Keeners Ranking Methode

Tabea Born

Überblick

1. Einführung
   - Motivation
   - Idee

2. Keeners Ranking Methode
   - Berechnung des Rankings
   - Beispiel

3. Was hat Google mit Keeners Ranking-Methode zu tun?

Teil 1:

EINFÜHRUNG

Motivation

TECHNISCHE UNIVERSITÄT DARMSTADT

- Streitigkeiten, welches College Football Team das Beste ist
- Umfragen und verschiedene mathematische Ranking-Methoden in Zeitungen sorgen für Verwirrung
- Viele Ranking-Methoden basieren auf Perron-Frobenius Theorem
- Ranking von (College Football) Teams als Möglichkeit, Studenten das Perron-Frobenius Theorem beizubringen

Idee

- Verwendung von nicht-negativen Statistiken zur Erstellung des Ratings

- Verknüpfen des *Ratings* eines Teams mit seiner *absoluten Stärke*
- *Absolute Stärke* eines Teams basiert auf seiner *relativen Stärke*

- Annahmen:
  1. Die *Stärke s* eines Teams wird durch Interaktionen mit Gegnern und deren Stärke ermittelt
  2. Das *Rating r* jedes Teams ist proportional zu seiner Stärke,

     d.h. für ein Team *i* gilt: $s_i = \lambda r_i$

Teil 2:

KEENERS RANKING METHODE

Schritt 1: Auswahl des Stärke-Attributs a_{ij}

- a_{ij} = Wert der Statistik, die bei einem Wettkampf von Team *i* gegen Team *j* erzeugt wurde
  - Beispiele:
    - Anzahl der Spiele, bei dem Team i gegen Team j gewonnen hat
    - Anzahl der Körbe / Tore / Punkte die Team i gegen Team j gemacht hat

- Sei a_{ij} die Anzahl der Punkte S_{ij} die Team i gegen Team j gemacht hat:
  - Rohe Daten nicht sinnvoll → Offensive Teams haben besseres Ranking
  - Besser: Totale Anzahl der erreichten Punkte verwenden $\boxed{a_{ij} = \frac{S_{ij}}{S_{ij} + S_{ji}}}$
  - Noch besser: Laplace'sche Regel anwenden $\boxed{a_{ij} = \frac{S_{ij} + 1}{S_{ij} + S_{ji} + 2}}$

Schritt 1: Beispiel

TECHNISCHE UNIVERSITÄT DARMSTADT

- Fußball-WM 2014: Gruppe G
  - Deutschland, Ghana, Portugal, USA
- Statistiken aus den WMen 2006, 2010 & 2014 herangezogen
- a_{ij} = Anzahl der Tore aller Spiele des Teams i gegen das Team j

$a_{ij} = S_{ij}$	Deutschland	Ghana	Portugal	USA
Deutschland	0	3	7	0
Ghana	2	0	0	5
Portugal	1	0	0	2
USA	0	4	2	0

Land	Gespielte Spiele (n_i)
DE	4
GH	5
PT	3
US	4

Schritt 1: Beispiel (Fortsetzung)

- Anwenden der Laplace'schen Regel:

$$a_{ij} = \frac{S_{ij} + 1}{S_{ij} + S_{ji} + 2}$$

a_{ij}	DE	GH	PT	US
DE	0	3	7	0
GH	2	0	0	5
PT	1	0	0	2
US	0	4	2	0

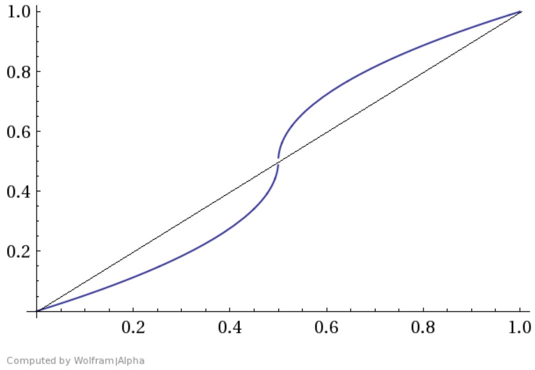
a_{ij}	DE	GH	PT	US
DE	0	4/7	4/5	½
GH	3/7	0	½	6/11
PT	1/5	½	0	½
US	½	5/11	½	0

Schritt 2: „Anpassen“ der Werte

- Kompensieren der „Weil wir es können“-Situation
- Nichtlineare *skewing*-Funktion:

$$h(x) = \frac{1}{2} + \frac{\operatorname{sgn}(x - \frac{1}{2})\sqrt{|2x-1|}}{2}$$

- Mäßigender Effekt auf starke Differenzen im oberen und unteren Bereich
- Künstliche Separierung von Werten, die nahe an ½ liegen



- Möglichkeit, das System für bestimmte Wettkämpfe anzupassen

Schritt 2: Beispiel

- *Skewing* in diesem Fall nicht notwendig!

- Alle Mannschaften etwa gleich stark
  - Teams waren alle bei jeder WM dabei
  - Teams haben meistens wenigstens das Achtelfinale erreicht

- WM-Siege aus vorherigen Jahrzehnten nicht aussagekräftig
  - Komplett andere Mannschaft / anderer Trainer
  - Evtl. Veränderung des Spielverhaltens?

Schritt 3: Normalisieren der Werte

$$a_{ij} \leftarrow \frac{a_{ij}}{n_i}$$

TECHNISCHE UNIVERSITÄT DARMSTADT

- Nicht jedes Team hat dieselbe Anzahl an Spielen gespielt

- Teams die mehrere Spiele gespielt haben, können mehr Punkte machen
  - Messung der Stärke wird beeinflusst

- ABER:
  - Skewing hat einen normalisierenden Effekt.
  - Wenn Skewing angewendet wird und es keine großen Unterschiede in der Anzahl von gespielten Spielen gibt, ist das Normalisieren nicht nötig

  ➢ Gefahr der Über-Normalisierung!

Schritt 3: Beispiel

- Da nicht alle Teams dieselbe Anzahl an Spielen gespielt haben, muss die Tabelle normalisiert werden:

a_{ij}	DE	GH	PT	US
DE	0	4/7	4/5	½
GH	3/7	0	½	6/11
PT	1/5	½	0	½
US	½	5/11	½	0

→

a_{ij}	DE	GH	PT	US
DE	0	1/7	1/5	1/8
GH	3/35	0	1/10	6/55
PT	1/15	1/6	0	1/6
US	1/8	5/44	1/8	0

Land	DE	GH	PT	US
Gespielte Spiele (n_i)	4	5	3	4

Schritt 4: Ranking berechnen – Basics

TECHNISCHE UNIVERSITÄT DARMSTADT

- Organisation der bearbeiteten Daten in eine Matrix:

  $A = \left[a_{ij}\right]_{m \times m}$, wobei m = Anzahl der Teams

  - ➢ Damit können Methoden der Matrixtheorie verwendet werden
  - ➢ Möglichkeit, ein numerisches Ranking zu erstellen

Schritt 4: Ranking berechnen – Rating

- $r_j(t)$ = numerisches Rating des Teams j zum Zeitpunkt t

- Explizite Zeitangabe t kann vernachlässigt werden

- Daher:

$$r = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{pmatrix} = \text{ der Rating-Vektor}$$

Schritt 4: Ranking berechnen – Stärke

- Keeners erste Annahme war:
  - Die *Stärke s* eines Teams wird durch Interaktionen mit Gegnern und deren Stärke ermittelt
- Relative Stärke:
  - Die relative Stärke eines Teams i gegen ein Team j ist definiert durch: $s_{ij} = a_{ij} r_j$
- Absolute Stärke:
  - Die absolute Stärke eines Teams i ist definiert durch: $s_i = \sum_{j=1}^{m} s_{ij} = \sum_{j=1}^{m} a_{ij} r_j$
  - Der Stärke-Vektor $s = (s_1 \quad s_2 \quad \cdots \quad s_m)^T$ kann auch so ausgedrückt werden:

$$s = \begin{pmatrix} \sum_j a_{1j} r_j \\ \sum_j a_{2j} r_j \\ \vdots \\ \sum_j a_{mj} r_j \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mm} \end{pmatrix} \begin{pmatrix} r_1 \\ r2 \\ \vdots \\ rm \end{pmatrix} = Ar$$

Schritt 4: Ranking berechnen – „Knackpunkt“

TECHNISCHE UNIVERSITÄT DARMSTADT

- Keeners 2. Annahme war:
  - Das *Rating r* jedes Teams ist proportional zu seiner Stärke, d.h. für ein Team *i* gilt: $s_i = \lambda r_i$
  - Also gilt $s = \lambda r$
- Auf der vorherigen Folie haben wir gesehen: $s = Ar$
- Daraus können wir folgende Gleichung schließen: $\boxed{Ar = \lambda r}$

Schritt 4: Ranking berechnen – „Knackpunkt“

$$\boxed{Ar = \lambda r}$$

- Lineare Algebra:
  - Proportionalitätsvektor λ = *Eigenwert*
  - Rating r = *Eigenvektor*

- Berechnung des Rankings = Berechnung des Eigenvektors?
  1. Für eine $m \times m$ - Matrix gibt es m verschiedene Eigenwerte λ
     - Wahl des Eigenwerts beeinflusst das Ranking
  2. Eigenwerte λ können komplexe Zahlen sein
     - Eigenvektoren enthalten komplexe Zahlen → unbrauchbar
  3. Eigenwerte λ können negative Zahlen sein → nicht optimal
  4. Berechnung der Eigenwerte und Eigenvektoren oft schwer
     - Beschränkung der Berechnung auf Näherungswerte bei großen Matrizen

Schritt 4: Ranking berechnen - Bedingungen

TECHNISCHE UNIVERSITÄT DARMSTADT

- Keener vermeidet genannte Probleme durch drei Bedingungen:

1. **Nicht-Negativität**
   - Die verwendeten Statistiken dürfen keine negativen Zahlen enthalten
   - $A = [a_{ij}] \geq 0$ ist eine nichtnegative Matrix

2. **Nicht-Reduzierbarkeit**
   - Zwei Teams i und j müssen miteinander verglichen werden können (auch wenn sie nicht gegeneinander gespielt haben)
   - $i \leftrightarrow k_1 \leftrightarrow k_2 \leftrightarrow \ldots \leftrightarrow k_p \leftrightarrow j$ mit $a_{ik1} > 0, \quad a_{k1k2} > 0, \quad \ldots, \quad a_{kpj} > 0$

3. **Primitivität**
   - Strengere Version der Nicht-Reduzierbarkeit
   - Alle Teams müssen durch <u>dieselbe</u> Anzahl an Spielen miteinander verbunden sein

Schritt 4: Ranking berechnen – Erzwingen von Nicht-Reduzierbarkeit und Primitivität

- Setze:

$$A \leftarrow A + E, \quad wobei \quad E = \varepsilon e e^T = \begin{pmatrix} \varepsilon & \varepsilon & \cdots & \varepsilon \\ \varepsilon & \varepsilon & \cdots & \varepsilon \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon & \varepsilon & \cdots & \varepsilon \end{pmatrix}$$

- Dabei ist *e* eine Spalte voller 1en und ε > 0 ist genügend klein gewählter Wert
- Effekt:
  - es wird ein künstliches Spiel zwischen allen Teams erzeugt
  - ε ist so klein gewählt, dass es die realen Spiele nicht beeinflusst
  - Jedes Team ist jetzt mit jedem anderen Team direkt verbunden

Schritt 4: Ranking berechnen – Perron-Frobenius Theorem

Sei $A = (a_{ij})$ eine reelle $n \times n$-Matrix mit positiven Einträgen $a_{ij} > 0$ und bezeichne $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A, so dass $|\lambda_1| > |\lambda_2| \geq \ldots \geq |\lambda_n|$. Dann gilt:

1. $\lambda_1 > |\lambda_2| \geq \ldots \geq |\lambda_n|$
2. Der Eigenraum zu λ_1 ist eindimensional.
3. Es gibt einen Eigenvektor zu λ_1, dessen Einträge alle positiv sind.

Schritt 4: Ranking berechnen

1. Prüfen, dass A nicht-reduzierbar ist

2. Eigenwerte und den Eigenvektor zum größten Eigenwert berechnen
   1. Mit Hilfe einer Software
      - In diesem Fall muss der Eigenvektor angepasst werden: $r = \frac{x}{\sum_{i=1}^{m} x_i}$
   2. Mit Hilfe der Potenzfunktion:

$$r = \lim_{k \to \infty} \frac{A^k x_0}{\sum_{i=1}^{m} A^k x_0}$$

Schritt 4: Beispiel

- Matrix A aus den Spieldaten herleiten:

a_{ij}	DE	GH	PT	US
DE	0	1/7	1/5	1/8
GH	3/35	0	1/10	6/55
PT	1/15	1/6	0	1/6
US	1/8	5/44	1/8	0

$$A = \begin{pmatrix} 0 & \frac{1}{7} & \frac{1}{5} & \frac{1}{8} \\ \frac{3}{35} & 0 & \frac{1}{10} & \frac{6}{55} \\ \frac{1}{15} & \frac{1}{6} & 0 & \frac{1}{6} \\ \frac{1}{8} & \frac{5}{44} & \frac{1}{8} & 0 \end{pmatrix}$$

- Eigenwerte λ der Matrix A berechnen:

$$\lambda_1 = -0{,}13217129761287308 - 0{,}05334107509050716 6i$$

$$\lambda_2 = -0{,}13217129761287308 + 0{,}053341075090507166i$$

$$\lambda_3 = -0{,}1107416332365628 3$$

$$\lambda_4 = 0{,}37508422846230 9$$

Schritt 4: Beispiel (Fortsetzung)

- Berechnen des Eigenvektors v für den Eigenwert $\lambda_4 \geq |\lambda_i|$:
  - $\lambda_4 = 0{,}37508422\ldots$

$$v = \begin{pmatrix} 0{,}5859 \\ 0{,}4094 \\ 0{,}5023 \\ 0{,}4867 \end{pmatrix}$$

- Rating-Vektor ermitteln:

$$r = \begin{pmatrix} 0{,}5859 \\ 0{,}4094 \\ 0{,}5023 \\ 0{,}4867 \end{pmatrix} \div \sum_{i=1}^{4} v_i = \begin{pmatrix} 0{,}2953 \\ 0{,}2063 \\ 0{,}2531 \\ 0{,}2453 \end{pmatrix}$$

Beispiel - Ergebnis

TECHNISCHE UNIVERSITÄT DARMSTADT

- Keine Spiele DE ⇔ US (seit 2002) oder GH ⇔ PT in einer WM

- Ranking der Mannschaften nach Keener:

Für WM 2006 + 2010 + 2014:

Team	Ranking
Deutschland	0,2953
Portugal	0,2531
USA	0,2453
Ghana	0,2063

Nur für WM 2014:

Team	Ranking
Deutschland	0,2886
USA	0,2661
Ghana	0,2403
Portugal	0,2050

- Was bedeutet das für die Spiele DE ⇔ US und GH ⇔ PT?

Was bedeutet das für die Spiele DE ⇔ US und GH ⇔ PT?

- Rankings kein objektives Abbild der Realität:
  - Keine lineare Relation zwischen Ranking und realen Spielen
  - Viele Faktoren werden nicht berücksichtigt (z.B. Heimvorteil, Verletzungen, …)
  - Aufstellung und Taktik der Mannschaft verändert sich
  - Spiele mit Gleichstand werden in dem Beispiel nicht berücksichtigt
  - …

➢Es ist nicht möglich, eine klare Aussage über den Spielausgang zu treffen

Teil 3:

WAS HAT GOOGLE MIT KEENERS RANKING-METHODE ZU TUN?

Was hat Google mit Keeners Rating-Methode zu tun?

- PageRank-Algorithmus

- Erstellt ein Ranking über Wichtigkeit von Webseiten
  - Genauer: Die Wichtigkeit einer Seite hängt von der Wichtigkeit der Seiten ab, die Links zu dieser Seite haben

- Logik ähnlich der von Keeners Rating-Methode
  - Es wird eine Hyperlink-Matrix definiert
  - Danach wird der Eigenwert bzw. der Eigenvektor zu Berechnung des Rankings ermittelt

- Perron-Frobenius Theorem hilft dabei, ein positives Ranking zu erhalten

VIELEN DANK FÜR EURE AUFMERKSAMKEIT

Quellen

- Literatur:
  - Who's #1? - The Science of Rating and Ranking, Kapitel 4
  - The Perron-Frobenius theorem and the ranking of football teams, James Keener
  - PageRank-Methode: Mathematik für Informatiker, Springer Verlag
- Bilder:
  - Keener: http://www.math.utah.edu/~keener/pictures_of_me/DSCN1349.JPG
  - Google: http://www.seo-united.de/blog/wp-content/uploads/2013/09/google_logo_2013.png
  - Football Team: http://dailyemerald.com/wp-content/uploads/2013/11/131116.RJK_.OEM_.FBC_.Utah_.1689.jpg
  - WM-Pokal: http://ssfkappishaeusern.de/assets/images/FifaPokal_WEB.JPG