

Аппроксимация фазовой траектории квазипериодических сигналов методом сферической регрессии

Аннотация: Решается задача аппроксимации фазовой траектории построенной по квазипериодическому временному ряду. Фазовая траектория представлена в сферической системе координат. Для ее аппроксимации используется метод сферической регрессии. Восстанавливается регрессия координат фазовой траектории на расстояние до центра координат. Учитывается зависимость от фазы квазипериодического сигнала. Находится пространство минимальной размерности, в котором фазовая траектория не имеет самопересечений с точностью до стандартного отклонения восстановленной траектории. Эксперимент проведен на двух наборах данных: показатели потребления электроэнергии в течение года и показатели акселерометра во время ходьбы.

Ключевые слова: временной ряд, фазовая траектория, сферическая регрессия, собственная размерность пространства, траекторная матрица

1 Введение

Работа посвящена аппроксимации квазипериодических временных рядов. Примерами таких рядов являются показатели акселерометра во время ходьбы, средняя температура за день в течение года, показатели потребления электроэнергии [1, 2].

Фазовая траектория временного ряда описывается его траекторной матрицей или матрицей Ганкеля. Эта матрица часто используется для анализа и прогноза временных рядов. Например, в методе анализа спектральных компонент (singular spectrum analysis, SSA) прогноз временного ряда основан на сингулярном разложении траекторной матрицы [3–5]. Так как на строки траекторной матрицы натянута траекторное пространство временного ряда, то она может использоваться для исследования его траекторного пространства. В методе сходящегося перекрестного отображения (convergent cross mapping, CCM) [6, 7] траекторные матрицы временных рядов используются для проверки наличия Липшицева отображения между фазовыми траекториями [8].

Размерность траекторного пространства может оказаться избыточна. Это приводит к неустойчивости прогностической модели и усложнению описания ряда. В этом случае необходимо производить отбор признаков [9, 10]. Снижение размерности траекторного пространства эквивалентно построению проекции фазовой траектории в некоторое его подпространство. В данной работе предлагается снизить размерность траекторного пространства с помощью сферической регрессии [11]. Согласно этому методу, необходимая информация об искомом подпространстве для выборки $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ извлекается из множества эмпирических направлений

$$\{\mathbf{x}_i - \mathbf{x}_j \mid i < j\} \quad (1)$$

и их квадратичных моментов. Элементами выборки являются точки в сферических координатах. Аналогичный подход используется в методе контурной регрессии [12]. В [13] описан

обобщенный метод сферической регрессии, где вместо множества (1) используются обобщенные эмпирические направления $\{\mathbf{x}_i - c\mathbf{x}_j \mid i < j, c > 0\}$. Преимуществами метода сферической регрессии является высокая точность и вычислительная эффективность.

Пусть для временного ряда \mathbf{s} построена проекция фазовой траектории в некоторое траекторное подпространство и задана аппроксимирующая модель. Назовем собственным подпространством траекторное подпространство минимальной размерности, при которой аппроксимирующая модель строит адекватное приближение ряда \mathbf{s} .

Фазовая траектория периодического временного ряда представляет собой пучок линий, где каждая линия описывает ряд в течение одного периода. Предлагается описывать траекторию диаметром этого пучка, а также линией, аппроксимирующей его ожидаемое значение. Ожидаемое значение пучка находится усреднением синфазных участков траектории. Диаметр пучка задается как его дисперсия вокруг среднего значения. Таким образом, если фазовая траектория не имеет самопересечений, то ее ожидаемое значение и диаметр задают тор в траекторном пространстве ряда. На рис. 1 представлен временной ряд, его фазовая траектория и ее ожидаемые значения, выделенные красным.

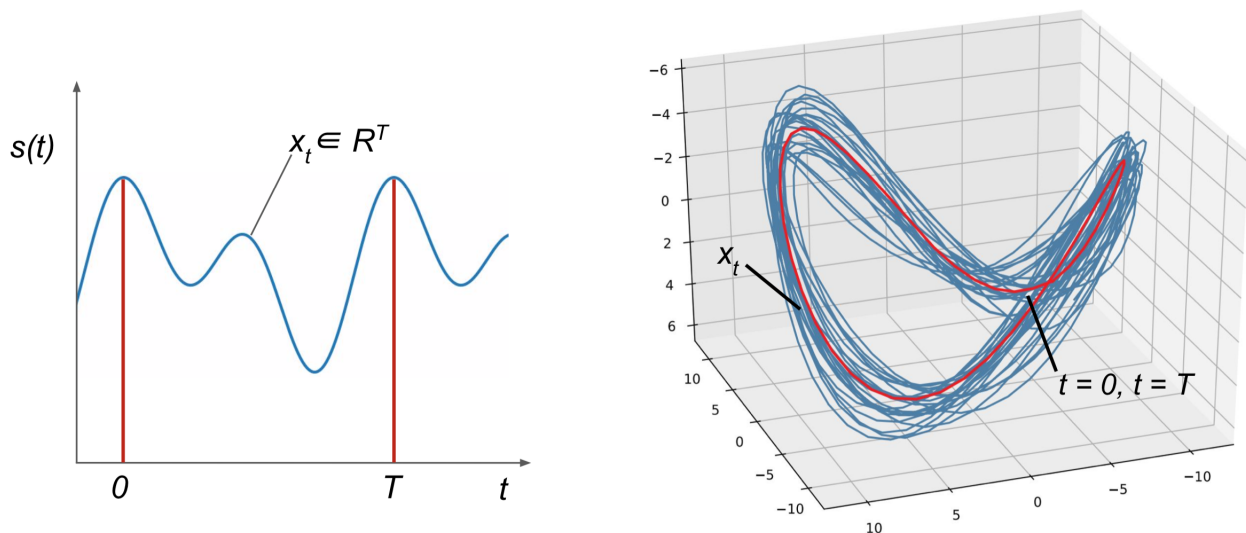


Рис. 1: Временной ряд и его фазовая траектория

Для временных рядов с переменной во времени структурой возникает задача разбиения их на сегменты, на которых их свойства не меняются [14, 15]. Для этой задачи используется описание фазовой траектории в виде тора. Если центр одной фазовой траектории статистически значительно выходит за пределы тора другой фазовой траектории на всем периоде, то делаем вывод, что эти траектории описывают различные по структуре ряды или что собственные пространства двух траекторий значительно различны.

В работе решается задача нахождения минимально необходимой размерности пространства фазовой траектории для построения адекватной аппроксимации ряда. Также исследуется наличие самопересечений фазовой траектории ряда. Наличие самопересечений означает, что ряд имеет более одного фундаментального периода. Это осложняет анализ ряда и изучение связи между рядами. Если в некотором подпространстве отсутствуют самопересечения фазовой траектории, то ряд напротив имеет простую структуру.

2 Представление квазипериодического сигнала в сферических координатах

Для заданного временного ряда $\mathbf{s} = [s_1, \dots, s_N]^\top$ построим траекторную матрицу

$$\mathbf{H}_s = \begin{bmatrix} s_1 & s_2 & \dots & s_{n-1} & s_n \\ s_2 & s_3 & \dots & s_n & s_{n+1} \\ \vdots & \vdots & & \vdots & \vdots \\ s_{N-n+1} & s_{N-n+2} & \dots & s_{N-1} & s_N \end{bmatrix}, \quad (2)$$

где n – ширина окна, не меньшая, чем предполагаемый период. Обозначим t -ю строку матрицы \mathbf{H}_s как \mathbf{s}_t . Тогда матрица \mathbf{H}_s принимает вид

$$\mathbf{H}_s = \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_m \end{bmatrix}, \quad \mathbf{s}_t = [s_t, s_{t+1}, \dots, s_{t+n-1}], \quad m = N - n + 1. \quad (3)$$

Заметим, что все векторы \mathbf{s}_t принадлежат n -мерному траекторному пространству $\mathbb{H}_s \subseteq \mathbb{R}^n$ ряда \mathbf{s} и образуют фазовую траекторию $\mathbf{x}(t) \in \mathbb{R}^n$ ряда \mathbf{s} . Предполагается, что размерность траекторного пространства избыточна. Это приводит к неустойчивости прогностических моделей ряда \mathbf{s} . Поэтому предлагается рассматривать не саму траекторию $\mathbf{x}(t)$, а ее проекцию

$$\mathbf{x}_p(t) \in \mathbb{R}^p, p \leq n,$$

в траекторное подпространство. Переход к траекторным подпространствам позволяет упростить структуру фазовой траектории и исследование ее свойств.

Для более компактного описания ряда \mathbf{x} предлагается аппроксимировать траекторию $\mathbf{x}(t)$ одной скалярной функцией от времени $\mathbf{x}^*(t)$ (см. рис. 1) и диаметром траектории. Для нахождения ожидаемого значения траектории усредним значения $\mathbf{x}(t)$ на отрезках времени $[1, T], [T + 1, 2T], \dots, [(K - 1)T + 1, KT]$:

$$\mathbf{x}^*(t) = \frac{1}{K} \sum_{k=0}^K \mathbf{x}(t + kT), \quad (4)$$

где T – период ряда \mathbf{s} , K – число целых периодов, укладывающихся в $[0, N]$. Таким образом $\mathbf{x}^*(t)$ – это матожидание пучка фазовой траектории

$$\mathbf{x}^*(t) = \mathbf{E}\mathbf{x}_t. \quad (5)$$

Диаметр траектории определим как дисперсию внутри пучка линий в траектории:

$$\mathbf{D}\mathbf{x}_t = \frac{2}{T} \sum_{t=1}^T \sqrt{\frac{1}{K} \sum_{k=1}^K \left(\mathbf{x}(t + kT) - \mathbf{x}^*(t) \right)^2}.$$

В результате траектория $\mathbf{x}(t)$ описывается с помощью матожидания $\mathbf{E}\mathbf{x}_t$ и диаметра $\mathbf{D}\mathbf{x}_t$.

Предполагается, что собственная размерность фазовой траектории $\mathbf{x}(t)$ ниже, а структура ее модели проще в сферических координатах. Построим отображение $\mathbf{x}_p(t)$ из декартовых координат в сферические:

$$\varphi : \mathbf{x}_p(t) \rightarrow \mathbf{z}_p(t) = [\alpha_1(t), \alpha_2(t), \dots, \alpha_{p-1}(t), r(t)]. \quad (6)$$

Необходимо построить регрессионную модель $f(\cdot)$, восстанавливающую переменную r по переменным $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{p-1}]$. Метод сферической регрессии позволяет понизить размерность задачи, строя отображение

$$g : \underbrace{[\alpha_1, \alpha_2, \dots, \alpha_{p-1}]}_{\boldsymbol{\alpha}} \rightarrow \underbrace{[\beta_1, \beta_2, \dots, \beta_{q-1}]}_{\boldsymbol{\beta}}, \quad q \leq p, \quad (7)$$

такое, что переменная r может быть восстановлена по $\boldsymbol{\beta}$:

$$f : \boldsymbol{\beta} \mapsto r'.$$

Предлагается исследовать зависимость сложности модели q (9) от фазовой траектории $\mathbf{x}(t)$. Сложность модели q определяется как минимальная сложность, при которой модель строит адекватную аппроксимацию

$$\mathbf{x}'_p(t) = \varphi^{-1}(\mathbf{z}'_p(t)) = \varphi^{-1}([\boldsymbol{\alpha}, r'(t)]).$$

фазовой траектории $\mathbf{x}_p(t)$ в смысле квадратичной ошибки

$$MSE(\mathbf{x}_p, \mathbf{x}'_p) = \frac{1}{N} \sum_{t=1}^N \|\mathbf{x}_p(t) - \mathbf{x}'_p(t)\|^2, \quad (8)$$

$$\min_{q \in [1, \dots, p]} : MSE(\mathbf{x}_p, \mathbf{x}'_p) < \theta. \quad (9)$$

Сложность модели q определяет размерность собственного подпространства ряда $\mathbf{s}(t)$.

3 Сферическая регрессия

В данном разделе представлено описание сферической регрессии, использующейся в отображении (7).

Пусть задана случайная независимая выборка $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)\}$ из распределения (\mathbf{X}, \mathbf{y}) , где $\mathbf{X} \in \mathbb{R}^{(p-1) \times n}$, $\mathbf{y} \in \mathbb{R}^n$. Здесь под (\mathbf{x}, y) понимаются не декартовы координаты, а сферические (6). Иными словами пара (\mathbf{x}, y) соответствует вектору

$$\mathbf{z}_p(t) = [\alpha_1(t), \alpha_2(t), \dots, \alpha_{p-1}(t), r(t)]$$

из (6). Для заданной размерности $q \leq p$ метод сферической регрессии строит проекцию \mathbf{X} в некоторое пространство $\mathbb{S}_{\mathbf{X}}^q$ размерности q . Обозначим его $\mathbb{S}_{\mathbf{X}}^q \subset \mathbb{R}^{q \times n}$.

Метод сферической регрессии находит центральное подпространство с помощью рассмотрения множества эмпирических направлений:

$$\{\mathbf{x}_i - \mathbf{x}_j \mid i < j\}.$$

Нормализуем \mathbf{X} :

$$\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu}), \quad \boldsymbol{\mu} = \mathbb{E}(\mathbf{X}), \quad \boldsymbol{\Sigma} = \text{cov}(\mathbf{X}).$$

Обозначим $(\tilde{\mathbf{Z}}, \tilde{\mathbf{y}})$ выборку из распределения (\mathbf{Z}, \mathbf{y}) , семплированную равномерно, равномошно и независимо от $\{(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_n, y_n)\}$. Введем $\mathbf{A}(\mathbf{y}, \tilde{\mathbf{y}}) \in \mathbb{R}^{n \times n}$:

$$\mathbf{A}(\mathbf{y}, \tilde{\mathbf{y}}) = \mathbb{E} \left[(\mathbf{Z} - \tilde{\mathbf{Z}})(\mathbf{Z} - \tilde{\mathbf{Z}})^\top \mid \mathbf{y}, \tilde{\mathbf{y}} \right].$$

Здесь каждое эмпирическое направление $(\mathbf{x}_i - \mathbf{x}_j)$ представлено с помощью $(\mathbf{z}_i - \tilde{\mathbf{z}}_j)(\mathbf{z}_i - \tilde{\mathbf{z}}_j)^\top$. Интуитивно направления из $\mathbf{Z} - \tilde{\mathbf{Z}}$, принадлежащие $\mathbb{S}_{\mathbf{Z}}^q$, будут сильнее зависеть от \mathbf{y} , чем направления, принадлежащие ортогональному дополнению $\mathbb{S}_{\mathbf{Z}}^q$. В [11] показано, что собственное пространство матрицы

$$\mathbf{G} = \mathbb{E} [2\mathbf{I}_p - \mathbf{A}(\mathbf{y}, \tilde{\mathbf{y}})]^2$$

является оценкой центрального подпространства $\mathbb{S}_{\mathbf{X}}^q$. Матрица \mathbf{G} представима в следующем виде:

$$\begin{aligned} \mathbf{G} = & 2\mathbb{E} \left[\mathbb{E}(\mathbf{Z}\mathbf{Z}^\top | \mathbf{y}) \right]^2 + 2\mathbb{E}^2 \left[\mathbb{E}(\mathbf{Z} | \mathbf{y}) \mathbb{E}(\mathbf{Z}^\top | \mathbf{y}) \right] + \\ & + 2\mathbb{E} \left[\mathbb{E}(\mathbf{Z}^\top | \mathbf{y}) \mathbb{E}(\mathbf{Z} | \mathbf{y}) \right] \cdot \mathbb{E} \left[\mathbb{E}(\mathbf{Z} | \mathbf{y}) \mathbb{E}(\mathbf{Z}^\top | \mathbf{y}) \right] - 2\mathbf{I}_p. \end{aligned}$$

Пусть $\Omega_{\mathbf{y}}$ — множество значений \mathbf{y} в выборке $\{(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_l, y_l)\}$. Обозначим $\{J_1, \dots, J_l\}$ — разбиение $\Omega_{\mathbf{y}}$ на подмножества. Оценка матриц \mathbf{G} и \mathbf{A} по выборке $\{(\mathbf{z}_1, y_1), \dots, (\mathbf{z}_l, y_l)\}$ делается следующим образом:

$$\begin{aligned} \hat{\mathbf{G}} &= \binom{l}{2}^{-1} \sum_{k < l} \left[2\mathbf{I}_p - \hat{\mathbf{A}}(J_k, J_{k'}) \right]^2, \text{ где} \\ \hat{\mathbf{A}}(J_k, J_{k'}) &= \frac{\sum_{i < j} (\mathbf{z}_i - \mathbf{z}_j) (\mathbf{z}_i - \mathbf{z}_j)^\top \mathbf{I}(y_i \in J_k, y_j \in J_{k'})}{\sum_{i < j} \mathbf{I}(y_i \in J_k, y_j \in J_{k'})}. \end{aligned}$$

Пусть $\lambda_1 \geq \dots \geq \lambda_p$ — собственные числа, v_1, \dots, v_p — собственные векторы матрицы \mathbf{G} . Тогда пространство $\mathbb{S}_{\mathbf{Z}}^q$ можно найти как пространство, натянутое на первые q собственных векторов матрицы \mathbf{G} , а пространство $\mathbb{S}_{\mathbf{X}}^q$ — как пространство натянутое на векторы $\Sigma^{-1/2}v_1, \dots, \Sigma^{-1/2}v_q$.

4 Анализ самопересечений фазовой траектории

Пусть для ряда $\mathbf{s}(t)$ найдено собственное подпространство \mathbb{S} с помощью сферической регрессии. Предлагается проверить, является ли проекция фазовой траектории \mathbf{x} в это подпространство самопересекающейся. Для этого необходимо рассмотреть матожидание пучка траектории $\mathbf{x}^*(t)$ и проверить, имеет ли траектория самопересечения в пределах диаметра.

Проверка самопересечений проекции фазовой траектории дает нам второе, эмпирическое определение собственного подпространства. Если оно выполняется, но при этом по (9) подпространство не является собственным, то будем увеличивать размерность подпространства. И наоборот, если найдено подпространство, удовлетворяющее (9), но при этом в нем фазовая траектория имеет самопересечения, то также необходимо увеличить его размерность.

5 Эксперимент

Целью эксперимента является анализ собственного подпространства и проверка наличия самопересечений в этом пространстве. Другими словами в эксперименте проведено сравнение определений собственного подпространств, приведенных в разделах 3 и 4. Эксперимент проведен на трех временных рядах: показатели акселерометра во время ходьбы $\mathbf{s}_1(t)$ и во время бега $\mathbf{s}_2(t)$, почасовые данные потребления электроэнергии в течение года $\mathbf{s}_3(t)$.

Построим проекции фазовых траекторий рядов s_1, s_2, s_3 в трехмерное пространство, обозначим их x_1, x_2, x_3 соответственно. Для каждой фазовой траектории определим ее матожидание.

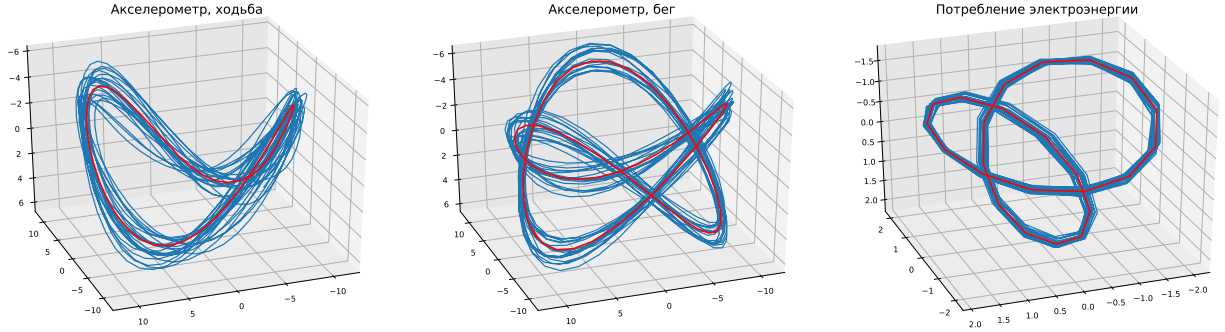


Рис. 2: Проекции фазовых траекторий и их матожидания

Для каждого ряда найдем собственное пространство удовлетворяющее (9). Для этого переберем его размерность $q \in [1, 15]$. Для каждого значения q применим сферическую регрессию и найдем ошибку 8. Полученные значения ошибки для всех трех рядов представлены на рис. 3.

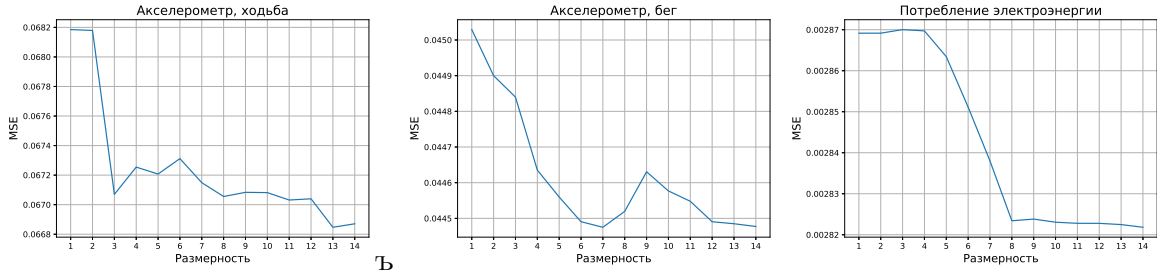


Рис. 3: Значение ошибки в зависимости от размерности модели

Таким образом для ряда s_1 собственная размерность равна 3, для ряда s_2 — 6, для ряда s_3 — 8. При этом фазовые траектории рядов s_1, s_3 не имеют самопересечений в трехмерном пространстве. Значит, согласно определениям из 3 и 4 собственная размерность s_1 равна $q_1 = 3$, а собственная размерность s_3 — $q_3 = 8$. Фазовая траектория ряда s_2 перестает самопересекаться в четырехмерном пространстве, значит $q_2 = 6$.

Для каждого из рядов построим аппроксимацию фазовой траектории моделью с найденной в предыдущем пункте размерностью. Исходные и восстановленные траектории изображены на рис. 4.

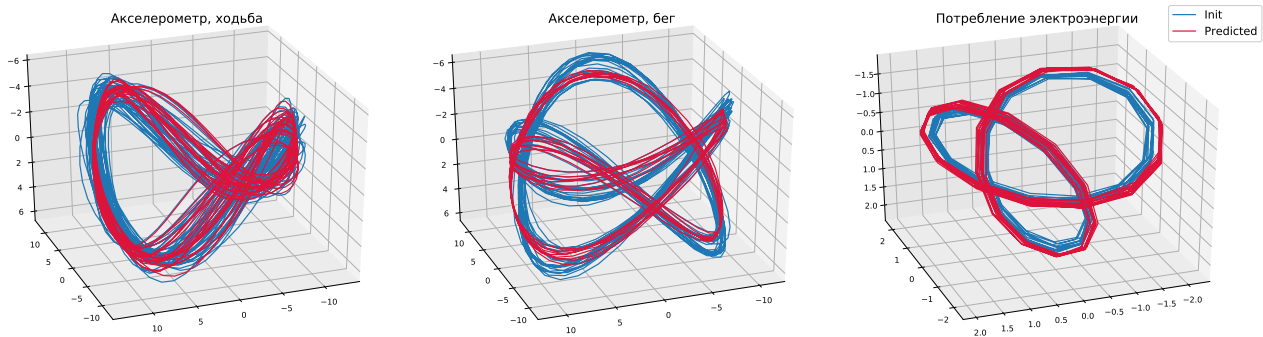


Рис. 4: Исходные и восстановленные фазовые траектории

6 Заключение

В работе решалась задача аппроксимации квазипериодических временных рядов, а также задача поиска собственного подпространства временного ряда. Для снижения размерности траекторного подпространства применялся метод сферической регрессии. В найденном подпространстве фазовая траектория дополнительно исследовалась на наличие самопересечений.

Эксперимент проводился на трех временных рядах: показатели акселерометра во время ходьбы и бега, потребление электроэнергии в течение года. Для каждого временного ряда найдено собственное подпространство и построена аппроксимирующая модель соответствующей размерности. Показано, что размерность фазовой траектории ходьбы равна трем, для бега шести, что соответствует ожиданиям экспертов.

Список литературы

- [1] M. Zhang and A. A. Sawchuk, “Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors,” in ACM International Conference on Ubiquitous Computing (UbiComp) Workshop on Situation, Activity and Goal Awareness (SAGAware), (Pittsburgh, Pennsylvania, USA), September 2012.
- [2] Данные потребления электроэнергии <https://sourceforge.net/p/mvr/code/HEAD/tree/data/EnergyConsumption.xls>
- [3] Golyandina N., Stepanov D. SSA-based approaches to analysis and forecast of multidimensional time series // 2005. Proceedings of the 5th St. Petersburg workshop on simulation. P. 293–298.
- [4] Golyandina N., Nekrutkin V., Zhigljavsky A. A. Analysis of time series structure: SSA and related techniques. – Chapman and Hall, 2002. 320 p.
- [5] Golyandina N., Zhigljavsky A. Singular Spectrum Analysis for time series. - Springer Science & Business Media, 2013. 120 p.
- [6] Sugihara G., May R., Hao Ye, Chih-hao H., Deyle E., Fogarty M., Munch S. Detecting causality in complex ecosystems // Science, 2012. Vol. 338. P. 1227079

- [7] Sugihara G., May R. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series // *Nature*, 1990. Vol. 344. No. 6268. P. 734–741.
- [8] Усманова, Карина Равиловна, Вадим Викторович Стрижов. "Модели обнаружения зависимостей во временных рядах в задачах построения прогностических моделей." *Системы и средства информатики* 29.2 (2019): 12-30.
- [9] A. Katrutsa and V. Strijov, "Stress test procedure for feature selection algorithms," *Chemometrics and Intelligent Laboratory Systems*, vol. 142, pp. 172–183, 2015.
- [10] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, p. 94, 2018.
- [11] Li, Bing and Wang, Shaoli // *On directional regression for dimension reduction.* // Taylor & Francis, 2007
- [12] Li, Bing, Hongyuan Zha, and Francesca Chiaromonte. "Contour regression: a general approach to dimension reduction." *The Annals of Statistics* 33.4 (2005): 1580-1616.
- [13] Yu, Z., Dong, Y. and Huang, M., 2014. General directional regression. *Journal of Multivariate Analysis*, 124, pp.94-104
- [14] Hušková, Marie, Zuzana Prášková, and Josef Steinebach. "On the detection of changes in autoregressive time series I. Asymptotics." *Journal of Statistical Planning and Inference* 137.4 (2007): 1243-1259.
- [15] Chen, Bin, and Yongmiao Hong. "Testing for smooth structural changes in time series models via nonparametric regression." *Econometrica* 80.3 (2012): 1157-1183.