Информационные процессы, Том 18, № 3, 2018, стр. 135–148

Онлайн агрегирование специализованных экспертных стратегий в задаче прогнозирования почасовых электрических нагрузок[1]

В.В. Вьюгин, В.Г. Трунов

Институт проблем передачи информации им. А.А. Харкевича РАН, Москва, Россия
Поступила в редколлегию 05.07.2018

Аннотация—Решается задача агрегирования в режиме онлайн прогнозов специализированных алгоритмов. В каждый момент времени каждый экспертный алгоритм выдает прогноз, сопровождаемый уровнем доверия, который представляет собой число между 0 и 1. Уровни доверия определяются на основе разделения временного ряда на области компетентности экспертных алгоритмов и предварительного их обучения внутри этих областей на основе ретроспективных данных. Рассмотрен адаптивный алгоритм, который агрегирует прогнозы экспертных стратегий и несет потери, не превосходящие (с точностью до некоторой величины, называемой регретом) потери наилучшей комбинации экспертов, распределенных по интервалу прогнозирования. Алгоритм применяется для прогнозирования почасовых электрических нагрузок на основе реальных данных.

КЛЮЧЕВЫЕ СЛОВА: предсказания с использованием экспертных стратегий, алгоритм AdaHedge, метод Fixed Share, эксперты специалисты, уровни доверия, прогнозирование потребления электроэнергии

1. ВВЕДЕНИЕ

Рассматривается задача агрегирования экспертных прогнозов в режиме онлайн (Prediction with Expert Advice – см. работы [18], [9], [20], [21] и др. Основная монография на эту тему – [6]). Имеется набор методов (экспертов, алгоритмов), которые на каждом шаге выдают прогнозы и несут потери вследствие расхождения этих прогнозов и исходов, которые появляются после того, как прогнозы были предъявлены.

В рамках данного подхода не делается никаких предположений о природе источника исходных данных. В целом, рассматривается повторяющаяся игра на предсказания, в которой каждая экспертная стратегия выдает свой прогноз, а агрегирующий алгоритм комбинирует эти прогнозы в один прогноз; после этого, предъявляется соответствующий исход и экспертные стратегии и агрегирующий алгоритм несут потери вследствие расхождения этих исходов и прогнозов.

На каждом раунде игры $t = 1, 2, \dots, T$ эксперты $i \in \{1, \dots, N\}$ представляют свои прогнозы $c_{i,t}$. После этого, агрегирующий алгоритм вычисляет свой прогноз γ_t. В простейшем случае этот прогноз равен выпуклой комбинации прогнозов экспертов $\gamma_t = \sum_{i=1}^N w_{i,t} c_{i,t}$, где $\mathbf{w}_t = (w_{1,t}, \dots, w_{N,t})$ – кумулятивные веса экспертов в момент времени t, $\sum\limits_{i=1}^N w_{i,t} = 1$ и $w_{i,t} \ge 0$ для всех i.

[1] Работа частично поддержана грантами РФФИ: проекты 16-01-00576 А и 16-29-09649 офи-м.

После того, как все прогнозы были сделаны, предъявляется исход ω_t, и все эксперты i и алгоритм вычисляют свои потери $l_{i,t} = \lambda(\omega_t, c_{i,t})$ и $a_t = \lambda(\omega_t, \gamma_t)$, где $\lambda(\omega, \gamma)$ – функция потерь. Рассматриваем выпуклые по γ функции потерь. В этом случае $a_t \le \sum\limits_{i=1}^{N} w_{i,t} l_{i,t} = h_t$.

Агрегирующий алгоритм модифицирует веса экспертных стратегий в конце каждого раунда игры используя потери экспертов. В классической постановке ([9], [20] и т.д.) используется метод экспоненциального взвешивания с постоянным параметром обучения η:

$$w_{i,t+1} = \frac{w_{i,t} e^{-\eta l_{i,t}}}{\sum\limits_{j=1}^{N} w_{j,t} e^{-\eta l_{j,t}}}. \tag{1}$$

Наша цель заключается в использовании такого метода взвешивания экспертных стратегий, который гарантирует что кумулятивные потери агрегирующего алгоритма не превосходят потери наилучшего эксперта или наилучшей выпуклой комбинации потерь экспертов.

Вектором сравнения называется такой вектор $\mathbf{q} = (q_1, \dots, q_N)$, для которого $q_1 + \dots + q_N = 1$ и все его компоненты неотрицательные. Мы будем сравнивать кумулятивные потери агрегирующего алгоритма и выпуклой комбинации кумулятивных потерь экспертов $\sum\limits_{t=1}^{T} (\mathbf{q} \cdot \mathbf{l}_t)$, где $\mathbf{l}_t = (l_{1,t}, \dots, l_{N,t})$ – вектор потерь экспертов на шаге t.

При более общем подходе, предложенном в [11], вектор сравнения изменяется со временем $\mathbf{q} = \mathbf{q}_t$ и мы сравниваем потери агрегирующего алгоритма с кумулятивной суммой $\sum\limits_{t=1}^{T} (\mathbf{q}_t \cdot \mathbf{l}_t)$. Важный частный случай состоит в следующем. Рассматриваются единичные векторы сравнения $\mathbf{q}_t$. Серия шагов, на которых делаются предсказания, делится на $k+1$ сегментов. В каждом таком сегменте выполнено $\mathbf{q}_t = \mathbf{q}_{t+1}$ для всех внутренних моментов времени t, т.е. такому сегменту соответствует один эксперт. Последовательность сегментов и соответствующих экспертов называется составным экспертом. Цель алгоритма изменяется – теперь он должен предсказывать так, чтобы быть не хуже каждого составного эксперта. Соответственно, модифицируется понятие регрета алгоритма – теперь это разность между потерями алгоритма и потерями наилучшей последовательности экспертов. Данное изменение позволяет точнее моделировать условия реальной жизни, когда природа исходов может меняться со временем и разные эксперты могут предсказывать с разной степенью успешности в зависимости от текущего тренда. Соответствующий алгоритм называется Fixed Share [11].

В работе [4] было предложено обобщение метода Fixed Share – метод MPP (Mixing Past Posteriors).

В большей части работ в области предсказания с использованием экспертных стратегий предполагается, что потери экспертных стратегий ограничены или используются функции потерь специального вида (см. [20], [6]). Например, предполагается, что $l_{i,t} \in [0, 1]$ для всех i и t.

Подобное ограничение неудобно в ряде практических приложений. В этой работе не предполагается ограниченность потерь экспертов. Алгоритм AdaHedge, предложенный в работе [19], свободен от подобных ограничений. Этот алгоритм представляет собой версию классического алгоритма Hedge [9], который настраивает параметр η в режиме онлайн в зависимости от результатов обучения агрегирующего алгоритма.

В работе [19] получена верхняя оценка регрета этого алгоритма. Пусть $l_t^- = \min\limits_{1 \le i \le N} l_{i,t}$ и $l_t^+ = \max\limits_{1 \le i \le N} l_{i,t}$ – наименьшие и наибольшие потери, понесенные экспертами на шаге t. При этом, не накладывается ограничений на величину и знак этих потерь. Определим $L_T^+ = \sum\limits_{t=1}^{T} l_t^+$ и $L_T^- = \sum\limits_{t=1}^{T} l_t^-$. Пусть также, $s_t = l_t^+ - l_t^-$ и $S_T = \max\{s_1, \ldots, s_T\}$.

Пусть $L_T^* = \min_{1\le i\le N} L_T^i$ – кумулятивные потери наилучшего эксперта и $H_T = \sum_{t=1}^T h_t$ – кумулятивные потери агрегирующего алгоритма за первые T шагов. Разность $R_T = H_T - L_T^*$ называется регретом.

В работе [19] получена верхняя оценка регрета алгоритма AdaHedge

$$R_T \le 2\sqrt{S_T \frac{(L_T^* - L_T^-)(L_T^+ - L_T^*)}{L_T^+ - L_T^-} \ln N} + \left(\frac{16}{3}\ln N + 2\right) S_T. \tag{2}$$

Если потери экспертов на каждом шаге ограничены, эта оценка приобретает форму $O(\sqrt{T\ln N})$.

В этой работе мы соединяем преимущества алгоритмов MPP (Fixed Share) и AdaHedge. Заметим, что версии алгоритмов MPP и Fixed Share, представленные в работах [11] и [4], использует постоянный параметр обучения. В то же время, алгоритм AdaHedge использует переменный настраиваемый в процессе работы (адаптивный) параметр обучения. В данной работе предлагается версия алгоритма Fixed Share с адаптивным параметром обучения. Благодаря этому, мы можем объединить алгоритмы AdaHedge и Fixed Share и получить верхнюю оценку регрета, подобную оценке (2).

В работе будет рассмотрено приложение предложенного алгоритма для последовательного краткосрочного (на один час вперед) прогнозирования электрических нагрузок. Для построения подобных прогнозов в работе [8] был использован метод специализованных (или спящих) экспертов. На каждом шаге только некоторые из экспертов активны и выдают предсказания, а остальные эксперты не активны и предсказаний не выдают. Анализ предметной области показывает, что пространство входных переменных можно разделить на области и выделить экспертов, которые выдают релевантные прогнозы только внутри этих областей. Например, эксперты могут специализироваться по сезонам, времени суток, по значению температуры.

Впервые метод агрегирования специализированных экспертов был предложен в работе [10] и изучался в работах [1], [5], [8], [16]. Согласно этому подходу, на каждом шаге t задается множество активных экспертов $E_t \subseteq \{1,\ldots,N\}$. Специализированный эксперт i выдает предсказания только на тех шагах $t = 1,2,\ldots,$ на которых $i \in E_t$. Агрегирующий алгоритм использует для построения своего прогноза на шаге t только прогнозы активных (не спящих) экспертов.

В данной работе мы добавляем к комбинации Fixed Share–AdaHedge “гладкое” обобщение метода специализированных экспертов. В каждый момент времени t предсказание каждого эксперта i сопровождается уровнем доверия, который представляет собой вещественное число $p_{i,t} \in [0,1]$. Подобная постановка была предложена в работе [3] и использовалась в работах [7], [14], [15].

В частности, $p_{i,t} = 1$ означает, что прогноз эксперта используется полностью, тогда как в случае $p_{i,t} = 0$ он вообще не принимается во внимание (эксперт не активен). В случае $0 < p_{i,t} < 1$ прогноз эксперта принимается во внимание только частично.

Например, с при переходе от одного времени года к другому, соответствующий специализированный эксперт постепенно теряет свою способность делать релевантные предсказания электрических нагрузок, поэтому естественно постепенно понижать уровень доверия к прогнозам этого эксперта.

В разделе 2 работы представлен алгоритм типа Fixed Share–AdaHedge, который использует уровни доверия к предсказаниям экспертов.

Верхние оценки регрета алгоритма представлены в теореме 1. Оценки даны для наихудшего случая, когда не делается никаких предположений о природе источника исходных данных.

В разделе 3 приведена технология построения экспертных стратегий, их обучения и определения уровней доверия экспертных прогнозов. Также представлены результаты численных экспериментов по краткосрочному прогнозированию электрических нагрузок по реальным данным и с использованием построенных алгоритмов.

Предложенный подход с использованием уровней доверия является более общим, чем подход с использованием специализированных экспертов из работы [8].

2. АЛГОРИТМ АГРЕГИРОВАНИЯ ЭКСПЕРТНЫХ ПРОГНОЗОВ

Потери экспертов будут вычисляться с помощью выпуклой по прогнозу γ функции потерь $\lambda(\omega,\gamma)$. Рассматриваем исходы $\omega \in \Omega$, где Ω – произвольное множество и прогнозы $\gamma \in G$, где G – некоторое линейное пространство. В экспериментах раздела 3 будет использоваться абсолютная функция потерь $\lambda(\omega,\gamma) = |\omega-\gamma|$, где ω и γ – произвольные неотрицательные вещественные числа.

Пусть на шаге t заданы прогнозы экспертов $\mathbf{c}_t = (c_{1_t},\dots,c_{N,t})$ и уровни доверия к этим прогнозам $\mathbf{p}_t = (p_{1,t},\dots,p_{N,t})$, где $p_{i,t} \in [0,1]$ для всех $1 \le i \le N$. Определим вспомогательные рандомизированные предсказания экспертов

$$\tilde{c}_{i,t} = \begin{cases} c_{i,t} \text{ с вероятностью } p_{i,t}, \\ \gamma_t \text{ с вероятностью } 1-p_{i,t}, \end{cases}$$

где γ_t – прогноз агрегирующего алгоритма.

После того как исход ω будет предъявлен, можно вычислить виртуальные потери экспертов $\tilde{l}_{i,t} = \lambda(\omega_t,\tilde{c}_{i,t})$. Математические ожидания виртуальных прогнозов и потерь экспертов равны $\hat{\mathbf{c}}_{i,t} = E_{\mathbf{p}_t}(\tilde{c}_{i,t}) = p_{i,t}c_{i,t} + (1-p_{i,t})\gamma_t$ и

$$\hat{l}_{i,t} = E_{\mathbf{p}_t}(\tilde{l}_{i,t}) = \sum_{i=1}^{N} p_{i,t}\lambda(\omega_t, c_{i,t}) + (1-p_{i,t})\lambda(\omega_t,\gamma_t)$$

при $1 \le i \le N$. Веса экспертов сначала модифицируется по формуле

$$w^m_{i,t} = \frac{w_{i,t}e^{-\eta\hat{l}_{i,t}}}{\sum_{s=1}^{N} w_{s,t}e^{-\eta\hat{l}_{s,t}}}, \tag{3}$$

а затем происходит модификация согласно методу Fixed Share. Окончательно, получаем вес $w_{i,t+1}$ (см. алгоритм ConfHedge).

Прогноз агрегирующего алгоритма определяется

$$\gamma_t = \sum_{i=1}^{N} w_{i,t}\hat{c}_{i,t}. \tag{4}$$

Используем метод неподвижной точки из работы [5] для устранения логического цикла в этих определениях:

$$\gamma_t = \sum_{i=1}^{N} w_{i,t}\hat{c}_{i,t} = \sum_{i=1}^{N} w_{i,t}(p_{i,t}c_{i,t} + (1-p_{i,t})\gamma_t) = \sum_{i=1}^{N} w_{i,t}p_{i,t}(c_{i,t}-\gamma_t) + \gamma_t.$$

Сокращаем одинаковые слагаемые из левой и правой частей этого равенства и получаем формулу для вычисления прогноза γ_t агрегирующего алгоритма:

$$\gamma_t = \frac{\sum_{i=1}^N p_{i,t} w_{i,t} c_{i,t}}{\sum_{i=1}^N p_{i,t} w_{i,t}}. \tag{5}$$

После того как γ_t вычислено, равенства (5) и (4) эквивалентны. Вычислим математическое ожидание потерь экспертов $\hat{l}_{i,t} = E_{\mathbf{p}_t}(\tilde{l}_{i,t}) = p_{i,t}\lambda(\omega_t, c_{i,t}) + (1-p_{i,t})\lambda(\omega_t, \gamma_t)$.

Выпуклая комбинация потерь экспертов равна $h_t = \sum_{i=1}^N w_{i,t}\hat{l}_{i,t} = \sum_{i=1}^N w_{i,t}(p_{i,t}l_t^i + (1-p_{i,t})a_t)$ (потери алгоритма распределения потерь экспертов).

После того как прогноз γ_t вычислен по формуле (5), веса экспертов $w_{i,t}^m$ вычисляются по формуле (3) используя ранее вычисленные значения $\hat{l}_{i,t}$.

Алгоритм ConfHedge

Полагаем $w_{i,1} = w_{i,0}^m = \frac{1}{N}$ при $i = 1, \ldots, N$, $\Delta_0 = 0$, $\eta_1 = \infty$.
FOR $t = 1, \ldots, T$
Получаем прогнозы экспертов $\mathbf{c}_t = (c_{1,t}, \ldots, c_{N,t})$ и соответствующие уровни доверия $\mathbf{p}_t = (p_{1,t}, \ldots, p_{N,t})$.
Вычисляем прогноз агрегирующего алгоритма $\gamma_t = \frac{\sum_{i=1}^N p_{i,t}w_{i,t}c_{i,t}}{\sum_{i=1}^N p_{i,t}w_{i,t}}$.
Получаем исход ω_t и вычисляем потери экспертов $\mathbf{l}_t = (l_{1,t}, \ldots, l_{N,t})$, где $l_{i,t} = \lambda(\omega_t, c_{i,t})$, $1 \le i \le N$, и потери агрегирующего алгоритма $a_t = \lambda(\omega_t, \gamma_t)$.
Модифицируем веса экспертов и параметр обучения:
Loss Update
Определим
$w_{i,t}^m = \frac{w_{i,t}e^{-\eta_t p_{i,t}(l_{i,t}-a_t)}}{\sum\limits_{s=1}^N w_{s,t}e^{-\eta_t p_{s,t}(l_{s,t}-a_t)}}$ for $1 \le i \le N$.
Mixing Update
Определяем веса экспертов для использования на следующем шаге
$w_{i,t+1} = \frac{\alpha_{t+1}}{N} + (1-\alpha_{t+1})w_{i,t}^m$ для всех $1 \le i \le N$, где $\alpha_t = \frac{1}{t}$ для всех t.
Learning Parameter Update
Определим $m_t = -\frac{1}{\eta_t}\ln\sum_{i=1}^N w_{i,t}e^{-\eta_t(p_{i,t}l_{i,t})+(1-p_{i,t})a_t)}$ и $\delta_t = h_t - m_t$, где $h_t = \sum_{i=1}^N w_{i,t}(p_{i,t}l_{i,t} + (1-p_{i,t})a_t)$, определим также $\Delta_t = \Delta_{t-1} + \delta_t$. Полагаем $\eta_{t+1} = \frac{\ln N}{\Delta_t}$.
ENDFOR

Из выпуклости функции потерь выполнено $a_t = \lambda(\omega_t, \gamma_t) = \lambda(\omega_t, \sum_{i=1}^N w_{i,t}\hat{c}_{i,t}) \le \sum_{i=1}^N w_{i,t}\hat{l}_{i,t} = h_t$. Пусть $H_T = \sum_{t=1}^T h_t$, $M_T = \sum_{t=1}^T m_t$, где $m_t = -\frac{1}{\eta_t}\ln\sum_{i=1}^N w_{i,t}e^{-\eta_t\hat{l}_{i,t}}$. Из выпуклости экспоненты $m_t \le h_t$, поэтому $\delta_t \ge 0$.

В теории предсказаний с использованием экспертных прогнозов эффективность работы агрегирующего алгоритма (в данном случае это ConfHedge) оценивается для произвольной последовательности исходов (в наихудшем случае). При этом, не делается никаких упрощающих предположений о природе источника этих исходов. Критерием эффективности агрегирующего алгоритма является разность между его кумулятивными потерями на произольной последовательности исходов и потерями наилучшего эксперта. Эта разность называется регретом алгоритма. Получаем оценки на регрет, не зависящие от последовательности исходов.

Далее, в теореме 1 представим две оценки на регрет алгоритма ConfHedge. Определим $L_T^{(\mathbf{q})} = \sum_{t=1}^T(\mathbf{q}_t \cdot \hat{\mathbf{l}}_t) = \sum_{t=1}^T\sum_{i=1}^N q_{i,t}(p_{i,t}l_{i,t} + (1-p_{i,t})h_t)$. Нижнюю и верхнюю границы этой суммы можно записать в виде $L_T^{(\mathbf{q}-)} \le L_T^{(\mathbf{q})} \le L_T^{(\mathbf{q}+)}$, где

$$L_T^{(\mathbf{q}-)} = \sum_{t=1}^{T}\sum_{i=1}^{N} q_{i,t}(p_{i,t}l_{i,t} + (1-p_{i,t})l_t^-),$$

$$L_T^{(\mathbf{q}+)} = \sum_{t=1}^{T}\sum_{i=1}^{N} q_{i,t}(p_{i,t}l_{i,t} + (1-p_{i,t})l_t^+).$$

В работах [19] и [23] были получены верхние оценки для величины Δ_T. В работе [23] также была получена оценка

$$M_T - L_T^{(\mathbf{q})} \le ((k+2)\ln T + k + 1)\Delta_T. \tag{6}$$

Из определения $H_T = M_T + \Delta_T$, поэтому $H_T - L_T^{(\mathbf{q})} \le \gamma_{k,T}\Delta_T$, где $\gamma_{k,T} = (k+2)(\ln T + 1)$.

Так как $a_t \le h_t$, получаем $\sum_{t=1}^{T} a_t \le H_T$. Из определения $a_t - (q_t \cdot \hat{l}_t) = \sum_{i=1}^{N} q_{i,t}p_{i,t}(a_t - l_{i,t})$.

Тогда регрет алгоритма равен

$$R_T^{(\mathbf{q})} = \sum_{t=1}^{T} a_t - L_T^{(\mathbf{q})} = \sum_{t=1}^{T}\sum_{i=1}^{N} q_{i,t}p_{i,t}(a_t - l_{i,t}).$$

В результате, из неравенства (6) и оценок для Δ_T получаем верхние оценки регрета (гарантийные оценки потерь алгоритма ConfHedge).

Теорема 1. *Для любого T и для любой последовательности векторов сравнения с не более чем k изменениями*

$$R_T^{(\mathbf{q})} \le \frac{1}{2}\gamma_{k,T}\sqrt{\sum_{t=1}^{T} s_t^2 \ln N} + \gamma_{k,T}\left(\frac{2}{3}\ln N + 1\right)S_T, \tag{7}$$

$$R_T^{(\mathbf{q})} \le \gamma_{k,T}\sqrt{S_T\frac{(L_T^+ - L_T^{(\mathbf{q}-)})(L_T^{(\mathbf{q}+)} - L_T^-)}{L_T^+ - L_T^-}\ln N} + \gamma_{k,T}\left(\left(\gamma_{k,T} + \frac{2}{3}\right)\ln N + 1\right)S_T, \tag{8}$$

где $\gamma_{k,T} = (k+2)(\ln T + 1)$.

Оценка (7) представляет собой обобщение оценки из работы [7], оценка (8) обобщает оценку (16) теоремы 8 из [19] на случай переменных векторов сравнения.[2]

Рассмотрим один важный частный случай предложенной схемы смешивания экспертных прогнозов. Пусть все векторы сравнения являются единичными: $\mathbf{q}_t = \mathbf{e}_{i_t}$ и $\mathbf{q}_{t-1} \ne \mathbf{q}_t$ для не более чем k шагов $1 < t \le T$. Рассматривается соответствующая цепочка экспертов (составной эксперт) $i_1, \dots, i_T$. Пусть также заданы уровни доверия $p_{i,t} \in \{0,1\}$, то есть рассматривается случай экспертов специалистов. Левая часть неравенства (8) имеет вид $\sum\limits_{t: p_{i_t,t}=1}^{T}(h_t - l_{i_t,t})$. Таким образом, для сравнения потерь алгоритма и потерь произвольной цепочки экспертов $i_1, \dots, i_T$

[2] Доказательство теоремы 1 в данной работе не приводится. Основная цель этой работы заключается в демонстрации практических приложений алгоритма ConfHedge.

в (7) и (8) учитываются только те моменты времени, когда соответствующие эксперты из этой цепочки были активны.

Соответствующие разности из правой части неравенства (8) имеют вид $L_T^+ - L_T^{(\mathbf{q}-)} = \sum\limits_{t:p_{i_t,t}=1}^{T}(l_t^+ - l_{i_t,t}) + \sum\limits_{t:p_{i_t,t}=0}^{T} s_t$ и $L_T^{(\mathbf{q}+)} - L_T^- = \sum\limits_{t:p_{i_t,t}=1}^{T}(l_t^+ - l_{i_t,t}) + \sum\limits_{t:p_{i_t,t}=0}^{T} s_t$.

3. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

При построении прогностических моделей, работающих в режиме онлайн и предназначенных для восстановления отклика, зависящего как от времени, так и от дополнительных экзогенных параметров, иногда можно заранее на основании физического смысла задачи выявить области пространства входных переменных, для которых изменения зависимой переменной от входных переменных заведомо описываются однородными моделями.

Рассмотрим в качестве примера задачу прогнозирования электрических нагрузок в локальной энергораспределительной сети. Так, зимой энергопотребление возрастает из-за включения обогревателей при похолодании, а летом из-за работы кондиционеров при повышении температуры; не менее значимое влияние на характер зависимостей оказывают такие факторы, как время года, время суток, календарные параметры (рабочие, выходные и праздничные дни). Это означает, что при формировании ансамбля прогностических моделей можно ориентироваться не только на универсальные алгоритмы прогнозирования, обучающиеся на всей обучающей выборке и выдающие свои прогнозы во всех точках пространства переменных, но и заранее выбрать области в пространстве входных переменных, для которых нужно строить специализированные модели на основе обучающих образцов, выбранных именно из этих областей, ориентированные на построение только локальных прогнозов. Может оказаться, что при таком подходе более простые модели смогут обеспечить более точные и надежные прогнозы при меньших объемах обучающих данных. При этом возникает задача формирования окончательного прогноза путем гладкого агрегирования разнородных моделей с учетом предыстории, текущего времени и изменяющихся внешних условий (экзогенных параметров).

Задачу прогнозирования и принятия решения ансамблем разнородных экспертов, каждый из которых выдает свой прогноз только в некоторой ограниченной области пространства переменных и ограниченном интервале времени, называют задачей с учетом спящих экспертов [8]. В таких задачах множество элементарных предсказателей (экспертов) расширяется за счет добавления экспертов-специалистов, которые принимают и выдают свои решения только при соблюдении некоторых, заранее заданных условий и спят, когда эти условия не выполняются.

В работе [13] предлагаются подходы к формированию ансамбля экспертов-специалистов, настроенных на конкретные сценарии типа жара-похолодание, солнечные-пасмурные дни и т.д. Можно можно предположить,что при достаточном разнообразии сценариев, соответствующие эксперты смогут уловить различные эффекты в потреблении энергии, которые затем могут быть объединены путем агрегирования этих экспертов.

В рамках данной работы предполагается, что выбор областей компетентности и алгоритмов построения прогностических моделей заранее выполняется специалистом в данной предметной области на основании его опыта и интуиции. Описание каждого эксперта включает следующие характеристики: алгоритм обучения (в данном эксперименте использовали два типа экспертов – прогностических моделей, а именно алгоритм случайных деревьев Random Forest и простую пошаговую линейную регрессию); правила формирования уровня доверия для данного эксперта (временные и календарные интервалы, описание областей в пространстве входных переменных, где данный алгоритм является компетентным); список предикторов (независимых переменных, используемых в данной модели).

Материалом для вычислительных экспериментов послужили данные конкурса GefCom2012, который проводился на платформе Kaggle [12]. Основная задача конкурса состояла в прогнозировании суточного хода почасовых электрических нагрузок в 20 разных регионах по данным температурных измерений на 11 метеорологических станциях [12]. В рамках конкурса для решения задачи применялись различные подходы и алгоритмы прогнозирования, наиболее успешные представлены в обзорной работе [12]. Базы данных доступны на сайте https://www.kaggle.com/datasets.

Основные исходные данные для конкурса GefCom2012 были предоставлены в виде таблицы TemperatureHistory с архивными записями температуры, измеренной на 11 метеорологических станциях, и таблицы LoadHistory с данными почасовой электрической нагрузки, измеренными на 20 энергораспределительных станциях региона. Данные были получены за каждый час периода с 01.01.2004 г. по 30.06.2008 г. Кроме того, можно было использовать дополнительную информацию о временах года, днях недели, а также рабочих и праздничных днях.

Для иллюстрации работы предложенных в статье методов агрегирования экспертов-специалистов выбрали упрощенную частную задачу, а именно, прогнозирование электрических нагрузок в одной из электрораспределительных сетей (Zone5) на один час вперед на основании предыстории и текущих календарных параметров. Для учета температурных изменений в качестве предикторов использовали предысторию измерений температуры только одной метеорологической станции (Station9), данные которой позволяли наилучшим образом предсказывать электрическую нагрузку в выбранной сети на обучающей части выборки.

Предварительная обработка исходных таблиц состояла в синхронизации строк матриц по времени, их объединении, а также добавлении календарных параметров (дни недели, рабочие, выходные или праздничные дни, времена года и т.д.). В результате предобработки формируется синхронизированный многомерный временной ряд, дополненный предысторией с некоторым заданным лагом для расширения пространства предикторов, а также текущей календарной информацией (дни недели, рабочие, выходные и праздничные дни). Для построения прогностических моделей использовали алгоритмы, которые, судя по результатам [12], хорошо зарекомендовали себя при решении задач данного конкурса.

В рамках рассматриваемой модельной задачи предполагается, что каждый час алгоритмы (эксперты) выдают свои прогнозы электрических нагрузок, а также их достоверности, определяемые вектором P_t размерности N, где N – число экспертов. Считаем, что вектор P_t становится известен заранее, перед выполнением агрегирования решений экспертов. На рис. 1 приведены для примера почасовые измерения электрической нагрузки в выбранной сети (Zone5) за май месяц 2005 г.

Как видим, реальные почасовые измерения отражают как периодичность по времени суток, так и существенно стохастический характер измеряемого показателя.

Следующая группа рисунков может служить обоснованием выбора ансамбля экспертов-специалистов (на основе формирования ограниченных областей в пространстве входных переменных). Так, на рис. 2 показаны усредненные кривые суточного хода электрических нагрузок за каждый из 4 сезонов 2004 – 2005 гг. Ширина полутоновых полос вокруг каждой кривой соответствует оценке ошибок среднего значения. Сплошными линиями показаны усредненные результаты для рабочих дней, а прерывистыми те же оценки для выходных дней указанных периодов. Следует обратить внимание на наличие явных различий в парах кривых на каждом из рисунков. Тем самым, оправдано применение разных экспертов (отдельных прогностических моделей) для рабочих и выходных дней недели.

Предварительный анализ данных для прогноза почасовых электрических нагрузок показывает, что основные временные и календарные факторы, влияющие на изменение характера потребления электроэнергии, это время года, рабочие или выходные дни и время суток.

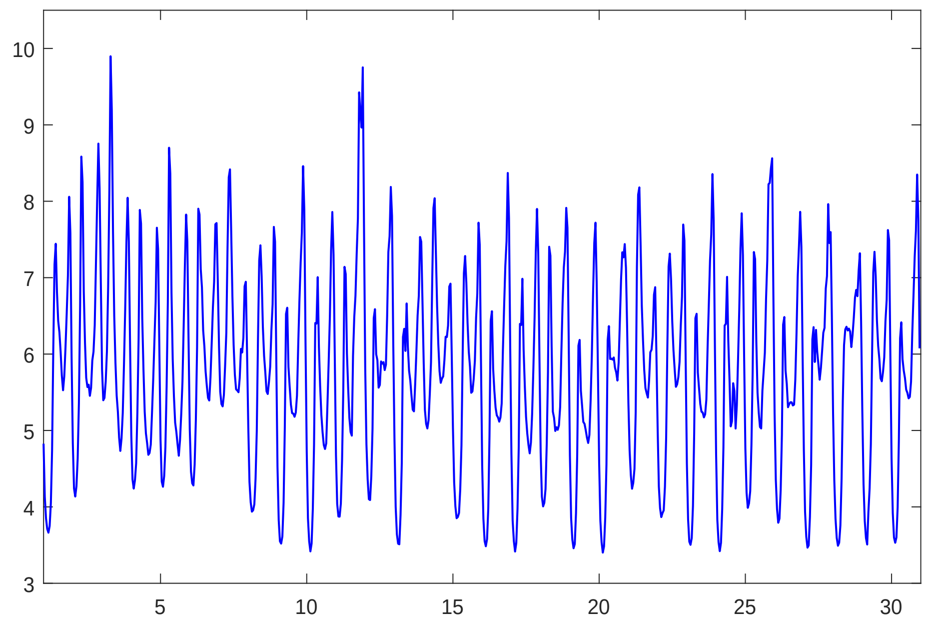

Рис. 1. График почасовых электрических нагрузок за все дни мая 2005 г.

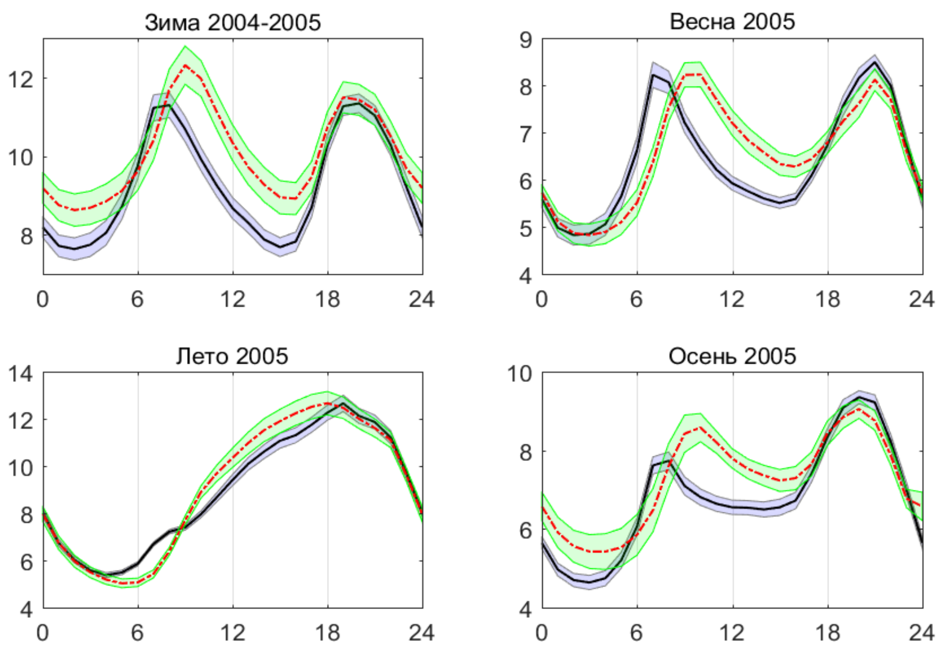

Рис. 2. Усредненные графики потребления электричества за каждый из четырех сезонов 2004–2005 гг. Полутоновая полоса вокруг каждой кривой отображает стандартную ошибку среднего значения. Сплошной линией показано среднее потребление электричества в рабочие дни недели, а прерывистой кривой – в выходные дни.

Воспользуемся этими легко формулируемыми календарными показателями для формирования простой схемы формирования ансамбля экспертов-специалистов. В качестве примера, на

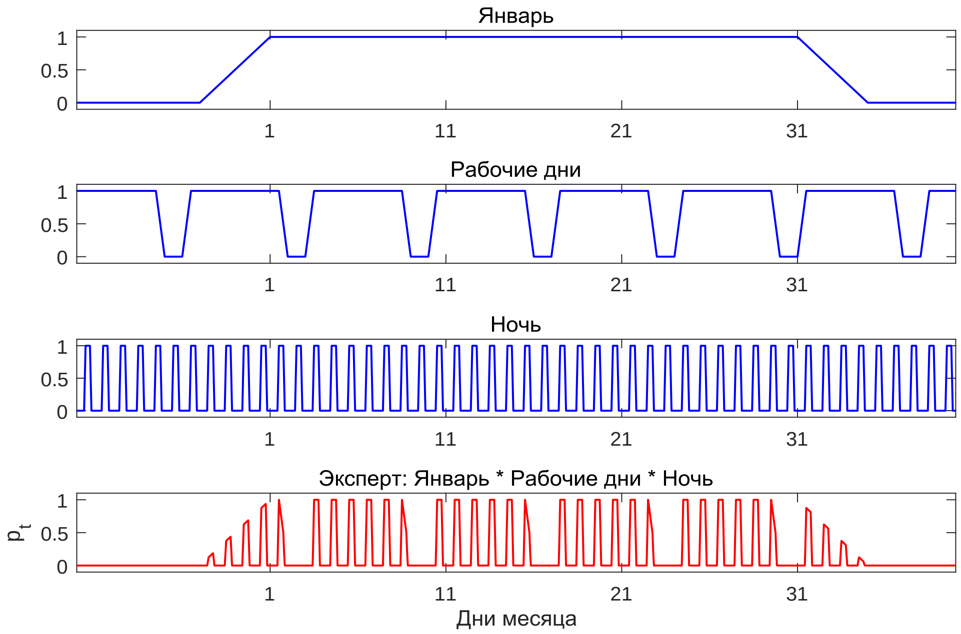

Рис. 3. Схема построения функции "компетентности" для эксперта типа Месяц года – Рабочие дни – Ночь.

рис. 3 показаны этапы построения функции компетентности для эксперта, ориентированного на формирование прогноза ночью с 0 часов до 6 утра в рабочие дни января месяца.

Заранее заданные функции компетентности экспертов используются на этапе обучения при формировании индивидуальных обучающих выборок для каждого эксперта, а на этапе прогнозирования – при агрегировании решений экспертов. Чтобы обеспечить более гладкое переключение между экспертами, функции принадлежности $p(t)$ формировали в виде трапеций, где на плато, соответствующем выбранному календарному интервалу, функция принимает значение 1, а на склонах линейно изменяется от 1 до 0. Ширина склонов задается параметром, зависящим от типа календарного параметра.

На этапе обучения (построения прогнозирующих моделей) для каждого из алгоритмов (экспертов) выполняются следующие действия: (1) По заданным для данного алгоритма правилам формирования уровня доверия, для всех элементов обучающей выборки вычисляется индекс компетентности, принимающий значения, близкие к 1, если данный образец с учетом значений предикторов подлежит рассмотрению данным экспертом, или близкие к 0, если эксперт на этом объекте не уверен в его надежности. На основании индекса компетентности для каждого алгоритма из полной обучающей выборки формируется индивидуальная обучающая выборка (корзина обучающих образцов). (2) На основании этой индивидуальной обучающей подвыборки производится построение прогностической модели.

На этапе тестирования прогностических алгоритмов на контрольной выборке данных построенные ранее модели из списка экспертов применяются к объектам тестирующей выборки для построения прогноза отклика. Для каждого объекта тестовой выборки, кроме собственно прогноза, оценивается уровень доверия, на основании которого агрегирующий алгоритм может решить, как использовать прогноз данного алгоритма (эксперта) при формировании окончательного мастер-прогноза. Заметим, что округляя индекс компетентности до значений

0,1, можно свести задачу гладкого смешивания к схеме агрегирования алгоритмов со спящими экспертами.

Схема онлайн оценки весов для смешивания спящих экспертов, предложенная в [8], основана на том, что на каждом шаге алгоритма спящим экспертам приписывают потери, равные средним потерям активных экспертов. При этом веса спящих экспертов остаются неизменными до тех пор, пока они не проснутся и не вступят в действие.

В настоящей работе предложен более гладкий вариант смешивания, когда все эксперты участвуют в принятии окончательного решения, но с учетом их индивидуальных уровней доверия, вычисляемых в каждой точке пространства входных переменных (предикторов).

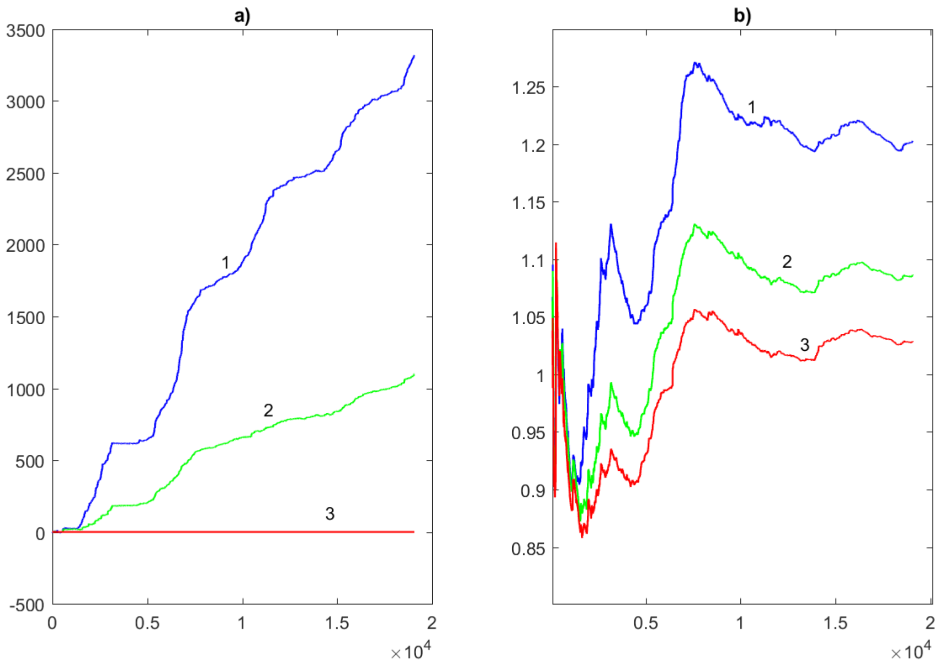

Рис. 4. (a) Эволюция разностей кумулятивных потерь $L^1_T - L^3_T$ и $L^2_T - L^3_T$: 1 – постоянно работающий неспящий эксперт (Random Forest), 2 – ConfHedge с использованием модели "спящих экспертов", 3 – ConfHedge с использованием гладких уровней доверия. (b) Средние кумулятивные потери (MAE) алгоритма Random Forest (1) и двух схем смешивания экспертов (2 и 3)

Для сравнения схемы гладкого смешивания со схемой спящих экспертов эксперименты по агрегированию экспертных решений проводили в два этапа. Сначала использовали только схему смешивания спящих и активных экспертов, т.е. уровень доверия принимал только два значения (0 или 1), а затем использовали алгоритм смешивания из раздела 2 данной работы.

Эволюция разностей кумулятивных потерь $L^1_T - L^3_T$ и $L^2_T - L^3_T$, где L^1_T – кумулятивные потери алгоритма Random Forest, который обучался на полной обучающей выборке, L^2_T, L^3_T – кумулятивные потери двух схем смешивания ("спящие эксперты" и "эксперты с гладкими уровнями доверия"), представлены на рис. 4a.

Средние абсолютные потери (MAE) $\frac{1}{T}L^1_T$ алгоритма Random Forest и двух схем смешивания: $\frac{1}{T}L^2_T$ и $\frac{1}{T}L^3_T$, представлены на рис 4b.[3]

[3] В экспериментах использовалась абсолютная функция потерь $\lambda(\omega,\gamma) = |\omega - \gamma|$.

4. ЗАКЛЮЧЕНИЕ

В данной работе, в рамках теории предсказания (принятия решений) с экспертами в режиме онлайн, предложен адаптивный алгоритм, который агрегирует решения экспертных стратегий и несет потери, не превосходящие (с точностью до некоторой величины, называемой регретом) потери наилучшей комбинации экспертов, произвольным образом распределенных по интервалу прогнозирования. При этом, единственным фактором, влияющим на величину регрета, является то, что число экспертов в такой комбинации не должно быть слишком большим. Для построения алгоритмов такого рода используется метод адаптивной модификации весов экспертов Fixed Share, который применяется к базовому алгоритму Hedge. Точнее, используется современная модификация этого алгоритма – алгоритм AdaHedge. Существенная особенность алгоритма AdaHedge заключается в том, что его основной параметр – параметр обучения – вычисляется адаптивно в процессе работы алгоритма (в режиме онлайн). Эта особенность определяет численную устойчивость алгоритма в условиях быстро изменяющегося лидерства экспертных стратегий.

Метод Fixed Share обычно формулируется для постоянного параметра обучения. В данной работе этот метод сформулирован для переменного параметра обучения, что позволило рассматривать его в комбинации с алгоритмом AdaHedge и получить оценку регрета комбинированного алгоритма.

Еще одним преимуществом алгоритма AdaHedge является также то, что он не требует ограниченности потерь экспертных стратегий и оценка его регрета проведена в отсутствие этих ограничений. Заметим также, что в рамках данного подхода не делается никаких стохастических предположений об источнике исходных данных. Все оценки регрета даются в терминах минимаксного типа.

В данной работе мы добавили к комбинации Fixed Share–AdaHedge “гладкое” обобщение метода специализированных экспертов. В каждый момент времени t предсказание каждого эксперта i сопровождается уровнем доверия $p_{i,t} \in [0,1]$.

В работе было рассмотрено приложение построенных алгоритмов для последовательного краткосрочного (на один час вперед) прогнозирования электрических нагрузок. Предложено обобщение предложенного в работе [8] подхода. Вместо специализованных (или спящих) экспертов, прогноз которых либо игнорируется либо полностью учитывается, рассмотрена “гладкая” характеристика учета прогноза эксперта – уровень доверия, с помощью которой прогноз эксперта может учитываться также частично.

Численные эксперименты по смешиванию экспертных решений, проведенных на реальных данных о величине электрических нагрузок, показали более высокую точность прогнозов при помощи алгоритма, учитывающего гладкие уровни доверия, по сравнению с алгоритмом, использующим метод специализированных экспертов.

СПИСОК ЛИТЕРАТУРЫ

1. D. Adamskiy, W.M. Koolen, A. Chernov, V. Vovk. A closer look at adaptive regret. *Lecture Notes in Artificial Intelligence*, 290–304, 2012.

2. C. Bishop. *Pattern Recognition and Machine Learning.* Springer, New York. 2006.

3. A. Blum, Y. Mansour. From external to internal regret. *Journal of Machine Learning Research*, 8:1307–1324, 2007.

4. O. Bousquet, M. Warmuth. Tracking a small set of experts by mixing past posteriors. *Journal of Machine Learning Research.* 3:363–396, 2002.

5. A. Chernov, V. Vovk. Prediction with expert evaluators' advice. *Lecture Notes in Artificial Intelligence.* 8–22, 2009.

6. N. Cesa-Bianchi, G. Lugosi. *Prediction, Learning, and Games.* Cambridge University Press. 2006.

7. N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning.* 66(2/3):321–352, 2007.

8. M. Devaine, P. Gaillard, Y. Goude, G. Stoltz. Forecasting electricity consumption by aggregating specialized experts. *Machine Learning.* 2013, 90 (2), 231–260.

9. 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.

10. Y. Freund, R.E. Schapire, Y. Singer, M.K. Warmuth. Using and combining pre- dictors that specialize. In: *Proc. 29th Annual ACM Symposium on Theory of Com- puting.* 334–343, 1997.

11. M. Herbster, M. Warmuth. Tracking the best expert. *Machine Learning*, 32(2) 151–178, 1998.

12. Tao Hong, P. Pinson, Shu Fan. Global Energy Forecasting Competition 2012, *International Journal of Forecasting* V.30(2): P.357–363, 2014.

13. P. Gaillard, Y. Goude. Forecasting electricity consumption by aggregating experts; how to design a good set of experts.*in Book Modeling and Stochastic Learning for Forecasting in High Dimensions*, Springer, 2015, pp. 95-?115.

14. P. Gaillard, Y. Goude, and G. Stoltz. A further look at the forecasting of the electricity consumption by aggregation of specialized experts. Technical report, 2011. pierre.gaillard.me/doc/GaGoSt-report.pdf.

15. P. Gaillard, G. Stoltz, and T. van Erven. A second-order bound with excess losses. In *JMLR: Workshop and Conference Proceedings.* 35, 176–196, 2014.

16. Y. Kalnishkan, D. Adamskiy, A. Chernov, T. Scarfe. Specialist Experts for Prediction with Side Information. *IEEE International Conference on Data Mining Workshop (ICDMW).* IEEE, 2015, 1470–1477.

17. J. Kivinen, M.K. Warmuth. Averaging expert prediction. In Paul Fisher and Hans Ulrich Simon, editors, *Computational Learning Theory: 4th European Conference (EuroColt '99)*, 153–167. Springer, 1999.

18. N. Littlestone, M. Warmuth. The weighted majority algorithm. *Information and Computation.* 108:212–261, 1994.

19. S. de Rooij, T. van Erven, P. Grunwald, W. Koolen. Follow the Leader. If You Can, Hedge If You Must. *Journal of Machine Learning Research.* 15:1281–1316, 2014.

20. 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.

21. V. Vovk, A game of prediction with expert advice. *Journal of Computer and System Sciences.* 56(2):153–173, 1998.

22. V. Vovk. Derandomizing stochastic prediction strategies. *Machine Learning.* 35(3):247–282, 1999.

23. V. V'yugin. Online Aggregation of Unbounded Signed Losses Using Shifting Experts. *Proceedings of Machine Learning Research* V.60:P.1–15, 2017. http://proceedings.mlr.press/v60/

Online aggregation of specialized experts strategies for short-term forecasting of electricity consumption

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

Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, 127051 Russia
e-mail: vyugin@iitp.ru

The problem of online aggregation of forecasts of specialized algorithms is considered. At each time moment, each expert algorithm issues a forecast accompanied by a confidence level, which is a number between 0 and 1. Confidence levels are determined on the basis of the division of the time series into areas of competence of the expert algorithms and their preliminary training within these areas using retrospective data. An adaptive algorithm is presented which aggregates forecasts of the expert strategies and suffer loss not exceeding (to within a certain value called regret) the loss of the best combination of experts distributed over the prediction interval. The algorithm is used for short-term prediction of the hourly electrical loads based on real data.

KEYWORDS: Prediction with Expert Advice, AdaHedge aalgorithm, Fixed Share method, Specialized experts, Confidence levels, Forecasting of electricity consumption