

Általános ember- és problémafüggetlen igazságossági mértéket megállapító algoritmus beosztásokra

Rónai-Kovács Martin

2022

Kivonat

Az alábbiakban egy általános ember- és problémafüggetlen igazságosságmegállapítási algoritmust határozzunk meg. Az ezzel kapcsolatos legnagyobb probléma, hogy a legtöbb beosztás esetén egyéni algoritmusok készítenek beosztásokat, egyéni be- és kimenetekre.

Megállapítunk általánosítási eljárásokat és fogalmakat, és ezeket használva definiáljuk a hasonlóságot bemenetek között és az igazságosságot, mint ezen hasonlóság megtartását egy beosztás elkészítése során¹.

Algoritmusunk eredményeként olyan mértéket kapunk a $[0; 1]$ intervallumon, mellyel összehasonlíthatunk két elkészült beosztást, melyek azonos kiinduló adathalmazból származnak.

1. Bevezetés

A beosztási problémák megoldása, lévén egy NP-teljes probléma, nagy körütekintést, hatékonyságot, és igen gyakran problémára specifikált megoldási módszert igényel. Emiatt kényszerűen el kell fogadnunk egy-egy elkészült beosztásról, hogy annál jobbat nem fogunk találni, megfelel a pszeudo-tökéletes eredmény is.

Ezzel az esetek nagy részében ki is lehet egyezni, azonban fontos kérdés, hogy milyen szempontból nem felel meg az adott beosztási eredmény. Egyes beosztásokban tapasztalható, hogy a beosztó algoritmus olyan szabályokat és egyszerűsítéseket tanul meg a már létező mintákból, amit mi nem szeretnénk figyelembe venni, például nemet, rasszot, kort, párkapcsolati státuszt, márkát, vagy bármi hasonlót. Ezt szeretnénk észrevenni.

A beosztási probléma sokrétűsége miatt azonban ennek minden problémára születhet egyéni meghatározása és kimutatási algoritmus, pont ugyanúgy, ahogy a beosztási algoritmus is általában egy-egy problémára születik meg.

Ezt szeretnénk mi áthidalni azzal, hogy megállapítsunk egy általánosított eljárást, amivel az igazságosság (amiről fentebb esett szó) mértéke megállapítható egy tetszőleges beosztási problémára készült beosztásról.

A módszer megalkotása során következő célokat határozzuk meg:

1. A módszer legyen problémafüggetlen, ideértve a problémához tartozó be- és kimenetek szemantikáját és formátumát.
2. A módszer legyen algoritmusfüggetlen, tehát ne vegye figyelembe a kapott be- és kimenet közti logikai kapcsolatot.
3. A módszer a lehető legkevesebb emberi beavatkozást igényelje. Ez lényegében azért lesz fontos szempont, mert feltételezzük az igazságosság objektív létét, illetve el szeretnénk kerülni, hogy a megválaszolatlan kérdéseket problémafüggő megoldásokkal helyettesítsük.
4. A módszer eredménye (kimenete) legyen szám-szerű; hiszen így állapítható meg, hogy egyes beosztások igazságosabbak, mint mások.

Amit mi keresünk, az egy $f(\{I_S, O_S\}) \in \mathbb{R}$ függvény, ahol $\{I_S, O_S\}$ egy tetszőleges beosztási probléma be- és kimenete. Ezen f megalkotásakor kell a fenti szempontokat megvalósítanunk.

2. Beosztások közti reláció

2.1. Definíció (Beosztás). A beosztás egy tetszőleges beosztási probléma egy bemenete és az arra (egy tetszőleges beosztáskészítő által) adott kimenet.

$$S = \{I_S, O_S\}, \quad (1)$$

ahol I_S az S beosztás bemenete, míg O_S az S beosztás kimenete.

Tegyük fel, hogy a beosztásainkat elkészítő folyamat egy σ függvény, tehát $\sigma(I_S) = O_S$. Logikus lehet

¹A „Fairness Through Awareness” definíció szerint ez az igazságosság

feltételezni azt, hogy eltérő bemenetekhez nem minden σ esetén lehet ugyanazt a kimenetet rendelni.

2.1. Tétel. Létezik olyan beosztó, ami nem képes azonos beosztást készíteni semmilyen két különböző bemenetre.

$$\exists \sigma : \forall I_S \neq I_\Sigma : \sigma(I_S) \neq \sigma(I_\Sigma) \quad (2)$$

Ennek jelentősége, hogy a beosztásokban az igazságosságot a kimenetek is szabják; ha egy kimenet eleve nem lehet elég megfelelő (mert olyan „szerencsétlen” bemenetbe tartozik), akkor a hozzá tartozó igazságossági mérték sem az adott beosztásról ad nekünk információt: előfordulhat, hogy egy megadott bemenetbe a lehető legoptimálisabb beosztás igazságossága is rosszabb, mint egy másik bemenetbe tartozó legrosszabb beosztásé.

Ennek következtében az igazságosságot nem feltétlenül mérhetjük össze két különböző bemenetnél.

2.2. Definíció (Izomorfia). Két beosztás pontosan akkor izomorf, ha azonos bemenetük van.

$$S \cong \Sigma \iff I_S = I_\Sigma \quad (3)$$

Megjegyzés. Két beosztás pontosan akkor izomorf, ha a beosztási algoritmus átalakításával elérhető, hogy a kimenetek azonosak legyenek.

Ezután a 2.1 tétel alapján következik az alábbi.

2.2. Tétel. Két beosztás igazságossága között csakis akkor létezik reláció, ha a beosztások izomorfak.

$$\exists \theta \in \{<, >, =\} : f(S) \theta f(\Sigma) \iff S \cong \Sigma \quad (4)$$

Azaz a trihotómia nem teljesül bármely két beosztás igazságosságára.

3. Általánosítás

Határozzuk meg pontosabban a bemenetet. Egy beosztás általában entitások egymáshoz rendeléseként fogható fel, talajdonképpen az előre ismert tulajdonságokhoz állapítunk meg relációkat és az ezekhez tartozó egyéb tulajdonságokat tudjuk még kiszámolni.

Példa: CPU-magok és feladatok esetén az entitásaink a magok és a feladatok. A kimenetünk az, hogy melyik feladatot melyik mag végzi el, illetve hogy melyik maghoz hány feladat tartozik (ez utóbbi egy számított tulajdonság). Kiszámítható még, hogy melyik magnak mekkora az összerhelése, de itt már számíthatunk idővel vagy feladatszámokkal is (Ez még problémákhoz vezet, lásd később).

Azonban a bemenetek leírása egyszerű: minden entitásunkat egy vektorként kell értelmezzük, melynek

elemei a paraméterértékek. Mivel többféle vektorra lehet szükség (mert egy CPU-magot nem ugyanúgy írunk le, mint egy feladatot), ezért a vektorok mérete eltérő.

Először határozzuk meg a beosztási problémát. Minden problémában entitások (beosztandók) és követelmények szerepelnek. Így adódik a következő egyszerű leírás:

3.1. Definíció (Beosztási probléma). Egy beosztási probléma definíciója a példányok leírásainak halmaza és a követelmények halmaza:

$$D_S = \{E_S, C_S\}, \quad (5)$$

ahol E_S a probléma entitásainak leírásait (entitásosztályok), C_S pedig a probléma követelményeit tartalmazza.

Megjegyzés. E_S és C_S pontos formájával nem kell foglalkoznunk, mert nem használjuk őket (ezek a σ függvény definiálásához kellenek); elég tudnunk, hogy E_S az entitásosztályok egy *halmaza*.

Megjegyzés. Az entitások itt nem nyers adatként jelennek meg. Úgy kell őket elképzelnünk, mint egy paraméterszámítási *függvény* eredményét ($p(r) = x$), ami a nyers adatokból releváns információkat nyer ki számunkra. Erre olyan esetek miatt van szükség, mint például záróvizsgabeosztásnál, ahol mind egy, mely hallgatókat vizsgáztatja egy oktató, a kérdés az, hány hallgatót vizsgáztat egy oktató.

3.2. Definíció (Parametrizációs függvény). A nyers adatokból igazságosság szempontjából releváns adatokat kinyerő transzformációs függvény neve parametrizációs függvény. Jele: p

$$p(r) = x, \quad (6)$$

ahol r az entitás nyers formája, x a számunkra fontos adatokat tartalmazó forma.

3.3. Definíció (Csoport). Azon entitások, melyek egy osztályba tartoznak, egy csoportot alkotnak. Egy $e \in E_S$ osztályhoz tartozó csoportot a következőképpen definiálunk:

$$G(e) = \bigcup_{x::e} x \quad (7)$$

Ahol az $x :: e$ jelentése, hogy x az e osztályleírásnak megfelel.

Példa: Órarendtervezéskor egy entitásosztály a *tanár*, és ennek egy példánya *Szeszlér Dávid*. *Szeszlér Dávid* benne van a *tanárok* csoportjában.

$$\text{Szeszlér} :: \text{tanár} \in E_{\text{órarend}}$$

$$\text{Szeszlér} \in G(\text{tanár})$$

3.4. Definíció (Homogenitás). Két entitás homogén egymással, ha azonos osztályhoz tartoznak.

$$x \sim y \iff x :: e \wedge y :: e \quad (8)$$

Az egyes entitások leírása bővül a beosztás (σ) hatatása után, ezért érdemes megkülönböztetni a bemeneti és kimeneti állapotokat.

3.5. Definíció (Be- és kimeneti állapotleírás). Egy x entitás bemeneti leírása x_{in} , kimeneti leírása x_{out} ; hasonlóan a be- és kimeneti osztályleírás e_{in} és e_{out} .

Ezen definíciók ismeretében már meg tudjuk határozni a bemenetet.

3.6. Definíció (Bemenet). A bemenet egy S beosztásra (I_S) megegyezik a csoportok halmazával.

$$I_S = \bigcup_{e \in E_S} G(e_{\text{in}}), \quad (9)$$

tehát I_S azonos méretű vektorok halmazainak halmaza.

Az egyes entitásozstályok külön bemenetként való kezeléséhez határozzuk meg a csoportonkénti bemenetet:

3.7. Definíció (Csoportonkénti bemenet). Az e entitásozstályhoz tartozó csoportonkénti bemenet $I_{S,e}$ megegyezik e_{in} csoportjával.

$$I_{S,e} = G(e_{\text{in}}) \quad (10)$$

Tehát meghatároztuk a beosztási problémánk bemenetének leírását. Tételezzük fel, hogy a kimenetet hasonlóan írhatjuk le, tehát a fenti definíciókban I_S felcserélhető O_S -sel, amennyiben az e_{in} -t is felcseréljük e_{out} -tal, stb.

3.1. Tétel (Normalizáció I.). Minden bemenetnek létezik normáltja, vagyis

$$\forall I_{S,e} : \exists \| I_{S,e} \| \quad (11)$$

Bizonyítás. Használjuk az alábbi lemmát:

3.2. Lemma (Entitás-vektor azonosság). Minden entitás leírható valós vektorként, azaz $x = \underline{x} \in \mathbb{R}^n$, ahol n az x entitás paramétereinek száma (ideértve a számított paramétereket is).

$\| I_{S,e} \|$ azonos a benne lévő entitások normáltjaival, hiszen azok egymástól függetlenek². Mivel ezek valós vektorok \mathbb{R}^n -ben (3.2), létezik normáltjuk, és így $I_{S,e}$ -nek is létezik normáltja. \square

²Ezt egyelőre feltételezzük, később foglalkozunk még ezzel

Megjegyzés. A tétel fontossága, hogy tetszőleges beosztási problémánál a bemenetet általánosítani tudjuk.

Következmény. Csakis akkor tudunk általánosítani, ha a bemeneti vektorok normálhatók.

4. Paraméterek függése

Az eddigiek alapján szeretnénk elérni, hogy az igazságossági függvény (f) bemenetei általánosak legyenek, és ehhez meghatároztuk a befogadható formákat (3.1). A teljes általánosításhoz azonban el kell végeznünk a normalizációt is, hogy tényleg tetszőleges entitásokkal tudjunk dolgozni.

A probléma itt az, hogy meg kell határoznunk, hogyan lehetne tulajdonságvesztés nélkül, a paraméterek egymáshoz való relációjának megtartásával együtt módosítani az entitásokat.

Mivel entitásozstályaink különbözőek, tudjuk, hogy

$$\| I_S \| = \bigcup_{e \in E_S} \| I_{S,e} \|, \quad (12)$$

így a teljes bemenetet tudjuk normálni, amennyiben az csoportokat egyesével tudjuk normálni. Ezt meg tudjuk tenni, mivel a csoport entitásai egymástól függetlenek (ahogy azt a (3.1) tétel bizonyításakor feltételeztük).

Megjegyzés. Logikusnak tűnhet, hogy (mivel valós vektorokról van szó) ezeket egész egyszerűen euklideszi hosszúsággal leoszszuk:

$$\| \underline{v} \| = \frac{1}{|\underline{v}|} \underline{v}$$

így kapva egy egység hosszú vektort tetszőleges \mathbb{R}^n térben. Azonban ezzel elkövethetjük azt a hibát, hogy két relatíve távoli pont távolságát 0-ra csökkentjük.

Példa: $\underline{u} = (0, 1, 1)$; $\underline{v} = (1, 1, 1)$; $\underline{w} = (10, 10, 10)$ vektorok esetén normálás során \underline{v} és \underline{w} távolsága 15.588-ról 0-ra csökken, míg \underline{v} és \underline{u} távolsága 1-ről 0.605-re csökken; tehát a $\underline{v} - \underline{u}$ és $\underline{v} - \underline{w}$ közti reláció megfordul, ha ezeket egyesével normáljuk (jelen példában feltételeztük, hogy két vektor távolsága megegyezik azok különbségével).

A példában látható jelenséget szeretnénk elkerülni, így máshogyan kell a normalizációt definiálnunk.

4.1. Definíció (Normalizáció). Egy \underline{x} entitás normáltja $\nu(\underline{x})$ (vagy $\| \underline{x} \|$). A normálás az azonos paraméterek közti relációt megtartja, de a vektort véges térbe transzformálja.

$$\nu : \mathbb{R}^n \rightarrow [r_1; r_2]^n, \quad (13)$$

ahol $r_1, r_2 \in \mathbb{R}$ és $r_1 < r_2$.

Megjegyzés. Az azonos paraméterek közti relációt meg kell tartanunk, hogy ne forduljon elő a fenti példában látott eset (reláció megfordulása). További előny, hogy a paraméter normáltja ez esetben függeni fog a többi entitás értékskálájától, de nem függ a többi paramétértől.

Megjegyzés. A továbbiakban $r_1 = 0$ és $r_2 = 1$ értékekkel használjuk a normalizáció fogalmát.

Hogy a (4.1) definíciónak megfeleljünk, feltehetjük, hogy a paramétereket egyenként normálhatjuk. Ennek azonban van egy feltétele, ami az alábbi tételben olvasható.

4.1. Tétel (Normalizáció II.). Amennyiben az entitás paraméterei függetlenek, az entitás normáltja megegyezik az entitás paramétereinek normáltjával:

$$\forall i \neq j : \underline{x}_i \not\propto \underline{x}_j \implies \|\underline{x}\|_i = \|\underline{x}\|_j \quad (14)$$

4.2. Tétel (Visszaállíthatóság). ν reverzibilis, $\exists \nu^{-1}$.

Bizonyítás. Mivel f csak adott bemenetre ad számszerű értéket (2.2), ismerjük a teljes lehetséges érték-készletet minden paraméterre. Ebben az esetben (tehát mindig) belátható, hogy

$$\nu_i(x) : [r_1; r_2] \rightarrow [0; 1],$$

ahol $r_1 = \arg_x \min(x_i)$ és $r_2 = \arg_x \max(x_i)$. Tudjuk, hogy minden entításra igaz, hogy azonos paramétereik között a relációs viszony megmarad (4.1), így ν_i^{-1} maga is megfelel a normalizáció fogalmának. Ekkor könnyen választhatunk egy olyan érték-készletet, amely az eredeti paraméter értelmezési tartománya. Ezzel megalkottuk ν inverzét. (És megállapítottuk, hogy ν bijektív). \square

Következmény. Ezzel beláttuk, hogy normalizáció során nem történik információvesztés.

5. Normalizációs problémák

A normalizációval beláttuk, hogy tetszőleges beosztási probléma (mely leírható a (3.1) formában) általánosítható.

Ennek azonban feltétele volt, hogy az osztályok, az entitások, és azok paraméterei egymástól függetlenek legyenek. Ez főként a p függvény (3.1) megalkotásában jelentős szempont.

A másik aggály, ami felmerülhet, hogy (véges pontosságú eszközök miatt) egyes adatok irrelevánssá válhatnak a normalizáció során. Például lehetnek olyan paraméterek, melyeket nem lineárisan, hanem logaritmikusan kell normálni, mivel maguk a paraméterek exponenciálisak. Ez a probléma elkerülhető, ha a

normalizációs (ν_i) függvényeket minden paraméterre körültekintően választjuk meg. Ez talán még automatizálható is regressziókkal.

Ha a normalizáció során elérjük, hogy a skálán elhelyezkedő értékek ugyanolyan eloszlásúak legyenek (lehetőleg lineárisak), akkor tetszőleges paraméter *tekinthető* lineárisnak. Ennek jelentősége a továbbiakban jelentős, hiszen ezzel elérjük, hogy két entitás távolságának számításakor a paraméterek homogének legyenek eloszlás szempontjából (tehát ne kelljen minden paraméterre meghatározni egy „távolságbefolyásolási tényezőt”, vagy hasonlót).

6. Távolságok

A távolságok nagy szerepet játszanak az igazságosság megállapításában, hiszen a beosztás előtti és utáni távolságok közti eltérésből deríthetjük ki a legszembe-tűnőbb anomáliákat.

A távolság felfogható ebben a problémában úgy, mint két entitás „hasonlótlanosságának” mértéke.

Példa: Ha csak 1GHz-es CPU-ink vannak, rendszerben van, hogy mind ugyanannyi feladatot kap; de ha van mellette egy 10GHz-es is, akkor az egyenlő szétosztás nem igazságos. Ez azért van, mert az 1GHz-es magok távolsága 0, de a 10GHz-es magtól való távolságuk 9(GHz).

6.1. Tétel (Állapotfüggetlen paraméterszámosság, ÁFPSZ). Egy $x :: e$ entitás x_{in} és x_{out} állapotai leírhatók azonos hosszúságú vektorokkal mindenféle torzulás nélkül, azaz

$$|x_{\text{in}}| = |x_{\text{out}}| = |e| \quad (15)$$

Ezt úgy tehetjük meg, hogy x_{in} valódi paraméterei mellé csatolunk $\dim(x_{\text{out}}) - \dim(x_{\text{in}})$ darab (konstans értékű) „álpamétert”.

6.1. Definíció (Résztávolság). A résztávolság egy $x \sim y :: e$ entitáspár azonos paramétereinek távolsága, azaz abszolút különbsége.

$$d_i(x, y) = |x_i - y_i| \quad (16)$$

Ez vonatkozik a példányok mindkét állapotára.

Megjegyzés. Amennyiben az entitásaink normáltak, $d_i(x, y) \in [0; 1]$ minden párra.

6.2. Definíció (Távolságfüggvény). Az igazságossági távolságfüggvény két azonosan paraméterezett entitásról ad olyan arányszámot, melyből azoknak különbözősége megállapítható. A beosztásban akkor állapíthatunk meg igazságtalanságot vagy igazságossági anomáliát, ha a be- és kimeneten mért távolságok

jelentősen eltérnek.

$$d : \mathbb{R}^{2|e|} \rightarrow \mathbb{R} \quad (17)$$

$$d(x, y) \propto \bigcup_{i=1}^{|e|} d_i(x, y) \quad (18)$$

6.2. Tétel (Bemeneti távolság megmaradása). Az ÁFPSZ használata nem befolyásolja a távolságokat.

Bizonyítás. Tekintsük az alábbi eseteket:

- (A) ÁFPSZ esetén a kimeneti vektorok nem módosulnak, így a távolság sem változik.
- (B) A bemeneti vektorokon eredetileg szereplő paraméterek távolsága nem változik.
- (C) A bemeneti vektoron megjelenő paraméterek egy konstans c értéket vesznek fel, tehát

$$d_i(x, y) = |c - c| = 0 \quad (19)$$

Tehát a megjelenő paraméterek távolsága nem módosítja az entitások távolságát, ameddig nem szorzatként definiáljuk a d függvényt. \square

Mivel a kimeneten több paraméter van, melyek eltérhetnek a bemeneti értéktől, feltehetjük, hogy

$$d_{in}(x, y) \leq d_{out}(x, y),$$

ahol $x \sim y$ példányok.

Ugyanezt megfogalmazhatjuk egyszerűbben, ha a be és kimeneten ugyanaz a függvény méri a távolságot:

$$d(x_{in}, y_{in}) \leq d(x_{out}, y_{out}) \quad (20)$$

Állapítsuk meg, miért fontos ez számunkra. Igazságosnak akkor tekintünk egy beosztást, ha a beosztás elkészítése során nem torzítjuk azokat a távolságokat, melyek már a bemeneten is léteznek. Ezt felfoghatjuk úgy is, mint a Lipschitz-tulajdonság betartását (valamely M konstansra):

$$\forall x \sim y : d(x_{in}, y_{in}) \leq M d(x_{out}, y_{out}) \quad (21)$$

Egy beosztástól azt szeretnénk, hogy a kimenetek távolságai arányosságban álljanak a bemeneti távolságokkal, tehát legyen igaz, hogy

$$d(x_{out}, y_{out}) \propto d(x_{in}, y_{in}). \quad (22)$$

6.3. Definíció (különbségváltozás, δ). Két egymással homogén entitás különbségváltozása megegyezik a bemeneti és kimeneti különbségük közti eltéréssel. Jele: δ .

$$\delta(x, y) = |d(x_{out}, y_{out}) - d(x_{in}, y_{in})| \quad (23)$$

Az eddigiekből következik, hogy a fairness-függvény értéke az egyes párok δ -jából lesz számítva; ezt az átszámítást végezzük esetünkben egy u_e (unfairness) függvény minden e osztályra. Ekkor adódik, hogy

$$f(S) \propto \bigcup_{e \in E_S} \left(f_e(S) \propto -u_e(S) \propto \bigcup_{x \sim y} \delta(x, y) \right) \quad (24)$$

ahol $x :: e$.

Az eddig megállapítottakat kell a fairness-függvény megalkotásakor figyelembe vennünk.

Az $u_e(S)$ unfairness függvényre könnyen adhatunk egy használható és kényelmes definíciót: határozzuk meg $u_e(S)$ az egyes „igazságtalanságok” összegeként:

6.4. Definíció (Osztályonkénti igazságtalanság).

$$u_e(S) = \sum_{x \sim y} \delta(x, y), \quad (25)$$

ahol $x :: e$ és $x \neq y$.

Ezután kényelmes $u(S)$ -t is meghatározni, mint az egyes osztályok igazságtalanságának összegét:

6.5. Definíció (Beosztás igazságtalansága).

$$u(S) = \sum_{e \in E_S} u_e(S) \quad (26)$$

Ezen definíciók használhatók f megalkotásához közvetlenül, amennyiben ismerjük értékkészleteiket. Mivel d_i egy tetszőleges entitás esetén is a $[0; 1]$ intervallumon mozog (4.1, 6.1), tudjuk, hogy d értéke véges. Ebből következik, hogy létezik egy \hat{d}_e maximális távolság minden $e \in E_S$ osztályban. δ definíciójából következik, hogy

$$\hat{\delta}_e = \hat{d}_e,$$

így meg tudjuk állapítani $u_e(S)$ szélsőértékeit (tudva, hogy minden δ egyszer szerepel és $\delta(x, x) = 0$):

$$\hat{u}_e(S) = \sum_{x \sim y} \hat{\delta}_e = \binom{|G(e)|}{2} \cdot \hat{\delta}_e \quad (27)$$

Ezután $\hat{u}(S)$ értéke:

$$\hat{u}(S) = \sum_{e \in E_S} \hat{u}_e(S) \quad (28)$$

Mivel minden S beosztásra $u(S) \in [0; \hat{u}(S)]$ (hiszen $\check{u}(S) = 0$), belátható az alábbi tény:

6.3. Tétel (Normált igazságtalanság). Egy S beosztás igazságtalanságának létezik normalizációja:

$$\exists \nu \circ u(S) : [0; \hat{u}(S)] \rightarrow [0; 1] \quad (29)$$

Az eddigiekből f egyszerűen definiálható:

6.6. Definíció (Általános FTA³ igazságossági mérték). Tetszőleges S beosztás igazságossági mértékét $f(S)$ adja, az alábbi formula szerint:

$$f(S) = 1 - \nu \circ u(S) \quad (30)$$

Így tehát van egy definíciónk f -re, mely nem használja d pontos meghatározását.

6.7. Definíció (Fairness-függvény). Általános, a 3.1 definíciónak megfelelő probléma igazságosságát az f függvény adja, ahol

$$f : \bigcup_{e \in E_S} \mathbb{R}^{|G(e)| \times |e|} \rightarrow [0; 1] \quad (31)$$

Ennek értéke:

$$f(S) = 1 - \frac{1}{\hat{u}(S)} \cdot \sum_{e \in E_S} \sum_{x \sim y} \delta(x, y) \quad (32)$$

A képlet (27) és (28) helyességén alapul, és feltételezi, hogy

$$\nu(u) = \frac{u - \tilde{u}}{\hat{u} - \tilde{u}} \quad (33)$$

Ez egy olyan skálán vehet fel értékeket, ahol a 0 a tökéletesen igazságtalan, míg 1 a tökéletesen igazságos beosztást jelenti⁴.

7. Súlyozás

Felmerülhetnek az igazságossággal kapcsolatban további kérdések, melyek megválaszolásával optimálisabb vagy akár gyorsabban számítható eredményeket kaphatunk. Ilyen lehet például, hogy minden entitás közti igazságosság kell-e, vagy vannak-e olyan paraméterek, amiket érzékenyebben kell kezelnünk, mint a többit?

Példa: Időszeletek egyformasága számít egy kész órarendnél?

Példa: A „hitelbírálat elfogadása” és a „rassz” vagy „párkapcsolati státusz” tulajdonság korrelálnak, de a többi tulajdonság nem.

Paraméterek fontosságának kijelentéséhez definiáljuk a súlyokat.

Megjegyzés. Innentől jelölje jobb felső indexben zárójellezett változó az entitás osztályát: $x^{(e)} \Leftrightarrow x :: e$.

³Fairness Through Awareness: beosztás előtti és utáni hasonlóság megtartása azonos az igazságossággal.

⁴Érdemes lehet belegondolni, hogy ezeket az értékeket egy-egy beosztás milyen szélsőséges tulajdonság teljesülése esetén veheti fel (például létezik-e olyan beosztás, ahol 1-nél több entitás is van, de $f(S) = 0$? Vagy van-e olyan igazságos beosztás, ami annyira igazságos, mintha be sem osztanánk az egyes entitásokat?).

7.1. Definíció (Súly). Egy osztály egy példányának $x_i^{(e)}$ paraméteréhez tartozó súlya azt fejezi ki, mennyire befolyásolja a fairséget egy $y \sim x$ példány $y_i^{(e)}$ paraméterétől való eltérés. Ezen súly minden esetben nem negatív értéket vesz fel.

$$\underline{w}^{(e)} \in [0; \infty)^{|e|} \quad (34)$$

7.2. Definíció (Súlyozott entitás). Egy entitás súlyozása a paraméterenkénti szorzat számításából adódik:

$$w^{(e)} \odot x^{(e)} \quad (35)$$

7.3. Definíció (Súlyozott (rész)távolság). A súlyozott távolság megegyezik a paraméterenkénti távolságok súlyokkal vett szorzatából számított távolsággal.

$$w_i d_i(x, y)$$

Megjegyzés. Ha $w^{(e)}$ minden eleme 1, akkor a súlyozás eredeti állapotot hagy, míg ha minden eleme 0, akkor az összes létező beosztást „igazságosra” módosítja.

7.1. Állítás. A távolságonkénti súlyozás helyettesíthető entitássúlyozással.

Bizonyítás.

$$\begin{aligned} d'_i(x, y) &= w_i d_i(x, y) \\ &= w_i |x_i - y_i| \\ &= |w_i x_i - w_i y_i| \\ &= d_i(w \odot x, w \odot y) \end{aligned} \quad (36)$$

□

Megjegyzés. Ennek hasznossága leginkább, hogy $d'(x, y)$ számításakor nem kell elvégeznünk a w -vel való elemenkénti szorzásokat minden (x, y) párra, elég a számítás előtt minden $w \odot x$ értékét kiszámítani.

7.1. Tétel (FTA - FTU⁵ átjárhatóság). Bináris súlyok használatával elérhető, hogy az FTU definíció szerint járjunk el, annak ellenére, hogy az FTA szerint haladt az eddigi gondolatmenetünk.

Ekkor az érdektelen változók súlyát 0-ra, a többi súlyt 1-re kell beállítanunk, így kapva egy f_1 fairness értéket; majd minden súlyt 1-re állítva kapunk egy f_2 értéket. Ezek különbsége mondja meg, hogy az FTU definíció szerint igazságos-e az adott beosztás:

$$|f_1 - f_2| < h_{FTU} \Leftrightarrow S \text{ igazságos,}$$

ahol h_{FTU} egy hibahatár.

⁵Fairness Through Unawareness: Az igazságosság teljesül akkor és csak akkor, ha az érzékeny változók ismeretében a beosztás elvégzése során születő távolságok hasonlóak annak a beosztásnak a távolságaihoz, amit az érzékeny változók „letakarásával” kapunk.

Alapvetően a súlyozás optimalizásra használható, azonban lehetőséget ad arra is, hogy bizonyos információkat „enyhítsünk”, ha azok esetleg redundánsak lennének. Eddig ezt azzal a feltevessel hidaltuk át, hogy az entitásainkat és azok paramétereit is páronként függetlennek tekinthettük, de így korlátoztuk I_S formátumát, ezzel szűkítve a lehetséges σ függvények, illetve a lehetséges D_S leírások halmazát.

Ennek áthidalása tehát kényelmes lenne, ezért vizsgáljuk meg, hogyan segíthetnek a súlyozások.

Paraméterek esetén, ha több is ugyanazt az információt hordozza magában, tehát valamely e -re

$$\forall x :: e : x_i^{(e)} \propto x_j^{(e)},$$

vagy máshogyan kifejezve $e_i \propto e_j$, akkor (hiába lehetséges, hogy eltérően definiálja p ezen értékeket) az információt redukálnunk kell 1 paraméterbe.

Példa: x_1 = foglalt időszávok száma; x_2 = foglalatlan időszávok száma.

7.2. Tétel. Súlyokkal elérhető, hogy a redundáns paraméterek információja ne legyen túlreprezentálva.

Bizonyítás. T.f.h. valamely $R \subset x^{(e)}$ egy olyan paraméterhalmaz, melynek minden eleme ugyanazt az információt hordozza, azaz teljesen redundáns paraméterek. (Ez nem azonos azzal, hogy két paraméter értéke megegyezik!)

Ekkor normálás után $\forall i : R_i \in [0; 1]$, tehát súlyuk a távolságszámításban azonos (figyelembe nem véve a $w^{(e)}$ együtthatókat).

Amit el szeretnénk érni, hogy az R -beli értékek összesen akkora súlyúak legyenek, mint egyetlen paraméter. Ezt (az azonos értékkészlet miatt) megtehetjük úgy, hogy $\frac{1}{|R|}$ együtthatót adunk nekik.

Alapvetően ezt akkor tehetjük meg, ha teljes a redundancia; de ezt feltételezhetjük, mivel minden egyes paraméter (a problémánk szempontjából) elemi információt hordoz (tehát nincs részleges függés). \square

7.4. Definíció (Redundanciaegyüttható). Amennyiben két (vagy több) paraméter egy e osztályban redundáns, redundanciaegyütthatójuk egy olyan érték, mellyel elérjük, hogy összesen akkora értékűek legyenek, mint egyetlen független paraméter.

$R(x_i^{(e)})$ azon paraméterek halmaza, melyek redundánsak $x_i^{(e)}$ -vel.

$$\forall x_i^{(e)} \propto x_j^{(e)} : x_j^{(e)} \in R(x_i^{(e)}) \quad (37)$$

Ezt használva könnyen definiálható a redundanciaegyüttható:

$$r_i^{(e)} = \frac{1}{|R(x_i^{(e)})|} \quad (38)$$

Példa: Ha $R = \{x_1, x_2, x_3\}$ halmaz tartalmazza az összes olyan paramétert, mely x_1 -gyel redundáns, akkor $r_1^{(e)} = \frac{1}{3}$.

7.5. Definíció (összesített együttható, korrigált távolság). Egy e_i paraméter összesített együtthatója $c_i^{(e)}$ megegyezik e_i összes együtthatójának szorzatával. Mi esetünkben ez:

$$c_i^{(e)} = r_i^{(e)} \odot w_i^{(e)} \quad (39)$$

Ezzel újradefiniálható a korrigált (rész)távolság is:

$$d'_i(x, y) = c_i d_i(x, y) \quad (40)$$

8. Súlyozási problémák

Láttuk, hogyan alkalmazható a súlyozás a fairség hatékonyabb megállapításához, de természetesnek vetjük, hogy a súlyok előre meg vannak állapítva. Azonban a kijelölt célok közül az elsőt (eddig tudásunk szerint) megszegi minden olyan eset, ahol létezik $w_i^{(e)} \neq 1$, mivel ez esetben kézzel kell beállítani w értékét (hiszen ez a paraméter „fontosságát” fejezi ki, tehát problémától és szemantikától nem független).

Megoldás lehet, ha találunk algoritmust ezen súlyok automatikus megállapítására, de ekkor elveszhet w eredeti célja. Itt inkább köztes megoldásra lesz szükség.⁶

$r^{(e)}$ meghatározása is hasonló problémákat vet fel, azonban a redundancia megállapítására léteznek módszerek, ez nem (feltétlenül) igényel emberi beavatkozást.

További probléma a redundanciaegyütthatókkal, hogy a fentebb leírtak szerint csak akkor használhatóak, ha a paraméterek elemi információt hordoznak; ami egy újabb követelményünk p megalkotásához.

Végül feltehetjük a kérdést, hogy ha részben meg szerettük volna oldani a függetlenség kérdését, akkor miért nem oldottuk meg teljesen? Jelenleg nincs módszerünk arra, hogy a nem elemi információk közti redundanciát kezeljük. Entitások között fennállhat együttmozgás, ha a reprezentációs modell úgy határozza meg a problémát.

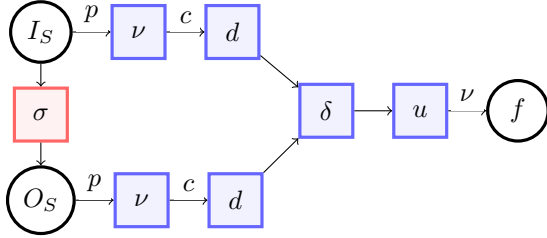
Példa: Zárovizsgabeosztás esetén szerepelhet a modellben *Vizsgáztató*, de nem szerepelhet *Elnök* és *Belső tag*, mert megeshet, hogy ezeket a szerepeket ugyanaz az ember tölti be, így (az utóbbi modellben) effektíve kétszer reprezentálunk egy példányt.

⁶Köztes megoldás lehet, ha véletlenszerűen inicializált értékekkel töltjük fel a vektorokat, kiszámoljuk az így kapott f -et, optimalizáljuk az értékeket egy $\pm\gamma$ -val való növeléssel, és ezt addig ismétljük, míg f értéke maximális lesz. A folyamat leállításakor megtekinthetjük a kiszámolt w értékeit, és megállapítjuk, reálisak-e.

9. Áttekintés

9.1. Teljes folyamat

Az eddig kijelentett fogalmak alapján megalkothatunk egy eljárást, hogy megállapítsuk f értékét tetszőleges beosztásra. Egy absztrakt verziót láthatunk az 1. ábrán.



1. ábra. A fairness (f) megállapítása

A folyamat első lépéseként feltételeztük, hogy a megkapott bemenet I_S -ből és O_S -ből áll, és minden egyes entitás egy p parametrizáló függvény hattatásával születik. A bemenet feldolgozása, azaz a mi algoritmusunk az ábrán kékkel jelölt függvények hattatása, ahol az élek az egyes függvények hattatásának sorrendjét jelzik. A folyamat lépései sorrendben:

1. Normalizáció, így minden paraméter értékkészlete a $[0; 1]$ intervallum.
2. Távolságok meghatározása mind a bemeneti adathalmazon, mind a kimenetén, minden egyes homogén entitáspárra. Itt nem határoztuk meg pontosan, mit értünk távolság alatt, csupán az egyes paraméterekre (tehát a paraméterenkénti távolság és az entitások távolsága közti összefüggést nem definiáltuk pontosan). Opcionálisan végrehajthatunk súlyozást a lépés során (előtt⁷); ennek különböző okai lehetnek (lásd fentebb).
3. Be- és kimeneti szálak összevonása, a különbségváltozások meghatározásával.
4. Egyszerű aggregált értéként megkapjuk u -t az egyes osztályok igazságtalansági mértékéből.
5. Megkapjuk a keresett f értéket.

9.2. A követelmények teljesülése

Vizsgáljuk meg a bevezetésben meghatározott követelmények teljesülését:

⁷Lásd: entitássúlyozás

1-2. Az ismertett folyamat során hagyunk szabad teret mind a problémadefiníciónak, a pontos távolságdefiníciónak, és a beosztási algoritmusnak is. Ezzel az első⁸ és második meghatározott kritériumot teljesítettük. Tegyük hozzá, hogy lehetséges bevonni a paraméterek szemantikáját, ha meghatározunk súlyokat.

3. A harmadik kritérium is teljesül, ha és amennyiben nem használunk súlyozást, vagy automatizált módszereket vetünk be annak pontos meghatározására.

4. A negyedik kritérium egyértelműen teljesül; a kérdés pusztán az lehet, hogy a kapott eredmény mekkora kifejezőerővel bír, mennyire megbízható, stb.

Tehát a kiinduláskor megállapított követelményeink (nagyreszt) teljesülnek.

10. További lehetőségek

Az itt leírtak alapján szükség lesz egy esettanulmányra, hogy az algoritmus helyességét és hasznosságát teszteljük. Ehhez szükség van még megfelelő tesztek előállítására is, itt értve a be- és kimeneteket.

A módszert lehet tovább is elemezni, hiszen több aggály, fejlesztési lehetőség is felmerült (lásd a problémás szekciókat).

A leghasznosabb további lépés pedig az lenne, ha találnánk módszert arra, hogy nem-izomorf beosztások igazsága közt relációt állítsunk.

⁸Muszáj megemlítenünk, hogy a problémadefiníció még mindig korlátozott, hiszen az entitáspáronként független modelt megköveteljük.