НЕЙРОІНФОРМАТИКА ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ

НЕЙРОИНФОРМАТИКА И ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

NEUROINFORMATICS AND INTELLIGENT SYSTEMS

УДК 519.71

Кучеренко Е. И.[1], Глушенкова И. С.[2], Глушенков С. А.[3]

[1]Д-р техн. наук, профессор, профессор кафедры искусственного интеллекта Харьковского национального университета радиоэлектроники, Харьков, Украина

[2]Канд. техн. наук, доцент кафедры геоинформационных систем, оценки земли и недвижимого имущества Харьковского национального университета городского хозяйства имени А.Н. Бекетова, Харьков, Украина

[3]Аспирант кафедры геоинформационных систем, оценки земли и недвижимого имущества Харьковского национального университета городского хозяйства имени А.Н. Бекетова, Харьков, Украина

НЕЧЕТКОЕ РАЗБИЕНИЕ ОБЪЕКТОВ НА ОСНОВЕ КРИТЕРИЕВ ПЛОТНОСТИ

Решена задача разбиения по критериям плотности в нечетком пространстве состояний при пересечении признаков. Объектом исследования являлись процессы разбиения заданной выборки объектов на подмножества. Предмет исследования составляют методы и алгоритмы нечеткого разбиения объектов на основе критериев плотности в сложных системах. Цель работы: развитие метода горной кластеризации Ягера-Филева на основе нечетких представлений для повышения эффективности решений. Предложен нечеткий метод разбиения, основанный на вычислении плотности распределения интегральных признаков объектов в нечетком пространстве состояний, который, в отличие от существующих, дополнительно функционирует в нечетком пространстве состояний и признаков. Описаны и обоснованы этапы метода нечеткого разбиения признаков с использованием нечеткого расстояния Хемминга. Была создана программа моделирования плотности распределения признаков на основе разработанного метода. Выполнен эксперимент по определению принадлежности объекта при пересечении областей нечеткого распределения признаков и предоставление результатов в виде логического вывода и графического материала. Результаты эксперимента позволяют рекомендовать предложенный метод для использования на практике. Перспективой дальнейших исследований является исследование и алгоритмизация метода, его адаптация в пространстве признаков предметных областей.

Ключевые слова: кластеризация, расстояние Хемминга, горная кластеризация, нечеткая логика.

НОМЕНКЛАТУРА

ГИС – геоинформационные системы;
C_i – размер множества разбиения;
D – детерминированное состояние;
$\{d_{ij}\}$ – множество расстояний;
E – ребро графа;
$\widetilde{F}$ – нечеткое состояние;
G^t – граф;
Int – интегральное распределение признаков;
$\{K_i\}$ – множество разбиений;
m – количество уровней иерархии;
$\{O_{ij}\}$ – множество объектов;
$\{O_{\alpha}, O_{\beta}\}$ – объекты локализованного пространства;
$P(x,y)$ – вероятностное распределение плотностей;
$\{p_{ij\alpha}\}$ – параметры множества;
R_i – радиус множества разбиения;
$S_{ij\beta}$ – площадь распределения плотности;
U – матрица распределения плотностей признаков по объектам;
Z_i – центры множества разбиения;
$\{a_{ij}\}$ – пространственные координаты;
$\Delta\mu(S_{ij\beta})$ – величина дискретизации;
$(\delta(\widetilde{O}_{\alpha}), \delta(\widetilde{O}_{\beta}))$ – относительное расстояние Хемминга;
δ_i – расстояние Хемминга;
$\eta(\widetilde{O})$ – квадратичный индекс нечеткости;
$\mu_{o_{\alpha}}, \mu_{o\beta}$ – множество значений функции принадлежности;
$\mu_{\widetilde{A}_i}$ – функция принадлежности;
$\mu(S_{ij\beta})$ – функция принадлежности плотности распределения;

DOI 10.15588/1607-3274-2016-1-4

ρ_α – плотность распределения признаков;

$\{\rho_{ij\beta}\}$ – множество распределения плотности признаков.

ВВЕДЕНИЕ

Важным аспектом классификации объектов является представление, структурирование и анализ огромных массивов информации, которые составляют основу функционирования и развития сложных систем. При анализе многомерных распределенных объектов требуются универсальные и надежные подходы, направленные на минимизацию критериев на множестве ограничений предметной области. Особенно это актуально при реализации геоинформационных систем (ГИС) и технологий.

Проблема принятия решений в таких системах является не тривиальной задачей [1] и характеризуется неопределенностью, которая может быть снижена за счет применения нечетких (fuzzy) знание-ориентированных технологий. Существующие решения [2] на основе построения кластеров являются актуальными в ГИС и технологиях. Однако наличие свойств пересечения кластеров часто приводит к трудностям классификации объектов производственных систем, что приводит к технологиям нечеткого разбиения пространства по заданным критериям.

Целью настоящих исследований является разработка и совершенствование подходов к оптимизации и классификации таких объектов на основе развития методов и алгоритмов нечеткой кластеризации и разбиения, а также систем на их основе.

1 ПОСТАНОВКА ЗАДАЧИ

Кластерный анализ – это задача разбиения заданной выборки объектов на подмножества, называемые кластерами, так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались [1]. Кластерный анализ представляет собой многомерную статистическую процедуру, выполняющую сбор данных, содержащих информацию о выборке объектов, и упорядочивающую объекты в сравнительно однородные группы нечетких разбиений. Множество разбиений $\{K_i\}, i \in I$ характеризуется:

– центрами $Z_i, i \in I$;

– размером C_i – $\{K_i\}, i \in I \,|\, C_i \neq \varnothing$;

– радиусами R_i – $\{K_i\}, i \in I \,|\, R_i \neq \varnothing$;

– множеством объектов $\{O_{ij}\}, j \in J$;

– множеством расстояний между кластерами $\{d_{ij}\}$, причем:

$$d(\widetilde{A}, \widetilde{A}) = 0;$$
$$d(\widetilde{A}, \widetilde{B}) = d(\widetilde{B}, \widetilde{A});$$
$$d(\widetilde{A}, \widetilde{C}) \le d(\widetilde{A}, \widetilde{B}) \circ d(\widetilde{B}, \widetilde{C});$$
$$d(\widetilde{A}, \widetilde{B}) \ge 0.$$

Множество объектов $\{O_{ij}\}, j \in J$ характеризуется множествами признаков распределения плотностей $\{\rho_{ij\beta}\}, \beta \in \mathrm{B}$, пространственных координат $\{a_{ij}\}$ и параметров $\{p_{ij\alpha}\}, \alpha \in \mathrm{A}$, причем $O_{ij} \in \{O_{ij}\},\ j \in J \,|\, \{a_{ij}\} \neq \varnothing, \{p_{ij}\} \neq \varnothing$.

Необходимо предложить методы и алгоритмы нечеткого разбиения объектов на основе критериев плотности в сложных системах и технологиях пространственно распределенных объектов предметных областей.

2 ОБЗОР ЛИТЕРАТУРЫ

Применение кластерного анализа в общем виде сводится к следующим этапам [3]:

– выборка объектов для кластеризации;

– определение множества переменных, по которым будут оцениваться объекты в выборке с нормализацией значений переменных;

– вычисление значений меры сходства между объектами;

– применение метода кластерного анализа для создания групп сходных объектов (кластеров);

– представление и интерпретация результатов анализа.

Для каждой пары объектов измеряется «расстояние» между ними – степень схожести. Существует множество метрик [3]: эвклидово расстояние; квадрат эвклидова расстояния; манхэттенское расстояние или расстояние Хемминга; расстояние Чебышева; степенное расстояние, которое является модификацией расстояния Эвклида.

Существуют следующие методы и алгоритмы кластеризации [3] (табл. 1):

– алгоритмы иерархической кластеризации;

– алгоритмы квадратичной ошибки;

– нечеткие алгоритмы;

– алгоритмы, основанные на теории графов;

Таблица 1 – Методы и алгоритмы кластеризации

Методы и алгоритмы кластеризации	Положительные качества	Отрицательные качества
алгоритмы иерархической кластеризации	представление результата в виде дендрограммы	необходима система полных разбиений
алгоритмы квадратичной ошибки	минимизация среднеквадратической ошибки разбиения	требуется задание количества кластеров
нечеткие алгоритмы	мягкое разбиение на кластеры	необходимо заранее знать количество кластеров
алгоритмы, основанные на теории графов	наглядность и возможность внесения различных модификаций	сложность подбора значащих коэффициентов
алгоритм выделения связных компонент	наглядность и возможность внесения различных модификаций	трудности управления количеством кластеров при помощи порога расстояния
алгоритм минимального покрывающего дерева	наглядность и возможность внесения различных модификаций	трудности управления количеством кластеров при помощи порога расстояния
послойная кластеризация	наглядность и возможность внесения различных модификаций	трудности управления количеством кластеров при помощи порога расстояния

– алгоритм выделения связных компонент;
– алгоритм минимального покрывающего дерева;
– послойная кластеризация.

Наличие множества методов и алгоритмов кластеризации не охватывает всей совокупности подходов и особенностей распределения признаков объектов, что свойственно распределению различной природы и требует дальнейших исследований.

3 МАТЕРИАЛЫ И МЕТОДЫ

Предлагаемые решения являются дальнейшим развитием метода горной кластеризации Ягера-Филева [4] на основе нечетких представлений.

Утверждение 1. Если задано множество площадей распределения плотностей признаков $\{\rho_{ij\beta}\}, \beta \in B$, пространственных координат $\{a_{ij}\}$ и параметров $\{p_{ij\alpha}\}, \alpha \in A$, а также задана плотность распределения признаков $\{\rho_{ij\beta}\}, \beta \in B$ в виде $\rho_\alpha, \alpha \in A$, для которых характерно $S_{ij}, i \in I, j \in J$, тогда критерием поиска плотности распределения признаков может быть

$$\rho_\alpha = \frac{\sum\limits_{i \in I,\, j \in J} \rho_{ij\beta} S_{ij}}{\sum\limits_{i \in I,\, j \in J} S_{ij}} \mu(S_{ij\beta}),\ \rho_{ij} > \rho^*, \qquad (1)$$

где ρ^* – порог плотности, $\mu(S_{ij\beta})$ – функция принадлежности плотности распределения (ρ_{ij}) по некоторой площади из $S_{ij\beta}$.

Особенностью метода является то, что он не требует задания количества кластеров и при числе признаков $(n = 2)$ поверхность распределения близка к горному рельефу. Кластеризация по горному методу не является нечеткой, однако, ее часто используют при синтезе нечетких правил на знаниях [4]. Особенностью горной кластеризации является следующее:

– на первом шаге горной кластеризации определяют точки, которые могут быть центрами кластеров;

– на втором шаге для каждой такой точки рассчитывается значение потенциала, показывающего возможность формирования кластера в ее окрестности. Плотность расположения объектов в окрестности потенциального центра кластера является функцией от значения его потенциала;

– на третьем шаге итерационно выбираются центры кластеров среди точек с максимальными потенциалами. Формируются также объекты кластеризации.

Важной особенностью, накладывающей ограничения на применение метода, является отсутствие возможности решений в нечетком пространстве состояний. Анализ возможных решений по совершенствованию вышеизложенного метода [4] позволил сформулировать следующее предположение: на первом этапе построение решетчатого гиперкуба следует дополнить функцией принадлежности распределения $\mu(S_{ij\beta})$.

Тогда этапы нового метода нечеткого разбиения признаков могут быть такие.

Этап 1. Формируем интегральное распределение признаков

$$Int = \frac{\sum\limits_{i \in I, j \in J, \beta \in B} \rho_{ij\beta} S_{ij\beta}}{\sum\limits_{i \in I, j \in J, \beta \in B} S_{ij\beta}}, \qquad (2)$$

исключаем из рассмотрения $\rho_{ij\beta}$, для которых $|I| \wedge |J| \wedge |B| = 0$, и уточняем интегральные признаки (2)

$$Int' = \frac{\sum\limits_{i \in I, j \in J, \beta \in B} \rho_{ij\beta} S_{ij\beta}}{\sum\limits_{i \in I, j \in J, \beta \in B} S_{ij\beta}}. \qquad (3)$$

Этап 2. Формируем область $S_{ij\beta_x}$, для которой значение квадратичного индекса нечеткости [5] принимает вид

$$\eta(\widetilde{O}) = \frac{2}{\sqrt{n}} e(\widetilde{O}, \overline{\overline{\widetilde{O}}}) \xrightarrow{a_{ij}} \min. \qquad (4)$$

Искомое значение (4) определяем в качестве первого с максимальной плотностью признаков потенциального центра разбиения $\{K_i\}, i = 1 | C_i \neq \varnothing$.

Этап 3. Формируем потенциальные центры разбиения. Для этого диапазоны изменения входных признаков разбиваем на n интервалов согласно (1), причем принимаем

$$\mu(S_{ij\beta}) = \mu(S_{i+1, j+1, \beta.}), \mu(S_{ij\beta}) = \exp(-k(x-u)^2), k > 0, \qquad (5)$$

где k – крутизна функции, u – центр гауссиана.

Параметры функции (5) являются элементами настройки.

Этап 4. Задав в (5) $\mu(S_{ij\beta}) = \mu(S_{ij\beta})_o$ и величину дискретизации $\Delta\mu(S_{ij\beta})$, находим итерационно, согласно (1), значения $x = x_o, ..., x_n$, таким образом, что

$$\rho_{ij} = \rho_o > \rho^*. \qquad (6)$$

Это определяет число и размер потенциальных пространств разбиения $\Delta S\alpha_{ij}$.

Этап 5. Уточняем из особенностей поверхности $\{S_{ij\beta}\}$ проекции значений $\{x_o, .., x_n\}$, $|x_o, .., x_n\}| = N$, что и определяет размер разбиения

$$C_i = x_{i+1} - x_i \qquad (7)$$

и радиус

$$R_i = \frac{x_{i+1} - x_i}{2} \qquad (8)$$

объектов.

Этап 6. Радиусы объектов уточняются на основе квадратичной нормы с учетом (6) и повторной реализацией этапов 1–5. Отметим, что $R_i \neq R_{i+1,\, C_i} \neq C_{i+1}$.

Этап 7. Осуществляем упорядочение признаков и формирование матрицы распределения плотностей признаков по объектам

$$U = \{K_i\}, i \in I \,|\, (C_i \neq \varnothing, O_{ij}, i \in I, j \in J),$$

$$if\ O_{ij} \in \{O_{ij}\}, j \in J \,|\, \{a_{ij}\} \neq \varnothing, \{p_{ij}\} \neq \varnothing. \qquad (9)$$

Этап 8. Останов.

Для реализации этапов 1–7 в знание-ориентированных технологиях целесообразно применить стратегию согласно [6], как решение

$$\{if\ x\, is\, \mu(x)\ then\ y\ is\ \mu(y)\}, \tag{10}$$

согласно

$$y' = \vee x' \wedge \mu(x,y) \tag{11}$$

с последующей (11) дефаззификацией [6].

Пусть существуют объекты $\{O_\alpha, O_\beta\} \subset \{O_{ij}\}, j \in J$ множества значений функций принадлежности, для которых

$$\mu_{o_\alpha} \cap \mu_{o_\beta} \neq \varnothing. \tag{12}$$

Существует некоторое пространство $S(x,y)$, которое принадлежит (12). Необходимо определить принадлежность пространства $S(x,y)$ одному из объектов из $\{O_\alpha, O_\beta\}$, если $\rho_{S(x,y)} \geq \rho^*$.

Утверждение 2. Если справедливо (12) и необходимо определить принадлежность пространства $S(x,y)$ одному из объектов из $\{O_\alpha, O_\beta\}$, при $\rho_{S(x,y)} \geq \rho^*$, то пространство $S(x,y)$ принадлежит одному из объектов $O_k \in \{O_\alpha, O_\beta\}$, согласно

$$(\delta(\widetilde{O}_\alpha), \delta(\widetilde{O}_\beta)) \xrightarrow{S(x,y),\, \rho_{S(x,y)} \geq \rho^*} \min. \tag{13}$$

Доказательство утверждения 2 очевидно, если принять в качестве критерия меру (13). В качестве критерия локализации пространства $S(x,y)$ на объектах $\{O_\alpha, O_\beta\}$ введем меру на основе нечеткого расстояния Хемминга [7]

$$\delta(\widetilde{O}_k, \widetilde{S}(x,y)) = \frac{d(\widetilde{O}_k, \widetilde{S}(x,y))}{n}, \tag{14}$$

которая справедлива при условии, что плоскость дискретизации функций принадлежности объектов в (14) определена в виде

$$n_{O_\alpha} = n_{O_\beta} = n. \tag{15}$$

Тогда, используя (14), (15), мы можем выявить свойство принадлежности (13) пространства $S(x,y)$ к объектам из $\{O_\alpha, O_\beta\}$.

Следствие утверждения 2. Если существует мера расстояния в виде расстояния Хемминга [14] причем

$$\{\mu_{\widetilde{A}_i}\} \cap \{\mu_{A_{i+1}}\} \neq \varnothing, \tag{16}$$

то следует рассматривать ряд случаев распределения плотностей.

Случай А. Пусть существует вероятностное распределение плотностей вещества на плоской поверхности

$$P(x,y) = e^{-k(x-a)^2}, \tag{17}$$

где a – центр гауссиана.

Определим пространство распределения плотностей на основе (17), где предложено формирование первого разбиения $S_{ij\beta_x}$ с центром разбиения

$$K_i \in \{K_i\}, i = 1 | C_i \neq \varnothing. \tag{18}$$

Тогда справедливо утверждение.

Утверждение 3. Если задана карта распределения с центрами a_1, a_2, находящимися в изолированном пространстве, и подвергающимися динамическим внешним факторам, то правило матричного отображения пространства имеет вид (19).

Приняв в (16) $a_1 = a_2$, уточнив (17), выполним анализ расположения точки $a(x,y) \in \widetilde{S}(x,y)$ согласно критерия

$$a_i \cap P(x,y) \neq \varnothing, \tag{19}$$

При нарушении (19), имеем возможное несоответствие согласно утверждения 3, что требует дальнейших исследований.

Рассмотрим случай, когда справедливо (18), (19).

Случай В. Приняв терм лингвистической переменной в виде $\mu_{A_1} \neq 0$, $\mu_{A2} \neq 0$, функции принадлежности определены в виде гауссианов:

$$\mu_{A_1} = e^{-k_1(x-a_1)^2}, k_1 > 0, \tag{20}$$

$$\mu_{A_2} = e^{-k_2(x-a_2)^2}, k_2 > 0. \tag{21}$$

Сформулируем утверждение 4.

Утверждение 4. Если существует (14), для которого справедливо (16), то нахождение минимального значения из

$$\alpha = \min(\delta(\mu_{\widetilde{A}_i}), \delta(\mu_{A_{i+1}})), i \in I \tag{22}$$

определяет принадлежность области $\widetilde{S}(x,y)$ соответствующему нечеткому разбиению при выполнении (16).

Справедливость утверждения 4 непосредственно следует из меры расстояния (14) и сущности операции пересечения функций принадлежности (16).

Используя положения (14), рассмотрим пересечение областей (20), (21), причем $k_1 > 0$, $k_2 > 0$, $k_1 < k_2$, $a_1 > a_2$, для которых справедливо (16).

Вычислениями определено, что $\delta_{A_1} < \delta_{A_2}$, тогда область $\widetilde{S}(x,y)$ принадлежит, согласно (22), разбиению (20).

Действительно, пространство $\widetilde{S}(x,y)$ попадает под влияние области с меньшим расстоянием Хемминга, что подтверждает справедливость (22) на функциях (20), (21).

Следствие 1 утверждения 4. В случае, если расстояния равны $\delta(14)$

$$\left|\{\mu_{\widetilde{A}_i}\} = \{\mu_{A_{i+1}}\}\right| \leq \varepsilon, \tag{23}$$

где ε – норма точности, то пространство $\widetilde{S}(x,y)$ не принадлежит ни к одному из источников.

Следствие 2 утверждения 4. Положения утверждения 4 справедливы при выполнении условия (6).

Тогда, учитывая положения утверждений 2–4, дополнительно к этапам 1–8, сформулируем дополнения:

5′ Если справедливо (12), то осуществляем уточнение нахождения дополнительных кластеров согласно (13)–(22);

5′′ Осуществляем контроль разбиения согласно этапов 1–7.

4 ЭКСПЕРИМЕНТЫ

Для эксперимента была создана программа моделирования плотности распределения признаков на основе разработанного метода при условии $\{\mu_{\tilde{A}_i}\} \cap \{\mu_{A_{i+1}}\} \neq \varnothing$.

UML – диаграмма классов представлена на рис. 1.

Программа PL.v.0.1, которая реализована в среде Python [8].

5 РЕЗУЛЬТАТЫ

В качестве исходных данных используются параметры, определяющие вид функции и местоположение в пространстве (рис. 2):

– координаты проекций источников признаков влияния и искомой точки на ось OX;

– значения крутизны функций принадлежности распределения плотности признаков от соответствующих источников;

– значение ограничителя в расчетах (ПДК).

ПДК – минимальное значение плотности признаков для изменения состояния искомой точки. ПДК выполняет роль логического оператора, который определяет дальнейший сценарий работы программы. Возможны 3 сценария работы программы. Сценарий 1: ПДК больше значений множеств в искомой точке, тогда точка не изменяет свое состояние и не принадлежит не к одной точке. Сценарий 2: ПДК меньше одного из значения множеств. Сценарий 3: ПДК меньше значений двух функций принадлежности в искомой точке, тогда для определения множества, к которому принадлежит точка, выполняется расчет расстояния Хемминга.

Во время работы программы выполняется расчет принадлежности точки к зонам влияния исходных множеств. В случае влияния обоих множеств на искомую точку выполняется расчет расстояния Хемминга и формирования результата путем логического вывода. Помимо выдачи текстового результата работы, программа выводит изображение с наглядным отображением всех объектов.

Таким образом, в результате эксперимента подтверждена справедливость утверждения 4 (рис. 2).

6 ОБСУЖДЕНИЕ

В работе рассмотрены методы нечеткого разбиения признаков по критерию плотности. Сформулирован метод, который является развитием метода горной кластеризации Ягера-Филева, что позволяет развивать метод на случай нечеткого пространства состояний. Дана оценка эффективности методов, основанных на вероятностном и нечетком распределении плотности признаков. Определена перспективность нечеткого разбиения по отношению к вероятностному распределению.

Действительно, используя положения (14) рассмотрено пересечение областей нечеткого распределения признаков. Вычислениями определено, что в случае, когда $\delta_{A_1} < \delta_{A_2}$, область $\widetilde{S}(x, y)$ принадлежит, согласно (22), разбиению (20).

Определено, что справедливо следствие утверждения 4 в случае, если расстояния равны (23).

Экспериментом подтверждена вычислительная сложность в виде полинома второго порядка

$$O(n) = A_0 + bx + cx^2, \tag{24}$$

для (24) справедливо

$$A_0 \neq 0, b \neq 0, c \neq 0.$$

Развитием метода является адаптация подходов к предметным областям путем дополнительного введения влияния различных внешних факторов на процесс нечеткого распределения пространства признаков.

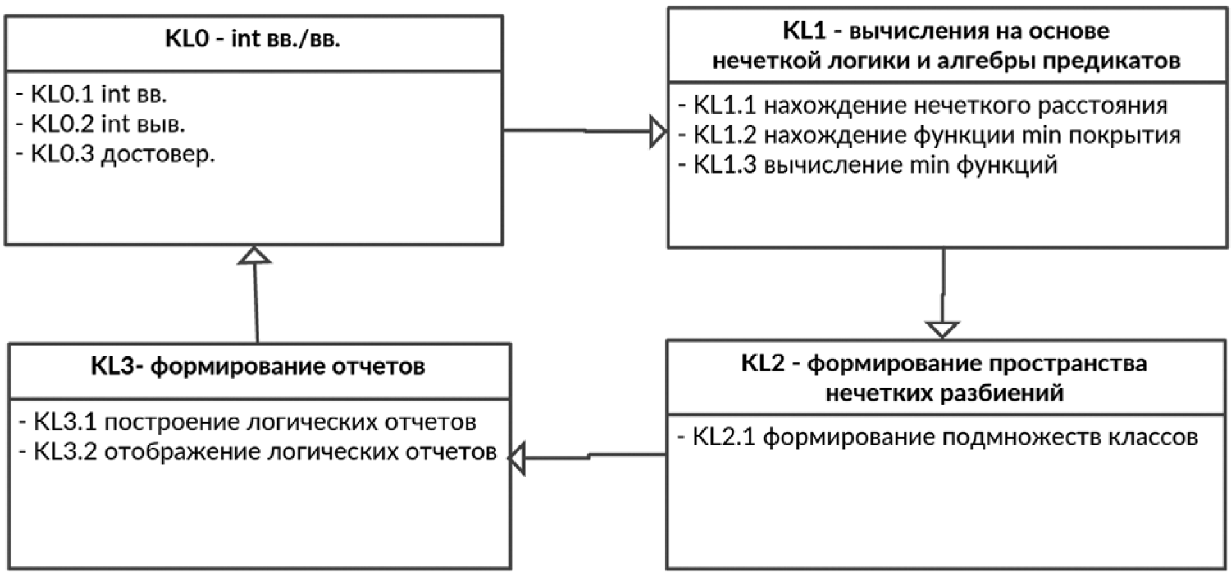

Рисунок 1 – UML диаграмма классов приложений

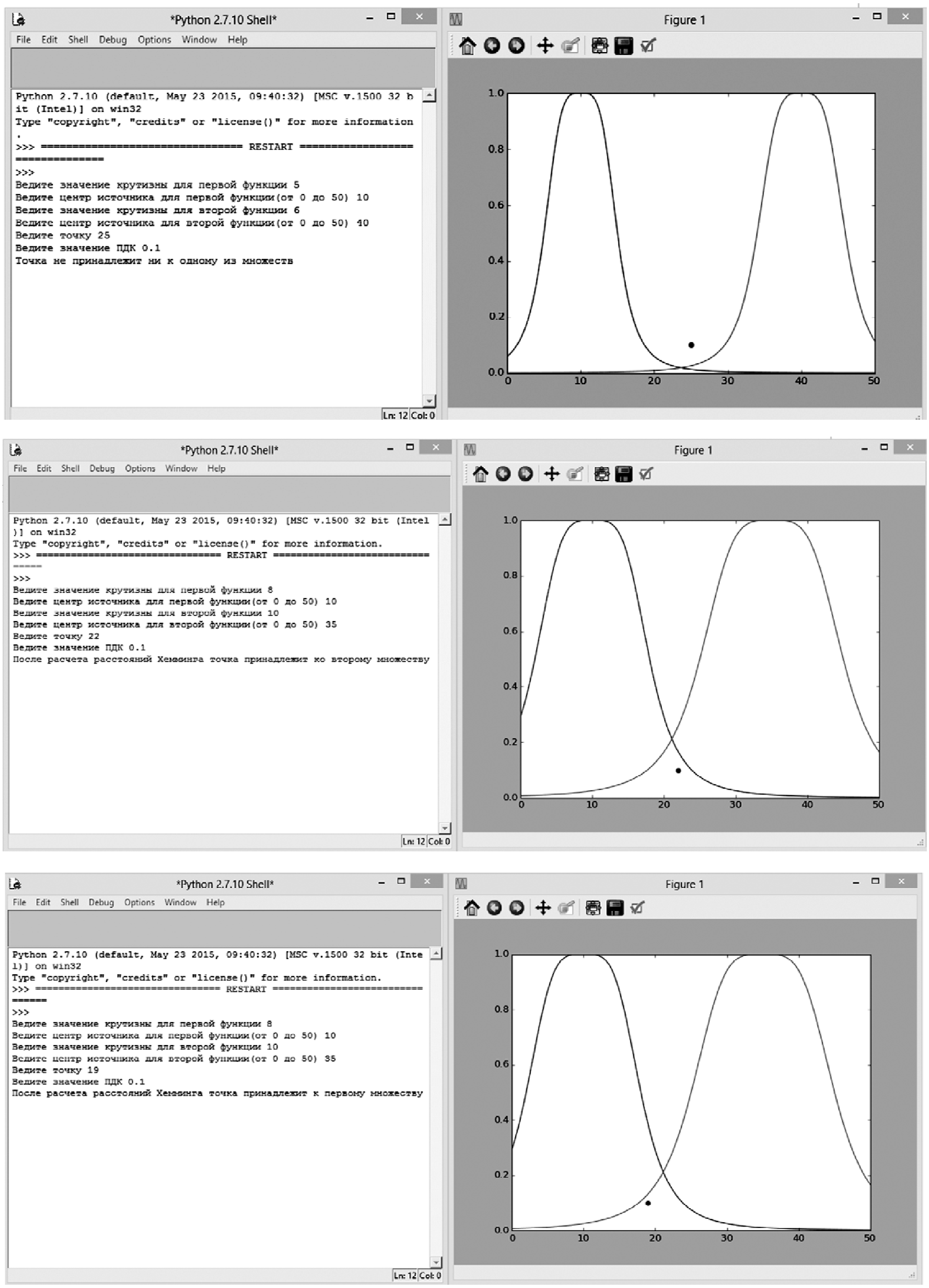

Рисунок 2 – Моделирование разбиений с различными сценариями

ВЫВОДЫ

Разработка знание-ориентированных интеллектуальных методов и моделей анализа сложных объектов является важной составляющей технологических пространственно распределенных процессов, функционирующих в условиях неопределенности. Знание-ориентированные методы направлены на моделирование и обработку детерминированных, вероятностных и нечетких знаний, как фактора повышения качества разбиений. Полученные результаты научных исследований позволили более полно, с высокой адекватностью реализовать разбиения по критериям плотности в нечетком пространстве состояний. Подход позволяет принципиально выделить нечеткую область при пересечении признаков. Таким образом, в работе предложено и исследовано:

1. Выполнен содержательный анализ методов и алгоритмов кластеризации объектов на множестве признаков. Определено, что наличие множества методов и алгоритмов кластеризации не охватывает всей совокупности подходов и особенностей распределения признаков плотности объектов, что свойственно распределению различной природы. В связи с этим важно рассмотрение подходов к нечеткому разбиению объектов на основе плотности их распределения в нечетком пространстве состояний.

2. Получил дальнейшее развитие новый нечеткий метод разбиения, основанный на вычислении плотности распределения интегральных признаков объектов в нечетком пространстве состояний, который, в отличие от существующих, дополнительно функционирует в нечетком пространстве состояний и признаков путем построения решетчатого гиперкуба с нечеткой функцией принадлежности, что позволяет рациональное разбиение признаков на множестве объектов.

3. Перспективой дальнейших исследований является исследование и алгоритмизация метода в задачах проектирования и эксплуатации систем, его адаптация в пространстве признаков предметных областей.

БЛАГОДАРНОСТИ

Работа выполнена в рамках исследований госбюджетной НИР «Нейро-фаззи системы для текущей кластеризации и классификации последовательностей данных в условиях их искривления отсутствующими и аномальными наблюдениями» (№ гос. регистрации 0113U000361). Авторами разработаны новые методы и модели, основанные на нечетком разбиении объектов на основе критериев плотности. Определены границы адекватности метода в программной среде ГИС и его вычислительная сложность, которая близка к квадратичной.

В рамках выполняемой НИР решены также задачи практической реализации и внедрения на реальных объектах.

СПИСОК ЛИТЕРАТУРЫ

1. Gan G. Data Clustering: theory, algorithms, and applications / G. Gan, C. Ma, J. Wu. – SIAM, Philadelphia, ASA, Alexandria, VA, 2007. – 466 p.
2. Cluster Analysis / [B. Everitt, S. Landau, M. Leese, D. Stahl]. – John Wiley &Sons Ltd, 2011. – 330 p.
3. Xu R. Clustering / R. Xu, D. C. Wunsch. John Wiley &Sons, Inc, 2009. – 358 p.
4. Yager R. Essentials of Fuzzy Modeling and Control / R. Yager, D. Filev. – USA : John Willey & Sons,1984. –387 p.
5. Борисов В. В. Нечеткие модели и сети / В. В. Борисов, В. В. Круглов, А. С. Федулов. – М. : Горячая линия, 2012. – 284 с.
6. Tsoukalas L. H. Fuzzy and Neural Approaches in Engineering / L. H. Tsoukalas, R. E. Uhrig. – New York : John Wiley&Sons.Inc, 1997. – 587 p.
7. Блейхут Р. Теория и практика кодов, контролирующих ошибки / Р. Блейхут. – М. : Мир, 1986. – 576 с.
8. Severance C. Python for Informatics [Electronic resource]. – Regime of access : http://do1.dr-chuck.com/py4inf/EN-us/book.pdf.

Статья поступила в редакцию 16.10.2015.
После доработки 28.10.2015.

Кучеренко Є. І.[1], Глушенкова І. С.[2], Глушенков С. А.[3]

[1]Д-р техн. наук, професор, професор кафедри штучного інтелекту Харківського національного університету радіоелектроніки, Харків, Україна

[2]Канд. техн. наук, доцент кафедри геоінформаційних систем, оцінки землі та нерухомого майна Харківського національного університету міського господарства імені О.М. Бекетова, Харків, Україна

[3]Аспирант кафедри геоінформаційних систем, оцінки землі та нерухомого майна Харківського національного університету міського господарства імені О.М. Бекетова, Харків, Україна

НЕЧІТКЕ РОЗБИТТЯ ОБ'ЄКТІВ НА ОСНОВІ КРИТЕРІЇВ ЩІЛЬНОСТІ

Розв'язано задачу розбиття за критеріями щільності в нечіткому просторі станів при перетині ознак. Об'єктом дослідження були процеси розбиття заданої вибірки об'єктів на підмножини. Предмет дослідження становлять методи й алгоритми нечіткого розбиття об'єктів на основі критеріїв щільності в складних системах. Мета роботи: розвиток методу гірської кластеризації Ягера-Філев на основі нечітких уявлень для підвищення ефективності рішень. Запропоновано нечіткий метод розбиття, заснований на обчисленні щільності розподілу інтегральних ознак об'єктів в нечіткому просторі станів, який, на відміну від існуючих, додатково функціонує в нечіткому просторі станів і ознак. Описано й обгрунтовано етапи методу нечіткого розбиття ознак із застосуванням нечіткої відстані Хеммінга. Було створено програму моделювання щільності розподілу ознак на основі розробленого методу. Виконано експеримент щодо визначення належності об'єкта при перетині областей нечіткого розподілу ознак та надання результатів у вигляді логічного виведення і графічного матеріалу. Результати експерименту дозволяють рекомендувати запропонований метод для використання на практиці. Перспективою подальших досліджень є дослідження та алгоритмізація методу, його адаптація в просторі ознак предметних областей.

Ключові слова: кластеризація, відстань Хеммінга, гірська кластеризація, нечітка логіка.

Kucherenko Ye. I.[1], Glushenkova I. S.[2], Glushenkov S. A.[3]

[1]Dr.Sc., Professor, Department of Artificial Intelligence, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

[2]PhD, Associate Professor, Department of Geoinformation Systems, Land Valuation and Real Property, O. M. Beketov National University of Urban Ekonomy in Kharkiv, Kharkov, Ukraine

[3]Postgraduate student, Department of Geoinformation Systems, Land Valuation and Real Property, O. M. Beketov National University of Urban Ekonomy in Kharkiv, Kharkiv, Ukraine

FUZZY PARTITIONING OF THE OBJECTS BASED ON THE CRITERIA OF DENSITY

The problem of the partition of the criteria in the fuzzy space density of states at the intersection of features. The object of research is the process of partitioning a given sample of objects into subsets. Subject of research methods and algorithms make fuzzy partition of objects based on the criteria density in complex systems. Objective: to develop a method of clustering mining Jager-Fileva based on fuzzy concepts to improve the effectiveness of the decisions. It was proposed fuzzy partitioning method based on the calculation of the density distribution of the integral attributes of the objects in a fuzzy space of conditions. The method, in contrast to existing, additionally operates in a fuzzy state space and features. Describe and justify the steps of the method of fuzzy partitioning features using fuzzy Hamming distance. It was created simulation program distribution density of features on the basis of this method. An experiments conducted to determine the affiliation of the object at the intersection of fuzzy areas of distribution and the provision of evidence of results in the form of inference and graphic material. The experimental results allow us to recommend the proposed method to be used in practice. Prospects for further research is to study and algorithmization method, its adaptation to the feature space domains.

Keywords: Clustering, Hamming distance, mountain clustering, fuzzy logic.

REFERENCES

1 Gan G., Ma C., Wu J. Data Clustering: theory, algorithms, and applicationsю SIAM, Philadelphia, ASA, Alexandria, VA, 2007, 466 p.

2. Everitt B., Landau S., Leese M., Stahl D. Cluster Analysis. John Wiley &Sons Ltd, 2011, 330 p.

3. Xu R., Wunsch D. C. Clustering. John Wiley &Sons, Inc, 2009, 358 p.

4. Yager R., Filev D. Essentials of Fuzzy Modeling and Control. USA, John Willey & Sons, 1984, 387 p.

5. Borisov V. V., Kruglov V. V., Fedulov A. S. Nechetkie modeli i seti. Moscow, Gorjachaja linija, 2012, 284 p.

6. Tsoukalas L. H., Uhrig R. E. Fuzzy and Neural Approaches in Engineering. New York, John Wiley&Sons.Inc, 1997, 587 p.

7. Blejhut R. Teorija i praktika kodov, kontrolirujushhih oshibki. Moscow, Mir, 1986, 576 p.

8. Severance C. Python for Informatics [Electronic resource]. Regime of access : http://do1.dr-chuck.com/py4inf/EN-us/book.pdf.