

Обробка інформації в складних організаційних системах

УДК 37.04:004 (355)

В.З. Андрушко

Національна академія Державної прикордонної служби України, Хмельницький

ВИБІР ПОКАЗНИКІВ НЕЙРО-НЕЧІТКОЇ МОДЕЛІ ПРОГНОЗУВАННЯ ІНТЕНСИВНОСТІ РУХУ НА ДЕРЖАВНОМУ КОРДОНІ

Запропоновано методику формалізованого вибору параметрів прогнозу інтенсивності руху осіб, транспортних засобів через державний кордон, яка дозволяє не лише провести оцінку пріоритетності, побудувати упорядковану послідовність параметрів у міру убавання їх значущості за заданими критеріями, а й здійснити розрахунок інформативності факторів, виключити малозначні і малоінформативні фактори. За її допомогою можна здійснити науково обґрунтований вибір інформативної сукупності параметрів для прогнозування, мінімізувати кількість параметрів для прогнозування без втрат в інформативності, зменшити розмірність масивів даних при обробці, а також скоротити витрати на їх прогноз.

Ключові слова: прогноз, параметри, факторний аналіз.

Вступ

Постановка проблеми у загальному вигляді. Однією з найважливіших частин розробки управлінського рішення є прогнозування.

Поява у Державній прикордонній службі України (далі – ДПСУ) потужних засобів збору й обробки інформації у складі інформаційно-телекомунікаційних систем «Гарт», які містять у базах даних відомості стосовно осіб, автотранспортних засобів, що перетнули державний кордон (далі – ДК), випадків порушення законодавства з питань перетинання ДК та інших показників, надає можливість автоматизувати процес прогнозування. Більшість цих даних можна інтерпретувати як часовий (чисельний) ряд.

Побудова адекватної моделі прогнозу передбачає вибір параметрів прогнозування та попередню обробку їх значень.

Маючи оброблені вихідні дані, не завжди одразу можна побудувати адекватну модель прогнозу, оскільки не всі параметри, за допомогою яких охарактеризована інтенсивність руху через ДК, на етапі побудови апріорної концептуальної моделі мають значення для побудови моделі; крім того, деякі параметри можуть чинити негативний вплив на модель.

Крім фільтрації і згладжування даних на практиці часто необхідно знайти значення деяких функцій, що характеризують певні властивості даних (наприклад, математичне сподівання, дисперсію), які будуть необхідні на наступних етапах прогнозу, тобто необхідно виконати статистичну обробку експериментальних даних.

На наступному етапі необхідно оцінити інформативність (значущість) параметрів (ознак) прогнозу і відібрати найбільш значущі з них для використання на наступних етапах прогнозу. Виконання цього етапу дозволяє не лише покращити модель, виключивши шкідливі та непотрібні параметри, але й спростити і здешевити процес прогнозу за рахунок скорочення вимірювань параметрів.

На жаль, при побудові концептуальної моделі прогнозу не завжди вдається вибрати такий набір ознак, який був би тісно пов'язаний з тими параметрами моделі, моделювання яких необхідно здійснити. Найчастіше деякі параметри, які тісно пов'язані з модельованими параметрами об'єкта пізнання, у набір ознак безпосередньо не включаються, а присутні в ньому побічно через інші параметри є пов'язаними з ними. Тому для з'ясування наявності таких прихованих параметрів і визначення їх зв'язку з параметрами, що входять в концептуальну модель, необхідно здійснювати етап аналізу параметрів прогнозу.

Аналіз останніх досліджень і публікацій, в яких започатковано вирішення даної проблеми та на які опирається автор. Загальне завдання прогнозування показників службової діяльності прикордонних структур частково розглядалось Кучковим А.Ф. і Лукашевичем Н.Ф. [1].

Завданням прогнозування на основі моделей часових (чисельних) рядів присвячено низку досліджень із використанням статистичних методів і суб'єктивних знань експертів, зокрема роботи Бокса Дж., Дженкінса Г., Боровікова В.П., Івченко Г.І. та інші [2; 3].

Недоліками вищезазначених підходів є: відсутність у моделей відомостей щодо структури й системи зв'язків реального об'єкта; труднощі побудови моделей за умови, що дані зберігаються в різних часових рядах і мають часові зміщення щодо один одного; недостатня точність прогнозу; значна чутливість отриманих результатів до недостатньої інформації; залежність результату прогнозу від кваліфікації аналітика в конкретній предметній сфері.

Зазначені недоліки можна долати із застосуванням апарата штучних нейро-нечітких мереж (далі – ШННМ), дослідження яких здійснено у [4].

Водночас відсутні дослідження стосовно вибору параметрів прогнозування із застосуванням ШННМ в оперативно-службовій діяльності.

Метою статті є розробка методичного підходу щодо вибору параметрів прогнозування інтенсивності руху осіб, транспортних засобів та вантажів у пунктах пропуску ДПСУ із застосуванням ШННМ.

Виклад основного матеріалу дослідження

Існують різні підходи до відбору інформативних ознак:

на основі дискретних методів пошуку в навчальній вибірці інформативної зони [5];

на основі методів кластеризації [6];

на основі припущення про нормальність розподілів об'єктів у кластерах [7];

на основі теоретико-інформаційного поняття ентропії [8];

на основі непараметричних оцінок щільності [9].

Розроблено та запропоновано формалізований підхід до вибору параметрів для прогнозування. Методологія вибору сукупності параметрів заснована на використанні факторного аналізу [10].

Об'єктом дослідження є інтенсивність руху осіб/транспортних засобів, стан якої можна описати сукупністю параметрів, що становлять собою вектор «службового» стану

$$\bar{Y} = (y_1, y_2, \dots, y_n),$$

де n – кількість параметрів системи (наприклад, кількість осіб, які перетнули пункт пропуску за одну добу, день тижня (включаючи святкові дні), різниця в ціні на «контрабандні товари» (сигарети, горілка, цукор, м'ясо тощо), курс валют, наявність у суміжній країні надзвичайної ситуації тощо).

Основним завданням є перехід від опису досліджуваної системи, заданої великим набором вимірюваних параметрів, до опису меншою кількістю максимально інформативних ознак, що відображають найбільш суттєві властивості системи. Передбачається, що кожен із параметрів y_i , що описує стан системи, може бути поданий як функція невеликої кількості загальних факторів p_i і характерного фактора u_i :

$$y_i = f(p_1, p_2, \dots, p_r, u_i). \quad (1)$$

Загальні фактори являють собою розрахункові змінні, тобто нові характеристики системи. Вихідною інформацією для проведення факторного аналізу є кореляційна матриця параметрів.

У подальшому використано такі умовні позначення:

j – номер об'єкту пропуску (особи, транспортні засоби, вантажі) ($j = 1, \dots, m$);

i – номер параметра контролю ($i = 1, \dots, n$);

k – номер виміру параметрів ($k = 1, \dots, s$).

Крок 1. Побудова кореляційної матриці $R_{ij} = \|r_{ik}\|$. На основі коваріаційної матриці будуватиметься кореляційна матриця параметрів $R_{ij} = \|r_{ik}\|$. Коефіцієнти парних кореляцій змінних розраховуються за формулою:

$$R_{ij} = \frac{C_{ij}^*}{\sqrt{C_{ii}^* \cdot C_{jj}^*}}, \quad (2)$$

де C_{ij}^* , C_{jj}^* , C_{ii}^* – коефіцієнти коваріації.

Отримана кореляційна матриця параметрів $R_{ij} = \|r_{ik}^h\|$ має відповідні значення елементів.

Крок 2. Побудова редуційованої кореляційної матриці $R_h = \|r_{ik}^h\|$, де $i, k = 1, \dots, n$. Для цього попередні оцінки спільностей h_i^2 отримують на основі обчислення квадрата множинної кореляції для кожної змінної за допомогою оберненої матриці R^{-1} за формулою:

$$h_i^2 = R_i^2 = 1 - \frac{1}{r_{ii}^{ii}}, \quad (3)$$

де r_{ii}^{ii} – діагональний елемент зворотної матриці.

Крок 3. Процедура виокремлення факторів і перехід до матриці факторного відображення $A = \|a_{il}\|$, елементами якої є факторні навантаження, де $i = 1, \dots, n$ – параметри; $l = 1, \dots, r$ – фактори.

Кореляційна матриця може бути відтворена за допомогою факторного відображення і кореляцій між факторами:

$$\begin{pmatrix} h_1^2 & r_{12} & \dots & r_{1n} \\ r_{21} & h_2^2 & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{n1} & \dots & \dots & h_n^2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1r} \\ a_{21} & a_{22} & \dots & a_{2r} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nr} \end{pmatrix} \times \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1r} & a_{2r} & \dots & a_{nr} \end{pmatrix} \quad (4)$$

або у матричному вигляді:

$$R_h = A \cdot A',$$

де R_h – редуційована кореляційна матриця;

h_i^2 – значення спільності;

A – матриця факторного відображення;

A' – транспонована матриця факторного відображення,

a_{il} – факторні навантаження.

Отримання результату рівняння (4) полягає у визначенні власних значень λ_i і власних векторів α_{il} матриці кореляцій.

Для знаходження векторів α_{il} і значень λ_i редуційованої кореляційної матриці $R_h = \left\| r_{ik}^h \right\|$ використовується ітераційний метод. Ітераційний процес починається з вибору вектора $\alpha^{(1)}$, елементи якого є першими наближеннями значень елементів власних векторів:

$$\beta^{(1)} = R_h \alpha^{(1)}, \quad (5)$$

$$\alpha^{(2)} = \beta^{(1)} / \max(\beta_i^{(1)}), \quad (6)$$

$$\alpha^{(k+1)} = \beta^{(k)} / \max(\beta_i^{(k)}), \quad (7)$$

де верхні індекси в скобках позначають шаг ітерації.

Формула (7) є загальною для k кроків ітерації.

Процес повторюється до тих пір, поки не досягаються збіжності до першого власного значення матриці $R_h = \left\| r_{ik}^h \right\|$: $\lambda_1 = \max(\beta_i^{(k)})$ і відповідному першому власному вектору $\alpha^{(k)}$.

Ітераційний процес завершується, коли $\alpha^{(k)}$ і $\alpha^{(k-1)}$ із достатньою точністю збігаються один з одним, тобто

$$\left| \alpha^{(k)} - \alpha^{(k-1)} \right| \leq \varepsilon.$$

Крок 4. Обчислення навантажень першого фактора a_{il} матриці факторного відображення $A = \left\| a_{il} \right\|$.

Навантаження матриці факторного відображення $A = \left\| a_{il} \right\|$ може бути визначено за компонентами власних векторів матриці $R_h = \left\| r_{ik}^h \right\|$ за формулою:

$$a_{il} = \alpha_{il} \sqrt{\lambda_1} / \sqrt{\alpha_{i1}^2 + \alpha_{i2}^2 + \dots + \alpha_{in}^2}, \quad (8)$$

де $i, l = 1, 2, \dots, n$.

Крок 5. Визначення матриці відтворених кореляцій R^+ з урахуванням тільки першого фактора і залишкової матриці кореляцій R_1 .

Матриця відтворених кореляцій R^+ визначається за формулою:

$$R^+ = a_1 \cdot a_1', \quad (9)$$

де a_1 – вектор-стовпчик навантажень першого фактора; a_1' – вектор-строка навантажень першого фактора.

Остаточну матрицю кореляцій R_1 , яка залишається після виокремлення першого фактора, визначаємо за формулою:

$$R_1 = R_h - R^+. \quad (10)$$

Крок 6. Виокремлення другого фактора і визначення другого власного значення λ_2 і другого власного вектора α_{i2} матриці R_h .

Розрахунок навантажень другого фактора матриці $A = \left\| a_{il} \right\|$.

Повторюємо процедуру обчислень з матрицею перших залишкових коефіцієнтів кореляції R_1 згідно з кроками 3 – 5 цього алгоритму.

Крок 7. Побудова матриці факторного відображення $A = \left\| a_{il} \right\|$, елементами якої є факторні навантаження, де $i = 1, \dots, n$ – параметри; $l = 1, \dots, r$ – фактори.

Матриця $A = \left\| a_{il} \right\|$ дозволяє визначити, скільки загальних факторів необхідно для відображення всіх кореляцій між змінними, а також навантаження кожного фактора для всіх змінних і навантаження всіх факторів для однієї змінної.

Крок 8. Розрахунок квадратів факторних навантажень a_{ik}^2 і дисперсій кожного фактора. Обчислення часткою дисперсії факторів згідно з матрицею факторного відображення $A = \left\| a_{il} \right\|$.

У результаті аналізу залишають у розгляді лише ті фактори, які мають максимальні дисперсії і забезпечують 90–95 % від повної дисперсії факторів. Фактори, що мають дисперсію менше 5 % від повної дисперсії, з розгляду виключаються.

Наприклад:

повна дисперсія	100 %
дисперсія першого фактора	91,23 %
дисперсія другого фактора	6,02 %
сумарна спільність	97,25 %
сумарна характерна дисперсія	2,75 %

Крок 9. Визначення коефіцієнтів пріоритетності параметрів здійснюється на основі вкладу факторних навантажень кожної змінної в загальний чинник, що забезпечує максимум дисперсії від повної дисперсії факторів.

На підставі проведених розрахунків видно, що на два перших фактори доводиться 97,25 % від пов-

ної дисперсії і лише 2,75 % припадає на сумарну характерну дисперсію, причому основна частина припадає на дисперсію першого фактора.

Тому оцінку значущості параметрів y_i виконують залежно від факторних навантажень у перший загальний фактор (табл. 1).

Таблиця 1

Значення коефіцієнта пріоритетності параметрів

Параметр y_i	y_6	y_4	y_1	y_3	y_2	y_5
Коефіцієнт пріоритетності	0,98889	0,95885	0,93705	0,92346	0,91083	0,69498
Ранг	1	2	3	4	5	6

Крок 10. Розрахунок інформативності I отриманої сукупності факторів і втрат інформації при виключенні окремих факторів.

Інформативність I сукупності двох перших факторів згідно з формулою:

$$I = \frac{\lambda_1 + \lambda_2}{\sum_{i=1}^n \lambda_i}$$

склала 0,9725 од. інф.

Втрати в інформативності при виключенні інших факторів складуть приблизно 2,75 % від повної інформативності. На підставі аналізу та розрахунків можна зробити висновок, що з усієї сукупності параметрів достатньо вибрати параметри з найбільшими коефіцієнтами пріоритетності, тобто y_6 , y_4 , для прогнозування інтенсивності руху. Інші параметри можна не враховувати, оскільки вони є сильно корельованими між собою і мають менші вагові коефіцієнти. Запропонована методика формалізованого вибору параметрів дозволяє не лише провести оцінку пріоритетності, побудувавши упорядковану послідовність параметрів у міру убування їх значущості за заданими критеріями, а й здійснити розрахунок інформативності факторів, виключити малозначні і малоінформативні фактори. За її допомогою можна здійснити науково обґрунтований вибір інформативною сукупністю параметрів для прогнозування інтенсивності руху через ДК, мінімізувати кількість параметрів для прогнозування без втрат в інформативності, зменшити розмірність масивів даних при обробці, а також скоротити витрати на їх прогноз. Для реалізації розрахунків і вибору параметрів за запропонованою методикою розроблено програмне забезпечення для ЕОМ.

Проведено оцінку ефективності розробленого методичного забезпечення прогнозування інтенсивності руху, включаючи розроблену методiku вибору параметрів та прогнозування. Для проведення порівняльної оцінки методик вибору параметрів або методик здійснення процесу прогнозування з розробленими методиками необхідно використовувати узагальнений показник, який включає в себе часткові показники параметрів. У загальному випадку

узагальнений показник може бути поданий у вигляді функції:

$$K = F(P_1, P_2, P_3, \dots, P_n), \quad (11)$$

де $P_1, P_2, P_3, \dots, P_n$ – часткові показники вибору параметрів.

Узагальнений показник має вигляд:

$$K = \sum_{i=1}^n P_i * W_i, \quad (12)$$

де W_1, W_2, \dots, W_n – вагові коефіцієнти.

При обчисленні узагальненого показника порівняльної оцінки необхідно враховувати значущість кожного часткового показника методик за допомогою вагових коефіцієнтів, які визначають методами експертних оцінок або за методикою аналізу ієрархій.

Як впливає з табл. 2, найбільш ефективною методикою для вибору параметрів прогнозу складових частин інтенсивності руху є методика на основі використання методу факторного аналізу.

Реалізація методичного забезпечення вибору параметрів призводить до підвищення достовірності прогнозування та зменшення економічних витрат.

Висновки

Отже, у статті було розроблено методiku формалізованого вибору сукупності параметрів для прогнозування інтенсивності руху через ДК осіб, транспортних засобів та вантажів, що дозволяє не лише провести оцінку пріоритетності з урахуванням їх взаємозв'язку, побудувавши упорядковану послідовність параметрів у міру убування їх значущості за заданими критеріями, а й здійснити розрахунок інформативності, вибрати найбільш значущі параметри, мінімізувати кількість параметрів без втрат в інформативності.

Дослідження засвідчили, що цей підхід надає можливість обґрунтовувати кількісний склад змін прикордонних нарядів у пунктах пропуску.

Перспективою подальших розвідок у даному напрямку є розробка методів пошуку вхідних параметрів моделі для прогнозу показників у діяльності ДПСУ з метою підвищення точності прогнозу.

Узагальнені показники методик вибору параметрів

Назва методик	Узагальнений показник		
	Особи	Транспортні засоби	Вантажі
Методика вибору параметрів на основі інформаційного підходу	0,65	0,67	0,73
Методика вибору параметрів на основі на основі методів кластеризації	0,59	0,50	0,68
Методика вибору параметрів на основі на основі непараметричних оцінок щільності	0,65	0,69	0,72
Методика вибору параметрів на основі методу факторного аналізу	0,70	0,71	0,75

Список літератури

1. Кучков А.Ф. Математическое моделирование служебно-боевых действий пограничных войск: учебник в 3-х томах / [А.Ф. Кучков, Н.Ф. Лукашевич, Г.П. Попов, В.В. Шумов]. – М.: Академия ФПС России, 1997. – Т. 1. – 195 с.; Т. 2. – 191 с.; Т. 3. – 240 с.

2. Бокс Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс. – М.: Мир, 1974.

3. Боровиков В.П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере / В.П. Боровиков, Г.И. Ивченко. – М.: Финансы и статистика, 2000. – С. 320.

4. Андрушко В.З. Методика прогнозування інтенсивності роботи пунктів пропуску із застосуванням нейронної моделі / В.З. Андрушко // Збірник наукових праць Національної академії Державної прикордонної служби України. Серія: військові та технічні науки / [голов. ред. Олексієнко Б.М.]. – Хмельницький: Вид-во НАДПСУ, 2014. – № 2(62). Частина II. – С. 6-17.

5. Дмитриев А.И. О математических принципах классификации предметов или явлений / А.И. Дмитриев,

Ю.И. Журавлев, Ф.П. Кренделев // Дискретный анализ. – Новосибирск: ИМ СО АН СССР, 1966. – Вып. 7. – С. 1-17.

6. Загоруйко Н.Г. Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. – Новосибирск: Изд-во Ин-та математики, 1999. – 270 с.

7. Воронцов К.В. Лекции по методам оценивания и выбора моделей [Электронный ресурс] / К.В. Воронцов. – 2007. – Режим доступа к ресурсу: www.ccas.ru/voron/download/Modeling.pdf.

8. Перегудов Ф.И. Введение в системный анализ / Ф.И. Перегудов, Ф.П. Тарасенко. – М.: Высшая школа, 2001. – 396 с.

9. Орлов А.И. Анализ нечисловой информации в социологических исследованиях / А.И. Орлов. – М.: Наука, 1985. – С. 58-92.

10. Иберла К. Факторный анализ / К. Иберла. – М.: Статистика, 1980. – 399 с.

Надійшла до редколегії 16.09.2015

Рецензент: д-р техн. наук, проф. О.С. Андрушук, Національна академія Державної прикордонної служби України імені Богдана Хмельницького, Хмельницький.

ВЫБОР ПОКАЗАТЕЛЕЙ НЕЙРО-НЕЧЕТКОЙ МОДЕЛИ ПРОГНОЗИРОВАНИЯ ИНТЕНСИВНОСТИ ДВИЖЕНИЯ НА ГОСУДАРСТВЕННОЙ ГРАНИЦЕ

В.З. Андрушко

Предложена методика формализованного выбора параметров прогноза интенсивности движения лиц, транспортных средств через государственную границу, которая позволяет не только провести оценку приоритетности, построить упорядоченную последовательность параметров по мере убывания их значимости по заданным критериям, но и осуществить расчет информативности факторов, исключить малозначительные и малоинформативные факторы. С ее помощью можно осуществить научно обоснованный выбор информативной совокупности параметров для прогнозирования, минимизировать количество параметров для прогнозирования без потерь в информативности, уменьшить размерность массивов данных при обработке, а также сократить расходы на их прогноз.

Ключевые слова: прогноз, параметры, факторный анализ.

THE CHOICE OF INDICATORS NEURO-FUZZY MODEL PREDICTION TRAFFIC INTENSITY ON THE STATE BORDER

V.Z. Andrushko

The technique formalized selection parameters of the forecast traffic of persons and vehicles across the state border, which allows for not only the assessment of priority, building an ordered sequence of parameters in descending order of importance of criteria, but also to carry out the calculation of informativeness factors exclude minor and uninformative factors. You can use it to implement evidence-based selection of informative set of parameters to predict, minimize the number of parameters to predict without loss of information content, to reduce the dimensionality of data sets in the processing, as well as reduce the costs and the forecast.

Keywords: forecast, parameters, factor analysis.