

КЛАСТЕРНЫЙ АНАЛИЗ Z-ИНФОРМАЦИИ НА ОСНОВЕ ЭТАЛОННОЙ СИСТЕМЫ НЕЧЕТКИХ ОПРЕДЕЛЕНИЙ ПРИНАДЛЕЖНОСТИ

О.М. Полещук

ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (Мытищинский филиал), Россия, 141005, Московская обл., г. Мытищи, ул. 1-я Институтская, д. 1
polshchuk@mgul.ac.ru

Разработан алгоритм кластеризации данных, представленных лингвистическими Z-числами. Обе компоненты чисел (оценки объектов и их достоверность) являются значениями лингвистических переменных. Кластеризация информации осуществлялась на основе нечетких эталонных высказываний о важности характеристик объектов, формализованных на основе лингвистических переменных. Оценки объектов и нечеткие эталонные высказывания использованы для определения нечетких рейтинговых оценок степени принадлежности объектов к кластерам. Разработанный в статье алгоритм улучшает алгоритм кластеризации, представленный автором ранее, поскольку сохраняет больше исходной информации из-за нового подхода к формализации данных и уменьшает нечеткость рейтинговых оценок объектов, тем самым уменьшая риски ошибок в задачах поддержки принятия решений.

Ключевые слова: Z-информация, алгоритм кластеризации, рейтинговая оценка, лингвистическая переменная

Ссылка для цитирования: Полещук О.М. Кластерный анализ Z-информации на основе эталонной системы нечетких определений принадлежности // Лесной вестник / Forestry Bulletin, 2024. Т. 28. № 2. С. 150–155. DOI: 10.18698/2542-1468-2024-2-150-155

Кластерный анализ является неотъемлемой частью аппарата обработки нечеткой информации (информации с нечеткими данными). Для формализации такой информации разработаны различные методы, в основе которых лежит понятие лингвистической переменной [1–6].

Одним из первых алгоритмов кластерного анализа нечетких данных был алгоритм нечеткой кластеризации *c*-средних (*c*-means) [7]. Этот алгоритм широко известен, часто используется и периодически улучшается, в том числе, в комбинации с другими алгоритмами. Так, например, в работе [8] этот алгоритм представлен в комбинации с генетическим алгоритмом, а в работах [9, 10] — улучшен в результате минимизации целевой функции с использованием алгоритма оптимизации роя частиц [11]. Новейшие модификации алгоритма нечеткой кластеризации *c*-средних представлены в работах [12, 13].

Для кластеризации нечетких данных разработаны алгоритмы на основе транзитивных нечетких отношений [14–20].

После определения Z-числа в 2011 г. профессором Лотфи Заде [21] и появления возможности учета достоверности информации с нечеткими данными, существующие алгоритмы кластеризации оказались непригодными для работы в новых условиях, поэтому возникла объективная необходимость в разработке алгоритмов

кластерного анализа Z-информации (информации с Z-числами).

В работе [22] представлен алгоритм нечеткой кластеризации, который объединяет преобразование Z-чисел в нечеткие числа и алгоритм нечеткой кластеризации *c*-средних. В работе [23] разработан алгоритм кластеризации Z-информации на основе системы нечетких правил вывода, а в работе [24] — на основе нечетких множеств второго типа. В работе [25] разработан кластерный алгоритм Z-информации на основе бинарных нечетких отношений, а в работе [26] — алгоритм на основе системы нечетких высказываний о принадлежности объектов к каждому из кластеров. В работе [26] нечеткие высказывания представлены в виде значений лингвистической переменной, а оценки объектов — в виде Z-чисел, которые трансформируются в агрегирующие интервалы для последующего кластерного анализа. В связи с этим возникает вопрос исключения потери информации в результате такой трансформации и необходимости разработки алгоритма, позволяющего этого избежать.

Цель работы

Цель работы — разработка алгоритма кластерного анализа Z-информации на основе эталонной системы нечетких определений принадлежности, позволяющего сохранить исходную информацию об объектах в процессе ее формализации и снизить нечеткость конечных результатов.

Основные понятия и определения

Лингвистической переменной называют систему $\{X, T(X), U, V, S\}$, где X — название переменной; $T(X) = \{X_l, l = 1, m\}$ — терм-множество переменной X , т. е. множество термов или названий лингвистических значений переменной X ; V — синтаксическое правило, порождающее названия значений лингвистической переменной X ; S — семантическое правило, которое ставит в соответствие каждому терму из множества $T(X)$ нечеткое подмножество универсального множества U [1].

Если множество U является подмножеством действительной прямой, то нечеткими формализациями термов являются нечеткие числа. Функции принадлежности нечетких чисел, соответствующих термам лингвистической переменной, определяются четырьмя параметрами, если используется нечеткое T -число $\tilde{A} = \{a_1, a_2, a_L, a_R\}$. Первые два параметра представляют собой абсциссы левой и правой вершин верхнего основания трапеции, а последние два параметра — длины левого и правого крыльев трапеции. Функции принадлежности нечетких чисел, соответствующих термам лингвистической переменной, определяются тремя параметрами, если используется треугольное нечеткое число $\tilde{A} = \{a, a_L, a_R\}$. Первый параметр — это абсцисса вершины треугольника, а последние два параметра — длины его левого и правого крыльев.

В работе [6] приведено определение агрегирующего отрезка $[c_1, c_2]$ для нечеткого числа $\tilde{A} = \{a_1, a_2, a_L, a_R\}$. [23]:

$$c_1 = \int_0^1 \frac{2a_1 - (1-\alpha)a_L}{2} 2\alpha d\alpha = a_1 - \frac{1}{6} a_L,$$

$$c_2 = \int_0^1 \frac{2a_2 - (1-\alpha)a_R}{2} 2\alpha d\alpha = a_2 - \frac{1}{6} a_R.$$

Z-числом называется упорядоченная пара нечетких чисел $Z = (\tilde{C}, \tilde{R})$, где число \tilde{R} достоверность (надежность) числа \tilde{C} [21].

Лингвистическим называется Z-число, компоненты которого являются значениями лингвистических переменных, т. е., если $Z = (\tilde{C}, \tilde{R})$ является лингвистическим Z-числом, то первое нечеткое число \tilde{C} равно одному из нечетких значений $\tilde{X}_l, l = 1, m$ некоторой лингвистической переменной X , а второе число \tilde{R} равно одному из нечетких значений $\tilde{R}_v, v = 1, V$ некоторой лингвистической переменной R .

В работе [27] представлено определение агрегирующего отрезка $[\delta_1, \delta_2]$ для Z-числа $Z = (\tilde{A}, \tilde{R})$, $\tilde{A} = \{c_1, c_2, c_L, c_R\}$, $\tilde{R} = \{r_1, r_2, r_L, r_R\}$,

$$\delta_1 = r_1 \left(c_1 - \frac{1}{6} c_L \right) - r_L \left(\frac{1}{6} c_1 - \frac{1}{12} c_L \right),$$

$$\delta_2 = r_2 \left(c_2 + \frac{1}{6} c_R \right) + r_R \left(\frac{1}{6} c_2 + \frac{1}{12} c_R \right).$$

Агрегирующие отрезки исходных Z-чисел использованы в работе [26] для построения алгоритма кластеризации.

Постановка задачи и ее решение

Предположим, что N объектов оцениваются в рамках характеристик X_1, X_2, \dots, X_k , при этом используются, соответственно, шкалы с уровнями $X_{il}, i = 1, k, l_i = 1, m_i$. Для оценок объектов известны степени их достоверности, которые определены нечетко в рамках лингвистической шкалы с уровнями $R_v, v = 1, V: R_1$ — «Не достоверно», R_2 — «Не очень достоверно», R_3 — «Достоверно», R_4 — «Очень достоверно», R_5 — «Абсолютно достоверно».

Для кластеризации объектов на основе полученных оценок характеристик X_1, X_2, \dots, X_k сформулирована эталонная система нечетких определений принадлежности объектов к P кластерам. В основе этой системы лежат нечеткие высказывания о важности характеристик объектов $X_i, i = 1, k$ для каждого из рассматриваемых кластеров. Для нечетких высказываний о важности характеристик объектов $X_i, i = 1, k$ использована лингвистическая шкала с уровнями «Абсолютно не важна», «Больше не важна, чем важна», «Не очень важна», «Больше важна, чем не важна», «Важна», «Очень важна».

Формализация всех используемых шкал для оценки объектов, достоверности этих оценок и важности характеристик выполнена на основе лингвистических переменных согласно методу работы [4].

Шкалы для оценки объектов в рамках характеристик $X_i, i = 1, k$ соответственно с уровнями $X_{il}, i = 1, k, l_i = 1, m_i$ формализованы с помощью лингвистических переменных со значениями $\tilde{X}_{il}, i = 1, k, l_i = 1, m_i$.

Достоверность оценок объектов формализована с помощью лингвистической переменной с термами $R_v, v = 1, 5: R_1$ — «Не достоверно», R_2 — «Не очень достоверно», R_3 — «Достоверно», R_4 — «Очень достоверно», R_5 — «Абсолютно достоверно» и соответствующими этим термам нечеткими числами $\tilde{R}_1 = (0, 0, 0, 25), \tilde{R}_2 = (0, 25, 0, 25, 0, 25), \tilde{R}_3 = (0, 5, 0, 25, 0, 25), \tilde{R}_4 = (0, 75, 0, 25, 0, 25), \tilde{R}_5 = (1, 0, 25, 0)$.

С помощью прямого опроса экспертов формализована лингвистическая шкала для оценки важности характеристик с термами «Абсолютно не важна», «Больше не важна, чем важна», «Не очень

важна», «Больше важна, чем не важна», «Важна», «Очень важна», которым в соответствие поставлены нечеткие числа $\tilde{C}_1, \tilde{C}_2, \dots, \tilde{C}_6$:

$$\begin{aligned} \tilde{C}_1 &= (0, 0, 0, 2), \tilde{C}_2 = (0, 2, 0, 2, 0, 2), \\ \tilde{C}_3 &= (0, 4, 0, 2, 0, 2), \tilde{C}_4 = (0, 6, 0, 2, 0, 2), \\ \tilde{C}_5 &= (0, 8, 0, 2, 0, 2), \tilde{C}_6 = (1, 0, 2, 0). \end{aligned}$$

Оценки объектов в рамках характеристик $X_i, i = 1, k$ формализованы с помощью лингвистических Z-чисел $Z_{ni} = (\tilde{X}_{ni}, \tilde{R}_{ni}), n = 1, N, i = 1, k$, обе компоненты которых являются значениями лингвистических переменных, описанных выше. Первая компонента $\tilde{X}_{ni} = (x_{ni}^1, x_{ni}^2, x_{ni}^L, x_{ni}^R), n = 1, N, i = 1, k$ равна одному из нечетких чисел $\tilde{X}_{li}, i = 1, k, l_i = 1, m_l$, вторая компонента $\tilde{R}_{ni} = (r_{ni}^1, r_{ni}^2, r_{ni}^L, r_{ni}^R), n = 1, N, i = 1, k$ равна одному из нечетких чисел $\tilde{R}_v, v = 1, V$.

Через $\tilde{C}_{pi} = (c_{pi}^1, c_{pi}^2, c_{pi}^L, c_{pi}^R), p = 1, P, i = 1, k$ обозначены нечеткие числа, которые формализуют степени важности характеристик $X_i, i = 1, k$ для p -го кластера. Каждое из чисел $\tilde{C}_{pi} = (c_{pi}^1, c_{pi}^2, c_{pi}^L, c_{pi}^R), p = 1, P, i = 1, k$ равно одному из чисел $\tilde{C}_1, \tilde{C}_2, \dots, \tilde{C}_6$:

С помощью объединения оценок объектов и степеней важности характеристик для каждого из кластеров получена информация по каждому из объектов в виде P совокупностей, каждая из которых соответствует определенному кластеру и состоит из k пар нечетких чисел (степени важности характеристик $X_i, i = 1, k$ для этого кластера) и Z-чисел (оценки объектов в рамках характеристика $X_i, i = 1, k$ с их достоверностью):

$$\begin{aligned} & \{(\tilde{C}_{p1}, Z_{n1}), (\tilde{C}_{p2}, Z_{n2}), \dots, (\tilde{C}_{pk}, Z_{nk})\} = \\ & = \{(\tilde{C}_{p1}, (\tilde{X}_{n1}, \tilde{R}_{n1})), (\tilde{C}_{p2}, (\tilde{X}_{n2}, \tilde{R}_{n2})), \dots, (\tilde{C}_{pk}, (\tilde{X}_{nk}, \tilde{R}_{nk}))\}, \\ & p = \overline{1, P}, n = \overline{1, N}. \end{aligned}$$

Агрегирующие отрезки $[\beta_1^{pi}, \beta_2^{pi}], p = 1, P, i = 1, k$ для нечетких чисел $\tilde{C}_{pi} = (c_{pi}^1, c_{pi}^2, c_{pi}^L, c_{pi}^R), p = 1, P, i = 1, k$ определены на основе работы [6] следующим образом:

$$\beta_1^{pi} = c_{pi}^1 - \frac{1}{6}c_{pi}^L, \beta_2^{pi} = c_{pi}^2 + \frac{1}{6}c_{pi}^R.$$

Агрегирующие отрезки $[\alpha_1^{ni}, \alpha_2^{ni}], n = 1, N, i = 1, k$ для нечетких чисел $\tilde{R}_{ni} = (r_{ni}^1, r_{ni}^2, r_{ni}^L, r_{ni}^R), n = 1, N, i = 1, k$ определены аналогично:

$$\alpha_1^{ni} = r_{ni}^1 - \frac{1}{6}r_{ni}^L, \alpha_2^{ni} = r_{ni}^2 + \frac{1}{6}r_{ni}^R.$$

Оценка для n -го объекта в рамках характеристики $X_i, i = 1, k$ определена в виде нечеткого числа

$$\begin{aligned} \tilde{A}_i^n &= (x_{ni}^1 \alpha_1^{ni}, x_{ni}^2 \alpha_2^{ni}, x_{ni}^L \alpha_1^{ni}, x_{ni}^R \alpha_2^{ni}), \\ n &= \overline{1, N}, i = \overline{1, k} \end{aligned}$$

Рейтинговая оценка для n -го объекта в рамках характеристик $X_i, i = 1, k$ определена в виде нечеткого числа

$$\begin{aligned} \tilde{A}^n &= \sum_{i=1}^k \tilde{A}_i^n = \\ &= \left(\sum_{i=1}^k x_{ni}^1 \alpha_1^{ni}, \sum_{i=1}^k x_{ni}^2 \alpha_2^{ni}, \sum_{i=1}^k x_{ni}^L \alpha_1^{ni}, \sum_{i=1}^k x_{ni}^R \alpha_2^{ni} \right), \\ n &= \overline{1, N}. \end{aligned}$$

Рейтинговая оценка принадлежности n -го объекта к p -му кластеру (в соответствии с эталонной системой нечетких определений) определена в виде нечеткого числа

$$\tilde{A}_p^n = \left(\sum_{i=1}^k x_{ni}^1 \alpha_1^{ni} \beta_1^{pi}, \sum_{i=1}^k x_{ni}^2 \alpha_2^{ni} \beta_2^{pi}, \sum_{i=1}^k x_{ni}^L \alpha_1^{ni} \beta_1^{pi}, \sum_{i=1}^k x_{ni}^R \alpha_2^{ni} \beta_2^{pi} \right), n = \overline{1, N}.$$

Объект с индексом l признан типичным представителем кластера с индексом p , если $\sup x_n: \mu_{A_p^n} = 1 \in A_p^l, \mu_{A_p^n}$ — функция принадлежности A_p^n . Степени принадлежности $\mu_p(n), n \neq l, n = 1, N$ других объектов к p -му кластеру найдены следующим образом:

$$\begin{aligned} \mu_p(n) &= \max \min_x (\mu_{A_p^l}(x), \mu_{A_p^n}(x)), \\ n &\neq l, n = \overline{1, N}. \end{aligned}$$

При определении рейтинговых оценок объектов, а также рейтинговых оценок принадлежности объектов к кластерам сохраняется больше исходной информации по сравнению с алгоритмом, представленным в работе [26], поскольку при формализации информации происходит трансформация Z-чисел не в отрезки, а в нечеткие числа на основе нечетких оценок объектов.

Пример. В работе [26] приведен пример, результаты которого сравним с результатами кластеризации по разработанному в настоящей статье алгоритму.

Рассмотрим три характеристики X_1, X_2, X_3 , оцениваемые у трех объектов в рамках шкал соответственно с уровнями:

$X_{11}, X_{12}, X_{13}, X_{21}, X_{22}, X_{23}, X_{24}, X_{31}, X_{32}, X_{33}$, которые формализованы с помощью следующих нечетких чисел:

$$\begin{aligned} \tilde{X}_{11} &= (0, 0, 1, 0, 0, 1), \tilde{X}_{12} = (0, 2, 0, 6, 0, 1, 0, 2), \\ \tilde{X}_{13} &= (0, 8, 1, 0, 2, 0), \tilde{X}_{21} = (0, 0, 0, 5, 0, 0, 1), \\ \tilde{X}_{22} &= (0, 1, 5, 0, 1, 5), \tilde{X}_{23} = (0, 3, 0, 7, 0, 1, 5, 0, 1), \end{aligned}$$

$$\tilde{X}_{24} = (0,9, 1, 0,2, 0), \quad \tilde{X}_{31} = (0, 0,2, 0, 0,2),$$

$$\tilde{X}_{32} = (0,4, 0,5, 0,2, 0,2), \quad \tilde{X}_{33} = (0,7, 1, 0,2, 0).$$

Оценки объектов получены с определенными уровнями достоверности, которые определены в рамках лингвистической шкалы с уровнями R_v , $v=1, V$: R_1 — «Не достоверно», R_2 — «Не очень достоверно», R_3 — «Достоверно», R_4 — «Очень достоверно», R_5 — «Абсолютно достоверно». Уровням R_v , $v=1, V$ в соответствие поставлены нечеткие числа $\tilde{R}_1 = (0, 0, 0,25)$, $\tilde{R}_2 = (0,25, 0,25, 0,25)$, $\tilde{R}_3 = (0,5, 0,25, 0,25)$, $\tilde{R}_4 = (0,75, 0,25, 0,25)$, $\tilde{R}_5 = (1, 0,25, 0)$.

Оценки объектов формализованы с помощью Z-чисел:

$$Z_{11} = (\tilde{X}_{12}, \tilde{R}_3), \quad Z_{12} = (\tilde{X}_{24}, \tilde{R}_5), \quad Z_{13} = (\tilde{X}_{32}, \tilde{R}_4),$$

$$Z_{21} = (\tilde{X}_{13}, \tilde{R}_5), \quad Z_{22} = (\tilde{X}_{21}, \tilde{R}_3), \quad Z_{23} = (\tilde{X}_{33}, \tilde{R}_2),$$

$$Z_{31} = (\tilde{X}_{11}, \tilde{R}_4), \quad Z_{32} = (\tilde{X}_{23}, \tilde{R}_5), \quad Z_{33} = (\tilde{X}_{32}, \tilde{R}_3).$$

Сформулированы эталонные высказывания о принадлежности объектов к трем кластерам: « X_1 очень важна, X_2 важна, X_3 больше важна, чем не важна», « X_1 важна, X_2 очень важна, X_3 больше важна, чем не важна», « X_1 больше важна, чем не важна, X_2 не очень важна, X_3 важна».

Найдены рейтинговые оценки объектов в рамках трех характеристик в соответствии с эталонной системой нечетких определений принадлежности:

$$\tilde{A}_1^1 = (0,848, 1,271, 0,207, 0,176),$$

$$\tilde{A}_1^2 = (0,812, 1,142, 0,202, 0,045),$$

$$\tilde{A}_1^3 = (0,291, 0,775, 0,045, 0,221),$$

$$\tilde{A}_2^1 = (1,127, 1,612, 0,145, 0,132),$$

$$\tilde{A}_2^2 = (0,704, 1,097, 0,182, 0,054),$$

$$\tilde{A}_2^3 = (0,419, 0,991, 0,109, 0,166),$$

$$\tilde{A}_3^1 = (0,596, 0,964, 0,208, 0,204),$$

$$\tilde{A}_3^2 = (0,552, 0,882, 0,142, 0,202),$$

$$\tilde{A}_3^3 = (0,248, 0,576, 0,024, 0,136).$$

Исходя из полученных результатов, имеем степени принадлежности $\mu_p(n)$, $n=1,3$, $p=1,3$ объектов к кластерам:

$$\mu_1(1) = 1; \mu_1(2) = 1; \mu_1(3) = 0,934; \mu_2(1) = 1;$$

$$\mu_2(2) = 0,614; \mu_2(3) = 0,753; \mu_3(1) = 1;$$

$$\mu_3(2) = 0,876; \mu_3(3) = 0,484.$$

В работе [26] с теми же исходными данными получены следующие результаты:

$$\tilde{A}_1^1 = (0,853, 1,316, 0,235, 0,354),$$

$$\tilde{A}_1^2 = (0,7940, 0,9545, 0,1757, 0,2273),$$

$$\tilde{A}_1^3 = (0,187, 0,807, 0,062, 0,229),$$

$$\tilde{A}_2^1 = (1,107, 1,617, 0,235, 0,354),$$

$$\tilde{A}_2^2 = (0,703, 0,917, 0,176, 0,227),$$

$$\tilde{A}_2^3 = (0,281, 1,062, 0,062, 0,229),$$

$$\tilde{A}_3^1 = (0,593, 0,947, 0,235, 0,354),$$

$$\tilde{A}_3^2 = (0,556, 0,737, 0,176, 0,227),$$

$$\tilde{A}_3^3 = (0,186, 0,603, 0,062, 0,227),$$

$$\mu_1(1) = 1; \mu_1(2) = 1; \mu_1(3) = 0,899; \mu_2(1) = 1;$$

$$\mu_2(2) = 0,588; \mu_2(3) = 0,905; \mu_3(1) = 1;$$

$$\mu_3(2) = 0,578; \mu_3(3) = 0,522.$$

Окончательные результаты кластерного анализа настоящей работы и работы [26] совпадают, но рейтинговые оценки объектов, имеют меньше нечеткости (нечеткие числа, формализующие оценки, имеют меньшие значения третьего и четвертого параметров), а поэтому распознаются с меньшим риском ошибок, что играет существенную роль при решении задач поддержки принятия решений.

Выводы

Алгоритмы кластеризации нечеткой информации, поступающей для дальнейшего анализа с определенной степенью достоверности, актуальны и востребованы, но не разработаны в должной степени для решения задач поддержки принятия решений в нечеткой среде.

Разработанный в статье алгоритм кластерного анализа Z-информации на основе эталонной системы нечетких определений принадлежности объектов к кластерам позволяет сохранить исходную информацию об объектах в процессе ее формализации и снизить нечеткость конечных результатов.

Для кластеризации оценки объектов формализованы в виде лингвистических Z-чисел, обе компоненты которых (сами оценки и степени их достоверности) являются значениями лингвистических переменных. Эталонная система нечетких определений принадлежности также формализована на основе лингвистической переменной. Формализации использованы для определения рейтинговых оценок принадлежности объектов к кластерам, которые представлены в виде нечетких чисел.

Работа алгоритма показана на примере. Проведен сравнительный анализ полученных результатов и результатов ранее разработанного алгоритма.

Представленный в статье алгоритм улучшает алгоритм кластеризации, представленный автором ранее, поскольку сохраняет больше исходной информации из-за нового подхода к формализации данных и уменьшает нечеткость рейтинговых оценок объектов, тем самым уменьшая риски ошибок в задачах поддержки принятия решений.

Список литературы / References

- [1] Zadeh L.A. Fuzzy logic and approximate reasoning // *Synthese*, 1975, v. 80, pp. 407–428.
- [2] Hwang C.L., Lin N.J. Group decision making under multiple criteria. Berlin: Springer, 1987, 400 p.
- [3] Borisov A.N., Krumberg O.A., Fedorov I.P. Decision making on the basis of fuzzy models: Examples of use. Riga: Zinatne, 1990, 184 p.
- [4] Poleshchuk O., Komarov E. Expert Fuzzy Information Processing. Studies in Fuzziness and Soft Computing, 2011, v. 268, pp. 1–239.
- [5] Ryjov A. Fuzzy Linguistic Scales: Definition, Properties and Applications. In: Reznik L., Kreinovich V. (eds) *Soft Computing in Measurement and Information Acquisition. Studies in Fuzziness and Soft Computing*, 2003, v. 127.
- [6] Poleshchuk O.M. Creation of linguistic scales for expert evaluation of parameters of complex objects based on semantic scopes // *International Russian Automation Conference (RusAutoCon – 2018)*, 2018, pp. 1–6.
- [7] Bezdek J.C. Selected applications in classifier design // *Pattern recognition with fuzzy objective function algorithms*, 1981, v. 2, pp. 203–239.
- [8] Bezdek J.C., Hathaway R.J. Optimization of fuzzy clustering criteria using genetic algorithm // *Proceedings of the IEEE Conference on Evolutionary Computation*, 1994, v. 2, pp. 589–594.
- [9] Runkler T.A., Katz C. Fuzzy clustering by particle swarm optimization // *Proceedings of the IEEE International Conference on Fuzzy Systems*, 2006, pp. 34–41.
- [10] Liu H.C., Yih J.M., Wu D.B., Liu S.W.: Fuzzy C-mean clustering algorithms based on Picard iteration and particle swarm optimization // *Proceedings of the International Workshop on Geoscience and Remote Sensing (ETT and GRS-2008)*, 2008, pp. 75–84.
- [11] Kennedy J., Eberhart R.C. Particle swarm optimization // *Proceedings of the IEEE International Joint Conference on Neural Networks*, 1995, v. 4, pp. 1942–1948.
- [12] Chen M., Ludwig A. Particle swarm optimization based fuzzy clustering approach to identify optimal number of clusters // *J. of Artificial Intelligence and Soft Computing Research*, 2014, v. 4, no. 1, pp. 43–56.
- [13] Phyo O., Chaw E.: Comparative Study of Fuzzy PSO (FPSO) Clustering Algorithm and Fuzzy C-Means (FCM) Clustering Algorithm // *National J. of Parallel and Soft Computing*, 2019, v. 1, no. 1, pp. 62–67.
- [14] Ruspini E.H. Numerical methods for fuzzy clustering // *Information Sciences*, 1970, v. 2, pp. 319–350.
- [15] Tamura S., Higuchi S., Tanaka K. Pattern classification based on fuzzy relations // *IEEE Transactions on Systems, Man and Cybernetics*, 1971, v. 1, pp. 61–66.
- [16] Zadeh L.A.: Similarity relations and fuzzy orderings // *Information Sciences*, 1971, v. 3, pp. 177–200.
- [17] Dunn J.C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters // *J. of Cybernetics*, 1973, v. 3, pp. 32–57.
- [18] Poleshchuk O., Komarov E. The determination of rating points of objects with qualitative characteristics and their usage in decision making problems // *International J. of Computational and Mathematical Sciences*, 2009, v. 3, no. 7, pp. 360–364.
- [19] Darwish A., Poleshchuk O. New models for monitoring and clustering of the state of plant species based on semantic spaces // *J. of Intelligent and Fuzzy Systems*, 2014, v. 3, no. 26, pp. 1089–1094.
- [20] Poleshchuk O.M., Komarov E.G., Darwish A. Assessment of the state of plant species in urban environment based on fuzzy information of the expert group // *Proceedings of the XX IEEE International Conference on Soft Computing and Measurements, (SCM-2017)*, 2017, pp. 651–654.
- [21] Zadeh L.A. A Note on Z-numbers // *Information Sciences*, 2011, v. 14, no. 181, pp. 2923–2932.
- [22] Jamal M., Khalif K., Mohamad S. The implementation of Z-numbers in fuzzy clustering algorithm for wellness of chronic kidney disease patients // *J. of Physics: Conference Series*, 2018, v. 1366, p. 012058.
- [23] Aliev R.A., Pedrycz W., Guirimov B.G., Huseynov O.H. Clustering method for production of Z-numbers based if-then rules // *Information Sciences*, 2020, v. 520, pp. 155–176.
- [24] Aliev R., Guirimov B. Z-number clustering based on general Type-II fuzzy sets // *Advances in Intelligent Systems and Computing*, 2018, v. 896, pp. 270–278.
- [25] Полещук О.М. Кластерный анализ экспертной информации на основе Z-чисел // *Лесной вестник / Forestry Bulletin*, 2022. Т. 26. № 1. С. 143–148. DOI: 10.18698/2542-1468-2022-1-143-148
Poleshchuk O.M. *Klasternyy analiz ekspertnoy informatsii na osnove Z-chisel* [Cluster analysis of expert information based on Z-numbers]. *Lesnoy vestnik / Forestry Bulletin*, 2022, vol. 26, no. 1, pp. 143–148. DOI: 10.18698/2542-1468-2022-1-143-148
- [26] Poleshchuk O. Clustering Z-information based on a system of fuzzy reference requirements // *E3S Web of Conferences*, 2023, v. 420, p. 06022.
- [27] Poleshchuk O.M. Expert group information formalization based on Z-numbers // *J. of Physics: Conference Series*, 2020, v. 1703, p. 012010.

Сведения об авторе

Полещук Ольга Митрофановна — д-р техн. наук, профессор, зав. кафедрой «Высшая математика и физика», ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (Мытищинский филиал), poleshchuk@mgul.ac.ru

Поступила в редакцию 07.11.2023.

Одобрено после рецензирования 20.12.2023.

Принята к публикации 06.02.2024.

CLUSTER ANALYSIS OF Z-INFORMATION BASED ON A REFERENCE SYSTEM OF FUZZY IDENTIFICATION

O.M. Poleshchuk

BMSTU (Mytishchi branch), 1, 1st Institutskaya st., 141005, Mytishchi, Moscow reg., Russia

poleshchuk@mgul.ac.ru

The paper develops an algorithm for clustering Z-information based on reference fuzzy identification of objects belonging to clusters. The information is represented by linguistic Z-numbers, both components of which (object evaluation and their validity) are values of linguistic variables. Reference fuzzy identification of affiliation is based on information about the importance of the characteristics assessed by objects, formalized on the basis of a linguistic variable. The object evaluation and fuzzy reference identification were used to determine fuzzy rankings of the degree to which objects belong to clusters. The algorithm developed in the article improves the clustering algorithm presented by the author earlier, since it preserves more initial information due to a new approach to data formalization and reduces the fuzziness of rating objects, thereby reducing the risks of errors in decision support tasks.

Keywords: Z-information, clustering algorithm, rating estimate, linguistic variable

Suggested citation: Poleshchuk O.M. *Klasternyy analiz Z-informatsii na osnove etalonnoy sistemy nechetkikh opredeleniy prinadlezhnosti* [Cluster analysis of Z-information based on a reference system of fuzzy identification]. *Lesnoy vestnik / Forestry Bulletin*, 2024, vol. 28, no. 2, pp. 150–155. DOI: 10.18698/2542-1468-2024-2-150-155

Author's information

Poleshchuk Ol'ga Mitrofanovna — Dr. Sci. (Tech.), Professor, Head of Higher Mathematics and Physics Department, of the BMSTU (Mytishchi branch) poleshchuk@mgul.ac.ru

Received 07.11.2023.

Approved after review 20.12.2023.

Accepted for publication 06.02.2024.