



UNIVERSITÀ DEGLI STUDI DI
MILANO-BICOCCA

F1801Q127

MODELLI PROBABILISTICI PER LE DECISIONI

Predicting soccer results

Studenti:

Basso Matteo

Ferri Marco

Matricole:

807628

807130

Giugno 2019

Indice

1	Abstract	2
2	Introduzione	2
2.1	Dominio di riferimento	2
2.2	Obiettivi dell'elaborato	2
2.3	Scelte di design per la creazione del dataset	3
2.4	Ipotesi e assunzioni	3
3	Dataset	4
3.1	Formato	4
3.2	Giocatori	4
3.3	Squadre	6
3.4	Partite	7
4	Applicazione web	8
4.1	Architettura	8
4.2	Interfaccia grafica	8
4.3	Predizione risultato	9
5	Conclusioni	10

1 Abstract

Il calcio rappresenta, soprattutto negli ultimi anni, lo sport maggiormente diffuso in vari paesi del mondo. Molte persone seguono con grande attenzione tutte le partite cercando di capire preventivamente il vincitore, per piacere personale oppure per giocare nel mercato delle scommesse. Ci si è posti quindi la domanda se tale predizione possa essere effettuata da un computer. L'obiettivo di questo progetto risulta dunque quello di predire il risultato delle partite di calcio basandosi sulla formazione e sulle caratteristiche dei giocatori coinvolti.

2 Introduzione

Viene qui presentata una visione generale del progetto, ovvero il dominio di riferimento e gli obiettivi che si pone, le scelte di design per la creazione del dataset ed eventuali ipotesi o assunzioni fatte durante lo sviluppo dell'elaborato.

2.1 Dominio di riferimento

Il calcio rappresenta, soprattutto negli ultimi anni, lo sport maggiormente diffuso in vari paesi del mondo. Molte persone seguono con grande attenzione tutte le partite cercando di capire preventivamente il vincitore, per piacere personale oppure per giocare nel mercato delle scommesse.

A seguito di questo fenomeno sono stati creati svariati portali che permettessero alle persone di entrare sempre di più in questo mondo. Un aspetto fondamentale per ogni piattaforma di questo genere è quello di predire il vincitore di un match simulando partite di calcio con particolari formazioni.

2.2 Obiettivi dell'elaborato

Nel seguente elaborato verranno esplicitati i criteri che hanno portato alla scelta della creazione del dataset e alle modalità di acquisizione dei dati dalla sorgente. Successivamente, saranno presentate alcune analisi qualitative e descritti i procedimenti svolti per l'integrazione e l'estrazione delle informazioni.

Successivamente all'esportazione del dataset verrà creato ed esplorato il training set per la creazione della rete Bayesiana, utilizzando diversi modelli di apprendimento, condotti degli esperimenti e analizzati i risultati.

L'obiettivo del progetto è quello di predire il team vincitore di un determinato match, incrociando le informazioni del match stesso con quelle dei giocatori presenti in campo.

Viene presentata infine una semplice interfaccia grafica utile a mostrare l'effettivo impiego della rete in un prodotto web.

2.3 Scelte di design per la creazione del dataset

Per la creazione del dataset utile per la creazione della rete è stato scelto di rappresentare i dati in formato CSV in quanto più semplice e pratico da utilizzare rispetto al database relazionale inizialmente fornito. Sarebbe infatti stato necessario dal formato relazione ricondursi ad una unica tabella per far sì che possa essere processata dal linguaggio R. Questo potrebbe portare tuttavia ad alcune problematiche relative la ristretta sintassi e funzionalità di SQLite.

2.4 Ipotesi e assunzioni

Al fine di sviluppare il seguente progetto e procedere quindi con la stesura del seguente elaborato, sono state fatte delle assunzioni che fungono da base portante.

Assumiamo per prima cosa una divisione del campo in varie aree per classificare i ruoli dei giocatori in difesa, centrocamp e attacco. Non è infatti fornita tale informazione nel dataset di partenza.

In secondo luogo supponiamo che le caratteristiche del team siano effettivamente determinate solo dai giocatori che ne fanno parte, rendendo di fatto poco utile tale informazione al variare dei giocatori. Non avendo a disposizione inoltre informazioni circa l'appartenenza di un giocatore ad un determinato team nel tempo, non è possibile creare delle squadre ben formate. Esse si potrebbero ricavare direttamente dai match svolti, tuttavia ciò comporterebbe la creazione di dati incoerenti col mondo reale. Non è infatti assicurato che un giocatore non abbia cambiato squadra nel tempo o non sia stato assegnato ad una squadra errata poiché presente in quella nuova solo in panchina.

3 Dataset

Di seguito viene presentata la descrizione del dataset.

3.1 Formato

Il dataset a disposizione per lo svolgimento dell'elaborato consiste in un database SQLite [3], ovvero un file con l'omonima estensione, in grado di essere condiviso e utilizzato su un qualsiasi dispositivo. Sia dal linguaggio di programmazione Javascript che da R è stato possibile accedervi semplicemente installando un'apposita libreria.

3.2 Giocatori

I giocatori sono espressi all'interno di 2 tabelle differenti nello schema del database considerato. La prima relazione denominata *Player* contiene informazioni di carattere generale e viene descritta nella tabella 1.

Tabella 1: Player

Campo	Descrizione
id	Id del record
player_api_id	Identificativo univoco del giocatore
player_name	Nome completo del giocatore
player_fifa_api_id	Id utile per reperire le informazioni da fifa
birthday	Data di nascita
height	Altezza in centimetri
weight	Peso in libbre

La seconda tabella, denominata *PlayerAttributes* contiene invece i punteggi associati alle abilità del giocatore. Essa contiene inoltre un campo che indica a quale data tali punteggi si riferiscono. Può essere infatti che essi cambino nel tempo e per questo è necessario selezionare quelli giusti in base alla data della partita. Nella tabella [?] viene descritta tale relazione.

Tabella 2: PlayerAttributes

Campo	Descrizione
id	Id del record
player_api_id	Identificativo univoco del giocatore
player_fifa_api_id	Id utile per reperire le informazioni da fifa
date	Data di riferimento del record
overall_rating	Valutazione generale
potential	Potenziale
preferred_foot	Piede preferito
attacking_work_rate	Punteggio di attacco
defensive_work_rate	Punteggio di difesa
crossing	Cross
finishing	Capacità di segnare
heading_accuracy	Tiro di testa
short_passing	Passaggio corto
volleys	Tiro al volo
dribbling	Dribbling
curve	Data di nascita
free_kick_accuracy	Accuratezza calcio di punizione
long_passing	Passaggio lungo
ball_control	Controllo di palla
acceleration	Accelerazione
sprint_speed	Velocità sprint
agility	Agilità
reactions	Altezza in centimetri
balance	Equilibrio
shot_power	Potenza di tiro
jumping	Salto
stamina	Stamina
strength	Forza
long_shots	Tiro lungo
aggression	Aggressività
interceptions	Intercettare
positioning	Posizionamento
vision	Vista
penalties	Falli commessi
marking	Capacità di marcare
standing_tackle	Contrasto
sliding_tackle	Scivolata
gk_diving	Tuffo portiere
gk_handling	Presa portiere
gk_kicking	Calcio da portiere
gk_positioning	Posizionamento portiere
gk_reflexes	Riflessi portiere

3.3 Squadre

Le squadre sono espresse in modo analogo ai player tramite 2 tabelle differenti, *Team* 4 e *TeamAttributes* 3 riportate di seguito.

Tabella 3: Team

Campo	Descrizione
id	Id del record
team_api_id	Identificativo univoco della squadra
team_name	Nome completo della squadra
team_fifa_api_id	Id utile per reperire le informazioni da fifa
date	Data di riferimento del record
buildUpPlaySpeed	Velocità di gioco
buildUpPlaySpeedClass	Descretizzazione velocità di gioco
buildUpPlayDribbling	Dribbling
buildUpPlayDribblingClass	Descretizzazione dribbling
buildUpPlayPassing	Passaggio
buildUpPlayPassingClass	Descretizzazione passaggio
buildUpPlayPositioningClass	Posizionamento della squadra, definito o libero
chanceCreationPassing	Creazione chance di passaggio
chanceCreationPassingClass	Descretizzazione creazione chance di passaggio
chanceCreationCrossing	Creazione chance di cross
chanceCreationCrossingClass	Descretizzazione creazione chance di cross
chanceCreationShooting	Creazione chance di tiro in porta
chanceCreationShootingClass	Descretizzazione creazione chance di tiro in porta
chanceCreationPositioningClass	Creazione chance di posizionamento
defencePressure	Pressing della difesa
defencePressureClass	Descretizzazione pressing della difesa
defenceAggression	Aggressività della difesa
defenceAggressionClass	Descretizzazione aggressività della difesa
defenceTeamWidth	Copertura della difesa
defenceTeamWidthClass	Descretizzazione copertura della difesa
defenceDefenderLineClass	Tipologia di difesa

Tabella 4: Team

Campo	Descrizione
id	Id del record
team_api_id	Identificativo univoco della squadra
team_name	Nome completo della squadra
team_fifa_api_id	Id utile per reperire le informazioni da fifa
team_short_name	Nome abbreviato della squadra
team_long_name	Nome completo della squadra

3.4 Partite

La tabella *Match* contiene tutte le informazioni relative le squadre che hanno giocato l'una contro l'altra, i giocatori coinvolti e il loro posizionamento in campo. I giocatori vengono numerati da 1 a 11 e il portiere viene sempre collocato in posizione [1;1]. Sono presente inoltre dei valori per indicare le peculiarità della partita che per brevità verranno omessi. Nella tabella 5 è presenta la descrizione dei vari campi.

Tabella 5: Match

Campo	Descrizione
id	Id del record
country_id	Id della nazione di riferimento
league_id	Id della lega di riferimento
season	Stagione calcistica
date	Data della partita
match_api_id	Id univoco della partita
home_team_api_id	Id del team in casa
away_team_api_id	Id del team in trasferta
home_team_goal	Goal del team in casa
away_team_goal	Goal del team in trasferta
home_player_i	Id dell'i-esimo giocatore in casa
away_player_i	Id dell'i-esimo giocatore in trasferta
home_player_Xi	Posizione dell'i-esimo giocatore in casa, sul lato corto del campo
home_player_Yi	Posizione dell'i-esimo giocatore in casa, sul lato lungo del campo
away_player_Xi	Posizione dell'i-esimo giocatore in trasferta, sul lato corto del campo
away_player_Yi	Posizione dell'i-esimo giocatore in trasferta, sul lato lungo del campo

4 Applicazione web

Al fine di dimostrare l'utilizzo della rete in un applicativo software, viene qui illustrato lo sviluppo e il funzionamento di un'applicazione web che permette la configurazione dei team coinvolti in un match e la conseguente predizione del vincitore.

4.1 Architettura

L'interfaccia web è stata sviluppata utilizzando l'architettura a 3 layer, separando frontend, backend e database.

Il database utilizzato risulta quello fornito inizialmente senza alcuna modifica. Esso consiste quindi in un file sqlite interrogabile e modificabile semplicemente tramite un web server. Esso risulta particolarmente utile per fornire i dettagli dei giocatori ed eventualmente dei team così che l'utente possa visualizzarli e sceglierli opportunamente.

Per lo sviluppo del backend è stato deciso di utilizzare l'engine javascript tramite il popolare progetto Node.js [1]. Esso è in grado di agire come middleware tra il frontend e il database, separando al meglio le logiche di manipolazione del dato. È inoltre incaricato di chiamare adeguatamente lo script R per la predizione del vincitore della partita e per svolgere l'inferenza.

Il frontend risulta sviluppato utilizzando la libreria javascript React.js [2]

4.2 Interfaccia grafica

Di seguito vengono mostrate brevemente le principali schermate dell'applicazione e il loro funzionamento.

All'avvio dell'app, utilizzando semplicemente un browser web, è possibile notare la schermata principale con 3 tab di selezione: *Home team*, *Away team* e *Results*. Mentre i primi 2 permettono di modificare le squadre, il terzo consente invece di visualizzare i risultati data la configurazione precedente.

La schermata di configurazione della squadra mostra sulla sinistra l'immagine di un campo da calcio su cui è possibile spostare i giocatori, semplicemente trascinandoli col mouse e opportunamente aggiunti tramite il pulsante sulla destra, creando così la formazione da utilizzare.

Una volta finita la configurazione cliccando sul tab dei risultati è possibile vedere in verde la squadra vincitrice e in rosso quella che invece è stata sconfitta.

Per non dover ricreare configurazioni dall'inizio ogni qualvolta si avvii l'applicazione inoltre, nella parte superiore destra dello schermo sono presenti dei tasti per salvare e caricare di precedenti.

4.3 Predizione risultato

Per predire il vincitore della partita, come mostrato precedentemente, è stato necessario utilizzare la rete Bayesiana ottenuta a seguito degli esperimenti. Tale rete è stata infatti salvata su un file che è stato possibile poi ricaricare nell'ambiente R tramite l'apposito comando. Il web server javascript è dunque in grado di chiamare lo script R da riga di comando fornendogli in input una rappresentazione in formato JSON dei giocatori in campo con le loro caratteristiche. Lo script non deve far altro che manipolare i dati che gli sono stati forniti e applicare l'inferenza.

Per mostrare inoltre il puro funzionamento della rete è stato sviluppato un semplice form che permette l'inserimento delle evidenze e fornisce in output le distribuzioni di probabilità degli altri nodi.

5 Conclusioni

Riferimenti bibliografici

[1] Node.js.

[2] React.

[3] Sqlite.