I} - \mathbf{1}\cdot\mathbf{1}^T$, $\mathbf{y} = (y_1,\ldots,y_m)^T$, and $\mathbf{I}, \mathbf{1}$ are the m-th order unit matrix and the m-th order column vector of all ones respectively. The operator $\frac{1}{m^2}S_x^*DS_x$ can be seen as a discrete version of the operator L. Therefore,

$$\langle S_x^*DS_x(f), f\rangle_K = \sum_{i,j=1}^{m}(f(x_i) - f(x_j))^2.$$

Note, that in the supervised learning regression the operator L appearing in (2) has the form

$$Lf = \int_X f(x)K_x d\rho_X(x).$$

Therefore regularized solution (and its discrete form) for regression problem differs from ranking problem solution, despite of fact that these two algorithms have the common solution $f_\rho(x)$.

IV. NUMERICAL ILLUSTRATIONS

In our study we analyze value of the self-monitoring of blood glucose taken three times during the day (before breakfast, lunch, dinner) and once around 22:00. Prediction horizon is nine hours, so the task is to estimate risk of hypoglycemia occurence during the night.

To illustrate this approach we use 2 data sets.

One of them is data from 34 diabetic subjects studied in the Montpellier University Hospital Center (France) and in the Department of Clinical and Experimental Medicine at the University of Padova (Italy) as a part of EU-project "DIAdvisor". This dataset contains the records on blood glucose concentration. The measurements were taken with the glucose meter at the same time during nearly 10 days in a row for each individual. We selected only those measurements corresponding to breakfast, lunch, dinner and bedtime. Onset of NH was detected from the system of continuous blood glucose measurements during a night. NH was observed in 27% of day-night outcomes.

Another one is Diabetes Data Set [15], containing the information for 70 sets of glucose, insulin, and lifestyle data recorded on diabetes patients during several weeks or months. The data on diabetes patients was obtained from two sources: an automatic electronic recording device and paper records. These sources differ in the way time of the record was fixed. For automatic device, there was an internal clock that would fixate the exact time of the record, whereas for the paper records the four fixed time slots were given to be associated with the data points: breakfast (08:00), lunch (12:00), dinner (18:00), and bedtime (22:00). Hence for paper records the timestamps were artificially categorized whereas electronic records contain actual timestamps. Moreover, paper records almost do not contain reports about nocturnal hypoglycemic symptoms. Accordingly, for our experiment we use only records with electronic inputs.

For both data sets we formed list, each row containing subject ID, day number, 4 blood glucose measurements and final value 1 if nocturnal hypoglycemia was observed, and 0 otherwise. For example,

Table I. EXAMPLE OF RECORD FROM DATASET

Subject ID	Day	Blood glucose measurements				Risk
2	1	192	158	200	127	1

Vector space of blood glucose measurements $\mathbb{R}^4$ was split into 20 clusters with k-means algorithm. So instead of vector from $\mathbb{R}^4$ we use number of corresponding cluster, or Voronoi cell. So, for example, vector $(192, 158, 200, 127)$ is transformed into 3.

Then for selected individual the ranking function f is constructed. For each other subject s and each cluster i we form rating $x_{s,i}$ as mean of final values for subject s at cluster i (this value is from 0 to 1). Missing values in the input features are populated with value -1. Vector of other patients' ratings $(x_{1,i}, x_{2,i}, \ldots, x_{u,i})$ forms input vector x_i. Then PCA was applied to reduce dimensionality of input space, for 30 components explained variance ratio exceeds 90%. Thus inputs are vectors from $\mathbb{R}^{30}$, and outputs are values in $[0, 1]$.

Training was carried out on half of the selected individual's records and testing was performed on the other half. The experiment was done for all subjects and the average performance recorded. As a hypothesis space $\mathcal{H}_K$ we used the RKHS generated by the universal Gaussian kernel

$$K(x, \bar{x}) = \exp(-\frac{\|x - \bar{x}\|_2^2}{\gamma}).$$

For each dataset, the width of the kernel γ and regularization parameter λ were optimized on a held-out sample.

We compare our method to algorithm based on the Low Blood Glucose Index (LBGI) that has been introduced in [6], [7] for measuring the risk of severe hypoglycemia, based on self-monitoring of blood glucose. LBGI is nonnegative value equal to mean of quadratic risk function calculated for each given blood glucose measurement. The main advantage of this algorithm is simplicity of its estimation. For details on the calculation of LBGI we refer to the papers by Kovatchev et al. [6], [7], where the index has been introduced and justified.

Both algorithms calculate risk of NH, that is some real numbers. To evaluate the performance of predictors in terms of Sensitivity (SE), Specificity (SP), Positive Predictive Value (PPV), Negative Predictive Value (NPV), and F1-score, we use various levels of risk threshold for both predictors, obtaining *true-false* outcomes ($1_{y>threshold}$). Results are presented in Table II.

Table II. COMPARATIVE PERFORMANCE OF PREDICTORS

Predictor	threshold	SE	SP	PPV	NPV	F1
LGBI	0.5	0.36	0.40	0.25	0.53	0.3
	1	0.27	0.55	0.25	0.58	0.26
	1.5–2	0.18	0.70	0.25	0.61	0.21
	2.5	0.18	0.75	0.29	0.63	0.22
	3	0	0.85	0	0.61	0
	3.5	0	0.90	0	0.62	0
	4–5	0	0.95	0	0.63	0
Proposed method	0.05	0.91	0.15	0.37	0.75	0.53
	0.10	0.82	0.20	0.36	0.67	0.5
	0.15	0.82	0.25	0.38	0.71	0.51
	0.20	0.55	0.45	0.35	0.64	0.43
	0.25	0.45	0.55	0.36	0.65	0.4
	0.30–0.35	0.45	0.60	0.38	0.67	0.42
	0.40–0.45	0.27	0.75	0.38	0.65	0.32
	0.50	0.27	0.90	0.6	0.69	0.38

For predicted risk values we use pairwise misrankings to compare to actual risk values in the test set of size t:

$$\frac{\sum_{i,j=1}^{t} 1_{y_i > y_j \wedge f(x_i) \le f(x_j)}}{\sum_{i,j=1}^{t} 1_{y_i > y_j}}.$$

Then experiment was performed for 20 selected individuals from each dataset, we report mean values and standard deviations of the pairwise misranking in Table III.

Table III. RANKING PERFORMANCE

Dataset	Pairwise Misranking		LBGI
	mean	standard deviation	
DIAdvisor	17.5%	7.6%	46%
Diabetes Data Set [15]	20.8%	4.5%	42%

V. CONCLUSIONS

Diabetes management is challenging domain for medical scientific research. Our algorithm for prediction of nocturnal hypoglycemia uses available self-monitoring data, this computational procedure can be incorporated in self-monitoring devices or applications. We conclude that the algorithm offers assessments of the risk of nocturnal hypoglycemia and it can be used for improvement in glycemic control. It is also worth to note that our method can be used as NH-indicators for both type 1 and type 2 diabetes patients.

However, further investigation is required to reduce complexity of ranking problem and to adjust partition of the data space into clusters. To adapt the algorithm to new situations and new observations the techniques for online machine learning may be applied.

ACKNOWLEDGMENT

This work started when the first author had been working in the scientific group of Prof. Dr. S.V. Pereverzyev at the Johann Radon Institute for Computational and Applied Mathematics (RICAM). Galyna Kriukova thanks him for the support and kind hospitality.

References

[1] The Diabetes Control and Complications Trial Research Group, “The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus,” *New England Journal of Medicine*, vol. 329, no. 14, pp. 977–986, 1993, pMID: 8366922. [Online]. Available: http://dx.doi.org/10.1056/NEJM199309303291401

[2] E. R. Seaquist, J. Anderson, B. Childs, P. Cryer, S. Dagogo-Jack, L. Fish, S. R. Heller, H. Rodriguez, J. Rosenzweig, and R. Vigersky, “Hypoglycemia and diabetes: A report of a workgroup of the american diabetes association and the endocrine society,” *Diabetes Care*, vol. 36, no. 5, pp. 1384–1395, 2013. [Online]. Available: http://care.diabetesjournals.org/content/36/5/1384

[3] G. Whincup and R. Milner, “Prediction and management of nocturnal hypoglycaemia in diabetes,” *Archives of Disease in Childhood*, vol. 62, no. 4, pp. 333–337, 1987.

[4] A. Davies, “Prediction and management of nocturnal hypoglycaemia in diabetes,” *Archives of Disease in Childhood*, vol. 62, no. 10, p. 1085, 1987.

[5] D. Cavan, R. Hovorka, O. Hejlesen, S. Andreassen, and P. Sönksen, “Use of the dias model to predict unrecognised hypoglycaemia in patients with insulin-dependent diabetes,” *Computer Methods and Programs in Biomedicine*, vol. 50, no. 3, pp. 241 – 246, 1996. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0169260796017531

[6] B. P. Kovatchev, D. J. Cox, L. A. Gonder-Frederick, D. Young-Hyman, D. Schlundt, and W. Clarke, “Assessment of risk for severe hypoglycemia among adults with iddm: validation of the low blood glucose index.” *Diabetes Care*, vol. 21, no. 11, pp. 1870–1875, 1998. [Online]. Available: http://care.diabetesjournals.org/content/21/11/1870

[7] D. J. Cox, L. Gonder-Frederick, L. Ritterband, W. Clarke, and B. P. Kovatchev, “Prediction of severe hypoglycemia,” *Diabetes Care*, vol. 30, no. 6, pp. 1370–1373, 2007. [Online]. Available: http://care.diabetesjournals.org/content/30/6/1370

[8] P. Tkachenko, G. Kriukova, M. Aleksandrova, O. Chertov, E. Renard, and S. V. Pereverzyev, “Prediction of nocturnal hypoglycemia by an aggregation of previously known prediction approaches: proof of concept for clinical application,” *Computer Methods and Programs in Biomedicine*, vol. 134, pp. 179–186, 2016.

[9] X. Su and T. M. Khoshgoftaar, “A survey of collaborative filtering techniques,” *Advances in Artificial Intelligence*, vol. 2009, p. 19, 2009, article ID 421425. [Online]. Available: http://dx.doi.org/10.1155/2009/421425

[10] M. Grabert, R. W. Holl, U. Krause, I. Melzer, and F. Schweiggert, *Predicting the Level of Metabolic Control Using Collaborative Filtering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 108–112. [Online]. Available: http://dx.doi.org/10.1007/3-540-45497-7_16

[11] S. Agarwal and P. Niyogi, “Generalization bounds for ranking algorithms via algorithmic stability,” *J. Mach. Learn. Res.*, vol. 10, pp. 441–474, Jun. 2009. [Online]. Available: http://dl.acm.org/citation.cfm?id=1577069.1577085

[12] G. Kriukova, S. Pereverzyev, and P. Tkachenko, “On the convergence rate and some applications of regularized ranking algorithms,” *Journal of Complexity*, vol. 33, pp. 14–29, 2016.

[13] S. Smale and D.-X. Zhou, “Learning theory estimates via integral operators and their approximations,” *Constr. Approx.*, vol. 26, pp. 153–172, 2007.

[14] G. Wahba, “Spline models for observational data,” *CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 59, 1990.

[15] [Online]. Available: http://archive.ics.uci.edu/ml/datasets/Diabetes