

УДК 004.8

О СОВМЕСТНОМ ИСПОЛЬЗОВАНИИ МЕТОДОВ КЛАСТЕРНОГО АНАЛИЗА МНОГОМЕРНЫХ ДАННЫХ

А.И. ЯКИМОВ, Е.М. БОРЧИК, В.В. БАШАРИМОВ

ГУВПО «Белорусско-Российский университет»,
пр. Мира, 43, Могилев, 212000, Беларусь

Поступила в редакцию 31 мая 2011

Предложена процедура проверки и уточнения результатов разделения многомерных наблюдений на кластеры с использованием нескольких методов кластерного анализа. В результате разбиения множества X на кластеры каждый из методов ставит в соответствие номерам элементов множества X соответствующие им номера кластеров. Показано, что в случае, если элементы множества X представляют собой наблюдения n параметров множества объектов, то результат кластеризации X может быть интерпретирован как матрица вероятностей принадлежности объектов определенным кластерам. Предложен критерий принадлежности объекта определенному кластеру, получена формула вычисления значений элементов обобщенной матрицы через элементы матриц вероятностей принадлежности объектов определенным кластерам. Экспериментальные исследования проведены с использованием методов кластерного анализа K-Means, Tree Clustering, Fuzzy Relation Clustering и программно-технологического комплекса имитации сложных систем BelSim, пакета статистической обработки данных STATISTICA, табличного процессора MS Excel.

Ключевые слова: кластерный анализ, K-Means, Tree Clustering, Fuzzy Relation Clustering, BelSim, STATISTICA, MS Excel.

Введение

В имитационном моделировании часто возникает необходимость анализа многомерных данных, полученных при проведении имитационных экспериментов, в частности – задачи разделения множеств данных $X \subset R^n$ на непересекающиеся подмножества. Для решения данной задачи используются методы кластерного анализа.

Пусть в ходе имитационных экспериментов получено множество наблюдений $X = \{x_i \mid x_i \in R^n, i = 1, \dots, m\}$, которое необходимо разбить на непересекающиеся подмножества (кластеры).

При решении такой задачи используются методы кластерного анализа, основанные на использовании матриц сходства, эвристических алгоритмов перебора, идей математического программирования, на оценивании функций плотности статистического распределения и др. [1, 2]. В результате анализа существующих методов кластеризации разработана их классификация [3, 4]. Для исследования выбраны методы кластеризации, являющиеся представителями основных методологических подходов к разделению исходного множества объектов на кластеры: K-Means, Tree Clustering, Fuzzy Relation Clustering (FRC).

Метод K-Means строит заданное количество кластеров, но требует охвата каждого кластера отдельным выпуклым множеством. Методы Tree Clustering и FRC не имеют этого ограничения, но не гарантируют построения заданного количества кластеров. Следует отметить, что метод FRC наиболее точен, но характеризуется трудоемкостью $O(n^4)$ от числа элементов.

Гарантированное разбиение множества X на кластеры предполагает использование нескольких методов кластеризации для проверки и уточнения результатов. Вначале разбиение

производится двумя методами. Если результаты разбиений не совпадают, то применяют третий метод. В результате разбиения множества X на кластеры каждый из методов ставит в соответствие номерам $i = 1, \dots, m$ элементов $x_i \in X$ соответствующие им номера кластеров $K_j, j \in 1, \dots, k$. При этом требуется решить задачу обобщения полученных результатов кластеризации множества наблюдений X несколькими методами.

Описание кластерного анализа данных

Кластерный анализ проводится на заранее подготовленных данных

$$X = \{x_i \mid x_i \in R^n, i = 1, \dots, m\}. \quad (1)$$

Пусть элементы $x_i \in X, i = 1, \dots, m$ представляют собой измерения n параметров объектов b_r множества

$$B = \{b_r \mid r = 1, \dots, |B|, |B| < m\}. \quad (2)$$

Тогда каждому $x_i \in X$ соответствует некоторый объект b_r и идентифицирующая информация об этом объекте.

Кластерный анализ методами K-Means и Tree Clustering проводится с использованием пакета STATISTICA 6.0. Метод кластерного анализа FRC реализован в составе программно-технологического комплекса имитации сложных систем BelSim [5].

В результате разбиения множества X на кластеры каждый из методов ставит в соответствие номерам $i = 1, \dots, m$ элементов $x_i \in X$ соответствующие им номера кластеров $K_j, j \in 1, \dots, k$.

Обобщение результатов кластеризации данных несколькими методами

Утверждение 1. Если элементы $x_i \in X, i = 1, \dots, m$ представляют собой измерения n параметров объектов $b_r \in B, r = 1, \dots, |B|, |B| < m$, то результат кластеризации множества X вида (1) может быть представлен в виде матрицы вероятностей принадлежности объектов b_r определенным кластерам $K_j, j \in 1, \dots, k$

$$P = \|p_{rj}\|, p_{rj} \in [0, 1], r = 1, \dots, |B|, j = 1, \dots, k, \quad (3)$$

где r – номер объекта, j – номер кластера, k – количество кластеров K_j, p_{rj} – вероятности принадлежности объектов b_r кластерам K_j .

Вероятности p_{rj} в (3) рассчитываются на основе классического определения вероятности как отношение количества случаев попадания объекта b_r в кластер K_j к общему количеству измерений, выполненных над объектом b_r . Сумма строчных элементов матриц P вида (3) постоянна и равна единице.

Определение 1. Объект $b_r \in B, r = 1, \dots, |B|$ является элементом кластера $K_j, j = 1, \dots, k$ тогда и только тогда когда он отнесен к данному кластеру по крайней мере L^* методами из L выбранных методов кластерного анализа, причем $L/2 < L^* \leq L, L \geq 3$.

Вероятности отнесения объектов к определенным кластерам одновременно L выбранными методами очень низкие, напротив – вероятности отнесения объектов к определенным кластерам, по крайней мере, одним из выбранных методов очень высоки. Поэтому рассматривается аналогия голосования большинством голосов – случай отнесения объекта к данному кластеру L^* методами из L выбранных методов кластерного анализа.

Утверждение 2. Пусть P_1, P_2, \dots, P_L – матрицы вида (3) вероятностей принадлежности объектов $b_r, r = 1, \dots, |B|$ определенным кластерам $K_j, j = 1, \dots, k$ согласно методам кластерного анализа M_1, M_2, \dots, M_L , соответственно. Тогда значения элементов $p_{rj} \in [0, 1]$ обобщенной

(в смысле *Определения 1*) матрицы P могут быть найдены суммированием коэффициентов $P_{v,L}$, $v = L^*, \dots, L$ производящей функции

$$\varphi(z) = \prod_{l=1}^L (q_{lrj} + p_{lrj} \cdot z) = \sum_{v=0}^L P_{v,L} \cdot z^v,$$

где p_{lrj} – элементы матрицы P_l , $l = 1, \dots, L$, $p_{lrj} \in [0, 1]$; $q_{lrj} = 1 - p_{lrj}$, $l = 1, \dots, L$, $r = 1, \dots, |B|$, $j = 1, \dots, k$, соответственно,

$$P_{rj} = \sum_{v=L^*}^L P_{v,L}. \quad (4)$$

Доказательство.

Пусть производится L независимых испытаний. Под испытанием будем понимать проведение кластерного анализа одним из L методов M_1, M_2, \dots, M_L .

Кластерный анализ разными методами проводится в одинаковых условиях, используются одни и те же исходные данные. Результаты обработки данных одним из методов не зависят от результатов анализа другими методами кластерного анализа. Указанные испытания независимы.

Введем в рассмотрение событие A – попадание объекта в результате испытания в определенный кластер. В каждом испытании событие A появляется со своими вероятностями.

Обозначим:

p_1, p_2, \dots, p_L – вероятности попадания объекта в определенный кластер в 1-м, 2-м, ..., L -м испытании, соответственно; $q_1 = 1 - p_1$, $q_2 = 1 - p_2, \dots, q_L = 1 - p_L$ – вероятности непопадания объекта в определенный кластер в 1-м, 2-м, ..., L -ом испытании, соответственно.

Вероятность того, что в этих L испытаниях событие A появится ровно 0, 1, 2, ..., L раз равна коэффициенту при соответствующей степени z производящей функции

$$\begin{aligned} \varphi(z) &= \prod_{l=1}^L (q_l + p_l \cdot z) = (q_1 + p_1 \cdot z)(q_2 + p_2 \cdot z) \cdot \dots \cdot (q_L + p_L \cdot z) = \\ &= q_1 q_2 \cdot \dots \cdot q_L + \dots + p_1 p_2 \cdot \dots \cdot p_L \cdot z^L = \sum_{v=0}^L P_{v,L} \cdot z^v, P_{v,L} \in [0, 1]. \end{aligned}$$

Таким образом, вероятность того, что в этих L испытаниях событие A не появится ни разу: $P_{0,L} = q_1 q_2 \cdot \dots \cdot q_L$; ..., ровно L раз из L : $P_{L,L} = p_1 p_2 \cdot \dots \cdot p_L$.

Тогда вероятность события Y – попадание объекта в определенный кластер, по крайней мере, в L^* испытаниях из L :

$$p(Y) = \sum_{v=L^*}^L P_{v,L}. \quad (5)$$

В результате кластеризации множества данных X вида (1) L методами кластерного анализа получены матрицы вероятностей P_1, P_2, \dots, P_L вида (3) принадлежности объектов определенным кластерам

$$P_l = \|p_{lrj}\|, p_{lrj} \in [0, 1], l = 1, \dots, L, r = 1, \dots, |B|, j = 1, \dots, k. \quad (6)$$

Для получения обобщенного результата кластеризации L методами, применяя к элементам p_{lrj} , $l = 1, \dots, L$ матриц P_1, P_2, \dots, P_L формулу (5), получим обобщенную матрицу

$$P = \|p_{rj}\|, p_{rj} \in [0, 1], r = 1, \dots, |B|, j = 1, \dots, k. \quad (7)$$

В общем виде элементы матрицы P могут быть вычислены по формуле

$$p_{rj} = \sum_{v=L}^L P_{v,L}, \quad r = 1, \dots, |B|, \quad j = 1, \dots, k,$$

где $P_{v,L}$ – коэффициенты производящей функции

$$\varphi(z) = \prod_{l=1}^L ((1 - p_{lrj}) + p_{lrj} \cdot z) = \sum_{v=0}^L P_{v,L} \cdot z^v, \quad r = 1, \dots, |B|, \quad j = 1, \dots, k.$$

Доказательство закончено.

Замечание 1. На элементы строк обобщенной матрицы P вида (7) налагается условие равенства единице их суммы. Поэтому в общем виде элементы обобщенной матрицы P вероятностей принадлежности объектов множества $B = \{b_r \mid r = 1, \dots, |B|\}$ определенным кластерам $K_j, j = 1, \dots, k$ могут быть вычислены по следующей формуле:

$$p_{rj} = \begin{cases} \sum_{v=L}^L P_{v,L} \mid j < k; \\ 1 - \sum_{j=1}^{k-1} p_{rj} \mid j = k. \end{cases} \quad (8)$$

Доказательство.

Введем в рассмотрение события C_1, C_2, \dots, C_k попадание объекта $b_r \in B$ в кластер K_1, K_2, \dots, K_k , соответственно.

При проведении испытания (отнесение методами кластерного анализа элементов $x_i \in X$ множества (1) к определенному кластеру) объект обязательно попадет, по крайней мере, в один из k кластеров, то есть произойдет одно из событий $C_r, r = 1, \dots, k$. Тогда $C_r, r = 1, \dots, k$ образуют полную группу событий и должно выполняться условие

$$p(C_1) + p(C_2) + \dots + p(C_k) = 1. \quad (9)$$

Таким образом, элементы p_{rj} обобщенной матрицы P вида (7) при $j = 1, \dots, k-1, r = 1, \dots, |B|$ могут быть вычислены по формуле (4); при $j = k, r = 1, \dots, |B|$ с учетом (9) – по формуле (8).

Доказательство закончено.

Замечание 2. Обобщаемые матрицы P_1, P_2, \dots, P_L должны иметь одну размерность. В случае разбиения множества X на $k_1 \leq k, k_2 \leq k, \dots, k_L \leq k$ кластеров, необходимо предварительно привести матрицы P_1, P_2, \dots, P_L к одной размерности $k = \max\{k_1, k_2, \dots, k_L\}$. Приведение матрицы к необходимой размерности возможно за счет ее дополнения столбцами с нулевыми вероятностями попадания объекта в добавленные кластеры.

В результате кластеризации множества X вида (1) L методами, получена обобщенная матрица $P = \|p_{rj}\|, r = 1, \dots, |B|, j = 1, \dots, k$, на основе анализа значений элементов которой определяется принадлежность объекта к определенному кластеру.

Определение 2. Объект $b_r \in B$ принадлежит кластеру K_{j_0} тогда и только тогда, когда вероятность принадлежности объекта кластеру в r -й строке обобщенной матрицы P максимальна: $p_{rj_0} = \max\{p_{rj} \mid j = 1, \dots, k\}$.

Экспериментальная часть

Проведем сравнение методов кластерного анализа на искусственно сгенерированном множестве данных с известной структурой. Пусть X – множество двумерных наблюдений, составленное из нескольких подмножеств $X = A_1 \cup A_2 \cup A_3 \cup A_4 \cup A_5 \cup A_6$ (см. рис. 1) таким образом, что кластеры выделяются визуально:

$$A_1 = \begin{cases} 0 < x < 1; \\ 0 < y < 1, \end{cases} \quad A_2 = \begin{cases} 0,5 < x < 2,5; \\ 2 < y < 3, \end{cases} \quad A_3 = \begin{cases} 2 < x < 3; \\ -1 < y < 2, \end{cases}$$

$$A_4 = \begin{cases} -1 < x < 2; \\ -2 < y < -1, \end{cases} \quad A_5 = \begin{cases} -2 < x < -1; \\ -1 < y < 3, \end{cases} \quad A_6 = \begin{cases} -1 < x < 0; \\ 4 < y < 5. \end{cases}$$

Области A_1, \dots, A_6 заполнены равномерно распределёнными на интервале $[0, 1]$ точками, сгенерированными в MS Excel с использованием функции $f(i) = \text{СЛЧИС}()$:

$$A_1 = \{(x, y) \mid x = f(i), y = f(i)\},$$

$$A_2 = \{(x, y) \mid x = 2 \cdot f(i) + 0,5, y = f(i) + 2\},$$

$$A_3 = \{(x, y) \mid x = f(i) + 2, y = 3 \cdot f(i) - 1\},$$

$$A_4 = \{(x, y) \mid x = 3 \cdot f(i) - 1, y = f(i) - 2\},$$

$$A_5 = \{(x, y) \mid x = f(i) - 2, y = 4 \cdot f(i) - 1\},$$

$$A_6 = \{(x, y) \mid x = f(i) - 1, y = f(i) + 4\}, \quad i = 1, \dots, 51.$$

Для каждой из областей A_1, \dots, A_6 сгенерировано одинаковое количество точек (наблюдений) $x_i \in X \subset R^2, i = 1, \dots, 51$ (см. рис. 2).

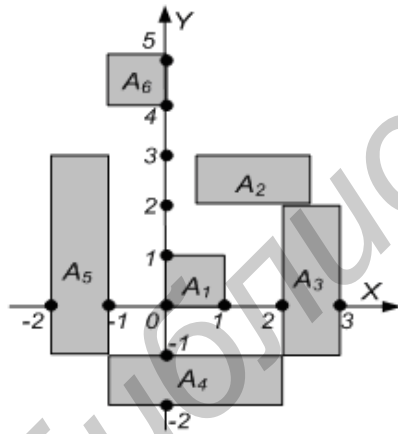


Рис. 1. Подмножества множества наблюдений X

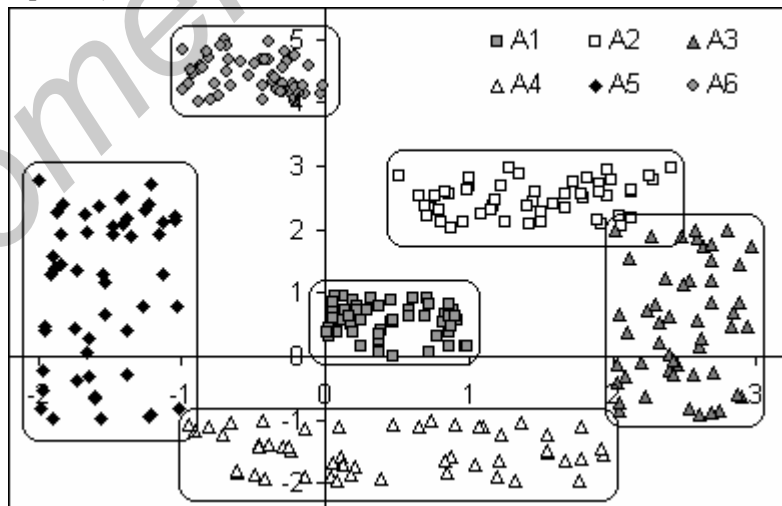


Рис. 2. Кластеризуемое множество наблюдений X

Количество элементов x_i множества $X \subset R^2$ вида (1): $|X| = 51 \cdot 6 = 306$.

Необходимо разделить объекты (наблюдения из областей A_1, \dots, A_6) на непересекающиеся подмножества (кластеры).

Гипотеза относительно числа кластеров: $k = 3$. Ожидается, что элементы подмножеств A_1 и A_6 будут выделены в отдельные кластеры; элементы областей A_2, \dots, A_5 имеют точки соприкосновения, поэтому будут объединены в один кластер.

Для решения данной задачи множество наблюдений $X = \{x_i \mid x_i \in R^2, i = 1, \dots, m\}$, $m = 306$ обрабатывается методами K-Means, Tree Clustering, FRC. Количество выбранных методов кластерного анализа $L = 3$.

Вначале разбиение производится методами K-Means и Tree Clustering.

Метод K-Means разделил множество X на 3 кластера таким образом, что в кластер K_1 преимущественно попала большая часть элементов подмножеств A_1, A_4 ; в кластер K_2 – большая часть элементов множеств A_2, A_3 ; к кластеру K_3 отнесено 50% элементов множества A_5 (см. рис. 3).

В результате разрезания на высоте 0,6 дендрограммы, построенной методом Tree Clustering, множество X разделено на 3 кластера. На основании анализа элементов матрицы Amalgamation Schedule пакета STATISTICA 6,0 определен состав кластеров. Кластер K_1 полностью состоит из элементов подмножества A_1 , кластер K_2 включает элементы $A_2 - A_5$, кластер K_3 полностью состоит из элементов подмножества A_6 (см. рис. 4).

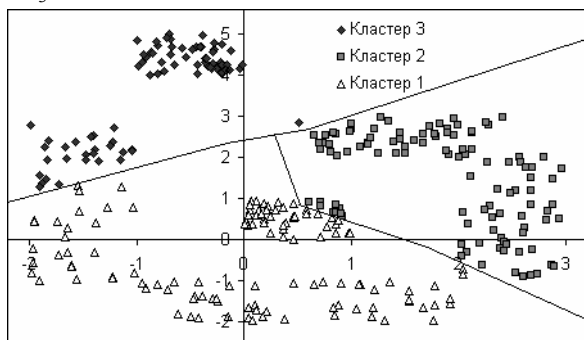


Рис. 3. Разделение множества X на кластеры методом K-Means

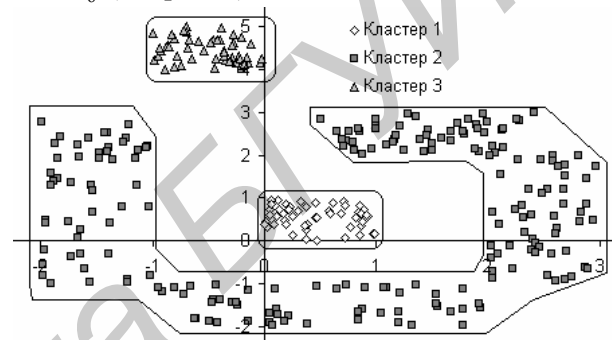


Рис. 4. Разделение множества X на кластеры методами Tree Clustering, FRC

Ниже приведены интерпретированные результаты кластеризации элементов множества X в соответствии с методами K-Means, Tree Clustering – матрицы P_1, P_2 вида (3) вероятностей принадлежности наблюдений из областей A_1, \dots, A_6 кластерам $K_j, j = 1, \dots, 3$:

$$P_1 = \begin{pmatrix} 0,82 & 0,18 & 0,00 \\ 0,00 & 0,98 & 0,02 \\ 0,04 & 0,96 & 0,00 \\ 1,00 & 0,00 & 0,00 \\ 0,49 & 0,00 & 0,51 \\ 0,00 & 0,00 & 1,00 \end{pmatrix}, P_2 = \begin{pmatrix} 1,00 & 0,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 0,00 & 1,00 \end{pmatrix}.$$

Элементы матриц P_1, P_2 рассчитаны как отношение количества случаев попадания элементов множества X в определенный кластер к общему количеству точек, сгенерированных для заданной области.

Если результаты разбиений не совпадают (как в данном случае), для уточнения результатов применяется третий метод кластерного анализа – метод FRC.

Методом FRC при значении параметра $\alpha = 0,85$ множество X разделено на 3 кластера. Результаты разбиения исходного множества методами FRC и Tree Clustering совпали (рис. 4). Ниже приведен интерпретированный результат кластеризации методом FRC элементов множества X – матрица P_3 вида (3) и матрица P , которая является обобщенной матрицей вероятностей принадлежности элементов областей A_1, \dots, A_6 определенным кластерам.

$$P_3 = \begin{pmatrix} 1,00 & 0,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 0,00 & 1,00 \end{pmatrix}, P = \begin{pmatrix} 1,00 & 0,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,00 & 0,00 & 1,00 \end{pmatrix}.$$

При значениях $L = 3$, $L^* = 2$ *Определение 1* принимает следующий вид: «Объект $b_r \in B$, $r = 1, \dots, 6$ (наблюдения из областей A_1, \dots, A_6) является элементом кластера K_j , $j = 1, \dots, 3$ тогда и только тогда, когда он отнесен к данному кластеру по крайней мере двумя из трех методов кластерного анализа».

Результаты и обсуждение

Обобщение результатов кластеризации методами K-Means, Tree Clustering, FRC и построение матриц P_1, P_2, P_3 проведено по формуле (8).

При $L = 3$, $L^* = 2$, $k = 3$, $|B| = 6$ с учетом производящей функции $\varphi(z) = \prod_{l=1}^3 ((1 - p_{lrj}) + p_{lrj} \cdot z) = \sum_{v=0}^3 P_{v,3} \cdot z^v$, формулы (4) $p_{rj} = \sum_{v=2}^3 P_{v,3}$, $r = 1, \dots, 6$, $j = 1, \dots, 3$ и формулы (9) расчет элементов обобщенной матрицы P по формуле (8) принимает вид:

$$p_{rj} = \begin{cases} (1 - p_{1rj})p_{2rj}p_{3rj} + p_{1rj}(1 - p_{2rj})p_{3rj} + p_{1rj}p_{2rj}(1 - p_{3rj}) + p_{1rj}p_{2rj}p_{3rj} & | j < 3; \\ 1 - \sum_{j=1}^2 p_{rj} & | j = 3, \end{cases}$$

где p_{lrj} – элементы матриц P_1, P_2, P_3 соответственно, $l = 1, \dots, 3$, $r = 1, \dots, 6$, $j = 1, \dots, 3$.

Например, элементы 1-й строки p_{11}, p_{12}, p_{13} матрицы $P = \|p_{rj}\|$, $r = 1, \dots, 6$, $j = 1, \dots, 3$ рассчитываются следующим образом:

$$p_{11} = (1 - 0,82) \cdot 1 \cdot 1 + 0,82 \cdot (1 - 1) \cdot 1 + 0,82 \cdot 1 \cdot (1 - 1) + 0,82 \cdot 1 \cdot 1 = 1 - 0,82 + 0,82 = 1;$$

$$p_{12} = (1 - 0,18) \cdot 0 \cdot 0 + 0,18 \cdot (1 - 0) \cdot 0 + 0,18 \cdot 0 \cdot (1 - 0) + 0,18 \cdot 0 \cdot 0 = 0;$$

$$p_{13} = 1 - 1 - 0 = 0.$$

Анализ обобщенного результата кластеризации по матрице P позволяет выделить 3 кластера: $K_1 = A_1$, $K_2 = A_2 \cup \dots \cup A_5$ и $K_3 = A_6$. Таким образом, ожидаемый результат кластеризации исходного множества со специально заданной структурой подтвердился.

Заключение

В результате разбиения множества X на кластеры каждый из методов кластеризации ставит в соответствие номерам $i = 1, \dots, m$ элементов $x_i \in X$ соответствующие им номера кластеров K_j , $j = 1, \dots, k$.

Результат кластеризации множества многомерных данных, соответствующих наблюдениям объектов $b_r \in B$ для каждого из методов кластерного анализа, можно представить в виде матрицы вероятностей принадлежности объектов b_r определенным кластерам K_j .

Обобщенный результат кластеризации несколькими методами получен посредством применения теорем сложения/умножения вероятностей. Анализ обобщенного результата кластеризации позволяет выявить группы объектов с эквивалентным поведением.

Проведено сравнение методов кластерного анализа на множестве X двумерных наблюдений, составленном из нескольких подмножеств таким образом, что кластеры выделяются визуально. Области, соответствующие подмножествам множества X заполнены одинаковым количеством равномерно распределенных на интервале $[0, 1]$ значений.

Проведена кластеризация множества X методами K-Means, Tree Clustering, FRC. Вначале произведено разбиение методами K-Means и Tree Clustering. Поскольку результаты разбиений не эквивалентны, для уточнения результатов применен третий метод кластеризации – FRC. Все три метода кластеризации разбили множество X на ожидаемое количество кластеров $k = 3$.

Анализ обобщенного результата кластеризации позволил выделить 3 кластера. Ожидаемый результат кластеризации исходного множества со специально заданной структурой подтвердился, что подтверждает работоспособность предложенной процедуры обобщения результатов кластеризации несколькими методами.

ABOUT JOINT USE OF METHODS OF MULTIDIMENSIONAL DATA CLUSTER ANALYSIS

A.I. YAKIMAU, E.M. BORCHYK, V.V. BASHARIMOV

Abstract

A procedure of validation and refinement of assignment of multidimensional observations into clusters is given. Every clustering method divides the set X into the clusters and assigns numbers of members of the set X to numbers of the clusters that corresponding to them. It is shown that if the members of the set X are observations of n parameters of a set of objects, then the result of clustering of X can be interpreted as a matrix of probability of object membership to clusters. Criterion of the object membership to the cluster is introduced; a formula to calculate values of elements of generalized matrix based on elements of the matrices of probability of object membership to clusters.

Clustering methods K-Means, Tree Clustering, Fuzzy Relation Clustering and software engineering kit for computer simulation BelSim, statistics and analytics software package STATISTICA, spreadsheet application Microsoft Excel are used to perform an experimental research.

Литература

1. Барсегян А.А. Методы и модели анализа данных: OLAP и Data Mining. СПб., 2004.
2. Паклин Н. [Электронный ресурс] Режим доступа: <http://www.basegroup.ru/library/analysis/clusterization/datamining/>.
3. Башаримов, В.В. // Информационные технологии, энергетика и экономика : сб. трудов 7-ой Межрег. науч.-техн. конф. студентов и аспирантов. 2010. Т.2. С. 21–26.
4. Отчет по НИР. №20091957. Могилев, 2011.
5. Якимов А.И. Технология имитационного моделирования систем управления промышленных предприятий. Могилев, 2010.