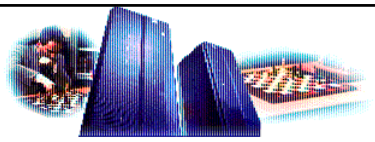
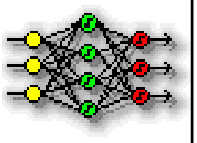




Sistemi Intelligenti
Learning: l'apprendimento degli agenti

Alberto Borghese

Università degli Studi di Milano
Laboratorio di Sistemi Intelligenti Applicati (AIS-Lab)
Dipartimento di Scienze dell'Informazione
borghese@dsi.unimi.it

Riassunto

- Gli agenti
- Il Reinforcement Learning
- Gli elementi del RL
- Un esempio: tris

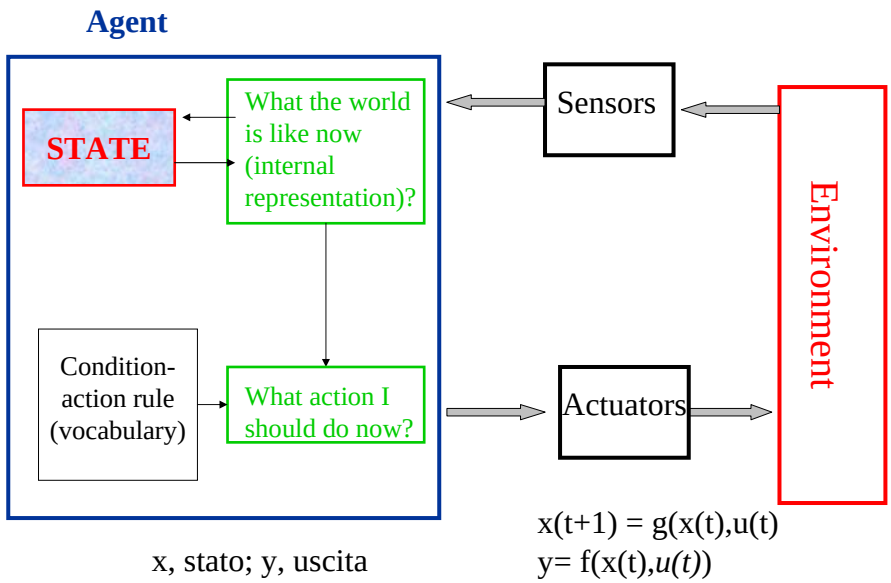
Agente

- Può scegliere un'azione sull'ambiente tra un insieme continuo o discreto.

- L'azione dipende dalla situazione. La situazione è riassunta nello stato del sistema.

- L'agente monitora continuamente l'ambiente (input) e modifica continuamente lo stato.

Schematic diagram of an agent



x, stato; y, uscita

$x(t+1) = g(x(t),u(t))$

$y = f(x(t),u(t))$

L'agente

- Inizialmente l'attenzione era concentrata sulla progettazione dei sistemi di "controllo". Valutazione, sintesi…
- L'intelligenza artificiale e la "computational intelligence" hanno consentito di spostare l'attenzione sull'apprendimento delle strategie di controllo.
- Macchine dotate di meccanismi (algoritmi, SW), per apprendere.

I vari tipi di apprendimento

Supervisionato (learning with a teacher). Viene specificato per ogni pattern di input, il pattern desiderato in input.

Non-supervisionato (learning without a teacher). I neuroni verranno associati a pattern di ingresso contigui. Clustering. Mappe neurali.

Apprendimento con rinforzo (reinforcement learning, learning with a distal teacher). L'ambiente fornisce un'informazione del tipo success or fail.

Apprendimento supervisionato

$$\min_{\{w\}} J(.) \qquad J = \left\| Y^D - g(W^{nuovo}; U) \right\|$$

Y^D è l'uscita desiderata nota.

- Si tratta di un problema di minimizzazione di una cifra di merito (J) sullo spazio di parametri W, che caratterizzano l'agente.

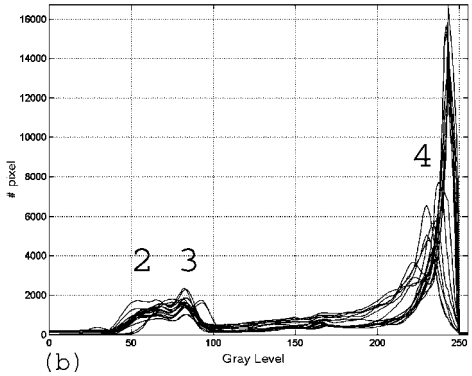
Soluzione iterativa:

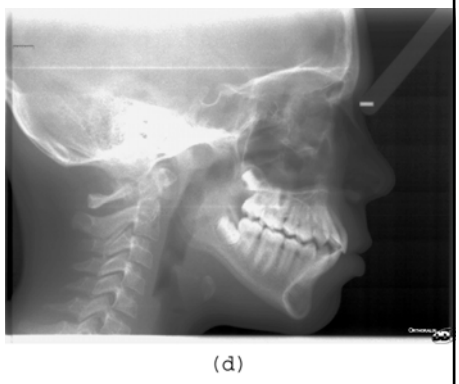
Obbiettivo: se esiste una soluzione, trovare ΔW in modo iterativo tale che l'insieme dei pesi W^{nuovo} ottenuto come:

$$W^{nuovo} = W^{vecchio} + \Delta W$$

dia luogo a un errore sulle uscite di norma minore che con $W^{vecchio}$

Esempio di apprendimento supervisionato





Y = g(esposizione, distanza…..; W).

Y = somma di Gaussiane (mixture model).

Apprendimento non-supervisionato: Classificazione

Descrizione numerica dell'oggetto:

altezza, colore, forma, posizione, ….

SPAZIO DEI CAMPIONI / DELLE CARATTERISTICHE

Classificatore

Classificazione dell'oggetto:

(classe A, classe B, …)

SPAZIO DELLE CLASSI

946

A cosa serve la classificazione?

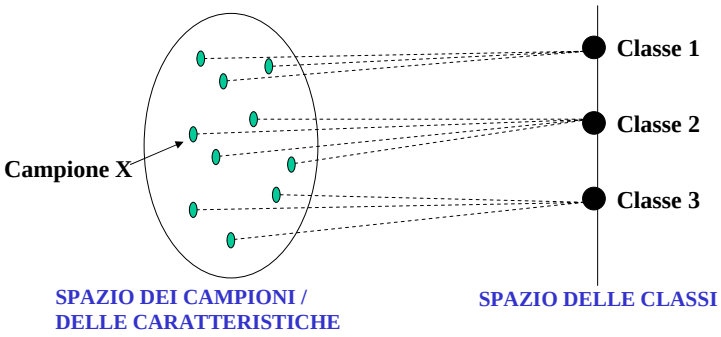
- Compressione dati (telecomunicazioni, immagini, …);
- segmentazione (bio)immagini;
- riconoscimento automatico;
- controllo robot;
- pattern recognition;
- ricostruzione superfici;
- …

1046

Classificazione

Un'interpretazione geometrica:

Mappatura dello spazio dei campioni nello spazio delle classi.



1146

Riassunto

- Gli agenti
- Il Reinforcement Learning
- Gli elementi del RL
- Un esempio: tris

Reinforcement learning

Nell'apprendimento supervisionato, esiste un "teacher" che dice al sistema quale è l'uscita corretta (learning with a teacher). Non sempre e' possibile.

Spesso si ha a disposizione solamente un'informazione numerica (a volte binaria, giusto/sbagliato successo/fallimento), puntuale.

Questa è un'informazione qualitativa**.**

L'informazione disponibile si chiama ***segnale di rinforzo.*** *Non dà alcuna informazione su come aggiornare i pesi. Non è possibile definire una funzione costo o un gradiente.*

Obbiettivo: creare degli agenti "intelligenti" che abbiano una "machinery" per apprendere dalla loro esperienza.

Reinforcement Learning: caratteristiche

- Apprendimento mediante interazione con l'ambiente. Un agente isolato non apprende.
- L'apprendimento è funzione del raggiungimento di uno o più obbiettivi.
- La ricompensa viene data puntualmente ad ogni istante di tempo.
- Le azioni vengono valutate con la ricompensa a lungo termine (**delayed reward**). Il meccanismo di ricerca delle azioni migliori è imparentato con la ricerca euristica: **trial-and-error**.
- L'agente sente l'input, modifica lo stato e genera un'azione che massimizza la ricompensa a lungo termine.

Exploration vs Exploitation

Esplorazione (exploration) dello spazio delle azioni per scoprire le azioni migliori. Un agente che esplora solamente raramente troverà una buona soluzione.

Le azioni migliori vengono scelte ripetutamente (exploitation) perchè garantiscono ricompensa (reward). Se un agente non esplora nuove soluzioni potrebbe venire surclassato da nuovi agenti più dinamici.

Occorre non interrompere l'esplorazione.

Occorre un approccio statistico per valutare le bontà delle azioni.

Exploration ed exploitation vanno bilanciate. Come?

Dove agisce un agente?

- L'agente ha un comportamento goal-directed ma agisce in un ambiente incerto.
- Esempio: planning del movimento di un robot.
- Un agente impara interagendo con l'ambiente. Planning può essere sviluppato mentre si impara a conoscere l'ambiente (mediante le misure operate dall'agente stesso). La strategia è vicina al trial-and-error.

Relazione con l'AI

- Gli agenti hanno dei goal da soddisfare. Approccio derivato dall'AI.
- Nell'apprendimento con rinforzo vengono utilizzati strumenti che derivano da aree diverse dall'AI:
  - Ricerca operativa.
  - Teoria del controllo.
  - Statistica.
- L'agente impara facendo. Deve selezionare i comportamente che ripetutamente risultano favorevoli.

Esempi

Un giocatore di scacchi. Per ogni mossa ha informazione sulle configurazioni di pezzi che può creare e sulle possibili contro-mosse dell'avversario.

Una gazzella in 6 ore impara ad alzarsi e correre a 40km/h.

Come fa un robot veramente autonomo ad imparare a muoversi in una stanza per uscirne?

Come impostare i parametri di una raffineria (pressione petrolio, portata….) in tempo reale, in modo da ottenre il massimo rendimento o la massima qualità?

Caratteristiche degli esempi

Parole chiave:

- Interazione con l'ambiente. L'agente impara dalla propria esperienza.
- Obbiettivo dell'agente.
- Incertezza o conoscenza parziale dell'ambiente.

Osservazioni:

- Le azioni modificano lo stato (la situazione), cambiano le possibilità di scelta in futuro (delayed reward).
- L'effetto di un'azione non si può prevedere completamente.
- L'agente ha a disposzione una valutazione globale del suo comportamento. Deve sfruttare questa informazione per migliorare le sue scelte. Le scelte migliorano con l'esperienza.
- I problemi possono avere orizzonte temporale finito od infinito.

Riassunto

- Gli agenti
- Il Reinforcement Learning
- Gli elementi del RL
- Un esempio: tris

I tue tipi di rinforzo

L'agente deve scoprire quale azione (policy) fornisca la ricompensa massima provando le varie azioni (trial-and-error).

"Learning is an adaptive change of behavior and that is indeed the reason of its existence in animals and man (K. Lorentz, 1977).

Rinforzo puntuale istante per istante, azione per azione (condizionamento classico).

Rinforzo puntuale "una-tantum" (condizionamento operante).

Il Condizionamento classico

L'agente deve imparare una (o più) trasformazione tra input e output. Queste trasformazioni forniscono un comportamento che l'ambiente premia.

Il segnale di rinforzo è sempre lo stesso per ogni coppia input – output.

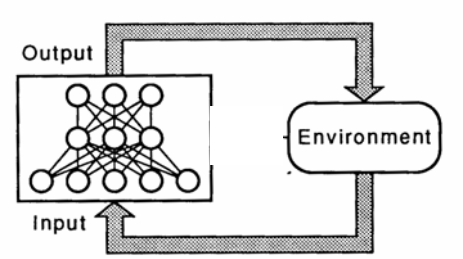
Esempio: risposte riflesse Pavolviane. Campanello (stimolo condizionante) prelude al cibo. Questo induce una risposta (salivazione). La risposta riflessa ad uno stimolo viene evocata da uno stimolo condizionante.

Stimolo-Risposta. Lo stimolo condizionante (campanello = input) induce la salivazione (uscita) in risposta al campanello.

Condizionamento operante

Reinforcement learning (operante).

Interessa un comportamento. Una sequenza di input / output che può essere modificata agendo sui parametri che definiscono il comportamento dell'agente. Il condizionamento arriva in un certo istante di tempo (spesso una-tantum) e deve valutare tutta la sequenza temporale di azioni, anche quelle precedenti nel tempo.



Gli attori del RL

Policy. Descrive l'azione scelta dall'agente: mapping tra input (stato ambiente) e azioni. Funzione di controllo. Le policy possono avere una componente stocastica. Viene utilizzato un modello adeguato del comportamento dell'agente.

Reward function. Ricompensa **immediata**. Associata all'azione intrapresa in un certo stato. Può essere data al raggiungimento di un goal (esempio: successo / fallimento). E' uno scalare associato allo stato dell'agente. Rinforzo primario.

Value function. "Cost-to-go". Ricompensa a **lungo termine**. Somma dei reward + costi associati alle azioni scelte istante per istante. Orizzonte temporale ampio. Rinforzo secondario.

- Quale delle due è più difficile da ottenere?
- L'agente agisce per massimizzare la funzione Value o Reward?

L'ambiente

Model of the environment. E' uno sviluppo relativamente recente. Da valutazione implicita dello svolgersi delle azioni future (trial-and-error) a valutazione esplicita mediante modello dell'ambiente della sequenza di azioni e stati futuri (planning).

Incorporazione di AI:

- Planning (pianificazione delle azioni).
- Viene rinforzato il modulo di pianificazione dell'agente.

-Incorporazione della conoscenza dell'ambiente:

-- Modellazione dell'ambiente (non noto o parzialmente noto).

Proprietà del rinforzo

L'ambiente o l'interazione può essere complessa.

Il rinforzo può avvenire solo dopo una più o meno lunga sequenza di azioni (delayed reward).

E.g. agente = giocatore di scacchi.
ambiente = avversario.

Problemi collegati:
temporal credit assignement.
structural credit assignement.

L'apprendimento non è più da esempi, ma dall'osservazione del proprio comportamento nell'ambiente.

Riassunto

•Reinforcement learning. L'agente viene modificato, rinforzando le azioni che sono risultate buone a lungo termine. E' quindi una classe di algoritmi iterativi.

•Self-discovery of a successful strategy (it does not need to be optimal!). La strategia (di movimento, di gioco) non è data a-priori ma viene appresa attraverso **trial-and-error**.

•Credit assignement (temporal and structural).

•Come possiamo procedere in modo efficiente nello scoprire una strategia di successo? Cosa vuol dire modificare l'agente?

Riassunto

- Gli agenti
- Il Reinforcement Learning
- Gli elementi del RL
- Un esempio: tris