
МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ОБЧИСЛЮВАЛЬНІ МЕТОДИ

doi: <https://doi.org/10.15407/emodel.42.04.003>
УДК 519.246.8

Г.А. Кравцов¹, канд. техн. наук, **А.Н. Примушко**², **В.И. Кошель**¹, аспирант

¹ Институт проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины
(Украина, 03164, Киев, ул. Генерала Наумова, 15,
hryhoriy.kravtsov@gmail.com, vlad.koshell@gmail.com),

² Национальный технический университет Украины
«Киевский политехнический институт им. Игоря Сикорского»
(Украина, 03056, Киев, пр-т Победы, 37,
arsentiy.prymushko@gmail.com)

Комбинированный авторегрессионно-нейросетевой метод прогнозирования временных рядов

Изложен концептуальный подход к построению модели прогнозирования временных рядов с помощью синергии авторегрессионных и нейросетевых моделей. К разрабатываемой модели поставлено ряд требований и условий, среди которых наиболее важным является требование построения модели без участия специалиста по машинному обучению. Условие, предъявляемое к временному ряду — дифференцируемость первого порядка — позволяет свести нестационарный ряд к стационарному. Детально описан механизм обучения. Получены первые практические результаты.

К л ю ч е в ы е с л о в а: временной ряд, прогнозирование, модель, авторегрессия, нейронные сети, механизм обучения.

Существуют достаточно эффективные методы и модели прогнозирования временных рядов [1,2], однако подготовка качественной модели прогнозирования требует участия эксперта и занимает значительное время. Это практически сводит к нулю эффективность известных методов, когда необходимо построить сотни и тысячи моделей за достаточно короткое время. Так, например, поставляя тысячи артикулов продукции в сотни магазинов, дистрибьюторы и ритейлеры, такие как Amazon и др., естественно, хотят прогнозировать продажи этих товаров во всех точках [1, 2].

Анализ математических моделей прогнозирования временных рядов, представленный в работе [3], систематизирован в табл. 1. Ссылаясь на авторитетные источники, автор работы [3] констатирует, что «на сегодняшний день наиболее распространенными моделями прогнозирования являются авторегрессионные модели (ARIMAX), а также ней-

© Кравцов Г.А., Примушко А.Н., Кошель В.И., 2020

росетевые модели (ANN)». Однако, как указывают эксперты [1,2], рассмотренные подходы не позволяют решить задачу разработки и адаптации большого числа моделей без участия человека в ограниченное время применительно к задачам ритейла.

Постановка задачи. Пусть даны временные ряды $X_{i,j}$, где $i \in [1, N]$; N — число связанных временных рядов; $j \in [1, T]$; T — число наблюдений. Требуется разработать способ прогнозирования значения $X_{1,T+1}$, т.е. целевой ряд $X_{1,j}$ и регрессоры $X_{i,j}$ (предикторы) при $i \in [2, N]$, учитывая сильные стороны авторегрессионных моделей и нейронных сетей и минимизируя недостатки каждого подхода в отдельности.

Таблица 1. Модели прогнозирования

Модель и метод	Достоинства	Недостатки
Регрессионные	Простота, гибкость, интуитивность моделирования; соответствие анализа и проектирования	Сложность определения функциональной зависимости; трудоемкость определения коэффициентов зависимости; отсутствие возможности моделирования нелинейных процессов (для нелинейной регрессии)
Авторегрессионные	Простота, интуитивность моделирования; единообразие анализа и проектирования; множество примеров применения	Трудоемкость и ресурсоемкость идентификации; невозможность моделирования нелинейностей; низкая адаптивность
Экспоненциального сглаживания	Простота моделирования; единообразие анализа и проектирования	Недостаточная гибкость; узость применения
Нейросетевые	Нелинейность моделей; масштабируемость, высокая адаптивность; единообразие анализа и проектирования; множество примеров применения	Отсутствие прозрачности; сложность выбора архитектуры; жесткие требования к обучающей выборке; сложность выбора алгоритма обучения; ресурсоемкость процесса обучения
На основе цепей Маркова	Простота моделирования; единообразие анализа и проектирования	Невозможность моделирования процессов с длинной памятью; узкая применимость моделей
На основе классификационно-регрессионных деревьев	Масштабируемость; быстрота и простота процесса обучения; возможность учитывать категориальные переменные	Неоднозначность алгоритма построения дерева; сложность останова

Условия, влияющие на построение модели:

1. Построение модели прогнозирования без участия человека.
2. Наиболее короткое время для построения одной модели и оценки ее способности давать значимый прогноз.
3. Максимальная интерпретируемость (объяснимость) метода построения модели и полученных результатов прогнозирования.

Допущения:

1. Одного дифференцирования $X_{i,j+1} - X_{i,j}$ достаточно, чтобы нестационарный ряд $X_{i,j}$ свести к стационарному.
2. Ряды $X_{i,j}$ являются «условно» периодическими, что в эконометрике соответствует понятиям сезонность или (и) цикличность.
3. Ряд должен содержать минимум два полных периода.
4. Любое значение целевого ряда может зависеть только от предыдущих значений целевого ряда и(или) предыдущих значений регрессоров (история значений определяет будущие значения).
5. Чем ближе рассматриваемое значение на временной оси к точке прогнозирования, тем большее влияние на прогноз оно оказывает.
6. Прогнозирование следующего значения сводится к прогнозированию величины изменения следующего значения относительно предыдущего.

Построение модели. Наиболее простым с точки зрения интерпретации инструментом является линейный коэффициент корреляции Пирсона, определяющий величину линейной взаимосвязи и изменяющийся в диапазоне от -1 до 1 . При этом минус единица означает полную (функциональную) линейную обратную взаимосвязь (две переменные связаны так, что при возрастании значений одной из них значения другой убывают); единица означает полную (функциональную) линейную положительную взаимосвязь; нуль свидетельствует об отсутствии линейной корреляции (но не обязательно взаимосвязи).

Коэффициент корреляции Пирсона рассчитываем по формуле

$$r_{i,j} = \frac{\sum_{k=1}^T (X_{i,k} - M(X_{i,k}))(X_{j,k} - M(X_{j,k}))}{\sqrt{\sum_{k=1}^T (X_{i,k} - M(X_{i,k}))^2} \sqrt{\sum_{k=1}^T (X_{j,k} - M(X_{j,k}))^2}}.$$

Для двух переменных построим коррелограмму:

$$r_{i,j}(h) = \frac{\sum_{k=h+1}^T (X_{i,k} - M'(X_{i,k}))(X_{j,k-h} - M'(X_{j,k-h}))}{\sqrt{\sum_{k=h+1}^T (X_{i,k} - M'(X_{i,k}))^2} \sqrt{\sum_{k=1}^{T-h} (X_{j,k} - M'(X_{j,k}))^2}}. \quad (1)$$

Очевидно, что при построении коррелограммы (1) используются «усеченные» ряды $X_{i,j}$ длиной $T-h$, где h — лаг. Следует заметить, что в (1) средние значения (со штрихом) взяты для усеченных рядов. Основная проблема такого усечения — потенциальная потеря информации в отсеченных частях рядов. Как следствие, чем выше лаг, тем ниже степень доверия такой оценке корреляции рядов.

Используя допущение 2, будем полагать T наибольшим из наблюдаемых периодов рядов, что при $j \in [1, T]$ может быть записано в виде $X_{i,j} = X_{i,((j-1) \bmod T)+1}$. Тогда (1) примет вид

$$r_{i,j}(h) = \frac{\sum_{k=h+1}^{T+h} (X_{i,((k-1) \bmod T)+1} - M(X_{i,k}))(X_{j,((k-h-1) \bmod T)+1} - M(X_{j,k}))}{\sqrt{\sum_{k=h+1}^{T+h} (X_{i,((k-1) \bmod T)+1} - M(X_{i,k}))^2} \sqrt{\sum_{k=h+1}^{T+h} (X_{j,((k-h-1) \bmod T)+1} - M(X_{j,k}))^2}}$$

Допущение 3 позволяет определить область допустимых значений для лага в виде $h \in [1, T/2]$. Потребуем, чтобы коррелограмма $r_{i,j}(h)$ содержала только значимые значения корреляции (на основании преобразования Фишера для аппроксимации точного распределения коэффициента корреляции). Пусть $r_{i,j}(h)$ при значениях лагов $h_{i,1}, \dots, h_{i,k}$, где $k \in [1, T/2]$, принимает значимые значения корреляции $r_{1,i}(h_{i,k})$. Будем полагать, что временные ряды $X_{i,j}$ оказывают влияние на ряды $X_{1,j}$ пропорционально значимым значениям корреляции $r_{1,i}(h_{i,k})$ (подразумевается и влияние $X_{1,j}$ на самого себя, т.е. авторегрессионность).

Коэффициент корреляции может быть представлен как величина влияния на изменения целевого ряда $X_{1,j}$ посредством фактора совокупного влияния:

$$R = \sum_i \sum_k |r_{1,i}(h_{i,k})|,$$

где $|r_{1,i}(h_{i,k})|$ — ненулевая значимая величина коэффициента корреляции по модулю; K_i — число ненулевых значений коэффициента корреляции между рядами $X_{1,j}$ и $X_{i,j}$.

Нормированное линейное влияние на целевой ряд $X_{1,j}$ значений рядов $X_{i,j}$ с лагами $h_{i,k}$ определяется по формуле $w_{i,k} = \frac{r_{1,i}(h_{i,k})}{R}$. Очевидно, что

$w_{i,k}$ имеет тот же знак, что и $r_{1,i}(h_{i,k})$. Решение в виде функции

$$X_{1,T+1} - X_{1,T} = f \left(\sum_i \sum_k w_{i,k} (X_{i,T-h_{i,k}+1} - X_{i,T-h_{i,k}}) \right) \quad (2)$$

соответствует допущениям 1 и 6.

Существует несколько особенностей, вытекающих из потенциально-го решения (2). Во-первых, решение (2) не чувствительно к нестационарным рядам, которые приводят к стационарным посредством первой разности (дифференцирования). Во-вторых, предлагаемая модель (2) есть не что иное, как перцептрон, где функция активации с областью значений $[-1; 1]$ смасштабирована максимальным значением, взятым по модулю разницы между двумя соседними элементами ряда $|X_{1,j} - X_{1,j-1}|$. В-третьих, очевидно, что модель (2) обладает памятью без учета фактора устаревания знаний о рядах.

Прежде всего, введем явным образом учет максимального изменения $X_{1,j} - X_{1,j-1}$ как функции расстояния между значениями (т.е. разницы между значениями, взятая по модулю). Основываясь на модели линейной регрессии, $|X_{1,j} - X_{1,j-1}|$ представим в виде

$$|X_{1,j} - X_{1,j-1}| = D_0 + D_1 j + e_j. \quad (3)$$

В уравнении (3) переменная D_1 указывает тренд изменения $|X_{1,j} - X_{1,j-1}|$ от $j \in [1, T]$:

$$D_1 = \frac{\sum_{j=2}^T (j - M(j)) (|X_{1,j} - X_{1,j-1}| - M(|X_{1,j} - X_{1,j-1}|))}{\sum_{j=2}^T (j - M(j))^2}.$$

Определим среднее значение для всех $j \in [1, T]$ на основании формулы суммы членов арифметической прогрессии $S_n = n \frac{a_1 + a_n}{2}$ и среднего значения члена арифметической прогрессии $M(n) = S_n/n$:

$$M(j) = \frac{1+T}{2},$$

что эквивалентно записи

$$M(j) = \frac{1}{T} \sum_{j=1}^T j.$$

Тогда

$$D_1 = \frac{\sum_{j=2}^T \left(j - \frac{1+T}{2} \right) (|X_{1,j} - X_{1,j-1}| - M(|X_{1,j} - X_{1,j-1}|))}{\sum_{j=2}^T \left(j - \frac{1+T}{2} \right)^2}. \quad (4)$$

Пусть в момент времени j_D значение $|X_{1,j} - X_{1,j-1}|$ максимально: $|X_{1,j_D} - X_{1,j_D-1}| \rightarrow \max$. Найдем постоянную составляющую тренда абсолютных значений колебаний в виде

$$D_0 = |X_{1,j_D} - X_{1,j_D-1}| - D_1 j_D. \quad (5)$$

Тогда (2) можно представить в виде

$$\begin{aligned} X_{1,T+1} - X_{1,T} &= \\ &= (D_1(T+1) + D_0) f \left(\sum_i \sum_k w_{i,k} (X_{i,T-h_{i,k}+1} - X_{i,T-h_{i,k}}) \right), \end{aligned} \quad (6)$$

где $-1 \leq f \left(\sum_i \sum_k w_{i,k} (X_{i,T-h_{i,k}+1} - X_{i,T-h_{i,k}}) \right) \leq 1$; $D_1(T+1) + D_0 \geq 0$; $f(\dots)$ — функция активации.

Модель (6) является моделью с бесконечной памятью. Очевидно, что более поздние наблюдения имеют большее влияние на поведение временного ряда, а требование иметь бесконечную память при реализации рассматриваемого метода прогнозирования средствами вычислительной техники сводит практическую ценность такого подхода к нулю.

В работе [4] показано, что естественный процесс забывания (уменьшение значимости устаревающей информации) по своей природе близок к процессу полураспада в физике, а следовательно,

$$A(X_{i,T-h}) = \frac{1}{(1+L_i)^h} A(X_{i,T}),$$

где L_i — скорость забывания, $L_i > 0$; $A(X_{i,j})$ — актуальность знаний о поведении ряда $X_{i,j}$ в момент j . Приняв, что $A(X_{i,T}) = 1$ (т.е. последнее наблюдение максимально актуально) и $A(X_{i,j})$ изменяется в пределах $[0; 1]$, актуальность наблюдения можно записать в виде

$$A(X_{i,T-h}) = \frac{1}{(1+L_i)^h}.$$

Введем в модель (6) актуальность знаний L_i и коэффициент коррекции влияния C_i целевого ряда и регрессоров на целевой ряд:

$$X_{1,T+1} - X_{1,T} = (D_1(T+1) + D_0) f \left(\sum_i C_i \sum_k \frac{w_{i,k}}{(1+L_i)^{h_{i,k}}} (X_{i,T-h_{i,k}+1} - X_{i,T-h_{i,k}}) \right). \quad (7)$$

Как видим, в модели (7) кроме параметров коррекции влияния использован коэффициент забывания L_i для каждого ряда. Очень большие по абсолютной величине значения параметра коррекции влияния C_i будут приводить к дестабилизации модели. Потенциальным решением может быть ограничение значений суммы

$$\sum_k \frac{w_{i,k}}{(1+L_i)^{h_{i,k}}} (X_{i,T-h_{i,k}+1} - X_{i,T-h_{i,k}})$$

в диапазоне $(-1; 1)$ или $[-1; 1]$, что логично сделать с помощью функции активации.

Перепишем (7) в виде

$$X_{1,T+1} - X_{1,T} = (D_1(T+1) + D_0) f_0 \left(\sum_i C_i f_i \left(\sum_k \frac{w_{i,k}}{(1+L_i)^{h_{i,k}}} (X_{i,T-h_{i,k}+1} - X_{i,T-h_{i,k}}) \right) \right). \quad (8)$$

Как видим, модель (8) представляет собой двухслойную нейронную сеть, в которой учтено следующее:

нестационарность целевого ряда, сводимого к стационарному первым дифференцированием;

гетеро- и гомоскедастичность целевого ряда вследствие граничного значения первой разницы как линейной функции от порядкового номера наблюдения (фактически от времени);

влияние значимых линейных зависимостей целевого ряда от регрессоров с учетом устаревания информации во времени.

Нейронная сеть с масштабируемым значением выхода обучается относительно параметров коррекции влияния C_i и скорости забывания прошлых знаний L_i и одновременно является авторегрессионной моделью (Auto-Regressive Artificial Neural Network, ARANN).

Определим функции активации. Критерием для их выбора является область значений $-1 \leq f(\dots) \leq 1$ или $-1 < f(\dots) < 1$. Рассмотрим варианты, где М означает монотонность функции, МП — монотонность произ-

водной, АТФ — возможность аппроксимации тождественной функцией вблизи начала координат (табл. 2).

Механизм обучения модели. Запишем (8) в общем виде для прогнозирования $X_{1,j}$:

$$\hat{X}_{1,j} = (D_1(j) + D_0) f_0 \left(\sum_i C_i f_i \left(\sum_k \frac{w_{i,k}}{(1 + L_i)^{h_{i,k}}} (X_{i,j-h_{i,k}} - X_{i,j-h_{i,k}-1}) \right) \right) + X_{1,j-1}, \quad (9)$$

где $\hat{X}_{1,j}$ — прогнозируемое значение $X_{1,j}$. Положим, что качество модели может быть оценено квадратичной функцией ошибки. Используя допущение 3 о наличии хотя бы двух периодов в целевом ряду, будем считать, что обучение целесообразно проводить на интервале $[\text{round}(T/2); T]$, так как модель предусматривает наличие циклов, поэтому будет использоваться вся информация из $X_{i,j}$:

$$E = \sum_{j=\text{round}(T/2)}^T (X_{1,j} - \hat{X}_{1,j})^2.$$

Потребуем, чтобы в результате обучения выполнялось условие $E \rightarrow \min$. Исходя из формального описания модели (9), будем осуществлять обучение методом градиентного спуска. Для этого нужно найти частные производные от функции ошибки по параметрам C_i и L_i ,

Таблица 2. Функции активации

Функция	Уравнение	Производная	Область значений	М	МП	АТФ
Сигмоида нормированная $\text{sigm}(x)$	$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$	$\frac{df(x)}{dx} = 2f(x)(1 - f(x))$	$(-1; 1)$	+	-	-
Гиперболический тангенс $\text{th}(x)$ или $\text{htan}(x)$	$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$\frac{df(x)}{dx} = 1 - f^2(x)$	$(-1; 1)$	+	-	+
Обратный квадратный корень	$f(x) = \frac{x}{\sqrt{1+x^2}}$	$\frac{df(x)}{dx} = \left(\frac{1}{\sqrt{1+x^2}} \right)^3$	$(-1; 1)$	+	-	+

т.е. $\frac{\theta E}{\theta C_i}$ и $\frac{\theta E}{\theta L_i}$. Очевидно, что

$$\frac{\theta E}{\theta C_i} = -2 \sum_{j=\text{round}(T/2)}^T (X_{1,j} - \hat{X}_{1,j}) \frac{\theta \hat{X}_{1,j}}{\theta C_i},$$

$$\frac{\theta E}{\theta L_i} = -2 \sum_{j=\text{round}(T/2)}^T (X_{1,j} - \hat{X}_{1,j}) \frac{\theta \hat{X}_{1,j}}{\theta L_i}$$

согласно правилу о нахождении производной от сложной функции, которое будем использовать. Если f и g — дифференцируемые функции от x , то сложная функция $y = f(g(x))$ дифференцируема по x и ее производная имеет вид

$$\frac{dy}{dx} = \frac{d}{dx} (f \circ g)(x) = \frac{d}{dx} f(g(x)) g'(x).$$

Найдем частные производные $\frac{\theta \hat{X}_{1,j}}{\theta L_i}$ и $\frac{\theta \hat{X}_{1,j}}{\theta C_i}$ в общем виде (т.е. без выбора конкретной функции активации). Пусть

$$F(L_i) = \sum_k \frac{w_{i,k}}{(1+L_i)^{h_{i,k}}} (X_{i,j-h_{i,k}} - X_{i,j-h_{i,k}-1}),$$

$$H(L_i, C_i) = \sum_i C_i f_i(F(L_i)).$$

Тогда

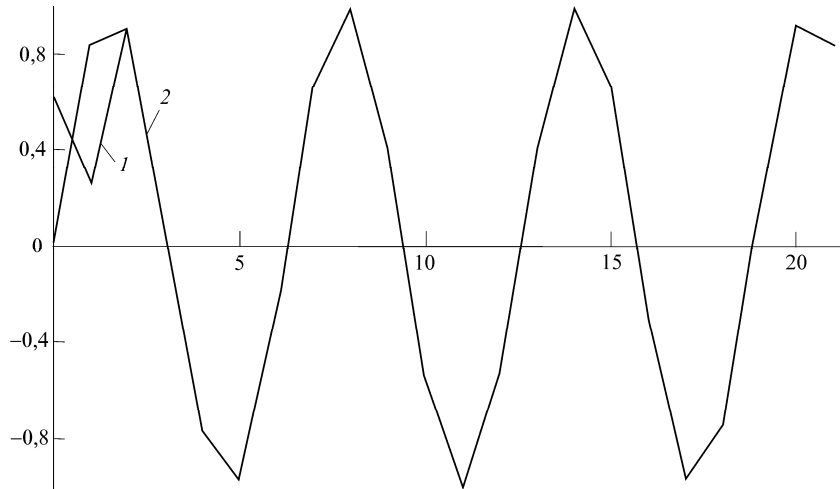
$$\frac{\theta \hat{X}_{1,j}}{\theta L_i} = (D_1(j) + D_0) f'_0(H(L_i, C_i)) \sum_i C_i f'_i(F(L_i)) \frac{\theta F(L_i)}{\theta L_i},$$

$$\frac{\theta \hat{X}_{1,j}}{\theta C_i} = (D_1(j) + D_0) f'_0(H(L_i, C_i)) \sum_i f_i(F(L_i)).$$

Находим частную производную:

$$\frac{\theta F(L_i)}{\theta L_i} = \frac{\theta}{\theta L_i} \sum_k w_{i,k} (X_{i,j-h_{i,k}} - X_{i,j-h_{i,k}-1}) (1+L_i)^{-h_{i,k}},$$

$$\frac{\theta F(L_i)}{\theta L_i} = - \sum_k h_{i,k} w_{i,k} (X_{i,j-h_{i,k}} - X_{i,j-h_{i,k}-1}) (1+L_i)^{-h_{i,k}-1}.$$



Результаты численного эксперимента прогнозирования временного ряда: 1 — модельная синусоида; 2 — оригинальная синусоида

По формулам (4) и (5) вычислим параметры линейной регрессии, которая определяет максимальное абсолютное значение первой разницы значений двух последовательных наблюдений временного ряда. При обучении модели будем использовать частные производные $\frac{\theta E}{\theta D_i}$, для D_1 и D_0 :

$$\frac{\theta E}{\theta D_i} = -2 \sum_{j=\text{round}(T/2)}^T (X_{1,j} - \hat{X}_{1,j}) \frac{\theta \hat{X}_{1,j}}{\theta D_i},$$

где

$$\frac{\theta \hat{X}_{1,j}}{\theta D_0} = f_0 \left(\sum_i C_i f_i \left(\sum_k \frac{w_{i,k}}{(1+L_i)^{h_{i,k}}} (X_{i,j-h_{i,k}} - X_{i,j-h_{i,k}-1}) \right) \right),$$

$$\frac{\theta \hat{X}_{1,j}}{\theta D_1} = j f_0 \left(\sum_i C_i f_i \left(\sum_k \frac{w_{i,k}}{(1+L_i)^{h_{i,k}}} (X_{i,j-h_{i,k}} - X_{i,j-h_{i,k}-1}) \right) \right).$$

Компьютерная модель, разработанная на основе изложенных формул, подтвердила эффективность предлагаемого подхода к временным рядам с явно выраженной периодичностью. На рисунке представлены результаты прогнозирования временного ряда. Первые практические результаты обнадеживают, но говорить о коммерческом использовании подхода пока рано. Предлагаемый подход основан на синергии авторегрессионных и нейросетевых моделей, что теоретически должно дать позитивный эффект.

Выводы

Выдвинутые требования и условия ограничили сферу применения предлагаемого подхода к построению модели прогнозирования временных рядов. Наиболее важным является требование построения модели без участия специалиста по машинному обучению, так как бывает необходимо на совокупности временных рядов построить сотни и тысячи моделей. Среди выдвигаемых условий к временному ряду отмечается дифференцируемость первого порядка, позволяющая свести нестационарный ряд к стационарному. Эксперты, работающие в области прогнозирования временных рядов, считают, что в большинстве случаев, применив первое дифференцирование, нестационарные ряды можно свести к стационарным. Предложенный механизм обучения прост в практической реализации и получены первые обнадеживающие практические результаты.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. *Рябенко Е.* Автоматическое прогнозирование временных рядов. Цикл лекций «Data Mining in Action», 17.12.2016. <https://www.youtube.com/watch?v=u433nrxdf5k>
2. *Sean Taylor.* Automatic Forecasting in Scale. // Join Statistical Meet up, 12.08.2015. <https://www.slideshare.net/seanjtaylor/automatic-forecasting-at-scale>
3. *Чучуева И.А.* Модель прогнозирования временных рядов по выборке максимального правдоподобия. Диссертация ... канд. физ.-мат. наук. Московский государственный технический университет им. Н.Э. Баумана. Москва, 2012, 153 с.
4. *Баканова С.А., Силкина Г.Ю.* Эволюция знаний: моделирование и прикладной анализ // Науч.-техн. ведомости Санкт-Петербургского государственного политехнического университета. Экономические науки. Санкт-Петербург, 2015, с. 173—182 <https://cyberleninka.ru/article/n/evolyutsiya-znaniy-modelirovanie-i-prikladnoy-naliz/viewer>

Получена 09.06.20

REFERENCES

1. Ryabenko, E. (2016), "Automatic time series forecasting", Tsikllektsiy «Data Mining in Action», available at: <https://www.youtube.com/watch?v=u433nrxdf5k> (accessed June 9, 2020).
2. Sean, T. (2015), "Automatic Forecasting in Scale", Join Statistical Meetup, available at: <https://www.slideshare.net/seanjtaylor/automatic-forecasting-at-scale> (accessed June 9, 2020).
3. Chuchuyeva, I.A. (2012), "Time Series Prediction Model for Maximum Credibility Sampling", Abstract of Cand. Sci. (Tech.) dissertation, 2012, Moscow State Technical University N.E. Bauman, Moscow, Russia.
4. Bakanova, S.A. and Silkina, G.Yu. (2015), "Knowledge Evolution: Modeling and Applied Analysis", Nauchno-tekhnicheskoyevedomosti Sankt-Peterburgskogogosudarstvennogopolitekhnicheskogouniversiteta. Ekonomicheskienauki, pp. 173-182, St. Petersburg, Russia, available at: <https://cyberleninka.ru/article/n/evolyutsiya-znaniy-modelirovanie-i-prikladnoy-analiz/viewer> (accessed June 9, 2020).

Received 09.06.20

Г.О. Кравцов, А.М. Примушко, В.І. Кошел

КОМБІНОВАНИЙ АВТОРЕГРЕСІЙНО-НЕЙРОМЕРЕЖЕВИЙ
МЕТОД ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Викладено концептуальний підхід до побудови моделі прогнозування часових рядів за допомогою синергії авторегресійних і нейромережєвих моделей. До розробленої моделі висунуто ряд вимог та умов, серед яких найважливішою є вимога побудови моделі без участі фахівця з машинного навчання. Умова, поставлена до часового ряду, а саме диференційованість першого порядку, дозволяє звести нестационарний ряд до стационарного. Детально описано механізм навчання. Отримано перші практичні результати.

К л ю ч о в і с л о в а: часовий ряд, прогнозування, модель, авторегресія, нейронні мережі, механізм навчання.

G.O. Kravtsov, A.N. Prymushko, V.I. Koshell

COMBINED AUTOREGRESSIVE-NEURAL
NETWORK METHOD FOR PREDICTING TIME SERIES

The authors propose an approach to constructing a time series forecasting model due to the synergy of autoregressive and neural network models. The authors put forward a number of requirements and conditions for the developed model. Among the requirements put forward, the central place is occupied by the requirement to build a model without the participation of a machine learning specialist. Among the conditions put forward to the time series, it should be noted the differentiability of the first order, which makes it possible to reduce the non-stationary series to the stationary one. The article describes the learning mechanism with detailed mathematical explanations. The approach outlined in the article is conceptual.

K e y w o r d s: time series, forecasting, model, autoregression, neural networks, learning mechanism.

КРАВЦОВ Григорий Алексеевич, канд. техн. наук, докторант Института проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины. В 2000 г. окончил Севастопольский военно-морской институт им. П.С. Нахимова. Область научных исследований — кибербезопасность смарт-грид, криптография, программирование, разработка распределенных гетерогенных вычислительных систем.

ПРИМУШКО Арсентий Николаевич, магистр Национального технического университета Украины «Киевский политехнический институт им. Игоря Сикорского», бакалаврат которого окончил в 2019 г. Область научных исследований — искусственный интеллект, машинное обучение, искусственные нейронные сети, программирование, системы навигации и ориентации.

КОШЕЛЬ Владимир Иванович, аспирант Института проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины. В 2002 г. окончил Харьковский национальный университет им. В.Н. Каразина. Область научных исследований — искусственный интеллект, интеллектуальный анализ данных, искусственные нейронные сети, обработка естественного языка.