Упорядоченные множества в анализе данных

Тема 1. Отношения и графы

С.О. Кузнецов

Цель курса

Познакомить с основами теории решеток и ее приложениями в анализе данных

Построение таксономий объектов и их визуализация

Машинное обучение (Machnie Learning)

Разработка данных (Data Mining), и в частности

Выявления знаний в текстах (Text Mining)

Обзор курса

1. Отношения и упорядоченные множества
   Основные понятия: отношение, частичный порядок
2. Решетки
   Основные темы: инфимум, супремум, типы решеток
3. Бинарные отношения и соответствия Галуа
   Основной пример: соответствия между объектами и их признаками
4. Анализ формальных понятий (АФП)
   Основная тема: содержание понятия, объем понятия, решетка понятий и ее диаграмма

Обзор курса

5. Импликации и функциональные зависимости
   Основная тема: точные зависимости в данных
6. Модели получения знаний из данных на основе ФАП
   Основные темы: Ассоциативные правила, машинное обучение, приложения в науках об обществе и науках о жизни

Бинарные отношения

Декартово (прямое) произведение множеств A и B: множество упорядоченных пар, первый элемент которых принадлежит A, а второй - B:

$$A \times B := \{(a, b) \mid a \in A, b \in B\}.$$

Бинарное отношение R из множества A в множество B - подмножество декартова произведения множеств A и B:
$R \subseteq A \times B$.
Инфиксная форма записи отношения R:
$aRb \Leftrightarrow (a, b) \in R \subseteq A \times B$.
Если $A = B$, то говорят, что R есть **отношение на множестве** A.
Тождественное отношение $I := \{(a, a) \mid a \in A)\}$.
Универсальное отношение $U := \{(a, b) \mid a \in A, b \in A)\}$.

Матричное представление отношений

Матрица бинарного отношения

Пусть $R \subseteq A \times A$. Отношение R представимо в виде матрицы

$$\begin{array}{c|ccc} R & \cdots & a_j & \cdots \\ \hline \vdots & & \vdots & \\ a_i & \cdots & \varepsilon_{ij} & \\ \vdots & & & \end{array}$$

$$\varepsilon_{ij} = \begin{cases} 1, & \text{если } (a_i, a_j) \in R \\ 0, & \text{если } (a_i, a_j) \notin R \end{cases}$$

Производные отношения $R \subseteq A \times A$

обратное отношение
$R^d := \{(a, b) \mid (b, a) \in R\}$

дополнение отношения
$R^c = \overline{R} := \{(a, b) \mid (a, b) \notin R\}$

отношение несравнимости
$I_R = A \times A \setminus (R \cup R^d) = (R \cup R^d)^c = R^c \cap R^{cd}$

произведение отношений
$P \cdot R = \{(x, y) \mid \exists z \quad (x, z) \in P, (z, y) \in R\}$

степень отношения
$R^n = \underbrace{R \cdot R \cdot \ldots \cdot R}_{n \text{ раз}}$

транзитивное замыкание отношения

$$R^T = R \cup R^2 \cup R^3 \cup \ldots = \bigcup_{i=1}^{\infty} R^i$$

Функции

Отношение $f \subseteq A \times B$ есть **функция** из A в B (обозначается $f : A \to B$) если для каждого $a \in A$ найдется $b \in B$, такое что $(a, b) \in f$ и

$$(a, b) \in f, (a, c) \in f \Rightarrow b = c$$

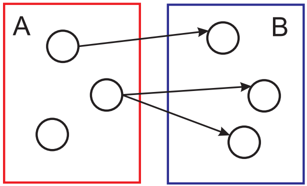
Функция $f : A \to B$ называется

инъекцией (отображением в), если $b = f(a_1)$ и $b = f(a_2) \Rightarrow a_1 = a_2$;

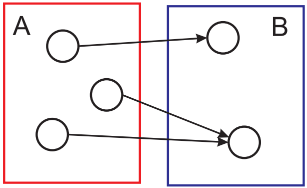
сюръекцией (отображением на), если для любого b из B существует a из A, такой что $b = f(a)$ (или $\forall b \in B \ \exists a \in A \ b = f(a)$);

биекцией, если она сюръекция и инъекция.

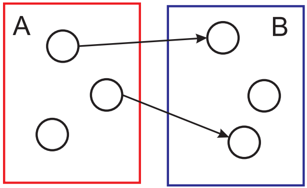
Свойства бинарных отношений



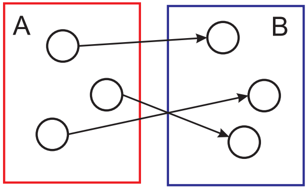
отношение, но не функция



сюръекция, но не инъекция



инъекция, но не сюръекция



биекция

Свойства бинарных отношений.

Пусть $R \subset A \times A$. Тогда R называется

рефлексивным, если $\forall a \in A\ aRa$
антирефлексивным, если $\forall a \in A \quad \neg(aRa)\ (\Leftrightarrow aR^c a)$
симметричным, если $\forall a, b \in A \quad aRb \Rightarrow bRa$
асимметричным, если $\forall a, b \in A \quad aRb \Rightarrow \neg(bRa)\ (\Leftrightarrow bR^c a)$
антисимметричным, если $\forall a, b \in A \quad aRb\ \&\ bRa \Rightarrow a = b$
транзитивным, если $\forall a, b, c \in A \quad aRb\ \&\ bRc \Rightarrow aRc$
полным, или линейным, если $\forall a, b \in A \quad a \neq b \Rightarrow aRb \vee bRa$.

Виды бинарных отношений

- **Толерантность** - рефлексивное и симметричное бинарное отношение;
- **Эквивалентность** - рефлексивное, симметричное и транзитивное бинарное отношение;
- **Квазипорядок** или **предпорядок** - рефлексивное и транзитивное бинарное отношение;
- **Частичный порядок** - рефлексивное, транзитивное и антисимметричное бинарное отношение;
- **Строгий порядок** - антирефлексивное и транзитивное бинарное отношение.

Представление отношений графами

Ориентированный граф (орграф) G - это пара вида (V, A), где V называется множеством **вершин** графа, а $A \subseteq V \times V$ называется множеством **дуг** (ориентированных ребер) графа G. Ориентированные графы с множеством вершин V представляют отношения на множестве V.

Представление отношений графами

Неориентированный граф - это пара вида $G = (V, E)$.
Множество V называется множеством **вершин** графа.
Множество $E = \{\{v, u\} \mid v, u \in V\} \cup E_0$, где
$E = \{\{v, u\} \mid v, u \in V\}$ - множество неупорядоченных пар элементов множества V, называется множеством **ребер**, а
$E_0 \subseteq V$ - множеством **петель**. Если $E_0 = \emptyset$, то G называется **графом без петель**.
Неориентированные графы с множеством вершин V представляют симметричные отношения на множестве V, т.е.
$R \subseteq V \times V$: $(a, b) \in R \Leftrightarrow (b, a) \in R$.

Матрица смежности (вершин) графа

(Ориентированный) граф $G = (V, E)$ можно представить в виде матрицы

$$\begin{array}{c|ccc} & \cdots & v_j & \cdots \\ \hline \vdots & & \vdots & \\ v_i & \cdots & \varepsilon_{ij} & \\ \vdots & & & \end{array}$$

$$\varepsilon_{ij} = \begin{cases} 1, & \text{если } (v_i, v_j) \in E \\ 0, & \text{если } (v_i, v_j) \notin E \end{cases}$$

В неориентированном графе $\varepsilon_{ij} = \varepsilon_{ji}$

Матричное представление графов

Матрица инцидентности графа

(Ориентированный) граф $G = (V, E)$ можно представить в виде матрицы

$$
\begin{array}{c|ccc}
 & \cdots & e_j & \cdots \\
\hline
\vdots & & \vdots & \\
v_i & \cdots & \varepsilon_{ij} & \\
\vdots & & &
\end{array}
\qquad
\varepsilon_{ij} = \begin{cases} -1, \text{ если } \exists v_k \in V\colon e_j \\ \;\;1, \text{ если } \exists v_k \in V\colon e_j \\ \;\;0, \text{ если } v_i \notin e_j \end{cases}
$$

Для неориентированного графа

$$
\varepsilon_{ij} = \begin{cases} 1, & \text{если } v_i \in e_j \\ 0, & \text{если } v_i \notin e_j \end{cases}
$$

Вершина v_i инцидентна дуге (ребру) e_j если $\varepsilon_{ij} \neq 0$.

Представление отношений графами

Ориентированный двудольный граф - это пара вида (V, A), где $V = V_1 \cup V_2$ и $V_1 \cap V_2 = \emptyset$ и $A \subseteq V_1 \times V_2$, т.е. любая дуга из A соединяет вершину из V_1 с вершиной из V_2. Множества V_1 и V_2 называются **долями** графа.

Ориентированные двудольные графы на долях V_1, V_2 представляют бинарные отношения из V_1 в V_2.

Представление отношений графами

Неориентированный двудольный граф - это пара вида (V, E), где V - множество вершин, $V = V_1 \cup V_2$ $V_1 \cap V_2 = \emptyset$, а $E = \{\{v_1, v_2\} \mid v_1 \in V_1, v_2 \in V_2\}$ - множество неориентированных ребер, соединяющих вершины из V_1 с вершинами из V_2.

Неориентированные двудольные графы на долях V_1, V_2 представляют симметричные бинарные отношения из V_1 в V_2.

(Неориентированный) **полный граф** - (неориентированный) граф $G = (V, E)$, в котором каждая пара вершин связана ребром.
Полный граф с множеством вершин V представляет универсальное отношение на множестве V.

Классы толерантности

Пусть дано множество M и отношение толерантности на $M \times M$.

Класс толерантности - максимальное подмножество элементов M, все пары которых принадлежат отношению.

Графовая интерпретация - максимально полный подграф неориентированого графа с петлями.

Классы эквивалентности и разбиения

Класс эквивалентности - подмножество элементов множества M, эквивалентных некторому элементу $x \in M$.

Графовая интерпретация - связанная компонента графа.

Разбиением множества M называется семейство множеств $\{M_1, \dots, M_n\}$, такое что

$$\bigcup_{i \in [1,n]} M_i = M, \quad \forall i, j \in [1, n] M_i \cap M_j = \emptyset.$$

Между разбиениями и эквивалентностями на множестве M существует взаимнооднозначное соответствие.

Части (неориентированных) графов

Граф $H = (V_H, E_H)$ есть **часть** (подграф) графа $G = (V_G, E_G)$ если и все вершины и ребра H являются, соответственно вершинами и ребрами G, т.е. $V_H \subseteq V_G$ и $E_H \subseteq E_G$.

Граф $H = (V_H, E_H)$ есть **(индуцированный) подграф** графа $G = (V_G, E_G)$ если H есть часть (подграф) G, а ребрами H являются все ребра G, обе концевые вершины которых лежат в H.

Пути и связность в неориентированных графах

маршрут (путь) - чередующаяся последовательность вершин и ребер графа вида $v_0, e_1, v_1, e_2, v_2, \ldots, e_k, v_k$, в которой каждые два соседних ребра имеют общую вершину.
цепь - маршрут, в котором все ребра различны.
простая цепь - маршрут в котором все вершины (а, следовательно, и ребра) различны.
Две вершины называются **связаными** если существует соединяющая их (простая) цепь.
Граф **связан** если все пары вершин связаны.
связная компонента графа - максимальное (по вложению) множество вершин графа, каждая пара которых связана.
Связаная компонента графа соответствует классу эквивалентности на множестве вершин по отношению "быть связаным".

Пути и связность в ориентированных графах

(ориентированный) маршрут или **(ориентированный) путь** - чередующаяся последовательность вершин и дуг графа вида $v_0, a_1, v_1, a_2, v_2, \ldots, a_k, v_k$, в которой конец любой дуги (кроме, быть может, последней) совпадает с началом следующей дуги, то есть $a_i = (v_{i-1}, v_i)$.
цепь - маршрут, в котором все дуги различны.
простая цепь - маршрут в котором все вершины (а, следовательно, и ребра) различны.
Вершина v_j **достижима** из вершины v_i если существует путь с началом в v_i и концом в v_j (считается, что вершина достижима из себя самой).
Граф **сильно связан** если любые две вершины достижимы друг из друга.
Граф **односторонне связан** если для любой пары вершин есть достижимость хотя бы в одну сторону.

Циклы

(ориентированный) цикл - (ориентированный) путь - путь, в котором первая и последняя вершины совпадают.
простой цикл - цикл, в котором любая вершина встречается не более одного раза.
неориентированный цикл (контур) в ориентрированном графе - последовательность вершин и дуг ориентированного графа, которая преобразовать в ориентированный цикл изменением ориентации некоторых дуг.
ациклический граф - ориентированный граф, не содержащий ориентированных циклов.

Ациклическими графами можно представлять квазипорядки и частичные порядки (при этом петли не изображаются), а также строгие порядки.

Деревья

(Неориентированное) дерево - неориентированный граф без циклов.
Ориентированное дерево - ориентированный граф, в которм нет циклов (как ориентированных так и неориентированных).
Корневое дерево дерево с выделенной вершиной, которая называется **корнем**.
В корневом дереве выделено направление от корня к листьям, поэтому корневое дерево можно считать ориентированным.
Звезда - дерево вида $G = (V, E)$, где $V = \{v_0\} \cup V_1$, $E = \{\{v_0, v_i\} \mid v_i \in V_1\}$.

Упражнения

- Определить какими свойствами обладает отношение $Q := \{(m, n) \mid m, n \in \mathbb{N} \;\&\; m = n^2\}$
- Найти число инъекций из конечного множества A в конечное множество B, считая что $|A| < |B|$.
- Найти число частей (неиндуцированных подграфов) конечного графа с множеством ребер E.
- Степенью вершины (неориентированного) графа называется число ребер, инцидентных данной вершине. Доказать, что в произвольном графе число вершин с нечетной степенью четно.

Литература по отношениям и графам

Ф.Т. Алескеров, Э.Л. Хабина, Д.А. Шварц, *Бинарные отношения, графы и коллективные решения*, М., ГУ-ВШЭ, 2006.

С.В. Судоплатов, Е.В. Овчинникова, *Дискретная математика*, Москва - Новосибирск, ИНФРА-М - НГТУ, 2007.

Г. Биркгоф, Т.К. Барти, *Современная прикладная алгебра*, М., Лань, 2005. (Главы 1,2)

А.А. Зыков, Основы теории графов, М., Наука, 1987.

О. Оре, *Теория графов*, М., Мир, 1965. (Главы 1,2,10)

Ф. Харари, *Теория графов*, М., Мир, 1973.