ВЫЯВЛЕНИЕ АНОМАЛИЙ В ПРОСТРАНСТВЕ ЭКОЛОГИЧЕСКИХ ПРИЗНАКОВ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ ОБНАРУЖЕНИЯ ЖИВЫХ ОБЪЕКТОВ В ЗДАНИИ

И.М. Куликовских[1]

[1]*Самарский национальный исследовательский университет имени академика С.П. Королева, Самара, Россия*

Аннотация

В данной работе рассматривается задача повышения точности обнаружения живых объектов в здании, описываемых пространством экологических факторов. Для решения поставленной задачи реализована модель логистической регрессии при условии неустойчивости оценок параметров для почти линейно разделимых классов. Создан алгоритм выявления аномалий, разрешающий компромисс между наличием выбросов и точностью распознавания. Эффективность предложенного алгоритма и целостность теоретических обоснований получили подтверждение при проведении вычислительных экспериментов.

Ключевые слова: выявление аномалий, логистическая регрессия, машинное обучение, преобразование Кокса–Бокса, система обнаружения, экологический фактор.

Цитирование: **Куликовских, И.М.** Выявление аномалий в пространстве экологических признаков для повышения точности обнаружения живых объектов в здании / И.М. Куликовских // Компьютерная оптика. – 2017. – Т. 41, № 1. – С. 126-133. – DOI: 10.18287/2412-6179-2017-41-1-126-133.

Введение

Задача обеспечения безопасности объектов и мониторинга территорий с использованием мультисенсорных систем (МС) является важной и актуальной [1, 2]. Повышение качества обнаружения несанкционированных вторжений возможно с помощью дополнения существующих МС видеоаналитическими системами [2]. Альтернативным подходом к решению данной задачи может быть создание и исследование механизмов взаимодействия объекта наблюдения и МС, окружающей среды или биологической системы. В этой связи следует упомянуть работы [3, 4], рассматривающие периодичность и синхронность коммуникации насекомых (сверчков) как средство контроля и прогнозирования незаконных вторжений на территорию.

Авторы работы [5], рассматривая задачу обнаружения живых объектов и обеспечения безопасности внутри здания, обращают внимание на специфику, связанную с соблюдением конфиденциальности и, таким образом, желательным исключением видеоконтроля в помещениях. Как отмечается в последней работе, более предпочтительным становится скрытый контроль, ориентированный на анализ и последующее построение модели тесного взаимодействия между окружающей средой, МС и живым объектом на основе экологических факторов [5, 6]. Также достоинством данного подхода является интегрирование систем обеспечения безопасности с интеллектуальными системами контроля в помещениях в контексте «умного дома» (*smart home*), направленных на распределение ресурсов и снижение эксплуатационных затрат [5 – 12].

В качестве адекватного математического аппарата для обнаружения объектов внутри здания на основе МС широкое применение нашли алгоритмы машинного обучения [5, 6, 9 – 12]. При этом, так как процесс обнаружения связан с непрозрачными механизмами взаимодействия между сенсорами и живыми объектами, для сокращения количества ложных срабатываний преимущественно делается выбор в пользу методов обучения по прецедентам [13].

Для создания набора прецедентов на основе данных с сенсоров, как правило, используется цифровая камера, которая с помощью алгоритмов распознавания аудио- и видеоизображений задаёт метки классов. На этапе тестирования модели классификатора камера исключается и решается задача обнаружения объектов на основе имеющихся данных, реализуя скрытый контроль. Тем не менее, необходимость этапа разметки данных вносит ряд неопределённостей.

Во-первых, помимо меток классов, видеоряд может быть использован для получения дополнительных признаков с целью построения более полной модели. Следовательно, сочетание и взаимное влияние используемых на этапе разметки признаков может значительным образом повлиять на формируемую модель взаимодействия факторов окружающей среды, совокупности сенсоров и живых объектов.

Во-вторых, использование результатов распознавания изображений, полученных с камер наблюдения, спутников или прочих источников, предполагает детальное рассмотрение аспектов преднамеренного и непреднамеренного искажения [14 – 18], что может существенным образом повлиять на конечный результат. Более того, выбросы и искажения (аномалии) могут возникнуть в результате работы МС. Как следствие, аномалии в размеченных таким способом данных носят интегральный характер.

Создание адекватных алгоритмов обнаружения аномалий [19 – 21] в пространстве экологических признаков могут значительно упростить решение данной проблемы. Такие алгоритмы не предполагают обработку потока изображений в режиме реального времени [22, 23], а анализируют лишь факт их искажения и влияние на конечный результат. В данной работе предлагается рассмотреть задачу выявления аномалий в пространстве экологических признаков с целью повышения точности обнаружения живых объектов в здании на примере набора данных, детально описанных в работе [5]. Представленные данные были получены с использованием МС и размечены с

помощью цифровой камеры на два класса: в случае обнаружения объекта и его отсутствия.

Другая проблема, которая затрагивается в данной работе, связана с построением модели обучения, устойчивой к наличию различных аномалий в исходных данных [13, 24]. Согласно результатам работы [5], лучшим классификатором для анализируемого набора данных является модель линейного дискриминантного анализа (LDA). Авторы работы отмечают невозможность реализации более простой модели логистической регрессии (LR), так как для некоторых наборов признаков классы являются линейно разделимыми, а, следовательно, модель LR расходится [25]. Удаление выбросов и несоответствий из исходных данных может усугубить данную проблему.

Тем не менее, модель LR предпочтительнее LDA ввиду ряда причин, а именно [13]: 1) даёт лучшие результаты, поскольку основана на менее жёстких гипотезах; 2) не вводит избыточную сущность, как LDA, которая сводит задачу классификации к более сложной задаче восстановления плотностей вероятностей. Таким образом, поставленную в данной работе задачу выявления аномалий необходимо дополнить возможностью последующего обучения с помощью более простой, по сравнению с ранее реализованными методами [5], моделью логистической регрессии.

Данная статья построена следующим образом. Параграф 1 посвящён формализации задачи обнаружения живых объектов в пространстве экологических признаков: описывает структуру набора данных и математическую постановку задачи. Параграф 2 описывает предлагаемый в работе алгоритм выявления аномалий в пространстве экологических признаков. В параграфе 3 приведены результаты вычислительных экспериментов. В заключении перечислены основные результаты, рекомендации по практическому использованию и дальнейшие направления исследований.

1. Формализация задачи обнаружения

Описание исходных данных

Исходные наборы данных для решения поставленной задачи доступны в ресурсе *UCI Machine Learning Repository* в параграфе *Occupancy Detection* [26]. Данные были собраны в помещении $5{,}85 \times 3{,}50 \times 3{,}53$ м с помощью МС, измеряющей *температуру*, *влажность*, *свет*, CO_2, *метки времени*, а также *метки классов*: объект обнаружен (класс 1) или нет (класс 0). Для получения меток классов использовалась цифровая камера, которая фиксировала факт присутствия объекта в помещении в заданный момент времени. Исходный набор признаков был дополнен *влагоёмкостью*, рассчитанной на основе измеренной температуры и относительной влажности. Более подробное описание процедуры сбора данных и метрологические характеристики МС представлены в работе [5].

Для формального описания исходных данных введём следующее определение.

Определение 1. *Пусть X – множество объектов, Y – множество допустимых ответов. Объекты описываются числовыми признаками $f_j: X \to \mathbf{R}$, .., где n – количество признаков. Тогда в русле работы* [13] *вектор $(x^j)_{j=1}^n \in \mathbf{R}^n$, где $x_j = f_j(x)$, называется пространством признаков объекта x.*

Для решения поставленной задачи доступны три набора исходных данных: обучающая выборка $X^l = (x_i, y_i)_{i=1}^l$ и две тестовых $X^{k_1} = (x_i, y_i)_{i=1}^{k_1}$, $X^{k_2} = (x_i, y_i)_{i=1}^{k_2}$. Объём выборки для каждого набора с распределением по классам y_i – в табл. 1.

Табл. 1. Объём выборки с распределением по классам

..	$\lvert X^r \rvert$, $X^r = (x_i, y_i)_{i=1}^r$		
	$y_i = \{0, 1\}$	$y_i = 0$	$y_i = 1$
l	8142	6414	1728
k_1	2664	1693	971
k_2	9751	7703	2048

Выборка $X^l = (x_i, y_i)_{i=1}^l$ была сформирована, когда дверь в помещение была преимущественно закрыта; выборка $X^{k_1} = (x_i, y_i)_{i=1}^{k_1}$ соответствует случаю закрытой двери; выборка $X^{k_2} = (x_i, y_i)_{i=1}^{k_2}$ – случаю открытой двери.

В работе [5] в результате проведенных исследований были сделаны следующие выводы: 1) для обеспечения хорошего качества обнаружения достаточно пары признаков; 2) расширение пространства признаков может привести к ухудшению результата. Таким образом, опираясь на результаты предыдущих исследований, рассмотрим пары признаков для решения поставленной задачи. Принимая во внимание *Определение 1*, обозначим пары в признаковом пространстве как $\{x^p, x^q\}$, где $\{p, q\} \subset j$, $j = \{1, n\}$. На рис. 1 приведено графическое представление обучающей выборки в пространстве признаков $n = 5$: $x^j = \{$*Температура*, *Влажность*, *Свет*, CO_2, *Влагоёмкость*$\}$.

Согласно [5], наилучший результат получен для пары $\{x^1, x^3\}$. Как видно из приведённой графической интерпретации, классы для данной пары являются *почти* линейно разделимыми и содержат выбросы различного происхождения. Исходя из данного показателя, совокупность пар признаков для последующего анализа может быть расширена: $\{x^1, x^3\}$, $\{x^2, x^3\}$, $\{x^3, x^4\}$, $\{x^3, x^5\}$. Поставим задачу бинарной классификации для выбранных пар признаков.

Математическая постановка задачи

Согласно цели данного исследования, необходимо реализовать устойчивую модель LR для классов, *почти* линейно разделимых ввиду наличия аномалий: 1) с целью упрощения ранее реализованной модели LDA; 2) с целью повышения точности обнаружения через создание алгоритма выявления аномалий. Для математической постановки задачи распознавания введём следующее определение.

Определение 2. *Пусть X – множество объектов, Y – множество допустимых ответов, а Θ – множество допустимых значений пространства параметров θ.*

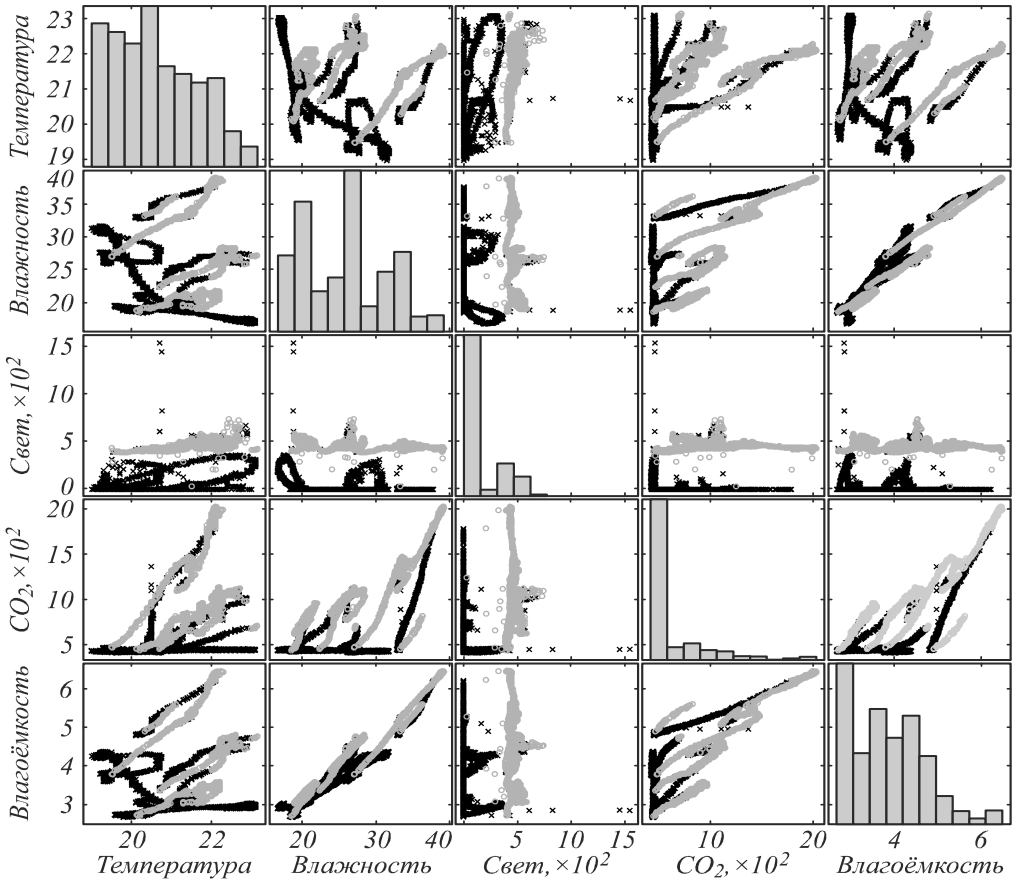

Рис. 1. Обучающая выборка в пространстве экологических признаков

Тогда в русле работы [13] *параметрическое семейство* $A=\{g(x,\theta)\,|\,\theta\in\Theta\}$, *где* $g: X\times Y\to Y$ *– фиксированная функция, называется моделью алгоритмов.*

Задача 1. *Пусть в качестве модели алгоритмов* $a\in A$ *выбрана модель логистической регрессии*

$$g(x,\theta)=1/(1+\exp(-\theta^T x)),$$

где $\theta\in\Theta\in\mathbf{R}^n$, $x\in X^l=\mathbf{R}^{n\times l}$, T – оператор транспонирования. *Тогда задача определения пространства параметров* θ *по выборке прецедентов* $X^l=(x_i,y_i)_{i=1}^l$, *где* $y_i=\{0,1\}$, *сводится к минимизации логарифмической функции потерь* $\ln L(\theta,X^l)$

$$\ln L(\theta,X^l)=-\sum_{i=1}^{l}\left[\theta^T x_i y_i-\ln\left(1+\exp(\theta^T x_i)\right)\right]. \qquad (1)$$

2. Алгоритм выявления аномалий

Гистограммы, представленные на рис. 1, позволяют сделать вывод о том, что исходные данные имеют распределение, отличное от нормального. К основным возможным причинам «ненормальности» данных можно отнести следующие: имеют экстремальные значения (выбросы); являются композицией распределений (описывают более одного процесса); содержат погрешности измерений в результате функционирования МС; отсортированы или представляют поднабор, последовательно описывающий процесс; содержат значения, близкие к нулю или предельному значению; наконец, по своей сути распределение, отличное от гауссовского. Таким образом, алгоритмы выявления аномалий направлены либо на устранение перечисленных отклонений в предположении «нормальности» данных, либо анализируют данные напрямую без приведения к нормальному распределению.

В данной работе предлагается алгоритм выявления аномалий с допущением, что распределение исходной последовательности может быть приближено к гауссовскому. Хотя подход к решению задачи является общим, новизну алгоритма определяет адаптация под специфику задачи, которая требует разрешения компромисса между наличием выбросов и точностью распознавания при условии линейной разделимости классов и устойчивости модели логистической регрессии.

Ввиду линейной разделимости классов, более целесообразно построить алгоритм как метод обучения по прецедентам [13, 19], так как очевидным выбором для разметки классов будет установка пороговой функции. Использование методов обучения без предварительной разметки [20] может привести к необоснованному усложнению алгоритма.

Анализируя причины аномалий в исходных данных, можно сделать вывод о наличии как выбросов, так и значений, близких к нулю и предельному значению (рис. 1), которые необходимо исключить. Кроме того, для дальнейшего обучения на скользящем контроле алгоритм должен производить нормирование данных при определении параметров многомерного гауссовского распределения

$$p(x,\mu,\Sigma)=\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right),$$

где $\mu\in\mathbf{R}^n$, $\Sigma\in\mathbf{R}^{n\times n}$. В процессе обучения необходимо оценить значение ε для отбрасывания аномальных значений согласно условию $p(x,\mu,\Sigma)<\varepsilon$ и значение ковариационной матрицы Σ на дополнительной диагонали α, т.е.

$$\Sigma=\begin{bmatrix}\sigma_1^2 & \alpha\\ \alpha & \sigma_2^2\end{bmatrix}$$

для пары анализируемых признаков ($n=2$) с линейно разделимыми классами.

При обучении модели логистической регрессии на этапе решения задачи обнаружения требуется произвести денормирование признаков. Хотя обучение на нормированных признаках повышает скорость сходимости при минимизации функции потерь (1), согласно результатам предварительных вычислительных экспериментов, нормирование приводит к худшим результатам на распознавании [27].

На этапе обучения необходимо использовать метрику F_1 для оценивания точности классификации. Данная метрика является гармоническим средним точности (*precision*) и полноты (*recall*) [28] и более целесообразна для несбалансированных выборок – как правило, при разметке класс с аномальными значениями значительно меньше [28, 29].

Наконец, алгоритм выявления аномалий должен содержать этап приведения распределения вероятностей исходных данных к нормальному. С этой целью в работе предлагается использовать преобразование Кокса–Бокса [30] в следующей модификации

$$x^{\lambda}=\begin{cases}((x+\beta)^{\lambda}-1)/\lambda, & \lambda\neq 0;\\ \ln(x+\beta), & \lambda=0,\end{cases}\qquad(2)$$

где λ – параметр преобразования, β – параметр смещения. Преобразование Кокса–Бокса считается наиболее подходящим, когда неизвестен тип распределения исходной выборки. При этом параметр λ может выбираться исходя из максимума логарифма правдоподобия или максимума величины коэффициента корреляции между квантилями отсортированной преобразованной выборки и выборки с нормальным распределением. Приведённая модификация данного преобразования связана с требованием условия $x+\beta>0$, так как исходный метод предполагает работу с положительными величинами.

Принимая во внимание все перечисленные аспекты, построим алгоритм выявления аномалий следующим образом.

Алгоритм 1.

1. *Выбрать пары* $\{x^p, x^q\}$, *где* $\{p,q\}\subset j$, $j=\{1,n\}$ *с почти линейно разделимыми классами.*
2. *Задать пороговую функцию* $T(x^p, x^q, t)$, *где* t *– значение порога, и параметр преобразования* (2) λ.
3. *Разметить данные для каждого класса:*
   - 3.1. *для класса* $y_i=0$*:*
     - 3.1.1. *исключить нулевые и граничные значения;*
     - 3.1.2. *разметить данные с порогом* t;
   - 3.2. *для класса* $y_i=1$*:*
     - 3.2.1. *разметить данные с порогом* t.
4. *Обучить модель на размеченной выборке* x^**:*
   - 4.1. *разбить выборку на обучающую и валидационную;*
   - 4.2. *оценить параметры в предположении нормального распределения* $p(x^*,\mu,\Sigma)$ *для каждого класса на скользящем контроле:*
     - 4.2.1. *нормировать пространство признаков* x^*;
     - 4.2.2. *преобразовать согласно* (2) $x^{\lambda}*$;
     - 4.2.3. *определить* μ,Σ;
     - 4.2.4. *оценить* ε *и* α *на метрике* F_1 *при условии* $p(x^{\lambda}*,\mu,\Sigma)<\varepsilon$;
     - 4.2.5. *построить плотность распределения* $x^{\lambda}*$ *и определить аномальные значения;*
     - 4.2.6. *денормировать пространство* $x^{\lambda}*$;
   - 4.3. *выявить аномальные значения и сформировать новый набор.*
5. *Протестировать модель обучения с* ε *и* α *на тестовых наборах.*

Варьируемые параметры в приведённом алгоритме, а именно пороговая функция $T(x^p, x^q, t)$ со значением порога t и параметр преобразования Кокса–Бокса λ, могут быть также оценены на скользящем контроле. Кроме того, указанные параметры могут быть использованы для обеспечения устойчивости модели логистической регрессии при решении дальнейшей задачи распознавания. Детальное рассмотрение вычислительных аспектов находится за пределами данной работы, но представляет интерес для дальнейших исследований.

3. Вычислительные эксперименты

Рассмотрим работу предложенного алгоритма выявления аномалий для повышения точности обнаружения с помощью логистической регрессии при условии линейной разделимости классов. С этой целью была создана эффективная программная реализация в системе GNU Octave 3.8.2 на MacBook Air 11 OS X EI Captain с процессором 1,3 GHz Intel Core i5 и памятью 4 GB 1600 MHz DDR3.

Для проведения сравнительного анализа с лучшими известными результатами [5] ниже приведены промежуточные этапы работы **Алгоритма 1** для пары $\{x^1, x^3\}$. На рис. 2 приведено графическое представление исходной выборки для пары $\{x^1, x^3\}$.

Ниже приведён результат работы алгоритма выявления аномалий для класса $y_i=0$ (рис. 3*а*) и для класса $y_i=1$ (рис. 3*б*).

Графические интерпретации были построены в результате работы **Алгоритма 1** со следующими параметрами: $t=400$, $\lambda=0$, $\beta_{y_i=0}=4{,}6$, $\beta_{y_i=1}=10$, $(\alpha,\varepsilon)_{y_i=0}=(0{,}008,\ 0{,}8080)$, $(\alpha,\varepsilon)_{y_i=1}=(0{,}007,\ 0{,}1242)$.

В табл. 2 представлены результаты решения задачи классификации для пар признаков $(p,q)\in\{(1,3),(2,3),(3,4),(3,5)\}$ с применением алгоритма выявления аномалий и без. Из табл. 2 следуют следующие выводы: 1) лучшее значение точности (выделено

жирным) получено для пары $\{x^1, x^3\}$, что не противоречит результатам предыдущих исследований [5]; 2) точность классификации, полученная для реализованной в данной работе модели LR (98,0 % для X^{k_1} и 99,35 % для X^{k_2}) сравнима с представленными для LDA в [5] (97,9 % для X^{k_1} и 99,33 % для X^{k_2}). Таким образом, предложенная эффективная программная реализация модели LR позволяет получить близкие по точности результаты классификации, но с применением более простой модели.

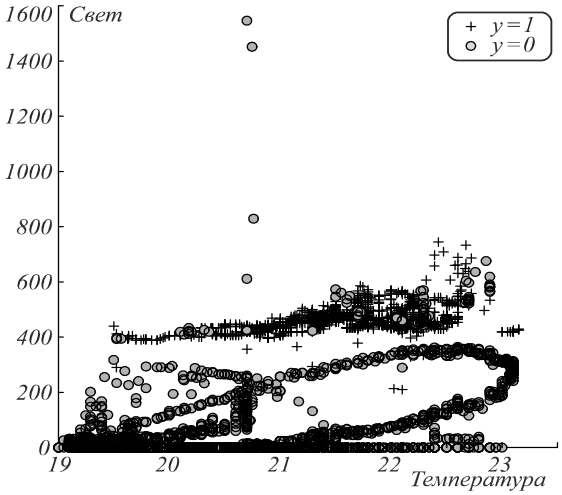

Рис. 2. Графическое представление для пары $\{x^1, x^3\}$

Табл. 2. Точность классификации без использования алгоритма выявления аномалий

(p,q)	X^l	X^{k_1}	X^{k_2}
(1,3)	**98,854**	**97,998**	**99,349**
(2,3)	94,768	97,898	98,205
(3,4)	91,759	97,935	94,770
(3,5)	93,908	97,110	95,365

Полученный результат может быть улучшен с помощью разработанного алгоритма выявления аномалий (табл. 3).

Табл. 3. Точность классификации с использованием алгоритма выявления аномалий

(p,q)	X^l	X^{k_1}	X^{k_2}
(1,3)	**100,00**	**99,960**	**99,956**
(2,3)	96,578	99,950	99,945
(3,4)	95,089	99,435	99,269
(3,5)	95,530	99,569	98,554

Применение алгоритма позволило повысить точность для всех пар признаков (p,q). Как ожидалось, наилучший результат получен для пары признаков $\{x^1, x^3\}$: на 1,96 % выше на наборе X^{k_1}; на 0,6 % выше на наборе X^{k_2}. Следует также обратить внимание на то, что значения точности на обучающей выборке для пар $(p,q) \in \{(2,3), (3,4), (3,5)\}$ ниже, чем на тестовых выборках. Данный результат может быть интерпретирован следующим образом: обучающая выборка комбинирует сценарии, когда дверь в помещение открыта и закрыта, тогда как тестовые рассматривают лишь один из предложенных сценариев. Следовательно, в зависимости от совокупности анализируемых признаков, влияние пропорции данных сценариев в обучающей выборке на конечный результат различно. Более точная калибровка свободных параметров алгоритма позволит исключить данный феномен.

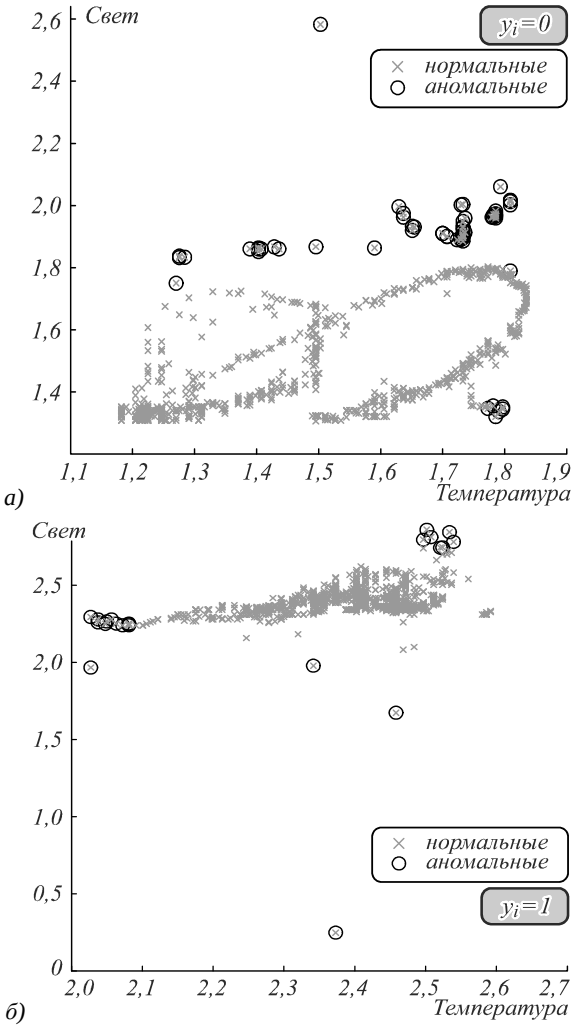

Рис. 3. Выявление аномальных значений для каждого класса

Заключение

В результате проведённых исследований:

1) предложена реализация более простой модели классификатора на основе LR, которая показывает результаты, сравнимые с лучшими для анализируемого набора;

2) создан алгоритм выявления аномалий, обеспечивающий устойчивость модели LR для *почти* линейно разделимых классов;

3) улучшены известные результаты на 1,96 % на наборе X^{k_1} и на 0,6 % на наборе X^{k_2} с помощью предложенного алгоритма выявления аномалий.

Таким образом, решена поставленная задача и достигнута цель исследования.

Дальнейшие направления исследований связаны с апробацией и адаптацией предложенного алгоритма выявления аномалий для расширенного набора эколо-

гических факторов. Значительный интерес также представляет задача многоклассовой классификации, нацеленная на более детальное распознавание признаков живого объекта и обнаружение группы объектов.

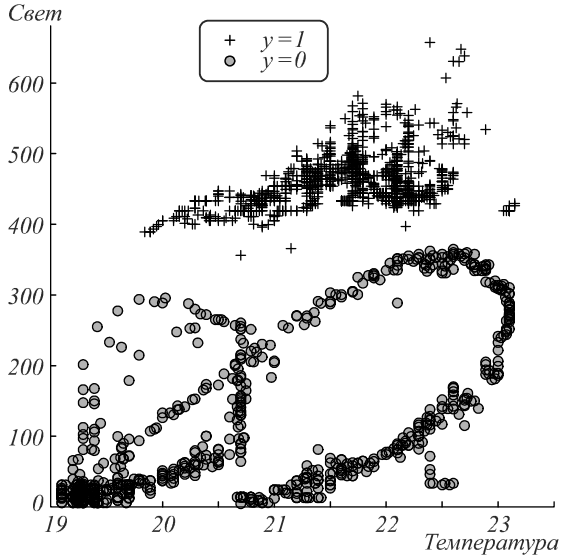

Рис. 4. Результат работы Алгоритма 1

Благодарности

Работа выполнена при государственной поддержке Министерства образования и науки РФ (грант № 074-U01). Автор выражает благодарность д.т.н., профессору С.А. Прохорову и к.ф.-м.н., профессору Л.П. Усольцеву, а также рецензентам за ценные замечания и рекомендации, способствующие повышению качества представления результатов исследований.

Литература

1. **Dulski, R.** Concept of data processing in multi-sensor systems for perimeter protection / R. Dulski, M. Kastek, P. Trzaskawka, T. Piątkowski, M. Szustakowski, M. Życzkowski // Proceedings of SPIE. – 2011. – Vol. 8019. – 8019X. – DOI: 10.1117/12.883965.
2. **Епифанцев, Б.Н.** Мультисенсорные системы мониторинга территорий ограниченного доступа: возможности видеоаналитического канала обнаружения вторжений / Б.Н. Епифанцев, А.А. Пятков, С.А. Копейкин // Компьютерная оптика. – 2016. – Т. 40, № 1. – С. 121-129. – DOI: 10.18287/2412-6179-2016-40-1-121-129.
3. **Souček, B.** Computers in Neurobiology and Behaviour / B. Souček, A.D. Carlson. – New York: John Wiley & Sons, 1976. – 324 p. – ISBN: 978-0471813897.
4. **Souček, B.** Event-train correlation and real-time microcomputer systems / B. Souček, S. Prokhorov // Microprocessing and Microprogramming. – 1983. – Vol. 11, Issue 1. – P. 23-29. - DOI: 10.1016/0165-6074(83)90161-8.
5. **Candanedo, L.M.** Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models / L.M. Candanedo, V. Feldheim // Energy and Buildings. – 2016. – Vol. 112. – P. 28-39. - DOI: 10.1016/j.enbuild.2015.11.071.
6. **Chen, Z.** A fusion framework for occupancy estimation in office buildings based on environmental sensor data / Z. Chen, M.K. Masood, Y.C. Soh // Energy and Buildings. – 2016. – Vol. 133. – P. 790-798. - DOI: 10.1016/j.enbuild.2016.10.030.
7. **Erickson, V.L.** OBSERVE: Occupancy-based system for efficient reduction of HVAC energy / V.L. Erickson, M.Á. Carreira-Perpiñán, A.E. Cerpa // Proceedings of 10th IEEE International Conference on Information Processing in Sensor Networks (IPSN), Stockholm, Sweden. – 2011. - P. 258-269.
8. **Erickson, V.L.** Occupancy modeling and prediction for building energy management / V.L. Erickson, M.Á. Carreira-Perpiñán, A.E. Cerpa // ACM Transactions on Sensor Networks (TOSN). – 2014. – Vol. 10, Issue 3. - 42. - DOI: 10.1145/2594771.
9. **Dong, B.** Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings [Electronical Resource] / B. Dong, B. Andrews. – 2014. – URL: www.ibpsa.org/proceedings/BS2009/BS09_1444_1451.pdf (Request Date 28.11.2016).
10. **Lam, K.P.** Occupancy detection through an extensive environmental sensor network in an open-plan office building / K.P. Lam, M. Höynck, B. Dong, B. Andrews, Y.-S. Chiou, R. Zhang, D. Benitez, J. Choi // Proceedings of Building Simulation 09, an IBPSA Conference, Glazgow, Scotland, July 27-30, 2009. – 2009. – P. 1452-1459.
11. **Hailemariam, E.** Real-time occupancy detection using decision trees with multiple sensor types / E. Hailemariam, R. Goldstein, R. Attar, A. Khan // Proceedings of 2011 Symposium on Simulation for Architecture and Urban Design, Boston, MA, USA. – 2011. – P. 141-148.
12. **Yang, Z.** A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations / Z. Yang, N. Li, B. Becerik-Gerber, M. Orosz // Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design, Orlando, FL, USA. – 2012. – P. 49-56.
13. **Воронцов, К.В.** Математические методы обучения по прецедентам (теория обучения машин) [Электронный ресурс] / К.В. Воронцов. – URL: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf (дата обращения 28.11.16).
14. **Brax, A.** An ensemble approach for increased anomaly detection performance in video surveillance data / A. Brax, L. Niklasson, R. Laxhammar // Proceedings of 12th International Conference on Information Fusion, Seattle, WA, USA. – 2009. – P. 694-701.
15. **Saini, D.K.** Techniques and challenges in building intelligent systems: anomaly detection in camera surveillance / D.K. Saini, D. Ahir, A. Ganatra // Proceedings of 1st International Conference on Information and Communication Technology for Intelligent Systems: Volume 2, Smart Innovation, Systems and Technologies 51. – 2016. – Part 1. – P. 11-21. – DOI: 10.1007/978-3-319-30927-9_2.
16. **Gashnikov, M.V.** Hyperspectral remote sensing data compression and protection / M.V. Gashnikov, N.I. Glumov, A.V. Kuznetsov, V.A. Mitekin, V.V. Myasnikov, V.V. Sergeev // Computer Optics. – 2016. – Vol. 40(5). – P. 689-712. – DOI: 10.18287/2412-6179-2016-40-5-689-712.
17. **Xu, Y.** Anomaly detection in hyperspectral images based on low-rank and sparse representation / Y. Xu, Z. Wu, J. Li, A. Plaza, Z. Wei // IEEE Transactions on Geoscience and Remote Sensing. – 2016. – Vol. 54(4). – P. 1990-2000. – DOI: 10.1109/TGRS.2015.2493201.
18. **Zhao, R.** Beyond background feature extraction: An anomaly detection algorithm inspired by slowly varying signal analysis / R. Zhao, B. Du, Li. Zhang, Le. Zhang // IEEE Transactions on Geoscience and Remote Sensing. – 2016. – Vol. 54(3). – P. 1757-1774. – DOI: 10.1109/TGRS.2015.2488285.
19. **Görnitz, N.** Toward supervised anomaly detection / N. Görnitz, M. Kloft, K. Rieck, U. Brefeld // Journal of Ar-

tificial Intelligence Research. – 2013. – Vol. 46. – P. 235-262. - DOI: 10.1613/jair.3623.
20. **Goldstein, M.** A Comparative evaluation of unsupervised anomaly detection algorithms for multivariate data / M. Goldstein, S. Uchida // PLoS One. – 2016. – Vol. 11(4). - e0152173 – DOI: 10.1371/journal.pone.0152173.
21. **Петренко, С.А.** Методы обнаружения вторжений и аномалий функционирования киберсистем / С.А. Петренко // Труды ИСА РАН. – 2009. – Т. 41. – С. 194-202.
22. **Kiryati, N.** Real time abnormal motion detection in surveillance video / N. Kiryati, T.R. Raviv, Y. Ivanchenko, S. Rochel // Proceedings of 19th Conference on Pattern Recognition (ICPR 2008), Tampa, Florida, USA. – 2008. – DOI: 10.1109/ICPR.2008.4761138.
23. **Buch, N.** Local feature saliency classifier for real-time intrusion moitoring / N. Buch, S. Velastin // Optical Engineering. – 2014. – 53(7). – 073108. – DOI: 10.1117/1/OE.53.7.073108.
24. **Неделько, В.М.** Регрессионные модели в задаче классификации / В.М. Неделько // Сибирский журнал индустриальной математики. – 2014. – Т. 17, № 1. – С. 86-98.
25. **Hastie, T.** The elements of statistical learning: Data mining, inference, and prediction / T. Hastie, R. Tibshirani, J. Friedman. – 2nd ed. – New York, NY, USA: Springer Science+Business Media, 2013. – 745 p. – ISBN: 978-0387848570.
26. UCI Machine Learning Repository. Occupancy Detection Data Set [Electronical Resource]. – URL: https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+ (Request Date 28.11.2016).
27. **Куликовских, И.М.** Формирование пространства признаков для обнаружения живых объектов в здании на основе экологических факторов / И.М. Куликовских // Известия Самарского научного центра РАН. – 2016. – Т. 18, № 4(4). – С. 754-759.
28. **Fawcett, T.** An introduction to ROC analysis / T. Fawcett // Pattern Recognition Letters. – 2006. – Vol. 27, Issue 8. – P. 861-874. - DOI: 10.1016/j.patrec.2005.10.010.
29. **Van Rijsbergen, C.J.** Information Retrieval / C.J. van Rijsbergen. – 2nd ed. – London: Butterworths-Heinemann, 1979. – 208 p. – ISBN: 978-0408709294.
30. **Box, G.E.P.** An analysis of transformations / G.E.P. Box, D.R. Cox // Journal of the Royal Statistical Society. Series B (Methodological). – 1964. – Vol. 26(2). – P. 211-252.

Сведения об авторе

Куликовских Илона Марковна работает доцентом на кафедре информационных систем и технологий Самарского национального исследовательского университета им. академика С.П. Королёва. В 2008 году окончила Самарский государственный аэрокосмический университет по специальности «Автоматизированные системы обработки информации и управления». В 2011 году защитила диссертацию на соискание степени кандидата наук по специальности «Математическое моделирование, численные методы и комплексы программ». Область научных интересов: цифровая обработка сигналов, статистическое обучение, машинное обучение, когнитивные вычисления, бихевиоризм. Имеет более 80 публикаций, среди которых 3 книги.

E-mail: *kulikovskikh.i@gmail.com* .

ГРНТИ: 28.23.20, 28.23.25
Поступила в редакцию 5 декабря 2016 г. Окончательный вариант – 7 января 2017 г.

ANOMALY DETECTION IN AN ECOLOGICAL FEATURE SPACE TO IMPROVE THE ACCURACY OF HUMAN ACTIVITY IDENTIFICATION IN BUILDINGS

I.M. Kulikovskikh [1]
[1] *Samara National Research University, Samara, Russia*

Abstract

This paper considers a problem of improving the accuracy of identifying human activity in buildings based on an ecological feature space. To solve this problem a model of logistic regression was implemented on the assumption of the unstable estimation of logistic regression parameters for near linearly separable classes. To reach a compromise between the presence of outliers and the accuracy of recognition an algorithm of anomaly detection was proposed. Computational experiments confirmed the effectiveness of the algorithm and its theoretical consistency.

Keywords: anomaly detection, logistic regression, machine learning, Cox-Box transformation, detection system, ecological feature.

Citation: Kulikovskikh IM. Anomaly detection in an ecological feature space to improve the accuracy of human activity identification in buildings. Computer Optics 2017; 41(1): 126-133. – DOI: 10.18287/2412-6179-2017-41-1-126-133.

Acknowledgements: This work was supported by the Ministry of Education and Science of the Russian Federation, grant 074-U01. The author would like to thank Dr. S. Prokhorov, Dr. L. Usoltsev, and the reviewers for the valuable comments and suggestions that led to improving the quality of presenting research results.

References

[1] Dubski R, Kastek M, Trzaskawka P, Piątkowski T, Szustakowski M, Życzkowski M. Concept of data processing in multi-sensor systems for perimeter protection. Proc SPIE 2011; 8019: 8019X. DOI: 10.1117/12.883965.
[2] Epifantsev BN, Pyatkov AA, Kopeykin SA. Multi-sensor systems for monitoring access to restricted areas: capabili-

ties of the intrusion detection video analytical channel. Computer Optics 2016; 40(1): 121-129. DOI: 10.18287/2412-6179-2016-40-1-121-129.
[3] Souček B, Carlson A.D. Computers in Neurobiology and Behaviour. New York: John Wiley & Sons; 1976. ISBN: 978-0471813897.
[4] Souček B, Prokhorov S. Event-train correlation and real-time microcomputer systems. Microprocessing and Microprogramming 1983; 11(1): 23-29. DOI: 10.1016/0165-6074(83)90161-8.
[5] Candanedo LM, Feldheim V. Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. Energy and Buildings 2016; 112: 28-39. DOI: 10.1016/j.enbuild.2015.11.071.
[6] Chen Z, Masood MK, Soh YC. A fusion framework for occupancy estimation in office buildings based on environmental sensor data. Energy and Buildings 2016; 133: 790-798. DOI: 10.1016/j.enbuild.2016.10.030.
[7] Erickson VL, Carreira-Perpiñán MA, Cerpa AE. OBSERVE: Occupancy-based system for efficient reduction of HVAC energy. Proc 10th IEEE International Conference on Information Processing in Sensor Networks (IPSN). Stockholm, Sweden 2011: 258-269.
[8] Erickson VL, Carreira-Perpiñán MA, Cerpa AE. Occupancy modeling and prediction for building energy management. ACM Transactions on Sensor Networks (TOSN) 2014; 10(3): 42. DOI: 10.1145/2594771.
[9] Dong B, Andrews B. Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings. Source: ⟨www.ibpsa.org/proceedings/BS2009/BS09_1444_1451.pdf ⟩.
[10] Lam KP, Höynck M, Dong B, Andrews B, Chiou YS, Zhang R, Benitez D, Choi J. Occupancy detection through an extensive environmental sensor network in an open-plan office building. IBPSA Building Simulation 2009: 1452-1459.
[11] Hailemariam E, Goldstein R, Attar R, Khan A. Real-time occupancy detection using decision trees with multiple sensor types. SimAUD'11 2011: 141-148.
[12] Yang Z, Li N, Becerik-Gerber B, Orosz M. A multi-sensor based occupancy estimation model for supporting demand driven HVAC operations. SimAUD'12 2012: 49-56.
[13] Vorontsov KV. Mathematical methods for supervised learning (machine learning theory) [In Russian]. Source: ⟨http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf⟩
[14] Brax A, Niklasson L, Laxhammar R. An ensemble approach for increased anomaly detection performance in video surveillance data. FUSION '09 2009: 694-701.
[15] Saini DK, Ahir D, Ganatra A. Techniques and challenges in building intelligent systems: anomaly detection in camera surveillance. 1st International Conference on Information and Communication Technology for Intelligent Systems: Volume 2, Smart Innovation, Systems and Technologies 51 2016; 1: 11-21. DOI: 10.1007/978-3-319-30927-9_2.
[16] Gashnikov MV, Glumov NI, Kuznetsov AV, Mitekin VA, Myasnikov VV, Sergeev VV. Hyperspectral remote sensing data compression and protection. Computer Optics 2016; 40(5): 689-712. DOI: 10.18287/2412-6179-2016-40-5-689-712.
[17] Xu Y, Wu Z, Li J, Plaza A, Wei Z. Anomaly detection in hyperspectral images based on low-rank and sparse representation. IEEE Trans on Geoscience and Remote Sensing 2016; 54(4): 1990-2000. – DOI: 10.1109/TGRS.2015.2493201.
[18] Zhao R, Du B, Zhang L, Zhang L. Beyond background feature extraction: An anomaly detection algorithm inspired by slowly varying signal analysis. IEEE Trans on Geoscience and Remote Sensing 2016; 54(3): 1757-1774. DOI: 10.1109/TGRS.2015.2488285.
[19] Görnitz N, Kloft M, Rieck K, Brefeld U. Toward supervised anomaly detection. Journal of Artificial Intelligence Research 2013; 46: 235-262. DOI: 10.1613/jair.3623.
[20] Goldstein M, Uchida S. A Comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. PLoS One 2016; 11(4): e0152173. DOI: 10.1371/journal.pone.0152173.
[21] Petrenko SA. Methods of intrusion detection and anomaly detection for cybersystems. Proceedings of ISA RAS 2009; 41: 194-202.
[22] Kiryati N, Raviv TR, Ivanchenko Y, Rochel S. Real time abnormal motion detection in surveillance video. ICPR 2008. DOI: 10.1109/ICPR.2008.4761138.
[23] Buch N, Velastin S. Local feature saliency classifier for real-time intrusion moitoring. Optical Engineering 2014; 53(7): 073108. DOI: 10.1117/1/OE.53.7.073108.
[24] Nedelko VM. Regression models in classification problems. Siberian Journal of Industrial Mathematics 2014; 17(1): 86-98.
[25] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning: Data mining, inference, and prediction: 2nd ed. New York, NY, USA: Springer Science+Business Media; 2013. ISBN: 978-0387848570.
[26] UCI Machine Learning Repository. Occupancy Detection Data Set. Source: ⟨https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+⟩.
[27] Kulikovskikh IM. Feature extraction to detect occupancy in buildings using ecological factors. Proceedings of SSC RAS 2016; 18(4): 754-759.
[28] Fawcett T. An introduction to ROC analysis. Pattern Recognition Letters 2006; 27(8): 861-874. DOI: 10.1016/j.patrec.2005.10.010.
[29] Van Rijsbergen CJ. Information Retrieval: 2nd ed. London: Butterworths-Heinemann; 1979. ISBN: 978-0408709294.
[30] Box GEP, Cox DR. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological) 1964; 26(2): 211-252.

Author's information

Ilona M. Kulikovskikh is an associate professor of Information Systems and Technologies department at Samara National Reseach University. In 2008 she defended her graduation work in Computer Science in SSAU. In 2011 she received her PhD in Applied Mathematics and Computer Science from SSAU. Her research interests are in the areas of signal processing, statistical learning, machine learning, cognitive computing, and behavior science. She is author of more than 80 publications. Among them are three co-authored books. E-mail: *kulikovskikh.i@gmail.com* .

Received December 5, 2016. The final version – January 7, 2017.