

Stabilné párovanie (Stable matching)

Je potrebné vytvoriť **páry** medzi dvomi množinami entít (napríklad muži a ženy, zamestnávateľia a uchádzači o zamestnanie, školy a študenti, doktori a pacienti, ap. - všade tam, kde existuje dvojstranný trh), pričom jeden člen každého páru je z jednej množiny entít a druhý člen páru je zase z druhej množiny entít. Každá entita môže byť zaradená iba v jednom páre. Toto párovanie je potrebné vytvoriť tak, aby neexistovali žiadne nestabilné páry, ktoré by sa mohli rozpadnúť a tým porušiť vytvorené priradenia. Zároveň je potrebné spárovať čo najviac entít, teda vytvoriť čo najviac párov (ak sa dá, tak spárovať všetky entity – samozrejme, že ak tie množiny entít sú nerovnako veľké, tak to nie je možné).

Každý člen nejakej množiny entít vyjadruje svoj vzťah k entitám z druhej množiny svojim **preferenčným zoznamom**, ktorý obsahuje entity z druhej množiny v preferovanom usporiadaní. Pri existencii množín entít $X=\{x_1, x_2, \dots, x_n\}$ a $Y=\{y_1, y_2, \dots, y_m\}$ potom preferenčný zoznam entity x_6 môže vyzeráť napríklad (predpokladáme, že množina Y obsahuje 10 entít)

$$PL(x_6) = y_8, [y_3, y_4, y_9], y_1, [y_2, y_5], y_6, -y_7, -y_{10}$$

znamenajúci, že

- y_8 je najpreferovanejší partner,
- y_1 je menej preferovaný ako y_4 alebo y_8 , avšak viac preferovaný ako y_2 či y_6
- y_3, y_4 a y_9 sú rovnako preferovaní (ale každý z nich je preferovaný viac ako y_1 a menej ako y_8)
- y_7 a y_{10} sú neprijateľné

Podobné preferenčné zoznamy má každá entita z množiny X (tieto zoznamy parciálne zotriedujú entity z Y) ako aj entita z množiny Y (reprezentujúce parciálne poradia entít z množiny X).

Nestabilné páry nesmú byť súčasťou vytvoreného párovania, inak by párovanie bolo nestabilné.

V párovaní vznikajú nestabilné páry vtedy, ak nejaký člen takéhoto páru má príležitosť spárovať sa s niekým iným, ktorého preferuje viac než svojho aktuálneho partnera. Príkladom jedného nestabilného páru je pár (x_1, y_1) v situácii, keď preferenčné zoznamy majú tvar

$$PL(x_1) = \dots, y_2, \dots, y_1, \dots$$

$$PL(y_1) = \dots, x_1, \dots, x_6, \dots$$

$$PL(y_2) = \dots, x_2, \dots, x_1, \dots$$

a entita y_2 je bez páru. V tomto prípade má entita x_1 tendenciu opustiť svojho aktuálneho partnera y_1 a vymeniť ho za nového partnera y_2 .

Iným prípadom je situácia dvoch párov, z ktorých každý samostatne je stabilný avšak spolu ako dvojica sú nestabilné. To nastáva vtedy, ak jeden člen prvého páru a jeden člen druhého páru sú schopní spolu vytvoriť tzv. **blokujúci pár** (čím sa pôvodné dva páry rozpadnú), ktorý nie je súčasťou aktuálneho párovania ale môže vzniknúť (a tým celé párovanie pokaziť). Je možné uvažovať dva typy takýchto blokujúcich párov. V prípade, že preferenčné zoznamy niektorých entít sú

$$PL(x_1) = \dots, y_6, \dots$$

$$PL(x_2) = \dots, y_5, \dots, y_4, \dots$$

$$PL(x_3) = \dots, y_6, \dots, y_5, \dots$$

$$PL(y_4) = \dots, x_2, \dots, x_6, \dots$$

$$PL(y_5) = \dots, x_2, \dots, x_3, \dots$$

$$PL(y_6) = \dots, [x_1, x_2, x_3], \dots$$

a aktuálne párovanie obsahuje dvojice

...

$$(x_1, y_6)$$

$$(x_2, y_4)$$

$$(x_3, y_5)$$

...
 tak blokujúcim párom prvého typu je pár (x_2, y_5) . Tieto dve entity sú aktuálne párované s entitami y_4 a x_3 , ale keďže obe x_2 aj y_5 si vedia polepšiť, tak svojich aktuálnych partnerov opustia a vytvoria spoločný pár, ktorý invaliduje páry (x_2, y_4) a (x_3, y_5) a tým aj celé aktuálne párovanie.

Neexistencia blokujúcich párov tohto typu je podmienkou **slabej stability** - párovanie, pre ktoré neexistuje žiadny takýto blokujúci pár (keď si jeho vznikom obaja členovia polepšia) sa označuje ako slabo stabilný.

Pri danom párovaní blokujúcim párom druhého typu je pár (x_3, y_6) . Tieto dve entity sú aktuálne párované s entitami y_5 a x_1 , ale keďže x_3 si vie polepšiť a y_6 si nepohorší (ale ocitne sa s novým partnerom na rovnakom stupni preferencie ako bol jeho starý partner), tak svojich aktuálnych partnerov opustia a vytvoria spoločný pár, ktorý invaliduje páry (x_1, y_6) a (x_3, y_5) a tým aj celé aktuálne párovanie.

Neexistencia blokujúcich párov predchádzajúceho aj tohto typu je podmienkou **silnej stability** - párovanie, pre ktoré neexistuje žiadny takýto blokujúci pár (keď si jeho vznikom jeden člen polepší a druhý člen si nepohorší) sa označuje ako silne stabilný.

Často je možné vytvoriť niekoľko rôznych stabilných párovaní pri danej ponuke entít s ich preferenčnými zoznamami. Potom prichádza do popredia otázka, ako je možné párovania ohodnotiť, aby sa dali porovnávať.

Pre definovanie nejakej miery je potrebné zaviesť vhodné značenie. Ak $PL(x_9)$ reprezentuje preferenčný zoznam entity x_9 , potom $PL(x_9)[y_2]$ bude reprezentovať poradie, na ktorom sa nachádza entita y_2 v preferenčnom zozname entity x_9 . Ak preferenčný zoznam entity x_9 je

$$PL(x_9) = [x_3, x_4, x_6, -x_1, -x_2]$$

tak $PL(x_9)[x_4] = 1$, $PL(x_9)[x_6] = 2$ a $PL(x_9)[x_2] = 4$ (na poradí odmietaných entít nezáleží, je jedno či budú na konci preferenčného zoznamu po jednom alebo v zátvorke ako jedna skupina – podstatné je, aby boli na konci a nenarúšali poradie entít, ktoré sú prípustné pre párovanie s vlastníkom daného preferenčného zoznamu).

Potom celková **cena šťastia** (happiness cost) je definovaná ako

$$\sum PL(x_i)[y_j] + \sum PL(y_j)[x_i]$$

pričom sa uvažuje dvojica indexov (i,j) iba vtedy, ak v párovaní je zaradený pár (x_i, y_j) .

Iná miera celková **cena rovnosti** (egalitarian cost) je definovaná ako

$$\sum PL(x_i)[y_j] - \sum PL(y_j)[x_i]$$

pričom sa uvažuje dvojica indexov (i,j) iba vtedy, ak v párovaní je zaradený pár (x_i, y_j) .

Existuje niekoľko rozlišovaných typov úlohy stabilného párovania:

- SM (Stable Matching) – preferenčné zoznamy udávajú úplné zotriedenie entít druhej množiny, nevylučujú žiadne entity
- SMI (Stable Matching with Incomplete lists) - preferenčné zoznamy udávajú úplné zotriedenie entít druhej množiny, niektoré entity entity sú vylúčené z párovania s vlastníkom preferenčného zoznamu
- SMT (Stable Matching with Ties) - preferenčné zoznamy udávajú iba parciálne zotriedenie entít druhej množiny, nevylučujú žiadne entity
- SMTI (Stable Matching with Ties and Incomplete Lists) - preferenčné zoznamy udávajú iba parciálne zotriedenie entít druhej množiny, niektoré entity entity sú vylúčené z párovania s vlastníkom preferenčného zoznamu

Sú aj ďalšie špeciálne prípady. V prípade „SM with master preference list“ sú entity zoradené do parciálneho poradia podľa nejakého objektívneho kritéria a každá entita vo svojom preferenčnom zozname reflektuje toto zoradenie (tie entity druhej množiny, ktoré sú prípustné pre vlastníka

preferenčného zoznamu, sú v preferenčnom zozname v rovnakom poradí ako je dané v master zozname). V prípade „MS with bounded preference lists“ sú preferenčné zoznamy obmedzené dĺžkou, ktorá nesmie byť väčšia ako nejaká hraničná hodnota (čo je typicky nejaká malá hodnota, napr. 2 alebo 3). V „N-sided SM“ namiesto dvoch množín vystupuje tých množín entít viac – veľmi často tri (napr. dodávateľ – obchodník – zákazník, darca – lekár – pacient, ap.) ale aj viac, dokonca s premenlivým počtom (napr. zdroj dát – niekoľko prenosových uzlov – prijímateľ).

Úloha

Vašou úlohou je pre zadaný zoznam entít a ich preferencií nájsť také silne stabilné párovanie, ktoré produkuje čo najväčší počet párov (primárne kritérium), pričom hodnota happiness cost bude čo najmenšia a hodnota egalitarian cost bude čo najbližšie k nule (tieto cenové kritériá sú použité ako sekundárne kritérium, uvažované v prípade riešenia s rovnakou hodnotou primárneho kritéria). Úlohu riešte pre 4 typy vstupov (SM, SMI, SMT, SMTI).

Vstupmi sú súbory s preferenciami: sm.txt, smi.txt, smt.txt a smti.txt kde každý z nich obsahuje dáta pre jeden typ problému.

Ako svoje riešenie dodajte 6 súborov:

- súbor so zdrojovým textom vášho algoritmu (predpokladá sa, že všetky štyri problémy budete riešiť tým istým algoritmom)
- súbor typu *.pdf s popisom riešenia (ako a prečo ste volili reprezentáciu, určovanie vhodnosti, generovanie iniciálnej populácie, genetické operátory, selekciu, náhradu, návrh hodnôt parametrov, ukončenie behu algoritmu), dosiahnutých výsledkov (pre každý z problémov dajte parametre odovzdávaného riešenia v tvare počet párov, hodnota ceny šťastia a ceny rovnosti) a svojich skúseností s tvorbou riešenia (s čím ste mali problém, čo bolo jednoduché, ...)
- súbor s vašim riešením (4 x – jeden súbor pre riešenie každého z problémov). V tomto súbore budú uvedené jednotlivé páry nájdeného párovania v tvare dvoch stĺpcov

m6 w12
m35 w80

...

pričom

- entity budú oddelené jednou medzerou, hneď za druhou entitou je znak nového riadku, pred prvou entitou nie je medzera ani v prípade jednociferej entity
- na poradí jednotlivých párov nezáleží
- je jedno či najprv bude entita m* a až po nej entita w* alebo naopak (poradie sa môže aj striedať v rámci jedného súboru)

Poznámka

Ako prvý sa riešeniu problémov tohto typu (so zameraním na ekonomickú oblasť) venoval Lloyd S. Shapley, ktorý za to získal v roku 2012 Nobelovu cenu za ekonómiu „for the theory of stable allocations and the practice of market design“. Algoritmus Gale-Shapley z roku 1962 rieši úlohu SM za podmienok, že obe množiny majú rovnaký počet entít.