

Онлайн агрегирование вероятностных прогнозов почасовых электрических нагрузок

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 01.06.2020

Аннотация—Рассматривается задача вероятностного прогнозирования временного ряда в режиме онлайн. Различие между вероятностным прогнозом (функцией распределения) и числовым исходом измеряется с помощью функции потерь (скорингового правила). В практической статистике для оценки расхождения между вероятностными прогнозами и количественными исходами часто используется скоринговое правило непрерывной ранжированной вероятности (Continuous Ranked Probability Score – CRPS). В работе рассматривается случай, когда несколько конкурирующих методов (экспертов) дают свои прогнозы в виде функций распределения. Прогнозы экспертов снабжены уровнями доверия. Предложен алгоритм для онлайн агрегирования этих функций распределения с учетом уровней доверия к прогнозам экспертов. Получены оценки дисконтированной с учетом уровней доверия ошибки предложенного алгоритма в форме сравнения кумулятивных потерь алгоритма и потерь экспертов. Разработана технология построения прогнозирующих экспертных алгоритмов и агрегирования их вероятностных прогнозов на примере задачи прогнозирования потребления электроэнергии на один или более часов вперед. Представлены результаты численных экспериментов на реальных данных.

КЛЮЧЕВЫЕ СЛОВА: предсказания с использованием экспертных стратегий, вероятностное прогнозирование в режиме онлайн, регрет, уровни доверия, агрегирующий алгоритм, скоринговое правило непрерывной ранжированной вероятности, CRPS.

1. ВВЕДЕНИЕ

Вероятностные прогнозы в форме распределения вероятностей будущих событий нашли широкое применение в таких областях, как метеорология, гидрология, экономика и т.д. Различие между вероятностным прогнозом и числовым исходом измеряется с помощью функции потерь (скорингового правила). Известным примером скорингового правила для непрерывных исходов являются правило непрерывной ранжированной вероятности (CRPS), которое определяется как

$$\text{CRPS}(F, y) = \int (F(u) - 1_{u \geq y})^2 du,$$

где $F(u)$ - функция распределения вероятностей, а y -исход (вещественное значение).¹

Мы рассматриваем постановку, 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 прогноз таким образом, чтобы минимизировать потери, исходя из имеющейся прошлой информации.

¹ Здесь $1_{u \geq y} = 1$, если $u \geq y$, в противном случае это 0.

Эта проблема решается в рамках теории предсказаний с использованием экспертных стратегий (Prediction with Expert Advice – PEA). Каждая экспертная стратегия использует свою специфическую предсказательную модель, описывающую и прогнозирующую наблюдаемые данные. При этом мы можем не знать природу источника, порождающего реальные исходы. Факторы, влияющие на релевантность экспертных моделей, могут быть скрыты от предсказателей, поэтому задача предсказания решается в рамках теоретико-игрового подхода, при котором не используется никаких предположений о том, какая из возможных моделей релевантна в данный момент времени.

Агрегирующий алгоритм в режиме онлайн объединяет прогнозы экспертных моделей в зависимости от их потерь в прошлом. Разность между кумулятивными 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 [1], [2], [3], [4] были развиты методы агрегирования в режиме онлайн точечных прогнозов. В этом случае эксперты выдают вещественные прогнозы в виде чисел или векторов, агрегирующий алгоритм выдает прогноз такого же вида.

В отличие от стандартной постановки задачи предсказания с использованием экспертных стратегий, в работах [5], [6] рассматривался случай, 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.

Качество экспертов и их прогнозов будем в данной работе измерять с помощью скорингового правила непрерывной ранжированной вероятности CRPS как функции потерь. На каждом временном шаге каждый эксперт выдает свой прогноз в форме распределения вероятностей. Агрегирующий алгоритм объединяет эти прогнозы в один консолидированный прогноз, который также является функцией распределения вероятностей. Эффективность агрегирующего алгоритма на любом временном интервале $[1, T]$ измеряется регретом, который представляет собой разницу между кумулятивными потерями агрегирующего алгоритма и кумулятивными потерями наилучшего эксперта.

Опубликовано много статей о вероятностных предсказаниях с использованием скорингового правила CRPS (напр., [7], [8], [9], [10], [11], [12], [13]). В подавляющем большинстве работ используются ансамблевые прогнозы и квантильные оценки функций распределения. Под ансамблями обычно понимают совокупность траекторий модели, вычисленных при измененных начальных условиях или возмущенных уравнениях модели. Ансамбль прогнозов может быть преобразован в кусочно-постоянную функцию распределения. Так, в работе [10] показано, что для вычисления CRPS можно интерпретировать ансамбль как квантильные оценки с неоднородными уровнями.

В работах [5], [6] предложены правила агрегирования прогнозов вероятностных моделей (экспертов). В данной работе прогнозы каждого эксперта сопровождаются уровнями доверия. Каждый уровень доверия представляет собой число, лежащее в интервале $[0, 1]$. Значение уровня доверия, равное 1, означает, что соответствующий экспертный прогноз полностью принимается во внимание агрегирующим алгоритмом, при значении уровня доверия 0 прогноз эксперта вообще им не учитывается, а при промежуточном значении уровня доверия соответствующий прогноз учитывается лишь частично.

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

Предлагаемые правила работают как в случае аналитических моделей, так и в случае использования детальных численных описаний функций распределения на основе ансамблевых или квантильных прогнозов.

В этой работе мы получаем теоретическую верхнюю границу дисконтированного (с учетом уровней доверия) регрета для специального случая, когда результаты измерения и распределения вероятностей находятся в конечном интервале $[a, b]$ вещественной прямой.

В разделе 2 представлены некоторые детали агрегирующего алгоритма Вовка [14], 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 3 мы обобщаем агрегирующий алгоритм на тот случай, когда прогнозы экспертов снабжены уровнями доверия. 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 ([16]). Отличительной особенностью данного конкурса является необходимость получения прогноза отклика в виде описания его вероятностного распределения.

Работа частично поддержана грантом РФФИ 20-01-00203.

2. ОСНОВНЫЕ ПОНЯТИЯ И УТВЕРЖДЕНИЯ

В этом разделе даются необходимые определения и вспомогательные результаты теории предсказаний с использованием экспертных стратегий.

2.1. Онлайн обучение и прогнозирование

Пусть Ω – множество исходов, Γ – множество прогнозов и $\lambda(f, y)$ – функция потерь, где $f \in \Gamma$ и $y \in \Omega$.² Также задано E – множество экспертов, обозначим $E = \{1, \dots, N\}$.

В теории предсказаний с использованием экспертных стратегий процесс обучения 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 оформляется в виде Протокола 1.

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 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

² Природа этих множеств будет уточнена позже.

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.

2.2. Агрегирующие алгоритмы АА и WA

Для агрегации прогнозов экспертов будем схему, предложенную в работах [3], [14]. Этот алгоритм 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:

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

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

$$w_{i,t}^* = \frac{w_{i,t}}{\sum_{j=1}^N w_{j,t}}. \quad (2)$$

Смешиваемость. В алгоритме АА используется так называемая функция суперпредсказания. Согласно [14] 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 $\bar{f} = (f_1, \dots, f_N)$ существует предсказание f такое, что

$$\lambda(f, y) \leq g(y) \text{ для всех } y, \quad (3)$$

где

$$g(y) = -\frac{1}{\eta} \ln \sum_{i=1}^N e^{-\eta \lambda(f_i, y)} q_i \quad (4)$$

– функция суперпредсказания. Предполагается, что $\eta > 0$.

Фиксируем некоторое правило для вычисления предсказания f и полагаем

$$f = \text{Subst}(\bar{f}, \mathbf{q}). \quad (5)$$

Функция Subst называется функцией подстановки.

В алгоритме АА на каждом шаге t рассматривается функция псевдопредсказания

$$g_t(y) = -\frac{1}{\eta} \ln \sum_{i=1}^N e^{-\eta \lambda(f_{i,t}, y)} w_{i,t}^*, \quad (6)$$

а прогноз предсказателя равен $f_t = \text{Subst}(\bar{f}_t, \mathbf{w}_t^*)$, где $\bar{f}_t = (f_{1,t}, \dots, f_{N,t})$ – 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, $\mathbf{w}_t^* = (w_{1,t}^*, \dots, w_{N,t}^*)$ – нормализованные веса и $\bar{f}_t = (f_{1,t}, \dots, f_{N,t})$ – предсказания экспертов на шаге t . Пусть также в Протоколе 1 предсказатель вычисляет свой прогноз $f_t = \text{Subst}(\bar{f}_t, \mathbf{w}_t^*)$. Согласно (3), $\lambda(f_t, y_t) \leq g_t(y_t)$ для всех t , где $g_t(y)$ определено по (6).

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 (1) 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

$$H_T \leq \sum_{t=1}^T g_t(y_t) = -\frac{1}{\eta} \ln W_{T+1} \leq L_T^i + \frac{\ln N}{\eta} \quad (7)$$

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 вогнутые 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 вогнутой, если для каждого y функция $\exp(-\eta \lambda(f, y))$ является вогнутой по f (см. [1]). Для экспоненциально вогнутой 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

$$f_t = \sum_{i=1}^N w_{i,t}^* f_{i,t}, \quad (8)$$

где веса $w_{i,t}^*$ вычисляются по формулам (1) и (2), алгоритмом WA.

Если в Протоколе 1 вычислять прогноз предсказателя по формуле (8), то также имеет место независимая от времени верхняя оценка регрета (7).

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. [3], [14] и [17]), она также является η -экспоненциально вогнутой при $0 < \eta \leq \frac{1}{2(b-a)^2}$ (см. [1]).

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

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

где $\bar{f} = (f_1, \dots, f_N)$ – прогнозы экспертов, $\mathbf{q} = (q_1, \dots, q_N)$ – произвольное вероятностное распределение (доказательство см. в [5]). При $a = 0$ и $b = 1$ полагаем $\eta = 2$.

2.3. Агрегирование вероятностных предсказаний

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 $F : [a, b] \rightarrow [0, 1]$.³

³ Функция распределения есть неубывающая функция $F(y)$ определенная на интервале $[a, b]$ такая, что $F(a) = 0$ и $F(b) = 1$. Она также непрерывна слева и у нее в каждой точке существует предел справа.

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

$$\text{CRPS}(F, y) = \int_a^b (F(u) - H(u - y))^2 du, \quad (9)$$

где $H(x)$ – функция Хевисайда: $H(x) = 0$ при $x < 0$ и $H(x) = 1$ при $x \geq 0$ ([18], [19], [20], [21] и так далее). 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 помещают 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 [5] было доказано, что функция потерь $\text{CRPS}(F, y)$ является η -смешиваемой при $0 < \eta \leq \frac{2}{b-a}$ и η -экспоненциально вогнутой при $0 < \eta \leq \frac{1}{2(b-a)}$ и поэтому к ней применимы оценки регрета из разделов 2, 2.1. Сформулируем точные утверждения.

Теорема 1. *Функция потерь $\text{CRPS}(F, y)$ 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

$$F(u) = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N q_i e^{-2(F_i(u))^2}}{\sum_{i=1}^N q_i e^{-2(1-F_i(u))^2}}, \quad (10)$$

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

Функция потерь CRPS является η -экспоненциально вогнутой при $0 < \eta \leq \frac{1}{2(b-a)}$. В этом случае, правило вычисления прогноза предсказателя (10) можно заменить на правило WA

$$F_t(u) = \sum_{i=1}^N q_i F_i(u), \quad (11)$$

Полное доказательство этой теоремы см. в работах [5] и [6].

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. Вычисляем прогноз предсказателя – функцию распределения

$$F_t(u) = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N w_{i,t}^* e^{-2(F_{i,t}(u))^2}}{\sum_{i=1}^N w_{i,t}^* e^{-2(1-F_{i,t}(u))^2}}, \quad (12)$$

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 $1 \leq i \leq N$

$$w_{i,t+1} = w_{i,t} e^{-\frac{2}{b-a} \text{CRPS}(F_{i,t}, y_t)} \quad (13)$$

ENDFOR

В случае, когда применяется алгоритм WA для вычисления прогноза предсказателя, правило (12) можно заменить на

$$F_t(u) = \sum_{i=1}^N w_{i,t}^* F_{i,t}(u), \quad (14)$$

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

$$w_{i,t+1} = w_{i,t} e^{-\frac{1}{2(b-a)} \text{CRPS}(F_{i,t}, y_t)}. \quad (15)$$

Оценка регрета приведенного алгоритма дается в следующей теореме.

Теорема 2. Для произвольного i имеет место верхняя оценка

$$\sum_{t=1}^T \text{CRPS}(F_t, y_t) \leq \sum_{t=1}^T \text{CRPS}(F_{i,t}, y_t) + \frac{b-a}{2} \ln N \quad (16)$$

для каждого T .

В случае, когда прогноз предсказателя вычисляется по правилу WA (14), а веса переопределяются по правилу (15), имеет место оценка

$$\sum_{t=1}^T \text{CRPS}(F_t, y_t) \leq \sum_{t=1}^T \text{CRPS}(F_{i,t}, y_t) + 2(b-a) \ln N. \quad (17)$$

Доказательство. Оценки (16) и (17) прямо следуют из оценки 7. $\triangle$

Отметим, что оценка регрета (17) в четыре раза хуже, чем оценка (16). Как показывают экспериментальные расчеты раздела 4, кумулятивные потери алгоритма АА несколько меньше чем потери алгоритма WA.

3. АА ДЛЯ ЭКСПЕРТОВ С УРОВНЯМИ ДОВЕРИЯ

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

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

Уровень доверия может определяться самим экспертом или каким-либо внешним образом.⁴

Мы используем уровни доверия в агрегирующем алгоритме на основе использования метода “фиксированной точки” из работы [23]. Сопоставим каждому уровню доверия $p_{i,t}$ 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

$$e^{-\eta\lambda(f_t, y)} \geq \sum_{i=1}^N E_{\mathbf{p}_{i,t}}[e^{-\eta\lambda(\tilde{f}_{i,t}, y)}] w_{i,t} \quad (18)$$

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 .

Запишем неравенство (18) в более детальной форме:

$$e^{-\eta\lambda(f_t, y)} \geq \sum_{i=1}^N E_{\mathbf{p}_{i,t}}[e^{-\eta\lambda(\tilde{f}_{i,t}, y)}] w_{i,t} = \quad (19)$$

$$\sum_{i=1}^N p_{i,t} w_{i,t} e^{-\eta\lambda(f_{i,t}, y)} + e^{-\eta\lambda(f_t, y)} \left(1 - \sum_{i=1}^N p_{i,t} w_{i,t} \right) \quad (20)$$

для всех ω . Таким образом, неравенство (18) эквивалентно неравенству

$$e^{-\eta\lambda(f_t, y)} \geq \sum_{i=1}^N w_{i,t}^* e^{-\eta\lambda(f_{i,t}, y)}, \quad (21)$$

где

$$w_{i,t}^* = \frac{p_{i,t} w_{i,t}}{\sum_{j=1}^N p_{j,t} w_{j,t}}. \quad (22)$$

Используем правило (5) для вычисления предсказания алгоритма АА: $f_t = \text{Subst}(\bar{f}_t, \mathbf{w}_t^*)$. Тогда (21) эквивалентно (19). Здесь Subst – функция подстановки, $\mathbf{w}_t^* = (w_{1,t}^*, \dots, w_{N,t}^*)$ и $\bar{f}_t = (f_{1,t}, \dots, f_{N,t})$.

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 $\hat{l}_{i,t}$.

⁴ Данная постановка была предложена в работе [22] и далее развивалась в работе [1].

Модифицируем Протокол 1 для АА с уровнями доверия.

Протокол 1a

FOR $t = 1, \dots, T$

1. Получаем предсказания $f_{i,t}$ экспертов и уровни доверия $p_{i,t}$, где $1 \leq i \leq N$.
2. Вычисляем прогноз предсказателя $f_t = \text{Subst}(\bar{f}_t, \mathbf{w}_t^*)$, где $\mathbf{w}_t^* = (w_{1,t}^*, \dots, w_{N,t}^*)$ – нормализованные веса, определенные по формуле (22).
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:

$$w_{i,t+1} = w_{i,t} e^{-\eta(p_{i,t}\lambda(f_{i,t}, y_t) + (1-p_{i,t})\lambda(f_t, y_t))} \quad (23)$$

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

$$\sum_{t=1}^T p_{i,t}(h_t - l_{i,t}) \leq \frac{\ln N}{\eta}. \quad (24)$$

Доказательство. Из выпуклости экспоненты и неравенства (18) получаем

$$e^{-\eta\lambda(f_t, y)} \geq \sum_{i=1}^N e^{-\eta E_{p_{i,t}}[\lambda(\tilde{f}_{i,t}, y)]} w_{i,t}^* = \sum_{i=1}^N e^{-\eta \hat{l}_{i,t}} w_{i,t}^*. \quad (25)$$

Пусть $m_t = -\frac{1}{\eta} \ln \sum_{i=1}^N w_{i,t}^* e^{-\eta \hat{l}_{i,t}}$. По (25) $h_t \leq m_t$. Перепишем правило (23) в виде

$$w_{i,t+1} = w_{i,t} e^{-\eta \hat{l}_{i,t}}. \quad (26)$$

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 (7), используя (25) и (26), 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 (24). $\triangle$

4. ЭКСПЕРИМЕНТЫ

При проведении экспериментов для вычисления прогнозов предсказателя использовались правила (алгоритмы) АА и WA. Эти правила удобно использовать в том случае, когда функции распределения, предложенные экспертами, заданы точным образом в виде формул или вычислительных процедур.

4.1. Агрегирование вероятностных предсказаний с уровнями доверия

В разделе 4.2 (далее) будут представлены результаты численных экспериментов с реальными данными для случая, когда предсказания экспертов снабжены уровнями доверия. Приведем модификацию протокола 2, которая использовалась в этих экспериментах.

Протокол 2а

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

FOR $t = 1, \dots, T$

1. Получаем прогнозы экспертов – функции распределения $F_{i,t}(u)$ и уровни доверия $p_{i,t}$, где $1 \leq i \leq N$.
2. Вычисляем прогноз предсказателя – функцию распределения $F_t(u)$, которая определена по правилу

$$F_t(u) = \frac{1}{2} - \frac{1}{4} \ln \frac{\sum_{i=1}^N w_{i,t}^p e^{-2(F_{i,t}(u))^2}}{\sum_{i=1}^N w_{i,t}^p e^{-2(1-F_{i,t}(u))^2}} \quad (27)$$

для алгоритма АА или по правилу

$$F_t(u) = \sum_{i=1}^N w_{i,t}^p F_{i,t}(u) \quad (28)$$

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 $1 \leq i \leq N$

$$w_{i,t+1} = w_{i,t} e^{-\eta(p_{i,t} \text{CRPS}(F_{i,t}, y_t) + (1-p_{i,t}) \text{CRPS}(F_t, y_t))}, \quad (29)$$

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

ENDFOR

Оценка регрета данного алгоритма дается неравенством (24) теоремы 3, 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 (27), и $\eta = \frac{1}{2(b-a)}$, если было использовано правило (28).

Также использовалась модификация Fixed Share (см. [24]), при которой правило (29) 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 $0 < \alpha < 1$. Полагаем $\alpha = 0.001$ в экспериментах.

4.2. 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 ([16]).

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 включал данные о почасовых электрических нагрузках за период с 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 включала данные за период с 1 января 2011 г. до декабря 2011 г.

Для решения поставленной задачи участники конкурса применяли различные подходы и алгоритмы вероятностного прогнозирования, наиболее успешные из которых представлены в обзорной работе [16]. Базы данных доступны на сайте [25]. В дополнение к этим данным можно было использовать календарную информацию о временах года, днях недели, а также рабочих и праздничных днях.

Судя по обзору основных результатов конкурса [16], его участники при построении прогностических вероятностных распределений, как правило, ограничивались рассмотрением только одного физически измеряемого показателя, а именно, в каждый момент времени использовали только значение средней температуры, вычисленное по данным всех 25 метеостанций. Кроме температурных данных, участникам конкурса были доступны для формирования прогноза только календарные показатели.

В данной работе мы тоже 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.

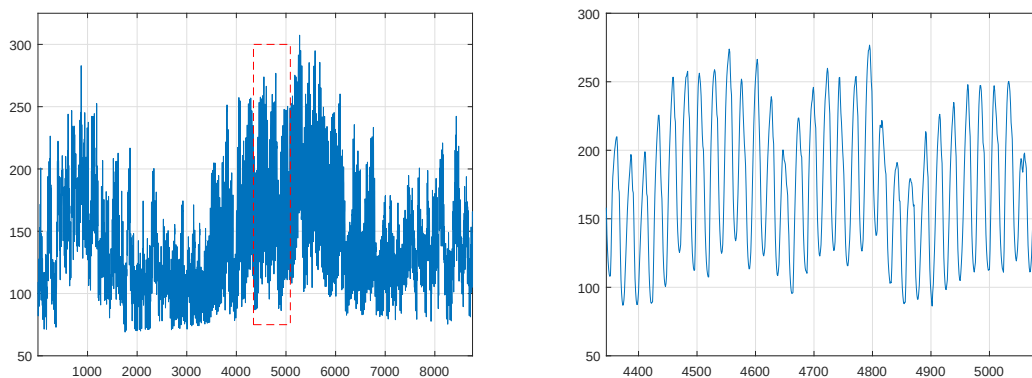


Рис. 1. Почасовые электрические нагрузки: слева за один год (2007) обучающей части выборки, справа – выделенный фрагмент (Июль 2007 года).

Технология проведения экспериментов. На рис. 1 приведены графики почасовых электрических нагрузок - слева за один год (2007) из шести лет обучающего периода, справа для наглядности – увеличенный фрагмент, соответствующий одному месяцу, выделенному прямоугольником на левом рисунке. Эти данные позволяют оценить календарную изменчивость величины электрических нагрузок в зависимости от сезона года и периодическую зависимость от

времени суток. Кроме этого, диаграммы рассеяния, представленные на рис. 2 и 3, демонстрируют характерные особенности зависимости величины потребления от температуры внутри каждого из календарных сегментов.

С учетом этих наблюдений, обучающая выборка была разбита на подвыборки, соответствующие временам года (зима, весна, лето, осень) и времени дня (утро, день, вечер, ночь). На каждой подвыборке был обучен эксперт, который определялся соответствующей функцией распределения, построенной по диаграмме рассеяния (такой как на рис. 2 и 3, детали см. ниже).

Такие эксперты (алгоритмы) называются специализированными. Иногда используют термин “спящие” эксперты, поскольку эти эксперты выполняют свои функции только при некоторых внешних условиях, а в остальное время “спят” и не участвуют в принятии решений. Примеры формирования и использования специализированных алгоритмов при точечном прогнозировании электрических нагрузок можно найти в работе [26].

На тестовой выборке применение специализированных экспертов регулировалось уровнями доверия к их прогнозам. Каждый эксперт имел уровень доверия 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 к его прогнозам вне его области компетентности понижался линейным образом от 1 до 0 (см. рис. 6). Поэтому в окрестностях различных подвыборок появляются конкурирующие между собой эксперты с различными ненулевыми уровнями доверия. Агрегирующий алгоритм (протокол 2а) в каждый момент времени объединял прогнозы экспертов (распределения вероятностей) с учетом этих уровней доверия.

Построение и обучение экспертов. Кроме алгоритмов прогнозирования общего профиля, которые обучаются и делают свои прогнозы на всех данных выборки, нередко рассматривают алгоритмы, работающие только при определенных сочетаниях прогностических параметров, например, ориентированные на выдачу прогнозов только в определенные сезоны и часы дня.

Отмеченная выше зависимость характера изменения электрических нагрузок от климатических сезонов и времени суток была использована для построения множества специализированных вероятностных прогностических алгоритмов.

На рисунке 2 приведена диаграмма рассеяния для показателей (час, потребление), построенная по всем данным обучающей части выборки.

Как видно из 3D диаграмм рассеяния в правой части рис. 2, сезон года и время суток проявляются в виде заметных регулярных изменений общего вида облаков рассеяния и их взаимного расположения. Для более детального анализа влияния этих параметров на следующем рис. 3 по данным обучающей выборки была построена 4×4 матрица диаграмм рассеяния для всех сочетаний 4-х сезонов года и четырех интервалов суток (ночь, утро, день, вечер).

Зависимость электрических нагрузок от температуры может существенно меняться с изменением сезона и времени суток. В частности, на диаграммах рассеяния для летнего и зимнего сезонов связь между потреблением электричества и температурой воздуха имеет достаточно простой, почти линейный характер. Для других сезонов года характерны нелинейные изменения. Кроме того, можно отметить сходство между весенними и осенними диаграммами. Как следует из взаимного расположения и достаточно простой формы облаков точек на рис. 2, совместные двумерные вероятностные распределения температуры и электрической нагрузки могут быть представлены в виде композиции нескольких (2 или 3) простых эллипсообразных облаков точек.

Заметим, что диаграммы рассеяния на рис. 2 и 3 были построены с учетом средней текущей температуры, взятой с упреждением в 1 час. Как было выше сказано, в практических задачах

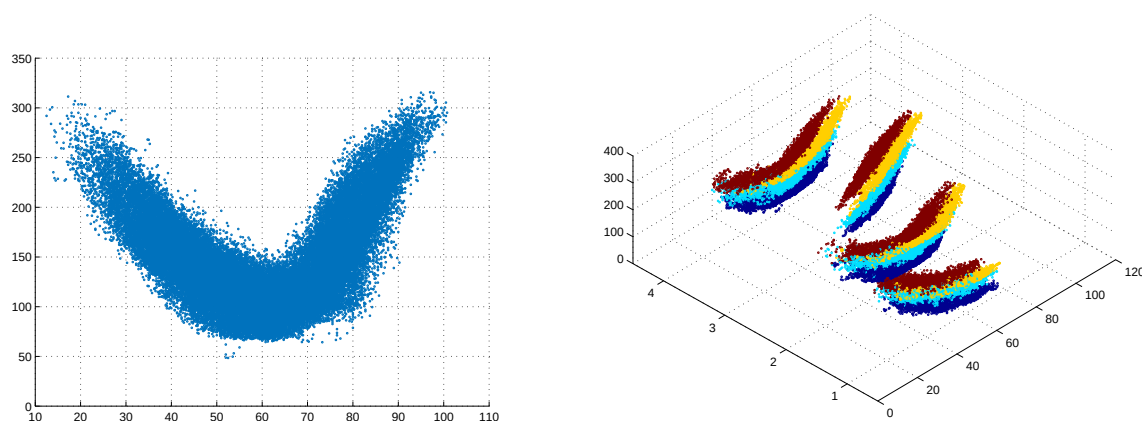


Рис. 2. Слева показано облако рассеяния (Температура, Нагрузка) для всех данных обучающей выборки, справа в 3D формате изображены диаграммы рассеяния тех же параметров, сгруппированные по сезонам (1-зима, 2-весна, 3-лето, 4-осень) и по времени суток. Временные интервалы суток (ночь, утро, день, вечер) выделены полутонами.

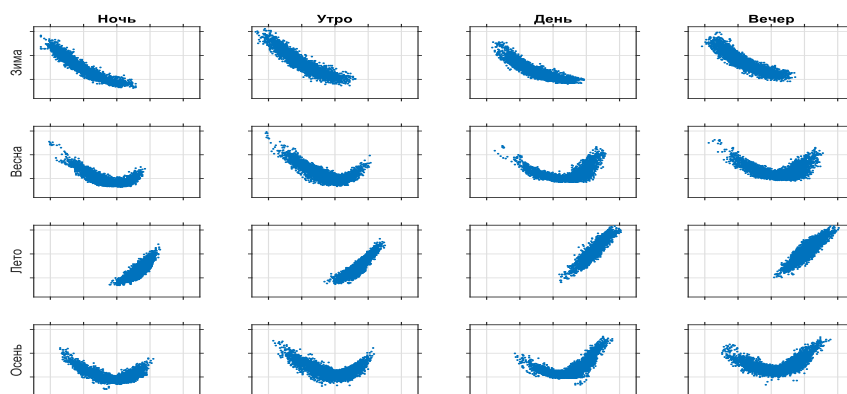


Рис. 3. Матрица диаграмм рассеяния для разных сочетаний календарных параметров (сезон года, время суток). Для удобства сравнения все диаграммы построены в одной и той же системе координат, где температура изменяется в интервале от 10 до 110 градусов (по Фаренгейту), а нагрузка в интервале от 20МГВ до 320МГВ.

горизонт прогноза (время упреждения) может изменяться в широком диапазоне от одного часа до десятков и даже сотен часов.

Для прогноза электрической нагрузки с упреждением на два или более часов необходимо иметь соответствующую модель для прогноза температуры. Это может быть один из известных методов предсказания значений одномерного случайного временного ряда, основанный на использовании авто-регрессии или более сложной модели с учетом сезона.

С учетом этих наблюдений, 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 на основе этих данных. Так, специализированный эксперт (Зима, Ночь) был сформирован на основе облака точек, представленных на соответствующей диаграмме рассеяния из рис. 3.

Вычислительный эксперимент можно условно разбить на следующие основные этапы:

1. Формируется множество экспертов-специалистов, путем разбиения пространства прогностических показателей на области, где связь между температурой и электрической нагрузкой может быть описана простой и относительно однородной зависимостью, составляющей его “специализацию”. Для этого формируется функция компетентности (доверия) или функция принадлежности эксперта, зависящая от текущего момента времени, который легко переводится в формальные календарные показатели (сезон года, час суток).

2. Для каждого эксперта генерируется индивидуальная обучающая подвыборка, составленная из тех точек полной обучающей выборки, которые отвечают его специализации. По данным подвыборки формируется прогностическая вероятностная модель на основе диаграмм рассеяния, представленных на рис. 2 и 3.

Один из наиболее простых и наглядных методов описания вероятностного распределения состоит в построении приближенного описании двумерного эмпирического распределения в виде вычислительной параметрической модели. Например, можно применить аппроксимацию на основе метода гауссовской смеси ([27]) или использовать ядерное восстановление плотности распределения ([28]).

Если построена эмпирическая модель двумерного распределения температура-нагрузка, то для получения оценки одномерного распределения при заданном значении температуры достаточно рассмотреть одномерное сечение двумерного распределения.

Приведенные диаграммы рассеяния могут служить основой для формирования множества прогнозирующих экспертов. Каждый из них обучается (строит прогностическую вероятностную модель) на точках выборки, относящихся к заранее заданному календарному сегменту, например, (Зима, Утро). Эти сегменты должны покрывать все возможные сочетания календарных показателей, присутствующих в данных. При формировании окончательного прогноза результаты экспертов, настроенных на соседних временных сегментах, могут комбинироваться для получения более гибкого описания изменений климатической обстановки.

Для создания априорного временного графика календарного взаимодействия между экспертами используется технология размытых множеств, в каждом из которых уровень доверия или компетентность эксперта описывается числовым параметром, изменяющимся в интервале $[0, 1]$, как это сделано в работе [15] для точечных прогнозов.

В общий список экспертов был включен один постоянно работающий Эксперт 1, который был обучен на основе всех данных обучающей выборки (из левого графика на рис. 2), Эксперты 2–5 были обучены на данных четырех сезонов, представленных на рис. 2. Эксперты 6–21 соответствуют диаграммам рассеяния, представленным на рис. 3. Вероятностное распределение каждого эксперта строилось на основе метода гауссовской смеси, В результате вероятностная модель каждого из экспертов была представлена как композиция двух или трех нормальных распределений.

В данной работе для иллюстрации предложенных методов мы рассматриваем частную проблему прогнозирования – краткосрочное прогнозирование функции распределения вероятностей электрических нагрузок на один час вперед по данным текущей температуры и учетом текущих календарных параметров.

Область компетентности каждого эксперта определяется значениями его функции (уровнями) доверия. На этапе обучения эксперта в тренировочную выборку включаются только те точки, где его уровень доверия строго равен 1. При прогнозировании область применения эксперта распространяются на все точки, где его уровень компетентности (доверия) больше 0. Таким образом, каждый эксперт конкурирует с другими экспертами, работающими на перекрывающихся интервалах. При этом, используется соответствующий алгоритм агрегирования экспертов с уровнями достоверности.

Диаграммы рассеяния (Температура, Нагрузка) для нескольких множеств календарных параметров, сгруппированным по четырем сезонам и по четырем дневным интервалам (каждый по 6 часов), представлены на рис. 3. Диаграммы построены по обучающей части выборки.

Как видно из расположения и достаточно простой формы облаков точек на рис. 2 и 3, совместные двухмерные вероятностные распределения температуры и электрической нагрузки могут быть представлены в виде композиции нескольких (2 или 3) двухмерных гауссовских распределений. В зависимости от выбранного набора календарных параметров на каждой из представленных диаграмм рассеяния можно выделить два или три температурных интервала таким образом, что на каждом из интервалов облако точек имеет простую эллипсообразную форму. Это дает основание для использования смеси нормальных распределений для вероятностного прогноза ожидаемой электрической нагрузки по данным краткосрочного прогноза температуры.

Зависимость электрических нагрузок от температуры существенно меняется в зависимости от сезона и времени дня. Кроме того, можно отметить сходство между весенними и осенними диаграммами. Приведенные иллюстрации позволяют сформировать наглядные прогностические алгоритмы для восстановления вероятностного распределения электрических нагрузок по измерениям температуры воздуха.

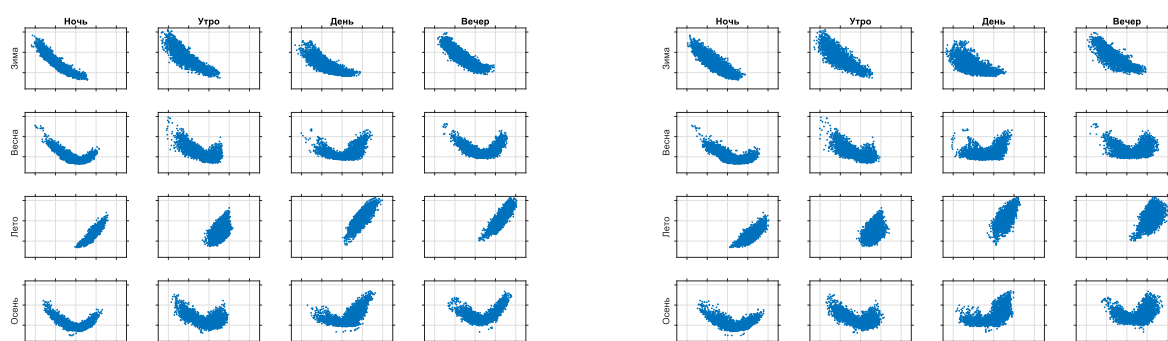


Рис. 4. Диаграммы рассеяния, полученные для временного упреждения с прогнозом на 4 и 8 часов вперед на основании только текущей температуры.

Диаграммы рассеяния на рис. 2 были построены с учетом средней по всем станциям температуры, которая наблюдалась 1 час тому назад. Как было выше сказано, в практических задачах горизонт прогноза (время упреждения) может изменяться в широком диапазоне от одного часа до десятков и даже сотен часов.

Для прогноза электрической нагрузки с упреждением на два или более часов необходимо иметь соответствующую модель для прогноза температуры. Это может быть один из известных методов предсказания значений одномерного случайного временного ряда на использовании авто-регрессии или более сложной модели с учетом сезонности и суточной периодичности.

Для иллюстрации влияния горизонта прогноза на ожидаемую точность на следующем рис. 4 показаны диаграммы рассеяния, полученные для других значений временного упреждения, а именно с прогнозом на 4 и 8 часов вперед на основании только текущей температуры. На них хорошо видно, как размывается прогноз по сравнению с прогнозом на 1 час вперед (рис. 3) с ростом упреждения по времени.

Результаты экспериментов. Для иллюстрации работы предложенных в статье методов агрегирования вероятностных прогнозов экспертов–специалистов была выбрана частная задача, а именно, прогнозирование функции распределения электрических нагрузок на один час вперед

на основании предыстории и текущих календарных параметров. Для учета температурных изменений в качестве предиктора использовали только температуру, усредненную по данным всех 25 метеорологических станций. Предварительная обработка исходных таблиц состояла в формировании объединенной выборки и добавлении календарных параметров. В результате предобработки был сформирован синхронизированный многомерный временной ряд, дополненный предысторией с некоторым заданным лагом для расширения пространства предикторов, а также текущей календарной информацией, необходимой для оценки доверительных уровней экспертов-специалистов.

Кривые дисконтированного регрета $T \rightarrow \sum_{t=1}^T p_{i,t}(h_t - l_{i,t})$ приведены на рис. 5 для алгоритма АА (слева на рисунке) и алгоритма WA (справа на рисунке) относительно каждого из 21 специализированного эксперта. Прямые линии представляют собой теоретическую оценку регрета.

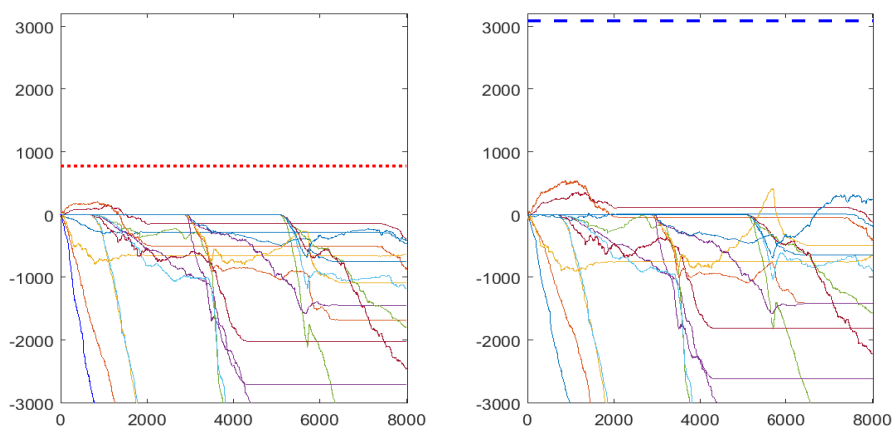


Рис. 5. Кривые дисконтированного регрета алгоритма АА (слева) and алгоритма WA (справа) относительно каждого из 21 специализированного эксперта. Штриховая линия представляет теоретическую оценку регрета.

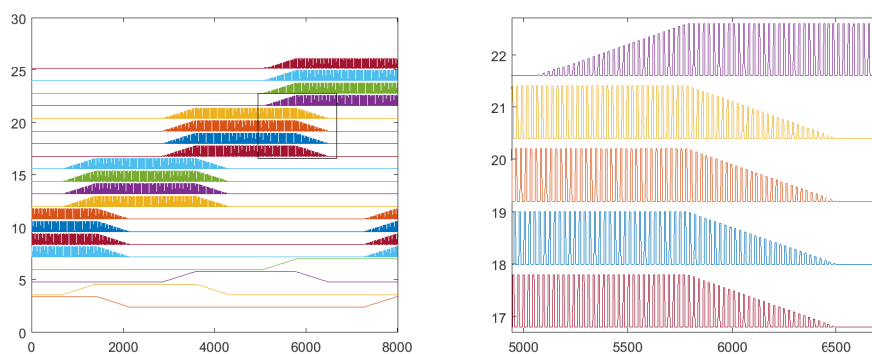


Рис. 6. Левая часть рисунка: уровни доверия для Экспертов 2-5 (сезонные эксперты) и 6-21 (Сезон, Время дня). Правая часть: увеличенный фрагмент рисунка из левой части.

Для обоснования роли параметров уровня доверия были проведены сравнительные эксперименты. Во время первого эксперимента все значения уровней достоверности для каждого

эксперта были равны 1. Во втором эксперименте алгоритмы АА и WA использовали прогнозы экспертов в пределах уровней их достоверности заданных внешним образом, примеры которых приведены на рис. 6.

Результаты двух методов АА и WA агрегации этих экспертов также представлены на рис. 7. В первом методе агрегирования уровни доверия всех экспертов были равны 1 (линии 2 и 3).

Во втором методе алгоритмы АА и WA используют специализированных экспертов, при этом, их уровни доверия устанавливаются извне. Кривые 4-5 отображают потери алгоритмов АА и WA для случая, когда уровни доверия к прогнозам экспертов соответствуют областям их компетентности и монотонно уменьшаются до нуля вне этих областей (см. рис. 6). Кривые 6-7 построены для случая, когда уровни доверия точно соответствуют областям компетентности (равны 1 внутри области и равны 0 вне этой области – эксперт активен внутри области и спит вне ее).

Результаты агрегации специализированных экспертов АА и WA представлены на рис. 7. Результаты проведенных экспериментов показывают, что использование уровней доверия специализированных экспертов повышает эффективность процесса онлайн-адаптации. Использование гладких уровней доверия повышает эту эффективность в большей степени, чем просто использование спящих экспертов.

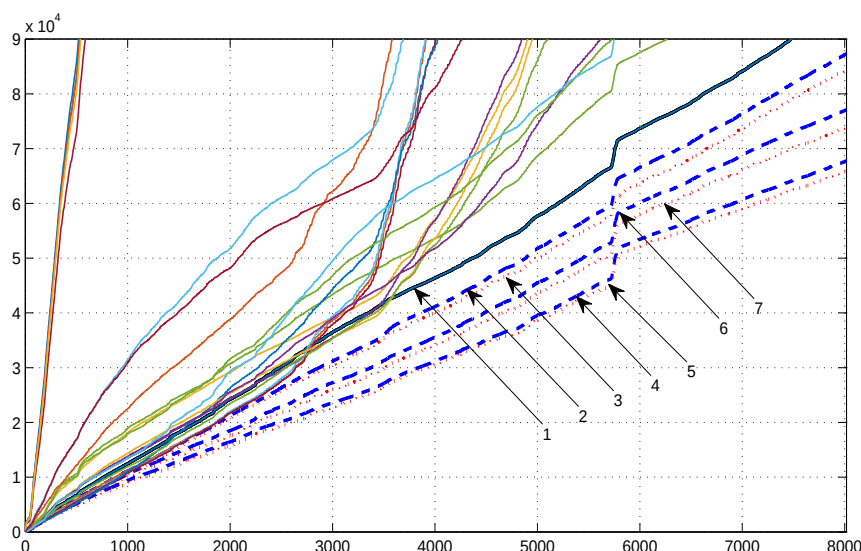


Рис. 7. Сравнительные результаты обучения с учетом специализации экспертов и без нее. Кумулятивные потери каждого из экспертов, при условии, что они работают все время, и результаты их агрегирования. 1 – Потери “универсального” алгоритма, обученного на всей выборке; 2 и 3 – результаты агрегации экспертов с помощью алгоритмов WA и АА без учета индивидуальных уровней доверия (уровни доверия всех экспертов равны 1); 4-5 – результаты агрегации алгоритмами WA и АА с использованием нетривиальных перекрывающихся значений уровней доверия; 6-7 – то же для случая, когда области компетентности экспертов не перекрываются.

На рис. 8 и 11 приведены функции распределения и плотности для прогнозов потребления для двух моментов суток 24 часа (полночь) и 12 часов (полдень).

На рис. 9, 10 и 12, 13 приведены межквантильные интервалы прогнозов электрических нагрузок и их фактические значения. Приведены графики для двух моментов суток 24 часа (полночь) и 12 часов (полдень). Тестовый период составляет более 330 суток.

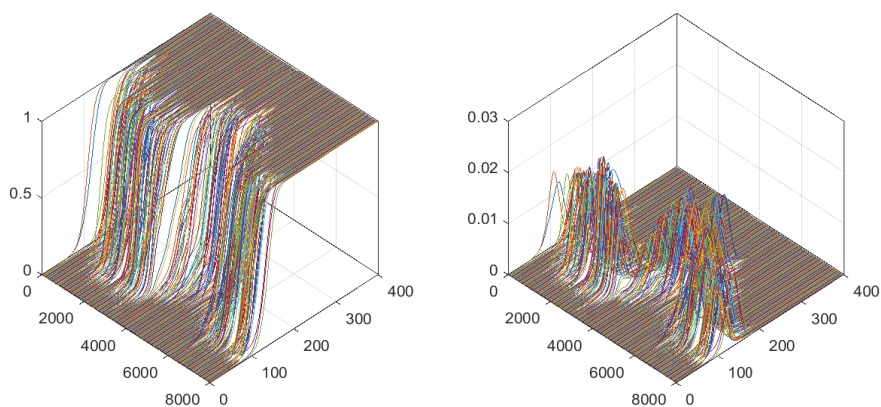


Рис. 8. Функции распределения и плотности прогнозов на 12 часов (полдень).

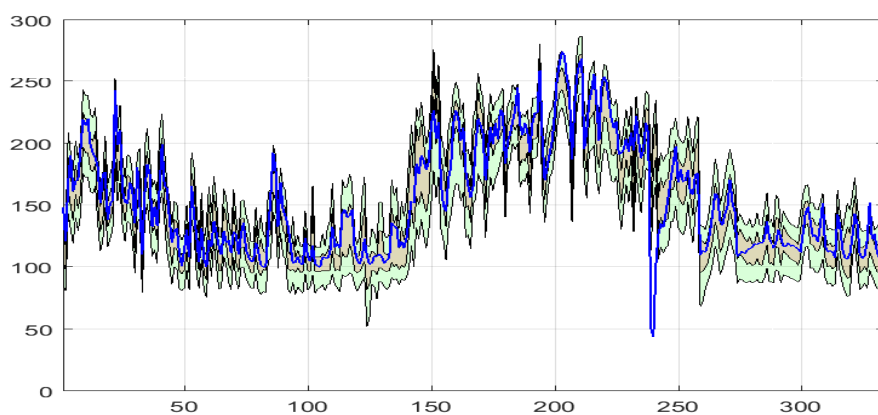


Рис. 9. Межквантильные интервалы (полутона) и фактические нагрузки (синяя линия) в 12 часов (полдень).

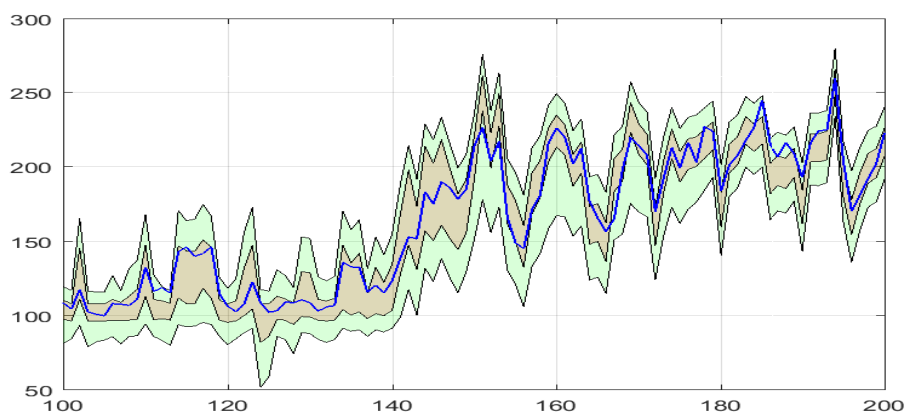


Рис. 10. Межквантильные интервалы (полутона) $[0.25, 0.75]$ и $[0.05, 0.95]$ и фактические нагрузки в 12 часов – полдень (сплошная линия). Фрагмент рис. 9 с 10 марта 2011 года до 20 июля того же года.

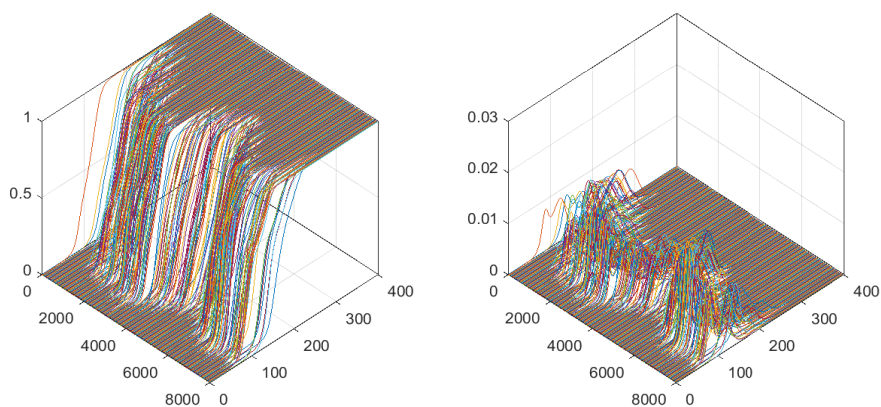


Рис. 11. Функции распределения и плотности прогнозов на 24 час (полночь).

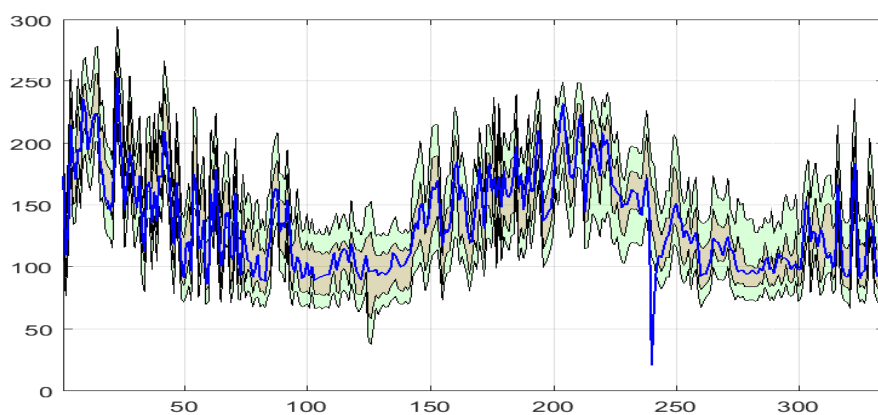


Рис. 12. Межквантильные интервалы (полутона) и фактические нагрузки (синяя линия) в 24 час. (полночь).

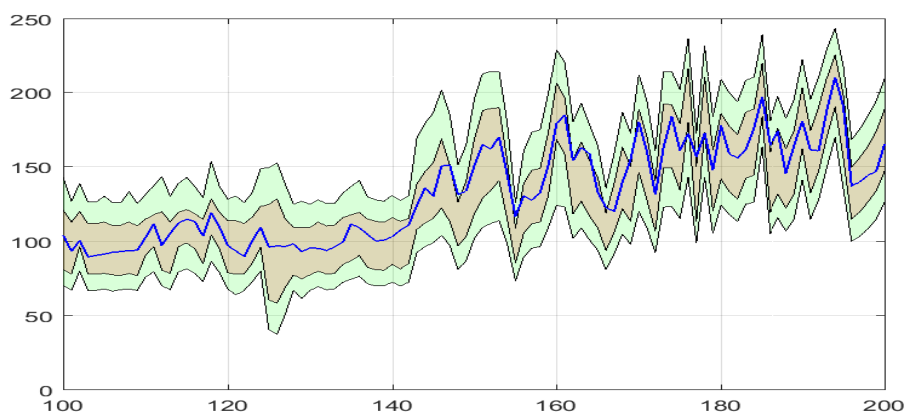


Рис. 13. Межквантильные интервалы $[0.25, 0.75]$ и $[0.05, 0.95]$ (полутона) и фактические нагрузки в 24 часа – полночь (сплошная линия). Фрагмент рис. 12 с 10 марта 2011 года до 20 июля того же года.

На увеличенных фрагментах сплошная линия соединяет истинные (измеренные) значения электрически нагрузок на соответствующий час суток. Широкая полутоновая полоса соответствует межквантильному интервалу $[0.05, 0.95]$. Внутри нее серая полоса соответствует межквантильному интервалу $[0.25, 0.75]$. Видно, что в основном серая полоса накрывает измеренные значения, но бывают и выбросы.

5. ЗАКЛЮЧЕНИЕ

В данной работе рассматривается проблема онлайн агрегирования вероятностных прогнозов. В этом случае популярным примером корректного скорингового правила для непрерывных исходов является скоринговое правило непрерывной ранжированной вероятности CRPS.

Представлен теоретический анализ возможностей использования функции непрерывной ранжированной вероятности CRPS для получения агрегированных прогнозов с использованием экспертных стратегий.

В работах [5] и [6] было показано, что функция потерь CRPS является смешиваемой и экспоненциально вогнутой. В данной работе агрегирующий алгоритм были введены дополнительные параметры, которые характеризуют уровни доверия к прогнозам экспертов. На основе этих свойств получено обобщение агрегирующего алгоритма АА [14] на случай вероятностных прогнозов с использованием уровней доверия.

Представлены два правила вычисления прогнозов (27) и (28) на основе алгоритмов АА и WA. Получены не зависящие от времени верхние границы дисконтированного с учетом уровней доверия регрета агрегирующих алгоритмов АА и WA.

Теоретические выводы проиллюстрированы результатами компьютерных экспериментов. Материалом для экспериментального исследования послужили данные конкурса Global Energy Forecasting – Competition 2014 (GEFCOM 2014, Track Load), который проводился на платформе Kaggle ([16]).

Разработана технология построения прогнозирующих экспертов и агрегирования их прогнозов с учетом уровней доверия.

Результаты численных экспериментов, проведенных на основе предложенных методов, показывают, что использование правил агрегирования прогнозов экспертов АА (27) и (28) приводят к существенно меньшим потерям, чем потери самих экспертов.

СПИСОК ЛИТЕРАТУРЫ

1. N. Cesa-Bianchi, G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.
2. Y. Freund, R.E. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*. 55:119–139, 1997.
3. 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.
4. N. Littlestone, M. Warmuth. The weighted majority algorithm. *Information and Computation*. 108:212–261, 1994.
5. В.В. Вьюгин, В.Г. Трунов. Онлайн агрегирование вероятностных прогнозов на основе скорингового правила непрерывной ранжированной вероятности. *Информационные процессы*, 19 (2), 92–112, 2019.
6. V. V'yugin, V. Trunov. Online Learning with Continuous Ranked Probability Score, *Proceedings of Machine Learning Research* 105: 163–177, 2019.
7. A. Jordan, F. Krüger, S. Lerch. Evaluating Probabilistic Forecasts with scoringRules, arXiv:1709.04743.

8. J. Bröcker, L.A. Smith. Scoring probabilistic forecasts: The importance of being proper. *Weather and Forecasting* 22. 382–388, 2007.
9. J. Bröcker, L.A. Smith. From ensemble forecasts to predictive distribution functions. *Tellus A* 60. 663–678, 2008.
10. J. Bröcker. Evaluating raw ensembles with the continuous ranked probability score. *Q. J. R. Meteorol. Soc.* 138, 1611–1617, July 2012 B
11. A.E. Raftery, T. Gneiting, F. Balabdaoui, M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* 133, 1155–1174, 2005.
12. K. Bogner, K. Liechti, M. Zappa. Technical note: Combining quantile forecasts and predictive distributions of streamflows. *Hydrol. Earth Syst. Sci.*, 21, 5493–5502, 2017.
13. J. Thorey, V. Mallet and P. Baudin. Online learning with the Continuous Ranked Probability Score for ensemble forecasting. *Quarterly Journal of the Royal Meteorological Society*, 143, 521–529, January 2017 A DOI:10.1002/qj.2940
14. V. Vovk, A game of prediction with expert advice. *Journal of Computer and System Sciences*. 56(2), 153–173, 1998.
15. V. V'yugin, V. Trunov. Online aggregation of unbounded losses using shifting experts with confidence. *Machine Learning*, 108(3): 425–444, 2019.
16. 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.
17. V. Vovk. Competitive on-line statistics. *International Statistical Review* 69, 213–248, 2001.
18. E.S. Epstein. A scoring system for probability forecasts of ranked categories. *J. Appl. Meteorol. Climatol.*, 8: 985–987, 1969.
19. G.W. Brier. Verification of forecasts expressed in terms of probabilities. *Mon. Weather Rev.* 78. 1–3, 1950.
20. I.J. Good. Rational Decisions. *Journal of the Royal Statistical Society B*, 14(1), 107–114, 1952. <https://www.jstor.org/stable/2984087>.
21. J.E. Matheson, R.L. Winkler. Scoring Rules for Continuous Probability Distributions. *Management Science*, 22(10), 1087–1096, 1976. doi:10.1287/mnsc.22.10.1087.
22. A. Blum, Y. Mansour. From external to internal regret. *Journal of Machine Learning Research*. 8:1307–1324, 2007.
23. A. Chernov and V. Vovk. Prediction with expert evaluators advice. In *Algorithmic Learning Theory, ALT 2009, Proceedings*, volume 5809 of LNCS, pages 8–22. Springer, 2009.
24. M. Herbster, M. Warmuth. Tracking the best expert. *Machine Learning*, 32(2): 151–178, 1998.
25. <https://www.kaggle.com/datasets>
26. M. Devaine, P. Gaillard, Y. Goude, G. Stoltz. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning*. 90(2): 231–260, 2013.
27. G. McLachlan, and D. Peel. *Finite Mixture Models*. Hoboken, NJ: Wiley and Sons, 2000.
28. Z.I. Botev, J.F. Grotowski, and D.P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38(5): 2916–2957, 2010.

Online Aggregation of Probabilistic Forecasts of Hourly Electrical Loads

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

The problem of probabilistic forecasting of a time series on-line is considered. The difference between the probabilistic forecast (distribution function) and the numerical outcome is measured using the loss function (scoring rule). In practical statistics, to assess the discrepancy between probabilistic forecasts and quantitative outcomes often use the scoring rule of continuous ranked probability (Continuous Ranked Probability Score – CRPS). The paper considers the case when several competing methods (experts) give their forecasts in the form of probability distribution functions. Forecasts provided by experts levels of confidence. An algorithm for online aggregation of these distribution functions taking into account levels of confidence of expert forecasts is proposed. Upper bounds of the regret of the proposed algorithm discounted taking into account confidence levels are obtained in the form of a comparison of cumulative losses of the algorithm loss of expert hypotheses. A technology has been developed for constructing predictive expert algorithms and aggregating their probabilistic forecasts using a tasks of forecasting electricity consumption for one or more hours in advance. The results of numerical experiments on real data are presented.

Keywords: predictions with expert advice, online probabilistic forecasting, regret, confidence levels, aggregation algorithm, continuous ranked probability score, CRPS