



СТВОРЕННЯ БАЗ НЕЧІТКИХ ЗНАТЬ ДЛЯ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ УПРАВЛІННЯ

Глоба Л.С.¹⁾, Терновой М.Ю.²⁾, Штогріна О.С.³⁾

- ¹⁾ Національний технічний університет України “Київський політехнічний інститут”,
Індустріальний пров., 2, Київ, 03056, Україна, e-mail: gls@densoft.com.ua
- ²⁾ Національний технічний університет України “Київський політехнічний інститут”,
Індустріальний пров., 2, Київ, 03056, Україна, e-mail: maximter@mail.ru
- ³⁾ Національний технічний університет України “Київський політехнічний інститут”,
Індустріальний пров., 2, Київ, 03056, Україна, e-mail: L_Shtogrina@mail.ru

Резюме: В роботі запропоновано підхід до створення бази нечітких знань, що є частиною інтелектуальної системи керування складними адміністративними системами. Він складається з двох основних етапів, а саме етапу формування структури та етапу настроювання параметрів функції належності. Запропонований підхід дозволяє прискорити створення бази нечітких знань для інтелектуальних систем керування.

Ключові слова: база нечітких знань, лінгвістична змінна, функція належності, нечітка нейрона мережа, NEFCLASS, генетичне програмування, складні адміністративні системи.

1. ВСТУП

Державні установи такі як міністерства та відомства відносяться до складних адміністративних систем (САДС). Вони мають розвинену ієрархічну структуру, що складається з багатьох підрозділів та підлеглих установ, які, як правило, географічно рознесені по великій території [1]. Інтелектуальні системи управління складають основу інформаційно-аналітичної діяльності для задач управління такими об'єктами, а обробка інформації є однією з найбільш складних частин процесу управління. Це обумовлено великою кількістю складних інформаційних потоків, що циркулюють в таких об'єктах, високою динамікою зміни їх структури, яка обумовлена законодавчими змінами та змінами керівництва, наявністю на кожному рівні суб'єктивних факторів, необхідністю врахування при обробці як кількісної так і якісної інформації.

Оскільки інформація характеризується тим чи іншим ступенем невизначеності, необхідно використовувати такі математичні методи і підходи, які включають засоби для представлення і оперування інформацією за наявності різних типів невизначеності у описанні об'єкту, а саме лінгвістичної невизначеності та необхідності врахування як кількісної так і якісної інформації. Серед таких методів та

підходів можна виділити три великі групи [2, 3, 4]: логічні, мережеві та еволюційне моделювання.

З наведених трьох основних типів методів лише логічні та мережеві методи можуть використовуватися для побудови моделі обробки інформації в умовах притаманних САДС, оскільки методи еволюційного моделювання не мають власної моделі знань.

Методи еволюційного моделювання є за своєю природою оптимізаційними методами та можуть використовуватися при побудові та настроюванні структури моделі [5, 6].

Для моделювання швидкоплинних локальних у часі процесів застосування математичних методів і моделей, які потребують навчання для виявлення закономірностей функціонування системи неможливо. У цьому випадку основним способом формування моделі в умовах невизначеності стає формалізація експертних знань і представлень у вигляді баз нечітких знань (БНЗ), використовуючи методи нечіткої логіки. Основна ідея полягає в тому, що настроюючи БНЗ, можна ідентифікувати нелінійні залежності з необхідною точністю. Перевагою даного підходу є також можливість врахування недетермінованих впливів, структурних і якісних змін у складі системи в процесі моделювання. Також дуже корисною особливістю цього підходу є можливість представлення результатів

моделювання у відносно простій і наочній формі, яка буде зрозумілою людині. До недоліків цього підходу можна віднести неупорядкованість та надлишковість початкової БНЗ, яка формується виходячи з експертних знань, та внаслідок цього важкість відповідей на питання “чому?” і “як?”.

Це ускладнює процес прийняття рішень, щодо необхідних напрямків управління для покращення ситуації.

При побудові моделей для проведення обробки інформації в САДС з відносно постійним складом учасників на середньо- і довгостроковому часовому інтервалі з'являється можливість застосовувати в якості вихідних даних моделі більш об'єктивну інформацію, а саме інформацію про стан системи в періоди її стійкості [7]. Тому додатково до сформованої за допомогою методів нечіткої логіки моделі можна використовувати методи теорії штучних нейронних мереж та еволюційних обчислень для проведення настроювання. Саме підхід до створення та настроювання БНЗ з використанням методів нечітких нейронних мереж та еволюційних обчислень запропоновано в даній роботі.

2. ОСНОВНА ЧАСТИНА

База нечітких знань є однією з головних складових частин інтелектуальних систем управління. Вона представляє собою об'єднання часткових БНЗ, кожна з яких визначає залежність результуючої лінгвістичної змінної (ЛЗ) від вхідних ЛЗ [8]. В свою чергу часткові БНЗ представляють собою сукупність правил виду “ЯКЩО .. ТА .. ТА .. АБО .. ТА .. ТА .. АБО .. ТА .. ТА .., ТО ...”. Кожне таке правило визначає залежність терму результуючої ЛЗ від термів вхідних ЛЗ.

В загальному випадку результуючих змінних може бути декілька, що обумовлено необхідністю розглядати різні аспекти проблеми. З іншого боку для визначення різних результуючих змінних можуть бути використані однакові вхідні змінні або їх частина. Тому побудова та настроювання БНЗ повинні проводитись одночасно для всіх ЛЗ.

Постановка задачі побудови та настроювання БНЗ наведена нижче.

Дано:

1) $\{x_i \mid i = \overline{1, v}\}$ – множина вхідних змінних;

2) $\{X_{li} \mid i = \overline{1, v}\}$ – множина ЛЗ, що відповідають $\{x_i \mid i = \overline{1, v}\}$;

3) $\{X_{zi}^z \mid i = \overline{1, v}, z = \overline{1, Z_{li}}\}$ – терм-множини ЛЗ з пункту 2, де

$z = \overline{1, Z_{li}}$ – номер терму вхідної ЛЗ X_{li} , Z_{li} – загальна кількість термів вхідної ЛЗ;

4) $\{\mu_{x_i} \mid i = \overline{1, v}, z = \overline{1, Z_{li}}\}$ – множина ФН, відповідних до термів ЛЗ з пункту 3, за якими буде проводитись фазифікація;

5) $\{X_{2i} \mid i = \overline{1, M_2}\}$ – множина результуючих ЛЗ;

6) $\{X_{2i}^z \mid z = \overline{1, Z_{2i}}\}$ – терм-множина ЛЗ з пункту 5;

7) $\{(x^1, X_2^1) \mid l = \overline{1, R_{Study}}\}$ – вибірка для побудови та настроювання, де $x^1 = \langle x_1^1, x_2^1, \dots, x_v^1 \rangle$ – 1-й набір значень вхідних змінних, $X_2^1 = \langle X_{21}^1, X_{22}^1, \dots, X_{2N}^1 \rangle$ – 1-й набір значень результуючих ЛЗ;

8) $\{(x^1, X_2^1) \mid l = \overline{1, R_{Test}}\}$ – вибірка для тестування

Знайти:

1) $\hat{R}_2 = \bigcup_1 \hat{R}_{2i}$ – базу нечітких знань, що

визначає залежність результуючих ЛЗ від вхідних ЛЗ;

2) Настроїти $\{\mu_{x_i} \mid i = \overline{1, v}, z = \overline{1, Z_{li}}\}$.

Для вирішення цієї задачі запропоновано підхід до побудови та настроювання БНЗ, структура якого наведена на рис. 1. Він складається з двох основних етапів: етапу формування структури та етапу настроювання параметрів функцій належності (ФН).

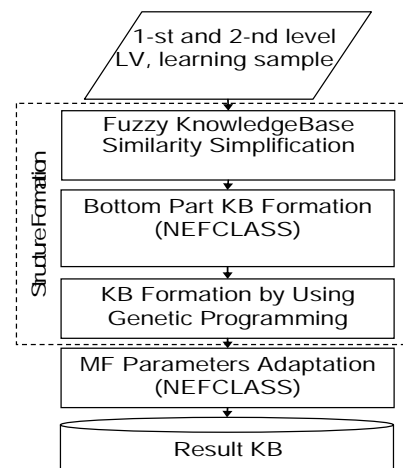


Рис. 1 – Етапи побудови та настроювання БНЗ.

Етап формування структури проводиться окремо для кожного терму кожної ЛЗ. Це обумовлено тим, що цей етап є незалежним для кожного терму. Часткова БНЗ в цьому випадку

буде отримана як об'єднання всіх правил для визначення кожного терму ЛЗ. А загальна БНЗ буде отримана як результат об'єднання всіх часткових БНЗ.

Етап формування структури в свою чергу виконується в три кроки.

Перший крок полягає в спрощенні за подібністю множин термів, що відповідають вхідним лінгвістичним змінним [9]. Для цього використовується міра подібності (1).

$$S(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

де A, B – нечіткі множини, $|\cdot|$ – потужність множини.

У випадку коли міра перевищує заданий поріг $\theta \in [0, 1]$ подібні нечіткі множини об'єднуються. Злиття зменшує кількість різних нечітких множин, що відповідають термам ЛЗ і використовуються при побудові БНЗ, і тим самим, підвищують прозорість при розв'язанні задачі класифікації. Коли всі терми однієї ЛЗ мають такі нечіткі множини, які відповідають універсальній множині, або коли злиття призводить до того, що терм-множина ЛЗ містить лише один терм, то такі ЛЗ видаляються з розгляду.

На другому кроці проводиться створення початкової структури БНЗ. Ця процедура виконується за допомогою алгоритмів, що базуються на методах та алгоритмах побудови нечіткої нейронної мережі NEFCLASS [10, 11].

Для використання цих алгоритмів БНЗ представляється у вигляді чотиришарової нечіткої нейронної мережі (рис. 2).

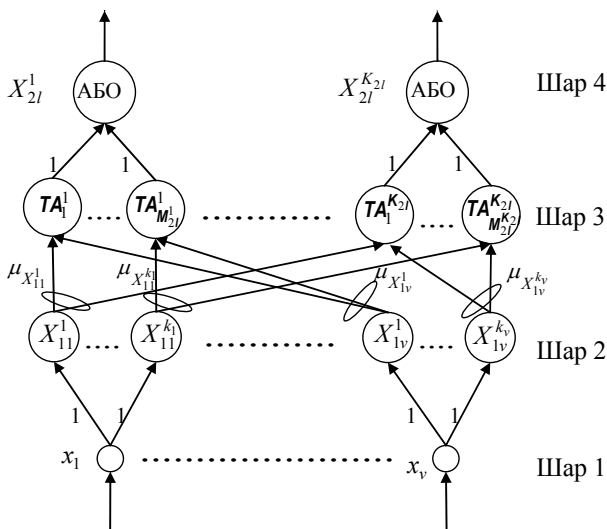


Рис. 2 – Нечітка нейронна мережа, що відповідає частковій БНЗ.

Структура мережі визначається наступним чином:

Шар 1 – вхідні нейрони, що відповідають входам об'єкту;

Шар 2 – нейрони першого схованого шару, що відповідають термам, які використовуються в частковій БНЗ, та належать до терм-множин ЛЗ першого рівня;

Шар 3 – нейрони другого схованого шару, які визначають мінімальне значення серед значень, що до них надійшли;

Шар 4 – нейрони вихідного шару, які визначають максимальне значення серед значень, що до них надійшли.

Вага зв'язків між нейронами рівнів задається: одиницею – вага між нейронами 1-го та 2-го шарів;

функціями належності входу до нечіткого терму – вага між нейронами 2-го і 3-го шару;

одиницею – вага між нейронами 3-го та 4-го шару.

Будемо розглядати обчислення результуючої змінної за даними виміру вхідних змінних як послідовну процедуру в багатшарових нейромережах, де вхідні сигнали поширюються в прямому напрямку (feedforward), але якщо дійсні значення виходів відрізняються від бажаних, то помилка поширюється у зворотному напрямку з урахуванням величин, розрахованих під час прямого ходу.

В якості функцій належності термів ЛЗ, що відповідають нейронам 2-го шару будуть використовуватися функції належності трикутного виду (2), які є універсальним апроксиматором [12]:

$$\mu: \mathfrak{R} \rightarrow [0, 1]$$

$$\mu(x) = \begin{cases} \frac{x-a}{b-a}, & x \in [a, b), \\ \frac{c-x}{c-b}, & x \in [b, c], \\ 0, & x \notin [a, c]. \end{cases} \quad (2)$$

Кожен нейрон 3-го шару TA_j^z ($j = \overline{1, M_{21}^z}$, $z = \overline{1, K_{21}}$) та пов'язані з ним нейрони другого шару можна розглядати як рядок кон'юнкції часткової нечіткої БЗ для визначення залежності результуючої ЛЗ X_{21} від вхідних ЛЗ, що визначає z-й терм цієї змінної, та переписати у вигляді правил "ЯКЩО.., ТО.." наступним чином (3)

$$\begin{aligned} & \text{ЯКЩО} \left(X_{1q_1} = X_{1q_1}^{k_{1q_1}^j} \right) \text{ТА} \left(X_{1q_2} = X_{1q_2}^{k_{1q_2}^j} \right) \dots \\ & \text{ТА} \left(X_{1q_s} = X_{1q_s}^{k_{1q_s}^j} \right) \text{ТО } X_{21} = X_{21}^z \end{aligned} \quad (3)$$

де q_s – номер вхідної ЛЗ, $s = \overline{1, S}$, $S = S_{2,1j}^z$ – номер ЛЗ у j -му правилі ($j = \overline{1, M_{21}^z}$) для визначення z -го терму ($z = \overline{1, K_{21}}$) 1-тої результуючої ЛЗ.

Третій крок є кінцевим в етапі формування структури БНЗ. Для отримання результуючої структури використовуються методи генетичного програмування (ГП), що входять до складу визначеної раніше більш широкої групи методів еволюційного моделювання.

Під час ГП кожна з частин побудованої нейронної мережі буде розглядатися як елемент початкового набору даних.

ГП оперує множиною правил для визначення терму ЛЗ. Ці правила подаються у вигляді структур даних – хромосом. Набір хромосом називається популяція. Тоді кожен елемент популяції представляє нечітку модель для визначення терму та може бути поданий у вигляді графу “ТА/АБО”.

Хромосоми оцінюються за допомогою функції відповідності на здатність їх розв’язувати проблему, тобто вірно класифікувати. Ця функція включає всі умови, які висуваються до розв’язку проблеми, тобто до БНЗ терму. Результат оцінювання, тобто ті хромосоми що залишилися, використовуються в процесі формування нової множини можливих розв’язків, нової популяції. Вибір відповідних хромосом називається селекцією. Під час неї вибираються хромосоми із найвищою функцією відповідності.

Для формування нових хромосом використовуються операції схрещування та мутації, які застосовуються до хромосом однієї популяції. Ці операції модифікують структуру нечіткої моделі.

В даній роботі пропонується використовувати два види операцій схрещування та мутації, а саме на рівні вузла АБО та рівні вузла ТА.

Схрещування забезпечує обмін інформацією між хромосомами та застосовується до випадково обраної пари хромосом з поточної популяції з ймовірністю p_c .

Операція схрещування, що здійснюється на рівні вузла АБО призводить до обміну одним або декількома вузлами ТА (рис. 3). Під час неї

обираються дві випадкові точки в хромосомах-предках А і Б. Відсічені частини з обох предків об’єднуються та створюється нащадок.

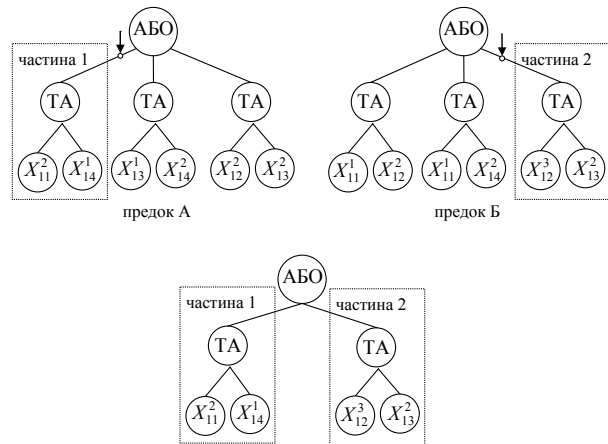


Рис. 3 – Операція схрещування на рівні входів вузла АБО.

Друга операція схрещування здійснюється на рівні входів вузла ТА. В цьому випадку процес обміну обмежується лише входами вузла ТА (рис. 4), тобто множина входів з випадково обраного вузла ТА заміщається множиною входів з випадково обраного вузла ТА іншого елемента популяції.

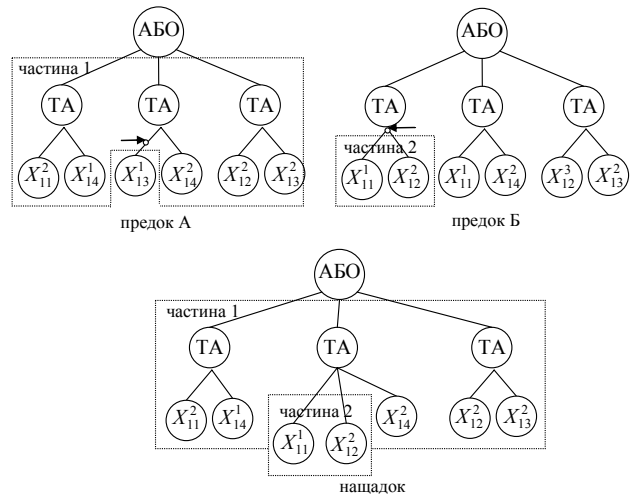


Рис. 4 – Операція схрещування на рівні входів вузла ТА.

Використовуючи два механізми схрещування вносяться зміни різного рівня в процес створення нащадків. Слід відмітити, що по своїй природі ГП в кожному новому випадку створює різні нащадки для одних й тих самих предків. Це може бути корисним в процесі пошуку оптимального рішення.

Мутація, з іншого боку, гарантує урізноманітнення, яке необхідно для еволюціонування. Вона застосовується з

ймовірністю p_m . В процесі мутації будується нова хромосома внесенням невеликих змін в структуру хромосоми-предка. Це все призводить до створення нової популяції та процес повторюється знову.

В даній роботі розглядається два типи мутації. В першому випадку мутація відбувається на рівні вузла АБО (рис. 5), в другому, на рівні вузла ТА (рис. 6).

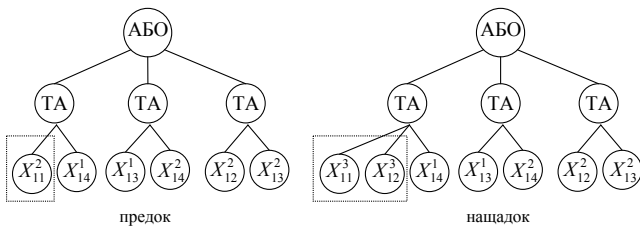


Рис. 5 – Мутація на рівні вузла ТА.

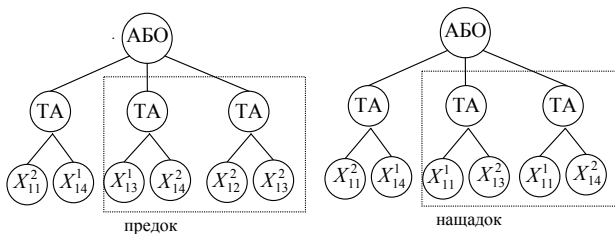


Рис. 6 – Мутація на рівні вузла АБО.

В кінці кожної епохи проходить селекція хромосом, яка полягає в оцінюванні за допомогою функції відповідності Fit_Fun (2) хромосом на здатність їх розв'язувати проблему, тобто вірно класифікувати, та формуванні нової популяції.

$$Fit_Fun = \frac{\sum_{p \in L} e_p}{|L|},$$

$$e_p = \begin{cases} 1, & (\mu_{X_{z1}^z}(p) \geq \alpha \text{ та } t_z = 1) \\ & \text{або } (\mu_{X_{z1}^z}(p) < \alpha \text{ та } t_z = 0) \\ 0, & (\mu_{X_{z1}^z}(p) \geq \alpha \text{ та } t_z = 0) \\ & \text{або } (\mu_{X_{z1}^z}(p) < \alpha \text{ та } t_z = 1) \end{cases} \quad (2)$$

де $|L|$ – потужність множини L , α – заданий поріг, при досягненні якого можна говорити, що даний вектор належить класу.

Всі операції повторюються до тих пір, поки не буде досягнута умова закінчення (3).

$$1 - Fit_Fun < \varepsilon \quad (3)$$

де ε – поріг, при досягненні якого можна говорити, що отримана часткова БНЗ задовольняє вимогам.

Важливим кроком в ГП є формування популяції. Всі хромосоми, що відповідають нечітким моделям генеруються випадково з обмеженням на те, що вузол ТА не може поєднувати два терми однієї лінгвістичної змінної. Таке саме обмеження накладається на сформовані за допомогою операції мутації нащадки. В якості генів, тобто частин з яких формуються хромосоми початкової популяції та які використовуються під час операції мутації обираються правила та їх окремі частини з бази знань сформованої за допомогою розглянутих вище алгоритмів.

Після проведення формування структури проводиться настроювання параметрів ФН. Для цього, сформовані для визначення кожного терму графі ТА/АБО об'єднуються таким чином, щоб отримати нечітку нейронну мережу (рис.2).

Після цього проводиться етап настроювання параметрів функції належності, який полягає в використанні алгоритмів навчання, що базуються на алгоритмах навчання нечіткої нейронної мережі NEFCLASS [11], де в якості критерію закінчення обирається відсоток не вірно класифікованих зразків.

Отримана кінцева нечітка нейронна мережа записується у вигляді набору нечітких правил: “ЯКЩО .. ТА .. ТА .. АБО .. ТА .. ТА .. АБО .. ТА .. ТА .., ТО ...”, які й складають часткові БНЗ.

Результуюча БНЗ буде отримана як результат об'єднання часткових БНЗ.

3. ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

Запропонований підхід було застосовано для формування БНЗ інтелектуальної системи управління державтоінспекції (ДАІ) для оцінки різних аспектів щодо стану аварійності та злочинності на дорогах України.

Даним, що зберігаються, було поставлено у відповідність такі лінгвістичні змінні:

X_{11} – кількість ДТП; X_{12} – кількість загиблих; X_{13} – кількість поранених; X_{14} – кількість ДТП

з вини водіїв у нетверезому стані; X_{15} – всього виявлено порушень правил дорожнього руху, допущених водіями; X_{16} – керування у нетверезому стані; X_{17} – перевищення швидкості; X_{18} – правил обгону; X_{19} –

створення аварійної ситуації; $X_{1,10}$ – правил переїзду перехресть; $X_{1,11}$ – зареєстровано викрадень транспортних засобів; $X_{1,12}$ – з числа зареєстрованих в поточному році розкрито; $X_{1,13}$ – всього розшукано транспортних засобів; $X_{1,14}$ – невиконання пішоходами правил дорожнього руху.

В БНЗ було виділено лінгвістичні змінні, що є результуючими, а саме X_{21} – ситуація по ДТП, X_{22} – ситуація по порушенням правил, X_{23} – ситуація, щодо викрадень транспортних засобів.

Для оцінки значень вхідних лінгвістичних змінних будемо використовувати єдину шкалу якісних термів: Н – низький, С – середній, В – високий. Слід відмітити, що кожен з цих термів представляє нечітку множину, задану за допомогою функції належності, яка в загальному випадку різна для однакових термів різних лінгвістичних змінних.

Для оцінки значень результуючих лінгвістичних змінних, тобто стан ситуації по різних напрямках використовують єдину шкалу якісних термів: Д – добрий, вС – вище середнього, С – середній, нС – нижче середнього, П – поганий.

Необхідно також підкреслити, що діапазон значень однакових вхідних змінних буде різнитися для різних регіонів. З метою використання однакових моделей було прийняте рішення проводити попередню обробку цих змінних, яка полягає в множенні на відповідний для кожної змінної та кожного регіону коефіцієнт, такий що в результаті значення вхідної змінної лежить в діапазоні від 0 до 1.

Були запропоновані наступні початкові функції належності для нечітких множин, що відповідають термам лінгвістичних змінних першого рівня (4).

В якості вихідних зразків, на яких проводилась побудова, настроювання і тестування була взята інформація щодо аварійності та злочинності по дорогах України для трьох регіональних центрів за один рік. Оскільки в базі міститься інформація по тижням, то набір містив 156 зразків. Для навчання було обрано 117 зразків, а для тестування 39.

$$\mu_H(x) = \begin{cases} 1 - 2 \cdot x, & x \in [0; 0,5], \\ 0, & x \notin [0; 0,5]. \end{cases} \quad (4)$$

$$\mu_C(x) = \begin{cases} 2 \cdot x, & x \in [0; 0,5), \\ 2 - 2 \cdot x, & x \in [0,5; 1], \\ 0, & x \notin [0; 1]. \end{cases}$$

$$\mu_B(x) = \begin{cases} 2 \cdot x - 1, & x \in [0,5; 1], \\ 0, & x \notin [0,5; 1]. \end{cases}$$

Розглянемо формування та настроювання бази знань для визначення лінгвістичної змінної другого рівня на прикладі БНЗ, яка визначає значення результуючої лінгвістичної змінної X_{21} “Ситуація по ДТП”.

Оскільки вигляд початкових функцій належності був заданий за формулою (4), а поріг $\theta = 0,7$, то на першому кроці не було внесено ніяких змін в множину ЛЗ та їх термів.

На другому кроці було сформовано 30 правил. В результаті тестування на навчальній вибірці було вірно класифіковано 100 %, а при використанні вибірки для тестування вірно класифіковано 95% (37 зразків).

Додаткове навчання за допомогою алгоритмів побудови мережі NEFCLASS, що проводилось на всій вибірці, привело до формування ще одного правила. Після цього тестування на всій вибірці дало результат вірної класифікації для всіх 156 зразків.

Далі, за допомогою методів та алгоритмів ГП, проводилось формування відповідних частин БНЗ для визначення кожного терму ЛЗ, яке розглянуто на прикладі результуючої ЛЗ “Ситуація по ДТП”.

Розмір початкової популяції дорівнював 20 хромосом для кожного терму та складався з правил, що були сформовані на попередньому кроці та їх складових частин. Було проведено 200 ітерацій (epoch). Поріг α дорівнював 0,5, а $\varepsilon = 0,01$.

Нижче розглянута побудова її частини для терму “Погано”.

Виходячи з сформованої часткової БНЗ, яка складалася з 5 правил з сумарною кількістю термів 20 був проведений етап формування структури за допомогою методів ГП. Розмір початкової популяції складав 20 хромосом, кількість ітерацій була задана в розмірі 200. В результаті була отримана база знань, що складається з 3 правил з сумарною кількістю термів 4.

Після проведення формування структури було проведено додаткове тестування на всій вибірці (156 зразків), яке дало 100% вірних результатів.

Відзначимо, що після застосування методів генетичного програмування БНЗ суттєва спростилася лише для випадку коли лінгвістична

змінна X_{21} приймає значення П (погано).

Після проведення етапу формування структури за допомогою ГП загальна кількість правил в БНЗ порівняно з початковою зменшилась в 1,4 рази, а сумарна кількість термів в 2,3 рази.

Етап настроювання параметрів для даного випадку не проводився, оскільки БНЗ давала вірне рішення для всієї навчальної вибірки.

4. ВИСНОВКИ

Запропонований в роботі підхід дає можливість істотно зменшити суб'єктивність і суперечливість експертних оцінок при формуванні бази нечітких знань. Внаслідок цього поліпшується якість та прозорість рішення задач обробки інформації в складних адміністративних системах.

Такий підхід також можна застосовувати як для вирішення задач оцінки стану в різноманітних організаційних структурах, так і для обробки інформації в суцільно технічних системах, таких як інформаційно-телекомунікаційні мережі [13, 14], ядерна енергетика тощо.

5. СПИСОК ЛІТЕРАТУРИ

- [1] Гольшев Л.К. *Сложные системы с развитой функцией информационно-аналитической поддержки управления. Элементы теории, методологии, практики.* – К., 2001. – 253 с.
- [2] Згуровский М.З., Панкратова Н.Д. *Системный анализ. Проблемы, методология, приложение.* – К: Наук. думка, 2005. – 744 с.
- [3] Ротштейн А.П. *Интеллектуальные технологии идентификации: нечеткие множества, генетические алгоритмы, нейронные сети.* – Винница: УНИВЕРСУМ, 1999. – 320 с.
- [4] Павлов А.А., Теленик С.Ф. *Информационные технологии и алгоритмизация в управлении.* – К.: Техніка, 2002. – 344 с.
- [5] Люгер Дж. Ф. *Искусственный интеллект: стратегии и методы решения сложных проблем, 4-е издание.* – М.: "Вильямс", 2003. – 864 с.
- [6] Корнеев В.В., Гарев А.Ф., Васютин С.В., Райх В.В. *Базы данных. Интеллектуальная обработка информации.* – М.: Нолидж, 2000. – 351 с.
- [7] Згуровский М.З., Доброногов А.В., Померанцева Т.Н. *Исследование социальных*

процессов на основе методологии системного анализа. – К.: Наукова думка, 1997. – 224 с.

- [8] Глоба Л.С., Терновой М.Ю. Оптимизация использования базы знаний экспертной системы введением иерархии правил. // *Труды седьмой международной научно-практической конференции "Современные информационные и электронные технологии" (СИЭТ-2006).* – Одесса, 2006. – С. 166.
- [9] Setnes M., Babuska R., Kaymak U., van Nauta Lemke H.R. Similarity measures in fuzzy rule base simplification, *IEEE Trans. SMC-B* 28, (1998), P. 376-386.
- [10] Зайченко Ю.П. *Основи проектування інтелектуальних систем. Навчальний посібник.* – К.: Видавничий Дім "Слово", 2004. – 352 с.
- [11] Nauck D., Kruse R. NEFCLASS – A Neuro-Fuzzy Approach for the Classification of Data. In K. M. George, Janice H. Carrol, Ed Deaton, Dave Oppenheim and Jim Hightower, editors, *Applied Computing 1995. Proc. 1995 ACM Symposium on Applied Computing, Nashville, Feb. 26-28, pages 461-465,* ACM Press, New York, February 1995.
- [12] Xiao-Jun Zeng; Singh, M.G Approximation theory of fuzzy systems-MIMO case. *IEEE Transactions on Fuzzy Systems.* Volume 3, Issue 2, May 1995 Page(s):219 - 235.
- [13] Ternovoy M. Intellectual Control Systems Fuzzy Knowledgebase-Building Technique // *Proceedings of the International Conference TCSET'2008.* – Lviv-Slavsko. – pp. 97-98.
- [14] Глоба Л.С., Терновой М.Ю., Штогріна О.С. *Технологія обробки інформації в гетерогенному інформаційно-телекомунікаційному середовищі. Електроніка і зв'язь. Тематический випуск "Проблеми електроніки", ч. 1.* – 2008. – С. 61-65.



Глоба Лариса Сергіївна. В 1978 році закінчила Харківський авіаційний інститут за спеціальністю "інженер-математик". В 1984 році захистила кандидатську дисертацію, а в 1996 році докторську дисертацію. Зараз обіймає посаду завідувача кафедри Інформаційно-телекомунікаційних мереж Інституту телекомунікаційних систем Національного технічного університету України "Київський політехнічний інститут". Має більше 140

наукових робіт.

Наукові інтереси: розподілені інформаційні систем, інтелектуальні систем підтримки прийняття рішень, системи виробництва товарної продукції.



Терновой Максим Юрійович. В 2002 році закінчив Національний технічний університет України "Київський політехнічний інститут" та отримав кваліфікацію магістра прикладної математики за спеціальністю інформатика. В 2007 році захистив кандидатську

дисертацію за спеціальністю "Автоматизовані системи управління і прогресивні інформаційні технології". Зараз обіймає посаду доцента кафедри Інформаційно-телекомунікаційних мереж Інституту телекомунікаційних систем Національного технічного університету України "Київський політехнічний інститут". Має близько 30 наукових робіт.

Наукові інтереси: інтелектуальні системи підтримки прийняття рішень, mobile grid, pervasive computing.



Штогріна Олена Сергіївна. В 2007 році закінчила Національний технічний університет України "Київський політехнічний інститут" та отримала кваліфікацію інженер математик за спеціальністю "Прикладна математика". Зараз працює асистентом кафедри Інформаційно-телекомунікаційних мереж

Інституту телекомунікаційних систем Національного технічного Університету України "Київський політехнічний інститут".

Наукові інтереси: mobile grid, distributed systems, pervasive computing.



FUZZY KNOWLEDGEBASE DESIGN FOR INTELLECTUAL SYSTEMS

Larisa S. Globa ¹⁾, Maksym Y. Ternovoy ²⁾, Olena S. Shtogryna ³⁾

¹⁾ National Technical University of Ukraine "Kiev Polytechnic Institute",
 Industrialnyy Al., 2, Kyiv, 03056, Ukraine, e-mail: gls@densoft.com.ua

²⁾ National Technical University of Ukraine "Kiev Polytechnic Institute",
 Industrialnyy Al., 2, Kyiv, 03056, Ukraine, e-mail: maximter@mail.ru

³⁾ National Technical University of Ukraine "Kiev Polytechnic Institute",
 Industrialnyy Al., 2, Kyiv, 03056, Ukraine, e-mail: L_Shtogryna@mail.ru

Abstract: *The approach to fuzzy knowledgebase designing for complex administrative intellectual systems is proposed in the paper. It consists of two phases: structure construction and membership function parameters adjustment. This approach allows accelerate the intellectual systems fuzzy knowledgebase construction.*

Keywords: *fuzzy knowledgebase, linguistic variable, membership function, fuzzy neural network, NEFCLASS, genetic programming, complex administrative systems.*

1. INTRODUCTION

State institutions such as different ministries and departments belong to complex administrative systems (CAS). Such systems have hierarchical structure and consist of many departments and subordinate organizations. Information-analytical activity in CAS is based on intellectual systems. The information processing is one of the most difficult part of such activity [1], because information has uncertainty. Consequently it's necessary to use such mathematical methods and algorithms which can represent and operate the data with different types of uncertainty in description of object. There are three big groups of such methods [2, 3, 4]: logical methods, network methods and evolutionary modeling. But only two first groups of these methods can be used for the model for information processing designing, because evolutionary modeling methods have no own knowledge model. But evolutionary modeling methods can be used as a part of knowledgebase designing process for the model structure optimization [5, 6].

The approach for fuzzy knowledgebase designing is proposed in the paper.

2. MAIN PART

Fuzzy knowledgebase is the one of the main part of intellectual systems. It is represented as union of partial fuzzy knowledgebases, each of which defines the dependence between the output linguistic variable (LV) and input linguistic variables [8].

The partial fuzzy knowledgebase is the set of rules of type: "IF .. AND .. AND .. ALSO .. AND .. AND .. ALSO .. AND .. AND .., THEN ..". Each of these rules defines dependence between output linguistic variable term and input variables terms.

In general case, the several output variables are determined by fuzzy knowledgebase. Each of these variables identifies the different aspects of the knowledge domain.

The problem definition of fuzzy knowledgebase construction and adjustment is described below.

Given:

- 1) $\{x_i | i = \overline{1, v}\}$ – the set of input variables;
- 2) $\{X_{li} | i = \overline{1, v}\}$ – the set of LVs, which corresponds to $\{x_i | i = \overline{1, v}\}$;
- 3) $\{X_{li}^z | i = \overline{1, v}, z = \overline{1, Z_{li}}\}$ – the set of terms of LVs from point 2;
- 4) $\{\mu_{X_{li}^z} | i = \overline{1, v}, z = \overline{1, Z_{li}}\}$ – the set of membership functions, which corresponds to LV's terms from point 3. These functions are used in fuzzyfication process;
- 5) $\{X_{2i} | i = \overline{1, M_2}\}$ – the set of output LVs;
- 6) $\{X_{2i}^z | z = \overline{1, Z_{2i}}\}$ – the set of terms of LVs from point 5;
- 7) $\{(x^1, X_2^1) | i = \overline{1, R_{Study}}\}$ – data samples for construction and adjustment,
 where $x^1 = \langle x_1^1, x_2^1, \dots, x_v^1 \rangle$ – 1-th sample of

input variables values,

$$X_2^1 = \langle X_{21}^1, X_{22}^1, \dots, X_{2N}^1 \rangle - 1\text{-th sample of}$$

output LVs values;

$$8) \{(x^1, X_2^1) | i = \overline{1, R_{\text{Test}}}\} - \text{samples for testing}$$

Find out:

$$1) \widehat{R}_2 = \bigcup_1 \widehat{R}_{21} - \text{fuzzy knowledgebase, which}$$

defines the dependence between output LVs and input LVs;

2) To adjust the membership functions parameters $\{\mu_{x_{ii}^z} | i = \overline{1, v}, z = \overline{1, Z_{ii}}\}$.

The structure of approach for solving fuzzy knowledgebase construction and adjustment problem is proposed on fig.1. It consists of two phases: structure construction and membership function's parameters adjustment phases.

Structure construction phase are made separately for each output linguistic variable term because this phase is independent for each term. The partial fuzzy knowledgebase could be obtained as union of all rules for each linguistic variable term. The whole fuzzy knowledgebase will be the result of all partial fuzzy knowledgebases union.

Structure construction phase is made in three steps.

The first step is the similarity simplification of input linguistic variables sets of terms [9]. It is made by using of similarity measure Eq. 1

Fuzzy sets are united when the measure exceed preset threshold $\theta \in [0, 1]$. The union decrease amount of different fuzzy sets which are correspond to linguistic variables terms and used in knowledgebase construction. Such approach increases the transparency during classification task solving.

The second step is the construction of fuzzy knowledgebase initial structure. It is made by using of fuzzy neural network NEFCLASS building based algorithms [10, 11]. Fuzzy knowledgebase is presented as four-layer fuzzy neural network (fig. 2).

Each of the part of build neural network will consider as an element of initial set of chromosome during genetic programming methods using.

The third step is the result structure construction. The genetic programming methods are used for the result structure obtaining. These methods are the part of evolutionary modeling methods.

There are two operation used: crossover (fig. 3, fig. 4) and mutation (fig. 5, fig. 6), each of which also divides into two types: OR-level and AND-level.

During each successive generation, the proportion of the existing population is selected to

breed a new generation. Individual solutions are selected by using fitness function Fit_Fun Eq. 2.

This generational process is repeated until a termination condition has been reached Eq.3.

After that the membership function's parameters adjustment phase is made by using of fuzzy neural networks NEFCLASS learning algorithms [11].

The obtained fuzzy neural network is written down in the form of fuzzy rules: "IF .. AND .. AND .. ALSO .. AND .. AND .. ALSO .. AND .. AND ..., THEN ..", which are the partial fuzzy knowledgebases.

The result fuzzy knowledgebase is obtained as union of partition fuzzy knowledgebases.

3. EXPERIMENTAL RESULTS

Proposed approach was used for Ukrainian State Motor Vehicle Inspectorate intellectual information system fuzzy knowledgebase designing, which is used for estimation of accident rate and criminality.

The information about accident rate and criminality from three cities during the year was used as data samples.

The information periodicity is one week, so the number of all samples was 156. There were chosen 117 samples for creation and adjustment, and 39 samples for testing. Initial membership functions were determined by Eq. 4.

The construction of the fuzzy knowledgebase part to determine the linguistic variable "Situation on traffic accident" is described in the following.

At the first step the threshold θ was set 0.7. Because of this there were no changes in the set linguistic variables terms.

There were constructed 31 rules at the second step. After that genetic programming methods were used to construct the partial fuzzy knowledgebases. The using of these methods allowed to decrease initial number of rules, which were obtained after second step, in 1.4 times and general number of terms in 2.3 times.

4. CONCLUSIONS

The proposed approach can decrease subjectivity and inconsistency of expert evaluations during fuzzy knowledgebase designing. That is why the data processing quality and transparency in complex administrative systems increase.

Such approach can also be used for solving information estimation tasks in different organizations and in technical systems like information-telecommunication networks [13, 14], nuclear power engineering and so on.