

УДК 519.254

## **РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ**

**О.Г. КОТОВА<sup>1\*</sup>, С.В. КОВАЛЕНКО<sup>2</sup>, С.М. КОВАЛЕНКО<sup>3</sup>**

<sup>1</sup> *магістрант кафедри системного аналізу та інформаційно-аналітичних технологій, НТУ «ХПІ», Харків, УКРАЇНА*

<sup>2</sup> *доцент кафедри системного аналізу та інформаційно-аналітичних технологій, канд. техн. наук, НТУ «ХПІ», Харків, УКРАЇНА*

<sup>3</sup> *доцент кафедри кібернетики, канд. техн. наук, ХНТУСГ ім. П. Василенка, Харків, УКРАЇНА*

\* *email: oly6287@gmail.com*

Повна і достовірна статистична інформація є тією необхідною основою, на якій базується процес управління економікою. Вся інформація, що має народногосподарську значимість, в кінцевому рахунку обробляється та аналізується за допомогою статистики. На всіх стадіях дослідження статистика використовує різні методи.

Майже в кожній області зустрічаються явища, які цікаво і важливо вивчати в їх розвитку і зміні у часі. Часові ряди, як правило, виникають в результаті вимірювання деякого показника. Це можуть бути як показники (характеристики) технічних систем, так і показники природних, соціальних, економічних та інших систем.

На сьогодні все більше набуває популярності такий напрям як Data Mining. Data Mining – це мультидисциплінарна область, що виникла і розвивається на базі таких наук як прикладна статистика, розпізнавання образів, штучний інтелект, теорія баз даних та інших. Однією з її задач є прогнозування даних, що є важливим елементом організації управління як окремими суб'єктами, так і економіки в цілому.

Основою для прогнозування служить інформація, що зберігається в базі даних та в подальшому може бути представлена у вигляді часових рядів. Часовий ряд – це послідовність спостережуваних значень будь-якої ознаки, упорядкованих у невипадкові моменти часу [1]. На базі цього існує напрям, що має назву Time Series Data Mining (Data Mining часових рядів). За однією з класифікацій часові ряди поділяються на стаціонарні та нестаціонарні. Стаціонарним часовим рядом називається такий ряд, який залишається в рівновазі щодо постійного середнього рівня. Решта часових рядів є нестаціонарними. У промисловості, торгівлі, в економіці, де прогнозування має важливе значення, багато часових рядів є нестаціонарними.

За оцінками зарубіжних і вітчизняних фахівців налічується понад сто методів та моделей прогнозування, в зв'язку з чим виникає задача вибору методів, які давали б адекватні прогнози для досліджуваних процесів або систем. Загалом методи прогнозування можна розділити на три широкі класи:

1. Прогнозування на основі суджень, тобто, прогнозування, що ґрунтується на суб'єктивних судженнях (оцінках), інтуїції, поглиблених знаннях конкретної області та іншій інформації, що має відношення до прогнозування процесу – так зване передбачення;

2. Методи прогнозування на основі використання часового ряду однієї змінної, тобто на основі авторегресії, авторегресії з ковзним середнім та авторегресії з ковзним середнім з моделлю тренду;

3. Методи прогнозування на основі використання часових рядів декількох змінних.

В ході виконання роботи був зроблений ретроспективний прогноз часового ряду тривалості життя в Україні за останні п'ятдесят п'ять років чотирма методами. При цьому виникає проблема вибору моделі. Тому більш універсальною представляється ідея спільного застосування декількох алгоритмів прогнозування [2].

Нехай для прогнозування часового ряду  $y(t)$  використовуються  $p$  алгоритмів,  $b_i(t)$  – прогноз  $i$ -го алгоритму на момент часу  $t$ . Знайдемо точний прогноз  $a(t)$  як лінійну комбінацію  $b_p$  [2]:

$$a(t) = \sum_{i=1}^p w_{it} b_i(t), \quad \sum_{i=1}^p w_{it} = 1, \quad t = 1, \dots, T. \quad (1)$$

Ваги алгоритмів  $w_{it}$ , що використовуються для прогнозування у момент часу  $t$ , обчислюються за даними передісторії  $[1, \dots, t-1]$ .

Адаптація вагових коефіцієнтів робиться за наступною формулою:

$$w_t = \operatorname{argmin} \sum_{\tau=1}^{t-1} \theta^{t-\tau-1} \left( \sum_{i=1}^p w_i b_i(\tau) - y(\tau) \right)^2 + \lambda \sum_{i=1}^p (w_i - w_{i,t-1})^2.$$

де мінімум береться по вектору ваг  $w = \mathbb{I}(w)_1, \dots, w_p$ , які відповідають умові нормування (1). Параметр  $\theta$  – це коефіцієнт «забування» передісторії,  $\lambda$  – коефіцієнт регуляції (згладжування). Другий доданок є штрафом за відхилення вектору ваг  $w_t$  від вектору ваг  $w_{t-1}$  у попередній момент часу.

Метою роботи було створення програмного забезпечення для аналізу статистичної інформації. У ході були розглянуті часові ряди, їх місце в новій області – Data Mining. Був проведений аналіз методів та моделей для прогнозування часових рядів. Розглянута проблема вибору адекватної моделі. Запропоновано алгоритм для динамічної адаптації композиції моделей прогнозування. Створено програмне забезпечення, що робить композиції на основі базових алгоритмів та за допомогою вагових коефіцієнтів, що динамічно адаптуються, знаходить найбільш точний прогноз.

#### Список літератури:

1. *Красс М.С.* Математика для экономистов / *М.С. Красс, Б.П. Чупрынов.* – СПб. : Питер, 2016 – 426 с.
2. *Дуброва Т.А.* Статистические методы прогнозирования в экономике. / *Т.А. Дуброва, М.Ю. Архипова.* – М. : МЭСИ, 2004 – 68 с.