บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยเรื่องการประยุกต์ใช้เทคนิคเหมืองข้อมูลที่เหมาะสมสำหรับจำแนกความรุนแรงของการบาดเจ็บจากอุบัติเหตุการจราจรทางบก ผู้วิจัยได้นำแนวคิดและงานวิจัยที่เกี่ยวข้องมาประยุกต์ใช้ ดังนี้

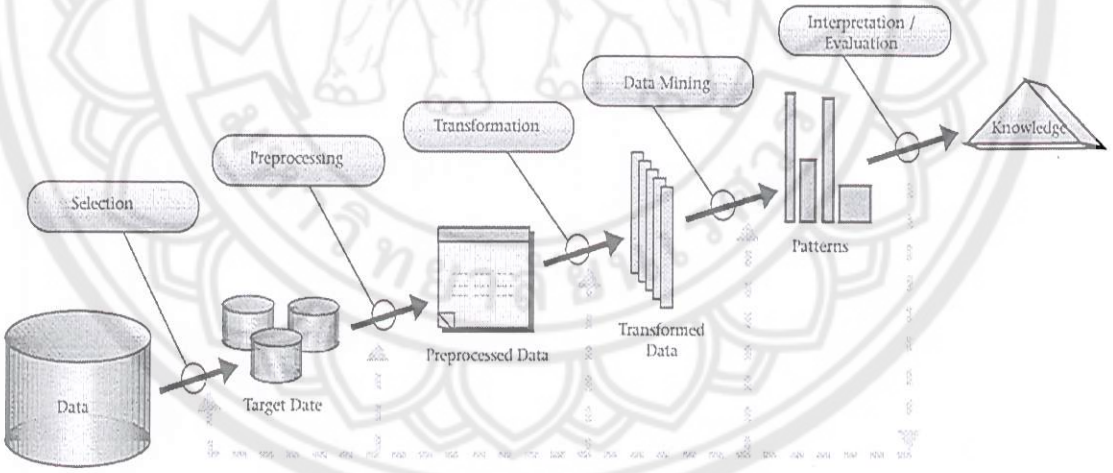
1. แนวคิดและทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล
2. แนวคิดและทฤษฎีเกี่ยวกับตัวพยากรณ์แบบต้นไม้ตัดสินใจ
3. แนวคิดและทฤษฎีเกี่ยวกับแบบจำลองแบบเบย์อย่างง่าย
4. แนวคิดและทฤษฎีเกี่ยวกับตัวพยากรณ์แบบโครงข่ายประสาทเทียม
5. แนวคิดและเทคนิคการแบ่งชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ
6. แนวคิดและทฤษฎีตัววัดประสิทธิภาพตัวพยากรณ์
7. งานวิจัยด้านอุบัติเหตุและการบาดเจ็บ

แนวคิดเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล โดยขั้นตอนในการดำเนินการต้องอาศัยวิธีการ หรือเทคนิคต่างๆ มาทำการวิเคราะห์ข้อมูลที่มีอยู่ เพื่อให้ได้ลักษณะของต้นแบบ (Modeling) ที่นำมาใช้อธิบายสภาพการณ์ที่เกิดขึ้น แล้วนำไปใช้สนับสนุนการตัดสินใจ (บุญเสริม กิจศิริกุล, 2545; การทำเหมืองข้อมูล Wikipedia, 2555; ปาลจิตต์ พันทวาที, 2552; เดอะ เพาเวอร์ สเตชั่น, 2555; Daniel T. Larose, 2005; กฤษณะ ไวยมัย, 2544; อดุลย์ ยิ้มงาม, 2555)

ไพฑูรย์ จันทร์เรือง (2550) ได้กล่าวว่า การทำเหมืองข้อมูล คือ ขบวนการทำงานในการกลั่นกรองข้อมูลจากฐานข้อมูลขนาดใหญ่เพื่อให้ได้สารสนเทศที่ยังไม่รู้ เป็นสารสนเทศที่สามารถนำไปใช้งานได้ ซึ่งเป็นสิ่งสำคัญที่ช่วยตัดสินใจในการทำธุรกิจ

การขุดค้น (Mining) เป็นขั้นตอนของการทำเหมืองข้อมูล ประสิทธิภาพของการทำเหมืองข้อมูลจะขึ้นอยู่กับปัจจัยที่เข้ามาเป็นตัวแปร ซึ่งปัจจัยที่สำคัญมีอยู่ 2 อย่าง คือ คุณภาพของชุดข้อมูล และ ความสามารถของอัลกอริทึมที่นำมาใช้ การทำเหมืองข้อมูลนั้นเป็นส่วนหนึ่งในกระบวนการการค้นหาความรู้ในฐานข้อมูล (Knowledge Discovery in Database : KDD) ซึ่งในที่นี้จะขอเรียกว่า KDD KDD นั้น เป็นวิธีที่ใช้จัดการข้อมูลที่มีขนาดใหญ่และมีความซับซ้อน เป็นกระบวนการในการค้นหาลักษณะแฝงของข้อมูลที่อยู่ในกลุ่มข้อมูลจำนวนมาก โดยการทำเหมืองข้อมูลเป็นขั้นตอนที่สำคัญในการค้นหาลักษณะที่น่าสนใจของข้อมูลเหล่านี้ การที่จะนำข้อมูลมาเข้าสู่การทำเหมืองข้อมูลนั้น ข้อมูลที่นำมาใช้จะต้องผ่านการรวบรวมและเตรียมข้อมูลที่จะใช้ โดยวิธีการเหล่านี้เป็นส่วนหนึ่งของ KDD กระบวนการของ KDD นั้นประกอบด้วยขั้นตอนต่างๆ ดังแสดงในภาพ 6 ดังนี้



ภาพ 6 กระบวนการของ KDD

ที่มา: Fayyad, et al., 1996

1. การคัดเลือกข้อมูล (Data Selection) คือ การเลือกเฉพาะข้อมูลที่ต้องการและเป็นประโยชน์ โดยการระบุถึงแหล่งข้อมูลที่จะนำมาใช้ และทำการดึงข้อมูลออกมาสำหรับการวิเคราะห์ ซึ่งจะแตกต่างไปตามวัตถุประสงค์ที่ได้กำหนดและลักษณะงานที่จะนำไปใช้ ลักษณะ

ของตัวแปรที่อยู่ในข้อมูล จะมี 2 ชนิดคือ ตัวแปรเชิงคุณภาพ (Qualitative) และ ตัวแปรเชิงปริมาณ (Quantitative)

2. การกรองข้อมูล (Data Cleaning) เป็นขั้นตอนที่แยกข้อมูลที่ไม่เกี่ยวข้อง ซ้ำซ้อนออกไป ทำให้ได้ข้อมูลที่มีคุณภาพและเป็นการแก้ไขข้อมูลให้ถูกต้องสมบูรณ์ เป็นกระบวนการที่ทำให้เกิดความมั่นใจในคุณภาพของข้อมูลที่จะนำมาวิเคราะห์

3. การแปลงรูปแบบข้อมูล (Data Transformation) ข้อมูลที่ผ่านการกรองข้อมูลแล้วจะถูกแปลงรูปแบบ หรือปรับค่าของข้อมูล โดยการลดรูปและจัดข้อมูลให้อยู่ในรูปแบบเดียวกัน เพื่อให้เหมาะสมสำหรับนำไปใช้วิเคราะห์ตามคุณสมบัติเฉพาะของแต่ละขั้นตอนวิธีที่ใช้ในการทำเหมืองข้อมูล เพื่อวิเคราะห์ตามอัลกอริทึมที่ใช้ในการทำเหมืองข้อมูลต่อไป

การจัดข้อมูลให้เป็นรูปแบบมาตรฐานเดียวกัน (Normalization) เป็นการแปลงข้อมูลโดยจะเปลี่ยนค่าข้อมูลให้อยู่ในรูปแบบมาตรฐานอันดับที่กำหนด การแปลงข้อมูลที่จะกล่าวถึงมี 2 วิธี

3.1. การแปลงค่าแบบ Min-Max Normalization ซึ่งเป็นเทคนิคที่ต้องทราบค่าสูงสุดและค่าต่ำสุดของข้อมูลที่จะทำให้ข้อมูลที่ได้นั้นอยู่ในช่วง 0 ถึง 1 ดังสมการ (1)

$$X^* = \frac{X - min}{\max - min} \quad (1)$$

จากสมการแทนค่าดังนี้

X^* คือ ค่าที่ได้จากการแปลงค่าให้อยู่ในรูปมาตรฐาน

X คือ ค่าปัจจุบันที่นำมาทำให้อยู่ในรูปมาตรฐาน

max คือ ค่าข้อมูลที่มีค่ามากที่สุดในชุดข้อมูล

min คือ ค่าข้อมูลที่มีค่าน้อยที่สุดในชุดข้อมูล

3.2. การแปลงค่าด้วยวิธีที่ทำให้อยู่ในรูปคะแนนมาตรฐาน Z (Z-scores) โดยในที่นี้ต้องทราบ ค่าเฉลี่ย (μ)และค่าส่วนเบี่ยงเบนมาตรฐาน (σ) ดังสมการ (2)

$$X^* = \frac{X - \mu}{\sigma} \quad (2)$$

ซึ่งขั้นตอนต่างๆ ที่กล่าวมาจะเป็นตัวกำหนดความถูกต้องและขอบเขตวัตถุประสงค์ที่ต้องการของฐานความรู้ได้ เนื่องจากถ้าเลือกข้อมูลไม่ดี ข้อมูลมีขยะมาปะปนมีการแปลงข้อมูลไม่ถูกต้องก็จะทำให้ฐานความรู้ที่ได้ไม่ถูกต้องตามไปด้วย (ดิษฐพล มั่นธรรม, 2553)

4. การทำเหมืองข้อมูล (Data Mining) คือ การเลือกใช้เทคนิคของเหมืองข้อมูลที่เหมาะสมสำหรับงานที่ต้องการ มาประมวลผลข้อมูลที่ผ่านการแปลงรูปแบบข้อมูลแล้วเพื่อดึงความรู้ หรือ ข้อมูลที่น่าสนใจ โดยสามารถใช้เทคนิคมากกว่าหนึ่งเทคนิคมาประมวลผลได้ ผลลัพธ์ที่ได้จากขั้นตอนนี้คือ ฐานความรู้ นอกจากนี้ประเภทของงานตามลักษณะของตัวพยากรณ์ที่ใช้ในการทำเหมืองข้อมูลสามารถแบ่งกลุ่มได้เป็น 2 ประเภทใหญ่ๆ คือ

4.1. การคาดคะเนลักษณะหรือประมาณค่าที่ชัดเจนของข้อมูลที่จะเกิดขึ้น โดยใช้พื้นฐานจากข้อมูลที่ผ่านมาในอดีต

4.2. การหาตัวพยากรณ์เพื่ออธิบายลักษณะบางอย่างของข้อมูลที่มีอยู่ โดยส่วนมากจะเป็นลักษณะการแบ่งกลุ่มให้กับข้อมูล

5. ขั้นตอนการแปลความหรือตีความ (Interpretation) คือ การนำฐานความรู้ที่ได้จากขั้นตอนเหมืองข้อมูลมาทดสอบ พิจารณาว่าถูกต้องตามความต้องการและมีประโยชน์หรือไม่ โดยอาจต้องทำการปรับแก้ค่าตัวแปรเพื่อกลับสู่ขั้นตอนเหมืองข้อมูลใหม่ หรือต้องเลือกข้อมูลชุดใหม่แม้กระทั่งทำการกรองข้อมูลใหม่อีกครั้ง หลังจากตรวจสอบความรู้จนถูกต้องแล้วจึงนำความรู้นี้ไปใช้ในการสืบค้นความรู้ การวิเคราะห์ หรือ การทำนายข้อมูลในอนาคตได้ต่อไป (Fayyad, et al, 1996; ปาลจิตต์ พันทวาที, 2552; ลลิตา ศรีชัยญา, 2553)

เทคนิคเหมืองข้อมูล

โดยทั่วไปประเภทของการทำเหมืองข้อมูลสามารถแบ่งได้เป็น 2 ประเภทใหญ่ๆ คือ

1. การสร้างตัวพยากรณ์เพื่อการทำนาย (Predictive Modeling) เป็นการคาดคะเนลักษณะหรือประมาณค่าที่ชัดเจนของข้อมูลที่จะเกิดขึ้นโดยใช้พื้นฐานจากข้อมูลที่ผ่านมาในอดีต ในที่นี้ ทุกข้อมูลจะมีคุณสมบัติหนึ่งเรียกว่าคำตอบ (Output) ซึ่งค่าของคุณสมบัตินี้จะเป็นค่าที่ใช้ในการทำนายผลของข้อมูล การหาอัลกอริทึมประเภทนี้จะมุ่งเน้นในการจำแนกข้อมูลออกเป็นกลุ่มตามค่าคุณสมบัติของคำตอบหรือค่าคำตอบนั้นๆ ซึ่งถ้าค่าคุณสมบัติของคำตอบมีค่าไม่ต่อเนื่องจะเรียกกระบวนการที่ใช้แบ่งแยกว่า การจำแนก (Classification) แต่ถ้าค่าคุณสมบัติของคำตอบมีค่าต่อเนื่องจะเรียกกระบวนการที่ใช้ทำนายคำตอบว่า การถดถอย (Regression) จัดเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นการเรียนรู้แบบมีเป้าหมายชัดเจนว่าต้องการศึกษา หรือจำแนกกลุ่มข้อมูลออกเป็นอย่างไร โดยนำเทคนิคและอัลกอริทึมต่างๆ มาใช้เพื่อจำแนกข้อมูลประมาณค่าของข้อมูล หรือ ทำนายลักษณะข้อมูลจากความสัมพันธ์ของสิ่งต่างๆ ที่เกี่ยวข้อง ซึ่งกระบวนการเรียนรู้จะพยายามหาตัวพยากรณ์ของตัวจำแนก (Classifier) ในการแบ่งข้อมูลจากชุดของตัวอย่างออกเป็นกลุ่ม เพื่อให้เกิดความผิดพลาดน้อยที่สุด เมื่อนำไปใช้กับข้อมูลจริงที่ไม่ได้

อยู่ในชุดข้อมูลตัวอย่าง โดยชุดข้อมูลตัวอย่างจำเป็นที่จะต้องทราบและได้รับการกำหนดว่าอยู่กลุ่มใดก่อนเข้ากระบวนการเรียนรู้สำหรับการจำแนกข้อมูล (Data Classification) (Witten and Eibe, 2005)

การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นการเรียนรู้เพื่อค้นหาลักษณะบางอย่างที่เหมือนกันของข้อมูลแต่ละรายการ โดยที่ไม่ทราบชัดเจนว่าข้อมูลรายการนั้นถูกจัดอยู่ในกลุ่มใด หรือมีทั้งหมดกี่กลุ่ม โดยกระบวนการเรียนรู้จะพยายามที่จะหารูปแบบความสัมพันธ์ภายในชุดข้อมูลตัวอย่างแล้วทำการวิเคราะห์ว่าควรจะจัดข้อมูลออกเป็นกี่กลุ่ม และข้อมูลใดควรจะอยู่ในกลุ่มใดแล้วทำการจัดกลุ่มข้อมูล (Clustering) (Witten and Eibe, 2005)

2. การสร้างตัวพยากรณ์เพื่อการบรรยาย (Descriptive Modeling) เป็นการหาตัวพยากรณ์เพื่ออธิบายลักษณะบางอย่างของข้อมูลที่มีอยู่ ซึ่งโดยส่วนมากจะเป็นลักษณะการแบ่งกลุ่มให้กับข้อมูล หรือการจัดกลุ่มข้อมูล (Clustering) หรืออาจเป็นการหาความสัมพันธ์ต่างๆ (Association) ซึ่งไม่ได้มีจุดมุ่งหมายเพื่อการทำนาย (บุญเสริม กิจศิริกุล, 2545; ปาลจิตต์ พันทวาที, 2552)

เทคนิคเหมืองข้อมูลยังสามารถจำแนกเป็นเทคนิคย่อยได้ 4 แบบ คือ

1. การสร้างกฎความสัมพันธ์ (Association Rule Discovery) เป็นเทคนิคเหมืองข้อมูลแบบการบรรยาย โดยหลักการทำงานคือการค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อไปวิเคราะห์ หรือทำนายปรากฏการณ์ต่างๆ ตัวอย่างของการประยุกต์ใช้ เช่น การวิเคราะห์ข้อมูลการขายสินค้า โดยเก็บข้อมูลจากระบบ ณ จุดขาย (Point Of Sale) หรือร้านค้าออนไลน์ แล้วพิจารณาสินค้าที่ผู้ซื้อมักจะซื้อพร้อมกัน เช่น ถ้าพบว่าคนที่ซื้อเทปวิดีโอมักจะซื้อเทปกาวด้วย ร้านค้าก็อาจจะจัดร้านให้สินค้าสองอย่างอยู่ใกล้กัน เพื่อเพิ่มยอดขาย หรืออาจจะพบว่าหลังจากคนซื้อหนังสือ ก แล้ว มักจะซื้อหนังสือ ข ด้วย ก็สามารถนำความรู้นี้ไปแนะนำผู้ที่กำลังจะซื้อหนังสือ ก ได้

2. การจำแนกและการพยากรณ์ (Classification and Prediction) เป็นเทคนิคเหมืองข้อมูลแบบการทำนาย ซึ่งการจำแนก (Classification) จะเป็นกระบวนการสร้างตัวพยากรณ์ (Model) จำแนกข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ โดยอาศัยลักษณะที่คล้ายคลึงกันหรือแตกต่างกันเป็นการจัดวัตถุที่เข้ามาใหม่เข้ากลุ่มที่จัดไว้แล้วให้เข้ากลุ่มได้ถูกต้อง และ การพยากรณ์ (Prediction) จะเป็นการทำนายหาค่าที่ต้องการจากข้อมูลที่มีอยู่ เป็นการหากฎเพื่อระบุประเภทของวัตถุจากคุณสมบัติของวัตถุ เช่น หาความสัมพันธ์ระหว่างผลการตรวจร่างกายต่าง ๆ กับการเกิดโรค โดยใช้ข้อมูลผู้ป่วยและการวินิจฉัยของแพทย์ที่เก็บไว้ เพื่อนำมาช่วยวินิจฉัยโรคของผู้ป่วย

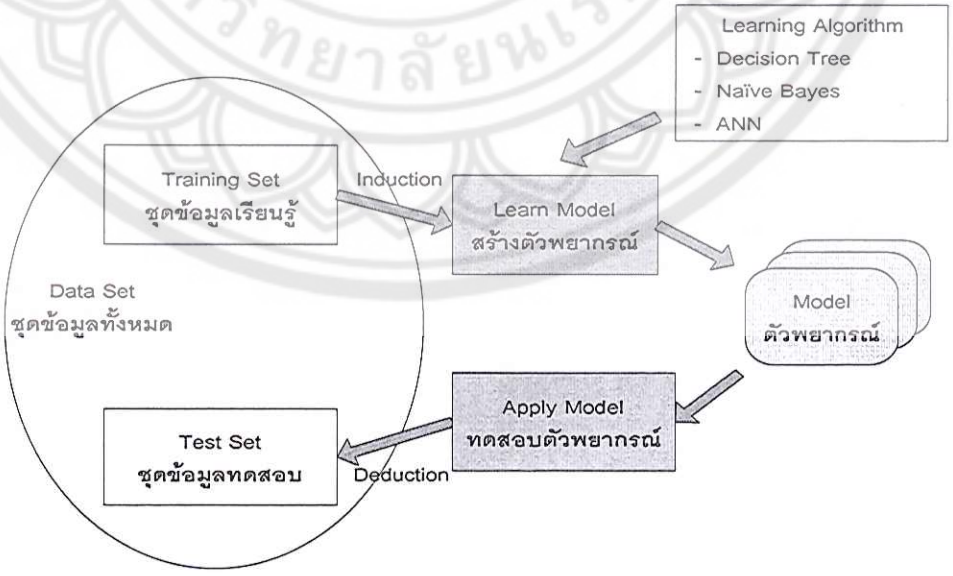
หรือการวิจัยทางการแพทย์ ในทางธุรกิจจะใช้เพื่อดูคุณสมบัติของผู้ที่จะก่อหนี้ดีหรือหนี้เสีย เพื่อประกอบการพิจารณาการอนุมัติเงินกู้

3. การจัดกลุ่มข้อมูล (Clustering) เป็นเทคนิคเหมืองข้อมูลแบบการบรรยาย เป็นเทคนิคในการลดขนาดข้อมูลด้วยการรวมกลุ่มตัวแปรที่มีลักษณะเดียวกันไว้ด้วยกัน โดยแบ่งข้อมูลที่มีลักษณะคล้ายกันออกเป็นกลุ่ม แบ่งกลุ่มผู้ป่วยที่เป็นโรคเดียวกันตามลักษณะอาการ เพื่อนำไปใช้ประโยชน์ในการวิเคราะห์หาสาเหตุของโรค โดยพิจารณาจากผู้ป่วยที่มีอาการคล้ายคลึงกัน

4. การบรรยายและจินตทัศน์ (Description and Visualization) เป็นเทคนิคเหมืองข้อมูลแบบการบรรยาย คือ การหาคำอธิบายถึงสิ่งที่จะเกิดขึ้นโดยอาศัยข้อมูลจากฐานข้อมูล และจินตทัศน์ คือ การนำเสนอข้อมูลในรูปแบบกราฟิก การนำเสนอจะสามารถทำได้มากกว่า 2 มิติ ซึ่งจะสร้างความละเอียดของการนำเสนอและสร้างความเข้าใจให้มากขึ้น ซึ่งเราอาจพบข้อมูลที่ซ่อนเร้นเมื่อดูข้อมูลชุดนั้นด้วยจินตทัศน์ (การทำเหมืองข้อมูล Wikipedia, 2555; อดิศักดิ์ พงษ์พูลผลศักดิ์ และพิณรัตน์ นุชโพธิ์, 2550; เชษฐา จิรไพศาลกุล, 2550)

ในงานวิจัยนี้ได้เลือกใช้เทคนิคเหมืองข้อมูลประเภทการจำแนก (Classification) เพื่อจำแนกระดับความรุนแรงของการบาดเจ็บจากอุบัติเหตุการจราจรทางบก

หลักการจำแนกข้อมูล (Classification) คือ การนำข้อมูลส่วนหนึ่งมาสอนระบบให้สามารถเรียนรู้ (Training) เพื่อจำแนกข้อมูลนั้นออกเป็นกลุ่มตามที่กำหนดไว้ ผลที่ได้จากการเรียนรู้ คือ ตัวจำแนกกลุ่มข้อมูล (Classifier) เมื่อมีข้อมูลที่ต้องการจำแนกกลุ่มเข้ามา ตัวจำแนกจะสามารถทำนายกลุ่มของข้อมูลที่ควรจะเป็นได้



ภาพ 7 หลักการจำแนกข้อมูล

ดังภาพ 7 ข้อมูลทั้งหมด จะถูกแบ่งเป็น 2 ชุด ได้แก่ ชุดข้อมูลเรียนรู้ (Training Set) และชุดข้อมูลทดสอบ (Test Set) โดยจะนำชุดข้อมูลเรียนรู้ไปทำการเรียนรู้เพื่อสร้างตัวพยากรณ์ตามเทคนิคเหมืองข้อมูลที่เลือกใช้ แล้วนำตัวพยากรณ์ที่ได้มาทดสอบโดยใช้ชุดข้อมูลทดสอบ เพื่อประเมินความแม่นยำของตัวพยากรณ์ที่ได้

โดยกลไกการพัฒนาตัวพยากรณ์เพื่อการจำแนกจะประกอบด้วย 4 ส่วนหลัก คือ

1. คัดเลือกลักษณะเฉพาะเพื่อนำมาพัฒนาตัวพยากรณ์
2. คัดเลือกข้อมูลเพื่อนำมาเป็นชุดข้อมูลเรียนรู้ และ ชุดข้อมูลทดสอบ
3. กระประยุกต์ใช้เทคนิคเหมืองข้อมูล
4. การประเมินความแม่นยำของตัวพยากรณ์

โดยงานวิจัยนี้สนใจศึกษาเทคนิคการจำแนกข้อมูลดังต่อไปนี้ เทคนิคการทำต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม และนาอีฟเบเซียน หรือ เบย์อย่างง่าย

แนวคิดและทฤษฎีเกี่ยวกับตัวพยากรณ์แบบต้นไม้ตัดสินใจ

เทคนิคต้นไม้ตัดสินใจ

ต้นไม้ตัดสินใจ (Decision Tree) เป็นเครื่องมือที่ช่วยกำหนดขอบเขตของปัญหาและช่วยสร้างทางเลือกที่เป็นไปได้ในการแก้ปัญหา (Quinlan, 1993) ซึ่งต้นไม้ตัดสินใจนั้นเป็นการนำข้อมูลมาสร้างตัวพยากรณ์ การพยากรณ์ในรูปของต้นไม้ตัดสินใจนั้นมีการทำงานแบบการเรียนรู้แบบมีผู้สอน (Supervised Learning) คือ การสร้างตัวพยากรณ์การจำแนกได้จากกลุ่มตัวอย่างของข้อมูลที่ได้กำหนดไว้ก่อนล่วงหน้า หรือที่เรียกว่า ชุดข้อมูลเรียนรู้ (Training Set) และนำตัวพยากรณ์ที่ได้ไปใช้กับข้อมูลทดสอบ (Test Set) เพื่อทดสอบความแม่นยำของตัวพยากรณ์ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจำแนกได้ ในกระบวนการเหมืองข้อมูลใช้ในการพยากรณ์ หรือการจำแนก แสดงแผนผังต้นไม้ตัดสินใจในรูปแบบของโหนด และ แสดงผลลัพธ์จากการกระทำหรือตัดสินใจในเงื่อนไขต่างๆ แล้วเชื่อมต่อกันเป็นเส้นที่แตกแขนงออกไป ต้นไม้ตัดสินใจเป็นวิธีที่ได้รับความนิยมเนื่องจากความไม่ซับซ้อนของอัลกอริทึม สามารถตีความและเข้าใจลักษณะของรูปแบบข้อมูลได้ง่าย (ชลนิศา สาระ, 2550)

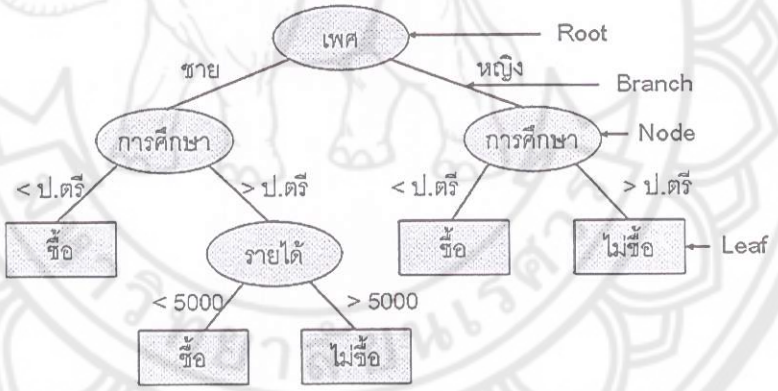
คุณลักษณะของต้นไม้ตัดสินใจ

คุณลักษณะของต้นไม้ตัดสินใจจะสามารถแสดงการเชื่อมโยงความสัมพันธ์ของปัญหาได้ชัดเจน และช่วยจัดการกับสถานการณ์ที่ซับซ้อนต่างๆ ให้อยู่ในรูปแบบที่กระชับขึ้น เพื่อช่วยให้เห็นภาพของปัญหาได้ชัดเจนขึ้น นอกจากนี้ยังมีโครงสร้างที่สามารถบอกผลลัพธ์ที่อาจเกิดขึ้นจากการ

เลือกทางเลือกต่างๆ สำหรับการตัดสินใจ และช่วยวิเคราะห์ลำดับการตัดสินใจแก้ปัญหาพร้อมทั้งวิเคราะห์ผลลัพธ์จากการตัดสินใจด้วยแนวทางต่างๆ

โครงสร้างต้นไม้

โครงสร้างต้นไม้จะประกอบด้วย ส่วนโหนด (Node) คือลักษณะเฉพาะต่างๆ ส่วนกิ่ง (branch) ของต้นไม้ คือค่าของลักษณะเฉพาะในแต่ละโหนด ส่วนใบ (Leaf) คือผลลัพธ์ที่ได้จากการจำแนก ขั้นตอนในการสร้างต้นไม้ตัดสินใจจะเริ่มจากการนำลักษณะเฉพาะแต่ละตัวมาคำนวนหาค่าเกณฑ์ ซึ่งโดยมากจะใช้เกณฑ์เอ็นโทรปี (Entropy) แล้วคำนวณหาค่า Information Gain เมื่อเริ่มต้น ลักษณะเฉพาะที่มีค่า Information Gain มากที่สุด จะได้รับเลือกเป็นโหนดราก (Root) และในรอบถัดมา โหนดชั้นถัดลงมาก็จะได้จากลักษณะเฉพาะที่มีค่า Information Gain มากที่สุด ซึ่งกลไกจะทำวนในลักษณะนี้จนกว่าจะพบเงื่อนไขการหยุดการแตกโหนด (บดินทร์ มิลินทางกูร และสุปียา เจริญศิริวัฒน์, 2553; Mittechell, 1997) โดยต้นไม้ที่ได้ในแต่ละจุดจะมีลักษณะจำเพาะเพื่อใช้ในการตรวจสอบการจำแนกข้อมูล ดังแสดงในภาพ 8



ภาพ 8 โครงสร้างต้นไม้ตัดสินใจ

การสร้างต้นไม้จะเริ่มสร้างจากบนลงล่าง โดยเริ่มจากโหนดราก (Root Node) คือ โหนดตัดสินใจแรก (Decision Node) การเลือกโหนดที่แตกต่างกันไปจะทำให้โครงสร้างต้นไม้เปลี่ยนไปตามเงื่อนไขที่เลือก

ขั้นตอนการสร้างต้นไม้

ต้นไม้ตัดสินใจนั้นสามารถสร้างได้จากอัลกอริทึมที่แตกต่างกันได้หลายวิธี เช่น กรีดีอัลกอริทึม หรือ ขั้นตอนเชิงละโมบ (Greedy Algorithm) ซึ่งเป็นการสร้างต้นไม้จากบนลงล่างแบบวนซ้ำด้วยวิธีการแบ่งปัญหาใหญ่เป็นปัญหาย่อย โดยส่วนใหญ่จะใช้กับข้อมูลที่เป็นแบบต่อเนื่อง

และจะมีตัวแปรตามหรือผลลัพธ์ได้ 1 ตัวแปรต่อตัวพยากรณ์ โดยส่วนมากขั้นตอนเชิงละโมบจะนิยมใช้ค่าสัมประสิทธิ์จีนี (GINI) ในการตัดสินใจเลือกคุณสมบัติของโหนด (พิจิตรา จอมศรี , 2549) นอกจากนี้ยังมีที่อัลกอริทึม ID3 ที่เป็นอัลกอริทึมที่นิยมใช้และเป็นพื้นฐานของการสร้างต้นไม้ตัดสินใจอื่นๆ อย่างเช่น C4.5 ซึ่งใช้ค่าเกณฑ์เอ็นโทรปี (Quinlan, 1993)

การตัดเล็มกิ่งต้นไม้ (Pruning)

เป็นขั้นตอนการตัดโหนดในต้นไม้ตัดสินใจเพื่อลดความซ้ำซ้อนและการรู้จำจากข้อมูลฝึกสอนมากเกินไป ทำให้สามารถจำแนกประเภทข้อมูลได้เร็วขึ้น แบ่งเป็น 2 รูปแบบ ดังนี้

1. การตัดเล็มก่อนการสร้างต้นไม้เสร็จเต็มรูป (Pre-Pruning) เป็นขั้นตอนการหยุดสร้างโหนดเพิ่ม เมื่อพบว่าคลาสที่พบในข้อมูลนี้เป็นคลาสประเภทเดียวกันหมด หรือพบว่าค่าของลักษณะเฉพาะนั้นมีค่าเดียวกันทั้งหมด

2. การตัดเล็มหลังการสร้างต้นไม้เสร็จเต็มรูป (Post-Pruning) เป็นขั้นตอนการตัดเล็มโหนดจากล่างขึ้นบน โดยเลือกต้นไม้ย่อยที่มีความลึกมากที่สุดแล้วทำการตัดต้นไม้ย่อยที่เลือกออก จากนั้นแทนต้นไม้ย่อยด้วยโหนดของใบซึ่งการเลือกคลาสที่ใส่ในโหนดใบนั้นได้มาจากการหาคลาสที่ปรากฏอยู่ในต้นไม้ย่อยที่ตัดทิ้งไปที่มีค่ามากที่สุด

ต้นไม้ตัดสินใจจะแสดงผลลัพธ์ในรูปโครงสร้างต้นไม้ ซึ่งสามารถแปลงเป็นกฎที่เข้าใจได้ง่าย โดยที่แต่ละโหนดของต้นไม้แสดงถึงลักษณะเฉพาะ แต่ละกิ่งแสดงถึงเงื่อนไขในการทดสอบ และ โหนดใบแสดงถึงประเภทข้อมูลที่กำหนดไว้ โดยลักษณะของโครงสร้างต้นไม้ที่มีขนาดเล็ก ไม่ซับซ้อน จะเป็นต้นไม้ตัดสินใจที่ดีที่สุด เหมาะสำหรับการนำไปใช้ในการตัดสินใจ (ลลิตา ศรีชัยญา, 2553)

อัลกอริทึมที่ใช้ในการสร้างต้นไม้ตัดสินใจ

อัลกอริทึม C4.5 เป็นอัลกอริทึมที่พัฒนาโดย J. Ross Quinlan (1993) โดยนำเอาอัลกอริทึม ID3 มาปรับปรุงให้มีความสามารถมากขึ้น ขณะที่อัลกอริทึม ID3 ได้ใช้ Entropy และ Gain ในการนำมาสร้างต้นไม้ตัดสินใจ โดยค่าน้อยสุดของ Gain จะเป็นค่าที่ใช้เป็นลักษณะเฉพาะในการแบ่งชุดข้อมูล ขณะที่ C4.5 ใช้วิธีการ Information Gain ซึ่งเพิ่มเติมการจัดการกับข้อมูลตัวเลข ข้อมูลที่ขาดไปและไม่สมบูรณ์ (Missing Value, Noisy Data) นำมาใช้ในการตัดต้นไม้ และการ Prune ด้วยการแทนกิ่ง (Branch) ที่ไม่ช่วยในการตัดสินใจด้วยโหนดใบ (Leaf Node) ที่ตัดสินใจได้ดีกว่า (ณัฐมณฑ์ สิริวัฒนานันท์, 2551) โดยใช้ค่าอัตราส่วนเกน (Gain Ratio) ซึ่งจะใช้เป็นตัวจำแนกในการตัดสินใจเลือกลักษณะเฉพาะที่จะใช้เป็นรากหรือโหนด ซึ่งในการทำงานขั้นตอนแรกคล้ายกับการทำงานด้วย ID3 คือต้องหาค่าสารสนเทศของข้อมูล คือ ค่า Gain จึงหาค่าสารสนเทศของการแบ่งหรือจำแนกข้อมูล (Split Information) ตามลักษณะของลักษณะเฉพาะแต่

ละตัวเพื่อนำมาคำนวณค่าอัตราส่วนเกน ซึ่งจะใช้เป็นหลักเกณฑ์ในการตัดสินใจสำหรับอัลกอริทึม C4.5 และถ้ายังไม่สามารถจัดกลุ่มข้อมูลให้เป็นกลุ่มเดียวกันทั้งหมด จะต้องสร้างต้นไม้ตัดสินใจต่อไป โดยพิจารณาเลือกแอททริบิวต์ที่จะมาเป็นโหนดต่อจากโหนดราก กระบวนการสร้างต้นไม้ตัดสินใจจะสิ้นสุดลงเมื่อโหนดใบเป็นกลุ่มของข้อมูลเดียวกันทั้งหมด (ดิษฐพล มั่นธรรม, 2553)

การคำนวณค่าคาดคะเนของข้อมูลตามหลักเกณฑ์ Entropy (Info(T)) ของชุดข้อมูล T แสดงใน (1).

$$Info(T) = -\sum_{i=1}^{k}((freq(C_i,T)/|T|) \cdot log_2(freq(C_i,T)/|T|)) \quad (1)$$

$freq$ คือ ความถี่, C_i คือ ค่าของลักษณะเฉพาะ

ค่าคาดคะเนของข้อมูล ที่แยกโดยการใช้ลักษณะประจำ X ซึ่งกำหนด X คือลักษณะประจำที่แบ่งออกเป็น n ตัว

$$info_x = \sum_{i=1}^{n}((|T_i|/|T|) \cdot info(T_i)) \quad (2)$$

$$Gain(X) = Info(T) - info_x(T) \quad (3)$$

$$Split - info(X) = -\sum_{i=1}^{n}((|T_i|/|T|)log_2(|T_i|/|T|)) \quad (4)$$

$$Gain - ratio(X) = gain(X)/Split - info(X) \quad (5)$$

ประสิทธิภาพของต้นไม้ตัดสินใจไม่ได้อยู่ที่การสร้างต้นไม้ตัดสินใจเพื่อให้สามารถจำแนกชุดข้อมูลเรียนรู้ได้อย่างถูกต้องเท่านั้น แต่ต้องสามารถจำแนกข้อมูลจากตัวอย่างใหม่ๆ ที่นอกเหนือจากนั้นได้อย่างถูกต้องด้วย ดังนั้นการสร้างต้นไม้ตัดสินใจจึงต้องมีชุดข้อมูลทดสอบที่จะใช้ตรวจสอบความถูกต้องของต้นไม้ตัดสินใจด้วย (DANIEL T. LAROSE,2006; M. Kantardzic, 2011).

ข้อจำกัดของต้นไม้ตัดสินใจ

1. การจำแนกกลุ่มแบบต้นไม้ตัดสินใจในกรณีเป็นข้อมูลที่มีค่าต่อเนื่อง เช่น ข้อมูลรายได้ ข้อมูลราคา ต้องทำการแปลงให้อยู่ในช่วงหรือตัดเป็นกลุ่มก่อน
2. เมื่ออัลกอริทึมเลือกว่าจะใช้ค่าไหนเป็นตัวจำแนกแล้วก็จะไม่สนใจค่าอื่นที่อาจมีความสำคัญเช่นกัน
3. การจัดการกับข้อมูลที่ไม่ทราบค่า อาจมีผลกระทบกับผลของต้นไม้ตัดสินใจ

4. ต้นไม้ที่มีระดับชั้นมากเกินไป จะทำให้ข้อมูลที่ผ่านโหนดแตกออกเป็นชิ้นเล็กชิ้นน้อย ซึ่งข้อมูลเหล่านั้นจะไม่มีประโยชน์ในการทำมาใช้ทำการวิเคราะห์

5. ปัญหาเรื่อง ความเฉพาะเจาะจงเกินไป (Overfitting) / การฝึกมากเกิน (Overtraining) เกิดจากการที่ตัวพยากรณ์ได้เรียนรู้เข้าไปถึงรายละเอียดของข้อมูลมากเกินไป จะทำให้เกิดโหนดที่เป็นส่วนเฉพาะเจาะจงกับกลุ่มข้อมูลที่ใช้ในการเรียนรู้ ซึ่งต้องหาวิธีการในการตัดกิ่งนี้ออกไป (เชษฐา จิรไพศาลกุล, 2550; นฤพนต์ ว่องประชานุกูล, 2548)

แนวคิดและทฤษฎีเกี่ยวกับแบบจำลองแบบเบย์อย่างง่าย

ตัวแบบจำลองแบบเบย์อย่างง่าย

ตัวจำแนกแบบนาอีฟเบเซียน หรือ เบย์อย่างง่าย (Naive Bayesian Classification) เป็นขั้นตอนวิธีหนึ่งที่มีประสิทธิภาพมากในการจำแนกข้อมูล เหมาะกับกรณีของเซตตัวอย่างมีจำนวนมากและคุณสมบัติของตัวอย่างไม่ขึ้นต่อกัน โดยการเรียนรู้ปัญหาที่เกิดขึ้นเพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลด้วยการใช้ทฤษฎีความน่าจะเป็นแบบมีเงื่อนไข

ทฤษฎีความน่าจะเป็นแบบมีเงื่อนไข ถ้าเหตุการณ์ A และ B เป็นเหตุการณ์ใดๆ ที่ไม่เป็นอิสระต่อกัน เราสามารถหาความน่าจะเป็นของเหตุการณ์ A เมื่อทราบเหตุการณ์ B ที่เกิดขึ้นแล้ว เรียกกว่า ความน่าจะเป็นแบบมีเงื่อนไข เขียนแทนด้วย $P(A|B)$ ซึ่งเกิดจากความน่าจะเป็นของเหตุการณ์ A และเหตุการณ์ B ที่เกิดขึ้นด้วยกัน เขียนแทนด้วย $P(A,B)$ หารด้วยความน่าจะเป็นของเหตุการณ์ B เขียนแทนด้วย $P(B)$ ดังสมการ (6)

$$P(A|B) = \frac{P(A,B)}{P(B)} \text{ เมื่อ } P(B) > 0 \qquad (6)$$

การแบ่งกลุ่มแบบนาอีฟเบเซียน (Naive Bayesian Classifier) เป็นการคำนวณหาความน่าจะเป็นของแต่ละกลุ่มข้อมูล ซึ่งในที่นี้เรียกว่า คลาส (Class) เมื่อกำหนดลักษณะประจำ (Attribute) และค่าของลักษณะประจำแต่ละตัวที่จะใช้ในการทำนาย ซึ่งเป็นการคำนวณหาค่าความน่าจะเป็นของแต่ละคำตอบที่เป็นไปได้ในคลาสแล้วทำการเปรียบเทียบกัน ค่าความน่าจะเป็นที่สูงที่สุดของคลาสใดๆ จะเป็นผลของการทำนายเพียงค่าเดียว โดยถือว่าลักษณะประจำแต่ละตัวมีความเป็นอิสระต่อกัน (Class Conditional Independence) ซึ่งมีการทำงานดังนี้

1. ให้ X เป็นเซตของเหตุการณ์หนึ่งๆ จะได้ $X = \{x_1, x_2, \dots, x_n\}$ เมื่อกำหนดให้ x_i คือค่าของ X ในลักษณะประจำที่ $A_1, A_2, \dots, A_n$ ตามลำดับ

2. สมมติให้มีคลาสที่เป็นไปได้จำนวน m คลาส คือ $C_1, C_2, \ldots, C_m$ สำหรับเหตุการณ์ X โดยการทำนายจะเลือกค่าความน่าจะเป็นของคลาสที่สูงที่สุดที่ทำให้เกิดเหตุการณ์ X มาเป็นผลของการทำนาย

$$P(C_i|X) > P(C_j|X) \tag{7}$$

สำหรับ $1 \leq j \leq m, j \neq i$

จากทฤษฎีของเบย์

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \tag{8}$$

3. เมื่อ $P(X)$ เป็นค่าคงที่สำหรับทุกคลาส ดังนั้น การพิจารณาจึงเลือกจากค่า $P(X|C_i)P(C_i)$ ที่สูงที่สุด โดย $P(C_i) = S_i/S$ เมื่อS_i คือจำนวนข้อมูลที่มีค่าคลาส C_i และ S คือจำนวนข้อมูลทั้งหมด

4. เนื่องจากนาอีฟเบเซียนมีเงื่อนไขว่าลักษณะประจำแต่ละตัวเป็นอิสระต่อกัน (Independent) หรือ ไม่มีความสัมพันธ์กัน ดังนั้น

$$P(X|C_i) = \prod_{k=1}^{n} P(X_k|C_i) \tag{9}$$

ค่าความน่าจะเป็น $P(x_1|C_i), P(x_2|C_i), \ldots, P(x_n|C_i)$ สามารถคำนวณได้จาก ลักษณะประจำของข้อมูลที่นำมาทำการเรียนรู้ โดยแบ่งเป็น 2 กรณี คือ

ถ้า A_k เป็นข้อมูลเชิงคุณภาพ $P(x_k|C_i) = \frac{S_{ik}}{S_i}$ เมื่อ S_{ik} คือจำนวนข้อมูลที่มีค่าคลาส C_i ซึ่งมีค่า x_k สำหรับลักษณะประจำที่ A_k และ S_i คือจำนวนข้อมูลที่มีค่าคลาส C_i

ถ้า A_k เป็นข้อมูลเชิงปริมาณ จะได้

$$P(X_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) \tag{10}$$

$$P(X_k|C_i) = \frac{1}{\sqrt{2\pi}\sigma_{C_i}} e^{-\frac{(x_k-\mu_{C_i})^2}{2\sigma_{C_i}^2}} \tag{11}$$

เมื่อ $g(x_k, \mu_{C_i}, \sigma_{C_i})$ คือ ฟังก์ชันเกาส์ (Gaussian Function) สำหรับลักษณะประจำ A_k μ_{C_i}และ σ_{C_i} คือ ค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐานของลักษณะประจำ A_k ที่มีค่าคลาส C_i ตามลำดับ

5. ในการทำนายคลาสของเหตุการณ์ X จะทำการคำนวณ $P(X|C_i)\,P(C_i)$ สำหรับคลาสทุกคลาส และเหตุการณ์ X จะถูกจัดอยู่ในคลาสที่มีค่า $P(X|C_i)\,P(C_i)$ สูงสุด (พยุง มีสัจ และณัฏฐา ห่อประชุม, 2548; สิทธิโชค มุกดาสกุลภิบาล, 2551)

แนวคิดและทฤษฎีเกี่ยวกับตัวพยากรณ์แบบโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Neural Network)

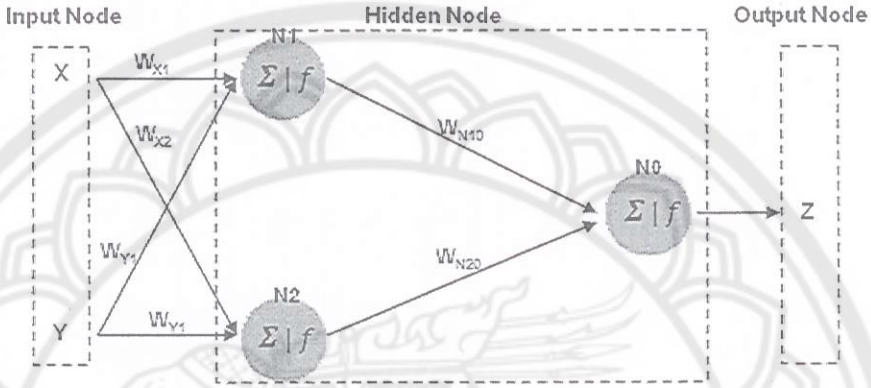
โครงข่ายประสาทเทียม หรือ นิวรอลเน็ต คือ โมเดลทางคณิตศาสตร์ หรือ โมเดลทางคอมพิวเตอร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (Connectionist) เป็นเทคโนโลยีที่มาจากงานวิจัยด้านปัญญาประดิษฐ์ (Artificial Intelligence : AI) เพื่อใช้ในการคำนวณค่าฟังก์ชันจากกลุ่มข้อมูล วิธีการของนิวรอลเน็ต (Artificial Neural Networks หรือ ANN) เป็นวิธีการที่ให้เครื่องเรียนรู้จากตัวอย่างต้นแบบ แล้วฝึก (Train) ให้ระบบได้รู้จักที่จะคิดแก้ปัญหาที่กว้างขึ้นได้ แนวคิดเริ่มต้นของเทคนิคนี้ได้มาจากการศึกษาโครงข่ายไฟฟ้าชีวภาพ (Bioelectric Network) ในสมอง ซึ่งประกอบด้วยเซลล์ประสาท (Neurons) และ จุดประสานประสาท (Synapses) ตามโมเดลนี้ ข่ายงานประสาทเกิดจากการเชื่อมต่อระหว่างเซลล์ประสาท จนเป็นเครือข่ายที่ทำงานร่วมกัน

โครงสร้างของนิวรอลเน็ตจะประกอบด้วยโหนดสำหรับ ข้อมูลนำเข้า (Input Value) และผลลัพธ์ (Output Value) การประมวลผลจะกระจายอยู่ในโครงสร้างเป็นชั้น ๆ ได้แก่ ชั้นข้อมูลเข้า (Input Layer) ชั้นข้อมูลซ่อน (Hidden Layers) และ ชั้นข้อมูลออก (Output Layer) มีการกำหนดค่าน้ำหนัก (Weight) ให้แก่เส้นทางนำเข้าของข้อมูลนำเข้าแต่ละตัว การประมวลผลของนิวรอลเน็ตจะอาศัยการส่งการทำงานผ่านโหนดต่าง ๆ ในชั้น (Layer) เหล่านี้ ในการเรียนรู้ของโครงข่ายประสาทเทียม ทำได้หลายวิธีแต่ในที่นี้จะขอกล่าววิธีอัลกอริทึมการแพร่ย้อนกลับ (Back-propagation Algorithm) ซึ่งเป็นวิธีที่นิยมนำมาใช้กับโครงข่ายประสาทเทียมแบบเพอร์เซ็พตรอน ในการสร้างการเรียนรู้สำหรับโครงข่ายประสาทเทียม เพื่อให้มีความคิดเสมือนมนุษย์

อัลกอริทึมการแพร่ย้อนกลับถูกนำเสนอโดย David E. Rumelhart, et al. (1986) หลักการนี้สามารถนำไปใช้แก้ปัญหาในลักษณะที่เป็นเชิงเส้น และ ปัญหาที่ไม่เป็นเชิงเส้นได้ ซึ่งเป็นที่นิยมมากที่สุดจนมาถึงปัจจุบัน การแพร่ย้อนกลับมาพื้นฐานมาจากกฎเดลต้า (Delta Rule)

กฎเดลต้า เป็นกฎที่ใช้ในการฝึกของเพอร์เซ็พตรอน (Perceptron) ต่อมาได้นำไปใช้ในการฝึกให้กับโครงข่ายประสาทเทียมที่มีลักษณะหลายชั้น โดยกฎของเดลต้าจะถูกใช้ในการปรับค่าน้ำหนักของเพอร์เซ็พตรอน

ชั้นเครือข่าย (Network Layer) เป็นหนึ่งในสถาปัตยกรรมของโครงข่ายประสาทเทียม โดยที่ชั้นเครือข่ายจะประกอบด้วย 3 ชั้น ได้แก่ โหนดข้อมูลเข้า (Input Node) โหนดข้อมูลซ่อน (Hidden Node) และ โหนดข้อมูลออก (Output Node)



ภาพ 9 โครงสร้างของชั้นเครือข่าย (Network Layer)

จากภาพ 9

X และ Y คือ ข้อมูลนำเข้า

Z คือ ผลลัพธ์

W_{ij} คือ น้ำหนักที่กำหนดให้เส้นทางนำเข้าข้อมูล

N_i คือ โหนด

$\sum$ คือ การรวมค่าที่ได้จากการคำนวณ

f คือ ฟังก์ชันที่ใช้เพื่อคำนวณค่า

การทำงานของโหนดข้อมูลเข้า จะทำหน้าที่แทนส่วนของข้อมูลดิบ ที่จะถูกป้อนเข้าสู่เครือข่าย

การทำงานของแต่ละโหนดข้อมูลซ่อนจะถูกกำหนดโดยการทำงานของโหนดข้อมูลเข้า และค่าน้ำหนักบนความสัมพันธ์ระหว่าง โหนดข้อมูลเข้า และ โหนดข้อมูลซ่อน

พฤติกรรมการทำงานของโหนดข้อมูลออก จะขึ้นอยู่กับการทำงานของโหนดข้อมูลซ่อน และ ค่าน้ำหนักระหว่างโหนดข้อมูลซ่อน และ โหนดข้อมูลออก

ประเภทของเครือข่ายนี้เราสามารถกำหนดการแทนค่าให้แก่ โหนดข้อมูลเข้าได้อย่างอิสระ ค่าน้ำหนักระหว่าง โหนดข้อมูลเข้า และ โหนดข้อมูลซ่อนจะถูกกำหนดเมื่อโหนดข้อมูลซ่อน

กำลังทำงาน ฉะนั้นเวลาที่แก้ไขค่าน้ำหนักโหนดข้อมูลซ่อนจะสามารถเลือกว่าอะไรคือค่าที่เราแทนเข้ามา

สามารถจำแนกออกเป็น 2 ประเภท คือ โครงข่ายแบบชั้นเดียว (Single-layer perceptron) และโครงข่ายแบบหลายชั้น (Multi-layer perceptron : MLPs)

1. โครงข่ายแบบชั้นเดียว (Single-layer perceptron) เป็นโครงข่ายประสาทเทียมอย่างง่ายที่มีเพียงชั้นรับข้อมูลป้อนเข้า และชั้นส่งข้อมูลออกเท่านั้น โหนดในชั้นรับข้อมูลป้อนเข้าทำหน้าที่รับข้อมูลเข้า แล้วส่งข้อมูลผ่านเส้นเชื่อมโยงต่างๆ ไปให้โหนดในชั้นส่งข้อมูลออก ความเข้มของสัญญาณ หรือ ปริมาณข้อมูลที่นำเข้าสู่โหนดในชั้นส่งข้อมูลออกจะขึ้นอยู่กับค่าน้ำหนักที่อยู่บนเส้นเชื่อมโยง

โหนดในชั้นส่งข้อมูลออกจะนำข้อมูลที่ได้รับมาคำนวณโดยใช้ฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันการแปลง (Transfer function) ที่เหมาะสมกับปัญหา แล้วส่งผลลัพธ์ที่ได้ออกมาเป็นข้อมูลส่งออก ถ้าผลลัพธ์ที่ต้องการเป็น "ใช่" หรือ "ไม่ใช่" เราจะต้องใช้ Threshold function

$$f(x) = \begin{cases} 1 \ if \ x \geq T \\ 0 \ if \ x < T \end{cases}, \ T = Threshold \ level \qquad (12)$$

หรือ ผลลัพธ์เป็นค่าตัวเลขที่ต่อเนื่อง เราต้องใช้ Continuous function เช่น Sigmoid function

$$f(x) = \frac{1}{1+e^{-\alpha x}} \qquad (13)$$

2. โครงข่ายแบบหลายชั้น (Multi-layer perceptron) เป็นโครงข่ายที่มีชั้นแอบแฝง (hidden layer) ตั้งแต่ 1 ชั้นขึ้นไป โครงข่ายแบบหลายชั้นจะใช้ในกรณีที่ปัญหามีความซับซ้อน ซึ่งโครงข่ายแบบชั้นเดียวไม่สามารถแก้ปัญหาได้ จึงเพิ่มจำนวนโหนดที่มีการคำนวณ หรือชั้นแอบแฝงให้กับโครงข่าย ตัวอย่างของโครงข่ายแบบหลายชั้น เช่น การแพร่ย้อนกลับ (Back propagation), เซลฟ์ออร์แกนไนซิ่งแมปช์ (Self organizing maps) และ เคาน์เตอร์พรอพะเกชัน (Counter propagation) เป็นต้น (โครงข่ายประสาทเทียม Wikipedia, 2555; ธัญรัตน์ สิทธิพล, 2552; Artificial Neural Network โครงข่ายประสาทเทียม, 2555; ธนาวุฒิ ประกอบผล, 2552; J. Scheetz, 2009)

แนวคิดและทฤษฎีเทคนิคการแบ่งชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ

วิธีการแบ่งชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ

1. การสุ่มเลือกข้อมูล (Sampling Data) จะทำการสุ่มเลือกข้อมูลแบบอิสระตามสัดส่วนของข้อมูลที่ต้องการ เช่น สุ่มเลือกข้อมูล 80% ของข้อมูลทั้งหมดเพื่อนำไปสร้างตัวพยากรณ์ หรือ เป็นชุดข้อมูลเรียนรู้ และนำข้อมูลที่เหลือจากการสุ่มเลือกข้อมูลมาใช้ทดสอบตัวพยากรณ์ หรือ เป็นชุดข้อมูลทดสอบ

2. k – folds Cross-Validation คือ การแบ่งชุดข้อมูลออกเป็น k กลุ่ม (k – Folds) แล้วทำการแบ่งชุดข้อมูล 1 กลุ่มไว้เพื่อทำการทดสอบ โดยนำชุดข้อมูลที่เหลือไปสร้างตัวพยากรณ์ จากนั้นเปลี่ยนชุดข้อมูลทดสอบเป็นกลุ่มอื่นแทนข้อมูลกลุ่มเดิม แล้วทำเหมือนเดิมจนครบ k รอบ แล้วนำเอาค่าความแม่นยำของการพยากรณ์ที่ได้มาหาค่าเฉลี่ย โดยวิธีนี้ข้อมูลทั้งหมดจะได้เป็นทั้งชุดข้อมูลเรียนรู้และชุดข้อมูลทดสอบ

แนวคิดและทฤษฎีตัววัดประสิทธิภาพตัวพยากรณ์

วิธีการประเมินความแม่นยำของตัวจำแนก

การวัดประสิทธิภาพตัวพยากรณ์ในกลไกทางเหมืองข้อมูลประกอบไปด้วยค่าต่างๆ ดังนี้

1. ค่าความถูกต้องในการจำแนกกลุ่ม (Accuracy)
2. ค่าความระลึก (Recall)
3. ค่าความแม่นยำ (Precision)
4. ค่าความเหวี่ยง (F-Measure)
5. ค่าผลบวกจริง (True Positive Rate)
6. ค่าผลบวกเท็จ (False Positive Rate)

ตาราง Confusion Matrix คือ คือการประเมินผลลัพธ์การทำนาย (หรือผลลัพธ์จากโปรแกรม) เปรียบเทียบกับผลลัพธ์จริงๆ ที่หาโดยคน

ตัวอย่าง การวัดประสิทธิภาพแบบผลลัพธ์ 3 ค่า คือ ค่าระดับ 1 ค่าระดับ 2 ค่าระดับ 3 แสดงในตาราง 1

ตาราง 1 ตัวอย่างการแสดงค่าความแม่นยำในการทำนายในรูปตาราง Confusion Matrix

		ค่าพยากรณ์			
	ระดับ	1	2	3	รวม
ค่าจริง	1	100	115	3	218
	2	44	1,083	56	1,183
	3	7	308	86	401
	รวม	151	1,506	145	1,802

เมื่อแทนค่าจะได้ความหมายดังนี้

100 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 1 ผลการจำแนกเป็นเท่ากับระดับ 1

115 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 1 ผลการจำแนกเป็นเท่ากับระดับ 2

3 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 1 ผลการจำแนกเป็นเท่ากับระดับ 3

44 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 2 ผลการจำแนกเป็นเท่ากับระดับ 1

1,083 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 2 ผลการจำแนกเป็นเท่ากับระดับ 2

56 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 2 ผลการจำแนกเป็นเท่ากับระดับ 3

7 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 3 ผลการจำแนกเป็นเท่ากับระดับ 1

308 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 3 ผลการจำแนกเป็นเท่ากับระดับ 2

86 คือ จำนวนผลผลิตที่มีจำนวนอยู่ในระดับที่ 3 ผลการจำแนกเป็นเท่ากับระดับ 3

จากนั้นนำผลลัพธ์ที่ได้จากตาราง 1 มาคำนวณเพื่อหาค่าความถูกต้องในการจำแนกกลุ่ม ค่าความระลึก ค่าความแม่นยำ ค่าความเหวี่ยง ค่าผลบวกจริง และค่าผลบวกเท็จ ดังนี้

ค่าความถูกต้องในการจำแนกกลุ่ม คือ ผลที่ได้จากการใช้อัลกอริทึมในการหาค่าความถูกต้องในการจำแนกคำตอบระดับผลผลิต ดังสมการ (14)

$$\text{ค่าความถูกต้องในการจำแนกกลุ่ม} = \frac{\text{จำนวนข้อมูลที่จำแนกได้ถูกต้อง}}{\text{จำนวนข้อมูลทั้งหมด}} \qquad (14)$$

$$Accuracy = \frac{100+1083+86}{151+1506+145} = 0.70$$

ค่าความระลึก หรือ ค่าความครบถ้วน (Recall) แสดงให้เห็นว่าตัวพยากรณ์ที่ได้มามีความแม่นยำเพียงใด คำนวณจากค่าของข้อมูลที่ผลลัพธ์ถูกต้อง โดยพิจารณาจากข้อมูลของ

ผลลัพธ์เดียวกัน อัลกอริทึมที่ให้ค่าความระลึก หรือ ค่าผลบวกจริงที่มีค่าสูงกว่า จะเป็นอัลกอริทึมที่ดีกว่า ดังสมการ (15)

$$\text{ค่าความระลึก} = \frac{\text{จำนวนข้อมูลที่จำแนกได้ถูกต้องของกลุ่มนั้น}}{\text{จำนวนข้อมูลที่อยู่ในกลุ่มนั้นจริง}} \quad (15)$$

ตัวอย่าง ค่าความระลึก ของระดับ 1 คือ

$$r = \frac{100}{100 + 115 + 3} = 0.45$$

ตัวอย่าง ค่าความระลึก ของระดับ 2 คือ

$$r = \frac{1083}{44 + 1083 + 56} = 0.91$$

ตัวอย่าง ค่าความระลึก ของระดับ 3 คือ

$$r = \frac{86}{7 + 308 + 86} = 0.21$$

ค่าความระลึกที่ได้มามีค่ายิ่งมากยิ่งดี แต่ค่าที่ได้จะไม่เกิน 1

ค่าความแม่นยำ (Precision) แสดงให้เห็นว่าตัวพยากรณ์ที่ได้มามีความแม่นยำเพียงใด คำนวณจากค่าของข้อมูลที่ผลลัพธ์ถูกต้องโดยพิจารณาจากจำนวนข้อมูลทั้งหมดที่ถูกจำแนกมีผลลัพธ์เดียวกัน ดังสมการ (16)

$$\text{ค่าความแม่นยำ} = \frac{\text{จำนวนข้อมูลที่จำแนกได้ถูกต้องของกลุ่มนั้น}}{\text{จำนวนข้อมูลที่ถูกจำแนกว่าเป็นกลุ่มนั้นทั้งหมด}} \quad (16)$$

ตัวอย่าง ค่าความแม่นยำของระดับ 1 คือ

$$p = \frac{100}{100 + 44 + 7} = 0.66$$

ตัวอย่าง ค่าความแม่นยำของระดับ 2 คือ

$$p = \frac{1083}{115 + 1083 + 308} = 0.71$$

ตัวอย่าง ค่าความแม่นยำของระดับ 3 คือ

$$p = \frac{86}{3 + 56 + 86} = 0.59$$

ค่าความแม่นยำที่ได้มามีค่ายิ่งมากยิ่งดี แต่ค่าที่ได้จะไม่เกิน 1

ค่าความเหวี่ยง (F – Measure) คือ ค่าเฉลี่ยที่ให้ความสำคัญกับค่าความแม่นยำและค่าความระลึกเท่าๆ กัน โดยจะคำนวณจากค่าความแม่นยำ (p) และ ค่าความระลึก (r) และนำมาหาค่าเฉลี่ยระหว่างค่าทั้งสองดังสมการ (17)

$$F(r,p) = \frac{2pr}{p+r} \tag{17}$$

ตัวอย่าง ค่าความเหวี่ยง ของระดับ 1 คือ

$$F(r,p) = \frac{2(0.662)(0.459)}{0.662 + 0.459} = 0.542$$

ตัวอย่าง ค่าความเหวี่ยง ของระดับ 2 คือ

$$F(r,p) = \frac{2(0.719)(0.915)}{0.719 + 0.915} = 0.806$$

ตัวอย่าง ค่าความเหวี่ยง ของระดับ 3 คือ

$$F(r,p) = \frac{2(0.593)(0.214)}{0.593 + 0.214} = 0.315$$

ค่าผลบวกจริง (True Positive Rate: TP Rate) หมายถึง ผลการทดสอบที่ให้ผลบวกแสดงว่าเป็นจำนวนคำตอบระดับผลผลิตที่ตอบตรงค่าจริง เทียบกับจำนวนค่าคำตอบระดับผลผลิตที่ตอบค่าจริงทั้งหมดแสดงค่าและอัลกอริทึมที่ให้ค่าของผลบวกจริงสูงกว่าจะเป็นอัลกอริทึมที่ดีกว่า ดังสมการ (15) ค่าผลบวกจริง และ ค่าความระลึก มีวิธีการคิดเหมือนกัน

ค่าผลบวกเท็จ (False Positive Rate: FP Rate) หรือ 1 – ค่าความจำเพาะ หมายถึง ผลการทดสอบที่ให้ผลบวก คือ จำนวนคำตอบระดับผลผลิตไม่ได้ตอบค่าจริง เทียบกับจำนวนคำตอบระดับผลผลิตไม่ได้ตอบค่าจริงทั้งหมด แสดงค่าและอัลกอริทึมที่ให้ค่าของผลบวกเท็จต่ำกว่าจะเป็นอัลกอริทึมที่ดีกว่า ดังสมการ (19)

$$\text{ค่าผลบวกเท็จ} = \frac{\text{จำนวนคำตอบระดับผลผลิตที่ไม่ได้ตอบค่าจริง}}{\text{จำนวนคำตอบระดับผลผลิตที่ไม่ได้ตอบค่าจริงทั้งหมด}} \tag{19}$$

ตัวอย่าง ค่าผลบวกเท็จ ของระดับ 1 คือ

$$FPRate = \frac{(44+7)}{((44+1083+56)+(7+308+86))} = 0.032$$

ตัวอย่าง ค่าผลบวกเท็จ ของระดับ 2 คือ

$$FPRate = \frac{(115 + 308)}{((100 + 115 + 3) + (7 + 308 + 86))} = 0.683$$

ตัวอย่าง ค่าผลบวกเท็จ ของระดับ 3 คือ

$$FPRate = \frac{(3 + 56)}{((100 + 115 + 3) + (44 + 1083 + 56))} = 0.042$$

จากการศึกษาผู้วิจัยได้นำวิธีการวัดประสิทธิภาพตัวพยากรณ์มาใช้เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของอัลกอริทึม (สิทธิโชค มุกดาสกุลภิบาล, 2551; ณพงศ์ วาณิชยพงศ์, 2553; รังสิพัฒน์ ยงยุทธวิชัย, 2555)

งานวิจัยด้านอุบัติเหตุและการบาดเจ็บ

อดิศักดิ์ พงษ์พูลผลศักดิ์ และพิณรัตน์ นุชโพธิ์ (2550) กล่าวว่า การตรวจสอบอุบัติเหตุเป็นการยากที่จะระบุว่าปัจจัยใดเป็นสาเหตุที่ก่อให้เกิดอุบัติเหตุ เนื่องจากการเกิดอุบัติเหตุไม่ได้มีสาเหตุมาจากปัจจัยใดปัจจัยหนึ่งเพียงอย่างเดียว แต่เกิดจากความสัมพันธ์ที่เกี่ยวเนื่องกันจากหลายๆ ปัจจัย จึงจำเป็นต้องพิจารณารายละเอียดของสาเหตุการเกิดอุบัติเหตุ จึงได้วิเคราะห์หาสาเหตุของอุบัติเหตุ โดยใช้เทคนิคการสร้างกฎความสัมพันธ์ ซึ่งเป็นเทคนิคหนึ่งในการทำเหมืองข้อมูล เพื่อทำนายโอกาสที่จะเกิดอุบัติเหตุจราจรทางถนน และนำไปใช้ในการดำเนินการแก้ไขเพื่อลดการเกิดอุบัติเหตุจราจรและความรุนแรงของอุบัติเหตุบนท้องถนน โดยพบว่าสาเหตุของการเกิดอุบัติเหตุจะขึ้นนอยู่กับ 4 ปัจจัย คือ คน ยานพาหนะ ถนน และสิ่งแวดล้อม และพบว่าระดับความรุนแรงแต่ละระดับนั้น เพศชายมีโอกาสเกิดอุบัติเหตุมากกว่าเพศหญิง และความรุนแรงระดับบาดเจ็บสาหัสมีโอกาสเกิดได้มากที่สุด

Mohamed A. Abdel-Aty and Hassan T. Abdelwahab (2004) ได้ประยุกต์ใช้โครงข่ายประสาทเทียมเพื่อใช้วิเคราะห์ความรุนแรงของการบาดเจ็บจากอุบัติเหตุ โดยได้ดัดแปลงโครงข่ายประสาทเทียมเป็น 2 รูปแบบ คือ Multilayer preception (MLP) และ Fuzzy adaptive resonance theory (ART) และเพิ่มวิธีการ Order Probit model แล้ววัดประสิทธิภาพของการพยากรณ์ความรุนแรงของการบาดเจ็บโดยใช้เงื่อนไขจากข้อมูลอุบัติเหตุในรัฐฟลอริดา (Florida) ประเทศสหรัฐอเมริกา ผลลัพธ์ที่ได้จากการทดสอบเปรียบเทียบแสดงให้เห็นถึงความแม่นยำในการพยากรณ์จาก MLP 73.5% , fuzzy ARTMAP 70.6% และ Order Probit model 61.7%

กวี เกื้อเกษมบุญ (2545) กล่าวว่า ปัจจัยที่เกี่ยวข้องกับการเกิดอุบัติเหตุสามารถจำแนกได้เป็น 4 กลุ่ม คือ คน ยานพาหนะ ถนน และสิ่งแวดล้อม พบว่าโดยรวมแล้วคนเป็นปัจจัยที่เกี่ยวข้องกับการเกิดอุบัติเหตุจราจรทางถนนมากที่สุด คือ คิดเป็นร้อยละ 95.62 รองลงมาได้แก่ ยานพาหนะ คิดเป็นร้อยละ 27.54 และถนนกับสิ่งแวดล้อม ตามลำดับ โดยปัจจัยที่มีผลโดยตรงต่อระดับความรุนแรงของอุบัติเหตุจราจรทางถนนมากที่สุด คือ พฤติกรรมการใช้รถใช้ถนน ซึ่งพฤติกรรมที่พบ คือ ไม่คาดเข็มขัดนิรภัย ขับรถเร็ว ไม่สวมหมวกนิรภัย ไม่ให้สัญญาณไฟในขณะจอด ชะลอ หรือเลี้ยว และขับรถตามคันอื่นในระยะกระชั้นชิด ตามลำดับ ขณะเดียวกันยังมีปัจจัยอีก 4 ปัจจัย ที่ส่งผลทางอ้อมต่อระดับความรุนแรงของอุบัติเหตุจราจรผ่านปัจจัยพฤติกรรมการใช้รถใช้ถนน คือ อุปกรณ์ควบคุมการจราจร สภาวะทางกาย สภาพยานพาหนะ และ สภาพแสงสว่าง ตามลำดับ ผลการวิจัยพบว่าปัจจัยข้างต้นมีลักษณะที่คล้ายกันมากเมื่อดูจากช่วงเวลาทั้งในและนอกเทศกาล และกรณีอุบัติเหตุที่เกิดขึ้นเนื่องจากผู้ขับขี่ดื่มเครื่องดื่มแอลกอฮอล์ พบว่าการดื่มเครื่องดื่มแอลกอฮอล์ และพฤติกรรมการใช้รถใช้ถนน มีผลโดยตรงต่อระดับความรุนแรงมากที่สุด