Agregacija ocjena u obrazovanju

Lavoslav Čaklović

23. veljače 2022.

Sadržaj

1 Uvod

Ocjena, u našem obrazovnom sustavu, predstavlja odraz učenikovog stupnja ovladanosti znanja iz nekog predmeta podučavanja i/ili vještina. Takva ocjena može se izraziti brojem na nekoj skali, obično je to skala 1–5 (1–6) ili vrijednošću na temperaturnoj ljestvici (1–100). Neki obrazovni sustavi, ruski na primjer, koriste uređene kategorijske vrijednosti A–F gdje je A najviša, a F najniža ocjena. Običaj je, barem u našem obrazovnom sustavu, ocjenjivanje na temperaturnoj ljestvici nazivati bodovanjem. Završna ocjena je ocjena nastala nekom procedurom agregiranja predmetnih ocjena, obično na polugodištu ili na kraju godine.

Iz gore rečenog proizlazi da termin 'ocjena' nije jednoznačan, tj. značenje tog termina je kontekstualno ovisno. Ako se smatra da je ocjena 'mjera' ovladavanja određenog gradiva, onda se tu implicitno podrazumijeva da ona mjeri učenika nekim 'metrom'. Taj 'metar' bi trebao biti *nepristran* u smislu da bi svako (ponovljeno) mjerenje učenikovog znanja trebalo dati istu ocjenu. Prevedeno na svakodnevni obrazovni riječnik:

> *Svaki nastavnik koji ispituje (istog) učenika na temelju istog testa (ili druge provjere znanja) trebao bi mu dati istu ocjenu.*

Takav metar u obrazovnom sustavu ne postoji iz jednostavnog razloga što je ocjena (1) subjektivna i (2) ne postoji zacrtana procedura donošenja ocjene koja bi osiguravala nepristranost. Ako mi ne vjerujete čitajte nacionalni kurikul(um).

U ovom tekstu djelomično ćemo se baviti konstrukcijom ocjenskog metra (skale) s posebnim naglaskom na *metodu potencijala* zasnovanu od autora, a opća teorija mjerenja nadilazi okvire ovog časopisa[1]. Nećemo se baviti smislenošću i opravdanošću ocjenjivanja u obrazovnom sustavu; uzmimo kao gotovu činjenicu da se ocjene u školama donose i da će tako zasigurno biti još neko vrijeme. Pozabavit ćemo se agregacijom već postojećih ocjena koja je dovoljno zanimljiva sama po sebi, tim više što nastavnici to rade svakodnevno, svaki na svoj način.

2 Agregacija ocjena

Pod agregacijom ocjena podrazumijeva se određivanje ocjene učeniku na temelju tablice profila svih učenika u njegovom okruženju/razredu. Pod profilom se smatra skup ocjena učenika/ispitanika na temelju raznih kriterija ili nastavnih predmeta. U tablici 1 dana su tri profila od tri učenika A, B, C, a I, II, III mogu biti predmeti, uzastopni testovi ili nešto treće.

[1] Znatiželjnog čitatelja upućujemo na autorovu knjigu *Teorija vrednovanja s naglaskom na metodu potencijala* i to na onaj dio koji se bavi konstrukcijom izmjerive funkcije vrijednosti.

Tablica 1: Skup učeničkih profila u formi tablice.

	Predmet		
	I	II	III
A	63	59	66
B	66	57	69
C	69	56	68

Najčešće korištena metoda agregiranja je *aritmetička sredina* bodova ili, što je ekvivalentno, *zbroj* bodova. Ako gledamo rang-listu učenika iz tablice 1 dobivenu tom metodom onda je to: $A \succcurlyeq B \succcurlyeq C$[2].

Mnogi nastavnici zagovaraju medijan kao agregiranu ocjenu bez dublje argumentacije. Ovdje ćemo prodiskutirati mane jedne i druge procedure.

2.1 O skalama

Čitateljima je, pretpostavljam, već poznat pojam skale, ako ne drugačije onda kao skup rezultata mjerenja. Ovdje ćemo detaljno pojasniti taj pojam i vidjeti po čemu se one razlikuju. Najčešće korištene skale su: *nominalna, ordinalna, omjerna* i *intervalna*. Takvu je tipologiju[3] skala uveo S. S. Stevens (1946), a karakterizirane su dozvoljenom grupom transformacija: identitet, monotona, sličnost i afina transformacija. U nazivu skale već je prisutna informacija o prirodi izmjerene vrijednosti, a prava pozadina takve tipologije leži u stupnju jedinstvenosti numeričke reprezentacije mjerenog entiteta.

Nominalna ili kvalitativna skala je skala čije vrijednosti su oznake (nazivi), i nema drugih dozvoljenih transformacija osim identitete. Primjer takve skale je sama Stivensova tipologija. Ona služi, kako i samo njeno ime govori za imenovanje objekata.

Ordinalna skala je invarijantna na rastuće transformacije: $w = \phi(u)$, gdje je $\phi : \mathbb{R} \mapsto \mathbb{R}$ strogo rastuća funkcija. Iako su ordinalne skale sveprisutne u socijalnim znanostima, izbjegavaju se u fizici i tehnici jer se smatra da predstavljaju vrlo slabu formu mjerenja. Odgovori na anketna pitanja koja rangiraju stupanj zadovoljstva nekom uslugom su na ordinalnoj skali.

Drugi tip skale je *omjerna* skala. To je skala koja je invarijantna na množenje pozitivnim brojem: $w = \alpha \cdot u, \ \alpha > 0$. Ako je izmerena vrijednost dobivena uspoređivanjem s mjernom jedinicom onda je ona izražena na omjernoj skali (masa, naboj, duljina štapa). Omjerna skala ima definirano ishodište. Tipičan primjer omjerne skale je izražavanje brzine gibanja nekog tijela ili odgovori na anketna

[2]Relaciju $\succcurlyeq$ čitaj: 'bolji ili jednako dobar'

[3]iako postoje i drugačije tipologije

pitanja tip: *"Koliko vremena dnevno provodite u šetnji?"* (A) 0.5–1 sat, (B) 1–2 sata, (C) 2–3 sata. Zaključci provedeni na omjernoj skali putem statističkih analiza su korektni jer su invarijantni na množenje pozitivnim brojem.

Intervalna skala dopušta pozitivne afine transformacije ili, izraženo formulom: ako je u varijabla koju mjerimo onda je varijabla

$$w = \alpha \cdot u + \beta, \ \alpha > 0, \ \beta \in \mathbb{R}$$

jednako korisna (interpretabilna) u našem modelu kao i u. Omjer vrijednosti na intervalnoj skali je besmislen, ali je zato kvocijent duljina intervala između dviju vrijednosti smislen i poprima vrijednosti na omjernoj skali. Kako interpretirati podatke na intervalnoj skali? Na primjer, ako je temperatura zraka 15^o na nekoj skali to ne znači da je vani toplo ili hladno jer toplinu doživljavamo u odnosu na temperaturu tijela.

Primjer intevalne skale je temperatura izražena u ^oC i F (Fahrenheiht). Konverzija jedne skale u drugu je dana formulom $F = 2 \cdot C + 30$. Temperaturna skala nosi u sebi još jedno svojstvo, a to je kvalitativna smislenost razlike temperature. Drugi ilustrativan primjer je *korisnost* u teoriji očekivane korisnosti. Korisnost je također invarijantna na pozitivne transformacije, ali razlika korisnosti nema substancijalnog smisla jer ta razlika nije interpretabilna u modelu. Uz intervalnu skalu veže se i *log-intervalna* skala koja je okarakterizirana dozvoljenim transformacijama oblika $x \mapsto sx^r, \ s, r \in \mathbb{R}$. Logaritam te transformacije je afina transformacija.

Osim spomenutih skala postoje još neke specijalne skale koje se koriste u kartografiji, a pojedini autori predlažu i drugačije tipologije mjernih skala u koje nećemo ulaziti.

Zašto je važno na kakvoj skali vršimo mjerenja? S izmjerenim vrijednostima manipuliramo i na temelju tih manipulacija donosimo određene zaključke. Zaključak ne bi smio varirati ako vršimo dozvoljene transformacije skale.

Centralna mjera skale. Centralna mjera skale predstavlja vrlo rudimentarnu agregaciju u kojoj se izmjereni podaci reprezentiraju jednim jedinim brojem. Primjer centralne mjere je: *aritmetička sredina* za intervalnu skalu i *median* za ordinalnu skalu.

Ako je $x = (x_i), \ i = 1, \ldots, n$ vektor izmjerenih vrijednosti onda se aritmetička sredina $\mu(x)$ računa po formuli

$$\mu(x) = \frac{1}{n} \sum_{i=1}^{n} x_i$$

i predstavlja točku minimimuma kvadratičnog odstupanja (devijance), tj.

$$\mu(x) = \arg\min_x \sum_{i=1}^{n} (x - x_i)^2.$$

Medijan je 2-kvantil izmjerenih podataka i predstavlja točku minimuma apsolutnog odstupanja[4], tj.

$$\mathrm{med}(x) = \arg\min_x \sum_{i=1}^{n} |x - x_i|.$$

Koju centralnu mjeru uzimamo za naše podatke ovisi o skali. Jedan od zahtjeva na centralnu mjeru je invarijantnost na transformaciju koja definira skalu, a drugi zahtjev treba odražavati usklađenost centralne mjere s algebarskom strukturom koju skala (objekti) posjeduje. U skalu također treba ugraditi i zahtjev invarijantnosti centralne mjere na labeliranje objekata (nazivi) što se formulira kao invarijantnost centralne mjere na permutiranje izmjerenih vrijednosti.

Aritmetička sredina. Centralna mjera $f(x)$ podataka na intervalnoj skali treba zadovoljavati sljedeće zahtjeve:

1. $f(x+\beta) = f(x) + \beta$ – invarijantnost na dodavanje konstante,
2. $f(\alpha x) = \alpha f(x)$ – invarijantnost na množenje brojem,
3. $f(x) = f(x_\sigma)$ za svaku permutaciju σ – invarijantnost na permutiranje objekata,
4. $f(x+y) = f(x) + f(y)$– invarijantnost na algebarsku operaciju zbrajanja.

Zahtjevi i) i ii) su ekvivalentni zahtjevu invarijantnosti na dozvoljene transformacije skale. Oznaka $x + \beta$ za sumu vektora i konstante znači dodavanje konstante svakoj komponenti vektora, $x + y$ je zbrajanje vektora po komponentame, a x_σ je vektor nastao iz x permutiranjem njegovih komponenti tj. $(x_\sigma)_i := x_{\sigma(i)},\ \forall i = 1, \ldots, n$.

Lemma 2.1. *Centralna mjera podataka na intervalnoj skali je aritmetička sredina* $\mu(x)$.

[4] Gledano na taj način, točka minimuma apslolutnog odstupanja je jednoznačna ako je broj podataka neparan ili srednji interval sortiranih podataka ako je broj podataka paran. U drugom slučaju neki uzimaju za medijan lijevi rub unutarnjeg intervala, desni rub ili sredinu intarvala.

Dokaz. Dokaz ćemo provesti u 2 koraka.
1. $\sum_i x_i = 0 \implies f(x) = 0$. Zaista,

$$\begin{aligned} f(x) &\underset{\text{iv)}}{=} \sum_{i=1}^{n} f(0,\dots,0,x_i,0,\dots,0) \\ &\underset{\text{ii)}}{=} \sum_{i=1}^{n} x_i f(0,\dots,0,1,0,\dots,0) \\ &\underset{\text{iii)}}{=} \left(\sum_{i=1}^{n} x_i\right) f(1,0,\dots,0) \\ &= 0. \end{aligned}$$

2. $f(x,0,\dots,0) = \frac{x}{n}$. Zaista,

$$\begin{aligned} f(x,,0,\dots,0) &\underset{\text{i)}}{=} f\left(\frac{(n-1)x}{n}, -\frac{x}{n}, \cdots, -\frac{x}{n}\right) + \frac{x}{n} \\ &= \frac{x}{n} \quad \text{(zbog 1. koraka.)} \end{aligned}$$

Sada je

$$f(x) = \left(\sum_{i=1}^{n} x_i\right) f(1,0,\dots,0) = \frac{1}{n}\sum_{i=1}^{n} x_i = \mu(x).$$

□

Medijan.

Lemma 2.2. *Centralna mjera podataka na ordinalnoj skali je medijan.*

Dokaz. Neka je $\phi(\cdot)$ strogo rastuća transformacija. Označimo $m = \mathrm{med}(x)$. Tvrdimo $\phi(m) = \mathrm{med}(\phi(x))$. Zbog monotonosti transformacije ϕ:

$$x_i \le m \le x_k \iff \phi(x_i) \le \phi(m) \le \phi(k_k)$$

Ako je $[x_i, x_k]$ srednji interval sortiranih podataka, onda je $[\phi(x_i), \phi(x_k)]$ srednji interval transformiranih sortiranih podataka, odakle slijedi $\mathrm{med}(\phi(x)) = \phi(m)$. □

Geometrijska sredina. Jednostavnosti radi promatrajmo situaciju u kojoj imamo samo dva mjerenja: x i y. Sljedeća svojstva očekujemo od centralne mjere f na omjernoj skali:

1. $f(\alpha x, \alpha y) = \alpha f(x, y),\ \alpha > 0$ – invarijantnost na promjenu mjerne jedinice,
2. $f(1/x, 1/y) = 1/f(x, y)$ – idempotentnost invertiranja,

Lemma 2.3. *Centralna mjera podataka na omjernoj skali je njihova geometrijska sredina.*

Dokaz. Dokaz ćemo provesti za $n = 2$, gdje je n broj podataka, a u općem slučaju dokaz ide indukcijom po broju podataka. Prvo ćemo dokazati da je $f(1, x) = \sqrt{x}$. Zaista,

$$f(1,x) \underset{\text{ii)}}{=} \frac{1}{f(1, \frac{1}{x})} = \frac{x}{x \cdot f(1, \frac{1}{x})} \underset{\text{i)}}{=} \frac{x}{f(x, 1)} \implies f(1, x) = \sqrt{x}.$$

Sada je

$$f(x, y) \underset{\text{i)}}{=} xf\left(1, \frac{y}{x}\right) = x\sqrt{\frac{y}{x}} = \sqrt{xy}.$$

□

2.2 Funkcija vrijednosti

U analizi tablice učeničkih profila manipuliramo s ulaznim podacima (brojevima) na način koji može biti manje ili više smislen, ovisno o tome održavaju li ti brojevi kvalitativne ili kvantitativne odnose među entitetima. Kvalitativne relacije su na primjer:

i) ovaj je put *duži* nego onaj, ili

ii) *više mi sviđa* slika a nego slika b.

Bez obzira kako nazivali taj kvalitativan odnos među objektima (duljina, sviđanje) mi ćemo ga nazvati generičkim nazivom *preferencija* u oznaci $\succcurlyeq$. Prirodno je zahtijevati da relacija preferencije bude tranzitivna, tj.

$$a \succcurlyeq b \;\&\; b \succcurlyeq c \implies a \succcurlyeq c.$$

Jednaku preferiranost među objektima nazivamo *ekvivalencijom* i označavamo

$$a \sim b \iff a \succcurlyeq b \;\&\; b \succcurlyeq a.$$

Kvantitativni modeli izražavaju kvalitativni odnos nekim brojem na način da se poštuje preferencija. Funkciju V koja to čini nazivamo *funkcijom vrijednosti* i ona zadovoljava

$$(\forall a, b)\ a \succcurlyeq b \iff V(a) \geq V(b).$$

Na primjeru puteva, funkcija vrijednosti je *duljina puta mjerena u kilometrima (ili satima)*, na primjeru slika, obično nismo u stanju slike postaviti na neku skalu. Sofisticiraniji modeli u tom slučaju pridružuju svakom paru objekata *intenzitet preferencije* i na temelju takvih podataka konstruiraju skalu na kojoj se izražavaju vrijednosti objekata (v. str. 20 pod nazivom Metoda potencijala).

2.3 Utjecaj mjerenja kod agregacije ocjena

Kao metodu agregacije ocjena uzmimo *sumu*. Pretpostavimo da je ocjenjivač bodovao testove tri ispitanika A, B, C kao u tablici 2 (najveći mogući broj bodova po predmetu je 100). Kvantitavni odnos između 66 i 63 pokazuje da je kandidat B

Tablica 2: Bodovanje ispitanika.

	Predmet			
	I	II	III	zbroj
A	63	59	66	188
B	66	57	69	192
C	69	56	68	193

postigao bolji uspjeh od kandidata A na testu ako se gleda samo predmet I. U posljednjem stupcu tabele dana je suma bodova svakog kandidata koja predstavlja *opći uspjeh* kandidata za sva tri predmeta kao sumu bodova.

Pretpostavimo da je neki drugi ispitivač također ocijenio iste testove i rezultat njegovog ocjenjivanja je skup profila predstavljen u tablici 3. Numeričke vrijed-

Tablica 3: Drugo moguće bodovanje ispitanika.

	Predmet			
	I	II	III	zbroj
A	59	65	46	170
B	60	61	48	169
C	61	58	47	166

nosti pridijeljene uspjehu ispitanika za svaki predmet predstavljaju isti kvalitativni poredak za oba ocjenjivača. Drugi ispitivač je bio nešto blaži u ocjenjivanju za drugi predmet i stroži za preostala dva. Suma bodova međutim daje inverzni poredak za opći uspjeh: C-B-A za prvog ispitivača i A-B-C za drugog ispitivača iako je kvalitativni odnos (poredak) među učenicima, kod oba ispitivača, za svaki ispit isti. Aritmetička sredina ne uvažava 'kvalitetu' kod agregacije i ako netko želi uvažiti kvalitativni odnos među učenicima u agregiranoj rang-listi, onda suma bodova nije dobra metoda.

Ova jednostavna analiza upućuje na općenitiji pristup agregaciji u kojem se daje lista zahtjeva (aksioma) koje bi metoda trebala zadovoljavati. Pri tome je moguće pretjerati sa zahtjevima, tj. da lista zahtjeva bude prestroga i da ne postoji metoda koja ih zadovoljava. Teorija odlučivanja prošlog stoljeća puna je takvih teorema nemogućnosti koji dokazuju nepostojanje metode za razne sustave aksioma. Posebno je važna i tražena takva situacija kad neki sustav aksioma jedinstveno određuje metodu. Jedan od primjera je *teorija očekivane korisnosti* koju ekonomska literatura i dan danas obilato koristi. Od teorema ne egzistencije spomenuo bih samo Arowljev teorem u teoriji izbora[5] koji je 'razočarao' zagovornike zapadne demokracije jet tvrdi da ne postoji metoda koja zadovoljava 'na oko jednostavne demokratske zahtjeve'.

2.4 Kvalitativna i kvantitativna smislenost agregacije

Definicija 2.4. Dvije skale su *ordinalno ekvivalentne* ako je jedna skala kompozicija druge sa strogo rastućom funkcijom i svaku od njih nazivamo *ordinalnom skalom*. Strogo rastuća transformacija naziva se *dopuštena transformacija* ordinalne skale.

Teorem 2.5. *Nejednakost između aritmetičkih sredina nije invarijantna na strogo rastuće transformacije. Drugim riječima, ako je V ordinalna funkcija onda*

$$\sum_{i=1}^{n} \lambda_i V(a_i) \geq \sum_{i=1}^{n} \mu_i V(a_i)$$

ne povlači da za svaku strogo rastuću funkciju h vrijedi

$$\sum_{i=1}^{n} \lambda_i h(V(a_i)) \geq \sum_{i=1}^{n} \mu_i h(V(a_i)),$$

gdje je $\lambda_i, \mu_i \geq 0$ i $\sum_{i=1}^{n} \mu_i = 1$, $\sum_{i=1}^{n} \lambda_i = 1$.

[5] Arrow, Kenneth J. (1951), *Social Choice and Individual Values*, John Wiley and Sons, New York.

Dokaz je kratak i ostavljamo ga čitatelju.

Teorem tvrdi da aritmetička sredina kao kvantitativni podatak nije smislena na ordinalnoj skali. U literaturi se koristi termin da ona nije *kvantitativno smislena*.

Ako broj bodova za svaki predmet, u tablicama 2 i 3 shvatimo kao odraz *poretka* kandidata na ispitu, onda je računanje prosjeka očigledno *kvantitativno besmislena* operacija jer su oba ocjenjivača dala isti poredak kandidata za svaki ispit (v. tablicu 4), a agregirani rangovi su različiti.

Tablica 4: Kvalitativni odnos učenika po predmetima. Rangovi.

	Predmet		
	I	II	III
A	1	3	1
B	2	2	3
C	3	1	2

Kako bismo interpretirali tu tvrdnju u kontekstu donošenja ocjena? Ako bismo dozvolili nastavnicima da svaki ima svoju skalu ocjena, na primjer jedan koristi skalu 1-2-3-4-5-6, a drugi koristi skalu 10-12-15-17-21-28, onda bi se moglo desiti da aritmetička sredina, kao metoda agregacije, u oba slučaja ne daje isti poredak kandidata. Jedan od načina rješavanja te nelagode je natjerati sve nastavnike da koriste istu skalu i da prođu ozbiljan trening u donošenju ocjena. Primjer takvog okruženja je ocjenjivanje figura na umjetničkom klizanju na primjer, a glavni razlog za takva nastojanja u teoriji izbora je onemogućavanje manipulacije[6].

2.5 Agregacija ordinalnih vrijednosti

Ocjene u hrvatskom obrazovnom sustavu poprimaju vrijednosti u uređenom skupu kategorija, a to što su kategorije označene brojevima u skupu $\{1, 2, 3, 4, 5\}$ nosi dodatne poteškoće. Stoga je opravdano pitanje kako agregirati ocjene ako aritmetička sredina nije kvantitativno smislena. Ovdje ćemo analizirati neke od metoda agregacije ordinalnih podataka.

2.5.1 Bordino pravilo

Tablica 4 predstavlja rang-listu kandidata po predmetima s tom razlikom što manji broj predstavlja lošiju poziciju na rang listi. Takva tablica predstavlja najsiroviju

[6] Pojam *manipulacija* je višeznačan i kontekstualno ovisan. U svakom konkretnom slučaju (metodi) treba eksplicitno reći o kakvoj manipulaciji se radi.

informaciju o kandidatima. Pitanje je možemo li ju iskoristiti kao takvu i donijeti negu agregiranu rang-listu na temelju tih podataka, a dodatno je pitanje koliko je ta rang-lista zadovoljavajuća.

Jedan način agregacije ordinalnih lista dao je Jean-Charles de Borda još davne 1770. godine. On je kritizirao način izbora u Francusku akademiju i predložio metodu, danas poznatu kao Bordino pravilo ili Bordin indeks.

Pravilo je jednostavno. Ako je u pitanju n kandidata onda se svakoj rang-listi prvo pridruži stupac numeričkih vrijednosti u skupu $\{1, 2, \ldots, n\}$ gdje se prvom mjestu na listi daje n bodova, drugom $n - 1$ i tako dalje kao što je urađeno u tablici 4. Tada se

svakom kandidatu pridijeli zbroj njegovih bodova iz svake liste.

Originalno Bordino pravilo, umjesto tablice oblika 4, koristi tablicu u kojoj su svi bodovi umanjeni za jedan. Jednostavno je pokazati da je krajnji rezultat, tj. rang-lista, u oba slučaja isti.

Tablica 5: Bordin indeks. Rang-lista učenika iz tablice 4.

	Borda
A	4
B	7
C	6

Na dodatno pitanje koliko je rezultat zadovoljavajući (prihvatljiv) teško je reći bez dodatnih kriterija što ovdje znači 'prihvatljivost'. Navodimo neke od tih kriterija koji se najčešće spominju u literaturi

Nezavisnost o irelevantnoj alternativi. U najkraćim crtama taj kriterij karakterizira agregativne procese u kojima dodavanje nove opcije (alternative) u račun ne mijenja poredak starih opcija.

Bordin indeks **ne zadovoljava** *nezavisnost o irelevantnoj alternativi.*

Pravilo većine. Kandidat na izborima koji ima više od 50% glasova je pobjednik.

Bordin indeks može **narušiti** to pravilo.

Monotonost. Ako izborne rezultate promijenimo na način da pobjednika još više favoriziramo on će i nakon toga biti pobjednik.

Bordin indeks **zadovoljava** monotonost.

Condorcetov kriterij. Ako postoji kandidat koji nadglasava svakog pojedinog protivnika (po broju listića) tada je on pobjednik izbora (tzv. Condorcetov pobjednik).

Bordin indeks može **narušiti** to Condorcetov kriterij.

Bordino pravilo se može jednoznačno okarakterizirati grupom kriterija koje nećemo ovdje navoditi. Štoviše, koliko je autoru ovog teksta poznato, postoje bar dvije grupe kriterija koje jednoznačno određuju Bordino pravilo.

2.5.2 Većinska procjena. Medijan.

Balinski i Laraki[7] suzili su ocjenjivačku skalu na opće prihvaćenu ordinalnu verbalnu skalu: *odličan (5), vrlo dobar (4), dobar (3), prihvatljiv (2), slab (1), nedovoljan (0)*, nadajući se da će time ublažiti mogućnost manipulacije. Za operator agregacije predlažu medijan profila jer su ocjene donesene na ordinalnoj skali. Ako je broj ocjena paran, onda se kao medijan uzima **manja vrijednost** od dvije srednje vrijednosti.

Jednostavnosti radi, u primjerima koji slijede umjesto verbalnih ocjena pisat ćemo brojeve, imajući u vidu da se radi o ordinalnoj skali. Primjeri su iz teorije izbora ali se lako mogu 'prevesti' u obrazovno okruženje.

2.5.3 Neke 'paradoksalne' situacije kod agregiranja medijanom

Evo nekoliko situacija u kojima rezultati agregacije pomoću medijana odudaraju od 'očekivanog rezultata', s napomenom da je teško definirati očekivani rezultat jer to može biti posljedica ustaljene prakse ili naše zablude. Iz tog razloga i termin 'paradoks' treba uzeti s rezervom.

Paradoks 1. *Dodavanje glasačkog listića koji svim kandidatima daje najlošiju ocjenu može poremetiti redoslijed kandidata.*

Promatrajmo dva kandidata A i B, čije su ocjene:

	ocjene	medijan
A	1, 2, 4, 4, 6	4
B	2, 3, 3, 6, 6	3

Dakle, A je pobjednik. Ako dodamo još jedan glasački listić s nedovoljnom ocjenom za oba glasača, tada su nove ocjene

[7] u svojoj knjizi *Election by Majority Judgement: Experimental Evidence*, Technical report. Laboratoire d'Econometrie 1, rue Descartes F-75005, Paris (2007)

	ocjene	medijan
A	0, 1, 2, 4, 4, 6	2
B	0, 2, 3, 3, 6, 6	3

odakle slijedi da je B sada pobjednik jer agregacija uzima donju granicu srednjih vrijednosti.

Paradoks 2. *Dokidanje* Pretpostavimo da su dvoje prijatelja Romeo i Julija prvoj tablici iz gornjeg paradoksa naknadno dodali svoje ocjene: $A \longleftarrow 5$, $B \longleftarrow 6$ (Romeo) i $A \longleftarrow 6$, $B \longleftarrow 5$ (Julija). Nova tablica ocjena je

	ocjene	medijan
A	1, 2, 4, 4, 5, 6, 6	4
B	2, 3, 3, 5, 6, 6, 6	5

Romeo i Julija su očekivali[8] da će se njihove ocjene 'poništiti', ali nisu, B je postao pobjednikom.

Paradoks 3. *Povećana podrška nekom kandidatu može ga od pobjednika pretvoriti u gubitnika.*

Pretpostavimo da su Romeo i Julija, dvoje novih glasača, dali kandidatu A iz prvog primjera ocjenu 6, a kandidatu B ocjenu 5. Nova tablica ocjena je:

	ocjene	medijan
A	1, 2, 4, 4, 6, 6, 6	4
B	2, 3, 3, 5, 5, 6, 6	5

Očigledno je A postao gubitnik.

To su samo neke od 'paradoksalnih' situacija. Ima ih još, a mogu se naći u spomenutoj knjizi od Balinskog i Larakija ili u autorovoj knjizi *Teorija odlučivanja s naglaskom na metodu potencijala.*

[8] Ovo očekivanje je tipično za ljude koji su naviknuti usrednjavati. Usrednjavanjem na intervalnoj skali je moguće 'poništiti' takve ocjene.

2.5.4 Bordin većinski indeks

Ordinalna agregacija pomoću medijana prirodno se uklapa u ordinalni kontekst ocjenjivanja, neosjetljiva je na manipulacije, ali ponekad nudi neočekivane i paradoksalne rezultate. S druge strane, aritmetička sredina izgleda da pati od izrazite mogućnosti manipulacije, s manje paradoksalnih mogućih ishoda, pa se prirodno javila ideja o ujedinjavanju obiju metoda. Jedan mogući hibrid nude Zahid i Swart[9]. Oni su eksperimentirali s lingvističkom skalom {*odličan, vrlo dobar, dobar, prihvatljiv, slab, nedovoljan*} i svoj indeks nazvali su *Bordin većinski indeks*[10].

Neka je $\{g_1, \ldots, g_k\}$ skup ocjena sa svojstvom $g_1 < g_2 < \cdots < g_k$ i a jedan od kandidata na izborima. Neka je n_i broj glasača koji su kandidatu a dali ocjenu $g_i,\ i = 1, \ldots, k$. Definirajmo *Bordin većinski indeks* $C(a)$ kandidata a formulom

$$C(a) := n_1 \cdot 0 + n_2 \cdot 1 + \ldots + n_k \cdot (k-1). \tag{BVI}$$

Funkcija C definirana je na skupu profila s vrijednostima u $\mathbb{N}^m$, gdje je m ukupan broj kandidata. Napomenimo da vrijednost $C(a)$ ne ovisi o vrijednostima $g_1, \ldots, g_k$, pa skala $\{g_1, \ldots, g_k\}$ može biti i ordinalna lingvistička skala.

Za razliku od Bordine metode, ovdje pojedini glasači ne daju rang-listu kandidata, nego ih vrednuju na zadanoj ocjenskoj ljestvici. Bordina metoda svakom glasaču nudi jedino mogućnost da kandidate rangira na nekoj individualnoj skali u svom mentalnom prostoru. Bordin većinski indeks nudi nešto drugačiji pristup, on od svakog glasača zahtijeva da kandidata ocijeni na unaprijed zadanoj ljestvici ocjena koja je (trebala bi biti) opće prihvaćena. Takav pristup zahtijeva jaku socijalnu povezanost glasača i zdrav razum u implementaciji zadane ocjenske skale. Mnogi teoretičari izbornih metoda smatraju da je to smjela pretpostavka. Izborna slika danas je nešto gdje je prisutan sukob interesa, prisutno je taktizirano glasovanje i postoji pritisak na glasače (mediji, stranačka pristranost) da koriste **tuđe procjene**. Takvo ponašanje je blisko, po svojoj formi, kupovanju glasova samo se tako ne zove. U obrazovnom kontekstu glasač je nastavnik koji učeniku (kandidatu) daje ocjenu. Socijalna povezanost unutar jednog obrazovnog sustava je izrazitija jer nastavnici surađuju i imali su iste učitelje.

Agregacija kod računanja Bordina većinskog indeksa samo formulom podsjeća na Bordinu agregaciju. Kod Bordine metode, bodovi koje dobiva svaki kandidat na individualnoj listi ovise o broju kandidata, što ovdje nije slučaj. Na prvi pogled, većinski indeks je više utilitaristički nego ordinalan po svojoj prirodi. Ordinalnost je prisutna jedino u izboru ocjenske skale.

[9] u svojoj knjizi *The Borda Majority Count.* Department of Philosophy, Tilburg University, the Netherlands (2010)

[10] eng. *Borda majority count.* Gotovo identičan tzv. *Range Voting* samo ima manju skalu ocjena.

Većinski indeks C ima sva dobra svojstva kao i utilitaristički indeks S i manje je osjetljiv na manipulacije jer je skala sužena s intervala na svega 6 mogućih ocjena. Sve primjedbe i paradoksi koje uzrokuje agregacija pomoću medijana sada ne stoje što trivijalno slijedi iz definicije većinskog indeksa. Štoviše, većinski indeks čuva konzistentnost prioriteta i pobjednika o čemu govori sljedeći teorem.

Teorem 2.6. *Većinski indeks čuva konzistentnost pobjednika, tj. ako neki kandidat pobjeđuje u svakom okrugu onda pobjeđuje i u cijelom elektoratu.*

Većinski indeks čuva konzistentnost prioriteta, tj. ako za dva kandidata a, b vrijedi $a \succcurlyeq b$ u svakom okrugu, onda je $a \succcurlyeq b$ u cijelom elektoratu.

Dokaz. Neka je I∪II particija izbornog tijela i $\succcurlyeq_I$, $\succcurlyeq_{II}$ relacije socijalne preferencije za oba dijela respektivno. Pretpostavimo da je

$$a \succcurlyeq_I b \succcurlyeq_I c \quad \text{i} \quad a \succcurlyeq_{II} b \succcurlyeq_{II} c.$$

Zbog $C = C_I + C_{II}$ slijedi

$$a \succcurlyeq b \succcurlyeq c,$$

što dokazuje i konzistentnost prioriteta i konzistentnost pobjednika. □

Primjer. Ovo je artificijelni primjer koji pokazuje da većinski indeks može dati drugačije rangiranje od aritmetičke sredine. Tablica 2 može se dobiti na sljedeći način.

Pretpostavimo da je učenik A na predmetu I dobio niz ocjena a_1, na predmetu II je dobio niz ocjena a_2 i na predmetu III je dobio niz ocjena a_3:

$$\begin{aligned} a_1 &= (2, 5, 4, 6, 3) \\ a_2 &= (4, 2, 7, 5, 2) \\ a_3 &= (1, 4, 5, 3, 6) \end{aligned}$$

Suma ocjena u nizovima je 63, 59, 66 respektivno što predstavlja prvi redak tablice. Ocjene za učenika B neka su:

$$\begin{aligned} b_1 &= (0, 5, 7, 5, 3) \\ b_2 &= (0, 4, 7, 7, 0) \\ b_3 &= (3, 4, 7, 8, 1) \end{aligned}$$

i za učenika C

$$\begin{aligned} c_1 &= (0, 1, 6, 6, 5) \\ c_2 &= (0, 1, 3, 5, 5) \\ c_3 &= (1, 2, 2, 8, 5). \end{aligned}$$

Ako niz a_1 zamijenimo s nizom $a_1' = (0, 3, 4, 5, 5)$ koji ima jednaku sumu ocjena, većinski indeks daje drugačiji poredak kandidata. To je zato jer većinski indeks uvažava broj ukupan ocjena koje je učenik dobio iz svih predmeta Za večinski indeks važna je povijest dobivanja ocjena, a ne samo njihov zbroj po predmetu kao što to radi aritmetička sredina.

2.5.5 Usporedba medijana i aritmetičke sredine

U tablici 6 učenici A, B, C ocijenjeni su na skali 0-1-2-3-4 (veća ocjena je bolja) iz pet predmeta. U predzadnjem retku je računat prosjek, a u zadnjem medijan. Primijetimo da je učenik A najbolji po aritmetičkoj sredini, a učenik B je najbolji po medijanu. Učenik C je, ako ga uspoređujemo s drugima prema broju uspješnijih ocjena[11], bolji i od A i od B (posljednja 3 retka tablice ocjena). Bordin većinski index (BVI) daje vrijednosti 18, 12, 28 što stavlja učenika C na prvo mjesto kao i kod metode većine. Na *Stanford Encyclopedia od Philosofy*[12] dan je izbor literature koja zastupa medijan, s jedne strane, i izbor literature koja je protiv medijana kao metode izbora, s druge strane. Želim ovdje istaknuti da nema nekog konsenzusa da je medijan bolji ili lošiji od aritmetičke sredine. Zapravo te dvije metode agregiraju različite podatke, medijan agregira položajne liste, a aritmetička sredina agregira originalne vrijednosti.

Tablica 6: text

	A	B	C
	4	3	1
	4	3	2
	2	0	3
	2	3	4
	1	0	2
BVI	18	12	28
ar. sred.	2.6	1.8	2.4
medijan	2	3	2

Što se aritmetičke sredine tiče, ona je podložna manipulaciji, dozvoljava kompenzaciju u ocjenjivanju i stoga obrazovni sustav unosi prolazni prag u ocjenjivanje. Posljedica toga je da nastavnici izjegavaju ocjene neposredno ispod praga.

[11] Condorcetova metoda većine

[12] https://plato.stanford.edu/entries/voting-methods/#VotiGrad

Navedimo još jedan primjer koji govori u prilog zdravog razuma u donošenju grupne rang-liste u slučaju kad postoji *preferencijalna zavisnost atributa.*

2.5.6 Preferencijalna (ne)zavisnost atributa

Agregiranje ocjena je posebno teško modelirati jer tu dolazi do izražaja efekt interakcije među atributima (kriterijima) koji narušava njihovu aditivnost. Evo primjera za ilustraciju.

Primjer 2.7. Četiri učenika pristupila su ispitima iz fizike, matematike i ekonomije i rezultati tih ispita dani su u sljedećoj tablici učeničkih profila:

	F	M	E
a	18	12	6
b	18	7	11
c	5	17	8
d	5	12	13

Neka je $\succcurlyeq$ neka relacija preferencije na studentima. Pitanje je ima li ta relacija aditivnu reprezentaciju, tj. može li suma atributnih funkcija vrijednosti oblika

$$v_F + v_M + v_E$$

predstavljati njenu funkciju vrijednosti. **Jedan od uvjeta** za takvo nešto je *preferencijalna nezavisnost* stributa koju je najbolje objasniti ilustracijom na našem primjeru.

Iz tablice vidimo da studenti a, b imaju jednak broj bodova iz fizike, a prema bodovima iz matematike i ekonomije a je bolji od b. Isto razmišljanje vrijedi i za studente c, d. Tada su matematika i ekonomija, kao atributi, *preferencijalno nezavisni* ako $a \succcurlyeq b \iff c \succcurlyeq d$. Očito je da aritmetička sredina odražava preferencijalnu nezavisnost atributa po konstrukciji.

Osim preferencijelne nezavisnosti postoji još nekoliko uvjeta poznatih kao uvjeti rješivosti za aditivnu reprezentaciju funkcije vrijednosti ali ih nećemo objašnjavati iz jednostavnog razloga jer su previše tehnički.

Pogledajmo sada kako razmišlja nastavnik koji boduje na ljestvici 0–20 ali s pragom prolaznosti 10. Tablica sugerira da studente a i b rangiramo ispred sudenata c i d. Student c je podbacio u dva predmeta i razumno ga je staviti na dno liste. Student a ima veći zbroj ocjena iznad praga od studenta b pa nastavnik smatra da bi studenta a trebalo rangirati ispred b. Slijedeći istu nit razmišljanja student d je bolji od studenta c jer ima ocjene iz matematike i ekonomije iznad praga, dok student c ima samo jednu takvu ocjenu. Stoga je razumno studenta d rangirati ispred c. Rezultat ovakvog razmišljanja je rangiranje

$$a \succcurlyeq b \succcurlyeq d \succcurlyeq c$$

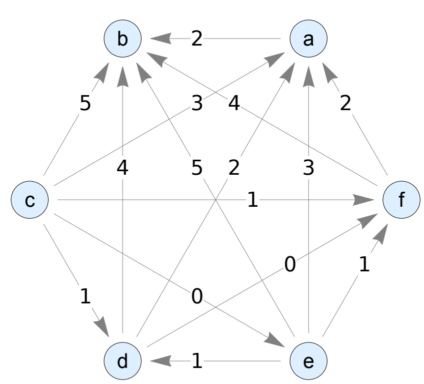
za koje atributi M i E nisu preferencijalno nezavisni i koje se ne može dobiti računanjem srednjeg općeg uspjeha (bez obzira na težine predmeta).

Istaknimo još jednom da aritmetička sredina posjeduje mogućnost kompenzacije, tj. da student može podbaciti u jednoj komponenti i nadoknaditi taj neuspjeh uspjehom u drugim komponentama. Ako ispitivač želi kažnjavati nerad, onda će uvesti prolazni prag u ocjenjivanje koji generira nelinearnu skalu i takvo je razmišljanje bilo ugrađeno u rješavanje gornjeg primjera.

3 Subjektivno mjerenje

Zamislite sebe kako razmišljate o nekim svojim postupcima iz prošlosti i pokušavate ustanoviti stupanj vlastite slobode odabira u svakom od njih. U nekim

Slika 1: Šest opcija



situacijama bili ste 'primorani', a u nekim drugim situacijama ste mogli postupiti i drugačije, bilo vam je dano na 'volju'. Jedan od načina da to učinite je taj da sami odredite skalu od 0–10 (ili 0–100) i na toj skali poredate postupke (opcije) po vlastitom nahođenju. Kod takvog vrednovanja vi se oslanjate na svoj osjećaj, iskustvo,…nazovite to kako želite, ali misaoni proces u pozadini takvog vrednovanja vam je nedostižan.

Druga mogućnost je da pokušate razmotriti dvije opcije i prednost date onoj u kojoj je bilo više slobode odabira. Ako možete tu prednost izraziti brojem još bolje. Učinite to za svake dvije opcije ili samo za one za koje ste sigurni u procjeni. Recimo da su vaše usporedbe u parovima kao na slici 1, gdje je strelica usmjerena prema opciji kojoj dajete prednost. Na toj slici postoji određena konzistentnost u procjenama jer za svake odabrane tri opcije, napr. $\{c, f, b\}$ suma brojeva duž zatvorenog puta $cfbc$ jednaka je $1+4-5=0$. Na tom putu ste dionicu bc prolazili u suprotnom smjeru od strelice i njen doprinos sumi iznosi -5. Upravo zbog navedene konzistentnosti nije teško zaključiti da je jedno moguće vrednovanje tih opcija

a	b	c	d	e	f
4	6	1	2	1	2

.

Za dokaz gornje tvrdnje povucimo paralelu prikazanog grafa s električnom mrežom u kojoj je razlika potencijala između dva čvora dana brojem uz odgovarajuću žicu (strelicu). Dovoljno je primijetiti da možemo fiksirati vrijednost potencijala V jednog čvora, napr. $V(c) = 1$, a potencijale ostalih čvorova odredimo uvažavajući danu razliku potencijala.

Što ako konzistentnost koju smo opisali nije zadovoljena? Tada se stvari kompliciraju (v. odjeljak 3.3).

3.1 Graf preferencije.

Sliku koje ste dobili, u literaturi nazivamo *usmjerenim grafom s težinama.* Taj graf može poslužiti za daljnju analizu vaših podataka. Jedna od njih je *metoda potencijala*, autorova zamisao, o kojoj će kasnije biti riječi.

U drugim konkretnim situacijama umjesto 'osjećaja' prepoznajemo kvalitete: duljinu, poželjnost, riskantnost, vrijednost, a entiteti koje uspoređujemo su fizički objekti ili misaoni konstrukti kao što su: akcije, scenariji, opcije…

U osnovi mjerenja stoji uspoređivanje objekata u parovima, što znači da za svaka dva entiteta ili opcije, mjerni instrument ili čovjek donositelj odluke treba biti u stanju reći koji je objekt dulji, poželjniji, riskantniji, kvalitetniji, vrijedniji. Rezultate takvih usporedbi nazivamo, u najapstraktnom smislu, *relacijom* na skupu entiteta; mi ćemo iz razloga konzistentnosti s literaturom tu relaciju nazivati *preferencijom* kao što smo to činili i u odjeljku 2.2 na str. 7.

3.2 Mjerenje razlike preferencije

Pretpostavimo da je $(S, \succcurlyeq)$ relacija preferencije. Na prvi pogled izgleda razumno reći da donositelj odluke preferira a u odnosu na b više nego što preferira c u odnosu na d ako i samo ako je spremniji odustati od b u zamjenu za a nego odustati od d u zamjenu za c.

Označimo s $(a \leftarrow b)$ prihvaćanje a u zamjenu za b ili kraće *zamjena.* Donositelj odluke sada ima dvije relacije slabe preferencije s kojima se mora suočiti, jedna je relacija $\succcurlyeq$ na objektima iz S, a druga je relacija $\succcurlyeq_e$ na skupu zamjena $(S \leftarrow S)$ koji je zapravo kartezijev produkt $(S \times S)$. Svaka relacija za sebe generira funkciju vrijednosti (uz određene uvjete), s napomenom da je prva definirana na skupu objekata, a druga na skupu zamjena. Nas zanima njihova usklađenost, odnosno odgovor na pitanje je li moguće naći takvu funkciju vrijednosti v na skupu objekata S koja zadovoljava:

$$a \succcurlyeq b \iff v(a) \geq v(b) \tag{1}$$

$$(a \leftarrow b) \succcurlyeq_e (c \leftarrow d) \iff v(a) - v(b) \geq v(c) - v(d). \tag{2}$$

Takvu funkciju, ako postoji, nazvat ćemo *izmjerivom funkcijom vrijednosti.* Očito je v ordinalna funkcija vrijednosti, dok (2) zahtijeva da je razlika $v(a) - v(b)$ ordinalna funkcija vrijednosti na skupu zamjena usklađena s relacijom $\succcurlyeq_e$ na zamjenama.

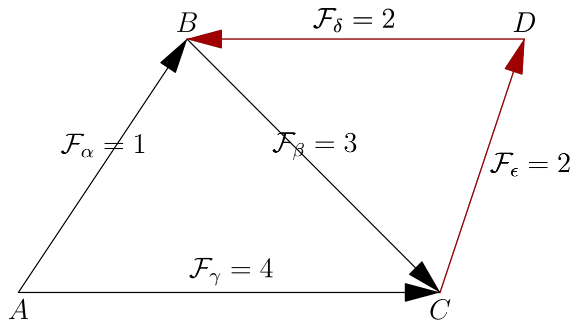
Mnogi teoretičari smatraju da ne treba donositelje odluke prisiljavati na zamišljanje zamjena jer da oni imaju urođeni osjećaj za intenzitet preferencije. Zamišljamo li relaciju preferencije kao usmjeren graf, onda je intenzitet preferencije broj koji je pridružen svakom luku u grafu kao što je to prikazano na slici 1. Intenzitet tako postaje funkcija na skupu zamjena koja može i ne mora biti funkcija vrijednosti na skupu zamjena usklađena s $\succcurlyeq_e$. Takva prezentacija slabe preferencije umnogome olakšava donositelju odluke da se koncentrira na parove i svakom paru (a, b) pridruži intenzitet preferencije[13]. Na kraju krajeva nije bitno kako ljudi zamišljaju zamjene, bitno je da to rade na konzistentan način.

3.3 Metoda potencijala

Metoda potencijala sjedinjuje fleksibilnost uspoređivanja u parovima i konstrukciju izmjerive funkcije vrijednosti. Umjesto da izgrađuje preferencije na zamjenama, donositelju odluke nudi se mogućnost iskazati to isticanjem *intenziteta preferencije* za zadani par alternativa kao što je to prikazano ne slici 2. Slika pokazuje *graf preferencije* koji nije potpun jer alternative A i D nisu uspoređene. Umjesto da govorimo o paru alternativa (B, A) mi ćemo govoriti o strelici (luku) koja izlazi iz A i ulazi u B. Odgovarajući intenzitet preferencije izražavamo brojem F_α koji na slici iznosi 1. Vrijednost $\mathcal{F}_\alpha = 0$ izražavala bi činjenicu da donositelj odluke

[13] Bilo po svom nahođenju, prihvaćanjem nekih egzaktno izmjerenih vrijednosti ili njihovom revizijom.

Slika 2: Primjer grafa preferencije



jednako preferira te dvije opcije. Katkada ćem luk α s odgovarajućom težinom s 1 označavati kao $B \xleftarrow{1} A$.

Intenzitet preferencije je nenegativan[14] broj kojeg donositelj odluke pridjeljuje preferenciji na nekoj skali u njegovoj mentalnoj slici. Za subjektivne preferencije uvriježeno je davati intenzitet u skupu $\{0, 1, 2, 3, 4\}$, što odgovara verbalnom intenzitetu: *jednaka*, *slaba*, *jaka*, *izrazita* i *apsolutna* preferencija, ali nisu isključeni niti drugačiji rasponi vrijednosti. Iskustvo pokazuje da je korisno povećati raspon skale intenziteta ako je broj alternativa veći.

Dakle, svakom luku $\alpha \in \mathcal{A}$ iz skupa lukova $\mathcal{A}$ donositelj odluke pridjeljuje nenegativan broj $\mathcal{F}_\alpha \geq 0$. Funkciju $\mathcal{F}$ definiranu na skupu lukova nazivamo *tokom preferencije* ili kraće *tokom*.

3.3.1 Konzistentnost toka.

Za razliku od preciznosti, koja je smislena kod mjerenja s mjernom jedinicom, ovdje ima smisla *konzistentnost* ulaznih podataka. Podgraf (na slici 2) određen lukovima α, β, γ je konzistentan jer je $\mathcal{F}_\alpha + \mathcal{F}_\beta = \mathcal{F}_\gamma$, dok podgraf određen lukovima β, ϵ, δ nije konzistentan jer je $F_\epsilon + \mathcal{F}_\delta + \mathcal{F}_\gamma = 7 \neq 0$. Tok je konzistentan ako je za svaki ciklus (zatvoren put) suma intenziteta preferencije duž lukova tog ciklusa jednaka nuli.

Algebarski se to može izraziti i postojanjem funkcije $X : V \longrightarrow \mathbb{R}$ takve da je za svaki luk $\alpha = (B, A)$

$$X(B) - X(A) = \mathcal{F}_\alpha. \tag{3}$$

U sustavu jednadžbi (3) imamo toliko jednadžbi koliko ima lukova u grafu. Donositelj odluke ne mora nužno dati konzistentan graf. Matematika kaže da bez

[14] Broj koji nije negativan. Može biti jednak 0 ili pozitivan.

obzira na konzistentnost toka, sljedeća jednadžba

$$A^\tau A\mathcal{F} = A^\tau\mathcal{F} \tag{4}$$

uvijek ima rješenje X i to rješenje predstavlja skalu s vrijednostima X_A, X_B, X_C, X_D na koju su smješteni objekti. Funkciju X nazivamo *potencijalom*, a metodu koja konstruira tu skalu nazivamo *metodom potencijala* (L. Čaklović, A. S. Kurdija (2017) *A universal voting system based on the Potential Method.*

Metoda potencijala omogućava mjerenje bilo kakvih objekata čija je međusobna 'udaljenost' izražena putem grafa preferencije.

4 Dualnost u ocjenjivanju

Jedna od primjena metode potencijala u kojoj dolazi do izražaja njena fleksibilnost je agregacija ocjena. Gotovo isti pristup koristi se u kontekstu strojnog učenja i umjetne inteligencije što nadilazi ciljeve ove knjige. U samoj jezgri tih problema nalazi se algoritam koji detaljizira Helmholzovu ideju da je mozak stroj za predviđanje i da revidira svoje iskustvo sadržano u memoriji na temelju ulaznih podataka (trenutne percepcije).

Metoda potencijala ima dvije novine koje joj omogućavaju tu fleksibilnost. To su: (1) *operator adaptacije* i (2) mogućnost uvažavanja *sinergije* bez eksplicitne definicije što bi sinergija trebala biti. Sinergija i adaptacija ne idu jedno bez drugog i teško ih je separirati jednako kao što je teško separirati prostor i vrijeme.

Samoprocjena. Adaptacija. Adaptacija je sposobnost sustava da procijeni stanje vlastitih parametara na temelju promijenjene okoline i svog vlastitog ustrojstva.

Svaki sustav (čovjek) ragira na okolinu koja mu nudi vrijednosti polaznih parametara ili mu određuje dinamiku promjene tih parametara. Operator adaptacije revidira, uvažavajući strukturu sustava, vrijednosti polaznih parametara i izračunava nove vrijednosti koje sustav koristi u procjeni vlastitog stanja i omogućava mu reakciju na promijenjeno stanje okoline.

Adaptacija nosi u sebi dualitet u odnosu između polaznih parametara i mogućih reakcija, tj. odgovora sustava na promjenu parametara. Tehnički termin za polazne parametre je *atributi*, a za moguće odgovore *alternative*. Evo konkretnog primjera.

Sinergija. Jedan od aspekata sinergije je i sinkronicitet (koherentnost) kako je to nazvao Jung. To je lijepa zamisao ali je s algoritamskog gledišta neupotrebljiva.

Mi u svakodnevnom životu konstatiramo sinkronicitet ali na to gledamo kao na slučajnost ili na čudo.

U gotovo svim današnjim modelima učenja koristi se statistika i vjerojatnostni modeli koji stoje u njenoj pozadini. Vjerojatnost je također jedna funkcija vrijednosti u koju je ugrađeno *svojstvo aditivnosti* događaja, slično nezavisnosti o irelevantnoj alternativi. U najkraćim crtama, to svojstvo zahtijeva da ako događajima A, B pridružimo njima disjunktni događaj C onda odnos vjerojatnosti od $A \cup C$ i $B \cup C$ ostaje nepromijenjen. Takvim procedurama je u samom startu otežana mogućnost da modeliraju i objasne one pojave u kojima se pojavljuje sinergija. Sinergija je pojava (konstelacija) koja drastično mijenja stanje svijeta ako je prisutna i teško ju je definirati u operativnom smislu.

Jednostavan primjer sinergije je detekcija bolesti pomoću simptoma. Za neke bolesti, ako se određeni simptomi pojave zajedno, iskusan liječnik će bez ikakvih laboratorijskih nalaza odrediti o kojoj se bolesti radi na temelju vlastitog iskustva.

Metoda potencijala omogućava uvažavanje sinergije ili na način da ju procjenjuju ljudi ili na način da proizlazi iz samih podataka. U kontekstu strojnog učenja metoda potencijala koristi operator samoprocjene (adaptacije). Detalje ostavljamo za neku drugu priliku.

4.1 Ocjenjivanje učenika.

U tablici 7 dane su ocjene učenika na ispitu iz `Fizike, Tjelesnog` i `Muzičkog`, a znak '*' znači da učenik nije ocijenjen iz odgovarajućeg predmeta. Predmeti

Tablica 7: Ocjene učenika na završnom ispitu.

	Fizika	Tjelesni	Muzički
Ivica	3	5	4
Marica	*	1	3
Trsek	2	*	2
Grozdek	4	3	*

učenja su *atributi*, a učenici su *alternative*.

Pitanje koje se postavlja je možemo li na temelju tih podataka rangirati učenike po uspješnosti? Takvo pitanje je nametnuto civilizacijskim dostignućem ili, da budem sarkastičan, nametnutim ponašanjem koje nam nudi naš obrazovni sustav.

4.1.1 Kompozitni tok tablice

Za sada nije preciziran nikakav odnos između atributa i alternativa, ali ako odlučimo rangirati učenike po uspješnosti tada treba precizirati važnost atributa. U ovom slučaju pretpostavljamo da predmeti podučavanja imaju podjednaku važnost jer bi se u suprotnom neki profesori mogli uvrijediti. Primijetimo da je tablica ocjena nepotpuna i kao takva nije pogodna kao input za sve do sada spominjane metode bez nekih dodatnih friziranja ulaznih podataka.

Kompozitni tok tablice nastaje agregacijom tokova za svaki predmet, tj. toka $\mathcal{F}_F$ za fiziku , toka $\mathcal{F}_T$ za tjelesni i toka $\mathcal{F}_M$ za muzički. Svaki od tih tokova ima iste čvorove, a to su alternative (učenici) koje ćemo jednostavnosti radi označiti s I, M, T, G. Lukovi toka $\mathcal{F}_F$ su: $\{I \xleftarrow{1} T, G \xleftarrow{1} I, G \xleftarrow{2} T\}$, lukovi toka $\mathcal{F}_T$ su: $\{I \xleftarrow{4} M, I \xleftarrow{2} G, G \xleftarrow{2} M\}$, a lukovi toka $\mathcal{F}_M$ su: $\{I \xleftarrow{1} M, I \xleftarrow{2} T, M \xleftarrow{1} T\}$. Agregirani graf ima vrhove I, M, T, G a njegove lukove čini unija svih lukova od atributnih tokova. Taj kompozitni graf ima paralelne lukove. Postoje tri načina kako analizirati taj graf: (1) kao multigraf u kojem paralelni lukovi ostaju, (2) paralelne lukove zamijeniti jednim koji ima intenzitet jednak **sumi** paralelnih lukova i (3) paralelne lukove zamijeniti jednim koji ima intenzitet jednak **srednjoj vrijednosti** paralelnih lukova. Svaka od navedenih metoda ima za i protiv, što nadilazi ciljeve ovog teksta. Mi ćemo uzetu sumu.

Rješenje Laplaceove jednadžbe dobivenog kompozitnog grafa (4) daje potencijal prikazan u tablici 8

Tablica 8: Rangiranje učenika iz tablice 7 na temelju jednako važnih atributa.

Učenici	Ivica	Grozdek	Trsek	Marica
potencijal	2.25	0.75	−1.5	−1.5
težina	0.665688	0.235356	0.0494776	0.0494776

4.1.2 Samoprocjena

A što ako odlučimo rangirati predmete na temelju ocjena (mjerenja) konkretnih učenika? To je jednako moguće, ali koji je smisao toga? Takav postupak bi se mogao interpretirati kao rangiranje *atributa* (predmeta) po stupnju ovladavanja gradivom. Što veća vrijednost na rang listi to je gradivo bolje savladano.

Ako *atribute* shvatimo kao primarne onda smo u stanju 'mjeriti' alternative, a atribut ovdje ima ulogu metra kojim mjerimo učenika. Ako *alternative* shvatimo kao primarne onda baždarimo metre (*atribute*) na konkretnim objektima mjerenja.

U svakom od spomenutih postupaka (mjerenje i baždarenje) potrebno je unaprijed zadati važnost (težinu) *atributa* odnosno *alternativa*. Mjerenje ćemo nazvati *primarnim*, a baždarenje *dualnim* postupkom.

Postoji i treća mogućnost kako iskoristiti podatke u tablici, a ta je da provedemo primarni postupak, za neke početne vrijednosti težina atributa, izračunamo težine alternativa i provedemo dualni postupak. Zatim ponovimo primarni postupak s novim, promijenjenim težinama, ponovimo dualni postupak i tako dalje. Osnovni korak u toj proceduri, a to je jedan primarni i jedan dualni postupak, nazivamo *operatorom samoprocjene*. To je adaptacija o kojoj smo govorili u početku.

Može se dokazati[15] da operator samoprocjene ima fiksnu točku a to su vrijednosti težina atributa, koje ostaju nepromijenjene prilikom samoprocjene. Te parametre interpretiramo kao važnost *atributa* i *alternativa* koje tablica nosi u sebi. Za podatke iz tablice 7 fiksna točka operatora samoprocjene je:

Muzički	Tjelesni	Fizika
0.339528	0.330731	0.32974

Ivica	Grozdek	Marica	Trsek
0.264151	0.261712	0.240225	0.233911

Prema rezultatu samoprocjene ispada da je `Muzički`, posmatrajući s gledišta ove grupe učenika, bolje ovladan od drugih predmeta. Velika razlika u težinama nastavnih predmeta ukazivala bi na problem čije uzroke treba tražiti van ove tablice.

4.2 Ordinalno rangiranje s metodom potencijala

Za razliku od drugih, metoda potencijala kao input koristi intenzitet preferencije što je u slučaju tablice 7 razlika ocjena. Ordinalnost, kao pristup, zaboravlja intenzitet preferencije i uvažava samo prioritet. Jedan od načina da se to uradi je taj da se svi intenziteti preferencije u kompozitnom grafu normiraju na 1. Ov-

Tablica 9: Ordinalno rangiranje učenika iz tablice 7 na temelju jednako važnih atributa.

Učenici	Ivica	Grozdek	Marica	Trsek
potencijal	3.75	1.25	−1.25	−3.75
težina	0.824028	0.145669	0.0257509	0.00455215

dje je to učinjeno na način da je kompozitni graf dobiven sumacijom, a zatim su svi intenziteti preferencije normirani na 1. Razlog tome je što normiranjem toka

[15] Čaklović, L. (2011). *Conflict Resolution. Risk-As-Feelings Hypothesis.* Labsi Working Papers, (35):1–16. (http://www.labsi.org/wp/labsi35.pdf).

prije računanja kompozitnog grafa može dovesti do situacije da se u kompozitnom grafu dokinu intenziteti na 0. U tom smislu ordinalnost metode potencijala, što se analize tablice tiče, je 'kontrolirana'.

Posljedica toga je rangiranje učenika dano u tablici 9. U usporedbi s tablicom 8 vidimo da ordinalno rangiranje daje prednost Marici o odnosu na Trseka. Ostali odnosu na rang-listi ostaju nepromijenjeni.

5 Umjesto zaključka

5.1 Agregacija

Agregacija ocjena nije tako jednostavna kako se čini. Aritmetička sredina nije preporučljiva ako se ocjena shvaća kao podatak na ordinalnoj skali jer razlika sredina nije invarijantna na rastuću transformaciju skale. Dodatni problem je što, ako se i računa srednja ocjena kao aritmetička sredina, nakon agregacije obrazovni sustav 'vraća' (zaokružuje) dobivenu vrijednost natrag na skalu, napr. 3.51 na 4 i daje ocjenu *vrlo dobar*. Takav postupak ničim nije opravdan osim uzrečicom: "Svi tako rade". Takvo ocjenjivanje je stresno za učenike, njihove roditelje i nastavnike. Jedan od načina je prekinuti s takvim načinom što bi se moglo postići reorganizacijom obrazovne prakse. Obrazovna praksa nije uzrok, ona je posljedica dubljih nedorečenosti u samom društvu. Osnovno pitanje je što društvo očekuje od obrazovanog pojedinca i je li obrazovanje koje ono nudi tom pojedincu poticajno ili ne za njegov osobni uzlet?

Vidjeli smo da je cilj metode agregacije donošenje rang-liste. Rang-lista se odnosi na grupu učenika koja je obuhvaćena ocjenjivanjem (razred, škola, država), a problem koji mi se čini otvorenim je kako interpretirati te rang liste u obrazovnom sustavu. To otvara drugo, smjelije pitanje, je li ocjena nužna i čemu zapravo služi. Ako je ona osnova za upis i viši razred i daljnje školovanje onda je za takav nastavak školovanja bitno predznanje, a ne ocjena. Predznanje se u principu odnosi na tematske jedinice i za nastavak školovanja treba provjeriti je li učenik ovladao tim tematskim jedinicama. Na primjer, ako je obrazovni cilj da učenik razumije i zna računati s postotcima, i ako to nije savladao onda nije 'zaslužio' da upiše srednju školu ili fakultet dok to ne savlada. Takvo viđenje stvari traži totalnu reviziju obrazovnog sustava i njegovog razumijevanja. To nije problem raznih kurikula već organizacije školstva.

Večina metoda koje smo prezentirali zadovoljava *aksiom o irelevantnoj alternativi*. To znači da u agregaciji učenikovih ocjena sudjeluju samo njegove ocjene, a ne i ocjene drugih učenika. Takav zahtjev odudara od prirode procedure donošenja individualne ocjene učeniku jer nastavnik uvijek uspoređuje učenike u parovima, pa čak i kad ispravlja test ili domaću zadaću. I procedura bodovanja zadatka na testu nije nikada fiksirana unaprijed jer ispravljač revidira svoje viđenje tokom ispravljanja testa.

5.2 Metoda potencijala

Metoda potencijala, za razliku od drugih metoda, uvažava i nepotpune podatke. Konkretno, ako neki učenik nije dobio ocjenu na jednom od testova ne mora mu se dati 0 bodova (ili ocjena 1) na tom testu. Podaci se mogu uvažiti i analizirati takvi kakvi jesu. To vrijedi i za ordinalno rangiranje također.

Osnova metode je uspoređivanje u parovima, a to je zahtjevna procedura i kod prikupljanja podataka i za numeričku obradu. Može se raditi i bez popratnog sofvera ali tada zahtijeva znanje iz linearne algebre. U današnjem, informatički rastućem okruženju to više ne bi trebao biti problem.

Svi izračuni, što se metode potencijala tiče, urađeni su pomoću softvera koji je u završnoj fazi izrade ali nije još za komercijalnu upotrebu. Neki dijelovi tog sofvera bit će dostupni akademskoj zajednici.

Ranije smo govorili o socijalnim aksiomima koji se mogu zahtijevati za metodu. Neki od tih aksioma diskutirani su u članku Čaklović, L. and Kurdija, A. S. (2017). *A universal voting system based on the Potential Method.* European Journal of Operational Research, 259:677–688. Možda nije na odmet napomenuti da su Bordina metoda i Condorcetova metoda specijalni slučajevi metode potencijala što je također dokazano u gore spomenutom članku.

5.3 Kako vidim budućnost ocjenjivanja

Vratimo se natrag na pitanje i problem donošenja individualne ocjene kao mjeru znanja i ovladavanja raznih umijeća, a za koji smo u samom početku rekli da nije tema ove rasprave. Ovdje pod ocjenom ne mislim na broj, to može biti logička vrijednost ili skup takvih vrijednost, procjena i samoprocjena izražena nekim kategorijskim varijablama, ili štošta drugo. Neka radni termin za te podatke bude *profil učenika.* Za takvo opisivanje učenika treba precizirana procedura i trening za kreatore takvog profila. Za donošenje prosudbe o tome je li učenik smije zakoračiti u viši stupanj obrazovnog sustava nije više zaslužan niz brojeva nego skup podataka. Takav skup podataka trebao bi biti dinamičan u smislu da se može revidirati kad je učenik na to spreman, a sustav mu to treba omogućiti.

Analiza takvih profila svakako treba biti hijerarhijska i višekriterijska jer su podaci kvalitativno različiti. Metoda potencijala i hijerarhijski modeli regresije trebale bi biti u stanju analizirati ih.

Jedna od uloga današnje ocjene, bar što se tiče upisa na fakultet, ima cilj razlikovanja učenika, ali je inflacija petica ukinula takvu tu njenu zadaću. Gore opisani pristup profiliranja učenika ne samo da onemogućava inflaciju petica već ukida i ocjenu kao mjeru usvojenog znanja i vještina. Nastavak školovanja omogućava učenikov profil, a ne ocjena. Treba znati rukovati s profilima, a ne ocjenama.