

科大教授創「倒置文法」 電腦翻譯快5,400倍 為Google網譯Siri奠基 吳德愷膺ACL首屆院士

香港文匯報訊(記者 劉景熙)全球一體化,要打破世界各地的文化隔膜,以高科技結合翻譯的技術是其中一座重要的橋樑。致力研究人工智能的香港科技大學計算機科學及工程學系副教授吳德愷,憑其於翻譯計算理論的革命性貢獻,早前更成為「國際計算語言協會 (ACL)」首屆院士,其成就被譽為媲美獲「電腦語言科技界的諾貝爾獎」。他針對翻譯時需要注意的語法特性,成功結合電腦運算模式,創出「倒置轉移文法」理論,將翻譯所需時間大大減少。以16組字詞的中英對譯為例,其工作令電腦翻譯速度提高達5,400倍,為近代先進系統如Google網上翻譯、iPhone人工智能程式Siri等背後的運算理論奠下重要基礎。

近年網上翻譯系統的使用十分方便快捷,很多熱門的搜尋網站大都附設翻譯程式,能整篇文章作對譯。吳德愷在接受本報訪問時解釋,這些可算是「次世代」的翻譯系統。由於不同語言的文法大不相同,翻譯系統不只需要掌握單字含義,如何能通順地譯出語句,是當中的大挑戰。

首代用字生硬 二代自然流暢

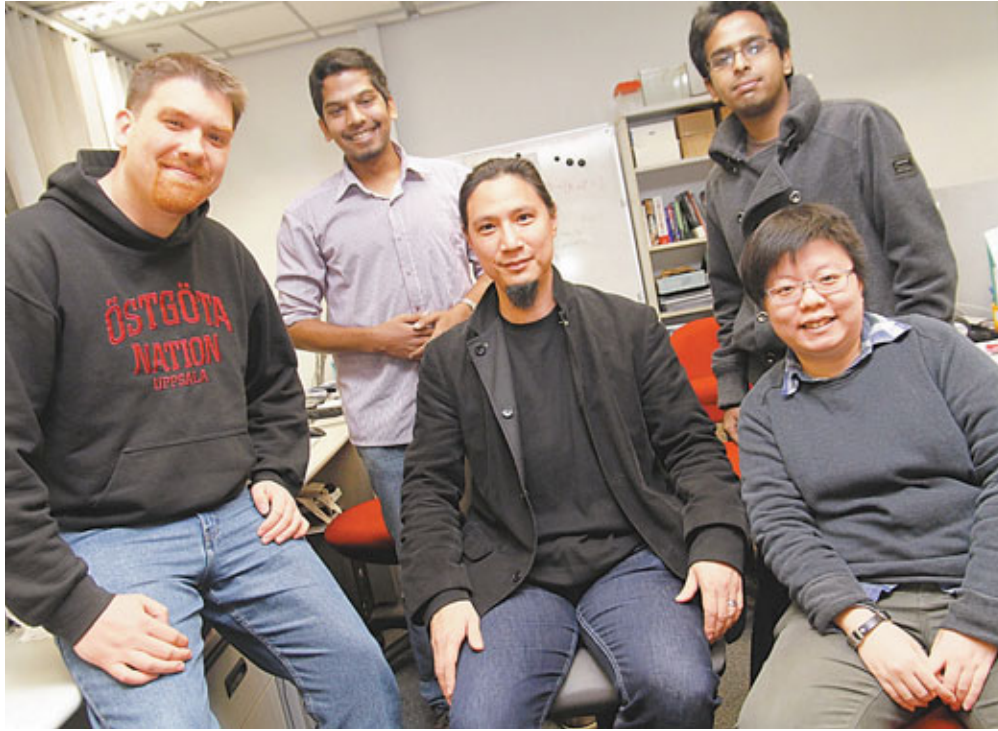
他說,早在上世紀七八十年代起,科學家一致認為要向電腦輸入大量文法規則和詞語意思,之後讓電腦「按本子辦事」即可,這被視為「第一代」人工智能系統。不過,該類系統翻譯時既欠精確且用字生硬,於是學界開始透過實際生活中的大量生活語例,輔以統計學的概率理論,使譯句更自然流暢,被視作「第二代」人工智能翻譯系統。

合併字詞單元 翻譯加速

吳德愷早年在美國攻博士時已是「概率理論」學派出身,他在上述「第二代」的基礎上,提出「倒置轉移文法 (Inversion Transduction Grammar, ITG)」理論模型,大幅強化翻譯效率。他以香港歷年立法局及立法會的中英翻譯語例為基礎,按概率理論,把中或英的字詞單元合併成經常出現而內容有機會倒置的詞組,從而大幅減少電腦運算時所須處理的單位 (unit) 數目,令翻譯速度明顯提高,譯文也更自然貼近原意,而非機械式直譯。

他解釋,如句子涉及4個字詞,翻譯時共有24個排序組合,而當應用ITG後,組合會減少至22個,效果雖不明顯。但由於排序組合數量會隨字詞的增加以幾何級數上升,當詞句越長組合量便會大增,6個字詞翻譯組合有720個,但ITG可將之減少近半至少於400個;如果要翻譯16個字的句子,組合數量便達驚人的約21萬億個,往往需要超級電腦才能有效處理;而ITG則可把運算的「單位」刪減至39億個,讓翻譯速度快上5,400倍(原理見說明圖)。吳德愷笑言:「16個字還只是一句很短的句子呢。」

事實上,以ITG為基礎,近代電腦翻譯理論出現劃時代的轉變:電腦要處理的數據量大為降低,不但提升演算效率,亦使一般家用電腦,同樣有條件應付日常的文章翻譯,大力促進互聯網時代知識的流通。亦因有關貢獻,吳德愷早前獲選為國際權威的ACL首批17名院士之一,也是當中唯一來自大中華地區的學者。



科大計算機科學及工程學系副教授吳德愷與他的研究團隊。 彭子文 攝

ITG與一般直譯對比說明圖

以句子「Where is the Secretary of Finance when needed」為例,當中the及of為虛詞,翻譯上沒有對照意思,共有6個字詞單元。

直接翻譯: (哪裡) (在) (司) (財政) (時) (有需要)
英文句子: (Where) (is) the (Secretary) of (Finance) (when) (needed)

正確排序: (財政) (司) (有需要) (時) (在) (哪裡)
按數學組合公式,所有隨機排序共有720個組合,即電腦要運算720次才能完全成功翻譯;字詞單元數越多,運算次數會以幾何級數提升。

「倒置轉移文法」理論:
從翻譯資料庫中以概率理論歸納出文法慣例,將字詞單元重編成字詞組合,而各字詞組合間的單元均可「倒置」,即:

英文句子: (Where is) (Secretary Finance) (when needed)

倒置轉移文法: (司財政/財政司) (時有需要/有需要時) (哪裡在/在哪裡)

正確排序: (財政) (司) (有需要) (時) (在) (哪裡)
應用「倒置轉移文法」理論後,排序組合數量大幅減少至少於400個,電腦需要運算的次數也大減,翻譯效能大增;當原句字詞單元數越多,翻譯時提速程度越高。以16字為例,隨機排序組合達21萬億個,但倒置轉移文法能將大量排序組合剔除至只剩下39億個,電腦運算有關組合的速度便提高約5,400倍。



吳德愷一直對父親是吳家瑋(圖右)一事保持低調,甚至改自己的姓氏拼音,避免張揚。資料圖片

虎父無犬子 改姓氏拼音免張揚

香港文匯報訊(記者 劉景熙)吳德愷的父親,是科大創校校長吳家瑋。對此,吳德愷一直保持低調,「我自少在美國長大,受那裡的教育影響很深。美國精神強調每個人不論出身自何家庭,都是平等的,你只是你自己。」

吳家瑋是一名物理學家,在美國學術界甚為有名,更在1983年出任三藩市州立大學校長,成為美國史上首位擔任美國著名大學校長的華人。吳德愷很早意識到這點,所以中學時故意改變自己的姓氏拼音,由原本「Woo」改為「Wu」,避免太張揚:「即使我在加州柏克萊大學唸了7年博士,我的指導教授也不知道這件事,直到畢業典禮時,我向他介紹我的家人,他當場嚇了一跳!」

共同回憶:到內地大學派電腦

兩父子的共同回憶,吳德愷印象最深的,是11歲那年隨父親到內地的經歷,「當時『文化大革命』剛結束,大學很多設備都受到破壞。父親和我帶上多部電腦,到大學各個學系派發,希望當時中國的學術發展可盡快重上軌道;之後我們又遊覽了18個內地城市。」

雖然兩人在工作上保持距離,父親總是對吳德愷言傳身教:「他和母親經常提醒我,人的價值不是建基於財富、樣貌,而是你有多努力學習、工作,如何貢獻社會、造福人類。」

大戰翻譯機 人類勝電腦

香港文匯報訊(記者 劉景熙)若電腦與人類進行翻譯大比拼,將是誰勝誰負?吳德愷曾嘗試找來一些學生與電腦進行測試;結果發現,只要給予充分時間,人類在準確、流暢度方面的表現會更勝一籌。

為了測試電腦翻譯的質素,吳德愷設計了一個實驗遊戲:由學生A做「挑戰者」,他要以中文提問,分別交予同學B及電腦翻譯成英文問題。兩者的翻譯結果,會再分別交予學生C及學生D作答。「挑戰者」需要推敲、分辨2個答案,各自從哪個問題而來;若果他成功辨認,將會獲得分數。

吳德愷整理數據後發現,如果每局只有3分鐘時限,電腦與人類表現相若,或僅只是略遜;但一旦把時限提升1倍至6分鐘,人類的優勢即時變得明顯,反勝電腦,「因為『挑戰者』有更多時間去設計問題,令電腦難以掌握意思;另一方面,負責翻譯及作答的同學亦有更充足的時間應付。」

走入商界 將科技帶進生活

香港文匯報訊(記者 劉景熙)吳德愷在1992年回流香港,到科大從事電腦語言翻譯研究。直到2000年,他曾經成立語言科技公司「Weniwen」,從象牙塔走入商界,希望學以致用,把科研帶入群眾生活,之後又完成科大行政人員工商管理碩士 (EMBA) 課程,惜公司最後在2002年破產。

對於這次「親自下海」的創業體驗,10年後的今天,他認為甚有價值。他特別提到EMBA的課堂生活,「很多同學都來自公司的管理層,在課堂上堂,可以左邊是公司的CEO,右邊則是集團副主席。討論時,他們你一言、我一語,加深我對商界運作細節的了解。此外,他們大多來自不同國家、民族,亦令我對世界不同文化有更多體會。」

批港抄襲成風 缺「市場創造者」

香港文匯報訊(記者 劉景熙)內地與香港近年均着力發展科技產業,大家都希望可產出「香港喬布斯」、「亞洲hotmail系統」。本身是加州柏克萊大學電腦科學博士的吳德愷批評,這些「目標」本身已不算創新,只是活在別人的陰影下:「香港最欠缺的,是『市場創造者』的氣魄!」

不重創新 只壓成本

吳德愷熟悉美國矽谷的情況,眼看香港推動科研發展多年,不無感慨:「當年不少同學熱衷創新,例如今天知名的網路搜尋器Google、網頁瀏覽器火狐 (Mozilla Firefox),他們都有份參與製作,這些產品的影響力,我們有目共睹。不過,香港的『創新風氣』,多只着重如何降低產品成本,忽略嶄新的想法。我在香港做科研,經常聽見別人說要打造出『亞洲Amazon、Hotmail』,但『照抄』西方已經存在的一套過來,然後只思考如何壓低成本,根本

是Copy Cat的行為!」

孕育創意 打造市場

香港的問題纏結,吳德愷認為在於心態短視,欠缺空間孕育創意,不能培養出帶領世界進步的「市場創造者」(Market Creator),「未有蘋果iPhone以前,沒有人想到電話是一部多功能的掌上電腦,包括上網、拍照、聽歌、打機等。」吳德愷強調,這些是科技創新的「範式轉移」(Paradigm shift),即是一種「根本、巨大的轉變」,然後改變了世界整個市場。

他認為,真正有創意的人,不會是別人的跟隨者 (follower),而是藉產品創新,打造一個全新的市場的「市場創造者」,帶領世界潮流。這會涉及很多資源和時間,雖然當中有些研究,最後可能沒有任何成果,但創新就是如此,建基於不斷的嘗試、挫折,才能應運而生。



吳德愷感慨香港科技發展多年,很多業內人士忽略嶄新想法,難以成為東方矽谷。 香港文匯報記者彭子文 攝

海外升學秘笈

海外留學生生活之「住」篇

在海外留學生生活的「衣食住行」中,同學最關注的便是留學海外的住宿安排。其實大部分遠赴海外升學的同學都是第一次離開父母獨自生活或初到一個陌生的國家居住,這也難怪他們會感到憂慮和不安。對於初來埗的海外生來說,一般有3個住宿選擇:親友家中、寄宿家庭 (Homestay) 和學校宿舍 (Hall of Residence),且各有優點。筆者在這裡為大家簡單分析以上住宿安排的好處,同學可因應自己的需要來作出選擇。

親友同住 難離中國人生活圈

如同學選擇住在親友家中,當地親友就算未能免費

留學住宿 三大選擇

接待,也一定不會收取太高的費用,可以省下不少生活費。此外,大家在生活和飲食文化上也大同小異,使學生更易適應海外生活。而且家長在心理上又會特別放心,不用擔心陌生家庭會否善待自己的子女,二來也可隨時了解子女的近況。不過,常言道:「相見好,同住難。」據筆者經驗,有逾半入住親友家中的學生,也會因為未能與親友融洽相處而要求轉到其他地方居住。而且,當地親友的生活習慣和生活圈子一般也離不開中國人聚集的地方,學生會錯失了許多體驗當地生活文化的機會,英語難有大進步。

入住寄宿家庭 體驗當地生活

寄宿家庭是其中一個最多留學生選擇的住宿方案,使學生能真正體驗當地的生活文化風俗,更營造了一個全

英語的生活環境,迫使學生使用英語與寄宿家庭成員對答,英語聆聽和會話自然能進步神速,讓學生可以更快融入當地生活和學習。不過,同學需盡量適應外國生活習慣上的差異,多了解寄宿家庭的生活模式和「家規」,與寄宿家庭成員多作溝通,避免產生誤會和發生衝突。值得一提的是,同學應拋開過往在香港的生活模式,千萬不要以將香港與之作比較,並需要學習如何與寄宿家庭成員互相遷就,這才是應有態度。

選擇學校宿舍 自由度高

至於專上學院或大學的宿舍,一般就近學校,以單人房或雙人房為主,而且大部分也只能接受18歲或以上的學生入住,故自由度較高,對學生沒有太多限制,學生必須自律。宿舍亦分為包膳食的及不包膳食

兩類,大部分學生也會選擇自行煮食,既健康又划算。入住的學生大多數來自世界各地,同學可學習如何與各國的海外學生相處,了解各地的生活文化,增廣見聞,加強和擴闊人際網絡。但要留意的是,由於宿舍位有限,校方多以「先到先得」形式分配,筆者建議同學盡早申請入住宿舍,並於成功獲得宿舍後馬上繳交留位費作實。

此外,有些同學亦問及自行租屋的方案。筆者的建議是,由於初到埗的同學對當地的社區環境並不了解,在「人生路不熟」的情況下,實在不適合自行租屋,宜先在首半年入住親友家中、寄宿家庭或學校宿舍,在適應了當地的生活和社區環境後,才考慮與同學們自行租住吧。

預告:「海外留學生生活之「住」篇」將於下期(3月12日)刊登,敬希垂注。

尚學堂 (Ada Tam)
(標題和小題為編者所加)

