

Оценка эффекта комбинирования моделей прогнозирования временных рядов

В.И. Гребнев, М.А. Немцев, А.П. Тюков³

¹*Волгоградский государственный университет*

²*Волгоградский государственный технический университет*

³*Волгоградский государственный технический университет*

Аннотация: В данной статье рассматриваются методы комбинирования моделей прогнозирования временных рядов с целью увеличения точности прогноза на основе анализа средних ошибок применительно к задаче прогнозирования на сутки вперед.

Ключевые слова: Временные ряды, комбинирование моделей, оценка точности, прогнозирование, точность прогноза.

Прогнозирование в современном мире является частью многих сфер жизни. А получение точного и стабильного прогноза является непростой задачей и требует обширных знаний в предметной области. Задача повышения точности прогноза чрезвычайно важна, и даже небольшое его улучшение при ежедневном использовании складывается в сэкономленные средства или ресурсы, а зачастую и первое и второе. Помимо настройки моделей и работы с данными есть дополнительные методы увеличения точности прогноза, одним из которых является комбинирование моделей, особенно ставший популярным в последнее время [1 – 3].

По типу прогнозной модели выделяют stochastic-model – стохастические, model-based – в основе которой лежит моделирование физические процессов, black-box – с применением машинного обучения [4]. Для оценки точности прогноза активно применяются такие метрики как MAPE, MAE, RMSE и др. [5-8].

В данной статье рассматривается комбинирование моделей на примере прогнозирования active energy (A_E) – солнечной энергии, полученной с использованием солнечных панелей и demand (D) – затраты энергетических ресурсов для объединения жилых помещений средних размеров [9].

В качестве базовых моделей, на основе которых, путем их комбинирования, будут получены ensemble-прогнозы, были использованы 3 типа моделей. Persistence модель - когда прогноз на завтра соответствует сегодняшним значениям. Так же этот прогноз удобно использовать в сравнении с другими моделями для оценки их эффективности, поскольку для его получения не требуется сил и времени прогнозиста. В качестве второй модели была применена библиотека Prophet от Facebook с открытым исходным кодом, в основе которой лежат стохастические алгоритмы. И третья модель GradientBoosting библиотеки scikit-learn языка python, в основе которой лежит дерево принятия решений.

Для построения прогноза на сутки вперед в качестве тренировочных данных используется временной ряд длиной 30 дней: дата (число, месяц, год и время) с шагом 15 минут и значение. Перед обучением была проведена обработка данных, в частности, были обработаны пропущенные значения в данных и удалены аномальные значения. Тип обучения sliding window, при котором происходит сдвиг данных на день для каждого нового прогнозируемого дня.

Для определения lags – важных для прогноза значений, на основе которых формируется набор обучающих данных, были получены с помощью statsmodels.tsa.stattools.pacf. Наиболее значимые для прогноза значения это последнее перед прогнозируемым значение (15 мин. назад), значение в это же время 1 день и неделю назад, а также время дня – значения от 0 до 95, соответствующих номеру пятнадцатиминутки. Далее подбирались параметры моделей с целью достижения минимального значения MAE. Таким образом, было сформировано 5 прогнозов: модель persistence, и по две модели с разными конфигурациями параметров Prophet_1, GB_1, а также Prophet_2 и GB_2 с погодными регрессорами. Результаты точности для пяти недель

прогнозов приведены в таблице № 1, где настроенные модель GB_1 на 41%, а GB_2 на 26% лучше, чем persistence для A_E и D соответственно.

Таблица № 1

Результаты точности прогнозов MAE

№	Прогноз	persistence, W	GB_1, W	GB_2, W	Prophet_1, W	Prophet_2, W
1	A_E	4246	2514.5	2551	4357.8	3335
2	D	2526	2042	1858	2528	2829

Далее 5 недель прогнозов были разделены на валидационную и тестовую выборки периодом 2 и 3 недели. На валидационной выборке подбираются веса, с помощью которых происходит комбинирование и далее на тестовых данных рассчитывается эффект комбинирования. Веса рассчитываются как средняя ошибка в конкретное временное значение дня в диапазоне от 0 до 95.

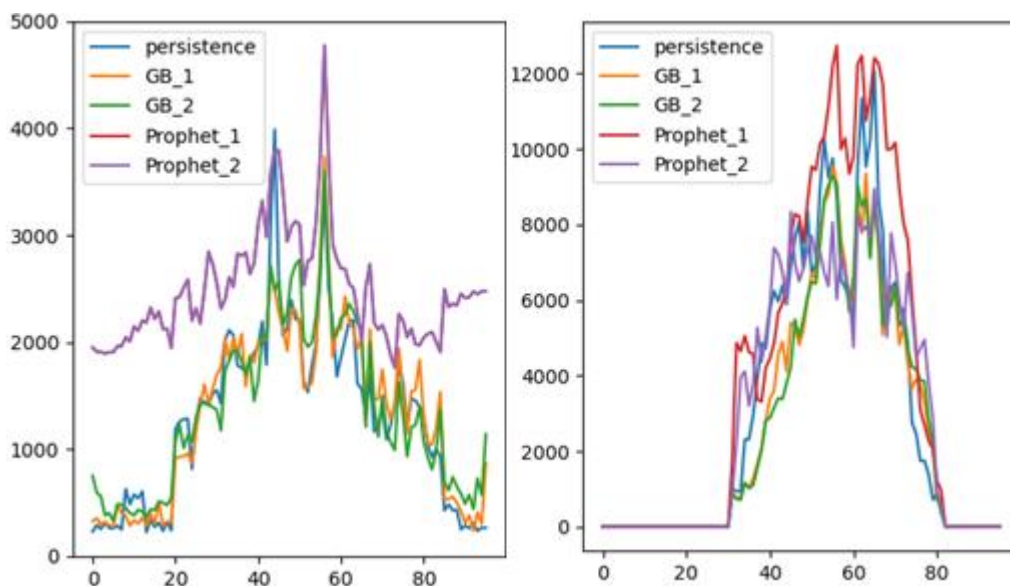


Рис. 1 . – Среднее значение ошибок (веса) для времени дня прогнозов demand и active energy

Ошибка для active energy ночью равна нулю, поскольку ночью энергия солнечными панелями не вырабатывается.

Далее рассматривается 3 вида комбинирования прогнозов [10]:

1. mean – среднее двух базовых прогнозов
2. weighted – где вклад прогноза с меньшей ошибкой больше, рассчитывается по формуле $(\frac{1}{w_{1,i}} * f_{1,j} + \frac{1}{w_{2,i}} * f_{2,j}) / (\frac{1}{w_{1,i}} + \frac{1}{w_{2,i}})$, где $w_{1,i}$ и $w_{2,i}$ веса, прогнозов $i = \overline{0, \dots, 95}$, $f_{1,j}$ и $f_{2,j}$ базовые прогнозы 1 и 2, где $j = \overline{0, \dots, n}$, где j – номер прогнозируемого значения временного ряда.
3. by_weight – берется значение базового прогноза у которого в текущий момент значение ошибки (веса) меньше.

Таким образом, были рассмотрены комбинации из двух прогнозов для пяти моделей persistence, Prophet_1, Prophet_2, GB_1 и GB_2. В результате лучшей комбинации методом комбинирования weighted удалось получить прогноз на 3% для demand и на 7% для active_energy точнее по метрике MAE в сравнении с лучшим прогнозом без комбинирования.

Литература

1. Wang J. , Zhu S. , Zhang W. , Lu H. Combined modeling for electric load forecasting with adaptive particle swarm optimization // Energy, 2010, vol.35, №4.
2. Xie J., Liu B., Lyu X., Hong T. Combining Load Forecasts from Independent Experts Energy, Conference: 2015 North American Power Symposium (NAPS) URL: [researchgate.net/ publication / 308847624_Combining_load_forecasts_from_independent_experts](https://researchgate.net/publication/308847624_Combining_load_forecasts_from_independent_experts).
3. Xie J., Hong T., Stroud J. Long-term Retail Energy Forecasting with Consideration of Residential Customer Attrition // IEEE Transactions on Smart Grid, 2015. p. 99.
4. Полуянович Н.К., Азаров Н.В. Синтез нейросетевой модели для прогнозирования термических процессов электроизоляционных материалов



силовых кабелей // Инженерный вестник Дона, 2020, №4. URL: ivdon.ru/ru/magazine/archive/N4y2020/6412.

5. Hyndman R., Koehler A. Another look at measures forecast accuracy // International Journal of Forecasting, 2006, pp. 679-688.

6. Thompson, P.A. An MSE statistic for comparing forecast accuracy across series // International J. Forecasting, 6, 1990, pp. 219–227.

7. Armstrong, J.S., Collopy, F. Error measures for generalizing about forecasting methods: empirical comparisons// International J Forecasting, 8, 1992 pp. 69–80.

8. Gardner, E. Evaluating forecast performance in an inventory control system // Management Science, 36, 1990, pp. 490–499.

9. Кузнецов Р.С., Тимофеев Ю.В. Механизмы вычислительного интеллекта при решении задачи автоматизации прогнозирования электроэнергии // Инженерный вестник Дона, 2020, №4. URL: ivdon.ru/ru/magazine/archive/N4y2020/6412/.

10. Smith D. Combination of forecasts in electricity demand prediction // International Journal of Forecasting, 8, 1989, pp. 349–356.

References

1. Wang J., Zhu S., Zhang W., Lu H. Energy, 2010, vol.35, №4.

2. Xie J., Liu B., Lyu X., Hong T. Energy Conference: 2015 North American Power Symposium (NAPS) URL: researchgate.net/publication/308847624_Combining_load_forecasts_from_independent_experts.

3. Xie J., Hong T., Stroud J. IEEE Transactions on Smart Grid, 2015. p. 99.

4. Polujanovich N.K., Azarov N.V. Inzhenernyj vestnik Dona, 2020, №4. URL: ivdon.ru/ru/magazine/archive/N4y2020/6412.

5. Hyndman R., Koehler A. International Journal of Forecasting, 2006, pp. 679-688.

6. Thompson, P.A. International J. Forecasting, 6,1990, pp. 219–227.



7. Armstrong, J.S., Collopy, F. International J Forecasting, 8,1992, pp. 69–80.
8. Gardner, E. Management Science, 36, 1990,pp. 490–499
9. Kuznetsov R.S., Timofeev U.V. Inzhenernyj vestnik Dona, 2020, №4. URL:
ivdon.ru/ru/magazine/archive/N4y2020/6412/.
10. Smith D. International Journal of Forecasting vol. 8, 1989, pp. 349–356.