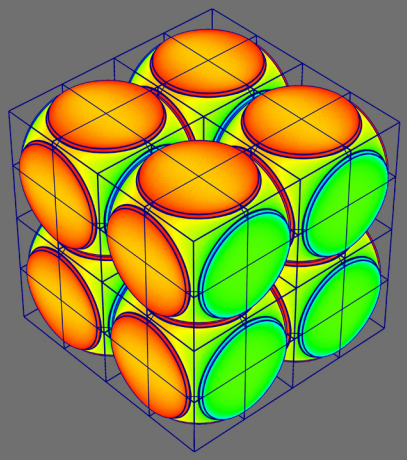
ISSN 2309-3684



Математическое моделирование и численные методы

Облакова Т.В., Касупович Э. Имплементация процедуры Кульбака-Лейблера к задаче о разладке во временных рядах различной природы. Математическое моделирование и численные методы, 2024, № 4, с. 111–127.

Источник: https://mmcm.bmstu.ru/articles/361/

Параметры загрузки:
IP: 216.73.216.253
04.11.2026 02:23:28

УДК 519.2 DOI: 10.18698/2309-3684-2024-4-111127

Имплементация процедуры Кульбака-Лейблера к задаче о разладке во временных рядах различной природы

© Т.В. Облакова, Э. Касупович

МГТУ им Н.Э. Баумана, Москва, Россия

Работа посвящена сравнительному изучению параметрических и непараметрических методов обнаружения разладки во временных рядах различной природы. Для решения задачи обнаружения аномалии в рядах с неизвестными статистическими характеристиками рассмотрена и реализована модель на основе оценки расхождения Кульбака-Лейблера между законами распределения. Для вычисления параметров линейной модели применена процедура оценки значимости Кульбака-Лейблера (KLIEP). С помощью оригинального программного кода реализован двухэтапный алгоритм решения полученной задачи условной оптимизации методом градиентного спуска, для оценки обучающей способности модели использован метод кросс-валидации. Проведен сравнительный анализ качества обнаружения момента разладки рассмотренным методом KLIEP и классической моделью кумулятивных сумм (CUSUM). При работе с моделированными данными обнаружены незначительные вариации в таких характеристиках моделей KLIEP и CUSUM, как среднее время обнаружения момента разладки и уровень ложной тревоги. Для реальной задачи обнаружения сбоя в режиме энергопотребления обе процедуры продемонстрировали 1-2 ложных тревоги, однако в KLIEP получилось более узкое временное окно (5 временных интервалов против 20), что в принципе позволяет значительно быстрее и без потери точности определять аномалию. При обнаружении разладок в ключевых показателях эффективности интернет сервиса получены схожие результаты. В целом показано, что применение модели KLIEP не ухудшает качество определения момента возникновения аномалии в сравнении с популярными моделями, использующими статистические характеристики ряда. На реальных данных продемонстрировано преимущество использования данного подхода, поскольку он не требует знания закона распределения временного ряда.

Ключевые слова: *статистическое моделирование, машинное обучение, временные ряды, задача о разладке, модель KLIEP*

Введение. Для реальных процессов, непрерывно генерирующих некоторые данные, существует необходимость в автоматическом отслеживании корректности работы программ и своевременном обнаружении разладки. Под разладкой понимается изменение статистических свойств временного ряда, начиная с некоторого момента времени. Приложения задачи о разладке весьма широки: обнаружение хакерских атак на интернет-сервисы, обнаружение аритмии и т.д. [1][2]. Основными требованиями к методам обнаружения аномалий системы являются скорость определения и полнота таких возникновений.

Временные ряды, порожденные реальными процессами, описываются сложными законами, статистические характеристики которых не всегда поддаются четкой структуризации [3,4]. В связи с этим, существует острая необходимость в разработке моделей, способных выявлять аномальное поведение системы без использования знаний о параметрах распределения. В данной работе исследуются различия реализованной модели на основе оценки значимости расхождения Кульбака-Лейблера (KLIEP) и классической процедуры вычисления кумулятивного отношения правдоподобия (CUSUM) [5-7].

Классическая постановка задачи. Рассмотрим процессы, в которых в некоторый момент θ происходит изменение статистических свойств временного ряда.

Введем гипотезы H^0 — разладка произошла в момент наблюдения θ и H^∞ — разладка не произошла вовсе. Классическая постановка задачи состоит в наблюдении бинарного статистического эксперимента [3]

$$\left(\Omega, F, \left(F_n\right)_{n\geq 0}; P^0, P^\infty\right),$$

где $\left(\Omega, F\right)$ — некоторое измеримое пространство;

$\left(F_n\right)_{n\geq 0}$ — фильтрация σ-алгебр,

$$F_0 = \left\{\varnothing, \Omega\right\}, F_0 \subseteq F_1 \subseteq \ldots \subseteq F = F_\infty;$$

P^0 — распределение вероятностей, соответствующее гипотезе H^0;

P^∞ — распределение вероятностей, соответствующее гипотезе H^∞.

Цель — построить вероятностную модель P^θ. Причем мера P^θ должна удовлетворять свойству $P^\theta(A) = P^\infty(A)$ при $0 < \theta < \infty$ если $A \in F_n$ и $n < \theta$, а также естественному условию:

$$P^\theta = \begin{cases} P^0, \theta = 0, \\ P^\infty, \theta = \infty. \end{cases} \tag{1}$$

Таким образом мера P^θ характеризует разладку в момент времени θ [5,6].

Опуская математические выкладки конструирования меры P^θ, запишем производную Радона-Никодима для момента времени θ

$$L_n^\theta = I(n < \theta)L_n^\infty + I(n \geq \theta)L_{\theta-1}^\infty \frac{L_n^0}{L_{\theta-1}^0} = \begin{cases} L_n^\infty, \theta > n, \\ L_{\theta-1}^\infty \dfrac{L_n^0}{L_{\theta-1}^0}, \theta \leq n. \end{cases} \tag{2}$$

Вычисление функций правдоподобия, входящих в формулу (2), требует знаний о законе распределения временного ряда.

Модель *KLIEP*. Пусть $D \subset \mathbb{R}^d$ — область определения случайных величин. Пусть $\left\{x_i^{tr}\right\}, i = 1...n_{tr}$ — независимые одинаково распределенные случайные величины с плотностью распределения $p_{tr}(x) > 0 \forall x \in D$ и $\left\{x_i^{te}\right\}, i = 1...n_{te}$ — независимые одинаково распределенные случайные величины с плотностью распределения $p_{te}(x)$.

Цель метода состоит в том, чтобы сразу, напрямую из данных, получить так называемую значимость Кульбака-Лейблера

$$w(x) = \frac{p_{te}(x)}{p_{tr}(x)}.$$

Процедура оценки значимости строится на основе линейной модели

$$\hat{w}(x) = \sum_{l=1}^{b} \alpha_l \varphi_l(x), \tag{3}$$

где α_l — параметры обучения, $l = 1,...b$;

$\varphi_l(x)$ — базисные функции такие, что $\varphi_l(x) \geq 0 \forall x \in D, l = 1,...b$;

b — размерность пространства базисных функций.

Оценка плотности распределения тестовой выборки тогда имеет вид [6]:

$$\hat{p}_{te}(x) = \hat{w}(x) p_{tr}(x).$$

Неизвестные параметры обучения α_l определяются минимизацией расстояния Кульбака-Лейблера между распределениями $p_{te}(x)$ и $\hat{p}_{te}(x)$ следующим образом:

$$D_{KL}\left(p_{te}(x), \hat{p}_{te}(x)\right) = \int_D p_{te}(x) \log \frac{p_{te}(x)}{\hat{w}(x) p_{tr}(x)} dx =$$
$$= \int_D p_{te}(x) \log \frac{p_{te}(x)}{p_{tr}(x)} dx - \int_D p_{te}(x) \log \hat{w}(x) dx.$$

Ввиду константности первого слагаемого, необходимо максимизировать второе слагаемое

$$J = \int_D p_{te} \log \hat{w}(x) dx \approx \frac{1}{n_{te}} \sum_{j=1}^{n_{te}} \log \hat{w}(x_j^{te}) = \frac{1}{n_{te}} \sum_{j=1}^{n_{te}} \log\left(\sum_{l=1}^{b} \alpha_l \varphi_l(x_j^{te})\right). \tag{4}$$

Поскольку $\hat{w}(x)$ — оценка значимости $w(x)$, являющейся положительной по определению, получаем ограничения на параметры

$\alpha_l \geq 0, l = 1,...b$. Дополнительно, $\hat{p}_{te}(x)$ — оценка плотности $p_{te}(x)$, нормирована на 1, что приводит к соотношению

$$1 = \int_D \hat{p}_{te}(x)dx = \int_D \hat{w}\,\hat{p}_{tr}(x)dx \approx \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \hat{w}(x_i^{tr}) = \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \sum_{l=1}^{b} \alpha_l \varphi_l(x_i^{tr})$$

Данная задача условной оптимизации называется процедурой оценки значимости Кульбака-Лейблера:

$$\begin{aligned} &\sum_{j=1}^{n_{te}} \log\left(\sum_{l=1}^{b} \alpha_l \varphi_l(x_j^{te})\right) \to \max_{\alpha_l}, \\ &\frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \sum_{l=1}^{b} \alpha_l \varphi_l(x_i^{tr}) = 1, \\ &\alpha_l \geq 0, l = 1,...b. \end{aligned} \tag{5}$$

Алгоритм численного решения задачи о разладке. Численное решение поставленной задачи достигается методом градиентного спуска [8,9], алгоритм которого представлен ниже:

1. Положим

$$A_{j,l} = \varphi_l(x_j^{te}),\ j = 1,...n_{te}, l = 1,...b.$$

2. Положим

$$b_l = \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \varphi_l(x_i^{tr}), i = 1,...n_{tr}, l = 1,...b.$$

3. Выберем $\alpha_l > 0, l = 1,...b$ и $0 < \lambda \ll 1$ — скорость градиентного спуска.
4. Шаг градиентного спуска

$$\alpha = \alpha + \lambda A^T \frac{1}{A\alpha}.$$

5. Условие сходимости

$$\alpha = \alpha + (1 - b^T\alpha)\frac{b}{b^T b},\ \alpha = \max(0, \alpha),\ \alpha = \frac{\alpha}{b^T\alpha}.$$

6. Повторяем пункты 4-5 до сходимости

$$\left\|\alpha - \alpha_{prev}\right\| < \Delta,$$

где Δ — некоторая константа.

Для оценки обучающей способности модели используется метод кросс-валидации, в результате которого выбираются лучшие параметры модели. Данный алгоритм, применительно к задаче оптимизации описан ниже:

1. Положим

$$M = \left\{ m_k \mid m_k = \left\{ \varphi_l^{(k)}(x) \right\} \right\}$$

— набор кандидатов, где k — номер кандидата.

2. Делим выборку $\left\{ x_i^{te} \right\}, i = 1...n_{te}$ на R подвыборок

$$\left\{ x_i^{te} \right\}_r, r = 1,...R .$$

3. Для каждого кандидата $m_k \in M$ выполняем пункты 4-5.

4. Применяем *KLIEP* к каждой $\left\{ x_i^{te} \right\}_r$ и считаем $\hat{J}_r$

5. Получаем усредненную оценку по всем подвыборкам

$$\hat{J} = \frac{1}{R} \sum_{r=1}^{R} \hat{J}_r .$$

6. Выбираем лучшую модель $\hat{m} = \arg\max \hat{J}(m_k)$.

7. Применяем KLIEP ко всей выборке $\left\{ x_i^{te} \right\}$ результирующей моделью $\hat{m}$.

В качестве базисных функций используются Гауссовские ядра с центрами [10-12]:

$$\hat{w}(x) = \sum_{l=1}^{n_{te}'} \alpha_l K_\sigma(x, x_l^{'te}), \tag{6}$$

где

$$K_\sigma(x, x') = \exp\left(-\frac{\left\| x - x' \right\|^2}{2\sigma^2} \right)$$

— Гауссовское ядро ширины σ.

Решение задачи обнаружения разладки реализуется в два этапа: обучение модели на исторических очищенных данных и ее последующее применение на входных данных. На этапе обнаружения для каждого момента времени θ область наблюдения делится на два окна размерами n_{te} и n_{tr}, левое и правое соответственно (рис. 1).

В моменты времени меньшие θ предполагается отсутствие разладки. Данные из этого окна имеют плотность распределения $p(x_i)$, обозначаемую $p_{te}(x_i)$. О распределении правого окна не делается никаких предположений, но при отсутствии разладки должно выполняться $p_{tr}(x) = \hat{w}(x) p_{te}(x)$. Задача таким образом сводится к проверке гипотезы об однородности распределения против его изменения.

$$H_0 : p(x_i) = p_{te}(x_i), \theta - n_{te} < i \le \theta + n_{tr},$$

$$H_1 : \begin{cases} p(x_i) = p_{te}(x_i), \theta - n_{te} < i \le \theta, \\ p(x_i) = p_{tr}(x_i), \theta < i \le \theta + n_{tr}, \end{cases}$$

где $p_{te}(x_i)$ — плотность распределения до аномалии;

$p_{tr}(x_i)$ — плотность распределения после аномалии.

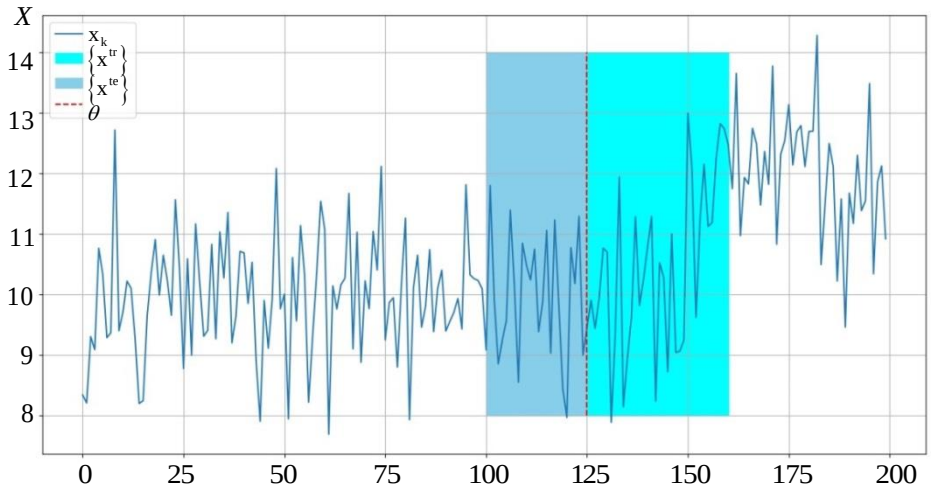


Рис. 1. Визуализация скользящих окон

Согласно выведенному ранее отношению правдоподобия определим накопленные за правое окно логарифмы этого отношения как функцию принятия решения

$$g_\theta = \sum_{i=\theta+1}^{\theta+n_{tr}} \ln \frac{p_{te}(x_i)}{p_{tr}(x_i)} = \sum_{i=\theta+1}^{\theta+n_{tr}} \ln w(x_i). \qquad (7)$$

Тогда θ будет считаться моментом разладки, если g_θ значимо отличается от усредненного значения статистик g_i из левого окна, где значимость определяется некоторой константой. Данное заключение записывается в виде условия следующим образом:

$$\left[\begin{array}{l} g_\theta < \dfrac{1}{n_{te}} \sum\limits_{i=\theta-n_{te}}^{\theta} g_i - T \\ g_\theta > \dfrac{1}{n_{te}} \sum\limits_{i=\theta-n_{te}}^{\theta} g_i + T \end{array} \right.,$$

где $T > 0$ — некоторый порог значимости.

Для неизвестных T и n_{tr} строится сетка параметров, из которого выбирается пара, удовлетворяющая критериям: наименьшее число ложноположительных обнаружений разладки, минимальный уровень порога T и минимальный размер окна n_{tr}.

Сравнительный анализ модели KLIEP на синтетических данных. Перед применением модели к реальным данным полезно проверить гипотезу на синтетических данных и получить метрики, по которым можно сделать выводы о реальной применимости модели. В качестве базовой модели предлагается взять классический метод CUSUM [5]. Алгоритм генерации данных выражается в следующих шагах:

1) Сгенерировать число разладок n;

2) Для каждого $\{i \mid 0 <= i <= n\}$ сгенерировать траекторию из случайного закона распределения;

3) Объединить полученные траектории в единый набор данных;

4) Применить модель обнаружения разладок к полученному ряду.

Таким образом будет эмулироваться изменение статистических свойств временного ряда и проверяться применимость модели KLIEP. На конкретном примере продемонстрируем применение моделей. Сгенерируем 3 момента разладки, обозначаемые красными вертикальными линиями на рисунке ниже.

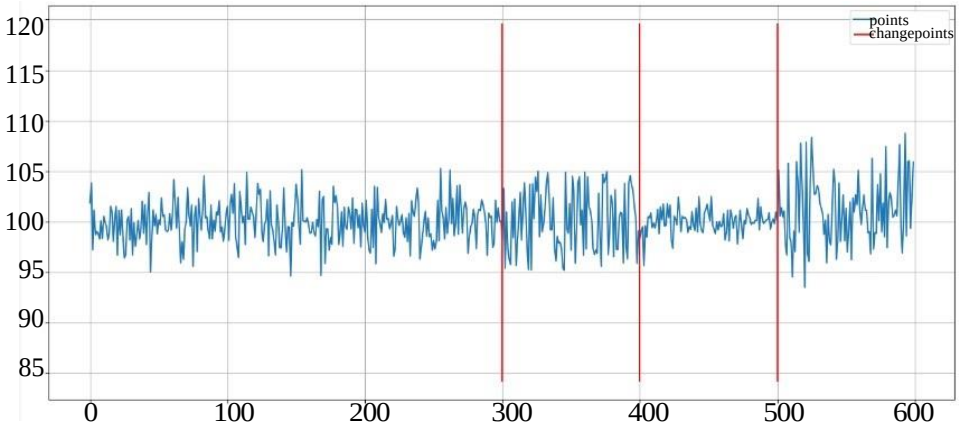


Рис. 2. Траектория из последовательных *N(100,2)*, *U(95,105)*, *100+t(11)* и *N(101,3)* законов распределения

Применяя метод CUSUM к сгенерированной траектории получаем следующие обнаружения разладок (рис. 3).

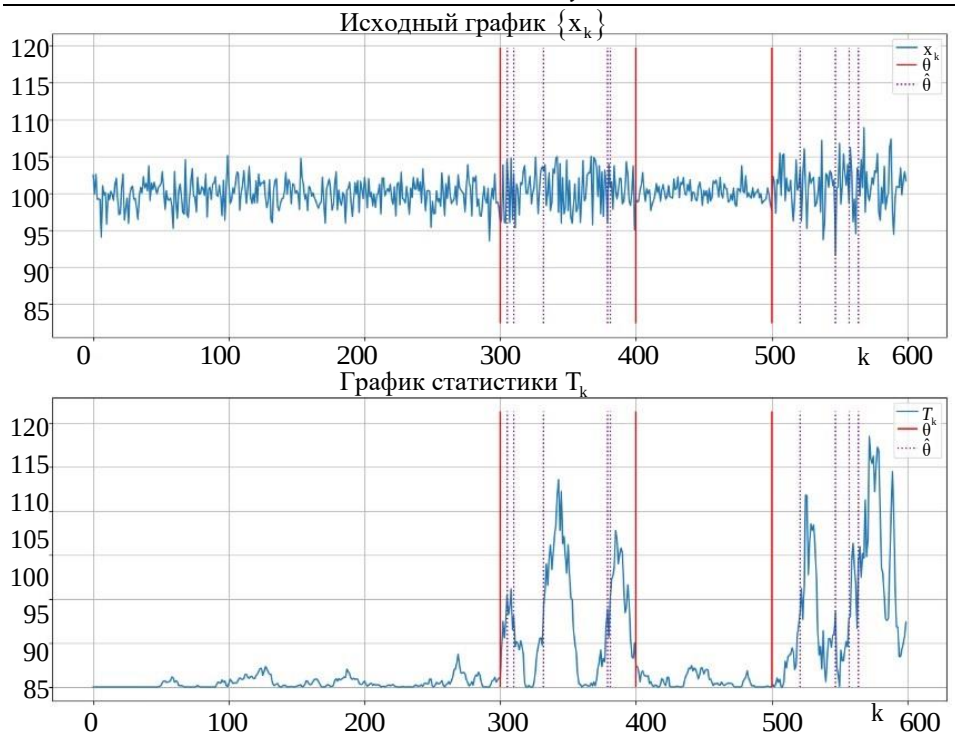


Рис. 3. Результат обнаружения разладок методом CUSUM на смоделированной траектории

В процессе обучения модели KLIEP были определены гиперпараметры модели: $\sigma = 125$ — Гауссовская ширина, $b = 17$ — размер пространства базисных функций, $T = 0.003$ — порог значимости и $n_{tr} = 25$ — размер окна определения аномалии. Применяя модель KLIEP с критерием (7) к аналогичному ряду, получаем результат (рис. 4).

Согласно полученному результату видно, что модели требуется больше времени на определение аномалии, однако обнаружение более полное, в отличие от метода CUSUM. Для более точного заключения смоделированы 25 различных траекторий данных и посчитаны метрики MAE и RMSE для временной задержки в обнаружении разладки и recall для определения полноты срабатываний. Результаты моделирования отражены в таблице ниже.

Таблица 1

Метрики MAE, RMSE и recall для методов CUSUM и KLIEP в задаче обнаружения разладок

Метод	MAE	RMSE	recall
CUSUM	14.713	27.638	0.646
KLIEP	23.925	43.086	0.781

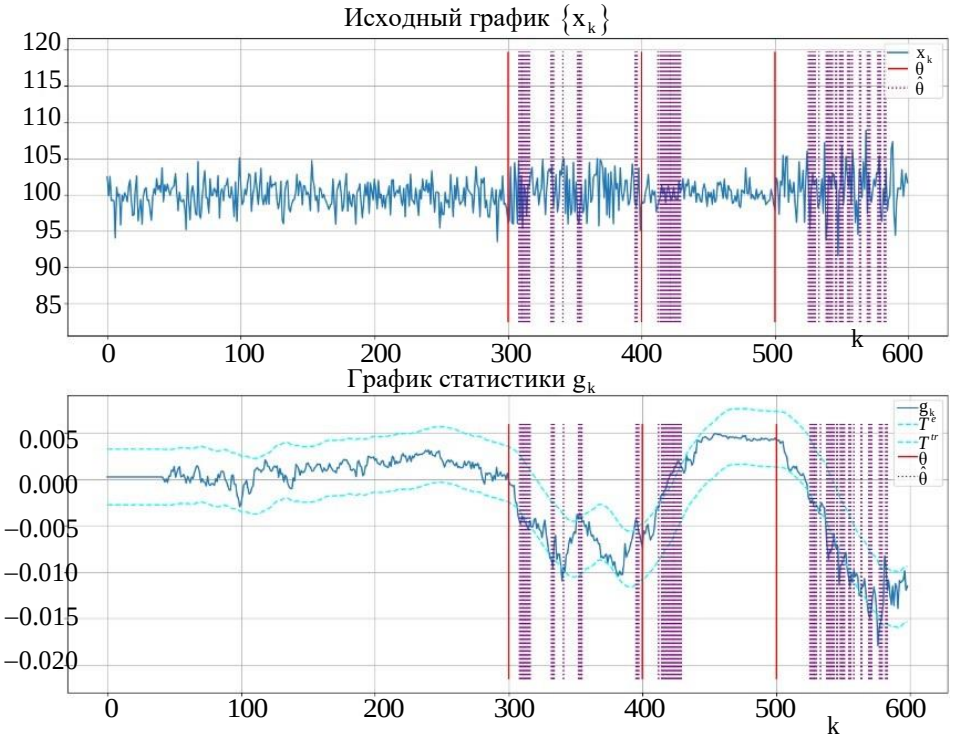


Рис. 4. Результат обнаружения разладок по смоделированной траектории моделью KLIEP

Из результатов можно сделать следующий вывод: метод CUSUM в среднем на 9.2 временных пункта быстрее обнаруживает разладку, однако полнота таких обнаружений ниже на 0.135. Задержка в обнаружении объясняется более комплексными аномалиями, которые смогла обнаружить модель KLIEP и на которые ей потребовалось больше времени.

Обнаружение разладок в энергопотреблении. В качестве данных использованы показания энергопотребления электрической нагнетательной машины, измеряемые в кВт/ч. Соответствующий график потребления можно наблюдать на рисунке ниже (рис. 5).

На рисунке видны частые всплески, характеризующие увеличение потребления энергии и пять резко выбивающихся скачков, в моменты которых произошла разладка. Применяя к этим данным модели CUSUM и KLIEP получаем следующие картины (рис. 6) и (рис. 7) соответственно.

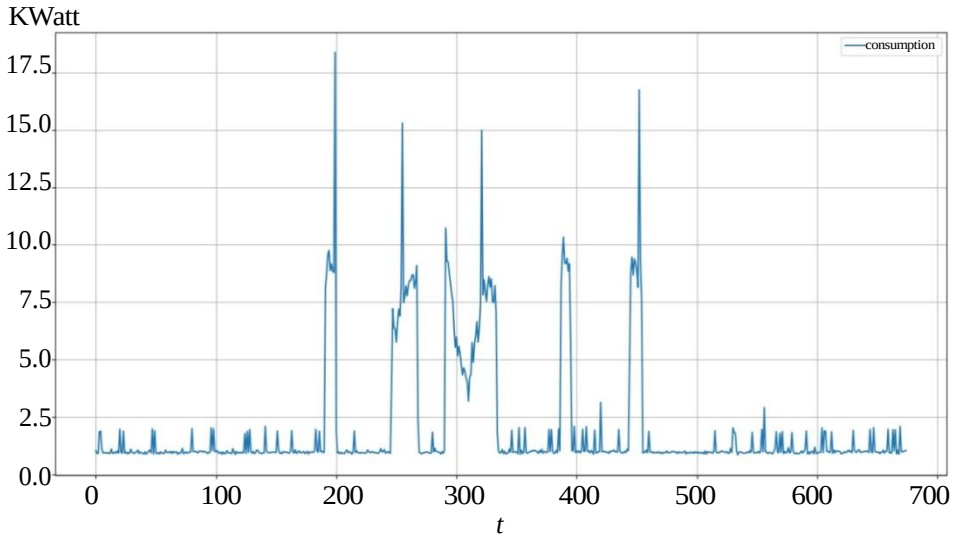


Рис. 5. График энергопотребления машины в течение некоторого времени

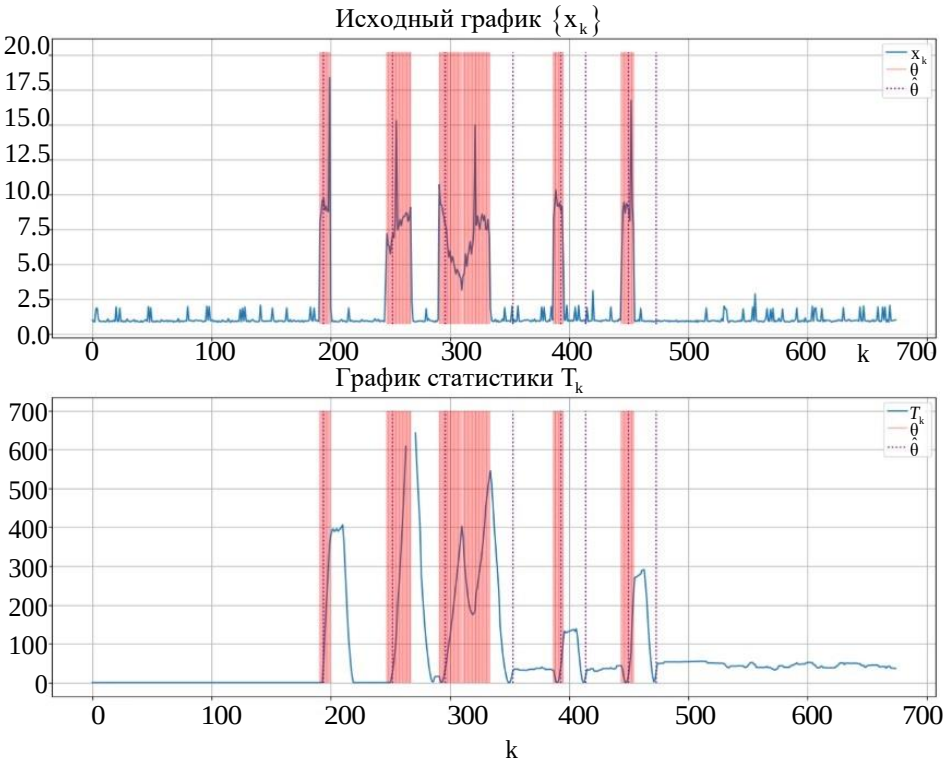


Рис. 6. Результат обнаружения разладок в энергопотреблении методом CUSUM

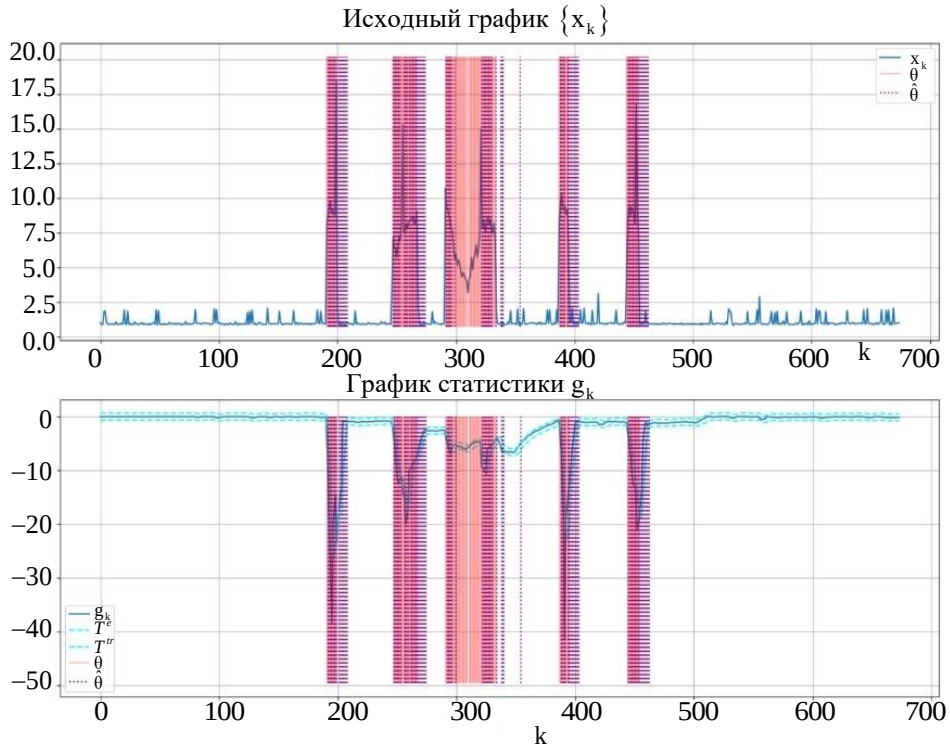


Рис. 7. Результат обнаружения разладок в энергопотреблении моделью KLIEP

Обе процедуры продемонстрировали 1-2 ложных обнаружения, однако KLIEP имеет более узкое временное окно (5 временных интервалов против 20), что позволяет ей значительно быстрее и без потери точности определять аномалию.

Обнаружение разладок в ключевых показателях эффективности интернет сервиса. С быстрым развитием интернета, особенно мобильного, веб-сервисы проникли во все сферы жизни общества. Поэтому обеспечение их стабильности становится все более важным. Персонал по эксплуатации отслеживает различные показатели эффективности (KPI) для определения стабильности сервиса, поскольку, если KPI ненормальный, то это означает, что есть проблема со связанными с ним приложениями. Основные трудности в обнаружении аномалий в KPI заключаются в низкой частотности реальных разладок и их разнообразии.

В качестве данных взят набор реальных данных о KPI различных интернет компаний (рис. 8). В них содержится временная метка (timestamp), значение показателя (value), факт наличия аномалии в данный момент времени (label) и идентификатор компании (KPI ID).

	timestamp	value	label	KPI ID
0	1493568000	1.901639	0	02e99bd4f6cfb33f
1	1493568060	1.786885	0	02e99bd4f6cfb33f
2	1493568120	2.000000	0	02e99bd4f6cfb33f
3	1493568180	1.885246	0	02e99bd4f6cfb33f
4	1493568240	1.819672	0	02e99bd4f6cfb33f

Рис. 8. Данные о KPI интернет компаний

Отображая на график данные об одной из компаний получаем следующий график (рис. 9).

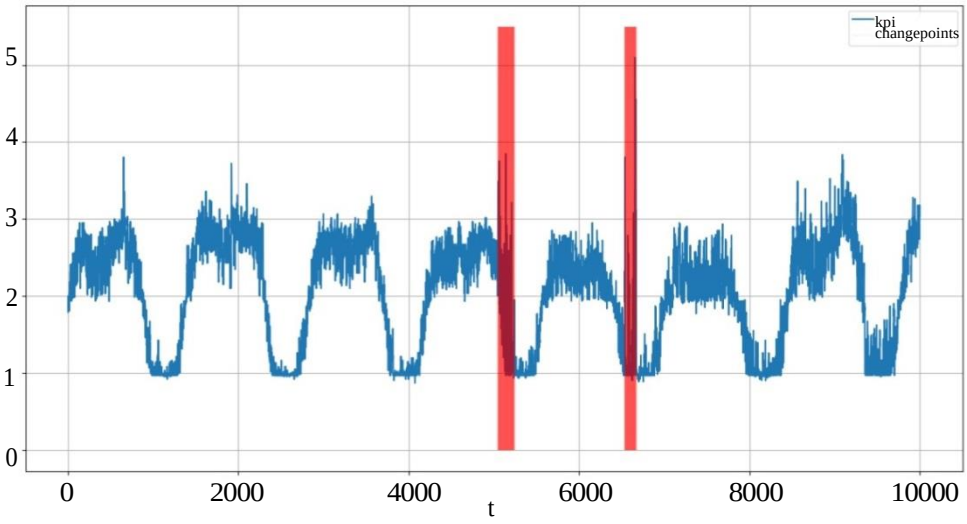


Рис. 9. KPI график с реальными разладками

Сложность обнаружения аномалий в подобных данных заключается в наличии сезонности. Поскольку на сырых данных обнаружить разладку не представляется возможным, необходимо извлечь сезонную компоненту. Результат извлечения периодичности можно увидеть на картине ниже (рис. 10).

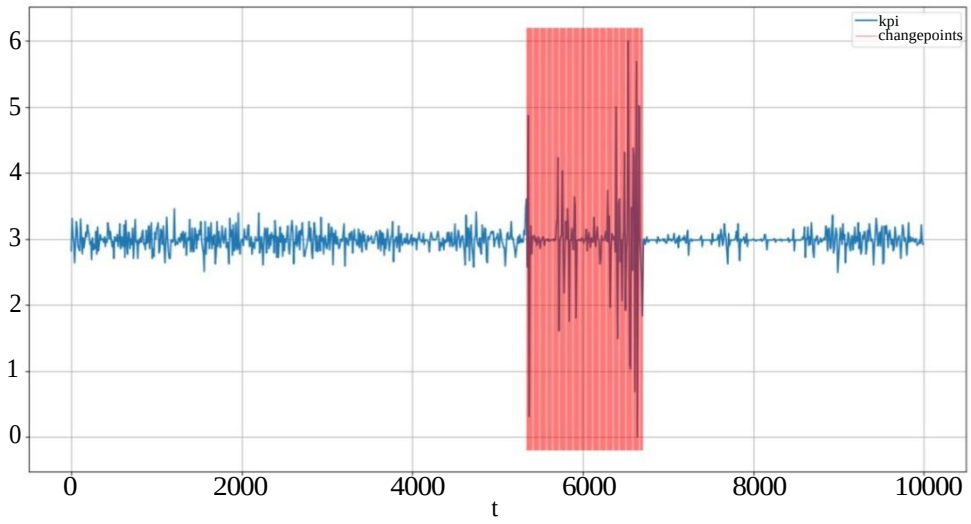


Рис. 10. Тестовая часть KPI графика с извлеченной периодической компонентой

Разделим преобразованные данные на тренировочные, без выявленной аномалии, и на тестовые, с присутствием реальной разладки. Затем обучим модель KLIEP на тренировочных данных. Лучшие результаты показала модель со следующими параметрами: $n_{tr} = 5$ — размер окна, $T = 0.3$ — порог значимости, $\sigma = 1$ — ширина гауссовского ядра, $b = 210$ — размер пространства базисных функций. На тестовых данных (рис. 11) модель обнаружила все разладки и не показала ни одного ложноположительного срабатывания.

Применяя метод CUSUM к тестовому набору получаем менее успешный результат (рис. 12). Параметры модели: размер окна 10, порог значимости 50. Метод так же обнаружил все разладки, однако после последней реальной разладки некоторое время метод показывает ложно-положительные обнаружения. Дополнительно к этому есть одно ложное срабатывание перед рядом действительных разладок.

Выводы. В ходе сравнительного анализа методов CUSUM и KLIEP было получено, что модель KLIEP предоставляет результаты не хуже классической статистической модели. При более точном сравнении на синтетических данных было обнаружено, что в среднем, модели требуется на 9.212 единиц времени больше для обнаружения разладки. Однако полнота таких обнаружений больше на 0.135, что означает, что модель чаще обнаруживает аномалию, чем не обнаруживает вовсе. На реальных данных было выявлено преимущество использования модели KLIEP, состоящее в более быстром и точном обнаружении разладки, без потери полноты. Дополнительно к этому, основное преимущество модели KLIEP состоит в том, что она не требует знания о статистических характеристиках исследуемого временного ряда.

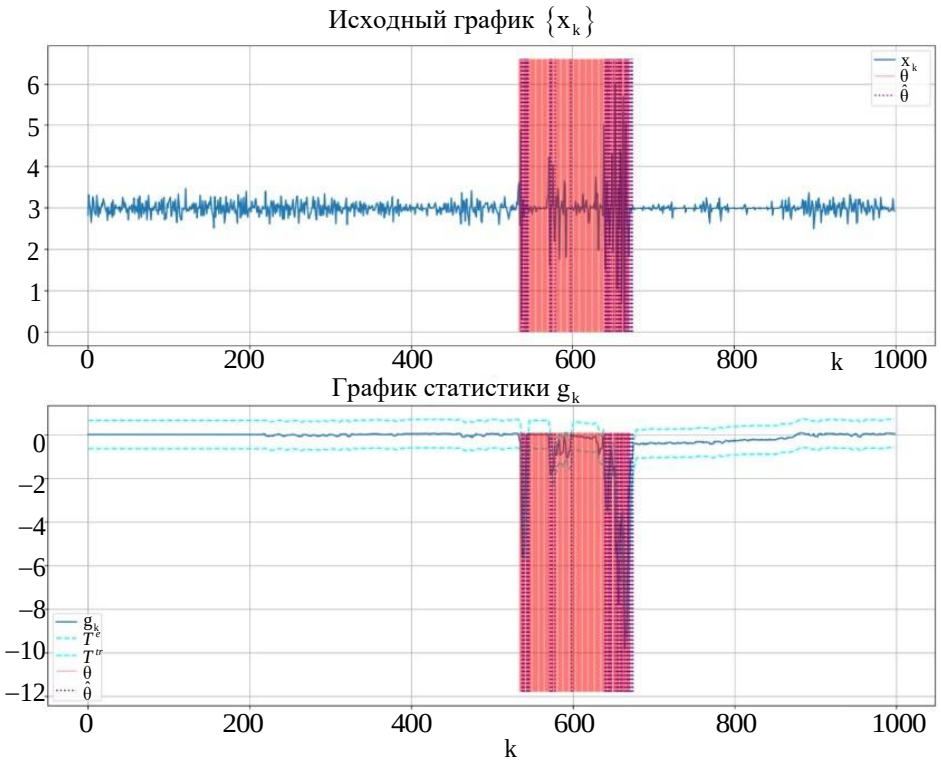


Рис. 11. Результат обнаружения разладок в KPI моделью KLIEP

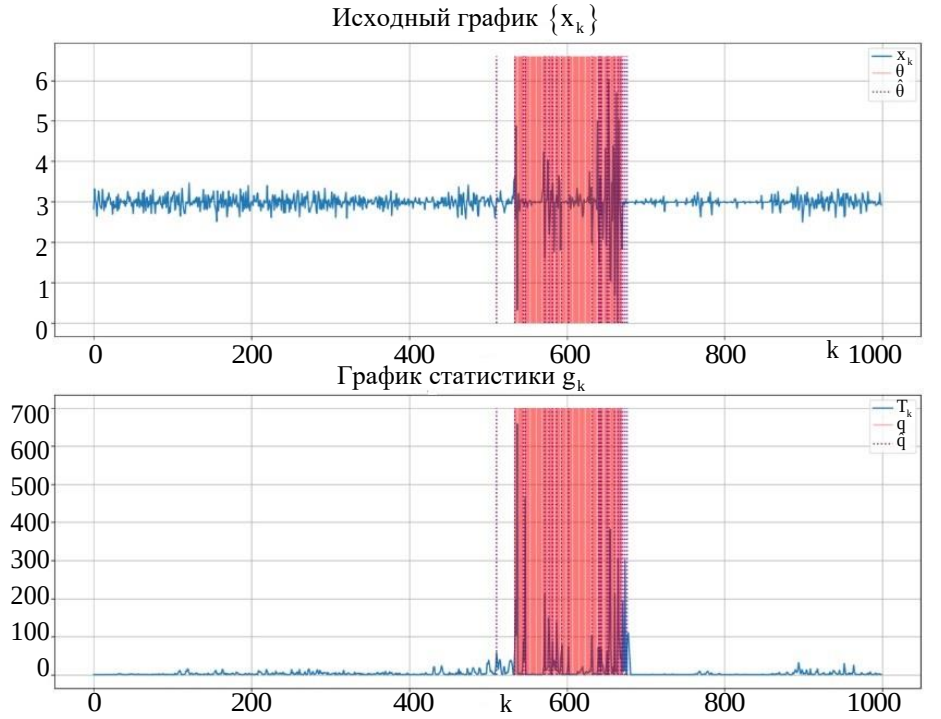


Рис. 12. Результат обнаружения разладок в KPI моделью CUSUM

Таким образом, KLIEP позволяет получить результаты не хуже, а в реальных задачах лучше, классических статистических моделей и увеличивает число областей, в которых данный метод может быть применен.

ЛИТЕРАТУРА

[1] Aminikhanghahi S., Cook D. J. A survey of methods for time series change point detection. *Knowledge and information systems,* 2017, vol. 51, No. 2, pp. 339-367.

[2] Itoh N., Kurths J. Change-point detection of climate time series by nonparametric method. *Proceedings of the world congress on engineering and computer science*, 2010, vol. 1, pp. 445-448.

[3] Облакова Т.В., Касупович Э. Численное исследование персистентных временных рядов на основе модели ARFIMA. *Математическое моделирование и численные методы*, 2022, № 4, с. 114–125.

[4] Облакова Т.В., Алексеев Д.С. Сравнительный анализ методов моделирования и прогнозирования временных рядов на основе теории фрактального броуновского движения. *Математическое моделирование и численные методы*, 2022, № 4, с. 48–62

[5] Ширяев А. *Стохастические задачи о разладке*, Litres, 2022, 393 с.

[6] Спивак В. С., Тартаковский А. Г. Байесовский подход к наибыстрейшему обнаружению разладки в частично наблюдаемых марковских случайных процессах. *Труды Московского физико-технического института*, 2021, т. 13, № 2 (50), с. 161-170.

[7] Спивак В. С. Численное сравнение наиболее популярных быстрых процедур обнаружения разладки. *Труды Московского физико-технического института*, 2020, т. 12, № 2 (46), с. 88-98.

[8] Sugiyama M. et al. Direct importance estimation for covariate shift adaptation. *Annals of the Institute of Statistical Mathematics*, 2008, vol. 60, pp. 699-746.

[9] Kanamori T., Hido S., Sugiyama M. A least-squares approach to direct importance estimation. The Journal of Machine Learning Research, 2009, vol. 10, pp. 1391-1445.

[10] Liu S. et al. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 2013, vol. 43, pp. 72-83.

[11] Desobry F., Davy M., Doncarli C. An online kernel change detection algorithm. *IEEE Transactions on Signal Processing*, 2005, vol. 53, No. 8, pp. 2961-2974.

[12] Chandola V., Vatsavai R. R. A gaussian process based online change detection algorithm for monitoring periodic time series. *Proceedings of the Eleventh SIAM International Conference on Data Mining*, 2011, pp. 95-106.

Статья поступила в редакцию 06.03.2024

Ссылку на эту статью просим оформлять следующим образом:

Облакова Т.В., Касупович Э. Имплементация процедуры Кульбака-Лейблера к задаче о разладке во временных рядах различной природы. *Математическое моделирование и численные методы,* 2024, № 4, с. 111–127.

Облакова Татьяна Васильевна — канд. физ.-мат. наук, доцент кафедры «Вычислительная математика и математическая физика» МГТУ им. Н.Э. Баумана. e-mail: obltvu@bmstu.ru

Касупович Эдин — бакалавр кафедры «Вычислительная математика и математическая физика» МГТУ им. Н.Э. Баумана. e-mail: mikkycat15@gmail.com

Implementation of the Kullback-Leibler procedure to the change point problem to time series of various nature

© T.V. Oblakova, E. Kasupovich

BMSTU, Moscow, 105005, Russia

The paper is devoted to methods to the comparative study of parametric and nonparametric methods for detecting anomaly in time series of various nature. To solve the problem of anomaly detection in series with unknown statistical characteristics, a model based on the estimation of Kulbak-Leibler divergence between distribution laws is considered and implemented. The Kullback-Leibler importance estimation procedure (KLIEP) is applied to calculate the parameters of the linear model. A two-stage algorithm for solving the obtained conditional optimization problem by the gradient descent method was implemented using the original software code, and the cross-validation method was used to evaluate the model's learning ability. A comparative analysis of the quality of fault moment detection by the considered KLIEP method and the classical cumulative sum model (CUSUM) was carried out. When working with simulated data, insignificant variations in such characteristics of KLIEP and CUSUM models as the average time of fault detection and false alarm rate were found. For the real power mode fault detection task, both procedures showed 1-2 false alarms, but KLIEP obtained a narrower time window (5 time intervals vs. 20), which in principle allows much faster and without loss of accuracy anomaly detection. Similar results were obtained when detecting discrepancies in key performance indicators of Internet service. In general, it is shown that the KLIEP model does not degrade the quality of anomaly detection compared to popular models that use statistical characteristics of the series. The advantage of using this approach is demonstrated on real data, since it does not require knowledge of the law of distribution of the time series.

***Keywords**: statistical modeling, machine learning, time series, the problem of the change point detection, the KLIEP model*

REFERENCES

[1] Aminikhanghahi S., Cook D. J. A survey of methods for time series change point detection. *Knowledge and information systems*, 2017, vol. 51, No. 2, pp. 339-367.

[2] Itoh N., Kurths J. Change-point detection of climate time series by nonparametric method. *Proceedings of the world congress on engineering and computer science*, 2010, vol. 1, pp. 445-448.

[3] Oblakova T.V., Kasupovich E. Numerical research of persistent time series based on the ARFIMA model. *Mathematical Modeling and Computational Methods*, 2022, No. 4, pp. 114–125.

[4] Oblakova T.V., Alekssev D.S. Comparative analysis of modeling methods and time series forecasting based on the theory of fractal Brownian motion. *Mathematical Modeling and Computational Methods*, 2022, No. 4, pp. 48–62

[5] Shiryaev A. *Stohasticheskie zadachi o razladke* [Stochastic discorded problems], Litres, 2022, 393 p.

[6] Spivak V. S., Tartakovsky A. G. Bayesian quickest changepoint detection approach to partially observe Markow processes. *Proceedings of Moscow Institute of Physics and Technology*, 2021, vol. 13, No. 2 (50), pp. 161-170.

[7] Spivak V. Numerical comparison of popular quickest changepoint detection procedures. *Proceedings of Moscow Institute of Physics and Technology*, 2020, vol. 12, No. 2 (46), pp. 88-98.

[8] Sugiyama M. et al. Direct importance estimation for covariate shift adaptation. *Annals of the Institute of Statistical Mathematics*, 2008, vol. 60, pp. 699-746.

[9] Kanamori T., Hido S., Sugiyama M. A least-squares approach to direct importance estimation. The Journal of Machine Learning Research, 2009, vol. 10, pp. 1391-1445.

[10] Liu S. et al. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 2013, vol. 43, pp. 72-83.

[11] Desobry F., Davy M., Doncarli C. An online kernel change detection algorithm. *IEEE Transactions on Signal Processing*, 2005, vol. 53, No. 8, pp. 2961-2974.

[1] Chandola V., Vatsavai R. R. A gaussian process based online change detection algorithm for monitoring periodic time series. *Proceedings of the Eleventh SIAM International Conference on Data Mining*, 2011, pp. 95-106.

Oblakova T.V., Cand. Sc. (Phys.-Math.), Assoc. professor of the Computational Mathematics and Mathematical Physics Department, Bauman Moscow State Technical University. e-mail: obltvu@bmstu.ru

Kasupovich E., Student of the Computational Mathematics and Mathematical Physics Department, Bauman Moscow State Technical University. e-mail: mikkycat15@gmail.com