MASSEY'S RANKING

AP

RATING AND RANKING

MOTIVATIONS

the ability to

- *rate* something (is this a warm day for June in London?), or to
- *rank* a set of elements (which is the coldest day of the month?)

is part of Science and Engineering since before Data Science.

Rating & ranking is a good framework to introduce Data Science techniques of general value and wide applicability.

Sports R&R is both fun and a huge Data Science market!

DEFINITION

A measure of value of the subject, as objective and replicable as possible.

E.g., temperature.

Normally, abilities are

- **latent**
- hard to measure
- time-dependent
- place-dependent

Exercise: take the Prof or Hobo? quiz!

yet, abilities are also

- hard to *transcend* (revert-to-the-mean effect, RTTM)
- relatively easy to perceive and project

EXAMPLE: FOOTBALL

- hard to guess the single score $\implies$ entertainment value
- easy for experts to guess the long-term effect $\implies$ different levels of enjoyment; RTTM: Revert To The Mean effect

Low scoring creates **randomness**

FORMALISATION

1-DIMENSIONAL RANKING

P : players, $|P| = n$

T : time instants

$r : P \times T \to \mathbb{R}$

A given rating function r creates a ranking (ρ) on a set:

$$\rho : P \times T \to [1..n]$$

$$\rho(p,t) = k \leftrightarrow |\{p_j : r(p_j,t) \leq r(p_i,t)\}| = k$$

$$\delta(p_i,p_j,t) = |r(p_i,t) - r(p_j,t)|$$

δ captures both similarity and distance

MULTI-DIMENSIONAL RANKING

Multi-dim. rating:

$$r_{multi} : P \times T \to \mathbb{R}^d$$

Often:

$$r_{multi}(p_i, t) : f(r_1(p_i, t), \ldots r_d(p_i, t))$$

Pareto dominance:

p_i dominates p_j (at time t) if on every dimension x

$$r_x(p_i, t) \geq r_x(p_j, t)$$

RATING IN GAMES

RATINGS IN GAMES

- score-based games are better-suited to create ratings
- yet effect of time and hardness of the proposed test match could be hard to assess.

SHOULD GAMES KEEP USER RATINGS?

YES:

- feeling of improvement

...

- a gauge for new features

...

- leads to rankings:
  - better matchmaking $\Longrightarrow$ entertainment value
  - fraud/anomaly detection?

NO: GAME PROWNESS AS SOCIAL RANKING?

The spectacle is a social relation mediated by images, not a collection of images.

«Le spectacle n’est pas un ensemble d’images, mais un rapport social entre des personnes, médiatisé par des images»

[Guy Debord, La Société du spectacle (1967), Thèse 4]

- a reflection of US culture?
- a turn-off for people who don’t feel competive?
- turns-off casual users?

SPORT RANKING/ESTIMATION

DOMAIN

- n teams play each other in a tournament
- final scores are recorded, e.g., Real Madrid–Borussia Dortmund: 2-0.
- predict the score for a match in the future.

-focus on predicting the score difference (eg, 2-0=2)

RUNNING EXAMPLE

	Duke	Miami	UNC	UVA	VT	Record	Point Differential
Duke		7-52	21-24	7-38	0-45	0-4	-124
Miami	52-7		34-16	25-17	27-7	4-0	91
UNC	24-21	16-34		7-5	3-30	2-2	-40
UVA	38-7	17-25	5-7		14-52	1-3	-17
VT	45-0	7-27	30-3	52-14		3-1	90

the win-loss balance and the points balance are second-level performance measures

they are not considered sufficient to create valuable ratings/rankings/predictions.

[IN]CREDIBLE ASSUMPTIONS

1. to each team a **latent** variable for *strength* is assigned

numerical **ratings** determine a **ranking** among teams (at t=end, so we can drop it)

and a prediction $Pr[a \to b] = \frac{\rho(a)}{\rho(a)+\rho(b)}$

2. strength/rating is immutable during the tournament

3. teams play each other exactly once during the tournament

Now, consider the score difference in each match, say i vs. j, defined as $s_i - s_j$

Define $\mathbf{y}_{m \times 1}$ as the vector of all score differences in matches

Assume (assumption 4) that strength/rating imbalance determines score difference:

$$r_i - r_j = s_i - s_j$$

$$X_{m\times n}\cdot\mathbf{r}_{n\times 1}=\mathbf{y}_{m\times 1}$$

$$\begin{bmatrix} 0 & 0 & +1 & 0 & -1 & 0 \\ & 0 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \ddots & \\ 0 & -1 & 0 & +1 & 0 & 0 \end{bmatrix}$$

$X_{m\times n}$ with $m >> n$ is overconstrained, no hope of finding a solution.

MASSEY'S RATINGS

DATA PREPARATION

Massey considered the equivalent formulation of

$$X_{m \times n} \cdot \mathbf{r}_{n \times 1} = \mathbf{y}_{m \times 1}$$

as

$$X^T \cdot X \cdot \mathbf{r} = X^T \cdot \mathbf{y}$$

Both sides are easier to work with.

On the right-and side, $X^T \cdot \mathbf{y}$ is the all-season points difference vector, called $\mathbf{p}$.

Notice that $\sum p_i = 0$.

On the left-hand side,

$$M_{n \times n} \; = \; X^T X$$

is squared, semidefinite and positive.

However, the rows sum to 0 and cols. are not independent: 0/∞ solutions ensue…

M. also noticed that *M* has a fixed structure and does not need to be re-computed all the times.

$$
\begin{bmatrix}
n-1 & 0 & -x & 0 & -y & 0 \\
 & n-1 & \ddots & \ddots & & \\
 & \ddots & \ddots & \ddots & \ddots & \\
 & \ddots & \ddots & \ddots & \ddots & \\
 & & \ddots & \ddots & \ddots & \\
0 & -z & 0 & -w & 0 & n-1
\end{bmatrix}
$$

$m_{i,i} = n - 1$ is the numbers of games i played,

$m_{i,j}$ is the negation of the no. of matches between i and j : here all values are set to -1.

$$\begin{pmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 4 & -1 & -1 & -1 \\ -1 & -1 & 4 & -1 & -1 \\ -1 & -1 & -1 & 4 & -1 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{pmatrix} = \begin{pmatrix} -124 \\ 91 \\ -40 \\ -17 \\ 0 \end{pmatrix}$$

MASSEY

1. drops the last row/match
2. replaces it with a row of 1s, and sets $p_n = 0$

(all ratings, positive and negative, will sum to 0)

$\overline{M} = M$ everywhere but for the last row which is full of 1s

$\overline{\mathbf{p}}$ is $\mathbf{p}$ everywhere but for the last el. $p_n = 0$.

1. now $\overline{M}$ is non-singular and invertible
2. solves

$$\overline{M}\mathbf{r} = \bar{\mathbf{p}}$$

to obtain an approximated rating for the teams.

The MSE solution to Massey's formula is a form of **regression.**

It can also be seen as $\mathbf{r} = (\overline{X^T X})^{-1} \overline{X^T \mathbf{y}}$.

OUTPUT

ratings sum to zero

values have no direct interpretation.

however, they effectively generate **a hierarchy.**

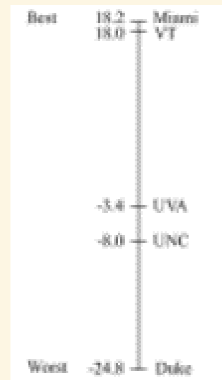
Team	Rating r	Rank
Duke	-24.8	5th
Miami	18.2	1st
UNC	-8.0	4th
UVA	-3.4	3rd
VT	18.0	2nd

VISUALISE THE RANKING

Ratings are not necessarily meaningful, as a result of the matrix preparation.

Rankings are meaningful:

Team	Rating r	Rank
Duke	-24.8	5th
Miami	18.2	1st
UNC	-8.0	4th
UVA	-3.4	3rd
VT	18.0	2nd



CONCLUSIONS

POINTS TO FOCUS ON

- rating and rating is the fun side of Data Science!

- *latent* variables that represent *non-measurable* skills
- those leave in a ***feature space*** possibly separated from the ***data space***
- yet they may get a numeric estimate, and inform our predictions
- Massey regresses on the latent variables

FURTHER READINGS