

Выбор суперпозиции моделей при прогнозировании грузовых железнодорожных перевозок

Двинских Дарина Михайловна

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

Научный руководитель д.ф.-м.н., В. В. Стрижов

2016 г.

Задача

Построить суперпозицию моделей для краткосрочного прогнозирования объемов железнодорожных грузовых перевозок

Проблема

- Отсутствие единой прогностической модели
- Большая волатильность временных рядов
- Наличие нулевых объемов перевозок

Требования к моделям

- Валидность
- Устойчивость
- Полнота

- Ю.И. Журавлев, К.В. Рудаков, А.Д. Корчагин, М.П. Кузнецов, А.П. Мотренко, С.С. Стенина, В.В. Стрижов, Создание системы прогнозирования объемов спроса на грузовые железнодорожные перевозки, 2016
- К.В. Рудаков, М.П. Кузнецов, А.П. Мотренко, М.М. Стенина, Д.О.Каширин, В.В. Стрижов, Выбор оптимальной модели прогнозирования грузовых железнодорожных перевозок, 2015

Суперпозиция

$\mathbf{x} = \{x_i\}_{i=1}^T$ — временной ряд,
 f и g - две базовые функции прогнозирования
(f прогнозирует ряд, а g — его остатки)

$$f : T \rightarrow \varepsilon, g : \varepsilon \rightarrow X$$

Тогда функция

$$s = f \circ g,$$

определенная равенством $f \circ g(\varepsilon) = g(f)(\mathbf{t})$, и такая, что
 $T \rightarrow X$, называется **суперпозицией** функций f и g

Посуточная загруженность железнодорожных путей:

$$\mathbf{X} = \{\{x_{j,i}\}_{j,i=1}^{8,T}\}$$

Вектор $\mathbf{x} = \{x_j\}_{i=j}^8$ имеет 8 компонент

1. дата погрузки
2. код станции отправления
3. код станции назначения
4. количество вагонов
5. код груза
6. род вагона
7. суммарный вес груза
8. признак маршрутной отправки

$D = \{t_i, x_i\}_{i=1}^T$ — регрессионная выборка, t_i — временные метки, x_i — объемы грузовых перевозок

Предположение, накладываемое на значения выборки

$$x_i = s(\hat{\mathbf{w}}, t_i) + \epsilon(t_i),$$

где $\hat{\mathbf{w}}$ — вектор оптимальных параметров, $\epsilon_i = \epsilon(t_i)$ — ошибка

$$\epsilon_i = s(\hat{\mathbf{w}}, t_i) - x_i$$

Требуется построить прогноз на t значений ряда вперед

$$\hat{x}_{T+h} = s(\hat{\mathbf{w}}, t),$$

где h — горизонт прогнозирования

Функция ошибки $S(\mathbf{w}|F, D)$ — функция, значение которой требуется минимизировать для получения оценок параметров $w \in \mathfrak{R}_s$ модели $s \in \mathfrak{F}$.

$$s, \hat{\mathbf{w}} = \arg \min_{\mathbf{w}} \sum_{t \in T} S(\mathfrak{F}, \mathfrak{R}_s),$$

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

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

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

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

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

$$\text{SS} = 1 - \frac{\text{MSE}_{\text{forecast}}}{\text{MSE}_{\text{history}}}$$

Анализ регрессионных остатков

1. Равенство нулю матожидания $E(\epsilon) = 0$
2. Постоянство дисперсии $D(\epsilon) = \sigma^2$
3. Нормальность остатков $\epsilon \sim \mathcal{N}(0, \sigma^2)$

Скользящее среднее

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

где n — ширина окна

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

Экспоненциальное сглаживание

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

где x — исходный ряд, z — сглаженный ряд, α — параметр сглаживания ряда, $\alpha \in (0, 1)$.

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

Метод Кростена

$\mathbf{d} = \{d_t\}_{t=1}^T$ — ненулевой спрос исходного временного ряда \mathbf{x} ,

$\mathbf{q} = \{q_t\}_{t=1}^T$ — интервалы между ненулевым спросом ряда \mathbf{x}

Экспоненциальное сглаживания обоих рядов

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

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

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

Модель ARIMA(p, d, q)

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

где ϵ_t - стационарный временной ряд

c, a_i, b_j - параметры модели

Δ^d - оператор разности временного ряда порядка d

Модель VAR

$$Y = XW,$$

где матрица X — матрица объект-признак, W — матрица весов, Y — матрица ответов

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

Тогда прогноз находится как

$$\hat{x} = xW$$

Таблица: Сопоставление различных прогностических моделей свойствам временных рядов

Алгоритмы	Свойства временных рядов		
	нестац.	нул. знач.	внеш. факторы
Скользящее среднее	-	+	-
Экспонен. сглаживание.	-	+	-
Метод Кростона	-	+	-
ARIMA	+	-	-
VAR	-	-	+

- 1 С помощью базовой функции f вычисляется $n = n(g)$ прогнозов конца истории $\hat{x}_t^f, \dots, \hat{x}_{t-n(g)+1}^f$ на одну точку.
- 2 Вычисляется $n(g)$ остатков $\hat{\varepsilon}_t, \dots, \hat{\varepsilon}_{t-n(g)+1}$ в виде разницы

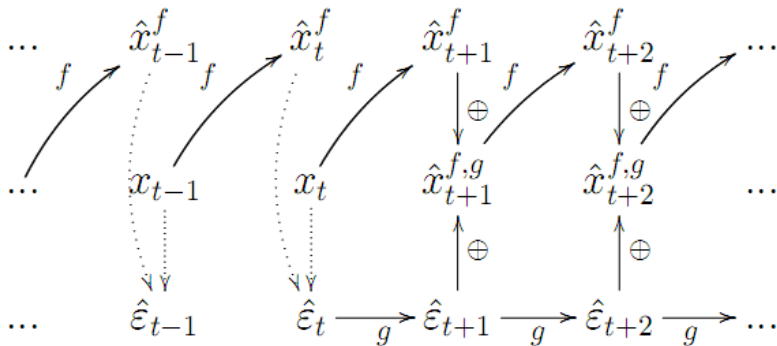
$$\hat{\varepsilon}_{t-k} = x_{t-k} - \hat{x}_{t-k}^f.$$

- 3 С помощью функции g прогнозируются остатки $\hat{\varepsilon}_{t+i}$ на $\max(i)$ отсчетов вперед.
- 4 Выполняется итеративный 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 .

Диаграмма прогнозирования исходного ряда и его остатков



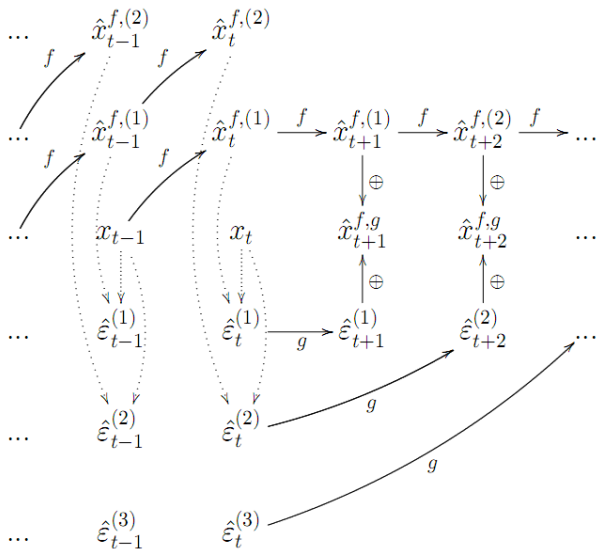
- 1 С помощью базовой функции f вычисляется ретроспективный прогноз $\hat{x}_{t+1}^{f,(1)}, \dots, \hat{x}_{t+i}^{f,(i)}$ с горизонтом прогнозирования i , каждый — на глубине i .
- 2 С помощью базовой функции f вычисляется $\max(i)$ наборов прогнозов конца истории, $\hat{x}_t^{f,(i)}, \dots, \hat{x}_{t-n(g)+1}^{f,(i)}$,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

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

- 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

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

Диаграмма прогнозирования исходного ряда и его остатков



Вычислительный эксперимент. Временные ряды

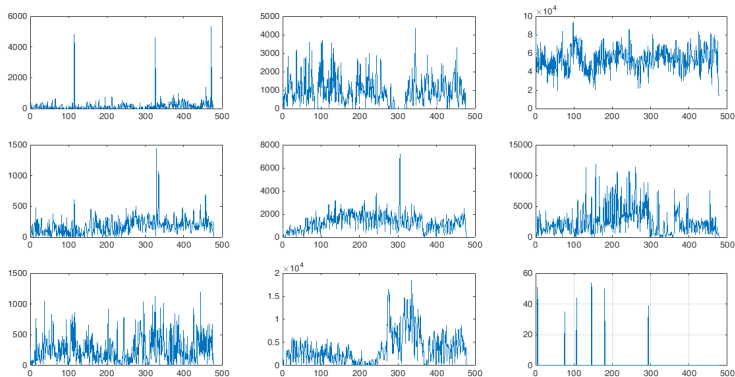
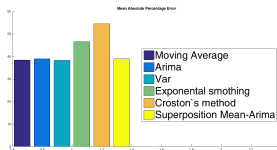
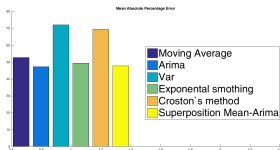
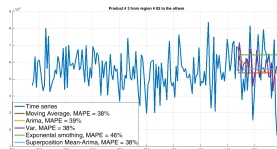
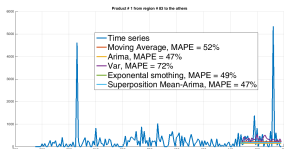
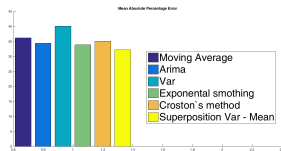
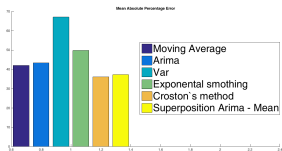
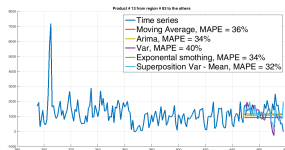
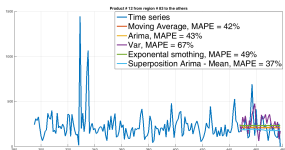


Рис.: Товары № 1, 2, 3, 12, 13, 15, 19, 30 и 34, отравленные с 83 региона

Вычислительный эксперимент



Вычислительный эксперимент



- Проведена предварительная агрегация временных рядов по станциям и районам
- Выбраны базовые модели для прогнозирования с учетом специфики временных рядов
- Построены суперпозиции моделей
- Проведен вычислительный эксперимент, сравнивающий качество базовых моделей и построенных суперпозиций
- Показано, что на некоторых рядах суперпозиция работает лучше