Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE

Martin Hána

Vlastní čísla nezáporných matic, vyhledávač Google a žebříčky nejlepších šachistů, tenistů a národních fotbalových týmů

Katedra algebry

Vedoucí bakalářské práce: Doc. RNDr. Jiří Tůma, DrSc.
Studijní program: Matematika, obor: Obecná matematika

2006

Děkuji Doc. RNDr. Jiřímu Tůmovi, DrSc. za odborné vedení práce.

Prohlašuji, že jsem svou bakalářskou práci napsal(a) samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 30.6.2006 Martin Hána

Obsah

1 Systém ELO **5**
1.1 Úvod 5
1.2 Zakladní formule systému ELO 6
1.3 Provizorní ohodnocení 8
1.4 Žebříčky FIDE a USCF 8
1.5 Další charakteristiky systému 9

2 Teorie ohodnocení Kendall-Wei **12**
2.1 Idea metody a vyhledávač Google 12
2.2 Existence nezáporného vlastního vektoru 14
2.3 Reducibilní matice a Kendall-Wei 20

3 Žebříček FIFA **23**
3.1 Současný oficiální FIFA žebříček 23
3.2 Žebříčky pomocí Kendall-Wei 24
3.3 Porovnání úspěchu předvídání 29

4 Přílohy **31**

Literatura **38**

Název práce: Vlastní čísla nezáporných matic, vyhledávač Google a žebříčky nejlepších šachistů, tenistů a národních fotbalových týmů
Autor: Martin Hána
Katedra (ústav): Katedra algebry
Vedoucí bakalářské práce: Doc. RNDr. Jiří Tůma, DrSc.
e-mail vedoucího: tuma@karlin.mff.cuni.cz

Abstrakt: V předložené práci studujeme dvě metody ohodnocování a jejich aplikace. První, systém ELO popíšeme při konkrétním použití na vytváření žebříčku mezinárodní organizace šachistů FIDE. Druhá metoda teorie ohodnocení Kendall-Wei využívá existenci vlastních vektorů matice, konktrétně Perronova vektoru. Jeho existenci a dalším trzením Perron-Frobeniovy teorie, jež použijeme při implementaci metody, se budeme věnovat ve druhé kapitole. Dále popíšeme současný způsob vytváření žebříčku národních fotbalových týmů FIFA a použijeme metodu Kendall-Wei na tvorbu vlastních žebříčků.
Klíčová slova: Kendall-Wei, Elo, FIFA

Title: Eigenvalues of non-negative matrices, Google Search Engine, and ratings of best chess players, tennis players and national soccer teams
Author: Martin Hána
Department: Department of Algebra
Supervisor: Doc. RNDr. Jiří Tůma, DrSc.
Supervisor's e-mail address: tuma@karlin.mff.cuni.cz

Abstract: In the present work we study two ranking methods and their applications. First one, ELO ranking system, is used for creating rating of chess players associated in organization FIDE. Second one, Kendall-Wei theory of ranking, use a existence of eigenvectors, specifically the Perron vector. Its existence and other theorems of Fobenius-Perron theory,which we will need by implementation, we write about in second chapter. We will describe recent way of creating ranking schedule of nation football teams FIFA and use Kendall-Wei method to create alternative ones.
Keywords: Kendall-Wei, Elo, FIFA

Kapitola 1

Systém ELO

1.1 Úvod

Systém Elo spadá do skupiny systémů používaných na ohodnocení prvků určité množiny na základě dat, které získáme, pokud porovnáváme dva jednotlivé prvky množiny. Tyto systémy mají široké použití, pro příklad můžeme uvést jejich využití v experimentální psychologii. Další užití ukážeme v kapitole o teorii Kendall-Wei. System Elo se také používá na tvorbu žebříčků jednotlivých hráčů sdružených v různých sportovních organizacích. Porovnání jednotlivých hráčů organizace nám přináší jejich vzájemné zápasy, pomocí nichž můžeme různým způsobem a pomocí různých kritérií určit preferenci jednoho hráče nad druhým.

Dále se budeme věnovat použití Elo systému na vytvoření žebříčku šachistů, kteří jsou členy mezinárodní šachistické organizace FIDE. Všechny informace jsou čerpány ze článků [1] a [2]. Od roku 1970 tato organizace vytváří svůj oficiální žebříček pomocí tohoto systému. Praktický užitek takového žebříčku je jasný. Je používán jako pomůcka při nasazování hráčů do turnaje, tak aby se předešlo vzájemným střetnutím silných hráčů už v počátečních kolech. Dále potom pomáhá hráčům a lidem, kteří se o sport zajímají, sledovat pokrok toho kterého hráče, kterého během času dosáhl.

Systém každému hráči přiřazuje číselné ohodnocení v rozsahu od nuly do 3000. Jeho změna v čase je dána pouze výsledky, kterých hráč při svých vystoupeních na různých turnajích dosáhl. Jak uvidíme později, číselné ohodnocení vytvořené pomocí systému Elo ,může po odehraném turnaji vzrůst ale i klesnout. Tím se liší například od systému, který používá pro svůj žebříček americká organizace hráčů bridge (ACBL). Ten po odehrané vyhraném zápase přidá hráči počet bodů odpovídají síle poraženého soupeře. Takový systém přímo netrestá za prohru a počet bodů v žebříčků může pouze růst. Ohodnocení v takovém případě není ale pouze vyjádření hráčových schopností, ale také frekvence, se kterou se účastní různých turnajů. Pro ilustraci jiných přístupů k tvorbě žebříčků se ještě zmíníme o žebříčku ATP, tedy o ohodnocení tenistů. Body získané z turnaje závisí hlavně na kvalitě turnaje, což je reprezentováno množstím pěněz, které je možné za úspěšné účinkovaní na turnaji získat, dále pak na tom, jak daleko hráč turnajem prošel, předtím než z něj byl vyřazen. Ohodnocení hráče je pak součet nejlepších 14 turnajových výsledků , to vše však pouze za za poslední rok. Pokud se hráč účastnil v dané době méně než 14 turnajů počítá se suma ze všech jeho vystoupení. Je zde tedy zastoupen i časový faktor . Pokud se hráč tedy například kvůli ztrátě formy neučastní dostatečně kvalitních turnajů, může jeho ohodnocení i klesnout.

1.2 Zakladní formule systému ELO

Každému hráči, který je členem organizace FIDE, přísluší určité ohodnocení. Toto ohodnocení je odhadem jeho skutečné síly. Ta může být ve skutečnosti odlišná. Pokud dojde ke zlepšení hráče, jeho staré ohodnocení už nevyjadřuje jeho skutečnou sílu a mělo by se tedy změnit. Systém by měl toto zlepšení zaznamenat a hráče posunout bodově vzhůru. Předpokládejme utkání dvou hráčů A a B s příslušejícími ohodnoceními R_A,R_B. Tyto hodnoty tedy reprezentují reálnou sílu obou soupeřů. Předpokládejme také zatím, že zápas je vždy rozhodnut pro jednoho z hráčů. Vždy tedy končí vítěstvím jednoho z hráčů, ten za vitěztví získává skóre 1, poražený tým nezíská nic (skóre nula). Pomocí formule

$$E = \frac{10^{R_A/400}}{10^{R_A/400} + 10^{R_B/400}} \tag{1}$$

získáme takzvané očekáváné skóre pro hráče A. Jeho interpretace je taková, že je dlohodobým průměrem. Pokud tedy tito dva hráči sehrají větší množství utkání a jejich schopnosti se přitom nezmění, pak hráč A dosáhne (při předchozím ohodnocení vitěství) skóre, které bude velmi blízko hodnotě E. Pokud je tedy například hodnota E rovná 0.25 , potom při velkém množství vzájemných zápasů hráč A vyhraje přibližně 25% všech utkání. Pokud připustíme i remízy, což je v šachu určitě dost častá situace, bude toto procento menší, protože část očekáváného skóre hráč získá za ně.

Problémem zůstává to, že pokud jde o hodnoty R_A a R_B, známe ve skutečnosti pouze odhady reálné síly jednotlivých hráčů. Ukazuje se, formule (1) skutečné výsledky nadhodnocuje. Průměrný počet bodů získaný hráčem A ze zápasů s hráčem B je menší než skóre očekávané formulí. Hlavním důvodem tohoto chování je právě to, že R_A a R_B nejsou odrazem reálné síly hráčů v každém utkání. Předpokládejme hráče A s reálnou silou 1900 a hráče B se silou 1700. Výkonnost hráče B bude v utkáních kolísat. Přibližně polovinu utkání tak hráč B odehraje se silou 1600 a polovinu se silou 1800. Očekávané skóre hráče A počítané formulí (1) je pro soupeře se silou 1700 0.76, pro soupeře se silou 1600 0.85 a pro soupeře se silou 1800 0.64. Protože polovinu utkání hráč B odehrál se sílou 1600 a polovinu se sílou 1800 je očekávané získané skóre hráče A $\frac{0.64+0.85}{2} = 0.745$, což je mémě než skóre, které očekáváme pokud by hráč B držel svou výkonnost stabilní na 1700.

Protože se výkonnost hráčů mění, a v některém okamžiku může být i špatně ohodnocena, musíme dostat nástroj, který by ohodnocení upravoval. Při úpravě můžeme dát různý důraz na předchozí ohodnocení(kteému odpovídají starší výsledky) a výsledky současné . Počítání ohodnocení ze všech i hodně starých výsledků by bylo náročné naopak jejich ignorování by nás připravilo o cenné informace. Kompromisem je tak následující formule.

$$r_{post} = r_{pre} + K(S - S_{ocek}) \tag{2}$$

Ohodnocení r_{post} je ohodnocení po provedení korekce, r_{pre} ohodnocení před jeho provedením, K je kladná konstanta. S je celkové skóre získané na turnaji a S_{ocek} očekávané skóre získáné pomocí formule(1),ve které nahradíme realné ohodnocení jejich odhady.

Podíváme se na interpretace členu $S - S_{ocek}$. Pokud je získané skóre větší než skóre, které jsme očekávali a které odpovídá ohodnocení r_{pred}, musí být reálná síla hráče větší a jeho ohodnocení se musí zvětšit. To odpovídá tomu, že je člen $S - S_{ocek}$ větší než nula a platí tedy v takovém případě $r_{post} > r_{pre}$. Pokud je $S - S_{ocek} < 0$ potom se ohodnocení hráče zmenší. Zároveň je vidět, že čím je absolutní hodnota rozdílu $S - S_{ocek}$ větší, tím horší předpověď jsme od formule (1) dostali. Formule(2) zabezpečí, že tím víc se v takovém případě posune hráčovo ohodnocení. Pokud byla predpověd přesná, hráčovo ohodnocení je přesné a nepotřebuje změnu. Tomu odpovídá nulový rozdíl $S - S_{ocek}$.

Faktor K v formuli(2) se dá interpretovat jako váha, kterou ve formuli dáváme posledním výsledkům, které jsou reprezentovány výrazem $S - S_{ocek}$. Čím větší tuto konstantu zvolíme, tím důležitější budou poslední hráčova vystoupení. Díky formuli(2) má tedy systém Elo schopnost sledovat hráčův vývoj v čase. Tento postup selhává, pokud je rozdíl získaného a očekávaného skóre příliž velký. To může nastat, jestliže se hráč dlouho žádného turnaje nezúčastnil a nedal tak možnost své ohodnocení zkorigovat. V takovém případě formule jenom špatně odhaduje jeho nové ohodnocení.

1.3 Provizorní ohodnocení

Pokud hráč teprve vstupuje do turnajů organizovaných oranizací, je mu přiděleno provizorní ohodnocení. Toto ohodnocení je počítáno pouze na základě výsledků několika jeho vystoupení a proto má menší schopnost odhadovat reálnou sílu hráče a následně tak předpovídat jeho výsledky. Organizace FIDE používá různé formule na výpočet provizorního ohodnocení v období šesti měsíců od doby, kdy se hráč začal turnajů účastnit. Jiná šachistická organizace, americká USCF počítá hráčovo ohodnocení jako provozorní, pokud sehrál méně než 20 her. Používá k tomu jiný způsob výpočtu než FIDE.

Jedním způsobem, jak počítat provozorní ohodnocení, ktré je blízké formuli(2) je její použití s velkou konstantou K.(např $K = 150$) Na začátku každému hráči přiřadíme ohodnocení založené například na jeho věku. Kvůli velké konstantě K pak dáváme velký důraz na následující výsledky, malý na počáteční ohodnocení, které ještě není zavedené a není dobrým odhadem reálné síly. Následné opakované použití formule(2) při dalších hráčových vystoupeních už reguluje jeho ohodnocení blíž k jeho skutečné výkonnosti.

1.4 Žebříčky FIDE a USCF

System ohodnocení Elo byl nejprve příjat v roce 1960 americkou šachistickou asociací USCF, organizací FIDE následně v roce 1970. Protože obě organizace počítají svoje žebříčky pomocí lehce odlišných algoritmů, vznikají odlišnosti i v ohodnoceních jednotlivých hráčů a v celkové ohodnocovací škále.

Ohodnocení FIDE je hráči přiřazeno pouze pokud je větší než 2000. Celá škála žebříčku sahá od těchto 2000 k přibližně 2800 bodů. Největší skóre dosáhl v roce 2000 Kasparov, který získal ohodnocení 2849, v součastnosti(květen 2006) je první hráč světa Topalov ohodnocen 2804 body. Algoritmus používaný FIDE se od původního algoritmu, který navrhl Elo liší. Hlavním rozdílem je způsob výpočtu očekávaného skóre. Zatímco Elo počítá skóre s každým ohodnocením soupeře zvlášť, FIDE do formule(1) dosazuje průměrné ohodnocení soupeřů. Takový postup nedává stejné výsledky a hlavně pokud jsou ohodnocení soupeřů více rozdílná, nedává přesnou předpověd. Dalším problémem je to, že kvůli bodové dvoutisícové hranici v průměru nadhodnocuje hráče, jejichž ohodnocení se kolem této hranice pohybuje. Například hráč s reálnou silou 1999 a lepšími výsledky na turnaji než je jeho reálná síla, se do žebříčku dostane. Na druhou stranu uvažujme hráče s reálnou silou 2001, který je kvůli svým výsledkům, které jsou slabší než by odpovídalo jeho reálné síle, podhodnocen. Takový do žebříčku zařazen není.

Ohodnocení USCF není omezené žádnou hranicí, nejmenší ohodnocení tak může být nulové. V takovém případě už hráč podle formulí(1) a (2) nemůže nic ztratit. V roce 1994 se ohodnocení hráčů pohybovalo mezi 45 a 2763 body. Průměrné skóre pak bylo 1490 a více než 95 procent hráčů měo své ohodnocení menší než 2200. Pokud bychom chtěli porovnat s žebříčkem FIDE ze stejné doby, průměrná hodnota skóre zde byla přibližně 2262 a pouze okolo 23% procent hráčů mělo své skóre menší než 2200 bodů.

1.5 Další charakteristiky systému

Ne všechny zápasy a turnaje se hrají za stejných podmínek. Jedním z rozdílných faktorů je časový limit, za který hráč musí stihnout ohlásit svůj tah. Tzv. rychlé šachy samozřejmě kladou jiné nároky na kvality hráče. Je zde důležitější schopnost rychle vymyslet strategii, která by možná při delším rozmýšlení byla hodnocena jako taktická chyba, zde ale vzhledem k časové tísni uspěje. Na konci 80. let tak byl vytvořen alternativní žebříček . Hranice, podle které se výsledky započítavají do toho kterého žebříčku, byla stanovena na 30 minut pro jednoho hráče na celou hru.

Problémem systémů ohodnocení jako je Elo je to, že se v principu jedná o systémy vytvářející ohodnocení pouze relativně. Přepokládejme pro příklad dva hráče se stejným ohodnocením 1500, které i přesně odhaduje jejich sílu. Oba odehrají spolu dostatek utkání a tedy získají stejný poloviční počet bodů. Za nějaký čas se utkají znovu ,oba se zlepší na stejnou úroveň reprezentovanou např. ohodnocením 1700. Oba mají ovšem stále svoje staré ohodnocení. Oba získají opět poloviční počet bodů a systém tedy žádnou změnu nepozná. Oběma hrácům zůstává ohodnocení 1500, i když jejich realné schopnosti už mu neodpovídají. Předpokládají se zde tedy pomalé změny, které nenastanou u velkého množství hráčů najednou. I přesto, pokud nyní budeme uvažovat dvě oddělené skupinky hráčů, kterým např. geografické podmínky brání se navzájem utkávat, může potom stejné ohodnocení v obou skupinkách odpovídat různým skutečným schopnostem. Tomu se dá vyhnout pouze tím, že se budou hráči obou skupinek pravidelně utkávat. Pokud se pak několik z nich zúčastní vzájemného turnaje, jejich ohodnocení se relativně zkoriguje a oni tuto změnu zanesou zpátky do svých skupin.

Porovnání hráčova ohodnocení s jeho reálnými schopnosti je velice problematické. Jeden důvod byl nastíněn v předchozím, další důležitější problém , se kterým je třeba se vypořádat, pokud chceme získat absolutní porovnání hráčových schopností z jeho ohodocení je deflace bodů v systému. V ideálním případě každá ztráta bodů vyústí ve stejný zisk jiného hráče. Toto vyplývá z formulí (1) a (2). Pokud je celkový počet bodů v systému konstatní , stejně jako počet hráčů, zůstává stejné i průměrné ohodnocení. V polovině 70.let začalo být evidentní, že se celkový počet bodů i průměrné ohodnocení v systému USCF začíná snižovat. Důvodem je vstup nových hráčů a ukončení aktivní kariéry u jiných. Deflace nastává, pokud schopnosti hráče, potažmo jeho ohodnocení, se v průměru s časem zvětšuje. Pokud je potom počet hráčů končících a začínajících přibližně stejný , celkový počet bodů, které systém ztratil musí větší než počet bodů, který noví hráči do systému přinesli. Jedním z pokusů o zastavení deflace bylo zavedení tzv. bonusových bodů polovině 70.let. Pokud se hráčovo ohodnocení(označme ho A) zvětšovalo dostatečně rychle, bylo k němu přičteno ještě určité množství bonusových bodů. Důvodem měla být neschopnost formule (2) dostatečně rychle reagovat na hráčovo zlepšení. Jeho soupeři(B) také bylo přičteno jisté množství tzv. bodů zpětné vazby. Vystoupení hráče B bylo totiž hodnocené jako vystupení s podhodnoceným soupeřem A. Při takovém vystoupení ztratil tento hráč B neprávem body, protože jeho skutečné skóre bylo menší než očekávané.To si můžeme ilustrovat na příkladu. Hráč A má ohodnocení 1600, ve skutečnosti ale sílu 1650, hráč B má ohodnocení i reálnou sílu 2000. Očekávané skóre pro hráče B počítané s nepřesným ohodnocením hráče A je 0.91, skóre počítané se skutečnou silou hráče B je 0.88.

V polovině 80.let bylo předchozí opatření zrušeno, protože se ukázalo, že se úbytek bodů kompenzuje příliš a bodů v systému začíná přibývat. Také nemělo takové opatření příliš oporu v statistickém pojetí žebříčku. Na konci 80.let bylo pak přijato opatření jiné, které se používá s obměněným parametrem dodnes. Jde o zavedení bodových hranic, pod které nemůže ohodnocení klesnout, pokud už předtím dosáhlo jisté úrovně. V současnosti je takovou hranicí první stovkou dělitelná hodnota menší než hodnota největšího dosaženého ohodnocení zmenšená o 100. Pokud tedy například hráč už jednou dosáhl ohodnocení 1875 nemůže jeho ohodnocení klesnout pod 1700. Toto nejenom vkládá to systému body navíc tak, že brání jejich odečtení u hráče, který narazil na svou hranici, ale může i bránit některým hráčům, aby si plánovaně snižovali svoje ohodnocení kvůli zařazení na turnaj s horšími hráči.

Několik dalších postupů se snaží o zachování absolutního porovnání pomocí žebříčku. Abychom mohli porovnávat tímto způsobem potřebujeme, aby hráči hodnocení například hodnotou 1500 měli stále stejné schopnosti. V ideální případě by tedy hráč hodnocený číslem 1500 měl stejnou reálnou sílu jako hráč, který byl takto ohodnocený například před deseti lety. Pokud by něco takového žebříček zajišťoval, bylo by možné sledovat dobře pokrok každého hráče.

Nejjednodušší způsob o tuto nápravu by bylo přičtení určitého množství bodů ke každénu ohodnocení. Množství by mohl určovat rozdíl medianů počítaných v různých obdobích přes všechny ohodnocení. Při zafixování medianu na hodnotě 1500 by polovina hráčů měla vždy ohodnocení menší a polovina větší než 1500. Další podobnou podmínkou je například taková, která by si vyžadovala pouze 1% procento hráčů s ohodnocením větším než 2200. Dalším možným přístupem je vytvoření určitých kritérií, které je možné považovat za absolutní, např. různých testů nebo počítačových šachových programů. Při použití testů se musí zajistit, aby hráči měli před jeho vypracováním stejné podmínky. Test tedy nemůže obsahovat všechny stejné otázky, jaké obsahoval v předchozím roce, tak aby se hráči nemohli otázky naučit předem. Na druhou stranu musí aspoň některé otázky být stejné, aby se zajistilo spravedlivé srovnání. Kompromisem je náhodný výběr několika otázek, které budou mít testy společné. Počítačový program, pokud se sám neumí učit, zlepšuje svou výkonnost pouze tehdy, pokud změníme jeho kód. Jako takový by nám mohl sloužit jako absolutní měřítko síly hráčů. Pokud necháme hrát počítačový program proti většímu množství lidí nebo jiných už ohodnocených programů, můžeme tak získat jeho ohodnocení. Stejně jako zvládnutí určených testových otázek i výsledky proti absolutně ohodnoceným programům určují sílu hráče absolutně a můžou tak pomoci při korekci žebříčku. Je nutné zajistit, aby hráči nehráli proti stejným programům příliš často a to ze stejného důvodu, jako nemůžeme používat stejné testové otázky. Samozřejmě otázkou zůstává, jestli hráči nehrají s počítačovými programy jinak než s lidkým partnerem a do jaké míry je tedy taková metoda použitelná.

Kapitola 2

Teorie ohodnocení Kendall-Wei

2.1 Idea metody a vyhledávač Google

Tato metoda stejně jako Elo porovnává význam a vytváří příslušné ohodnocení určité množiny, na základě preferencí jednoho prvku množiny nad druhým. Tuto preferenci určujeme podle vlivu, který na sebe prvky vzájemně mají. Teorie ohodnocení Kendall-Wei má mnoho různých aplikací ve statistice, ekonomii i v různých jiných oborech, používá ji jak uvidíme například vyhledávač Google.

Demonstrovat jak metoda funguje, budeme na příkladě tvorby žebříčku sportovních týmů na základě jejich vystoupení na společném turnaji.

Budeme používat incideční matici turnaje A. Jednotlivé týmy označíme čísly $1 \dots n$. Prvek a_{ij} matice A můžeme pak definovat různými způsoby podle míry preference určené vzajemným zápasem týmu i s týmem j. V našem případě budeme a_{ij} definovat jako 1 pokud tým i porazil tým j. Jinak budeme tento prvek definovat jako nulový. Všechny prvky na diagonále jsou nulové.

Hledáme metodu, která každému týmu přiřadí ohodnocení, které je přímo uměrné součtu ohodnocení týmů, které náš tým na turnaji porazil. Taková metoda hodnotí každé vítězství různě podle kvality soupeře, který byl poražen.

Definice ohodnocení týmu pomocí jiných ohodnocení zní jako definice kruhem ,jak uvidíme později není v metodě problém a to z toho důvodu, že nedefinujeme přímo ohodnocení ,pouze vzajemné vztahy mezi nimi a ohodnoceni poté hledáme mezi číselnými vektory, které tyto podmínky splňují.

Pro ohodnocení i-teho týmu x_i platí tedy podmínka

$$x_i = K \sum_{j=1}^{n} a_{ij} x_j$$

Vidíme, že problém můžeme převést na nalezení vlastních čísel a vektorů matice turnaje. Pokud totiž λ je vlastní číslo a v vlastní vektor matice pak platí

$$A.v = \lambda v$$

a tedy

$$v_i = \frac{1}{\lambda} \sum_{j=1}^{n} a_{ij} v_j \tag{3}$$

Pokud je $\lambda \neq 0$ našli jsme hledané ohodnocení.

Problémem zůstavá to, jestli existuje vektor pro ohodnocení vhodný. Předpokládáme, že prvek a_{ij} definujeme nezáporně. Dále tedy pracujeme s nezápornou maticí A. Vlastní čísla matice jsou jako kořeny charakteristického polynomu čísla komplexní. Vektor, jehož složky nejsou čistě reálné, není pro ohodnocení vhodný. Stejně tak není vhodný reálný vektor, jehož složky nemají stejné znaménko. V takovém případě by tým s kladným hodnocením byl „potrestán" za výhru s týmem ohodnoceným záporně. Prostor vlastních vektorů dále nemusí mít dimenzi 1. V případě dimenze rovné 1 je vektor ohodnocení definován až na násobek reálným číslem jednoznačně, v případě více dimenzí tomu tak není. Dále se musíme vyhnout situaci,kdy je spektrálni polměr roven 0. V takovém případě nemá výraz ohodnocení (3) smysl. V druhé podkapitole ukážeme, že pro určitou skupinu tzv. ireducibilních matic existuje kladné vlastní číslo s příslušejícím kladným vlastním vektorem. Prostor vlastních vektorů ireducibilní matice je navíc jednodimenzionální. Pro nezáporné matice se poznatek zeslabuje na existenci nezáporného vlastního čísla, kterému přísluší nezáporný vlastní vektor. Dimenze prostoru vlastních vektorů navíc může být i větší než 1.

Jedním z dalších použití teorie Kendall-Wei je její použití internetovým vyhledávačem Google. Ten tuto metodu používá na hodnocení důležitosti vyhledaných stránek. Matici A v tomto případě sestavíme tak , že bude obsahovat prvek 1 na pozici a_{ij} pokud ze stránky s indexem j míří odkaz na stránku s indexem i, jinak bude obsahovat pouze nulové prvky. Důležitost stránky je tedy přímo úměrná součtu důležitosti stránek, který obsahují odkaz na ni. Pokud je tedy počet odkazů na dvě stránky stejný, je jako důležitější hodnocena stránka, na niž se odkazují jiné důležité stránky.

2.2 Existence nezáporného vlastního vektoru

Definice 2.2.1 *Nezáporná čtvercová matice A se nazývá reducibilní, jestliže existuje permutační matice P taková, že matice PAP^T má formu*

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

kde B a D jsou čtvercové podmatice.

V opačném případě je matice ireducibilní.

Matice 1×1 je podle definice ireducibilní.

Definice 2.2.2 *Grafem asociovaným nezáporné matici $n \times n$ rozumíne orientovaný graf G(V,E), kde $V = \{1, \ldots, n\}$ a z vrcholu i vede hrana do vrcholu j, právě když je $a_{ij} \neq 0$.*

Definice 2.2.3 *Orientovaný graf je silně souvislý, právě když pro každou uspořádanou dvojici (i,j) existuje v grafu cesta z vrcholu i do vrcholu j.*

Pokud P je permutační matice, pak PAP^T odpovídá v grafové interpretaci přejmenování vrcholů. Matice A a $B = PAP^T$ mají tedy izomorfní příslušející grafy. Platí následující tvrzení:

Věta 2.2.1 *Nezáporná matice je ireducibilní, pravě když je asociovaný graf silně souvislý.*

Důkaz $\Longrightarrow$ Pokud asociovaný graf není silně souvislý, existují vrcholy i, j takové, že neexistuje cesta z i do j. Rozdělme nyní množinu V vrcholů grafu na dvě disjunktní podmnožiny V_1 a V_2. Vrchol $i \in V_1$, vrchol $j \in V_2$, vrchol $v \neq i, v \neq j$ náleží V_1 právě když do něj z vrcholu i vede cesta. Pokud do něj taková cesta nevede, náleží tedy množině V_2. Pro každý vrchol $v_1 \in V_1$ platí, že z něj nevede hrana do žádného vrcholu $v_2 \in V_2$. Předpokládejme existenci cesty z vrcholu v_1 do vrcholu v_2 ,$v_1 \in V_1$,$v_2 \in V_2$.. Protože podle definice V_1 existuje cesta z vrcholu i do vrcholu v_1, existuje cesta z i do v_2, což se spor s definicí V_2. Neexistence hrany mezi dvěma vrcholy je v matici reprezentována jako 0. Z toho už vyplývá, že existuje permutační matice P taková, že PAP^T má formu

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

Nulová podmatice má rozměry $|V_1| \times |V_2|$ a platí $|V_1| + |V_2| = n$,podmatice B a D jsou tedy čtvercové.
$\Longleftarrow$ Pokud je matice $n \times n$ reducibilní, má PAP^T pro vhodnou permutační matici P tvar

$$A = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

kde D a B jsou čtvercové matice. Nechť má nulová podmatice rozměry $k_1 \times k_2$. Pak nutně $k_1 + k_2 = n$. Protože úprava PAP^T odpovídá přejmenování vrcholů asociovaného grafu, existují dvě disjunktní podmožiny vrcholů V_1 a V_2, takové, že $|V_1| = k_1$, $|V_2| = k_2$ a neexistuje hrana mezi libovolnými vrcholy $v_1 \in V_1$,$v_2 \in V_2$. Graf tedy nemůže být silně souvislý.

Tvrzení, které potřebujeme k tomu, abychom měli zajistěnou existenci kladného (v případě ireducibilních matic) vlastního vektoru je částí Perron-Frobeniovy teorie. V roce 1907 objevil Perron v své době neočekané spektrální vlasnosti kladných matic. Frobenius jeho výsledky později zobecnil na případ nazáporných ireducibilních matic. Všechny následující věty a důkazy jsou čerpány z knihy [3].

Lemma 2.2.1 *Nechť A je ireducibilní matice a v je nazáporný vektor, který má přesně k nenulových souřadnic,* $1 \leq k \leq n-1$*. Potom má vektor* $(I_n + A)v$ *více než k kladných souřadnic.*

Důsledek 2.2.1 *Pokud A je ireducibilní* $n \times n$ *matice a y nezáporný nenulový vektor, potom* $(I_n + A)^{n-1}y > 0$.

Důsledek 2.2.2 *Nezáporná čtvercová* $n \times n$ *matice A je ireducibilní,právě tehdy pokud platí* $(I_n + A)^{n-1} > 0$.

Důsledek 2.2.3 *Nezáporný vlastní vektor nezáporné ireducibilní matice musí být kladný.*

Jako E^n budeme dále označovat množinu $\{(x_1, x_2, \ldots, x_n) \in \mathbf{P^n} | \sum x_i = 1\}$, kde $\mathbf{P^n}$ značí množinu všech nazáporných vektorů délky n.

Definice 2.2.4 *Nechť A je nezáporná matice o velikosti* $n \times n$ *. Potom definuje funkci* f_A *z množiny* $\mathbf{P^n}$ *do množiny nezáporných čísel takto:*

$$f_A(x) = \min_{x_i \neq 0} \frac{(Ax)_i}{x_i}$$

pro každý nenulový vektor $v \in \mathbf{P^n}$.

Funkce f_A se nazývá Collatz-Wielandtova funkce asocianovaná s maticí A.

Lemma 2.2.2 *Nechť A je ireducibilní nezáporná matice, f_A je Collatz-Wielandtova funkce asocianovaná s maticí A. Pak platí následující :*
(i) funkce f_a je homogenní stupně 0, platí tedy $f_A(x) = f_A(tx)$ pro $t > 0$ a každý nenulový nezáporný vektor x.
(ii) pokud je x nezáporný nenulový vektor a σ je největší reálné číslo pro které

$$Ax - \sigma x \geq 0,$$

potom $\sigma = f_a(x)$.
(iii) pokud x je nezáporný nenulový vektor a $y = (I_n + A)^{n-1}x$, potom $f_A(y) \geq f_A(x)$.

Věta 2.2.2 *Nechť A je ireducibilní nezáporná $n \times n$ matice. Potom funkce f_a nabývá na množině E^n svého maxima.*

Důkaz Nechť

$$G = (I_n + A)^{n-1}E^n = \{y| \ y = (I_n + A)^{n-1}x, x \in E^n\}$$

Potom je G kompaktní množina. Dále podle Důsledku 2.2.2 jsou všechny vektory náležící množině G kladné. Protože je funkce f_a na množině kladných vektorů spojitá, nabývá na G svého maxima v nejakém jeho bodě y^0. Potom vektor $x^0 = y^0/\sum_{i=1}^{n} y_i^0$ kde y_i^0 jsou jednotlivé složky vektoru y^0, náleží množině E^n. Nechť potom x je libovolný vektor náležící množině E^n , $y = (I_n + A)^{n-1}x$. Potom dostáváme
$f_A(x) \leq f_A(y)$ podle Lemmatu 2.2.2 (iii)
$f_A(y) \leq f_A(y^0)$ díky maximalitě vektoru y^0 ve množině G
$f_A(y^0) = f_A(x^0)$ podle Lemmatu 2.2.2 (i).
Proto funkce f_A nabývá svého maxima na množině E^n v bodě x^0.

Důvod proč jsme převáděli v přechozím důkazu hledání maxima na množině E^n na množinu G je ten, že funkce f_A není na množině E^n spojitá. Nemůžeme tak využít tvrzení o existenci extrému spojité funkce na kompaktu. Funkce f_A je zřejmě spojitá na množině kladných vektorech, což jsme využili i v předchozím důkazu. To že není spojitá na hranici množiny E^n můžeme demonstrovat na následujícím příkladu.

Nechť

$$A = \begin{pmatrix} 2 & 2 & 1 \\ 2 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}$$

a $x(\varepsilon) = (1, 0, \varepsilon)/(1+\varepsilon)$, *kde* $\varepsilon \geq 0$. Potom

$$Ax(\varepsilon) = (2+\varepsilon, 2+\varepsilon, \varepsilon)/(1+\varepsilon),$$

a funkce

$$f_A = \min(\frac{2+\varepsilon}{1}, \frac{\varepsilon}{\varepsilon}) = 1.$$

Potom ale $f_A(x(0)) = 2 \neq 1 = \lim_{x \to 0} f_A(x(\varepsilon))$ a funkce f_A tedy není spojitá.

Věta 2.2.3 *Iredcibilní nezáporná matice A má reálné kladné vlastní číslo r (tvz. maximální vlastní číslo) , pro které platí*

$$r \geq |\lambda_i|$$

pro každé vlastní číslo λ_i matice A. Vlastnímu číslu r přísluší kladný (maximální) vlastní vektor .

Důkaz Věta 2.2.2 nám dává existenci $r = \max\{f_A(x)| \ x \in E^n\}$. Dokážeme, že právě toto r je hledaným maximálním vlastním číslem.

Nejprve ukážeme, že platí $r > 0$. Víme, že $r > f_A(u)$ pro každé $u \in E^n$. Za u vybereme vektor $u = (1, 1, \ldots, 1)/n$. Potom

$$r > f_A(u) = \min \frac{(Au)_i}{u_i} = \min \sum_{j=1}^{n} a_{ij} > 0$$

Nyní ukážeme, že r je skutečně vlastní číslo matice A. Podle Lemmatu 2.2.2 máme

$$Ax^0 - rx^0 \geq 0.$$

Chceme dokázat rovnost. Předpokládejme tedy pro spor, že je $Ax^0 - rx^0 \neq 0$.. Podle Důsledku 2.2.1 je

$$(I_n + A)^{n-1}(Ax^0 - rx^0) > 0,$$

tedy platí

$$Ay^0 - ry^0 > 0$$

kde $y^0 = (I_n + A)^{n-1}x^0$.

Protože se zde jedná o ostrou nerovnost, můžeme určitě nalést takové ε, že platí

$$Ay^0 - (r + \varepsilon)y^0 \geq 0$$

Potom ale podle

$$r + \varepsilon \leq f_A(y^0)$$

a proto

$$r < f_A(y^0).$$

Podle předchozí věty jsou maxima množin G a E^n a došli jsme tedy ke sporu.

Dokázali jsme nejenom to, že r je vlastní číslo, ale také to, že pokud pro nějaké x platí $Ax - rx \geq 0$. pak x je vlastní vektor příslušející r. Podle Důsledku 2.2.3 musí být kladný.

Zbývá tedy dokázat,

$$r \geq |\lambda_i|$$

kde λ_i jsou všechny vlastní čísla matice, tedy že r je rovno spektrálnímu poloměru.

Nechť λ_i je vlastní číslo matice A s příslušejícím vlastním vektorem $z = (z_1, z_2, \ldots, z_n)$. Potom platí

$$\lambda_i z_t = \sum_{j=1}^{n} a_{tj} z_j \qquad t = 1 \ldots n.$$

a tedy z trojúhelníkové nerovnosti i

$$|\lambda_i|\, |z_t| = \sum_{j=1}^{n} |a_{tj}|\, |z_j| \qquad t = 1 \ldots n.$$

V vektorovém zápisu pak

$$|\lambda|\, |z| \leq A|z|.$$

Podle Lemmatu 2.2.2 (ii) a definice r pak platí

$$|\lambda_i| \leq f_A(|z|) \leq r,$$

čímž je věta dokázaná.

V následujícím trvzení se využívá toho, že vlastní čísla a vektory jsou spojité funkce vzhledem k prvkům matice. Pokud přejdeme od ireducibilních (speciálně kladných) matic k nezáporným, dostaneme obecnější a slabší tvrzení.

Věta 2.2.4 *Nechť A je nezaporná $n \times n$ matice. Potom A má nezáporné vlastní číslo r, pro které platí*

$$r \geq |\lambda_i|$$

pro každé vlastní číslo λ_i matice A. Vlastnímu číslu r přílsuší nezáporný vlastní vektor.

Dalším tvrzením, které se hodí při aplikaci teorie Kendall-Wei je následující trzení dokazující jednorozměrnost prostoru vlastních vektorů. Předpokladem je ale ireducibilita matice A. Ve třetím pododdíle ukážeme příklad reducibլní matice, jejíž prostor vlastních vektorů má dimenzi větší než 1.

Věta 2.2.5 *Prostor vlastních vektorů ireducibilní matice A příslušející maximálnímu vlastnímu číslu r má dimenzi 1.*

Podle důsledku 2.2.3 pokud x je nezáporný vektor příslušející maximálnímu vlastnímu číslu r potom platí $x > 0$. Předpokládejme, že y vlastní vektor matice A, pro který předpokládáme pouze to, že je různý od nuly. Potom platí podle trojúhelníkové nerovnosti

$$A|y| \geq r|y|,$$

Ve větě 2.2.3 jsme dokázali, že $|y|$ je vlastní vektor matice A. Tím jsme dokázali i to, že původní vektor y nemůže mít nulovou souřadnici.

Mějme nyní libovolné dva vlastní vektory matice A: $x = (x_1, x_2, \ldots, x_n)$ a $y = (y_1, y_2, \ldots, y_n)$. Vektor $y_1 x - x_1 y$ leží ve prostoru vlastních vekotrů a jeho první souřadnice je nulová. Z přechozího vyplývá, že platí

$$y_1 x - x_1 y = 0$$

a vektory x a y jsou lineárně závislé.

Jako poslední v této podkapitole uvedeme tvrzení které dává horní a dolní odhad velikosti maximálního vlastního čísla r.

Věta 2.2.6 *Nechť A je nezáporná matice s maximalním vlastním číslem r a řadkovými součty $r_1, r_2, \ldots, r_n$. Potom*

$$\rho \leq r \leq \sigma$$

kde $\rho = \min_i r_i$ a $\sigma = \max_i r_i$.

2.3 Reducibilní matice a Kendall-Wei

V této části uvedeme několik příkladů, na které můžeme narazit, pokud není splněna podmínka ireducibility matice. Věta 2.2.4 pro nezápornou matici zaručuje maximální nezáporné vlastní číslo. Může tedy nastat případ, kdy je spektrální poloměr roven 0. Uvažujme například matici

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Všechny vlastní čísla této matice jsou rovny nule a tedy neexistuje žádné ohodnocení, protože výraz (3) nemá smysl. Dá se dokázat, že spektrální poloměr incidenční matice turnaje je roven nule, právě když graf asociovaný k matici neobsahuje orientovaný cyklus.

Ireducibilita matice nám zaručí, že dimenze prostoru vlastních čísel příslušných maximálnímu vlastnímu číslu je rovna 1. V obecném případě nezáporné matice to neplatí. Můžeme uvažovat matici:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Taková matice má maximální vlastní číslo rovné 1. Jeho algebraická násobnost je 3, a prostor vlastních vektorů příslušejících k maximálnímu vlastnímu číslu je dimenze 2. Jeho bázi tvoří vektory

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

Výběr ohodnocení tak není jednoznačný.

V některých případech není zlom mezi ireducibilními a reducibilními maticemi nijak ostrý a blízko k sobě mají i maximální vlastní vektory. Uvažujme například incidenční matici turnaje

$$A = \begin{pmatrix} 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 3 & 0 & 0 \\ 8 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 15 & 0 \\ 0 & 0 & 0 & 0 & 0 & 15 \\ 0 & 0 & 0 & 15 & 0 & 0 \end{pmatrix} \tag{4}$$

Matice je reducibilní. Vidíme, že jsou zde vytvořeny dva bloky týmů, první blok je reprezentován prvními třemi řádky, druhý posledními. Týmy v obou blocích jsou podle výsledků vyrovnané. Každý tým vyhrál stejný počet utkání se soupeřem ze svého bloku. Jediné vzájemné zápasy, který byl mezi týmy z různých bloků odehrány, jsou reprezentovány prvky a_{24},a_{35}. Vyhráli je týmy z prvního bloku. Můžeme tedy předpokládat, že výkonnost prvního bloku je větší než výkonnost druhého.

Pokud ale spočítáme maximální vlastní číslo a vektor, dostaneme toto ohodnocení.

r=15

$$\begin{pmatrix} 0.1928 \\ 0.3615 \\ 0.3028 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Vidíme, že jsou zde oproti naší úvaze nadhodnocené týmy z druhé skupiny. Důvodem je pouze větší počet sehraných utkání(obecněji větší počet získaných bodů). Více se podobným situacím budeme věnovat ve třetí kapitole. Zde pouze ukážeme, že problém neodstraníme pouze tím, že budeme počítat pouze s ireducibilními maticemi. Pokud totiž budeme uvažovat matici turnaje

$$A = \begin{pmatrix} 0 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & 3 & 0 & 0 \\ 8 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 15 & 0 \\ 0 & 0 & 1 & 0 & 0 & 15 \\ 0 & 0 & 0 & 15 & 0 & 0 \end{pmatrix} \tag{5}$$

vidíme, že je to matice ireducibilní. Od matice(4) se liší pouze prvkem a_{53}. Její maximální vlastní číslo a vektor se také od případu matice(4) příliš neliší.

r=15.1

$$\begin{pmatrix} 1.000 \\ 0.972 \\ 0.963 \\ 0.426 \\ 0.456 \\ 0.438 \end{pmatrix}$$

Vidíme, že nadhodnocení týmu druhého bloku je zde podobné.

Kapitola 3

Žebříček FIFA

3.1 Současný oficiální FIFA žebříček

Současný žebříček FIFA (platný 30.června 2006) byl zaveden v roce 1993. V současnosti je na žebříčku přes 200 členských zemí. Body do žebříčku tým získává za každý zápas, v závislosti na kvalitě soupeře (jeho postavení na žebříčku před zápasem) a na dalších faktorech, které zmíníme. Rozpětí bodů, které je množné získat za jeden zápas je 0 až 30.

Hlavním kritériem ohodnocení zisku z utkání je samozřejmě to, jestli skončil výhrou, remízou nebo prohrou uvažovaného týmu. V případě zápasu se silným soupeřem je možné získat body i za prohru.

Dalším kritériem je počet vstřelených a obdržených branek. I zde rozhoduje síla soupeře o počtu přidělených bodů. Aby FIFA podpořila útočnou hru, góly vstřelené jsou v jejím žebříčku hodnoceny jako důležitější než góly obdržené. Dále postupně klesá důležitost brankového rozdílu, góly, které zvyšují gólový rozdíl za rozhodnutého stavu jsou tak méně hodnotné.

Malý bodový zisk je přidáván týmu, který hraje na cizím hřišti.

Důležitost utkání je dělena do pěti kategorií. Určuje koeficient, kterým je předchozí bodový zisk násoben. Kategorie a koeficienty podle vzrůstající důležitosti jsou následující: přatelské zápasy(1), kvalifikace kontinentálního šampionátu(1.5),kvalifikace mistrovství světa(1.5), zápasy z kontinentálních šampionátů(1.75), zápasy z poháru FIFA(1.75) a zápasy mistrovství světa(2).

Dále je brána v potaz síla jednotlivých zemských regionů. Tato síla je počítána na základě zápasů nejsilnějších týmů jednotlivých oblastí. Koeficienty jsou následující:

UEFA(Evropa) x 1.00
CONMEBOL(Jižní Amerika) x 0.99
CAF(Afrika) x 0.96
CONCACAF(Střední a Severní Amerika) x 0.94
AFC(Asie) x 0.93
OFC(Oceánie) x 0.93

Koeficintem se násobí zisk z utkání dvou týmů ze stejného regionu. Pokud se utkají dva týmy z různých oblastí, násobí se průměrem obou příslušných koeficientů.

FIFA se ve svém žebříčku snaží vyhnout přecenění týmů, které hrají velké množství zápasů. Proto proběhne následují výpočet:

1.vybere se nejlepších 7 utkání z předchozích 52 týdnů
2.spočítá se celkový počet bodů ze všech za rok odehraných utkání
3.spočítá se celkový počet bodů z vybraných sedmi
4.počet bodů ze všech zápasů se vydělí jejich počtem, poté násobí 7
5.spočte se průměr hodnot z bodů 3 a 4

Počítání průměru v bodě 5 na druhou stranu zohledňuje více nejlepší výsledky týmu.

Posledním kritériem je stáří započítavaného výsledku. Započítávají se výsledky z posledních 8 let, přitom jejich důležitost rok od roku lineárně klesá. Výsledky mladší než rok jsou násobeny koeficientem 1, výsledky staré 7 let koefientem 1/8.

3.2 Žebříčky pomocí Kendall-Wei

Klíčovým problémem je sestavení incidenční matice tedy ohodnocení utkání mezi i-tým a j-tým týmem, které určuje prvky a_{ij} a a_{ji}.

Důležitost utkání může být hodnocena různými kritérii. Hlavním kritériem je i zde vítěztví remíza nebo prohra v utkání. Za výhru získá vítěz celé ohodnocení utkání, u remízy si toto ohodnocení oba soupeři rozdělí a oba získají po polovině.

Dalším důležitým kritériem je příležitost, při které se oba týmy utkávají. Utkání mistrovství světa můžme hodnotit jinak než kvalifikace konfederačních mistrovství a ty zase jinak než přáteské zápasy. Míra preference jednotlivých událostí bude vyjádřena koeficientem, kterým se bude násobit předchozí ohodnocení utkání.

Koeficienty můžeme také ohodnotit rozdíl branek v utkání nebo to, jestli tým hrál doma nebo na hřišti soupeře.

Různý koeficient může být podobně přiřazen utkáním hraným v různé době. Všechny žebříčky budeme vždy počítat z výsledků utkání starých 4 roky. Větší koeficient pak ale přiřadíme utkání hraném v posledním roce uvažovaného období, menší pak v roce prvním.

Pro každou použitou metodu budeme počítat žebříček počítat za období 1999-2002, 2000-2003, 2001-2004 a 2002-2005. Kontrolu úspěšnosti žebříčku pak na období půl roku následující po posledním roku období. U prvního období tak například budeme brát výsledky z časového úseku leden 2003-červenec 2003. Tato doba je nastavena tak aby nebyla příliš dlouhá, protože výkonnost týmů se s časem mění a žebříček by potom neodrážel skutečnou sílu ani její aktualní odhad podle použité metody. Na druhou stranu nesmi být příliš krátká, aby se zabezpečila dostečně výmluvná statistika.

Problémem aplikace metody Kendall-Wei na žebříček Fifa je to, že počet zápasů odehraný různými týmy za období 4 let, se velmi liší a také to, že většina (asi 82 procent) všech zápasů se odehraje v rámci jednoho regionu. Tabulka 3 ukazuje u jednotlivých týmů počet odehraných zápasů a počet bodů, které tým z těchto zápasů získal. Za vítěství si tým připsal 1 bod, za remízu 1/2. Zobrazeno je 66 týmů s největším počtem zápasů. Stejné bodování platí pro tabulku 1, která ukazuje bodové zisky ze vzájemných utkání jednotlivých fotbalových konfederací. Obě dvě tabulky počítají s výsledky za období leden 2002-červenec 2006.

Budeme se věnovat čtyřem konktrétním variantám metody.

1.varianta

Při první incideční matici sestavíme následujícím způsobem. Počáteční ohodnocení je 1, násobené podle data utkání koeficientem :

- letošní rok - koeficient 1
- 1 rok staré výsledky - koeficient 0.8
- 2 roky staré - koeficient 0.6
- 3 roky staré - koeficient 0.4

Dalším počítaným kritériem je důležitost utkání.

- utkání na mistrovství světa - koef. 2
- utkání konfederačního mistrovství - koef. 1
- utkání kvalifikace na mistrovství světa nebo na kontinentální mistrovství- koef. 0.9
- jiné utkání (přáteské nebo z různých turnajů)

Kritéria jako rozdíl obdržených branek nebo výhoda domácího prostředí nebudeme používat. První zvýhodňuje (alespoň v této implementaci regiony), kde padá více branek(viz příklad dvou různě silných bloků z předchozí podkapitoly), co se týče druhého je výmluvnost takového kritéria v současném fotbale diskutabilní a žebříček s takovým kritériem pak méně úspěšný.

2.varianta

Další tři metody mají společné to, že se snaží omezit vliv počtu utkání na ohodnocení týmu.

Pokud se vrátíme k příkladu matice 5 z předchozí podkapitoly, vidíme, že týmy z druhého bloku jsou ohodnoceny výše než týmy z bloku prvního přesto, že v přímé konfrotaci vychází hůře.

Uvedeme jeden jiný příklad podobný příkladu matice 5 . Větší počet utkání zde nadhodnocuje tři poslední týmy. Je zde vidět, že situace nemusí být dovedena do krajnosti, aby větší počet utkání zvýhodňoval.

$$\begin{pmatrix} 0 & 2 & 1 & 1 \\ 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 4 \\ 0 & 3 & 4 & 0 \end{pmatrix}$$

První tým odehrál pouze 5 utkání a 4 z nich vyhrál, každého soupeře porazil častěji než on jeho. Ostatní jeho soupeři odehráli 11,10 resp.16 utkání. Druhý a třetí tým odehrál vysoký počet výsledkově vyrovnaných utkání s posledním týmem. To je všechny tři stavi v ohodnocení výše.

$$\begin{pmatrix} 0.5983 \\ 0.8005 \\ 0.8357 \\ 1 \end{pmatrix}$$

A.

Uvažujeme situaci kdy za výhru tým získá jeden bod, za remízu 1/2 bodu. Metodou bude získaný počet bodů vydělit celkovým počtem utkání. Součet bodů v řádku matice bude odpovídat poměru bodů získaných ku počtu bodů, které tým, jemuž řádek odpovídá,získat mohl ve všech svých utkáních. Pokud tým všechny zápasy vyhrál, bude tak součet v řádku roven 1. Zároveň se v jednom řádku neporuší vzájemné poměry mezi bodovými zisky z jednotlivých utkání. Toto opatření je namířeno proti týmům, které hrají zápasů přespříliš, naopak ale pomáhá týmům, které jich odehrají velmi malý počet. Tototo nešvaru se zbavíme, pokud zavedeme strop na počet utkání. Pokud tým odehraje víc než je daná hranice, budou jeho bodové zisky z jednotlivých utkání nejprve vyděleny jeho počtem odehraných utkání a potom vynásobeny zápasovým stropem. Výsledný součet v řádku u takového týmu vyjadřuje jeho průměrný bodový zisk za počet zápasů, který je roven zápasovému stropu. Na druhou stranu, pokud tým odehraje míň utkání než je strop, jeho zisky se nijak nepřepočítávají.

V implementaci na žebříček Fifa budeme za strop považovat medián počtu utkání všech týmů, které se ve sledovaném období zapojily do hry.

Podíváme se, jak funguje nastíněná metoda na předchozí příklady. Nebudeme zde používat zápasový strop, protože ten zde vzhledem k malému počtu týmů nemá opodstatnění. Matici pro lepší přehlednost vynásobíme 10, vlastní vektory se nezmění. Ohodnocení u příkladu z předchozí podkapitoly je nyní počítáno z matice:

$$\begin{pmatrix} 0 & 5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4.211 & 1.579 & 0 & 0 \\ 4 & 0 & 0 & 0 & 1.500 & 0 \\ 0 & 0 & 0 & 0 & 4.545 & 0 \\ 0 & 0 & 0.294 & 0 & 0 & 4.412 \\ 0 & 0 & 0 & 5 & 0 & 0 \end{pmatrix}$$

Bodové zisky v jednotlivých blocích se srovnaly. Ohodnocení nyní zohlední vzájemné zápasy a týmy prvního bloku postaví výše.

$$\begin{pmatrix} 2.2822 \\ 2.2194 \\ 2.19846 \\ 0.9725 \\ 1.0403 \\ 1 \end{pmatrix}$$

Druhý příklad je reprezentován maticí:

$$\begin{pmatrix} 0 & 4 & 2 & 2 \\ 0.9091 & 0 & 0 & 3.6364 \\ 0 & 1 & 0 & 4 \\ 0 & 1.8750 & 2.500 & 0 \end{pmatrix}$$

Ohodnocení je následující:

$$\begin{pmatrix} 1.8021 \\ 1.1080 \\ 1.073 \\ 1 \end{pmatrix}$$

B.

Uvažovat budeme i situaci, kdy jsou k tvorbě matice používané různé koeficienty(jako u varianty 1). Zisk z utkání tak může být větší i menší než jedna. Stejně jako u varianty 2A budeme ale dělit bodové zisky počtem zápasů. Stejně budeme určovat a počítat se zápasovým stropem.

3.varianta

Pokud použijeme 1.variantu a koeficiety násobící bodové a půlbodové zisky, můžeme uvažovat následující průměrný přepočet. Dělit nebudeme počtem zápasů, ale celkovým počtem bodů, které tým mohl získat. Opět můžeme nastavit bodový strop, v našem případě to bude medián možných bodových zisků všech týmů. Se stropem budeme pracovat podobně jako u varianty 2A. Pokud tým všechny své utkání vyhraje, je vliv koeficientů nulový, pokud je jeho úspěšnost menší než stoprocentní, má vyšši bodový zisk z utkání na celkový součet bodů v řádku pozitivní vliv.

3.3 Porovnání úspěchu předvídání

Za kvalitní žebříček (ohodnocení) považujeme takový, které nejlépe předvídá výsledky budoucích utkání. Výsledek předpovídáme jako nerozhodnodný nebo jako vitěství jednoho z týmů. Úspěch budeme vyjadřovat v procentech správně předpovězených výsledků. Problémem je předpovídání remíz.

Musíme zde nastavit bodovou hranici na žebříčku, která bude remízy i ostatní výsledky předvídat nejlépe. Pokud pak absolutní hodnota rozdílu ohodnocení dvou týmů bude menší než tato hranice, předpovíme remízu, pokud bude větší, budeme předpokládat, že vyhraje tým s větším ohodnocením. Protože ve většině případů získáme žebříček jednoznačný až na násobek, budeme v takovém případě pracovat s vektorem z jednotkové sféry. Při testování úspěšnosti v jednotlivých obdobích můžeme hranici optimalizovat, abychom získali co největší uspěšnost.

Tato hranice musí být ve skutečnosti samozřejmě uniformní pro všechny období. Cílem je zde co největší průměrná uspěšnost za sledovaná čtyři období. Tuto hranici bychom pak také použili pro případné předvídání v budoucnosti.

Při počítání úspěchu žebříčku FIFA jsem neměl k dispozici starší bodové ohodnocení týmů z jednotlivých období, pouze jejich pořadí na žebříčku. Hranici remízy jsem v takovém případě nastavoval ne podle ohodnocení ale podle tohoto pořadí. Takový přístup může samozřejmě vést k jiné úspěšnosti.

Tabulka 10 ukazuje úspěšnost jednotlivých variant ohodnocení na základě metody Kendall-Wei a žebříčku Fifa v předvídání ve sledovaných čtyřech obdobích a pak také průměrnou úspěšnost za celé čtyřleté období. Ta není průměrem přechozích, protože je zde optimalizována a uniformizována hranice remizy , která je jinak pro každé období jiná.

Největší úspěšnost z uvažovaných třech metod má metoda 2A -57 procent a metoda 3-56 procent. Ostatní dvě stejně jako žebříček Fifa mají úspěšnost kolem 50 procent, stejnou jako předvídání na základě náhody. Žebříčky dvou nejúspěšnějších variant za období 2002-2005 jsou uvedeny v příloze jako tabulky 8 a 9.

Mohli bychom předpokládat, že týmy, které se umístily na žebříčku na předních pozicích, mají stabilnější výkonnost. Můžeme tedy zkusit omezit se v předpovídání výsledků pouze na týmy, jejichž ohodnocení je vyšší než zvolená hranice ,teoreticky bychom mohli dostat vyšší úspěšnost. Počítali bychom se zápasy, v nichž figuruje aspoň jeden tým s vyšším ohodnocením. Jak ukazují grafy 4-7, zobrazující závislost úspěšnosti na zvolené hranici, není žádná taková pravidelnost z dosažených výsledků průkazná. V příloze jsou uveden případ 3.varianty, výsledky jsou podobné i u ostaních metod. Je vidět,že předpoklad stabilnější výkonnosti silných týmů není vždy splněn.

Kapitola 4

Přílohy

	Uefa	Caf	Conmebol	Acf	Concacaf	Ofc
Uefa	896	31	38.5	96	27	3.5
Caf	23	747	12	28	9	0
Conmebol	19.5	6	385	16	39.5	1
Acf	56	27	16	631	10	15.5
Concacaf	29	9	51.5	28	134	1
Ofc	0.5	0	0	3.5	0	68

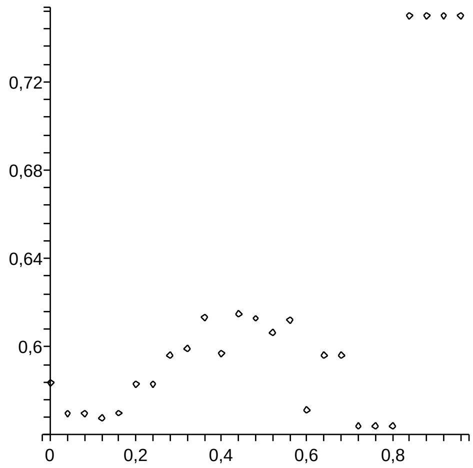
Tabulka 1. vzájemné bodové zisky mezi konfederacemi ze zápasů v období leden 2002-cervenec 2006

Uefa	Evropa
Caf	Afrika
Conmebol	Střední a Severní Amerika
Acf	Asie
Concacaf	Jižní Amerika
Ofc	Oceánie

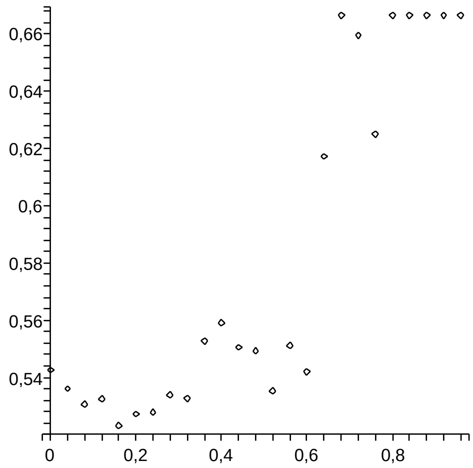
Tabulka 2. Konfederace a příslušné světadíly

Bahrain	43	81	France	40.5	53
Trinidad and Tobago	38.5	74	Honduras	26.5	53
Mexico	48.5	72	Tunisia	29.5	51
Kuwait	37	72	Portugal	35	51
Korea Republic	40.5	72	Iran	36.5	51
Japan	46	71	Finland	26.5	51
USA	50	71	Zimbabwe	31.5	50
Costa Rica	36	70	Syria	23.5	50
Saudi Arabia	44.5	67	Oman	31.5	50
Brazil	48	66	Italy	34.5	50
South Africa	36	64	Morocco	34	50
Guatemala	31	63	China PR	28	50
Germany	39	60	Panama	19	50
Jamaica	33	60	Denmark	34.5	50
Jordan	34.5	59	Spain	38	49
Uganda	28	58	Greece	32	49
Egypt	39.5	58	England	33.5	49
Kenya	31.5	58	Paraguay	27	49
Ecuador	28	56	Uruguay	25.5	48
Turkey	34	56	Senegal	32	48
Estonia	24	56	Singapore	25	48
Qatar	29.5	56	Hungary	20.5	46
Sweden	33	56	Czech Republic	36	46
Poland	36.5	55	Netherlands	33	46
Sudan	25.5	54	Croatia	29.5	46
Argentina	37.5	54	Botswana	20	46
Colombia	26.5	54	Republic of Ireland	30.5	45
Zambia	30.5	53	Mali	28	45
Latvia	24.5	53	Thailand	20	44
Iraq	30.5	53	Rwanda	20.5	44
United Arab Emirates	23	53	Slovakia	23	44
Norway	31.5	53	Congo DR	22.5	43
Nigeria	35	53	Serbia and Montenegro	20	42

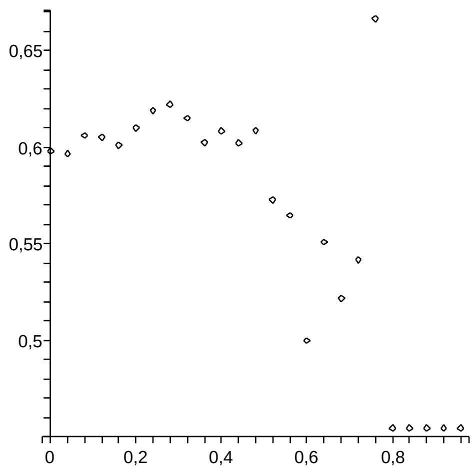
Tabulka 3. Nejvyšší počet zápasů a příslušné bodové zisky za období leden 2002-cervenec 2006



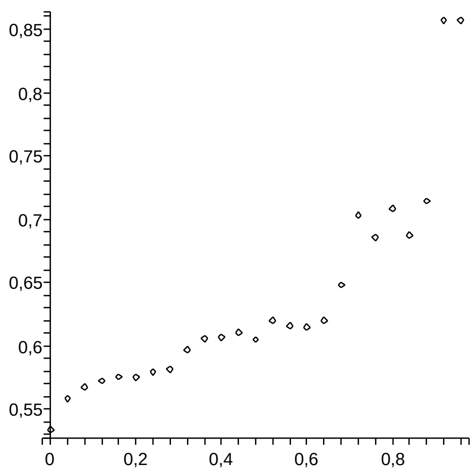
Graf 4. Závislost úspěšnosti předvídání na zvolené bodové hranici 3.varianta, 1.období (žebříček počítán z období 1999-2002, úspěšnost v období leden 2003-červenec 2003)



Graf 5. Závislost úspěšnosti předvídání na zvolené bodové hranici 3.varianta, 2.období (žebříček počítán z období 2000-2003, úspěšnost v období leden 2004-červenec 2004)



Graf 6. Závislost úspěšnosti předvídání na zvolené bodové hranici 3.varianta, 3.období (žebříček počítán z období 2001-2004, úspěšnost v období leden 2005-červenec 2005)



Graf 7. Závislost úspěšnosti předvídání na zvolené bodové hranici 3.varianta, 4.období (žebříček počítán z období 2002-2005, úspěšnost v období leden 2006-červenec 2006)

1	Argentina	1	34	Norway	.4527791527
2	Brazil	.8238116601	35	Honduras	.4371521718
3	France	.8073169383	36	Poland	.4344245741
4	Spain	.7891772422	37	Romania	.4344174776
5	England	.7762622345	38	Tunisia	.4341149073
6	Czech Republic	.6843017238	39	Russia	.4246370665
7	Uruguay	.6838122234	40	Bulgaria	.4224656258
8	Denmark	.6836237283	41	Bolivia	.4171304793
9	Colombia	.6800642225	42	Wales	.3964058367
10	Italy	.6598251485	43	Côte d'Ivoire	.3937843532
11	Netherlands	.658514209	44	Ukraine	.392786487
12	Republic of Ireland	.6578324714	45	Morocco	.3853445627
13	Greece	.6279223604	46	Egypt	.3829360866
14	Paraguay	.6275004453	47	South Africa	.3709243146
15	Germany	.6198527736	48	Guatemala	.3660300289
16	Mexico	.6138359219	49	Finland	.3642554892
17	Cameroon	.5776326356	50	Iran	.3544564713
18	Turkey	.5742235599	51	Slovenia	.3487572963
19	Peru	.5723827321	52	Switzerland	.3391166725
20	Portugal	.5704036211	53	Serbia and Montenegro	.3385753077
21	Ecuador	.5600665989	54	Belarus	.3367749228
22	Senegal	.5536618339	55	Libya	.3278975017
23	USA	.5514873781	56	Angola	.3218222755
24	Croatia	.5369068023	57	Togo	.3139175669
25	Sweden	.5361161569	58	Israel	.3001006168
26	Japan	.5192552275	59	Bosnia-Herzegovina	.2995316659
27	Venezuela	.514132127	60	Canada	.2986943666
28	Korea Republic	.5139457899	61	Austria	.29510853
29	Belgium	.511270677	62	Congo DR	.289218333
30	Chile	.4981615414	63	Zimbabwe	.2825371804
31	Australia	.4598947976	64	Jamaica	.2776224308
32	Costa Rica	.4554450906	65	Panama	.2774089678
33	Nigeria	.4551108805	66	Mali	.2772442354

Tabulka 8. Žebříček fotbalových týmů , období 2002-2005, varianta 2A

1	Argentina	1	34	Nigeria	.419552061
2	France	.7782045461	35	Bolivia	.4195416895
3	Brazil	.7377119374	36	Norway	.4194508835
4	Spain	.7184836827	37	Poland	.414466401
5	Colombia	.7119696992	38	Tunisia	.4075415667
6	England	.701651823	39	Romania	.4046242971
7	Uruguay	.6773484166	40	Côte d'Ivoire	.3950908859
8	Italy	.6104584941	41	Morocco	.3793031582
9	Paraguay	.6047473962	42	Guatemala	.3730258384
10	Denmark	.6026062784	43	Bulgaria	.3689250565
11	Czech Republic	.5825163752	44	Russia	.3634132359
12	Republic of Ireland	.5791235762	45	Ukraine	.3593420677
13	Ecuador	.5617311454	46	Egypt	.3520226471
14	Netherlands	.5602731953	47	Finland	.3402488473
15	Mexico	.5596888139	48	South Africa	.3376065677
16	Peru	.5494869866	49	Slovenia	.3315171228
17	Greece	.5486793787	50	Wales	.3269777113
18	Germany	.5435685051	51	Belarus	.3213431999
19	Turkey	.5161174918	52	Angola	.3159650633
20	Cameroon	.5072879624	53	Togo	.3148917586
21	Portugal	.5031802598	54	Iran	.314761665
22	USA	.5022253336	55	Israel	.2909083173
23	Venezuela	.5021551491	56	Libya	.2862363601
24	Japan	.489137249	57	Serbia and Montenegro	.2851116544
25	Sweden	.4844877571	58	Bosnia-Herzegovina	.2727042387
26	Croatia	.4842101778	59	Canada	.2722176515
27	Korea Republic	.483530478	60	Jamaica	.2707316456
28	Senegal	.4792204353	61	Panama	.2695710404
29	Chile	.477801419	62	Switzerland	.2685981318
30	Australia	.4693940154	63	Austria	.2680189932
31	Costa Rica	.4649547637	64	Zimbabwe	.2652173274
32	Belgium	.4517765622	65	Congo DR	.264536632
33	Honduras	.4424537562	66	Scotland	.2573454189

Tabulka 9. Žebříček fotbalových týmů , období 2002-2005, varianta 3

metoda	1.období	2.období	3.období	4.období	celková úspěšnost
1	0.489	0.569	0.531	0.542	0.493
2A	0.540	0.593	0.546	0.600	0.570
2B	0.516	0.573	0.531	0.598	0.506
3	0.532	0.583	0.543	0.598	0.561
Fifa	0.505	0.560	0.522	0.598	0.497

Tabulka 10. Porovnání úspěšnosti předpovědí u jednotlivých variant ohodnocení

Literatura

[1] Mark E. Glickman, Albyn C. Jones: *Rating the Chess Rating Systems*, Math.bu.edu/people/mg/papers/chance.ps

[2] Mark E. Glickman: *A comprehensive Guide to Chess Ratings*, http://math.bu.edu/people/mg/papers/acjpaper.pdf

[3] Henryk Minc: *Nonnegative Matrices*, A Wiley-Interscience Publication, 1988, str.5-25