

PURELY NEURAL MACHINE TRANSLATION

ISSUE 1



NEURAL MACHINE TRANSLATION (NMT): LET'S GO BACK TO THE ORIGINS

Each of us have experienced or heard of deep learning in day-to-day business applications. What are the fundamentals of this new technology and what new opportunities does it offer?

The concept of deep learning is closely linked to artificial neural networks. These networks have completely changed the way humans work with machines. In the classic approach, we give instructions to the computer by breaking up a problem into a sequence of tasks and algorithms. In contrast, in an artificial neural network, no indications are given to the computer on how to process the problem. Instead, we provide the data that the machine then uses to learn by itself and find solutions on its own.

Today artificial neural networks and deep learning bring powerful solutions to several domains such as image recognition, big data analysis, digital assistants, and natural language processing. These solutions have already been deployed on a large scale by Google (AlphaGo, automatic captioning...), Microsoft (Cortana...) and Facebook.

In the last two years, a lot of research has been conducted on artificial neural networks as applied to natural language processing. Results are shared among an open source community in which SYSTRAN actively participates. Impressive results are produced almost daily.

WHAT MAKES NMT A TECHNOLOGICAL BREAKTHROUGH?



Unlike statistical (SMT) or ruled-based (RMT) engines, NMT engines process the entire sentence, paragraph or document. The entire chain is processed end-to-end with no intermediate stages between the source sentence and the target. The NMT engine models the whole process of machine translation through a unique artificial neural network.

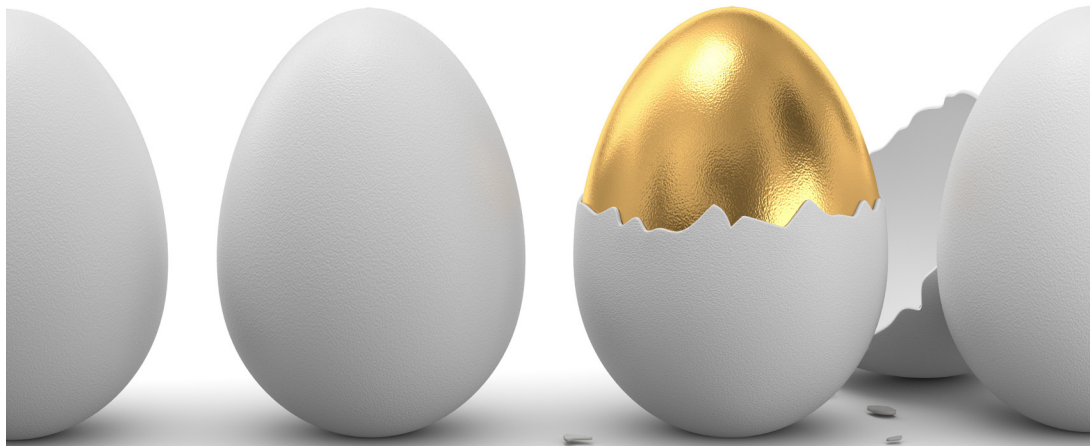
However, similar to the human brain, some complementary neural sub-networks activate themselves within this unique neural network as the translation is being generated:

- a first subnetwork addresses the source sentence to extract its meaning,

- a second specialized in syntactic (grammar) or semantics (words meaning) analysis enriches understanding,
- a third contextualizes the content,
- another focuses on keywords...

All these subnetworks communicate with the engine and allow it to ultimately choose the best translation with a quality overachieving the current state of the art!

WHAT MAKES SYTRAN'S OFFERING UNIQUE?



Since its creation, SYSTRAN devoted all its R&D effort to creating machine translation and natural language processing solutions at the cutting edge of technology. As the pioneer of free online translation tools, enterprise server solutions for business, embedded applications on mobile, it is no surprise that SYSTRAN leads the way launching today the engine that brings the greatest technological leap in the history of machine translation.

We noticed that a neural network trains itself on the data provided. Unlike in previous generations of engines where a huge volume of data was mandatory, the neural network feeds itself on enriched data. The quality and the wealth of these data largely count on their quantity. The expertise that SYSTRAN has been acquiring for over 40 years, has made it possible today to provide artificial neural networks with data enriched by terminology and annotated resources.

Artificial neural networks have a terrific potential but they also have limitations, particularly to understand rare words.

SYSTRAN mitigates this weakness by combining artificial neural network and its current terminology technology that will feed the machine and improve its ability to translate.

SYSTRAN exploits the capacity NMT engines have to learn from qualitative data by allowing translation models to be enriched each time the user submits a correction. SYSTRAN has always sought to provide solutions adjusted to the terminology and business of its customers by training its engines on customer data. Today SYSTRAN offers a self-specialized engine, which is continuously learning on the data provided.

It is important to point out that graphic processing units (GPUS) are required to operate the new engine. Also, to quickly make available this technology, SYSTRAN will provide the market with a ready-to-use solution using an appliance (that is to say hardware and software integrated into a single offering). In addition, the overall trend is that desktops will integrate GPUs in the near future as some smartphones already do (the last iPhone can manage neural models). As size is becoming less and less of an issue, NMT engines will easily be able to run locally on an enterprise server.

THE ADVENTURE HAS JUST BEGUN...



The technology is ready and the results are amazing. A demonstrator will be online during October. The next step for SYSTRAN is to deliver this breakthrough technology to its customers. Several of them have already been participating in a beta test program to give back feedback in real situations. Customers can expect a smooth and transparent change. The noticeable change is the quality of the output translations, and what a change!

ABOUT SYSTRAN

To help organizations enhance multilingual communication and increase productivity, SYSTRAN delivers real-time language solutions for internal collaboration, search, eDiscovery, content management, online customer support and e-Commerce.

With the ability to facilitate communication in 140+ language combinations, SYSTRAN is the leading choice of global companies, Defense and Security organizations, and Language Service Providers. SYSTRAN has also been the technological choice of Samsung for its embedded translation application, S-Translator, available on the Galaxy S and Note series, and of Sonico for the latest version of its mobile application, iTranslate.

Since its early beginnings, SYSTRAN has been pioneering advances in Machine Translation and Natural Language Processing and delivers today a new generation of engines leveraging the last technological innovations from Artificial Neural Networks and Deep Learning models.

For more information, visit www.systrangroup.com

CONTACT

Gaëlle BOU
Marketing & Communication Director
gaelle.bou@systrangroup.com
+33 1 44 82 49 50

ABOUT THE PNMT INSIGHT SERIES

Project "PNMT" (Purely Neural Machine Translation) was this year's flagship project for the researchers and developers at SYSTRAN

SYSTRAN brings its expertise to the sector in several ways: contributing to research on neural models; applying its know-how in terminology to increase the potential of Neural Machine Translation; and industrializing technology to make it available to companies, organizations and individuals.

We will keep you posted each month and share best practices, research paper, customers insights...

