

DOI: 10.47026/1810-1909-2023-2-76-84

УДК 621.316.91

ББК 31.247

А.Т. ГРИГОРЬЕВ, Н.А. КУЗНЕЦОВ, Е.В. СЛАВУТСКАЯ

**МЕТОД «ДЕРЕВО РЕШЕНИЙ»
ДЛЯ СТАТИСТИЧЕСКОГО КОНТРОЛЯ
ВЗАИМОСВЯЗЕЙ ПАРАМЕТРОВ
МНОГОМЕРНЫХ ИНФОРМАЦИОННЫХ ПОТОКОВ**

***Ключевые слова:** многомерные информационные потоки, машинное обучение, дерево решений, классификация, распознавание связей.*

***Цель исследования** – показать возможности методов машинного обучения для анализа внутрисистемных связей многомерных данных. В современных системах автоматизированного управления технологическими процессами, и, в частности, в электроэнергетике, необходим непрерывный мониторинг информационного обмена. Потоки данных несут случайный многомерный характер и передаваемые по каналам связи параметры имеют разные диапазоны варьирования и размерность. В этих условиях актуальным является развитие методов статистического контроля внутрисистемных связей таких данных.*

***Материалы и методы.** Для решения задачи в работе применен метод машинного обучения «дерево решений». Возможности подхода продемонстрированы при анализе взаимосвязей данных, моделирующих поток, содержащий 27 случайных параметров с разной размерностью. Проверка проводилась на выборке от 100 до 500 значений каждого из параметров.*

***Результаты.** Показано, что статистический контроль может осуществляться без рассмотрения структуры самого дерева решений по таким показателям, как процент распознавания связей, диапазоны разбиения значений параметров при классификации, значимость отдельных параметров (атрибутов).*

***Выводы.** Поскольку алгоритм не требует большой выборки значений анализируемых параметров, статистический контроль может осуществляться в скользящем временном окне. Показано, что подход может быть использован для анализа информационного обмена в системах автоматизированного управления технологическими процессами.*

Введение. Метод «дерево решений» является одним из методов машинного обучения [3, 6, 8] для задач интеллектуального анализа данных (Data Mining) [2, 5]. Он используется в основном для решения классификационных задач [11, 13, 18]. Метод находит всё более широкое применение в задачах интеллектуальной энергетики [12, 14, 17], в частности в задачах релейной защиты и автоматизации, определения места повреждения, классификации аварийных режимов [16, 20]. Построение «деревьев решений» позволяет визуализировать иерархическую классификацию данных в виде разветвлённой структуры, состоящей из узлов и листьев, аналогичной структуре папок и файлов в системе windows. Таким образом анализируется разветвлённая структура связей между изучаемыми параметрами (атрибутами) [3, 9]. Качество классификации определяется по проценту распознавания связей и зависит от многих факторов. Одним из достоинств метода «дерево решений» является то, что результат не зависит от размеров выборки и длительности обучения в такой степени, как, например, при обучении искусственных нейронных сетей [4, 7, 10].

В настоящей работе показано, что метод «дерево решений» может использоваться для анализа взаимосвязи параметров в многомерных данных при информационном обмене, в частности в электроэнергетике. Его применение может быть реализовано без построения собственно самого дерева решений. Для статистического контроля взаимосвязей параметров многомерных информационных потоков могут использоваться такие показатели, как значимость отдельных признаков и процент распознавания при классификации целевых атрибутов по этим признакам. Метод может использоваться в «скользящем временном окне» [21].

Метод «дерево решений» для классификации многомерных данных. В качестве примера работы алгоритма рассмотрим многомерный массив, содержащий 27 случайных признаков (атрибутов), по каждому из которых выборка составляла $N = 496$ и $N = 168$ значений (строк). Особенностью тестовых данных являлось то, что первые семь атрибутов, которые использовались в качестве целевых, имели вещественные значения, остальные 20 атрибутов принимали случайные целочисленные значения в узком диапазоне от 1 до 6. Соответствующие примеры приведены в табл. 1 и 2.

Таблица 1

Пример значений целевых атрибутов

Номер в выборке	Атрибуты						
	COL1	COL2	COL3	COL4	COL5	COL6	COL7
1	12,4	13,3	11,0	12,1	11,8	16,0	11,5
...
N	13,6	15,3	10,0	12,8	11,8	18,0	14,5

Таблица 2

Пример значений входных признаков

Номер в выборке	Атрибуты																										
	COL 8	COL 9	COL 10	COL 11	COL 12	COL 13	COL 14	COL 15	COL 16	COL 17	COL 18	COL 19	COL 20	COL 21	COL 22	COL 23	COL 24	COL 25	COL 25	COL 27							
1	4	4	4	3	4	2	3	2	1	1	2	2	2	2	1	1	1	1	1	1	1						
...						
N	6	1	5	2	1	3	2	3	2	3	3	2	1	3	1	1	1	1	1	1	1						

Поскольку данные, между которыми необходимо контролировать взаимосвязи (в статистическом смысле), имеют разную числовую размерность, проверка при помощи корреляционного анализа даёт результаты очень низкой статистической значимости. Это требует для оценки внутрисистемных связей данных использования методов машинного обучения, возможность применения которых не зависит от размерности анализируемых случайных признаков. При анализе могут быть использованы разные наборы атрибутов, что позволяет осуществлять моделирование взаимосвязи случайных неоднородных

данных разной размерности. Это качественно соответствует потребностям в анализе информационного обмена в цифровой электроэнергетике [15]. Расчеты проводились на отечественной аналитической платформе DEDUCTOR.

Результаты применения методики. Классификация и оценка взаимосвязей признаков проводились при целевых атрибутах в виде первых семи признаков COL1 – COL7, имеющих вещественные значения (см. табл. 1). В качестве входных атрибутов использовались 20 целочисленных признаков COL8 – COL27 (см. табл. 2). При анализе и построении дерева решений целевой атрибут разбивается на определённое количество классов, которое выбирается при постановке задачи. Границы классов (диапазонов) при этом в платформе DEDUCTOR, которая использовалась в работе, выбираются автоматически. Значения целевого атрибута разбивались на три-пять классов. Качество классификации и её точность могут быть оценены по таблицам сопряжённости, пример которых приведён на рис. 1.

Фактически	Классифицировано					Итого
	11,35 <=...< 13,65	13,65 <=...< 14,225	14,225 <=...< 14,5125	до 11,35	от 14,5125	
11,35 <=...< 13,65	304	2		4		310
13,65 <=...< 14,225	10	14		4		28
14,225 <=...< 14,5125	4		20	2		26
до 11,35	16	2		86		104
от 14,5125	10				18	28
Итого	344	18	20	96	18	496

а

Фактически	Классифицировано			Итого
	11,35 <=...< 12,2125	до 11,35	от 12,2125	
11,35 <=...< 12,2125	96	4	12	112
до 11,35	6	88	16	104
от 12,2125	3	6	286	290
Итого	110	90	286	496

б

Целевой атрибут: COL1Метка			
№	Номер	Атрибут	Значимость, %
1	14	COL21Метка	14,709
2	4	COL11Метка	10,634
3	12	COL19Метка	9,787
4	20	COL27Метка	9,717
5	3	COL10Метка	8,236
6	2	COL9Метка	7,849
7	18	COL25Метка	7,776
8	1	COL8Метка	5,219
9	9	COL16Метка	3,954
10	8	COL15Метка	3,178
11	7	COL14Метка	2,939
12	17	COL24Метка	2,742
13	10	COL17Метка	2,675
14	6	COL13Метка	2,557
15	11	COL18Метка	2,429
16	16	COL23Метка	1,529
17	19	COL26Метка	1,456
18	13	COL20Метка	1,269
19	15	COL22Метка	0,826
20	5	COL12Метка	0,521

в

Рис. 1. Таблицы сопряженности при 5 и 3 классах целевой функции (а, б) и гистограмма значимости входных атрибутов при классификации COL1 по 5 классам (в)

Числовые данные в табл. 1, 2 соответствуют количеству примеров из общей выборки входных атрибутов, для которых связи с целевой функцией в разных ее диапазонах распознаны правильно или неправильно. На диагонали таблиц обозначено количество правильно распознанных (классифицированных) примеров. Как видно из таблиц, для целевого атрибута COL1 при разбиении его значений на пять и три класса процент неправильно распознанных связей составляет

соответственно 10,9% и 9,3%. В данном случае процент неправильно распознанных связей составляет порядка 10% в обоих случаях при классификации по разному числу классов. Однако из таблицы для 5 классов (рис. 1, а) видно, что можно выделить диапазоны целевого атрибута в которых распознавание составляет 100%. Таким образом, таблица сопряжённости позволяет оценить качество классификации в разных диапазонах целевого атрибута и, соответственно, устойчивость связи между входными и выходными данными.

Более конкретную информацию о структуре связи можно получить из самого дерева решений, но при 20 входных признаках его структура оказывается очень сложной и с трудом поддается интерпретации. Поэтому важная информация может быть получена о структуре связей без анализа самого дерева решений. Такую информацию предоставляет гистограмма значимости входных атрибутов при классификации целевого признака. Пример такой гистограммы, соответствующий разбиению COL1 на 5 классов, приведён на рис 1, в.

Как следует из гистограммы рис. 1, в, 18 из 20 входных признаков имеют связи с COL1 со значимостью от 14,7% до 1,27%. Это дает возможность выделить признаки, которые связаны с целевым атрибутом наиболее устойчиво. То есть можно построить иерархию связей даже без анализа структуры самого дерева решений.

Для иллюстрации этого на рис. 2 приведены данные, полученные при классификации другой выборки из 168 значений тех же 27 признаков. В отличие от рис. 1 здесь при том же целевом атрибуте COL1 входными признаками являются все остальные 26. Из рис. 2 видно, что в этом случае аналогичные матрица сопряжения (уровень распознавания связей при классификации) и гистограмма значимости входных признаков существенно меняются. Можно обнаружить следующие особенности при сравнении таблиц сопряженности рис. 1, а–б и рис. 2, а–б:

1) процент нераспознанных при классификации связей сильно отличается (10,9% и 2,4% соответственно);

2) если в первом случае безошибочно классифицируются связи для значений целевого атрибута $COL1 > 14,225$, то во втором случае все неправильно распознанные связи соответствуют $11,0325 < COL1 < 13,8675$;

3) если при классификации по трем классам в первом случае (рис. 1, б) нераспознанные связи распределены относительно равномерно по всему диапазону варьирования целевого атрибута, то во втором случае (рис. 2, б) значения $COL1 < 11,64$ распознаются безошибочно;

4) процент неправильно распознанных связей в зависимости от количества классов и диапазонов разбиения целевого атрибута может значительно меняться и служить контрольным статистическим показателем.

При сравнении гистограмм рис. 1, в и рис. 2, в также может быть получена значительная дополнительная информация:

1) по гистограмме, представленной на рис. 2, из 20 целочисленных признаков COL–COL27 (см. рис. 1) значимыми остаются только три – COL8, COL10, COL12, – остальные значимые признаки – признаки из первой семёрки вещественных атрибутов;

Фактически	Классифицировано					Итого
	11,0325 <=...< 13,8675	13,8675 <=...< 14,07	14,07 <=...< 15,0825	до 11,0325	от 15,0825	
11,0325 <=...< 13,8675	112					112
13,8675 <=...< 14,07	2	12				14
14,07 <=...< 15,0825	2		8			10
до 11,0325				22		22
от 15,0825					10	10
Итого	116	12	8	22	10	168

а

Фактически	Классифицировано			Итого
	11,64 <=...< 13,8675	до 11,64	от 13,8675	
11,64 <=...< 13,8675	90		4	94
до 11,64	2	38		40
от 13,8675	8		26	34
Итого	100	38	30	168

б

Целевой атрибут: COL1 Метка			
№	Номер	Атрибут	Значимость, %
1	6	COL7 Метка	43,636
2	2	COL3 Метка	12,263
3	11	COL12 Метка	12,027
4	3	COL4 Метка	8,711
5	5	COL6 Метка	5,971
6	1	COL2 Метка	5,481
7	4	COL5 Метка	4,181
8	8	COL9 Метка	3,206
9	7	COL8 Метка	2,373
10	9	COL10 Метка	2,090
11	21	COL22 Метка	0,000
12	22	COL23 Метка	0,000
13	19	COL20 Метка	0,000
14	20	COL21 Метка	0,000
15	25	COL26 Метка	0,000
16	26	COL27 Метка	0,000
17	23	COL24 Метка	0,000
18	24	COL25 Метка	0,000
19	13	COL14 Метка	0,000
20	14	COL15 Метка	0,000
21	10	COL11 Метка	0,000
22	12	COL13 Метка	0,000
23	17	COL18 Метка	0,000
24	18	COL19 Метка	0,000
25	15	COL16 Метка	0,000
26	16	COL17 Метка	0,000

в

Рис. 2. Таблицы сопряженности при 5 и 3 классах целевой функции (а, б) и гистограмма значимости входных атрибутов при классификации COL1 по 5 классам (в). На входе – 26 атрибутов

2) выделяется очень высокая значимость признака COL7 при классификации COL1, что говорит о том, что эти атрибуты связаны между собой наиболее устойчиво;

3) характерно изменение значимости признака COL12. Если на рис. 2 он входит в тройку наиболее значимых с вкладом 12%, то на рис. 1 он вообще не дает вклада в классификацию по COL1 (в гистограмме рис. 1 он находится в числе последних двух с нулевым вкладом). Как было показано ранее [1], это соответствует опосредованной связи целочисленного параметра COL12 с COL1. То есть их связь обеспечивается через один из вещественных атрибутов COL2–COL6. И выделить этот «связующий» атрибут можно при его использовании в качестве целевой функции. В данном случае этим признаком является COL3, при классификации которого COL12 обеспечивает до 15% вклада.

Возможность оценки промежуточных (медианных) связей случайных данных является безусловным достоинством метода «дерево решений». Следует отметить, что значимость отдельных атрибутов и процент распознавания связей классификации в разных условиях могут являться характерными параметрами для статистического анализа информационных потоков. Резкое изменение этих

параметров будет соответствовать изменению внутрисистемных связей в информационном потоке, что может быть использовано, например, для поиска аномалий т.д. [15, 19].

Процент правильно распознанных связей при классификации данных по разным целевым атрибутам существенно меняется.

На рис. 3 показано качество результатов классификации по четырём из семи целевых атрибутов COL2–COL7 в зависимости от количества диапазонов, на которые значения этих атрибутов разбиваются. Как видно из рис. 3, если для COL4 и COL7 уровень нераспознанных связей остаётся приблизительно на 9–11%, то для COL2 и COL3 качество распознавания существенно меняется.

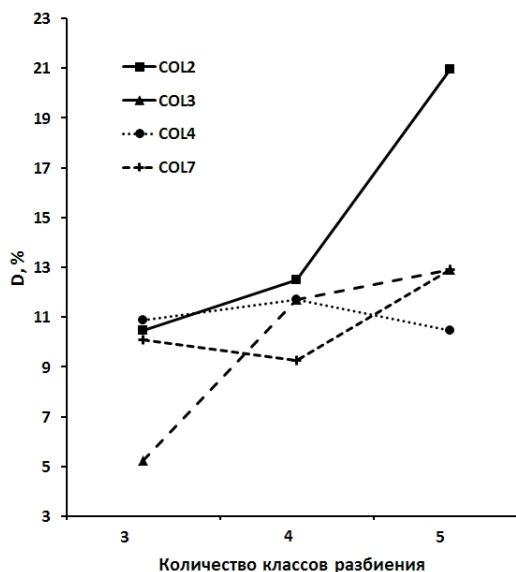


Рис. 3. Изменения процента нераспознанных связей для разных целевых атрибутов

Выводы. Метод «дерево решений», как один из методов машинного обучения и интеллектуального анализа данных, позволяет контролировать взаимосвязи между случайными параметрами в неоднородных многомерных потоках информации. В работе показано, что даже при разной размерности случайных данных оценки их внутренних взаимосвязей могут проводиться без построения собственно самого «дерева решений», по качеству классификации. Статистическими критериями для оценок могут служить значимость отдельных признаков, процент правильно классифицированных связей, границы диапазонов классификации. Поскольку метод не требует значительной выборки, он может применяться в скользящем временном окне для анализа многомерных информационных потоков с данными, которые имеют разные размерности, в частности – при анализе информационного обмена в цифровой электроэнергетике. Соответствующий подход может использоваться для мониторинга информационных потоков, обнаружение аномалий, для задач обеспечения информационной безопасности.

Литература

1. Вертикальный системный анализ данных психодиагностики учащихся с использованием метода «дерево решений» / *Е.В. Славутская, Л.А. Славутский, В.С. Абриков и др.* // Science for Education Today. 2020. № 3. С. 87–107. DOI: <http://dx.doi.org/10.15293/2658-6762.2003.05>.
2. Дюк В., Самойленко А. Data Mining: учебный курс. СПб.: Питер, 2001. 386 с.
3. Левитин А.В. Ограничения мощи алгоритмов: Деревья принятия решения // Алгоритмы. Введение в разработку и анализ (глава 10). М.: Вильямс, 2006. С. 409-417.
4. Славутская Е.В., Славутский Л.А. О выборе структуры искусственных нейросетей и алгоритмов анализа психодиагностических данных // Казанский педагогический журнал. 2020. № 5(142). С. 202–211.

5. Шумиков В.К., Мاستицкий С.Э. Классификация, регрессия и другие алгоритмы Data Mining с использованием R [Электронный ресурс]. URL: <https://analytics.github.io/data-mining>.
6. Adriaens, F., Lijffijt J., De Bie T. Subjectively interesting connecting trees and forests. *Data Min Knowl Disc*, 2019, vol. 33, pp. 1088–1124. DOI: <https://doi.org/10.1007/s10618-019-00627-1>.
7. Andreev O.N., Slavutskiy A.L., Slavutskii L.A. Neural network in a sliding window for power grids signals structural analysis. *IOP Conference Series: Earth and Environmental Science*, 2022, vol. 990. DOI: [10.1088/1755-1315/990/1/012054](https://doi.org/10.1088/1755-1315/990/1/012054).
8. Breiman L., Friedman J.H., Olshen R.A., Stone C.J. Classification and regression trees. Monterey CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984, 366 p.
9. Genrikhov I.E., Djukova E.V., Zhuravlev V.I. On full regression decision trees. *Pattern Recognit. Image Anal.*, 2017, vol. 27, pp. 1–7. DOI: <https://doi.org/10.1134/S1054661817010047>.
10. Grossberg S. A Path Toward Explainable AI and Autonomous Adaptive Intelligence: Deep Learning, Adaptive Resonance, and Models of Perception, Emotion, and Action. *Front. Neurorobot*, 2020, vol. 14. DOI: <https://doi.org/10.3389/fnbot.2020.00036>.
11. Holena M., Pulc P., Kopp M. Classification Methods for Internet Applications. Springer. 2020. DOI: [10.1007/978-3-030-36962-0](https://doi.org/10.1007/978-3-030-36962-0).
12. Ivanov S.O., Nikandrov M., Lariukhin A. Neuro Algorithm Accuracy Evaluation for The Anomalies Detecting in Overcurrent Protection Operation. In: International Ural Conference on Electrical Power Engineering (UralCon), 2021, pp. 116–120. DOI: [10.1109/UralCon52005.2021.9559614](https://doi.org/10.1109/UralCon52005.2021.9559614).
13. Kantardzic M. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011, 550 p.
14. Kulikov A.L., Loskutov A.A., Mitrovic M. Improvement of the technical excellence of multiparameter relay protection by combining the signals of the measuring fault detectors using artificial intelligence methods. International Scientific and Technical Conference Smart Energy Systems (SES-2019), 2019, vol. 124. DOI: [10.1051/e3sconf/201912401039](https://doi.org/10.1051/e3sconf/201912401039).
15. Laruhin A., Nikandrov M., Slavutskii L. Anomalous modes recognizing secondary equipment in electric power industry: adaptive neuro algorithms. In: Proceedings – 2019 International Ural Conference on Electrical Power Engineering (UralCon 2019), 2019, pp. 399–403.
16. Moloi K., Akumu A.O. Power distribution fault diagnostic method based on machine learning technique. In: 2019 IEEE PES/IAS PowerAfrica, Abuja, Nigeria, 2019, pp. 238–242. DOI: [10.1109/PowerAfrica.2019.8928633](https://doi.org/10.1109/PowerAfrica.2019.8928633).
17. Quintero-Zuluaga J.F. et al. Decision Tree-Based Automated Test-Bed for Performance Validation of Line Protection Relays Using a Hardware-in-the-Loop Architecture. In: 2020 IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020), Cali, Colombia, 2020, pp. 1–6. DOI: [10.1109/ColCACI50549.2020.9247877](https://doi.org/10.1109/ColCACI50549.2020.9247877).
18. Rumelhart D.E., Hinton G.E., Williams R.J. Learning Internal Representations by Error Propagation. *Parallel Distributed Processing*. Cambridge, MA: MIT Press, 1986, vol. 1, pp. 318–362.
19. Samantaray S.R., Kamwa I., Joos G. Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame. *IET Generation, Transmission & Distribution*, 2010, vol. 4(12), pp. 1334–1348. DOI: [10.1049/iet-gtd.2010.0201](https://doi.org/10.1049/iet-gtd.2010.0201).
20. Sheng Y., Rovnyak S.M. Decision tree-based methodology for high impedance fault detection. In: *IEEE Transactions on Power Delivery*, 2004, vol. 19, no. 2, pp. 533–536. DOI: [10.1109/TPWRD.2003.820418](https://doi.org/10.1109/TPWRD.2003.820418).
21. Slavutskiy A., Slavutskii L., Slavutskaya E. Neural Network for Real-Time Signal Processing: the Nonlinear Distortions Filtering. International Ural Conference on Electrical Power Engineering (UralCon), 2021, pp. 84–88. DOI: [10.1109/UralCon52005.2021.9559619](https://doi.org/10.1109/UralCon52005.2021.9559619).

ГРИГОРЬЕВ АЛЕКСАНДР ТАЛЪЯНОВИЧ – техник-программист, АО «Научно-производственный комплекс “Элара” имени Г.А.Ильенко», Россия, Чебоксары (sashagrigorev@vk.com).

КУЗНЕЦОВ НИКИТА АЛЕКСЕЕВИЧ – студент IV курса, факультет радиоэлектроники и автоматики, Чувашский государственный университет, Россия, Чебоксары (alca@mail.ru).

СЛАВУТСКАЯ ЕЛЕНА ВЛАДИМИРОВНА – доктор психологических наук, профессор кафедры психологии и социальной педагогики, Чувашский государственный педагогический университет имени И.Я. Яковлева, Россия, Чебоксары (slavutskayaev@gmail.com; ORCID: <https://orcid.org/0000-0002-3759-6288>).

Alexandr T. GRIGORIEV, Nikita A. KUZNETSOV, Elena V. SLAVUTSKAYA
THE «DECISION TREE» METHOD FOR STATISTICAL CONTROL
OF PARAMETERS INTERRELATIONS
IN MULTIDIMENSIONAL INFORMATION FLOWS

Key words: multidimensional information flows, machine learning, decision tree, classification, link recognition.

The purpose of the study is to show the possibilities of machine learning methods for analyzing intra-system connections of multidimensional data. In modern automated process control systems and in particular, in the electric power industry, continuous monitoring of information exchange is necessary. Data flows are random and the parameters transmitted via communication channels have different ranges of variation and dimension. In these conditions, particularly relevant is the development of statistical control methods of such data intra-system connections.

Methods. To solve the problem, the machine learning method "decision tree" is used. The possibilities of the approach are demonstrated by analyzing the data interconnections which model a stream containing 27 random parameters with different dimensions. The test was carried out on a sample of 100 to 500 values of each of the parameters.

Results. It is shown that statistical control can be carried out without considering the structure of the decision tree itself, according to such indicators as the percentage of links recognition, ranges of splitting of parameter values during classification, the significance of individual parameters (attributes).

Conclusions. Since the algorithm does not require a large sample of the analyzed parameters values, statistical control can be carried out in a sliding time window. It is shown that the approach can be used to analyze information exchange in the automated control system.

References

1. Slavutskaya E.V., Slavutskii L.A., Abrukov V.S. et al. *Vertikal'nyy sistemnyy analiz dannykh psikhodiagnostiki uchashchikhsya s ispol'zovaniyem metoda «derevo resheniy»* [Vertical system analysis of students' psycho diagnostic data using the 'Decision Tree' method]. *Science for Education Today*, 2020, vol. 10, no. 3, pp. 87–107. DOI: <http://dx.doi.org/10.15293/2658-6762.2003.05>.
2. Duke V., Samoylenko A. *Data Mining: uchebnyi kurs* [Data Mining: training course]. St. Petersburg, Peter Publ., 2001, 386 p.
3. Levitin A.V. *Ogranicheniya moshchi algoritmov: Derev'ya prinyatiya resheniya. Algoritmy. Vvedeniye v razrabotku i analiz* [Algorithm Power Constraints: Decision Trees, Algorithms. Introduction to Design and Analysis (Chapter 10)]. Moscow, Williams Publ., 2006, pp. 409–417.
4. Slavutskaya E., Slavutskiy L. *O vybore struktury iskusstvennykh neyrosety i algoritmov analiza psikhodiagnosticheskikh dannykh* [On choosing the artificial neural networks structure and the algorithms for psycho diagnostic data analyzing]. *Kazan pedagogical journal*, 2020, no. 5(142), pp. 202–210. DOI: 10.34772/KPJ.2020.142.5.026.
5. Shitikov V.K., Mastitsky S.E. *Klassifikatsiya, regressiya i drugiye algoritmy Data Mining s ispol'zovaniyem R*. [Classification, regression and other data mining algorithms using R. Available at: <https://ranalytics.github.io/data-mining>.
6. Adriaens, F., Lijffijt J., De Bie T. Subjectively interesting connecting trees and forests. *Data Min Knowl Disc*, 2019, vol. 33, pp. 1088–1124. DOI: <https://doi.org/10.1007/s10618-019-00627-1>.
7. Andreev O.N., Slavutskiy A.L., Slavutskii L.A. Neural network in a sliding window for power grids signals structural analysis. IOP Conference Series: Earth and Environmental Science, 2022, vol. 990. DOI: 10.1088/1755-1315/990/1/012054.
8. Breiman L., Friedman J.H., Olshen R.A., Stone C.J. *Classification and regression trees*. Monterey CA: Wadsworth & Brooks/Cole Advanced Books & Software, 1984, 366 p.
9. Genrikhov I.E., Djukova E.V., Zhuravlev V.I. On full regression decision trees. *Pattern Recognit. Image Anal.*, 2017, vol. 27, pp. 1–7. DOI: <https://doi.org/10.1134/S1054661817010047>.
10. Grossberg S. A Path Toward Explainable AI and Autonomous Adaptive Intelligence: Deep Learning, Adaptive Resonance, and Models of Perception, Emotion, and Action. *Front. Neurobot*, 2020, vol. 14. DOI: <https://doi.org/10.3389/fnbot.2020.00036>.

11. Holena M., Pulc P., Kopp M. Classification Methods for Internet Applications. Springer. 2020. DOI: 10.1007/978-3-030-36962-0.
12. Ivanov S.O., Nikandrov M., Lariukhin A. Neuro Algorithm Accuracy Evaluation for The Anomalies Detecting in Overcurrent Protection Operation. In: International Ural Conference on Electrical Power Engineering (UralCon), 2021, pp. 116–120. DOI: 10.1109/UralCon52005.2021.9559614.
13. Kantardzic M. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011, 550 p.
14. Kulikov A.L., Loskutov A.A., Mitrovic M. Improvement of the technical excellence of multi-parameter relay protection by combining the signals of the measuring fault detectors using artificial intelligence methods. International Scientific and Technical Conference Smart Energy Systems (SES-2019), 2019, vol. 124. DOI: 10.1051/e3sconf/201912401039.
15. Laruhin A., Nikandrov M., Slavutskii L. Anomalous modes recognizing secondary equipment in electric power industry: adaptive neuro algorithms. In: Proceedings – 2019 International Ural Conference on Electrical Power Engineering (UralCon 2019), 2019, pp. 399–403.
16. Moloi K., Akumu A.O. Power distribution fault diagnostic method based on machine learning technique. In: 2019 IEEE PES/IAS PowerAfrica, Abuja, Nigeria, 2019, pp. 238–242. DOI: 10.1109/PowerAfrica.2019.8928633.
17. Quintero-Zuluaga J.F. et al. Decision Tree-Based Automated Test-Bed for Performance Validation of Line Protection Relays Using a Hardware-in-the-Loop Architecture. In: 2020 IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020), Cali, Colombia, 2020, pp. 1–6. DOI: 10.1109/ColCACI50549.2020.9247877.
18. Rumelhart D.E., Hinton G.E., Williams R.J. Learning Internal Representations by Error Propagation. Parallel Distributed Processing. Cambridge, MA-MIT Press, 1986, vol. 1, pp. 318–362.
19. Samantaray S.R., Kamwa I., Joos G. Ensemble decision trees for phasor measurement unit-based wide-area security assessment in the operations time frame. IET Generation, Transmission & Distribution, 2010, vol. 4(12), pp. 1334–1348. DOI: 10.1049/iet-gtd.2010.0201.
20. Sheng Y., Rovnyak S.M. Decision tree-based methodology for high impedance fault detection. In: IEEE Transactions on Power Delivery, 2004, vol. 19, no. 2, pp. 533–536. DOI: 10.1109/TPWRD.2003.820418.
21. Slavutskiy A., Slavutskii L., Slavutskaya E. Neural Network for Real-Time Signal Processing: the Nonlinear Distortions Filtering. International Ural Conference on Electrical Power Engineering (UralCon), 2021, pp. 84–88. DOI: 10.1109/UralCon52005.2021.9559619.

ALEXANDER T. GRIGORIEV – Programmer Technician, The Ilyenko Elara Research and Production Complex (ELARA JSC), Russia, Cheboksary (sashagrigorev@vk.com).

NIKITA A. KUZNETSOV – 4th year Student, Faculty of Radio Electronics and Automation Chuvash State University, Russia, Cheboksary (alca@mail.ru).

ELENA V. SLAVUTSKAYA – Doctor of Psychological Sciences, Professor, Department of Psychology and Social Pedagogy, I.Ya. Yakovlev Chuvash State Pedagogical University, Russia, Cheboksary (slavutskayaev@gmail.com; ORCID: <https://orcid.org/0000-0002-3759-6288>).

Формат цитирования: Григорьев А.Т., Кузнецов Н.А., Славутская Е.В. Метод «дерево решений» для статистического контроля взаимосвязей параметров многомерных информационных потоков // Вестник Чувашского университета. – 2023. – № 2. – С. 76–84. DOI: 10.47026/1810-1909-2023-2-76-84.