VYHODNOTENIE RELIABILITY HODNOTENIA OLYMPIÁDY V INFORMATIKE

RNDr. Michal Forišek
Katedra informatiky, Fakulta matematiky, fyziky a informatiky Univerzity Komenského, Mlynská Dolina, 842 48 Bratislava
e-mail: forisek@dcs.fmph.uniba.sk

Pomocou Item Response Theory navrhujeme teoretický model na vyhodnotenie reliability výsledkov Olympiády v informatike. Navrhnutý model používame na konkrétnych dátach pre krajské kolá 22. a 23. ročníka Olympiády v informatike.

Kľúčové slová: *item response theory, olympiáda v informatike, obtiažnosť úloh, reliabilita hodnotenia*

1. ÚVOD

Cieľom výskumu prezentovaného v tomto článku bolo analyzovať reliabilitu výsledkov žiakov v Olympiáde v informatike a formulovať závery, ktoré povedú k jej zlepšeniu.

V kapitolách 2, 3 a 4 prezentujeme nami použitý teoretický model, následne v kapitole 5 uvádzame analýzu konkrétnych dát pre školské roky 2006/07 a 2007/08.

2. MODEL OBTIAŽNOSTI ÚLOHY

Našu analýzu postavíme na modernej teórii testovania – Item Response Theory (IRT). V tejto kapitole uvedieme stručný prehľad použitých výsledkov, ako aj vlastné úpravy existujúcich modelov. Podrobnejší popis tejto teórie nájdete napr. v [1] a [5].

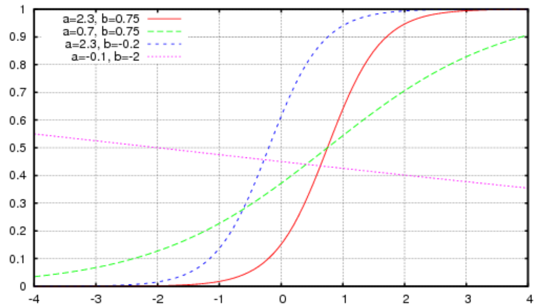
(Lingvistická poznámka: Stretli sme sa s viacerými rôzne zlými slovenskými prekladmi, asi najlepší je „teória odpovedí na položku“.)

Základným predpokladom, z ktorého budeme vychádzať, je, že každý z subjektov, v našom prípade teda súťažiacich žiakov, má určitú úroveň schopnosti riešiť algoritmické úlohy. Budeme predpokladať, že úroveň tejto schopnosti subjektu s sa dá popísať reálnym číslom, ktoré budeme označovať θ_s. Keďže ide o latentnú schopnosť, hodnoty θ_s nevieme merať priamo, ale vieme získavať ich odhady prostredníctvom testov.

Náš model obtiažnosti úlohy postavíme na 2-parametrovom logistickom (2PL) modeli. Klasický 2PL model je navrhovaný pre úlohy s binárnym (v teórii testovania sa niekedy používa pojem „dichotomickým“) hodnotením – subjekt buď úlohu vyriešil alebo nie. V klasickom 2PL modeli popisujeme obtiažnosť úlohy dvoma parametrami: koeficient obtiažnosti b, a koeficient diskriminácie a. Pre úlohu s parametrami (a,b) je pravdepodobnosť, že ju subjekt so schopnosťou θ vyrieši, daná logistickou funkciou:

$$Pr(\theta, a, b) = \frac{1}{1+e^{-a(\theta-b)}}$$

Príklady priebehu tejto funkcie pre rôzne dvojice (a,b) sú znázornené na obr. 1.



Obr. 1 Logistická funkcia pre 2PL model.

V našom prípade tento model ale nemôžeme priamo použiť, keďže Olympiáda v informatike používa iné ako binárne hodnotenie. Model, ktorý použijeme, by mal čo najpresnejšie odzrkadľovať použitý spôsob hodnotenia.

Úlohy v Olympiáde v informatike sú hodnotené na diskrétnej škále od 0 do 10 bodov, primárnym kritériom hodnotenia je efektívnosť nájdeného algoritmu. Pre veľkú väčšinu zadaných úloh (vrátane tých skúmaných v nasledujúcej kapitole) platia nasledovné hranice:

- každé korektné riešenie získa aspoň 4 body
- každé efektívne riešenie aspoň 7 bodov
- riešenia podobné vzorovému aspoň 9 bodov

Riešenie úlohy väčšinou vyžaduje viacero myšlienkových krokov, ktoré vedú k postupnému zlepšeniu nájdeného riešenia. Na základe tohto pozorovania sme sa rozhodli modelovať takúto úlohu ako tri samostatné binárne hodnotené úlohy. Následne vieme očakávaný počet bodov (na škále 0-3, nie 0-10) vyjadriť ako súčet pravdepodobností vyriešenia jednotlivých podúloh:

$$E(Body(\theta)) = \sum_{i=1}^{3} \frac{1}{1+e^{-a_i(\theta-b_i)}}$$

3. FISHEROVA INFORMÁCIA

Vždy, keď pozorujeme hodnotu náhodnej premennej, získavame tým informáciu. Pri testovaní môžeme túto informáciu použiť na získanie lepšieho odhadu skrytých parametrov systému. Intuitívne môžeme mieru informácie definovať ako prevrátenú hodnotu presnosti, s akou vieme daný parameter odhadnúť. Formálne túto intuíciu zachytáva Fisherova definícia štatistickej informácie ako očakávanej hodnoty variancie skóre. Známym výsledkom je, že pre 2PL model je:

$$I(\theta)=a^2 Pr(\theta,a,b)(1-Pr(\theta,a,b))$$

Znamená to teda, že zaradenie úlohy s parametrami (a,b) do testu nám poskytne najviac informácie pre subjekty, pre ktoré je $Pr(\theta,a,b)=1/2$, a teda $\theta=b$.

Keďže predpokladáme, že výsledky subjektu pre rôzne položky sú navzájom nezávislé náhodné premenné, platí, že informácia získaná z rôznych položiek testu je aditívna.

Zamerajme sa teraz na krajské kolo Olympiády v informatike. Z tohto kola najlepších približne 30 riešiteľov z celého Slovenska postupuje na celoštátne kolo. Reliabilita krajského kola je teda schopnosť čo najpresnejšie identifikovať 30 súťažiacich, ktorí sú v riešení algoritmických úloh najlepší. Znamená to teda, že našim cieľom pri výbere súťažných úloh by malo byť primárne vytvoriť takú sadu úloh, ktorá nám umožní čo najpresnejšie určiť schopnosti subjektov v okolí postupovej hranice, t. j. 30-teho miesta.

4. ODHAD PARAMETROV

Pochopiteľne, aj keby náš teoretický model presne popisoval realitu (čo samozrejme nie je pravda), presné parametre popisujúce obtiažnosť úloh nám nik nedá. Najbežnejším spôsobom, ako približné hodnoty týchto parametrov získať pred samotným použitím testu je kalibrácia – dáme úlohy testovacej vzorke subjektov, ktorých úrovne schopnosti považujeme za známe, a na základe ich výsledkov vypočítame najpravdepodobnejšie hodnoty parametrov úloh. Tento postup však v našom článku nepotrebujeme použiť, preto sa ním nebudeme podrobnejšie zaoberať.

Pri zodpovedaní našej otázky si vystačíme s posteriórnym určením parametrov úloh. Použijeme metódu maximum likelihood estimate.

(Keďže aj „likelihood“ aj „probability“ sa do slovenčiny prekladajú ako „pravdepodobnosť“, kvôli potrebe odlíšiť tieto dva pojmy budeme používať anglický pojem „likelihood“.)

Základná myšlienka tohto prístupu je nasledovná: Majme náhodnú premennú X, ktorej rozloženie pravdepodobnosti fy závisí od parametra y. Namerali sme premennú X a dostali sme výsledok x. V tejto situácii môžeme likelihood $L(y)$ parametra y definovať ako podmienenú pravdepodobnosť $P(X=x \mid y)$. Zdôrazňujeme, že hodnota $L(y)$ nie je pravdepodobnosť, že y je skutočná hodnota skrytého parametra. Hodnota y, pre ktorú funkcia $L(y)$ nadobúda maximum, sa nazýva maximum likelihood estimate premennej y.

V našom prípade budeme chcieť zostrojiť maximum likelihood estimate pre vektory parametrov θ_i, a_i, a b_i za predpokladu, ktorým sú výsledky subjektov pri riešení dotyčných úloh.

5. ANALÝZA DÁT Z PRAXE

Analyzovali sme dáta zodpovedajúce dvom školským rokom. V každom školskom roku sme mali k dispozícii 44 úloh – 4 série Korešpondenčného seminára z programovania (KSP) po 10 úloh a 4 úlohy krajského kola Olympiády v informatike (OI), kategórie A. Počet subjektov, ktoré sa do súťaží zapojili, bol 177 v šk. r. 06/07, resp. 201 v šk. r. 07/08.

Na modelovanie obtiažnosti úloh sme použili model popísaný v kapitole 2, teda každú úlohu modelujeme ako tri nezávislé binárne hodnotené položky. Vzorky odpovedí (response patterns) pre tieto položky sme zostrojili tak, že za správne vyriešenú sme položku označili u tých riešiteľov, ktorí za ňu získali dostatočne veľa bodov.

Napr. pre úlohu z OI prvú položku zodpovedajúcu úlohe správne vyriešili riešitelia, ktorí získali aspoň 3, druhú tí, ktorí aspoň 6, a aj tretiu tí, ktorí získali aspoň 9 bodov z 10. Pre úlohy z KSP, ktoré používalo škálu od 0 do 15 bodov, sme použili hranice 4, 8, resp. 13 bodov.

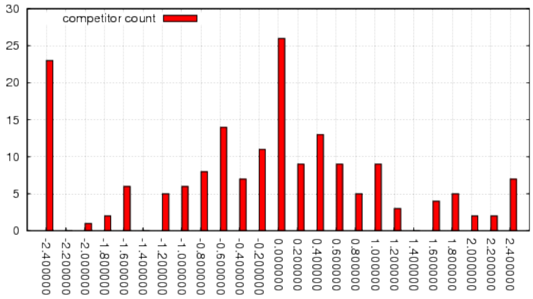
Na základe týchto vzoriek odpovedí sme zostrojili maximum likelihood estimates pre schopnosť θ_i každého z účastníkov a pre parametre a_i a b_i každej z položiek. Použili sme na to metódu hill climbing numerickej optimalizácie, náš program začal s základným odhadom že všetky θ_i sú rovné 0 a všetky $(a_i,b_i)=(1,0)$. Následne až do okamihu dostatočnej konvergencie striedavo dopočítal maximum likelihood estimate pre schopnosti a pre parametre položiek na základe predchádzajúceho odhadu pre opačný typ parametrov.

Použitá formula pre likelihood schopnosti θ konkrétneho subjektu na základe parametrov položiek a správnosti jeho odpovedí na ne:

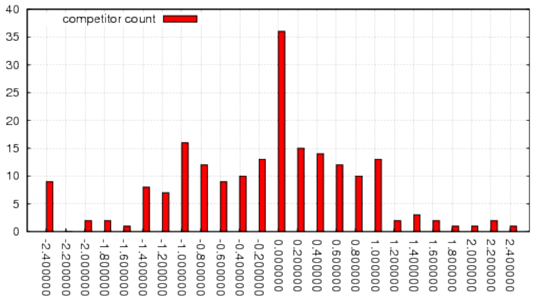
$$L(\theta)=\Pi_{i=1}^{N} p_i^{s_i}(1-p_i)^{1-s_i}$$

kde $p_i=Pr(\theta,a_i,b_i)$ a s_i je 0 ak položku i nevyriešil a 1 ak áno. Maximum likelihood estimate parametra θ sme počítali ako maximum tejto funkcie (presnejšie jej logaritmu, ktorý je aditívny, má maximum v rovnakom bode a výpočet je numericky stabilnejší.) Formula pre opačný krok vyzerá analogicky.

Na obr. 2a, 2b je vyššie uvedeným postupom získané rozloženie schopností riešiteľov pre jednotlivé školské roky. Všimnite si, že rozdelenie podľa očakávania nezodpovedá Gaussovej krivke: na ľavej strane je vidieť, že existuje skupina riešiteľov, ktorí sa do súťaže zapoja, no úlohy sú nad ich sily. Hranica 30. miesta v celej vzorke subjektov zodpovedala v šk. r. 06/07 približne hodnote $\theta=0{,}95$ a v šk. r. 07/08 hodnote $\theta=0{,}8$. Hranica 30. miesta v množine účastníkov krajského kola bola v oboch prípadoch mierne nižšia.

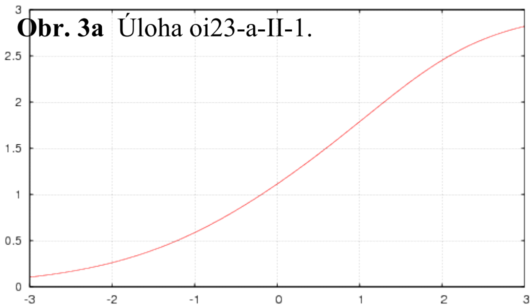


Obr. 2a Rozloženie schopností v šk. r. 06/07.

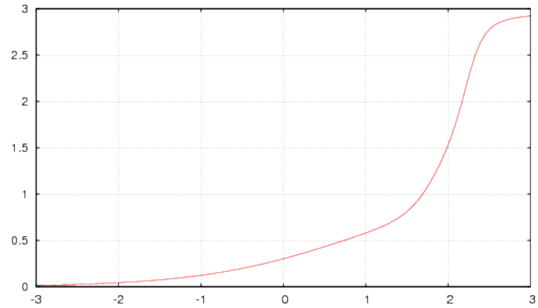


Obr. 2b Rozloženie schopností v šk. r. 07/08.

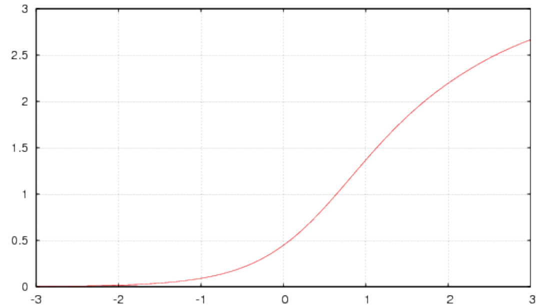
Na obr. 3a, 3b, 3c sú grafy znázorňujúce očakávaný počet bodov (na škále 0-3, nie 0-10) pre tri úlohy krajského kola OI v šk. r. 07/08. Vidíme, že priebehy týchto funkcií sú výrazne rozličné. V prvom prípade ide o relatívne ľahkú, vyváženú úlohu, v ktorej očakávaný počet bodov približne lineárne zodpovedá schopnosti riešiteľa. V druhom prípade ide o ťažkú úlohu, kde je prudko netriviálne nájsť ľubovoľné efektívne riešenie, v dôsledku čoho je očakávaný počet bodov vysoký len na samom konci spektra. V treťom prípade ide o teoretickú úlohu, ktorá vyžaduje určitú úroveň už na pochopenie zadania, preto očakávaný počet bodov začína byť nenulový až v okolí $\theta=0$.



Obr. 3a Úloha oi23-a-II-1.

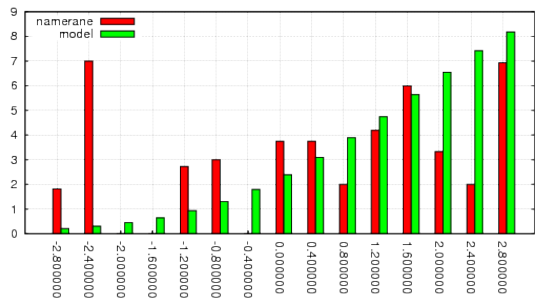


Obr. 3b Úloha oi23-a-II-3.

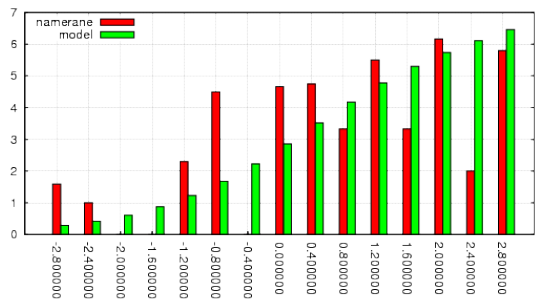


Obr. 3c Úloha oi23-a-II-4.

Na obrázkoch 4a, 4b je porovnanie skutočného priemerného počtu bodov a očakávaného počtu bodov pre dve vybrané úlohy. Riešiteľov sme rozdelili do skupín podľa ich odhadnutých schopností. Výrazné odchýlky reality od hodnôt predpovedaných modelom (pre $\theta=-2{,}4$ a $\theta=2{,}4$ na obr. 4a, resp. $\theta=2{,}4$ na obr. 4b) všetky zodpovedajú skupinám, do ktorých padol len jeden riešiteľ. Skupiny, ktoré v grafe nemajú vôbec znázornený priemerný počet bodov, sú prázdne.



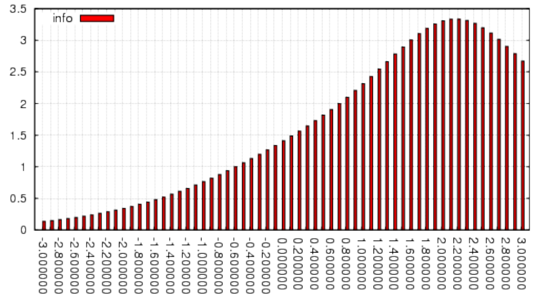
Obr. 4a Úloha oi22-a-II-1.



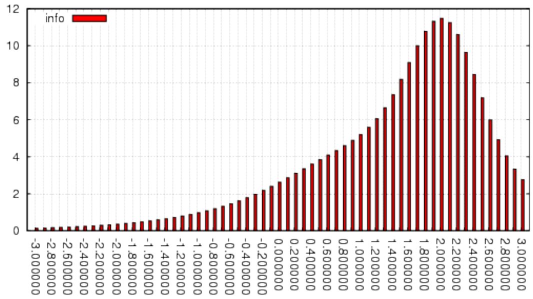
Obr. 4b Úloha oi22-a-II-3.

Hlavným záverom, ktorý nás pri našom výskume zaujíma, je funkcia udávajúca množstvo získanej informácie v závislosti od schopnosti riešiteľa pre celé krajské kolo dokopy (4 úlohy, t. j. 12 binárnych položiek v našom modeli.)

Priebeh tejto funkcie pre krajské kolá v šk. rokoch 06/07 a 07/08 je znázornený na obr. 5a, 5b.



Obr. 5a Krajské kolo 22. ročníka OI, kat. A



Obr. 5b Krajské kolo 23. ročníka OI, kat. A

Vidíme, že v oboch prípadoch je maximum medzi $\theta=2,0$ a $\theta=2,2$. Teda z výsledkov uvedených krajských kôl vieme najpresnejšie odhadnúť schopnosti žiakov, ktorí patria do tohto intervalu. Tu však ide o absolútnu špičku súťažiacich, a to nie je to, čo by sme od krajského kola požadovali. Ako sme už uviedli vyššie, hlavným cieľom krajského kola by malo byť čo najpresnejšie identifikovať množinu najlepších, ktorí by mali postúpiť do celoštátneho kola. Na to by ale bolo omnoho vhodnejšie zvoliť také úlohy, pre ktoré by test information funkcia mala maximum v okolí očakávanej hranice schopností potrebných na postup. (Naopak, skúmané sady úloh by boli výborné pre použitie na celoštátnom kole, kde je hlavným cieľom identifikovať víťaza.)

Ďalej považujeme za potrebné uviesť ešte jedno súvisiace pozorovanie. Hlavným dôvodom, prečo Olympiáda v informatike používa 10-bodovú škálu, je prítomnosť sekundárnych kritérií hodnotenia. Okrem samotnej myšlienky riešenia sa hodnotí aj dôkaz jej správnosti, kvalita popisu riešenia, odhad časovej a pamäťovej zložitosti a prítomnosť drobných chýb. Tieto faktory nemajú na hodnotenie až taký výrazný vplyv ako samotná efektívnosť nájdeného riešenia.

V nami skúmaných krajských kolách však voľba súťažných úloh spôsobila, že medzi bodovým hodnotením efektívnosti algoritmov u súťažiacich, ktorí sa schopnosťou pohybovali okolo hranice postupu, boli len malé rozdiely. Dôsledkom toho mali na konečné umiestnenie v tejto časti výsledkovej listiny vplyv práve spomínané sekundárne faktory. Dovoľujeme si teda tvrdiť, že spomedzi súťažiacich, ktorých schopnosti sa pohybovali okolo hranice postupu, postúpili na celoštátne kolá tí, ktorí dokázali svoje riešenia poriadnejšie vypracovať. Rozhodnutie, či v tomto prípade ide o pozitívum alebo negatívum, prenechávame na čitateľa.

6. ZÁVER

Vypracovali sme teoretický model, pomocou ktorého sme úspešne dokázali modelovať zložitosť algoritmickej úlohy. Pomocou uvedeného modelu sme analyzovali dáta pochádzajúce z reálnych súťaží. Dospeli sme k záveru, že skladba úloh, ktoré sa v súčasnosti používajú na krajskom kole Olympiády v informatike, je vhodná na určenie najlepších riešiteľov. Bolo by však lepšie zmeniť náročnosť použitej sady tak, aby sme pomocou nej čo najlepšie dokázali odhadnúť schopnosti tých riešiteľov, ktorí sú v okolí hranice na postup. Odporúčame teda v budúcnosti použiť širšie spektrum úloh s nižším priemerom koeficientov náročnosti.

Ako možnú tému ďalšieho výskumu vidíme otázku súvisu bodového hodnotenia a motivácie riešiteľov. Ide o komplexnú problematiku: uvedieme napr. situáciu, kedy zaradenie ľahkej úlohy síce prospeje motivácii riešiteľov („aspoň niečo som vyriešil"), avšak zvyšuje volatilitu výsledného umiestnenia riešiteľov (chyba pri riešení ľahkej úlohy sa nemusí dať „dohnať" bodmi za ostatné.)

LITERATÚRA

[1] BAKER, F. B. and KIM, S.: Item Response Theory: Parameter Estimation Techniques. CRC 2004. ISBN 978-0824758257.

[2] CORMACK, G.: Random Factors in IOI 2005 Test Case Scoring. Informatics in Education 5/2006. ISSN 0868-4952.

[3] KEMKES, G. et al.: Objective Scoring for Computing Competition Tasks. Information technologies at school, 2006.

[4] GLEASON, J.: An Evaluation of Mathematics Competitions Using Item Response Theory. Notices of the ACM, 2008.

[5] PARTCHEV, I.: A visual guide to item response theory. 2004. Dostupné na http://www.metheval.uni-jena.de/irt/VisualIRT.pdf

[6] Slovenská komisia OI: Organizačný poriadok Olympiády v informatike. 2006. Dostupné na http://www.olympiady.sk/index.php?www=sp_file&id_item=555