DOI:10.3969/j.issn.1005-202X.2016.04.012

医学信号处理与医学仪器

基于支持向量机和自回归积分滑动平均模型组合的血糖值预测

余丽玲[1],陈婷[2],金浩宇[1],徐彬锋[1]
1. 广东食品药品职业学院,广东 广州 510520;2. 广东省食品药品职业技术学校,广东 广州 510663

【摘 要】根据动态血糖监测系统采集糖尿病患者血糖值,有效预测血糖值是治疗糖尿病的前提。为了预测糖尿病患者未来一段时间内的血糖值,本文根据最小方差将支持向量机(SVM)和自回归积分滑动平均(ARIMA)进行组合得到新的预测模型。为了验证本文方法的有效性,采用多组临床实验数据进行实验,同时对比ARIMA模型、SVM模型、神经网络模型结果。实验结果表明本文方法预测血糖值精度明显提高,弥补单一预测模型方法的不足,发挥了两种模型各自优势。

【关键词】动态血糖监测系统;糖尿病;血糖值;支持向量机;自回归积分滑动平均;最小方差
【中图分类号】R319;R197.39 【文献标识码】A 【文章编号】1005-202X(2016)04-0381-04

Prediction of blood glucose level based on model combining support vector machine and autoregressive integrated moving average

YU Li-ling[1], CHEN Ting[2], JIN Hao-yu[1], XU Bin-feng[1]
1. Guangdong Food and Drug Vocational College, Guangzhou 510520, China; 2. Guangdong Food and Drug Vocational and Technical School, Guangzhou 510663, China

Abstract: The continuous glucose monito ring system (CGMS) is used to collect the blood glucose level of diabetics. Effectively predicting the blood glucose level is the premise for the treatment of diabetes. A prediction model combining support vector machine (SVM) and autoregressive integrated moving average (ARIMA) was proposed in the paper to predict the blood glucose level over a period of time. Multiple sets of clinical trial data were processed for verifying the effectiveness of proposed method and comparing the results of ARIMA model, SVM model, neural network model. The experimental result showed the precision of the propose method for predicting the blood glucose level increased significantly, covering the shortage of signal prediction model and taking the advantage of two models.

Key words: continuous glucose monitoring system; diabetes; blood glucose level; support vector machine; autoregressive integrated moving average; minimum variance

前 言

随着经济发展和生活水平提高,糖尿病患者越来越多。长期的糖尿病会引起一系列并发症,甚至会带来生命危险。如何及时预测,有效控制血糖波动,是治疗糖尿病的主要任务。一个良好的血糖预测算法不仅可以减少低血糖或者高血糖发生,同时还可结合胰岛素泵使用,调节胰岛素剂量以及寻找最优的开泵时间,使得糖尿病患者的血糖值控制更加准确。目前,动态血糖监测系统(Continuous Glucose Monitoring System, CGMS)作为一种新的血糖检测手段,已广泛应用于临床[1]。CGMS是一种通过葡萄糖感应器监测皮下组织间液葡萄糖值而反映血糖水平的监测技术,能够全面、详细地显示体内血糖变化及波动趋势[2]。研究学者发现利用CGMS采集的数据进行未来一段时间的血糖预测,可以有效降低高血糖或低血糖发生的可能性,因此血糖预测已成为治疗糖尿病的一个研究热点。现有的血糖预测策略主要分为两个方向:一种是基于生理模型的预测;另一种是基于数据的预测[3]。由于人体生理机制比较复杂,影响血糖的因素很多,很难建立精准的血糖预测模型,同时研究发现复杂的人体生理血糖预测模型,会导致预测的血糖具有一定的时延。常用的血糖预测方法是基于数据的预测,它是完全基于历史血糖数据,不考虑机体的先验知识,通过建立数学模型预测未来一段时间血糖。现有的预测方法主

【收稿日期】2015-11-18
【基金项目】广东省科技计划项目(2015A020214016)
【作者简介】余丽玲(1988-),广东韶关人,助教,研究方向:动态血糖监测,E-mail: lilingyu124@126.com。

要有自回归模型[4]、支持向量机[5-6]、人工智能神经网络[7]、极限学习机算法[8]等，这些方法都能预测血糖变化趋势，但预测的精准性及普适性还有待提高。针对糖尿病患者血糖数据的复杂性与不稳定性，本文提出一种支持向量机（Support Vector Machines, SVM）和自回归积分滑动平均模型（Autoregressive Integrated Moving Average Model, ARIMA）相组合（SVM-ARIMA）的血糖值预测方法。

1 SVM-ARIMA预测模型

1.1 SVM模型

SVM模型是一种建立在VC维理论和结构风险最小原理基础上的机器学习算法，在解决小样本、非线性等问题上具有十足的优点。

假设给定样本数据集$\{(x_i,y_i),i=1,2,\ldots,k\}$，$x_i$为输入的训练样本，$y_i$为输出数据，$k$为样本大小。由于大部分样本数据呈非线性关系，使用非线性函数$\phi(x)$将原输入空间的x_i映射到某一高维空间，然后在高维空间进行线性回归，以获得原空间非线性回归效果。则有：

$$f(x)=w^T\phi(x)+b \tag{1}$$

其中，w为权值向量，b为偏差。

SVM回归一般采用ε不敏感损失函数来度量经验风险，该函数的定义为[9]：

$$L_\varepsilon=\begin{cases}|f(x)-y|-\varepsilon; & |f(x)-y|\geq\varepsilon\\ 0 & ;|f(x)-y|<\varepsilon\end{cases} \tag{2}$$

在考虑训练w和经验风险最小化原则下，可将SVM回归问题转化为约束优化求解问题：

$$\begin{aligned}&\min\frac{1}{2}\|w\|^2+C\sum_{i=1}^{n}\xi_i^2\\&s.t.y_i=w^T\phi(x_i)+b+\xi_i\end{aligned} \tag{3}$$

其中C为惩罚参数，其中ξ_i,ξ_i^*为松弛变量。

最后，SVM回归估计函数表达式为[10]：

$$f(x)=\sum_{i=1}^{l}(a_i-a_i^*)k(x,x_i)+b \tag{4}$$

其中，函数$f(x)$由a_i,a_i^*参数决定，$k(x,x_i)$为核函数，不同核函数可构造出不同SVM。考虑到糖尿病患者血糖值变化特点都不一样，本文采取局部优化算法，自适应寻找最优惩罚系数C、不敏感损失函数参数及核函数。本文方法采用以下4种核函数进行辨识建模预测，它们的表现形式为：

（1）线性内核：$k(x_i,x_j)=x_i\cdot x_j$；

（2）多项式内核：$k(x,x_i)=[(x_i\cdot x_j)+c]^d$；

（3）径向基（RBF）内核：

$k(x_i,k_j)=\exp(-\|x-y\|^2/2\sigma^2)$；

（4）S形内核：$k(x_i,x_j)=\tan ch(g*x_i*x_j+c)$。

其中，d为多项式阶数，c为补偿参数，g为S形内核斜率。

1.2 ARIMA预测模型

ARIMA模型是一种常用时间序列预测方法，适用于解决线性问题。ARIMA模型由自回归、差分、滑动平均三部分组成，其中AR代表自回归，I代表使非平稳序列转化为平稳序列的差分运算，MA代表滑动平均。该模型常记作ARIMA（p,d,q），其中p、d、q分别表示自回归阶数、差分阶数、移动平均阶数。ARIMA模型的数学表达式为：

$$\nabla^d x_t=\sum_{i=1}^{p}\varphi_i\nabla^d x_{t-i}+w_t+\sum_{j=1}^{q}\theta_j x_{t-j} \tag{5}$$

其中，x_t是血糖数据组成的时间序列，φ_i是自回归项的参数，θ_j是滑动平均项的参数，w_t为高斯白噪声序列。∇^d用于对血糖数据进行差分处理，它的数学表达式为：

$$\nabla^d x=(x_t-x_{t-1})-(x_{t-1}-x_{t-2})-\ldots-(x_{t-d+1}-x_{t-d}) \tag{6}$$

ARIMA处理过程包括以下步骤：（1）糖尿病患者血糖是在不断波动，需要对采集的血糖数据进行差分处理，使其转换为平稳化时间序列。（2）通过自相关函数和偏自相关函数确定（p, d, q）的选取，从（1, 1, 1）逐渐递增尝试，并结合最小信息准则确定最合适的模型[11]。（3）估计参数，并检验参数的显著性以及ARIMA模型合理性；如果检验通不过，则继续识别。（4）利用选中的预测模型，预测未来血糖值。

1.3 SVM-ARIMA模型

SVM模型通过运用核函数，较好地解决了时间序列的非线性等问题，同时SVM回归模型还结合了一定的人体生理模型。对于线性时间序列，ARIMA模型的预测简单精准。由于糖尿病患者血糖值具有非线性、时变性和模糊性特征，单纯使用SVM或者ARIMA模型进行血糖值预测都有可能导致误差过大。本文利用最小方差原则将SVM和ARIMA模型组合起来预测血糖值。该组合模型SVM-ARIMA集结了SVM和ARIMA模型所包含的信息，既发挥了SVM、ARIMA的优点，又弥补了各自不足，同时其误差方差小于任一分量的误差方差，从而改善了预测能力。

假设$\hat{x}_1$为SVM预测值，$\hat{x}_2$为ARIMA预测值，两种预测方法的误差值分别为e_1和e_2，取β_1和β_2为相应的加权系数，而且$\beta_1+\beta_2=1$，则SVM-ARIMA模型的预测值$\hat{x}_t$为：

$$\hat{x}_t=\beta_1\hat{x}_1+\beta_2\hat{x}_2 \tag{7}$$

SVM-ARIMA模型的预测误差$e=\beta_1e_1+\beta_2e_2$，则误差的方差为：

$$D(e)=\beta_1^2D(e_1)+\beta_2^2D(e_2)+2\beta_1\beta_2\operatorname{cov}(e_1,e_2) \tag{8}$$

其中，$\mathrm{cov}(e_1,e_2)$ 为协方差，对 $D(e)$ 求极小值，可得到 β_1：

$$\beta_1=\frac{D(e_1)-\mathrm{cov}(e_1,e_2)}{D(e_1)+D(e_2)-2\,\mathrm{cov}(e_1,e_2)} \tag{9}$$

2 实 验

本文方法采用MATLAB编程实现，为验证所提出方法的有效性，对20位糖尿病患者的血糖进行了预测。由于CGMS系统采集的血糖数据受到皮下传感器稳定性和外部因素的影响，存在着噪声干扰，因此需要对历史的血糖数据进行平滑滤波预处理。卡尔曼滤波能有效地减少信号中的噪声，同时滤波后的数据变化平稳，与实际值吻合性较好，而且能够减少预测值和实际值之间的时延[12-14]。本文采用卡尔曼滤波对采集的血糖数据进行预处理。

CGMS系统每天可记录288个葡萄糖值，一般能够监测人体72 h内动态血糖变化。本文任意选取每位患者24 h血糖数据，其中前228个血糖数据作为训练样本，对血糖数据进行数学建模，采用最后60个作为测试样本。图1显示其中1位糖尿病患者24 h血糖数据。图2显示采用本文方法、ARIMA模型、SVM三种模型对该患者最后60个血糖值进行预测。从图2可看到，本文方法预测的血糖值更贴近真实值，预测准确性优于其它两种模型。

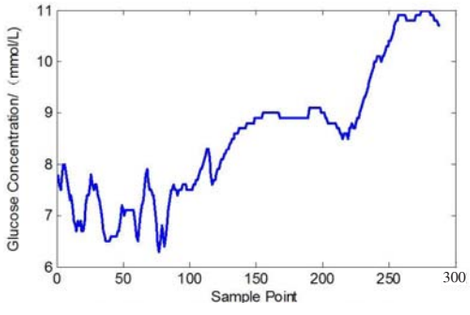


图1 24 h采集的病人血糖浓度

Fig.1 Glucose concentration of patient in 24 h

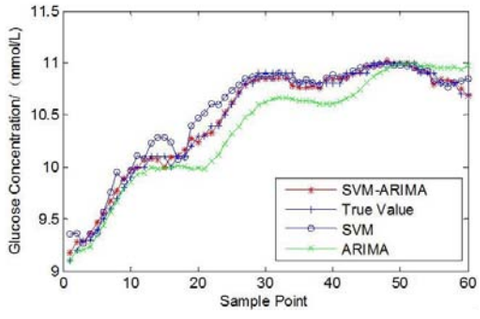


图2 比较3种方法预测的血糖值

Fig.2 Comparison of blood glucose level predicted by three methods

SVM: Support vector machine; ARIMA: Autoregressive integrated moving average

本文采用平均绝对相对误差（Mean Absolute Error，MAE）和均方误差（Mean Squared Error，MSE）两个性能指标评价预测结果。

$$MAE=\frac{1}{N}\sum_{k=1}^{N}|\hat{x}(k)-x(k)| \tag{10}$$

$$MSE=\frac{1}{N}\sum_{k=1}^{N}(\hat{x}(k)-x(k))^2 \tag{11}$$

其中，$\hat{x}(k)$ 为预测的血糖值，N 为血糖数据长度。

神经网络模型是目前预测血糖值比较理想的方法[15]。本文将ARIMA方法、SVM方法、神经网络方法作为基准方法与本文方法进行比较。图3显示采用神经网络模型、ARIMA模型、SVM-ARIMA三种方法对其中4位病人进行血糖值预测，从图上可以看到，3种预测方法都能够较好地把握血糖的变化趋势，但神经网络模型、ARIMA模型预测的血糖值精度不高。同时实验发现，神经网络模型难以找到最优的隐含层数，存在过拟合和局部极小值。ARIMA模型比较适合于血糖值变化平稳的预测，血糖波动大的预测精度会下降。图4对比了4种方法预测10位病人血糖值的精度，从结果可以看到本文方法的MAE和MSE值都小于其它3种方法，说明SVM-ARIMA血糖预测精度要高于单一模型，同时本文方法的血糖预测精度要高于神经网络模型，是一种有效的血糖预测方法。

3 结 论

由于糖尿病患者血糖的非线性、复杂性与不稳定性，本文利用CGMS系统提供的血糖数据，提出一种融合SVM模型和ARIMA模型的组合预测模型SVM-ARIMA进行血糖预测。该模型结构简单，融合了SVM模型和ARIMA模型的预测优点，能够较好地模拟血糖值中的非线性规律以及线性趋势。将本文方法与ARIMA、SVM、神经网络方法进行对比，SVM-ARIMA模型能较好地反映血糖波动趋势，显著提高血糖预测精度，是一种有效的血糖预测方法。

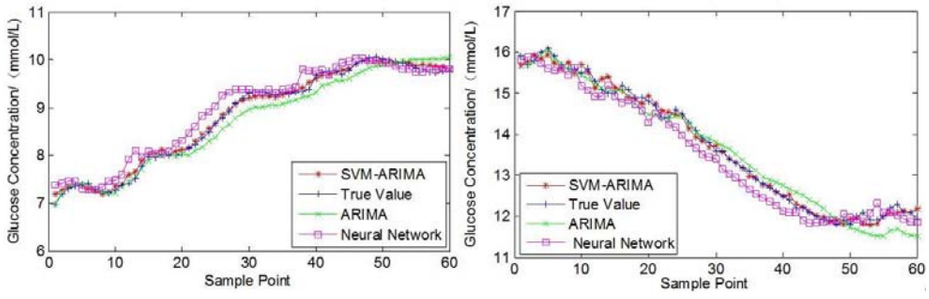
【参考文献】

[1] HERMANIDES J, PHILLIP M, DEVRIES J H. Current application of continuous glucose monitoring in the treatment of diabetes [J]. Diabetes Care, 2011, 34(2): 197-201.

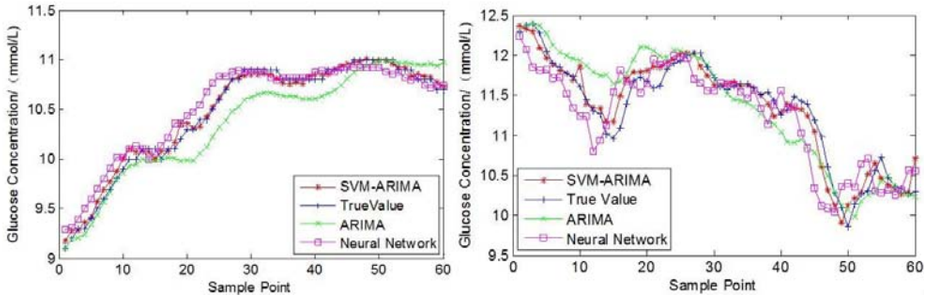
[2] REBRIN K, STEIL G M, *van* ANTWERP W P, et al. Subcutaneous glucose predicts plasma glucose independent of insulin: implications for continuous monitoring[J]. Am J Physiol, 1999, 277(3 Pt 1): E561-E571.

[3] 吴向伟. 数据驱动的血糖预测方法及其应用[D]. 北京: 北京化工大学, 2013.
WU X W. Data-driven blood glucose prediction Algorithms and their application [D]. Beijing: Beijing University of Chemical Technology, 2013.

[4] GANI A, GRIBOK A V, RAJARAMAN S, et al. Predicting subcutaneous glucose concentration in humans: data-driven glucose modeling[J]. IEEE Trans Biomed Eng, 2009, 56(2): 246-254.



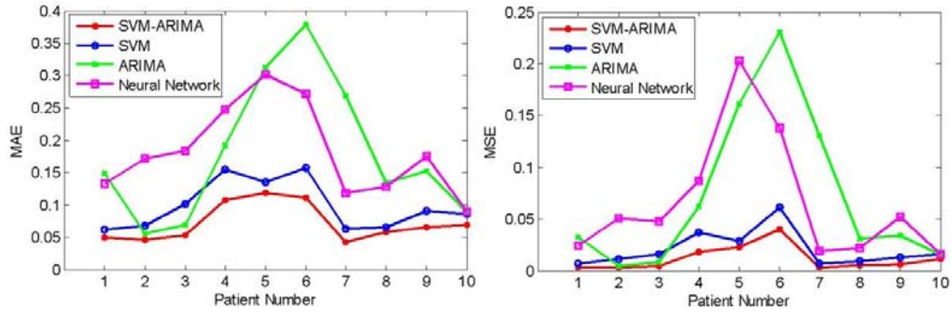
a: Case 1 b: Case 2



c: Case 3 d: Case 4

图3 不同方法血糖值预测效果比较

Fig.3 Comparison of predicting performance of different methods



a: Mean absolute error (MAE) b: Mean squared error (MSE)

图4 不同血糖预测方法的精度

Fig.4 Precision of different methods for predicting blood glucose level

［5］ WANG Z, LAI L, XIONG D, et al. Study on predicting method for acute hypotensive episodes based on wavelet transformand support vector machine［C］. International Conference on Biomed Engineering and Informatics, 2010: 1041-1045.

［6］ GEORGA E I, PROTOPAPPAS V C, ARDIGO D, et al. Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression［J］. IEEE J Biomed Health Inform, 2013, 17(1): 71-81.

［7］ ZECCHIN C, FACCHINETTI A, SPARACINO G, et al. A new neural network approach for short-term glucose prediction using continuous glucose monitoring time-series and meal information［C］. Engineering in Medicine and Biology Society, 2011 Annual International Conference of the IEEE, 2011: 5653-5656.

［8］ 莫雪. 数据驱动的血糖预测方法研究［D］. 北京: 北京化工大学, 2014.
MO X. Study on data-drive glucose prediction methods［D］. Beijing: Beijing University of Chemical Technology, 2014.

［9］ 蒋喆. 支持向量机在电力负荷预测中的应用研究［J］. 计算机仿真, 2010, 27: 282-285.
JIANG Z. Research power load forecasting baseon support vector machines［J］. Computer Simulation, 2010, 27: 282-285.

［10］邵忻. 一种新的基于ARIMA-SVM 网络流量预测研究［J］. 计算机应用研究, 2012, 29(5): 1901-1903.
SHAO X. Application of ARIMA-SVM model in network traffic prediction［J］. Application Research of Computers, 2012, 29(5): 1901-1903.

［11］YE R, SUGANTHAN P N, SRIKANTH N, et al. A hybrid ARIMA-DENFIS method for wind speed forecasting［C］. IEEE International Conference on Fuzzy Systems, 2013: 1-6.

［12］KNOBBE E J, BUCKINGHAM B. The extended kalmanfilter for continuous glucose monitoring［J］. Diabetes Technol Ther, 2005, 7 (1): 15-27.

［13］王延年, 雍永强, 贾晓灿, 等. 融合ARIMA和REFNN的血糖预测［J］. 计算机工程与设计, 2015, 36(6): 1652-1656.
WANG Y N, YONG Y Q, JIA X C, et al. Prediction of blood glucose fusing ARIMA and RBFNN［J］. Computer Engineering and Design, 2015, 36(6): 1652-1656.

［14］FACCHINETTI A, SPARACINO G, COBELLI C. An online self-tunable method to denoise CGM sensor data［J］. IEEE Trans Biomed Eng, 2010, 57(3): 634-641.

［15］PAPPADA S M, CAMERON B D, ROSMAN P M, et al. Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes［J］. Diabetes Technol Ther, 2011, 13(2):135-141.

（编辑：黄开颜）