PAPER • OPEN ACCESS

Energy-efficient algorithm for classification of states of wireless sensor network using machine learning methods

To cite this article: M N Yuldashev *et al* 2018 *J. Phys.: Conf. Ser.* **1015** 032153

IOP Conf. Series: Journal of Physics: Conf. Series **1015** (2018) 032153 doi:10.1088/1742-6596/1015/3/032153

Energy-efficient algorithm for classification of states of wireless sensor network using machine learning methods

M N Yuldashev[1], **A I Vlasov**[1], **A N Novikov**[2]

[1] Bauman Moscow State Technical University, Department of Design and Technology of Electronic Devices, 5/1, 2-ya Baumanskaja Str., Moscow, 105005, Russian Federation
[2] Orel State University, Department Service and repair of machines, 95, Komsomol'skaya Str., Orel, 302026, Russian Federation

Abstract. This paper focuses on the development of an energy-efficient algorithm for classification of states of a wireless sensor network using machine learning methods. The proposed algorithm reduces energy consumption by: 1) elimination of monitoring of parameters that do not affect the state of the sensor network, 2) reduction of communication sessions over the network (the data are transmitted only if their values can affect the state of the sensor network). The studies of the proposed algorithm have shown that at classification accuracy close to 100%, the number of communication sessions can be reduced by 80%.

1. Introduction
Most studies of wireless sensor networks (WSNs) are aimed at improving such features as efficiency, reliability and security of the system. This article focuses on improving the energy efficiency of WSNs. Since most of the network nodes operate autonomously, the nature of the use of energy resources directly affects the life cycle of the network. A significant part of the energy of network nodes is consumed by the process of communication via wireless communication. A simple solution to reduce the number of communication sessions generates a technical contradiction: “With an increase in the number of communication sessions, the energy consumption of nodes increases, which leads to a reduction in the time of autonomous operation of the node; with a decrease – the probability of missing important events increases, which may lead to incorrect decision making” [1].

The paper contains the study of the method for solving the above-mentioned technical contradiction by a dynamic change in the data transmission frequency depending on the trend of the sensor readings change.

2. Materials and methods
To determine the state of the network, the data of the sensors are used. The proposed method for increasing the energy efficiency of the sensor network is based on the hypothesis that there is a subset of controlled parameters whose contribution to the final decision is insignificant or does not exist. Thus, excluding the above mentioned data will not affect the functioning of the system and the energy will be used more efficiently [2].

To test this hypothesis, the simulation of the sensor network operation was performed. The sensor network consists of 20 sensors, each of which every minute provides data on 4 parameters of the environment: temperature, pressure, humidity and illumination. The dynamics of the change in readings was performed with the help of Gaussian processes [3]. The results of the initial data are shown in Figure 1.

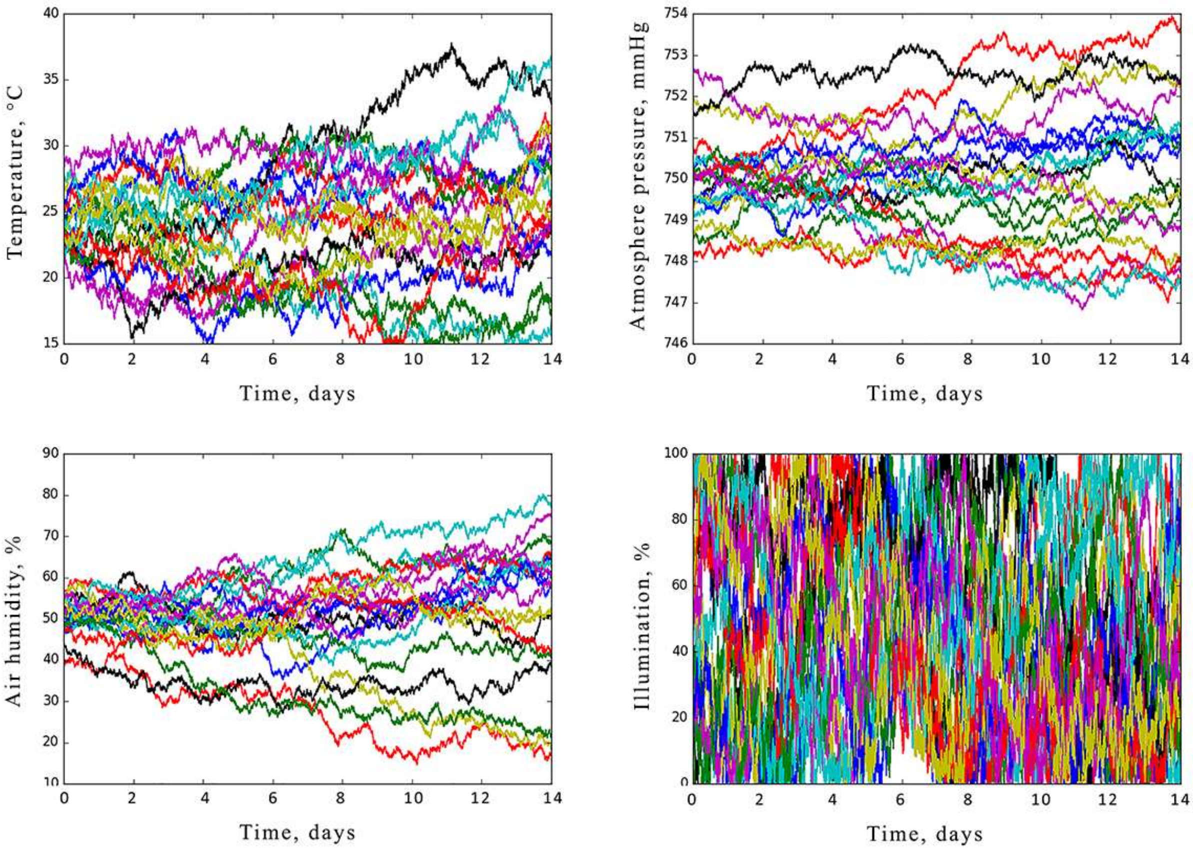

Figure 1. Readings of the initial data of the sensor network according to the parameters: temperature, pressure, humidity and illumination.

Next, a linear algorithm for classifying network states was developed, which compared the temperature and humidity data at a particular point of time with one of 5 states: 0 – normal, 1 – high temperature, 2 – ignition, 3 – high humidity, 4 – supercooling. At the same time, data on pressure and illumination were not used. The result of the classification is shown in Figure 2.

As a classification model, a decision tree was selected. The selection is conditioned by the fact that in similar models the calculation process is based on conjunctive patterns consisting of a set of predicates for input data, which corresponds to the process of decision making by the operator. In addition, decision trees allow performing classification with gaps in the input data, which takes place in the sensor networks.

At the first stage, 80% of the data, on the basis of which the decision tree was constructed using the ID3 algorithm, was selected randomly. At the second stage, the remaining 20% of the data was classified using the model obtained. To assess the reliability of the constructed model, the accuracy was calculated:

$$Accuracy = (T) / (T + F) = 4030 / 4032 = 99\%,$$

where T – the number of correct responses, F – the number of incorrect responses.

Thus, the constructed model showed that it can classify the state of the sensor network with a probability of 99%.

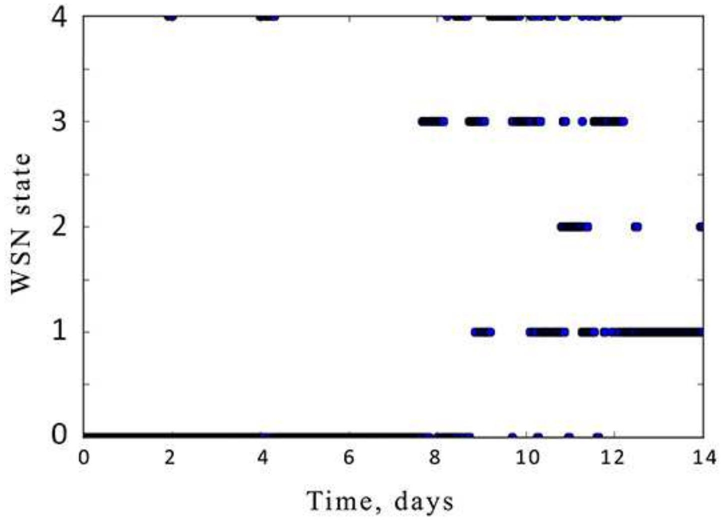

Figure 2. The graph of the distribution of states of the sensor network (*0 – normal, 1 – high temperature, 2 – ignition, 3 – high humidity, 4 – supercooling).*

The decision tree contains a set of predicates, each of which is associated with one of the parameters of the input data; at the same time, if the parameter does not have a predicate, it does not affect the classification process. For the current model, a graph of the distribution of weights for the parameters of the input data was made (see Figure 3).

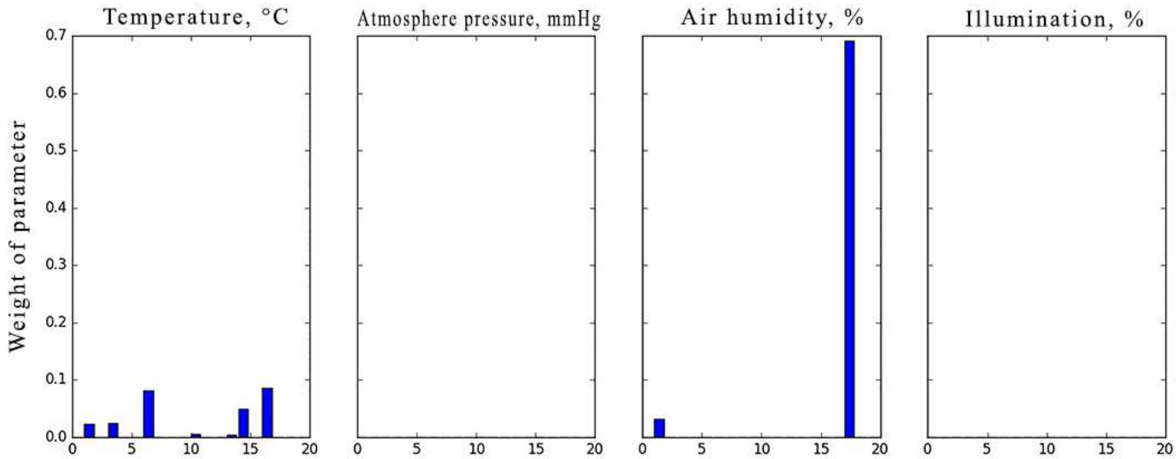

Figure 3. Distribution of the weights of the decision tree with respect to input parameters.

Further, zero-weight readings were excluded from the test sample, a new model of the decision tree was trained, and testing was conducted. In this case, the classification accuracy was obtained:

$$Accuracy = (T) / (T + F) = 4030 / 4032 = 99\%.$$

As a result of excluding parameters with zero weights from the input data, similar accuracy is obtained. On the basis of the results obtained, it can be concluded that the energy consumption by the process of monitoring the parameters with zero weights is not justified and can be eliminated. It is also possible to exclude parameters whose weights are very close to zero. In this case, the classification accuracy can be slightly lower. The search for a compromise between the accuracy and the number of parameters depends on the problem being solved and in each case must be solved individually.

3. Energy-efficiency for data transmission between the elements of the sensor network

The control module makes a decision about the state of the sensor network based on the set of the data transmitted from the sensors. The process of computing the network state using the decision tree is idempotent, that is, if the same input data is provided, the result is identical. Thus, if the data, transmitted from the sensors, have not changed, then the network state remains the same.

The calculation of the network state is carried out by means of the alternate comparison of the readings of the sensors with the predicates obtained earlier in the training of the model. The comparison continues until a leaf of the tree with the value of the particular state of the system is achieved [4]. An example of a decision tree obtained when training using the data, generated in Section 2, is shown in Figure 4.

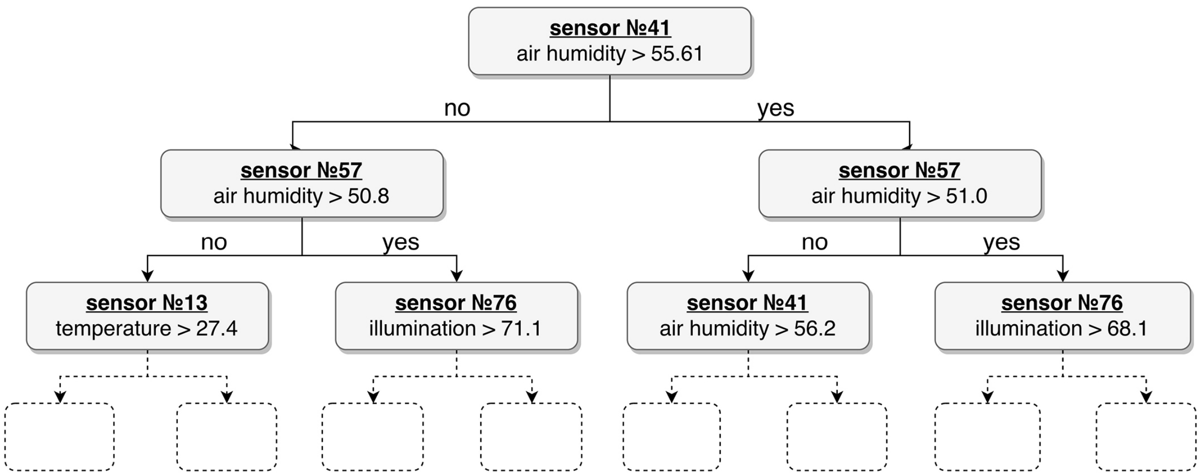

Figure 4. The decision tree (the first 3 levels).

The proposed algorithm for energy efficient data transmission is that each sensor transmits data only when it can influence the decision of the main module. Vertical predicates form numerical ranges of values, within which the state of the network will be unchanged. Thus, if the current sensor reading is within this range, then the data may not be sent. In this case, the control module should understand that if a particular sensor does not transmit data, then its readings have not changed significantly and it is possible to use the previous value [5-8].

As a result of the application of this algorithm, the computational load is distributed: before the control module makes a decision about the state of the system, each sensor independently makes a decision whether its current state is important when the main module makes a decision.

The sensor decision-making algorithm for data transmission is shown in Figure 5.

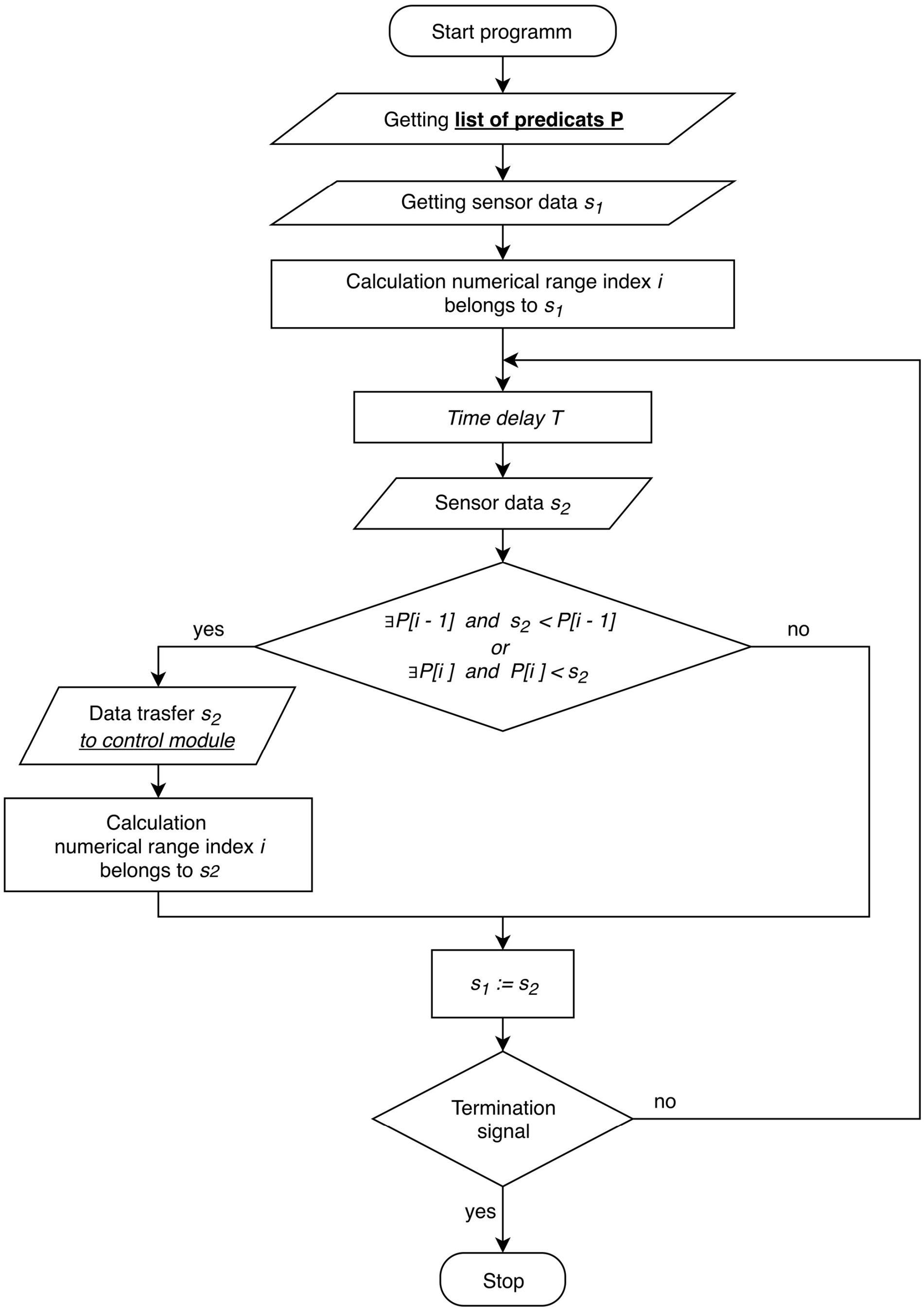

Figure 5. An algorithm of decision making by the sensor about data transmission.

The energy efficiency of the proposed algorithm is due to the reduction in the number of data transmission sessions. As previously noted, data transmission consumes most of the energy.

When sensor makes a decision about data transfer, information about the numerical ranges, within which the state of the system is unchanged, is required. To do this, the following steps must be performed:

– to train the model on previously known data;
– from a set of predicates, to select a subset relating to a particular sensor;
– to sort the predicates in ascending order (the predicates in the tree are arranged in a chaotic order);
– to send the obtained list of predicates to the sensor.

To study the energy efficiency of the proposed algorithm, an experiment was conducted using a network of pressure sensors [9] and an algorithm for classifying network states on the basis of data generated in Section 2. Based on the results of the study, it is possible to conclude that the application of the proposed algorithm allowed reducing the number of communication sessions by 80%. The second advantage of this method is that time delay T for each sensor can be determined separately depending on the dynamics of the monitored parameter (for example, the temperature tends to change faster than the pressure). The proposed solutions can be used in the implementation of energy efficient information transmission systems in the management of mobile systems [10–11], as part of the concept of “Internet of Things” with the support of predictive repair functions [12]. Thus, energy saving, as well as timely informing on significant changes in the environment, is provided.

4. Conclusion

The procedure for minimizing the set of monitored parameters allows saving energy resources due to the elimination of monitoring the parameters that do not affect the process of classification of states of the wireless sensor network. The second advantage of this technique is that in order to obtain the most accurate classification, it is possible to initially control all possible parameters of the external environment, and then to select only the minimum set of necessary parameters.

The proposed data transmission algorithm reduces the number of communication sessions by approximately 80%. This is achieved by the property of the decision tree, when the process is classified by predicate comparison. Thus, the sensors, based on the numerical ranges of a particular monitored parameter, can independently determine whether it is important to send data at the moment or not.

5. Acknowledgments

The work was performed with partial financial support under Agreement No. 2.4176.2017/PCh.

References

[1] Adamov A P, Adamova A A and Yuldashev M N 2016 Classification of states of a wireless sensor network using machine learning methods *Problems of Development of Promising Micro- and Nanoelectronic Systems (MES)* **2** 248–51

[2] Vlasov A I, Ivanov V V and Kosolapov I A 2011 Methods of anticipatory forecasting of the state of a broadband communication network *Software Products and Systems* **1** 3–6

[3] Vlasov A I and Yuldashev M N 2017 Gaussian processes in regression analysis of wireless sensor network states with allowance for electromagnetic interference *Electromagnetic Compatibility Technologies* **3(62)** 35–43

[4] Ivanov A M and Vlasov A I 2012 Verification of software models of communication networks *Mechanical Engineering and Computer Technologies* **10** 24

[5] Vasconcelos D R *et al*. 2017 Mathematical model for a collaborative indoor position system (IPS) and movement detection of devices within IoT environment *Proc. of the Symposium on Applied Computing* (ACM) pp 602–8

[6] Boser B E, Guyon I M and Vapnik V N 1992 A training algorithm for optimal margin classifiers *COLT '92 Proc. of the Fifth Annual Workshop on Computational Learning Theory* ed D Haussler (ACM Press) pp 144–52

[7] Witten I H and Frank E 2005 *Data Mining: Practical Machine Learning Tools and Techniques* (Second Edition) (Morgan Kaufmann)

[8] Mitchell T 2007 *Machine Learning* (McGraw-Hill Science/Engineering/Math)

[9] Andreev K A, Shakhnov V A and Vlasov A I 2016 Silicon pressure transmitters with overload protection *Automation and Remote Control* **77** 1281–85

[10] Yudin A, Kolesnikov M, Vlasov A and Salmina M 2017 Project oriented approach in educational robotics: from robotic competition to practical appliance *Advances in Intelligent Systems and Computing* **457** 83–94

[11] Yudin A V and Vlasov A I 2011 Distributed control system in mobile robot application: general

approach, realization and usage *Communications in Computer and Information Science* **156** 180–92

[12] Yudin A V, Salmina M A, Shakhnov V A, Usov K A and Vlasov A I 2017 Design methods of teaching the development of internet of things components with considering predictive maintenance on the basis of mechatronic devices *International Journal of Applied Engineering Research* **12** 9390–96