Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

АПАНАСЕНКО ДМИТРО ВОЛОДИМИРОВИЧ

УДК 519.874:519.852:519.816

МЕТОДИ КЛАСТЕРИЗАЦІЇ НЕЧІТКИХ ДАНИХ СПЕЦІАЛЬНОГО ВИГЛЯДУ ТА ЇХ ЗАСТОСУВАННЯ

01.05.04 – системний аналіз і теорія оптимальних рішень

Автореферат
дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ – 2019

Дисертацією є рукопис

Робота виконана на кафедрі системного аналізу та теорії прийняття рішень факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка МОН України.

Науковий керівник: доктор фізико-математичнич наук, професор
Івохін Євген Вікторович,
Київський національний університет імені Тараса Шевченка,
професор кафедри системного аналізу та теорії прийняття рішень.

Офіційні опоненти: доктор технічних наук, професор
Андрущак Ігор Євгенович,
Луцький національний технічний університет МОН України,
завідувач кафедри комп'ютерних технологій;

кандидат технічних наук, доцент
Олецький Олексій Віталійович,
Національний університет «Києво-Могилянська академія» МОН України, доцент кафедри мультимедійних систем.

Захист відбудеться "24" червня 2019 року о 15 год. 30 хв. на засіданні спеціалізованої вченої ради Д 26.001.35 Київського національного університету імені Тараса Шевченка за адресою: м.Київ, проспект Академіка Глушкова, 4-Д, факультет комп'ютерних наук та кібернетики, ауд.01.

З дисертацією можна ознайомитися у Науковій бібліотеці імені М.Максимовича Київського національного університету імені Тараса Шевченка за адресою: 01601, м.Київ, вул.Володимирська, 58, зал №12.

Автореферат розіслано "___" травня 2019 р.

Вчений секретар
спеціалізованої вченої ради Д 26.001.35 Зінько П.М.

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. Інформаційні технології впевнено інтегрувалися в життя сучасної людини. На сьогоднішній день вони надають можливість для здійснення online-покупок, геопозиціонування, розробки власних додатків, спілкування. Особливе місце у сфері активного впровадження інформаційних технологій займають явища, що відбуваються у соціальному середовищі.

Традиційна задача групування даних у відносно однорідні групи у цьому випадку не тільки є актуальною в сучасних умовах, а й постійно отримує новий розвиток в силу необхідності оперативного опрацювання величезних потоків інформації та вирішення проблеми побудови інформаційних співтовариств. Задача, яка виникає в самих різних сферах соціуму включає в себе проблеми виділення кластерів і знаходження еталонних вузлів в сукупності даних. При цьому, в реалізації методів кластеризації важливу роль мають характеристики належності даних до групи, способи обчислення відстаней між елементами, вагові коефіцієнти, що залежать від конкретної предметної області і спостережень, отриманих в рамках цієї області. Дослідженню проблем кластеризації та виділення однорідних за різною сукупністю параметрів груп присвячено роботи И.Д.Манделя, Д.А.Губанова, Ю.М. Плотинського, Д.А.Новікова, B. Durand'a, P.Odell'a, В.Е.Котова, Л.А.Гладкова, W.T.Williams'a, G.N.Lance, M.Jamba, А.Г.Чхартішвілі, К.В.Воронцова, та інших. При цьому потрібно зауважити, що важливою особливістю соціальної інформації про події та явища, яка характерно проявляється у сучасних інформаційних мережах, є постійна модифікація складу даних.

Необхідно також зауважити на наявність у сукупностях явищ та груп соціального середовища широкої неоднорідності та суттєвої внутрішньої розшарованості. Це зумовлено нерівномірністю розвитку окремих одиниць сукупності і своєрідністю умов, у яких вони функціонують (природних, технологічних тощо). Одні сукупності поділяються на чітко визначені, ізольовані класи (групи, типи), іншим властива латентна, прихована структура. Серйозною проблемою виступає також обсяг інформації, що потребує обробки, швидкість пошуку та аналізу даних в умовах надзвичайно великих інформаційних потоків.

Тому зрозумілою та дуже важливою стає задача створення ефективних технологій, що дозволяють проводити комплексний аналіз різнорідних за обсягом, складом і якістю даних, які генеруються соціальним середовищем.

Практичне використання математичних і програмних засобів для дослідження соціальних процесів вимагає створення прикладних автоматизованих інформаційних (прогнозних) систем та систем підтримки прийняття рішень, що є особливо важливим для проведення системного аналізу умов виникнення та розвитку подій соціального середовища. Існуючі системи мають, як правило, вузькі спеціалізовані варіанти застосування, що не завжди підходить для роботи з недостатньо формалізованими ситуаціями, серед яких особливої уваги заслуговують проблеми аналітичної обробки інформаційних потоків. Тому важливою прикладною задачею залишається розробка і впровадження програмних систем аналітичної обробки інформації та систем підтримки прийняття рішень на основі адаптації відомих математичних методів системного аналізу з урахуванням умов інформаційної невизначеності та нечіткості.

Аналіз задач кластеризації та дослідження процесів структурних змін в соціальних групах і в спільнотах знайшли своє відображення у роботах F.Herrera, M.Lozano, D.Whitley, P.Клемента, S. Openshaw, Т.В.Панченка, В.М.Курейчика, Ю.Р.Цоя, В.Г.Редька та інших. Результати дослідження проблем кластеризації з урахуванням невизначеності та неточності вихідних даних наведені у роботах A.Jain'a, M. Murty, P. Flynn'a, J. Kogan'a, C. Nicholas'a, M. Teboulle, Д.А. Вятченина та ін.

Актуальність і перспективність задач сукупного аналізу динаміки структур соціальних спільнот, необхідність вирішення задач групування даних, встановлення та дослідження зв'язків на основі графів і текстових даних, що отримуються з соціальних мереж, з метою розв'язування сучасних проблем формалізації та дослідження потоків інформаційного розповсюдження та впливу визначили вибір теми, мети і задач дисертаційної роботи.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота є складовою частиною наукових робіт, які ведуться на кафедрі системного аналізу та теорії прийняття рішень і в науково-дослідному секторі "Проблем системного аналізу" Київського національного університету імені Тараса Шевченка. Дослідження виконувалися в рамках науково-дослідної теми №16БФ015-02 "Розробка нових математичних методів системного аналізу і теорії оптимальних рішень та їх застосування" (державний номер реєстрації 0116U002529, термін виконання 2016–2018г.г., в рамках програми "Інформатизація суспільства") і науково-дослідної теми «Розробка і впровадження інформаційної та організаційної системи заходів по забезпеченню інноваційної спрямованості науково-дослідних робіт в Київському національному університеті імені Тараса Шевченка», НДР № 08БП 013-01 (за напрямом підпрограми "Інформаційні технології в науці та навчальному процесі").

Мета і задачі дослідження. Метою роботи є розробка та практичне впровадження методів і алгоритмів розв'язування задач кластеризації даних, що подаються сукупністю нечітких трикутних чисел. Розв'язуються задачі створення та застосування методів дослідження динаміки в процесах кластеризації, розробки математичного забезпечення для обробки нечіткої інформації спеціального вигляду, а також розв'язання на основі розроблених методик задачі оптимального розподілу потужностей каналів передачі даних у компьютерній мережі. Результати дослідження можуть бути використані для розв'язування різноманітних прикладних задач аналізу інформаційних процесів із застосуванням систем підтримки прийняття рішення.

Об'єктом дослідження є процеси аналізу та групування інформації, способи формалізації невизначеності та алгоритми обробки нечітких даних спеціального вигляду.

Предмет дослідження – методи формалізації та дослідження динаміки у процесах кластеризації даних, математичні методи та програмні засоби аналізу нечіткої інформації, їх використання для оптимального розподілу потужностей каналів в задачах передачі даних.

В основу *методів дослідження* покладено системний підхід до дослідження процесів формалізації та аналізу динаміки складних процесів і систем. При розв'язанні поставлених задач були використані методи системного аналізу, методи дослідження теорії нечітких множин, методи дослідження динамічних систем, теорії оптимізації та теорії прийняття рішень.

Наукова новизна одержаних результатів. У процесі розв'язання поставлених задач отримано нові наукові результати, які полягають у такому:

- запропоновано новий підхід для формалізації невизначеності у вигляді складених нечітких чисел та побудови множин скалярного та векторного рівня;

- для аналізу процесів та проведення групування нечіткої інформації запропоновано новий підхід, що базується на застосуванні множин рівня спеціального вигляду та оригінальній методиці обчислення відстані;
- вперше запропоновано методи кластеризації даних, поданих у вигляді сукупності складених нечітких чисел;
- набула подальшого вдосконалення методика застосування генетичного алгоритму для роз'язання задачі кластеризації;
- сформульовано нову оптимізаційну задачу математичного програмування для розрахунку раціональної кількості груп маршрутизаторів для забезпечення оптимальної потужності каналів передачі даних;
- розроблено та впроваджено новий метод розв'язання задачі розрахунку оптимальної потужності каналів передачі даних у комп'ютерній мережі.

Практичне значення одержаних результатів полягає в розробці нового конструктивного способу формалізації невизначеності, у дослідженні та розв'язанні ряду практичних задач з нечіткими вхідними даними. Ці результати можуть бути використані при створенні ефективних програмних систем для аналізу та реалізації процесів групування інформації, для підтримки прийняття рішень при дослідженні систем, що функціонують в умовах невизначеності.

Розроблені математичні методи та підходи використані для розрахунку оптимальної кількості маршрутизаторів для забезпечення передачі даних у комп'ютерній мережі Інформаційно-обчислювального центру Київського національного університету імені Тараса Шевченка. Результати дисертаційного дослідження використовуються у навчальному процесі на факультеті комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка при викладенні спеціальних курсів «Невизначеність та групування інформації» та «Методи дослідження динаміки нечітких систем» (спеціальність «системи і методи прийняття рішень»).

Особистий внесок здобувача. Дисертація є самостійною науковою працею, в якій висвітлені власні ідеї та розробки автора, що дозволили досягти поставленої мети. Наукові положення, пропозиції та рекомендації, що виносяться на захист, отримані здобувачем самостійно. У спільних роботах із науковим керівником Є.В.Івохіним, вибір методів дослідження та доведення основних результатів виконано автором, особистий внесок якого полягає у наступному: у роботі [2] формалізовано процеси кластеризації складених нечітких даних на основі множин скалярного рівня; у роботі [3] адаптовано математичні методи групування складених нечітких даних на основі множин векторного рівня; у роботі [4] запропоновано методику розв'язання трьохіндексної транспортної задачі шляхом зведення її до відповідної дворівневої задачі оптимізації; у роботі [5] запропоновано узагальнення методики кластеризації нечітких даних на основі множин векторного рівня, наведено числові результати розрахунків; у роботі [6] викладено результати практичного застосування методів групування для розв'язування задачі розподілу потужностей каналів передачі даних в комп'ютерній мережі.

Апробація результатів дисертації. Матеріали дисертаційної роботи доповідалися та обговорювалися на наукових конференціях та семінарах: VII міжнародній науково-практичній конференції «Сучасні проблеми математичного моделювання, прогнозування та оптимізації» (Кам'янець-Подільський, квітень, 2016); міжнародній школі-семінарі «Теорії прийняття рішень» (Ужгород, вересень, 2016); Міжнародній науковій конференції "Problems of Decision Making under

Uncertainties” (Вільнюс, Литва, серпень, 2017; Прага, Чехія, серпень, 2018); Міжнародній науковій конференції ім. Т.А. Таран ”Интеллектуальний анализ информации” (IAI) (Київ, травень, 2017); Міжнародній науково-практичній конференції «Обчислювальний інтелект» (Київ, травень, 2017); на наукових семінарах факультету комп’ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка.

Публікації. Основні наукові твердження, висновки і результати дисертаційної роботи опубліковані в 11 наукових працях. З них – 6 наукових статей, у тому числі 5 у фахових виданнях [1–5], 1 стаття у виданні, яке входить до наукометричної бази даних SCOPUS [5], 5 – праці та тези наукових конференцій [7–11].

Структура дисертації. Дисертаційна робота складається з вступу, п’яти розділів, висновків, списку використаних джерел з 85 найменувань (на 8 сторінках) і п’яти додатків (на 11 сторінках). Загальний обсяг роботи становить 147 сторінок, з них 133 сторінки – основна частина, що містить 7 рисунків та 5 таблиць.

ОСНОВНИЙ ЗМІСТ РОБОТИ

У **вступі** до дисертаційної роботи обґрунтовано актуальність теми дослідження, вказано на зв’язок роботи з науковими програмами, темами, планами, сформульовано мету та задачі дослідження, охарактеризовано використані в роботі методи, вказано наукову новизну одержаних результатів та їх практичне значення, а також наведено дані про апробацію отриманих в роботі результатів та особистий внесок здобувача, наведено список публікацій.

Перший розділ дисертаційної роботи присвячений огляду літератури з наукових досліджень на основі нечітких множин та з проблем кластеризації даних. В розділі наведено аксіоматику нечітких множин, основні поняття і визначення нечітких величин і нечітких чисел, описано узагальнення поняття нечітких чисел у вигляді складених нечітких чисел. Проведено огляд використання нечітких множин та математичних моделей нечітких систем. Відмічено, що перспективним напрямком формалізації нечіткості є використання трикутних нечітких величин, основною ідеєю визначення яких є побудова інтервалів можливих значень невизначеного (невідомого) параметра з застосуванням лінійної функції належності.

Сформульовано задачі групування (кластеризації) інформації, що отримується внаслідок статистичних спостережень. Визначено вид групувань у вигляді набору кластерів, наведено проблеми групування нечітких даних.

У першому підрозділі наведено загальні підходи до формалізації нечіткості. Відмічено, що у фундаментальних та прикладних дослідженнях однією з найпоширеніших стає теорія нечітких множин Заде, яка забезпечує можливість теоретико-множинного представлення неточних понять.

Наведемо основні поняття. Нехай $\Re$ – скінченновимірний нормований простір.

Означення 1.1. Нечіткою множиною $\widetilde{A}$ універсальної множини $X \subseteq \Re$, називається сукупність пар $\widetilde{A} = \{(\mu_{\widetilde{A}}(x), x)\}$, де $\mu_{\widetilde{A}} : X \to [0,1]$ – відображення множини X в одиничний відрізок [0,1], яке називається функцією належності нечіткої множини $\widetilde{A}$.

Означення 1.2. Нечітким відображенням $\widetilde{R}$ з X у довільний скінченновимірний простір Y називається нечітка підмножина $\widetilde{R}$ в $X \times Y$ з функцією належності:

$$\mu_{\widetilde{R}}(x, y) : X \times Y \to [0,1], x \in X, y \in Y . \quad (1)$$

Розглянемо в якості універсальної множини X множину дійсних чисел R^1, тобто $X = R^1$.

Означення 1.3. Нечітким трикутним числом $\widetilde{A}$ називається впорядкована трійка чисел *(a, b, c)*, котрі визначають функцію належності $\mu_{\widetilde{A}}(x)$ вигляду:

1. $\mu_{\widetilde{A}}(x) = \dfrac{x-a}{b-a}, x \in [a,b]$;

2. $\mu_{\widetilde{A}}(x) = \dfrac{c-x}{c-b}, x \in [b,c]$; (2)

3. $\mu_{\widetilde{A}}(x) = 0, x \notin [a,c]$.

Нечітке трикутне число виду *(a, b, b)*, яке називають лівим нечітким трикутним числом, визначається функцією належності вигляду

$$\mu_{\widetilde{A}}(x) = 0, x < a\,; \quad \mu_{\widetilde{A}}(x) = \frac{x-a}{b-a}, x \in [a,b]; \; \mu_{\widetilde{A}}(x) = 1, x > b\,, \tag{3}$$

а нечітке трикутне число виду *(b, b, c)*, яке називають правим нечітким трикутним числом, – функцією належності

$$\mu_{\widetilde{A}}(x) = 1, x < b\,; \mu_{\widetilde{A}}(x) = \frac{c-x}{c-b}, x \in [b,c]; \; \mu_{\widetilde{A}}(x) = 0, x > c\,. \tag{4}$$

У четвертому підрозділі наведено аксіоматику складених нечітких трикутних чисел.

Припустимо, що задано нечіткі трикутні числа $\widetilde{b}_1,\ldots,\widetilde{b}_m$, $\widetilde{b}_i = \{(a_i, b_i, c_i)\}$, $a_i \le b_i \le c_i$, $i = \overline{1,m}$ з лінійними функціями належності $\mu_{\widetilde{b}_1}(x)$, $\ldots,\mu_{\widetilde{b}_m}(x)$ вигляду (2), що визначені на відповідних універсальних множинах $X_1,\ldots,X_m$, де $X_i \subseteq R^1, i = \overline{1,m}$. При цьому, очевидно, що для носіїв нечітких чисел справедливе включення $[a_i, c_i] \subseteq X_i$, $i = \overline{1,m}$.

З довільних елементів цієї множини можна сформувати множину

$$\widetilde{b}^m = \{(x^1, \mu_{\widetilde{b}_1}(x^1)), (x^2, \mu_{\widetilde{b}_2}(x^2)), \ldots, (x^m, \mu_{\widetilde{b}_m}(x^m))\}, \;\; x^i \in [a_i, c_i]\,, \; i = \overline{1,m}\,. \tag{5}$$

Означення 1.4. Довільну множину $\widetilde{b}^m$ вигляду (5) будемо називати складеним нечітким числом на множині нечітких чисел $\widetilde{b}_1$, …, $\widetilde{b}_m$ трикутного вигляду.

Необхідно відзначити, що складене нечітке число не є в загальному випадку нечіткою множиною, тому що для нього неможливо визначити універсальну множину, на якій воно задається. Нескладно пересвідчитися, що складене нечітке число являє собою вектор розмірності m, що складається з елементів нечітких чисел. Крім того, цей агрегат також не слід ототожнювати з нечіткою множиною $\widetilde{A}^m = \{(x, \mu_{\widetilde{A}^m}(x))\}$, $x = (x^1, x^2, \ldots, x^m) \in R^m$, яке визначає поняття нечіткого вектора, для побудови котрого використовується універсальна множина вигляду $\bigotimes_{i=1}^{m} X_i \subseteq R^m$ з єдиною функцією належності $\mu_{\widetilde{A}^m}(x)$.

Сукупність складних нечітких чисел $\widetilde{b}^m$ заданого розміру $N \ge 1$ надалі позначатимемо через $K(\widetilde{b}^m)$, $|K(\widetilde{b}^m)| = N$.

Означення 1.5. Множину

$$L_0^m = \{x^1 \in X_1, x^2 \in X_2, \ldots, x^m \in X_m : \mu_{\widetilde{b}_1}(x^1) > 0,\; \mu_{\widetilde{b}_2}(x^2) > 0, \ldots, \mu_{\widetilde{b}_m}(x^m) > 0\}\,,\; L_0^m \subseteq \mathop{\times}_{i=1}^{m} X_i\,, \tag{6}$$

назвемо носієм складеного нечіткого числа $\widetilde{b}^m$.

Означення 1.6. Множину

$$L^m(\alpha) = \{x^1 \in X_1, x^2 \in X_2, \ldots, x^m \in X_m : \mu_{\widetilde{b}_1}(x^1) \ge \alpha,\; \mu_{\widetilde{b}_2}(x^2) \ge \alpha, \ldots, \mu_{\widetilde{b}_m}(x^m) \ge \alpha\}\,,\; L^m(\alpha) \subseteq L_0^m\,, \tag{7}$$

назвемо множиною скалярного рівня $\alpha \in (0,1]$ складеного нечіткого числа $\widetilde{b}^m$.

Означення 1.7. Множину

$$L^m(\alpha_1,...,\alpha_m)=\{x^1\in X_1, x^2\in X_2,...,x^m\in X_m : \mu_{\tilde b_1}(x^1)\ge\alpha_1,\ \mu_{\tilde b_2}(x^2)\ge\alpha_2,...,\mu_{\tilde b_m}(x^m)\ge\alpha_m\}, \quad (8)$$

$L^m(\alpha_1,...,\alpha_m)\subseteq L_0^m$, назвемо множиною векторного рівня $(\alpha_1,...,\alpha_m)$, $\alpha_i\in(0,1]$, $i=\overline{1,m}$, складеного нечіткого числа $\tilde b^m$.

Означення 1.8. Складене нечітке число $\tilde b^{m|0}$ вигляду $\tilde b^{m|0}=\{(x^1,0),(x^2,0),...,(x^m,0)\}$, $\forall x^i\in X_i$, $i=\overline{1,m}$, назвемо повністю виродженим, а складене нечітке число $\tilde b^{m|j}=\{(x^1,\mu_{\tilde b_1}(x^1)),...,(x^j,0),...,(x^m,\mu_{\tilde b_m}(x^m))\}$, $\mu_{\tilde b_i}(x^i)>0$, $\forall x^i\in X_i$, $i=\overline{1,m}$, $i\ne j$ – виродженим за j-ю компонентою, $j=\overline{1,m}$.

Означення 1.9. Складене нечітке число $\tilde b_j^{m-1}$, побудоване з елементів нечітких чисел $\tilde b_1$, …, $\tilde b_{j-1}$, $\tilde b_{j+1}$, …, $\tilde b_m$, назвемо зрізом сукупності складених нечітких чисел $K(\tilde b^m)$ за j-ю компонентою, $j=\overline{1,m}$.

Припустимо, що задано два невироджених складених нечітких числа

$$\tilde u^m=\{(x^1,\mu^u_{\tilde b_1}(x^1)),(x^2,\mu^u_{\tilde b_2}(x^2)),...,(x^m,\mu^u_{\tilde b_m}(x^m))\},$$

$$\tilde v^m=\{(y^1,\mu^v_{\tilde b_1}(y^1)),(y^2,\mu^v_{\tilde b_2}(y^2)),...,(y^m,\mu^v_{\tilde b_m}(y^m))\}.$$

Обчислимо величину $\gamma=\min\limits_{i=\overline{1,m}}\min\left\{\mu^u_{\tilde b_i}(x^i),\mu^v_{\tilde b_i}(y^i)\right\}$, яка є мінімальним значенням серед величин мір належності окремих елементів обох чисел $\tilde u^m,\tilde v^m$. Це значення дозволяє побудувати дві множини скалярного рівня γ у вигляді звичайних множин $L^m_{\tilde u}(\gamma)$, $L^m_{\tilde v}(\gamma)$, що визначають точки носія L_0^m складеного нечіткого числа $\tilde b^m$, між якими може бути обчислена відстань $d=\left\|L^m_{\tilde u}(\gamma)-L^m_{\tilde v}(\gamma)\right\|$.

Означення 1.10. Нечітка величина за Вонгом

$$\tilde W(\tilde b^m)=\{(d,\gamma): d=\left\|L^m(\gamma)\right\|,\ \gamma=\min_{i=\overline{1,m}}\mu_{\tilde b_i}(x^i)\in(0,1], x^i\in X_i,\ i=\overline{1,m}\ \}, \quad (9)$$

де $\|\cdot\|$ – евклідова норма вектора в R^m, визначає метрику на множині $K(\tilde b^m)$.

Таким чином, кожне складене нечітке число $\tilde b^m$ «вимірюється» за допомогою нечіткої величини $\tilde W(\tilde b^m)$, а для визначення нечіткої відстані $\tilde\rho(\tilde u^m,\tilde v^m)$ між довільними складеними нечіткими числами $\tilde u^m,\tilde v^m$ можна застосувати нечітку величину

$$\tilde\rho(\tilde u^m,\tilde v^m)=\{(d,\gamma): d=\left\|L^m_{\tilde u}(\gamma)-L^m_{\tilde v}(\gamma)\right\|, \gamma=\min_{r\in\{u,v\},\, i=\overline{1,m}}\mu^r_{\tilde b_i}(\cdot)\}, \quad (10)$$

де $L^m_{\tilde u}(\gamma)$, $L^m_{\tilde v}(\gamma)$ – множини скалярного рівня $\gamma\in(0,1]$ складених нечітких чисел $\tilde u^m,\tilde v^m$ відповідно.

У разі використання множин векторного рівня побудова такої метрики вимагає більшої формалізації. Припустимо, що для заданих чисел $\tilde u^m,\tilde v^m$ побудовано множини рівня $L^m_{\tilde u}(\alpha_1,...,\alpha_m)$ і $L^m_{\tilde v}(\alpha_1,...,\alpha_m)$. Це означає, що виконані співвідношення $\mu^u_{\tilde b_i}(x^i)\ge\alpha_i$, $\mu^v_{\tilde b_i}(y^i)\ge\alpha_i$, $i=\overline{1,m}$. За цими даними можна побудувати складене нечітке число вигляду $\tilde D^m=\{(|x^i-y^i|,\alpha_i), i=\overline{1,m}\}$.

Нескладно помітити, що необхідна для цієї побудови процедура формування компромісного розв'язку буде полягати у розв'язанні задачі двокритеріальної оптимізації вигляду

$$|x^i-y^i|\to\max_{i=\overline{1,m}},\ \alpha_i\to\max_{i=\overline{1,m}}, \quad (11)$$

розв'язок якої $d=\arg\max\limits_{i=\overline{1,m}}|x^i-y^i|$, $\gamma=\arg\max\limits_{i=\overline{1,m}}\alpha_i$, дозволяє визначити відстань між множинами $L^m_{\tilde u}(\alpha_1,...,\alpha_m)$ і $L^m_{\tilde v}(\alpha_1,...,\alpha_m)$ у вигляді нечіткої величини за Вонгом $\tilde W(\tilde D^m)=\{(d,\gamma)\}$.

Таким чином, нечітка відстань $\tilde{\rho}(\tilde{u}^m, \tilde{v}^m)$ між довільними складеними нечіткими числами $\tilde{u}^m, \tilde{v}^m$ при використанні множин векторного рівня можна визначити як нечітку величину $\tilde{W}(\tilde{D}^m)$, що отримується в результаті розв'язання задачі (11).

Запропоновано принцип порівняння відстаней між складеними нечіткими числами. В його основу покладено спосіб обробки нечітких відношень, введений Орловським С.А. Відповідно до цього, якщо задано $\tilde{\rho}_1 = \tilde{\rho}(\tilde{u}_1^m, \tilde{v}_1^m)$, $\tilde{\rho}_2 = \tilde{\rho}(\tilde{u}_2^m, \tilde{v}_2^m)$ — нечіткі відстані між складеними нечіткими числами $\tilde{u}_1^m, \tilde{v}_1^m$ і $\tilde{u}_2^m, \tilde{v}_2^m$ відповідно, то можна визначити нечітку відстань, яка є меншою у розумінні нечіткого відношення переваги "<".

Для знаходження «меншої» відстані вважається, що нечітке відношення переваги $g(a,b)$ для довільних нечітких елементів $a \in \tilde{A}, b \in \tilde{B}$ за умови $\mu_g(a,b) = \min\{\mu_{\tilde{A}}(a), \mu_{\tilde{B}}(b), \mu_{\tilde{A}\times\tilde{B}}(a,b)\} > 0$ визначає співвідношення у вигляді $a < b$ із ступенем $g(a,b) = \mu_g(a,b)$.

За його допомогою можна порівняти відстані $\tilde{\rho}_1 = \tilde{\rho}(\tilde{u}_1^m, \tilde{v}_1^m)$ і $\tilde{\rho}_2 = \tilde{\rho}(\tilde{u}_2^m, \tilde{v}_2^m)$: вираз $g(\tilde{\rho}_1, \tilde{\rho}_2)$ визначає ступінь того, наскільки $\tilde{\rho}_1$ «менше», ніж $\tilde{\rho}_2$. Це також надає можливість визначити "найближчий" до заданого складеного нечіткого числа $\tilde{Z}^0$ елемент $\tilde{Z}^*$, де $\tilde{Z}^*$ – складене нечітке число, значення функцій належності якого для кожного $x^i \in X_i$, $i = \overline{1,m}$, визначаються із співвідношень

$$\mu_{\tilde{b}_i}^{z^*}(x_i) = \min_{x \in X_i}(1 - \mu_T(x_i, x)) = 1 - \max_{x \in X_i} \mu_T(x_i, x),\ i = \overline{1,m},$$

де через *T* позначено нечітке відношення строгої переваги, що відповідає $g(a,b)$, $a, b \in X_i$, $i = \overline{1,m}$

$$\mu_T(a,b) = \begin{cases} 0, \text{ якщо } \mu_g(a,b) < \mu_g(b,a) \\ \mu_g(a,b) - \mu_g(b,a), \text{інакше} \end{cases}.$$

У шостому підрозділі проаналізовано процес групування (кластеризації) даних як етап обробки статистичної інформації. Групування за своєю сутністю полягає у розподілі наявної сукупності даних на групи за істотними для них ознаками. Для забезпечення процесу науково обґрунтованого групування наведено єдині правила, вимоги та принципи, яких необхідно дотримуватися. Визначено основні види групувань та основні питання методології їх проведення.

У сьомому підрозділі розглянуто загальні методологічні принципи кластерного аналізу та методик його проведення. Наведено короткий огляд проблем динамічної кластеризації. Зазнчено, що динамічний підхід у задачах групування – напрямок у вивченні даних, отриманих за спостереженнями процесів, що відбуваються у соціальному середовищі. У цьому випадку об'єктами досліджень є зміни в мережевій структурі, що відбуваються у часі. Основна увага приділяється визначенню приблизних кластерів користувачів та їх часових змін.

Другій розділ дисертаційної роботи присвячено визначенню та застосуванню методик групування складених нечітких множин (СНЧ) трикутного вигляду. В розділі розглянуто і розв'язано задачу кластеризації сукупності СНЧ. Запропонований підхід базується на модифікації існуючих методів кластеризації, основна ідея якої полягає в обчисленні та застосуванні значень відстані між елементами відповідних множин заданого скалярного або векторного рівня сукупності складених нечітких чисел. Запропоновані алгоритми дозволяють формалізувати пошук кластерних центрів сукупності і реалізувати процедури групування даних.

Математичні методи групування даних, заданих у формі складених нечітких чисел, викладено у підрозділах 1–4. Наведено алгоритми кластеризації даних, що подаються у вигляді сукупності складених нечітких чисел з $K(\tilde{b}^m)$. Розроблені алгоритми узагальнюють найбільш відомі підходи у розв'язанні задачі кластеризації: методи C-means, пікового групування та різницевого групування, Густафсона-Кесселя.

Припустимо, що задана деяка сукупність невироджених складених нечітких чисел $\{\tilde{A}^{m(j)}, j=\overline{1,p}\}$ из $K(\tilde{b}^m)$. В подальшому будемо вважати, що використовується скалярний рівень $0<\gamma\le 1$ або векторний рівень $\gamma=(\gamma_1,...,\gamma_m)$, $\gamma_i\in(0,1]$, які визначають гарантований рівень нечіткості даних, що розглядаються.

В алгоритмі *нечіткого групування C-means* припускається, що задану сукупність можна згрупувати в k кластерів. Визначаємо k центрів кластерів у вигляді набору складених нечітких чисел

$$\tilde{C}^{m(i)}=\{(x^{1(i)},\mu_{\tilde{A}_1}(x^{1(i)})),(x^{2(i)},\mu_{\tilde{A}_2}(x^{2(i)})),...,\ (x^{m(i)},\mu_{\tilde{A}_m}(x^{m(i)}))\}\ ,\ i=\overline{1,k}\,.$$

Припустимо, що складені нечіткі числа $\tilde{A}^{m(j)}$, $j=\overline{1,q}\,, q\le p$, мають непорожні множини рівня $L^m(\bar{\gamma})$, які представляють собою звичайні вектори $S(\tilde{A}^{m(j)})=\{x^{1(j)},x^{2(j)},...,x^{m(j)}\}$, $j=\overline{1,q}$. Зрозуміло, що вектори $S(\tilde{A}^{m(j)})$ будуть належати до різних груп з центрами $S(\tilde{C}^{m(i)})$, $i=\overline{1,k}$, відповідно, зі ступенем належності u_{ij}, $i=\overline{1,k}$, $j=\overline{1,q}$.

Функцію помилок, котра відповідає такому представленню, можна визначити як суму часткових помилок у визначенні належності центрам з урахуванням значень ступенів належності u_{ij}, $i=\overline{1,k}$, $j=\overline{1,q}$, відповідно:

$$E(S)=\sum_{i=1}^{k}\sum_{j=1}^{q}u_{ij}\left\|S(\tilde{C}^{m(i)}),S(\tilde{A}^{m(j)})\right\|\ . \tag{12}$$

Тоді задача групування буде полягати у знаходженні мінімальних значень функції помилок $E(S)$ з p обмеженнями типу $\sum_{i=1}^{k}u_{ij}=1$, $j=\overline{1,q}$. Розв'язок даної задачі подається у вигляді

$$S(\tilde{C}^{m(i)})=\sum_{j=1}^{q}u_{ij}S(\tilde{A}^{m(j)})\Big/\sum_{j=1}^{q}u_{ij}\ , \tag{13}$$

$$u_{ij}=1/\sum_{t=1}^{k}\left(\frac{\xi_{ij}^2}{\xi_{tj}^2}\right)^{\frac{1}{m-1}}, \tag{14}$$

де $\xi_{ij}=\xi(S(\tilde{C}^{m(i)}),S(\tilde{A}^{m(j)}))=\left\|S(\tilde{C}^{m(i)}),S(\tilde{A}^{m(j)})\right\|$, $i=\overline{1,k}$, $j=\overline{1,q}$ – евклідова відстань між парами векторів, які визначають центри $\tilde{C}^{m(i)}, i=\overline{1,k}$ і складені нечіткі числа $\tilde{A}^{m(j)}$, $j=\overline{1,q}$, відповідно.

Остаточно алгоритм формулюється таким чином:

1. Задати величину скалярного рівня $0<\gamma\le 1$ або вектор величин рівня $\gamma=(\gamma_1,...,\gamma_m)$, $\gamma_i\in(0,1]$.
2. Визначити значення коефіцієнтів u_{ij}, $i=\overline{1,k}$, $j=\overline{1,q}$ з урахуванням умов нормування.
3. Визначити k центрів кластерів у відповідності з (2.2).
4. Обчислити значення функції помилок у відповідності з формулою (12). Якщо її значення не перевищує деякої граничної величини або за умови незначного зменшення значення помилки щодо попередньої ітерації, то обчислення припиняються. Поточні значення центрів кластерів складають розв'язок задачі. У протилежному випадку необхідно перейти до п'ятого пункту.
5. Розрахувати нові значення u_{ij}, $i=\overline{1,k}$, $j=\overline{1,q}$ за формулами (14) і перейти до другого пункту.

У алгоритмі *пікового групування*, що запропонований Р. Єгером та Д. Фільовим, розглянуто міру щільності розміщення векторів $S(\tilde{A}^{m(j)})$, $j=\overline{1,q}$, для чого генеруються так звані пікові функції. При застосуванні p вхідних векторів створюється сітка, що рівномірно накриває простір векторів $S(\tilde{A}^{m(j)})$, $j=\overline{1,q}$. Вузли цієї сітки розглядаються як потенційні нечіткі центри

$$\tilde{C}^{m(i)}=\{(x^{1(i)},\mu_{\tilde{A}_1}(x^{1(i)})),(x^{2(i)},\mu_{\tilde{A}_2}(x^{2(i)})),...,(x^{m(i)},\mu_{\tilde{A}_m}(x^{m(i)}))\},\ i=\overline{1,k},$$

і для кожного з них розраховується величина пікової функції $M(S(\tilde{C}^{m(i)}))$, $i=\overline{1,k}$,

$$M(S(\tilde{C}^{m(i)}))=\sum_{j=1}^{p}\exp\left(-\left\|S(\tilde{C}^{m(i)}),S(\tilde{A}^{m(j)})\right\|^{2b}\Big/2\sigma^2\right), \quad (15)$$

де $\left\|S(\tilde{C}^{m(i)}),S(\tilde{A}^{m(j)})\right\|$, $i=\overline{1,k}$, $j=\overline{1,p}$ – евклідові відстані між парами векторів, що визначають центри $\tilde{C}^{m(i)}, i=\overline{1,k}$ та складені нечіткі числа $\tilde{A}^{m(j)}$, $j=\overline{1,p}$, відповідно, коефіцієнт σ – константа, що індивідуально підбирається для кожної конкретної задачі, а b – показник ступеню узагальненої функції Гауса.

Величина функції $M(S(\tilde{C}^{m(i)}))$, $i=\overline{1,k}$, розглядається як оцінка висоти пікової функції. Вона пропорційна кількості векторів $S(\tilde{A}^{m(j)})$, $j=\overline{1,q}$, що знаходяться в околі потенційного центру $S(\tilde{C}^{m(i)})$, $i=\overline{1,k}$. Мале значення $M(S(\tilde{C}^{m(i)}))$, $i=\overline{1,k}$, свідчить про те, що центр $S(\tilde{C}^{m(i)})$, $i=\overline{1,k}$, розташований в області з невеликою кількістю векторів $S(\tilde{A}^{m(j)})$, $j=\overline{1,q}$. Слід звернути увагу на те, що коефіцієнт σ має незначний вплив на кінцеві пропорції між $M(S(\tilde{C}^{m(i)}))$, $i=\overline{1,k}$, для різних значень $S(\tilde{C}^{m(i)})$, $i=\overline{1,k}$, тому підбір його величини не є критичним.

Метод *різницевого групування* даних – це модифікація алгоритму пікового групування, в якій усі навчальні вектори $S(\tilde{A}^{m(j)})$, $j=\overline{1,q}$, розглядаються в якості k потенційних центрів, $k=q$. Пікова функція $D(S(\tilde{A}^{m(i)}))$, $i=\overline{1,k}$, у цьому алгоритмі задається у вигляді

$$D(S(\tilde{A}^{m(i)}))=\sum_{j=1}^{p}\exp\left(-\left\|S(\tilde{A}^{m(i)}),S(\tilde{A}^{m(j)})\right\|^{2b}\Big/\left(r_a/2\right)^2\right),\ i=\overline{1,k}. \quad (16)$$

Значення коефіцієнта r_a визначає сферу сусідства. На значення $D(S(\tilde{A}^{m(i)}))$, $i=\overline{1,k}$, істотно впливають лише ті вектори $S(\tilde{A}^{m(j)})$, $j=\overline{1,q}$, що розташовані в межах цієї сфери. За великої щільності точок в околі $S(\tilde{A}^{m(i)})$, $i=\overline{1,k}$, (потенційного центра) значення функції $D(S(\tilde{A}^{m(i)}))$, $i=\overline{1,k}$, велике. І, навпаки, її мале значення свідчить про те, що в околі $S(\tilde{A}^{m(i)})$, $i=\overline{1,k}$, незначна кількість даних. Така точка вважається «невдалим» кандидатом в центри. Після розрахунку значень пікової функції для кожної точки $S(\tilde{A}^{m(j)})$, $j=\overline{1,q}$ відбирається вектор $S(\tilde{A}^{m})$, для якого міра щільності $D(S(\tilde{A}^{m}))$ виявилась найбільшою.

Алгоритм Густафсона-Кесселя є вдосконаленням алгоритму C-means, що полягає у введенні у формулу розрахунку відстані між окремими векторами коефіцієнтів масштабування. Це реалізовується на основі довільної квадратної симетричної додатно визначеної матриці, усі власні значення якої $\lambda_i, i=\overline{1,m}$, є дійсними та додатніми. Множина власних векторів, що відповідють таким власним числам, формує ортогональну базу багатомірного простору, у якому довжини нормованих власних векторів масштабуються з коефіцієнтами $\sqrt{\lambda_i}, i=\overline{1,m}$.

Розв'язування задачі розміщення центрів за даним алгоритмом відбувається аналогічно алгоритму C-means шляхом багатократного перерахунку коефіцієнтів належності u_{ij}, $i=\overline{1,k}$, $j=\overline{1,q}$, за формулою (14) та координат центрів $S(\tilde{C}^{m(i)})$, $i=\overline{1,k}$, за формулою (13), але з урахуванням масштабування при розрахунку відстаней.

Запропоновані алгоритми дозволяють реалізувати процедури групування даних в межах попередньо заданої або автоматично згенерованої по ходу алгоритму кількості кластерів. Розглянуто умови їх конструктивного застосування в задачах групування. Наведено приклади використання підходу при розв'язанні практичних задач, проаналізовано результати чисельних експериментів.

Третій розділ присвячено дослідженню етапів роботи генетичного алгоритму. Наведено сутність операторів генетичного алгоритму, визначено його переваги та недоліки. Увагу зосереджено на використанні генетичного алгоритму для кластеризації сукупності складених нечітких чисел трикутного вигляду. В розділі запропоновано спосіб кодування ознак, заданих цілими числами та числами з плаваючою крапкою, з використанням кодів Грея. Відзначено, що наведена схема генетичного алгоритму дозволяє вирішувати задачу формування відносно однорідних груп, які подаються сукупністю складених нечітких чисел.

У першому підрозділі розглянуто основну ідею та етапи роботи генетичних алгоритмів. Генетичні алгоритми – це стохастичні евристичні оптимізаційні методи, головна ідея яких взята з теорії еволюційного розвитку видів. Механізмом еволюції є природний відбір, суть якого полягає в тому, що більш пристосовані особини мають більше шансів на виживання. У підрозділі два розглянуто обґрунтування доцільності використання генетичного алгоритму, його переваги та недоліки.

Третій підрозділ присвячено опису застосування генетичного алгоритму для задач кластеризації даних як задачі мінімізації розбіжностей в групах елементів вхідної сукупності. Розглянуто особливості кодування ознак об'єктів у схемі генетичного алгоритму: оригінальний спосіб подання ознак у вигляді цілих чисел (генотипу); подання ознак, заданих числами з плаваючою крапкою; спосіб визначення фенотипу об'єкта за його генотипом.

У четвертому підрозділі наведено схему генетичного алгоритму для кластеризації сукупності складених нечітких чисел. Припускається, що задана деяка сукупність невироджених складених нечітких чисел $\{\tilde{A}^{m(j)}, j=\overline{1,p}\}$ из $K(\tilde{b}^m)$. Вважаємо, як і раніше, що використовується скалярний рівень $0<\gamma\le 1$ або векторний рівень $\gamma=(\gamma_1,...,\gamma_m), \gamma_i \in (0,1]$, які визначають гарантований рівень нечіткості даних, що розглядаються.

Набір значень генотипу за кожною ознакою, що визначає фенотип, буде визначатися за допомогою скалярного або векторного рівня нечіткості. Без обмеження загальності усі значення, які подаються числами з плаваючою крапкою, належать інтервалу [0,1], а їх кодування проводиться за вищезгаданою схемою. Кожна хромосома формується у вигляді бітової послідовності фіксованої довжини, кожна з ділянок якої має однаковий розмір і відповідає значенням генів. Значення окремих генів утворюють першу нитку хромосоми. Двійкові рядки значень генів характеризуються особистими величинами мір належності відповідно до схем побудови множин скалярного $0<\gamma\le 1$ або векторного рівня $\gamma=(\gamma_1,...,\gamma_m), \gamma_i \in (0,1]$. Значення відповідних мір належності утворюють другу нитку кожної з хромосом, які визначають різнотипність особин фенотипу.

Реалізуючи оператори схрещування, мутації та інверсії за процедурами, розробленими у розділі, отримуємо схему функціонування генетичного алгоритму у випадку групування нечітких даних, поданих у вигляді складених нечітких чисел $\tilde{A}^{m(j)}$, $j=\overline{1,p}$:

1. Сформувати множини скалярного або векторного рівня для нечітких чисел, які визначають гени. Вважаємо, що вони є не порожніми. Побудувати сукупність особин (хромосом), які представляють собою звичайні вектори $A_j=\{x_1^j, x_2^j,..., x_m^j\}$, $j=\overline{1,p}$.
2. Ініціювати початковий момент часу *t = 0*. Випадковим чином сформувати початкову популяцію, що складається з *k* особин $B_0=\{A_1,A_2,...,A_k\}$.

3. Обчислити пристосованість кожної особини $F_{A_i} = fit(A_i)$, $i = \overline{1,k}$, і популяції в цілому $F_t = fit(B_t)$, використовуючи задану фітнес-функцію. Значення цієї функції визначає наскільки відповідає особина, описана даною хромосомою, для розв'язування задачі.
4. Вибрати особину A_{c1} з популяції $A_{c1} = Get(B_t)$.
5. З певною ймовірністю кросовера P_c вибрати другу особину з популяції $A_{c2} = Get(B_t)$ і виконати оператор кросовера $A_c = Crossover(A_{c1}, A_{c2})$.
6. З певною ймовірністю мутації P_m виконати оператор мутації $A_c = mutation(A_c)$.
7. З певною ймовірністю інверсії P_i виконати оператор інверсії $A_c = inversion(A_c)$.
8. Розмістити отриману хромосому в новій популяції *insert* (B_{t+1}, A_c).
9. Виконати операції, починаючи з пункту 3, k раз.
10. Збільшити номер поточної епохи $t = t + 1$.
11. Якщо виповнилося умова зупинки, то завершити роботу,
    інакше перехід на крок 2.

У **четвертому розділі** роботи запропоновано формалізацію динаміки кластерних структур, що формуються на основі змінних у часі сукупностей складених нечітких чисел. В якості моделі процесу змін у сукупності розглянуто нечітку різницеву систему, стани якої в довільний момент часу є складеними нечіткими числами. Також розглянуто методику вибору рішень за умов подання вихідних даних у вигляді сукупності складених нечітких чисел та за наявності набору критеріїв, множини значень яких задаються числовими інтервалами з лінійними функціями належності.

Припускається, що функціонування процесів та явищ нечітких систем описується нечіткими різницевими рівняннями:

$$\tilde{X}(t+1) = \tilde{X}(t) \circ \tilde{R}\ , \tag{17}$$

де $\tilde{R}$ – нечітке відображення з X в X з функцією належності $\mu_{\tilde{R}}(x, y)$, $x, y \in X$. Операція "$\circ$" означає максимальну композицію, яка для (18) має вигляд:

$$\mu_{\tilde{X}_{t+1}}(y) = \max_{x \in X}[\min(\mu_{\tilde{X}(t)}(x), \mu_{\tilde{R}}(x, y))] \quad . \tag{18}$$

У третьому підрозділі запропоновано формалізацію динаміки кластерних структур для задачі кластеризації сукупностей складених нечітких чисел.

Нехай задано m нечітких чисел $\tilde{b}_i$, $i = \overline{1,m}$, визначених на універсальних множинах $X_1, ..., X_m$, відповідно, $X_i \subseteq R^1$, $i = \overline{1,m}$. Множина $T = \{0,1,2,..., M\}$, де M – наперед задане додатне число, визначає дискретні моменти часу. Позначимо через $T_r = \{t_0, t_1, t_2, t_3, ..., t_r\} \subset T$ впорядковану за зростанням послідовність моментів часу з T.

Обмежуючи функціонування деякого процесу, будемо описувати стани нечіткої різницевої системи в довільний момент часу t_k, $k = 0,1,2,...$, у вигляді складених нечітких чисел $\tilde{X}^m(t_k) = \{(x_k^1, \mu_{\tilde{b}_1}(x_k^1)), ..., (x_k^m, \mu_{\tilde{b}_m}(x_k^m))\}, k = 0,1,2,...$.

Перепишемо систему (17) у вигляді:

$$\tilde{X}_{k+1}^m = \tilde{\mathrm{R}}_k \circ \tilde{X}_k^m,\ k = 0,1,2,... \quad , \tag{19}$$

де $\tilde{X}^m(t_0) = \tilde{X}_0^m$ – СНЧ, що є початковим станом системи, $\tilde{X}^m(t_k) = \tilde{X}_k^m$ – СНЧ, що описують стани системи в моменти часу $t_k \in T$, $k = 1,2,...$, $\tilde{\mathrm{R}}(t_k) = \tilde{\mathrm{R}}_k$ – деякі нечіткі відображення з R^1 в R^1, що визначають переходи системи, $k = 0,1,2,...$, "$\circ$" – операція максимальної композиції, результат якої визначається за формулою (18).

Означення 4.1. Траєкторією системи (19) за i-ю компонентою, $i=\overline{1,m}$, будемо називати послідовність $\{x^i(t_k)=x_k^i\}$, $k=0,1,2,...$, для її елементів справедливі співвідношення:

$$x_{k+1}^i = R_k(x_k^i),\ k=0,1,2,...\ ,\ i=\overline{1,m}, \text{ де } R_k \subseteq supp\ \tilde{R}_k,\ k=0,1,2,...\ . \qquad (20)$$

Означення 4.2. Траєкторію системи (19) $\{\bar{x}_k^i\}$, $k=0,1,2,...$, $i=\overline{1,m}$, за i-ю компонентою, $i=\overline{1,m}$, назвемо регулярною, якщо для її елементів справедлива умова:

$$\mu_{\tilde{b}_i}(\bar{x}_k^i)=1,\ k=0,1,2,..., \text{ для заданого } i=\overline{1,m}. \qquad (21)$$

Нехай у момент t_0 задано довільну сукупність СНЧ $\tilde{S}_j^m$, $j=\overline{1,s}$, з $K(\tilde{b}^m)$. Динамічний процес, що визначає зміни елементів сукупності $\tilde{S}_j^m$, $j=\overline{1,s}$, у часі, може бути формалізований за допомогою нечіткої різницевої моделі (19). Для побудови кластерних структур розглянемо різні випадки нечітких станів системи $\tilde{X}_0^m=\tilde{S}_j^m$, $j=\overline{1,s}$, $\tilde{X}_k^m$, $k=1,2,...$.

1) Задана сукупність має регулярну траєкторію за довільною компонентою i, $i=\overline{1,m}$. Це означає, що існує СНЧ $\tilde{S}_p^m$, $p=\overline{1,s}$, з заданої сукупності, для якого траєкторія $\{\bar{x}_k^i\}$, $\bar{x}_k^i \in \operatorname{supp}\tilde{b}_i, k=0,1,2,...$, за компонентою i, $i=\overline{1,m}$, є регулярною траєкторією моделі (19) з початковими даними у вигляді $\tilde{X}_0^m=\tilde{S}_p^m$.

Тоді зрозуміло, що СНЧ $\tilde{X}_k^m$, $k=0,1,2,...$, i-а компонента яких задається величинами $\bar{x}_k^i$, $k=0,1,2,...$, можна розглядати як кандидатів на центри динамічних кластерів в будь-який момент часу $t_k \in T_r$ для поточного набору СНМ сукупності.

У частинному випадку для i-ї компоненти окремих СНЧ з поточного набору сукупності на кожному кроці динамічного процесу виконуються співвідношення $\bar{x}_{k+1}^i=\bar{x}_k^i$, $k=0,1,2,...$. Тоди i-а компонента центрів динамічних кластерів визначається значенням $\bar{x}_k^i$, $k=0,1,2,...$, і процедура групування даних зводиться до формування у часі кластерних структур для модифікованих сукупностей СНЧ, які є зрізами $K(\tilde{b}^m)$ за i-ю компонентою.

2) Припустимо далі, що у сукупності немає регулярної траєкторії за будь-якою компонентою. Це означає, що для довільного початкового стану системи (19), які визначаються елементами заданої сукупності, для будь-якої траєкторії $\{x_k^i\}$, $x_k^i \in \operatorname{supp}\tilde{b}_k, k=0,1,2,...$, $i=\overline{1,m}$, існує момент часу $t_d^i \in T_r$, що виконується нерівність $\mu_{\tilde{b}_i}(x_d^i)<1$. За таких умов розглянемо два випадки.

а) Нехай числова множина $Y_0=\{y_0^1,...,y_0^n\}$, $n \le s$, складається з переліку значень носіїв i-ї компоненти, $i=\overline{1,m}$, початкових станів $\tilde{S}_p^m$, $p=\overline{1,s}$, системи (19), для яких існує такий момент часу $t_d^i \in T_r$, що $\mu_{\tilde{b}_i}(x_d^i)=1$.

У якості i-ї компоненти центрів динамічних кластерних структур на кожному кроці процесу змін сукупності складених нечітких множин може бути рекомендовано значення, яке є результатом оптимального вибору з використанням правила розрахунку відстані між окремими компонентами елементів кластерів або результатом формального вибору між величинами, для яких у цей момент часу $t_d^i \in T_r$ виконується умова $\mu_{\tilde{b}_i}(x_d^i)=1$.

б) Нарешті припустимо, що неможливо виділити множину Y_0 значень носіїв i-ї компоненти, $i=\overline{1,m}$, початкових станів $\tilde{S}_p^m$, $p=\overline{1,s}$, системи (19), для яких існує момент часу $t_d^i \in T_r$, що $\mu_{\tilde{b}_i}(x_d^i)=1$. В цьому випадку потрібно говорити про пряме застосування методів та алгоритмів кластеризації для кожного поточного переліку складених нечітких множин заданої сукупності $\tilde{S}_j^m$, $j=\overline{1,s}$.

У третьому підрозділі запропоновано алгоритм прийняття рішень з складеними нечіткими початковими даними.

Припустимо, що вибір рішення в деякій прикладній сфері здійснюється за набором критеріїв $K_1,\ldots,K_p$, а множини значень критеріїв задаються числовими інтервалами $X_i \subset R^1$, $i=\overline{1,p}$. Ці інтервали можна розглядати в якості носіїв нечітких чисел трикутного вигляду $\tilde{b}_1,\ldots,\tilde{b}_p$, $\tilde{b}_i=\{(a_i,b_i,c_i)\}$, $a_i \le b_i \le c_i$, $i=\overline{1,p}$ з лінійними функціями належності $\mu_{\tilde{b}_1}(x),\ldots,\mu_{\tilde{b}_p}(x)$ вигляду (1), що визначені на відповідних універсальних множинах. Тому будь-який розв'язок представляється у вигляді СНЧ

$$\tilde{S}^p=\{(x^1,\mu_{\tilde{b}_1}(x^1)),(x^2,\mu_{\tilde{b}_2}(x^2)),\ldots,(x^p,\mu_{\tilde{b}_p}(x^p))\}\in K(\tilde{S}^p), \tag{22}$$

де кожне значення $x^i \in X_i$, $i=\overline{1,p}$, є кількісним показником, що відповідає критерію K_i, $i=\overline{1,p}$.

Складені нечіткі числа $\tilde{S}^p$ з $K(\tilde{S}^p)$ будемо називати альтернативами. Таким чином, кожна альтернатива $\tilde{S}^p$ з відповідними значеннями функцій належності характеризує нечітке представлення особи, що приймає рішення, про прийнятність окремих елементів розв'язку для прийняття відповідного рішення у конкретній ситуації.

Розглянемо ситуацію прийняття рішень в умовах невизначеності, для цього використаєм нечіткий підхід. Нехай є сукупність алгоритмів, яку позначимо через $S=\{S_1,S_2,\ldots,S_n\}$, та нехай величини кожного з критеріїв K_i, $i=\overline{1,p}$, що характеризує альтернативи, задаються дискретним набором значень $\{b_1^i,\ldots,b_{n_i}^i\}$, $i=\overline{1,p}$, з відповідного числового інтервалу $X_i=[a^i,c^i]\subseteq R^1$, $a^i \le s_1^i \le \ldots \le s_{n_i}^i \le c^i$, $i=\overline{1,p}$. Тоді кожен елемент альтернатив у ситуації прийняття рішень може бути формалізований у вигляді нечіткого числа

$$\tilde{b}_i=\{(b_1^i,\mu_{\tilde{b}_i}(b_1^i)),\ldots,(b_{n_i}^i,\mu_{\tilde{b}_i}(b_{n_i}^i))\},\ i=\overline{1,p}, \tag{23}$$

значення функції належності для якого визначатимуть рівні відповідності значень окремих критеріїв конкретному ситуаційному рішенню. Без обмеження загальності нехай функції належності лінійні, а нечіткі числа $\tilde{b}_i$, $i=\overline{1,p}$, – трикутні. Тоді будь-яке рішення S_q, $q=\overline{1,n}$, у проблемній ситуації подається у вигляді СНЧ:

$$\tilde{S}_q^p=\{(x_1^q,\mu_{\tilde{b}_1}(x_1^q)),(x_2^q,\mu_{\tilde{b}_2}(x_2^q)),\ldots,(x_p^q,\mu_{\tilde{b}_p}(x_p^q))\},\ x_i^q\in X_i,\ i=\overline{1,p},\ q=\overline{1,n}.$$

Таким чином, задаючи вихідну інформацію для ситуації прийняття рішень у вигляді сукупності альтернатив, що описуються як СНЧ:

$$\tilde{U}_j^p=\{(u_j(x^1),\mu_{\tilde{b}_1}(u_j(x^1))),(u_j(x^2),\mu_{\tilde{b}_2}(u_j(x^2)),\ldots,(u_j(x^p),\mu_{\tilde{b}_p}(u_j(x^p))\},\ j=\overline{1,r},$$

де $u_j(x^i)$, $j=\overline{1,r}$, $i=\overline{1,p}$, — конкретні значення критеріальних функцій $K_1,\ldots,K_p$, t — кількість спостережень, зводимо процедуру вибору рішень до аналізу та взаємного співставлення наборів складених нечітких чисел $\tilde{U}_j^p$, $j=\overline{1,r}$, та $\tilde{S}_q^p$, $q=\overline{1,n}$. Задача в цій постановці може розглядатися як задача групування (кластеризації) нечітко визначених даних з заданими *n* центрами кластерів, для вирішення якої використовуються різні методики.

Дійсно, якщо сукупність складених нечітких чисел $\tilde{S}_q^p$, $q=\overline{1,n}$ розглядати як центри нечітких кластерів, то задача буде полягати у визначенні груп вихідних даних, що утворюються цими центрами. Визначення даних, які відповідають центрам кластеризації, можливе за умов визначення способу розрахунку відстаней у сукупності складених нечітких множин.

Для двох довільних складених нечітких чисел

$$\tilde{U}^p=\{(x^1,\mu_{\tilde{b}_1}^U(x^1)),(x^2,\mu_{\tilde{b}_2}^U(x^2)),\ldots,(x^p,\mu_{\tilde{b}_p}^U(x^p))\}\ \text{та}\ \tilde{V}^p=\{(x^1,\mu_{\tilde{b}_1}^V(x^1)),(x^2,\mu_{\tilde{b}_2}^V(x^2)),\ldots,(x^p,\mu_{\tilde{b}_p}^V(x^p))\}$$

можна побудувати множини скалярного рівня $\gamma>0$ $L_{\tilde{U}^p}^p(\gamma)$ та $L_{\tilde{V}^p}^p(\gamma)$ або векторного рівня $(\alpha_1,\ldots,\alpha_p)$, $\alpha_i\in(0,1]$, $i=\overline{1,p}$, $L_{\tilde{U}^p}^p(\alpha_1,\ldots,\alpha_p)$ та $L_{\tilde{V}^p}^p(\alpha_1,\ldots,\alpha_p)$ у якості звичайних множин, між якими визначена евклідова відстань d .

Як було визначено у першому розділі, для знаходження нечіткої відстані $\rho(\tilde{U}^p, \tilde{V}^p)$ між довільними складеними нечіткими числами $\tilde{U}^p, \tilde{V}^p$ можна використати нечітку величину за Вонгом, а порівняння відстаней між двома довільними парами СНЧ провести за допомогою нечіткого відношення переваги, формалізованому у розділі 1.

Використовуючи введене поняття відстані між СНЧ та спосіб порівняння відстаней можна сформулювати задачу нечіткого групування в кластерів даних, поданих у вигляді сукупності СНЧ $\tilde{U}_j^p$, $j = \overline{1, r}$ з $K(\tilde{S}^p)$. Для розв'язування задачі кластеризації можна застосувати алгоритми, наведені у розділі 2.

У п'ятому розділі роботи розроблені методи та алгоритми використано для розв'язання прикладних задач, що характеризуються неточністю або невизначеністю даних. Запропоновано варіант реалізації генетичного алгоритму для розв'язання задачі складання розкладу з урахуванням побажань викладачів щодо часу проведення занять. Побудовано математичну модель для визначення оптимального розподілу потужностей каналів передачі даних у локальній мережі. Наведено приклади розв'язування практичних задач визначення пропускних здатностей каналів мережі інформаційно-обчислювального центру для ситуацій, в яких потужності з'єднань описуються сукупністю даних у вигляді сукупності СНЧ.

У першому підрозділі розглянуто застосування генетичного алгоритму для задачі складання розкладу. Формується розклад у вигляді двовимірного масиву, у стовпцях якого розміщується інформація про викладача, дисципліну, що він викладає, і навчальну групу. Перелік занять має вигляд $Z = \{z_i\}$, $z_i = \{z_i^p, t_i^d, t_i^g\}$, $i = \overline{1, I}$, де I – кількість навчальних позицій у навчальному плані, z_i^p – номер викладача, z_i^d – номер дисципліни за навчальним планом, z_i^g – номер групи, для якої проводяться заняття, $i = \overline{1, I}$.

Часова сітка навчального процесу також задається двовимірним масивом, стовпці якого містять інформацію про тиждень, день і час. Множину інтервалів представимо у вигляді $T = \{t_k\}$, $t_k = \{t_k^w, t_k^d, t_k^p\}$, $k = \overline{1, K}$, де K – кількість навчаль-них тижнів, t_k^w – номер тижня, t_k^d – кількість навчальних днів у тижні, $t_k^d = \overline{1, K_d}$, t_k^p – номер навчальної години протягом дня, $t_k^p = \overline{1, K_p}$, $k = \overline{1, K}$. З метою задоволення побажань викладачів стосовно часу проведення занять задається блок переваг, який представлений у вигляді двовимірного масиву, що формується для кожного викладача і містить дані щодо бажаних часових інтервалів для навчання. Будемо вважати, що ця інформація подається у формі нечітких числових множин, кожен елемент яких $t_k^p = \overline{1, K_p}$, $k = \overline{1, K}$, додатково описується рівнем задоволення викладача вибором того чи іншого часового інтервалу в межах робочого тижня. Ці множини без обмеження загальності можна розглядати як спосіб нечіткого трикутного (трапецієвидного) подання часового розкладу роботи викладача з урахуванням його побажань.

В результаті роботи буде отримано варіант розкладу, що враховує умови навчального та часового забезпечення навчального процесу. Варіант зберігається у хромосомі, яка подається у вигляді асоціативного масиву, що зв'язує переліки занять та часових інтервалів. Індекси масиву – номери занять, значення – номери часових інтервалів.

Кожна хромосома складається з заданої кількості генів, яка дорівнює сумі кількості атрибутів з переліку занять та одного значення варіативного атрибуту для ведення інформаційного наповнення хромосоми. Таким показником є часові інтервали, які з урахуванням нечіткості в побажаннях часу проведення занять викладача в межах робочого тижня утворюють з даними переліків занять складені нечіткі числові множини. Це означає, що вхідні дані для побудови розкладу подаються сукупністю складених нечітких чисел, для кожного елемента якої визначено оцінку якості у вигляді штрафу. Для побудови оптимального розкладу потрібно знайти такий варіант, що забезпечує мінімальне значення критерію P втрат його якості:

$$P = f(t) = \sum_{n=1}^{N} c_n w_n(t) \rightarrow \min_{T}, \tag{24}$$

де N – кількість частин розкладу, які впливають на його якість, c_n – оцінки ступеня неефективності частини розкладу, w_n – значення коефіцієнтів штрафу за невиконання частини розкладу, $\tau \in T$, $n = \overline{1, N}$. Однією з найбільш важливих вимог до розкладу є показник урахування побажань викладачів щодо свого розкладу занять.

Остаточно маємо схему генетичного алгоритму складання розкладу, який деталізує загальну схему підрозділу 3.4 і складається з таких кроків:

1. Формування переліку занять.
2. Формування множини часових інтервалів.
3. Формування нечітких множин переваг часу викладання.
4. Формування множини скалярного рівня для нечітких чисел, які визначають варіативний ген розв’язків.
5. Генерація початкової популяції хромосом:

5.1. Генерація екземпляру хромосоми;

5.2. Перевірка хромосоми на допустимість;

5.3. Додавання допустимого екземпляру до початкової популяції;

5.4. Перехід до 4.1.

6. Визначення штрафних оцінок хромосом у початковій популяції.
7. Вибір найкращої хромосоми з найменшою оцінкою, а також ще однієї з 10 кращих.
8. Виконання операції кросовера для обраних двох хромосом.
9. Перевірка отриманих хромосом на допустимість.
10. Якщо отримано допустимі хромосоми, оцінюємо їх:

10.1. Якщо отримані хромосоми краще існуючих, додаємо їх у популяцію замість найгірших;

10.2. В іншому випадку:

10.2.1. Виконується мутація;

10.2.2. Перевіряється допустимість і проводиться оцінювання;

10.2.3. Якщо отримані хромосоми краще існуючих, додаємо їх у популяцію замість найгірших.

11. Повторюємо пункти 6–9.
12. Зупиняємо виконання алгоритму у будь-який момент. Результатом роботи алгоритму буде хромосома з найменшою оцінкою штрафів

Застосуванню процесу кластеризації в задачах оптимізації розподілу потужностей каналів передачі даних у комп’ютерних та інформаційних мережах присвячено другий підрозділ. У ньому сформульовано виробничо-транспортну задачу для оптимізації розподілу потужностей каналів, що мінімізує сумарні витрати, котрі пов’язані з організацією передачі даних до пунктів споживання (абонентів мережі).

Якщо ввести позначення: c_{ik}^{p} – питома вартість виробництва продукції i-им виробником за допомогою k-го способу; c_{ij}^{t} – питома вартість перевезень продукції від i-го виробника j-му споживачу; b_j – величина попиту j-го споживача; a_{ik} – кількість продукції, виготовленої i-м виробником k-м способом; z_{ik} – інтенсивність використання k-го способу i-м виробником протягом частини періоду, який приймається рівним 1; x_{ij} – кількість продукції, перевезеної i-м виробником j-му споживачу; $i=\overline{1,N}$, $j=\overline{1,M}$, $k=\overline{1,S}$, тоді виробничо-транспортна задача може бути записана у вигляді дворівневої задачі оптимізації:

$$f(z)=\sum_{i=1}^{N}\sum_{k=1}^{S}c_{ik}^{p}z_{ik}\to\min \tag{25}$$

за умов

$$g(x)=\sum_{j=1}^{M}\sum_{i=1}^{N}c_{ij}^{t}x_{ij}\to\min \tag{26}$$

при обмеженнях

$$\sum_{i=1}^{N}x_{ij}=b_j,\ j=\overline{1,M}, \tag{27}$$

$$\sum_{j=1}^{M}x_{ij}\le\sum_{k=1}^{S}a_{ik}z_{ik},\ i=\overline{1,N}, \tag{28}$$

$$\sum_{k=1}^{S}z_{ik}\le 1,\ i=\overline{1,N}, \tag{29}$$

$$z_{ik}\ge 0,\ x_{ij}\ge 0,\ i=\overline{1,N},\ j=\overline{1,M},\ k=\overline{1,S}. \tag{30}$$

Запропоновано схему для знаходження розв’язку дискретної виробничо-транспортної задачі (25)–(30), яку використано для розв’язання задачі ефективного розподілу потужностей каналів передачі даних між вузлами мережі провайдерів та користувачів з урахуванням потреб і переваг абонентів, так і можливостей провайдерів. Передбачається, що відомі необхідні обсяги абонентів мережі в отриманні тих чи інших типів інформації. Задано побажання абонентів і можливості провайдерів щодо здатності по передачі інформації тому чи іншому абоненту або вузлу. Визначено умови оцінювання ефективності розподілу каналів (відносно їх пропускної здатності). Структуру мережі та інформації, що розподіляється у ній, визначено виробничо-транспортну задачу з лінійними функціями цілі та деякими конструктивними обмеженнями:

- інформація розподіляється від провайдера до абонентів через комутаційні вузли за постійно підключеними каналами зв’язку;
- кожен абонент мережі обслуговується одним або декількома комутаційними вузлами;
- кількість інформації, що розподіляється, для комутаційних вузлів та абонентів може бути обмеженою як зверху (принципові обмеження можливостей провайдера), так і знизу (мінімальна потреба абонентів у необхідній інформації).

Запропоновано застосування алгоритмів кластеризації в задачі оптимізації розподілу потужностей каналів передачі даних. Розв’язано задачу ефективного розподілу потужностей за наявності заданої кількості комунікаційних серверів різної пропускної здатності, залучивши до обслуговування різних груп користувачів окремий сервер.

Для цього розв’язана задача групування користувачів за показниками бажаних швидкостей з’єднаннь окремих показників. До кожної групи вносилися користувачі з інтервально заданими запитами щодо швидкості з’єднання. Величини швидкостей задавалися у вигляді трикутних нечітких чисел. Результати отримано методом групування C-means.

ВИСНОВКИ

Дисертаційна робота присвячена розробці та практичному впровадженню методів і алгоритмів розв'язування задач кластеризації даних, що подаються сукупністю нечітких трикутних чисел, створенню підходів для дослідження динаміки в процесах кластеризації, розробці математичного забезпечення для обробки нечіткої інформації спеціального вигляду, яке може бути використано у процесах створення та впровадження систем підтримки прийняття управлінських рішень в умовах невизначеності.

Основними результатами дисертаційної роботи є:

- запропоновано новий підхід для формалізації невизначеності у вигляді складених нечітких чисел та побудови множин скалярного та векторного рівня;
- для аналізу процесів та проведення групування нечіткої інформації запропоновано новий підхід, що базується на застосуванні множин рівня спеціального вигляду та оригінальній методиці обчислення відстані;
- вперше запропоновано методи кластеризації даних, поданих у вигляді сукупності нечітких чисел;
- сформульовано нову оптимізаційну задачу математичного програмування для розрахунку раціональної кількості груп маршрутизаторів для забезпечення оптимальної потужності каналів передачі даних;
- розроблено та впроваджено новий метод розв'язання задачі розрахунку оптимальної потужності каналів передачі даних у комп'ютерній мережі;
- набула подальшого вдосконалення методика застосування генетичного алгоритму для роз'язання задачі кластеризації.

Розроблені математичні методи та підходи використані для розрахунку оптимальної кількості маршрутизаторів для забезпечення передачі даних у комп'ютерній мережі Інформаційно-обчислювального центру Київського національного університету імені Тараса Шевченка та для розв'язання задачі складання розкладу занять навчального підрозділу на основі використання генетичного алгоритму.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових фахових виданнях України:

1. Апанасенко Д.В. Про методи кластеризації множин складених нечітких чисел Д.В.Апанасенко Вісник КНУ імені Тараса Шевченка. Серія: Кібернетика. – 2015. – №1(15). – С.5–8.
2. Івохін Є.В. Про застосування методів кластеризації нечітких даних спеціального вигляду Є.В.Івохін, Д.В.Апанасенко Вісник Київського університету ім. Тараса Шевченка. Серія: Фізико-математичні науки. – 2013. – №1. – С.167–171.
3. Івохін Є.В. Про кластеризацію складених нечітких чисел на основі множин векторного рівня Є.В.Івохін, Д.В.Апанасенко Вісник КНУ імені Тараса Шевченка. Серія: Фізико-математичні науки. – 2016. – №2. – С.90–93.
4. Івохін Є.В. Про деякі моделі та методи розв'язування нечітких трьохіндексних транспортних задач Є.В.Івохін, Д.В.Апанасенко Вісник КНУ імені Тараса Шевченка. Серія: Фізико-математичні науки. – 2017. – №2. – С.74–78.
5. Ивохин Е.В. Кластеризация совокупности составных нечетких чисел на основе множеств скалярного и векторного уровней Е.В.Ивохин, Д.В.Апанасенко Проблемы управления и информатики. – 2018. – №5. – С.136–147.

Статті у наукових виданнях:

6. Івохін Є.В. Про зведення виробничо-транспортної задачі до дворівневої задачі оптимізації та її застосування Є.В.Івохін, Д.В.Апанасенко, В.О.Навродський Вісник КНУ імені Тараса Шевченка. Серія: Економіка. – 2018. – №3(198). – С.48–53.

Тези наукових доповідей:

7. Івохін Є.В. Про моделювання розповсюдження реклами в межах цільової аудиторії Є.В.Івохін, Ю.О.Науменко, Д.В.Апанасенко Сборник трудов межд. научной конференции имени Т.А.Таран “Интеллектуальний анализ информации” (IAI-2017), Київ, 17–19 травня 2017. – С.96–100.

8. Івохін Є.В. Про реалізацію процесу кластеризації на множині складених нечітких чисел Є.В.Івохін, Д.В.Апанасенко Матеріали VII міжнар. конф. "Сучасні проблеми математичного моделювання, прогнозування та оптимізації", (Камя’нець-Подільський національний університет імені Івана Огієнка, березень 2016). – Камя’нець-Подільський, 2016. –С. 84.

9. Ivohin E. About hybrid mathematical model of information diffusion process E.Ivohin, D.Apanasenko, Yu.Naumenko Abstracts XXXII International Conference “Problems of descision making under uncertainties”, (August 27–31, 2018, Prague, Czech Republic). Prague, 2018. – P.59.

10. Івохін Є.В. Один спосіб кластеризації складених нечітких чисел на основі множин векторного рівня Є.В.Івохін, Д.В.Апанасенко Тези VIII міжн. школи–сем. "Теорія прийняття рішень", (Ужгород, 26 вересня – 01 жовтня 2016). – Ужгород, 2016. – С.128.

11. Івохін Є.В. Про підходи до моделювання розповсюдження реклами як процесу агрегації, обмеженої дифузією Є.В.Івохін, Д.В.Апанасенко, Ю.О.Науменко Матер. IV міжн. наук. –техн. конф. “Обчислювальний інтелект” (ОІ-2017), (Київ, 16–18 травня 2017р.). – Київ, 2017. – С.116–117.

АНОТАЦІЯ

Апанасенко Дмитро Володимирович. Методи кластеризації нечітких даних спеціального вигляду та їх застосування. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня кандидата технічних наук із спеціальності 01.05.04 – системний аналіз і теорія оптимальних рішень. – Київський національний університет імені Тараса Шевченка МОН України, Київ, 2018.

Дисертаційна робота присвячена дослідженню та розробці моделей, методів і алгоритмів для розв’язання задач кластеризації сукупності даних, що подаються у вигляді складених нечітких трикутних чисел. Запропоновано нову методику застосування складених нечітких чисел для формалізації невизначеності.

Розроблено методи розв’язання задачі кластеризації станів нечіткої системи, що описуються сукупністю складених нечітких чисел трикутного вигляду. Запропоновано використання генетичного алгоритму для кластеризації даних у вигляді сукупності складених нечітких чисел.

Сформульовано принцип формалізації динаміки кластерних структур, що формуються на основі змінних у часі сукупностей складених нечітких чисел.

Запропоновано методику вибору рішень за умов подання вихідних даних у вигляді сукупності складених нечітких чисел та за наявності набору критеріїв, множини значень яких задаються числовими інтервалами з лінійними функціями належності.

Розроблено варіант реалізації генетичного алгоритму для розв’язання задачі складання розкладу. Побудовано математичну модель для визначення оптимального розподілу потужностей каналів передачі даних у локальній мережі, запропоновано новий підхід для розв’язання

задачі розподілу потужностей як виробничо-транспортної задачі. Отримано розв'язок задачі групування користувачів за рівнями споживання мережевих потужностей та задачі ефективного розподілу ресурсів каналів передачі даних серед груп користувачів, що обслуговуються.

Ключові слова: складені нечіткі числа, нечіткі трикутні числа, задача кластеризації, генетичний алгоритм, динамічна кластеризація, пропускна здатність, модель виробничо-транспортної задачі, розподіл ресурсів, дворівнева задача оптимізації.

АННОТАЦИЯ

Апанасенко Дмитрий Владимирович. Методы кластеризации нечетких данных специального вида и их применение. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 01.05.04 – системный анализ и теория оптимальных решений. – Киевский национальный университет имени Тараса Шевченко МОН Украины, Киев, 2018.

Диссертационная работа посвящена исследованию и разработке моделей, методов и алгоритмов для решения задач кластеризации совокупности данных, представленных в виде составных нечетких треугольных чисел. В работе предложена новая методика использования составных нечетких чисел для формализации неопределенности, проведена формализация процедур кластеризации информации в условиях неопределенности, усовершенствована схема применения генетического алгоритма для решения задач кластеризации.

Исследована методика группирования составных нечетких чисел треугольного вида. Разработаны методы решения задачи кластеризации состояний нечеткой системы, описываемых совокупностью составных нечетких чисел. Предложено использование генетического алгоритма для кластеризации данных в виде совокупности составленных нечетких чисел треугольного вида.

Сформулирован принцип формализации динамики кластерных структур, которые формируются на основе переменных во времени совокупностей составных нечетких чисел.

Предложена методика выбора решений в условиях представления исходных данных в виде совокупности составных нечетких чисел и при наличии набора критериев, множества значений которых задаются числовыми интервалами с линейными функциями принадлежности.

Предложен вариант реализации генетического алгоритма для решения задачи составления расписания. Построена математическая модель для определения оптимального распределения мощностей каналов передачи данных в локальной сети, предложен новый подход для решения задачи распределения мощностей как производственно-транспортной задачи. Получено решение задачи группировки пользователей по уровням потребления сетевых мощностей и задачи эффективного распределения ресурсов каналов передачи данных среди групп обслуживаемых пользователей.

Ключевые слова: составные нечеткие числа, нечеткие треугольные числа, задача кластеризации, генетический алгоритм, динамическая кластеризация, пропускная способность, модель производственно-транспортной задачи, распределение ресурсов, двухуровневая задача оптимизации.

ANNOTATION

Apanasenko Dmytro Volodymyrovych. Methods of clustering fuzzy data of a special kind and their application. – Manuscript.

Candidate’s thesis on Technics, speciality 01.05.04 – system analysis and optimal decision theory. – Taras Shevchenko National University of Kyiv, Ministry of Education and Science of Ukraine, Kyiv, 2018.

The thesis is devoted to the development and practical application of models, methods and algorithms for solving tasks of clustering of a set of data presented in the form of composite fuzzy triangular numbers were carried out. The paper proposes a new method for the use of composite fuzzy numbers to formalize uncertainty, formalizes clustering information procedures under uncertainties, investigates the dynamics of the process of cluster structures development, and improves the application of the genetic algorithm for the clusterization problem.

The methods of grouping composite fuzzy numbers of a triangular type are investigated. Methods of solving the problem of clustering states of fuzzy systems are developed. The conditions of their constructive application in the problems of grouping are considered.

The basic stages of the work of the genetic algorithm are analyzed, the essence of the operators of the genetic algorithm is presented. A method for coding the characters given by integers and floating point numbers using Gray codes is proposed, and the scheme of the operation of the main operators of the genetic algorithm under the proposed presentation and data encoding is presented. The scheme of grouping fuzzy triangular data based on the genetic algorithm is formulated.

The paper proposes the formalization of the dynamics of cluster structures, which are formed on the basis of time-varying sets of composite fuzzy numbers. As a model of the process of changes, a totally unclear difference system is considered, the states of which at any time are composed by fuzzy numbers. The concept of the trajectory and the regular trajectory of the system are formulated. The methods of grouping composite fuzzy numbers of a triangular type are investigated. Methods of solving the problem of clustering states of fuzzy systems, which are described by a set of composite fuzzy numbers, are developed. The conditions of their constructive application in the grouping problems are considered.

The method of choosing solutions under the conditions of presentation of the initial data in the form of a set of composite fuzzy numbers and in the presence of a set of criteria, the set of values of which are given in numerical intervals with linear membership functions are considered.

Examples of practical problems are given, in the solution of which the uncertainty of data is taken into account. The variant of implementation of the genetic algorithm for solving the scheduling problem with the consideration of teachers' wishes regarding the time of conducting training sessions is offered. The process of distribution of network resources in a two-level system of Internet access is formalized based on the use of a continuous-discrete model and the description of data in the form of composite fuzzy numbers. A mathematical model for determining the optimal distribution of capacities of data transmission channels in a local network is constructed, algorithms are proposed for solving the problem of distribution of channel capacities. The solution of the task of grouping users by the levels of consumption of network capacities and the corresponding task of efficient allocation of resources of channels of data transmission among the user groups served.

Key words: composite fuzzy numbers, fuzzy triangle numbers, clusterization task, genetic algorithm, dynamic clausterization, network capacity, model of production and transportaion problem, resource distribution, a two-level optimization task.