Савченко Е.А.

ЭКСПРЕСС-ПРОГНОЗ УРОВНЯ ГЛЮКОЗЫ В КРОВИ ПО КОМБИНАТОРНОМУ АЛГОРИТМУ МГУА

Рассматривается задача нахождения краткосрочного экспресс-прогноза по малой выборке данных для задачи домашнего мониторинга диабета. Сравнивается три варианта прогноза: по временному, аналоговому и по смешанному аналогово-временному рядам данных. Учет аналогов повышает точность и смещение прогнозирующей модели почти в два раза. Рассчитаны прогнозы уровня глюкозы в крови на целый день, соответствующие четырем измерениям в определенное время дня.

Введение

Решается задача прогноза состояния больного диабета по данным мониторинга, который ведет сам больной в домашних условиях, следуя определенной инструкции врача. Поэтому данные зачастую содержат много пропусков и составить представительную выборку данных несколько затруднительно [1]. В работе поставлена цель, получить прогноз уровня глюкозы в крови больного по достаточно короткой выборке данных, имея только шесть дней наблюдения за больным, или найдя всего пять аналогов текущего состояния в предыстории болезни.

Методика моделирования

Для нахождения прогноза применен комбинаторный алгоритм МГУА, который выполняет полный перебор постепенно усложняющихся структур моделей и, оценивая их по внешнему критерию, выбирает модель, дающую минимум этого критерия. В качестве внешнего критерия используется критерий регулярности, вычисляемый по формуле:

$$ER = RR_{B/A} = \frac{\sum_{i=1}^{n_B} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y_i})^2} \longrightarrow \min \qquad (1)$$

где $RR_{B/A}$ – критерий регулярности для модели, коэффициенты которой получены на обучающей выборке А, а точность рассчитана на проверочной выборке В, n_B – количество строк в проверочной выборке данных, y_i – действительное значение выходной величины, $\hat{y}_i$ – значений выходной величины, рассчитанное по модели, $\overline{y_i}$ - среднее значение выходной величины.

Полином с наилучшим значением внешнего критерия является полиномиальной моделью оптимальной сложности. Часто бывает, что таких наиболее точных моделей оказывается несколько [2], тогда используется доопределение модели по перекрестному критерию смещения (Cross-Bias Criterion). Расчет его подобен расчету перекрестного критерия ошибки (Cross-Validation), в котором каждая строка по очереди исключается из

обучающей выборки, а остальные представляют собой тестовую выборку, по которой рассчитывается ошибка. Аналогично для перекрестного критерия смещения, строки, по одной, исключаются из выборки данных, и на них рассчитывается квадрат ошибки. При числе строк выборки равном n, получим столько же квадратов ошибки на одной строке: er_1^2, er_2^2, … ,er_n^2. Далее рассчитываем среднее значение квадратов ошибки на всех остальных строках, кроме одной, исключенной. По формуле (2) получим n значений квадратов средней ошибки ER_1^2, ER_2^2, . . . , ER_n^2.

$$\overline{ER_i}^2 = \frac{1}{N}\sum_{i=1}^{N} er_i^2 \quad , \; i = 1, 2, \ldots, \text{n.} \qquad (2)$$

Разность квадрата ошибки на одной строке и среднего значения квадратов ошибки на всех остальных строках должна быть минимальной, как показано в формуле (3). Эта разность и будет смещением модели.

$$BS^2 = \frac{1}{N}\sum_{i=1}^{N}(er_i^2 - \overline{ER}^2) \longrightarrow \min \text{ , где i = 1, 2, …, n.} \qquad (3)$$

По минимум критерия смещения и выбирается единственная оптимальная модель.

В статье рассматривается три варианта прогноза. Первый по временному ряду, когда данные взяты из исходной выборки данных, последовательно за шесть дней наблюдения, и по ним находится прогноз уровня глюкозы в крови на утро следующего дня. Второй по данным аналогового ряда, когда наблюдения взяты из предыстории одного больного. Аналогом называется объект, характеристический вектор (состав аргументов) которого близок к характеристическим векторам других объектов в пространстве всех измеряемых признаков. Выбраны недели, близкие к текущей в пространстве признаков. Каждая неделя описывается характеристическим вектором, который содержит средние за неделю значения всех переменных.

Расстояние вычисляется как квадрат евклидовой нормы вектора отклонений в пространстве всех m переменных, указанных в исходной выборке данных:

$$L_{ij}^2 = (x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \ldots + (x_{mi} - x_{1mj})^2.$$

Таким образом, выбраны недели, аналоги текущей, в пространстве m признаков, и из каждой недели взят один день наблюдения, предшествующий дню прогноза. По ним так же найдено значение уровня глюкозы в крови больного в тот же день, что и в первом прогнозе.

В классическом регрессионном анализе для достижения точности результата оценивания параметров рекомендуется, чтобы число наблюдений или реализаций стационарного процесса, т.е. число строк исходной выборки данных, было в три и более раз больше числа наблюдений переменных [3]. Усреднение оценок коэффициентов модели начинается при числе строк превышающем число переменных.

Третий вариант прогноза, это смешанный аналогово-временной ряд данных, который содержит данные нескольких дней нескольких аналогичных друг другу недель. Для прогноза используется только две строки наблюдений для двух аналогов, т.е. недель, наиболее близких к текущей неделе наблюдения. Составлена выборка, содержащая шесть строк по два наблюдения из каждой недели. Рассчитан прогноз уровня глюкозы в крови утром следующего дня.

Допустимость применения такого алгоритма можно объяснить так. Модели, получаемые как по временному, так и по аналоговому ряду, имеют одну и ту же выходную переменную, и одно и тоже множество аргументов-кандидатов. Следовательно, условные уравнения, получаемые по указанным двум способам, имеют одинаковую природу.

Описание задачи

Характеристический вектор пациента включает в себя все показатели, регистрируемые в карте больного. Здесь указывается пол больного, его возраст, рост, вес и тому подобные показатели [2]. В характеристический вектор больного входят также четыре значения уровня глюкозы, измеряемые за день: x_{6k} - в 6 часов, перед завтраком, x_{12k} - в 12, перед обедом, x_{17k} - в 17 часов, и x_{22k} - перед отходом ко сну в 22 часа, а также дозы инсулина, назначенные больному в течение суток: x_{R1k} – доза инсулина утром в 6 часов, x_{R2k} – доза инсулина в 12 часов, x_{R3k} – доза инсулина в 17 часов, x_{R4k} – доза инсулина перед сном в 22 часа. По всем этим параметрам, из всех данных мониторинга больного диабетом за 6 месяцев, выбраны 42 дня болезни, которые использовались для построения трех выборок данных, содержащих по 6 строк наблюдения.

Исходная выборка данных представлена в таблице 1. В ней приведены данные шести недель мониторинга одного больного с 27 июля по 6 сентября 1993 года.

Как уже говорилось, приводятся три варианта прогноза. Первый во временной области, т.е. прогноз по шести дням наблюдения с 31 августа по 5 сентября. Необходимо найти уровень глюкозы в крови утром 6 сентября. Второй в аналоговой области, т.е. по шести дням, выбранным из недель, аналогов текущей недели наблюдения. Из каждой недели взят один день наблюдения, воскресенье. Нужно найти прогноз уровня глюкозы в

крови утром 6 сентября. Третий прогноз в смешанной аналогово-временной области, т.е. для прогноза взяты шесть строк, по две строки для трех аналогов, т.е. недель, близких к наблюдаемой в пространстве всех признаков. Необходимо также найти прогноз уровня глюкозы утром 6 сентября.

Таблица 1. Исходные данные наблюдения одного пациента за 42 дня наблюдения.

	№	Дата	День Недели	$x_{6,k+1}$	$x_{12,k+1}$	$x_{17,k+1}$	$x_{22,k+1}$	x_{6k}	x_{12k}	x_{17k}	x_{22k}	x_{R1k}	x_{R2k}	x_{R3k}	x_{R4k}
5-й аналог	1	27.07.93	Вт	18,20	13,80	11,90	8,20	6,30	12,60	7,80	8,40	17	9	11	16
	2	28.07.93	Ср	4,70	5,80	10,30	8,10	18,20	13,80	11,90	8,20	17	9	13	16
	3	29.07.93	Чт	5,60	10,00	13,10	3,70	4,70	5,80	10,30	8,10	14	7	11	16
	4	30.07.93	Пт	10,30	14,60	10,50	4,60	5,60	10,00	13,10	3,70	17	9	13	16
	5	31.07.93	Сб	6,10	11,00	13,70	12,30	10,30	14,60	10,50	4,60	17	9	13	16
	6	01.08.93	Вс	8,20	9,10	9,50	10,00	6,10	11,00	13,70	12,30	17	9	13	16
	7	02.08.93	Пн	6,20	8,10	6,10	4,90	8,20	9,10	9,50	10,00	17	9	13	16
4-й аналог	8	03.08.93	Вт	12,40	15,20	9,10	5,00	6,20	8,10	6,10	4,90	15	9	13	16
	9	04.08.93	Ср	11,70	14,50	13,50	13,60	12,40	15,20	9,10	5,00	17	9	13	14
	10	05.08.93	Чт	5,10	6,90	6,80	10,10	11,70	14,50	13,50	13,60	17	8	11	14
	11	06.08.93	Пт	9,90	6,90	9,10	11,90	5,10	6,90	6,80	10,10	17	9	13	14
	12	07.08.93	Сб	8,30	7,90	16,30	13,90	9,90	6,90	9,10	11,90	17	9	13	14
	13	08.08.93	Вс	7,60	9,50	11,70	12,90	8,30	7,90	16,30	13,90	17	9	13	14
	14	09.08.93	Пн	10,70	15,50	11,50	29,40	7,60	9,50	11,70	12,90	17	9	13	12
2-й аналог	15	10.08.93	Вт	18,50	17,60	18,60	13,50	10,70	15,50	11,50	29,40	17	10	13	12
	16	11.08.93	Ср	9,70	17,80	22,70	13,20	18,50	17,60	18,60	13,50	17	10	13	12
	17	12.08.93	Чт	14,70	16,30	12,30	22,50	9,70	17,80	22,70	13,20	17	10	13	12
	18	13.08.93	Пт	18,50	14,00	14,70	17,20	14,70	16,30	12,30	22,50	17	10	13	12
	19	14.08.93	Сб	8,90	8,70	17,80	23,00	18,50	14,00	14,70	17,20	18	10	13	12
	20	15.08.93	Вс	15,40	12,40	9,70	7,90	8,90	8,70	17,80	23,00	18	10	13	12
	21	16.08.93	Пн	10,50	10,60	5,50	7,00	15,40	12,40	9,70	7,90	20	10	14	12
1-й аналог	22	17.08.93	Вт	13,50	24,60	4,40	13,60	10,50	10,60	5,50	7,00	20	12	14	12
	23	18.08.93	Ср	14,90	16,40	5,80	13,70	13,50	24,60	4,40	13,60	20	12	14	12
	24	19.08.93	Чт	22,50	11,40	6,10	13,70	14,90	16,40	5,80	13,70	20	10	14	12
	25	20.08.93	Пт	10,00	18,50	18,10	19,10	22,50	11,40	6,10	13,70	20	12	14	12
	26	21.08.93	Сб	17,20	14,00	5,80	6,40	10,00	18,50	18,10	19,10	20	12	12	12
	27	22.08.93	Вс	16,30	8,50	7,90	11,30	17,20	14,00	5,80	6,40	19	12	12	12
	28	23.08.93	Пн	15,20	11,50	8,00	6,80	16,30	8,50	7,90	11,30	19	11	14	12
3-й аналог	29	24.08.93	Вт	14,00	21,70	24,40	9,80	15,20	11,50	8,00	6,80	19	9	12	12
	30	25.08.93	Ср	21,10	10,40	11,00	15,30	14,00	21,70	24,40	9,80	19	11	14	12
	31	26.08.93	Чт	14,20	24,70	20,70	10,20	21,10	10,40	11,00	15,30	17	11	14	12
	32	27.08.93	Пт	14,30	19,60	16,00	9,60	14,20	24,70	20,70	10,20	19	11	14	12
	33	28.08.93	Сб	26,90	28,10	20,90	16,70	14,30	19,60	16,00	9,60	18	11	14	12
	34	29.08.93	Вс	26,90	24,50	16,50	7,20	26,90	28,10	20,90	16,70	18	11	13	12
	35	30.08.93	Пн	16,50	16,80	19,60	10,90	26,90	24,50	16,50	7,20	18	11	13	12
Текущая неделя	36	31.08.93	Вт	16,30	21,40	16,70	13,00	16,50	16,80	19,60	10,90	18	10	13	12
	37	01.09.93	Ср	15,40	20,50	20,40	20,60	16,30	21,40	16,70	13,00	18	10	13	12
	38	02.09.93	Чт	13,70	15,80	17,40	14,40	15,40	20,50	20,40	20,60	17	10	13	12
	39	03.09.93	Пт	17,00	14,60	17,70	20,00	13,70	15,80	17,40	14,40	17	10	13	12
	40	04.09.93	Сб	9,60	12,10	16,40	12,10	17,00	14,60	17,70	20,00	15	10	13	12
	41	05.09.93	Вс	8,30	13,60	18,40	14,90	9,60	12,10	16,40	12,10	17	10	13	12
	42	06.09.93	Пн	16,60	16,30	13,20	9,40	8,30	13,60	18,40	14,90	18	10	13	12

Прогноз во временной области

Требуется получить прогноз уровня глюкозы в крови больного диабетом, который будет у него в понедельник утром 6 сентября, по данным шести предшествующих дней наблюдения. Поскольку мы хотим использовать выборку данных, содержащую только шесть строк, а этого мало для получения точной модели по комбинаторному алгоритму МГУА, в выборку были добавлены средние точки [4]. Средними называются точки, координаты которых равны средним значениям координат точек выборки. Для их генерации перебираются все пары точек или строк выборки: 1-2, 1-3, …, 5-6. Для каждой пары находятся среднегеометрические значения всех переменных, и в выборку данных записывается новая строка $x_{i-j} = \frac{x_i \cdot x_j}{2}$, где *i* и *j* номера строк, для которых генерируется средняя точка.

Так для шести строк исходной выборки сгенерировано 15 новых строк, всего 21 строка новой выборки, которая и используется для прогноза.

С учетом средних точек, по комбинаторному алгоритму МГУА, находим следующую прогнозирующую модель:

$y_{пн,6} = x_{6,пн} = -138{,}33 + 1{,}7096\, x_{6вс} - 1{,}614\, x_{17вс} - 1{,}214\, x_{22вс} + 1{,}538\, x_{R1вс} + 8{,}968\, x_{R2вс}$

где $y_{пн,6}$ – прогноз уровня глюкозы, который будет утром в понедельник 6 сентября.

Показатели модели следующие: ER = 1,41; BS = 0,89, где ER – точность модели на проверочной выборке; BS – смещение модели.

Ошибка прогноза для понедельника 6 сентября:

$\Delta y_{пн,6} = | y_{пн,6} - \hat{y}_{пн,6} | = |16{,}6 - 15{,}9| = 0{,}7.$

Прогноз в аналоговой области по данным одного дня наблюдения

Дата выдачи прогноза та же, только для получения прогноза использовались данные всего за один день 5 сентября, и пять аналогов, т.е. воскресенья: 29, 22, 15, 8 и 1 августа.

По комбинаторному алгоритму с добавлением средних точек получена следующая прогнозирующая модель:

$y_{пн,6} = x_{6,пн} = -298{,}678 + 0{,}391\, x_{6,вс} + 0{,}207\, x_{12,вс} + 8{,}462\, x_{R1,вс} + 11{,}617\, x_{R3,вс} + 0{,}452\, x_{R4,вс}$

Показатели ее точности следующие: ER = 0,256; BS = 0,005.

Ошибка прогноза для понедельника 6 сентября:

$\Delta y_{пн,6} = | y_{пн,6} - \hat{y}_{пн,6} | = |16{,}6 - 16{,}3| = 0{,}3.$

Прогноз по смешанной аналогово-временной выборке данным

При том же времени выдачи, требуется получить прогноз, пользуясь данными только за два дня: субботу и воскресенье, выбранных из трех недель наблюдения – данной, наблюдаемой недели и двух ее аналогов. Выборка содержит шесть дней наблюдения больного 14 и 15, 21 и 22 августа и 4 и 5 сентября.

Используя также средние точки, по комбинаторному алгоритму МГУА получаем уравнение прогнозирующей модели:

$y_{пн,6} = x_{6,пн} = 155{,}313 - 0{,}775\, x_{12,вс} - 0{,}127\, x_{17,вс} + 0{,}415\, x_{R3,вс} - 10{,}832\, x_{R4,вс}$

Показатели ее точности следующие: ER = 0,236; BS = 0,002.

Ошибка прогноза для понедельника 6 сентября:

$\Delta y_{пн,6} = | y_{пн,6} - \hat{y}_{пн,6} | = |16{,}6 - 16{,}4| = 0{,}2.$

Сравнение результатов прогноза

Сравнение результатов прогноза во временной области и прогноза по аналоговым данным и смешанным аналогово-временным данным показало, что учет аналогов снижает ошибку почти в два раза, но главное состоит в том, что решается задача экспресс-прогноза, для которого может быть использована весьма короткая выборка данных.

Прогноз уровня глюкозы в крови в течение одного дня наблюдения

Ставится задача дать прогноз уровня глюкозы в крови в течение одного дня лечения в определенные моменты измерения уровня глюкозы: в 6, 12, 17 и 22 часа. Прогноз рассчитан для варианта аналогово-временного ряда, давшего лучший результат в предыдущем случае. Прогноз утреннего значения уровня глюкозы был приведен в четвертом разделе. Аналогично рассчитаны прогнозы и для остальных измерений уровня глюкозы.

Модель для прогноза в понедельник 6 сентября, в 12 часов дня:

$x_{12,пн} = -24{,}87 - 0{,}027\, x_{6,вс} + 0{,}565\, x_{12,вс}\ 0{,}392\, x_{17,вс} + 0{,}371\, x_{R1,вс} + 2{,}202\, x_{R3,вс} - 0{,}79\, x_{R4,вс}$

Показатели ее точности следующие: ER = 0,2559; BS = 0,0005.

Ошибка этого прогноза:

$\Delta y_{пн,12} = | y_{пн,12} - \hat{y}_{пн,12} | = |16{,}3 - 15{,}6| = 0{,}7.$

Модель для прогноза в понедельник 6 сентября, в 17 часов дня:

$x_{17,пн} = -101{,}03 - 0{,}05\, x_{6,вс} + 0{,}127\, x_{12,вс}\ 0{,}427\, x_{17,вс} - 0{,}234\, x_{22,вс} - 6{,}039\, x_{R1,вс} + 2{,}809\, x_{R2,вс} - 1{,}15\, x_{R4,вс}$

Показатели ее точности следующие: ER = 0,239; BS = 0,0003.

Ошибка этого прогноза:

$\Delta y_{пн,17} = | y_{пн,17} - \hat{y}_{пн,17} | = |13{,}2 - 12{,}23| = 0{,}95.$

Модель для прогноза в понедельник 6 сентября, в 22 часа:

$x_{22,пн} = 398{,}13 – 0{,}053\ x_{6,вс} - 0{,}005\ x_{17,вс} + 0{,}24\ x_{22,вс} - 9{,}20\ x_{R1,вс} - 0{,}24\ x_{R2,вс} – 16{,}43\ x_{R3,вс}$

Показатели ее точности следующие: ER = 0,458; BS = 0,0006.

Ошибка этого прогноза:

$\Delta y_{пн,22} = | y_{пн,22} - \hat{y}_{пн,22} | = |9{,}4 – 6{,}0| = 3{,}4.$

Поскольку прогноз в 22 часа оказался неточным, этот прогноз был пересчитан прогноз по аналоговому ряду данных.

Модель для прогноза в понедельник 6 сентября, в 22 часа по аналоговому ряду:

$x_{22,пн} = 10{,}288 – 0{,}021\ x_{6,вс} - 0{,}013\ x_{17,вс} – 0{,}014\ x_{R4,вс}$

Показатели ее точности следующие: ER = 1,18; BS = 0,08.

Ошибка этого прогноза:

$\Delta y_{пн,22} = | y_{пн,22} - \hat{y}_{пн,22} | = |9{,}4 – 10{,}16| = 0{,}76.$

Выводы

Можно сделать вывод, что прогноз по аналогово-временному ряду дает возможность найти значения уровня глюкозы в крови в течение дня с небольшой ошибкой. Учет аналогов дает возможность предсказывать значение уровня глюкозы в крови по одному или двум текущим наблюдениям за больным, если этот больной уже наблюдался и имеется некоторая предыстория его болезни, из которой можно выбрать аналоги недели или одного дня наблюдения. Этот и дает возможность получения экспресс-прогноза, т.е. прогноза, не требующего длительного наблюдения за больным. Повышение точности прогноза может дать выбор из предыстории болезни максимального числа дней-аналогов текущего дня наблюдения.

Литература

1. Ивахненко А.Г., Савченко Е.А., Ивахненко Г.А., Гергей Т. Применение алгоритмов МГУА для восстановления пропущенных данных и прогноза уровня глюкозы в крови при надомном мониторинге диабета // *Проблемы управления и информатики*, № 1, 2002, с. 25-33.
2. А.Г.Ивахненко, Е.А.Савченко, Г.А.Ивахненко, А.Б.Надирадзе, А.О.Рогов Индуктивный метод выбора модели с минимальной ошибкой и наименьшим смещением для решения интерполяционных задач искусственного интеллекта // 6-я международная конференция «Распознавание образов и анализ изображений: новые информационные технологии» (РОАИ-6-2002). Великий Новгород, 21-26 окт. 2002 г.: Тр. Конф.: В 2 т. / НовГУ им. Ярослава Мудрого. - Великий Новгород, 2002. – Т. 1.с. 240-245.
3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. - М.: Статистика, 1973.- 392 с.
4. Ивахненко А.Г., Ивахненко Г.А., Савченко Е.А. Концепция последовательных алгоритмических приближений (спусков) к точному решению интерполяционных задач искусственного интеллекта // Кибернетика и вычислительная техник, № 127, 2000, стр. 47-58.
5. Ивахненко А.Г., Петухова С.А. и др. Объективный выбор оптимальной кластеризации выборки данных при компенсации неробастных случайных помех // Автоматика, №3, 1993, с. 46- 58.

Савченко Е.А.

ЭКСПРЕСС-ПРОГНОЗ УРОВНЯ ГЛЮКОЗЫ В КРОВИ ПО КОМБИНАТОРНОМУ АЛГОРИТМУ МГУА

Рассматривается задача нахождения краткосрочного экспресс-прогноза по небольшой выборке данных для домашнего мониторинга диабета. Сравнивается три варианта прогноза: по временному ряду данных, по аналоговому ряду и по смешанному аналогово-временному ряду. Учет аналогов повышает точность и смещение прогнозирующей модели почти в два раза. Кроме того, показана возможность прогноза уровня глюкозы в крови на целый день (четыре измерения в определенное время дня).

Савченко Є. А.

ЕКСПРЕС-ПРОГНОЗ РІВНЯ ГЛЮКОЗИ В КРОВІ ЗА КОМБІНАТОРНИМ АЛГОРИТМОМ МГУА

Розв’язується задача знаходження короткострокового прогнозу експрес-прогнозу за невеликою вибіркою даних для домашнього моніторингу діабету. Порівнюються три варіанти прогнозу: за часовим рядом, за аналоговим рядом та за змішаним часово-аналоговим рядом. Урахування аналогів підвищує точність та зміщення прогнозуючої моделі майже удвічі. Крім того, показано можливість прогнозу рівня глюкози у крові на цілий день (чотири вимірювання у певний час дня).

Savchenko E.A.

ECSPRESS-PREDICTION OF THE CONTENT OF THE BLOOD GLUCOSE LEVEL BY COMBINATORIAL GMDH ALGORITHM

The problem of a short-term express-prediction after a small data sample is considered for home monitoring of diabetes. Three variants of prediction are compared: by time series, b y analog series and by composed time-analog series. Considering analogs increases the accuracy and bias of the prediction model in nearly two times. In addition, possibility of prediction of the blood glucose level is shown on the whole day (four measurements in defined time of a day).