

Titolo tesi: A neuro-computational model for reward-based motor learning

Candidato : Gianmarco Ragonetti

Tutors: Prof. Angelo Marcelli, Prof. Andrea Viggiano

La tesi seguente tratta i modelli computazionali del sistema nervoso impiegati nell'apprendimento motorio per rinforzo. Il contributo originale di questo lavoro consiste in una metodologia di confronto dei risultati di esperimenti per valutare i tassi di apprendimento (*learning rates*) umani con i risultati provenienti da un modello computazionale derivato da un'analisi approfondita della letteratura.

Reward e *punishers* sono stimoli particolari in grado di indirizzare al meglio o al peggio la prestazione dell'azione da imparare. Ciò avviene perché possono rafforzare o indebolire le connessioni tra una combinazione di stimoli sensoriali in ingresso e una combinazione di attivazioni motorie in uscita, attribuendo loro un valore.

Un *reward/punisher* può provenire da esigenze innate (fame, sete, ecc.) segnalate da strutture cablate nel cervello (*ipotalamo*), ma potrebbe anche provenire da uno stimolo inizialmente neutrale (proveniente da corteccia o stimoli sensoriali) che acquisisce la capacità di produrre valore dopo l'apprendimento (ad esempio il valore dei soldi, l'approvazione). Abbiamo chiamato i primi *primary values*, mentre i secondi *learned values*. L'efficacia di uno stimolo come reinforcer/punisher dipende dal contesto specifico in cui l'azione ha luogo (*motivazione*).

Si afferma che i valori innati guidano l'apprendimento attraverso attivazioni della *dopamina* e che i valori appresi acquisiscono questa abilità dopo accoppiamenti ripetuti con valori primari innati, in un paradigma di condizionamento classico (*classic conditioning*).

In base ad alcune ipotesi, proponiamo un modello computazionale composto da:

- Un blocco che rappresenta la corteccia cerebrale (*Actor*) e consiste in un *mapping* di combinazioni sensoriali (*corteccia posteriore*) e possibili azioni (*corteccia motoria*). I pesi della rete corrispondono alla probabilità di un movimento, data una combinazione sensoriale in ingresso. Le ricompense/punizioni alterano queste probabilità che attraverso una regola di selezione che è stata implementata nei *Basal Ganglia* per la selezione delle azioni;
- Un blocco per la produzione di valori (*Critic*): abbiamo valutato due diversi scenari. Nel primo abbiamo considerato solo il blocco per i valori innati, costituito da *VTA (Ventral Tegmental Area)*, *Lateral Hypothalamus* (ricompense innate) e *Lateral Habenula* (punizioni innate)

Nel secondo scenario abbiamo aggiunto le strutture per l'apprendimento di valori, l'*Amygdala*, che impara a produrre un'attivazione della dopamina all'insorgenza di uno stimolo inizialmente neutro e il *Ventral Striatum*, che impara a prevedere il verificarsi della ricompensa innata, annullando l'attivazione della dopamina.

La ricompensa innata è fondamentale per il sistema di apprendimento dei valori: anche in un sistema ben addestrato, se lo stimolo che fa da reward appreso non è più in grado di prevedere un premio innato di stimolo (perché è in ritardo o in assenza), e se ciò accade frequentemente potrebbe perdere capacità di rinforzo/indebolimento. Questo fenomeno è chiamato *estinzione dell'acquisizione* e dipende rigorosamente dal contesto (*motivating operation*).

La validazione del modello è iniziata con Emergent, che fornisce un modello biologicamente accurato delle reti e dei meccanismi di apprendimento dei neuroni, è stato effettuato un porting in Matlab, più versatile, per dimostrare la capacità del sistema di apprendere per un compito specifico.

In questo semplice compito, il sistema deve imparare tra due possibili azioni, dato un gruppo di stimoli di cardinalità crescente: 2, 4 e 8.

Abbiamo valutato l'attività nei due scenari descritti, uno con ricompense innate e uno con ricompense apprese.

Infine, sono stati condotti diversi esperimenti per valutare il tasso di apprendimento umano: i volontari dovevano imparare a premere i pulsanti della tastiera corretti quando gli stimoli visivi apparivano sul monitor, per ottenere una ricompensa uditiva e visiva.

Gli esperimenti sono stati accuratamente progettati in modo tale da rendere comparabili il risultato della semplice rete neurale simulata con quelle degli utenti umani. La strategia è stata quella di selezionare un insieme ridotto di risposte e una serie di stimoli visivi il più semplice possibile (*edges*), evitando così il problema di una rappresentazione gerarchica complessa dell'informazione, collassandole in un unico livello. Sui risultati è stato quindi effettuato il *fitting* con funzioni esponenziale e iperbolica. Entrambi i modelli mostrano che il tasso di apprendimento umano è lento rispetto alla rete artificiale e diminuisce con il numero di stimoli da imparare.