Основы теории принятия решений

Лекция 3

Содержание

1. Алгоритм Delta-1
2. Алгоритм Gamma-1
3. Выбор алгоритмов таксономии.
4. Пример 1.
5. Примеры прикладных задач таксономии:

- прогнозирование успеваемости;
- ранжирование объектов.

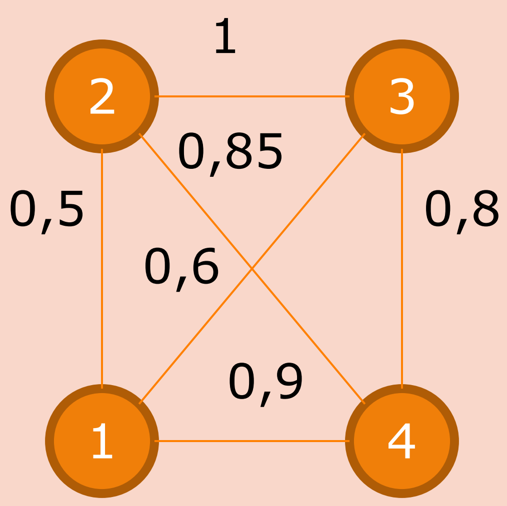
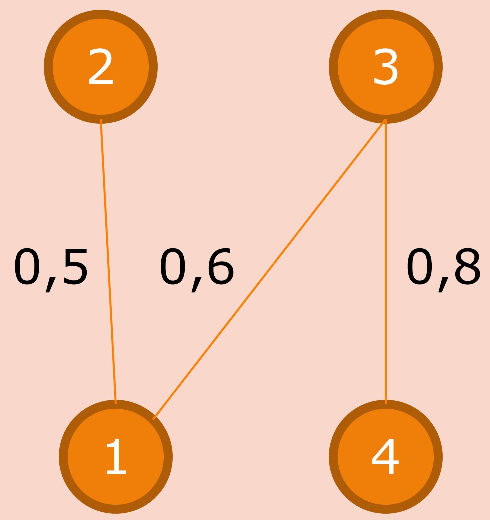
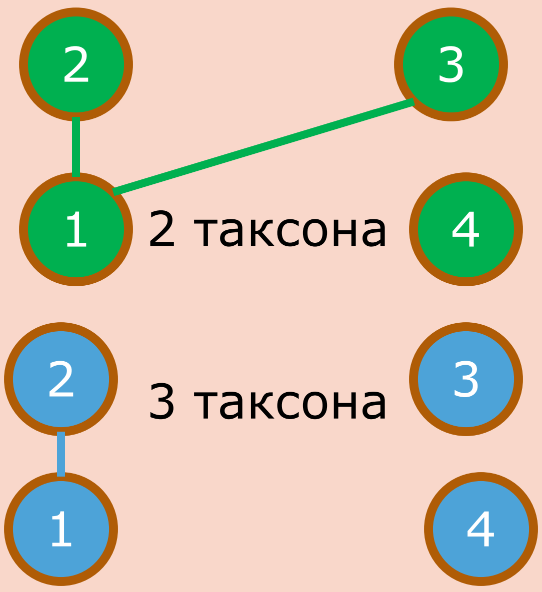
Таксономия в λ-пространстве с заданным числом таксонов

- **Цель**: распределить по W таксонам N объектов с неоднородными характеристиками.
- **Реализация**: алгоритм Delta-1.
- **Отличие от алгоритма Forel-2**: неоднородность характеристик объектов.

Алгоритм DELTA1

- Шаг 1. Ищется λ - расстояние между каждой парой объектов.
- Шаг 2. Строится полный взвешенный неориентированный граф G(X,U), вершины
- которого отвечают объектам, а каждого рёбра (p,q) равен расстоянию между Xp и Xq.
- Шаг 3. Алгоритмом Прима ищется минимальное связывающее подмножество рёбер, остальные рёбра удаляются.
- Шаг 4. Полученный граф обозначить G(X,U0).
- Шаг 5. i=1
- Шаг 6. Выбор ребра (p,q) с максимальным весом.
- Шаг 7. Ребро (p,q) отбрасывается: U0 = U0\(p,q).
- Шаг 8. Если i =W, то перейти к шагу 10, нет - к шагу 9.
- Шаг 9. i=i+1, перейти к шагу 6.
- Шаг 10. Конец алгоритма.

ПРИМЕР

<table>
<tr><th>Исходный граф (вес ребра = λ – расстоянию)</th><th>Остов графа, полученный алгоритмом Прима</th><th>Таксономии при W=2 и W=3</th></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
</table>

САМОСТОЯТЕЛЬНО

- Пользуясь DELTA 1, распределить по двум таксонам 5 объектов, каждый из которых обладает двумя разнородными характеристиками, заданными таблицей:

№	Признак 1	Признак 2
1	20	1
2	30	3
3	40	4
4	45	6
5	50	5

САМОСТОЯТЕЛЬНО

- Пользуясь алгоритмом DELTA 1, распределить по n таксонам 5 объектов, каждый из которых обладает двумя разнородными характеристиками, заданными на следующих двух слайдах

Персональные задания 1-9

Вариант 1			Вариант 2			Вариант 3		
n=2	Признак 1	Призна к 2	n=3	Признак 1	Признак 2	n=2	Призна к 1	Призна к 2
1	20	5	1	25	4	1	30	4
2	30	3	2	30	5	2	25	5
3	35	4	3	35	3	3	35	3
4	40	3	4	40	3	4	40	2
5	50	2	5	45	2	5	55	1
Вариант 4			Вариант 5			Вариант 6		
n=2	Признак 1	Призна к 2	n=3	Признак 1	Признак 2	n=2	Призна к 1	Призна к 2
1	15	5	1	10	5	1	10	4
2	20	3	2	20	4	2	15	5
3	25	4	3	25	3	3	20	3
4	35	2	4	30	3	4	25	2
5	40	1	5	50	2	5	40	1
Вариант 7			Вариант 8			Вариант 9		
n=2	Признак 1	Призна к 2	n=3	Признак 1	Признак 2	n=2	Призна к 1	Призна к 2
1	15	5	1	20	3	1	20	2
2	20	4	2	15	4	2	30	3
3	30	3	3	10	5	3	50	1
4	35	2	4	30	3	4	10	5
5	50	1	5	50	2	5	15	4

Персональные задания 10-18

Вариант 10			Вариант 11			Вариант 12		
n=3	Признак 1	Признак 2	n=4	Признак 1	Признак 2	n=3	Признак 1	Признак 2
1	20	5	1	25	4	1	30	4
2	30	3	2	30	5	2	25	5
3	35	4	3	35	3	3	35	3
4	40	3	4	40	3	4	40	2
5	50	2	5	45	2	5	55	1
Вариант 13			Вариант 14			Вариант 15		
n=3	Признак 1	Признак 2	n=4	Признак 1	Признак 2	n=3	Признак 1	Признак 2
1	15	5	1	10	5	1	10	4
2	20	3	2	20	4	2	15	5
3	25	4	3	25	3	3	20	3
4	35	2	4	30	3	4	25	2
5	40	1	5	50	2	5	40	1
Вариант 16			Вариант 17			Вариант 18		
n=3	Признак 1	Признак 2	n=4	Признак 1	Признак 2	n=3	Признак 1	Признак 2
1	15	5	1	20	3	1	20	2
2	20	4	2	15	4	2	30	3
3	30	3	3	10	5	3	50	1
4	35	2	4	30	3	4	10	5
5	50	1	5	50	2	5	15	4

САМОСТОЯТЕЛЬНО

- Изменить алгоритм Delta-1 таким образом, чтобы минимизировать верхнюю границу числа объектов, принадлежащих одному таксону (т.е. сделать распределение объектов между таксонами равномерным).
- Пользуясь приведенными выше персональными заданиями решить эту задачу при условии, что число таксонов $n=2$.

Парное сравнение алгоритмов таксономии алгоритмом Gamma-1.

Обозначения и определения

- Назначение алгоритма Gamma-1 заключается в том, чтобы попарно сравнивать различные алгоритмы таксономии. Для формального описания этого подхода далее используются следующие обозначения:
- S_i - таксономия, полученная i -м алгоритмом; «p» и «q»,- объекты;
- $r_i(p,q)$ - расстояние между «p» и «q», полученное i-м алгоритмом:
- (Очевидно, что $\forall$ p, $r_i(p,p)=0$).
- Величины $r_i(p,q)$ образуют матрицу μi (mxm матрица). $r_i(p,q)=0$, если p и q принадлежат одному таксону и $r_i(p,q)=1$, p и q принадлежат разным таксонам.

Алгоритм Gamma-1

- Шаг 1. Генерация матрицы μ1.
- Шаг 2.. Генерация матрицы μ2.
- Шаг 3. Определение максимального числа несовпадающих элементов $\beta = m(m-1)$.
- Шаг 4. Генерация новой матрицы μ3, каждый элемент которой $r_3(p,q)$ равен
- $r_3(p,q) = |r_1(p,q)-r_2(p,q)|/ \beta$.
- Шаг 5. Вычисление критерия F, равного сумме всех $r_3(p,q)$ и представляющего собой нормированное расстояние Хемминга между μ1 и μ2.
- Шаг 6. Конец алгоритма.

Пример: парное сравнение алгоритмов таксономии

- Пользуясь алгоритмом Gamma-1, матрицы μ1 и μ2 которого соответственно равны:

μ1				μ2			
0	1	0	1	0	1	1	1
0	0	0	1	0	0	0	1
1	0	0	0	0	0	0	1
1	0	1	0	0	1	0	0

определить величину F.

ОПРЕДЕЛЕНИЕ ВЕЛИЧИНЫ “F”

Матрица μ_3 каждый элемент которой следует умножить на 1/[4(4-1)] = 1/12.

0	**0**	**1**	**0**
0	**0**	**0**	**0**
1	**0**	**0**	**1**
1	**1**	**1**	**0**

F = 6/12 = 0.5

Выбор алгоритма таксономии

Пусть S_i - таксономия m объектов, полученная *i*-м алгоритмом, $F_{i,j}$ - нормированное расстояние Хемминга между i-м и j-м алгоритмами таксономии, полученное алгоритмом Gamma 1. Тогда характеристикой каждого i-го алгоритма F_i является сумма: $F_i = \sum_j F_{i,j}.$

Лучшим является q-й алгоритм, для которого справедливо: $F_q = \min_i F_i.$

Пример 1: условия

Определить, пользуясь Gamma 1, наилучший и наихудший из трех алгоритмов таксономии, матрицы μ1, μ2 и μ3 которых соответственно равны:

μ1=

0	0	1	1
0	0	1	1
0	1	0	0
1	1	0	0

μ2=

0	1	0	1
0	0	1	1
0	1	0	1
1	0	1	0

μ3=

0	1	1	1
1	0	0	0
1	0	0	0
1	0	0	0

Пример 1: решение

Вычисление характеристик каждого i-го алгоритма F_i (i=1,2,3):

$M_{12}=$

0	1	1	0
0	0	0	0
0	0	0	1
0	1	1	0

$F_{12}=0{,}41$; $M_{13}=$

0	1	0	0
1	0	1	1
1	1	0	0
0	1	1	0

$F_{13}=0{,}667$;

$M_{23}=$

0	0	1	0
1	0	1	1
1	1	0	1
0	0	1	0

$F_{23}=0{,}667$. $F_1= F_{12} +F_{13}=1{,}077$;

$F_2= F_{12} +F_{23}=1{,}077$;

$F_3= F_{13} +F_{23}=1{,}334$;

Лучшие алгоритмы- первый и второй, худший – третий.

САМОСТОЯТЕЛЬНО

- Определить наилучшую из четырех таксономий, представленных матрицами $\mu 1$, $\mu 2$, $\mu 3$ и $\mu 4$ на следующих трех слайдах. Номера вариантов указаны в правом столбце матрицы персональных данных.

Варианты 1 - 7

μ1				μ2				μ3				μ4				№
1	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	1
0	1	1	0	0	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	
1	0	1	1	0	0	1	1	1	0	1	0	0	0	1	1	2
0	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	
0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	3
0	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	0	1	1	
1	0	0	0	0	0	1	1	1	0	1	0	0	0	1	1	4
0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	0	1	0	1	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	0	
1	0	1	1	0	0	1	1	1	0	1	0	0	0	1	1	5
0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	0	1	1	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	0	1	
1	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	6
0	1	1	0	0	0	0	1	1	0	0	1	0	0	1	0	
1	0	1	1	1	1	0	0	0	1	1	0	1	0	1	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	
1	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	7
0	0	1	0	0	1	0	1	1	1	0	1	1	0	0	0	
1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	

Варианты 8 - 14

μ1				μ2				μ3				μ4				№
0	0	0	1	0	0	1	1	0	0	1	0	0	0	1	1	8
0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	0	0	0	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	0	0	1	1	0	0	1	1	0	
1	0	1	1	0	0	1	1	1	0	1	0	0	0	1	1	9
0	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	
0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	10
0	0	1	0	1	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	0	1	1	
1	0	0	0	0	0	1	1	1	0	1	0	0	0	1	1	11
0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	0	1	0	1	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	0	
1	0	1	1	0	0	1	1	0	0	1	0	0	0	1	1	12
0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	0	1	1	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	0	1	
1	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	13
0	1	1	0	0	0	0	1	1	0	0	1	0	0	1	0	
1	0	1	1	1	1	0	0	0	1	1	0	1	0	1	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	
1	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	14
0	0	1	0	0	1	0	1	1	1	0	1	1	0	0	0	
1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	0	

Варианты 15 - 21

μ1				μ2				μ3				μ4				№
0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	15
0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	0	0	0	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	0	0	1	1	1	0	1	1	1	
1	0	1	1	0	0	1	1	1	0	1	0	0	0	1	1	16
0	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	1	1	1	1	
0	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	17
0	1	1	0	1	0	0	1	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	1	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	0	1	1	
1	0	0	0	0	0	1	1	1	0	1	0	0	0	1	1	18
0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	0	1	0	1	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	0	
1	0	1	1	0	0	1	1	1	0	1	0	0	0	1	1	19
0	1	1	0	0	0	0	0	1	0	0	1	0	0	1	0	
1	0	0	1	1	1	1	0	0	1	1	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	0	1	
1	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	20
0	1	1	0	0	0	0	1	1	0	0	1	0	0	1	0	
1	0	1	1	1	1	0	0	0	1	1	0	1	0	1	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	
1	0	0	1	0	0	1	1	1	0	1	0	0	0	1	1	21
0	0	1	0	0	1	0	1	1	1	0	1	1	0	0	0	
1	0	0	1	1	1	1	0	0	1	0	0	1	0	0	0	
0	0	1	0	0	1	1	1	0	1	1	1	0	1	1	1	

Примеры прикладных задач таксономии

- Прогнозирование успеваемости.
- Ранжирование студентов.

Прогнозирование успеваемости – содержательная постановка задачи.

- Задана матрица, содержащая данные об оценках 3-х студентов по трем дисциплинам и одного – по первым двум. Требуется для последнего студента найти аналога среди первых двух, чтобы спрогнозировать его успеваемость по третьей дисциплине.

Решение задачи прогнозирования оценки первого ученика по 3-й дисциплине

Исходная матрица Нормированная матрица

№	Дисциплины		
	1	2	3
1	5	3	
2	4	4	5
3	4	3	3
4	5	4	4

№	Дисциплины		
	1	2	3
1	1	0	
2	1/2	1/2	5
3	1/2	0	3
4	1	1/2	4

Расстояния от первого ученика до остальных: r(1,2)=0,7; r(1,3)=0,5; r(1,4)=0,5. Прогнозируемая оценка: 3,5. Выбранные аналоги – третий и четвертый студенты.

Ранжирование студентов по успеваемости - условия

- Задана матрица М, содержащая данные об оценках 5-х студентов по трем дисциплинам. Требуется ранжировать их относительно отличника.
- М =

№	Дисц.1	Дисц.2	Дисц.3
1	4	5	3
2	5	4	4
3	3	5	5
4	4	4	4
5	2	5	5

Предложите *a priori* Вашу версию ранжирования.

Ранжирование студентов по успеваемости - нормирование

- Нормированная матрица М1 (шестой студент – эталон):

М1 =

Студенты	Дисциплины		
№	**1**	**2**	**3**
1	0,67	1	0,33
2	1	0,67	0,67
3	0,33	1	1
4	0,67	0,67	0,67
5	0	1	1
6	**1**	**1**	**1**

Ранжирование студентов по успеваемости - упорядочение

Расстояния от i-го студента до шестого ($0<i<6$):

- $r(1,6)=0,74868$;
- $r(2,6)=0,46669$;
- $r(3,6)=0,67$;
- $r(4,6)=0,5715$;
- $r(5,6)=1$.

Ранжирование студентов:

- $\pi = \{2, 4, 3, 1, 5\}$

САМОСТОЯТЕЛЬНО

- Ранжировать относительно двоечника учеников, успеваемость которых описывается матрицей М:

М =

Ученик	Дисц.1	Дисц.2	Дисц.3
1	5	4	3
2	2	3	4
3	5	5	4
4	3	3	2
5	3	4	2
6	5	5	3
7	2	2	4

САМОСТОЯТЕЛЬНО

- Определить прогноз оценки первого ученика по третьей дисциплине, полагая, что:
- Эта оценка неизвестна;
- Исходные данные приведены в матрице М на предыдущем слайде.

- **Смотрите прикладные аспекты теории принятия решений:**
- https://postnauka.ru/courses/28275