บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีที่มีเนื้อหาเกี่ยวข้องกับทฤษฎีในวิทยานิพนธ์ฉบับนี้มีดังนี้ ต้นไม้ตัดสินใจ, การรวมการจำแนกแบ็กกิง และการรวมการจำแนกด้วยวิธีการลงคะแนนแบบทั่วไปโดยตัวอย่างเรียนรู้ การเตรียมข้อมูลทดสอบสำหรับต้นไม้แต่ละต้น ส่วนงานวิจัยที่เกี่ยวข้องได้แก่ งานวิจัยอัลกอริทึมเพื่อนบ้านที่ใกล้ที่สุด *k* ตัว งานวิจัยที่ใช้กฎที่ใกล้ที่สุดและการคำนวณหาค่าความใกล้เคียงของกฎ

2.1 **ทฤษฎีที่เกี่ยวข้อง** (Theory)

ทฤษฎีที่เกี่ยวข้องกับวิทยานิพนธ์ฉบับนี้ประกอบไปด้วยต้นไม้ตัดสินใจและการรวมการจำแนกที่แบ่งออกได้เป็นการรวมต้นไม้ตัดสินใจวิธีแบ็กกิง, การจำแนกวิธีบูสต์และ การรวมการจำแนกด้วยการลงคะแนนแบบทั่วไปโดยตัวอย่างเรียนรู้

2.1.1 ต้นไม้ตัดสินใจ (Decision Tree)

ต้นไม้ตัดสินใจเป็นหนึ่งในวิธีที่ใช้อย่างกว้างขวาง (Quinlan, 1993) ในการหาคำตอบจากข้อมูลที่ถือได้ว่าเป็นวิธีการประเมินคำตอบที่ทำงานได้ดีแม้มีข้อมูลรบกวนมาก อัลกอริทึมในการสร้างต้นไม้ตัดสินใจได้แก่ การแบ่งแยกและการถดถอยของต้นไม้ (Classification and Regression Trees, CART) การนำของต้นไม้ตัดสินใจ (Induction of Decision Trees, ID3) (Quinlan, 1986, pp 81-106) และ ซี 4.5 (Quinlan, 1993) โดยวิธีการสร้างต้นไม้ตัดสินใจเหล่านี้มุ่งไปที่การสร้างต้นไม้ที่มีขนาดเล็ก ที่ให้ผลดีกว่าหรือเท่ากับต้นไม้ที่มีขนาดใหญ่

รูปแบบของต้นไม้ตัดสินใจอยู่ในรูปของการคาดเดาและการอธิบาย เหตุผลที่เรียกว่าต้นไม้ตัดสินใจเนื่องจากรูปแบบของผลลัพธ์ที่ได้อยู่ในรูปของแผนภูมิต้นไม้ที่ง่ายต่อการเรียนรู้และการทำความเข้าใจ ดังนั้นต้นไม้ตัดสินใจจึงเป็นที่นิยมในการนำมาประยุกต์ใช้เพราะสามารถแสดงความสัมพันธ์ของข้อมูลและผลลัพธ์อย่างชัดเจน ยิ่งไปกว่านั้นต้นไม้ตัดสินใจยังสามารถนำมาแปลงเป็นกฎต่างๆ (Quinlan, 1986, pp 81-106) ได้ เช่น ถ้าอากาศ ร้อน และอุณภูมิ สูง และความชื้น สูง และลม ไม่แรง ผลลัพธ์คือ ไม่เล่น

การแทนต้นไม้ตัสินใจ (Decision Tree Representation)

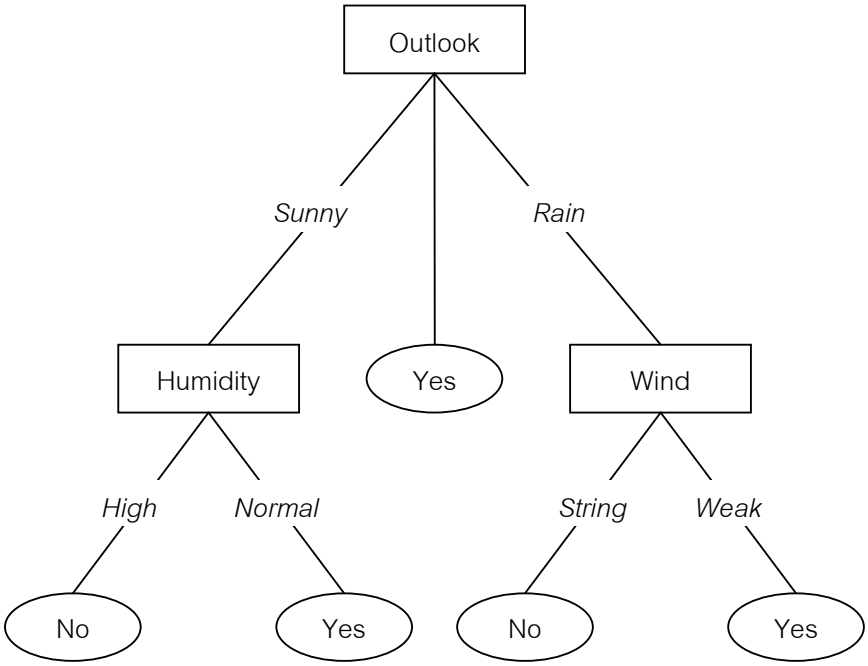
ต้นไม้ตัดสินใจจะแสดงข้อมูลเรียงจากส่วนรากของต้นไม้ตัดสินใจไปยังส่วนของใบ (Leaf Node) ในลักษณะของต้นไม้คว่ำ การคาดเดาผลลัพธ์เริ่มจากรากของต้นย้ายไปตามกิ่งและใบของต้นเรื่อยๆ ด้วยการตรวจสอบคุณลักษณะ (Attribute, แต่ละบัพภายในของต้นไม้, ใบ) และค่าคุณลักษณะ (Attribute Value, แต่ละกิ่งของต้นไม้สัมพันธ์กับค่าคุณลักษณะ, กิ่ง) จนถึงบัพสุดท้ายที่ให้ค่าผลลัพธ์ของต้นไม้ตัดสินใจ

ตารางที่ 2.1 แสดงตัวอย่างข้อมูลการเล่นกีฬาเทนนิสซึ่งประกอบไปด้วยคุณลักษณะ 5 คุณลักษณะคือ สภาพอากาศ (Outlook), อุณภูมิ (Temperature), ความชื้น (Humidity) และสภาพลม (Wind) โดยมีผลลัพธ์ของข้อมูลคือเล่นเทนนิสหรือไม่เล่นเทนนิส (Play Tennis)

ตารางที่ 2.1 : ตัวอย่างข้อมูลการเล่นเทนนิส

Outlook	Temperature	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Hot	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Normal	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

ภาพที่ 2.1 : ตัวอย่างต้นไม้ตัดสินใจของการทำนายเล่นเทนนิสหรือไม่



อัลกอริทึมพื้นฐานของต้นไม้ตัดสินใจ

อัลกอริทึมในการสร้างต้นไม้ตัดสินใจมีหลายอัลกอริทึมด้วยกัน ส่วนมากเป็นการทำงานแบบบนลงล่าง (Top-down) เช่น การแบ่งแยกและการถดถอยของต้นไม้, การนำของต้นไม้ตัดสินใจและ ซี 4.5 ทั้งนี้วิทยานิพนธ์ฉบับนี้ใช้อัลกอริทึมของ ซี 4.5 ในการทำงาน

ปัญหาที่เหมาะสมกับต้นไม้ตัดสินใจ

แม้ว่าอัลกอริทึมในการสร้างต้นไม้ตัดสินใจมีหลายวิธีที่ถูกสร้างเพื่อให้สามารถรองรับการทำงานและความต้องที่แตกต่างกัน แต่โดยส่วนมากต้นไม้ตัดสินใจเหมาะกับการใช้งานการตัดสินใจที่มีคุณสมบัติดังนี้

1 ตัวอย่างของปัญหาถูกอธิบายโดย คู่ของคุณลักษณะและค่าคุณลักษณะ เช่น คุณลักษณะ ท้องฟ้า (Outlook) และค่าคุณลักษณะ มีแดด (Sunny), มีเมฆมาก (Overcast), ฝนตก (Rain) เป็นต้น

2 ผลลัพธ์ที่ต้องการเป็นค่าที่ไม่ต่อเนื่อง

3 อาจจำเป็นต้องใช้สมมติฐานแบบเลือก (Disjunctive Hypothesis)

4 ข้อมูลในการเรียนรู้มีสิ่งรบกวน (Noise) ได้ เนื่องจากต้นไม้ตัดสินใจทนต่อสิ่งรบกวนที่ทำให้เกิดข้อผิดพลาดที่เกิดจากคุณลักษณะและคลาสของตัวอย่าง

5 ข้อมูลในการเรียนรู้อาจไม่มีค่าคุณลักษณะได้

คาร์ต (CART)

คาร์ตมาจากคำว่า Classification and Regression Trees วิธีการสร้างต้นไม้ตัดสินใจอยู่ในรูปแบบของการแตกข้อมูลออกเป็นสองส่วนจากบัพพ่อ และแต่ละบัพจะมีบัพลูกเพียงสองบัพเท่านั้น ดังนั้นคำตอบของคำถามมีเพียงสองคำตอบคือ ใช่หรือไม่ใช่ การสร้างต้นไม้แบบนี้จะสร้างโดยวิธีนี้ไปเรื่อยๆ ดังนั้นบัพลูกของวิธีคาร์ตจะมีลักษณะที่เหมือนๆ กัน จากนั้นได้รับการพัฒนาให้สามารถรองรับบัพได้มากขึ้น โดยมีวิธีการสร้างต้นไม้ตัดสินใจอื่นๆ ที่ใช้วิธีการแตกบัพที่หลากหลายที่แตกข้อมูลได้เร็ว โดยวิธีคาร์ตได้ปรับข้อมูลเพื่อให้เหมาะสมดังนี้

1 กรณีที่ไม่มีค่าข้อมูล (Missing Value) ให้แทนตัวแปรในการแตกค่าแทนที่

2 กรณีที่คาดเดาผิดให้ปรับค่าเพื่อช่วยหลีกเลี่ยงค่าความผิดพลาด

3 หยุดสร้างต้นไม้ตัดสินใจเมื่อ

1) มีข้อมูลอย่างเดียวในแต่ละบัพลูก

2) ข้อมูลทั้งหมดในแต่ละบัพลูกมีการกระจายของการคาดเดาเหมือนกัน

3) จำนวนระดับของต้นไม้ที่มากที่สุดถูกให้ค่าโดยผู้ใช้ได้ทำให้ต้นไม้ที่สร้างไม่สามารถใช้กับข้อมูลจริงได้ เพราะต้นไม้ที่ได้จะขึ้นกับลักษณะเฉพาะของชุดข้อมูล

ไอดีทรี (ID3)

ไอดีทรีเป็นหนึ่งในอัลกอริทึมต้นไม้ตัดสินใจที่ถูกนำมาใช้อย่างกว้างขวาง อัลกอริทึมนี้สร้างต้นไม้ตัดสินใจที่มีขนาดเล็กกว่าต้นไม้ตัดสินใจเดิมที่มีขนาดใหญ่ ด้วยการเลือกคุณลักษณะที่ดีที่สุดในการจำแนกตัวอย่างจากนั้นใช้ค่าการเพิ่มสารสนเทศ (Information Gain) ช่วยในการวัดค่าคุณลักษณะที่ดีที่สุด โดยเริ่มต้นอธิบายการวัดที่ใช้ในทฤษฎีสารสนเทศ (Information Theory) ที่เรียกว่าเอนโทรปี (Entropy) อธิบายลักษณะกลุ่มของตัวอย่างที่ไม่บริสุทธิ์ โดยกำหนดให้

S คือ กลุ่มของตัวอย่างที่บรรจุทั้งตัวอย่างบวกและตัวอย่างลบ

$Entropy(S)$ คือ จำนวนบิตที่ใช้ในการอธิบายเหตุการณ์ที่ต้องการเข้ารหัสคลาสตัวอย่างบวกและตัวอย่างลบ

p_+ คือ สัดส่วนตัวอย่างบวกใน S

p_- คือ สัดส่วนตัวอย่างลบใน S

$$Entropy(S) \equiv -p_+ \log_2 p_+ - p_- \log_2 p_-$$

โดยค่า $Entropy(S)$ จะมีค่าเท่ากับ 0 ก็ต่อเมื่อสมาชิกทั้งหมดใน S มีคลาสเหมือนกันทั้งหมด ดังนั้นถ้าตัวอย่างทั้งหมดเป็นตัวอย่างบวกจะได้ว่า

$$Entropy(S) \equiv -1\log_2(1) - 0\log_2(0)$$

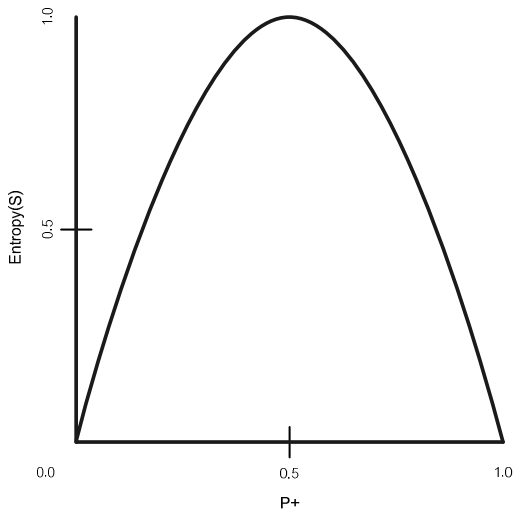
จากตาราง 2.1 มีตัวอย่างทั้งหมด 14 ตัวอย่าง เป็นตัวอย่างบวก 9 ตัวอย่างและตัวอย่างลบ 5 ตัวอย่างดังนั้น ค่าเอนโทรปีมีค่าดังนี้

$$Entropy(S) = \left(-\frac{9}{14}\right)\log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right)\log_2\left(\frac{5}{14}\right)$$

$$= 0.940 \; bits$$

กรณีที่เป็นตัวอย่างบวกทั้งหมดหรือตัวอย่างลบทั้งหมดค่า $Entropy(S)$ มีค่าเท่ากับ 1 ดังนั้นสรุปได้ว่าค่า $Entropy$ จะมีค่าตั้งแต่ 0 ถึง 1 ดังภาพที่ 2.2

ภาพที่ 2.2 : ความสัมพันธ์ของค่าเอนโทรปี



ค่าการเพิ่มสารสนเทศ (Information Gain)

เมื่อค่าเอนโทรปีแสดงการปนเปื้อนของตัวอย่างเรียนรู้ เราใช้ค่าการเพิ่มสารสนเทศอธิบายการวัดของประสิทธิภาพของคุณลักษณะในข้อมูลเรียนรู้เพื่อให้ทราบว่าคุณลักษณะใดดีกว่า โดยให้ $Gain(S,A)$ คือค่าการเพิ่มสารสนเทศของค่าคุณลักษณะ A

$$Gain(S,A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

or

$$Gain(S,A) \equiv Entropy(S) - E(A)$$

$$E(A) \equiv \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

กำหนดให้

S คือตัวอย่างทั้งหมด

$Values(A)$ คือกลุ่มของค่าที่เป็นไปได้ทั้งหมดสำหรับค่าคุณลักษณะ A

S_v คือ กลุ่มย่อยของ S ของค่าคุณลักษณะ A

A มีค่า v เช่น $S_v = s \in S | A(s) = v$

จากสมการ $Gain(S,A)$ ในเทอมแรกเป็นค่าเอนโทรปีของกลุ่มตัวอย่างเริ่มต้นทั้งหมด ในส่วนเทอมที่สองคือ ค่าที่คาดหวังของเอนโทรปีหลังจากที่ S ถูกแบ่งโดยใช้ค่าคุณลักษณะ A ยกตัวอย่างเช่น การหาค่าเพิ่มสารสนเทศของคุณลักษณะ Wind ในตาราง 2.1 ประกอบไปด้วยตัวอย่างบวก 9 ตัวอย่าง ตัวอย่างลบ 5 ตัวอย่าง และค่าคุณลักษณะของคุณลักษณะ Wind มี 2 ค่าคือ Weak และ Strong ดังนั้นค่าเพิ่มสารสนเทศของค่าคุณลักษณะมีค่าเท่ากับ

$$Values(Wind) = Weak, Strong$$
$$S = [9+, 5-]$$
$$S_{Wind=Weak} \leftarrow [6+, 2-]$$
$$S_{Wind=Weak} \leftarrow [3+, 3-]$$

$$Gain(S,Wind) = Entropy(S) - \sum_{v\in\{Weak,Strong\}} \frac{|S_v|}{[S[} Entropy(S_v)$$

$$= Entropy(S) - \left(\frac{8}{14}\right) Entropy(S_{Wind=Weak}) - \left(\frac{6}{14}\right) Entropy(S_{Wind=Strong})$$

$$= Entropy(S) - \left(\frac{8}{14}\right) 0.811 - \left(\frac{6}{14}\right) 1.00$$

$$= 0.940 - \left(\frac{8}{14}\right) 0.811 - \left(\frac{6}{14}\right) 1.00$$

$$= 0.048$$

คำนวณ $Gain(S,A)$ จนครบทุกคุณลักษณะจากนั้นเลือกค่าคุณลักษณะที่มีค่า $Gain(S,A)$ ที่ดีที่สุด จากตัวอย่างนี้คำนวณค่าการเพิ่มสารสนเทศได้ดังนี้

$Gain(S,Outlook) = 0.246$
$Gain(S,Temperature) = 0.029$
$Gain(S,Humidity) = 0.151$
$Gain(S,Wind) = 0.048$

ดังนั้นเมื่อเลือกคุณลักษณะ Outlook ซึ่งมีค่าการเพิ่มสารสนเทศที่ดีที่สุดเป็นคุณลักษณะแรกในการจำแนกตัวอย่างดังภาพที่ 2.3

จากนั้นคำนวณค่าเพิ่มสารสนเทศที่ดีที่สุดไปเรื่อยๆ จนกระทั่งไม่สามารถเพิ่มบัพลงในต้นไม้ได้ เช่น

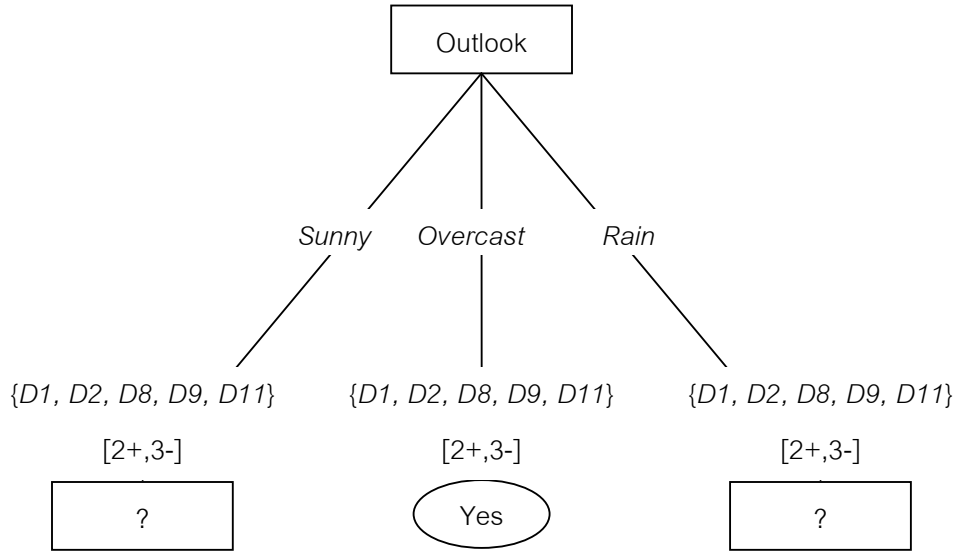
$$S_{Sunny} = \{D1, D2, D8, D9, D11\}$$

$$Gain(S_{Sunny}, Humidity) = 0.970 - \left[\left(\frac{3}{5}\right)0.0 - \left(\frac{2}{5}\right)0.0\right] = 0.970$$

$$Gain(S_{Sunny}, Temperature) = 0.970 - \left[\left(\frac{2}{5}\right)0.0 - \left(\frac{2}{5}\right)1.0 - \left(\frac{1}{5}\right)0.0\right] = 0.570$$

$$Gain(S_{Sunny}, Wind) = 0.970 - \left[\left(\frac{2}{5}\right)1.0 - \left(\frac{3}{5}\right)0.918\right] = 0.019$$

ภาพที่ 2.3 : ผลลัพธ์ต้นไม้ตัดสินใจจากการหาค่าเพิ่มสารสนเทศขั้นแรก



ซี 4.5 (C4.5)

ซี 4.5 เป็นอัลกอริทึมต้นไม้ตัดสินใจที่อิงกับอัลกอริทึมไอดีทรี โดยปรับปรุงคุณสมบัติหลายอย่างให้มีความสามารถเพิ่มขึ้นดังนี้

- การเลือกคุณลักษณะที่เหมาะสมในการเลือก
- จัดการชุดเรียนรู้ที่ไม่มีค่าคุณลักษณะ
- จัดการคุณลักษณะที่ค่าเป็นค่าต่อเนื่องได้

2.1.2 การเตรียมชุดตัวอย่างเรียนรู้บูตสแตรป

บูตสแตรปเป็นวิธีการเตรียมชุดตัวอย่างเรียนรู้ (Training Set) สำหรับต้นไม้แต่ละต้นด้วยการสุ่มตัวอย่าง *m* ตัวอย่างจากชุดตัวอย่างเรียนรู้ตั้งต้นขนาด *m* ดังนั้นชุดตัวอย่างเรียนรู้ใหม่ที่ได้อาจมีตัวอย่างจากชุดตัวอย่างเรียนรู้ตั้งต้นจำนวนมากตัวอย่าง ในขณะที่บางตัวอย่างอาจไม่ปรากฏเลย

จากตารางที่ 2.2 ชุดข้อมูลเริ่มต้นประกอบด้วยหมายเลข 1 ถึง หมายเลข 8 เมื่อเตรียมชุดทดสอบด้วยวิธีบูตสแตรป ผลปรากฏว่าชุดทดสอบที่ 1 มีตัวอย่างหมายเลข 7 ซ้ำกัน 2 ครั้ง ในขณะที่ไม่มีตัวอย่างหมายเลข 8 ชุดทดสอบที่ 2 มีครบทุกตัวอย่างหมายเลข 1 ถึง หมายเลข 8 แต่ชุดตัวอย่างที่ได้ไม่มีการเรียงลำดับ ชุดทดสอบที่ 3 มีตัวอย่างหมายเลข 2 ซ้ำกัน 3 ครั้ง ในขณะ

ที่ไม่มีตัวอย่าง หมายเลข 1, หมายเลข 4 และ หมายเลข 8 และชุดทดสอบที่ 4 มีตัวอย่างหมายเลข 4 ซ้ำกัน 3 ครั้ง ในขณะที่ไม่มีตัวอย่างหมายเลข 2 และ หมายเลข 7

ตารางที่ 2.2 : ชุดข้อมูลที่ได้ของวิธีแบ็กกิง

ตัวอย่างชุดข้อมูลเริ่มต้น	
ชุดทดสอบเริ่มต้น:	1, 2, 3, 4, 5, 6, 7, 8
ตัวอย่างชุดข้อมูลของวิธีแบ็กกิง	
ชุดทดสอบที่ 1:	2, 7, 8, 3, 7, 6, 3, 1
ชุดทดสอบที่ 2:	7, 8, 5, 6, 4, 2, 7, 1
ชุดทดสอบที่ 3:	3, 6, 2, 7, 5, 6, 2, 2
ชุดทดสอบที่ 4:	4, 5, 1, 4, 6, 4, 3, 8

2.1.3 การรวมตัวจำแนก (Classifier Ensemble)

การรวมของการจำแนกที่เป็นที่รู้จักได้แก่ วิธีแบ็กกิง (Breiman, 1996) และวิธีบูสต์ โดยภาพที่ 2.4 แสดงการรวมตัวจำแนก โดยนำผลลัพธ์ที่ได้จากตัวจำแนกแต่ละตัวรวมเป็นผลลัพธ์ (the Ensemble Output)

การรวมการจำแนกวิธีแบ็กกิง (Bagging Classifier)

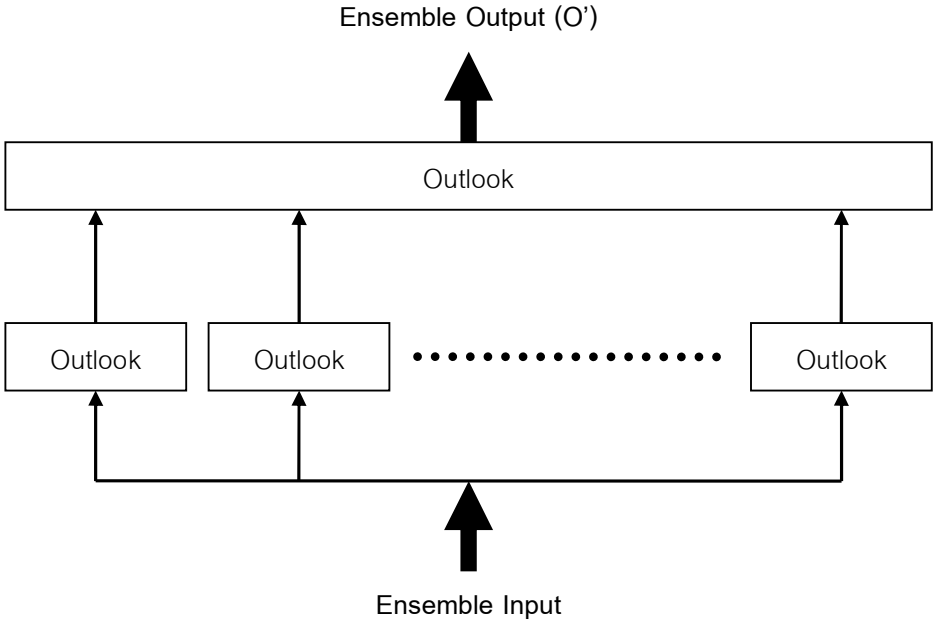
แบ็กกิงมาจาก Bootstrap Aggregating (Brieman, 1996a, pp 123-140) ใช้วิธีบูตสแตรปในการเตรียมชุดตัวอย่างเรียนรู้สำหรับต้นไม้แต่ละต้น โดยที่การสุ่มตัวอย่างแต่ละชุดข้อมูลทำเพื่อนำชุดข้อมูลที่ได้มาใช้เป็นข้อมูลเข้าอัลกอริทึมสร้างต้นไม้ตัดสินใจแต่ละต้น จำนวนชุดข้อมูลที่นำมาใช้มีจำนวนเท่ากับจำนวนชุดข้อมูลในชุดข้อมูลเริ่มต้น การสร้างการจำแนกสุดท้ายสร้างจากการนำคลาสที่ถูกทำนายบ่อยที่สุดมาใช้สร้างการจำแนกสุดท้ายหรือคำตอบสุดท้าย โดยการสร้างการจำแนกสุดท้ายนี้วิธีแบ็กกิงให้ค่าน้ำหนักในแต่ละตัวอย่างและแต่ละกฎเท่ากันทั้งหมด

การจำแนกวิธีบูสต์ (Boosting Classifier)

การรวมของการจำแนกด้วยวิธีบูสต์เน้นที่การสร้างชุดต้นไม้ตัดสินใจ โดยชุดทดสอบที่ใช้ในแต่ละต้นไม้ตัดสินใจถูกเลือกอิงกับประสิทธิภาพของต้นไม้ตัดสินใจก่อนหน้าคือ ตัวอย่างที่

คาดเดาผิดจากการจำแนกก่อนหน้าจะถูกเลือกบ่อยครั้งมากกว่าตัวอย่างที่คาดเดาถูกต้อง ดังนั้น หลักการสำคัญของวิธีบูสต์คือ พยายามสร้างต้นไม้ตัดสินใจให้ได้ค่าคาดเดาที่ดีกว่าประสิทธิภาพการรวมของต้นไม้ตัดสินใจปัจจุบัน

ภาพที่ 2.4 : การรวมตัวจำแนก



จากตารางที่ 2.3 จะเห็นว่าชุดข้อมูลเริ่มต้นมีหมายเลข 1 ถึง หมายเลข 8 เมื่อสุ่มชุดข้อมูลพบว่าชุดทดสอบที่ 4 มีตัวอย่างหมายเลข 1 ซ้ำกัน 5 ครั้ง แสดงว่าตัวอย่างหมายเลข 1 เป็นตัวอย่างที่มีการคาดเดาผิดจากการจำแนกก่อนหน้า ดังนั้นจะเห็นได้ว่าวิธีบูสต์ไม่สามารถใช้สร้างต้นไม้ตัดสินใจแต่ละต้นพร้อมๆ กันได้เนื่องจากในการสุ่มข้อมูลสำหรับแต่ละต้นต้องอิงกับข้อมูลที่ได้จากต้นไม้ก่อนหน้าต่างกับแบ็กกิงที่สามารถทำได้เนื่องจากไม่ต้องอิงกับข้อมูลที่ได้จากต้นไม้ก่อนหน้า

วิธีบูสต์มีสองรูปแบบที่เป็นที่รู้จักกันดีคือ อาร์คกิงเอ็กโฟร์ (Arcing-x4) (Brieman, 1996b) และเอดาบูสต์ (AdaBoost) (Freund and Schapire, 1996, pp 148-156)

วิธีอาร์คกิงเอ็กโฟร์ (มาจากคำว่า Adaptively resample and combine) คล้ายวิธีแบ็กกิง เริ่มจากเลือกชุดเรียนรู้ขนาด n สำหรับการจำแนกครั้งที่ $k+1$ โดยเลือกตัวอย่างฝึกเรียนรู้

ต้นฉบับจำนวน n ชุด ต่างจากวิธีแบ็กกิงตรงที่ค่าความน่าจะเป็นในการเลือกตัวอย่างแต่ละตัวอย่างของวิธีอาร์คชิงเอ็กโฟร์ไม่เท่ากัน โดยค่าความเป็นไปได้ขึ้นกับจำนวนการจำแนกผิดในการจำแนกก่อนหน้า ซึ่งถ้ามีจำนวนมาก ค่าความน่าจะเป็นที่จะถูกเลือกจะมีค่ามากด้วย ส่วนวิธีเอดาบูสต์ (ย่อมาจาก Adaptive Boosting หรือเรียกว่าอาร์คกิงเอฟเอส Arcing-fs โดย Breiman) สามารถทำได้สองรูปแบบคือ ก) เลือกชุดตัวอย่างโดยอิงกับความเป็นไปได้ของตัวอย่าง ข) พิจารณาตัวอย่างทั้งหมดและค่าน้ำหนักแต่ละตัวอย่าง (โดยตัวอย่างที่มีค่าน้ำหนักสูงกว่ามีผลกับความผิดพลาดมาก)

ตารางที่ 2.3 : ชุดข้อมูลที่ได้ของวิธีบูสต์

ตัวอย่างชุดข้อมูลเริ่มต้น	
ชุดทดสอบเริ่มต้น:	1, 2, 3, 4, 5, 6, 7, 8
ตัวอย่างชุดข้อมูลของวิธีบูสต์	
ชุดทดสอบที่ 1:	2, 7, 8, 3, 7, 6, 3, 1
ชุดทดสอบที่ 2:	1, 4, 5, 4, 1, 5, 6, 4
ชุดทดสอบที่ 3:	7, 1, 5, 8, 1, 8, 1, 4
ชุดทดสอบที่ 4:	1, 1, 6, 1, 1, 3, 1, 5

วิธีอาร์คกิงเอ็กโฟร์และวิธีเอดาบูสต์ เริ่มต้นให้ค่าความเป็นไปได้ในแต่ละตัวอย่างเท่ากับ $1/n$ (ดังนั้นค่าความเป็นไปได้รวมทั้งหมดเท่ากับ 1 โดยไม่มีค่าติดลบ) แล้วคำนวณค่าความเป็นได้ใหม่หลังจากที่แต่ละตัวจำแนกที่ถูกฝึกได้เพิ่มเข้าไปในการรวมแล้ว โดยวิธีเอดาบูสต์รวมการจำแนกตั้งแต่ C_1 ถึง C_k (C คือตัวจำแนก) โดยใช้วิธีให้ค่าน้ำหนัก (Weight Voting) โดยที่ C_k มีค่าน้ำหนัก จาก $\log(B_k)$ (B_k มีค่าเท่ากับ $\frac{1-E_k}{E_k}$ โดยค่า E_k คือผลรวมของค่าความเป็นได้ของตัวอย่างที่จัดกลุ่มผิดของการจำแนก C_k ปัจจุบัน) แต่วิธีอาร์คชิง ค่าความเป็นไปได้ที่ $k+1$ มาจาก $p^{k+1}(n) = \frac{1+m(n)^4}{\sum(1+m(n)^4)}$ โดยที่ $m(n)$ เป็นจำนวนการจำแนกผิดของกรณีที่ n โดย $C_1,...,C_k$

การรวมการจำแนกด้วยการลงคะแนนแบบทั่วไปโดยตัวอย่าง

การสร้างการจำแนกสุดท้ายสร้างจากการลงคะแนนคลาสของตัวอย่างที่นำเข้าต้นไม้เพื่อการเรียนรู้และถูกจำแนกได้ด้วยกฎที่จำแนกตัวอย่างใหม่ได้ คลาสที่มีคะแนนมากที่สุดถือเป็นคำตอบสุดท้าย โดยการสร้างการจำแนกสุดท้ายนี้ การรวมการจำแนกด้วยการลงคะแนนแบบทั่วไปโดยตัวอย่างให้ค่าน้ำหนักในแต่ละตัวอย่างและแต่ละกฎเท่ากันทั้งหมด

2.2 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องประกอบไปด้วย โดยงานวิจัยที่ศึกษาอัลกอริทึม เพื่อนบ้านที่ใกล้ที่สุด *k* ตัว, การใช้กฎที่ใกล้ที่สุด และงานวิจัยที่ใช้การคำนวณหาค่าความใกล้เคียงระหว่างกฎ

2.2.1 เพื่อนบ้านที่ใกล้ที่สุด *k* ตัว

เป็นการจำแนกบนพื้นฐานของตัวอย่างเรียนรู้ที่ใกล้ที่สุดคือ คำตอบของการตัดสินใจมาจากการลงคะแนนของเพื่อนบ้าน k หลังหรือตัวอย่างเรียนรู้ที่ใกล้ที่สุดกับตัวอย่างใหม่ k ตัวอย่าง โดยปกติแล้ว k เป็นจำนวนเต็มบวกที่มีค่าน้อย โดยถ้าให้ $k=1$ นั่นคือการให้ตัวอย่างที่ใกล้ที่สุดเป็นคำตอบ นอกจากนี้นิยมกำหนดค่า k เป็นจำนวนคี่เมื่อเป็นการจำแนกที่มีสองคำตอบเพื่อหลีกเลี่ยงการมีคะแนนเท่ากัน

2.2.2 กฎที่ใกล้ที่สุด

Francisco ได้นำเสนอการสร้างต้นไม้ตัดสินใจที่สามารถกำหนดขนาดของตัวจำแนกได้ชื่อว่า SCALLOP [4] ซึ่ง SCALLOP นี้จะเตรียมชุดของกฎที่เป็นประโยชน์ต่อผู้ใช้และจะปรับโครงสร้างความรู้ทุกครั้งที่มีการอ่านตัวอย่างใหม่เข้ามา ด้วยการเพิ่มกฎที่น่าสนใจและลบข้อมูลที่ล้าสมัย โดยจะนำกฎที่อยู่ใกล้ที่สุดสองกฎที่มีความสัมพันธ์กันมาร่วมพิจารณาด้วย

2.2.3 การลดจำนวนกฎ

Lemuel นำเสนอการลดจำนวนกฎที่ได้จากการเตรียมข้อมูลแบบบูธสสแตปปิง โดยมีขั้นตอนดังต่อไปนี้

1 สร้างบทสรุปกฎ (Summarize Rule)

เมื่อเตรียมข้อมูลด้วยบูตสแตรปเรียบร้อย Lemuel ใช้ระบบบรูท (Brute System) วิเคราะห์ข้อมูลและสร้างกฎทั้งหมด แล้วนำเอา N กฎ (50 กฎสำหรับการทดลองของ Lemuel) ที่ดีที่สุด โดยใช้ค่า Extended-Laplace Accuracy ในการคัดเลือก เปรียบเทียบกับกฎที่เหลือ เพื่อกำหนด “กฎใกล้จะเหมือนกัน” (Nearly Identical Rule) ทั้งนี้กฎซ้ำจากชุดข้อมูล (Replications) อื่นๆ จะต้องมีคุณลักษณะและเครื่องหมายเหมือนกับคุณลักษณะของกฎหลัก และค่าของคุณลักษณะที่เป็นค่าไม่ต่อเนื่องต้องเป็นค่าเดียวกัน ในขณะที่คุณลักษณะที่เป็นค่าต่อเนื่องสามารถมีค่าต่างกันได้ จึงจะถือว่า ใกล้จะเหมือน (Nearly Identical) กับกฎหลัก ตัวอย่างเช่น

IF CPTCode = 29 AND Height < 164 AND HeartRateVariability >= 28.6
THEN NauseaGreaterThanMild = yes
ใกล้จะเหมือนกับ
IF CPTCode = 29 AND Height < 158 AND HeartRateVariability >= 25.7
THEN NauseaGreaterThanMild = yes
แต่ไม่ใกล้จะเหมือนกับ
IF CPTCode = 29 AND Height >= 140 AND HeartRateVariability >= 21.3
THEN NauseaGreaterThanMild = yes
และไม่ใกล้จะเหมือนเช่นกันกับ
IF CPTCode <> 27 AND Height < 164 AND HeartRateVariability >= 31.5
THEN NauseaGreaterThanMild = yes

บทสรุปกฎจะถูกสร้างขึ้นเมื่อมีกฎใกล้จะเหมือนกันเกิดขึ้น โดยบทสรุปจะเก็บค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของความถูกต้องลาปลาสขยาย (Extended-Laplace Accuracy) และขอบเขตครอบคลุม (Coverage) (เช่น ร้อยละของกรณีที่รองรับกับกฎดั้งเดิม), จำนวนของชุดข้อมูลที่มีกฎใกล้จะเหมือนกันที่สนับสนุนบทสรุปกฎ และสถิติพื้นฐานเกี่ยวกับการผันแปรของคุณลักษณะที่เป็นค่าต่อเนื่องที่มีอยู่ในกฎ

2 การนำบทสรุปกฎมาใช้และการคัดเลือกระยะของคุณลักษณะที่เป็นค่าไม่ต่อเนื่อง (continuous attribute range filtering)

บทสรุปกฎที่ถูกสร้างขึ้น จะถูกนำมาพิจารณาโดย กำหนดระดับจำนวนชุดข้อมูลที่สนับสนุนกฎหรือระดับสนับสนุน (Supporter) เพื่อจำกัดจำนวนบทสรุปกฎที่แสดง ทั้งนี้ระดับระดับสนับสนุนจะขึ้นอยู่กับชุดข้อมูลและจำนวนกฎที่พบด้วย ดังด้านล่างแสดงตัวอย่างบทสรุปกฎ

SummaryRuleID: 7820, SourceRuleID: NVPreo0048005237
IF CPT = 29 AND HeartRateVariability >= 28.6 (4.5) AND Height < 164 (6.0) THEN
NauseaGreaterThanMild = Yes
(2.87x as likely) Accuracy: 51.2 (10.9), Coverage: 5.9 (1.1), 10/10

ค่าเฉลี่ยของค่าคุณลักษณะที่เป็นค่าต่อเนื่องจะตามด้วยส่วนเบี่ยงเบนมาตรฐานซึ่งถูกแสดงในวงเล็บ และอัตราส่วนโอกาสที่เป็นไปได้ระหว่าง ความถูกต้องลาปลาสขยายของกฎและความแพร่หลายของผลลัพธ์สำหรับชุดข้อมูลทั้งหมดชี้ให้เห็นว่ามีจำนวนกี่ครั้งที่เป็นไปได้ที่ผลลัพธ์เกิดขึ้นสำหรับประชากรที่ตอบสนองกฎที่มีอยู่ก่อนจะถูกแสดงในวงเล็บเช่นกันในบรรทัดถัดมา พร้อมกับแสดงความถูกต้องลาปลาสขยายและขอบเขตครอบคลุม

บทสรุปกฎที่มีคุณลักษณะเป็นค่าต่อเนื่องเหมือนกันสองคุณลักษณะจะถูกทดสอบทางสถิติเพื่อกรองเอากฎที่ด้อยออกไป ตัวอย่างเช่น พิจารณากฎที่กำหนดระยะของตัวแปร Height โดยกำหนดขอบเขตบนและขอบเขตล่าง

IF Height >= 159 AND Height < 163
    AND Phase2Recovery < 27
THEN NauseaGreaterThanMild = Yes

ตัวอย่างด้านบนตัวแปร Height ถูกจำกัดอยู่ในระยะที่ค่อนข้างแคบและไม่ได้คาดไว้ นอกจากนี้กฎใกล้จะเหมือนกันยังเกิดขึ้นทั้ง 10 ชุดข้อมูลและบทสรุปกฎเป็นดังต่อไปนี้

SourceRuleID: NVPreo0048005115, SummaryRuleID: 7851
IF Height >= 158.6 (1.2) AND Height < 162.9 (0.7)
    AND Phase2Recovery < 29.4 (6.24)
THEN NauseaGreaterThanMild = Yes
(2.89x as likely) Accuracy: 51.7 (8.0), Coverage: 6.6 (2.6), 10/10

ในกรณีนี้ควรคำนึงถึงระยะห่างระหว่างสองขอบเขตที่น้อย แต่ด้วยส่วนเบี่ยงเบนมาตรฐานที่มีค่าน้อยเช่นกันชี้ให้เห็นว่าบทสรุปกฎนี้มีความสำคัญ และควรเก็บกฎนี้ไว้ ในขณะที่โดยปกติแล้วควรขจัดกฎที่มีกำหนดระยะอย่างไร้ความหมายเนื่องจากมีส่วนเบี่ยงเบนมาตรฐานที่มีค่ามากโดยเทียบเคียงกับระยะห่างระหว่างขอบเขตบนและขอบเขตล่าง

การทดสอบความแตกต่างระหว่างค่าเฉลี่ยถูกนำมาประยุกต์ใช้ สมมติฐานที่ไม่มีค่า (the null hypothesis) คือ ระยะห่างระหว่างค่าเฉลี่ยของขอบเขตทั้งสองไม่มีความสำคัญ และควบคุมทั้งการกระจายประชากรปกติ (a normal population distribution) และค่าความแปรปรวนเท่ากัน (equal variance)

3 ความใกล้เคียงระหว่างกฎ (Similarity Between Rules)

เมื่อสร้างกฎและบทสรุปกฎแล้ว ผู้วิเคราะห์ระบบ (the analyst) จะต้องการเปรียบเทียบกฎเพื่อตัดสินใจถ้ากฎค้นพบกฎจากพื้นที่ของปัญหาที่ต่างกันหรือค่าการเปลี่ยนที่ค่อนข้างน้อยจากสัดส่วนที่เล็กกว่าของปริภูมิปัญหา (problem space) มากไปกว่านี้ ผู้วิเคราะห์ระบบอาจต้องการเปรียบเทียบกฎที่มีผลลัพธ์ที่ต่างกันเพื่อดูว่าพวกมันเกี่ยวข้องกันในคุณลักษณะหรือไม่

คุณลักษณะของกฎในการจำแนกแบ่งได้เป็นสองประเภทคือ คุณลักษณะที่เป็นค่าต่อเนื่อง (Continuous) และค่าคุณลักษณะที่เป็นค่าไม่ต่อเนื่อง (Discrete) โดยคุณลักษณะที่เป็นค่าต่อเนื่องจะเป็นตัวเลขที่มีค่าต่อเนื่องกันไปตามความมากน้อยของค่าเช่น อายุ, ส่วนสูง, น้ำหนัก เป็นต้น ส่วนค่าคุณลักษณะที่เป็นค่าไม่ต่อเนื่องจะไม่มีความสัมพันธ์ในลักษณะดังกล่าวแต่จะเป็นในลักษณะของการเป็นตัวแทนเช่น สี, รายได้สูงหรือต่ำหรือปานกลาง, ลักษณะรูปร่าง เป็นต้น

Gower ได้นำเสนอสมการเพื่อคำนวณหาค่าความใกล้เคียงระหว่างตัวแปรทั้งที่เป็นค่าต่อเนื่องและค่าไม่ต่อเนื่อง:

$$S_{ij} = \frac{\sum_{k=1}^{p} W_{ijk} S_{ijk}}{\sum_{k=1}^{p} W_{ijk}}$$

ในสมการนี้ S_{ijk} คือ ค่าความใกล้เคียงระหว่างคุณลักษณะที่ k ของกฎที่ i กับกฎที่ j และ W_{IJK} จะมีค่าได้สองค่าคือ 1 หรือ 0 ค่าใดค่าหนึ่ง ขึ้นอยู่กับว่าคุณลักษณะที่ k นั้นมีความเกี่ยวข้องในการเปรียบเทียบนั้นหรือไม่

การคำนวณค่าความใกล้เคียงของคุณลักษณะที่เป็นค่าต่อเนื่อง

เริ่มพิจารณาค่าความใกล้เคียงของคุณลักณะ k ระหว่างกฎ X_i และกฎ X_j โดยกำหนดให้

$X_{ik\min}$ เป็นค่าของคุณลักณะ k ที่น้อยที่สุดของกฎแรก
$X_{jk\min}$ เป็นค่าของคุณลักณะ k ที่น้อยที่สุดของกฎที่สอง
$X_{ik\max}$ เป็นค่าของคุณลักณะ k ที่มากที่สุดของกฎแรก
$X_{jk\max}$ เป็นค่าของคุณลักณะ k ที่มากที่สุดของกฎที่สอง

$R_{ik} = X_{ik\max} - X_{ik\min}$ เป็นระยะห่างค่าคุณลักณะ k ของกฎแรก
$R_{ik} = X_{ik\max} - X_{ik\min}$ เป็นระยะห่างค่าคุณลักณะ k ของกฎที่สอง
R_k เป็นระยะห่างค่าคุณลักณะ k ที่เป็นได้ทั้งหมด

และ

$X_{jk\min} > X_{ik\max}$ ดังนั้นกฎ X_j จะมีขอบเขตบนของค่าคุณลักษณะ k สูงกว่าของกฎ X_i

จากข้อกำหนดเหล่านี้เราสามารถคำนวณหาค่าความใกล้เคียงของคุณลักษณะซึ่งแบ่งได้เป็นสามหมวดดังนี้

- คุณลักษณะไม่ใกล้เคียงกัน: $X_{ik\max} > X_{jk\min}$

$$S_{ijk} = \frac{X_{ik\max} - X_{jk\min}}{R_k - R_{ik} - R_{jk}} \qquad \text{----------(1)}$$

คุณลักษณะที่ไม่มีส่วนทับซ้อนกันนี้จะให้ผลลัพธ์เป็นจำนวนลบ โดยจะยิ่งเข้าใกล้ -1 มากขึ้นเมื่อแต่ละคุณลักษณะมีค่าอยู่คนละด้านบน R_k ดังตัวอย่างจากภาพที่ 2.5 คำนวณค่าความใกล้เคียงได้

$$S_{ijk} = \frac{90-130}{(200-45)-(90-70)-(150-130)} = -0.35$$

- คุณลักษณะมีความใกล้เคียงกัน: ค่าของคุณลักษณะมีส่วนที่ทับซ้อนกัน นั่นคือ

$$S_{ijk} = \frac{X_{ik\max} - X_{jk\min}}{(R_{ik} + R_{jk})/2} \qquad \text{----------(2)}$$

คุณลักษณะที่มีส่วนที่ทับซ้อนกันจะให้ค่าความใกล้เคียงเป็นค่าบวกและเท่ากับ 1 เมื่อกฎครอบคลุมค่าคุณลักษณะในขอบเขตเดียวกัน ดังตัวอย่างจากภาพที่ 2.6 คำนวณค่าความใกล้เคียงได้

$$S_{ijk} = \frac{105 - 100}{((105 - 85) + (120 - 100))/2} = 0.25$$

- คุณลักษณะปรากฏในกฎใดกฎหนึ่งในขณะที่ไม่ปรากฏในอีกกฎหนึ่ง: ในกรณีที่กฎหนึ่งไม่ได้กำหนดค่าคุณลักษณะ จะถือว่าคุณลักษณะนั้นมีค่าเป็นขอบเขตทั้งหมดที่คุณลักษณะนั้นครอบคลุม ดังนั้นการคำนวณจะเป็นการคำนวณโดยส่วนที่ทับซ้อนกัน

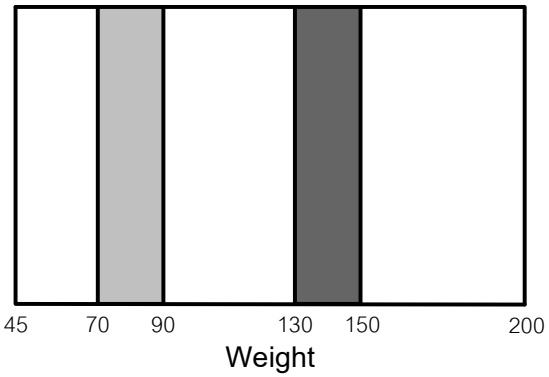
$$S_{ijk} = \frac{R_{ik}}{(R_{ik} + R_{k})/2} \qquad \text{----------(3)}$$

เมื่อคุณลักษณะได้ถูกกำหนดโดยกฎ i แต่ไม่ได้ถูกกำหนดโดยกฎ j ดังภาพที่ 2.5 ถ้าเฉพาะกฎที่ต่ำกว่าคือ น้ำหนักอยู่ระหว่าง 70 กก. และ 90 กก. ถูกกำหนด ในขณะที่อีกกฎไม่ได้กำหนดคุณลักษณะนี้จะได้

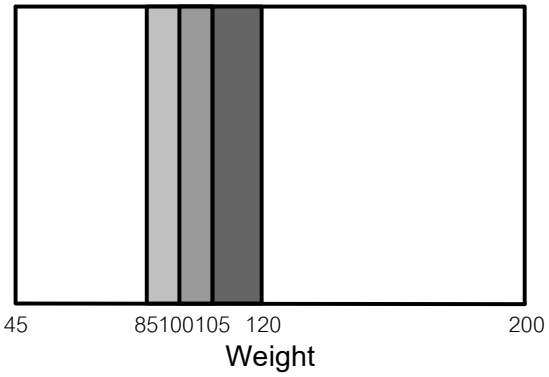
$$S_{ijk} = \frac{90 - 70}{((90 - 70) + (200 - 45))/2} = 0.23$$

ค่าต่ำสุดของค่าความใกล้เคียงคือ -1 โดยแต่ละค่าของคุณลักษณะจะมีค่าอยู่ตรงที่สุดของขอบเขตคนละด้าน และค่าสูงสุดของค่าความใกล้เคียงคือ +1 เมื่อคุณลักษณะมีขอบเขตเป็นค่าเดียวกัน

ภาพที่ 2.5 : ขอบเขตค่าคุณลักษณะของสองกฎที่ไม่มีส่วนทับซ้อนกัน



ภาพที่ 2.6 : ขอบเขตค่าคุณลักษณะของสองกฎที่มีส่วนทับซ้อนกัน



การคำนวณค่าความใกล้เคียงของคุณลักษณะที่เป็นค่าไม่ต่อเนื่อง

คุณลักษณะที่เป็นค่าไม่ต่อเนื่องจะไม่มีความเกี่ยวข้องกับขอบเขตของค่าและการทับซ้อนของค่าแต่จะให้ความสนใจกับจำนวนของค่าไม่ต่อเนื่องที่เป็นไปได้แทน โดยคำนวณจากเศษส่วนที่มีตัวเศษเป็นจำนวนของค่าไม่ต่อเนื่องที่มีอยู่ร่วมกันในทั้งสองกฎหารด้วยจำนวนค่าไม่ต่อเนื่องที่เป็นไปได้ทั้งหมด และตัวส่วนคือ ค่าเฉลี่ยระหว่างจำนวนที่ครอบคลุมโดยคุณลักษณะนั้นของทั้งสองกฎ เราสามารถแบ่งได้เป็นเจ็ดหมวดดังนี้ (แต่ละหมวดแสดงตัวอย่างด้วยการใช้คุณลักษณะ Anesthetic Agent ที่มีค่าไม่ต่อเนื่องที่เป็นไปได้ทั้งหมดหกค่าคือ Desflurane, Isoflurane, Lidocaine, Propofol, Sevoflurane และ Unspecified)

- มีกฎเดียวเท่านั้นที่กำหนดคุณลัษณะและใช้เครื่องหมาย “=” ในการอธิบายในขณะที่อีกกฎไม่มี:

$$S_{ijk} = \frac{1/N}{(1/N+1)/2} = \frac{2}{1+N} \qquad \text{----------(4)}$$

เมื่อ N เป็นจำนวนค่าไม่ต่อเนื่องที่เป็นไปได้ของคุณลักษณะ

Rule 1: $Agent = Desflurane$

Rule 2:

Similarity : $\frac{2}{7}$

- มีกฎเดียวเท่านั้นที่กำหนดคุณลัษณะและใช้เครื่องหมาย “≠” ในการอธิบายในขณะที่อีกกฎไม่มี:

$$S_{ijk} = \frac{N-1/N}{(((1/N)/N)+1)/2} = \frac{2N-2}{2N-1} \qquad \text{----------(5)}$$

เมื่อ N เป็นจำนวนค่าไม่ต่อเนื่องที่เป็นไปได้ของคุณลักษณะ

Rule 1: $Agent \neq Isoflurane$

Rule 2:

Similarity : $\frac{10}{11}$

แต่ในกรณีที่คุณลักษณะหนึ่งนั้นใช้เครื่องหมาย “≠” อธิบายค่าคุณลักษณะมากกว่าหนึ่งครั้งการคำนวณจะเป็น

$$S_{ijk} = \frac{N-M/N}{((N-M)/N+1)/2} = \frac{2N-2M}{2N-M} \qquad \text{----------(6)}$$

เมื่อใช้เครื่องหมาย “≠” อธิบายคุณลักษณะนั้น M ครั้ง

Rule 1: $Agent \neq Isoflurane$ AND $Agent \neq Desflurane$

Rule 2:

Similarity : $\frac{4}{5}$

- เกิด $X_{ik} = X_{jk}$, ค่าคุณลักษณะเป็นค่าเดียวกันและเครื่องหมายที่ใช้อธิบายเป็นเครื่องหมายเดียวกัน:

$$S_{ijk} = 1 \qquad \text{----------(7)}$$

Rule 1: $Agent = Isoflurane$

Rule 2: $Agent = Isoflurane$

Similarity : 1

Rule 1: $Agent \neq Isoflurane$

Rule 2: $Agent \neq Desflurane$

Similarity : 1

- เกิด $X_{ik} = X_{jk}$ และค่าคุณลักษณะเป็นค่าเดียวกันแต่เครื่องหมายที่ใช้อธิบายต่างกัน:

$$S_{ijk} = -1 \qquad \text{----------(8)}$$

Rule 1: $Agent = Isoflurane$

Rule 2: $Agent \neq Isoflurane$

Similarity : -1

- เกิด $X_{ik} \neq X_{jk}$ และค่าของคุณลักษณะมีค่าต่างกันแต่เครื่องหมายที่ใช้อธิบายเหมือนกัน:

$$S_{ijk} = -1 \qquad \text{----------(9)}$$

Rule 1: $Agent = Isoflurane$

Rule 2: $Agent = Desflurane$

Similarity : -1

- กฎหนึ่งใช้เครื่องหมาย “=” ในการอธิบายคุณลักษณะในขณะที่กฎอีกข้อใช้เครื่องหมาย “≠” ในการอธิบาย โดยที่ทั้งสองกฎมีค่าคุณลักษณะต่างกัน:

$$S_{ijk} = \frac{1/N}{\left(1/N + \frac{N-1}{N}\right)/2} = \frac{2}{N} \qquad \text{----------(10)}$$

Rule 1: $Agent = Isoflurane$

Rule 2: $Agent \neq Desflurane$

Similarity : $\frac{1}{3}$

จากสมการนี้จะถูกนำมาขยายต่อเมื่อเครื่องหมาย “≠” ถูกใช้อธิบายมากกว่าหนึ่งครั้งในกฎหนึ่งข้อ ดังนี้

$$S_{ijk} = \frac{1/N}{\left(1/N + \frac{N-M}{N}\right)/2} = \frac{2}{1+N-M} \qquad \text{----------(11)}$$

เมื่อใช้เครื่องหมาย “≠” อธิบายคุณลักษณะนั้น M ครั้ง

Rule 1: $Agent = Isoflurane$

Rule 2: $Agent \neq Desflurane$ AND $Agent \neq Sevoflurane$

Similarity : $\frac{2}{5}$

(สมการที่ 4 เป็นกรณีของสมการที่ 11 เมื่อ $M = 0$)

- กฎทั้งสองใช้เครื่องหมาย “≠” ในการอธิบายคุณลักษณะเดียวกันและใช้อธิบายมากกว่าหนึ่งครั้ง ค่าความใกล้เคียงจะขึ้นอยู่กับสัดส่วนของจำนวนค่าคุณลักษณะที่ถูกใช้อธิบายในทั้งสองกฎลบกับสัดส่วนของจำนวนค่าคุณลักษณะที่ไม่ถูกใช้อธิบายในทั้งสองกฎ:

$$S_{ijk} = \frac{(N-(M_1 \,\mathrm{Y}\, M_2))/N - (M_1 \oplus M_2)/N}{\left(\frac{N-M}{N} + \frac{N-M_2}{N}\right)/2} \qquad \text{----------(12)}$$

$$S_{ijk} = \frac{(N-(M_1 \text{ Y } M_2))/N - ((M_1 \text{ Y } M_2)-(M_1 \text{ I } M_2))/N}{\left(\frac{N-M}{N} + \frac{N-M_2}{N}\right)/2}$$

$$S_{ijk} = \frac{2(N-2(M_1 \text{ Y } M_2)+(M_1 \text{ I } M_2))}{2N-M_1-M_2} \qquad \text{----------(13)}$$

Rule 1: $Agent \neq Isoflurane$ AND $Agent \neq Desflurane$

Rule 2: $Agent \neq Desflurane$ AND $Agent \neq Sevoflurane$

Similarity : $\frac{1}{4}$

(สมการที่ 6 เป็นกรณีของสมการที่ 13 เมื่อ M_1 หรือ M_2 ค่าใดค่าหนึ่งมีค่าเป็น 0)

4 การสเกลหลายมิติ (Multidimensional Scaling, MDS)

คำนวณหาค่าความใกล้เคียงระหว่างบทสรุปกฎ แล้วนำบทสรุปกฎที่ผ่านระดับสนับสนุน (ในข้อ 2) เข้าสู่อะเรย์และสร้างเมตริกซ์ของค่าความใกล้เคียงเพื่อใช้เป็นข้อมูลเข้า การสเกลหลายมิติให้ผู้วิเคราะห์ระบบเห็นภาพความใกล้เคียงของกฎว่าแสดงให้เห็นการจับกันเป็นกลุ่มใหญ่ของกฎและมีจำนวนกลุ่มที่น้อยหรือมีการกระจายที่เหมือนมากกว่า

การสเกลหลายมิติของเมตริกซ์ค่าความใกล้เคียงของบทสรุปกฎด้วย เอสพีเอสเอส (SPSS) เวอร์ชั่น 10.0, พีออกซ์สกอล (PEOXSCAL) เวอร์ชั่น 1.0 และใช้การทำสเกลโทรเกอร์สัน (Trogerson Scaling) ในการสร้างสถานะเริ่มต้นและกำหนดคำตอบของปัญหาในรูปสองมิติ

ผลการทดลองพบว่า บทสรุปกฎแสดงค่าที่เปลี่ยนแปลงได้ในขอบเขตที่รวมกันและคาดหวังโดยผู้ชำนาญการเฉพาะทาง (domain expert) และลดจำนวนของกฎที่ต้องถูกวิเคราะห์ นอกจากนี้ระดับสนับสนุนระดับสูงทำให้บทสรุปกฎมีความเสถียรมาก และให้ความมั่นใจในด้านปัญหาการเกินพอดี (over-fitted construct of machine learning) สุดท้ายการสเกลหลายมิติเป็นเครื่องมือที่มีประโยชน์มากในการหาชุดของกฎ