

А.И. Якимов, Е.М. Борчик, В.В. Башаримов

## ОПРЕДЕЛЕНИЕ КЛАССОВ ЦЕЛЕВЫХ ФУНКЦИЙ МЕТОДАМИ КЛАСТЕРНОГО АНАЛИЗА

Во введении указан объект исследования – целевые функции (ЦФ) с эквивалентным поведением алгоритма оптимизации (АО). При решении задач имитационного моделирования используются АО, основанные на случайном поиске, которые имеют множество параметров, требующих дополнительной настройки для эффективной работы самого АО. С целью повышения эффективности стохастических алгоритмов многопараметрической оптимизации в имитационном моделировании предложена процедура определения классов целевых функций, основанная на применении нескольких методов кластерного анализа. В основной части указывается, что одним из этапов методики настройки АО в программно-технологическом комплексе имитации сложных систем BelSim является определение классов целевых функций (ЦФ) с эквивалентным поведением АО. Для определения классов ЦФ, в пределах которых известны функциональные зависимости для получения оптимального сочетания параметров АО, используются методы кластерного анализа. При решении практической задачи проводится эксперимент над целевой функцией для того, чтобы определить класс, к которому она относится. Поиск решения задачи проводится с помощью алгоритма, параметры которого вычисляются на основе функциональных зависимостей данного класса. Экспериментальные исследования проведены с использованием программно-технологического комплекса имитации сложных систем BelSim, пакета статистической обработки данных STATISTICA, табличного процессора MS Excel. Научная новизна работы заключается в том, что получена формула, позволяющая обобщить результаты кластеризации исходного множества данных несколькими методами кластерного анализа. В заключении приведен анализ полученных результатов исследования ЦФ с учетом их особенностей. Полученные результаты могут быть применены в имитационном моделировании и обработке данных.

**Ключевые слова:** целевая функция, алгоритм оптимизации, кластерный анализ, K-Means, Tree Clustering, Fuzzy Relation Clustering, BelSim, STATISTICA, MS Excel.

**Введение.** Одним из этапов методики настройки алгоритма оптимизации (АО) в программно-технологическом комплексе имитации сложных систем (ПТКИ) BelSim [1, с. 105–143] является определение классов целевых функций (ЦФ) с эквивалентным поведением АО.

Пусть задано множество ЦФ  $\{F_i\}$ ,  $i = 1, \dots, n$  и АО с  $m$  параметрами и  $r$  откликами. Чувствительность откликов АО при изменении значений параметров для заданной ЦФ разная [2, с. 38–40]. Необходимо разделить ЦФ на непересекающиеся подмножества (классы), для которых результаты анализа чувствительности в пределах одного класса эквивалентны.

Для решения данной задачи предложено использовать методы кластерного анализа. В литературе описывается множество различных методов кластеризации, основанных на использовании матриц сходства, эвристических методов перебора, идей математического программирования, оценивании функций плотности статистического распределения и др. [3, с. 21–26;

---

**Якимов Анатолий Иванович**, канд. техн. наук, доц., доц. каф. «Автоматизированные системы управления» Белорусско-Российского университета (Беларусь).

**Адрес для корреспонденции:** пр. Мира, 43, 212030, г. Могилёв, Беларусь; e-mail: ykm@tut.by

**Борчик Екатерина Михайловна**, аспирант каф. «Автоматизированные системы управления» Белорусско-Российского университета (Беларусь); науч. рук. – А.И. Якимов, канд. техн. наук, доц., доц. каф. «Автоматизированные системы управления» Белорусско-Российского университета (Беларусь).

**Адрес для корреспонденции:** пр. Мира, 43, 212030, г. Могилёв, Беларусь; e-mail: katrinb15@gmail.com

**Башаримов Вячеслав Владимирович**, аспирант каф. «Автоматизированные системы управления» Белорусско-Российского университета (Беларусь); науч. рук. – А.И. Якимов, канд. техн. наук, доц., доц. каф. «Автоматизированные системы управления» Белорусско-Российского университета (Беларусь).

**Адрес для корреспонденции:** пр. Мира, 43, 212030, г. Могилёв, Беларусь; e-mail: basharimovvv@tut.by

4, с. 149–168; 5]. В результате анализа существующих методов кластеризации разработана их классификация [6, с. 64–66]. Для исследования выбраны следующие методы кластеризации, являющиеся представителями основных методологических подходов к разделению исходного множества объектов на классы: K-Means, Tree Clustering, Fuzzy Relation Clustering (FRC).

Предложена процедура определения классов целевых функций для алгоритмов оптимизации с использованием нескольких методов кластерного анализа для проверки и уточнения результатов. Вначале разбиение производится методами K-Means и Tree Clustering. Если результаты разбиений не совпадают, то применяется метод FRC.

В результате разделения кластеризуемого множества данных (чувствительности откликов АО при изменении значений параметров для заданных ЦФ и определенного АО) на  $k$  классов каждый из методов ставит в соответствие номерам элементов множества соответствующие им номера классов. При этом требуется обобщение полученных результатов кластеризации множества данных несколькими методами с последующим определением классов целевых функций.

**Содержательное описание основных методов кластерного анализа.** Метод кластеризации Tree Clustering входит в наиболее распространенную группу эвристических методов кластеризации, основывающихся на иерархической агломеративной процедуре, смысл которой заключается в следующем. В начале кластеризации все объекты считаются отдельными кластерами. На первом шаге алгоритма определяются два наиболее близких или сходных объекта, которые объединяются в один кластер, общее количество которых сокращается на единицу. Итеративный процесс повторяется, пока на последнем шаге все классы не объединятся.

Последовательность объединения поддается геометрической интерпретации в виде графа-дерева (дендрограммы). На оси абсцисс дендрограммы откладываются имена наблюдений, а по оси ординат – расстояние объединения наблюдений в классы. Таким образом, чем выше расположена ветвь дерева на дендрограмме, тем позднее было проведено объединение объектов. Дендрограмму можно использовать для определения того или иного количества отдельных классов путем «разрезания» дендрограммы горизонтальной линией в месте, где наблюдается максимальный скачок в изменении межкластерного расстояния [6, с. 64–66].

Метод кластеризации K-Means применяется, если уже имеются гипотезы относительно числа кластеров. В общем случае метод K-Means строит ровно заданное количество  $k$  различных классов. Алгоритм метода носит итеративный характер, и на каждой итерации требуется рассчитывать матрицу расстояний между объектами. Это при большом числе объектов может потребовать серьезных вычислительных ресурсов.

Имеет недостатки в K-Means сам подход с идеей поиска классов сферической или эллипсоидной формы. Подход хорошо работает, когда данные в пространстве образуют компактные сгустки, отличимые друг от друга. Если данные имеют вложенную форму, то метод не справится с такой задачей [4, с. 162–166].

Метод кластеризации Fuzzy Relation Clustering представляет собой подход к кластеризации конечного набора элементов произвольного метрического пространства на основании разбиения множества на классы эквивалентности по нечеткому отношению  $R_\alpha$ , позволяющему определить для каждого значения  $\alpha$  в диапазоне от 0 до 1 отношение эквивалентности на исходном множестве. При этом два элемента входят в один класс эквивалентности тогда и только тогда, когда между ними есть последовательность попарно «близких» друг к другу элементов.

Полученные при помощи описанного метода классы могут иметь произвольную геометрическую форму, в том числе состоять из невыпуклых множеств. Это является значительным преимуществом подхода перед другими методами кластеризации, например, модификациями K-Means.

Недостатком метода является большое время выполнения, характеризуемое порядком  $O(n^4)$  от числа элементов. Устранение данного недостатка возможно за счет последовательной кластеризации специальным образом определённых подмножеств исходного множества с последующим объединением результатов [4, с. 174–204].

### **Шаги процедуры определения классов целевых функций в ПТКИ BelSim.**

#### **Шаг 1.** Предварительное исследование АО.

Проводится анализ чувствительности [2, с. 38–40] откликов АО при изменении значений параметров АО для каждой ЦФ. Результатом анализа чувствительности являются матрицы вида

$$S = \left\| \|s_{ij}\|, s_{ij} \in R, i = 1, \dots, r, j = 1, \dots, m \right\|, \quad (1)$$

где  $s_{ij}$  – чувствительность  $i$ -го отклика при изменении  $j$ -го параметра АО.

Для проведения кластерного анализа матрицы вида (1) преобразуются к вектору вида:

$$S = (s_{11}, s_{12}, \dots, s_{1m}, \dots, s_{r1}, \dots, s_{rm}). \quad (2)$$

В итоге для каждой ЦФ проводится  $t$  раз анализ чувствительности и получается множество векторов вида (2):

$$\{S_i\}, i = 1, \dots, t \cdot n. \quad (3)$$

### Шаг 2. Кластерный анализ.

Для разделения на классы множества векторов (3) используются методы кластерного анализа, например, K-Means, Tree Clustering и FRC. Гарантированное разбиение множества (3) на классы предполагает использование нескольких методов кластеризации для проверки и уточнения результатов.

В результате разбиения множества (3) на  $k$  классов каждый из методов ставит в соответствие номерам  $i = 1, \dots, t \cdot n$  элементов множества (3) соответствующие им номера классов  $K_j, j = 1, \dots, k$ .

### Шаг 3. Интерпретация закономерностей, найденных методами кластерного анализа.

Поскольку векторы  $S_i \in R^{m \cdot r}, i = 1, \dots, t \cdot n$  кластеризуемого множества (3) представляют собой результаты статистической обработки экспериментальных данных  $n$  ЦФ, то итог кластеризации множества (3) можно представить в виде матрицы вероятностей принадлежности ЦФ определенным классам:

$$P = \left\| p_{ij} \right\|, i = 1, \dots, n, j = 1, \dots, k, \quad (4)$$

где  $i$  – номер ЦФ,  $j$  – номер класса  $K_j, k$  – количество классов,  $p_{ij} \in [0, 1]$  – вероятность принадлежности  $i$ -й ЦФ классу  $K_j$ .

Вероятности  $p_{ij}$  в (4) рассчитываются на основе классического определения вероятности как отношение количества случаев попадания ЦФ  $F_i$  в класс  $K_j$  к общему количеству экспериментов, проведенных над данной ЦФ.

### Шаг 4. Обобщение результатов кластеризации данных несколькими методами.

**Определение 1.** ЦФ  $F_i$  является элементом класса  $K_j$  тогда и только тогда, когда она отнесена к данному классу, по крайней мере, двумя из трех выбранных методов кластеризации.

Вероятности отнесения ЦФ к определенным классам одновременно тремя выбранными методами очень низкие, напротив – вероятности отнесения ЦФ к определенным классам, по крайней мере, одним из выбранных методов очень высоки. Поэтому рассматривается аналогия голосования большинством голосов – случай отнесения ЦФ к данному классу в соответствии с хотя бы двумя из трех выбранных методов кластерного анализа.

**Утверждение.** Пусть  $P_1, P_2, P_3$  – матрицы вида (4) вероятностей принадлежности ЦФ  $F_i, i = 1, \dots, n$  определенным классам в соответствии с методами кластерного анализа K-Means, Tree Clustering, FRC соответственно. Тогда значения элементов  $p_{ij} \in [0, 1]$  обобщенной (в смысле

Определения 1) матрицы  $P$  могут быть найдены суммированием коэффициентов  $P_{2,3}, P_{3,3}$  производящей функции

$$\varphi(z) = \prod_{l=1}^3 (q_{lij} + p_{lij} \cdot z) = \sum_{v=0}^3 P_{v,3} \cdot z^v,$$

где  $q_{lij} = 1 - p_{lij}$ ,  $p_{lij} \in [0, 1]$ ,  $l = 1, \dots, 3$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, k$  – элементы матриц  $P_1, P_2, P_3$ , соответственно,

$$p_{ij} = (1 - p_{1ij})p_{2ij}p_{3ij} + p_{1ij}(1 - p_{2ij})p_{3ij} + p_{1ij}p_{2ij}(1 - p_{3ij}) + p_{1ij}p_{2ij}p_{3ij}. \quad (5)$$

**Д о к а з а т е л ь с т в о.** Пусть производятся три независимых испытания. Под испытаниями будем понимать проведение кластерного анализа одним из трех методов: K-Means, Tree Clustering, FRC.

Кластерный анализ разными методами проводится в одинаковых условиях. На вход подаются одни и те же исходные данные. Результаты обработки одним из методов не зависят от результатов работы других методов кластерного анализа. Указанные испытания независимы.

Введем в рассмотрение событие  $A$  – попадание ЦФ в результате испытания в определенный класс. В каждом испытании событие  $A$  появляется со своими вероятностями.

Обозначим:

–  $p_1, p_2, p_3$  – вероятности попадания ЦФ в определенный класс в 1-м, 2-м, 3-м испытании соответственно;

–  $q_1 = 1 - p_1, q_2 = 1 - p_2, q_3 = 1 - p_3$  – вероятности не попадания ЦФ в определенный класс в 1-м, 2-м, 3-м испытании соответственно.

Вероятность того, что в этих трех испытаниях событие  $A$  появится ровно 0, 1, 2, 3 раза равна коэффициенту при соответствующей степени  $z$  производящей функции:

$$\begin{aligned} \varphi(z) &= \prod_{l=1}^3 (q_l + p_l \cdot z) \\ \varphi(z) &= (q_1 + p_1 \cdot z)(q_2 + p_2 \cdot z)(q_3 + p_3 \cdot z) = q_1q_2q_3 + (p_1q_2q_3 + q_1p_2q_3 + q_1q_2p_3)z + \\ &+ (q_1p_2p_3 + p_1q_2p_3 + p_1p_2q_3)z^2 + p_1p_2p_3z^3 = \sum_{v=0}^3 P_{v,3} \cdot z^v. \end{aligned}$$

Таким образом, вероятность того, что в этих трех испытаниях событие  $A$  не появится ни разу:  $P_{0,3} = q_1q_2q_3$ ; появится ровно один раз:  $P_{1,3} = p_1q_2q_3 + q_1p_2q_3 + q_1q_2p_3$ ; ровно два раза из трех:  $P_{2,3} = q_1p_2p_3 + p_1q_2p_3 + p_1p_2q_3$ ; ровно 3 раза из трех:  $P_{3,3} = p_1p_2p_3$ .

Тогда вероятность события  $Y$  – попадание ЦФ в определенный класс, по крайней мере, в двух испытаниях из трех  $p(Y) = P_{2,3} + P_{3,3}$ , или

$$p(Y) = q_1 \cdot p_2 \cdot p_3 + p_1 \cdot q_2 \cdot p_3 + p_1 \cdot p_2 \cdot q_3 + p_1 \cdot p_2 \cdot p_3. \quad (6)$$

В результате кластеризации множества наблюдений (3) тремя методами кластерного анализа получены матрицы вероятностей принадлежности ЦФ определенным классам  $P_1, P_2, P_3$  вида (4) с элементами  $p_{ij}, l = 1, \dots, 3$ :

$$P_l = \left\| p_{lij} \right\|, p_{lij} \in [0, 1], i = 1, \dots, n, j = 1, \dots, k, l = 1, \dots, 3.$$

Для получения обобщенного результата кластеризации тремя методами, применяя к элементам  $p_{lij}, l = 1, \dots, 3$  матриц  $P_1, P_2, P_3$  формулу (6), получим матрицу:

$$P = \left\| p_{ij} \right\|, p_{ij} \in [0, 1], i = 1, \dots, n, j = 1, \dots, k.$$

В общем виде элементы матрицы  $P$  могут быть вычислены по формуле

$$p_{ij} = (1 - p_{1ij})p_{2ij}p_{3ij} + p_{1ij}(1 - p_{2ij})p_{3ij} + p_{1ij}p_{2ij}(1 - p_{3ij}) + p_{1ij}p_{2ij}p_{3ij}, i = 1, \dots, n, j = 1, \dots, k. \quad (7)$$

Доказательство закончено.

**Замечание 1.** На элементы строк обобщенной матрицы  $P$  налагается условие равенства единице их суммы. Поэтому в общем виде элементы обобщенной матрицы  $P$  вероятностей принадлежности ЦФ  $F_i, i = 1, \dots, n$  определенным классам  $K_j, j = 1, \dots, k$  могут быть вычислены по следующей формуле:

$$p_{ij} = \begin{cases} (1 - p_{1ij})p_{2ij}p_{3ij} + p_{1ij}(1 - p_{2ij})p_{3ij} + p_{1ij}p_{2ij}(1 - p_{3ij}) + p_{1ij}p_{2ij}p_{3ij}, j < k; \\ 1 - \sum_{j=1}^{k-1} p_{ij}, j = k. \end{cases} \quad (8)$$

**Доказательство.** Введем в рассмотрение события  $C_1, C_2, \dots, C_k$  – попадание ЦФ  $F_i, i = 1, \dots, n$  в класс  $K_1, K_2, \dots, K_k$  соответственно.

При проведении испытания (отнесении методами кластерного анализа элементов (3) к определенному классу) ЦФ обязательно попадет хотя бы в один из  $k$  классов, то есть произойдет одно из событий  $C_j, j = 1, \dots, k$ . Тогда события  $C_j, j = 1, \dots, k$  образуют полную группу событий и должно выполняться условие

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

Таким образом, элементы  $p_{ij}$  обобщенной матрицы  $P$  при  $j = 1, \dots, k - 1, i = 1, \dots, n$  могут быть вычислены по формуле (7) при  $j = k, i = 1, \dots, n$ , с учетом (9) – по формуле (8).

Доказательство закончено.

**Замечание 2.** Обобщаемые матрицы  $P_1, P_2, P_3$  должны иметь одну размерность. В случае разбиения множества (3) на  $k_1 \leq k, k_2 \leq k, k_3 \leq k$  классов, необходимо предварительно привести матрицы  $P_1, P_2, P_3$  к одной размерности  $k = \max\{k_1, k_2, k_3\}$ . Приведение матрицы к необходимой размерности возможно за счет ее дополнения столбцами с нулевыми вероятностями попадания объекта в добавленные классы.

**Шаг 5.** Определение классов ЦФ.

В результате кластеризации множества (3) тремя методами получена обобщенная матрица  $P = \left\| p_{ij} \right\|, i = 1, \dots, n, j = 1, \dots, k$ , на основе анализа которой определяется принадлежность ЦФ определенному классу. Полагается, что ЦФ  $F_i$  принадлежит классу  $K_{j_0}$  тогда и только тогда, когда вероятность принадлежности ЦФ классу  $K_{j_0}$  в обобщенной матрице  $P$  максимальна.

**Экспериментальная часть.** Пусть выбраны следующие тестовые ЦФ (рисунки 1–6)  $F_i, i = 1, \dots, n (n = 5)$ :  $F_1$  – Easom,  $F_2$  – Six-hump,  $F_3$  – Branin RCos,  $F_4$  – Modified Branin RCos,  $F_5$  – Goldstein&Price [7].

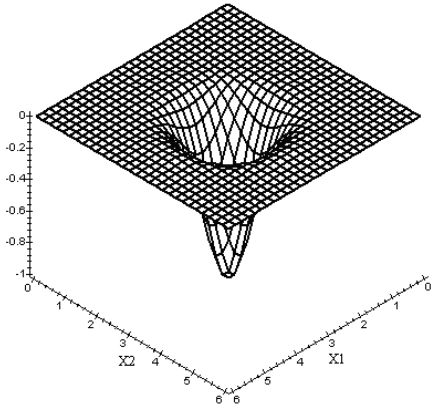


Рисунок 1 – Тестовая ЦФ  $F_1$  – функция Easom

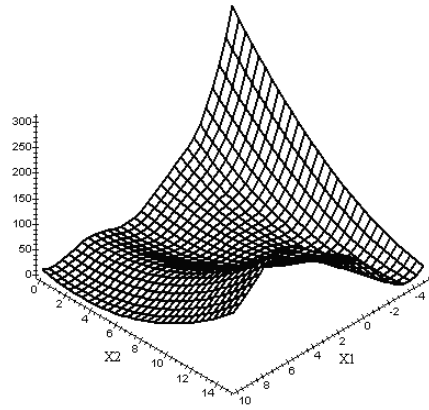


Рисунок 2 – Тестовая ЦФ  $F_3$  – функция Branin RCos

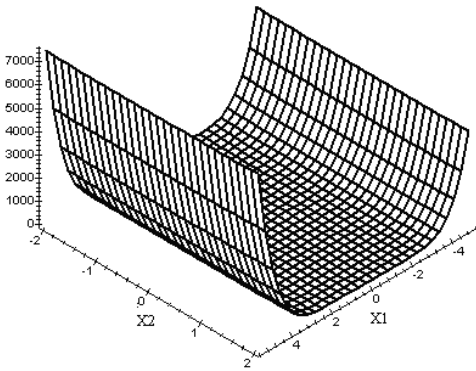


Рисунок 3 – Тестовая ЦФ  $F_2$  – функция Six-hump

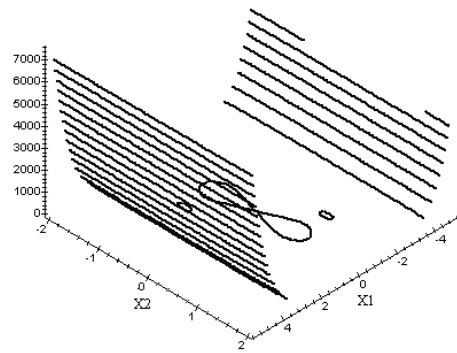


Рисунок 4 – Линии уровня тестовой ЦФ  $F_2$  – функции Six-hump

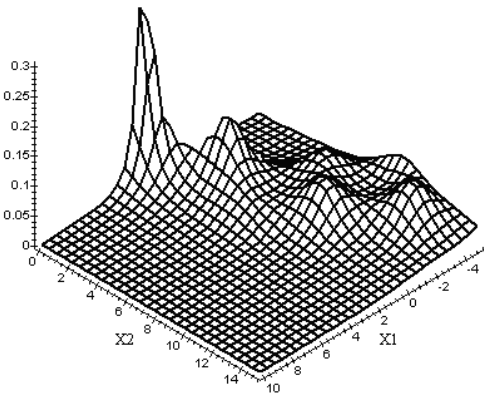


Рисунок 5 – Тестовая ЦФ  $F_4$  – функция Modified Branin RCos

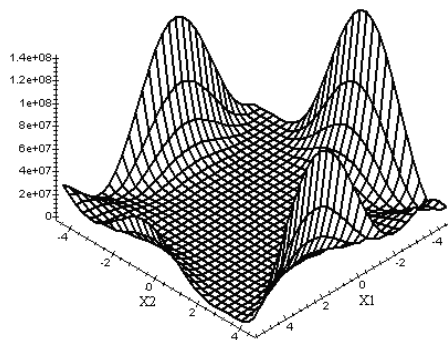


Рисунок 6 – Тестовая ЦФ  $F_5$  – функция Goldstein&Price



В качестве АО выбран генетический алгоритм (ГА).

Параметры ГА:  $X_1$  – вероятность кроссовера,  $X_2$  – вероятность мутации,  $X_3$  – доля родителей от размера популяции,  $X_4$  – математическое ожидание для распределения количества точек разрыва. Количество параметров АО  $m = 4$ .

Отклики ГА:  $Y_1$  – лучшее решение поколения,  $Y_2$  – худшее решение поколения,  $Y_3$  – среднее значение приспособленности поколения,  $Y_4$  – дисперсия приспособленности поколения,  $Y_5$  – лучшее решение,  $Y_6$  – количество запусков функции качества. Таким образом, количество откликов АО с использованием ГА  $r = 6$ .

Значения параметров ГА заданы на трех уровнях [2, с. 38–40] (таблица 1), где <"-1"> означает "-1"-й уровень эксперимента, <"0"> – соответственно "0"-й, <"1"> – "1"-й уровень эксперимента.

Таблица 1 – Значения параметров АО на уровнях эксперимента

Уровни\Параметры	$X_1$	$X_2$	$X_3$	$X_4$
<"-1">	0,01	0,01	0,5	2
<"0">	0,5	0,5	0,75	5
<"1">	0,99	0,99	1	8

Необходимо разделить ЦФ на классы.

Для решения поставленной задачи в ПТКИ BelSim выполняются шаги процедуры определения классов целевых функций.

Элементы  $S_i$  множества (3), полученные в результате предварительного исследования АО на шаге 1 процедуры, представляют собой  $4 \cdot 6 = 24$ -мерные вектора действительных чисел. Для каждой из пяти ЦФ проведен  $t = 31$  эксперимент по анализу чувствительности, общее количество элементов  $S_i$  множества (3) составляет  $|\{S_i\}| = 5 \cdot 31 = 155$ .

Для разделения на классы множества векторов (3) (шаг 2 процедуры) используются методы кластерного анализа K-Means, Tree Clustering и FRC. Кластерный анализ методами K-Means и Tree Clustering проводится с использованием пакета STATISTICA 6.0. Метод кластерного анализа FRC реализован в составе ПТКИ BelSim. Гипотеза относительно числа классов:  $k = 3$ .

Метод K-Means делит данные ровно на  $k$  классов. Ожидаемое количество классов:  $k = 3$ . По запросу пакета STATISTICA может быть построен график средних значений для каждого класса (Graph of means); выведены по отдельности члены каждого класса (members of each cluster& distances) и др.

Полная классификация для всех случаев данных (Save classifications and distances), или соответствие номерам  $i = 1, \dots, m$  элементов множества (3) номеров классов  $K_j$ ,  $j \in 1, \dots, k$ , сохраняется для дальнейшего анализа.

Для определения методом Tree Clustering в пакете STATISTICA 6.0 количества классов  $k_1 \leq k$  ( $k$  – ожидаемое количество классов) достаточно построения дендрограммы (Horizontal hierarchical tree plot), матрицы расстояний  $D$  (Distance matrix), матрицы Amalgation Schedule. Полная классификация для всех случаев данных может быть получена на основе анализа данных матрицы расстояний и матрицы Amalgation Schedule.

Для кластеризации множества (3) методом FRC данные передаются в блок кластерного анализа ПТКИ BelSim. Задается значение параметра  $\alpha \in [0, 1]$  метода FRC. С учетом варьирования значений параметра  $\alpha \in [0, 1]$  метод FRC позволяет построить  $k_2 \leq k$  классов, где  $k$  – ожидаемое количество классов. Полная классификация, построенная методом FRC для всех случаев данных, сохраняется для дальнейшего анализа.

Все три метода кластерного анализа разбили множество (3) на 3 класса.

Вначале разбиение производится методами K-Means и Tree Clustering. Интерпретированные результаты кластеризации элементов множества (3) (шаг 3 процедуры) – матрицы  $P_1, P_2$  вероятностей принадлежности ЦФ  $F_1, \dots, F_5$  классам  $K_j, j = 1, \dots, 3$  в соответствии с методами K-Means и Tree Clustering приведены ниже. Элементы матриц  $P_1, P_2$  рассчитаны как отношение количества случаев попадания элементов множества (3) в определенный класс к общему количеству экспериментов, проведенных над ЦФ.

Если результаты разбиений не совпадают, как в данном случае (см. элементы  $p_{22}, p_{23}$  матриц  $P_1, P_2$ ), то применяется метод FRC. Интерпретированный результат кластеризации методом FRC множества (3) (шаг 3 процедуры) – матрица  $P_3$  – приведен ниже.

$$P_1 = \begin{pmatrix} 0,97 & 0,00 & 0,03 \\ 0,06 & 0,03 & 0,90 \\ 0,00 & 1,00 & 0,00 \\ 0,19 & 0,81 & 0,00 \\ 1,00 & 0,00 & 0,00 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0,97 & 0,00 & 0,03 \\ 0,06 & 0,94 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,19 & 0,81 & 0,00 \\ 1,00 & 0,00 & 0,00 \end{pmatrix}, \quad P_3 = \begin{pmatrix} 0,97 & 0,00 & 0,03 \\ 0,06 & 0,94 & 0,00 \\ 0,00 & 1,00 & 0,00 \\ 0,19 & 0,81 & 0,00 \\ 1,00 & 0,00 & 0,00 \end{pmatrix},$$

$$P = \begin{pmatrix} 1,00 & 0,00 & 0,00 \\ 0,01 & 0,88 & 0,11 \\ 0,00 & 1,00 & 0,00 \\ 0,10 & 0,90 & 0,00 \\ 1,00 & 0,00 & 0,00 \end{pmatrix}.$$

Матрица  $P$  – является обобщенной матрицей вероятностей принадлежности ЦФ определенным классам. Элементы матрицы  $P$  рассчитаны на основании (8) (шаг 4 процедуры). Например, элементы 2-й строки  $p_{21}, p_{22}, p_{23}$  матрицы  $P = \|p_{ij}\|, i = 1, \dots, 5, j = 1, \dots, 3$  рассчитываются следующим образом:

$$p_{21} = (1 - 0,06) \cdot 0,06 \cdot 0,06 + 0,06 \cdot (1 - 0,06) \cdot 0,06 + 0,06 \cdot 0,06 \cdot (1 - 0,06) + 0,06 \cdot 0,06 \cdot 0,06 = 0,01,$$

$$p_{22} = (1 - 0,03) \cdot 0,94 \cdot 0,94 + 0,03 \cdot (1 - 0,94) \cdot 0,94 + 0,03 \cdot 0,94 \cdot (1 - 0,94) + 0,03 \cdot 0,94 \cdot 0,94 = 0,88,$$

$$p_{23} = 1 - 0,01 - 0,88 = 0,11.$$

**Результаты и их обсуждение.** В начале исследования предполагалось выделить три класса ЦФ. Процедура определения классов ЦФ в ПТКИ BELSIM выявила наличие двух классов (шаг 5 процедуры):

– класс  $K_1$  (с наименьшим значением чувствительности) со 100 % вероятностью принадлежности классу включает функции Easom и Goldstein&Price;

– класс  $K_2$  включает функции Six-hump, Branin RCos, Modified Branin RCos с вероятностями принадлежности классу 88, 100 и 90 % соответственно.



Ранее проводилось предварительное исследование – сравнение методов кластерного анализа на множестве двумерных наблюдений, составленном из нескольких подмножеств таким образом, что классы выделялись визуально. Области соответствующих подмножеств исходного составного множества заполнялись одинаковым количеством равномерно распределенных на интервале  $[0, 1]$  значений. Анализ обобщенного результата кластеризации позволяет выделить ожидаемое количество классов. Таким образом, ожидаемый результат кластеризации исходного множества со специально заданной структурой подтвердился.

**Заключение.** В проведенном исследовании целевых функций близость внутри классов результатов вида (2) показывает наличие структурированности в данных, что позволяет выделить классы ЦФ с эквивалентным поведением.

Анализ особенностей ЦФ показывает следующее: в класс  $K_1$  входят унимодальные ЦФ с наличием пологих участков, в класс  $K_2$  – мультимодальные ЦФ с множественными локальными оптимумами. Эти результаты показывает работоспособность предложенной процедуры.

#### СПИСОК ЛИТЕРАТУРЫ

1. Якимов, А.И. Технология имитационного моделирования систем управления промышленных предприятий: монография / А.И. Якимов. – Могилев: Беларус.-Рос., 2010. – 304 с.
2. Максимей, И.В. Задачи и модели исследования операций / И.В. Максимей [и др.]. – Гомель: БелГУТ, 1999. – 150 с.
3. Башаримов, В.В. Выбор методов кластерного анализа при решении задач оптимизации в имитационном моделировании / В.В. Башаримов, Е.М. Борчик, А.И. Якимов // Информационные технологии, энергетика и экономика (информационные технологии, математическое моделирование технологических процессов, электроника): сб. трудов 7 межрегион. науч.-техн. конф. студентов и аспирантов, 8–9 апр. 2010 г.: в 3 т. – Смоленск: ГОУ ВПО МЭИ(ТУ), 2010. – Т. 2. – С. 21–26.
4. Методы и модели анализа данных: OLAP и Data Mining / А.А Барсегян и [и др.] – СПб.: БХВ–Петербург, 2004. – 336 с.
5. Паклин, Н. Алгоритмы кластеризации на службе Data Mining [Электронный ресурс] / Н. Паклин // Технологии анализа данных. – 2011. – Режим доступа: <http://www.basegroup.ru/library/analysis/clusterization/datamining/>. – Дата доступа: 12.03.2011.
6. Методы, средства и технологии исследования временных последовательностей статистических данных в имитационном моделировании: отчет о НИР (заключ.) / Беларус.-Рос. ун-т; рук. Е.А. Якимов; исполн.: Р.В. Петров [и др.]. – Могилев, 2011. – 126 с. – № ГР 20091957. – Инв. № Ф09М-171.
7. Pohlheim, H. Example Functions (single and multi-objective functions) [Электрон. ресурс] / H. Pohlheim // GEATbx: Genetic and Evolutionary Algorithm Toolbox for use with Matlab. – 2006. – Part 2. Parametric Optimization. – Режим доступа: <http://www.geatbx.com/docu/fcnindex-01.html>. – Дата доступа: 16.02.2011.

Поступила в редакцию 27.05.11.

In the introduction an object of investigation is pointed – criterion functions (CF) with similar optimization algorithms (OA) behavior. Optimization algorithms (OA), based on random search, can be used for simulation problem solving. Such algorithms have a set of parameters that require precise fine tuning in order to be efficiently applied to. Identification of classes of criterion functions, based on several clustering methods application, is given in this paper. The introduced technique can be used to improve efficiency of stochastic algorithms for multi-parameter optimization. The basic part. Identification of classes of criterion functions (CF) with similar OA behavior is one of phase of the fine tuning in software engineering kit for computer simulation of complex system BelSim. Clustering methods are used to assign a set of criterion functions into classes in which functional dependences are defined for the fine tuning of algorithm's parameters. In practice, an experiment is run on a criterion function to identify a class the criterion function belongs to. The finding of the problem solution is performed using the optimization algorithm in which the algorithm's parameters are adjusted based on the functional dependence of that class. Software engineering kit for computer simulation BelSim, statistics and analytics software package STATISTICA, spreadsheet application Microsoft Excel are used to perform an experimental research. Scientific advance is the formula to generalize results of several clustering methods. The conclusion shows the analysis of the results of CF research taking the CF features into account. The research results can be applied in the simulation and data processing.

**Keywords:** criterion function, optimization algorithm, cluster analysis, K-Means, Tree Clustering, Fuzzy Relation Clustering, BelSim, STATISTICA, MS Excel.

## References

1. Yakimov A.I. Technology of imitating modeling of control systems of the industrial enterprises [*Tekhnologiya imitatsionnogo modelirovaniia sistem upravleniia promyshlennykh predpriatii: monografiia*]. Mogilev, 2010, 304 p.: il.
2. Maksimei I.V. [i dr.]. Problems and models of research of operations [*Zadachi i modeli issledovaniia operatsii*]. Gomel, 1999, 150 p.: il.
3. Basharimov V.V., Borchik E.M., Yakimov A.I. A choice of cluster's analys methods at the decision of problems of optimization in imitating modeling [*Vybor metodov klasterного analiza pri reshenii zadach optimizatsii v imitatsionnom modelirovanii*]. *Informatsionnye tekhnologii, energetika i ekonomika (informatsionnye tekhnologii, matematicheskoe modelirovanie tekhnologicheskikh protsessov, elektronika)*: sb. trudov 7 Mezhd. (mezhd.) nauch.-tekhn. konf. studentov i aspirantov, 8-9 apr. 2010 g.: v 3 t. Smolensk, 2010, vol. 2, pp. 21-26.
4. Methods and models of the analysis of the data: OLAP and Data Mining [Metody i modeli analiza dannykh: OLAP i Data Mining] (Barsegian A.A. i [i dr.]). Saint Petersburg, 2004, 336 p.: il.
5. Paklin N. Algorithms of clusters on service Data Mining [*Algoritmy klasterizatsii na sluzhbe Data Mining*] [Elektron. resurs]. *Tekhnologii analiza dannykh*, 2011. Available at: <http://www.basegroup.ru/library/analysis/clusterization/datamining/>. Accessed: 12.03.2011.
6. Methods, means and technologies of research of time sequences of the statistical data in imitating modeling [*Metody, sredstva i tekhnologii issledovaniia vremennykh posledovatel'nostei statisticheskikh dannykh v imitatsionnom modelirovanii: otchet o NIR (zakliuch.)*]. Belorus.-Ros. un-t ; ruk. E.A. Yakimov; ispohn.: R.V. Petrov [i dr.]. Mogilev, 2011, 126 p. Bibliogr.: s. 124-126, no. GR 20091957. Inv. no. F09M-171.
7. Pohlheim H. Example Functions (single and multi-objective functions) [Elektron. resurs]. GEATbx: Genetic and Evolutionary Algorithm Toolbox for use with Matlab. 2006. Part 2. Parametric Optimization. Available at: <http://www.geatbx.com/docu/fcindex-01.html>. Accessed: 16.02.2011.



*Уважаемые авторы!*

*Болеe подробно требования к оформлению материалов, а также условия для  
принятия материалов см. на сайте журнала*

<http://vesnik.grsu.by>