УДК 004
DOI https://doi.org/10.32838/2663-5941/2022.3/06

Завгородній В.В.
Державний університет інфраструктури та технологій

Завгородня Г.А.
Державний університет інфраструктури та технологій

Валявська Н.О.
Державний університет інфраструктури та технологій

Герасименко О.О.
Державний університет інфраструктури та технологій

Калюжний О.В.
Державний університет інфраструктури та технологій

Степовий А.В.
Державний університет інфраструктури та технологій

ПОШУК АНОМАЛІЙ У ДАНИХ ЗА ДОПОМОГОЮ МАШИНОГО НАВЧАННЯ

Робота присвячена питанню пошуку аномалій у даних, зокрема практичному аспекту даного питання. Цей напрямок у машинному навчанні є досить новим, тому в ньому багато відкритих завдань та серйозних викликів.

В основі машинного навчання лежить припущення, що дані для навчання, тестування та застосування взяті з одного й того самого розподілу. На жаль, у процесі застосування моделі це припущення може порушуватися, що призводить до незрозумілих наслідків – зсуву розподілу. Особливо такі порушення небезпечні в галузях, що вимагають швидкого та точного прийняття рішень.

Невизначеність у даних виникає через складну структуру даних, шумів і неоднозначність, а невизначеність у знаннях виникає у зв'язку з обмеженою кількістю прикладів, використовуваних для навчання моделі, або з відсутністю доступу до певної області даних.

Для вирішення завдання пошуку аномалій пропонується побудувати модель, яка б за певним прецедентом видаватиме значення, що трактується як міра аномальності. Після чого обирається певний поріг: всі прецеденти зі значенням аномальності менше оголошуються звичайними прецедентами, а інші – аномаліями.

Для оцінки якості роботи запропонованого методу пошуку аномалій використано підхід, що полягає у змішуванні двох наборів даних. Для цього беруться тестові частини двох наборів: одного, на навчальній частині якого навчалася модель, та другого, який модель ніколи не бачила. Ці тестові частини змішуються, а потім метод пошуку аномалій тестується як бінарний класифікатор: він відокремлює набір даних, який бачила модель, від набору даних, який модель бачить вперше.

Напрямок градієнтних методів є перспективним і конкурентоспроможним у порівнянні з описаними в літературі алгоритмами. У цій галузі грамотна постановка задачі та методика тестування рішень важлива не менше, ніж винахід нових рішень.

Ключові слова: *машинне навчання, аномалії даних, пошук аномалій, метод градієнта, класифікація зображень.*

Постановка проблеми. В основі машинного навчання лежить припущення, що дані для навчання, тестування та застосування взяті з одного й того самого розподілу. На жаль, у процесі застосування моделі це припущення може порушуватися, що призводить до незрозумілих наслідків – зсуву розподілу [1; 2]. Особливо такі порушення небезпечні в галузях, що вимагають швидкого та точного прийняття рішень: медицина, фінанси, безпілотні автомобілі.

Системи машинного навчання часто просто ігнорують зсуви розподілу та продовжують

працювати у штатному режимі, не припускаючи, що відповіді на запити можуть бути невалідними [3; 4]. Щоб зробити роботу системи точнішою та зрозуміти причину походження некоректних даних, можна відловлювати такі порушення – потрібно лише додати можливість пошуку аномалій.

Машинне навчання і, зокрема, нейронні мережі вже глибоко проникли у багато сфер життя суспільства: рекомендації у соціальних мережах, медична діагностика, пошук таксі, торгівля на біржі, безпілотні автомобілі тощо. У деяких із цих галузей ціна помилки не велика, а в деяких – може коштувати людського життя [2; 5]. Тому дуже важливо, щоб практичне застосування алгоритмів машинного навчання було максимально безпечним.

Пошук аномалій у даних – це окремий випадок ширшого завдання з оцінки невизначеності передбачення. Вона поділяється на оцінку невизначеності у даних (*aleotoric uncertainty*) та оцінку невизначеності у знаннях (*epistemic uncertainty*). Невизначеність у даних виникає через складну структуру даних, шумів і неоднозначність, а невизначеність у знаннях виникає у зв'язку з обмеженою кількістю прикладів, використовуваних для навчання моделі, або з відсутністю доступу до певної області даних. Дана стаття наголошує в першу чергу на практичному аспекті процесу пошуку аномалій у даних.

Аналіз останніх досліджень і публікацій. Для того, щоб застосовувати пошук аномалій до довільних завдань машинного навчання, насамперед треба вирішити завдання класифікації. Розглянемо вирішення цього завдання на прикладі класифікації зображень, оскільки це досить широка, але водночас проста область моделювання.

Для завдання класифікації зсуви розподілу можуть бути, наприклад, такими:

- додавання шуму до зображень, що надходять;
- прецеденти, що належать класам, які були відсутні в навчальній вибірці;
- прецеденти, що належать класу, який був в навчальній вибірці, але представлені у новій текстурі або формі.

Завдання пошуку аномалій можна сформулювати так: потрібно побудувати якусь модель M, яка б за прецедентом x видавала значення $M(x)$, яке можна трактувати як міру аномальності. Після чого обирається певний поріг λ: всі прецеденти зі значенням аномальності менше λ оголошуються звичайними прецедентами, а інші – аномаліями. Тобто, щоб перевірити, чи є x аномалією, потрібно перевірити, чи правильна нерівність $M(x) \le \lambda$.

Найпоширенішим способом оцінити якість роботи вищеописаного методу пошуку аномалій є змішування двох наборів даних. Для цього беруться тестові частини двох наборів: одного (інлаєра), на навчальній частині якого навчалася модель, та другого (аномалії), який модель ніколи не бачила. Ці тестові частини змішуються, а потім метод пошуку аномалій тестується як бінарний класифікатор: він повинен відокремити набір даних, який бачила модель, від набору даних, який модель бачить вперше. За метрику якості можна брати будь-які метрики для бінарної класифікації, наприклад класичні метрики *ROC-AUC* або *PR-AUC*.

Це стандартна процедура апробації методів пошуку аномалій [6–9]. Проте ця процедура має істотний недолік: вона перевіряє лише те, як добре метод знаходить аномалії, але ніяк не враховує якість вирішення вихідної задачі класифікації на прецедентах, оголошених інлаєрами. У даній статті пропонується новий спосіб оцінювання пошуку аномалій, який враховує цей недолік.

Для вирішення завдань пошуку аномалій у даних були розглянуті наступні існуючі рішення:

- *Maximum Softmax Probability* (*MSP*) – це найпростіший та інтуїтивно зрозумілий метод [6]. В якості $M(x)$ використовується негативна максимальна *softmax*-ймовірність. Таким чином, якщо нейронна мережа видає якийсь клас із досить високою ймовірністю, то прецедент x оголошується інлаєром.
- *ODIN* – це модифікований метод *MSP*, який використовує додатковий препроцесінг зображення [7]. Такий самий препроцесінг використовується у запропонованому в даній статті градієнтному методі.
- Ансамблеві методи – це методи, що використовують кілька навчених різними генераторами випадкових значень нейронних мереж однакової архітектури [8; 9]. Їх передбачення агрегуються певним чином, щоб отримати міру аномальності прецеденту. Ансамблеві методи математично обґрунтовані, але надто важкі для інтеграції у виробничі системи.

Постановка завдання. Машинне навчання припускає, що вхідні дані в процесі експлуатації моделі беруться з того самого розподілу, з якого були взяті дані на етапі навчання. Насправді, це припущення виконується вкрай рідко: у таких випадках спостерігається зсув розподілу, що може призвести до практично будь-яких наслідків. Метою даної статті є дослідження зсувів розподілу, які можуть спостерігатися, та способів їх моделювання.

Виклад основного матеріалу дослідження. Для пошуку аномалій у даних при рішенні завдань класифікації зображень необхідне застосування певних тестів продуктивності. Найпопулярнішим тестом для продуктивності класифікації зображень є набір даних *ImageNet-1k* [10]. На його основі було створено тести для пошуку аномалій у завданні класифікації зображень. Для проведення досліджень було використано наступні набори даних:

- *ImageNet-O* – це деяка підмножина набору даних *ImageNet-22k*, яка не перетинається з *ImageNet-1k* [11]. Цей набір є зсувом, в якому додаються нові класи, які модель раніше не бачила.
- *ImageNet-R* складається із зображень, що належать класам оригінального набору даних *ImageNet-1k*, але представлені в інших текстурах та формах [12].
- *ImageNet-A* – це природні приклади, тобто зображення, які дуже складно коректно класифікувати нейронними мережами, навіть за умови присутності даних класів зображень в оригінальному наборі даних *ImageNet-1k* [11].
- *ImageNet-C* – це зашумлена версія тестової частини *ImageNet-1k* [13]. Набір даних складається з кількох видів шумів та кількох рівнів сили шуму. У роботі використовується лише *Frosted Glass Blur* з рівнем 5. Вибір конкретного шуму зумовлений тим, що у ньому *ResNet-50* показує найгірший результат класифікації.

Використання аналізу градієнтів для вирішення поставленого завдання, пояснюється двома підходами:

- *Influence functions* пояснює зміну прогнозу нейронної мережі при вилученні певного прецеденту з набору даних [14]. Це реалізується формулюванням першої та другої похідної за вагою моделі.
- *Neural Tangent Kernel* аналізує поведінку нескінченно широких нейронних мереж з погляду простору похідних за вагою моделі [15].

Завдання класифікації об'єктів a між класами $K = (1,...,k)$ найчастіше вирішують за допомогою мінімізації крос-ентропії:

$$-\sum_{k=1}^{K} r_{a,k} \log p_{a,k} , \qquad (1)$$

де r – розподіл істинних значень, а p – розподіл ймовірностей прогнозів моделі.

Евклідову норму градієнта крос-ентропії за вагою моделі можна використовувати як міру аномальності:

$$M(x) = \left\| \nabla_{\Theta} \left(-\sum_{k=1}^{K} r_{a,k} \log p_{a,k} \right) \right\|_2 , \qquad (2)$$

де Θ – оцінка наближення невідомого параметра на основі деяких даних.

Інтуїтивно це можна розуміти так: якщо прецедент є аномалією, модель не знає, що з ним робити. Тоді градієнт буде досить великим, оскільки прецедент несе у собі велику кількість інформації. Якщо ж прецедент є інлаєром, він принесе у собі малу кількість нової інформації, що виявиться у невеликому значенні градієнта.

Припустимо, що нейронна мережа $f = (f_1,...,f_k)$ навчена класифікувати K класів. Для кожного входу x нейронна мережа призначає мітку, обчислюючи *softmax* вихід для кожного класу. Далі можна докладніше розписати градієнт крос-ентропії w, щоб декомпозувати його у добуток двох множників:

$$\nabla_{\Theta} w(x,\hat{\Theta}) = -\nabla_{\Theta} \log \frac{\max_{k'} e^{f_{k'}(x,\Theta)}}{\sum_{k=1}^{N} e^{f_k(x,\Theta)}}\Bigg|_{\hat{\Theta}} = V(x,\hat{\Theta}) \cdot D(x,\hat{\Theta}) , , \qquad (3)$$

де $\hat{\Theta}$ – точкова оцінка наближення невідомого параметра.

V – частина – це максимум *softmax*-ймовірності, а D – частина – це множник, що відповідає за похідну за вагами моделі. Така декомпозиція дозволяє відокремити два джерела інформації одне від одного. У якості градієнтного методу можна використовувати добуток $(V(x,\Theta) \cdot D(x,\Theta))$ або ж просто $D(x,\Theta)$.

$$V(x,\Theta) = \frac{1}{1+\sum_{k \neq \hat{k}} e^{\left(f_k(x,\hat{\Theta}) - f_{\hat{k}}(x,\hat{\Theta})\right)}} \qquad (4)$$

$$D(x,\Theta) = -\sum_{k \neq \hat{k}} e^{(f_k(x,\hat{\Theta}) - f_{\hat{k}'}(x,\hat{\Theta}))} \frac{\partial (f_{k'}(x,\hat{\Theta}) - f_k(x,\Theta))}{\partial \Theta}\Bigg|_{\hat{\Theta}} \qquad (5)$$

Для пошуку аномалій використаємо алгоритм *ODIN*, який реалізує спеціальний препроцесінг, що складається з двох частин: внесення шуму в *softmax* і вихідну картинку [7]. Замість звичайного *softmax* використаємо *softmax* із температурою T:

$$D(x,\Theta) = -\sum_{k \neq \hat{k}} e^{(f_k(x,\hat{\Theta}) - f_{k'}(x,\hat{\Theta}))} \frac{\partial (f_{k'}(x,\hat{\Theta}) - f_k(x,\Theta))}{\partial \Theta}\Bigg|_{\hat{\Theta}} \qquad (6)$$

Замість прецеденту x візьмемо

$$x_p = x - \gamma sign\left(-\nabla_x \log V_{k'}(x,T)\right) \qquad (7)$$

Потім до прецеденту x_p застосовується описане вище обчислення норми градієнта, щоб отримати функцію $M(x)$.

Оптимальні значення параметрів γ та T підбираються методом перебору по сітці на невеликій валідаційній вибірці.

Для експериментів було обрано дві архітектури нейронних мереж: *ResNet-18* та *ResNet-50*. Для експериментів з *ResNet-50* було навчено чотири

Таблиця 1

ImageNet*-подібні набори даних. Значення *ROC-AUC

	Dataset	***MSP***	***ODIN***	***Ensemble***	***G-part***	***SG-part***
ResNet-18	*ImageNet-O*	0,484 ± 0,006	0,628 ± 0,007	—	0,760 ± 0,006	0,804 ± 0,006
	ImageNet-R	0,776 ± 0,001	0,840 ± 0,001	—	0,852 ± 0,001	0,833 ± 0,002
	ImageNet-A	0,818 ± 0,003	0,848 ± 0,002	—	0,853 ± 0,002	0,837 ± 0,002
	ImageNet-O	0,943 ± 0,001	0,974 ± 0,001	—	0,985 ± 0,001	0,950 ± 0,001
ResNet-50	*ImageNet-O*	0,470 ± 0,006	0,599 ± 0,007	0,612 ± 0,006	0,735 ± 0,006	0,769 ± 0,006
	ImageNet-R	0,803 ± 0,002	0,863 ± 0,001	0,856 ± 0,001	0,877 ± 0,001	0,856 ± 0,001
	ImageNet-A	0,839 ± 0,002	0,872 ± 0,002	0,885 ± 0,002	0,877 ± 0,002	0,859 ± 0,002
	ImageNet-O	0,937 ± 0,001	0,969 ± 0,001	0,976 ± 0,001	0,982 ± 0,001	0,958 ± 0,001

моделі з різним генераторами випадкових значень для тестування ансамблевих методів (табл. 1).

З таблиці 1 видно, що запропоновані методи (*G-part* та *SG-part*) перевершують решту бейзлайнів у семи випадках з восьми. *G* показує себе краще практично за всі бейзлайни, у той час як *SG* обганяє *G* на *ImageNet-O*, але йому не вистачає якості на *ImageNet-A/R/C* щодо інших бейзлайнів. Це досить несподіваний ефект, якому поки що немає пояснень. Таким чином, *G-part* вже достатньо для переваги над іншими рішеннями, але його результат на якихось доменах можна покращити, додавши інформацію від останнього шару мережі, тобто *softmax*-прогнозування моделі.

Варто зазначити, що єдиний бейзлайн, котрому програли градієнтні методи, – це ансамблі. Порівняння з ними дещо некоректне, оскільки це набагато «важчі» методи ніж *MSP*, *ODIN* та запропоновані *G-part* та *SG-part*. Ансамблі практично неможливо застосовувати на практиці, на відміну від інших, «легших» методів, які можна без проблем вбудовувати в реальні завдання. Таким чином, у цьому програші немає нічого незвичайного, але немає й нічого страшного.

Важливо додати, що *ODIN* та градієнтні методи вимагають налаштування гіперпараметрів на валідаційній вибірці, що накладає деякі обмеження на використання методу. Проте існує низка ідей, які можуть дозволити позбутися настроювання параметрів.

Висновки. У даному дослідженні розглянуто питання пошуку аномалій у даних. Цей напрямок у машинному навчанні є досить новим, тому в ньому багато відкритих завдань та серйозних викликів. Напрямок градієнтних методів є перспективним і конкурентоспроможним у порівнянні з описаними в літературі алгоритмами. У цій галузі грамотна постановка задачі та методика тестування рішень важлива не менше, ніж винахід нових рішень.

Список літератури:

1. V. Mukhin, Y. Komaga, V. Zavgorodnii, A. Zavgorodnya, O. Herasymenko and O. Mukhin, "Social Risk Assessment Mechanism Based on the Neural Networks," 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), 2019, pp. 179-182. DOI: https://doi.org/10.1109/ATIT49449.2019.9030519.

2. Valerii Zavgorodnii, Anna Zavgorodnya, Vladyslav Maiko, Valerii Malikov, & Dmytro Zhuk. (2018). METHODS AND MODELS FOR ASSESSMENT OF RELIABILITY OF STRUCTURAL-COMPLEX SYSTEMS. World Science, (11(39), 5-14. DOI: https://doi.org/10.31435/rsglobal_ws/30112018/6227

3. V. Mukhin, Y. Kornaga, M. Bazaliy, V. Zavgorodnii, I. Krysak and O. Mukhin, "Obfuscation Code Technics Based on Neural Networks Mechanism," 2020 IEEE 2nd International Conference on System Analysis & Intelligent Computing (SAIC), 2020, pp. 1-6. DOI: https://doi.org/10.1109/SAIC51296.2020.9239247.

4. Valerii Zavgorodnii, Anna Zavgorodnya, Vladyslav Plisenko, Nikita Provatorov, & Pavlo Kudientsov. (2019). METHODS MODELING SYSTEMS FOR THE IMPROVEMENT OF THEIR RELIABILITY. International Academy Journal Web of Scholar, 1(9(39), 3-11. DOI: https://doi.org/10.31435/rsglobal_wos/30092019/6683

5. Mukhin, V., Zavgorodnii, V., Barabash, O.V., Mykolaichuk, R.A., Kornaga, Y., Zavgorodnya, A., & Statkevych, V. (2020). Method of Restoring Parameters of Information Objects in a Unified Information Space Based on Computer Networks. International Journal of Computer Network and Information Security, vol.12, no.2, pp. 11–21. DOI: https://doi.org/10.5815/ijcnis.2020.02.02

6. Hendrycks, D., & Gimpel, K. (2016). A baseline for detecting misclassified and out-of-distribution examples in neural networks. DOI: https://doi.org/10.48550/arXiv.1610.02136

7. Liang, S., Li, Y., & Srikant, R. (2017). Enhancing the reliability of out-of-distribution image detection in neural networks. DOI: https://doi.org/10.48550/arXiv.1706.02690

8. Malinin, A., Mlodozeniec, B., & Gales, M. (2019). Ensemble distribution distillation. DOI: https://doi.org/10.48550/arXiv.1905.00076

9. Malinin, A., & Gales, M. (2018). Predictive uncertainty estimation via prior networks. Advances in neural information processing systems, 31. DOI: https://doi.org/10.48550/arXiv.1802.10501

10. Russakovsky, O., Deng, J., Su, H. et al. ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 115, 211–252 (2015). DOI: https://doi.org/10.1007/s11263-015-0816-y

11. Hendrycks, D., Zhao, K., Basart, S., Steinhardt, J., & Song, D. (2021). Natural adversarial examples. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 15262-15271). DOI: https://doi.org/10.48550/arXiv.1907.07174

12. Hendrycks, D., Basart, S., Mu, N., Kadavath, S., Wang, F., Dorundo, E., ... & Gilmer, J. (2021). The many faces of robustness: A critical analysis of out-of-distribution generalization. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 8340-8349). DOI: https://doi.org/10.48550/arXiv.2006.16241

13. Hendrycks, D., & Dietterich, T. (2019). Benchmarking neural network robustness to common corruptions and perturbations. DOI: https://doi.org/10.48550/arXiv.1903.12261

14. Koh, P. W., & Liang, P. (2017, July). Understanding black-box predictions via influence functions. In International conference on machine learning (pp. 1885-1894). PMLR. DOI: https://doi.org/10.48550/arXiv.1703.04730

15. Jacot, A., Gabriel, F., & Hongler, C. (2018). Neural tangent kernel: Convergence and generalization in neural networks. Advances in neural information processing systems, 31. DOI: https://doi.org/10.48550/arXiv.1806.07572

Zavgorodnii V.V., Zavgorodnya A.A., Valyavska N.O., Gerasimenko O.O., Kalyuzhny O.V., Stepovy A.V. **SEARCHING FOR ANOMALIES IN MACHINE LEARNING DATA**

The work is devoted to the issue of finding anomalies in the data, in particular the practical aspect of this issue. This direction in machine learning is quite new, so it has many open tasks and serious challenges.

Machine learning is based on the assumption that data for training, testing and application are taken from the same distribution. Unfortunately, in the process of applying the model, this assumption can be violated, which leads to unclear consequences – a shift in the distribution. Such violations are especially dangerous in industries that require quick and accurate decision-making.

Uncertainty in data arises from complex data structures, noise, and ambiguity, and uncertainty in knowledge arises from the limited number of examples used to teach the model or the lack of access to a particular area of data.

To solve the problem of finding anomalies, it is proposed to build a model that would, according to a certain precedent, give a value that is interpreted as a measure of anomaly. Then a certain threshold is chosen: all precedents with the value of an anomaly are declared less than ordinary precedents, and others – anomalies.

To assess the quality of the proposed method of finding anomalies, an approach was used, which consists in mixing two sets of data. For this purpose, test parts of two sets are taken: one, on the training part of which the model studied, and the second, which the model has never seen. These test pieces are mixed, and then the anomaly search method is tested as a binary classifier: it separates the data set that the model saw from the data set that the model sees for the first time.

The direction of gradient methods is promising and competitive in comparison with the algorithms described in the literature. In this area, the competent formulation of the problem and the method of testing solutions is no less important than the invention of new solutions.

Key words: *machine learning, data anomalies, anomaly search, gradient method, image classification.*