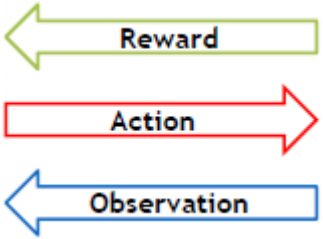
Fase di attacco di una squadra di calcio: Approximated Dynamic Programming per il Decision Making

Mauro Samarelli
834196

Introduzione

AGENT



ENVIRONMENT

Obiettivo

- Conoscenza dell'**expected reward** di un'azione effettuata in una zona del campo
- Conoscenza del **valore della funzione di stato** relativa ad una zona del campo
- Conoscenza del **valore della funzione di stato-azione** relativa ad una zona del campo per ogni azione
- Analisi dell'**apprendimento per rinforzo** svolto durante un match
- Definizione dell'**optimal path** per raggiungere il target partendo da una zona del campo

Premessa

Lo studio tralascia i seguenti fattori che incidono sull'esito di un match:

- Strategie difensive della squadra avversaria
  (ma considerabili con studio pregresso)
- Condizione fisico-atletica dei calciatori
- Condizione psicologica dei calciatori
- Fattori esterni
  (condizioni metereologiche, condizioni del campo da gioco, influenza del pubblico o altro)

Oggetto di studio

Match di Qualificazione a Euro 2020

3

0

L'evento si presta bene allo studio della **fase di attacco della Spagna tramite apprendimento per rinforzo**, in quanto contrappone il suo stile di gioco fatto da una serie numerosa di passaggi (azioni elementari da una zona all'altra del campo) allo stile di gioco della Svezia, estremamente difensivo.

Operazioni preliminari

- Divisione concettuale in zone del campo da gioco

- Tagging delle azioni elementari tramite un software di video-analysis

LONGOMATCH
BY FLUENDO

Definizione di stato

S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, TARGET}

Il sistema squadra si trova nello **stato s** quando ha il possesso della palla nella relativa zona di campo.

Ogni zona può essere uno **stato di partenza** (possesso recuperato) o **uno stato di arrivo** (possesso perso), ad eccezione dello *stato TARGET* che può essere solo uno stato finale.

Definizione di azione

Il sistema squadra effettua un'**azione a** per transitare da uno stato ad un'altro, quando il calciatore in possesso della palla decide di effettuare un gesto tecnico per far muovere la palla da una zona ad un'altra del campo.

A = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, TARGET}

Ogni a ∈ A rappresenta anche lo **stato di arrivo s'** dell'**azione a** partendo dallo **stato s**.

transizioni possibili = |S – {TARGET}| x |S| = 18 x 19 = 342

Peso di un’azione

TIPOLOGIA	ESITO	
	Positivo	Negativo
Passaggio orizzontale	0,5	-2
Passaggio in avanti	1	-1
Retropassaggio	0,25	-4
Azione chiave (passaggio filtrante, dribbling)	2	-0,5
Assist o tentativo	5	-0,25
Cambio di gioco sulla fascia	1,25	-1
Conduzione della palla	1	-1,25
Conclusione in porta	2	NA
Occasione da gol creata	8	NA
Rigore conquistato	10	NA
Gol segnato	10	NA

Expected Reward

Osservando uno specifico periodo del match è possibile calcolare il **valore atteso del reward** che il sistema squadra ottiene effettuando un'**azione a** nello **stato s** e **transitando in s'**, definito da a.

E[R|s, a] = score medio azioni positive * prob azione positiva
+ score medio azioni negative * prob azione negativa

$\forall s \in S - \{TARGET\}$ e $\forall a \in A$

- $> 0 \rightarrow$ **REWARD**
- $< 0 \rightarrow$ **PUNISHMENT**

Probabilità di transizione

Osservando uno specifico periodo del match è possibile calcolare la probabilità di transizione da uno **stato s** ad uno **stato s'**, tramite un'**azione a**.

P(s'|s, a) = # transizioni da s a s' / # transizioni da s

$\forall s \in S - \{TARGET\}\ e\ \forall a \in A$

Fase di Exploration

La fase di **Exploration** coincide con il periodo del match che va dal suo inizio al 15’ (scelta progettuale), in cui il sistema squadra attua la politica **“act-to-observe”** per avere informazioni sugli expected reward (inizialmente nulli) → in gergo calcistico “studio dell’avversario”

Value Iteration

Start of match: $V(s) \leftarrow 0 \; \forall \; s \in S$

checkpoint $\leftarrow 1$

Repeat

for all $s \in S - \{TARGET\}$

for all $a \in A$

value of state-action function

expected reward

prob of transition

value of state function

$$Q(s, a) \leftarrow E[r \mid s, a] + \gamma \sum_{s' \in S} P(s' \mid s, a) * V(s')$$

discount factor (=1)

$V(s) \leftarrow \max_a Q(s, a)$

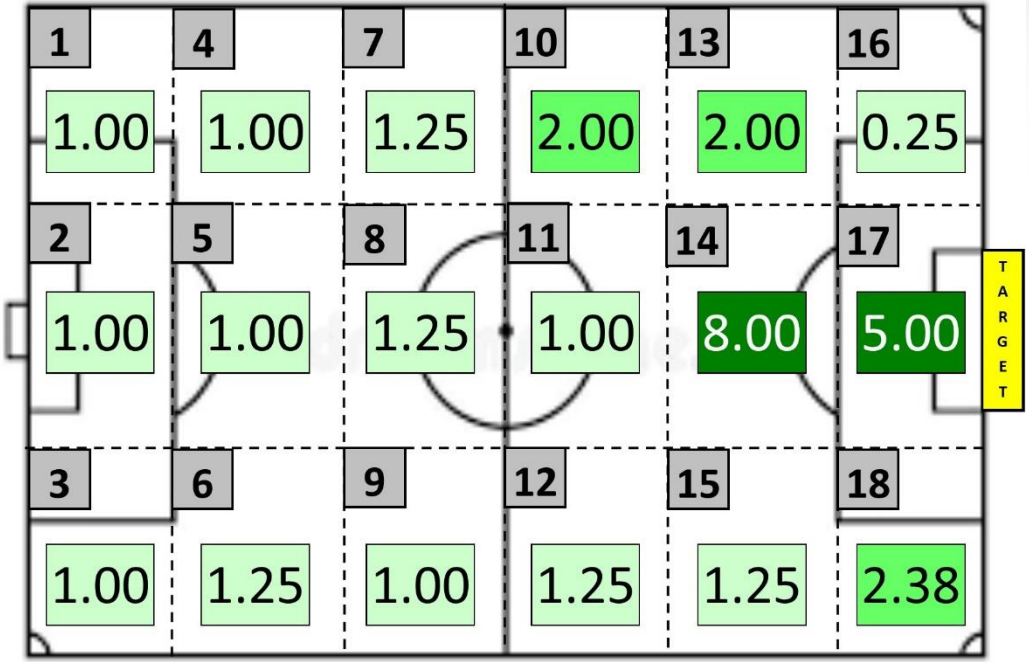
$V(TARGET) \leftarrow \max V(s)$

checkpoint $\leftarrow$ checkpoint + 1

Until Time(checkpoint) = End of match

NB: la fase di Exploration termina a checkpoint = 1 (primo step)

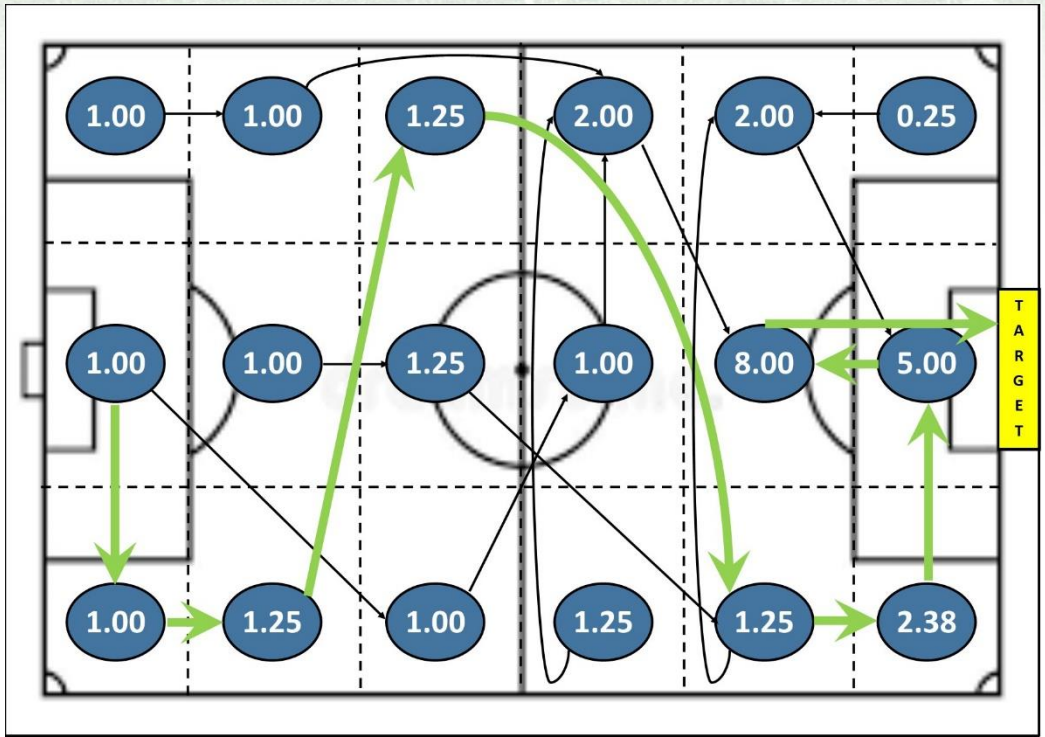
Exploration: V(s)



Rappresentazione grafica su campo da calcio diviso in zone dei valori della funzione di stato relativi al checkpoint 1 (15' → fine Exploration)

V(s) indica «quanto è buono avere il possesso di palla in una specifica zona del campo», quindi si tratta di un **indicatore quantitativo statico**

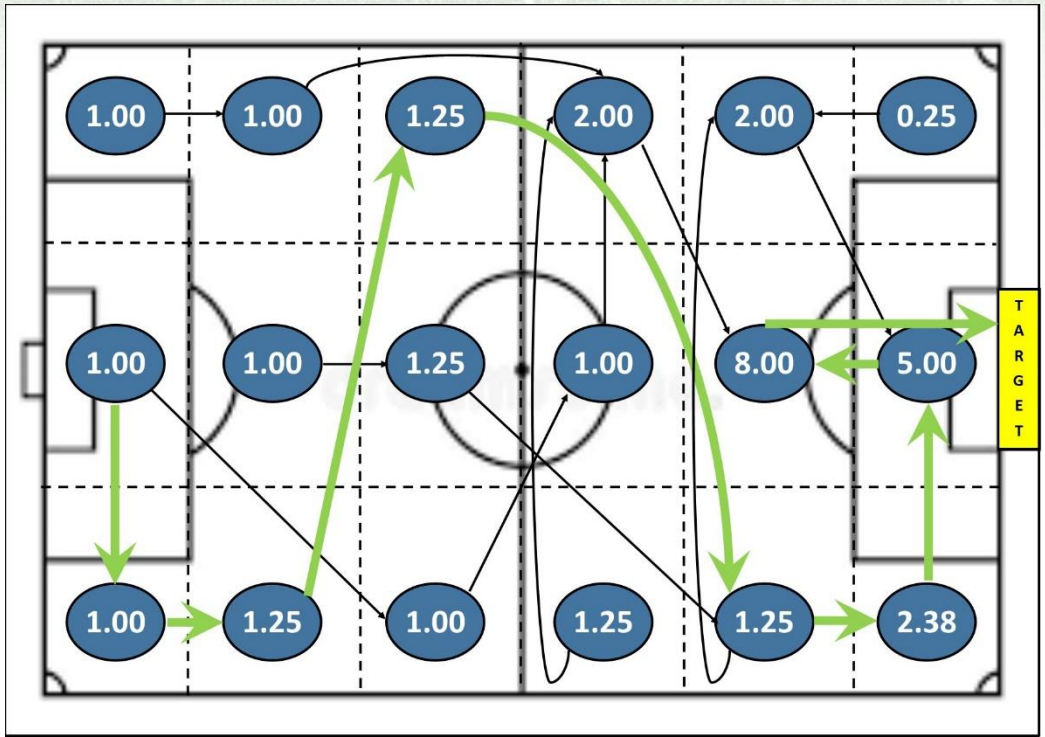
Exploration: Q(s, a)



Rappresentazione grafica su campo da calcio diviso in zone dei valori della funzione di stato-azione relativi al checkpoint 1 (15' → fine Exploration)

Q(s, a) indica «quanto è buono avere il possesso di palla in una specifica zona del campo e decidere di eseguire un'azione a per portare la palla nello stato s', definito da a», quindi si tratta di un **indicatore quantitativo-qualitativo dinamico**

Exploration: Q(s, a)



indica la policy ottimale $\pi^*(s)$ per cui $V^{\pi^*}(s) = \max_{a \in A} Q^{\pi^*}(s, a)$

indica una transizione all'interno dell'**optimal path** ottenuto con processo backward che **massimizza Q(s,a) per ogni transizione**, collegando un'ipotetica rimessa del portiere alla conclusione verso il target e vietando di tornare in zone già considerate

Fase di Exploitation

La fase di **Exploitation** coincide con il periodo del match che va dal 15’ alla sua fine (scelta progettuale), in cui il sistema squadra attua la politica **“act-observe-improve to get target”**, ovvero aggiorna il suo apprendimento dopo ogni azione di attacco con lo scopo di migliorare l’efficacia di quelle future e arrivare al target (segnare un gol)

Q - Learning

End of exploration: Q ← Q[Exploration]

checkpoint ← 2

Repeat

for all s ∈ S – {TARGET}

for all a ∈ A

NEW value of state-action function

OLD value of state-action function

learning rate [ck-1; ck] = # transition from s_t with a_t / # transition from s_t

expected reward from ck-1 to ck

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha(s_t, a_t) * [E[r_{t+1}] + \gamma * \max a_{t+1} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$$

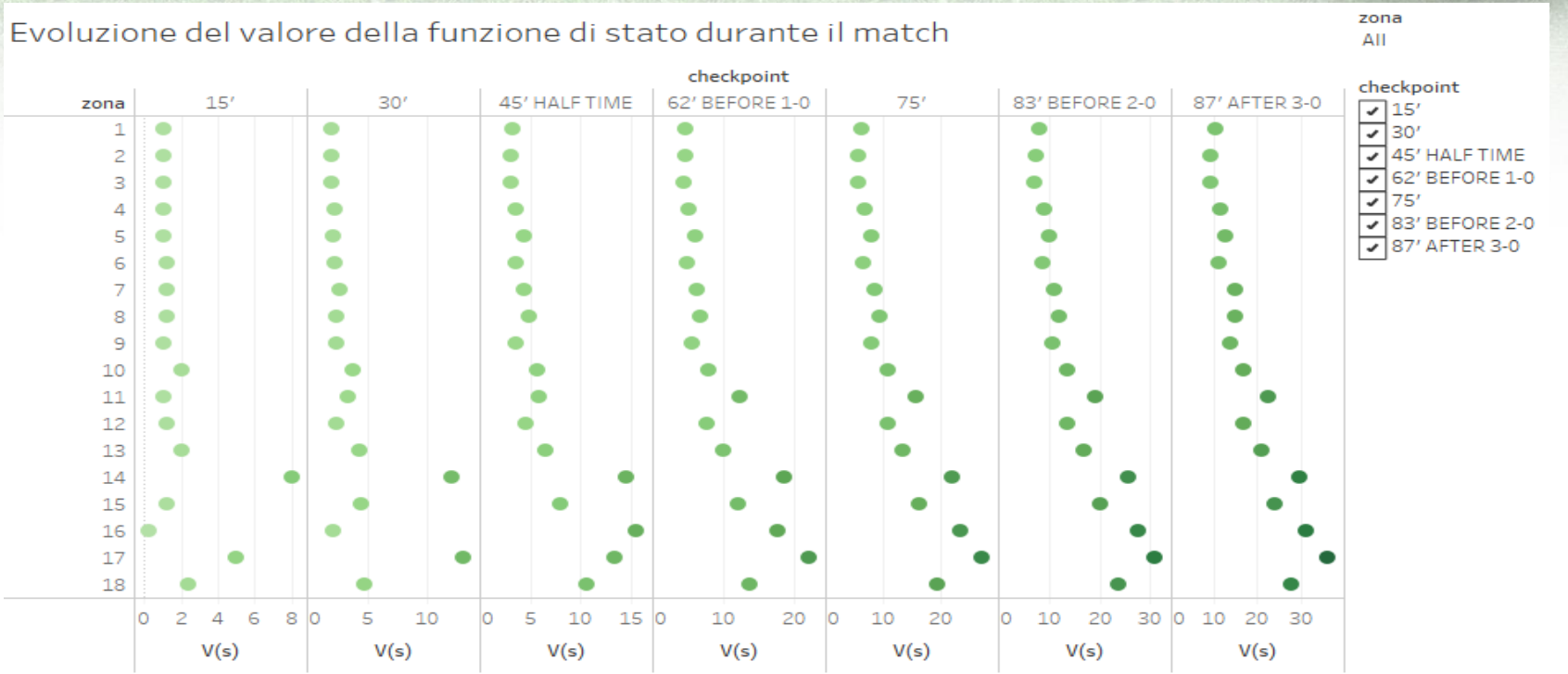
max value of state-action function from ck-1 to ck

checkpoint ← checkpoint + 1

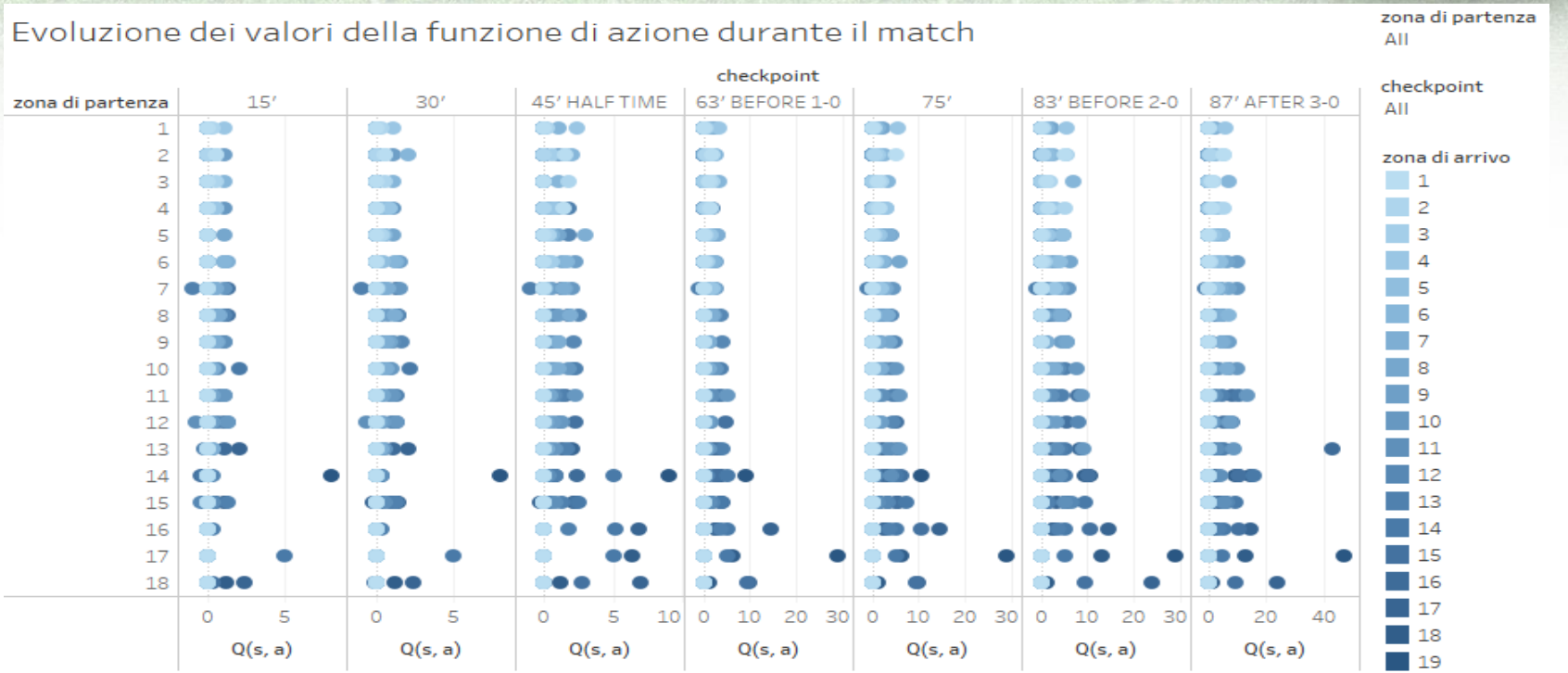
Until Time(checkpoint) = End of match

NB: per il calcolo di $Q(s_{t+1}, a_{t+1})$ vengono utilizzati i valori di V(s) del checkpoint precedente ottenuti con il Value Iteration

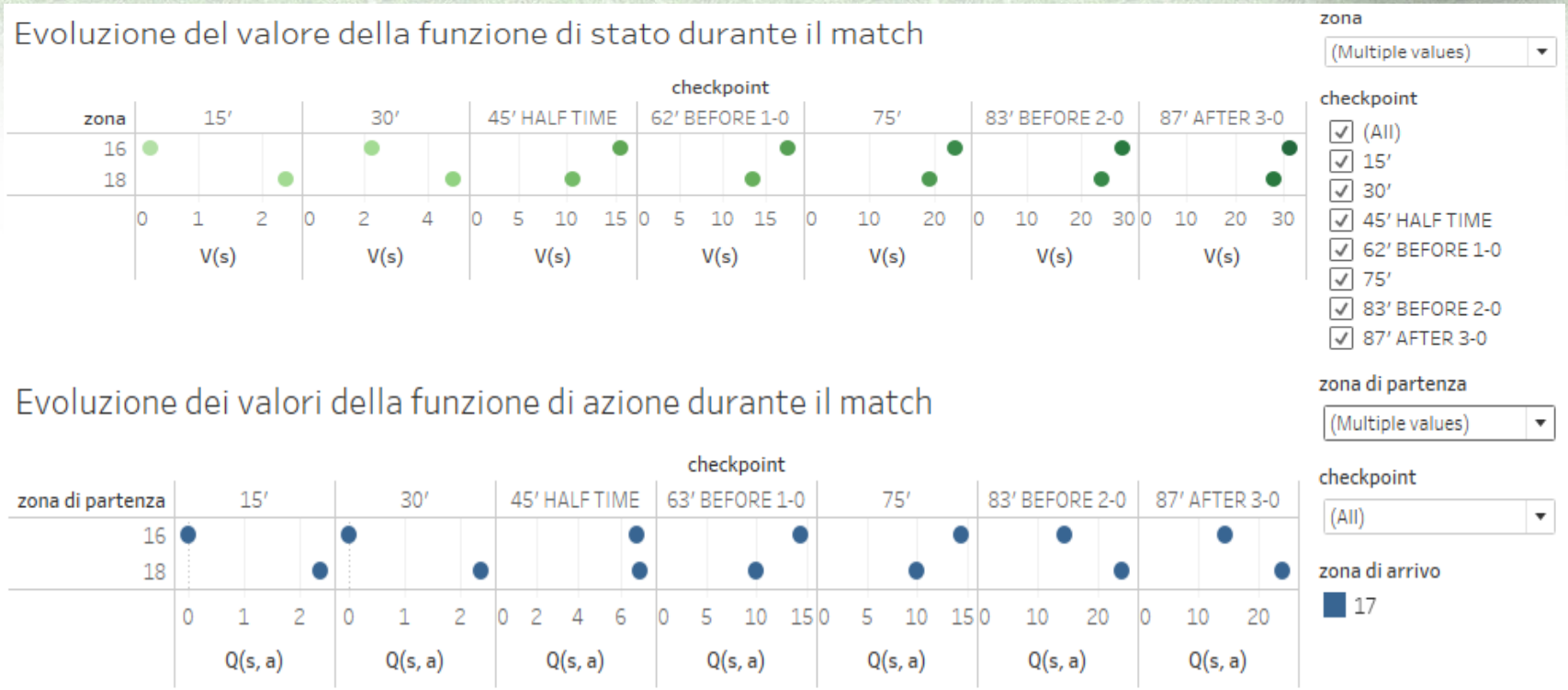
Evoluzione di V(s)



Evoluzione di Q(s, a)



Discussione



Conclusioni

PUNTI DI FORZA:

1) Patrimonio di informazioni significativo e di valore per la preparazione di un match
2) Analisi approfondita dei motivi che hanno portato al verificarsi di eventi decisivi per il risultato di un match, i quali solitamente vengono classificati come *«episodi casuali che cambiano la partita»*, senza tener conto di ciò che è avvenuto fino a quel momento

LIMITI:

1) Operazione di tagging delle azioni elementari molto onerosa
2) Trasmissione comunicativa del valore dell'algoritmo allo staff tecnico, tramite linguaggio semplice, comprensibile e con gergo calcistico (in caso contrario il rischio di non interesse è molto elevato)
3) Progettazione di un'architettura complessa per ricavare le informazioni in tempo reale durante un match

GRAZIE PER L’ATTENZIONE