Quasi-periodic time series clustering for human activity recognition*

A. V. Grabovoy[1], **V. V. Strijov**[2]

Annotation: This paper analyses the periodic signals in the time series to recognize human activity by using a mobile accelerometer. Each point in the timeline corresponds to a segment of historical time series. This segments form a phase trajectory in phase space of human activity. The principal components of segments of the phase trajectory are treated as feature descriptions at the point in the timeline. The paper introduces a new distance function between the points in new feature space. To reval changes of types of the human activity the paper proposes an algorithm. This algorithm clusters points of the timeline by using a pairwise distances matrix. The algorithm was tested on synthetic and real data. This real data were obtained from a mobile accelerometer.

Key words: time series; clustering; segmentation; recognition of physical activity; principal component method.

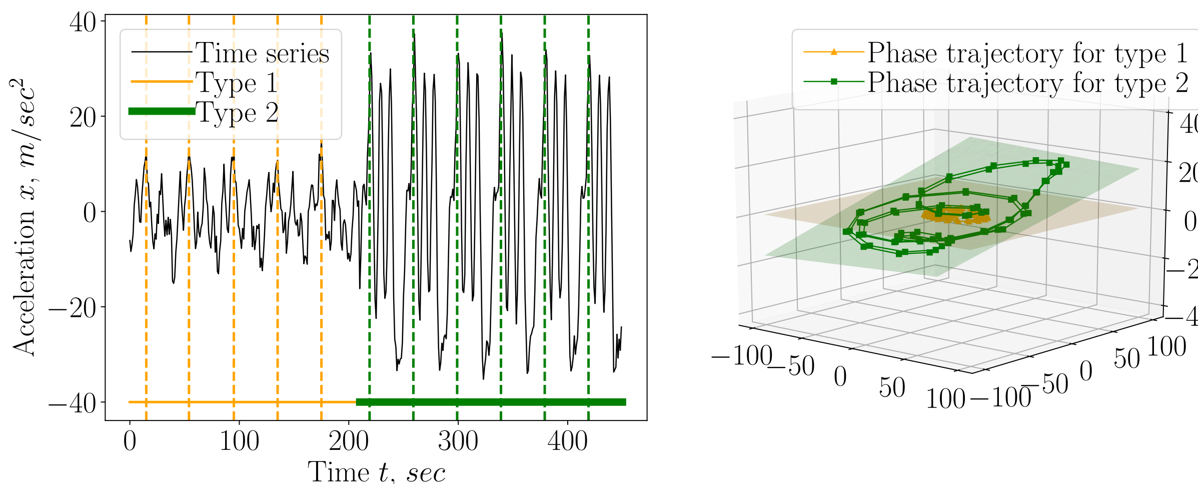
*This research was supported by RFBR, project ..., and by Government of the Russian Federation, agreement ...

[1]Moscow Institute of Physics and Technology, grabovoy.av@phystech.edu

[2]Moscow Institute of Physics and Technology, strijov@ccas.ru

1 Introduction

Analysis of a person's physical activity is carried out by using mobile phones, smart watches etc [1, 2]. These devices are using an accelerometer, gyroscope and magnetometer. The main purpose of this work is to label and recognize human activity during the time [12, 4], and also search for the beginning of a periodic signal [6, 5]. Examples of an action segment is a step, a step of running, a single squat, a single jump, etc. Current work considers a sequence that consists of at least two consecutive segments that correspond to the same type of human activity.



(a) (b)

Figure 1: A time series with clustering: a) time series with assessor markings on clusters and markings at beginning of each quasi-periodic segment; b) projection of phase trajectories on the two principal components

Time series are objects of complex structure. The method of constructing feature vectors for points is very important for their clustering. In this article, the object of analysis, as well as clustering, is a point on the time axis. The paper investigates the problem of clustering points in a time series. *Clustering* is a process in which all points in a time series are labeled with a label from the finite set of labels. Each label corresponds to one characteristic physical action. A *segment* is a part of a time series that corresponds to one characteristic human physical action, for example: a step with two legs when walking, or a step with two legs when running. A sequence of segments that forms a quasi-periodic sequence is called a *chain*. A sequence of points $\{b_t\}_{t=1}^{N}$ is called *quasi-periodic* with period of T, if for all t there is a Δ, such that:

$$b_t \approx b_{t+T+\Delta}, \quad |\Delta| \ll T. \tag{1.1}$$

An example of clustering and splitting a series into segments is shown in fig. 1.a. The time series is divided into two characteristic type of segments, which are marked Type 1 and Type 2. Also, this time series contains two quasi-periodic chains of actions.

The proposed solution of the clustering problem for samples in a time series consists of two stages. First, an algorithm for local approximation of time series by using the principal component method [11] is proposed to obtain a feature description of points in the time series. *Local approximation* of the time series means that only a certain neighborhood of a point is used to describe the features of their points. The two main components of the phase trajectory segment are considered as a feature description of a time series point. The fig. 1.b shows the first two principal components of the phase trajectories, and also shows the projection of the phase trajectories on these components. The trajectories relate to different physical actions, which are denoted Type 1 and Type 2, in the time series. The planes that are generated by these basic components are different. That means that Type 1 and Type 2 are two different actions. Secondly, the distance function between points in the new feature space is considered. This function is the distance between two bases of some subspaces within the phase space of a time series. It can be explained by using fig 1.b. The function is considered between two planes, which are defined by two different bases for the Type1 and Type2 segments. Points can be clustered by using a pairwise distance matrix. The segmentation problem is solved by using the main components [6] of the phase trajectory in each cluster separately.

The method has several assumptions. It is assumed that the periods of different segments differ slightly. Minimum and maximum periods of segments are known. The number of different segments in the time series is known too. It is also assumed that the type of segments in time does not change often.

The quality analysis of the proposed clustering method is carried out on synthetic and real data. A synthetic data constructed by using the sum of the first few terms of the Fourier series with random coefficients. But the experiment on the segmentation of the time series was carried out on simple sinusoidal signals with random amplitude and frequency. Real data was received by using a mobile accelerometer, which took readings during exercise of a person.

2 Related work

The paper [1] describes a method for constructing a feature description based on expertly defined generating functions. The article [7] proposes a method for constructing features based on the data generation hypothesis. A combined feature description is introduced in the paper [8] based on these methods. The paper [9] consider the problem of construction a feature space. The article proposes a criterion for the redundancy of the selected features.

The paper [12] is the closest to our research. They proposed method for human physical activity recognition. To extract the fundamental period they construct the phase trajectory matrix, applying a technique of the principal component analysis. This method allows you to find and classify segments in time series with great accuracy. But this method can work only with time series, in which all segments belong to the same characteristic action.

The Article[6] is one of the closest to our research. The article searches for the beginning of a segment within a quasiperiodic signal, which consists of only one chain of actions. This method is based on the study of the phase space, namely, the search for a stable hyperplane that divides the phase space into two equal parts. The points that are close to this hyperplane are selected as the beginning and ending of the segment. The article [6] proposes to project the phase space into two principal components, after which the beginning of each segment should be highlighted. This method finds the beginning of a segment in the case when the time series consists of one type of signal.

The paper [5] is also close to our research. The article proposes a method for searching for a periodic structure in a series by using the LSTM model with attention mechanism. This method has the same disadvantage as the previous method [6]. This method finds the beginning of a segment in the case when the time series consists of one type of signal.

3 Time series clustering problem

There given a time series

$$\mathbf{x} \in \mathbb{R}^N, \tag{3.1}$$

where N is a number of samples in the time series. The time series consists of a sequence of segments,

$$\mathbf{x} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_M], \tag{3.2}$$

where $\mathbf{v}_i$ is a segment from the set of segments $\mathbf{V}$. It is supposed that the time series $\mathbf{x}$ contains any $\mathbf{v}_i$ from $\mathbf{V}$. For all i either $[\mathbf{v}_{i-1}, \mathbf{v}_i]$ or $[\mathbf{v}_i, \mathbf{v}_{i+1}]$ is a chain of action. The fig. 2.a shows an example of segments $\mathbf{v}_i$. Each segments $\mathbf{v}_i$ has no more than T elements, $|\mathbf{v}_i| \leq T$. The set V has K different types of segments, $|\mathbf{V}| = K$. The fig. 2.a shows an example of segments, which are constructed the chain of action $[\mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}]$.

The figure shows an example of 1, 2 segments that belong to segments of the same type.

Construct the mapping

$$a : t \to \mathbb{Y} = \{1, \cdots, K\}, \tag{3.3}$$

where $t \in \{1, \cdots, N\}$ is a point in the timeline which defines the time series $\mathbf{x}$. Let the map a satisfy the following properties:

$$\begin{cases} a(t_1) = a(t_2), & \text{if time moments } t_1, t_2 \text{ relate to similar type of segments,} \\ a(t_1) \neq a(t_2), & \text{if time moments } t_1, t_2 \text{ relate to different type of segments.} \end{cases} \tag{3.4}$$

Each point t corresponds to a label y_t from the set $\mathbb{Y} = \{1, \cdots, K\}$:

$$\mathbf{y} \in \mathbb{Y}^N. \tag{3.5}$$

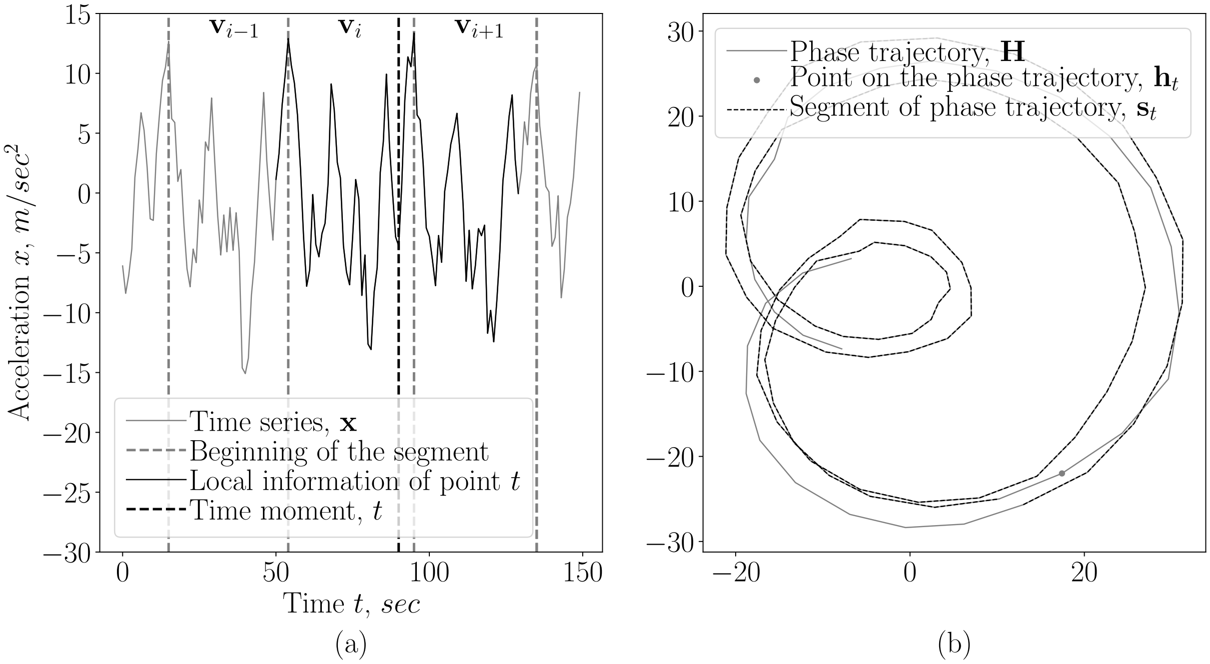


Figure 2: A time series and example of local information of points: a) time series example of local information of time point $t = 90$; b) projection of phase trajectories on the two principal components and example of segment $\mathbf{s}_t$ in time point $t = 90$

The error for the time series is

$$S = \frac{1}{N} \sum_{t=1}^{N} [y_t = a(t)], \tag{3.6}$$

where t is a point of time, y_t is its label corresponds to a point in time t for time series $\mathbf{x}$.

4 Time series samples clustering

Construct the phase trajectory $\mathbf{H}$ with the time series $\mathbf{x}$:

$$\mathbf{H} = \{\mathbf{h}_t | \mathbf{h}_t = [x_{t-T}, x_{t-T+1}, \cdots, x_t],\ T \leq t \leq N\}, \tag{4.1}$$

where $\mathbf{h}_t$ is a point in the phase trajectory.

Split the phase trajectory into segments by using the maximum length of segments:

$$\mathbf{S} = \{\mathbf{s}_t | \mathbf{s}_t = [\mathbf{h}_{t-T}, \mathbf{h}_{t-T+1}, \cdots, \mathbf{h}_{t+T-1}],\ T \leq t \leq N - T\}, \tag{4.2}$$

where $\mathbf{s}_t$ is a segment of phase trajectory. The fig. 2.b shows an example of a phase trajectory segment. The segments have all local information about the time series, as it contains all the information on the period up to some time point t and information about the period after the time point t.

The principal components $\mathbf{W}_t$ for T-dimensional segments $\mathbf{s}_t$ are considered as a feature description of a point t in a time series. The segment $\mathbf{s}_t$ is projected onto subspace with two dimension by using the principal component method $\mathbf{z}_t = \hat{\mathbf{W}}_t\mathbf{s}_t$:

$$\mathbf{W} = \{\mathbf{W}_t | \mathbf{W}_t = [\mathbf{w}_t^1, \mathbf{w}_t^2]\}, \quad \hat{\mathbf{W}} = \{\hat{\mathbf{W}}_t | \hat{\mathbf{W}}_t = [\lambda_t^1\mathbf{w}_t^1, \lambda_t^2\mathbf{w}_t^2]\}, \quad \mathbf{\Lambda} = \{\boldsymbol{\lambda}_t | \boldsymbol{\lambda}_t = [\lambda_t^1, \lambda_t^2]\}, \tag{4.3}$$

where $[\mathbf{w}_t^1, \mathbf{w}_t^2]$ and $[\lambda_t^1, \lambda_t^2]$ are the basis vectors and eigenvalues obtained by using the principal component method for the phase trajectory segment $\mathbf{s}_t$.

Construct the distance function between points $\mathbf{W}_{t_1}, \mathbf{W}_{t_2}$ in the time series $\mathbf{x}$ for their clustering:

$$\rho(\mathbf{W}_1, \mathbf{W}_2) = \max\left(\max_{\mathbf{e}_2 \in \mathbf{W}_2} d_1(\mathbf{e}_2), \max_{\mathbf{e}_1 \in \mathbf{W}_1} d_2(\mathbf{e}_1)\right), \tag{4.4}$$

where $\mathbf{e}_i$ is the basic vector of space $\mathbf{W}_i$, and $d_i(\mathbf{e})$ is the distance from vector $\mathbf{e}$ to the subspace $\mathbf{W}_i$.

If all subspaces $\mathbf{W}_t$ have dimension two, then the distance function $\rho(\mathbf{W}_1, \mathbf{W}_2)$ has the following interpretation:

$$\rho(\mathbf{W}_1, \mathbf{W}_2) = \max_{\{\mathbf{a},\mathbf{b},\mathbf{c}\} \subset \mathbf{W}_1 \cup \mathbf{W}_2} V(\mathbf{a}, \mathbf{b}, \mathbf{c}), \tag{4.5}$$

where $\mathbf{W}_1 \cup \mathbf{W}_2$ is a concatenation of bases, $V(\mathbf{a}, \mathbf{b}, \mathbf{c})$ is the volume of parallelepiped built on vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$, which are columns of matrix $\mathbf{W}_1 \cup \mathbf{W}_2$.

Construct the distance function between eigenvalues:

$$\rho(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) = \sqrt{(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2)^\mathsf{T}(\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2)}. \tag{4.6}$$

Construct the distance between two points in time t_1, t_2 by using equations (4.5-4.6), and consider the matrix of pairwise distances between pairs of points in the time series:

$$\rho(t_1, t_2) = \rho(\mathbf{W}_1, \mathbf{W}_2) + \rho(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2), \quad \mathbf{M} = \mathbb{R}^{N \times N}, \tag{4.7}$$

where $\mathbf{M}$ is a matrix of pairwise distances between all pairs of points t in the time series $\mathbf{x}$. The pairwise distance matrix $\mathbf{M}$ is used for clustering points t of the time series (3.3).

5 Computation experiment

5.1 Points clustering

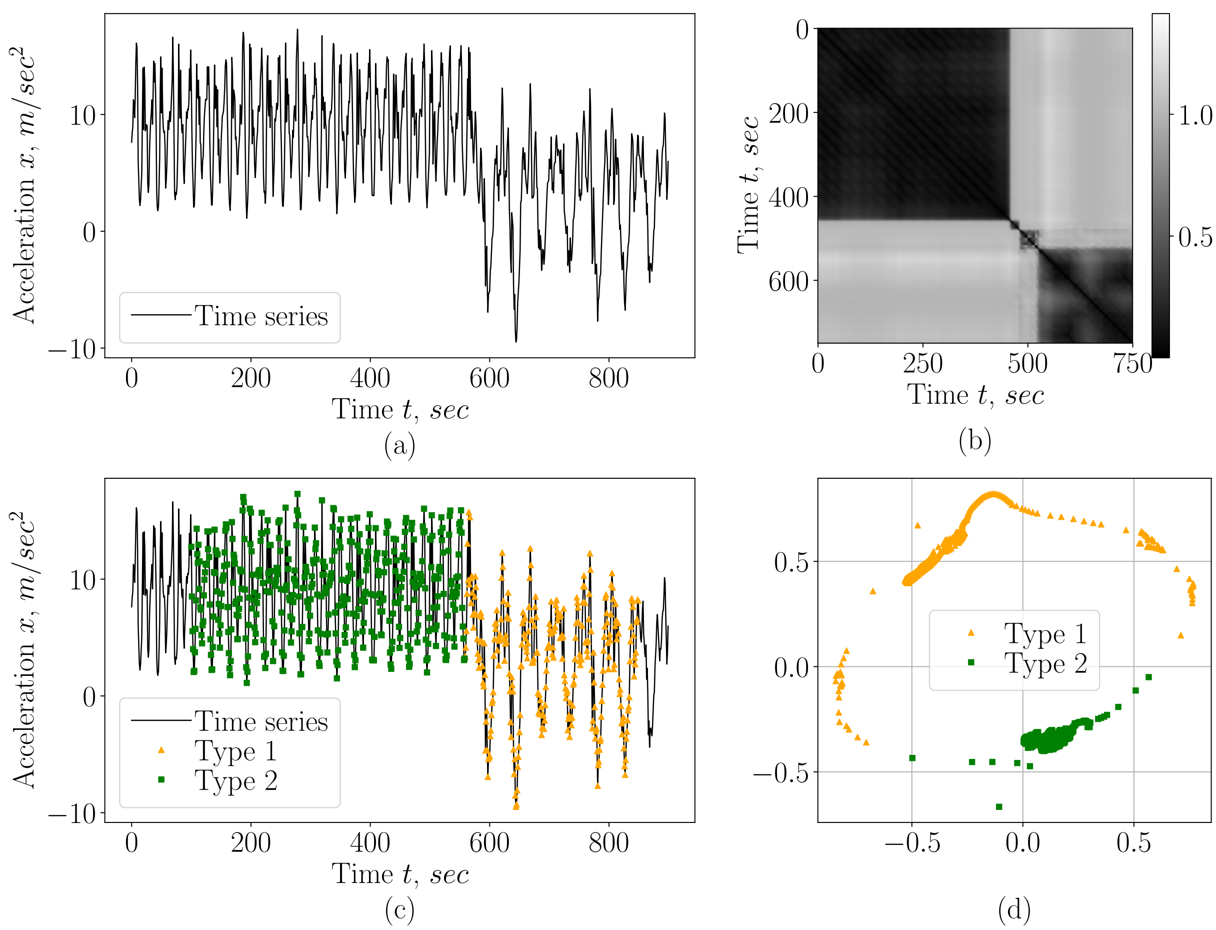


Figure 3: The result of the clustering algorithm for the time series "Physical Motion 2": a) a time series; b) the pairwise distance matrix for the time series; c) an example of clustering; d) Multidimensional Scaling of the points into two dimensional subspace by using pairwise distance matrix

In the experiment, the clustering of points in the time series was carried out using matrices of pairwise distances (4.7). The experiment was carried out on real and synthetic data, which are described in the table 1. "Physical Motion" is a real time series obtained by using a mobile accelerometer. Synthetic time series were constructed by using the first few terms of the Fourier series with random coefficients from the standard normal distribution. The generation of synthetic time series consisted of two stages. At the first stage, short segments $\mathbf{v}$ were generated to build a set of all segments $\mathbf{V}$. The second stage is the following random process:

$$\mathbf{x} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_M] + \boldsymbol{\varepsilon}, \quad \begin{cases} \mathbf{v}_1 \sim \mathcal{U}(\mathbf{V}), & \\ \mathbf{v}_i = \mathbf{v}_{i-1}, & \text{with probability } \frac{3}{4}, \\ \mathbf{v}_i \sim \mathcal{U}(\mathbf{V}), & \text{with probability } \frac{1}{4} \end{cases} \tag{5.1}$$

where $\mathcal{U}(\mathbf{V})$ is a uniform distribution on objects from the set $\mathbf{V}$, and $\boldsymbol{\varepsilon}$ is gaussian noise.

The result of the algorithm for the real time series is shown in the fig. 3. The fig 3.b shows the matrix of pairwise distances for points in the timeline. Time series points can be easily visualised on a plane by using a pairwise distance matrix and a Multidimensional Scaling method [10]. The visualisation is shown in the fig. 3.d. The fig 3.c shows the result of clustering points. More results of clustering are presented in the table 1.

5.2 Time series segmentation

The time series segmentation is shown on the example of real data obtained by using a mobile accelerometer. Segmentation is carried out by using the method that is presented in the work [6]. The method is used for each action within the time series separately. The fig. 4 shows the result of the segmentation algorithm for the time series "Physical Motion 2". The segmentation result for a cluster "Type 1" is much better than for a cluster "Type 2". This result is easily explained by using fig. 4.b and fig. 4.c. The fig. 4.b is a phase trajectory for cluster "Type 1". This phase trajectory has no self-intersections in contrast to the phase trajectory shown in the fig. 4.c. As a result, for time series with a simple phase trajectory, this method performs the time series segmentation well.

Table 1: Description of time series in the experiment

Series, $\mathbf{x}$	Length, N	Number of segments, K	Period, T	Error, S
Physical Motion 1	900	2	50	0.03
Physical Motion 2	900	2	35	0.08
Physical Motion 3	900	2	30	0.09
Physical Motion 4	800	2	50	0.01
Synthetic 1	2000	3	40	0.008
Synthetic 2	2000	2	40	0.06
Synthetic 3	2000	2	40	0.03
Synthetic 4	2000	2	40	0.03
Synthetic 5	2000	2	40	0.04
Simple	1000	2	135	0.14

6 Conclusion

The paper considered the problem of finding periodic structures within a time series. A method based on local reduction of the phase space dimension was considered. An algorithm for searching for segments was proposed. The algorithm is based on the principal component method for local dimension reduction. Also introduced is the function of the distance between local basis at each time instant. Local bases were interpreted as a features description of a point in the time series.

During the experiment on the real and synthetic data showed that the proposed method for measuring the distance between the basis well separates points that are related to different type of action, which leads to good clustering of time series points. The results of the experiment are shown in the table 1. The experiment was carried out segmentation of time series by using the method [6] for each cluster separately.

The proposed method has few disadvantages associated with a large number of assumption on a time series. These restrictions will be relaxed in subsequent papers. It is also

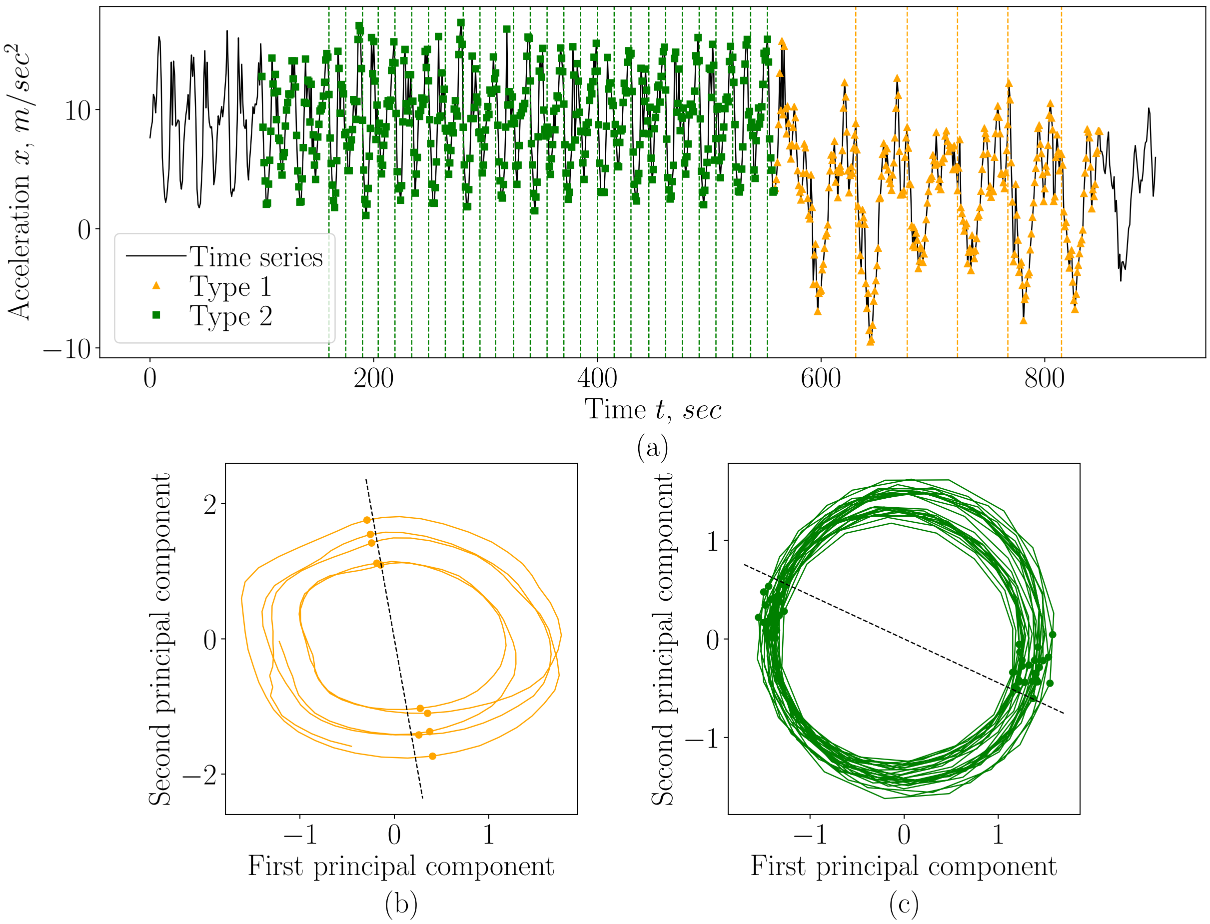


Figure 4: The result of the segmentation algorithm for the time series "Physical Motion 2": a) a time series segmentation; b) a phase trajectory for cluster "Type 1"; c) a phase trajectory for cluster "Type 2"

planned to solve the problem of finding the minimum dimension of the phase space for which the phase trajectory will not have self-intersections.

References

[1] *J. R. Kwapisz, G. M. Weiss, S. A. Moore* Activity Recognition using Cell Phone Accelerometers // Proceedings of the Fourth International Workshop on Knowledge Discovery from Sensor Data, 2010. Vol. 12. P. 74–82.

[2] *W. Wang, H. Liu, L. Yu, F. Sun* Activity Recognition using Cell Phone Accelerometers // Joint Conference on Neural Networks, 2014. P. 1185–1190.

[3] *A. D. Ignatov, V. V. Strijov* Human activity recognition using quasiperiodic time series collected from a single tri-axial accelerometer. // Multimedial Tools and Applications, 2015.

[4] *A. Olivares, J. Ramirez, J. M. Gorris, G. Olivares, M. Damas* Detection of (in)activity periods in human body motion using inertial sensors: A comparative study. // Sensors, 12(5):5791–5814, 2012.

[5] *Y. G. Cinar and H. Mirisaee* Period-aware content attention RNNs for time series forecasting with missing values // Neurocomputing, 2018. Vol. 312. P. 177–186.

[6] *A. P. Motrenko, V. V. Strijov* Extracting fundamental periods to segment biomedical signals // Journal of Biomedical and Health Informatics, 2015, 20(6). P. 1466 - 1476.

[7] *Y. P. Lukashin* Adaptive methods for short-term forecasting // Finansy and Statistik, 2003.

[8] *M. P. Kuznetsov, N. P. Ivkin* Time series classification algorithm using combined feature description // Machine Learning and Data Analysis, 2015, 11(1). P. 1471–1483.

[9] *V. V. Strijov, A. M. Katrutsa* Stresstes procedures for features selection algorithms. // Schemometrics and Intelligent Laboratory System, 2015.

[10] *I. Borg, P. J. F. Groenen* Modern Multidimensional Scaling. — New York: Springer, 2005. 540 p.

[11] *D. L. Danilov, A. A. Zhiglovsky* Main components of time series: method "Gesenitsa". — St. Petersburg University, 1997.

[12] *A.D. Ignatov, V. V. Strijov* Human activity recognition using quasi-periodic time series collected from a single triaxial accelerometer. // Multimedia Tools and Applications, 2015, P. 1–14.