

Разработка и исследование свойств системы прогнозирования объемов грузовых железнодорожных перевозок

М. М. Стенина, М. П. Кузнецов, А. П. Мотренко,
Д. О. Каширин, А. Д. Корчагин,
К. В. Рудаков, В. В. Стрижов

Вычислительный центр
ФИЦ ИУ Российской академии наук

24 сентября 2015 г.

Проект выполнен при поддержке МО РФ, 14.604.21.0041

Решается проблема повышения эффективности транспортировки грузов. Для решения задачи выполняется прогноз потребностей у заказчиков РЖД в узлах погрузки/разгрузки с учетом временных интервалов доставки.

Цель проекта: повышение качества прогнозирования загруженности железнодорожных узлов за счет учета экзогенных факторов.

Требуется построить прогноз отправления/погрузки грузов

в заданном периоде:

- | | |
|--------------------------|---------------------------|
| 1) на месяц посуточно, | 4) на год помесечно, |
| 2) на месяц подекадно, | 5) на год поквартально, |
| 3) на квартал помесечно, | 6) на период больше года, |

с разложением

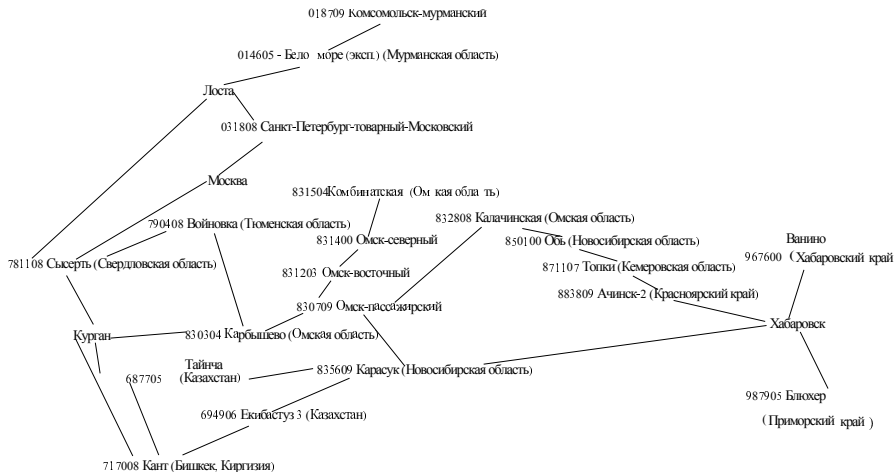
- 1) по группам грузов,
- 2) по парам станций/регионов,
- 3) по комбинированному разложению, учитывающему перечисленные варианты.

Запись о перемещении блока вагонов между парой станций включает

- 1 **коды станций** отправления и назначения (78 регионов, ~ 4000 станций, ~ 100 сортировочных),
- 2 **код груза** (нефть и нефтепродукты, сахар, продукты перемола и т.д., учтены 43 наименования),
- 3 **род вагонов** (полувагоны, крытые вагоны, цистерны, платформы, прочие).

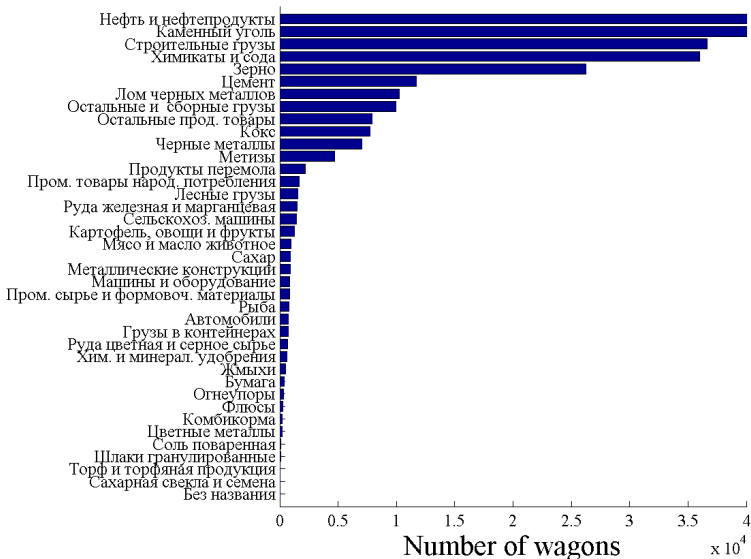
Дата погрузки	Ст. отправления	Ст. назначения	Число вагонов	Код груза	Род вагона	Вес груза, т.
2007-01-01	020108	932902	1	1	216	56
2007-01-01	032105	840109	1	19	040	63
2007-01-01	035508	843408	2	3	070	120

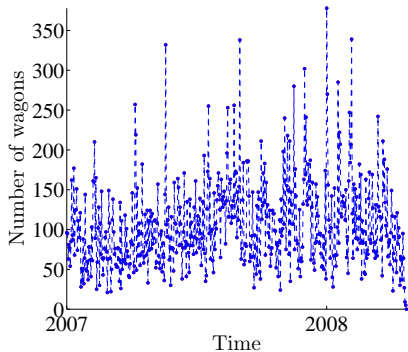
Топология основных железнодорожных путей



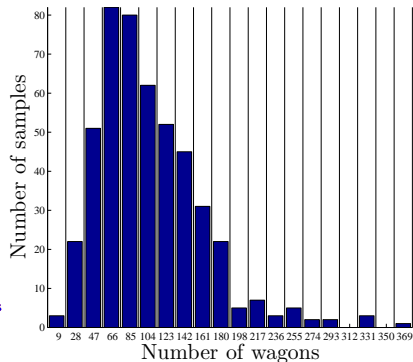
Загруженность узла различными типами грузов

Число вагонов, прибытие на станцию в течение года





(a) Временной ряд



(b) Гистограмма

Требуется построить прогнозы значений семейства временных рядов, связанных в иерархическую многоуровневую структуру и описывающих объемы погрузки ряда грузов в заданных узлах или на парах узлов с разным уровнем детализации.

Прогностическая модель должна удовлетворять следующим требованиям.

- Прогнозы должны быть точны — обеспечивать минимум возможное значение заданной функции потерь.
- Должны быть учтены внешние факторы, влияющие на объемы грузоперевозок.
- Прогнозы должны удовлетворять условию согласованности — структуре иерархии.
- Прогнозы должны удовлетворять физическим ограничениям — лежать в заданном интервале для каждого временного ряда.

Пусть значения ряда $x = [x_1, \dots, x_T]$ принадлежат одному распределению с плотностью $p(u)$.

Требуется спрогнозировать \hat{x} следующего значения x_{T+1} временного ряда, минимизирующий ожидание заданной функции потерь $l(\hat{x}, x_{T+1})$

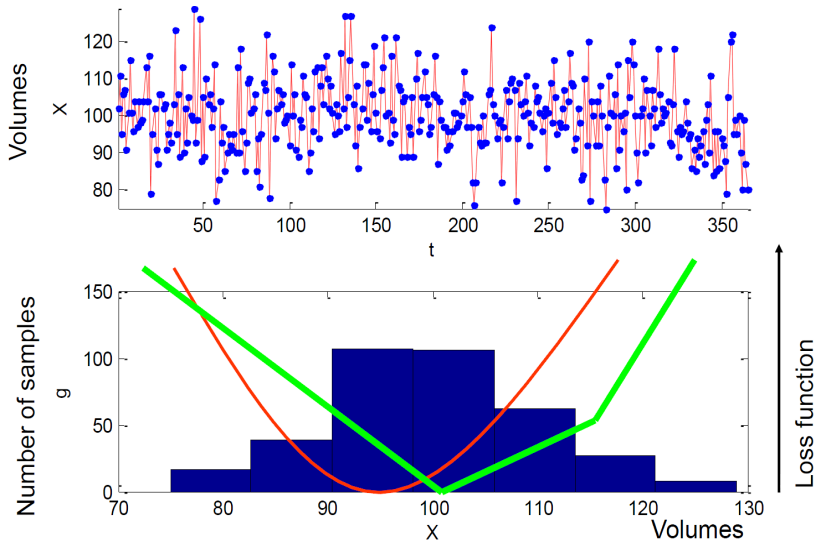
$$\hat{x} = \underset{c}{\operatorname{argmin}} \int_{u_{\min}}^{u_{\max}} l(c, u) p(u) du = \underset{c}{\operatorname{argmin}} L(c).$$

Для некоторых функций потерь при известной $p(u)$ прогноз вычислим аналитически

$$l(\hat{x}, x_{T+1}) = (\hat{x} - x_{T+1})^2: \hat{x} = c^* = E(x),$$

$$l(\hat{x}, x_{T+1}) = |\hat{x} - x_{T+1}|: \hat{x} = c^* = \operatorname{median} p(u).$$

Гистограммное прогнозирование, иллюстрация



Для оценки плотности $p(u)$ построим гистограмму $\hat{p}(u)$

с n интервалами (u_i, u_{i+1}) , где

$u_0 = u_{min} \leq u_1 \leq \dots \leq u_{n-1} \leq u_n = u_{max}$ — концы отрезков постоянства функции, $u_{i+1} - u_i = b$

h_1, \dots, h_n — значения гистограммы. Тогда $L(c) \approx L_{hist}(c)$, где

$$L_{hist} = \int_{u_{min}}^{u_{max}} l(c, u) \hat{p}(u) du = \sum_{i=1}^n h_i \int_{u_{i-1}}^{u_i} l(c, u) du.$$

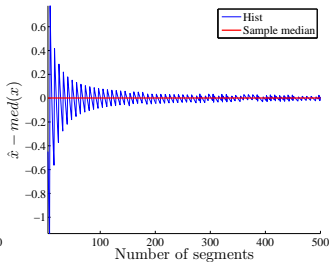
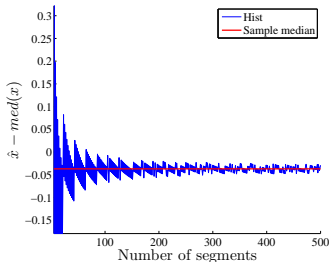
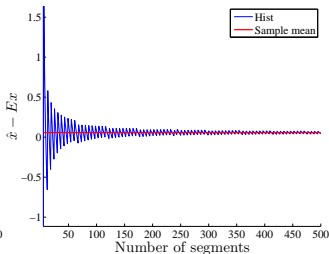
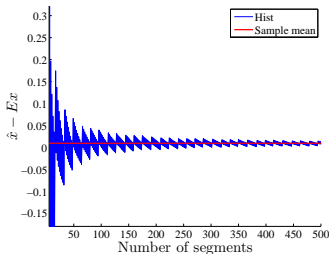
Упростим интегрирование $l(c, u)$, перейдем к свертке

$$\int_{u_{i-1}}^{u_i} l(c, u) du \approx l\left(c, \frac{u_i + u_{i-1}}{2}\right) (u_i - u_{i-1} = b) l\left(c, \frac{u_i + u_{i-1}}{2}\right).$$

Приближенное решение выбирается из множества точек

$$c^* \in \left\{ \frac{u_1 + u_0}{2}, \dots, \frac{u_n + u_{n-1}}{2} \right\}.$$

Сверху: квадратичная функция потерь, снизу: абсолютная функция потерь. Слева: $\mathcal{N}(0, 1)$, $\chi^2(3)$.



Временной ряд описывается моделью $ARIMA(p, d, q)$, если ряд его разностей

$$\nabla^d x_t = (1 - L)^d x_t, \text{ где } L^\tau x_t = x_{t-\tau}, \tau = 1, \dots, d$$

описывается моделью $ARMA(p, q)$:

$$x_t = \alpha + \varphi_1 x_{t-1} + \dots + \varphi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q},$$

где $\alpha = \mu(1 - \varphi_1 - \dots - \varphi_p)$, $\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q$ — константы, ε_t — гауссов белый шум с нулевым средним и постоянной дисперсией.

Модель ARIMA с мультипликативной сезонностью.

Временной ряд описывается моделью $SARIMA(p, d, q) \times (P, D, Q)_s$, если

$$\Phi_P(L^s)\varphi(L)\nabla_s^D \nabla^d x_t = \alpha + \Theta_Q(L^s)\theta(L)\varepsilon_t,$$

$$\Phi_P(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_P L^{Ps},$$

$$\Theta_Q(L^s) = 1 + \Theta_1 L^s + \Theta_2 L^{2s} + \dots + \Theta_Q L^{Qs}.$$

- 1 Модель $ARIMA(p, d, q)$ применима для прогнозирования нестационарных временных рядов, но только в случае симметричной функции потерь.
- 2 Модель Hist применима в случае несимметричных функций потерь, но только для прогнозирования стационарных временных рядов.

Комбинированный прогноз найдем в виде декомпозиции

$$\hat{x} = \hat{x}^{ns} + \hat{x}^s,$$

где прогноз нестационарной части \hat{x}^{ns} определяется моделью $SARIMA(p, d, q) \times (P, D, Q)_s$,
стационарной части \hat{x}^s — моделью Hist.

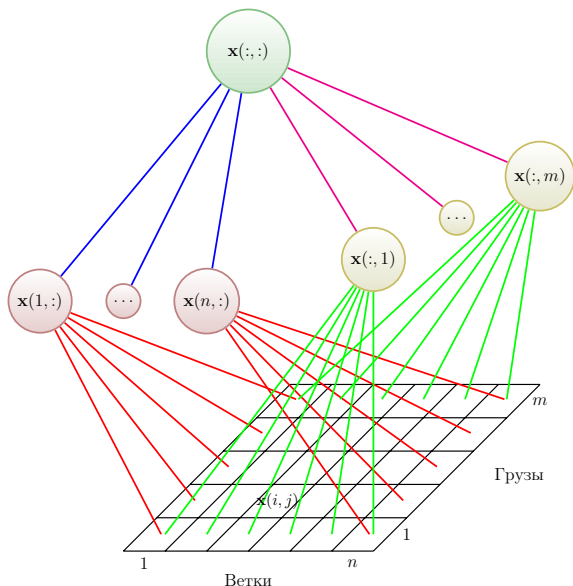
Вход: временной ряд $\mathbf{x} = \{x_1, \dots, x_T\}$, функция потерь $l(\hat{x}, x_{T+1})$;

Выход: прогноз \hat{x} ;

- 1: подобрать подходящую для временного ряда модель *ARIMA* по методологии Бокса-Дженкинса;
 - 2: вычислить прогноз нестационарной компоненты \hat{x}^{ns} с помощью выбранной модели *ARIMA*;
 - 3: вычислить регрессионные остатки $\mathbf{r} = \{r_1, \dots, r_T\}$ для выбранной модели *ARIMA*;
 - 4: задать число столбцов в гистограмме для алгоритма *Hist*;
 - 5: вычислить прогноз стационарной компоненты \hat{x}^s с помощью алгоритма *Hist*;
 - 6: $\hat{x} = \hat{x}^{ns} + \hat{x}^s$;
-

В таблице приведены средние потери прогнозирования временного ряда цен на сахар.

Алгоритм	Квадратичная функция потерь	Абсолютная функция потерь	Асимметричная функция потерь
<i>ARIMA</i>	0.127	0.265	0.340
<i>ARIMA + Hist(20)</i>	0.128	0.267	0.260
<i>ARIMA + Hist(50)</i>	0.127	0.266	0.267
<i>ARIMA + Hist(300)</i>	0.127	0.265	0.266
<i>ARIMA + Hist(500)</i>	0.127	0.265	0.266



$$x_t(:, :) = \sum_{i=1}^n x_t(i, :);$$

$$x_t(:, :) = \sum_{j=1}^m x_t(:, j);$$

$$x_t(i, :) = \sum_{j=1}^m x_t(i, j),$$

$$i = 1, \dots, n;$$

$$x_t(:, j) = \sum_{i=1}^n x_t(i, j),$$

$$j = 1, \dots, m;$$

$$t = 1, \dots, T.$$

Прогнозы, полученные
 дого временного ряда
 симо, могут не удовл
 структуре иерархии, т. е.
 согласованными.

Срез иерархии, вектор независимых и вектор согласованных прогнозов:

$$\chi_t = \begin{pmatrix} x_t(:, :) \\ \dots \\ x_t(n, 1) \\ \dots \\ x_t(n, m) \end{pmatrix}, \quad \hat{\chi} = \begin{pmatrix} \hat{x}(:, :) \\ \dots \\ \hat{x}(n, 1) \\ \dots \\ \hat{x}(n, m) \end{pmatrix}, \quad \hat{\varphi} = \begin{pmatrix} \hat{y}(:, :) \\ \dots \\ \hat{y}(n, 1) \\ \dots \\ \hat{y}(n, m) \end{pmatrix}.$$

Условие согласованности $\mathbf{S}\chi_t = \mathbf{0}$, $t = 1, \dots, T$,

где \mathbf{S} — матрица связей, имеет размер $(2 + n + m) \times (1 + n + m + nm)$ и записывается в виде

$$\mathbf{S} = \begin{pmatrix} -1 & 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ -1 & 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ \hline 0 & -1 & \dots & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\ \dots & & \dots & & & \dots & & & & \dots & & & & & \dots & & \\ 0 & 0 & \dots & -1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \\ \hline 0 & 0 & \dots & 0 & -1 & \dots & 0 & 1 & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \\ \dots & & \dots & & & \dots & & & & \dots & & & & & \dots & & \\ 0 & 0 & \dots & 0 & 0 & \dots & -1 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Дано Матрица связей \mathbf{S} , множества \mathcal{A} , \mathcal{B} и вектор независимых прогнозов $\hat{\chi}$

$$\hat{\chi} \notin \mathcal{A}, \quad \hat{\chi} \in \mathcal{B}.$$

Требуется построить вектор согласованных прогнозов $\hat{\varphi}$, который удовлетворяет следующим требованиям:

- $\hat{\varphi} \in \mathcal{A}$, $\mathcal{A} = \{\chi \in \mathbb{R}^d \mid \mathbf{S}\chi = \mathbf{0}\}$ — согласованность;
- $\hat{\varphi} \in \mathcal{B}$ — физические ограничения;
- $l_h(\chi_{T+1}, \hat{\varphi}) \leq l_h(\chi_{T+1}, \hat{\chi})$ для любого среза действительных значений $\chi_{T+1} \in \mathcal{A} \cap \mathcal{B}$ — качество.

Игрок, выбирающий вектор согласованных прогнозов $\hat{\varphi}$, играет с природой, выбирающей срез иерархии в момент времени $(T + 1)$. Цель игрока — минимизировать свои потери при любом ходе природы.

	Стратегия	Потери
Игрок	$\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}$	$L(\hat{\varphi}, \chi_{T+1}) = I_h(\chi_{T+1}, \hat{\varphi}) - I_h(\chi_{T+1}, \hat{\chi})$
Природа	$\chi_{T+1} \in \mathcal{A} \cap \mathcal{B}$	$-L(\hat{\varphi}, \chi_{T+1})$

Равновесие Нэша в антагонистической игре — это пара стратегий $(\hat{\varphi}, \chi_{T+1})$, таких что для любых стратегий $\hat{\varphi}'$, χ'_{T+1} выполнено неравенство

$$L(\hat{\varphi}, \chi'_{T+1}) \leq L(\hat{\varphi}, \chi_{T+1}) \leq L(\hat{\varphi}', \chi_{T+1}).$$

Цена игры (Дж. Нэш)

$$V = \min_{\hat{\varphi}} \max_{\chi_{T+1}} L(\hat{\varphi}, \chi_{T+1}) = \max_{\chi_{T+1}} \min_{\hat{\varphi}} L(\hat{\varphi}, \chi_{T+1})$$

определена тогда и только тогда, когда в игре существует равновесие Нэша.

Теорема 1 (Стенина, 2014)

Пусть $\mathcal{A} \cap \mathcal{B} \neq \emptyset$ и для функции потерь l_h выполнено

- 1 $l_h(\chi_{T+1}, \hat{\chi}) \geq 0$ для произвольных векторов χ_{T+1} , $\hat{\chi}$,
причем $l_h(\chi_{T+1}, \hat{\chi}) = 0 \Leftrightarrow \chi_{T+1} = \hat{\chi}$;
- 2 существует проекция $\chi_{proj} = \operatorname{argmin}_{\chi \in \mathcal{A} \cap \mathcal{B}} l_h(\chi, \hat{\chi})$;
- 3 для всех $\chi \in \mathcal{B}$ и для всех $\psi \in \mathcal{A} \cap \mathcal{B}$ выполняется неравенство $l_h(\psi, \chi) \geq l_h(\psi, \chi_{proj}) + l_h(\chi_{proj}, \chi)$.

Тогда

- пара стратегий $(\chi_{proj}, \chi_{proj})$ является равновесием Нэша в антагонистической игре, описывающей задачу согласования прогнозов;
- пара $(\chi_{proj}, \chi_{proj})$ является седловой точкой функции $L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$.

Теорема 2: цена игры (Стенина, 2014)

При выполнении требований теоремы 1 цена игры определена и равна

$$V = \min_{\hat{\varphi}} \max_{\chi_{T+1}} L(\hat{\varphi}, \chi_{T+1}) = \max_{\chi_{T+1}} \min_{\hat{\varphi}} L(\hat{\varphi}, \chi_{T+1}) = -l_h(\chi_{proj}, \hat{\chi}) \leq 0.$$

Теорема 3: согласованные прогнозы (Стенина, 2014)

При выполнении требований теоремы 1 использование в качестве вектора согласованных прогнозов $\hat{\varphi}$ вектора

$$\hat{\varphi} = \chi_{proj} = \operatorname{argmin}_{\chi \in A \cap B} l_h(\chi, \hat{\chi})$$

гарантирует, что вектор согласованных прогнозов будет удовлетворять требованиям согласованности и качества и физическим ограничениям.

Задача согласования прогнозов сводится к решению оптимизационной задачи.

Вход: вектор независимых прогнозов $\hat{\chi}$, матрица связей \mathbf{S} , множества \mathcal{A} и \mathcal{B} , функция потерь $l_h(\cdot, \cdot)$;

Выход: вектор согласованных прогнозов $\hat{\varphi}$;

$$1: \hat{\varphi} = \operatorname{argmin}_{\chi \in \mathcal{A} \cap \mathcal{B}} l_h(\chi, \hat{\chi});$$

Свойства алгоритма, согласно теоремам 1, 2, 3 [Стенина, 2015].

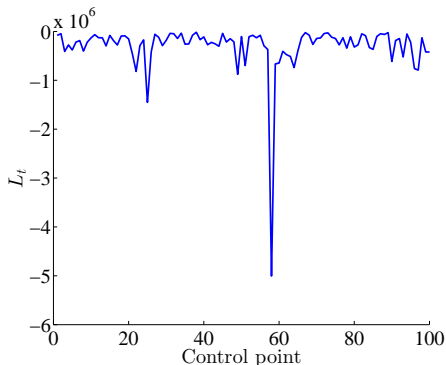
- Позволяет согласовывать прогнозы, одновременно обеспечивая выполнение физических ограничений и неухудшение качества прогнозирования.
- Не требует оценки погрешности независимых прогнозов и их несмещенности.
- На независимые прогнозы накладываются только физические ограничения.
- Работает с иерархическими структурами любой сложности.
- Для решения оптимизационной задачи можно использовать стандартные методы.

Для согласования прогнозов $H = 100$ последних точек истории решалась оптимизационная задача $\hat{\varphi} = \operatorname{argmin}_{\chi \in A \cap B} \|\chi - \hat{\chi}\|_2^2$.

Изображена величина

$$L_t = \|\chi_t - \hat{\varphi}\|_2^2 - \|\chi_t - \hat{\chi}\|_2^2, \quad t = (T - H + 1), \dots, T.$$

Во всех контрольных точках потери уменьшились.



Функция потерь $l_h(x_t, \hat{x}) = \|x_t - \hat{x}\|_2^2$.

Средние потери прогнозирования отгрузки в узлах РЖД, $\times 10^8$

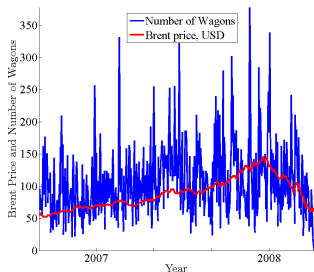
Уровень иерархии	Независимые прогнозы	Восходящее согласование ¹	Оптимальная регрессия ²	Модиф. теор.-игр. согл. (веса 700)
Вся иерархия	10.038	9.999	10.035	9.969
Верхний уровень	2.858	2.868	2.856	2.840
Средний уровень, ветки	2.549	2.486	2.545	2.487
Средний уровень, грузы	2.338	2.351	2.340	2.348
Нижний уровень	2.294	2.294	2.294	2.294

¹ Albert B. Schwarzkopf, Richard J. Tersine, John S. Morris *Top-down versus bottom-up forecasting strategies*. The International Journal Of Production Research, 26(11):1833–1843, 1988.

² Rob J. Hyndman, Roman A. Ahmed, George Athanasopoulos, Han Lin Shang. *Optimal combination forecasts for hierarchical time series*. Computational Statistics and Data Analysis, 55(9):2579–2589, 2011.

Задача повышения точности прогноза решается путем введения в прогностическую модель следующих факторов:

- сезонности и данных о погоде (для учета перевозок сельскохозяйственной продукции),
- биржевых цен на перевозимые грузы на российских и зарубежных биржах.



На графике синей линией показано суммарное количество вагонов на фиксированной станции, красной линией — цена за баррель нефти. Цена за баррель является фактором, используемым для прогнозирования объема поставок.

Использование экспертных высказываний о влиянии внешних событий на объем погрузок

Задан набор экспертных утверждений о влиянии внешних факторов на грузоперевозки в порядковых шкалах (например, степень влияния низкая, высокая, либо фактор не оказывает влияния).

Вид фактора, влияющего на объем грузоперевозок	Группы грузов и отрасли, на которые оказывается влияние	Степень и характер влияния
Мировые и внутренние цены на соответствующие активы	Нефть и нефтепродукты, черные металлы, цветные металлы, удобрения, уголь и др.	На экспортные перевозки влияние сильное
Курс рубля к доллару	Грузы, отправляемые на экспорт (нефть и нефтепродукты, металлы, уголь)	Степень влияния для экспортных перевозок зачастую высокая
Сезонность производства природного-климатического характера	Зерно, овощи, бахчевые культуры	Степень влияния высокая. Динамика перевозки связана со сбором урожая

При построении прогностической модели используются экспертные высказывания о влиянии внешних факторов $S = \{s_1, \dots, s_M\}$ на прогнозируемые ряды x .

В связи с возможной несогласованностью экспертных высказываний и исторических данных решается задача определения их достоверности $\mu = \mu(x, s_i)$.

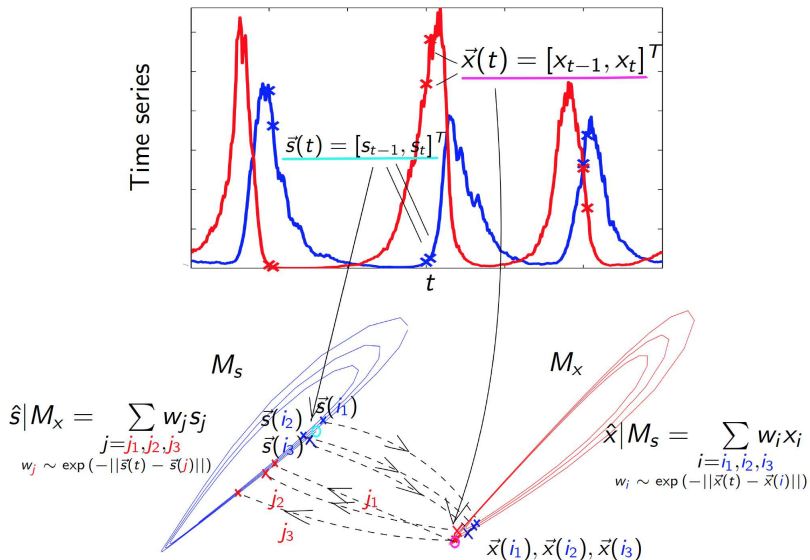
Тест Грэнжера:

$$\mu(x, s) \sim \frac{ESS_x - ESS_{x,s}}{ESS_{x,s}} \frac{T - k_{x,s}}{k_{x,s} - k_x},$$

$k_{x,s}$, k_x — порядок расширенной $\hat{x} = \hat{x}(x, s)$ и базовой $\hat{x} = \hat{x}(x)$ моделей прогнозирования, $ESS_{x,s}$, ESS_x — суммы квадратов регрессионных остатков расширенной и базовой моделей:

$$ESS = \sum_{t=1}^T (\hat{x}_t - x_t)^2.$$

Метод сходящегося перекрестного отображения



Определим множества

$$M_x = \{\vec{x} \mid \vec{x}(t) = [x_{(t-D)}, \dots, x_t]^T\}$$

$$M_s = \{\vec{s} \mid \vec{s}(t) = [s_{(t-D)}, \dots, s_t]^T\}$$

Пусть i_1, i_2, \dots, i_{D+1} — индексы ближайших соседей точки $\vec{s}(T+1)$ множества M_s . Тогда прогноз $\hat{x} \mid M_s$ сходящегося перекрестного отображения есть

$$\hat{x} \mid M_s = \sum_{d=1}^{D+1} w_d x_{i_d}, \quad w_d = \frac{u_d}{\sum_d u_d}, \quad u_d = \exp\left(-\frac{\|\vec{x}(T+1), \vec{x}(i_d)\|}{\|\vec{x}(T+1), \vec{x}(i_1)\|}\right).$$

Решение о наличии связи принимаем при больших значениях

$$\mu(\mathbf{x}, \mathbf{s}) \sim \rho(\hat{x}, x_{T+1}) = \frac{1}{\sigma_x \sigma_{\hat{x}}} E(\hat{x} - E\hat{x})(x_{T+1} - E x_{T+1})$$

Решение о достоверности экспертного высказывания о зависимости объемов погрузок от внешних факторов принимается на основе теста Грэнжера и анализа перекрестных отображений исследуемых временных рядов.

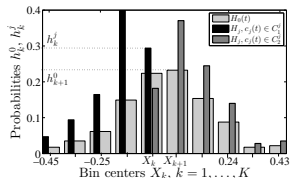
Исследуемые ряды	Высказывание эксперта	Оценка достоверности	Решение о наличии связи
«Нефть и нефтепродукты» — «Цены на нефть»	Связь сильная	0.99	Положительное
«Сахар» — «Цены на сахар»	Связь сильная	0.97	Положительное
«Продукты перемола» — «Цены на пшеницу»	Связь сильная	0.46	Отрицательное

Для повышения качества прогнозирования моделью Hist путем учета экзогенных временных рядов $\mathbf{s}^j, j = 1, \dots, n$ уточним гистограмму $p^0(T) = [h_1^0, \dots, h_K^0]$ прогнозируемого временного ряда $\mathbf{x} = [x_1, \dots, x_T]^T$ линейной комбинацией

$$\hat{p}(T) = w_0 p_0(T) + \sum_{j=1}^n w_j p_j(T),$$

где $\hat{p}^j(T) = [h_1^j, \dots, h_K^j]$ — гистограммы прогнозируемого временного ряда \mathbf{x} , условные по экзогенным временным рядам $\mathbf{s}^j = [s_1^j, \dots, s_{T-\tau}^j]$.

	s_1^j	...	s_g^j	...	s_N^j	\sum_g
X_1	p_{11}	...	$p_{1g} = h_1^j \cdot p_g$...	p_{1N}	h_1^0
X_2	p_{21}	...	$p_{2g} = h_2^j \cdot p_g$...	p_{2N}	h_2^0
...
X_K	p_{K1}	...	$p_{Kg} = h_k^j \cdot p_g$...	p_{KN}	h_K^0
\sum_k	p_1^j	...	p_g^j	...	p_N^j	1



Включение экзогенных временных рядов s^j в модель контролируется вектором параметров $\mathbf{w} = [w_0, \dots, w_n]^T$, максимизирующим правдоподобие модели

$$\mathbf{w} = \underset{\substack{\mathbf{w} \in [0,1]^{|\mathcal{J}|} \\ \sum_{j \in \mathcal{J}} w_j = 1}}{\operatorname{argmax}} \frac{1}{|\mathcal{J}|} \sum_{t=1}^T \log \left(\sum_{j \in \mathcal{J}} w_j h_k^j(t) \right), \text{ где } \mathcal{J} = \{j : w_j \neq 0\}.$$

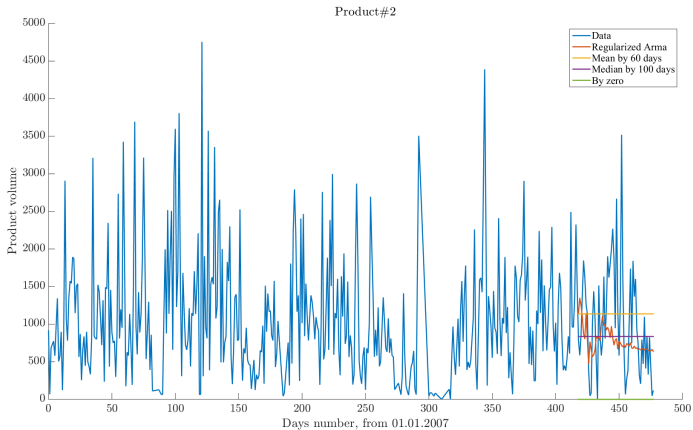
В таблице приведены результаты выбора s^j , увеличивающих качество прогнозирования с достоверностью не меньше 95% и относительное уменьшение потерь $\Delta L/L_0$.

Группа грузов	c_j	$\Delta L/L_0$
Нефть и нефтепродукты	Свинец, 2 τ	0.27873
Черные металлы	Свинец, τ	0.29659
Металлические конструкции	Свинец, τ	0.1044
Метизы	Свинец, τ	0.43858
Хмикаты и сода	Свинец, 2 τ	0.41155
Строительные грузы	Свинец, τ	0.34462
Шлаки гранулированные	Свинец, τ	0.17258
Огнеупоры	Свинец, 2 τ	0.086756
Цемент	Свинец, τ	0.31271

- Эксперимент проводился на реальных данных о посуточной загрузженности железнодорожных узлов.
- Было проведено сравнение следующих алгоритмов:
 - 1 Базовый алгоритм прогнозирования средним.
 - 2 Модель авторегрессии-скользящего среднего.
 - 3 Алгоритм гистограммного прогнозирования.
 - 4 Алгоритм, учитывающий экзогенные факторы.
- Прогноз выполнялся в разрезе:
 - 1 временном (на день, неделю и месяц вперед)
 - 2 пространственном (по парам станций и парам регионов).
- В качестве функций ошибок рассматривались функции
 - 1 абсолютной стандартной ошибки $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$,
 - 2 нормированной абсолютной ошибки $MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}$.

Временной ряд — суммарное значение перевозок каменного угля по всем парам станций.

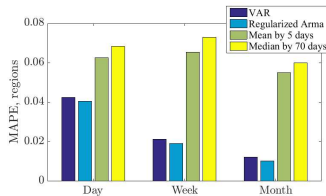
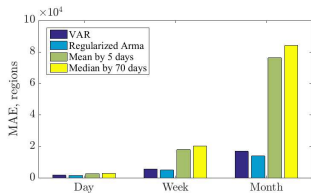
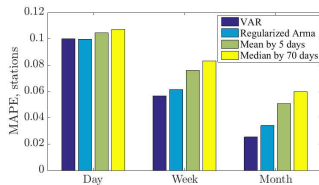
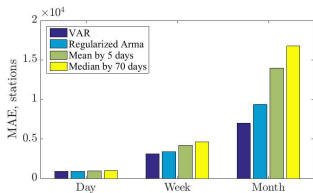
Прогноз строился на 60 последних дней.



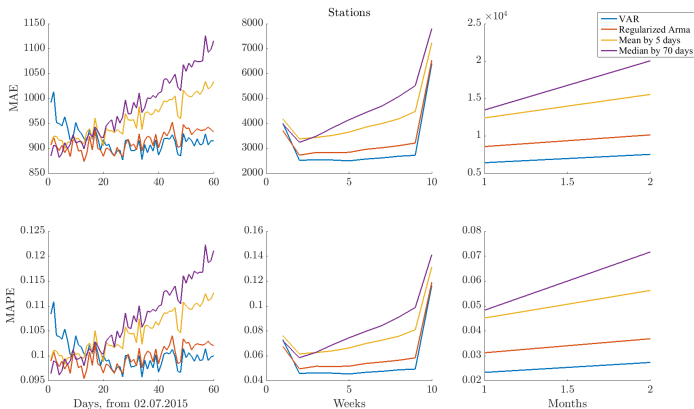
	Stations			Regions		
	D	W	M	D	W	M
MAE						
VAR	917.104	3114.073*	6986.463*	1946.955*	5855.633	17053.388
Regularized Arma	913.336*	3381.220	9366.639	1853.329	5168.458*	14056.409*
Mean by 5 days	959.840	4195.548	13987.420	2892.025	18165.117	76272.813
Hist	984.188	4623.588	16761.591	3168.165	20385.037	84069.813
MAPE						
VAR	0.100	0.057*	0.025*	0.042	0.021	0.012
Regularized Arma	0.100*	0.062	0.034	0.040*	0.019*	0.010*
Mean by 5 days	0.105	0.076	0.051	0.063	0.065	0.055
Hist	0.107	0.083	0.060	0.068	0.073	0.060

- В разрезе по парам регионов лучшие результаты продемонстрировал алгоритм, учитывающий экзогенные факторы.
- В разрезе по парам станций — модель авторегрессии.

Результаты прогнозирования с агрегированием во времени



По оси абсцисс отложен временной отсчет ретроспективного прогнозирования (на 60 дней, 9 недель или 2 месяца вперед).



Для модели авторегрессии и метода, учитывающего экзогенные факторы, ошибка ретроспективного прогноза возрастает медленнее.

Разработана система прогнозирования объемов железнодорожных грузоперевозок.

- 1 Прогнозы выполняются по одному узлу или по парам узлов (шахматка).
- 2 Выполняется агрегирование прогноза по времени: дни, недели, месяцы.
- 3 Выполняется агрегирование по типам грузов, регионам.
- 4 Прогноз согласуется по иерархическому агрегированию.
- 5 Учет внешних факторов может повысить качество прогноза.