

SMALL DATA: Solution for the next era of AI

(สรุปจากบทสัมภาษณ์ของ Andrew Ng ในวารสาร IEEE Spectrum ฉบับวันที่ 9 กุมภาพันธ์ 2565)

ศาสตราจารย์ ดร.นุญเจริญ ศิริเนาวกุล: ประธานเจ้าหน้าที่บริหารสำนักเคเอกซ์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี

ในช่วงทศวรรษที่ผ่านมา คำว่า **Big Data** กลายเป็นคำสามัญที่นักบริหารและนักวิชาการนิยมนำมาใช้ เพื่อชี้ให้เห็นประโยชน์ของข้อมูลขนาดใหญ่ แต่ Andrew Ng ศาสตราจารย์มหาวิทยาลัย Stanford และเป็นนักธุรกิจด้าน AI ผู้ร่วมก่อตั้งบริษัทหนึ่งซึ่งเป็นที่รู้จักกันเป็นอย่างดีคือ Coursera นอกจากนี้เขาคือหนึ่งในผู้บุกเบิกงานด้าน Big Data ของโลก กำลังออกมาชี้ให้เห็นความสำคัญของ **Small Data**

AI ที่ถูกสอน (trained) ด้วยข้อมูลขนาดใหญ่หรือ Big Data เพื่อสร้างระบบที่ฉลาดเฉพาะอย่างได้ เช่น การแปลภาษานั้น เป็นการนำ Big Data ของคู่ภาษาป้อนให้กับ AI ให้มากพอจนเกิดเป็นโมเดลพื้นฐาน (Fundamental Model) ที่สามารถนำไปใช้แปลภาษาได้ ยิ่งข้อมูลของคู่ภาษาถูกป้อนเข้าไปมากเท่าไร การแปลจะยิ่งถูกต้องมากขึ้น หรือ AI ที่สามารถรู้จำหน้าคนได้ (Face Recognition) ก็ต้องอาศัยข้อมูลใบหน้าจำนวนหลายร้อยล้านภาพเพื่อสร้างโมเดลพื้นฐาน ซึ่งสร้างรายได้อย่างมหาศาลให้หลายธุรกิจที่นำระบบนี้ไปใช้



ที่ผ่านมา Big Data ทำให้ AI สามารถแก้ปัญหาในหลายด้าน เช่น การแปลภาษา การจำแนกชนิดของภาพ การเข้าใจเสียงพูด แต่สำหรับบางปัญหา เช่น การนำ VDO มาสอน AI อาจจะทำได้ยาก เนื่องจากข้อมูลของ VDO ในแต่ละช่วงเวลาจะมีจำนวนมาอย่างน้อย 24-30 ภาพต่อวินาที ทำให้ต้องใช้พลังและเวลาของคอมพิวเตอร์อย่างมหาศาล ดังนั้นการใช้ข้อมูลขนาดเล็ก หรือ Small Data ที่มีประสิทธิภาพน่าจะเป็นทางออกที่ดี

ตัวอย่างการใช้ข้อมูลขนาดเล็ก เช่น ระบบรู้จำคำเสียงที่สามารถทำงานได้ดีมากกับสถานการณ์ทั่วไป แต่ทำงานได้แย่มากเมื่อมีเสียงรบกวนเป็น background noise ปนเข้ามา แทนที่เราจะสอน AI ใหม่ (Retrain) ด้วยการนำ Big Data ที่มี background noise เป็นเสียงรบกวนเพิ่มเข้ามา ซึ่งจะทำให้เสียทั้งเวลาและทรัพยากรจำนวนมาก เราเพียงแต่นำเสียงที่มี background noise นั้นแยกออกมาสอน AI ต่างหาก แล้วทำให้มันเป็นระบบย่อย ก็จะได้ AI ตัวใหม่ที่ใช้กับการรู้จำเสียงที่มี background noise ได้

เช่นเดียวกับกับระบบตรวจจับข้อผิดพลาดสินค้าบนสายพาน โดยเฉพาะสำหรับโรงงานที่มีข้อจำกัดในการใช้ข้อมูลขนาดยักษ์ เราสามารถอาศัยวิศวกรที่ชำนาญสัก 50 คน ช่วยกันเลือกภาพหรือ VDO อย่างรอบคอบ เพื่อเป็นตัวอย่างข้อมูลใช้สอน AI เราก็สามารถใช้ประโยชน์จากระบบนี้ได้เช่นกัน

แม้ว่าจะมีการใช้ Small Data กับ AI มาแล้วกว่า 20 ปี แต่การทำให้อะไรนี้เป็นศาสตร์ทางวิศวกรรมที่เป็นระบบและเกิดความแม่นยำในการสร้างความสำเร็จยังเป็นเรื่องที่ท้าทาย เราเรียกศาสตร์ที่ศึกษาการใช้ Small Data กับ AI นี้ว่า Data-centric AI

ในระยะเวลาที่ผ่านมา การพัฒนา Algorithm ของ AI ได้ทำให้เกิดระบบที่ดีจำนวนมาก แต่นับจากนี้เป็นต้นไป นักวิจัยด้าน AI อาจจะหันมาทำงานด้าน Data-centric AI โดยเฉพาะอย่างยิ่งการหาวิธีเลือก Small Data ที่มีประสิทธิภาพมาใช้งาน

