บทที่ 4

ผลการทดลองและวิจารณ์

วิทยานิพนธ์นี้ได้ทดลองกับฐานข้อมูลที่เป็นมาตรฐานจาก University of California at Irvine [26] เป็นฐานข้อมูลระดับสากลได้แก่ ฐานข้อมูลดอกไม้ไอริส (Iris Flower Database) ฐานข้อมูลโรคมะเร็งเต้านม (Wisconsin Breast Cancer Database) ฐานข้อมูลโรคหัวใจจากโปรเจ็กต์สเตตล็อคยุโรป (Heart Disease from European Statlog Project Database) และฐานข้อมูลโรคเบาหวาน (Pima Indians Diabetes Database) ซึ่งใช้แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง (KESOM) [30] ความรู้ที่สกัดได้อยู่ในรูปของกฎทั่วไป และแบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (KESOM_MEP) [31] ความรู้ที่สกัดได้อยู่ในรูปของกฎภาษาธรรมชาติ

4.1 ฐานข้อมูลดอกไม้ไอริส

ฐานข้อมูลดอกไม้ไอริส (Iris Flower Database) ประกอบด้วยดอกไม้ 3 ชนิด คือ ดอกซีโทซ่า (Setosa) ดอกเวอร์ซิคัลเลอร์ (Versicolor) และดอกเวอร์จินิคา (Virginica) แต่ละชนิดมีข้อมูลจำนวน 50 ตัว รวมทั้งหมด 150 ตัว ซึ่งมี 4 ตัวแปรเข้า คือ x_1 คือความยาวกลีบเลี้ยง (Sepal length), x_2 คือความกว้างกลีบเลี้ยง (Sepal width), x_3 คือความยาวกลีบดอก (Petal length), และ x_4 คือความกว้างกลีบดอก (Petal width) ดังตารางที่ 4.1

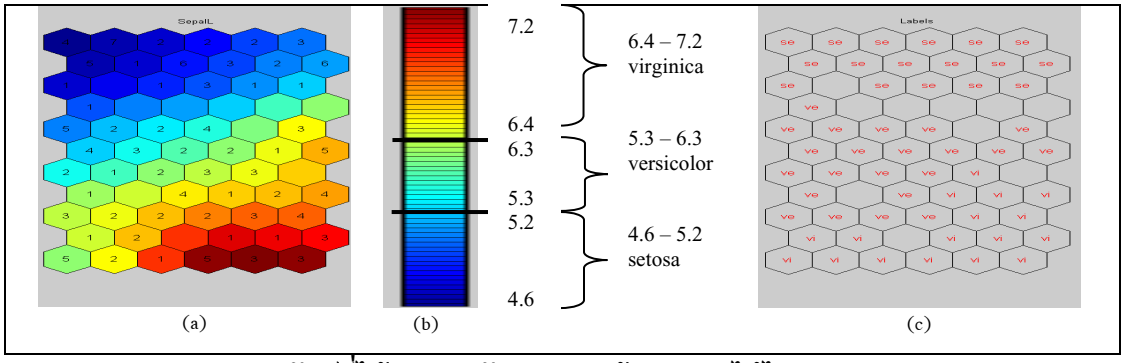
ตารางที่ 4.1 รายละเอียดของฐานข้อมูลดอกไม้ไอริส

ตัวแปร	ชื่อตัวแปร	ชนิด	ค่าที่เป็นไปได้
x_1	Sepal length	Numeric	4.3-7.9
x_2	Sepal width	Numeric	2.0-4.4
x_3	Petal length	Numeric	1.0-6.9
x_4	Petal width	Numeric	0.1-2.5
c	class	Nominal Scale	1 = Setosa 2 = Versicolor 3 = Virginica

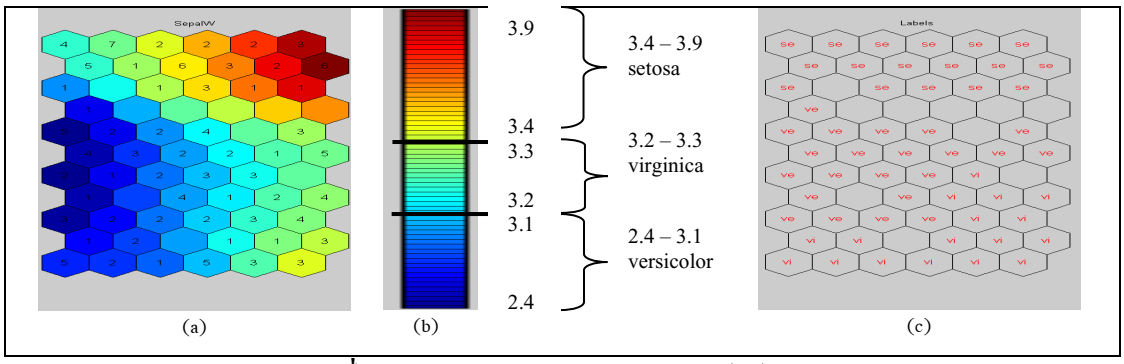
4.1.1 แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง (KESOM)

สำหรับการสกัดความรู้จากฐานข้อมูลดอกไม้ไอริสที่อยู่ในรูปกฎทั่วไปนั้น มีการทำงาน 4 ขั้นตอนหลัก ตามรายละเอียดดังนี้

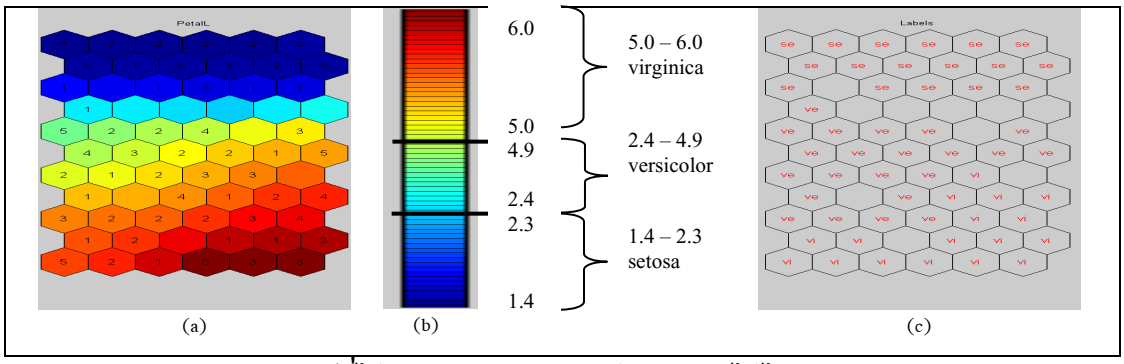
ขั้นตอนหลักที่ A1 กระบวนการแผนที่การจัดกลุ่มเอง (Self-Organizing Map Process) ทำการแบ่งกลุ่มข้อมูลด้วย Self-Organizing Map โดยใช้ Matlab SOM Toolbox เป็นเครื่องมือ จะได้ Component Plane, Color Bar และ Label ซึ่งเป็นเครื่องมือสำหรับวิเคราะห์ผลลัพธ์ที่ได้จากการจัดกลุ่ม ดังภาพประกอบ 4.1 ถึง ภาพประกอบ 4.4 สำหรับ Component Plane เป็นผลลัพธ์ที่ได้จากการจัดกลุ่มข้อมูลในแต่ละตัวแปรเข้า ของข้อมูล และมี Color Bar ที่แสดงค่าข้อมูลที่ได้จากการจัดกลุ่มข้อมูล สำหรับ Label นั้นเป็นส่วนที่แสดงกลุ่มที่ได้จากการจัดกลุ่มข้อมูล



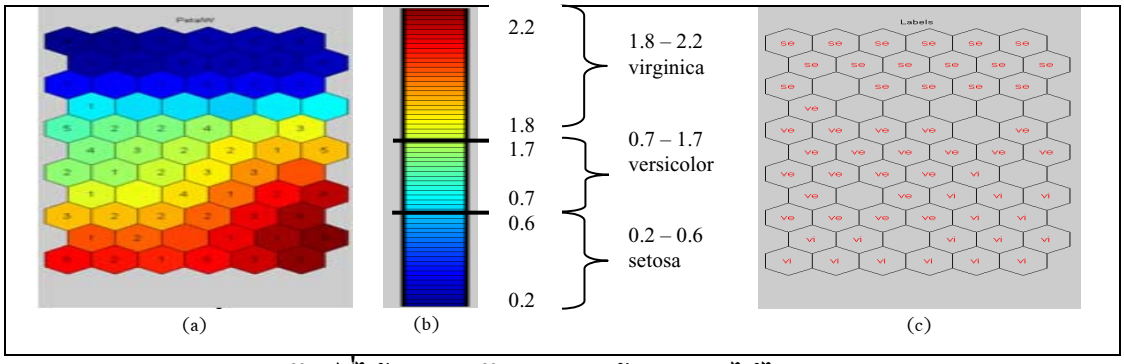
ภาพประกอบ 4.1 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลดอกไม้ไอริส (a) Component Plane ของ x_1 (b) ช่วงค่า Color Bar ของ x_1 (c) Label ของกลุ่มทั้ง 3 กลุ่ม (se, ve, vi)



ภาพประกอบ 4.2 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลดอกไม้ไอริส (a) Component Plane ของ x_2 (b) ช่วงค่า Color Bar ของ x_2 (c) Label ของกลุ่มทั้ง 3 กลุ่ม (se, ve, vi)



ภาพประกอบ 4.3 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลดอกไม้ไอริส (a) Component Plane ของ x_3 (b) ช่วงค่า Color Bar ของ x_3 (c) Label ของกลุ่มทั้ง 3 กลุ่ม (se, ve, vi)



ภาพประกอบ 4.4 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลดอกไม้ไอริส (a) Component Plane ของ x_4 (b) ช่วงค่า Color Bar ของ x_4 (c) Label ของกลุ่มทั้ง 3 กลุ่ม (se, ve, vi)

จากภาพประกอบ 4.1(a) ถึง ภาพประกอบ 4.4(a) แสดง Component Plane ของแต่ละข้อมูลเข้า คือ Sepal length, Sepal width, Petal length และ Petal width ตามลำดับ ภาพประกอบ 4.1(b) ถึงภาพประกอบ 4.4(b) แสดง Color Bar ที่บอกค่าข้อมูลที่ได้จากการจัดกลุ่มข้อมูล ดังตารางที่ 4.2 และภาพประกอบ 4.1(c) ถึง ภาพประกอบ 4.4(c) แสดง Label ที่บอกกลุ่มของข้อมูล

ตารางที่ 4.2 รายละเอียด Color Bar ที่ได้จากการจัดกลุ่มข้อมูล

ตัวแปร	ชื่อตัวแปร	Color Bar ที่ได้จากการจัดกลุ่ม
x_1	Sepal length	4.6-7.2
x_2	Sepal width	2.4-3.9
x_3	Petal length	1.4-6.0
x_4	Petal width	0.2-2.2

ขั้นตอนหลักที่ A2 การวิเคราะห์ Component Plane และ Label (Component Plane and Label Analysis) เป็นการหากลุ่มที่เป็นไปได้ ซึ่งกำหนดให้ดัชนีค่าสีมีค่าตั้งแต่ 1-64 และกำหนดให้มีค่าสีทั้งหมด 6 ค่า ดังตารางที่ 4.3 หลังจากนั้นทำการเปลี่ยน Component Plane เป็น เมทริกซ์ Component Plane ดังตารางที่ 4.4 ถึง 4.7 เป็นการแทน Component Plane ของแต่ละตัวแปรเข้าด้วยเมทริกซ์ Component Plane (CP Matrix) ซึ่งเป็นเมทริกซ์ที่มี p แถว และ q คอลัมน์ ค่า x คือ ดัชนีค่าสี, y คือ กลุ่ม และ z คือ ค่าความถี่ ดังอย่างเช่น $CP_{1,6}[21,2,4]$ ถูกกำหนดด้วย Component Plane แถวที่ 1 คอลัมน์ที่ 6 ซึ่งมีดัชนีค่าสี คือ 21 อยู่ในกลุ่ม 2 และมีความถี่ของข้อมูล คือ 4

ตารางที่ 4.3 ค่าสี

สี	ค่าสี
น้ำเงิน	101
ฟ้า	102
เขียว	103
เหลือง	104
ส้ม	105
แดง	106

ตารางที่ 4.4 เมทริกซ์ Component Plane ของ x_1

$CP_{1,1}[1,1,4]$	$CP_{1,2}[3,1,5]$	$CP_{1,3}[5,1,1]$	$CP_{1,4}[13,2,1]$	$CP_{1,5}[16,2,5]$
$CP_{2,1}[25,2,2]$	$CP_{2,2}[34,2,1]$	$CP_{2,3}[36,2,3]$	$CP_{2,4}[37,3,1]$	$CP_{2,5}[33,3,5]$
$CP_{3,1}[5,1,1]$	$CP_{3,2}[7,0,0]$	$CP_{3,3}[16,0,0]$	$CP_{3,4}[20,2,2]$	$CP_{3,5}[25,2,3]$
$CP_{4,1}[37,0,0]$	$CP_{4,2}[39,2,2]$	$CP_{4,3}[44,3,2]$	$CP_{4,4}[40,3,2]$	$CP_{4,5}[5,1,2]$
$CP_{5,1}[10,1,1]$	$CP_{5,2}[18,0,0]$	$CP_{5,3}[23,2,2]$	$CP_{5,4}[29,2,2]$	$CP_{5,5}[33,2,2]$
$CP_{6,1}[44,2,2]$	$CP_{6,2}[53,0,0]$	$CP_{6,3}[53,3,1]$	$CP_{6,4}[8,1,2]$	$CP_{6,5}[12,1,3]$
$CP_{7,1}[21,0,0]$	$CP_{7,2}[26,2,4]$	$CP_{7,3}[33,2,2]$	$CP_{7,4}[36,2,3]$	$CP_{7,5}[43,3,1]$
$CP_{8,1}[58,3,1]$	$CP_{8,2}[63,3,5]$	$CP_{8,3}[12,1,2]$	$CP_{8,4}[16,1,2]$	$CP_{8,5}[17,0,1]$
$CP_{9,1}[32,0,0]$	$CP_{9,2}[39,3,1]$	$CP_{9,3}[39,3,3]$	$CP_{9,4}[44,3,2]$	$CP_{9,5}[50,3,3]$
$CP_{10,1}[64,3,3]$	$CP_{10,2}[15,1,3]$	$CP_{10,3}[18,1,6]$	$CP_{10,4}[21,1,1]$	$CP_{10,5}[33,0,0]$
$CP_{11,1}[45,2,5]$	$CP_{11,2}[43,0,0]$	$CP_{11,3}[46,3,4]$	$CP_{11,4}[50,3,4]$	$CP_{11,5}[56,3,3]$

ตารางที่ 4.5 เมทริกซ์ Component Plane ของ x_2

$CP_{1,1}$[27,1,4]	$CP_{1,2}$[27, 1,5]	$CP_{1,3}$[17, 1,1]	$CP_{1,4}$[7, 2,1]	$CP_{1,5}$[1, 2,5]
$CP_{2,1}$[1, 2,2]	$CP_{2,2}$[3,2,1]	$CP_{2,3}$[2,2,3]	$CP_{2,4}$[8,3,1]	$CP_{2,5}$[10,3,5]
$CP_{3,1}$[34, 1,1]	$CP_{3,2}$[25,0,0]	$CP_{3,3}$[19,0,0]	$CP_{3,4}$[7,2,2]	$CP_{3,5}$[11,2,3]
$CP_{4,1}$[12, 0,0]	$CP_{4,2}$[8,2,2]	$CP_{4,3}$[12,3,2]	$CP_{4,4}$[11,3,2]	$CP_{4,5}$[38,1,2]
$CP_{5,1}$[35, 1,1]	$CP_{5,2}$[30,0,0]	$CP_{5,3}$[17,2,2]	$CP_{5,4}$[18,2,2]	$CP_{5,5}$[15,2,2]
$CP_{6,1}$[15, 2,2]	$CP_{6,2}$[18,0,0]	$CP_{6,3}$[14,3,1]	$CP_{6,4}$[45,1,2]	$CP_{6,5}$[48,1,3]
$CP_{7,1}$[36, 0,0]	$CP_{7,2}$[24,2,4]	$CP_{7,3}$[24,2,2]	$CP_{7,4}$[20,2,3]	$CP_{7,5}$[23,3,1]
$CP_{8,1}$[25, 3,1]	$CP_{8,2}$[20,3,5]	$CP_{8,3}$[53,1,2]	$CP_{8,4}$[57,1,2]	$CP_{8,5}$[49,0,1]
$CP_{9,1}$[30, 0,0]	$CP_{9,2}$[29,3,1]	$CP_{9,3}$[25,3,3]	$CP_{9,4}$[29,3,2]	$CP_{9,5}$[27,3,3]
$CP_{10,1}$[29,3,3]	$CP_{10,2}$[61,1,3]	$CP_{10,3}$[64,1,6]	$CP_{10,4}$[58,1,1]	$CP_{10,5}$[47,0,0]
$CP_{11,1}$[31,2,5]	$CP_{11,2}$[30,0,0]	$CP_{11,3}$[33,3,4]	$CP_{11,4}$[32,3,4]	$CP_{11,5}$[36,3,3]

ตารางที่ 4.6 เมทริกซ์ Component Plane ของ x_3

$CP_{1,1}$[1,1,4]	$CP_{1,2}$[2, 1,5]	$CP_{1,3}$[8, 1,1]	$CP_{1,4}$[23, 2,1]	$CP_{1,5}$[31, 2,5]
$CP_{2,1}$[38, 2,2]	$CP_{2,2}$[44,2,1]	$CP_{2,3}$[48,2,3]	$CP_{2,4}$[51,3,1]	$CP_{2,5}$[52,3,5]
$CP_{3,1}$[2, 1,1]	$CP_{3,2}$[7,0,0]	$CP_{3,3}$[23,0,0]	$CP_{3,4}$[33,2,2]	$CP_{3,5}$[37,2,3]
$CP_{4,1}$[45, 0,0]	$CP_{4,2}$[49,2,2]	$CP_{4,3}$[53,3,2]	$CP_{4,4}$[54,3,2]	$CP_{4,5}$[1,1,2]
$CP_{5,1}$[7, 1,1]	$CP_{5,2}$[22,0,0]	$CP_{5,3}$[35,2,2]	$CP_{5,4}$[40,2,2]	$CP_{5,5}$[42,2,2]
$CP_{6,1}$[50, 2,2]	$CP_{6,2}$[57,0,0]	$CP_{6,3}$[59,3,1]	$CP_{6,4}$[1,1,2]	$CP_{6,5}$[2,1,3]
$CP_{7,1}$[21, 0,0]	$CP_{7,2}$[36,2,4]	$CP_{7,3}$[43,2,2]	$CP_{7,4}$[45,2,3]	$CP_{7,5}$[50,3,1]
$CP_{8,1}$[60, 3,1]	$CP_{8,2}$[64,3,5]	$CP_{8,3}$[1,1,2]	$CP_{8,4}$[2,1,2]	$CP_{8,5}$[6,0,1]
$CP_{9,1}$[39, 0,0]	$CP_{9,2}$[45,3,1]	$CP_{9,3}$[48,3,3]	$CP_{9,4}$[53,3,2]	$CP_{9,5}$[56,3,3]
$CP_{10,1}$[64,3,3]	$CP_{10,2}$[2,1,3]	$CP_{10,3}$[2,1,6]	$CP_{10,4}$[6,1,1]	$CP_{10,5}$[25,0,0]
$CP_{11,1}$[46,2,5]	$CP_{11,2}$[50,0,0]	$CP_{11,3}$[54,3,4]	$CP_{11,4}$[57,3,4]	$CP_{11,5}$[61,3,3]

ตารางที่ 4.7 เมทริกซ์ Component Plane ของ x_4

$CP_{1,1}$[1,1,4]	$CP_{1,2}$[2, 1,5]	$CP_{1,3}$[8, 1,1]	$CP_{1,4}$[23, 2,1]	$CP_{1,5}$[29, 2,5]
$CP_{2,1}$[33, 2,2]	$CP_{2,2}$[38,2,1]	$CP_{2,3}$[44,2,3]	$CP_{2,4}$[51,3,1]	$CP_{2,5}$[56,3,5]
$CP_{3,1}$[2, 1,1]	$CP_{3,2}$[8,0,0]	$CP_{3,3}$[22,0,0]	$CP_{3,4}$[31,2,2]	$CP_{3,5}$[34,2,3]
$CP_{4,1}$[39, 0,0]	$CP_{4,2}$[44,2,2]	$CP_{4,3}$[50,3,2]	$CP_{4,4}$[55,3,2]	$CP_{4,5}$[2,1,2]
$CP_{5,1}$[8, 1,1]	$CP_{5,2}$[22,0,0]	$CP_{5,3}$[32,2,2]	$CP_{5,4}$[36,2,2]	$CP_{5,5}$[37,2,2]
$CP_{6,1}$[45, 2,2]	$CP_{6,2}$[52,0,0]	$CP_{6,3}$[54,3,1]	$CP_{6,4}$[3,1,2]	$CP_{6,5}$[3,1,3]
$CP_{7,1}$[21, 0,0]	$CP_{7,2}$[34,2,4]	$CP_{7,3}$[41,2,2]	$CP_{7,4}$[41,2,3]	$CP_{7,5}$[49,3,1]
$CP_{8,1}$[57, 3,1]	$CP_{8,2}$[56,3,5]	$CP_{8,3}$[3,1,2]	$CP_{8,4}$[3,1,2]	$CP_{8,5}$[7,0,1]
$CP_{9,1}$[37, 0,0]	$CP_{9,2}$[44,3,1]	$CP_{9,3}$[47,3,3]	$CP_{9,4}$[56,3,2]	$CP_{9,5}$[57,3,3]
$CP_{10,1}$[59,3,3]	$CP_{10,2}$[3,1,3]	$CP_{10,3}$[4,1,6]	$CP_{10,4}$[7,1,1]	$CP_{10,5}$[24,0,0]
$CP_{11,1}$[44,2,5]	$CP_{11,2}$[51,0,0]	$CP_{11,3}$[60,3,4]	$CP_{11,4}$[62,3,4]	$CP_{11,5}$[64,3,3]

หลังจากนั้นเปลี่ยน Color Bar เป็นเมทริกซ์ Color Bar ดังตารางที่ 4.8 ถึง 4.11 ตัวอย่างเช่น $CB_{1,2}$ [101,4.77,1] ถูกกำหนดด้วย Color Bar ของแถวที่ 1 คอลัมน์ที่ 2 มีค่าสี คือ 3 (สีเขียว) ซึ่งมีค่าข้อมูลเข้า 5.7 และอยู่ในกลุ่มที่เป็นไปได้ คือ กลุ่ม 1 สำหรับการหากลุ่มที่เป็นไปได้ของเมทริกซ์ Color Bar นั้นจะทำการพิจารณากลุ่มและค่าความถี่จากเมทริกซ์ Component Plane

ตารางที่ 4.8 เมทริกซ์ Color Bar ของ x_1

$CB_{1,1}$[101,4.67,1]	$CB_{1,2}$[101,4.77,1]	$CB_{1,3}$[101,4.86,1]	$CB_{1,4}$[101,5.18,1]	$CB_{1,5}$[102,5.31,2]
$CB_{2,1}$[102,5.68,2]	$CB_{2,2}$[103,6.01,2]	$CB_{2,3}$[103,6.11,2]	$CB_{2,4}$[103,6.14,2]	$CB_{2,5}$[103,5.98,2]
$CB_{3,1}$[101,4.87,1]	$CB_{3,2}$[101,4.95,1]	$CB_{3,3}$[102,5.30,2]	$CB_{3,4}$[102,5.44,2]	$CB_{3,5}$[102,5.68,2]
$CB_{4,1}$[103,6.16,2]	$CB_{4,2}$[103,6.23,2]	$CB_{4,3}$[105,6.42,2]	$CB_{4,4}$[104,6.25,2]	$CB_{4,5}$[101,4.86,1]
$CB_{5,1}$[101,5.08,1]	$CB_{5,2}$[102,5.40,2]	$CB_{5,3}$[102,5.60,2]	$CB_{5,4}$[102,5.84,2]	$CB_{5,5}$[103,5.84,2]
$CB_{6,1}$[105,6.43,2]	$CB_{6,2}$[106,6.80,3]	$CB_{6,3}$[106,6.79,3]	$CB_{6,4}$[101,4.99,1]	$CB_{6,5}$[101,5.13,1]
$CB_{7,1}$[102,5.51,2]	$CB_{7,2}$[102,5.71,2]	$CB_{7,3}$[103,5.97,2]	$CB_{7,4}$[103,6.11,2]	$CB_{7,5}$[105,3,2]
$CB_{8,1}$[106,6.60,3]	$CB_{8,2}$[106,6.98,3]	$CB_{8,3}$[101,7.19,1]	$CB_{8,4}$[102,5.13,2]	$CB_{8,5}$[102,5.28,2]
$CB_{9,1}$[103,5.94,2]	$CB_{9,2}$[103,6.23,2]	$CB_{9,3}$[103,6.22,2]	$CB_{9,4}$[105,6.45,2]	$CB_{9,5}$[106,6.67,3]
$CB_{10,1}$[106,7.26,3]	$CB_{10,2}$[102,5.27,2]	$CB_{10,3}$[102,5.40,2]	$CB_{10,4}$[102,5.49,2]	$CB_{10,5}$[103,5.97,2]
$CB_{11,1}$[105,6.47,5]	$CB_{11,2}$[105,6.39,2]	$CB_{11,3}$[105,6.50,2]	$CB_{11,4}$[106,6.67,3]	$CB_{11,5}$[106,6.90,3]

ตารางที่ 4.9 เมทริกซ์ Color Bar ของ x_2

$CB_{1,1}$[102,3.06,2]	$CB_{1,2}$[102,3.06,2]	$CB_{1,3}$[102,2.82,2]	$CB_{1,4}$[101,2.61,2]	$CB_{1,5}$[101,2.46,2]
$CB_{2,1}$[101,2.45,2]	$CB_{2,2}$[101,2.51,2]	$CB_{2,3}$[101,2.49,2]	$CB_{2,4}$[101,2.62,2]	$CB_{2,5}$[101,2.66,2]
$CB_{3,1}$[103,3.21,3]	$CB_{3,2}$[102,3.00,2]	$CB_{3,3}$[102,2.87,2]	$CB_{3,4}$[101,2.60,2]	$CB_{3,5}$[101,2.69,2]
$CB_{4,1}$[101,2.71,2]	$CB_{4,2}$[101,2.62,2]	$CB_{4,3}$[101,2.71,2]	$CB_{4,4}$[101,2.69,2]	$CB_{4,5}$[103,3.29,3]
$CB_{5,1}$[103,3.24,3]	$CB_{5,2}$[102,3.11,2]	$CB_{5,3}$[102,2.82,2]	$CB_{5,4}$[102,2.86,2]	$CB_{5,5}$[102,2.79,2]
$CB_{6,1}$[102,2.79,2]	$CB_{6,2}$[102,2.84,2]	$CB_{6,3}$[101,2.77,2]	$CB_{6,4}$[105,3.46,1]	$CB_{6,5}$[105,3.52,1]
$CB_{7,1}$[103,3.26,3]	$CB_{7,2}$[102,2.99,2]	$CB_{7,3}$[102,2.98,2]	$CB_{7,4}$[102,2.91,2]	$CB_{7,5}$[102,2.97,2]
$CB_{8,1}$[102,3.00,2]	$CB_{8,2}$[102 2.91,2]	$CB_{8,3}$[106,3.64,1]	$CB_{8,4}$[106,3.73,1]	$CB_{8,5}$[105,3.55,1]
$CB_{9,1}$[102,3.11,2]	$CB_{9,2}$[102,3.09,2]	$CB_{9,3}$[102,3.02,2]	$CB_{9,4}$[102,3.09,2]	$CB_{9,5}$[102,3.06,2]
$CB_{10,1}$[102,3.10,2]	$CB_{10,2}$[106,3.83,1]	$CB_{10,3}$[106,3.91,1]	$CB_{10,4}$[106,3.76,1]	$CB_{10,5}$[105,3.51,1]
$CB_{11,1}$[103,3.15,3]	$CB_{11,2}$[102,3.12,0]	$CB_{11,3}$[103,3.18,3]	$CB_{11,4}$[103,3.17,3]	$CB_{11,5}$[103,3.27,3]

ตารางที่ 4.10 เมทริกซ์ Color Bar ของ x_3

$CB_{1,1}$[101,1.41,1]	$CB_{1,2}$[101,1.49,1]	$CB_{1,3}$[101,1.94,1]	$CB_{1,4}$[102,.05,2]	$CB_{1,5}$[103,3.65,2]
$CB_{2,1}$[103,4.14,2]	$CB_{2,2}$[105,4.56,2]	$CB_{2,3}$[105,4.87,2]	$CB_{2,4}$[106,5.10,3]	$CB_{2,5}$[106,5.13,3]
$CB_{3,1}$[101,1.51,1]	$CB_{3,2}$[101,1.90,1]	$CB_{3,3}$[102,3.02,2]	$CB_{3,4}$[103,3.76,2]	$CB_{3,5}$[103,4.10,2]
$CB_{4,1}$[105,4.65,2]	$CB_{4,2}$[105,4.94,2]	$CB_{4,3}$[106,5.27,3]	$CB_{4,4}$[106,5.31,3]	$CB_{4,5}$[101,1.45,1]
$CB_{5,1}$[101,1.86,1]	$CB_{5,2}$[102,2.95,2]	$CB_{5,3}$[103,3.90,2]	$CB_{5,4}$[104,4.30,2]	$CB_{5,5}$[105,4.42,2]
$CB_{6,1}$[106,5.04,3]	$CB_{6,2}$[106,5.56,3]	$CB_{6,3}$[106,5.70,3]	$CB_{6,4}$[101,1.45,1]	$CB_{6,5}$[101,1.52,1]
$CB_{7,1}$[102,2.92,2]	$CB_{7,2}$[103,4.02,2]	$CB_{7,3}$[105,4.50,2]	$CB_{7,4}$[105,4.64,2]	$CB_{7,5}$[106,5.04,3]
$CB_{8,1}$[106,5.77,3]	$CB_{8,2}$[106,6.03,3]	$CB_{8,3}$[101,1.48,1]	$CB_{8,4}$[101,1.52,1]	$CB_{8,5}$[101,1.80,1]
$CB_{9,1}$[103,4.19,2]	$CB_{9,2}$[105,4.67,2]	$CB_{9,3}$[105,4.87,2]	$CB_{9,4}$[106,5.26,3]	$CB_{9,5}$[106,5.47,3]
$CB_{10,1}$[106,6.07,3]	$CB_{10,2}$[101,1.51,1]	$CB_{10,3}$[101,1.52,1]	$CB_{10,4}$[101,1.81,1]	$CB_{10,5}$[102,3.19,2]
$CB_{11,1}$[105,4.71,2]	$CB_{11,2}$[106,4.99,3]	$CB_{11,3}$[106,5.33,3]	$CB_{11,4}$[106,5.55,3]	$CB_{11,5}$[106,5.78,3]

ตารางที่ 4.11 เมทริกซ์ Color Bar ของ x_4

$CB_{1,1}$[101,0.18,1]	$CB_{1,2}$[101,0.23,1]	$CB_{1,3}$[101,0.44,1]	$CB_{1,4}$[102,0.90,2]	$CB_{1,5}$[102,1.11,2]
$CB_{2,1}$[103,1.25,2]	$CB_{2,2}$[103,1.40,2]	$CB_{2,3}$[105,1.60,2]	$CB_{2,4}$[106,1.81,3]	$CB_{2,5}$[106,1.96,3]
$CB_{3,1}$[101,0.24,1]	$CB_{3,2}$[101,0.42,1]	$CB_{3,3}$[102,0.89,2]	$CB_{3,4}$[103,1.16,2]	$CB_{3,5}$[103,1.25,2]
$CB_{4,1}$[103,1.44,2]	$CB_{4,2}$[105,1.60,2]	$CB_{4,3}$[106,1.79,3]	$CB_{4,4}$[106,1.94,3]	$CB_{4,5}$[101,0.22,1]
$CB_{5,1}$[101,0.42,1]	$CB_{5,2}$[102,0.86,2]	$CB_{5,3}$[103,1.21,2]	$CB_{5,4}$[103,1.35,2]	$CB_{5,5}$[103,1.36,2]
$CB_{6,1}$[105,1.64,2]	$CB_{6,2}$[106,1.84,3]	$CB_{6,3}$[106,1.91,3]	$CB_{6,4}$[101,0.25,1]	$CB_{6,5}$[101,0.28,1]
$CB_{7,1}$[102,0.85,2]	$CB_{7,2}$[103,1.27,4]	$CB_{7,3}$[104,1.48,2]	$CB_{7,4}$[104,1.50,2]	$CB_{7,5}$[105,1.75,2]
$CB_{8,1}$[106,2.00,3]	$CB_{8,2}$[106,1.96,3]	$CB_{8,3}$[101,0.27,1]	$CB_{8,4}$[101,0.28,1]	$CB_{8,5}$[101,0.40,1]
$CB_{9,1}$[103,1.37,2]	$CB_{9,2}$[105,1.59,2]	$CB_{9,3}$[105,1.69,2]	$CB_{9,4}$[106,1.98,3]	$CB_{9,5}$[106,2.02,3]
$CB_{10,1}$[106,2.08,3]	$CB_{10,2}$[101,0.28,1]	$CB_{10,3}$[101,0.29,1]	$CB_{10,4}$[101,0.40,1]	$CB_{10,5}$[102,0.96,2]
$CB_{11,1}$[105,1.60,2]	$CB_{11,2}$[106,1.81,3]	$CB_{11,3}$[106,2.09,3]	$CB_{11,4}$[106,2.18,3]	$CB_{11,5}$[106,2.25,3]

ขั้นตอนหลักที่ A3 การสกัดคุณลักษณะที่สำคัญ (Feature Extraction Process) เป็นหาค่าความถูกต้องของแต่ละตัวแปรเข้า แสดงได้ดังตารางที่ 4.12 หลังจากนั้นหาค่าความถูกต้องที่ยอมรับได้ ดังสมการ (3.1) ซึ่งได้ค่าความถูกต้องที่ยอมรับได้ (ϕ) คือ 89.7 (94.7-5.0) ถ้าค่าความถูกต้องจากแต่ละตัวแปรเข้ามากกว่าหรือเท่ากับ ϕ แล้ว ให้ยอมรับค่าความถูกต้องเหล่านั้น และเก็บตัวแปรเหล่านั้นไว้ใน accepted accuracy list ซึ่งคือ x_3 และ x_4

ตารางที่ 4.12 ค่าความถูกต้องของแต่ละตัวแปรเข้า

ตัวแปร	ค่าความถูกต้อง	Accepted Accuracy List
x_1	64.0	No
x_2	22.0	No
x_3	94.7 (maximum accuracy)	Yes
x_4	94.7 (maximum accuracy)	Yes

ขั้นตอนหลักที่ A4 การสร้างกฎความรู้ (Generated Knowledge Based) เป็นการสร้างกฎที่เป็นไปได้ ตามสมการ (3.2) ซึ่งได้กฎทั้งหมด 6 กฎ และมีค่าความถูกต้อง 94.7% ดังภาพประกอบ 4.5

กฎที่ได้จากการสกัดความรู้ด้วย KESOM

R1 : If petal length <= 1.9 then setosa

R2 : If petal length <= 4.9 then versicolor

R3 : If petal length >= 5.0 then viginica

R4 : If petal width <= 0.4 then setosa

R5 : If petal width <= 1.7 then versicolor

R6 : If petal width >= 1.8 then viginica

ค่าความถูกต้อง 94.7%

ภาพประกอบ 4.5 กฎและค่าความถูกต้องจากฐานข้อมูลไอริส

4.1.2 แบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (KESOM_MEP)

สำหรับการสกัดความรู้จากฐานข้อมูลดอกไม้ไอริสที่อยู่ในรูปของกฎภาษาธรรมชาตินั้น มีการทำงาน 5 ขั้นตอนหลัก ตามรายละเอียดดังนี้

ขั้นตอนหลักที่ B1 กระบวนการแผนที่การจัดกลุ่มเอง (Self-Organizing Map Process) ทำการแบ่งกลุ่มข้อมูลด้วย Self-Organizing Map โดยใช้ Matlab SOM Toolbox เป็นเครื่องมือ จะได้ Component Plane, Color Bar และ Label ซึ่งเป็นเครื่องมือสำหรับวิเคราะห์ผลลัพธ์ที่ได้จากการจัดกลุ่ม ดังภาพประกอบ 4.1 ถึง ภาพประกอบ 4.4

ขั้นตอนหลักที่ B2 หลักเอ็นโทรพีค่าต่ำสุด (Minimization Entropy Principle Algorithm) เป็นขั้นตอนการหา potential threshold point หรือจุดแบ่งข้อมูลของแต่ละตัวแปรเข้า ซึ่งมีรายละเอียดขั้นตอนดังนี้

หาค่า $R_{\min}$ คือ ค่าน้อยที่สุด, หาค่า $R_{\max}$ คือ ค่ามากที่สุด, หาค่า σ คือ ค่าเบี่ยงเบนมาตรฐาน และคำนวณช่วงค่าที่ใช้ในการแทนตัวแปร จากสมการ (3.3) ซึ่ง x_1อยู่ในช่วงค่า [4.0,8.0] x_2 อยู่ในช่วงค่า [2.1,4.3] x_3 อยู่ในช่วงค่า [-0.2,7.7] และ x_4 อยู่ในช่วงค่า [-0.5,3.0] จะเห็นได้ว่าทั้ง x_3 และ x_4 นั้นเป็นไปไม่ได้ที่มีช่วงค่าเป็นลบ ดังนั้นช่วงค่าที่ถูกต้อง คือ [0,7.7], และ [0,3.0] ตามลำดับ

ตารางที่ 4.13 ค่าน้อยที่สุดและค่ามากที่สุดของแต่ละกลุ่ม

ตัวแปร	Setosa		Versicolor		Virginica	
	Min	Max	Min	Max	Min	Max
x_1	4.7	5.5	5.2	6.6	6	7.3
x_2	2.8	3.9	2.5	3.2	2.6	3.3
x_3	1.4	1.9	3.0	5.3	4.9	6.1
x_4	0.2	0.4	0.9	1.8	1.7	2.3

จากตารางที่ 4.13 แสดงค่ามากที่สุดและค่าน้อยที่สุด และช่วงค่าของแต่ละกลุ่มทั้ง 4 ตัวแปรเข้า จะเห็นได้ว่า ทั้ง x_1 และ x_2 มีช่วงค่าที่คาบเกี่ยวกันทุกกลุ่ม ในขณะที่ x_3 และ x_4 ช่วงค่าของกลุ่ม Setosa ไม่คาบเกี่ยวกับกลุ่มอื่น ๆ แต่ช่วงค่าของกลุ่ม Vesicolor และ Virginica มีช่วงค่าที่คาบเกี่ยวกัน เมื่อทราบว่าช่วงค่าใดที่คาบเกี่ยวกันหรือไม่คาบเกี่ยวกันแล้ว เรียงลำดับจากค่าน้อยไปมากในแต่ละตัวแปรเข้า แล้วหาค่า potential threshold point หรือจุดแบ่งข้อมูล ดังตารางที่ 4.14

ตารางที่ 4.14 potential threshold point ของแต่ละตัวแปรเข้า

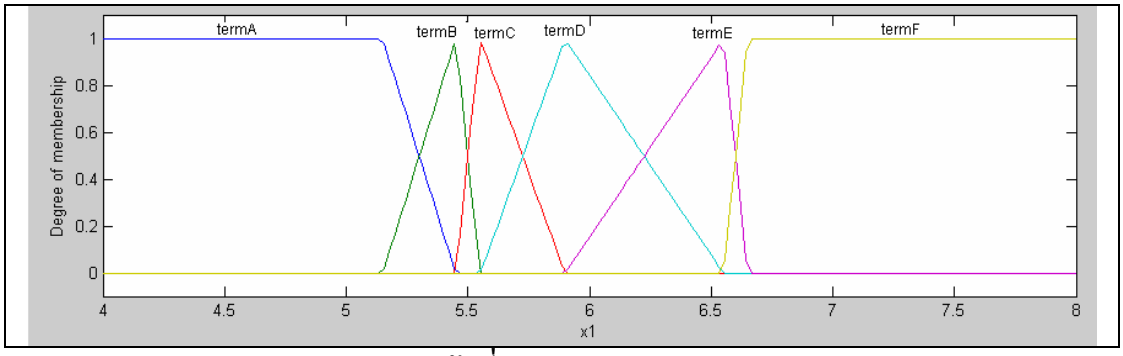
ตัวแปร	Potential threshold point	กลุ่ม
x_1	4.75, 4.85, 4.95, 5.05,5.15, 5.25, 5.35, 5.45, 5.55, 5.65, 5.75, 5.9, 6.05, 6.15, 6.25, 6.35, 6.45, 6.55	Setosa, Vesicolor, Virginica
x_2	2.55, 2.65, 2.75, 2.85, 2.95, 3.05, 3.15, 3.25, 3.35, 3.45, 3.55, 3.65, 3.75, 3.85	Setosa, Vesicolor, Virginica
x_3	1.45, 1.65, 1.85	Setosa
	3.35, 3.75, 3.85, 3.95, 4.05, 4.15, 4.25, 4.35, 4.45, 4.55, 4.65, 4.75, 4.85, 4.95, 5.05, 5.2, 5.4, 5.6, 5.75, 5.9, 6.05	Vesicolor, Virginica
x_4	0.25, 0.35	Setosa
	1.0, 1.25, 1.35, 1.45, 1.55, 1.65, 1.75, 1.85, 1.95, 2.05, 2.15, 2.25	Vesicolor, Virginica

หลังจากนั้นหาค่า entropy ของแต่ละ potential threshold point จากสมการ (3.4) ถึง (3.12) โดยใช้ขั้นตอนวิธีหลักเอ็นโทรพีค่าต่ำสุด เลือกค่า potential threshold point ที่มีค่า entropy น้อยที่สุดเก็บไว้ใน member_list ดังตารางที่ 4.15

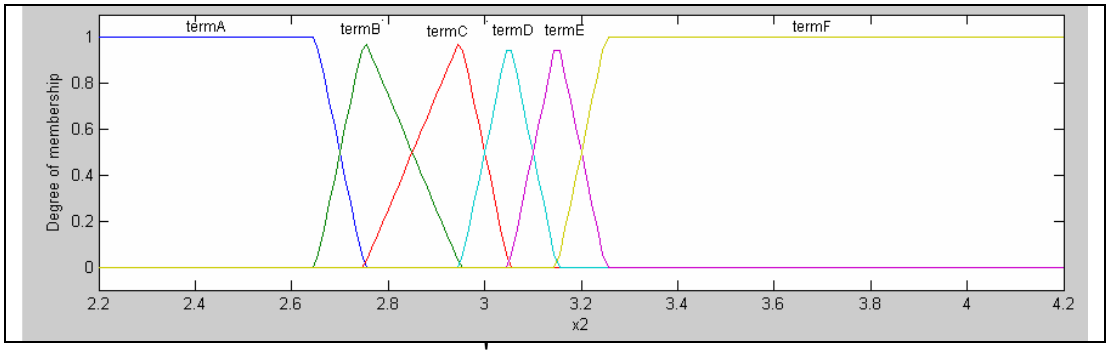
ตารางที่ 4.15 potential threshold point ที่เก็บใน member_list

ตัวแปร	Member_list	กลุ่ม
x_1	5.15, 5.45, 5.55, 5.9, 6.55, 6.65	Setosa, Vesicolor, Virginica
x_2	2.65, 2.75, 2.95, 3.05, 3.15, 3.25	Setosa, Vesicolor, Virginica
x_3	1.9, 3.0	Setosa
	4.85, 5.05	Vesicolor, Virginica
x_4	0.4, 0.9	Setosa
	1.65, 1.85, 2.25	Vesicolor, Virginica

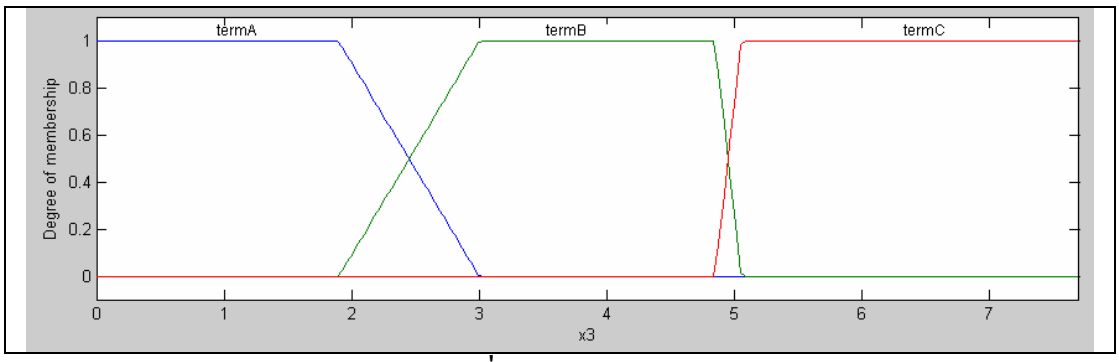
ขั้นตอนหลักที่ B3 การสร้างฟัซซี่เซต (Creating Fuzzy Set) นำค่า potential threshold point จาก member_list ไปสร้างฟัซซี่เซตและเทอมเซต ตามสมการ (3.15) ถึง (3.15) ดังภาพประกอบ 4.6 ถึงภาพประกอบ 4.9



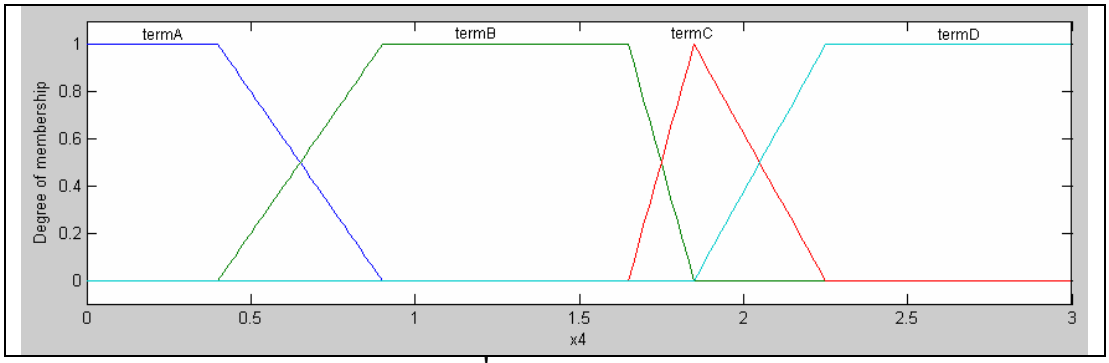
ภาพประกอบ 4.6 ฟัซซี่เซต และเทอมเซต 6 เทอม ของ x_1



ภาพประกอบ 4.7 ฟัซซี่เซต และเทอมเซต 6 เทอม ของ x_2



ภาพประกอบ 4.8 ฟัซซี่เซต และเทอมเซต 3 เทอม ของ x_3



ภาพประกอบ 4.9 ฟัซซี่เซต และเทอมเซต 4 เทอม ของ x_4

ขั้นตอนหลักที่ B4 การสร้างรูปแบบภาษาธรรมชาติ (Creating Linguistic Term Using Membership Function) แปลงค่าสี จาก Color Bar ของข้อมูลทุกตัวแปรเพื่อสร้างเป็นช่วงค่าภาษาธรรมชาติโดยดูค่าความเป็นสมาชิกที่เป็นไปได้ว่าอยู่ในช่วงค่าใด ดังตารางที่ 4.16 ถึง 4.19 โดยสมการที่ใช้ในการประมาณค่าความเป็นสมาชิกของแต่ละช่วงค่าภาษาธรรมชาติ คือ ในกรณีฟังก์ชันความเป็นสมาชิกเป็นรูปสี่เหลี่ยมคางหมู พิจารณาดังสมการ (3.16) และในกรณีฟังก์ชันความเป็นสมาชิกเป็นรูปสามเหลี่ยม พิจารณาดังสมการ (3.17)

ตารางที่ 4.16 ช่วงค่าที่เป็นไปได้ของ x_1

Interval	Linguistic Term
4.0–5.3	termA
5.4–5.5	termB
5.6–5.7	termC
5.8–6.2	termD
6.3–6.6	termE
>6.6	termF

ตารางที่ 4.17 ช่วงค่าที่เป็นไปได้ของ x_2

Interval	Linguistic Term
2.2–2.6	termA
2.7–2.8	termB
2.9	termC
3.0	termD
3.1	termE
>3.1	termF

ตารางที่ 4.18 ช่วงค่าที่เป็นไปได้ของ x_3

Interval	Linguistic Term
0.0–2.4	termA
2.5–4.9	termB
>4.9	termC

ตารางที่ 4.19 ช่วงค่าที่เป็นไปได้ของ x_4

Interval	Linguistic Term
0.0-0.6	termA
0.7-1.7	termB
1.8-2.0	termC
>2.0	termD

ขั้นตอนหลักที่ B5 การสร้างกฎความรู้ (Knowledge Extraction Using Rough Set) เป็นการสร้างกฎโดยใช้ราฟเซต สำหรับการสร้างกฎถ้า-แล้ว จากหลักการประมาณค่าขอบเขตล่างดังสมการ (3.18) และการประมาณค่าขอบเขตบนดังสมการ (3.19) กฎที่ยอมรับได้ต้องมีค่าค่าความมั่นใจมากกว่า 50% ดังแสดงได้ตามภาพประกอบ 4.10 ถึงภาพประกอบ 4.13

ตัวแปรเข้า x_1

Lower Approximation with Certainty Value

R1 : If x1 = termC then class 2.
Certainty Value = 100.0

R2 : If x1 = termF then class 3.
Certainty Value = 100.0

Upper Approximation with Certainty Value

R3 : If x1 = termA then class 1.
Certainty Value = 88.2

R4 : If x1 = termD then class 2.
Certainty Value = 75.0

ค่าความถูกต้องทั้งหมด 58.0

ภาพประกอบ 4.10 กฎและค่าความถูกต้องที่ได้จาก x_1

ตัวแปรเข้า x_2

Lower Approximation with Certainty Value

ไม่มี

Upper Approximation with Certainty Value

R1 : If x2 = termA then class 2.

Certainty Value = 71.4

R2 : If x2 = termB then class 2.

Certainty Value = 60.0

R3 : If x2 = termC then class 2.

Certainty Value = 60.0

R4 : If x2 = termD then class 1.

Certainty Value = 84.6

ค่าความถูกต้องทั้งหมด 49.3

ภาพประกอบ 4.11 กฎและค่าความถูกต้องที่ได้จาก x_2

ตัวแปรเข้า x_3

Lower Approximation with Certainty Value

R1 : If x3 = termA then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R2 : If x3 = termB then class 2.

Certainty Value = 95.2

R3 : If x3 = termC then class 3.

Certainty Value = 88.9

ค่าความถูกต้องทั้งหมด 94.7

ภาพประกอบ 4.12 กฎและค่าความถูกต้องที่ได้จาก x_3

ตัวแปรเข้า x_4

Lower Approximation with Certainty Value

R1 : If x4 = termA then class 1.

Certainty Value = 100.0

R2 : If x4 = termD then class 3.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R3 : If x4 = termB then class 2.

Certainty Value = 91.3

R4 : If x4 = termC then class 3.

Certainty Value = 90.0

ค่าความถูกต้องทั้งหมด 96.0

ภาพประกอบ 4.13 กฎและค่าความถูกต้องที่ได้จาก x_4

จากนั้นเลือกเฉพาะตัวแปรเข้าที่มีค่าความถูกต้องสูงที่สุด ซึ่งกฎที่ได้จาก "x_4" เป็นกฎที่ยอมรับได้ ดังภาพประกอบ 4.14 โดยที่ R1 และ R2 คือกฎที่อยู่ใน lower approximation R3 และ R4 คือกฎที่อยู่ใน upper approximation มีค่าความถูกต้องทั้งหมด 96.0

กฎที่ได้จากการสกัดความรู้ด้วย KESOM_MEP

Lower Approximation with Certainty Value

R1 : If x4 = termA then class 1.

Certainty Value = 100.0

R2 : If x4 = termD then class 3.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R3 : If x4 = termB then class 2.

Certainty Value = 91.3

R4 : If x4 = termC then class 3.

Certainty Value = 90.0

ค่าความถูกต้องทั้งหมด 96.0

ภาพประกอบ 4.14 กฎและค่าความถูกต้องที่ได้จากสกัดความรู้

4.1.3 เปรียบเทียบผลการทดลองและวิจารณ์ผลการทดลอง

แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง (KESOM)

ผลการทดลองจากฐานข้อมูลไอริส แบบจำลองนี้สามารถสกัดความรู้ที่อยู่ในรูปของกฎทั่วไปได้ 6 กฎ ซึ่งกฎนี้เลือกมาจากตัวแปรเข้าที่มีค่าความถูกต้องของกฎมากที่สุด โดยกฎที่ได้นั้นไม่มีตัวเชื่อมเงื่อนไข

แบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (KESOM) มีผลการทดลองดังนี้

1. ผลการทดลองจากฐานข้อมูลไอริส แบบจำลองนี้สามารถสกัดความรู้ที่อยู่ในรูปของกฎภาษาธรรมชาติ โดยที่จำนวนของกฎขึ้นอยู่กับระดับค่าความมั่นใจ (Certainty Value) ของกฎที่ได้จากราฟเซต ซึ่งโดยทั่วไปแล้วกฎที่ได้จากการประมาณค่าขอบเขตล่างจะเป็นกฎที่ยอมรับได้ ที่มีระดับค่าความมั่นใจ 100% แต่กฎที่ได้จากการประมาณค่าขอบเขตบนต้องมีระดับค่าความมั่นใจมากกว่า 50% จึงจะเป็นกฎที่ยอมรับได้ จะเห็นได้ว่าระดับค่าความมั่นใจนั้นเป็นตัวกรองเพื่อให้ได้กฎที่น้อยและมีความสำคัญเท่านั้น ดังภาพประกอบ 4.9 ถึงภาพประกอบ 4.12

2. กฎที่ได้จากแบบจำลองนี้มาจากตัวแปรเข้าที่มีค่าความถูกต้องมากที่สุดโดยกฎจะไม่มีตัวเชื่อมเงื่อนไข

3. รูปแบบภาษาธรรมชาติหรือเทอมเซตได้มาจากขั้นตอนหลักเอ็นโทรพีค่าต่ำสุด ซึ่งเป็นวิธีการสร้างเทอมเซตจากข้อมูลของตัวแปรเข้าต่าง ๆ อย่างอัตโนมัติ ดังภาพประกอบ 4.5 ถึงภาพประกอบ 4.8

ตารางที่ 4.20 แสดงผลการเปรียบเทียบการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง KESOM และ KESOM_MEP กับวิธีการอื่น ๆ พบว่าวิธีการสกัดความรู้ KESOM ได้จำนวนกฎในรูปกฎทั่วไป 6 กฎ โดยไม่มีตัวเชื่อมเงื่อนไข ซึ่งมีค่าความถูกต้อง 94.7% จะเห็นได้ว่า KESOM มีจำนวนกฎและจำนวนตัวเชื่อมเงื่อนไขที่น้อยกว่า ReFuNN [7] วิธีการสกัดความรู้ KESOM_MEP ได้จำนวนกฎที่เป็นแบบกฎภาษาธรรมชาติ 4 กฎ โดยไม่มีตัวเชื่อมเงื่อนไข ซึ่งมีค่าความถูกต้อง 96.0% จะเห็นได้ว่า KESOM_MEP ได้จำนวนกฎน้อยกว่า จำนวนตัวเชื่อมเงื่อนไขน้อยกว่า และมีค่าความถูกต้องมากกกว่า ReFuNN [7] ที่เป็นกฎภาษาธรรมชาติ และได้ผลการทดลองที่ดีกว่า C-MLP2LN [7] ที่เป็นกฎทั่วไป

ตารางที่ 4.20 ผลการทดลองเมื่อเปรียบเทียบการสกัดความรู้ด้วย KESOM และ KESOM_MEP กับขั้นตอนวิธีอื่น ๆ ของฐานข้อมูลดอกไม้ไอริส

Method	#Rules	#Condition	Fuzzy Rule	Accuracy (%)
KESOM	**6**	**0**	**No**	**94.7%**
KESOM_MEP	**4**	**0**	**Yes**	**96.0%**
ReFuNN [7]	9	26	Yes	95.7%
C-MLP2LN [7]	2	2	No	95.7%

4.2 ฐานข้อมูลโรคมะเร็งเต้านม

ฐานข้อมูลโรคมะเร็งเต้านม (Wisconsin breast cancer database) ประกอบด้วยข้อมูล 2 กลุ่มคือ กลุ่มผู้ที่ไม่เป็นโรคมะเร็งเต้านม (Benign) มี 458 คน และกลุ่มผู้ที่เป็นโรคมะเร็งเต้านม (Malignant) มี 241 คน รวมทั้งหมด 699 คน ตัวแปรข้อมูลเข้ามีทั้งหมด 9 ตัวแปร x_1 คือ Clump thickness x_2 คือ Uniformity of cell size x_3 คือ Uniformity of cell shape x_4 คือ Marginal adhesion x_5 คือ Single epithelial cell size x_6 คือ Bare nuclei x_7 คือ Bland chromatin x_8 คือ Normal nucleoli และ x_9 คือ Mitoses ดังตารางที่ 4.21

ตารางที่ 4.21 รายละเอียดของฐานข้อมูลโรคมะเร็งเต้านม

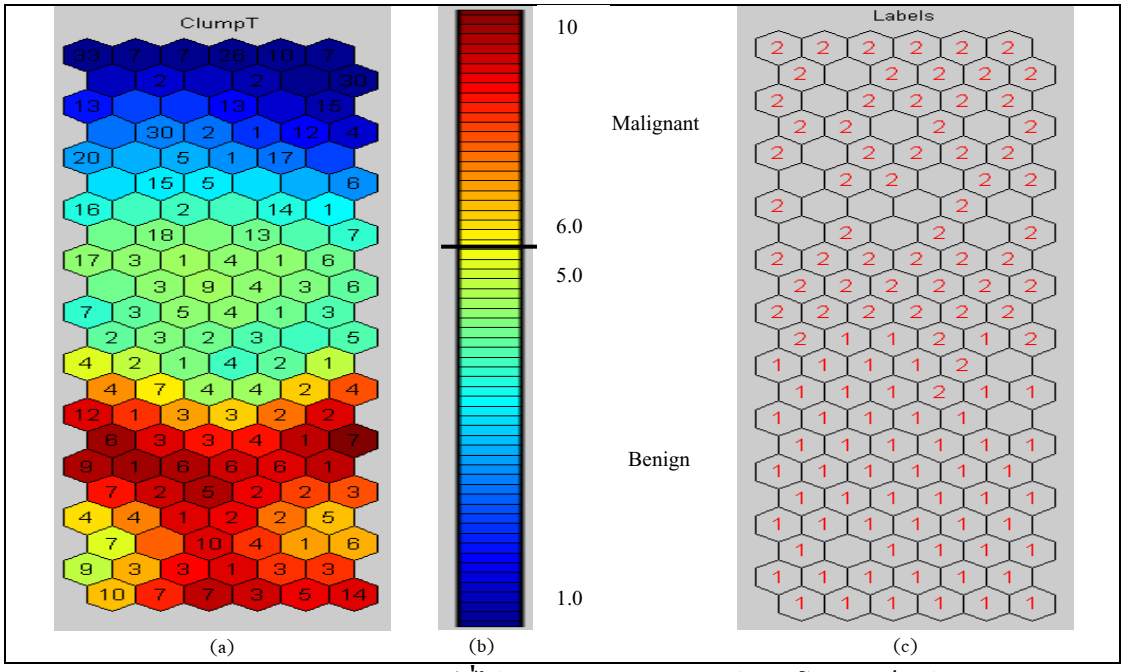
ตัวแปร	ชื่อตัวแปร	ชนิด	ค่าที่เป็นไปได้
x_1	clump thickness	Nominal Scale	1-10
x_2	uniformity of cell size	Nominal Scale	1-10
x_3	uniformity of cell shape	Nominal Scale	1-10
x_4	marginal adhesion	Nominal Scale	1-10
x_5	single epithelial cell size	Nominal Scale	1-10
x_6	bare nuclei	Nominal Scale	1-10
x_7	bland chromatin	Nominal Scale	1-10
x_8	normal nucleoli	Nominal Scale	1-10
x_9	mitoses	Nominal Scale	1-10
c	class	Nominal Scale	1 = Malignant 2 = Benign

4.2.1 แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง (KESOM)

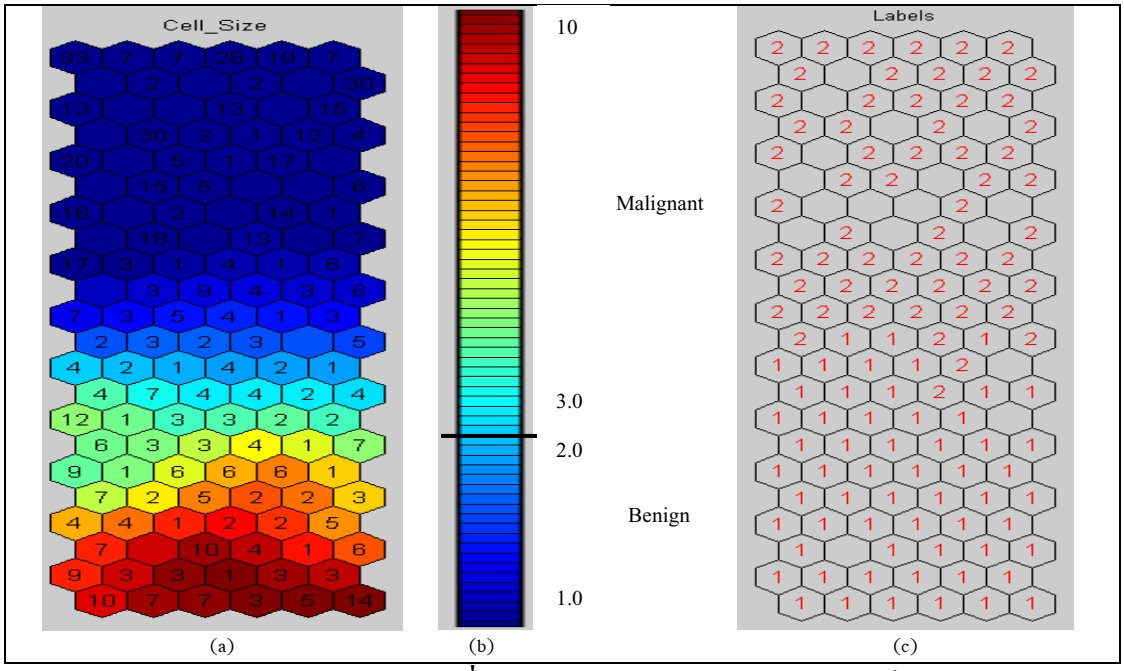
สำหรับการสกัดความรู้จากฐานข้อมูลโรคมะเร็งเต้านมที่อยู่ในรูปกฎทั่วไปนั้น มีการทำงาน 4 ขั้นตอนหลัก ตามรายละเอียดดังนี้

ขั้นตอนหลักที่ A1 กระบวนการแผนที่การจัดกลุ่มเอง (Self-Organizing Map Process) ทำการแบ่งกลุ่มข้อมูลด้วย Self-Organizing Map โดยใช้ Matlab SOM Toolbox เป็นเครื่องมือ จะได้ Component Plane, Color Bar และ Label ซึ่งเป็นเครื่องมือสำหรับวิเคราะห์ผลลัพธ์ที่ได้จากการจัดกลุ่ม ดังภาพประกอบ 4.15 ถึง ภาพประกอบ 4.23 สำหรับ Component Plane เป็นผลลัพธ์ที่ได้จากการจัดกลุ่มข้อมูลในแต่ละตัวแปรเข้า ของข้อมูล และมี Color Bar ที่

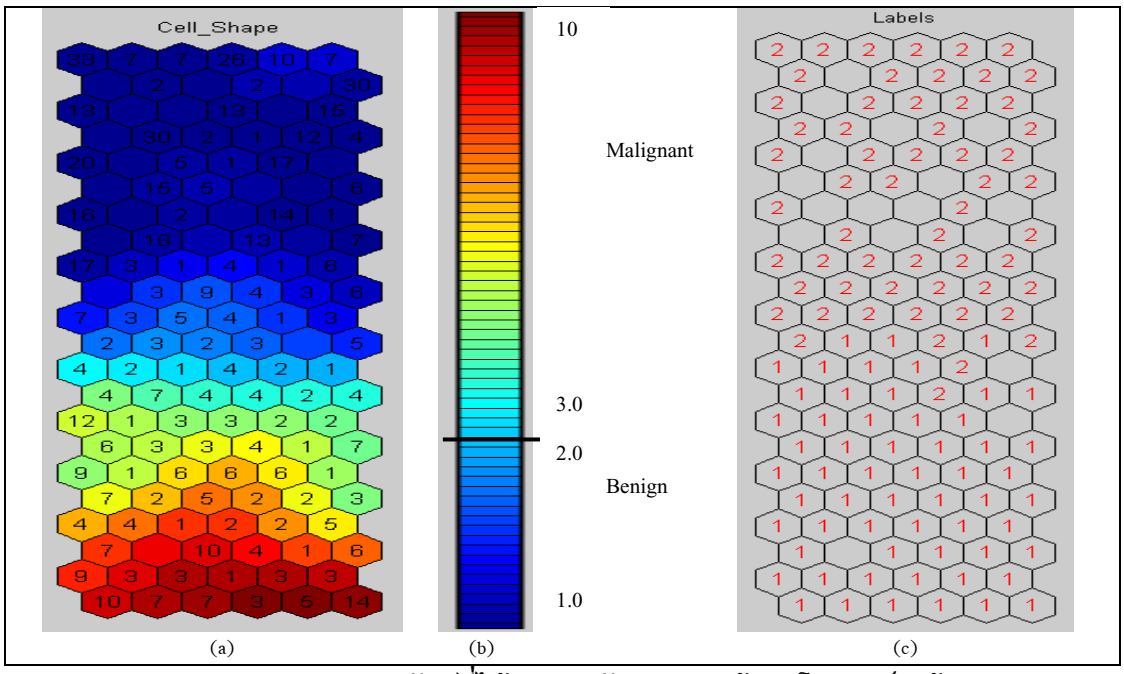
แสดงค่าข้อมูลที่ได้จากการจัดกลุ่มข้อมูล สำหรับ Label นั้นเป็นส่วนที่แสดงกลุ่มที่ได้จากการจัดกลุ่มข้อมูล



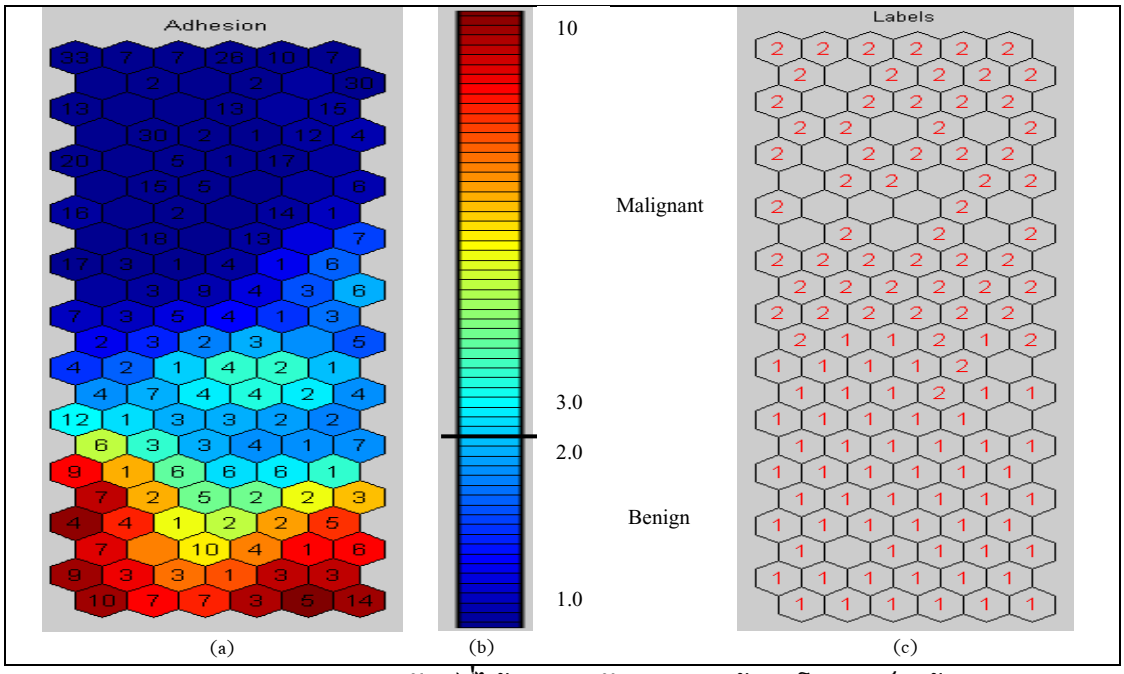
ภาพประกอบ 4.15 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_1 (b) ช่วงค่า Color Bar ของ x_1 (c) Label ของกลุ่มทั้ง 2 กลุ่ม



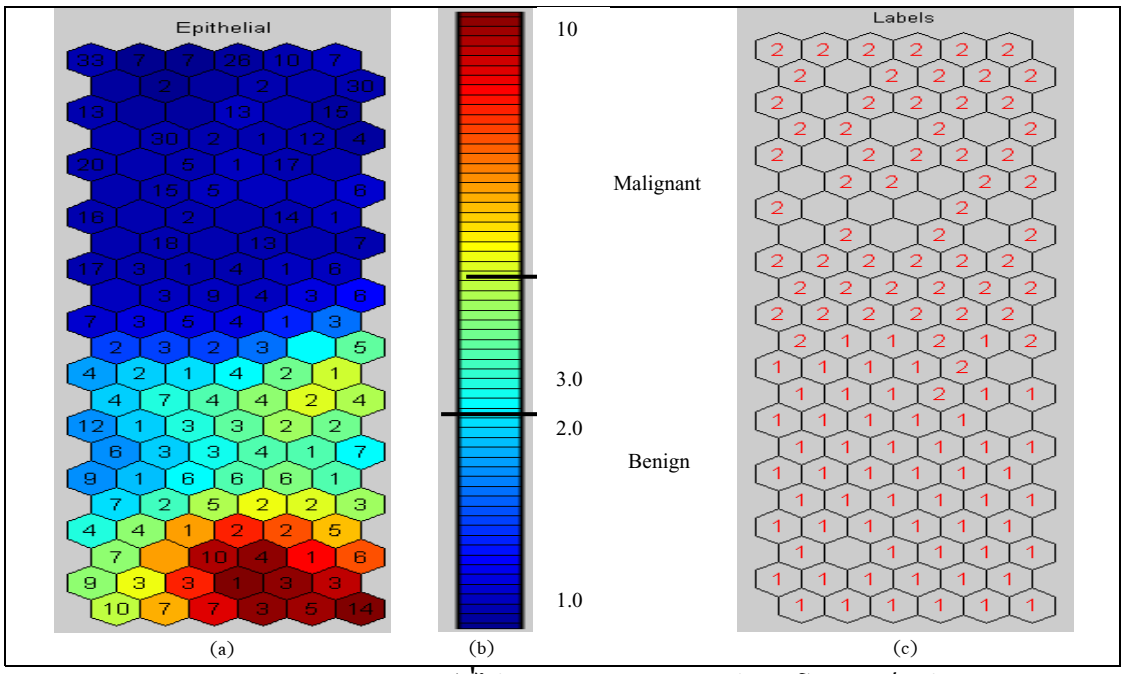
ภาพประกอบ 4.16 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_2 (b) ช่วงค่า Color Bar ของ x_2 (c) Label ของกลุ่มทั้ง 2 กลุ่ม



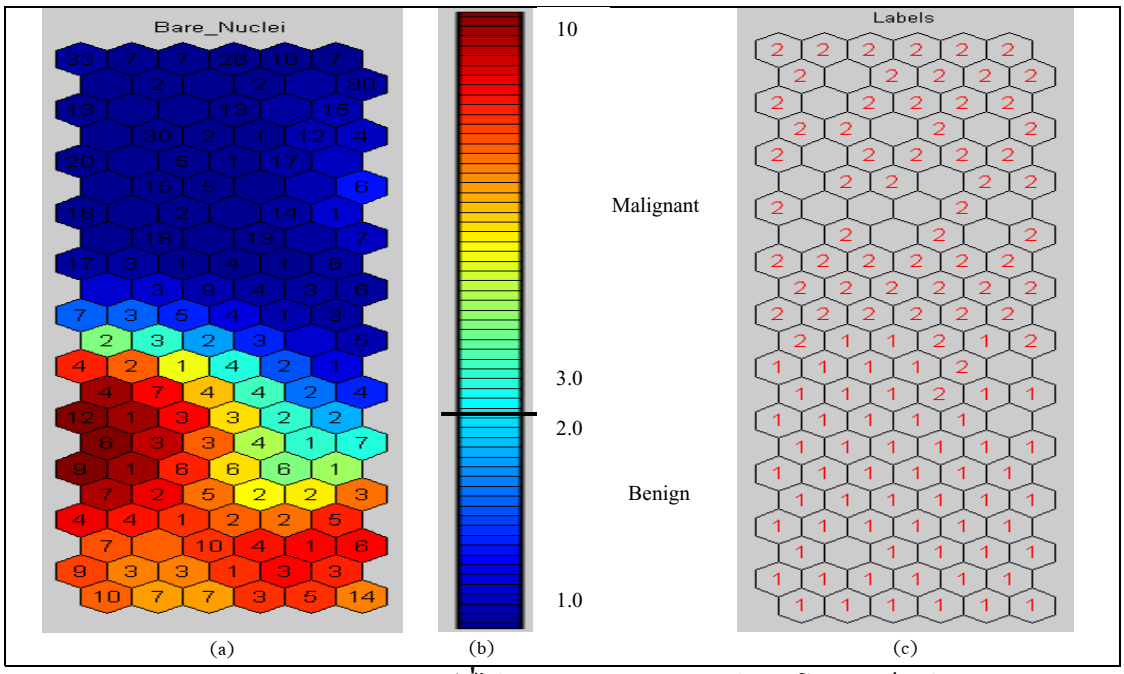
ภาพประกอบ 4.17 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_3 (b) ช่วงค่า Color Bar ของ x_3 (c) Label ของกลุ่มทั้ง 2 กลุ่ม



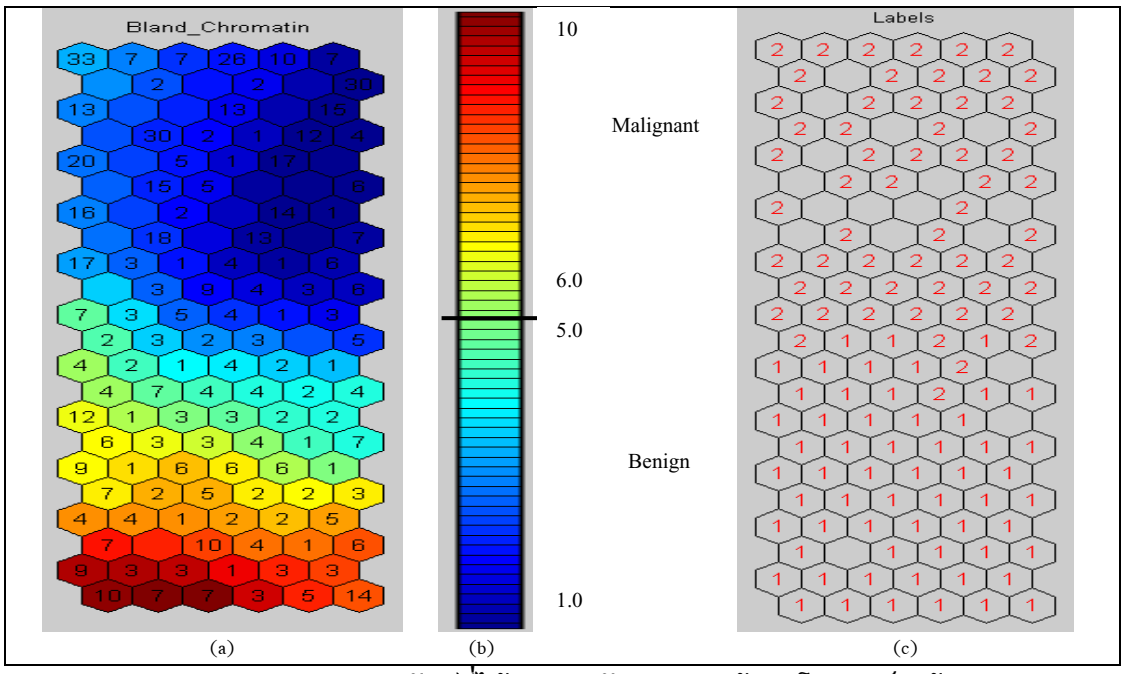
ภาพประกอบ 4.18 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_4 (b) ช่วงค่า Color Bar ของ x_4 (c) Label ของกลุ่มทั้ง 2 กลุ่ม



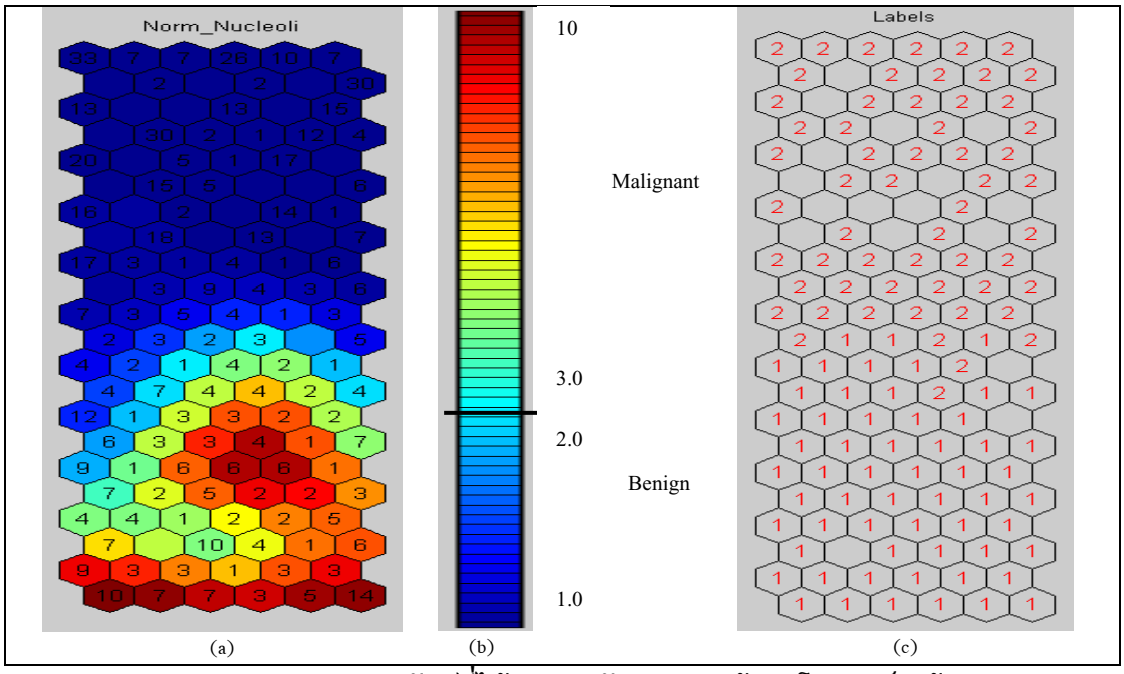
ภาพประกอบ 4.19 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_5 (b) ช่วงค่า Color Bar ของ x_5 (c) Label ของกลุ่มทั้ง 2 กลุ่ม



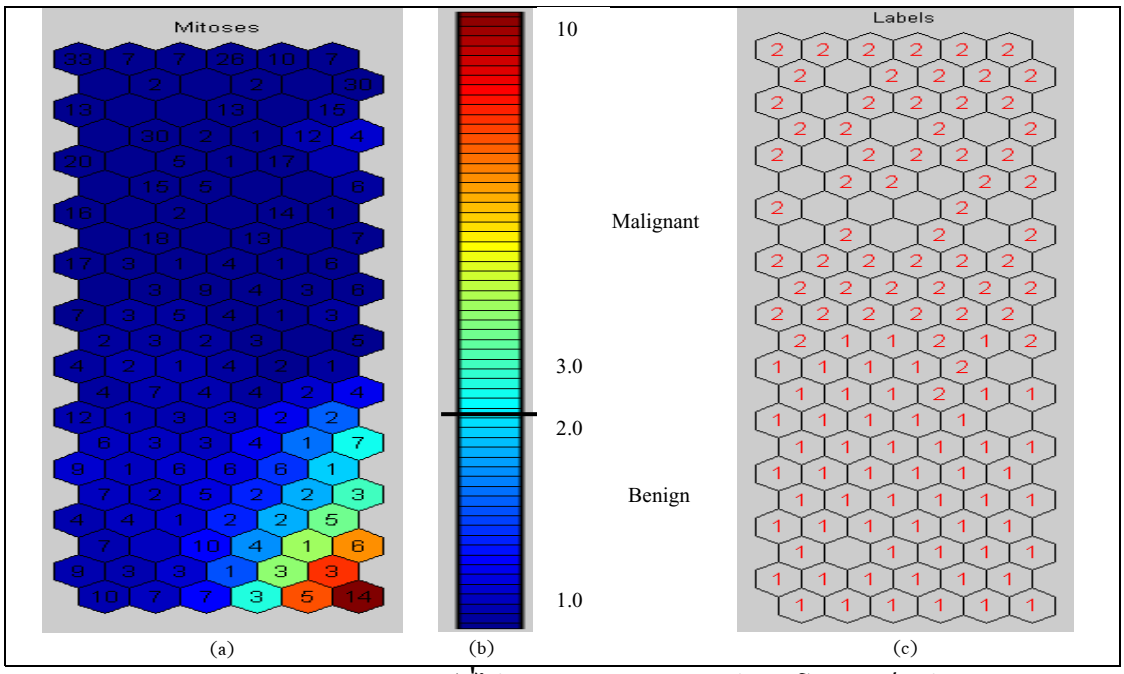
ภาพประกอบ 4.20 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_6 (b) ช่วงค่า Color Bar ของ x_6 (c) Label ของกลุ่มทั้ง 2 กลุ่ม



ภาพประกอบ 4.21 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_7 (b) ช่วงค่า Color Bar ของ x_7 (c) Label ของกลุ่มทั้ง 2 กลุ่ม



ภาพประกอบ 4.22 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_8 (b) ช่วงค่า Color Bar ของ x_8 (c) Label ของกลุ่มทั้ง 2 กลุ่ม



ภาพประกอบ 4.23 ผลลัพธ์ที่ได้จากการจัดกลุ่มฐานข้อมูลโรคมะเร็งเต้านม
(a) Component Plane ของ x_9 (b) ช่วงค่า Color Bar ของ x_9 (c) Label ของกลุ่มทั้ง 2 กลุ่ม

ขั้นตอนหลักที่ A2 การวิเคราะห์ Component Plane และ Label (Component Plane and Label Analysis) เป็นการเปลี่ยน Component Plane และ Color Bar ให้อยู่ในรูปแบบของเมทริกซ์ Component Plane และเมทริกซ์ Color Bar หลังจากนั้นทำการหากลุ่มที่เป็นไปได้มากที่สุด โดยพิจารณาที่กลุ่มและความถี่ของเมทริกซ์ Component Plane

ขั้นตอนหลักที่ A3 การสกัดคุณลักษณะที่สำคัญ (Feature Extraction Process) หาค่าความถูกต้องของแต่ละตัวแปรเข้า แสดงได้ดังตารางที่ 4.22 หลังจากนั้นหาค่าความถูกต้องที่ยอมรับได้ ตามสมการ (3.1) ซึ่งได้ค่าความถูกต้องที่ยอมรับได้ (ϕ) คือ 87.4 (92.4-5.0) ถ้าค่าความถูกต้องจากแต่ละตัวแปรเข้ามากกว่าหรือเท่ากับ ϕ แล้ว ให้ยอมรับค่าความถูกต้องเหล่านั้น และเก็บตัวแปรเหล่านั้นไว้ใน accepted accuracy list ซึ่งคือ x_2, x_3, x_5, x_6 และ x_8

ตารางที่ 4.22 ค่าความถูกต้องของแต่ละตัวแปรเข้า

ตัวแปร	ค่าความถูกต้อง	Accepted Accuracy List
x_1	60.4	No
x_2	92.4	Yes
x_3	90.8	Yes
x_4	85.8	No
x_5	89.6	Yes
x_6	89.4	Yes
x_7	77.4	No
x_8	89.6	No
x_9	76.3	Yes

ขั้นตอนหลักที่ A4 การสร้างกฎความรู้ (Generated Knowledge Based) เป็นการสร้างกฎที่เป็นไปได้ ตามสมการ (3.2) ซึ่งได้กฎทั้งหมด 1 กฎ และมีค่าความถูกต้อง 92.4% ดังภาพประกอบ 4.24

กฎที่ได้จากการสกัดความรู้ด้วย KESOM

R1 : If uniformity of cell size > 3 then malignant else benign

ค่าความถูกต้อง 92.4%

ภาพประกอบ 4.24 กฎและค่าความถูกต้องจากฐานข้อมูลโรคมะเร็งเต้านม

4.2.2 แบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (KESOM_MEP)

สำหรับการสกัดความรู้จากฐานข้อมูลโรคมะเร็งเต้านมที่อยู่ในรูปของกฎภาษาธรรมชาตินั้น มีการทำงาน 5 ขั้นตอนหลัก ตามรายละเอียดดังนี้

ขั้นตอนหลักที่ B1 กระบวนการแผนที่การจัดกลุ่มเอง (Self-Organizing Map Process) ทำการแบ่งกลุ่มข้อมูลด้วย Self-Organizing Map โดยใช้ Matlab SOM Toolbox เป็นเครื่องมือ จะได้ Component Plane, Color Bar และ Label ซึ่งเป็นเครื่องมือสำหรับวิเคราะห์ผลลัพธ์ที่ได้จากการจัดกลุ่ม ดังภาพประกอบ 4.15 ถึง ภาพประกอบ 4.23

ขั้นตอนหลักที่ B2 หลักเอ็นโทรพีค่าต่ำสุด (Minimization Entropy Principle Algorithm) เป็นขั้นตอนการหา potential threshold point หรือจุดแบ่งข้อมูลของแต่ละตัวแปรเข้า ซึ่งมีรายละเอียดขั้นตอนดังนี้

หาค่า $R_{\min}$ คือ ค่าน้อยที่สุด, หาค่า $R_{\max}$ คือ ค่ามากที่สุด, หาค่า σ คือ ค่าเบี่ยงเบนมาตรฐาน และคำนวณช่วงค่าที่ใช้ในการแทนตัวแปร จากสมการ (3.3) ซึ่ง x_1อยู่ในช่วงค่า [0,10.7] x_2 อยู่ในช่วงค่า [0,12.0] x_3 อยู่ในช่วงค่า [0,11.8] x_4 อยู่ในช่วงค่า [0,10.4] x_5 อยู่ในช่วงค่า [0,10.0] x_6 อยู่ในช่วงค่า [0,12.8] x_7 อยู่ในช่วงค่า [0,10.2] x_8 อยู่ในช่วงค่า [0,11.0] และ x_9 อยู่ในช่วงค่า [0,9.9]

ตารางที่ 4.23 ค่าน้อยที่สุดและค่ามากที่สุดของแต่ละกลุ่ม

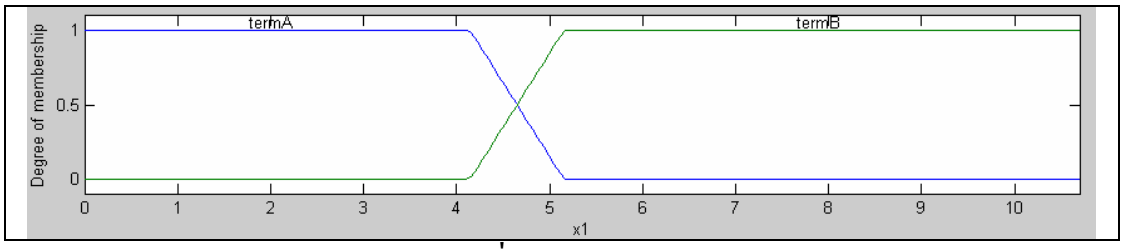
ตัวแปร	Benign		Malignant	
	Min	Max	Min	Max
x_1	1.0	5.1	4.2	8.5
x_2	1.0	3.8	2.6	9.2
x_3	1.0	4.2	2.5	9.1
x_4	1.0	3.8	2.2	8.1
x_5	1.7	4.9	2.7	8.1
x_6	1.0	4.9	1.8	9.7
x_7	1.1	4.3	2.5	8.1
x_8	1.0	5.9	1.9	8.4
x_9	1.0	1.9	1.1	8.6

จากตารางที่ 4.23 แสดงค่ามากที่สุดและค่าน้อยที่สุด และช่วงค่าของแต่ละกลุ่มทั้ง 9 ตัวแปรเข้า จะเห็นได้ว่า ทั้ง x_1 ถึง x_9 มีช่วงค่าที่คาบเกี่ยวกันทุกกลุ่ม เมื่อทราบว่าช่วงค่าใดที่คาบเกี่ยวกันแล้ว เรียงลำดับจากค่าน้อยไปมากในแต่ละตัวแปรเข้า แล้วหาค่า potential threshold point หรือจุดแบ่งข้อมูล หลังจากนั้นหาค่า entropy ของแต่ละ potential threshold point จากสมการ (3.4) ถึง (3.12) โดยใช้ขั้นตอนวิธีหลักเอ็นโทรพีค่าต่ำสุด เลือกค่า potential threshold point ที่มีค่า entropy น้อยที่สุดเก็บไว้ใน member_list ดังตารางที่ 4.24

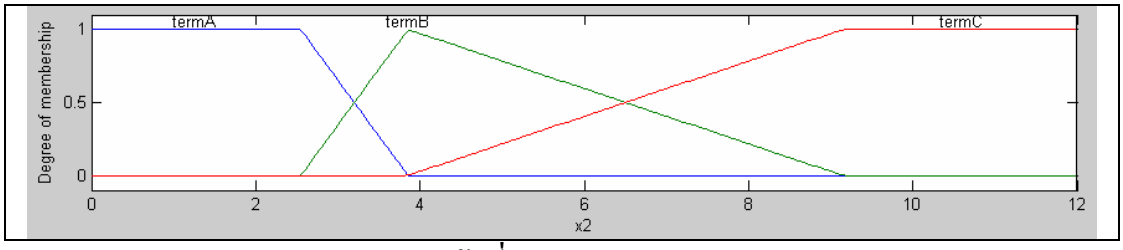
ตารางที่ 4.24 potential threshold point ที่เก็บใน member_list

ตัวแปร	Member_list	กลุ่ม
x_1	4.15, 5.15	Benign, Malignant
x_2	2.55, 3.85, 9.15	Benign, Malignant
x_3	2.95, 4.25, 9.05	Benign, Malignant
x_4	2.15, 3.15, 3.85, 8.00	Benign, Malignant
x_5	2.65, 3.25, 4.95, 8.05	Benign, Malignant
x_6	1.75, 2.45, 3.00, 5.00, 9.65	Benign, Malignant
x_7	3.35, 4.40, 8.05	Benign, Malignant
x_8	1.80, 2.15, 4.75, 5.85, 5.95, 8.35	Benign, Malignant
x_9	1.15, 1.25, 2.00, 7.90	Benign, Malignant

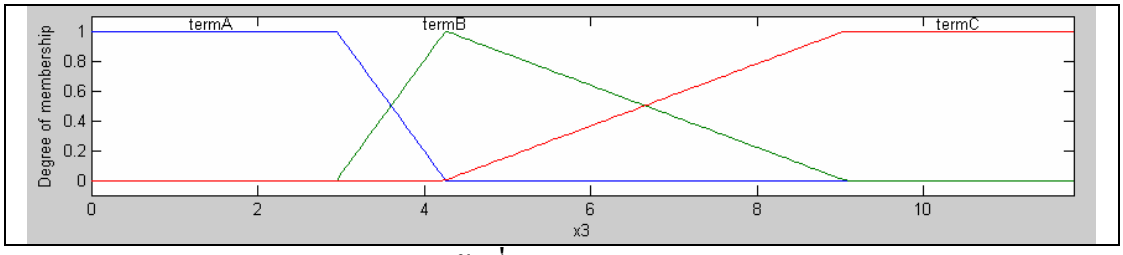
ขั้นตอนหลักที่ B3 การสร้างฟัซซี่เซต (Creating Fuzzy Set) นำค่า potential threshold point จาก member_list ไปสร้างฟัซซี่เซตและเทอมเซต ตามสมการ (3.15) ถึง (3.15) ดังภาพประกอบ 4.25 ถึงภาพประกอบ 4.33



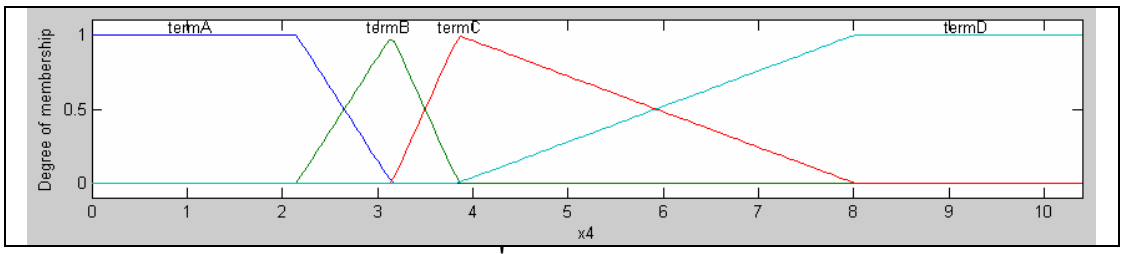
ภาพประกอบ 4.25 ฟัซซี่เซต และเทอมเซต 2 เทอม ของ x_1



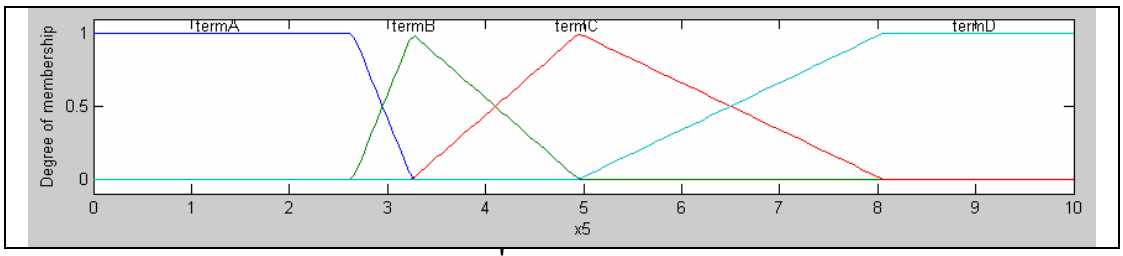
ภาพประกอบ 4.26 ฟัซซี่เซต และเทอมเซต 3 เทอม ของ x_2



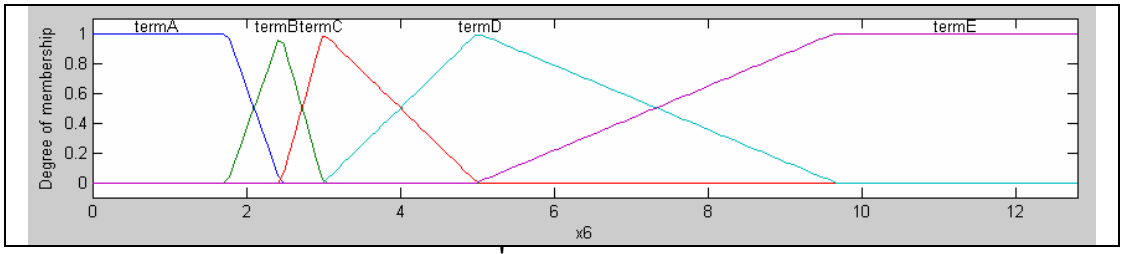
ภาพประกอบ 4.27 ฟัซซี่เซต และเทอมเซต 3 เทอม ของ x_3



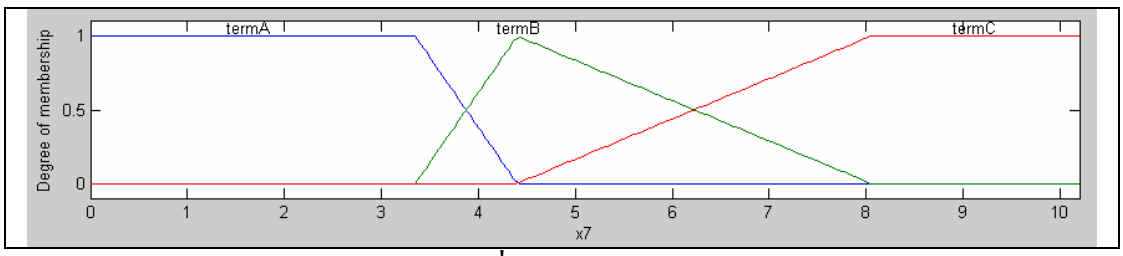
ภาพประกอบ 4.28 ฟัซซี่เซต และเทอมเซต 4 เทอม ของ x_4



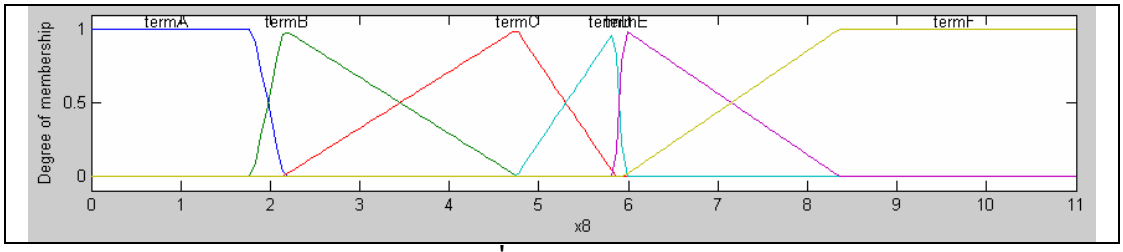
ภาพประกอบ 4.29 ฟัซซี่เซต และเทอมเซต 4 เทอม ของ x_5



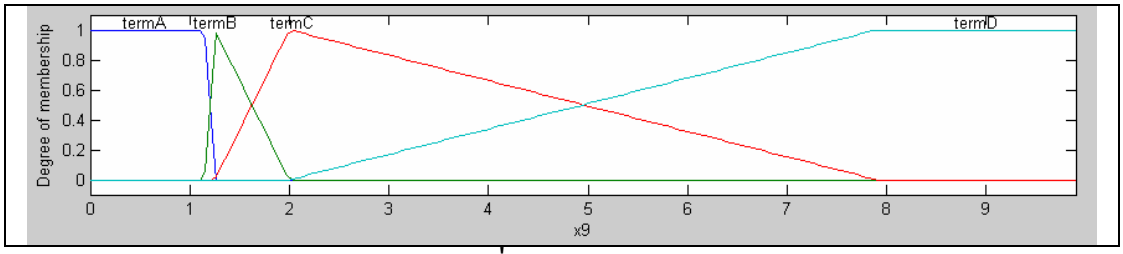
ภาพประกอบ 4.30 ฟัซซี่เซต และเทอมเซต 5 เทอม ของ x_6



ภาพประกอบ 4.31 ฟัซซี่เซต และเทอมเซต 3 เทอม ของ x_7



ภาพประกอบ 4.32 ฟัซซี่เซต และเทอมเซต 6 เทอม ของ x_8



ภาพประกอบ 4.33 ฟัซซี่เซต และเทอมเซต 4 เทอม ของ x_9

ขั้นตอนหลักที่ B4 การสร้างรูปแบบภาษาธรรมชาติ (Creating Linguistic Term Using Membership Function) แปลงค่าสี จาก Color Bar ของข้อมูลทุกตัวแปรเพื่อสร้างเป็นช่วงค่าภาษาธรรมชาติโดยดูค่าความเป็นสมาชิกที่เป็นไปได้ว่าอยู่ในช่วงค่าใด ดังตารางที่ 4.25 ถึง ตารางที่ 4.33 โดยสมการที่ใช้ในการประมาณค่าความเป็นสมาชิกของแต่ละช่วงค่าภาษาธรรมชาติ คือ ในกรณีฟังก์ชันความเป็นสมาชิกเป็นรูปสี่เหลี่ยมคางหมู พิจารณาดังสมการ (3.16) และในกรณีฟังก์ชันความเป็นสมาชิกเป็นรูปสามเหลี่ยม พิจารณาดังสมการ (3.17)

ตารางที่ 4.25 ช่วงค่าที่เป็นไปได้ของ x_1

Interval	Linguistic Term
0.0-4.6	termA
>4.6	termB

ตารางที่ 4.26 ช่วงค่าที่เป็นไปได้ของ x_2

Interval	Linguistic Term
0.0-3.1	termA
3.2-6.5	termB
>6.5	termC

ตารางที่ 4.27 ช่วงค่าที่เป็นไปได้ของ x_3

Interval	Linguistic Term
0.0–3.5	termA
3.6–6.6	termB
>6.6	termC

ตารางที่ 4.28 ช่วงค่าที่เป็นไปได้ของ x_4

Interval	Linguistic Term
0.0–2.6	termA
2.7–3.4	termB
3.5–5.9	termC
>5.9	termD

ตารางที่ 4.29 ช่วงค่าที่เป็นไปได้ของ x_5

Interval	Linguistic Term
0.0–2.9	termA
3.0–4.0	termB
4.1–6.5	termC
>6.5	termD

ตารางที่ 4.30 ช่วงค่าที่เป็นไปได้ของ x_6

Interval	Linguistic Term
0.0–2.0	termA
2.1–2.7	termB
2.8–3.9	termC
4.0–7.3	termD
>7.3	termE

ตารางที่ 4.31 ช่วงค่าที่เป็นไปได้ของ x_7

Interval	Linguistic Term
0.0-3.8	termA
3.9-6.2	termB
>6.2	termC

ตารางที่ 4.32 ช่วงค่าที่เป็นไปได้ของ x_8

Interval	Linguistic Term
0.0-1.9	termA
2.0-3.4	termB
3.5-5.3	termC
5.4-5.9	termD
6.0-7.1	termE
>7.1	termF

ตารางที่ 4.33 ช่วงค่าที่เป็นไปได้ของ x_9

Interval	Linguistic Term
0.0-1.2	termA
1.3-1.6	termB
1.7-4.9	termC
>4.9	termD

ขั้นตอนหลักที่ B5 การสร้างกฎความรู้ (Knowledge Extraction Using Rough Set) เป็นการสร้างกฎโดยใช้ราฟเซต สำหรับการสร้างกฎถ้า-แล้ว จากหลักการประมาณค่าขอบเขตล่างดังสมการ (3.18) และการประมาณค่าขอบเขตบนดังสมการ (3.19) กฎที่ยอมรับได้ต้องมีค่าค่าความมั่นใจมากกว่า 50% ดังแสดงได้ตามภาพประกอบ 4.34 ถึงภาพประกอบ 4.42

ตัวแปรเข้า x_1

Lower Approximation with Certainty Value

ไม่มี

Upper Approximation with Certainty Value

R1 : If x1 = termA then class 2.

Certainty Value = 95.2

R2 : If x1 = termB then class 1.

Certainty Value = 77.8

ค่าความถูกต้องทั้งหมด 80.4

ภาพประกอบ 4.34 กฎและค่าความถูกต้องที่ได้จาก x_1

ตัวแปรเข้า x_2

Lower Approximation with Certainty Value

R1 : If x2 = termC then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R2 : If x2 = termA then class 2.

Certainty Value = 96.5

R3 : If x2 = termB then class 1.

Certainty Value = 93.1

ค่าความถูกต้องทั้งหมด 92.7

ภาพประกอบ 4.35 กฎและค่าความถูกต้องที่ได้จาก x_2

ตัวแปรเข้า x_3

Lower Approximation with Certainty Value

R1 : If x3 = termC then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R2 : If x3 = termA then class 2.

Certainty Value = 94.9

R3 : If x3 = termB then class 1.

Certainty Value = 96.9

ค่าความถูกต้องทั้งหมด 92.3

ภาพประกอบ 4.36 กฎและค่าความถูกต้องที่ได้จาก x_3

ตัวแปรเข้า x_4

Lower Approximation with Certainty Value

R1 : If x4 = termD then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R2 : If x4 = termA then class 2.

Certainty Value = 94.6

R3 : If x4 = termB then class 1.

Certainty Value = 82.4

R4 : If x4 = termC then class 1.

Certainty Value = 94.1

ค่าความถูกต้องทั้งหมด 85.8

ภาพประกอบ 4.37 กฎและค่าความถูกต้องที่ได้จาก x_4

ตัวแปรเข้า x_5

Lower Approximation with Certainty Value

R1 : If x5 = termD then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R2 : If x5 = termA then class 2.

Certainty Value = 98.1

R3 : If x5 = termB then class 1.

Certainty Value = 87.5

R4 : If x5 = termC then class 1.

Certainty Value = 90.3

ค่าความถูกต้องทั้งหมด 89.6

ภาพประกอบ 4.38 กฎและค่าความถูกต้องที่ได้จาก x_5

ตัวแปรเข้า x_6

Lower Approximation with Certainty Value

R1 : If x6 = termA then class 2.

Certainty Value = 100.0

R2 : If x6 = termE then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R3 : If x6 = termB then class 2.

Certainty Value = 66.7

R4 : If x6 = termC then class 2.

Certainty Value = 60.0

R5 : If x6 = termD then class 1.

Certainty Value = 90.5

ค่าความถูกต้องทั้งหมด 90.1

ภาพประกอบ 4.39 กฎและค่าความถูกต้องที่ได้จาก x_6

ตัวแปรเข้า x_7

Lower Approximation with Certainty Value

R1 : If x7 = termC then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R2 : If x7 = termA then class 2.

Certainty Value = 93.1

R3 : If x7 = termB then class 1.

Certainty Value = 91.7

ค่าความถูกต้องทั้งหมด 90.7

ภาพประกอบ 4.40 กฎและค่าความถูกต้องที่ได้จาก x_7

ตัวแปรเข้า x_8

Lower Approximation with Certainty Value

R1 : If x8 = termA then class 2.

Certainty Value = 100.0

R2 : If x8 = termD then class 1.

Certainty Value = 100.0

R3 : If x8 = termF then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R4 : If x8 = termB then class 1.

Certainty Value = 83.3

R5 : If x8 = termC then class 1.

Certainty Value = 87.5

R6 : If x8 = termE then class 1.

Certainty Value = 93.8

ค่าความถูกต้องทั้งหมด 86.1

ภาพประกอบ 4.41 กฎและค่าความถูกต้องที่ได้จาก x_8

ตัวแปรเข้า x_9

Lower Approximation with Certainty Value

R1 : If x9 = termD4 then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R2 : If x9 = termA then class 2.

Certainty Value = 84.4

R3 : If x9 = termB then class 1.

Certainty Value = 91.3

R4 : If x9 = termC then class 1.

Certainty Value = 95.7

ค่าความถูกต้องทั้งหมด 79.0

ภาพประกอบ 4.42 กฎและค่าความถูกต้องที่ได้จาก x_9

จากนั้นเลือกเฉพาะตัวแปรเข้าที่มีค่าความถูกต้องสูงที่สุด ซึ่งกฎที่ได้จาก "x_2" เป็นกฎที่ยอมรับได้ ดังภาพประกอบ 4.43 โดยที่ R1 คือกฎที่อยู่ใน lower approximation R2 และ R3 คือกฎที่อยู่ใน upper approximation มีค่าความถูกต้องทั้งหมด 92.7

กฎที่ได้จากการสกัดความรู้ด้วย KESOM_MEP

Lower Approximation with Certainty Value

R1 : If x2 = termC then class 1.

Certainty Value = 100.0

Upper Approximation with Certainty Value

R2 : If x2 = termA then class 2.

Certainty Value = 96.5

R3 : If x2 = termB then class 1.

Certainty Value = 93.1

ค่าความถูกต้องทั้งหมด 92.7

ภาพประกอบ 4.43 กฎและค่าความถูกต้องที่ได้จากสกัดความรู้

4.2.3 เปรียบเทียบผลการทดลองและวิจารณ์ผลการทดลอง

แบบจำลองการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง (KESOM)

ผลการทดลองจากฐานข้อมูลโรคมะเร็งเต้านม แบบจำลองนี้สามารถสกัดความรู้ที่อยู่ในรูปของกฎทั่วไปได้ 1 กฎ ซึ่งกฎนี้เลือกมาจากตัวแปรเข้าที่มีค่าความถูกต้องของกฎมากที่สุด โดยกฎที่ได้นั้นไม่มีตัวเชื่อมเงื่อนไข

แบบจำลองการสกัดความรู้จากแผนที่การจัดกลุ่มเองโดยใช้หลักเอ็นโทรพีค่าต่ำสุด (KESOM_MEP)

1. ผลการทดลองจากฐานข้อมูลโรคมะเร็งเต้านม แบบจำลองนี้สามารถสกัดความรู้ที่อยู่ในรูปของกฎภาษาธรรมชาติ โดยที่จำนวนของกฎขึ้นอยู่กับระดับค่าความมั่นใจ (Certainty Value) ของกฎที่ได้จากราฟเซต ซึ่งโดยทั่วไปแล้วกฎที่ได้จากการประมาณค่าขอบเขตล่างจะเป็นกฎที่ยอมรับได้ ที่มีระดับค่าความมั่นใจ 100% แต่กฎที่ได้จากการประมาณค่าขอบเขตบนต้องมีระดับค่าความมั่นใจมากกว่า 50% จึงจะเป็นกฎที่ยอมรับได้ จะเห็นได้ว่าระดับค่าความมั่นใจนั้นเป็นตัวกรองเพื่อให้ได้กฎที่น้อยและมีความสำคัญเท่านั้น ดังภาพประกอบ 4.34 ถึงภาพประกอบ 4.42

2. กฎที่ได้จากแบบจำลองนี้มาจากตัวแปรเข้าที่มีค่าความถูกต้องมากที่สุดโดยกฎจะไม่มีตัวเชื่อมเงื่อนไข

3. รูปแบบภาษาธรรมชาติหรือเทอมเซตได้มาจากขั้นตอนหลักเอ็นโทรพีค่าต่ำสุด ซึ่งเป็นวิธีการสร้างเทอมเซตจากข้อมูลของตัวแปรเข้าต่าง ๆ อย่างอัตโนมัติ ดังภาพประกอบ 4.25 ถึงภาพประกอบ 4.33

ตารางที่ 4.34 แสดงผลการเปรียบเทียบการสกัดความรู้โดยใช้แผนที่การจัดกลุ่มเอง KESOM และ KESOM_MEP กับวิธีการอื่น ๆ พบว่าวิธีการสกัดความรู้ KESOM ได้จำนวนกฎในรูปกฎทั่วไป 1 กฎ โดยไม่มีตัวเชื่อมเงื่อนไข ซึ่งมีค่าความถูกต้อง 92.7% จะเห็นได้ว่า KESOM มีจำนวนกฎ จำนวนตัวเชื่อมเงื่อนไขที่น้อยกว่า และมีค่าความถูกต้องมากกว่า NEFCLASS [7] ที่เป็นกฎภาษาธรรมชาติ วิธีการสกัดความรู้ KESOM_MEP ได้จำนวนกฎที่เป็นแบบกฎภาษาธรรมชาติ 3 กฎ โดยไม่มีตัวเชื่อมเงื่อนไข ซึ่งมีค่าความถูกต้อง 92.7% จะเห็นได้ว่า KESOM_MEP ได้จำนวนกฎน้อยกว่า จำนวนตัวเชื่อมเงื่อนไขน้อยกว่า และมีค่าความถูกต้องมากกว่า NEFCLASS [7] ที่เป็นกฎภาษาธรรมชาติ

ตารางที่ 4.34 ผลการทดลองเมื่อเปรียบเทียบการสกัดความรู้ด้วย KESOM และ KESOM_MEP กับขั้นตอนวิธีอื่น ๆ ของฐานข้อมูลโรคมะเร็งเต้านม

Method	**#Rules**	**#Condition**	**Fuzzy Rule**	**Accuracy (%)**
KESOM	**1**	**0**	**No**	**92.4%**
KESOM_MEP	**3**	**0**	**Yes**	**92.7%**
NEFCLASS [1]	3	24	Yes	92.7%