

INTELLIGENZA ARTIFICIALE

Il pc che capisce le emozioni

L'Università di Rochester ha ideato un software che analizza la voce. Così il deep learning sfrutta le potenzialità dei computer

di Marco Passarello

Tra le funzioni più accattivanti dei nuovi iPhone c'è Siri, l'assistente vocale in grado di conversare quasi come se fosse una persona. È un "quasi" ancora molto pesante: basta poco per rendersi conto che Siri non capisce davvero ciò che diciamo. Tuttavia l'effetto è sorprendente, ed è uno dei primi frutti di una tecnica dalla quale ci si aspettano importanti progressi nel campo dell'intelligenza artificiale: il *deep learning*.

Il *deep learning* («apprendimento profondo») deriva da studi cominciati già negli anni 80, anche se ha conosciuto grandi sviluppi solo di recente, da quando i processori grafici (Gpu) forniscono la potenza di calcolo parallelo necessaria. Si basa sulle reti neurali, circuiti costruiti in modo analogo alle sinapsi del cervello, che possono "apprendere" modificando le proprie connessioni fino al momento in cui a ogni ingresso corrisponde l'uscita desiderata.

Il *deep learning* presuppone che la materia da apprendere sia la somma di fattori concomitanti, sui quali la rete neurale viene messa in grado di effettuare generalizzazioni, estendendo ad altri fattori ciò che ha già appreso su alcuni. Inoltre i fattori appresi possono essere utilizzati per la comprensione di fattori di livello più elevato. Per esempio, per comprendere una lingua straniera la rete non tenta di apprendere subito il significato delle parole, ma identifica per prime le singole lettere, poi le mette a comporre delle sillabe, e solo in ultimo usa le sillabe apprese per identificare le parole. Le reti di questo tipo possono apprendere senza supervisione, a patto di fornire loro grandi quantità di dati e lasciare poi che compiano da sole classificazioni e generalizzazioni.

Un esperimento del genere è quello condotto nei laboratori di Google, dove a una rete neurale formata da 16mila processori è stato affidato il compito di classificare milioni di immagini casuali

provenienti da YouTube. Senza istruzioni esplicite, la rete è stata in grado di formarsi da sola il concetto di "gatto" e isolare le immagini contenenti gatti, con una precisione che, pur non elevata (15,8% di riconoscimenti corretti) è comunque molto superiore a quella di esperimenti precedenti.

La potenzialità è quella di espandere le capacità dei computer di comprendere la realtà, consentendo di affidargli compiti che finora richiedevano un operatore umano, come la traduzione simultanea o la sorveglianza. Per esempio, nel corso di un esperimento condotto nei laboratori di intelligenza artificiale dell'Università di Lugano, un software basato sul *deep learning* si è dimostrato capace di riconoscere segnali stradali con una precisione superiore a quella dei 32 esperti umani usati come termine di confronto. E un'equipe di studenti guidata dal professor Geoffrey E. Hinton dell'Università di Toronto ha recentemente vinto un premio sponsorizzato dalla Merck per un software in grado di riconoscere molecole farmacologicamente attive, grazie a un programma di *deep learning* che ha sbaragliato la concorrenza, nonostante l'insieme dei dati su cui basarsi fosse piuttosto ridotto.

Questo è bastato perché alcuni, tra cui il «New York Times», annunciasse l'arrivo di una nuova era per l'intelligenza artificiale, che sarebbe finalmente in grado di raggiungere i risultati promessi negli anni '80 e mai realizzati. Altri sono più prudenti: il professor Gary Marcus, psicologo e teorico del linguaggio,

dalle pagine del «New Yorker» ammonisce che l'intelligenza è la somma di molti tipi di apprendimento, e che il *deep learning* da solo non permetterà di raggiungerla, «non più di quanto costruire una scala più lunga permetta di raggiungere la luna».

Comunque, anche se la "vera" intelligenza dovesse essere ancora lontana, farsi capire a voce dai computer diventerà sempre più facile. In occasione del seminario Ieee sulla tecnologia del linguaggio parlato tenutosi a Miami il mese scorso, l'Università di Rochester (New York) ha presentato un programma in grado di percepire le emozioni di chi parla. La tecnica prescinde dalla comprensione del linguaggio, sfruttando solo fattori come l'altezza e il volume della voce per classificare l'umore del parlante come "felice", "triste", "spaventato", "disgustato" o "neutro". Il risultato è corretto nell'81% dei casi per una voce già nota. Ora si tenta di arrivare a indovinare anche l'umore di voci sconosciute.

L'obiettivo è un assistente vocale che adegui i suoi suggerimenti al nostro umore, senza che gli si dica nulla. Non sarà come i robot di «Star Wars», ma poco ci manca.

Le reti neurali apprendono da sole, a patto di fornire una grande quantità di dati

* PAROLA CHIAVE

Deep learning

Nell'ambito dell'apprendimento automatico, si definisce *deep learning* quell'apprendimento che si basa su diversi livelli di rappresentazione, corrispondenti a gerarchie di fattori o concetti, dove i concetti di livello inferiore possono essere usati per definire quelli di livello superiore. È una tecnica concepita decenni fa, ma che solo ora, grazie l'aumento della potenza di calcolo disponibile, sta raggiungendo grandi risultati nel campo dell'intelligenza artificiale.





Oltre il linguaggio. Il programma dell'Università di Rochester analizza l'altezza e il volume della voce per comprendere le emozioni (nella foto la locandina del film «Carnage», che si basa su personaggi con passioni contrastanti)