

Analýza obtížnosti logických úloh na základě modelů lidského chování

Petr Jarušek, Radek Pelánek

Fakulta informatiky, Masarykova univerzita
Botanická 68a, 602 00, Brno
xjarusek@fi.muni.cz, pelanek@fi.muni.cz

Abstrakt

Naše práce se zaměřuje na studium obtížnosti logických úloh Sokoban a Sudoku pro člověka. I zadání s velice podobnými charakteristikami vykazují při experimentech až desetinásobné rozdíly v čase potřebném k vyřešení úlohy. Proč jsou rozdíly ve zdánlivě podobných úlohách tak markantní? Jak tento rozdíl předpovědět? V této práci představujeme dva výpočetní modely napodobující lidské chování při řešení úloh a ukazujeme, že je lze využít pro předpovídání obtížnosti studovaných úloh.

1 Úvod

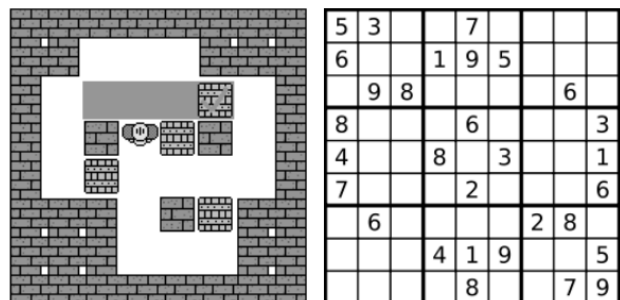
Obtížnost logických úloh byla doposud studována především na základě vlastností a charakteristik konkrétních zadání. Náš přístup se naproti tomu zaměřuje na vývoj modelů, které napodobují lidské chování při řešení úloh. Z chování modelu následně usuzujeme, zda bude úloha pro člověka jednoduchá nebo obtížná. Tento přístup umožňuje zohlednit strukturu jednotlivých problémů a lépe odhadovat jejich obtížnost.

Výzkumníci zabývající se otázkou, jak lidé řeší problémy [4,7], doposud studovali logické úlohy Hanojské věže [2], Chinese ring puzzle [1] a Patnáctku [3]. Výzkum Hanojských věží ukazuje, jaký vliv má na obtížnost identických problémů jejich reprezentace. V našem případě jsou však zadání reprezentována stejným způsobem. Druhý přístup odvozuje obtížnost od

velikosti zadání, případně počtu tahů optimálního řešení. I tento přístup se ukázal v našem případě jako nedostatečný.

2 Motivace a výběr úloh

Pro náš výzkum jsme vybrali dvě logické úlohy – Sokoban a Sudoku. Výsledky experimentů s lidmi ukázaly, že rozdíl v průměrném čase na vyřešení velice podobných zadání úlohy Sokoban kolísá až 50-ti násobně; pro úlohu Sudoku potom až 10-ti násobně. Jak vysvětlit tuto rozdílnost?



Obrázek 1. Ukázka úloh Sokoban a Sudoku

Sokoban je logická hra, jejímž cílem je dopravit pomocí skladníka všechny bedny v místnosti na cílové pozice. Skladník smí tlačit v jeden okamžik pouze jednu bednu. Jde o zástupce úloh typu „procházení stavového prostoru“. Každé rozložení beden v místnosti a pozice skladníka určují jeden stav. Posunutím bedny se pak posouváme do dalšího stavu. V praxi se podobné úlohy řeší například v oblasti verifikace protokolů.

Sokoban je navíc obtížná úloha pro samotné počítače [6], které nejsou schopny řešit její rozsáhlejší zadání. Lidé naproti tomu dokáží řešit jednoduchá i složitá zadání úlohy, avšak pro některé zadání hledají řešení neúměrně dlouho. Jak efektivně zkombinovat silné stránky lidí i počítačů? Průzkum této otázky může odstartovat zajímavé experimenty na poli spolupráce člověka s počítačem.

Cílem úlohy Sudoku je doplnit čísla do tabulky 9x9 políček při dodržení podmínky vyplnění každého z čísel 1-9 do každého řádku i sloupce tabulky. Úloha je pro počítače triviální, avšak pro lidi obtížnost jednotlivých zadání značně kolísá. Jde o úlohu typu „problém splnění podmínek“, praktické aplikace tohoto typu úloh nalezneme v rozvrhování, kryptografii, interpretaci obrazu, řízení letového provozu.

3 Internetový sběr dat

Specifikem našeho výzkumu je sběr dat pomocí internetu. Většina výzkumů zkoumajících řešení problémů lidmi provádí experimenty pod pečlivým dohledem badatelů. Tím je sice zaručena vysoká kvalita dat, nicméně to znamená, že je reálné pracovat jen s několika málo úlohami a maximálně s desítkami řešitelů. Pro náš výzkum jsme použili odlišný přístup a data sbírali pomocí webových portálů umožňujících hraní logických her online. Nevýhodou sběru dat z internetu je absence přímé kontroly nad respondenty. Sebraná data jsou však řádově rozsáhlejší a tudíž robustnější. Tímto způsobem je navíc možné experiment uskutečnit prakticky zadarmo.

Pro úlohu Sokoban jsme vytvořili vlastní webový portál, na kterém mohli lidé řešit 35 zadání úlohy Sokoban. Portál zaznamenával všechny tahy hráčů a zapisoval je do databáze. Experimentu se zúčastnilo 291 řešitelů, kteří při řešení úloh strávili 785 hodin a úspěšně vyřešili 2071 zadání.

Pro úlohu Sudoku jsme využili existující webové portály. Data byla převzata ze tří webových serverů. Dva z nich poskytly cca 1000 úloh, přičemž každou úlohu řešilo alespoň 100, resp. 1000 lidí. K dispozici máme v těchto případech pouze celkový čas potřebný na vyřešení úlohy. Z třetího serveru jsme získali 20 her včetně záznamů jednotlivých her. Níže uvádíme výsledky pro data z portálu sudoku.org.uk, pro data z ostatních portálů jsou výsledky podobné.

Jako metriku pro posouzení obtížnosti úloh jsme použili medián času na vyřešení, který je odolnější vůči extrémním hodnotám než průměrný čas. Každopádně medián i průměr dávají velice blízké výsledky, navíc medián času také dobře koreluje s počtem kroků a počtem hráčů, kteří úlohu úspěšně dokončili, a je tudíž dostatečně robustní metodou k odhadu obtížnosti.

4 Abstraktní výpočetní model

Náš přístup k odhadu obtížnosti je založen na využití abstraktních výpočetních modelů, které simulují chování lidí při řešení daného problému. Modely nesimulují, jak lidé o úloze přemýšlí, ale pouze se snaží kopírovat co nejvěrněji postup řešení. Modely, které popisujeme, jsou velmi abstraktní a mají minimum parametrů a pravidel specifických pro konkrétní problém. Díky tomu jsou potenciálně přenositelné na podobné typy problémů.

Intuitivní metriky pro posouzení obtížnosti úloh jsou postaveny na konkrétních vlastnostech jednotlivých zadání (počet beden, počet předem vyplněných políček), případně na analýze optimální cesty ze startu do cíle. Pracují tedy jen s určitou informací o úlohách. Abstraktní modely naproti tomu berou svým chováním v potaz celkovou strukturu problému. Při řešení se rozhodují neoptimálně, prozkoumávají stavový prostor (Sokoban), případně zohledňují,

kolik různých kroků je možné v daném stavu učinit a jak jsou tyto kroky náročné (Sudoku).

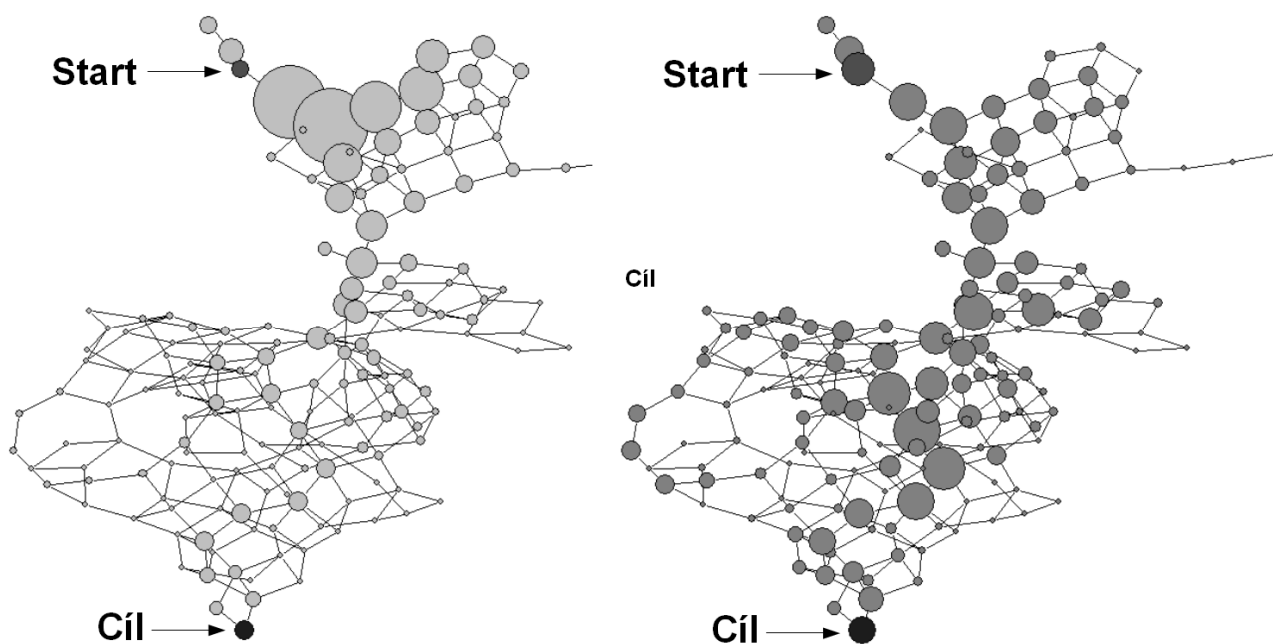
Obtížnost úlohy pak z chování modelu, podobně jako u lidí, odvozujeme od počtu kroků k vyřešení (celkový čas), případně od počtu kroků a jejich náročnosti (analogie času stráveného v daném stavu).

5 Model řešení úlohy Sokoban

Na řešení úlohy Sokoban lze nahlížet jako na průchod stavovým prostorem. Stav je reprezentován pozicí beden v místnosti a pozicí skladníka. Pohyb ve stavovém prostoru je možný posunutím bedny o jedno políčko. Ve stavovém prostoru je jasně vyznačený startovní a cílový stav. Stavy, ze kterých je možné úlohu vyřešit (tj. úspěšně dojít do cílového stavu) označíme jako živé. Stavy, ze kterých již úlohu není možné vyřešit označíme jako mrtvé.

Pro analýzu získaných dat jsme pohyb řešitelů ve stavovém prostoru vizualizovali pomocí metod pro automatizované vykreslování grafů, viz obrázek 2., který ilustruje pohyb řešitelů, a modelu nad stavových prostorem úlohy z obrázku 1. Na základě vizualizací jsme formulovali pozorování, která jsme následně kvantifikovali.

Z analýzy například vyplynulo, že lidé tráví většinu času (85%) v živých stavech, tedy ve stavech, ze kterých je možné dosáhnout cílového stavu. Navíc při pohybu v blízkosti cílového stavu volí častěji optimální cestu k cíli, zatímco v blízkosti startovního stavu mají větší sklon pohybovat se po neoptimální cestě. Obě pozorování jsme zohlednili při návrhu výpočetního modelu.



Obrázek 2. Chování lidí a modelu ve stavovém prostoru zadání úlohy Sokoban z obrázku 1. Každý kruh odpovídá jednomu stavu hry, velikost kruhu určuje počet návštěv daného stavu. Levý obrázek zobrazuje pohyb lidí, pravý pohyb modelu. Jsou zobrazeny pouze živé stavy (stavy, ze kterých lze dosáhnout cíle).

Model, na rozdíl od člověka, pracuje pouze se stavovým prostorem hry. Protože lidé tráví v mrtvých stavech relativně málo času, nechali jsme model pracovat pouze nad prostorem živých stavů. Na počátku je model umístěn do startovního stavu a následně se pomocí jednoduchých pravidel pohybuje po stavovém prostoru. Jakmile dosáhne cílového stavu, je umístěn opět na start.

V každém kroku zvažuje model následníky aktuálního stavu s . Každému následníkovi přiřadí model skóre a zvýší hodnotu rozdaného skóre SumScore. Následující stav s' vybere model dle pravděpodobnostní distribuce:

$$P(s') = \text{score}(s') / \text{SumScore}$$

Pro výpočet skóre jsme použili jednoduchou funkci, která rozlišuje dva typy následníků – přibližující se k cíli a vzdalující se od cíle.

Nechť pro aktuální stav s platí:

s' – následník s

$d(s)$ – vzdálenost s od cíle

B – nastavitelný parametr modelu (bonus)

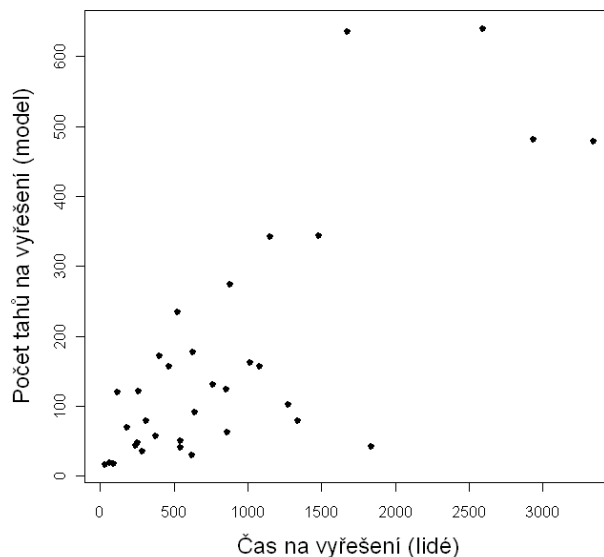
Potom skóre následníka s' odpovídá:

$$\text{score}(s') = \begin{cases} 0 & d(s') = \infty \\ d(s) & d(s') \neq \infty, d(s') \geq d(s) \\ d(s) + B & d(s') < d(s) \end{cases}$$

První případ ukazuje, že model navštěvuje pouze živé stavy. Třetí případ ilustruje, že stavy vedoucí blíže k cíli, obdrží také nastavitelný bonus B a mají tedy vyšší šanci, že budou vybrány. Díky použití vzdálenosti od cíle ve vzorci poroste význam bonusu B s tím, jak se model přiblíží blíže k cíli. Tedy čím blíže bude model k cíli, tím méně náhodně se bude chovat.

Pro $B = 0$ odpovídá model náhodné procházce po stavovém prostoru. S rostoucí hodnotou bonusu B se chování modelu přibližuje k optimální cestě ze startu do cíle.

Pro kalibrování modelu (hledání ideální hodnoty vstupního parametru K) jsme prováděli srovnání stavů navštívených modelem a lidmi (obrázek 2. ilustruje toto srovnání vizuálně pro konkrétní model). Zjistili, že model se chová nejlépe pro hodnotu parametru $B = 0,25$. Pokud použijeme tuto verzi modelu pro predikci obtížnosti, dosáhneme docela dobré korelace ($r = 0,77$, viz též obrázek 3.). Pro srovnání: intuitivní metrika založená na délce optimální cesty od startu do cíle dosahuje výrazně nižší korelace ($r = 0,30$).



Obrázek 3. Korelace průměrného počtu tahů k vyřešení úloh modelem a mediánu času lidí ($r = 0,77$). Parametr modelu B byl nastaven na hodnotu $0,25$.

6 Model řešení úlohy Sudoku

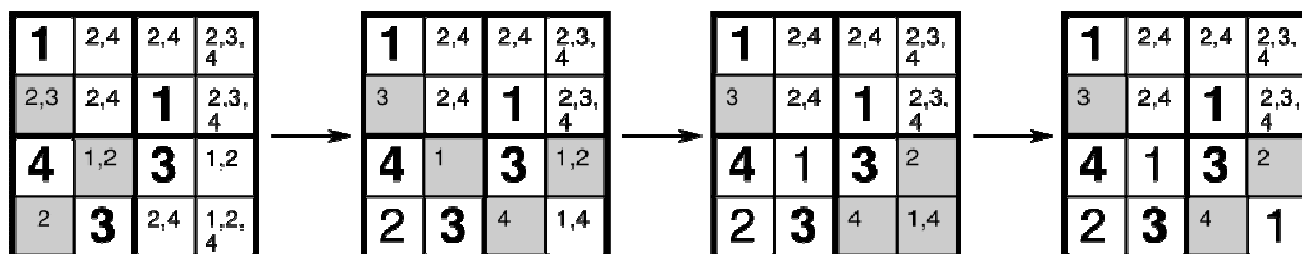
V případě Sudoku se jako přirozená metrika obtížnosti jeví počet vyplněných polí v zadání. Tato intuitivní metrika však funguje špatně ($r = 0,20$) a pro odhad obtížnosti je opět nutno použít výpočetní model simulující chování člověka.

Vzhledem k popularitě Sudoku existuje řada modelů a metrik pro odhad obtížnosti. Základ vždy spočívá v simulování „logických technik“ používaných člověkem. Model vždy určí, která čísla jde do mřížky „logicky“ doplnit, jedno z nich náhodně vybere a pokračuje dál. Běžně používané modely v sobě obsahují řadu technik specializovaných pro Sudoku a mají mnoho parametrů, které určují relativní obtížnost dílčích technik. Tyto parametry jsou většinou nastaveny autorem modelu podle jeho pocitu, nikoli na základě exaktního vyhodnocení.

My jsme pro Sudoku opět navrhli jednoduchý abstraktní model, který používá minimum specifických technik a je tak potenciálně přenositelný i na jiné problémy. Model používá pouze dvě základní techniky pro odvození hodnoty políčka (v anglické literatuře standardně označované jako „hidden single“ a „naked single“).

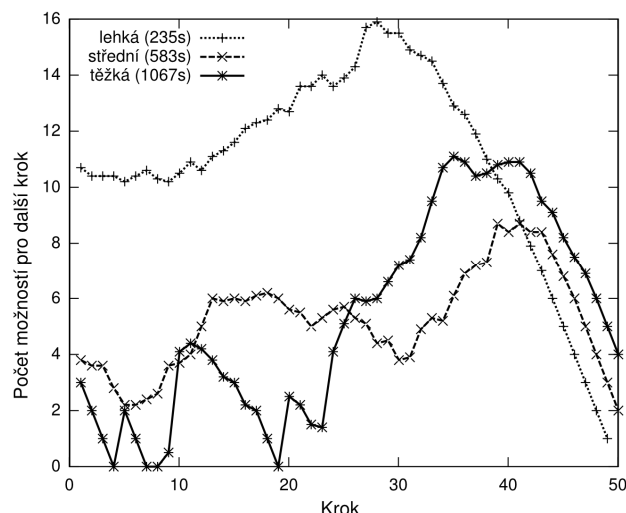
Příklad běhu modelu je na obrázku 4. Pokud základní techniky nestačí pro doplnění žádného čísla, model systematicky hledá políčko, kde lze co nejjednodušeji vyvrátit všechny špatné kandidáty. Tento postup aproximuje složitější logické techniky a přitom není nijak specifický pro Sudoku a neobsahuje žádné ad hoc parametry.

Celková obtížnost úlohy se běžně stanovuje pouze na základě obtížnosti dílčích kroků (dílčích logických technik). Můžeme se však



Obrázek 4. Běh modelu na Sudoku velikosti 4x4. Tučně psaná čísla tvoří zadání, malá čísla jsou seznamy kandidátů pro daná políčka, šedě označená políčka jde v dané situaci odvodit pomocí jednoduchých technik. V každém kroku model doplní číslo do náhodně vybraného šedého políčka.

pokusit zohlednit i „celkovou strukturu problému“. Obrázek 5. zachycuje průběh řešení tří úloh rozdílné obtížnosti. Osa y znázorňuje „počet šedých polí“ ve významu z obrázku 4., tj. počet polí, na která je možné doplnit další číslo pomocí jednoduchých technik.



Obrázek 5. Průběh řešení pro 3 zadání různé obtížnosti (průměrné hodnoty z 30 běhů modelu).

Vidíme, že obtížnost úlohy souvisí nejen s náročností dílčích technik, ale také s tím, jak je jednoduché najít tu část zadání, kde lze tyto techniky použít. Obrázek 5. lze jednoduše převést do numerické metriky tím, že počítáme průměrnou hodnotu zobrazenou na ose y (přesně řečeno na základě analýz jsme určili, že ideální je počítat průměrnou hodnotu z prvních 25 kroků).

S využitím tohoto jednoduchého modelu dosahujeme srovnatelných výsledků jako běžně používané modely s mnoha parametry ($r \sim 0,85$). Pomocí lineární kombinace několika modelů pak lze dokonce dosáhnout velmi přesných odhadů obtížnosti ($r = 0,95$).

7 Výhledy

V této práci jsme představili dva modely simulující chování lidí při řešení úloh Sokoban a Sudoku naznačili jejich využití pro odhadování obtížnosti dílčích zadání. Přestože se modely řídí pouze jednoduchými pravidly, prokazují dobrou korelaci s chováním lidí. V případě Sokobanu odhadl model obtížnost úlohy s korelací 0,67 a v případě Sudoku s korelací 0,85. Jak lze tyto výsledky aplikovat v praxi? Pro ilustraci uvádíme několik možných aplikací výpočetních modelů.

Hromadné ohodnocení obtížnosti úloh – pomocí modelů seřadíme velké množství úloh podle obtížnosti bez nutnosti řešení jednotlivých úloh lidmi.

Odhalení podvádění – pomocí modelu je možné například detekovat nestandardní chování při řešení Sudoku, čímž můžeme odhalit podvádění při internetových soutěžích.

Umělá inteligence – napodobení chování lidí při řešení problémů je užitečné například při simulaci soupeře v počítačových hrách.

V budoucnosti bychom rádi ověřili přenositelnost výpočetních modelů na další logické úlohy. Poznatky o obtížnosti se rovněž chystáme aplikovat při návrhu výukového systému postaveném na doporučování úloh vhodné obtížnosti.

Literatura

- [1] R. Pelánek and P. Jarušek: Human Problem Solving: Sokoban Case Study. Technical report, Masarykova Univerzita, Fakulta informatiky, 2010.
- [2] K. Kotovsky and H.A. Simon: What Makes Some Problems Really Hard: Explorations in the Problem Space of Difficulty. *Cognitive Psychology* (1990) 143 – 83.
- [3] Kotovsky, K., Hayes, J., Simon, H.: Why are some problems hard? Evidence from tower of Hanoi. *Cognitive psychology* 17 (1985) 248–294.
- [4] Z. Pizlo and Z. Li: Solving combinatorial problems: The 15-puzzle. *Memory and Cognition* (2005).
- [5] H.A. Simon and A. Newell: *Human problem solving*, Prentice Hall, 1972.
- [6] J.G. Greeno: Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology* (1974) 270–292.
- [7] A. Junghanns: Pushing the limits: New developments in single-agent search. *PhD thesis*, University of Alberta, Department of Computing Science, 1999.
- [8] A. Newell and H.A. Simon: *GPS, a program that simulates human thought*. *Computers and thought* (1963) 279–293.