Nerozhodnuteľnosť a neúplnosť

12. prednáška
Logika pre informatikov a Úvod do matematickej logiky

Ján Kľuka, Ján Mazák, Jozef Šiška

Letný semester 2023/2024

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

Obsah 12. prednášky

Nerozhodnuteľnosť a neúplnosť

Cieľom tejto prednášky je jemne doplniť filozofický, matematický, algoritmický a historický kontext týkajúci sa logiky.

Nie je nutné, aby ste uvedené veci podrobne ovládali — ide o veľké idey, nezaujímame sa o technické detaily.

Nerozhodnuteľnosť a neúplnosť

Logika 1. a 2. rádu

Zhrnutie: logika 1. rádu

Uvažujme jazyk logiky 1. rádu a v ňom teóriu T a formulu F.

Veta 17.1 (Korektnosť logiky 1. rádu)

Ak $T \vdash F$, tak $T \vDash F$.

Veta 17.2 (Úplnosť logiky 1. rádu)

Ak $T \vDash F$, tak $T \vdash F$.

Veta 17.3 (Kompaktnosť logiky 1. rádu)

Ak je teória T nesplniteľná, existuje konečná podmnožina T, ktorá je nesplniteľná.

Ako vyjadriť indukciu

Predstavme si, že chceme pridať axiómu umožňujúcu využívať v dôkazoch matematickú indukciu. Nech P^1 je predikát a S^1 funkčný symbol pre nasledovníka. Chceli by sme

$$\forall P \quad \Big(\big(P(0) \wedge \forall x\big(P(x) \to P(S(x))\big)\big) \to \forall x\, P(x)\Big).$$

Toto je však formula logiky 2. rádu — v logike 1. rádu nič ako $\forall P$ pre predikát P nemáme. Jediné, čo nám ostáva, je pridať axiómu

$$\big(P(0) \wedge \forall x\big(P(x) \to P(S(x))\big)\big) \to \forall x\, P(x),$$

kde P je predikát **vyjadriteľný formulou** v našom konkrétnom prvorádovom jazyku $\mathcal{L}$. To má nevýhody:

- dôkaz využívajúci túto axiómu bude platiť len pre jeden konkrétny predikát;
- dôkazy sa budú týkať len predikátov vyjadriteľných v $\mathcal{L}$.

Ako vyjadriť tranzitívny uzáver

Majme binárnu reláciu reprezentovanú predikátom R. Pre každé $n \geq 1$ označme $R_n(a, b)$ formulu

$$\exists x_1\, \exists x_2 \cdots \exists x_n \big(R(a, x_1) \wedge R(x_1, x_2) \wedge \cdots \wedge R(x_n, b)\big)$$

(jednotlivé formuly R_n sú konečné, ale je ich nekonečne veľa — pre každé n by sme mohli náš jazyk rozšíriť o predikát, ktorý R_n definuje).

Ďalej nech $R_0 = R$ a nech

$$R^*(a, b) \quad \text{vtt} \quad \text{existuje } n \text{ také, že platí } R_n(a, b).$$

Naše R^* vyjadruje tranzitívny uzáver relácie R.

Dá sa však R^* vyjadriť prvorádovou formulou?

Ako vyjadriť tranzitívny uzáver

Predpokladajme, že R^* sa dá vyjadriť prvorádovou formulou.

Uvažujme nekonečnú prvorádovú teóriu

$$T = \{\neg R_0(a,b), \neg R_1(a,b), \ldots, \neg R_n(a,b), \ldots, \quad R^*(a,b)\}.$$

1. T nemá model: ak $R^*(a,b)$, tak pre nejaké n platí $R_n(a,b)$.
2. Každá konečná podmnožina T má model: ak n je najväčší z indexov R_n obsiahnutých v podmnožine, vyhovujúci model je napr. $D = \mathbb{N}$, $i(R) = \{(0,1),(1,2),\ldots,(n+1,n+2)\}$, $i(a) = 0$, $i(b) = n+2$.

Tieto dve pozorovania sú však v spore s kompaktnosťou logiky 1. rádu.

Formulou logiky 1. rádu tranzitívny uzáver nevieme vyjadriť.

Ako vyjadriť tranzitívny uzáver

V logike 2. rádu by to už šlo:

$$R^*(a,b) \leftrightarrow$$
$$\forall P \left(\Big(\forall x \left(R(a,x) \rightarrow P(x) \right) \wedge \forall x \, \forall y \left(P(x) \wedge R(x,y) \rightarrow P(y) \right) \Big) \rightarrow P(b) \right)$$

Tento popis je založený na tom, že vlastnosť „b *je* R*-potomkom* a“ interpretujeme ako „b dedí od a každú vlastnosť, ktorú majú všetci priami R-potomkovia a a zároveň sa zachováva cez R“.

Syntax a sémantiku logiky 1. rádu možno rozšíriť tak, aby umožňovala kvantifikovanie cez predikáty; dostaneme tak logiku 2. rádu.

Použitie formúl 2. rádu zvyšuje vyjadrovaciu silu jazyka, ale zároveň stratíme kompaktnosť a úplnosť.

Nerozhodnuteľnosť a neúplnosť

Peanova aritmetika

Peanova aritmetika

Jednou z možností, ako formalizovať aritmetiku, je Peanova aritmetika (PA). Jazyk vychádza z prvorádovej logiky s rovnosťou, indivíduová konštanta je 0, funkčné symboly sú S^1, $+^2$, $\cdot^2$ (nasledovník, sčítanie, násobenie). Prirodzené čísla reprezentujeme ako termy, napr. 3 je $S(S(S(0)))$. Axiómy:

$$\begin{aligned}
&\forall x \;\; 0 \neq S(x) \\
\forall x\, \forall y \;\; &(S(x) \doteq S(y) \to x \doteq y) \\
&\forall x \;\; x + 0 \doteq x \\
\forall x\, \forall y \;\; &x + S(y) \doteq S(x + y) \\
&\forall x \;\; x \cdot 0 \doteq 0 \\
\forall x\, \forall y \;\; &x \cdot S(y) \doteq (x \cdot y) + x
\end{aligned}$$

Peanova aritmetika

K uvedeným axiómam treba pridať axiómu pre indukciu. Máme dve možnosti:

- Formuly 1. rádu.
  Nevýhoda: neštandardné modely (aj s nespočítateľnou doménou); ich existencia vyplýva o.i. z vety o kompaktnosti.
- Formula 2. rádu.
  Výhoda: všetky modely sú izomorfné s prirodzenými číslami.

Peanova aritmetika

Prečo len tri funkčné symboly? Nebolo by výhodné pridať napr. umocňovanie?

Netreba: prvorádová Peanova aritmetika umožňuje formulou popísať **každú *vypočítateľnú* funkciu**.

Označme $\overline{x}$ reprezentáciu prirodzeného čísla x v PA (ako termu vznikajúceho opakovanou aplikáciou symbolu S). Funkciu f^1 možno reprezentovať akousi formulou F takou, že $F(\overline{x}, \overline{y})$ je pravda vtt $f(x) = y$.

Navyše, ak $f(x) = y$, v PA možno dokázať $F(\overline{x}, \overline{y})$,
a ak $f(x) \neq y$, možno dokázať $\neg F(\overline{x}, \overline{y})$.

Nerozhodnuteľnosť a neúplnosť

Nerozhodnuteľnosť

Nerozhodnuteľnosť platnosti formuly vo FOL

Veta 17.4

Platnosť formuly v prvorádovej logike je nerozhodnuteľná.

Dôkaz: redukciou na problém zastavenia.

Ukážeme, ako by sme vedeli rozhodnúť, či daný Turingov stroj M zastaví, ak by sme vedeli rozhodovať platnosť prvorádových formúl.

Nerozhodnuteľnosť platnosti formuly vo FOL

Jazyk našej logiky bude obsahovať

- indivíduová konštanta ε pre prázdny reťazec;
- unárny funkčný symbol a^1 pre každé písmeno a v abecede;
- binárny predikát f_q pre každý stav q TS M.

Naša interpretácia tohto jazyka:

- $a(w)$ označuje reťazec aw;
- $f_q(x, y)$ indikuje, že M dosiahne na danom vstupe stav q, pričom na páske je reťazec $\overline{x}y$ (x v opačnom poradí) a hlava M je na prvom znaku y.

Nerozhodnuteľnosť platnosti formuly vo FOL

Krok výpočtu zachytáva formula

$$\forall x\, \forall y \quad f_q(x, a(y)) \to f_{q'}(b(x), y)$$

M prečíta z pásky a, zapíše b, prejde zo stavu q do stavu q' a posunie hlavu doprava. (Takúto formulu pridáme do teórie popisujúcej činnosť M pre každú dvojicu a, b.)

Pre pohyb hlavy doľava máme formulu

$$\forall x\, \forall y \quad f_q(c(x), a(y)) \to f_{q'}(x, c(b(y))).$$

Nerozhodnuteľnosť platnosti formuly vo FOL

Pre polohu hlavy na ľavom okraji pásky pridáme

$$\forall y \quad f_q(\varepsilon, a(y)) \to f_{q'}(\varepsilon, b(y))$$

(hlava sa nehýbe, len prepisuje znak na páske).

Podobne doriešime aj polohu na pravom okraji, keď sa hlava posunie na časť pásky, kam sa ešte nezapisovalo.

Konkrétne detaily závisia od uvažovaného variantu Turingovho stroja (páska môže byť obojstranne nekonečná apod.).

Nerozhodnuteľnosť platnosti formuly vo FOL

Začiatok výpočtu z počiatočného stavu q_0 na slove w popisuje formula $f_{q_0}(\varepsilon, w)$ a zastavenie stroja v (jedinom) akceptačnom stave q_{acc} vyjadruje formula

$$F_M = f_{q_0}(\varepsilon, w) \wedge T \rightarrow \exists x \, \exists y \;\; f_{q_{\text{acc}}}(x, y),$$

kde T je konjunkcia implikácií popisujúcich povolené prechody TS M.

Každý akceptačný výpočet M vieme prerobiť na dôkaz F_M (stačí opakovane používať modus ponens na príslušné implikácie). Keďže prvorádová logika je korektná, tak ak M zastaví, F_M je platná formula.

Naopak, ak F_M je platná, tak je pravdivá v každej intepretácii (štruktúre), čiže aj v tej našej týkajúcej sa TS M. Pritom premisy F_M sú v nej splnené, preto musí byť splnený aj záver $\exists x \, \exists y \, f_{q_{\text{acc}}}(x, y)$, takže M zastaví.

Čiastočná rozhodnuteľnosť platnosti formuly

Veta 17.5

Platnosť formuly v prvorádovej logike (so spočítateľným jazykom) je čiastočne rozhodnuteľná.

Dôkaz: stačí enumerovať všetky dôkazy v danom jazyku.

Formula je platná vtt jej negácia je nesplniteľná. Dôsledky:

Veta 17.6

(1) Nesplniteľnosť formuly v prvorádovej logike je nerozhodnuteľná a čiastočne rozhodnuteľná.
(2) Splniteľnosť v prvorádovej logike nie je ani čiastočne rozhodnuteľná.

Dôkaz (2): ak by sme splniteľnosť vedeli rozhodovať čiastočne, môžeme paralelne spustiť testovanie splniteľnosti aj nesplniteľnosti, a jeden z týchto výpočtov by musel skôr či neskôr skončiť, čím by sme vedeli rozhodovať nesplniteľnosť formuly, a to je spor.

Monadic FOL

Monadická prvorádová logika (MFOL) je prvorádová logika, v ktorej nemáme funkčné symboly a predikáty majú len jeden argument. (V takomto jazyku možno vyjadriť bežné sylogizmy a počas väčšiny 19. storočia sa verilo, že postačuje na formalizáciu uvažovania.)

Veta 17.7 (1915)

Pravdivosť formuly v MFOL je rozhodnuteľná.

Ak pridáme čo len jeden predikát arity 2, stratíme rozhodnuteľnosť.

Tento príklad ilustruje, že zvoliť jazyk s veľkou vyjadrovacou silou (napr. pre databázový systém) neraz prináša algoritmické problémy, preto to nie je samozrejmá voľba.

Nerozhodnuteľnosť a neúplnosť

Deskriptívna zložitosť

Deskriptívna zložitosť

Pri klasickom pohľade na zložitosť algoritmov skúmame, koľko krokov spraví Turingov stroj v závislosti od veľkosti vstupu, resp. či Turingov stroj s dodatočnými obmedzeniami vôbec vie problém riešiť.

Alternatívny pohľad: akú zložitú logickú formulu potrebujeme na popis daného problému (jazyka akceptovaného Turingovým strojom)?

Príklady rôzne zložitých formúl:

- prvorádová logika (FOL)
- druhorádová logika (SOL) — kvantifikujeme aj predikáty/množiny objektov
- existential SOL: $\exists X_1\, \exists X_2 \cdots \exists X_n\ F$, F je formula FOL
- universal SOL: $\forall X_1\, \forall X_2 \cdots \forall X_n\ F$, F je formula FOL

Deskriptívna zložitosť

Existencia trojuholníka v grafe sa dá vyjadriť vo FOL:

$$\exists x\, \exists y\, \exists z\ \big(V(x) \wedge V(y) \wedge V(z) \wedge \\ \wedge\, x \neq y \wedge y \neq z \wedge z \neq x \wedge E(x,y) \wedge E(y,z) \wedge E(z,x)\big)$$

Pri testovaní, či konkrétny graf spĺňa túto formulu, nerozhodujeme o platnosti formuly vo všeobecnosti (to je nerozhodnuteľný problém), ale len vyhodnocujeme jej splnenie v konkrétnej interpretácii: predikát V popisuje vrcholy, E hrany.

Pri vyhodnotení kvantifikátora stačí preskúmať všetky vrcholy, čiže pre k kvantifikátorov vo formule preveríme $O(|V|^k)$ možností. Preto tento problém patrí do triedy P (zjemnenie úvah umožňuje dokázať príslušnosť do LOGSPACE).

Deskriptívna zložitosť

Existencia 3-farbenia grafu: NP-complete;
nevyjadriteľné vo FOL, vyjadriteľné v existential SOL:

$$\exists R\, \exists G\, \exists B \big(\forall x\, \forall y\, (E(x,y) \to \neg R(x) \vee \neg R(y)) \wedge \cdots \big)$$

(R, G, B sú predikáty vyjadrujúce jednotlivé farby).
Pri vyhodnocovaní formuly pre konkrétny graf za doménu zoberieme vrcholy grafu. Existencia predikátu zodpovedá existencii podmnožiny všetkých vrcholov; túto podmnožinu vieme nedeterministicky uhádnuť a vyhodnotenie prvorádovej časti formuly už pridá len polynomiálny faktor (exponent závisí od počtu vnorených kvantifikátorov), preto tento problém patrí do triedy NP.

Deskriptívna zložitosť

- FOL: trieda zložitosti AC^0 (vlastná podmnožina LOGSPACE)
- FOL + tranzitívny uzáver: non-deterministic LOGSPACE
- FOL + least fixed point operator: P
  (súvisí s databázami: dotazy sú podmnožinou prvorádových formúl a navyše je povolená rekurzia počítaná seminaivnou evaluáciou, čiže ako least fixed point)

- existential SOL: NP
- universal SOL: co-NP
- SOL: PH (obsahuje NP aj coNP, ale vlastná podmnožina PSPACE)
- SOL + least fixed point operator: EXPTIME

Nerozhodnuteľnosť a neúplnosť

Neúplnosť

Efektívna axiomatizovateľnosť

Ako formalizovať, že niečo „možno vypočítať“?

Vieme vymýšľať rôzne konkrétne modely na počítanie a porovnávať, čo dokážu. Ukazuje sa, že všeobecné modely počítania s prirodzenými číslami (nie naschvál oslabené dodatočnými reštrikciami) vedú k tomu istému pojmu vypočítateľnosti.

- Turingove stroje
- frázové gramatiky
- Minského registrové automaty
- lambda kalkul
- čiastočné rekurzívne funkcie

Church-Turing Thesis

Ak sa funkcia z $\mathbb{N}$ do $\mathbb{N}$ dá vypočítať,
dá sa vypočítať na Turingovom stroji.

Efektívna axiomatizovateľnosť

Z hľadiska logiky nás zaujímajú formálne systémy, s ktorými možno algoritmicky pracovať:

- konečná množina symbolov
- algoritmy na prácu s termami
- algoritmy na prácu s formulami
- jednoznačné interpretovanie formuly v danej štruktúre
- rozhodnuteľnosť, či niečo je axióma
  (resp. enumerácia axióm/formúl v teórii)
- rozhodnuteľnosť, či niečo je dôkaz
  (resp. enumerácia dôkazov)

Teórii s týmito vlastnosťami (resp. formálnemu systému, ktorého je súčasťou) hovoríme **efektívne axiomatizovateľná**
(v angličtine *effectively*, nie efficiently).

Sémantická a negačná úplnosť

Doterajší pojem úplnosti, tzv. **sémantická úplnosť** formálneho systému, hovoril o tom, že pravdivé veci možno dokázať:
Pre každú teóriu T a uzavretú formulu F, ak $T \vDash F$, tak $T \vdash F$.

Zaujímavá je však aj **negačná úplnosť** teórie T: pre každú uzavretú formulu F (v jazyku tejto teórie) platí $T \vdash F$ alebo $T \vdash \neg F$.

Sémantická a negačná úplnosť

Uvažujme výrokovú logiku s atómami A, B, C a teóriu $T = \{A \wedge B\}$.
Sémantická úplnosť vyplýva z vlastností výrokovej logiky.

Avšak napr. formula C je nezávislá od teórie T, neplatí $T \models C$ ani $T \models \neg C$. Preto $T \not\vdash C$ a $T \not\vdash \neg C$, čiže T nie je negačne úplná.

Sémantická a negačná úplnosť

Je ľahké vytvárať teórie, ktoré sú negačne neúplné:
stačí vynechať zopár podstatných axióm/formúl.

Ak sa však snažíme nájsť teóriu T, ktorá čo najúplnejšie popisuje nejakú časť sveta vyjadriteľnú v jazyku $\mathcal{L}$, chceli by sme, aby pre každú uzavretú formulu jazyka $\mathcal{L}$ platilo buď $T \models F$ alebo $T \models \neg F$. Preto by sme chceli vedieť efektívne rozhodnúť, či $T \vdash F$ alebo $T \vdash \neg F$. Hľadáme teda negačne úplné teórie.

Prvá Gödelova veta o neúplnosti

Veta 17.8

Nech T je prvorádová teória v jazyku $\mathcal{L}$ s týmito vlastnosťami:

- *je konzistentná (nevyplýva z nej nepravda);*
- *je efektívne axiomatizovateľná;*
- *obsahuje aritmetiku.*

Potom T nie je negačne úplná
(t. j. existuje tvrdenie v $\mathcal{L}$, ktoré nemožno z T dokázať ani vyvrátiť).

Ak do T pridáme to nerozhodnuteľné tvrdenie, znova bude podľa tejto vety existovať ďalšie, a problému sa tak nedokážeme zbaviť (jedine pridaním tak veľa vecí, že porušíme predpoklady).

Toto prekvapivé tvrdenie z r. 1931 znamenalo krach Hilbertovho programu z r. 1900 (snahy o kompletnú axiomatizáciu matematiky v štýle Euklidových základov).

Prvá Gödelova veta o neúplnosti

Idea dôkazu — Gödelova veta G: „toto je veta, ktorá sa nedá dokázať“.

Aritmetiku potrebujeme na to, aby sme vedeli jednoznačne očíslovať všetky možné formuly a tiež všetky možné dôkazy (efektívna axiomatizovateľnosť zaručuje, že formuly aj dôkazy sa dajú enumerovať, je ich teda spočítateľne veľa). Potom vieme kľúčovú vetu G zapísať ako prvorádovú formulu.

Ak by sa G dala dokázať, je nepravdivá, čo je spor s konzistentnosťou. Ak sa G dokázať nedá, je pravdivá. Ale potom $\neg G$ je nepravdivá, a teda sa nemôže dať v konzistentnom systéme dokázať. Teória T teda nie je negačne úplná.

Druhá Gödelova veta o neúplnosti

Veta 17.9

Ak S je konzistentný efektívne axiomatizovateľný formálny systém obsahujúci aritmetiku (napr. PA), jeho konzistentnosť nemožno dokázať v rámci S.

Čiže dôkaz treba spraviť neformálnou metaúvahou alebo v rámci iného „silnejšieho“ formálneho systému. Ako však dokážeme konzistentnosť tohto druhého systému? …