КІБЕРНЕТИКА
та КОМП'ЮТЕРНІ
ТЕХНОЛОГІЇ

Досліджуються проблема адекватності та питання введення метрик і псевдометрик у нечіткому кластерному аналізі. Запропоновано аксіоматичне визначення нечітких кластерів на основі нечіткого відношення схожості на базі коефіцієнта лінгвістичної кореляції. Введено поняття нечіткого кластера рівня α. Визначено відстань між нечіткими кластерами рівня α з використанням порогової конорми. Запропонований підхід може бути основою для розробки алгоритмів розв'язування задач кластерного аналізу та обґрунтування змістовної інтерпретації результатів досліджень.

***Ключові слова**: нечітка множина, конорма, метрика, псевдометрика, нечітке відношення схожості, нечіткий кластер.*

УДК 519.8 DOI:10.34229/2707-451X.23.1.3

І.І. РЯСНА

НЕЧІТКИЙ КЛАСТЕРНИЙ АНАЛІЗ: ПСЕВДОМЕТРИКИ ТА НЕЧІТКІ КЛАСТЕРИ

Вступ. Нечіткий кластерний аналіз ґрунтується на визначенні нечітких відношень схожості та несхожості. При цьому значна увага приділяється питанням введення метрик і псевдометрик для нечітких множин та визначенню відстаней між формально введеними нечіткими кластерами [1–5]. Огляд підходів до побудови мір схожості для інтервальних нечітких множин типу 2 і типу 1 наведено у роботах [6, 7], у яких викладено методології, що ґрунтуються на нечіткій логіці, а об'єктами досліджень є слова і речення природної мови (computing with words). Для нечітких інтуїціонівістських множин питання введення метрик розглядаються в [8]. Огляд підходів до формалізації задач нечіткого кластерного аналізу та найбільш повний перелік методів та алгоритмів їхнього розв'язування наведено у [9]; розглядаються як випадки нечітких даних, так і розробка нечітких методів та алгоритмів для розв'язування задач з чіткими даними. Використанню кластерного аналізу в інтелектуальних системах присвячена монографія [10], у якій ґрунтовно викладені питання застосування чітких та нечітких алгоритмів для задач нечіткого кластерного аналізу.

У випадках, коли відсутні вихідні дані, достатні для статистичного аналізу або використовується інформація, отримана від експертів, пропонуються нечіткі моделі задач, що ураховують різні види невизначеності та більш аргументовано відображають реальні ситуації у системах логістики, моніторингу, спостереження, виконанні завдань з обслуговування тощо.

Прикладом можуть бути задачі кластерного аналізу, які виникають при формалізації та розв'язуванні задач маршрутизації груп безпілотних літальних апаратів (БПЛА). Розвиток технологій розробки та використання БПЛА продовжує здійснювати суттєві зміни у військовій тактиці та стратегії. Наразі набутий оперативний досвід демонструє великі перспективи, а використання БПЛА зробить військові сили більш ефективними, зменшить витрати та ризики [11, 12].

Застосування команд безпілотників як технології породжує багато питань, які привертають увагу дослідників.

Задачі побудови кластерів БПЛА з прив’язкою до місць розташування чи їхнього обслуговування часто виникають як складові сучасних технологій при оптимізації маршрутів груп БПЛА. Так, у дослідженні [13] розміщення базової станції БПЛА проводиться за допомогою методів Fuzzy C-Means and Fuzzy C-Means – Center of Gravity відповідно до імовірнісних вимог у регіонах, де мережа зв'язку перервана. У роботі [14] запропонована нова модель для кластеризації місць доставки та варіантів маршрутів: команда однотипних безпілотників перевозиться однією вантажівкою до координаційних точок для задоволення вимог клієнтів. Автори розглянули дві різні політики для визначення координаційних точок: 1) обмеження місць зупинки вантажівок розташуванням клієнтів; 2) дозвіл місць зупинки будь-де в регіоні доставки.

Нечітка логіка використовується у різноманітних алгоритмах через її властивість відігравати роль універсального апроксиматора різних властивостей та нелінійних характеристик. Однак потрібно здійснити багато спроб, щоб отримати найкращий набір функцій належності та бази правил, які ефективно працюватимуть для певного застосування. Цей процес можна спростити, використовуючи евристичний алгоритм пошуку, такий як генетичний алгоритм. В роботі [15] нечітка логіка у генетичному алгоритмі застосована для призначення завдань взаємодіючим БПЛА, класифікованим як полігон-задача відвідування кількох комівояжерів. Ця задача має багато застосувань, у тому числі в задачі маршрутизації зграї БПЛА. Пропонується метод нечіткої кластеризації для генетичного алгоритму, який є специфічним для розглядуваної задачі та більш ефективним порівняно з кластеризацією *k*-середніх і *c*-середніх. Наведено два різні алгоритми, засновані на нечіткій логіці для генетичного алгоритму: один оцінює відстань, яку подолав кожний БПЛА, щоб кластеризувати простір пошуку, а інший використовує функцію вартості, яка є апроксиматором для відстані, що призводить до скорочення часу обчислення. Два підходи порівнюються один з одним. Обговорюється можливість масштабування алгоритму для збільшення кількості цілей. Результати порівнюються для маленьких і великих полігонів-багатокутників.

В розглянутих роботах оцінки в основному ґрунтуються на нечіткій логіці, яка ураховує фактори, які отримуються з різних джерел. Проте, багатовимірні випадки визначення параметрів і характеристик об’єктів за наявності їхніх вимірювань за різними шкалами не розглядаються. Розглянемо ці питання.

Задача кластеризації неформально формулюється так [16]: згрупувати об’єкти із заданої множини в підмножини (які називаються кластерами) так, щоб більш схожі за певними характеристиками об’єкти відносилися до однієї і тієї ж підмножини, а менш схожі – до різних підмножин. Схожість визначається як бінарне рефлексивне та симетричне відношення на заданій множині об’єктів X. Однак, результатом процедури кластеризації має бути розбиття множини об’єктів на підмножини, що не перетинаються, тобто класів еквівалентності. У роботі [16] показано, що такий підхід базується на встановленні ізоморфізму неізоморфних структур.

Для розв'язування задач кластерного аналізу часто використовуються методи нечіткої кластеризації, складовою яких є операція транзитивного замикання матриці схожості заданої множини об'єктів [17, 18]. Результатом цієї операції є нечітке відношення еквівалентності. Такий підхід або його модифікації використовуються як у випадках прямої експертної оцінки схожості у випадку, коли двоїсте відношення несхожості не є метрикою [18, 19], так і в тих випадках, коли початкові дані подані матрицею відстаней між об'єктами [20].

Відомо, що доповнення нечіткого відношення еквівалентності є ультраметрикою, специфіку якої неможливо змістовно проінтерпретувати як властивість отримуваних класів еквівалентності.

Визначення поняття нечітких кластерів. Будемо ототожнювати нечіткі множини з їхніми функціями належності.

Нехай на скінченній множині *X* задано нечітке відношення схожості $\tau : X \times X \to [0,1]$ – нечітке рефлексивне та симетричне відношення: $\forall x \in X\ \tau(x,x) = 1$; $\forall x, y \in X \quad \tau(x,y) = \tau(y,x)$, $\tau(x,y) \in [0,1]$. Матрицю, яка відповідає нечіткому відношенню τ позначимо R_τ.

Означення 1. C_x-кластером назвемо нечітку підмножину $C_x = \{(y, c_x(y))\}$ множини *X*, функція належності якої визначається так: $c_x(y) = \tau(x,y)$, $x, y \in X$.

Функція належності кластера C_x – це відображення $c_x : X \to [0,1]$, при якому $c_x(x) = \tau(x,x) = 1$. Елемент $x \in X$ назвемо центром C_x-кластера. У матричному поданні $c_x(y)$ – це рядок матриці схожості R_τ, який визначає схожість елемента $x \in X$ з усіма елементами $y \in X$.

Позначимо $\mathbf{C} = \{C_x | x \in X\}$. Оскільки вважаємо, що всі елементи множини X різні (не тотожні), то не тотожними є і елементи множини кластерів **C**, тому $|X| = |\mathbf{C}|$; між цими множинами можна визначити бієктивне відображення. Вочевидь, не тотожні кластери $C_x, C_y \in \mathbf{C}$ можуть співпадати або не співпадати.

Ядром $Ker\, C_x$ нечіткої підмножини C_x є чітка підмножина множини X така, що

$$Ker\, C_x = \{y \mid y \in X, c_x(y) = 1\}.$$

Носій підмножини C_x визначається так:

$$Supp\ C_x = \{y \mid y \in X, c_x(y) > 0\}.$$

Оскільки $\tau(x,y)$ – нечітке відношення схожості, то $\rho(x,y) = 1 - \tau(x,y)$, $(x,y) \in X \times X$, називається нечітким відношенням несхожості (або відмінності) на $X \times X$ [21]. Для будь-якого нечіткого відношення несхожості виконуються такі умови: $\forall x, y \in X\ \rho(x,y) \in [0,1]$; $\forall x \in X\ \rho(x,x) = 0$ (антирефлексивність); $\forall x, y \in X\ \rho(x,y) = \rho(y,x)$ (симетричність).

Визначимо особливості належності та неналежності нечіткому кластеру ($y \in C_x$, $y \notin C_x$):

$$y \in C_x \Leftrightarrow y \in Supp\, C_x \Leftrightarrow c_x(y) > 0 \Leftrightarrow \rho(x,y) < 1;$$

$$y \in Ker\, C_x \Leftrightarrow c_x(y) = 1 \Leftrightarrow \rho(x,y) = 0;$$

$$y \notin C_x \Leftrightarrow c_x(y) = 0 \Leftrightarrow \rho(x,y) = 1.$$

Тоді, $\forall y \in C_x\ \&\ y \notin Ker\, C_x$ маємо $0 < c_x(y) < 1$, $0 < \rho(x,y) < 1$.

Розглянемо умови, за яких відношення несхожості визначає метрику або псевдометрику.

Метрикою називається числова функція $d(x,y)$, яка визначена на декартовому добутку $X \times X$ та задовольняє трьом аксіомам:

1) $d(x,y) = 0 \Leftrightarrow x = y$ (аксіома тотожності);
2) $d(x,y) = d(y,x)$ (аксіома симетрії);
3) $d(x,y) \le d(x,z) + d(z,y)$ (нерівність трикутника).

З цих аксіом випливає, що $d(x,y) \ge 0$.

Зауважимо, що $x = y$ означає, що *x* є тотожним *y*, а $x \ne y$ означає, що елементи *x*, *y* є різними (не тотожними).

Значення функції $d(x,y)$ для фіксованих елементів упорядкованої пари (x,y) називають також відстанню між цими елементами.

Упорядковану послідовність m елементів $l_m = \left(x_1, ..., x_m\right)$ називають шляхом, який проходить через ці елементи [22]. Довжина цього шляху визначається так:

$$D\left(l_m\right) = \sum_{i=1}^{m-1} d\left(x_i, x_{i+1}\right),$$

де $d\left(x_i, x_{i+1}\right)$ – відстань між елементами x_i, x_{i+1}.

Якщо існують такі елементи $x \neq y$, що $d(x, y) = 0$, то функція $d(x, y)$ називається псевдометрикою. Для псевдометрики імплікація $d(x, y) = 0 \Rightarrow x = y$ не виконується, а перша аксіома має вигляд $d(x, x) = 0$.

Псевдометрика визначає метрику на множині класів еквівалентності (фактор-множині) $X / \sim$, де «$\sim$» – відношення еквівалентності: $x \sim y \Leftrightarrow\ d\left(x, y\right) = 0.$

Відношення несхожості за означенням задовольняє першим двом аксіомам псевдометрики: $\forall x \in X\ \rho(x, x) = 0,\ \ \forall x, y \in X\ \ \rho(x, y) = \rho(y, x)$.

Нерівність трикутника повинна виконуватися як для метрики, так і для псевдометрики:

$$\rho(x, z) \leq \rho(x, y) + \rho(y, z) \quad \forall x, y, z \in X\ , \tag{1}$$

а виконання умови (1) є еквівалентним виконанню нерівності

$$\left|\rho(x, z) - \rho(y, z)\right| \leq \rho(x, y) \quad \forall x, y, z \in X\ . \tag{2}$$

Вочевидь, для випадку $\rho(x, z) - \rho(y, z) > 0$ нерівність (2) виконується. Для випадку $\rho(x, z) - \rho(y, z) < 0$ нерівність (2) також виконується, тому що для протилежного випадку, а саме, $\left|\rho(x, z) - \rho(y, z)\right| = \rho(y, z) - \rho(x, z) > \rho(x, y),$ нерівність трикутника не буде виконуватись: $\rho(y, z) >$ $> \rho(x, y) + \rho(x, z) = \rho(y, x) + \rho(x, z).$ Переходячи до відношення схожості отримаємо умову виконання нерівності трикутника:

$$\left|\tau\left(x, z\right) - \tau(y, z)\right| \leq 1 - \tau(x, y) \quad \forall x, y, z \in X. \tag{3}$$

Відношення схожості, для якого виконується умова (3), названо в [16] відношенням подібності. У подальшому, дотримуючись позначення, яке введено в [16], нечіткі кластери, що породжуються відношенням подібності називатимемо *R*-кластерами або просто кластерами, для функцій належності яких, згідно (3), має місце така нерівність:

$$\left|c_x\left(z\right) - c_y(z)\right| \leq 1 - \tau(x, y) \quad \forall x, y, z \in X\ . \tag{4}$$

Покажемо, що в залежності від структури кластерів доповнення відношення схожості визначатиме метрику або псевдометрику.

Лема 1. Якщо ядра різних кластерів $C_x, C_y \in \mathbf{C},\ \ x, y \in X$, перетинаються, то функції належності цих кластерів співпадають, а відношення $\rho(\cdot,\cdot) = 1 - \tau(\cdot,\cdot)$ визначає псевдометрику.

Доведення. Якщо $Ker\, C_x \bigcap Ker\, C_y \neq \varnothing$, то існує $y \neq x$ (y, x – не тотожні), для якого $c_x\left(y\right) = 1,$ тоді $\tau\left(x, y\right) = c_x\left(y\right) = 1,$ а з нерівності (4) $\left|c_x\left(z\right) - c_y(z)\right| \leq 1 - \tau(x, y)$ випливає $\left|c_x(z) - c_y(z)\right| = 0$, тобто

$$c_x(z) = c_y(z) \quad \forall z \in X\ .$$

У цьому випадку $\rho\Big(\left(x, y\right) | \, x \in C_x, y \in C_y\Big) = 0,$ а $y \neq x$ ($y,\ x$ не є тотожними), тобто $\rho\left(x, y\right)$ є псевдометрикою.

Лема доведена.

Лема 2. Якщо ядра різних кластерів $C_x, C_y \in \mathbf{C}$, $x, y \in X$, попарно не перетинаються, то відношення $\rho(x,y)$ є метрикою.

Доведення. Якщо $Ker\, C_x \cap Ker\, C_y = \varnothing,\ \ y \neq x,$ то $c_x(y) \neq 1.$ Отже, $\rho(x,y) = 1 - c_x(y) > 0$, тому виконується аксіома тотожності: $\rho(x,y) = 0$ тоді і тільки тоді, коли $x = y$, тобто $\rho(x,y)$ є метрикою.

Лема доведена.

Надалі вважаємо, що $\rho(x,y)$ є псевдометрикою, оскільки в загальному випадку не можна виключати можливість перетину ядер різних кластерів.

Означення 2. Нечітким відношенням схожості рівня α, $\alpha \in (0,1]$ назвемо нечітке відношення схожості $\tau^{(\alpha)}(x,y)$, функція належності якого $\forall x, y \in X$ задовольняє таким умовам:

$$\tau^{(\alpha)}(x,y) = \begin{cases} \tau(x,y), & \text{якщо } \tau(x,y) \geq \alpha, \\ 0, & \text{якщо } \tau(x,y) < \alpha. \end{cases} \tag{5}$$

Вочевидь, що $\tau^{(\alpha)}(x,y) \leq \tau(x,y)\ \ \forall x, y \in X$.

Означення 3. $C_x^{(\alpha)}$-кластером назвемо нечітку підмножину $C_x^{(\alpha)} = \left\{\left(y, c_x^{(\alpha)}(y)\right)\right\}$ множини X, $\alpha \in (0,1]$, функція належності якої визначається так: $c_x^{(\alpha)}(y) = \tau^{(\alpha)}(x,y)$, $x, y \in X$.

Елемент $x \in X$ називатимемо центром або ідентифікатором $C_x^{(\alpha)}$-кластера. Позначимо множину таких кластерів $\mathbf{C}^{(\alpha)} = \left\{C_x^{(\alpha)} \middle| x \in X\right\}$. Нехай $\delta = 1 - \alpha$.

Зауважимо, що $Supp\, C_x^{(\alpha)} \subseteq Supp\, C_x$, а також

$$y \in C_x^{(\alpha)} \Leftrightarrow y \in Supp\, C_x^{(\alpha)} \Leftrightarrow c_x^{(\alpha)}(y) \geq \alpha \Leftrightarrow \rho(x,y) \leq \delta = 1 - \alpha\ ;$$

$$y \in Ker\, C_x^{(\alpha)} \Leftrightarrow c_x^{(\alpha)}(y) = 1 \Leftrightarrow \rho(x,y) = 0\,;$$

$$y \notin C_x^{(\alpha)} \Leftrightarrow c_x^{(\alpha)}(y) = 0 \Leftrightarrow \rho(x,y) > \delta\,;$$

$$\forall y \in C_x^{(\alpha)}\ \&\ y \notin Ker\, C_x^{(\alpha)},\ \alpha \leq c_x^{(\alpha)}(y) < 1,\ 0 < \rho(x,y) \leq \delta\,.$$

Означення 4. Нечітким відношенням несхожості рівня δ, $\delta \in [0,1)$, назвемо нечітке відношення несхожості $\rho^{(\delta)}(x,y)$, функція належності якого $\forall x, y \in X$ задовольняє умовам

$$\rho^{(\delta)}(x,y) = \begin{cases} \rho(x,y), & \text{якщо } \rho(x,y) \leq \delta, \\ 1, & \text{якщо } \rho(x,y) > \delta. \end{cases} \tag{6}$$

Тоді, $\rho^{(\delta)}(x,x) = 0$, $\rho^{(\delta)}(x,y) = \rho^{(\delta)}(y,x)$, $\rho^{(\delta)}(x,y) \geq \rho(x,y)$.

Оскільки $\delta = 1 - \alpha,$ то згідно (5), (6) отримуємо, що $\tau^{(\alpha)}(x,y) + \rho^{(\delta)}(x,y) = 1.$

Порогові трикутні конорми. Конорма за Я. Лукасевичем визначається так [21]:

$$S_L(u,v) = \min(1, u+v), \tag{7}$$

де $u, v \in [0,1]$, $S_L(u,v) \in [0,1]$.

Формулу (7) перепишемо як $S_L(u,v) = \begin{cases} u+v, & \text{якщо } (u+v) \le 1, \\ 1, & \text{якщо } (u+v) > 1. \end{cases}$

Розвиваючи цей підхід, введемо порогову трикутну конорму так:

$$S_L^{(\delta)}(u,v) = \begin{cases} u+v, & \text{якщо } u+v \le \delta, \\ 1, & \text{якщо } u+v > \delta. \end{cases} \tag{8}$$

Нехай $\rho(x,y)$ – інваріантна псевдометрика, яку визначено за допомогою коефіцієнта лінгвістичної кореляції [23], який введено для визначення нечіткої міри схожості об'єктів, що мають як кількісні, так і якісні характеристики. Значення цих характеристик можуть бути отримані з різних джерел і визначені за різними шкалами (за класифікацією С. Стівенса).

Визначення S_L-псевдометрики на $X \times X$.

Теорема 1. Функція

$$\hat{\rho}(x,y) = \min_{z \in X} S_L(\rho(x,z), \rho(z,y)), \quad x, y \in X, \tag{9}$$

визначає обмежену псевдометрику на $X \times X$, де $S_L(a,b) = \min(1, a+b)$ – конорма за Лукасевичем.

Доведення. Згідно (7)

$$S_L(\rho(x,z), \rho(z,y)) = \min(1, \rho(x,z) + \rho(z,y))$$

є довжиною шляху, що визначається елементами $x, z, y \in X$.

Так як $\rho(y,y) = 0$, то $S_L(\rho(x,y), \rho(y,y)) = \rho(x,y) = \hat{\rho}(x,y)$ – відстань між двома елементами x, y.

Покажемо, що функція (9) визначає псевдометрику, тобто, що для неї виконується нерівність (аналог нерівності трикутника):

$$S_L(\rho(x,y), \rho(y,y)) \le S_L(\rho(x,z), \rho(z,y)) \quad \forall x, y, z \in X. \tag{10}$$

Так як $\forall x, y, z \in X \quad \rho(x,y) \le \rho(x,z) + \rho(z,y)$, то враховуючи, що $S_L(a,b) = \min(1, a+b)$, отримаємо

$$\min(1, \rho(x,y) + 0) = \rho(x,y) \le \min(1, \rho(x,z) + \rho(z,y)) \quad \forall x, y, z \in X.$$

Отже, виконується аналог нерівності трикутника (10): відстань між двома елементами x, y не більше довжини шляху через елементи x, z, y, тобто виконується третя аксіома псевдометрики.

Оскільки $\forall x, y \in X \ \hat{\rho}(x,y) = \rho(x,y)$, $\rho(x,z) = \rho(z,x)$, $\rho(z,y) = \rho(y,z)$, то $\hat{\rho}(x,y) = \hat{\rho}(y,x)$.

Отже, виконується аксіома симетричності (друга аксіома псевдометрики).

Перша аксіома псевдометрики також виконується, так як $\rho(x,x) = 0$, а $\hat{\rho}(x,x) = \rho(x,x)$, отже, $\hat{\rho}(x,x) = 0$, за означенням $\forall x, y, z \in X \ S_L(\rho(x,z), \rho(z,y)) \le 1$, тоді $\hat{\rho}(x,y) \le 1$.

Отже, $\hat{\rho}(x,y) = \min_{z \in X} S_L(\rho(x,z), \rho(z,y))$ є обмеженою псевдометрикою.

Теорема доведена.

Означення 7. Псевдометрику $\hat{\rho}(x, y)$ назвемо S_L-псевдометрикою на $X \times X$.

Наведемо такий приклад. Нехай $\rho(x,y) = 0{,}8$, $\rho(x,z) = 0{,}5$, $\rho(z,y) = 0{,}4$; $\delta = 0{,}7$;

тоді $\rho^{(\delta)}(x,y) = 1$, $\rho^{(\delta)}(x,z) = 0{,}5$, $\rho^{(\delta)}(z,y) = 0{,}4$, та $S_L(\rho(x,z), \rho(z,y)) = 0{,}9 < \rho^{(\delta)}(x,y) = 1$.

Таким чином, при застосуванні S_L-псевдометрики нерівність трикутника для відношення несхожості рівня δ не виконується.

Визначення $S_L^{(\delta)}$-псевдометрики на $X \times X$.

Теорема 2. Функція

$$\hat{\rho}^{(\delta)}(x,y) = \min_{z \in X} S_L^{(\delta)}\left(\rho^{(\delta)}(x,z), \rho^{(\delta)}(z,y)\right), \quad x,y \in X, \tag{11}$$

визначає обмежену псевдометрику на $X \times X$, де $S_L^{(\delta)}(\cdot,\cdot)$ – порогова конорма.

Доведення.

1. $S_L^{(\delta)}(\rho(x,z),\rho(z,y))$ – довжина шляху, що визначається елементами x,z,y , і за формулою (8) маємо:

а) якщо $\rho^{(\delta)}(x,z)+\rho^{(\delta)}(z,y) \le \delta$, то $S_L^{(\delta)}(\cdot,\cdot) = \rho^{(\delta)}(x,z)+\rho^{(\delta)}(z,y)$;

б) якщо $\rho^{(\delta)}(x,z)+\rho^{(\delta)}(z,y) > \delta$, то $S_L^{(\delta)}(\cdot,\cdot) = 1$.

Крім того, відстань між елементами $x,y \in X$ дорівнює:

$$S_L^{(\delta)}\left(\rho^{(\delta)}(x,y), \rho^{(\delta)}(y,y)\right) = S_L^{(\delta)}\left(\rho^{(\delta)}(x,y), 0\right) = \rho^{(\delta)}(x,y) = \hat{\rho}^{(\delta)}(x,y). \tag{12}$$

Покажемо, що для псевдометрики $S_L^{(\delta)}(\cdot,\cdot)$ виконується нерівність (аналог нерівності трикутника):

$$S_L^{(\delta)}\left(\rho^{(\delta)}(x,y), \rho^{(\delta)}(y,y)\right) \le S_L^{(\delta)}\left(\rho^{(\delta)}(x,z), \rho^{(\delta)}(z,y)\right), \quad x,y,z \in X. \tag{13}$$

Згідно (6) $\rho^{(\delta)}(x,y) \ge \rho(x,y)$, $\rho^{(\delta)}(x,z) \ge \rho(x,z)$, $\rho^{(\delta)}(z,y) \ge \rho(z,y)$.

Розглянемо такі можливі випадки.

А). Якщо $\rho(x,z)+\rho(z,y) > \delta$, то згідно (6) $\rho^{(\delta)}(x,z)+\rho^{(\delta)}(z,y) > \delta$, а згідно (8) $S_L^{(\delta)}\left(\rho^{(\delta)}(x,z), \rho^{(\delta)}(z,y)\right) = 1.$

При цьому, якщо $\rho(x,y) \le \delta,$ то згідно з (6) $\rho^{(\delta)}(x,y) = \rho(x,y),$ $\rho^{(\delta)}(x,y) < 1$, тому згідно з (8) $S_L^{(\delta)}\left(\rho^{(\delta)}(x,y), \rho^{(\delta)}(y,y)\right) < 1,$ отже нерівність (13) виконується. Якщо $\rho(x,y) > \delta,$ то згідно з (6) $\rho^{(\delta)}(x,y) = 1,$ $S_L^{(\delta)}\left(\rho^{(\delta)}(x,y), \rho^{(\delta)}(y,y)\right) = 1$, тобто нерівність (13) також виконується.

Б). Якщо $\rho(x,z)+\rho(z,y) \le \delta,$ то $\rho(x,z) \le \delta,$ $\rho(z,y) \le \delta.$ Тоді з (6) випливає $\rho^{(\delta)}(x,z) = \rho(x,z)$, $\rho^{(\delta)}(z,y) = \rho(z,y)$.

Так як $\rho(x,z)+\rho(z,y) \le \delta,$ то з нерівності $\rho(x,y) \le \rho(x,z)+\rho(z,y)$ випливає $\rho(x,y) \le \delta,$ тому згідно з (6) $\rho^{(\delta)}(x,y) = \rho(x,y).$ Отже, згідно з (8) з урахуванням того, що $\rho^{(\delta)}(x,z) = \rho(x,z),$ $\rho^{(\delta)}(z,y) = \rho(z,y)$ маємо:

$$S_L^{(\delta)}\left(\rho^{(\delta)}(x,z),\rho^{(\delta)}(z,y)\right)=\rho(x,z)+\rho(z,y)\geq\rho(x,y)=\rho^{(\delta)}(x,y),$$

а так як $\rho^{(\delta)}(x,y)=S_L^{(\delta)}\left(\rho^{(\delta)}(x,y),\rho^{(\delta)}(y,y)\right)$, то

$$S_L^{(\delta)}\left(\rho^{(\delta)}(x,y),\rho^{(\delta)}(y,y)\right)\leq S_L^{(\delta)}\left(\rho^{(\delta)}(x,z),\rho^{(\delta)}(z,y)\right),$$

отже, нерівність (13) також виконується.

Таким чином доведено, що аналог нерівності трикутника завжди виконується (третя аксіома псевдометрики).

2. Так як $\forall x,y\in X$ $\hat{\rho}^{(\delta)}(x,y)=\rho^{(\delta)}(x,y)$, згідно з (6) $\rho^{(\delta)}(x,y)=\rho^{(\delta)}(y,x)$, $\rho^{(\delta)}(x,z)=\rho^{(\delta)}(z,x)$, то $\hat{\rho}^{(\delta)}(x,y)=\hat{\rho}^{(\delta)}(y,x)$.

Тобто виконується друга аксіома псевдометрики (симетричність).

3. Так як за означенням $\rho^{(\delta)}(x,x)=0$, а $\hat{\rho}^{(\delta)}(x,x)=\rho^{(\delta)}(x,x)$, то $\hat{\rho}^{(\delta)}(x,x)=0$, тобто виконується перша аксіома псевдометрики.

4. За означенням $\forall x,y,z\in X$ $S_L^{(\delta)}\left(\rho^{(\delta)}(x,z),\rho^{(\delta)}(z,y)\right)\leq 1$.

Отже,

$$\hat{\rho}^{(\delta)}(x,y)=\min_{z\in X}S_L^{(\delta)}\left(\rho^{(\delta)}(x,z),\rho^{(\delta)}(z,y)\right)$$

обмежена псевдометрика.

Теорема доведена.

Означення 8. Псевдометрику $\hat{\rho}^{(\delta)}(x,y)$, що визначається за пороговою трикутною конормою $S_L^{(\delta)}(\cdot,\cdot)$, називатимемо $S_L^{(\delta)}$-псевдометрикою на $X\times X$.

Визначення відстані між нечіткими кластерами рівня α. Відстань між кластерами $C_x^{(\alpha)},C_y^{(\alpha)}\in\mathbf{C}^{(\alpha)}$ будемо ототожнювати з мінімальною відстанню між елементами цих кластерів:

$$d\left(C_x^{(\alpha)},C_y^{(\alpha)}\right)=\min_{u\in X,v\in X}S_L^{(\delta)}\left(\rho^{(\delta)}(x,v),\rho^{(\delta)}(y,u)\right),$$

де, згідно теореми 2, $\rho^{(\delta)}(x,v)=1-c_x^{(\alpha)}(v)$ – відстань між центром x та довільним елементом $\left(v,c_x^{(\alpha)}(v)\right)$ кластера $C_x^{(\alpha)}$, і $\rho^{(\delta)}(y,u)=1-c_y^{(\alpha)}(u)$ – відстань між центром y та довільним елементом $\left(u,c_y^{(\alpha)}(u)\right)$ кластера $C_y^{(\alpha)}$.

Теорема 3. За $S_L^{(\delta)}$-псевдометрикою відстань між кластерами $C_x^{(\alpha)},C_y^{(\alpha)}\in\mathbf{C}^{(\alpha)}$ дорівнює відстані $\rho^{(\delta)}(x,y)$ між центрами x,y цих кластерів.

Доведення. Покажемо, що відстань між центрами x,y кластерів $C_x^{(\alpha)},C_y^{(\alpha)}\in\mathbf{C}$ не більше відстані між довільними елементами, відповідно, кластерів $C_x^{(\alpha)},C_y^{(\alpha)}\in\mathbf{C}$ за $S_L^{(\delta)}$-псевдометрикою.

А). Нехай $x, y \in X$ – центри кластерів $C_x^{(\alpha)}$, $C_y^{(\alpha)}$, відповідно, тоді $\rho^{(\delta)}(x,y) = 1 - c_x^{(\alpha)}(y)$, $\rho^{(\delta)}(y,x) = 1 - c_y^{(\alpha)}(x)$ і згідно з теоремою 2

$$S_L^{(\delta)}\left(\rho^{(\delta)}(x,y), \rho^{(\delta)}(y,y)\right) = \rho^{(\delta)}(x,y),$$

де $\rho^{(\delta)}(x,y) = \rho^{(\delta)}(y,x)$ – відстань між центрами кластерів $C_x^{(\alpha)}$, $C_y^{(\alpha)}$.

Згідно (6), якщо $\rho(x,y) \le \delta$, то $\rho^{(\delta)}(x,y) = \rho(x,y) \le \delta$; якщо $\rho(x,y) > \delta$, то $\rho^{(\delta)}(x,y) = 1$.

Б). Для шляху між елементами $\left(z, c_x^{(\alpha)}(z)\right) \in C_x^{(\alpha)}$ та $\left(z, c_y^{(\alpha)}(z)\right) \in C_y^{(\alpha)}$, $x, z, y \in X$, на основі теореми 2 отримаємо:

$$\rho^{(\delta)}(x,y) \le S_L^{(\delta)}\left(\rho^{(\delta)}(x,z), \rho^{(\delta)}(z,y)\right),$$

де $\rho^{(\delta)}(x,z)$ – відстань між центром *x* та елементом $\left(z, c_x^{(\alpha)}(z)\right)$ кластера $C_x^{(\alpha)}$, $\rho^{(\delta)}(z,y)$ – відстань між центром *y* та елементом $\left(z, c_y^{(\alpha)}(z)\right)$ кластера $C_y^{(\alpha)}$.

В). Для шляху між елементами $x, z, u, y \in X$ виконується така нерівність (чотирикутника):

$$\rho(x,y) \le \rho(x,z) + \rho(z,y) \le \rho(x,z) + \rho(z,u) + \rho(u,y),$$

де $\rho(z,u) + \rho(u,y)$ – довжина шляху між елементами $z, u, y \in X$.

Аналогічно, на основі теореми 2, за псевдометрикою $S_L^{(\delta)}(\cdot,\cdot)$ буде виконуватись нерівність:

$$\rho^{(\delta)}(x,y) \le S_L^{(\delta)}\left(\rho^{(\delta)}(x,z), \rho^{(\delta)}(z,y)\right) \le$$

$$\le S_L^{(\delta)}\left(\rho^{(\delta)}(x,z), S_L^{(\delta)}\left(\rho^{(\delta)}(z,u), S_L^{(\delta)}\left(\rho^{(\delta)}(u,y), \rho^{(\delta)}(y,y)\right)\right)\right),$$

де $S_L^{(\delta)}\left(\rho^{(\delta)}(z,u), S_L^{(\delta)}\left(\rho^{(\delta)}(u,y), \rho^{(\delta)}(y,y)\right)\right)$ – довжина шляху між центрами z, u, y, відповідно, кластерів $C_z^{(\alpha)}, C_u^{(\alpha)}, C_y^{(\alpha)} \in \mathbf{C}^{(\alpha)}$.

Отже, за $S_L^{(\delta)}$-псевдометрикою відстань між центрами x, y кластерів $C_x^{(\alpha)}, C_y^{(\alpha)} \in \mathbf{C}^{(\alpha)}$ не більше відстані між довільними елементами кластерів $C_x^{(\alpha)}, C_y^{(\alpha)} \in \mathbf{C}^{(\alpha)}$: $\left(z, c_x^{(\alpha)}(z)\right) \in C_x^{(\alpha)}$, $\left(u, c_y^{(\alpha)}(u)\right) \in C_y^{(\alpha)}$.

Теорема доведена.

Висновки. Аксіоматично введені поняття кластерів та нечітких кластерів рівня α, які визначені як нечіткі множини елементів, схожих з певними елементами заданої множини, при виконанні умови: відношення несхожості повинно бути інваріантною псевдометрикою. Виконання цієї умови забезпечується використанням коефіцієнта лінгвістичної кореляції при обчисленні нечітких відношень. На основі визначення нечіткого кластера рівня α та порогової конорми введено інваріантну $S_L^{(\delta)}$-псевдометрику та визначено відстань між нечіткими кластерами рівня α. При цьому, на відмінність від методів нечіткої кластеризації, складовою яких є операція транзитивного замикання матриці схожості заданої множини об'єктів, не спотворюються зв'язки між об'єктами, тому забезпечується прозорість інтерпретації отриманих кластерів та можливість уточнення результатів при подальших дослідженнях їхньої структури. Введені поняття можуть бути закладені в основу побудови алгоритмів нечіткої кластеризації при визначенні багатовимірних характеристик заданої множини об'єктів.

Список літератури

1. Balopoulos V., Hatzimichailidis A.G., Papadoupoulos B.K. Difference and similarity measures for fuzzy operators. *Information Sciences*. 2007. 177. P. 2336–2348. https://doi.org/10.1016/j.ins.2007.01.005
2. Chaudur B.B., Rosenfeld A. On a metric distance between fuzzy sets. *Pattern Recognition Letters.* 1996. 17. 11. P. 1157–1160. https://doi.org/10.1016/0167-8655(96)00077-3
3. Gardner Andrew, et al. Measuring distance between unordered sets of different sizes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 137–143. https://doi.org/10.1109/CVPR.2014.25
4. Kosub S. A note on the triangle inequality for the jaccard distance. CoRR, abs/1612.02696, 2016.
5. Williams J.S, Steele N. Difference, distance and similarity as a basis for fuzzy decision support based on prototypical decision classes. *Fuzzy Sets and Systems*. 2002. 131. P. 35–46. http://dx.doi.org/10.1016/S0165-0114(01)00253-6
6. Wu D. and Mendel J.M. A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets. *Information Sciences.* 2009. **179** (8). P. 1169–1192. https://doi.org/10.1016/j.ins.2008.12.010
7. Wu D. and Mendel J.M. A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets. *Information Sciences.* 2008. **178** (2). P. 381–402. https://doi.org/10.1016/j.ins.2008.12.010
8. Szmidt E., Kacprzyk J. Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems.* 2000. 114. P. 505–518. https://doi.org/10.1016/S0165-0114(98)00244-9
9. D'Urso P., Gil M.Á. Fuzzy data analysis and classification. *Adv. Data Anal. Classif.* 2017. 11. P. 645–657. https://doi.org/10.1007/s11634-017-0304-z
10. Zgurovsky M.Z., Zaychenko Y.P. The Fundamentals of Computational Intelligence: System Approach. Springer: Switzerland, 2016. 375 p.
11. Яснопольська В. Розвідка, перехоплення цілей і керування вогнем: які безпілотники задіяно у війні в Україні. https://fakty.com.ua/ua/svit/20220428-osnovne-pryznachennya-rozvidka-najpopulyarnishi-modeli-bezpilotnykiv-v-ukrayini-ta-rosiyi/
12. Mumtaz Karatas, Ertan Yakıcı, Nasuh Razi Military Facility Location Problems: A Brief Survey. 2018. https://doi.org/10.4018/978-1-5225-5513-1.ch001
13. Sevdik G., Esnaf S., Baytürk E. Facility Location for Unmanned Aerial Vehicle Base Stations to Provide Uninterrupted Mobile Communication After Earthquakes. In: Durakbasa, N.M., Gençyılmaz, M.G. (eds) Digital Conversion on the Way to Industry 4.0. ISPR 2020. Lecture Notes in Mechanical Engineering. Springer, Cham. https://doi.org/10.1007/978-3-030-62784-3_5
14. Salama M., Srinivas S. Joint optimization of customer location clustering and drone-based routing for last-mile deliveries. *Transportation Research Part C: Emerging Technologies.* 2020. Vol. 114. P. 620–642. https://doi.org/10.1016/j.trc.2020.01.019
15. Ernest N., Sathyan A., Cohen K. Genetic Fuzzy Single and Collaborative Tasking for UAV Operations. In: *Multi-Rotor Platform-based UAV Systems.* P. 217–242. https://doi.org/10.1016/B978-1-78548-251-9.50011-X
16. Руспини Э.Г. Последние достижения в нечетком кластер-анализе. В кн. Нечеткие множества и теория возможностей. Последние достижения. М.: Радио и связь, 1986. С. 114–132.
17. Zadeh L.A. Similarity relations and fuzzy ordering. *Information Sciences.* 1971. Vol. 3. P. 177–200. https://doi.org/10.1016/S0020-0255(71)80005-1

18. Tamura S., Higuchi S., Tanaka K. Pattern classification based on fuzzy relations. *IEEE Trans. Systems, Man and Cybernetics.* 1971. v. SMC-1. P. 61–66. https://doi.org/10.1109/TSMC.1971.5408605
19. Yang M.-S., Shih H.-M. Cluster analysis based on fuzzy relations. *Fuzzy Sets and Systems.* 2001. Vol. 120. P. 197–212. https://doi.org/10.1016/S0165-0114(99)00146-3
20. Барсегян А.А. и др. Методы и модели анализа данных: OLAP и Data Mining. СПб: БХВ-Петербург, 2004. 336 с.
21. Нечеткие множества в моделях управления и искусственного интеллекта. М.: Наука, 1986. 312 с.
22. Кривий С.Л. Дискретна математика. Вибрані питання. Київ: Видавничий дім „Києво-Могилянська академія", 2007. 572 с.
23. Hulianytskyi L., Riasna I. On Fuzzy Similarity Relations for Heterogeneous Fuzzy Sets. *II International Scientific Symposium «Intelligent Solutions» IntSol-2021, September 28–30, 2021, Kyiv-Uzhhorod, Ukraine IntSol.* P. 48–59. https://ceur-ws.org/Vol-3018/Paper_5.pdf

Одержано 27.11.2022

Рясна Ірина Іванівна,
кандидат фізико-математичних наук, старший науковий співробітник
Інституту кібернетики імені В.М. Глушкова НАН України, Київ.
https://orcid.org/0000-0003-1370-3066

УДК 519.8

І. Рясна

Нечіткий кластерний аналіз: псевдометрики та нечіткі кластери

Інститут кібернетики імені В.М. Глушкова НАН України, Київ
Листування: riasn2080@gmail.com

Вступ. Задачі кластеризації виникають у різних сферах людської діяльності. У випадках, коли відсутні вихідні дані, достатні для статистичного аналізу або використовується інформація, отримана від експертів, пропонуються нечіткі моделі задач, що ураховують різні види невизначеності та більш аргументовано відображають реальні ситуації, які моделюють системи різного призначення. Особливу увагу привертають проблеми інваріантності у задачах з різнотипними даними, виміряними за різними шкалами за класифікацією С. Стівенса. Відомо, що при розв'язанні задач кластерного аналізу з використанням операції транзитивного замикання у відношенні еквівалентності, яке отримується, змінюються такі зв'язки між об'єктами, як схожість та несхожість. Тому, необхідно ураховувати проблему адекватності при розробці моделей та алгоритмів для розв'язання задач нечіткого кластерного аналізу.

Мета роботи. Провести аналіз проблеми адекватності результатів нечіткого кластерного аналізу щодо введення метрик і псевдометрик на нечітких множинах за наявності кількох якісних та кількісних характеристик об'єктів. Запропонувати підхід, що забезпечує адекватність псевдометрики, тобто забезпечує інваріантність відносно допустимих перетворень значень нечітких ознак, а також забезпечує розбиття об'єктів на класи еквівалентності без спотворення відстані між ними.

Результати. Запропоновано аксіоматичні визначення нечіткого кластера та нечіткого кластера рівня α, які введено як нечіткі множини елементів, схожих з певними елементами заданої множини, при виконанні умови: відношення несхожості повинно бути інваріантною псевдометрикою. Ця умова забезпечується використанням коефіцієнта лінгвістичної кореляції при обчисленні нечітких відношень схожості та несхожості. На основі визначення нечіткого кластера рівня α та порогової конорми визначено відстань між нечіткими кластерами рівня α.

Висновки. Запропонований підхід може бути основою для розробки алгоритмів розв'язання задач кластерного аналізу. При цьому забезпечується змістовна інтерпретація отриманих кластерів та можливість уточнення результатів при подальших дослідженнях їхньої структури.

Ключові слова: нечітка множина, конорма, метрика, псевдометрика, нечітке відношення схожості, нечіткий кластер.

UDC 519.8

Iryna Riasna

Fuzzy Cluster Analysis: Pseudometrics and Fuzzy Clusters

V.M. Glushkov Institute of Cybernetics of the NAS of Ukraine, Kyiv
Correspondence: riasn2080@gmail.com

Introduction. Clustering problems arise in various spheres of human activity. In cases where there are no initial data sufficient for statistical analysis or information obtained from experts is used, fuzzy models are proposed that take into account different types of uncertainty and more argumentatively reflect real situations that model systems of different purposes. Particular attention is drawn to invariance in problems with different types of data measured in different scales according to the classification of S. Stevens. It is known that when solving cluster analysis problems using the transitive closure operation with respect to the equivalence that is obtained, such connections between objects as similarity and dissimilarity are changed. Therefore, it is necessary to take into account the problem of adequacy when developing models and algorithms for solving problems of fuzzy cluster analysis.

The purpose of the paper is an analyzing the problem of adequacy of the results of fuzzy cluster analysis on the introduction of metrics and pseudometrics on fuzzy sets in the presence of several qualitative and quantitative characteristics of objects. Propose an approach that ensures the adequacy of pseudometrics, that is, provides invariance with respect to permissible transformations of the values of fuzzy features, and also ensures the division of objects into equivalence classes without distorting the distance between them.

Results. Axiomatic definitions of a fuzzy cluster and a fuzzy α level cluster are proposed, which are introduced as fuzzy sets of elements similar to certain elements of a given set, if the condition is met: the dissimilarity ratio must be an invariant pseudometric. This condition is ensured by the use of the linguistic correlation coefficient when calculating fuzzy relations of similarity and dissimilarity. Based on the definition of a fuzzy cluster of α level and threshold conorm, the distance between fuzzy clusters of α level is determined.

Conclusions. The proposed approach can be the basis for the development of algorithms for solving cluster analysis problems. This provides a meaningful interpretation of the obtained clusters, and the possibility of clarifying the results in further studies of their structure.

Keywords: fuzzy set, conorm, metric, pseudometric, fuzzy similarity relation, fuzzy cluster.