

Исследование эффективности применения моделей на базе интервальных временных рядов с сезонностью для прогнозирования состояния сложных технических объектов

Р.Р. Имильбаев

УфаСистемаГаз, Уфа

Аннотация: В статье приводятся результаты исследования эффективности применения различных моделей, ориентированных на интервальные временные ряды (ИВР), для прогнозирования поведения параметров газораспределительных сетей с учетом реальных данных, получаемых в процессе их непрерывного контроля. Обосновывается необходимость учета в ИВР фактора сезонности, вызванного периодическими колебаниями уровня соответствующей величины. Выполняется сравнительный анализ свойств специальной интервальной модификации модели на базе экспоненциального сглаживания, нейросетевых и гибридных моделей прогнозирования применительно к ИВР с сезонностью, отмечаются их достоинства и недостатки.

Ключевые слова: интервальный временной ряд, модель на базе экспоненциального сглаживания, нейросетевая модель, долгосрочная краткосрочная память, гибридная модель.

Введение

Временной ряд, составленный из интервальнозначных величин (для упрощения далее именуется «интервальный временной ряд», или ИВР), – это последовательность интервальных данных, соотнесенных с определенными моментами времени. В свою очередь, операции над интервальными данными в последние годы принято рассматривать как действия, принадлежащие области анализа символьных данных (SDA – Symbolic Data Analysis) [1]. Эта область предполагает использование переменных, значения которых заданы с точностью до множеств (интервалов, гистограмм и др.). К построению ИВР можно прийти по результатам формализации задач в самых разных сферах деятельности. Например, применительно к финансовым рынкам известны так называемые «японские свечи» – интервалы, ограниченные максимальными и минимальными ценами на бирже в течение заданного периода времени [2]. В технике показана целесообразность использования ИВР при формировании прогнозов неблагоприятных воздействий на ветроэнергетические установки

[3]. Модели на базе ИВР могут быть применены даже для такой специфической цели, как описание надежности программного обеспечения

[4]. Наличие инструментов для работы с интервальными данными позволяет учесть неопределенность значений тех или иных параметров и принимать управленческие решения, справедливые для всего множества этих значений. Сказанное относится и к методам прогнозирования поведения переменных, информация о которых представлена в виде ИВР.

Хотя на сегодняшний день предложен ряд моделей прогнозирования на базе ИВР [5-8], они, тем не менее, пока не охватывают случай интервальных рядов с «сезонностью», вызванной периодическими колебаниями уровня соответствующей величины. Между тем, сезонность характерна для параметров многих систем и объектов. В частности, ее нетрудно заметить, наблюдая за изменением контролируемых параметров (входных и выходных давлений) компонентов газораспределительных сетей (ГС), так как потребление газа колеблется в зависимости от месяца и времени суток.

Автором данной публикации были разработаны специальные, рассчитанные на учет фактора сезонности модификации методов прогнозирования, нашедших ранее применение при анализе временных рядов с «точечными» (т.е. обычными, неинтервальными) величинами, а также некоторых алгоритмов, ориентированных на ИВР, но не принимавших сезонность во внимание. Полученные при этом модели относятся к различным классам: модель на базе экспоненциального сглаживания, две разновидности нейросетевых моделей, а также гибридные модели, предусматривающие комбинированную реализацию нейросетевых и иных алгоритмов. В настоящей статье приводятся результаты исследования эффективности предложенных подходов применительно к прогнозированию поведения параметров ГС с учетом реальных данных, собираемых в процессе их непрерывного контроля.

Подходы к представлению интервальных данных

Первоначально охарактеризуем наиболее часто применяемые подходы к представлению данных, которые сгруппированы в виде ИВР для последующего построения моделей прогнозирования.

Анализ по отдельно (независимо) рассматриваемым границам интервалов. В этом случае ИВР символически записывается как $I_i = [\underline{Y}(i), \bar{Y}(i)]$, где $i = 1, 2, \dots, n$, $\bar{Y}(i)$, $\underline{Y}(i)$ – нижняя и верхняя границы интервала значений параметра Y применительно к временной отметке с номером i . Далее все этапы анализа выполняются для двух независимых временных рядов (ВР), составленных соответственно из значений верхней и нижней границы интервала [5]:

$$I_i^U = \bar{Y}(i), I_i^L = \underline{Y}(i).$$

При этом используются классические модели прогнозирования, предназначенные для «точечных» данных.

Использование среднего значения интервала и отклонения от среднего. Здесь также предлагается рассматривать ИВР с точки зрения двух независимых множеств [6], которые в данном случае формируются несколько иным способом. Интервалы I_i представляются с помощью средних значений I_i^c и расстояний от середин интервалов («радиусов») I_i^r :

$$I_i^c = \frac{\bar{Y}(i) + \underline{Y}(i)}{2}, I_i^r = \frac{\bar{Y}(i) - \underline{Y}(i)}{2}.$$

На базе полученных ВР I_i^c , I_i^r с помощью классических моделей для «точечных» данных формируются прогнозы \hat{I}_i^c и \hat{I}_i^r . С использованием результатов этих прогнозов строится ИВР $\hat{I}_i = [\hat{I}_i^L, \hat{I}_i^U]$, $i = 1, 2, \dots, n$, в котором верхние \hat{I}_i^U и нижние \hat{I}_i^L границы прогнозируемых интервалов определяются с помощью соотношений:

$$\hat{I}_i^U = \hat{I}_i^c + \hat{I}_i^r, \hat{I}_i^L = \hat{I}_i^c - \hat{I}_i^r.$$

Применение векторного представления интервалов. В ряде работ, посвященных моделям прогнозирования на базе экспоненциального сглаживания применительно к ИВР [5, 6, 8], предлагается представлять интервальнозначные величины в виде векторов $\mathbf{I}_i = \left\| \underline{Y}(i), \overline{Y}(i) \right\|^T$, $i = 1, 2, \dots, n$, где T – символ транспонирования. Дальнейшие вычисления выполняются с матрицами и векторами по соответствующим математическим правилам.

Исследование моделей прогнозирования на базе ИВР

С целью определения модели, формирующей наиболее точные прогнозы на базе ИВР с сезонной составляющей, было проведено сравнительное экспериментальное исследование. Оценивалась эффективность предложенных автором «интервальных» модификаций нейросетевых моделей на основе классического многослойного персептрона (MLP¹) [6] и сети долгосрочной краткосрочной памяти (LSTM¹) [9], а также модели экспоненциального сглаживания с сезонной составляющей (HoltWinters¹) [10]. Опыт использования нейросетевых моделей для решения задач прогнозирования в системах мониторинга в целом известен [11], однако наличие интервальной неопределенности в задании исходных данных порождает немало особенностей в их реализации. Кроме того, рассматривались гибридные модели, в рамках которых варьировалось представление интервалов с помощью средних значений и отклонения от них («точечный» подход - Hybrid) либо в виде векторов («векторный» подход - Hybrid¹). Точность формирования прогнозов характеризовалась перечисленными ниже показателями.

MSE^I (mean square error) – среднеквадратичная мера ошибки, используемая для выявления больших неточностей в работе модели:

$$MSE^I = \frac{1}{2m} \sum_{j=1}^m [(I_j^U - \hat{I}_j^U)^2 + (I_j^L - \hat{I}_j^L)^2],$$

где m – количество сравниваемых значений.

$MAPE^I$ (interval mean absolute percentage error) – «интервальная» версия средней абсолютной ошибки в процентах. Позволяет оценить ошибку прогноза в процентах относительно реальных значений:

$$MAPE^U = \frac{100}{m} \sum_{j=1}^m \left| \frac{I_j^U - \hat{I}_j^U}{I_j^U} \right|, \quad MAPE^L = \frac{100}{m} \sum_{j=1}^m \left| \frac{I_j^L - \hat{I}_j^L}{I_j^L} \right|, \quad MAPE^I = \frac{MAPE^U + MAPE^L}{2},$$

где $MAPE^U$ и $MAPE^L$ – значения средней ошибки для верхних и нижних границ интервалов соответственно.

ARV^I (interval average relative variance) – средняя относительная дисперсия интервалов, которая позволяет оценить результаты прогнозирования модели исходя из предположения, что значения в будущем будут соответствовать медиане значений ряда на предыдущих шагах:

$$ARV^I = \frac{\sum_{j=1}^m (\mathbf{I}_j - \hat{\mathbf{I}}_j)^T (\mathbf{I}_j - \hat{\mathbf{I}}_j)}{\sum_{j=1}^m (\mathbf{I}_j - \bar{\mathbf{I}})^T (\mathbf{I}_j - \bar{\mathbf{I}})} = \frac{\sum_{j=1}^m (I_j^U - \hat{I}_j^U)^2 + \sum_{j=1}^m (I_j^L - \hat{I}_j^L)^2}{\sum_{j=1}^m (I_j^U - \bar{I}^U)^2 + \sum_{j=1}^m (I_j^L - \bar{I}^L)^2},$$

где \bar{I}^U и \bar{I}^L – медианы (средние значения) временных рядов, составленных применительно к верхней и нижней границам интервала соответственно, $\bar{\mathbf{I}} = [\bar{I}^L, \bar{I}^U]$.

На основе перечисленных показателей предлагается формировать комплексный нормированный показатель. Для этого значения каждого показателя сначала приводятся к диапазону от 0 до 1:

$$x_{норм}^j(x^j) = 1 - \frac{x^j - x_{мин}^j}{x_{макс}^j - x_{мин}^j}, \quad j = 1, 2, 3,$$

где x^j – текущее значение j -го показателя точности моделей (при этом x^1 соответствует $MAPE^I$, x^2 – ARV^I и x^3 – MSE^I), $x_{мин}^j$ и $x_{макс}^j$ – его наименьшее и наибольшее значения.

Далее комплексный показатель находится в виде линейной свертки:

$$R = \frac{1}{3} \sum_{j=1}^3 x_{\text{норм}}^j (x^j). \quad (1)$$

Методика исследования эффективности построения прогнозов

С целью построения ИВР использовались эксплуатационные данные относительно величин выходного низкого давления газа для газораспределительных пунктов в составе контролируемой ГС (максимальные и минимальные значения этого давления с 50-ти объектов мониторинга за каждый час наблюдений в течение недели). Каждый отдельно взятый набор данных из исходной выборки был предварительно разделен на две части. Первая часть (85% данных) включала показания за 6 суток; соответствующие данные применялись в качестве обучающих с целью поиска оптимальных весовых коэффициентов в алгоритмах моделей. Оставшиеся 15% данных (показания за последние сутки) использовались в качестве контрольных для определения итоговых показателей точности. Прогноз осуществлялся на 3 часа (краткосрочный, или оперативный) и 24 часа вперед (долгосрочный, или стратегический).

Результаты исследования эффективности построения прогнозов

Средние значения показателей эффективности применения моделей для краткосрочного (оперативного) прогноза приведены в таблице 1.

Таблица № 1

Средние показатели точности для краткосрочных прогнозов

Показатель	MLP ¹	LSTM ¹	HoltWinters ¹	Hybrid	Hybrid ¹
<i>MAPE¹</i>	1,1703	1,1268	1,0052	0,9829	1,2513
<i>ARV¹</i>	1,3107	1,0819	0,9550	0,9143	1,5687
<i>MSE¹</i>	18,998	18,276	13,767	12,081	18,909

Как следует из результатов, наилучшие средние показатели относительно других моделей прогнозирования имеет гибридная модель на основе «точечного» подхода. Тем не менее, если вычислить комплексный показатель точности (1), то можно установить, что модель HoltWinters¹ имеет

незначительное «отставание» от гибридной модели, имеющей максимально возможное значение. Рис. 1 (левая диаграмма) показывает, что только интервальная версия «экспоненциальной» модели и гибридная модель на основе «точечного» подхода имеют значения показателя ARV^I меньше единицы. Это свидетельствует о более точных прогнозах по сравнению с наивной моделью прогнозирования на основе средних значений.

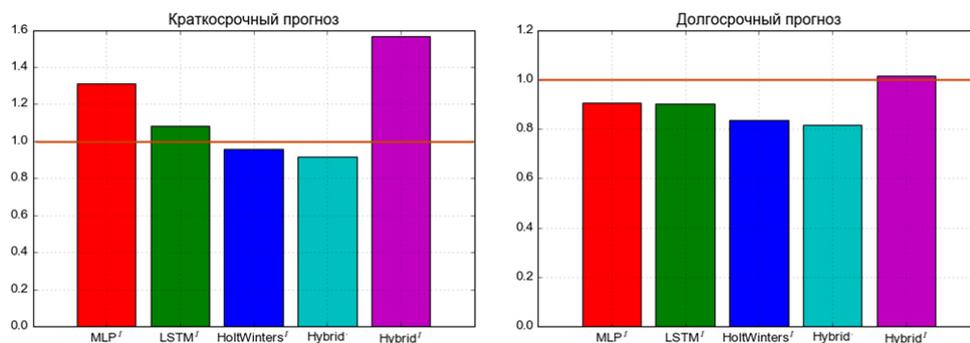


Рис. 1. – Сравнение средних значений показателя ARV^I

Средние значения показателей эффективности для долгосрочного (стратегического) прогноза приведены в таблице 2.

Таблица № 2

Средние показатели точности для долгосрочных прогнозов

Показатель	MLP ^I	LSTM ^I	HoltWinters ^I	Hybrid	Hybrid ^I
$MAPE^I$	1,3571	1,2904	1,1938	1,2115	1,3381
ARV^I	0,9063	0,8997	0,8364	0,8140	1,0163
MSE^I	24,166	21,884	19,189	19,419	23,168

Из представленных результатов видно, что интервальная версия «экспоненциальной» модели HoltWinters^I при построении долгосрочных прогнозов имеет незначительное превосходство в точности относительно гибридной модели на основе «точечного» подхода. Как показывает правая диаграмма на рис. 1, все модели прогнозирования, за исключением гибридной модели на основе «векторного» подхода, имеют значения показателя ARV^I ниже единицы. Таким образом, можно утверждать, что гибридная модель прогнозирования на основе «векторного» подхода строит

менее точные прогнозы относительно наивной модели прогнозирования на основе средних значений при формировании прогнозов любой глубины.

Значения комплексного показателя точности (1) для краткосрочных и долгосрочных прогнозов, полученных с помощью исследуемых моделей, а также итоги ранжирования этих моделей по данному показателю представлены в сводной таблице 3.

Таблица № 3

Итоги ранжирования моделей прогнозирования по комплексному показателю

Прогноз	1	2	3	4	5
Краткосрочный	Hybrid (1,000)	HoltWinters ¹ (0,870)	LSTM ¹ (0,437)	MLP ¹ (0,231)	Hybrid ¹ (0,004)
Долгосрочный	HoltWinters ¹ (0,963)	Hybrid (0,948)	LSTM ¹ (0,481)	MLP ¹ (0,181)	Hybrid ¹ (0,105)

Из этой таблицы видно, что гибридная модель на основе «точечного» подхода точнее других моделей формирует краткосрочные прогнозы. В то же время, интервальная модификация «экспоненциальной» модели (HoltWinters¹) точнее остальных моделей строит долгосрочные прогнозы. При этом можно отметить существенное превосходство в точности двух вышеназванных моделей прогнозирования по отношению к другим моделям, что отражено на диаграммах их сравнения по комплексному показателю (Рис. 2).

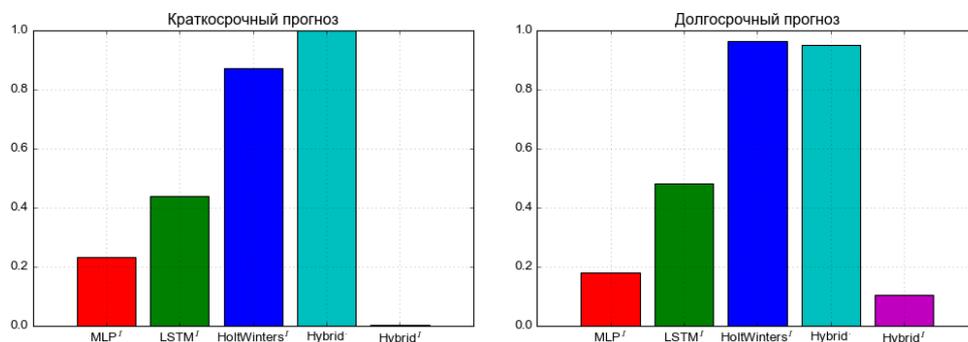


Рис. 2. – Сравнение моделей по комплексному показателю точности

Полученные выводы дают возможность подойти с обоснованных позиций к формированию модулей прогнозирования в составе систем

телеметрического контроля ГС [12]. В то же время, они могут быть применены и при построении систем мониторинга более широкого назначения [13].

Литература

1. Billard L., Diday E. Symbolic Data Analysis: Conceptual Statistics and Data Mining. Chichester: Wiley, 2006. 330 p.
2. Kamo T. Hybrid Approach to the Japanese Candlestick Method for Financial Forecasting // Expert Systems and Applications. 2009. Vol.36. No.3. pp. 5023-5030.
3. Ak R., Vitelli V., Zio E. An Interval-Valued Neural Network Approach for Uncertainty Quantification in Short-Term Wind Speed Prediction // IEEE Transactions on Neural Networks and Learning Systems. 2015. Vol.26. No.11. pp. 2787-2800.
4. Krymsky V.G., Ivanov I.V. Application of Interval-Valued Probabilities and Unified Scheme of Non-Homogeneous Poisson Process Models to Software Failure Prognostics // Safety and Reliability of Complex Engineered Systems. – Podofilini et al. (Eds.).- London: Taylor & Francis Group, 2015. pp. 2403 – 2411.
5. Maia A.L.S, De Carvalho F.A.T., Ludermir T.B. Forecasting Models for Interval-Valued Time Series // Neurocomputing, 2008. Vol.71. No. 16-18. pp. 3344-3352.
6. Maia A.L.S., De Carvalho F.A.T. Holt's Exponential Smoothing and Neural Network Models for Forecasting Interval-Valued Time Series // International Journal of Forecasting, 2011. Vol.27. No.3. pp. 740-759.
7. Arroyo, J., San Roque A.M., Mate C., Sarabia A. Exponential Smoothing Methods for Interval Time Series // Proceedings of the 1st European Symposium on Time Series Prediction. - Espoo: Multiprint Oy / Otamedia. 2007. pp.231–240.

8. Xiong, T., Li C., Bao Y. Interval-valued Time Series Forecasting Using a Novel Hybrid Holt^I and MSVR Model // *Economic Modelling*, 2017. Vol. 60. Issue C. pp. 11-23.

9. Hochreiter, S., Schmidhuber J. Long Short-Term Memory // *Neural Computation*, 1997. Vol. 9. pp. 1735–1780.

10. Имильбаев Р.Р., Крымский В.Г., Юнусов А.Р. Использование интервальных временных рядов для прогнозирования состояния газораспределительной сети // *Электротехнические и информационные комплексы и системы*, 2016. Т.12. №4. С. 62-72.

11. Плуготаренко Н.К., Варнавский А.Н. Применение нейронных сетей для построения модели прогнозирования состояния городской воздушной среды // *Инженерный вестник Дона*. 2012. №4 (часть 2). URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1351.

12. Крымский В.Г., Жалбеков И.М., Имильбаев Р.Р., Юнусов А.Р. Автоматизация управления технологическими процессами в газораспределительных сетях: проблемы, тенденции и перспективы // *Электротехнические и информационные комплексы и системы*. 2013. Т.9. №2. С.70-79.

13. Стуженко Н.И., Илиев А.Г., Стуженко И.А., Костромина Е.И., Шеметов А.И. Теоретические основы построения региональных систем мониторинга // *Инженерный вестник Дона*, 2015. №4. URL: ivdon.ru/ru/magazine/archive/n4y2015/3282.

References

1. Billard L., Diday E. *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Chichester: Wiley, 2006. 330 p.

2. Kamo T. Hybrid Approach to the Japanese Candlestick Method for Financial Forecasting. *Expert Systems and Applications*. 2009. Vol.36. No.3. pp. 5023-5030.

3. Ak R., Vitelli V., Zio E. An Interval-Valued Neural Network Approach for Uncertainty Quantification in Short-Term Wind Speed Prediction. IEEE Transactions on Neural Networks and Learning Systems. 2015. Vol.26. No.11. pp. 2787-2800.
 4. Krymsky V.G., Ivanov I.V. Application of Interval-Valued Probabilities and Unified Scheme of Non-Homogeneous Poisson Process Models to Software Failure Prognostics // Safety and Reliability of Complex Engineered Systems. – Podofilini et al. (Eds.). London: Taylor & Francis Group, 2015. pp. 2403 – 2411.
 5. Maia A.L.S, De Carvalho F.A.T., Ludermir T.B. Forecasting Models for Interval-Valued Time Series. Neurocomputing, 2008. Vol.71. No. 16-18. P. 3344-3352.
 6. Maia, A.L.S., De Carvalho F.A.T. Holt's Exponential Smoothing and Neural Network Models for Forecasting Interval-valued Time Series. International Journal of Forecasting, 2011. Vol. 27. No.3. pp. 740-759.
 7. Arroyo, J., San Roque A.M., Mate C., Sarabia A., Exponential Smoothing Methods for Interval Time Series. Proceedings of the 1st European Symposium on Time Series Prediction, 2007. pp.231–240.
 8. Xiong, T., Li C., Bao Y. Interval-valued Time Series Forecasting Using a Novel Hybrid Holt¹ and MSVR Model. Economic Modelling, 2017. Vol.60. Issue C. pp.11-23.
 9. Hochreiter, S., Schmidhuber J. Long Short-term Memory. Neural Computation, 1997. Vol.9. pp. 1735–1780.
 10. Imilbaev R.R., Krymsky V. G., Yunusov A.R. Electrical and Data Processing Facilities and Systems (Rus), 2016, No. 4. pp. 62–72.
 11. Plugotarenko N.K. Varnavskiy A.N. Inzhenernyj vestnik Dona (Rus), 2012, №4 (part 2). URL: ivdon.ru/ru/magazine/archive/n4p2y2012/1351.
-



12. Krymsky V. G., Zhalbekov I.M., Imilbaev R.R., Yunusov A.R. Electrical and Data Processing Facilities and Systems (Rus), 2013, No. 2, pp. 70–79.
13. Stuzhenko N.I., Iliev A.G., Stuzhenko I.A., Kostromina E.I., Chemetov A.I. Inzhenernyj vestnik Dona (Rus), 2015, №4. URL: ivdon.ru/en/magazine/archive/n4y2015/3282.