



Общероссийский математический портал

М. М. Стенина, В. В. Стрижов, Согласование прогнозов при решении задач прогнозирования иерархических временных рядов, *Информ. и её примен.*, 2015, том 9, выпуск 2, 75–87

DOI: 10.14357/19922264150209

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.97.14.81

23 января 2025 г., 16:46:35



СОГЛАСОВАНИЕ ПРОГНОЗОВ ПРИ РЕШЕНИИ ЗАДАЧ ПРОГНОЗИРОВАНИЯ ИЕРАРХИЧЕСКИХ ВРЕМЕННЫХ РЯДОВ*

М. М. Стенина¹, В. В. Стрижов²

Аннотация: Рассматривается задача одновременного прогнозирования набора временных рядов, объединенных в иерархическую многоуровневую структуру. Требуется, чтобы полученные прогнозы удовлетворяли физическим ограничениям и структуре иерархии. Предложен алгоритм согласования прогнозов иерархических временных рядов GTOp (Game-theoretically optimal reconciliation), гарантирующий неухудшение качества прогнозов после проведения процедуры согласования по сравнению с качеством прогнозов, полученных для каждого временного ряда независимо. Подход базируется на поиске равновесия Нэша в антагонистической игре заданного вида и сводит задачу согласования прогнозов к задаче оптимизации с ограничениями типа равенства и неравенства. Доказывается, что при выполнении ряда общих предположений о свойствах структуры иерархии, физических ограничений и функции потерь в игре существует равновесие Нэша в чистых стратегиях. Работа алгоритма демонстрируется на разных типах иерархических структур с использованием данных посуточной загруженности железнодорожных узлов.

Ключевые слова: иерархические временные ряды; согласование прогнозов временных рядов; антагонистическая игра; равновесие Нэша

DOI: 10.14357/19922264150209

1 Введение

Рассматривается задача одновременного прогнозирования набора временных рядов, связанных в иерархическую многоуровневую структуру, в которой временные ряды каждого следующего (более высокого) уровня формируются путем поэлементного суммирования некоторой части (возможно, всех) временных рядов предыдущего уровня. Используемые при решении этой задачи принцип равновесия и регрессионные методы широко обсуждаются в научной литературе, в том числе на страницах журнала «Информатика и её применения» [1, 2].

Задача прогнозирования иерархических временных рядов возникает в различных прикладных областях. В [3] описан конкурс, проведенный в 2012 г. на Kaggle [4], в котором одной из задач было прогнозирование иерархических временных рядов с требованием согласования прогнозов по иерархии. В работе [5] обсуждается задача прогнозирования туристической активности по регионам и целям поездов. В статье [6] решается задача прогнозирования потребительского спроса на различные группы товаров в ряде магазинов. В [7]

решается задача согласования прогнозов объемов железнодорожных перевозок различных типов грузов по ряду железнодорожных веток.

В связи с высоким интересом к задаче разными авторами предлагаются различные подходы к ее решению. Как правило, сперва получают каким-либо образом прогнозы всех (или некоторой части) временных рядов независимо друг от друга, а затем корректируют (согласуют) эти прогнозы.

Самыми простыми и самыми распространенными способами согласования являются нисходящий (top-down) и восходящий (bottom-up) подходы [8–15]. Нисходящий подход предполагает получение прогноза на верхнем уровне иерархии (агрегированный временной ряд), а затем деагрегирование этого прогноза на следующий (более низкий) уровень иерархии на основании долей, наблюдаемых в истории. Восходящий подход использует прогнозы временных рядов нижнего уровня иерархии (неагрегированных), из которых получает прогнозы рядов из верхних уровней путем агрегирования. Также встречаются подходы, комбинирующие нисходящий и восходящий.

Нет единой точки зрения на то, какой из этих подходов позволяет получать более точные про-

* Работа выполнена при финансовой поддержке РФФИ (проект 13-07-13139).

¹Московский физико-технический институт, Национальный исследовательский университет «Высшая школа экономики», mmedvednikova@gmail.com

²Вычислительный центр Российской академии наук им. А. А. Дородницына, strijov@ccas.com

гнозы. Наиболее ранние исследования проведены в работе [8], где авторы считают, что агрегированные данные содержат много ошибок и поэтому нисходящее прогнозирование дает более точные прогнозы. К таким же выводам приходят авторы работ [12, 13]. В [15] также утверждается, что агрегированные прогнозы более точны. С другой стороны, в [9, 10] обсуждается, что основные потери информации происходят при агрегировании и поэтому восходящий подход предпочтительнее. В [11] сравниваются оба подхода к согласованию прогнозов и утверждается, что восходящий предпочтительнее при выполнении некоторых условий на структуру иерархии и горизонт прогноза. В [14] исследуется смещение и устойчивость прогнозов, получаемых с помощью обоих подходов, и заключается, что восходящий надежнее, за исключением случаев с пропусками значений и выбросами на нижних уровнях иерархии.

Авторы статьи [5] обобщают нисходящий и восходящий подходы к согласованию иерархических прогнозов и предлагают оптимальное согласование с использованием регрессии, позволяющей согласовывать одновременно по всем уровням иерархии любой сложности. Однако предложенный ими алгоритм согласования предполагает, что полученные независимые прогнозы являются несмещенными оценками и что ошибки прогнозов временных рядов удовлетворяют структуре иерархии, т. е. являются согласованными. Оба эти требования являются достаточно строгими и сильно ограничивают применимость этого способа согласования.

Предложенный в [7] способ согласования прогнозов не требует несмещенности независимых прогнозов и согласованности ошибок. При этом, как продемонстрировано в статье, он не уступает методу из [5] по качеству согласованных прогнозов. Однако и этот метод имеет ряд недостатков. Для его использования необходимо оценивать погрешность независимых прогнозов, что не всегда удается корректно сделать. Метод не гарантирует, что качество согласованных прогнозов не будет уступать качеству независимых. И согласование происходит поэтапно по узлам иерархии, что не позволяет учесть сразу всю информацию о взаимосвязи между временными рядами.

В настоящей статье предлагается обобщение алгоритма согласования из [7] с сохранением его преимуществ и устранением недостатков. Предлагается алгоритм согласования прогнозов ГТОр, основанный на идеях из [16]. Алгоритм ГТОр не требует оценки погрешности независимых прогнозов временных рядов, не требует несмещенности независимых прогнозов и имеет теоретическое обоснование улучшения качества прогнозов

после проведения согласования. Задача согласования прогнозов рассматривается как поиск равновесия Нэша [17, 18] в антагонистической игре игрока, выбирающего согласованные прогнозы, с природой, которая выбирает действительные значения временных рядов, и сводится к решению оптимизационной задачи с ограничениями типа равенства и неравенства. Вид равновесия Нэша задает параметры оптимизационной задачи.

Работа предлагаемого алгоритма демонстрируется на данных о посуточном отправлении вагонов с 37 типами грузов с 98 веток Российских железных дорог (РЖД). Проведены эксперименты по согласованию прогнозов для различных типов иерархических структур и показано, что на практике действительно наблюдается улучшение качества прогнозов после проведения согласования алгоритмом ГТОр.

Статья включает следующие разделы. В разд. 2 вводятся необходимые обозначения, в разд. 3 формулируется задача согласования прогнозов иерархических временных рядов. Раздел 4 содержит необходимые определения и факты из теории игр, описание антагонистической игры, соответствующей задаче согласования прогнозов, и доказательство существования в этой игре равновесия Нэша. Раздел 5 описывает оптимизационную задачу, к которой сводится задача согласования прогнозов, а также преимущества и недостатки алгоритма. В разд. 6 приводятся примеры функций потерь, для которых применим алгоритм ГТОр. В разд. 7 приведены результаты экспериментов. В разд. 8 подводятся итоги и делаются выводы.

2 Обозначения

В этом разделе вводится система обозначений, которая будет использоваться в настоящей работе. Будем обозначать временной ряд через вектор \mathbf{x} , элементы временного ряда будем снабжать индексом t , $t = 1, \dots, T$, где T — длина истории.

$$\mathbf{x} = \{x_t\}_{t=1}^T.$$

Общее количество временных рядов во всей иерархии будем обозначать d . Далее для наглядности будем рассматривать иерархию рядов, изображенную на рис. 1. Она содержит один временной ряд на верхнем уровне и n рядов на нижнем. Для этой иерархии $d = 1 + n$.

Временные ряды нижнего уровня будут обозначаться $\mathbf{x}(i)$, $i = 1, \dots, n$, где n — число временных рядов на нижнем уровне. Временной ряд верхнего уровня обозначается как $\mathbf{x}(\cdot)$. Во избежание

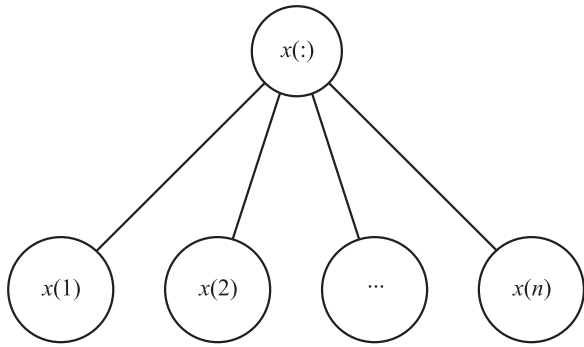


Рис. 1 Плоская двухуровневая иерархия

путаницы условимся использовать нижние индексы для обозначения отсчетов времени, а индексы в скобках использовать для обозначения положения временных рядов в структуре иерархии. Элементы временных рядов, составляющих плоскую двухуровневую иерархию, обозначаются соответственно $x_t(i)$, $i = 1, \dots, n$, $x_t(:)$. Их соотношение задается формулой

$$x_t(:) = \sum_{i=1}^n x_t(i), \quad t = 1, \dots, T. \quad (1)$$

Будем называть соотношение (1) **условием согласованности**. Прогнозы этих временных рядов будем обозначать «шляпками», опуская нижние индексы, чтобы избежать излишне громоздких обозначений. Прогнозироваться будет всегда $(T + 1)$ -е значение временного ряда $\hat{x}(i)$, $i = 1, \dots, n$, $\hat{x}(:)$. Согласованные прогнозы будут также обозначаться без нижних индексов: $\hat{y}(i)$, $i = 1, \dots, n$, $\hat{y}(:) = \sum_{i=1}^n \hat{y}(i)$.

Запишем все временные ряды в матрицу, каждая строка которой соответствует одному временному ряду. Для иерархии с рис. 1 эта матрица будет размера $(1 + n) \times T$ и выглядеть следующим образом:

$$X = \begin{pmatrix} x_1(:) & x_2(:) & \dots & x_T(:) \\ x_1(1) & x_2(1) & \dots & x_T(1) \\ \dots & \dots & \dots & \dots \\ x_1(n) & x_2(n) & \dots & x_T(n) \end{pmatrix}.$$

Будем называть срезом иерархии в момент времени t столбец матрицы X , соответствующий моменту времени t . Для удобства записи введем векторы, соответствующие срезу иерархии в момент времени t , прогнозам и согласованным прогнозам. В этих векторах значения, соответствующие разным временным рядам, записаны в столбец, начиная с верхнего уровня иерархии и заканчивая нижним уровнем:

$$\chi_t = \begin{pmatrix} x_t(:) \\ x_t(1) \\ \vdots \\ x_t(n) \end{pmatrix}; \quad \hat{\chi}_t = \begin{pmatrix} \hat{x}(:) \\ \hat{x}(1) \\ \vdots \\ \hat{x}(n) \end{pmatrix}; \quad \hat{\varphi}_t = \begin{pmatrix} \hat{y}(:) \\ \hat{y}(1) \\ \vdots \\ \hat{y}(n) \end{pmatrix}.$$

Условие (1) для векторов χ_t и $\hat{\varphi}_t$ запишем, введя матрицу связей размером $1 \times (n + 1)$:

$$S = (-1 \ 1 \ \dots \ 1).$$

Тогда условие согласованности запишется кратко

$$S\chi_t = 0; \quad S\hat{\varphi}_t = 0.$$

В случае, когда иерархия имеет более сложную структуру, чем на рис. 1, векторы χ_t , $\hat{\chi}_t$ и $\hat{\varphi}_t$ имеют размерность d , матрица X имеет ровно d строк, временные ряды в ней записываются от верхних уровней к нижним. А размерность матрицы связи S равна $c \times d$, где c — число узлов в графе иерархии или, другими словами, количество ограничений-равенств, наложенных на элементы срезов иерархии χ_t .

3 Задача согласования прогнозов иерархических временных рядов

Сформулируем задачу согласования прогнозов иерархических временных рядов. Для этого введем еще ряд необходимых обозначений.

Пусть дан набор из d временных рядов, значения которых записаны в матрицу X размера $d \times T$:

$$X = (\chi_1 \ \chi_2 \ \dots \ \chi_T),$$

где каждый столбец χ_t соответствует срезу в момент времени t , а каждая строка x_i — одному временному ряду. Пусть структура иерархии задана матрицей связи S так, что для всех $t = 1, \dots, T$ выполнено условие согласованности:

$$S\chi_t = 0.$$

Пусть даны прогнозы значений $\hat{\chi}_t$ для всех временных рядов в момент времени $T + 1$ и задана функция суммарных потерь при прогнозировании иерархии

$$l_h(\hat{\chi}_T, \chi_{T+1}). \quad (2)$$

Определим множество

$$\mathcal{A} = \{\chi \in \mathbb{R}^d \mid S\chi = 0\}, \quad (3)$$

где χ — произвольный d -мерный вектор, а S — заданная матрица связи. Отметим, что все срезы χ_t , $t = 1, \dots, T$, лежат в множестве \mathcal{A} . Также в нем должны лежать согласованные прогнозы $\hat{\varphi}_t$.

В ряде задач прогнозы должны удовлетворять некоторым ограничениям, связанным с физической природой прогнозируемой величины. В связи с чем введем множество

$$\mathcal{B} = \{\chi \in \mathbb{R}^d \mid \chi(i) \in [A_i, B_i], \\ A_i, B_i \in [-\infty, +\infty], i = 1, \dots, d\}, \quad (4)$$

где χ — произвольный d -мерный вектор; A_i и B_i задают отрезок, в котором должна находиться i -я компонента этого вектора. Например, в случае $A_1 = \dots = A_d = 0, B_1 = \dots = B_d = +\infty$ вектор χ лежит в положительном октанте, а конечные значения $A_1, \dots, A_d, B_1, \dots, B_d$ могут задавать интервалы, в которых должны лежать прогнозы. Отсутствие каких-либо ограничений задается значениями $A_1 = \dots = A_d = -\infty, B_1 = \dots = B_d = +\infty$.

Введя все необходимые дополнительные обозначения, можно сформулировать задачу поиска согласованных прогнозов. Требуется найти вектор прогнозов $\hat{\varphi}$, удовлетворяющий следующим требованиям.

Согласованность: вектор прогнозов $\hat{\varphi}$ должен удовлетворять структуре иерархии, заданной матрицей связи S , т. е. $\hat{\varphi} \in \mathcal{A}$.

Ограничения: вектор прогнозов $\hat{\varphi}$ должен удовлетворять наложенным ограничениям, т. е. $\hat{\varphi} \in \mathcal{B}$.

Качество: общие потери при использовании согласованных прогнозов не должны превышать общие потери при использовании независимых прогнозов, т. е. $l_h(\hat{\varphi}, \chi_{T+1}) \leq l_h(\hat{\chi}, \chi_{T+1})$.

4 Задача согласования прогнозов как поиск равновесия в антагонистической игре

В этом разделе задача согласования прогнозов, сформулированная выше, рассматривается как антагонистическая игра. Такое представление не влияет на решение задачи согласования прогнозов и направлено лишь на достижение наглядности и интерпретируемости полученных результатов. В первой части раздела приводятся необходимые определения и факты из теории игр [17–19], во второй части вводится антагонистическая игра, соответствующая задаче согласования прогнозов, в третьей части формулируется и доказывается теорема о существовании в этой игре равновесия Нэша, которая является обобщением теоремы, доказанной в [16], и приводится следствие из этой теоремы, однозначно определяющее выбор оптимального вектора согласованных прогнозов.

4.1 Понятие антагонистической игры

Определение 1. Система

$$\Gamma = (\mathcal{M}, \mathcal{N}, L),$$

где \mathcal{M} и \mathcal{N} — непустые множества и функция $L: \mathcal{M} \times \mathcal{N} \rightarrow \mathbb{R}$, называется антагонистической игрой (игрой с нулевой суммой) в нормальной форме. Элементы $\mu \in \mathcal{M}$ и $\nu \in \mathcal{N}$ называются стратегиями игроков 1 и 2 соответственно, L — функцией потерь игрока 1. Потери игрока 2 полагаются равными $-L(\mu, \nu)$.

Определение 2. Говорят, что игра разыгрывается в чистых стратегиях, если оба игрока из имеющихся наборов действий \mathcal{M} и \mathcal{N} выбирают по одному действию μ и ν соответственно.

Определение 3. Введем на множествах стратегий \mathcal{M} и \mathcal{N} вероятностные распределения $p(\mu)$ и $q(\nu)$ соответственно:

$$\int_{\mathcal{M}} p(\mu) d\mu = 1; \quad \int_{\mathcal{N}} q(\nu) d\nu = 1.$$

Распределения $p(\mu)$ и $q(\nu)$ задают смешанные стратегии в игре Γ , если игрок 1 выбирает действие в соответствии с распределением $p(\mu)$ и игрок 2 выбирает действие в соответствии с распределением $q(\nu)$.

Чистые стратегии являются частным случаем смешанных. Поэтому далее будут рассматриваться смешанные стратегии, за исключением специально оговоренных моментов. Обозначать стратегии будем $p(\cdot)$ и $q(\cdot)$.

Математическое ожидание потерь игрока 1 при паре смешанных стратегий $p(\cdot)$ и $q(\cdot)$ обозначим через

$$\bar{L}(p(\cdot), q(\cdot)) = \int_{\mathcal{M}} \int_{\mathcal{N}} p(\mu)q(\nu)L(\mu, \nu) d\mu d\nu.$$

Игрок 1 преследует цель минимизировать эту величину при любых действиях игрока 2.

Определение 4. Пара стратегий $(p(\cdot), q(\cdot))$ называется равновесием Нэша в смешанных стратегиях в игре Γ , если для любых $p(\cdot)'$ и $q(\cdot)'$ выполнено неравенство:

$$\bar{L}(p(\cdot), q(\cdot)') \leq \bar{L}(p(\cdot), q(\cdot)) \leq \bar{L}(p(\cdot)', q(\cdot)).$$

При равновесии Нэша ни одному из игроков не выгодно отклоняться от равновесной стратегии, если второй продолжает придерживаться равновесной стратегии. При этом игрок 1 минимизирует свои потери в ситуации, когда игрок 2 действует наиболее выгодным для себя образом. Отметим

также, что равновесие Нэша является седловой точкой функции $\bar{L}(p(\cdot), q(\cdot))$.

Теорема 1. *В антагонистической игре равновесие Нэша существует тогда и только тогда, когда определена величина*

$$V = \min_{p(\cdot)'} \max_{q(\cdot)'} \bar{L}(p(\cdot)', q(\cdot)') = \max_{q(\cdot)'} \min_{p(\cdot)'} \bar{L}(p(\cdot)', q(\cdot)').$$

Величина V называется ценой игры. Доказательство этой теоремы в настоящей статье не приводится, при желании его можно найти в [17–19].

4.2 Антагонистическая игра, описывающая задачу согласования прогнозов

Вернемся к рассмотрению введенных в разд. 3 множеств \mathcal{A} (3) и \mathcal{B} (4). Напомним, что множество \mathcal{A} содержит все d -мерные векторы, удовлетворяющие структуре иерархии, заданной матрицей связи S . Множество \mathcal{B} содержит d -мерные векторы, удовлетворяющие ограничениям-неравенствам рассматриваемой задачи прогнозирования.

Будем рассматривать антагонистическую игру Γ , в которой игрок 1 выбирает вектор согласованных прогнозов $\hat{\varphi}$ из множества $\mathcal{A} \cap \mathcal{B}$, которые одновременно удовлетворяют структуре иерархии, заданной матрицей связи S , и ограничениям-неравенствам, задающим множество \mathcal{B} , игрок 2 также выбирает вектор действительных значений χ_{T+1} , удовлетворяющих структуре иерархии и физическим ограничениям (можно считать игрока 2 природой).

Определим множества стратегий \mathcal{M} и \mathcal{N} игроков 1 и 2 как пересечение множеств $\mathcal{A} \cap \mathcal{B}$. Функцию потерь игрока 1 определим с помощью функции потерь по иерархии (2):

$$L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi}),$$

где вектор независимых прогнозов $\hat{\chi}$ считается заданным и не зависит от действий, выбираемых игроками. Считается, что $\hat{\chi} \in \mathcal{B}$. Такой выбор функции потерь игрока 1 связан с тем соображением, что при $L(\hat{\varphi}, \chi_{T+1}) = 0$ качество согласованных прогнозов не хуже, чем качество независимых прогнозов, а при $L(\hat{\varphi}, \chi_{T+1}) < 0$ и вовсе превосходит его.

Таким образом, получена антагонистическая игра

$$\Gamma = (\mathcal{A} \cap \mathcal{B}, \mathcal{A} \cap \mathcal{B}, l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})), \quad (5)$$

где игрок 1 выбирает вектор согласованных прогнозов $\hat{\varphi}$ из множества $\mathcal{A} \cap \mathcal{B}$, а игрок 2 выбирает вектор

действительных значений элементов временных рядов χ_{T+1} из множества $\mathcal{A} \cap \mathcal{B}$. При этом первый игрок преследует цель минимизировать свои потери, выраженные функцией $L(\hat{\varphi}, \chi_{T+1})$, при любых действиях игрока 2, т. е. при любом векторе действительных значений χ_{T+1} . Эта цель достигается, если игрок 1 воспользуется стратегией, входящей в равновесие Нэша.

4.3 Существование равновесия Нэша

В этой части раздела будет показано, что при выполнении ряда естественных требований к множествам \mathcal{A} (3) и \mathcal{B} (4) и функции суммарных потерь при прогнозировании иерархии l_h (2) в антагонистической игре (5), описывающей задачу согласования прогнозов, существует равновесие Нэша в чистых стратегиях. Также будет показано, что соответствующее этому равновесию значение функции потерь игрока 1 неположительно, что гарантирует неухудшение качества прогнозов при переходе от независимых прогнозов к согласованным. Рассмотрим эти требования.

Определение 5. Множество $\mathcal{C} \subseteq \mathbb{R}^d$ называется выпуклым [20], если для любых $\chi_1 \in \mathcal{C}$ и $\chi_2 \in \mathcal{C}$ и любого $0 \leq \alpha \leq 1$ выполнено

$$\alpha\chi_1 + (1 - \alpha)\chi_2 \in \mathcal{C}.$$

Заметим, что множества \mathcal{A} (3) и \mathcal{B} (4) выпуклы и замкнуты.

Предположение 1. *Будем предполагать, что пересечение множеств \mathcal{A} и \mathcal{B} не пусто: $\mathcal{A} \cap \mathcal{B} \neq \emptyset$.*

Требование непустого пересечения этих множеств естественно, так как в противном случае неразрешима задача поиска вектора согласованных прогнозов $\hat{\varphi}$, который должен одновременно принадлежать обоим множествам. Также отметим, что множество $\mathcal{A} \cap \mathcal{B}$ является выпуклым и замкнутым как пересечение двух выпуклых и замкнутых множеств [20].

Предположение 2. *Будем считать, что функция суммарных потерь l_h (2) неотрицательна и равна нулю только при равенстве аргументов:*

$$l_h(\chi_{T+1}, \hat{\chi}) \geq 0 \text{ для всех } \chi_{T+1}, \hat{\chi};$$

$$l_h(\chi_{T+1}, \hat{\chi}) = 0 \Leftrightarrow \chi_{T+1} = \hat{\chi}.$$

Равенство аргументов $\chi_{T+1} = \hat{\chi}$ соответствует случаю, когда прогноз полностью совпадает с действительными значениями. В этом случае потери равны нулю. Во всех остальных случаях потери при прогнозе положительные.

Определение 6. Проекцией точки $\chi_0 \in \mathbb{R}^d$ на множество $C \subseteq \mathbb{R}^d$, инициированной функцией расстояния f , называется точка

$$\chi_{\text{proj}} = \arg \min_{\chi \in C} f(\chi, \chi_0).$$

Предположение 3. Пусть существует проекция точки из \mathbb{R}^d , соответствующей вектору независимых прогнозов $\hat{\chi}$, на выпуклое и замкнутое множество $A \cap B$, инициированная функцией суммарных потерь l_h :

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

Предположение 4. Пусть $\chi_{\text{proj}} = \arg \min_{\chi \in A \cap B} l_h(\chi, \hat{\chi})$.

Будем предполагать, что для всех $\chi \in B$ и для всех $\psi \in A \cap B$ выполняется неравенство

$$l_h(\psi, \chi) \geq l_h(\psi, \chi_{\text{proj}}) + l_h(\chi_{\text{proj}}, \chi).$$

Для пояснения этого требования рассмотрим частный случай, когда $\mathbb{R}^d = \mathbb{R}^3$, $A \subset \mathbb{R}^2$, $B \subset \mathbb{R}^3$, $A \cap B \subset \mathbb{R}^2$ и l_h — квадрат метрики Евклида (рис. 2). Точки ψ , χ , χ_{proj} образуют треугольник. Обозначим θ угол при вершине χ_{proj} и запишем теорему косинусов:

$$l_h(\psi, \chi) = l_h(\psi, \chi_{\text{proj}}) + l_h(\chi_{\text{proj}}, \chi) - 2\sqrt{l_h(\psi, \chi_{\text{proj}})l_h(\chi_{\text{proj}}, \chi)} \cos \theta.$$

Поскольку χ_{proj} является проекцией, то угол θ не может быть острым. Он прямой, если проецируемая точка χ находится «над» множеством $A \cap B$, и тупой, если точка находится «в стороне». Таким образом получаем, что $\cos \theta \leq 0$, а значит, последнее слагаемое в теореме косинусов неотрицательное. Исключая его и заменяя знак равенства на знак нестрогого неравенства, получаем, что предположение 4 соответствует естественным свойствам проекции.

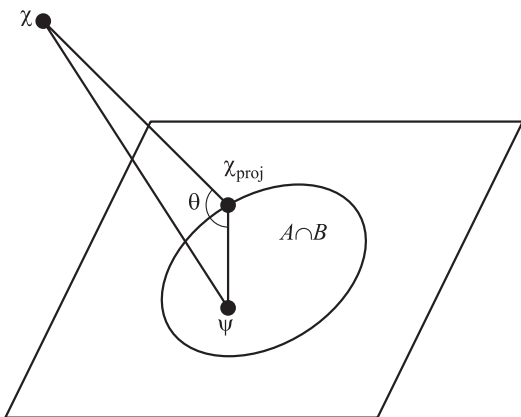


Рис. 2 Пояснение к предположению 4

Введя предположения 1–4, сформулируем теорему.

Теорема 2. Пусть выполнены предположения 1–4. Тогда пара стратегий $(\chi_{\text{proj}}, \chi_{\text{proj}})$ является равновесием Нэша в игре (5) и седловой точкой функции $L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$. Цена игры при этом равна $V = -l_h(\chi_{\text{proj}}, \hat{\chi})$.

Доказательство. Найдем седловую точку функции $L(\hat{\varphi}, \chi_{T+1})$ в соответствии с определением 4 и теоремой 1. Найдем максимум этой функции по второму аргументу при $\hat{\varphi} = \chi_{\text{proj}}$. Для этого воспользуемся предположением 4:

$$l_h(\psi, \chi) \geq l_h(\psi, \chi_{\text{proj}}) + l_h(\chi_{\text{proj}}, \chi).$$

Применяя неравенство к функции потерь L игрока 1 (подставляем $\psi = \chi_{T+1}$, $\chi = \hat{\chi}$), получаем:

$$\begin{aligned} L(\chi_{\text{proj}}, \chi_{T+1}) &= l_h(\chi_{T+1}, \chi_{\text{proj}}) - l_h(\chi_{T+1}, \hat{\chi}) \leq \\ &\leq l_h(\chi_{T+1}, \hat{\chi}) - l_h(\chi_{\text{proj}}, \hat{\chi}) - l_h(\chi_{T+1}, \hat{\chi}) = \\ &= -l_h(\chi_{\text{proj}}, \hat{\chi}) \end{aligned}$$

для всех $\chi_{T+1} \in A \cap B$. Заметим также, что из предположения 2 вытекает

$$L(\chi_{\text{proj}}, \chi_{\text{proj}}) = l_h(\chi_{\text{proj}}, \chi_{\text{proj}}) - l_h(\chi_{\text{proj}}, \hat{\chi}) = -l_h(\chi_{\text{proj}}, \hat{\chi}).$$

Приходим к выводу, что

$$L(\chi_{\text{proj}}, \chi_{T+1}) \leq L(\chi_{\text{proj}}, \chi_{\text{proj}})$$

для всех $\chi_{T+1} \in A \cap B$.

Следовательно, максимум по второму аргументу достигается при $\chi_{T+1} = \chi_{\text{proj}}$.

Минимум по первому аргументу при $\chi_{T+1} = \chi_{\text{proj}}$ находим, используя предположение 2, из соотношения

$$\begin{aligned} \arg \min_{\hat{\varphi} \in A \cap B} L(\hat{\varphi}, \chi_{\text{proj}}) &= \\ &= \arg \min_{\hat{\varphi} \in A \cap B} l_h(\chi_{\text{proj}}, \hat{\varphi}) - l_h(\chi_{\text{proj}}, \hat{\chi}). \end{aligned}$$

Второе слагаемое не зависит от $\hat{\varphi}$, а по предположению 2 функция суммарных потерь неотрицательна и обращается в ноль только при равенстве аргументов, поэтому получаем

$$\arg \min_{\hat{\varphi} \in A \cap B} L(\hat{\varphi}, \chi_{\text{proj}}) = \chi_{\text{proj}}.$$

Таким образом, получаем, что

$$L(\chi_{\text{proj}}, \chi_{T+1}) \leq L(\chi_{\text{proj}}, \chi_{\text{proj}}) \leq L(\hat{\varphi}, \chi_{\text{proj}}), \quad \chi_{T+1}, \hat{\varphi} \in A \cap B.$$

Следовательно, $(\chi_{\text{proj}}, \chi_{\text{proj}})$ — седловая точка функции L . И эта пара является равновесием Нэша в игре (5), и цена игры выражается как

$$\begin{aligned} V &= \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} \max_{\chi_{T+1} \in \mathcal{A}} L(\hat{\varphi}, \chi_{T+1}) = \\ &= \max_{\chi_{T+1} \in \mathcal{A}} \min_{\hat{\varphi} \in \mathcal{A} \cap \mathcal{B}} L(\hat{\varphi}, \chi_{T+1}) = \\ &= L(\chi_{\text{proj}}, \chi_{\text{proj}}) = -l_h(\chi_{\text{proj}}, \hat{\chi}). \end{aligned}$$

Следствие 1. *Использование в качестве вектора согласованных прогнозов $\hat{\varphi}$ проекции вектора независимых прогнозов $\hat{\chi} \in \mathbb{R}^d$ на множество $\mathcal{A} \cap \mathcal{B}$, инициированной функцией суммарных потерь l_h , гарантирует значение функции суммарных потерь не большее, чем при использовании вектора независимых прогнозов $\hat{\chi}$. Доказательство.* По теореме 2 цена игры (5) равна значению функции потерь игрока 1 в точке $(\chi_{\text{proj}}, \chi_{\text{proj}})$ и неположительна в силу предположения 2

$$V = L(\chi_{\text{proj}}, \chi_{\text{proj}}) = -l_h(\chi_{\text{proj}}, \hat{\chi}) \leq 0.$$

А выбор рассматриваемой функции потерь игрока 1 $L(\hat{\varphi}, \chi_{T+1}) = l_h(\chi_{T+1}, \hat{\varphi}) - l_h(\chi_{T+1}, \hat{\chi})$ был обусловлен тем, что ее знак совпадает со знаком изменения суммарных потерь при переходе от вектора независимых прогнозов $\hat{\chi}$ к вектору согласованных прогнозов $\hat{\varphi}$. Следовательно, при $\chi_{T+1} = \chi_{\text{proj}}$ суммарные потери при согласованных прогнозах меньше, чем при независимых.

Согласно определению 4 равновесия Нэша

$$L(\chi_{\text{proj}}, \chi_{T+1}) \leq L(\chi_{\text{proj}}, \chi_{\text{proj}}) \leq 0$$

для любых $\chi_{T+1} \in \mathcal{A} \cap \mathcal{B}$. Поэтому при любом векторе действительных значений $\chi_{T+1} \in \mathcal{A} \cap \mathcal{B}$ согласованные прогнозы $\hat{\varphi}$ оказываются предпочтительнее независимых прогнозов $\hat{\chi}$.

5 Алгоритм согласования прогнозов ГТОр

Согласно следствию 1 оптимальным выбором вектора согласованных прогнозов $\hat{\varphi}$ является проекция вектора независимых прогнозов $\hat{\chi}$ на множество $\mathcal{A} \cap \mathcal{B}$, инициированная функцией суммарных потерь l_h . Множество $\mathcal{A} \cap \mathcal{B}$ содержит векторы размерности d , удовлетворяющие структуре иерархии, так как множество \mathcal{A} задается ограничениями-равенствами, порожденными матрицей связи иерархии S . В то же время $\mathcal{A} \cap \mathcal{B}$ содержит d -мерные векторы, удовлетворяющие ограничениям-неравенствам из множества \mathcal{B} . Таким образом, задача поиска проекции — это оптимизационная задача с ограничениями типа равенства и неравенства:

$$\begin{cases} l_h(\chi, \hat{\chi}) \rightarrow \min_{\chi}, \\ \chi \in \mathcal{A} \text{ (ограничения-равенства)}; \\ \chi \in \mathcal{B} \text{ (ограничения-неравенства)}. \end{cases} \quad (6)$$

Алгоритм согласования прогнозов иерархических временных рядов ГТОр заключается в решении оптимизационной задачи (6). Достоинства этого алгоритма заключаются в том, что он требует от вектора независимых прогнозов $\hat{\chi}$ лишь принадлежности множеству \mathcal{B} , и не требует несмещенности независимых прогнозов, а следовательно, для получения независимых прогнозов можно использовать любой алгоритм прогнозирования. Также ГТОр не требует оценки погрешностей независимых прогнозов. Самое важное, что ГТОр обеспечивает неухудшение качества прогнозирования при замене независимых прогнозов на согласованные прогнозы. При этом на структуру иерархии, ограничения-неравенства на прогнозы и функцию суммарных потерь накладываются лишь общие ограничения, гарантирующие существование решения оптимизационной задачи (6). Еще одно достоинство алгоритма ГТОр в том, что он позволяет согласовывать прогнозы для иерархий любой сложности одновременно по всем уровням, учитывая все связи в иерархии и решая одну оптимизационную задачу. От сложности иерархии и количества временных рядов и уровней в оптимизационной задаче зависит число переменных и ограничений.

6 Дивергенция Брегмана

В этом разделе будет описано семейство функций двух переменных, удовлетворяющих предположениям 1–4, и приведен ряд примеров функций из этого семейства, которые можно использовать в задаче согласования прогнозов в качестве функции суммарных потерь l_h (2). Все эти функции двух переменных называются дивергенциями Брегмана [19, 21] и порождаются функциями одной переменной, обладающими следующими свойствами.

Определение 7. Функцией Лежандра [19] называется функция $F: \mathcal{B} \rightarrow \mathbb{R}$, которая удовлетворяет следующим условиям:

- $\mathcal{B} \subseteq \mathbb{R}^d$ — непустое множество, и внутренность \mathcal{B} выпукла;
- F — строго выпуклая функция с непрерывной первой производной на множестве \mathcal{B} ;
- если $\chi_1, \chi_2, \dots \in \mathcal{B}$ — последовательность, сходящаяся к граничной точке \mathcal{B} , то $\|\nabla F(\chi_n)\| \rightarrow \infty$ при $n \rightarrow \infty$.

Определение 8. Дивергенцией Брегмана, порожденной функцией Лежандра $F: \mathcal{B} \rightarrow \mathbb{R}$, называется неотрицательная функция $D_F: \mathcal{B} \times \text{int}(\mathcal{B}) \rightarrow \mathbb{R}$, определенная как

$$D_F(\mathbf{u}, \mathbf{v}) = F(\mathbf{u}) - F(\mathbf{v}) - (\mathbf{u} - \mathbf{v})\nabla F(\mathbf{v}).$$

Свойства дивергенции Брегмана:

- для всех \mathbf{u} и \mathbf{v} выполнено $D_F(\mathbf{u}, \mathbf{v}) \geq 0$. Это следует из выпуклости функции F ;
- $D_F(\mathbf{u}, \mathbf{v})$ выпукла по первому аргументу \mathbf{u} , но необязательно выпукла по второму аргументу \mathbf{v} ;
- для любых $\alpha, \beta \in \mathbb{R}$ и любых функций Лежандра F_1 и F_2 выполнено

$$D_{\alpha F_1 + \beta F_2}(\mathbf{u}, \mathbf{v}) = \alpha D_{F_1}(\mathbf{u}, \mathbf{v}) + \beta D_{F_2}(\mathbf{u}, \mathbf{v}).$$

Определение 9. Пусть $F: \mathcal{B} \rightarrow \mathbb{R}$ — функция Лежандра и $\mathcal{A} \subset \mathbb{R}^d$ — замкнутое выпуклое множество, такое что $\mathcal{A} \cap \mathcal{B} \neq \emptyset$. Проекция Брегмана \mathbf{w}' точки $\mathbf{w} \in \text{int}(\mathcal{B})$ на множество \mathcal{A} — это

$$\mathbf{w}' = \arg \min_{\mathbf{u} \in \mathcal{A} \cap \mathcal{B}} D_F(\mathbf{u}, \mathbf{w}).$$

Теорема 3. Для всех функций Лежандра $F: \mathcal{B} \rightarrow \mathbb{R}$, для всех замкнутых выпуклых множеств $\mathcal{A} \subset \mathbb{R}^d$, имеющих непустое пересечение $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, и для всех точек $\mathbf{w} \in \text{int}(\mathcal{B})$ проекция Брегмана точки \mathbf{w} на множество \mathcal{A} существует и единственна.

Доказательство этой теоремы приведено в [21].

Теорема 4. Пусть F — функция Лежандра. Для всех $\mathbf{w} \in \text{int}(\mathcal{B})$ и для всех замкнутых выпуклых множеств $\mathcal{A} \subseteq \mathbb{R}^d$ с непустым пересечением $\mathcal{A} \cap \mathcal{B} \neq \emptyset$, если \mathbf{w}' — проекция Брегмана точки \mathbf{w} на множество \mathcal{A} ($\mathbf{w}' = \arg \min_{\mathbf{v} \in \mathcal{A} \cap \mathcal{B}} D_F(\mathbf{v}, \mathbf{w})$), то верно неравенство:

$$D_F(\mathbf{u}, \mathbf{w}) \geq D_F(\mathbf{u}, \mathbf{w}') + D_F(\mathbf{w}', \mathbf{w})$$

для всех $\mathbf{u} \in \mathcal{A}$.

Доказательство этого факта можно найти в [19].

Соотнесем перечисленные свойства дивергенции Брегмана и предположения 1–4. Определение 8 и свойство 1 дивергенции Брегмана обеспечивают выполнение предположения 2 о знаке функции суммарных потерь. В определении 9 и теореме 3 предполагается выпуклость и замкнутость множеств \mathcal{A} и \mathcal{B} и их непустое пересечение, как и в предположении 1. Теорема 4 гарантирует выполнение предположения 3 о выпуклости множеств и существовании и единственности проекции. Наконец, теорема 4 гарантирует выполнение предположения 4. Следовательно, для функций

суммарных потерь l_h (2), являющихся дивергенциями Брегмана, выполнены все условия теоремы 2. Использование в качестве вектора согласованных прогнозов $\hat{\varphi}$ проекции Брегмана вектора независимых прогнозов $\hat{\chi} \in \mathcal{B}$, где множество \mathcal{B} определено по формуле (4) и является выпуклым и замкнутым, имеющим непустое пересечение с множеством \mathcal{A} (3), на множество $\mathcal{A} \cap \mathcal{B}$, включающее векторы прогнозов, удовлетворяющих структуре иерархии, гарантирует неухудшение качества прогнозов.

Следующие функции являются дивергенциями Брегмана и могут быть использованы в качестве функций суммарных потерь (2) при согласовании прогнозов.

Квадрат евклидова расстояния

$D_F(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|^2$ — канонический пример дивергенции Брегмана, порождается функцией $F(\mathbf{u}) = \|\mathbf{u}\|^2$.

Квадрат расстояния Махаланобуса

$D_F(\mathbf{u}, \mathbf{v}) = (1/2)(\mathbf{u} - \mathbf{v})^T Q(\mathbf{u} - \mathbf{v})$ — обобщение евклидова расстояния, порождается квадратичной формой $F(\mathbf{u}) = (1/2)\mathbf{u}^T Q\mathbf{u}$.

Обобщенная дивергенция Кульбака–Лейблера

$D_F(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^n u_i \log(u_i/v_i) - \sum_{i=1}^n u_i + \sum_{i=1}^n v_i$ порождается функцией $F(\mathbf{u}) = \sum_{i=1}^n u_i \log u_i - \sum_{i=1}^n u_i$.

Расстояние Itakura–Saito

$D_F(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^n ((u_i/v_i) - \log(u_i/v_i) - 1)$ порождается функцией $F(\mathbf{u}) = -\sum_{i=1}^n \log u_i$.

7 Эксперимент

В экспериментальной части рассматриваются данные о посуточной загруженности железнодорожных узлов РЖД. Из имеющихся данных были сформированы временные ряды, описывающие отправление 37 различных типов груза со станций 98 железнодорожных веток посуточно. Рассмотрены два вида иерархии.

В случае плоской двухуровневой иерархии решены задачи согласования прогнозов отправления грузов по отдельности и всех грузов в сумме для каждой ветки, а также решены задачи согласования прогнозов отправления груза с каждой ветки и суммарного отправления груза со всех веток для каждого типа груза.

В случае неплоской трехуровневой иерархии решена задача согласования прогнозов всех имеющихся временных рядов. Демонстрируется уменьшение значения функции суммарных потерь при переходе от независимых прогнозов к согласованным. В качестве функции суммарных потерь был

Вид записи базы данных железнодорожных перевозок

Дата погрузки	Станция отправления	Станция назначения	Количество вагонов	Код груза	Род вагона	Суммарный вес груза	Признак маршрутной отправки
2007-01-01	020108	932902	1	1	216	56	9

использован квадрат евклидового расстояния. Независимые прогнозы были получены алгоритмом Hist, описанным в работах [7, 22].

Экспериментальные данные. В эксперименте использованы данные о посуточной загруженности железнодорожных узлов РЖД с 1 января 2007 г. по 22 апреля 2008 г. В таблице приведен пример записи базы данных.

Коды станций представляют собой шестизначные числа. Станции, в коде которых две первые цифры совпадают, входят в одну железнодорожную ветку. Станций отправления — 1566, станций назначения — 1902, веток — 98. Код груза — натуральное число от 1 до 37; также имеются перевозки, где код груза не указан. Род вагона — натуральное число, в имеющихся данных 75 различных родов вагонов.

Иерархическая структура. Экспериментальные данные удовлетворяют структуре, изображенной на рис. 3.

Как видно из рисунка, иерархия не является плоской и содержит три уровня временных рядов. Временные ряды нижнего уровня этой иерархии

имеют два индекса, соответствующих номеру ветки и коду груза: $x(i, j)$, $i = 1, \dots, n$, $j = 1, \dots, m$, где n — число веток, а m — количество грузов. На среднем уровне — два семейства временных рядов. Временные ряды, соответствующие суммарному отправлению всех грузов с каждой ветки, обозначаются $x(i, :)$, $i = 1, \dots, n$. Ряды среднего уровня, соответствующие суммарному отправлению со всех веток каждого из грузов, обозначаются $x(:, j)$, $j = 1, \dots, m$. Временной ряд верхнего уровня обозначается $x(:, :)$.

Условие согласованности для трехуровневой иерархии задается равенствами ($t = 1, \dots, T$):

$$\left. \begin{aligned} x_t(:, :) &= \sum_{i=1}^n x_t(i, :); \\ x_t(:, j) &= \sum_{i=1}^n x_t(i, j); \\ x_t(i, :) &= \sum_{j=1}^m x_t(i, j), \quad i = 1, \dots, n; \\ x_t(i, j) &= \sum_{i=1}^n x_t(i, j), \quad j = 1, \dots, m. \end{aligned} \right\} \quad (7)$$

Векторная запись срезов иерархии, независимых прогнозов и согласованных прогнозов имеет размерность $d = 1 + n + m + nm$ и выглядит следующим образом:

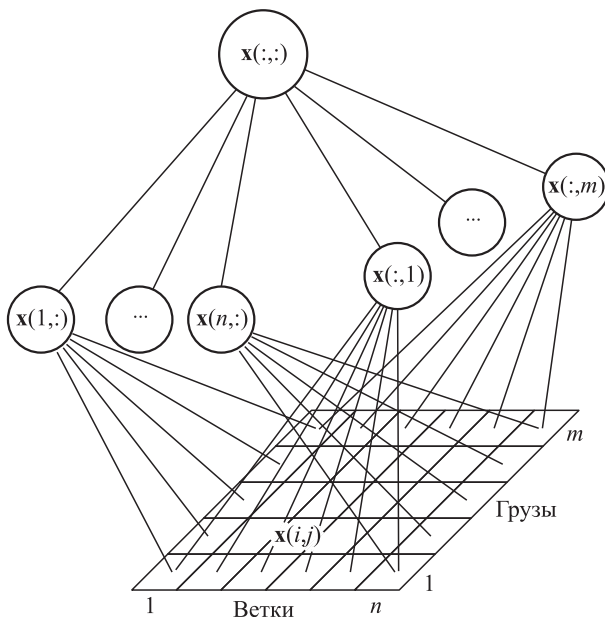


Рис. 3 Неплоская трехуровневая иерархия

$$\chi_t = \begin{pmatrix} x_t(:, :) \\ x_t(1, :) \\ \vdots \\ x_t(n, :) \\ x_t(:, 1) \\ \vdots \\ x_t(:, m) \\ x_t(1, 1) \\ \vdots \\ x_t(1, m) \\ \vdots \\ x_t(n, 1) \\ \vdots \\ x_t(n, m) \end{pmatrix}; \quad \hat{\chi} = \begin{pmatrix} \hat{x}(:, :) \\ \hat{x}(1, :) \\ \vdots \\ \hat{x}(n, :) \\ \hat{x}(:, 1) \\ \vdots \\ \hat{x}(:, m) \\ \hat{x}(1, 1) \\ \vdots \\ \hat{x}(1, m) \\ \vdots \\ \hat{x}(n, 1) \\ \vdots \\ \hat{x}(n, m) \end{pmatrix};$$

$$\hat{\varphi} = \begin{pmatrix} \hat{y}(:, :) \\ \hat{y}(1, :) \\ \vdots \\ \hat{y}(n, :) \\ \hat{y}(:, 1) \\ \vdots \\ \hat{y}(:, m) \\ \hat{y}(1, 1) \\ \vdots \\ \hat{y}(1, m) \\ \vdots \\ \hat{y}(n, 1) \\ \vdots \\ \hat{y}(n, m) \end{pmatrix}.$$

Матрица связей для условия (7) имеет размер $(2 + n + m) \times (1 + n + m + nm)$:

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

Требуется, чтобы все прогнозы были неотрицательны, поэтому множество \mathcal{B} (4) задается как

$$\mathcal{B} = \{\chi \in \mathbb{R}^d \mid \chi(i) \in [0, +\infty], i = 1, \dots, d\}.$$

В качестве функции суммарных потерь l_h (2) используется квадрат евклидова расстояния.

Оптимизационная задача. Задача, решенная для согласования прогнозов рассматриваемой иерархии, имеет вид:

$$\begin{cases} \|\chi - \hat{\chi}\|^2 \rightarrow \min_{\chi}, \\ S\chi = 0, \\ \chi \geq 0. \end{cases}$$

Результаты эксперимента. Для прогноза были использованы временные ряды, описывающие суммарный вес отправленных грузов разных типов по каждой ветке. Значения временных рядов нижнего уровня иерархии были отнормированы на отрезок $[0; 1]$ по формуле

$$x_t^{\text{norm}}(i, j) = \frac{x_t(i, j) - m(i, j)}{M(i, j) - m(i, j)},$$

где

$$m = \min_{t=1, \dots, T} x_t(i, j), \quad M = \max_{t=1, \dots, T} x_t(i, j).$$

Прогнозы были построены и согласованы для 100 последних точек истории. Для каждого отсчета времени строился вектор независимых прогнозов $\hat{\chi}$ и вектор согласованных прогнозов $\hat{\varphi}$. Для каждого вектора вычислялось значение функции суммарных потерь, затем вычислялись потери игрока 1 в игре (5), равные разности суммарных потерь при использовании согласованных прогнозов и суммарных потерь при использовании независимых прогнозов:

$$L_t = l_h(\hat{\varphi}, \chi_t) - l_h(\hat{\chi}, \chi_t), \quad t = T - 100 + 1, \dots, T.$$

Значения этой величины изображены на рис. 4. Теоретические выкладки подтверждаются на практике. Для плоских двухуровневых иерархий есть случаи, когда суммарные потери при переходе к согласованным прогнозам не изменяются, и случаи, когда суммарные потери уменьшаются. Для неплоской трехуровневой иерархии суммарные потери во всех контрольных точках уменьшаются.

8 Заключение

Предложен алгоритм ГТОр для согласования прогнозов иерархических временных рядов. Алгоритм не требует оценки погрешностей независимых прогнозов и не требует их несмещенности. Для любого набора независимых прогнозов временных рядов алгоритм согласования не ухудшает качество прогнозирования. Возможна работа с иерархическими структурами любой сложности. Все свойства алгоритма согласования прогнозов ГТОр подтверждаются на практике.

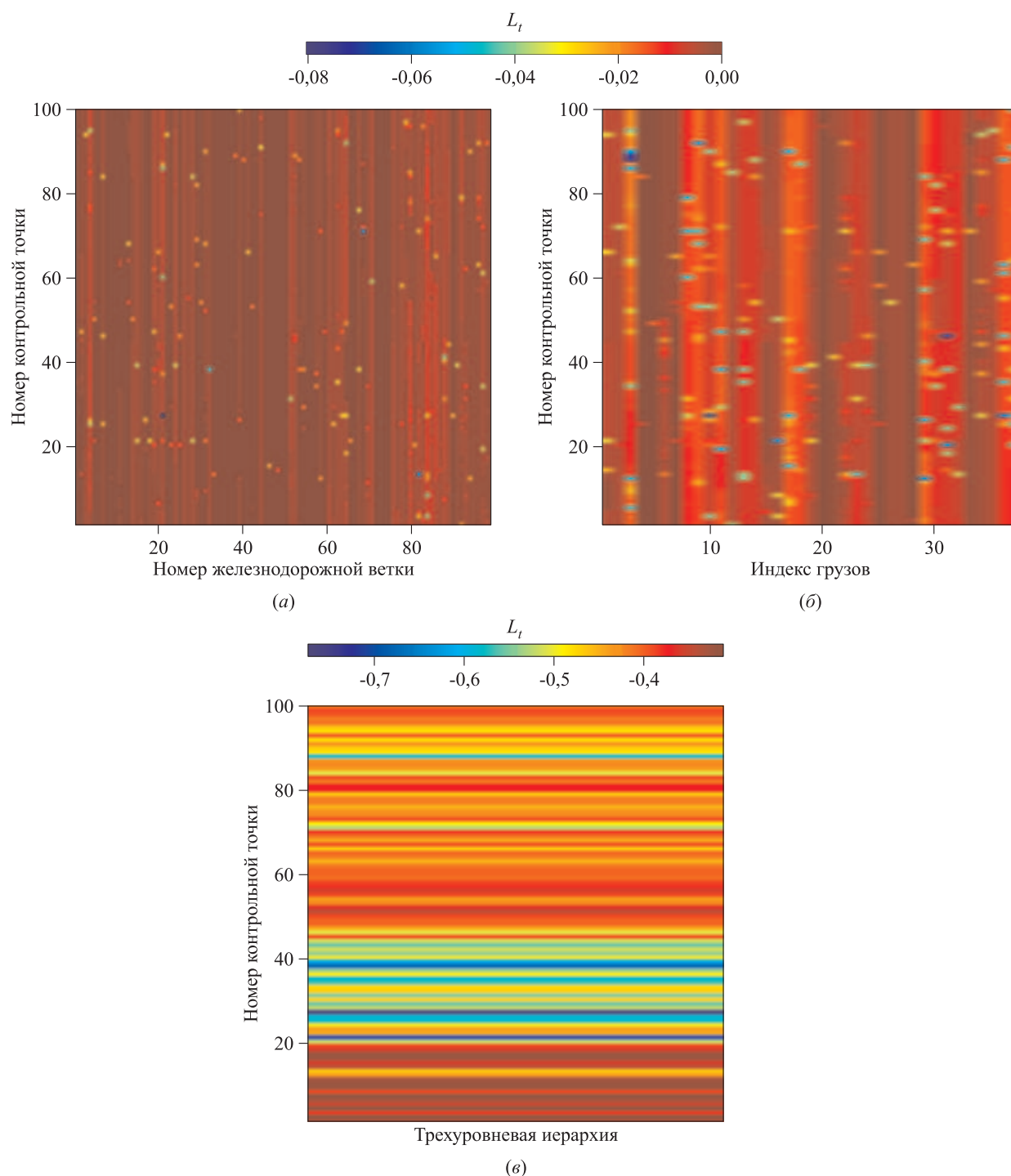


Рис. 4 Изменение суммарных потерь L_t : (а) для каждой ветки; (б) для каждого груза; (в) вся иерархия

Литература

1. Токмакова А. А., Стрижов В. В. Оценивание гиперпараметров линейных и регрессионных моделей при отборе шумовых и коррелирующих признаков // Информатика и её применения, 2012. Т. 6. Вып. 4. С. 66–75.
2. Васильев Н. С. Использование принципа равновесия для управления маршрутизацией в транспортных сетях // Информатика и её применения, 2014. Т. 8. Вып. 1. С. 28–35.
3. Hong T., Pinson P., Fan S. Global energy forecasting competition 2012 // Int. J. Forecasting, 2014. Vol. 30. No. 2. P. 357–363.

4. Kaggle. <https://www.kaggle.com>.
5. Hyndman R. J., Ahmed R. A., Athanasopoulos G., Shang H. L. Optimal combination forecasts for hierarchical time series // *Comput. Stat. Data Anal.*, 2011. Vol. 55. No. 9. P. 2579–2589.
6. Кузнецов М. П., Мафусалов А. А., Животовский Н. К., Зайцев Е. Ю., Сунгуров Д. С. Сглаживающие алгоритмы прогнозирования // *Машинное обучение и анализ данных*, 2011. Т. 1. Вып. 1. С. 104–112.
7. Стенина М. М., Стрижов В. В. Согласование агрегированных и детализированных прогнозов при решении задач непараметрического прогнозирования // *Системы и средства информатики*, 2014. Т. 24. Вып. 2. С. 21–34.
8. Grunfeld Y., Griliches Z. Is aggregation necessarily bad? // *Rev. Econ. Stat.*, 1960. Vol. 42. No. 1. P. 1–13.
9. Orcutt G. H., Watts H. W., Edwards J. B. Data aggregation and information loss // *Am. Econ. Rev.*, 1968. Vol. 58. No. 4. P. 773–787.
10. Edwards J. B., Orcutt G. H. Should aggregation prior to estimation be the rule? // *Rev. Econ. Stat.*, 1969. Vol. 51. No. 4. P. 409–420.
11. Shlifer E., Wolff R. W. Aggregation and proration in forecasting // *Manage. Sci.*, 1979. Vol. 25. No. 6. P. 594–603.
12. Fogarty D. W., Blackstone J. H., Hoffman T. R. Production and inventory management. — 2nd ed. — Cincinnati, OH, USA: South-Western Publication Co., 1990. 880 p.
13. Narasimhan S. L., McLeavey D. W., Billington P. J. Production planning and inventory control. — 2nd ed. — Englewood Cliffs, NJ, USA: Prentice Hall, 1995. 716 p.
14. Schwarzkopf A. B., Tersine R. J., Morris J. S. Top-down versus bottom-up forecasting strategies // *Int. J. Prod. Res.*, 1998. Vol. 26. No. 11. P. 1833–1843.
15. Fliedner G. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation // *Comput. Oper. Res.*, 1999. Vol. 26. No. 10–11. P. 1133–1149.
16. Van Erven T., Cugliari J. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. 2013. <https://hal.inria.fr/hal-00920559>.
17. Петросян Л. А., Зенкевич Н. А., Семин Е. А. Теория игр. — М.: Университет, 1998. 301 с.
18. Меньшиков И. С. Лекции по теории игр и экономическому моделированию. — 2-е изд., испр. и доп. — М.: Контакт Плюс, 2010. 336 с.
19. Cesa-Bianchi N., Lugosi G. Prediction, learning, and games. — Cambridge: Cambridge University Press, 2006. Vol. 1. 403 p.
20. Boyd S., Vandenberghe L. Convex optimization. — Cambridge: Cambridge University Press, 2009. 732 p.
21. Bregman L. M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming // *USSR Comput. Math. Math. Phys.*, 1967. Vol. 7. No. 3. P. 200–217.
22. 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. Т. 1. Вып. 4. С. 448–465.

Поступила в редакцию 27.10.14

FORECASTS RECONCILIATION FOR HIERARCHICAL TIME SERIES FORECASTING PROBLEM

M. M. Stenina¹ and V. V. Strijov²

¹Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation

²Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation

Abstract: The hierarchical time series forecasting problem is researched. Time series forecasts must satisfy the physical constraints and the hierarchical structure. In this paper, a new algorithm for hierarchical time series forecasts reconciliation is proposed. The algorithm is called GTOp (Game-theoretically optimal reconciliation). It guarantees that the quality of reconciled forecasts is not worse than the quality of self-dependent forecasts. This approach is based on Nash equilibrium search for the antagonistic game and turns the forecasts reconciliation problem into the optimization problem with equality and inequality constraints. It is proved that the Nash equilibrium in pure strategies exists in the game if some assumptions about the hierarchical structure, the physical constraints, and the loss function are satisfied. The algorithm performance is demonstrated for different types of hierarchical structures of time series.

Keywords: hierarchical time series; reconciliation of time series forecasts; antagonistic game; Nash equilibrium

DOI: 10.14357/19922264150209

Acknowledgments

The research was financially supported by the Russian Foundation for Basic Research (project 13-07-13139).

References

1. Tokmakova, A. A., and V. V. Strizhov. 2012. Otsenivanie giperparametrov lineynykh i regressionnykh modeley pri otbore shumovykh i korrelirovannykh priznakov [Estimation of linear model hyperparameters for noise or correlated feature selection problem]. *Informatika i ee Primeneniya — Inform. Appl.* 6(4):66–75.
2. Vasil'ev, N. S. 2014. Ispol'zovanie printsipa ravnovesiya dlya upravleniya marshrutizatsiyey v transportnykh setyakh [Equilibrium principle application to routing control in packet data transmission networks]. *Informatika i ee Primeneniya — Inform. Appl.* 8(1):28–35.
3. Hong, T., P. Pinson, and S. Fan. 2014. Global energy forecasting competition 2012. *Int. J. Forecasting* 30(2):357–363.
4. Kaggle. Available at: <https://www.kaggle.com> (accessed May 20, 2015).
5. Hyndman, R. J., R. A. Ahmed, G. Athanasopoulos, and H. L. Shang. 2011. Optimal combination forecasts for hierarchical time series. *Comput. Stat. Data Anal.* 55(9):2579–2589.
6. Kuznetsov, M. P., A. A. Mafusalov, N. K. Zhivotovskiy, E. Yu. Zaytsev, and D. S. Sungurov. 2011. Sglazhivayushchie algoritmy prognozirovaniya [Smoothing forecast algorithms]. *Mashinnoe obuchenie i analiz dannykh* [J. Machine Learning Data Anal.] 1(1):104–112.
7. Stenina, M. M., and V. V. Strizhov. 2014. Soglasovanie agregirovannykh i detalizirovannykh prognozov pri reshenii zadach neparametricheskogo prognozirovaniya [Reconciliation of aggregated and disaggregated time series forecasts in nonparametric forecasting problem]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 24(2):21–34.
8. Grunfeld, Y., and Z. Griliches. 1960. Is aggregation necessarily bad? *Rev. Econ. Stat.* 42(1):1–13.
9. Orcutt, G. H., H. W. Watts, and J. B. Edwards. 1968. Data aggregation and information loss. *Am. Econ. Rev.* 58(4):773–787.
10. Edwards, J. B., and G. H. Orcutt. 1969. Should aggregation prior to estimation be the rule? *Rev. Econ. Stat.* 51(4):409–420.
11. Shlifer, E., and R. W. Wolff. 1979. Aggregation and proration in forecasting. *Manage. Sci.* 25(6):594–603.
12. Fogarty, D. W., J. H. Blackstone, and T. R. Hoffman. 1990. *Production and inventory management*. 2nd ed. Cincinnati, OH: South-Western Publication Co. 880 p.
13. Narasimhan, S. L., D. W. McLeavey, and P. J. Billington. 1995. *Production planning and inventory control*. 2nd ed. Englewood Cliffs, NJ: Prentice Hall. 716 p.
14. Schwarzkopf, A. B., R. J. Tersine, and J. S. Morris. 1998. Top-down versus bottom-up forecasting strategies. *Int. J. Prod. Res.* 26(11):1833–1843.
15. Fliedner, G. 1999. An investigation of aggregate variable time series forecast strategies with specific subaggregate time series statistical correlation. *Comput. Oper. Res.* 26(10–11):1133–1149.
16. Van Erven, T., and J. Cugliari. 2013. Game-theoretically optimal reconciliation of contemporaneous hierarchical time series forecasts. Available at: <https://hal.inria.fr/hal-00920559> (accessed May 20, 2015).
17. Petrosyan, L. A., N. A. Zenkevich, and E. A. Semina. 1998. *Teoriya igr* [Games theory]. Moscow: Knizhnyy Dom Universitet. 301 p.
18. Men'shikov, I. S. 2010. *Lektsii po teorii igr i ekonomicheskomu modelirovaniyu* [Games theory and economics modeling lectures]. 2nd ed. Moscow: OOO Kontakt Plyus. 336 p.
19. Cesa-Bianchi, N., and G. Lugosi. 2006. *Prediction, learning, and games*. Cambridge: Cambridge University Press. Vol. 1. 403 p.
20. Boyd, S., and L. Vandenberghe. 2009. *Convex optimization*. Cambridge: Cambridge University Press. 732 p.
21. Bregman, L. M. 1967. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* 7(3):200–217.
22. Val'kov, A. S., E. M. Kozhanov, M. M. Medvednikova, and F. I. Khusainov. 2012. Neparametricheskoe prognozirovaniye zagruzhennosti sistemy zheleznodorozhnykh uzlov po istoricheskim dannym [Nonparametric forecasting of railroad stations occupancy according to historical data]. *Mashinnoe Obuchenie i Analiz Dannykh* [J. Machine Learning Data Anal.] 1(4):448–465.

Received October 27, 2014

Contributors

Stenina Mariya M. (b. 1991) — student, Moscow Institute of Physics and Technology, 9 Institutskiy Per., Dolgoprudny, Moscow Region 141700, Russian Federation; mmedvednikova@gmail.com

Strijov Vadim V. (b. 1967) — Doctor of Science in physics and mathematics, leading scientist, Dorodnicyn Computing Center, Russian Academy of Sciences, 40 Vavilov Str., Moscow 119333, Russian Federation; strijov@gmail.com