

Ufficio Stampa della Provincia autonoma di Trento

Piazza Dante 15, 38122 Trento

Tel. 0461 494614 - Fax 0461 494615

uff.stampa@provincia.tn.it

COMUNICATO n. 702 del 30/03/2017

Premio conferito alla 42esima conferenza internazionale per la ricerca nel settore del riconoscimento vocale e del “signal processing” ICASSP 2017 , che si è tenuta da 5 al 9 marzo a New Orleans, USA

Premiato ricercatore trentino alla più importante conferenza mondiale sul riconoscimento vocale

Una concorrenza globale e ai massimi livelli. Sono stati oltre 2.600 i lavori sottomessi - e di questi 1.317 accettati - alla 42esima conferenza internazionale per la ricerca nel settore del riconoscimento vocale e del “signal processing” ICASSP 2017, che si è tenuta da 5 al 9 marzo a New Orleans, USA.

Ad aggiudicarsi uno dei quattro premi in palio per i migliori lavori presentati dagli studenti PhD - l’“IBM Best Paper Student Award” - è stato il ricercatore **Mirco Ravanelli**, trentunenne trentino in forza al gruppo “SHINE” del Centro ICT della Fondazione Bruno Kessler di Trento.

Mirco Ravanelli lavora in FBK da oltre sei anni. “Sono entrato in Fondazione per la compilazione della mia tesi di laurea (*ndr.* Ingegneria delle telecomunicazioni)” - racconta. “Ora sono all’ultimo anno del mio dottorato e questo premio rappresenta un importante riconoscimento per me. L’esperienza dello scorso anno è stata determinante. Ho trascorso sei mesi a Montréal, in Canada, nel laboratorio di **Yoshua Bengio**, uno dei guru del “deep learning”. **Bengio** ha un H-index di 92 e per lavorare nel suo laboratorio c’è una richiesta altissima. Pensate che lo scorso anno ha ricevuto circa 700 domande, ma ne ha accettate solo 20. Io ero tra questi”. “Durante la mia **Mobility** (*ndr.* il programma di FBK che permette ai ricercatori e alle ricercatrici di trascorrere un periodo di studio e lavoro all’estero)” – conclude **Ravanelli** – “ho avuto la fortuna di lavorare in questo prestigioso laboratorio, dove ho sviluppato nuove metodologie di ricerca e collaborato con scienziati di fama mondiale”.

Lo studio di **Ravanelli** e dei colleghi premiato a New Orleans, che titola “A network of deep neural networks for distant speech recognition” (*Una rete di reti neurali per il riconoscimento vocale a distanza*), sebbene sviluppato nell’ambito del riconoscimento vocale, definisce un **nuovo paradigma** particolarmente flessibile, che può trovare applicazione in moltissimi altri settori, tra i quali la robotica, dove diversi sistemi devono cooperare nel migliore dei modi per raggiungere un unico obiettivo comune.

“E’ un risultato di massima eccellenza scientifica” – aggiunge **Maurizio Omologo**, responsabile dell’unità di ricerca **SHINE** (Speech-Acoustic scene Analysis and Interpretation). “L’articolo premiato ad **ICASSP 2017** descrive il lavoro svolto da Mirco l’anno scorso, durante il suo periodo di **Mobility** a Montréal, in buona parte basato su idee e strategie delineate fin dal *qualifying exam* al termine del primo anno di dottorato”.

Abstract dello studio

A network of deep neural networks for distant speech recognition - Una rete di reti neurali per il riconoscimento vocale a distanza)

Autori Mirco Ravanelli, Philemon Brakel, Maurizio Omologo, Yoshua Bengio

Costruire computer in grado di riconoscere il parlato rappresenta un passo fondamentale per lo sviluppo delle future interfacce uomo-macchina e più in generale per lo sviluppo di un'intelligenza artificiale. Per questi motivi, negli ultimi anni l'interesse verso questo tipo di tecnologie è cresciuto moltissimo, portando anche a numerose applicazioni commerciali come gli **assistenti vocali di Siri di Apple, Google Voice e Alexa di Amazon**. Tuttavia, la maggior parte dei sistemi attuali fornisce prestazioni soddisfacenti solamente in ambienti a bassa rumorosità e con parlatori che interagiscono vicino ad un microfono, rendendo di grande interesse lo studio di tecnologie più robuste in grado di funzionare anche a distanza dal parlatore, in ambienti complessi caratterizzati da rumori di vario tipo e riverbero.

Il lavoro presentato alla conferenza internazionale [ICASSP 2017 \(42th International Conference on Acoustics, Speech, and Signal Processing\)](#) approccia questo problema utilizzando **tecniche avanzate di deep learning**. Il *deep learning* è una tecnologia che sta drasticamente rivoluzionando il mondo dell'intelligenza artificiale e permette di sviluppare sistemi che, a partire dai cosiddetti "**big data**", sono in grado di risolvere, con una precisione mai raggiunta in passato, moltissimi problemi come, l'identificazione di particolari oggetti in immagini e video, la traduzione di un testo in un'altra lingua oppure, appunto, il riconoscimento del parlato.

Più precisamente il lavoro propone di cambiare in maniera significativa la modalità attraverso cui questa tecnologia viene impiegata negli attuali sistemi di riconoscimento del parlato a distanza. Anche i sistemi più evoluti, infatti, sono basati sulla cascata di diverse tecnologie che vengono tipicamente sviluppate in maniera indipendente, senza nessuna garanzia circa la loro effettiva compatibilità ed il loro adeguato livello di comunicazione e cooperazione.

Gli attuali sistemi potrebbero dunque essere paragonati ad una squadra di calcio che, anche se composta da ottimi giocatori singoli, non sarà mai in grado di vincere il campionato se questi giocatori non sono in grado di giocare bene assieme. La "rete di *deep neural networks*" proposta nello studio rivede dunque drasticamente l'architettura degli attuali sistemi, "rompendo" questa cascata inserendo tutte le tecnologie coinvolte nel riconoscimento vocale all'interno di una rete dove c'è piena comunicazione e cooperazione fra i vari elementi. L'attenzione della ricerca si è concentrata non solo sulla definizione di una nuova architettura, ma soprattutto sullo studio di innovative tecniche per addestrare in modo automatico questa rete di componenti. Il lavoro ha portato alla definizione di un nuovo algoritmo di addestramento (chiamato "*backpropagation through network*") in cui tutti gli elementi nella rete vengono addestrati in modo congiunto e progressivamente imparano in modo automatico ad interagire, comunicare e cooperare fra loro.

Il lavoro, sebbene sviluppato nell'ambito del riconoscimento vocale, definisce un paradigma particolarmente generale e flessibile, che può trovare applicazione in moltissimi altri settori applicativi, tra i quali la robotica, dove diversi sistemi devono cooperare nel migliore dei modi per raggiungere un unico obiettivo comune.

()