

Agents and Learning

Alberto Borghese

Università degli Studi di Milano
Laboratorio di Sistemi Intelligenti Applicati (AIS-Lab)
Dipartimento di Informatica
alberto.borghese@unimi.it



A.A. 2018-2019

1/27

<http://borghese.di.unimi.it>



Riassunto



- **Gli agenti**
- L'apprendimento nelle macchine

A.A. 2018-2019

2/27

<http://borghese.di.unimi.it>



L'agente



- E' un'entità in grado di eseguire delle azioni sull'ambiente che lo circonda in funzione di quanto percepito attraverso dei sensori.
- 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 e più in generale di comportamento.
- **Macchine dotate di meccanismi (algoritmi, SW), per apprendere.**



Why agents are important?

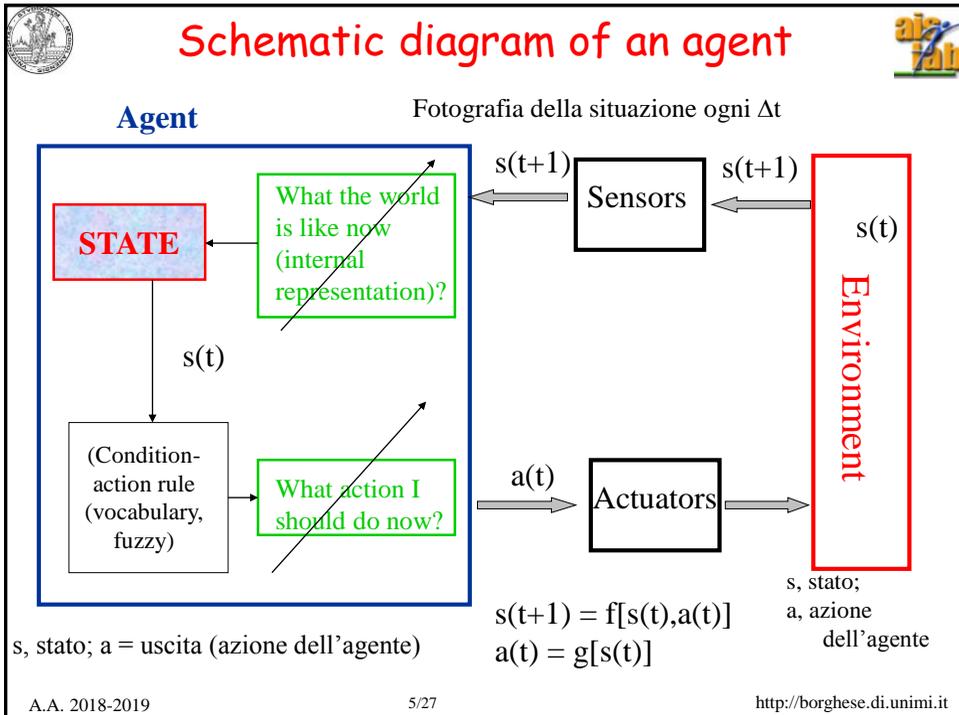


Agente (software): essere software che svolge servizi per conto di un altro programma, solitamente in modo automatico ed invisibile. Tali software vengono anche detti agenti intelligenti

“They are seen as a natural metaphor for conceptualising and building a wide range of complex computer systems (the world contains many passive objects, but it also contains very many *active* components as well);

They cut across a wide range of different technology and application areas, including telecoms, human-computer interfaces, distributed systems, WEB and so on;

They are seen as a natural development in the search for ever-more powerful abstractions with which to build computer systems.“



How agents solve a problem

Formulate a problem. Through analysis. State, action, identification.

Solve the problem (by searching).

Implement the solution (execute).

Evaluate the implemented solution.

- ◆ Success or fail? Adequate or not adequate?
- ◆ How much adequate? How to measure the success or failure of the performance?
- ◆ Optimization of the performance to create better agents.

- Solve a problem = achieve a given goal (= reach a final state or avoid certain states)
- An agent can examine different sequences of actions (deterministic or stochastic response by the environment) and search the best sequence.

A.A. 2018-2019 6/27 <http://borghese.di.unimi.it>



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 (stato); l'ambiente modifica continuamente lo stato.
- La scelta dell'azione è non banale e richiede un certo grado di "intelligenza".
- L'agente ha una memoria "intelligente". Non può tenere in memoria tutto quanto successo nel passato.



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?



Ambiente



- L'agente ha un comportamento goal-directed ma agisce in un **ambiente incerto** non noto a priori o parzialmente noto.
- Esempio: planning del movimento di un robot.
- **Un agente impara interagendo con l'ambiente.** Spesso l'ambiente non è noto o è solo parzialmente noto (o non è modellizzabile facilmente).
- **Credit Assignment.**



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? (cf. competizione Robocare@home).

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

Come imparare a mantenere verticale il pendolo rovesciato mediante una forza che agisce sul vagone sul quale il pendolo è incernierato?



Apprendimento incrementale



$$\begin{aligned}x(t+1) &= f[x(t), a(t)] && \text{Ambiente} \\ a(t) &= \mathbf{g}[x(t)] && \text{Agente}\end{aligned}$$

Ad ogni passo viene corretta la funzione $g(\cdot)$ per migliorare le prestazioni.

2 modalità:

- Per trial: ad ogni passo viene analizzato un valore di $x(t)$ ed eventualmente di $a(t)$ e calcolata una variazione dei parametri.
- Per epoche: ad ogni passo vengono analizzati tutti i dati e da questi calcolata una variazione dei parametri.



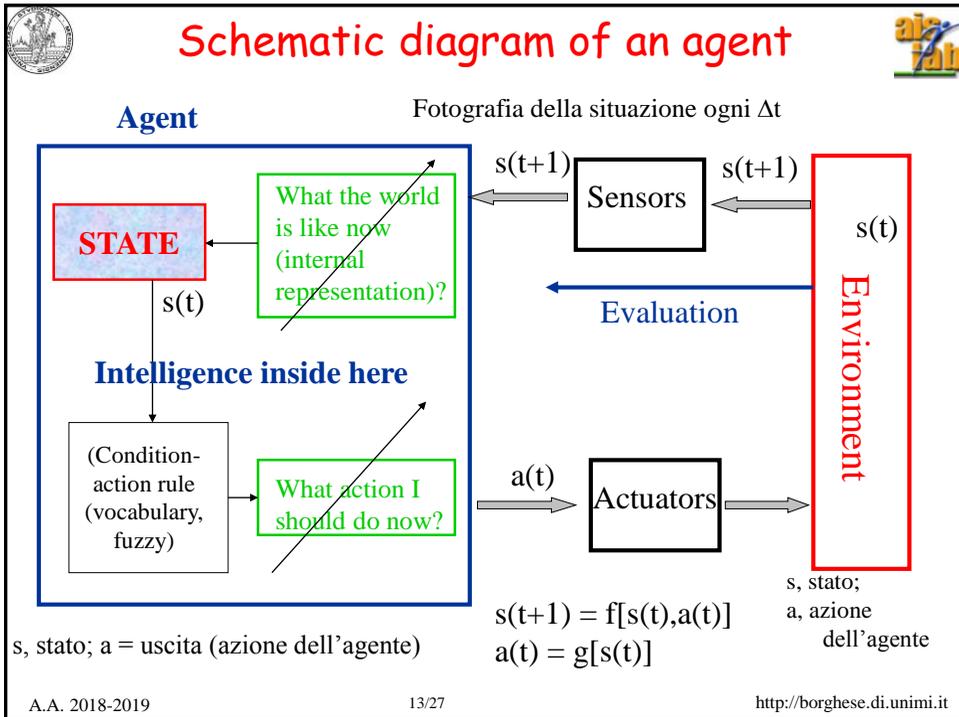
Complessità di $g(\cdot)$



La funzione $g(\cdot)$ viene tipicamente scritta come $g(x | \mathbf{w})$. La forma di $g(\cdot)$ viene data dai parametri che possono variare. $g(\cdot)$ è plastica, non viene fornita in partenza, ma viene modificata con l'interazione con l'ambiente.

Maggiore è il numero di parametri, maggiori sono i gradi di libertà e quindi la variabilità della funzione $g(\cdot)$.

Troppi parametri possono portare a **over-fitting** dei dati.





Come apprende una macchina



$$a(t) = \mathbf{g}[x(t)]$$

La funzione $g(\cdot)$ non viene fissata all'inizio, ma l'agente è dotato di meccanismi che possono modificare la forma di $g(\cdot)$ con l'interazione con l'ambiente.

Espliciteremo questo scrivendo: $g(x(t) | w)$, con w parametri, ovvero quantità che possono variare e dare una forma opportuna alla funzione $g(\cdot)$.

L'apprendimento ha quindi come obiettivo i parametri w , che vengono fatti variare in modo che $g(\cdot)$ rappresenti il comportamento desiderato o ottimale.



I vari tipi di apprendimento Apprendimento non supervisionato



$$\begin{array}{ll} x(t+1) = f[x(t), a(t)] & \text{Ambiente} \\ a(t) = \mathbf{g}[x(t)] & \text{Agente} \end{array}$$

Non-supervisionato (learning without a teacher). Estrazione di similitudine statistiche tra pattern di input.

Clustering = raggruppamento

Viene analizzato solo l'input: $x(t)$ e trovate similitudini tra i diversi elementi.



I vari tipi di apprendimento

Apprendimento supervisionato



$$\begin{array}{ll} x(t+1) = f[x(t), a(t)] & \text{Ambiente} \\ a(t) = \mathbf{g}[x(t)] & \text{Agente} \end{array}$$

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

Viene confrontato il valore prodotto dall'agente, $a(s_t)$, con $a_{\text{des}}(s_t)$ desiderato, la differenza viene utilizzata per correggere la funzione $g(\cdot)$.

a) Regressione: a partire da un insieme di dati costruisco un modello (funzione $g(\cdot)$) che predica in modo robusto $a(s_t)$ per valori di s_t non ancora esaminati.

b) Classificazione: la funzione $g(\cdot)$ s_t a una tra un certo numero di classi predefinite.



Ruolo del teacher



$$\begin{array}{ll} x(t+1) = f[x(t), a(t)] & \text{Ambiente} \\ a(t) = \mathbf{g}[x(t)] & \text{Agente} \end{array}$$

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

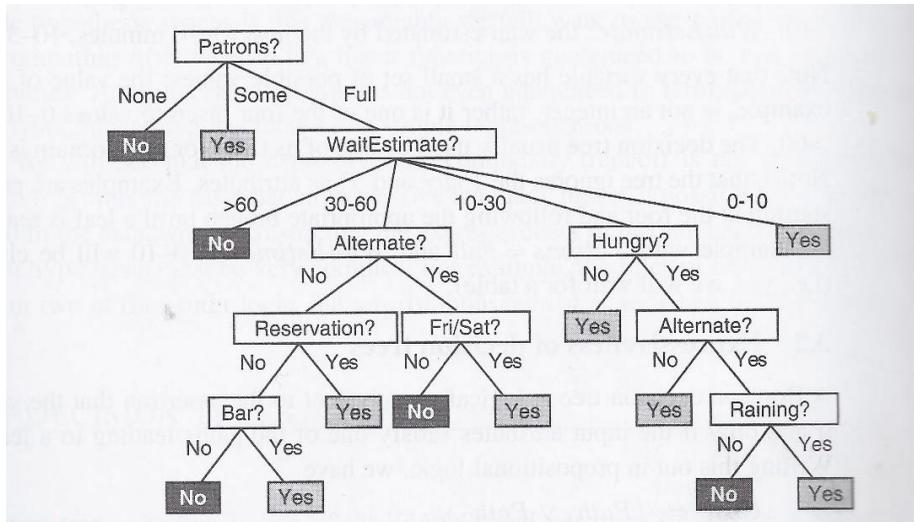
Il teacher fornisce degli esempi: coppie di valori: $\{x_t, a_{\text{des}}(x_t)\}$.

x_t e a_t possono essere conosciuti con un'approssimazione (errore, conoscenza incompleta, distribuzione stocastica...).

Il teacher non conosce tutte le possibili coppie: $\{x_t, a(x_t)\}$.



Albero di decisione per classificazione



Algoritmo per scegliere se fermarsi in un ristorante o meno

A.A. 2018-2019

19/27

<http://borghese.di.unimi.it>



I dati misurati sul campo



Examp	Input attributes										Goal (wait)
	Alt	Bar	Fri	Hun	Pat	Price	Rain	Res	Type	Est	
x ₁	Yes	No	No	Yes	Some	\$\$\$	No	Yes	Fr	0-10	y ₁ = Y
x ₂	Yes	No	No	Yes	Full	\$	No	No	Thai	30-60	y ₂ = N
x ₃	No	Yes	No	No	Some	\$	No	No	Burg	0-10	y ₃ = Y
x ₄	Yes	No	Yes	Yes	Full	\$	Yes	No	Thai	10-30	y ₄ = Y
x ₅	Yes	No	Yes	No	Full	\$\$\$	No	Yes	Fr	>60	y ₅ = N
x ₆	No	Yes	No	Yes	Some	\$\$	Yes	Yes	I	0-10	y ₆ = Y
x ₇	No	Yes	No	No	None	\$	Yes	No	Burg	0-10	y ₇ = N
x ₈	No	No	No	Yes	Some	\$\$	Yes	Yes	Thai	0-10	y ₈ = Y
x ₉	No	Yes	Yes	No	Full	\$	Yes	No	Burg	>60	y ₉ = N
x ₁₀	Yes	Yes	Yes	Yes	Full	\$\$\$	No	Yes	I	10-30	y ₁₀ = N
x ₁₁	No	No	No	No	None	\$	No	No	Thai	0-10	y ₁₁ = N
x ₁₂	Yes	Yes	Yes	Yes	Full	\$	No	No	Burg	30-60	y ₁₂ = Y

A.A. 2018-2019

20/27

<http://borghese.di.unimi.it>



How to classify the error introduced by a model?



Is the model good enough?

Does it have enough parameters?

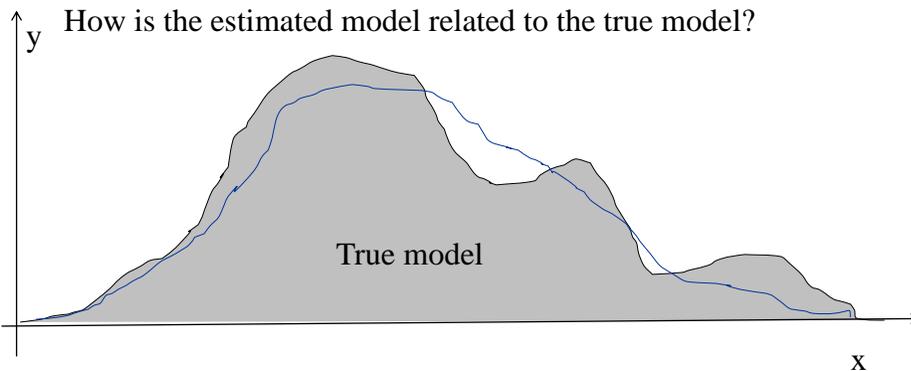
Does it cover the input domain (in all dimensions)?

This is not enough to obtain a good model!!

The model should be properly tuned to the data



How to classify the error introduced by a model?



Bias and variability trade-off

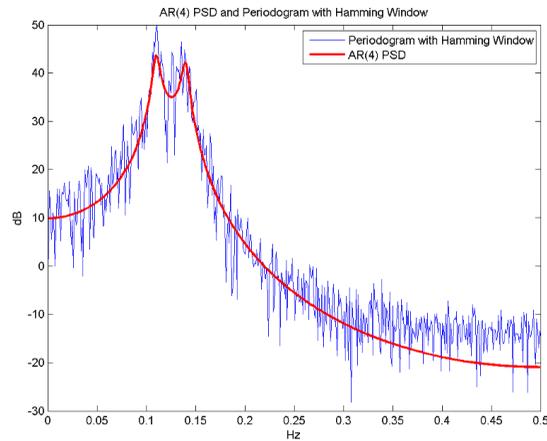
Bias is the distance of the model curve from the true unknown curve.
It is associated to model error.



Variability



How are the measured points related to the estimated model?



Given $P_{mes}(x_{mes}, y_{mes})$ and $y = f(x)$, the error is measured as: $dist(y_{mes}, f(x_{mes}))$, for instance Euclidean distance. It is associated to measurement error.

If variability goes to zero, bias increases and overfitting arises.

In a good model, variability tends to the statistics of the measurement noise.



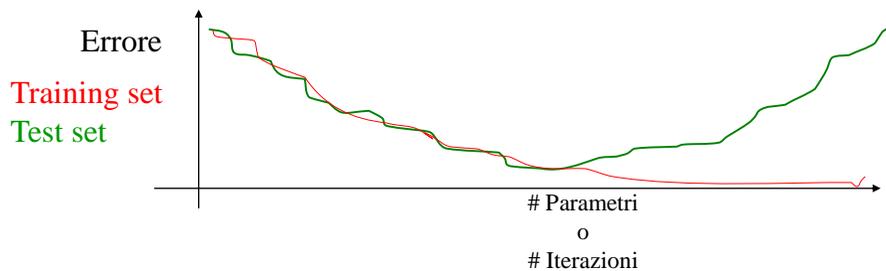
Scelta empirica - cross-validation



Cross-Validation - Errore sull'insieme di training = Errore sull'insieme di test.

Si vuole evitare che il modello si specializzi troppo sui pattern di training e non sia in grado di interpolare correttamente.

*Il numero di parametri viene aumentato fino a quando **entrambi** gli errori diminuiscono.*





I vari tipi di apprendimento

Apprendimento con rinforzo



$$x(t+1) = f[x(t), a(t)] \quad \text{Ambiente}$$
$$a(t) = g[x(t)] \quad \text{Agente}$$

Apprendimento con rinforzo (reinforcement learning, learning with a critic, learning with a **distal teacher**).

L'ambiente fornisce un'informazione puntuale, **di tipo qualitativo**, sull'interazione: ad esempio success or fail.

E' l'agente che trasforma questa informazione qualitativa in un segnale di valutazione interno e da questo in un segnale di «errore» per correggere la funzione $g(\cdot)$.

Segnali di rinforzo: caduta, vinta/persa una partita, successo in un'interazione....



Learning without a teacher



L'agente osserva l'ambiente e non sono previste azioni / goal.

La situazione dell'ambiente viene analizzata e situazioni simili vengono raggruppate in una stessa categoria.

Clustering.



Riassunto



- Gli agenti
- L'apprendimento nelle macchine