

ПРОГНОЗУВАННЯ ЕЛЕКТРОНАВАНТАЖЕННЯ З ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Анотація: Розглянуто використання методів машинного навчання, а саме методу Gradient Boosting для короткострокового прогнозування електронавантаження для задачі оперативного керування режимами розподільних мереж в умовах впровадження концепції SMAR GRID. Встановлено, що метод Gradient Boosting може бути використаний для складання прогнозу електронавантаження на кожні дві години з точністю до 5% від фактичного обсягу навантаження. Проведені порівняння прогнозних значень з фактичними дозволяють говорити про адекватність обраної моделі мережі і її використанні на практиці для налагодження ефективної роботи розподільної мережі.

Ключові слова: прогнозування, електричне навантаження, SMART GRID, машинне навчання, розподільна мережа, ансамблеві методи, керуючий вплив

Abstract: The use of machine learning methods, namely the Gradient Boosting method for short-term power forecasting for the problem of operational control of distribution network modes under the conditions of SMAR GRID concept implementation, is considered. It has been established that the Gradient Boosting method can be used to estimate electricity consumption every two hours with an accuracy of up to 5% of the actual amount of electricity consumed. The comparisons of the predicted values with the actual ones make it possible to speak about the adequacy of the chosen network model and its applicability in practice for establishing efficient operation of the distribution network.

Вступ

Однією з головних задач, яку необхідно вирішувати для розподільних мереж – це пошук та вибір оптимальних місць розмикання. Метою вирішення цієї задачі є те, що даний захід дозволяє позитивно вплинути на величину втрат електроенергії краще, ніж будь-який інший захід. Концепція SMART GRID передбачає впровадження джерел розподіленої генерації в існуючі мережі, тому постає задача не тільки пошуку місць розмикання, а і оперативного керування конфігурацією мережі в реальному часі.

Виходячи з цього прогнозування графіків електричного навантаження стає все більш важливою задачею, що дає змогу отримати ефективний інструмент для складання балансів, планування та управління режимами роботи розподільних мереж, задоволення зростаючого попиту на електроенергію. Вирішення цих проблем особливо актуальне при умові трансформації енергосистеми згідно концепції SMART GRID, що, у якості важливої складової, передбачає появу джерел розподіленої генерації в складі існуючих мереж.

На даний час розроблено багато методів прогнозування, які відрізняються один від одного в залежності від виду вирішуваних задач при планування роботи системи. Згідно [1], існує наступна класифікація методів прогнозування електричного навантаження:

- оперативне – від кількох хвилин до декількох годин в межах поточної доби;
- короткострокове – від однієї до декількох діб;
- довгострокове поточне – від одного до декількох місяців, кварталів і до одного року;
- довгострокове – від одного до п'яти років;
- перспективне – на п'ять років і більше.

В зарубіжній літературі [2,3] приведена подібна класифікація з певними відмінностями:

- very short-term load forecasting (найбільш короткострокове прогнозування) – з інтервалом до 24-х годин;
- short-term load forecasting (короткострокове прогнозування) – з інтервалом від 24-х до 168-и годин (від однієї доби до тижня);
- mid-term forecasting (середньострокове прогнозування) – з інтервалом від одного тижня до року;
- long-term forecasting (довгострокове прогнозування) – з інтервалом більше одного року.

Про точність прогнозування можна судити в залежності від математичного апарату, який застосовується в тій чи іншій моделі та від низки факторів, які безпосередньо впливають на вихідний результат. Взагалі, формування електричного навантаження являється комплексним нестационарним стохастичним процесом [1,4], що передбачає наявність регулярної та випадкової складових [5,6]. Регулярна складова залежить від набору факторів, які сезонно коливаються в одних і тих же межах, характерних для певної місцевості, які залежать від метеорологічних умов, особливостей функціонування підприємств, характеру життя населення та ін. Нерегулярна складова залежить від появи в сучасних мережах джерел розподіленої генерації, характер вироблення електроенергії яких напряму залежить від погодних умов. В цілому, фактори, які впливають на електричне навантаження можна розділити на чотири групи: часові, метеофактори, економічні і т.д. [7].

Найсуттєвішими являються метеоумови, такі як температура та освітленість, що визначають сезонний характер електроспоживання [9,10]. Це обумовлено роботою різної опалювальної та охолоджувальної техніки в зимовий та літній періоди та тривалістю світлового дня в різні пори року. В свою чергу інші метеофактори (швидкість та направленість вітру, сонячна інсоляція) визначають об'єми генерації нетрадиційних джерел. Як показано в [8], вони суттєво визначають добові коливання та нерівномірність процесів споживання та генерації енергії.

Виходячи з цього, виникає проблема оптимального розподілення потужності для задоволення попиту та ефективного функціонування мережі. Тому, метою побудови моделі прогнозу є отримання даних не тільки в залежності від сезонності, а і з урахуванням нетрадиційних джерел, що дасть змогу для організації управління режимами роботи мережі в реальному часі.

Мета та завдання

Згідно прийнятої концепції SMART GRID наша держава перейшла до етапу побудови принципово нової моделі функціонування електричних мереж, що передбачає зміну її структури і, як наслідок, зміну підходів до її функціонування та керування.

Відповідно до концепції керування має здійснюватися за умови автоматичності, оперативності та на базі інтелектуальних технологій з метою надійного та якісного електропостачання. Процес імплементації на практиці виявляється досить складним, ресурсовитратним і проходить на протязі значного проміжку часу. Тому при побудові сучасної мережі та протягом перехідного періоду важливим стає впровадження новітніх інформаційних та оперативних комплексів.

Базуючись на сучасних інформаційних комплексах можливо налагодити ефективне збирання та обробку даних про електроспоживання, генерацію, з високою точністю аналізувати режими роботи мережі та приймати оперативні рішення по їх керуванню та, як наслідок, зменшити втрати електроенергії та підвищити надійність та якість електропостачання. Це вимагає визначення і застосування адекватних аналітичних критеріїв і підходів до оцінки, аналізу та управління режимами мереж.

Відповідно до вищеприведеного, метою статті є розробка методики, на основі якої оперативний персонал енергокомпаній зможе здійснювати управління режимами розподільної

мережі з двостороннім живленням 6-10 кВ шляхом зміни конфігурації мережі і ґрунтуючись на мінімізації втрат активної потужності. Для вирішення поставленого завдання необхідно розробити модель прогнозування електричного навантаження у вузлі розподільної мережі з джерелами розподіленої генерації.

Матеріал і результати досліджень

В якості моделі для прогнозування електричного навантаження обрано один з ансамблевих методів машинного навчання, принцип роботи яких полягає у об'єднанні так званих «слабких учнів» з низькою здатністю прогнозування у «сильного учня», який володіє в свою чергу високою точністю прогнозу.

Метод Gradient Boosting заснований на принципі мета-алгоритмів ансамблевих методів, так званому boosting, який передбачає послідовне навчання «слабких учнів» адаптивним способом, коли наступний «слабкий учень» залежить від попереднього. На виході ми отримуємо модель, яка детермінованим чином об'єднує всіх «слабких учнів» у одного «сильного» для отримання результуючого прогнозу.

Gradient Boosting - це метод машинного навчання для задач регресії і класифікації, який створює модель прогнозування у вигляді безлічі моделей слабого прогнозування, зазвичай дерев рішень. Він будує модель поетапно, як і інші методи, і узагальнює їх, дозволяючи оптимізувати довільну диференційовану функцію втрат, яка є мірою того, наскільки добре навчена модель передбачає цільову змінну на навчальній вибірці. Типовими прикладами функції втрат є середньоквадратична помилка (MSE) в задачах регресії і логістична функція втрат (logistic loss) [10].

Алгоритм Gradient Boosting вимагає функціонування таких компонентів:

1. Функція втрат: щоб зменшити помилки в прогнозуванні нам потрібно оптимізувати функцію втрат. На відміну від AdaBoost, неправильний результат не має великої ваги при підвищенні градієнта. Задачею алгоритму являється зменшення функції втрат шляхом усереднення результатів від «слабких учнів».

2. «Слабкий учень»: для підвищення градієнта нам потрібні «слабкі учні», щоб виконати прогноз. Щоб отримати реальні значення в якості вихідних даних, ми використовуємо дерева рішень. Щоб отримати найбільш підходящу точку поділу, ми створюємо дерева жадібним чином, завдяки чому модель відповідає набору даних.

3. Адитивна модель: при підвищенні градієнта ми намагаємося зменшити втрати, додаючи дерева рішень. Крім того, ми можемо мінімізувати частоту помилок, скорочуючи параметри. В цьому випадку ми проектуємо модель таким чином, що додавання дерева не змінює існуюче дерево [11-12].

Алгоритм розрахунку:

Задаємося вихідними даними, які представляють собою масив з прогнозними даними та фактори, від яких вони залежать: $\{(x_i, y_i)\}_{i=1}^n$

При вирішенні задачі прогнозування необхідно враховувати диференційовану функцію втрат: $L(y_i, F(x_i)) = \frac{1}{2}(y_i - F(x_i))^2$, похідна якої дорівнює: $L'(y_i, F(x_i)) = -(y_i - F(x_i))$

Крок 1: Ініціалізація моделі з постійним значенням:

$$F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \gamma) \quad (1)$$

Крок 2: для $m=1$ до M :

а) Розраховується:

$$r_{im} = - \left[\frac{\partial L(y_i, F(x_i))}{\partial L(F(x_i))} \right]_{F(x)=F_{m-1}(x)}, \text{ для } i = 1, \dots, n \quad (2)$$

б) Побудова дерева рішень для значень r_{im} і створення кінцевої області R_{jm} для $j = 1, \dots, J_m$.

в) Для $j = 1, \dots, J_m$ розраховується:

$$\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma) \quad (3)$$

г) Оновлення результатів:

$$F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm}) \quad (4)$$

Оцінка точності прогнозової моделі

Для оцінки прогнозу електричного навантаження використовуються статистичні оцінки помилки прогнозування часового ряду. Найпростіший показник - відхилення фактичного значення від прогнозу в кількісному вигляді.

У практиці розраховують помилку прогнозування по кожній окремій позиції, а також розраховують середню помилку прогнозування. Наступні показники помилки відносяться саме до показників середніх помилок прогнозування.

До них відносяться:

1) MAPE - середня абсолютна помилка у відсотках:

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (5)$$

2) MAE – середня абсолютна помилка:

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (6)$$

3) MSE – середньоквадратична помилка:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

4) Коефіцієнт детермінації R^2 :

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (8)$$

5) Медіана абсолютна помилка:

$$MedAE = \operatorname{median}(|y_1 - \hat{y}_1|, \dots, |y_n - \hat{y}_n|) \quad (9)$$

6) Середньоквадратична логарифмічна помилка:

$$MSLE = \frac{1}{n} \sum_{i=1}^n (\log(1 + y_i) - \log(1 + \hat{y}_i))^2 \quad (10)$$

Результати моделювання. В даній роботі, як було сказано вище, розглядається короткострокове прогнозування в межах декількох годин і тільки для тих ділянок електричної мережі, в яких спостерігається суттєва зміна значення електричного навантаження. Для цих цілей розмір тренувальної послідовності дорівнює тижню (п'ять робочих днів і два вихідних), розмір тестової послідовності дорівнює одному дню. Розміри послідовностей вибиралися з тих міркувань, що при прогнозуванні на кілька годин модель повинна бути натренована на даних, які повинні включати в себе інформацію відносно графіків навантаження як робочих днів, так і вихідних. Вибирати розмір послідовності більше тижня не має сенсу, так як більш ранні дані можуть негативно вплинути на точність моделі короткострокового прогнозування.



Рисунок 1 – Вихідні дані електричного навантаження за одну добу

Проаналізувавши добові графіки електричного навантаження можна зробити висновок, що для вирішення поставленого завдання, розробки моделі короткострокового прогнозування, модель повинна працювати на тих ділянках графіка, де є якесь виражене зміни значення електричного навантаження. Для того щоб перевірити і протестувати дану модель для вирішення поставленого завдання вибираємо чотири часових відрізка на добовому графіку і проводимо прогнозування.

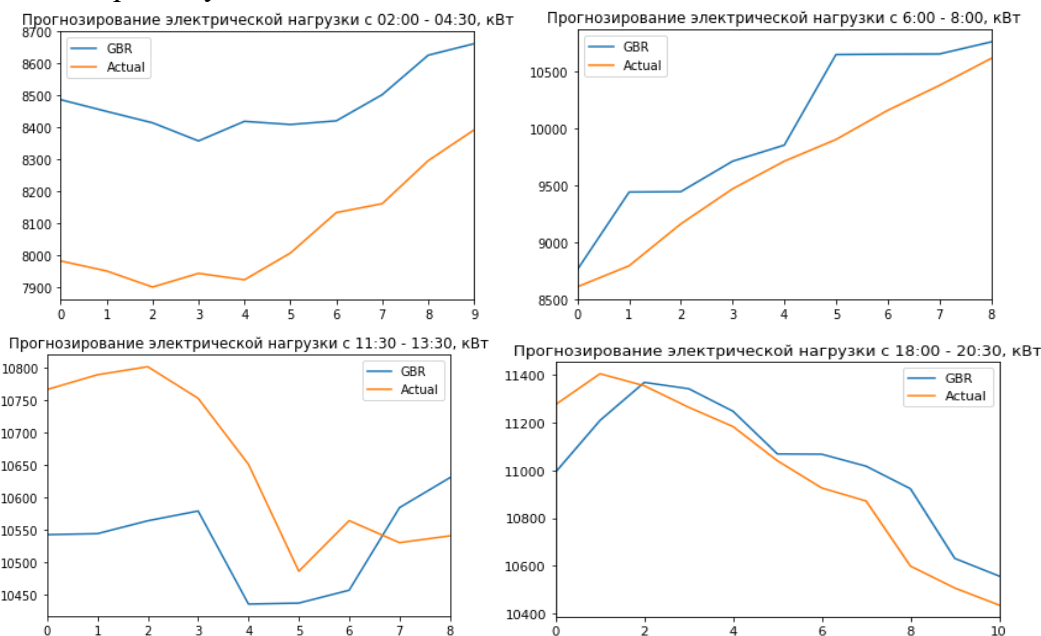


Рисунок 2 – Результати прогнозування на 4 часових відрізках

Таблиця 1 – Оцінка продуктивності моделі на даних чотирьох часових відрізків

Часовий відрізок	R ²	MSE	MAPE (%)	MAE	MLSE	MedAE
02:00-4:30	0,5236	171902	4,78	404,7	0,0025	407,3
06:00-08:00	0,635	166293	3,45	347,21	0,0017	276,13
11:30-13:30	0,714	29788	1,472	155,08	0,0002	173,76
18:00-20:30	0,556	27954	1,261	138,57	0,00023	124,39

Висновки

У ході дослідження було проаналізовано можливість використання методу Gradient Boosting для вирішення задачі короткострокового прогнозування електричного навантаження в реальному часі. За результатами розрахунку можна зробити висновок про адекватність даної моделі для цілей прогнозування. Даний метод може використовуватися як один із можливих методів, на основі яких можливо просканувати стан мережі і визначити доцільність її реконфігурації.

Даний алгоритм реалізовано в програмному середовищі Python і на основі даних розрахунку можна зробити висновок про його працездатність для виконання поставленої задачі.

Список літератури

1. Седов А.В. Системы контроля, распознавания и прогнозирования электропотребления: модели, методы, алгоритмы и средства/ А.В. Седов, И.И. Надтока. – Ростов-н/Д: Изд-во Ростовского университета, 2002. 320 с.
2. Mishra S. Short term load forecasting using computation intelligence methods: Thesis for the degree of Master of technology electronics and communication engineering/Mishra Sanjib. – National Institute of Technology, India, Rourkela, 2008. 89 p.
3. Yang, J. Power System Short-term Load Forecasting: Thesis for Ph.d degree / Yang Jingfei. – Elektrotechnik und Informationstechnik der Technischen Universitat, Germany, Darmstadt, 2006. 139 p.
4. Бокс, Дж. Анализ временных рядов. Прогноз и управление. В 2 ч. Ч. 1/Дж. Бокс, Г. Дженкинс; под ред. В.Ф. Писаренко. – М.: Мир, 1974. 406 с.
5. Кильдишев Г.С. Анализ временных рядов и прогнозирование / Г.С. Кильдишев, А.А. Френкель. – М.: Статистика, 1973. 104 с.
6. Дуброва Т.А. Статистические методы прогнозирования: учеб. пособие для вузов / Т.А. Дуброва. – М.: ЮНИТИ-ДАНА, 2003. 206 с.
7. Гросс, Дж. Краткосрочное прогнозирование нагрузки / Дж. Гросс, Ф.Д. Гальяны// ТИИЭР. – 1987. – Т.75, №12. 6-21 с.
8. Губский С.О. Краткосрочное прогнозирование электропотребления в операционной зоне регионального диспетчерского управления с учетом фактора освещенности: дис. ... канд. техн. Наук: 05.14.02/Губский Сергей Олегович. – Новочеркасск, 2012. 231 с.
9. Влияние колебаний метеорологических факторов на энергопотребление энергообъединений / Б.И. Макоклюев, В.С. Павликов, А.И. Владимиров, Г.И. Фефелова // Энергетик. 2003. №6. 14-19 с.
10. Макоклюев Б.И. Учет влияния метеорологических факторов при прогнозировании электропотребления энергообъединений/ Б.И. Макоклюев, В.Ф. Еч / Энергетик. 2004. №6. 15-16 с.

11. Smola A. and Schölkopf B. A Tutorial on Support Vector Regression Neuro COLT, Technical Report NC-TR-98-030. 2003 URL: <http://library.nlu.edu.ua/Biblioteka/sait/nauka/gost/spisok-DSTU.pdf>
12. Jonathan H. Exploring XGBoost. An introduction into classification and feature selection with Ensemble Learning and XGBoost. Jan 17, 2019. URL: <https://towardsdatascience.com/exploring-xgboost-4baf9ace0cf6>
13. Debaditya Ch., Hazem E. Advanced machine learning techniques for building performance simulation: a comparative analysis. 2-6 p. DOI: [10.1080/19401493.2018.1498538](https://doi.org/10.1080/19401493.2018.1498538)