บทที่ 3

แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง

วิทยานิพนธ์นี้ใช้เทคนิคแผนที่การจัดกลุ่มเองมาสร้างแบบจำลองการสกัดความรู้ ซึ่งมี 2 แบบจำลอง คือ แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง (Knowledge Extraction Using Self-Organizing Map : KESOM) ความรู้ที่สกัดได้อยู่ในรูปของกฎทั่วไป มีการทำงาน 4 ขั้นตอนหลัก คือ ขั้นตอนหลักที่ A1 : กระบวนการแผนที่การจัดกลุ่มเอง (Self-Organizing Map Process) ขั้นตอนหลักที่ A2 : การวิเคราะห์ Component Plane และ Label (Component Plane and Label Analysis) ขั้นตอนหลักที่ A3 : การสกัดคุณลักษณะที่สำคัญ (Feature Extraction Process) และ ขั้นตอนหลักที่ A4 : การสร้างกฎความรู้ (Generated Knowledge Based) ส่วนแบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (Knowledge Extraction from Self-Organizing Map Using Minimization Entropy Principle Algorithm : KESOM_MEP) ความรู้ที่สกัดได้อยู่ในรูปของกฎภาษาธรรมชาติ มีการทำงาน 5 ขั้นตอนหลัก คือ ขั้นตอนหลักที่ B1 : กระบวนการแผนที่การจัดกลุ่มเอง (Self-Organizing Map Process) ขั้นตอนหลักที่ B2 : หลักเอ็นโทรพีค่าต่ำสุด (Minimization Entropy Principle Algorithm) ขั้นตอนหลักที่ B3 : การสร้างฟัซซี่เซต (Creating Fuzzy Set) ขั้นตอนหลักที่ B4 : การสร้างรูปแบบภาษาธรรมชาติ (Creating Linguistic Term Using Membership Function) และขั้นตอนหลักที่ B5 : การสร้างกฎความรู้ (Knowledge Extraction Using Rough Set)

3.1 แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง (KESOM)

แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง สกัดความรู้ได้ในรูปของกฎทั่วไป แสดงได้ดังภาพประกอบ 3.1 ประกอบด้วย 4 ขั้นตอนหลัก ดังนี้ ขั้นตอนหลักที่ A1 : กระบวนการแผนที่การจัดกลุ่มเอง (Self-Organizing Map Process) ขั้นตอนหลักที่ A2 : การวิเคราะห์ Component Plane และ Label (Component Plane and Label Analysis) ขั้นตอนหลักที่ A3 : การสกัดคุณลักษณะที่สำคัญ (Feature Extraction Process) และ ขั้นตอนหลักที่ A4 : การสร้างกฎความรู้ (Generated Knowledge Based) รายละเอียดของแบบจำลอง KESOM แสดงได้ดังภาพประกอบ 3.2

ขั้นตอนหลักที่ A1 : Self-Organizing Map Process
ขั้นตอนหลักที่ A2 : Component Plane and Label Analysis
ขั้นตอนหลักที่ A3 : Feature Extraction Process
ขั้นตอนหลักที่ A4 : Generated Knowledge Based

ภาพประกอบ 3.1 แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง (KESOM)

ขั้นตอนหลักที่ A1 : Self-Organizing Map Process

1.1 สำหรับแต่ละตัวแปรเข้า d_i ทำการจัดกลุ่มข้อมูลโดยใช้ SOM เมื่อ $1 \leq i \leq n$ และ n คือ จำนวนตัวแปรเข้าทั้งหมด

ผลลัพธ์ที่ได้จากการจัดกลุ่ม คือ component plane, color bar และ label

↓

ขั้นตอนหลักที่ A2 : Component Plane and Label Analysis

2.1 กำหนดให้ $CP_{pq}[x,y,z]$ คือ เมทริกซ์ Component Plane เมื่อ CP_{pq} คือ Component plane แถวที่ p และคอลัมน์ที่ q, x คือ ดัชนีค่าสี, y คือ กลุ่มจาก component plane และ z คือ ค่าความถี่จาก component plane

2.2 กำหนดให้ $CB_x[a,b,c]$ คือ เมทริกซ์ Color Bar เมื่อ $1 \leq x \leq m$ โดยที่ a คือ ค่าสี, b คือ ค่าข้อมูลเข้า, c คือ กลุ่ม และ m คือ จำนวนทั้งหมดของสี

2.3 หากลุ่มที่เป็นไปได้ของแต่ละดัชนีค่าสี เมื่อ w คือ ดัชนีของกลุ่ม, cl คือ จำนวนทั้งหมดของกลุ่ม และ $freq_class_w$ คือ ความถี่ของกลุ่ม c

```
For v = 1  to m do
    For w = 1 to cl, freq_class_w = 0 /* reset the frequency all class to zero
    For i = 1 to p
      For j = 1 to q
        if v = CP_ij[x]
        then  w = CP_ij[y],
                freq_class_w =freq_class_w  + CP_ij[z]
              /* the component plane value match the index color
    CB_v[c] = w   /*set CB with class c which has maximum frequency
    End For
```

↓

ขั้นตอนหลักที่ A3 : Feature Extraction Process

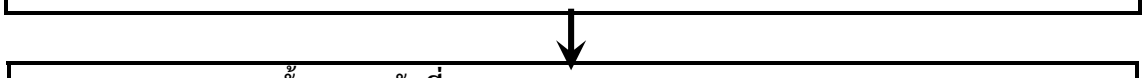
3.1 หาค่าความถูกต้องที่มากที่สุดจากแต่ละตัวแปรเข้า

3.2 กำหนดขอบเขตค่าความผิดพลาด

3.3 กำหนดให้ ϕ คือ ค่าความถูกต้องที่ยอมรับได้

3.4 ϕ = ค่าความถูกต้องที่มากที่สุด – ขอบเขตค่าความผิดพลาด

3.5 ถ้าค่าความถูกต้องจากแต่ละตัวแปรเข้ามากกว่าหรือเท่ากับ ϕ แล้ว ให้ยอมรับค่าความถูกต้องเหล่านั้น และเก็บไว้ใน accepted accuracy list (A)



ขั้นตอนหลักที่ A4 : Generated Knowledge Based

4.1 จาก accepted accuracy list (A) กำหนดให้ *s* คือ จำนวนของ element ทั้งหมดใน list(A) และ *r* คือ จำนวนของ element ทั้งหมดที่นำมา *"and"* กัน

4.2 สร้างกฎที่เป็นไปได้โดยใช้ตัวกระทำการ *"and"* ดังสมการ

$$\binom{s}{r} = \frac{s!}{(s-r)!r!},\ 2 \le r \le s$$

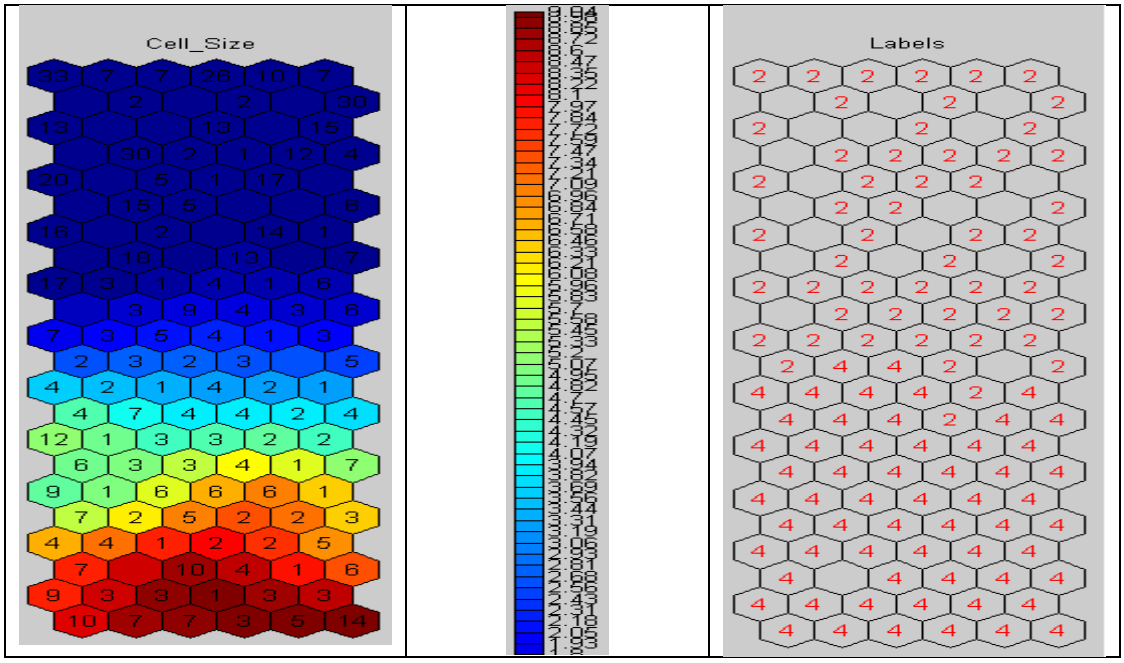
4.3 ถ้าค่าความถูกต้องใหม่ของกฎ > ค่าความถูกต้องที่มากที่สุดแล้ว ให้ยอมรับกฎใหม่

ภาพประกอบ 3.2 รายละเอียดแบบจำลองการสกัดความรู้โดยใช้
แผนที่การจัดกลุ่มเอง (KESOM)

ขั้นตอนหลักที่ A1 เป็นการจัดกลุ่มข้อมูลโดยใช้วิธีการแผนที่การจัดกลุ่มเองจากโปรแกรม Matlab SOM Toolbox ผลลัพธ์ที่ได้จากขั้นตอนนี้ คือ Component Plane ของแต่ละตัวแปรเข้า, Color Bar ที่แสดงค่าข้อมูลเข้าของแต่ละตัวแปรเข้า และ Label ที่แทนด้วยกลุ่มของแต่ละตัวแปรเข้า ดังภาพประกอบ 3.3

ขั้นตอนหลักที่ A2 เป็นการหากลุ่มที่เป็นไปได้โดยการเปลี่ยน Component Plane, Color Bar และ Label เป็นเมทริกซ์ Component Plane และเมทริกซ์ Color Bar ดังภาพประกอบ 3.4 และ 3.5 สำหรับภาพประกอบ 3.4 เป็นการแทน Component Plane ของแต่ละตัวแปรเข้า ด้วยเมทริกซ์ Component Plane (CP Matrix) ซึ่งเป็นเมทริกซ์ที่มี *p* แถว และ *q* คอลัมน์ ค่า *x* คือ ดัชนีค่าสี, *y* คือ กลุ่ม และ *z* คือ ค่าความถี่ ดังอย่างเช่น $CP_{12}[3,1,20]$ ถูกกำหนดด้วย Component Plane แถวที่ 1 คอลัมน์ที่ 2 ซึ่งมีดัชนีค่าสี คือ 3 อยู่ในกลุ่ม 1 และมีความถี่ของข้อมูล คือ 20 สำหรับภาพประกอบ 3.5 แสดงให้เห็นถึงการแทน Color Bar ด้วยเมทริกซ์ Color Bar (CB Matrix) ซึ่ง *a* คือ ค่าสี, *b* คือ ค่าข้อมูลเข้า และ *c* คือ กลุ่มที่เป็นไปได้ และสำหรับการหากลุ่มที่เป็นไปได้นั้น จะพิจารณากลุ่มและค่าความถี่จาก CP Matrix เพื่อทำการหากลุ่มที่เป็นไปได้ของแต่ละดัชนีค่าสี ตัวอย่างเช่น CB_2 [3,5.7,1] ถูกกำหนดด้วย Color Bar

ของตำแหน่งที่ 2 มีค่าสี คือ 3 (สีเขียว) ซึ่งมีค่าข้อมูลเข้า 5.7 และอยู่ในกลุ่มที่เป็นไปได้ คือ กลุ่ม 1



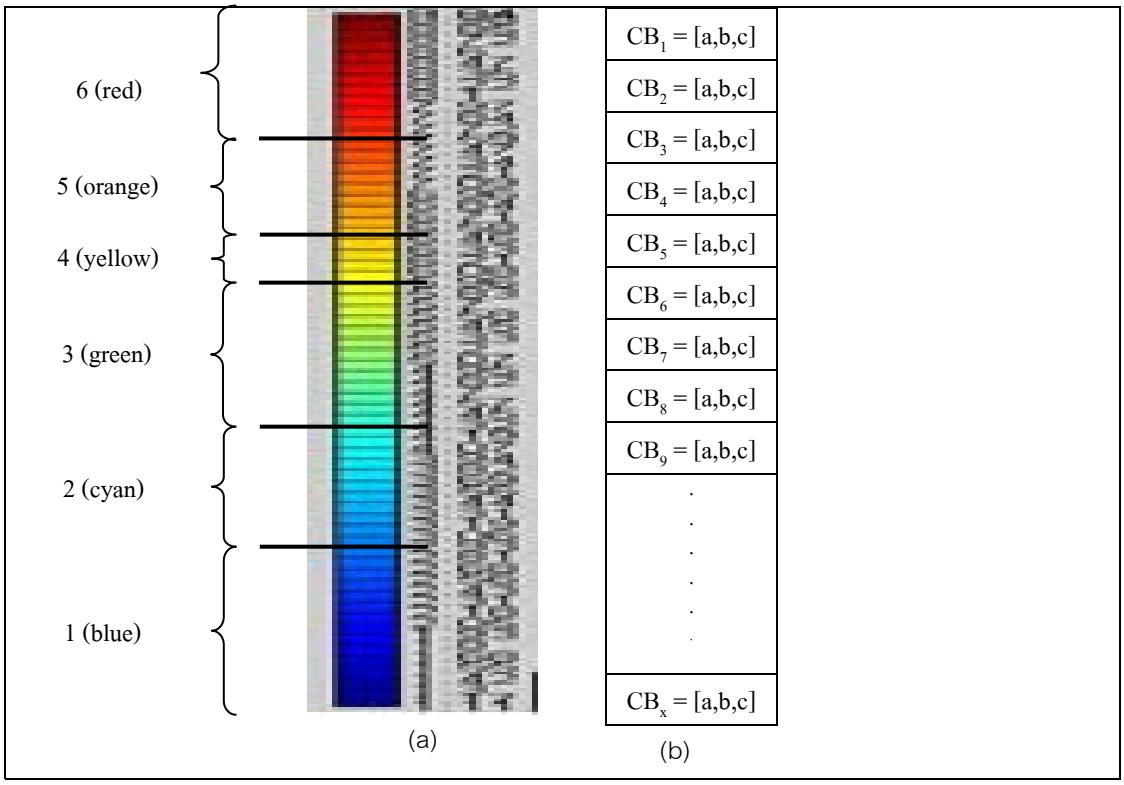
(a) (b) (c)

ภาพประกอบ 3.3 ผลลัพธ์จาก Matlab SOM Toolbox

(a) Component Plane (b) Color Bar (c) Label

$CP_{11}[x,y,z]$	…	$CP_{1q}[x,y,z]$
$CP_{21}[x,y,z]$	…	$CP_{2q}[x,y,z]$
.	.	.
.	.	.
$CP_{p1}[x,y,z]$	…	$CP_{pq}[x,y,z]$

ภาพประกอบ 3.4 เมทริกซ์ Component Plane



ภาพประกอบ 3.5 (a) Color Bar (b) เมทริกซ์ Color Bar

ขั้นตอนหลักที่ A3 เป็นการหาตัวแปรเข้าที่มีค่าความถูกต้องมากที่สุด หลังจากนั้นกำหนดขอบเขตค่าความผิดพลาดซึ่งในที่นี้มีค่า คือ 5% และกำหนดให้ ϕ คือ ค่าความถูกต้องที่ยอมรับได้ ซึ่งคำนวณได้จากสมการ (3.1)

$$\phi = \text{ค่าความถูกต้องที่มากที่สุด} - \text{ขอบเขตค่าความผิดพลาด} \qquad (3.1)$$

ถ้าค่าความถูกต้องจากแต่ละตัวแปรเข้ามากกว่าหรือเท่ากับ ϕ แล้ว ให้ยอมรับค่าความถูกต้องเหล่านั้น และเก็บไว้ใน accepted accuracy list (A) ตัวอย่างเช่น กำหนดให้มีจำนวนตัวแปรเข้าทั้งหมด 4 ตัวแปร ซึ่งแต่ละตัวแปรเข้า มีค่าความถูกต้อง ดังนี้ d_1 = 69.3%, d_2 = 55.3%, d_3 = 94.7% และ d_4 = 96.0% ซึ่ง d_4 เป็นตัวแปรเข้าที่มีค่าความถูกต้องมากที่สุด คือ 96% หลังจากนั้นทำการหาค่าความถูกต้องที่ยอมรับได้ที่คำนวณได้จากสมการ (3.1) ซึ่งค่าความถูกต้องที่ยอมรับได้ (ϕ) คือ 91% (96.0% – 5% = 91%) หลังจากนั้นพิจารณาค่าความถูกต้องแต่ละตัวแปรเข้า ถ้าตัวแปรเข้าที่มีค่าความถูกต้องมากกว่าหรือเท่ากับ ϕ แล้ว ให้ยอมรับค่าความถูกต้องเหล่านั้น และเก็บไว้ใน accepted accuracy list (A) ดังนั้น จากตัวอย่างข้างต้น ตัวแปรเข้าที่ถูกเก็บใน accepted accuracy list (A) คือ d_3 และ d_4 (A = d_3, d_4)

ขั้นตอนหลักที่ A4 เป็นการสร้างกฎที่เป็นไปได้จากตัวกระทำการ *"and"* กำหนดให้ s คือ จำนวนของ element ทั้งหมดใน list(A) และ r คือ จำนวนของ element ทั้งหมดที่นำมา *"and"* กันดังสมการ (3.2)

$$\binom{s}{r} = \frac{s!}{(s-r)!r!},\ 2 \le r \le s \qquad (3.2)$$

ถ้าค่าความถูกต้องใหม่ของกฎมากกว่าค่าความถูกต้องที่มากที่สุดแล้ว ให้ยอมรับกฎใหม่นั้น ตัวอย่างเช่น จาก accepted accuracy list (A) = [d_3, d_4] ซึ่ง s = 2 และ r = 2 และจากสมการ (3.2) ทำให้ได้กฎที่เกิดจากตัวกระทำการ *"and"* เพียงกฎเดียว คือ d_3 *and* d_4 ที่มีค่าความถูกต้อง 98.0% ดังนั้นจึงยอมรับกฎใหม่ที่เกิดขึ้นจากตัวกระทำการ *"and"* เพราะว่าค่าความถูกต้องของกฎใหม่ที่ได้นั้นมากกว่าค่าความถูกต้องเดิม (98.0% > 96.0%)

3.2 แบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (KESOM_MEP)

แบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (Knowledge Extraction from Self-Organizing Map Using Minimization Entropy Principle Algorithm : KESOM_MEP) สามารถสกัดความรู้ได้ในรูปของกฎภาษาธรรมชาติ แสดงได้ดังภาพประกอบ 3.6 ประกอบด้วย 5 ขั้นตอนหลัก ดังนี้ ขั้นตอนหลักที่ B1 : กระบวนการแผนที่การจัดกลุ่มเอง (Self-Organizing Map Process) เป็นการจัดกลุ่มข้อมูลโดยใช้แผนที่การจัดกลุ่มเอง ขั้นตอนหลักที่ B2 : หลักเอ็นโทรพีค่าต่ำสุด (Minimization Entropy Principle Algorithm) เป็นการหาจุดแบ่งข้อมูลเพื่อเตรียมไปสร้างฟัซซี่เซต ขั้นตอนหลักที่ B3 : การสร้างฟัซซี่เซต (Creating Fuzzy Set) เป็นการสร้างฟัซซี่เซตจากจุดแบ่งข้อมูล ขั้นตอนหลักที่ B4 : การสร้างรูปแบบภาษาธรรมชาติ (Creating Linguistic Term Using Membership Function) เป็นการสร้างรูปแบบภาษาธรรมชาติโดยใช้ฟังก์ชันความเป็นสมาชิก และขั้นตอนหลักที่ B5 : การสร้างกฎความรู้ (Knowledge Extraction Using Rough Set) เป็นการสร้างกฎความรู้โดยใช้ราฟเซต รายละเอียดของแบบจำลอง KESOM_MEP แสดงได้ดังภาพประกอบ 3.7

ขั้นตอนหลักที่ B1 : Self-Organizing Map Process
ขั้นตอนหลักที่ B2 : Minimization Entropy Principle Algorithm
ขั้นตอนหลักที่ B3 : Creating Fuzzy Set
ขั้นตอนหลักที่ B4 : Creating Linguistic Term Using Membership Function
ขั้นตอนหลักที่ B5 : Knowledge Extraction Using Rough Set

ภาพประกอบ 3.6 แบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (KESOM_MEP)

ขั้นตอนหลักที่ B1 : Self-Organizing Map Process

1.1 สำหรับแต่ละตัวแปรเข้า d_i ทำการจัดกลุ่มข้อมูลโดยใช้ SOM เมื่อ $1 \leq i \leq n$ และ n คือ จำนวนตัวแปรเข้าทั้งหมด

ผลลัพธ์ที่ได้จากการจัดกลุ่ม คือ component plane, color bar และ label

↓

ขั้นตอนหลักที่ B2 : Minimization Entropy Principle Algorithm

นำผลลัพธ์คือ ค่าสี จาก color bar ของข้อมูลทุกตัวแปร มาทำงานดังต่อไปนี้

2.1 หา $R_{\min}$ คือ ค่าน้อยที่สุด, หา $R_{\max}$ คือ ค่ามากที่สุด, หา σ คือ ค่าเบี่ยงเบนมาตรฐาน และคำนวณช่วงค่าที่ใช้ในการแทนตัวแปรคือ $[R_{\min} - \sigma, R_{\max} + \sigma]$

2.2 ในแต่ละกลุ่มของตัวแปรใด ๆ หา $C_{\min}$ คือ ค่าน้อยที่สุดของกลุ่ม และ $C_{\max}$ คือ ค่ามากที่สุดของกลุ่ม

2.3 คำนวณหาค่า potential threshold point (x)

2.3.1 คำนวณหาค่า entropy ของแต่ละ potential threshold point จากสมการ

$$S(x) = p(x)S_p(x) + q(x)S_q(x)$$

โดยที่ $S(x)$ คือ ค่า entropy ของ potential threshold point, $S_p(x)$ คือ ค่า entropy ของขอบเขต p, $S_q(x)$ คือ ค่า entropy ของขอบเขต q, $p(x)$ คือ ค่าสัดส่วนของข้อมูลทั้งหมดในขอบเขต p และ $q(x)$ คือ ค่าสัดส่วนของข้อมูลทั้งหมดในขอบเขต q

2.3.2 เลือก potential threshold point ที่มีค่า entropy น้อยที่สุด มาเก็บไว้ member_list (ptp)

2.3.3 กำหนดช่วงค่าใหม่ในการหาค่า entropy ที่น้อยที่สุดต่อไปโดยใช้ค่าขอบเขต p และ ขอบเขต q

2.3.4 และทำซ้ำข้อ 2.3.1 จนกระทั่งค่า entropy เท่ากับศูนย์ หรือได้ค่า potential threshold point เป็นค่าเดิม

↓

ขั้นตอนหลักที่ B3 : Creating Fuzzy Set

สร้างฟัซซี่เซต จาก member_list $(\text{ptp}_j), 1 \le j \le n$ โดยที่ n คือจำนวนทั้งหมดของ potential threshold point ใน member_list

3.1 สร้างช่วงค่าแรกเป็นรูปสี่เหลี่ยมคางหมู กำหนดเทอมเซตของช่วงค่าแรก คือ term1 มีค่าช่วง คือ $[R_{\min}, ptp_2]$

3.2 สร้างช่วงค่าที่สองถึงช่วงค่าที่ $n-1$ เป็นรูปสามเหลี่ยม และกำหนดเทอมเซตจาก term2 ถึง mf(n-1) โดยที่ $1 \le k \le n-2$ มีค่าช่วง คือ $[ptp_k, ptp_{k+2}]$,..., $[ptp_{n-2}, ptp_n]$ ตามลำดับ

3.3 สร้างช่วงค่าสุดท้ายเป็นรูปสี่เหลี่ยมคางหมู กำหนดเทอมเซต ช่วงสุดท้ายคือ term(n) มีค่าช่วง คือ $[ptp_{n-1}, R_{\max}]$

ขั้นตอนหลักที่ B4 : Creating Linguistic Term Using Membership Function

แปลงค่าสี จาก color bar ของข้อมูลทุกตัวแปรจากข้อ 1.1 และ เทอมเซตจากข้อ 3 เพื่อสร้างเป็นช่วงค่าภาษาธรรมชาติโดยดูค่าความเป็นสมาชิกที่เป็นไปได้ว่าอยู่ในช่วงค่าใด

4.1 สมการที่ใช้ในการหาค่าฟังก์ชันความเป็นสมาชิก คือ

4.1.1 ในกรณีที่ฟังก์ชันความเป็นสมาชิกเป็นรูปสี่เหลี่ยมคางหมู

$$\mu_A(v_r) = \begin{cases} \frac{1}{b-a}(v_r - a); a \le v_r \le b, a \ne b \\ 1; b < v_r \le c \\ \frac{1}{d-c}(d - v_r); c < v_r \le d, c \ne d \\ 0 \text{ ; สำหรับกรณีอื่น ๆ} \end{cases}$$

4.1.2 ในกรณีที่ฟังก์ชันความเป็นสมาชิกเป็นรูปสามเหลี่ยม

$$\mu_A(v_r) = \begin{cases} \frac{1}{b-a}(v_r - a); a \le v_r \le b, a \ne b \\ \frac{1}{c-b}(c - v_r); b < v_r \le c, b \ne c \\ 0 \text{ ; สำหรับกรณีอื่น ๆ} \end{cases}$$

4.2 สำหรับช่วงค่าภาษาธรรมชาติที่สามารถแทนได้ 2 ค่าเช่นมีโอกาสเป็นทั้ง “term1” และ “term2” ให้เลือกช่วงค่าภาษาธรรมชาติของค่า color bar ของข้อมูลตัวแปร ที่มีระดับความเป็นสมาชิกมากกว่า

ขั้นตอนหลักที่ B5 : Knowledge Extraction Using Rough Set

5.1 กำหนด ϕ คือ ค่าความถูกต้องที่ยอมรับได้ โดยที่ $\phi > 50\%$

5.2 สร้างกฎ ถ้า-แล้ว จากหลักการประมาณค่าขอบเขตล่าง ดังนี้

$$\underline{R}X = \{x \in U : [x]_R \subseteq X\}$$

5.3 สร้างกฎ ถ้า-แล้ว จากหลักการประมาณค่าขอบเขตบน ดังนี้

$$\overline{R}X = \{x \in U : [x]_R \cap X \ne \phi\}$$

ภาพประกอบ 3.7 รายละเอียดแบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (KESOM_MEP)

ขั้นตอนหลักที่ B1 เป็นการจัดกลุ่มข้อมูลโดยใช้วิธีการแผนที่การจัดกลุ่มเองจากโปรแกรม Matlab SOM Toolbox ผลลัพธ์ที่ได้จากขั้นตอนนี้ คือ Component Plane ของแต่ละตัวแปรเข้า, Color Bar ที่แสดงค่าข้อมูลเข้าของแต่ละตัวแปรเข้า และ Label ที่แทนด้วยกลุ่มของแต่ตัวแปรเข้า

ขั้นตอนหลักที่ B2 เป็นขั้นตอนการหา potential threshold point หรือจุดแบ่งข้อมูลของแต่ละตัวแปรเข้า ซึ่งมีรายละเอียดขั้นตอนดังนี้

2.1 หาค่า $R_{\min}$ คือ ค่าน้อยที่สุด, หาค่า $R_{\max}$ คือ ค่ามากที่สุด, หาค่า σ คือ ค่าเบี่ยงเบนมาตรฐาน และคำนวณช่วงค่าที่ใช้ในการแทนตัวแปร ดังสมการ (3.3)

$$[R_{\min} - \sigma, R_{\max} + \sigma] \tag{3.3}$$

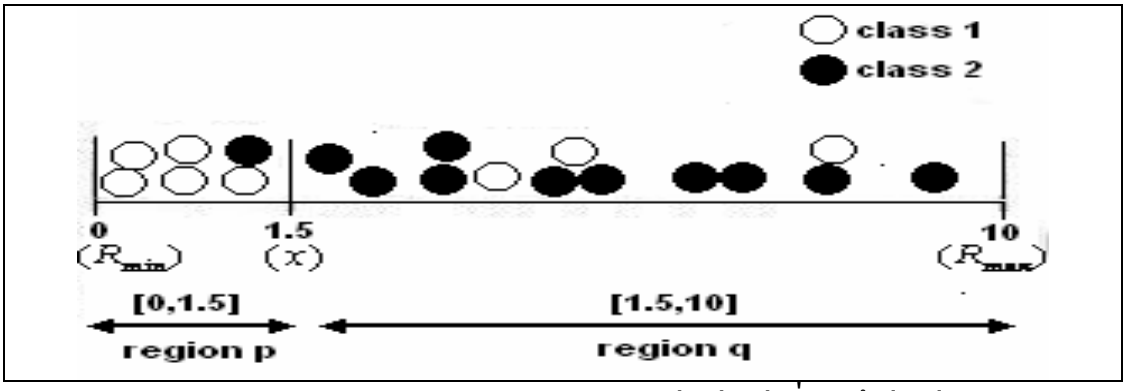
2.2 พิจารณาค่าแยกตามกลุ่ม หาค่าน้อยที่สุดของกลุ่ม $C(l)_{min}$ และค่ามากที่สุดของกลุ่ม $C(l)_{max}$ของแต่ละตัวแปร โดย $1 \le l \le class$ และ $class$ คือ จำนวนกลุ่มทั้งหมด ตัวอย่างเช่น มีจำนวน 2 กลุ่ม กลุ่มที่ 1 มีค่าความเข้มของสีจาก color bar คือ {6, 6, 3, 3, 9, 5, 4, 7, 8, 9} และกลุ่มที่ 2 มีค่าความเข้มของสีจาก color bar คือ {2, 1, 3, 2, 3, 4, 4, 5, 4, 5} ดังนั้นค่าน้อยที่สุดของกลุ่มที่ 1 $C(1)_{min}$ คือ 3 และค่ามากที่สุดของกลุ่ม 1 $C(1)_{max}$ คือ 9 ค่าน้อยที่สุดของกลุ่ม 2 $C(2)_{min}$ คือ 1 และค่ามากที่สุดของกลุ่ม 2 $C(2)_{max}$ คือ 5 ตามลำดับ

2.3 คำนวณหาค่า potential threshold point จากข้อมูลแต่ละตัวแปร กำหนดให้ ค่า x คือ potential threshold point

1) เรียงลำดับข้อมูลค่าความเข้มของสีจากค่าน้อย ไปมากแยกตามกลุ่ม ตัวอย่างกลุ่มที่ 1 คือ {3, 3, 4, 5, 6, 6, 7, 8, 9, 9} และ กลุ่มที่ 2 คือ {1, 2, 2, 3, 3, 4, 4, 4, 5, 5} เป็นต้น

2) หาค่า Potential threshold point หรือเป็นจุดแบ่งข้อมูล ซึ่งพิจารณาจากการคาบเกี่ยวกันของกลุ่มข้อมูล ดังนี้ ในกรณีช่วงค่าของแต่ละกลุ่ม*ไม่คาบเกี่ยวกัน* กำหนดค่า potential threshold point คือค่ากลางของตัวเลขสองตัวที่ติดกันในกลุ่มนั้น ๆ แต่ในกรณีที่มีช่วงค่าของกลุ่ม*ข้อมูลคาบเกี่ยวกัน*ให้นำค่าจากทั้งสองกลุ่มข้อมูลมา Union กันโดยไม่ต้องพิจารณาค่าซ้ำ แล้วกำหนดค่า potential threshold point คือ ค่ากลางของตัวเลขสองตัวที่อยู่ติดกัน จากตัวอย่างข้างต้น เนื่องจากกลุ่มที่ 1 และ 2 มีช่วงค่าข้อมูลคาบเกี่ยวกันตั้งแต่ 3 ถึง 5 จะได้ $C(1) \cup C(2)$ คือ {1, 2, 3, 4, 5, 6, 7, 8, 9} และค่า potential threshold point คือ ค่ากลางของตัวเลขสองตัวที่ติดกันคือ {1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5} = {x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8} ตามลำดับ

2.3.1 คำนวณหาค่า entropy ของแต่ละ potential threshold point กำหนดให้ $S(x)$ คือ ค่า entropy ของแต่ละ potential threshold point ค่า entropy ของแต่ละ potential threshold pointซึ่งคำนวณได้ดังสมการ (3.4) ในการพิจารณาจะต้องพิจารณาค่าช่วงของ *ขอบเขต p* หมายถึง ค่าระหว่าง R_{min}และ x และ *ขอบเขต q* หมายถึง ค่าระหว่าง x และ R_{max} คำนวณได้จากดังสมการที่ (3.5) และ (3.6) ตามลำดับ ตัวอย่างกำหนดให้ x มีค่า potential threshold point = 1.5 ภาพประกอบ 3.8 แสดงค่าขอบเขต p คือ [0,1.5] และขอบเขต q คือ [1.5,10]



ภาพประกอบ 3.8 Potential threshold point ระหว่างช่วงค่าต่ำสุดถึงช่วงค่าสูงสุด

$$S(x) = p(x)S_p(x) + q(x)S_q(x) \quad (3.4)$$

ขอบเขต p = $[R_{\min} - \sigma, x]$ (3.5)

ขอบเขต q = $[x, R_{\max} + \sigma]$ (3.6)

จากสมการที่ (3.4) p(x) คือ สัดส่วนของข้อมูลทั้งหมดในขอบเขต p, $q(x)$ คือ สัดส่วนของข้อมูลทั้งหมดในขอบเขต q โดยที่ $p(x)+q(x) = 1$ เสมอ

โดยแต่ละขอบเขตสามารถกำหนดค่า entropy ของแต่ละขอบเขตได้ กำหนดให้ $S_p(x)$ คือ ค่า entropy ของขอบเขต p และ $S_q(x)$ คือ ค่า entropy ของขอบเขต q ซึ่งสามารถคำนวณได้จากสมการที่ (3.7) และ (3.8) ตามลำดับ

$$S_p(x) = -(p_1(x)\ln p_1(x) + p_2(x)\ln p_2(x)) \quad (3.7)$$

$$S_q(x) = -(q_1(x)\ln q_1(x) + q_2(x)\ln q_2(x)) \quad (3.8)$$

โดยที่ค่า $p_k(x)$ คือ ความเป็นไปได้ที่ข้อมูลกลุ่ม k ที่อยู่ในขอบเขต p และ $q_k(x)$ คือ ความเป็นไปได้ที่ข้อมูลกลุ่ม k อยู่ในขอบเขต q คำนวณได้ดังสมการที่ (3.9) ถึง (3.12) ตามลำดับ

$$p_k(x) = \frac{n_k(x)+1}{n(x)+1} \quad (3.9)$$

$$p(x) = \frac{n(x)+1}{n+1} \quad (3.10)$$

$$q_k(x) = \frac{m_k(x)+1}{m(x)+1} \quad (3.11)$$

$$q(x) = \frac{m(x)+1}{n+1} \quad (3.12)$$

โดยที่ $n_k(x)$ คือ จำนวนข้อมูลกลุ่ม k ที่อยู่ในขอบเขต p, $m_k(x)$ คือ จำนวนข้อมูลกลุ่ม k ที่อยู่ในขอบเขต q, $n(x)$ คือ จำนวนข้อมูลทั้งหมดที่อยู่ในขอบเขต p, $m(x)$ คือ จำนวนข้อมูลทั้งหมดที่อยู่ในขอบเขต q, และ n คือ จำนวนข้อมูลทั้งหมดที่อยู่ในขอบเขต p และขอบเขต q รวมกัน

ตัวอย่างการคำนวณหาค่า entropy ของ potential threshold point ($x=1.5$) จากภาพประกอบ 3.8 จากขอบเขต p, ค่า $n = 19$, $n_1(1.5) = 5$, $n_2(1.5) = 1$ และ $n(1.5) = 6$ จากขอบเขต q, ค่า $m_1(1.5) = 3$, $m_2(1.5) = 10$, $m(1.5) = 13$ จากสมการ (3.9) ถึง (3.12) จะได้ค่า $p_1(1.5) = 0.86$, $p_2(1.5) = 0.29$, $p(1.5) = 0.35$, $q_1(1.5) = 0.29$, $q_2(1.5) = 0.79$ และ $q(1.5) = 0.65$ จากสมการ (3.7) และ (3.8) จะได้ค่า $S_p(1.5) = 0.49$ และ $S_q(1.5) = 0.55$ คำนวณค่า entropy จากสมการ (3.4) ได้ $S(1.5) = 0.53$ หมายถึง ค่า entropy ของ potential threshold point 1.5 มีค่าเท่ากับ 0.53

2.3.2 เก็บค่า potential threshold point หรือค่า x ที่มีค่า entropy น้อยที่สุดไว้ใน member_list

2.3.3 กำหนดช่วงค่าใหม่ในการหาค่า entropy ที่น้อยที่สุดต่อไป โดยใช้ค่าขอบเขต p และ ขอบเขต q

2.3.4 ทำซ้ำข้อ 2.3.1 จนกระทั่งค่า entropy มีค่าเท่ากับศูนย์ หรือคำนวณได้ค่า potential threshold point เป็นค่าเดิม

ขั้นตอนหลักที่ B3 นำค่า potential threshold point จาก member_list(ptp_j) ซึ่ง $1 \le j \le n$ โดยที่ n คือจำนวนทั้งหมดของ potential threshold point ไปสร้างฟัซซี่เซต จากตัวอย่างข้างต้น member_list มีทั้งหมด 4 ค่า คือ 1.5, 3.5, 6.5 และ 8.5

1. สร้างช่วงค่าแรกเป็นรูปสี่เหลี่ยมคางหมูจากช่วงค่า $[R_{min}, ptp_2]$ ดังสมการ (3.13) โดยที่ R_{min} คือ ช่วงค่าน้อยที่สุด และ ptp_j คือ potential threshold point ใน member_list กำหนดเทอมเซต คือ “term1” ดังภาพประกอบ 3.9

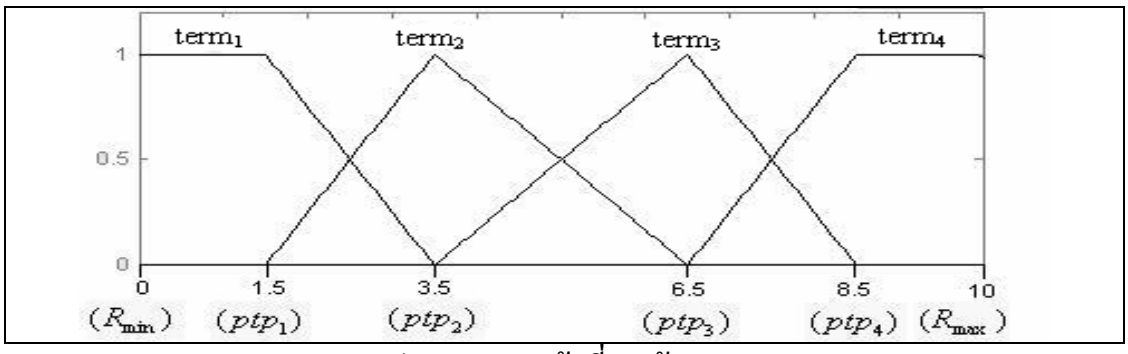
$$\mu_A(v_r) = \begin{cases} 1;\ R_{\min} \le v_r \le ptp_1 \\ \dfrac{1}{(ptp_2 - ptp_1)}(ptp_2 - v_r);\ ptp_1 < v_r \le ptp_2 \\ 0;\ \text{สำหรับกรณีอื่นๆ} \end{cases} \quad (3.13)$$

2. สร้างช่วงค่าที่สองถึงช่วงค่าที่ *n-1* เป็นรูปสามเหลี่ยมจากช่วงค่า $[ptp_k,ptp_{k+2}]$,.., $[ptp_{n-2},ptp_n]$ ดังสมการ (3.14) โดยที่ ptp_j คือ potential threshold point ใน member_list กำหนดเทอมเซตจาก term2 ถึง term(n-1) โดยที่ $1 \le k \le n-2$ ดังภาพประกอบ 3.9 จะเห็นได้ว่ามี 2 ช่วงค่าคือ ช่วงค่า $[ptp_1,ptp_3]$ หรือเท่ากับ [1.5,6.5] มีเทอมเซต คือ "term2" และช่วงค่า $[ptp_2,ptp_4]$ หรือเท่ากับ [3.5,8.5] มีเทอมเซต คือ "term3"

$$\mu_A(v_r) = \begin{cases} \dfrac{1}{(ptp_{k+1} - ptp_k)}(v_r - ptp_k);\ ptp_k \le v_r \le ptp_{k+1} \\ \dfrac{1}{(ptp_{k+2} - ptp_{k+1})}(ptp_{k+2} - v_r);\ ptp_{k+1} < v_r \le ptp_{k+2} \\ 0;\ R_{\min} \le v_r \le ptp_1, v_r > ptp_n \end{cases} \quad (3.14)$$

3. สร้างช่วงสุดท้ายเป็นรูปสี่เหลี่ยมคางหมูจากช่วงค่า $[ptp_{n-1},R_{max}]$ ดังสมการ (3.15) โดยที่ R_{max} คือ ช่วงค่ามากที่สุด และ ptp_j คือ potential threshold point ใน member_list กำหนดเทอมเซต คือ "term(n)" ดังภาพประกอบ 3.9

$$\mu_A(v_r) = \begin{cases} \dfrac{1}{(ptp_n - ptp_{n-1})}(v_r - ptp_{n-1});\ ptp_{n-1} \le v_r \le ptp_n \\ 1;\ ptp_n < v_r \le R_{max} \\ 0;\ \text{สำหรับกรณีอื่นๆ} \end{cases} \quad (3.15)$$



ภาพประกอบ 3.9 ฟัซซีเซตกับเทอมเซต

ขั้นตอนหลักที่ B4 แปลงค่าสี จาก color bar ของข้อมูลทุกตัวแปรเพื่อสร้างเป็นช่วงค่าภาษาธรรมชาติโดยดูค่าความเป็นสมาชิกที่เป็นไปได้ว่าอยู่ในช่วงค่าใด โดยสมการที่ใช้ในการประมาณค่าความเป็นสมาชิกของแต่ละช่วงค่าภาษาธรรมชาติ คือ ในกรณีฟังก์ชันความเป็นสมาชิกเป็นรูปสี่เหลี่ยมคางหมู พิจารณาดังสมการ (3.16) และในกรณีฟังก์ชันความเป็นสมาชิกเป็นรูปสามเหลี่ยม พิจารณาดังสมการ (3.17)

$$\mu_A(v_r) = \begin{cases} \frac{1}{b-a}(v_r - a); a \le v_r \le b, a \ne b \\ 1;\ b < v_r \le c \\ \frac{1}{d-c}(d - v_r); c < v_r \le d, c \ne d \\ 0; \text{สำหรับกรณีอื่นๆ} \end{cases} \tag{3.16}$$

$$\mu_A(v_r) = \begin{cases} \frac{1}{b-a}(v_r - a); a \le v_r \le b, a \ne b \\ \frac{1}{c-b}(c - v_r); b < v_r \le c, b \ne c \\ 0; \text{สำหรับกรณีอื่นๆ} \end{cases} \tag{3.17}$$

กำหนดให้ $a \le b \le c \le d$ เมื่อ a, b, c, d เป็นpotential threshold point ใดๆ ที่เก็บอยู่ใน member_list และ v_r เป็นค่าสี จาก color bar ของข้อมูลทุกตัวแปร สำหรับช่วงค่าภาษาธรรมชาติที่สามารถแทนได้ 2 ค่า เช่น มีโอกาสเป็นทั้ง “term1” และ “term2” ให้เลือกช่วงค่าภาษาธรรมชาติของค่า Color Bar ของข้อมูลตัวแปร ที่มีระดับความเป็นสมาชิกมากกว่า ตัวอย่างจากภาพประกอบ 3.9 ถ้าต้องการทราบว่า “3” อยู่ในช่วงค่าภาษาธรรมชาติ “term1” หรือ “term2” จากสมการ (3.16) ช่วงค่าภาษาธรรมชาติ “term1” กำหนดให้ v_r =3, a =0, b =1.5, c =3.5, d =6.5 ดังนั้น ระดับค่าความเป็นสมาชิก = 0.25 สำหรับช่วงค่าภาษาธรรมชาติ “term2” จากสมการ (3.17) กำหนดให้ v_r =3, a =1.5, b =3.5, c =6.5 ดังนั้น ระดับค่าความเป็นสมาชิก = 0.75 แสดงว่า “3” อยู่ในช่วงค่าภาษาธรรมชาติ “term2” เนื่องจากมีระดับค่าความเป็นสมาชิกมากกว่า

ขั้นตอนหลักที่ B5 สำหรับการสร้างกฎถ้า-แล้ว จากหลักการประมาณค่าขอบเขตล่างดังสมการ (3.18) และการประมาณค่าขอบเขตบนดังสมการ (3.19) กฎที่ยอมรับได้ต้องมีค่าความมั่นใจมากกว่า 50%

$$\underline{B}X = \{x \in U : B(x) \subseteq X\} \quad (3.18)$$

$$\overline{B}X = \{x \in U : B(x) \cap X \neq \phi\} \quad (3.19)$$

ตารางที่ 3.1 แสดงตัวอย่างค่าสี จาก Color Bar ของข้อมูลตัวแปร d_1 นำข้อมูลจากตารางที่ 3.1 แปลงเป็นความสัมพันธ์ระหว่างกลุ่ม (class) กับหมายเลขแถว (No.) ดังตารางที่ 3.2(a) และแปลงเป็นความสัมพันธ์ระหว่างช่วงค่าภาษาธรรมชาติ (linguistic interval) กับหมายเลขแถว (No.) ดังตารางที่ 3.2(b)

ตารางที่ 3.1 ตัวอย่างข้อมูลของตัวแปรเข้า d_1

No.	Linguistic Interval	Class
1	$term_1$	1
2	$term_1$	1
3	$term_2$	1
4	$term_2$	2
5	$term_2$	2

ตารางที่ 3.2 ความสัมพันธ์ (a) class กับ No. (b) linguistic interval กับ No.

Class	No.
1	1,2,3
2	4,5

(a)

Linguistic Interval	No.
$term_1$	1,2
$term_2$	3,4,5

(b)

หากฎจากการประมาณค่าขอบเขตล่างและค่าความมั่นใจของกฎจากสมการ (3.18) ได้ กฎ R1 มีค่าความมั่นใจ 100% การประมาณค่าขอบเขตบนและค่าความมั่นใจของกฎจากดังสมการ (3.19) จะได้ กฎ R2 และ R3 ที่มีค่าความมั่นใจ 67% และ 33% ตามลำดับดังนี้

R1: If d_1 = $term_1$ then class 1 (100%)

R2: If d_1 = $term_2$ then class 2 (67%)

R3: If d_1 = $term_2$ then class 1 (33%)

กฎที่สามารถยอมรับได้ต้องมีค่าความมั่นใจมากกว่า 50% ดังนั้นกฎที่ยอมรับได้มีเพียง 2 กฎ คือ R1 และ R2 จากตัวแปรเข้าที่ได้จากการเลือกกฎที่มีค่าความถูกต้องสูงที่สุด