

Класифікація методів машинного навчання у MicroGrid

Крилов А. В., ORCID [0000-0002-7203-300X](https://orcid.org/0000-0002-7203-300X)

Кафедра електронних пристроїв та систем

Факультет електроніки

Національний технічний університет України

«Київський політехнічний інститут імені Ігоря Сікорського»

Київ, Україна

Анотація—У статті розглянуто класифікацію й застосування методів машинного навчання в системі розподіленої генерації енергії MicroGrid. Проведено порівняльний аналіз існуючих методів машинного навчання та їх відповідність задачам керування у MicroGrid. Проведено розробку програмної складової реалізації дерева прийняття рішень для розв'язання задачі класифікації даних машинного навчання.

Ключові слова — MicroGrid; машинне навчання; дерево прийняття рішень.

I. ВСТУП

Важливим напрямком розвитку сучасних електротехнічних систем з розподіленою генерацією електричної енергії є впровадження концепції Інтернету речей [1]. Зокрема, для об'єктів типу «розумний будинок» в рамках цієї концепції розглядаються процеси, що забезпечують виконання наступних функцій керування та регулювання:

- керування освітленням;
- керування параметрами мікроклімату;
- керування споживанням електричної енергії;
- моніторинг біотелеметричних показників.

Звичайно, цей перелік є далеко не повним, до того ж спектр задач керування залежить від функціонального призначення та особливостей системи [2].

Тож надалі термін «розумний будинок» замінимо на MicroGrid. Так як параметри об'єкту, що розглядається, притаманні більш системі розподіленої генерації енергії. А саме не тільки керування кінцевими приладами, а й аналіз даних з датчиків, лічильників та інших приладів.

Що більшим є коло задач, кількість параметрів, які підлягають вимірюванню, аналізу та обробці в системі прийняття рішень, то більш актуальним є залучення складних алгоритмів, серед яких важливе місце займають методи штучного інтелекту та машинного навчання [2]. Про актуальність розробки та подальшого впровадження таких методів свідчить темп зростання продажів смартфонів та інших гаджетів, які поєднують в собі дедалі більшу кількість датчиків і дають великі можливості для вивчення, отримання і аналізу величезної кількості різноманітних даних.

II. СУЧАСНИЙ СТАН ДОСЛІДЖЕНЬ ТА АНАЛІЗ ПУБЛІКАЦІЙ

Сучасні системи керування електроспоживанням використовують автоматизоване керування виконавчими пристроями з метою забезпечення автоматизації процесів прийняття рішень. При цьому на перший план виходять потреби кінцевого споживача, які мають задовольнятися в першу чергу, забезпечуючи пріоритетність рівня «комфортності» або задоволеності людини із збереженням функцій забезпечення її життєдіяльності та виконання інших робочих процесів у системі. Однак присутність людини як споживача послуг системи «розумного будинку» обумовлює в першу чергу складність побудови алгоритмів прийняття рішень внаслідок суб'єктивності та непередбачуваності її поведінки. Саме тому методи, які в тій чи іншій мірі наближають поведінку технічної системи до поведінки, притаманній людині, мають величезну перспективу.

Характерні особливості людини як користувача системи вимагають здійснення попереднього налаштування та адаптації системи під ці особливості. Тому важливим етапом є навчання, яке здійснюється на попередньому етапі. Саме тут на допомогу приходять методи машинного навчання.

У машинному навчанні зустрічаються дві проблеми - це недонавчання і перенавчання алгоритмів. Інтуїтивно зрозуміло, що ці проблеми можуть виникати через недостатність або навпаки - надлишковість даних, пов'язану з кількістю і зашумленістю даних. В. Вапніку вдалося математично точно сформулювати зазначені проблеми і дати оцінки складності алгоритмів навчання, правда, в той час ці результати виявилось важко застосувати на практиці.

Теорія статистичного навчання вводить поняття розмірності Вапніка-Червоненкіса (VC-розмірність), яке універсальним чином відображає складність будь-якого алгоритму для певного набору даних.



Також в даній теорії вводиться принцип структурної мінімізації ризику (SRM – Structured Risk Minimization), при якому навчання відбувається в два етапи. На першому етапі обирається підходяща складність моделі з використанням VC-розмірності, а на другому – сам процес навчання.

З тих пір і до сьогоднішнього дня всі методи ґрунтуються на засадах, що запровадив Вапнік [3].

На разі машинне навчання застосовується в морському [4] й наземному транспорті. Активно машинне навчання застосовується в кібер безпеці для розпізнавання обличч. Але контроль за комп'ютерними системами на сьогоднішній день не є досить розвинутою темою, тому це є дуже актуальною [5].

Метою статті є дослідження складних інтелектуальних методів прийняття рішень у керуванні електротехнічними системами з розподіленою генерацією MicroGrid, а також розробка методу класифікації робочих режимів MicroGrid [6].

III. МЕТОДИ МАШИННОГО НАВЧАННЯ

У методах машинного навчання надзвичайно важливим фактором є якість власне навчання системи на накопичених тестових вибірках. Ця якість залежить не тільки від характеристики вибірки (повнота, зкомпенсованість), але й від методу навчання [2]. Рис. 1 ілюструє класифікацію методів навчання за призначенням.

Таблиця 1 МЕТОДИ МАШИННОГО НАВЧАННЯ ТА ЇХ ВІДПОВІДНІСТЬ ЗАДАЧАМ КЕРУВАННЯ У MICROGRID

Метод Machine Learning	Задача керування	Шлях розв'язання	Примітка
Лінійний дискримінантний аналіз	Класифікація /кластеризація робочих режимів та ситуацій, що вимагають прийняття рішень.	Обчислення дискримінантного значення для всіх класів; вибір класу з найбільшим значенням.	Передбачається, що дані мають нормальний розподіл, тому необхідна процедура попереднього видалення аномалій
Дерева прийняття рішень	Прогнозування режимів	Пошук «листових вузлів» - змінних, які використовуються для прогнозування	Рішення формуються у вузлах системи, на певних стадіях обробки.
Мережі векторного квантування	Пошук схожих режимів та ситуацій з метою передбачення за зразком «найближчого сусіда»	Формування набору кодових векторів (LVQ – Learning Vector Quantization)	Кодові вектори обираються на початку випадковим чином і в подальшому адаптуються так, щоб найкращим чином узагальнити весь набір даних

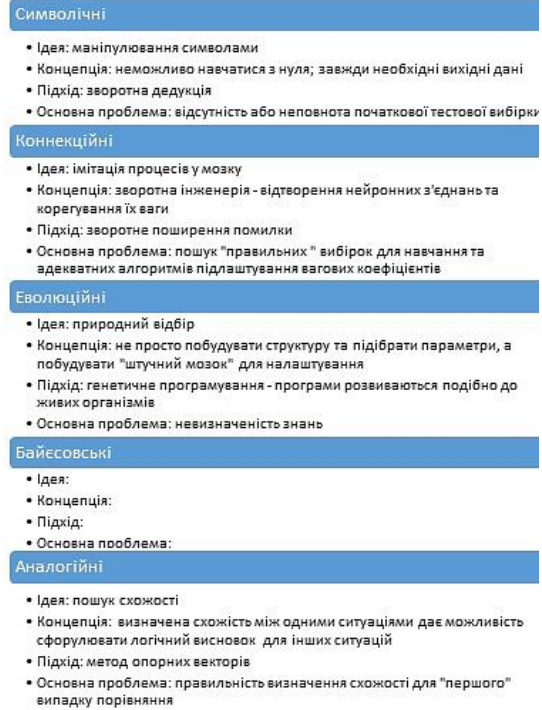


Рис. 1 Класифікація методів навчання[8].

Відповідно до особливостей MicroGrid, а також особливостей знаходження та активності в ньому людини встановимо певні відповідності між методами машинного навчання та задачами керування (див. Таблиця 1) [7]

Надалі в роботі обирається метод «Дерева прийняття рішень». Так як цей метод є доцільним для використання в системах з багатьма точками прийняття рішень.

IV. РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

Об'єктом програми є система розподіленої генерації енергії MicroGrid, а дані в системі являються змінними значеннями, які треба класифікувати. Дані являють собою інформацією зчитаною з датчиків. За представленими даними з датчиків визначається, до якого типу значення відноситься кластер даних. Спочатку дані треба систематизувати у наступному форматі:

```
path = "%шлях до файлу%/BigData.csv"
data = read(path, delimiter=",")
data.head()
```

Результатом виконання програми є класифікація даних з великої бази даних. На підставі цього виконується подальший аналіз. Він полягає у відсортуванні даних з датчиків системи (змінних даних) від зайвої інформації мережі(статичних даних), за результатами якого у системі приймаються рішення для вироблення керуючих сигналів виконавчих пристроїв, що безпосередньо змінюють робочі режими відповідних підсистем MicroGrid. У таблиці даних, у першому стовпці значення Par показують, до якого параметру



належать дані, а інші стовпці - ознаки, за якими їх можна розрізнити [8].

Далі вирішується основна задача – задача класифікації.

Нижче наведено частину програмного коду, створеного за допомогою Python® [8], де використано дані які згенеровані машиною завчасно.

Етапи вирішення задачі класифікації:

- Створення початкової навчальної вибірки;
- Навчання з випадково обраними початковими значеннями параметрів і відповідних їм класів.
- Оцінка якості навчання та готовності працювати з тестовими вибірками.

```
X = data.values[:, 1:14]
```

```
y = data.values[:, 0:1]
```

```
from sklearn.cross_validation import train_test_split
as train
```

```
X_train, X_test, y_train, y_test = train(X, y,
test_size=0.6)
```

```
from sklearn.ensemble import
RandomForestClassifier
```

```
clf = RandomForestClassifier(n_estimators=100,
n_jobs=-1)
```

```
clf.fit(X_train, y_train)
```

```
clf.score(X_test, y_test)
```

Створюються масиви, де X – інформація з датчиків, y – класи датчиків. Потім, щоб зібрати тестову і навчальну вибірку з вихідних даних, скористаємося зручною функцією крос-валідації `train_test_split`, реалізованої в програмному середовищі Scikit-learn [9]. З готових вибірок імпортуємо `RandomForestClassifier` з `ensemble` в Scikit-learn. Цей клас містить у собі всі необхідні для навчання і тестування методи і функції. Надаємо змінній `clf` (classifier) клас `RandomForestClassifier`, потім викликом функції `fit()` запускаємо процес навчання з класу `clf`, де `X_train` - ознаки категорій `y_train`. Тепер можна використовувати вбудовану в клас метрику `score`, щоб визначити точність передбачених для `X_test` категорій по істинним значенням цих категорій у тестовій множині. При використанні даної метрики виводиться значення точності від 0 до 1, де 1 відповідає 100% збігу.

При ініціалізації `clf` для `RandomForestClassifier` виставлено значення `n_estimators = 100`, `n_jobs = -1`, де перший відповідає за кількість параметрів в BigData, а другий - за кількість ядер, які беруть участь у роботі процесора (при -1 задіяні всі ядра, за замовчуванням 1). Так як програма працює з даними BigData і нам ніде взяти тестуючу вибірку, використовується `train_test_split` для «розумного» розбиття даних на навчальну вибірку і системою тестування.

Для вирішення завдань на класифікацію важливим фактором є вибір найкращих параметрів для навчальної вибірки категорій [9].

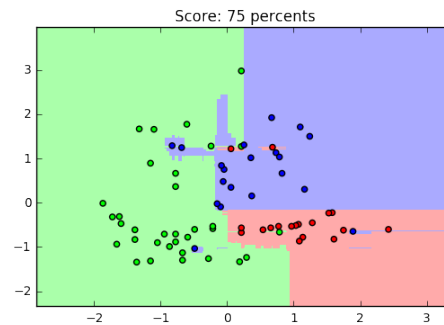


Рис. 2 Результат класифікації навчальної вибірки.

Для наочного перегляду результату навчання на даній вибірці було залишено тільки два параметри – тип й значення даних [4], значення яких розміщено у двовимірному просторі (див. Рис. 2).

Кожен колір на малюнку відповідає одному типу даних. Шкала вказана для зручності класифікації даних. В результаті роботи програми кожен параметр (точка) групується з ідентичними даними, що дає змогу розділити дані.

ВИСНОВОК

Отже, використання методу «дерево прийняття рішень» машинного навчання спрощує роботу з великою кількістю даних, що дозволяє класифікувати дані за допомогою алгоритму написаного на Python. Розроблена програма дає можливість розділити інформацію з різних датчиків і використовувати їх для подальших розрахунків та формування алгоритмів керування виконавчих пристроїв.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] L. V. Kantorovich and M. . Gavuri, “Primeneniye matematicheskikh metodov v voprosakh analiza gruzopotokov [The use of mathematical methods in the analysis of cargo flows],” *Probl. povysheniya Eff. Rab. Transp.*, pp. 110–138, 1949.
- [2] D. Dantsig and F. Vol’f, “Algoritm razlozheniya dlya zadach lineynogo programmirovaniya [Decomposition Algorithm for Linear Programming Problems],” *Matematika*, vol. 8, no. 1, pp. 151–160, 1964.
- [3] O. I. Aven, S. Y. Lovetskiy, and G. Y. Moiseyenko, *Optimizatsiya transportnykh potokov [Traffic optimization]*. Moskva: Nauka, 1985.
- [4] A. V. Krylov and Y. S. Yamnenko, “Rishennya transportnoyi zadachi za dopomohoyu shtuchnoho intelektu dlya mors’koho transportu [Solving the transport problem using artificial intelligence for maritime transport],” *Shipbuild. Mar. Infrastruct.*, vol. 1, no. 11, pp. 116–122, 2019,
URL: <http://smi.nuos.mk.ua/archive/2019/1/14.pdf>
- [5] V. G. Galaburda, *Optimal’noye planirovaniye gruzopotokov [Optimal freight traffic planning]*. Moskva: Transport, 1985.
- [6] A. L. Lur’ye, “Algoritm resheniya setevoy transportnoy zadachi s ograniчениem propusknykh sposobnostey metodom uslovno-optimal’nykh planov [An algorithm for solving a network transportation problem with bandwidth limitation by the method of conditionally optimal plans],” in *Materialy Konferentsii po opytu i perspektivam primeneniya matematicheskikh metodov i EMM v planirovani.*, 1962, pp. 3–13.
- [7] V. Y. Zhuykov et al., *Pidvyshchennya efektyvnosti system z vidnovlyuvanymy dzherelamy enerhiyi [Improving the efficiency of systems with renewable energy sources]*. Ky’iv: Kafedra, 2018, ISBN: 978-617-7301-48-5.



[8] G. De Jong, H. F. Gunn, and W. Walker, "National and international freight transport models: An overview and ideas for further development," *Transp. Rev.*, vol. 24, no. 1, pp. 103–124, 2004.

[9] L. A. Beklaryan and N. K. Khachatryan, "Traveling wave type solutions in dynamic transport models," *Funct. Differ. Equations*, vol. 13, no. 2, pp. 125–155, 2006.

UDC 621:314

Use of Machine Learning Methods in MicroGrid

A. V. Krylov, ORCID [0000-0002-7203-300X](https://orcid.org/0000-0002-7203-300X)

Department of Electronic Devices and Systems

Faculty of Electronics

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Kyiv, Ukraine

Abstract—Use of Internet of Things (IoT) concept plays the important role in modern electrotechnical systems with distributed generation. In particular, for the objects like "smart house" in the bounds of this concept some processes are considered that provide fulfillment of such control and regulation functions like following: light control; control of electrical energy consumption; monitoring of biotelemetrical parameters. Of course, this list is far from the complete one. Additionally, the range of control tasks depends on functionality and features of the system. The more is range of the tasks and number of different parameters to be measured, analyzed and processed – the more urgent become use of complicated control algorithms. Among them methods of artificial intelligent has the important place. The actuality is additionally proven by the further increasing of sales volume of the smartphones and other gadgets that combine more and more sensors and give more and more possibilities for the learning, getting and analyzing of the huge volume of different-kind data. Modern systems of energy consumption control use automated regulation of the devices and automated process of decision making. During this, demands of the end consumer get more priority because of necessity to provide admirable level of "comfortability" while all demands of safety and life support should be answered too. However, the presence of the human like a consumer of the services of smart house leads to the complexity of control algorithms creation due to subjectivity and impossibility to predict human behavior. Thus, the methods that can more or less imitate human behavior in technical system have the huge perspective. The aim of the paper is investigation of complicated intellectual methods of decision-making in MicroGrid generation as well as creation of the method able to classify working regimes in MicroGrid. In machine learning methods it is highly important to provide necessary quality of the learning based on accumulated data sets. This quality depends not only on characteristics of dataset itself (like completeness and level of compensation) but also on the learning technique. There is classification of the learning methods was fulfilled in the paper. Among them we considered methods based on following concepts: symbolism; connectivity; evolution; Bayesian probability; analogies. For each method and each ideology we pointed the idea, the concept, the approach and the most problem. Such classification allows to understand better what kind of learning is more preferable on this or that stage or in this or that situation. Then correspondence between the machine learning methods and control tasks in MicroGrid were investigated and presented. The software for the simple situation of light and heating control were elaborated and presented in the paper. Visualization of the machine learning result were executed and the areas of different classes were defined and shown.

Keywords — *MicroGrid; machine learning; decision-making tree.*

