

УДК 621.311.001.57

ББК 31.27-05

Л.А. СЛАВУТСКИЙ, Е.В. СЛАВУТСКАЯ

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ КАК АЛЬТЕРНАТИВА ФАКТОРИЗАЦИИ МНОГОМЕРНЫХ ДАННЫХ

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

Многомерные случайные данные и информационные потоки часто имеют разную или ограниченную числовую размерность. При анализе внутрисистемных связей таких данных корреляционный и факторный анализ неэффективны.

Цель исследования – оценить возможности комбинированного использования метода «дерево решений» и аппарата искусственных нейронных сетей для анализа многомерных случайных данных.

Материалы и методы. Методы машинного обучения применены для классификации многомерных случайных данных, имеющих разную числовую размерность и статистическое распределение. В качестве программного обеспечения использована аналитическая платформа «Deductor». Экспериментальный массив данных содержит 27 случайных параметров. Системный анализ проводился на выборке от 200 до 500 значений каждого из параметров.

Результаты исследования. Показано, что предлагаемый подход к системному анализу многомерных информационных потоков обладает рядом преимуществ по сравнению с традиционным корреляционным и факторным анализом. Он не накладывает ограничений на статистические распределения, позволяет работать с ограниченной выборкой данных, осуществлять выделение наиболее значимых параметров.

Выводы. Комбинированное использование методов машинного обучения позволяет существенно уменьшить обучающую выборку без потери точности вычислений. Для технических приложений это даёт возможность получать и анализировать информацию в динамике в реальном времени с использованием стандартного микропроцессорного оборудования. Результаты могут найти применение в задачах информационного обмена и кибербезопасности электроэнергетики.

Введение. Многомерные случайные данные могут рассматриваться как объект системного анализа, что актуально для широкого круга задач [3, 7, 16, 18]. Для системного анализа таких данных можно пользоваться как традиционными статистическими методами, так и методами машинного обучения [15, 19–21]. Если случайные данные и информационные потоки имеют разную числовую размерность, то традиционные методы, такие как корреляционный и факторный анализ, для установления и идентификации внутрисистемных связей могут оказаться неприемлемыми. Регрессионный анализ также имеет аналогичные ограничения [12, 14]. Использование этих традиционных инструментов для анализа связей данных затруднено, если параметры имеют разные диапазоны варьирования и размерность. Корреляционный анализ и факторный анализ, основанный на корреляционных коэффициентах, позволяют идентифицировать только монотонные связи и не дают информации об иерархии связей, о причинно-следственных закономерностях [27]. В этом случае для статистического контроля многомерных информационных потоков (в электроэнергетике, в частности [4, 6, 8]) могут использоваться методы машинного обучения [2, 22].

Цель настоящей работы – оценить возможность комбинированного использования метода «дерево решений» и нейронных сетей прямого распространения для анализа многомерных данных.

Материалы и методы. В настоящей работе предлагается комбинированное использование метода «дерево решений» и аппарата искусственных нейронных сетей (ИНС) в виде многослойного персептрона (MLP) для анализа внутрисистемных связей и уменьшения количества значимых случайных параметров в информационном потоке. Совместное использование этих двух методов связано со следующими особенностями. MLP позволяет решать как регрессионные так и аппроксимационные задачи, его использование даёт возможность установить связи между входными и выходными данными нейронной сети, но без изменения её структуры не позволяет выделить наиболее значимые случайные параметры [1, 28]. Метод «дерево решений» даёт возможность решать только классификационные задачи [9,11], однако позволяет оценить значимость отдельных входных атрибутов при установлении связи с целевой функцией [5]. Для многомерных данных в том случае, когда эти данные имеют очень разную числовую размерность, группировка параметров, выделение наиболее значимых из них и оценка внутрисистемных связей актуальны для широкого круга задач. К таким областям относятся задача информационного обмена для технических приложений, социология, психология и т.д.

Без ограничения общности предлагаемый подход проверялся при обработке многомерных психодиагностических данных. Использовались результаты тестирования 496 респондентов-родителей дошкольников при помощи двух тестов: PARI (7 признаков P1-P7 с вещественными значениями в диапазоне от 6.0 до 18.0 [25] и тест АСВ Э.Г. Эйдемиллера (20 целочисленных признаков E1–E20 в диапазоне от 0 до 6) [13].

На рис. 1 в качестве примера переведены графики варьирования показателей и статистические распределения результатов. Если в первом случае гистограмма распределение близка к нормальному распределению Гаусса, то во втором случае распределение несимметрично и отличается от нормального значительно. В этих условиях корреляционный анализ показал, что связи между данными разных тестов имеют уровни значимости не выше 0,1. Факторный анализ, основанный на коррекционных связях, в этих условиях также оказывается неэффективным, – из 27 исходных признаков по критерию Кайзера [17] выделяется не менее 10 факторов. При этом их суммарный вклад общую дисперсию не превышает 60%.

Для выделения наиболее значимых признаков в настоящей работе при классификации по разным целевым функциям сначала использовался метод «дерево решений». При этом выделялось несколько признаков по уровню значимости. А затем для проверки полученных результатов и оценки связи между входными данными и целевой функцией использовался персептрон с одним скрытым слоем (см. рис. 2).

Структура персептрона простейшая, позволяющая менять целевую функцию на его выходе, количество нейронов в скрытом слое и количество входных нейронов.

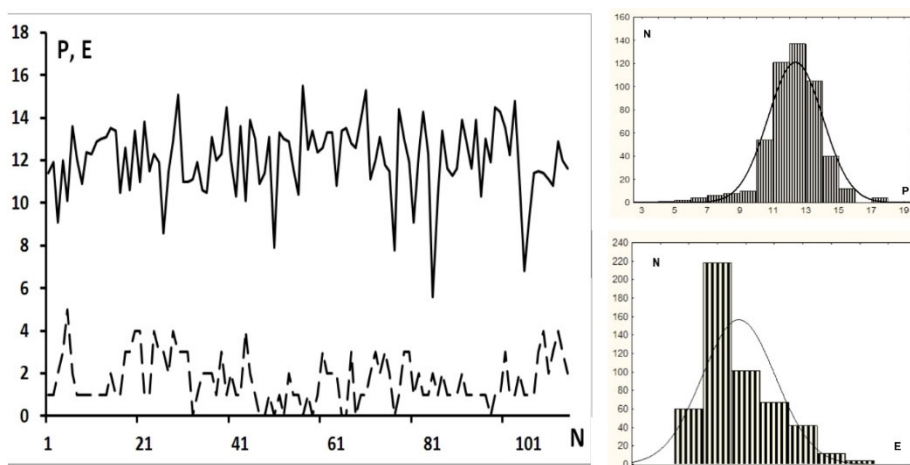


Рис. 1. Примеры записей данных с разной размерностью и статистическим распределением

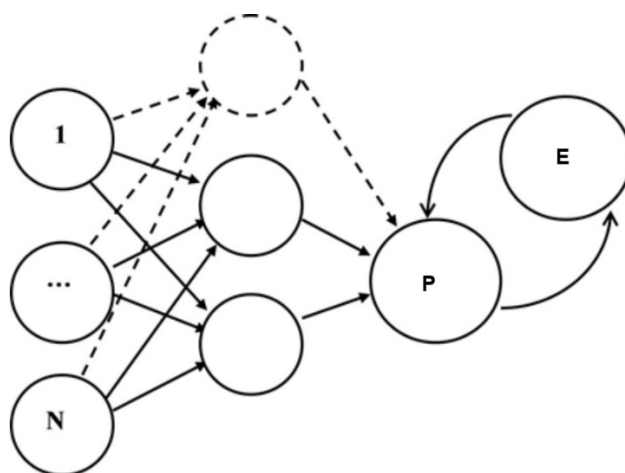


Рис. 2. Структура персептрона с заменой целевой функции

Кроме того, качество обучения ИНС при установлении связей между многомерными данными проверялось по следующей процедуре. Для каждого из 27 случайных параметров рассчитывалось среднее значение, а если значение данных выше или ниже среднего, то им присваивались величины 1 или 0. Это можно рассматривать как классификацию данных по двум классам. Сравнение полученных при помощи ИНС результатов с анализом исходных данных дает дополнительную информацию о внутрисистемных связях.

Результаты исследования. В рамках использованного авторами подхода метод «дерево решений» не требует построения структуры самого «дерева». Используется гистограмма значимости входных атрибутов при классификации по разным выходным данным. На рис. 3 приведён пример соответствующей гистограммы при классификации 26 признаков P2–E20 по целевому атрибуту P1.

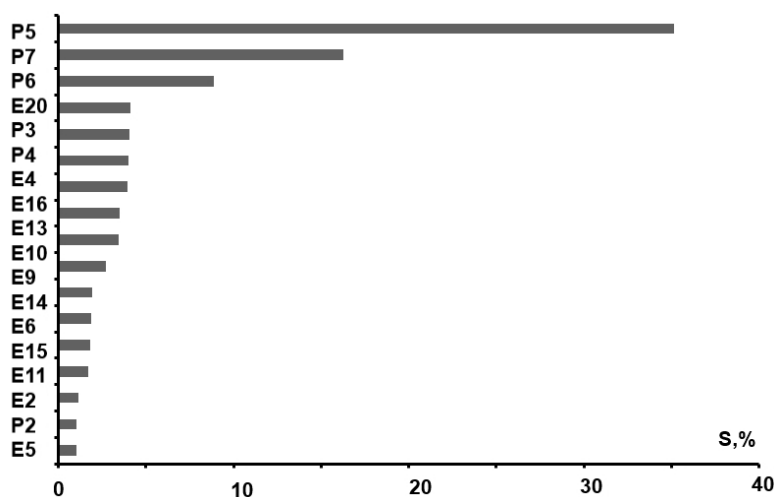


Рис. 3. Гистограмма значимости входных признаков при классификации по атрибуту P1

Как видно из рис. 3, в гистограмме можно выделить 7 входных параметров, значимость которых превышает 4 процента. При этом E20 и E4 имеют отличную от P1 размерность. Эти 7 параметров могут использоваться для дальнейшего анализа и классификации внутрисистемных связей при помощи ИНС. Сравнение соответствующих результатов (ошибок при обучении) перцептрона при 26 входных признаках и выделенных выше семи входных признаках показано на рис. 4.

Как видно из рис. 4, результаты обучения двух ИНС, отличающихся только количеством входных нейронов, очень сильно разнятся. Во втором случае максимальные и средние ошибки при обучении ИНС оказываются почти на порядок меньше. То есть связи между разнородными данными и целевой функцией значительно устойчивее.

Таким образом, предварительное выделение значимых признаков при помощи метода «дерево решений» даёт возможность значительно улучшить условия использования аппарата искусственных нейронных сетей на втором этапе анализа [23]. Кроме того, значительное увеличение точности вычислений ИНС подтверждает адекватность работы аппарата «дерево решений».

Таким образом, предлагаемый подход позволяет решать сразу две задачи с комбинированным использованием методов машинного обучения. С одной стороны, можно анализировать связи между случайными данными, имеющими разные цифровые шкалы, с другой – уменьшить количество значимых признаков. А слово «значимые» в данном случае означает иерархический анализ внутрисистемных связей многомерных данных. Кроме того, возможность исключить «незначимые» параметры означает, что их проявление соответствует возникновению аномалий в информационных потоках. В этом случае результаты представляют значительный интерес, в частности для задач информационного обмена и кибернетической безопасности электроэнергетики [8, 26].

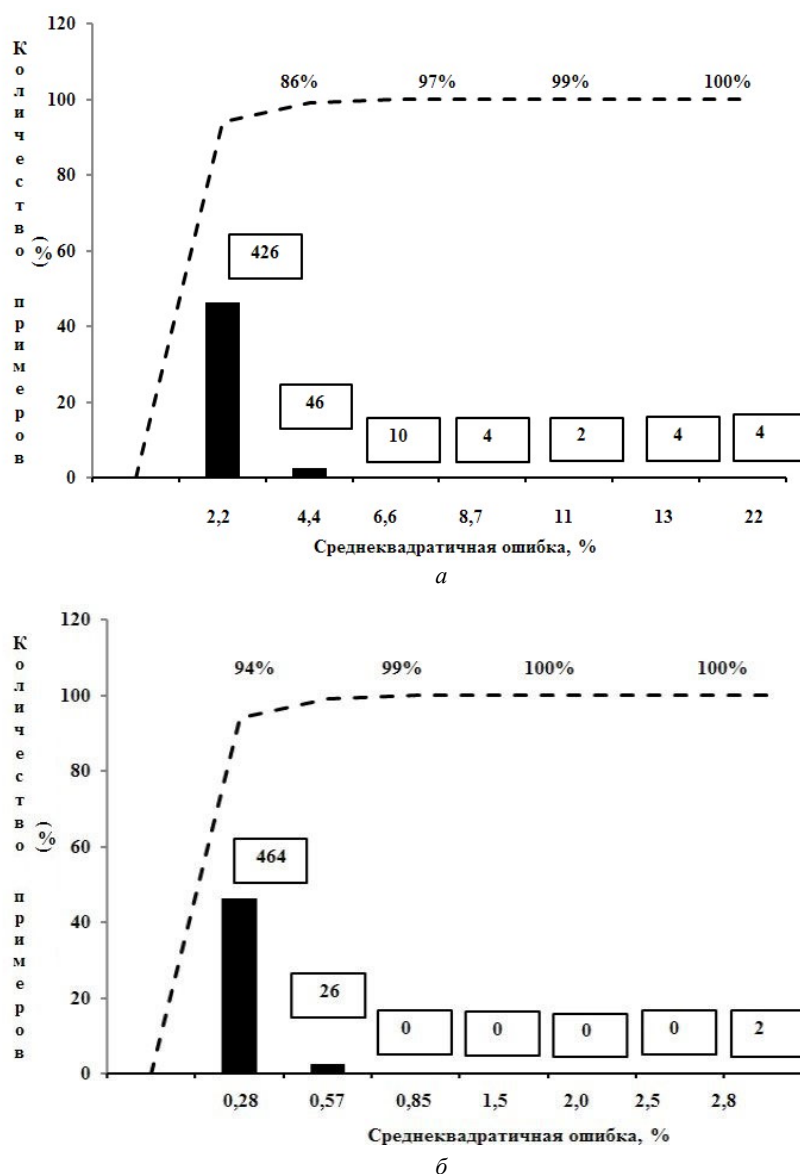


Рис. 4. Результат обучения ИНС при целевой функции P1 для 26 (а) и 7 (б) входных параметров

Уменьшение количества признаков на входе ИНС с увеличением качества ее обучения дает двойной эффект. С одной стороны, из рассмотрения исключаются данные, слабо влияющие на структуру внутрисистемных связей. С другой стороны, это позволяет уменьшить общее количество нейронов в ИНС и увеличить скорость вычислений. Это очень важно для обработки информационных потоков и сигналов в реальном времени с использованием микропроцессорной техники в скользящем временном окне [1, 10, 24].

Методы машинного обучения как альтернатива факторизации многомерных данных в информационном потоке имеют следующие преимущества:

– факторный анализ в скользящем временном окне осуществлять трудно в связи с тем, что число дискретных отчетов для факторного анализа должно как минимум в несколько раз превышать количество параметров в информационном потоке;

– факторный анализ требует интерпретации факторов. Если набор значимых параметров в факторах меняется, требуется дополнительная обработка, осуществить которую в реальном времени затруднительно;

– при факторном анализе группировка параметров производится по всем признакам, без обозначения целевого атрибута или целевой функции. Это ограничивает возможности интерпретации результатов.

Кратное уменьшение количества нейронов в MLP дает еще одно важное преимущество при анализе многомерных данных. Размер выборки при обучении ИНС должен значительно превышать количество весовых коэффициентов или связей между нейронами. Для метода «дерево решений» такого жесткого ограничения по размерам выборки нет. То есть предлагаемое комбинированное использование метода «дерево решений» и ИНС позволяет существенно уменьшить обучающую выборку без потери точности вычислений.

Выводы. Таким образом, методы машинного обучения могут эффективно использоваться как альтернатива многократной факторизации многомерных данных при использовании самых простых и достаточно хорошо изученных соответствующих инструментов. Показано, что подход применим при работе с ограниченной выборкой данных. Это, в свою очередь, даёт возможность получать и анализировать информацию в динамике, для технических приложений – в реальном времени, поскольку MLP, состоящий из нескольких десятков нейронов, легко использовать в стандартном микропроцессорном оборудовании.

Литература

1. Андреев О.Н., Славутский А.Л., Алексеев В.В. Структурный анализ электротехнических сигналов при рекуррентном использовании многослойного персептрона // Электротехника. 2022. № 8. С. 41–44. DOI 10.53891/00135860_2022_8_41.

2. Андреев О.Н., Ксенофонтов С.И., Славутский А.Л. Моделирование и нейросетевая обработка сигналов при переходных процессах в электротехнических комплексах. Чебоксары: Чуваш. гос. пед. ун-т, 2023. 212 с.

3. Афанасьев А.Ю., Макаров В.Г., Ханнанова В.Н. Идентификация параметров трехфазного асинхронного двигателя при изменении начальных значений оценок в широком диапазоне // Известия высших учебных заведений. Проблемы энергетики. 2015. № 11–12. С. 87–96.

4. Бульчев А.В., Охоткин Г.П., Силанов Д.Н. Цифровая система релейной защиты в распределительных электрических сетях // Электротехника. 2020. № 8. С. 31–35.

5. Вертикальный системный анализ данных психодиагностики учащихся с использованием метода «дерево решений» / Е.В. Славутская, Л.А. Славутский, В.С. Аbruков и др. // Science for Education Today. 2020. № 3. С. 87–107. DOI: <http://dx.doi.org/10.15293/2658-6762.2003.05>.

6. Воробьев Е.С., Антонов В.И., Наумов В.А. Функциональная совместимость устройств РЗА мультивендорных цифровых подстанций // Релейная защита и автоматизация. 2019. № 4(37). С. 42–45.

7. Дюк В., Самойленко А. Data Mining: учебный курс. СПб.: Питер, 2001. 386 с.

8. Коцеев М.И., Ларюхин А.А., Славутский А.Л. Использование адаптивных нейроалгоритмов для распознавания аномальных режимов систем вторичного оборудования электроэнергетики // Вестник Чувашского университета. 2019. № 1. С. 47-58.
9. Левитин А.В. Ограничения мощи алгоритмов: Деревья принятия решения // Алгоритмы. Введение в разработку и анализ (глава 10). М.: Вильямс, 2006. С. 409-417.
10. Лямец Ю.Я., Воронов П.И., Мартынов М.В., Маслов А.Н. Обучение релейной защиты на малом окне наблюдения // Электричество. 2017. № 3. С. 28-33.
11. Adriaens F., Lijffijt J., De Bie T. Subjectively interesting connecting trees and forests. *Data Min Knowl Disc*, 2019, vol. 33, pp. 1088-1124. DOI: 10.1007/s10618-019-00627-1.
12. 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.
13. Eidemiller E.G., Yustitsky V.V. Family psychotherapy: The basic principles and practical experience. *International Journal of Family Psychiatry*, 1989, vol. 10(3-4), pp. 325-337.
14. 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: 10.1134/S1054661817010047.
15. Grossberg S. Toward Autonomous Adaptive Intelligence: Building Upon Neural Models of How Brains Make Minds. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, vol. 51, no. 1, pp. 51-75. DOI: 10.1109/TSMC.2020.3041476.
16. Holena M., Pulc P., Kopp M. Classification Methods for Internet Applications. Springer. 2020. DOI: 10.1007/978-3-030-36962-0.
17. Kaiser H.F. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 1960, vol. 20, pp. 141-151.
18. Kantardzic M. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011, 550 p.
19. Kulikov A., Loskutov A., Bezdushniy D., Petrov I. Decision Tree Models and Machine Learning Algorithms in the Fault Recognition on Power Lines with Branches. *Energies*, 2023, vol. 16, p. 5563.
20. 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. In: International Scientific and Technical Conference Smart Energy Systems (SES-2019), 2019, vol. 124. DOI 10.1051/e3sconf/201912401039.
21. Leonowicz Z., Jasinski M. Machine Learning and Data Mining Applications in Power Systems. *Energies*, 2022, vol. 15, p. 1676. DOI: 10.3390/en15051676.
22. Quintero-Zuluaga J.F., Viana-Villa J.P., Villegas D. Decision Tree-Based Automated Test-Bed for Performance Validation of Line Protection Relays Using a Hardware-in-the-Loop Architecture. In: IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020), Cali, Colombia, 2020, pp. 1-6. DOI: 10.1109/ColCACI50549.2020.9247877.
23. 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.
24. 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.
25. Schaefer E.S., Bell R.Q. Development of a parental attitude research instrument. *Child Develop.*, 1958, vol. 29, pp. 339-361.
26. Singh V.K., Govindarasu M.A. Cyber-Physical Anomaly Detection for Wide-Area Protection Using Machine Learning. *IEEE Transactions on Smart Grid*, 2021, vol. 12, no. 4, pp. 3514-3526. DOI: 10.1109/TSG.2021.3066316.
27. Slavutskaya E.V., Slavutskii L.A., Nikolaev E.L. Neural Network Models for the Analysis and Visualization of Latent Dependencies: Examples of Psycho Diagnostic Data Processing. Knowledge in the Information Society: Joint Conferences XII Communicative Strategies of the Information Society (CSIS2020) and XX Professional Culture of the Specialist of the Future (PCSF2020). Cham, Springer Verlag, 2021, pp. 61-70. DOI 10.1007/978-3-030-65857-1_7.
28. Slavutskaya E., Slavutskii L., Zakharova A. Integrated Use of Data Mining Techniques for Personality Structure Analysis. Technology, Innovation and Creativity in Digital Society. Springer Nature Switzerland, 2022, vol. 345, pp. 522-533. DOI 10.1007/978-3-030-89708-6_44.

СЛАВУТСКИЙ ЛЕОНИД АНАТОЛЬЕВИЧ – доктор физико-математических наук, профессор кафедры автоматизации и управления в технических системах, Чувашский государственный университет, Россия, Чебоксары (lenya@slavutskii.ru; ORCID: <https://orcid.org/0000-0001-6783-2985>).

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

Leonid A. SLAVUTSKII, Elena V. SLAVUTSKAYA
MACHINE LEARNING METHODS AS AN ALTERNATIVE
TO FACTORIZATION OF MULTIDIMENSIONAL DATA

Key words: multidimensional information flows, machine learning, decision tree, classification, links recognition, neural networks.

Multidimensional random data and information flows often have different or limited numerical dimensions. When analyzing the intra-system relationships of such data, correlation and factor analysis are ineffective.

The purpose of the study is to evaluate the possibilities of the combined use of the "decision tree" method and the artificial neural networks for the analysis of multidimensional random data.

Materials and methods. Machine learning methods are used to classify multidimensional random data with different numerical dimensions and statistical distribution. The analytical platform "Deductor" is used as the software. The experimental data set contains 27 random parameters. The system analysis was carried out on a sample of 200 to 500 values of each parameter.

Results. It is shown that the proposed approach to the system analysis of multidimensional information flows has a number of advantages over traditional correlation and factor analysis. It does not impose restrictions on statistical distributions, allows one to work with a limited data sample, and select the most significant parameters.

Conclusions. The combined use of machine learning methods allows one to significantly reduce the training sample without losing the calculations accuracy. For technical applications, this makes it possible to receive and analyze information dynamically, in real time using standard microprocessor equipment. The results can be applied in the tasks of information exchange and cybersecurity of the electric power industry.

References

1. Andreev O.N., Slavutskiy A. L., Alekseev V.V. *Strukturnyy analiz elektricheskikh signalov s periodicheskim ispol'zovaniyem mnogosloynnogo perseptrona* [Structural Analysis of Electrical Signals with Recurrent Use of a Multilayer Perceptron]. *Russian Electrical Engineering*, 2022, vol. 93, no. 8, pp. 529–532. DOI 10.3103/S1068371222080028.
2. Andreev O.N., Slavutskiy A.L., Ksenofontov S.I. *Modelirovaniye i neyrosetevaya obrabotka signalov perekhodnykh protsessov v elektrotekhnicheskikh kompleksakh* [Modeling and neural network signal processing transients processes in electrical engineering complexes]. Cheboksary, 2023, 212 p.
3. Afanasiev A.Yu., Makarov V.G., Khannanova V.N. *Identifikatsiya parametrov trekhfaznogo asinkhronnogo dvigatelya pri izmenenii nachal'nykh znacheniy otsenok v shirokom diapazone* [Identification of parameters of three-phase asynchronous motor when changing the initial values of the estimates in a wide range]. *Power engineering: research, equipment, technology*, 2015. no. 11-12, pp. 87–96.
4. Bulychev A.V., Okhotkin G.P., Vasiliev S.A. *Tsifrovaya sistema releynoy zashchity v elektricheskikh raspredelitel'nykh setyakh* [A Digital Relay Protection System in Electrical Distribution Networks]. *Russian Electrical Engineering*, 2020, vol. 91, no. 8, pp. 495–499. DOI 10.3103/S1068371220080064.

5. Slavutskaya E.V., Slavutskii L.A., Abrukov V.S. *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: 10.15293/2658-6762.2003.05.
6. Vorobyev E.S., Antonov V.I., Naumov V.A. *Funktsional'naya sovmestimost' ustroystv RZA mul'tivendornykh tsifrovyykh podstantsiy* [Interoperability of relay protection and automation devices across multivendor digital substations]. *Relay protection and automation*, 2019, no. 4(37), pp. 42–45.
7. Duke V., Samoylenko A. *Data Mining: uchebnyi kurs* [Data Mining: training course]. St. Petersburg, Piter Publ., 2001, 386 p.
8. Koshcheev M.I., Laryukhin A.A., Slavutskiy A.L. *Ispol'zovanie adaptivnykh neuroalgoritmov dlya raspoznavaniya anomal'nykh rezhimov sistem vtorichnogo oborudovaniya elektroenergetiki* [Using adaptive neuroalgorithms to recognize anomalous modes of secondary equipment systems in the electrical power industry]. *Vestnik Chuvashskogo universiteta*, 2019, no. 1, pp. 47–58.
9. 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.
10. Liamets Y.YA., Voronov P.I., Martynov M.V., Maslov A.N. *Obucheniye releynoy zashchity na malom okne nablyudeniya* [Training of relay protection with small observation window]. *Elektrichestvo*, 2017, no. 3, pp. 28–33.
11. Adriaens F., Lijffijt J., De Bie T. Subjectively interesting connecting trees and forests. *Data Min Knowl Disc*, 2019, vol. 33, pp. 1088–1124. DOI: 10.1007/s10618-019-00627-1.
12. 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.
13. Eidemiller E.G., Yustitsky V.V. Family psychotherapy: The basic principles and practical experience. *International Journal of Family Psychiatry*, 1989, vol. 10(3-4), pp. 325–337.
14. 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: 10.1134/S1054661817010047.
15. Grossberg S. Toward Autonomous Adaptive Intelligence: Building Upon Neural Models of How Brains Make Minds. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, vol. 51, no. 1, pp. 51–75. DOI: 10.1109/TSMC.2020.3041476.
16. Holena M., Pulc P., Kopp M. *Classification Methods for Internet Applications*. Springer, 2020. DOI: 10.1007/978-3-030-36962-0.
17. Kaiser H.F. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 1960, vol. 20, pp. 141–151.
18. Kantardzic M. *Data mining: concepts, models, methods, and algorithms*. John Wiley & Sons, 2011, 550 p.
19. Kulikov A., Loskutov A., Bezdushniy D., Petrov I. Decision Tree Models and Machine Learning Algorithms in the Fault Recognition on Power Lines with Branches. *Energies*, 2023, vol. 16, p. 5563.
20. 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. In: International Scientific and Technical Conference Smart Energy Systems (SES-2019), 2019, vol. 124. DOI 10.1051/e3sconf/201912401039.
21. Leonowicz Z., Jasinski M. Machine Learning and Data Mining Applications in Power Systems. *Energies*, 2022, vol. 15, p. 1676. DOI: 10.3390/en15051676.
22. Quintero-Zuluaga J.F., Viana-Villa J.P., Villegas D. Decision Tree-Based Automated Test-Bed for Performance Validation of Line Protection Relays Using a Hardware-in-the-Loop Architecture. In: IEEE Colombian Conference on Applications of Computational Intelligence (IEEE ColCACI 2020), Cali, Colombia, 2020, pp. 1–6. DOI: 10.1109/ColCACI50549.2020.9247877.
23. 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.
24. 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.

25. Schaefer E.S., Bell R.Q. Development of a parental attitude research instrument. *Child Develop.*, 1958, vol. 29, pp. 339–361.

26. Singh V.K., Govindarasu M.A. Cyber-Physical Anomaly Detection for Wide-Area Protection Using Machine Learning. *IEEE Transactions on Smart Grid*, 2021, vol. 12, no. 4, pp. 3514–3526. DOI: 10.1109/TSG.2021.3066316.

27. Slavutskaya E.V., Slavutskii L.A., Nikolaev E.L. Neural Network Models for the Analysis and Visualization of Latent Dependencies: Examples of Psycho Diagnostic Data Processing. Knowledge in the Information Society: Joint Conferences XII Communicative Strategies of the Information Society (CSIS2020) and XX Professional Culture of the Specialist of the Future (PCSF2020). Cham, Springer Verlag, 2021, pp. 61–70. DOI 10.1007/978-3-030-65857-1_7.

28. Slavutskaya E., Slavutskii L., Zakharova A. Integrated Use of Data Mining Techniques for Personality Structure Analysis. Technology, Innovation and Creativity in Digital Society. Springer Nature Switzerland, 2022, vol. 345, pp. 522–533. DOI 10.1007/978-3-030-89708-6_44.

LEONID A. SLAVUTSKII – Doctor of Physical and Mathematical Sciences, Professor, Department of Automation and Control in Technical Systems, Chuvash State University, Russia, Cheboksary (lenya@slavutskii.ru; ORCID: <https://orcid.org/0000-0001-6783-2985>).

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: *Славутский Л.А., Славутская Е.В.* Методы машинного обучения как альтернатива факторизации многомерных данных // Вестник Чувашского университета. – 2024. – № 2. – С. 141–150. DOI: 10.47026/1810-1909-2024-2-141-150.