OUTLIERS DETECTION METHOD FOR DATA BASED ON MULTI-VALUED PREDICATE LOGIC

L.A. LYUTIKOVA[1,2], M.A. SHOGENOV[1]

[1] Institute of Applied Mathematics and Automation –
branch of the FSBSE "Federal Scientific Center
"Kabardin-Balkar Scientific Center of the Russian Academy of Sciences"
360000, KBR, Nalchik, Shortanov street, 89 "А"
E-mail: ipma@niipma.ru
[2] FSBSI " Federal scientific center
"Kabardin-Balkar scientific center of the Russian academy of sciences"
360002, KBR, Nalchik, 2 Balkarova st.
E-mail: lab.msb@mail.ru

The paper discusses a method for analyzing data that will be used in solving machine learning problems to find noise and inaccuracies, distortions in these data that impede the construction of an adequate model. Data of this kind is called outliers. The proposed approach uses methods and algorithms based on multi-valued logic systems. Multivalued logic can be used in the case of multidimensional heterogeneous features that characterize the objects of the original subject area. To conduct a qualitative data analysis, the following procedure is proposed in the work: a multi-valued logical function is constructed for the analyzed data, which finds all possible classes on the subject area under consideration; Further, the analysis of objects that did not fall into the constructed classes for a number of signs; and the hypothesis that these objects are emissions is tested. In the work, a hypothesis test is a sequence of logical rules for restoring the original dependencies presented in the training set. The proposed approach was considered for classification problems, in the case of multidimensional discrete features, where each feature can take k-different values and be equivalent in importance to class identification.

Keywords: object, class, knowledge base, emissions, informative weight

REFERENCES

1. Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka GrabskaBarwinska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, et al. Hybrid computing using a neural network with dynamic external memory. Nature, 2016. № 538(7626). Pp. 471-476.

2. Ashley I. Naimi, Laura B. Balzer Stacked generalization: an introduction to super learning // European Journal of Epidemiology (2018) 33:459-464.

3. Fan Yang Zhilin Yang William W. Cohen Differentiable Learning of Logical Rules for Knowledge Base Reasoning // Advances in Neural Information Processing Systems. Volume 2017-December, 2017, Pp. 2320-2329.

4. Peter Flach Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, 2012. 396 p.

5. Rahman Akhlaqur & Tasnim Sumaira. Ensemble Classifiers and Their Applications: A Review // International Journal of Computer Trends and Technology. (2014). Vol. 10. No 1. Pp. 31-35.

6. Serafini L., Garcez A.A. Logic tensor networks: Deep learning and logical reasoning from data and knowledge. Preprint arXiv:1606.04422, 2016 - arxiv.org

7. Shibzukhov Z.M. Correct Aggregation Operations with Algorithms // Pattern Recognition and Image Analysis. 2014, Vol. 24, No. 3. Pp. 377-382.

8. Shibzukhov, Z.M. On the principle of empirical risk minimization based on averaging aggregation functions. Doklady Mathematics, 2017. Volume 96, Issue 2. Pages 494-497.

9. Zhiting Hu, Xuezhe Ma, Zhengzhong Liu, Eduard Hovy, Eric Xing. Harnessing Deep Neural Networks with Logic Rules // Computer Science Learning. 2016. P. 2410-2420. arxiv:1603.06318.

10. Dyukova E.V., Zhuravlev Yu.I., Prokof`ev P.A. *Metody` povy`sheniya e`ffektivnosti logicheskix korrektorov* [Methods for increasing the efficiency of logical correctors] // *Mashinnoe obuchenie i analiz danny`x* [Machine Learning and Data Analysis]. 2015. T. 1. № 11. Pp. 1555-1583.

11. Zhuravlev Yu.I. *Ob algebraicheskom podxode k resheniyu zadach raspoznavaniya ili klassifikacii* [On the algebraic approach to solving recognition or classification problems] // *Problemy` kibernetiki* [Problems of Cybernetics].1978. T. 33. Pp. 5-68.

12. Lyutikova L.A., Shmatova E.V. *Analiz i sintez algoritmov raspoznavaniya obrazov s ispol`zovaniem peremenno-znachnoj logiki* [Analysis and synthesis of pattern recognition algorithms using variable-valued logic] // *Informacionny`e texnologii* ["Information Technologies"]. Tom 22. № 4. 2016. Pp. 292-297.

Lyutikova Larisa Adolfovna, candidate of physical and mathematical Sciences; Head of the Department of Neuroinformatics and machine learning; Institute of Applied Mathematics and Automation of Kabardin-Balkar Scientific Centre of RAS.

360000, KBR, Nalchik. Shortanov street, 89 A;

Senior staff scientist, Laboratory of Neurocognitive Autonomous Intelligent Systems Kabardin-Balkar Scientific Centre of RAS.

360000, KBR, Nalchik, Inessa Armand street 37-a.

Тел. 8(963)166-40-14.

E-mail: lylarisa@yandex.ru

Shogenov Mukhamed Aslanovich, postgraduate student of the Department of Neuroinformatics and machine learning; Institute of Applied Mathematics and Automation of Kabardin-Balkar Scientific Centre of RAS.

 360000, KBR, Nalchik, Shortanov street, 89 A.

Ph. 8-909-490-74-40

E-mail: brainmerge@gmail.com