

Rangiranje ekipa i prognoziranje ishoda u rukometu korištenjem PageRank algoritma

DUŠAN MUNĐAR¹ I DAMIR HORVAT²

Sažetak – Rad pridonosi poznavanju primjene matematičkih modela u analitici sportskih rezultata. U radu rangiramo ekipe koje su igrale na Europskom rukometnom prvenstvu u Poljskoj od 15. do 31. siječnja 2016. i prognoziramo pobjednike rukometnih utakmica. Dominaciju među ekipama modeliramo metodom za rangiranje web stranica – PageRank algoritam – koju ukratko opisujemo. Rezultati modela služe za rangiranje i prognoziranje. Rangiranje dobiveno algoritmom uspoređujemo s ostvarenim plasmanom na kraju prvenstva. Preciznost prognoziranja provjeravamo uspoređujući prognoze s ishodima odigranih utakmica.

Ključne riječi – analitika, prognoziranje, rangiranje, sport

I. Uvod

Analitika podataka može poslužiti u više svrha: za opis sadašnje situacije (deskriptivna analitika), za uspješnije prognoziranje budućih događaja (prediktivna analitika) te, u najboljem slučaju, kao pomoć za donošenje kvalitetnijih odluka (preskriptivna analitika). Analitikom rezultata utakmica dobivamo nove spoznaje o okolnostima nastanka događaja – relativnu snagu momčadi – i predviđanje budućih događaja – uspješnije prognoziranje pobjednika utakmice.

Međunarodna sportska događanja obično izazovu velik interes javnosti. Europsko rukometno prvenstvo odigrano u Poljskoj primjer je takvog događaja. Na kraju prvenstva ljubitelji sporta često se pitaju je li prvenstvo osvojila najbolja ekipa. Odgovor na prethodno pitanje nudimo nakon obrade rezultata utakmica i primjene PageRank algoritama. U nastavku ukratko opisujemo metodu, interpretiramo rezultate i provjeravamo iskoristivost rezultata za prognoziranje ishoda utakmice.

¹Dušan Mundar, Fakultet organizacije i informatike, Varaždin

²Damir Horvat, Fakultet organizacije i informatike, Varaždin

Ciljevi rada:

- i. ukratko opisati ideju i jedan način provedbe PageRank algoritma;
- ii. predstaviti ideju primjene PageRank algoritma za rangiranje rukometnih ekipa;
- iii. prikazati mogućnost primjene rezultata algoritma za prognoziranje pobjednika rukometne utakmice;
- iv. ocijeniti preciznost prognoziranja temeljenog na vrijednostima dobivenima PageRank algoritmom.

II. PageRank algoritam

PageRank algoritam objavila je skupina autora 1999. godine. Dva autora tog rada, Larry Page i Sergey Brin, godinu dana prije objave rada pokrenuli su Google. Početna uspješnost te kompanije povezuje se s navedenim algoritmom.

PageRank algoritam osmišljen je s ciljem određivanja relevantnosti web stranice, a u novije vrijeme koristi se i u druge svrhe, kao primjerice rangiranje nogometnih ekipa. Ideja je bila da se pri pretraživanju pojma web tražilicom, web stranice izlistaju uvažavajući međusobnu relevantnost stranica. PageRank algoritam bazira se na izračunu svojstvenog vektora pridruženog po modulu najvećoj svojstvenoj vrijednosti kvadratne matrice. Algoritam možemo intuitivno shvatiti ukoliko web stranice zamislimo kao vrhove usmjerenog grafa. Tada svaka web stranica predstavlja jedan vrh usmjerenog težinskog grafa. Luk iz vrha Q u vrh P postoji ukoliko stranica Q sadrži link na stranicu P . Za određivanje ranga stranica ključnu ulogu imaju izlazni stupnjevi vrhova. Za vrh P , izlazni stupanj vrha $d^+(P)$ je broj lukova koji počinju u vrhu P . Rang web stranice P tada je realni broj

$$r(P) = \sum_{Q \in B_p} \frac{r(Q)}{d^+(Q)}$$

pri čemu je B_p skup svih web stranica koje imaju link na web stranicu P . Rang stranice P ovisi o rangovima svih stranica Q koje imaju linkove na stranicu P i o ukupnom broju linkova na pojedinoj stranici Q . U slučaju kad je matrica susjedstva usmjerenog grafa ireducibilna, do ranga stranica se može doći iterativnim postupkom. Rang $r_{k+1}(P)$ web stranice P u $(k+1)$ -om koraku dobije se na temelju ranga u k -tom koraku preko formule

$$r_k(P) = \sum_{Q \in B_p} \frac{r_{k+1}(Q)}{d^+(Q)}$$

Za početnu vrijednost ranga $r_0(P)$ može se uzeti vrijednost $1/n$, gdje je n broj stranica koje se rangiraju. Postupak se ponavlja dok komponente vektora

$$r_k = (r_k(P_1), r_k(P_2), \dots, r_k(P_n))$$

ne počnu konvergirati k traženom vektoru r . Više o matematičkom aspektu PageRank algoritma mogli ste pročitati u na početku spomenutom radu [2].

III. Primjena PageRank algoritma za rangiranje rukometnih ekipa

Europsko rukometno prvenstvo igralo se u Poljskoj u siječnju 2016. godine. Na prvenstvu je sudjelovalo 16 momčadi. Prvenstvo se igralo u ciklusima: prvo u skupinama, pa razigravanje po skupinama, te na kraju polufinalne utakmice i utakmice za prvih osam mjesta. Ukupno je odigrano 48 utakmica. Rezultati utakmica dostupni su na stranici www.rezultati.com. Tablica 1. prikazuje broj golova koje je pojedina ekipa zabila nekoj drugoj ekipi. Hrvatska i Norveška te Španjolska i Njemačka odigrale su međusobno po dvije utakmice pa je za njih unesen broj postignutih golova u obje utakmice. Ukoliko se igrao produžetak, unesen je konačni rezultat nakon produžetka. Na primjer, za utakmicu Norveška : Njemačka kod koje je rezultat nakon regularnog vremena bio 27 : 27, unesen je rezultat nakon produžetka 33 : 34.

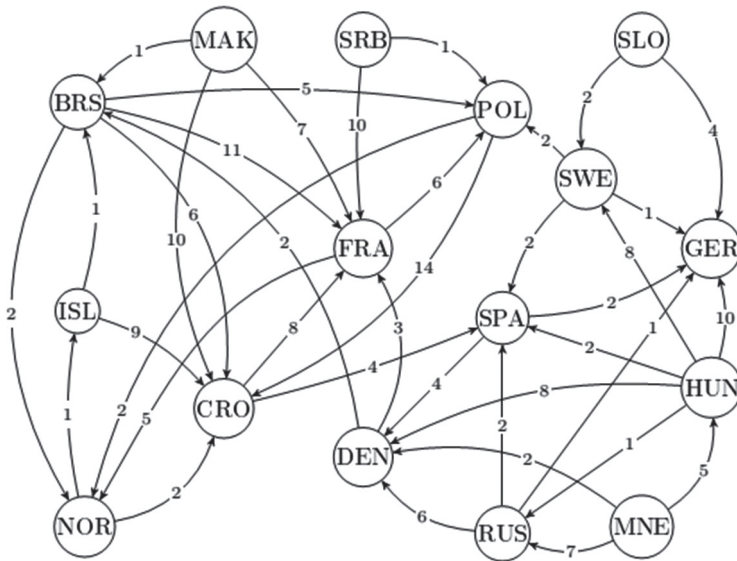
Za potrebe primjene PageRank algoritma, svaka ekipa predstavljena je kao jedan vrh u usmjerenom grafu. Luk od ekipe A prema ekipi B postoji ukoliko je ekipa B pobijedila ekipu A . Dakle, luk ide od slabije momčadi prema boljoj, težina luka je razlika u golovima. Ukoliko je između A i B odigrano više utakmica, smjer i težinu luka određuje prosječna gol-razlika u međusobnim utakmicama. Matematički iskazano, ako je vrijednost

$$v(A, B) = \frac{g(B, A) - g(A, B)}{n(A, B)}$$

veća od nule, gdje je $g(A, B)$ broj golova koje je momčad A zabila momčadi B i $n(A, B)$ broj utakmica koje su odigrale ekipe A i B , onda postoji luk od A prema B i ima težinu $v(A, B)$.

Primjerice, Hrvatska je zabila Poljskoj 37 golova, dok je Poljska zabila Hrvatskoj 23 gola te su stoga vrhovi CRO i POL, koji predstavljaju navedene ekipe, povezani te postoji luk od POL prema CRO s težinom 14. Slika 1 prikazuje usmjereni težinski graf dobiven na temelju razlike postignutih golova među ekipama.

Rangiranje ekipa temeljeno na algoritmu i pozicije na kraju prvenstva prikazani su u Tablici 2. Na temelju relativne dominacije među ekipama dobivene algoritmom dobije se da je Hrvatska umjesto trećeg mjesta zavrijedila drugo mjesto. Neka od iznenađenja prvenstva potvrđena su metodom kao slučajna: Poljska je prema dominaciji trebala završiti na trećem mjestu (umjesto na sedmom); Španjolska je prema dobivenoj dominaciji zavrijedila šesto mjesto (umjesto drugog); najbolja ekipa trebala je biti Francuska koja je završila na petom mjestu. Dakle, na temelju gol-razlika svih utakmica analiziranih PageRank algoritmom, zaključujemo da je najbolja momčad prvenstva bila Francuska. Prema rezultatima, Njemačku bi se pozicioniralo na četvrto mjesto.



Slika 1. Usmjereni graf dominacije među ekipama dobiven na temelju razlike postignutih golova

Izvor podataka: www.rezultati.com

IV. Prognoziranje pobjednika utakmice

Temeljem vrijednosti dobivenih algoritmom možemo pristupiti prognoziranju pobjednika u međusobnom dvoboju dviju ekipa. Očekivanim pobjednikom smatramo ekipu koja je ostvarila veću PageRank vrijednost (bolje rangiranu ekipu). Tablica 3. prikazuje ishode nekoliko utakmica koje su bile odigrane na prvenstvu i prognozu pobjednika dobivenu rangiranjem opisanim algoritmom.

Usporedimo li stvarne ishode svih odigranih utakmica s prognoziranim ishodima koji su dobiveni PageRank algoritmom, možemo procijeniti točnost prognoziranja. Algoritam je omogućio točno određivanje pobjednika u $1+25+9=35$ od 48 slučajeva, odnosno u približno 71% slučajeva. Preciznost prognoziranja najslabija je za utakmice koje su odigrane izjednačenim rezultatom (20%). Izuzmemo li te utakmice iz razmatranja, preciznost prognoziranja naraste na približno 79%.

EKIPA 2																	
	BRS	CRO	DEN	FRA	GER	HUN	ISL	MAK	MNE	NOR	POL	RUS	SLO	SPA	SRB	SWE	UKUPNO
EKIPA 1	BRS	0	21	0	23	0	0	39	30	0	27	27	0	0	0	0	167
	CRO	27	0	0	24	0	0	37	34	0	62	37	0	0	29	0	250
	DEN	0	0	0	26	23	30	0	0	30	0	0	31	0	27	0	195
	FRA	34	32	29	0	0	0	0	30	0	4	25	0	0	0	36	210
	GER	0	0	25	0	0	29	0	0	0	34	0	30	25	53	0	223
	HUN	0	0	22	0	19	0	0	0	32	0	0	26	0	29	0	142
	ISL	38	28	0	0	0	0	0	0	0	26	0	0	0	0	0	92
	MAK	29	24	0	23	0	0	0	0	0	31	23	0	0	0	27	157
	MNE	0	0	28	0	0	27	0	0	0	0	0	21	0	0	0	76
	NOR	29	58	0	29	33	0	25	31	0	0	30	0	0	0	0	235
	POL	32	23	0	31	0	0	0	24	0	28	0	0	0	0	29	193
	RUS	0	0	25	0	29	27	0	0	28	0	0	0	0	23	0	160
SLO	0	0	0	0	21	0	0	0	0	0	0	0	0	24	0	66	
SPA	0	33	23	0	49	31	0	0	0	0	0	25	24	0	0	209	
SRB	0	0	0	26	0	0	0	27	0	0	28	0	0	0	0	81	
SWE	0	0	28	0	26	22	0	0	0	0	24	28	23	22	0	173	
UKUPNO	189	219	180	182	200	166	101	176	90	232	194	161	72	207	92	168	

Tablica 1. Broj golova koji je je Ekipa 1, ekipa zapisana u retku, zabilježila Ekipi 2 zapisanoj u stupcu na Europskom rukometnom prvenstvu 2016.

Izvor podataka: www.rezultati.com

Ekipa	Pozicija ekipa na prvenstvu	Pozicija ekipe prema PageRank rangiranju	PageRank vrijednost
BRS	9-16	12	0,019
CRO	3	2	0,176
DEN	6	7	0,078
FRA	5	1	0,178
GER	1	4	0,104
HUN	9-16	11	0,019
ISL	9-16	8	0,036
MAK	9-16	13-16	0,015
MNE	9-16	13-16	0,015
NOR	4	5	0,097
POL	7	3	0,111
RUS	9-16	10	0,022
SLO	9-16	13-16	0,015
SPA	2	6	0,078
SRB	9-16	13-16	0,015
SWE	8	9	0,024

Tablica 2. Rangiranje ekipa

U trećoj koloni Tablice 3. upisan je prirodni logaritam omjera PageRank vrijednosti $\ln(v_1/v_2)$ za ekipe koje igraju utakmicu. Ako je vrijednost tog broja pozitivna, prognozira se pobjeda prve ekipe; ako je nula izjednačeno, a pobjeda druge ekipe prognozira se ukoliko je vrijednost negativna. Osim za prognoziranje pobjednika, navedena vrijednost može poslužiti za kvantificiranje nadmoći bolje ekipe. Što je broj po apsolutnoj vrijednosti veći, sigurniji smo u prognozu koju dajemo. Primjerice, ukoliko razmotrimo samo slučajeve za koje vrijedi da je apsolutna vrijednost logaritma omjera PageRank vrijednosti veća od 1 (proizvoljno određena granica), prognozu dajemo za 25 od 48 utakmica. Ishodi utakmica i prognoze ishoda tih utakmica prikazani su u Tablici 5. Točnost prognoziranja porasla je u tom slučaju na 84%.

Ekipa 1	Ekipa 2	$\ln(v_1/v_2)$	Pobjednik utakmice	Bolja ekipa (PageRank)
Danska	Mađarska	1,412	1	1
Španjolska	Švedska	1,179	1	1
Njemačka	Slovenija	1,936	1	1
Rusija	Crna Gora	0,383	1	1
Francuska	Poljska	0,472	1	1

Hrvatska	Island	1,587	1	1
Bjelorusija	Island	-0.639	2	2
Makedonija	Norveška	-1,867	0	2
Njemačka	Švedska	1,466	1	1
Crna Gora	Danska	-1,649	2	2

Tablica 3. Prognoziranje pobjednika

		PROGNOZIRANI ISHOD (PREMA PAGERANK VRIJEDNOSTIMA)			
		izjednačeno	pobjeda prve ekipe	pobjeda druge ekipe	ukupno
ISHOD UTAKMICE	izjednačeno	1	1	3	5
	pobjeda prve ekipe	0	25	5	30
	pobjeda druge ekipe	0	4	9	13
	ukupno	1	30	17	48

Tablica 4. Stvarni i prognozirani ishodi utakmica

Od 25 slučajeva točno je prognozirano u 21 slučaj. Izuzmemo li utakmice koje su odigrane izjednačeno, točnost prognoziranja je približno 95% (21 od 22). Procijenjena vjerojatnost da ekipa za koju smo prognozirali pobjedu izgubi je mala. Realiziran je samo 1 od 25 takva slučajeva. Radi se o utakmici između Islanda i Norveške za koju je vrijednost 1.002, a rezultat je bio 26 : 25.

		PROGNOZIRANI ISHOD (PREMA PAGERANK VRIJEDNOSTIMA)			
		izjednačeno	pobjeda prve ekipe	pobjeda druge ekipe	ukupno
ISHOD UTAKMICE	izjednačeno	0	0	3	3
	pobjeda prve ekipe	0	16	1	17
	pobjeda druge ekipe	0	0	5	12
	ukupno	0	16	9	25

Tablica 5. Stvarni i prognozirani ishodi utakmica

V. Zaključak

Sportski turniri imaju razne sustave natjecanja. Zanimljivost turnira je veća što rezultat više ovisi o sposobnosti ekipe u nekom trenutku, a manje o sreći. Sustav natjecanja trebao bi omogućiti rangiranje ekipa prema sposobnosti. Događa se često da osvojeno mjesto nije jednako poziciji koju bi ekipa trebala imati na temelju relativne sposobnosti. PageRank algoritam jedan je način kojim se može dobiti bolja slika o relativnoj snazi momčadi. U radu smo prikazali kako iskoristiti PageRank algoritam za usporedbu ekipa koje su igrale na Europskom rukometnom prvenstvu. Konačni raspored po završetku turnira razlikuje se od rangiranja dobivenog primjenom algoritma. Prema našem rangiranju među tri najbolje ekipe trebale su se naći Francuska, Hrvatska i Poljska. Rezultati modela poprilično dobro služe za prognoziranje, a posebno ako izuzmemo slučajeve u kojima zbog sličnih sposobnosti na rezultat značajno utječe sreća.

Literatura:

1. Anjela Y. Govan, Carl D. Meyer, Ranking national football league teams using google's pagerank, <https://www.ncsu.edu/crsc/reports/ftp/pdf/crsc-tr06-19.pdf>, (7. 7. 2016.)
2. Page, L. et al., *The PageRank Citation Ranking: Bringing Order to the Web.*, Stanford InfoLab, 1999. Dostupno na: <http://ilpubs.stanford.edu:8090/422>, (7. 7. 2016.)
3. Rezultati, <http://www.rezultati.com/rukomet/europa/europsko-prvenstvo-2016/>, (7. 7. 2016.)