Gaussian Process Regression with Categorical Inputs for Predicting the Blood Glucose Level

Jakub M. Tomczak(✉)

Department of Computer Science, Faculty of Computer Science and Management, Wrocław University of Science and Technology, Wrocław, Poland
jakub.tomczak@pwr.edu.pl

Abstract. In diabetes treatment, the blood glucose level is key quantity for evaluating patient's condition. Typically, measurements of the blood glucose level are recorded by patients and they are annotated by symbolic quantities, such as, date, timestamp, measurement code (insulin dose, food intake, exercises). In clinical practice, predicting the blood glucose level for different conditions is an important task and plays crucial role in personalized treatment. This paper describes a predictive model for the blood glucose level based on Gaussian processes. The covariance function is proposed to deal with categorical inputs. The usefulness of the presented model is demonstrated on real-life datasets concerning 10 patients. The results obtained in the experiment reveal that the proposed model has small predictive error measured by the Mean Absolute Error criterion even for small training samples.

Keywords: Gaussian process · Categorical data · Diabetes · Nonparametric regression

1 Introduction

Diabetes is reported to be one of the most dangerous chronic disease that afflicted around 171 million people in the world in the year 2000 [26]. The increasing number of diabetics entails growing total costs of a treatment, *i.e.*, pharmacological treatment, hospitalization, laboratory test, medical visits, and constant patient health monitoring. Therefore, there is a need to propose personalized therapy to lower the costs and make the disease bearable for patients [22]. In diabetes treatment, the blood glucose level is crucial quantity for evaluating patient's condition. Understanding the influence of different factors on the glucose level and possibility to predict its values for new measurements would give opportunity to design therapy-effective decision-support systems.

First mathematical models for diabetes aimed at understanding the biochemical processes governing the blood glucose level. The mechanistic models were proposed to explain dependencies between the glucose level and insuline and food ingestion [9]. However, such approach fails in several apsects. First, it is troublesome to propose correct relations using dynamical systems. Second, in practice it

© Springer International Publishing AG 2017
J. Świątek and J.M. Tomczak (eds.), *Advances in Systems Science*, Advances in Intelligent Systems and Computing 539, DOI 10.1007/978-3-319-48944-5_10

is almost impossible to force patients to record all meals represented by ingested calories. Additionally, patients have tendency to forget to report all important information which results in unreliable models. Third, sometimes parameters of the models do not represent any physical quantities and hence the justification of applying the mechanistic models is put into question. Fourth, there are many external factors which affect the blood glucose level, *e.g.*, lifestyle, which cannot be included in the mechanistic models. Fifth, except numeric values of the blood glucose level, we can have access to a non-numeric (*symbolic*) description of the measurement, *e.g.*, day of a week, period of day, measurement's code representing insuline dose, or food ingestion. Typically, such information cannot be included in the mechanistic models.

All these issues cause a demand to formulate new models that allow to predict the blood glucose level for symbolic data. In other words, we need to propose a model for a relation between non-numeric inputs (symbolic variables representing measurements) and numeric output (the blood glucose level). The problem of symbolic variables is an important issue in the modern modelling and different types of symbolic data are distinguished. There are *categorical (nominal) variables* that take values in a finite set with no ordering, and *ordinal variables* that take values in a finite set with an ordering between the values but no metric notion is appropriate [1]. Moreover, there are *structural variables* that take values in sets of mathematical structures, *e.g.*, graphs [4].

The theoretical inquires about symbolic data for classification or regression models are forced by many practical applications, *e.g.*, nominal data in credit scoring [27] and medicine [22], structural data in biology [10], biochemistry [21], and chemistry [25]. Therefore, practice requires developing new models to cope with symbolic data [4]. There are methods for clustering, see [14], dimensionality reduction, see [18], mixture models, see [16], classifiers, *e.g.*, logistic regression [5], and regression models, *e.g.*, CART [6].

In this paper, we cope with the regression problem with categorical inputs, in which mechanistic models fail completely. Moreover, because of the specificity of the domain, we would like to apply a non-parametric model in order to avoid proposing explicit parametrization of the model. In machine learning, one of the most successful non-parametric regression model is *Gaussian process regression* [19]. It has been applied to numerous applications, *e.g.*, biosystems [3], discovering biomarkers in microarray gene expression data [7], chemical plants [17], non-linear system identification [24], predicting Quality-of-Service in Web service systems [23]. Additionally, Gaussian process regression allows to find a relation between any kind of inputs and output because similarity between objects is expressed by a kernel function. Hence, the core of the approach is to define appropriate kernel function for symbolic inputs [12,20].

The contribution of the paper is the following. First, the Gaussian process regression as the predictive model for the blood glucose level is outlined. Second, the covariance function for categorical inputs is presented. Third, the mean function for categorical inputs is proposed. Fourth, the learning of hyperparameters is outlined. Fifth, the experiment with real-life data is conducted.

2 Methodology

Let us consider a dataset $\mathcal{D}$ of N measurements of patient's blood glucose level. Each observation is represented by measurement's description, denoted as a vector of D categorical (nominal) variables $\mathbf{x} \in \mathcal{X}$,[1] and measured blood glucose level, $y \in \mathbb{R}_+$. The variables $\mathbf{x}$ will be called *inputs* and y – *output*. Further, we write $\mathbf{X}$ to denote a matrix of training inputs, and $\mathbf{y}$ – a vector of training outputs.

2.1 Gaussian Process Regression Model

In the regression problem it is assumed that there exists a mapping between inputs and output, denoted by $f(\mathbf{x})$, with an additive Gaussian noise

$$y = f(\mathbf{x}) + \varepsilon, \tag{1}$$

where ε is a zero mean Gaussian random variable with variance σ^2, that is, $\varepsilon \sim \mathcal{N}(\cdot|0, \sigma^2)$. If the mapping f is parameterized by $\mathbf{w} \in \mathbb{R}^D$, there exist a set of features $\boldsymbol{\phi}$ transforming the original input space to a new space, and the mapping is linear with respect to parameters, that is, $f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x})$, then such model is known as *linear regression model* [5].

The linear regression models have limiations because they require explicit form of the features and the number of parameters. Therefore, it would be beneficial to assume that the mapping f is unknown and try to induce it from data. In the probabilistic (Bayesian) framework it is accomplished by treating the mapping f as a latent variable that results in obtaining a flexible non-parametric regression model. In fact, this is the idea standing behind the *Gaussian processes* [19]. The final regression model is the following:

$$\begin{aligned} y &= f(\mathbf{x}) + \varepsilon \\ f &\sim \mathcal{GP}(\cdot|\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \\ \varepsilon &\sim \mathcal{N}(\cdot|0, \sigma^2) \end{aligned} \tag{2}$$

where $\mathcal{GP}$ denotes the Gaussian process, $\mu(\mathbf{x})$ is the mean function, $k(\mathbf{x}, \mathbf{x}')$ is the covariance function.

There are two components in the model to be determined: the covariance function and the mean function for categorical inputs. Both issues are crucial to allow calculating similarities between measurements' descriptions. In order to solve these problems we need to propose proper kernel function for the covariance function and a domain-specific expression for the mean function.

[1] Further, in the experiment, we will consider only three inputs ($D = 3$) which are typical in the diabetes treatment, namely, day of a week, period of a day, and a measurement code. However, the presented idea is given in a general case for any number of inputs.

Covariance Function for Categorical Inputs. The covariance function of two function values corresponding to the inputs $\mathbf{x}$ and $\mathbf{x}'$ is a kernel function[2] [5,19], denoted by $k(\mathbf{x}, \mathbf{x}')$. Our goal is to propose a proper kernel function for the categorical inputs in the context of diabetes. Here, we restrict ourselves to nominal variables, however, in general, there are many possible kernels for other types of symbolic data like strings, trees, and graphs [12]. We propose to apply the following kernel function for nominal inputs:

Proposition *(Covariance Function for Categorical Inputs).* *Let* $\mathbf{x}$ *be a vector of categorical inputs,* $x_d \in \mathcal{X}_d$, $\text{card}\{\mathcal{X}_d\} < \infty$, *for all* $d = 1 \dots D$ *and there is no ordering between the values, and* $\delta_d(\mathbf{x}, \mathbf{x}')$ *be the Kronecker's delta,*

$$\delta_d(\mathbf{x}, \mathbf{x}') = \begin{cases} 1, \, if \, x_d = x'_d, \\ 0, \, otherwise. \end{cases} \tag{3}$$

Then the following function

$$k(\mathbf{x}, \mathbf{x}') = \prod_{d=1}^{D} \delta_d(\mathbf{x}, \mathbf{x}') \tag{4}$$

is a valid kernel function.

Proof. First, let us prove that the Kronecker's delta is a kernel function. From the definition of kernel function we need to show that for any set $\{x_n\}_{n=1}^N$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} x_{d,i} \; \delta_d(\mathbf{x}_i, \mathbf{x}_j) \; x_{d,j} \geq 0.$$

The Kronecker's delta returns 1 if the two values are equal and 0 otherwise, thus we get a sum of squares of those objects which have equal values. The sum of squares for any objects is nonnegative that yields the Kronecker's delta is a valid kernel function.

Second, we need to prove that the product of Kronecker's deltas is also a valid kernel function. We use the fact that product of any valid kernels is also a kernel [5,20]. Hence, we get that the proposed kernel function for categorical inputs (4) is a valid kernel function. □

Our proposition of the kernel function could be presented in a simpler form as the Kronecker's delta for whole vectors $\mathbf{x}$ and $\mathbf{x}'$. However, we present the kernel in the given form (4) because of two reasons. First, we aim at distinguishing our proposition to the one proposed in [8] which is a sum of Kronecker's deltas (and kernels for continuous variables). Second, it is easier to interpret the product of the Kronecker's deltas for each input as a partition of the input space into single conjunctions of values.

[2] Kernel function is a symmetric function and the Gram matrix whose elements are given by $k(\mathbf{x}_n, \mathbf{x}_m)$ is positive semidefinite for any set $\{\mathbf{x}_n\}_{n=1}^N$ [20].

Mean Function. It is common practice to use Gaussian processes with a zero mean function [19]. However, explicit modelling of the mean function allows us to incorporate additional information about the considered phenomenon. In the case of diabetes and the categorical inputs we can take advantage on the character of the input space which is finite and propose different mean values for different combinations of values of selected or all inputs. In other words, the mean function can be parameterized as follows: for each combination of inputs' values a fixed nonnegative real number is assigned, namely

$$\mu(\mathbf{x}) = \mu_{\mathbf{x}}, \tag{5}$$

where $\mu_{\mathbf{x}} \in \mathbb{R}_+$ is a fixed value for given inputs $\mathbf{x}$.

The form of the proposed mean function requires to calculate as many values of mean as the cardinality of the input space which grows exponentially. However, we can limit the number of mean values by considering only selected inputs, *e.g.*, the code of the measurement, and do not include others in the calculations, *e.g.*, day of a week.

2.2 Prediction

Let us take a test measurement $\mathbf{x}_t$ for which we want to predict an output y_t. The similarities between the new observation $\mathbf{x}_t$ and the training examples $\mathbf{X}$ are defined by the kernel function $k(\mathbf{x}_t, \mathbf{x}_n)$ as in Eq. (4). We write $\mathbf{k}_t$ to denote the vector of covariances between the test point and the N training examples, and $\mathbf{K}$ is the kernel matrix for $\mathbf{X}$. According to the regression model in Eq. (2) we get the predictive distribution for given $\mathbf{x}_t$ with the following mean [19]:

$$\mu_t = \mu(\mathbf{x}_t) + \mathbf{k}_t^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu}), \tag{6}$$

where $\boldsymbol{\mu}$ is a vector of means for $\mathbf{X}$, and variance [19]:

$$\text{var}(\mathbf{x}_t) = k(\mathbf{x}_t, \mathbf{x}_t) - \mathbf{k}_t^\top (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{k}_t. \tag{7}$$

The predictive distribution is Gaussian distribution and thus the most probable value is chosen as the prediction, *i.e.*, $y_t = \mu_t$. Additionally, we can provide the uncertainty of the prediction which is equivalent to the variance $\text{var}(\mathbf{x}_t)$.

2.3 Learning

In practical applications, we need to determine a covariance function, a mean function and values of free parameters (*hyperparameters*), *e.g.*, the noise variance σ^2. While selection of the covariance function and mean function may be accomplished basing on the considered domain, setting the hyperparameters requires application of techniques of *model selection* [19]. We refer to the determination of values of the hyperparameters as *learning*.

Learning Mean Function. The mean function is parameterized by $\mu_{\mathbf{x}}$ for all possible combinations of values of selected inputs. Therefore, we propose to calculate mean values of outputs for all possible $\mathbf{x}$ as follows:

$$\mu_{\mathbf{x}} = \begin{cases} \dfrac{\sum\limits_{\mathbf{x}_n \in \mathbf{X}} \mathbb{1}\{\mathbf{x}_n = \mathbf{x}\} y_n}{\sum\limits_{\mathbf{x}_n \in \mathbf{X}} \mathbb{1}\{\mathbf{x}_n = \mathbf{x}\}}, & \text{if } \sum\limits_{\mathbf{x}_n \in \mathbf{X}} \mathbb{1}\{\mathbf{x}_n = \mathbf{x}\} > 0, \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

where $\mathbb{1}\{\cdot\}$ is the indicator factor.

Rationale behind the formula in Eq. (8) is that the prediction is made as a mean value of the same situations in the past. Such approach represents an assumption that patient's life is repeatable and her customs are essentially the same during a week. This is a manner how to incorporate *context* of daily routines into the model. On the other hand, the correlations among past observations are introduced by the covariance function.

Learning Covariance Function. In the literature, there are several approaches to model selection, *e.g.*, cross-validation, approximate methods like Laplace's Approximation, Variational Bayes, Expactation Propagation [19]. However, in this work we use the procedure based on the maximization of the marginal likelihood [5,19]. Once we have determined mean values using (8), we deal with one parameter only, *i.e.*, the variance of the noise σ^2. Let us denote the difference between outputs and means by $\bar{\mathbf{y}} = \mathbf{y} - \boldsymbol{\mu}$. Then the objective function is the log likelihood function in the following form:

$$\ln p(\mathbf{y}|\mathbf{X}, \sigma^2) = -\frac{1}{2}\bar{\mathbf{y}}^\top (\mathbf{K} + \sigma^2\mathbf{I})^{-1}\bar{\mathbf{y}} - \ln|\mathbf{K} + \sigma 2\mathbf{I}| - \frac{N}{2}\ln 2\pi, \tag{9}$$

where $|\cdot|$ denotes the determinant of a matrix. Next, we need to calculate derivative of (9) w.r.t. σ^2 which leads to the following equation:

$$\begin{aligned} \frac{\partial}{\partial\sigma^2}\ln p(\mathbf{y}|\mathbf{X}, \sigma^2) = &-\frac{1}{2}\bar{\mathbf{y}}^\top (\mathbf{K} + \sigma^2\mathbf{I})^{-1}\frac{\partial(\mathbf{K} + \sigma^2\mathbf{I})}{\partial\sigma^2}(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\bar{\mathbf{y}} + \\ &-\frac{1}{2}\mathrm{tr}\Big((\mathbf{K} + \sigma^2\mathbf{I})^{-1}\frac{\partial(\mathbf{K} + \sigma^2\mathbf{I})}{\partial\sigma^2}\Big), \end{aligned} \tag{10}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Notice that $\frac{\partial(\mathbf{K}+\sigma^2\mathbf{I})}{\partial\sigma^2} = \mathbf{I}$ which yields

$$\begin{aligned} \frac{\partial}{\partial\sigma^2}\ln p(\mathbf{y}|\mathbf{X}, \sigma^2) = &-\frac{1}{2}\bar{\mathbf{y}}^\top (\mathbf{K} + \sigma^2\mathbf{I})^{-1}(\mathbf{K} + \sigma^2\mathbf{I})^{-1}\bar{\mathbf{y}} + \\ &-\frac{1}{2}\mathrm{tr}\big((\mathbf{K} + \sigma^2\mathbf{I})^{-1}\big). \end{aligned} \tag{11}$$

Solving the optimization problem with the objective function given by (11) and without constraints requires inverting $(\mathbf{K} + \sigma^2\mathbf{I})$. The computational cost is proportional to $O(N^3)$ and hence gradient-based optimization techniques can be successfully applied for reasonably small N.[3]

[3] By *reasonably small* we mean up to $N = 1000$.

3 Experiment

To evaluate the Gaussian process regression for predicting the blood glucose level a real-life datasets are used [11]. The original data covers 70 patients. Diabetes patient records were obtained from two sources: an automatic electronic recording device and paper records. The automatic device had an internal clock to time stamp events, whereas the paper records only provided "logical time" slots (breakfast, lunch, dinner, bedtime). Each patient's medical history corresponds to a period from 20 to 149 days of measurements, depending on a patient.

3.1 Data Description

Original diabetes files consist of four information per record: (i) date, (ii) time, (iii) code (categorical), (iv) blood glucose level (numeric). The code describes the measurement, *e.g.*, regular insulin dose, pre-lunch glucose measurement, typical meal ingestion, typical exercise activity, and others (details can be found in [11]).

The original records were transformed into the following inputs: x_1 – day of a week, $\mathcal{X}_1$ consists of the following values: Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday, x_2 – part of a day, $\mathcal{X}_2$ consists of the following values: from 4:00 until 10:00, from 10:00 until 16:00, from 16:00 until 22:00, and from 22:00 until 4:00, x_3 – measurement code, $\mathcal{X}_3$ consists of 20 values, *e.g.*, insulin dose, measurement before breakfest.

In the experiment only 10 out of 70 patient records were used from which the smallest number of examples was 926 (116 days), and the biggest number was 1327 (149 days). The rest of records consist of too small number of observations to conduct statistically reliable experiments.

3.2 Experiment Details

Evaluation Metric. In order to evaluate the performance of the Gaussian process regression and compare it with other models we use the Mean Absolute Error (MAE). It is reported that this evalution metric is less sensitive to outliers in comparison to other metrics, *e.g.*, mean square error and root mean square error, and thus is preferred in forecast accuracy assessement [15].

Predictive Models. In the experiment, the prediction of the blood glucose level is made according to the following models:

1. *Mean Prediction* (MPred) is a model which always returns $y_{MP} = \frac{1}{N}\sum_{n=1}^{N} y_n$ as a prediction. This model is a baseline for comparing models in the experiment.
2. *Classification and Decision Tree* (CART) is a regression model which can be used for symbolic inputs. In this approach the input space is recursively partitioned due to given criterion which results in a tree-structured model. At each leaf a mean value of objects covered by decision criteria at each node up to the root. For details see [6].

3. *Gaussian process regression* (GP) model with the covariance function in the form (4), and the mean function (5) calculated only for the input x_3, that is, the measurement code.[4] This assumption results in 20 mean values to be determined. The prediction for new object is made according to (6).

Experiment Setting. The considered data in the experiment consists of 10 datasets representing different patients' records. Each dataset formulates time series. We decided to fix test set to contain last 300 examples in the time series (the most recent examples). However, the training set consists of varying number of observations equal 100, 200, 300, 400, 500, and 600. This aspect allows us to analyze the sensitivity of the considered models to different number of observations. Additionally, we use 100 examples before training set as a validation set to determine the noise parameter σ^2.

Implementation Details. The experiment was carried out in MATLAB environment. The GP regression model with the proposed covariance function and mean function were implemented in MATLAB. For CART model the built-in MATLAB implementation was used. In order to determine the hyperparameter σ^2 MATLAB optimization function was used with the objective function given by (11) which was calculated basing on data included in the validation set.

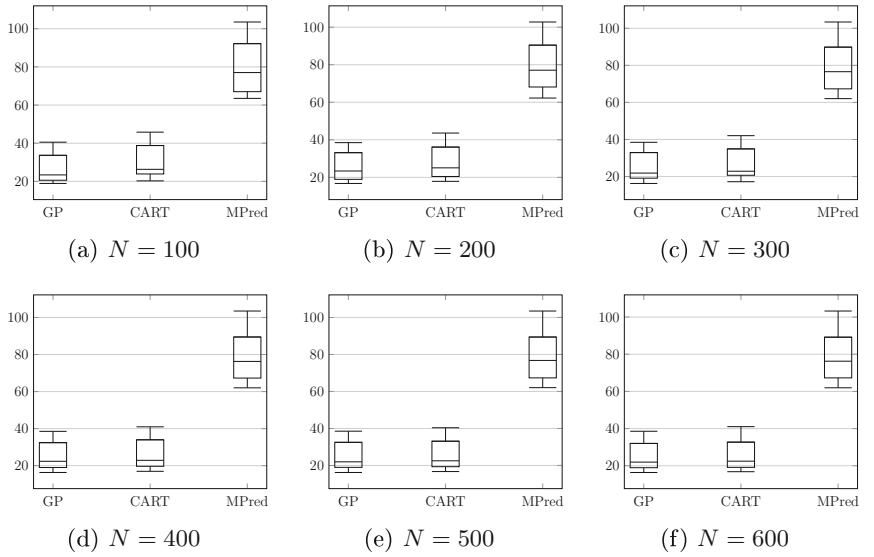
3.3 Results and Discussion

The results of the prediction of the blood glucose level averaged over 10 patients are presented as boxplots in Figs. 1 and 2. In the Fig. 1 the predictive models are compared for different size of the training set. In the Fig. 2 detailed performance on all predictive models are presented as a function of varying number of the training examples.

CART and Gaussian process regression performed significantly better than the baseline. CART and GP obtained mean MAE at the level of 23–27 mg/dl (see Fig. 2(a)), and 21.5-23 mg/dl (see Fig. 2(b)), respectively, whilst MPed – 77–78 mg/dl (see Fig. 2(c)). GP model behaved more stable for varying size of the training set than CART and it was enough to have $N = 300$ of training examples to obtain the best predictive accuracy (see Fig. 1). Moreover, for $N = 100$ and $N = 200$ the GP model performs only slightly worst (about 1.5 mg/dl) than for N greater than 300 (see Fig. 2(a)).

There are two main conclusions following from the experiment. First, both CART and GP models achieved good prediction accuracy and thus could be succesfully used in real-life applications. Second, the GP regression obtained better results than CART for smaller training samples. This issue is especially important from the practical point of view because of smaller memory requirements and lower computational costs. These aspects are crucial in modern eHealth systems [2], *e.g.*, as mobile services [13].

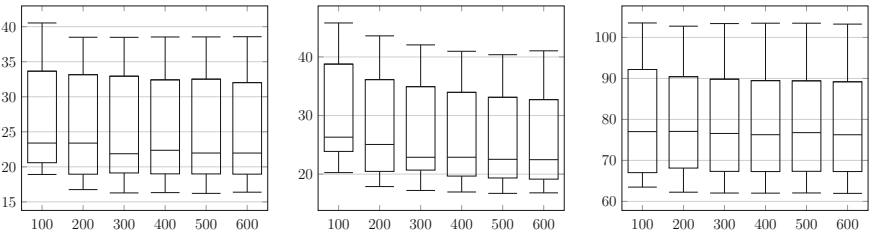
[4] We have omitted the day of a week and the part of a day because of two reasons. First, we wanted to have less parameters of the mean function. Second, in the preliminary experiments, including also x_1 and x_2 resulted in no significant change in the performance of the GP.



(a) $N = 100$ (b) $N = 200$ (c) $N = 300$

(d) $N = 400$ (e) $N = 500$ (f) $N = 600$

Fig. 1. Boxplots representing comparison between methods using MAE evaluation metric for changing number of training examples N. GP stands for Gaussian process regression, CART – Classification and Regression Tree, MPred – prediction with mean value basing on training examples.



(a) Gaussian process (b) CART (c) Mean prediction

Fig. 2. Boxplots for each method with changing number of training examples (x-axis) and MAE evaluation metric (y-axis).

4 Conclusion

In this paper, we have presented the model based on Gaussian processes for the prediction of the blood glucose level. Considering the specificity of the problem, *i.e.*, the symbolic character of inputs, the covariance function and the mean function have been proposed. The learning of the hyperparameter, *i.e.*, mean values and the variance of the noise, has been presented using maximization of the marginal likelihood. At the end, the experiment with 10 real-life datasets has been conducted. The results indicate high predictive accuracy of the proposed

approach (see Figs. 1 and 2). Moreover, our model can be easily implemented in mobile eHealth systems and this would be a focus of our future work.

Acknowledgements. The research is partially supported by the grant co-financed by the Ministry of Science and Higher Education in Poland.

References

1. Agresti, A.: An Introduction to Categorical Data Analysis. Wiley-Interscience, New York (2007)
2. Alemdar, H., Ersoy, C.: Wireless sensor networks for healthcare: a survey. Comput. Netw. **54**(15), 2688–2710 (2010)
3. Ažman, K., Kocijan, J.: Application of Gaussian processes for black-box modelling of biosystems. ISA Trans. **46**, 443–457 (2007)
4. Billard, L., Diday, E.: From the statistics of data to the statistics of knowledge: symbolic data analysis. J. Am. Stat. Assoc. **98**(462), 470–487 (2003)
5. Bishop, C.: Pattern Recognition and Machine Learning. Elsevier, Amsterdam (2006)
6. Breiman, L., Friedman, J., Olshen, R., Stone, C., Steinberg, D., Colla, P.: CART: Classification and Regression Trees. Wadsworth, Belmont (1983)
7. Chu, W., Ghahramani, Z., Falciani, F., Wild, D.: Biomarker discovery in microarray gene expression data with Gaussian processes. Bioinforma **21**(16), 3385–3393 (2005)
8. Daemen. A., De Moor, B.: Development of a kernel function for clinical data. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2009), pp. 5913–5917. IEEE (2009)
9. De Gaetano, A., Arino, O.: Mathematical modelling of the intravenous glucose tolerance test. J. Math. Biol. **40**, 136–168 (2000)
10. Fischer, I., Meinl, T.: Graph based molecular data mining - an overview. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 5, pp. 4578–4582. IEEE (2004)
11. Frank, A., Asuncion, A.: UCI machine learning repository (2010). http://archive.ics.uci.edu/ml
12. Gärtner, T.: A survey of kernels for structured data. ACM SIGKDD Explor. Newsl. **5**(1), 49–58 (2003)
13. Grzech, A., Juszczyszyn, K., Swiatek, P., Mazurek, C. Sochan, A.: Applications of the future internet engineering project. In: International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel & Distributed Computing (SNPD), pp. 635–642. IEEE (2012)
14. Huang, Z.: Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Min. Knowl. Discov. **2**(3), 283–304 (1998)
15. Hyndman, R., Koehler, A.: Another look at measures of forecast accuracy. Int. J. Forecast **22**(4), 679–688 (2006)
16. Iannario, M.: Preliminary estimators for a mixture model of ordinal data. Adv. Data Anal. Classif. **6**, 163–184 (2012)
17. Likar, B., Kocijan, J.: Predictive control of a gas-liquid separation plant based on a Gaussian process model. Comput. Chem. Eng. **31**, 142–152 (2007)
18. Makosso-Kallyth, S., Diday, E.: Adaptation of interval PCA to symbolic histogram variables. Adv. Data Anal. Classif. **6**, 1–13 (2012)

19. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. MIT Press, London (2006)
20. Shawe-Taylor, J., Cristianini, N.: Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge (2004)
21. Srinivasan, A., King, R.D.: Feature construction with inductive logic programming: a study of quantitative predictions of biological activity aided by structural attributes. Data Min. Knowl. Discov. **3**(1), 37–57 (1999)
22. Tomczak, J., Gonczarek, A.: Decision rules extraction from data stream in the presence of changing context for diabetes treatment. Knowl. Inf. Syst. **34**, 521–546 (2013)
23. Tomczak, J., Świątek, J., Latawiec, K.: Gaussian process regression as a predictive model for quality-of-service in web service systems. arXiv preprint arXiv: 1207.6910 (2012)
24. Turner, R., Deisenroth, M.P., Rasmussen, C.E.: System identification in Gaussian process dynamical systems. In: Görür, D. (ed.) NIPS Workshop on Nonparametric Bayes. Whistler, Canada (2009)
25. Węglarz-Tomczak, E., Vassiliou, S., Mucha, A.: Discovery of potent and selective inhibitors of human aminopeptidases erap. 1 and erap. 2 by screening libraries of phosphorus-containing amino acid and dipeptide analogues. Bioorg. Med. Chem. Lett. **26**(16), 4122–4126 (2016)
26. World Health Organization. Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia. Report of a WHO/IDF Consultation (2006)
27. Zięba, M., Świątek, J.: Ensemble classifier for solving credit scoring problems. IFIP AICT **372**, 59–66 (2012)