Assessing the Probability of the Soccer Match Outcome

Presented by Xinyuan Zhao, Shanqing Yu, Arman Aydemir

2020/8/7





Questions to answer	1
Data preparation work	2
Tools used	3
Technique used	4
Knowledge gained	5
Application	6



Question to answer

The Project Description

The goal of this project is primarily to create a prediction model which can assess the probabilities of the soccer match outcome.

Questions to answer

- 1.Are there significant betting opportunities for soccer betting?
- 2. Is it possible to build a successful model for predictions of the results for soccer matches



Data preparation work 2

Data preparation work



Data Cleaning

Only the information about the numbers of the goal scored, when and where the score happens is relevant, other information such as the referee's name, the number of yellow or red cards will be discarded.



Data Preprocessing

The ratio of frequencies of home wins, draws and away wins will be used to determine the home advantage. The dependency between home and away score will also be checked



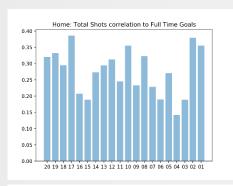
Data Integration

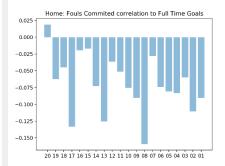
Data of matches in the 20 years in Premier League have been integrated.

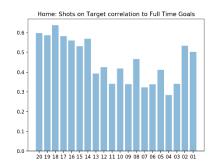
Attributes Chosen

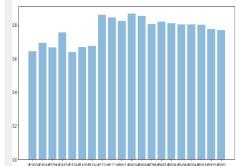
The primary attributes we are using are-Date - Interval (String)

- Home Team, Away Team -Nominal (String)
- Final Result Nominal (Char)
- Goals, Shots On Target,
 Total Shots, Fouls, Yellow
 Cards, Red Cards (for both
 Home and Away) Nominal
 (Int)











Tools used

Attributes Chosen

The tool we used can be listed as follow:

{1}pandas

{2}numpy

{3}scrapy (A web scraper to scrap online data)

{4}sklearn

{5}matplotlib



Technique used

Dixon – Coles Model

Dixon- Coles is a model given by Dixon and Cole in 1983. The whole model is based on the likelihood function.

$$\mathrm{L}(\alpha_{i},\beta_{i},\rho,i=1,2,3,4..) = \prod_{k \in A_{t}} \left\{ \tau_{\lambda_{k}}\left(x_{k},y_{k}\right) \exp\left(-\lambda_{k}\right) \lambda_{k}^{x_{k}} \exp\left(-\lambda_{k}\right) \mu_{k}^{x_{k}} \right\}^{\left\{\phi(t-t_{k})\right\}}$$

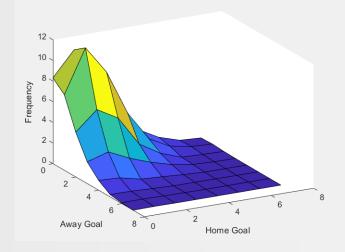
And in Dixon and Coles model, they divided the strength of a team into two parts: attack ability and defend ability, and consider them are independent Poisson variables, then make them two individual input of likelihood function.

Assumptions of Dixon Coles Model

There is two assumptions of Dixon Coles model. The first one is that the frequency of different score satisfy Poisson Distribution

We count the frequency of different scores and find it satisfy Poisson Distribution. Then we paint a figure about it.

Table: Empirical estimates of each score probability									
		AWAY GOAL							
		0	1	2	3	4	5	6	7
	0	8.31	7.32	4.34	2.11	0.96	0.24	0.09	0
Н	1	10.68	11.52	6.18	2.73	0.86	0.21	0.1	0.02
	2	8.34	8.81	5.02	1.77	0.39	0.12	0.05	0
M E	3	4.21	4.48	2.15	1.03	0.27	0.04	0.02	0
G O	4	1.93	1.67	0.91	0.46	0.13	0.03	0	0
A L	5	0.82	0.59	0.18	0.12	0.03	0.01	0	0
	6	0.18	0.22	0.1	0.03	0.01	0	0	0
	7	0.06	0.11	0.04	0.01	0.01	0	0	0



Assumptions of Dixon Coles Model

The second assumption is that home goals and away goals are independent Poisson variables.

$$\frac{\widetilde{f}(i,j)}{\widetilde{f}_{H}(i)\,\widetilde{f}_{A}(j)}$$

Table: Estimates of the ratios of the observed joint probability function and the empirical probability function									
		AWAY GOAL							
		0	1	2	3	4	5	6	7
	0	1.03	0.9	0.98	1.09	1.54	1.6	1.48	0
	1	0.96	1.03	1.01	1.03	1	1.02	1.19	3.1
H O	2	0.99	1.04	1.08	0.88	0.6	0.73	0.78	0
M E	3	1	1.06	0.93	1.02	0.83	0.49	0.63	0
G O	4	1.09	0.94	0.94	1.09	0.98	0.87	0	0
A L	5	1.36	0.97	0.55	0.8	0.62	0.86	0	0
	6	0.98	1.18	0.94	0.65	0.67	0	0	0
	7	0.76	1.38	0.92	0.53	1.63	0	0	0

Under most case, we can find the number of home goals and away goals are independent



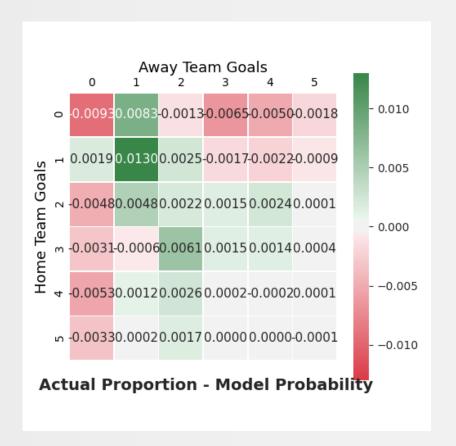
Knowledge gained

Departure from Poisson Model

Correlation parameter

$$\tau\left(x,y,\gamma\right) = \begin{cases} 1 - \lambda\mu\rho & if x = y = 0\\ 1 + \lambda\rho & if x = 0, y = 1\\ 1 + \mu\rho & if x = 1, y = 0\\ 1 - \rho & if x = y = 1\\ 1 & otherwise \end{cases}$$

$$max(-1/\lambda, 1/\mu) \le \rho \le min(-1/\lambda\mu, 1)$$



Parameter Estimation

Given the likelihood function, we can calculate the attack and defend parameter in 19-20 season.

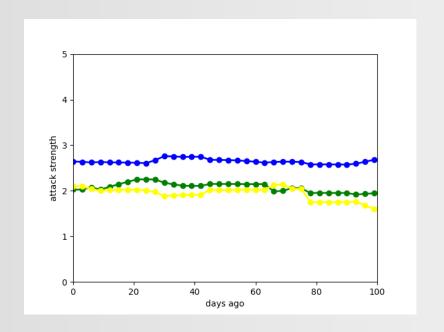
Maximum likelihood function

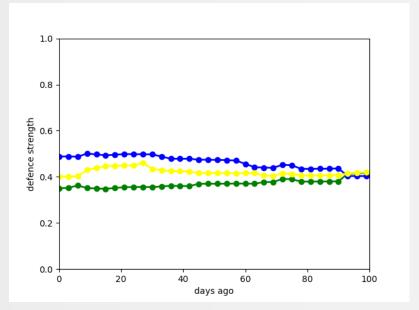
$$\begin{split} L\left(\alpha_{i},\beta_{i},\rho,i=1,2,3,4..\right) &= \prod_{k \in A_{t}} \left\{ \tau_{\lambda_{k}}\left(x_{k},y_{k}\right) \right. \\ &\left. exp\left(-\lambda_{k}\right) \lambda_{k}^{x_{k}} exp\left(-\lambda_{k}\right) \mu_{k}^{x_{k} \left\{\phi\left(t-t_{k}\right)\right\}} \right. \end{split}$$

Table 1: Attack parameter α and Defend parameter β in 19-20 season

Team	$log(\alpha)$	log (β)
Arsenal	1.134194461	-0.9388646131
Aston Villa	0.8428578212	-0.6189790245
Bournemouth	0.8120741353	-0.6519458441
Brighton	0.7809206483	-0.8319848409
Burnley	0.8736217223	-0.9169827088
Chelsea	1.3406671569	-0.8095874718
Crystal Palace	0.5402903267	-0.9276208686
Everton	0.904040090	-0.7978082567
Leicester	1.3057729257	-1.0774228343
Liverpool	1.53881068376	-1.2883023793
Man City	1.7197223805	-1.2127458303
Man United	1.2833265314	-1.2167372006
Newcastle	0.7560667438	-0.7732833065
Norwich	0.38629771197	-0.5236456429
Sheffield United	0.7571901683	-1.1712131162
Southampton	1.046302694	-0.7286538126
Tottenham	1.216391527	-0.9507146131
Watford	0.7122211110	-0.6644120721
West Ham	1.0118122088	-0.692856079
Wolves	1.0374189490	-1.1199782123
Mean	1.0000	-0.8957

Fluctuations of the Parameter





Predictions

Table 2: The prediction of given teams

Home Team	Away Team	Probability of Home Win	Probability of Away Win	Probability of Drawn
Arsenal	Watford	0.68435	0.10981	0.20591
Burnley	Brighton	0.51685	0.18444	0.29869
Chelsea	Wolves	0.39325	0.3345	0.26822
Crystal Palace	Tottenham	0.18543	0.53088	0.28367
Everton	Bournemouth	0.56457	0.18075	0.25467
Leicester	Man United	0.34908	0.37817	0.27274
Man City	Norwich	0.95967	0.00547	0.02986
Newcastle	Liverpool	0.14106	0.64365	0.21222
Southampton	Sheffield United	0.39826	0.28137	0.32036
West Ham	Aston Villa	0.58609	0.18475	0.22913

Thank You

