

Н. Д. Уваров¹, М. П. Кузнецов², А. С. Малькова³, К. В. Рудаков⁴, В. В. Стрижов⁵

ВЫБОР СУПЕРПОЗИЦИИ МОДЕЛЕЙ ПРИ ПРОГНОЗИРОВАНИИ ГРУЗОВЫХ ЖЕЛЕЗНОДОРОЖНЫХ ПЕРЕВОЗОК*

Рассматривается задача выбора оптимальной прогностической системы моделей для краткосрочного прогнозирования объемов железнодорожных грузовых перевозок. Исторические данные представляют собой временные ряды, отражающие объемы перевозок различных типов грузов между станциями. Свойствами данных временных рядов являются их высокая волатильность, зашумленность и нестационарность. Для построения прогнозов предлагается создать систему, которая бы выбирала оптимальную суперпозицию прогностических моделей с учетом особенности исторических данных. В качестве возможных моделей для включения в суперпозицию рассматриваются модели скользящего среднего, экспоненциального и ядерного сглаживания, векторной авторегрессии, интегрированные модели авторегрессии и скользящего среднего, а также метод Кростона и LSTM.

Ключевые слова: временные ряды, прогнозирование, суперпозиция, прогностические модели, ассимметричное распределение, регрессионные остатки.

1. Введение. Российские железные дороги играют ключевую роль в экономике страны. Железнодорожный транспорт занимает второе место, после трубопроводного, по грузообороту. На сегодняшний день около 80% своей прибыли РЖД получают именно от грузовых перевозок [1]. Поэтому перед РЖД ставится задача по привлечению грузов с конкурирующих видов транспорта. В связи с этим прогнозирование объемов перевозимых грузов имеет большое практическое значение для повышения экономической эффективности загрузки железнодорожных путей и их рационального использования. В текущей экономической ситуации дальнейшее использование экстенсивных методов развития производства и сферы услуг не приводит к желаемой экономической отдаче. Использование интенсивного подхода, связанного с привлечением современных научных методов анализа данных и оптимизации, должно улучшить показатели эффективности работы РЖД и увеличить величину добавленной стоимости услуг железнодорожных грузоперевозок.

¹Факультет УПМ МФТИ, студент, E-mail: nikita.uvarov@phystech.edu

²Yahoo! Research, исследователь, E-mail: mikhail.kuznecov@phystech.edu

³Факультет УПМ МФТИ, студент, E-mail: malkova@phystech.edu

⁴Факультет ВМК МГУ, проф., д.ф.-м.н., акад. РАН, E-mail: rudakov@ccas.ru

⁵ФИЦ ИУ РАН, проф., д.ф.-м.н., E-mail: strijov@ccas.ru

*Работа выполнена при финансовой поддержке РФФИ, проект 17-20-01212.

Не все временные ряды подчиняются гипотезе о нормальном распределении шумовой компоненты [2]. Для повышения качества прогноза полезно использование отдельной модели. Так, в [3] для прогнозирования остатков строится марковская цепь, что улучшает точность прогноза электроэнергии, генерируемой ветряной фермой, на 10%. Распространен подход коррекции Фурье-спектра остатков во временных рядах туристической отрасли [4].

Целью работы является построение прогностической модели, способной описывать временные ряды с асимметрично распределенным шумом. Предлагается использовать идею суперпозиции двух моделей, изложенную в [5]. Одновременно используются две модели: одна из них предсказывает основное, обычно гладкое, поведение временного ряда, а вторая корректирует прогноз с целью минимизации ошибки от наличия шума. В качестве обучающей выборки для второй модели выступают ошибки прогнозирования первой модели. При прогнозировании на более чем один отсчет в будущее, возникает возможность использования для построения прогноза первой модели как ее собственного прогноза, так и скорректированного прогноза суперпозиции. В вычислительном эксперименте на исторических данных качество прогноза суперпозиции сравнивается с качеством прогноза базовых алгоритмов, а на синтетических данных исследуется зависимость качества прогноза от меры асимметрии шума.

Исследуемыми подходами к прогнозированию временных рядов являются модели авторегрессии и скользящего среднего ARIMA [6] анализа спектральных компонент SSA [7], а также подходы на основе нейронных сетей, особенно LSTM [8]. Из более простых подходов исследуются модели экспоненциального [9] и ядерного [10] сглаживания. Перечисленные алгоритмы составляют семейство базовых алгоритмов, используемых для построения суперпозиций в данной работе. Оптимальная комбинация алгоритмов зависит от решаемой задачи, и для ее нахождения перебираются все возможные пары моделей: строится матрица качества прогноза суперпозиций.

Исторические данные, предлагаемые для исследования, представляют собой временные ряды объемов грузоперевозок по российским железным дорогам [11]. Имеется информация о дате погрузки товара, станции отправления, станции назначения, количестве вагонов, коде перевозимого груза, роде вагонов, в которых этот груз перевозился, суммарном весе груза в тоннах и признаке маршрутной отправки. Ряды отличаются высокой волатильностью, неравномерностью, а также зашумленностью. Для некоторых временных рядов свойственна нестационарность. Кроме того, между некоторыми парами станций перевозки производятся крайне нерегулярно, и число нулевых измерений в таких парах превышает число ненулевых. Проверка временных рядов на стационарность выполняется с помощью теста Дики-Фуллера.

Данные о посуточной загруженности железнодорожных путей представлены в [12], таблицей в которой числится информация о времени перевозки, станциях, между которыми выполняются перевозки, а также коде и объеме перевозимого груза. представлена годовая история перевозок между 78 регионами и 4000 станциями по 43 наименованиям, среди которых нефть, руда, торф, автомобили, хлопок, сахар, зерно.

Помимо временных рядов РЖД, в эксперименте участвуют временные ряды пассажирских авиаперевозок [13], цены на электроэнергию в Германии [14], а также синтетическая выборка, представляющая собой синусоидальный временной ряд с аддитивным шумом, получаемым из распределения Вальда.

2. Задача построения оптимальной суперпозиции. Поставим задачу построения оптимальной прогностической модели в виде суперпозиции двух функций прогноза ряда и прогноза ошибки. Моделью $f \circ g$ назовем параметрическую функцию, которая по временному ряду $\mathbf{x} = [x_1, \dots, x_t]^\top$ и фиксированным параметрам вычисляет значение \hat{x}_{t+1} ряда в момент времени $t + 1$.

Прогностической моделью называется суперпозиция, где прогноз

$$\hat{x}_{t+1} = f \circ g(\hat{\mathbf{w}}, x_t, x_{t-1}, \dots, x_1) \quad (1)$$

получен суперпозицией моделей f и g . Параметры $\hat{\mathbf{w}}$ оптимизированы согласно функции ошибки.

Прогностические модели выбираются из набора моделей-претендентов $f, g \in \mathfrak{F}$. Набор моделей описан в разделе 4. Невязка (регрессионный остаток) прогнозирования ε — разности $\varepsilon_e = x_t - \hat{x}_t$ между фактическими значениями ряда и прогнозом $x_{t+1} = f \circ g(x_t, x_{t-1}, \dots, x_1) + \varepsilon_{t+1}$. Предполагается, что математическое ожидание остатков не равно нулю, $\mathbf{E}(\varepsilon) \neq 0$, и их дисперсия постоянна, $\mathbf{D}(\varepsilon) = \sigma^2$.

Функция f прогнозирует тренд и сезонность, а функция g прогнозирует остаток. Конечный прогноз \hat{x}_{t+1} складывается с учетом прогноза обеих функций. Предполагается, что функция g повышает точность прогноза, полученного с помощью функции f :

$$\begin{aligned} f : x_t &\rightarrow \hat{x}_{t+1}^f, \\ g : x_{t+1}^f &\rightarrow \hat{x}_{t+1}^{f:g}. \end{aligned}$$

Требуется построить поточечный прогноз (1) и прогноз на r отсчетов вперед $[\hat{x}_{t+1}, \dots, \hat{x}_{t+r}]^\top$. Назовем *запросом на прогнозирование* число r , которое задает число точек, на которые необходимо сделать прогноз.

Назовем *горизонтом прогнозирования* число h , $h \leq r$, при котором расстояние между распределением двух случайных величин \hat{x} и x значимо после шага под номером

h

$$\|\hat{x}_{t+h-1} - x_{t+h-1}\| \leq \mu, \text{ при условии } \|\hat{x}_{t+h} - x_{t+h}\| > \mu. \quad (2)$$

Для получения оценок параметров $\hat{\mathbf{w}}$ модели требуется минимизировать значение функции ошибки S

$$\hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{\tau=1}^t S(\mathbf{w}, x_{\tau} - \hat{x}_{\tau}).$$

Ошибка прогнозирования S оценивается с помощью MAPE (5).

3. Суперпозиция при ретроспективном прогнозе. В данной работе предлагаются два способа построения суперпозиции и алгоритм поточечного прогнозирования. Метод параллельного прогноза и метод прогноза с накоплением основаны на уточнении базового добавлением прогноза ошибок. В них используются две функции для прогноза: одна — для базового прогноза, вторая — для ошибок. Их отличие заключается в том, что в методе параллельного прогноза учитывается только предыдущее значение и ошибки, и базового прогноза, а в методе прогноза с накоплением — несколько предыдущих. В алгоритме поточечного прогнозирования прогнозируется следующее значение ряда, которое затем используется как элемент истории.

3.1. Метод параллельного прогноза. Задан временной ряд \mathbf{x} и две функции f и g . Алгоритм (3) состоит в следующем.

1. С помощью функции f вычисляются n прогнозов конца истории $\hat{x}_t^f, \dots, \hat{x}_{t-n+1}^f$ по точкам.
2. Вычисляются n остатков $\hat{\varepsilon}_t, \dots, \hat{\varepsilon}_{t-n+1}$ в виде разницы

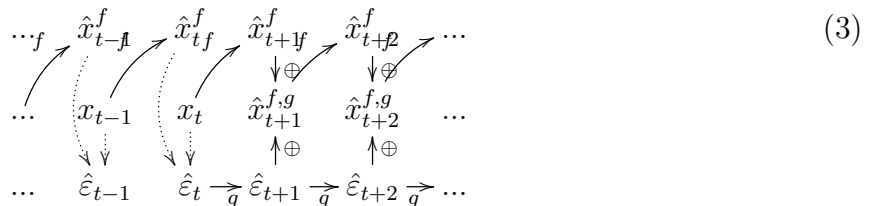
$$\hat{\varepsilon}_{t-k} = x_{t-k} - \hat{x}_{t-k}^f,$$

где n — длина исторических значений временного ряда.

3. С помощью функции g прогнозируются остатки $\hat{\varepsilon}_{t+i}$ на h отсчетов вперед.
4. Выполняется подсчет конечных прогнозов

$$\hat{x}_{t+i}^{f,g} = \hat{x}_{t+i}^f + \hat{\varepsilon}_{t+i}$$

с последовательным подсчетом прогноза базовой функцией f на одну точку \hat{x}_{t+i}^f .



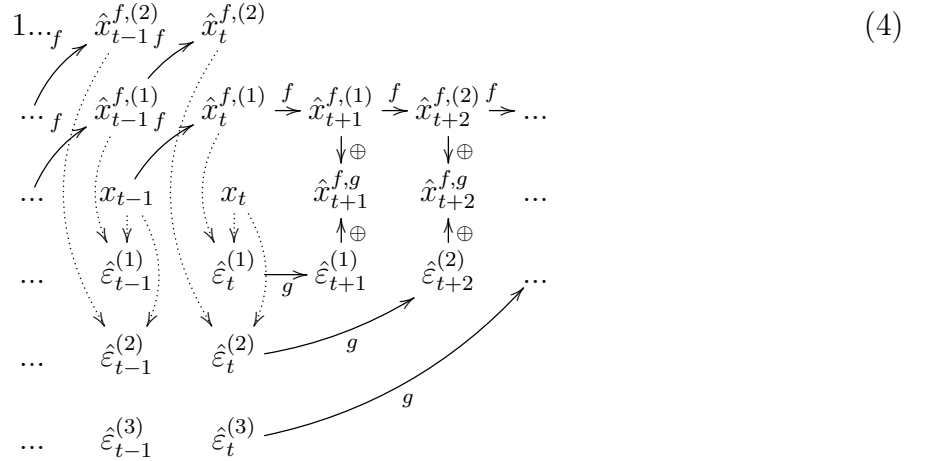
3.2. Метод прогноза с накоплением. Задан временной ряд \mathbf{x} и две функции f и g , с помощью которых делается прогноз на глубине прогнозирования i . Глубина прогнозирования i — количество предшествующих элементов заданного временного ряда, на основе которых делается прогноз. Алгоритм (4) состоит в следующем.

1. С помощью базовой функции f вычисляется ретроспективный прогноз $\hat{x}_{t+1}^{f,(1)}, \dots, \hat{x}_{t+i}^{f,(i)}$ с горизонтом прогнозирования i , каждый — на глубине i .
2. С помощью базовой функции f вычисляется r наборов прогнозов конца истории, $\hat{x}_t^{f,(i)}, \dots, \hat{x}_{t-n+1}^{f,(i)}$, каждый набор — на глубине i , $i = 1, \dots, r$.
3. Вычисляется r наборов остатков $\hat{\varepsilon}_t^{(i)}, \dots, \hat{\varepsilon}_{t-n+1}^{(i)}$

$$\hat{\varepsilon}_{t-k}^{(i)} = x_{t-k} - \hat{x}_{t-k}^{f,(i)}, \quad i = 1, \dots, r.$$

4. С помощью функции g прогнозируются остатки $\hat{\varepsilon}_{t+i}^{(i)}$, каждый прогноз выполняется на одну точку и использует вычисленную последовательность $\hat{\varepsilon}_t^{(i)}, \dots, \hat{\varepsilon}_{t-n(g)+1}^{(i)}$.
5. Выполняется подсчет конечных прогнозов

$$\hat{x}_{t+i}^{f,g} = \hat{x}_{t+i}^{f,(i)} + \hat{\varepsilon}_{t+i}^{(i)}.$$



3.2. Алгоритм поточечного прогнозирования по запросу. Зафиксируем время t_0 , считаем известными значения ряда $\mathbf{x}_{1:t_0} = [x_1, x_2, \dots, x_{t_0}]$. Вектор \mathbf{x}_{t_0+r} назовем прогнозом с запросом на горизонт прогнозирования r , $t_0+r < T$, где T — длина истории ряда.

1. С помощью базовой функции f вычисляется прогноз \hat{x}_{t+1} следующего элемента.
2. Используя полученное значение \hat{x}_{t+1} и известную историю с помощью базовой функции f вычисляется прогноз \hat{x}_{t+2} .

3. Шаги повторяются, пока не будет достигнут запрос на прогнозирование r .

3.3. Алгоритм блочного прогноза. Для учета качества прогнозирования на запросах, превышающих единицу, предлагается строить блочный прогноз.

Для оценки качества прогноза временной ряд разбивается на обучающую $\mathbf{x}_{1 \div t_0}$ и проверочную $\mathbf{x}_{t_0+1 \div t_1}$ части. Для вычисления средней ошибки строится совокупность из $t_1 - t_0$ прогнозов, в которой i -ый прогноз ($0 \leq i < t_1 - t_0$) строится по известному ряду $\mathbf{x}_{1 \div t_0+i}$ на 1 отсчет вперед.

Зафиксируем длину блока r — максимальный запрос на прогнозирование, который не превышает горизонта прогнозирования используемых алгоритмов. Для сравнения с проверочной частью временного ряда составим следующий временной ряд-прогноз:

1. По известной обучающей части ряда $\mathbf{x}_{1 \div t_0}$ построим прогноз на r отсчетов вперед алгоритмом поточечного прогнозирования.
2. Будем теперь считать первые r значений проверочной части ряда известными. По $\mathbf{x}_{1 \div t_0+r}$ построим прогноз на r отсчетов вперед.
3. Повторим шаг 2, каждый раз считая очередные r проверочных значений известными, пока не дойдем до конца проверочной части.

В полученном блочном прогнозе одинаково часто встречаются значения, полученные в результате прогнозирования на 1, 2, ..., r отсчетов вперед. Таким образом, его качество является усредненным качеством прогнозирования алгоритма на запросах от 1 до r .

4. Вычислительный эксперимент. Цель эксперимента — анализ качества и устойчивости прогноза с использованием суперпозиции прогностических моделей при условии асимметричного распределения регрессионных остатков. Используются следующие выборки: временные ряды грузоперевозок [11], потребление электроэнергии [15], цена на электричество [14], число пассажиров авиалиний [13], а также синтетическая (синус, зашумленный распределением Вальда).

При тестировании суперпозиций исследуются прогнозы, построенные базовыми моделями и по их суперпозициям. Приводятся графические иллюстрации прогнозов, а также численные значения ошибки. В качестве функций ошибки взяты функция MAPE (5) и функция MSE (6):

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\varepsilon_i}{x_i} \right|, \quad (5)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2. \quad (6)$$

Во всех экспериментах измеряется качество блочного прогноза (3.3).

Для каждой из выборок (railroads, airline, synthetic, german; результаты для последних трех размещены в [12]) построены:

1. Общий вид временного ряда и разбиение на обучающую и проверочную части.
2. Прогнозы всех основных моделей с оптимальными структурными параметрами. Размер блока во всех экспериментах равен 10. Все прогнозы строятся с помощью алгоритма (3.3).
3. Прогноз лучшей суперпозиции и основной модели этой суперпозиции.
4. Матрица качества прогноза суперпозиций. По строкам отложены основные модели, по столбцам – модели прогнозирования; первый столбец (None) соответствует основной модели, без суперпозиции. Ячейки, обведенные рамками, соответствуют наилучшим прогнозам для данной основной модели (то есть минимальному значению в строке). Глобальный минимум обозначен более широкой рамкой.

5. Результаты эксперимента: railroads.

На рис. 2 видно, что распределение ошибок близко к нормальному, поэтому использование суперпозиции не улучшает качество прогноза.

6. Горизонт прогнозирования в экспериментах. Для определения горизонта прогнозирования (2) построим зависимость ошибки MAPE (5), а также ее выборочного стандартного отклонения (представлена на графике в виде доверительной верхней границы), от запроса прогнозирования в проведенных экспериментах. Горизонт прогнозирования будем определять графически по правилу сломанной трости (рис. 4).

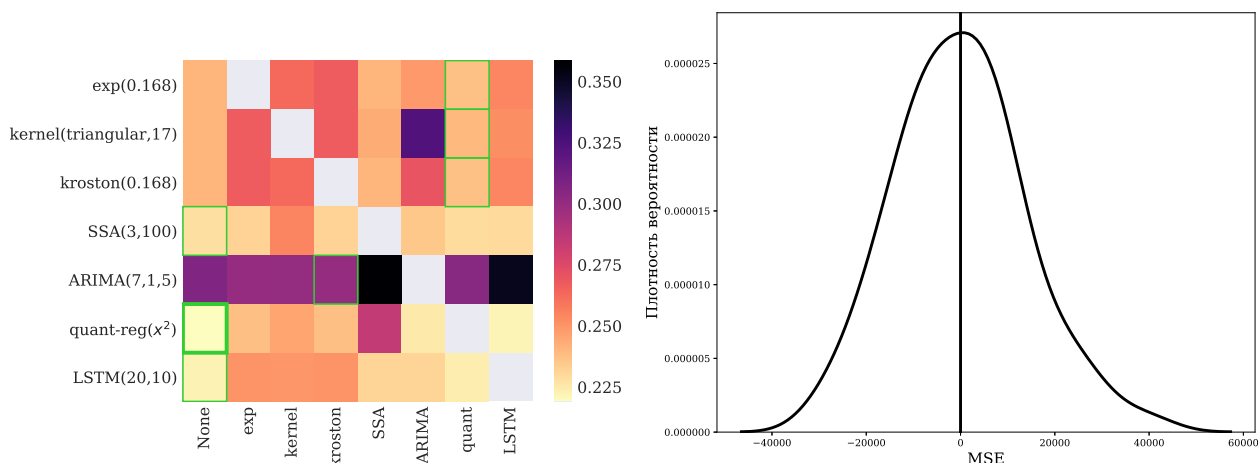


Рис. 1: Матрица качества прогноза суперпозиций (MAPE)

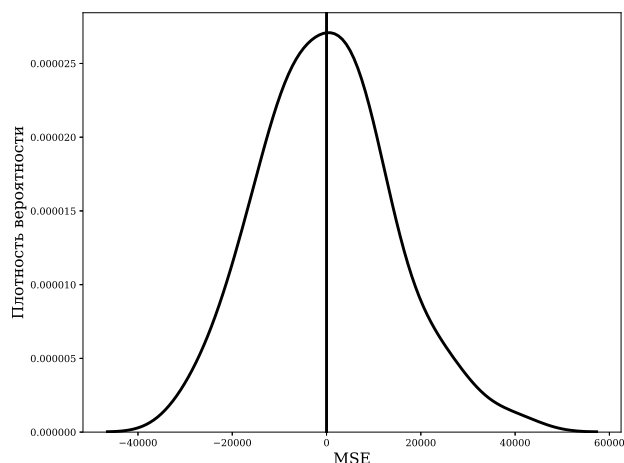


Рис. 2: Эмпирическая функция распределения ошибок лучшей основной модели (LSTM)

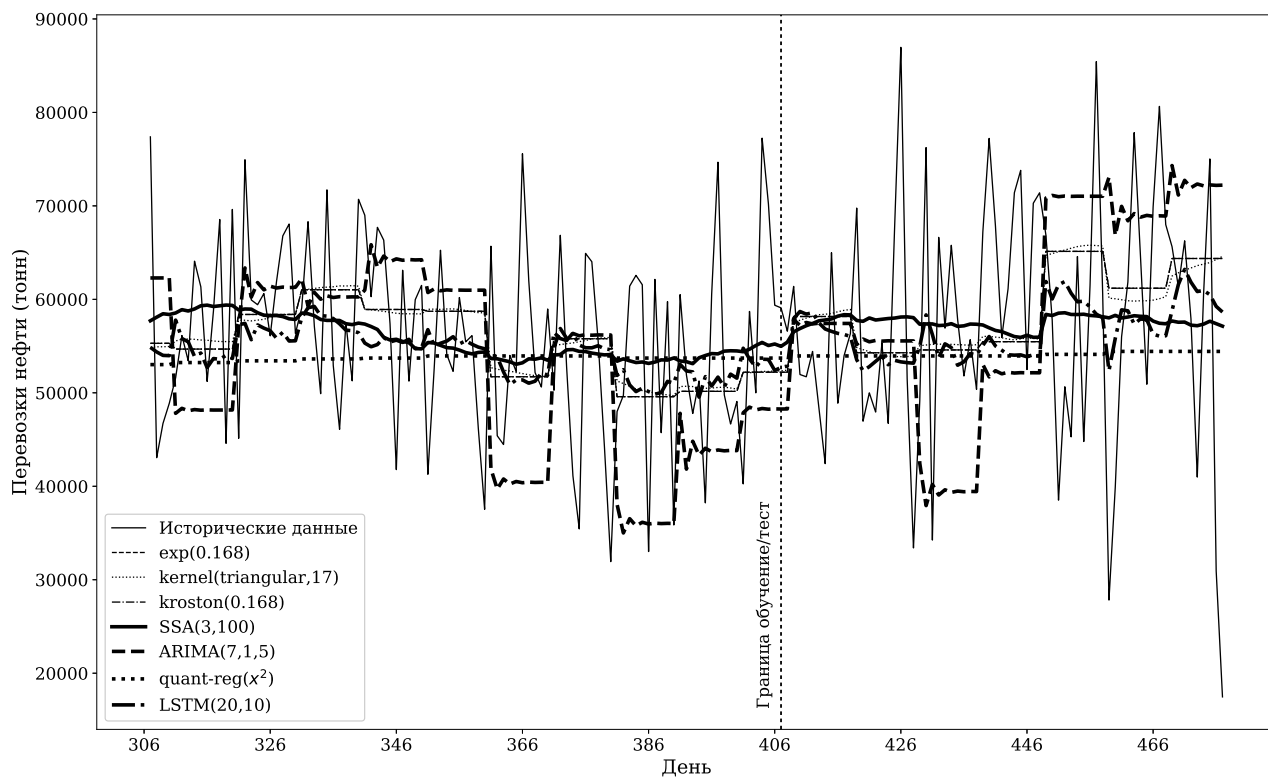


Рис. 3: Прогнозы основных моделей

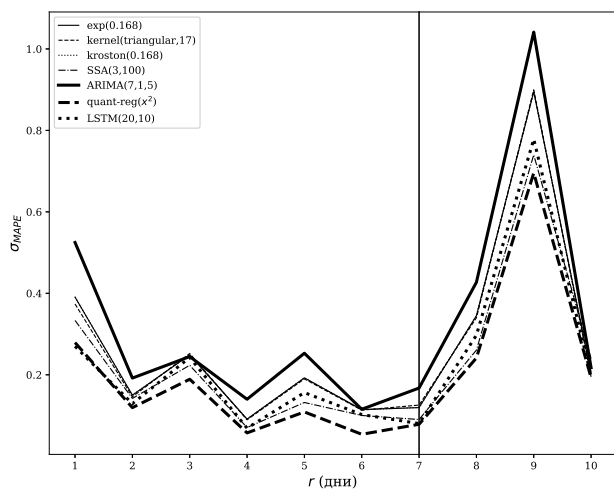


Рис. 4: Правило сломанной трости: резкий рост выборочного стандартного отклонения ошибки свидетельствует о достижении горизонта прогнозирования

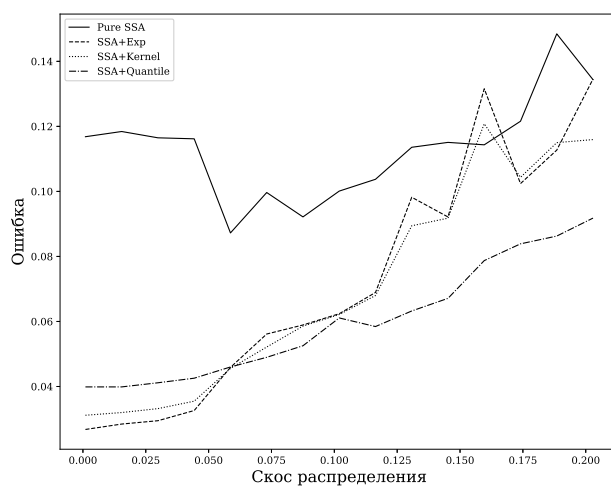


Рис. 5: Зависимость качества прогноза от “скоса” распределения

Таким образом, горизонт прогнозирования равен 7, а использованная длина блока 10, превышает его незначительно.

7. Эксперимент: зависимость качества от асимметричности распределения.

Для распределения шума временного ряда используется однопараметрическое подмножество распределений Вальда, задаваемое параметром l , по которому параметры распределения Вальда определяются как $\mu = 1 + \frac{l}{30}$, $\lambda = \frac{l^2}{4}$. Для наглядной интерпретации, в качестве меры скоса распределения введем

$$S = |\mathbb{P}(x > x_{\text{н.в.}}) - \mathbb{P}(x < x_{\text{н.в.}})|,$$

где $x_{\text{н.в.}}$ — наиболее вероятное значение случайной величины, соответствующее максимуму плотности вероятности и положению пика на рис. 5. Очевидно, $S \in [0, \frac{1}{2})$, причем значение 0 соответствует распределению, вероятностная масса которого расположена симметрично от наиболее вероятного значения, например, нормальному, а значения, близкие к $\frac{1}{2}$, соответствуют сильно скошенным распределениям. Использована кусочно-линейная асимметричная функция ошибки

$$l(x) = \begin{cases} x, & x > 0 \\ -\frac{1}{3}x, & x < 0 \end{cases}.$$

8. Заключение. Рассмотрена задача построения суперпозиции моделей при прогнозировании временных рядов. В качестве базовых алгоритмов рассмотрены экспоненциальное сглаживание, ядерное сглаживание, метод Кростена, SSA, ARIMA, квантильная регрессия и LSTM. Для рассматриваемых выборок: железнодорожных грузовых перевозок, пассажирских авиаперевозок, цен на электроэнергию и синтетической — построены матрица качества суперпозиции и распределение остатков модели, а также определен горизонт прогнозирования. Вычислительный эксперимент показал, что использование суперпозиции может повысить качество прогноза относительно качества базовых моделей. Отдельно исследована зависимость ошибки от степени асимметричности распределения шумовых остатков и показано, что с ростом асимметричности суперпозиции достигают меньшей ошибки по сравнению с базовыми моделями. Исходный код эксперимента размещен в [16].

СПИСОК ЛИТЕРАТУРЫ

- [1] Буряк Е.В., Кульпина В.П., Голяшев А.В., Лобанова А.А. Динамика грузоперевозок в России // Бюллетень социально-экономического кризиса в России. 2015. 1. № 4. С. 12-15.
- [2] Lijuan L., Liu H., Wu J. et al. A novel model for wind power forecasting based on Markov residual correction // Renewable Energy Congress. 2015. N 6. P. 1-5.

- [3] Cohen D. A., Lys T. Z. A note on analysts' earnings forecast errors distribution // Journal of Accounting and Economics. 2003. **36**. N 1-3. P. 147-164.
- [4] Kan M. L., Lee Y. B., Chen W. C. Apply grey prediction in the number of Tourist // Genetic and Evolutionary Computing. 2010. N 4. P. 481-484.
- [5] Журавлев Ю. И., Рудаков К. В., Корчагин А. Д., Кузнецов М. П., Мотренко А. П., Стенина М. М., Стрижов В. В. Методы прогнозирования временных рядов на примере железнодорожных грузоперевозок // Вестник Российской академии наук. 2016. **86**. № 2. С. 186–188.
- [6] Guha B., Bandyopadhyay G. Gold price forecasting using ARIMA model // Journal of Advanced Management Science Vol. 2016. **4**. N 2. P. 117-121.
- [7] Голяндина Н. Э. Метод “Тусеница”-SSA: прогноз временных рядов. М.: Издательский центр “Академия”, 2004.
- [8] Laptev N., Yosinski J., Li L. E. et al. Time-series extreme event forecasting with neural networks at Uber // International Conference on Machine Learning. 2017. N 34. P. 1-5.
- [9] Kalekar P. S. Time series forecasting using holt-winters exponential smoothing // Kanwal Rekhi School of Information Technology. 2004. P. 1-13.
- [10] Ralaivola L., D'Alché-Buc F. Time series filtering, smoothing and learning using the kernel Kalman filter // Neural Networks, proceedings of IEEE International Joint Conference. 2005. **3**. P. 1449-1454.
- [11] Временные ряды грузоперевозок. URL: <http://svn.code.sf.net/p/mlalgorithms/code/Group474/Uvarov2017SuperpositionForecasting/data> (дата обращения: 27.03.2018).
- [12] Уваров Н.Д., Кузнецов М.П., Малькова А.С., Рудаков К.В., Стрижов В.В. Приложение к работе “Выбор суперпозиции моделей при прогнозировании грузовых железнодорожных перевозок”. 2018. URL: <http://svn.code.sf.net/p/mlalgorithms/code/Group474/Uvarov2017SuperpositionForecasting/doc/addition.pdf> (дата обращения: 27.03.2018).
- [13] Box G. E., Jenkins G. M., Reinsel G. C. et al. Time series analysis, forecasting and control. New York: John Wiley & Sons, 1976.
- [14] Цена на электричество в Германии. URL: http://svn.code.sf.net/p/mlalgorithms/code/Group474/Uvarov2017SuperpositionForecasting/data/electricity_price_german/GermanSpotPrice.csv (дата обращения: 27.03.2018).
- [15] Потребление электроэнергии (Польша). URL: <http://gdudek.el.pcz.pl/files/PL.xls> (дата обращения: 27.03.2018).

- [16] Исходный код эксперимента к работе “Выбор суперпозиции моделей при прогнозировании грузовых железнодорожных перевозок” (модуль на Python и Jupyter-ноутбук). URL: <http://svn.code.sf.net/p/mlalgorithms/code/Group474/Uvarov2017SuperpositionForecasting/code/> (дата обращения: 27.03.2018).