Московский физико-технический институт
(Государственный университет)

Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

ДИПЛОМНАЯ РАБОТА СТУДЕНТКИ 274 ГРУППЫ

«Выбор суперпозиции моделей при прогнозировании грузовых железнодорожных перевозок»

Выполнила:
студентка 4 курса 274 группы
Двинских Дарина Михайловна

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

Москва, 2016

Содержание

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, волатильность, зашумленность.*

1 Введение

Актуальность темы. 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 загруженности железнодорожных путей и их рационального использования. В статье [2] 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. Построить аналитическую систему, представляющую собой суперпозицию моделей, которая бы давала адекватные, качественные и как можно более точные прогнозы по объемам перевозимых грузов.

Новизна. 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.

Сеймество моделей. В качестве возможных прогностических моделей предлагается использовать среднее значение, скользящее среднее (MA), экспоненциальное сглаживание, метод Кростона, интегрированную модель авторегрессионного скользящего среднего (ARIMA) и векторную авторегрессию (VAR). Модели выбирались на основе логических соображений, учитывавшие специфику временных рядов, их свойства, а также в соответствии со статьей [6],в которой приведен сравнительный анализ моделей в применении к данной задаче. В статье [7] изложена идея построения суперпозиции. Моделью будем называть параметрическую функцию, которая по временному ряду (или набору временных рядов, заданных на одном и том же временном промежутке) и фиксированным параметрам вычисляет значения временного ряда в следующий момент времени. Допустимыми моделями будем считать те, которые учитывают специфику исторических данных, то есть ограничения на модели должны пересекаться со свойствами временных рядов, задающих объемы грузоперевозок. Под адекватностью модели обычно подразумевают соответствие модели моделируемому объекту или процессу. Ввиду невозможности полностью воссоздать процесс моделью, адекватность проверяется не для всех свойств, а только для тех, которые считаются наиболее существенными для исследования. Многообразие моделей для решения большинства задач представлено в классических учебниках [3,4]. В статье в статье [5] отмечается, что задача выбора модели является одной из самых актуальных в регрессионном анализе.

Данные. 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 волатильность, неравномерность, а также зашумленность. Для некоторых временных рядов свойствена 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 выполняются с помощью теста Дики-Фуллера [8].

Запрос на прогнозирование. Запрос на прогноз содержит время, код груза, а также станцию отправления и станцию назначения. Система в свою очередь возвращает количество перевозимого груза заданного типа между фиксированными станциями. Данные могут быть агрегированы в соответствии с запросом на прогнозирование. Допускаются следующие вариации: по времени (по дням, неделям, месяцам), по месту (пара станций, пара районов), по грузам (один тип груза, суммарно все грузы).

Агрегация временных рядов порождает иерархическую структуру данных, в которой временные ряды каждого следующего (более высокого) уровня формируются путем поэлементного суммирования. На первом уровне иерархии располагаются перевозки по каждому грузу и по каждому району в отдельности. Второй уровень иерархии отражает суммарные объемы перевозок по грузам или по районам. И, наконец, последний, третий, уровень иерархии отвечает за общий суммарный прогноз по всем грузам и всем районам. В соответствии с иерархической структурой, прогнозы должны удовлетворять требованию согласованности. Данное ограничение требует чтобы, суммарный прогноз по различным разрезам иерархии должен совпадать с общим прогнозом. Так, сумма прогнозов объемов спроса на отдельные типы грузов должна совпадать с прогнозом суммарного объема спроса по всем грузам. Поэтому прогнозы, удовлетворяющие условию согласованности по уровням иерархии, будем называть согласованными прогнозами.

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. Проводя анализ остатков, можно сделать вывод о качестве модели и ее адекватности. Остатки модели должны быть независимы и не автокореллированными. Последнее проверяется с помощью теста Дарбина Уотсона [8]. Также качество прогностической модели можно оценить с помощью специальных функций ошибок таких, как среднее арифметическое модулей остатков MAE (mean absolute error), среднее арифметическое модулей относительных остатков MAPE (mean absolute percentage error), среднее отклонение модулей остатков PMAD, среднеквадратичная ошибка MSE, корень среднеквадратичной ошибки RMSE и сила прогноза SS. Функции потерь вычисляются на основе знаний значений исторического временного ряда и построенного ретроспективного прогноза.

2 Выбор модели

Для разработки суперпозиции моделей прогнозирования объемов спроса на грузовые железнодорожные перевозки необходимо провести исследование алгоритмов прогнозирования. Предлагается, исходя из анализа временных рядов для каждого из 43-х товаров, определить наиболее подходящие алгоритмы прогнозирования, которые будут включены в систему. В зависимости от запроса на прогноз, учитывающий горизонт прогнозирования и степень агрегации данных, может изменяться выбор алгоритмов, включаемых в систему.

Сложность прогнозирования обусловлена высокой волатильностью исторических данных по объёмам перевозок. Применение экспоненциального сглаживания [4] к такому ряду позволяет выровнять ряд и определить его тенденцию.

2.1 Экспоненциальное сглаживание

Один из простейших и распространенных приемов выравнивания ряда. Экспоненциальное сглаживание можно представить как фильтр, на вход которого последовательно поступают члены исходного ряда, а на выходе формируются текущие значения экспоненциальной средней.

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

где $\mathbf{x}$ — исходный ряд, $\mathbf{z}$ — сглаженный ряд. Здесь α — параметр сглаживания ряда, $\alpha \in (0,1)$. Чем меньше α, тем в большей степени фильтруются, подавляются колебания исходного ряда и шума.

Также выровнять ряд можно используя модель среднего значения и модель скользящего среднего,последняя усредняет значения ряда в окне шириной n

2.2 Прогноз средним значением

Модель усредняет последние k отсчетов временного ряда. Также можно усреднять значения по всему ряду.

$$\hat{x}_{T+1} = \sum_{t=T-k+1}^{T} x_t.$$

2.3 Скользящее среднее

Модель используется для усреднения значений исходного временного ряда. Скользящие средние обычно используются с данными временных рядов для сглаживания краткосрочных колебаний и выделения основных тенденций или циклов.

$$z_t = \frac{x_t + x_{t+1} + \ldots + x_{t+n-1}}{n} = \sum_{i=0}^{n-1} x_{t+i},$$

где n — ширина окна. Тогда прогноз на горизонт прогнозирования h

$$\hat{x}_{T+h} = z_T.$$

Кроме того, прогнозирование перевозок осложняется присутствием значительного количества нулевых значений, которые означают, что в течение определенного момента времени груз не перевозился. Таким образом, временные ряды, отвечающие за перевозки товаров, представляют собой не непрерывную зависимость объема товаров от времени, а дискретную зависимость. Для прогнозирвоания прерывистых рядов используют метод Кростона [10], 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.

2.4 Метод Кростена

$\mathbf{d} = \{d_t\}_{t=1}^{T}$ — временной ряд ненулевого спроса,

$\mathbf{q} = \{q_t\}_{t=1}^{T}$ — временной ряд для интрвалов времени между ненулевым спросом.

Экспоненциальное сглаживания обоих рядов:

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

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

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

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

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

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. Одной из наиболее гибкой и простой в использовании моделью является векторная авторегрессия VAR(Vector Autoregression) [11], применяемая для прогнозирования многомерных временных рядов.

2.5 Модель VAR

Модель динамики нескольких временных рядов, в которой текущие значения этих рядов зависят от прошлых значений этих же временных рядов.

$$\mathbf{Y} = \mathbf{X}\mathbf{W},$$

где матрица $\mathbf{X}$ — матрица объектов-признак, $\mathbf{W}$ — матрица весов, $\mathbf{Y}$ — матрица ответов.

Откуда веса находятся транспонированием матриц

$$\mathbf{W} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}.$$

Тогда вектор прогнозов находится как

$$\hat{\mathbf{x}} = \mathbf{x}\mathbf{W}.$$

Но модель VAR применима только для стационарных временных рядов, в случае, когда исследуемые временные ряды нестационарны, необходимо использовать дополнительные инструменты, учитывающие коинтеграцию временных рядов. Также на практике широко применяются такие модели, как ARIMA (Autoregression Integrated Moving Average) и скользящее среднее MA [11,12].

2.6 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:

$$\nabla^d x_t = c + \sum_{i=1}^{p} a_i \nabla^d x_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} + \epsilon_t,$$

где ϵ_t - стационарный временной ряд,

c, a_i, b_j - параметры модели,

∇^d - оператор разности временного ряда порядка d.

Также данная модель интерпретируется как ARMA(p+d, q)— модель с d единичными корнями.

Модель ARIMA позволяет с хорошим 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) . Однако 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.

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.

Таблица 1: Сопоставление различных прогностических моделей свойствам временных рядов

Алгоритмы	Свойства временных рядов				
	сезонность	тренд	нестационарность	нулевые значения	внешние факторы
Среднее	-	-	-	+	-
Скользящее среднее	-	-	-	+	-
Сглаживание.	-	-	-	+	-
Метод Кростона	-	-	-	+	-
ARIMA	+	+	+	-	-
VAR	+	+	-	-	+

3 Суперпозиция при ретроспективном прогнозе

Рассмотрим суперпозицию двух видов. Оба основаны на уточнении базового добавлением прогноза ошибок.

3.1 Метод 1

Пусть дан временной ряд $\mathbf{x} = \{x_i\}_{i=1}^T$ и две базовые функции прогнозирования f и g (f прогнозирует ряд, а g – его остатки). Алгоритм состоит в следующем:

1. С помощью базовой функции f вычисляется n прогнозов конца истории $\hat{x}_t^f, ..., \hat{x}_{t-n+1}^f$ на одну точку.

2. Вычисляется n остатков $\hat{\varepsilon}_t, ..., \hat{\varepsilon}_{t-n+1}$ в виде разницы

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

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

4. Выполняется итеративный подсчет конечных прогнозов

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

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

$$\begin{tikzcd}
\cdots & \hat{x}^{f}_{t-1} \arrow[dd, dotted, bend right] & \hat{x}^{f}_{t} \arrow[dd, dotted, bend right] & \hat{x}^{f}_{t+1} \arrow[d, "\oplus"] & \hat{x}^{f}_{t+2} \arrow[d, "\oplus"] & \cdots \\
\cdots \arrow[ur, bend left, "f"] & x_{t-1} \arrow[ur, bend left, "f"] \arrow[d, dotted] & x_{t} \arrow[ur, bend left, "f"] \arrow[d, dotted] & \hat{x}^{f,g}_{t+1} \arrow[ur, bend left, "f"] & \hat{x}^{f,g}_{t+2} \arrow[ur, bend left, "f"] & \cdots \\
\cdots & \hat{\varepsilon}_{t-1} & \hat{\varepsilon}_{t} \arrow[r, "g"'] & \hat{\varepsilon}_{t+1} \arrow[u, "\oplus"'] \arrow[r, "g"'] & \hat{\varepsilon}_{t+2} \arrow[u, "\oplus"'] \arrow[r, "g"'] & \cdots
\end{tikzcd}$$

3.2 Метод 2

1. С помощью базовой функции f вычисляется ретроспективный прогноз $\hat{x}^{f,(1)}_{t+1}, ..., \hat{x}^{f,(i)}_{t+i}$ с горизонтом прогнозирования i, каждый — на глубине i.

2. С помощью базовой функции f вычисляется h наборов прогнозов конца истории, $\hat{x}^{f,(i)}_{t}, ..., \hat{x}^{f,(i)}_{t-n+1}$, каждый набор — на глубине i, $i = 1, ..., h$.

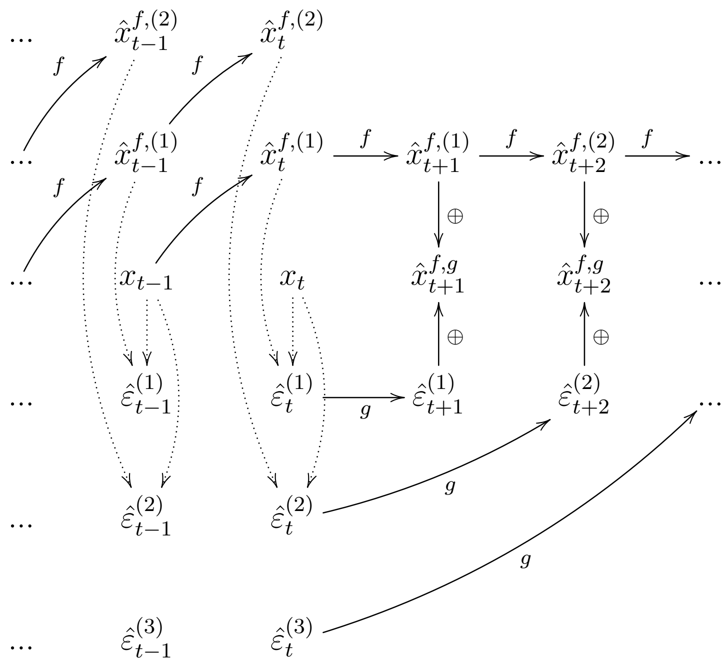
3. Вычисляется h наборов остатков $\hat{\varepsilon}^{(i)}_{t}, ..., \hat{\varepsilon}^{(i)}_{t-n+1}$

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

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

5. Выполняется подсчет конечных прогнозов

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



4 Постановка задачи

4.1 Описание данных

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

4.2 Задача построения оптимальной суперпозиции

Рассмотрим временной ряд $\mathbf{x} = \{x_i\}_{i=1}^T$ из системы временных рядов.

$f, g \in \mathcal{F}$ - две базовые функции прогнозирования из допустимого семейства моделей, отображающие допустимые параметры и значения времени в значение объемов перевозок

$$f, g : (\mathbf{w}, \mathbf{t}) \mapsto \hat{\mathbf{x}},$$

В семейство $\mathcal{F}$ включены модели: среднее значение, скользящее вреднее (MA), экспоненциальное сглаживание, метод Кростона, интегрированная модель авторегрессионного скользящего среднего (ARIMA) и векторная авторегрессия (VAR)

$$f : x_t \to \hat{x}_{t+1}^f,$$

$$g : x_{t+1}^f \to \hat{x}_{t+1}^{f,g}.$$

В таких обозначениях функция f является базовой функцией прогнозирования, а функция g - второстипенная функция, прогнозирующая ошибки. Конечный прогноз складывается с учетом прогноза обеих функций. Роль функции g заключается в уточнении прогноза, полученного с омощью базовой функции.

Предполагается, что значения временного ряда можно приблизить с допустимой точностью

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

Требуется построить прогноз

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

и ретроспективный прогноз с горизонтом прогнозирования h

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

На остатки модели накладываются следующие ограничения:

1. Равенство нулю матожидания: $\mathbf{E}(\epsilon) = 0$

2. Постоянство дисперсии: $\mathbf{D}(\epsilon) = \sigma^2$

3. Нормальность остатков: $\epsilon \backsim N(0, \sigma^2)$

Функция ошибки $S(\mathbf{w}|F, D)$ — функция, значение которой требуется минимизировать для получения оценок параметров $\hat{\mathbf{w}}$ модели

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \sum_{t\in\mathcal{T}} S(\mathbf{w}, x_t),$$

$$\epsilon = x_t - \hat{x}_t,$$

где $\hat{x}_t = f \circ g(\hat{\mathbf{w}}, x_t, x_{t-1}, \ldots, x_1)$ - прогноз, построенный суперпозицией моделей f и g.

Сравнения моделей и определение их адекватности проводятся по ретроспективным прогнозам.

Качество прогнозирования для ретроспективного прогноза оценивается с помощью MAE, MAPE, PMAD, MSE, RMSE и SS.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |\epsilon_i|$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{\epsilon_i}{x_i}\right|$$

$$\text{PMAD} = \sum_{i=1}^{n} |\epsilon_i| \, (\sum_{i=1}^{n} |x_i|)^(- 1)$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2$$

$$\text{RMSE} = \frac{1}{\sqrt{n}}\sqrt{\sum_{i=1}^{n} \epsilon_i^2}$$

$$SS = 1 - \frac{MSE_{forecast}}{MSE_{history}}$$

5 Вычислительный эксперимент

5.1 Предобработка данных данных

На первом этапе вычислительного эксперимента была проведена агрегация данных по станциям и регионам. Ниже представлен пример временных рядов для товаров, отправленных с 83-го региона, среди которых числятся каменный уголь, кокс, нефть, метизы, лом черных металлов, строительные грузы, зерно и бумага.

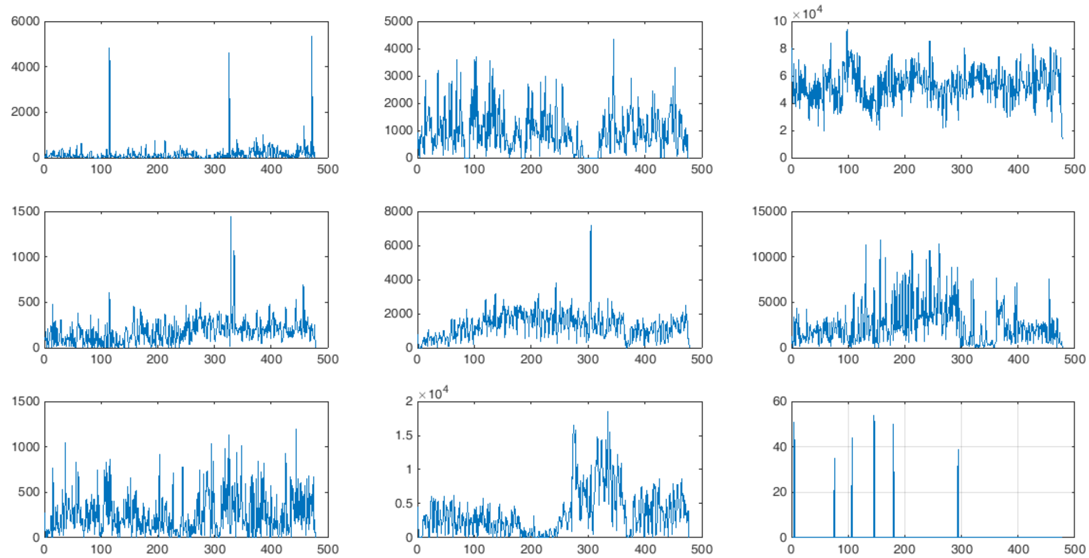


Рис. 1: Товары № 1, 2, 3, 12, 13, 19, 30, 34 и 39 отправленные с 83 региона

5.2 Тестирование суперпозиций

Рассматриваются временные ряды для различных товаров и исследуется прогнозы, построенных базовыми моделями и их суперпозициями. Приводятся графические иллюстрации прогнозов, а также численные значения ошибки. В качестве функции ошибки взята функция MAPE (mean absolute percentage error).

Таблица 2: Нефть

	MA	ARIMA	VAR	MA-MA	ARIMA-ARIMA	VAR-MA
MAPE, %	16,97	17,04	25,52	18,29	17,75	22,07

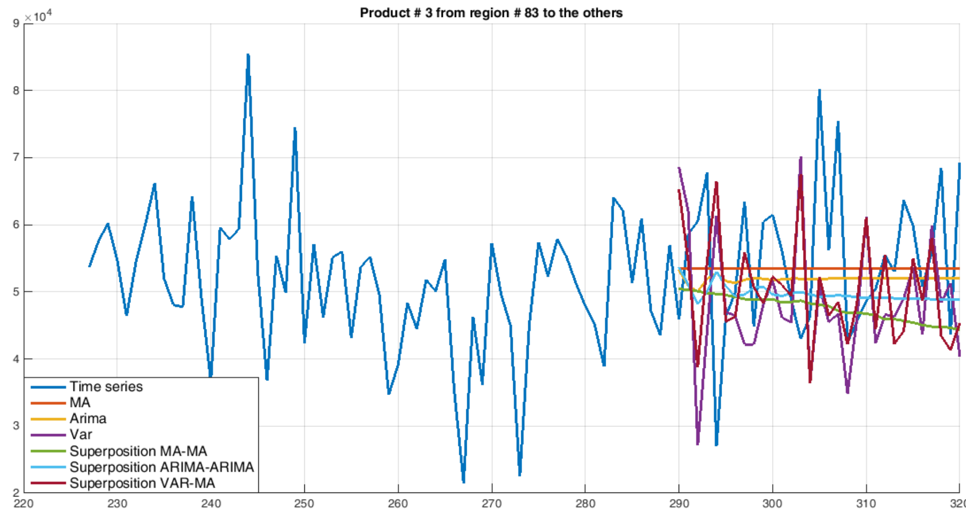


Рис. 2: Нефть, отправленная с 83 региона

Таблица 3: Метизы

	MA	ARIMA	Croston's method	MA-MA	ARIMA-MA
MAPE, %	42,06	37,66	36,18	41,60	37,25

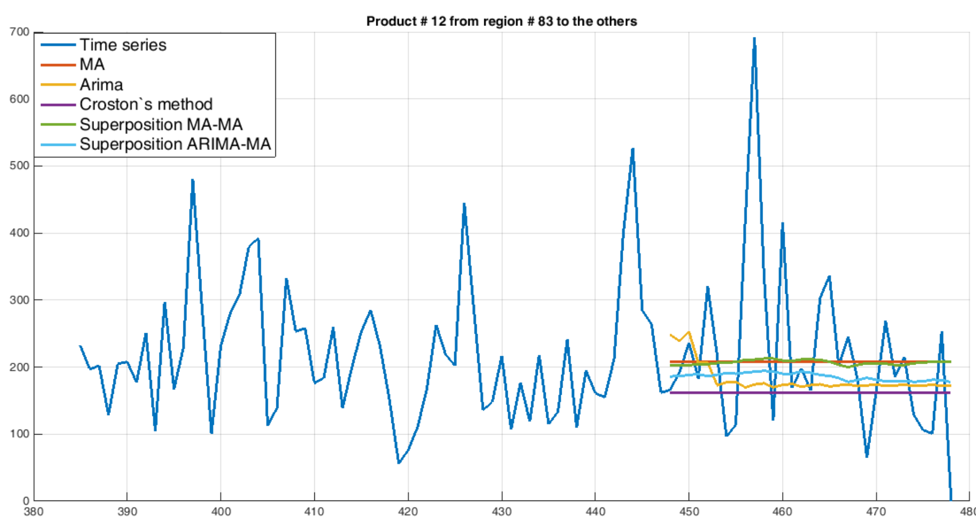


Рис. 3: Метизы, отправленные с 83 региона

Таблица 4: Каменный уголь

	Exponental smothing	ARIMA	VAR	MA	MA-MA	VAR-MA
MAPE, %	49,37	45,47	52,80	72,14	48,51	58,52

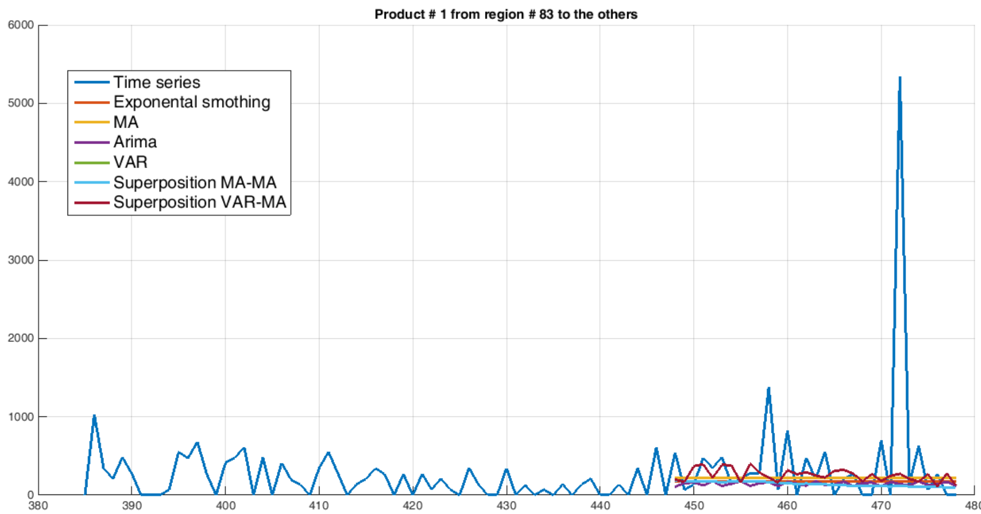


Рис. 4: Каменный уголь, отправленный с 83 региона

6 Заключение

Рассмотрена задача построения суперпозиции моделей при прогнозировании железнодорожных грузовых перевозок. В семейство моделей, из которого строилась суперпозиция, были включены следующие модели: скользящее среднее (MA), экспоненциальное сглаживание, интегрированная модель авторегриссионного скользящего среднего(ARIMA) и векторная авторегресси (VAR). По ним были построены следующие суперпозиции: MA-MA, ARIMA-ARIMA, MA-ARIMA, ARIMA-MA, VAR-VA. 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 гргрузов железнодорожных перевозок. 2014.

[3] В. Хардле. *Прикладная непараметрическая регрессия.* 1993.

[4] Ю.П. Лукашин. *Адаптивные методы краткосрочного прогнозирования.* М.:Финансы и статистика, 2003.

[5] В. В. Cnhb;jd. Порождение и выбор моделей в задачах регрессии и классификации. 2014.

[6] А.П.Мотренко М.М. Стенина. К.В. Рудаков, М.П. Кузнецов. Выбор оптимальной модели прогнозирования объемов грузовых железнодорожных перевозок. 2015.

[7] А. Д. Корчагин М. П. Кузнецов А. П. Мотренко М. М. Стенина В. В. Стрижов Ю. И. Журавлев, К. В. Рудаков. Создание системы прогнозирования объемов спроса на грузовые железнодорожные перевозки.

[8] М. Вербик. *Путеводитель по современнной эконометрике.* М.:Научная книга, 2008.

[9] В.В. Стрижов А.М. Мотренко, К.В. Рудаков. Учет влияния экзагенных факторов при непараметрическом прогнозировании временных рядов. 2015.

[10] R. J. HYNDMAN L. SHENSTONE. Stochastic models underlying croston's method for intermittent demand forecasting. *Journal of Forecasting*, 2005.

[11] Я.Р. Магнус. *Эконометрика.* М.:Дело, 2004.

[12] С.А. Айвазян. *Основы эконометрики.* М.:Юнитидана, 2001.