RANGSOROLÁSI ALGORITMUSOK

(Iványi Antal előadásának kivonata: Szeged, 2003. április 22.)

A hagyományos *versenyeken* n játékos körmérkőzést játszik. A mérkőzések *győztese* egy pontot, vesztese nulla pontot kap.

A játékosok rangsorolása történhet például a megszerzett pontok száma alapján (ekkor a pontszámok szerint rendezünk), vagy minimális számú inverzióra (vesztes megelőzi)

Legyenek $a,\ b, m$ és n egész számok, melyekre $0 \le a \le b$, $2 \le m < n$, $1 \le b$.

Egy A (a, b, m, n)**-verseny** egy olyan $T_{kmn} = [t_{ij}]_{n\times n}$ $n \times n$ méretű mátrix, amelynek az elemei egész számok és

$$a \le t_{ij} \le b \ \ (1 \le i,\ j \le n),$$

$$t_{ii} = 0 \ \ (1 \le i \le n),$$

$$k \le t_{ij} + t_{ji} \le m \ \ (1 \le i,\ j \le n,\ i \ne j). \tag{1}$$

Tegyük fel, hogy adott n-re előállítottuk az összes **sport** sorozatot. Ezek definíciója: azok a helyi sorozatok, amelyekhez van sport mátrix. Az eddigi szimulációs adatok szerint a monoton sorozatoknak körülbelül nyolcadrésze sport sorozat és a sport sorozatoknak körülbelül kétharmada focisorozat.

Adott sorozat esetén $O(n)$ idő alatt el tudjuk dönteni, hogy a sorozat sport sorozat-e. Tehát a monoton sorozatok mintegy 92 %-át gyorsan ki tudjuk zárni.

A következő három "negatív szűrő" (azaz sorozatokat kizáró algoritmus) már több időt igényel, bár kétségtelenül polinomiálisak.

Az egyik a **döntetlenekkel** foglalkozik. A maradékokból adódó kötelező döntetleneken felül kiosztandó döntetleneket lehetőleg egyenletesen osztja szét, majd ezeket megpróbálja párosítani.

Ennek a módszernek erősebb változata a nyeretlen, illetve veretlen csapatok közötti, valamint a csapatok más részhalmazain belüli kötelező döntetleneket is figyelembe veszi.

A másik a **győzelmeket és vereségeket** vizsgálja: kioszthatók-e úgy a szabad vereségek, illetve győzelmek, hogy a nagy számú győzelmeknek és vereségeknek legyen párja.

A harmadik a sorozatok **felbontásán** alapul: bizonyos feltételek esetén egy sorozat úgy bontható rövidebb sorozatokra, hogy azok könnyen ellenőrizhető tulajdonságaiból következtethetünk az eredeti sorozat tulajdonságaira.

Az eddigi tapasztalatok szerint ez a három szűrő nagyon hatásos: a felbontott sorozatok túlnyomó részéhez sikerült pont mátrixot készíteni.

2. A fő ötletek

Az algoritmusokat úgy fogalmazzuk meg, hogy adott sorozat ellenőrzését végezzék – bár célunk gyakran több sorozat (például adott számú csapat esetén lehetséges sport sorozatok) ellenőrzése.

2.1. Döntetlenek egyenletes elosztása

Egy $\mathbf{q} = q[1], \ldots, q[n]$ sport sorozat döntetlenjeit úgy vizsgáljuk, hogy először kitöltjük a $n \times 4$ méretű OSM kötelező sport mátrixot. Ennek első három oszlopa a szokásos: a kötelező döntetlenek, győzelmek és vereségek száma. A negyedik oszlop az adott csapat hiányzó **csomagjainak** számát adja meg. Egy csomag vagy 3 döntetlent, vagy 1 győzelmet és 2 vereséget tartalmaz, ezért a hiányzó csomagok száma $S - 3OW - OD$. A kiosztandó döntetlen csomagok száma $NDP[1] = 2(3NM - S - OD)/3$.

Legyen $POS[1]$ azon csapatok száma, amelyeknek adható még döntetlen, azaz amelyekre $OSM[i, 4] > 0$.

Ha $NDP[1]/POS[1] \ge 1$, akkor menjünk végig azokon a csapatokon, amelyek készek még döntetlent vállalni, és adjunk nekik min $\lfloor NDP[1]/POS[1] \rfloor, OSM[i, 4]$ $(i = 1, 2, \ldots, n)$ döntetlen csomagot. Az odaadott csomagok számát a DP vektorban jegyezzük meg, és $NDP[1]$ csökkentésévek számítsuk ki $NDP[2]$ értékét.

Mindezt addig imételjük a $POS[j]$ $(j = 2, \ldots)$ értékekkel, amíg $NDP[j]/POS[j] \ge 1$. Az új POS értékeket természetesen az $OSM[i, 4] - NP[j - 1]$ értékek alapján határozzuk meg.

Ha $NDP[j] = 0$, akkor kész vagyunk. Ha $0 < NDP[j]/POS[1] < 1$, akkor osszuk ki úgy a maradék csomagokat a pozitív $OSM[i, 4]$ értékű csapatok között, hogy először azok kapnak (pontszámuk növekvő sorrandjében), akiknél $OSM[i, 2] = 0$, azután azok, akiknek a pontszáma egy maradékot ad hárommal osztva, végül azok, akiknél a maradék kettő.

Ezután az NP tömbnek megfelelően állítsuk elő a $n \times 3$ méretű SM sport mátrixot, majd annak döntetlen oszlopában szereplő értékek alapján egy ND vektorban a döntetlenszámok csökkenő sorozatát. Ezután ezt a vektort úgy kezeljük, hogy rendre töröljük legnagyobb $ND[i]$ $(i = 1, \ldots, n)$ elemét, egyidejűleg eggyel csökkentjük a következő $ND[i]$ elemet.

Ha ilymódon nem sikerült kinulláznunk az ND vektort, akkor a vizsgált sorozat rossz. Ha sikerült, akkor jön a "szigorú döntetlenvizsgálat". Ennek során egyrészt a kötelező sport mátrix szerint veretlen, illetve nyeretlen csapatok belső döntetlenjeit, másrészt az $SW[i]$ és $SB[i]$ értékek vizsgálata alapján adódó belső döntetleneket kivesszük a sport mátrix döntetlenjei közül, és ezután próbáljuk meg a Havel-Hakim párosítást.

2.2. Kötelező döntetlenek

Ha egy sorozat kezdőszelének összege kicsi, vagy zárószeletének összege nagy, akkor abból következtetéseket vonhatunk le mindkét szelet belső döntetlenjeinek számára nézve. Legyen $IDW[i]$ az i kis pontszámú játékos belső döntetlenjeinek a száma, $OIDW[i]$ az i kis pontszámú játékos kötelező belső döntetlenjeinek száma, $IDB[i]$ az i nagy pontszámú játékos belső döntetlenjeinek a száma és $OIDB[i]$ az i nagy pontszámú játékos kötelező belső döntetlenjeinek száma.

Az i kis pontszám összege $SW[i]$, az $n - i$ nagy pontszám összege $SB[n - i]$.

A legegyszerűbb eset, hogy ha $SW[i] = 2NM[i]$, akkor az első i csapat minden belső meccse döntetlennel végződött, és ezért $q[1] = q[2] = \ldots = q[i] = i - 1$. Ezért a kötelező sport mátrixba beírhatjuk ezeket a döntetlenszámokat (külön jelezve, hogy ezek "foglaltak").

Általában:

$$IDW[i] \geq OIDW[i] \geq 3NM[i] - SW[i], \tag{2}$$

$$IDB[i] \geq 3(n - i) + 2NM[n - i] - 2((SW[i] - 3NM[i]), \tag{3}$$

Másrészt az is igaz, hogy ha egy sportmátrixban $k \geq 2$ nyeretlen csapat van, akkor

$$IDW[i] \geq OIDW[i] \geq \binom{k}{2}. \tag{4}$$

Hasonló állítások a nagy pontszámokra is igazak.

2.3. Győzelmek és vereségek egyenletes elosztása

Egy sportmátrix helyreállíthatóságának szükséges feltétele, hogy bármely k $(1 \leq k \leq n)$ csapat győzelmeihez tudjunk vereségeket párosítani – figyelembe véve, hogy ehhez a többi $n - k$ csapat egyenként legfeljebb k vereséggel, a kiválasztott csapatok pedig egyenként legfeljebb $k-1$, összesen pedig legfeljebb $\binom{k}{2}$ vereséggel járulhatnak hozzá.

Természetesen a vereségekre hasonló szükséges feltétel érvényes.

Ezt a feltételt alkalmazhatjuk például a kötelező sportmátrixra úgy, hogy rendre a $ow[1]$, $ow[1] + ow[2], \ldots, ow[1] + \ldots + ow[n]$ összegekhez próbálunk megfelelő számú vereséget keresni – a 2 vereségből és 1 győzelemből álló csomagok legkedvezőbb kiosztását feltételezve.

2.4. Felbontás

A **felbontás** alapja az, hogy ha egy sorozat kezdőszelének összege kicsi, vagy zárószeletének összege nagy, akkor abból szükséges feltételeket vezethetünk le mindkét szeletre nézve.

1) Ha $SW[i] = 2\binom{i}{2} = i(i - 1)$, akkor minden meccs eredménye döntetlen, ezért minden pontszám $i - 1$.

2) $SW[i] = i(i - 1) + 1$ és $i = n$ csak úgy lehetséges, hogy a belső döntetlenek mellett van egy belső győzelem, ezért $q[1] = i - 2, q[i] = i + 1$, a többi pontszám pedig $i - 1$.

Ha $SW[i] = i(i - 1) + 1$ és $i < n$, akkor a többletpontszám származhat egy külső döntetlenből is, ekkor $q[i] = i$, a többi pontszám pedig $i - 1$.

3) $SW[i] = i(i - 1) + 2$ és $i = n$ csak két belső győzelemmel lehetséges, amihez legalább 3 csapat kell. Vagy egy csapat nyer 2 meccset (ekkor $q[1] = q[2] = i - 2$), vagy egy csapat veszít és nyer 1-1 meccset (ekkor $q[1] = i - 2,\ q[i - 1] = i, q[i] = i + 1$) vagy egy csapat veszít 2 meccset (ekkor $q[1] = i - 3,\ q[i - 1] = q[i] = i + 1$) vagy 2 csapat nyer 1-1 meccset különböző ellenfelek ellen (ehhez 4 csapat kell és ekkor $q[1] = q[2] = i - 2,\ q[i - 1] = q[i] = i + 1$).

Ha $SW[i] = i(i-1)+2$ és $i < n$, akkor a 2 többletpont egyike vagy mindkettő külső döntetlenből származik. Utóbbi a $q[i] = i+1$ vagy a $q[i-1] = q[i] = i$ módon fordulhat elő, míg előbbi a $q[1] = i-2,\ q[i] = i+2$ vagy $q[1] = i-2,\ q[i-1] = i,\ q[i] = i+1$ vagy $q[i] = i+2$ módon lehetséges.

4) Ha $SW[i] = i(i-1)+3$ és $i = n$, akkor 3 belső győzelem van.

Ha $SW[i] = i(i-1)+3$ és $i < n$, akkor a 3 többletpontból 3 (1 győzelem vagy 3 döntetlen), 2 (2 döntetlen), vagy 1 (1 döntetlen) kívülről is származhat.

5) Ha $SB[i] = 3i(n-i) + 3NM[i]$, akkor a nagypontosok legyőzték a többieket, egymás között pedig nem játszottak döntetlent. Ezért a $q[1], \ldots, q[n-i]$ sorozatnak jónak kell lennie, a $(q[n-i+1] - 3(n-i)), \ldots, (q[n]-3(n-i))$ sorozatnak a Landau-feltétel szerint jónak kell lennie.

6) Ha $SB[i] = 3i(n-i) + 3NM[i] - 1$, akkor a nagypontosok legyőzték a többieket, egymás között pedig egy döntetlent játszottak. Ezért a kis pontszámokra az előző feltétel vonatkozik, a nagyoknál két pontszámnak kell egy maradékot adnia mod 3. Ezek egyikét eggyel csökkentve, a másikat kettővel növelve az előzőekhez hasonlóan alkalmazzuk a Landau-feltételt.

7) Ha $SB[i] = 3i(n-i) + 3NM[i] - 2$, akkor vagy egy külső döntetlennel, vagy két belső döntetlennel vesztettek pontokat a nagy pontszámúak.

Mivel a legfeljebb 3 többletpontot tartalmazó kezdőszeleteket megvizsgáljuk, ezért természetes, hogy algoritmusunk $n = 3$ csapatig csak a foci sorozatokat állítja elő, hiszen a korábbi ellenőrzések legfeljebb 3 többletpontot engednek meg.

3. Algoritmusok

A Draws algoritmus bemenő adatai: egy sorozat vizsgálata esetében n, a csapatok száma és $\mathbf{q} = q[1], \ldots, q[n]$, a vizsgálandó sorozat (egyébként a vizsgálandó sorozatok száma és maguk a sorozatok).

Az algoritmus kimenő adatai: egy sorozat esetén a sorozatot minősítő üzenet (egyébként $\sigma(n)$, a talált sorozatok száma, valamint a sport sorozatokat tartalmazó mátrix).

Az algoritmust a [2]-ben alkalmazott pszeudokód segítségével adjuk meg. Az alkalmazott munkaváltozók:

$i,\ j$: indexek (i mindig egytől n-ig változik);

$NM[i] = \binom{i}{2}$: i csapat egymás között lejátszott mérkőzéseinek száma;

$NM = \binom{n}{2}$: az összes mérkőzés száma;

$SW[i]$: az i legkisebb pontszám összege;

$SW = SW[n]$: az összes pontszám összege;

$LW[i]$: az i legkisebb pontszámú csapat által biztosan kidobott pontok száma;

$OW[i]$: a vizsgált $q[i]$ pontszámhoz szükséges győzelmek becsült száma;

$OD[i]$: a vizsgált $q[i]$ pontszámhoz szükséges döntetlenek becsült száma;

$OL[i]$: a vizsgált $q[i]$ pontszámhoz szükséges vereségek száma;

$SOW[i]$: a vizsgált kezdőszelethez szükséges győzelmek becsült száma;

$SOD[i]$: a vizsgált kezdőszelethe szükséges döntetlenek becsült száma;

$SOL[i]$: a vizsgált kezdőszelethez szükséges vereségek száma;

$SADP[i]$: a vizsgált kezdőszeletben megengedett kiegészítő döntetlen csomagok száma;

$OIP[i]$: a vizsgált játékos kötelező belső pontjainak száma;

$OIS[i]$: a vizsgált kezdőszelet elemeinek kötelező összege;

$OIL[i]$: a vizsgált kezdőszelet kötelező belső vereségeinek becsült száma;

σ: az adott csapatszámhoz tartozó sport sorozatok száma;

$S[j, i]$: a j-edik sport sorozat i-edik eleme;

DRAWS(n, **q**) **algoritmus definíciója**

$\triangleright$ kiszámítjuk a kötelező sport mátrix elemeit
$SW[0] \leftarrow 0$
$SOD[0] \leftarrow 0$
$SOW[0] \leftarrow 0$
$SOL[0] \leftarrow 0$
$SADP[0] \leftarrow 0$
$NM = n(n-1)/2$
$NM[0] = 0$
for $i \leftarrow 1$ **to** n **do**
    $OD[i] \leftarrow (q[i]) - 3\lfloor q[i]/3 \rfloor)$

$OW[i] \leftarrow 0$
$OL[i] \leftarrow 0$
$NM[i] = NM[i-1] + i - 1$
if $q[i] \geq n-1$ **then do**
$OW[i] \leftarrow (\lceil (q[i] - (n-1))/2 \rceil$
if $(q[i] - n + 1)/2 = \lfloor (q[i] - n + 1)/2 \rfloor$ **then do** $OL[i] \leftarrow 1$
if $q[i] < n-1$ **then do** $OL[i] \leftarrow n - 1 - q[i]$
$ADP[i] \leftarrow (q[i] - OD[i] - 3OW[i])$
$SW[i] \leftarrow SW[i-1] + q[i]$
$SOD[i] \leftarrow SOD[i-1] + OD[i]$
$SOW[i] \leftarrow SOW[i-1] + OW[i]$
$SOL[i] \leftarrow SOL[i-1] + OL[i]$
$SADP[i] \leftarrow SADP[i-1] + ADP[i]$
$\triangleright$ egyenletesen kiosztjuk a döntetleneket
$S \leftarrow SW[n]$
$ND \leftarrow 3NM - S$
$NW \leftarrow S - 2NM$
$NL \leftarrow NW$
$NDP \leftarrow 2(S - 3SOW[n] - SOD[n])/3$
while $NDP > 0$ **do**

4. Példák

Nézzünk néhány példát. 4 csapatra 62 sport sorozat van, közülük 40 foci sorozat. A kiszűrendők egy részében a mod 3 vett maradékok minden döntetlent igényelnek és jelzik a hibát. Az **(1,3,3,9)** sorozatban egy csomagot kell kiosztani – ezt a jelöltek (a 3 pontos csapatok) közül a legkisebb indexű kapja. Az így kapott (1,3) megvalósíthatatlan. A **(2,3,3,6)** sorozatban a két csomagot a 3 pontosok kapják – a (2,3,3) párosíthatatlan. A **(3,3,3,4)** sorozat esetében a 3 csomagot $d[i] = 0$ miatt a 3 pontosok kapják – az (1,3,3,3) nem párosítható. Az egyszerű döntetlenellenőrzés 4 játékosig tökéletesen szűr.

Öt játékosnál már az **(1,1,8,9,9)** döntetlensorozata (1,1,2) – ez párosítható. Az 1 pontosak kötelező döntetlenjét figyelembe vevő (0,2) már mutatja, hogy rossz a sorozat. A **(2,2,3,10,10)** döntetlensorozata (1,1,2) – párosítható. A két 10 pontos közötti kötelező döntetlen miatt (2) adódik, ami már nem párosítható. Hasonló a helyzet a **(2,3,3,8,10)** sorozattal. A javított döntetlenellenőrzés 5 játékosig tökéletes. A megvizsgált esetekben 6 játékosra is jól szűrt.

7 játékosra viszont például az **(1,3,5,7,11,15,15)** döntetlensorozata párosítható. Mivel a pontszámok összege 57, 15 győzelemmel és 6 döntetlen lehet a pont mátrixban. A kötelező sportmátrix
5 - 1
5 - 1
3 2 1
1 1 1
- 2 1
- - 3
- 1 5.

Ehhez 14 győzelemre, 6/2 döntetlenre és 13 vereségre van szükség. Eszerint 1 vereségcsomagot és 2 döntetlencsomagot kell kiosztanunk (az első 2 és az utolsó sorban már van 6 eredmény – a többi sor 1-1 csomagot fogad). A legnagyobb győzelemszám 5 – ehhez adottak a vereségek csomag nélkül is. Az első 2 győzelemszám összege 10. Ehhez az első 2 sor 1, a többi sor összesen 7 vereséggel tud hozzájárulni. Az 1 csomagot legjobban például a negyedik sorban tudjuk felhasználni – de ez is csak 1 új vereség – tehát csak 9 felhasználható vereséget találtunk, a sorozat rossz.

Ezt a sorozatot felbontással is kizárhatjuk. Az első 3 csapatnak csak 3 többletpontja van, az utolsó 2 csapat pedig 3 pontot veszített. A 3 többletpont és (1,3,5) kezdőszelet kétféleképpen érhető el: két belső győzelemmel (1,3,4)-ből, vagy egy belső győzelemmel (1,2,4)-ből. Előbbi esetben az 1 és a 3 pontos eredményeit ismerjük, ezért az (1,1,5,9,9) sorozat marad, ami nem helyi. A második esetben az 1 pontos eredményei ismertek, (1,1,1,5,9,9) marad, ami nem teljes sorozat. A sorozatot kizártuk, ezért a 2 nagy pontszámmal nem is foglalkozunk – azonban az esetszétválasztás szükségessége miatt az ilyen kizárás időigénye problémát jelenthet.

A **(3,3,3,10,13,13,13)** döntetlensorozata egyértelműen párosítható, belső döntetlen nincs. A pontösszeg 58, ezért 16 győzelem és 5 döntetlen van. A kötelező sportmátrix
4 1 1
4 1 1
4 1 1
2 1 -
- - 3
- - 3

- - 3.

Ehhez 14 győzelem és 4/2 döntetlen szükséges, így még két vereség csomagot oszthatunk ki. Az első három csapat győzelemigénye legalább 12. Ehhez ők 3 vereséggel tudnak hozzájárulni. A többi csapat csomag nélkül legfeljebb 6 vereséget tud adni, és ehhez a csomagok csak kettővel tudnak hozzájárulni. Tehát a sorozat rossz.

Ismét jó a felbontás is. Az első 3 játékosnak 3 többletpontja van. Ez származhat a 3 belső döntetlenes (2,2,2)-ből, vagy a belső döntetlenek nélküli (3,3,3)-ból. Utóbbi esetben az (1,4,4,4) sorozat marad, ez pedig nem sport sorozat. Előbbi esetben egy csapatnak sem ismerjük pontosan az eredményeit, viszont a kötelező sportmátrixot ki tudjuk egészíteni:

4 1 1
4 1 1
4 1 1
- 3 3
- 3 3
- 3 3.

Ehhez azonban nincs elég döntetlen, tehát ez a sorozat is rossz. Ismét az esetszétválasztás okozhat gondot.