วารสารวิชาการสาธารณสุข
ปีที่ 32 ฉบับที่ 3 พฤษภาคม - มิถุนายน 2566

Journal of Health Science
Vol. 32 No. 3, May - June 2023

มุมวิธีวิจัย

Methodology Corner

อัลกอริทึมของการเรียนรู้ของเครื่องสำหรับงานวิจัยด้านสุขภาพ

รุจิรา อธิบาย วท.บ.*; หทัยรัตน์ โกษียาภรณ์ วท.ม.*; ชาฮีดา วิริยาทร วท.ม.*; อานนท์ คุณากรจรัสพงศ์วท.ม.; ระพีพงศ์ สุพรรณไชยมาตย์ Ph.D.*****

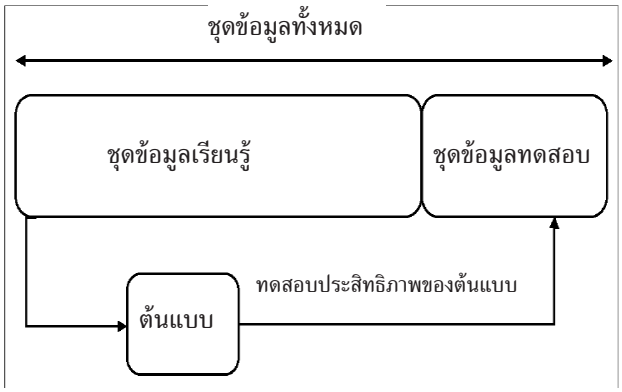
** สำนักงานพัฒนานโยบายสุขภาพระหว่างประเทศ; ** หน่วยวิจัยเพื่อขับเคลื่อนนโยบายสุขภาพ คณะแพทยศาสตร์ ศิริราชพยาบาล; *** กองระบาดวิทยา กรมควบคุมโรค*

การเรียนรู้ของเครื่อง (machine learning) เป็นหนึ่งในสาขาของปัญญาประดิษฐ์ (artificial intelligence) ซึ่งเกี่ยวกับการเรียนรู้รูปแบบของข้อมูลเพื่อการจำแนกหรือทำนายเหตุการณ์[1] สามารถนำไปประยุกต์ใช้ในหลากหลายด้านรวมถึงงานวิจัยด้านสุขภาพ เช่น การพยากรณ์โรคและการวินิจฉัยโรค[1] โดยขั้นตอนในการเรียนรู้ของเครื่องสามารถจำแนกออกเป็น 2 ประเภทหลัก ได้แก่ (1) การเรียนรู้แบบมีผู้สอน (supervised learning) และ (2) การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning)[1]

การเรียนรู้แบบมีผู้สอน (supervised learning)

การเรียนรู้แบบมีผู้สอน คือขั้นตอนการวิเคราะห์ข้อมูลโดยใช้การเรียนรู้จากชิ้นงาน (task-driven approach) ที่ประกอบไปด้วยตัวแปรนำเข้า และตัวแปรผลลัพธ์ของข้อมูลที่ถูกจำแนกแล้ว (labeled data) เพื่อนำไปแบ่งกลุ่มหรือทำนายข้อมูลอีกชุดหนึ่งที่ยังไม่เคยถูกจำแนกมาก่อน วิธีดังกล่าว จะแบ่งข้อมูลออกที่จำแนกแล้วเป็น 2 ชุด คือ (1) ชุดข้อมูลเรียนรู้ (training set) และ 2) ชุดข้อมูลทดสอบ (test set)[1] โดยอัตราส่วนที่ได้รับความนิยมในการแบ่งชุดข้อมูล คือ 80:20 ทั้งนี้ การวิเคราะห์ผ่านชุดข้อมูลเรียนรู้จะเป็นการพัฒนาต้นแบบที่นำมาใช้ในการทดสอบกับชุดข้อมูลทดสอบ[1] เพื่อให้ได้ต้นแบบที่มีประสิทธิภาพ ดังแสดงในภาพที่ 1

ภาพที่ 1 ตัวอย่างการวิเคราะห์ข้อมูลของการเรียนรู้แบบมีผู้สอน



ตัวอย่างการวิเคราะห์ข้อมูลการเรียนรู้แบบมีผู้สอน

- การจำแนกหมวดหมู่ (classification) หมายถึง การสร้างต้นแบบเพื่อทำนายค่าผลลัพธ์ที่เป็นตัวแปรจัดกลุ่ม วิธีนี้นิยมนำมาใช้เมื่อผู้วิจัยต้องการแยกข้อมูลออกจากกัน[1] เช่น พัฒนาต้นแบบการแยกคุณลักษณะเด่นของคลื่นไฟฟ้าหัวใจที่บ่งบอกถึงภาวะหัวใจเต้นปกติและผิดปกติ โดยจำแนกหมวดหมู่คลื่นไฟฟ้าหัวใจตามคุณลักษณะเด่นต่าง ๆ ซึ่งสามารถนำไปใช้พัฒนานวัตกรรมหรือระบบการดูแลสุขภาพอัจฉริยะในการวินิจฉัยและทำนายการเกิดโรคหัวใจ[2]

- การวิเคราะห์แบบถดถอย (regression) หมายถึง การสร้างต้นแบบเพื่อทำนายค่าผลลัพธ์ที่เป็นตัวแปรต่อ

เนื่อง[1] เช่น ต้นแบบการทำนายอาการผู้ป่วยโรคหลอดเลือดสมองจากระดับน้ำตาลในเลือด ความดันโลหิต อัตราการเต้นหัวใจ และอุณหภูมิร่างกายที่พบในผู้ป่วย 48 ชั่วโมงหลังเกิดอาการครั้งแรก เพื่อให้บุคลากรทางการแพทย์ใช้ข้อมูลเหล่านี้เพื่อให้มีผลลัพธ์การรักษามีประสิทธิภาพมากขึ้น[3]

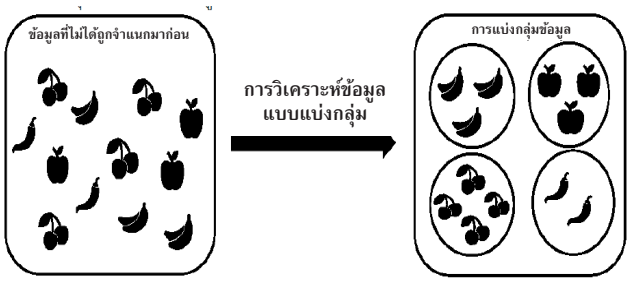
ก่อนที่จะนำต้นแบบไปใช้ จำเป็นต้องมีการพิจารณาค่าประสิทธิภาพของต้นแบบที่สร้างขึ้น โดยวิธีที่ใช้กันอย่างแพร่หลายสำหรับการจำแนกหมวดหมู่ คือ การประเมินค่าความถูกต้อง ค่าความแม่นยำ ค่า recall และค่าเอฟ (F-measure)[1] และสำหรับการวิเคราะห์แบบถดถอย ผู้วิจัยสามารถประเมินค่าความผิดพลาดจากค่า mean squared error (MSE) ค่า root mean square error (RMSE) และค่า mean absolute error (MAE)[1]

การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning)

เป็นขั้นตอนการวิเคราะห์ข้อมูลโดยใช้การเรียนรู้จากข้อมูลที่ตัวแปรผลลัพธ์ไม่ได้ถูกจำแนกมาก่อน (unlabeled data)[1] วิธีดังกล่าวจะพิจารณาความสัมพันธ์หรือคุณลักษณะพื้นฐานจากตัวแปรนำเข้าเพื่อจัดกลุ่มข้อมูลที่มีคุณลักษณะเหมือนหรือคล้ายกันไว้ด้วยกัน ตัวอย่างการวิเคราะห์ข้อมูลการเรียนรู้แบบไม่มีผู้สอนที่พบได้บ่อย คือ

- การแบ่งกลุ่มข้อมูล (clustering) ซึ่งหมายถึง การแบ่งกลุ่มจากความเหมือนและความแตกต่าง โดยชุดข้อมูลที่มีคุณลักษณะเฉพาะที่เหมือนกันมากที่สุดจะถูกจัดกลุ่มเข้าด้วยกัน และแต่ละกลุ่มจะมีลักษณะเฉพาะที่แตกต่างจากกลุ่มอื่น[1] ดังแสดงในภาพที่ 2

ภาพที่ 2 ตัวอย่างการวิเคราะห์ข้อมูลการเรียนรู้แบบไม่มีผู้สอน



ตัวอย่างเช่น การจัดกลุ่มประชากรที่มีโรคเบาหวานประเภทที่สองโดยมีอายุ เพศ ดัชนีมวลกาย เส้นรอบวงของเอว ความดันโลหิต ประวัติการเป็นโรคเบาหวานในครอบครัวเป็นตัวแปรทำนาย ซึ่งพบว่าแต่ละกลุ่มมีตัวแปรทำนายที่แตกต่างกัน ข้อมูลดังกล่าวสามารถนำมาใช้ในการออกแบบมาตรการลดปัจจัยเสี่ยงในกลุ่มผู้ป่วยโรคเบาหวานประเภทที่สองที่จำเพาะกับแต่ละกลุ่มได้[4]

จะเห็นได้ว่า การเรียนรู้ของเครื่องสามารถเข้ามามีบทบาทในงานวิจัยด้านสุขภาพ ซึ่งแตกต่างจากการวิเคราะห์ทางสถิติแบบเดิม เนื่องจากการเรียนรู้ของเครื่องมีความสามารถในการเรียนรู้จากข้อมูลและสร้างต้นแบบในการวิเคราะห์ข้อมูลได้เองแทนการใส่ต้นแบบเข้าไปโดยนักวิจัย การเรียนรู้ของเครื่องจึงเหมาะกับข้อมูลที่มีความหลากหลาย ซับซ้อน และเปลี่ยนแปลงอย่างรวดเร็ว[1] อย่างไรก็ตาม ความแม่นยำของการวิเคราะห์ข้อมูลด้วยการเรียนรู้ของเครื่องจะขึ้นอยู่กับปริมาณและคุณภาพของชุดข้อมูลที่ป้อนให้ต้นแบบเรียนรู้ และเป็นที่คาดการณ์ว่าในอนาคตจะมีการพัฒนาและใช้ประโยชน์จากการเรียนรู้ของเครื่องในบริบทด้านทางสาธารณสุขเพิ่มขึ้นในยุคการเปลี่ยนแปลงของเทคโนโลยีดิจิตัล (digital disruption)

เอกสารอ้างอิง

1. Panesar A. Machine learning and AI for healthcare. Coventry, UK: Springer; 2019.
2. Hassaballah M, Wazery YM, Ibrahim IE, Farag A. ECG heartbeat classification using machine learning and meta-heuristic optimization for smart healthcare systems. Bioengineering 2023;10(4):429.
3. Zhang Q, Xie Y, Ye P, Pang C. Acute ischaemic stroke prediction from physiological time series patterns. Australas Med J 2013;6(5):280-6.
4. Rodrigo MCL, Manuel CC, Cecilia AR, Antonio BO. Clusters of people with type 2 diabetes in the general population: unsupervised machine learning approach using national surveys in Latin America and the Caribbean. BMJ Open Diabetes Research Care 2021;9(1): e001889.