

Лекция 2. Основные понятия Теории Измерений. Представление различных типов данных эмпирическими системами.

Рассмотрим роль и смысл числового представления.

Смысл состоит в том, чтобы значениям величины приписать числа так, чтобы исходные отношения и операции преобразовывались в некоторые «простые» и «удобные» числовые отношения и операции. В этом случае по значениям числовых отношений и операций легко определяются значения исходных отношений и операций.

Пример с коромысловыми весами, линейкой и законом Гука ...

Пусть наша исследуемая реальность или, в частном случае, величина представлена эмпирической системой.

Определение 1. Эмпирической системой $\mathfrak{S} = \langle A; \Omega_{\mathfrak{S}} \rangle$ сигнатуры Ω будем называть алгебраическую систему сигнатуры Ω , в которой основное множество A – множество объектов, а $\Omega_{\mathfrak{S}} = \{P_0^{\mathfrak{S}}, P_1^{\mathfrak{S}}, \dots, P_n^{\mathfrak{S}}, \rho_1^{\mathfrak{S}}, \dots, \rho_m^{\mathfrak{S}}, c_0^{\mathfrak{S}}, c_1^{\mathfrak{S}}, c_2^{\mathfrak{S}}, \dots\}$ – множество отношений, операций и констант сигнатуры Ω , интерпретируемых в системе понятиях предметной области ♦

Сигнатура

$$\Omega = \langle P_0, P_1, \dots, P_n, \rho_1, \dots, \rho_m, c_0, c_1, c_2, \dots \rangle,$$

где P_i , $i \leq n$ – предикатные символы;

ρ_j , $j \leq m$ – символы операций;

c_l , $l \in I$ – символы констант ($I = \emptyset$, $I \subset \omega$, $\omega = \{0, 1, 2, \dots\}$).

Эмпирическую систему мы никогда не имеем реально, мы знаем систему аксиом, которой она удовлетворяет. Пусть знания о некоторой величине, свойстве, признаке сформулированы в виде некоторой системы аксиом Σ сигнатуры Ω .

Определение 2. Величиной будем называть эмпирическую систему $\mathfrak{S} = \langle A; \Omega_{\mathfrak{S}} \rangle$ сигнатуры Ω , удовлетворяющую системе аксиом Σ ♦

Определение 3. Числовыми системами называются системы $\mathfrak{R} = \langle Re, \Omega_{\mathfrak{R}} \rangle$ сигнатуры Ω , где Re – множество действительных чисел, где

$$\Omega_{\mathfrak{R}} = \{P_1^{\mathfrak{R}}, \dots, P_n^{\mathfrak{R}}, \rho_1^{\mathfrak{R}}, \dots, \rho_m^{\mathfrak{R}}, c_0^{\mathfrak{R}}, c_1^{\mathfrak{R}}, c_2^{\mathfrak{R}}, \dots\}$$

множество отношений, операций и констант, определенных на Re ♦

Определение 4. Шкалой (числовым представлением) величины $\mathfrak{S} = \langle A; \Omega_{\mathfrak{S}} \rangle$ называется сильный гомоморфизм эмпирической системы $\mathfrak{S} = \langle A; \Omega_{\mathfrak{S}} \rangle$ в числовую систему $\mathfrak{R} = \langle Re, \Omega_{\mathfrak{R}} \rangle$ той же сигнатуры Ω отображающий значения величины A в числа μ : $A \rightarrow Re$ и удовлетворяющий условиям:

$$1. P_i^{\mathfrak{S}}(a_1, \dots, a_{mi}) \Leftrightarrow P_i^{\mathfrak{R}}(\mu a_1, \dots, \mu a_{mi}), i = 0, 1, \dots, n;$$

$$2. \mu \rho_j^{\mathfrak{S}}(a_1, \dots, a_{mj}) = \rho_j^{\mathfrak{R}}(\mu a_1, \dots, \mu a_{mj}), j = 1, \dots, m;$$

$$3. \mu c_l^{\mathfrak{S}} = c_l^{\mathfrak{R}}, l \in I.$$

Сильный гомоморфизм означает, что, если предикат $P_i^{\mathfrak{R}}(\mu a_1, \dots, \mu a_{mi})$ истинен на некотором наборе $\langle \mu a_1, \dots, \mu a_{mi} \rangle$, то существует набор $\langle b_1, \dots, b_{mi} \rangle$, $\mu b_1 = \mu a_1, \dots, \mu b_{mi} = \mu a_{mi}$, на котором предикат $P_i^{\mathfrak{S}}(b_1, \dots, b_{mi})$ также истинен ♦

Введем обозначения:

$AC(\Sigma)$ – множество алгебраических систем удовлетворяющих системе аксиом Σ ;

$F(\mathfrak{S}, \mathfrak{R})$ – множество шкал величины \mathfrak{S} отображающих ее в заданную числовую систему \mathfrak{R} .

В теории измерений исследуются три основные проблемы.

Проблема существования. Для данной системы аксиом Σ величины найти достаточно простую и удобную числовую систему \mathfrak{R} такую, что бы можно было доказать, что для любой величины $\mathfrak{S} \in AC(\Sigma)$ существует шкала, т.е. $F(\mathfrak{S}, \mathfrak{R}) \neq \emptyset$ ♦

Из формулировки проблемы существования следует, что система аксиом Σ должна быть достаточно сильной, что бы для выбранной числовой системы \mathfrak{X} можно было построить шкалу для любой величины $\mathfrak{Z} \in AC(\Sigma)$.

Доказательство, что любая эмпирическая система, удовлетворяющая системе аксиом, сильным гомоморфизмом отображается в выбранную числовую систему, предъявляет сильные требования к системе аксиом. Приходится включать в нее аксиомы, не поддающиеся экспериментальной проверке, как, например, аксиома Архимеда, а также «чисто технические» аксиомы, не изменяющие множества экспериментально проверяемых следствий. Такие аксиомы часто отражают свойства числовой системы, а не свойства величин. Это противоречит содержанию систем аксиом, как результатам экспериментального анализа свойств величин.

Пример: Система с отношениями $\mathbf{A} = \langle A, < \rangle$ называется интервальным порядком, если на A выполнены аксиом:

1. $\neg(a < a)$ – нерефлексивность;
2. $(a < b) \& (c < d) \Rightarrow (a < d) \vee (c < b)$.

Теорема (Фишберн): Если $\mathbf{A} = \langle A, < \rangle$ интервальный порядок, и A не более чем счётно, то существуют функции $U, V: A \rightarrow \text{Re}$ такие, что :

1. $V(a) > 0$;
2. $a < b \Leftrightarrow U(a) + V(a) < U(b)$.

Проблема единственности: Для выбранной числовой системы \mathfrak{X} определить все множество шкал $F(\mathfrak{Z}, \mathfrak{X})$ величины $\mathfrak{Z} \in AC(\Sigma)$ ♦

Это множество можно определить, найдя группу допустимых преобразований шкалы.

Обозначим через $\Gamma(\mathfrak{X})$ – группу всех автоморфизмов числовой системы \mathfrak{X} на себя.

Известно, что, если $\mu \in F(\mathfrak{Z}, \mathfrak{X})$, то $\gamma\mu \in F(\mathfrak{Z}, \mathfrak{X})$, $\gamma \in \Gamma(\mathfrak{X})$ тоже шкала.

Группа $\Gamma(\mathfrak{X})$ называется **группой допустимых преобразований шкал** $F(\mathfrak{Z}, \mathfrak{X})$.

Таблица 1. Числовые типы данных.

Допустимые преобразования	Группы допустимых преобразований	Шкалы
$x \rightarrow f(x)$,	$f: \text{Re} \rightarrow (\text{на}) \text{Re}$, взаимно-однозначные преобразования	Номинальная
$x \rightarrow f(x)$,	$f: \text{Re} \rightarrow (\text{на}) \text{Re}$ монотонные преобразования	Порядка
$x \rightarrow rx + s$, $r > 0$	Позитивная аффинная группа	Интервалов
$x \rightarrow tx^r$, $t, r > 0$	Степенная группа	Логарифмически-интервальная
$x \rightarrow x + s$	Группа сдвига	Разностей
$x \rightarrow tx$, $t > 0$	Группа подобия	Отношений
$x \rightarrow x$	Тождественная группа	Абсолютная

Проблема адекватности. Числовые утверждения должны быть инвариантны относительно произвола в выборе шкал величин из $F(\mathfrak{Z}, \mathfrak{X})$ или это эквивалентно тому, что числовые утверждения должны быть инвариантны относительно допустимых преобразований шкал.

В законах физики проверяется совпадение размерностей.

Точно также методы извлечения знаний должны быть инвариантны относительно допустимых преобразований шкал.

Представление известных типов данных эмпирическими системами.

Анализ эмпирического содержания данных должен начинаться с представления соответствующих данных эмпирическими системами.

Например, Что представляет собой физическая величина температуры, если рассмотреть ее использование в медицине, почвоведении или кулинарии? В этих областях это косвенные измерения некоторых других величин, эмпирические системы которых сначала надо определить и затем постараться определить шкалу этой величины.

Покажем, каким образом такие известные типы данных, как:

матричное представление бинарных отношений;

матрицы упорядочений;

матрицы близости;

матрицы объект-признак;

могут быть представлены эмпирическими системами.

Эти типы данных встречаются в таких областях, как экспертное оценивание, социология, психология, психофизика, геология, медицина, сельское хозяйство и т.д. Все эти области характеризуются тем, что в них встречаются признаки и величины самой разнообразной природы.

Данный раздел преследует следующие цели:

1. Показать, что эмпирические системы являются достаточно общим и более естественным способом представления данных.
2. Привести для каждого типа данных, используя представление их эмпирическими системами, относящиеся к ним результаты теории измерений. Эти результаты включают в себя системы аксиом и теоремы представления и единственности, указывающие, какие числовые представления для данных систем аксиом существуют и какие у них шкалы. По шкалам величин можно определять группы допустимых преобразований, что позволяет корректно применять методы анализа данных, инвариантные относительно соответствующих групп допустимых преобразований.
3. Для каждого типа данных привести основные существующие в настоящее время методы их обработки и провести критику их применимости.

Рассмотрим сначала данные, в которых многоместные отношения возникают естественным образом в силу специфики самого объекта исследования. Как отмечается в работах, источником информации часто являются суждения человека. Как показали многие эксперименты, человек более правильно и с меньшими затруднениями отвечает на вопросы качественного, в частности, сравнительного характера, чем количественного. В различных дисциплинах человек называется по-разному: как эксперт в экспертных оценках, как испытуемый в психологии и психофизике, как респондент в социологии, как пациент в медицине и т.д.

1. Матричное представление бинарных отношений. Бинарное отношение $P(a,b)$, определенное на множестве объектов $A = \{a_1, \dots, a_m\}$, задается матрицей (e_{ij}) , $i, j = 1, \dots, m$; где $e_{ij} = 1(0)$ означает, что $P(a_i, a_j)$ истинно (ложно). Такой матрицей можно задать произвольное бинарное отношение на множестве A .

Такое представление широко используется ввиду его привычности и простоты. Наиболее часто используются отношения эквивалентности, квазипорядка, частичного порядка и лексикографического порядка. Данные, включающие эти отношения, встречаются в следующих задачах:

1.1. Отношение эквивалентности: Задает некоторое разбиение множества объектов. С его помощью задают: номинальные признаки (признаки в шкале наименований), в частности признаки, определяющие принадлежность к образу в распознавании образов; результаты

классификации, таксономии и кластеризации, полученные как опросом экспертов, так и применением машинных методов.

1.2. Отношения порядка и квазипорядка. Любой признак измеримый в шкале порядка, задает некоторое отношение порядка, например, шкала Морса твердости минералов или шкала силы ветра. Упорядочения объектов экспертами. Упорядочения, получаемые методами ранжирования.

1.3. Отношения частичного и древовидного порядка. Возникают в лингвистике при построении дерева связей. В иерархической классификации, при задании вложенных классов или таксонов. В психологии и других областях, при задании дерева целей. В социологии отмечается, что для социологических данных более типичны отношения частичного порядка и толерантности, чем порядка и квазипорядка. В психологии также возникают не транзитивные предпочтения.

Матрица бинарного отношения фиксирует некоторое бинарное отношение P , которое включается в словарь Ω_3 эмпирической системы $\mathfrak{S} = \langle A; \Omega_3 \rangle$.

Приведем результаты теории измерений, относящиеся к одному бинарному отношению P .

1.4. Отношение толерантности $a, b \in A$:

1. $P(a, a)$;

2. $P(a, b) \Leftrightarrow P(b, a)$.

Числового представления не существует.

1.5. Отношение эквивалентности $a, b, c \in A$:

1. $P(a, a)$;

2. $P(a, b) \Leftrightarrow P(b, a)$;

3. $P(a, b) \& P(b, c) \Rightarrow P(a, c)$.

Числовое представление существует в виде нумерации классов.

1.6 Отношение частичного порядка, для любых $a, b, c \in A$:

1. $P(a, a)$;

2. $P(a, b) \& P(b, c) \Rightarrow P(a, c)$.

Числового представления не существует.

1.7. Отношение интервального упорядочения: для любых $a, b, c, d \in A$:

1. $\neg P(a, a)$;

2. $P(a, b) \& P(c, d) \Rightarrow (P(a, d) \vee P(c, b))$.

Числовое представление существует.

Существуют две вещественнозначные функции $U, V: A \rightarrow \text{Re}^+$,

такие, что для любых $a, b \in A$.

$P(a, b) \Leftrightarrow (U(a) + V(a)) < U(b)$.

1.8. Отношение полупорядка. Отношение P называется отношением полупорядка, если оно является отношением интервального порядка и для любых $a, b, c, d \in A$ удовлетворяет аксиоме:

3. $P(a, b) \& P(b, c) \Rightarrow P(a, d) \vee P(d, c)$.

Числовое представление существует.

Существует вещественнозначная функция $U: A \rightarrow \text{Re}$ такая, что для любых $a, b \in A$,

$P(a, b) \Leftrightarrow (U(a) + 1) < U(b)$.

1.9. Отношение древесного порядка. Отношение P называется отношением древесного порядка, если для любых $a, b, c \in A$ удовлетворяет аксиоме:

1. $(a < b) \& (a < c) \Rightarrow (b < c) \vee (c < b)$;

2. Существует наибольший элемент.

Числового представления не существует.

1.10. Отношение слабого порядка (квазисерии, предпорядки).

Для любых $a, b, c \in A$ удовлетворяет аксиомам:

1. $P(a, b) \vee P(b, a)$;
2. $P(a, b) \& P(b, c) \Rightarrow P(a, c)$.

Если упорядоченная система $\langle A; P \rangle$ имеет счетную базу, то числовое представление существует.

Не все из приведенных отношений имеют числовые представления. Поэтому не всегда данные, содержащие бинарные отношения, можно представить в некотором числовом пространстве.

Рассмотрим, какие в настоящее время существуют методы обработки бинарных отношений. Большинство методов использует для обработки матриц расстояния или меры близости между матрицами. Эти расстояния и меры вводятся либо, исходя из систем аксиом, либо из статистических предположений и свойств самих отношений, как, например, коэффициенты Стьюарта, ранговой корреляции Кендала, Спирмена, Юла, информационные меры и т.д. Введение расстояний и мер близости связано с определенными дополнительными предположениями. К методам, использующим расстояния, относятся методы анализа структуры связей между объектами, методы классификации, методы построения регрессии и другие.

2. Матрицы упорядочений: (r_{ij}) , $i = 1, \dots, m$; $j = 1, \dots, n$; r_{ij} - оценка i -го объекта по j -му признаку. Такие матрицы могут выражать либо упорядочения k объектов n экспертами, либо упорядочения k объектов по n ранговым признакам. Такие матрицы обрабатываются методами многомерного шкалирования, а также некоторыми из методов обработки матричного представления бинарных отношений.

Поставим в соответствие каждому признаку j отношение P_j определенное следующим образом:

$$P_j(a_{i1}, a_{i2}) \Leftrightarrow r_{i1j} < r_{i2j}.$$

В теории измерений разработано много систем аксиом, определяющих взаимодействие нескольких отношений порядка.

3. Матрицы близости. Пусть дано некоторое множество объектов $A = \{a_1, \dots, a_m\}$. Матрицей близости для этих объектов называется матрица (g_{ij}) , $i, j = 1, \dots, m$; g_{ij} - числовые оценки меры близости (сходства или различия) в порядковой шкале (имеет смысл только сравнение величин $g_{i1j1} < g_{i2j2}$). Такие матрицы возникают в различных областях при сравнении или оценке экспертом двух объектов в некотором отношении.

Матрицы близости обрабатываются методами многомерного неметрического шкалирования. Целью этих методов является представление объектов точками в некотором метрическом пространстве (Евклидовом или Римановом) минимальной размерности так, чтобы расстояния t_{ij} между ними с точностью до порядка соответствовали бы величинам g_{ij} . После применения методов многомерного шкалирования мы получаем представление данных в метрическом пространстве.

Определим на множестве A отношение $g_{i1i2} < g_{i3i4}$.

В теории измерений эмпирические системы, включающие подобные четырехместные отношения, обозначаются как $M = \langle A^*; \leq \rangle$, где $A^* \subset A \times A$, \leq - бинарное отношение упорядочения, определенное на A^* . Приведем некоторые результаты теории измерений относящиеся к таким эмпирическим системам.

3.1. Шкала положительных разностей.

Существует гомоморфизм $\Phi: A^* \rightarrow \text{Re}$, $A \neq \emptyset$, такой, что для любых пар $(a, b), (b, c), (c, d)$ из A^* :

а) $(a, b) \leq (c, d) \Leftrightarrow \Phi(a, b) \leq \Phi(c, d)$

б) $\Phi(a, c) = \Phi(a, b) + \Phi(b, c)$.

Отображение Φ единственно с точностью до положительного множителя (шкала отношений).

3.2. Шкала алгебраических разностей. $A^* = A \times A$.

Существует гомоморфизм $\Phi: A \rightarrow \text{Re}$ такой, то для любых $a, b, c, d \in A$:

$$a) (a, b) < (c, d) \Leftrightarrow (\Phi(a) - \Phi(b)) < (\Phi(c) - \Phi(d)).$$

Отображение Φ , обладающее этим свойством, единственно с точностью до лог-линейных преобразований (шкала интервалов).

3.3. Шкала разностей равных конечных промежутков. $A^* = A \times A$, A - конечно, $A^* \neq \emptyset$.

Существует гомоморфизм $\Phi: A \rightarrow \mathbb{N}$ (натуральные числа), такой, что для любых $a, b, c, d \in A$:

$$(a, b) \leq (c, d) \Leftrightarrow \Phi(a) - \Phi(b) \leq \Phi(c) - \Phi(d).$$

Отображение Φ единственно с точностью до линейных преобразований (шкала интервалов).

3.4. Шкала абсолютных разностей. $A^* = A \times A$. Существует гомоморфизм $\Phi: A \rightarrow \text{Re}$ такой, что:

$$(a, b) < (c, d) \Leftrightarrow |\Phi(a) - \Phi(b)| < |\Phi(c) - \Phi(d)|.$$

Отображение Φ единственно с точностью до линейных преобразований (шкала интервалов).

4. Матрица объект-признак. (x_{ij}) , $i = 1, \dots, m$; $j = 1, \dots, n$; $x_{ij} = x_j(a_i)$ числовое значение j -го признака x_j на i -ом объекте. Признаки могут быть самыми произвольными: как количественными, так и качественными. Тот факт, что такая матрица получена в результате некоторых измерений (опросов, экспериментов, обследований и т.д.), говорит о том, что существует n измерительных процедур x_j . Такие измерения называют приборными или косвенными измерениями.

Рассмотрим, как можно определить эмпирическую систему приборных измерений.

Для каждого прибора x_j и некоторого числового отношения $R(y_1, \dots, y_k)$, определенного на Re , можно определить следующее эмпирическое отношение на множестве объектов A :

$$P_j^R(a_1, \dots, a_k) \Leftrightarrow R(x_j(a_1), \dots, x_j(a_k)).$$

Это отношение может не иметь эмпирической интерпретации. Прибор $x_j(a)$ имеет эмпирическую интерпретацию, но связь его значений отношением R может уже не иметь эмпирическую интерпретацию. Поэтому нужно найти такие числовые отношения на Re , для которых отношение P_j^R имело бы эмпирическую интерпретацию.

Предположим, что мы перебрали некоторые, наиболее распространенные числовые отношения и нашли, что отношения $P_j^{R^1}, \dots, P_j^{R^k}$ имеют эмпирическую интерпретацию. Данное множество отношений не пусто, так как по крайней мере отношение

$$P_j^=(a_1, a_2) \Leftrightarrow x_j(a_1) = x_j(a_2)$$

имеет эмпирическую интерпретацию состоящую в том, что на объектах a_1 и a_2 величина x_j принимает одно и то же значение. Отношение $P_j^=$, как правило, является отношением эквивалентности.

В теории измерений известно много систем аксиом, использующих для некоторых величин только отношение $P_j^=$ и приводящих, тем не менее, к сильным шкалам. Поэтому наличие в языке эмпирических систем одного лишь отношения $P_j^=$ может много дать.

Определим множество отношений для приборного измерения x_j как множество $\{P_j^{R^1}, \dots, P_j^{R^1}\}$.

Эмпирической системой для матрицы объект-признак будет система, включающая объединение всех отношений для всех приборных измерений.