Continuous data stream clustering based on a generalized model of art family neural network

I L Kashirina, K A Fedutinov, T V Azarnova and Yu V Bondarenko
Department of Applied Mathematics, Informatics and Mechanics, Voronezh State University, 1, University Square, Voronezh, 394018, Russia
E-mail: kash.irina@mail.ru

Abstract. The article considers peculiarities of continuous data stream clustering with the help of ART family neural networks. A generalized model of operation of an arbitrary ART family network is proposed. A methodology for planning a practical data analysis using the proposed generalized model is formulated. Implementation of the model stages for two different ART family networks is considered in detail: ART-2a and Fuzzy ART. A general approach to solving data clustering problems using the generalized model is proposed. Methods for assessment of the effectiveness of the continuous data stream clustering results are considered. Recommendations are developed for configuring the ART networks hyperparameters depending on the nature of the analyzed data.

1. Introduction
Clustering is a fundamental data analysis method. It has been widely used for pattern recognition, feature extraction, vector quantization, image segmentation, function approximation and data mining [1–3]. Clustering identifies common structures present in the dataset based on some measure of similarity. And the data stream clustering algorithms continuously process the incoming objects. Such algorithms must satisfy the following conditions [4]:

– return the results in real time;
– rapidly adapt to the changing data;
– be scalable to the number of the incoming objects;
– provide a data structure that does not expend excessively with processing great number of objects;
– detect the presence of emissions.

Moreover, the continuous data stream clustering algorithms must solve the so-called stability-plasticity dilemma. This problem lies in creation of a mechanism capable of both adapting to new images and preserving the existing images in its functioning [5]. Stability is an important aspect of continuous data stream analysis. Stable data attribution, in particular, means that if the same input vector is presented several times to the system input, it will sequentially be recognized as one belonging to the same category. Another important aspect of data stream analysis is plasticity, which determines the system’s adaptability to new data. Plasticity is a property of continuously learning systems. Stability and plasticity are, in fact, the conflicting algorithmic requirements.
In classical clustering algorithms (K-means, EM-algorithm etc.), the stability-plasticity dilemma is not solved [4]. One of the approaches developed in an attempt to solve this problem is creation of Adaptive Resonance Theory (ART), a theory developed by Carpenter and Grossberg [1, 5–7] and describing the neural networks family. ART networks are capable of generating stable clusters via self-organization in response to arbitrary sequences of input patterns. The ART family networks view the clustering problem as a task of adaptation in the operational mode according to the stability-plasticity principle; at the same time, a topology of neural network is formed and training of neural network is performed. The classifier is built via alternately adding or removing one neuron into/from the network structure, and the change of the classifier’s architecture does not involve retraining of the entire neural network, but only of its fragment associated with the given neuron, which accelerates the performance of the classification operation.

The ability of ART networks to efficiently process dynamic data makes them attractive candidates for template recognition in large rapidly changing datasets being created in real conditions. ART applications encompass many areas, including recognition of sonar signals [6], processing the information from sensors of modern technical systems [5], text clustering consisting in grouping of text documents in accordance with their content [3] etc.

However, many researchers note disadvantages of ART networks when working with infinite stream data [3, 4, 6] that reduce their efficiency. First, the result of clustering with the help of the ART family model is influenced by the order of presentation of the training sample [3, 6]. Another problem is the monotonous increase in the number of clusters in the process of network operation [4]. This article will discuss the approaches that allow overcoming these disadvantages.

As of today, the ART neural networks family includes about 10 different models based on the observance of the stability-plasticity principle. ART-1 network works only with binary input vectors. ART-2 and ART-2 clusterize continuous input vectors, and ART-2 operates considerably faster, but the qualitative results are sometimes inferior to ART-2 implementation. ART-3 network has a multi-layered architecture. During the layer-to-layer transition, it contrasts the input images and stores them in the form of increasingly general categories. ARTMAP network combines two slightly modified ART-1 or ART-2 networks into a single structure that learns with the teacher. Fuzzy-ART is a hybrid network that uses fuzzy logic in data processing.

At the same time, another problem associated with the practical use of ART family networks is that the authors of these models often describe them in an insufficiently formalized manner. They pay much attention to the concept of their functioning failing to give a clear idea of a specific model’s architecture and to provide structured algorithms for their training. This is why the researchers who use ART networks in their tasks, starting from the declared concept of a particular model, describe their architecture and training algorithms in absolutely different ways [3,8,9]. Thus, it can be stated that despite the diversity of models, and their relative simplicity, the literature lacks any common approach to description of ART family networks. Bridging this gap is among the goals of this article.

2. Materials and methods

2.1. Generalized model of ART network functioning

Let us formulate the general principles ART family network functioning and highlight their common stages. Most of ART family networks actually consist of one layer of neurons, and in the process of network operation each of them turns into a prototype of a cluster that corresponds to it. ART network expresses its classification decision in the form of excitation of one of the neurons of this layer. If the input vector does not correspond to any of the stored images, then a new cluster is created by storing an image identical to the new input vector. If it is determined that the input vector is similar to one of the previously stored prototypes from the point of view of a certain similarity criterion, this prototype will change (“learn”) under the influence of
the new input vector in such a manner as to become similar to this input vector. The stored image will not change if the current input vector is not sufficiently similar to it. Thus, the stability-plasticity dilemma is solved. The new image can create additional clusters; however, it cannot destroy the existing memory.

The results of the system analysis of the existing training algorithms for ART networks [1, 3, 5–10] allows presenting the network operation in a general form, namely in the form of the following generalized model [11].

**Stage 1. Initializing the network and its parameters.** At this stage, one working neural layer is created, which will store the resulting neurons representing the prototypes of clusters. The network has a regular layer of input neurons associated with all the prototype layer neurons with weights $w_{ij}$. Also, initial values of the network control parameters are set at this stage, such as learning rate $\beta$ or similarity parameter $\rho$. With the help of these parameters, the network operation is adjusted in accordance with the desired result.

**Stage 2. Preprocessing of input vectors.** Prior to applying the vector onto the input neurons layer, it is processed based on the specific type of network and basic knowledge of the nature of the input data. This can be normalization, noise removal, complementation and other transformations necessary for functioning of the corresponding subtype of networks.

**Stage 3. Selecting clustering category (using the choice function).** At this stage, the network performs a primary assessment of the input vector using a certain similarity function based on one of the existing metrics of inter-vector distance.

**Stage 4. Verifying similarity.** At this stage, some neuron from the prototype layer is activated, and its weights, along with the input vector, are applied to the input of the matching function that is responsible for a more detailed analysis of similarity and imparts a nonlinearity to the recognition process. If the matching function successfully verifies the prototype and the vector, the next stage comes on — prototype training. Otherwise, the network deactivates the neuron that has reacted to the choice function (removes it from consideration at the next iteration) and restarts the selection process.

**Stage 5. Training.** At this stage is performed the adjustment of weights of the prototype of the cluster that caused the choice function to react with the help of a corresponding training function. If none of the existing prototypes has been successfully verified by the compliance function, then a new cluster is created based on the unrecognized vector. A new neuron is activated whose weights are initialized by the coordinates of the given vector.

**Stage 6. Neural layer postprocessing.** When the training process is over, postprocessing of the entire neuron layer can be provided for in the network; for example, when a limit is set for the network with respect to the total number of clusters. The clusters can be combined or separated in accordance with the required conditions.

Thus, when solving a certain clustering problem, it is necessary to decide about selecting the variation of an ART family network model (which depends on the selected data structure) and perform the presented algorithmic stages. As an example, let us consider the implementation of the stages described in the scheme for two different ART networks.

2.2. Implementation of the generalized model stages for ART-2 network

ART-2 network has the following parameters: $\beta$ — learning rate, $\theta$ — normalization parameter and $\rho$ — similarity criterion [5], [1]. As inputs, the network uses arbitrary dimension vectors with real components. Input vectors $Y^i$ ($i = 1, 7$), before being applied to the network input, are normalized to a unit length, following which the coordinates of these vectors, whose value has become less than parameter $\theta$, are equated to 0, and the vectors are normalized once again. Thus, random small perturbations in the initial data that reduce the clustering quality are suppressed. The cosine function of the angle between the corresponding vectors is used as a similarity function, which in the case of normalized vectors is equal to their scalar product.
\[ T_k = M_k = (Y^i, w^k) = \sum_{j=1}^{J} y_{ij}w^k_j, \]  

where \( T_k \) denotes the value of the choice function for the \( k \)-th cluster, \( M_k \) is the corresponding value of the matching function, \( w^k \) is the current prototype vector of the \( k \)-th cluster, \( J \) is the input vector dimension. Thus, at the 3\(^{rd} \) stage of the presented generalized model, a cluster with number \( s = \arg \max_k T_k \) is selected, and at the 4\(^{th} \) stage condition \( M_s > \rho \) is verified for it, where \( \rho \) is the so-called similarity criterion, which is the conditional assessment of the required homogeneity of the vectors within the cluster. Note that with the functions set in this way there is no need for a mechanism for resetting (temporary deactivation) of neurons not meeting the compliance condition. It simplifies the algorithm structure and accelerates its functioning.

If the cluster prototype satisfies the similarity condition, training of this prototype is performed by changing its weights in order to make it more similar to the input vector:

\[ w^{k+1} = \frac{(1 - \beta)w^k + \beta \cdot Y^i}{\| (1 - \beta)w^k + \beta \cdot Y^i \|}, \]

where \( \beta \) is a learning rate parameter. After this, a new vector of the prototype weights is normalized. Parameter \( \beta \) is selected depending on whether it is required to carry out the identification of long-term patterns (small parameter values) or to monitor the data change trends (high values). Selection of similarity level \( \rho \) is more complicated since the general idea of the initial data structure may be lacking in general case. This results in the main disadvantage of ART-2a network, namely the uncontrolled increase in the number of clusters during operation with large parameter values, and the merging of data into one large cluster with small ones. This is why, when using this network, it is necessary to gradually decrease the learning rate parameter in order to stabilize the number of clusters.

2.3. Implementation of the model stages for Fuzzy ART network

Another way to work with real input vectors is Fuzzy ART network, which requires data scaling within the limits \([0, 1]\). Network operation is regulated by parameters \( \alpha, \beta, \rho \) where \( \beta \) and \( \rho \) are used by analogy with ART-2a network, while \( \alpha \) is an ultra-small number (of order of \( 10^{-6} \)), which prevents the ultimate degeneration of prototypes. One of the possible options of implementing ”fuzzy AND” is used as a choice function:

\[ T_k = \frac{|Y^i \wedge w^k|}{\alpha + |w^k|}, \]

where operator \( \wedge \) is defined as \( (p \wedge q)_j = \min(p_j, q_j), j = \overline{1, J} \), and the norm is \( |p| = \sum_i p_i \).

Unlike ART-2a, a reset mechanism is used in this network — a temporary cluster deactivation when the similarity condition is not met. This is why, the selection of the cluster at the stage of selecting the clustering category is defined as \( s = \arg \max_k T_k, k \in Q \), \( Q \) is the set of nodes active (not deactivated) at this stage, following which the selected prototype is verified against the input vector using the matching function.

The matching function is determined as follows:

\[ M_s = \frac{|Y^i \wedge w^s|}{|Y^i|}. \]  

The cluster is selected when condition \( M_s \geq \rho \) is met, where \( \rho \) is a similarity criterion. In case the condition is violated, the cluster is marked as inactive and the choice function is called
If there are no active clusters left, a new one is created with weights equivalent to the input vector. Training function: the weights of the verified prototype are modified as

\[ w_{t+1} = (1 - \beta) w_t + \beta (Y^t \land w^t) \]  

(5)

Since in this option the prototype weights are not normalized after training, a problem arises associated with the training function. The used operator of the "fuzzy AND" results in a constant trend to decrease the vector coordinates values. This decrease, under certain conditions, may result in degeneration of the cluster prototypes. In order to solve the vector degeneration problem at the preprocessing phase, it is suggested that the so-called complementary coding be applied to the input vectors – the \( J \)-dimensional vector turns into \( 2J \)-dimensional via completing it with \( J \) components: \( y_{i(J+j)} = 1 - y_{ij} \), which preserves its amplitude information while equitably reflecting the minima and maxima.

2.4. Solving the data clustering problems using ART family networks

The general approach to solving the data clustering problems using the two considered ART family networks is as follows.

Stage 1. Selecting the preprocessing method. At this stage, preliminary processing of input data is performed using one or more methods.

1.1) Coordinate ranking. With the data on the importance or priority of individual input attributes at hand, it is possible to somehow rank the sequence of coordinates, and then to scale the original vectors in accordance with these priorities (for example, the first coordinate is scaled to segment \([0; 1]\); the second one is scaled to segment \([0; 0.95]\); and so on). This will obviously change the clustering logic in accordance with our wishes.

1.2) Normalization. In ART-2a network, it is preferable to deal with normalized input vectors. Work [8] suggests using normalization for Fuzzy ART networks as well, completing each vector with one more coordinate equal to its length before normalization. The methodology has its pros and cons described in the corresponding work.

1.3) Noise removal. Noises (erroneous values) in coordinates are easily eliminated by introducing into the network logic a parameter that defines the limits of the possible data element values, they being exceeded, the respective coordinates will be scaled.

1.4) Complementation. This is a process that doubles the dimensionality of the input vector, which allows preventing the degeneration of prototypes when using the selection and training functions close in its logic to those used in a Fuzzy ART network.

Stage 2. Defining the choice, matching and training functions.

After preprocessing, a respective set of choice, matching and training functions is defined. The generalized model proposed in section 2.1 allows easily replacing these functions at any time during processing, since the format and general logic of the network operation are essentially similar for all cases.

Stage 3. Selecting the postprocessing method.

This stage allows compensating for the disadvantages of a certain ART family model. Let us dwell on some postprocessing examples.

For example, one of the common disadvantages of the family is that it is difficult to select similarity parameter \( \rho \) or its analogue — some integral estimation of the vectors’ similarity. Its wrong selection may result, in extreme cases, in merging of all data into one cluster or an uncontrolled increase in the number of categories.

As a means to overcome this problem, we can use the postprocessing method aimed at limiting the number of clusters. Since the number of clusters required to break up the initial data is not always obvious in advance, it is suggested that the following approach be used to determine the desired similarity parameter. The upper limit of the permissible number of clusters is set.
At the beginning of operation, parameter $\rho$ is set close to 1. If the next vector does not satisfy the matching function for any of the existing clusters, and their number is already equal to the limiting one, it is suggested that the following steps be performed.

3.1) The current input vector is added to the existing clusters as one more new cluster.

3.2) $R_{ml} = \max\{R_{ij}\}, \forall i, j = 1, K, i \neq j$ is sought, where index pair $(m, l)$ is the number of clusters whose prototypes are closest in the sense of matching functions, $R_{ij}$ is the distance between the prototypes of clusters $i$ and $j$. Then objects of cluster $l$ are trained by the prototype of cluster $i$ in accordance with the set training function.

3.3) Cluster $m$ is deleted.

3.4) If found $R_{ml} < \rho$, then $\rho = R_{ml}$ is set.

The described approach is quite universal; it can apparently be used to modify both ART-2a and Fuzzy ART networks.

2.5. Clustering effectiveness estimation

Solution of the clustering problem is essentially ambiguous since there is no definitely best criterion of the clustering quality. Estimation of the clustering efficiency is complicated by the fact that true the marks of the clusters are not always known even for some parts of the objects, and therefore metrics are needed allowing estimating the clustering quality using only the unmarked sample. For evaluation of the efficiency of ART family networks in data stream clustering problems, it is proposed to use Silhouette Coefficient [12] as such a metric. Initially, silhouette $s$ is determined for each object:

$$s = \frac{b - a}{\max(a, b)},$$  \hspace{1cm} (6)

where $a$ is the average distance from this object to all other vectors in the same cluster, and $b$ is the average distance from the given object to all vectors from the nearest cluster (different from the one in which the object itself is located). Coefficient of a certain data set silhouette is the average value of silhouettes of all objects in the given set. The silhouette coefficient shows how much the average distance between the objects inside the clusters differs from the average distance to the objects of other proximate clusters. The value of the silhouette coefficient lies in the range $[-1, 1]$. The higher the coefficient value, the better the performed clustering.

A disadvantage of the described clustering efficiency estimation approach is that the silhouette coefficient depends on the shape of the clusters and is more informative in case of convex clusters.

If the reference clustering of at least some set of objects is known, then it is advisable to use metrics that characterize the degree of predictive clustering proximity to it. One of such metrics is Rand index [9]. Let $n$ be the number of marked objects in the sample. Let $a$ denote the number of marked object pairs having the same marks and assigned by the algorithm to the same cluster, and let $b$ denote the number of marked object pairs having different marks and assigned by the algorithm to different clusters. Then Rand index is calculated by formula:

$$RI = \frac{2(a + b)}{n(n - 1)}.$$  \hspace{1cm} (7)

That is, Rand index is the share of object pairs for which the partitions, both original and resulting from the algorithm operation, are ”matched”. The values close to unity indicate that the algorithm forms a cluster structure that is close to the reference clustering. The index value decreases with the increase of the number of errors.

A considerable advantage of Rand index is that it disregards the shape of clusters and can be used to estimate the cluster structuring quality for a wide range of practical problems.
3. Results and discussion

3.1. ART-2. Experiments

As part of this study, a computational experiment was performed with ART-2a and Fuzzy ART networks for processing a continuous data stream. The data stream was simulated using scikit-learn machine learning library, which includes various random sample generators that can be used to create artificial datasets of controlled size and complexity [12]. The make_blobs procedure allows creating multi-class datasets, while allocating one or more normally distributed clusters to each class. With the help of this procedure, a stream of 10-dimensional data containing three clusters was generated, one of which is "easily separable", and the other two are closely spaced. The following results were obtained in the course of the computational experiment with ART-2a network.

1. For similarity parameter values $0 < \rho < 0.88$, the network combines all input elements into one cluster, regardless of the parameter of learning rate $\beta$.
2. Further, for $0.88 \leq \rho < 0.984$, a partitioning into two clusters takes place, that is the network performs separation of the "easily separable" cluster 1, again, regardless of the learning rate parameter.
3. Next stage occurs at values $\rho = 0.984$ and $\beta = 0.9$. With these parameters, the 3rd "hardly separable" cluster was formed for the first time. However, with these parameter values the network does not converge in 500 iterations; "rattling" of the clusters’ limits is observed. But if value $\beta$ is smoothly reduced down to 0.1, then in 10 iterations we obtain a practically "ideal" (corresponding to the model one) distribution of vectors.
4. Finally, at value $\rho = 0.99$, value $\beta = 0.3$ is already sufficient for the data to be split into 3 clusters. Model clustering can be achieved in the same manner, by smoothly reducing value $\beta$ down to 0.1. At $\beta > 0.3$, the network increases the number of prototypes up to 4, splitting the set in a more detailed manner. At the same time, the subsequent rate decrease will already not help reduce the number of clusters, but it will impart more density to the created clusters from the point of view of prototypes, and it will ensure convergence over a small number of iterations.

In case of "correct" selection of parameters in this particular example we managed to obtain a data partitioning into 3 clusters, completely corresponding to the model one. However, this occurred at a low learning rate and required a large number of iterations, which is not always advisable in the online processing conditions. Thus, ART-2a technology is suitable for analyzing long-term trends over a large amount of data or a data stream of great saturation.

3.2. Fuzzy ART. Experiments

The following results were obtained during the computational experiment with Fuzzy ART network.

1. Up to value $\rho = 0.68$, the network combines all the input vectors into one cluster. It should be noted that the critical parameter value is located in a completely different area than when using ART-2a.
2. At $0.68 < \rho < 0.85$, the network begins to divide the set into two clusters, one being "easily separable" and the other one consisting of the remaining elements. However, due to degeneration of the prototypes, in case of continuation of the learning process, generation of redundant clusters is observed, or combination of the previously separated ones into a single one.
3. At $0.85 < \rho < 0.9$, distribution into 3 or 4 clusters occurs. The allocation of individual small clusters is still retained at average values of $\beta$, but for high $\beta$ the network already gives "quite good" (in the sense of proximity to the model distribution) results.
4. Finally, with further increase of $\rho > 0.9$, the network generates an increasingly greater number of clusters, dividing the space into many small groups.
On the whole, it should be noted that set partitioning into clusters using Fuzzy ART network, on average, demonstrates worse compliance with the "model" classification than ART-2a does, but it has such positive features as stability and convergence, as well as weak dependence on selection of the learning rate.

The computational experiment also revealed that choice, training and matching functions used in Fuzzy ART lead to appearance of networks' high sensitivity to the order of the supplied data. The more chaotic the data stream, especially at the beginning of the operation, the better the distribution over clusters.

### 3.3. Recommendations for processing of continuous data stream

Let us provide some general recommendations practically obtained during the computational experiment with ART-2 and Fuzzy ART networks for continuous data stream processing.

Network selection in each particular case largely depends on the data type and the operation mode according to which it will function. ART networks, as noted above, are sensitive to the input presentation order and, depending on this order, can generate different clustering results. This is why, performing several analysis iterations in the "batch" mode is recommended at the initial stage.

Selection of learning rate \( \beta \) parameter depends on the type of the problem being solved: a low learning rate allows revealing general consistent patterns in the data, while a high learning rate allows catching the current trends promptly.

The following approach is suggested in order to adjust similarity parameter \( \rho \). Initially, the upper limit for permissible number of clusters \( K \) should be specified. At the beginning of the network’s functioning, it is appropriate to set parameter \( \rho \) to 1 (or close to 1). If the next input vector does not satisfy the matching function for any of the existing clusters, and their number is already equal to the limiting one, the value of parameter \( \rho \) decreases until it becomes possible to either assign the vector to one of the clusters or combine any two clusters into one.

It is recommended that the initial analysis of the "basic package" of data be carried out using Fuzzy-ART network, which ensures rapid convergence of cluster prototypes. If, even for low values of similarity of parameter \( \beta \), the "basic package" data demonstrate rapid and stable data separation into clusters up to sufficiently large values of \( \rho \), then Fuzzy-ART network should be used for further analysis of the main data stream, and the number of clusters should be limited to the one that by 2–3 exceeds the desired one, in order to prevent the uncontrolled increase in their number in the event of noise.

If the network converges very slowly at small values of \( \beta \), and a continuous "migration" of clusters is observed, Fuzzy-ART model may only be applied to the data arriving in a volumetric yet discrete stream (for example, n times per day), because they can be processed in a "batch" way in this case. If the data arrive in a continuous stream in the described situation, ART-2a network should be used for simulation. Parameter \( \beta \) will be self-regulated by the network, and the use of the training function from ART-2a network will result in the acquisition of meaningful values by the prototypes. The number of clusters will vary inconsiderably, which will allow tracking the emerging trends.

If it turns out that "bursts" become characteristic of the data stream with the appearance of atypical input vectors that are stored in new clusters and lead to exhaustion of the limit on the number of clusters, then, depending on the problem being solved, the operator may either study these clusters (for example, if data anomalies search problem is being solved in the data), or remove them (as a noise component), or lessen the requirements as to the number of clusters, which, in the conditions of the input stream chaotic character, will lead to the absorption of these mini-clusters by the main ones at certain moments.
4. Conclusion
Cluster analysis is one of the key methods in the spheres of processing continuous streams of large data volumes. This article presents the following results that allow increasing the performance of data clustering systems built on the ART family networks on a real-time basis.

1. A generalized model of ART family neural networks has been constructed.
2. Practical implementation of the stages of ART-2A and Fuzzy ART networks generalized model has been demonstrated.
3. Results of a computational experiment for continuous data stream processing by ART-2A and Fuzzy ART networks have been provided that allow making conclusions about these networks’ behavior.
4. A methodology for planning a practical analysis of a continuous data stream using the proposed generalized model has been formulated.
5. Recommendations have been developed for configuring the hyperparameters of ART networks depending on the nature of the data available.

Thus, using the obtained results, it is possible to form an efficient cluster structure of a continuous data stream with the help of ART family networks.

References
[1] Grossberg S 2013 Neural Networks 37 1–47
[2] Kashirina I L, Lvovich Y E and Sorokin S O 2017 Informacionnye technologii 23 228–232
[3] Massey L 2005 Proceedings of 2005 international joint conference on neural network
[4] Skovajsova L and Rojcek M 2011 IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI) (Budapest) pp 439–442
[5] Carpenter G and Grossberg S 2003 Adaptive Resonance Theory. The Handbook of Brain Theory and Neural Networks (Cambridge, MA: MIT Press)
[6] Carpenter G, Grossberg S and Rosen D 1991 Neural Networks 4 759–771
[7] Carpenter G, Grossberg S and Reynolds J 1991 Neural Networks 4 565–588
[8] Keskin G A and Ozkan C 2009 QUality and reliability engineering international 25 647–661
[9] Wasserman P 1989 Neural Computing. Theory and Practice (New York: Van Nostrand Reinhold)
[10] Rosenberg A and Hirschberg J 2007 In Interspeech-2007 2777–2780
[11] Kashirina I L, Lvovich Y E and Sorokin S O 2015 Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Sistemnyj analiz i informacionnye technologii 3 62–67
[12] Pedregosa et al 2011 JMLR 12 2825–2830