TEORIE RACIONÁLNÍHO USUZOVÁNÍ

Logické metody a algoritmy usuzování pro systémy umělé inteligence a reprezentace znalostí

Petr Jirků

14. října 2003

Univerzita Karlova,
Filozofická fakulta, Katedra logiky
CZ–116 42 Praha 1, Celetná 20
E-mail: Petr.Jirku@ff.cuni.cz
URL: http://www.cuni.cz/~jirkup

Předmluva

Reprezentace znalostí v logice je jedním z centrálních témat studia dynamiky znalostí v rámci umělé inteligence. Jde vlastně o studium změn epistemických stavů, což ovšem už není jen záležitost aktuálních zkoumání v kontextu umělé inteligence. Taková zkoumání mají v širším rámci formální logiky dost dávnou tradici, kterou je možné sledovat od Aristotelových *Druhých analytik*. V průběhu posledních dvou tisíciletí formální logika, pěstovaná právě v aristotelské tradici, uspokojivě vysvětlila dedukci. Koncept dedukce je ale statický. Je sice znamenitým nástrojem porozumění logické struktuře úsudků, ale nestačí k tomu, abychom dobře porozuměli její dynamice.

Aktuálním úkolem logiky na tomto poli je vzít do úvahy další vlastnosti konceptu *formálního logického systému* a reflektovat nejen korektnost argumentů vzhledem k pravdivosti tvrzení, ale též vlastnosti nositelů znalostí a schopnosti nositelů znalostí odvozovat, a to, pokud možno, ideálně a racionálně. Naším cílem je tudíž studium rozšíření logických systémů o takové vyjadřovací a odvozovací schopnosti. Nejde nám ale samozřejmě o libovolná rozšíření. Jde nám především o přístupy, které je možné označit ještě jako *racionální*. Nebudeme se ani příliš věnovat psychologickým aspektům argumentace, budeme se pohybovat stále ještě v rámci formální logiky, ale přece jen poněkud rozvolníme jak syntaktický koncept odvození, tak sémantický koncept vyplývání a budeme hledat racionální a dobře zdůvodněné cesty aktualizace znalostí a porozumění struktuře jejich důvodů a důsledků. Takové cesty usuzování jsou často označovány, ne právě šťastně, jako metody nemonotónního usuzování. Rozumí se tím takové argumenty, které na rozdíl od deduktivních, mohou být zpochybněny nebo dokonce falzifikovány ve světle nových položek znalostí a nově získaných faktů. A to je právě tématem studia dynamiky epistemických stavů.

Moderní historie zkoumání teorií racionálního usuzování je dvacetiletá. V roce 1980 vyšlo specializované číslo časopisu *Artificial Intelligence*, které obsahovalo mj. články Raymonda Reitera [114], Drew Dermotta, Jon Doylea a dalších. V osmdesátých letech pak vychází publikace Petra Gärdenforse [47] *Knowledge in Flux* a několik relevantních textů ve třech svazcích publikace *Handbook of Philosophical Logic* (editoři D. Gabbay a F. Güenthner) [38]. A konečně v devadesátých letech pod redakcí Dov Gabbaye a dalších vychází postupně několikasvazkový *Handbook of Logic in Artificial Intelligence and Logic Programming* [41].

Pro porozumění tomuto textu nepředpokládáme žádné zvláštní znalosti speciálních logických kalkulů, pouze se odvoláváme na některé základní koncepty výrokové a predikátové logiky či teorií prvního řádu,

které je možné si připomenout např. v dnes už klasických publikacích Josepha R. Shönfielda [125] nebo Richarda L. Epsteina [33]. V českém jazyce vyšla v nakladatlství Karolinum kniha Antonína Sochora [129] *Klasická matematická logika.*[1]

Autor je zavázán díky *Open Society Fund*, který mu v roce 1995 umožnil měsíční pobyt v Exeter College v Oxfordu, kde mohl shromáždit velkou část materiálu převážně prostřednictvím *Bodleyho* knihovny a nakladatelství a vydavatelství Blackwell. První impulzy pro tuto práci vzešly ovšem při autorově půlročním pobytu v Laboratoři reprezenlace znalostí v logice na univerzitě ve švédském Linköpingu. Větší část této práce pak vznikla s podporou grantu GA ČR 401/98/0383 Alternativy klasické logiky.

Některé části této práce pak byly prezentovány na přednáš-

[1] V místech, kde jsou využita některá pro laika možná méně známá matematická tvrzení např. z teorie množin, jsou vždy provedeny odkazy na příslušnou literaturu.

kách a v seminářích oboru logika na Filozofické fakultě UK v letech 1999 a 2002. Na těchto přednáškách a seminářích studenti navrhli mnohá zlepšení textu i způsobu výkladu. Zvláštní poděkování patří Mgr. Martinovi Špalkovi a panu Janovi Filippimu za grafické provedení obrázků a za pečlivé přečtení textu.

14. října 2003 *Petr Jirků*

Obsah

Část I

Dedukce

Kapitola 1

Úvod

Porozumět cestám usuzování je jeden z nejdůležitějších úkolů studia umělé inteligence a reprezentace znalostí. Je to zároveň otázka porozumění lidské racionalitě. Usuzování je vlastní každému člověku, který při svém rozhodování a konání používá rozum. To je naštěstí stále ještě obvyklé, i když ne vždy. Navíc, ne vždy tak činíme vědomě a se znalostí logicky korektních postupů. Někdy to nevadí, jindy to ale může být fatální. Vždy ovšem stojí za to uvědomovat si, *co to vlastně děláme, když používáme nějaké formální vzorce usuzování*.

V této publikaci budeme systematicky studovat teorie vytvářené pro reprezentaci znalostí v logice, chápané ovšem v širokém slova smyslu deduktivních teorií.

Pro náš účel začneme rekapitulací nejvýznamnějších vlastností dedukce, která je nejznámější a nejlépe prozkoumanou metodou odvozování zkoumanou v rámci logiky.

Čtenáři ještě později v této kapitole předložíme koncepty, které jsou nezbytné pro porozumění následujícím kapitolám o nemonotónním usuzování a usuzování s neúplnou informací. Většina těchto pojmů je vyložena v tradičním tarskiánském stylu, i když gentzenovský či hilbertovský přístup je zde stejně oprávněný. Pro naše účely je ale na tomto místě podstatné popsat ty nejobecnější vlastnosti dedukce, které jsou nezávislé na vyjadřovací síle prostředků reprezentace znalostí, tedy nezávislé především na jazyce.

Jsou nejméně dva dobré důvody proč provádět výzkum v umělé inteligenci a filozofické logice prostřednictvím formálního aparátu. Zaprvé to jsou *deskriptivní charakteristiky* takových zkoumání a požadavek, aby ona zkoumání byla dostatečně přesná a zřetelná. Zadruhé je to jistá normativnost zdůrazňovaná v tomto kontextu např. Donaldem Nutem [100]. To se týká naší snahy precizovat některé dosud jen intuitivně a nezřetelně strukturované, ale často spíše nedostatečně strukturované koncepty. Reformulace znalostí ve vhodném formálním prostředí je cestou, která poskytne odpověď na jednu z nejdůležitějších otázek o roli dedukce v racionálním usuzování. Chceme se vyrovnat s otázkou

- *Proč klasická deduktivní logika není dostatečně adekvátní pro porozumění každodennímu usuzování, tedy tomu, co obvykle Angličané označují termínem* commonsense reasoning*?*

Po více než dvě tisíciletí se logika zabývala převážně "věčnými pravdami" a nebrala vůbec do úvahy čas. Navíc tomistická absolutizace aristotelské logiky po celé milenium vytvářela bariéru pro porozumění dynamice znalostí, protože klasický princip dyadické pravdivostní funkcionality pokrývá pouze nepatrnou část logické povahy významu výrazů (formulí) a porozumění logické korektnosti argumentů. Odpověď na otázku, co je to *logicky korektní argument*, je ale stále předmětem zkoumání [73]. Filozofická reformulace, o kterou se zde budeme snažit, musí samozřejmě vzít v úvahu na jedné straně jazykové užití oněch konceptů a na straně druhé cíle, k nimž mají být relevantní důvody hledány.

1.1 Deduktivní usuzování

V tomto odstavci osvětlíme dnes už klasický tarskiánský koncept *logické konsekvence* definovaný nezávisle na vyjadřovacích charakteristikách jazyka. Připomeneme vztah logického důsledku a souvisejících metamatematických pojmů jako bezespornost, nezávislost, deduktivní uzavřenost, deduktivní úplnost apod.

Deduktivní usuzování je obvykle charakterizováno několika málo vlastnostmi [133], mezi něž patří např. i monotónnost, která může být v tradiční terminologii volně vyjádřena *pravidlem proporcionality*:

Více premis, více závěrů.

Ospravedlněním tohoto principu je skutečnost, že klasická logika se po více než dvě tisíciletí zabývala téměř výlučně již dříve zmíněnými "*věčnými pravdami*", zatímco epistemické stavy, tj. kontextově závislé znalosti, které se v čase mění, nepopisovala vůbec anebo nepopisovala adekvátně. V současné době se v rámci umělé inteligence, ale i v jiných oblastech lidské činnosti, kde informace a znalosti hrají důležitou roli, setkáváme se znalostmi, jejichž platnost je spíše dočasná. To se například jistě týká aktualizace znalostních databází a každodenního usuzování. V této souvislosti je zajímavá zkušenost z tvorby znalostních systémů, která ukazuje na to, že při zpracování vysoce odborných znalostí expertů jsme relativně úspěšní zatímco obtíže, které vznikají, chceme-li pochopit a efektivně reprezentovat logickou strukturu každodenních znalostí a inferencí, se zdají být téměř nepřekonatelné. Je zřejmé, že takový typ usuzování, který vystihuje dynamiku znalostí, nebývá z hlediska pravdivosti extenzionální a pravdivost složených tvrzení není funkcí pravdivostí jejich skladebných částí. To bude předmětem našich úvah ve třetí kapitole, ale k tomu, abychom plně pochopili tento typ argumentace, je třeba mít dostatečný základ klasické logické dedukce.

1.1.1 Operace logického důsledku

Nechť Fle je libovolná neprázdná množina. Prvky této množiny budeme nazývat formule, ale prozatím nebudeme předpokládat nic o struktuře této množiny ani o struktuře jejích prvků. Postupně budeme ovšem brát do úvahy další vlastnosti množiny formulí, tj. vlastnosti jazyka, jako např. vyjadřovací sílu jazyka, v němž jsou formule konstruovány. Budeme to ale činit jen v míře nezbytné pro naše cíle. Aristoteles nás naučil chápat tvrzení, která o světě vyslovujeme, jako soustavu důvodů a důsledků. Přitom pojem logického důsledku je zde primární, proto jím začneme. O důvodech bude pak řeč zejména v kapitole o abdukci.

Definice 1.1.1 Operaci logického důsledku *na dané množině* Fle *definujeme jako zobrazení*

$$Cn : \mathcal{P}(Fle) \longrightarrow \mathcal{P}(Fle),$$

které je reflexivní, monotónní a tranzitivní.[1]

To může být vyjádřeno následujícími podmínkami: Pro každé tři množiny formulí X, Y a $Z \subseteq Fle$ platí

- $X \subseteq Cn(X)$ (reflexivnost, idempotence),
- $X \subseteq Y$ implikuje $Cn(X) \subseteq Cn(Y)$ (monotónnost),
- $Cn(Cn(X)) \subseteq Cn(X)$ (tranzitivita).

Když $\alpha \in Cn(X)$, řekneme, že α *je důsledkem* množiny formulí X.

První podmínku (reflexivnost) můžeme chápat tak, že to, co předpokládáme, je odvoditelné. O monotónnosti jsme se už zmínili a tranzitivnost se v této souvislosti zdá být také přirozená

[1] Zde $\mathcal{P}(Fle)$ označuje tzv. potenční množinu množiny Fle, tj. množinu všech podmnožin množiny Fle.

a zřejmá. Říká totiž, že když už jsme jednou odvodili všechny důsledky z dané množiny X, tak každé další odvozování nepřinese již nic nového.

Přijmeme-li tyto tři požadavky, pak snadno ověříme, že pro každé X, Y a $Z \subseteq Fle$ bude platit:

1. $Cn(X) \subseteq Cn(Cn(X))$ [2].
2. $X \subseteq Cn(Y)$ implikuje $Cn(X) \subseteq Cn(Y)$.
3. $X \subseteq Cn(Y)$ a $Y \subseteq Cn(Z)$ implikuje $X \subseteq Cn(Z)$.
4. $X \subseteq Y \subseteq Cn(X)$ implikuje $Cn(X) = Cn(Y)$.
5. $Cn(X \cap Y)) = (Cn(X) \cap Cn(Y)) \subseteq Cn(X)$.
6. $Cn(X \cup Y) = Cn(Cn(X) \cup Y) = Cn(X \cup Cn((Y)) =$
   $= Cn(Cn(X) \cup Cn(Y))$.

Poznamenejme, že z tranzitivnosti operace konsekvence a z reflexivnosti vyplývá, že $Cn(Cn(X)) = Cn(X)$.

Připomeňme jet, že monotónnost operace logické konsekvence může být ekvivalentně vyjádřena podmínkou

$$Cn(X) \subseteq Cn(X \cup Y),$$

kde Y je libovolná množina formulí.

Dvěma hlavními příklady logické konsekvence jsou koncept syntaktického důsledku a koncept sémantického důsledku.

Syntaktický důsledek

Nechť Fle je množina dobře utvořených formulí daného jazyka, pro jednoduchost např. množina dobře utvořených formulí výrokového počtu, a nechť dále $Ax \subseteq Fle$ je množina axiomů a R je množina obvyklých odvozovacích pravidel. Potom zobrazení

[2] Tato vlastnost spolu s tranzitivitou dává rovnost $Cn(X) = Cn(Cn(X))$, což je obvyklejší.

$$c_1 : \mathcal{P}(Fle) \longrightarrow \mathcal{P}(Fle),$$

definované tak, že $c_1(X)$ je minimální množina, která obsahuje množinu X, axiomy Ax a která je navíc uzavřená na odvozovací pravidla z R, je operace konsekvence.

Tato definice je přitom dost univerzální v tom smyslu, že ji lze využít nejen pro jednoduché dvouhodnotové formule typu ano/ne (pravda/nepravda), ale i pro formule s více než dvěmi pravdivostními hodnotami, jak je dnes obvyklé např. při studiu tzv. fuzzy logik. Když totiž prvky množiny Fle budeme chápat jako dvojice tvaru

[syntaktický tvar formule, pravdivostní hodnota]

a pravidla odvozování jako $(n+1)$-tici dvojic výše uvedeného tvaru, tak naše definice logické konsekvence zůstane beze změny v platnosti.

Sémantický důsledek

Druhým významným příkladem je sématický koncept logického důsledku. Když například Fle je množina formulí jazyka prvního řádu a $\mathcal{M}$ je třída modelů pro Fle v obvyklém slova smyslu, pak operace c_2 definovaná tak, že $\varphi \in c_2(X)$, když každý model $m \in \mathcal{M}$, který je modelem každé formule $\psi \in X$ je také modelem formule φ.

Tyto příklady ukazují, že náš abstraktní koncept operároru logické konsekvence zahrnuje jak syntaktickou tak sémantickou stránku logického důsledku.

Následující dvě lemmata jsou jednoduchá, ale velmi důležitá. Druhé říká, že libovolná operace logického důsledku, která je restringovaná na pevně danou podmnožinu A množiny Fle, je opět operace konsekvence (na té restringované množině formulí A).

Lemma 1.1.1 *(Fundamentální lemma) Nechť $S \subseteq \mathcal{P}(Fle)$ je neprázdná třída množin. Potom zobrazení $f : \mathcal{P}(Fle) \longrightarrow \mathcal{P}(Fle)$*

definované tak, že $f(X) = \bigcap Y_\xi$ *pokud existuje aspoň jedno* $Y_\xi \in S$ *takové, že* $X \subseteq Y_\xi$ *a* $f(X) = Fle$ *jindy, je operace konsekvence na* Fle.

Lemma 1.1.2 *(Restrikční lemma) Nechť*

$$g : \mathcal{P}(Fle) \longrightarrow \mathcal{P}(Fle)$$

je takové zobrazení, že pro každé $X \subseteq A$ *platí* $g(X) = A \cap Cn(X)$, *kde* $A \subseteq Fle$. *Potom* g *je reflexivní, monotónní a tranzitivní.*

Důkazy obou lemmat se získají snadno přímým ověřením všech vlastností operace konsekvence pro zobrazení g i pro zobrazení f.

Fundamentální lemma vlastně říká, že operace, charakterizující logickou dedukci, může být určena systémem množin formulí, který pak pro libovolnou množinu vymezí její důsledky jako průnik všech jejích nadmnožin patřících právě do daného systému. To mj. využijeme v následujícím odstavci k tomu, abychom ukázali vzájemnou definovatelnost některých důžitých metamatematických pojmů jako je deduktivní uzavřenost a deduktivní úplnost.

Protože zobrazení f ve fundamentálním lemmatu je jednoznačně určeno množinou S, budeme takto definovanou operaci konsekvence označovat symbolem $Cn[S]$.

Definice 1.1.2 *Operaci konsekvence* Cn *na* Fle *nazveme* kompaktní *(nebo též* finitní*), když* $Cn(X) = \bigcup Cn(Y_\xi)$, *kde sjednocení je přes všechny konečné množiny* $Y_\xi \subseteq X$.

Jinými slovy, Cn je kompaktní operace konsekvence, jestliže pro libovolnou formuli $\alpha \in Cn(X)$ existuje konečná množina $Y \subseteq X$, že $\alpha \in Cn(Y)$.

Tato vlastnost může být zobecněna pro jazyky s nespočetně mnoha formulemi. Předpokládejme, že množina dobře utvořených formulí Fle má kardinalitu κ. Operaci konsekvence Cn nazveme κ-kompaktní, když $Cn(X)$ se získá jako sjednocení uzávěrů všech podmonožin $Y_\xi \subseteq X$ mohutnosti nejvýše κ.

Definice 1.1.3 *Množina formulí $X \subseteq Fle$ se nazývá* bezesporná*, nebo též* konzistentní*, (vzhledem k operaci Cn), jestliže*

$$Cn(X) \neq Fle.$$

Jinak se nazývá sporná.

Tato definice, původně navržená Emilem Léonem Postem, má oproti jiným definicím (např., že $Cn(X)$ obsahuje nějakou formuli spolu s její negací) tu přednost, že je nezávislá na zvoleném jazyce. Speciálně nevyžaduje koncept negace.

Definice 1.1.4 *Množina formulí X se nazývá* deduktivně uzavřená *(vzhledem k operaci Cn), když $Cn(X) \subseteq X$ a nazývá se* úplná*, je-li to maximální bezesporná množina (vzhledem k Cn), tj. když každá její nadmnožina je sporná.*

Fakt, že množina formulí X je vůči operaci Cn bezesporná, deduktivně úplná, resp. deduktivně uzavřená, označíme po řadě symboly $cons(X)$, $compl(X)$ a $closed(X)$.

Připomeneme ještě dobře známá fakta: Pro danou operaci konsekvence Cn na Fle a pro každé dvě množiny formulí $X, Y \subseteq F$ platí

1. Množina X je deduktivně uzavřená, tj. $closed(X)$, právě tehdy, když $Cn(X) = X$.

2. Pro každou množinu formulí X platí $closed(Cn(X))$. Navíc $Cn(X)$ je nejmenší deduktivně uzavřená množina obsahující X.

3. Jestliže $closed(X)$ a $closed(Y)$, potom i $closed(X \cap Y)$. Pro sjednocení to ovšem obecně neplatí.

Definice 1.1.5 *Nechť $S \subseteq \mathcal{P}(Fle)$ je libovolná třída množin formulí. Symbolem σ_{Cn} označíme třídu všech množin uzavřených vzhledem k operaci Cn a symbolem $\theta(S)$ třídu všech množin $Z \in S$ maximálních vzhledem k inkluzi restringované na S. Třídu S nazveme* regulární, *když pro každou množinu formulí $X \subseteq Fle$ a pro každou formuli $\alpha \in Fle$ platí, že $\alpha \in Cn[S](X)$ nebo když existuje $Y \in \theta(S)$ taková, že $X \subseteq Y$ a $\alpha \notin Y$.*

Tyto definice nám nyní umožňují vyslovit následující teorém.

Teorém 1.1.1 *(O vzájemné definovatelnosti) Pro každou operaci konsekvence Cn platí $Cn = Cn[\sigma_{Cn}]$. Navíc, kdykoli $\theta(\sigma_{Cn})$ je regulární, tak $Cn[\sigma_{Cn}] = Cn[\theta(\sigma_{Cn})]$.*

Důkaz získáme z vlastností množin σ_{Cn} a $\theta(\sigma_{Cn})$. Viz např. [58].

Z předchozího teorému pak bezprostředně vyplývá, že za podmínky regularity množina všech maximálních konzistentních množin formulí jednoznačně určuje operaci konsekvence na Fle.

1.1.2 Relace logického důsledku

Logická konsekvence může být studována možná přirozenějším způsobem též jako *relace* $\vdash$ mezi formulemi (přesněji mezi množinou formulí a formulí, kterou označujeme jako její důsledek), která splňuje následující podmínky:

1. reflexivnost
   $X \vdash \varphi$ kdykoli $\varphi \in X$.

2. monotónnost
   Jestliže $X \vdash \varphi$ a $X \subseteq Y$, pak $Y \vdash \varphi$.

3. tranzitivnost (řez)
   $X \vdash \varphi$ právě tehdy, když $X \cup \{\psi_i; i \in I \text{ a } X \vdash \psi_i\} \vdash \varphi$.
   Ve finitistické verzi:
   $X \vdash \psi$ a $X \cup \{\psi\} \vdash \varphi$ právě tehdy, když $X \vdash \varphi$.

4. kompaktnost
   $X \vdash \varphi$ právě když $X_0 \vdash \varphi$ pro nějaké konečné $X_0 \subseteq X$.
   Ve finitistické verzi:
   $X \vdash \varphi$ právě tehdy, když $X \vdash \psi$ a $X \cup \{\psi\} \vdash \varphi$.

To je gentzenovský styl definování relace logického důsledku.

Výše uvedené podmínky nejsou nezávislé. Snadno nahlédneme, že například z kompaktnosti již plyne monotónnost. Ta bude hrát významnou roli v dalších kapitolách, kde právě monotónnost bude pod naším drobnohledem.

Tento druh formálního popisu dedukce ale není jediný možný. Jsou užívány i jiné popisy, které jsou ovšem *de facto* variantami právě zavedeného konceptu. Pro úplnost bychom měli uvést, že např. David Hilbert chápal dedukci jako relaci množiny axiomů a formule, která je odvozována z prázdné množiny předpokladů. Pak hovoříme o hilbertovských kalkulech.

1.2 Abstraktní operátory logické konsekvence

V tomto oddíle budeme analyzovat abstraktní operace [3] na množině Fle, tj. zobrazení z $\mathcal{P}(Fle)$ do $\mathcal{P}(Fle)$. Začneme s monotónními operátory, ale pro potřeby výkladu nemonotónní inference v kap. 3 a dalších budeme analyzovat nemonotónní operátory a budeme je konfrontovat s takovými vlastnostmi jako kompaktnost (nebo κ-kompaktnost), reflexivnost, atd.

[3] Takové operace také nazýváme *operátory* na Fle.

Nechť Fle je neprázdná množina formulí daného jazyka.

Definice 1.2.1 Operátor f na Fle *je zobrazení*

$$f : \mathcal{P}(Fle) \longrightarrow \mathcal{P}(Fle).$$

Tento obecný koncept operátoru na množině Fle souvisí s dobře známým pojmem obecné topologie na dané množině. To nám umožní nalézt další užitečné analogie, musíme ale přitom být opatrní při využití výsledků získaných ze znalostí topologických uzávěrových operátorů, protože aditivnost, tak důležitá v topologii, zde obecně neplatí, a tudíž ty topologické výsledky, které se o ni opírají, musí být znovu ověřovány. Na druhé straně, řadu z topologických zkoumání s úspěchem využijeme. Například pojem *okolí*, jak je užíván v hausdorfovské tradici, může být přirozeným způsobem definován pro obecné operátory.[4]

Definice 1.2.2 *Nechť* f *je libovolný operátor na* Fle. *Množina* $N \in \mathcal{P}(Fle)$ *se nazývá* okolí *množiny* $X \in \mathcal{P}(Fle)$ *jestliže* $X \cap f(Fle - N) = \emptyset$.

Definice 1.2.3 Horní iterativní sekvence *pro* f *začínající v* $X \subseteq Fle$ *je definována následovně*

$f \uparrow 0(X) = X$

$f \uparrow (\alpha + 1)(X) = f(f \uparrow \alpha(X))$

$f \uparrow \lambda(X) = \bigcup\{f \uparrow \alpha(X) : \alpha < \lambda\}$ *pro* λ *limitní.*

Dolní iterativní sekvence *pro* f *začínající v* $X \subseteq Fle$ *je pak definována takto*

$f \downarrow 0(X) = X$

$f \downarrow (\alpha + 1)(X) = f(f \downarrow \alpha(X))$

$f \downarrow \lambda(X) = \bigcap\{f \downarrow \alpha(X) : \alpha < \lambda\}$ *pro* λ *limitní.*

[4] Jiná rozumná možnost, jak zobecnit pojem operátoru, spočívá v definici operátoru na částečně uspořádané množině (poset - partially ordered set).

Definice 1.2.4 *Operátor f na Fle nazveme* monotónní*, jestliže pro všechny množiny $X, Y \subseteq Fle$ takové, že $X \subseteq Y$ platí $f(X) \subseteq f(Y)$. A nazývá se* kompaktní*, když pro každé $X \subseteq Fle$,*

$$f(X) = \bigcup\{f(Y) : Y \subseteq X, \ a\ Y\ je\ konečná\}.$$

Je zřejmé, že každá operace logické konsekvence na Fle je monotónní operátor na Fle.

Zde je jednoduchý příklad: Nechť $Fle = \omega$ je množina všech přirozených čísel a nechť $f(X) = \{\mathcal{P}(X) : X \subseteq F\}$. Potom f je monotónní a kompaktní operátor na Fle a $\emptyset$ je jeho jediný pevný bod, tj. platí

$$f(\emptyset) = \emptyset.$$

Následující tvrzení je dobře známá charakterizace monotónních operátorů, kterou podali Tarski a Knaster.

Teorém 1.2.1 *Nechť f je monotónní operátor na množině Fle. Potom f má na Fle nejmenší pevný bod (fixpunkt). Jestliže navíc f je kompaktní, tak ten nejmenší fixpunkt operátoru f je roven $f \uparrow \omega(\emptyset)$.*

1.2.1 Operátory založené na pravidlech

Výroková logika (PC - *propositional calculus*) je vhodným příkladem formálního systému dokazování založeného na pravidlech, který je generován množinou axiomů a množinou odvozovacích pravidel. Budeme zkoumat, co se stane, když takový systém rozšíříme o další pravidla. Pravidla budeme yapisovat ve tvaru $(\alpha_1, \ldots, \alpha_n/\gamma)$, kde $(\alpha_1, \ldots, \alpha_n$ jsou premisy a $\gamma)$ je závěr pravidla.

Nechť R je libovolná konečná množina (možná dokonce i sporná) odvozovacích pravidel výrokového počtu. Symbolem $Cn^R(X)$

označíme množinu všech formulí odvoditelných z množiny $X \subseteq Fle$ ve formálním systému PC s přidanými pravidly R.

Potom pro každé $X \subseteq Fle$ platí
$Cn(X) \subseteq Cn^R(X)$, a $Cn^R(Cn^R(X)) = Cn^R(X)$.

Teorém 1.2.2 *Pro každé X, množina $Cn^R(X)$ je nejmenší množina obsahující X, která je uzavřená na všechna pravidla PC a všechna pravidla z R.*

Definice 1.2.5 *Nechť R je množina odvozovacích pravidel. Pro každé $X \subseteq Fle$ definujeme operátor*

$$\Gamma^R(X) = Cn(X \cup \{\gamma : (\alpha/\gamma) \in R \text{ a } \alpha \in X\}) \ .$$

Teorém 1.2.3 *Nechť R je konečná množina pravidel. Operátor Γ^R je reflexivní (někdy též nazývaný progresivní,* srovnej [86]*), monotónní a kompaktní.*

Teorém 1.2.4 *Pro každé $X \subseteq F$ platí $Cn^R(X) = \Gamma^R \uparrow \omega(X)$.*

To znamená, že množina $Cn^R(X)$ je nejmenší pevný bod operátoru Γ^R obsahující množinu X.

Zde je přímá souvislost s korektními logickými pravidly (vzhledem k PC). Když R je množina korektních pravidel, tak pro každou množinu X, $Cn(X) = Cn^R(X)$ a pro každou konzistentní množinu X je $Cn^R(X)$ konzistentní. Na druhé straně, když R obsahuje nekorektní pravidlo r, pak existuje taková konzistentní množina X, že $Cn^{\{r\}}(X)$ je sporná. [5] To je zřejmý a velmi významný fakt, který říká, že pro libovolnou množinu X formulí a množinu R pravidel, která obsahuje logicky nekorektní pravidlo, konzistenci nelze zachovat. To ale ještě neznamená, že konzistence nemůže být udržena, omezíme-li se na vybranou třídu

[5] Pro jazyky s konjunkcí lze bez újmy na obecnosti snadno dokázat, že stačí, když pravidlo $r = (\varphi/\psi)$ obsahuje pouze jednu premisu.

teorií. Pak můžeme získat systém, který je silnější než klasický PC. Jedna z těchto možností bude diskutována v kap. 4.

1.3 Svaz operací konsekvence

Teď popíšeme algebraickou strukturu všech možných operací logické konsekvence na dané množině formulí. Pro ten účel zavedeme ještě několik pomocných pojmů včetně báze operace konsekvence.

Nechť Fle je opět pevně daná množina dobře utvořených formulí nějakého jazyka. V dalším bude $\mathcal{C}_{Fle}$ označovat třídu všech operací logické konsekvence na Fle. Nepůjde tedy teď o všechny abstraktní operátory, ale pouze o ty, které jsou reflexivní, monotónní a tranzitivní.

Řekneme, že operátor Cn_1 je deduktivně silnější než operátor Cn_2, jestliže pro každou množinu $X \subseteq Fle$ platí $Cn_2(X) \subseteq Cn_1(X)$. V takovém případě píšeme $Cn_1 \succeq Cn_2$. Je zřejmé, že relace $\succeq$ je částečné uspořádání na $\mathcal{C}_{Fle}$. Je také zřejmé, že operace definovaná vztahem $Cn_\top(X) = Fle$ pro každé $X \subseteq F$ je nejsilnější a $Cn_\perp(X) = X$ je nejslabší operace konsekvence na Fle.

1.3.1 Báze operace konsekvence

Definice 1.3.1 *Třída množin $S \subseteq \mathcal{P}(Fle)$ se nazývá* báze pro *Cn, jestliže pro každou množinu $X \subseteq Fle$ platí*

$$Cn(X) = Cn[S](X).^6$$

Dvě báze S a S' jsou ekvivalentní, jestliže generují tutéž operaci konsekvence na Fle.

[6] Připomeňme, že $Cn[S]$ označuje operaci konsekvence generovanou třídou S, takže $Cn[S](X)$ je uzávěr množiny X vzhledem k $Cn[S]$.

Nechť $Int(S)$ označuje uzávěr systému S na průniky. Potom S a $Int(S)$ jsou ekvivalentní báze pro Cn.

Lemma 1.3.1 *Nechť $S \subseteq \mathcal{P}(Fle)$. Potom $\sigma_{Cn[S]} = Int(S)$.*

Důkaz: $Cn[S](X)$ je sjednocením všech nadmnožin Y_ξ množiny Z, které zároveň patří do systému S. Teď je třeba rozebrat dva případy. Zaprvé, když X je v $Int(S)$, tak $Cn[S](X) = X$ a tudíž $X = \sigma_{Cn[S]}$. Odtud pak dostáváme, že $Int(X) \subseteq \sigma_{Cn[S]}$. Zadruhé, když $X \in \sigma_{Cn[S]}$, tak potom opět $Cn[S](X) = X$. To je možné pouze tehdy, když $X \in S$ nebo když X je průnikem množin z S, tj. když $X \in Int(S)$ a tudíž $\sigma_{Cn[S]} \subseteq Int(S)$.

Toto lemma spolu s teorémem o vzájemné definovatelnosti nám umožňují vyslovit následující tvrzení.

Teorém 1.3.1 *Nechť $S \subseteq \mathcal{P}(Fle)$ a nechť $Cn[S]$ je operace konsekvence s bází S. Potom $Int(S)$ je (ve smyslu inkluze) největší báze ekvivalentní S. Jestliže navíc S je regulární, tak $\theta(S)$ je nejmenší báze ekvivalentní S.*

Teorém 1.3.2 *Třída $\mathcal{C}_{Fle}$ všech operací konsekvence na Fle spolu s částečným uspořádáním $\succeq$ určeným deduktivní silou je úplný distributivní svaz.*

Důkaz Nejprve se dohodneme na několika označeních a ve shodě s předcházejícím bude symbol σ^X_{Cn} označovat třídu všech nadmnožin množiny X, které jsou uzavřené vůči Cn. Nechť $S^X_{uni} = \bigcup \sigma^X_{Cn}$ a $S^X_{int} = \bigcap \sigma^X_{Cn}$. Potom můžeme definovat dvě nekonečné operace *spojení* $\sqcup$ a *průseku* $\sqcap$ tak, že

$$(\sqcap Cn)(X) = \bigcap S^X_{\cup},$$
$$(\sqcup Cn)(X) = \bigcap S^X_{\cap}.$$

Není těžké ověřit, že jsme tímto definovali opět dvě operace konsekvence. Zřejmě pro libovolné dva systémy S_1, $S_2 \subseteq \mathcal{P}(Fle)$ platí, že $S_1 \subseteq S_2$ právě tehdy, když $Cn[S_2] \preceq Cn[S_1]$. Takže můžeme uzavřít, že $\sqcap$ je nejslabší a $\sqcup$ je nejsilnější operace konsekvence na Fle. Můžeme také psát

$$\sqcap = inf\{Cn \ ; \ Cn \in \mathcal{C}_{Fle}\},$$
$$\sqcup = sup\{Cn \ ; \ Cn \in \mathcal{C}_{Fle}\}.$$

Distributivita plyne z vlastností operací množinového sjednocení a průniku.

Definice 1.3.2 *Řekneme, že třída množin $S \subseteq \mathcal{P}(Fle)$ je κ-usměrněná, jestliže pro každou třídu $S' \subseteq S$ takovou, že*

$$card(S') \leq \kappa$$

existuje $s \in S$ tak, že $s' \subseteq s$.

Lemma 1.3.2 *Nechť $card(Fle) \geq \kappa$ a nechť Cn je libovolná operace konsekvence na Fle, potom existuje aspoň jedna κ-usměrněná třída $S \subseteq \sigma_{Cn}$.*

Lemma 1.3.3 *Nechť $S_1, S_2 \subseteq \mathcal{P}(Fle)$ jsou κ-induktivní třídy formulí. Potom i $S_1 \cap S_2$ je κ-induktivní.*

Lemma 1.3.4 *Nechť $\kappa \geq 2$ je regulární číslo. Potom Cn je κ-kompaktní právě tehdy, když σ_{Cn} je κ-induktivní.*

Lemma 1.3.5 *Nechť $\kappa \geq 2$ je regulární číslo. Potom jak průsek tak i spojení dvou κ-kompaktních operací konsekvence jsou opět κ-kompaktní operace konsekvence.*

Z předcházejících lemmat už snadno odvodíme následující tvrzení.

Teorém 1.3.3 *Třída všech κ-kompaktních operací konsekvence na Fle je úplný podsvaz svazu všech operací konsekvence na Fle.*

Jestliže zkoumáme libovolnou třídu operací logické konsekvence, pak tento teorém nám umožňuje soustředit pozornost právě jen na horní a dolní závoru této množiny operací. Wójcicki [142] ukázal, že analogicky i strukturální[7] operace konsekvence tvoří úplný podsvaz.

[7]Operace konsekvence se nazývá strukturální, když pro všechny substituce e platí $eCn(X) \subseteq Cn(eX)$.

Kapitola 2

Automatizace logické dedukce

Problém automatizace logického usuzování je obvykle chápán jako otázka, zda proces deduktivního usuzování může být automatizován, tj. zda odvozování může provádět stroj. V širokém slova smyslu je to úloha, která má různé podoby a byla ve dvacátém století podrobně zkoumána např. v rámci teorií prvního řádu. Otázku, zda (logické) úsudky mají povahu "vykalkulovatelných" závěrů si už kladli mnozí filozofové a matematici mezi nimiž je třeba jmenovat především Gottfrieda Wilhelma Leibnize (1646–1716), ale naši pozornost si zaslouží i méně známý Raymundus Lullus (asi 1235–1315), vlastním jménem Menendéz y Pelao, který se při svých misijních cestách z Mallorky do Palestiny snažil nalézt nejobecnější principy vědy pomocí důmyslného mechanického zařízení, jakéhosi logického stroje. Leibnizovou snahou zase bylo nahlížet na veškerou vědu jako na kalkul. Byl zakladatelem logicismu a lze ho považovat za předchůdce formální logiky v dnešním slova smyslu právě s ohledem na důraz, který kladl na aspekty kalkulovatelnosti.

V užším slova smyslu je to otázka, zda věty vyjádřené v jazyce prvního řádu mohou být dokazovány automaticky strojem, např. počítačovými programy. Pro porozumění dedukci je odpověď na tuto otázku jednou z nejdůležitějších v kontextu klasické Turingovy otázky [138], zda stroje mohou myslet.

V této kapitole podáme výklad těch základních pojmů, které pak v dalším budeme potřebovat pro zkoumání odvození, která jsou sice velmi užitečná, ale nejsou striktně deduktivní.

2.1 Herbrandova věta

Standardní odvozovací pravidla logické dedukce (ať už ve formulaci Gentzenově či Hilbertově) nejsou vhodná pro automatizaci důkazů. Od třicátých let logikové znají metodu *zamítání* (angl. *refutation method*), která je založena na významném výsledku [53], kterého dosáhl v roce 1929 mladý francouzský logik Jacques Herbrand. Tento výsledek je na jeho počest uváděn pod názvem Herbrandova věta. V kontextu automatického dokazování dává Herbrandova věta do souvislosti syntaktický koncept dokazatelnosti a sémantický koncept splnitelnosti (resp. nesplnitelnosti) některých teorií.

Především si uvědomíme, že místo toho, abychom přímo hledali důkaz formule φ v teorii T, převedeme úlohu na zkoumání nesplnitelnosti pozměněné teorie, přesněji teorie, která vznikne z původní teorie rozšířením o negaci dokazované formule. To lze vyjádřit následovně:

Teorém 2.1.1 *Formule φ prvního řádu logicky vyplývá z teorie T právě tehdy, když teorie $T \cup \{\neg\varphi\}$ je nesplnitelná.*

Metoda dokazování je tak založena na tom, že místo toho, abychom hledali přímý důkaz, snažíme se vyvrátit negaci dokazované formule.

Pro ten účel nejprve dokazovanou formuli upravíme ekvivalentně do podoby vhodné pro stroj. Formuli v takovém tvaru budeme nazývat klauzulí (*clausal form*).

Definice 2.1.1 Klauzule *je literál nebo disjunkce literálů nebo prázdný výraz.* Literál *je atomická formule nebo negace atomické formule.*

Je žádoucí znát odpověď na otázku, zda je možné ke každé formuli (a stačí umět to pro uzavřené formule neboli sentence) nalézt ekvivalentní formuli v klauzulárním tvaru. Na tyto otázky dává odpověď následující lemma.

Lemma 2.1.1 *Každá formule jazyka prvního řádu může být ekvivalentně přepsána na formuli prvního řádu, která je v klauzulárním tvaru.*

Postup, jak převést formuli do klauzulárního tvaru lze popsat následovně:

1. Nejprve nalezneme prenexní normální tvar dané formule.
2. Eliminujeme redundantní kvantifikátory.
3. Eliminujeme implikace a ekvivalence tím, že je nahradíme pomocí negací, konjunkcí a disjunkcí.
4. Negace přesouváme vpravo a kvantifikátory vlevo.
5. Skolemizací eliminujeme všechny existenční kvantifikátory. (Formule $\exists x \forall y P(x, y)$ může být nahrazena formulí $\forall y P(a, y)$ a formule $\forall y \exists x P(x, y)$ může být nahrazena $\forall y P(f(y), y)$. Zde Skolemova funkce přímo reprezentuje individuum zmiňované v původní formuli.)

6. A konečně, eliminujeme všechny obecné kvantifikátory.

Nyní můžeme pro danou množinu K klauzulí definovat *Herbrandovo univerzum* $\mathcal{H}(K)$, což je množina výrazů (termů) jazyka množiny K, které splňují dvě následující podmínky:

1. Do univerza $\mathcal{H}(K)$ patří všechny individuální konstanty z jazyka klazulí K.
2. Pro každou n–ární funkci $f \in K$ také term $f(t_1, \ldots, t_n)$ patří do $\mathcal{H}(K)$.
3. Množina $\mathcal{H}(\mathcal{K})$ je nejmenší množina splňující podmínky 1 a 2.

Příklad: Formule $\exists x \forall y P(x, y)$ může být ekvivalentně přepsána na formuli $\forall y P(a, y)$ zavedením nové konstanty a do jazyka, o nějž nám jde, zatímco formule $\forall y \exists x P(x, y)$ může být ekvivalentně přepsána na formuli $\forall y P(f(y)), y)$, kde $f(y)$ je Skolemova funkce reprezentující přímo objekty, jejichž existence je garantována původní formulí. Takto mohou být postupně eliminovány všechny existenční kvantifikátory.

Teď už jsme připraveni vyslovit Herbrandovu větu.

Teorém 2.1.2 *(Herbrandova věta) Množina klauzulí je nesplnitelná právě tehdy, když existuje konečná množina jejích základních (angl.* ground*) klauzulí, která je vÅrokov logicky sporná.*

Příklad: Nechť teorie $T = \{\forall x(P(x) \Rightarrow Q(x)), \forall y(P(f(y)))\}$ a nechť je dána formule $\varphi = \forall y(Q(f(y)))$. Otázkou je, zda platí $T \vdash \varphi$?

K tomu je třeba dokázat nesplnitelnost množiny

$$K = \{\forall x(P(x) \Rightarrow Q(x)), \forall y(P(f(y))), \neg \forall y Q(f(y))\}.$$

Třetí formuli z množiny K klauzulí můžeme ekvivalentně přepsat na formuli $\exists y \neg Q(f(y))$ a tu dále, použitím Skolemovy konstanty přepíšeme na formuli $\neg Q(f(a))$. Klauzulární tvar množiny K pak bude následující

$$\{\neg P(x) \vee Q(x), P(f(y)), \neg Q(f(a))\ \}$$

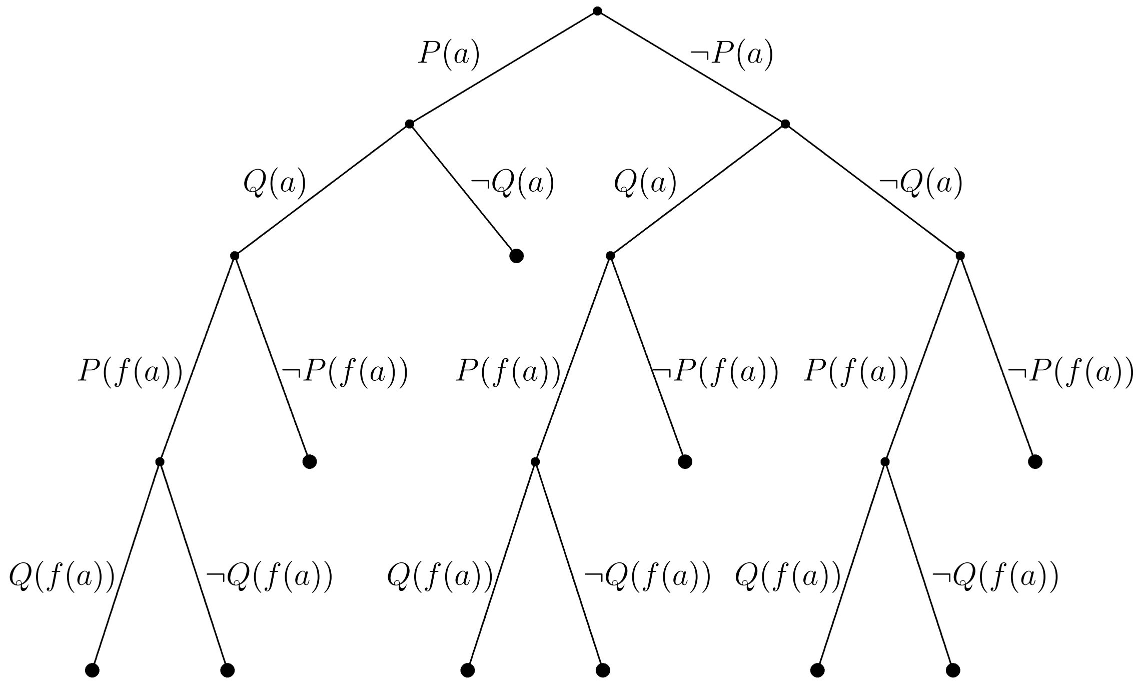
a Herbrandovým univerzem bude nekonečná množina obsahující termy $\{a, f(a), f(f(a)), \ldots\}$. Z Herbrandova univerza množiny klauzulí K a z predikátů P, Q užitých ve formulích množiny K, vytvoříme *sémantický strom* pro množinu K (Viz obr. 2.1.1.)

Jeho konstrukce je jednoduchá. Vezmeme první predikát v pořadí (v našem případě je to predikát P a první term Herbrandova univerza (v našem případě je to konstanta a) a ptáme se, zda formule $P(a)$ anebo formule $\neg P(a)$ je splnitelná vzhledem k množině klauzulí K. Pokud ano, prodloužíme danou větev sémantického stromu o další atomickou formuli, tentokrát utvořenou z termu a a dalšího predikátu v pořadí (v našem případě Q) a opět se ptáme, zda takto rozšířená množina formulí dává s množinou K klauzulí spor či nikoli. Herbrandova věta nám zaručuje, že pokud se na každé větvi takto konstruovaného stromu objeví spor, je množina K formulí nesplnitelná.

Avšak takový strom bývá často velmi rozsáhlý, což je přirozený důvod k tomu, abychom hledali výpočtově realističtejší postup. Navíc v případě, že množina klauzulí sporná není, v sémantickém stromu se objeví nekonečná větev a procedura se nezastaví.

V šedesátých letech, Julia A. Robinson [117] vyvinul nový postup, který je obecně znám pod názvem *rezoluční metoda* [1]. Ve stručnosti jej popíšeme v následujícím odstavci.

[1]Název metody je odvozen z angl. *resolution*, což zde lze přeložit slovy *nové řešení*.



Obr. 2.1.1: Sémantický strom pro množinu klauzulí K

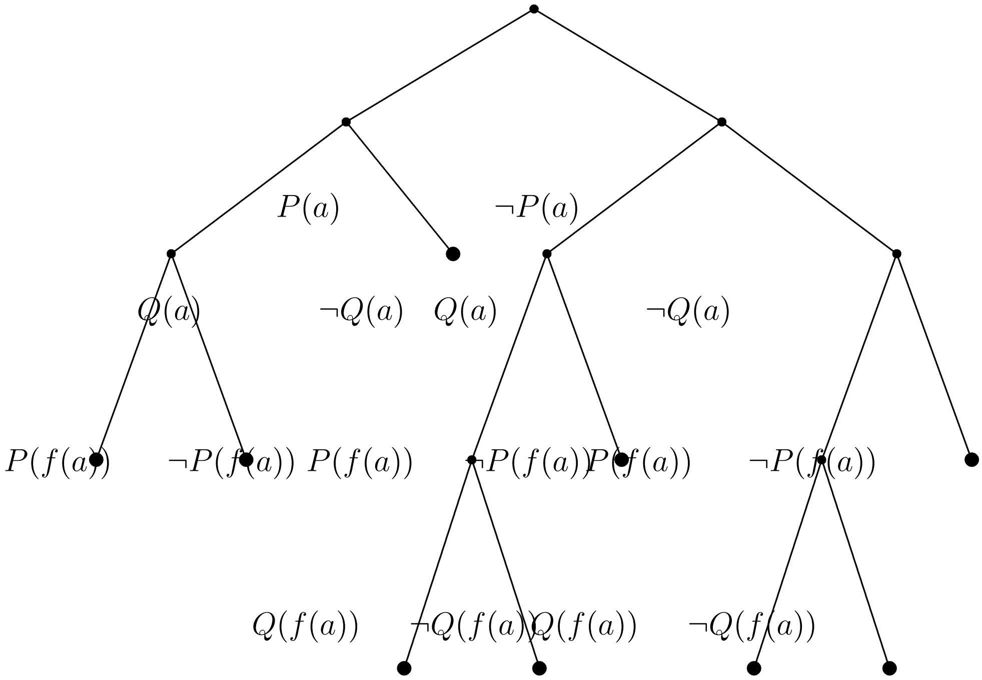
2.2 Robinsonův rezoluční princip

Idea, jak urychlit dokazování formulí pomocí rezoluce je tato: V procesu dokazování můžeme do úvahy brát právě odvozené formule, které přidáváme k původním klauzulím a využíváme je s výhodou při dalším odvozování. Samo odvozování se pak děje na základě nového odvozovacího pravidla. Ukážeme to nejprve na jednoduchém přikladě (cf. [25], [63]).

Předpokládejme, že množina klauzulí $\{\neg P(x), Q(x)\}$ je splněna interpretací i. Potom v téže interpretaci je splněna i každá instance každé klauzule, speciálně tedy i klauzule

$$\{\neg P(f(a)), Q(f(a))\}.$$

Avšak literál $Q(f(a))$ v takovém případě nemůže být splněn současně se třetí klauzulí. To je dobrým důvodem k tomu, abychom tento literál přidali k původní množině klauzulí a opět vytvořili sémantický strom (viz obr. 2.2.2.).



Obr. 2.2.2: Nový sémantický strom

Obr. 2.2.2 Nový sémantický strom pro K.

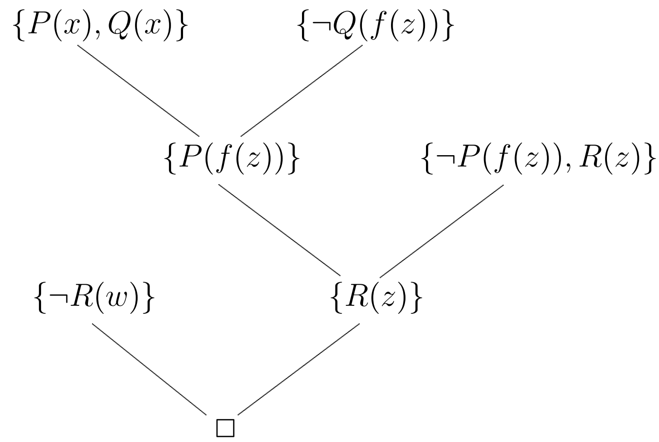
Ze stromu na obr. 2.2.2 je ihned vidět, že spor se objevil o jednu úroveň dříve než v předchozím případě.

Odvození lze znázornit schematem, které nazýváme rezoluční pravidlo:

$$\frac{\neg P(x) \vee Q(x), \quad \neg Q(f(a))}{\neg P(f(a))}$$

Klauzuli $\neg P(f(a))$ zde nazýváme *rezolventou* dvou horních klauzulí, obě premisy nazýváme *rodičovskými klauzulemi.* Rezoluční pravidlo tak kombinuje *substituci*, pravidlo *modus ponens* a různé druhy tautologií.

Příklad Množina následujících čtyř klauzulí
$P(x) \vee Q(x)$,
$\neg Q(f(z))$,
$\neg P(f(z)) \vee R(z)$,
$\neg R(w)$
je nesplnitelná, jak je patrno z odvození na obr. 2.2.3.



Obr. 2.2.3: Rezoluční strategie

2.3 Logické programy

Již delší čas je logikům známo, že v obecném případě není možné zkonstruovat algoritmus, který by pro libovolnou množinu formulí K a danou formuli φ prvního řádu rozhodl zda φ je z množiny K dokazatelná či nikoli, tj. zda platí $K \vdash \varphi$. To je dobře známý výsledek Alonzo Churche dosažený nejprve v elementární teorii čísel v roce 1936 a pak též v predikátové logice 1956 [24]. Tento sice negativní důsledek ale ještě neznamená, že bychom měli snah po automatizaci deduktivního dokazování zanechat. Spíše to znamená, že cesta jak dokazovat s podporou stroje nebude zvládnuta snadno jedním univerzální algoritmem, ale že bude nutné hledat užší třídy úloh, které možná takto zvládnutelné budou. Může se totiž ukázat, že omezíme-li se na formule předem vymezeného typu, bude možné dokazovací postupy přeci jen přenechat stroji, tedy automatizovat. V šedesátých a sedmdesátých letech minulého století byla tato možnost detailně zkoumána a výsledky vedly k tomu, že predikátová logika restringovaná na formule tzv. Hornovy logiky má tu žádanou vlastnost rozhodnutelnosti. To byl základ velkého rozmachu logických programů. Hornova logika tvoří velmi zajímavý fragment predikátové logiky, který, i když je omezením, dává stále ještě prostor značné vyjadřovací síly. Znalosti vyjádřené v rámci Hornovy logiky pak mají tu přednost, že odvození mohou být plně automatizována.

2.3.1 Hornovy klauzule

Zajímavou podtřídu třídy všech klauzulí tvoří klauzule speciálního typu, které se dají zapsat jako formule tvaru

$$\varphi_1 \wedge \ldots \wedge \varphi_n \Rightarrow \psi.$$

Třída takových klauzulí je zajímavá právě tím, že je rozhodnutelná a hodí se tudíž pro automatické dokazování.

Definice 2.3.1 *Klauzule, která obsahuje nejvýše jeden pozitivní literál, se nazývá* Hornova klauzule*.*

Je vhodné rozlišit čtyři druhy Hornových klauzulí: Jednotkové klauzule jsou ty, které neobsahují žádný negativní literál, programové (někdy též podmíněné) klauzule jsou ty, které obsahují pozitivní a aspoň jeden negativní literál. Cílové klauzule jsou ty, které obsahují pouze negativní literály. A konečně prázdné klauzule neobsahují ani pozitivní ani negativní literál a tudíž vyjadřují spor.

Definice 2.3.2 Logickým programem *nazýváme libovolnou neprázdnou množinu Hornových klauzulí.*

2.3.2 Substituce a unifikace

Definice 2.3.3 *Substituce θ je konečná množina tvaru*

$$\{v_1/t_1, \ldots, v_n/t_n\},$$

kde každé v_i $(i = 1, \ldots, n)$ je proměnná a každé t_i je term.

Substituce je tedy zobrazení z množiny proměnných do množiny termů jazyka. Poznamenejme ještě, že vzhledem k tomu, že každá proměnná je term, můžeme přejmenování proměnných považovat též za substituci.

Definice 2.3.4 Výraz *je term, literál, nebo konjunkce anebo disjunkce literálů.* Jednoduchý výraz *je term nebo atom.* Instancí výrazu *rozumíme výraz, na který byla aplikována substituce. Instance výrazu se nazývá* základní *(ground), neobsahuje-li žádné proměnné.*

Substituce je možné obvyklým způsobem skládat. Všimněme si, že skládání substitucí není komutativní, což znamená, že záleží na pořadí v jakém substituce provádíme. *Identická substituce* ι je taková substituce, že aplikována na daný výraz vydá opět ten původní výraz. Někdy se nazývá přejmenování proměnných.

Teď můžeme definovat unifikátor pro konečnou množinu jednoduchých výrazů.

Definice 2.3.5 *Substituce σ je* unifikátorem pro množinu S *jestliže $S\sigma$ je jednoprvková množina.*

Unifikátor σ pro S se nazývá nejobecnější unifikátor pro S, jestliže pro každý unifikátor θ pro S existuje taková substituce γ, že $\theta = \sigma\gamma$.

Příklad: Když σ_1, σ_2 jsou nejobecnější unifikátory pro množinu výrazů $\{e_1, \ldots, e_n\}$, tak $e_i\sigma_1$ je variantou $e_i\sigma_2$. To znamená, že výraz $e\sigma_1$ může být získán z $e\sigma_2$ pouhým přejmenováním proměnných. To mj. znamená, že nejobecnější unifikátory jsou až na přejmenování proměnných určeny jednoznačně.

Unifikační algoritmus

1. Polož $k = 0$ a $\sigma_0 = \epsilon$.

2. Jestliže $S\sigma_k$ je jednoprvková množina, tak STOP a σ_k je $nu(S)$. Jinak nalezni množinu neshod D_k pro množinu $S\sigma_k$.

3. Jestliže v D_k existují takové proměnné v a t, že v se nevyskytuje v t, polož $\sigma_{k+1} = \sigma_k\{v/t\}$. Polož $k = k + 1$. Go to 2. Jinak STOP; S není unifikovatelná.

Množina neshod v množině výrazů se získá tak, že nalezneme první pozici zleva, v níž neshoda nastala a z každého výrazu v

množině vyčleníme podvýraz začínající symbolem na této pozici. Množina všech takových podvýrazů je množina neshod.
Příklad: $S = \{P(f(x), h(y), a), P(f(x), z, a), P(f(x), h(y), b)\}$. Množinou neshod je zde množina $\{h(y), z\}$.

Teorém 2.3.1 *(Unifikační teorém) Pro libovolnou konečnou množinu V jednoduchých výrazů se unifikační algoritmus zastaví když V je unifikoavatelná a vydá nejobecnější unifikátor. V případě, že V není unifikovatelná, algoritmus se také zastaví a vydá o tom zprávu.*

Důkaz. Viz. např. [82].

Příklady: Nechť u, w, x, y, z jsou proměnné, a, b jsou konstanty, f, g jsou jednoargumentové funkční symboly a t je term. Potom následující dvojice výrazů se unifikují takto:

1. $P(x, y), P(t, f(z))$
   $\{x/t, y/f(z)\}$
2. $Q(a, y, f(y)), Q(y, y, u)$
   $\{z/a, y/a, u/f(a)\}$
3. $R(x, g(x)), R(y, y)$
   substituce nexistuje
4. $F(a, x), F(y, b)$
   $\{x/b, y/a\}$
5. $Q(x, f(y), z), Q(g(w), u, g(w))$
   $\{a/g(w), y/y, z/g(w), u/f(y)\}$
6. $P(g(x), y), P(u, f(w))$
   $\{x/x, y/f(w), u/g(w), w/w\}$

Všimněme si, že v unifikačním algoritmu se v bodu 3. vyžaduje, aby se proměnná, za níž se v substituci substituuje, nevyskytovala v termu, který je právě substituován. Tato kontrola výskytu termu (*occur-check*) je důležitá v teoretických úvahách, je ale výpočtově velmi náročná, proto se v praktických implementací logických programů od ní obvykle upouští a řešení se ponechává na programátorovi.

Poznamenejme ještě, ě ani paralelismus nemůže urychlit unifikační proces. (Hûet: Journal of Logic Programming) [55].

2.3.3 Varianty rezolučního principu

2.3.4 Lineární rezoluce s vytčeným prvkem

See [82].

2.3.5 Negace jako neúspěch

See [40].

2.4 Stratifikované logické programy

Definice 2.4.1 *Logický program p (tj. množina formulí tvaru $c \leftarrow \alpha \wedge \beta$) se nazývá stratifikovaný* stratified *právě když je možné dekomponovat množinu S všech predikátů vyskytujících se v p na třídu disjunktních množin $S_1, \ldots, S_p$ (nazývanou* stratum*) tak, že pro každou klauzuli*

$c \leftarrow a_1 \wedge \ldots \wedge a_n \wedge \neg b_1 \ldots \wedge \neg b_m$ *(kde a_i, b_j jsou atomy)*

v programu p

$$stratum(a_i) \leq stratum(c),$$
$$stratum(b_j) < stratum(c),$$

kde $stratum(a) = k$ pokud predikátový symbol $a \in S_k$.

Každá taková dekompozice se nazývá stratifikace programu *p.*

Stratifikace vlastně určuje priority mezi predikáty. Literál se může vyskytnout v těle klauzule pouze v případě, že patří k predikátům, jejichž stratum je vyšší než stratum predikátu v hlavě klauzule. Intuitivně, negace může být použita ve stratifikovaném programu pouze tehdy, když patří k již dříve definovanému predikátu.

Nechť Γ je acyklická ISA–hierarchie. $\Gamma \models x \rightarrow y$ nebo $\Gamma \models x \not\rightarrow y$, bude vyjadřovat skutečnost, že teorie dědění vlastností dovoluje z Γ odvodit formuli $x \rightarrow y$ resp. formuli $x \not\rightarrow y$. To znamená, že v odpovídajícím logickém programu se vyskytnou literály ξ a $\neg\xi$ jako součást uzlů v síti znalostí

2.5 Anotované logické programy

Anotování logických programů je další účinnou metodou, která umožňuje využít rozlišení argumentů v definicích predikátů na vstupní a výstupní, jak je to obvyklé v klasických imperativních jazycích.

Ideu snadno vyložíme na příkladu spojování seznamů standarní procedurou `append`, která je definována následovně

```
append([],list,list).
append([h|t1],t2,[h|t]) <-- append(t1,t2,t).
```

Anotování proměnných určuje, které pozice v predikátu jsou považovány za vstupní a které za výstupní. Např. výraz

```
append(i,i,o)
```

znamená, že první i druhý argument je považován za vstupní a třetí za výatupní argument. V anotaci `append(i,o,i)` je výstupní argument na druhé pozici. Neformálně můžeme říci, že v prvním případě se ptáme na to, co je výsledkem spojení dvou seznamů, ve druhém případě se ptáme na seznam, který je třeba připojit k prvnímu seznamu, abychom dostali seznam, který je konkretizován ve třetím argumentu.

Definice 2.5.1 *Nechť c je klauzule anotovaného logického programu. Výskyt proměnné X na vstupní pozici hlavy klauzule nebo na výstupní pozici libovolného literálu v těle klauzule se nzývá* definující výskyt *proměnné X v klauzuli c.*

Logický program, ve kterém pro každou klauzuli c má každá proměnná vyskytující se v klauzuli c definující výskyt v c, se nazývá jednoduchý logický program.

Teorém 2.5.1 *Pro jednoduché logické programy platí: V každém kroku výpočtu jsou všechny vstupní pozice jeho aktuálních (pod)cílů základní. Po úspěšném splnění aktuálního (pod)cíle jsou všechny výstupní pozice základní.*

Za zmínku stojí ještě to, že, jak ukázali Deransant a Maluszyński [26], nejde o omezení vyjadřovací síly jazyka logických programů, neboť každý Turingův stroj může být simulován jednoduchým programem. Navíc předností anotovaných logických programů je to, že mohou být korektně počítány bez kontroly výskytu (occur check) proměnné v termu při použití unifikace [82].

2.5.1 Reprezentace znalostí v logických programech

V tomto odstavci se ještě zmíníme o systémech usuzování založených na znalostech s neúplnou informací. Budeme předpokládat, že znalosti jsou reprezentovány logickými programy, takže vlastně můžeme spíše hovořit o deduktivních databázích. Začneme s odvozeními s nejjednodušším tvarem sémantických sítí, které se někdy nazývají *hierarchie dědičnosti* (angl. *inheritance hierarchies*).

Hierarchie dědičnosti jsou zajímavým příkladem znalostního systému s možností aktualizace znalostí i odvození. Znalosti jsou reprezentovány pravidly popisujícími vztahy dědění vlastností (ISA–hierarchies) mezi koncepty sémantické sítě [119], [137], [54], [50]. Sítě s ISA-hierarchiemi jsou obvykle definovány jako acyklický graf atomických formulí, které jsou spojeny pozitivními nebo negativními hranami, které vyjadřují právě vztahy dědění vlastností. Pozitivní hrany jsou tvaru *"x je y"* a můžeme je interpretovat a číst jako výrok *"x normálně y"* nebo *"x je typicky y"*. Negativní hrany jsou tvaru *"x je ne-y"*. Intuitivně: *"x je normálně ne-y"* nebo *"x je typicky ne-y"*.

Zde je několik příkladů sítí s nejednoznačnou interpretací:

Příklad 1. (Nixon diamond)

$n \to r, n \to q, q \to p, r \not\to p.$

Chcete-li, můžete jména uzlů (základních výroků) číst takto: $n = Nixon, r = republikn, q = quaker, p = pacifista.$

Příklad 2. (Problém barvy slonů)

$c \to a, a \to s, s \ \to g, c \to k, k \ \to s, k \ \not\to g.$

Příklad 3.

$a \to b, b \to d, d \not\to e, a \not\to d, b \to c, c \to e.$

Příklad 4.

$a \to b, a \to c, b \to d, c \not\to d, c \to e, d \to e, d \to f, e \not\to f.$

Ve většině metod používaných pro odvozování v těchto systémech je preferována specifičtější informace. Tak může být dosaženo toho, že pro každý acyklický graf (tj. graf, který neobsahuje žádný orientovaný cyklus) lze vytvořit právě jednu extenzi.

2.6 Logické jazyky pro systémy znalostí

Jazyky pro báze znalostí sestávají z těchto komponent:

K - množiny formulí popisující bázi znalostí
Q - jazyk otázek
A - jazyk odpovědí

Systém odpovídání dotazů (QA-system) je funkce

$$answ : K \times Q \rightarrow A$$

Příklady.
Teorie prvního řádu
$K = Q = A, In =$ klasická logická konsekvence Cn (neboli relace $\vdash$, která je **monotónní oprací** (relací na množině formulí.

Relační databáze
$K =$ množina základních atomických formulí (pozitivních faktů), které jsou reprezentovány tabulkami či relacemi,
$Q = SQL$,
$A = \{ano, ne\}$,
Operace inference je **nemonotónní**, protože negace $\neg$ je interpretována jako množinový rozdíl v relační algebře.

Jednoduchá deduktivní databáze (Hornovy klauzule)
$K =$ množina pozitivních faktů a pravidel tvaru

$$\alpha \Leftrightarrow \lambda_1 \wedge \ldots \wedge \lambda_n,$$

kde α je atomická formule, λ_i jsou literály a $\neg$ je interpretována jako neúspěch (negation-as-failure),
$Q =$ množina atomických formulí,

$A = \{ano, ne\}$,
In = lineární rezoluce s vytčeným prvkem ([?]).

Disjunktivní deduktivní databáze
K = množina disjunkcí literálů a pravidel jako v jednoduchých logických databázích,
Q = množina literálů,
$A = \{ano, ne\}$ se substitucí,
In = lineární rezoloce (srovnej Minker 93).

Obecné logické programy
Jsou ekvivalentní uzavřeným teoriím prvního řádu.
K i Q jsou množiny obecných klauzulí,
In je klasická logická relace dokazatelnosti $\vdash$.

Část II

Alternativní inference

Kapitola 3

Nededuktivní odvození

Usuzování není jen deduktivní proces, dokonce ne především, a to i tehdy, když jde o postupy, které bychom byli ještě ochotni označit jako racionální a podložené argumenty. Uvedeme alespoň příklady typických nededuktivních úsudků, které by měly být analyzovány právě z hlediska racionality.

3.1 Induktivní usuzování

Deduktivní usuzování, které bylo předmětem předchozích dvou kapitol, není jediným typem usuzování. To je dobře známo. Vedle dedukce je induktivní usuzování považováno za hodnotné, i když ne tak dobře zabezpečené jako dedukce. Pro induktivní úsudky je charakteristické, že závěry nelze tvrdit s definitivní platností. Jejich pravdivost není opřena pouze na pravdivosti všech premis daného argumentu.

Uvidíme to na příkladu. Uvažujme tento úsudek:

Dané pravidlo: Všichni králíci v tomto klobouku jsou bílí.
Daný fakt: Tito králíci jsou z tohoto klobouku.

Závěr: Tito králíci jsou bílí.

Je to typické deduktivní pravidlo, podle kterého je závěr odvozen na základě obecné věty (*major premise*) a specifického faktu (*minor premise*). Je přitom plně zaručeno, že za předpokladu pravdivosti premis je pravdivý i závěr úsudku.

Naproti tomu úsudek

Daný fakt: Všichni tito králíci jsou z tohoto klobouku.
Daný fakt: Všichni tito králíci jsou bílí.

Závěr (pravidlo): Všichni králíci v tomto klobouku jsou bílí.

je typickým příkladem induktivního argumentu. Závěr zde má tvar obecné věty. To ale znamená, že vždy ještě může být nalezen protipříklad, který případně vyvrátí tentativní závěr.

Tématika induktivního usuzování je tradičně zkoumána z různých hledisek, často v kontextu pravděpodobnostního usuzování.

3.2 Abduktivní usuzování

Abduktivní usuzování je obvykle chápano jako hledání nejlepšího plauzibilního vysvětlení dané množiny faktů [104], [105].

Zde je příklad:

(*Dané pravidlo*) Všichni králíci v tomto klobouku jsou bílí.
(*Daný fakt*) Tito králíci jsou bílí.

(*Závěr*) Tito králíci jsou z tohoto klobouku.

V tomto připadě závěrem je *hypotetický* fakt.

Provádíme-li abduktivní úsudek v přirozeném jazyce, obvykle v takovém případě dodáváme formulaci typu ... *to je proto, že* tito králíci jsou z tohoto klobouku.

Abduktivní inference patří k důležitému typu úsudků a proto jí věnujeme ve třetí části této publikace samostatnou kapitolu.

3.3 Pravděpodobnostní usuzování

Pravděpodobnostní usuzování je důležité v těch situacích, kdy jsou naše znalosti neurčité a neučitost je vedena pravděpodobnostními charakteristikami. Například jsou známa marginální pravděpodobnostní rozdělení zkoumaných jevů či veličin a my chceme vědět víc o celé struktuře.

Typickým příkladem pravděpodobnostního úsudku je následující úvaha:

Příklad: **************************************

3.4 Usuzování z neurčitých informací

Fuzzy usuzování je pokus formalizovat usuzování zatížené neurčitostí, které ale nemá nutně pravděpodobnostní interpretaci. V současné době je to asi nejrychleji se rozvíjející oblast logických zkoumání, kde významnou roli hraje i pražská logická škola. [52].

3.5 Usuzování podle analogie

Patří k nejméně prozkoumaným, ale často používaným, typům úsudků. Vychází z neúplné informace a nepřesně definované podobnosti. Významnější roli hraje především při vysvětlování. (Viz. např. J. F. Nilson [**?**].

3.6 Metaúsudky

Úsudky o úsudcích, to je další velice významný typ úsudku. Porozumění metaúsudkům je však stále jen v počátcích. Zde je příklad:

(*Fakt*) Znám všechny své bratry.
(*Fakt*) Nemám žádnou informaci o tom, že Pavel je můj bratr.

(Závěr) Pavel není můj bratr.

Zdůvodnění bývá třeba následující: Kdyby to byla pravda, tak bych o tom musel něco vědět. To je typická argumentace, kterou používáme při absenci znalosti či nějaké pozitivní informace. Týká se to možností usuzování z neúplných znalostí anebo usuzování na základě negativní informace. O tom viz dále.

Je mnoho zajímavých vlastností rozmanitých typů úsudků, které by měly být v souvislosti s odvozováním a s dynamikou znalostí studovány. Patří k nim jistě neúplnost, revidovatelnost, falzifikovatelnost, aproximativnost, neurčitost, vágnost, podmíněná korektnost, a další vlastnosti. Otvírá se nám tím bohatá škála témat pro studium takových typů úsudků, které sice nezachovávají pravdivost odvozených tvrzení striktně, ale mají společné to, že stále ještě jde o odvození opřená o nějaké racionální argumenty.

Naším cílem je jednak mapovat rozmanitost racionální inference, jednak hledat ty vlastnosti, které jsou společné většině z nich, a přispět tak k porozumění konceptu racionální konsekvence.

Kapitola 4

Nemonotónní usuzování

Deduktivní usuzování je hlavním tématem klasické logiky, protože zachovává pravdivost odvozených tvrzení. Jestliže však chceme v rámci logiky reprezentovat znalosti a vztahy mezi jednotlivými položkami znalostí, musíme se zabývat i postupy, které nejsou striktně deduktivní a které jsme alespoň v přehledu uvedli v předchozí kapitole. K takovým postupům patří i tentativní odvození, která nám umožňují překlenout někdy absenci informace, někdy se vypořádat s dynamikou znalostí, tj. faktem, že mnoho položek znalostí má jen dočasnou platnost. To je typické pro tzv. *nemonotónní usuzování* .

Nemonotónní usuzování je často charakterizováno jako takový druh odvození, který může být zpochybněn nově příchozí informací. Je to tedy přesně ten druh inference, která má co činit s dynamikou znalostí.

Během posledních dvou desetiletí byla značná pozornost věnována detailnímu studiu mnoha systémů nemonotónního usuzování. Na počátku osmdesátých let dvacátého století, kdy vyšlo speciální číslo časopisu *Artificial Intelligence*, se zájem o tuto problematiku nápadně zvýšil. Brzy se ukázalo, že jde o otázky jak zajímavé tak také obtížné. Množství autorů přicházelo s mnoha dílčími problémy a brzy jsme měli před sebou velmi rozdílné systémy. Naším cílem v této kapitole bude referovat o těch systémech nemonotónního usuzování, které nejvíce ovlivnily náš pohled na strukturu znalostí, zejména na znalosti podmíněné.

4.1 Reiterova logika defaultů

4.1.1 Výchozí pojmy

Logika nemonotónního usuzování, či přesněji, formální systém navržený Raymondem Reiterem v již zmíněném speciálním čísle časopisu Artificial Intelligence[114] patří k těm systémům, které nejsilněji ovlivnily náš náhled na racionální usuzování, které ale nezachovává pravdivost v klasickém logickém slova smyslu. Reiterova logika defaultů se později stala jakýmsi referenčním systémem, s nímž jsou ostatní systémy založené na pravidlových principech porovnávány, i když byly později navrženy různé modifikace. To se týkalo zejména pojmu extenze teorie s defaulty. Pro logiky defaultů je obvyklé, že vyjadřovací síla jazyka teorie prvního řádu se rozšíří o speciální odvozovací pravidla nazývaná *defaulty.* Defaulty mají tvar

$$d = (\alpha(x); \beta(x)/\gamma(x))$$

kde $\alpha(x), \beta(x)$, a $\gamma(x)$ jsou formule jazyka prvního řádu.
Tato pravidla jsou obvykle interpretována takto:

Jestliže $\alpha(x)$ platí a $\beta(x)$ můžeme konzistentně předpokládat, pak odvoď $\gamma(x)$.

Formule $\alpha(x)$ v defaultovém pravidle d se obvykle nazývá předpoklad (angl. *precondition* nebo *prerequisite*), $\beta(x)$ se nazývá ospravedlnění (*justification*) a formule $\gamma(x)$ se nazývá závěr (*consequent*) pravidla.

Příklady:
$d_1 = (\text{pták}(x); \text{létá}(x)/\text{létá}(x))$
$d_2 = (\text{poraněné křídlo}(x); \neg\text{létá}(x)/\neg\text{létá}(x))$

Default d_1 pak můžeme interpretovat slovy *Když* x *je pták a je možné konzistentně předpokládat, že létá, pak uzavřeme, že létá.* Default d_2 pak říká, že *má-li poraněné křídlo, pak nelétá.*

V dalším budeme předpokládat, že množina faktů i defaultových pravidel teorie sestávají z formulí prvního řádu, nebo speciálně ze sentencí. Definice základních pojmů budou ale předkládány pokud možno nezávisle na jazyce a nezávisle na zvolené operaci logického důsledku, která je předpokládána v zázemí reprezentace znalostí. Předpokládáme tedy, že jazyk je dán, tj. je dána množina F dobře utvořených formulí a je dána deduktivní (monotónní) operace logické konsekvence Cn definovaná na F, která je idempotentní, monotónní a tranzitivní, jak jsme o tom pojednávali v první kapitole. Teorii s defaulty budeme chápat jako množinu formulí (většinou jako fakty vyjádřené formulemi prvního řádu) spolu s množinou D defaultů, které budeme nejčastěji formalizovat trojicemi uzavřených formulí prvního řádu.

Definice 4.1.1 *Default se nazývá* uzavřený*, když neobsahuje žádné volné proměnné (v obvyklém smyslu jak jsou definovány v rámci teorií prvního řádu); jinak se nazývá* otevřený*. Default se nazývá* normální*, jestliže jeho ospravedlnění a závěr jsou logicky ekvivalentní formule (vzhledem k operaci* Cn*).*

Naším hlavním cílem v tomto odstavci je definovat pojem *extenze* teorie s defaulty, který je analogií deduktivního uzávěru v klasickém smyslu. Protože jsme dosud na defaultová pravidla nekladli žádné požadavky jako tomu bylo v případě deduktivních pravidel (klasická deduktivní pravidla musejí být logicky korektní, což znamená, že musejí zachovávat pravdivost tj., že od pravdivých premis musejí vést k pravdivým závěrům), setkáváme se s tím, že teorie s defaulty může mít více extenzí anebo naopak mohou existovat "rozumné" teorie, které nemají žádnou extenzi.

Definice 4.1.2 *Množina formulí E se nazývá* extenze *teorie $T = [F, D]$ s fakty F a defaulty D, jestliže je to nejmenší deduktivně uzavřená množina formulí, která obsahuje F a pro každý default $d \in D$ $(d = (\alpha; \beta/\gamma))$ takový, že $\alpha \in Cn(E \cup \{\beta\}) \neq Fle$ (tj. β může být konzistentně přidáno k*[1] *výsledné množině E) $\gamma \in E$.*

Poznámka: Když jazyk reprezentace znalostí obsahuje negaci a operace logického důsledku je charakterizována standardní relací dokazatelnosti (tj. relací $\vdash$), potom podmínka, že extenze teorií s defaulty mají být uzavřené na všechny aplikovatelné defaulty může být vyjádřena, možná zřetelněji, tak, že pro každý default $d = (\alpha; \beta/\gamma)$ platí:
$E \vdash \neg\beta$ nebo $\gamma \in E$.

Definice 4.1.3 *Nechť $T = [F, D]$ je teorie s defaulty, F množina faktů, D množina defaultů. Nechť $\Gamma(X)$ je nejmenší taková množina, že:*

- $F \subseteq \Gamma(X)$
- $Cn(\Gamma(X)) = \Gamma(X)$
- *Když $(\alpha; \beta_1, \ldots, \beta_n/\gamma) \in D$ a $\alpha \in \Gamma(X)$ a $\beta_1, \ldots, \beta_n$ jsou konzistentní s $\Gamma(X)$ (vzhledem k Cn), tak $\gamma \in \Gamma(X)$.*
  Nebo, alternativně (pokud v jazyce máme negaci):
  Jestliže $(\alpha; \beta_1, \ldots, \beta_n/\gamma) \in D$ a $\alpha \in \Gamma(X)$ a $\neg\beta_1, \ldots, \neg\beta_n \notin \Gamma(X)$, potom $\gamma \in \Gamma(X)$.

Množina E formulí je extenzí *teorie T s defaulty D, jestliže $\Gamma(E) = E$, tj. když E je pevným bodem zobrazení Γ.*

[1]Termín “konzistentně přidáno k” zde zřejmě znamená, že konzistence se myslí vzhledem k operaci deduktivní konsekvence, která tvoří logické zázemí.

Není těžké ověřit všechny podmínky kladené na extenzi v předešlé definici (třetí podmínka vyplývá z minimality Γ operátoru).

Teorém 4.1.1 *Nechť $T = [F, D]$ je teorie s defaulty D a E je množina uzavřených formulí prvního řádu. Definujeme*

$E_0 = F$

Pro $i \geq 0$:

$E_{i+1} = Cn(E_i) \cup \{\gamma : (\alpha; \beta_1, \ldots, \beta_n/\gamma) \in D, \alpha \in E_i$ *a* $\beta_1, \ldots, \beta_n$ *jsou konzistentní s* E *(nebo* $\neg\beta_1, \ldots, \neg\beta_n \notin E)\}$.

Potom E je extenze teorie $T = [F, D]$ s defaulty právě tehdy, když $E = \bigcup_{i=0}^{\infty} E_i$.

Jinými slovy, E je množina (uzavřených) formulí, která obsahuje F a která je uzavřená jak na logické důsledky tak na defaultová pravidla. Zde je vhodné poznamenat, že nehledě na to, že množina E je zmíněna v definici E_i, výsledná definice extenze není definicí kruhem, ale je to naprosto korektní definice. Nedává ovšem přímý návod.

Důkaz: Nejprve dokážeme implikaci zleva doprava: Nechť E je extenze. Indukcí můžeme dokázat, že pro každé $i \geq 0 : E_i \subseteq E$ a tudíž $\bigcup_{i=0}^{\infty} E_i \subseteq E$ a vzhledem k tomu, že $E = \Gamma(E)$ dostáváme, že $E = \bigcup_{i=0}^{\infty} E_i \subseteq E$.
Implikace zprava doleva: Předpokládejme teď, že $E = \bigcup_{i=0}^{\infty} E_i \subseteq E$. Podobně jako v předchozí části důkazu můžeme ukázat, že pro všechna $i \geq 0$ platí $E_i \subseteq \Gamma(E)$, takže $E = \bigcup_{i=0}^{\infty} E_i \subseteq E \subseteq \Gamma(E)$.

Q. E. D.

Teorém 4.1.2 *(Minimalita extenze) Nechť E_1 a také E_2 jsou dvě extenze teorie $T = [F, D]$ s defaulty. Jestliže $E_1 \subseteq E_2$, potom $E_1 = E_2$.*

Důkaz: Podle předchozího teorému máme

$$E_1 = \bigcup_{i=0}^{\infty} E'_i,\ E_2 = \bigcup_{i=0}^{\infty} E''_i.$$

Jelikož $E_1 \subseteq E_2$, stačí dokázat, že také $E_2 \subseteq E_1$. To dokážeme indukcí podle konstrukce extenze. Snadno ověříme, že $E''_0 \subseteq E'_0$, jelikož obě množiny se rovnají F. Předpokládejme navíc, že $E''_i \subseteq E'_i$ a uvažujme formuli $\gamma \in E''_{i+1}$. Chceme ukázat, že také $\gamma \in E'_{i+1}$. Je jistě pravda, že když $\gamma \in Cn(E''_{i+1})$, tak potom $E''_i \subseteq E'_i$ a $\gamma \in Cn(E'_i) \subseteq E'_{i+1}$; jinak existuje takový default $d = (\alpha; \beta_1, \ldots, \beta_n/\gamma)$, že $\alpha \in E''_i$ a $\neg\beta_i, \ldots, \neg\beta_n \notin E''$. Protože $E''_i \subseteq E'_i$ a $\alpha \in E''$, dostáváme $\alpha \in E'_i$ a $\neg\beta_i, \ldots, \neg\beta_n \notin E$. Tudíž $\gamma \in E'_{i+1}$.

Q. E. D.

Důsledek 4.1.1 *Teorie s defaulty má inkonzistentní extenzi právě tehdy, když sama teorie T je inkonzistentní.*

Teorém 4.1.3 *Nechť E je extenze teorie $T = [F, D]$ s defaulty. Nechť $X \subseteq E$. Potom E je také extenzí teorie s fakty $F \cup X$ a s defaulty D.*

Důkaz: (Indukcí).

Tento teorém říká, že Reiterův koncept extenze teorie s defaulty splňuje podmínku opatrné monotónnosti, kdy můžeme přidávat fakty z extenze, aniž bychom tím pozměnili vztahy vzájemné odvoditelnosti.

4.1.2 Příklady extenzí teorií s defaulty

Začněme s extenzemi různých teorií s defaulty, z nichž většina pochází již od Reitera. (Srovnej Reiter [114]).

E1. **Teorie se dvěmi extenzemi**:
Nechť T je teorie s fakty
$F = \{b \Rightarrow \neg a\ \wedge \neg c\}$ a defaulty

$D = \{(; a/a), (; b/b), (; c/c)\}$. potom tato teorie má dvě různé extenze:

$E_1 = Cn(F \cup \{a, c\})$
$E_2 = Cn(F \cup \{b\})$.

Poznamenejme, že
$Cn(F) = \{b\wedge\neg a\wedge\neg c, \neg b\wedge\neg a\wedge\neg c, \neg b\wedge a\wedge c, \neg b\wedge\neg a\wedge c, \neg b\wedge a\wedge\neg c\}$

E2. **Teorie s jedinou extenzí**:
$F = \emptyset$,
$D = \{(; a/\neg b), (; b/\neg c), (; c/d)\}$. Tato teorie má právě jednu extenzi.

$E = Cn(\{\neg b, d\})$.

E3. **Teorie se třemi různými extenzemi**:
$F = \{b, c \Rightarrow d \vee a, a \ \wedge c \Rightarrow \neg e\}$.
$D = \{(; a/a), (; c/c), (d \vee a; e/e), (c \wedge e; \neg a, d \vee a\ /f)\}$. Tato teorie má tyto tři extenze:

$E_1 = Cn(F \cup \{a, c\})$
$E_2 = Cn(F \cup \{a, e\})$
$E_3 = Cn(F \cup \{c, e, f\})$.

E4. **Teorie se dvěmi extenzemi, které navíc podporují různé ontologie**:
$F = \emptyset$,
$D = \{(a; (\exists x)P(x)/(\exists x)P(x)), (; a/a), (; \neg a\ /\neg a)\}$.

$E_1 = Cn(\{\neg a\})$
$E_2 = Cn(\{a, (\exists x)P(x)\})$.

E5. **Teorie bez extenze**:
$F = \{a\}$,
$D = \{(a; b \wedge c/c), (c; \neg b/\neg b)\}$.

E6. **Jiný příklad teorie bez extenze**:
$F = \emptyset$,
$D = \{(; a/\neg a)\}$.

Pravda, to je poněkud "patologický" default, který paradoxně spojuje pozitivní ospravedlnění výroku s negativním závěrem o témže výroku.

Následující příklad ukazuje, že přidání nových faktů k dané množině faktů může mít vliv dokonce na počet extenzí.

E7. Uvažujme množinu defaultů: $D = \{(; a/a), (a \vee b; \neg a/\neg a)\}$ a dvě množiny faktů:
$F_1 = \emptyset$
$F_2 = \{a \vee b\}$.

Potom teorie $T_1 = [F_1, D]$ má právě *jednu* extenzi $E = Cn(F_1 \cup \{a\})$, zatímco teorie $T_2 = [F_2, D]$ má *dvě* extenze $E' = Cn(F_1 \cup \{a\})$ a $E'' = Cn(F_2 \cup \{\neg a, b\})$.

4.1.3 Omezení defaultů

Teorie s libovolnými defaulty jsou příliš složité a mají některé nežádoucí vlastnosti, které se na této hladině obecnosti dají jen těžko dobře charakterizovat. Hlavní problém v případě defaultů, jejichž formát není omezen, spočívá v tom, že defaultové pravidlo může být aplikováno dokonce i v situaci, kdy o závěru tohoto pravidla je již známo, že je nepravdivý. Abychom se takovým situacím vyhnuli, požadujeme, aby platnost závěru pravidla byla obsažena již v ospravedlnění pravidla. To vedlo ke studiu tzv. *seminormálních* defaultů, tj. pravidel tvaru

$$(\alpha; \beta \wedge \gamma/\gamma).$$

Je to dost dobré omezení tvaru pravidel, které nemá podstatný vliv na jejich vyjadřovací sílu. Jürgen Dix [28] ale ukázal, že některé důležité vlastnosti jako např. kumulativnost a existence extenze nelze zaručit dokonce ani pro seminormální pravidla s prázdnými premisami (angl. *prerequisite-free*), tj. s pravidly tvaru

$$(true \;\; ; \beta \wedge \gamma/\gamma)$$

Abychom mohli dobře formulovat následující teorém o semimonotónnosti, zavedeme nejprve jednoduchý ale důležitý koncept množiny defaultů generujících extenzi E vzhledem k teorii $T = [F, D]$.

Definice 4.1.4 *Nechť $T = [F, D]$ je teorie s uzavřenými defaulty a E je její extenze. Množinu $gd(E, [F, D])$ defaultů* generujících *extenzi E vzhledem k T definujeme následovně:*

$$gd(E, [F, D]) = \{d \in D;\ \alpha \in E \ \ a \ \neg\beta \notin E\}.$$

Je zřejmé, že když E je extenze teorie $T = [F, D]$, tak E můžeme chápat jako deduktivní uzávěr množiny formulí, který obsahuje fakta z F spolu se všemi závěry generujících defaultů, tj.

$$E = Cn(F \cup consq(gd(E, T))),$$

kde $consq(D) = \{\gamma; (\alpha; \beta/\gamma) \in D\}$.

Teorém 4.1.4 *(Semimonotónnost)*

Nechť $T = [F, D]$ je teorie s uzavřenými normálními defaulty. $D' \subseteq D$ a E' je extenze teorie $T' = [F, D']$. Potom T má takovou extenzi E, že

1. *$E' \subseteq E$ a*
2. *$gd(E', T') \subseteq gd(E, T)$.*

To znamená, že gd "odděluje zrno od plev", tj. vybírá pouze ty skutečně užitečné defaulty.

Důkaz: Definujeme

$F_0 = F$ a pro $ige0$ definujeme

$F_{i+1} = Cn(F_i) \cup \{\gamma; (\alpha; \gamma/\gamma) \in D$, kde $\alpha \in F_i$ a $\neg\gamma \notin E\}$.

Ukážeme, že množina

$E = \bigcup_{i=0}^{\infty} F_i$ je extenze teorie $T = [F, D]$ (podle Teorému 4.1.1).

Avšak teorém o semimonotónnosti neplatí pro libovolné defaulty. Zde je protipříklad: Nechť $T = [F, D]$ je taková teorie, že $F = \emptyset$ a $D = \{(;\beta/\gamma)\}$. T má pouze jednu extenzi $E = Cn(\{\gamma\})$. Přidáme-li nový default $d' = (;\neg\beta/\neg\beta)$ do D obdržíme teorii $T' = [F, D']$, $D' = D \cup \{d'\}$ s právě jednou extenzí $E' = Cn(\{\neg\beta\})$, ale $E \not\subseteq E'$.

Zajímavou třídu teorií s "rozumnými" defaulty tvoří teorie s tzv. *normálními* defaulty. Připomeňme, že normálním defaultem rozumíme pravidlo, jehož ospravedlnění i závěr tvoří táž formule, nebo obecněji, jsou to logicky ekvivalentní formule. To znamená, že normální defaulty jsou pravidla tvaru

$$(\alpha; \gamma/\gamma)$$

Teorie obsahující pouze normální defaulty mají dobré a žádoucí vlastnosti, především vlastnost, která zaručuje existenci extenze. To právě říká následující teorém.

Teorém 4.1.5 *Každá teorie s defaulty, která obsahuje pouze uzavřené normální defaulty, má aspoň jednu extenzi.*

Důkaz: Nechť $T = [F, D]$ je teorie s normálními defaulty. Když F je inkonzistentní, tak potom tato teorie má inkonzistentní extenzi (srovnej s důsledkem teorému 4.1.2 o minimalitě extenzí). Předpokládejme teď, že F je konzistentní. Požadovanou extenzi zkonstruujeme takto:
$E_0 = F$
Pro každé $i \geq 0$, nechť F_i je maximální množina uzavřených formulí, taková, že

1. $E_i \cup F_i$ je konzistentní a

2. jestliže $\varphi \in F_i$, tak existuje takový default $d = (\alpha; \gamma/\gamma) \in D$, že $\varphi = \gamma$ a $\alpha \in E_i$.

Definujme $E_{i+1} = Cn(E_i) \cup F_i$ a $E = \bigcup_{i=0}^{\infty} E_i$.

To, že E je extenzí teorie T, můžeme dokázat tak, že ověříme, že

$F_i = \{\gamma : (\alpha; \gamma/\gamma) \in D$, kde $\alpha \in E_i$ a $\neg\gamma \notin E\}$ a aplikací teorému 3.1.1.

Q. E. D.

Ano, název normální defaulty je velmi dobře vybrán, protože každý "skutečně normální" default má tuto formu, zatímco ostatní formy jsou více či méně "patologické" jako třeba ve výše uvedeném příkladu E5. Zkuste nalézt opravdu *přirozeně* formulovaný příklad defaultového pravidla, které není normální.

Teorém 4.1.6 *(Ortogonalita extenzí) Jestliže teorie $T = [F, D]$ s uzavřenými normálními defaulty má dvě různé extenze E', E'' (tj. $E' \neq E''$), potom množina $E' \cup E''$ je inkonzistentní.*

Důkaz: Podle teorému 4.1.1 $E' = \bigcup_{i=0}^{\infty} E'_i$ a $E'' = \bigcup_{i=0}^{\infty} E''_i$, kde

$E'_0 = F$ a pro $i \geq 0$ $E'_{i+1} = Cn(E'_i) \cup \{\gamma; (\alpha; \gamma/\gamma) \in D$, kde $\alpha \in E'_i$ a $\neg\gamma \notin E'\}$.

Analogicky pro E'': Protože E' a E'' jsou různé, tj. $E' \neq E''$, a $E'_0 = E''_0$, tak musí existovat takové $i \geq 0$, že $E'_i = E''_i$ a $E'_{i+1} \neq E''_{i+1}$. Takže existuje takový default d, že $\alpha \in E'_i = E''_i$ a pro který $\neg\gamma \notin E'$ a současně $\gamma \in E'_{i+1}$, ale $\gamma \notin E''_{i+1}$. Avšak když $\alpha \in E'_i$ a $\gamma \notin E''_{i+1}$, tak $\neg\gamma \in E''$. Tudíž $\gamma \in E'$ a $\neg\gamma \in E''$. To znamená, že $E' \cup E''$ je inkonzistentní.

Q. E. D.

Důsledek 4.1.2 *: Jestliže $T = [F, D]$ je taková teorie s uzavřenými normálními defaulty, že $F \cup consq(D)$ je bezesporná, pak T má právě jednu bezespornou extenzi.*

Důkaz: Předpokládejme opak, tj. že T má dvě různé extenze E_1, E_2. Potom

$E_1 = Cn(F \cup consq(gd(E_1, T)))$
$E_2 = Cn(F \cup consq(gd(E_2, T)))$
Nyní je zřejmé, že
$E_1 \subseteq Cn(F \cup consq(D))$
$E_2 \subseteq Cn(F \cup consq(D))$

Avšak předpoklad, že množina $F \cup consq(D)$ je konzistentní, je ve sporu s tím, že sjednocení $E_1 \cup E_2$ obou extenzí je sporné.
Q. E. D.

Konečně ukážeme, že pro danou množinu formulí počet extenzí teorie s normálními defaulty neklesá s rostoucím počtem defaultů. To lze dobře vyjádřit následujícím teorémem.

Teorém 4.1.7 *Nechť $T = [F, D]$ a $D' \subseteq D$ a nechť E_1' a E_2' jsou dvě různé extenze teorie $T' = [F, D']$. Potom T má dvě různé extenze E_1, E_2 takové, že $E_1' \subseteq E_1$ a $E_2' \subseteq E_2$.*

Důkaz: Vzhledem k semimonotónnosti existují dvě extenze E_1 a E_2 takové, že $E_1' \subseteq E_1$ a $E_2' \subseteq E_2$. Předpokládejme, že $E_1 = E_2$. Potom $E_1' \cup E_2' \subseteq E_1$, ale $E_1' \cup E_2'$ je sporná. Takže i E_1 musí být sporná a tudíž i T (i.e. F) musí být sporná.
Q. E. D.

Poznámka: Jak ještě uvidíme později, analogická tvrzení pro nemonotónní logiku McDermott–Doyleovu neplatí.

4.1.4 Problémy v teoriích s defaulty

Stále ale zůstávají různé problémy usuzování v teoriích s defaulty, které pocházejí z mimologických zdrojů. K těm nejdůležitějším problémům patří nemožnost rozlišit mezi standardními a defaultovými závěry. Navíc není zřejmé, jak přeložit defaulty z přirozeného jazyka do zamýšlené formální reprezentace (normální, seminormální, obecné defaulty). Z výpočtového hlediska je to především fakt, že širší třídy teorií s defaulty v jazyce prvního řádu nejsou algoritmicky zvládnutelné [114]. To se ovšem týká i seminormálních teorií (Etherington 1988, Selman and Kautz 1988).

Jsou dobré důvody k tomu, zkoumat modifikace konceptu extenze, které by měly vhodnější vlastnosti. Takové výzkumy se v nedávné době uskutečnily. Například Witold Łukaszewicz [84] studoval variantu Reiterovy logiky defaultů, která zaručuje existenci extenzí spolu s vlastností semimonotónosti.[2] Ale opět, Poole nalezl příklad (*broken arms*[3]) ukazující, že jsou situace, které nejsou adekvátně traktovány ani konceptem extenze modifikovaným podle Łukaszewicze.

Zkušenost s Brewkovou [16] kumulativní logikou defaultů ukazuje, že stále existují otevřené problémy. To znamená, že existují problémy, které nelze řešit uniformním způsobem. Správná cesta snad povede k rovnováze mezi omezením defaultů a žádoucí volností jazyka pro vyjádření defaultů.

Následující Makinsonův příklad dobře ilustruje problém:
$d_1 = (true; p/p)$
$d_2 = (p \vee q; \neg p/\neg p)$.
Avšak znovu je to jistý druh patologie.

[2]Łukaszewiczova varianta je založena na odlišném řešení konfliktů mezi aplikovatelnými defaulty. V Reiterově přístupu jsou všechna aplikovatelná pravidla “nucena” se aplikovat, tj. všechna pravidla se zúčastní konstruování extenze, někdy se ale simultánní aplikovatelnost pravidel ukáže jako nemožná či nerealizovatelná. To se stane např. v situaci, kdy závěry defaultů spolu se základními fakty a závěry jiných defaultů jsou v kontradikci s některými vlastními ospravedlněními nebo popírají některé z již aplikovaných defaultů, anebo když závěr defaultu je ve sporu s větami odvozenými ze základních faktů a závěrů defaultových pravidel.

3****************************

4.1.5 Věta o reprezentaci pro defaultové logiky

Pro $T = [F, D]$, nechť $ext(T)$ je třída extenzí teorie T s defaulty D a fakty F. Je zřejmé, že lze nalézt jiné teorie, které mají stejnou třídu extenzí. Takové teorie, nazveme ekvivalentní (symbolicky $T \approx T'$) tj. $T_1 = [F_1, D_1]$ a $T_2 = [F_2, D_2]$ jsou *ekvivalentní*, jestliže $ext(T_1) = ext(T_2)$.

Příklad: Nechť $T_1 = [F_1, D_1]$ je teorie s defaulty v jazyce nultého řádu (s následujícími výrokovými konstantami p, q, r, s, z, t) kde $F_1 = \{p\}$, $D_1 = \{(p; \neg r/q), (q; s\ /t)\}$. Teorie T_1 má právě jednu extenzi $Cn(\{p, q, t)\}$ a teorie $T_2 = [F_2, D_2]$ s prázdnou množinou faktů $F_2 = \emptyset$ a s defaulty $D_2 = \{(; /p), (p; \neg r/q), (q; s\ /t)\}$ má tutéž extenzi. Tudíž $T_1 \approx T_2$.

Snadno ověříme, že pro každé dvě množiny F_1, F_2: Jestliže $Cn(F_1) = Cn(F_2)$, potom pro libovolnou množinu D defaultů $[F_1, D] \approx [F_2, D]$.

Teorém 4.1.8 *Nechť $E \subseteq \mathcal{L}$. Potom E je extenze teorie $T = [F, D]$ právě když E je extenze teorie $T' = [F, D \cup \mathcal{D}(F)]$ kde $\mathcal{D}(F) = \{(; /\varphi) : \varphi \in F\}$.*

Důsledek 4.1.3 *Každá teorie s defaulty je reprezentovatelná třídou teorií defaultů s prázdnou množinou faktů, tj. $F = \emptyset$.*

Je otázka, zda je či není možné pro každou teorii T najít ekvivalentní teorii T', která obsahuje pouze defaulty bez normálních předpokladů[4]. Tento problém byl pozitivně vyřešen Bonattim a Eiterem v roce 1995 [15]. Marek, Treur, a Truszczyński [**?**] ukázali, že pro každou bezespornou teorii defaultů T je možné zkon-

[4]Default s prázdnou množinou předpokladů se nazývá *prerequisite-free*.

struovat teorii s normálními prerequisite-free defaulty T', která je ekvivalentní T.[5]

Definice 4.1.5 Monotónní část *množiny defaultů D je množina pravidel, definovaných rovností*

$$m\text{-}part(D) = \{(p(d); /c(d)) : d \in D\}.$$

Ve skutečnosti je monotónní část množiny D množinou defaultů bez ospravedlnění, anebo lépe, množina defaultů zbavených všech svých ospravedlnění.

Definice 4.1.6 *Řekneme, že defaultové pravidlo d je* aplikovatelné *vzhledem k množině X formulí (je X-aplikovatelné), jestliže $\neg\beta_i \notin Cn(X)$ pro každé $\beta \in j(d)$.*

Definice 4.1.7 Redukt D_X *množiny pravidel D vzhledem k X definujeme rovností*

$$D_X = m\text{-}part(\{d \in D : d \text{ je } X\text{-aplikovatelné}\}).$$

Teorém 4.1.9 *Množina formulí E je* extenzí *teorie $T = [F, D]$ právě když $E = Cn^{D_T}(F)$.*

Definice 4.1.8 *Nechť $\mathcal{E}$ je třída množin formulí v $\mathcal{L}$, tj. $\mathcal{E}$ v $\mathcal{P}(\mathcal{P}(\mathcal{F}))$. Řekneme, že $\mathcal{E}$ je* reprezentovatelná *teorií $T = [F, D]$, když $ext(T) = \mathcal{E}$.*

Teorém 4.1.10 *Třída $\mathcal{E}$ množin formulí (v jazyce $\mathcal{L}$) je* reprezentovatelná *teorií $T = [F, D]$ s množinou D defaultů, právě tehdy, když $\mathcal{E}$ je množina ortogonální*[6] *vzhledem k množině, která je konečně generovaná průniky prvků z $\mathcal{E}$.*

Důkaz: Viz Marek, Treur, and Truszczyński in [86].

[5] Navíc, rozsah takové teorie T' je polynomiální vzhledem k původní teorii.

[6] Třída množin je ortogonální, když pro libovolnou dvojici množin X, Y neplatí $X \subseteq Y$ ani $Y \subseteq X$.

4.2 Interpretace defaultů pomocí procesů

Znovu teorie s defaulty

Co jsou to defaulty? Podle publikace Grigoris Antoniou: Nonmonotonic reasoning. The MIT Press, Cambridge, Mass 1998.

Default je pravidlo tvaru

$d = \frac{\alpha(\overrightarrow{x});\beta(\overrightarrow{x})}{\gamma(\overrightarrow{x})}$ je schema defaultů

α, β, γ - uzavřené formule
$T = [F, D]$, kde D je spočetná množina defaultů

Př.

$$\frac{pt\acute{a}k(x); l\acute{e}t\acute{a}(x)}{l\acute{e}t\acute{a}(x)}$$

není default (co je to otevřený default?).

Schema defaultů vymezuje množinu defaultů

$$\frac{\alpha\sigma; \beta\sigma}{\gamma\sigma}$$

pro všechny takové základní (ground) substituce σ že přiřazují hodnotu všem proměnným vyskytujícím se ve schematu. To znamená, že volné proměnné jsou interpretované jako univerzálně kvantifikované přes celé schema.

Proč není vhodné interpretovat otevřené defaulty jako univerzálně kvantifikované? Jestliže všechna x jsou ptáci a o každém x lze předpokládat, že létá, pak uzavřeme, že všechna x létají. To je neintuitivní, znamená to, že default může být aplikován pouze když každý objekt je pták a není-li znám nelétající pták.

Existenční interpretace: Jestliže existuje pták a ex. x tž. létá, pak uzavřeme, že ex. létající objekt.

Při této interpretaci nelze uzavřít např., že $létá(x)$ na základě faktu $pták()$. Můžeme pouze uzavřít, že $\exists x létá(x)$.

Dosud nebyla nalezena vhodná interpretace otevřených defaultů.

$\Pi = [d_0, d_1, \ldots]$ uspořádaná množiny defaultů (bez opakování).

$\Pi[k]$ počáteční úsek délky k .

$In(\Pi) = Cn(F \cap c(d) : d$ se vyskytuje v $\Pi)$

$Out(\Pi) = \{\neg\beta : \beta \in just(d)$ pro nějaké d vyskytující se v Π $\}$

Příklad.

$T = [F, D]$

$F = \{a\}$,a je výroková konstanta.

$D = \left\{d_1 = \frac{a;\neg b}{\neg b}, d_2 = \frac{b:c}{c}\right\}$

1) pro

$$\Pi = [d_1]$$

máme

$$In(\Pi) = Cn(\{a, \neg b)\}$$

$$Out(\Pi) = \{b\}$$

2) pro

$$\Pi = [d_2, d_1]$$

máme

$$In(\Pi) = Cn(\{a, c, \neg b\})$$

$$Out(\Pi) = \{[\neg c, b\}$$

Okamžitá (current knowledge) znalost:

$$b \notin In(\lceil]) = Cn(F) = Cn(\{a\})$$

Okamžitá znalost před aplikací d_2.

PROCES: (vzhledem k T)

$\Pi = \lceil d_0, d_1, \ldots, d_k, \ldots, d_n]$ je proces vzhledem k T iff d_k je aplikovatelné na $In(\Pi[k])$ pro každé k tž. d_k se vyskytuje v Π.

Definice

- Π je úspěšný právě tehdy, kdyĹ

  $$In(\Pi) \cap Out(\Pi) = \emptyset,$$

  jinak je neúspěšný. (To znamená, že žádná funkce tvaru $\neg\beta$ není prvkem množiny okamžitých znalostí). Tudíž bylo konzistentní předpokládat β.

- Π je uzavřený právě když každý default $d \in D$, který je aplikovatelný na $In(\Pi)$ se již vyskytuje v Π (intuitivně F je uzavřená na defaulty).

EXTENZE

Množina funkcí E je extenzí teorie $T = [F, D]$ iff existuje uzavřený a úspěšný proces Π pro T, tž. $E = In(\Pi)$.

$$T = [F, D], \quad F = \{a\}$$

$$D = \left\{ d_1 = \frac{a; \neg b}{d}, \frac{true; c}{b} \right\}$$

$\Pi_1 = [d_1]$ je úspěšný, ale není uzavřený, neboť d_2 je možno aplikovat na $In(\Pi) = Cn(\{a, d\})$.

$\Pi_2 = [d_1, d_2]$ je uzavřený, ale není úspěšný, neboť ex. b

$$In(\Pi_2) = Cn(\{a, d, b\})$$

$$Out(\Pi_2) = \{b, \neg c\}$$

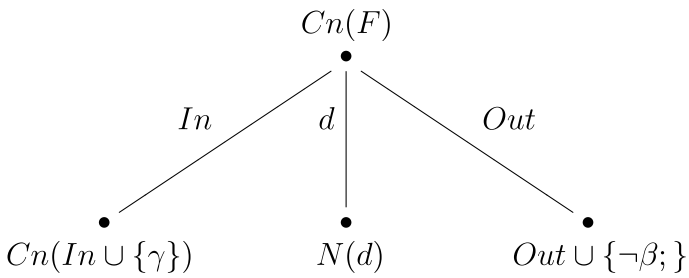
$\Pi_3 = [d_2]$ je uzavřený i úspěšný (vzhledem k T). takže $In(\Pi_3) = Cn(\{a, b\}$ je (jediná) extenze teorie T.

Lemma 4.2.1 *Nekonečný proces* Π *je uzavřený iff každý default, který je aplikovatelný na* $In(\Pi[k])$ *pro nekonečně mnoho přirozených čísel k je (již)* Π.

Důkaz
Využije se věta o kompaktnosti predikátového počtu.
spravedlivost, nestrannost - concurrent programming.

- β je konzistentní s $In(\Pi)$
- β je konzistentní s $In(\Pi[k])$ pro nekonečně mnoho k
- β je konzistentní s $In(\Pi[k])$ pro všechna $k > k'$ (pro nějaké k').

PROCESS TREE
Uzel N je expandován pouze, když $In(N) \cap Out(N) = \emptyset$.



PŘÍKLADY

$$T = [F, D], \quad F = \emptyset, \quad D = \left\{ d = \frac{true; a}{\neg a} \right\}$$

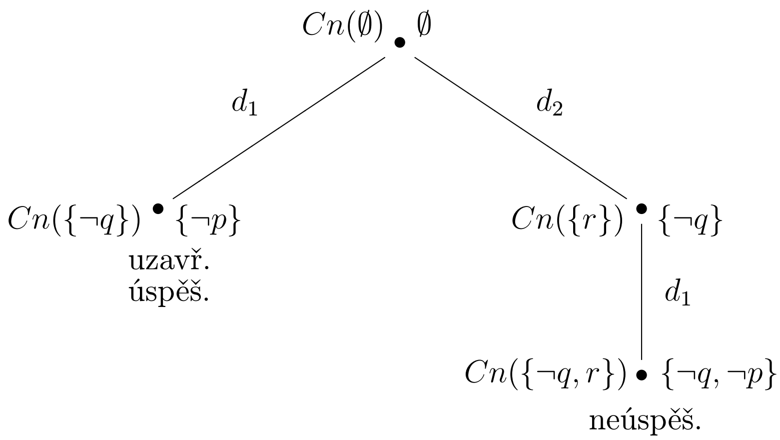
$$\begin{tikzcd} Cn(\emptyset)\ \bullet\ \emptyset \arrow[d, dash, "d"'] \\ Cn(\{\neg a\})\ \bullet\ \{\neg a\} \end{tikzcd}$$

Neexistuje rozšíření.

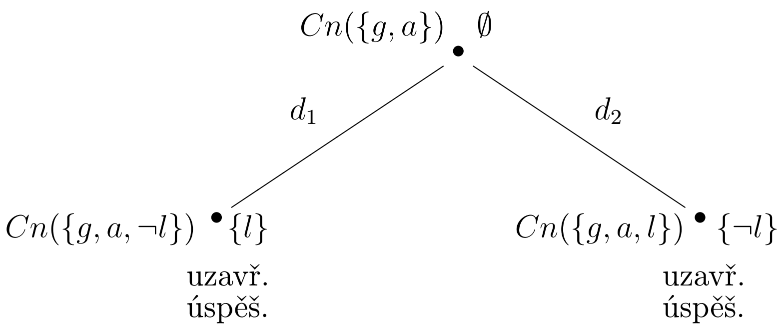
$$T = [F, D], \quad F = \{\neg p, q\}, \quad D = \left\{ d = \frac{q; \neg r}{p} \right\}$$

$$\begin{tikzcd} Cn(\{\neg p, q\})\ \bullet\ \emptyset \arrow[d, dash, "d"'] \\ \text{inconsistentní } r\ \bullet\ \{r\} \end{tikzcd}$$

$$T = [F, D], \quad F = \emptyset, \quad D = \left\{ d_1 = \frac{true; p}{\neg q}, d_2 = \frac{true; q}{r} \right\}$$



$$T = [F, D], \quad F = \{g, a\}, \quad D = \left\{ d_1 = \frac{g; \neg c}{\neg c}, d_2 = \frac{a; c}{c} \right\}$$



4.3 McDermottova–Doyleova logika

McDermottovo a Doyleovo společné řešení [89] je nemonotónní logika, která patří mezi modální logiky. Pro tento účel se zavádějí modality "zdůvodňování" (arguing) anebo "konzistence", které jsou přímo vyjádřeny v objektovém jazyce. Jestliže např. chceme formalizovat výrok *"Typický Švéd je protestant."* můžeme toho dosáhnout tak, že modální operátor $\Diamond$ bude formalizovat výraz "*může být konzistentně předpokládáno, že ...*" a tudíž náš výrok bude dobře formalizován formulí prvního řádu

$$\forall x(S(x) \wedge \Diamond P(x) \Rightarrow P(x)).$$

4.3.1 Výchozí pojmy

Nechť A je pevně daná množina formulí prvního řádu, kterým budeme říkat axiomy a nechť X je libovolná množina takových formulí. V dalším bude Cn znamenat operaci klasického logického důsledku v nějaké modální logice, např. S5, S4 apod.

Definice 4.3.1 *Množinu $Ass_A(X)$* presupozicí teorie X vzhledem k A *pak definujeme tak, že*

$$Ass_A(X) = \{\Diamond\varphi : \neg\varphi \notin X\} - Cn(A)$$

a nemonotónní operátor N_A *pro množinu X a axiomy A je definován rovností*

$$N_A(X) = Cn(A \cup Ass_A(X)).$$

Zhruba řečeno, pevný bod operátoru N_A je "uzávěr" množiny výchozích axiomů A s přidaným pravidlem (*possibilitatum*), tj. pravidlem, zavedení operátoru možnosti:

$$\frac{\neg\varphi \notin Cn(A)}{\Diamond\varphi}$$

Definice 4.3.2 Nemonotónní teorie T *je pevný bod operátoru* N_A*, tj. je to taková množina formulí, pro kterou platí rovnost* $N_A(T) = T$*.*

Tato definice nemonotónní teorie (vše definováno vzhledem k axiomům A) zaručuje, že T obsahuje všechny axiomy, všechny logické důsledky axiomů a maximální počet formulí tvaru $\Diamond\varphi$, které mohou být bezesporně přidány a samozřejmě též všechny logické důsledky všech těchto tří množin formulí.

Příklady: Nechť

$A_1 = \emptyset$. Potom existuje právě jeden pevný bod, který obsahuje všechny tautologie spolu s formulemi tvaru $\Diamond\varphi$, kde φ nevede ke sporu.

$A_2 = \{\Diamond\varphi \Rightarrow \varphi\}$. Zde existuje právě jeden pevný bod, který obsahuje jak formuli $\Diamond\varphi$ tak formuli φ.

$A_3 = \{\Diamond\varphi \Rightarrow \varphi, \neg\varphi\}$. I v tomto případě existuje právě jeden pevný bod uzávěrové operace N_A, který ale teď neobsahuje ani formuli $\Diamond\varphi$ ani φ. A poznamenejme, že $A_2 \subset A_3$.

$A_4 = \{\Diamond\varphi \Rightarrow \neg\psi, \Diamond\psi \Rightarrow \neg\varphi\}$. Zde existují dva pevné body. F_1, který obsahuje formuli $\Diamond\varphi$, ale nikoli formuli $\Diamond\psi$ ani $\neg\varphi$ a F_2, který obsahuje $\Diamond\psi$ i $\neg\varphi$, ale ne formuli $\Diamond\varphi$ ani $\neg\psi$.

$A_5 = \{\Diamond\varphi \Rightarrow \neg\varphi\}$. V tomto případě neexistuje žádný pevný bod. Je sice pravdou, že axiom není kontradiktorický, ale je poněkud "pathologický". Srovnej "patologické defaulty Reiterovy logiky.

Konečně můžeme definovat množinu všech *nemonotónních* důsledků daných axiomů.

Definice 4.3.3 *Nechť* A *je množina formulí a nechť* $fixp(A)$ *je množina všech pevných bodů pro* A*. Množinu* $C(A)$ *všech ne-*

monotónních důsledků axiomů A definujeme následující rovností

$$C(A) = \{\varphi : \varphi \in F \ \ a \ \varphi \in fixp(A)\}.$$

Poznámka: Nechť výraz $\Box\varphi$ je standardní zkratka pro $\neg\Diamond\neg\varphi$. McDermott ukázal [90], že jeho nemonotónní verze modální logiky, pro kterou platí

1. $\Box\varphi \Rightarrow \varphi$
2. $\Box(\varphi \Rightarrow \psi) \Rightarrow (\Box\varphi \Rightarrow \Box\psi)$
3. $\Diamond\varphi \Rightarrow \Box\Diamond\varphi$

je ekvivalentní monotónní verzi dobře známé modální Lewisovy logiky S5. To mj. znamená, že je třeba obrátit pozornost ke slabším modálním systémům, např. k systému S4 nebo v tomto kontextu důležitějšímu Gödelově systému T, který je vlastní částí systému S5 (platí v něm výše uvedené podmínky 1 a 2). O modálních systémech viz kap. XX.

4.3.2 Vztah McDermottovy–Doyleovy logiky k Reiterově logice defaultů

Existuje přímá cesta jak “přeložit” defaultová pravidla do modální logiky: Pravidlo

$$d = (\alpha(x); \beta(x)/\gamma(x))$$

se převede na formuli

$$\forall x(\alpha(x) \wedge \Diamond\beta(x) \Rightarrow \gamma(x)).$$

Avšak obě definice extenze, tj. definice extenze v Reiterově logice defaultů a v McDermottově–Doyleově logice se významně od sebe odlišují. Například teorie $T = [F, D]$ s jednoprvkovou množinou faktů $F = \{a \vee c\}$ a s rovněž jednoprvkovou množinou defaultů $D = \{(a; b/b)\}$ má podle Reitera extenzi $E = Cn(\{a \vee c)\}$, zatímco odpovídající McDermott–Doyleova modální teorie má extenzi, která obsahuje formuli $b \vee d$.

Poznamenejme, že existují teorie s defaulty, které mají extenze, zatímco jim odpovídající modální teorie extenze nemají a naopak.

Doplnit příklady!!!

4.4 Epistemické logiky

Řecké slovo $\varepsilon\pi\iota\sigma\tau\eta\mu\eta$ (epistémé), které znamená *znalost* je velmi staré. Jeho význam, který je v současnosti tolik populární především ve znalostním inženýrství, byl zkoumán skoro všemi filozofickými školami, které aspoň trochu něco znamenaly v širokém filozofickém proudu racionálního myšlení. Studium znalostí, jejich struktury, původ a omezení, má v západní filozofii více než dvě milenia dlouhou tradici. Nyní, v poněkud odlišném kontextu, je koncept znalostí znovu objevován právě znalostními inženýry z velmi praktických důvodů. Co ale znalosti opravdu jsou, to je stále tématem diskusí.

4.4.1 Logická rekonstrukce znalostí

Není nic užitečnějšího pro praxi než dobrá teorie.
Dines Bjørner

V moderní epistemologii je obvyklé, že teorie znalostí je rekonstruována jako jistý druh modální logiky. V moderní formální matematické tradici je epistemická logika chápána jako formální modální logika se speciálními epistemickými axiomy. Modální logika, která je v takovém případě zde v pozadí úvah, se týká jak modality možnosti tak modality nutnosti, které byly zkoumány již ve starověku. Ve všech epistemologických přístupech jsou tyto modality duální. To znamená, že jsou vzájemně definovatelné "ne nutně φ" znamená totéž co "možná $\neg\varphi$". Další vlastnosti epistemických modalit se ale mohou lišit.

Logické studium epistemických aspektů znalostí přináší, jak jinak, řadu otázek. Někteří lidé se domnívají, že logika je tím jediným nezbytným teoretickým základem, jiní dokonce věří, že dobrá modální logika je právě tím nástrojem, který je vhodný jak pro porozumění znalostem, tak pro jejich efektivní zpracování. Stále ale není shoda v tom, jaké epistemické modality jsou ty pravé pro porozumění znalostem a přesvědčením především tehdy, kdy uvažujeme o jejich změnách.

Naskýtá se tak poněkud skromnější otázka:

> *Co můžeme získat ze starších filozofických zkoumání a ze současných zkoumání čistě formálních matematických modalit?*

Odpovědět na tuto otázku adekvátně znamená poučit se též z vývoje odpovídajících logických konceptů ve starších dobách.

4.4.2 Starověké a středověké zdroje epistemických modalit

Pohlédneme-li zpět k počátkům formální logiky, tj. do aristotelských dob, nalézáme modality nutnosti a možnosti spolu s modalitou kontingence. Nutnost a možnost jsou čistě logické koncepty vztahující se k věčným pravdám. Jak už jsme se zmínili v první kapitole, byl to převažující filozofický náhled silně později ovlivněný nejen autoritou Aristotelovou, ale později především autoritou Tomáše Akvinského. Co je však relevantní právě pro moderní informační obsah (ve smyslu mimologických znalostí) sentence, to je koncept kontingence. Žádný z obou logických konceptů samostatně, ani oba dohromady nestačí. To je patrné při každém hledání vhodné formalizace.

Ve středověkém chápání byly obecně přijímány následující aristotelské principy pro modality nutnosti a možnosti:

> Jestliže není možné φ, potom $\neg\varphi$.
> (*Ab non posse ad non esse valet consequentia.*)
>
> Předpokládáme-li $\neg\varphi$, pak není možné φ.
> (*Unumquodque, quando est, oportet esse.*)

Pokusíme-li se formalizovat tyto principy v jednoduchém výrokovém jazyku prvního řádu, dostaneme následující formule, které ukazují nežádoucí důsledky, podle nichž modální i nemodální formule jsou *de facto* ekvivalentní:

$$\neg\Diamond\varphi \Rightarrow \neg\varphi$$

$$\neg\varphi \Rightarrow \neg\Diamond\varphi$$

Tento přístup k formalizaci průkazně ukazuje na to, že formalizace nebude jednoduchá.

Jiný přirozený princip říká, že existují výroky, pro které mohou nastat obě možnosti, tj. je možné, že φ a je též možné, že $\neg\varphi$. Formálně, existují takové formule, že $\Diamond\varphi \wedge \Diamond\neg\varphi$. Tento princip vyjadřuje kontingenci každé elementární neboli prvotní propozice.

Jako obvykle můžeme přidat dvě inferenční pravidla čisté modální logiky *de dicto*:

1. Z formule $\Box(\varphi \Rightarrow \psi)$ odvoď $\Box\varphi \Rightarrow \Box\psi$.
2. Z formule $\Box(\varphi \Rightarrow \psi)$ odvoď $\Diamond\varphi \Rightarrow \Diamond\psi$.

Operátory $\Box$ a $\Diamond$ jsou duální operátory, každý z nich může být definován pomocí negace z toho druhého, jak už jsme si všimli dříve. Znamená to, že koncept negace je v tomto kontextu mimořádně důležitý.

Již letmý pohled zpět ukáže, že i ve středověké filozofii koncepty *znalostí a přesvědčení* (belief) byly široce považovány za částečně analogické nutnosti a možnosti.[7] Tato tradice pokračuje i v našich dnech.

V otázkách k Aristotelovým Prvním analytikám (Pseudo-Scotus) můžeme nalézt různé epistemické koncepty jako např. *verum,*

[7] Viz reprezentativní publikaci Simo Knuuttila: *Modalities in Medieval Philosophy*, Routledge, London and New York, 1993.

falsum, per se, scitum, dubitum, opinatum, creditum, apparens, notum, volitum, dilectum [8], některé z nich silně připomínající moderní pojmy kredibility, fuzzyfikace a další koncepty studované v současném znalostním inženýrství a příslušných modálních logikách.

Koncept *znalosti* by podle středověkých představ měl též splňovat princip:

$$K_a\varphi \Rightarrow B_a\varphi$$ [9]

někdy nazývaný *Occamovo jádro* epistemické logiky.

Začneme-li s pojmen přesvědčení (*believability*) B jako s pojmem výchozím, pak znalost či vědomost (*knowledge*) lze definovat takto

$$K_a\varphi =_{def} B_a\varphi \wedge \varphi \ .$$

To znamená, že znalost lze chápat jako pravdivé přesvědčení[10]. Někteří autoři jako např. právě Occam, preferovali následující jemnější definici

$$K_a\varphi =_{def} B_a\varphi \wedge \varphi \wedge J_a\varphi \ ,$$

kde $J_a\varphi$ znamená výrok *Agent a je přesvědčem* (justified in believing)*, že* φ.

Podobnými otázkami se zabývali i mnozíı dal vsí filozofové a logikové. Například Bernard Bolzano později ve Wissenschaftslehre $ 307 též zavedl přesnější definice pojmů *znalost*, *ignorance*, a *klam (fallacy)*. A v jiném paragrafu popisuje koncepty *measure of belief* a *measure of disbelief*, termíny dobře známé tvůrcům tzv. expertních systémů.[11]

[8]Viz. John Duns Scotus: *Opera Omnia I.*

[9]Zde formuli $K_a\varphi$ je třeba číst jako *Agent a ví, že* φ. Analogicky formuli $B_a\varphi$ je třeba číst jako *Agent a je přesvědčen (věří), že* φ.

[10]Tento předpoklad neznamená nutně kompatibilitu s moderním pojmem "fuzzyness".

[11]Viz. Znalostní expertní systém MYCIN diskutovaný ve [152].

Obraťme nyní svoji pozornost k současné tradici ve formální matematické modální logice, která může být interpretováma jako epistemická logika.

Někteří logikové pracující v epistemických logikách, tj. v logice znalostí nebo v logice přesvědčení obvykle začínají úvahami o úsudcích s modalitami splňujícími tři následující a snad přirozené podmínky (srovnej [158]):

- Agent, tj. nositel znalostí, má plnou schopnost introspekce. To znamená, že když ví/věří, že φ, pak též $K\varphi/B\varphi$ (pozitivní introspekce) a když neví/nevěří, že φ, pak také $K\neg\varphi/B\neg\varphi$ (negativní introspekce).

- Agent v úsudcích používá modální epistemickou logiku tak, že začíná z počátečních předpokladů a z modálních formulí, které vyjadřují to, co je mu známo nebo o čem je přesvědčen nebo co může předpokládat či co je pro něj konzistentní předpokládat.

- Agent přiřazuje jednotlivé nekontradiktorické položky znalostí anebo celé množiny (avšak jen konzistentní množiny) znalostí či přesvědčení základní epistemické logice nebo logice přesvědčení.

Definice 4.4.1 *Stabilní množina znalostí je taková teorie, která je uzavřená jak na pozitivní tak i na negativní introspekci.*

Všechny výše uvedené podmínky jsou často užívány v různých nemonotónních systémech jako kupř. v defaultových logikách, logikách s0mantických omezení (circumscriptions) a jiných. Avšak tyto podmínky nejsou realistické v případě praktických aplikací, protože předpokládají logickou omniscienci. Logická omniscience není relistická, neboť usuzovací kapacita skutečných lidských ale i umělých agentů je omezená. Předpoklad, že agent může okamžitě ustavit všechny logické důsledky daných znalostí či přesvědčení naráží buď na velkou (výpočtovou) složitost, nebo je extrémně "time-consuming", nebo je dokonce intraktabilní.

Navíc je třeba vzít v úvahu, že potřebujeme logiku znalostí, která zohlední neúplnost jednotlivých položek našich znalostí či přesvědčení spolu s omezenou schopností odvozovat korektní závěry jak ze znalostí tak z logického zázemí.

4.4.3 Základní modální epistemická logika K

Začneme s epistemickou logikou ve výrokovém jazyku (jazyku nultého řádu) s přidaným modálním symbolem K možná opatřeným podle potřeby indexy agentů. Tak formule $K_a\varphi$ bude čtena jako: *Agent a ví, že* φ. V případě kdy půjde jen o jednoho agenta, indexy nebudeme užívat. Potom přijmeme následující axiomy.

Axiomy:

(**PC**) Všechny tautologie výrokové logiky

(**K**) $K\varphi \wedge K\ (\varphi \Rightarrow \psi) \Rightarrow K\ \psi$
Axiom logické racionality (uzavřenost na logickou implikaci znalostí)

Inferenční pravidla: (Obě pravidla dohromady doplněná uniformní substitucí.)

(**MP**) *Modus ponens*
Z formulí φ a $\varphi \Rightarrow \psi$ odvoď ψ.

(**NEC**) *Necesitace, tj. zavedení modálního operátoru K*
Z formule φ odvoď $K\varphi$.

Některé *silnější modální epistemické logiky* s dalšími přidanými axiomy, které jsou často diskutovány v literatuře):

(**T**) $K\varphi \Rightarrow \varphi$
Znalostní axiom (znalost implikuje pravdivost)

(**D**) $K\varphi \Rightarrow \neg K\ \neg\varphi$
Axiom konzistence znalostí
Jiná přirozená a někdy používaná formulace: $\neg K\ (\varphi \wedge \neg\varphi)$.

(**4**) $K\varphi \Rightarrow K\ K\ \varphi$
Pozitivní introspekce

(**5**) $\neg K\ \varphi \Rightarrow K\ \neg K\ \varphi$
Negativní introspekce

Snadno lze nahlédnout, že axiom **5** je silnější než axiom **4**. To odpovídá známé hierarchii modálních systémů Carl I. Lewise, jmenovitě systémům **S4** a **S5**.

Další, méně frekventované axiomy jsou známy ze zkoumání různých druhů modalit, a to nejen epistemických. Podrobnosti může čtenář nalézt např. v publikaci [166].

Přítomnost nebo nepřítomnost axiomu **T** je v kontextu epistemick7ch logik významná, protože odlišuje *logiky znalostí* od *logik přesvědčení*. Znalost musí být vždy pravdivá, zatímco přesvědčení může být i nepravdivé. Navíc, axiom **T** garantuje konzistenci znalostí jejich nositele.

Teď je třeba poznamenat, že konverze axiomů **D** a **4**, tj. formule

$$K\neg\varphi \Rightarrow \neg K\ \varphi \text{ a } \neg K\ K\ \varphi \Rightarrow \neg K\ \varphi$$

jsou vlastně instancemi axiomu **T**. Takže z axiomů **4** a **5** spolu s jejich konverzemi dostáváme

$$KK\ \varphi \Leftrightarrow K\ \varphi \text{ a } K\neg K\ \varphi \Leftrightarrow \neg K\ \varphi.$$

Tyto principy se zdají být racionálními, když modalitu K interpretujeme epistemologicky.

K tomu ještě poznamenejme, že je pouhou technickou záležitostí přidat pravidlo monotónnosti a pravidlo kongruence:

MON Z formule $\varphi \Rightarrow \psi$ odvoď $K\varphi \Rightarrow K\ \psi$.
Pravidlo monotónnosti

CGR Z formule $\varphi \Leftrightarrow \psi$ odvoď $K\varphi \Leftrightarrow K\ \psi$.
Pravidlo kongruence
a pravidlo

RKn Z formule $\varphi_1 \wedge \ldots \wedge \varphi_n \Rightarrow \psi$ odvoď formuli
$K\varphi_1 \wedge \ldots K\ \ \wedge K\ \varphi_n \Rightarrow K\ \psi$ pro všechna n.
Pravidlo zavedení znalostního operátoru

Dosud by nebylo obtížné generalizovat tyto axiomy pro multiagentové systémy parametrizací znalostního operátoru individuálními agenty. Ale to není adekvátní řešení, protože reální (lidští) agenti nemají obvykle shodné znalosti. Můžeme hovořit pouze o společných znalostech (common knowledge), tj. takové části znalostí, která je společná všem agentům. O tom bude ještě řeč později.

Jiný aspekt neadekvátnosti standardní modální epistemické logiky spočívá v tom, že žádný reálný agent nezná aktuálně všechny logické důsledky všech svých vlastních znalostí.

Někteří autoři navíc navrhují poněkud odlišné interpretace formule $K_a\varphi$. Nejslibnější jsou interpretace jako "agent a implicitně zná φ", "agent a možná zná φ", "agent a je schopen využít informaci obsaženou ve φ" a případně další interpretace. Porozumět takovým interpretacím dobře, znamená většinou zeslabit dosud diskutované "klasické" epistemické systémy.

Jindy je naopak vhodné přidat další axiomy jako třeba

(**B**) $\varphi \Rightarrow K \neg K \neg\varphi$

(**W5**) $\neg K \neg K\ \varphi \Rightarrow (\varphi \Rightarrow K\ \varphi)$

(**F**) $(\varphi \wedge \Diamond K\ \psi) \Rightarrow K(\Diamond\varphi \vee \psi)$

Axiom **F** hraje důležitou roli ve stabilní sémantice logických programů v nemonotónním systému S4F neboť disjunktivní defaultová logika může být vnořena do nemonotónní modální logiky [158].

V literatuře o epistemických logikách a v literatuře o modálních logikách obecně jsou nejčastěji studovány následující systémy:

Číslo	Standardní název modální logiky	Množina axiomů
1	N	žádné axiomy
2	T	**T**
3	B	**B**
4	5	**5**
5	K	**K**
6	D	**K, D**
7	KD45	**K, D, 4, 5**
8	T	**K, T**
9	S4	**K, T, 4**
10	S4F	**K, T, 4, F**
11	S4W5	**K, T, 4, W5**
12	S5	**K, T, 4, 5**

Tab. 1

Je zřejmé, že pouze logiky 2 a 8 až 12 mohou pretendovat na to, aby byly nazývány logikami znalostí, zatímco logiky 5 až 7 jsou normální logiky přesvědčemí.

Logika N je nejslabší modální logika a logika K je nejslabší logika znalostí. Protože obě jsou velmi slabé nemají reálnou šanci, aby mohly být rozumně interpretovány epistemicky. Jen málo je známo o logikách přesvědčení, které jsou extenzemi axiomu **D**.

4.4.4 Kripkovské sémantiky

Pro modální logiky obecně existuje dobře propracovaná sémantika známá jako kripkovská símantika. Ta sestává z tzv. kripkovského rámce, tj. neprázdné množiny W možných světů a binární relace R (relace dosažitelnosti), taková, že $R \subseteq W \times W$. Různé modální logiky jsou pak charakterizovány různými relacemi dosažitelnosti, které pro každý aktuální možný svět udávají světy z něj dosažitelné.

Je známo, že logika K je charakterizována třídou všech kripkovských modelů. To je také důvod, proč všechny modální logiky, které obsahují axiom **K** se nazývají normální modální logiky, zatímco všechny ostatní se nazývají subnormální. Takže všechny logiky znalostí jsou normální logiky vzhledem ke kripkovským modelům.

Logika T je charakterizována třídou všech reflexivních kripkovských modelů, logika S4 třídou všech reflexivních a tranzitních kripkovských modelů, logika S5 třídou všech kripkovských modelů, jejíž relace dosažitelnosti je relací ekvivalence, logika KD45 třídou kripkovských modelů, které jsou tranzitivní, eukleidovské a nemají žádný dead ends, logika S4F by ... , logika SW5 by ... .

Z hlediska počítačových aplikací je důležité, že většina z výše uvedených modálních logik je rozhodnutelná.[12]

[12] Je to např. známo, pro teorie K, T, S4, S5, KD, KD4, KD45. Viz např. [144], [151] a další.

4.4.5 Paradox poznatelnosti a logicky omniscientní agenti

Je několik verzí dobře známého paradoxu poznatelnosti (knowability paradox) [146]. Verze, která říká, že každá pravdivá věta je poznatelná, může být snadno odvozena z formule $\varphi \Rightarrow \Diamond K\ \varphi$, která říká, říká, že každá pravdivá sentence je poznatelná, což je opět dobře známá anti-realistická téze, řečeno filozofickou terminologií (Viz. [159]).

Když přidáme distributivnost operátoru K vzhledem ke konjunkci, tj.

$$K(\varphi \wedge \psi) \Rightarrow (K\ \varphi \wedge K\ \psi)$$

a tvrzení, že znalost implikuje pravdivost, tj. axiom **T**, že pouze pravdivé výroky mohou být známy a formuli $\neg\Diamond(K\ \varphi \wedge \neg K\ \varphi)$ přijdeme k neakceptovatelnému tvrzení, že každá sentence je známa.

Paradox znalostního operátoru K lze odvodit různými cestami. Zde je jedna z mnoha syntaktických cest.

1. $K(\varphi \wedge \neg K\ \varphi) \Rightarrow (K\ \varphi \wedge K\ \neg K\ \varphi)$ (distribuce znalostního operátoru vzhledem k implikaci)
2. $K(\varphi \wedge \neg K\ \varphi) \Rightarrow K\ \varphi$ (z formule 1)
3. $K(\varphi \wedge \neg K\ \varphi) \Rightarrow K\ \varphi(\neg K\ \varphi)$ (z formule 1)
4. $K(\neg K\ \varphi) \Rightarrow \neg K\ \varphi$ (axiom T)
5. $K(\varphi \wedge \neg K\ \varphi) \Rightarrow \neg K\ \varphi$ (z formulí 3 a 4)
6. $K(\varphi \wedge \neg K\ \varphi) \Rightarrow (K\ \varphi \wedge \neg K\ \varphi)$ (z formulí 2 a 5)
7. $\neg K\ (\varphi \wedge \neg K\ \varphi)$ (z formule 6)

8. $\Box\neg K\ (\varphi \wedge \neg K\ \varphi)$ (ze 7 a zavedením oprátoru nutnosti)

9. $\neg\Diamond(\varphi \wedge \neg K\ \varphi)$ (z formule 8 a duality $\Box$ a $\Diamond$)

10. $(\varphi \wedge \neg K\ \varphi) \Rightarrow \Diamond K\ (\varphi \wedge \neg K\ \varphi)$

11. $\neg(\varphi \wedge \neg K\ \varphi)$ (z formulí 9 a 10)

Bod 11 lze ekvivalentně přeformulovat na výrok $\varphi \Rightarrow K\ \varphi$, podle kterého, je-li φ pravdivý výrok, pak φ je znám, což je jistě nežádoucí důsledek.

Jsou i různé cesty jak paradox poznatelnosti odstranit. Jednou z těchto cest je různé užití operátoru negace. Negace se totiž v kontextu epistemických sentencí chová odlišně než klasická logická negace. Chová se spíše jako *podmínka uzavřeného světa*, která pochází původně ze světa teorie databází. Jestliže agent neví φ, předpokládá se $\neg\varphi$. Tento předpoklad má různé varianty. Nejvýznemnější i nejznámnější je užívána v logických programech pod názvem negace jako neúspěch (negation-as-failure).

Na druhé straně, někteří autoři navrhují řešit paradox poznatelnosti pravdivých tvrzení v logice RN4, což je logika, která je epistemickým rozšířením relevantní verze Nelsonovy parakonzistentní logiky N4. [].

Jinou cestou je rozlišení interní a externí negace a vnitřních a vnějších znalostí v topologickém jazyku [161].

Usuzování lidských agentů a dokonce i arteficiálních agentů není logicky omniscientní, tj. nejsou schopni odvodit logické důsledky okamžitě.

> Vzniká tak otázka: Jak odstranit požadavek logické omniscience [146] a tak vytvářet realističtější teorie racionálně usuzujících agentů?

V publikaci [153] autoři zdůrazňují, že

> "When people reason deductively, they start with some information – either evidence of the senses or a verbal description – and thez assess whether a given conclusion follows validly from this information. In real life there is often no given conclusion, and so they generate a conclusion for themselves. Logic alone is insufficient to characterize intelligent reasoning in this case, because any set of premises yield an infinite number of valid conclusions. Most of them are banal, such as the conjunction of a premise with itself, and no sane individual, apart from a logician, would dream of drawing such conclusions. Hence, when individuals make a deduction indaily life, they must be guided by more than logic. They draw useful conclusions. The evidence suggests that they tend to maintain the information conveyed by the premises, to re-express it more parsimoniously, and to establish something not directly asserted in a premise. If nothing meets these constraints, they declare that there is no valid conclusion."

Taková rekonstrukce by měla být nejen nemonotónní, jak je tomu v různých kontextově závislých systémech jako např. v logice defaultů [158], [143]. Ta má být nejen tolerantní vůči inkonzistenci, jak je tomu např. v paraconsistentních logikách [159], ale závěry dosažené při derivování dalších znalostí by měly být jak koherentní tak relevantní daným otázkám. Srovnej [164].

Důvod spočívá mj. v explanatorní síle. Formálně lze tak problém koherence chápat jako úkol disjunktně rozdělit konečnou množinu E událostí opatřenými vahami pro pozitivní a negativní omezení $C = C_+ \cup C_-$ chápanou jako dvojice prvků z E na dvě množiny A (acceptance) a R (rejection) tak, že je maximalizována shoda s následujícími podmínkami koherence:

1. Jestliže $[e_i, e_j] \in C_+$, potom $e_i \in A$ právě tehdy, když $e_j \in A$.

2. Jestliže $[e_i, e_j] \in C_-$, potom $e_i \in A$ právě tehdy, když $e_j \in R$.

Zde je ovšem potíž v tom, že obecně je tato úloha výpočtově obtížná, což znamená, že je ve skutečnosti nerealizovatelná v tom smyslu, že neexistuje algoritmus, který by byl jak dostatečně efektivní, tak i korektní. Jsou ale cesty jak hledat vhodné aproximace např. pomocí neuronových sítí.

4.5 Hierarchie derivovatelností znalostí a přesvědčení

Vraťme se zpět ke standardním logickým přístupům. Když použijeme nějaký druh nemonotónní logiky jako logiku, která tvoří logický základ teorie, můžeme rozlišit různé vztahy derivovatelnosti. Takové vztahy pak musejí být brány do úvahy v procesu aktualizace znalostí.

Definice 4.5.1 *Nechť cn ae klasická monotónní operace logické konsekvence v daném jazyce (tj. množina dobře utvořených formulí) a cnn je nemonotónní operace důsledku nad týmž jazykem. Potom pro teorii T a formuli φ definujeme, že formule φ je*

diskutovatelná*, když existuje takový pevný bod operace cnn, který obsahuje φ;*

představitelná*, když $\neg\varphi \notin cnn(T)$ a*

nepochybná*, když $\neg\varphi$ není diskutovatelná.*

Lze ukázat, že mezi třídami takto definovaných formulí platí následující inkluze, které jsou vlastními inkluzemi:

DOKAZATELNÉ $\subseteq$ DISKUTOVATELNÉ $\subseteq$ PŘEDSTAVITELNÉ
DOKAZATELNÉ $\subseteq$ NEPOCHYBNÉ $\subseteq$ PŘEDSTAVITELNÉ.

Definice 4.5.2 *Nechť T je konzitentní teorie vzhledem k cn. Potom formule φ se bude nazývat*

bezpečná *vzhledem k T, jetliže $\varphi \in cnn(T')$ pro všechny takové konzistentní teorie T', že $T \subseteq T'$;*

vynutitelná*, jestliže $\neg\varphi$ není bezpečná;*

plauzibilní *(nebo* předpokládatelná (assumable)*) vzhledem k T, jetliže $T \cup \varphi$ je konzistentní;*

nekontroverzní*, když $\neg\varphi$ není plauzibilní;*

realizovatelná*, když existuje taková teorie T', že $T \subseteq T'$ a $\varphi \in T'$;*

nezamítnutelná*, když $\neg\varphi$ není realizovatelná.*

Analogicky lze ukázat, že

DISKUTOVATELNÉ $\subseteq$ PLAUZIBILNÍ $\subseteq$ REALIZOVATELNÉ $\subseteq$ VYNUTITELNÉ
PŘEDSTAVITELNÉ $\subseteq$ VYNUTITELNÉ

Tyto skutečnosti dokumentují obrovskou složitost možných hierarchií různých typů “odvoditelnosti”, která musí být brána v úvahu pracujeme-li s různými úrovněmi znalostí a přesvědčení dokonce i v případě jediného nositele znalostí. To nám může přivést k novým podnětům pro studium vhodných a realističtějších systémů.

4.6 Autoepistemické logiky

Ego scio, me nihil scire. (Vím, že nic nevím.)
Sokrates

V autoepistemických logikách se snažíme modelovat (podobně jako v McDermottově a Doyleově logice) opět přesvědčení (*beliefs*) ideálního racionálního nositele znalostí, který stav svých znalostí reflektuje. Předpokládáme, že ten, kdo usuzuje, je schopen introspekce. Znamená to, že náš odvozovací systém by měl být rovněž schopen provádět logická odvození, která se týkají stavu znalostí. Systém by měl být schopen rozlišit, co nositel znalostí ví a co nikoli. Měl by být schopen logických introspektivních inferencí. Idealizace se pak týká jak toho, že jsou odvozovány pouze takové závěry, které jsou očekávané na základě výchozích základních faktů, tak také toho, že všechny takové závěry berou v úvahu právě ony introspektivní rysy argumentů, které mají charakter jednak pozitivní jednak negativní introspekce.[13] Idea autoepistemické logiky v této verzi pochází od R. M. Stalnakera [130] a byla později rozvinuta R. C. Moorem [93].

Abychom dobře porozuměli autoepistemickým logikám, je třeba rozlišit dvě úrovně: externí a interní.

Interní logika se týká aktivního nositele (*agenta*) znalostí či přesvědčení a způsobů jak svá přesvědčení vyjadřuje, jak vyjadřuje své vlastní znalosti a inference. Externí logiku vytváří pozorovatel k tomu, aby popsal přesvědčení jiného nositele či nositelů znalostí a přesvědčení a aby porozuměl jejich inferencím.

V autoepistemické logice obvykle opět zbohatíme jazyk o jeden unární modální operátor $\Box$ (píšeme $\Box\varphi$), který je interpretován jako "*agent věří, že* φ" nebo "*agent je přesvědčen, že* φ

[13] To je hlavní odlišnost od nemonotónní logiky McDermotta a Doylea, která reflektuje pouze negativní introspekci.

anebo duální operátor $\Diamond$ (píšeme $\Diamond\varphi$), který bývá interpretován jako "*agent nevěří, že* $\neg\varphi$" nebo *agent si není jist, že* $\neg\varphi$. Výroková varianta jazyka autoepistemické logiky vznikne z jazyka $\mathcal{L}_0$ klasické výrokové logiky přidáním následujícího transformačního pravidla[14]:

$$\text{Jestliže } \varphi \in \mathcal{L}, \text{ potom } \Box\varphi \in \mathcal{L}.$$

Příklad: Když φ, ψ jsou výroky, tak i např. výrazy $\Box\varphi$, $\Box(\varphi \wedge \Box\psi)$, $\Box\Box\Box\varphi$ jsou správně utvořené formule. Řekneme, že v první z nich má modální operátor $\Box$ hloubku 1, ve druhé má hloubku 2, a ve třetí formuli má hloubku 3. Hloubkou formule tedy myslíme hloubku zanoření modality $\Box$ ve formuli. Čtenář jistě snadno zformuluje vhodnou induktivní definici.

Poznamenejme, že výrok $\Box\varphi$ je chápán jako *de facto* atomický výrok, jehož pravdivostní hodnota nijak nesouvisí s pravdivostní hodnotou výroku p. To mj. znamená, že může existovat takový model $\mathcal{M}$, že $\mathcal{M} \models \varphi$ a zároveň $\mathcal{M} \not\models \Box\varphi$.

Můžeme tedy říci, že mezi formulemi φ, $\Box\varphi$, a formulemi φ a $\Diamond\varphi$ není žádná sémantická souvislost. Pouze platí vzájemný vztah mezi modalitami, tj. $\Box\varphi$ je totéž co $\neg\Diamond\neg\varphi$ a podobně $\Diamond\varphi$ je totéž co $\neg\Box\neg\varphi$.

Definice 4.6.1 Autoepistemickou teorií T *rozumíme libovolnou množinu formulí jazyka* $\mathcal{L}$*.*

Výraz autoepistemická teorie T budeme také zkracovat symbolem ae-teorie T.

Výroková interpretace autoepistemické teorie T je přiřazení pravdivostních hodnot formulím v T, které respektuje syntaktická pravidla výrokové logiky a které přiřazuje libovolnou hodnotu výroku tvaru $\Box\varphi$.

---

[14] Většina výsledků dosažených v autoepistemických logikách může být ovšem formulována a dokázána i pro formule jazyka prvního řádu.

Standardní výrokový model autoepistemické teorie T je taková výroková interpretace, ve které jsou všechny formule v T pravdivé.

Definice 4.6.2 *Model autoepistemické teorie se nazývá* autoepistemický*, pokud libovolná formule φ tvaru $\Box\chi$ je pravdivá v modelu právě když $\varphi \in T$.*

Nechť $\emptyset \neq A \subseteq Fle$ jazyka $\mathcal{L}$ je množina přesvědčení (základních faktů), tj. je to množina autoepistemických formulí. Definujeme, co znamená, že teorie T respektuje základní výchozí fakta či přesvědčení (beliefs) a co znamená, že tato teorie je sémanticky úplná vůči těmto přesvědčením.

Definice 4.6.3 *Autoepistemická teorie T* respektuje *množinu premis A, jestliže každá ae–interpretace teorie T, jejíž model množiny A je také modelem teorie T.*

Definice 4.6.4 *Autoepistemická teorie T se nazývá* korektní *vzhledem k množině premis A, když každá formule $\varphi \in T$ je pravdivá ve všech modelech respektujících T.*

Opět jinými slovy, jestliže $\varphi \in T$, pak $A \models_T \varphi$.

Definice 4.6.5 *O autoepistemické teorii T řekneme, že je* sémanticky úplná*, když obsahuje všechny formule pravdivé v autoepistemickém modelu teorie T.*

Jinými slovy, jestliže $T \models_T \varphi$, pak $\varphi \in T$.

Definice 4.6.6 Extenze *množiny A premis je teorie*
$$T = \{\varphi : A \models_T \varphi\}.$$

Příklady: Nechť p, q jsou výrokové konstanty.

$A_1 = \{p\}$. Extenze těchto premis obsahuje všechny logické důsledky množiny $\{p\}$ a žádnou jinou formuli bez modality. Tato extenze ale obsahuje formule tvaru $\Box\varphi$, kde φ je důsledek premis p a $\neg\Box\varphi$, kde φ nepatří do extenze premis A.

$A_2 = \{\Box p\}$ nemá žádnou extenzi. To dokážeme redukcí k absurdu: Nechť E je extenze premis A. Potom buď $p \in E$ nebo $p \notin E$, ale to druhé nemůže být. Takže předpokláme, že $p \in E$. Můžeme však konstruovat interpretaci m, ve které je A pravda a p je v ní nepravdivé. Není tedy možné, aby $A \models_E p$, takže $p \notin E$, což je spor s předpokladem.

$A_3 = \{\Box p \Rightarrow p\}$ má takové dvě extenze, že $p \in E_1$ a $p \notin E_2$.

$A_4 = \{\neg\Box p \Rightarrow q, \neg\Box q \Rightarrow p\}$ má takové dvě extenze, že

$\Box p \in E_1$ a $\Box q \notin E_1$

$\Box p \notin E_2$ a $\Box q \in E_2$.

Teď ještě potřebujeme charakterizovat autoepistemické teorie syntakticky. Chceme charakterizovat teorie, které jsou sémanticky úplné a respektují danou množinu A axiomů (např. základních přesvědčení). To znamená, že chceme charakterizovat maximální množiny, které respektují teorii T a které ideální racionální agent může z A odvodit. K tomu využijeme Stalnakerův koncept *stability* [130] teorie. Nejdřív ale zavedeme pomocná označení. Pro každou množinu formulí zavedeme tyto zkratky:

$$\begin{array}{rcl} \Box X & = & \{ \Box\varphi : \varphi \in X \} \\ \neg\Box X & = & \{ \neg\Box\varphi : \varphi \in X \} \\ \Box\bar{X} & = & \{ \Box\varphi : \varphi \notin X \} \\ \neg\Box\bar{X} & = & \{ \neg\Box\varphi : \varphi \notin X \} \end{array}$$

Definice 4.6.7 *Autoepistemickou teorii T nazveme* stabilní*, jestliže má následující vlastnosti:*

1. *s každou (konečnou) podmnožinou $X \subseteq T$ obsahuje též všechny její logické důsledky, tj. $Cn(X) \subseteq T$;*

2. *s každou formulí $\varphi \in T$ obsahuje také $\Box\varphi$;*

3. *pro každou formuli $\varphi \notin T$ také $\Box\varphi \notin T$.*

Definice 4.6.8 *Množina T formulí je* stabilní extenzí *základních přesvědčení (axiomů) A, když*

$$T = \{\varphi : \varphi \in Cn(A \cup \Box T \cup \neg\Box\bar{T}).$$

Tyto definice nám už teď umožní vyslovit následující teorém.

Teorém 4.6.1 *Autoepistemické extenze premis A jsou právě autoepistemické teorie, které jsou korektní a úplné vůči A.*

4.6.1 Překlad defaultů do autoepistemické logiky

Teď ještě předložíme přímou interpretaci defaultů v termínech introspectivních formulí autoepistemické logiky. Nechť $T = [F, D]$ je teorie s (uzavřenými) defaulty ve smyslu Reiterově a nechť A je autoepistemická teorie.

Definice 4.6.9 Standardním překladem *teorie $T = [F, D]$ do autoepistemické logiky A rozumíme takový překlad formulí a defaltů teorie T na formule autoepistemické teorie A, že default $(\alpha; \beta_1, \ldots, \beta_n/\gamma)$ se přeloží na formuli $(\Box\alpha \wedge \neg\beta_1 \wedge \ldots \wedge \neg\beta_n) \Rightarrow \gamma$.*

Takový překlad můžeme parafrázovat následovně: “Jestliže vím (je známo), že α je pravdivá a jestliže nemám žádnou informaci o tom, že některá z formulí $\beta_1, \ldots, \beta_n$ neplatí, potom γ musí platit.

Mohou být ovšem i jiné překlady. Například můžeme uvažovat o tom, že premisa α se přeloží na formuli $\Box\alpha$ a když závěr γ defaultu je prvkem "fixpunktu", bude přeložen na formuli γ. Důvod, proč první překlad je lepší, naznačí následující příklad": Uvažujme množinu $F = \emptyset$ a dva defaulty
$d_1 = (\alpha; \beta/\gamma)$
$d_2 = (; \beta/\alpha \Rightarrow \gamma)$.

Potom teorie T_1 s fakty F a defaulty D_1 nemá extenzi, zatímco teorie T_2 s nějakou neprázdnou množinou faktů, ale s "přeloženými" defaulty d_2 má extenzi $E = Cn(\{\alpha \Rightarrow \gamma\})$.

Extenze teorie v autoepistemické logice je plně určena svým nemodálním jádrem. Z tohoto důvodu je možné srovnávat extenze teorií s defaulty právě podle jejich jader autoepistemických obrazů (tj. ve standardním překladu).

4.6.2 Autoepistemické logiky a možné světy

Významnou úlohu má v autoepistemických logikách a sémantika možných světů Lewisova modálního systému S5.[15] Moore totiž ukázal[?], že autoepistemická teorie T je množinou formulí, které jsou pravdivé v každém světě nějaké úplné struktury $\mathcal{S}$ pro systém S5 právě tehdy, když je to stabilní autoepistemická teorie.

Definice 4.6.10 *Modelem možných světů pro autoepistemickou teorii rozumíme dvojici $\mathcal{S} = [S, val]$, kde S je úplná struktura pro systém S5 a $val : VAR \mapsto \{0, 1\}$ je ohodnocení (tj. zobrazení z výrokových proměnných do pravdivostních hodnot), které určuje, co je pravda v aktuálním světě.*

Příklady: Nechť $A = \{\neg\Box p \Rightarrow q\}$, kde p a q jsou atomické výroky, je množina počátečních přesvědčení.

[15] Podrobněji o modálním systému S5 viz kap. XX.

1. Předpokládejme, že T obsahuje výrok q, ale že neobsahuje p. V tomto případě úplná struktura pro modální systém S5 charakterizující teorii T je
$$S = \{\{p, q\}, \{\neg p, q\}\}.$$
Formule $\neg\Box p \Rightarrow q$ je potom pravdivá pouze ve dvou případech:
$$val_0 = \{p, q\}, \quad val_1 = \{\neg p, q\}.$$
V tomto případě kterékoli z obou ohodnocení odpovídající možnému světu ve struktuře pro $\mathcal{S}$ je teorie T se stabilní expanzí základních přesvědčení A.

2. Předpokládejme, že teorie T obsahuje p, ale neobsahuje q. Potom $S = \{\{p, q\}, \{p, \neg q\}\}$. Druhé ohodnocení, tj. $val_1 = \{\neg p, q\}$ je ovšem autoepistemická interpretace teorie T ve které je pravda A. Avšak protože val_1 neodpovídá žádnému možnému světu ve struktuře $\mathcal{S}$, teorie T nemůže být stabilním rozšířením množiny A.

3. Předpokládejme konečně, že teorie T neobsahuje ani p, ani q. Potom $S = \{\{p, q\}\}$. Tentýž argument (se stejným ohodnocením) nás vede k závěru, který je analogický předchozímu příkladu.

4.7 Preferenční modely

4.7.1 Minimální modely a omezení (circumscription)

Začněme jednoduchým příkladem: Jsou dány dva fakty

$$\text{Marie} \neq \text{Jan a zrzavá(Marie)}$$

a chceme získat další závěry o Janovi. V našich zeměpisných šířkách je vlastnost "být zrzavý" málo frekventovaná, tak chceme spíše (nemonotónně) odvodit, že

$$\neg \text{ zrzavý(Jan)}.$$

Nechvt $\mathcal{M}_1 a \mathcal{M}_2$ jsou modely formule φ a nechť P je predikátový symbol. Řekneme, že $\mathcal{M}_1 \leq_P \mathcal{M}_2$ když

1. $\mathcal{M}_1, \mathcal{M}_2$ mají tutéž doménu;
2. Všechny predikátové symboly, kromě symbolu P, které se vyskytují ve formuli φ mají stejnou extenzi v obou modelech.
3. Extenze predikátu P v $\mathcal{M}_1$ je částí extenze predikátu P v $\mathcal{M}_2$.

Definice 4.7.1 *Model $\mathcal{M}$ nazveme* minimální *(vzhledem k uspořádání $\leq_P$), jestliže pro všechny modely $\mathcal{M}'$ takové, že* *$\mathcal{M}' \leq_P \mathcal{M}$ platí $\mathcal{M}' = \mathcal{M}$.*

Definice 4.7.2 *Formule α* m–vyplývá z *β (vzhledem k predikátu P), jestliže α je pravdivá ve všech modelech formule β minimálních v uspořádání $\leq_P$.*

Teď ovšem vzniká otázka jak charakterizovat m–vyplývání syntakticky.

Jedna z možností je omezit původní premisy a tím eliminovat nežádoucí modely. Nemonotónní důsledky formule φ pak mohou být získány jako monotónní důsledky formule φ a nějaké dodatečné informace. Naším cílem je nalézt takové *skryté předpoklady* charakterizující ona omezení (circumscription), které nám umožní odvodit právě minimální model formule φ s omezením. Obvykle omezením predikátu P rozumíme formuli α druhého řádu a tvaru

$$(\alpha(\Phi) \wedge ((\forall x)\Phi(x) \Rightarrow P(x))) \Rightarrow ((\forall x)P(x) \Rightarrow \Phi(x)),$$

kde $x = \{x_1, \ldots, x_n\}$ a $\alpha(\Phi)$ je výsledek substituce všech výskytů predikátu P ve formuli α parametrem Φ. To lze interpretovat jako "kontext" v němž se predikátový symbol P vyskytuje.

Při odvozování důsledků pak použijeme všechny instance výše uvedeného schematu. Poznamenejme ještě, že parametr Φ může být nahrazen jiným n–argumentovým predikátem.

4.7.2 Preferenční uspořádání

I. Odvození založená na částečném pre-uspořádání

Teď si všimneme logických systémů, které jsou založené na *a priori* daném částečném pre-uspořádání znalostí. Budeme zde studovat modely nemonotónních operací konsekvence vhodné pro (tentativní) důsledky odvozované s vyuřitím (částečného) pre-uspořádání definovaného na formulích.

Jak jsme uvedli již dříve, ve třicátých letech minulého století Alfred Tarski zavedl klasický koncept (monotónní) operace logického důsledku. Současné výzkumy v nemonotónních odvození jsou většinou chápány jako jistý druh kondicionálu: *Formule φ je nemonotónně odvoditelná z ψ, jestliže je logicky odvoditelná z φ a ze "skrytých" (nebo vhodných) formulí χ.* Diskuse se pak týjá obecných vlastností takové uzávěrové operace, které by nemonotónní odvození měla intuitivně splňovat. (Srovnej Chisholm

1940, Hintikka 1962, Lewis 1973, Adams 1975, ...)

Většina definicí silně závisí na vyjadřovací síle použitého jazyka. Hlavním naším cílem je uťinit tento koncept nezávislým na jazyce jak jen to je možné. Speciálně chceme mít tento koncept nezávislým na různých interpretacích negace.

Nechť F je algebra (úplný svaz) dobře utvořených formulí daného jazyka. $Cn : \mathcal{P}(F) \mapsto \mathcal{P}(F)$ – nechť je (monotónní) operace logické konsekvence (podle Tarského [133]). Chceme definovat (nemonotónní) operaci C, která je založena na "informačním" (posibilistickém nebo expektačním) uspořádání (Fariñas del Cerro, Gärdenfors, Wójcicki) formulí z F.

- expektační uspořádání
- posibilistické uspořádání
- probabilistické uspořádání (komparativní, kvalitativní)
- ...
- uspořádání na modelech
- uspořádání na množinách modelů
- ...

II. Odvození založená na uspořádáních

Definice 4.7.3 *Úplné pre-uspořádání (reflexivní a tranzitivní)* $\leq_e$ *se nazývá* uspořádání očekávání (expectation ordering)*, jestliže*
1. $\psi \in Cn(\varphi)$ *implikuje* $\varphi \leq_e \psi$ *(dominance) a*
2. $\varphi \vee \psi \leq_e \varphi$ *mebo* $\varphi \vee \psi \leq_e \psi$ *(disjunctiveness).*

Tento koncept je duální ke konceptu *uspořádání možností (possibility ordering)*, které se liší od uspořádání očekávání pouze ve druhé podmínce, tj.

2'. $\varphi \leq_p \varphi \wedge \psi$ nebo $\psi \leq_p \varphi \wedge \psi$, (srovnej Fariñas [35]).

Poznámka 1: Tranzitivnost, dominance a disjunktivnost implikují souvislost, tj. $\varphi \leq_p \psi$ nebo $\psi \leq_p \varphi$.
Poznámka 2: Je-li v jazyce negace, tak $\varphi \leq_e \psi$ právě když $\neg\varphi \leq_p \psi$.

Definice 4.7.4 *Let* $Cn = \vdash$. *Operace konsekvence založená na* pre-uspořádání $C_{\leq}$ *může být definována následovně:* $\psi \in C_{\leq}(\{\varphi\})$ *právě tehdy, když buď* $\varphi \vdash \psi$ *mebo existuje taková formule* χ, *že* $\varphi \wedge \chi \vdash \psi$ *a* $\neg\chi \leq \varphi$.

Tato definice je *korektní*, protože v případě, že není žádné uspořádání (tj. $\leq = \emptyset$) na formulích, tak $C_{\leq}$ se redukuje na $\vdash$. Tato definice může být zobecněna následovně:

Definice 4.7.5 C *se nazývá* nemonotónní operace konseqvence *royšiřující* Cn, *jestliže*
1. $Cn(X) \subseteq C(X)$ *a*
2. $\varphi \in C(X)$ *právě když buď* $\varphi \in Cn(X)$ *nebo existuje množina* H *skrytých formulí, taková, že* $\varphi \in Cn(X \cup H)$ *a* $inf(X) \geq sup(\hat{H})$, *kde* $\hat{H}$ *je množina negativních obrazů prvků z* H.

III. Odvození založená na neúplných uspořádáních

Když $\leq$ je ťásteťné uspořádání, tak (*credulous* nebo *liberální*)[16] nemonotónní operace konsekvence může být definována takto: $\varphi \in C(X)$ jestliže existuje takové totální (úplné) rozšíření $\sqsubseteq$ ťásteťného uspořádání $\leq$, že $\varphi \in C_{\sqsubseteq}(X)$.

Některé důležité vlastnosti

- *Supraklasikalita:*
  Jestliže $\varphi \in Cn(X)$, potom $\varphi \in C(X)$.

[16] Skeptická operace konsekvence bere v úvahu všechny možné extenze.

- *(Levá) logická ekvivalence:*
  Jestliže $Cn(X) = Cn(Y)$ a $\varphi \in C(X)$, potom $\varphi \in Cn(Y)$.

- **Zachování konzistence:**
  Jestliže $C(X) = F$, potom $Cn(X) = F$.

- **And:**
  Jestliže $X \subseteq C(Z)$ a $Y \subseteq C(Z)$, potom
  $X \cup Y \subseteq C(Z)$.

- **Or:**
  Jestliže $Z \subseteq C(X)$ a $Z \subseteq C(Y)$, potom
  $Z \subseteq C(X \cup Y)$.

- **Kumulativita:**
  Jestliže $\varphi \in C(\{\psi\})$ a $\psi \in Cn(\{\varphi\})$, potom
  $\chi \in C(\{\varphi\})$ právě tehdy, když $\chi \in C(\{\psi\})$.

(cf. [35]).

Základní množina vlastností

- Slabá kondicionalizace:
  Jestliže $X \subseteq C(Y)$, potom $sup(X) \Rightarrow inf(Y)$
  Jestliže $\beta \in C(\alpha)$, potom $\alpha \Rightarrow \beta \in C(\emptyset)$

- Slabá racionální monotónnost:
  Jestliže $\neg\alpha \notin C(\emptyset)$ a $\alpha \Rightarrow \beta$, potom $\in C(\alpha)$

- Zachování konzistence

Rozšířená množina vlastností

- Kumulativnost

- Or

- Racionální (opatrná) monotónnost

Výše uvedené vlastnosti však nejsou nezávislé. Například: Kumulativita je ekvivalentní řezu + opatrné monotónnosti. Kumulativita je ekvivalentní reciprocitě.

Teorém 4.7.1 *Každá nemonotonní operace konsekvence C založená na částečném pre-uspořádání splňuje následující vlastnosti:* **supraklasikalitu, levou logickou ekvivalenci, vlastnost and, zachovávání konzistence a vlastnost or** *a* kumulativnost.

Avšak *racionální (opatrná) monotónnost* (tj. Když $X \cup \{\varphi\}$ je konzistentní a $\psi \in C(X)$, tak $\psi \in C(X \cup \{\varphi\})$ nemůže být splněna.

IV. Pre-uspořádání na modelech

Nechť $W \neq \emptyset$ je množina možných světů a nechĽ $\geq$ je ťásteťné uspořádání možných světů. Jsou různé cesty jak rozšířit uspořádání $\geq$ na uspořádání na $\mathcal{P}(W)$, tj. *množiny* možných světů tak, aby byl splněn požadavek, že extenze zachovávají uspořádání světů:

- $W_1 \sqsupset W_2$ jestliže pro všechny světy $w \in W_1$ a pro všechna $v \in W_2$: $w \geq v$;

- zdůrazněním symetrické diference množin světů
  $W_1 \sqsupset W_2$ jestliže pro všechny světy $w \in W_1 - W_2$ existuje nějaké $v \in W_2 - W_1$ takové, že $v \sqsupset w$;

- zdůrazněním počtu možných světů;

- ...

V. Preferenční struktury

Nechť $K \neq \emptyset$ je množina elementárních výroků (výrokových konstant) a nechť $\mathcal{M} = [\mathrm{W}, \geq, \pi]$, kde W je množina možných světů, $\geq$ je částečné pre-uspořádání na W, a π přiřazuje každému světu $w \in W$ pravdivostní ohodnocení všech elementárních výroků v **K** tak, že

$$\mathcal{M} \models \varphi \succ \psi \text{ když } \| \varphi \|_{\mathcal{M}} \geq \| \psi \|_{\mathcal{M}}.$$

Poznámka: $\mathcal{M} \models \neg(\neg\varphi \succ false)$ právě když $\| \neg\varphi \|_{\mathcal{M}} = \emptyset$ právě když $\| \varphi \|_{\mathcal{M}} = W$.

VI. LPP - Logika s částečným pre-uspořádáním

A1: Všechny instance tautologií PC

A2: $\neg(\varphi \succ \varphi)$

A3: $((\varphi \vee \psi) \succ \chi) \wedge ((\varphi \vee \chi) \succ \psi) \Rightarrow$
$(\varphi \succ (\psi \vee \chi))$

A4: $(\Box(\varphi \Rightarrow \varphi_1) \wedge \Box(\psi_1 \Rightarrow \psi) \wedge (\varphi \succ \psi)) \Rightarrow \varphi_1 \succ \psi_1$

Poznámka: $\succ$ je irreflexivní, kvalitativkí a uspořádávající.

Inferenční pravidla:

Z φ *a* $\varphi \Rightarrow \psi$ *odvoď* ψ. (Modus ponens)

Odvoď $\Box\varphi$ *z každé výrokové tautologie.* (Generalizace/Necesitace) ∎

Teorém 4.7.2 *Logic LPP is sound and complete axiomatization with respect to pre-ordered preferential structures.*

VII. LTP - Logika s totálním pre-uspořádáním

Definujeme **LTP** = **LPP** plus axiom **A5**.

A5: $(\varphi \succ \psi) \Rightarrow ((\varphi \succ \chi) \vee (\chi \succ \psi))$

Teorém 4.7.3 *Logika LTP je korektní a úplnou axiomatizací vzhledem k totálně pre-uspořádaným preferenčním strukturám.*

Poznámka: V **LTP** může být axiom **A4** nahrazen transitivitou a vlastností sjednocení pro $\succ$. (Srovnej. Lewis)

Union property: $(\varphi \succ \psi) \wedge (\varphi \succ \chi) \rightarrow \varphi \succ (\psi \vee \chi)$

Příklad: (Halpern) Let K = $\{p, q\}$. Formule
$(p \succ (\neg p \wedge q)) \wedge \neg((p \wedge q) \succ (\neg p \wedge q)) \wedge \neg((p \wedge \neg q) \succ (\neg p \wedge q))$
je splněna ve struktuře $(W, \geq)$ s W = $\{w_1, w_2, w_3, w_4\}$ taková, že $w_1 \geq w_3, w_2 \geq w_4$ a $p \wedge q$ je pravdivá ve w_1, $p \wedge \neg q$ je pravda ve w_2, a $\neg p \wedge q$ je pravdivá ve w_3 i ve w_4. Avšak není splnitelná v žádné struktuře, kde existuje aspoň jeden svět, v němž je splněna formule $\neg p \wedge q$.

4.8 Zamítnutelné argumenty

Idea zamítnutelného usuzování pochází od Donalda Nuta [100]. Je založena na rozlišení dvou druhů pravidel, *logických* (nebo absolutních) pravidel a *empirických* (nebo zamítnutelných) pravidel a vzhledem k tomu také mezi absolutními a zamítnutelnými závěry. To znamená, že máme dva druhy znalostí, ale toto rozdělení není absolutní. Je spíše závislé na kontextu (nebo lépe řečeno, problémově závislé). Které pravidlo bude chápáno jako absolutní a které jako zamítnutelné může být velmi závislé na oblasti aplikace. Podle dobře známého metodologického principu "*de omnibus dubitandum est*" libovolná položka znalostí může být zamítnuta (nebo aspoň zpochybněna). Jakmile ale fixujeme úlohu, můžeme rozlišit silná a slabá pravidla vzhledem k právě řešené úloze.

Začneme specifikací jazyka. Pro jednoduchost budeme prezentovat výrokovou verzi ačkoli verze pro jazyky prvního řádu s proměnnými je také možná. Budeme předpokládat, že znalosti sestávají z konečného počtu *faktů* a konečného počtu *pravidel*. Z formálního hlediska fakty jsou literály (tj. pozitivní a/nebo negativní atomické sentence). To znamená, že v našem jazyce bychom měli být schopni vyjádřit jakýsi druh negace. Takže budeme předpokládat, že libovolném faktu a je přiřazen fakt *neg* a, který je *opakem* faktu a, a naopak. Jinými slovy, množina {a, neg a} je *inkonzistentní*. Jak ale uvidíme později, přítomnost dvou opačných fakt nemá za následek to, že by měl být odvozen *libovolný* fakt. To znamená, že logický *princip kontradikce*, neboli pravidlo $(p \wedge \neg p) \Rightarrow q$, zde pro opačné fakty neplatí. Tato čistě syntaktická charakterizace inkonzistentních množin formulí stačí pro to, abychom dostatečně přesně definovali proceduru zamítnutelného usuzování. Aby byl náš jazyk vhodný pro praktické aplikace, budeme za opačné fakty považovat i fakty opírající se o *ne-*

kompatibilní predikáty jako jsou např. *a je zelený* a *a je bezbarvý*, které nejsou kontradiktorické ve striktním logickém smyslu.

Pravidla jsou dvou druhů:

$[\alpha_1, ..., \alpha_n] \rightarrow \gamma$ (silná, absolutní)
$[\alpha_1, ..., \alpha_n] \Rightarrow \gamma$ (slabá, zamítnutelná)

kde $[\alpha_1, ..., \alpha_n]$ označuje konečnou (možná dokonce i prázdnou) množinu faktů;

γ je fakt (pozitivní nebo negativní).

První druh pravidel (tj. absolutní či striktní pravidla vyhledem k aplikaci), druhý pro empirická zamítnutelná pravidla. Například pravidlo "*Birds fly but penguins and ostriches do not fly*" může být ekvivalentně vyjádřeno pomocí těchto tří formálních pravidel:

[bird(X)] $\Rightarrow$ flies(X)
[penguin(X)] $\rightarrow$ neg flies(X)
[ostrich(X)] $\rightarrow$ neg flies(X)

Pravidla jsou jednoduchá, každé s *jedinou*[17] podmínkou. Samozřejmě, že ale můžeme zapisovat pravidla s více než jedním předpokladem. Například pravidlo

[bird(X), neg penguin(X), neg ostrich(X)] $\rightarrow$ flies(X).

Jak ale uvidíme, toto nebude cesta jak překládat předchozí tři pravidla na toto poslední. Naším cílem je dosáhnout dobrého formalismu pro rozhodování mezi zamítnutelnými a nezamítnutelnými položkami báze znalostí. (Připomeňme ještě, že symbol "," použitý v seznamu podmínek jako separátor, zde vlastně hraje

[17]Vzhledem k tomu, že v antecedentu pravidla je *seznam* podmínek, můžeme také formulovat pravidla bez podmínek, tj. s prázdnou množinou podmínek.

roli konjunkce.

Nechť **K** je okamžitý stav znalostí (pravidel a faktů). Jako obvykle formuli α nazveme *logicky odvoditelnou* z okamžitého stavu znalostí **K** (symbolicky: $\mathbf{K} \vdash \alpha$), jestliže může být získána z **K** konečným počtem aplikací logických pravidel inference v **K**. Definice zamítnutelné empirické odvoditelnosti je rovněž založena na zřetězování pravidel, je ale poněkud komplikovanější, protože musí vzít ohled na možnou zamítnutelnost odvození. Takže formule α se nazývá *empiricky odvoditelná* z okamžitého stavu znalostí **K** (symbolicky: $\mathbf{K} \rightsquigarrow \alpha$), jestliže je splněna aspoň jedna z následujících podmínek:

- α je logicky odvoditelná;
- existuje takové logické pravidlo $[\beta_1, ..., \beta_n] \rightarrow \alpha$ v **K**, že každá formule β_i je empiricky odvoditelná z **K** a α není zpochybnitelná okamžitým stavem znalostí **K**;
- existuje takové empirické pravidlo $[\beta_1, ..., \beta_m] \Rightarrow \alpha$ v množině **K**, že každá formule β_j je empiricky odvoditelná z **K** a α není ani zpochybnitelná množinou **K** ani zamítnutelná množinou formulí $\{\beta_1, ..., \beta_n\}$.

K dokončení naší definice empirické odvoditelnosti musíme ještě určit kdy je formule α *zamítnuta* okamžitým stavem znalostí. Začneme několika pomocnými definicemi.

Řekneme, že množina formulí X je *deduktivně silnější* než množina Y, jestliže Y je logickým důsledkem X (tj. libovolná formule $\alpha \in Y$ může být získána z X zřetezením logických pravidel) a X není logickým důsledkem Y. Řekneme také, že formule α *je napadnutelná (contradicted)* (okamžitým stavem znalostí), jestliže existuje formule β, taková, že $\mathbf{K} \vdash \beta$ a $\beta = \text{neg}\ \alpha$, nebo když formule α a β jsou *nekompatibilní.*

Nyní můžeme říci, že formule α *je zamítnuta množinou formulí* X vzhledem k okamžitému stavu znalostí $\mathbf{K}$, jestliže existuje formule $\gamma \in \mathbf{K}$, která napadá α a takové empirické pravidlo

$[\beta_1, ..., \beta_n] \Rightarrow \gamma$

že každá formule β_i je empiricky odvoditelná a množina X *není* deductivně silnější než množina $\{\beta_1, ..., \beta_n\}$. (Připomeňme, že výraz X *není deduktivně silnější* p5esn2 ynamen8, že $X \nvdash Y$ nebo $Y \vdash X$.)

Fakt: Z $\mathbf{K} \vdash \alpha$ vyplývá $\mathbf{K} \rightsquigarrow \alpha$, ale ne nutně naopak.

4.8.1 Interakce pravidel

Všimněme si nejprve následujích dvou pravidel

[a] $\Rightarrow$ c
[b] $\Rightarrow$ neg c

Vzniká otázka jak odvoditelnost výroku c závisí na a a na b. Jestliže v bázi znalostí $\mathbf{K}$ nejsou už jiná pravidla, tak odvoditelnost c plně závisí jen na a a b. Samozřejmě, že z a může být (empiricky) odvozeno c, zatímco z b je (empiricky) odvozeno *neg c*. Konečně když předpokládáme jak a a b , tak podle naší definice korektně odvodíme c a *neg c*. V tomto případě není žádná položka znalostí (pravidlo a/nebo fakt), která by mohla zamítnout jeden z obou faktů.

Bude-li však přidáno další pravidlo, které říká něco o logickém nebo empirickém vztahu výroků a a b, potom nás procedura usuzování povede k jiným závěrům. Podejme několik příkladů, které budou ilustrovat jak odvoditelnost výroku závisí právě na vzájemném vztahu mezi formulemi, které jsou důvodem pro odvozovanou formuli.

Příklad 1

[a] $\Rightarrow$ c
[b] $\Rightarrow$ neg c
[a] $\rightarrow$ b

Potom {a} $\rightsquigarrow$ c. Jistě, *neg c* nemůže být odvozeno z {a} neboť existuje logická implikace mezi a a b, která ukazuje, že množina {a} je deduktivně silnější než {b}. Takže *neg c* musí být zamítnuto množinou {b}, protože b není důvodem pro výrok *neg c*, jež by měl být nezávislý na a.

Příklad 2

[a] $\Rightarrow$ neg c
[b] $\Rightarrow$ c
[a] $\rightarrow$ b

Potom {a} $\rightsquigarrow$ neg c.

Jestliže však v obou příkladech nahradíme silné logické pravidlo

[a] $\rightarrow$ b

empirickým pravidlem, potom ani c ani *neg c* nejsou odvoditelné.

Následující dva příklady bází znalostí vedou ke kontradiktorním výsledkům, protože jak c tak *neg c* jsou empiricky odvoditelné z a.

Příklad 3

[a] $\Rightarrow$ c
[b] $\rightarrow$ neg c
[a] $\Rightarrow$ b

Příklad 4

[a] $\Rightarrow$ neg c

[b] $\rightarrow$ c
[a] $\Rightarrow$ b

Interakce mezi pravidly nám dovolují pracovat s rozmanitými druhy pravidel často diskutovanými ve znalostních systémech umělé inteligence. Tak například můžeme simulovat:

Úsudky s výjimkami těmito pravidly

$[\alpha_1, ..., \alpha_n] \Rightarrow \gamma$
$[exception_1] \rightarrow$ neg γ
...
$[exception_m] \rightarrow$ neg γ

Není těžké si uvědomit, že stejné výsledky dostaneme, když pravidla pro výjimky zapíšeme pomocí pouze empirických pravidel. Neměli bychom ale přehlédnout, že taková pravidla nelze ekvivalentně vyjádřit jediným pravidlem tvaru
$(\alpha_1 \wedge ... \wedge \alpha_n) \wedge (\neg e_1 \vee ... \vee \neg e_m) \rightarrow \gamma.$
Důvod je tento: *Množina* pravidel dovoluje za předpokladu osamoceného faktu *Tweety is a bird*, uzavřít - *na základě okamžitého stavu znalostí* -, že *Tweety flies*, ale později, když se dovíme, že *Tweety is also a penguin*, revidujeme dosavadní závěr. Takový *postup* odvození ovšem nemůže být modelován výše uvedenou formulí, protože ta vyžaduje znát aspoň jednu výjimku na samém začátku usuzování.

Usuzování s omezujícími pravidly, tj. usuzování s pravidly tvaru "*From* $\alpha_1, ..., \alpha_n$ *odvoď* γ *ledaže by* $\beta_1, ..., \beta_m$" pomocí pravidel

$[\alpha_1, ..., \alpha_n] \Rightarrow \gamma$
$[\alpha_1, ..., \alpha_n, \beta_1, ..., \beta_m] \Rightarrow$ neg γ

Usuzování s defaulty, tj. s pravidly typu "*Jestliže není nic jiného řečeno, předpokládáme, že* γ".

$[\,] \Rightarrow \gamma$

I. Dopředné řetězení

Donald Nute [99] navrhl implementovat zamítnutelné usuzování pomocí logických programů. Autor této publikace využil Nuteho myšlenku při realizaci systému D–Expert, tj. prologovského programu jakožto argumetátorského systému pro diagnostické rozhodování [60].

Hlavní problém prologovské implementace zamítnutelného usuzování spočívá ve značné výpočtové složitosti. Tento problém byl zkoumán v diplomové práci Martina Špalka.

```
% Program for non-monotonic reasonig via defeasible rules

:-   op(900,fx,neg).
:-   op(1100,xfy,->>).
:-   op(1100,xfy,=>>).
:-   op(1100,xfy,?->>).

d(X) :- nl,write('A moment, please. Just computing! '),
        nl,fail.
d(X) :- logically_derivable(X);empirically_derivable(X).

logically_derivable([]).
logically_derivable([H|T]) :-
   logically_derivable(H),
   logically_derivable(T).
logically_derivable(L) :-
   not list(L),
   (fact(L);L).
logically_derivable(L) :-
   not list(L),
   ((Antecedent ->> L)),
```

```
   logically_derivable(Antecedent).

empirically_derivable([]).
empirically_derivable([H|T]) :-
   empirically_derivable(H),
   empirically_derivable(T).
empirically_derivable(L) :-
   not list(L),
   logically_derivable(L).
empirically_derivable(L) :-
   not list(L),
   ((Antecedent ->> L)),
   empirically_derivable(Antecedent),
   not contradicted(L).
empirically_derivable(L) :-
   not list(L),
   ((Antecedent =>> L)),
   empirically_derivable(Antecedent),
   not contradicted(L),
   not defeated(L,Antecedent).

contradicted(L) :-
   contrary(L,Opposite),
   logically_derivable(Opposite).

contrary(L,Opposite) :-
   negation(L,Opposite).
contrary(L,Opposite) :-
   incompatible(L,Opposite).
contrary(L,Opposite) :-
   incompatible(Opposite,L).

negation(neg Atomic_formula,Atomic_formula).
negation(Atomic_formula,neg Atomic_formula) :-
```

```
   not functor(Atomic_formula,neg,1).

defeated(L,Antecedent_1) :-
   contrary(L,Opposite),
   ((Antecedent_2 =>> Opposite)),
   empirically_derivable(Antecedent_2),
   not better_informed(Antecedent_1,Antecedent_2).
defeated(L,Antecedent_1) :-
   contrary(L,Opposite),
   ((Antecedent_2 ?->> Opposite)),
   empirically_derivable(Antecedent_2),
   not better_informed(Antecedent_1,Antecedent_2).

better_informed(Antecedent_1,Antecedent_2) :-
   relative_consequence(Antecedent_2,Antecedent_1),
   not relative_consequence(Antecedent_1,Antecedent_2).

relative_consequence([],Premises).
relative_consequence([H|T],Premises) :-
   relative_consequence(H,Premises),
   relative_consequence(T,Premises).
relative_consequence(L,Premises) :-
   not list(L),
   member(L,Premises).
relative_consequence(L,Premises) :-
   not list(L),
   ((Antecedent ->> L)),
   relative_consequence(Antecedent,Premises).

% Auxiliary predicates

list([]).
list([H|T]).
```

```
member(H,[H|_]).
member(H,[_|T]) :-
   member(H,T).

append([],X,X).
append([H|T],X,[H|Y]) :- append(T,X,Y).

q :- halt.

retractall(X) :- retract(X),fail.
retractall(X) :- retract((X:-Y)),fail.
retractall(_).
```

4.8.2 Souvislost s jinými přístupy

I. Skeptická teorie dedičnosti

Na první pohled, zapomeneme-li na rozdíl mezi dvěma druhy spojení (přesněji, budeme-li pracovat pouze s empirickými pravidly), pak restringovaná třída znalostních bází se zamítnutelným usuzováním je podobná skeptické teorii dědičnosti v nemonotónních sémantických sítích [54]. Přesněji, síť tvrzení ve smyslu teorie děděných vlastností lze chápat jako množinu empirických pravidel (každé pouze s pozitivními předpoklady). Tudíž, *isa*–relace v sémantických sítích může být traktována jako speciální případ (empirických) pravidel. Snadno porovnáme různé dobře známé konfliktní sítě, jako např. *Nixon diamond* a nalezneme stejné závěry. Jsou však příklady, které dopadnou jinak. Uveďme aspoň tento příklad sítě pravidel:

[a] $\Rightarrow$ b
[b] $\Rightarrow$ c
[a] $\Rightarrow$ neg c

Podle skeptické teorie dědění vlastností *neg c* můžeme odvodit z faktu a, tj. $a \rightsquigarrow neg\ c$, protože kratší cesta je preferována, ale algoritmus zamítnutelného usuzování nedovoluje uzavřít ani na c ani na *neg c*. To mj. ukazuje na to, že procedura zamítnutelného odvození je opatrnější, nebo jinak, že procedura skeptického odvození v hierachiích dědění vlastností není tak skeptická.

II. Hypotetické usuzování

Pokusy vytvořit systémy pro aktualizaci úsudků ve stylu teorie zamítnutelných odvození úzce souvisejí s *hypotetickým usuzováním* [IM 90], kde *fakty* odpovídají *logickým* položkám znalostí a *hypotézy* odpovídají *empirickým* položkám znalostí, tj. empirickým pravidlům a empirickým faktům. Připomeňme, že empirické (zamítnutelné) fakty mohou být vyjádřeny jako empirická pravidla tvaru

$[\ ] \Rightarrow \text{fact},$

tj. jako empirická pravidla s prázdnými předpoklady. V takovém případě může být libovolný empirický fakt zamítnut jinou položkou znalostí, třeba i empirickou.

Nechť $\mathbf{L}$ je množina logických pravidel a faktů. Když je dána množina pozorování (observací) $\mathbf{O}$, argumentační algoritmus se snaží hledat množinu hypotéz $\mathbf{H}$, která je podmnožinou empirických znalostí splňujících následují dvě podmínky

- $\mathbf{L} \cup \mathbf{H} \vdash \mathbf{O}$,
- $\mathbf{L} \cup \mathbf{H}$ není sporná.

To znamená, že $\mathbf{H}$ je množina konzistentních hypotéz, které vysvětlují množinu observací $\mathbf{O}$. Tento přístup může být zobecněn tak, že první podmínku nahradíme podmínkou

- $\mathbf{L} \cup \mathbf{H} \models \mathbf{O}$

a druhá podmínka je nahrazena požadavkem, že

- **H** není zamítnuta množinou **L**.

Hypotetický druh usuzování úzce souvisí se strojovým učením a induktivním logickým programováním. Je to ale především abduktivní druh inference, který bude podrobněji popsán v kapitole 7.

4.9 Obecné vlastnosti nemonotónních systémů

Deduktivní důsledek
operace $c : \mathcal{P}(F) \longrightarrow \mathcal{P}(F)$ (*tarskiánský styl*)
relation $\vdash$ on $\mathcal{P}(F) \times F$ (*Gentzen style*)

- incluze
  $X \subseteq c(X)$
- monotónnost
  $X \subseteq Y$, pak $c(X) \subseteq c(Y)$
- idempotence
  $c(c(X)) \subseteq c(X)$
- kompaktnost
  $c(X) = \bigcup c(X_i)$ pro všechny konečné množiny $X_i \subseteq X$
- $\ldots$
- reflexivnost
  $X \vdash \varphi$ kdykoli $\varphi \in X$
- monotónnost
  Jestliže $X \vdash \varphi$ a $X \subseteq Y$, potom $Y \vdash \varphi$
- tranzitivnost (cut)
  $X \vdash \varphi$ právě když $X \cup \{\psi_i; i \ \in I$ a $X \vdash \psi_i\}$
  finitistická verze: $X \cup \{\psi\} \vdash \varphi$ právě když $X \vdash \varphi$
- kompaktnost
  $X \vdash \varphi$ iff $X_0 \vdash \varphi$ pro nějakou končnou množinu $X_0 \subseteq X$
  finitistická verze: $X \vdash \varphi$ iff $X \vdash \psi$ a $X \cup \{\psi\} \vdash \varphi$

- ...

Nonmonotónní inference
operace $i : \mathcal{P}(F) \longrightarrow \mathcal{P}(F)$

- reflexivnost (inkluze)
  $X \subseteq i(X)$
- supraklasikalita
  $c(X) \subseteq i(X)$
- kumulativní transitivity (řez)
  Jestliže $X \subseteq Y \subseteq i(X)$, potom $i(Y) \subseteq i(X)$
- kumulativní monotónnost (také opatrná monotónnost)
  Jestliže $X \subseteq Y \subseteq i(X)$, potom $i(X) \subseteq i(Y)$
- absorbce
  $c(i(X)) = i(X) = i(c(X))$
- zachování konzistence
  $c(X) \neq F$, pak $i(X) \neq F$
- distributivnost
  $i(X) \cap i(Y) \subseteq i(c(X) \cap c(Y))$
  speciálně:
  $i(X) \cap i(Y) \subseteq i(X \cap Y)$ kdykoli $X = c(X), Y = c(Y)$
- reciprocita
  Jestliže $X \subseteq i(Y)$ a $Y \subseteq i(X)$, potom $i(X) = i(Y)$
- ...

Tyto vlastnosti nejsou nezávislé. Není obtížné např. ověřit, že:

- Platí-li inkluze, potom kumulativnost *je ekvivalentní* reciprocitě.
- Supraklasikalita *a* idempotence *implikují* $c(i(X)) = c(X)$.
- Supraklasikalita *a* kumulativivnost *implikuje* absorpci.

Podle Dov Gabbaye [39] “operace inference”, která je reflexivní a má vlastnost kumulativní monotónnosti a kumulativní tranzitivnosti je nejslabší *racionální* operace inference, která je nemonotónní.

Naším záměrem je zde ukázat jak jsou tyto vlastnosti propojeny vzájemně mezi sebou. Avšak otázka, která kombinace vlastností nejlépe charakterizuje libovolnou rationální inferenci zůstává stále otevřenou otázkou.

Kapitola 5

Usuzování z negativní informace

5.1 Negace jako nemožnost

Překvapení je to, co nastane
když selže zdravý rozum.

Někteří lidé říkají, že ve znalostech je síla a moc. Ale to se týká každých znalostí a tudíž i negativních znalostí. Co to ale je negativní znalost? Je to poznání toho, že něco není možné díky tomu, že naše síly jsou chabé, nebo je to absence znalostí anebo poznání toho, že něco není pricipiálně možné. Vzpomňme na paradoxní otázky typu zda existuje kámen, který všemocný bůh neuzvedne.

5.1.1 Nemožnost a překvapení

Začneme jednoduchou klasifikací různých druhů negativních znalostí a potom podáme argument pro tvrzení, že negativní znalost je vlastně kontrapozicí mezi přirozeným světem a světem překvapení. Tím máme na mysli především výsledky o nemožnosti, o tom, že něco neplatí, něco není možné uskutečnit. Vždy, když se v rámci vědeckého zkoumání setkáváme s negativními výsledky vede to obvykle k překvapivému obratu. Překvapení se tak stává motorem věckého zkoumání.

Uvažujme o následujícím přehledu několika ne právě systematicky vybraných ale dobře známých příkladů výsledků ve vědě, kterých lidé dosáhli za poslední dvě milenia. Každý z nich vyjadřuje nějaký negativní poznatek, nějakou znalost o tom, že něco není možné.

- trisekce úhlu (*omezené prostředky*)
- nesouměřitelnost strany čtverce s její úhlopříčkou (*podivný prostor, neexistuje ratio, poměr dvou přirozených čísel, který by vzjadřoval jejich vztah*)
- vězňovo dilemma (*zde řešení úloh s neúplnou informací*)
- nemožnost perfektní demokratické volby, Arrowův teorém (*vyhodnocujeme* de facto *něco jiného než chceme*)
- problém obchodního cestujícího (*složitost výpočtů*)
- perpetuum mobile (*skrytý princip, který zajišťuje bezespornost teorie*)
- Gödelovy výsledky - nemožnost dokázat bezespornost elementární aritmetiky prostředky jejího jazyka (*koncept pravdy je bohatší než koncept důkazu*)
- relace neurčitosti v kvantové fyzice (*není možné současně změřit dvojici veličin*)

- myšlenkové experimenty (většinou ve fyzice: Maxwellův démon, Schödigerova kočka, Einsteinův výtah, Newtonova kytice)

Naskýtá se otázka, zda je neco společného ve výše uvedených úlohách.

5.2 Negace ve formálních systémech

V tomto odstavci probereme různé přístupy k porozumění negaci ve formálních systémech. Porornost zaměříme na rozlišení interní a externí a externí negace a budeme se snažit pochopit, jak tyto dva koncepty ovlivňují koncepty možnosti a nemožnosti.

Negace

Negace je mentální konstrukt či jazykový jev. Z hlediska logiky se týká naší schopnosti rozllišovat pravdu od nepravdy. Ale co je to nepravda a proč je důležitější než pravda? Je nepravda vždy vzjadřována pomocí negace? V přirozených jazycích nalezneme mnoho příkladů, kdy tomu tak není.

Setkáváme se s různými významy konceptu negace.

Pozitivní logika

Here we start discussion on the concept of classical propositional logic with respect to negation. The begining is a set of axioms of the positive propositional logic as it was presented for example in Grzegorczyk [**?**]. Adding two axioms concerning negation we can obtain intuitionistic set of axioms. And again, adding another axiom, we complete the set of axioms for classical propositional logic. It seems to be obvious. But how it corresponds to results concerning the fact that intuitionistic logic is contained (in the sense of *Enthaltsein*) in classical logic but also the opposite is true, i.e. classical logic is contained in intuitionistic logic, as well.

Negace jako inkonzistence

V tomto případě, který je blízký Sókratovskému koncepru racionality je negace chápána jako sponost, tj. inkonzistence. Ale sponost s čím? Obvykle je formula jazyka považována za nepravdivou v daném systému, když její přidání do systému vede v systému ke sporu (Johannson: minimální logika, Heyting: intuitionistická logika)

Negace jako neúspěch

Negace chápaní jako neúspěch *negation-as-failure* je koncept důležitý pro procedurální chápání nepravdy. Mzslí se samozřejmě na neúspěch při odvozování. Je přitom ovšem o neúspěch či selhání, které není dáno případnou neschopnotí toho, kdo odvozuje, ale o principiální nemožnost odvození. V procedurální sémantice logických programů je negace chápána právě jako selhání při dokazovaní. (Gabbay [**?**], Kowalski).

Theory completion

Konstruktivní negace

negace může být též chápána jako konstruktivní nepravda *constructive falsity.* (Fitch, Markov) Je vyžadovám přímý důkaz. Důkaz musí být zkonstruován. Nepřipouštějí se nepřímé důkazy, zejména pak ne tehdy, jde-li o nekonečné univerzum.

Externí negace

Zajímavým a podnětným je porozumění konceptu negace v rámci informační sémantiky, kdy negace je pojímána jako nedostatek informace či absence znalostí [?].

Kapitola 6

Dynamika znalostí

Život je to, co se vám přihodí,
když máte právě jiné plány.
Anonym

Nyní nás zaujme *proces změny znalostí.* Naše znalosti obvzkle pocházejí z rozmanitých zdrojů a je tudíž přirozené, že se každá položka báze znalostí může v čase měnit. Nebudeme se ale zabávat otázkami adekvátnosti – to je úloha, která jde nejen za rámec této publikace, ale i za rámec exaktního formálního zkoumání.

Proto raději začneme s nějakou množinou formulí, o níž budeme předpokládat, že reprezentují daný epistemický stav (okamžitý stav) znalostí. Cílem je hledat racionální bázi pro popis procesu změn epistemických stavů.

Budeme se opět snažit učinit naše zkoumání co nejméně závislým na vyjadřovacích prostředcích i na prostředcích reprezentace znalostí, takže nám prozatím postačí předpokládat, že je fixován jazyk a logika (tj. je pevně dána operace logické konsekvence Cn) definovaná na dané množině dobře utvořených formulí.

O množině znalostí teď budeme uvažovat jako o přesvědčeních (*beliefs*) a samozřejmě, že přijmeme nějaké předpoklady o struktuře znalostí a o jejím nositeli. Uvažujeme-li o nositeli nějakého souboru přesvědčení, máme hned zpočátku aspoň dvě možnosti náhledu na jeho schopnosti inferemce logických důsledků. Z pohledu ideálního racionálního nositele znalostí (přesvědčení) je vhodné epistemický stav chápat jako množinu přesvědčení, která je uzavřená na logické důsledky. To je samozřejmě značná idealizace, která ale odpovídá našemu tématu, zatímco bez tohoto předpokladu se dostáváme do zcela odlišné oblasti zkoumání, které se netýkají primárně konceptu racionalty, ale spíše psychologických aspektů usuzování.

Proto v dalším budeme předpokládat, že množina přesvědčení (epistemický stav) je uzavřená na logické důsledky, tj. na operaci Cn.

6.1 Změny znalostí

6.1.1 Pomocné definice

Nechť $\alpha \in F$ je libovolná formule a X je množina přesvědčení. Pak můžeme rozlišit tři případy (vzhledem k X):

- *α je akceptována (množinou X)*, jestliže $\alpha \in Cn(X)$.
- *α je zamítnuta (množinou X)*, jestliže $X \cup \{\alpha\}$ je inkonzistentní, tj. $Cn(X \cup \{\alpha\})F$.
- *α je nezávislá (na X)*, jestliže není akceptovaná ani zamítnutá množinou X. (Jinými slovy, α nezávisí na X kdzž je možné ji konzistentně přidat k X.)

6.1.2 Aktualizace znalostí

Kapitola 7

Abduktivní usuzování

Abdukce je velmi účinné rozšíření deduktivní inference. V této kapitole se budeme zabývat abduktivním usuzováním ve vztahu k deduktivnímu, induktivnímu a nemonotónnímu usuzování, jak jsme o nich pojednávali doposud. Termín *abduktivní usuzování* je velmi starý [1], ale systematicky byl studován až v tomto století Charlesem Sandersem Peircem [**?**] a dalšími [105], [46], [69], [68].

V aristotelské logice je abdukce ($\alpha\pi\alpha\gamma\omega\gamma\eta$, *apagogé*) chápána jako sylogismus jehož hlavní premisa je jistá, ale vedlejší premisa je jen pravděpodobná. Podle Peirce je to druh inference, který přináší explanatorní hypotézy spíš než výsledek deduktivního či induktivního pravidla. Abduktivní usuzování v tomto smyslu je usuzování z neúplných znalostí nebo hypotetické usuzování. V protikladu k dedukci se abdukce týká plauzibilnosti závěrů, nikoli jejich platnosti.

[1] Srovnej Aristotelés [**?**].

7.1 Výchozí pojmy

Jak jsme poznali již dříve, různé typy usuzování jsou často velmi závislé na vyjadřovací síle užitého jazyka. Nechť tedy jazyk je fixován, tj. nechť F je třída dobře utvořených formulí. Opět zatím nepředpokládáme nic o její struktuře ani o struktuře jejích prvků, tedy formulí. Podáme tentativní formální definici abdukce, která bude nezávislá na užitém jazyce a na prostředcích reprezentace znalostí.

Budeme předpokládat, že jsou dány nejaké znalosti B tvořící znalostní zázemí (background knowledge) a formule nebo množina formulí G. Abduktivní inferenci definujeme jako úlohu nalézt množinu hypotéz H, která je konstruována v prostoru formulí A, které někdy nazýváme *abducibles*, což jsou také formule daného jazyka. Není ovšem nutné, aby formule ze všech tří množin byly nutně definovány v témže jazyce, i když společná část je patrně nezbytná. Jen těžko bychom hledali rozumnou definici abdukce mezi formulemi, které jsou navzájem irelevantní.

Definice 7.1.1 *Abdukce je úloha nalézt takovou množinu hypotéz H, pro kterou platí*

1. $B \cup H \models G$, *neboli* $G \in Cn(B \cup H)$
2. $B \cup H$ *je konzistentní*
3. $H \subseteq A$ *a* $G \cap H = \emptyset$
4. $G \notin Cn(B)$
5. *Neexistuje taková množina* $H' \subset H$, *že* $G \in Cn(B \cup H')$. *(Minimalita vysvětlení)*

Jde tedy o tentativní přijetí množiny hypotéz. Abdukci lze v tomto kontextu chápat jako inferenci, která směřuje k nalezení

nejlepšího vysvětlení. Přitom se obvykle požaduje minimalita, což je princip, který je často uváděn jako Occamova břitva. Uvedeme to na jednoduchém příkladě:
Nechť báze znalostí obsahuje tato dvě tvrzení:

$p \to r$
$p \wedge q \to r$

V této situaci je $\{p \wedge q\}$ možným vysvětlením faktu r, ale není to vysvětlení minmální, tím je fakt $\{p\}$.

Dalším častým požadavkem je fundovanost vysvětlení (angl. *basicality*), termín zavedený Feyerabendem. Řekneme, že vysvětlení je fundované, když je není možné vysvětlit v termínech jiných vysvětlení. Opět uvedeme příklad:

$m(b) \to m(t)$
$k(t) \to m(t)$
$m(t) \to m(b)$

kde $m(x)$ je výraz, který interpretujeme jako predikát x je mokrý, t, - trávník a b - boty jsou logické konstanty a $k(x)$ znamená, že x byl kropen.

7.2 Závislostní sítě

Statickou část znalostí obvykle v logice reprezentujeme formulemi nějakého formálního jazyka. Tak o teoriích mluvíme jak v rámci teorií nultého řádu, prvního řádu i vyšších řádů. Pro účely odvozování, které je založeno na aplikaci pravidel to obvykle postačí. Cena, kterou za to platíme spočívá v tom, že položky znalostí, které spolu mohou souviset nejsou asociovány, mohou být od sebe velmi "vzdáleny". V mnoha situacích ale, i když říkáme, že pracujeme s formulemi, *de facto* pracujeme se sítěmi formulí, ačkoli si to možná ani zřetelně neuvědomujeme. Chceme-li modelovat dynamiku znalostí, bývá výhodné inferenční mechanismy formulovat přímo na sítích. Závislostní sítě, někdy též nazývané systémy aktualizace znalostí či pravd anebo důvodů (TMS - *truth maintenance systems*, RMS - *reason maintenance systems*), mají velmi úzký vztah k abduktivnímu usuzování. Mají rozmanité podoby, jednu z nich teď probereme podrobněji, abychom poznali jejich možnosti.

James Goodwin [49] v osmdesátých letech minulého století vytvořil systém aktualizace sítí znalostí, který umožňuje reprezentovat pravidla s vyjímkami. Základními stavebními kameny jsou výroky (sentence), které jsou propojovány závislostmi do sítí. Jednoduchá závislost je znázoněna na obrázku XX2 a XX3.

Takovou závislost můžeme vyjádřit slovy:

"Jestliže α, potom γ, ledaže by β.

Výroky mohou být označeny jako pravdivé, nepravdivé anebo neohodnocené. Závislosti mohou být považovány za platné (označeny symbolem IN), nebo za neplatné (označení OUT) anebo jsou neohodnocené. Jde tedy vlastně o trojhodnotový přístup.

Definice 7.2.1 *Závislost je* platná*, když všechny předpoklady (monotónní předchůdci) jsou pravdivé a všechna omezení (nemonotónní předchůdci) jsou nepravdivá tvrzení. Závislost je* neplatná*, když aspoň jeden z předpokladů je nepravdivý nebo aspoň jedno z omezení je pravdivé.*

Všimněme si, že platnost a neplatnost nejsou jedno logickou negací druhého.

Mějme nyn dáno nějaké ohodnocení výroků a závislostí v síti. Naším cílem je hledat konzistentní ohodnocení výroků a závislostí v dané síti.

Definice 7.2.2 *Síť je konzistentní, když každý její uzel je konzistentní.*

Snadno ověříme, že některé sítě lze konzistentně ohodnotit i když obsahují cykly. Cykly rozlišíme na liché a suché podle toho kolik je v cyklu závislostí. Na obr. XX6 a XX5 je nejjednodušší lichý a nejjednodušší sudý cyklus.

Nejjednodušší lichý cyklus nemá konzistentní ohodnocení, ale nejjednodušší sudý cyklus má konzistentní ohodnocení a vzhledem k symetrii má dvě různá konzistentní ohodnocení.

Existují ovšem sítě (viz např. obr. XXX), které lze ohodnotit i více než dvěma různými způsoby. Síť na obr. XXX reprezentuje navzájem se vylučující výroky.

Je-li dána síť, která je částečně ohodnocena, je naším cílem nalézt konzistentní ohodnocení celé sítě, které je rozšířením původního ohodnocení, pokud ovšem vůbec takové ohodnocení existuje. To je množné tehdy, když vyloučíme liché cykly. Síť bez lichých cyklů lze vždy konzistentně ohodnotit. [2]

7.3 Abdukce a logika defaultů

[2]Pro některé sítě s lichými cykly ovšem konzistentní ohodnocení nalézt lze. Pro naše účely si situaci poněkud zjednodušujeme a liché cykly vyloučíme.

Kapitola 8

Usuzování a učení

Teď se budeme zabývat procesem učení, který může být v jistém smyslu chápán jako opak usuzování. Půjde nám o základní pojmy důležité pro pochopení strojového učení z logického hlediska a ukážeme, jak prostředky reprezentace znalostí, tj. především výběr jazyka, a logické zázemí ovlivňují porozumění procesu učení a samozřejmě nám půjde i o efektivnost algoritmů učení a konečně porovnáme algoritmy učení s algoritmy nemonotónní inference.

8.1 Úvod

Začneme definicí učení na abstraktní množině formulí. Náš náhled je náhledem formální logiky, což mimo jiné znamená, že se snažíme vystihnout ty vlastnosti procesu učení, které jsou společné různým postupům, široce studovaným např. v kontextu umělé inteligence, ale pro které zatím není obecně přijatá definice.

Učení zde chápeme jako *proces zlepšování znalostí* na základě příkladů konceptu (pozitivních a negativních), kterému chceme porozumět a naučit se ho. To obvykle znamená, že už máme nějaké výchozí znaloti (background knowledge) B a aspoň jednu množinu E příkladů konceptu, kterému se chceme naučit.[1] Naším cílem je obohatit výchozí znalosti B tím, že zkonstruujeme takovou množinu H formulí, že každý příklad může být (logicky) odvozen z množiny $B \cup H$. Proces učení může být tedy považován za černou skříňku, jejímž vstupem jsou příklady konceptu a výstupem popis naučeného konceptu. Mělo by tedy být zřejmé, že pro adekvátní porozumění myšlence učení z příkladů jsou velmi důležité minimálně dva parametry. Jsou to

- specifikace jazyka (prostředek reprezentace znalostí),
- vymezení pojmu (logické) odvoditelnosti (operace logické konsekvence).

Naším cílem je ukázat jak logika a prostředky reprezentace znalostí ovlivňují sám proces učení, ale nejen ten proces samotný, nýbrž i jeho výsledek. Zaměříme se na učení z příkladů, i když pojmy, kterými zde začínáme, budou užitečné nejen při učení z příkladů.

8.2 Výchozí pojmy

Začneme jako v první kapitole s neprázdnou (obvykle spočetnou, ale možná konečnou) množinou F dobře utvořených formulí daného jazyka J. Libovolnou operaci

$$Cn : \mathcal{P}(F) \longrightarrow \mathcal{P}(F)$$

jsme nazvali *operace logického důsledku* na F, je-li reflexivní, monotónní a tranzitivní.

[1] Později budeme obvykle používat dvě množiny příkladů: E^+ pro pozitivní příklady, E^- pro negativní příklady.

Jestliže bude taková operace na F fixována, řekneme, že je dána logika.

Nechť $B \neq \emptyset$ je teorie, která bude hrát roli *základních znalostí.* Nechť E je neprázdná *množina příkladů* taková, že $E = E^+ \cup E^-$, kde množina E^- je možná i prázdná. Je ale rozumné předpokládat, že množiny pozitivních a negativních příkladů jsou disjunktní, tj. $E^+ \cap E^- = \emptyset$.

Tato definice je přirozená, předpokládáme-li ideálního učitele, ale je realističtější předpokládat, že velké trénovací množiny E mohou obsahovat i konfliktní příklady. Tato skutečnost byla v dosavadních systémech strojového učení jen málo zohledňována.[2]

V abstraktní obecné úrovni definujeme *učení z příkladů* jako úlohu zkonstruovat takovou množinu formulí $H \subseteq F$, že

- $B \cap H = \emptyset$
- $E^+ \subseteq Cn(B \cup H)$
- $E^- \cap Cn(B \cup H) = \emptyset$.

Množinu formulí H budeme nazývat *hypotézy.*

Poznámka. Toto je nejobecnější, či lépe, nejabstraktnější definice učení z příkladů. Její výhodou je opět nezávislost na zvoleném jazyce a na použité logice. Tak např. jazyk pro reprezentaci znalostí může sestávat z Hornových klauzulí, formulí prvního řádu anebo jiných nestandardních schémat reprezentace apod. Významným aspektem našeho přístupu je ovšem *uniformita* prostředků reprezentace, což zde znamená, že jazyk pro základní znalosti, pro hypotézy i pro příklady je týž.

Na druhé straně velká obecnost konceptu učení dovoluje i triviální řešení typu $H = E^+$. Abychom se jim vyhnuli, je třeba

---

[2] O systémech s konfliktními množinami příkladů (současně pozitivními i negativními) pojednává např. [45].

formulovat dodatečné podmínky na hledané hypotézy. Například omezíme prostor hypotéz na ty, které jsou v jistém smyslu maximální či nejobecnější, ale stále ještě relevantní pro proces učení.

K tomu, abychom mohli tyto pojmy definovat přesně, musíme obohatit strukturu jazyka použitého pro reprezentaci znalostí, protože tyto věci silně závisejí na vyjadřovací síle (např. na kombinatorické složitosti) použitého jazyka a v případě jazyků prvního řádu také na počtu predikátů. Tyto požadavky budeme podrobněji diskutovat v další kapitole. Nyní můžeme pouze říci, že prostor hypotéz může být přirozeně uspořádán množinovou inkluzí. Čím více formulí hypotéza obsahuje, tím je specifičtější. Čtenář si může představit všechny formule v hypotéze H spojené logickou konjunkcí a potom místo množinové inkluze mluvit o podformulích (nebo o délce formulí). Tato struktura jazyka ovšem také nestačí, takže navíc předpokládáme, že existuje částečné uspořádání $\preceq$ množiny F všech správně vytvořených formulí, které popisuje jejich relaci vzhledem obecnosti. Řekneme, že formule ψ je zobecněním formule ϕ, pokud $\varphi \preceq \psi$.[3]

8.2.1 Základní vlastnosti abstraktního konceptu učení

Některé vlastnosti abstraktního konceptu učení byly studovány např. Muggletonem v [94]. V tomto odstavci je shrneme v mírně pozměněné, ale obecné formě pro abstraktní množiny formulí a připojíme seznam základních vlastností, které jsou nezávislé na

[3]Například univerzální formule $f(X)$ je obecnější, než kterákoli z následujících jejích konkretizací jako např. $f(a), f(f(a))$, ale třeba i než $f(f(Y))$. Jiným zdrojem pro zobecňování je částečné uspořádání na extenzích predikátů (atributů) jako např. *trojúhelník* $\prec$ *mnohostěn* $\prec$ *planimetrické seskupení*. Zde jsou dva atributy uspořádány vzhledem k jejich oborům možných hodnot.

jazyce.

Především je rozumné předpokládat *netriviálnost*, která říká, že

$$E^+ \cap B = \emptyset.$$[4]

Tato podmínka vyjadřuje, že žádný pozitivní příklad nemůže být logicky odvozen z výchozích znalostí. Poznamenejme, že díky tomu, že výchozí znalosti jsou teorií, může být podmínka netriviálnosti vyjádřena ekvivalentně jako $E^+ \cap Cn(B) = \emptyset$.

Pokud $E^- \neq \emptyset$, tak podmínka vyjadřující, že negativní příklady nesmějí být ve sporu s výchozími znalostmi, může být obecně vyjádřena následujícím způsobem:

$$Cn(E^- \cup B) \neq F.$$

Tuto nerovnost budeme nazývat *vlastnost konzistence.*[5]

Opět je zřejmé, že podmínka konzistence je relativní vzhledem k dané operaci logického důsledku, která blíže specifikuje "theorem prover". Navíc, podmínka konzistence může být vyjádřena dokonce i pro jazyky bez negace.

Definice 8.2.1 *Řekneme, že hypotéza H* pokrývá *množinu formulí G pokud $G \subseteq Cn(B \cup H)$.*

Použitím této terminologie můžeme říci, že *učení z příkladů (vzhledem k B) je úloha najít hypotézu H, která pokrývá všechny pozitivní příklady a nepokrývá žádný negativní příklad.*

Hypotéza H, která pokrývá všechny pozitivní příklady se často nazývá *úplná*; pokud nepokrývá žádný negativní příklad, nazývá se *konzistentní* (vzhledem k E). Úplná konzistentní hypotéza se nazývá *přijatelná* (vzhledem k E).[6]

[4]Tato podmínka se často nazývá *nutná (necessity)* nebo někdy též *přednostní nutná (prior necessity)* podmínka. Viz např. [94].

[5]V [94] se tato podmínka nazývá *silná (strong)* nebo *pozdější podmínka konzistence* (*posterior consistency*).

[6]V publikaci [91] se nazývá *rozlišující (discriminant).*

Není těžké ukázat, že existují úlohy učení, které nelze vyřešit žádnou přijatelnou hypotézou.

Následující lemma tvrdí, že pokud zvětšíme počet příkladů tak, že pozitivní příklady jsou logickými důsledky hypotézy H a negativní příklady nemohou být logicky odvozeny z $B \cup H$, tak nové příklady nemusejí být brány v úvahu, neboť nepřispívají žádnou informací podstatnou k řešení úlohy..

Lemma. Princip skládání platí pro každou přijatelnou hypotézu[7], tj. nechť H je přijatelná hypotéza vzhledem k E a E_1 je množina přidaných příkladů taková, že $E_1 \cup E \neq \emptyset$. Pak, pokud platí $E_1^+ \subseteq Cn(H)$ a $E_1^- \cap Cn(B \cup H) = \emptyset$, tak H je hypotéza přijatelná také pro $E \cup E_1$.

Důkaz. Ihned vyplývá z faktu, že $Cn(B \cup H) = Cn(Cn(B) \cup Cn(H))$.

Soubor $\mathcal{H}$ všech přijatelných hypotéz vzhledem k E je podmnožina množiny $\mathcal{P}(F)$, tj. $\mathcal{H} \in \mathcal{P}(\mathcal{P}(F))$ taková, že každá hypotéza $H \in \mathcal{H}$ pokrývá všechny pozitivní příklady a nepřipouští žádný negativní příklad z E. Pokud se podíváme na soubor $\mathcal{H}$ jako na *prostor stavů*, tak hypotézy odpovídají stavům a (induktiní) usuzovací pravidla odpovídají operacím.[8]

V praktických aplikacích ovšem, i když úplné konzistentní řešení existuje, je často složité najít přijatelnou hypotézu H. Pak je nutné zkoumat některé aproximace. Toto je typická empirická metoda učení. Aproximativní přístup budeme probírat v následujícím odstavci.

Hypotézu H_1 nazveme *specifičtější než* H_2 (vzhledem k množině příkladů E), pokud množina pozitivních příkladů pokrytá H_1 je podmnožinou pozitivních příkladů pokrytých H_2. Analogicky definujeme relaci *"... obecnější než ..."*.

Výsledek učení pak může být definován jako nejspecifičtější

[7]Srovnej [37].

[8]V [135] se tento prostor nazývá *prostor možností (version space).*

hypotéza H taková, že každá formule $\xi \in H$ je *generalizací* nějakého příkladu $\varepsilon \in E^+$, tj. $\varepsilon \preceq \xi$. Nicméně se můžeme ptát, zda nejspecifičtější hypotéza existuje či nikoli. To bude samozřejmě záviset na vyjadřovací síle použitého jazyka.

8.2.2 Aproximativní přístupy k učení

Doposud jsme explicitně neuvedli pojem učitele, jehož role je v učení z příkladů rozhodující. Předpokládáme, že učitel formuluje úlohu učení a dává příklady vyučovanému. Nicméně vyučovaný nemůže pojmout kompletní znalosti učitele, hledá proto nejlepší hypotézu na základě obvykle malých množin příkladů a učitel mu poskytuje nějaké testy, které ukazují, jak dobrá nebo špatná je jeho hypotéza. Někdy dává učitel příklady jeden za druhým, v takovém případě hovoříme o *sekvenčním* nebo *přírůstkovém* učení.

To, že učitel dává vyučovanému nějaké příklady, může být formálně vyjádřeno jako význačná znalost $K \subseteq F$ (ideální hypotéza), která je známa pouze učiteli. Učitel pak může zkontrolovat, zda hypotéza vyučovaného souhlasí s K.

Tento odstavec můžeme uzavřít tím, že řekneme, že učení z příkladů může být formálně charakterizováno jako zobrazení

$$\mathcal{L} : \mathcal{P}(F) \times \mathcal{P}(F) \longrightarrow \mathcal{P}(\mathcal{P}(F))$$

takové, že pro každé $B \subset F$ a $E \subset F$ vytváří soubor hypotéz $\mathcal{H}$ splňující *kritérium kvality*, jak bylo definováno dříve a které reprezentuje prohledávací proceduru v prostoru hypotéz. Takové zobrazení nazýváme *konstruktor znalostí*.[9]

[9] V kontextu konceptového učení by bylo přirozenější nazývat jej *konceptový konstruktor*.

8.3 Prostředky reprezentace znalostí

V této části představíme jazyky (prostředky pro reprezentaci znalostí) používané nejčastěji v oblasti učení z příkladů. Jejich výčet ale pochopitelně nebude vyčerpávající. Začneme s nejjednoduššími (ale nejčastěji studovanými) jazyky, tj. s páry atribut-hodnota a rozhodovacími stromy, a poté budeme pokračovat s prostředky reprezentace s větší výrazovou silou a následně i s větší kombinatorickou složitostí formulí.

8.3.1 Tradiční prostředky

Obecně rozlišujeme dva hlavní druhy popisných jazyků pro “tradiční” učící se systémy: atributový popis a strukturální (relační) popis.

- Atributový popis

Jazyk atributového popisu pro reprezentování příkladů sestává z atributů, jejich oborů a symbolu, řekněme $\times$, pro neznámou hodnotu (ekvivalentní s proměnnou). Příklad je reprezentován tzv. seznamem atributů, což je uspořádaná množina párů atribut-hodnota.

Každý atribut má svůj obor možných hodnot. V závislosti na organizaci oboru atributu rozlišujeme tři základní typy atributů: nominální, lineární a strukturovaný. Toto rozlišení závisí pochopitelně na vyučovaném problému. Obor hodnot nominálního atributu obsahuje nezávislé symboly, tj. nepředpokládáme na něm *apriori* žádnou strukturu. Množina hodnot lineárního atributu je uspořádaná množina, obvykle množina čísel. Na oboru hodnot strukturovaného atributu je definována obvykle grafová struktura orientovaného stromu, tzv. *hiearchický strom*.

Atributový přístup bohužel není “čistým” popisem jazyka, protože neposkytuje jednotnou reprezentaci příkladů a hypotéz. Většina strojových učících se systémů používajících atributový přístup reprezentuje hypotézu pomocí rozhodovacího stromu. Tento strom je jednoduchá rekurzivní struktura charakterizovaná takto: (i) každý list stromu je asociován s třídou, a (ii) každý uzel odpovídá atributu a jeho větve odpovídají množinám vzájemně různých možných hodnot tohoto atributu.

Soubor TDIDT je nejznámějším a široce používaným souborem strojově se učících algoritmů, který využívá atributový přístup. Tyto algoritmy konstruují rozhodovací strom rekurzívním způsobem od kořene k listům, rozdělujíce původní množinu na menší a menší podmnožiny (tzv. metoda *rozděl a panuj divide et conquer*). Pokud všechny příklady právě uvažované podmnožiny náleží do jedné třídy, tak je zkonstruován seznam a označen touto třídou. Jinak je vybrán atribut přinášející největší množství informace a uvažovaná podmnožina je rozdělena na podmnožiny tak, že příklady v každé podmnožině vykazují stejnou hodnotu zvoleného atributu.

- Strukturální (též relační) popis

Strukturální popisy popisují příklady v termínech jejich komponent, jejich relací a hodnot jejich atributů. Mezi nimi je také obvykle znázorněn hiearchický rozklad příkladu až na elementární (dále nerozložitelné) části. Jazyk strukturálního popisu obvykle obsahuje relační a atributové funktory, části a hodnoty atributů. Selektor (více či méně analogický atomické formuli) je buď relační selektor (relační funktor s částmi vyjadřujícími vztah mezi těmito částmi, nebo údaj o pravdivostní hodnotě), nebo atributový selektor (atributový selektor s/bez svých argumentových částí, spolu s hodnotou atributu.

Ve většině strojových učících se empirických aplikací se konjunkce selektorů (ekvivalentní sémantické síti) nazývá komplex [91]. Výhoda strukturálního přístupu spočívá v tom, že obojí -

příklady i prvky hypotéz (formule) - mohou být reprezentovány jako komplexy selektorů (jak bylo definováno výše) a hypotéza tedy jako disjunkce komplexů. Důvod, proč tomu tak je, vyplývá z obecného způsobu konstrukce komplexů: učící se algoritmus dokládá existenci konzistentních komplexů, které nemusejí být úplné. Proto musí být o množině konzistentních komplexů vždy dokázáno, že jejich disjunnkce pokrývá všechny pozitivní příklady.

Soubor AQ strojových učících se algoritmů silně využívá tento strukturální popis, pro který byla vyvinuta modifikace logiky prvního řádu, nazvaná Anotovaný predikátový kalkulus. In their environment, a complex involved in a concept description (hypothesis) is identical to that for examples with two extensions:

1. Pravá velikost libovolného atributového selektoru může obsahovat vnitřní disjunkci svých hodnot nebo interval svých hodnot. Shrnuto, atributový selektor je tvaru $A = R$, kde A je atributový funktor, vnitřní disjunkce nebo interval. Pro zjednodušení můžeme uvažovat selektory s operátory $<, >, \leq, \geq$, pokud byly definovány pomocí intervalů, kde jedna ze závor intervalu bude maximální (resp. minimální) hodnota oboru atributu.

2. Komplex může také obsahovat negace selektorů, atributových i relačních.

Vedle systémů AQx existují jiné algoritmy využívající myšlenku komplexů. Systém CNx algoritmů je mezi nimi zřejmě nejznámější a poskytuje nejširší aplikace. Všechny tyto algoritmy používají pokrývací paradigma. Jeho idea odpovídá filozofii reprezentace hypotéz jako množin formulí (disjunkcí komplexů). Pokrývací algoritmus zkouší vytvořit komplex, který by pokrýval maximální možné množství pozitivních příkladů a byl také významný a předvídající (z určitého úhlu pohledu). Dokud nejsou všechny pozitivní příklady pokryté, pokrývací algoritmus zkouší pokrýt zbývající příklady stejným způsobem, tedy vytvářením disjunkcí komplexů.

Částečné uspořádání $\geq$ na formulích v obou přístupech uvažuje formuli s proměnnou jako nejobecnější. Pro atributy (atributové selektory) existují ještě další kroky k zobecnění, závisející na typu atributu. Zobecnění nominálního páru atribut-hodnota je interval hodnot atributu, a další zobecnění je opět charakterizováno neznámou hodnotou. Jak jsme již zmínili dříve, pár atribut-hodnota strukturovaného atributu může být zobecněn záměnou jeho hodnoty (např. trojúhelníku) hodnotou obecnější v hiearchickém stromu konceptů (např. mnohostěn).

Protože pravidla vytváření jsou široce používané a dobře chápané prostředky pro representování znalostí (zejména v expertních systémech), je rozumné změnit výše uvedené reprezentace hypotéz na množiny takových pravidel. Disjunkce komplexů můžou být snadno přetransformovány v rozhodovací pravidla tím, že vytvoříme jednotlivá pravidla pro každý komjplex originální popisu díky ekvivalenci mezi komplexy a podmínkami pravidel. Rozhodovací strom může být snadno přetransformován na množinu rozhodovacích pravidel popsáním každé cesty skrze strom z kořene do každého listu konjunkcí atributových selektorů: tato konjunkce tvoří podmínku pravidla a třída asociovaná s listem značí pravou stranu rozhodovacího pravidla.

8.3.2 Induktivní logické programy

Jak jsme viděli již dříve, Hornovské klauze jsou nejlépe prozkoumaným prostředkem pro reprezentaci znalostí v oblasti logického programování.

- Induktivní logické programování.

[94]

- Disjunktivní logické programování

Minker and Ruiz, ISMIS'92 [92].

8.3.3 Další prostředky

Přístupy založené na neklasických logikách. Téměř všechny přístupy, ve kterých je použita nejistota k vyjádření míry náležení do třídy, zde mohou být zmíněny. Existuje mnoho článků o tomto tématu, zmiňme nejméně [?].

Jiným zajímavým příkladem použití neklasické logiky je systém KEX (Knowledge Explorer, tzn. Průzkumník znalostí) [12] a metoda analýzy kombinace dat [?], kde je využita spojitost s vícehodnotovou Łukasiewiczovou logikou. Ta je zajímavá tím, že se jedná o jedinou úplnou vícehodnotovou logiku.

8.4 Znovu o usuzování

Usuzování založené na pravidlech je velmi důležitým prostředekem i v kontextu učení. Doposud jsem obsáhli pouze takové operace důsledku, které jsou klasické ve striktně logickém smyslu. Nicméně, logická dedukce je vzhledem k učení příliš silná, specielně vlastnost monotónnosti logické dedukce není realistická, protože v učících se systémech induktivní usuzovací pravidla jsou nejčastěji používána ke konstrukci nových položek znalostí. Zaměříme tedy větší pozornost na ostatní usuzovací systémy, které jsou v literatuře zmíněny jako *nemonotónní* systémy nebo nemonotónní logiky. Nejprve ovšem shrneme v následujícím paragrafu nejdůležitější vlastnosti klasických deduktivních systémů.

8.4.1 A opět logická inference

Jak jsme již zmínili v části 2, libovolná dedukce (operace důsledku) je reflexivní, monotónní a tranzitivní operace na potenční množině dobře vytvořených formulí. Taková relace na formulích byla základním tématem formálního studia logiky v tomto století. Studium operací logického důsledku v nejobecnější formě začal Alfred Tarski a pokračuje dodnes. Viz např. [59].

Jedna z nejdůležitějších vlastností logického důsledku je jeho *kompaktnost* v tomto smyslu: Pro všechna $X \subseteq F$ platí podmínka

$$Cn(X) = \bigcup_{Y_i \subseteq X} Cn(Y_i),$$

kde Y_i jsou všechny konečné podmnožiny X. Je sndadné ověřit, že když Cn je kompaktní operace důsledku, je také monotónní, tzn. čtyři podmínky požadované v definici operace d ˚ usledku nejsou nezávislé. Tento fakt použijeme v dalším odstavci, kde uvedeme operace, které nejsou monotónní. Vzhledem k tomu, že kompaktnost znamená, že důkazy jsou konečné posloupnosti formulí, je monotónnost kombinovaná s kompaktností v tomhle silném smyslu nadbytečná.[10]

Typická operace deduktivního důsledku je syntakticky definována počáteční množinou formulí (čast nazývaných *axiomy*) a konečnou množinou odvozovacích pravidel. $Cn(X)$ je pak definováno jako nejmenší nadmnožina X obsahující axiomy a uzavřená vzhledem ke všem odvozovacím pravidlům.[11]

Je-li takto dána operace důsledku, můžeme množinu X nazývat *deduktivně uzavřenou*, pokud $Cn(X) = X$. Množina X pak sluje *konzistentní (vzhledem k Cn)*, pokud $Cn(X) \neq F$.

[10]Toto je pravda, ale nadbytečnost zmizí při oslabení definice kompaktnosti: to lze snadno nahrazením symbolu rovnosti symbolem podmnožiny.

[11]Jako cvičení je ponecháno ověření faktu, že nejen syntaktická definice logického důsledku splňuje tři vlastnosti Cn, tj. reflexivitu, monotónnost a transitivitu. Čtenář může ověřit, že pokud definujeme operaci na sémantických modelech obvyklým způsobem (tj. formule $\varphi \in Cn(X)$ tehdy a jen tehdy, když každý model X je také modelem φ), získáme opět operaci důsledku.

Mnoho lidí věří, že logické odvozování je nutné k libovolné rozumné rekonstrukci znalostí. V umělé inteligenci (zejména ve strojovém učení) však často potřebujeme usuzovat způsobem, který není striktně deduktivní, tj. operace přiřazující každé množině formulí množinu jejich důsledk ˚ u není monotónní. To je typické např. pro racionální usuzování (commonsense reasoning), které je důležité v procesu učení. Důvodem, proč tomu tak je, je možnost, že vznik nové položky znalostí může zbavit platnosti část znalostí, kterou jsme akceptovali dříve.

8.4.2 Učení a usuzování – dvě strany téže mince

Z naší diskuze o učení a usuzování lze vidět, že na učení a usuzování může být pohlíženo jako na dvě strany téže mince.

Jak vyplývá z části 2, učení může být charakterizováno jako úloha

- *je-li dána výchozí znalost B a (nějaké) příklady E, nalezněte hypotézu H takovou, že $E^+ \subseteq Cn(B \cup H)$.*

Na druhou stranu, z části 4 plyne, že na usuzování může být pohlíženo jako na úlohu

- *je-li dána výchozí znalost B and a hypotéza (defaults) H, najděte maximální (nebo všechny maximnální) konzistentní množinu $C \subset F$ možných důsledků, tj. $C \subseteq Cn(B \cup H)$.*

V úloze učení je znalost (B, E) dána (je nezpochybnitelná, fixovaná) a hledáme hypotézu, která by ji (v nějakém dobře definovaném smyslu) vysvětlila. Protože toto hledání je obvykle doprovázeno pravidly, která mohou zachovávat nepravdu na úkor pravdy, musíme být připraveni modifikovat naši hypotézu v případě, že se objeví nějaká nová nezpochybnitelná položka znalostí.

Část III

Logiky pro racionální usuzování

Kapitola 9

Kondicionální logiky

V této kapitole podáme základy kondicionálních logik jak z hlediska formální sémantiky, tak z hlediska dokazování. Studium kondicionálních logik je relativně nové téma filozofické logiky. Základní ideje pocházejí od [130] a [99]. Důvodem, proč se těmito logikami zde musíme zabývat, je fakt, že většinu nemonotónních systémů, jak jsme je pojednávali v předešlé kapitole, lze chápat jako preferenční logiky a ty právě jsou *de facto* logikami kondicionálními, což vede některé autory k tvrzení, že to jsou právě kondicionální logiky, které jsou těmi pravými logikami nemonotónní inference.

9.1 Motivace

Začneme motivací pro kondicionální logiky.

9.2 Dokazování v kondicionálních logikách

Formální systém kondicionálních logik budeme budovat jako teorii nultého řádu, tj. výrokovou logikou, která bude sestávat ze spočetné množiny výrokových proměnných, obvyklých extenzionálních (pravdivostně hodnotových) výrokových symbolů ($\neg, \wedge, \vee, \Rightarrow$ $, \Leftrightarrow$) a jednoho speciálního binárního symbolu $>$ pro kondicionál. Množina Fle dobře utvořených formulí je definována obvyklým způsobem.

Následují nejčastěji používané axiomy:

ID: $\phi > \phi$
MP: $(\phi > \psi) \Rightarrow (\phi \Rightarrow \psi)$
MOD: $(\neg\phi > \phi) \Rightarrow (\psi > \phi)$
CSO: $((\phi > \psi) \wedge (\psi > \phi)) \Rightarrow ((\phi > \chi)) \Leftrightarrow (\psi > \chi))$
CV: $((\phi > \psi) \wedge \neg(\phi > \neg\chi)) \Rightarrow ((\phi \wedge \chi) > \psi)$
CS: $(\phi \wedge \psi) \Rightarrow (\phi > \psi)$
CA: $((\phi > \psi) \wedge (\chi > \psi)) \Rightarrow ((\phi \vee \chi) > \psi)$
CEM: $(\phi > \psi) \vee (\phi > \neg\psi)$

Pravidla inference jsou tato:

RCEC: Z formule $\phi \leftrightarrow \psi$ odvoď formuli $(\chi > \phi) \leftrightarrow (\chi > \psi)$.
RCK: Z formule $(\phi_1 \wedge \ldots \wedge \phi_n)$ odvoď formuli $[(\chi > \phi_1 \wedge \ldots \wedge (\chi > \phi_n)] \Rightarrow (\chi > \psi)$, pro $n \geq 0$.

Nejčastěji zkoumanými systémy axiomů, které byly v literatuře zkoumány a používají dvou výše uvedených pravidel, jsou:

V = [ID, MOD, CSO, CV]
VW = [ID, MP, MOD, CSO, CV]
VC = [ID, MP, MOD, CSO, CV, CS]
SS = [ID, MP, MOD, CSO, CA, CS]
C2 = [ID, MP, MOD, CSO, CV, CEM]

Avšak následující tři formule nejsou teorémy žádného z pěti výše uvedených systémů:

Tranzitivita:
$[(\varphi > \chi) \wedge (\chi > \psi)] \Rightarrow (\varphi > \psi)$
Kontrapozice:
$(\varphi > \neg\psi) \Rightarrow (\psi > \neg\varphi)$
Zesílení antecedentu (levá monotónnost):
$(\varphi > \psi) \Rightarrow [(\varphi \wedge \chi) > \varphi]$

9.2.1 Stalnackerův model teorie kondicionalů

Stalnackerův model sestává z množiny $W \neq \emptyset$ možných světů, binární (reflexivní) relace $R \subseteq W \times W$, partiální selekční funkce $s : Fle \times W \longrightarrow W$ a funkce $\|\| : Fle \longrightarrow \mathcal{P}(W)$.

Definice 9.2.1 *Kondicionál $\psi > \phi$ je pravdivý, když ψ je pravdivý ve světě, který je nejpodobnější (*most like*) aktuálnímu světu, v němž ψ je pravda.*

Znamená to, že při vyhodnocování kondicionálu přidáme antecedent, tedy formuli ψ k množině přesvědčení (*belief set*) a potom tentativně modifikujeme náš svět jen natolik, aby bylo možno akceptovat novou znalost. Jde tedy o *princip minimální změny.*

Kondicionální logika definovaná Stalnackerovou teorií modelů je nejmenší kondicionální logika, která je uzavřená na dvě pravidla odvozování:
RCEC: Z $\phi \leftrightarrow \psi$ odvoď $(\chi > \phi) \leftrightarrow (\chi > \psi)$
RCK: Z $(\phi_1 \wedge \ldots \wedge \phi_n)$ odvoď $[(\chi > \phi_1 \wedge \ldots \wedge (\chi > \phi_n)] \Rightarrow (\chi > \psi)$, pro $n \geq 0$
a která navíc obsahuje všechny instance formulí: ID, MP, MOD, CSO, CV, CEM. Tato axiomatizace Stalnackerovy kondicionální logiky se v literatuře obvykle nazývá **C2**.

9.2.2 Systémy sfér

Systém sfér je funkce $g : W \longrightarrow \mathcal{P}(\mathcal{P}(W))$ z možných světů do rodiny množin možných světů, která každému možnému světu w přiřazuje do sebe zanořenou množinu $g(w)$ množin světů uzavřených na sjednocení a konečné průniky.

Jestliže $s \in g(w)$, pak každý svět $w \in s$ je považován za uzavřený nebo podobnější w než libovolný svět $w' \notin s$. Pravdivostní podmínky pro kondicionál $\phi > \psi$ potom mohou být vyjádřeny následovně

SOS: $w \in \|\phi > \psi\|$ právě tehdy, když $\bigcap g(w) \cap \|\phi\| = \emptyset$ nebo existuje takový svět $s \in g(w)$, že $s \cap \|\phi\| \neq \emptyset$ a $s \subseteq \|\phi \Rightarrow \psi\|$.

Znamená to, že kondicionál $\phi > \psi$ je pravdivý ve světě w právě když neexistuje sféra kolem w připouštějící formuli ϕ nebo existuje sféra kolem w připouštějící ψ, v níž každý svět pro ϕ je také světem pro ψ.

9.2.3 Problém disjunktivních antecedentů

9.2.4 Problém času a gramatických časů

Kapitola 10

Vícehodnotové a modální logiky

Vícehodnotové a modální logiky jsou rozšířením klasické dvouhodnoté logiky, v nichž se zkoumají takové situace, kdy nejsme schopni určit pravdivostní hodnotu výroku s jistotou, anebo se studují pojmy jako modality možnosti a nutnosti. K nim patří i kondicionální logiky, v nichž jde o studium podmíněných výroků, což je problematika dobře známá např. programátorům.

10.1 Vícehodnotovost a neurčitost

10.1.1 Trojhodnotová logika

V tomto odstavci se na výroky budeme dívat trochu jinak než jako na dvouhodnotová tvrzení typu *ano/ne*. Abychom postihli i situace, kdy nevíme, zda je výrok A pravdivý či nepravdivý, zavedeme třetí "pravdivostní" hodnotu, která bude znamenat *nevím*, a kterou označíme symbolem $\times$.

Výroková spojka negace $\neg$ bude pak definována následující tabulkou:

φ	$\neg\varphi$
1	0
0	1
$\times$	$\times$

Tato tabulka jistě není překvapující. Pro klasické pravdivostní hodnoty pravda a nepravda se shoduje s nám dobře známou tabulkou negace, pro hodnotu $\times$ přirozeným způsobem dává negaci opět hodnotu $\times$, protože

Nevím-li nic o pravdivostní hodnotě výroku φ, nevím samozřejmě nic ani o hodnotě výroku $\neg\varphi$

.

Jde tedy o informační pojetí negace.

Na tomto místě však musíme připomenout, že jsme se dopustili jisté nedůslednosti v označení. Symbol negace zde vlastně změnil svůj význam, jde o modifikovaný pojem, takže kdybychom chtěli být naprosto důslední, měli bychom správně tuto novou, trojhodnotovou negaci označovat jiným symbolem, např $\neg_3$. Pokud ovšem nevznikne nebezpečí nedorozumění, nebudeme tak činit, zvláště když naše trojhodnotová negace je rozšířením klasické dvouhodnotové negace. Podobně budeme postupovat i u ostatních výrokových spojek. Faktickou odlišnost nových spojek však budeme mít stále na mysli. To, co víme o klasické disjunkci, nás vede k této tabulce trojhodnotové disjunkce, kterou bychom opět měli značit spíše symbolem $\vee_3$:

φ	ψ	$\varphi \vee \psi$
1	1	1
1	0	1
1	$\times$	1
0	1	1
0	0	0
0	$\times$	$\times$
$\times$	1	1
$\times$	0	$\times$
$\times$	$\times$	$\times$

Poznámka: Všimněme si ještě, že tuto "dlouhou" tabulku můžeme zapsat i jinou, úspornější, formou

$\varphi \vee \psi$	1	$\times$	0
1	1	1	1
$\times$	1	$\times$	$\times$
0	1	$\times$	0

Trojhodnotovou spojku konjunkce můžeme teď zavést dvěma způsoby. Opět tabulkou, anebo jako zkratku za složenou formuli

$$\neg(\neg\varphi \vee \neg\psi) \ .$$

Snadno se přesvědčíme, že při této definici bude tabulka trojhodnotové konjunkce následující:

φ	ψ	$\varphi \wedge \psi$
1	1	1
1	0	0
1	$\times$	$\times$
0	1	0
0	0	0
0	$\times$	0
$\times$	1	$\times$
$\times$	0	0
$\times$	$\times$	$\times$

Při konstrukci tabulky konjunkce jsme postupovali analogickým způsobem jako při konstrukci pravdivostních tabulek klasické výrokové logiky, ale s tím rozdílem, že výchozí tabulky pro negaci a disjunkci byly jiné.

Podobně jako v klasické dvouhodnotové logice můžeme i zde uvažovat o všech myslitelných jedno-, dvou- a popř. i více-argumentových výrokových spojkách. Zůstaňme na chvíli u jednoargumentových. Ve dvouhodnotové logice, jak jsme viděli, jsou čtyři (negace, asserce, true a false). V trojhodnotové logice takových spojek ovšem můžeme definovat $3^2 \times 3$, tedy 27. To je úctyhodné číslo![1] V běžném jazyce je rozhodně nepoužíváme všechny. Některé jsou ovšem velmi zajímavé, a proto se jim budeme věnovat podrobněji v následujícím odstavci.

Zatím jsme definovali pouze tři trojhodnotové spojky. Otázkou ovšem zůstává, jak definovat implikaci. Tady se ale cesty rozcházejí. Záleží totiž na tom, jak budeme rozumět pojmům implikace, tautologie a vyplývání. Touto problematikou vícehodnotových logik se zabývali zejména Łukasiewicz a Kleene.

Uvedeme dvě různé tabulky implikace. Budeme je rozlišovat

[1] Čtenář si jistě snadno vytvoří tabulku všech sedmadvaceti jednoargumentových spojek.

pomocí indexů podle jejich autorů Łukasiewicze a Kleeneho:

φ	ψ	$\Rightarrow_K$	$\Rightarrow_L$
1	1	1	1
1	0	0	0
1	$\times$	$\times$	$\times$
$\times$	1	1	1
$\times$	0	$\times$	$\times$
$\times$	$\times$	$\times$	1
0	1	1	1
0	0	1	1
0	$\times$	1	1

Jak je vůbec možné, že máme dvě různé definice implikace? To souvisí s motivací pro tyto pojmy, a každý z nich tedy vyjadřuje různé věci. Łukasiewiczova koncepce je historicky starší a je vedena úvahami o budoucích událostech. U výroků “Dnes bylo krásně.” či “Ted prší.” jsme schopni jednoznačně rozhodnout o jejich pravdivosti či nepravdivosti, avšak u některých výroků, které se týkají budoucích událostí, to tak snadné být nemusí. Když např. řekneme, že “Za rok budu v Praze.”, tak mohu mít v pravdivost tohoto výroku větší či menší důvěru, ale o pravdivosti či nepravdivosti nemůžeme nic říci s jistotou. Je možné, že tomu tak bude, je ale též možné, že tomu tak nebude. Naproti tomu Kleeneho koncept trojhodnotové implikave je motivován absencí informace: “Když nevím, tak nevím.”

Různé definice implikace mají pochopitelně za důsledek platnost různých formulí. Tak např. ani v Łukasiewiczově logice ani v Kleeneho logice neplatí zákon vyloučení třetího (*princip tercium non datur*), ani zákon sporu. V Kleeneho logice navíc neplatí zákon dvojité negace, tj.

$$\neg\neg\varphi \Rightarrow \varphi \ .$$

(protože pro $val(\varphi) = \times$ nabývá hodnoty $\times$), zatímco v Łukasiewiczově logice tento princip platí. Čtenář si už jistě všiml, že ale hovořit o tom, že nějaký princip známý z klasické dvouhodnotové logiky v Kleeneho logice neplatí je vlastně triviální, protože zde neexistují žádné tautologie. To je dáno právě tím, že implikace, jejímž antecedentem i konsekventem je hodnota $\times$ nabývá opět hodnoty $\times$.

V dalších odstavcích se seznámíme s tím, jak trojhodnotový přístup může být využit k definici modalit.

Ideu trojhodnotových logik lze též snadno zobecnit na vícehodnotové logiky. Pak ovšem vzniká otázka, co budeme rozumět pod pojmem tautologie. Jedna možnost je striktní, že to jsou právě jen ty formule, které (stejně jako v klasické logice) nabývají hodnoty 1 při každém ohodnocení proměnných. Jiná možnost je hovořit o *A-pravdivých* formulích, tj. formulích, které pro libovolné ohodnocení proměnných nabývají hodnoty z vytčené množiny A pravdivostních hodnot.

10.1.2 Externí negace a operátory jistoty a možnosti

V předešlém odstavci jsme navrhli jedno možné rozšíření klasické negace vzhledem ke třetí pravdivostní hodnotě. Trojhodnotovou spojku negace můžeme kombinatoricky právě vzhledem ke třetí pravdivostní hodnotě chápat ještě dvěma dalšími způsoby: optimisticky (výslednou hodnotou bude 1) nebo pesimisticky (výslednou hodnotou bude 0). Obě takto definované negace vlastně eliminují třetí pravdivostní hodnotu. Někdy se jim také říká *externí negace*.[2] Proto nadále, budeme-li hovořit o externí negaci a nebude-li řečeno jinak, budeme mít vždy na mysli optimistickou externí negaci.

[2]V dalším ještě uvidíme, že stačí, aby jazyk obsahoval jednu z nich, druhou pak snadno dodefinujeme.

Tabulka externí optimistické negace (označíme ji symbolem $\sim$) tedy bude tato:

φ	$\sim\varphi$
1	0
0	1
$\times$	1

zatímco tabulka pesimistické negace (označíme ji symbolem $\sim^p$) bude tato:

φ	$\sim^p\varphi$
1	0
0	1
$\times$	0

Následující tabulka pak přehledně ukazuje souvislost mezi trojhodnotovou negací a externí (optimistickou) trojhodnotovou negací:

φ	$\sim\varphi$	$\neg\varphi$	$\neg\neg\varphi$	$\sim\sim\varphi$	$\neg\sim\varphi$	$\sim\neg\varphi$
1	0	0	1	1	1	1
0	1	1	0	0	0	0
$\times$	1	$\times$	$\times$	0	0	1

Pomocí externí a interní negace nyní můžeme definovat další symboly jazyka $\Box$, $\Diamond$, které budeme chápat jako užitečné zkratky. Symbol $\Box$ pro *externí asserci* či *jistotu*, nebo také *nutnost* a symbol $\Diamond$ pro *možnost*:

Definice 10.1.1 *Definiční rovnosti pro $\Box$ a $\Diamond$ jsou*

$\Box\varphi =_{def} \neg\sim\varphi$

$\Diamond\varphi =_{def} \sim\neg\varphi$

Zřejmě platí

$\neg\Box\varphi = \sim \varphi$

$\neg\Diamond\varphi = \sim\sim \neg\varphi$

To, co vyplývá z definic operátorů jistoty a možnosti, lze přehledně shrnout do těchto tabulek:

φ	$\Box\varphi$	$\sim\Box\varphi$	$\Box\sim\varphi$	$\neg\Box\varphi$	$\Box\neg\varphi$
1	1	0	0	0	0
0	0	1	1	1	1
$\times$	0	1	1	1	0

φ	$\Diamond\varphi$	$\sim\Diamond\varphi$	$\Diamond\sim\varphi$	$\neg\Diamond\varphi$	$\Diamond\neg\varphi$
1	1	0	0	0	0
0	0	1	1	1	1
$\times$	1	0	1	0	1

Všimněme si, že se obě tabulky liší pouze ve třetím řádku, což znamená, že oba operátory $\Box$ a $\Diamond$ jsou opět rozšířením klasických (dvouhodnotových) operátorů nutnosti a možnosti. Z těchto tabulek můžeme vyčíst řadu ekvivalencí (v klasickém, tj. dvouhodnotovém smyslu). Lze ovšem definovat i trojhodnotové ekvivalence. Jednak ve striktním chápání (dvě formule jsou ekvivalentní, mají-li touž pravdivostní hodnotu) nebo v optimistickém pojetí (dvě formule jsou ekvivalentní, je-li možné, aby měly touž pravdivostní hodnotu).

Čtenář snadno ověří, že analogických výsledků dosáhneme také, když vyjdeme od externí pesimistické negace $\sim^p$.

Definice 10.1.2 *Definiční rovnosti pro $\Box$ a $\Diamond$ jsou pak tyto:*

$\Diamond\varphi =_{def} \neg \sim^p \varphi$

$\Box\varphi =_{def} \sim^p \neg\varphi.$

10.1.3 Axiomatizace a odvozování v trojhodnotové logice

Podobně jako v klasické výrokové logice byl hledán systém axiomů, ze kterého by bylo možno odvodit všechny tautologie trojhodnotové logiky v Łukasiewiczově smyslu. Takové systémy nalezli Łukasiewiczovi žáci, Wajsberg a Slupecki.

Wajsbergův systém trojhodnotové logiky obsahuje tyto axiomy

Axiom T1 $\varphi \Rightarrow (\psi \Rightarrow \varphi)$
Axiom T2 $(\varphi \Rightarrow \psi) \Rightarrow ((\psi \Rightarrow \chi) \Rightarrow (\varphi \Rightarrow \chi))$
Axiom T3 $((\varphi \Rightarrow \neg\varphi) \Rightarrow \varphi) \Rightarrow \varphi$
Axiom T4 $(\neg\varphi \Rightarrow \neg\psi) \Rightarrow (\psi \Rightarrow \varphi)$

a obvyklá odvozovací pravidla, tj. modus ponens a pravidlo o substituci. Wajsbergův axiomatický systém je úplný a bezesporný.

10.1.4 Axiomatizace a odvozování v modálních logikách

V tomto odstavci si všimneme modální logiky z hlediska její axiomatické výstavby. Samozřejmě nejjednodušším systémem je modální výroková logika. Její jazyk vznikne z jazyka klasické dvouhodnotové logiky prostým přidáním dalšího symbolu $\Box$, který z výroku vytváří nový výrok, což znamená, že když V je výrok, tak $\Box V$ je také výrok, který budeme ve shodě s tím, co bylo řečeno v předchozích odstavcích, číst *nutně platí V*.

Axiomatických systémů výrokové logiky je známa celá řada. Historicky nejstarším, snad proto že patří k nejjednodušším, je systém S5 [3], který (viz dále) vedle axiomů klasické výrokové logiky obsahuje tyto axiomy týkající se pojmu nutnosti:

$\Box\varphi \Rightarrow \varphi$

[3] Označení pochází od C. I. Lewise.

$\neg\Box\neg\varphi \Rightarrow \Box\neg\Box\neg\varphi$

a distributivnost nutnosti vůči implikaci, tj.

$\Box(\varphi \Rightarrow \psi) \Rightarrow (\Box\varphi \Rightarrow \Box\psi)$

a pro odvozování *modus ponens* a nové pravidlo *zavedení operátoru nutnosti*

$$\frac{\varphi}{\Box\varphi}.$$

Aby však bylo zřejmé, o čem je řeč, je nezbytné říci něco bližšího o sémantice modálních logik. Ta byla podána v Kripkeho publikaci [79].

Kripke vychází z Leibnizova pojetí *možných světů*, což pro nás z logického hlediska bude libovolná neprázdná (nekonečná) množina, kterou označíme W. Pro nás v tuto chvíli bude důležité, že na množině W je definována binární relace $R \subseteq W \times W$, která určuje dosažitelnost či alternativnost možných světů. To tedy znamená, že ke každému světu $w \in W$ je určena množina možných světů z něj dosažitelných (s ním alternativních). Potom řekneme, že formule $\Box\varphi$ je pravdivá ve světě w, jestliže φ je pravdivá ve všech světech dosažitelných z w. Analogicky, $\Diamond\varphi$ je pravdivá ve světě w, jestliže existuje svět alternativní světu w, v němž je pravdivá formule φ.

Nyní je zřejmé, že ale záleží na tom, jaké vlastnosti má konkrétní relace R dosažitelnosti (či alternativnosti).

Byly studovány různé systémy modálních logik vzhledem k vlastnostem relace dosažitelnosti. Mezi nejznámější patří systémy $K, D, T, S4$ a $S5$. Pro ně mj. platí ([139]): Systém K neklade žádné omezení na relaci alternativnosti, je tedy nejslabším modálním systémem. Jestliže od relace R požadujeme, aby ke každému světu existoval aspoň j eden alternativní svět, pak mluvíme o systému D, v němž platí

$$\Box\varphi \Rightarrow \Diamond\varphi.$$

Jestliže R je reflexivní, jde o systém T a platí

$$\Box\varphi \Rightarrow \varphi,$$

Jestliže navíc R je tranzitivní, jedná se o systém $S4$, v němž platí

$$\Box\varphi \Rightarrow \Box\Box\varphi,$$

A konečně, jestliže R je reflexivní, symetrická a tranzitivní, tj. je-li to relace ekvivalence, jedná se o Lewisův systém $S5$, ve kterém navíc platí formule

$$\Diamond\varphi \Rightarrow \Box\Diamond\varphi.$$

Z vlastností relací R je zřejmé, že každý následující systém modální logiky v pořadí $K, D, T, S4, S5$ je rozšířením předchozího.

Zajímavá je interpretace termínu možnosti, když R je lineárním uspořádáním. Pak lze takovou modální logiku interpretovat jako časovou logiku (*tense logic*). Jestliže R je uspořádání reprezentované konečnými stromy, lze termín nutnosti interpretovat jako dokazatelnost (v nějakém formálním kalkulu).

Teď se vrátíme k axiomatizaci modálních logik. Uvidíme, že takových systémů bude více. Budeme pracovat s následujícími axiomy:

Axiom M $\Diamond\varphi \Leftrightarrow \neg\Box\neg\varphi$
Axiom K $\Box(\varphi \Rightarrow \psi) \Rightarrow (\Box\varphi \Rightarrow \Box\psi)$
Axiom T $\Box\varphi \Rightarrow \varphi$
Axiom E $\Diamond\varphi \Rightarrow \Box\Diamond\varphi$
(varianta) $\neg\Box\varphi \Rightarrow \Box\neg\Box\varphi$
Axiom D $\Box\varphi \Rightarrow \Diamond\varphi$
Axiom 4 $\Box\varphi \Rightarrow \Box\Box\varphi$
Axiom B $\neg\varphi \Rightarrow \Box\neg\Box\varphi$ *(tzv. Brouwerova formule)*

Dobře známé Lewisovy systémy S5 a S4 budou pak definovány takto:

S5 = K + T + E (+ M)

S4 = K + T + 4 (+ M)

Pro tyto a další modální systémy například platí následující vztahy obsažení:
$Cn(K) \subset Cn(T) \subset Cn(S4) \subset Cn(S5)$
$Cn(K) \subset Cn(T) \subset Cn(B) \subset Cn(S5)$
Poznámka: První axiom, který jsme zde označili **M**, je možno chápat spíše jako *definici* modality možnosti na základě nutnosti a negace. Ostatní označení je v literatuře o modalitách obvyklé.

Odvozovací pravidlo:
Z formule φ *odvoď* $\Box\varphi$.

Teď se ještě podrobněji seznámíme s modálním systémem S5. Ten požívá mezi ostatními modálními systémy jakési výsadní postavení; důvod lze hledat např. v tom, že sémantická relace alternativnosti mezi možnými světy je zde relace ekvivalence.

Následující formule jsou příklady **teorémů** platných v systému S5:

$\vdash \varphi \Rightarrow \Diamond\varphi$
$\vdash \Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi$
$\vdash \neg\Diamond(\varphi \wedge \neg\varphi)$
$\vdash \Box\Box\varphi \Rightarrow \varphi$
$\vdash \Diamond\Box\varphi \Rightarrow \Box\varphi$
$\vdash \Box\varphi \Rightarrow \Box\Box\varphi$
$\vdash \varphi \Rightarrow \Diamond\Diamond\varphi$
$\vdash \neg\Diamond\varphi \Leftrightarrow \Box\neg\varphi$
$\vdash \Diamond\varphi \Leftrightarrow \neg\Box\varphi$
$\vdash \neg\Diamond\neg\Diamond(\varphi \vee \neg\varphi)$
$\vdash \Box(\varphi \wedge \psi) \Leftrightarrow (\Box\varphi \wedge \Box\psi)$
$\vdash \neg\Diamond(\varphi \vee \psi) \Leftrightarrow (\neg\Diamond\varphi \wedge \neg\Diamond\psi)$

Některé z nich postupně dokážeme:
Dokažme: $\vdash \varphi \Rightarrow \Diamond\varphi$ (1)

1. $\vdash \Box\neg\varphi \Rightarrow \neg\varphi$ (axiom T)
2. $\vdash \neg\neg\varphi \Rightarrow \neg\Box\neg\varphi$ (kontrapozice)
3. $\vdash \varphi \Rightarrow \neg\Box\neg\varphi$ (dvojí negace)
4. $\vdash \Diamond\varphi \Leftrightarrow \neg\Box\neg\varphi$ (axiom M)
5. $\vdash \neg\Box\neg\varphi \Rightarrow \Diamond\varphi$ (fi část u 4)
6. $\vdash \varphi \Rightarrow \Diamond\varphi$ (3,5 hypotetický sylogismus)

Dokažme: $\vdash \Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi$ (2)

1. $\vdash \Diamond\neg\varphi \Leftrightarrow \neg\Box\neg\neg\varphi$ (axiom M, substituce $\neg\varphi$ za φ)
2. $\vdash \Diamond\neg\varphi \Leftrightarrow \neg\Box\varphi$ (dvojí negace)
3. $\vdash \Diamond\neg\varphi \Rightarrow \neg\Box\varphi$ (if část u 2)
4. $\vdash \neg\Box\varphi \Rightarrow \Diamond\neg\varphi$ (fi část u 2)
5. $\vdash \neg\neg\Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi$ (kontrapozice)
6. $\vdash \Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi$ (dvojí negace)

Dokažme: $\vdash \neg\Diamond(\varphi \wedge \neg\varphi)$ (3)

1. $\vdash \neg(\varphi \wedge \neg\varphi)$ (tautologie)
2. $\vdash \Box\neg(\varphi \wedge \neg\varphi)$ (odvozovací pravidlo)
3. $\vdash \Box\neg(\varphi \wedge \neg\varphi) \Leftrightarrow \neg\Diamond\neg\neg(\varphi \wedge \neg\varphi)$
   (předchozí teorém **(2)**, substituce $\neg(\varphi \wedge \neg\varphi)$ za φ)

4. $\vdash \Box\neg(\varphi \wedge \neg\varphi) \Rightarrow \neg\Diamond\neg\neg(\varphi \wedge \neg\varphi)$ (if část u 3)

5. $\vdash \neg\Diamond\neg\neg(\varphi \wedge \varphi)$ (2, 4 modus ponens)

6. $\vdash \neg\Diamond(\varphi \wedge \neg\varphi)$ (dvojí negace)

Dokažme: $\vdash \Box\Box\varphi \Rightarrow \varphi$ (4)

1. $\vdash \Box\varphi \Rightarrow \varphi$ (axiom T)

2. $\vdash \Box(\Box\varphi \Rightarrow \varphi)$ (odvozovací pravidlo)

3. $\vdash \Box(\Box\varphi \Rightarrow \varphi) \Rightarrow (\Box\Box\varphi \Rightarrow \Box\varphi)$ (axiom K)

4. $\vdash \Box\Box\varphi \Rightarrow \Box\varphi$ (2, 3 modus ponens)

5. $\vdash \Box\Box\varphi \Rightarrow \varphi$ (hypotetický sylogismus 4, 1)

Dokažme: $\vdash \Diamond\Box\varphi \Rightarrow \Box\varphi$ (5)

1. $\vdash \Diamond\Box\varphi \Leftrightarrow \neg\Box\neg\Box\varphi$ (axiom M)

2. $\vdash \Diamond\Box\varphi \Rightarrow \neg\Box\neg\Box\varphi$ (if část 1)

3. $\vdash \Box\varphi \Leftrightarrow \neg\Diamond\neg\varphi$ (teorém (2))

4. $\vdash \Box\varphi \Rightarrow \neg\Diamond\neg\varphi$ (if část teorému 3)

5. $\vdash \neg\neg\Diamond\neg\varphi \Rightarrow \neg\Box\varphi$ (kontrapozice 4)

6. $\vdash \Diamond\neg\varphi \Rightarrow \neg\Box\varphi$ (dvojí negace)

7. $\vdash \Box(\Diamond\neg\varphi \Rightarrow \neg\Box\varphi)$ (odvozovací pravidlo)

8. $\vdash \Box(\Diamond\neg\varphi \Rightarrow \neg\Box\varphi) \Rightarrow (\Box\Diamond\neg\varphi \Rightarrow \Box\neg\Box\varphi)$ (axiom K)

9. $\vdash \Box\Diamond\neg\varphi \Rightarrow \Box\neg\Box\varphi$ (7, 8 modus ponens)

10. $\vdash \neg\Box\neg\Box\varphi \Rightarrow \neg\Box\Diamond\neg\varphi$ (kontrapozice 9)

11. $\vdash \Diamond\Box\varphi \Rightarrow \neg\Box\Diamond\neg\varphi$ (2, 10 hypotetický sylogismus)

12. $\vdash \Diamond\neg\varphi \Rightarrow \Box\Diamond\neg\varphi$ (axiom E)

13. $\vdash \neg\Box\Diamond\neg\varphi \Rightarrow \neg\Diamond\neg\varphi$ (kontrapozice 12)

14. $\vdash \Diamond\Box\varphi \Rightarrow \neg\Diamond\neg\varphi$ (11, 13 hypotetický sylogismus)

15. $\vdash \neg\Diamond\neg\varphi \Rightarrow \Box\varphi$ (fi část 3)

16. $\vdash \Diamond\Box\varphi \Rightarrow \Box\varphi$ (14, 15 hypotetický sylogismus)

Dokažme: $\vdash \Box\varphi \Rightarrow \Box\Box\varphi$ (6)

1. $\vdash \Box\varphi \Rightarrow \Diamond\Box\varphi$ (teorém 1, substituce $\Box\varphi$ za φ)

2. $\vdash \Diamond\Box\varphi \Rightarrow \Box\Diamond\Box\varphi$ (axiom E)

3. $\vdash \Box\varphi \Rightarrow \Box\Diamond\Box\varphi$ (1,2 hypotetický sylogismus)

4. $\vdash \Diamond\Box\varphi \Rightarrow \Box\varphi$ (teorém 5)

5. $\vdash \Box(\Diamond\Box\varphi \Rightarrow \Box\varphi)$ (odvozovací pravidlo)

6. $\vdash \Box(\Diamond\Box\varphi \Rightarrow \Box\varphi) \Rightarrow (\Box\Diamond\Box\varphi \Rightarrow \Box\Box\varphi)$ (axiom K)

7. $\vdash \Box\Diamond\Box\varphi \Rightarrow \Box\Box\varphi$ (5,6 modus ponens)

8. $\vdash \Box\varphi \Rightarrow \Box\Box\varphi$ (3,7 hypotetický sylogismus)

10.2 Vlastnosti vícehodnotových a modálních logik

V tomto odstavci si všimneme základních vlastností neklasických logických systémů, a pak uvedeme několik zajímavých vlastností těchto systémů, které se opírají o pojem *stupně informačního uspořádání*.

10.2.1 Úplnost a rozhodnutelnost modální logiky

V rámci Kripkeho sémantiky platí: Systém modální logiky S5 je podobně jako výrokový kalkul úplný a rozhodnutelný. (Viz např. [**?**].)

10.2.2 Další vlastnosti logických kalkulů

Nejprve několik definic.

Definice 10.2.1 *Logický systém je*

1. perzistentní*, když pravdivé (resp. nepravdivé) formule zůstávají pravdivými (resp. nepravdivými), i když jsou přidány další formule.*

2. koherentní*, jestliže libovolná formule nemůže být současně pravdivá i nepravdivá v témže modelu.*

3. determinovaný*, jestliže každá formule je determinovaná, tj. pravdivost nebo nepravdivost formule je jednoznačně určena v úplném modelu.*

4. spolehlivý*, jestliže pravdivost (resp. nepravdivost) formule v částečném modelu má za následek její pravdivost (resp. nepravdivost) i v každém informačním zúplnění.*

Můžeme ověřit, že platí následující tvrzení:

Teorém 10.2.1 *Jestliže systém je perzistentní a determinovaný, pak je spolehlivý.*

Teorém 10.2.2 *Kleeneho trojhodnotová logika je koherentní, determinovaná, perzistentní, a tudíž i spolehlivá.*

10.2.3 Varianty modálních logik

V posledních desetiletích se intenzivně studují varianty modálních logik, které jsou motivovány rozmanitými idejemi od epistemických po výpočtové. Pro orientaci uvedeme alespoň odkazy na hlavní směry. Důležité zde je to, že z jednoho společného východiska pak v závislosti na určité motivaci vytváříme aplikačně zaměřené systémy. Tak např. ve všech uvedených systémech použijeme něco, co bychom mohli nazvat de Morganovy zákony modálních logik, tj. $\Diamond\varphi =_{df} \neg\Box\neg\varphi$. Musíme si ale být vědomi toho, jak se proměňuje koncept negace. Mezi nejintenzivněji studované patří například:

Epistemická logika
$\Box\varphi \ldots$ je známo, že φ
$\Diamond\varphi \ldots$ opak tvrzení φ není znám

Logika přesvědčení (logic of beliefs)
$\Box\varphi \ldots$ věří se, že (panuje názor, že) φ
$\Diamond\varphi \ldots$ v opak φ se nevěří (nepanuje názor)

Deontická logika
$\Box\varphi \ldots$ musí být φ
$\Diamond\varphi \ldots$ je (morálně) dovoleno, že φ

Časová logika (tense logic)
$\Box\varphi\ldots$ vždy bude pravda, že φ
$\Diamond\varphi\ldots$ někdy bude pravda φ

Dynamická (algoritmická) logika
$\Box\varphi\ldots$ po každém ukončení běhu programu je pravda φ
$\Diamond\varphi\ldots$ existuje běh programu, po jehož ukončení je pravda φ

Literatura

[1] **Adámek, Jiří**: Matematické struktury. SNTL, Praha 1982.

[2] **Adams, E.**: The Logic of Conditionals. D. Reidel, Dordrecht 1975.

[3] **Alchurrón, C. E. – Makinson, David**: The logic of theory change: Contraction functions and their associated revision functions. *Theoria 18* (1982), 14–37.

[4] **Amati, Gianni – Carlucci-Aiello, Luigia – Gabbay, Dov – Rirri, Fiora**: A Proof Theoretical Approach to Default Reasoning I: Tablaux for Default Logic. *J. Logic Comput. 6 no. 2* (1996) 205-231.

[5] **Amati, Gianni – Carlucci-Aiello, Luigia – Gabbay, Dov – Pirri, Riora**: Intuitionistic Autoepistemic Logic. *Studia Logica 59* (1997), 103–120.

[6] **Anderson, J. G. – Belnap, Nuel G.**: Entailment. The Logic of Relevance and Necessity. I. Princeton University Press, Princeton 1975.

[7] **Antoniou, G.**: Nonmonotonic Reasoning. The MIT Press, Cambridge, MA 1997.

[8] **Arieli, Ofer – Avron, Arnon**: General Patterns for Nonmonotonic Reasoning. From Basic Entailment to Plausible Reasoning Relations. L. J. of the IGPL, Vol. 8, no. 2, 119-148, 2000.

[9] **Audibert, Laurent – Lhoussaine, Cédric – Schlechta, Karl**: Distance Based Revision of Preferential Logics. *Int. Journal of the IGPL*, Vol. 7, no. 4 (1999) 429-446.

[10] **Bell, John:** The Logic of Time. D. Reidel, Dordrecht 1983.

[11] **van Benthem, Johan**: A Manual of Intensional Logic. Lecture Notes CSLI 1988.

[12] **Berka, Petr**: Expertní systémy. VŠE, Praha 1998.

[13] **Besnard Philippe**: An Introduction to Default Logic. Springer-Verlag, Berlin 1989.

[14] **Bochman, Alexander**: A Logical Theory of Nonmonotonic Inference and Belief Change. Springer 2001.

[15] **Bonatti, P. A. – Eiter, T.**: Querying Disjunctive Databases through Nonmonotonic Logic. In *Proceedings of the 5th International Conference on Databases Theory – ICDT 95*, Springer-Verlag, Berlin 1995, LNCS 893.

[16] **Brewka, Gerhard**: Cumulative Default Logic: In Defense of Nonmonotonic Inference Rules. *Artificial Intelligence 50* (1991) 183-205.

[17] **Busch, Douglas**: *Sequent formalization of three-valued logic.* In: Patrich Doherty (ed.) Partiality, Modality, and Nonmonotonicity. Folli 1996.

[18] **Carnielli, Walter A. - Coniglio, Marcelo E. - Lofredo D'Ottaviano, Itala M. (eds.)**: Paraconsistency. The Logical Way to the Inconsistent. Marcel Dekker, Inc. New York, Basel 2002.

[19] **Mc Carthy, John**: Circumscription: A Form of a Non-monotonic Reasoning. *Artificiall Intelligence 13* (1980) 27-39.

[20] **Chellas, Brian F.**: Modal Logic (An Introduction). Cambridge University Press, Cambridge 1988.

[21] **Chaitin, Gregory J.**: The unknowable. Springer 1999.

[22] **Chagrov, Alexander – Zakharyaschev, Michael**: Modal Logic. Oxford Science Publications, Clarendon Press, Oxford 1997.

[23] **Chang, C. L. – Lee, R. C.**: Symbolic Logic and Mechanical Theorem Proving. Academic Press, New York 1973.

[24] **Church, Alonzo**: Introduction to Mathematical Logic. Princeton University Press, 1996.

[25] **Delahaye, Jean-Paul**: Formal Methods in Artificial Intelligence. John Wiley & Sons, New York 1986.

[26] **Deransant – Maluszyński**: Anotated Logic Programs.

[27] **Dubois, Didier – Hájek, Petr – Prade, Henri**: Knowledge-driven versus data-driven logics. Journal of Logic, Language, and Information 9 (2000) 65-89.

[28] **Dix, Jürgen**: Default Theories of Poole-Type and a Method for Constructing Cummulative Versions of Default Logic. In B. Neumann (ed.), *Proc. of the 10th European Conf.*

on Artificial Intelligence ECAI '92. John Willey & Sons, 1992, 289-293.

[29] **Doyle, J.**: A truth–maintenance system. *Artificial Intelligence, 12* (1979) 231–272.

[30] **Duc, Ho Ngoc**: Reasoning about rational, but not logically omniscient, agents. J. Logic Computat. 7 (1997) no. 5, 633-648.

[31] **Dudley, Underwood**: *What to do when the trisector comes.* The Mathematical Intelligencer. Vol. 5, No. 1 (1983).

[32] **Elkan, Ch.**: A rational reconstruction odf nonmonotonic truth maintenance systems. *Artificiall Intelligence 43* (1990) 219–234.

[33] **Epstein**: Predicate Logic. Oxford??, 19??.

[34] **Etherington, David W. – Reiter, Raymond**: On Inheritance Hierarchies. In: Proc. AAAI-83, Washington, D. C., 1983, 104-108.

[35] **Fariñas del Cerro, L. – Herzig, A. – Lang, J.**: From ordering–based nonmonotonic reasoning to conditional logics. Aritficial Intelligence 66 (1994) 375–393.

[36] **Feys, R.**: Modal Logics. Gauthier-villars, Paris 1965.

[37] **Flach, Peter A.**: Towards a logical theory of inductive learning. Proceedings of the conferrence *Inductive Logic Programming*, Porto, 1991.

[38] **Gabbay, Dov - Guenthner, F. (eds.)**: Handbook of Philosophical Logic. Vol. I–IV. D. Reidel Publ., Dordrecht, 1983.

[39] **Gabbay, Dov**: Theoretical foundations for nonmonotonic reasoning in expert systems. In K. Appt,(ed.): *Logic and Models of Concurrent Systems.* Springer-Verlag, Berlin 1985.

[40] **Gabbay, Dov**: What is Negation in a System? In: Drake, F. R. and Truss, J. K. (eds.) Logic Colloquium '86, Elsevier 1988, 95-112.

[41] **Gabbay, Dov M. - C. J. Hogger, and J. A. Robinson (eds.)**: Handbook of Logic in Artificial Intelligence and Logic Programming. Oxford Science Publications. Clarendon Press, Oxford 1994.

[42] **García, Alejandro J. – Smari Guillermo R. – Chesñevar, Carlos I.**: An Argumentative Framework for Reasoning with Inconsistent and Incomplete Information. In: ECAI '98 Workshop W6 Practical Reasoning and Rationality. Brighton 1998, 13-19.

[43] **Gelfond, M. – Lifschitz, V.**: The stable model semantics for logic programming. In: Fifth International Symposium in Logic Programming, 2, Seattle, WA (MIT Press, Cambridge, MA), 1988, 1070–1080.

[44] **Goldstein, Martin – Hain, Judah**: The Incompleteness Phenomenon. A New Course in Mathematical Logic. AK Peters Ltd., Natick, Massachusetts, 1998, 249 p.

[45] **Gomolińska, Anna**: On Logic of Acceptance and Rejection. In Proc. *Non–Classical Logics in Computer Science*, 1993.

[46] **Gupta, Anil – Belnap, Nuel**: The Revision Theory of Truth. A Bradford Book, The NIT Press, Cambridge, Massachussetts, London, England 1993.

[47] **Gärdenfors, Peter**: Knowledge in Flux. MIT Press, 1988.

[48] **Gärdenfors, Peter**: Propositional logic based on the dynamics of belief. The Journal of Symbolic Logic 50 (1985), no. 2, 390-394.

[49] **Goodwin, James W.**: A theory and system for non–monotonic reasoning. Linköping University, Dept. of Computer and Information Science, Dissertation no. 165, 1987.

[50] **Grégoire, E.**: About the logical Interpretation of ambiguous inheritance hierarchies. University of Maryland Technical Technical Report CS–2452, may 1990.

[51] **Grzegorczyk, Andrzej**: Zarys logiki matematycznej. PWN, Warzsawa 1961. 477 p.

[52] **Hájek, Petr – Godo, Luis**: Deductive systems of fuzzy logic (a tutotial). 1998. 48 p.

[53] **Herbrand, Jaques**: 1929.

[54] **Horty, J. F. – Thomason, R. H. – Touretzky, D. S.**: A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial Intelligence 42* (1990) 311–348.

[55] **Hûet**: The Journal of Logic Programming 197?

[56] **Hughes, G. E. - Cresswell, M. J.**: An Introduction to Modal Logic. Methuen Co., New York, 1969.

[57] **Imielinski, T.**: Results on translating defaults to circumscription. *Artificial Intelligence 32*, no. 1 (1987) 131–146.

[58] **Jirků, Petr**: Theory of Logical Consequence. Acta Universitatis Carolinae, Studia Logica II, Praha 1974, 11-31.

[59] **Jirků, Petr - Materna, Pavel**: Znalosti, logika, usuzování. In: Proceedings SOFSEM '90, Janské Lázně 1990, 187-206.

[60] **Jirků, Petr**: Defeasible reasoning. Linköping University, Dept. of Computer and Information Science, LiTH-IDA-R-90-10, May 1990.

[61] **Jirků, Petr**: On consequence and inference operations. LiTH-IDA-R-90-23 Research Report, Dept. Comp. Inf. Sci., Linköping University, 1990. 18 p.

[62] **Jirků, Petr - Štěpánek, Petr - Štěpánková, Olga**: Programování v jazyku Prolog. SNTL, Praha 1991. 251 p.

[63] **Jirků, Petr**: Rule–based defeasible reasoning. In: Proceedings conference *LOGICA '93*. Liblice, May 1993. FILOSOFIA, Praha 1994, 130–135.

[64] **Jirků, Petr**: Reasoning and Dynamics of Knowledge. Proceedings conference *LOGICA '94*, Liblice, June 12–16. FILOSOFIA, Praha 1995, 201–209.

[65] **Jirků, Petr**: Reasoning about incomplete knowledge. An axiomatic approach. Proceedings *WUPES '97*, Prague 1997, –.

[66] **Jirků, Petr**: How to understand negatives. Proceedings conference *LOGICA 2000*, Liblice, June XX–XX. FILOSOFIA, Praha (v tisku).

[67] **Jirků, Petr**: On interrelatons among extensional three-valued logics. Logic Colloquium '01, Vienna 2001 (v tisku).

[68] **Josephson, J. R.**: On the "Logical Form" of Abduction. 1994.

[69] **Kakas, A. C. – R. A. Kowalski – F. Toni**: Abductive Logic Programming. *Journal of Logic Programming and Computation*, Vol. 2, (1993) no. 6, pp. 719–770..

[70] **Kapitan, Tomis**: Peirce and the Autonomy of Abductive Reasoning. *Erkenntnis 59* (1992) 1–26.

[71] **Kleene, Stephen Cole**: Introduction to Metamathematics. D. van Nostrand Company, Princeton New Jersey, 1952.

[72] **Kleene, Stephen Cole**: Mathematical Logic. John Wiley & Sons, New York 1968.

[73] **Kolář, Petr**: Argumenty filozofické logiky. Filosofia, Praha 1999.

[74] **Konar, Amit**: Artificial Intelligence and Soft Computing, Behavioral and Cognitive Modeling of the Humen Brain. CRC Press 2000. 786 p.

[75] **Konolige, Kurt**: A Deduction Model of Belief (Research Notes in Artificial Intelligence Series). Pitman, London 1986.

[76] **Konolige, Kurt**: On the Relation between Default Theories and Autoepistemic Logic. Proceedings IJCAI-87, Milan 1987, 394–401.

[77] **Kramosil, Ivan**: A note on deductive rules with nagative premises. AIJCAI 75, Tbilisi, 53-56.

[78] **Kraus, S. – Lehmann, D. – Magidor, M.**: Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence 44* (1990) 167–207.

[79] **Kripke, Saul**: A completeness theorem in modal logic. *Journal of Symbolic logic 24* (1959), 1-14.

[80] **Lifschitz, V.**: On the satisfiability of circumscription. *Artificial Intelligence 28* (1986) 17–27.

[81] **Lin, Jinxin**: A semanticx for reasoning consistently in the presence of inconsistency. Artificial Intelligence 86 (1996) 75-95.

[82] **Lloyd, John W.**: Foundations of Logic Programming. Springer–Verlag, Berlin – Heidelberg – New York – Tokyo, 1985. 212 p.

[83] **Long, Derek**: Reasoning by Analogy and Causality: A Model and Application.

[84] **Łukaszewicz, Witold**: Non–monotonic reasoning. Formalization of commonsense reasoning. Ellis Horwood, New York – London – Toronto – Sydney – Tokyo – Singapore 1990.

[85] **Makinson, D.**: General theory of cumulative inference. In: Proceedings 2nd International Workshop on Nonmonotonic Reasoning. Grassau 1988. Springer-Verlag, LNCS 346, 1989.

[86] **Marek, V. Wiktor – Truszczyński, Miroslaw**: Nonmonotonic Logic. Contex–Dependent Reasoning. Springer–Verlag, Berlin Heidelberg New York London Paris Tokyo Hong Kong Barcelona Budapest 1993.

[87] **Martinez, N. G. – Petrovich, A.**: Uniqueness of the Implication for totally ordered MV-algebras. Annals of Pure and Applied Logic, Vol. 108 (2000) 261–268.

[88] **Mařík, V. - Štěpánková, O. - Lažanský, J. (eds.)**: Umělá inteligence (2). Academia, Praha 1997. 373 p.

[89] **McDermott, Drew – Doyle, Jon**: Non-Monotonic Logic I. Artificial Intelligence 13 (1980) 41–72.

[90] **McDremott, Drew**: Nonmonotonic Logic II: Nonmonotonic Modal Theories. Journal fo the Association for Computing Machinery 29 (1982) 33–57.

[91] **Michalski, R. S. – Carbonell, J. G. – Mitchell, T. M.**: Machine Learning. An Artificial Intelligence Approach. Tioga Publ. Comp., Palo Alto 1983. 572 p.

[92] **Minker, J. – Ruiz, C.**: On Extended Disjunctive Logic Programs. In Komorowski, J. – Raś, Z. W. (eds.): *Methodologies for Intelligent Systems.* 7th INt. Symposium ISMIS'93, Springer–Verlag, 1993, 1–19.

[93] **Moore, R. C.**: Semantical Considerations on Nonmonotonic Logic. Proceedings IJCAI-83, Karlsruhe, FRG, 1983, 272-279.

[94] **Muggleton, Stephen**: Inductive Logic Programming: derivations, successes and shortcommings. In: *Machine Learning: EMCL–93.* LNAI 667, Springer–Verlag 1993, 21–37.

[95] **Nebel, B.**: Based Revision Operations and Schemes: Semantics, Representations, and Complexity.

[96] **Nebel, Berhard**: How hard is it to revise a belief base. In: D. M. Gabbay and Ph. Smets (eds.) Handbook of Defeasible Reasoning and Uncertainty Management Systems, Vol. 3, 77-145. Kluwer, Netherlands 1998.

[97] **Niemalä. Ilkka**: A dwecision method for nonmonotonic reasoning based on autoepistemic reasoning. Journal of Automated Reasoning 14 1995 3-42.

[98] **Nilsson, Nils J.**: Principles of Artificial Intelligence. Tioga, Palo Alto 1980.

[99] **Nute, David**: Conditional logic. In: Gabbay D. and Guenthner (eds.), Handbook of Philosophical Logic, Vol. II, D. Reidel, Dordrecht 1984, 387–439.

[100] **Nute, David**: Defeasible reasoning: A philosophical analysis in Prolog. In: James H. Fetzer (ed.) Aspects of Artificial Intelligence. Kluwer Academic Publishers, The Netherlands, 1988, 251–288.

[101] **Ohlbach, Hans Jürgen**: Translation methods for non/classical logics: An Overwiew. Bull. of the IGPL 1 (1993) no. 1, 69-89.

[102] **Ohlbach, Hans Jürgen – Koehler, Jana**: Modal logics, description logics and arithmetic reasoning. *Artificial Intelligece*, 109 (1999) 1-34.

[103] **Ohlbach, Hans Jürgen – Reyle Uwe (eds.)**: Logic, Language and Reasoning. Kluwer, Dordrecht 1999.

[104] **Peirce, Charles, Sanders**: Abduction and Induction. Dover 1955.

[105] **Peng, Yun – Reggia James A.**: Abductive Inference Models for Diagnostic Problem–Solving. Springer–Verlag, New York – Berlin – Heidelberg – London – Paris – Tokyo – Hong Kong 1990. 284 p., 25 figs.

[106] **Peot, Mark, A. – Shachter, Ross, D.**: Fusion and propagation with multiple observations in belief networks. Artificial Intelligence 48 (1991) 299-318.

[107] **Pople, H.**: On the Mechaniyation of Abductive Logic. *Proceedings of the International Joint Conference on Artificial Intelligence* IJCAI 1973, 147–152.

[108] **Przymusiński, T. C.**: On the Relationship between Logic Programming and Non-monotonic Reasoning. Proceedings AAAI-88, 1988, 444-448.

[109] **Quinlan, J. Ross**: Learning logical definitions from relations. *Machine Learning*, (5) 1990, 239–266.

[110] **de Raedt, Luc**: Interactive concept–learning. PhD thesis, Dept. of Computer Science, Katholic University, Leuven, 1991, 213 p.

[111] **Rasiova, Helena - Sikorski, Roman**: The Mathematics of Metamathematics. PWN, Warszawa 1957.

[112] **Reinfrank, Michael**: Fundamentals and logical foundations of truth maintenance. Linköping Studies in Sciece and Technology, Dissertations no. 221, 1989.

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

[114] **Reiter, Raymond**: A Logic for Default Reasoning. Artificiall Intelligence 13 (1980) 81-132.

[115] **Rieger, Ladislav**: Algebraic Methods of Mathematical Logic. Academia, Praha 1967.

[116] **Robinson, Alan – Voronkov, Andrei (eds.)**: Handbook of Automated Reasoning. Vol. I, II. Elsevier, Amsterdam and MIT Press, Cambridge, Mass. 2001.

[117] **Robinson, Julia A.**: A Machine Oriented Logic Based on the Resolution Principle. Journal ACM 12 (1965), 23-41.

[118] **Sandewall, Erik**: An approach to the frame problem and its implementation. In: B. Meltzer and D. Mitchie (eds.): Machine Intelligence 7. Edinburgh University Press, Edinburgh 1972.

[119] **Sandewall, Erik**: Nonmonotonic inference rules for multiple inheritance with exceptions. Proceedings of the IEEE, vol. 74, no. 10, October 1986, 1345–1353.

[120] : Some completeness results for stoppered and ranked classical preferential models. J. Logic Computat., 6 (1996), no. 4, 599-622.

[121] **Schurz, Gerhard**: Probabilistic semantics for Delgrande's conditional logic and counterexamlpe to his default logic. Artificial Intelligence 102 (1998) 81–93.

[122] **Schurz, Gerhard - Uršič, Marko**: Bezoend Classical Logic. Philosophical and Computational Investigations in Deductive Reasoning and Relevance. Academia Verlag, Sankt Augustin 1999. 190 p.

[123] **Shoham, Yoav**: Nonmonotonic Logics: Meaning and Utility. Proc. IJCAI-87. Milano 1987, 388-393.

[124] **Shoham, Yoav**: Reasoning about Change. MIT Press, 1988.

[125] **Shönfield, J. R.**: Mathematical logic. Addison-Wesley Publishing Comp., Reading (Mass.) 1967.

[126] **Shwaytser, H.**: A Necessity Condition for Learning from Positive Examples. *Machine Learning* (5) 1990, 101–113.

[127] **Singh, Munindar, P. – Asher, Nicholas, M.**: A logic of intensions and beliefs. Journal of Philosophical Logic 22 (1993) 513-544.

[128] **Simon, Herbert A.**: Does Scientific Discovery Have a Logic? *Philosophy of Science 4* (1973) 471-480.

[129] **Sochor, Antonín**: Matematická logika. Karolinum - nakl. Univerzity Karlovy, Praha 2000 (v tisku).

[130] **Stalnaker, R.**: A theory of conditionals. In: N. Rescher (ed.) *Studies in Logical Theory.* American Philosophical Quarterly Monograph Series, no.2, Blackwell, Oxford 1968.

[131] **Stalnacker, R. C.**: A note on nonmonotonic logic. Dept. of Philosophy, Cornell University, 1980.

[132] **Šefránek, Ján**: Inteligencia ako výpočet, Vydavatelstvo IRIS, Bratislava, 2000.

[133] **Tarski, Alfred**: Logic, Semantics, Metamathematics. Papers from 1923 to 38, translated by J. H. Woodger. Clarendon Press, Oxford 1956.

[134] **Thayse, André (ed.)**: From Modal Logic to Deductive Databases. Introducing a Logic Based Approach to Artifical Intelligence. John Wiley & Sons, Chichester – New York – Brisbane – Toronto – Singapore, 1989.

[135] **Thayse, André (ed.)**: From Natural Language Processing to Logic for Expert Systems (A Logic Based Approach to Artificial Intelligence). John Wiley & Sons, Chichester – New York – Brisbane – Toronto – Singapore, 1991.

[136] **Thirunarayan, Krishnaprasad**: A theory of nonmonotonic inheritance based on annotated logic. Artificial Intelligence 60 (1993) 23–50.

[137] **Touretzky, D.**: The Mathematics of inheritance. Pitman Research Notes in Artificial Intelligence, London 1986.

[138] **Turing, Alan**: Can Computers Think? 1950.

[139] **Turner, Raymond**: Truth and Modality for Knowledge Representation. Pitman, London 1990.

[140] **Urquhart, A.**: Many-valued logic. In: D. M. Gabbay and F. Guenther (eds.) Handbook of Philosophical Logic, vol III, D. Reidel 1986, 71-116.

[141] **Wagner, Gert**: Vivid logic. Knowledge-based reasoning with two kinds of negation. LNAI 765, Springer Verlag 1994.

[142] **Wójcicki, Ryszard**: Theory of Logical Calculi (Basic Theory of Consequence Operations). Kluwer Academic Publishers, Dordrecht 1988.

[143] **Bochman, A.**: A Logical Theory of Nonmonotonic Inference and Belief Change. Springer 2001.

[144] **Chellas, Brian**: Modal Logic: An Introduction. Cambridge University Press 1980.

[145] **Dretske, F.**: Epistemic Operators. The Journal of Philosophy **67** (1970), 1007-1022

[146] **Duc, Ho Ngoc**: Reasoning about rational, but not logically omniscient, agents. J. Logic. Comput., **7** (1997) No. 5, 633-648.

[147] **Edgington, D.**: The Paradox of Knowability. Mind **94** (1985) 557-568.

[148] **Fagin, R., Halpern, J. Y., Moses, Y. and Vardi, M. Y.**: Reasoning about Knowledge. Cambridge: MIT Press, 1995.

[149] **Haack, S.**: Evidence and Inquiry. Towards a reconstruction in Epistemology. Blackwell, Oxford 1993.

[150] **Hendrix, V., F.**: Active Agents. In: Hendrix, Pedersen (eds.) Ψ News, Vol. 2, 0ctober 2002 The Danish Network for Philosophical Logic and Its Applications. 5–40.

[151] **Hintikka, Jaakko**: Knowledge and Belief: An Introduction to the Logic of the Two Notions. Cornell Univ. Press, Ithaca, 1962.

[152] **Hájek, Petr and Petr Jirků**: Glance and Mysery of Expert Systems (Lesk a bída expertních systémů) (in Czech), Sofsem (1985) 131–173.

[153] **Johnson-Laird, P. N. and R. M. J. Byrne**: Deduction. Lawrence Erlbaum Assoc. Exeter 1991.

[154] **Kripke, Saul**: Semantical Analysis of Modal Logic I: Normal Propositional Calculi. Zeitschrift für Mathematsche Logik und Grundlagen der Mathematik 8 (1963) 67-96.

[155] **Kripke, Saul**: Semantical Analysis of Modal Logic II: Non-Normal Propositional Calculi. In: Addison, Henkin, and Tarski (eds.) The Theory of Models.1965, 206-220.

[156] **Levesque, H.**: A logic of implicit and explicit belief. Proceedings AAAI-84 (Austin TX, 1984) 198-202.

[157] **Lomuscio, A. and Ryan, M.**: Ideal agents sharing (some!) knowledge. ECCAI '98, 13th European Conference on Artificial Intelligence 1998, 557-561.

[158] **Marek, V. W. and M. Truszczyński**: Nonmonotonic Logic. Context-Dependent Reasoning. Springer/Verlag 1993.

[159] **Priest, G.**: An Introduction to Non-Classical Logic. Cambridge University Press, 2001.

[160] **Segerberg, Krister**: An Essay in Classical Modal Logic. Vols. 1-3, University of Uppsala, 1971.

[161] **de Swart, H. and C. Rauszer**: Different Approaches to Knowledge, Common Knowledge and Auman's Theorem.

[162] Shoham, Yohav and Y. Moses: Belief as defeasible knowledge. Proceedings IJCAI-89, Morgan Kaufmann, 1992, 214-228.

[163] **Thagard, P. and Vergeurgt, K.**: Coherence as Constraints Satisfaction. Cognitive Science 22 (1998) 1-24.

[164] **Thagard, P. – Eliasmith, Ch. – Rusnock, P. – Shelley, C.**: Knowledge and Coherence. Waterloo. URL: http://cogsci.uwaterloo.ca/Articles/Pages/epistemic.html 2002.

[165] **Thayse, A. (ed.)**: From Modal Logic to Deductive Databases. Introducing a Logic Based Approach to Artificial Intelligence. Joh Willey & Sons, 1989.

[166] **Turner, R.**: Truth and Modality for Knowledge Representation. Pitman 1990.