

ПОИСК ПАРЕТО-ОПТИМАЛЬНЫХ ГРАНИЦ ИНТЕРВАЛЬНЫХ ПРОГНОЗНЫХ МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ СЛОЖНЫХ СИСТЕМ

Червяков А.А.

Федеральное казначейство Министерства финансов Российской Федерации, Москва, Россия
achervyakov@roskazna.ru

Гусев А.А.

Национальный исследовательский университет ИТМО, Санкт-Петербург, Россия
aagusev@itmo.ru

Тарасов А. В., Никольчев Е.В.

МИРЭА — Российский технологический университет, Москва, Россия
nikulchev@mirea.ru

Аннотация. Информационно-аналитическое обеспечение цифровых платформ оперирует с большим количеством прогнозов, в том числе, построенных в форме интервалов, к которым предъявляются высокие требования. В работе обсуждается подход к поиску границ интервальных прогнозных оценок на основе парето-оптимальных решений в условиях заданных критериев.

Ключевые слова: информационно-аналитическое обеспечение, интервальный прогноз, модели LUBE.

Введение

Крупномасштабные цифровые платформы и системы аккумулируют сотни тысяч данных от систем, имеющих динамическую и структурную сложность. Основой принятия решений, формирования законов управления, создание информационно-аналитических отчетов является прогнозирование поведения динамических систем, полученных на основе математических, имитационных, интеллектуальных моделей. Одной из часто встречающихся особенностей крупномасштабных систем является не столько точности прогноза, сколько достоверность и надежность границ прогнозных оценок. Например, в задаче резервирования ресурсов, если реальное значение будет выше верхней оценки прогноза (ресурса зарезервировано недостаточно), то система не сможет функционировать. Таким образом, построение моделей, обеспечивающих построение надежных оценок, прогнозируемых данных имеет важное значение для создания информационно-аналитического обеспечения крупномасштабных систем.

Задача прогнозирования представляет собой задачу поиска значений наблюдаемого процесса в будущем на количество отсчетов (для дискретных систем, временных рядов) или интервал времени (для непрерывных систем), называемый горизонтом прогноза. В отличие от точечного, интервальный прогноз (ПИ) строит числовое значение диапазона, в котором будет находиться система в следующий момент времени [1, 2]. При использовании ПИ происходит размытие точности прогноза, но в тоже время снижается степень неопределенности, и обеспечивается робастность модели по выходу. На основе исторических данных строится прогнозный интервал, представляющих собой два непересекающихся графика на заданный горизонт прогноза, первый график – верхняя граница, второй график – нижняя граница прогноза. Ищется два преобразования – исходного графика в график верхней границы и график нижней границы [3]. В полученный диапазон, может быть включено бесконечное количество функций, в том числе и соответствующий реальным данным.

В настоящее время ПИ, как наиболее интуитивно понятный подход к моделированию, привлек большое внимание, поскольку такой подход является интуитивно понятным [4]. Например, метеопрогноз представляется, традиционно, как диапазон прогнозируемых температур на каждый момент времени или по суткам. Методы широко применяется как для погодных данных, прогнозирования ветра [5], в энергетике [6] и в других прикладных областях [3]. Общность применения можно охарактеризовать как: (1) задача обладает ярко выраженными трендами, сезонными компонентами; (2) является портфелем нескольких сигналов; (3) к прогнозам, предъявляются требования к достоверной надежности верхней или нижней границы.

В последние годы методы вероятностного прогнозирования широко изучались для эффективной количественной оценки неопределенностей. Вероятностный прогноз генерирует функции плотности вероятности [7], квантили [8] или интервалы [9] в прикладных задачах прогнозирования. Традиционные методы интервального прогнозирования, такие как нечеткий вывод [10], функция бета-распределения [11] и гауссовский процесс, авторегрессионные интегрированные модели скользящего среднего, логарифмически нормальные процессы для получения вероятностных

прогнозов [2]. Особенности методов является высокая математическая вычислительная сложность, и гипотезы о видах распределения.

Одним из ключевых методов является непараметрический метод нахождения оценок нижней и верхней границ ПИ, которые получил название LUBE (Lower Upper Bound Estimation) [12, 13]. Согласно методу LUBE ставится многокритериальная задача с противоречивыми критериями качества, компромисс между которыми определяется фронтом множества Парето, который ищется численными методами.

В настоящей статье рассматривается поиск парето-оптимального решения задачи поиска границ прогнозного интервала для динамически-сложных систем.

1. Критерии качества

В результате применение LUBE формируются верхняя и нижняя границы ПИ. Таким образом, модель LUBE можно рассматривать как прямое сопоставление входных данных с ПИ без каких-либо предположений о распределении данных. Суть метода LUBE заключается условиями задания противоречивых критериев качества, например: с одной стороны необходимо обеспечить гарантированное попадание прогнозируемых значений в прогнозируемый интервал, с другой — нельзя расширять интервал до бесконечности, оценки должны быть конечны и достаточно узки. Рассмотрим некоторые критерии оценки ПИ [3].

Критерий PISCP — указывает вероятность того, что будущие значения будут охвачены нижними и верхними границами. Очевидно, что большее значение PISCP означает, что построенные ПИ могут более точно отражать будущие неопределенности. PISCP определяется следующим образом:

$$PISCP = \frac{1}{N} \sum_{i=1}^N a_i, \quad (1)$$

где N — объем выборки данных, а a_i представляет собой двоичную переменную, определяемую по формуле:

$$a_i = \begin{cases} 1, & y_i \notin [y_{i-}, y_{i+}]; \\ 0, & y_i \in [y_{i-}, y_{i+}]; \end{cases}$$

значения y_{i-}, y_{i+} предполагаемые верхняя и нижняя границы.

Чтобы гарантировать точность ПИ, требуется, чтобы PISCP не превышал заранее определенный уровень. В целом PISCP считается очень важным показателем ПИ, который представляет точность ПИ, то есть вероятность того, что целевое значение перекрывается верхним и нижние границы ПИ. Хотя PISCP является ключевым показателем точности ПИ, эффективность PISCP нивелируется расширением интервала, поскольку большой PISCP может быть легко получен с чрезвычайно широким интервалом прогноза. При этом широкие интервалы не могут обеспечивать значимую информацию или обеспечивать эффективное управление или мониторинг работы системы.

Критерий PINAW вводится для оценки эффективности ПИ:

$$PINAW = \frac{1}{N \cdot W} \sum_{i=1}^N (y_{i+} - y_{i-}), \quad (2)$$

где W — диапазон ширины ПИ. Целью нормирования является объективная оценка ширины ПИ от абсолютных значений.

Критерий CWC предлагается путем свертки двух критериев PISCP и PINAW:

$$CWC = PINAW(1 + \gamma(PISCP)^{-\eta(PISCP-\mu)}), \quad (3)$$

где γ представляет собой двоичное значение, определяемое как:

$$\gamma = \begin{cases} 0, & \text{if } PISCP \geq \mu; \\ 1, & \text{if } PISCP \leq \mu. \end{cases}$$

Здесь два параметра γ, μ для CWC отражают требование вероятности ПИ, которое может быть рассчитано по предварительно определенному доверительному уровню $(1 - \alpha)$; η — штрафной коэффициент, когда полученный PI не может удовлетворить требованию вероятности покрытия.

Критерий Винклера (WS) — рассчитывается как взвешенная сумма ширины и ошибки оценки ПИ. Например, выход реального значения за границы интервала может привести к поломке оборудования, когда границы соответствуют критичным значениям технологий. Поэтому для обеспечения надежности при построении ПИ критерий WS позволяет учитывать погрешность оценки. Обычно при заданном уровне достоверности ПИ с небольшим абсолютным значением WS имеют высокое качество. Однако WS не различает вклад между средней шириной и ошибкой оценки.

В некоторых реальных приложениях, лица, принимающие решения, хотят знать ошибку оценки ПИ, чтобы оценить операционный риск. Чтобы смягчить вышеуказанные недостатки, в этом исследовании предлагается новый индекс оценки, называемый

Критерий PИЕЕ определяется как ошибка оценки ПИ:

$$PИЕЕ = \frac{1}{N \cdot W} \sum_{i=1}^N E_i, \quad (4)$$

где

$$E_i = \begin{cases} (y_i - y_{i-}), & \text{if } y_i \geq y_{i+} \\ 0, & \text{if } y_{i-} \leq y_i \leq y_{i+} \\ (y_{i-} - y_i), & \text{if } y_{i-} \geq y_i \end{cases}$$

Основываясь на PИЕЕ, можно измерить ошибку оценки ПИ отдельно со средней шириной ПИ. Таким образом, лица, принимающие решения, могут снизить риск, выходящий за рамки ПИ. Хотя PИЕЕ предлагается для измерения точности ПИ, эффективность ПИ игнорируется. Например, небольшой PИЕЕ может быть достигнут при использовании широкого ПИ. В LUBE критерий PИЕЕ оптимизируется вместе с PINAW для получения высококачественных ПИ.

Кроме критериев (1)–(4) существуют и другие различные критерии, и их сочетания. В работе будем искать парето-оптимальное решение противоречивой задачи поиска границ ПИ с использованием критериев (1) и (2).

2. Методы и результаты

В качестве временного ряда с сложной динамикой рассмотрим известный и широкодоступный датасет — индекса S&P 500 с суточной стоимостью на момент закрытия торгов.

Применим разложение на сезонные компоненты и построим прогнозные модели ARIMA в форме интервалов, используя функции Python.

Методика построения прогнозных ПИ имеет вид:

1. Сезонное разложение временного ряда.
2. Построение модели ARIMA с автоматическим поиском коэффициентов.
3. Построение доверительных интервалов для каждого из временного ряда.
4. Построение прогнозирования (пример для одной из компонент приведен на рис. 1)

В данном случае метод выдает прогнозные значения в форме расширяющихся интервалов, причем верхняя и нижняя границы имеют зеркальную симметрию относительно точек прогноза.

Таким образом, имеем три ряда, для каждого можно построить свою интервальную модель. Шумовая компонента служит для быстрых прогнозов и учитывая неопределенность достаточно широкий диапазон, чтобы обеспечить качественную ПИ. Для трендового графика можно использовать узкие интервалы.

Для формирования границ ПИ рассмотрим задачу поиска компромиссного парето-оптимального решения по критериям, заданных уравнением (3). Фронт множества Парето будем строить на основе эволюционного алгоритма [14], для этого выделим две обещающих выборки, для которых оценим значение CWC_1 и CWC_2 . Результат построения фронта Парето приведен на рис. 2. Решение МР - это точка на фронте Парето CWC_1 и CWC_2 .

В соответствии с [14] выбирается решение соответствующее точки, принадлежащих фронту Парето, которое характеризуется самым высоким PИСР, определяемым в соответствии с (1). Результат построения прогнозной модели приведен на рис. 3.

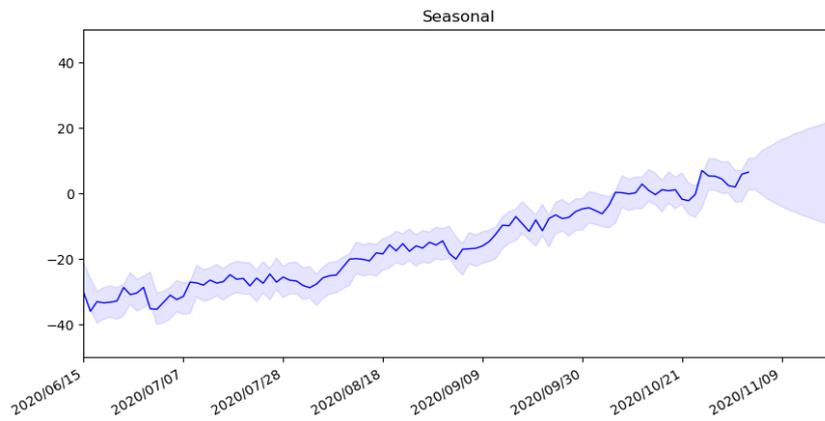


Рис. 1. Прогнозный интервал для трендовой компоненты

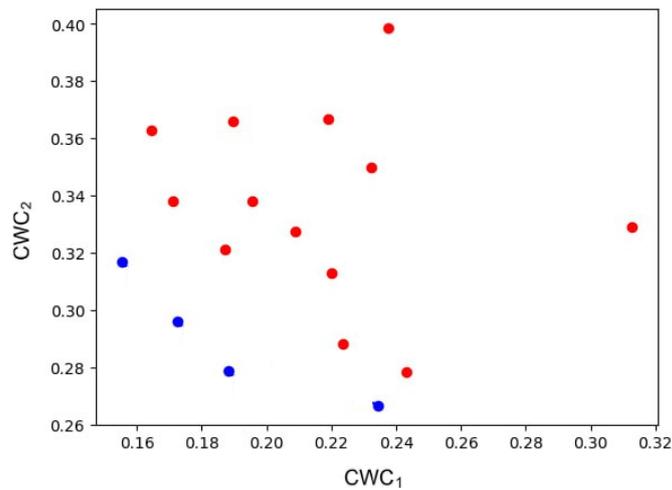


Рис. 2. Фронт множества Парето (синие точки)

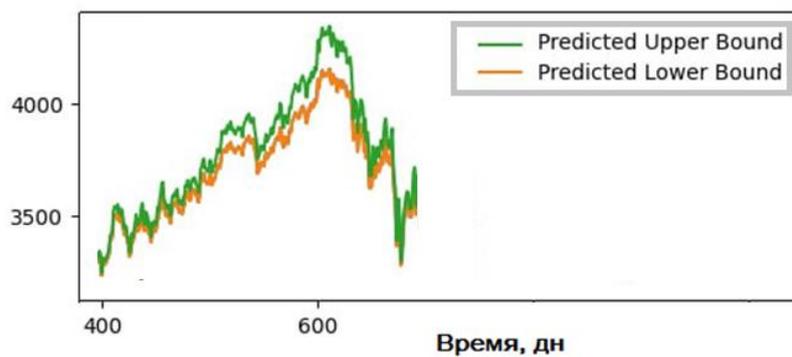


Рис. 3. Прогнозная модель по методу LUBE

3. Заключение

В настоящей статье рассмотрено решение задачи поиска компромиссного парето-оптимального решения значений границ прогнозного интервала для динамически-сложных систем в условиях заданных противоречивых критерия качества. Приведенные результаты демонстрируют эффективность разработанного инструмента и перспективность его использования в информационно-аналитическом обеспечении крупномасштабных цифровых платформ и систем.

Литература

1. Червяков А.А., Никольцев Е.В. Робастное интервальное прогнозирование временных рядов // International Journal of Open Information Technologies. — 2023. — Т. 11, №. 4. — С. 122-128.

2. *Никульчев Е.В., Червяков А.А.* Построение робастных интервальных моделей прогнозирования динамики структурно-сложной системы // Труды НГТУ им. П.Е. Алексеева. — 2023. — № 1. — С. 33-41.
3. *Nikulchev E., Chervyakov A.* Prediction Intervals: A Geometric View // *Symmetry*. — 2023. — Vol. 15. N. 4. — P. 781.
4. *Nikulchev E., Chervyakov A.* Development of Trading Strategies Using Time Series Based on Robust Interval Forecasts // *Computation*. — 2023. — Vol. 116 N. 5. — P. 99.
5. *Zhou M., Wang B., Guo S., Watada J.* Multi-objective prediction intervals for wind power forecast based on deep neural networks // *Information Sciences*. — 2021. — Vol. 550. — P. 207-220.
6. *Cao Z., Wan C., Zhang Z., Li F., Song Y.* Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting // *IEEE Transactions on Power Systems*. — 2019. — Vol.35, N. 3. — P. 1881-1897
7. *Guan Y., Li D., Xue S., Xi Y.* Feature-fusion-kernel-based Gaussian process model for probabilistic long-term load forecasting // *Neurocomputing*. — 2021. — Vol. 426. — P. 174-184.
8. *Zhang W., Quan H., Srinivasan D.* An improved quantile regression neural network for probabilistic load forecasting // *IEEE Trans. Smart Grid*. Jul. — 2019. — Vol, 10, N. 4. — P. 4425–4434.
9. *Wan C., Xu Z., Pinson P., Dong Z.Y., Wong K.P.* Probabilistic forecasting of wind power generation using extreme learning machine // *IEEE Trans. Power Syst.* — 2014. — Vol. 29, N. 3. — P. 1033–1044.
10. *Marín L. G., Cruz N., Sáez D., Sumner M., Núñez, A.* Prediction interval methodology based on fuzzy numbers and its extension to fuzzy systems and neural networks // *Expert Systems with Applications*. — 2019. — Vol. 119. —P. 128-141.
11. *Yuan X., Chen C., Jiang M., Yuan Y.* Prediction interval of wind power using parameter optimized beta distribution based LSTM model // *Appl. Soft. Comput.* — 2019. — Vol. 82. — P. 105550.
12. *Khosravi A., Nahavandi S., Creighton D., Atiya A.F.* Lower upper bound estimation method for construction of neural network-based prediction intervals // *IEEE Trans. Neural Netw. Learn. Syst.* — 2011. — Vol. 22, N. 3. — P. 337–346.
13. *Quan H., Srinivasan D., Khosravi A.* Short-term load and wind power forecasting using neural network-based prediction intervals // *IEEE Trans. Neural Netw. Learn. Syst.* — 2014. — Vol., 25, N. 2. — P. 303–315.
14. *Taormina R., Chau K. W.* ANN-based interval forecasting of streamflow discharges using the LUBE method and MOFIPS // *Engineering Applications of Artificial Intelligence*. — 2015. — Vol. 45. — P. 429-440.