

羅致翹博士：了解機器翻譯 分析錯譯原因

轉載文章 | 2021/1/23 — 11:48



Google 翻譯曾將「China breaks promise」譯成「中國信守諾言」。

【文：羅致翹博士】

在 2019 年中及剛過去的週末，Google Translate 兩次被發現翻譯結果有災難性的錯誤。由於兩次錯誤都牽涉香港和「鄰近地區」，事件引起網上不少騷動。一時之間草木皆兵，大家以為 Google Translate 要不是故意偏袒，就是受到惡意攻擊。本文旨在簡單介紹 Google Translate 的基本運作和提供有關兩次錯誤的部分觀察，讓讀者加深了解以釋除疑慮。

機器翻譯 (Machine Translation, MT) 的前世今生

機器翻譯一直是人工智能研究領域中的長青題目。自上世紀七十年代起，由語言學家人手編寫「規則式機器翻譯」(Rule-based, RBMT)，到九十年代利用機器學習 (Machine Learning) 的「統計式機器翻譯」(Statistical, SMT)，到 2015 年機器翻譯突破性發展成「神經元式機器翻譯」(Neural, NMT)。

RBMT 受限於人手編寫規則，所以系統的規模和表現有限，只有在預設的領域 (domain)、話題 (topic) 和句式 (style) 方有較好的表現。

SMT 將整個翻譯過程分拆為多個統計模組，在大量的訓練材料中數算統計模組的參數，情況有如機器從很多前人的翻譯作品中學會了翻譯，因此這種過程被稱為機器學習。【圖1】簡單說明 SMT 最基本的統計模組，如結果的字數、位置等。訓練材料來自翻譯員編寫、網上搜集及用戶回饋。

SMT 的優勢在於翻譯的話題和句式比 RBMT 廣泛：只要詞彙和譯文在訓練材料中出現得愈頻密，SMT 就愈能在翻譯時從記憶中拼湊出結果。SMT 最主要的局限是對語意完全缺乏理解，單純地硬背自學的詞彙表。因此，SMT 的表現很快就到了樽頸，在往後的日子發展變得緩慢。

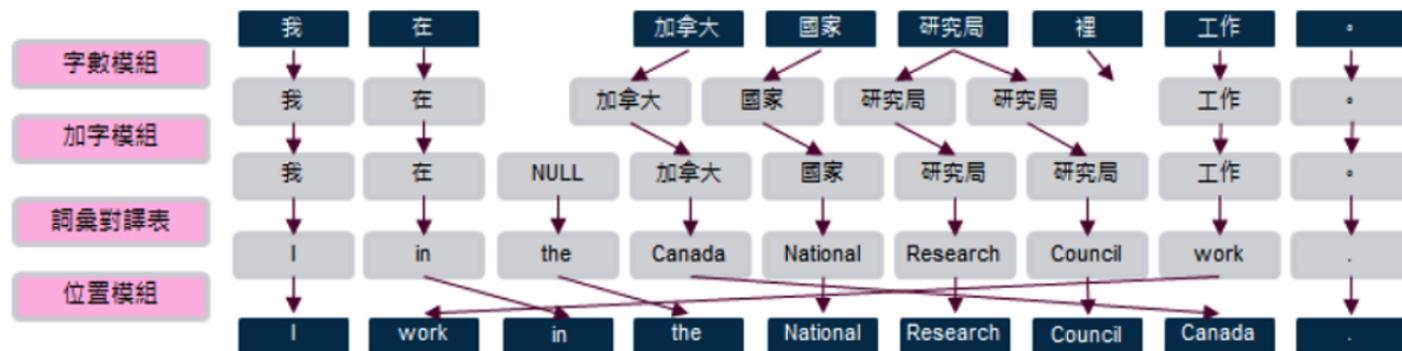


圖1. SMT 基本示範。作者提供

初代 Google Translate 就是 SMT 系統。值得一提的是，縱使 Google Translate 的用戶介面看似支援多種語言互譯，但實際上每一個翻譯方向都是一個獨立的系統，而且大部分翻譯方向都要通過以英文作為中介的連鏈管道 (daisy chaining pipeline)。例如，德文譯向中文，系統會先將德文譯為英文的中介結果，繼而將英文的中介結果翻譯成中文譯文。

而 NMT 就是利用了神經網絡 (neural network) 模擬大腦腦細胞 (神經元, neuron) 的運作模式，將原文視為輸入網絡的刺激，再通過已經訓練及學習後的神經網絡得出譯文，就像腦細胞受到刺激而作出反應一樣。

【圖 2】所見 NMT 將原文文字嵌入成向量 (vector)，通過神經網絡中神經元的運算後，輸出譯文的第一個字。已輸出的譯文會被視為刺激的一部分，重新輸入到網絡，用以猜想下一個要輸出的譯文文字。

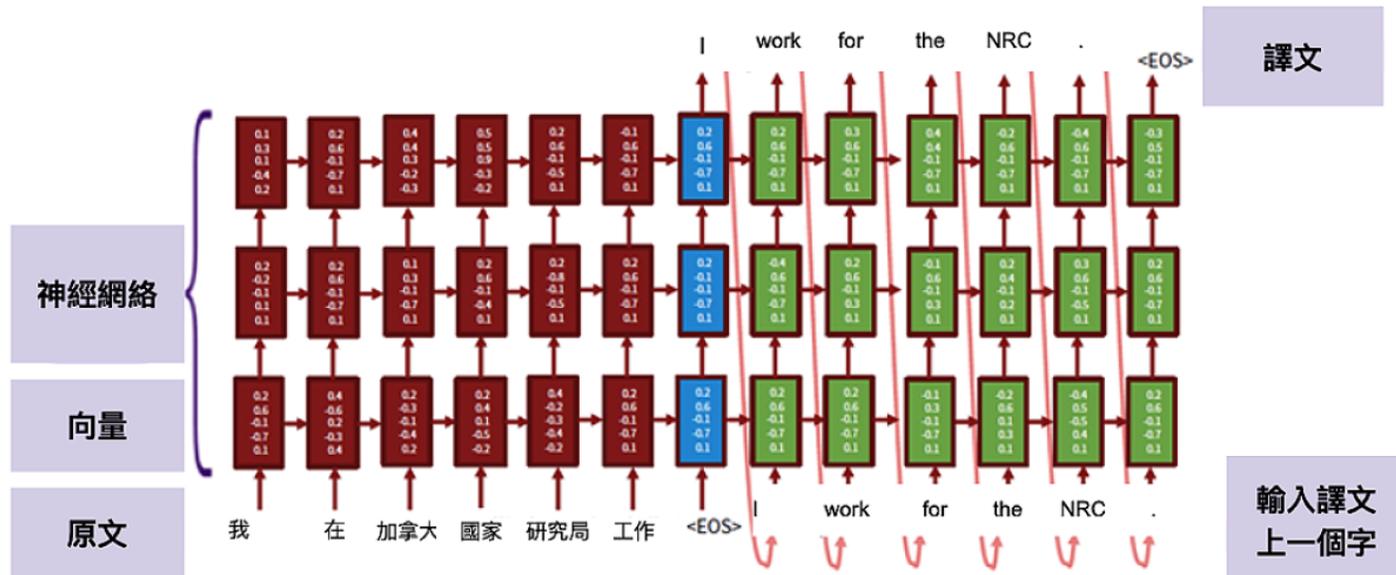


圖2. NMT 圖解。作者提供 Based on Luong et al. 2016. Neural Machine Translation. Proceedings of the 54th Annual Meeting of the ACL: Tutorial

神經網絡學習需要經過大量矩陣運算，而傳統中央處理器 (Central Processing Unit, CPU) 運算耗時甚多。近年圖像處理器 (Graphics Processing Unit, GPU) 的運算能力比 CPU 有所提升，GPU 更能以高速執行矩陣運算，使神經網絡學習得以成為可實現的系統。

有部分證據顯示神經網絡中的向量空間 (vector space) 能夠學習語義 (semantics/meaning)，因為意思類別相近字詞的向量位置非常相近。機器翻譯在 NMT 世代迎來大幅度的進步，譯文顯得更準確更流暢。2016 至 17 年間 Google Translate 已發展成 NMT。

但凡事有得必有失，SMT 自學的詞彙表其實是系統人員為 SMT 除錯 (debug) 的重要工具。NMT 摒棄了詞彙表，人員便失去了除錯工具，系統亦失卻了 SMT 可解釋、可追查的特性。NMT 系統非常龐大，訓練材料非常多，系統內深層的參數已經超越人類可閱讀或理解的層面，NMT 系統的除錯因而變得非常困難。

文意極性 (Polarity) 是 MT 常見錯誤出處

極性即是意思正反方向 (指句子的肯定和否定性)。人們常以為「保持原文極性」這個翻譯要求是很容易達到，但事實上這要求對機器而言是非常困難的。RBMT 需要人手編寫所有令極性轉變的條件，這是一個近乎不可能的任務。SMT 將原文一個字詞翻譯成的字數化為統計機率，使過程中可能忽略了當中幾個字。

要是被忽略的字剛好令原文極性轉變，如「不會」，翻譯結果就會出現災難性的錯誤。NMT 會自行將意思類別相近的字放在向量空間中相近的位置，例如「傷心」與「高興」兩詞雖然極性相反，但他們在向量空間中的距離會比與「咖啡」或「踢波」近得多。可是這樣的「安排」卻令 NMT 容易因位

置偏差而產生極性錯誤。

NMT 的穩健性 (Robustness) 問題

NMT 經常出現穩健性問題，原文稍有串錯字或不合文法，系統都會失控，造成災難性的錯誤。

2019 年 6 月，Google Translate 將 “So sad to see Hong Kong become China” 錯譯成「很高興看到香港成為中國」，事件正正反映 NMT 穩健性備受原文句子的文法錯誤所影響。【圖 3】所見，最後一句原文句子因較少文法錯誤，所以翻譯結果比較理想。

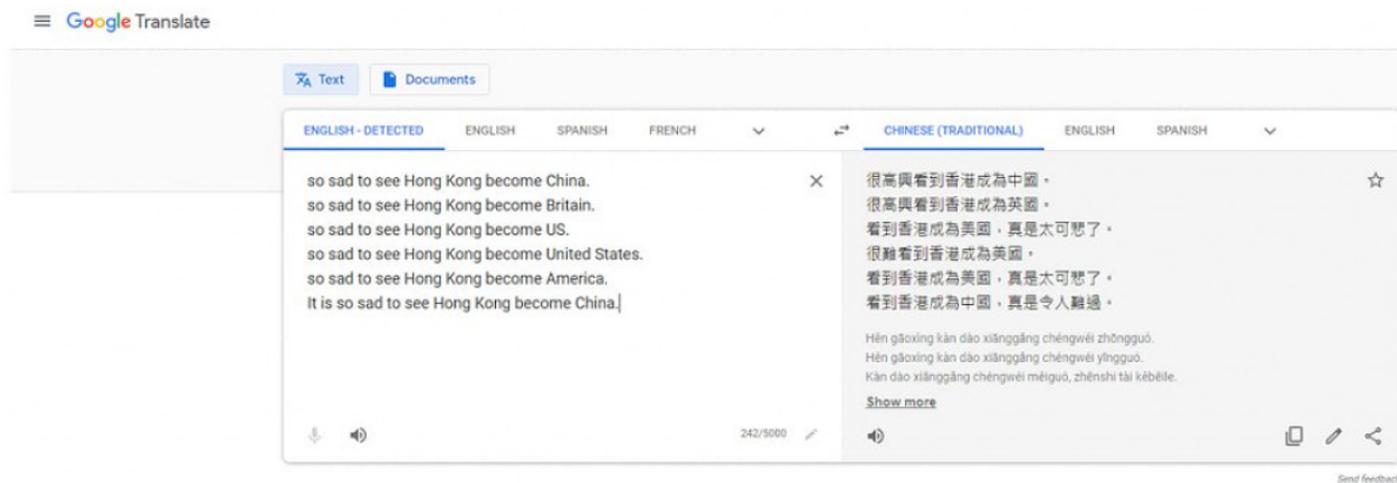


圖3. 黑盒測試 —— 原文不合文法，圖為2019年測試，其後已修正。作者提供

情感字詞如 happy、sad、angry，在 NMT 系統裏面的嵌入位置非常接近。只要原文稍有錯誤，系統便無法識別原意的準確位置，繼而單憑猜測選擇附近的位置。

因此在系統失控的情況下，系統不會完全脫序地翻譯成「咖啡」、「奶茶」等完全不相干的字眼。

【圖 4】所見，每多輸入一個字母，翻譯結果隨即跳轉。這顯示「高興」與「生氣」（及「可悲」）在輸出向量空間 (output vector space) 中位置非常接近。當原文出現文法錯誤時，就會影響到 NMT 嵌入原文文意 (context) 的準確性，然後解碼時所輸出向量就落在「高興」、「生氣」、「可悲」附近的位置，稍有改變即使結果跳轉。

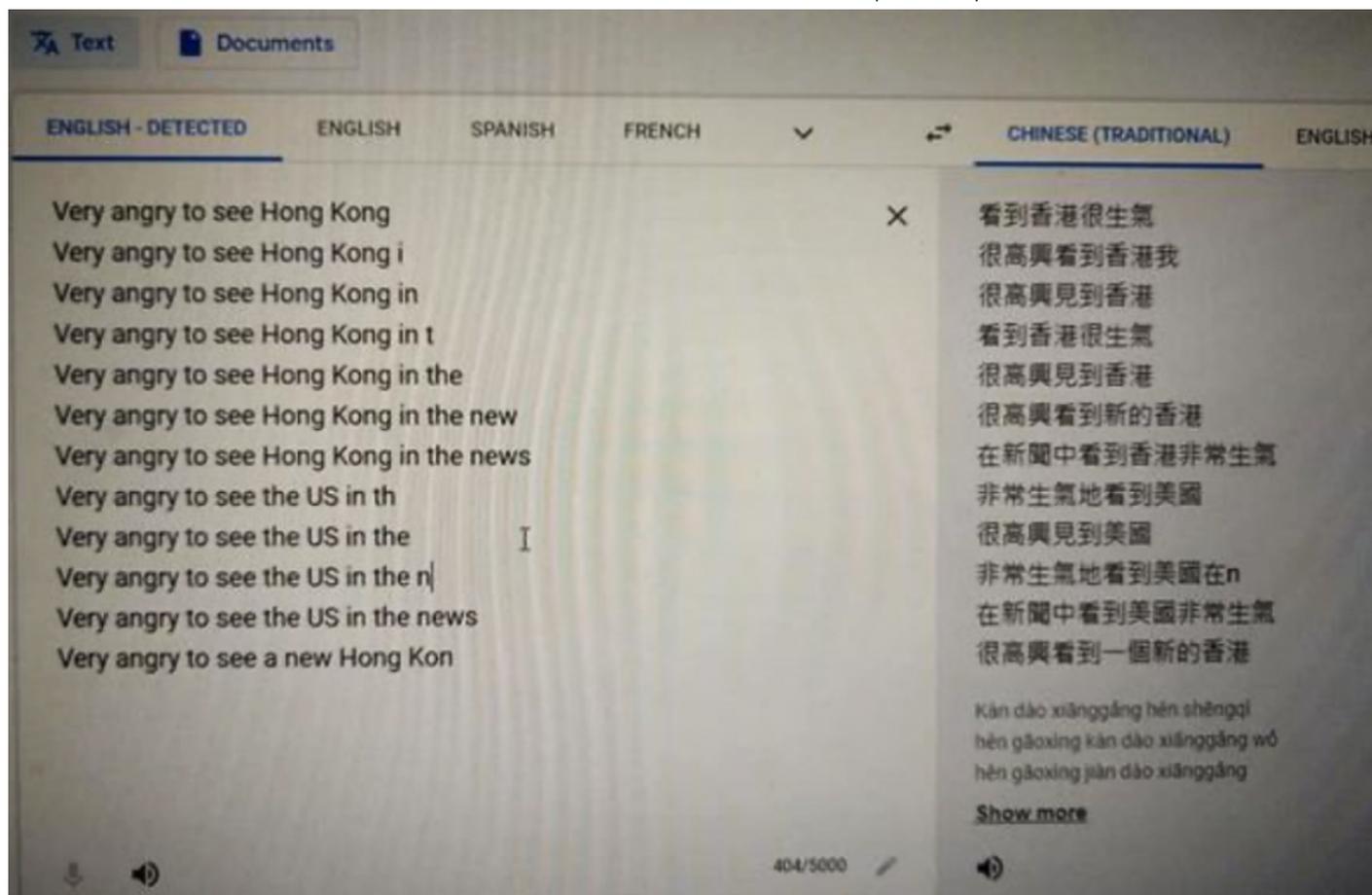


圖4. 黑盒測試 —— 翻譯結果敏感跳轉。作者提供

至於今年“China breaks its promise”錯譯成「中國信守承諾」事件中，經測試將China代為 Hong Kong、Taiwan、Singapore都會出現同樣的翻譯錯誤 —— 這四個地名的嵌入位置接近，因此這翻譯錯誤算是合理表現。

但更嚴重的錯誤在“China breaks its promises”，將 China 以多個不同國家取代，如UK、Thailand、Korea、Russia、Kazakhstan 等等都出現相同錯誤。

另一點觀察比較特別：“China breaks its promise”文法無誤，但若以 break 取代 breaks，雖然原文文法錯誤，但翻譯卻跳轉成正確結果。今年這個錯誤顯然比 2019 年的較有可能源於錯誤的訓練材料。但由於大量地名都出現錯譯，而且錯與對之間原文只差一個字母，這顯示問題源於系統穩定性的可能，大於系統訓練材料因用戶回饋機制而受到惡意毒害攻擊（poisoning attack）。

NMT 穩健性的缺陷尚未修正，但並不代表不能修正。

NMT 的穩健性已經引起科學界關注，世界各地的科學家都正在為拆解這個缺陷而努力。自 2019 年起，Conference on Machine Translation 新增了穩健性的競賽項目，以相同開發條件為前提，建造最穩健的系統。

大學與商業實驗室紛紛投入資源，研究如何令 NMT 在面對雜亂的原文時表現更穩定。而筆者作為翻譯質素評測及估測研究員，曾聯同美國馬利蘭大學學院市分校研究如何找出含有誤導的翻譯結果，目標旨在提醒用戶可能收到含有誤導的翻譯，不宜盡信。縱然現時只可作臨時修補——見一個洞補一個，但筆者相信在一眾科學家的努力下，治本的方法終會出現。

科學要以證據支持，來找出可能性最高的結論。上述兩個例子即使非經嚴謹科學實驗論證，但簡單黑盒測試已可提供足夠觀察和證據，顯示錯譯源於翻譯系統的穩定性。

筆者與大家一同面對在香港備受壓迫的生存空間，同樣感受大家對生命與生活的不安。但在徬徨午夜裏、在迷霧裏，我們不應自己嚇自己至杯弓蛇影、草木皆兵。願科學的實事求是成為你我心裡其中一道號角聲，願榮光歸於生我育我的香港。

作者簡介: 羅致翹博士，香港科技大學畢業，受聘於加拿大國家研究局。專注研究及評測機器翻譯，更不時於業內頂級國際學術會議發表文章，在機器翻譯質素評測方面得到頻繁關注及引用。此文為羅博士個人專業分析，與加拿大國家研究局無關。

關於我們 客戶查詢 技術支援 私隱政策 承印人：Best Pencil (HK) Ltd
地址：九龍觀塘開源道55號開聯工業中心B座14樓06室

© 2021 立場新聞. All rights reserved.