

Egyedek tipizálása II.

4. kutatói pályázat (nyílt, pénzdíjas) - meghosszabbítása

Pályadíj (fennmaradó): 35 000 Ft

Kiírás időpontja: 2010. december 7. Módosítás időpontja: 2011. február 8.

Határidő: 2011. február 28

Az „Egyedek tipizálása” pályázatra 2011. január 31.-ig beérkezett megoldások közül a bírálók Tőkés Ábel pályaművét ítélték a legjobb részmegoldásnak és a pályadíj 30%-val díjazták. Gratulálunk!

Ugyanakkor a pályázatot a pályaművek alapján pontosítva 2011. február 28.-i **VÉGSŐ** határidőig meghosszabbítottuk és az érdeklődőket továbbra is arra biztatjuk, hogy folytassák a munkát. A

A pályamunka (dokumentáció) elektronikus formában az Egyesület fényposta címére nyújtandó be: ettesust0@gmail.com.

Ezen felül minden, a pályázaton indulni kívánó érdeklődőt arra kérünk, hogy részvételi szándékát jelezze a megadott email címen. Tesztadatokat kizárólag személyes megkeresésre tudunk küldeni.

Az eredeti és a módosított kiírás elektronikus formában elérhető: <http://ette-sust.hu/?q=paly4>

A kiírással kapcsolatos kérdések és válaszok elérhetősége: <http://ette-sust.hu/?q=paly4-forum>

A feladat leírása

Vannak egyedek (Egyed 1 ... Egyed 10), melyek különböző tulajdonságokkal (Tul. 1 ... Tul. 10) bírnak. Egy példa erre az alábbi táblázatban található.

	Tul. 1	Tul. 2	Tul. 3	Tul. 4	Tul. 5	Tul. 6	Tul. 7	Tul. 8	Tul. 9	Tul. 10
Egyed 1	x	x			x	x		x	x	x
Egyed 2			x			x	x			
Egyed 3						x	x			
Egyed 4		x	x	x				x		
Egyed 5	x	x			x	x			x	x
Egyed 6										
Egyed 7	x	x	x			x	x		x	x
Egyed 8			x			x	x			
Egyed 9		x	x	x				x		
Egyed 10		x	x	x				x		

A pályázat célja olyan **matematikai módszer, algoritmus** kidolgozása, definiálása, mellyel olyan, egymást nem kizáró általánosításokat (Tip1 ... TipN - típusokat) lehet bevezetni, amivel a jelenleginél jóval egyszerűbben jellemezhetők ezek az egyedek.

Az általánosításra vonatkozó szabályok:

1. Egy egyed tartozhat több típushoz is, de cél, hogy egy egyed minél kevesebb típushoz tartozzon és leginkább, hogy maximálva legyen egy egyed esetén ezeknek a száma.
2. Cél, hogy a populáció minél kevesebb típussal leírható legyen
3. Kis mértékben megengedett a kivételek száma, vagyis Típ1, de Tul.7 nélkül, vagy Típ2 + Tul.9. De a kivételek száma/aránya a maximálható legyen egyedenként és a teljes populáció viszonylatában.
4. A határértékek a modellben legyenek paramétereizhetők.

A fenti példa egy lehetséges megoldása:

	Tul. 1	Tul. 2	Tul. 3	Tul. 4	Tul. 5	Tul. 6	Tul. 7	Tul. 8	Tul. 9	Tul. 10
Egyed 1	Típ1	Típ1			Típ1	Típ1		X	Típ1	Típ1
Egyed 2			Típ2			Típ2	Típ2			
Egyed 3			-			Típ2	Típ2			
Egyed 4		Típ3	Típ3	Típ3				Típ3		
Egyed 5	Típ1	Típ1			Típ1	Típ1			Típ1	Típ1
Egyed 6										
Egyed 7	Típ1	Típ1	Típ2		-	Típ1Típ2	Típ2		Típ1	Típ1
Egyed 8			Típ2			Típ2	Típ2			
Egyed 9		Típ3	Típ3	Típ3				Típ3		
Egyed 10		Típ3	Típ3	Típ3				Típ3		

Pályázati feltételek.

A pályamunka a matematikai módszer, algoritmus leírását kell, hogy tartalmazza, valamint egy olyan futtatási eredményt, ami 20 egyed*20 tulajdonság esetén 3-5 Típussal, és 3-6 kivétellel leírható (becsült érték, ettől el lehet térni). Fontos, hogy a dokumentáció tartalmazza a modell megvalósításának módját, ami alapján a bíráló bizottság képes kell, hogy legyen a működés ellenőrzésére. Az elbírálásnál szempont a helyes működés mellett a rugalmasság, paramétereizhetőség, illetve az egyszerűség, reprodukálhatóság is.

A pályadíjat a legjobb munka nyeri, azt csak kivételes esetben osztjuk meg. Szükség esetén részmegoldásokat is értékelünk. Egyéb kérdésekben az Egyesület általános pályázati szabályzata érvényes. A pályázat kiírása és elbírálását az ETTE szakmai közreműködésével a kiíró végzi. A beérkezett feladatok pályázatok elbírálása a beérkezéskor kezdődik és a határidőn belül leadott, vagy azt követő első sikeres pályázati értékelésig, ennek hiányában a végső határidőig tart.

Kiegészítések – Kérdések és Válaszok

Amelyik kalickákban nincs X..

Kérdés:

Amelyik kalickákban nincs X, az a tulajdonság nem teljesül, vagy csak nem tudjuk, hogy teljesül-e? Pl. Egyed 1 rendelkezhet-e Tul. 3-al, vagy biztosan tudjuk róla, hogy nem rendelkezik vele?

Válasz:

Ahol nincs X, az a tulajdonság az adott egyed esetén egyértelműen nem teljesül. (Vagyis egyedeknél nincs "mindegy"/"don't care" állapot.) Ugyanakkor a típusok esetén ez másként értelmezendő.

Típus megadása hogyan történik?

Kérdés:

Egy típus megadható pl.

Tul.1 = Y

Tul.2 = N

Tul.3 = W

Tul.4 = Y

Tul.5 = Y

Tul.6 = N

Tul.7 = W

Tul.8 = W

Tul.9 = Y

Tul.10 = Y

alakban? Itt Y a YES (igen), N a NO (nem), W a WHATEVER (mindegy) jelölése. Vagy csak Y és W szerepelhet? Vagy csak Y és N szerepelhet?

Válasz:

A típus szempontjából ennél a feladatnál csak valamely tulajdonság megléte számít Y, a többi W, vagyis ("mindegy"/"don't care")

A típusok egyértelműen meghatározzák-e...

Kérdés:

Ha pl. Egyed 1-nek megmondjuk a típusait, abból a Tul. 1, Tul. 2, ...

Tul. 10-ről mindről egyértelműen ki kell derülnön, hogy Egyed 1 rendelkezik-e vele? Tehát az is ki kell derülnön, hogy melyik tulajdonságokkal nem rendelkezik?

Válasz:

A típus + hiba listának pontosan definiálnia kell az egyedeket.

A túl sok típus elkerülésére a típusoktól önmagukban csak egy jó közelítést várunk el. Ugyanakkor a

teljes egyezés elérésére megengedjük a "hibákat", amikből persze szintén ne legyen túl sok. Az optimumot kell megkeresni.

Egy egyed tartozhat több típushoz...

A kiírásban szerepel, hogy egy egyed tartozhat több típushoz is. Mit is jelent ez? Mi ennek a célja, háttere.

Példa:

Tegyük fel, hogy az E1 egyed T1, T2, T3 tulajdonságokkal bír, ez alapján ő Y1 típusú. (Nevezzük az érthetőség kedvéért Y1-et kék szeműnek.)

Tegyük fel, hogy az E2 egyed T4, T5, T6 tulajdonságokkal bír, ez alapján ő Y2 típusú. (Nevezzük az érthetőség kedvéért Y2-t szőkének.)

Tegyük fel, hogy az E3 egyed T1, T2, T3 és T4, T5, T6 tulajdonságokkal bír, ez alapján ő lehetne egy Y3 típus. (Nevezzük az érthetőség kedvéért Y3-t szőke-kékszeműnek.) Ugyanakkor sokkal célszerűbb E3 jellemzésére Y1-et és Y2-t együttesen használni. A feladat céljának ez utóbbi megközelítés felel meg jobban.

Típusok jellemzői

Valószínűleg lehetne a feladatban erre a funkcióra a típus megnevezésnél jobb szót találni. Ugyanakkor történelmi okokból ezen a szóhasználaton most már nem változtatunk, inkább pontosan definiáljuk, hogy mit értünk ez alatt és milyen jellegzetességekkel rendelkezik.

A **típus** egy olyan **elemi tulajdonság kombináció**, ami az esetek döntő többségében együtt jár, tehát az egyednek vagy mind (többségében) megvan, vagy mind (többségében) hiányzik.

Milyen irodalmat érdemes ...

Kérdés:

Milyen irodalmat érdemes megnézni a feladat megoldásához?

Válasz:

Javasolnánk átnézni egy pár adatbányászati, elemzési témát, mint:

- Clustering - http://en.wikipedia.org/wiki/Cluster_analysis
- Boolean factor analysis
- Unsupervised learning - http://en.wikipedia.org/wiki/Unsupervised_learning
- esetleg: Dimensional Reduction - http://en.wikipedia.org/wiki/Dimensionality_reduction

témákat. A feladat természetesen ezek nélkül is megoldható, ugyanakkor ezek az olvasmányok nagyban segíthetik mind az ismeretek bővítését, mind a professzionálisabb megoldás készítését.

Típus megadása hogyan történik?

Kérdés:

Egy típus megadható pl.

Tul.1 = Y

Tul.2 = N

Tul.3 = W

Tul.4 = Y

Tul.5 = Y

Tul.6 = N

Tul.7 = W

Tul.8 = W

Tul.9 = Y

Tul.10 = Y

alakban? Itt Y a YES (igen), N a NO (nem), W a WHATEVER (mindegy) jelölése. Vagy csak Y és W szerepelhet? Vagy csak Y és N szerepelhet?

Válasz:

A típus szempontjából ennél a feladatnál csak valamely tulajdonság megléte számít Y, a többi W, vagyis ("mindegy"/"don't care")

A típusok egyértelműen meghatározzák-e...

Kérdés:

Ha pl. Egyed 1-nek megmondjuk a típusait, abból a Tul. 1, Tul. 2, ...

Tul. 10-ről mindről egyértelműen ki kell derülnön, hogy Egyed 1 rendelkezik-e vele? Tehát az is ki kell derülnön, hogy melyik tulajdonságokkal nem rendelkezik?

Válasz:

A típus + hiba listának pontosan definiálnia kell az egyedet.

A túl sok típus elkerülésére a típusoktól önmagukban csak egy jó közelítést várunk el. Ugyanakkor a teljes egyezés elérésére megengedjük a "hibákat", amikből persze szintén ne legyen túl sok. Az optimumot kell megkeresni.

Amelyik kalickákban nincs X...

Kérdés:

Amelyik kalickákban nincs X, az a tulajdonság nem teljesül, vagy csak nem tudjuk, hogy teljesül-e? Pl. Egyed 1 rendelkezhet-e Tul. 3-al, vagy biztosan tudjuk róla, hogy nem rendelkezik vele?

Válasz:

Ahol nincs X, az a tulajdonság egyértelműen nem teljesül. (Vagyis egyedeknél nincs "mindegy"/"don't care" állapot.)