Fuzzy logic usage for the data processing in the Internet of Things networks

Larysa Globa Igor Sikorsky Kyiv Polytechnic Institute Igor Sikorsky Kyiv Polytechnic Institute Igor Sikorsky Kyiv Polytechnic Institute Kyiv, Ukraine lgloba@hotmail.com

Vasyl Kurdecha Kyiv, Ukraine kvv.vasyl@gmail.com

Ivan Ishchenko Kyiv, Ukraine ivanishchenkoo@gmail.com

Andrii Zakharchuk Igor Sikorsky Kyiv Polytechnic Institute Kyiv, Ukraine 96andriy@ukr.net

Nataliia Kunieva Igor Sikorsky Kyiv Polytechnic Institute Kyiv, Ukraine kuneva.nataly@gmail.com

Abstract-An overloaded Internet of Things (IoT) network needs data processing mechanisms to analyze the state of the system from the input parameters. There is not enough simple and statistical data analysis methods when operating with a large volume of data. The paper deals with the approach to big data analyses that use of fuzzy logic and fuzzy knowledge base for big data processing. The mathematical methods that allow to obtain efficiently the required solution in conditions of high performance requirements are considered. The proposed approach is quite complex on computational costs in training but it allows further to use not complex algorithms for big data processing. This is very important for Internet of Things systems.

Keywords-Internet of Things, fuzzy logic, fuzzy knowledge base, data processing, peer-to-peer connection.

I. INTRODUCTION

The Internet of Things (IoT) paradigm enables interconnected among devices-anytime, anywhere on the planetproviding the Internets advantages in all aspects of daily life. Analysts predict that the IoT will comprise up to 26 billion interconnected devices by 2020 [1]. Modern computing and distributed information systems (for example, cloud computing systems, networks function on the dynamic architectures of VANET and MANET, smart-type networks, intelligent controllers and sensors networks, etc.) are complex telecommunication networks that include many different types of network devices, which are integrated into the information and computing block, operate with high network and computational load. The conventional Internet has proved valuable in almost all trends by giving people the ability to interact with global information and services.

Extensive data has led to an explosive growth in the popularity of a modern data processing methods and the effective utilization of the bandwidth of the telecommunication network, because the amount of information has become much larger and it is by its nature and content becoming more diverse and large.

The feature of these systems is that they should control of the devices functioning or make decisions about the behavior of the system in different situations. Developing such kind of

the systems requires powerful computers usage. The feature of IoT systems is the impossibility of using complex calculations, since the client device is quite often a microcontroller that can not perform complex operations. Therefore, the paper deals with the approaches to processing a large number of data generated by the IoT system, and discuss possible approaches for managing IoT systems that can directly functioning on the client devices.

A. Maintaining the Integrity of the Specifications

Currently, when all mobile and most consumer digital devices have network interfaces that are used to exchange user data, branching and multicomponent of such networks, heterogeneity and virtually unlimited volume of nodes can determine the new phenomenon of network technologies -Internet of Things.

According to the Cisco IBSG Consulting Group the number of devices which are connected to the Internet by 2020 will increase to 50 billion [2]. For the implementation of the Internet of Things systems, we need to further develop specialized methods put forward for such systems. Goals that can be used for verification, and transmitted in accordance with the rules included in the system. The usual mathematical methods are too complicated for IoT systems, so there is a need for simpler algorithms that can provide the quality of the received technical solutions. In addition, it is important that during the adoption of IoT systems, it is often not possible to use Boolean logic, because sometimes it is not possible to describe the system parameter with two possible states.

The mathematical apparatus of fuzzy logic is used when choosing the best moment to increase or decrease the temperature of the environment, taking into account the external conditions. In addition, fuzzy logic can help reduce the energy consumption of urban buildings to achieve zero energy potential [3]. The basic idea of fuzzy logic is that the intellectual way of reasoning based on the natural language of human communication can't be described by traditional mathematical formulas. A formal approach is characterized

by a strict uniqueness of the interpretation, and all that is associated with the use of natural language, has a significant interpretation. Fuzzy logic is intended to formalize human abilities to inaccurate or approximate considerations that allow more adequately describe situations with uncertainty [4].

L. Zade introduces the concept of fuzzy set and, along with this, he also proposes to generalize classical logic with an infinite number of truth values. In the theory of fuzzy logic, the true meanings of statements can take any truth values from the actual interval numbers [0; 1]. This provision allows us to construct a logical system in which judgments can be made with uncertainty and to assess the degree of truth of statements. One of the concepts of fuzzy logic is the concept of elementary fuzzy statement. An elementary fuzzy statement is a narrative sentence, which expresses a complete opinion about which one can judge whether it is true or false with some definite degree of confidence.

II. BASIC CONCEPTS OF THE FUZZY SETS THEORY

A. Structure identification of the fuzzy system

Identification of fuzzy systems is based on observational data, but it is still not possible to exclude from this process the participation of an expert who solves the following tasks: determining the type of fuzzy system (Singleton, Mamdani or Takagi-Sugeno), choice of t-normal functions for fuzzy logic operations, choice of methods for fuzzy output (for a system like Mamdani) [5].

The choice of the system type is determined by the task that needs to be solved. If the problem of interpolation or approximation is solved, and accuracy is a determining factor, then the choice should be made in favor of the fuzzy system of the Takagi-Sugeno type. If the task of obtaining knowledge of data (in the form of linguistic rules) or the search for associative connections on a plurality of data is solved then for these purposes it is necessary to use a fuzzy system of the type Mamdani. The undoubted advantage of such models is their comprehensiveness and interpretation. The disadvantage is that in high computational costs. The type of the Singleton system can be used both for solving problems of approximation and obtaining knowledge [6].

Fuzzy model is defined as a system with n input variables defined on the input field of attributes, and one output variable Y defined on the original domain of DY. The exact value that the input variable X_i , accepts is denoted as x_i , and y is used for the output variable Y.

The fuzzy area of definition of the *i*-th input variable X_i is denoted by, where p_i – is the number of linguistic terms on which the input variable is defined, $LX_{i,k}$ defines the membership function and the name of the *k*-th linguistic term. Similarly – fuzzy field of definition of the output variable, q – number of fuzzy values, LY_i – membership function and the name of the original linguistic term.

The rule base in a fuzzy system like Mamdani is the set of fuzzy rules of the form:

$$R_j: LX_{1,j_1} \text{ AND } \dots \text{ AND } LX_{n,j_n} \to LY_j.$$
 (1)

The fuzzy j - th rule in the Singleton system has the form:

$$R_j: LX_{1,j_1} \text{ AND } \dots \text{ AND } LX_{n,j_n} \to r_j,$$
 (2)

where r_j is the number by which y is evaluated. Fuzzy j - th rule in the Takagi model – Sugeno has the form:

$$R_j: LX_1, j_1 \dots LX_n, j_n \rightarrow r_0 j + r_{1j} x_1 + \dots + r_{nj} x_n, \quad (3)$$

where the output y is evaluated by a linear function.

An important role is played by the membership function $\mu_{LXij}(x_i)$, which indicate the degree of belonging to a clear variable x_i fuzzy concept LX_{ij} . The training of the system is based on observation tables or test functions f(x). In general, the construction of a fuzzy system consists of the following main stages:

- Expert evaluation the type of fuzzy system and its associated parameters.
- Structure identification the choice of variables (X, DX, Y, DY, pi, q) and fuzzy rules R_j
- Parameters identification the search for optimal values of all parameters involved in the fuzzy system: searching of the values of the consequent (THAN-parts of the rule) and the parameters of membership functions in the antecedent (IF-part) of each rule (FX_i, FY) based on the given criteria of quality and optimization method selected criteria.
- Verifying the correctness of the model.

To make decisions in a fuzzy system, it is proposed to use the process of identifying a structure - the definition of the structural characteristics of a fuzzy system, such as the number of fuzzy rules, the number of linguistic terms, to which the incoming variables are divided. This identification is performed with the help of a fuzzy cluster analysis or a method of selection [7].

The usage of cluster analysis algorithms is aimed at dividing the set of data into clusters so that each of them has the closest objects. Fuzzy clustering is one of the most interesting methods for identifying possible groups and testing hypotheses about the data structure. Methods of fuzzy clustering allow one and the same object to belong simultaneously to several clusters, but with varying degrees of affiliation. Usually, each cluster is characterized by some prefix, which is described by the cluster center and some additional information, such as the size and form of the cluster. There are a number of clustering techniques. The main methods of fuzzy clustering: fuzzy decision trees, fuzzy Petri nets, fuzzy self-organizing maps [7].

Fuzzy clustering in many situations is more "natural" than clear, for example, for objects located on the border of clusters. The most common: the algorithm of fuzzy self-organization of c-means and its generalization in the form of the Gustafson-Kessel algorithm.

B. Algorithm of fuzzy self-organization c-means

Purpose: clustering large sets of numeric data.

Advantages: fuzzy when assigning objects to the cluster, allows you to identify objects that are on the border of the cluster.

Disadvantages: computational complexity, specifying the number of clusters at the input of the algorithm, there is uncertainty with objects that are significantly removed from the centers of all clusters.

In the classical algorithm of k-means (c-means), elements are chosen using the ordinary Euclidean distance between the vector x and the center of the cluster c. With this assignment, the distance between two vectors of a set of points equidistant from the center takes the form of a sphere with the same scale along all the axes. But if the data creates groups whose form differs from the spherical one or if the scales of the individual vector coordinates are very different, in this case the metric becomes inadequate. In this case, the quality of clustering can be significantly improved by using an improved version of the self-organization algorithm, which is called the Gustavsson-Kessel algorithm [8].

C. Fuzzy decision trees

Fuzzy decision trees are used in Data Mining to solve classification problems and to solve the regression problem when it is necessary to know the degree of belonging to a particular outcome. They can be used in various fields: in banking for solving the problem of scoring, in medicine for diagnosing various diseases, in the industry for quality control of products and so on.

Unconditional advantage of this approach is the high accuracy of classification achieved by combining the advantages of fuzzy logic and decision trees. The learning process takes place quickly, and the result is simple for interpretation. Since the algorithm is capable of issuing for the new object not only the class, but also the degree of belonging to it, it allows to control the classification threshold [9].

However, the complexity of applying the algorithm is the necessity of having a representative set of learning examples, otherwise the decision tree generated by the algorithm will slightly reflect the reality and, as a result, produce wrong results.

D. Fuzzy Petri nets

In the Petri time networks, conditions are represented by a set of positions, and their execution is represented by marking the corresponding position. Placing a certain number of labels in the given position after a specified time. In the basis of the study of the listed properties lies the feasibility analysis. Methods for analyzing the properties of Petri nets are based on using graphs of achievable (covering) markings, solving the equation of network states and calculating linear invariants of positions and transitions. Also, auxiliary reduction methods are used, which make it possible to reduce the size of the Petri net while preserving its properties, and the decomposition that divide the source network into subnets [10]. A feature of Petri nets is the possibility of presenting fuzzy processes and the dynamics of their interaction. The disadvantage is that many parameters, indicators and characteristics are not taken into account, without which it is difficult to imagine the real processes of practical implementations. In addition, the possibility of specifying the indistinctness of the marking and the components of the incidence function is limited, which significantly limits the possibilities of the researcher.

E. The result of the fuzzy algorithms

All of the considered algorithms have a common disadvantage of the algorithm complexity and this is very important as the IoT network has a rigid frame of time delays in processing information, obtaining inference, etc. Because of this problem, an approach is taken to form the basis of fuzzy knowledge base which allows using previously obtained knowledge about the behaviour of the system or process, which allows us to do the analysis of data based on certain rules [11].

With this approach, the algorithm for processing information for large data is as follows:

- The behavior of the system is determined based on the input set of data obtained in the form of rules or relationships that can be represented in the form of fuzzy knowledge base.
- According to the obtained model of the behavior of the system of fuzzy logic output conducts its state or development trends. This approach is most applicable for the IoT system because it allows you to reduce the number of computational operations for the system.

The formed fuzzy knowledge base, which displays the dependencies of the system with known input parameters, can be used to determine its system states. This is possible due to fuzzy logic output, which in turn can be implemented in different ways.

A fuzzy logical inference is the approximation of the dependence $y = f(x_1, x_2..., x_n)$ with the help of a fuzzy knowledge base and operations on fuzzy sets. In order to fulfill a fuzzy logical inference, the following conditions are required [12]:

- There should be at least one rule for each term of the output variable.
- For any term of an input variable there should be at least one rule in which this term is used as a prerequisite.
- There should be no contradictions and correlations between the rules

In "Fig. 1", shows the sequence of actions using the process of fuzzy logical inference

The process of fuzzy logical inference is a procedure or algorithm for obtaining fuzzy inference based on fuzzy conditions or prerequisites.

Systems of fuzzy logical inference can be considered as a separate case of so-called production fuzzy systems. In such systems, the conditions and logical inference of different rules are formed in the form of fuzzy statements made regarding the values of some linguistic variables. The development and

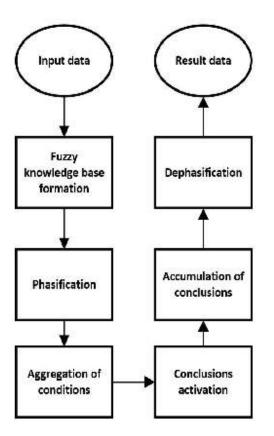


Figure 1. The sequence of actions using the process of fuzzy logical inference.

application of fuzzy logical inference systems consists of several stages, implementation of which is carried out with the help of the basic provisions of the theory of fuzzy sets.

Input of the fuzzy logic system are variables that carry information, which are obtained in some way, for example, by measuring some physical parameter of the system. These parameters are considered as real variables in the management processes [13]. The output of the system is a formed control variable of the fuzzy logic inference. Thus, fuzzy logic systems change the value of the input variables of the control process into output variables based on certain fuzzy rules. For IoT systems, the most effective is the application of fuzzy logic inference in the form of a set of rules such as fuzzy products, which are written in the form:

$$Rule < \# >: IF "\beta1 is \alpha1", THAN "\beta2 is \alpha2"$$
(4)

In the equation (4), the fuzzy statement " $\beta 1$ is $\alpha 1$ " is a condition of this rule. The false statement " $\beta 2$ is $\alpha 2$ " is the fuzzy inference of this rule. They are formulated in terms of fuzzy linguistic statements. It is assumed that " $\beta 1 \neq \beta 2$ ". The main stages of obtaining a fuzzy logical inference of the Mamdani system and the details of each stage are discussed in more detail below:

• Fuzzy logic base formation. The rules base of fuzzy logical systems is intended for a formal description of

empirical knowledge or experts knowledge in a problem area and is a set of fuzzy rules of the form:

$$Rule_{1}: IF "Condition_{1}", THAN "Conclusion_{1}"$$
$$Rule_{n}: IF "Condition_{n}", THAN "Conclusion_{n}"$$
(5)

where F_i (*i* belongs to 1, 2, ..., *n*) are the coefficients of certainty or the weighting coefficients of the corresponding rules. They can acquire values from the interval [0; 1]. Unless otherwise specified, $F_i = 1$. The rule base is given if it has set many rules for fuzzy products, as well as many input linguistic variables and the set of output linguistic variables.

- Phasification is a process or procedure for obtaining the values of the functions of fuzzy sets (terms) belonging to the basis of given input data. As a result of this phase completion, for all input variables, the specific values of the membership functions for each of the linguistic terms used in the set of conditions of the rules base of the fuzzy logical inference should be determined.
- Aggregation of conditions is the procedure that determines the conditions truth degree for each of the rules of the fuzzy logical inference system. When the rule condition has a simple form, its truth is equal to the corresponding value of the function of the input of the input variable to the term used in this condition.
- Activation is the process of finding the degree of truth of each of the underlying conditions of fuzzy rules. Before the start of this stage it is assumed that the known degree of truth and weight coefficient F_i for each rule. Next, each of the outputs of the rules of the fuzzy logic inference system is considered. When the inference of the rule is one vague statement, the degree of its truth is equivalent to the algebraic product of the corresponding degree of truth of the condition of the weight coefficient.
- Accumulation is the process of finding an accessory function for each of the output linguistic variables. The purpose of the accumulation phase is to combine all the degrees of truth of the outputs to obtain the function of belonging to each of the output variables. The conditions that relate to the same source linguistic variable belong to different rules of the system of fuzzy logical inference.
- Dephasification (clarification) is a procedure for determining the usual (strict) value for each of the output linguistic variables. The purpose of this phase is to obtain, using the results of the accumulation of all output linguistic variables, the usual quantitative value of each of the output variables. This value can be used by special devices that do not belong to the fuzzy logic system.

After analyzing the main stages of the fuzzy logical inference, we can conclude that the most important stage of the logical output is the phase of defazification, since it is the stage in which the numerical value of the resulting variable is determined. But there are several ways to carry out this phase. The use of one algorithm for various systems is at least not efficient, as it is unacceptable for some systems to use heavy mathematical operations, and the requirements for the accuracy of calculations are insignificant. This in turn allows you to use more simple algorithms. In particular, this is especially important if you implement a pre-processing of information on the client part of the Internet of Things system. Therefore, it is simply necessary to somehow choose the algorithm of data processing [13].

Developments in this direction have been under way for a long time, and already a number of systems and regulators that are capable of self-learning are built and the choice of algorithms for processing in such systems goes back to the background, as during the training of the system the parameters of its accessories functions change to achieve the maximum possible accuracy. Such a process is quite long, it is most efficient to run on servers, and if there are several fuzzy knowledge bases in the system, and for each one it is necessary to undergo training, even for a powerful computer, this will be a difficult task. And as we are talking about IoT systems, to such databases, besides, quite a lot will be completely different. In addition, such an approach is not entirely appropriate for its use in systems with pre-processing information on client devices of the IoT system.

Each of the algorithms can be used in the real system, the main difference between these algorithms is the accuracy of the calculation and the operating time, they are different in each algorithm. But this is not the only criterion for analysis, a more detailed analysis can be seen in the table:

 Table I

 CRITERIONS OF CHOOSING DEPHASIFICATION ALGORITHMS

Method	Complexity	Sensitivity	Precision
Medium the maximum	low	absent	middle
First of the maximum	middle	absent	middle
The last maximum	low	middle	below average
Centre of Gravity	high	significant	high
Centre of Area	very high	significant	high
Height middle	low	high	above average

An overloaded Internet of Things (IoT) network needs data processing to analyze the state of the system from the input parameters [14]. When working with large volumes of data, there is not enough simple and statistic data analysis. Needed a more sophisticated data analysis mechanism [15].

III. CONCLUSIONS

Thus, the analysis of the mathematical methods of data analysis that can be carried out in IoT systems allows us to draw the following conclusions:

• Different data analysis methods allow you to obtain different data, which in turn are close to real ones, which is evidence that all these algorithms can be used in different systems. For real systems it is necessary to carry out the analysis of algorithms for the formation of fuzzy knowledge base and application of fuzzy logical inference, as well as to conduct the training of the system during the knowledge base formation and during the selection of algorithms for fuzzy logical inference separately on real data sets).

- The task of choosing the correct fuzzy logical inference algorithm remains relevant.
- A possible direction for further development of this topic may be the search for selection criteria, and the construction of a system that, with the help of the developed criterion, will be able to choose the best configuration of the system of fuzzy databases.

REFERENCES

- R. Lea, M. Blackstock, Smart cities: An iot-centric approach, in: Proceedings of the 2014 International Workshop on Web Intelligence and Smart Sensing, IWWISS'14, ACM,NewYork, NY, USA, 2014, pp. 12–15.
- [2] M. Nikravesh, Evolution of fuzzy logic: From intelligent systems and computation to human mind, Stud. Fuzziness Soft Comput. 2007, pp. 37–53.
- [3] G. Cueva-Fernandez, J. Pascual Espada, V. García-Díaz, R. Gonzalez-Crespo, Fuzzy decision method to improve the information exchange in a vehicle sensor tracking system, Appl. Soft Comput. 2015, pp. 708–716.
- [4] Sugeno M., Yasukawa T. A fuzzy-logic-based approach to qualitative modeling // IEEE Trans. on Fuzzy Systems, 1993, vol. 1, pp. 7–31.
- [5] Emami M.R., Turksen I.B., Goldenberg A.A. A unified parameterized formulation of reasoning in fuzzy modeling and control // Fuzzy Sets and Systems, 1999, vol. 108, pp. 59–81.
- [6] Espinosa J., Vandewalle J., Wertz V. Fuzzy logic, identification and predictive control. — London: Springer-Verlag, 2005, 263 p.
- [7] Kurdecha V.V., Ishchenko I.O., Zakharchuk A.G. An approach to the Internet of Things system architecture // CADSM'2017 14th International conference the experience of design and application of CAD systems in microelectronics, 2017, pp. 1–3.
- [8] D.F. Larios, J. Barbancho, F.J. Molina, C. León, Lis: Localization based on an intelligent distributed fuzzy system applied to a wsn, Ad Hoc Networks, 2012, pp. 604–622.
- [9] Globa L.S., Ishchenko I.O., Zakharchuk A.G. Intelligent support system for e-Health // 20th International Multi-Conference on Advanced Computer Systems Miedzyzdroje, Poland, pp. 1–4, October 2016.
- [10] Fuzzy knowledge bases integration based on ontology Ternovoy, M., Shtogrina, O. 2012 Modern Problems of Radio Engineering, Telecommunications and Computer Science - Proceedings of the 11th International Conference, TCSET'2012, pp. 1–8.
- [11] Method of heterogeneous information resources structuring and systematizing for Internet portals development Globa, L., Novogrudska, R., Oriekhov, O. 2013 IEEE EuroCon, pp. 10–20.
- [12] Menard M. Fuzzy clustering and switching regression models using ambiguity and distance rejects / M. Menard // Fuzzy Sets and Systems, 2001, pp. 363–399.
- [13] Chuang C.C. TSK fuzzy modeling for function approximation with outliers / C. C. Chuang, S. F. Su, S. S. Chen // IEEE Trans. on Fuzzy Systems, 2001, pp. 810–821.
- [14] Andrzej Piegat, Fuzzy Modeling and Control //BINOMIAL. Knowledge lab, 2013, pp. 179–400.
- [15] Koucheryavy A. The QoS Estimation for Physiological Monitoring Service in the M2M network / A. Koucheryavy, Y. Al-Naggar // Internet of Things and its Enablers (INTHITEN): Proceedings. Conference, State University of Telecommunication (SUT), St. Petersburg, Russia, 2013, pp. 133-139.

ИСПОЛЬЗОВАНИЕ НЕЧЕТКОЙ ЛОГИКИ ДЛЯ ОБРАБОТКИ ДАННЫХ В ІоТ СЕТИ Глоба Л.С., Курдеча В.В., Ищенко И.А.,

Захарчук А.Г., Кунева Н.В.

Национальный технический университет Украины "Киевский политехнический институт имени Игоря Сикорского"

Современные ІоТ (Интернет вещей) системы продуцируют значительные объемы данных, называемые «большими» данными. Объемы «больших» данных возрастают экспоненциально, поэтому для ІоТ систем нужны эффективные механизмы обработки и анализа данных, быстрого определения состояния системы по входным параметрам.

Особенностью ІоТ систем является то, что они должны контролировать работу устройств или принимать решения о поведении системы в разных ситуациях. При создании программного обеспечения для ІоТ систем следует учитывать невозможность использования сложных вычислений, поскольку клиентское устройство довольно часто является микроконтроллером, который не может выполнять сложные операции. В связи с этим в статье рассматриваются подходы к обработке большого количества данных, генерируемых ІоТ системой, обсуждаются возможные подходы к созданию алгоритмов управления ее состоянием, которые могут непосредственно функционировать на клиентских устройствах или узлах с малой вычислительной производительностью.

Анализ математических методов обработки и анализа данных, которые может использовать IoT система, позволяет сделать такие выводы:

- различные методы анализа данных позволяют получать выводы, которые, в свою очередь, близки к реальным, что свидетельствует о том, что все множество алгоритмов можно использоваться в разных системах, однако временные и вычислительные затраты при этом будут значительными;
- не все алгоритмы и методы учитывают возможную нечеткость границ изменения состояния системы и могут формировать правила перехода между этими состояниями;
- анализ методов обработки «больших» данных показывает, что недостаточно простых и статистических методов анализа данных при работе с большим объемом данных. В статье предлагается подход к анализу «больших» данных, использующий методы нечеткой логики и нечеткую базу знаний для их последующей обработки. Рассмотрены математические методы, позволяющие эффективно получить требуемое решение в условиях высоких требований к производительности.

Однако, в каждом конкретном случае для создания реальных систем необходим анализ наиболее эффективных алгоритмов формирования нечеткой базы знаний и проведения нечеткого логического вывода, а также предварительное обучение системы во время формирования базы нечетких знаний и при выборе алгоритмов для проведения нечеткого логического вывода. Проведение обучения всей системы на реальных наборах данных возможно выполнить в автономном режиме. Таким образом, предлагаемый подход не смотря на то, что является довольно сложным по затратам на обучение на значительных реальных наборах данных, все же позволяет использовать не сложные алгоритмы для обработки «больших» данных, что очень важно для IoT систем. Возможным направлением дальнейшего развития этих исследований является поиск критериев выбора наилучшей конфигурации системы нечетких баз данных при ее формировании.