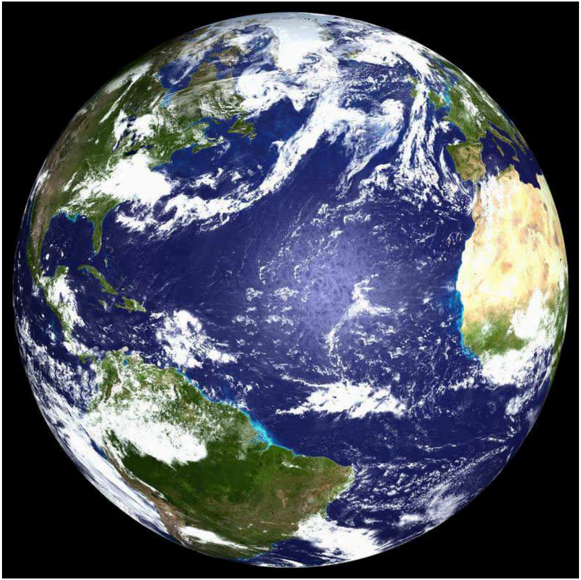
Struktury nezávislosti v pravděpodobnostních modelech

Václav Kratochvíl

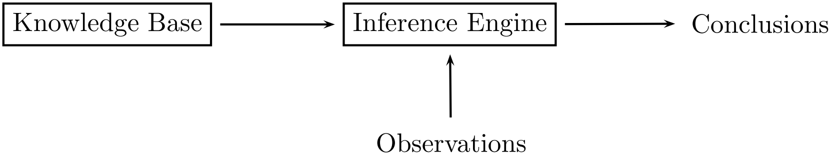
ÚTIA AVČR

11.04. 2013

Motivace



Motivace


Knowledge Base
Inference Engine
Conclusions
Observations

Pravděpodobnostní modely

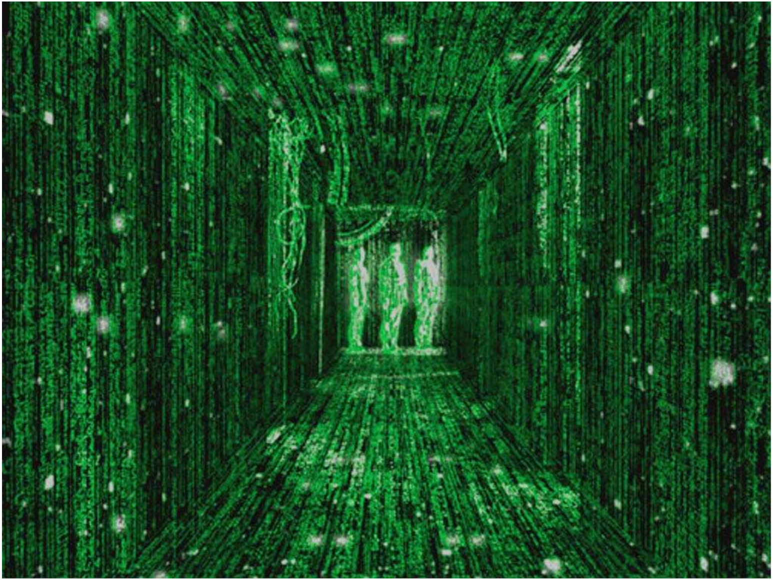
Nejistota

- pravděpodobnostní distribuce (alternativy: possibility theory, theory of evidence, fuzzy přístupy, ...

Multidimenzionalita

- koncept podmíněné nezávislosti

Motivace



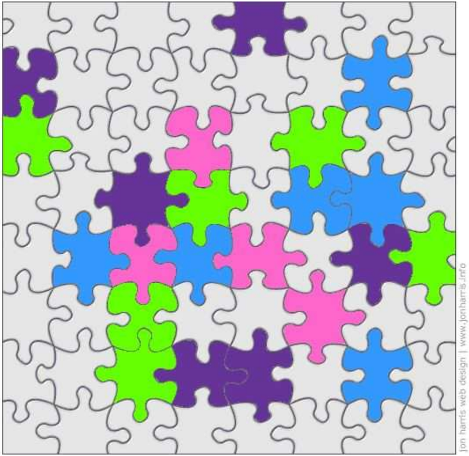
Motivace

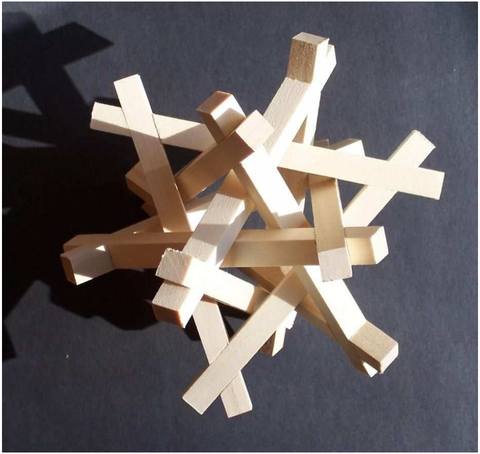
Table: Co jde

	X_1=.05			X_1=.20			X_1=.40		
X_3	X_2=0	X_2=0.5	X_2=1	X_2=0	X_2=0.5	X_2=1	X_2=0	X_2=0.5	X_2=1
0	0	0.0291	0.0291	0.0735	0.0735	0.0734	0.0185	0.0113	0.0080
1	0.0221	0	0.0221	0.0571	0.1571	0.0570	0.0058	0.0023	0.0807
2	0.0973	0.0073	0.0973	0.0080	0.0079	0.0807	0.0447	0.0541	0.5403

Multidimensionalita

- diskrétní pravděpodobnostní distribuce - kontingenční tabulka
- 100 bin. proměnných $\Rightarrow \quad 2^{100} - 1$ čísel
- faktorizace - rozložit mnohodimenzionální distribuci do málodimensionálních faktorů (komponent)


jon harris web design | www.jonharris.info



Podmíněná nezávislost

Theorem

Let $U, V, Z \subset N$ be such that $U \neq \emptyset \neq V$. Then for any probability distribution $\pi(UVZ)$;

$$U \perp\!\!\!\perp V | Z[\pi]$$

if and only if there exists functions

$$\psi_1 : \times_{i \in UZ} \mathbf{X}_i \to [0, +\infty); \psi_2 : \times_{i \in VZ} \mathbf{X}_i \to [0, +\infty)$$

such that $\forall x_{UVZ} \in \times_{i \in UVZ}\ \pi(x_{UVZ}) = \psi_1(x_{UZ})\psi_2(x_{VZ})$.

10^3 versus $2 * 10^2$

Example

Známka z testu je nezávislá na délce přípravy podmíněně počtem dosažených bodů.

Idea pravděpodobnostních modelů

- uložit systém podmíněných nezávislostí - jak?
- uložit odpovídající faktory
- problém: elementární operace jako projekce a podmiňování jsou komplikované pokud je distribuce reprezentované faktory

Semi-grafoid

symmetry	$U \perp\!\!\!\perp V \mid Z \Leftrightarrow V \perp\!\!\!\perp U \mid Z$
decomposition	$U \perp\!\!\!\perp VW \mid Z \Rightarrow U \perp\!\!\!\perp V \mid Z \ \& \ U \perp\!\!\!\perp W \mid Z$
weak union	$U \perp\!\!\!\perp VW \mid Z \Rightarrow U \perp\!\!\!\perp W \mid VZ$
contradiction	$U \perp\!\!\!\perp W \mid VZ \ \& \ U \perp\!\!\!\perp V \mid Z \Rightarrow U \perp\!\!\!\perp VW \mid Z$

Neexistuje polynomiální reprezentace takovýchto struktur

Grafoid

symmetry	$U \perp\!\!\!\perp V \mid Z \Leftrightarrow V \perp\!\!\!\perp U \mid Z$
decomposition	$U \perp\!\!\!\perp VW \mid Z \Rightarrow U \perp\!\!\!\perp V \mid Z \;\&\; U \perp\!\!\!\perp W \mid Z$
weak union	$U \perp\!\!\!\perp VW \mid Z \Rightarrow U \perp\!\!\!\perp W \mid VZ$
contradiction	$U \perp\!\!\!\perp W \mid VZ \;\&\; U \perp\!\!\!\perp V \mid Z \Rightarrow U \perp\!\!\!\perp VW \mid Z$
intersection	$U \perp\!\!\!\perp W \mid VZ \;\&\; U \perp\!\!\!\perp V \mid WZ \Rightarrow U \perp\!\!\!\perp VW \mid Z$

- lze reprezentovat např. grafem
- některé nezávislostní struktury nezachytí
- zachytí alespoň podmnožinu nezávislostí

Reprezentace nezávislostních struktur

Grafy

- neorientované grafy
- acyklické orientované grafy (DAG)
- řetězcové grafy

Algebraické struktury (kompozicionální modely)

Čtení nezávislostí z grafu

Separační kritéria

- neorientovaný graf - klasické grafové separační kritérium
- DAG: D-separační kritérium, Moralizační kritérium

D-separační kritérium

$u \rightarrow z \leftarrow v$: z je kolizní uzel.
$u \rightarrow z \rightarrow v$: z není kolizní uzel.
$u \leftarrow z \leftarrow v$: z není kolizní uzel.
$u \leftarrow z \rightarrow v$: z není kolizní uzel.

Cesta z U do V je blokovaná pokud $\exists$ nekolizní uzel ze Z, nebo $\exists$ uzel který není ze Z a ani nemá potomka ze Z.

U a V jsou separovány Z pokud je každá cesta z U do V blokovaná Z.

Strukturální ekvivalence

Čtení složité, proto má velký význam jak poznat zda grafy reprezentují stejné nezávislosti. Charakteristické vlastnosti, unikátní reprezentant třídy ekvivalentních grafů.

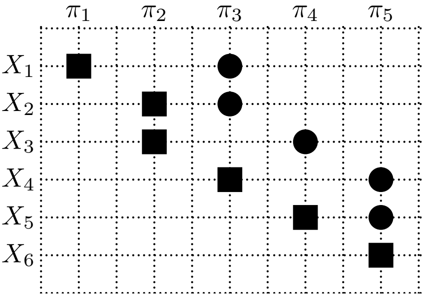
- neorientované grafy - musí být stejné
- DAG - různé grafy, stejný systém nezávislostí
- řešení - stejné hrany, stejné imorality
- esenciální graf

Aritmetické reprezentace

Kompozicionální model - $\pi_1(U_1) \triangleright \pi_2(U_2) \triangleright \ldots \triangleright \pi_n(U_n)$

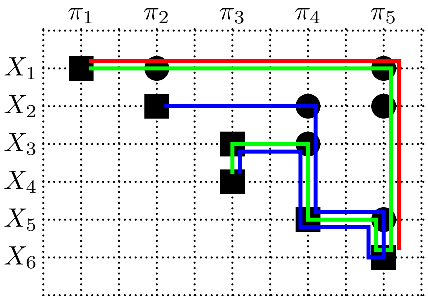
- skládání různých i nekonzistentních informací (pravd. distribucí)
- nepodmíněné nezávislosti vložené na začátku zachovává, později je může učinit podmíněně závislé.
- struktura $(U_1, U_2, \ldots, U_n)$
- ekvivalentní DAG

Persegram



- zobrazení alg. struktury
- separační kritéria
- problém ekvivalence

Separační kritéria - persegram



- červená $X_1 \perp\!\!\!\perp X_3 X_4 X_5 | X_2$
- zelená $X_1 \not\!\perp\!\!\!\perp X_3 X_4 X_5 | X_2 X_6$
- modrá $X_4 X_3 \not\!\perp\!\!\!\perp X_2 | X_6$

Persegram

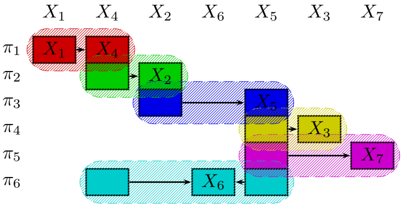


Figure: Side projection

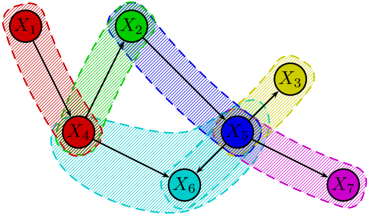
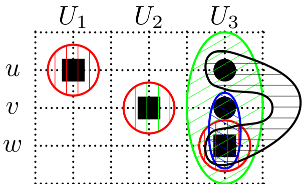
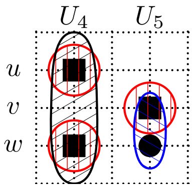


Figure: Horizontal projection

Netriviální množiny

cardinality	1	2	3
$\mathcal{N}(\mathcal{P}_1)$	u,v,w	{u,w},{v,w}	{u,v,w}
$\mathcal{N}(\mathcal{P}_2)$	u,v,w	{u,w},{v,w}	-





Imsety

- $\{0,1\}$-vektor velké dimenze (potenční množina)
- odpovídá netriviálním množinám
- učení odpovídá maximalizaci kritéria kvality $Q(G,D)$, G-graf, D-databáze
- Q - rozložitelné, skore-ekvivalentní
- každé skore-ekvivalentní kritérium je affinní funkce těchto vektorů
- $Q(G,D) = s_D^Q - \langle t_D^Q, u_G \rangle$