**Facebook Develops Machine Translation System for 100 Languages**

**脸书研发可翻译100种语言的机器翻译系统**

Facebook has developed the first machine learning model that can translate between any two of 100 languages without going into English first.

脸书研发出第一个机器学习模式的翻译系统。它可以在100种语言中的任意两种语言之间进行翻译，而无需先译成英语。

Facebook says the new multilingual machine translation model was created to help its more than two billion users worldwide. The company is still testing the translation system – which it calls M2M-100 - and hopes to add it to different products in the future.

脸书表示这一新型多语种机器翻译模式是为了帮助其全球20多亿用户。该公司仍在测试这个被称为“M2M-100”的翻译系统，并希望在未来将其添加到不同的产品中。

The social media service says it has made the system open source -- meaning its computer code will be freely available for others to copy or change.

这家社交媒体服务公司表示，他们已将该系统开源——这意味着它的计算机代码将可以免费供他人复制或更改。

Angela Fan, a research assistant at Facebook, explained the new machine translation model this week on one of the company’s websites. She said its development represented a “milestone” in progress after years of “foundational work in machine translation."

本周脸书公司的研究助理安吉拉•范在该公司的一家网站上对这一新的机器翻译模式做了解说。她表示，它的研发是数年“机器翻译基础工作”发展中的一个“里程碑”。

Fan said the model produces better results than other machine learning systems that depend on English to help in the translation process. The other systems use it as an intermediate step -- like a bridge -- to translate between two non-English languages.

范女士还表示，与其他依靠英语帮助翻译的机器学习系统相比，这种模式能得出更好的翻译效果。其他系统会利用英语作为一个中间步骤，就像一座桥梁一样，在两种非英语语言之间进行翻译。

One example would be a translation from Chinese to French. Fan noted that many machine translation models begin by translating from Chinese to English first, and then from English to French. This is done “because English training data is the most widely available," she said. But such a method can lead to mistakes in translation.

其中一个例子就是汉语翻译成法语。安吉拉指出，很多机器的翻译模式都是先把中文翻译成英文，然后再把英文翻译成法文。她说，这样做“是因为英语的可用数据应用最为广泛”，但这种翻译方法会导致翻译出现错误。

"Our model directly trains on Chinese to French data to better preserve meaning,” Fan said. Facebook said the system outperformed English-centered systems in a widely used system that uses data to measure the quality of machine translations.

安吉拉还指出：“我们的翻译模型是直接使用中文到法文的培训数据，它能更好地保留原文的含义。” 脸书表示，在一个使用数据衡量机器翻译质量的广泛应用的系统中，新系统的性能明显优于以英语为中心的翻译系统。

Facebook says about two-thirds of its users communicate in a language other than English. The company already carries out an average of 20 billion translations every day on Facebook’s News Feed. But it faces a huge test with many users publishing massive amounts of content in more than 160 languages.

脸书称，他们有大约三分之二的用户用英语以外的语言交流。该公司在脸书每天的动态消息中已有日均200亿次的翻译。但它也面临着一个巨大的考验——数量庞大的用户群用160多种语言发布大量内容。

The development team trained, or directed, the new model on a data set of 7.5 billion sentence pairs for 100 languages. In addition, the system was trained on a total of 2,200 language directions. Facebook said this is 10 times the number on the best machine translation models in the past.

开发团队在一个包含100种语言的75亿个句子的数据集上研发培育出这种新模型。此外，该系统还接受了总共2200个语言方向的培训。脸书表示，这一数量是已有的最好的机器翻译模式的10倍。

One difficulty the team faced was trying to develop an effective machine translation system for language combinations that are not widely used. Facebook calls these “low-resource languages.” The data used to create the new model was collected from content available on the internet. But there is limited internet data on low-resource languages.

研究小组面临的一个难题是为没有被广泛使用的语言组合开发一个有效的机器翻译系统。脸书称这些语言为“低资源语言”。用于创建新翻译模型的数据是从互联网上收集的，然而关于低资源语言的互联网数据资源很有限。

To deal with this problem, Facebook said it used a method called back-translation. This method can create “synthetic translations” to increase the amount of data used to train on low-resource languages.

脸书表示，为解决这一问题它使用了一种叫做“回译”的方法。这种方法可以通过创建“合成翻译”来增加用于低资源语言培训的数据量。

For now, the company says, it plans to continue exploring new language research methods while working to improve the new model. No date has been set for launching the translation system on Facebook.

该公司还表示，目前它计划继续探索新的语言研究方法同时努力改进新的翻译模式。目前还没有确定在脸书上推出这一翻译系统的日期。

But Angela Fan said the new system marks an important step for Facebook, especially for the times we live in. "Breaking language barriers through machine language translation is one of the most important ways to bring people together, provide authoritative information on COVID-19, and keep them safe from harmful content," she said.

但安吉拉·范表示，新系统标志着脸书迈出了重要一步，尤其是对我们所处的时代而言。“通过机器语言翻译系统来打破语言障碍是将人们聚集起来、提供有关新冠肺炎的权威信息并使人们免受有害内容影响的最重要途径之一。”

I’m Bryan Lynn.

布莱恩·林恩报道。

**Facebook Develops Machine Translation System for 100 Languages**

Facebook has developed the first machine learning model that can translate between any two of 100 languages without going into English first.

Facebook says the new multilingual machine translation model was created to help its more than two billion users worldwide. The company is still testing the translation system – which it calls M2M-100 - and hopes to add it to different products in the future.

The social media service says it has made the system open source -- meaning its computer code will be freely available for others to copy or change.

Angela Fan, a research assistant at Facebook, explained the new machine translation model this week on one of the company’s websites. She said its development represented a “milestone” in progress after years of “foundational work in machine translation."

Fan said the model produces better results than other machine learning systems that depend on English to help in the translation process. The other systems use it as an intermediate step -- like a bridge -- to translate between two non-English languages.

One example would be a translation from Chinese to French. Fan noted that many machine translation models begin by translating from Chinese to English first, and then from English to French. This is done “because English training data is the most widely available," she said. But such a method can lead to mistakes in translation.

"Our model directly trains on Chinese to French data to better preserve meaning,” Fan said. Facebook said the system outperformed English-centered systems in a widely used system that uses data to measure the quality of machine translations.

Facebook says about two-thirds of its users communicate in a language other than English. The company already carries out an average of 20 billion translations every day on Facebook’s News Feed. But it faces a huge test with many users publishing massive amounts of content in more than 160 languages.

The development team trained, or directed, the new model on a data set of 7.5 billion sentence pairs for 100 languages. In addition, the system was trained on a total of 2,200 language directions. Facebook said this is 10 times the number on the best machine translation models in the past.

One difficulty the team faced was trying to develop an effective machine translation system for language combinations that are not widely used. Facebook calls these “low-resource languages.” The data used to create the new model was collected from content available on the internet. But there is limited internet data on low-resource languages.

To deal with this problem, Facebook said it used a method called back-translation. This method can create “synthetic translations” to increase the amount of data used to train on low-resource languages.

For now, the company says, it plans to continue exploring new language research methods while working to improve the new model. No date has been set for launching the translation system on Facebook.

But Angela Fan said the new system marks an important step for Facebook, especially for the times we live in. "Breaking language barriers through machine language translation is one of the most important ways to bring people together, provide authoritative information on COVID-19, and keep them safe from harmful content," she said.

I’m Bryan Lynn.