

Artificial intelligence goes bilingual—without a dictionary

S science.org/content/article/artificial-intelligence-goes-bilingual-without-dictionary



Automatic language translation has come a long way, thanks to neural networks—computer algorithms that take inspiration from the human brain. But training such networks requires an enormous amount of data: millions of sentence-by-sentence translations to demonstrate how a human would do it. Now, two new papers show that neural networks can learn to translate with no parallel texts—a surprising advance that could make documents in many languages more accessible.

"Imagine that you give one person lots of Chinese books and lots of Arabic books—none of them overlapping—and the person has to learn to translate Chinese to Arabic. That seems impossible, right?" says the first author of one study, Mikel Artetxe, a computer scientist at the University of the Basque Country (UPV) in San Sebastián, Spain. "But we show that a computer can do that."

Most machine learning—in which neural networks and other computer algorithms learn from experience—is "supervised." A computer makes a guess, receives the right answer, and adjusts its process accordingly. That works well when teaching a computer to translate between, say, English and French, because many documents exist in both languages. It doesn't work so well for rare languages, or for popular ones without many parallel texts.

The two new papers, both of which have been submitted to next year's International Conference on Learning Representations but have not been peer reviewed, focus on another method: unsupervised machine learning. To start, each constructs bilingual dictionaries without the aid of a human teacher telling them when their guesses are right. That's possible because languages have strong similarities in

the ways words cluster around one another. The words for table and chair, for example, are frequently used together in all languages. So if a computer maps out these co-occurrences like a giant road atlas with words for cities, the maps for different languages will resemble each other, just with different names. A computer can then figure out the best way to overlay one atlas on another. Voilà! You have a bilingual dictionary.

The new papers, which use remarkably similar methods, can also translate at the sentence level. They both use two training strategies, called back translation and denoising. In back translation, a sentence in one language is roughly translated into the other, then translated back into the original language. If the back-translated sentence is not identical to the original, the neural networks are adjusted so that next time they'll be closer. Denoising is similar to back translation, but instead of going from one language to another and back, it adds noise to a sentence (by rearranging or removing words) and tries to translate that back into the original. Together, these methods teach the networks the deeper structure of language.

There are slight differences between the techniques. The UPV system back translates more frequently during training. The other system, created by Facebook computer scientist Guillaume Lample, based in Paris, and collaborators, adds an extra step during translation. Both systems encode a sentence from one language into a more abstract representation before decoding it into the other language, but the Facebook system verifies that the intermediate "language" is truly abstract. Artetxe and Lample both say they could improve their results by applying techniques from the other's paper.

In the only directly comparable results between the two papers—translating between English and French text drawn from the same set of about 30 million sentences—both achieved a bilingual evaluation understudy score (used to measure the accuracy of translations) of about 15 in both directions. That's not as high as Google Translate, a supervised method that scores about 40, or humans, who can score more than 50, but it's better than word-for-word translation. The authors say the systems could easily be improved by becoming semisupervised—having a few thousand parallel sentences added to their training.

In addition to translating between languages without many parallel texts, both Artetxe and Lample say their systems could help with common pairings like English and French if the parallel texts are all the same kind, like newspaper reporting, but you want to translate into a new domain, like street slang or medical jargon. But, "This is in infancy," Artetxe's co-author Eneko Agirre cautions. "We just opened a new research avenue, so we don't know where it's heading."

"It's a shock that the computer could learn to translate even without human supervision," says Di He, a computer scientist at Microsoft in Beijing whose work influenced both papers. Artetxe says the fact that his method and Lample's—uploaded to arXiv within a day of each other—are so similar is surprising. "But at the same time, it's great. It means the approach is really in the right direction."
