

Автоматическое выявление значений факторов риска заболеваний с помощью методов искусственного интеллекта и технологии Интернета вещей

А. И. Молодченков^{1, II, III}, О. Г. Григорьев^I, Я. Н. Шарафутдинов^{II}

^I Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», г. Москва, Россия

^{II} Московский физико-технический институт, Московская область, г. Долгопрудный, Россия

^{III} Российский университет дружбы народов, г. Москва, Россия

Аннотация. В работе описываются алгоритмы, которые позволяют автоматически получать значения факторов риска ряда заболеваний на основе данных, поступающих с браслетов, умных часов, весов и других устройств, относящихся к Интернету вещей. В настоящее время такие устройства позволяют измерять показатели состояния здоровья и образа жизни человека. На основании этих показателей можно определить значения факторов риска для различных заболеваний. В статье, на примере определения среднего значения пульса в покое, описан подход по применению методов искусственного интеллекта для получения значений факторов риска на основе анализа данных, поступающих с носимых устройств.

Ключевые слова: здоровьесбережение, факторы риска, искусственный интеллект, базы знаний, профилактика, Интернет вещей

DOI 10.14357/20718632210109

Введение

В Федеральном исследовательском центре «Информатика и управление» Российской академии наук в сотрудничестве с рядом коммерческих организаций был разработан прототип системы автоматизированной оценки рисков заболеваний и выработки профилактических рекомендаций для их уменьшения [1]. На риски заболеваний влияют различные факторы, которые можно разделить на такие группы, как: Болезни и семейный анамнез, Образ жизни, Питание, Физиология, Лабораторные и функциональные исследования, Среда и социальный статус, События жизни. В настоящее время

разработанный прототип содержит более 250 различных факторов риска. Показатели для их оценки должны сниматься с разной периодичностью. Некоторые показатели, например, пульс, должны сниматься несколько раз в день, другие, например, личностные черты - раз в пять лет. Понятно, что постоянно вручную вводить такое количество показателей практически невозможно. Поэтому актуальной задачей является автоматизация процесса получения значений различных показателей.

В связи с активным развитием и распространением технологий удаленного мониторинга и Интернета вещей становится возможным

получение значений ряда показателей здоровья человека в режиме реального времени.

В качестве средств, относящихся к технологии Интернета вещей, для сбора данных о здоровье человека, в основном используются фитнес браслеты, умные часы, электронные весы, умные холодильники, камеры видеонаблюдения и другие гаджеты. Самыми распространенными являются носимые устройства. Почти все они имеют однотипный набор датчиков для съема значений показателей здоровья человека. Рассмотрим наиболее распространенные в настоящее время модели носимых устройств: Xiaomi Mi Smart Band 4, Honor Band 5, Samsung Gear Fit2 Pro, Huawei Band 4 Pro, Huawei Band 4, Samsung Galaxy Fit, Xiaomi Mi Band 3, Amazfit Cor 2, Garmin Vivosmart 4, Polar A370.

Почти во всех вышеперечисленных устройствах имеются акселерометр и пульсометр. В некоторые устройства, например, Xiaomi Mi Smart Band 4, встроен гироскоп.

Основными показателями, которые считывают фитнес браслеты, являются: количество шагов, частота пульса, давление, качество сна, количество сжигаемых калорий, содержание кислорода в крови, показатель уровня стресса. В отдельных устройствах реализованы некоторые дополнительные возможности. Например, Garmin Vivosmart 4 собирает минуты интенсивной физической активности пользователя, пройденное расстояние, Xiaomi Mi Smart Band 4 позволяет построить диаграмму сердечного ритма. Кроме этого, фитнес браслетами предоставляются данные GPS, показатели здоровья для разных стадий активности человека: бег на улице, бег на беговой дорожке, езда на велосипеде, велотренажер и т.д.

Используя значения всех этих показателей можно своевременно получать значения факторов риска различных заболеваний.

1. Расчет значений основных показателей, поступающих с носимых устройств

С помощью носимых устройств можно получить следующие показатели:

1. Количество сожженных калорий.
2. Наличие стресса.
3. Классификация образа жизни.

4. Классификация качества сна на основе сердечного ритма.

5. Определение показателей сна.

6. Классификация утомляемости.

7. Определение параметров телосложения и возрастных признаков.

Для определения количества сожженных калорий в работе [2] используются размеченные данные с акселерометра и тип физической активности (1 - сидячий, 2 - бег или ходьба, 3 – повседневная активность). Перед использованием методов машинного обучения для подсчета калорий применяется алгоритм определения типа активности на основе данных с акселерометра. Также происходит предобработка данных, при которой суммируются значения ускорений в течение одной минуты по каждой из осей. В статье [2] были использованы линейная и нелинейная регрессия, а именно: регрессионное дерево решений M5. В Табл. 1 представлены значения точности использованных моделей.

В статье [3] предлагается рассчитывать количество сожженных калорий на основе параметров человека по следующим формулам:

1. для мужчин:

$$((-55.0969 + (0.6309 \times HR) + (0.1988 \times W) + (0.2017 \times A)) / (4.184 \times 60 \times T));$$

2. для женщин:

$$((-20.4022 + (0.4472 \times HR) - (0.1263 \times W) + (0.074 \times A)) / (4.184 \times 60 \times T));$$

где HR – частота сердцебиения, W – вес, A – возраст, T – температура.

В Табл. 2 приведены примеры расчета количества сожженных калорий.

В работе [4] описан способ применения методов машинного обучения для подсчета уровня стресса на основе значения вариабельности частоты сердечных сокращений. В качестве моделей машинного обучения использовались логистическая регрессия и метод опорных векторов (SVM). Количество признаков и размер тренировочного набора равны соответственно 270 и 318. Оценка качества полученных моделей приведена в Табл. 3.

Что касается образа жизни, то тут чаще всего авторы предлагают алгоритмы определения – ведет человек здоровый образ жизни или нет. Так, в работе [4] была построена модель на основе логистической регрессии. Входными признаками являются количество потребляемых и

Табл. 1. Таблица методов определения калорий

	Linear Regression	Non-linear Regression (Decision Tree M5)
Коэффициент корреляции	0.87	0.91
Среднеквадратическая ошибка	1.02	0.88
Средняя абсолютная ошибка	0.63	0.59
Максимальное абсолютное отклонение	4.21	3.36

Табл. 2. Примеры расчёта количества сожжённых калорий

Пол	Вес	Возраст	Частота сердцебиения	Количество сожжённых калорий
мужской	65	22	160	906
мужской	60	22	155	847
женский	50	22	145	728
женский	47	22	145	719

Табл. 3. Значения точности методов определения наличия стресса

	Тренировочная выборка	Тестовая выборка
Logistic Regression	100%	66%
SVM	97%	68%

сжигаемых калорий. Точность построенной модели составляет 82,42%.

Качество сна предлагается определять, как хорошее, плохое и среднее [5]. Для качественной интерпретации и объяснения статуса качества сна авторы статьи используют глобальный алгоритм индукции правил покрытия LERS (Learning from Examples based on Rough Sets) [6]. Общее количество собранных суточных наборов данных составляет 2829.

Авторы статьи [7], на основании анализа медицинской литературы и признаков, получаемых с носимых устройств, предлагают следующие обобщённые показатели сна: качество сна, уровень сонливости, хронотип (определяет склонность ко сну в определенное время в течение 24-часового периода). В качестве признаков используются данные с носимого устройства: продолжительность сна, фазы сна, циклы сна и бодрствования, минимальная частота сердечных сокращений и средняя температура. Алгоритм прогнозирования обучается на начальном этапе с использованием данных, собранных с носимого устройства, и ответа пользователя на вопрос о

его/ее ощущении качества сна. Пользователь должен выбрать свое собственное чувство, используя шкалу Лайкерта (с 5 баллами): 0 - очень усталое; 25-усталое; 50- среднее; 75-счастливое; 100-очень счастливое (что является целевым признаком). Необходимо минимум пять суточных проб, чтобы получить первое прогнозируемое значение качества сна, которое улучшается, когда увеличивается количество суточных проб. После первоначальной обучающей тренировки ежедневно выводится индикатор SQ (sleep quality) со значением от 0 до 100. Этот индикатор измеряет уровень расслабления и восстановления сна, на основе следующих показателей: продолжительность сна, переход в состояние сна, бодрствование, минимальная ЧСС и среднее значение температуры тела.

Для обучения были использованы алгоритмы:

1. K* [8].
2. Linear Regression.
3. SMOreg [9].
4. Gaussian processes classifier [10].
5. Multilayer perceptron [11].
6. Locally weighted learning [12].

Значения точности моделей, определяемые по нормированной средней абсолютной ошибке и по нормированному стандартному отклонению (в скобках) приведены в Табл. 4.

Авторы статьи [7] определили сонливость как состояние, близкое ко сну, в тех ситуациях, когда пользователь должен быть активным. Поэтому были использованы два вида данных. Первый включает данные от датчиков частоты сердечных сокращений, температуры и акселерометра, собранные, когда человек спит ночью. Второй включает сбор данных с тех же датчиков в моменты, когда пользователь чувствует себя действительно активным. Использовались следующие методы [7]: C4.5, Adaboost(C4.5), KNN, Random forest, Naive Bayes, OneR, ZeroR.

Результаты по метрике Ассигасу для различных моделей обучения приведены в Табл. 5.

Хронотип определяет и продолжительность сна и бодрствования в часах. Для расчета этого показателя используется таблица эквивалентности, связывающая время начала и окончания сна со значением в шкале утренней и вечерней. Началу и концу сна присваиваются значения согласно Табл. 6, на основе которых определяется хронотип в диапазоне -4 (-2 начало

Для определения утомляемости строится модель бинарной классификации с применением метода опорных векторов (SVN) и искусственных нейронных сетей (ANN): утомлен или бодр. Для построения множества признаков использовались следующие данные [7]: частота сердечных сокращений, температура запястья и проводимость кожи. Эти данные собираются с фитнес-часов и предварительно обрабатываются. На их основе создаются новые признаки во временной и частотной областях. Из всех признаков были выбраны те, которые вносят наибольший вклад в результаты классификации:

1. средняя частота сердечных сокращений;
2. стандартное отклонение частоты сердечных сокращений;
3. средняя температура;
4. стандартное отклонение температуры;
5. спектральная плотность мощности сердечного пульса;
6. спектральная плотность мощности температуры.

Для построения признаков использовалось окно размером 30 секунд. Спектральная плотность мощности рассчитывалась с помощью методов Welch [13] и Burg [14].

Табл. 4. Значения точности моделей машинного обучения по определению качества сна

K*	Linear Regression	SMOreg	Gaussian processes	Multilayer Perceptron	LWL
0.155 (0.02)	0.19 (0.023)	0.2025 (0.023)	0.195 (0.02)	0.2025 (0.043)	0.1825 (0.023)

Табл. 5. Значения точности моделей машинного обучения по определению сонливости

C4.5	Adaboost (C4.5)	Knn	Random Forest	Naive Bayes	OneR	ZeroR
97.26 %	97.84 %	97.12 %	97.51 %	83.55 %	90.98 %	63.97 %

Табл. 6. Таблица эквивалентности с утренней и вечерней шкалами

Тип	Время начала сна	Значения	Время окончания сна	Значения
Очень раннее утро	-/21:30	2	-/05:00	2
Раннее утро	21:30/22:45	1	05:00/06:30	1
Переходное время	22:45/00:45	0	06:30/08:30	0
Вечер	00:45/02:00	-1	08:30/10:00	-1
Поздний вечер/ночь	02:00/-	-2	10:00/-	-2

Использовались две модели: SVM и нейронная сеть (ANN).

Искусственная нейронная сеть ANN создается и обучается с тремя наборами на основании 80% от доступного размеченного набора данных, в то время как остальные 20% записей используются для тестирования. Предлагаемая ANN с прямой связью, основана на входном слое из шести узлов, одном скрытом слое и единственном выходном модуле с передаточной функцией касательно-сигмоидной формы. Кроме того, было рассмотрено много конфигураций для определения структуры ANN с наилучшей производительностью. Это включало изменение количества скрытых слоев и узлов в каждом слое, а также изменение алгоритмов обучения и функции передачи решения.

На основе того же набора обучающих данных с помощью алгоритма SVM была построена модель, которая может классифицировать данные без предобработки. В работе [7] SVM также используется для классификации собранного набора данных по двум классам: бодрость и усталость. Значения характеристик каждого класса нормализованы. Этап классификации реализуется с использованием набора сформированного данных, чтобы классифицировать операторов в состоянии бодрости и в состоянии усталости.

Было проведено восемь испытаний случайно выбранных записей из отдельных собранных наборов данных для валидации, обучения и тестирования, чтобы рассчитать точность (precision) классификации для обоих классификаторов. Табл. 7 показывает точность классификации для ANN и SVM. Подход SVM показывает преимущество показателей точности по сравнению с подходом ANN.

Табл. 7. Таблица, иллюстрирующая точность моделей ANN и SVM

	ANN	SVM
Fatigue	94,7%	97,2%
Alert	88,3%	91,3%

2. Вычисление значений факторов риска заболеваний

Авторами этой статьи ведутся разработки алгоритмов по определению значений факторов

риска на основе данных, поступающих с носимых устройств. Рассмотрим один из таких алгоритмов на примере определения среднего значения пульса в покое. Этот показатель является фактором риска для целого ряда заболеваний. Известно, что носимые устройства снимают значения показателей постоянно. В современные устройства встроен датчик, который позволяет определять моменты, когда человек не двигается. Считается, что в этот период он находится в покое и вычисляется значение пульса в покое. Однако при получении этих данных с таких мест их хранения, как Google fit, Apple health, Samsung health, предоставляются временные ряды, в которых не выделены промежутки состояния человека в покое. Поэтому вычисление среднего значения пульса, когда человек находится в покое на такого рода данных – это не тривиальная задача. Авторам не удалось найти публикации, в которых бы описывалось решение этой задачи. Все публикации направлены на решение задач использования значений показателя пульса в разных состояниях для определения сердечно-сосудистых заболеваний у пациентов.

Для корректной работы алгоритма вычисления среднего значения пульса в покое на основе значения временных рядов необходимо максимально убрать из временного ряда те значения пульса, когда человек спал или находился не в состоянии покоя. Для разработки алгоритма решения этой задачи использовались значения временных рядов следующих показателей, которые считываются с помощью гаджета (браслета, умных часов и т.п.): пульса, качества сна, количества шагов, интенсивности физической нагрузки, количества сжигаемых калорий (все показатели снимаются раз в минуту).

Общая методика заключается в определении периода бодрствования и выделения в этом периоде интервалов, когда пользователь находился в состоянии покоя, с дальнейшим сглаживанием результатов. Через `rest_state` будем обозначать двумерный массив, который соответствует временному ряду состояния объекта, где значение члена ряда определяется с интервалом в минуту как: 0 – состояние движения, 1 – состояние покоя, 2- состояние сна. Для каждого показателя (пульс, качество сна, количество шагов, количество сжигаемых калорий, интенсивность физической нагрузки) создается

массив значений. Обозначим эти массивы следующим образом: 'Heartrate' - массив сердечного пульса, 'Sleep' - массив качества сна, 'Steps' - массив количества шагов, 'Calories' - массив количества сжигаемых калорий, 'Intensity' - массив интенсивности физической нагрузки. Все вышеупомянутые массивы являются двумерными, первый столбец (назовём его 'value') содержит значения показателей за минуту, второй столбец (назовём его 'Time') отображает время, в которое был снят показатель.

На первом шаге массив `rest_state` заполняется на основе значений временного ряда сна: 2, если объект спал, 0 – в ином случае. Блок-схема алгоритма изображена на Рис. 1.

Далее ячейки столбца 'value' массива 'rest_state' с нулевым значением заполняются значением 1, если в соответствующие минуты значение временного ряда количества шагов равно 0. Блок-схема соответствующего алгоритма изображена на Рис. 2.

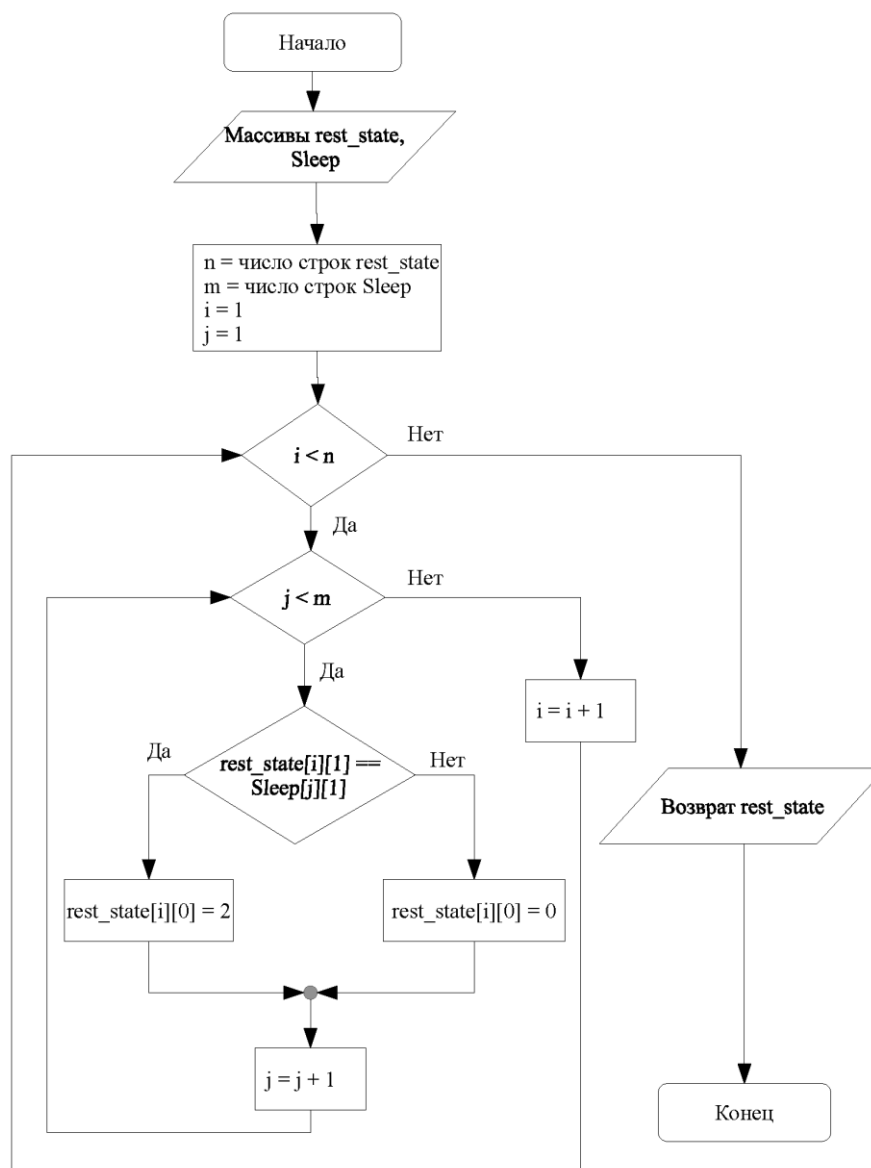


Рис. 1. Блок-схема алгоритма заполнения массива `rest_state` на основе значений временного ряда сна

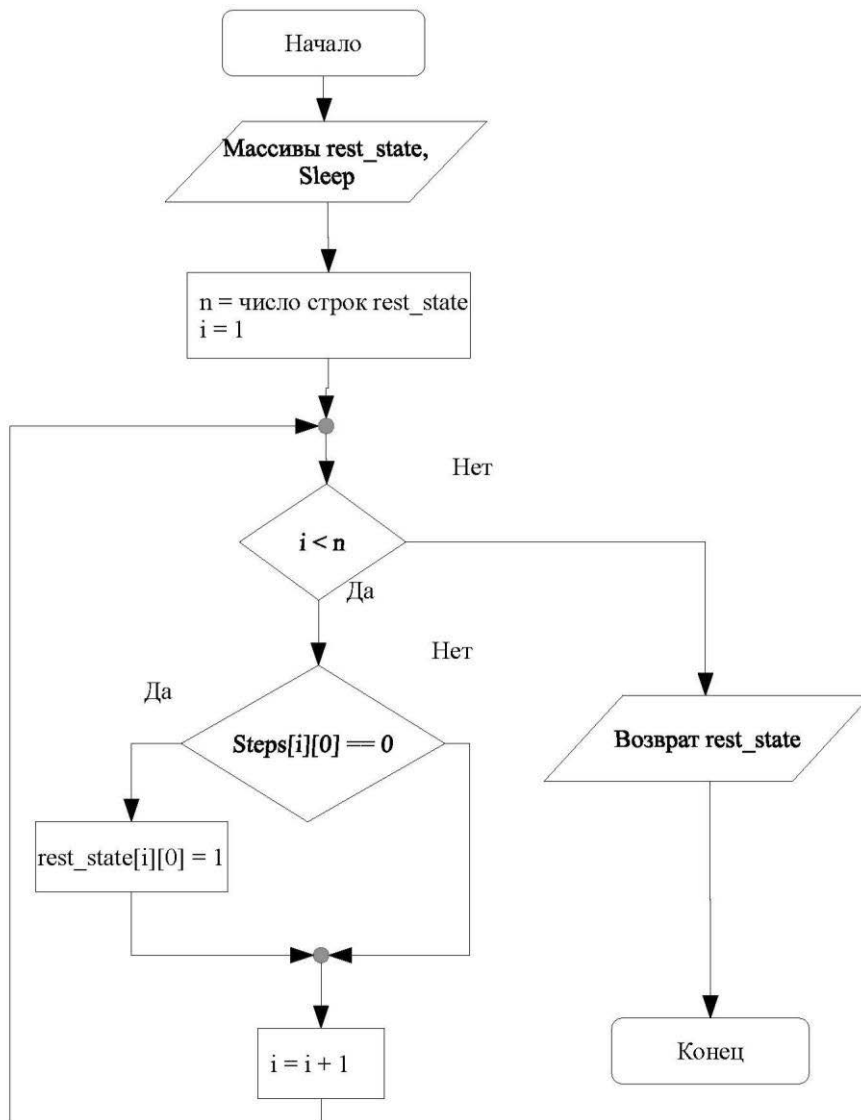


Рис. 2. Блок-схема второго шага алгоритма

На третьем шаге отбираются ячейки массива 'rest_state' со значением 1 и им присваивается значение 0, если интенсивность нагрузки человеческой активности превышает 1 (на основании соответствующих массивов временных рядов). Алгоритм этого шага изображен на Рис. 3.

На четвертом шаге отбираются ячейки массива 'rest_state' со значением 1 и им присваивается значение 0, если соответствующее значение временного ряда «количество сжигаемых калорий» превышает 2.282 калорий. Число 2.282 было вычислено экспериментальным путем. Алгоритм этого шага изображен на Рис. 4.

На следующем шаге производится разбиение значений временного ряда сердечного пульса на значения в состоянии покоя и движения. Для каждого значения пульса, которое не попадает на период сна, вычисляется n_{pos} - количество значений пульса, выше текущего и помеченных как в состоянии покоя (определяется временным рядом 'rest_state'), и n_{neg} - количество значений пульса, ниже текущего и помеченных как в состоянии движения. Если $n_{pos} \geq n_{neg}$, то соответствующей данной минуте ячейке массива rest_state присваивается значение 1, иначе - 0. Алгоритм изображен на Рис. 5.

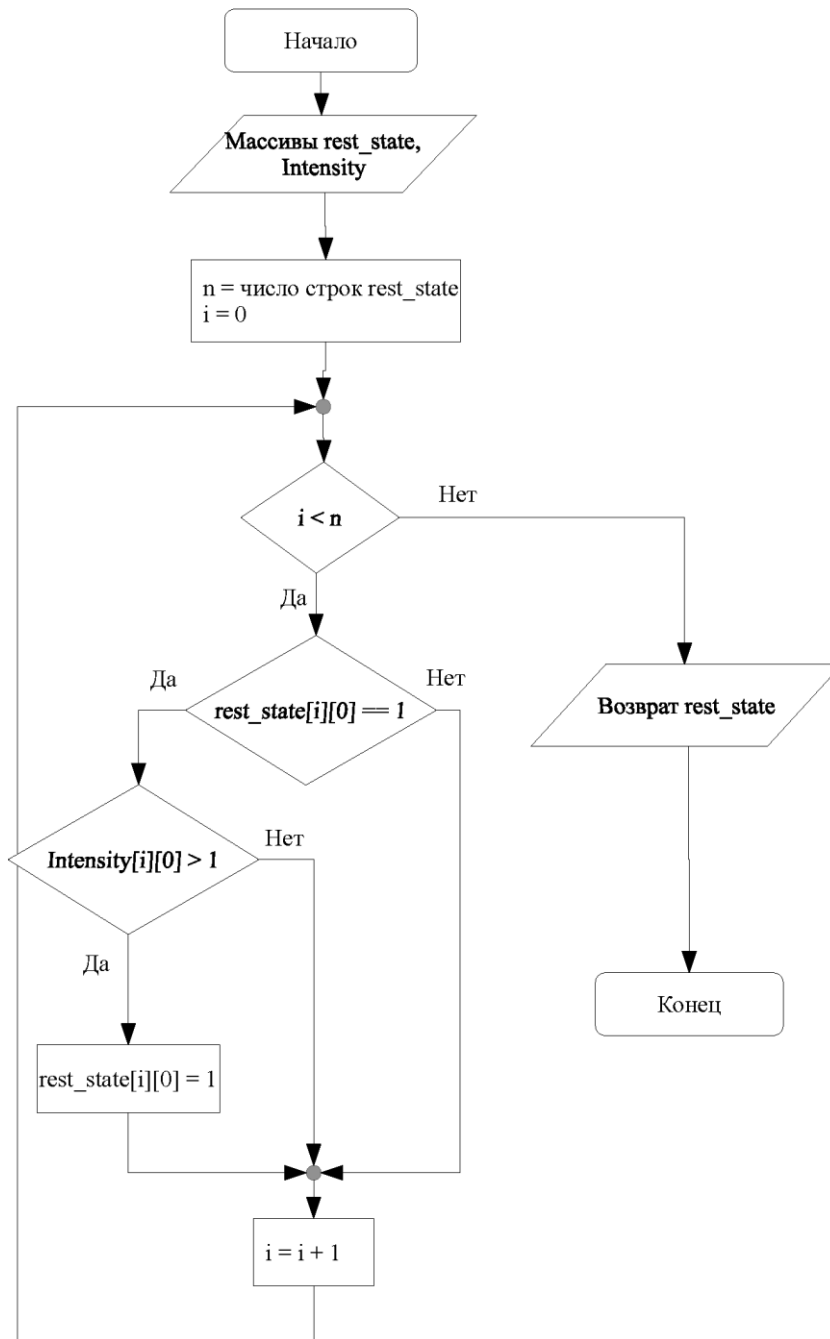


Рис. 3. Блок-схема третьего шага алгоритма

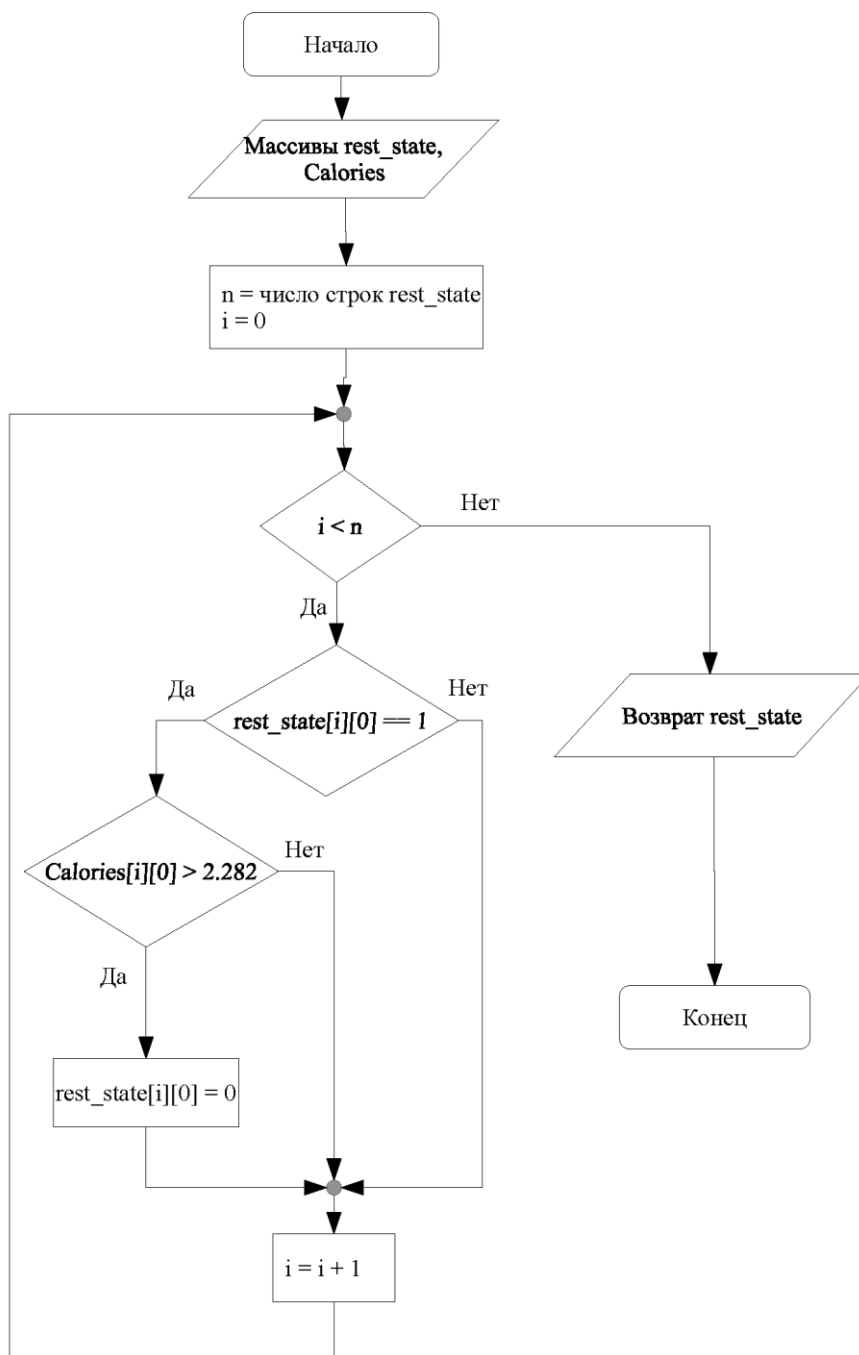


Рис. 4. Блок-схема четвертого шага алгоритма

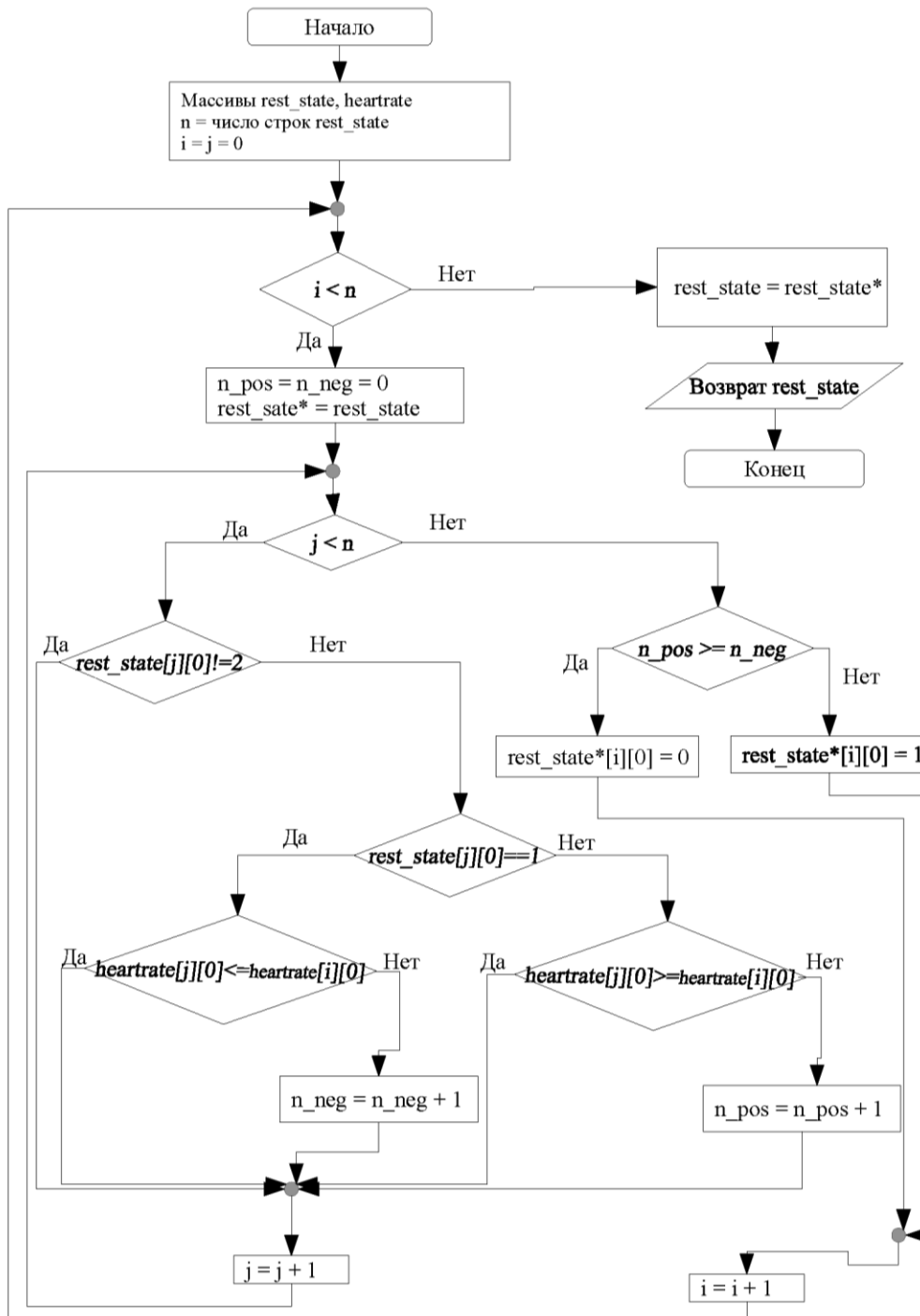


Рис. 5. Блок-схема пятого шага алгоритма

На последнем шаге алгоритма для сглаживания результата и коррекции случайных ошибок используется функция оконного фильтра. Она применяется к массиву 'rest_state'. Размер окна равен 3 минутам (иными словами учиты-

вает влияние состояний минутой позже и минутой раньше). Каждой ячейке массива 'rest_state' присваивается значение, встречающееся в окне большее количество раз. Алгоритм этого шага изображен на Рис. 6.

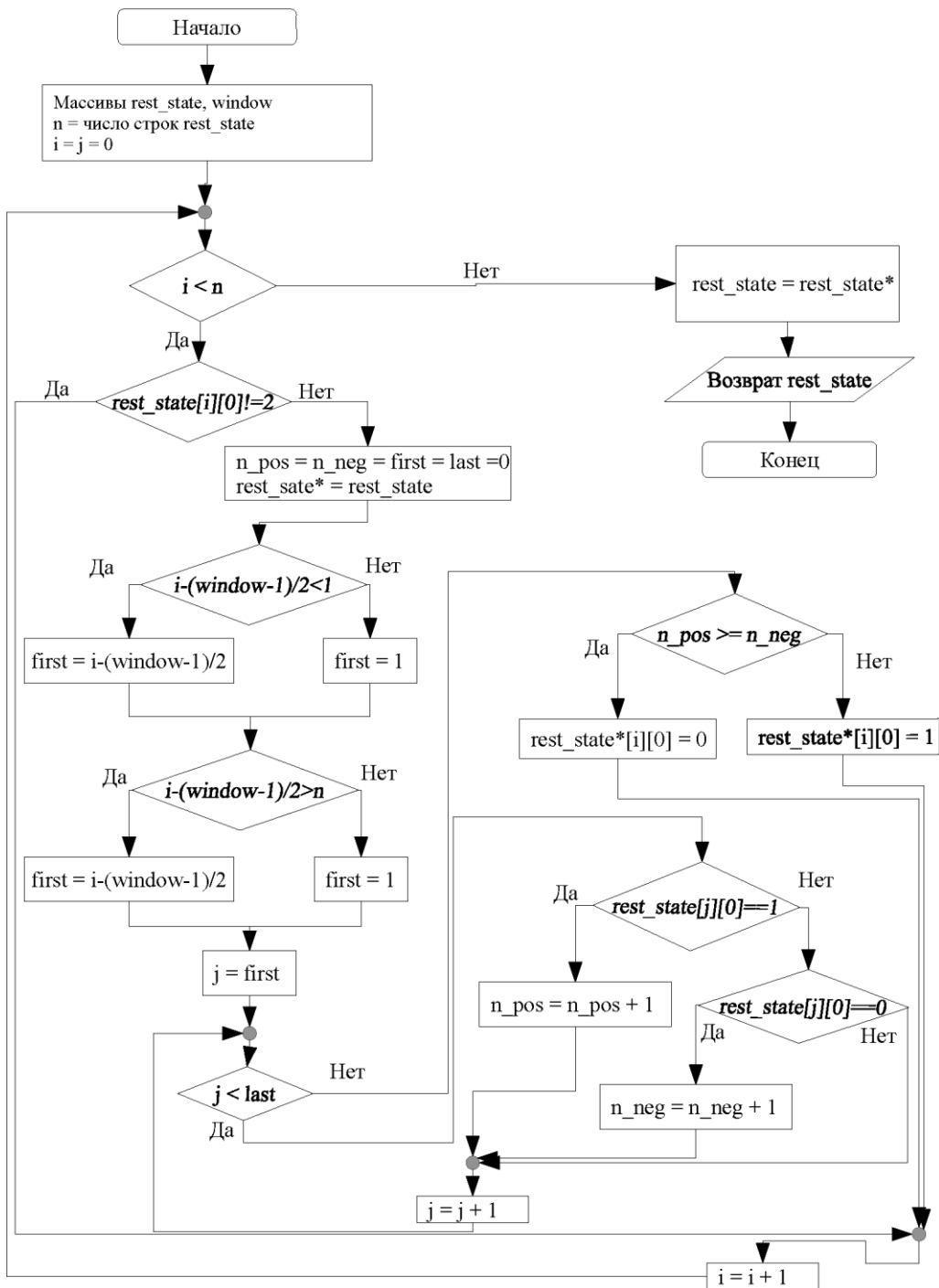


Рис. 6. Блок-схема шестого шага алгоритма

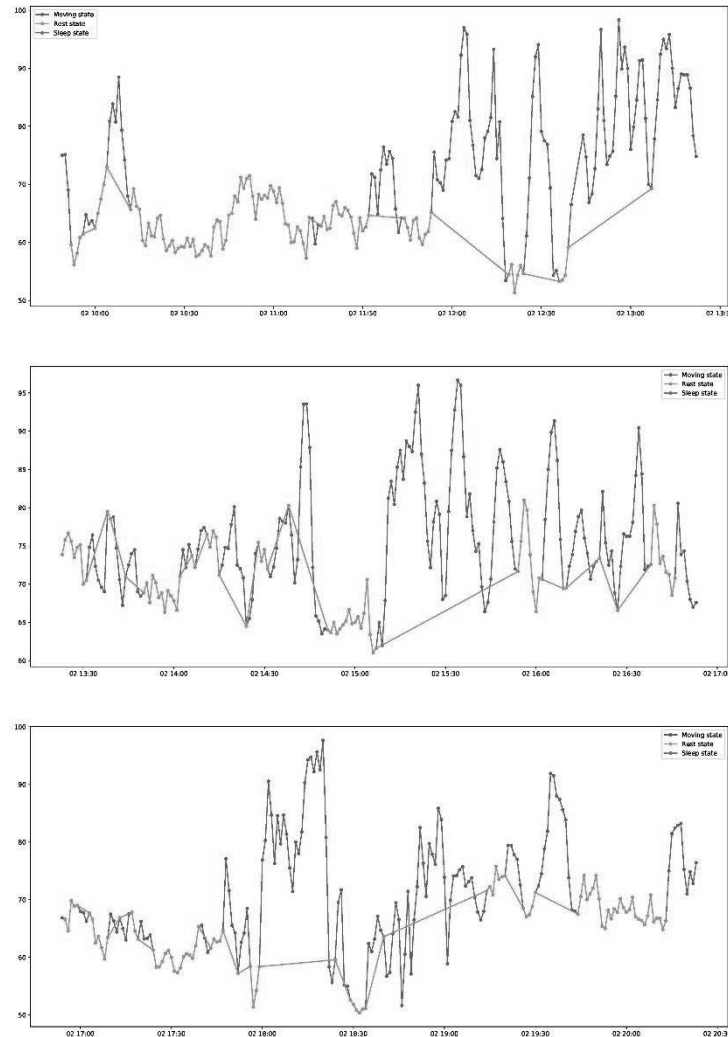


Рис. 7. Пример работы алгоритма

На графиках изображено значение пульса в зависимости от времени

Далее, для всех ячеек массива `rest_state`, равных 1, отбираются значения пульса и вычисляется среднее. На Рис. 7 изображен пример работы алгоритма. Состояние покоя (оранжевым цветом) показано графиком без разрывов - место разрыва (при переключении на другое состояние) заменено на прямую линию.

Для оценки эффективности работы алгоритма был использован набор данных Fitbase [15], который содержит последовательности временных рядов, снятых ежедневно в течение недели, куда входят и значения пульса в покое. Для оценки качества результатов работы разработанного алгоритма каждое значение пульса в покое в наборе данных было заменено интервальным значением

со значением $\Delta = \pm 2$ от его значения. Например, если значение пульса в покое равно 57, то оно заменяется на интервал от 55 до 59. Если результат работы алгоритма попадает в этот интервал, то считается, что значение вычислено корректно, иначе нет. В результате точность работы алгоритма составила 93%

Заключение

В статье описаны методики расчёта различных показателей здоровья на основе данных, получаемых с носимых устройств. Приведен алгоритм расчета значения одного из факторов риска заболеваний на основе значений показателя

телей здоровья, получаемых с помощью носимых устройств. Результаты экспериментальных исследований показали высокую эффективность работы разработанного алгоритма.

Литература

1. Grigoriev O. G., Molodchenkov A. I. Technology of Personalized Preventive Recommendation Formation Based on Disease Risk Assessment // 17th Russian Conference on Artificial Intelligence, RCAI 2019 (Ulyanovsk; Russian Federation; 21-25 October).2019. Springer, Communications in Computer and Information Science, Vol.1093. P. 298-309. DOI: 10.1007/978-3-030-30763-9_25
2. E. Tapia, "Using machine learning for real-time activity recognition and estimation of energy expenditure" Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2008.
3. G. K. Reddy and K. L. Achari A non invasive method for calculating calories burned during exercise using heartbeat // IEEE 9th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, 2015, pp. 1-5.
4. Y. Jain, D. Chowdhury and M. Chattopadhyay Machine Learning Based Fitness Tracker Platform Using MEMS Accelerometer // International Conference on Computer, Electrical & Communication Engineering (ICCECE), Kolkata, 2017, pp. 1-5.
5. R. Choi, W. Kang and C. Son, Explainable sleep quality evaluation model using machine learning approach // 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Daegu, 2017, pp. 542-546.
6. Grzymala-Busse J. W. LERS-a system for learning from examples based on rough sets //Intelligent decision support. – Springer, Dordrecht, 1992. – pp. 3-18.
7. de Arriba-Pérez, F., Caeiro-Rodríguez, M. & Santos-Gago, J.M. How do you sleep? Using off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators. // J Ambient Intell Human Comput 9, 897–917 (2018). <https://doi.org/10.1007/s12652-017-0477-5>.
8. Cleary JG, Cleary JG, Trigg LE (1995) K*: An Instance-based Learner Using an Entropic Distance Measure. In: Proc 12th Int Conf Mach Learn, pp 108–114.
9. Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK (2000) Improvements to the SMO algorithm for SVM regression. IEEE Trans Neural Netw 11:1188–1193
10. MacKay D (1998) Introduction to Gaussian processes. NATO ASI Ser F Comput Sys Sci 168:133–166
11. Ruck DW, Rogers SK, Kabrisky M et al (1990) The multi-layer perceptron as an approximation to a Bayes optimal discriminant function. IEEE Trans Neural Netw 1:296–298.
12. Frank E, Hall M, Pfahringer B (2003) Locally weighted naive bayes. In: Proc Ninet Conf Uncertain Artif Intell, pp 249–256.
13. Welch, P. D. (1967), "The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms" (PDF), IEEE Transactions on Audio and Electroacoustics, AU-15 (2): 70–73, Bibcode:1967ITAE...15...70W, doi:10.1109/TAU.1967.1161901.
14. Orfanidis, S. J. Introduction to Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1995.
15. Fitbase (электронный ресурс) url: <https://www.fitabase.com/resources/knowledge-base/exporting-data/example-data-sets/> (дата обращения 01.02.2021).

Молодченков Алексей Игоревич. Научный сотрудник отдела «Систем поддержки принятия клинических решений» Федерального исследовательского центра «Информатика и управление» РАН, доцент кафедры информационных технологий Российского университета дружбы народов. Автор более 70 печатных работ. Адрес: 117312, г. Москва, пр-т 60-летия Октября, 9. E-mail: aim@isa.ru

Григорьев Олег Георгиевич. Главный научный сотрудник Федерального исследовательского центра «Информатика и управление» РАН. Автор более 80 печатных работ. Адрес: 117312, г. Москва, пр-т 60-летия Октября, 9. E-mail: olegpolikvart@yandex.ru

Шарафутдинов Якуб Насырьянович. Студент магистратуры Московского физико-технического института. Адрес: 141700 Московская область, г. Долгопрудный, Институтский переулок, 9. E-mail: sharafutdinov.yan@phystech.edu

Automatic Calculation of Disease Risk Factors Values Using Artificial Intelligence Methods and Internet of Things Technology

A.I. Molodchenkov^{1, II, III}, O. G. Grigoriev^I, Ya.N. Sharafutdinov^{II}

^I Federal Research Centre "Computer Science and Control" RAS, Moscow, Russia

^{II} Moscow Institute of Physics and Technology, Dolgoprudny, Moscow region, Russia

^{III} Russian People's Friendship University, Moscow, Russia

Abstract. The paper describes algorithms for automatic calculation of the values of risk factors for diseases on the base of data received from bracelets, smart watches, scales and other Internet of Things

devices. At the present time such devices allow us to measure different parameters of a person's health and lifestyle. It is possible to calculate the values of risk factors for various diseases, on the base of these parameters. The article describes an algorithm to calculate the average heart rate at rest values as one of risk factors using data received from wearable devices.

Keywords: health, risk factors, artificial intelligence, knowledge base, prevention, Internet of Things.

DOI 10.14357/20718632210109

References

1. Grigoriev O. G., Molodchenkov A. I. Technology of Personalized Preventive Recommendation Formation Based on Disease Risk Assessment // 17th Russian Conference on Artificial Intelligence, RCAI 2019 (Ulyanovsk; Russian Federation; 21-25 October).2019. Springer, Communications in Computer and Information Science, Vol.1093. P. 298-309. DOI: 10.1007/978-3-030-30763-9_25
2. E. Tapia, "Using machine learning for real-time activity recognition and estimation of energy expenditure" Ph.D. dissertation, Massachusetts Inst. Technol., Cambridge, MA, USA, 2008.
3. G. K. Reddy and K. L. Achari A non invasive method for calculating calories burned during exercise using heartbeat // IEEE 9th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, 2015, pp. 1-5.
4. Y. Jain, D. Chowdhury and M. Chattopadhyay Machine Learning Based Fitness Tracker Platform Using MEMS Accelerometer // International Conference on Computer, Electrical & Communication Engineering (ICCECE), Kolkata, 2017, pp. 1-5.
5. R. Choi, W. Kang and C. Son, Explainable sleep quality evaluation model using machine learning approach // 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Daegu, 2017, pp. 542-546.
6. Grzymala-Busse J. W. LERS-a system for learning from examples based on rough sets //Intelligent decision support. – Springer, Dordrecht, 1992. – pp. 3-18.
7. de Arriba-Pérez, F., Caeiro-Rodríguez, M. & Santos-Gago, J.M. How do you sleep? Using off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators. // J Ambient Intell Human Comput 9, 897–917 (2018). <https://doi.org/10.1007/s12652-017-0477-5>
8. Cleary JG, Cleary JG, Trigg LE (1995) K*: An Instance-based Learner Using an Entropic Distance Measure. In: Proc 12th Int Conf Mach Learn, pp 108–114
9. Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK (2000) Improvements to the SMO algorithm for SVM regression. IEEE Trans Neural Netw 11:1188–1193
10. MacKay D (1998) Introduction to Gaussian processes. NATO ASI Ser F Comput Sys Sci 168:133–166
11. Ruck DW, Rogers SK, Kabrisky M et al (1990) The multi-layer perceptron as an approximation to a Bayes optimal discriminant function. IEEE Trans Neural Netw 1:296–298.
12. Frank E, Hall M, Pfahringer B (2003) Locally weighted naive bayes. In: Proc Ninet Conf Uncertain Artif Intell, pp 249–256.
13. Welch, P. D. (1967), "The use of Fast Fourier Transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms" (PDF), IEEE Transactions on Audio and Electroacoustics, AU-15 (2): 70–73, Bibcode:1967ITAE...15...70W, doi:10.1109/TAU.1967.1161901
14. Orfanidis, S. J. Introduction to Signal Processing. Englewood Cliffs, NJ: Prentice-Hall, 1995.
15. Fitbase <https://www.fitbase.com/resources/knowledge-base/exporting-data/example-data-sets/>

Molodchenkov A. I. PhD. Federal Research Centre "Computer Science and Control" RAS, Peoples friendship University, Moscow, 60-let Oktyabrya av., 9, 119333 Russia. E-mail: aim@isa.ru

Grigoriev O. G. Doctor of sciences. Federal Research Centre "Computer Science and Control" RAS, Moscow, 60-let Oktyabrya av., 9, 119333 Russia. E-mail: olegpolikvart@yandex.ru

Sharafutdinov Ya. N. Student, Moscow Institute of Physics and Technology. Institutsky lane 9, Dolgoprudny, Moscow region, 141700 Russia. E-mail: sharafutdinov.yan@phystech.edu