

Erötetická epistemická logika pro DZD

Michal Kováč

Motivace

- Chybí obecná logická teorie pro DZD
- Analytické otázky často neformální a vágní
- Hodila by se flexibilní báze znalostí a otázek
- Ever-miner



Obsah přednášky

- Epistemická logika
- Logika otázek
- Proces DZD a logika
- Možné využití do budoucna
- Kam směřuji já

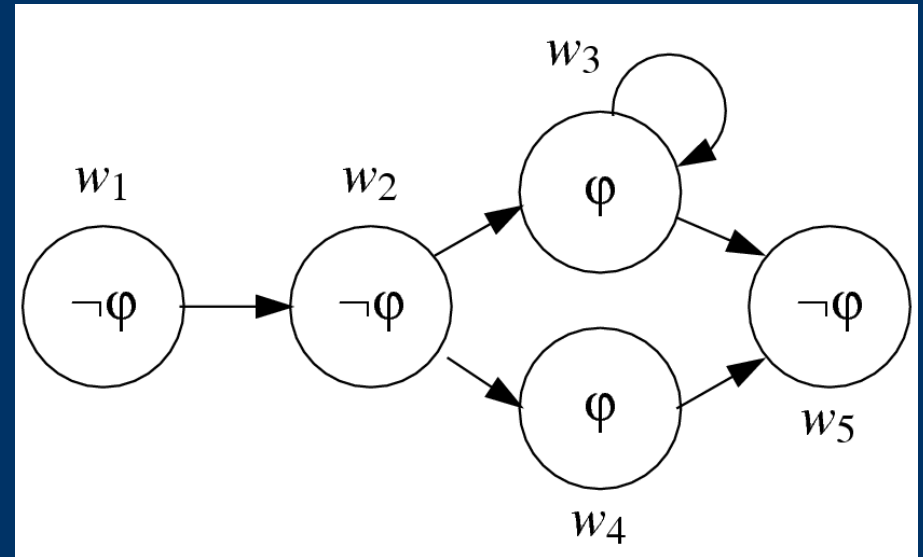


Epistemická logika

- Epistemická logika S5 s více agenty
 - $q ::= a \mid \neg p \mid \dots \mid K_i p$
 - $M_i p$ lze definovat jako $\neg K_i \neg p$
 - Skupinová znalost: E_G (všichni v G ví) a C_G (všichni v G ví a ví o této skupinové znalosti)

Kripkeho rámce a model

- Možné světy
- Relace R_i pro každé K_i
- Uvnitř standardní logika
- $K_i p$ – ve všech R_i následujících platí p
- U S5 je R_i ekvivalence



Axiomatika

- Výroková logika +
- $K_i(p \rightarrow q) \rightarrow (K_i p \rightarrow K_i q)$
- $K_i p \rightarrow p$
- $M_i p \rightarrow K_i M_i p$

Sémantika

- Kripkeho ráamec: $F = \langle S, R_1, \dots, R_n \rangle$
- Kripkeho model: $\langle F, v \rangle$
- Běžné spojky klasicky ve světě
- $(M, s) \models K_i p$ iff pro každé r , že $sR_i r$ platí $(M, r) \models p$
- R_i jsou ekvivalence

Znalosti o datech

- Epistemické rozšíření observačního kalkulu
 - $K_{\text{Honza}}(\text{Tlak}(\text{vysoký}) \& \text{BMI}(\text{nizký})) \Rightarrow_{0,8} \text{Riziko}(A)$
 - Navíc vhodné rozlišovat báze:
 $K_{\text{Honza}}(\text{Tlak}(\text{vys.}) \& \text{BMI}(\text{nizký})) \Rightarrow_{0,8(\text{Praha})} \text{Riziko}(A)$
 - Bylo by možné fuzzy a další modality
 - Dolování nad znalostmi:
 $K_{\text{Honza}}(\text{Riziko}(A)) \& K_{\text{Pepa}}(\text{Riziko}(A))$
 $\Rightarrow_{0,8} K_{\text{Oldřich}}(\text{Riziko}(A))$
-
-

Logika otázek

- Hamblinova tvrzení:
 - Odpověď na otázku je výrok.
 - Znat význam otázky znamená vědět, co je považováno za odpověď na ni.
 - Možné odpovědi na otázku tvoří vyčerpávající množinu vzájemně se vylučujících možností.
 - Otázka je formule ve tvaru: $?_i(a_1, \dots, a_n)$
 - a_1, \dots, a_n jsou syntakticky nezávislé formule
 - $n \leq 2$
-
-

Vyplývání otázek

- Wisniewského erotetický důsledek (podproblém):
 - 1) Každá odpověď na novou otázku implikuje nějakou alespoň částečnou odpověď na původní otázku.
 - 2) Přenášení korektnosti: Jestliže má původní otázka pravdivou odpověď, musí mít nějakou pravdivou odpověď i otázka nová.
 - Z $?(A \& B, \neg(A \& B))$ nevyplývá $?(A, \neg A)$
 - Z $?(A \& B, \neg(A \& B))$ vyplývá ?
 $(A \& B, \neg A \& B, A \& \neg B, \neg A \& \neg B)$
 - Z $?(A \& B, \neg A \& B, A \& \neg B, \neg A \& \neg B)$ vyplývá ?
 $(A, \neg A)$
-
-

Vyplývání otázek 2

- Pomocí vyplývání otázek lze
 - Rozložit otázku na dílčí podotázky
 - Vytvořit dotazovací scénář složený pouze ze zjišťovacích otázek

Dotazitelnost

- Netriviálnost: „Není rozumné se ptát na otázku, když odpověď je již známá.“
- Přípusnost: „Všechny přímé odpovědi musí být možné“
- Souvislost: „Alespoň jedna přímá odpověď musí být správná.“

Dotazateľnosť agentů

- $(\forall p \in dQ^i)((M, s) \not\models K_i p)$
- $(\forall p \in dQ^i)((M, s) \models M_i p)$
- $(M, s) \models K_i (\bigvee_{p \in dQ_i} p)$

Proces dobývání znalostí

- Znalosti
- Analytická otázka
- Zadání úlohy
- Řešení úlohy
- Odpověď na analytickou otázku
- Analytická zpráva
- Rozšíření znalostí



Analytické otázky

- Použití logiky otázek nad observačním kalkul.
 - Problém: veliké množství možných odpovědí
 - Řešení: možnost zapsat pomocí schémat
 - Příklad: „Jak příznaky podstatně zvyšují pravděpodobnost choroby?“
 - $?(P(\&(\text{příznaky}) \rightarrow \text{nemoc}))$
 - Nicméně: Analytika nezajímají věci, co již zná, tedy: $?(P(\&(\text{příznaky}) \rightarrow \text{nemoc}) - \text{známé})$
 - (Jiné otázky mohou být dál od zadání úlohy)
-
-

Problémy analytické otázky

- Vágnost jazyka
- Co člověk to jiné chápání otázky
- Při otázce ne vždy si dokážeme představit všechny možné odpovědi
- Příklad:
 - Jaké platí zajímavé výroky o pacientech v Čáslavi.
 - Jak rozpoznat včas možnost výskytu nemoci XY.

Zadání úlohy

- Pro zodpovězení analytické otázky je potřeba často více úloh a naopak samotné úlohy často odpovídají více, než je třeba.
 - GUHA zadání – jazyk Ferda – upravený lambda kalkulus, blízko k intenzionální logice
 - Mezistupeň: Otázka z logiky otázek nad observačním kalkulem, kde není požadavkem netriviálnost ani přípustnost z pohledu analytika. Otázka jako množina možných výsledků Ferda úlohy.
-
-

Analytická otázka vs. zadání

- Zadání využívá pouze možné programové prostředky
- Formálnost – např. kategorizace, mapování atributů
- Je třeba rozdělit otázku na podotázky
- Vztah n ku n
- GUHA dnes
 - rozdělena na procedury
 - neumožňuje obecné množiny

Řešení úlohy

- Možno vidět jako odpověď
- Redundance: Obsahuje známé věci, věci, na které se neptalo, a opakující se znalosti
- Chybějící znalosti: často se zjistí nevhodnost zadání až na základě řešení
 - Částečně řešitelné lepší zadavatelností

Odpověď na analytickou otázku

- Formule observačního predikátového kalkulu
 - Je třeba získat filtrováním ve znalostech získaných z úloh a vyplýváním
 - Problém: pravdivost odpovědi: je třeba důkaz
 - Důkaz pravdivosti: řešení úloh
-
-

Podmínky pravdivost odpovědi analytické otázky

- Odpověď je z množiny možných přímých odpovědí otázky
- Mohly vyjít všechny možné odpovědi
- Nebylo nevhodně vyfiltrováno
- Využilo se jen předpokladů tazajícího se
- Byl udělán vhodný výběr nejlepší odpovědi



Poloodpovědi

- Částečná odpověď, zúžení množiny odpovědí
- Lze se dívat jako na odpověď s předpoklady



Analytická zpráva

- Důkaz pravdivosti odpovědi na analytickou otázku
 - Pětice $\langle A, Z, T, P, O \rangle$, která obsahuje
 - A: Analytickou otázku
 - Z: Využité znalosti a předpoklady
 - T: Množina trojic $\langle U, Z_t, R \rangle$
 - P: Množina podzpráv $\langle A_p, Z_p, T_p, P_p, O_p \rangle$
 - O: Odpověď na otázku
-
-

Správnost zprávy

- Odpověď musí být přípustná
 - Sjednocení možných řešení úloh a podzpráv musí umožňovat všechny možné odpovědi na otázku
 - R musí obsahovat tvrzení, které platí z U , a to všechny podstatné pro řešení odpovědi se znalostmi Z_t
 - Z_t a Z_p jsou podmnožiny Z
 - $A + Z +$ množina R a O_p musí implikovat O
-
-

Lokální a globální zprávy

- Lokální: nejsou podzprávy
- Globální: nejsou úlohy jsou jen podzprávy



Rozšiřování znalostí

- Potřeba zachytit znalosti
- Prohlížet a rozšiřovat
- Využívat k dobývání a filtrování



Analytická otázka a znalost

- Znalosti vedou k otázkám
 - Lze generovat zajímavé otázky za znalostí?
- Otázky obsahují znalost



Další

- Vyplývání mezi zprávami
- Vyplývání mezi znalostmi
- Vyplývání otázek za znalostí



Teorie v praxi 1

- Mějme bázi znalostí a otázek o lékařství
 - Lékař Pepa by rád věděl jak míry ovlivňují tlak
 - Lékař Pepa do báze vloží tedy otázku:
 - $K_{\text{Pepa}}(? \text{ (&(míry) } \rightarrow \text{tlak}))$
 - V bázi jsou již následující znalosti (ukáže se):
 - K_{All} (hmotnost neovlivňuje systolický tlak)
 - K_{Pepa} (obvod pasu veliký $\Rightarrow_{0,7}$ systolický k.t. vysoký)
 - M_{Pepa} (obvod pasu veliký $\Rightarrow_{0,7}$ diastolický k.t. vysoký)
 - K_{Jarda} (obvod pasu veliký $\Rightarrow_{0,6}$ diastolický k.t. vysoký)
-
-

Teorie v praxi 2

- Systém může vygenerovat různé úlohy
 - 4FT
 - Konjunkce míry různé kvantifikátory jeden z tlaků
 - KL
 - Opět různé kvantifikátory
 - Na generování zadání se dá dívat jako na formu nemonotónní dedukce
 - Znalosti navíc můžou jak zvětšovat tak zmenšovat velikost úlohy
 - K_{All} (hmotnost neovlivňuje systolický tlak) zmenšuje
-
-

Teorie v praxi 3

- Výsledky nejen odpovídají na otázku, ale odpovídají na domněnku. Domněnku je možné vidět jako otázku na míru fuzzy pravdivosti.
 - M_{Pepa} (obvod pasu veliký $\Rightarrow_{0,7}$ diastolický k.t. vysoký)
 - obvod pasu veliký $\Rightarrow_{0,6}$ diastolický k.t. Vysoký
- Výsledky jsou nové obecné znalosti
- Je třeba pamatovat i na data nad kterými je dolováno
 - Problém průběžně rostoucích dat

Chybějící části systému

- Obecná rychlá báze znalostí
- Obecný zadávač do báze
- Prohlížeč báze
- Dedukční a souvislostní systém nad bází
- Dedukční systém na zadání DZD
- Import výsledků do báze
- Generátor pro návrh zajímavých otázek



Sewebar & Joomla stav

- Jsou řešeny podobné problémy
- Báze není obecná a různá úložiště
- Zadávač a prohlížeč není obecný
- Dedukce u zrodu
- Import funkční



Cíle do budoucna

- 1) Jazyk pro zápis analytických otázek a znalostí
 - Problém velikost formulí
 - Pokud možno logika
 - Pokud možno obecné
 - 2) Teorie kolem logiky otázek vhodná pro GUHA praxi
 - Tvorba zadání z otázek
 - Otázky ze znalostí a otázek
 - 3) Jednoduchá implementace kolem Ferdy
 - Báze znalostí
 - Příklady generování zadání a pod.
-
-