

Онлайн агрегирование систем конформного прогнозирования

В.В. Вьюгин, В.Г. Трунов

*Институт проблем передачи информации им. А.А.Харкевича,
Российская академия наук, Москва, Россия
e-mail: vyugin@iitp.ru*

Поступила в редколлегию 10.04.2023

Аннотация—Рассматривается задача вероятностного прогнозирования в режиме онлайн. Вероятностные прогнозы получены в результате применения систем конформного прогнозирования. Системы конформного прогнозирования представляют собой новый метод для построения достоверных прогнозов на основе точечных прогнозов регрессионного алгоритма. В работе рассматривается случай, когда в каждый момент времени несколько конкурирующих методов (экспертов) выдают свои прогнозы в виде функций конформного распределения. Вероятностные прогнозы экспертов объединяются с помощью агрегирующего алгоритма в один вероятностный прогноз на каждом шаге процесса прогнозирования, при этом прогнозы экспертов могут использоваться с дисконтом.

Разработана технология построения прогнозирующих экспертных алгоритмов и агрегирования их вероятностных прогнозов на примере задачи прогнозирования почасовой нагрузки электрической сети в зависимости от прогноза температуры. Представлены результаты численных экспериментов на реальных данных, проведен сравнительный анализ метода конформных предсказаний и метода гауссовских смесей.

КЛЮЧЕВЫЕ СЛОВА: машинное обучение, вероятностное прогнозирование в режиме онлайн, конформные прогнозирующие системы, калибровка прогнозов, предсказания с использованием экспертных стратегий, регрет, уровни компетентности, агрегирующий алгоритм, скоринговое правило непрерывной ранжированной вероятности CRPS, прогнозирование нагрузки электрической сети.

DOI: 10.53921/18195822_2023_23_2_185

1. ВВЕДЕНИЕ

Вероятностное прогнозирование важно для многих приложений, поскольку прогнозирующее распределение описывает неопределенность прогноза, а также обеспечивает возможность вычисления вероятностей любых событий, связанных с прогнозируемым параметром.

Для иллюстрации предлагаемых методов построения прогнозирующих моделей и их агрегирования используются реальные данные задачи прогнозирования нагрузок в электрической сети (см. также [1], [2]).

В первой части работы будут рассматриваться методы прогнозирования числовых характеристик y объектов x сложной природы, где $x \in \mathcal{R}^k$, (в простейшем случае $k = 1$) и $y \in \mathcal{R}$. Предполагается, что пары "объект–метка" (x, y) генерируются некоторым вероятностным источником (распределением), причем, пары (x, y) независимы и одинаково распределены (i.i.d). Конкретная форма этого распределения вероятностей может быть нам неизвестна и в дальнейшем не будет использоваться.

Имеется большое число методов точечного, интервального и вероятностного прогнозирования. Первая часть этой работы основана на недавно предложенном непараметрическом подходе, использующим системы конформного прогнозирования [3]. Системы конформного прогнозирования предназначены для того, чтобы делать достоверные вероятностные прогнозы для тестовых меток. Они развивают исследования прогностических (или доверительных) распределений [4] и разработаны на основе идеи машинного обучения [5].

Система конформного прогнозирования строит прогнозное распределение вероятностей только на основе соотношения между текущим и прошлыми точечными прогнозами произвольного алгоритма.

В работах [7], [8], [3], [9] были введены расщепленные конформные прогностические системы, а также кросс-конформные прогностические системы [10]. Система первого типа разбивает обучающие данные на две части – тренировочную и калибровочную выборки, система второго типа использует процесс перекрестной проверки для расчета соответствующей статистики. Мы будем использовать системы первого типа.

Система конформного прогнозирования строится следующим образом. Обучающая выборка разделяется на тренировочную и калибровочную. Тренировочная выборка используется для построения прогнозирующего правила (алгоритма), который, используя объект x , способен производить точечный прогноз его метки y . Калибровочная выборка позволяет на основе сравнения точечного прогноза с прошлыми прогнозами и известными исходами перестроить прогноз алгоритма в (прогнозное) распределение вероятностей. Тем самым появляется возможность оценить неточность и неопределенность точечного прогноза. Например, используя это распределение вероятностей, можно строить доверительные интервальные прогнозы для y .

В данной работе тренировочная выборка используется только на начальном этапе обучения – при построении алгоритма регрессии. Этот алгоритм будет использоваться в дальнейшем для построения конформных прогнозирующих систем.

В данной работе системы конформного прогнозирования будут строиться в режиме онлайн. В каждый момент времени t система получает очередной тестируемый объект x_t , для которого необходимо определить функцию распределения его метки y_t . Для этого используется калибровочная выборка $\tilde{z}_1^m = (\tilde{x}_1, \tilde{y}_1), \dots, z_m = (\tilde{x}_m, \tilde{y}_m)$, состоящая из наблюдаемых объектов и их меток, используются также соответствующие прогнозы алгоритма для этих меток. Калибровочная выборка образуется на начальном этапе обучения и может пополняться в каждый момент времени на основе прошлых данных, т.е. в режиме онлайн. В простейшем варианте калибровочная выборка может быть одна и та же для всех моментов времени.

Распределение вероятностей для тестовой метки строится с помощью специальной статистики – меры конформности (или неконформности) пары (x, y) относительно калибровочной выборки $\tilde{z}_1^m$.

Распределение, предсказанное для тестовой метки, должно обладать свойством валидности [3], другими словами, это означает, что распределение должно обладать статистической совместимостью с реализациями, т.е., отражать истинное положение дел.

Основное предположение для обеспечения такой валидности заключается в том, что элементы калибровочной выборки и тестируемое значение должны быть независимо и одинаково распределены относительно некоторого вероятностного распределения на парах (x, y) , которое может быть нам неизвестно. По калибровочной выборке для любого x в качестве прогноза рассматривается функция распределения вероятностей $F_x(y)$.

В целях достижения более точного прогнозирования используется метод построения предсказаний с использованием экспертных стратегий. На этапе обучения обучающая часть вре-

менного ряда разбивается на участки однородности, которые определяются исходя из свойств предметной области.

На этапе обучения регрессионная зависимость строится отдельно по каждому элементу разбиения. Конформные прогностические системы, которые строятся на основе таких зависимостей предназначены для получения прогноза в соответствующих областях пространства входных переменных. Мы рассматриваем их как вероятностные прогнозы соответствующих экспертных стратегий (экспертов).

В наших экспериментах почасовые данные о нагрузке электрических сетей разбиваются на сезоны (зима, весна, лето, осень) и времена суток (утро, день, вечер, ночь). На этапе обучения регрессионная зависимость строится отдельно по каждому элементу разбиения (области компетентности эксперта). Полученные на очередном шаге конформные прогностические системы позволяют строить вероятностные прогнозы соответствующих экспертных стратегий (экспертов) на этом шаге.

Хотя каждый эксперт обучается в своей области компетентности, мы будем частично использовать его прогнозы также и в соседних областях, имея ввиду, что его прогностические способности падают постепенно вне его области компетентности. Таким образом, в каждой точке прогнозирования мы можем иметь прогнозы сразу нескольких экспертов. Возникает задача агрегации (вероятностных) прогнозов таких экспертов с учетом степени их компетентности.

Вторая часть работы посвящена реализации постановки, когда несколько конкурирующих методов дают прогнозы в режиме онлайн в виде функций распределения вероятностей. Эти прогнозы могут привести к большим или меньшим потерям. Задача агрегирующего алгоритма состоит в том, чтобы объединить эти прогнозы в один агрегированный вероятностный прогноз таким образом, чтобы минимизировать потери, исходя из имеющейся прошлой информации.

Эта проблема решается в рамках теории предсказаний с использованием экспертных стратегий (Prediction with Expert Advice). Каждая экспертная стратегия использует свою специфическую предсказательную модель, описывающую и прогнозирующую наблюдаемые данные.

Агрегирующий алгоритм объединяет прогнозы экспертных моделей в зависимости от их потерь в прошлом. В качестве функции потерь используется скоринговое правило непрерывной ранжированной вероятности (Continuous Ranked Probability Score – CRPS), с помощью которого измеряется различие между функцией распределения и числовым исходом, см. [11], [12], [13], [14], [15].

Разность между кумулятивными (суммарными) потерями агрегирующего алгоритма и кумулятивными потерями произвольного экспертного алгоритма, накопленными за весь период прогнозирования, называется регретом. Цель агрегирующего алгоритма заключается в том, чтобы минимизировать регрет относительно любой экспертной стратегии.

В работах [16], [17], [18], [19], [20] были развиты методы агрегирования в режиме онлайн точечных прогнозов. В этом случае эксперты выдают прогнозы в виде чисел, агрегирующий алгоритм выдает прогноз такого же вида.

В работах [21], [22], [23] рассматривалась постановка, когда каждый эксперт представляет функцию вероятностного распределения, а не точечный прогноз. Агрегирующий алгоритм также представляет свой прогноз в форме функции распределения вероятностей, рассчитанной с использованием вероятностных прогнозов экспертов.

В отличие от этих работ, в которых вероятностное распределение каждого эксперта строилось на этапе обучения и далее уже не изменялось, в данной работе вероятностное распределение эксперта строится на каждом шаге прогнозирования с использованием калибровочной выборки, которая может пополняться новыми данными. Это позволяет более точно учитывать возможные изменения параметров источника данных.

В данной работе в каждый момент времени каждому эксперту придается свой уровень компетентности, который представляет собой число, лежащее в интервале $[0, 1]$. Значение уровня компетентности, равное 1, означает, что соответствующий экспертный прогноз полностью принимается во внимание агрегирующим алгоритмом, при значении уровня компетентности 0 прогноз эксперта вообще им не учитывается, а при промежуточном значении уровня компетентности соответствующий прогноз учитывается лишь частично в зависимости от значения уровня компетентности.

В работе представлены формулы для прямого расчета агрегированного распределения вероятностей на основе распределений вероятностей, представленных экспертами в качестве их прогнозов с учетом уровней их компетентности.

В разделе 2.1 приведены основные понятия и определения метода конформных предсказаний. В разделе 2.2 представлены некоторые детали агрегирующего алгоритма Вовка [18], а также его расширение на случай вероятностных прогнозов. В разделе 2.3 представлен метод вычисления агрегированного распределения вероятностей на основе распределений, представленных экспертами. В разделе 2.4 мы обобщаем агрегирующий алгоритм на тот случай, когда прогнозы экспертов снабжены уровнями компетентности. В этом разделе вводится понятие дисконтированного регрета и получены его верхние оценки для агрегирующего алгоритма.

Эффективность предложенных методов продемонстрирована результатами численных экспериментов, приведенных в разделе 4. Материалом для экспериментального исследования послужили данные конкурса Global Energy Forecasting – Competition 2014 (GEFCOM 2014, Track Load), который проводился на платформе Kaggle [2].

2. ОСНОВНЫЕ ПОНЯТИЯ И МЕТОДЫ

2.1. Конформные предсказания

Пусть $\mathcal{X}$ – измеримое пространство, которое мы будем называть пространством объектов. Пространство наблюдения определяется как $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$; его элемент $z = (x, y)$, где $x \in \mathcal{X}$ и $y \in \mathcal{Y}$, интерпретируется как наблюдение, состоящее из объекта x и его метки y .

Задача “обучения с учителем” состоит в том, чтобы, учитывая обучающие данные, состоящие из наблюдений $z_i = (x_i, y_i)$, $i = 1, \dots, n$ и новый (тестовый) объект $x_{n+1} \in \mathcal{X}$, предсказать соответствующую метку y_{n+1} .

Классическая постановка метода конформных предсказаний имеет следующий вид [5]. Конформные предикторы используют предыдущие данные для определения размера доверительного множества Γ_n^ϵ для новых предсказаний. Пусть имеется алгоритм f , который генерирует предсказания $\hat{y}_{n+1}$ и задан уровень значимости ϵ . Тогда описываемый метод позволяет получить множество Γ_n^ϵ , включающее в себя эти предсказания, которое также содержит истинное значение y_{n+1} с уровнем значимости $1 - \epsilon$. Этот метод может применяться к любому алгоритму машинного обучения. Предсказания алгоритма обладают свойством валидности, если при увеличении числа исходов (x_n, y_n) и соответствующих прогнозов $\hat{y}_{n+1} = f_{z_1^n}(x_{n+1})$, с вероятностью 1 доля ошибок будет стремиться к ϵ .¹

Мерой конформности (мерой соответствия) называется измеримая функция $A : \cup_{n=1}^\infty Z^{n+1} \rightarrow \mathcal{R}$, инвариантная относительно перестановок обучающих наблюдений: для любого n , для любой перестановки π множества $\{1, \dots, n\}$ и для любой последовательности $z_1^n = z_1, \dots, z_n \in Z^n$ и любого $z_{n+1} \in Z$ выполнено

$$A(z_1, \dots, z_n, z_{n+1}) = A(z_{\pi(1)}, \dots, z_{\pi(n)}, z_{n+1}).$$

¹ Под ошибкой понимается результат прогноза, когда $y_{n+1} \notin \Gamma_n^\epsilon$.

Интерпретация меры конформности A заключается в том, что значение $A(z_1, \dots, z_n, z_{n+1})$ измеряет, насколько новое наблюдение z_{n+1} “похоже” на ранее полученные наблюдения $(z_1, \dots, z_n)$.

Простейший пример меры конформности для случая регрессии (где $\mathcal{Y} = \mathcal{R}$),

$$A(z_1, \dots, z_n, x_{n+1}) = y_{n+1} - \hat{y}_{n+1}, \quad (1)$$

где $\hat{y}_{n+1}$ – предсказание метки, вычисленное рассматриваемым алгоритмом на основе известных значений $z_1, \dots, z_n, x_{n+1}$, y – тестируемая метка.

Определим

$$\alpha_i^y = A(z_1, \dots, z_{i-1}, z_{i+1}, \dots, (x_{n+1}, y), z_i) \text{ при } i = 1, \dots, n, \quad (2)$$

$$\alpha_{n+1}^y = A(z_1, \dots, z_n, (x_{n+1}, y)). \quad (3)$$

Конформный преобразователь, соответствующий мере конформности A , определяется как

$$C^A(z_1^n, (x_{n+1}, y)) = \frac{1}{n+1} |\{i : 1 \leq i \leq n+1, \alpha_i^y \leq \alpha_{n+1}^y\}|, \quad (4)$$

где посредством $|D|$ обозначается количество элементов конечного множества D . Легко видеть, что при предположении, что все пары $z_i = (x_i, y_i)$ генерируются независимо одним и тем же распределением вероятностей P , выполнено

$$P\{C^A(z_1^n, (x_{n+1}, y)) \leq \epsilon\} \leq \epsilon. \quad (5)$$

Действительно, левая часть неравенства (5) есть вероятность того, что число α_{n+1}^y находится среди подмножества самых больших чисел набора $\alpha_1^y, \dots, \alpha_{n+1}^y$ доли $\leq \epsilon$. Так как все варианты упорядочения этого набора равновероятны, вероятность события (5) не превосходит ϵ .

Наша цель – построить на основе величины (4) функцию распределения, т.е., эта величина должна быть равномерно распределена. Для этого в работе [8] была проведена следующая модификация определения (4). Пусть $\tau \in [0, 1]$ – равномерно распределенная независимая от всех z_i случайная величина.

Рандомизированный конформный преобразователь, соответствующий мере конформности A , определяется как

$$C^A(z_1^n, (x_{n+1}, y), \tau) = \frac{1}{n+1} |\{i : 1 \leq i \leq n+1, \alpha_i^y < \alpha_{n+1}^y\}| + \frac{\tau}{n+1} |\{i : 1 \leq i \leq n+1, \alpha_i^y = \alpha_{n+1}^y\}|, \quad (6)$$

где τ – равномерно распределенная на $[0, 1]$ случайная величина.

Предсказательная система (6) калибруется по вероятности, если

$$P\{C^A(z_1^n, (x_{n+1}, y), \tau) \leq \epsilon\} = \epsilon, \quad (7)$$

где P – комбинированная вероятность по z_i , x_{n+1} и $\tau \sim U$ (равномерное распределение на интервале $[0, 1]$). Все эти предположения выполнены для мер конформности (1).

Свойство (7) используется для построения доверительных множеств прогнозов.² Каждое y , для которого

$$C^A(z_1^n, (x_{n+1}, y), \tau) > \epsilon, \quad (8)$$

рассматривается как значение прогноза с уровнем значимости $1 - \epsilon$: $y \in \Gamma_n^\epsilon$.

² Доверительное множество является интервалом, если функция (6) монотонная по y .

Условие (8) (и определение (6)) означает, что значение счетчика конформности тестовой пары (x, y) не меньше чем значения счетчиков конформности у доли ϵ всех пар обучающей выборки.

В [6] доказано, что величины $p_n = C^A(z_1^n, (x_{n+1}, y), \tau)$ независимо и одинаково распределены по P и U , что, вместе с (7), обеспечивает валидность предсказаний – согласно закону больших чисел с вероятностью 1 доля ошибок при выборе y в пределе не превосходит ϵ .

Расщепленные системы конформного прогнозирования. Нахождение доверительных множеств конформных прогнозов с помощью правила (8) даже в случае конечного множества $\mathcal{Y}$ представляет собой вычислительно трудную задачу. В работах [7], [8], [9] были предложены расщепленные системы конформного прогнозирования, которые значительно более эффективны в вычислительном отношении.³ В этой постановке вся обучающая выборка разбивается на тренировочную $z_1^n = z_1, \dots, z_n$ и калибровочную $\tilde{z}_1^m = \tilde{z}_1, \dots, \tilde{z}_m$.

По тренировочной выборке строится базовый алгоритм, который может выдавать прогнозы $\hat{y} = f_{z_1^n}(x)$ для каждого x .

По калибровочной выборке определяется мера конформности $A(z_1^n, (x, y))$, например в виде (1), и определяются счетчики конформности α_i элементов калибровочной выборки, также счетчик конформности α^y произвольной тестовой пары (x, y) :

$$\alpha_i = A(z_1^n, \tilde{z}_i), \text{ при } i = 1, \dots, m. \quad (9)$$

$$\alpha^y = A(z_1^n, (x, y)). \quad (10)$$

Соответствующий рандомизированный конформный преобразователь определяется как

$$C_{z_1^n, \tilde{z}_1^m, x, \tau}^A(y) = \frac{1}{m+1} |\{i : 1 \leq i \leq m, \alpha_i < \alpha^y\}| + \frac{\tau}{m+1} |\{i : 1 \leq i \leq m, \alpha_i = \alpha^y\}|, \quad (11)$$

где τ – случайная величина, равномерно распределенная на интервале $[0, 1]$.

Расщепленная конформная предсказательная система (11) также калибруется по вероятности в смысле (7), см. [9].

Расщепленная конформная предсказательная система (11) позволяет вместо доверительных множеств прогнозов построить конформное распределение вероятностей прогнозов соответствующей метки. Реализация схемы построения вероятностного расщепленного конформного предсказания в режиме онлайн приведена в разделе 3 далее.

2.2. Предсказания с использованием экспертных стратегий

В предыдущем разделе были приведены методы построения функций вероятностного распределения для каждого эксперта. Возникает задача агрегирования этих функций распределения в одну результирующую функцию распределения.

В этом разделе даются необходимые определения и вспомогательные результаты теории предсказаний с использованием экспертных стратегий. Подробное изложение и доказательства основных утверждений приводятся в работах [21], [15], [22], [23].

Онлайн обучение и прогнозирование. Пусть Ω – множество исходов, Γ – множество прогнозов и $\lambda(f, y)$ – вещественная, принимающая неотрицательные значения, функция потерь, где $f \in \Gamma$ и $y \in \Omega$.⁴ Также задано E – множество экспертов, обозначим $E = \{1, \dots, N\}$.

³ Хотя и менее точны в приложениях, как отмечено в этих работах.

⁴ Природа этих множеств будет уточнена позже.

В теории предсказаний с использованием экспертных стратегий процесс онлайн обучения и прогнозирования представляется в виде игры с полной информацией. На каждом шаге $t = 1, 2, \dots$ каждый эксперт $i \in E$ представляет прогноз $f_{i,t} \in \Gamma$, после этого, предсказатель представляет свой прогноз $f_t \in \Gamma$. После того, как прогнозы были сделаны, объявляется исход $y_t \in \Omega$. Каждый эксперт i несет потери $\lambda(f_{i,t}, y_t)$, предсказатель несет потери $\lambda(f_t, y_t)$. Указанная последовательность действий оформлена ниже в виде Протокола 1. Каждый игрок может использовать любую информацию, которая предшествует его ходу.

Протокол 1

FOR $t = 1, \dots, T$

1. Эксперты представляют прогнозы $f_{i,t}$, где $1 \leq i \leq N$.
2. Предсказатель представляет свой прогноз f_t .
3. Объявляется исход y_t и вычисляются потери $\lambda(f_{i,t}, y_t)$ экспертов и потери $\lambda(f_t, y_t)$ предсказателя.

ENDFOR

Пусть $H_T = \sum_{t=1}^T \lambda(f_t, y_t)$ – суммарные (кумулятивные) потери предсказателя и $L_T^i = \sum_{t=1}^T \lambda(f_{i,t}, y_t)$ – суммарные потери эксперта i за первые T шагов. Разность $R_T^i = H_T - L_T^i$ называется регретом относительно эксперта i , $R_T = H_T - \min_i L_T^i$ – регрет относительно несущего наименьшие потери (наилучшего) эксперта. Цель предсказателя – выдавать предсказания так, чтобы минимизировать регрет.

Агрегирующие алгоритмы АА и WA. Для агрегации прогнозов экспертов будем рассматривать схему, предложенную в работах [17], [18], [19]. Эта схема также использовалась в работах [21], [15] и [22].

Каждый из алгоритмов АА и WA приписывает веса экспертам в зависимости от накопленных ими потерь. На первом шаге используются веса $w_{i,1} = \frac{1}{N}$ для всех i . На последующих шагах веса $w_{i,t}$ вычисляются по правилу:

$$w_{i,t+1} = w_{i,t} e^{-\eta \lambda(f_{i,t}, y_t)} \text{ при } t = 1, 2, \dots, \quad (12)$$

где $\eta > 0$ – параметр обучения. Производится нормализация весов:

$$w_{i,t}^* = \frac{w_{i,t}}{\sum_{j=1}^N w_{j,t}}. \quad (13)$$

Смешиваемость. В алгоритме АА для построения агрегированного прогноза используется так называемая функция суперпредсказания. Согласно [18] функция потерь называется η -смешиваемой, если для любого распределения вероятностей $\mathbf{q} = (q_1, \dots, q_N)$ на множестве E всех экспертов ($\sum_{i=1}^N q_i = 1$ и $q_i \geq 0$ для всех i) и для любых их предсказаний $\mathbf{f} = (f_1, \dots, f_N)$ существует предсказание f такое, что

$$\lambda(f, y) \leq g(y) \text{ для всех } y, \quad (14)$$

где

$$g(y) = -\frac{1}{\eta} \ln \sum_{i=1}^N e^{-\eta \lambda(f_i, y)} q_i \quad (15)$$

– функция суперпредсказания.

Фиксируем некоторое правило для вычисления предсказания f и полагаем

$$f = \text{Subst}(\mathbf{f}, \mathbf{q}). \quad (16)$$

Функция Subst называется функцией подстановки.

В алгоритме АА на каждом шаге t рассматривается функция псевдопредсказания

$$g_t(y) = -\frac{1}{\eta} \ln \sum_{i=1}^N e^{-\eta \lambda(f_{i,t}, y)} w_{i,t}^*, \quad (17)$$

при этом, прогноз предсказателя равен $f_t = \text{Subst}(\mathbf{f}_t, \mathbf{w}_t^*)$, где $\mathbf{f}_t = (f_{1,t}, \dots, f_{N,t})$ – прогнозы экспертов, $\mathbf{w}_t^* = (w_{1,t}^*, \dots, w_{N,t}^*)$ – набор их нормализованных весов.

Оценка регрета для АА. Пусть функция потерь $\lambda(f, y)$ является η -смешиваемой, где $\eta > 0$, $\mathbf{w}_t^* = (w_{1,t}^*, \dots, w_{N,t}^*)$ – нормализованные веса и $\mathbf{f}_t = (f_{1,t}, \dots, f_{N,t})$ – предсказания экспертов на шаге t . Пусть также в Протоколе 1 предсказатель вычисляет свой прогноз $f_t = \text{Subst}(\mathbf{f}_t, \mathbf{w}_t^*)$. Согласно (14), $\lambda(f_t, y_t) \leq g_t(y_t)$ для всех t , где $g_t(y)$ определено по (17).

Пусть $H_T = \sum_{t=1}^T \lambda(f_t, y_t)$ – суммарные потери предсказателя и $L_T^i = \sum_{t=1}^T \lambda(f_{i,t}, y_t)$ – суммарные потери эксперта i . По определению $g_t(y_t) = -\frac{1}{\eta} \ln \frac{W_{t+1}}{W_t}$, где $W_t = \sum_{i=1}^N w_{i,t}$ и $W_1 = 1$. По правилу вычисления весов (12) получим $w_{i,t+1} = \frac{1}{N} e^{-\eta L_t^i}$.

Сокращая одинаковые по модулю и противоположные по знаку слагаемые, получим независимую от времени верхнюю оценку

$$H_T \leq \sum_{t=1}^T g_t(y_t) = -\frac{1}{\eta} \ln W_{T+1} \leq L_T^i + \frac{\ln N}{\eta} \quad (18)$$

для произвольного эксперта i . Таким образом, существует стратегия предсказателя, которая гарантирует верхнюю оценку регрета $R_T \leq \frac{\ln N}{\eta}$ для всех T .

Экспоненциально выпуклые функции потерь и алгоритм WA. Допустим, что множество предсказаний образует линейное пространство, т.е., прогнозы можно складывать и умножать на вещественные числа. Функция потерь $\lambda(f, y)$ называется η -экспоненциально выпуклой, если для каждого y функция $\exp(-\eta \lambda(f, y))$ является вогнутой по f (см. [16]). Для экспоненциально выпуклой функции неравенство $\lambda(f, y) \leq g(y)$ выполнено по определению, если прогноз предсказателя вычисляется как взвешенное среднее прогнозов экспертов $f = \sum_{i=1}^N q_i f_i$, где $\mathbf{q} = (q_1, \dots, q_N)$ – произвольное вероятностное распределение на множестве всех экспертов, $f_1, \dots, f_N$ – предсказания экспертов. Называем правило вычисления прогнозов на шаге t в виде взвешенного среднего

$$f_t = \sum_{i=1}^N w_{i,t}^* f_{i,t}, \quad (19)$$

где веса $w_{i,t}^*$ вычисляются по формулам (12) и (13), алгоритмом WA.

Если в Протоколе 1 вычислять прогноз предсказателя по формуле (19), то также имеет место независимая от времени верхняя оценка регрета (18).

Квадратичная функция потерь $\lambda(\gamma, \omega) = (\gamma - \omega)^2$, где $\omega \in [a, b]$, является η -смешиваемой при $0 < \eta \leq \frac{2}{(b-a)^2}$ (см. [17], [18] и [19]), она также является η -экспоненциально выпуклой при $0 < \eta \leq \frac{1}{2(b-a)^2}$ (см. [16]).

Для квадратичной функции потерь при $\omega \in [0, 1]$ выражение для f можно вычислить по формуле (алгоритм АА):

$$f_t = \text{Subst}(\mathbf{f}, \mathbf{q}) = \frac{1}{2} - \frac{1}{2\eta} \ln \frac{\sum_{i=1}^N q_i e^{-\eta f_i^2}}{\sum_{i=1}^N q_i e^{-\eta (f_i - 1)^2}},$$

где $\mathbf{f} = (f_1, \dots, f_N)$ – прогнозы экспертов, $\mathbf{q} = (q_1, \dots, q_N)$ – произвольное вероятностное распределение (доказательство см. в [21]). При $a = 0$ и $b = 1$ полагаем $\eta = 2$.

2.3. Агрегирование вероятностных прогнозов

Функция потерь CRPS. Пусть в Протоколе 1 множество исходов представляет собой интервал $\Omega = [a, b]$ вещественной прямой, где $a < b$, а множеством предсказаний Γ является множество всех функций распределения вероятностей на этом интервале: $F : [a, b] \rightarrow [0, 1]$.⁵

Оценка качества предсказания F при известном исходе y измеряется с помощью функции непрерывной ранжированной вероятности (функции потерь)

$$\text{CRPS}(F, y) = \int_a^b (F(u) - H(u - y))^2 du, \quad (20)$$

где $H(x)$ – функция Хевисайда: $H(x) = 0$ при $x < 0$ и $H(x) = 1$ при $x \geq 0$. Здесь для простоты изложения мы рассматриваем интегрирование по конечному интервалу. Такое определение ближе к практическим приложениям и проще для теоретического анализа.

Функция потерь CRPS измеряет расхождение между вероятностным прогнозом F и точной функцией распределения $H(u - y)$, которая помещает всю массу вероятности на истинный исход y . Наименьшее значение $\text{CRPS}(F, y)$, равное 0, достигается, когда F сконцентрирована на y , во всех остальных случаях $\text{CRPS}(F, y) > 0$.

Рассмотрим игру на предсказания с использованием экспертных прогнозов для случая, когда прогнозы экспертов и предсказателя представляют собой функции распределения вероятностей.

На каждом шаге t игры каждый эксперт $i \in \{1, \dots, N\}$ представляет прогноз – функцию распределения вероятностей $F_{i,t}(u)$, предсказатель вычисляет свой прогноз $F_t(u)$. После этого, появляется исход $y_t \in [a, b]$ и эксперты и предсказатель несут потери $\text{CRPS}(F_{i,t}, y_t)$ и $\text{CRPS}(F_t, y_t)$.

Цель предсказателя состоит в том, чтобы, независимо от того какие бы исходы не предъявлялись, предсказывать так, чтобы в любой момент времени T его суммарные потери $L_T = \sum_{t=1}^T \text{CRPS}(F_t, y_t)$ были бы не больше чем суммарные потери $L_T^i = \sum_{t=1}^T \text{CRPS}(F_{i,t}, y_t)$ любого эксперта i с точностью до величины порядка $o(T)$.

В работе [21] было доказано, что функция потерь $\text{CRPS}(F, y)$ является η -смешиваемой при $0 < \eta \leq \frac{2}{b-a}$ и η -экспоненциально выпуклой при $0 < \eta \leq \frac{1}{2(b-a)}$ и поэтому к ней применимы оценки регрета из раздела 2.2.

Сформулируем точные утверждения.

⁵ Функция распределения вероятностей есть неубывающая функция $F(y)$ определенная на интервале $[a, b]$ такая, что $F(a) = 0$ и $F(b) = 1$. Она также непрерывная слева и у нее в каждой точке существует предел справа.

Теорема 1. *Функция потерь CRPS(F, y) является η -смешиваемой при $0 < \eta \leq \frac{2}{b-a}$.*

При $\eta = \frac{2}{b-a}$ прогноз предсказателя $F(u)$ вычисляется по прогнозам экспертов $F_i(u)$, где $1 \leq i \leq N$, и по вероятностному распределению $\mathbf{q} = (q_1, \dots, q_N)$ на множестве всех экспертов по правилу АА

$$F(u) = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N q_i e^{-2(F_i(u))^2}}{\sum_{i=1}^N q_i e^{-2(1-F_i(u))^2}}, \quad (21)$$

Легко проверить, что $F(u)$ есть функция распределения вероятностей.

Функция потерь CRPS является η -экспоненциально выпуклой при $0 < \eta \leq \frac{1}{2(b-a)}$. В этом случае, правило вычисления прогноза предсказателя (21) можно заменить на правило WA

$$F_t(u) = \sum_{i=1}^N q_i F_i(u), \quad (22)$$

Полное доказательство этой теоремы см. в [21].

Уточним протокол 1 для игры с вероятностными предсказаниями и функцией потерь CRPS для АА.

Протокол 2

Определим $w_{i,1} = \frac{1}{N}$ при $1 \leq i \leq N$.

FOR $t = 1, \dots, T$

1. Получаем прогнозы экспертов – функции распределения вероятностей $F_{i,t}(u)$, где $1 \leq i \leq N$.
2. Составляем прогноз предсказателя – функцию распределения

$$F_t(u) = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N w_{i,t}^* e^{-2(F_{i,t}(u))^2}}{\sum_{i=1}^N w_{i,t}^* e^{-2(1-F_{i,t}(u))^2}}, \quad (23)$$

где $w_{i,t}^* = \frac{w_{i,t}}{\sum_{j=1}^N w_{j,t}}$.

3. Наблюдаем исход y_t и вычисляем потери экспертов

$$\text{CRPS}(F_{i,t}, y_t) = \int_a^b (F_{i,t}(u) - H(u - y_t))^2 du$$

при $1 \leq i \leq N$ и потери предсказателя

$$\text{CRPS}(F_t, y_t) = \int_a^b (F_t(u) - H(u - y_t))^2 du.$$

4. Переопределяем веса экспертов $1 \leq i \leq N$

$$w_{i,t+1} = w_{i,t} e^{-\frac{2}{b-a} \text{CRPS}(F_{i,t}, y_t)} \quad (24)$$

ENDFOR

В случае, когда применяется алгоритм WA для вычисления прогноза предсказателя, правило (23) можно заменить на

$$F_t(u) = \sum_{i=1}^N w_{i,t}^* F_{i,t}(u), \quad (25)$$

где $w_{i,t}^* = \frac{w_{i,t}}{\sum_{j=1}^N w_{j,t}}$ – нормализованные веса. При $\eta = \frac{1}{2(b-a)}$ соответствующие веса определяются по формуле

$$w_{i,t+1} = w_{i,t} e^{-\frac{1}{2(b-a)} \text{CRPS}(F_{i,t}, y_t)}. \quad (26)$$

Оценка регрета алгоритма АА имеет вид (см. ([22])): Для произвольного i имеет место верхняя оценка

$$\sum_{t=1}^T \text{CRPS}(F_t, y_t) \leq \sum_{t=1}^T \text{CRPS}(F_{i,t}, y_t) + \frac{b-a}{2} \ln N \quad (27)$$

для каждого T .

В случае, когда прогноз предсказателя вычисляется по правилу WA (25), а веса переопределяются по правилу (26), имеет место оценка

$$\sum_{t=1}^T \text{CRPS}(F_t, y_t) \leq \sum_{t=1}^T \text{CRPS}(F_{i,t}, y_t) + 2(b-a) \ln N. \quad (28)$$

Отметим, что оценка регрета (28) в четыре раза хуже, чем оценка (27). Как показывают экспериментальные расчеты раздела 4, кумулятивные потери алгоритма АА несколько меньше чем потери алгоритма WA.

2.4. АА для экспертов с уровнями компетентности

В этом разделе мы несколько расширим постановку задачи предсказания с использованием экспертов. Пусть в каждый момент времени t прогнозы $\mathbf{f}_t = (f_{1,t}, \dots, f_{N,t})$ экспертов снабжаются уровнями компетентности $\mathbf{p}_t = (p_{1,t}, \dots, p_{N,t})$.

Уровень компетентности – это вещественное число из интервала $[0, 1]$. Если уровень компетентности $p_{i,t} = 0$, то соответствующий эксперт “спит” на шаге t , т.е., его прогноз не принимается во внимание агрегирующим алгоритмом. Если $p_{i,t} < 1$, то мы будем использовать его прогноз $f_{i,t}$ только частично с дисконтом.

Уровень компетентности может определяться самим экспертом или каким-либо внешним образом.⁶

Сопоставим каждому уровню компетентности $p_{i,t}$ распределение вероятности $\mathbf{p}_{i,t} = (p_{i,t}, 1 - p_{i,t})$ на двухэлементном множестве. Определим вспомогательный рандомизированный прогноз эксперта i :

$$\tilde{f}_{i,t} = \begin{cases} f_{i,t} & \text{с вероятностью } p_{i,t}, \\ f_t & \text{с вероятностью } 1 - p_{i,t}, \end{cases}$$

где f_t – прогноз предсказателя, который будет вычислен позже.

Наша цель – определить прогноз f_t так, чтобы

$$e^{-\eta \lambda(f_t, y)} \geq \sum_{i=1}^N E_{\mathbf{p}_{i,t}}[e^{-\eta \lambda(\tilde{f}_{i,t}, y)}] w_{i,t} \quad (29)$$

для каждого y . Здесь $E_{\mathbf{p}_{i,t}}$ – математическое ожидание относительно распределения вероятности $\mathbf{p}_{i,t}$. Также $w_{i,t}$ – вес эксперта i на шаге t .

⁶ Данная постановка была использована в работе [20].

Запишем неравенство (29) в более детальной форме:

$$e^{-\eta\lambda(f_t, y)} \geq \sum_{i=1}^N E_{\mathbf{p}_{i,t}}[e^{-\eta\lambda(\tilde{f}_{i,t}, y)}] w_{i,t} = \quad (30)$$

$$\sum_{i=1}^N p_{i,t} w_{i,t} e^{-\eta\lambda(f_{i,t}, y)} + e^{-\eta\lambda(f_t, y)} \left(1 - \sum_{i=1}^N p_{i,t} w_{i,t} \right) \quad (31)$$

для всех ω . Таким образом, неравенство (29) эквивалентно неравенству

$$e^{-\eta\lambda(f_t, y)} \geq \sum_{i=1}^N w_{i,t}^* e^{-\eta\lambda(f_{i,t}, y)}, \quad (32)$$

где

$$w_{i,t}^* = \frac{p_{i,t} w_{i,t}}{\sum_{j=1}^N p_{j,t} w_{j,t}}. \quad (33)$$

Используем правило (16) для вычисления предсказания алгоритма АА: $f_t = \text{Subst}(\mathbf{f}_t, \mathbf{w}_t^*)$. Тогда (32) эквивалентно (30). Здесь Subst – функция подстановки, $\mathbf{w}_t^* = (w_{i,1}^*, \dots, w_{i,N}^*)$ и $\mathbf{f}_t = (f_{1,t}, \dots, f_{i,N})$.

Пусть $h_t = \lambda(f_t, y_t)$ – потери предсказателя на шаге t , $\tilde{l}_{i,t} = \lambda(\tilde{f}_{i,t}, y_t)$ – потери рандомизированного эксперта i , $\hat{l}_{i,t} = E_{\mathbf{p}_{i,t}}[\tilde{l}_{i,t}]$ – математическое ожидание потерь рандомизированного эксперта. Под виртуальным экспертом понимаем эксперта, который несет потери $\hat{l}_{i,t}$.

Модифицируем Протокол 1 с уровнями компетентности.

Протокол 1a

FOR $t = 1, \dots, T$

1. Получаем предсказания $f_{i,t}$ экспертов и уровни компетентности $p_{i,t}$, где $1 \leq i \leq N$.
2. Вычисляем прогноз предсказателя $f_t = \text{Subst}(\mathbf{f}_t, \mathbf{w}_t^*)$, где $\mathbf{w}_t^* = (w_{1,t}^*, \dots, w_{N,t}^*)$ – нормализованные веса, определенные по формуле (33).
3. Получаем истинное значение исхода y_t и вычисляем потери $l_{i,t} = \lambda(f_{i,t}, y_t)$ экспертов и потери $\lambda(f_t, y_t)$ предсказателя.
4. Перестраиваем веса (виртуальных) экспертов:

$$w_{i,t+1} = w_{i,t} e^{-\eta(p_{i,t}\lambda(f_{i,t}, y_t) + (1-p_{i,t})\lambda(f_t, y_t))} \quad (34)$$

ENDFOR

Так как по определению потери виртуального эксперта равны $\hat{l}_{i,t} = p_{i,t} l_{i,t} + (1 - p_{i,t}) h_t$, будет $h_t - \hat{l}_{i,t} = p_{i,t} (h_t - l_{i,t})$. Назовем эту величину дисконтированным регретом относительно эксперта i на шаге t . Мы будем измерять качество нашего алгоритма с помощью суммарного дисконтированного регрета относительно эксперта i .

Теорема 2. Для любого $1 \leq i \leq N$ имеет место верхняя оценка суммарного дисконтированного регрета относительно эксперта i :

$$\sum_{t=1}^T p_{i,t} (h_t - l_{i,t}) \leq \frac{\ln N}{\eta}. \quad (35)$$

Доказательство. Из выпуклости экспоненты и неравенства (29) получаем

$$e^{-\eta\lambda(f_t, y)} \geq \sum_{i=1}^N e^{-\eta E_{\mathbf{P}_{i,t}}[\lambda(\tilde{f}_{i,t}, y)]} w_{i,t}^* = \sum_{i=1}^N e^{-\eta \hat{l}_{i,t}} w_{i,t}^*. \quad (36)$$

Пусть $m_t = -\frac{1}{\eta} \ln \sum_{i=1}^N w_{i,t}^* e^{-\eta \hat{l}_{i,t}}$. По (36) $h_t \leq m_t$. Перепишем правило (34) в виде

$$w_{i,t+1} = w_{i,t} e^{-\eta \hat{l}_{i,t}}. \quad (37)$$

Напомним, что $W_T = \sum_{t=1}^T w_{i,t}$, $W_1 = 1$ и $m_t = \frac{1}{\eta} \ln \frac{W_{t+1}}{W_t}$. Также как в (18), используя (36) и (37), получим

$$\sum_{t=1}^T h_t \leq \sum_{t=1}^T m_t = -\frac{1}{\eta} \ln W_{T+1} \leq \sum_{t=1}^T \hat{l}_{i,t} + \frac{\ln N}{\eta}$$

для любого i . Так как $h_t - \hat{l}_{i,t} = p_{i,t}(h_t - l_{i,t})$, получаем неравенство (35). $\triangle$

3. ПОСТРОЕНИЕ КОНФОРМНЫХ РАСПРЕДЕЛЕНИЙ В РЕЖИМЕ ОНЛАЙН

При проведении экспериментов для вычисления прогнозов экспертов использовались конформные прогнозирующие системы. Агрегация прогнозов производилась с помощью алгоритмов АА и WA. Проводилось сравнение результатов с результатами метода прогнозирования на основе гауссовых смесей GMM.

Построение регрессионного алгоритма.

Для построения экспертных стратегий весь массив исторических данных, состоящий из пар (x, y) , где x – температура, y – нагрузка на сеть, сначала делится на попарно непересекающиеся интервалы временных отрезков (сезон, время дня), которые также будут называться областями компетентности соответствующих экспертов. В каждой области компетентности определяется обучающая выборка, которая делится на тренировочную $z_1^m = z_1, \dots, z_m$ и калибровочную $\tilde{z}_1^m = \tilde{z}_1, \dots, \tilde{z}_1^k$ выборки. Каждая пара (x, y) соответствует некоторому моменту времени.

Каждый эксперт обучается на своей тренировочной выборке z_1^m ⁷ и будет калиброваться в дальнейшем на элементах выборки $\tilde{z}_1^m$ или на ее расширении, которое будет определяться в режиме онлайн. ⁸

Подчеркнем, что каждый эксперт обучается и калибруется на данных из своей области компетентности.

Онлайн калибровка предсказаний эксперта. Процедура конформной калибровки будет строить систему расщепленного конформного прогнозирования. В каждый очередной момент времени t получаем тестируемый объект x_t (в экспериментах – это значение температуры и дополнительная информация в виде вектора), для которого мы должны построить функцию распределения величины потребления с использованием калибровочной выборки.

В каждый момент времени имеется калибровочная выборка $\tilde{z}_1^m = (\tilde{x}_1, \tilde{y}_1), \dots, \tilde{z}_m = (\tilde{x}_m, \tilde{y}_m)$, элементы которой принадлежат области компетентности эксперта. ⁹

⁷ По этой выборке строится уравнение регрессии, которое позволяет вычислять точечные предсказания.

⁸ По этой выборке на основании точечных предсказаний регрессионного алгоритма в результате онлайн калибровки будет строиться конформная предсказательная система, которая по x выдает функцию распределения вероятностей y .

⁹ Более простой вариант, когда в качестве калибровочной выделяется часть обучающей выборки, которая служит в качестве калибровочной на всех последующих шагах.

При этом сохраняется основное предположение о том, что элементы калибровочной выборки $\tilde{z}_1^m$ и пара (x_t, y_t) должны быть независимо и одинаково распределены относительно некоторого вероятностного распределения на парах.¹⁰

Меры конформности элементов калибровочной выборки определяются согласно определениям раздела 2.1.

Онлайн агрегирование экспертных конформных функций распределения. В разделе 4 (далее) будут представлены результаты численных экспериментов с реальными данными. Приведем модификацию протокола 2, которая использовалась в этих экспериментах.

Протокол 2а

FOR $i = 1, \dots, N$ *Предварительная обработка

Используя обучающую выборку, точнее, ее тренировочную часть $z_1^n = z_1, \dots, z_n$ из области компетентности эксперта i , строим алгоритм регрессии $y = f_i(x)$.

ENDFOR *Конец цикла предварительной обработки

Определим $w_{i,1} = \frac{1}{N}$ при $1 \leq i \leq N$.

FOR $t = 1, \dots, T$ *Главный цикл

1. Получаем тестируемый объект x_t и вычисляем вероятностные прогнозы экспертов – функции распределения $F_{i,t}(y)$ для $i = 1, \dots, N$.

FOR $i = 1, \dots, N$ *Построение конформных распределений экспертов.

Формируем калибровочную выборку $\tilde{z}_1^m = \tilde{z}_1, \dots, \tilde{z}_m$ – из области компетентности соответствующего эксперта.¹¹

Используем меру конформности (1) для случая регрессии:

$$A(z_1^n, (x, y)) = y - \hat{y},$$

где $\hat{y} = f_i(x)$ – предсказание метки y , вычисленное алгоритмом f_i для объекта x .

Вычисляем значения счетчиков конформности α_s , при $s = 1, \dots, m$: $\alpha_s = A(z_1^n, (\tilde{x}_s, \tilde{y}_s))$ и располагаем эти числа в возрастающем порядке:

$$\alpha_{(1)} < \dots < \alpha_{(k)}.$$

Пусть $n_j = |\{s : \alpha_s = \alpha_{(j)}\}|$ при $j = 1, \dots, k$.

$m_j = \sup\{y : \alpha^y < \alpha_{(j)}\}$ и $M_j = \inf\{y : \alpha^y > \alpha_{(j)}\}$, где $\alpha^y = A(z_1^n, (x, y))$.

Определим функцию конформного предсказательного распределения:

$$Q_{z_1^n, \tilde{z}_1^m, x_t, \tau}(y) = \begin{cases} \frac{\tau}{m+1} & \text{если } y < m_1, \\ \frac{n_1 + \dots + n_{j-1} + \tau n_j + \tau}{m+1} & \text{если } m_j < y < M_j, j = 1, \dots, k, \\ \frac{n_1 + \dots + n_j + \tau}{m+1} & \text{если } M_j < y < m_{j+1}, j = 1, \dots, k-1, \\ \frac{n_1 + \dots + n_k + \tau}{m+1} = \frac{m+\tau}{m+1} & \text{если } y > M_k. \end{cases}$$

В работе [9] доказано, что при некоторых дополнительных предположениях на меру конформности функция Q является функцией распределения вероятностей.

Полагаем $F_{i,t}(y) = Q_{z_1^n, \tilde{z}_1^m, x_t, \tau}(y)$ – функция конформного распределения вероятностей для эксперта i . Оценку расхождения между функцией конформного распределения и исходом y_{n+1} будем измерять с помощью функции CRPS.

ENDFOR *Конец цикла построения конформных распределений экспертов

¹⁰ Конкретная форма распределения при этом не принимается во внимание.

¹¹ Метод CP на каждом шаге использует калибровочную выборку, определенную на начальном этапе обучения. Метод CP+ использует начальную калибровочную выборку, пополненную новыми парами из области компетентности эксперта, которые наблюдаются на текущем шаге t .

2. Агрегирование распределений вероятностей экспертов.

Получаем уровни компетентности $p_{i,t}$, где $1 \leq i \leq N$. Задаем агрегированный прогноз предсказателя – функцию распределения $F_t(y)$, которая определена по правилу

$$F_t(y) = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N w_{i,t}^p e^{-2(F_{i,t}(y))^2}}{\sum_{i=1}^N w_{i,t}^p e^{-2(1-F_{i,t}(y))^2}} \quad (38)$$

для алгоритма АА или по правилу

$$F_t(y) = \sum_{i=1}^N w_{i,t}^p F_{i,t}(y) \quad (39)$$

для алгоритма WA, где

$$w_{i,t}^p = \frac{p_{i,t} w_{i,t}}{\sum_{j=1}^N p_{j,t} w_{j,t}}.$$

3. Наблюдаем исход y_t и вычисляем потери $\text{CRPS}(F_{i,t}, y_t)$ экспертов $1 \leq i \leq N$, а также потери $\text{CRPS}(F_t, y_t)$ предсказателя.

4. Определяем веса (виртуальных) экспертов $1 \leq i \leq N$

$$w_{i,t+1} = w_{i,t} e^{-\eta(p_{i,t} \text{CRPS}(F_{i,t}, y_t) + (1-p_{i,t}) \text{CRPS}(F_t, y_t))}, \quad (40)$$

где $\eta = \frac{2}{b-a}$ для АА и $\eta = \frac{1}{2(b-a)}$ для WA.

ENDFOR *Конец главного цикла

Оценка регрета данного алгоритма дается неравенством (35) теоремы 2, где $h_t = \text{CRPS}(F_t, y_t)$, $l_{i,t} = \text{CRPS}(F_{i,t}, y_t)$ и $\eta = \frac{2}{b-a}$, если для вычисления прогноза предсказателя было использовано правило (38), и $\eta = \frac{1}{2(b-a)}$, если было использовано правило (39).

Также использовалась модификация Fixed Share (см. [24]) правил вычисления весов, при которой правило (40) заменено на двух-уровневую схему

$$w_{i,t}^\mu = \frac{w_{i,t} e^{-\eta(p_{i,t} \text{CRPS}(F_{i,t}, y_t) + (1-p_{i,t}) \text{CRPS}(F_t, y_t))}}{\sum_{j=1}^N w_{j,t} e^{-\eta(p_{j,t} \text{CRPS}(F_{j,t}, y_t) + (1-p_{j,t}) \text{CRPS}(F_t, y_t))}},$$

$$w_{i,t+1} = \frac{\alpha}{N} + (1 - \alpha) w_{i,t}^\mu,$$

где $0 < \alpha < 1$. То же самое делаем для правила (26). Полагаем $\alpha = 0.001$ в экспериментах.¹²

4. ПРОГНОЗИРОВАНИЕ РАСПРЕДЕЛЕНИЙ ПОЧАСОВЫХ ЭЛЕКТРИЧЕСКИХ НАГРУЗОК

Материалом для экспериментального исследования послужили данные конкурса Global Energy Forecasting – Competition 2014 (GEFCOM 2014, Track Load), который проводился на платформе Kaggle ([2]). Целью GEFCOM2014-L была оценка квантилей (более точно, всех 100 процентилей) для вероятностного распределения почасовых электрических нагрузок. При этом горизонт прогноза изменялся в широком диапазоне от одного часа до одного месяца. Основной блок обучающей выборки включает данные о почасовых электрических нагрузках за период с

¹² Детали подбора параметра α приведены в работе [23].

января 2005 г. до конца 2010 г. и данные о почасовых температурных измерениях на 25 метеорологических станциях с января 2001 до сентября 2010 г. (за 117 месяцев). Тестовая выборка включает данные за период с 1 января 2011 г. до декабря 2011 г. Базы данных доступны на сайте <https://www.kaggle.com/datasets>

В данной работе мы ограничимся только данными усредненной по метеостанциям текущей температуры и ее предыстории и календарными показателями (сезон года и время суток). Предварительный анализ и содержательная визуализация данных обучающей выборки позволяют выбрать и обосновать методы вероятностного прогнозирования, уместные при решении данной задачи.

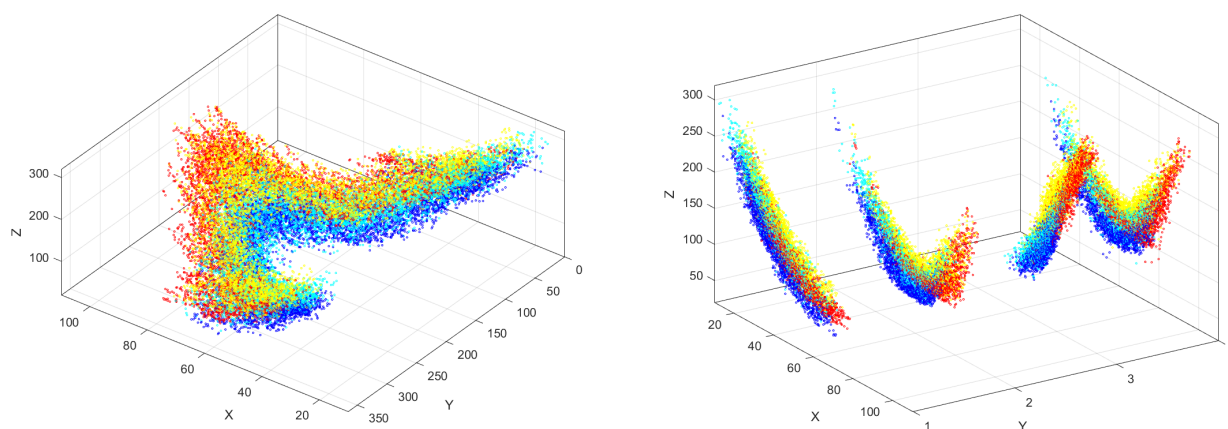


Рис. 1. 3D диаграммы рассеяния “Температура–Электрическая нагрузка” отражающие взаимосвязь между почасовыми измерениями температуры воздуха и электрической нагрузкой в сети.

На рисунке 1 представлены в двух ракурсах 3D диаграммы рассеяния, которые позволяют обосновать выбор используемых в работе экспертов. На левом рисунке показано облако рассеяния точек “Температура”, “Нагрузка”, “День года”. Приведены данные за 5 лет, представленные в обучающей выборке. Цветом показаны периоды суток (ночь, утро, день, вечер). По оси X отложены значения температуры, а по оси Y – все дни года.

На правом рисунке показаны те же данные, но точки диаграммы рассеяния из левого рисунка, относящиеся к одному и тому же сезону года (зима, весна, лето, осень) “схлопнуты” в плоские диаграммы рассеяния, отличающиеся, как можно заметить, по форме облаков, что может служить основанием для выбора сезонных экспертов. Наконец, в каждом из 4-х сезонов на правом рисунке видно разделение по времени суток, что оправдывает выбор соответствующих экспертов типа сезон–период суток.

4.1. Технология проведения экспериментов

В этом разделе мы представляем методологию, которая была использована для вероятностного прогнозирования электрической нагрузки на электрическую сеть.

Области компетентности экспертов. С учетом вида 3D-диаграмм рассеяния (см. рис. 1), обучающая выборка была разбита на подвыборки, соответствующие временам года (зима, весна, лето, осень) и времени дня (утро, день, вечер, ночь).

Область компетентности каждого эксперта i в момент времени t определяется значениями его коэффициентов (уровней) компетентности $p_{i,t}$. На этапе обучения эксперта i в тренировочную выборку включаются только те точки t , где его уровень компетентности $p_{i,t} = 1$.

При прогнозировании область применения эксперта распространяется на все моменты времени, где его уровень компетентности больше 0. Таким образом, каждый эксперт конкурирует с другими экспертами, работающими на перекрывающихся интервалах. При этом используется соответствующий алгоритм АА (или WA) агрегирования экспертов с учетом уровней их компетентности.

Построение и обучение экспертов.

С учетом анализа диаграмм, рассеяния, приведенных на рис. 1, полная обучающая выборка была разбита на подвыборки, соответствующие четырем временам года (зима, весна, лето, осень) и четырем интервалам времени суток (утро, день, вечер, ночь). Каждой из подвыборок был поставлен в соответствие специализированный эксперт, который формировал свою вероятностную прогностическую модель (методом GMM или CP) на основе этих данных.

В общий список экспертов был включен постоянно работающий эксперт, который был обучен на основе всех данных обучающей выборки. Эксперты 2–5 были обучены на данных четырех сезонов, представленных на рис. 1. Эксперты 6–21 соответствуют более детальным разбиениям типа сезон года–период суток.

Произведено сравнение двух подходов к построению вероятностных прогностических моделей экспертов:

1. GMM – аппроксимация двумерного облака точек с помощью нескольких гауссовских компонент как это было сделано в работах [15] и [22].

2. CP (CP+) – конформное прогнозирование на основе точечных прогнозов полиномиальной регрессии и последующем использовании (постоянной – CP, или пополняемой – CP+) калибровочной выборки для построения прогнозного вероятностного распределения.

На тестовой выборке применение специализированных экспертов регулировалось их уровнями компетентности $p_{i,t}$. Эксперт i имеет уровень компетентности 1, т.е. $p_{i,t} = 1$ в пределах соответствующей подвыборки (области его компетентности). Так как при переходе к следующей календарной подвыборке, этот эксперт теряет свою прогностическую способность постепенно, уровень доверия $p_{i,t}$ к его прогнозам вне его области компетентности понижался линейным образом от 1 до 0 в расширенной части интервала (см. рис. 2). Поэтому в окрестностях границ соответствующих подвыборок появляются конкурирующие между собой эксперты с различными ненулевыми уровнями компетентности. Агрегирующий алгоритм (протокол 2а) в каждый момент времени объединяет прогнозы экспертов (распределения вероятностей) с учетом этих уровней компетентности.¹³

Как видно из диаграмм рассеяния на рис. 1, сезон года и время суток проявляются в виде заметных регулярных изменений общего вида облаков рассеяния и их взаимного расположения.

Результаты работы систем конформного прогнозирования (CP и CP+) сравниваются с результатами применения метода построения вероятностных распределений по эмпирическим данным – метода гауссовых смесей (GMM), см. работы [15], [22].

Метод гауссовских смесей GMM. Метод гауссовских смесей на этапе обучения строит распределение вероятностей для каждой подвыборки на основе диаграммы рассеяния пар “температура – нагрузка”. Функции распределения, выдаваемые методом GMM для каждого эксперта, агрегируются с помощью алгоритмов АА и WA (Протокол 2а) в функцию распределения результирующего прогноза.

Системы конформного прогнозирования. Те же обучающие выборки были использованы для построения и оценки методов конформных распределений CP (CP+). Для каждого

¹³ Как показано в работах [15], [22], использование расширенных областей компетентности приводит к более точным прогнозам и соответствующему уменьшению кумулятивных потерь агрегирующих алгоритмов.

эксперта на начальном этапе обучения область компетентности эксперта разделяется на две части – тренировочную и калибровочную. Сначала по тренировочной выборке строится полиномиальное регрессионное приближение – для этого использовались кубические полиномы. Примеры точечного прогнозирования электрических нагрузок можно найти в работе [20]. Калибровочная выборка используется на последующих шагах для построения прогнозного вероятностного распределения.

В режиме онлайн производятся следующие действия:

На каждом шаге периода прогнозирования на основе текущих прогнозов температуры и калибровочной выборки, определенной на начальном этапе обучения, строится распределение вероятностей величины нагрузки.¹⁴ Этот метод обозначается как СР.

При методе СР+ первоначальная калибровочная выборка на каждом шаге пополняется новыми парами из наблюдаемой части области компетентности эксперта.¹⁵

Агрегирование функций распределения вероятностей экспертов. Функции распределения, выдаваемые конформными предсказательными системами для каждого эксперта, агрегируются с помощью алгоритмов АА и WA (Протокол 2а) в функцию распределения результирующего прогноза.

4.2. Результаты экспериментов

Для иллюстрации работы предложенных в статье методов построения конформных прогнозирующих систем и агрегирования соответствующих вероятностных прогнозов экспертов-специалистов была выбрана частная задача, а именно, прогнозирование функции распределения электрических нагрузок на один час вперед на основании прогноза температуры и текущих календарных параметров.¹⁶

Пример задания уровней компетентности экспертов приведен на рисунке 2.

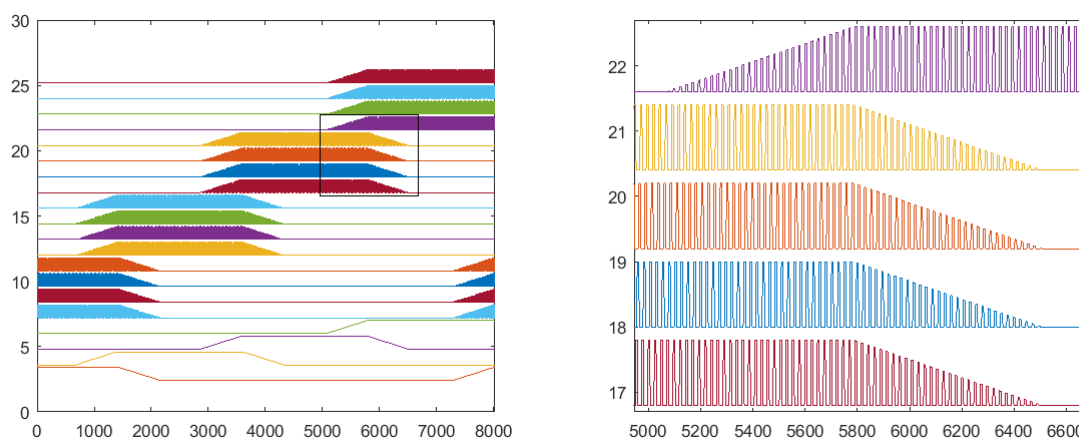


Рис. 2. Левая часть рисунка: уровни компетентности для Экспертов 2-5 (сезонные эксперты) и 6-21 (Сезон, Время дня). Правая часть: увеличенный фрагмент рисунка из левой части.

¹⁴ Здесь используется предположение, что распределение вероятностей, генерирующее данные, наблюдаемые в любом подмножестве области компетентности эксперта является одним и тем же.

¹⁵ Таким образом учитываются небольшие изменения параметров генерирующего распределения.

¹⁶ Заметим, что приведенная технология и соответствующие алгоритмы позволяют вычислять прогнозы нагрузки в любой будущий момент времени, для которого имеется прогноз температуры.

На рисунке 3 показаны графики средних (по времени) значений накопленных CRPS-потерь: $t \rightarrow \frac{1}{t} \sum_{s=1}^t \text{CRPS}(F_{i,s}, y_s)$, для экспертов 1–21, построенных методом СР и потери двух агрегаторов WA и AA, которые применяются без учета уровней компетентности экспертов ($p_{i,t} = 1$ для всех i и t). На всем протяжении тестового периода потери агрегатора AA несколько меньше, чем потери агрегатора WA.

В этом эксперименте эксперты, которые строились по своим областям компетентности, производят прогнозы во всех точках тестовой выборки, а алгоритмы AA и WA агрегируют эти прогнозы в одном случае без учета уровней их компетентности, а во втором случае – с учетом уровней компетентности (см. рис. 3).

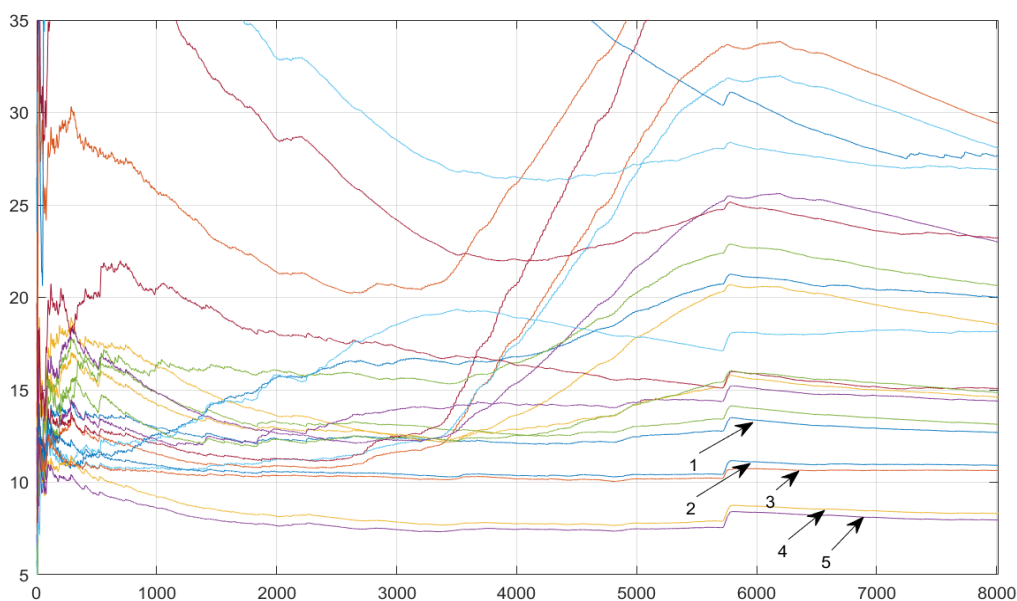


Рис. 3. Средние значения накопленных CRPS-потерь экспертов, построенных методом СР, и потери агрегаторов WA и AA этих экспертов, применяемых без учета их уровней компетентности (помечены цифрами 2 и 3). Для сравнения на том же рисунке показаны потери агрегаторов WA и AA, построенных с учетом уровней компетентности экспертов (цифры 4 и 5). Цифрой 1 помечена кривая потеря эксперта, работающего постоянно.

На рисунке 4 приведены средние (по времени) значения накопленных CRPS-потерь агрегаторов AA и WA для экспертов, построенных методом GMM, СР и СР+, применяемых с учетом уровней компетентности экспертов. На рисунке наблюдается скачок всех графиков потерь, который произошел в результате блокаута, который произошел в окрестности 5700-го часа.

Алгоритм AA при агрегировании экспертов, построенных методами СР и СР+, приводит к меньшим потерям чем алгоритм AA, примененный к экспертам, построенным методом GMM, причем применение метода СР+ для построения экспертов приводит к меньшим потерям чем применение метода СР.

По-прежнему, для всех методов построения экспертов на всем протяжении тестового периода потери агрегатора AA несколько меньше, чем потери агрегатора WA.

Кривые дисконтированных регретов алгоритма AA относительно экспертов, построенных методом СР, приведены на рис. 5. Эти кривые показывают преимущество агрегирующего ал-

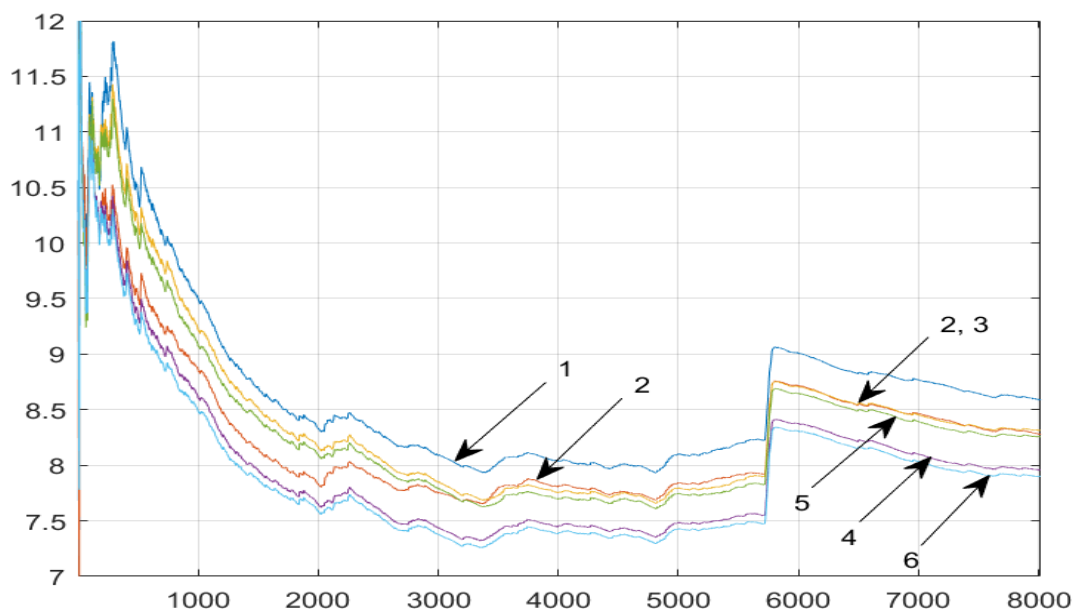


Рис. 4. Средние значения накопленных CRPS-потерь агрегаторов AA и WA для экспертов, построенных методами GMM, CP и CP+. Цифрами помечены кривые потерь агрегаторов 1-WA((GMM), 2-AA((GMM), 3-WA(CP), 4-AA(CP), 5-WA(CP+), 6-AA(CP+), применяемых с учетом уровней компетентности экспертов.

горитма (имеет меньшие потери) по сравнению с каждым из экспертов, работающих в своей области компетентности.

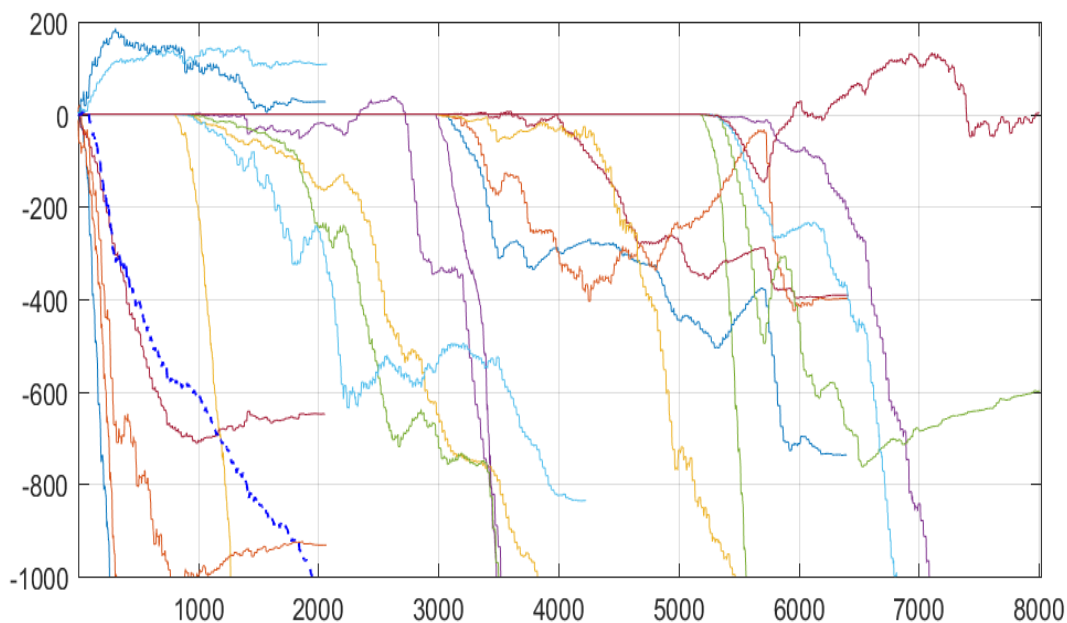


Рис. 5. Дисконтированные регреты алгоритма AA (учитывающего уровни компетентности экспертов) относительно экспертов, построенных методом CP.

3D-изображение прогнозных плотностей распределения, полученных алгоритмом АА при агрегации экспертов, построенных методом СР, приведено на рис.6.

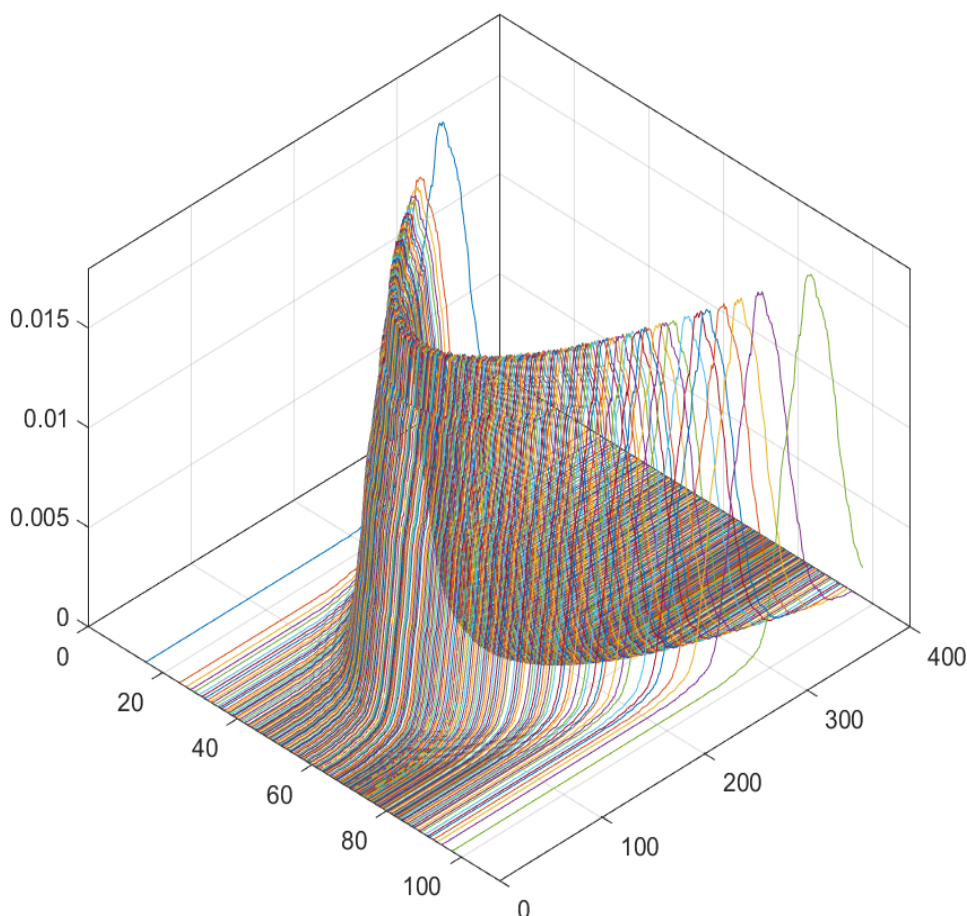


Рис. 6. Плотности распределений, построенных алгоритмом АА при агрегации экспертов, построенных методом СР с учетом их уровней компетентности.

Аналогичные графики потерь, регретов и плотностей распределения для экспертов, построенных методом GMM, приведены в работе [22].

Сравнение кривых потерь, приведенных рис.4, показывает, что агрегатор АА, примененный к экспертам, построенным методами СР и СР+, в среднем имеет меньшие потери, чем он имеет при использовании экспертов, построенных методом GMM, причем метод СР+ приводит к меньшим потерям чем метод СР.

5. ЗАКЛЮЧЕНИЕ

В данной работе представлена компьютерная технология вероятностного прогнозирования нагрузок в электрической сети, которая использует методы построения конформных вероятностных прогнозов (СР и СР+) и методы агрегирования этих прогнозов в режиме онлайн. Проведены экспериментальные расчеты на реальных данных. Производилось сравнение результа-

тов прогнозирования методами CP и CP+ с результатами метода построения вероятностных прогнозов на основе гауссовских смесей GMM. Для оценки качества вероятностных прогнозов использовалось скоринговое правило непрерывной ранжированной вероятности CRPS.

В работе предложен алгоритм для прогнозирования нагрузок электрической сети в режиме онлайн с использованием систем конформного прогнозирования CP (CP+). Получены гарантийные оценки его эффективности. Проведен сравнительный анализ эффективности методов конформных предсказаний CP и CP+ и ранее использовавшегося метода гауссовских смесей GMM для построения функций распределения вероятностей экспертных стратегий.

Метод прогнозирования GMM для построения эксперта использует аппроксимацию двумерного облака точек “температура – нагрузка” с помощью нескольких гауссовских компонент. Система прогнозирования на основе метода GMM представляет собой распределение, аппроксимирующее это облако точек.

Системы конформного прогнозирования CP и CP+ строят прогнозное распределение вероятностей только на основе соотношения между текущим и прошлыми точечными прогнозами регрессионного алгоритма.

Результаты расчетов показывают, что применение методов CP и CP+ для формирования экспертных прогнозов приводит к меньшим потерям агрегирующих алгоритмов чем метод GMM, причем метод CP+ приводит к заметно меньшим потерям чем метод CP.

Также показано, что агрегатор AA имеет меньшие средние потери чем агрегатор WA на всем тестируемом периоде, что согласуется с оценками регрета этих алгоритмов.

СПИСОК ЛИТЕРАТУРЫ

1. M. Devaine, P. Gaillard, Y. Goude, G. Stoltz. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*. 90(2): 231–260, 2013.
2. Tao Hong, Pierre Pinson, Shu Fanc, Hamidreza Zareipour, Alberto Troccoli, Rob J. Hyndman. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting* 32: 896–913, 2016.
3. Vovk, V., Shen, J., Manokhin, V., Xie, M.g., 2019. Nonparametric predictive distributions based on conformal prediction. *Machine Learning*. 108, 445–474.
4. Jieli Shen, Regina Liu, and Minge Xie. Prediction with confidence – a general framework for predictive inference. *Journal of Statistical Planning and Inference*, 195:126–140, 2018.
5. Vovk, V., Gammerman, A., Shafer, G., 2005. Algorithmic learning in a random world. Springer Science and Business Media.
6. V. Vovk, I. Nourtdinov, V. Manokhin, A. Gammerman. Conformal predictive distributions with kernels, On-line Compression Modelling project (New Series), <http://alrw.net>, Working Paper 20, January 2019 (first posted October 2017).
7. V. Vovk, I. Nourtdinov, V. Manokhin, A. Gammerman, Conformal predictive distributions with kernels, in: Braverman Readings in Machine Learning. Key Ideas from Inception to Current State. Springer. 103–121.
8. V. Vovk, I. Nourtdinov, V. Manokhin, A. Gammerman. Computationally efficient versions of conformal predictive distributions. *Neurocomputing*. 397:292–308 (2020).
9. V. Vovk, I. Petej, P. Toccaceli, A. Gammerman, E. Ahlberg, L. Carlsson, Conformal calibration, in: Conformal and Probabilistic Prediction and Applications, *Proceedings of Machine Learning Research*. 84–99.
10. V. Vovk, I. Nourtdinov, V. Manokhin, A. Gammerman. Cross conformal predictive distributions. *Proceedings of Machine Learning Research*, COPA 91:37–51, 2018.

11. A. Jordan, F. Krüger, S. Lerch. Evaluating Probabilistic Forecasts with Scoring Rules, arXiv:1709.04743.
12. J. Bröcker, L.A. Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting* 22. 382–388, 2007.
13. J. Bröcker, L.A. Smith. From ensemble forecasts to predictive distribution functions. *Tellus A* 60. 663–678, 2008.
14. J. Bröcker. Evaluating raw ensembles with the continuous ranked probability score. *Q. J. R. Meteorol. Soc.* 138, 1611–1617, July 2012
15. V. V'yugin, V. Trunov. Online Learning with Continuous Ranked Probability Score, *Proceedings of Machine Learning Research* 105: 163–177, 2019.
16. N. Cesa-Bianchi, G. Lugosi. Prediction, Learning, and Games. Cambridge University Press, 2006.
17. V. Vovk, Aggregating strategies. In M. Fulk and J. Case, editors, *Proceedings of the 3rd Annual Workshop on Computational Learning Theory*, 371–383. San Mateo, CA, Morgan Kaufmann, 1990.
18. V. Vovk, A game of prediction with expert advice. *Journal of Computer and System Sciences*. 56(2), 153–173, 1998.
19. V. Vovk. Competitive on-line statistics. *International Statistical Review* 69, 213–248, 2001.
20. V. V'yugin, V. Trunov. Online aggregation of unbounded losses using shifting experts with confidence. *Machine Learning*, 108(3): 425–444, 2019.
21. В.В. Вьюгин, В.Г. Трунов. Онлайн агрегирование вероятностных прогнозов на основе скорингового правила непрерывной ранжированной вероятности. *Информационные процессы*, 19 (2), 92–112, 2019.
22. В.В. Вьюгин, В.Г. Трунов. Онлайн агрегирование вероятностных прогнозов почасовых электрических нагрузок. *Информационные процессы*, 20, № 2, 2020, стр. 171–192
23. Vladimir V'yugin, Vladimir Trunov. Online aggregation of probability forecasts with confidence. *Pattern Recognition* 121, January 2022, 108193.
24. M. Herbster, M. Warmuth. Tracking the best expert. *Machine Learning*, 32(2): 151–178, 1998.

Online Aggregation of Conformal Forecasting Systems

V.V. V'yugin, V.G. Trunov

The problem of online probabilistic time series forecasting is considered. Probabilistic forecasts are obtained as a result of the application of conformal forecasting systems. The conformal forecasting system is defined on the basis of point forecasts of the regression algorithm. The corresponding probability distribution function is used to assess the degree of reliability of the algorithm's predictions.

The paper considers the case when at each moment of time several competing methods (experts) present their forecasts in the form of distribution functions. These distributions are constructed online from point predictions using the conformal prediction method. The probabilistic forecasts of experts are combined using an aggregating algorithm into one probabilistic forecast at each stage of the forecasting process, while expert forecasts can be used at a discount.

A technology has been developed for constructing predictive expert algorithms and aggregation of their probabilistic forecasts on the example of the problem of forecasting the hourly load of an electrical network depending on the temperature forecast. The results of numerical experiments on real data are presented, a comparative analysis of the method of conformal predictions and the method of Gaussian mixtures is carried out.

KEYWORDS: machine learning, online probabilistic forecasting, conformal forecasting systems, forecasts calibration, prediction with expert advice, regret, confidence levels, aggregating algorithm, continuous ranked probability score, CRPS forecasting loads in an electrical network