

Щука Р.В., Іванов С.С., Терентьев О.М., Орешко В.А., Яворський Я.Б.,  
Кузенко А.Ю.

КПІ ім. Ігоря Сікорського, ІПСА, Київ, Україна

## Побудова предиктивної моделі прогнозування злочинності на основі методу подібних траєкторій в системі SAS

**Вступ.** Однією з наріжних задач розвитку соціальної сфери України, що тривалий термін перебуває під тиском несприятливих політичних та соціальних явищ, спричинених антитерористичною операцією в окремих районах Донецької та Луганської областей, залишається підвищення соціальної безпеки населення. У зв'язку із цим, дослідження феномену зростання злочинності сьогодні не тільки набувають великого значення у правовому житті суспільства, а й виступають одним із вирішальних чинників у справі успішного здійснення реформ на шляху розбудови якісно нової української держави і забезпечення безпеки її громадян. Успішне вирішення останніх уявляється можливим завдяки новітнім інформаційним технологіям та створенню на їх базі інформаційно-аналітичних систем.

Результати вивчення публікацій в галузі методологічних підходів до моделювання злочинності та злочинної поведінки у кримінологічній науці [1–6] дозволяють дійти висновку, що питанням створення відносно дешевих інформаційних систем для аналізу і прогнозування злочинності приділяється недостатньо уваги.

**Запропонований підхід.** У даній роботі в якості одного зі шляхів вирішення проблеми запропоновано використання розробленого К. Мортеном методу подібних траєкторій. Аналіз технологічного програмного забезпечення й додатків класу Business Intelligence показав, що згаданий метод не має програмної реалізації на платформі SAS. Це надало можливість висунути гіпотезу про те, що створення програмного продукту в SAS Enterprise Guide із використанням мови програмування SAS Base зможе посприяти підвищенню точності прогнозування наслідків соціально-економічних процесів, зокрема – рівня злочинності.

**Метод подібних траєкторій.** Нехай у ході соціального експерименту сформовано апріорний ряд спостережень  $Y(t) = y(1), y(2), \dots, y(i), \dots, y(n)$ , де  $y(i), \dots, y(n)$  – значення спостережуваної величини, що їх зафіксовано у ході експерименту з певним часовим кроком  $\Delta t_i = const$ , зображеним на рисунку 1:

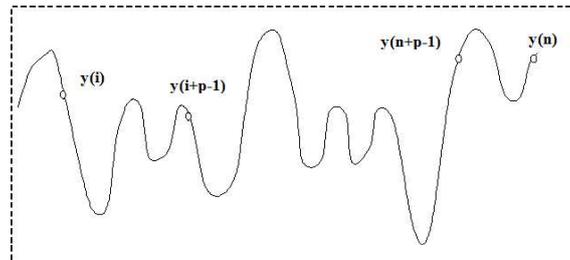


Рис. 1. Приклад спостережень

Обчислимо “перші різниці”  $\Delta y(i) = y(i+1) - y(i)$ , між сусідніми точками спостережень (різниці між сусідніми значеннями апріорного ряду). Для цього виберемо ділянку траєкторії, що передують прогнозованій точці на крок розміром  $(p-1)$ .

Сутність подальшого алгоритму полягатиме у наступному. Спочатку визначимо певну кількість символів алфавіту  $B = b_k$ , де  $k$  – довжина алфавіту.

Кількість символів в алфавіті є визначальною під час розбиття відсортованого ряду на часткові інтервали: кожному значенню ряду у відповідність ставиться власний символ. Таким чином числовий ряд спостережень  $\Delta y(i)$  набуває символічних значень  $b_1 = \text{“A”}$ ,  $b_2 = \text{“B”}$ ,  $b_3 = \text{“C”}$  і т.д.

Дане дослідження було здійснено в рамках проекту NUKR. SFPP G4877 “Modelling and Mitigation of Social Disasters Caused by Catastrophes and Terrorism”

На наступному кроці визначаємо опорний ланцюжок  $\langle e_i \rangle$  – останні значення символічного ряду, які передують прогнозованій точці  $y(n)$ ; весь ряд розбивається на часткові вектори спостережень  $\langle e_i; e_{i+1} \rangle$ ; вектор  $i$  є опорним. Далі підраховуємо кількість ланцюжків, що співпадають з вектором  $i$  та дивимось, яких значень набуває ряд на наступних кроках.

Частота появи того чи іншого ланцюжка відповідає ймовірності повторення відповідних їм послідовностей. Іншими словами, обчислюється умовна ймовірність  $P(e_{i+1}|e_{i-n} \dots e_{i-1})$  появи символу  $e_{i+1}$  за умови, що йому передували символи:  $e_{i-n}, e_{i-n+1}, \dots, e_{i-1}$ .

За допомогою обчислених ймовірностей можемо зробити ймовірнісний прогноз виникнення якогось символу за умови що відомі ланцюжки попередніх символів. Після отримання прогнозу, робимо зворотній перехід від символічного ряду  $B = b_k$  до числового  $\Delta y(i)$ .

**Результати.** Алгоритм, що описує метод подібних траєкторій, був реалізований в середовищі SAS Enterprise Guide. В результаті роботи програми були побудовані прогнозні моделі для тестових даних зі стандартної бібліотеки SAS AIR, що описує історію продажів авіабілетів на основі якої можна порівняти результати роботи нових методів з уже існуючими, та даних МВС України по кількості зареєстрованих правопорушень у місті Дніпро. Результати прогнозування наведені у таблиці 1:

Табл. 1. Статистики прогнозних моделей

|   | RMSE     | MAPE     | DW     | R <sup>2</sup> |
|---|----------|----------|--------|----------------|
| Тестовий графік AIR $ql = 3, ww = 3$                        | 9881     | 16,5016  | 0,8085 | 0,3459         |
| Тестовий графік AIR $ql = 3, ww = 7$                        | 3009     | 9,7480   | 0,9967 | 0,6732         |
| Графік кількості правопорушень у м. Дніпро $ql = 5, ww = 3$ | 1935     | 114,1657 | 1,0323 | 0,7267         |
| Графік кількості правопорушень у м. Дніпро $ql = 7, ww = 3$ | 960,6833 | 82,1263  | 0,9600 | 0,8642         |

де  $ww$  – це ширина опорного вектора, а  $ql$  – довжина алфавіту.

**Висновки.** З аналізу результатів таблиці 1 можна зробити наступні висновки:

- при фіксованій довжині опорного вектора значення RMSE та MAPE спадають за зростання алфавіту, тобто точність прогнозу покращується;
- при фіксованій довжині опорного вектора значення критерію R<sup>2</sup> наближається до одиниці при збільшенні кількості букв в алфавіті, тобто зростає точність прогнозу;
- за результатами роботи методу на даних числових рядах не уявляється можливим зробити остаточний висновок стосовно автокореляції за статистикою DW.

Результати дослідження у вигляді програмної розробки будуть використані при створенні Ситуаційно-аналітичного центру при ННК ІПСА, для вирішення задач сценарного аналізу та моделювання.

**Література.** 1. Бова А.А. Перспективи використання інтелектуального аналізу даних (Data Mining) у кримінологічних дослідженнях [Текст] / А.А. Бова // Науковий вісник Національної академії внутрішніх справ України. – 2002. – № 5. – С. 32–35. 2. Жеребкин В.Е. Логический анализ понятий права: монография [Текст]: / В.Е. Жеребкин. – К.: Издательское объединение «Вища школа», 1976. – 151 с. 8. 3. Закалюк А.П. Курс сучасної української кримінології: теорія і практика: У 3-х кн. [Текст]: / А.П. Закалюк. – К.: Видавничий Дім «Ін Юре», 2008. – Кн. 3: Практична кримінологія. – 320 с. 4. Кравчук С.Й. Економічна злочинність в Україні [Текст] / С.Й. Кравчук. – К.: Кондор, 2009. – 282 с. 5. Кримінологія: Загальна та Особлива частини: підручник / І.М. Даньшин, В.В. Голіна, М.Ю. Валуйська та ін.; за заг. ред. В.В. Голіни. – 2-ге вид. перероб. і доп. [Текст]: – Х.: Право, 2009. – 288 с. 6. Чорноус Ю.М. Теорія і практика криміналістичного забезпечення досудового слідства у справах про злочини міжнародного характеру [Текст]: монографія / Чорноус Ю.М.; Нац. акад. внутр. справ. – К.: Скіф, 2012. – 447 с.