

Мулеца О. Ю.

АДАПТАЦІЯ МЕТОДУ НЕЧІТКИХ С-СЕРЕДНІХ ДО ЗАДАЧІ ВИЗНАЧЕННЯ СТРУКТУРИ СОЦІАЛЬНИХ ГРУП

Розглядається задача визначення структури соціальних груп як задача нечіткої кластеризації об'єктів. Пропонується адаптація методу нечітких с-середніх для розв'язання сформульованої задачі, яка полягає у визначенні нових способів обчислення відстані між об'єктами та встановлення центрів кластерів. Ефективність методу досягається шляхом представлення ознак об'єктів через відповідні лінгвістичні змінні.

Ключові слова: структура соціальної групи, нечітка кластеризація, лінгвістична змінна, метод нечітких с-середніх.

1. Вступ

Соціальна структура суспільства, як сукупність взаємопов'язаних та взаємодіючих між собою соціальних груп є об'єктом багатьох наукових досліджень [1, 2]. Вивчення її властивостей дозволяє розв'язувати важливі соціальні, політичні, економічні, медичні та інші задачі [1–3]. Основними елементами соціальної структури суспільства є соціальні групи осіб, об'єднаних за деякими спорідненими ознаками: етнічними, демографічними, професійними, територіальними тощо. Результати дослідження структури соціальних груп є основою для здійснення прогнозування виникнення та розвитку різних соціальних процесів, планування та виконання профілактичних і запобіжних дій до попередження виникнення негативних явищ у суспільстві загалом та в колективі зокрема, прогнозування розвитку ринків збуту товарів та отримання характеристики ринку трудових ресурсів тощо. Серед параметрів, що характеризують структуру соціальної групи є її кількісний та якісний склад. Звичайно, задачі визначення структури соціальної групи розв'язують шляхом аналізу результатів соціологічних досліджень, які, як правило, є високовартісними, ресурсомісткими та розтягнутими в часі. Проте, на практиці часто виникають задачі, пов'язані з визначенням кількісних характеристик таких груп, які характеризуються закритістю та невеликою чисельністю, наприклад, групи підвищеного ризику інфікування ВІЛ [3–5]. Виявлення та опитування представників цих груп є складною, а подекуди і не вирішуваною проблемою. Таким чином, актуальним є розробка підходів та методів визначення структури деяких соціальних груп, які б базувалися на використанні доступних даних таких як результати експертних опитувань.

Серед задач, які виникають в ході дослідження соціальної структури суспільства важливою є задача його поділу на соціальні групи, яка математично може бути представлена як задача кластеризації [6].

2. Аналіз літературних даних та постановка проблеми

Задача кластеризації полягає у визначенні груп об'єктів, що є найбільш близькими один до одного за

деяким критерієм [7–11]. Більшість методів кластеризації базуються на аналізі матриць коефіцієнтів подібності, в якості яких виступають, як правило відстані. Серед методів кластерного аналізу варто виокремити наступні:

- статистичні [12, 13], які базуються на припущенні, що кластери добре описуються деяким сімейством імовірнісних розподілів;
- деревовидні [14, 15], в результаті роботи яких відбувається побудова дерева кластерів;
- нейромережні [9, 16, 17], в основі яких лежить використання карт Кохонена;
- генетичні алгоритми [9, 11], які використовують принципи природного відбору, селекції та мутації.

Особливе місце серед методів кластеризації посідає метод нечітких с-середніх [18], який базується на апараті теорії нечітких множин [19]. В основі цього методу лежить положення про те, що кожен об'єкт може належати різним кластерам з різною ступенем належності.

Характерною особливістю задач визначення структури соціальних груп є те, що вхідні дані переважно мають нечисловий характер. Цей фактор ускладнює застосування класичних методів кластеризації, робить їх сильно залежними від правил переводу нечислових даних у числові та використовуваних метрик.

Іншою особливістю класу задач, що розглядається є те, що при виконанні розбиття групи населення на окремі соціальні групи необхідно брати до уваги мету розв'язання задачі, тобто природу таких груп, а, отже, по-різному враховувати ознаки соціально-демографічного портрету особи.

Також, важливим є те, що в ряді прикладних задач пов'язаних з дослідженням соціальних явищ неможливо отримати чіткі однозначні розв'язки. Повсякчас зустрічаються випадки, коли один і той же об'єкт може різною мірою належати до різних груп, що необхідно враховувати при розв'язанні таких задач.

Зазначені факти обумовлюють необхідність розвитку теорії нечіткого кластерного аналізу на клас задач, пов'язаних з дослідженням структури суспільства.

3. Об'єкт, ціль та задачі дослідження

Об'єктом дослідження є технології розв'язання задач визначення структури соціальних груп.

Метою дослідження є підвищення ефективності процесів прийняття рішень при розв'язанні задач, що виникають в управлінні, виробництві, медицині, шляхом розробки моделей та методів кластеризації в умовах невизначеності.

- В межах визначеної мети були поставлені такі задачі:
 - виконати аналіз задачі визначення структури соціальних груп та побудувати її вербальну постановку;
 - побудувати математичну модель задачі визначення структури соціальної групи як задачі кластеризації;
 - розробити метод нечіткої кластеризації для задачі визначення структури соціальної групи на основі експертних оцінок та з використанням апарату теорії нечітких множин.

4. Задача визначення структури соціальної групи як задача нечіткої кластеризації

Вербальна постановка задачі визначення структури соціальної групи може бути виконана таким чином: у заданій соціальній групі здійснити групування її членів за певними соціально-демографічними ознаками; визначити кількісний та представницький склад кожної складової заданої групи. Представники групи характеризуються значеннями ознак, що входять в їх соціально-демографічний портрет. До таких ознак, як правило відносять стать, вік, освітній рівень, професійну діяльність, сімейний стан, соціальний рівень, релігійну належність тощо.

Критерії групування осіб залежать від задачі, яка розв'язується. Це може бути групування осіб за їх споживчими вподобаннями, за купівельною спроможністю, за характером суспільної поведінки тощо. Складність здійснення такого групування пов'язана з тим, що через відсутність чітких описів груп, на які розбивається задана, можливі випадки, в яких для деяких осіб виникне ситуація одночасної належності до різних груп. З огляду на це представимо поставлену задачу к задачу нечіткої кластеризації, математичну модель якої побудуємо таким чином:

Нехай дано множину об'єктів $X = \{X_i = (x_{i1}, x_{i2}, \dots, x_{iM}), i = \overline{1, N}\}$, які характеризуються за множиною ознак $K = \{K_1, K_2, \dots, K_M\}$, тобто x_{ij} — значення j -ої ознаки для i -го об'єкта. Необхідно розбити дану множину на G нечітких кластерів за заданим критерієм, тобто побудувати алгоритм Θ , виконання якого дозволило б визначати ступені належності об'єкта до кожного з кластерів, тобто:

$$X_i \xrightarrow{\Theta} (\mu_1(X_i), \mu_2(X_i), \dots, \mu_G(X_i)),$$

де $\mu_g(X_i)$ — ступінь належності об'єкта X_i кластеру під номером g , $g = \overline{1, G}$, при чому $\sum_{g=1}^G \mu_g(X_i) = 1$.

Для розв'язання сформульованої задачі на практиці часто використовується метод нечітких s -середніх. Основною складністю застосування цього методу в да-

ному випадку є те, що, як правило, значна частина координат векторів X_i приймає нечислові значення, що приводить до необхідності введення спеціальних функцій відстані між об'єктами. Також, при здійсненні розбиття об'єктів на кластери необхідно забезпечити врахування особливостей самих кластерів.

В дослідженні пропонується адаптований метод нечітких s -середніх.

5. Алгоритм адаптованого методу нечітких s -середніх для розв'язання задачі визначення структури соціальної групи

На початку роботи процедури кластеризації необхідно розв'язати такі задачі:

1. *Задача впорядкування ознак.* Вербальна постановка задачі є такою: необхідно для кожної ознаки з множини K встановити ступінь її впливу на можливість входження об'єкта до того чи іншого кластеру.

Математично задача полягає у ідентифікації функціональної залежності Ω , за допомогою якої б здійснювалося відображення:

$$\Omega: K \rightarrow \mathbb{R}^+,$$

та визначалася б ступінь впливу ознаки на формування кластерів.

Один із способів обчислення значень функції Ω базується на результатах експертних опитувань та зводиться до послідовних розв'язань задач числової оцінки об'єкта. Серед методів визначення числової оцінки об'єкта можна виокремити статистичний метод, утилітарний та егалітарний методи, евристичні методи тощо [18, 20, 21]. Позначимо $\alpha_j = \Omega(K_j)$, $j = \overline{1, M}$.

2. *Задача побудови лінгвістичних змінних.* Для кожної ознаки з множини K будемо лінгвістичну змінну [9, 19]. Побудову можна виконувати на основі висновків експертів. Позначимо через $\langle I_j, T_j, Q_j, S_j, P_j \rangle$ параметри, що відповідають лінгвістичній змінній K_j , $j = \overline{1, M}$; R_j — потужність множини T_j ; $\mu_{T_j}(x_j)$ — функція належності для r -го терму множини T_j , $r = \overline{1, R_j}$.

Після розв'язання вказаних задач переходимо до запуску ітераційної процедури адаптованого методу нечітких s -середніх. Алгоритм методу є таким:

Крок 1. Задаємо параметр зупинки алгоритму $\varepsilon \in (0; 1)$ та ступінь нечіткості m , $1 \leq m < \infty$ [18]. При $m = 1$ отримуємо чітке розбиття, тобто розбиття, у якому кожен об'єкт відноситься тільки до одного кластеру. Із збільшенням параметру m збільшується ступінь розмитості виділених кластерів.

Для кожного об'єкта із множини $X = \{X_i = (x_{i1}, x_{i2}, \dots, x_{iM}), i = \overline{1, N}\}$ довільним чином задаємо ступінь належності кожному з кластерів та формуємо матрицю $U^0 = (u_{ig})$, $u_{ig} \in [0; 1]$, де u_{ig} — початкова ступінь належності i -го об'єкта g -му кластеру ($i = \overline{1, N}, g = \overline{1, G}$), та $\sum_{g=1}^G u_{ig} = 1, \forall i \in \{1, \dots, N\}$.

Визначимо функцію відстані між двома об'єктами $X_{i_1}, X_{i_2} \in X$ таким чином:

$$d(X_{i_1}, X_{i_2}) = \frac{\sum_{j=1}^M \left(\alpha_j \cdot \frac{1}{R_j} \sum_{r=1}^{R_j} |\mu_{rT_j}(x_{i_1j}) - \mu_{rT_j}(x_{i_2j})| \right)}{\sum_{j=1}^M \alpha_j}, \quad (1)$$

де x_{i_1j}, x_{i_2j} — відповідні компоненти векторів X_{i_1}, X_{i_2} .

Крок 2. Визначаємо значення функцій належності для компонент векторів, що є поточними центрами кластерів за формулою:

$$\mu_{rT_j}(c_j^{(g)}) := \frac{\sum_{i=1}^N (u_{ig})^m \cdot \mu_{rT_j}(x_{ij})}{\sum_{i=1}^N (u_{ig})^m},$$

де $C^{(g)} = (c_1^{(g)}, c_2^{(g)}, \dots, c_M^{(g)})$ — поточний центр g -го кластеру ($g = \overline{1, G}$).

Крок 3. Знаходимо відстані від об'єктів до центрів кластерів за формулою (1). Позначимо $d_{ig} = d(X_i, C^{(g)})$, $i = \overline{1, N}$, $g = \overline{1, G}$.

Крок 4. Обчислюємо коефіцієнти матриці U за формулами:

якщо $\exists g_1 \in \{1, 2, \dots, G\}$: $d_{ig_1} = 0$,

то $u_{ig_1} = 1$, $u_{ig} = 0$, $\forall g \neq g_1$, $g = \overline{1, G}$;

якщо $\forall g = \overline{1, G}$ $d_{ig} > 0$, то $u_{ig} = \frac{1}{\sum_{i=1}^N \left(\frac{d_{ig}}{d_{ig}} \right)^{\frac{2}{m-1}}}$.

Крок 5. Перевіряємо умови зупинки алгоритму. Тобто, якщо $\|U - U^*\|^2 < \epsilon$, де U^* — матриця, отримана на попередньому кроці, то алгоритм закінчено, інакше переходимо до кроку 2.

6. Сфери застосування методу

Адаптований метод нечіткої кластеризації може використовуватися в задачах дослідження соціальної структури суспільства. Числові експерименти розв'язання таких задач запропонованим методом показали, що кількість ітерацій залежить від початкових значень функції належності, а результат об'єднання об'єктів у кластери, — від визначення вагових коефіцієнтів ознак, що впливають на групування. Незважаючи на такі особливості, при вдалому підборі параметрів та залученні кваліфікованих експертів, метод дозволяє ефективно визначати структуру соціальних груп без необхідності проведення додаткових експертиз, опитувань тощо.

Також, результати розбиття групи осіб на підгрупи можуть бути основою для розв'язання задач побудови соціально-демографічного портрету особи-представника певної групи, що виникають в політології, економіці, медицині та інших галузях.

7. Висновки

В роботі досліджено задачу визначення структури соціальних груп. Проведений аналіз задачі виявив її особливості, пов'язані з характером вхідних даних та складностями, які виникають при визначенні відстані між об'єктами.

Виконано вербальну та математичну постановку задачі визначення структури соціальної групи як задачі нечіткої кластеризації, яка передбачає можливість виконання такого розбиття на кластери, при якому кожен об'єкт може одночасно належати декільком кластерам з різними ступенями належності.

Розроблено адаптований метод нечітких s -середніх для розв'язання сформульованої задачі. В методі використано результати експертних опитувань щодо оцінки впливу окремих ознак на можливість входження об'єктів у кластери. Врахування таких оцінок дозволяє проводити розбиття у відповідності до мети задачі та природи самих кластерів. Ознаки об'єктів, за якими проводиться розбиття на кластери, представлені у вигляді лінгвістичних змінних. Таке перетворення дозволяє в подальшому проводити порівняння нечислових даних та задавати метрику на множині об'єктів.

Запропоновано правила для обчислення відстані між об'єктами та визначенні центрів кластерів, які базуються на врахуванні відстаней між ознаками, яким відповідають нечіткі множини лінгвістичних змінних.

В дослідженні зазначені класи задач, для яких ефективним є використання запропонованого методу.

Література

- Огаренко, В. М. Соціологія малих груп [Текст]: підручник для студ. вузів / В. М. Огаренко, Ж. Д. Малахова; Гуманіт. ун-т, Запорізький ін-т держ. та муніципального управління. — Київ: Центр навчальної літ., 2005. — 291 с.
- Жоль, К. К. Соціологія [Текст] / К. К. Жоль. — К.: Либідь, 2005. — 440 с.
- Мироноук, І. С. Результати визначення оціночної чисельності представників уразливої щодо інфікування ВІЛ групи населення (жінки секс-бізнесу) в Закарпатській області [Текст] / І. С. Мироноук, В. Й. Шатило, І. Я. Гуцол, В. В. Брич // Україна. Здоров'я нації. — 2011. — № 2(18). — С. 133–137.
- Мироноук, І. С. Застосування інформаційних технологій для вирішення задачі оцінки чисельності представників груп високого ризику інфікування ВІЛ [Текст] / І. С. Мироноук, О. Ю. Мулеса // Вісник соціальної гігієни та організації охорони здоров'я України. Науково-практичний журнал. — 2013. — № 3(57). — С. 55–63.
- Мироноук, І. С. Визначення поведінкових ризиків ВІЛ-інфікованих жителів Закарпаття, пов'язаних з трудовою міграцією [Текст] / І. С. Мироноук // Туберкульоз. Легеневі хвороби. ВІЛ-інфекція. — 2012. — № 2(9) — С. 97–102.
- Мулеса, О. Ю. Інформаційні технології кількісного оцінювання груп ризику інфікування вірусом імунодефіциту людини [Текст] / О. Ю. Мулеса // Східно-Європейський журнал передових технологій. — 2013. — № 5/4(65). — С. 10–15. — Режим доступу: \www/URL: <http://journals.uran.ua/eejet/article/view/18327>
- Ту, Дж. Принципи розпознавання образів [Текст] / Дж. Ту, Р. Гонсалес. — М.: Мир, 1978. — 412 с.
- Мандель, І. Д. Кластерний аналіз [Текст] / І. Д. Мендель. — М.: Финансы и статистика, 1988. — 320 с.
- Снитюк, В. Є. Прогнозування. Моделі. Методи. Алгоритми [Текст]: навч. посіб. / В. Є. Снитюк. — К.: Маклаук, 2008. — 364 с.
- Gorban, A. N. Method of Elastic Maps and its Applications in Data Visualization and Data Modelling [Text] / A. N. Gorban, A. Yu. Zinoyev // Int. Journal of Computing Anticipatory Systems, CHAOS. — 2002. — Vol. 12. — P. 353–369.

11. Снитюк, В. Эволюционная кластеризация сложных объектов и процессов [Текст] / В. Снитюк // XI-th International Conference «Knowledge-Dialogue-Solution» — Varna, 2005. — Т. 1. — С. 232–237.
12. Батуркин, С. А. Статистические алгоритмы кластеризации данных в адаптивных обучающих системах [Текст] / С. А. Батуркин, Е. Ю. Батуркина, В. А. Зименко, И. В. Сигинов // Вестник РГРТУ. — Рязань, 2012. — Вып. 31, № 1. — С. 82–85.
13. Фукунага, К. Введение в статистическую теорию распознавания образов [Текст] / К. Фукунага. — М.: Наука. Главная редакция физико-математической литературы, 1979. — 368 с.
14. Плюта, В. Сравнительный многомерный анализ в эконометрическом моделировании [Текст] / В. Плюта. — М.: Финансы и статистика, 1989. — 175 с.
15. Котов, А. Кластеризация данных [Электронный ресурс] / А. Котов, Н. Красильников. — 2 октября 2006 г. — Режим доступа: \www/URL: <http://yury.name/internet/02ia-seminar-note.pdf>
16. Kohonen, T. Self-Organization and Associative Memory [Text] / T. Kohonen // Springer Series in Information Sciences. — Springer Berlin Heidelberg, 1989. — 312 p. doi:10.1007/978-3-642-88163-3
17. Горбаченко, В. И. Сети и карты Кохонена [Электронный ресурс] / В. И. Горбаченко. — Режим доступа: \www/URL: <http://gorbachenko.self-organization.ru/index.html>
18. Смешко, Ю. В. Об одном критерии для выбора значения экспоненциального веса в алгоритме классификации нечетких С-средних [Текст] / Ю. В. Смешко // Молодежь и наука. — Красноярск, 2012. — С. 19–25.
19. Орловский, С. А. Проблемы принятия решений при нечеткой исходной информации [Текст] / С. А. Орловский. — М.: Наука. Главная редакция физико-математической литературы, 1981. — 208 с.
20. Волошин, О. Ф. Теорія прийняття рішень [Текст]: навч. посібн. / О. Ф. Волошин, С. О. Машенко. — К.: Видавничо-поліграфічний центр «Київський університет», 2010. — 366 с.
21. Маляр, М. М. Схема паралельно-послідовного відсіву варіантів для задачі вибору [Текст] / М. М. Маляр, О. Ю. Швалагін // Східно-Європейський журнал передових технологій. — 2011. — № 1/4(49). — С. 39–42. — Режим доступу: \www/URL: <http://journals.urau.ua/eejet/article/view/1911>

АДАПТАЦИЯ МЕТОДА НЕЧЕТКИХ С-СРЕДНИХ К ЗАДАЧЕ ОПРЕДЕЛЕНИЯ СТРУКТУРЫ СОЦИАЛЬНЫХ ГРУПП

Рассматривается задача определения структуры социальных групп как задача нечеткой кластеризации объектов. Предлагается адаптация метода нечетких с-средних для решения сформулированной задачи, которая заключается в определении новых способов вычисления расстояния между объектами и определения центров кластеров. Эффективность метода достигается путем представления признаков объектов через соответствующие лингвистические переменные.

Ключевые слова: структура социальной группы, нечеткая кластеризация, лингвистическая переменная, метод нечетких с-средних.

Мулеса Оксана Юрїївна, кандидат технічних наук, доцент, кафедра кібернетики і прикладної математики, ДВНЗ «Ужгородський національний університет», Україна, e-mail: mulesa.oksana@gmail.com.

Мулеса Оксана Юрьевна, кандидат технических наук, доцент, кафедра кибернетики и прикладной математики, ГВУЗ «Ужгородский национальный университет», Украина.

Mulesa Oksana, Uzhgorod national university, Ukraine, e-mail: mulesa.oksana@gmail.com