Согласование прогнозов в задачах прогнозирования иерархических временных рядов (пример)

Мария Стенина

Московский физико-технический институт
Физтех-школа прикладной математики и информатики
Кафедра интеллектуальных систем

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

61-я Всероссийская научная конференция МФТИ
Москва, 2020

Цель работы

Задача

Построить прогнозы семейства временных рядов, связанных в иерархическую многоуровневую структуру и описывающих объемы погрузки ряда грузов в заданных узлах РЖД с разным уровнем детализации.

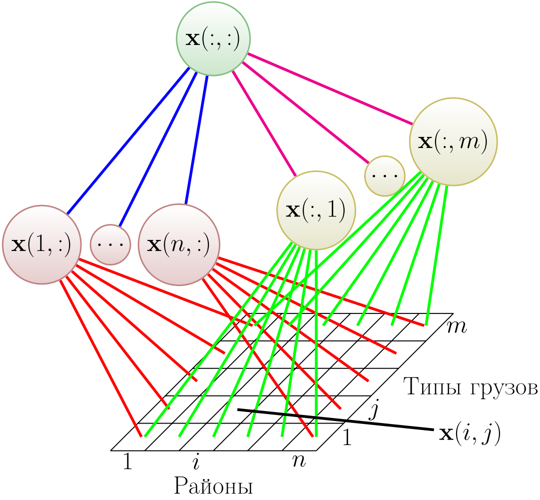
Требования к модели

- прогнозы должны быть точны — обеспечивать минимально возможное значение заданной функции потерь;
- прогнозы должны удовлетворять физическим ограничениям — лежать в заданном интервале для каждого временного ряда;
- прогнозы должны удовлетворять условию согласованности (структуре иерархии).

Проблема согласования прогнозов

Прогнозы, полученные для каждого временного ряда независимо, могут не удовлетворять структуре иерархии, т. е. не быть *согласованными*.

Условие согласованности прогнозов



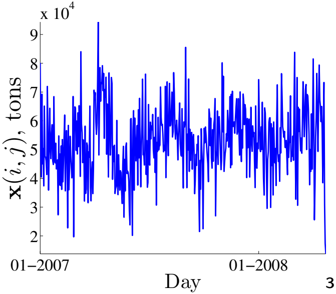
$$x_t(:,:) = \sum_{i=1}^{n} x_t(i,:);$$

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

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

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

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



Существующие решения

Алгоритмы согласования прогнозов

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.

Базовые публикации

1. Tim Van Erven and Jairo Cugliari. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. 2013.
2. M. M. Стенина, В. В. Стрижов. *Согласование агрегированных и детализированных прогнозов при решении задач непараметрического прогнозирования.* Системы и средства информатики, 24(2):21–34, 2014.

Обозначения: структура иерархии

Срез иерархии, вектор независимых и вектор согласованных прогнозов:

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

Условие согласованности $\quad \mathbf{S}\boldsymbol{\chi}_t = \mathbf{0}, \ t = 1, \dots, T,$

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

$$\mathbf{S} = \left(\begin{array}{c|ccc|ccc|cccc|c|cccc}
-1 & 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
-1 & 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
\hline
0 & -1 & \dots & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 & \dots & 0 & 0 & \dots & 0 \\
\dots & & \dots & & & \dots & & & & \dots & & & & & \dots & \\
0 & 0 & \dots & -1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 1 & 1 & \dots & 1 \\
\hline
0 & 0 & \dots & 0 & -1 & \dots & 0 & 1 & 0 & \dots & 0 & \dots & 1 & 0 & \dots & 0 \\
\dots & & \dots & & & \dots & & & & \dots & & & & & \dots & \\
0 & 0 & \dots & 0 & 0 & \dots & -1 & 0 & 0 & \dots & 1 & \dots & 0 & 0 & \dots & 1
\end{array}\right).$$

Обозначения: ограничения и функция потерь

Множество векторов, удовлетворяющих условию согласованности

$$\mathcal{A} = \{\boldsymbol{\chi} \in \mathbb{R}^d \mid \mathbf{S}\boldsymbol{\chi} = \mathbf{0}\}.$$

$$\boldsymbol{\chi}_1, \dots, \boldsymbol{\chi}_{T+1},\ \hat{\boldsymbol{\varphi}} \in \mathcal{A}, \quad \hat{\boldsymbol{\chi}} \notin \mathcal{A}.$$

Прогнозы и исторические значения временных рядов удовлетворяют физическим ограничениям

$$\boldsymbol{\chi}_1, \dots, \boldsymbol{\chi}_{T+1},\ \hat{\boldsymbol{\chi}},\ \hat{\boldsymbol{\varphi}} \in \mathcal{B},$$

$$\mathcal{B} = \{\boldsymbol{\chi} \in \mathbb{R}^d \mid \boldsymbol{\chi}(i) \in [A_i, B_i],\ \text{для всех } i = 1, \dots, d\},$$

$$A_i, B_i \in [-\infty; +\infty] \text{ для всех } i = 1, \dots, d.$$

Для задачи прогнозирования объемов погрузки

$$A_i = 0,\ B_i = +\infty, \quad i = 1, \dots, d.$$

Качество прогнозов оценивается с помощью функции потерь, которая зависит от вектора прогнозов и среза иерархии в момент времени $(T+1)$

$$l_h(\boldsymbol{\chi}_{T+1}, \hat{\boldsymbol{\chi}}).$$

Постановка задачи согласования прогнозов

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.

Согласование как антагонистическая игра

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	$\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}$	$L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$
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.

Существование равновесия Нэша и выбор согласованных прогнозов

Теорема 1 (Стенина, 2014)

Пусть $\mathcal{A} \cap \mathcal{B} \neq \varnothing$ 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 $l_h(\chi_{T+1}, \hat{\chi}) = 0 \quad \Leftrightarrow \quad \chi_{T+1} = \hat{\chi}$;
2. существует проекция $\chi_{proj} = \underset{\chi \in \mathcal{A} \cap \mathcal{B}}{\arg\min}\, l_h(\chi, \hat{\chi})$;
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 $L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$.

Оптимизационная задача

Теорема 2: цена игры (Стенина, 2014)

При выполнении требований теоремы 1 цена игры определена и равна
$V = \min_{\hat{\varphi}} \max_{\boldsymbol{\chi}_{T+1}} L(\hat{\varphi}, \chi_{T+1}) = \max_{\boldsymbol{\chi}_{T+1}} \min_{\hat{\varphi}} L(\hat{\varphi}, \chi_{T+1}) = -l_h(\chi_{proj}, \hat{\chi}) \leq 0.$

Теорема 3: согласованные прогнозы (Стенина, 2014)

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

$$\hat{\varphi} = \chi_{proj} = \arg\min_{\chi \in \mathcal{A} \cap \mathcal{B}} l_h(\chi, \hat{\chi})$$

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

Задача согласования прогнозов сводится к решению оптимизационной задачи.

Примеры функций потерь — дивергенции Брегмана

Удовлетворяют условию теоремы 1

- $l_h(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|^2$ — квадрат евклидового расстояния,
- $l_h(\mathbf{u}, \mathbf{v}) = \frac{1}{2}(\mathbf{u} - \mathbf{v})^T Q(\mathbf{u} - \mathbf{v})$ — квадрат расстояния Махаланобиуса,
- $l_h(\mathbf{u}, \mathbf{v}) = \sum\limits_{i=1}^{d} u_i \log \frac{u_i}{v_i} - \sum\limits_{i=1}^{d} u_i + \sum\limits_{i=1}^{d} v_i$ — обобщенная дивергенция Кульбака-Лейблера,
- $l_h(\mathbf{u}, \mathbf{v}) = \sum\limits_{i=1}^{d} \left(\frac{u_i}{v_i} - \log \frac{u_i}{v_i} - 1 \right)$ — расстояние Itakura-Saito,
- $l_h(\mathbf{u}, \mathbf{v}) = \sum\limits_{i=1}^{d} |u_i|^a - \sum\limits_{i=1}^{d} |v_i|^a - a \sum\limits_{i=1}^{d} sign(v_i)|v_i|^{a-1}(u_i - v_i)$, $a > 1$,
- $l_h(\mathbf{u}, \mathbf{v}) = \frac{2}{a^2} \sum\limits_{i=1}^{d} (e^{au_i} - e^{av_i}) - \frac{2}{a} \sum\limits_{i=1}^{d} e^{av_i}(u_i - v_i)$, $a \neq 0$.

Алгоритм теоретико-игрового оптимального согласования

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: $\hat{\varphi} = \arg\min\limits_{\chi \in \mathcal{A} \cap \mathcal{B}} l_h(\chi, \hat{\chi})$;

Свойства алгоритма, согласно теоремам 1, 2, 3.

- 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.

Модифицированный алгоритм теоретико-игрового оптимального согласования

- Решение оптимизационной задачи $\hat{\varphi} = \arg\min\limits_{\chi \in \mathcal{A} \cap \mathcal{B}} l_h(\chi, \hat{\chi})$ является вектором, наиболее близким к $\hat{\chi}$ в терминах функции потерь l_h, удовлетворяющим условию согласованности и физическим ограничениям.

- Прогнозы некоторых временных рядов могут быть более точными, чем прогнозы других. При согласовании более точные прогнозы должны подвергаться коррекции в меньшей степени.

- Предлагается ввести функцию потерь согласования $l_r(\chi, \hat{\chi})$, штрафующую разницу между согласованным $\hat{y}(i)$ и независимым $\hat{x}(i)$ прогнозом временного ряда тем сильнее, чем более точен независимый прогноз $\hat{x}(i)$.

Анализ свойств алгоритма теоретико-игрового оптимального согласования

Теорема 4 (Стенина, 2015)

Если l_r **удовлетворяет условию теоремы 1**, то

$$\hat{\varphi} = \arg\min_{\chi \in \mathcal{A} \cap \mathcal{B}} l_r(\chi, \hat{\chi})$$

1. удовлетворяет условию согласованности;
2. удовлетворяет физическим ограничениям;
3. обеспечивает $l_r(\chi_{T+1}, \hat{\varphi}) \leq l_r(\chi_{T+1}, \hat{\chi})$;
4. существуют веса для l_r, такие что
   $l_h(\chi_{T+1}, \hat{\varphi}) \leq l_h(\chi_{T+1}, \hat{\chi})$;
5. не требуется, чтобы l_h удовлетворяла условию теоремы 1.

Варианты выбора весов для l_r**:**

- обратно пропорциональные погрешностям независимых прогнозов для каждого временного ряда;
- с помощью кросс-валидации.

Сравнительнй анализ алгоритмов согласования

Теорема 5 (Стенина, 2015)

Пусть $l_r(\boldsymbol{\chi}, \hat{\boldsymbol{\chi}}) = w_{(:,:)}\big(\chi(:,:) - \hat{x}(:,:)\big)^2 +$
$+ w_{(1,:)}\big(\chi(1,:) - \hat{x}(1,:)\big)^2 + \ldots + w_{(n,:)}\big(\chi(n,:) - \hat{x}(n,:)\big)^2 +$
$+ w_{(:,1)}\big(\chi(:,1) - \hat{x}(:,1)\big)^2 + \ldots + w_{(:,m)}\big(\chi(:,m) - \hat{x}(:,m)\big)^2 +$
$+ w_{(1,1)}\big(\chi(1,1) - \hat{x}(1,1)\big)^2 + \ldots + w_{(n,m)}\big(\chi(n,m) - \hat{x}(n,m)\big)^2$ —
квадрат взвешенного евклидового расстояния, причем
$w_{(:,:)} = w_{(1,:)} = \ldots = w_{(n,:)} = w_{(:,1)} = \ldots = w_{(:,m)} = 1$,
$w_{(1,1)} = \ldots = w_{(n,m)}$.

Тогда при увеличении весов нижнего уровня $w_{(1,1)}, \ldots, w_{(n,m)} \to +\infty$ вектор согласованных прогнозов $\hat{\boldsymbol{\varphi}} = \underset{\boldsymbol{\chi} \in \mathcal{A} \cap \mathcal{B}}{\arg\min}\, l_r(\boldsymbol{\chi}, \hat{\boldsymbol{\chi}})$ приближается к вектору прогнозов, полученному алгоритмом восходящего согласования[1].

[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.

Цель эксперимента и данные

Цель

Сравнить качество прогнозов, полученных предложенными алгоритмами согласования, с качеством независимых прогнозов и согласованных прогнозов, полученных при помощи существующих алгоритмов согласования, для различных типов иерархических структур.

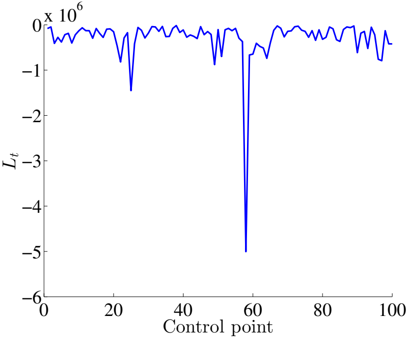
Данные

- **Трехуровневая иерархия:** данные о посуточной загруженности узлов РЖД. 37 типов грузов, 98 ЖД веток.
- **Двухуровневая иерархия:** данные о почасовом потреблении электроэнергии в 20 регионах Канады (Global Energy Forecasting Competition 2012).

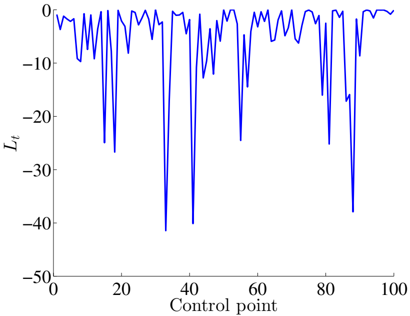
Сравнение качества согласованных и независимых прогнозов

Для согласования прогнозов $H = 100$ последних точек истории решалась оптимизационная задача $\hat{\varphi} = \underset{\boldsymbol{\chi} \in \mathcal{A} \cap \mathcal{B}}{\arg\min} \|\boldsymbol{\chi} - \hat{\boldsymbol{\chi}}\|_2^2$.

Изображена величина
$L_t = \|\boldsymbol{\chi}_t - \hat{\varphi}\|_2^2 - \|\boldsymbol{\chi}_t - \hat{\boldsymbol{\chi}}\|_2^2, \quad t = (T - H + 1), \ldots, T.$



Для РЖД



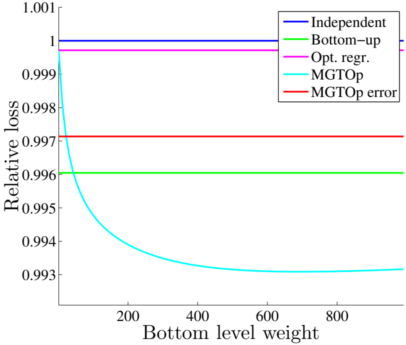
Для электроэнергии

Во всех контрольных точках потери уменьшились.

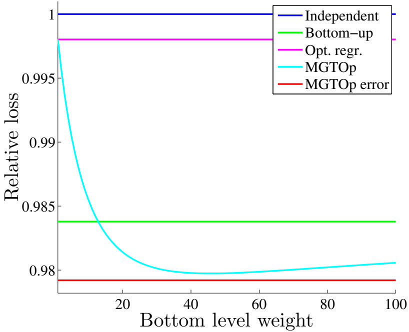
Зависимость функции потерь от значения весов согласования

$$\text{Relative loss}\,(\text{algorithm}) = \frac{\sum\limits_{t=T-H+1}^{T} \|\boldsymbol{\chi}_t - \hat{\boldsymbol{\varphi}}_{\text{algorithm}}\|_2^2}{\sum\limits_{t=T-H+1}^{T} \|\boldsymbol{\chi}_t - \hat{\boldsymbol{\chi}}\|_2^2}.$$

$$\hat{\boldsymbol{\varphi}}_{\text{теор.игр.}} = \mathop{\arg\min}_{\boldsymbol{\chi} \in \mathcal{A} \cap \mathcal{B}} l_r(\boldsymbol{\chi}, \hat{\boldsymbol{\chi}}), \qquad l_r(\boldsymbol{\chi}, \hat{\boldsymbol{\chi}}) = \sum_{i=1}^{d} w_i \big(\chi(i) - \hat{x}(i)\big)^2.$$



Для РЖД



Для электроэнергии

Сравнение качества согласованных и независимых прогнозов на разных уровнях иерархии

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

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

Уровень иерархии	Независимые прогнозы	Восходящее согласование[2]	Оптимальная регрессия[3]	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

Средние потери прогнозирования потребления электроэнергии

Уровень иерархии	Независимые прогнозы	Восходящее согласование	Оптимальная регрессия	Модиф. теор.-игр. согл. (погрешности)
Вся иерархия	2727	2683	2722	2670
Верхний уровень	2083	2039	2076	2029
Нижний уровень	644	644	646	642

[2] 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.

[3] 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.

Выносится на защиту

1. Предложен алгоритм теоретико-игрового оптимального согласования прогнозов иерархических временных рядов, который сводит задачу согласования к задаче оптимизации.
2. Доказано, что алгоритм не требует несмещенности независимых прогнозов и оценок их погрешностей. На независимые прогнозы накладываются только физические ограничения.
3. Показано, что алгоритм позволяет работать с иерархическими структурами любой сложности.
4. Результаты экспериментов подтверждают, что качество прогнозов, согласованных с помощью предложенного алгоритма и его модификации, превосходит качество независимых прогнозов, а также качество согласованных прогнозов, полученных с помощью существующих алгоритмов согласования.

Публикации

1. **Stenina M. M., Kuznetsov M. P., Strijov V. V. Ordinal classification using Pareto fronts // Expert Systems with Applications. 2015. No 42(14). P. 5947–5953.**
2. **Стенина М. М., Стрижов В. В. Согласование прогнозов при решении задач прогнозирования иерархических временных рядов // Информатика и ее применения. 2015. №9(2). С. 77–89.**
3. **Газизулина Р. К., Стенина М. М., Стрижов В. В. Прогнозирование объемов железнодорожных перевозок по парам веток // Системы и средства информатики. 2015. №25(1). С. 142–155.**
4. **Стенина М. М., Стрижов В. В. Согласование агрегированных и детализированных прогнозов при решении задач непараметрического прогнозирования // Системы и средства информатики. 2014. №24(2). С. 21–34.**
5. **Медведникова М. М., Стрижов В. В. Построение интегрального индикатора качества научных публикаций методами ко-кластеризации // Известия ТулГУ. Естественные науки. 2013. №1.**
6. **Кузнецов М. П., Стрижов В. В., Медведникова М. М. Алгоритм многоклассовой классификации объектов, описанных в ранговых шкалах // НТВСПбГПУ. 2012. №5. С. 92–94.**
7. **Медведникова М. М. Стрижов, В. В., Кузнецов М. П. Алгоритм многоклассовой монотонной Парето-классификации с выбором признаков // Известия ТулГУ. 2012. №3. С. 132–141.**
8. Вальков А. С., Кожанов Е. М., Медведникова М. М., Хусаинов Ф. И. Непараметрическое прогнозирование загруженности системы железнодорожных узлов по историческим данным // Машинное обучение и анализ данных. 2012. №4. С. 448–465.