

АЛГОРИТМИ МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ГЕНЕРАЦІЇ СЕС

Мороз О.М., д.т.н., проф., e-mail: moroz.an@ukr.net

Сотнік О.В., аспірант, e-mail: sidi.leha@gmail.com, (ДБТУ, м. Харків, Україна)

The thesis discusses machine learning algorithms for forecasting the generation of solar power plants (SPPs), including linear models, support vector machines (SVM/SVR), ensemble methods (Random Forest, XGBoost) and neural networks (LSTM, CNN). Their advantages, limitations and optimal applications are identified, and a combined approach is proposed to improve the accuracy of forecasts depending on the type of data and time horizon.

Прогнозування генерації сонячної електростанції (СЕС) є критично важливим завданням для ефективного управління енергетичними системами, включаючи локальні мережі (ЛЕС) та об'єднані енергосистеми. Точний прогноз дозволяє оптимізувати баланс між виробництвом і споживанням електроенергії, зменшити витрати на резервні потужності та підвищити стабільність енергопостачання. Для вирішення цієї задачі використовуються різноманітні алгоритми машинного навчання, які можна класифікувати наступним чином:

- лінійні моделі (лінійна регресія, Ridge, Lasso);
- методи на основі опорних векторів (SVM/SVR);
- дерева рішень та ансамблеві методи (Random Forest, XGBoost, LightGBM);
- нейронні мережі (MLP, RNN/LSTM, CNN).

Кожен з цих підходів має свої переваги, обмеження та оптимальні сфери застосування. Одним із найпростіших алгоритмів для прогнозування, який описує залежність між вхідними ознаками (наприклад, інтенсивність сонячного випромінювання, температура, час доби) та цільовою змінною (генерація енергії) є лінійна регресія:

$$Y^*(X) = a_1x_1 + a_2x_2 + \dots + a_mx_m + b,$$

де $Y^*(X)$ - прогнозоване значення, x - вектор з m вхідних ознак, a - вектор з m коефіцієнтів регресії, b – випадкова похибка, розподіл якої в загальному випадку залежить від незалежних змінних (зсув).

Лінійна регресія ефективна, коли залежність між даними близька до лінійної, але для складніших випадків використовуються інші методи.

Метод опорних векторів (Support Vector Machine, SVM) може застосовуватися як для класифікації, так і для регресії (SVR – Support Vector Regression). Він базується на ідеї знаходження гіперплощини, яка:

- у випадку класифікації – максимізує відстань між класами (маргінал);
- у випадку регресії – мінімізує відхилення прогнозованих значень від реальних у межах заданої смуги (ϵ -труби).

SVR особливо ефективний для короткострокового прогнозування, коли дані мають складні нелінійні залежності, але їх обсяг невеликий (наприклад, прогноз на основі погодних умов за останні кілька днів). Дерева рішень будуються шляхом рекурсивного розбиття даних на основі правил, що базуються на

вхідних ознаках. У контексті прогнозування генерації СЕС:

– кожен вузол дерева відповідає умові (наприклад, «чи інсоляція $> 500 \text{ Вт/м}^2$?»).

– листя дерева містять прогнозовані значення генерації.

Для підвищення точності використовуються ансамблеві методи:

– бегінг (Random Forest) – будує множину дерев на різних підвибірках даних, усереднюючи їхні прогнози, зменшує дисперсію та ризик перенавчання;

– бустинг (XGBoost, LightGBM) – послідовно покращує прогнози, зосереджуючись на помилках попередніх моделей, є ефективним для складних залежностей між даними.

Ансамблеві методи особливо ефективні для середньострокового прогнозування (добові/тижневі прогнози), коли важливо враховувати взаємодію багатьох факторів (метеорологічні умови, стан обладнання тощо). Для прогнозування СЕС часто використовуються нейронні мережі, зокрема:

– багатошарові перцептрони (MLP) – базові моделі з прихованими шарами, які можуть навчатися нелінійним залежностям;

– рекурентні нейронні мережі (RNN, LSTM) – ефективні для роботи з часовими рядами (наприклад, прогнозування генерації на основі минулих даних);

– згорткові нейронні мережі (CNN) – можуть аналізувати просторові дані (наприклад, супутникові знімки хмарності).

Вибір алгоритму залежить від характеру даних:

– лінійна регресія – для простих залежностей;

– SVM – для невеликих, але складних даних;

– ансамблі дерев (Random Forest, XGBoost) – для високої точності;

– нейронні мережі – для складних часових або просторових даних.

На основі аналізу визначено переваги та обмеження досліджених алгоритмів для прогнозування генерації СЕС, які наведені в таблиці.

Алгоритм	Найкраще підходить для	Обмеження
Лінійна регресія	Прості лінійні залежності, швидкий прогноз	Низька точність для складних даних
SVM/SVR	Невеликі, але складні набори даних	Повільний для великих даних
Random Forest	Середньострокові прогнози, стабільність	Вимагає налаштування гіперпараметрів
XGBoost/LightGBM	Висока точність, складні залежності	Схильність до перенавчання
LSTM/CNN	Довгострокове прогнозування, обробка зображень	Висока обчислювальна складність

Оптимальним підходом для прогнозування генерації СЕС є комбінація методів, наприклад:

– LSTM для аналізу часових рядів (генерація минулих періодів);

– XGBoost для врахування метеорологічних факторів;

– CNN для обробки супутникових даних про хмарність.

Таким чином, вибір алгоритму залежить від специфіки даних, часового горизонту прогнозування та доступних обчислювальних ресурсів.