

ВИКОРИСТАННЯ МЕТОДІВ КЛАСТЕРИЗАЦІЇ ДЛЯ АНАЛІЗУ СПОЖИВАННЯ ЕЛЕКТРОЕНЕРГІЇ

Ювченко К.А. – гр. МгіТ-24, магістрант, yuvchenko.k.a@gmail.com
Київський національний університет технологій та дизайну

У сучасних умовах цифровізації енергетичного сектору, зокрема з впровадженням систем "розумних" лічильників (smart meters), накопичуються величезні масиви даних про споживання електроенергії. Ці дані містять цінну інформацію про моделі поведінки споживачів, однак їхній безпосередній аналіз утруднений через обсяг і складність. Традиційні методи аналізу, орієнтовані на агреговані показники, не дозволяють виявити приховані закономірності та сегментувати споживачів за індивідуальними характеристиками. У цьому контексті методи машинного навчання, зокрема кластеризація, стають ключовим інструментом для перетворення сирих даних на корисні знання, що необхідні для прийняття ефективних управлінських рішень.

Мета роботи: дослідити потенціал методів кластеризації для виявлення типових профілів навантаження, сегментації споживачів та підвищення ефективності функціонування енергетичних систем.

Основна частина

1. Джерела даних та їх попередня обробка

Вихідними даними для аналізу є часові ряди споживання електроенергії, отримані від "розумних" лічильників. Дані можуть мати різну часову деталізацію (хвилинні, погодинні, добові). Типовий набір даних включає:

- Ідентифікатор споживача.
- Відмітки часу.
- Обсяг спожитої електроенергії (кВт·год).

Передобробка даних є критично важливим етапом і включає:

1. **Очищення даних:** усунення викидів, аномальних значень та заповнення пропусків (наприклад, методами інтерполяції або використанням середніх значень). [1, с. 4]

2. **Нормалізація/стандартизація:** приведення даних до єдиного масштабу для усунення впливу абсолютних величин споживання та забезпечення коректної роботи алгоритмів кластеризації. Найпоширенішими методами є Z-score стандартизація та Min-Max нормалізація. [1, с. 6, 2, с. 479,]

3. **Агрегація та трансформація:** формування добових або тижневих профілів навантаження для кожного споживача. Для кращої інтерпретації

можуть використовуватися ознаки, такі як пікове споживання, нічне споживання, співвідношення споживання у вихідні та робочі дні.

2. Методи кластеризації для аналізу енергоспоживання

Кластеризація - це задача неконтрольованого навчання, мета якої - розбити набір даних на групи (кластери) таким чином, щоб об'єкти всередині одного кластера були схожими між собою, а об'єкти з різних кластерів - відрізнялися.

Найбільш прийнятними для аналізу енергоспоживання є такі алгоритми:

1. **K-means (метод k-середніх):** Найпопулярніший алгоритм через свою простоту та ефективність. Він прагне мінімізувати дисперсію всередині кластерів. Його недоліком є необхідність заздалегідь задавати кількість кластерів (k) та чутливість до викидів. Для вибору оптимального k використовуються метрики, такі як Silhouette Score або Elbow Method. [1, с. 6, 3, с. 122]

2. **Ієрархічна кластеризація:** Будує дерево кластерів (дендрограму), що дозволяє аналізувати дані на різних рівнях деталізації. Може бути агломеративною (об'єднання дрібних кластерів) та дивізійною (розділення великих кластерів). Цей метод не вимагає заздалегідь заданої кількості кластерів і є більш інтуїтивно зрозумілим для інтерпретації результатів. [3, с. 71]

3. **DBSCAN (Щільностно-орієнтований просторовий кластеринг застосовно для зашумлених даних):** Ефективний для виявлення кластерів довільної форми та ігнорування викидів (шумових точок). Його перевагою є те, що він не вимагає фіксації кількості кластерів, але вимагає підбору параметрів щільності (`eps` і `min_samples``). [3, с. 220]

3. Інтерпретація результатів та практичне застосування

Після виконання кластеризації необхідно проаналізувати отримані кластери:

1. **Характеристика профілів навантаження:** для кожного кластера будується усереднений профіль споживання та аналізуються його характеристики (час піку, нічний провал, стабільність).
2. **Сегментація споживачів:** на основі аналізу кластери можуть бути інтерпретовані як сегменти споживачів [4, с.8], наприклад:
 - a) **"Ранковий пік":** споживачі з високим навантаженням у ранкові години.
 - b) **"Денний пік":** домогосподарства з активним споживанням вдень.
 - c) **"Вечірній пік":** домогосподарства з активним споживанням у вечірній час.
 - d) **"Нічні":** споживачі з високим навантаженням у нічний час.

е) "**Стабільні**": споживачі з відносно рівномірним графіком протягом доби (наприклад, промислові підприємства з безперервним циклом).

Практична цінність результатів включає:

- **Персоналізація тарифів:** енергокомпанії можуть розробляти диференційовані тарифні плани, орієнтовані на конкретні сегменти споживачів (наприклад, нічні тарифи для "нічних" споживачів).
- **Управління навантаженням на мережу:** прогнозування та вирівнювання графіків навантаження шляхом цілеспрямованого впливу на певні групи споживачів (програми Demand Response), що є особливо актуальним у існуючій ситуації з постійними атаками на як генеруючи так і розподільні потужності енергосистеми України.
- **Виявлення аномалій:** порівняння реального профілю споживача з типовим профілем його кластера дозволяє швидко виявляти несправності обладнання, несанкціоноване підключення або зміни у моделі поведінки.
- **Планування розвитку мереж:** дані про концентрацію певних типів споживачів у різних географічних районах допомагають оптимізувати інвестиції в мережеву інфраструктуру.

Висновки

Методи кластеризації є потужним та ефективним інструментом для аналізу великих масивів даних про споживання електроенергії, що дозволяє перейти від аналізу середніх значень до глибокої сегментації споживачів.

Успішне застосування цих методів вимагає ретельної попередньої обробки даних та обґрунтованого вибору алгоритму кластеризації, враховуючи специфіку набору даних та цілей дослідження.

Результати кластеризації – типізовані профілі навантаження – мають значний практичний потенціал для підвищення економічної ефективності роботи енергокомпаній, оптимізації роботи енергосистеми та розробки персоналізованих послуг для кінцевих споживачів.

Подальший розвиток напряму пов'язаний із застосуванням більш складних методів, таких як глибинне навчання для кластеризації часових рядів, а також інтеграцією даних про споживання електроенергії з іншими джерелами даних (наприклад, погодними умовами, соціально-демографічними характеристиками).

Список використаних джерел:

1. Tureczek A., Nielsen P., Madsen H. Electricity Consumption Clustering Using Smart Meter Data. *Energies*. 2018. Vol. 11, no. 4. P. 859. URL: <https://doi.org/10.3390/en11040859> (date of access: 20.10.2025).

2. Phyto E. E., Myat E. E. Efficient K-Means Clustering Algorithm Using Feature Weight and Min-Max Normalization. *International Journal of Science and Engineering Applications*. 2018. Vol. 7, no. 12. P. 479–482. URL: <https://doi.org/10.7753/ijsea0712.1001> (date of access: 20.10.2025).
3. Cluster analysis / B. S. Everitt et al. Chichester, UK : John Wiley & Sons, Ltd, 2011. URL: <https://doi.org/10.1002/9780470977811> (date of access: 20.10.25).
4. Haben S., Singleton C., Grindrod P. Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data. *IEEE Transactions on Smart Grid*. 2016. Vol. 7, no. 1. P. 136–144. URL: <https://doi.org/10.1109/tsg.2015.2409786> (date of access: 20.10.2025).