

**P** Ricerca | Università di Pisa | SoBigData

# La partita predittiva di Big data

L'analisi dei dati può fornire in anticipo informazioni su performance e infortuni

di **Paolo Cintia, Luca Pappalardo e Stefano Mastini**

● a finale di Champions League tra Juventus e Real Madrid: che cosa sarà cambiato in quel tragico secondo tempo? La cocente sconfitta del Brasile contro la Germania con un imprevedibile 7 a 1, all'ultimo mondiale. E, in tempi più recenti, la leggendaria vittoria del Leicester City in Premier League: una squadra che a inizio anno non era neanche accreditata per un posto tra le prime 5, riesce a sopraffare i blasoni di Manchester, Liverpool, Chelsea e via dicendo. Come è potuto accadere tutto ciò?

Proprio il "mistero" che si cela dietro la cavalcata trionfale del Leicester City nella Premier 2015/2016 ha incuriosito non solo i semplici appassionati, ma anche un gruppo internazionale di ricercatori, per la precisione di data scientist, scienziati dei dati. Alla prossima conferenza internazionale sul Knowledge Discovery and Data mining (KDD) verrà presentata un'analisi approfondita della stagione del Leicester City di Claudio Ranieri, partita per partita, azione per azione, avversario per avversario. Un dataset composto da centinaia di gigabyte con le posizioni di ciascun giocatore in ogni partita, la sua attività in campo (passaggi, tiri, ecc.) e molto altro. La forza del Lei-

cester, spiegano i ricercatori, può essere compresa attraverso il concetto di expected goal, una misura che stima la probabilità che un tiro finisca in rete considerando l'azione che l'ha generato, la situazione del tiratore, la distanza e l'angolo rispetto alla porta. Sotto la luce dei dati, il "miracolo" diventa facilmente comprensibile e risiede in due semplici quantità, la differenza tra goal fatti ed expected goal generati e quella tra goal subiti ed expected goal concessi all'avversario. Quest'ultima, per il Leicester, è stata la più bassa di sempre. O meglio, da quando il calcio è entrato nell'era dei dati.

Pur con ritardori rispetto ad altri sport, da circa un lustro il calcio ha cominciato a produrre numeri in quantità da capogiro. Lo scorso anno la Fifa ha finalmente autorizzato i calciatori a indossare i sensori biometrici utilizzati in allenamento (Gps, accelerometri, cardiofrequenzimetri) anche durante le partite ufficiali. Una vera e propria rivoluzione. Dalla mole di dati prodotti dalle partite, il mondo della ricerca scientifica sta ottenendo risultati via via più interessanti, in ambiti che vanno dalla simulazione di strategie di gioco alla predizione degli infortuni. Un infortunio è un costo notevole per una società ed uno stress fisico e psicologico per il calciatore: riuscire a prevederlo in tempo significa risparmiare cifre importanti e soprattutto comprendere qualcosa in più sui metodi di allenamento. Un esperimento che abbiamo condotto all'Università di Pisa, in collaborazione con il Barcelona Football Club, ha recentemente dimostrato che, con l'aiuto dell'intelligenza artificiale, gli infortuni dei calciatori possono essere previsti accuratamente e, quindi, evitati. Simulare l'impatto di una strategia di gioco è un altro problema matematico di non poco conto. Undici giocatori, indi-

pendenti, che si muovono "contro" altri undici giocatori: le combinazioni possibili hanno un numero di zeri notevole. Con l'analisi dei dati si può ridurre un po' l'incertezza, osservando ad esempio i movimenti tipici del singolo giocatore, o della singola squadra. Si può cominciare a capire le azioni e i comportamenti tipici delle squadre vincenti, così da studiarle e riprodurle.

Combinando i Big Data e l'intelligenza artificiale possiamo aprire scenari interessanti e prima impensabili, come prevedere il futuro valore di mercato dei calciatori o i trasferimenti tra le squadre di calcio, oppure sviluppare algoritmi per valutare in modo "data-driven" la performance di un calciatore durante una partita. Un altro studio che stiamo sviluppando al Cnr in collaborazione con la Northeastern University di Boston, mira a comprendere le relazioni tra il giudizio umano e la performance di un calciatore, o, in altre parole, come vengono assegnati i voti del fantacalcio. Quali sono le giocate che più influenzano i giornalisti che assegnano i voti? È possibile creare un giornalista robot in grado di assegnare voti ai giocatori in modo indistinguibile rispetto ad un giornalista umano?

Queste e molte altre attività di ricerca sullo sport sono possibili grazie al progetto europeo SoBigData, guidato da Università di Pisa e Cnr. Il progetto SoBigData prevede borse di studio per il primo corso di dottorato in Data Science, con scadenza delle domande di iscrizione ad agosto. Un'occasione per attrarre giovani ricercatori intenzionati a svelare "misteri", non solo calcistici, scandagliando il mondo dei Big Data.

© RIPRODUZIONE RISERVATA

