

# 1. Обзор литературы

Метод обзора: открыл google scholar и там взял самые цитируемые статьи, ссылающиеся на две оригинальные [25] [26] . Сейчас граница выставлена на 50 цитирований. Это 4 книги и около 150 статей. Статистика примерно следующая:

- 25 источников предлагают какие-либо усовершенствования,
- 103 источника либо используют оригинальный isoforest в своём пайплайне, либо сравнивают свои алгоритмы с ним, либо просто упоминают о его существовании,
- 19 источников осуществляют обзоры разных методов и подходов, упоминая или описывая в том числе изофорест,
- 7 источников не упоминают изофорест вообще,
- 2 источника найти не удалось.

## Обсуждаемые усовершенствования изоляционного леса

Здесь обсудим, какие усовершенствования предлагают различные авторы в доступной литературе.

### Статический лес

В статье [36] авторы предлагают проводить выборку пороговых значений для фич неравномерно от минимума до максимума, а по некоторому построенному на основе данных распределению. Распределение строится на основе линейной комбинации ядер, где ядра в центре имеют меньшую плотность вероятности, чем на краях. Таким образом вероятность выбрать из пороговое значение из плотного региона будет меньше, а разделение – лучше.

Статьи [45] [44] посвящены ”глубокому лесу”. Идея повторяет глубокие нейронные сети. Несколько раз каскадом повторяем одну и ту же операцию – строим лес, считаем скоры, приписываем скоры в дополнение к исходным фичам. Как именно происходит обучение я не догнал, но в целом подход хороший. Гиперпараметров значительно меньше, чем у нейронов.

Ещё пара статей по дальнейшей проработке глубокого леса: [37] .

Тут [34] в целом ничего интересного, кроме мысли о том, что изофорест можно применять в т.ч. и к категориальным данным. При этом сами они применяют его очень странно.

В статье [38] коллеги зачем-то выкидывают деревья из изофореста, но зато доказывают интересное наблюдение: средняя длина изолирующего пути не зависит от размерности для равномерно распределённых данных. Заодно выводят по-человечески ту самую формулу для средней длины пути.

Авторы [24], [42], [9] [10] применяют по сути изоляционный лес к географическим координатам машин, чтобы понять, когда машина уехала куда-то вообще не туда. У их подхода есть интересная особенность. Они не строят деревьев на траекториях, а прямо задают вопрос ”сколько делений выдержит конкретна эта траектория, если мы будем её изолировать от вот этого сабсемпла траекторий”.

Поступают они совершенно аналогично [38] , что намекает, что эта мысль у людей бродит. И правильно бродит – не всегда есть возможность нарезать данные в какой-то

очевидной топологии. Например, если у нас данные состоят сплошь из категориальных данных. Как делить тогда? Или если данные разреженные – тот же вопрос. Не всегда вообще в данных присутствуют все измерения. Такой подход в принципе может упростить работу с инженерией фич.

В статьях [6] [7] авторы создают гибрид между LOF и изофорестом. Выглядит любопытно. Но это уже совсем не деревья получаются. Никакой вероятностной интерпретацией и не пахнет. Да и во много опирается на понятие расстояния, от которого изофорест удачно дистанцировался.

Интересная попытка обогатить изофорест – это extended isolation forest [20]. Вместо выбора случайной фичи выбирается случайное направление. Это хорошая заявка на успех, но почему-то по тестам в той же статье получается не суперархизикарно.

Считаю, что причина в том, что идея не доведена до логического завершения. Посмотрел код – у них вместо порогового значения скалярного произведения зачем-то выбирается случайная точка из гиперкуба, ограничивающего данные. Явно есть возможность нахалевку доработать и получить результат наверняка более впечатляющий. Либо узнать нечто новое про изоляционный лес.

Чёрт, это уже сделано в [23]. Буквально год назад.

Статья [41] делает что-то относительно интересное и издалека похожее на изоляционный лес. Но только издалека. Они предлагают двухэтапный подход к поиску аномалий, где в качестве первого этапа используют построение дерева с минимизацией межгрупповой дисперсии. Этим деревом отсеивают наиболее плотные участки данных. Вторым этапом считают какие-то своеобразные метрики, основанные на расстояниях. Частично интересно, но расстояния всё обламывают.

Вот тут [5] интересное предложение состоит в том, что можно считать скор не как длину пути от корня дерева к листу с аномалией, а как долю семплов, которую отделяет финальная ветвь. Получается аналог LOF, т.к. задаёт именно локальный скор семпла. По утверждению авторов получается не хуже самого изофореста. Стоит взять на вооружение.

Вот здесь [11] какое-то не очень очевидное применение изофореста для задачи поиска полезных ископаемых. Упоминаю в этом разделе, т.к. такое применение может наметнуть на неочевидную интерпретацию метода.

Статья [43] предлагает интересный наркоманистый способ использования изофореста для поиска релевантных данных. Точнее даже две интересных роли для изофореста. Первая роль – это трансформация фич. Генерируется пачка деревьев, и скор от каждого дерева является новой фичей. В целом это соответствует каким-то мыслям про Deep Forest.

Вторая роль – это высчитывание скоров ”похожести”. Если у нас есть какой-то семпл, для которого мы хотим найти похожих, сперва считаем веса по всем фичам. Веса принимают значения от -1 (для самых аномальных данных) до 1 (для самых регулярных). Потом с этими весами считаются средневзвешенные скоры у всех остальных данных. Похожие данные имеют скор больше, чем непохожие.

В статье [15] предлагают использовать двойной изоляционный лес для поиска аномалий. Строятся два леса – один на исходных данных, а другой на данных после прокручивания через PCA. Выбросом считается объединение выбросов этих лесов. Видимо, они пришли к такому алгоритму чисто перебором. Какие-то аномалии хорошо отлавливал один лес, а какие-то – другой.

Авторы [35] решили удивить мир тем, что распараллелили изофорест на спарке и хаду-пе. Выглядит немного странно. А что, кто-то сомневался, что это можно будет легко сделать?

В статье [28] под флагом оптимизации сделали из изофореста ещё один вариант слу-

чайного леса.

## **Оценка плотности**

Отдельное направление – density estimation. Началось оно похоже ещё до изофореста, где-то в районе книги [33]. Там автор формулирует среднеквадратичный критерий, по которому много кто в последующем приближает исходную плотность вероятности кусочно-постоянной функцией. И даже предлагает иерархический метод для описания плотностей распределений, т.к. деревья. Всю основу заложил.

Какой-то дальнейший толчок дала статья [21], но первый значимый шаг был сделан в [31]. Там авторы проработали целевую функцию среднеквадратичной близости между исходным распределением и кусочно-постоянным приближением к нему. И каждое деление в дереве осуществляют, локально минимизируя целевую функцию. Подход жадный, гарантий минимума не даёт, но определённая субоптимальность наверняка имеется.

В статье [40] приводится подход, где куча всего изменено, а доказательная база слабовата. Приводится как раз идея оценивать собственно сами распределения, а не матожидание логарифма. Даже какая-то теория прилеплена, но не видно доказательств, что это лучше обычного изофореста в смысле поиска аномалий. Если приглядеться к характерным числам на графиках, то более-менее очевидно становится, что поиск аномалий у RS-Forest'a будет хуже.

В статьях [4] [3] автор предлагает отрезать от деревьев куски, которые мало вносят в расхождение между плотностями распределения с учётом регуляризации (tree pruning). Отдельную технологию разводит на тему того, как бы подобрать правильный коэффициент в регуляризации.

Статья [17] предлагает вместо простого критерия деления нод из оригинальной статьи про DET использовать MAP-оптимизацию для поиска среди всех деревьев. Придумывают приоры, рисуют несколько разных лайклихудов. Математическим отжигом гуляют по деревьям и ищут сколько-то оптимальные. В целом идея огонь. Результаты похоже скромные, т.к. картинки не впечатляют. Можно лучше.

Статьи [39] [12] делают нечто мутное. ЯННП. Простите меня.

## **Кластеры аномалий**

Много кто заинтересован в поиске целых кластеров аномалий. Глубоко в эти статьи ещё не вчитывался.

Авторы оригинальной статьи предлагали накрутить поиск аномальных кластеров в [27].

Что именно сделали в статье [22] я искренне не понял. Они что-то как-то порезали на кластеры. Как – хрен знает. Вчитываться в детали пока не собираюсь, т.к. направление мыслей не впечатляет.

## **Стриминг**

Про стриминг вообще чёртовы наркоманы пишут. Либо они прикидываются.

Авторы [14] предлагают резать входные данные на блоки и на основе этих блоков строить новые леса. Написано достаточно мутно, поэтому не ясно ни то, как они считают скорости блоков, ни то, при каких обстоятельствах они считают, что нашли аномалию. Очень скользкие типы.

Очередная статья про стриминг [29] мутная. Режут данные на разные виды аномальностей – подальше от регулярных данных и поближе, покрывают шарами те, которые поближе, дорастивают деревья. Какая-то откровенная дичь, которую я вообще не затащил.

Ещё про стриминг – [18] . Здесь уже интереснее. Сперва авторы меняют алгоритм построения дерева – зачем-то вешают разные веса на разные размерности в зависимости от диаметра данных на этой размерности. Таким образом вновь убивая независимость изофореста от определения расстояния. Но зато у них какие-то свои прикольные мысли на тему того, как считать скор аномалий и как жить в стриме.

Ещё про стриминг – [40] . Та самая статья, обсуждающая то, как оценивать именно плотность вероятности. В названии сказано про стриминг, но конкретно стриминга внутри там очень немного.

### **Активное обучение**

В статье [30] предлагают попробовать интересный вид уменьшения размерностей – модернизировать PCA по SVD-разложению минимаксом. Если верить их графикам, штука очень даже хорошо рабочая.

Что по лесу, то авторы делают важное наблюдение, что при наличии парочки размеченных аномалий уже можно собрать статистики по тому, какие фишки позволяют эти аномалии дискриминировать (аналогичную мысль повторяют авторы статьи [16] ).

Отдельно интересно, что эти коллеги фитят F-распределение к скорам аномальности семплов на изоляционном лесе. Это какой-то теоретический результат?

Статьи [32] и [13] обсуждаются здесь.

### **Общесметодическое**

Отдельно от всех упомяну статью [8] , которая обсуждает, как вообще измерять и сравнивать перфоманс алгоритмов поиска аномалий. Статья не впечатляющая на первый взгляд, но возможно что-то полезное для наших дел можно оттуда подчерпнуть.

Другая статья для отдельного упоминания – [2] . Авторы обсуждают вопросы bias-variance tradeoff, bagging и subsampling в применении к поиску аномалий.

Ещё отдельно интересно почитать повнимательнее книгу [1] . Там хоть всё по поверхности, но зато ссылок много, к которым следует приглянуться.

Есть какая-то странная мысль про то, как на самом деле стоит изображать найденные выбросы. Типа как найти среди двумерных графиков тот набор, который наилучшим образом объяснит, что именно отличает выбросы от регулярных данных [19] .

### **Список источников**

1. Aggarwal C. C., Sathe S. Outlier Ensembles. — Cham : Springer International Publishing, 2017. — ISBN 978-3-319-54764-0 978-3-319-54765-7. — DOI: 10.1007/978-3-319-54765-7. — URL: <http://link.springer.com/10.1007/978-3-319-54765-7> (дата обр. 19.07.2022).
2. Aggarwal C. C., Sathe S. Theoretical Foundations and Algorithms for Outlier Ensembles // ACM SIGKDD Explorations Newsletter. — 2015. — 29 сент. — Т. 17, № 1. — С. 24—47. — ISSN 1931-0145. — DOI: 10.1145/2830544.2830549. — URL: <https://doi.org/10.1145/2830544.2830549> (дата обр. 19.07.2022).

3. *Anderlini L.* Density Estimation Trees as fast non-parametric modelling tools // *Journal of Physics: Conference Series*. — 2016. — Окт. — Т. 762. — С. 012042. — ISSN 1742-6588, 1742-6596. — DOI: 10.1088/1742-6596/762/1/012042. — arXiv: 1607.06635. — URL: <http://arxiv.org/abs/1607.06635> (дата обр. 18.08.2022).
4. *Anderlini L.* Density Estimation Trees in High Energy Physics. — 03.02.2015. — URL: <http://arxiv.org/abs/1502.00932> (дата обр. 18.08.2022).
5. *Aryal S.* [и др.]. Improving iForest with Relative Mass // *Advances in Knowledge Discovery and Data Mining* / под ред. V. S. Tseng [и др.]. — Cham : Springer International Publishing, 2014. — С. 510—521. — ISBN 978-3-319-06605-9. — DOI: 10.1007/978-3-319-06605-9\_42.
6. *Bandaragoda T. R.* [и др.]. Efficient Anomaly Detection by Isolation Using Nearest Neighbour Ensemble // 2014 IEEE International Conference on Data Mining Workshop (2014 IEEE International Conference on Data Mining Workshop (ICDMW)). — Shenzhen, China : IEEE, 12.2014. — С. 698—705. — ISBN 978-1-4799-4274-9 978-1-4799-4275-6. — DOI: 10.1109/ICDMW.2014.70. — URL: <http://ieeexplore.ieee.org/document/7022664/> (дата обр. 20.07.2022).
7. *Bandaragoda T. R.* [и др.]. Isolation-based anomaly detection using nearest-neighbor ensembles // *Computational Intelligence*. — 2018. — Т. 34, № 4. — С. 968—998. — ISSN 1467-8640. — DOI: 10.1111/coin.12156. — URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/coin.12156> (дата обр. 20.07.2022).
8. *Campos G. O.* [и др.]. On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study // *Data Mining and Knowledge Discovery*. — 2016. — 1 июля. — Т. 30, № 4. — С. 891—927. — ISSN 1573-756X. — DOI: 10.1007/s10618-015-0444-8. — URL: <https://doi.org/10.1007/s10618-015-0444-8> (дата обр. 19.07.2022).
9. *Chen C.* [и др.]. iBOAT: Isolation-Based Online Anomalous Trajectory Detection // *IEEE Transactions on Intelligent Transportation Systems*. — 2013. — Июнь. — Т. 14, № 2. — С. 806—818. — ISSN 1558-0016. — DOI: 10.1109/TITS.2013.2238531.
10. *Chen C.* [и др.]. Real-Time Detection of Anomalous Taxi Trajectories from GPS Traces // *Mobile and Ubiquitous Systems: Computing, Networking, and Services* / под ред. A. Puiatti, T. Gu. — Berlin, Heidelberg : Springer, 2012. — С. 63—74. — ISBN 978-3-642-30973-1. — DOI: 10.1007/978-3-642-30973-1\_6.
11. *Chen Y., Wu W.* Isolation Forest as an Alternative Data-Driven Mineral Prospectivity Mapping Method with a Higher Data-Processing Efficiency // *Natural Resources Research*. — 2019. — 1 янв. — Т. 28, № 1. — С. 31—46. — ISSN 1573-8981. — DOI: 10.1007/s11053-018-9375-6. — URL: <https://doi.org/10.1007/s11053-018-9375-6> (дата обр. 20.07.2022).
12. *Cousins C., Riondato M.* CaDET: interpretable parametric conditional density estimation with decision trees and forests // *Machine Learning*. — 2019. — Сент. — Т. 108, № 8/9. — С. 1613—1634. — ISSN 0885-6125, 1573-0565. — DOI: 10.1007/s10994-019-05820-3. — URL: <http://link.springer.com/10.1007/s10994-019-05820-3> (дата обр. 18.08.2022).
13. *Das S.* [и др.]. Incorporating Expert Feedback into Active Anomaly Discovery. —

14. *Ding Z., Fei M.* An Anomaly Detection Approach Based on Isolation Forest Algorithm for Streaming Data using Sliding Window // IFAC Proceedings Volumes. — 2013. — 1 янв. — Т. 46, № 20. — С. 12—17. — (3rd IFAC Conference on Intelligent Control and Automation Science ICONS 2013). — ISSN 1474-6670. — DOI: 10.3182/20130902-3-CN-3020.00044. — URL: <https://www.sciencedirect.com/science/article/pii/S1474667016314999> (дата обр. 19.07.2022).
15. *Elnour M.* [и др.]. A Dual-Isolation-Forests-Based Attack Detection Framework for Industrial Control Systems // IEEE Access. — 2020. — Т. 8. — С. 36639—36651. — ISSN 2169-3536. — DOI: 10.1109/ACCESS.2020.2975066.
16. *Gavai G.* [и др.]. Detecting Insider Threat from Enterprise Social and Online Activity Data // Proceedings of the 7th ACM CCS International Workshop on Managing Insider Security Threats. — New York, NY, USA : Association for Computing Machinery, 16.10.2015. — С. 13—20. — (MIST '15). — ISBN 978-1-4503-3824-0. — DOI: 10.1145/2808783.2808784. — URL: <https://doi.org/10.1145/2808783.2808784> (дата обр. 19.07.2022).
17. *Goh S. T., Rudin C.* Cascaded High Dimensional Histograms: A Generative Approach to Density Estimation. — 12.02.2016. — URL: <http://arxiv.org/abs/1510.06779> (дата обр. 18.08.2022).
18. *Guha S.* [и др.]. Robust Random Cut Forest Based Anomaly Detection on Streams // Proceedings of The 33rd International Conference on Machine Learning (International Conference on Machine Learning). — PMLR, 11.06.2016. — С. 2712—2721. — URL: <https://proceedings.mlr.press/v48/guha16.html> (дата обр. 19.07.2022).
19. *Gupta N.* [и др.]. Beyond Outlier Detection: LookOut for Pictorial Explanation // Machine Learning and Knowledge Discovery in Databases / под ред. М. Berlingerio [и др.]. — Cham : Springer International Publishing, 2019. — С. 122—138. — ISBN 978-3-030-10925-7. — DOI: 10.1007/978-3-030-10925-7\_8.
20. *Hariri S., Kind M. C., Brunner R. J.* Extended Isolation Forest // IEEE Transactions on Knowledge and Data Engineering. — 2021. — Апр. — Т. 33, № 4. — С. 1479—1489. — ISSN 1558-2191. — DOI: 10.1109/TKDE.2019.2947676.
21. *Hooker G.* Diagnosing extrapolation: tree-based density estimation // Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04 (the 2004 ACM SIGKDD international conference). — Seattle, WA, USA : ACM Press, 2004. — С. 569. — DOI: 10.1145/1014052.1014121. — URL: <http://portal.acm.org/citation.cfm?doid=1014052.1014121> (дата обр. 20.08.2022).
22. *Karczmarek P.* [и др.]. K-Means-based isolation forest // Knowledge-Based Systems. — 2020. — 11 мая. — Т. 195. — С. 105659. — ISSN 0950-7051. — DOI: 10.1016/j.knosys.2020.105659. — URL: <https://www.sciencedirect.com/science/article/pii/S0950705120301064> (дата обр. 20.07.2022).
23. *Lesouple J.* [и др.]. Generalized isolation forest for anomaly detection // Pattern Recognition Letters. — 2021. — Сент. — Т. 149. — С. 109—119. — ISSN 01678655. — DOI: 10.1016/j.patrec.2021.05.022. — URL: <https://linkinghub.elsevier.com/retrieve/pii/S0167865521002063> (дата обр. 19.07.2022).

24. *Lin Q.* [и др.]. Disorientation detection by mining GPS trajectories for cognitively-impaired elders // *Pervasive and Mobile Computing*. — 2015. — Май. — Т. 19. — С. 71—85. — ISSN 15741192. — DOI: 10.1016/j.pmcj.2014.01.003. — URL: <https://linkinghub.elsevier.com/retrieve/pii/S1574119214000200> (дата обр. 20.07.2022).
25. *Liu F. T., Ting K. M., Zhou Z.-H.* Isolation Forest // 2008 Eighth IEEE International Conference on Data Mining (2008 Eighth IEEE International Conference on Data Mining (ICDM)). — Pisa, Italy : IEEE, 12.2008. — С. 413—422. — ISBN 978-0-7695-3502-9. — DOI: 10.1109/ICDM.2008.17. — URL: <http://ieeexplore.ieee.org/document/4781136/> (дата обр. 16.07.2021).
26. *Liu F. T., Ting K. M., Zhou Z.-H.* Isolation-Based Anomaly Detection // *ACM Transactions on Knowledge Discovery from Data*. — 2012. — Март. — Т. 6, № 1. — С. 1—39. — ISSN 1556-4681, 1556-472X. — DOI: 10.1145/2133360.2133363. — URL: <https://dl.acm.org/doi/10.1145/2133360.2133363> (дата обр. 16.07.2021).
27. *Liu F. T., Ting K. M., Zhou Z.-H.* On Detecting Clustered Anomalies Using SCiForest // *Machine Learning and Knowledge Discovery in Databases* / под ред. J. L. Balcázar [и др.]. — Berlin, Heidelberg : Springer, 2010. — С. 274—290. — ISBN 978-3-642-15883-4. — DOI: 10.1007/978-3-642-15883-4\_18.
28. *Liu Z.* [и др.]. An Optimized Computational Framework for Isolation Forest // *Mathematical Problems in Engineering*. — 2018. — 11 июня. — Т. 2018. — e2318763. — ISSN 1024-123X. — DOI: 10.1155/2018/2318763. — URL: <https://www.hindawi.com/journals/mpe/2018/2318763/> (дата обр. 01.08.2022).
29. *Mu X., Ting K. M., Zhou Z.-H.* Classification Under Streaming Emerging New Classes: A Solution Using Completely-Random Trees // *IEEE Transactions on Knowledge and Data Engineering*. — 2017. — Авг. — Т. 29, № 8. — С. 1605—1618. — ISSN 1558-2191. — DOI: 10.1109/TKDE.2017.2691702.
30. *Puggini L., McLoone S.* An enhanced variable selection and Isolation Forest based methodology for anomaly detection with OES data // *Engineering Applications of Artificial Intelligence*. — 2018. — 1 янв. — Т. 67. — С. 126—135. — ISSN 0952-1976. — DOI: 10.1016/j.engappai.2017.09.021. — URL: <https://www.sciencedirect.com/science/article/pii/S095219761730235X> (дата обр. 20.07.2022).
31. *Ram P., Gray A. G.* Density estimation trees // *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11 (the 17th ACM SIGKDD international conference)*. — San Diego, California, USA : ACM Press, 2011. — С. 627. — ISBN 978-1-4503-0813-7. — DOI: 10.1145/2020408.2020507. — URL: <http://dl.acm.org/citation.cfm?doid=2020408.2020507> (дата обр. 18.08.2022).
32. *Siddiqui M. A.* [и др.]. Feedback-Guided Anomaly Discovery via Online Optimization // *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. — New York, NY, USA : Association for Computing Machinery, 19.07.2018. — С. 2200—2209. — (KDD '18). — ISBN 978-1-4503-5552-0. — DOI: 10.1145/3219819.3220083. — URL: <https://doi.org/10.1145/3219819.3220083> (дата обр. 19.07.2022).

33. *Silverman B. W.* Density Estimation for Statistics and Data Analysis. — Routledge, 1986. — ISBN 978-1-315-14091-9. — DOI: 10.1201/9781315140919. — URL: <https://www.taylorfrancis.com/books/mono/10.1201/9781315140919/density-estimation-statistics-data-analysis-silverman> (дата обр. 20.08.2022).
34. *Sun L.* [и др.]. Detecting Anomalous User Behavior Using an Extended Isolation Forest Algorithm: An Enterprise Case Study. — 21.09.2016. — URL: <http://arxiv.org/abs/1609.06676> (дата обр. 20.07.2022).
35. *Tao X.* [и др.]. A parallel algorithm for network traffic anomaly detection based on Isolation Forest // International Journal of Distributed Sensor Networks. — 2018. — 1 нояб. — Т. 14, № 11. — С. 1550147718814471. — ISSN 1550-1329. — DOI: 10.1177/1550147718814471 URL: <https://doi.org/10.1177/1550147718814471> (дата обр. 01.08.2022).
36. *Tokovarov M., Karczmarek P.* A probabilistic generalization of isolation forest // Information Sciences. — 2022. — 1 янв. — Т. 584. — С. 433—449. — ISSN 0020-0255. — DOI: 10.1016/j.ins.2021.10.075. — URL: <https://www.sciencedirect.com/science/article/pii/S0020025521010999> (дата обр. 19.07.2022).
37. *Utkin L. V., Kovalev M. S., Meldo A. A.* A deep forest classifier with weights of class probability distribution subsets // Knowledge-Based Systems. — 2019. — 1 июня. — Т. 173. — С. 15—27. — ISSN 0950-7051. — DOI: 10.1016/j.knosys.2019.02.022. — URL: <https://www.sciencedirect.com/science/article/pii/S0950705119300838> (дата обр. 01.08.2022).
38. *Vinh N. X.* [и др.]. Discovering outlying aspects in large datasets // Data Mining and Knowledge Discovery. — 2016. — 1 нояб. — Т. 30, № 6. — С. 1520—1555. — ISSN 1573-756X. — DOI: 10.1007/s10618-016-0453-2. — URL: <https://doi.org/10.1007/s10618-016-0453-2> (дата обр. 20.07.2022).
39. *Wells J. R., Ting K. M.* A simple efficient density estimator that enables fast systematic search. — 12.09.2017. — URL: <http://arxiv.org/abs/1707.00783> (дата обр. 18.08.2022).
40. *Wu K.* [и др.]. RS-Forest: A Rapid Density Estimator for Streaming Anomaly Detection // 2014 IEEE International Conference on Data Mining (2014 IEEE International Conference on Data Mining). — 12.2014. — С. 600—609. — DOI: 10.1109/ICDM.2014.45.
41. *Yu X., Tang L. A., Han J.* Filtering and Refinement: A Two-Stage Approach for Efficient and Effective Anomaly Detection // 2009 Ninth IEEE International Conference on Data Mining (2009 Ninth IEEE International Conference on Data Mining). — 12.2009. — С. 617—626. — DOI: 10.1109/ICDM.2009.44.
42. *Zhang D.* [и др.]. iBAT: detecting anomalous taxi trajectories from GPS traces // Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11 (the 13th international conference). — Beijing, China : ACM Press, 2011. — С. 99. — ISBN 978-1-4503-0630-0. — DOI: 10.1145/2030112.2030127. — URL: <http://dl.acm.org/citation.cfm?doid=2030112.2030127> (дата обр. 19.07.2022).
43. *Zhou G.-T.* [и др.]. Relevance feature mapping for content-based multimedia information retrieval // Pattern Recognition. — 2012. — 1 апр. — Т. 45, № 4. — С. 1707—1720. — ISSN 0031-3203. — DOI: 10.1016/j.patcog.2011.09.016. — URL: <https://www.sciencedirect.com/science/article/pii/S0031320311003955> (дата обр. 20.07.2022).



44. *Zhou Z.-H., Feng J.* Deep forest // National Science Review. — 2019. — 1 янв. — Т. 6, № 1. — С. 74—86. — ISSN 2095-5138. — DOI: 10.1093/nsr/nwy108. — URL: <https://academic.oup.com/nsr/article/6/1/74/5123737> (дата обр. 19.07.2022).
45. *Zhou Z.-H., Feng J.* Deep Forest: Towards An Alternative to Deep Neural Networks // Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (Twenty-Sixth International Joint Conference on Artificial Intelligence). — Melbourne, Australia : International Joint Conferences on Artificial Intelligence Organization, 08.2017. — С. 3553—3559. — ISBN 978-0-9992411-0-3. — DOI: 10.24963/ijcai.2017/497. — URL: <https://www.ijcai.org/proceedings/2017/497> (дата обр. 19.07.2022).