

ANALIZA ALGORITMILOR DE TESTARE ADAPTIVĂ

Deinego Nona

În lucrare este prezentat algoritmul testării adaptative asistată de calculator, avantajele și dezavantajele testării adaptative.

В данной работе представлен алгоритм адаптивного тестирования, достоинства и недостатки адаптивного тестирования.

This article is devoted to the algorithm of CAT (computer adaptive testing) and describes its advantages and disadvantages.

Introducere

Testarea adaptativă asistată de calculator (CAT – computer adaptive testing) este una din cele mai puternice "consecințe" a unor aplicații a testării adaptative, începute cu Binet în 1905. Testele adaptative sunt compuse din itemi, selectați dintr-o colecție de itemi, numită bancă de itemi (item bank). Itemii sunt aleși în așa fel, încât să se potrivească nivelului de capacitate a persoanei testate. În caz că acesta reușește să răspundă corect, imediat îi este prezentat un item puțin mai dificil, iar dacă nu răspunde corect atunci i se administrează un item puțin mai ușor. Această tehnică converge repede spre o secvență de itemi ce oscilează monoton în același diapazon mic de dificultate, care formează nivelul efectiv de capacitate a celui examinat. Testul se termină atunci când capacitatea celui examinat este determinată cu precizia necesară.

Avantajele testării adaptative computerizate sunt: teste mai scurte, mai rapide, testare mai flexibilă, o mai bună echilibrare a conținutului testului pentru toate nivelele de capacitate. Dezavantajele: prețurile mari ale echipamentelor, programatura săracă, slaba familiarizare a multor persoane cu calculatorul, o aparentă diferență a testelor pentru diferite persoane.

1. Noțiuni generale

Testele întotdeauna au fost construite în așa mod, încât să se conforme cerințelor persoanelor examinate și performanțelor așteptate în cazul testării unui grup. Este recunoscut faptul că administrarea unui item prea ușor, prea simplu pentru cei examinați este o pierdere de timp, determinând din partea candidaților un comportament nedorit cum ar fi comiterea greșelilor din neatenție, alegerea intenționată a răspunsurilor incorecte care pot fi plauzibile. Pe de altă parte, itemii care-s prea dificili, de obicei, iarăși conduc la rezultate neinformative, departe de adevăr, deoarece candidații, parcă intuind zădărnicia încercărilor, nu denotă o atitudine serioasă față de itemi, încercând să ghicească răspunsul corect.

Adaptarea testului în conformitate cu performanțele individuale ale fiecărei persoane examinate la început a fost privită, desemnată ca ceva problematic și poate chiar nedorit. Apare întrebarea – cum sunt comparați candidații, dacă fiecare răspunde la teste diferite?

Alfred Binet (1905), grație testelor sale de inteligență, a avansat cel mai mult în acest domeniu. În preocupările sale despre diagnosticul candidaților individuali și nu a unui grup, dispărea nevoia ca fiecărui candidat să-i fie administrat același test. El a realizat că poate crea testul pentru un individ, având la îndemână o simplă stratagemă - categorizarea (aranjarea) itemilor în conformitate cu dificultatea lor. El începea a examina candidații pornind de la un subset de itemi, aleși de el, cu dificultatea cea mai potrivită (intuitiv) pentru capacitatea candidatului dat. Dacă candidatul răspundea corect, Binet îi oferea un nou subset de itemi succesivi mai dificili, până când candidatul nu termina prin a greși prea frecvent. În caz că candidatul nu reușea la primul set de itemi, atunci Binet îi administra un set de itemi succesivi mai ușori până când candidatul răspundea corect aproape de fiecare dată. Folosind această informație, Binet estima capacitatea candidatului.

Modalitatea de testare practică de Lord Flexilevel și variantele sale, cum ar fi cea a lui Henning și Lewis prezintă variante rafinate ale metodei lui Binet.

2. Esența testării adaptative computerizate

Vom considera un caz particular. Fie că dispunem de un banc de itemi construit din itemi dihotomici. Fiecare item are un grad de dificultate, care variază în diapazonul 0–100 unități. Fie că itemul "2+2=4", prezent în acest banc, are dificultatea de 5 unități. Atunci candidații pentru care acest item este ușor au capacitatea mai mare de 5 unități. Candidații care au 50 % șanse de a răspunde corect la acest item au capacitatea 5 unități. În acest caz, pentru a caracteriza acest tip de itemi, se folosește termenul de "targeted on", ceea ce înseamnă că acest item este "așintit" spre capacitatea celui examinat.

Alegerea de calculator a primului item nu este un moment critic pentru măsurare, dar poate influența starea psihologică a candidatului. Administrarea unui item prea dificil îi poate determina celui examinat o stare de disperare, și, ca urmare, – rezultate slabe, neadecvate. Administrarea unui item prea ușor îi poate inhiba toată seriozitatea celui examinat, cauzând apariția greșelilor din neatenție.

Fie că capacitatea candidatului este de 50 unități. În exemplul nostru este rațional de a stabili dificultatea primului item administrat de calculator egală cu 30 unități. Dacă candidatul va reuși la acest item, atunci calculatorul va fi nevoit să selecteze un item mai dificil, de 40 unități. Dacă candidatul iarăși reușește, calculatorul alege un item mai dificil – de 50 unități. Atunci probabilitatea că candidatul va reuși la acest item este de 0,5. Fie că el greșește. Atunci calculatorul alege un item mai ușor decât acesta de 50 unități, însă mai dificil decât ultimul reușit, adică de 40 unități. Va fi ales un item de 45 unități dificultate. Candidatul reușește. Calculatorul alege un item de 48 unități. Candidatul iarăși reușește. Acum, observând că succesele candidatului variază între 40 - 48 unități, este posibil că nereușita candidatului la itemul de 50 unități a fost întâmplătoare. Calculatorul alege un item de 52 unități. Candidatul reușește. Următorul item este de 54 unități. Candidatul greșește, imediat i se propune un item de 51 unități dificultate.

Acest process continuă și calculatorul "devine" tot mai "sigur" că capacitatea candidatului este în vecinătatea de 50 unități. Cu cât mai mulți itemi i se vor administra, cu atât mai precis va fi estimată capacitatea. Programele de testare conțin diferite criterii, "condiții de finisare", pentru stoparea administrării testului. Atunci când aceste condiții sunt satisfăcute, testul este oprit și calculatorul prezintă imediat rezultatele celui testat.

3. Teoria psihometrică și algoritmi pe calculator

3.1. Modelul de măsură ales

Un concept esențial, comun tuturor testelor de capacitate sau atitudine, este că capacitățile și atitudinile pot fi comparate în aceeași dimensiune, pe aceeași axă, putând, astfel, afirma că cineva a avut un mai mare succes decât altul la un test oarecare. Această unidimensionalitate simplifică testarea adaptativă, deoarece permite ca itemii să poată fi "mai dificili" și "mai facili", și candidații – "mai puțin capabili", și "mai capabili", indiferent de ce itemi sunt comparați cu ce candidați. Modelul de măsurare necesar și suficient de a construi o variabilă unidimensională este modelul dihotomic a lui Rash.

Modelul dihotomic a lui Rash prezintă o simplă relație dintre candidat și item. Fiecare candidat este caracterizat de un nivel de capacitate, exprimat printr-un număr, aflat pe axa infinită a capacităților. În mod simir, fiecare item este caracterizat printr-un grad de dificultate de asemenea exprimat printr-un număr plasat pe axa infinită a capacităților. La nivel matematic itemii sunt plasați pe axă în punctele unde candidații se așteaptă la 50 % success.

Relația dintre candidați și itemi în modelul dihotomic a lui Rash:

$$\ln(P_{ni1}/P_{ni0}) = B_n - D_i, \quad (1)$$

unde:

- P_{ni1} - probabilitatea că candidatul n va reuși la itemul i ;
- P_{ni0} - probabilitatea că candidatul n nu va reuși la itemul i ;
- B_n - capacitatea candidatului n ;
- D_i - dificultatea itemului i .

Mărimea $\ln(P_{ni1}/P_{ni0})$ reprezintă unitatea de măsură numită *logit*.

3.2. Elaborarea algoritmului

Metodele testării adaptative sunt foarte simple și evidente. Inițial este estimată sau ghicită capacitatea candidatului. Este administrat un item cu o dificultate echivalentă cu această capacitate. În cazul în care candidatul reușește estimarea capacității este mărită și viceversa. Următorul item este administrat în corespundere cu noua capacitate estimată. Procesul se repetă.

3.3. Condiții de finisare

Decizia când trebuie finisat testul este un element crucial în proiectarea testării. Dacă testul e prea scurt, atunci estimarea capacității nu este destul de exactă. Dacă testul e prea lung, atunci se irosesc două elemente importante: timpul și resursele, iar o parte din itemi sunt expuși fără rost. Candidatul, la rândul său, poate obosi și performanța lui poate scădea, determinând rezultate incorecte.

Testarea se oprește când:

- bancul de itemi este epuizat;
- a fost administrat numărul maxim de itemi;
- capacitatea este măsurată cu o precizie suficientă. Eroarea tipică este de 0.2 logit;
- măsura capacității se află destul de "departe" de criteriul de trecere-eșuare. Aceasta poate apărea când capacitatea estimată e la cel puțin două erori standarde de estimare "depărtare" de la criteriu;
- candidatul se manifestă neobișnuit. Răspunde prea repede sau prea încet, folosește aceeași "mască" când răspunde.

Testarea nu poate fi oprită până când:

- nu a fost administrat numărul minim de itemi;
- nu au fost "acoperite" toate domeniile subiectelor în care au fost clasificați itemii;
- au fost administrați suficienți itemi pentru a nu da nimic de bănuț celui testat și a-i păstra dorința de a continua. În caz că cineva va atinge prea repede criteriul de trecere-eșuare, acest fapt, observat fiind de ceilalți, care încă n-au finisat, le poate schimba atitudinea, determinând în unele cazuri rezultate slabe.

3.4. Un algoritm simplu de testare adaptativă

Acest algoritm este propus de Wright (1988). El are o utilizare restrânsă la clasificarea performanțelor, spre exemplu, la sfârșitul fiecărui modul de învățare.

1. Dificultatea itemului $D=0$
2. Numărul de itemi administrați $L=0$
3. Suma dificultăților folosite $H=0$
4. Numărul răspunsurilor corecte $R=0$
5. Determinarea următorului item aproape de dificultatea D
6. Atribuirea actualei variației a acestui item variabilei D
7. Administrarea acestui item
8. Obținerea răspunsului
9. Notarea răspunsului
10. Numărul de itemi administrați $L=L+1$
11. Sumarea dificultăților folosite $H=H+D$
12. Dacă răspunsul este incorect, se modifică dificultatea : $D=D-2/L$
13. Dacă răspunsul este corect, se modifică dificultatea: $D=D+L/2$
14. Dacă răspunsul e corect, atunci numărul răspunsurilor corecte $R=R+1$

15. Dacă nu se poate decide conform criteriului de trecere-eșuare, atunci se trece la pasul 5

16. Dacă se poate decide, dacă a trecut sau a eșuat, se calculează răspunsurile greșite:

$W=L-R$

17. Se estimează măsura: $B=H/L+\ln(R-W)$

18. Se estimează eroarea standardă : $S=(L/(R*W))$

19. Se compară B cu criteriul de trecere-eșuare T

20. Dacă $(T-S)<B<(T+S)$, atunci se trece la pasul 5

21. Dacă $(B-S)>T$, atunci a trecut

22. Dacă $(B-S)>T$, atunci a eșuat

3.5. Construirea bancului de itemi

O pre-rechizită indispensabilă a testelor adaptative computerizate este bancul de itemi. Un banc de itemi este o colecție de itemi, împreună cu textul itemului, detaliile despre răspunsul corect și cele incorecte și dificultatea lor estimată curent. Mai pot fi prezenți indici despre apartenența itemului la careva subiect, domeniu.

Inițial, dificultatea itemilor este calculată după formula:

$$D=\ln((100-x)/x), \quad (2)$$

unde x este procentajul de succes obținut la acest item de un număr de persoane testate fără calculator, prin intermediul testelor "creion-hîrtie". Sau dificultatea poate fi stabilită intuitiv, în urma analizei cantitative a itemilor.

Un avantaj al algoritmilor de testare adaptativă este ușurința cu care pot fi introduși în banc itemii noi. Inițial aceștia pot fi administrați pentru candidat împreună cu cei existenți, însă fără a fi folosiți la estimarea capacității celui testat, mereu estimând dificultatea acestora după răspunsurile candidaților. Apoi, în final, se includ ca itemi obișnuiți.

Concluzii

Avantajele testării adaptative

Testarea adaptativă:

- evită administrarea întrebărilor "neinteresante". Itemii prea simpli deopotrivă cu cei prea dificili provoacă un comportament nedorit, cum ar fi ghicirea răspunsurilor, neatenția;
- poate fi mai scurtă;
- testele pot fi rapid dezvoltate și implementate. Testarea poate începe imediat după ce au fost introduși itemii. Rezultatele sunt expuse, de asemenea, imediat;
- poate servi drept o bună experiență.

Studiind problema testării adaptative, Rudner (1998) [1] a identificat următoarele avantaje:

- este simplu de administrat;
- rezultatele sunt disponibile imediat;
- testele sunt "reglate" individual, astfel cel examinat nu este nevoit să-și aștepte colegii de testare pentru a trece la o nouă secție;
- e nevoie de mai puțin timp pentru a le administra, deoarece e nevoie de mai puțini itemi pentru a obține o precizie acceptabilă. Testele adaptative pot reduce timpul de testare cu mai mult de 50% , menținând același nivel de veridicitate a rezultatelor;
- testele scurte reduc oboseala, factor ce poate afecta semnificativ rezultatele testării.

Problemele nerezolvate, legate de testarea adaptativă au fost identificate de Fairtest (1992) [2]

- testele computerizate au o serie de constrângeri. E nevoie de mai mult timp pentru a citi de pe monitor decât de pe hîrtie, uneori e mai complicat să depistezi greșala de pe monitor decât de pe hîrtie;
- majoritatea testelor computerizate afișează numai un item;
- există instituții de învățământ ce nu dispun de un număr suficient de calculatoare;
- prețul mare al testărilor pe calculator.

Limitările testării adaptative, identificate de Rudner [3]:

- testarea adaptativă nu este aplicabilă pentru toate subiectele;

- la fiecare examinat ce primește un set diferit de itemi poate fi observată o diferență;
- cei examinați nu pot modifica itemii precedenți.

Bibliografie

1. <http://www.rash.org/memo69>
2. <http://www.psych.umn.edu/psylabs/CATCentral/>
3. <http://www.ercial.net/scripts/cat>

Prezentat la 16.04.2004