

*В даній роботі розглядаються питання формування моделі прогнозування кореспонденцій потоків на залізничних лініях при впровадженні швидкісних пасажирських поїздів. Запропоновано процедуру настроювання моделі прогнозування на основі генетичного алгоритму з дійсним кодуванням. Проведені експериментальні дослідження підтверджують, що запропонована процедура настроювання є стійкою, прийнятно складною та дозволяє підвищити точність прогнозування*

*Ключові слова: залізничний транспорт, швидкісні перевезення, прогнозування, генетичний алгоритм*

*В данной работе рассматриваются вопросы формирования модели прогнозирования корреспонденций потоков на железнодорожных линиях при внедрении скоростных пассажирских поездов. Предложена процедура настройки модели прогнозирования на основе генетического алгоритма с действительным кодированием. Проведенные экспериментальные исследования подтверждают, что предложенная процедура настройки модели является устойчивой, приемлемо сложной и позволяет повысить точность прогнозирования*

*Ключевые слова: железнодорожный транспорт, скоростные перевозки, прогнозирование, генетический алгоритм*

# ПРОЦЕДУРА ФОРМУВАННЯ МОДЕЛІ ПРОГНОЗУВАННЯ ПАСАЖИРОПОТОКІВ НА ЗАЛІЗНИЧНИХ ЛІНІЯХ

**Л. О. Пархоменко**

Асистент

Кафедра управління експлуатаційною  
роботою

Українська державна академія залізничного  
транспорту

майд. Фейєрбаха, 7, м. Харків, Україна, 61050

E-mail uermp@ukr.net

## 1. Вступ і актуальність теми

Одним із найбільш важливих етапів в плануванні розвитку пасажирського швидкісного руху на залізничному транспорті є проведення прогнозування попиту на перевезення. Від достовірності довгострокових прогнозів формування пасажиропотоків на залізничних лініях залежить ефективність інвестиційних проєктів з підвищення швидкості руху пасажирських поїздів. Найбільш прийнятним напрямком розробки прогнозів є застосування різних математичних моделей [1, 2].

Враховуючи, що при прогнозуванні обсягів перевезень на період від п'яти і більше років динаміка показника може кардинально змінитися із-за впливу великої кількості факторів, застосування одномірних математичних моделей, що засновані на аналізі часових рядів, є неприйнятним. За таких умов при довгостроковому прогнозуванні поширена практика використання багатовимірних регресійних моделей [3], що дозволяють врахувати вплив найбільш значимих факторів на обсяги перевезень.

## 2. Постановка задачі

В практиці прогнозування транспортних кореспонденцій набули широкого застосування так звані "гравітаційні" моделі [4, 5]. В основі даних моделей лежить сформульована більше ста років назад теорія, за якою обсяги перевезень між двома містами підпорядковуються закону обернених квадратів, відповідно

до закону гравітації [4]. На сучасному етапі наукових досліджень найбільш популярні версії гравітаційних моделей використовують ймовірнісну постановку, крім того моделі доповнюються різними варіантами показників, що підвищують точність прогнозування [6].

Враховуючи відсутність реалізації гравітаційних моделей з прийнятною точністю (похибка складає понад 20%) для залізничного транспорту дана робота присвячена продовженню досліджень [7] щодо реалізації моделі прогнозування кореспонденцій в умовах впровадження швидкісних пасажирських поїздів на основі нечітких реляційних обчислень.

## 3. Процедура настроювання моделі прогнозування пасажиропотоків на основі нечітких реляційних обчислень за допомогою генетичного алгоритму

Математичну модель задачі прогнозування кореспонденцій в умовах впровадження швидкісних пасажирських поїздів запропоновано описати за допомогою нечітких реляційних рівнянь виду [7]

$$R \circ S = T, \quad (1)$$

де  $R_{m \times p}$  – матриця нечітких входів,  $S_{p \times n}$  – матриця нечітких відношень,  $T_{m \times n}$  – матриця нечітких виходів на дискретних чітких множинах  $X, Y, Z$  кінцевих потужностей  $m, p, n$  відповідно; правило композиції  $\circ$ , засновано на максимінному критерії.

Входи моделі представлено як лінгвістичні змінні (ЛЗ) [8].

Перші два параметри моделі можна уявити як ЛЗ  $\Lambda_1 = \{\text{населеність міста } M_1\}$  та  $\Lambda_2 = \{\text{населеність міста } M_j\}$ , значення яких визначаються на єдиній шкалі з трьох термів  $T_{1,2} \forall \{ \text{“мала”}-t_1, \text{“середня”}-t_2, \text{“велика”}-t_3 \}$ . Для формалізації термів запропоновано використати функції належності (ФН), що формуються із застосуванням кусочно-лінійної апроксимації. Крайні терми  $t_1$  та  $t_3$  описуються ФН, що мають трапецієподібну конфігурацію відповідно відкрити ліворуч та праворуч (рис. 1, а), тоді як терм  $t_2$  описується трикутною залежністю.

Наступні параметри можна записати лінгвістичними змінними  $\Lambda_3 = \{\text{індекс тривалості подорожі}\}$ ,  $\Lambda_4 = \{\text{індекс транспортної доступності}\}$  та  $\Lambda_5 = \{\text{індекс відвідуваності міста прибуття}\}$  із наступними терм-множинами  $T_{3,4,5} \forall \{ \text{“низький”}-t_1, \text{“середній”}-t_2, \text{“високий”}-t_3 \}$  та відповідними трикутними функціями належності (рис. 1, б). Схематична візуалізація графіків функцій належності нечітких термів вхідних змінних  $\Lambda_{1-5}$  наведено на рис. 1.

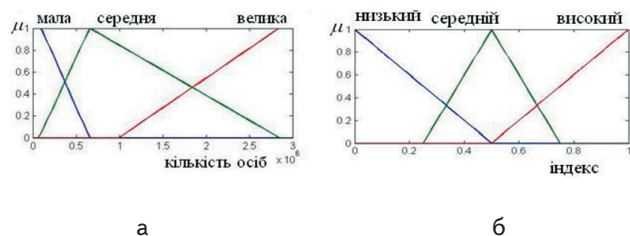


Рис. 1. Графіки функцій належності нечітких термів вхідних змінних, що відображаються залежністю нечіткого ступеня належності  $\mu$  нечіткій множині термів від всіх можливих значень відповідно кількості осіб та індексу на інтервалі  $[0, 1]$ ; а – терм-множина ЛЗ  $\Lambda_{1-2}$ ; б – терм-множина ЛЗ  $\Lambda_{3-5}$

Приймаючи до уваги, що при фазифікації вхідних параметрів кожному з них ставиться у відповідність по три нечітких значення функції належності до кожного з термів, загальний розмір вхідної вектор-строки становитиме  $p = 3 \times 5 = 15$ , тобто

$$R = (\underbrace{\mu_{(t_1)}, \mu_{(t_2)}, \mu_{(t_3)}}_{\Lambda_1}, \underbrace{\mu_{(t_1)}, \mu_{(t_2)}, \mu_{(t_3)}}_{\Lambda_2}, \dots, \underbrace{\mu_{(t_1)}, \mu_{(t_2)}, \mu_{(t_3)}}_{\Lambda_5}), m=1, p=15. \quad (2)$$

Вихідні дані моделі запропоновано формалізувати також у вигляді ЛЗ  $\Omega = \{\text{обсяги кореспонденції пасажирів із міста } i \text{ до міста } j\}$  з терм-множиною  $T_\Omega \forall \{ \text{“низькі”}-t_1, \text{“середні”}-t_2, \text{“високі”}-t_3 \}$  з трикутними функціями належності. За таких умов вихідний вектор має розмірність  $n = 3$ , тобто

$$T = (\underbrace{\mu_{(t_1)}, \mu_{(t_2)}, \mu_{(t_3)}}_{\Omega}), m=1, n=3. \quad (3)$$

Для визначення залежності між вхідними параметрами та виходом моделі необхідно знайти матрицю нечітких відношень  $S_{p=15 \times n=3}$ . Отже, в рівнянні (1) є невідомим  $S$ . За аналогією з матричною алгеброю дане рівняння можна назвати лівим нечітким реляційним рівнянням [9]. В загальному вигляді очевидно, що для композиції максимум-мінімум існує єдине максимальне і декілька мінімальних рішень. Зрозуміло, іноді рішень немає. З практичної точки зору застосування для рішення нечітких рівнянь аналітичних

методів призведе до виникнення ситуації, коли рівняння нерозв'язне, в такому випадку прогнозування теж неможливе, що недопустимо. Для уникнення даного недоліку в роботі запропоновано застосувати наближений метод рішення нечіткого реляційного рівняння на основі генетичного алгоритму з дійсним кодуванням [10].

Крім того, по суті нечітке відношення в рівнянні (1) є реляційною базою правил, що повинна відображати не тільки один із варіантів залежності вхідних параметрів від виходу моделі, але й цілий спектр таких варіантів. Це дозволить при зміні вхідних параметрів моделі знайти загальний вплив кожного із них на вихід моделі. За таких умов в роботі запропоновано застосувати так званий метод навчання моделі прогнозування на основі навчальної вибірки [11], яка пов'язує вхід з виходом залежності, що досліджується. Під навчальною вибіркою слід розуміти два вектори рівняння (1), які відповідають одному із варіантів залежності вхідних факторів від обсягів кореспонденцій пасажирів між різними містами. Навчальну вибірку можна позначити наступним чином:

$$\langle R_i, y_i \rangle, \quad i = \overline{1, K}, \quad (4)$$

де  $R_i$  – вектор-строка нечітких входів рівняння (1), що відповідає  $i$ -му варіанту даних,  $R_{i \times 15} \forall \mu_{(t)} \in [0, 1]$ ;  $y_i$  – обсяг кореспонденцій пасажирів між містами, що відповідає  $i$ -му варіанту даних, пас;  $K$  – обсяг вибірки. Слід зазначити, що вхідний вектор навчальної вибірки є фазифікованим відповідно до нечітких термів лінгвістичних змінних, що наведені вище, тоді як вихід представлено чітким значенням відповідно до реальних даних щодо обсягів кореспонденцій.

Навчання моделі полягає у знаходженні мінімуму середньої відносної похибки розбіжності між результатами прогнозування та фактичними даними із навчальної вибірки

$$F = \frac{1}{K} \sum_{i=1}^K \left| \frac{y_i - y_i^f}{y_i} \right| \times 100\% \rightarrow \min, \quad (5)$$

де  $y_i$  – вихідний вектор навчальної вибірки,  $y_i^f$  – результат прогнозування моделі при значенні входів із  $i$ -ї строки вибірки  $\langle R_i, y_i \rangle$ .

Результат прогнозування моделі  $y_i^f$  запропоновано знаходити на основі операції дефазифікації нечіткої множини, що відповідає вектору  $T$ , який отримано на основі прямого рішення рівняння (1) з нечітким вхідним вектором з навчальної вибірки  $R_i$  та знайденим за допомогою еволюційного моделювання відношенням  $S_{15 \times 3}$ . На першому етапі приведення до чітких значень виходу рівняння (1) проводиться операція знаходження мінімального зрізу нечітких множин відповідно до значень вектора  $T$ . Після чого за допомогою операції максимум проводиться об'єднання усічених нечітких множин трьох термів ЛЗ  $\Omega = \{\text{обсяги кореспонденції пасажирів із міста } i \text{ до міста } j\}$ . На останньому етапі над знайденою об'єднаною нечіткою множиною проводиться операція дефазифікації на основі методу центру тяжіння [8].

Враховуючи, що вихід моделі залежить від параметрів функцій належності в роботі запропоновано для

підвищення точності настроювання моделі прогнозування застосувати в межах генетичного алгоритму (ГА) спосіб настроювання функцій належності термів вихідної ЛЗ  $\Omega$ . Для цього додатково до хромосоми ГА введено п'ять генів, що відповідають параметрам функцій належностей терм-множини ЛЗ  $\Omega$ . За таких умов загальна хромосома ГА складається з двох частин:

$$Ch_j = (C^1, C^2), \tag{6}$$

де  $C^1 = (s_{1,1}, s_{1,2}, s_{1,3}, s_{2,1}, s_{2,2}, s_{2,3}, \dots, s_{i,j}, \dots, s_{p,n})$  – частина хромосоми, що відповідає значенням відношення  $S_{рхн}$  в рівнянні (1),  $s_{i,j} \in [0,1]$ ;  $C^2 = (c_{t_1}, a_{t_2}, b_{t_2}, c_{t_2}, a_{t_3})$  – частина хромосоми, що відповідає параметрам ФН;  $a_{t_2}, a_{t_3}$  – перший параметр трикутної функції належності відповідно термів  $t_2$  та  $t_3$ ;  $b_{t_2}$  – другий параметр ФН терма  $t_2$ ;  $c_{t_1}, c_{t_2}$  – третій параметр ФН відповідно термів  $t_1$  та  $t_2$ .

На кожен з даних параметрів ФН накладено обмеження, що забезпечують лінійну впорядкованість елементів терм-множин ЛЗ  $\Omega$  (рис. 2). Слід зазначити, в роботі для зменшення складності моделі параметри ФН  $a_{t_1}, b_{t_1}, b_{t_3}, c_{t_3}$  прийняті постійними.

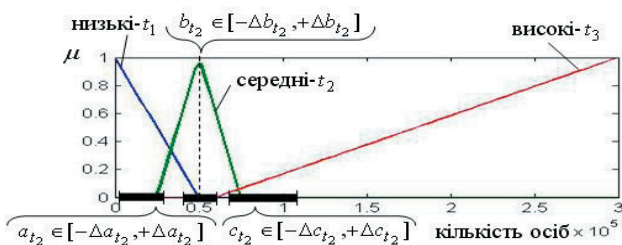


Рис. 2. Залежність нечіткого ступеня належності  $\mu$  нечіткій множині термів лінгвістичної змінної  $\Omega$  від всіх можливих значень кількості осіб та схематичне представлення інтервалів зміни параметрів функції належності

Для рішення задачі настроювання моделі прогнозування на основі нечітких реляційних рівнянь в роботі використано генетичний алгоритм, в якому був застосований арифметичний кросовер та випадкова мутація [12]. В якості фітнес функції була прийнята функція (5). Запропонована процедура еволюційного моделювання дозволила при популяції рівній 40 та за 12000 генерацій знайти достатньо оптимальне рішення, при якому відносна похибка моделі прогнозування на тестовій вибірці становить не більше 10,0144%. Результати роботи генетичного алгоритму з дійсним кодуванням для настроювання моделі прогнозування наведено на рис. 3, а, б.

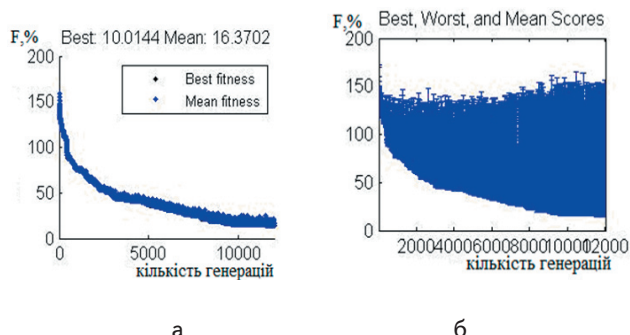


Рис. 3. Графіки залежності значень цільової функції F: а – найкращих значень функції F; б – найменших, найбільших, і середніх значень функції F для кожного покоління від кількості генерацій алгоритму

Візуалізація функцій належності терм-множини вихідної лінгвістичної змінної  $\Omega = \{\text{обсяги кореспонденції пасажирів із міста і до міста j}\}$  після процедури настроювання моделі прогнозування наведено на рис. 4.

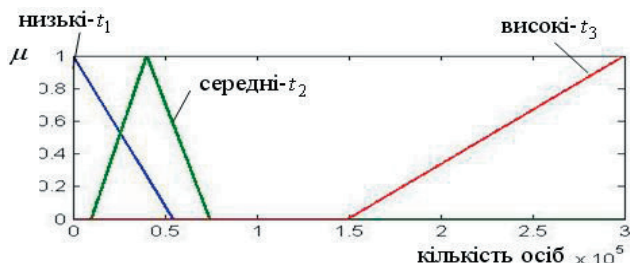


Рис. 4. Залежність нечіткого ступеня належності  $\mu$  нечіткій множині термів лінгвістичної змінної  $\Omega$  від всіх можливих значень кількості осіб після проведення процедури настроювання моделі прогнозування

#### 4. Висновки

Запропонована процедура настроювання моделі прогнозування на основі нечітких реляційних обчислень є стійкою, прийнятно складною та дозволяє підвищити точність прогнозування. На базі запропонованого підходу є можливим в перспективі реалізувати автоматизований програмний комплекс прогнозування кореспонденцій між містами на період стратегічного планування. Даний програмний комплекс є необхідною складовою для формування системи підтримки прийняття рішень щодо вибору стратегії розвитку швидкісного пасажирського руху на залізничній мережі.

#### Література

1. Лукашин, Ю. П. Адаптивные методы краткосрочного прогнозирования [Текст] / Ю. П. Лукашин. – М.: Статистика, 1979. – 254 с.
2. Прохорченко, А. В. Удосконалення системи прогнозування пасажиропотоків на тактичному рівні планування перевезень [Текст] / А. В. Прохорченко, О. М. Данко, С. М. Журко Збірник наукових праць УкрДАЗТ. – 2009. – Вип. 102. – С. 60-67.
3. Єріна, А. М. Статистичне моделювання та прогнозування [Текст]: навч. посібник / А. М. Єріна. – К.: КНЕУ, 2001. – 170 с.
4. Carrothers, G. A. P. An historical review of the gravity and potential concepts of human interaction [Text] / G. A. P. Carrothers // J. of the American Instit. Planners. – 1956. –V. 22. – P. 94-102.

5. Doganis, R. Traffic forecasting and the gravity model [Text] / R. Doganis // Flight International. – 1966. – P.547–549.
6. Погребняк, Е. Б. Анализ методов формирования матрицы корреспонденций транспортной сети города [Текст] / Е. Б. Погребняк, Н. И. Самойленко // Коммунальное хозяйство городов. – Харьковская национальная академия городского хозяйства, 2006. – № 69. – С. 121-126.
7. Пархоменко, Л. О. Розроблення моделі прогнозування кореспонденцій пасажирів в умовах впровадження залізничного швидкісного пасажирського сполучення на основі нечітких реляційних обчислень [Текст] / Л. О. Пархоменко // Збірник наукових праць УкрДАЗТ. – 2012. – Вип. 131. – С. 109-115.
8. Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы [Текст] / Д. Рутковская, М. Пилинский, Л. Рутковский; пер.с польск. И. Д. Рудинского. – М.: Горячая линия. – Телеком, 2004. – 452 с.
9. Peeva, K. Fuzzy Relational Calculus Theory: Applications and Software [Text] / Ketty Peeva, Yordan Kyosev // World Scientific Publishing Co. Pte. Ltd. – 2005. – 291 p.
10. Wright, A. "Genetic algorithms for real parameter optimization"[Text] / A. Wright // Foundations of Genetic Algorithms. – 1991. – V. 1. – P. 205-218.
11. Ротштейн, А. П. Идентификация нелинейной зависимости нечеткой базой знаний с нечеткой обучающей выборкой [Текст]/ А. П. Ротштейн, С. Д. Штовба // Кибернетика и системный анализ. – 2006. – No2. – С. 17-24.
12. Кисляков, А. В. Генетические алгоритмы: операторы скрещивания и мутации [Текст] / А. В. Кисляков // Информационные технологии. – 2001. – №1. – С. 29-34.

*Виконано побудову математичних моделей для задач кількісного оцінювання груп ризику інфікування вірусом імунодефіциту людини. Проведено аналіз технологій ідентифікації та кластеризації для розв'язання задачі оцінювання кількості представників груп ризику та задачі визначення міри належності особи до такої групи*

*Ключові слова: група ризику інфікування вірусом імунодефіциту людини, задача кластеризації, задача ідентифікації*

*Выполнено построение математических моделей для задач количественной оценки групп риска инфицирования вирусом иммунодефицита человека. Проведен анализ технологий идентификации и кластеризации для решения задачи оценки числа представителей групп риска и задачи определения степени принадлежности лица к такой группе*

*Ключевые слова: группа риска инфицирования вирусом иммунодефицита человека, задача кластеризации, задача идентификации*

УДК 004.94

# ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ КІЛЬКІСНОГО ОЦІНЮВАННЯ ГРУП РИЗИКУ ІНФІКУВАННЯ ВІРУСОМ ІМУНОДЕФІЦИТУ ЛЮДИНИ

О. Ю. Мулеса

Викладач

Кафедра кібернетики і прикладної математики  
Ужгородський національний університет  
вул. Університетська, 14, м. Ужгород, Україна, 88000  
E-mail: mulesa.oksana@gmail.com

## 1. Вступ

Світова спільнота об'єднує свої зусилля для подолання епідемії СНІДу [1]. Так, Об'єднаною програмою Організації Об'єднаних Націй з ВІЛ/СНІДу було розроблено 10 конкретних цільових показників на 2015 рік для направлення сумісних дій, серед яких [1]:

1. Знизити передачу ВІЛ статевим шляхом на 50% до 2015 року.

2. Знизити передачу ВІЛ серед споживачів ін'єкційних наркотиків (СІН) на 50% до 2015 року.

3. Викоринити нові випадки ВІЛ-інфікування серед дітей до 2015 року та значно знизити рівень материнської смертності внаслідок СНІДУ тощо.

Україна є однією з країн Східної Європи, найбільш охоплених епідемією ВІЛ-інфекції/СНІДу [2]. Згідно з останніми офіційними даними, масштаби епідемії ВІЛ-інфекції/СНІДу в Україні мають зростаючу динаміку, незважаючи на комплекс заходів з