

Лабораторная работа №6

КЛАССИФИКАЦИЯ ОБЪЕКТОВ МЕТОДОМ ИЕРАРХИЧЕСКОГО ГРУППИРОВАНИЯ

Цель работы: Изучить правила построения иерархических группировок, а также метод классификации объектов на основе иерархических группировок.

Порядок выполнения работы

1. Ознакомление с теоретической частью лабораторной работы.
2. Реализация классификации объектов с помощью иерархий.
3. Оформление отчета по лабораторной работе.

Исходные данные:

1. n – количество объектов группирования.
2. Таблица расстояний между объектами. Таблица заполняется автоматически случайными значениями.

Выходные данные: иерархии, построенные по критериям минимума и максимума. Результаты работы программы должны представляться в графическом виде.

Примечание: Результат работы представляется графически.

Методы распознавания, где классы известны заранее и разделяющие функции вырабатывались в процессе обучения, сильно влияют на выбор признаков и критериев разделения, от которых зависит получаемый результат.

Для того чтобы уменьшить влияние первоначальных сведений, их обогащают дополнительной информацией. Например, уточняют пространственные или временные отношения (общепринятое пространственное отношение: глаза на лице находятся выше носа); находят существующие отношения между исследуемыми объектами (в частности с помощью графов). Такие действия называются символическим описанием, которое получается в результате процедуры группирования, выполняющей роль и процедуры классификации.

Искомое символическое представление может иметь вид иерархической структуры, дерева минимальной длины или символического описания классов. Иерархия строится на основе понятия расстояния. Метод состоит в том, чтобы разработать последовательность разделений рассматриваемого множества на подгруппы, одна из которых обладает некоторым свойством, неприсущим другим. Искомая иерархия основывается на предъявляемых выборах. Поскольку их число весьма велико, иногда на одном и том же множестве исходных данных могут быть получены различные иерархии.

Рассмотрим правила построения иерархических группировок. Пусть X – множество, состоящее из m реализаций $\{X_1, X_2, \dots, X_m\}$, а $P(X)$ – множество

всех его частей: $P(X) = \{0, X_1, \{X_1, X_2\}, \{X_1, X_3\}, \dots, X_m\}$. Иерархией H называется подмножество, удовлетворяющее следующим условиям:

1. $X \in H$;
2. $\forall x_i \in X, x_i \in H$;
3. $\forall h, h' \in H$, если $h \cap h' \neq \emptyset$, то либо $h \subset h'$, либо $h' \subset h$.

На практике чаще всего используется иерархия, обозначаемая вещественной функцией, откладываемой вдоль оси ординат. Эта функция называется расстоянием в широком смысле слова, поскольку она не связана с евклидовым расстоянием между двумя точками. Выбор расстояния обуславливает построение иерархии.

Существует ряд алгоритмов для построения иерархических группировок и иерархий на их основе. Рассмотрим пример построения иерархии по критерию минимума. В этом случае иерархические группы A и B объединяются, если $d(A, B) = \min\{d(A, p), d(B, q)\}$.

Даны четыре атома (x_1, x_2, x_3, x_4) , расстояния между ними приведены в табл. 1.

Таблица 1

признаки	x_1	x_2	x_3	x_4
x_1	0	5	0,5	2
x_2	5	0	1	0,6
x_3	0,5	1	0	2,5
x_4	2	0,6	2,5	0

$d(x_1, x_3) = 0,5$ – минимальное расстояние, содержащиеся в таблице, следовательно, оно становится первым иерархическим объединением и обозначается $d(x_1, x_3) = \{a\}$, после чего элементы x_1 и x_3 в явном виде больше не участвуют в дальнейшем построении иерархии. Вместо них используется группировка a . Расстояния от нее до остальных элементов определяются следующим образом: $d\{a, x_2\} = \min\{d(x_1, x_2), d(x_3, x_2)\} = 1$;

$$d\{a, x_4\} = \min\{d(x_1, x_4), d(x_3, x_4)\} = 2.$$

Продолжая процесс сокращения, выделяем новую группировку $(x_2, x_4) = b$, в результате остаются две группы a и b , объединяемые окончательно в $c = \{a, b\} = 1$. На рис. 1 показано дерево, соответствующее исходным данным.

Для тех же исходных данных на рис. 2 приведена иерархия, построенная по критерию максимума. В этом случае иерархические группы A и B объединяются, если $d(A, B) = \max\{d(A, p), d(B, q)\}$.

Рассмотрим один из возможных вариантов построения такой иерархии. Заменим числа в таблице расстояний их обратными значениями, т. е. числами вида $1/5, 1/0,5, 1/2$ и т. д. После чего можно опять построить иерархию по критерию минимума, однако в этом случае самому маленькому числу будет

соответствовать максимальное число из исходных данных и т. д. по убыванию всех значений.

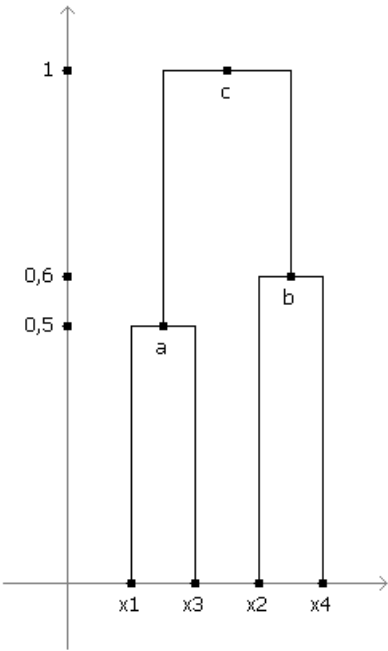


Рис. 1. Результирующее иерархическое дерево, построенное по критерию минимума

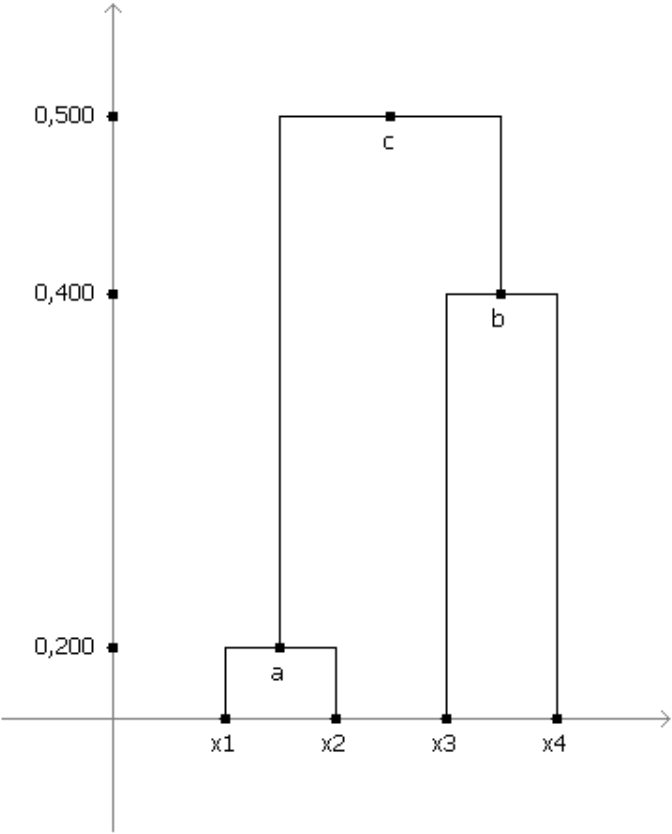


Рис. 2. Результирующее иерархическое дерево, построенное по критерию максимума