เดลินิวส์

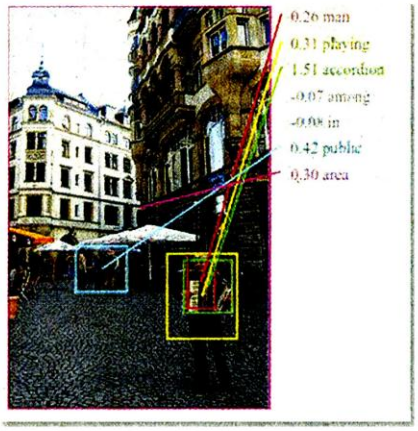
ฉบับที่ 23,925 วันพุธที่ 15 เมษายน พ.ศ. 2558 หน้า 23

รอบรู้ไอที รอบโลกเทคโนโลยี

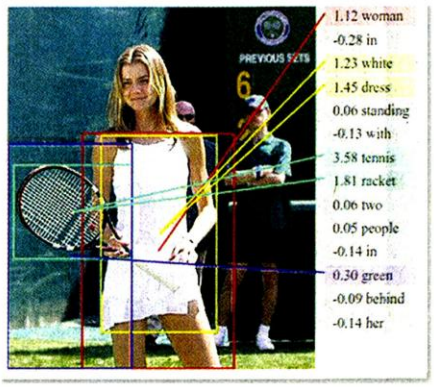
เปลี่ยนรูปภาพหนึ่งภาพให้กลายเป็นประโยค

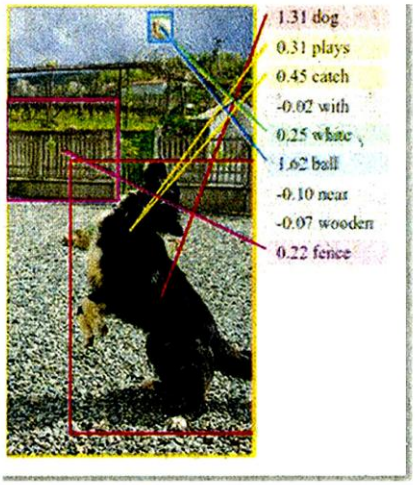
ผมเชื่อว่าในอดีตคุณผู้อ่านประจำคอลัมน์วันพุธของผมคงจะเคยได้ยินสำนวนภาษาอังกฤษที่ว่า "A picture is worth a thousand words" หรือแปลเป็นภาษาไทยได้ว่า "ภาพหนึ่งภาพแทนค่ามากกว่าคำพูดเป็นพันคำ" ซึ่งนั่นเป็นสำนวนคำพูดใช่ไหมครับ แต่คุณผู้อ่านทราบไหมครับว่า ปัจจุบันเทคโนโลยีสมัยใหม่จะทำให้รูปภาพหนึ่งภาพของพวกเราสามารถแปลงกลับไปเป็นประโยคหรือคำพูดได้จริงแล้ว และแน่นอนครับ เทคโนโลยีที่หนีไม่พ้นในการทำงานวิจัยทางด้านนี้ก็ต้องเป็นคอมพิวเตอร์วิทัศน์ (Computer Vision) ที่ผมเคยเขียนถึงบ่อย ๆ

โดยงานวิจัยนี้เป็นของมหาวิทยาลัยสแตนฟอร์ด (Stanford University) นำทีมโดย รศ.ดร.หลี่ เฟยเฟย (Li Fei-Fei) ได้พัฒนาระบบซอฟต์แวร์ที่สามารถทำความเข้าใจองค์ประกอบของภาพที่เราถ่ายรูปออกมา แล้วพิจารณาว่ามีอะไรอยู่บ้างในรูปภาพ พร้อมทั้งบรรยายรูปภาพนั้น ๆ ออกมาเป็นประโยคภาษาอังกฤษด้วยถ้อยคำที่เป็นธรรมชาติ หรือเรียกง่าย ๆ ว่าเราสามารถทำให้คอมพิวเตอร์บรรยายรูปภาพหนึ่งภาพให้ได้ออกมาเป็นประโยคหนึ่งประโยค (หรือมากกว่า) นั่นเองแหละครับ



อย่างไรก็ตามคุณผู้อ่านทราบไหมครับว่าการทำให้คอมพิวเตอร์เข้าใจรูปภาพได้เสมือนมองด้วยตามนุษย์ร้อยเปอร์เซ็นต์นั้นไม่ใช่เรื่องง่ายเลยสำหรับวิศวกรคอมพิวเตอร์ (แม้จะมีการพยายามวิจัยเรื่องนี้จากมหาวิทยาลัยชั้นนำทั่วโลกกันมานานหลายทศวรรษแล้วก็ตาม) เพราะว่าเราไม่ใช่เพียงแค่ติดตา (กล้อง) ให้กับคอมพิวเตอร์แล้วก็เสร็จ แต่เราจำเป็นต้องสอนให้คอมพิวเตอร์รู้จักคำจำกัดความของวัตถุต่างๆ ซึ่งขั้นตอนนี้ก็เปรียบได้เหมือนกับสมองของมนุษย์ ซึ่งแม้ว่าจะทำได้ แต่ทำให้ได้สมบูรณ์แบบนั้นไม่ใช่เรื่องง่ายซะทีเดียว





หนึ่งในทีมวิจัยที่พยายามจะแก้ปัญหานี้ก็คือทีมวิจัยจากกูเกิลครับ โดยเมื่อปีที่แล้วกูเกิลก็ได้ออกงานวิจัยมางานหนึ่งในงานประกวด Large Scale Visual Recognition Challenge 2014 โดยเขาได้พัฒนาโครงข่ายประสาทเทียม (Artificial Neural Network) เพื่อสร้างระบบซอฟต์แวร์เรียนรู้รูปร่างของสิ่งของหรือวัตถุต่าง ๆ เพื่อให้สามารถระบุได้ว่าสิ่งของที่ปรากฏในภาพนั้นคืออะไร อยู่ที่ตำแหน่งไหน มีลักษณะอย่างไร และผลการทดลองที่ได้ก็เรียกว่าดีในระดับหนึ่งเลยทีเดียว

อย่างไรก็ตาม แม้กูเกิลจะสามารถถอดข้อมูลเหล่านี้ออกมาจากรูปภาพได้ แต่ก็คงยังไม่อาจเรียกว่าเป็นการบรรยายภาพถ่ายได้ ดังนั้นมหาวิทยาลัยสแตนฟอร์ดจึงมีการผนวกเอางานวิจัยของกูเกิลนี้มาต่อยอด บูรณการ โดยใช้โครงข่ายประสาทเทียมเพื่อเรียนรู้วิธีการแยกแยะสิ่งต่าง ๆ ในภาพ หลังจากนั้นก็นำเอาข้อมูลที่ได้มาเรียบเรียงให้เป็นภาษาธรรมชาติมาปรับใช้งานร่วมกัน โดยทีมวิจัยนั้นทำโดยการป้อนตัวอย่างภาพถ่ายพร้อมประโยคบรรยายภาพให้โครงข่ายประสาทเทียมได้เรียนรู้ว่าการบรรยายภาพที่ดีนั้นควรเป็นอย่างไร เรียกว่าสอนให้จำหรือบางคนก็เรียกว่าการจำแบบ (Pattern Recognition) จนเมื่อป้อนข้อมูลด้วยปริมาณข้อมูลที่มากพอ สอนให้ระบบรู้แบบได้มากพอ ก็จะทำให้ระบบสามารถบรรยายภาพออกมาเป็นประโยคได้

แน่นอนครับ ว่าผลการทดลองที่ได้ก็ยังไม่สมบูรณ์แบบร้อยเปอร์เซ็นต์ซะทีเดียว ซึ่งถ้าโครงข่ายประสาทเทียมยังขาดข้อมูลตัวอย่างของการฝึกสอนสำหรับภาพถ่ายแล้ว ก็ยังมีบางภาพที่ยังบรรยายเป็นประโยคไม่ถูกต้องทั้งหมด เรียกว่าแม้ขนาดงานวิจัยจะเป็นของมหาวิทยาลัยชั้นนำของโลก ก็ยังคงมีขีดจำกัดอยู่ แต่อย่างว่าล่ะครับ เป็นเรื่องปกติของการวิจัย การพัฒนา การคิดองค์ความรู้ใหม่ เพราะสุดท้ายผมเชื่อว่าถ้าเราไม่กล้าลุกขึ้นมาคิด ลุกขึ้นมาพัฒนา ลุกขึ้นมาต่อยอดสรรค์สร้างสิ่งใหม่ ๆ อย่างสร้างสรรค์แล้ว นวัตกรรมสุดยอดของโลกศตวรรษที่ 21 ของพวกเราก็คงยากที่จะเกิดมาหล่อเลี้ยงขับเคลื่อนสังคมเทคโนโลยีสมัยใหม่ของพวกเรา หรือคุณผู้อ่านว่าจริงไหมล่ะครับ.

พศ.ดร.ชุติสันต์ เกิดวิบูลย์เวช
สถาบันบัณฑิตพัฒนบริหารศาสตร์ (นิด้า)
chutisant.ker@nida.ac.th