

Программные средства анализа качественных данных на основе детерминационного анализа

Нестругина В. В.

д. т. н., проф. Воронина И. Е.

Направления обработки качественных данных

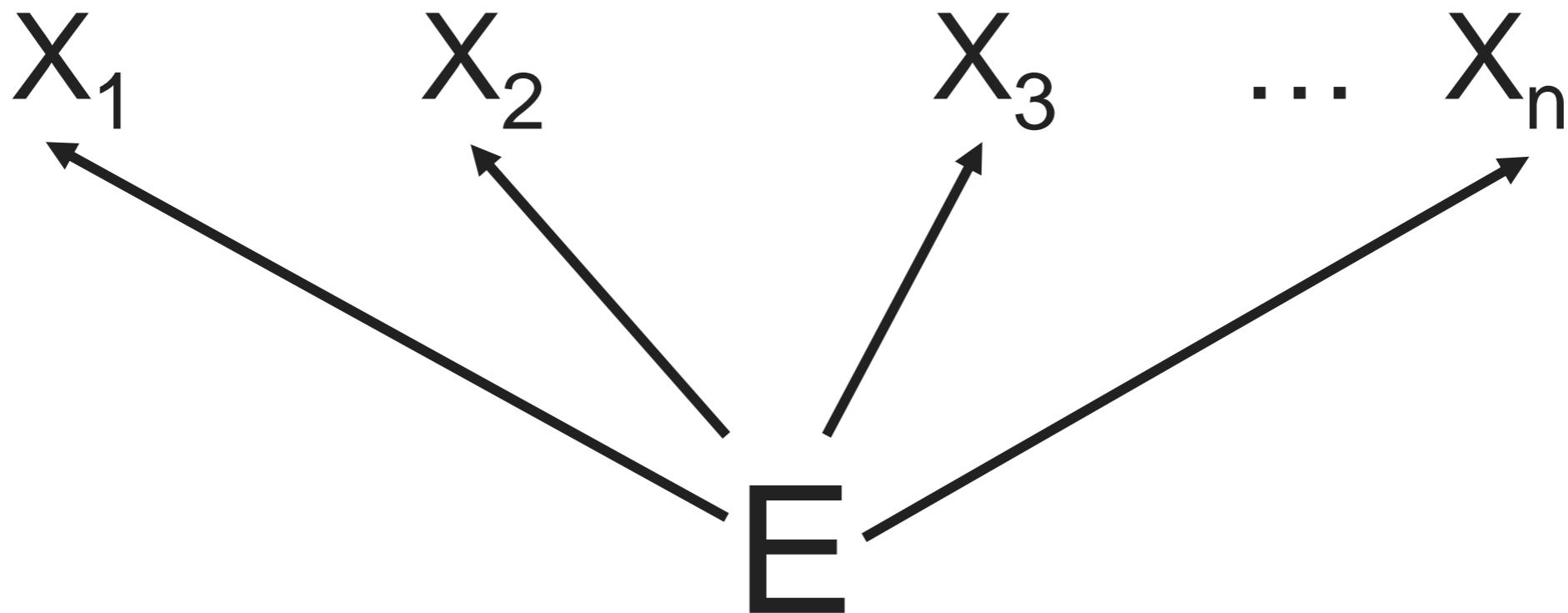
1. Квантификация значений качественной переменной.
2. Использование интегральных показателей.
3. Детерминационный анализ (анализ детерминаций **a**→**b**)

Постановка задачи

Разработать систему для анализа качественных данных на основе детерминационного анализа. Для этого необходимо решить следующие задачи:

1. Обеспечить поддержку решения основных задач детерминационного анализа:
 - получение объяснения заданного свойства
 - получение уточнений для определенного решения
 - получение дополнений для совокупности решений задачи 1
 - определение существенности контекста, объясняющих и объясняемых свойств
 - построение объясняющей и объясняемой типологий
 - проверка объяснительных возможностей и объяснимости типологии
2. Обеспечить поддержку агрегированных задач на базе основных

Веер отображений



Веер отображений – совокупность отображений вида $E \rightarrow X_i$, $i = \overline{1, n}$

E – множество объектов

X_i – множество значений переменной x_i

i – индекс, нумерующий переменные.

Таблица сопряженности

		у			
Y	$y^{(3)}$	0	0	60	
	$y^{(2)}$	0	50	70	
	$y^{(1)}$	50	0	70	
		$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	X

Интенсивность и ёмкость

()

(|)

—

()

—

Приращение интенсивности

() () ()

Основная задача детерминационного анализа

$((()))$
 $((()))$

Сравнение детерминационного анализа с известными статистическими методами

Метод	Принцип номинальности	Принцип конкретности	Принцип ограниченной статистичности
Детерминационный анализ	+	+	+
Статистическая связь по определению класса вероятностно-статистических методов	+	+	—
Методы исследования связей на основе критерия	+	—	—
Расстояние по Хеммингу между разбиениями объектов	+	—	—
Уравнение регрессии	—	—	—
Метод главных компонент	—	—	—

Пример применения системы

Вопрос 1: Кто наиболее склонен с высоким оценкам – мужчины или женщины?

X: код 2 – пол

- 1: женский
- 2: мужской

Y: код 29 – оценка, значение – 3,49–4/4 (код 5)

Найденное D-отношение:

2	->	29	I	C	N(X&Y)	N(X)	N(Y)
2	->	5	0.218391	0.760000	19	87	25
1	->	5	0.103448	0.240000	6	58	25

Пример применения системы

Вопрос 2: Можно ли объяснить успех мужчин уровнем образования родителей?

11	I	ΔI
1	0.241379	0.022989
4	0.428571	0.210181

Приращение интенсивности при уровне образования матери (код 11) для мужчин

Коды:

- 1 – начальная школа
- 2 – средняя школа
- 3 – старшая школа
- 4 – бакалавриат
- 5 – магистратура
- 6 – докторантура

12	I	ΔI
2	0.300000	0.081609
3	0.266667	0.048276

Приращение интенсивности при уровне образования отца (код 12) для мужчин

Пример применения системы

Вопрос 2: Можно ли объяснить успех мужчин уровнем образования родителей?

11	12	I	ΔI
1	3	0.333333	0.114943
1	4	0.250000	0.031609
1	1	0.250000	0.031609
3	2	1.000000	0.781609
4	2	1.000000	0.781609
4	3	1.000000	0.781609
4	1	0.500000	0.281609

Коды:

1 – начальная школа

2 – средняя школа

3 – старшая школа

4 – бакалавриат

5 – магистратура

6 – докторантура

Приращение интенсивности при учете
уровня образования и матери, и отца
для мужчин

Пример применения системы

Вопрос 3: Можно ли построить типологическое свойство на основе привычек в учебе мужчин?

22	25	26	->	29
2	3	1	->	3

22	25	26	->	29
1	1	2	->	4
2	3	3	->	4

22	25	26	->	29
1	2	3	->	5
1	1	3	->	5

Таблица для
типологического свойства

X_1 : 22 – посещение занятий
1 – всегда
2 – иногда
3 – никогда

X_2 : 25 – ведение записей
1 – никогда
2 – иногда
3 – всегда

X_3 : 26 – внимательное прослушивание лекций
1 – никогда
2 – иногда
3 – всегда

Y : 29 – средний балл (max = 4)
1 – 2,50–2,99
2 – 3,0–3,49
3 – 3,5–4

Результаты работы

Реализованы программные средства, обеспечивающие поддержку решения основных и агрегированных задач детерминационного анализа:

- получение объяснения заданного свойства
- получение уточнений для определенного решения
- получение дополнений для совокупности решений задачи 1
- определение существенности контекста, объясняющих и объясняемых свойств
- построение объясняющей и объясняемой типологий
- проверка объяснительных возможностей и объяснимости типологии