

М. Р. Братищенко, Т. В. Філімончук, Г. В. Майстренко, В. І. Сітніков

Харківський національний університет радіоелектроніки, Харків, Україна

АНАЛІЗ МЕТОДІВ ВИЯВЛЕННЯ АНОМАЛІЙ У ДАНИХ ПРО СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ

Анотація. Актуальність. Сучасні люди постійно шукають нові способи використання енергії для покращення свого життя, тому попит на неї зростає. У більшості випадків компаніям і галузям важко контролювати всі свої пристрої одночасно, що може призвести до втрати електроенергії в будь-який час. В результаті операційні витрати будуть більшими, ніж необхідно. Крім того, втрата електроенергії сприяє глобальному потеплінню через вивільнення вуглецю, коли енергія генерується шляхом спалювання вугілля, газу та нафти. Отже, потрібні рішення для вирішення цих проблем. **Метою даної роботи** є аналіз існуючих методів виявлення аномалій в даних задля вирішення проблеми надмірного споживання електроенергії та попередження про критичні значення в показниках електроенергії, споживаної різноманітними пристроями та електрообладнанням. **Об'єктом дослідження** є процес виявлення нетипових значень або значних відхилень в показниках споживання електроенергії за такими параметрами, як: напруга, сила, частота струму, потужність. **Предметом дослідження** є моделі та методи виявлення аномалій в даних. **Результати.** Після ретельного аналізу кожного з перелічених методів виявлення аномалій відкриваються нові можливості для вирішення проблеми енергоспоживання. Наприклад: об'єднання декількох методів в один; розробка моделі машинного навчання на основі одного або декількох методів, тренування на тестових даних і в перспективі оброблення реальних даних про енергоспоживання з метою визначення нетипових значень, з можливістю фіксування дати та часу виникнення аномалій, а також побудовою різноманітних графіків на основі цієї інформації. **Висновок.** Завдяки розглянутим методам виявлення аномалій можна запобігти великому споживанню електроенергії для досягнення енергозбереження, нагадувати користувачам про визначення несправних електроприладів або змінювати неправильні схеми споживання електроенергії, знижувати витрати користувачів на енергоспоживання та сприяти обізнаності щодо безпеки споживання електроенергії.

Ключові слова: виявлення аномалій, енергоспоживання, статистичні дані, машинне навчання, кластеризація, методи.

Вступ

На сьогоднішній день людство використовує електроенергію багато в яких сферах свого життя, починаючи від побутових електроприладів, і закінчуючи електрообладнанням на підприємствах та електростанціях. Проте треба пам'ятати, що надмірне споживання електроенергії це погано, а особливо якщо є різкі скачки напруги – можуть погоріти електроприлади або електрообладнання вийде з ладу. Виявлення аномалій може запобігти великому споживанню електроенергії для досягнення енергозбереження, нагадувати користувачам про виявлення несправних електроприладів або змінювати неправильні схеми споживання електроенергії, знижувати витрати користувачів на енергоспоживання та сприяти обізнаності щодо безпеки споживання електроенергії.

На початку аналізу досліджень стосовно існуючих методів та рішень щодо врегулювання енергоспоживання та ідентифікації нетипових значень варто ознайомитись із статистичними даними про енергоспоживання у різних країнах світу за останні роки.

У статті [1] представлені та проаналізовані дані, взяті з кількох досліджень про енергоспоживання будівель у США, ЄС та БРІК країн (Бразилія, Росія, Індія, Китай). Більшість поточних досліджень споживання енергії стосується статистики конкретної країни. Однак міжнародні порівняння корисні для виявлення історичних, фактичних тенденцій та тенденцій споживання енергії. Дані представлені у звітах Світового банку, Програми ООН з навколишнього середовища, Міжурядової групи експертів зі змін клімату та міжнародного енергетичного агентства. Вони порівнюються з національними звітами, а також із до-

слідженнями. Цей аналіз показує, що країни БРІК вже подолали загальне енергоспоживання розвинутих країн. Але розширення їх будівельного фонду викликає нагальну потребу в енергоефективності в будівлях. Водночас, можна зробити висновок, що заходів, прийнятих у розвинених країнах, недостатньо для того, щоб гарантувати значне скорочення енергії споживання в будівлях та на підприємствах.

У статті [2] проаналізовано виявлення аномалій у даних щодо споживання електроенергії за допомогою 2 методів: Isolation Forest та Gaussian Naïve Bayes. Автори докладно описують сутності цих методів, наводять формули для розрахунку ймовірностей виникнення аномалій у даних за період часу і пропонують моделі, які можливо натренувати на основі цих методів. Результатом стали аналіз цих даних споживання електроенергії з відповідними графіками. Проведене моделювання показало, що було досягнуто збалансованої оцінки точності щонайменше 0,8947.

Підхід Isolation Forest використовувався для позначення нормальних та ненормальних даних, і він успішно ідентифікував аномальні стрибки на основі поведінки моделі споживання енергії, а не просто на величині споживання. У виявленні аномалій Gaussian Naïve Bayes дав задовільну продуктивність у визначенні ненормальних, а також нормальних точок. Прогнозування споживання електроенергії та ідентифікація аномалій є критично важливими у функціонуванні енергомережі, а обробка багатозмінних часових рядів є складною задачею. Автори статті [3] представили модель, яка поєднує підходи Transformer та K-means. Кожні 23 години навчальні дані розділяються на k кластерів за допомогою кластеризації K-середніх. У той же час ці навчальні дані використовуються для навчання

моделі Transformer для прогнозування споживання електроенергії за наступну годину, при цьому прогнозоване значення поміщається в навчений кластер K-середніх, а центроїд кластера виступає як остаточне прогнозоване значення. Нарешті, для визначення аномалії, було зроблене порівняння очікуваного значення з фактичними результатами тесту. Експериментальні результати у вигляді графіків доводять, що модель забезпечує точність прогнозування з меншою похибкою та високою ефективністю виявлення аномалій.

В дослідженні [4] автори охопили багато питань, починаючи від теоретичних відомостей щодо аномалій в даних та їх виявлення, аналіз існуючих математичних методів та алгоритмів машинного навчання, і закінчуючи розробкою концепції фреймворку для виявлення аномалій споживання електроенергії за допомогою хмарних обчислень. На початку було класифіковано 3 види аномалій: точкові, контекстуальні та колективні. Далі проведено аналіз методів для виявлення аномалій, які найбільш підходять під кожен вид. На основі аналізу автори запропонували своє бачення, яким чином можна визначати нетипові показники у даних про споживання електроенергії за допомогою хмарних обчислень. Наприкінці дослідження було приділено увагу точності виявлення аномалій та можливим способам покращення цього показника.

Автори публікації [5] застосували нову техніку лямбда-архітектури до системи виявлення аномалій у даних споживання електроенергії, щоб підтримувати пакетне оновлення моделей та моніторинг в реальному часі. Вони запропонували алгоритм викриття для пошуку аномалій на основі історії споживання за допомогою контрольованого навчання та статистичних алгоритмів. Крім того, система підтримує персоналізовану службу оповіщення шляхом встановлення порогового значення для ненормальних показників споживання енергії. Було оцінено точність виявлення аномалій алгоритму на наборі даних реального світу та масштабованість системи на великому наборі синтетичних даних. Результати підтвердили ефективність запропонованої системи з лямбда-архітектурою.

Метою роботи є аналіз існуючих методів виявлення аномалій в даних задля вирішення проблеми надмірного споживання електроенергії та попередження про критичні значення в показниках.

Основна частина

Виявлення аномалій – це метод розпізнавання даних, які відрізняються від звичайних. Аномалії в даних – це ситуації, які не відповідають визначеній звичайній моделі поведінки. Для того щоб проаналізувати та ідентифікувати аномалії треба сформувати набір даних за певний період часу (або використовувати вже готові дані). Також можливий аналіз даних в реальному часі. Аномалії, загалом, також відомі як викиди, девіанти, неузгодженості або винятки. Як правило, аномалії бувають наступних типів:

- точкові аномалії, які є найпростішим і дуже поширеним випадком. Точкові аномалії часто представляють екстремум, нерегулярність або відхилення, що трапляються випадковим чином і не мають особливого значення;

- контекстуальні аномалії, які часто виявляються в часових рядах та просторових даних. Це випадки, які можна розглядати як аномальні у якомусь конкретному контексті. Але варто зазначити, що спостереження однієї і тієї ж точки в різних контекстах не завжди дасть ознаки аномальної поведінки;

- колективні аномалії – це група корельованих, взаємопов'язаних або послідовних випадків, коли кожен конкретний екземпляр сам по собі не повинен бути аномальним, але їх колективне виникнення є аномальним.

Існують три основні категорії методів виявлення аномалій: контрольована, напівконтрольована та без нагляду.

Методи виявлення аномалій без нагляду (неконтрольовані алгоритми) визначають аномалії на непозначеному наборі даних, виходячи з припущення, що більшість зразків у цьому наборі є нормальними, і шукаючи зразки, що виглядають якнайменше відповідними решті набору даних.

Методи контрольованого виявлення аномалій вимагають набору даних, що позначено як «нормальні» або «аномальні», та включають навчання класифікатора (ключовою відмінністю від інших задач класифікації є притаманно незбалансований характер виявлення викидів).

Методи напівконтрольованого виявлення аномалій створюють модель, що представляє нормальну поведінку, виходячи із заданого нормального навчального набору даних, і потім перевіряють правдоподібність того, що тестовий екземпляр було породжено вивченою моделлю.

Метод Isolation Forest. Isolation Forest – це один із алгоритмів неконтрольованого машинного навчання, який використовується для виявлення аномалій у наборі даних [2]. На відміну від керованих алгоритмів машинного навчання, Isolation Forest не потребує жодних міток чи класифікації для даних, які потрібно проаналізувати. Алгоритм або відокремлює аномалії, розглядаючи аномалії як випадки, які менш імовірні, або приписує значення, які дуже відрізняються від зазвичай приписуваних. На рис. 1 можна побачити алгоритм Isolation Forest для визначення аномалій в даних щодо споживання електроенергії. Класифіковані значення 0 використовувалися для позначення нормального споживання енергії, тоді як 1 вказувало на аномальне споживання енергії.

Метод Clustering of K-means. K-Means – це один із алгоритмів, що використовується для кластеризації та розбиває дані на кілька груп. Алгоритм K-Means є одним із методів неієрархічної кластеризації даних, який може групувати дані в кілька кластерів на основі подібності даних [6]: дані з однаковими характеристиками в один кластер, а дані, які мають різні характеристики, групуються в інші кластери. Щоб визначити мітку кластера будь-яких даних, обчислюється відстань між даними з кожним кластером центру.

Є кілька способів, якими можна скористатися для обчислення відстані, наприклад Евклідова відстань, відстань Манхеттен і відстань Чебічі. Метод K-Means спрямований на мінімізацію суми квадратів відстаней між усіма точками та центром кластера.

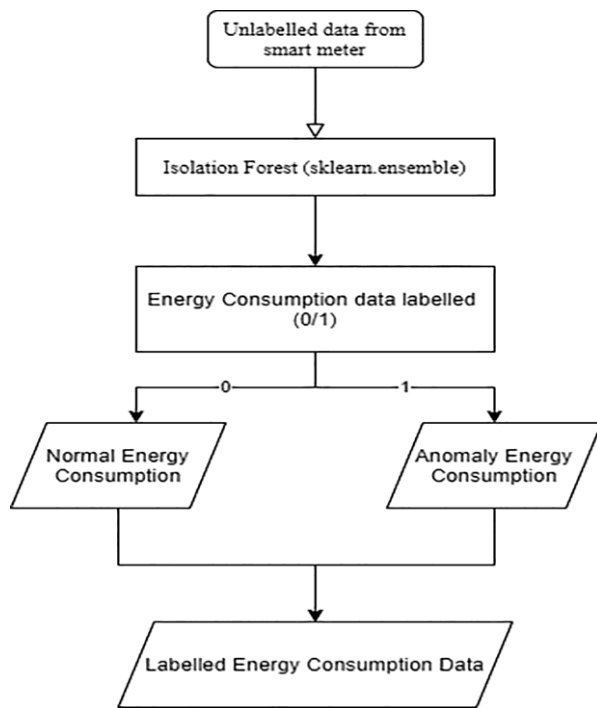


Рис. 1. Блок-схема маркування даних за алгоритмом Isolation Forest

Ця процедура складається з наступних кроків.

Крок 1: обираємо k із заданих n шаблонів як початкові центри кластерів. Призначаємо кожний шаблон, що залишився, до одного з k кластерів; шаблон призначається найближчому центру/кластеру.

Крок 2: обчислюємо центри кластерів на основі поточного призначення шаблонів.

Крок 3: відносимо кожен із n шаблонів до їх найближчого центру/кластеру.

Крок 4: якщо немає змін у призначенні шаблонів кластерам протягом двох послідовних ітерацій, завершуємо процедуру, інакше повертаємось до кроку 2.

В [6] автори взяли тестові дані з показниками споживання електроенергії одним підприємством на протязі 1 року і розділили на 4 частини (на кожен пору року). На цих даних і було проведено тренування та апробація результатів кластеризації. Для кожного сценарію було виміряно сумарну квадратичну помилку (SSE) та кількість ітерацій. SSE описує значення стандартного відхилення кожного кластера до центру обробки даних. Більше значення SSE означає, що ступінь подібності даних в одному кластері нижче. Кількість ітерацій описує довжину кластерів на протязі процесу формування. Результати кластеризації наведено в табл. 1.

Метод Support Vector Machine. Support vector machine (SVM) є керованим алгоритмом машинного навчання, який часто використовують для класифікації. SVM використовує гіперплощини в багатовимірному просторі, щоб розділити точки даних на класи. SVM зазвичай застосовується, коли в проблему залучено більше ніж один клас. Однак у виявленні аномалій він також використовується для проблем одного класу. Модель натренована визначати «норму» і може зрозуміти, чи належать незнайомі дані до цього класу, чи являють собою аномалію.

Таблиця 1 – Результати кластеризації

№	Сценарій	SSE	Кількість ітерацій
1	4 кластери без усунення аномалій	0.174	20
2	4 кластери з усуненням аномалій	0.752	14
3	5 кластерів без усунення аномалій	0.134	21
4	5 кластерів з усуненням аномалій	0.509	22

Багато факторів сприяли високій популярності SVM сьогодні. Наприклад: рішення розріджене, що робить його дійсно ефективним у порівнянні з іншими підходами на основі ядра. Також він може використовувати нелінійне перетворення в формі ядра, яке навіть дозволяє розглядати SVM як техніку зменшення розмірності [7]. Однокласовий SVM розроблений для випадків, коли відомий лише один клас і треба виявляти будь-що поза цим класом. Це відомо як визначення новизни і відноситься до автоматичної ідентифікації непередбачених або аномальних явищ, тобто викиди, вбудовані у велику кількість нормальних даних. На відміну від традиційного SVM, One-Class SVM (OC-SVM) дозволяє визначити межу, яка досягає максимальне розділення між зразками «відомий клас» та «походження». Лише невелика частина точок даних може лежати з іншого боку межі прийняття рішення: ці точки даних вважаються викидами (рис. 2).

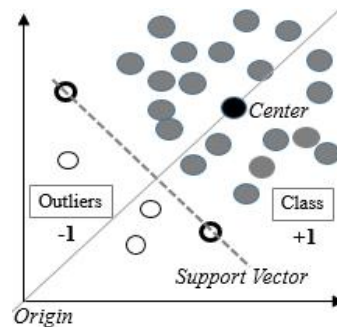


Рис. 2. Класифікація точок відносно опорного вектора класу

Метод Local Outlier Factor. Алгоритм LOF – це неконтрольований метод виявлення викидів, який оцінює унікальність кожної події на основі відстані від k -найближчих сусідів [8]. Алгоритм LOF здатний виявляти викиди незалежно від розподілу даних, оскільки він робить певні припущення щодо їх розподілу. Основна ідея алгоритму полягає в тому, що щільність навколо стороннього об'єкта суттєво відрізняється від щільності навколо своїх сусідів. Це перевага, коли дані, які аналізуються, не позначені або неможливо позначити через великий обсяг даних. Такий варіант поширено в комп'ютерних мережах де ряд генерованих мережних пакетів дуже високий.

На рис. 3 наведено схему обробки даних та виявлення аномалій. Перший крок складається з навчання NSL-KDD та розділення наборів даних на звичайні та атакуючі. Тільки для навчання використовуються записи, що відповідають звичайним даним. Далі, необхідно видалити атрибути 8 та 20 зі звичай-

ного набору даних, оскільки всі значення цих атрибутів дорівнюють 0, тобто вони не мають передбачуваної сили. Наступний крок це переведення даних на стандартизовану Z шкалу, де середнє значення буде дорівнювати 0, а стандартне відхилення 1. Це необхідно задля визначення допустимих значень та відхилень (аномалій). Даний процес застосовується до всіх числових значень і номінальних атрибутів (2 – тип протоколу, 3 – сервіс, 4 – прапор). Після обробки отримуємо 75 атрибутів. Перед застосуванням алгоритму LOF необхідно вказати два параметри: кількість найближчих сусідів k та граничне значення, щоб виявляти чи запис виходить за межі чи ні. При виборі кількість найближчих сусідів, рекомендується щоб k -значення дорівнюватиме квадратному кореню з усіх даних, які використано для модельного навчання. За алгоритмом LOF записи, які мають граничне значення відхилення більше 1, вважаються викидами. Для нормальної обробки даних існує шість порогових значень: Th_c (cleaning) = {1,5; 1,75; 2; 3; 5 та 10}.

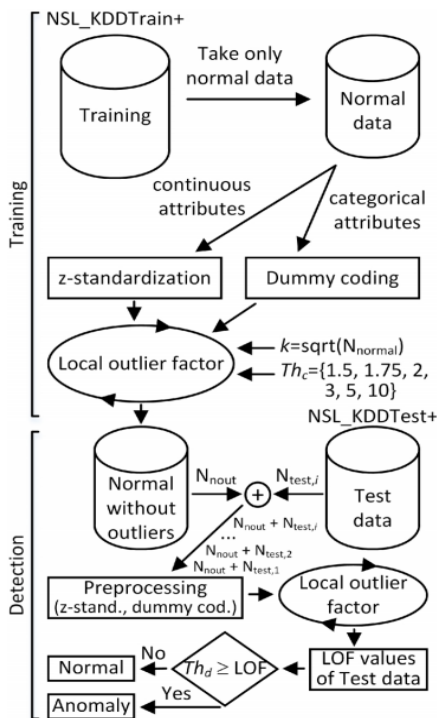


Рис. 3. Обробка даних та виявлення аномалій в LOF

Алгоритм LOF застосовується з обраною кількістю найближчих сусідів k та граничним значенням. На основі розрахованих значень, записи зі значеннями граничного відхилення, які вище або дорівнюють Th_c видаляються. Тоді k -значення для нормального типу даних, з якого були викиди видаляються, перераховуються, і знову виконується алгоритм LOF. Цикл повторюється до тих пір, поки не залишиться жодних записів, які будуть перевищувати встановлене граничне значення. Після навчання, отриманий набір даних далі використовується для виявлення аномалій.

Після фази навчання було підготовлено шість наборів даних, які складаються з даних нормального типу з видаленими аномаліями. Кількість викидів, знайдених та видалених за допомогою обраних

порогових значень наведено на рис. 4. Числа в дужках – це відсотки записів, які було видалено з набору навчальних даних.

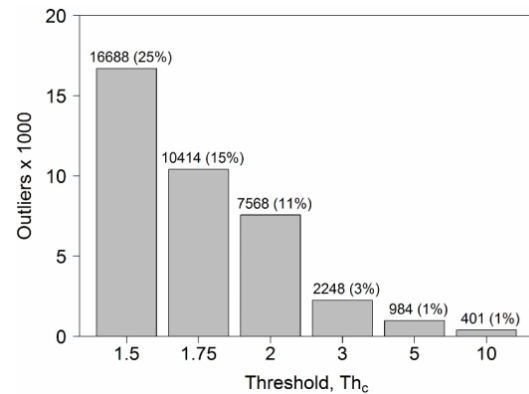


Рис. 4. Кількість аномалій, які було видалено з нормальних даних використовуючи значення Th_c

З рисунку видно, що перші три набори даних із пороговими значеннями 1,5, 1,75 та 2,0 повинні мати найбільший вплив на результати виявлення, оскільки ці набори даних зменшуються на 25, 15 та 11 відсотків відповідно.

Метод Unsupervised Niche Clustering. Unsupervised Niche Clustering (UNC) – це надійний метод кластеризації, який використовує еволюційний алгоритм зі стратегією заняття ніш [9]. Даний алгоритм допомагає знаходити кластери за допомогою стійкої функції пристосованості, в той час як техніка заняття ніш дозволяє створювати та підтримувати ніші (кластери-кандидати). Оскільки UNC базується на генетичній оптимізації, він набагато менш сприйнятливий до неоптимальних рішень, ніж традиційні методи. Основною перевагою алгоритму є здатність обробляти шум і автоматично визначати кількість кластерів. Автори статті [10] поєднали UNC з теорією нечітких множин для виявлення аномалій і застосували його для ідентифікації мережних вторгнень. Вони пов'язані з кожним кластером, згенерованим за допомогою UNC. Це функції-члени, які відповідають гаусовій формі, використовуючи еволюційний центр та радіус кластера. Такі функції приналежності кластерів визначатимуть рівень нормалізації вибірки даних.

Метод Regression Model-Based. Виявлення аномалій на основі регресійних моделей є підкатегорією параметричних методів [11], що включає низку методів, які широко застосовуються до даних часових рядів. Ці методи базуються на двоетапному підході. Спочатку на навчальних даних будується регресійна модель. Потім отримана модель використовується на тестових послідовностях для обчислення залишків, наприклад, різниці між прогнозованим значенням і реальним значенням. На основі залишків остаточно визначаються оцінки аномалій. До цієї категорії можна віднести методи виявлення аномалій, які засновані на традиційних моделях прогнозування часових рядів, таких як векторна авторегресія (VAR) [12] та авторегресійне інтегроване ковзне середнє (ARIMA).

Висновки

В результаті проведених авторами досліджень було зроблено аналіз популярних існуючих рішень для визначення аномалій в показниках щодо споживання електроенергії. При чому дослідження було проведено як з метою теоретичного ознайомлення з принципом роботи того чи іншого методу, так і експериментальною перевіркою тестових даних задля визначення коректності та точності виявлення аномалій. Звернута увага на обмеження даних методів обробки даних, а також на факти, висновки, рекомендації, закономірності з раніше відомих досліджень.

Після ретельного огляду кожного методу можна зробити висновок, що кожен з них підходить для вирішення поставленої задачі – потрібно лише розробити модель на основі обраного методу машинного навчання (або об'єднати декілька методів). Потім провести «тренування» на тестових даних, після чого

можна аналізувати вже реальні дані за певний проміжок часу про енергоспоживання за такими критеріями як напруга (В), сила струму (А), потужність (Вт), частота (Гц). Було б доцільно не тільки обробляти дані та визначати нетипові значення, а ще і формувати це у вигляді графіків задля наочного бачення аномальних значень. В ідеальному сценарії результатом досліджень є розробка моделі, яка буде обробляти ці дані в реальному часі.

З точки зору як теоретичної, так і прикладної перевірки цей аналіз має вагомий значення. Адже тема енергетики була, є і буде актуальною. Тим паче попередження про різкі показники напруги дозволять попередити вихід зі строю електроприладів та завчасно їх вимкнути. Або у випадку збору статистичних даних це дасть змогу зрозуміти, коли і як часто показники мали пікові значення, що дасть розуміння щодо подальшого обслуговування та експлуатації електрообладнання.

СПИСОК ЛІТЕРАТУРИ

- Berardi U. (2015), "Building energy consumption in US, EU, and BRIC countries", Department of Architectural Science, Faculty of Engineering and Architectural Science, № 118, P. 128-136.
- Jia Yan Lim, Wooi-Nee Tan, Yi-Fei Tan (2022), "Anomalous energy consumption detection using a Naïve Bayes approach", Faculty of Engineering, Multimedia University, Cyberjaya, Selangor, 63100, Malaysia, № 1, P. 4.
- Zhang J., Zhang H., Ding S., Zhang X. (2021), "Power Consumption Predicting and Anomaly Detection Based on Transformer and K-Means", College of Mathematics and Inf. Technology, Hebei University. Vol. 9, Article № 779587, P. 3-7.
- Longji Feng, Shu Xu, Linghao Zhang, Jing Wu, Jidong Zhang, Chengbo Chu, Zhenyu Wang and Haoyang Shi (2020), "Anomaly detection for electricity consumption in cloud computing: framework, methods, applications, and challenges", *Wireless Com Network*, № 194, P. 2-10, doi: <https://doi.org/10.1186/s13638-020-01807-0>
- Liu X., Iftikhar N., Nielsen P.S., Heller A. (2016), "Online Anomaly Energy Consumption Detection Using Lambda Architecture", *Big Data Analytics and Knowledge Discovery*, vol 9829, P. 193-209, doi: https://doi.org/10.1007/978-3-319-43946-4_13
- Yasirli Amril, Amanda Lailatul Fadhillah, Fatmawati, Novi Setianil, Septia Ranil (2016), "Analysis Clustering of Electricity Usage Profile Using K-Means Algorithm", *IOP Conf. Series*. № 105, P.2-7, doi: 10.1088/1757-899X/105/1/012020
- Lamrini B., Gjini A., Daudin S., Armando F., Pratomarty P., Travé-Massuyès L. (2018) "Anomaly Detection Using Similarity-based One-Class SVM for Network Traffic Characterization", *Un-té de Toulouse, CNRS, Toulouse, France.*, № 1, P. 2-4.
- Auskalnis J., Paulauskas N., Baskys A. (2018), "Application of Local Outlier Factor Algorithm to Detect Anomalies in Computer Network", *Elektronika I Elektrotechnika*, 24(3), P. 96-99, doi: <https://doi.org/10.5755/j01.eie.24.3.20972>
- Nasraoui O., Leon E., Krishnapuram R. (2005), "Unsupervised Niche Clustering: Discovering an Unknown Number of Clusters in Noisy Data Sets", *Evolutionary Computation in Data Mining*, Springer Berlin Heidelberg. Volume 1, P. 2-30.
- Lizabeth Leon., Olfa Nasraoui., Jonatan Gomez. (2007), "Anomaly detection based on unsupervised niche clustering with application to network intrusion detection", *Proc. of the IEEE Conference on Evolutionary Computation*. Volume 1, P. 502.
- Chandola V., Banerjee A., Kumar V. (2009), "Anomaly detection: A survey", *ACM Computing Surveys*, № 41, P. 1-58.
- Melnyk I., Matthews B., Valizadegan H., Banerjee A., Oza N. (2016), "Vector Autoregressive Model-Based Anomaly Detection in Aviation Systems", *JAIS Aerospace Information Systems.*, 13, P. 161-173.

Received (Надійшла) 22.12.2023

(Accepted for publication) Прийнята до друку 17.01.2024

Analysis of methods for detecting anomalies in electricity consumption data

Mykyta Bratyschenko, Tetiana Filimonchuk, Halyna Maistrenko, Vitalii Sitnikov

Abstract. Topicality. Modern people are constantly looking for new ways to use energy to improve their lives, so the demand for it is growing. In most cases, it is difficult for companies and industries to control all their devices at the same time, which can lead to a loss of electricity at any time. As a result, operating costs will be higher than necessary. In addition, the loss of power contributes to global warming through the release of carbon when energy is generated by burning coal, gas, and oil. Thus, solutions are needed to address these issues. **The purpose of this work** is to analyse existing methods for detecting anomalies in data to solve the problem of excessive electricity consumption and to warn of critical values in the indicators of electricity consumed by various devices and electrical equipment. **The object of the study** is the process of detecting atypical values or significant deviations in electricity consumption by such parameters as voltage, current strength and frequency, power. **The subject** of the study is models and methods for detecting anomalies in data. **Results.** After a thorough analysis of each of the above anomaly detection methods, new opportunities for solving the energy consumption problem open up. For example: combining several methods into one; developing a machine learning model based on one or more methods, training on test data and, in the future, processing real energy consumption data to identify atypical values, with the ability to record the date and time of anomalies, and build various graphs based on this information. **Conclusion.** The anomaly detection methods discussed here can prevent high electricity consumption to achieve energy savings, remind users to identify faulty electrical appliances or change incorrect electricity consumption patterns, reduce users' energy costs, and promote awareness of electricity safety.

Keywords: anomaly detection, energy consumption, statistical data, machine learning, energy efficiency, clustering, methods.