УДК 004.942:519.876 DOI: 10.20535/2523-4455.2018.23.4.143310

Виявлення аномальної поведінки людини у MicroGrid на базі машинного навчання

Комаревич О. М., ORCID 0000-0002-0206-3393
e-mail komarevich.alex@gmail.com
Хохлов Ю. В., к.т.н. доц., ORCID 0000-0002-2034-6979
e-mail ykhokhlov@gmail.com
Ямненко Ю. С., д.т.н. проф., ORCID 0000-0002-9796-6420
e-mail petergerya@yahoo.com
Кафедра промислової електроніки
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського» kpi.ua
Київ, Україна

Реферат—У застосуванні до системи розподіленої генерації MicroGrid розглянуто задачу виявлення аномалій поведінки користувача. Задача вирішується із залученням методів машинного навчання (Machine Learning), зокрема, методу детектування аномалій (Anomaly Detection). В якості ключових параметрів для найпростішого випадку задачі виявлення аномальної поведінки розглянуто усереднене електроспоживання за п'ятихвилинні проміжки часу, а також кількість спрацювань датчика руху, встановленого у приміщенні MicroGrid.

Бібл. 12, рис. 4, табл. 2.

Ключові слова — MicroGrid; Anomaly Detection; Machine Learning.

I. Вступ

Розвиток електронних систем та інтелектуальних засобів керування та обміну інформацією призвів до формування нової інформаційно-енергетичної концепції SmartGrid [1], [2], в рамках якої розглядаються локальні електротехнічні об'єкти – MicroGrid, що містять певний набір джерел та навантажень, які, як правило, підключені до централізованої електромережі. В рамках MicroGrid постає цілий ряд задач керування та узгодженого функціонування пристроїв, що призводить до необхідності оперування великими інформаційними потоками [3], [4].

У порівнянні з повністю автоматичними системами, MicroGrid характеризується присутністю людини, що обумовлює присутність та вплив суб'єктивних факторів – втручання в процес функціонування електротехнічного обладнання та вплив дій людини на робочі режими всієї системи. Наявність людського фактору додатково збільшує обсяг даних, що підлягають аналізу та обробці. У свою чергу це змушує дослідників звертатися до спеціалізованих методів роботи з великими даними (Big Data) [4]. Велика кількість та різноманітність цих методів включає підходи, інструменти та методи обробки, ефективні в умовах безперервного зростання обсягів даних, та отримання результатів, придатних для сприйняття людиною. Важливим та ефективним інструментом обробки великих обсягів даних є методи машинного навчання (Machine Learning) [5], [6]. Як галузь інформатики, Machine Learning використовує статистичні методи для забезпечення навчання комп'ютерних систем із поступовим покращенням продуктивності вирішення конкретних задач без явного програмування.

До аномальних режимів у MicroGrid можна віднести:

1) Аварійні випадки.
2) Вихід технічних параметрів за допустимі межі.
3) Незвична активність, нетипова поведінка людини – «людський фактор».

Оскільки перші два випадки відносяться до суто технічних, то їх відпрацювання здійснюється за рахунок конструкторських особливостей інженерної розробки електротехнічного комплексу і здійснюється відомими методами [1], [2]. Що стосується третього випадку, то саме він є характерним для MicroGrid і являє собою найбільший інтерес з точки зору застосування методів машинного навчання, оскільки наявність «людського фактору» веде до великої кількості можливих сценаріїв, які підлягають обробці та подальшому використанню в якості навчальної вибірки. При цьому традиційні методи обробки можуть виявитися нездатними обробити настільки великі обсяги даних, що це дозволяє віднести їх до категорії Big Data.

Дослідження в напрямку виявлення нетипових (аномальних) поведінкових моделей людини є актуальними для систем психофізіологічного моніторингу літніх людей, осіб, що потребують постійного нагляду та піклування, перебувають у постстресових

станах або знаходяться під впливом надзвичайних чи екстремальних ситуацій [7].

Серед сукупності методів машинного навчання найбільш придатним є метод детектування аномалій (Anomaly Detection) [6], [8]. В статті розглядається задача застосування Anomaly Detection як одного із засобів Machine Learning для визначення нетипових моделей поведінки людини (аномальної поведінки) у MicroGrid.

II. Оцінка обсягу даних

Інформаційна інфраструктура сучасних комплексів MicroGrid оснащена величезною кількістю датчиків, що фіксують різні типи подій та утворюють інформаційну картину функціонування технічних пристроїв і підсистем, а також переміщення та дії людини. Кількість спрацьовувань n двопозиційних датчиків (що фіксують бінарні події типу «вкл/викл») можна розрахувати за формулою:

$$N = 2^n \sum_{m=1}^{n} C_n^m + 1,$$

де множник 2 визначає спрацювання у двох режимах, $C_n^m = \frac{n(n-1)\ldots\left[n-(m-1)\right]}{m!}$ - кількість сполучень з n елементів по i. Наприклад, при кількості датчиків n = 24 кількість різних інформаційних станів у базі даних може досягати десятків тисяч. При аномальній поведінці людини спрацьовувань датчиків може бути ще більше. Тому задача обробки та аналізу даних з усіх датчиків ускладнюється, а її розв'язок традиційними методами стає неможливим через необхідність оперувати з Big Data. Саме це обумовлює залучення методів машинного навчання для вирішення поставленої задачі.

III. Anomaly Detection

В аналізі даних методами машинного навчання є два напрямки, які займаються пошуком аномалій: «Детектування викидів» (Outlier Detection) і «Детектування новизни» (Novelty Detection) [6], [8]. У першому випадку детектується вихід значень досліджуваних параметрів за межі області, яка в результаті аналізу була встановлена як «нормальна». У другому випадку реєструється поява «нового об'єкту» - відсутнього у наявній базі даних, але не аномального.

Як «викид», так і «новий об'єкт» відрізняється від об'єктів наявної навчальної вибірки. Проте на відміну від викиду, у майбутньому цей об'єкт повинен бути включеним до навчальної вибірки. Таким чином, межі «області нормальності» динамічно змінюються, вміщуючи виявлені нові об'єкти.

Наприклад, при аналізі виміряних значень температури навколишнього середовища відкидаються аномально великі або маленькі значення, що відноситься до детектування викидів. Якщо ж алгоритм аналізу передбачає оцінку кожного нового значення у термінах «схожості» на наявні значення у навчальній вибірці – це детектування новизни.

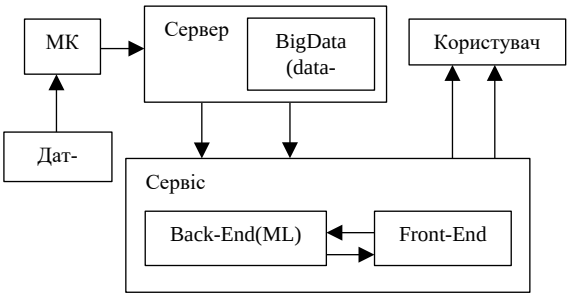

Рис. 1 Схема програмно-апаратного модуля детектування аномалій

В термінах задачі виявлення аномальної поведінки людини важливим є детектування як викидів значень на координатній площині параметрів поведінки, так і детектування новизни – поява нових поведінкових шаблонів, що є відмінними від наявних у навчальній вибірці та підлягають обробці як новий варіант «норми».

Викиди виникають внаслідок:

- помилок у даних (неточності вимірювання, округлення, невірного запису);
- наявності завад, що спричиняють невірну класифікацію об'єктів, тобто їх помилкове віднесення до певних груп;
- присутності об'єктів «сторонніх» вибірок (наприклад, даних з датчика, що вийшов з ладу).

Суб'єктивні фактори, в свою чергу, здатні обумовити появу викидів внаслідок помилкової інтерпретації або несанкціонованого втручання людини у роботу технічного обладнання.

При аналізі поведінкових характеристик постає задача вибору з усієї сукупності параметрів, що описують поведінку та переміщення людини, саме тих вирішальних параметрів, на підставі яких можна детектувати викиди або новизну поведінки.

У найпростішому випадку одним з таких вирішальних параметрів може виступати кількість спрацьовувань датчика руху в одному і тому самому приміщенні протягом певного невеликого інтервалу часу τ (наприклад, τ = 5 хв), а другим – усереднене значення енергії споживання за цей самий інтервал. Такий підхід дозволить виявити випадки швидкого «безсистемного» переміщення людини, що може бути наслідком знаходження у стані афекту, наляканості чи психологічного розладу, адже однакова кількість спрацювань одного датчику протягом 5 хвилин розглядається як «аномалія», а протягом години – як «норма».

Незвичну активність людини можна детектувати за допомогою аналізу даних з датчиків (рис. 1), які підключаються до мікроконтролера (МК), що надсилає інформацію на сервер. В свою чергу, сервер передає дані у сервіс обробки та аналізу, а результат обробки відображується користувачу.

Сервіс складається з блоку «Back-end (ML)», в якому відбувається власне машинне навчання —

обробка даних, формування навчальної вибірки та подальший аналіз, результат якого надходить до блоку «Front-end». Блок «Front-end» відповідає за відображення результатів машинного навчання користувачеві.

Слід зазначити, що весь процес проходить без втручання користувача, якому надається можливість лише переглядати виявлені викиди – аномалії.

Послідовність дій при детектуванні аномалій поведінки людини у MicroGrid містить три основних етапи:

1) Вимірювання і накопичення даних (Big Data).
2) Аналіз даних – виявлення «областей нормальності».
3) Застосування результатів навчання для аналізу поточної активності.

На першому етапі відбувається збір даних для моделювання та адаптації інформаційного середовища, що досліджується. Спрацювання відповідних датчиків розцінюються як події, фіксуються у часі та заносяться у базу даних. Історія спостережуваних подій датчиків відображає дії, які відбуваються в навколишньому середовищі і можуть бути використані для виявлення часто повторюваних моделей поведінкової активності. При цьому визначаються різні варіанти активності повсякденного життя з урахуванням зовнішніх факторів (час доби, сезонність, вихідні/святкові дні, тощо), а також виявляються аномальні стани.

Формат даних, занесених у таблицю історії подій, містить 5 полів: дата, час, сенсор, нотаток і стан. Приклад фрагменту такої історії подій наведено у табл. 1, де відображено події відкривання та закривання дверей (Open_Door), а також спрацювання датчиків, що мають умовні позначення у загальній специфікації MicroGrid – L5643 та D556K.

На рис. 2 зображено логічну схему послідовності дій машинного навчання на другому (аналіз та навчання) та третьому (тестування поточного режиму) етапах.

Блок «парсер» як інструмент синтаксичного аналізу перетворює вхідні дані в структурований формат з введенням додаткових атрибутів для підвищення ефективності обробки набору даних.

Робота класифікатора SVM (Support Vector Machine) контролюється заданими моделями та алгоритмами навчання, які аналізують тренувальні та тестові дані, що використовуються для класифікації та регресійного аналізу. Враховуючи набір навчальних прикладів, кожен з яких позначається як такий, що належить до однієї з двох категорій, SVM-модель являє собою представлення тестового прикладу як точки в просторі, мапованому таким чином, що приклади окремих категорій діляться явним розривом [6]. Нові приклади потім вносяться у той самий простір і беруть участь у подальшому аналізі вже як елементи навчальної вибірки.

ТАБЛИЦЯ 1 ІСТОРІЯ ПОДІЙ

Дата	Час	Сенсор	Статус	Прим.	Стан
06.06	15:55:20	K154M	OPEN	Open_ Door	begin
06.06	15:57:10	L5643	ON		
06.06	15:57:40	L5643	OFF		
06.06	15:58:10	D556K	ON		
06.06	15:58:35	K154M	CLOSE	Open_ Door	end
06.06	15:58:50	D556K	OFF		

ТАБЛИЦЯ 2 ФІКСАЦІЯ ВИРІШАЛЬНИХ ОЗНАК

№	Час	W, Вт	Кількість, N
1	00:00-00:05	20	**50**
2	00:05-00:10	30	40
3	00:10-00:15	25	35
4	00:15-00:20	27	40
…	…	…	…
127	10:30-10:35	20	**10**
128	10:35-10:40	25	30
…	…	…	…

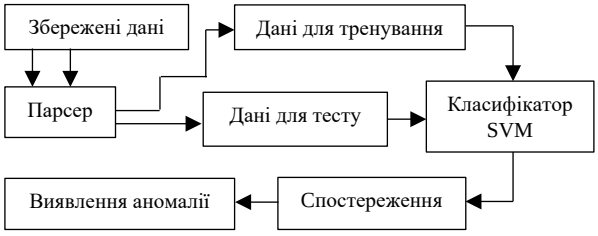

Рис. 2 Схема послідовності дій машинного навчання

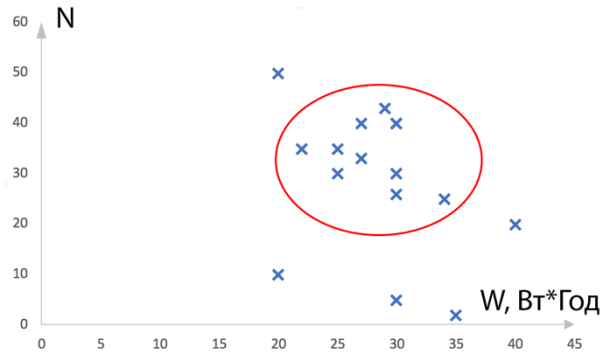

Рис. 3 Визначення «зони нормальності»

За допомогою алгоритму Anomaly Detection [8] сервіс виявляє аномалії і попереджає про них користувача. Оскільки мікроконтролер передає дані через інтерфейс Rest API, доцільно використовувати Web-інтерфейс в якості Front-end частини [9].

Для моделювання найпростішого варіанту аналізу поведінкових моделей активності людини за допомогою методу Anomaly Detection в якості вирішальних параметрів було обрано: **Ознака 1** - кількість N спрацьовувань датчика руху протягом кожного з послідовних періодів тривалістю 5 хвилин, **Ознака 2 –** усереднене електроспоживання *W* за цей період. Значення цих ознак за період спостереження утворюють зразкову матрицю – множину двоелементних векторів ($K = 2$ – кількість ознак, що беруться до розгляду).

Приклад накопичених даних значень цих ознак наведено у табл. 2.

Отримані результати наносяться на координатну площину параметрів *N* та *W*. У подальшому графічно-аналітичними методами Anomaly Detection визначається «зона нормальності» (рис. 3).

Зазначимо, що для формування цієї зони використовуються різні підходи [6], [8], [10], зокрема, з урахуванням експертних оцінок та даних попередніх періодів спостережень. Наявні на графіку відхилення вважаються аномаліями і потребують втручання інших сервісів інформаційно-програмного забезпечення MicroGrid для подальшого відпрацювання – активації тривоги, сповіщення користувача, тощо.

Виявлення меж «зони нормальності» може бути полегшено за рахунок попередньої кластеризації [11]. Тоді кластери суттєво меншого розміру, швидше за все, відносяться до аномалій.

У реальній системі MicroGrid кількість точок спостереження, що складають навчальну вибірку, і відповідно, кількість рядків у табл. 2 буде збільшуватися. Постійне накопичення даних дозволяє збільшувати обсяг навчальної вибірки та відповідно динамічно змінювати «зону нормальності».

Важливим питанням при аналізі великих даних та визначенні «зони нормальності» є вимірювання відстані між точками в координатній площині вирішальних параметрів та від точки до встановленої області. Для цього використовуються кореляційні та метричні методи [10], [12], що дозволяють визначити сусідні точки та віднести (або не віднести) їх до аномалій в залежності від відстані – метрики (рис. 4).

Інтуїтивно зрозуміло, що «викид» має мало «сусідів» (точок, що знаходяться на малій відстані від нього), а типова точка, що належить «зоні нормальності» - багато. Евклідова відстань як проста метрика може бути застосована лише у випадку порівняння тестового прикладу з векторами навчальної вибірки попарно. Враховуючи велику кількість даних, що підлягає аналізу, доцільніше використовувати відстань Махалонобіса [12], яка дозволяє порівнювати відстань від одного вектору до центру тяжіння множини векторів навчальної вибірки:

$$D_{mahal}(X) = \sqrt{(X-\mu)^T COV^{-1}(X-\mu)},$$

де X – тестовий вектор даних, $\mu = (\mu_1, \mu_2, \dots \mu_K)$ – вектор середніх значень ознак векторів даних навчальної вибірки, COV – коваріаційна матриця.

Таким чином, відстань Махалонобіса вказує на те, чи є тестовий вектор викидом відносно номінального набору векторів навчальної вибірки.

Слід зазначити, що отриманий розподіл точок всередині та зовні «зони нормальності» не є точним, безперечним та статичним, а підлягає постійному уточненню по мірі накопичення даних і тестових прикладів.

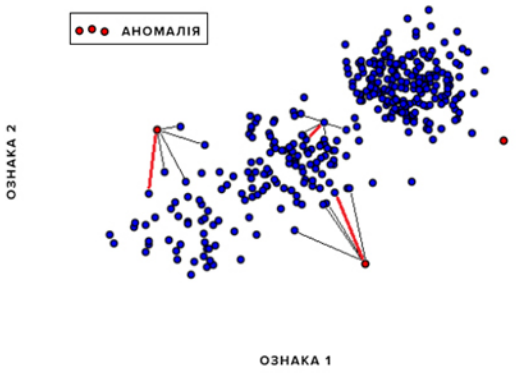

Рис. 4 Точки аномалій

Висновки

Таким чином, актуальна задача виявлення аномалій моделей поведінки людини як користувача сервісів MicroGrid вирішується із залученням методів машинного навчання, зокрема, Anomaly Detection. Це дозволяє своєчасно реагувати на виявлені випадки аномальної поведінки, залучаючи відповідні сервіси інформаційно-керуючого забезпечення MicroGrid.

Надзвичайно важливими аспектами цієї задачі є вибір «вирішальних параметрів», що беруться до уваги при аналізі, а також формування «зони нормальності», межі якої в подальшому динамічно змінюються з урахуванням поповнення навчальної вибірки новими значеннями.

Перелік посилань

[1] T. M. Baziuk, I. V. Blinovv, O. F. Butkevich, I. S. Honcharenko, S. P. Denysiuk, and et. al., *Intelektual'ni elektrychni merezhi: elementy ta rezhymy [Intelligent Electric Networks: Elements and Modes]*. Kyiv, Ukraine: The National Academy of Sciences of Ukraine The Institute of Electrodynamics, 2016, **ISBN:** 978-966-02-7913-1.

[2] D. Q. Oliveira, A. C. Zambroni de Souza, A. B. Almeida, M. V. Santos, B. I. L. Lopes, and D. Marujo, “Microgrid management in emergency scenarios for smart electrical energy usage,” in *2015 IEEE Eindhoven PowerTech*, 2015, pp. 1–6, **DOI:** 10.1109/PTC.2015.7232309.

[3] J. Yamnenko, T. Tereshchenko, L. Klepach, and D. Palii, “Forecasting of electricity consumption in SmartGrid,” in *2017 International Conference on Modern Electrical and Energy Systems (MEES)*, 2017, pp. 208–211, **DOI:** 10.1109/MEES.2017.8248891.

[4] S. Singh and N. Singh, “Big Data analytics,” in *2012 International Conference on Communication, Information & Computing Technology (ICCICT)*, 2012, pp. 1–4, **DOI:** 10.1109/ICCICT.2012.6398180.

[5] A. Gulenko, M. Wallschlager, F. Schmidt, O. Kao, and F. Liu, “Evaluating machine learning algorithms for anomaly detection in clouds,” in *2016 IEEE International Conference on Big Data (Big Data)*, 2016, pp. 2716–2721, **DOI:** 10.1109/BigData.2016.7840917.

[6] Xueqin Zhang, Chunhua Gu, and Jiajun Lin, “Support Vector Machines for Anomaly Detection,” in *2006 6th World Congress on Intelligent Control and Automation*, 2006, pp. 2594–2598, **DOI:** 10.1109/WCICA.2006.1712831.

[7] Y. S. Yamnenko and T. O. Tereshchenko, “Spectral methods for processing biotelemetrical data,” *Electron. Commun.*, vol. 21, no. 4, pp. 38–43, Nov. 2016, **DOI:** 10.20535/2312-1807.2016.21.4.81904.

[8] X. Song, M. Wu, C. Jermaine, and S. Ranka, “Conditional Anomaly Detection,” *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 5, pp. 631–645, 2007, **DOI:** 10.1109/TKDE.2007.1009.

[9] H. M. Abdullah and A. M. Zeki, “Frontend and Backend Web Technologies in Social Networking Sites: Facebook as an Example,” in *2014 3rd International Conference on Advanced Computer Science Applications and Technologies*, 2014, pp. 85–89, **DOI:** 10.1109/ACSAT.2014.22.

[10] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF,” in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD ’00*, 2000, pp. 93–104, **DOI:** 10.1145/342009.335388.

[11] P. S. Badase, G. P. Deshbhratar, and A. P. Bhagat, “Classification and analysis of clustering algorithms for large datasets,” in *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, 2015, pp. 1–5, **DOI:** 10.1109/ICIIECS.2015.7193191.

[12] R. Lin, E. Khalastchi, and G. A. Kaminka, “Detecting anomalies in unmanned vehicles using the Mahalanobis distance,” in *2010 IEEE International Conference on Robotics and Automation*, 2010, pp. 3038–3044, **DOI:** 10.1109/ROBOT.2010.5509781.

Надійшла до редакції 28 липня 2018 р.

УДК 004.942:519.876

Выявление аномального поведения человека в MicroGrid на базе машинного обучения

Комаревич А. Н., ORCID 0000-0002-0206-3393
e-mail komarevich.alex@gmail.com
Хохлов Ю. В., к.т.н. доц., ORCID 0000-0002-2034-6979
e-mail ykhokhlov@gmail.com
Ямненко Ю. С., д.т.н. проф., ORCID 0000-0002-9796-6420
e-mail petergerya@yahoo.com
Кафедра промышленной электроники
Национальный технический университет Украины
«Киевский политехнический институт имени Игоря Сикорского» kpi.ua
Киев, Украина

***Реферат*—В применении к системе распределенной генерации MicroGrid рассмотрена задача выявления аномалий поведения пользователя. Задача решается с привлечением методов машинного обучения (Machine Learning), в частности, метода детектирования аномалий (Anomaly Detection). В качестве ключевых параметров для простейшего случая задачи обнаружения аномального поведения рассмотрено усредненное электропотребление за пятиминутные промежутки времени, а также количество срабатываний датчика движения, установленного в помещении MicroGrid.**

Библ. 12, рис. 4, табл. 2.

Ключевые слова — MicroGrid; Anomaly Detection; Machine Learning

UDC 004.942:519.876

Detecting Abnormal Human Behavior in Microgrid Based on Machine Learning

O. M. Komarevych, ORCID 0000-0002-0206-3393
e-mail komarevich.alex@gmail.com
Yu. V. Khokhlov, PhD Assoc.Prof., ORCID 0000-0002-2034-6979
e-mail ykhokhlov@gmail.com
Yu. S. Yamnenko, Dr.Sc.(Eng.) Prof., ORCID 0000-0002-9796-6420
e-mail petergerya@yahoo.com

Department of Industrial Electronics
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” kpi.ua
Kyiv, Ukraine

***Abstract*—Development of electronic systems and intellectual controls and information exchange has led to the formation of a new informatics and energy concept SmartGrid, which addresses the local electrical engineering objects - MicroGrid. MicroGrid contains a certain set of sources and loads that are typically connected to a centralized power grid. As part of MicroGrid, there are a number of management and device-specific tasks that result in the need to handle large information flows. Compared to fully automated systems, Microgrid is characterized by the presence of a person, which leads to the need to take into account subjective factors - the presence of a person, its interference in the functioning of electrical equipment and the impact of his actions on the working modes of the whole system. The presence of a human factor leads to the need to operate large volumes of data. In turn, it forces researchers to turn to specialized methods of working with large data (Big Data). Big Data is a series of approaches, tools and methods for processing structured and unstructured data of large volumes to produce results that are suitable for human perception and effective in the conditions of their continuous increase. To solve the problem of processing large amounts of data, it is necessary to use methods of machine learning (Machine Learning). Machine Learning as a branch of informatics uses statistical techniques to give computer systems the ability to "learn" (that is, to gradually improve performance in a particular task) without explicit programming. In this paper, Anomaly Detection was considered as one of the Machine Learning tools for defining anomalies in MicroGrid. One of the regime in MicroGrid that was considered as anomaly is non-usual activity, non-typical behavior of the person – “human factor”. Research directed to revealing of non-typical (abnormal) behavioral models is topical for the systems of psycho-physiological monitoring of elderly people, persons who need constant observation and care, persons in post-stress condition or under the extraordinary conditions. Subjective factors, in turn, can cause the appearance of emissions as a result of false interpretation or unauthorized human intervention in the work of technical equipment. In the analysis of behavioral characteristics, the task of choosing from the whole set of parameters describing the behavior and movement of a person, the very decisive parameters on the basis of which it is possible to detect emissions or novelty of behavior. In the application to distributed generation MicroGrid, the problem of detecting user behavior abnormalities is considered. The problem is solved by the use of Machine Learning methods, in particular, the Anomaly Detection method. As the key parameters for the simplest case of detecting abnormal behavior, we consider the average power consumption at five-minute intervals, as well as the number of triggers of the motion sensor installed in the MicroGrid indoor space.**

Ref. 12, fig. 4, tabl. 2.

Keywords — MicroGrid; Anomaly Detection; Machine Learning