

Проблема невизначеності фаззифікатора і «ширини зони» в РСМ-алгоритмі

Ольга Серпінська, ст. викладач каф. ІТППМ¹ (ORCID: 0000-0003-3589-2267), Олександр Терентьєв проф., д.т.н., зав. каф. ІТППМ¹ (ORCID: 0000-0002-8148-5323)

¹ Київський національний університет будівництва і архітектури, 03037, м. Київ, проспект Повітряних Сил, 31, Україна

АНОТАЦІЯ

Запропоновано вирішення проблеми невизначеності фаззифікатора і «ширини зони» в РСМ (possibilistic c -means) -алгоритмі. РСМ-алгоритм дозволяє поліпшити результати кластеризації, що отримані із застосуванням FCM (fuzzy c -means)-алгоритму, у випадку, якщо множина об'єктів містить атипові об'єкти (об'єкти-шуми), за рахунок послаблення властивості кластерної відносності і з урахуванням властивості типовості.

Ключові слова: фаззифікатор, ширина зони, РСМ-алгоритм, кластеризація, FCM-алгоритм, нечіткі множини, хромосома змінної довжини, нечіткі ступені приналежності, центр кластеру.

1. ВСТУП

FCM-алгоритм працює добре, якщо множина об'єктів містить кластери подібного об'єму гіперсферичної форми або подібної щільності. Якщо об'єм кластерів збільшується або кількість об'єктів в кожному кластері зменшується, то нечітка ступінь приналежності для об'єктів в кластері буде змінюватися. В FCM – алгоритмі збільшення (зменшення) нечіткої ступені приналежності для об'єктів може призвести до небажаного зміння результатів кластеризації. РСМ-алгоритм дозволяє поліпшити результати кластеризації, що отримані із застосуванням FCM-алгоритму, у випадку, якщо множина об'єктів містить атипові об'єкти (об'єкти-шуми), за рахунок послаблення властивості кластерної відносності і з урахуванням властивості типовості. Однак результати кластеризації з використанням цього алгоритму також сильно залежать від вибору фаззифікатора m і значень «ширини зони» η_j ($j = 1 \dots c$), якщо кластери мають суттєво різну щільність або суттєво різний об'єм. Таким чином, можна говорити про існування невизначеності при виявленні типовості об'єкта кластеру в РСМ-алгоритмі і необхідності в реалізації методу розрахунку значень ступені типовості об'єктів кластером за рахунок управління невизначеністю при виборі параметрів РСМ-алгоритму – значень фаззифікатора m або «ширини зони» η_j ($j = 1 \dots c$).

2. ПРОБЛЕМА НЕВИЗНАЧЕНОСТІ ФАЗЗИФІКАТОРА І «ШИРИНИ ЗОНИ» В РСМ-АЛГОРИТМІ

Як і FCM-алгоритм на основі НМТ1, (нечітких множин 1-го типу) РСМ-алгоритм на основі НМТ1 є ітераційним алгоритмом, який реалізує обчислення функцій типовості об'єктів центрам кластерів. В РСМ-алгоритмі на основі НМТ1 функції типовості оцінюють «сумісність об'єкта x_i з центром кластера v_j » (можливість того, що об'єкт належить кластеру), яка залежить від абсолютної відстані між центром кластера v_j ($j = 1, c$) і об'єктом x_i ($i = 1 \dots n$), а також від вибору фаззифікатора m і «ширини зони» η_j .

$$\eta_1 = \eta_2 = 0,5$$

На рисунку 1 представлені графіки ФТ (функції типовості) для випадку двох кластерів при різних комбінаціях значень фаззифікатора m і «ширини зон» $\eta_1 = \eta_2 = 0,5$.

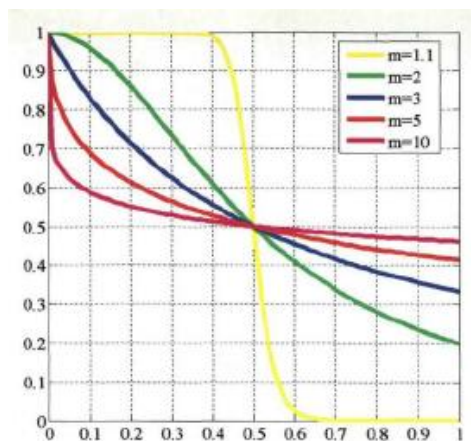


Рисунок 1. Графік функцій типовості для РСМ-алгоритму на основі НМТ1 при різних значеннях фаззифікатора m і «ширини зони»

Вочевидь, що без правильного вибору значень фаззифікатора m і «ширини зони» η_j ($j = 1 \dots c$) неможливо отримати адекватні результати кластеризації. При цьому, як і при використанні FCM-алгоритму на основі ПМТ1, використання РСМ-алгоритму на основі НМТ1 не завжди є ефективним у разі, коли кластери мають істотно різну щільність або істотно різний об'єм.

3. РОЗШИРЕННЯ МНОЖИНИ ОБ'ЄКТІВ КЛАСТЕРИЗАЦІЇ НА ІНТЕРВАЛЬНІ НЕЧІТКІ МНОЖИНИ ДРУГОГО ПОРЯДКУ ДЛЯ РСМ-АЛГОРИТМУ

Невизначеність в РСМ-алгоритмі на основі ІНМТ2 може бути виражена як за допомогою двох різних значень фаззифікатора m (як в FCM-алгоритмі на основі ІНМТ2) і фіксованого значення «ширини зони» η_j ($j = 1 \dots c$) для кожного кластера, так і за допомогою двох різних значень для «ширини зони» η_j ($j = 1 \dots c$) для кожного кластера при фіксованому значенні фаззифікатора m [2]. При представленні невизначеності в РСМ-алгоритмі на основі ІНМТ2 (нечіткої множини 2-го типу) за допомогою двох

різних значень фазифікатора m : m_1 і m_2 (як в FCM-алгоритмі на основі ІНМТ2) і фіксованого значення «ширини зони» η_j ($j = 1 \dots c$) для кожного кластера «нижня» і «верхня» інтервальні ФТ об'єкта x_i можуть бути записані таким чином:

$$w_j(x_i) = \begin{cases} \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}} \leq \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}} \\ \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}} > \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}} \end{cases} \quad (1)$$

$$\bar{w}_j(x_i) = \begin{cases} \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}} > \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}} \\ \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}} \leq \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}} \end{cases} \quad (2)$$

На рисунку 2 наведений приклад «відбитка невизначеності» для випадку двох кластерів на основі формул (1) і (2) для комбінації $\eta_j = 0,5$, $m_1 = 2$, $m_2 = 5$.

При представленні невизначеності в РСМ-алгоритмі на основі ІНМТ2 за допомогою двох різних значень «ширини зони» η_j : η_1 і η_2 ($j = 1 \dots c$) для кожного кластера η_j при фіксованому значенні фазифікатора m «нижня» і «верхня» інтервальні ФТ об'єкта x_i можуть бути записані у вигляді [1]:

$$\bar{w}_j(x_i) = \begin{cases} \frac{1}{1 + \left(\frac{d_{ji}}{\eta_1}\right)^{\frac{2}{m-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_1}\right)^{\frac{2}{m-1}}} > \frac{1}{1 + \left(\frac{d_{ji}}{\eta_2}\right)^{\frac{2}{m-1}}} \\ \frac{1}{1 + \left(\frac{d_{ji}}{\eta_2}\right)^{\frac{2}{m-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_1}\right)^{\frac{2}{m-1}}} \leq \frac{1}{1 + \left(\frac{d_{ji}}{\eta_2}\right)^{\frac{2}{m-1}}} \end{cases} \quad (3)$$

$$w_j(x_i) = \begin{cases} \frac{1}{1 + \left(\frac{d_{ji}}{\eta_1}\right)^{\frac{2}{m-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_1}\right)^{\frac{2}{m-1}}} \leq \frac{1}{1 + \left(\frac{d_{ji}}{\eta_2}\right)^{\frac{2}{m-1}}} \\ \frac{1}{1 + \left(\frac{d_{ji}}{\eta_2}\right)^{\frac{2}{m-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_1}\right)^{\frac{2}{m-1}}} > \frac{1}{1 + \left(\frac{d_{ji}}{\eta_2}\right)^{\frac{2}{m-1}}} \end{cases} \quad (4)$$

В формулах (3) і (4) η_{j_1} і η_{j_2} - дві «ширини зони», що визначають різні відстані, на яких можливий ступінь приналежності дорівнює 0,5.

На рисунку 3 наведений приклад «відбитка невизначеності» для випадку двох кластерів на основі формул (3) і (4) для комбінації $m = 3$, $\eta_{j_1} = 0,25$ і $\eta_{j_2} = 0,75$.

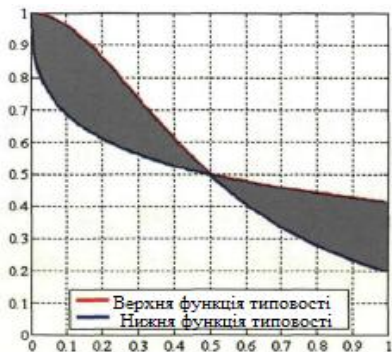


Рисунок 2. «Відбиток невизначеності» для РСМ-алгоритму на основі ІНМТ2 для комбінації $\eta_j = 0,5$, $m_1 = 2$, $m_2 = 5$

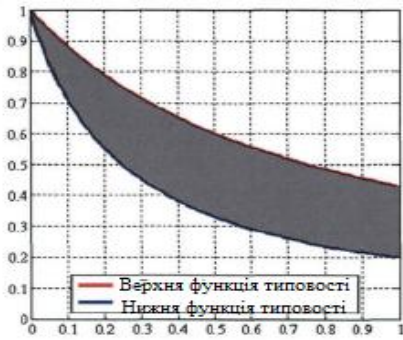


Рисунок 3. «Відбиток невизначеності» для РСМ-алгоритму на основі ІНМТ2 для Комбінацій $m=3$, $\eta_{j_1} = 0,25$, і $\eta_{j_2} = 0,75$

Використання різних значень для фазифікатора m визначає різні цільові функції в РСМ-алгоритмі на основі ІНМТ2 при $m=m_1$ і $m = m_2$:

$$J_{m_1}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{m_1} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{m_1} \quad (5)$$

$$J_{m_2}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{m_2} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{m_2} \quad (6)$$

Використання різних значень «ширини зони» η_j ($j = 1 \dots c$) визначає різні цільові функції в РСМ-алгоритмі на основі ІНМТ2 при $\eta_j = \eta_{j_1}$ і $\eta_j = \eta_{j_2}$

$$J_{\eta_1}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{m_1} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_{j_1}^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{m_1} \quad (7)$$

$$J_{\eta_2}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{m_2} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_{j_2}^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{m_2} \quad (8)$$

Обчислення ФТ $w_j(x_i)$ і координат центрів кластерів реалізується з використанням ітераційного алгоритму Карника - Менделя.

Пошук оптимальних параметрів РСМ-алгоритму на основі ІНМТ2, що забезпечують адекватні результати кластеризації, може бути реалізований з використанням ГА [2].

Пошук оптимальних параметрів РСМ-алгоритму на основі ІНМТ2, що забезпечують адекватні результати кластеризації неможливий без застосування будь-якого оптимізаційного алгоритму, наприклад, генетичного алгоритму (ГА).методу нечіткої кластеризації [3].

Список літератури

[1] Terentiev O.O. Ph.D., professor, Serpinska O. I., assistant. Clustering methods in the tasks of assessing the technical condition of buildings under conditions of uncertainty. 2021. II Scientific and practical conference "Distributed software systems and technologies". November 19, 2021, Kyiv, KNUBA, p. 8-9.

[2] Jae Hyuk Cho, Data Clustering for Fuzzyfier Value Derivation, 2022, Vol.12, pp.280-286/doi: 10.5772/intechopen.87799.

[3] Terentiev O.O., Serpinska O. I, Ryabchun Yu.V. Application of fuzzy logic to assess the technical condition of construction objects. VIII international scientific-practical conference "Management of the development of technologies". Kyiv, p. 45-46.