

УДК 004.855.5

ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ ДЛЯ АНАЛИЗА ЭЛЕКТРОПОТРЕБЛЕНИЯ ПРЕДПРИЯТИЯ НА ОСНОВЕ НЕЙРОНОЙ СЕТИ LSTM

Рубан Е.А., студент гр. ИИМ-241, II курс
Научный руководитель: Протодяконов А.В., к.н., доцент
Кузбасский государственный технический университет
имени Т.Ф. Горбачева
г. Кемерово

В условиях цифровизации промышленности и роста объёмов данных предприятия всё чаще используют интеллектуальные методы анализа информации для повышения эффективности управления производственными процессами. Особенно актуальной задачей является прогнозирование энергопотребления. Неправильная оценка будущих значений нагрузки может приводить к перегрузке энергосистемы, неэффективному использованию энергетических ресурсов и увеличению эксплуатационных затрат. В условиях динамически изменяющихся производственных процессов традиционные статистические методы анализа временных рядов часто оказываются недостаточно точными, поскольку они плохо учитывают сложные нелинейные зависимости и долгосрочные взаимосвязи между наблюдениями.

Использование LSTM-моделей позволяет анализировать динамику временных рядов, выявлять скрытые закономерности в данных и строить прогнозы на заданный временной период. Это делает применение таких моделей перспективным инструментом для разработки интеллектуальных систем поддержки принятия решений на предприятиях.

Для обучения и тестирования модели используем набор данных, содержащий значения временного ряда, отражающего динамику энергопотребления. Данные представлены в виде последовательности наблюдений, зафиксированных через каждый час с периода 2004 по 2018 год. На рис. 1. приведено 28 строк исходных данных.

Основные характеристики датасета:

1. Дата и время наблюдения, позволяющие анализировать последовательность событий.
2. Энергопотребление - основные показатели, которые используются для обучения модели. Варьируется от 9,58 тысяч до 25,7 тысяч мегаватт (MW). В среднем значения около 13, 5 мегаватт.

1	Datetime	AEP_MW
2	31.12.2004 1:00	13478
3	31.12.2004 2:00	12865
4	31.12.2004 3:00	12577
5	31.12.2004 4:00	12517
6	31.12.2004 5:00	12670
7	31.12.2004 6:00	13038
8	31.12.2004 7:00	13692
9	31.12.2004 8:00	14297
10	31.12.2004 9:00	14719
11	31.12.2004 10:00	14941
12	31.12.2004 11:00	15184
13	31.12.2004 12:00	15009
14	31.12.2004 13:00	14808
15	31.12.2004 14:00	14522
16	31.12.2004 15:00	14349
17	31.12.2004 16:00	14107
18	31.12.2004 17:00	14410
19	31.12.2004 18:00	15174
20	31.12.2004 19:00	15261
21	31.12.2004 20:00	14774
22	31.12.2004 21:00	14363
23	31.12.2004 22:00	14045
24	31.12.2004 23:00	13478
25	01.01.2005 0:00	12892
26	30.12.2004 1:00	14097
27	30.12.2004 2:00	13667
28	30.12.2004 3:00	13451
29	30.12.2004 4:00	13379

Рис.1. Данные для анализа

Далее проведём подготовку данных. Она выполняется поэтапно и состоит из следующих этапов:

Первый этап - очистка данных. Удаляются или заполняются пропущенные значения и аномальные наблюдения.

Второй этап - нормализация данных. Производится масштабирование данных в диапазон значений от 0 до 1, для повышения эффективности обучения нейронной сети.

Третий этап - формирование последовательностей. Временной ряд преобразуется в набор обучающих последовательностей фиксированной длины. Например, если размер окна равен n , то входные данные имеют вид (1):

$$(x_{t-n}, \dots, x_{t-1}, x_t), \quad (1)$$

Четвертый этап - разделение данных. Данные разделяются на обучающую и тестовую выборку.

Проведем обучение модели на обучающей выборке с использованием алгоритма оптимизации градиентного спуска.

Основные параметры обучения:

- количество эпох - 100;
- размер батча (порция данных, которую модель обрабатывает за один раз) - 64;
- скорость обучения (размер шага) - 0.001;

- количество нейронов в LSTM-слое -128 (в LSTM три слоя).

Обученную модель проверим на тестовой выборке: сопоставим то, что она предсказала, с фактическими данными. Результаты сравнения — на рис. 2, а сам график прогнозов и реальных значений — на рис. 3.

ТЕСТОВЫЕ ЗНАЧЕНИЯ И СОПОСТАВЛЕННЫЕ ПРЕДСКАЗАНИЯ				
№	Реальные (MW)	Предсказанные (MW)	Ошибка (MW)	Ошибка (%)
1	14229.33	14287.95	58.62	0.41
2	14368.33	14504.16	135.83	0.95
3	14224.33	14281.13	56.79	0.40
4	13705.00	13729.40	24.40	0.18
5	12968.00	12950.01	17.99	0.14
6	12206.00	12187.72	18.28	0.15
7	11601.33	11591.59	9.74	0.08
8	11256.00	11202.21	53.79	0.48
9	11060.00	11101.21	41.21	0.37
10	11107.67	11124.05	16.38	0.15
11	11457.33	11462.92	5.58	0.05
12	12283.00	12294.56	11.56	0.09
13	13287.67	13224.20	63.46	0.48
14	13985.67	14033.24	47.58	0.34
15	14291.33	14243.13	48.20	0.34
16	14295.67	14329.93	34.27	0.24
17	14321.33	14284.82	36.52	0.25
18	14261.33	14296.05	34.72	0.24
19	14189.33	14216.66	27.33	0.19
20	14116.00	14057.88	58.12	0.41

Рис.2. Тестовые значения и сопоставленные предсказания

Проанализировав полученные результаты, можно сделать несколько выводов:

1. Модель имеет высокую точность прогноза. Т.е. ошибки предсказаний малы — максимальная ошибка около 1%, в то время как большинство значений находятся ниже 0,5%. В следствии чего можно сказать, что модель LSTM улавливает динамику нагрузки очень точно.

2. Модель стабильна. Проанализировав рис. 2. можно сказать, что ошибки распределены равномерно, нет резких выбросов или аномальных предсказаний. Даже при небольших колебаниях нагрузки модель сохраняет надёжность.

3. Модель имеет небольшую склонность к недооценке или переоценке значений. Можно заметить, что в некоторых точках (например, №2 рис. 2.) предсказание немного превышает реальное значение, в других (например, №13 рис. 2.) — немного меньше. Однако, общая амплитуда отклонений незначительна и не влияет на практическое использование.

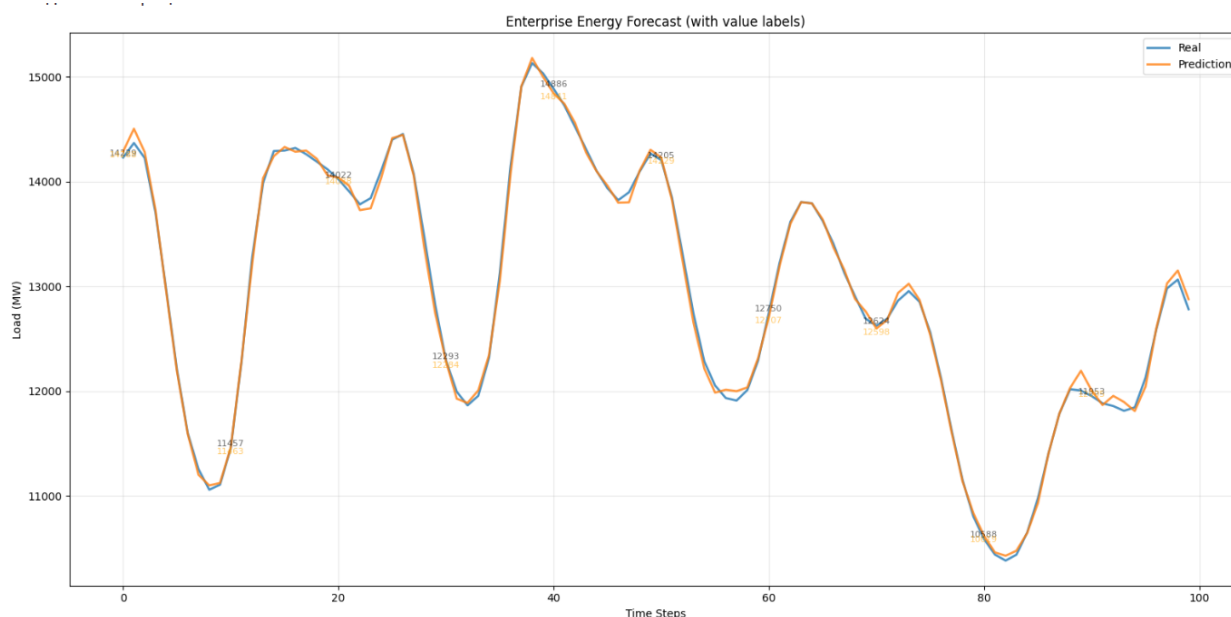


Рис.3. График сравнения реальных и предсказанных

Для оценки точности прогнозирования используем стандартные метрики ошибок.

MAE (Mean Absolute Error) - средняя абсолютная ошибка определяется как(2):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (2)$$

где:

N - количество наблюдений (размер выборки), y_i - фактическое значение (реальное значение временного ряда), $\hat{y}_i$ - прогнозируемое значение модели.

Данная метрика показывает среднее отклонение прогнозируемых значений от фактических.

RMSE (Root Mean Square Error) - корень из средней квадратичной ошибки рассчитывается по формуле (3):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (3)$$

где:

N - количество наблюдений (размер выборки), y_i - фактическое значение (реальное значение временного ряда), $\hat{y}_i$ - прогнозируемое значение модели.

RMSE более чувствителен к большим ошибкам прогнозирования. Поэтому если есть большие промахи - RMSE растёт.

MAPE (Mean Absolute Percentage Error) - средняя абсолютная процентная ошибка определяется как(4):

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right|, \quad (4)$$

где:

N - количество наблюдений (размер выборки), y_i - фактическое значение (реальное значение временного ряда), $\hat{y}_i$ - прогнозируемое значение модели.

Данная метрика показывает среднюю ошибку прогнозирования в процентах.

После проведения тестирования модели мы получили значения метрик:

- $MAE = 41,47 \text{ MW}$ - модель в среднем ошибается примерно на 0,3% от нагрузки
- $RMSE = 54,78 \text{ MW}$ – максимальная средняя ошибка относительно нагрузки равна 0,41%
- $MAPE = 0,29\%$ - уровень очень точного прогноза.

В заключение хочется сказать, что разработанная модель LSTM для прогнозирования потребления электроэнергии показала высокую точность и надёжность. Она способна давать прогнозы с минимальными отклонениями от реальных значений, что делает её эффективным инструментом для планирования и оптимизации энергоресурсов.

Список литературы:

1. Селезнёв М. LSTM нейронные сети: краткое объяснение принципов работы [электронный ресурс]. — URL: <https://habr.com/ru/companies/wunderfund/articles/331310/>.
2. Методы предварительной обработки данных для дальнейшего исследования. [Электронный ресурс]. - Режим доступа: <https://birdyx.ru/blog/show/data-pre-processing>.
3. Нормализация данных. [Электронный ресурс]. - Режим доступа: <https://habr.com/ru/articles/527334/>.
4. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение [текст]. — М.: ДМК Пресс, 2017. — 652 с.