Zkušenosti ze stáže – Clermont-Ferrand

Lukáš Havrlant

DEPARTMENT OF COMPUTER SCIENCE
PALACKÝ UNIVERSITY, OLOMOUC

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Základní informace o stáži

- Datum konání stáže: 20.1.2014–1.3.2014
- Navštívené pracoviště: Université Blaise Pascal, ISIMA, Clermont-Ferrand, Francie
- Zahraniční garant: prof. Engelbert Mephu Nguifo

Informace o univerzitě

- ISIMA = Institut supérieur d'informatique, de modélisation et de leurs applications

- Laboratoř má přes 300 studentů a kolem 25 doktorandů.

- Zabývá se především rozhodovacími algoritmy a jejich modely.

- Laboratoř je silně propojena s komerční sférou.

Průběh stáže

- V laboratoři probíhala spousta přednášek, nicméně téměř všechny ve francouzštině.
- Setkal jsem se s prof. Abdoulaye Baniré Diallo, který tou dobou působil na Université du Québec à Montréal a navštívil jsem jeho přednášku *A probabilistic framework for gene tree reconstruction and reconciliation to a known species tree with an explicit macroevolutionary model*.
- Během stáže jsem přednesl dvě přednášky, *Search engine based on FCA* a *Bit-vector encoding and matrix decomposition*
- O druhé přednášce jsme dále diskutovali s profesorem Mephu a ostatními doktorandy.

Součin dvou binárních matic

Definition (Součin dvou binárních matic)

Mějme $n \times k$ matici A a $k \times m$ matici B, součin $A \circ B$ je definován jako

$$(A \circ B)_{i,j} = \bigvee_{l=1}^{k} A_{il} \cdot B_{lj}$$

Example

$$\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 \cdot 1 \vee 1 \cdot 1 & 0 \cdot 1 \vee 1 \cdot 0 \\ 1 \cdot 1 \vee 1 \cdot 1 & 1 \cdot 1 \vee 1 \cdot 0 \end{pmatrix} =$$

$$= \begin{pmatrix} \max(0 \cdot 1, 1 \cdot 1) & \max(0 \cdot 1, 1 \cdot 0) \\ \max(1 \cdot 1, 1 \cdot 1) & \max(1 \cdot 1, 1 \cdot 0) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Dekompozice matic

- Obecně: dekompozice matic je rozklad matice na součin více maticí.

- Má smysl omezit všechny matice na binární matice, tj. matice, které obsahují pouze nuly a jedničky.

- Popis problému: máme vstupní $n \times m$ binární matici I a chceme nalézt dvě binární $n \times k$ a $k \times m$ matice A a B tak, aby platilo $I = A \circ B$ a k chceme co možná nejmenší.

- Číslo k se pak nazývá scheinův rank matice.

Pokrytí relace I

Nechť $\mathcal{R}$ je množina konceptů v kontextu $\langle X, Y, I \rangle$. Zřejmě:

$$\bigcup \mathcal{R} = I.$$

Problém: nalezení minimální množiny $\mathcal{F} \subseteq \mathcal{R}$ takové, že $\bigcup \mathcal{F} = I$.

$$I = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \qquad I_1 = \begin{pmatrix} \mathbf{1} & \mathbf{1} & 0 \\ \mathbf{1} & \mathbf{1} & 0 \\ 0 & 1 & 1 \end{pmatrix}, I_2 = \begin{pmatrix} 1 & \mathbf{1} & 0 \\ 1 & \mathbf{1} & 0 \\ 0 & \mathbf{1} & 1 \end{pmatrix}, I_3 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & \mathbf{1} & \mathbf{1} \end{pmatrix}.$$

Platí, že: $I_1 \cup I_2 \cup I_3 = I$, ale také $I_1 \cup I_3 = I$. Množina $\{I_1, I_3\}$ je minimální množina, která pokrývá I.
Pokud nalezneme pokrytí matice I, jsme schopni nalézt také rozklad matice I.

Bit-vector kódování uspořádané množiny

- Libovolnou uspořádanou množiny jsme schopni reprezentovat pomocí bit-vector kódování.
- Každý prvek můžeme reprezentovat jako bit-vector, jako podmnožinu nějakého universa. Přitom jsme schopni nalézt takové kódování c, aby pro prvky $x, y \in P$ z uspořádané množiny P platilo: $x \leq y \Leftrightarrow c(x) \subseteq c(y)$.
- Dotaz $x \leq_P y$ pak snadno vypočteme jako bitovou operaci $x \& y =^? x$.
- Nejmenší počet bitů, který potřebujeme na zakódování uspořádané množiny P je rovný 2-dimenzi množiny P.

Dimeze konceptuálního svazu

- Uspořádanou množinu reprezentovanou maticí můžeme vidět jako formální kontext.
- Platí přitom, že
$$\dim_2 \mathbf{P} = \dim_2 \langle \mathcal{B}(P, P, \leq), \leq_C \rangle$$
- Důsledek: můžeme spočítat $\dim_2 \langle \mathcal{B}(P, P, \leq), \leq_C \rangle$ namísto $\dim_2 \mathbf{P}$.

Ferrerova dimenze

Definition (Ferrerova 2-dimenze)

Nechť $L = \langle \mathcal{B}(X, Y, I), \leq_C \rangle$ je konceptuální svaz. Nechť $\mathcal{R}$ je minimální množina konceptů, která pokrývá relaci $X \times Y \setminus I$. Pak Ferrerova 2-dimenze svazu L je $|\mathcal{R}|$. Označíme ji $\operatorname{fdim}_2 C$.

Theorem

Pro uspořádanou množinu $\mathbf{P} = \langle P, \leq \rangle$*:*

$$\dim_2 \mathbf{P} = \operatorname{fdim}_2 \mathcal{B}(P, P, \leq).$$

Důsledky

1. Velikost minimální množiny, kterou jsme schopni zakódovat uspořádanou množinu P je rovná $\dim_2 P$.
2. $\dim_2 P = \operatorname{fdim}_2 \mathcal{B}(P, P, \leq)$.
3. $\operatorname{fdim} \mathcal{B}(P, P, \leq)$ je rovna velikosti minimální množiny konceptů $\mathcal{R}$, která pokrývá komplementární relaci $>$.
4. Množinu $\mathcal{R}$ můžeme použít pro dekompozici matice $>$.

Závěr: Velikost minimální množiny pro zakódování uspořádané množiny P je rovna Scheinovu ranku matice $>$.