

**«Выбор суперпозиции моделей при прогнозировании грузовых
железнодорожных перевозок»**

Выполнила:
студентка 5 курса 374 группы
Малькова Александра Сергеевна

Научный руководитель:
д.ф-м.н., н.с. ВЦ РАН
Стрижов Вадим Викторович

Москва, 2017

Содержание

1	Введение	3
2	Постановка задачи	4
2.1	Описание данных	4
2.2	Задача построения оптимальной суперпозиции	5
3	Суперпозиция при ретроспективном прогнозе	6
3.1	Метод параллельного прогноза	7
3.2	Метод прогноза с накоплением	7
3.3	Алгоритм поточечного прогнозирования на r отсчётов вперёд	8
4	Набор моделей-претендентов	9
4.1	Экспоненциальное сглаживание	9
4.2	Ядерное сглаживание	10
4.3	Метод Кростена	10
4.4	Модель "Гусеница" Singular Spectrum Analysis	11
4.5	Модель ARIMA(p, d, q)	11
4.6	Квантильная регрессия	12
5	Вычислительный эксперимент	13
5.1	Список выборок	13
5.2	Список моделей	13
5.3	Структурные параметры	13
5.4	Тестирование суперпозиций	13
6	Заключение	15
	Список литературы	16

Аннотация

Рассматривается задача выбора оптимальной аналитической системы моделей для краткосрочного прогнозирования объемов железнодорожных грузовых перевозок. Исторические данные представляют собой временные ряды, отражающие объемы перевозок различных типов грузов между различными станциями. Характерными свойствами данных временных рядов являются их неравномерность по станциям и грузам, высокая волатильность, зашумленность, а также нестационарность, присущая некоторым рядам. Для построения качественных прогнозов предлагается создание такой системы, которая бы являлась суперпозицией прогностических моделей и учитывала особенности и свойства исторических данных. В качестве возможных моделей для включения в суперпозицию были рассмотрены модели среднего значения, скользящего среднего, экспоненциального сглаживания, векторной авторегрессии, интегрированная модель авторегрессионного скользящего среднего, а также метод Кростона. Приведены сравнительные характеристики моделей.

Ключевые слова: *временные ряды, прогнозирование, суперпозиция, прогностические модели, асимметричное 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 [?]. 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 статье [?] обосновывается необходимость в усовершенствовании и оптимизации работы, связанной с использованием железных путей и говорится, что 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.

Цель работы. Предложить метод построения суперпозиции моделей, которая бы давала адекватные, качественные и как можно более точные прогнозы по объемам перевозимых грузов.

Предложен метод построения суперпозиции, опирающийся на корректировке исходного прогноза. Алгоритм базируется на уточнении прогноза, полученного базовой функцией прогнозирования с помощью добавки прогноза для ошибок, полученного второй функцией.

Семейство моделей. В качестве возможных прогностических моделей предлагается использовать среднее значение, скользящее среднее (МА), экспоненциальное сглаживание, метод Кростона, интегрированную модель авторегрессионного скользящего среднего (ARIMA). В статье [?] приведен сравнительный анализ моделей в применении к данной задаче. В статье [?] изложена идея построения суперпозиции. В статье [?] отмечается, что задача выбора модели является актуальной в регрессионном анализе.

Данные. 1. Синтетическая выборка (+ как порождались)

а) временной ряд: ежедневные данные, недельная периодика [10];

б) временной ряд: часовые данные, суточная и недельная периодика [11], [12].

В данной работе будут использоваться остатки с распределением Леви-Хинчина:

$$\Phi(q) = e^{iq\beta - \gamma[1 + \beta\theta \text{sign}(q) \text{tg}(\pi\alpha/2)]|q|^\alpha}, \Phi(q) = e^{iq\beta\gamma|q| - \beta\gamma\theta q \ln|q|},$$

где $\text{sign}(q) = q/|q|$ - знак q , параметры $0 < \alpha \leq 2$, $|\theta| \leq 1$, $\gamma \geq 0$. Первое распределение - четырехпараметрическое, второе - предел первого при $\alpha \rightarrow 1$.

Рассматриваются следующие гипотезы распределения остатков:

* Нормальное распределение: $\varepsilon \sim N(0, \sigma^2)$

* Распределение Леви-Хинчина: $\varepsilon \sim \Phi(\alpha, \beta, \gamma, \theta)$

2. Исторические данные, которые предлагаются для исследования, представляют собой временные ряды, отражающие объемы грузоперевозок по российским железным дорогам. 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 [?].

3. Физико-химические измерения.

Качество прогнозирования. Сравнение рассматриваемых моделей предлагается проводить, выполняя ретроспективные прогнозы. Под ретроспективным прогнозом понимают прогноз, сделанный на основе знаний только части временного ряда (прогноз прошлого). Остатками временного ряда назовем разницу исторических и спрогнозированных значений. Сводя анализ остатков, можно сделать вывод о качестве модели и ее адекватности. Остатки модели должны быть независимы и не автокоррелированными. Последнее проверяется с помощью теста Дарбина Уотсона [?]. Также качество прогностической модели можно оценить с помощью специальных функций ошибок таких, как среднее арифметическое модулей относительных остатков MAPE (mean absolute percentage error), среднее MAPE (mean asymmetric percentage error) и сила прогноза SS. Функции потерь вычисляются на основе знаний значений исторического временного ряда и построенного ретроспективного прогноза.

2 Постановка задачи

2.1 Описание данных

Выборка. Временной ряд – это последовательность значений, описывающих протекающий во времени процесс, измеренных в последовательные моменты времени, обычно через равные промежутки.

Выборка

Подряд - временной сегмент.

Значения прогнозов вычисляются посредством модели. Моделью называется параметрическая функция, которая по временному ряду (или набору временных рядов одинаковой длины) и фиксированным параметрам вычисляет единственную точку – прогноз значения временного ряда в требуемый момент времени.

Понятие суперпозиции моделей. Суперпозиция моделей - объединение моделей прогнозирования. Конечный прогноз складывается с учётом всех используемых моделей прогнозирования. Таким образом, прогноз базовой функции уточняется с помощью функций, объединённых с ней в суперпозицию.

Структурные параметры. Пусть \mathcal{P} - множество параметров модели. Тогда оптимальный набор параметров α определяется как $\alpha = \operatorname{argmin}()$

Длина окна

Длина гусеницы

Данные о посуточной загруженности железнодорожных путей в системе представлены 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.

дата погрузки	код станции отправления	код станции назначения	код груза	суммарный вес груза
2007-01-01	14605	831504	1	56
2007-01-01	135602	165504	18	63
2007-01-01	830304	814208	13	246

2.2 Задача построения оптимальной суперпозиции

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

Прогностической моделью называется суперпозиция где прогноз

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

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. Предлагается приблизить временной ряд суперпозицией

$$x_{t+1} = f \circ g(x_t, x_{t-1}, \dots, x_1) + \varepsilon_{t+1}.$$

Невязка (регрессионный остаток) прогнозирования ε — разности $\varepsilon = x_t - \hat{x}_t$ между фактическими значениями ряда и прогнозом. Предполагается, что математическое ожидание остатков не равно нулю ($E(\varepsilon) \neq 0$) и их дисперсия постоянна ($D(\varepsilon) = \sigma^2$). 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

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

и прогноз на r отсчётов вперед $[\hat{x}_{t+1}, \dots, \hat{x}_{t+r}]^T$. Назовём *запросом на прогнозирование* число r , которое задаёт 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 статистически значимо после шага под номером h

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

Для получения оценок параметров $\hat{\mathbf{w}}$ модели требуется минимизировать значение функции ошибки $S(\mathbf{w}|F, D)$

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

Ошибка прогнозирования S оценивается с помощью MAPE (9) и SS (11). Формулы приведены в разделе 4.

3 Суперпозиция при ретроспективном прогнозе

Предлагается алгоритм поточечного прогнозирования (1) и два способа построения суперпозиции. 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 2 — 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 (2) состоит в следующем:

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

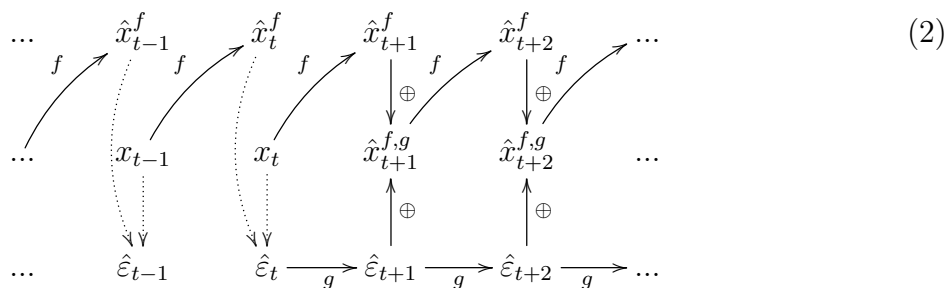
$$\hat{\epsilon}_{t-k} = x_{t-k} - \hat{x}_{t-k}^f,$$

где n — длина исторических значений временного ряда.

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

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

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



3.2 Метод прогноза с накоплением

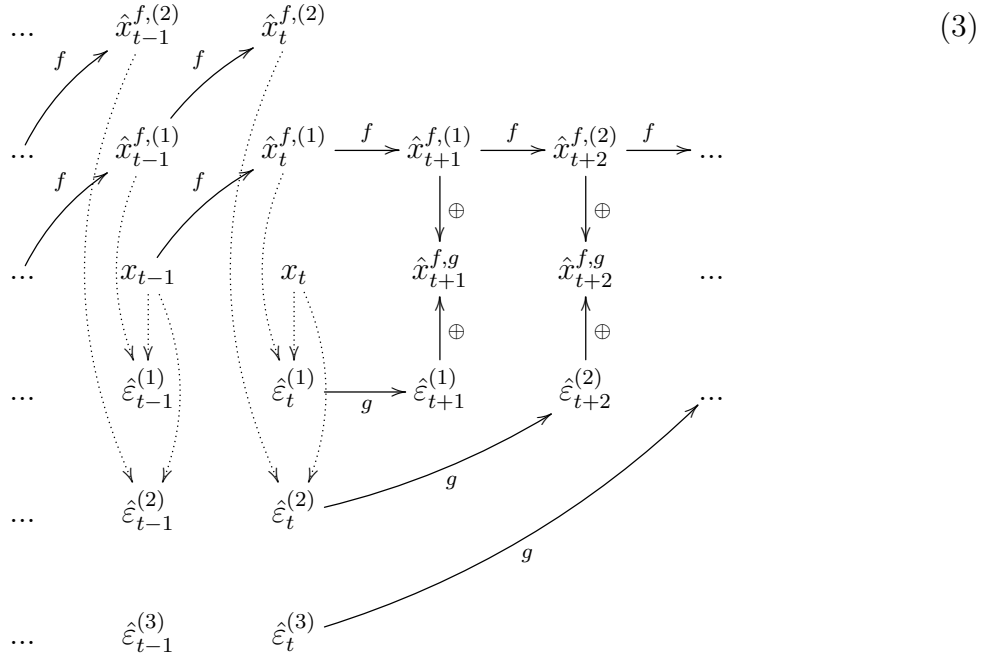
Задан временной ряд \mathbf{x} и две функции f и g , с помощью которых делается прогноз на глубине прогнозирования i . Глубина прогнозирования i — количество предшествующих элементов заданного временного ряда, на основе которых делается прогноз. Алгоритм (3) состоит в следующем.

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

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



3.3 Алгоритм поточечного прогнозирования на r отсчётов вперёд

Зафиксируем время t_0 , считаем известными значения ряда $\mathbf{x}_{1 \div t_0} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{t_0}]$. Вектор \mathbf{x}_{t_0+r} назовём 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

ряда.

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 до тех пор, пока не будет достигнут запрос на прогнозирование r .

На рисунке 1 ось t - ось единиц прогноза; t_1 - рассматриваемый момент времени, t_0 - момент времени, на который делается прогноз, t_i, \dots, t_2 - история, которая используется для прогноза следующей точки.

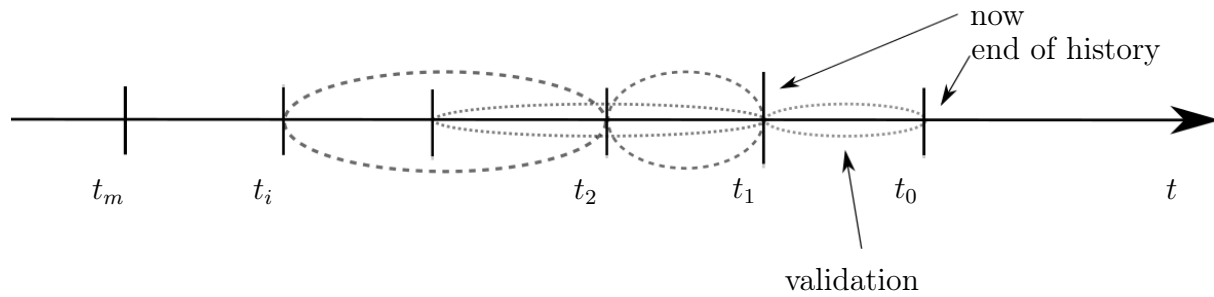


Рис. 1: Схема работы алгоритма прогнозирования на r отсчётов вперёд

4 Набор моделей-претендентов

4.1 Экспоненциальное сглаживание

Сложность прогнозирования обусловлена высокой волатильностью исторических данных по объёмам перевозок. Применение экспоненциального сглаживания [?] к такому ряду позволяет выровнять ряд и определить его тенденцию.

Экспоненциальное сглаживание представляется фильтром, на вход которого последовательно поступают члены исходного ряда, а на выходе формируются текущие значения экспоненциальной средней

$$z_t = \alpha x_t + (1 - \alpha)z_{t-1},$$

Чем меньше α — параметр сглаживания ряда, $\alpha \in (0, 1)$, тем в большей степени фильтруются, подавляются изменения исходного ряда и шума.

Таким образом,

$$\hat{x}_{t+1} = \alpha x_t + (1 - \alpha)\hat{x}_{t+1} \quad (4)$$

4.2 Ядерное сглаживание

Задан временной ряд x . Существует неизвестная целевая зависимость $y^* : t \rightarrow \mathbf{x}$, значения которой известны только на объектах обучающей выборки $X^l = (t_i, x_i)_{i=1}^l$. Требуется построить алгоритм $a : t \rightarrow \mathbf{x}$, аппроксимирующий целевую зависимость y^* .

С помощью функции ядра зададим веса объектов:

$$\omega_i(t) = K\left(\frac{\rho(t, t_i)}{h}\right),$$

где h - ширина окна сглаживания, $\rho(t, t_i) = |t - t_i|$ - функция расстояния (в данном случае она будет принимать целые значения - расстояние между номерами отсчётов), K - ядро (гладкая, невозрастающая, ограниченная функция $[0, +\infty) \rightarrow [0, +\infty)$).

По формуле Надарая-Ватсона

$$a_h(t, X^l) = \frac{\sum_{i=1}^l x_i K\left(\frac{\rho(t, t_i)}{h}\right)}{K\left(\frac{\rho(t, t_i)}{h}\right)}.$$

В данной работе используется ядро Епанечникова: $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$.

$$\hat{x}_{t+1} = a_h(t + 1, X^l) \quad (5)$$

4.3 Метод Кростена

Для прогнозирования прерывистых рядов используют метод Кростена [?], в соответствии с которым исходный ряд разбивается на два ряда, один из которых описывает ненулевой объем перевозимых товаров, а второй — интервалы между ненулевыми перевозками (промежутки времени, в течение которых перевозки отсутствовали). Затем каждый ряд прогнозируется в отдельности экспоненциальным сглаживанием.

$\mathbf{x} = \{x_t\}_{t=1}^T$ — временной ряд ненулевого спроса, $\mathbf{q} = \{q_t\}_{t=1}^T$ — временной ряд для интервалов времени между ненулевым спросом.

Экспоненциальное сглаживания обоих рядов:

$$z_t = \alpha x_t + (1 - \alpha)z_{t-1},$$

$$p_t = \alpha q_t + (1 - \alpha)p_{t-1},$$

$$\hat{x}_t = \frac{z_t}{p_t}.$$

Здесь α — параметр сглаживания (как в обычном экспоненциальном сглаживании), $\alpha \in (0, 1)$. Когда значение \hat{x} оказывается превышенным, вводят модификацию, в соответствии с которой прогноз находится как

$$\hat{x}_t = \left(1 - \frac{\alpha}{2}\right) \frac{z_t}{p_t}. \quad (6)$$

В основе метода Кростона лежит предположение о том, что все перевозки являются случайной величиной с нормальным распределением, а факт того, что перевозка произошла является случайной величиной с распределением Бернулли. Это достаточно сильное предположение и выполняется не для всех рядов, поэтому зачастую рассматривают модифицированный алгоритм, именуемый подходом Виллемейна, опирающийся на процедуру бутстрепа (bootstrap = resampling).

Существенным недостатком данных моделей является то, что они не учитывают влияние внешних факторов. Так, например, экономические факторы, такие как цены товаров, курс рубля, а также сезонность влияют на объем спроса товаров. Наиболее применяемыми на практике моделями, позволяющими учесть внешние факторы, являются регрессионные модели.

4.4 Модель "Гусеница" Singular Spectrum Analysis

Поставим в соответствие временному ряду хтраекторную матрицу \mathbf{X} , т.е. матрицу следующего вида:

$$\begin{bmatrix} x_1 & x_2 & \dots & x_l \\ x_2 & x_3 & \dots & x_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m-l+1} & x_{m-l+2} & \dots & x_m \end{bmatrix}$$

Построим сингулярное разложение матрицы $\mathbf{X}^T \mathbf{X}$:

$$\mathbf{X}^T \mathbf{X} = \mathbf{V} \mathbf{H} \mathbf{V}^T, \mathbf{H} = \text{diag}(h_1, \dots, h_l),$$

где h_1, \dots, h_l — собственные числа матрицы $\mathbf{X}^T \mathbf{X}$, отвечающие за величины различных частот спектра временного ряда \mathbf{x} .

4.5 Модель ARIMA(p, d, q)

Интегрированная модель авторегрессии и скользящего среднего является расширением моделей ARMA для нестационарных временных рядов, которые можно преобразовать в стационарные взятием разностей некоторого порядка от исходного временного ряда. Модель ARIMA(p, d, q) означает, что разности временного ряда порядка d подчиняются модели ARMA(p, q).

Для нестационарного временного ряда x_t имеет вид:

$$\nabla^d x_t = c + \sum_{i=1}^p a_i \nabla^d x_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t, \quad (7)$$

где ε_t — стационарный временной ряд, c, a_i, b_j — параметры модели, ∇^d — оператор разности временного ряда порядка d .

Модель (??) интерпретируется как ARMA($p + d, q$) — модель с d единичными корнями.

Модель ARIMA позволяет с приемлемым качеством прогнозировать как стационарные временные ряды, так и нестационарные, временные ряды с трендом, а также при небольшой модификации и ряды с сезонной компонентой (SARIMA) и внешними факторами (ARIMAX). В случае большого количества нулевых значений регрессионные модели не подходят. Поэтому предлагается рассмотреть суперпозицию нескольких алгоритмов, чтобы избежать ошибочных прогнозов в случае, когда значительную часть временного ряда составляют нулевые значения.

4.6 Квантильная регрессия

Задан набор временных рядов, в котором временной ряд $\mathbf{x} = \{x_i\}_{i=1}^T$, и горизонт отсрочки прогноза h (число отсчетов от конца временного ряда до точки прогноза, включительно). Требуется спрогнозировать следующую точку x_{T+1} временного ряда \mathbf{x} так, чтобы выполнялось условие оптимальности функции потерь и свертки гистограммы (1), построенной по значениям временного ряда.

Гистограмма \mathcal{H} — набор пар

$$\mathcal{H} = \{(y_k, g_k)\}_{k=1}^K, \quad (8)$$

где K — число интервалов $[y_k^{min}, y_k^{max}]$ со средним значением y_k , на которые разбита ось значений ряда \mathbf{x} , g_k — высота столбца гистограммы на интервале \bar{y}_k , равная взвешенной сумме количества точек ряда, попавших в этот интервал.

Предполагается, что ряд $\mathbf{x} = \{x_i\}_{i=1}^T$ стационарен, то есть совместное распределение вероятностей T наблюдений $x_1, x_2, x_3, \dots, x_T$ совпадает с распределением $x_{1+\tau}, x_{2+\tau}, x_{3+\tau}, \dots, x_{T+\tau}$ при любых T и τ .

Ни одна из рассмотренных допустимых моделей не удовлетворяет всем свойствам временных рядов по грузовым перевозкам, поэтому предлагается рассмотреть суперпозицию алгоритмов для достижения наилучшего результата. Гипотеза: одна модель не может учесть все особенности, скрытые в данных, поэтому можно рассмотреть суперпозицию нескольких моделей.

5 Вычислительный эксперимент

Цель эксперимента. Анализ качества и устойчивости прогноза с использованием суперпозиции прогностических моделей при условии асимметричного распределения регрессионных остатков.

5.1 Список выборок

В данной работе используются следующие выборки:

- Временные ряды грузоперевозок [10]
- Потребление электроэнергии (Польша) [11]
- Цена на электричество [12]

От Никиты Уварова Ввести предположение о сезонности (не более двух циклов сезонности)

5.2 Список моделей

Для построения прогноза используются следующие модели прогнозирования, включенные в семейство \mathfrak{F} :

- Экспоненциальное сглаживание: по формуле (4)
- Ядерное сглаживание: по формуле (5)
- Метод Кростена: по формуле (6)
- Модель „Гусеница“(SSA): по формуле (??)
- Интегрированная модель авторегрессионного скользящего среднего (ARIMA): по формуле (7)
- Квантильная регрессия: по формуле (??)
- Модель LSTM.

5.3 Структурные параметры

5.4 Тестирование суперпозиций

Рассматриваются временные ряды, указанные в пункте 5.1, и исследуются прогнозы, построенные базовыми моделями и их суперпозициями. Приводятся графические иллюстрации прогнозов, а также численные значения ошибки. В качестве метрик качества взяты следующие функции: функция MAPE (9), функция MSE (10), функция

SS (11). MSE_{forecast} - средний квадрат ошибки, посчитанный на прогнозных данных, MSE_{history} - средний квадрат ошибки, посчитанный на исходных данных.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{\varepsilon_i}{x_i} \right|, \quad (9)$$

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2, \quad (10)$$

$$\text{SS} = 1 - \frac{MSE_{\text{forecast}}}{MSE_{\text{history}}}. \quad (11)$$

Добавить кусочно-линейную функцию ошибки

6 Заключение

Рассмотрена задача построения суперпозиции моделей при прогнозировании железнодорожных грузовых перевозок. В семейство моделей, из которого строилась суперпозиция, были включены следующие модели: ядерное сглаживание, экспоненциальное сглаживание, интегрированная модель авторегрессионного скользящего среднего (ARIMA) и векторная авторегрессия (VAR). Со ними были построены следующие суперпозиции: MA-MA, ARIMA-ARIMA, MA-ARIMA, ARIMA-MA. Вычислительный эксперимент показал, что суперпозиция способна прогнозировать ряд с наименьшей по сравнению с базовыми моделями ошибкой. Ввиду того, что не существует единой прогностической системы, которая бы давала адекватные и качественные прогнозы для всех временных рядов, приходится подбирать прогностическую модель для каждого ряда в отдельности. Таким образом, был предложен еще один метод прогнозирования - суперпозиция, построенная двумя методами, описанными в работе, которая наряду с базовыми моделями может быть прогностической моделью.

Список литературы

- [1] Е. Буряк, В. Кульпина, А. Голяшев, А. Лобанова. Динамика грузоперевозок в России. *Бюллетень социально-экономического кризиса в России*, 2015.
- [2] О.В. Коришева. Управление экономической устойчивостью транспортных компаний в сфере грузов железнодорожных перевозок. 2014.
- [3] В. Хардле. *Прикладная непараметрическая регрессия*. 1993.
- [4] В.В. Стрижов. *Порождение и выбор моделей в задачах регрессии и классификации*. 2014.
- [5] А.П. Мотренко, М.М. Стенина, К.В. Рудаков, М.П. Кузнецов. *Выбор оптимальной модели прогнозирования объёмов грузовых железнодорожных перевозок*. 2015.
- [6] А.Д. Корчагин, М.П. Кузнецов, А.П. Мотренко, М.М. Стенина, В.В. Стрижов, Ю.И. Журавлёв, К.В. Рудаков. *Создание системы прогнозирования объёмов спроса на грузовые железнодорожные перевозки*.
- [7] М. Вербик. *Путеводитель по современной эконометрике*. М.: Научная книга, 2008.
- [8] В.В. Стрижов, А.П. Мотренко, К.В. Рудаков. *Учёт влияния экзогенных факторов при непараметрическом прогнозировании временных рядов*. 2015.
- [9] R.J. Hyndman, L. Shenstone. *Stochastic models underlying croston's method for intermittent demand forecasting*. *Journal of Forecasting*, 2005
- [10] Временные ряды грузоперевозок
- [11] Потребление электроэнергии (Польша)
- [12] Цена на электричество