



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	Applicazione web	4
3.1	Architettura	4
3.2	Interfaccia grafica	4
3.3	Predizione risultato	5
4	Conclusioni	6

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 aeree per classificare i ruoli dei giocatori in difesa, centrocampio 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 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.

3.1 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]

3.2 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.

3.3 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.

4 Conclusioni

Riferimenti bibliografici

[1] Node.js.

[2] React.