Evolúcia poznania a dynamika usudzovania

Ján Šefránek

Ústav informatiky, Fakulta matematiky, fyziky a informatiky,
Univerzita Komenského, Bratislava

Abstrakt.[1] Hlavným cieľom článku je ukázať, ako umelá inteligencia prispela k prehĺbeniu chápania kogníciе tým, že rozpoznala dôležitosť výpočtového modelovania nemonotónneho usudzovania a tzv. zdravého rozumu (common sense). Pred vznikom umelej inteligencie sa nevenovala dostatočná pozornosť komplexu problémov, týkajúcich sa každodenného usudzovania. Nepochybne však ide o dôležitú stránku kogníciе.

Výskum v oblasti nemonotónneho usudzovania viedol k pochopeniu, že nemonotónnosť je iba symptómom podstatnejších vlastností. Hypotetickosť usudzovania, vynucovanie revízií a súvislosť s evolúciou poznania sa považujú za tie, čo patria medzi kľúčové.

Jadrom článku je informácia o autorovom výskume v oblasti dynamických Kripkeho štruktúr, ktoré umožňujú špecifikovať hypotetické usudzovanie sémanticky. Ide o štruktúry, do ktorých je priamo zavedená nejaká dynamika (zmeny). Táto konštrukcia dovoľuje sémanticky charakterizovať autoepistemickú logiku Przymusinského AELKB, do ktorej možno izomorfne vnoriť mnohé rozhodujúce formalizácie nemonotónneho usudzovania. Ďalej možno vytvoriť sémantiku pre dynamické logické programy, modelujúce evolúciu poznania (báz poznatkov). Nakoniec sa prezentuje informácia o teórii kombinujúcej rýchle a omylné usudzovanie na jednej strane s možnosťou konzultovať ideálny usudzovací stroj (niektorú z logík), ak to je potrebné, na druhej strane.

1 Úvod

Jozef Kelemen robil prvé priekopnícke kroky na pôde slovenskej umelej inteligencie nie tak dávno, niekedy v prvej polovici sedemdesiatych.

Svet umelej inteligencie vtedy veril, že vedecká disciplína, ktorá sa tak razantne hlási k životu, podstatne zosilní schopnosti počítačov. Málokto však vtedy predpokladal, že výskum v umelej inteligencii povedie k dôležitým zmenám v tradičnejších vedách o kognícii.[2]

Dnes to vyzerá tak, že výraznejší vplyv na zosilnenie schopností počítačov mal technologický pokrok a naopak, umelá inteligencia vynútila nové smery výskumu, nové pohľady vo vedách o kognícii.

---

[1] Ďalšie koordináty autora: e-mail `sefranek@fmph.uniba.sk`
`http://www.ii.fmph.uniba.sk/~sefranek`

[2] The Oxford Dictionary of Current English: „**cognition** – knowing, perceiving, or conceiving as an act or faculty distinct from emotion and volition; result of this“.

Klasici umelej inteligencie John McCarthy a Patrick Hayes vo veľmi vplyvnom článku [1] hľadajú v analytickej filozofii a filozofickej logike metódy a pojmy, ktoré by sa dali využiť v umelej inteligencii.

Vývoj v najbližších rokoch po uverejnení tohto článku (1969) však ukázal, že tradičná filozofická logika sa nevenovala mnohým otázkam dôležitým pre výpočtové modelovanie inteligencie. Jedna z týchto otázok je kardinálneho významu: Filozofická logika si nevšímala dostatočne dynamiku poznania a usudzovania. Ďalej sa budem venovať práve prínosu umelej inteligencie pre pochopenie (alebo skromnejšie: pre úsilie o pochopenie) dynamických aspektov usudzovania. Výskum v tejto oblasti je dôležitý aj pre pochopenie tzv. zdravého rozumu (common sense), každodenného ľudského usudzovania (poznamenajme, že vedecké usudzovanie je spravidla iba jeho kultivovanou formou).

Navyše, tento text má tri ciele.

Prvý z nich je v dosť širokom zmysle slova kultúrny. V slovenskom kultúrnom prostredí, domnievam sa, je zakorenená predstava, že formálne modely usudzovania a poznania, vytvorené v logike a umelej inteligencii nehovoria nič relevantného z hľadiska poznania kognície. Sú to iba vymyslené umelé systémy. Ťažko si možno predstaviť, že by kapitola s názvom Cognitive Systems vyšla v nejakej encyklopédii na Slovensku a jej obsahom by bol stručný prehľad formalizmov charakterizujúcich rôzne druhy usudzovania [2]. V prvej časti práce sa budem stručne venovať tvrdeniu, že

- *logika je fakt* (význačná črta) kognície,
- formálne modely usudzovania a poznania sú prijateľným nástrojom poznania kognície.

Druhý z cieľov je v užšom zmysle slova kultúrny. V slovenskom (odbornom) prostredí prevláda názor, že formálne modely usudzovania a poznania zafixujú objekt svojho záujmu, abstrahujú od zmien a sú schopné podať iba statický obraz (zodpovedajúci uzavretej matematickej teórii, dedukujúcej z axióm). Mojím cieľom je informovať o tom, že výskum výpočtových modelov inteligencie prirodzene viedol k rešpektovaniu dynamiky poznania a k pokusom túto dynamiku zachytiť formálnymi prostriedkami. Sústredím sa hlavne na motiváciu pre takéto výskumy.

Tretím cieľom je informovať o vlastných výskumoch v spomínanej oblasti.

Prvým dvom cieľom sa venujem ešte v úvode. Tretiemu cieľu vo zvyšku textu. Výklad je ne-formálny. Čitateľ, zaujímajúci sa o „technické“ detaily, nájde odkazy na literatúru.

1.1 Logika ako fakt kognície

Sila ľudskej kognície (zodpovedajúcich procesov a schopností) je daná predovšetkým silou úsudku.

Ak sme schopní využiť dispozície ľudskej kognície, nepotrebujeme zhromažďovať viac informácií ako stačí na to, aby sme si odvodili nevyvrátiteľné potrebné dôsledky. Vieme dedukovať.

Na druhej strane, ak sme zhromaždili nejaké množstvo – zdánlivo nesúvisiacich, nesúrodých – informácií, vieme v nich nájsť poriadok. Vieme vysloviť hypotézy, vysvetľujúce a systematizujúce tie informácie, ktoré sme zhromaždili. Toto vysvetlenie nie je nevyvrátiteľné. Po získaní ďalších informácií prijatú hypotézu často odmietneme alebo revidujeme. Spravidla však hypotézy, aj keď ich neskôr odmietneme, produktívne orientujú naše myslenie i konanie. Vieme usudzovať hypoteticky.

Hypotetické usudzovanie nás niekedy vedie do slepej uličky. Sila ľudskej kognície je však aj v tom, že sa dokáže učiť z tých situácií, keď sa dostala do slepej uličky.

Logikou ľudia nazvali schopnosť odvodzovať z nejakého množstva zistení podstatne väčšie množstvo dôsledkov a nachádzať poriadok v zozbieraných poznatkoch. I malé deti dokážu rozpoznať, kedy zafungovala logika. Skutočné úspechy v učení sú tie, ktoré sa získali logikou – schopnosťou nachádzať poriadok.

Toto je fakt.

Úlohou kognitívnej vedy je tento fakt pochopiť. Každá metóda a paradigma, ktorá prispeje k jeho pochopeniu, si zasluhuje pozornosť.

Ak opíšeme usudzovanie do takých detailov, ktoré umožnia realizovať niektoré usudzovacie úlohy pomocou počítača, nesporne to treba považovať za krok k pochopeniu kognície. (To, že tento opis je v tvare formalizovanej teórie nezvyšuje, ale ani neznižuje jeho cenu.) Takéto opisy budeme volať výpočtovými modelmi usudzovania.

1.2 Výpočtové modelovanie usudzovania

Výpočtové modely dedukcie sa konštruovali už pred vznikom počítačov. Skúsenosti s ich realizáciou na počítačoch obohatili poznanie o metódy a heuristiky vedúce k zefektívneniu implementácií. Napriek tomu, sú tu neprelomiteľné zásadné bariéry. Sú nimi

1. nerozhodnuteľnosť, prípadne praktická nezvládnuteľnosť základných výpočtových úloh, spätých s dedukciou,

2. veľké nároky na ľudskú prácu pri formulovaní netriviálnych a pre ciele umelej inteligencie dôležitých usudzovacích úloh v tvare potrebnom na užitočnú dedukciu z axióm (z báz poznatkov).

Tieto okolnosti viedli k záveru, že výpočtové modely usudzovania nemožno založiť na dedukcii (resp. vyčerpať využitím dedukcie).

Od polovice sedemdesiatych rokov sa objavujú početné pokusy o pochopenie a výpočtovú realizáciu nededuktívneho usudzovania. Cieľom týchto výskumov bolo navrhnúť také formalizmy, ktoré dokážu

- úspornejšie a hutnejšie vyjadriť, reprezentovať poznatky,

- usudzovať hypoteticky.

Hypotetické usudzovanie je možné aj vtedy, keď nemáme k dispozícii úplnú informáciu, potrebnú na dedukciu. Vziklo charakteristické heslo o skákaní k záverom ('jumping to conclusions' [3]). Výskum hypotetického usudzovania nutne viedol k tomu, že vážna pozornosť sa začala venovať revíziám.

Za viac ako dve desaťročia výskumu sa dosiahli mnohé hlboké výsledky. Výskum sa sústredil na nemonotónosť tohoto druhu usudzovania. To znamená, že usudzovanie sa charakterizovalo nasledujúcou vlastnosťou: existujú také množiny viet A a B, pre ktoré platí $A \subseteq B$, ale neplatí, že všetky dôsledky množiny A sú aj dôsledkami množiny B. Môžeme to zapísať nasledovne: $Cn(A) \not\subseteq Cn(B)$, kde $Cn(X)$ predstavuje množinu všetkých dôsledkov množiny A. Cn budeme ďalej nazývať operátorom odvodenia.

Najznámejšie formalizácie nemonotónneho usudzovania charakterizuje aspoň jedna z nasledujúcich myšlienok:

1. rešpektovanie nejakého globálneho kontextu,
2. preferovanie istých informácií pred inými.

Globálny kontext je dominantný pre hypotetické usudzovanie už na intuitívnej úrovni. Inteligentný a kritický človek prijíma nejakú domnienku vtedy, keď dobre zapadá do sústavy jeho poznatkov. Keď si uvedomuje, že v jeho poznatkoch sú nejaké medzery alebo rozpory, usiluje sa „poupratovať" vo svojich poznatkoch a prijať domnienky, ktoré vytvorený poriadok umožňujú alebo podporujú. „Poupratované" má až vtedy, keď nenachádza argumenty proti týmto domnienkam.

Preferovanie niektorých informácií je zdôvodniteľné tým, že spomedzi všetkých mysliteľných hypotéz (ktorých je nesmierne mnoho) si ceníme iba relevantné. Zavedenie preferencií má byť nástrojom na špecifikovanie relevantnosti.

Oba typy prístupov[3] viedli k výpočtom podstatne ťažším ako sú výpočty, potrebné na realizáciu dedukcie.

V tomto texte sa nebudeme venovať otázke vypočítateľnosti a zložitosti výpočtov. Namiesto toho sa sústredíme na otázku, čo je podstatné na nemonotónnom usudzovaní. Nemonotónnosť, zdá sa, nie je to podstatné, čo by sme si mali všímať na hypotetickom usudzovaní.[4] Jednou z podstatných vlastností je dynamika a nevyhnutnosť túto dynamiku analyzovať, ak chceme pochopiť usudzovanie do takej miery detailnosti, aby sme mohli tvoriť výpočtové modely usudzovania (common sensu).

Pod dynamikou poznania a usudzovania rozumieme:

- poznanie sa vyvíja: ak v (diskrétnom) čase t poznatky agenta α môžeme charakterizovať (pôvodnou) bázou poznatkov KB_t^α, potom v nejakom neskoršom čase $t + i$ môžeme očakávať, že jeho báza poznatkov KB_{t+i}^α sa bude líšiť od pôvodnej (ab-

[3]Najznámejšími formalizáciami, ktoré možno zastrešiť uvedenými nálepkami (rešpektovanie globálneho kontextu, preferovanie relevantného), sú cirkumskripcia [4], defaultové teórie [5], autoepistemická logika [6] a predpoklad uzavretého sveta [7]. Koncom osemdesiatych rokov sa ukázalo, že všetky spomínané prístupy možno charakterizovať pomocou logického programovania [8], [9]. Podrobnú informáciu (v slovenčine) o formalizáciách hypotetického usudzovania poskytuje [10].

[4]Môžeme zacitovať van Benthema [11]: nemonotónnosť je skôr symptómom ako esenciou neštandardného usudzovania. Pod neštandardnosťou mal van Benthem na mysli problémy, ktoré neštudovala logika štandardne.

strakcia. ktorú s obľubou používajú logici, nie je namieste: zmenu bázy poznatkov nie je adekvátne chápať ako vytvorenie inej bázy poznatkov),[5]

- úsudok pri tomto prechode od jednej bázy k druhej funguje tak, že sa usiluje „znovunastoliť poriadok“ hľadaním relevantných hypotéz,
- samozrejme, nové hypotézy vedú opäť k novej báze poznatkov; dynamické usudzovanie, t.j. usudzovanie, ktoré mení bázu poznatkov, stimuluje, ovplyvňuje vývoj (evolúciu) poznania,
- nutnou črtou tohto druhu usudzovania sú revízie a nemonotónnosť.

Zvyšok tohto článku je venovaný autorovým prácam o dynamických Kripkeho štruktúrach [13, 14, 15, 16], ktorých cieľom je pochopiť nemonotónnosť (blížiť sa k jej chápaniu) práve optikou nasmerovanou na dynamiku poznania.

2 Dynamické Kripkeho štruktúry

Predstavme si nemonotónny operátor Cn a dve množiny viet A, B, pre ktoré platí $A \subset B$ a $Cn(A) \not\subseteq Cn(B)$. To znamená, že existujú vety ϕ, ψ také, že $\phi \in B - A$ a $\psi \in Cn(A) - Cn(B)$. Môžeme tomu rozumieť tak, že ϕ patrí medzi nové vety, rozširujúce A a ψ medzi tie, ktoré treba odstrániť z $Cn(A)$ kvôli vloženiam do A.

Práve opísanú situáciu charakterizujeme sémanticky pomocou dynamickej Kripkeho štruktúry. Najprv si objasníme základnú myšlienku Kripkeho štruktúry. Na to budeme potrebovať neformálne vysvetlenia základných pojmov.

Predpokladáme nejaký jazyk J. Budeme predpokladať ďalej, že pre J máme definovaný pojem atomárnej formuly. Tento jazyk je formálny v tom zmysle slova, že o význame jeho výrazov nič nepredpokladáme. *Interpretácia* jazyka je priradenie nejakých objektov výrazom jazyka. Interpretáciu jazyka J si budeme reprezentovať nejakým zoznamom atomárnych formúl (atómov). Idea za tým je, že zoznam atómov $\mathcal{A}$ predstavuje všetky atómy, ktoré sú pravdivé vďaka interpretácii $\mathcal{I}$.

Model nejakej množiny formúl je taká interpretácia, v ktorej je každá formula z tejto množiny pravdivá. Veľmi dôležitý je pojem vyplývania. Budeme hovoriť, že formula ϕ *vyplýva* z množiny formúl M, ak ϕ je pravdivá v každom modeli množiny M, značenie: $M \models \phi$. Keď na triedu (relevantných) modelov sformulujeme rôzne obmedzenia, dostávame rôzne druhy vyplývania.

V našich Kripkeho štruktúrach budeme pracovať s nejakou triedou možných interpretácií nejakého jazyka. Voláme ich možnými svetmi. Kripkeho štruktúry dovoľujú vidieť nejaké súvislosti medzi interpretáciami. Z daného možného sveta sa možno dostať do iných možných svetov vďaka relácii dostupnosti. Relácie dostupnosti umožňujú rozlíšiť niečo, čo by sme mohli chápať ako epistemické alternatívy. Formálne, Kripkeho

[5]Treba pripomenúť, že aj niektorí filozofickí logici – možno pod vplyvom informatiky a umelej inteligencie – začali v osemdesiatych rokoch študovať zmenu teórií [12].

štruktúra je dvojica (W, ρ), kde W je množina možných svetov[6] a ρ je binárna relácia dostupnosti.

Teraz načrtneme, ako možno charakerizovať revízie z prvého odstavca tejto časti pomocou dynamických Kripkeho štruktúr (DKS). DKS je dvojica $(\mathcal{F}, \mathcal{K})$. $\mathcal{K}$ je Kripkeho štruktúra, pomocou ktorej je definovaný (nejaký) vzťah vyplývania $\models$. Tento vzťah môžeme chápať ako sémantickú špecifikáciu operátora Cn. $\mathcal{F}$ je množina nejakých transformácií možných svetov. Ak $f \in \mathcal{F}$, potom $f : W \longrightarrow W$.

Predstavme si teraz $w \in W$ a $f(w) \in W$ také, že $w \not\models \phi$, ale $f(w) \models \phi$. To znamená, že f zodpovedá nejakému vloženiu. Podobne, predpokladajme, že $w \models \psi$ a $f(w) \not\models \psi$. To znamená, f zodpovedá aj nejakému odstráneniu (povedzme, že vynútenému spomínaným vložením).

Základná intuícia spojená s dynamickými Kripkeho štuktúrami je teda nasledovná: DKS pozostáva z dvoch zložiek, z jednej statickej (je ňou obyčajná Kripkeho štruktúra) a z druhej dynamickej (je ňou množina transformácií možných svetov na iné možné svety). Pre nemonotóne usudzovanie sú kľúčové tie situácie, keď sa získavajú nové poznatky a keď kvôli tomuto pridaniu nových poznatkov sme nútení revidovať niektoré poznatky, ktoré sme akceptovali predtým. DKS poskytujú sémantickú charakterizáciu práve týchto situácií. Transformácia jedného možného sveta na iný možný svet predstavuje zmenu v našom poznaní. Môžeme to vidieť tak, že transformácie zodpovedajú generovaniu hypotéz a revíziám (nejaká hypotéza môže byť pravdivá vo svete $f(w)$, ale nie vo w a naopak). Transformácia niekedy „mení“ aj reláciu dostupnosti – presnejšie: pre dané svety w_1, w_2, reláciu dostupnosti ρ a transformáciu f môže platiť: $(w_1, w_2) \in \rho$, $(f(w_1), f(w_2)) \notin \rho$ alebo naopak. To znamená – v prípadoch, keď je operátor odvodenia Cn špecifikovaný v závislosti na relácii dostupnosti, táto transformácia môže spôsobiť nemonotónnosť Cn. V istom zmysle slova DKS je konštrukciou vysvetľujúcou nemonotónnosť usudzovania. Z hľadiska DKS: ak nemonotónnosť je symptóm, potom generovanie hypotéz (získavanie nových poznatkov) a revízie vynútené týmto generovaním sú esenciou (neštandardného, hypotetického, na kontexte závislého, vyvíjajúceho sa usudzovania).

Dynamické Kripkeho štruktúry boli zavedené v [13], kde sa s ich pomocou charakterizovali vloženia a odstránenia viet do/z databázy a predpoklad uzavretého sveta. Celá konštrukcia a spôsob, ako s jej pomocou možno vysvetliť nemonotónnosť usudzovania vedie k predpokladu, že DKS poskytujú (pomerne) univerzálnu sémantickú charakterizáciu nemonotónneho (hypotetického) usudzovania.

Tento predpoklad, že DKS je vhodným nástrojom na unifikujúcu sémantickú charakterizáciu hypotetického usudzovania sa podarilo potvrdiť v [14]. Przymusinského autoepistemická logika poznania a viery [17] (ďalej AELKB) je silný formalizmus, ktorý zovšeobecňuje mnohé známe formalizácie nemonotónneho usudzovania. V [14] je navrhnutá sémantická charakterizácia AELKB pomocou DKS. Pomerne komplikovanú konštrukciu tu nebudeme uvádzať. Sústredíme sa iba na to, že v [14] sa používa obohatená predstava o dynamickosti. Doposiaľ sme dynamiku v DKS chápali tak, že sme zaviedli

[6] My ich stotožňujeme s interpretáciami. Zaužívanejší prístup je komplikovanejší – interpretácie priraďujú nejaké hodnoty možným svetom a formulám. Sú možné aj ďalšie komplikácie pri zavádzaní pojmu Kripkeho štruktúry. Niektoré z nich sú použité aj v citovaných prácach [14, 15, 10].

nejaké transformácie na možných svetoch. Spomínané obohatenie spočíva v tom, že dovoľujeme modifikovať aj reláciu dostupnosti.

Predstavme si Kripkeho štruktúru (W, ρ). Predpokladajme ďalej, že pomocou relácie dostupnosti definujeme pravdivosť nejakých formúl. Konkrétne, v AELKB je formula tvaru $B\phi$ (možné čítanie: veríme, predpokladáme, že ϕ) pravdivá v možnom svete $w \in W$, ak ϕ je pravdivá v každom svete w', dostupnom z w, teda $(w, w') \in \rho$. Takto definovaná pravdivosť vedie niekedy k nekonzistentnostiam. Jedným zo spôsobov ich odstránenia je modifikovanie relácie ρ. Metóda revízie ρ, použitá v [14] je založená na myšlienke, ktorú navrhli [18] pre iné ciele.

Týmto spôsobom sa modifikuje sémantická štruktúra, zodpovedajúca teórii v jazyku (logiky) AELKB. Modifikovaná štruktúra mení množinu odvoditeľných hypotéz tvaru $B\phi$. Vyplývanie, špecifikované takouto sémantikou, môžeme volať *dynamickým preferenčným vyplývaním*. Ak niektoré fakty bázy poznatkov protirečia odvoditeľným presvedčeniam, potom dynamicky modifikujeme danú sémantickú štruktúru, na základe ktorej je špecifikovaný vzťah vyplývania. Uvedený prístup k revíziám AELKB-teórií motivuje zovšeobecnenie DKS. DKS možno obohatiť o zobrazenia z relácií dostupnosti na iné relácie dostupnosti.

Prv, než prejdeme k ďalšej téme, zhrňme. DKS sú sémantickou štruktúrou, ktorá umožňuje jednak (a obvyklým spôsobom) charakterizovať vyplývanie (rôzneho druhu). Okrem toho však umožňujú sémanticky charakterizovať zmeny v poznaní – transformácie medzi možnými svetmi zodpovedajú získavaniu alebo odstraňovaniu poznatkov. Zdôrazniť treba ich schopnosť charakterizovať odstraňovanie poznatkov, vynútené nejakými novými poznatkami. Vďaka schopnosti vyjadriť revízie sa dá charakterizovať aj nemonotónne správanie operátora odvodenia. Máme teda sémantiku, zachytávajúcu usudzovanie, ktoré je dynamické v už spomínanom zmysle: je to usudzovanie, meniace bázu poznatkov, stimulujúce vývoj (evolúciu) poznania.

Pojem dynamickej Kripkeho štruktúry sa dá ďalej modifikovať a zosilňovať. Ďalšou možnosťou, ako zaviesť dynamiku do Kripkeho štruktúr, je zadefinovať na nich nejakú operáciu. S touto metódou sa zoznámime v nasledujúcej časti.

3 Dynamické logické programovanie

Evolúcia poznania je jedným z kľúčových problémov z hľadiska nemonotónneho usudzovania. Už sme uviedli dôvody, prečo treba považovať nemonotónnosť iba za symptóm vývoja poznania (pribúdania nových poznatkov, generovania hypotéz, vynucovania revízií, falzifikácie starších presvedčení).

Cieľom dynamického logického programovania je reprezentácia dynamicky sa meniaceho poznania, dynamických báz poznatkov. Základná myšlienka dynamického logického programovania [19] je táto: Je daný program P. Ten predstavuje počiatočnú bázu poznatkov. Modifikujeme ho pomocou iného programu U. Ten predstavuje bázu poznatkov, špecifikujúcu modifikácie pôvodnej bázy poznatkov. Modifikujúci program môže predstavovať poznatky o zmenách v prostredí, ktoré nastali časom, môže špecifikovať revízie presvedčení alebo aj zmenu hľadiska. Výsledkom takejto modifikácie je nový program $P \oplus U$. Uvedený prípad možno zovšeobecniť na postupnosť programov $P \oplus U_1 \oplus \cdots \oplus U_n$ (alebo inú komplikovanejšiu štruktúru, napríklad graf, pozri [20]).

Za poznámku stojí grafová štruktúra na programoch. Každý z programov prirodzene predstavuje nejaký modul (modulárnej) bázy poznatkov. Jednotlivé moduly je vhodné aktivovať v nejakých kontextoch (po nejakých udalostiach). Navyše, samotná grafová štruktúra môže byť dynamická: po nejakých udalostiach, v nejakých kontextoch sa modifikuje samotný graf.

Keď modifikujeme programy, kotrých súvislosti sú definované nejakou komplikovanou štruktúrou programov (báz poznatkov), dostaneme nejaký výsledný (modifikovaný) program. Ten zodpovedá relevantnej časti modulárnej bázy poznatkov, relevantnej vzhľadom na daný kontext, udalosť a pod.

Po dôležitých intuíciách, spätých s dynamickým logickým programovaním sa vráťme k nemu samotnému. Programy, ktoré sa používajú v dynamickom logickom programovaní, sa nazývajú zovšeobecnenými logickými programami.

Sú to množiny formúl tvaru $L \leftarrow L_1, \ldots, L_k$. Jedno možné čítanie: Ak chceš odvodiť L, musíš najprv odvodiť L_1 a L_2 $\ldots$ a L_k. Druhé čítanie: Ak platí L_1 a L_2 $\ldots$ a L_k, potom platí L.

Tieto formuly nazývame aj pravidlami. L nazývame hlavou pravidla a $L_1, \ldots, L_k$ jeho telom.

Pritom L a všetky L_i sú buď atómy alebo ich negácie (nazývame ich *literálmi*). Negácia, ktorá sa tu používa, sa líši od klasickej negácie. Ide o tzv. defaultovú negáciu: Predpokladá sa, že platí $not\ A$, pokiaľ nie je dokázané, že platí A.

Pravidlá modifikujúceho programu špecifikujú nejaké vloženia do (a odstránenia z) pôvodného programu. Defaultová negácia v hlave pravidla slúži na špecifkáciu odstránení.

V [19] sa analyzovala možnosť špecifikovať $P \oplus U$ vychádzajúc z modifikácií interpretácií. Prednosťou takéhoto prístupu je, že dovoľuje modifikáciu špecifikovať sémanticky. Dovoľuje tak abstrahovať od náhodilostí, závislých na nepodstatných syntaktických rozdieloch. Pod *modifikáciou interpretácií* rozumieme to, že modifikovaný program (teóriu) zamýšľame špecifikovať na základe modifikovanej sémantickej štruktúry. Môžeme to symbolicky vyjadriť ako

$$Mod(P') = Update_U(Mod(P)).$$

$Mod(X)$ je množina všetkých relevantných modelov programu X. $Update_U(M)$ je modifikácia triedy modelov M, determinovaná programom U. Preto, P' je modifikáciou P pomocou U práve vtedy, keď množina všetkých relevantných modelov P' sa zhoduje s množinou modifikovaných relevantných modelov programu P.[7]

Podľa [19] modifikovaný program $P \oplus U$ nezávisí iba na sémantike programov P a U, ale aj na ich syntaxi. Tento uzáver demonštrovali príkladom, ukazujúcim, že modifikácie interpretácií musia často brať do úvahy aj argumentačné (zdôvodňovacie) závislosti v pôvodných bázach znalostí. Ak chceme odstrániť nejaký literál, musíme odstrániť alebo neutralizovať aj všetky jeho zdôvodnenia (argumenty v jeho prospech).

[7] V práci [19] sa používajú stabilné modely. Považujú sa za relevantné modely, charakterizujúce sémantiku zovšeobecnených programov. Nebudeme tu však zaťažovať definíciou stabilných modelov.

V [15] bola navrhnutá kripkeovská sémantika zovšeobecneného logického programu. Táto sémantika dovoľuje zaznamenať aj argumentačné (zdôvodňovacie) závislosti medzi literálmi. Kripkeho štruktúra asociovaná s logickým programom P vyzerá takto: Možnými svetmi sú všetky (čiastočné) interpretácie programu P, t.j. konzistentné množiny literálov. Okrem toho, je tu jeden možný svet, zastupujúci všetky „nemožné, absurdné“ interpretácie, teda nekonzistentné množiny literálov.

Relácia dostupnosti ρ_1 je definovaná takto: z možného sveta w možno dostúpiť do sveta w', ak v P existuje pravidlo $L \leftarrow L_1, \dots, L_k$ také, že $L_1, \dots, L_k$ je pravdivé vo w (hovoríme, že dané pravidlo je aplikovateľné na w) a L je pravdivé vo w' (okrem toho, vo w' je pravdivé všetko, čo je pravdivé aj vo w). Ľahko si predstavíme graf, ktorý zodpovedá doteraz zavedenej konštrukcii: možné svety sú jeho vrcholy, dvojice $(w, w') \in \rho_1$ hranami.

Pokračujeme v konštrukcii: Keď máme možný svet w taký, že žiadne pravidlo nie je aplikovateľné na w, potom môžeme dostúpiť do takého možného sveta, ktorý okrem všetkých literálov z w obsahuje aj všetky ostatné negatívne literály, vyjadriteľné v jazyku programu P. Presnejšie: Nech na w nie je aplikovateľné žiadne pravidlo z P. Potom $(w, v) \in \rho_2$, pričom $v = w \cup \{not\ A : A \notin w\}$. Hrany z ρ_2 nazývame zúplňovacími.

Kripkeho štruktúra asociovaná s programom P je dvojica $(W, \rho_1 \cup \rho_2)$, kde W je množina všetkých konzistentných množín literálov (vrátane prázdnej množiny) spolu s jediným reprezentantom všetkých nekonzistentných množín literálov. Budeme ju značiť ako $\mathcal{K}_P$. $\mathcal{K}_P$ dovoľuje elegantne a jednoducho identifikovať stabilné modely programu P.

V súvislosti s dynamickým logickým programovaním je dôležité, že možno zadefinovať operáciu na Kripkeho štruktúrach, ktorá predstavuje modifikáciu jednej štruktúry inou a sémanticky špecifikuje modifikované programy. Nasledujúci výklad tejto operácie sa líši od pôvodnej definície. Opiera sa o modifikácie navrhnuté v [21].

Predpokladajme teraz Kripkeho štruktúry $\mathcal{K}_P$ a $\mathcal{K}_U$ asociované s programami P a U. Možné svety oboch štruktúr sú interpretáciami toho istého jazyka (v ktorom možno sformulovať P i U). Zadefinujeme na nich operáciu, ktorej výsledkom je Kripkeho štruktúra $\mathcal{K}_{P \oplus U}$. Táto operácia zachováva $\mathcal{K}_U$ (okrem zúplňujúcich hrán). Z $\mathcal{K}_P$ zachováva všetko, okrem zúplňujúcich hrán a takých hrán, ktoré sú v konflikte s nejakou hranou z $\mathcal{K}_U$. Napríklad: Predpokladajme, že v $\mathcal{K}_U$ vedie hrana z w do u, pričom u obsahuje literál L. Ďalej predpokladajme, že v $\mathcal{K}_P$ vedie hrana z w do v, pričom v obsahuje komplementárny literál $\overline{L}$. Potom hrana z w do v je v konflikte s hranou z w do u. Všetky ρ_1-hrany z $\mathcal{K}_P$, ktoré sú v konflikte s nejakou hranou z $\mathcal{K}_U$, odmietneme.

Výsledkom operácie je Kripkeho štruktúra, konštruovaná takto:

1. obsahuje presne tie isté možné svety ako $\mathcal{K}_P$ a $\mathcal{K}_U$,
2. každá ρ_1-hrana z $\mathcal{K}_U$ je jeho hranou,
3. každá neodmietnutá ρ_1-hrana z $\mathcal{K}_P$ je jeho hranou,
4. zúplňujúce hrany sa dodajú všade, kde to je možné.

Táto sémantika je vo vývoji, jej modifikácie a zovšeobecnenia sa študujú v [21]. Čo je dôležité, ide o sémantiku, schopnú zachytávať argumentačné, zdôvodňovacie závislosti medzi literálmi.

Pred prechodom k poslednej téme opäť zhrnutie. Pokus modelovať evolúciu báz poznatkov pomocu dynamického logického programovania sa zdá byť veľmi perspektívnym. Ide o flexibilný a mnohotvárny nástroj. Zvlášť cenná je možnosť modelovať zmenu hľadiska pri modularizovanej báze poznatkov. Ak súvislosti jednotlivých modulov špecifikujeme pomocou dynamického grafu a modifikácie tohoto grafu dokážeme urobiť závislými na zmenách relevantných kontextov, sme schopní formalizovať ideu dynamického preferenčného vyplývania. Samozrejme, zmeny bázy poznatkov, sú prirodzeným výsledkom tohto prístupu.

V nasledujúcej časti použijeme kripkeovskú sémantiku dynamického logického programovania ako „implementačné prostredie" pre matematickú teóriu rýchleho (a možno chybného) usudzovania, ktoré má k dispozícii možnosť konzultovať „ideálneho usudzovača", zodpovedajúceho nejakej logickej formalizácii [16].

4 Rýchle usudzovanie a logika

Cieľom článku [16] je nanovo pochopiť myšlienku skákania k záverom (jumping to conclusions, [3]). Už sme uviedli, že formalizácie, inšpirované týmto heslom, viedli k problémom, ťažkým z výpočtového hľadiska. Je to dôsledkom logickej rigoróznosti a dôkladnosti: Hľadajú sa množiny presvedčení, ktoré sú „dokonale upratané", t.j. konzistentné, uzavreté vzhľadom na dedukciu a „nasýtené" (podľa nejakého dobre definovaného kritéria). V množine všetkých modelov sa hľadajú tie, ktoré sú najpreferovanejšie (vzhľadom na nejakú reláciu preferencie).

Samozrejme, za túto dôkladnosť a rigoróznosť sa platí výpočtovou náročnosťou problémov, ktoré treba riešiť.

Návrh z [16] spočíva v „kopírovaní" spôsobu ako uvažujú ľudia. Ľudia spravidla minimalizujú odvodzovanie. Keď rozmýšľajú, používajú vzory, schémy, ktoré si osvojili (vďaka genetickej výbave alebo tým, že sa ich naučili). Závery, ku ktorým dochádzame, keď rozmýšľame, často „vidíme" na prvý pohľad. Niekedy ani nie sme schopní sformulovať vety, na ktorých by sme naše závery založili. To však neznamená, že tieto závery nemôžu byť správne. Na druhej strane, nie sú ani imúnne voči omylom. Vtedy, keď ľudia rozpoznajú, že sa mýlia, alebo keď ich usudzovanie vedie k iným problémom, začínajú argumentovať, analyzovať dôvody.

Formálna teória, prezentovaná v [16] kombinuje rýchle usudzovanie na základe vzorov s dôkladným usudzovaním podľa nárokov logiky. Toto výpočtovo náročnejšie usudzovanie sa aktivuje vtedy, keď sa identifikujú nejaké problémy pri rýchlom usudzovaní.

Na tomto mieste nepôjdeme do detailov, iba neformálne vysvetlíme základné myšlienky. Vzor (belief pattern) sa definuje ako nejaká množina literálov (nemusí byť konzistentná, ani úplná, t.j. nemusí obsahovať každý atóm alebo jeho negáciu). Usudzovanie je aktivované nejakými udalosťami. Udalosť je v navrhnutej formalizácii reprezentovaná správou v nejakom komunikačnom jazyku. Predstava za tým je taká, že prostredie, v ktorom sa nachádzajú usudzujúce agenty, vysiela agentom nejaké impulzy. Tie impulzy reprezentujeme správami. Na relevantné správy agenty odpovedajú. Agenta chápeme ako čiastočnú funkciu, odpovedajúcu na niektoré správy.

Usudzujúce agenty sú agenty vybavené nejakou dodatočnou štruktúrou. Majú k dispozícii nejakú množinu vzorov. Ovládajú istý komunikačný jazyk. Sú schopné v nejakých stavoch aktivovať relevantný vzor. Vedia odpovedať na nejaké typy správ. Vedia modifikovať relevantný vzor v závislosti na aktuálnom vzore a danej udalosti. V prípade, keď sa udeje udalosť istého typu, vedia konzultovať „ideálny usudzovací stroj". Predpokladom takéhoto konzultovania je prístup k nejakej báze poznatkov.

Uvedené schopnosti usudzujúceho agenta sa formalizujú dvojicami tvaru (Σ, F). Tieto dvojice nazývame protokolmi. Σ je postupnosť správ (udalostí), F je nejaká funkcia. Úlohou tejto funkcie je „implementovať" nejakú usudzovaciu úlohu, ak sa vyskytla postupnosť udalostí (správ) Σ.

Tieto základné myšlienky sa dajú pomerne jednoducho formalizovať. V [16] bola navrhnutá dvojkroková formalizácia. V prvom kroku sa abstraktne definoval všeobecný rámec – elementárne výrazy komunikačného jazyka a typy navrhovaných funkcií.

V druhom kroku sa všeobecný rámec „implementoval" v prostredí dynamického logického programovania. Detailnejšie sa vymedzujú protokoly a funkcie, ktoré adekvátne reagujú na udalosti. Konzultácie ideálneho usudzovacieho stroja spočívajú v aktivovaní nejakej bázy poznatkov (nejakých modulov štrukturovanej množiny zovšeobecnených programov) a reakcie na udalosti sa špecifikujú pomocou sémantiky dynamických logických programov.

Možno badať istú paralelu medzi rýchlym usudzovaním na základe vzorov a usudzovaním nad Kripkeho štruktúrami asociovanými s programami. Tie si totiž môžeme reprezentovať nejakými grafmi, ktorých uzlami sú objekty toho istého typu ako vzory (belief patterny). Mnohé usudzovacie problémy sa dajú nad Kripkeho štruktúrami riešiť s použitím niektorých uzlov (možných svetov) – tých, ktoré zodpovedajú stabilným modelom. Ďalšie (revízie) možno riešiť spájaním nejakých grafov.

Treba poznamenať, že za istých okolností možno funkcie definované na vzoroch a udalostiach trénovať (učiť), napríklad pomocou evolučných algoritmov.

Môžeme povedať, že navrhnutá teória umožňuje aproximovať ideálne usudzovanie (vrátane hypotetického) rýchlym skákaním k záverom na základe vzorov aktivovaných nejakými udalosťami. Evolúciu poznania si môžeme predstaviť ako postupné vylepšovanie tejto aproximácie.

Uzavrime tým, že model kombinovaného používania vzorov a usudzovacích strojov je dynamickým modelom. Evolúcia poznania sa dostáva do centra pozornosti v týchto situáciách:

- nové udalosti aktivujú ďalšie vzory,
- problémy, ktoré sa vyskytnú pri odpovedaní na niektoré správy, sa riešia tak, že tie isté správy sa pokúšajú zodpovedať inak, s použitím inej „usudzovacej výbavy" – ďalších vzorov, explicitne sformulovaných poznatkov, usudzovacích strojov,
- existuje mechanizmus modifikujúci explicitne sformulované poznatky,
- správanie usudzujúceho agenta sa vyvíja – funkcie, ktoré aktivujú vzory, explicitne sformulované poznatky alebo usudzovacie stroje sa môžu meniť (ich správanie sa môže vylepšovať napríklad pomocou evolučných algoritmov).

5 Závery

Vráťme sa k otázke, postavenej na začiatku tohto textu.

Ide o vplyv výskumu nemonotónneho (hypotetického) usudzovania (common sensu) v umelej inteligencii na

- pochopenie kognície,
- tradičnejšie disciplíny, skúmajúce kogníciu,
- pochopenie evolúcie poznania, umožnenej dynamikou usudzovania.

Už sme zdôraznili, že umelá inteligencia prišla s novými pohľadmi na kogníciu. Tradičnejšie prístupy nevenovali dostatočnú pozornosť mnohým otázkam kardinálneho významu. Tu sme si všímali jednu z týchto otázok: dynamiku poznania a usudzovania.

Okolnosť, že báza poznatkov znalostného systému si musí udržať identitu aj po modifikáciách, viedla k nutnosti dynamického pohľadu na bázy poznatkov (teda na teórie, po malej abstrakcii). Tým sa nastolil problém používať usudzovanie, ktoré spôsobuje zmeny teórií. Nevyhnutným predpokladom toho je, pochopiť toto usudzovanie a opísať ho tak detailne, aby sa dalo implementovať na počítači. Fundamentálnym problémom je tu problém adekvátnej sémantickej charakterizácie takéhoto usudzovania.

Táto – možno povedať – technologická motivácia inšpirovala výskum týkajúci sa samotnej podstaty ľudského usudzovania.

Výskum nemonotónneho usudzovania od konca sedemdesiatych rokov dosiahol mnoho hlbokých výsledkov (od teoretických základov, cez experimentálne implementácie až po priemyselné aplikácie). Oblasť je naďalej objektom intenzívneho vývoja. Stále je tu mnoho ťažkých neriešených problémov. Nie je vylúčené, že niektoré dôležité typy hypotetického usudzovania ešte neboli predmetom záujmu. Obzvlášť dôležité sú problémy vypočítateľnosti a výpočtovej zložitosti, hľadanie výpočtovo zvládnuteľných verzií ťažkých problémov.

Typickou črtou (ľudskej) kognície je schopnosť pochopiť a meniť svoje okolie pomocou vlastnej mentálnej výbavy.[8] Táto schopnosť je nemysliteľná bez hypotetického usudzovania. Vďaka nemu môžeme hovoriť o evolúcii nášho poznania. Na hypotetickom usudzovaní a evolúcii poznatkov je dôležité to, že nutne musí vykročiť von z mentálneho prostredia. O tom, ktoré úsudky sú zdravé a produktívne, nerozhodujeme my (naše mentálne procesy a stavy). Preto výskum, spočívajúci v konštrukcii výpočtových modelov usudzovania a v ich nasadzovaní na problémy reálneho sveta je užitočným nástrojom verifikácie (a falzifikácie) našich predstáv o zdravosti a produktívnosti hypotetického usudzovania.

Tento výskum je aj výzvou pre viaceré disciplíny, ktorých cieľom je pochopiť kogníciu. Metóda konštrukcie výpočtových modelov ponúka v tomto smere konkurencieschopný nástroj.

[8] Samozrejme, má to svoje pozitívne i negatívne dôsledky.

Referencia

[1] McCarthy, J., Hayes, P.: Some philosophical problems from the standpoint of Artificial Intelligence. *Machine Intelligence 4*, B. Meltzer, D. Michie (eds.), Edinburgh University Press, 1969, 463-502

[2] Dix, J.: Cognitive Systems. In John G. Webster (ed.), *Encyclopedia of Electrical Engineering*, vol. 3 (of 24), 531-540, John Wiley & Sons, 1999

[3] McCarthy, J.: Epistemological problems of Artificial Intelligence. em Proceedings of IJCAI77, Cambridge, Massachusets, 1977, pp.1038-1044

[4] McCarthy, J.: Circumscription – a form of non-monotonic reasoning. *Artificial Intelligence* **13** (1980), 27-39

[5] Reiter, R.: A logic for default reasoning. *Artificial Intelligence*, **13** (1980), 81-132

[6] Moore, R.C.: Semantic considerations on nonmonotonic logic. *Artificial Intelligence*, **25** (1985), 75-94

[7] Reiter, R.: On closed world data bases. In H. Gallaire and J.Minker (eds.), *Logic and Data Bases*, Plenum, New York, 1978

[8] Przymusinski, T.: Three-valued nonmonotonic formalisms and semantics of logic programs. *Artificial Intelligence* **49** (1991), 309-343

[9] Marek, W., Truszczynski, M.: *Nonmonotonic logic: context-dependent reasoning.* Springer, 1993

[10] Šefránek, J.: *Inteligencia ako výpočet.* IRIS, 2000

[11] van Benthem, J.: Semantic parallels in natural language and computation. *Logic Colloquium'87*, North Holland, 1989, 331-375

[12] Gärdenfors, P., Rott, H.: Belief Revision. In: *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol.4. Oxford 1995.

[13] Šefránek, J.: Dynamic Kripke structures. *CAEPIA'97.* Actas de la VII Conferencia de la Associación Española para la Inteligencia Artificial, Malaga 1997, 271-283

[14] Šefránek, J.: Belief, knowledge, revisions, and a semantics of non-monotonic reasoning. In M. Gelfond, N.Leone, G. Pfeifer (eds.), *Logic Programming and Nonmonotonic Reasoning*, Springer, LNAI 1730, 1999, 63-77

[15] Šefránek, J.: A Kripkean semantics for dynamic logic programming. In: M. Parigot, A. Voronkov (eds.), *Logic for Programming and Automated Reasoning*, Springer, LNAI 1955, 2000, 469-486

[16] Šefránek, J.: Jumping to conclusions. Rukopis, 2001

[17] Przymusinski, T.: Autoepistemic logic of knowledge and belief. *Artificial intelligence* **95** (1), 1997

[18] Buccafurri, F., Eiter, T., Gottlob, G., Leone, N.: Enhancing symbolic model checking by AI techniques. IFIG Research Report 9701, September 1997

[19] Alferes, J., Leite, J., Pereira, L., Przymusinska, H., Przymusinski, T.: Dynamic logic programming. *Proc. of KR'98*, Morgan Kaufmann, 1998

[20] Dell'Acqua, P., Leite, J., Pereira, L.: Evolving multi-agent viewpoints – an architecture.
`http://centria.di.fct.pt/~lmp/publications/Biblio.html`

[21] Mariničová, E., diplomová práca (ukončená bude v apríli 2001)