บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 แผนภาพตัดสินใจทวิภาค (Binary Decision Diagrams : BDDs)

แผนภาพตัดสินใจทวิภาคเป็นโครงสร้างข้อมูลแบบกราฟไม่มีวงที่ระบุทิศทาง (Directed Acyclic Graph : DAG) ซึ่งใช้แทนฟังก์ชันบูลีน [1-4] โดยมีส่วนประกอบดังรูปที่ 2.1 คือ

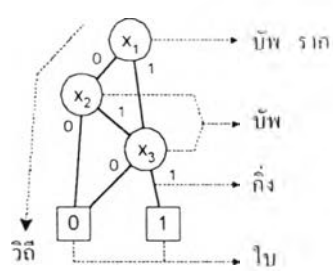
(1) บัพ (node / non-terminal) แสดงถึง ตัวแปรของฟังก์ชันบูลีน โดยเรียกบัพแรกของแผนภาพตัดสินใจทวิภาคว่าราก

(2) กิ่ง (branch) แสดงถึง ค่าของตัวแปร โดยแต่ละตัวแปรมี 2 กิ่งเสมอ คือ กิ่งที่มีค่าของตัวแปรเป็น 0 และ 1

(3) ใบ (leaf / terminal) แสดงถึง ค่าคำตอบของฟังก์ชันบูลีน ซึ่งมีค่าคำตอบที่เป็นไปได้ 2 ค่า คือ 0 และ 1

(4) วิถี (path) คือ เส้นทางที่เชื่อมต่อกันของบัพและกิ่งจากรากถึงใบ โดยทุกวิถีของแผนภาพตัดสินใจทวิภาคจะต้องมีลำดับของตัวแปรที่เหมือนกันทั้งหมด

(5) ขนาด (size) คือ จำนวนบัพทั้งหมดของแผนภาพตัดสินใจทวิภาค



รูปที่ 2.1 แผนภาพตัดสินใจทวิภาค

จากรูปที่ 2.1 แสดงถึงแผนภาพตัดสินใจทวิภาคของฟังก์ชัน 3 ตัวแปร คือ x_1 x_2 และ x_3 โดยมี**ขนาดของแผนภาพตัดสินใจทวิภาค**เท่ากับ 3 บัพ และวิถีทั้งหมด 5 วิถี คือ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิถีที่ 1 : เมื่อตัวแปร $x_1 = 0$ และ $x_2 = 0$ จะให้ค่าคำตอบของฟังก์ชันเป็น 0

วิถีที่ 2 : เมื่อตัวแปร $x_1 = 0$ $x_2 = 1$ และ $x_3 = 0$ จะให้ค่าคำตอบของฟังก์ชันเป็น 0

วิถีที่ 3 : เมื่อตัวแปร $x_1 = 0$ $x_2 = 1$ และ $x_3 = 1$ จะให้ค่าคำตอบของฟังก์ชันเป็น 1

วิถีที่ 4 : เมื่อตัวแปร $x_1 = 1$ และ $x_3 = 0$ จะให้ค่าคำตอบของฟังก์ชันเป็น 0

วิถีที่ 5 : เมื่อตัวแปร $x_1 = 1$ และ $x_3 = 1$ จะให้ค่าคำตอบของฟังก์ชันเป็น 1

เมื่อพิจารณาวิถีทั้งหมด จะเห็นได้ว่าแผนภาพตัดสินใจทวิภาคดังกล่าวมีลำดับตัวแปรเป็น x_1 x_2 และ x_3 ตามลำดับ ซึ่งสามารถแสดงแบบย่อได้คือ (x_1, x_2, x_3)

ขั้นตอนการสร้างแผนภาพตัดสินใจทวิภาคอย่างง่าย สามารถทำได้โดยการเปลี่ยนรูปแบบการแทนข้อมูลจากตารางค่าความจริงไปเป็นกราฟ ซึ่งในทุกๆวิถีของกราฟจะต้องมีลำดับของตัวแปรที่เหมือนกัน จากนั้นก็นำกราฟดังกล่าวมาลดทอนส่วนที่ซ้ำ โดยการใช้กฎการลดทอน (reduction rule) ดังนี้

(1) กำจัดใบที่ซ้ำกัน (remove duplicate terminals) เป็นการยุบใบที่ซ้ำกันให้เหลือเพียงใบเดียว และเปลี่ยนทิศทางของกิ่งที่เชื่อมต่อกับใบที่ถูกยุบไปเชื่อมต่อกับใบที่เหลือแทน

(2) กำจัดบัพที่ซ้ำกัน (remove duplicate non-terminals) เป็นการยุบบัพที่ซ้ำกันให้เหลือเพียงบัพเดียว และเปลี่ยนทิศทางของกิ่งที่เชื่อมต่อกับบัพที่ถูกยุบไปเชื่อมต่อกับบัพที่เหลือแทน

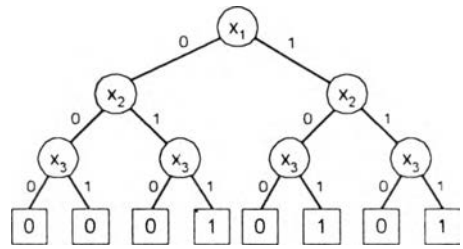
(3) กำจัดส่วนที่ซ้ำซ้อน (remove redundant tests) เป็นการตัดบัพที่กิ่ง 0 และ 1 ของมันเชื่อมต่อที่เดียวกัน และเปลี่ยนทิศทางของกิ่งที่เชื่อมต่อกับบัพที่ถูกตัดให้เชื่อมต่อกันโดยตรงแทน

ตารางที่ 2.1 ตารางค่าความจริงของฟังก์ชัน $f(x_1,x_2,x_3) = (x_1 + x_2) \cdot x_3$

x_1	x_2	x_3	f
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	0
1	1	1	1

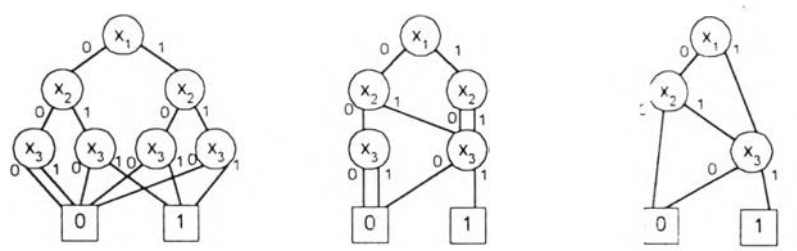
จากตัวอย่างตารางค่าความจริงในตารางที่ 2.1 สามารถเปลี่ยนรูปแบบการแทนข้อมูลเป็นกราฟที่มีลำดับตัวแปร คือ (x_1, x_2, x_3) ได้ดังรูปที่ 2.2 โดยตารางค่าความจริงจะแสดงทุกค่าคำตอบของฟังก์ชัน

สำหรับตัวแปร x_1 x_2 และ x_3 ตามแนวตั้ง แต่แผนภาพตัดสินใจจะแสดงทุกค่าคำตอบของฟังก์ชันสำหรับตัวแปร x_1 x_2 และ x_3 ตามแนวนอน



รูปที่ 2.2 กราฟที่มีลำดับตัวแปร (x_1,x_2,x_3) ของฟังก์ชัน $f(x_1,x_2,x_3) = (x_1 + x_2) \cdot x_3$

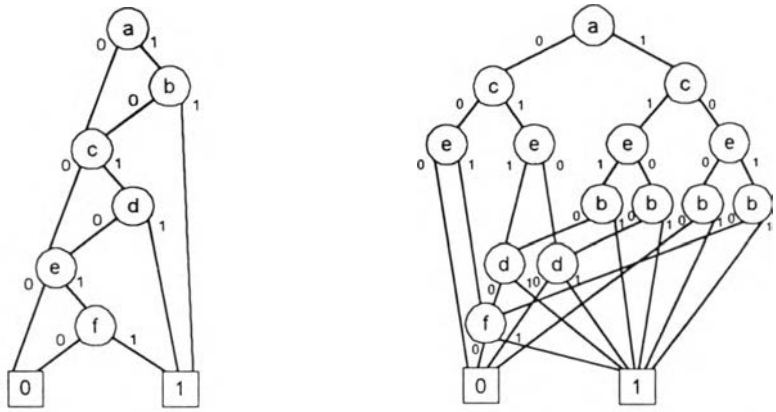
เมื่อนำกราฟในรูปที่ 2.2 มาลดทอนส่วนที่ซ้ำด้วยการใช้กฎการลดทอนจะได้ดังรูปที่ 2.3 โดยการกำจัดใบที่ซ้ำกันในรูปที่ 2.3(1) สามารถกำจัดใบที่มีค่า 0 และ 1 ที่ซ้ำกันได้ 4 และ 2 ใบตามลำดับ การกำจัดบัพที่ซ้ำกันในรูปที่ 2.3(2) สามารถกำจัดบัพของตัวแปร x_3 ที่ซ้ำกันได้ 2 บัพ และการกำจัดส่วนที่ซ้ำซ้อนดังรูปที่ 2.3(3) สามารถกำจัดบัพที่ซ้ำซ้อนได้ 2 บัพสำหรับตัวแปร x_2 และ x_3



(1) กำจัดใบที่ซ้ำกัน (2) กำจัดบัพที่ซ้ำกัน (3) กำจัดส่วนที่ซ้ำซ้อน

รูปที่ 2.3 การลดทอนส่วนที่ซ้ำของกราฟสำหรับฟังก์ชัน $f(x_1,x_2,x_3) = (x_1 + x_2) \cdot x_3$

การสร้างแผนภาพตัดสินใจทวิภาคด้วยวิธีดังกล่าวสามารถสร้างได้หลายแบบ โดยที่แต่ละแบบจะมีขนาดของแผนภาพแตกต่างกันไปตามลำดับของตัวแปร ดังรูปที่ 2.4 ซึ่งแสดงตัวอย่างของการแทนแผนภาพตัดสินใจทวิภาคสำหรับฟังก์ชัน $f(a,b,c,d,e,f) = a \cdot b + c \cdot d + e \cdot f$ เดียวกัน แต่ใช้ลำดับของตัวแปรที่ต่างกัน ซึ่งส่งผลให้แผนภาพตัดสินใจทวิภาคที่ได้มีขนาดต่างกันด้วย นั่นคือ แผนภาพตัดสินใจทวิภาคในรูปที่ 2.4(1) มีลำดับตัวแปร คือ (a,b,c,d,e,f) และมีขนาดเท่ากับ 6 บัพ แต่แผนภาพตัดสินใจทวิภาคในรูปที่ 2.4(2) มีลำดับตัวแปร คือ (a,c,e,b,d,f) และมีขนาดเท่ากับ 14 บัพ ดังนั้นสิ่งที่สำคัญของการสร้างแผนภาพตัดสินใจทวิภาคที่มีขนาดเล็ก คือ การหาลำดับตัวแปรที่ดี



(1) ลำดับตัวแปรคือ (a, b, c, d, e, f)　　(2) ลำดับตัวแปรคือ (a, c, e, b, d, f)

รูปที่ 2.4 แผนภาพตัดสินใจทวิภาคของฟังก์ชัน f(a,b,c,d,e,f) = a·b + c·d + e·f

2.2 การพัฒนาทีละขั้น (Gradual Improvement)

การพัฒนาทีละขั้นเป็นวิธีการหาลำดับของตัวแปรที่ดี กายหลังการสร้างแผนภาพตัดสินใจทวิภาค โดยใช้การแลกเปลี่ยนลำดับของตัวแปร เพื่อลดขนาดของแผนภาพตัดสินใจทวิภาคให้เล็กลง โดยขั้นตอนวิธีสำหรับการพัฒนาทีละขั้น ได้แก่ AD2 AD3 AD4 AR ARSA และ SIFTING

(1) <u>ขั้นตอนวิธี ADm</u> (เมื่อ m = 2 3 และ 4) เป็นการแลกเปลี่ยนตัวแปร m ตัวที่อยู่ติดกันของแผนภาพตัดสินใจทวิภาคที่มีตัวแปรทั้งหมด n ตัว โดยการเรียงสับเปลี่ยน (permutation) ลำดับของตัวแปร m ตัวนั้น แล้วสร้างแผนภาพตัดสินใจทวิภาคที่ได้จากการเรียงสับเปลี่ยนทั้งหมด m! แผนภาพ ดังตัวอย่างรูปที่ 2.5 2.6 และ 2.7 จากนั้นจะเลือกแผนภาพตัดสินใจทวิภาคที่เล็กที่สุดภายใน m! นั้น กระบวนการนี้จะถูกทำซ้ำ จนกระทั่งไม่สามารถหาแผนภาพตัดสินใจทวิภาคที่เล็กกว่าได้ [10]

ขั้นตอนวิธี AD2 (m = 2) เป็นการแลกเปลี่ยนตัวแปร 2 ตัวที่อยู่ติดกัน ซึ่งจำนวนลำดับตัวแปรของการเรียงสับเปลี่ยนมีค่าเท่ากับ 2 ลำดับ รูปที่ 2.5 แสดงตัวอย่างการเรียงสับเปลี่ยนของขั้นตอนวิธี AD2 สำหรับตัวแปร x_3 และ x_4

การเรียงสับเปลี่ยนของตัวแปร (x_3, x_4) = { (x_3, x_4) (x_4, x_3) }

$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	เริ่มต้น
$x_1, x_2, x_4, x_3, x_5, x_6, x_7$	สลับที่ (x_3, x_4)

รูปที่ 2.5 การเรียงสับเปลี่ยนของขั้นตอนวิธี AD2 สำหรับตัวแปร x_3 และ x_4

ขั้นตอนวิธี AD3 (m = 3) เป็นการแลกเปลี่ยนตัวแปร 3 ตัวที่อยู่ติดกัน ซึ่งจำนวนลำดับตัวแปรของการเรียงสับเปลี่ยนมีค่าเท่ากับ 6 ลำดับ รูปที่ 2.6 แสดงตัวอย่างการเรียงสับเปลี่ยนของขั้นตอนวิธี AD3 สำหรับตัวแปร x_2 x_3 และ x_4

การเรียงสับเปลี่ยนของตัวแปร (x_2, x_3, x_4) = { (x_2, x_3, x_4) (x_2, x_4, x_3) (x_4, x_2, x_3)
(x_4, x_3, x_2) (x_3, x_4, x_2) (x_3, x_2, x_4) }

$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	เริ่มต้น
$x_1, x_2, x_4, x_3, x_5, x_6, x_7$	สลับที่ (x_3, x_4)
$x_1, x_4, x_2, x_3, x_5, x_6, x_7$	สลับที่ (x_2, x_4)
$x_1, x_4, x_3, x_2, x_5, x_6, x_7$	สลับที่ (x_2, x_3)
$x_1, x_3, x_4, x_2, x_5, x_6, x_7$	สลับที่ (x_4, x_3)
$x_1, x_3, x_2, x_4, x_5, x_6, x_7$	สลับที่ (x_4, x_2)

รูปที่ 2.6 การเรียงสับเปลี่ยนของขั้นตอนวิธี AD3 สำหรับตัวแปร x_2 x_3 และ x_4

ขั้นตอนวิธี AD4 (m = 4) เป็นการแลกเปลี่ยนตัวแปร 4 ตัวที่อยู่ติดกัน ซึ่งจำนวนลำดับตัวแปรของการเรียงสับเปลี่ยนมีค่าเท่ากับ 24 ลำดับ รูปที่ 2.7 แสดงตัวอย่างการเรียงสับเปลี่ยนของขั้นตอนวิธี AD4 สำหรับตัวแปร x_1 x_2 x_3 และ x_4

การเรียงสับเปลี่ยนของตัวแปร (x_1, x_2, x_3, x_4) = { (x_1, x_2, x_3, x_4) (x_2, x_1, x_3, x_4) (x_2, x_1, x_4, x_3)
(x_1, x_2, x_4, x_3) (x_1, x_4, x_2, x_3) (x_1, x_4, x_3, x_2)
(x_4, x_1, x_3, x_2) (x_4, x_1, x_2, x_3) (x_4, x_2, x_1, x_3)
(x_2, x_4, x_1, x_3) (x_2, x_4, x_3, x_1) (x_4, x_2, x_3, x_1)
(x_4, x_3, x_2, x_1) (x_4, x_3, x_1, x_2) (x_3, x_4, x_1, x_2)
(x_3, x_4, x_2, x_1) (x_3, x_2, x_4, x_1) (x_2, x_3, x_4, x_1)
(x_2, x_3, x_1, x_4) (x_3, x_2, x_1, x_4) (x_3, x_1, x_2, x_4)
(x_3, x_1, x_4, x_2) (x_1, x_3, x_4, x_2) (x_1, x_3, x_2, x_4) }

รูปที่ 2.7 การเรียงสับเปลี่ยนของขั้นตอนวิธี AD4 สำหรับตัวแปร x_1 x_2 x_3 และ x_4

$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	เริ่มต้น
$x_2, x_1, x_3, x_4, x_5, x_6, x_7$	สลับที่ (x_1, x_2)
$x_2, x_1, x_4, x_3, x_5, x_6, x_7$	สลับที่ (x_3, x_4)
$x_1, x_2, x_4, x_3, x_5, x_6, x_7$	สลับที่ (x_2, x_1)
$x_1, x_4, x_2, x_3, x_5, x_6, x_7$	สลับที่ (x_2, x_4)
$x_1, x_4, x_3, x_2, x_5, x_6, x_7$	สลับที่ (x_2, x_3)
$x_4, x_1, x_3, x_2, x_5, x_6, x_7$	สลับที่ (x_1, x_4)
$x_4, x_1, x_2, x_3, x_5, x_6, x_7$	สลับที่ (x_3, x_2)
$x_4, x_2, x_1, x_3, x_5, x_6, x_7$	สลับที่ (x_1, x_2)
$x_2, x_4, x_1, x_3, x_5, x_6, x_7$	สลับที่ (x_4, x_2)
$x_2, x_4, x_3, x_1, x_5, x_6, x_7$	สลับที่ (x_1, x_3)
$x_4, x_2, x_3, x_1, x_5, x_6, x_7$	สลับที่ (x_2, x_4)
$x_4, x_3, x_2, x_1, x_5, x_6, x_7$	สลับที่ (x_2, x_3)
$x_4, x_3, x_1, x_2, x_5, x_6, x_7$	สลับที่ (x_2, x_1)
$x_3, x_4, x_1, x_2, x_5, x_6, x_7$	สลับที่ (x_4, x_3)
$x_3, x_4, x_2, x_1, x_5, x_6, x_7$	สลับที่ (x_1, x_2)
$x_3, x_2, x_4, x_1, x_5, x_6, x_7$	สลับที่ (x_4, x_2)
$x_2, x_3, x_4, x_1, x_5, x_6, x_7$	สลับที่ (x_1, x_2)
$x_2, x_3, x_1, x_4, x_5, x_6, x_7$	สลับที่ (x_4, x_1)
$x_3, x_2, x_1, x_4, x_5, x_6, x_7$	สลับที่ (x_2, x_3)
$x_3, x_1, x_2, x_4, x_5, x_6, x_7$	สลับที่ (x_2, x_1)
$x_3, x_1, x_4, x_2, x_5, x_6, x_7$	สลับที่ (x_2, x_4)
$x_1, x_3, x_4, x_2, x_5, x_6, x_7$	สลับที่ (x_3, x_1)
$x_1, x_3, x_2, x_4, x_5, x_6, x_7$	สลับที่ (x_4, x_2)

รูปที่ 2.7 การเรียงสับเปลี่ยนของขั้นตอนวิธี AD4 สำหรับตัวแปร x_1 x_2 x_3 และ x_4 (ต่อ)

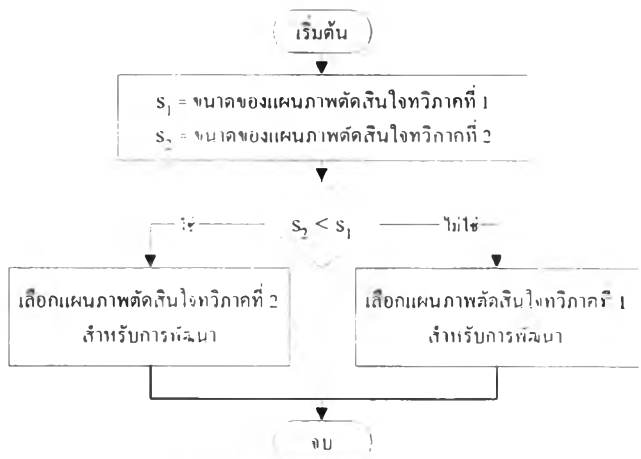
(2) <u>ขั้นตอนวิธี AR และ ARSA</u> เป็นการแลกเปลี่ยนตัวแปร 2 ตัวใดๆแบบสุ่มของแผนภาพตัดสินใจทวิภาคที่มีตัวแปรทั้งหมด n ตัว แล้วทดลองสร้างแผนภาพตัดสินใจทวิภาคใหม่ที่ได้จากการแลกเปลี่ยนตัวแปรทั้งสองตัว จากนั้นเปรียบเทียบขนาดของแผนภาพที่ได้ โดยขั้นตอนวิธี AR จะเลือกแผนภาพตัดสินใจทวิภาคที่เล็กกว่าเท่านั้นสำหรับการพัฒนา ดังรูปที่ 2.8 ส่วนขั้นตอนวิธี ARSA จะยอมให้เลือกแผนภาพตัดสินใจทวิภาคที่ใหญ่ขึ้นได้บ้างสำหรับการพัฒนา ดังรูปที่ 2.9 โดยจะใช้ความน่าจะเป็นตามเทคนิคของการจำลองการหลอม (simulated annealing) [12,13] ดังสมการที่ 2.1 สำหรับการเลือกแผนภาพตัดสินใจทวิภาคที่ใหญ่ขึ้น กระบวนการนี้จะถูกทำซ้ำจนครบจำนวนครั้งของ

การแลกเปลี่ยนตัวแปร และในที่สุดแผนภาพตัดสินใจทวิภาคที่เล็กที่สุดในระหว่างการพัฒนาจะถูกเลือกเป็นคำตอบ [10]

$$p = e^{-\Delta E/T} \tag{2.1}$$

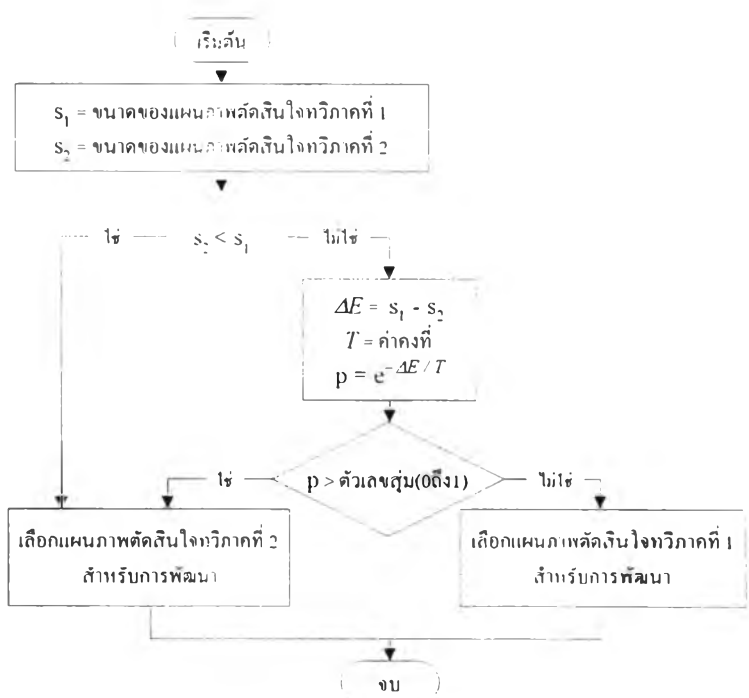
โดย ΔE คือ ค่าของการศึกษาสำนึกที่เปลี่ยนไป

T คือ ค่าคงที่



หมายเหตุ แผนภาพตัดสินใจทวิภาคที่ 2 ได้จากการแลกเปลี่ยนตัวแปรสองตัวใดๆของแผนภาพตัดสินใจทวิภาคที่ 1

รูปที่ 2.8 การเลือกแผนภาพตัดสินใจในการพัฒนาสำหรับวิธี AR



หมายเหตุ แผนภาพตัดสินใจทวิภาคที่ 2 ได้จากการแลกเปลี่ยนตัวแปรสองตัวใดๆของแผนภาพตัดสินใจทวิภาคที่ 1

รูปที่ 2.9 การเลือกแผนภาพตัดสินใจในการพัฒนาสำหรับวิธี ARSA

(3) ขั้นตอนวิธี SIFTING เป็นวิธีการหาตำแหน่งที่เหมาะสมสำหรับตัวแปรหนึ่งของแผนภาพตัดสินใจทวิภาคที่มีตัวแปรทั้งหมด n ตัว โดยการเลื่อนตำแหน่งของตัวแปรนั้นไปยัง n ตำแหน่งที่เป็นไปได้ แล้วสร้างแผนภาพตัดสินใจทวิภาคที่ได้จากการเลื่อนทั้งหมด n แผนภาพ ดังตัวอย่างในรูปที่ 2.10 จากนั้นจะเลือกแผนภาพตัดสินใจทวิภาคที่เล็กที่สุดภายใน n นั้น กระบวนการนี้จะถูกทำซ้ำจนกระทั่งไม่สามารถหาแผนภาพตัดสินใจทวิภาคที่เล็กกว่าได้ [11]

การเลื่อนตำแหน่งของตัวแปรหนึ่งไปยัง n ตำแหน่ง สามารถทำได้โดยการแลกเปลี่ยนตำแหน่งกับตัวแปรก่อนหน้ามันไปเรื่อยๆจนกระทั่งตัวแปรนั้นเป็นตัวแรก และแลกเปลี่ยนตำแหน่งกับตัวแปรหลังมันไปเรื่อยๆจนกระทั่งตัวแปรนั้นเป็นตัวสุดท้าย ดังรูปที่ 2.10(1) และ 2.10(2) ตามลำดับ

$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	เริ่มต้น
$x_1, x_2, x_4, x_3, x_5, x_6, x_7$	สลับที่ (x_3, x_4)
$x_1, x_4, x_2, x_3, x_5, x_6, x_7$	สลับที่ (x_2, x_4)
$x_4, x_1, x_2, x_3, x_5, x_6, x_7$	สลับที่ (x_1, x_4)

(1) การเลื่อนตัวแปร x_4 ขึ้น

$x_1, x_2, x_3, x_4, x_5, x_6, x_7$	เริ่มต้น
$x_1, x_2, x_3, x_5, x_4, x_6, x_7$	สลับที่ (x_4, x_5)
$x_1, x_2, x_3, x_5, x_6, x_4, x_7$	สลับที่ (x_4, x_6)
$x_1, x_2, x_3, x_5, x_6, x_7, x_4$	สลับที่ (x_4, x_7)

(2) การเลื่อนตัวแปร x_4 ลง

รูปที่ 2.10 การเลื่อนตำแหน่งตัวแปรของขั้นตอนวิธี SIFTING สำหรับตัวแปร x_4

กำหนดให้ n เป็นจำนวนตัวแปรทั้งหมดของแผนภาพตัดสินใจทวิภาค และเมื่อพิจารณาขั้นตอนวิธีการพัฒนาทีละขั้นทั้งหมด สามารถคำนวณความซับซ้อนของแต่ละขั้นตอนวิธีได้ดังตารางที่ 2.2 คือ

(1) ขั้นตอนวิธี ADm (เมื่อ m = 2 3 และ 4) เป็นการแลกเปลี่ยนตัวแปร m ตัวที่อยู่ติดกัน ซึ่งในแต่ละรอบของการแลกเปลี่ยนนั้น จะมีจำนวนครั้งของการสลับที่ตัวแปรเท่ากับ m!-1 ครั้งเสมอ โดยกรณีที่ดีที่สุดจะมีจำนวนรอบของการแลกเปลี่ยนเป็นจำนวน n-m+1 ครั้ง (ซึ่ง n-m+1 เป็นจำนวนกลุ่มทั้งหมดของการเลือกตัวแปร m ตัวที่อยู่ติดกัน จากตัวแปรทั้งหมด n ตัว) และโดยเฉลี่ยแล้วจำนวนรอบของการแลกเปลี่ยนจะเท่ากับ $n\cdot(n-m+1)$ ครั้ง ดังนั้นจำนวนครั้งของการสลับที่ตัวแปรทั้งหมดโดยเฉลี่ยจะเท่ากับ $(m!-1)\cdot(n\cdot(n-m+1))$

(2) ขั้นตอนวิธี AR และ ARSA เป็นการแลกเปลี่ยนตัวแปร 2 ตัวใดๆแบบสุ่ม ซึ่งในแต่ละรอบของการแลกเปลี่ยนนั้น จะมีจำนวนครั้งของการสลับที่ตัวแปรโดยเฉลี่ยเท่ากับ n ครั้ง และจะมีจำนวนรอบของการแลกเปลี่ยนเท่ากับ L ครั้งเสมอ (เมื่อ L เป็นจำนวนครั้งของการแลกเปลี่ยนตัวแปรที่กำหนด) ดังนั้นจำนวนครั้งของการสลับที่ตัวแปรทั้งหมดโดยเฉลี่ยจะเท่ากับ $n\cdot L$

(3) ขั้นตอนวิธี SIFTING เป็นการเลื่อนตำแหน่งตัวแปรหนึ่งไปยัง n ตำแหน่งที่เป็นไปได้ ซึ่งในแต่ละรอบของการแลกเปลี่ยนนั้น จะมีจำนวนครั้งของการสลับที่ตัวแปรเท่ากับ n–1 ครั้งเสมอ โดยกรณีที่ดีที่สุดจะมีจำนวนรอบของการแลกเปลี่ยนเป็นจำนวน n ครั้ง (ซึ่ง n เป็นจำนวนตัวแปรทั้งหมด) และโดยเฉลี่ยแล้วจำนวนรอบของการแลกเปลี่ยนจะเท่ากับ n·n ครั้ง ดังนั้นจำนวนครั้งของการสลับที่ตัวแปรทั้งหมดโดยเฉลี่ยจะเท่ากับ (n)·(n·n)

ตารางที่ 2.2 ความซับซ้อนของขั้นตอนวิธีการพัฒนาทีละขั้น

ขั้นตอนวิธีการพัฒนาทีละขั้น	จำนวนครั้งของการสลับที่ตัวแปรสำหรับ 1 รอบของการแลกเปลี่ยน	จำนวนรอบของการแลกเปลี่ยน		จำนวนครั้งของการสลับที่ตัวแปรทั้งหมดโดยเฉลี่ย
		ดีที่สุด (best case)	*โดยเฉลี่ย (average case)*	
ADm	$\Theta(m!)$	$\Theta(n-m)$	$O(n^2)$	$O(m!\cdot n^2)$
AR และ ARSA	$O(n)$	$\Theta(L)$	$\Theta(L)$	$O(n\cdot L)$
SIFTING	$\Theta(n)$	$\Theta(n)$	$O(n^2)$	$O(n^3)$

2.3 การเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learning : DTL)

การเรียนรู้ต้นไม้ตัดสินใจเป็นวิธีการเรียนรู้จากตัวอย่างที่อาศัยการจัดหมวดหมู่ (classification) จากตัวอย่างที่เรียกว่าข้อมูลสอน (training data) และสร้างเป็นต้นไม้ตัดสินใจ (decision tree) [12-15] โดยการเรียนรู้ต้นไม้ตัดสินใจจะถูกใช้ในการอนุมานแบบอุปนัย (inductive inference) เช่น การวินิจฉัยโรค การประเมินความเสี่ยง เป็นต้น

ข้อมูลสอนประกอบด้วยแถวแสดงถึงตัวอย่าง และสดมภ์แสดงถึงคุณลักษณะ ซึ่งมี 2 ชนิด ดังตัวอย่างในตารางที่ 2.3 คือ

(1) คุณลักษณะแบ่งพวก (category attribute) เป็นคุณลักษณะที่กำหนดว่าตัวอย่างถูกจัดอยู่พวกไหน (class) โดยจะมีเพียงคุณลักษณะเดียวเท่านั้นในแต่ละชุดข้อมูลสอน

(2) คุณลักษณะไม่แบ่งพวก (non-category attribute) เป็นคุณลักษณะที่บอกถึงลักษณะต่างๆของตัวอย่าง ซึ่งสามารถมีได้หลายคุณลักษณะในแต่ละชุดข้อมูลสอน

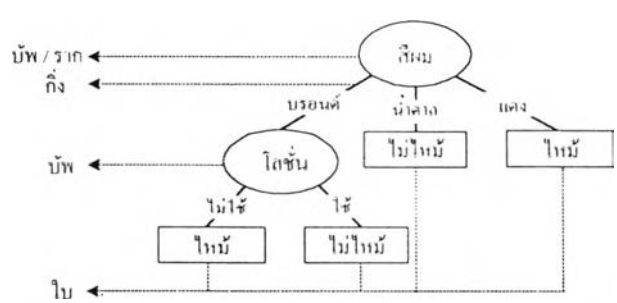
ตารางที่ 2.3 ข้อมูลสอนของปัจจัยที่ทำให้ผิวหนังไหม้

		คุณลักษณะไม่แบ่งพวก			คุณลักษณะแบ่งพวก
คุณลักษณะ	สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
ตัวอย่าง	บรอนด์	ปานกลาง	เบา	ไม่ใช้	ไหม้
	บรอนด์	เตี้ย	ปานกลาง	ไม่ใช้	ไหม้
	บรอนด์	สูง	ปานกลาง	ใช้	ไม่ไหม้
	บรอนด์	เตี้ย	เบา	ใช้	ไม่ไหม้
	น้ำตาล	เตี้ย	ปานกลาง	ใช้	ไม่ไหม้
	น้ำตาล	สูง	หนัก	ไม่ใช้	ไม่ไหม้
	น้ำตาล	ปานกลาง	หนัก	ไม่ใช้	ไม่ไหม้
	แดง	ปานกลาง	หนัก	ไม่ใช้	ไหม้

จากตารางที่ 2.3 แสดงถึงคุณลักษณะต่างๆที่มีผลต่อผิวหนังไหม้ โดยมีคุณลักษณะแบ่งพวกเป็นผิวหนัง ซึ่งมีค่าที่เป็นไปได้ คือ {ไหม้, ไม่ไหม้} และมีคุณลักษณะไม่แบ่งพวกเป็นสีผม ส่วนสูง น้ำหนัก และโลชั่น ซึ่งมีค่าที่เป็นไปได้ คือ {บรอนด์, น้ำตาล, แดง} {เตี้ย, ปานกลาง, สูง} {เบา, ปานกลาง, หนัก} และ{ใช้, ไม่ใช้} ตามลำดับ

ต้นไม้ตัดสินใจเป็นโครงสร้างข้อมูลแบบกราฟ โดยมีส่วนประกอบ ดังรูปที่ 2.11 คือ

(1) บัพ แสดงถึง คุณลักษณะไม่แบ่งพวก และเรียกบัพแรกของต้นไม้ตัดสินใจว่าราก
(2) กิ่ง แสดงถึง ค่าที่เป็นไปได้ตามคุณลักษณะไม่แบ่งพวก
(3) ใบ แสดงถึง ค่าที่เป็นไปได้ของคุณลักษณะแบ่งพวก ซึ่งเป็นค่าคำตอบของการตัดสินใจ
(4) วิถี คือ เส้นทางที่เชื่อมต่อกันของบัพและกิ่งจากรากถึงใบ
(5) ขนาด คือ จำนวนบัพทั้งหมดของต้นไม้ตัดสินใจ



รูปที่ 2.11 ต้นไม้ตัดสินใจของปัจจัยที่ทำให้ผิวหนังไหม้

จากรูปที่ 2.11 แสดงถึงต้นไม้ตัดสินใจที่ได้จากการเรียนรู้ตัวอย่างของข้อมูลสอนในตารางที่ 2.3 ซึ่งประกอบไปด้วยบัพ 2 บัพ คือ สีผมและโลชั่น โดยมีขนาดของต้นไม้ตัดสินใจเท่ากับ 2 บัพ และวิถีทั้งหมด 4 วิถี คือ

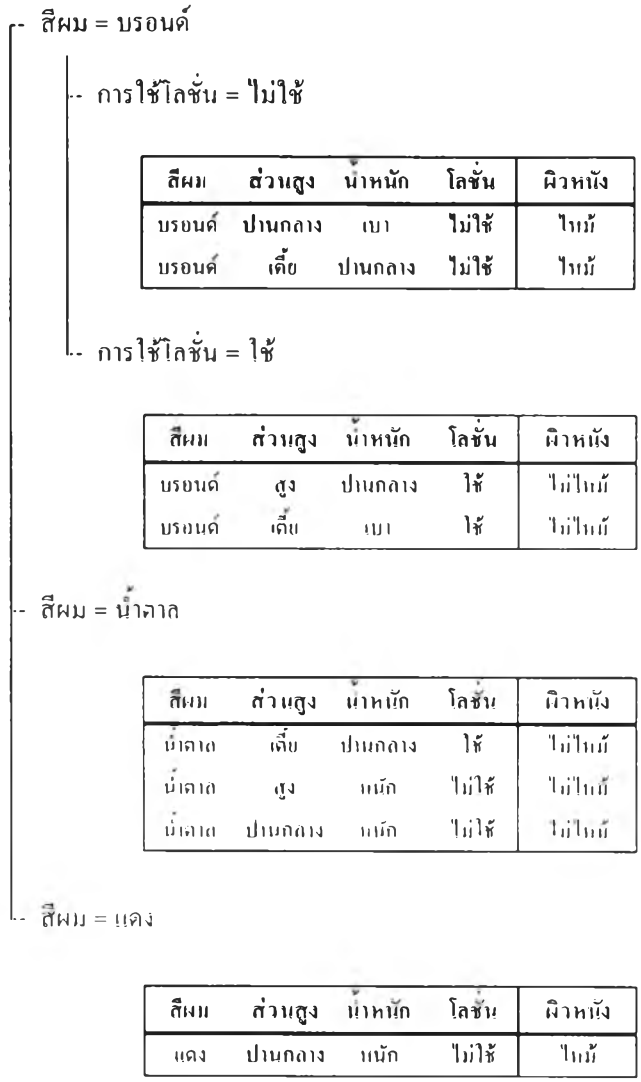
วิถีที่ 1 : เมื่อสีผม = บรอนด์ และโลชั่น = ไม่ใช้ จะให้ค่าของคุณลักษณะผิวหนัง คือ ไหม้
วิถีที่ 2 : เมื่อสีผม = บรอนด์ และโลชั่น = ใช้ จะให้ค่าของคุณลักษณะผิวหนัง คือ ไม่ไหม้
วิถีที่ 3 : เมื่อสีผม = น้ำตาล จะให้ค่าของคุณลักษณะผิวหนัง คือ ไม่ไหม้
วิถีที่ 4 : เมื่อสีผม = แดง จะให้ค่าของคุณลักษณะผิวหนัง คือ ไหม้

การสร้างต้นไม้ตัดสินใจทำได้โดยการเลือกคุณลักษณะไม่แบ่งพวกทีละ 1 คุณลักษณะ และแบ่งตัวอย่างออกเป็นกลุ่มย่อยตามค่าคุณลักษณะไม่แบ่งพวกนั้น ในกรณีที่กลุ่มย่อยประกอบไปด้วยตัวอย่างหลายพวก จะต้องทำการแบ่งกลุ่มย่อยนั้นต่อไป โดยการเลือกคุณลักษณะไม่แบ่งพวกตัวใหม่ และทำเช่นนี้ไปเรื่อยๆจนกระทั่งกลุ่มย่อยที่ได้จากการแบ่งเป็นพวกเดียวกัน

จากตัวอย่างข้อมูลสอนในตารางที่ 2.3 เมื่อแบ่งตัวอย่างตามคุณลักษณะของสีผม พบว่ากลุ่มตัวอย่างที่มีสีผมเป็นสีน้ำตาลและแดงประกอบด้วยตัวอย่างพวกเดียวกัน (ตัวอย่างมีค่าของคุณลักษณะของผิวหนังที่เหมือนกัน) ส่วนกลุ่มตัวอย่างที่มีสีผมบรอนด์ประกอบด้วยตัวอย่างหลายพวก (ตัวอย่างมีค่าของคุณลักษณะของผิวหนังที่ต่างกัน) และเมื่อแบ่งกลุ่มตัวอย่างนี้ต่อไปด้วยคุณลักษณะของโลชั่น จะพบว่าแต่ละกลุ่มย่อยประกอบด้วยตัวอย่างที่เป็นพวกเดียวกัน ดังรูปที่ 2.12 และต้นไม้ตัดสินใจที่ได้เป็นไปดังรูปที่ 2.11

การสร้างต้นไม้ตัดสินใจด้วยวิธีดังกล่าวสามารถสร้างต้นไม้ตัดสินใจได้หลายต้น ขึ้นอยู่กับการเลือกคุณลักษณะไม่แบ่งพวกที่แตกต่างกัน โดยต้นไม้ตัดสินใจที่ต้องการควรมีขนาดเล็ก ดังนั้นสิ่งที่สำคัญคือการเลือกคุณลักษณะไม่แบ่งพวกที่ใช้แบ่งตัวอย่างแล้ว ทำให้จำนวนครั้งของการแบ่งหรือจำนวนบัพน้อยๆ

การเรียนรู้ต้นไม้ตัดสินใจเป็นการใช้ศึกษาสำนึกช่วยในการเลือกคุณลักษณะไม่แบ่งพวกในแต่ละบัพของต้นไม้ตัดสินใจ โดยจะอาศัยหลักการของทฤษฎีสารสนเทศ (information theory) ในการคำนวณค่าความสามารถในการแยกตัวอย่าง และเรียกค่าที่คำนวณได้ว่าค่า *Gain* ดังสมการที่ (2.2) ค่า *Gain* จะมีค่าอยู่ระหว่าง 0 กับ 1 โดยค่า *Gain* เท่ากับ 0 แสดงถึงความสามารถในการแยกตัวอย่างของคุณลักษณะไม่แบ่งพวกนั้นดีที่สุด สามารถแยกตัวอย่างเป็นพวกเดียวกันทั้งหมด และความสามารถในการแยกตัวอย่างจะลดลงเมื่อค่า *Gain* เพิ่มขึ้น



รูปที่ 2.12 การแบ่งตัวอย่างของปัจจัยที่ทำให้ผิวหนังไหม้

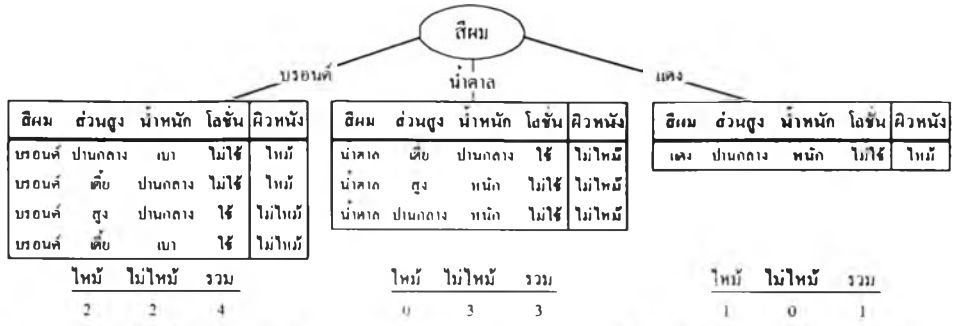
$$Gain(\text{คุณลักษณะไม่แบ่งพวก}) = \sum_{b}\left[\left(\frac{n_b}{n_t}\right)\times\left(\sum_{c}-\frac{n_{bc}}{n_b}\log_2\frac{n_{bc}}{n_b}\right)\right] \qquad (2.2)$$

โดย n_b เป็นจำนวนตัวอย่างในกิ่ง b

n_t เป็นจำนวนตัวอย่างทั้งหมด

n_{bc} เป็นจำนวนตัวอย่างของพวก c ในกิ่ง b

จากตัวอย่างของข้อมูลสอนในตารางที่ 2.3 เมื่อทดลองแบ่งตัวอย่างและคำนวณค่า *Gain* สำหรับแต่ละคุณลักษณะไม่แบ่งพวกของสีผม ความสูง น้ำหนัก และโลชั่นเป็นไปดังรูปที่ 2.13 จะเห็นได้ว่าการแบ่งตัวอย่างด้วยคุณลักษณะของสีผมมีค่า *Gain* ที่น้อยที่สุด ดังนั้นคุณลักษณะของสีผมจะถูกเลือกเป็นบัพแรกของต้นไม้ตัดสินใจ



สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
บรอนด์	ปานกลาง	เบา	ไม่ใช้	ไหม้
บรอนด์	เตี้ย	ปานกลาง	ไม่ใช้	ไหม้
บรอนด์	สูง	ปานกลาง	ใช้	ไม่ไหม้
บรอนด์	เตี้ย	เบา	ใช้	ไม่ไหม้

ไหม้	ไม่ไหม้	รวม
2	2	4

สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
น้ำตาล	เตี้ย	ปานกลาง	ใช้	ไม่ไหม้
น้ำตาล	สูง	หนัก	ไม่ใช้	ไม่ไหม้
น้ำตาล	ปานกลาง	หนัก	ไม่ใช้	ไม่ไหม้

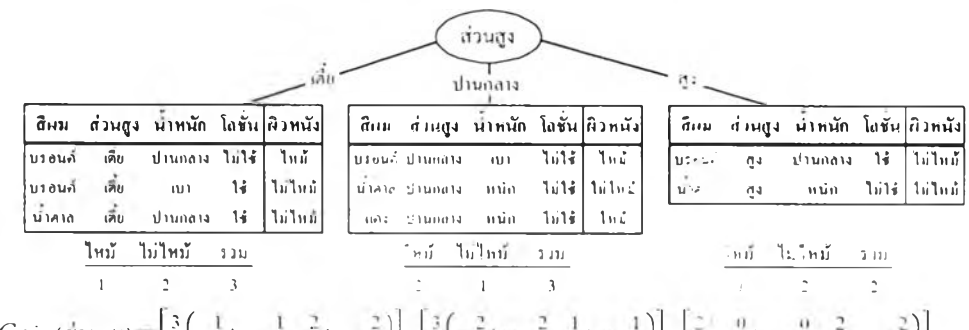
ไหม้	ไม่ไหม้	รวม
0	3	3

สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
แดง	ปานกลาง	หนัก	ไม่ใช้	ไหม้

ไหม้	ไม่ไหม้	รวม
1	0	1

$$Gain(\text{สีผม}) = \left[\frac{4}{8}\left(-\frac{2}{4}\log_2\frac{2}{4}-\frac{2}{4}\log_2\frac{2}{4}\right)\right]+\left[\frac{3}{8}\left(-\frac{0}{3}\log_2\frac{0}{3}-\frac{3}{3}\log_2\frac{3}{3}\right)\right]+\left[\frac{1}{8}\left(-\frac{1}{1}\log_2\frac{1}{1}-\frac{0}{1}\log_2\frac{0}{1}\right)\right]=0.50$$

(1) แบ่งตัวอย่างโดยการใช้คุณลักษณะของสีผม



สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
บรอนด์	เตี้ย	ปานกลาง	ไม่ใช้	ไหม้
บรอนด์	เตี้ย	เบา	ใช้	ไม่ไหม้
น้ำตาล	เตี้ย	ปานกลาง	ใช้	ไม่ไหม้

ไหม้	ไม่ไหม้	รวม
1	2	3

สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
บรอนด์	ปานกลาง	เบา	ไม่ใช้	ไหม้
น้ำตาล	ปานกลาง	หนัก	ไม่ใช้	ไม่ไหม้
แดง	ปานกลาง	หนัก	ไม่ใช้	ไหม้

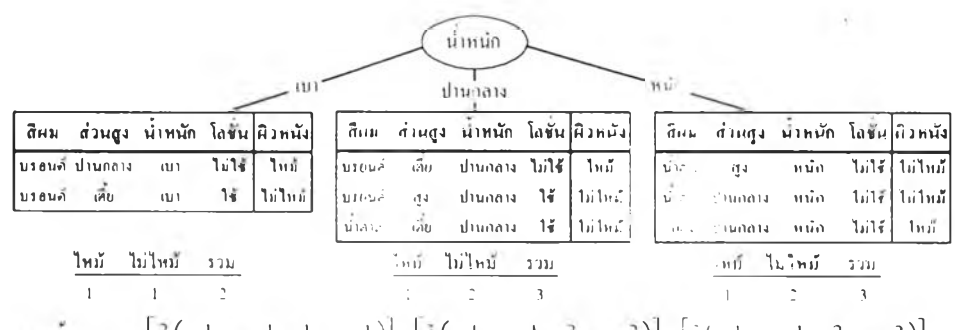
ไหม้	ไม่ไหม้	รวม
2	1	3

สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
บรอนด์	สูง	ปานกลาง	ใช้	ไม่ไหม้
น้ำตาล	สูง	หนัก	ไม่ใช้	ไม่ไหม้

ไหม้	ไม่ไหม้	รวม
0	2	2

$$Gain(\text{ส่วนสูง}) = \left[\frac{3}{8}\left(-\frac{1}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{2}{3}\right)\right]+\left[\frac{3}{8}\left(-\frac{2}{3}\log_2\frac{2}{3}-\frac{1}{3}\log_2\frac{1}{3}\right)\right]+\left[\frac{2}{8}\left(-\frac{0}{2}\log_2\frac{0}{2}-\frac{2}{2}\log_2\frac{2}{2}\right)\right]=0.69$$

(2) แบ่งตัวอย่างโดยการใช้คุณลักษณะของส่วนสูง



สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
บรอนด์	ปานกลาง	เบา	ไม่ใช้	ไหม้
บรอนด์	เตี้ย	เบา	ใช้	ไม่ไหม้

ไหม้	ไม่ไหม้	รวม
1	1	2

สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
บรอนด์	เตี้ย	ปานกลาง	ไม่ใช้	ไหม้
บรอนด์	สูง	ปานกลาง	ใช้	ไม่ไหม้
น้ำตาล	เตี้ย	ปานกลาง	ใช้	ไม่ไหม้

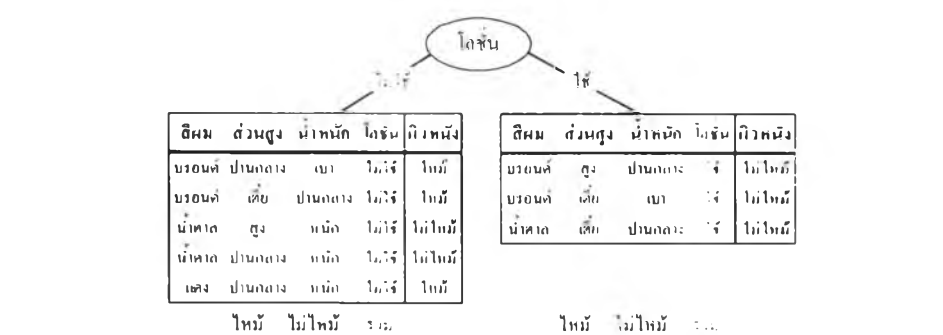
ไหม้	ไม่ไหม้	รวม
1	2	3

สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
น้ำตาล	สูง	หนัก	ไม่ใช้	ไม่ไหม้
น้ำตาล	ปานกลาง	หนัก	ไม่ใช้	ไม่ไหม้
แดง	ปานกลาง	หนัก	ไม่ใช้	ไหม้

ไหม้	ไม่ไหม้	รวม
1	2	3

$$Gain(\text{น้ำหนัก}) = \left[\frac{2}{8}\left(-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}\right)\right]+\left[\frac{3}{8}\left(-\frac{1}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{2}{3}\right)\right]+\left[\frac{3}{8}\left(-\frac{1}{3}\log_2\frac{1}{3}-\frac{2}{3}\log_2\frac{2}{3}\right)\right]=0.94$$

(3) แบ่งตัวอย่างโดยการใช้คุณลักษณะของน้ำหนัก



สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
บรอนด์	ปานกลาง	เบา	ไม่ใช้	ไหม้
บรอนด์	เตี้ย	ปานกลาง	ไม่ใช้	ไหม้
น้ำตาล	สูง	หนัก	ไม่ใช้	ไม่ไหม้
น้ำตาล	ปานกลาง	หนัก	ไม่ใช้	ไม่ไหม้
แดง	ปานกลาง	หนัก	ไม่ใช้	ไหม้

ไหม้	ไม่ไหม้	รวม
1	2	3

สีผม	ส่วนสูง	น้ำหนัก	โลชั่น	ผิวหนัง
บรอนด์	สูง	ปานกลาง	ใช้	ไม่ไหม้
บรอนด์	เตี้ย	เบา	ใช้	ไม่ไหม้
น้ำตาล	เตี้ย	ปานกลาง	ใช้	ไม่ไหม้

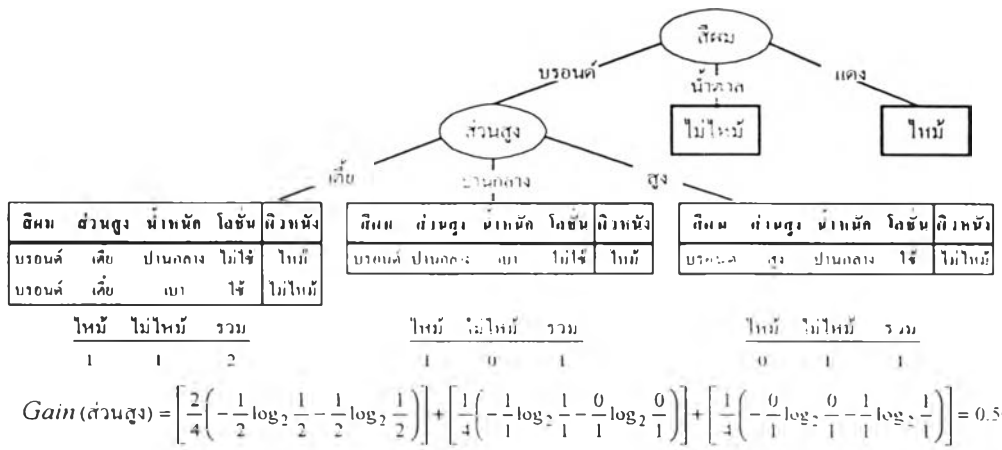
ไหม้	ไม่ไหม้	รวม
1	2	3

$$Gain(\text{โลชั่น}) = \left[\frac{5}{8}\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)\right]+\left[\frac{3}{8}\left(-\frac{0}{3}\log_2\frac{0}{3}-\frac{3}{3}\log_2\frac{3}{3}\right)\right]=0.61$$

(4) แบ่งตัวอย่างโดยการใช้คุณลักษณะของโลชั่น

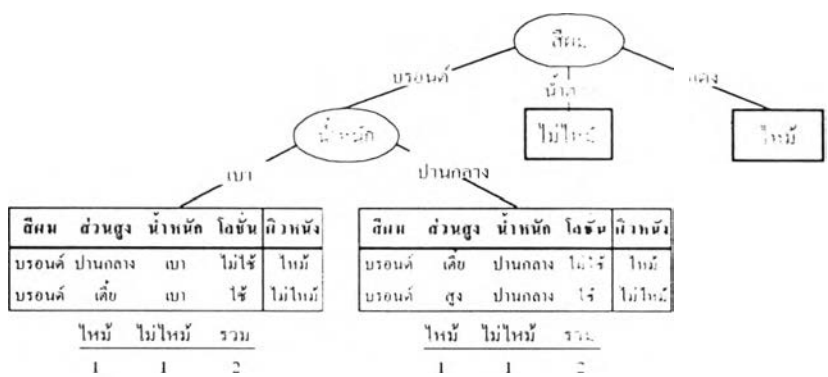
รูปที่ 2.13 การแบ่งตัวอย่างของปัจจัยที่ทำให้ผิวไหม้สำหรับแต่ละคุณลักษณะไม่แบ่งพวก (ระดับที่1)

เมื่อพิจารณาการแบ่งตัวอย่างด้วยคุณลักษณะของสีผม พบว่ากลุ่มตัวอย่างที่มีสีผมเป็นสีน้ำตาลและแดงสามารถแยกตัวอย่างเป็นพวกเดียวกัน แต่กลุ่มตัวอย่างที่มีสีผมเป็นสีบรอนด์ยังไม่สามารถแยกตัวอย่างเป็นพวกเดียวกันได้ ดังนั้นกลุ่มตัวอย่างที่มีสีผมบรอนด์จะถูกแบ่งตัวอย่างต่อไปด้วยคุณลักษณะไม่แบ่งพวกที่เหลือ ดังรูปที่ 2.14



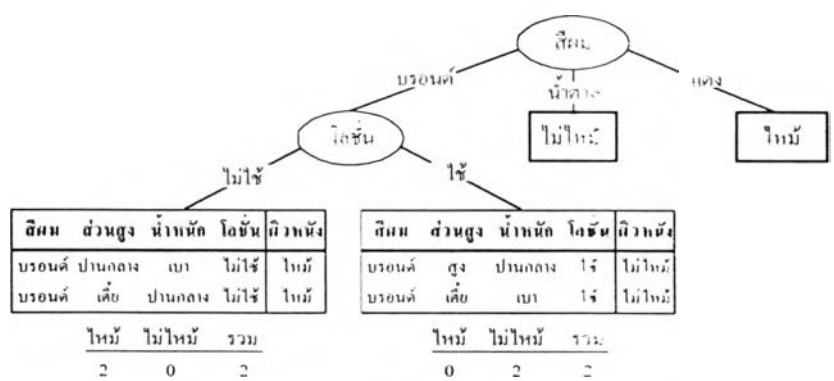
$$Gain(\text{ส่วนสูง}) = \left[\frac{2}{4}\left(-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}\right)\right]+\left[\frac{1}{4}\left(-\frac{1}{1}\log_2\frac{1}{1}-\frac{0}{1}\log_2\frac{0}{1}\right)\right]+\left[\frac{1}{4}\left(-\frac{0}{1}\log_2\frac{0}{1}-\frac{1}{1}\log_2\frac{1}{1}\right)\right] = 0.50$$

(1) แบ่งตัวอย่างโดยการใช้คุณลักษณะของส่วนสูง



$$Gain(\text{น้ำหนัก}) = \left[\frac{2}{4}\left(-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}\right)\right]+\left[\frac{2}{4}\left(-\frac{1}{2}\log_2\frac{1}{2}-\frac{1}{2}\log_2\frac{1}{2}\right)\right] = 1.00$$

(2) แบ่งตัวอย่างโดยการใช้คุณลักษณะของน้ำหนัก



$$Gain(\text{โลชั่น}) = \left[\frac{2}{4}\left(-\frac{2}{2}\log_2\frac{2}{2}-\frac{0}{2}\log_2\frac{0}{2}\right)\right]+\left[\frac{2}{4}\left(-\frac{0}{2}\log_2\frac{0}{2}-\frac{2}{2}\log_2\frac{2}{2}\right)\right] = 0.00$$

(3) แบ่งตัวอย่างโดยการใช้คุณลักษณะของโลชั่น

รูปที่ 2.14 การแบ่งตัวอย่างของปัจจัยที่ทำให้ผิวไหม้สำหรับแต่ละคุณลักษณะไม่แบ่งพวก (ระดับที่2)

จากรูปที่ 2.14 เมื่อทดลองแบ่งกลุ่มตัวอย่างที่มีสีผมบรอนด์ และคำนวณค่า *Gain* สำหรับแต่ละคุณลักษณะไม่แบ่งพวกของความสูง น้ำหนัก และโลชั่น พบว่าการแบ่งตัวอย่างด้วยคุณลักษณะของโลชั่นมีค่า *Gain* ที่น้อยที่สุด ดังนั้นคุณลักษณะของโลชั่นจะถูกเลือกบัพต่อไปของต้นไม้ตัดสินใจ และเมื่อพิจารณาการแบ่งตัวอย่างด้วยคุณลักษณะของสีผมและโลชั่น พบว่ากลุ่มตัวอย่างที่ได้จากการแบ่งด้วยคุณลักษณะทั้งสอง สามารถแยกตัวอย่างเป็นพวกเดียวกันทั้งหมด และต้นไม้ตัดสินใจที่ได้เป็นดังรูปที่ 2.11