POLITECNICO
MILANO 1863

L'intelligenza artificiale diventa più sostenibile!

Il Politecnico di Milano sviluppa una nuova generazione di acceleratori di calcolo

Milano, 12 febbraio 2020 – Un gruppo di ricerca del **Politecnico di Milano** ha sviluppato un nuovo circuito di calcolo che permette di eseguire operazioni avanzate, tipiche delle reti neurali su cui si basa l'intelligenza artificiale, in una sola operazione.
I risultati di performance in termini di velocità e consumo di energia pongono le basi per una nuova generazione di acceleratori di intelligenza artificiale con maggiore efficienza energetica e migliore sostenibilità a livello globale. Lo studio è stato recentemente pubblicato sulla prestigiosa **Science Advances**.

Riconoscere un viso o un oggetto, oppure interpretare correttamente una parola o un motivo musicale sono operazioni oggi possibili sui più comuni gadget elettronici, come un normale smartphone, grazie all'intelligenza artificiale. Perché avvengano, le reti neurali necessitano però di un opportuno addestramento (training) così energeticamente oneroso che, secondo alcuni studi, la carbon footprint dell'addestramento di una complessa rete neurale può eguagliare il consumo di 5 automobili in tutto il loro arco vitale.

Per ridurre i tempi e i consumi del training, si è reso necessario sviluppare circuiti radicalmente diversi dall'approccio convenzionale e che mappano più fedelmente la struttura delle reti neurali e le caratteristiche delle sinapsi biologiche. Un tipico esempio è il concetto di computing in memoria, dove i dati vengono elaborati direttamente all'interno della memoria, esattamente come nel cervello umano.

A partire da questa intuizione, i ricercatori del Politecnico hanno sviluppato un nuovo circuito, che riesce ad eseguire una funzione matematica, nota come regressione, in una sola operazione. Per questo scopo è stata utilizzata una memoria resistiva, anche nota con il nome di memristor, che riesce a memorizzare un dato qualsiasi (come ad esempio il valore di un'azione in un certo istante) nel valore della sua resistenza. Disponendo questi elementi di memoria in una matrice di dimensioni di pochi micron (milionesimi di metro), il gruppo del Politecnico di Milano è riuscito ad eseguire una regressione lineare su un gruppo di dati. Questa operazione è capace di determinare la retta che meglio descrive una sequenza di dati, permettendo ad esempio di prevedere l'andamento della borsa sulla base di un semplice modello lineare. È stata anche dimostrata la regressione logistica, che permette di classificare un dato all'interno di un database. Questa funzione è fondamentale nei cosiddetti sistemi di

Ufficio Relazioni con i Media
Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano

T +39 02 2399 2229
C. +39 366 6211436
relazionimedia@polimi.it
www.polimi.it

raccomandazione, che sono uno strumento di marketing fondamentale per gli acquisti sul web.

Lo studio “One-step regression and classification with cross-point resistive memory arrays (DOI: 10.1126/sciadv.aay2378): https://advances.sciencemag.org/content/6/5/eaay2378).

Artificial intelligence is becoming sustainable!
The Politecnico di Milano develops a new generation of computing accelerators

Milan, February 12, 2020 - A research group from **Politecnico di Milano** has developed a new computing circuit that can execute advanced operations, typical of neural networks for artificial intelligence, in one single operation.
The circuit performance in terms of speed and energy consumption paves the way for a new generation of artificial intelligence computing accelerators that are more energy efficient and more sustainable on a global scale. The study has been recently published in the prestigious **Science Advances**.

Recognizing a face or an object, or correctly interpreting a word or a musical tune are operations that are today possible on the most common electronic gadgets, such as smartphones and tablets, thanks to artificial intelligence. For this to happen, complicated neural networks needs to be appropriately trained, which is so energetically demanding that, according to some studies, the carbon footprint that derives from the training of a complex neural network can equal the emission of 5 cars throughout their whole life cycle.

To reduce the time and energy consumption of the training, one should develop circuits that are radically different from the conventional approach and that are able to mimic more accurately the structure of the neural networks and the characteristics of the biological synapses. A typical example is the concept of in-memory computing, where data are processed directly within the memory, exactly like in the human brain.

Based on this analogy, the research group at Politecnico di Milano have developed a novel circuit that can execute a mathematical function known as regression in just one operation. For this purpose they use a resistive memory, also known as memristor, a device that can memorize any datum (for example the value of a share at a certain time) in the value of its resistance. By arranging these memory elements within an array with the size of a few micrometer (a few millionths of a meter), the group at

Politecnico di Milano has been able to execute a linear regression on a group of data. This operation is capable of determining the straight line that best describes a sequence of data, allowing, for instance, to predict the trend in the stock market based on a simple linear model. Logistical regression, that allows to classify data within a database, has also been demonstrated. This function is essential for the so-called recommendation systems, that are a crucial marketing tools for online purchases.

“One-step regression and classification with cross-point resistive memory arrays (DOI: 10.1126/sciadv.aay2378) study: https://advances.sciencemag.org/content/6/5/eaay2378).