

УДК 004.942

ОПРЕДЕЛЕНИЕ ПРИНАДЛЕЖНОСТИ ОБЪЕКТОВ БАЗОВОЙ ИНФОРМАЦИИ ДЛЯ ИНФОРМАЦИОННЫХ ПРОЦЕССОВ СЛОЖНЫХ СИСТЕМ К ЗАДАННОМУ КЛАССУ В ЗАВИСИМОСТИ ОТ ЗНАЧЕНИЙ ХАРАКТЕРИСТИК

¹Сумин В.И., ²Смоленцева Т.Е., ¹Дыбова М.А.

¹ФКОУ ВПО «Воронежский институт ФСИН России»,

Воронеж, e-mail: viktorsumin51@yandex.ru;

²ФГБОУ ВПО «Липецкий государственный технический университет», Липецк

Статья посвящена решению задачи по разработке метода группировки объектов системы на классы, позволяющего определять схожие по своим характеристикам объекты, входящие в сложную систему, которые относятся к определяемому классу. Рассмотрен процесс формирования алгоритма по разбиению информации на группы объектов, который состоит из формирования первоначального разбиения объектов с учетом их размерности, а также проверяется условие на принадлежность рассматриваемых объектов к выбранному классу. Сформированы первоначальные данные по распределению объектов управления на соответствующие классы. Для определения групп объектов, совпадающих по значениям параметров, использовали метод по группированию объектов сложной системы на классы, состоящий из этапов разбиения объектов по группам с учетом всех параметров и формирования групп из объектов сложной системы по соответствующим значениям параметров за текущий период.

Ключевые слова: управляющие воздействия, базовая информация, центр тяжести кластеров

IDENTIFY OBJECTS BASIC INFORMATION FOR INFORMATION PROCESSES OF COMPLEX SYSTEMS TO THE SPECIFIED CLASS, DEPENDING ON THE VALUES OF CHARACTERISTICS

¹Sumin V.I., ²Smolentseva T.E., ¹Dybova M.A.

¹FSE VPO «Voronezh Institute of the Federal penitentiary service of Russia»,

Voronezh, e-mail: viktorsumin51@yandex.ru;

²FGBOU VPO «Lipetsk state technical University», Lipetsk

The article is devoted to the task of developing a method of grouping objects into classes, allowing to determine similar characteristics of the objects in a complex system, which belong to the class being defined. The process of the formation of the algorithm by splitting the information into groups of objects, which consists of generating an initial partition objects taking into account their dimension, and also checked the condition on the considered objects belonging to the selected class. Formed initial data on the distribution of control objects to these classes. To define groups of objects that match the values of the parameters used method for grouping objects of a complex system into classes consisting of the steps of partitioning objects into groups taking into account all the parameters and the formation of groups of objects of a complex system according to the respective values of the parameters for the current period.

Keywords: control actions, basic information, the center of gravity of the clusters

Метод группировки объектов системы на классы позволит определять подобные по своим составляющим объекты, входящие в сложную систему, которые относятся к определяемому классу [6].

Для определения всех объектов, которые подобны по своим характеристикам, заданным в рассматриваемой системе, возможно применение итеративного метода кластерного анализа.

Базовая информация (БИ) представима в виде следующего множества $\{P_{ij}, A_i, B_j\}$, где $\overline{i} = \overline{1, I}$ – индекс объектов, т.е. носителей; $\overline{j} = \overline{1, J}$ – индекс всех характеристик выбранных рассматриваемых объектов; P_{ij} – количественное значение j -й характеристики

для i -го объекта; A_i – наименование i -го объекта; B_j – наименование j -й характеристики.

Для требуемых значений характеристик лицо, принимающее решение (ЛПР), определяет центр тяжести формируемого класса объектов P_j^{\ominus} .

Все элементы P_{ij} разделим на группы – принадлежащие и не принадлежащие классу с точностью T_1 .

Процесс разбиения информации на группы объектов представлен следующим образом:

0. Формирование исходного разбиения осуществляется на основе вспомогательной информации с использованием $\{r_i, d_i, a_i, k_i\}$ размерностью $\overline{i} = \overline{1, I}$:

r_i – смешанный момент корреляции Карла Пирсона или угловая мера

$$r_i = \sum_{j=1}^J (P_{i,j} - P_j^{\ominus}) / \sqrt{\sum_{j=1}^J (P_{i,j} - P_j^{\ominus})^2}; \quad (1)$$

d_i – евклидово расстояние от заданного центра тяжести до $P_{i,j}$

$$d_i = \sqrt{\sum_{j=1}^J (P_j^{\ominus} - P_{i,j})^2}; \quad (2)$$

a_i – индекс объекта в соответствии с $P_{i,j}$;
 k_i – признак принадлежности i -го объекта к выбранному классу (0 – принадлежит, 1 – не принадлежит, изначально все $k_i = 1$).

1. При первичном разбиении на группы необходимо выполнить следующие два этапа.

1.1. Вначале:

- первоначально $i = 1$;
- определяется значение s , где s – среднее расстояние между всеми элементами – d_i

$$s = \frac{\sum_{i=1}^{I-1} (d_i - d_{i+1})}{I}. \quad (3)$$

1.2. Рассчитывается расстояние между текущим элементом группы и следующим $\Delta d = d_i - d_{(i+1)}$.

1.3. Проверка условия на принадлежность объекта к выбранному классу осуществляется в процессе выполнения следующих шагов:

- Если $\Delta d \leq s$, то $k_{(i+1)} = 0$, $i = i + 1$ и:
 - если $i \leq I$, то переходим к пункту 1.2;
 - если $i > I$, то осуществляется переход к пункту 1.4.

Если $\Delta d > a$, то $i = i + 1$ и переход к пункту 1.2.

1.4. Далее необходимо объединить классы на основе смешанного момента корреляции Карла Пирсона r_i .

1.5. Вначале:

- элементы $\{r_i, k_i\}$ распределяются по возрастанию элементов k_i и r_i соответственно;
- определяем $i = 1$.

1.6. Определяем пороговое значение α , на основании данного параметра определяется принадлежность $i + 1$ объекта к рассматриваемому классу [2, 4]:

$$\alpha = (r_i - r_{(i+1)}).$$

Если $\alpha = 0$, то $i = i + 1$ и α вычисляется заново.

Если $\alpha < 0$, то $\alpha = |\alpha|$ и $i = 1$.

Если $\alpha > 0$, то $i = 1$.

1.7. Проверка на выполнение условия, признака окончания всех объектов класса.

Если $k_i = 0$, то переход к пункту 1.8.

Если $k_i = 1$, то переход к пункту 1.10.

1.8. Определяется расстояние между текущим и следующим элементами:

$$\Delta r = r_i - r_{(i+1)}.$$

1.9. Проверяется принадлежность $i + 1$ объекта к рассматриваемому классу:

Если $\Delta r > \alpha$, то $k_{(i+1)} = 1$, $i = i + 1$, и:

- если выполняется условие $i \leq I$, то переходим к пункту 1.7;
- если выполняется условие $i > I$, то переход к пункту 1.10.

1.10. Второй этап завершен. В результате $P_{i,j}$ разбиваются на две группы.

2. Вычисляются $\overline{P}_{k,j}$ – центры тяжести полученных групп:

$$\overline{P}_{k,j} = \frac{\sum_{(1,k_{k_i})}^I P_{a_i,j}}{\sum_{(1,k_{k_i})}^I 1}$$

$K = 2$ – индекс полученных групп.

3. Проверяется, находится ли каждый объект в ближайшей группе.

3.1. Первоначально $i = 1$, $n = 0$.

3.2. Вычисляется квадрат отклонения объекта a_i от центра тяжести групп:

$$F_{ka_i} = \sum_{j=1}^J (P_{a_i,j} - \overline{P}_{k,j})^2, \quad (4)$$

где $\overline{k} = 2$ – индекс полученных групп;
 $j = \overline{1}, \overline{J}$ – индекс характеристики, участвовавшей в формировании результата $P_{i,j} a_i$ объекта.

3.3. Если достигается $\min(F_{rai})$ при $k = k_p$, то рассматриваемый объект – a_i находится в ближайшей группе и изменение класса не происходит.

Если достигается $\min(F_{kai})$ при $k \neq k_p$, то объект – a_i соответственно не находится в ближайшей группе и поэтому $k_i = k$ и $n = n + 1$ [5].

- 3.4. Увеличивается $i = i + 1$ и выполняется:
 - если $i > I$, то закончился просмотр всех объектов и переход к пункту 4;
 - если $i \leq I$, то переход к пункту 3.2.

4. Если: $\frac{n}{I} \cdot 100 \geq T_t$ требуемая точность не достигнута, то осуществляется возврат к пункту 2.

Если $\frac{n}{I} \cdot 100 \leq T_t$, то требуемая точность итеративного процесса достигнута. Получено окончательное разбиение $P_{i,j}$ по классам.

При разбиении объектов управления на классы первоначальными данными являются:

$i = 1, I$ – индекс объекта управления;

$j = \overline{1, J}$ – индекс характеристики объекта управления;

P_{ij} – количественное значение j -й характеристики i -го;

A_i – наименование i -го объекта;

B_j – наименование j -й характеристики;

T_1 – требуемая точность разбиения в процентах;

P_j^0 – центр тяжести интересующего класса объектов.

Результатом разбиения объектов на группы являются:

$k_i = 0$ – признак, что i -й объект принадлежит к выбранной группе;

a_i – индекс объекта в соответствии с P_{ij} .

Оценка по выбору управляющих воздействий должна основываться на сравнении получаемых результатов за наблюдаемый промежуток времени с соответствующими значениями параметров из предыдущего периода, основной задачей которых является повышение эффективности сложной системы [3, 7]. Если итоговые значения принятых управляющих воздействий (УВ) не ухудшились, то УВ принимаются как эффективные и соответственно неэффективные в противном случае. Для сравнения результатов УВ обоих периодов необходимо вначале разбить их на близкие по своим значениям характеристик из группы и произвести оценку по групповым параметрам, что позволит уменьшить размерность решаемой задачи [1, 2].

Для выбора из всех объектов, совпадающих по значениям параметров групп объектов, используем методику группирования объектов сложной системы на соответствующие классы. Оценка эффективности выбранных УВ осуществляется следующим образом:

1. Формирование групп объектов по всем параметрам за предыдущий период.

Формируются первоначальные данные, необходимые для использования методики.

Из первичной информации, циркулирующей в сложной системе из множества $\{P_{ij}, A_i, B_j\}$, выделяем информацию за предыдущий период:

P_{ij} – количественное значение j -й характеристики i -го объекта;

$i = \overline{1, I}$ – индекс объектов управления, носителей первичной информации;

$j = \overline{1, J}$ – индекс характеристик;

A_i – идентификационные данные i -го объекта;

B_j – наименование j -й характеристики.

ЛПП определяет:

T_1 – требуемая точность разбиения в процентах;

K_0 – требуемое количество классов разбиения.

Получим следующие результаты:

\overline{K} – количество полученных классов;

$P_{k,j}$ – центры тяжести полученных классов;

K_i – номер класса, к которому принадлежит i -й объект;

a_i – индекс объекта в соответствии с P_{ij} .

2. Формирование групп объектов по значениям их характеристик за рассматриваемый период времени.

Формируются исходные данные, необходимые для использования методики [4, 6].

Из первичных данных выделяется информация за исследуемый период:

P_{ij} – количественное значение j -й характеристики i -го объекта;

$i = \overline{1, I}$ – индекс объектов;

$j = \overline{1, J}$ – индекс характеристик объектов.

A_i, B_j, T_1, K_0 – определены в пункте 1.

В результате работы данного алгоритма будут получены следующие характеристики:

\overline{K} – количество полученных классов;

$P_{k,j}^1$ – центры тяжести полученных классов;

k_i^1 – номер класса, к которому принадлежит i -й объект;

a_i^1 – индекс объекта в соответствии с P_{ij} ;

3. Исследование двух полученных групп классов.

K – количество классов, полученных в пункте 1;

$P_{k,j}$ – центры тяжести данных классов;

k_i – номер класса, к которому принадлежит i -й объект;

a_i – индекс объекта P_{ij} .

В пункте 2 получены:

\overline{K} – количество полученных классов;

$P_{k,j}^1$ – центры тяжести полученных классов;

k_i^1 – номер класса, к которому принадлежит i -й объект;

a_i^1 – индекс объекта в соответствии с P_{ij} .

3.1. Для сравнения полученных классов объектов преобразуются центры тяжести классов $\overline{P}_{k,j}$ и $P_{k,j}^1$, чтобы получить оценки сформированных групп классов независимыми от участвующих в анализе характеристик объектов следующим образом:

$$P_k = \sum_{j=1}^J W_j \overline{P}_{k,j}; \quad (5)$$

$$\overline{P}_k^1 = \sum_{j=1}^J W_j P_{k,j}^1, \quad (6)$$

где W_j – коэффициенты значимости j -го параметра исследуемого управляющего

объекта, которые определяет ЛПР; $k = \overline{1, K}$ – индекс полученных классов; $j = \overline{1, J}$ – индекс параметров, участвующих в анализе.

Полученные элементы (a_j, k_i) и (a_i^1, k_i^1) упорядочиваются в соответствии с a_i и a_i^1 .

3.2. Первоначально $S = 0, i = 0, n = 0$.

3.3. Если $a_i = a_i^1$ и $k_i \neq k_i^1$, тогда $S = S + (\overline{P_{k_i^1}} - P_{k_i})$ и $n = n + 1$;

3.4. $i = i + 1$.

Если $i < 1$, то осуществляется переход к пункту 3.3.

Если $i \geq I$, то переходим к пункту 3.5.

3.5. $\Theta = S/n \cdot 100$.

3.6. Оценка эффективности принятых в исследуемый период УВ:

– если $\Theta \geq 0$, то УВ принимались эффективные;

– иначе – неэффективные.

Список литературы

1. Жиликов Е.Г., Ломазов В.А., Ломазова В.И. Компьютерная кластеризация совокупности аддитивных математических моделей взаимосвязанных процессов // Вопросы радиоэлектроники. Сер. ЭВТ. – 2011. – Вып. 1. – С. 115–119.
2. Журавлев Ю.И., Рязанов В.В., Сенько О.В. «Распознавание». Математические методы. Программная система. Практические применения. – М.: Фазис, 2006.
3. Замятина О.М. Моделирование систем. – М.: Изд-во ТПУ, 2009. – С. 204.
4. Мендель, И.Д. Кластерный анализ. – М.: Финансы и статистика, 1988. – 176 с.
5. Саати Т.В. Аналитическое планирование: организация систем. – М.: Радио и связь: Изд-во Радио и связь, 1998. – 224 с.
6. Сумин В.И., Смоленцева Т.Е. Моделирование обучения с использованием временных рядов наблюдений: моно-

графия // Издательско-полиграфический центр «Научная книга», 2014. – 104 с.

7. Сумин В.И., Об алгоритмах и моделях, данных в решениях задач принятия решения / В.В. Цветков // Научные ведомости Белгород. гос. ун-та. Сер. История Политология Экономика Информатика. – 2010. – № 13 (84). – Вып. 15/1. – С. 120–128.

References

1. Zhiljakov E.G., Lomazov V.A., Lomazova V.I. Komp'yuternaja klasterizacija sovokupnosti additivnyh matematicheskih modelej vzaimosvjazannyh processov // Voprosy radioelektroniki. Ser. JeVT. 2011. Vyp. 1. pp. 115–119.
2. Zhuravlev Ju.I., Rjazanov V.V., Senko O.V. «Raspoznvanie». Matematicheskie metody. Programmaja sistema. Prakticheskie primenenija. M.: Fazis, 2006.
3. Zamjatina O.M. Modelirovanie sistem. M.: Izd-vo TPU, 2009. pp. 204.
4. Mendel I.D. Klasternyj analiz. M.: Finansy i statistika, 1988. 176 p.
5. Saati T.V. Analiticheskoe planirovanie: organizacija sistem. M.: Radio i svjaz: Izd-vo Radio i svjaz, 1998. 224 p.
6. Sumin V.I., Smolenceva T.E. Modelirovanie obucheniija s ispolzovaniem vremennyh rjadov nabljudenij: monografija // Izdatel'sko-poligraficheskij centr «Nauchnaja kniga», 2014. 104 p.
7. Sumin V.I., Ob algoritmah i modeljah, dannyh v reshenijah zadach prinjatija reshenija / V.V. Cvetkov // Nauchnye vedomosti Belgorod. gos. un-ta. Ser. Istorija Politologija Jekonomika Informatika. 2010. no. 13 (84). Vyp. 15/1. pp. 120–128.

Рецензенты:

Филатов Г.Ф., д.ф.-м.н., профессор кафедры математики, Военный учебно-научный центр Военно-воздушных сил, «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина», г. Воронеж;

Обуховский В.В., д.ф.-м.н., профессор, заведующий кафедрой высшей математики, Воронежский государственный педагогический университет, г. Воронеж.