Poznatky a usudzovanie

Ján Šefránek
Ústav informatiky MFF UK

Revidované: 14.12.2000

Abstract

Medzi centrálne kognitívne schopnosti človeka patrí spracovanie poznatkov, usudzovanie. Prvá prednáška pripraví východiskové predstavy a základné pojmy.

Pre inteligentné usudzovanie je charakteristická schopnosť tvoriť užitočné hypotézy a revidovať ich, ak je to potrebné. Druhá prednáška sa bude venovať práve hypotetickému usudzovaniu a revíziám.

Text bude ešte revidovaný.

1 Úvod

Tento text tvorí základ dvoch prednášok v rámci celouniverzitného kurzu o Kognitívnej vede a v rámci projektu dištančného vzdelávania v tejto oblasti. Cieľom celého kurzu je priblížiť poslucháčom základné tendencie a prístupy k skúmaniu kognície (myslenia, poznania, vnímania, ...).

Tento text pristupuje k téme z hľadiska umelej inteligencie (a logiky). Sústredíme sa na problematiku usudzovania (získavania nejakých poznatkov z iných poznatkov na základe istých "myšlienkových operácií"). Budeme sa zaujímať o to, ako náročné je vytvoriť tak dôkladný opis (rôznych typov) usudzovania, aby bolo možné usudzovanie zveriť počítaču a aj o to, k ako ťažkým výpočtom to vedie.

Najprv budem motivovať prístup, ciele a použitú optiku (hľadiská), časť 2. Potom (v časti 3) sa budeme venovať analýze dedukcie. Ukážeme si, že už veľmi jednoduchá predstava o dedukcii vedie k nepríjemne zložitým (nezvládnuteľným, prípadne až nemožným) výpočtom. Predstavíme si aj ďalšie dôvody, prečo dedukcia nestačí na modelovanie a pochopenie usudzovania. Po takejto príprave prejdeme k analýze niektorých zjednodušení dedukcie, umožňujúcich zvládnuteľný výpočet (časť 4) a k analýze hypotetického usudzovania (5). Na záver sa budeme pýtať (pokúsime sa zhrnúť), čo nám o kognícii môže povedať tu použitý prístup.

Dúfam, že text nepredpokladá žiadne vedomosti (presahujúce stredoškolské vzdelanie).

Ospravedlňujem sa za rozdeľovanie slov na konci riadku v tomto texte. Toto rozdeľovanie robil automat bez dostatočného vybavenia. V poslednej verzii textu sa budem venovať aj rozdeľovaniu. Text bude ešte revidovaný.

2 Motivácia

Iba na rozohriatie začnime slovníkovou definíciou kognície (nejde o pojem, ktorý by bol u nás vžitý). Navyše, neexistuje – zdá sa – zodpovedajúci slovenský ekvivalent. V Oxfordskom slovníku angličtiny sa význam slova *cognition* vysvetľuje ako *knowing, perceiving, or conceiving as an act or faculty distinct from emotion and volition; result of this* [4].

Samozrejme, ide o veľmi širokú oblasť. Naša pozornosť bude zameraná iba na schopnosť usudzovať (na základe istých poznatkov odvodzovať ďalšie poznatky).

Môžeme sa oprieť o pozorovateľné javy. Človek disponuje nejakými poznatkami, niečo vie: toto je môj sused, tento študent ešte neurobil u mňa skúšku, tento chlap má na topánke popol. Dá sa aj pozorovať, že z toho, čo človek vie, dokáže odvodiť dôsledky (niekedy triviálne, niekedy zaujímavé, niekedy významné): z pozorovania, že chlap má na topánke popol, pričom vytrvalo prší, možno odvodiť, že niečo nie je v poriadku [1]. V centre našej pozornosti bude usudzovanie, poznatky nás budú zaujímať iba ako objekty, s ktorými sa pri usudzovaní pracuje.

Budeme abstrahovať od fyziologického, materiálneho podložia usudzovania (mozog, nervová sústava) a aj od evolúcie tejto schopnosti. Je to naprosto legitímny postoj – napríklad, ani pri štúdiu tatranských kamzíkov nás nemusí (ale môže) zaujímať atomárna úroveň, ani evolučné aspekty (hoci môžu).

Na čo zameriame pozornosť, sú

- kritériá, umožňujúce odlíšiť úspešné, zdravé usudzovanie, dobrú argumentáciu od chybného usudzovania,
- aby sme zvládli prvú úlohu, potrebujeme adekvátne odlíšiť rôzne druhy usudzovania,
- napokon, budeme zvedaví, ako detailne sa dá opísať (zdravé) usudzovanie; či takýto opis môže byť až tak detailný, aby slúžil ako východisko pre riadenie usudzovania (tak ako si to predstavoval Leibniz) alebo pre jeho automatizáciu (tak ako sa usiluje umelá inteligencia).

Cieľ, ktorý je ukrytý za uvedenou formuláciou fokusu: Pokúsime sa demonštrovať, že k pochopeniu kogníce prispieva aj symbolová umelá inteligencia, založená na logike. Sústredíme sa pritom na analýzu a rekonštrukciu usudzovania, najmä hypotetického.

Dôkladné pochopenie usudzovania je extrémne komplikovaný problém. Iba jeden pekný príklad – v už citovanej poviedke *Případ Dra Mejzlíka* [1] Čapek konštruuje autoreflexiu detektíva, ktorý nečakane a obdivuhodne rýchlo identifikoval lupiča. Detektíva trápilo, ako vlastne prišiel na svoj objav. Príbeh, ktorý Čapek skonštruoval, ukazuje, že to mohla byť náhoda, tušenie, inštinkt, podvedomý alarm, intuícia, telepatia, vyššie schopnosti, detektívny bystrozrak, zvyk, mechanická rutina, pozorovanie, logika, úsudok na základe skúsenosti, zistenie, že niečo nie je v poriadku, popol na topánkach muža, kráčajúceho vo vytrvalom daždi.

Ide skutočne o zamotaný komplex problémov, ktoré sú (alebo mohli by byť) predmetov viacerých vied, pokúšajúcich sa o pochopenie kogníce. My si budeme všímať jednoduchšiu stránku tohoto komplexu. Budeme si všímať to, čo možno vyjadriť vetami nejakého jazyka. Napríklad jednoduché pozorovania, ako popol na topánkach muža, kráčajúceho vo vytrvalom daždi. Budeme sa pýtať, čo takéto pozorovania môžu znamenať. Niekedy pôjde o zistenie, že niečo nie je v poriadku. Inokedy o potrebu doplniť alebo revidovať náš obraz situácie. Často je potrebné práve opísané úlohy (či procesy) dotiahnuť do tvaru veľmi detailnej argumentácie, kde každý krok je nesporný (alebo sú zrejmé miesta, kde argumentáciu možno modifikovať, ak sa dostaneme do rozporov).

Ešte raz zdôraznime, že situáciu si uľahčíme – budeme si všímať iba vety nejakého jazyka. Na týchto vetách (alebo množinách viet) budeme zavádzať nejaké operácie (zodpovedajúce odvodzovaniu).

Hľadisko, z ktorého k téme pristúpime, objasníme po analýze metafory *čínskej izby* [5].

2.1 Metafora čínskej izby

Predstavme si, že ktosi (napríklad Searle) je zamknutý v izbe. Jeho materinským jazykom je angličtina a vôbec neovláda čínštinu. Má prístup k trom sadám čínskych znakov: Druhá z nich podáva nejaký príbeh, tretia z nich kladie otázky, týkajúce sa tohoto príbehu. Prvá z nich obsahuje čínske písmená. Okrem toho má Searle k dispozícii množstvo pravidiel (sformulovaných v angličtine), ktoré určujú, *akými kombináciami čínskych písmen (z prvej sady) má odpovedať na aké otázky (na základe daného príbehu).*

Predstavme si aj, že takáto hra umožní odpovedať čínsky na čínske otázky tak dokonale, že vonkajší pozorovateľ mimo tejto izby nie je schopný odlíšiť práve opísané odpovedanie od odpovedí toho, kto ovláda čínštinu. [5]

2.1.1 Dôsledky podľa Searlea

Searle rozlišuje slabú a silnú umelú inteligenciu [5, 2]. Slabá pristupuje k výpočtovému modelovaniu ako k nástroju, využiteľnému pri štúdiu myslenia (mysle) a umožňujúcemu rigorózne testovanie hypotéz. Môže tým prispieť k objasneniu ľudského správania.

Naproti tomu, pozíciu silnej umelej inteligencie Searle charakterizuje ako presvedčenie o tom, že vhodne naprogramovaný počítač skutočne je mysľou, že takýto počítač môže rozumieť a zažívať ďalšie kognitívne stavy. Svoju argumentáciu zameral proti takémuto poňatiu umelej inteligencie.

Searle vidí markantný rozdiel medzi porozumením čínštine a angličtine v našom (jeho) príklade. Na rozdiel od angličtiny (ktrej Searle perfektne rozumie), odpovede v čínštine sú iba výsledkom manipulácií neinterpretovaných formálnych symbolov. Searle dôvodí, že takéto manipulácie neumožňujú pochopiť čínske vety.

Záver: Tak, ako človek z nášho príkladu, ani počítač nedokáže rozumieť, pochopiť, čo robí. Totižto aj počítač iba manipuluje neinterpretované formálne symboly.

"Filozofia", ktorá je za tým:

Pokiaľ niekto robí niečo mechanicky podľa nejakých pravidiel, pokiaľ vykonáva iba syntaktické operácie, nemožno povedať, že skutočne rozumie.

Počítač – na rozdiel od človeka – vykonáva iba formálne, syntaktické operácie, bez väzby na nejaké sémantické poznanie, na poznanie sveta a bez intencie dosiahnuť isté výsledky tým, že niečo robí.

Podľa Searlea pochopenie je vlastnosťou ľudského mozgu, ktorý je schopný vyvolať (zapríčiniť) mentálne procesy. Za istých okolností je schopný produkovať vnemy, akcie, pochopenie, učenie a iné intencionálne fenomény.

Zhrňme:

1. programy sú iba manipulácie neinterpretovaných symbolov,
2. syntax nepostačuje na vytvorenie sémantiky,
3. myseľ kauzálne závisí na (ľudskom) mozgu,
4. iba myseľ vytvára mentálne obsahy (sémantiku) a nutne ich vytvára.

Z toho:

- každý systém, ktorý je schopný vytvoriť myseľ, musí mať aspoň takú kauzálnu silu ako (ľudský) mozog,
- programy nestačia na vytvorenie sémantiky a mysle.

2.1.2 Iný pohľad na Searlovu metaforu

Mojím cieľom nie je kritika a vyvracanie Searlovej argumentácie, ale jej otočenie iným, domnievam sa, že produktívnejším, smerom. Kým prejdem k spomínanej konštrukcii, neodpustím si však niekoľko poznámok.

Do očú bije porovnanie filozofa a spisovateľa. Searle ponúka bezstarostné a jednoduché riešenie. Iba ľudský mozog produkuje myseľ (na základe definície).[1] Mentálne procesy sú nositeľom (konštruktérom)

[1] Searle sa pokúša o zdôvodnenie takéhoto tvrdenia. Žiaľ toto tvrdenie patrí do triedy nevyvrátiteľných, ak nedefinujeme myseľ inak ako kauzálny účinok ľudského mozgu.

sémantiky. U Searla dominuje naivná a neanalyzovaná dôvera v mentálne procesy; naivný predpoklad "porozumenia" vďaka mentálnym procesom. Čapek ponúka komplikovaný obraz, ukazuje množstvo nepochopených procesov a otvorených problémov. U Čapka dominuje skepticizmus a *hľadanie*.

Podľa Patricka Hayesa, jedného zo zakladateľov umelej inteligencie, jadro kognitívnej vedy (a teda úlohy pochopiť kogníciu) by sa dalo vyjadriť (zhrnúť) ako "starostlivé a detailné vysvetlenie, prečo je Searleov argument čínskej izby hlúpy, skutočne hlúpy" [3].

Nebudem sa o to pokúšať, iba natočím pozornosť na dva problémy (možno toto natočenie niečo povie aj o Searlovom argumente, aj o Hayesovej charakterizácii jadra kognitívnej vedy).

1. Pokúsim sa pochopiť, čo metafora čínskej izby znamená pre "slabú" umelú inteligenciu. I keď Searle svoju argumentáciu nezameriava proti slabej (opatrnej) umelej inteligencii, dôsledky budú zaujímavé. Budú relevantné aj pre analýzu mysle a porozumenia, prípadne produktívnych abstrakcií nad nimi.
2. Pokiaľ ide o ambície "silnej" umelej inteligencie a problematiku mysle, porozumenia a sémantiky, pokúsim sa ukázať, že je užitočné *rozlišovať mentálne procesy a významy*. Po takomto rozlíšení zavrhneme predstavu, že sémantiku možno budovať iba nad mentálnymi procesmi. (Mohli by sme sa pokúsiť aj o demonštráciu, že nad mentálnymi procesmi sémantiku vybudovať nemožno, ale to naším cieľom nebude.) Významy nám budú predstavovať produktívnu abstrakciu nad mysľou. Pokúsime sa zdôvodniť, že táto abstrakcia je užitočná pre pochopenie mysle. Domnievam sa, že takýmto spôsobom možno skonštruovať univerzálnejší (a zaujímavejší) pojem ako "mysel'".

I Jedným z výskumných cieľov, ktoré možno zaradiť do slabej umelej inteligencie, je bádanie v oblasti automatického prekladu, v oblasti komunikácie (s počítačom) v prirodzenom jazyku, získavania zhrnutí z obsiahlejších dokumentov (napísaných v prirodzenom jazyku) a množstvo ďalších úloh z výpočtovej lingvistiky. Súvislosť so Searlovou metaforou je zrejmá.

Je však možný tak detailný popis čínštiny, dokumentov napísaných v čínštine a tak detailný opis sveta (znalosti o ňom potrebujeme, aby sme mohli reprezentovať dokumenty, príbehy), že na jeho základe sa dá pomocou pravidiel odpovedať na otázky v čínštine tak, že tieto odpovede sú neodlíšiteľné od odpovedí človeka, ktorého materinskou rečou je čínština?

Predpokladám, že sa dá urobiť seriózna analýza (možno aj dôkaz), ktorá spochybňuje vytvorenie tak dôkladného opisu. Tu sa uspokojím iba s konštatovaním, že je to nesmierne ťažká úloha, nad hranicami súčasného poznania. Odporúčam čitateľovi vyskúšať, ako odpovedá na otázky v prirodzenom jazyku (v angličtine) systém START, vytvorený na špičkovom svetovom pracovisku, MIT:
`http://www.ai.mit.edu/projects/infolab/start-system.html`

Zhrňme: Tak detailný opis komplikovanejších reálnych oblastí a inteligentných správaní, ktorý by umožnil modelovanie pomocou pravidiel (alebo pomocou iných formálnych prostriedkov) je ťažká úloha, spravidla na hraniciach súčasného poznania (alebo nad nimi). Preto konštrukcia takýchto modelov prispieva k *pochopeniu* týchto oblastí. Úloha opísať jazyk tak dôkladne, aby program mohol zabezpečiť preklad z jedného prirodzeného jazyka do iného, vedie k presnejšiemu, podrobnejšiemu a aj hlbšiemu pochopeniu jazyka. K hlbšiemu, než bol ten, ktorým pred vznikom tejto paradigmy[2] ľudia disponovali. Iba podotýkam, že ich pochopenie bolo menej hlboké napriek tomu, že aj vtedy sa mohli pýšiť mozgom, mentálnymi procesmi a intencionalitou.

Je teda čas prejsť k analýze pochopenia, teda k druhému natočeniu našej optiky.

II Už slovníkové objasnenie slova kognícia [4] ukazuje, že ide o viacvrstenú oblasť (a pojem). Videli sme, že oxfordský slovník rozlišoval (na kognícii) činnosti (procesy), schopnosti (vykonávať tieto činnosti) a výsledky (týchto činností).

[2] Hovorím o paradigme: pochopiť znamená (podľa metafory čínskej izby) opísať do takého detailu, že zodpovedajúce procesy (modely) možno realizovať i počítačom (nám pôjde o modely usudzovania; spracovania znalostí).

My budeme rozlišovať aspoň tri úrovne:

- neurofyziologickú,
- mentálnu,
- obsahovú.

Čo hovorí metafora čínskej izby o kognícii? Sústreďme sa na pojem pochopenia (porozumenia). Naša analýza bude mať dôsledky pre sémantiku (a aj pre náš prístup k usudzovaniu). Je nesporné, že pochopenie (naša schopnosť chápať) je možné vďaka (nášmu) neurofyziologickému vybaveniu. Je nesporné aj to, že toto neurofyziologické vybavenie zapríčiňuje isté mentálne procesy a mentálne procesy sú aktivované aj vtedy, keď niečo chápeme (alebo sa usilujeme pochopiť).

Napriek tomu, pochopenie je vždy pochopením NIEČOHO.

Pozrime sa bližšie na toto niečo. Keď niečo opisujeme (usilujeme sa pochopiť), hovoríme o nejakých objektoch (materiálnych alebo ideálnych). Ibaže, na sémantickú analýzu nášho hovorenia (našich opisov) nevystačí, keď si budeme všímať iba tieto objekty (pokiaľ si nemôžeme dovoliť pre nejaké ciele zjednodušiť situáciu). Po Fregem nemožno (bez prijateľného alebo neprijateľného zjednodušenia) sémantiku budovať iba na objektoch, označených jazykovými výrazmi. Ten istý objekt možno označiť jazykovými výrazmi, ktorých zmysel je rôzny (autor Povídek z jedné kapsy vs. autor Hovorů s T.G. Masarykem). Existujú jazykové výrazy, ktoré neoznačujú nijaký objekt, ale majú zmysel, rozumieme im (miesiželezo; množina stabilných modelov logického programu $P = \{a_i \leftarrow \neg b_i, b_i \leftarrow \neg a_i : i = 1, \ldots, n\}$, ktorá je iba lineárne väčšia než je rozsah P).

Zmysel si môžeme voľne charakterizovať ako kritérium, ktoré rozhoduje o tom, či nejaký objekt je označený nejakým jazykovým výrazom. Toto kritérium môžeme modelovať nejakou funkciou. V tomto texte budeme pracovať s veľmi jednoduchou verziou takejto funkcie, budeme ju volať interpretáciou. Netvrdím, že takéto funkcie sú to najvhodnejšie, čím možno charakterizovať sémantiku prirodzeného jazyka.

Treba zdôrazniť, že zmysel (význam) nie je redukovateľný na mentálne procesy. Určite existuje množstvo významov, ktoré doposiaľ "neaktivoval" vo svojej mysli žiadny človek.

Zhrňme Keď hovoríme o pochopení, hovoríme predovšetkým o kritériu, umožňujúcom rozpoznávať, rozlišovať, o zmysle.

Vôbec sa nemôžeme brániť tomu, keď detailný opis nejakej sémantiky má tvar nejakého formálneho modelu, s ktorým je schopný pracovať aj počítač.

Pre analýzu pochopenia je kľúčové postulovať čosi, čo by sme mohli nazvať ríšou významov. Významy predstavujú abstrakciu nad mysľou a sú dôležité aj pre pochopenie mysle. Formálne modely ríše významov môžu slúžiť ako východisko pre špecifikáciu počítačových programov (aj takých, ktoré dokážu usudzovať).

Aké dôsledky to má pre naše rozprávanie o poznatkoch a usudzovaní? Ako produktívna je táto abstrakcia nad mysľou, pokiaľ ide o usudzovanie?

Charakterizácia hľadiska Po analýze Searlovej metafory je našou úlohou charakterizovať hľadisko, z ktorého pristúpime k téme usudzovania. Chceme sa dostať od (subjektívnych) mentálnych stavov a procesov k intersubjektívnym významom.

Zastavme sa pri vete (typu) "Počítače nerozumejú slovám, s ktorými manipulujú". Čo to vlastne znamená? Určite môžeme povedať, že počítače nemajú mentálne stavy a procesy. Čo viac však možno povedať s istotou o analyzovanej vete? Ľudia, napriek mentálnym procesom, ktoré sú im vlastné, robia chyby, plynúce z neznalosti významov. Naopak, počítače môžu odhaliť významy, ktoré si ľudia ešte pred týmto odhalením neuvedomovali.

Preto hľadisko, z ktorého pristúpime k analýze usudzovania charakterizujeme takto:

- opierame sa o konštrukciu detailných formálnych modelov,

- preferujeme sémantickú špecifikáciu týchto modelov,
- sémantická špecifikácia je založená opäť na nejakom formálnom modeli.

Formálne modely považujeme za užitočné nástroje, ktoré myseľ používa, aby pochopila komplikované procesy, objekty a situácie. Samozrejme, schopnosť poodstúpiť od takéhoto modelu a skúmať hranice jeho použiteľnosti, je tiež užitočnou vlastnosťou mysle. Domnievam sa však, že touto vlastnosťou myseľ disponuje práve preto, že má prístup do ríše významov, čo jej umožňuje spájať to, čo nie je priamo spojené a izolovať prvky alebo aspekty nejakých celkov (preniesť ich do iného kontextu).[3]

Začneme dedukciou.

3 Dedukcia

Načrtneme veľmi jednoduchý model – prvorádovú dedukciu. Opiera sa o jednoduchú predstavu o svete. Svet tvoria

- objekty (množina U),
- vlastnosti objektov a vzťahy medzi objektami.

Nad touto predstavou sa konštruuje zodpovedajúci jazyk (nazýva sa prvorádovým):

- konštanty, premenné,
- predikátové symboly.

Príklad 1

- *konštantny* – jakub, 7, *premenné* – X, Y
- *predikátové symboly* chlapec, vek
- *vety* – *chlapec(jakub), vek(jakub, 7); iná možnosť zápisu: vek(jakub) = 7;* rozšírenie jazyka – funkčné symboly.

Term je konštanta alebo premenná alebo vznikne aplikáciou funkčného symbolu na jednoduchšie termy.

Príklad 2 *Príklady termov:* jakub, 7, X. *Ak vek pokladáme za funkčný symbol, potom aj vek(jakub) je term, v našom príklade ho považujeme za "meno tojho istého objektu" ako je objekt "pomenovaný" termom 7.*

Príklad 3 *Predikátový symbol* chlapec *je unárny, vytvára vety, ak k nemu pripojíme jediný term (jediný argument). Predikátový symbol* vek *je binárny.*

Počet termov, ktoré možno k predikátovému symbolu jednoznačne určuje jeho *aritu.*

Logika, rozšírenie jazyka:

- pre všetky (niektoré) objekty platí nejaká veta: $\forall X \exists Y\, vek(X,\, Y)$, $\exists X\, chlapec(X)$,
- ak ϕ, ψ sú vety
  - $\neg\phi$ (*nie je pravda, že* ϕ),
  - $\phi \wedge \psi$ (ϕ a ψ),
  - $\phi \rightarrow \psi$ (ak ϕ, tak ψ)

  sú tiež vety.

Poznámka 1 *Jazyk* výrokovej *logiky je podstatne jednoduchší. Obsahuje iba nulárne predikátové symboly (ich arita je* 0*). Nazývame ich výrokové symboly (výrokové premenné), značiť ich budeme ako* $p, q, \ldots$. *Okrem toho sa vo výrokovej logike používajú symboly (spojky)* $\neg, \wedge, \rightarrow \ldots$.

[3]Podľa Newella kognitívna veda je pokus o vyrovnanie sa s vplyvom informatiky (umelej inteligencie) na vedy o kognícii, pričom tento pokus je zasiahnutý červom filozofovania.

Príklad 4 *Príklady výrokovologických viet:* $p, p \leftarrow q, ((p \wedge q) \vee r) \leftarrow t, q \rightarrow p$.

Vety $p \leftarrow q$ *a* $q \rightarrow p$ *sú ekvivalentné, obe čítame: ak* q*, potom* p*. Preferovať budeme zápis* $p \leftarrow q$*, ktorý sa zaužíval v logickom programovaní (ak chcem vypočítať* p*, prv musím vypočítať* q*).*

Úloha logiky:

- špecifikovať, čo znamená, že veta ϕ vyplýva z množiny viet V ($V \models \phi$)
- navrhnúť metódu, ako preveriť, či veta ϕ vyplýva z množiny viet V

Ako sa dá špecifikovať vzťah $\models$ (fakt, že $V \models \phi$) ? Pozorovaním ? Simuláciou ?

Sémantickým rozhodnutím (samozrejme, ovplyvneným pozorovaniami, skúsenosťou s úspešnými a neúspešnými dedukciami). Úspešné dedukcie sú tie, čo z pravdivých predpokladov vždy dostanú pravdivé závery; neúspešné tie, čo niekedy z pravdivých predpokladov odvodia nepravdivé závery.

Interpretácia je funkcia, ktorá priraďuje každej vete jazyka nejakú pravdivostnú hodnotu. Kvôli jednoduchosti budeme predpokladať v tejto chvíli iba dve hodnoty – pravdu (t) a nepravdu (f).

Model množiny viet S je taká intepretácia, ktorá priraďuje každej vete z S hodnotu t. (Modelom vety ϕ je model množiny viet, obsahujúcej iba ϕ, t.j. model $\{\phi\}$.)

Ak S je množina viet a ϕ veta, potom $S \models \phi$ má tento význam: každý model množiny S je aj modelom vety ϕ.

Možno navrhnúť čisto syntaktický systém pozostávajúci z neinterpretovaných symbolov a manipulácií s nimi tak, aby sme dostali syntakticky definovaný vzťah odvodenia. Navyše, pri vhodnej konštrukcii syntaktického systému presne pokryjeme sémanticky definovaný vzťah $\models$:

úplnosť ak platí $S \models \phi$, potom dokážeme syntaktickými prostriedkami odvodiť ϕ z S,

korektnosť ak odvodíme ϕ z S, potom $S \models \phi$.

Zamontujme teraz do toho výpočty. Možno naprogramovať nejaký počítač tak, aby bol schopný rozhodnúť pre každú dvojicu (S, ϕ), či platí (alebo neplatí), že $S \models \phi$? Žiaľ, pre prvorádové jazyky to nie je možné. Ide o *nerozhodnuteľný* problém – dokázateľne neexistuje algoritmus, ktorý by ho riešil. (Problém môžeme nazvať problémom vyplývania.) Problém je aspoň *polorozhodnuteľný*: Existuje algoritmus s nasledovným správaním. Ak dostane vstup (S, ϕ), pre ktorý platí $S \models \phi$, potom rozhodne, že $S \models \phi$, (ak pre ne tento vzťah neplatí, algoritmus nemusí zastaviť).

Aj oveľa jednoduchší problém, problém vyplývania vo výrokovej logike, spôsobuje vážne ťažkosti z hľadiska výpočtového. Existuje síce algoritmus, ktorý ho umožní rozhodovať (známa tabuľková metóda), výpočet podľa tohoto (ale aj podľa ľubovoľného iného) algoritmu nie je zvládnuteľný v rozumnom čase pre vstupy väčšieho rozsahu.[4]

Teda, pre veľmi jednoduchú predstavu o svete a jazyku, vhodnom na opis takéhoto sveta, sa dostávame do vážnych problémov s vypočítateľnosťou dedukcie. Táto správa je nepríjemnejšia pre doslovné chápanie Searlovej metafory ako pre paradigmu konštrukcie formálnych modelov usudzovania. Uvedené negatívne charakterizácie vypočítateľnosti treba považovať za veľmi cenné hodnotenia náročnosti, zložitosti istých problémov. Tento vhľad, toto pochopenie by nebolo možné bez konštrukcie detailného formálneho modelu.

Ak nepatrne skomplikujeme spomínanú predstavu o svete a o jazyku, vhodnom na jeho opis, situácia bude ešte horšia. Budeme predpokladať aj vlastnosti vlastností, vzťahy medzi vlastnosťami, vlastnosti vzťahov a vzťahy medzi vzťahmi. V jazyku budeme mať predikátové symboly (druhého rádu), ktorých argumentami môžu byť predikátové symboly (prvého rádu). Kvantifikovať môžeme aj predikátové symboly (prvého rádu). V takomto jazyku sa možno pokúsiť o axiomatizáciu aritmetiky (sme v ňom schopní vyjadriť princíp indukcie). Platí však veľmi nepríjemný výsledok (Gödel): nie je

[4] Ak príjmeme jednu zo základných hypotéz súčasnej informatiky, $P \neq NP$. Táto hypotéza predpokladá, že trieda problémov algoritmicky rozhodnuteľných v čase polynomickom vzhľadom na rozsah vstupov sa nerovná triede problémov algoritmicky rozhodnuteľných na abstraktnom zariadení, zvanom nedeterministický Turingov stroj, v čase polynomickom vzhľadom na rozsah vstupov.

možné vytvoriť, skonštruovať množinu A axióm, z ktorých by sa dali dokázať všetky vety pravdivé pre prirodzené čísla (neexistuje algoritmus, ktorý by vyčíslil, vypísal všetky axiómy z A). To znamená, že nie je možné vytvoriť úplný axiomatický systém aritmetiky.

V ďalších častiach textu sa budeme pýtať, či existujú také zjednodušenia predstavy o dedukcii, ktoré sú výpočtovo zvládnuteľné. Budeme sa aj pýtať, či dedukcia stačí na všetky problémy, ktoré chceme riešiť usudzovaním. Zavedieme pojem hypotetického usudzovania. Vytváranie hypotéz (domnienok) a ich revidovanie patrí medzi kľúčové kognitívne schopnosti. Čitateľa, ktorý sa zaujíma o podrobnejšie informácie o problémoch, ktoré nasledujú, odkazujem na svoju knihu *Inteligencia ako výpočet* [6].

Napokon, pokúsime sa o istý nadhľad, o analýzu usudzovania v kontexte kognície, o posúdenie miesta a úlohy formálnych a výpočtových modelov spracovania poznatkov z tohto hľadiska.

4 Limitované logické jazyky

4.1 Relačné databázy

Začneme príkladom.

Príklad 5 *Predstavme si jednoduchú databázu, pozostávajúcu z troch tabuliek. Prvá z nich registruje dvojice osôb, rodiča a dieťa. Ďalšie dve obsahujú množiny mužov a žien. Možno treba čitateľa upozorniť, že každé z použitých mien, napríklad* anna, *je menom jedinej osoby.*

RODIČ	*DIEŤA*
anna	*miro*
anna	*zuza*
anna	*štefan*
karol	*miro*
karol	*zuza*
milan	*štefan*
zuza	*eva*
ondrej	*eva*

MUŽ
miro
štefan
karol
milan
ondrej

ŽENA
anna
zuza
eva

Napríklad druhý riadok v prvej tabuľke zaznamenáva, že Anna je rodičom Zuzy.

Všetky databázové vety, ktoré máme zaznamenané v intuitívne dobre pochopiteľných a manipulovateľných tabuľkách možno vyjadriť aj v jazyku logiky.

Databázu z príkladu 5 teraz možno vidieť takto (zdá sa, že prostý pohľad na oba príklady a vhodná skladačka by umožnili malým deťom vzájomne transformovať obe formulácie):

Príklad 6 *Databáza:*

$$\begin{array}{lll}
rodi\check{c}(anna, miro) & mu\check{z}(miro) & \check{z}ena(anna) \\
rodi\check{c}(anna, zuza) & mu\check{z}(\check{s}tefan) & \check{z}ena(zuza) \\
rodi\check{c}(anna, \check{s}tefan) & mu\check{z}(karol) & \check{z}ena(eva) \\
rodi\check{c}(karol, miro) & mu\check{z}(milan) & \\
rodi\check{c}(karol, zuza) & mu\check{z}(ondrej) & \\
rodi\check{c}(milan, \check{s}tefan) & & \\
rodi\check{c}(zuza, eva) & & \\
rodi\check{c}(ondrej, eva) & &
\end{array}$$

Tabuľky z príkladu 5 možno považovať za interpretácie predikátových symbolov $rodi\check{c}, mu\check{z}, \check{z}ena$. Presnejšie, množina dvojíc zo všetkých riadkov tabuľky RODIČ je touto interpretáciou priradená predikátovému symbolu *rodič*, množina všetkých hodnôt z tabuľky MUŽ je priradená predikátovému symbolu *muž*, množina všetkých hodnôt z tabuľky ŽENA je priradená predikátovému symbolu *žena*. Množina databázových viet z príkladu 6 je pravdivá v tejto intepretácii. (Táto interpretácia je jej modelom.)

Teraz si priblížime usudzovanie, typické pre databázy. Každá databáza obsahuje okrem explicitne zaznamenaných viet (riadkov, priamo obsiahnutých v nejakej tabuľke), aj implicitné znalosti. Tie si dokážeme z explicitne reprezentovaného obsahu odvodiť. Napríklad, z informácií o rodičovskom vzťahu si môžeme odvodiť informáciu o vzťahu byť starým rodičom. Takáto informácia nie je v tabuľke z príkladu 5 priamo (explicitne) obsiahnutá, je tam však implicitná, dá sa z uvedenej tabuľky vypočítať, odvodiť. Databázové systémy poskytujú vyjadrovacie prostriedky (databázové jazyky), pomocou ktorých môžeme formulovať požiadavky na získavanie (odvodzovanie) implicitných znalostí.

Jazyk, v ktorom budeme otázky formulovať, sa nazýva *relačný kalkul*.

Príklad 7 *Naša otázka bude:*

$$\{(X,Y) \mid \exists Z(rodi\check{c}(X,Z) \wedge rodi\check{c}(Z,Y))\}.$$

Pýtame sa na množinu dvojíc (X,Y), pre ktoré existuje Z, ktoré je dieťaťom X-a a súčasne rodičom Y-a.

V relačných databázach sa používa jednoduchý, limitovaný jazyk na reprezentáciu poznatkov. Výhodou za toto "uskromnenie sa" je výpočtová efektívnosť inferencie nad vetami tohto jazyka. Odpovede na otázky adresované databáze sa dajú vypočítať v rozumnom čase. Relačné databázy možno považovať za aplikáciu fragmentu logiky, v ktorom sa dajú realizovať efektívne výpočty. Z nášho hľadiska je dôležité všimnúť si, že odpovedanie na otázky nie je založené na vzťahu vyplývania ($\models$). Nepožadujeme, aby odpoveď na otázku bola pravdivá vo všetkých modeloch databázy, stačí nám preveriť, či platí v danej (aktuálnej) databáze.

Ide o dôležitý trend pri implementácii niektorých typov usudzovania: niekedy nepotrebujeme zisťovať, či nejaká veta ϕ vyplýva z nejakej množiny viet V, stačí nám zistiť, či ϕ je pravdivá v (aktuálnom) modeli množiny V. Hovoríme, že miesto odvodzovania (dokazovania) stačí testovať modely.

4.2 Hornovská logika

Iným užitočným ohraničením jazyka logiky je *hornovská logika*.

Opäť využijeme náš rozbehnutý príklad. Na vzťah starého rodičovstva sme sa v príklade 7 pýtali iba nepriamo (bez použitia tohto pomenovania). Podobne, na databázu uvedenú v príkladoch 5 a 6 nemôžeme priamo adresovať (s použitím vhodných predikátových symbolov) otázky na také rodinné vzťahy, ako sú otec, matka, brat, starý otec, strýko atď. Dosiahneme to však, keď rozšírime náš jazyk o pravidlá.

Príklad 8 *Otázku z príkladu 7 možno preformulovať takto: X je starým rodičom Y, ak existuje Z také, že X je jeho rodičom a súčasne Z je rodičom Y. Táto formulácia má tvar pravidla: ak platia nejaké predpoklady, tak platí nejaký dôsledok.*

Pravidlo si zapíšeme takto:

$$star\acute{y}\_rodi\check{c}(X,Y) \leftarrow rodi\check{c}(X,Z), rodi\check{c}(Z,Y).$$

Vety, ktoré neobsahujú žiadne logické symboly (kvantifikátory $\forall, \exists$ alebo spojky $\neg, \wedge, \leftarrow, \rightarrow, \vee, \ldots$) nazveme *atómami*. *Pravidlo* je každá veta tvaru $A \leftarrow A_1, \ldots, A_n$, kde A, A_i sú atómy (spravidla nie základné). (Treba si uvedomiť, že databázové fakty, atómy, sú špeciálnym prípadom pravidiel, kde $n = 0$. Každý atóm A z príkladu 6 môžeme písať v tvare $A \leftarrow$.)[5]

[5]Niektoré formálnosti, "technikality" nemožno obísť. Dovoľujú totiž dôjsť k teoreticky dôležitým záverom.

Príklad 9 *Do databázy môžu patriť okrem faktov uvedených v príklade 6 a pravidla z príkladu 8 aj nasledujúce pravidlá:*

$$\begin{aligned}
otec(X,Y) &\leftarrow rodi\check{c}(X,Y), mu\check{z}(X),\\
matka(X,Y) &\leftarrow rodi\check{c}(X,Y), \check{z}ena(X),\\
brat(X,Y) &\leftarrow matka(M,X), matka(M,Y), mu\check{z}(X),\\
str\acute{y}ko(X,Y) &\leftarrow rodi\check{c}(R,Y), brat(X,R).
\end{aligned}$$

Aj po zavedení pravidiel do relačných databáz sa nedostaneme von zo sféry zvládnuteľných výpočtov.

Pod názvom *problém splniteľnosti* budeme rozumieť tento problém: je daná množina M viet výrokovej logiky. Treba rozhodnúť, či existuje interpretácia (priradenie hodnôt t alebo f všetkým atómom), pre ktorú každá veta z M nadobúda hodnotu t.

Kým vo všeobecnosti je problém splniteľnosti formúl výrokového počtu NP-úplný, pre hornovskú logiku (pre množiny pravidiel) existuje algoritmus, ktorý tento problém rozhodne v lineárnom čase (vzhľadom na rozsah vstupov).

Aj vo všeobecnom prípade majú výpočty v hornovskej logike slušné výpočtové vlastnosti. Množiny pravidiel (v tu zavedenom význame) bez funkčných symbolov nazývame aj deduktívnymi databázami (datalogovskými programami). Aj po tomto rozšírení dostávame odpovede na otázky v rozumnom čase (polynomickom vzhľadom na rozsah vstupov).

Za tento zisk z hľadiska výpočtovej efektívnosti sa platí nižšou odvodzovacou silou: v jazyku hornovských kláuz neodvodíme všetky dôsledky, ktoré možno odvodiť vo všeobecnom prípade: neodvodíme napríklad $p \vee q$, z pravidla $p \rightarrow q$ neodvodíme implikáciu $\neg q \rightarrow \neg p \ldots$

Na záver tejto časti iba poznamenajme, že sú známe aj ďalšie obmedzenia výrokovej logiky alebo predikátovej logiky prvého rádu. ktorých výpočtové vlastnosti sú lepšie než výpočtové vlastnosti nelimitovaných formalizmov.

5 Hypotetické usudzovanie

V počiatkoch umelej inteligencie vládla predstava, že cesta k úspechu povedie touto cestou:

Stav vecí vyjadríme (nejakú oblasť reprezentujeme) v jazyku logiky (dostaneme tak množinu viet V). Problémy z tejto oblasti riešime tak, že logickú reprezentáciu problému, označme ju ako ϕ, sa pokúsime dokázať z V.

Skoro sa ukázalo, že tadeto cesta nepovedie.

Jednou z príčin boli negatívne výpočtové vlastnosti takto postavenej úlohy (už o nich vieme). Tak sa objavila predstava, že miesto dedukcie, logického dôkazu treba hľadať metódy *rýchleho skákania k záverom* ("jumping to conclusions"). Inými slovami: človek málokedy robí dôkladnú dedukciu. Veľmi často vie rýchlo prehliadnuť situáciu a prijať adekvátny záver, hypotézu.

Ďalšou z príčin hľadania alternatívneho spôsobu usudzovania (miesto dedukcie) je obvyklá neúplnosť poznania. Ak máme rozhodnúť o pravdivosti vety ϕ (alebo dokonca množiny viet M) odvodzovaním, málokedy máme k dispozícii všetky relevantné vety, ktoré by nám to umožnili. (Nevieme skonštruovať množinu viet V takú, aby platilo, že $V \models \phi$, prípadne, aby pre každú vetu ϕ z M platilo $V \models \phi$. Napokon, videli sme, že už v aritmetike prirodzených čísel toto nie je možné, ak M je množina všetkých viet pravdivých pre prirodzené čísla.) To znamená, že dedukciu treba popreplietať s generovaním (navrhovaním) hypotéz.

Navyše, ani človek neusudzuje tak, že by jeho závery boli "vynútené" akceptovnými predpokladmi. Dedukcia je skôr okrajový prípad ľudského usudzovania. Spravidla pri našich úsudkoch berieme do úvahy iba kontexty, ktoré považujeme za relevantné, nie všetky možné kontexty (všetky modely).

Budeme sa teda venovať hypotetickému usudzovaniu, usudzovaniu, pri ktorom záver nemusí byť nutne pravdivý, ak sú pravdivé všetky predpoklady. Tento druh usudzovania využíva tzv. common sense.

Základné inutície, ktoré motivujú štúdium takéhoto druhu usudzovania:

- neúplnosť (našich) poznatkov,
- dôležité je analyzovať usudzovanie na pozadí evolúcie poznatkov; poznatky majú svoju dynamiku,
- pre evolúciu poznatkov sú typické revízie; (naše) poznatky sú obvykle revidovateľné,
- závery, ku ktorým prichádza common sense, často závisia na kontexte,
- pri usudzovaní často treba zohľadňovať výnimočné prípady,
- usudzovanie často vedie k chybným záverom,
- dôležitou črtou common sensu je zvládanie, riešenie rozporov (nekonzistentností),
- pri usudzovaní je dôležité rešpektovať približnosť, aproximatívnosť, neurčitosť, vágnosť.

6 Základné prístupy k sémantike hypotetického usudzovania

Môžeme rozlíšiť dva základné prístupy k štúdiu hypotetického usudzovania. Prvý spočíva v rozlišovaní medzi relevantnými a menej relevaantnými konextami. Definuje sa pomocou nejakej relácie preferencie na modeloch. Na základe tohoto prístupu sa pri usudzovaní akceptujú také dôsledky, ktoré sú podopreté najviac preferovanými modelmi predpokladov.

Druhý prístup spočíva v realizácii nejakej predstavy "poupratovania" množiny presvečení – definuje sa taká množina presvedčení, ktorá je v istom zmysle slova nasýtená (nič viac ani nič menej nemôže obsahovať) a je aj v dobre definovanom zmysle slova koherentná.

6.1 Preferenčné vyplývanie

Opäť začneme tabuľkou.

Príklad 10 *Predstavme si informácie o odchodoch vlakov.*

KAM	*KEDY*
kúty	*4:50*
brno	*5:20*
žilina	*6:35*
košice	*6:58*
zvolen	*7:45*
kúty	*8:11*
budapest	*8:50*
praha	*9:09*

Informáciu z tabuľky si môžeme vyjadriť množinou viet v jazyku logiky.
$V = \{$

odchod(kúty,4:50)
odchod(brno,5:20)
odchod(žilina,6:35)

odchod(košice,6:58)
odchod(zvolen,7:45)
odchod(kúty,8:11)
odchod(budapest,8:50)
odchod(praha,9:09) }

Zdravý rozum (common sense) z uvedených viet usudzuje, že $\neg odchod(kúty, 8:00)$.

Tento záver akceptujeme napriek tomu, že nevyplýva z uvedenej množiny viet V. *Ak predikát odchod interpretujeme množinou dvojíc* $M = \{(kúty, 4:50), (brno, 5:20), (žilina, 6:35), (košice, 6:58), (zvolen, 7:45), (kúty, 8:11), (budapest, 8:50), (praha, 9:09), (kúty, 8:00)$, *táto interpretácia je modelom množiny* V, *ale nie je modelom vety* $\neg odchod(kúty, 8:00)$.

To znamená, že uvedený záver nemôžeme získať dedukciou. Napriek tomu, s takýmto typom usudzovania nemáme (spravidla) problémy.

Práve opísaný spôsob usudzovania preferuje niektoré kontexty (modely) pred inými. Interpretácia (model) z nášho príkladu je v istom zmysle slova príliš silná (a neadekvátna). Množina M, ktorú táto interpretácia priradila predikátovému symbolu *odchod* obsahuje aj dvojicu $(kúty, 8:00)$, ktorá nie je využitá na "potvrdenie" žiadnej vety z V. Preto budeme *preferovať* niektoré modely pred inými modelmi.

Každú interpretáciu si môžeme reprezentovať množinou atómov, neobsahujúcich žiadne premenné. Myšlienka: uvedená množina atómov je pravdivá pri danej interpretácii a žiadne ďalšie atómy pri tejto interpretácii nie sú pravdivé. Odteraz môžeme interpretácie porovnávať: niektoré interpretácie sú podmnožinami iných interpretácií. Spomedzi všetkých modelov množiny viet V z nášho príkladu budeme preferovať tie modely, ktoré sú minimálne (vzhľadom na vzťah podmnožiny). Sú to presne tie modely, ktoré predpokladajú iba toľko, koľko stačí na to, aby všetky vety z V boli pravdivé a nič viac (nepredpokladajú).

Zavedieme aj pojem preferenčného vyplývania: veta ϕ preferenčne vyplýva z množiny viet V práve vtedy, keď v každom najpreferovanejšom modeli množiny V je ϕ pravdivá. Značenie: $V \models_{\leq} \phi$, kde $\leq$ predstavuje usporiadanie, na ktorom sú preferencie založené (napríklad vzťah byť podmnožinou).

Príklad 10 demonštruje spôsob usudzovania, ktorý je známy pod názvom predpoklad uzavretého sveta (closed world assumption, CWA): ak vychádzame z množiny viet V (z databázy DB), potom $\neg A$, negácia základného atómu A, sa prijíma vtedy, keď A nevyplýva z V (A neplatí v DB).

Ide o elementárny príklad hypotetického usudzovania: hypotézu $\neg A$ prijímame, ak nie je známe, že by bola vyvrátená (že by platilo A).

Na príklade CWA si môžeme priblížiť niektoré dôležité vlastnosti hypotetického usudzovania.

Nemonotónnosť Ak doplníme cestovný poriadok o nový spoj, o vetu $odchod(a, t)$, potom hypotézu $\neg odchod(a, t)$, ktorú sme mohli odvodiť z predchádzajúceho stavu "databázy" už viac nemôžeme odvodiť.

Všeobecne: Zavedieme si najprv pojem *operátora odvodenia*, budeme ho označovať Cn. Ak máme množinu viet A, potom $Cn(A)$ predstavuje množinu všetkých dôsledkov množiny A (pri spôsobe odvodzovania Cn, o ktorom v tejto chvíli nič nepredpokladáme). Usudzovanie na základe prvorádovej logiky môžeme značiť ako Cn_{FOL}, na základe predpokladu uzavretého sveta ako Cn_{CWA}.

Teraz môžeme definovať nemonotónny operátor odvodenia. Ak existujú množiny viet A, B také, že $A \subset B$ (A je podmnožinou B) a $Cn(A) \not\subset Cn(B)$ ($Cn(A)$ nie je podmnožinou $Cn(B)$), potom hovoríme, že Cn ne nemonotónny.

Cn_{CWA} je nemonotónny. Cn_{FOL} je monotónny: vždy, keď platí $A \subset B$, platí aj $Cn_{FOL}(A) \subset Cn_{FOL}(B)$.

Implicitné revízie S každým nemonotónnym operátorom odvodenia (aj s Cn_{CWA}) je v dobre definovanom zmysle slova asociovaná nejaká revízia: To, že platí $A \subset B$ a súčasne $Cn(A) \not\subset Cn(B)$

znamená, že existuje veta ϕ, ktorá patrí do množinového rozdielu $B \setminus A$ (patrí do B, ale nepatrí do A) a existuje veta ψ, ktorá patrí do $Cn(A) \setminus Cn(B)$. Môžeme to chápať ako *vkladanie* nových viet do A, čím získame B a toto vkladanie nám vynucuje *odstraňovanie* nejakých viet z $Cn(A)$.

6.2 Stabilné množiny presvedčení

Od príkladu 10 sa teraz posunieme k ďalšiemu typu hypotetického usudzovania. Cestovný poriadok platí v *obvyklých prípadoch*. V mimoriadnych situáciách (pred a po sviatkoch, pri výnimočných udalostiach na trati) nemusí platiť, sú možné nejaké výnimky, zmeny. Vtedy známu (strašiu) informáciu o odchodoch vlakov berieme s rezervou a usilujeme sa zistiť nové informácie. Tento spôsob usudzovania formalizujú defaultové teórie.

Defaultové teórie používajú tzv. defaultové pravidlá. Ak chceme napríklad vyjadriť, že odchod do Kútov v obvyklom prípade nie je o 8 : 00, urobíme to takto:

$$\frac{: \neg odchod(k\acute{u}ty, 8:00)}{\neg odchod(k\acute{u}ty, 8:00)}.$$

Uvedený zápis môžeme čítať nasledovne: pokiaľ nie je vyvrátené, že vlak do Kútov neodchádza o 8.00, potom akceptujme, že vlak do Kútov neodchádza o 8:00.

Všeobecný tvar defaultového pravidla:

$$\frac{\alpha : \beta}{\gamma}.$$

Defaultové pravidlo hovorí: vtedy, keď platí α a nie je známe, že by neplatilo β (nemáme nijaké argumenty proti β, nevieme dokázať $\neg\beta$), akceptujeme γ.

Veta "Prednáška z kognitívnych vied býva obvykle vo štvrtok" sa dá pomocou defaultového pravidla vyjadriť takto:

$$\frac{: predn\_CogSc\_\check{s}tv}{predn\_CogSc\_\check{s}tv}.$$

Inak: ak nemáme nijakú informáciu, ktorá by vyvrátila, že prednáška z kognitívnych vied bude vo štvrtok, tak prednáška z kognitívnych vied bude vo štvrtok.

Defaultové pravidlá však nemôženme chápať ako pravidlá odvodzovania. Predstavme si, že máme pravidlá

$$\frac{p : \neg q}{r}, \qquad \frac{: t}{q}.$$

Potom z predpokladu p môžeme odvodiť aj r aj q, ak použijeme najprv prvé a potom druhé pravidlo, ale r nemôžeme odvodiť, ak použijeme najprv druhé pravidlo.

Defaultové pravidlá sa používajú na inú, dôležitú charakterizáciu hypotetického usudzovania. Intuitívne tomu môžeme hovoriť "poupratovanie" presvedčení.[6]

Myšlienku upratovania postihuje pojem extenzie (defaultovej teórie). Defaultová teória (v jazyku J) je dvojica (E, D), kde E je množina viet jazyka J a D je množina defaultových pravidiel, v ktorých sa vyskytujú vety z J. Množina E predstavuje vety, ktorým veríme, ktoré sú overené. D predstavuje nejaký "generátor hypotéz". *Extenzia* je taká množina viet $\mathcal{E}$, pre ktorú platí:

- $E \subseteq \mathcal{E}$ (všetky vety z E patria do extenzie),
- $\mathcal{E}$ je uzavretá vzhľadom na dedukciu (obsahuje všetky svoje deduktívne dôsledky),
- napokon kľúčová vlastnosť: ak je nejaké defaultové pravidlo aplikovateľné na $\mathcal{E}$, potom jeho dôsledok patrí do $\mathcal{E}$; o niečo presnejšie: ak máme defaultové pravidlo

[6] Dr. Mejzlík v Čapkovej poviedke v istej situácii konštatoval "tady není něco v pořádku".

$$\frac{\alpha : \beta}{\gamma}$$

a α platí v $\mathcal{E}$, ďalej v $\mathcal{E}$ nie sú známe argumenty proti β (v $\mathcal{E}$ neplatí $\neg\beta$), potom $\mathcal{E}$ obsahuje γ.

Extenzia je stabilná množina presvedčení v tomto zmysle slova: Nech $\mathcal{E}$ je extenzia, potom $\mathcal{E}$ nezmeníme, ak k nej pridávame deduktívne dôsledky a/alebo dôsledky aplikovateľných defaultových pravidiel.

Dôležité je aj upozorniť, že jedna defaultová teória môže mať viac extenzií (niektoré však nemajú žiadnu extenziu). Existenciu viacerých extenzií si môžeme vykladať ako *alternatívnosť* hypotetického usudzovania. Neexistenciu žiadnej extenzie ako "neupratateľnosť", ako nemožnosť navrhnúť stabilnú množinu presvedčení, ktorá by zodpovedala danej defaultovej teórii.

6.3 Jumping to conclusions?

Nemonotónne usudzovanie sa intenzívne začalo študovať s cieľom pochopiť, opísať a implementovať rýchle usudzovanie, podobné na common sense. Žiaľ, aj preferenčné vyplývanie, aj hľadanie stabilných množín presvedčení je výpočtovo ešte ťažšie ako dedukcia.

V prípade preferenčného vyplývania treba kontrolovať reláciu preferencie na modeloch, v prípade stabilných množín presvedčení treba generovať množiny presvedčení a preverovať, či sú stabilné. V oboch prípadoch treba okrem operácií, používaných pri dedukcii, robiť ešte niečo (nákladné na vynaložené zdroje) navyše.

To platí vo všeobecnosti. Iba poznamenáme, že existujú nejaké špeciálne prípady, jednoduché formalizmy s limitovaným jazykom, ktoré dovoľujú zvládnuteľný výpočet (aj v jednom aj v druhom prístupe).

Zdá sa teda, že hypotetické usudzovanie nemôže splňovať naraz tri požiadavky:

- bude používať pomerne silný jazyk (aspoň výrokovologický),
- bude sa usilovať o istý druh dôkladnosti (koherenciu, úplnosť, nasýtenosť, preferovanie relevantnej informácie),
- bude efektívne vypočítateľné.

V záverečnej časti sa budeme venovať istému východisku. Súčasne načrtneme, v akom zmysle slova je predložená teória hypotetického usudzovania aj charakterizáciou kognície.

7 Ďalšie typy hypotetického usudzovania

Nasleduje iba letmá informácia o niektorých typoch hypotetického usudzovania.

Usudzovanie o hierarchiách Človek svoje poznatky často organizuje do hierarchií. Niekedy sú tieto hierarchie pomerne komplikované – pripúšťajú výnimky, niektoré hierarchické závislosti sú "mäkké" (neplatia vždy), vyskytujú sa v nich konflikty.

Pri štúdiu odvodzovania v hierarchiách sa uplatnili oba tu spomínané prístupy – ak sa vyskytne v hierarchii konflikt a jedno z konfliktných tvrdení je opreté o viac špecifickejšiu informáciu, *preferuje* sa toto tvrdenie. Rozhodnutie o preferencii však nie je jednoduché – niektoré tvrdenia sa vzájomne *neutralizujú*, ak sú konfliktné a ak o žiadnom z nich nemožno povedať, že je opreté špecifickejšiu informáciu. Ak by argumentácia, opierajúca sa o špecifickejšiu informáciu obsahovala aj neutralizované tvrdenie, nemožno ju považovať za platnú. Rozhodnutia o vzájomnej neutralizácii a o preferovaní špecifickejšej informácie možno však urobiť až na základe všetkých možných argumentácií, opierajúcich sa o danú hierarchiu. To znamená, že sa robí istý druh globálneho "upratania" (vytvára sa stabilná množina presvedčení, v ktorej nie sú vzájomne neutralizované presvedčenia).

Ďalšíámi druhmi hypotetického usudzovania sú:

Indukcia Zovšeobecňovanie z pozorovaní – v zásade ide o preferovanie všeobecnejšej informácie, spomedzi všetkých zovšeobecnení danej množiny pozorovaní sa preferujú tie najslabšie zovšeobecnenia.

Abdukcia Vysvetľovanie – ak je daná množina pozorovaní a nejaká teória, hľadáme hypotézy, ktoré po pridaní k danej teórii umožnia odvodiť dané pozorovania; možno povedať, že ide o istý druh "upratania".

Logické programovanie Logické programovanie je dôležitý formalizmus, ktorý výrazne posunul štúdium hypotetického usudzovania.

S logickými programami sme sa v tomto texte už stretli: boli to datalogovské databázy (množiny pravidiel).

Plnú silu logického programovania dosiahneme, ak dovolíme používať aj negáciu. Logický program je množina viet tvaru $A \leftarrow L_1, \ldots, L_k$, kde A je atóm a L_i sú atómy alebo ich negácie (negáciu atómu nazvime negatívnym literálom).

Práve negácie vnášajú do hry hypotetické usudzovanie. Negácia *not* A sa vyhodnocuje takto: preveruje sa, či platí A, ak sa v konečnom čase dokáže, že A neplatí, prijíma sa *not* A. Ide teda vlastne o pokus vyvrátiť hypotézu *not* A, ak je teno pokus neúspešný (nezistí sa, že platí A), hypotéza sa akceptuje.

Tento druh negácie je známy pod menom negácia ako konečné zlyhanie. Sémantika tejto negácie spôsobuje isté problémy. Jedným z najznámejších a najúspešnejších pokusov o sémantickú charakterizáciu tejto negácie je sémantika stabilných modelov. Myšlienka je asi táto: Predpokladajme logický program P. Ak vhodne navrhneme nejakú množinu negatívnych literálov (hypotéz) a odvodíme všetky ich dôsledky pomocou P, dostaneme stabilný model. Stabilné modely logických programov integrujú oba prístupy, o ktorých sme hovorili. Sú to minimálne modely a sú to aj stabilné množiny (ak S je stabilný model a aplikujeme na S pravidlá z P, dostaneme presne S – S je množina dôsledkov pravidiel, aplikovateľných na S).

V rámci logického programovania možno realizovať aj indukciu, aj abdukciu. Navyše, prostriedkami sémantík logického programovania možno adekvátne charakterizovať známe prístupy k revíziám poznatkov (presvedčení).

Zosilnenia jazyka logického pogramovania (disjunktívne logické programy, logické programy, používajúce aj klasickú negáciu, epistemické logické programy, dynamické logické programy, ...) umožňujú charakterizovať komplikovanejšie typy usudzovania.

Dôležité je spomenúť, že prostriedkami logického programovania možno ukázať ekvivalenciu rôznych (pôvodne nezávisle navrhnutých a študovaných) formalizácií nemonotónneho usudzovania.[7]

8 Záver

Najprv poznámka o tom, že mnohé dôležité typy usudzovania som tu ani nespomenul (analógia, usudzovanie za prítomnosti neurčitosti a vágnosti, usudzovanie založené na prípadoch, ...). Na tak limitovanom priestore to ani nemohlo byť rozumným cieľom.

Kľúčové je teraz venovať sa podstatnej otázke. Venovali sme sa nejakým teoretickým, abstraktným prístupom k štúdiu usudzovania. Celkom prirodzene a oprávnene sa dá spýtať, čo to má spoločné s kogníciou a ako to umožňuje pochopiť kogníciu.

Vychádzal som z toho, že súčasťou pochopenia kognície je sémantická špecifikácia "zdravého" usudzovania. Táto špecifikácia nemá byť založená na mentálnych procesoch a stavoch, ale na sfére, ktorá je intersubjektívna, nazval som ju metaforicky "ríšou významov". Načrtol som základné

[7] Opäť pripomínam, že podstatne detailnejšiu informáciu o hypotetickom usudzovaní možno nájsť v [6].

prístupy k takto chápanej sémantickej špecifikácii zdravého, racionálneho usudzovania (rôznych jeho typov).

Samozrejme, reba sa opýtať, nakoľko sa táto idealizácia podobá na reálne usudzovanie. V istom zmysle slova tu reálne usudzovanie nie je zaujímavé, lebo veľmi často býva nekorektné, chybné, neracionálne. Idealizáciu, ktorej sme sa tu venovali, môžeme nazvať IDEÁLNYM USUDZOVAČOM (aby sme nepoužívali toto nešikovné pomenovanie, vystačíme so skratkou IU).

Na druhej strane, núka sa tu isté pozorovanie. Aj veľmi racionálne rozmýšľajúci človek rozmýšľa inak ako IU. Aplikuje nejaké šablóny, vzory. Rýchlo, skutočne rýchlo akceptuje užitočné závery. Niekedy, keď narazí na problém, začne argumentovať. Niekdy mu stačí niekoľko málo argumentov, aby modifikoval vzory, šablóny, ktoré používa a ďalej úspešne aplikoval tieto modifikácie. Výnimočne je nútený podrobne, zdĺhavo a detailne odvodzovať. Môžeme teda vysloviť hypotézu i výskumný cieľ: silné, adekvátne teórie reálnej kognície by sa dali budovať tak, že budú charakterizovať REÁLNEHO USUDZOVAČA (skratka RU), ktorého základnou ambíciou nie je okamžitá dôkladnosť a úplnosť jeho usudzovania, skôr rýchlosť, získaná vďaka používaniu overených vzorov, šablón. Tieto vzory, šablóny možno reprezentovať napríklad stabilnými množinami presvedčení, podopretými nejakými dobre overenými, "naučenými" "teóriami" (bázami poznatkov). V prípade neúspechu, omylov, ťažkostí, problémov pri aplikovaní týchto šablón RU využíva schopnosť krátkej argumentácie. Keď ani to nestačí "volá" IU a s jeho pomocou urobí dôkladnú argumnetáciu. Táto redstava umožní rekonštruovať rýchle usudzovanie a súčasne udržať "dotyk" so zdravým, dôkladným, sémanticky špecifikovaným (ideálnym) usudzovaním. Ak chceme aproximovať zdravé usudzovanie pomocou RU, dôležité je vedieť, čo chceme aproximovať, teda ako vyzerá IU.

References

[1] Čapek, K. *Případ Dra Mejzlíka.* Povídky z jedné kapsy.

[2] Gardner, H. *The Mind's New Science. A History of the Cognitive revolution.* Basic Books Inc., New York, NY, 1987

[3] Hayes, P. *Introduction.* In: P.J.Hayes, M.M.Lucas (eds.), Proceedings of the Cognitive Curricula Conference, vol. 2, Rochester, 1982

[4] *The Oxford Dictionary of Current English. Oxford University Press.* 1984

[5] Searle *Minds, Brains, and Programs.*

[6] Šefránek, J. *Inteligencia ako výpočet.* IRIS, Bratislava 2000