Perronova-Frobeniova věta a hodnocení (fotbalových) týmů

1. Úvod

Fotbalová liga skončila a na prvním místě se umístil tým FC Viktoria Plzeň. Ale někteří lidé si myslí, že Sparta, nebo Jablonec byli lepší, tak proč mistrovský titul přebírala Plzeň? V tomto článku si ukážeme některé pokročilejší metody hodnocení, k tomu využijeme *Perronovu-Frobeniovu* větu, *vlastní čísla* a *vlastní vektory* nezáporných *ireducibilních* matic.

V tomto textu popisované metody hodnocení mají širokou škálu uplatnění - nemusí se používat pouze pro fotbalové týmy, ale dají se uplatnit např. na sestavení žebříčku, který tenista byl v historii nejlepší, pro porovnávání zubních past apod.

Na konci práce ukážu konkrétní aplikaci na českou fotbalovou ligu ročník 2014/2015.

2. Turnaje

V turnaji s n hráči se systémem každý s každým odehraje každý účastník jeden zápas (popř. dva, pokud se hraje i odveta) proti ostatním $n-1$ hráčům. Pro jednoduchost uvažujme turnaj, kde se nehraje na remízy. Výsledky zápasů budeme zapisovat do turnajové matice $A = (a_{ij})_{n \times n}$ následovně: nejprve účastníkům přiřadíme čísla $1, 2, ..., n$, potom definujeme prvek na místě i, j v matici A takto:

$$a_{ij} = \begin{cases} 1 & \text{pokud tým } i \text{ porazil tým } j \\ 0 & \text{pokud tým } i \text{ prohrál s týmem } j. \end{cases}$$

Pokud $i \neq j$, pak je právě jeden z prvků a_{ij}, nebo a_{ji} nenulový a $a_{ii} = 0$. Kdybchom uvažovali zápasy s remízami, tak by se složky matice A vhodně upravily - např. 2 body za výhru, 1 za remízu a žádný za prohru.

Jako nejjednodušší způsob hodnocení týmů se nabízí vytvořit pořadí podle počtu vítězství. Počet vítězství týmu i dostaneme jako součet hodnot i-tého řádku. Vektorem skóre $\mathbf{s}$ označíme vektor součtů řádků - tedy $\mathbf{s} = A\mathbf{1}$ *. Potom řekneme, že tým i je silnější než tým j, jestliže $\mathbf{s}_i > \mathbf{s}_j$. Pro jaké turnaje dává tento způsob hodnocení jednoznačné řešení?

Nechť H_n označuje horní trojúhelníkovou matici s nulami na hlavní diagonále a samými jedničkami nad hlavní diagonálou. Takováto matice odpovídá turnaji, kde i-tý tým porazí j-tý, kdykoliv je $i < j$. Vektor skóre potom vypadá takto $(n-1, n-2, ..., 1, 0)^T$, každý tým má unikátní skóre a tedy tento vektor vytváří jednoznačné výsledné skóre. Jediné turnajové matice, kde vektor skóre unikátně ohodnocuje týmy, jsou právě takové, které jsou permutací matice H_n.

Jak vyřešit hodnocení týmů v turnaji, kde dva a více týmů mají stejné skóre?

3. Relativní síla týmu

Každému týmu budeme chtít přiřadit jeho *relativní sílu* na základě odehraných zápasů proti ostatním týmům. Do *síly* týmu i budeme brát v úvahu nejen výsledek střetnutí proti ostatním týmům, ale i *sílu* protivníků, se kterým tým i získal alepoň bod. Naším cílem nyní je ukázat, že vektor sil $\mathbf{v}$ dostaneme jako řešení rovnice

$$A\mathbf{v} = r\mathbf{v}, \tag{1}$$

kde r je vlastní číslo matice A. Tedy, že vektor sil je kladným vlastním vektorem turnajové matice A. K tomu budeme potřebovat několik pojmů.

Ireducibilní matice Čtvercová matice M se nazývá ireducibilní, pokud je tvaru 1×1 nebo pokud neexistuje žádná *permutační matice*† P taková, že matice PMP^T je v blokově horním diagonálním

*Vektor $\mathbf{1}$ značí vektor příslušné velikosti složený ze samých jedniček, tj. $\mathbf{1}=(1, 1, ..., 1)^T$

†Permutační maticí nazýváme matici, která vznikla z jednotkové matice záměnou sloupců

tvaru s dvěma nebo více čtvercovými bloky, tj.:

$$PMP^T \neq \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}, \tag{2}$$

kde B a D jsou čtvercové bloky rozměru alespoň 1. Pro lepší představu přikládám ještě jednu ekvivalentní definici -nahlížejme na matici M jako na matici sousednosti orientovaného grafu. Potom matice M je ireducibilní právě tehdy, když je její graf D_A *silně souvislý*[*].

Není-li matice ireducibilní, tak potom je reducibilní.

Předpokládejme nyní, že A je $n \times n$ turnajová matice a existuje celé číslo k a permutační matice P taková, že

$$PAP^T = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix}, \tag{3}$$

kde X má velikost $k \times k$ a Z je $(n-k) \times (n-k)$ - matice A je tedy reducibilní. Jelikož A je turnajová matice, tak jsou i PAP^T, X a Z turnajové; Y má velikost $k \times (n-k)$ a skládá se ze samých jedniček. Skóre prvních k týmů (myšleno od shora, viz PAP^T) musí být alespoň $n-k$, posledních $(n-k)$ skóre může být maximálně $n-k-1$. Proto prvních k týmů má větší skóre než posledních $n-k$ a tedy každý z prvních k týmů se celkově umístí lépe než zbylých $n-k$ týmů.

Odtud vyplývá, že pokud chceme ohodnocovat týmy v takovémto turnaji, tak se stačí zaměřit na dvě skupiny týmů odpovídajícím dvěma diagonálním blokům v PAP^T. Pokud jeden z těchto dvou bloků není ireducibilní, tak nalezneme permutační matici, která tento blok převede na horní blokově trojúhelníkový tvar, jehož diagonální bloky jsou už ireducibilní [1]. A tedy k pochopení jak porovnat týmy v obecném turnaji, se zužuje na porozumění hodnocení týmů v turnaji s ireducibilní turnajovou maticí.

Spektrální poloměr Spektrálním poloměrem čtvercové matice A označujeme maximum z absolutních hodnot vlastních čísel matice A. Často se spektrální poloměr matice A značí $\rho(A)$, tj.

$$\rho(A) = \max\{|\lambda| : \lambda \text{ je vlastní číslo matice } A\} \geq 0$$

Právě nabyté pojmy uplatníme na velmi důležitou větu, která nám umožní poměrně snadný výpočet *relativní síly* týmů.

Perronova-Frobeniova věta[1]

Nechť A je nezáporná ireducibilní matice řádu n. Potom platí

1. existuje kladné reálné vlastní číslo r matice A (zvané Perronovo-Frobeniovo vlastní číslo, nebo také Perronův kořen) takové, že pro všechny ostatní vlastní čísla matice A platí

$$|\lambda| \leq r, \qquad \text{tj. } r \text{ je spektrální poloměr matice } A,$$

2. algebraická i geometrická násobnost r je jedna,

3. vlastní vektor příslušný k r je kladný [†],

4. vlastní vektor $\mathbf{v} = (v_1, ..., v_n)^T$ příslušný r, jehož (eukleidovská) norma je 1, je určený jednoznačně a nazývá se *Perronův vektor*,

5. libovolný kladný (resp. záporný) vlastní vektor matice A je lineární kombinací (v tomto případě tedy násobkem) Perronova vektoru.

[*]Orientovaný graf $D = (V, E)$ se nazývá *silně souvislý*, pokud pro každou dvojici vrcholů $i, j \in V$ existuje orientovaná cesta z i do j i z j do i

[†]Kladným vektorem rozumíme vektor, jehož všechny složky jsou kladné

Prozkoumáme, jakou roli hraje *Perronův vektor* při hodnocení týmů.

Jelikož silný tým by měl porážet týmy, které porazí hodně týmů, tak mějme, že *síla* týmu i (značme x_i) bude určena součtem skóre týmů, které tým i porazil. . Potom tedy

$$x_i = \sum_{j\text{:i porazil j}} s_j, \tag{4}$$

kde s_j je skóre j-tého týmu. Jelikož i porazí j právě, když $a_{ij} = 1$, tak dostáváme, že *síla* týmu i se dá vyjádřit takto:

$$x_i = \sum_{j\text{:i porazil j}} s_j = \sum_{j=1}^{n} a_{ij} s_j = \sum_{j=1}^{n} \left(a_{ij} \sum_{k=1}^{n} a_{jk} \right) = \sum_{k=1}^{n} \sum_{j=1}^{n} a_{ij} a_{jk}, \tag{5}$$

což je přesně součet všech prvků v i-tém řádku matice A^2! Tedy $A^2\mathbf{1}$ je vektor, jehož i-tá složka odpovídá součtu skóre všech týmů, které i porazil.

Zamysleme se nyní nad tím, jaké vektory hodnocení jsme již uvažovali - nejprve to byl vektor $A\mathbf{1}$, jehož složky jsou příslušné počty poražených týmů, což můžeme brát jako první stupeň měření síly týmů, dále potom vektor $A^2\mathbf{1}$, jehož složky odpovídají příslušným sumám skóre poražených týmů, zajišťující druhý stupeň měření síly týmů. Zdá se logické uvažovat vektor $A^3\mathbf{1}$, jehož složky odpovídají příslušným součtům součtů skóre poražených týmu. Podobně můžeme uvažovat vektory $A^4\mathbf{1}$, $A^5\mathbf{1}$, nebo dokonce $A^k\mathbf{1}$, kde k je libovolné kladné celé číslo. S rostoucím k se na složkách vektoru $A^k\mathbf{1}$ více projevují vztahy mezi jednotlivými týmy (více se projevuje, jaké soupeře porazil soupeř, kterého tým i porazil). Proto, čím větší k, tím přesnější způsob ohodnocení týmů. Jelikož složky vektoru $A^k\mathbf{1}$ se zvětšují poměrně rychle a nás zajímá pouze relativní síla týmů (vůči ostatním), tak můžeme velikost složek (v poměru) libovolně upravovat, např. aby vektor $A^k\mathbf{1}$ byl jednotkový. Jak velké zvolit k, abychom dostali přesné výsledky relativní síly týmů? Jak jsme psali výše - čím větší, tím přesnější výsledek bude. Nyní ukážeme, jak spočítat hledaný vektor síly $\mathbf{v}$ pro $k \to \infty$.

Mějme ireducibilní turnajovou matici A. Potom posloupnost vektorů

$$\frac{A\mathbf{1}}{||A\mathbf{1}||}, \frac{A^2\mathbf{1}}{||A^2\mathbf{1}||}, \dots, \frac{A^k\mathbf{1}}{||A^k\mathbf{1}||}, \quad k \in \mathbb{N}, \tag{6}$$

konverguje k Perronově vektoru $\mathbf{v}$ matice A, neboť posloupnost (6) přesně odpovídá *mocninné metodě*, která slouží k nalezení vlastního vektoru příslušného vlastnímu číslu s největší absolutní hodnotou - a jelikož jsou splněny předpoklady Perronovy věty, tak posloupnost v (6) konverguje ke kladnému vektoru příslušného největšímu vlastnímu číslu matice A, tj. k Perronově vektoru. Tedy vektor relativních sil týmů, značíme $\mathbf{v}$, dostaneme jako řešení rovnice $A\mathbf{v} = r\mathbf{v}$, kde r je vlastní číslo matice A s největší absolutní hodnotou. Tato metoda hodnocení hráčů, tj. hledání jejich relativní síly, se nazývá **Kendall-Wei ranking**[3,4].

Podobně můžeme pracovat se sloupci matice A a spočítat *levý Perronův vektor* (značme $\mathbf{w}$), tj. vektor s kladnými složkami splňující rovnici

$$\mathbf{w}^T A = r\mathbf{w}^T \text{nebo ekvivalentně } A^T\mathbf{w} = r\mathbf{w}. \tag{7}$$

Jelikož sumy sloupců označují počet proher, tak potom složky vektoru $\mathbf{w}$ udávají relativní slabost týmů.

Jiná metoda hodnocení (metoda, kterou představil C. Ramanujacharyulu[5]) vytváří hodnocení týmů na základě poměrů relativní síly a relativní slabosti týmů - odtud její název **Power-Weakness Ratio (PWR)**. Výhodou PWR oproti předchozím metodám je to, že výhra a prohra jsou ceněny stejně, neboť *síla* je vydělena *slabostí.*

4. Aplikace

Poznámka Pro počítání vlastních vektorů využívám software *MATLAB*

4.1

Uvažujme ping-pongový turnaj o čtyřech hráčích, kde každý hráč hrál proti každému jeden zápas. Výsledky jsou zaneseny do turnajové matice M (za výhru 1 bod, za prohru 0):

$$M = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \tag{8}$$

kde první řádek odpovídá hráči A, druhý hráči B, třetí hráči C a poslední řádek patří hráči D. První zmíněná metoda v tomto článku počítala vektor skóre, který v tomto turnaji vypadá takto:

$$\mathbf{s} = (2, 1, 2, 1)^T. \tag{9}$$

Vidíme, že hráči A a C mají stejný počet bodů, stejně tak hráči B a D. Takovýto výsledek pořadatele turnaje moc nepotěší, protože by musel předat dvě ceny pro vítěze. Jak si s takovou situací poradí *Kendall-Wei ranking*?

Určíme relativní síly hráčů. K tomu nám stačí spočítat *Perronův vektor* $\mathbf{v}$, jehož složky budou relativní síly týmů vyjadřovat. Jak již víme, tak *Perronův vektor* je kladný vlastní vektor matice M příslušný vlastnímu číslu s největší absolutní hodnotou, tedy:

$$\mathbf{v} = (0.6256, 0.3213, 0.5516, 0.4484)^T. \tag{10}$$

Při použití metody *Kendall-Wei ranking* již rovnosti mezi hráči nejsou - nejsilnější tým je A, následovaný týmy C, D a poslední skončí B.

Pro použití metody *PWR* potřebujeme navíc spočítat vektor slabostí $\mathbf{w}$. Ten spočítáme dle rovnice (7) a dostaneme:

$$\mathbf{w} = (0.4484, 0.5516, 0.3213, 0.6256)^T. \tag{11}$$

Vektor $\mathbf{w}$ ilustruje jednu zajímavou vlastnost KW-metody - záměna výher za prohry a naopak, nemusí nutně otočit pořadí. Výsledný vektor $\mathbf{t}$ poměrů relativní síly ku relativní slabosti a tedy výsledek podle PWR metody potom je:

$$\mathbf{t} = (1.3951, 0.5824, 1.7167, 0.7167)^T. \tag{12}$$

Poslední dva výsledky (vektor $\mathbf{w}$ určující relativní slabost týmů a vektor $\mathbf{t}$ vyjadřující hodnocení metodou PWR) hodnotí hráče C lépe než A. Je to odměna pro hráče C za vyrovnané výsledky, tj. neprohrát se slabým soupeřem.

4.2

Poslední srovnání, jak jednotlivé metody fungují, jsem připravil na české fotbalové lize ročníku 2014/2015. Z výsledků sestavíme turnajovou matici jako v předešlém případě - pro lepší přehlednost sestavíme dvě tabulky - první bude odpovídat turnajové matici, druhá tabulka bude obsahovat příslušné ohodnocení stejnými třemi metodami (resp. čtyřmi, počítáme-li i relativní slabost) jako v sekci 4.1. Postup zde nebudu uvádět, neboť byl v předchozích sekcích dostatečně vysvětlen.

První tabulka obsahuje body, kolik který tým na jaký uhrál

ČFL 2014/2015	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
FC Viktoria Plzeň	0	6	6	3	4	6	4	3	6	3	3	4	6	6	6	6
AC Sparta Praha	0	0	4	6	6	3	4	4	6	6	6	3	1	6	6	6
FK Jablonec	0	1	0	4	6	3	4	6	4	4	4	6	6	6	6	4
FK Mladá Boleslav	3	0	1	0	3	3	3	4	4	6	6	2	3	3	2	3
1. FK Příbram	1	0	0	3	0	4	2	0	3	4	3	4	4	3	6	6
Dukla Praha	0	3	3	3	1	0	3	1	1	6	1	4	2	4	6	3
FK Teplice	1	1	1	3	2	3	0	3	4	3	4	4	4	1	1	3
Bohemians Praha	3	1	0	1	6	4	3	0	0	4	4	1	0	4	1	6
1. FC Slovácko	0	0	1	1	3	4	1	6	0	6	3	1	3	4	0	4
FC Vysočina Jihlava	3	0	1	0	1	0	3	1	0	0	6	4	4	3	6	4
Sk Slavia Praha	3	0	1	0	3	4	1	1	3	0	0	6	4	0	4	4
FC Slovan Liberec	1	3	0	2	1	1	1	4	4	1	0	0	4	3	4	4
FC Baník Ostrava	0	4	0	3	1	2	1	6	3	1	1	1	0	4	3	3
Zbrojovka Brno	0	0	0	3	3	1	4	1	1	3	6	3	1	0	3	4
FC Hradec Králové	0	0	0	2	0	0	4	4	6	0	1	1	3	3	0	1
Dynamo Č. Budějovice	0	0	1	3	0	3	3	0	1	1	1	1	3	1	4	0

Druhá tabulka obsahuje výsledky hodnocení metodami, které jsme v článku probírali. Sloupec *body* obsahuje skóre podle kterého je v naší lize tvořena výsledková tabulka, zároveň to odpovídá trochu vylepšené první metodě, kterou jsme v tomto textu popisovali - při rovnosti bodů je rozhodující skóre. V závorkách u pořadí je posun oproti oficiálním výsledkům; znaménko - znamená propad a znaménko + naopak polepšení

ČFL 2014/2015	body	pořadí	KW ranking	pořadí	slabost	pořadí	PWR	pořadí
FC Viktoria Plzeň	72	1	0.4355	1	0.1029	2 (-1)	4.2322	1
AC Sparta Praha	67	2	0.3768	2	0.1045	3 (-1)	3.6057	3 (-1)
FK Jablonec	64	3	0.3495	3	0.0916	1 (+2)	3.8155	2 (+1)
FK Mladá Boleslav	46	4	0.2702	4	0.2133	5 (-1)	1.2667	4
1. FK Příbram	43	5	0.2256	6 (-1)	0.2118	4 (+1)	1.065	5
Dukla Praha	41	6	0.2381	5 (+1)	0.2394	6	0.9946	6
FK Teplice	38	7	0.2239	8 (-1)	0.2486	7	0.9006	7
Bohemians Praha	38	8	0.2249	7 (+1)	0.2603	8	0.8640	8
1. FC Slovácko	37	9	0.2058	9	0.2631	10 (-1)	0.7822	9
FC Vysočina Jihlava	36	10	0.2002	10	0.2697	11 (-1)	0.7423	10
Sk Slavia Praha	34	11	0.1989	11	0.2857	13 (-2)	0.6962	12 (-1)
FC Slovan Liberec	33	12	0.1920	13 (-1)	0.2605	9 (+4)	0.7370	11 (+1)
FC Baník Ostrava	33	13	0.1975	12 (+1)	0.2874	14 (-1)	0.6870	13
Zbrojovka Brno	33	14	0.1769	14	0.2828	12 (+2)	0.6256	14
FC Hradec Králové	25	15	0.1408	15	0.3343	15	0.4211	15
Dynamo Č. Budějovice	22	16	0.1263	16	0.3564	16	0.3544	16

I přesto, že bodový rozdíl ve středu tabulky je poměrně malý, tak KW ranking i PWR až na malé změny odpovídají oficiálnímu hodnocení. Je to dáno tím, že při rovnosti bodů rozhoduje skóre, které by silnější tým měl mít lepší než slabší. Zajímavý je posun Jablonce v pořadí u *slabosti* a PWR na první, respektive na druhé místo. Toto je dáno tím, že Jablonec neztrácel body proti slabým soupeřům - viz sloupec 3 první tabulky.

V těchto hodnoceních se nachází míra subjektivity, kterou jsme zatím přehlíželi - kolik bodů přiřadit za jaký výsledek? Dát 3 body za výhru, 1 za remízu a 0 za prohru, nebo jsou lepší jenom 2 body za výhru? Na tohle neexistuje jednoznačná odpověď - záleží na osobních, resp. pořadatelových preferencích - 3 body za vítězství by měly týmy více tlačit do útočení, neboť remíza a prohra znamenají v tomto systému poměrně velkou ztrátu, naopak v systému se dvěma body při vítězství se může proti silnému soupeři vyplatit hrát na remízu.

Metody Kendall-Wei i PWR se také dají používat k sestavení žebříčku hráčů/týmů, kteří spolu nehráli. Postup by byl stejný, jak jsme popisovali v sekci 3.

Závěr

V tomto textu jsme si představili tři metody hodnocení - klasické hodnocení podle počtu bodů, poté jsme si důkladně vysvětlili metodu Kendall-Wei ranking založenou na porovnávání relativních sil hráčů a na konec Ramanujacharyuluvu metodu PWR (Power-Weakness Ratio), která každému týmu přiřadí hodnotu poměru mezi jeho relativní sílou a slabostí. Prvně jmenovaná metoda je užitečná pro porovnávání týmů, u kterých nedochází k rovnosti bodů. Pokud při použití této metody dochází k častým remízám (více týmů se stejným skóre), je vhodné použít některou ze zbylých dvou metod - Kendall-Wei ranking nebo PWR. Mně osobně se více zamlouvá metoda PWR, neboť v hodnocení týmu i je započtena jak síla týmů, které i porazil, tak i s jak slabými soupeři prohrál.

5. Odkazy v textu

[1] Tvrzení, že každou čtvercovou matici lze pomocí vhodné permutační matice převést na horní blokově trojúhelníkový tvar, jehož diagonální bloky jsou již ireducibilní, přejato odtud:
`https://is.muni.cz/th/63999/prif_m/Diplomka.txt`: Věta 1.2.2.

[2] Znění Perronovy-Frobeniovy věty převzato odsud:
Lekture 12, The Perron-Frobenius theorem, Shlomo Sternberg:
`http://www.math.harvard.edu/library/sternberg/slides/1180912pf.pdf`

[3] M. G. Kendall, Further contributions to the theory of paired comparisons, Biometrics 11 (1955)

[4] T. H. Wei, The algebraic foundations of ranking theory, Cambridge University Press (1952)

[5] C. Ramanujacharyulu, Analysis of preferential experiments, Psychometrica 29 (1964)

6. Zdroje

[1] James P. Keener, The Perron-Frobenius Theorem and the ranking of football teams, SIAM Review, Vol. 35, No. 1 (Mar., 1993)

[2] Carolyn Eschenbach et al., Properties of Tournaments among Well-Matched Players, The American Mathematical Monthly, Vol. 107, No. 10 (Dec,. 2000)

[3] H. A. David, Ranking the players in a round robin tournament, Review of the International Statistical Institute, Vol. 39, No. 2 (1971)

[4] Herbert S. Wilf, Searching the web with eigenvectors:
`https://www.math.upenn.edu/~wilf/website/KendallWei.pdf`

[5] Drew Armstrong, Tournaments:
`http://www.math.umn.edu/~reiner/Classes/Tournaments.pdf`

[6] Jiří Tůma, přednáška O vyhledávači Google (2.4.2015):
`http://www.karlin.mff.cuni.cz/~tuma/Aplikace15/Ukazky2015_Google_4.pdf`

Články [1]-[3] jsou dostupné v digitální podobě na `http://www.jstor.org/`

Tento text vytvořil Daniel Štumpf pro Ukázky aplikací matematiky, MFF UK 2015