УДК 681.518

Л.А. Демидова, С.Б. Титов

ПОДХОД К ПРОБЛЕМЕ НЕЧЕТКОЙ КЛАСТЕРИЗАЦИИ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ ВЫБОРА ЦЕЛЕВОЙ ФУНКЦИИ

Рассматривается проблема управления неопределенностью выбора целевой функции при использовании FCM и PCM алгоритмов кластеризации. Для улучшения результатов кластеризации предлагается реализовать расширение множества объектов на интервальные нечеткие множества второго типа, а для поиска оптимальной комбинации параметров алгоритма кластеризации – использовать генетический алгоритм.

Ключевые слова: *нечеткая кластеризация, FCM-алгоритм, PCM-алгоритм, интервальные нечеткие множества второго типа, целевая функция.*

Введение. Методы кластеризации на основе целевых функций используются для минимизации расстояния между образцом и прототипом кластера (таким как точка, линия, гиперэллипсоид и т.п.) и определения параметров прототипа – центра или радиуса.

Если множество объектов состоит из компактных кластеров и каждый кластер разумно отделим от других, желаемый результат кластеризации может быть получен с помощью четкого алгоритма c-средних. Однако в практических задачах множества объектов редко являются такими. Наиболее известными алгоритмами кластеризации, основанными на учете того или иного вида неопределенности, являются: алгоритм нечетких c-средних (fuzzy c-means – FCM-алгоритм), алгоритм возможностных c-средних (possibilistic c-means – PCM-алгоритм).

FCM-алгоритм позволяет получить адекватные результаты кластеризации в случае, когда множество объектов содержит пересекающиеся кластеры [1, 2]. Результаты кластеризации основаны на функциях принадлежности, использующих относительные расстояния объектов относительно центров кластеров. Однако FCM-алгоритм работает хорошо, если множество объектов содержит кластеры подобной плотности и подобного объема гиперсферической формы.

PCM-алгоритм позволяет улучшить результаты нечеткой кластеризации в случае, если множество объектов содержит атипичные объекты, за счет ослабления свойства кластерной относительности и учета свойства типичности. Однако результаты кластеризации с использованием этого алгоритма также сильно зависят от выбора фаззификатора m и значений «ширины зоны» η_j ($j=\overline{1,c}$), если кластеры в множестве объектов имеют существенно разную плотность или существенно разный объем.

Классическим алгоритмам кластеризации типа FCM-алгоритма на основе нечетких множеств первого типа – НМТ1 (или просто нечетких множеств) и PCM-алгоритма на основе НМТ1 присуща неопределенность, связанная с выбором параметров соответствующих целевых функций. Так, например, можно говорить о неопределенности фаззификатора m в FCM-алгоритме на основе НМТ1 и о неопределенности фаззификатора m и «ширины зоны» η_j ($j=\overline{1,c}$, c – количество кластеров) в PCM-алгоритме на основе НМТ1. Кроме того, может существовать неопределенность в выборе алгоритма кластеризации, а, значит, неопределенность в задании функций принадлежности (типичности), отражающих кластерную относительность (как в FCM-алгоритме на основе НМТ1) и кластерную типичность (как в PCM-алгоритме на основе НМТ1).

Расширение множества объектов кластеризации на интервальные нечеткие множества второго типа позволяет управлять неопределённостями такого рода. Алгоритмы кластеризации на основе интервальных нечетких множеств второго типа (ИНМТ2) были предложены Хвангом (Hwang C.) и Рхи (Rhee F.C.-H.) в 2007 г. [2].

Неопределенность в выборе целевой функции. Неопределенность в выборе целевой функции может быть продемонстрирована на примере целевых функций, используемых в FCM-алгоритме и PCM-алгоритме. Целевые функции в FCM-алгоритме и PCM-алгоритме основаны на свойствах кластерной относительности и кластерной типичности соответственно.

Таким образом, существует неопределенность при задании принадлежности (типичности) объекта кластеру с учетом кластерной относительности в FCM-алгоритме и кластерной типичности в PCM-алгоритме. Иногда бывает трудно определить, какой алгоритм применить для кластеризации множества объектов (FCM-алгоритм или PCM-алгоритм). Если множество объектов содержит объекты-шумы, то PCM-алгоритм может дать лучшие результаты кластеризации, чем FCM-алгоритм. Однако при применении PCM-алгоритма может иметь место частичное пересечение кластеров, размещенных близко друг к другу. Следовательно, это может привести к худшим результатам кластеризации, чем для FCM-алгоритма. Таким образом, существует необходимость в реализации метода управления неопределенностью значений степеней принадлежностей (типичности) объектов кластерам, который бы учитывал свойство кластерной относительности для FCM-алгоритма и свойство кластерной типичности для PCM-алгоритма. При разработке такого метода необходимо рассмотреть следующие свойства [2]:

1) свойство кластерной относительности: уменьшение эффекта шумовых объектов за счет соответствующего задания степеней принадлежности, которое выполняется для шумовых объектов, обладающих кластерной типичностью;

2) свойство кластерной типичности: уменьшение чувствительности при инициализации за счет соответствующего задания степеней принадлежности, которое учитывает кластерную относительность для предохранения от частичного пересечения кластеров, которое происходит в PCM-алгоритме.

Соответствующее управление указанными выше неопределенностями позволит улучшить результаты кластеризации на основе целевых функций.

Расширение множества объектов на интервальные нечеткие множества второго типа при комбинировании FCM-алгоритма и PCM-алгоритма. Функции принадлежности для FCM-алгоритма и функции типичности для PCM-алгоритма при неопределенности выбора целевой функции можно вычислять с использованием относительного и абсолютного расстояний для объектов и центров кластеров в соответствии с нижеприведенными формулами:

$$u_j(x_i) = \frac{1}{\sum\limits_{t=1}^{c}\left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m-1}}}, \tag{1}$$

$$w_j(x_i) = \frac{1}{1+\left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m-1}}}. \tag{2}$$

При этом выполняется расширение множества объектов на ИНМТ2.

При определении интервальных первичных функций принадлежности объекта x_i для FCM-PCM-алгоритма на основе ИНМТ2 (при комбинировании FCM-алгоритма и PCM-алгоритма на основе ИНМТ2) «нижняя» и «верхняя» интервальные функции принадлежности могут быть представлены с использованием формул (1) и (2) [2]:

$$\bar{u}_j(x_i) = \begin{cases} \dfrac{1}{\sum\limits_{t=1}^{c}\left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m-1}}}, \text{если } \dfrac{1}{\sum\limits_{t=1}^{c}\left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m-1}}} > \dfrac{1}{1+\left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m-1}}} \\ \dfrac{1}{1+\left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m-1}}}, \text{если } \dfrac{1}{\sum\limits_{t=1}^{c}\left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m-1}}} \le \dfrac{1}{1+\left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m-1}}} \end{cases}, \tag{3}$$

$$\underline{u}_j(x_i) = \begin{cases} \dfrac{1}{\sum\limits_{t=1}^{c}\left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m-1}}}, \text{если } \dfrac{1}{\sum\limits_{t=1}^{c}\left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m-1}}} \le \dfrac{1}{1+\left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m-1}}} \\ \dfrac{1}{1+\left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m-1}}}, \text{если } \dfrac{1}{\sum\limits_{t=1}^{c}\left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m-1}}} > \dfrac{1}{1+\left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m-1}}} \end{cases}. \tag{4}$$

В качестве целевых функций следует рассматривать функции, определяемые с помощью следующих формул.

$$J(U,V) = \sum_{j=1}^{c}\sum_{i=1}^{n}\left(u_j(x_i)\right)^m \cdot d_{ji}^2, \tag{5}$$

$$J(W,V) = \sum_{j=1}^{c}\sum_{i=1}^{n}\left(w_j(x_i)\right)^m \cdot d_{ji}^2 + \\ +\sum_{j=1}^{\tilde{n}}\eta_j^2 \cdot \sum_{i=1}^{n}\left(1-w_j(x_i)\right)^m. \tag{6}$$

При этом в обеих целевых функциях фаззификатор m имеет одно и то же значение.

На рисунках 1, 2 и 3 приведены примеры ИНМТ2 в соответствии с формулами (3) и (4) для случая двух кластеров при $(m,\eta_j)=(2,\ 0{,}25)$, $(m,\eta_j)=(2,\ 0{,}5)$ и $(m,\eta_j)=(2,\ 0{,}75)$ соответственно.

Управление неопределенностью фаззификатора m осуществляется с помощью вычисления центров кластеров и получения четкого разбиения для принятия конечного решения о результатах кластеризации [2].

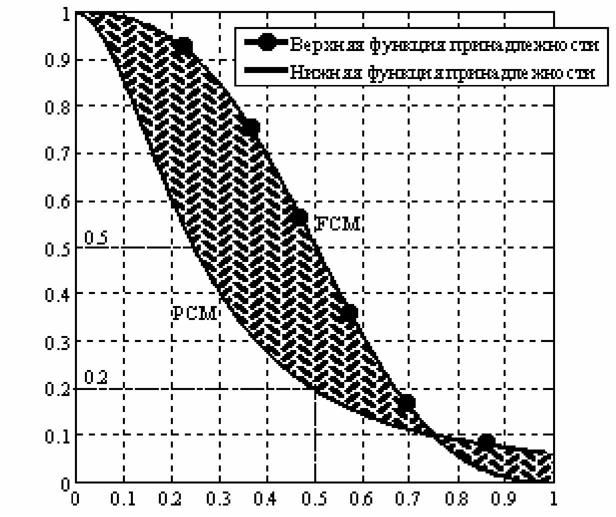

Рисунок 1 – «Отпечаток неопределенности» для FCM-PCM-алгоритма ($m = 2$ и $\eta_j = 0{,}25$)

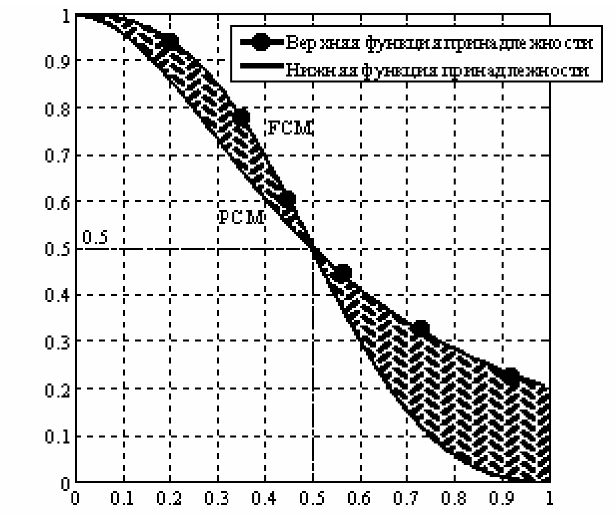

Рисунок 2 – «Отпечаток неопределенности» для FCM-PCM-алгоритма ($m = 2$ и $\eta_j = 0{,}5$)

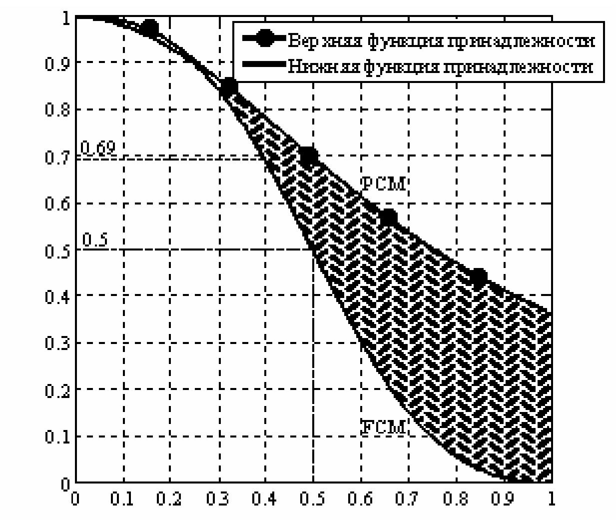

Рисунок 3 – «Отпечаток неопределенности» для FCM-PCM-алгоритма ($m = 2$ и $\eta_j = 0{,}75$)

Центроид НМТ1 для n объектов может быть вычислен по формуле:

$$v_X = \sum_{i=1}^{n}\left(x_i \cdot u(x_i)\right) \Big/ \sum_{i=1}^{n} u(x_i). \qquad (7)$$

По принципу расширения центроид НМТ2 $\widetilde{X}$ вычисляется как:

$$v_{\widetilde{X}} = \sum_{u(x_1)\in J_{x_1}} \ldots \sum_{u(x_n)\in J_{x_n}} \frac{f(u(x_1))\cdot \ldots \cdot f(u(x_n))}{\left(\left(\sum_{i=1}^{n} x_i \cdot u(x_i)\right) \Big/ \sum_{i=1}^{n} u(x_i)\right)}. \quad (8)$$

Для ИНМТ2 в формуле (8) можно заменить все $f(u(x_i))$ ($i = \overline{1,n}$) на 1 и использовать эту формулу для нечеткой кластеризации, введя дополнительно в формулу нечеткую степень m:

$$v_{\widetilde{X}} = \sum_{u(x_1)\in J_{x_1}} \ldots \sum_{u(x_n)\in J_{x_n}} 1 \Bigg/ \left(\sum_{i=1}^{n} x_i \cdot u(x_i)^m \Big/ \sum_{i=1}^{n} u(x_i)^m\right). \quad (9)$$

При оценке центров кластеров целесообразно использовать итерационный алгоритм Карника-Менделя [2].

Итерационный алгоритм Карника-Менделя. Алгоритм Карника-Менделя позволяет определить два «вложенных» НМТ1 – L и R° – внутри FOU ИНМТ2 $\widetilde{X}$. Множества L и R имеют минимально и максимально возможный центроиды v_L и v_R в $\widetilde{X}$ соответственно. Пусть для каждого из объектов кластеризации x_i ($i = \overline{1,n}$) заданы значения всех q характеристик, измеренные в некоторой количественной шкале: $x_i = (x_{i1},\ldots,x_{iq})$.

Алгоритм Карника-Менделя для поиска максимума v_R центра кластера v_j имеет вид:

1. Для множества объектов $x_i = (x_{i1},\ldots,x_{iq})$ ($i = \overline{1,n}$) вычисляются функции принадлежности в соответствии с формулами (10) и (11).

$$\overline{u}_j(x_i) = \begin{cases} u_j^1(x_i), & \text{если } u_j^1(x_i) > u_j^2(x_i) \\ u_j^2(x_i), & \text{если } u_j^1(x_i) \le u_j^2(x_i) \end{cases}, \quad (10)$$

$$\underline{u}_j(x_i) = \begin{cases} u_j^1(x_i), & \text{если } u_j^1(x_i) \le u_j^2(x_i) \\ u_j^2(x_i), & \text{если } u_j^1(x_i) > u_j^2(x_i) \end{cases}, \quad (11)$$

где $u_j^p(x_i) = 1 \Big/ \sum_{k=1}^{c}\left(d_{ji}/d_{ki}\right)^{\frac{2}{m_p - 1}}$, $\mathrm{p} = 1,2$.

2. Выбирается значение фаззификатора m (любое из m_1 и m_2).

3. Вычисляется центроид $v'_j = (v'_{j1},\ldots,v'_{jq})$ по формуле (7) и $u_j(x_i) = \left(\overline{u}_j(x_i) + \underline{u}_j(x_i)\right)/2$.

4. Выполняется сортировка индексов n объектов ($i = \overline{1,n}$) по каждой характеристике l ($l = \overline{1,q}$) по возрастанию: $x_{11},\ldots,x_{n1}$; …; $x_{1q},\ldots,x_{nq}$.

5. По каждой характеристике l ($l=\overline{1,q}$) выполняется поиск индекса k ($1 \le k \le n-1$) такого, что $x_{kl} \le v'_{jl} \le x_{(k+1)l}$.

6. Для всех n объектов вычисляются функция принадлежности: если $i \le k$, то $u_j(x_i)=\underline{u}_j(x_i)$; иначе – $u_j(x_i)=\overline{u}_j(x_i)$.

7. Вычисляется центроид для v''_{jl} по формуле (9).

8. Если $v'_{jl}=v''_{jl}$, то l-я координата центра j-го кластера считается вычисленной и осуществляется переход к шагу 9. В противном случае полагается, что $v'_{jl}=v''_{jl}$, и осуществляется переход к шагу 5 для уточнения l-й координаты j-го кластера.

9. Если $l<q$, то номер координаты увеличивается на единицу: $l=l+1$ и осуществляется переход к шагу 5 для уточнения l-й координаты j-го кластера. Если $l=q$, считается, что процедура вычисления максимума v_R центра кластера v_j завершена и $v_R=v'_j=(v'_{j1},...,v'_{jq})$.

Минимум v_L центра кластера v_j вычисляется аналогичным образом с заменой действий шага 6 на следующее: если $i \le k$, то $u_j(x_i)=\overline{u}_j(x_i)$; иначе – $u_j(x_i)=\underline{u}_j(x_i)$.

Результирующее интервальное НМТ1 может быть записано в виде: $v_j=1{,}0/[v_L,v_R]$. Четкое значение для центра кластера v_j находится с помощью операции «понижения типа» [2]:

$$v_j=(v_L+v_R)/2. \tag{12}$$

«Четкое разбиение» в FCM-алгоритме на основе ИНМТ2 находится в соответствии с правилом:

«Если ($u_j(x_i)>u_t(x_i)$) для $t=\overline{1,...,c}$ и $j \ne t$, (13)

то x_i относится к кластеру j».

Перед выполнением «четкого разбиения» необходимо выполнить «понижение типа» для функций принадлежности объектов $u_j(x_i)$, используя левые и правые функции принадлежности ($u_j^L(x_i)$ и $u_j^R(x_i)$). «Понижение типа» может быть выполнено как:

$$u_j(x_i)=\left(u_j^R(x_i)+u_j^L(x_i)\right)/2,\ j=\overline{1,...,c}, \tag{14}$$

$$u_j^R(x_i)=\sum_{l=1}^{q}u_{jl}(x_i)\Big/q,$$

$$u_{jl}(x_i)=\begin{cases}\overline{u}_j(x_i),\ \text{если } x_{il} \text{ использует } \overline{u}_j(x_i) \text{ для } v_j^R\\ \underline{u}_j(x_i)-\text{иначе,}\end{cases}$$

$$u_j^L(x_i)=\sum_{l=1}^{q}u_{jl}(x_i)\Big/q,$$

$$u_{jl}(x_i)=\begin{cases}\overline{u}_j(x_i),\ \text{если } x_{il} \text{ использует } \overline{u}_j(x_i) \text{ для } v_j^L\\ u_j(x_i)-\text{иначе.}\end{cases}$$

Генетический алгоритм поиска оптимальной комбинации значений фаззификатора и «ширины зоны» в алгоритме кластеризации на основе интервальных нечетких множеств второго типа. Метод кластеризации, реализующий FCM-PCM-алгоритм на основе ИНМТ2, учитывающий одновременно свойства кластерной относительности и кластерной типичности, и генетический алгоритм (ГА), позволяет значительно сократить время поиска оптимальной комбинации значений параметров алгоритма кластеризации и обеспечить получение адекватных результатов кластеризации.

При реализации FCM-PCM-алгоритма на основе ИНМТ2 возникает задача поиска оптимальной комбинации значения фаззификатора m и значений «ширины зоны» η_j ($j=\overline{1,c}$).

В этом случае хромосома может быть представлена в виде:

$$s=(m,\eta_1,...,\eta_c), \tag{15}$$

где $m \in (1,\ m_{max}]$; m_{max} – некоторое действительное число, определяющее максимальное значение фаззификатора; η_j – «ширина зоны» j-го кластера ($j=\overline{1,c}$); $\eta_j \in [\eta_j^{min},\eta_j^{max}]$; $\eta_j^{min}>0$, $\eta_j^{max}>0$, $\eta_j^{min}<\eta_{max}$, $\eta_j^{max}\le\eta_{max}$, $\eta_j^{min}<\eta_j^{max}$, η_{max} – некоторое действительное число, определяющее максимальное значение «ширины зоны».

В качестве функции соответствия для ГА в общем случае следует использовать общий гиперобъем H, а в частном случае – для множества объектов, содержащего кластеры гиперсферической формы, – индекс Sph [3].

Общий гиперобъем H рассчитывается по следующей формуле:

$$H=\sum_{j=1}^{c}\left(det(R_j)\right)^{\frac{1}{2}} \to min, \tag{16}$$

$$R_j=\frac{1}{n_j}\cdot\sum_{i=1}^{n}(x_i-v_j)\cdot(x_i-v_j)^T, \tag{17}$$

где R_j – ковариационная матрица j-го кластера; n_j – количество объектов кластеризации, отнесенных к j-му кластеру; v_j – вектор координат центра j-го кластера; x_i – вектор координат (оценок по критериям) i-го объекта; $det(R_j)$ – определитель ковариационной матрицы j-го кластера; n – количество объектов; c – количество кластеров; $i=\overline{1,n}$, $j=\overline{1,c}$.

При реализации ГА следует использовать одноточечное скрещивание, а количество мутирующих генов не должно превышать 10-20 % от длины хромосомы. Выбор хромосом-родителей для выполнения операции скрещивания осуществляется с использованием алгоритма вероятностного отбора. При выполнении операции скрещивания выбирается вероятность скрещивания R_c и генерируется случайное число N_c. Если $R_c > N_c$, то случайным образом выбирается точка скрещивания z и выполняется скрещивание. При выполнении операции мутации выбирается вероятность мутации R_m и генерируется случайное число N_m. Если $R_m > N_m$, то случайным образом выбирается точка мутации z и выполняется мутация. При реализации данного ГА не требуется осуществлять проверку каких-либо условий при выполнении операций скрещивания и мутации.

ГА поиска оптимальной комбинации значения фаззификатора m и значений «ширины зоны» η_j ($j=\overline{1,c}$) имеет следующий вид.

1) Случайным образом создается популяция размером P.

2) При $g<G$ (G и g – максимальное и текущее количество поколений ГА соответственно) реализуется FCM-PCM-алгоритм на основе ИНМТ2 и вычисляется значение функции соответствия по формуле (16) для каждой хромосомы и создается $P/2$ пар хромосом-родителей.

3) Выполняются операции скрещивания и мутации для текущей популяции. Для хромосом-отпрысков реализуется FCM-PCM-алгоритм на основе ИНМТ2 и вычисляются значения функции соответствия по формуле (16).

4) Создается новая популяция размером $(P+R_c \cdot P)$, дополненная хромосомами-отпрысками в количестве $R_c \cdot P$, затем $R_c \cdot P$ хромосом с худшими значениями функции соответствия по формуле (16) отбрасываются. Если $g<G$, осуществляется переход к шагу 2.

5) Выбирается лучшая хромосома, которая минимизирует функцию соответствия по формуле (16). Для каждого объекта определяется его принадлежность к кластерам.

При реализации ГА одновременно с популяцией хромосом вида (15) существуют «популяции» значений функции соответствия по формуле (16), координат центров кластеров и степеней принадлежности объектов центрам кластеров.

Экспериментальные результаты. Ниже приведен пример кластеризации множества объектов на три кластера с использованием FCM-PCM-алгоритма на основе ИНМТ2 и ГА поиска оптимальной комбинации значений фаззификатора m и значений «ширины зоны» η_j ($j=\overline{1,3}$).

На рисунке 4 показано множество объектов, содержащее три кластера существенно разного объема (объекты разных кластеров помечены маркерами разной формы). Кластеры представляют собой множества объектов, координаты которых были сгенерированы с использованием нормального закона распределения с центрами (10, 50), (50, 50) и (90, 50).

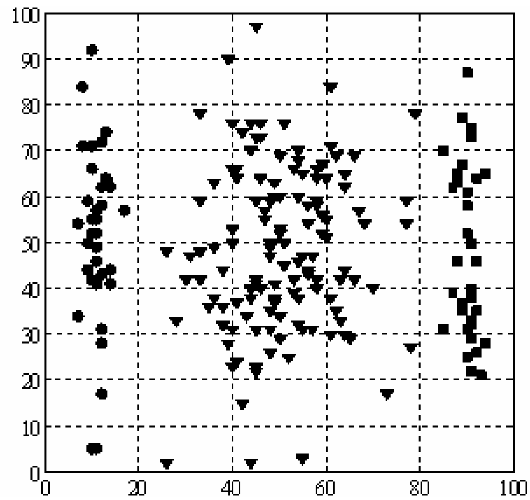

Рисунок 4 – Исходное множество объектов кластеризации

При этом при генерации первой и второй координат объектов первого и третьего кластеров использовались нормальные законы распределения, имеющие одинаковые дисперсии.

Координаты объектов второго кластера имеют существенно большую дисперсию по обеим координатам. Первый и второй кластеры содержат по 35 объектов, а второй кластер – 130 объектов.

При использовании FCM-алгоритма на основе НМТ1 были ошибочно классифицированы 13 объектов (6,5 % от мощности множества объектов кластеризации) (рисунок 5). Применение FCM-PCM-алгоритма на основе ИНМТ2 позволило улучшить результаты кластеризации: в этом случае были ошибочно классифицированы только четыре объекта (2 % от мощности множества объектов кластеризации) (рисунок 6). На рисунках 5 и 6 центры кластеров отмечены «белыми» маркерами треугольной формы.

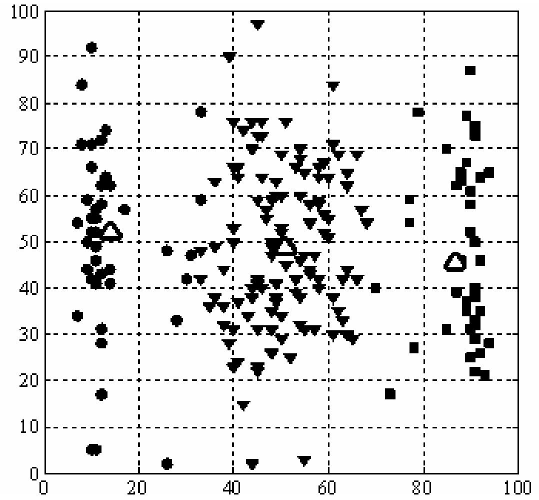

Рисунок 5 – Результаты кластеризации с использованием FCM-алгоритма на основе НМТ1 при $m = 2$
Центры кластеров:
первый кластер: (13,854673, 51,999137)
второй кластер: (50,718757, 48,734540)
третий кластер: (86,748131, 45,534273)

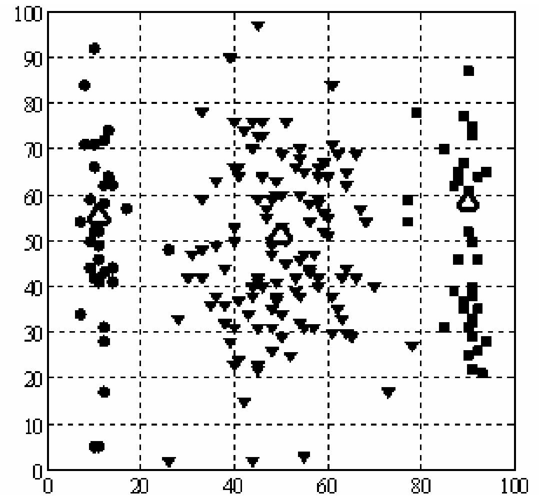

Рисунок 6 – Результаты кластеризации для FCM-PCM-алгоритма на основе ИНМТ2
($m = 135{,}794039$ **и** $\eta_1 = 78{,}737059$;
$\eta_2 = 122{,}235233$; $\eta_3 = 75{,}345363$)
Центры кластеров:
первый кластер: (11,013806, 54,996726)
второй кластер: (49,771084, 50,961819)
третий кластер: (89,981587, 57,994123)

В таблице приведены результаты кластеризации с использованием алгоритмов кластеризации на основе НМТ1 и ИНМТ2, подтверждающие эффективность использования алгоритмов кластеризации на основе ИНМТ2 для множеств объектов существенно разной плотности или существенно разного объема.

Результаты кластеризации с использованием различных алгоритмов кластеризации

Алгоритм кластеризации	Значение функции соответствия	Ошибочно классифицированные объекты
–	(для трех сгенерированных кластеров) 279,725273	0 объектов
Алгоритм четких c -средних	350,616421	9 объектов (4,5%)
FCM на основе НМТ1 при $m = 2$	(нечеткий общий гиперобъем) 541,268379	13 объектов (6,5 %)
FCM на основе ИНМТ2	330,143524	5 объектов (2,5 %)
PCM на основе ИНМТ2 (при фиксированных значениях фаззификаторов m_1 и m_2, определенных с помощью FCM-алгоритма)	327,775548	5 объектов (2,5 %)
FCM-PCM на основе ИНМТ2	319,708034	4 объекта (2 %)

При использовании алгоритмов кластеризации на основе ИНМТ2 удалось уменьшить количество ошибочно классифицированных объектов с минимального количества в 9 объектов (4,5 % от мощности множества объектов) для четкого алгоритма c -средних до 4-5 объектов (1,5-2,5 % от мощности множества объектов) для алгоритмов кластеризации на основе ИНМТ2.

Заключение. В статье рассмотрена проблема управления неопределенностью выбора целевой функции при использовании FCM- и PCM-алгоритмов кластеризации множества объектов. Предложен метод кластеризации, разработанный для FCM-PCM-алгоритма на основе ИНМТ2, который позволяет учесть одновременно свойства кластерной относительности и кластерной типичности объектов и в ряде случаев улучшить результаты кластеризации.

Применение ГА позволяет найти оптимальную комбинацию значений фаззификатора и «ширины зоны» в FCM-PCM-алгоритме на основе ИНМТ2, обеспечивающую лучшие результаты кластеризации, что подтверждается минимальным значением функции соответствия H .

На практике предлагаемый метод кластеризации может быть применен, например, для решения задачи классификации объектов недвижимости [4, 5].

Библиографический список

1. *Демидова Л.А., Кираковский В.В., Пылькин А.Н.* Алгоритмы и системы нечеткого вывода в задачах диагностики городских инженерных коммуникаций. – М.: Радио и связь, Горячая линия – Телеком, 2005. 592 с., ил.

2. *Hwang C., Rhee F.C-.H.* Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means // IEEE Trans. on Fuzzy Systems. 2007. vol. 15. № 1. – P. 107–120.

3. *Демидова Л.А., Кираковский В.В.* Методы кластеризации объектов на основе нечетких множеств второго типа и генетического алгоритма. // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. Информатика. Телекоммуникации. Управление. – СПб., 2008. – № 6 (69). – С. 136–142.

4. *Титов С.Б.* Применение теории нечетких множеств второго типа к решению задачи классификации жилой недвижимости. // Проблемы информатики в образовании, управлении, экономике и технике: материалы VII Всероссийской научно технической конференции. – Пенза: Пензенский государственный педагогический университет им. В.Г. Белинского. – 2008. – С.263–265.

5. *Титов С.Б., Демидова Л.А.* Классификация объектов жилой недвижимости с использованием алгоритма нечетких c-средних на основе нечетких множеств второго типа. // Математическое и программное обеспечение вычислительных систем: материалы межвузовского сборника научных трудов. – Рязань: Рязанский государственный радиотехнический университет. – 2008. – С. 129–135.