Scientific Journal of Riga Technical University
Computer Science. Information Technology and Management Science
2011
Volume 49

Processing Short Time Series with Data Mining Methods

Arnis Kirshners[1], Arkady Borisov[2], *[1-2]Riga Technical University*

***Abstract* – This article examines several data mining approaches that perform short time series analysis. The basis of the methods is formed by clustering algorithms with or without modifications. The proposed methods implement short time series analysis when the numbers of the observations are not equal and the historical information is short. The inspected approaches are offered for solving complex tasks where statistical analysis methods cannot be applied or their functioning does not provide the necessary efficiency. The proposed methods are based on *grid-based clustering* and *k-means* algorithm modifications.**

***Keywords* – Clustering, data mining, grid-based clustering, k-means, prototypes, short time series**

I. Introduction

In statistics time series analysis process is based on searching for relations in a longer period of time, analyzing changes of values over time; as a result, the future values of the object are predicted. These processes can be related to variations of economic indexes, e.g., currency rate fluctuations [1]. Similar processes are analyzed, e.g. when determining tourist inflow for the next season [2] etc. All these analysis processes are similar because they change system parameters when environment functioning changes are made. It is postulated that time series are stationary – the longer observations of this system are, the larger the probability to effectively discover the regularities is [3].

But in real life there are tasks that involve time series with comparatively short length and it is almost impossible to find the regularities using methods of statistical analysis. These tasks include: product life cycle analysis [4], [5], e-service analysis at the beginning of their introduction [6], analysis of textile item sales when there is a wide range of products and short product life expectancy [7], analysis of gene expressions in bioinformatics [8] etc. Therefore it is important to create an approach that would solve the problem of short time series analysis when there are relatively few observations.

There are many different time series analysis methods. These include a large variety of statistical models like discriminant analysis and logic regression but the use of these models is limited when the analyzed data is complex and non-linear [9], but exactly situations like these are in real life domains. Thus, machine learning methods are considered to be more suitable to characterize these situations [10].

The goal of this article is to propose several alternative solutions for tasks that involve short time series, using data mining methods and algorithms. Data mining at this moment is in the stage of growth and popularity because this field of science is at the crosspoint of sciences like artificial intelligence, data base management systems, statistics, bioinformatics, etc. All these science branches have to operate with large amounts of data and this is where data mining becomes useful in helping to acquire new unknown knowledge with its methods and algorithms. This knowledge includes searching for relations in short time series where data mining methods and algorithms help in searching for similarity or difference between them. Similarities and differences are usually based on one of the measures or metrics that are used for comparison of objects using specific criteria.

II. Used methods

Short time series describe a set of subsequent events that are arranged in a specific time and order. The number of periods in a short time series is small and is represented using equation $T=\{T_1,T_2,\ldots,T_n\}$. Short time series analysis is done with the goal of finding the structure of series and forecast future values. The process of analysis includes finding time series structures and relations that cover noisy, peak, seasonal and cyclic values. To carry out data analysis in short time series, it is crucial to ensure integrity of the studied objects. Based on the specifics of a task the right data pre-processing approaches have to be chosen in a way that provides preparation of the initial data for data analysis process. In the data analysis process, the most appropriate clustering algorithm has to be chosen and in order to do so, it is important to initially inspect a list of problems that can occur in data analysis process. The attention should be paid to the following processes:

- The complexity of parameter choice that serves as a base for clustering;
- The complexity of clustering method choice; it is important to know the structure of methods and specifics of their functioning to make the right decision;
- The problem of choosing the number of clusters; if the probable number of clusters is not known, it is important to make a series of experiments searching the set of possible cluster numbers and choosing the most appropriate based on one of the clustering performance evaluations;
- The problem of clustering result interpretation; the shape of clusters is determined by the choice of merging technique; but it should be considered that specific clustering methods tend to form clusters of different shapes that cannot be used in particular tasks, e.g., if the clusters overlap and a small region of object space is

saturated because these shapes can make a wrong impression of clustering results.

A. Clustering of Short Time Series with Different Lengths

There are tasks related to processing of short time series with different lengths like forecasting of new electronic services. Electronic services are accessed using Internet and browsing tools. The historical request information of the used services is stored for future data analysis. When a new service is introduced, it is important to know its demand, e.g., in its first days or a month. An important step in these tasks is the data pre-processing because the data available in the short time series holds historical information of different lengths, see Fig. 1. To perform an adequate short time series analysis it is necessary to transform time series. The transformations provide the modification of time series making the beginning of the analyzed data even, see Fig. 1. As a result of the transformations the meaning and structure of data is changed.

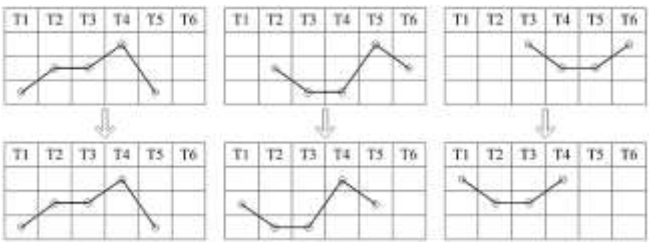

Fig. 1. Data transformations

Data clustering can be implemented using the *k-means* algorithm [11] and [12] or a modified grid-based clustering method [4]. To use grid-based clustering method for time series with different lengths, they should be transformed to form pairs, i.e. the excess periods at the end of the time series should be dismissed, see Fig. 2. To form a clustering grid, it is necessary to choose a number of clusters, e.g., consisting of nine clusters or a field of size 3x3. Then grid size is to be computed using mathematical functions floor (-3,01) = -4 and ceiling (1,30) = 2 correspondingly for the minimum and maximum values of the data set, see Table I values in bold.

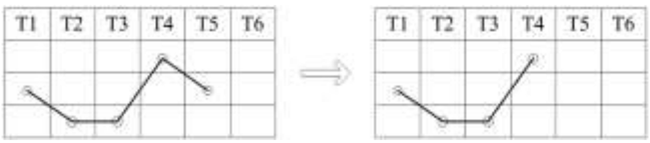

Fig. 2. Period alignment in pairs

TABLE I

NORMALIZED DISTANCES OF DEMAND DATA SET

ID	T1	T2	T3	T4	T5	T6
1	-0,88	**1,30**	0,79	-1,08		
2	-1,77	-2,19	-1,81	-1,27	-1,11	-1,20
3	**-3,01**	-1,75	-0,74	0,36	0,46	0,51
4	-1,12	-1,06	-1,06	-1,04		

The size of the clustering grid is set on the basis of the obtained results; when merging time points of each data set object by twos ID1=T1 and T2; T3 and T4; ID2= T1 and T2; T3 and T4; T5 and T6 etc. gives the value represented in the clustering grid that is constructed in two-dimensional space as a point, see Fig. 3. Numbers in the right-hand bottom corners represent the number of the corresponding cluster.

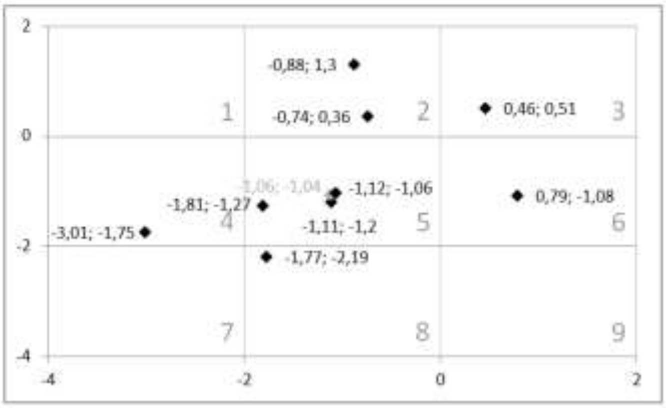

Fig. 3. Placement of pairs in a clustering grid

To continue with the time series analysis it is necessary to reduce grid-based clustering results to a sequence of cluster numbers with discrete values and to align the lengths of time series. The new values with cluster numbers are depicted in Table II and the free spaces are filled with „0" values initially appending letter „C" to all values. If the initial maximum number of periods in a time series was six, after reducing the results it only consisted of three blocks.

TABLE II

NUMBERS OF BLOCKS REDUCED AS A RESULT OF CLUSTERING

ID	B1	B2	B3
1	C2	C6	C0
2	C8	C5	C5
3	C4	C2	C3
4	C5	C5	C0

The proposed approach can also be implemented using the *k-means* algorithm. The classification algorithm is chosen using experiments that determine the mean absolute, the mean squared and the total error of clustering [9]. Clustering algorithm divides short time series into groups based on one of the metrics between the analyzed time series. Based on the obtained block numbers and the aligned length of the time series it is possible to run further data analysis, e.g., classification to find new knowledge and relations. Similar approaches were used for solving product life cycle transition phase problems related to product being in different stages of development [13].

B. Short Time Series Clustering Using Profiles

There are tasks where object life has seasonal features and that cannot be solved using classic time series analysis methods because the object is in market for approximately three to nine months. This type of task is solved in the fashion industry when selling famous clothing brands. A product is marketed for a certain period of time – a season, and then withdrawn from the turnover. This type of product does not have historical data that can serve as a basis for long-term forecast because the form, type, colour, fabric and other parameters change from one season to another. When solving this or similar tasks the pattern an analyst is looking for is a demand profile of the product representing the behaviour of the product or product group in a certain period of time. The profile describes the mean values of a group or cluster at each moment of time.

Data pre-processing consists of normalization using life curve [7] that ensures neutralization of dominating values and whose efficiency was experimentally proven by [14]. The basis of the approach is data clustering that builds profiles based on the found object groups using a modified *k-means* algorithm. The merge of short time series into groups or clusters is carried out based on comparison of curve structures at each moment of time using a distance metric [11], see Fig. 4. The chosen metric is *Euclidean* distance [12] because some comparative studies on metrics [15] show that there is no significance in the chosen metric because the difference is trivial and it practically does not influence the results of clustering.

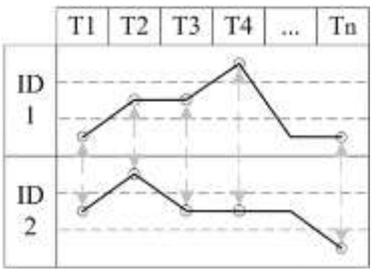

Fig. 4.Comparing short time series

As a result of *k-means* clustering algorithms modification, a variable – maximum number of clusters, was introduced; it is calculated as $\sqrt{n}$, where n – number of records in the data set. This variable increases the efficacy of the algorithm by reducing the number of iterations required to cluster a data set. The classic *k-means* algorithm continues its work until finding the distance between every object and the center of a cluster or centroid. Then the most appropriate number of clusters for data set clustering is determined using mean absolute error (MAE) that is calculated using equation (1):

$$MAE = \frac{\left(\frac{d_{11}+d_{12}+\cdots+d_{1n}}{c_1}\right)+\left(\frac{d_{21}+d_{22}+\cdots+d_{2n}}{c_2}\right)+\cdots+\left(\frac{d_{z1}+d_{z2}+\cdots+d_{zn}}{c_z}\right)}{C_n}, \quad (1)$$

where

d_{11}-d_{zn} – the distance between the corresponding object and centroid;

c_n – number of records in the cluster;

C_n – number of clusters.

Based on calculations using equation (1) and the assumption that the experiments were carried out using 10 clusters, the obtained mean absolute error results of clustering, see Fig.5., show that the most appropriate number of clusters to describe

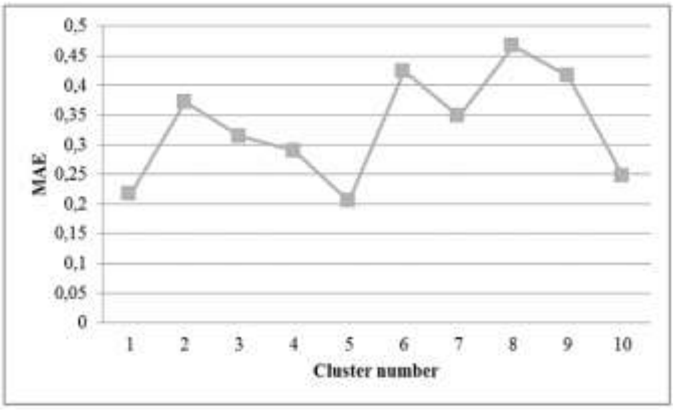

Fig. 5. Determining the most appropriate number of clusters using *MAE* evaluation

that data set is five because the minimum clustering error was achieved at this number of clusters. After determining the memberships of objects, a profile is made for each cluster, see Fig. 6, that characterizes the mean values of demand at each time point. The obtained profiles can serve in the further data analysis as value scale for the particular group of objects that was found in the clustering process in a particular period of time. These prototypes can be used to easily make future forecasts for a new product or a product group, e.g., using classification algorithms.

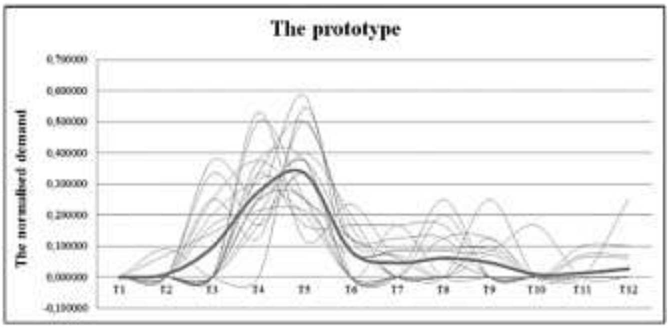

Fig. 6. Construction of a prototype (bold line) in a cluster

C. Clustering of Short Time Series in Medical and Pharmacological Tasks

This approach provides processing of short time series when analyzing behaviour of heart contraction force in a certain period of time. In medicine this method can be used in cardiology to cure heart ischemia by analyzing changes in heartbeat when a specific period of time has passed after introducing a stimulating substance. Similar tasks are solved in pharmacology when analyzing laboratory examination data for a research; like data obtained in a model of isolated heart ischemia - reperfusion [16], [17] using *Wistar* rats that were fed the studied substances for a certain period of time. The aim of this research was to assess the efficiency of the studied substances in heart cell protection from ischemic - reperfusion damages by determining the size of the heart necrosis (necrotic tissue). The isolated heart experiments can be conditionally divided into three stages: heart adaptation period lasting for 20 minutes when a probe is inserted into the isolated heart to measure heart contraction power and heart rhythm; occlusion – the banding of the downward coronary artery lasting for 40 minutes; reperfusion – unfastening of blood vessels that lasts for 120 minutes. Heart contraction power and heart rate are registered in the occlusion stage using specific medical apparatus that records data with 60 second intervals, see Fig. 7. As a result of data selection, the obtained data set holds time series consisting of heart contraction power readings at 40 time points.

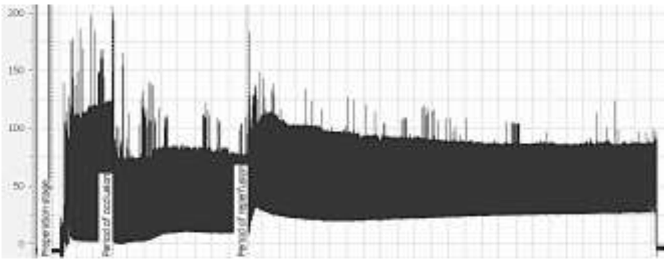

Fig. 7. Registering heart contraction power in the isolated heart experiment

The used equipment provides dismissal of noisy values and calculation of mean values after each period of time. The values of the acquired data set are time series but due to the length of observations being too short and inconsistent these values are called short time series. When analyzing short time series (continuous data) it is practically impossible to determine functional relations of the process in them therefore this type of tasks are considered difficult to formalize and cannot be solved using classical statistical methods, thus data mining methods and algorithms are suggested.

Data pre-processing is based on time series selection and data normalization. In data selection process the first and the last of the selected time series values are dismissed because they can hold faulty readings related to tying and untying of the coronary arteries. Data normalization can be implemented using various approaches, e.g. *normalization using life curve* [7] or *Z-estimate normalization with standard deviation* that is calculated using equation (2):

$$T_i^{'} = \frac{T_i - \overline{T_i}}{\sigma_{T_i}}, \qquad (2)$$

where T_i – value of the time series T at the moment i;
$\overline{T_i}$ – mean value of the time series;
? ?$_{T_i}$ – standard deviation.

As a result of data pre-processing, short time series with equal length and number of periods are selected see Fig. 8. These time series are cleared of noisy values and normalized. During the process of clustering, groups of similar objects called clusters are found. Clustering can be implemented using *k-means*, *modified k-means* [14], maximum likelihood [18] or agglomerative hierarchical algorithms [12]. Only a complex verification of the proposed clustering algorithms at different numbers of clusters can determine the most appropriate clustering approach and the optimal number of clusters to describe the data set.

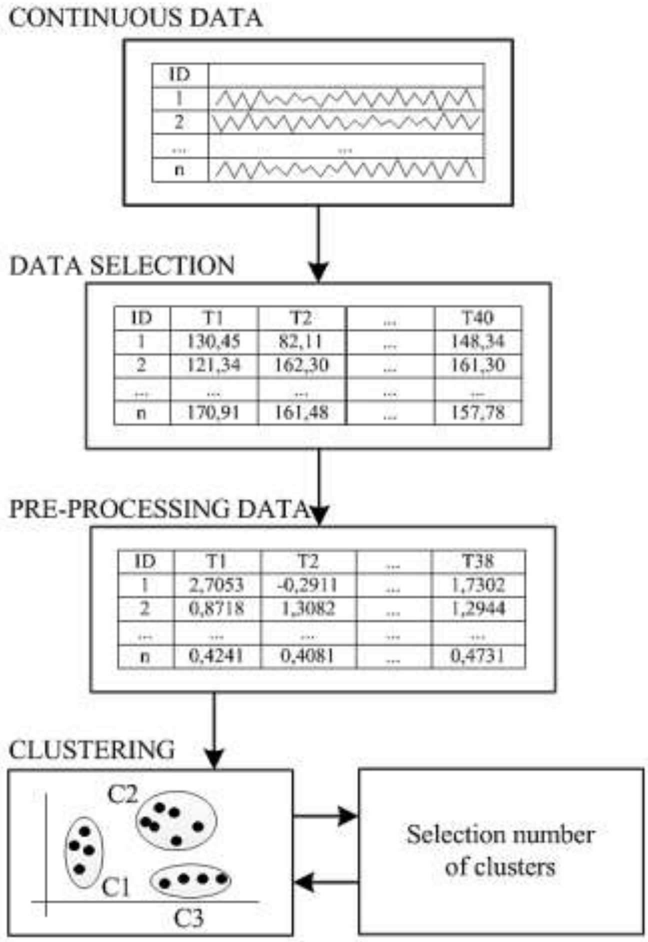

Fig. 8. Data analysis process

It is necessary to scrutinize the obtained clustering results and evaluate the impact of the induced clustering error on the acquired results.

III. Analysis and Evaluation of Clustering Results

After clustering of a data set the results should be analyzed to answer the following questions:

- Are the created clusters essential or accidental?
- Is the division stable and reliable?
- Are there relationships between clustering results and variables that were not recognized in clustering?
- Are the obtained results interpretable?

The acquired clustering results have to be evaluated using formal and informal methods. The formal methods are built upon structure of the algorithm used in the clustering; but informal include classification performance assessment like:

- Result analysis using training and test data sets;
- Cross-validation;

- Clustering data set with changed succession of observations;
- Clustering the data with some records excluded;
- Clustering few records.

To improve the validity of the results, several of the mentioned clustering evaluation methods should be used when comparing the results. If, however, several clustering result evaluation methods are not used, it does not point to the incorrectness of the obtained results. The use of several approaches and the comparison of results are considered to be an indication of qualitative clustering.

There are also many problems associated with cluster analysis. Just as any other method, clustering approaches and algorithms have their weaknesses and restrictions. For example, when carrying out clustering, one must remember that reduction of data dimensionality can cause noise and divest some characteristics of objects.

IV. Conclusions

The short time series clustering approaches investigated in the article provide division of the analyzed data into clusters when solving various practical tasks. The used pre-processing approaches show that data can be qualitatively prepared for clustering process by implementing data cleaning, transformation and normalization processes.

Clustering data with time series of uneven length enables solving of specific tasks with a small amount of historical information. The alignment of short time series lengths using coupled merging method, data transformation and discretization allows clustering observation values that could not be clustered using classic data analysis methods that are not designed to analyze short time series with different numbers of periods. The discretization proposed in the approach provides wider range of used methods, e.g., the obtained results can be further processed using classification.

The approach that is based on prototype construction can be used as a basis for forecasting system that determines the demand prognosis of a new product for next periods, e.g., 12 months. The acquired forecast that is presented as a prototype helps decision makers to make a decision when ordering new products because the graphical representation of the prototype is easy to perceive and interpret. The proposed approach does not involve complex calculations and the obtained results are easily interpretable for users.

The clustering approach designated for solving medicine and pharmacology tasks implements heart contraction power analysis. This approach is distinct because of the specific equipment used in data acquisition and pre-processing dismissing noisy values. This approach shows that selection of short time series in a specific period of time is possible and it depicts heart rate of a patient or an animal in period of occlusion. On the basis of the pre-processed values it is possible to implement data preparation for further analysis – clustering. The clustering algorithms mentioned in this article can be used to carry out the clustering process. In this task it is also important to determine the most appropriate number of clusters required for data clustering.

V. Further Research

The planned future research is related to comparison of the proposed approaches to methods used in statistics like exponential smoothing, exponential smoothing with trend, forecasting with weighted average and the moving average method. This comparison is necessary to prove the preferential of data mining methods in short time series analysis employing the used data sets.

VI. Acknowledgments

This work has been supported by the European Social Fund within the project «Support for the implementation of doctoral studies at Riga Technical University».

References

[1] J. J. Flores and R. Loaeza, *"*Financial time series forecasting using a hybrid neural-evolutive approach," *Proceedings of the XV SIGEF International Conference*, Lugo, Spain, 2009, pp. 547-555.

[2] E. de Alba and M. Mendoza, „Bayesian Forecasting Methods for Short Time Series," *Foresight: The International Journal of Applied Forecasting, International Institute of Forecasters*, Issue 8, 2007, pp. 41-44.

[3] J. S. Armstrong, F. Collopy and J. T. Yokum, "Decomposition by causal forces: A procedure for forecasting complex time series," *International Journal of Forecasting*, 21, 2005, pp. 25-36.

[4] A. Kirshners and A. Sukov, "Rule induction for forecasting transition points in product life cycle data," *Scientific Proceedings of Riga Technical University, Information Technology and Management Science*, Issue 5, Vol.36, RTU, Riga, 2008, pp. 170-177.

[5] S. Parshutin, L. Aleksejeva and A. Borisov, "Forecasting Product Life Cycle Phase Transition Points with Modular Neural Networks Based System," *Proceedings of 9th Industrial Conference on Data Mining ICDM`2009*, Springer-Verlag, 2009, LNAI 5633, pp. 88-102.

[6] A. Kirshners and Y. Kornienko, "Time-series data mining for e-service application analysis," *Scientific Proceedings of Riga Technical University, Information Technology and Management Science*, Issue 5, Vol. 40, RTU, Riga, 2009, pp. 94-100.

[7] S. Thomassey and A. Fiordaliso, "A hybrid sales forecasting system based on clustering and decision trees," Decision Support Systems, Volume 42, Issue 1, 2006, pp. 408-421.

[8] J. Ernst, G. J. Nau and Z. Bar-Joseph, "Clustering short time series gene expression data," Bioinformatics, Vol. 21, No. suppl_1, 2005, pp. 159-168.

[9] D. C. Montgomery, C. L. Jennings and M. Kulachi, "*Introduction to time series analysis and forecasting."* Wiley-interscience, 2008.

[10] M. Devisscher, B. De Baets and I. Nopens, "Pattern discovery in intensive care data through sequence alignment of qualitative trends: proof of concept on a diuresis dataset." *Proceedings of the ICML/UAI/COLT 2008 Workshop on Machine Learning for Health-Care Applications*, Helsinki, Finland, 2008.

[11] I. H. Written and E. Frank, "*Data mining: practical machine learning tools and techniques - 2nd edition."* Amsterdam etc.: Morgan Kaufman, 2005.

[12] A. Barsegyan, M. Kupriyanov, V. Stepanenko and I. Holod, "*Technologies of Data Analysis: Data Mining*," Visual Mining, Text Mining, OLAP. St.Petersburg, 2007.

[13] Sukov A. Recognition of transitions between different phases of the product life cycle // *Proceedings of the 20th European Modelling & Simulation Symposium, The 20th European Modelling & Simulation Symposium, EMSS 2008*, September 17-19, 2008, Amantea, Italy, pp. 81-86.

[14] A. Kirshners, S. Parshutin and A. Borisov, "Combining clustering and a decision tree classifier in a forecasting task," Automatic Control and Computer Sciences, Vol. 44, N 3, 2010, pp. 124-132.

[15] P. Grabust "The choice of metrics for clustering algorithms," *Proceedings of the 8th International Scientific and Practical Conference*, June 20-22, 2011, Rezekne, Latvia, pp. 70-76.

[16] E. Liepinsh, R. Vilskersts, E. Skapare, B. Svalbe, J. Kuka, H. Cirule et al., "Mildronate decreases carnitine availability and up-regulates glucose

uptake and related gene expression in the mouse heart," Life Sci, 2008, 83: 613–619.
[17] E. Liepinsh, R. Vilskersts, L. Zvejniece, B. Svalbe, E. Skapare, J. Kuka et al., "Protective effects of mildronate in an experimental model of type 2 diabetes in Goto-Kakizaki rats," British Journal of Pharmacology, 2009, 157: 1549–1556.
[18] G. McLachlan and T. Krishnan, "The EM algorithm and extensions." Wiley series in probability and statistics. John Wiley & Sons, 1997.

Arnis Kirshners, Mg. sc. ing., Doctoral student, Lecturer and Systems Analyst, Department of Modelling and Simulation, Riga Technical University. He received his diploma from Riga Technical University. His research interests include data mining and knowledge extraction, bioinformatics, intelligent systems, programming, database and evolutionary computing. Previous publications: A. Kirshners, S. Parshutin, A. Borisov, "Combining clustering and a decision tree classifier in a forecasting task," Automatic Control and Computer Sciences, Vol. 44, N 3, 2010, pp. 124-132. Contact information: 1 Kalku Street, Riga Technical University, Riga, LV-1658, phone: +371 67089530, e-mail: arnis.kirsners@rtu.lv.

Arkady Borisov is Professor of Computer Science at the Faculty of Computer Science and Information Technology, Riga Technical University. He holds a Doctor of Technical Sciences degree in Control in Technical Systems and the Dr. habil. sc. comp. degree. His research interests include fuzzy sets, fuzzy logic, computational intelligence and bioinformatics. Previous publications: A. Kirshners, S. Parshutin, A. Borisov, "Combining clustering and a decision tree classifier in a forecasting task," Automatic Control and Computer Sciences, Vol. 44, N 3, 2010, pp. 124-132. He has 220 publications in the field. Contact information: 1 Kalku Street, Riga Technical University, Riga, LV-1658; phone: +371 67089530, e-mail: arkadijs.borisovs@cs.rtu.lv.

Arnis Kiršners, Arkadijs Borisovs A. Īsu laika rindu apstrāde ar datu ieguves metodēm
Rakstā tiek aplūkotas vairākas datu ieguves pieejas, kas realizē īsu laika rindu apstrādi. Metožu pamatā ir izmantoti klasterizācijas algoritmi ar vai bez modifikācijām. Piedāvātās pieejas risina īsu laika rindu analīzi pie nevienāda novērojumu skaita un pie īsas vēsturiskās informācijas. Aplūkotās pieejas tiek piedāvātas sarežģītu uzdevumi izpildei, kur nav iespējams pielietot statiskās analīzes metodes vai arī to izmantošana ir neefektīva. Piedāvātās pieejas balstās uz *klasterizācijas režģa* un *k-vidējo* algoritmu modifikācijām. Pie nevienāda laika rindu garuma tiek aplūkota klasterizācijas pieeja, kas izlīdzina laika periodus ar datu transformācijas palīdzību, apvienojot tos pāros un diskretizējot pāra vērtības, kā arī aizstājot trūkstošās lauku vērtības. Šāda veida uzdevumi ir pielietojami, analizējot nepieciešamību ieviest jaunu interneta pakalpojumu iedzīvotāju vidū vai analizējot produkta dzīves cikla pārejas fāzi. Īsu laika rindu klasterizācija ar paraugmodeļa izveidi klasteros realizē pieprasījuma prognozēšanas uzdevumus, nosakot vēsturiskos pieprasījuma datos līdzīgas objektu grupas, uz kuru bāzes tiek veidoti paraugmodeļi. Balstoties uz paraugmodeļiem, var uzskatāmi interpretēt atrastās objektu grupas uzvedību noteiktā laika posmā. Iegūtie paraugmodeļi ir pielietojami prognozēšanas uzdevumos, nosakot jaunā produkta iespējamo pieprasījumu nākotnē. Pieejas pamatā, kas realizē īsu laika rindu apstrādi medicīnas un farmakoloģijas uzdevumu risināšanā, ir sirds kontrakcijas spēka rezultātu analīze, kas iegūta medicīnas vai farmakoloģiskajos laboratoriskajos pētījumos, lai atrastu līdzīgas objektu grupas to tālākai izmantošanai datu analīzes procesā. Piedāvātā pieeja realizē datu apstrādi no specializētas medicīniskās aparatūras, veicot laika periodu atlasi, izveidojot laika rindas, kuras ar datu ieguves pirmapstrādes metodēm tiek sagatavotas klasterizācijas procesam. Klasterizācijas procesā tiek atrastas līdzīgas objektu grupas un noteikts optimālākais klasteru skaits datu kopas klasterizācijai. Tiek sniegti ieteikumi, kas jāņem vērā, klasterizējot īsas laika rindas un veicot iegūto rezultātu novērtējumu.

Арнис Киршнерс, Аркадий Борисов. Обработка коротких временных рядов методами добычи данных
В статье рассматривается несколько методов добычи знаний из данных, предназначенных для обработки коротких временных рядов. В основе методов лежат алгоритмы кластеризации с модификациями и без них. Предлагаемые методы выполняют анализ коротких временных рядов при различном количестве наблюдений во множестве данных и краткой исторической информации. Предложенные подходы решают сложные задачи, где невозможно использовать методы статического анализа или их применение не является эффективным. Предложенные подходы основаны на модификациях таких алгоритмов кластерного анализа, как *k-means* и *grid-based clustering*. При различной длине временных рядов применяется метод кластеризации, который выравнивает периоды времени с помощью преобразования данных. Временные ряды делятся на кортежи, состоящие из двух смежных значений, после чего происходит кластеризация кортежей и каждый кортеж заменяется номером ближайшего кластера, тем самым выполняя дискретизацию временных рядов. Временной ряд доводится до необходимой длины, после дискретизации заполняя недостающие значения условной константой. Данный подход можно применить при решении задачи прогнозирования спроса. При помощи кластеризации коротких временных рядов в исторических данных, выявленные прототипы применяются для интерпретации поведения схожих объектов на определенном промежутке времени, например, выявляя потенциальный спрос на новый продукт в будущем. Предлагаемый подход обработки коротких временных рядов также можно применить для анализа данных из медицины и фармакологии, например, данных о силе сердечных сокращений, полученных в лаборатории медицинских или фармакологических исследований, с целью выявить закономерности и сформировать прототипы для их дальнейшего использования в процессе анализа новых данных. Предлагаемый подход реализует обработку данных, полученных с применением специализированного медицинского оборудования. Отбираются периоды времени и формируются короткие временные ряды, которые далее подвергаются предобработке перед процессом кластеризации. Сформулированы рекомендации, которые следует учитывать при кластеризации коротких временных рядов, проводя анализ оценки полученных результатов.