Perspectives of intellectual processing of large volumes of astronomical data using neural networks

A A Gorbunov¹, E A Isaev¹,², V A Samodurov¹,²

¹National Research University Higher School of Economics
²Pushchino Radio Astronomy Observatory of Astro Space Center of P.N. Lebedev Physics Institute, RAS, Pushino, 142290

agorbunov@hse.ru, eisaev@hse.ru, vsamodurov@hse.ru

Abstract. In the process of astronomical observations vast amounts of data are collected. BSA (Big Scanning Antenna) LPI used in the study of impulse phenomena, daily logs 87.5 GB of data (32 TB per year). This data has important implications for both short-and long-term monitoring of various classes of radio sources (including radio transients of different nature), monitoring the Earth's ionosphere, the interplanetary and the interstellar plasma, the search and monitoring of different classes of radio sources. In the framework of the studies discovered 83096 individual pulse events (in the interval of the study highlighted July 2012 - October 2013), which may correspond to pulsars, twinkling springs, and a rapid radio transients. Detected impulse events are supposed to be used to filter subsequent observations. The study suggests approach, using the creation of the multilayered artificial neural network, which processes the input raw data and after processing, by the hidden layer, the output layer produces a class of impulsive phenomena.

1. Formulation of the problem

The basis for the research in this study is the data recorded by Big Scanning Antenna (BSA) LPI. BSA LPI daily registers 87.5 GB of data (32 TB per year). For structuring the flow of "big data" researchers use a technique based on the methods of mathematical statistics [1,2]. In some cases, automatic classification applied for statistical processing produces incorrect results and the resulting impulses have to be checked and sorted visually [1]. In this regard, there is a need for more effective classification methods, allowing to obtain more reliable results.

Among the papers considering the problems of classification of impulse phenomena are noticeable developments using a neural network. For example, in [3] shows the results of using multilayer perceptron and self-organizing Kohonen network to solve the problem of recognition of the impulses of the meteors. There is also a patented method that allows determining the structure and demodulation of the signal with unknown structure by using complex multi-stage study characteristics [4]. At a certain stage of this method supposes the use of Kohonen neural network [4].

It should be noted that the main focus of studies aimed at determining the possibilities of using neural networks to classify impulse phenomena. With the current state of research, it remains unclear whether it is possible to use classifiers based on neural networks to process large amounts of astronomical data. In the work [1] were identified 83096 impulse phenomena. It is required to determine the possibility of a multilayer neural network for classification the raw data. Conducting...
practical experiments for training artificial neural networks with thousands of elements shows that it takes several tens of hours to the network began to show signs of intelligent behavior [5].

The aim of this work is to develop the neural network for structuring and processing large amounts of astronomical data.

To achieve the goal the following tasks are set:

• Analysis of the structure of the input/output data necessary for the design of a neural network;
• Conducting experiments using different types of neural networks.

2. Analysis of the structure of a stream of input/output data

In order to determine the number of neurons in the input layer of the neural network, it is necessary to explore the structure of the data stream. Astronomical data [1] processed by two recorders, consisting of 6 modules. Each module registers 8 signals (rays). The signal for each beam is divided into 6 bands of frequency ranges. Additionally, it registers a signal with a common bandwidth of the radio telescope. The calculation of the statistical characteristics of the input data for each band (from 1 to 6), for each of the 96 beams is produced every 5 seconds.

Thus, to classify the 5-second observations requires a neural network with a number of inputs 96*6 (number of rays*number of bands) or 576.

The number of neurons in the output layer is determined by the number of observed classes of impulse phenomena. The result of statistical processing of data, registered by BSA LPI during the period from 2012 to 2013, has been allocated the following classes of phenomena:

• impulses of technogenic character of various kinds;
• hardware failures data;
• impulses having the character of single flickering radio sources of different types;
• impulses having well-defined space character.

Thus, the output layer consists of 4 neurons, the outputs of which will indicate a particular class of observable data.

3. Conducting experiments using different types of neural networks

Experiments on the recognition of different classes of impulse phenomena were carried out for a multilayer perceptron and self-organizing Kohonen network. Neural network data types were used in [3] to identify observations corresponding to the meteors. For training neural networks used 5 minute recording at a frequency of 2.5 kHz. The result of this research is a neural network, is able to divide the observations into 2 classes: the meteors and non-meteors.

The basis of the experiments, the results of which are given in the framework of this work are the classes of impulse phenomena identified during observations in the period from 2012 to 2013. The complex flow of astronomical data (see section 2) must be classified by the neural network. Statistical processing allowed to reveal 83096 different phenomena, which can be divided into 4 classes (see section 2). This sample is used as training dataset. It is expected that after training, the neural network will be able to divide the observations into 4 classes.

Implementation and experiments with neural networks have been performed using MATLAB tools.

3.1 Multilayer perceptron

A multilayer perceptron [5] consists of input, hidden, output layer. The hidden layer may consist of one or more layers. The input layer has no synaptic weights, and thus not trained. A multilayer perceptron is trained with the teacher, that is, the input is a vector of observations, and in the output, and layer checks the consistency of the result of the neural network by the previously known answer.

If the expected response does not match the output result, the adjustment of synaptic weights of the hidden layer is produced. Thus, using the examples you can get from a neural network solving the task: the classification of astronomical observations. The neural network is trained using learning algorithms: backpropagation, gradient descent, Levenberg–Marquardt.
The results of applying different learning algorithms are presented in table 1.

**Table 1.** The results of training the neural network

| Algorithm                        | Training sample size | Test sample size | Recognition rate (training sample), % | Recognition rate (test sample), % |
|----------------------------------|----------------------|-----------------|----------------------------------------|----------------------------------|
| Backward propagation             | 62322                | 20774           | 84.69%                                 | 69.43%                           |
| Gradient descent                 |                      |                 | 93.41%                                 | 59.64%                           |
| Levenberg–Marquardt algorithm    |                      |                 | 98.73%                                 | 88.97%                           |

Pre-fetch from 83096 classified observations, divided into training (75%) and test sets (25%). The best classification result of test data achieved when using the training algorithm of Levenberg – Marquardt. Percentage of recognition data to the training sample is not increased 100% in order to avoid overfitting. The duration of the learning process was 1000 epochs.

### 3.2 Kohonen network

In contrast to the multilayer perceptron that uses the algorithms of supervised learning, Kohonen network [5,6] does not require the filing of the expected values of the classification at the output layer. The Kohonen network consists of an input and topological layers.

As a result of self-organization of a network of neurons in the topological layer are grouped together, forming clusters. Thus, when the input of the unclassified observations-organized neural network determines its membership to one of the clusters. In our case, for self-organizing network, observations from 4 classes of impulse phenomena forming the training sample is sequentially applied to the input Kohonen neural network and thereby cause the neurons to be grouped into clusters.

### 4. Evaluation of results

After training the neural network (p.3.1, 3.2) were used for the classification of astronomical observations. The classification results are shown in tables 2, 3.

**Table 2.** The results of the classification using the Kohonen network

| Class                                | Recognized observations | Unrecognized observations | Recognized as the objects of another class |
|--------------------------------------|-------------------------|---------------------------|---------------------------------------------|
| impulses of technogenic character of various kinds (3116) | 2939                    | 177                       | 14                                          |
| hardware failures data (4155)        | 3871                    | 283                       | 5                                           |
| impulses having the character of single flickering radio sources of different types (5194) | 4957                    | 237                       | 50                                          |
| impulses having well-defined space character (8310) | 8105                    | 204                       | 26                                          |

Using the methods of classification, based on methods of mathematical statistics [1], misclassified remain 10-15 % of cases.
Table 3. The results of classification using multilayer perceptron

| Impulses Classification                                                                 | Recognized Observations | Unrecognized Observations | Recognized as the Objects of another Class |
|----------------------------------------------------------------------------------------|--------------------------|---------------------------|------------------------------------------|
| Impulses of technogenic character of various kinds (3116)                              | 2732                     | 385                       | 57                                       |
| Hardware failures data (4155)                                                          | 3639                     | 516                       | 30                                       |
| Impulses having the character of single flickering radio sources of different types (5194) | 4656                     | 537                       | 63                                       |
| Impulses having well-defined space character (8310)                                    | 7558                     | 751                       | 47                                       |

The use of Kohonen neural network shows the smallest percentage of unrecognized data (average 5%) as compared to the use of the classifier based on multilayer perceptron (average 11%) and in comparison with the methods of statistical processing [1].

5. Conclusion

In this work, we investigate the question of the applicability of different technologies of neural networks for classification of large volumes of astronomical observations. As the basis of research used data from round-the-clock observation was BSA LPI in the period 2012 – 2013 the analysis of the structure of the flow of astronomical data. Based on the results of the research presented in [1] determined the resulting classes is required to recognize using neural network technology.

Experiments on learning the two types of neural networks: multilayer perceptron and self-organizing Kohonen network. The classification accuracy of astronomical observations made with neural networks is higher than when using the methods of [1], based on the methods of mathematical statistics.

References

[1] Samodurov V A, Dumsky D V, Isaev E, Rodin A E, Kazancev A N, Fedorova V A, Belyatskij Y A 2016 The daily 110 Mhz radio wave sky survey: statistical analysis of impulse phenomena from observation in 2012-2013 Odessa Astronomical Publications 29 pp 167-170.
[2] Taylor G B et al: 2012 First Light for the First Station of the Long Wavelength Array J. Astron. Instrum. 1 pp 1–56
[3] Victor S R, Ca’ta’lin B 2015 Automatic Analysis of Radio Meteor Events Using Neural Networks J Earth Moon Planets. pp 101-113
[4] Kuzovnikov A V, Anzhina V A, Pashkov A E et al 2010 Sposob opredeleniya struktury i demodulyatsii signala s neizvestnoy strukturoy Patent RF № 2386165 / 10
[5] Tadeusevich R I et al 2011 Elementarnoe vvedenie v tehnologiyu neyronnykh setey Goryachaya liniya telekom
[6] Kohonen T 1990 The self-organizing map. Proc. IEEE 78 pp 1464–1480
[7] Konovalenko A A et al 2016 Astronomicheskie issledovaniya s pomoshch’yu malorazmernykh nizkochastotnykh radioteleskopov novogo pokoleniya Radiofizika i radioastronomiya 21 pp 83-131.
[8] Tyul’bashev S A, Tyul’bashev V S 2015 Otkrytie novyh pul’sarov na radioteleskope BSA FIAN Astronomicheskiy tsirkulyar 1624
[9] Rodin A E, Samodurov V A, Oreshko V V 2015 Bystryy metod obnaruzheniya periodicheskikh radioistochnikov na BSA FIAN Astronomicheskiy tsirkulyar 1629