

ACQUISIZIONE DI NUOVE CATEGORIE IN RETI NEURALI ATTRAVERSO SIMBOLI "FONDATI"

A. Greco (1), T. Riga (1), A. Cangelosi (2)

(1) DISA, Università di Genova; (2) University of Plymouth, UK

Introduzione

Uno dei maggiori problemi incontrati dalla metodologia simulativa in psicologia è quello del *symbol grounding*, cioè di rendere il significato dei simboli effettivamente intrinseco al sistema e non dipendente da una interpretazione esterna ad esso. Questo problema può essere risolto usando la metodologia connessionista. Esistono diversi modelli che sono in grado di estrarre le caratteristiche invarianti degli stimoli presentati e ad associarvi etichette "linguistiche" (Plunkett, Sinha, Moller & Strandsby, 1992; Cangelosi, Greco & Harnad, 2000). Tuttavia, avere una collezione di nomi associati a certe categorie concettuali non è condizione sufficiente per poter sviluppare un linguaggio ed esprimere significati più complessi attraverso la combinazione di simboli in proposizioni. I simboli che non fossero *fondati* su precedenti rappresentazioni di origine sensomotoria sarebbero vuoti, come avverrebbe nel caso in cui si volesse comprendere il significato di un termine di una lingua sconosciuta consultando un dizionario scritto solo in quella lingua: la definizione conterrebbe altri simboli bisognosi di definizione e così via in un circolo vizioso all'infinito (Harnad, 1990).

Lo scopo del modello qui presentato è di studiare attraverso reti neurali in che modo sia possibile conferire un fondamento alle descrizioni simboliche di stimoli mai incontrati da un sistema, attraverso l'uso di simboli già fondati in precedenza. In sostanza viene studiata la possibilità di trasferire il fondamento (*grounding transfer*) da un simbolo all'altro. Le simulazioni proposte sono un'estensione di un precedente modello (Cangelosi, Greco & Harnad, 2000).

Metodo

Stimoli e procedura. Dieci reti neurali sono state addestrate a categorizzare e denominare forme animali (es. cavallo, tartaruga) e loro caratteristiche (es. strisce, puntini), costituite da immagini di 50 X 50 pixel proiettate sui campi recettivi di una retina costituita da 7 X 7 unità. L'addestramento è stato svolto in tre fasi. Nella prima fase (*categorizzazione*), le reti hanno appreso a classificare 216 stimoli visivi (4 forme animali e 4 caratteristiche di texture, ciascuna in 27 varianti) nelle 8 categorie corrispondenti. Nella seconda fase (*denominazione delle categorie di base*), le reti sono state addestrate ad associare gli stessi stimoli retinici con i nomi delle loro categorie. In queste prime due fasi le reti hanno appreso sulla base dell'esperienza diretta per tentativi ed errori, supervisionata attraverso un feedback correttivo (algoritmo di *backpropagation*). Nella terza fase (*apprendimento di alto livello*), sono state presentate alle reti soltanto stringhe di simboli, costituite dalla combinazione di nomi di forme animali e di loro caratteristiche (ad es. cavallo, strisce; zebra), che definivano 4 categorie di alto livello.

Le reti sono state infine testate presentando altri 180 stimoli retinici contenenti effettivamente le combinazioni di forme e caratteristiche definite nella terza fase. Il risultato atteso era che, anche se si trattava di stimoli nuovi per le reti, se il *grounding transfer* si fosse realizzato, le reti sarebbero state in grado di categorizzarli correttamente.

Architettura delle reti. Sono state usate dieci reti feed-forward costituite da tre strati (input, output e intermedio), i cui pesi iniziali erano casuali. Lo strato di input era suddiviso in due gruppi di unità: la retina di 7 X 7 campi recettivi, dunque costituita di 49 unità; 12 unità linguistiche corrispondenti alle 8 categorie di base (4 forme animali e 4 caratteristiche di texture) e alle 4 categorie di alto livello. Lo strato nascosto, costituito da 6 unità, era connesso a quello di input in modo che le prime tre unità nascoste ricevessero input dalla periferia della retina e le altre tre unità dalla parte centrale della retina stessa. Lo strato di output, come quello di input, era suddiviso in due gruppi di unità: il primo conteneva le 8 unità che rappresentavano l'output sensorimotorio e il secondo conteneva 12 unità linguistiche strutturate in maniera simile a quelle di input. Lo strato di output era connesso con quello nascosto in modo da ottenere, grazie alla particolare connettività della retina, una struttura modulare, in cui tre unità si specializzano nel rappresentare forme animali e le altre tre nel rappresentare loro caratteristiche.

Risultati

Tutte e dieci le reti hanno appreso i tre compiti con successo. L'errore medio per la fase di categorizzazione è stato 0.07, per la denominazione di base 0.15 e per l'apprendimento di alto livello 0.04. Per quanto riguarda la fase di test, è stata calcolata la frequenza di risposte corrette alle immagini corrispondenti agli stimoli di alto livello sia per l'output sensorimotorio che linguistico. Tutte e dieci le reti hanno categorizzato e denominato correttamente gli stimoli del test, dimostrando che il fondamento realizzato per i simboli al livello di base si era trasferito con successo alla categorie di alto livello.

Allo scopo di controllare se questo risultato potesse dipendere da qualche altra variabile non controllata piuttosto che dal *grounding transfer*, abbiamo ripetuto le fasi di training delle reti escludendo la fase di denominazione di base, facendo seguire l'apprendimento di alto livello subito dopo la prima categorizzazione. In questa simulazione di controllo nessuna delle 10 reti ha fornito risultati corretti.

Conclusioni

Questi risultati sostengono l'approccio al problema del *symbol grounding* basato su modelli interamente connessionisti. Nel nostro modello la stessa rete elabora sia il fondamento sensorimotorio dei simboli, sia la generazione di nuove categorie attraverso la combinazione di simboli. L'organizzazione modulare delle unità nascoste suggerisce che è importante che il fondamento sensorimotorio sia separato per le diverse caratteristiche alla base della classificazione. Stiamo lavorando allo sviluppo del modello in varie direzioni, esplorando l'uso di stimoli e caratteristiche diversi, di gerarchie di categorie più complesse e di nuove architetture.

Riferimenti bibliografici

Cangelosi A., Greco A., & Harnad S. (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science*, **12**(2), 143-162.

Cangelosi A. & Harnad S. (in press). The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*

Harnad S. (1990). The Symbol Grounding Problem. *Physica D*, **42**, 335-346

Plunkett K., Sinha C., Møller M.F., & Strandsby O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science*, **4**, 293-312.