NFL 美式足球分析競賽

第一組

2019/6/16

團隊介紹:

05115246 社三 B 黃聖崴

07170144 巨一 A 陳勤

07170101 巨一 A 塗晨睿

07170128 巨一 A 周嘉竑

07170135 巨一 A 蔡欣成

07170171 巨一 A 陳嘉豪

07170121 巨一 A 許哲聞

07170138 巨一 A 簡呈澔

壹、專案介紹

一、Kaggle 資料科學競賽平台介紹

Kaggle 是一個數據建模和數據分析競賽平台。通常由各大企業或廠商提供資訊、獎金發包給此平台。然後有興趣的資訊科學家或業餘玩家可以進行資料分析。並且會即時的知道自己的預測精準度，最後在比賽時間截止前得出最佳模型的獲勝可得到獎金。2017 年 google 宣布要收購 kaggle。

二、本專案使用之 kaggle 資料科學比賽介紹

本專案的比賽內容為 NFL(美國國家橄欖球聯盟)發包。獎金為 8 萬美金，約 240 萬台幣。本競賽目的並非進行機器學習，而是要經過資料分析得出結果後，提出針對現有規則的修改，已達到使運動員減少受傷(尤其是腦震盪)的目標。

三、本專案之資訊管理應用系統動機與內容介紹

由於本專案需要完成規則修改，之外還須提出四個查詢模組。我們預計都將針對球員的受傷問題等進行分析，為了方便處理官方提供的大量表格資料。因此我們需要使用資料庫管理系統進行資料的整理、分析。我們採用 MySQL Server 連線 PHP My admin 的方式進行資料處理。此外過大的檔案會利用 MySQL Workbench 進行匯入。

針對規則修改的部分，我們將先處理各種球員受傷原因以及脈絡來進行分析已提出規則修改。此外還包刮以下模組:

1. 哪個球隊的主場受傷人數最多，可得知哪些球隊戰術上可能是較為激進的。
2. 各類型天氣分別的受傷次數，因為美式足球大多是戶外場地，不同天氣可能造成的事故機率不同。
3. 各種部位的受傷次數，可得知哪種受傷方式最多，並思考如何避免。
4. 各節數受傷的人數，如此一來可得知比賽在哪個環節將最激烈。
5. 查詢資料中給出的踢球在各節的發生次數，用來計算傷停加時或是延長比賽。

貳、研究方法

一、關聯式資料庫介紹

關聯式資料庫的定義，指的是由兩個或兩個以上的資料表組合而成，資料表之間透過相同的欄位值來連結。

組成元素包含資料庫(Schema)、資料表(Table)、欄位(Column)、欄位值(Value)、關聯(Relation)

使用方法將各種資料依照性質的不同，例如本專案就是將玩家資料、比賽資料、球員受傷資料分別存放在幾個不同的表格中，表格與表格之間的關係則以共同的欄位值相互連結例如球員資料與球員受傷資料都有 GISSID 可以連結。

二、Rmarkdown 文件介紹

Rmarkdown 可以呈現成 Html 型式也能轉乘 PDF 檔案，這種文件常用在資料科學的報告上。原因在於其一目了然、方便、有效率的功能，資料科學家不必再將程式碼複製到 word 而是使用 Rmarkdown 就能清楚呈現。凡是寫在 R 上的程式碼結果都會幫你呈現在該文件上，並且也能直接在程式碼或結果下方寫下自己的註解和解釋。也能利用一些功能進行美化、資料視覺化。

參、研究資料

一、kaggle 比賽資料集說明

1. Game Data: 比賽級別數據，指定賽季類型（前，中，後），週數、主辦城市和隊伍。每個比賽都是使用的唯一標識 Gamekey。

2. Play Information: 描述遊戲類型，參與隊伍，得分和每個遊戲的簡要敘述。PlayID 會對應比賽唯一識別的 GameKey。並且 PlayID 不是唯一的。

3. Player Punt Data: 球員級數據，指定每個球員的比賽擔任位置。每個玩家用 GSISID 進行識別。

4. Play Player Role Data: 該數據集指定每個比賽中的每個球員擔任的角色。用 Gamekey,PlayID 和唯一定義 GSISID 進行識別。

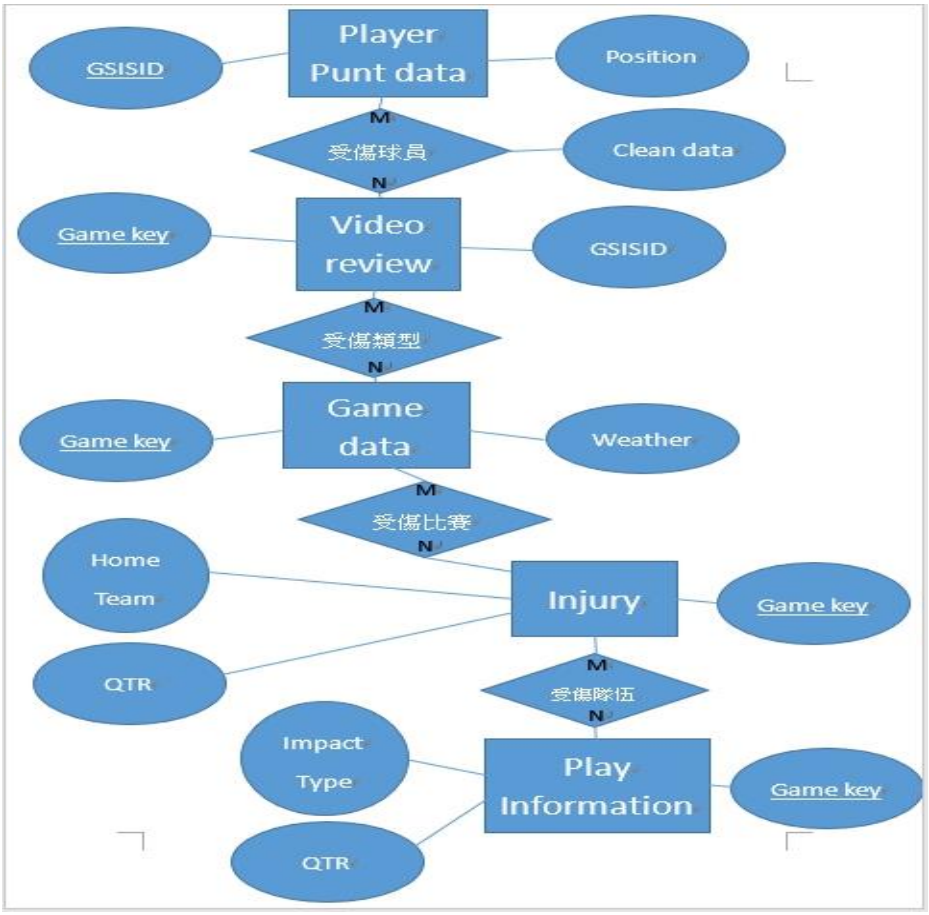
5. Video Review: 提供有關腦震盪產生事件的詳細描述。影片評論數據僅適用於可識別傷害發生的情況。每個視頻審核的情況下可以使用 GameKey，PlayID 和 GSISID。進行識別，並提供了對比賽事件的簡要敘述。

6. NGS: Next Gen Stats: 描述每個玩家在遊戲過程中的移動。BIOCORE 處理 NGS 數據以產生相關的速度和方向數據。該 NGS 的數據，使用 GameKey，PlayID 和 GSISID 進行識別。每個遊戲的玩家數據作為遊戲持續時間的時間（時間）的函數提供。

二、本專案整理之資料表說明

我們將重點放在球員受傷的分析因此使用資料有 Gamedata、video_review、Play Information、Player_Punt_Data 以及 video_review 中細分出的 injury 資料進行分析。

三、本專案資料庫 ERD 實體關係圖



A caption

肆、資訊管理應用系統

一、本專案之資訊管理應用系統架構說明

本專案的資訊管理應用系統目的為以下六種

1. 完成比賽目標，修改規則
2. 哪個球隊的主場受傷人數最多
3. 各類型天氣分別的受傷次數哪種最多
4. 各種部位的受傷次數哪種最多
5. 各節數受傷的人數哪節最多
6. 查詢資料中給出的踢球在各節的發生次數

開始操作

安裝套件

```
#install.packages("DBI")
library(DBI)
#install.packages("RMySQL")
library(RMySQL)
#install.packages("rstudioapi")
```

連結資料庫

```
connect<-DBI::dbConnect(drv=RMySQL::MySQL(),
                        host="127.0.0.1",
                        port=8889,
                        user="guest2",
                        password="12345678",
                        dbname="nfl")
```

1.完成比賽目標，修改規則

創建需要的資料表格–clean_data(已完成創建)

```
#rs<-dbSendQuery(connect,
#"CREATE TABLE clean_data
#SELECT p.GSISID as p_ID, p.Position,v.GSISID as v_ID,v.Primary_Impact_
Type
#FROM player_punt_data as p
#INNER JOIN video_review as v
#ON p.GSISID = v.GSISID
#ORDER BY `p`.`Position` ASC
#")
#計算哪一個球員的位置最容易受傷
rs1<-dbSendQuery(connect,
"SELECT position,COUNT(position)
FROM clean_data
GROUP BY position
")
data_1<-dbFetch(rs1)
knitr::kable(data_1)
```

position	COUNT(position)
CB	7
FB	1
FS	4
ILB	9
LS	4
MLB	3
OLB	6
P	1
RB	5
S	2
SS	1
TE	9
WR	6

得出的結果為 ILB(內線衛)–阻擋進攻 9(防守方)、TE(邊鋒)—側邊進攻 9(進攻方)有趣的是這兩個位置的選手將會是死對頭互相阻擋。

2017 年三月，NFL 通過了新的“頭盔防守”判罰標準，旨在限制球員低頭用頭盔接觸對手的動作。這個動作容易造成腦震盪因此被停止。

現有規定擒抱要抱腰部以上，但沒有規定方向，後面的模組會提到的最常受傷的方式有兩種，一種是頭部對撞，一種是頭部對身體撞擊。兩者都會造成腦震盪—這也是這場比賽最需要防止的。尤其上述兩個位置的選手最常頭部對撞。因此我們認為必須修改擒抱持球者的方式，必須從側邊腰部以上進行擒抱，如此一來可以減少頭部對撞的狀況，並且側邊的體積較小，應該也能有效減少頭部撞擊身體的狀況。

2.哪個球隊的主場受傷人數最多

計算哪一個隊伍擔任主場時受傷人數最多

```
rs2<-dbSendQuery(connect,
"SELECT home_team,count(home_team)
FROM injury
group by home_team")
data_2<-dbFetch(rs2)
knitr::kable(data_2)
```

home_team	count(home_team)
Atlanta Falcons	1
Baltimore Ravens	2
Buffalo Bills	1
Carolina Panthers	1
Chicago Bears	1
Cleveland Browns	1
Denver Broncos	2
Detroit Lions	1
Houston Texans	2
Indianapolis Colts	2
Jacksonville Jaguars	1
Kansas City Chiefs	3
Miami Dolphins	2
New England Patriots	1
New Orleans Saints	1
New York Giants	1
New York Jets	2
Oakland Raiders	1
Philadelphia Eagles	1
San Francisco 49ers	1
Seattle Seahawks	2
Tennessee Titans	3
Washington Redskins	4

由此結果可以看出哪個隊伍擔任主場時容易造成傷害，可推估某些隊伍可能比較激進。

3.各類型天氣分別的受傷次數哪種最多

計算哪一種天氣時最容易受傷

```
rs3<-dbSendQuery(connect,
"SELECT GameWeather  , COUNT(GameWeather)
FROM game_data
INNER JOIN video_review ON game_data.GameKey=video_review.GameKey
GROUP by GameWeather ORDER BY COUNT(GameWeather) DESC")
data_3<-dbFetch(rs3)
knitr::kable(data_3)
```

GameWeather	COUNT(GameWeather)
Sunny	12
Cloudy	7
Partly Cloudy	5
Clear	4
	3
Mostly cloudy	3
Clear and warm	1
Controlled Climate	1
Partly Cloudy, Chance of Rain 80%	1

由此結果可以看出晴天最容易受傷，可能是因為溫度較高使得選手情緒較高漲或暴躁。

4.各種部位的受傷次數哪種最多

計算各種部位的受傷次數哪種最多

```
rs4<-dbSendQuery(connect,
"SELECT Primary_Impact_Type ,
COUNT(Primary_Impact_Type)
FROM video_review
GROUP BY Primary_Impact_Type")
data_4<-dbFetch(rs4)
knitr::kable(data_4)
```

Primary_Impact_Type	COUNT(Primary_Impact_Type)
Helmet-to-body	17
Helmet-to-ground	2
Helmet-to-helmet	17
Unclear	1

由此結果可以看出哪種傷害類型最多，連結第一模組，可得證頭部對撞、頭部撞擊身體是最多的，同時在比賽描述中也應是最嚴重的傷害，因此需要避免。

5.各節數受傷的人數哪節最多

計算哪一個節數受傷的人數哪節最多

```
rs5<-dbSendQuery(connect,
"SELECT Qtr ,
COUNT(Qtr)
FROM injury
GROUP BY Qtr")
data_5<-dbFetch(rs5)
knitr::kable(data_5)
```

Qtr	COUNT(Qtr)
1	5
2	11
3	14
4	7

由此結果可以看出第二、三節在比賽中相對容易受傷。

6. 查詢資料中給出的踢球在各節的發生次數

```
rs6<-dbSendQuery(connect,
"Select quarter,count(quarter)
from play_information
group by quarter")
data_6<-dbFetch(rs6)
knitr::kable(data_6)
```

quarter	count(quarter)
1	1628
2	1815
3	1588
4	1622
5	28

表中的第五列通常是因為傷停加時或是延長。

伍、結論與未來建議

我們最後得出結論，關於比賽內容最多也最嚴重的受傷便是頭部的對撞，以及以頭部撞擊身體。由於該競賽早已結束，其得出的解果便是限制低頭衝撞對手。然而卻未解決頭部對撞的問題，以及哪些位置的成員最容易遭受傷害，因此我們補充了針對最容易受傷的 TE、ILB 位置選手頭部對撞的規則解決方案。

然而我們能力有限沒辦法做出非常優秀的計算模型，因此只能針對狀況進行主觀推測，未來有興趣者可以針對這個方向進行壹些資料的演算。

此外模組 2 中我們未考慮客隊，因此未來可進行客隊與主隊的比較來得出隊伍戰術風格可以使這個查詢模組更為可靠。