การลดหลักเกณฑ์เชื่อมโยงจากการผ่านตัววัดค่าสนับสนุนแบบอ่อนที่ขึ้นอยู่กับความเชื่อมั่น ประพจน์แย้งสลับที่และหลักเกณฑ์กำหนดทิศทาง

นางสาวธนาพร ธัญญะเศรษฐ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคณนา ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2551

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

WEAK SUPPORT ASSOCIATION RULE REDUCTION BASED ON CONTRAPOSITIVE CONFIDENCE AND DIRECTION SETTING RULES

Miss Tanaporn Tunyaset

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Computational Science
Department of Mathematics
Faculty of Science
Chulalongkorn University
Academic Year 2008
Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การลดหลักเกณฑ์เชื่อมโยงจากการผ่านตัววัดค่าสนับสนุนแบบอ่อนที่ขึ้นอยู่กับความเชื่อมั่นประพจน์แย้งสลับที่และหลักเกณฑ์กำหนดทิศทาง
โดย	นางสาวธนาพร ธัญญะเศรษฐ์
สาขาวิชา	วิทยาการคณนา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร. กรุง สินอภิรมย์สราญ

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิทยาศาสตร์

(ศาสตราจารย์ ดร. สุพจน์ หารหนองบัว)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. จารุโลจน์ จงสถิตย์วัฒนา)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ผู้ช่วยศาสตราจารย์ ดร. กรุง สินอภิรมย์สราญ)

..... กรรมการ

(รองศาสตราจารย์ ดร. พรชัย สาตรวาหา)

..... กรรมการ

(อาจารย์ ดร. สิริพันธุ์ สงวนสินธุกุล)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ธนาพร ธัญญะเศรษฐ์ : การลดหลักเกณฑ์เชื่อมโยงจากการผ่านตัววัดค่าสนับสนุนแบบอ่อนที่ขึ้นอยู่กับความเชื่อมั่นประพจน์แย้งสลับที่และหลักเกณฑ์กำหนดทิศทาง. (WEAK SUPPORT ASSOCIATION RULE REDUCTION BASED ON CONTRAPOSITIVE CONFIDENCE AND DIRECTION SETTING RULES) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ. ดร. กรุง สินอภิรมย์สราญ, 64หน้า.

การวิเคราะห์หลักเกณฑ์เชื่อมโยง เป็นหนึ่งในเทคนิคการทำเหมืองข้อมูลที่ค้นหาหลักเกณฑ์น่าสนใจหรือความสัมพันธ์ที่น่าสนใจในข้อมูล เพื่อตัดสินความน่าสนใจตัววัดสองค่าคือค่าสนับสนุนและค่าความเชื่อมั่นมักถูกนำมาใช้ ปกติผู้เชี่ยวชาญจำเป็นต้องกำหนดค่าสนับสนุนน้อยสุดและค่าความเชื่อมั่นน้อยสุดให้เหมาะสม เพื่อกรองหลักเกณฑ์เชื่อมโยงที่ไม่น่าสนใจออก ปัจจุบันตัววัดอีกคู่ซึ่งถูกนำเสนอคือค่าสนับสนุนแบบอ่อนและค่าความเชื่อมั่น ค่าสนับสนุนแบบอ่อนเป็นตัววัดที่อธิบายความน่าจะเป็นของเหตุการณ์ซึ่งไม่ขัดแย้งกับหลักเกณฑ์ค่าสนับสนุนแบบอ่อนน้อยสุดที่กำหนดต้องสูงกว่า 0.5 เพราะหลักเกณฑ์ที่ยอมรับข้อมูลที่ขัดแย้งเกินกว่า 50% ไม่มีประโยชน์ ด้วยการกำหนดค่าสนับสนุนแบบอ่อนน้อยสุดและค่าความเชื่อมั่นน้อยสุด ขั้นตอนวิธีดับบลิวเอสสามารถค้นหาหลักเกณฑ์ที่มีค่าความเชื่อมั่นสูงออกมาได้ จากหลักดังกล่าวเป็นสาเหตุให้ปริมาณหลักเกณฑ์ที่ได้ออกมามีปริมาณมาก เพื่อลดปริมาณหลักเกณฑ์ที่มากเหล่านี้ งานวิจัยนี้จึงเสนอวิธีกรองสองรูปแบบคือการกรองด้วยค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุดและการกรองด้วยหลักเกณฑ์กำหนดทิศทาง ขั้นตอนวิธีซีซีดับบลิวเอสประยุกต์ค่าสนับสนุนแบบอ่อนน้อยสุด กับค่าความเชื่อมั่นน้อยสุดพร้อมกับค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุด ในขณะที่ขั้นตอนวิธีดีเอสอาซีซีดับบลิวเอสเพิ่มการประยุกต์แนวคิดหลักเกณฑ์กำหนดทิศทางกับขั้นตอนวิธีซีซีดับบลิวเอส เพื่อเปรียบเทียบระหว่างขั้นตอนวิธีทั้งหมดตัวประเมินความไวถูกเลือกมาใช้ประเมิน ผลการทดลองที่ได้แสดงให้เห็นว่า ขั้นตอนวิธีดีเอสอาซีซีดับบลิวเอสมีประสิทธิภาพในการลดจำนวนหลักเกณฑ์ที่ได้ออกมาปริมาณมากในขณะที่ค่าความไวไม่แตกต่างกับขั้นตอนวิธีอื่น ๆ

ภาควิชา คณิตศาสตร์ ลายมือชื่อนิสิต.....

สาขาวิชา วิทยาการคณนา ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ปีการศึกษา 2551

4872308023 : MAJOR COMPUTATIONAL SCIENCE

KEYWORDS : ASSOCIATION RULE / MINIMUM SUPPORT / MINIMUM CONFIDENCE / MINIMUM WEAK SUPPORT / SENSITIVITY

TANAPORN TUNYASET : WEAK SUPPORT ASSOCIATION RULE REDUCTION BASED ON CONTRAPOSITIVE CONFIDENCE AND DIRECTION SETTING RULES. ADVISOR : ASST. PROF. KRUNG SINAPIROMSARAN, Ph.D., 64 pp.

Association analysis is one of the data mining techniques that extracts interesting rules or interesting relationship within data. In order to determine interestingness, two measures, the support and confidence are used. Normally, experts must set the appropriate minimum support and minimum confidence to filter uninteresting association rules out. Recently, a pair of measures are presented called the weak support and confidence. The weak support is the measure that describes the probability of events which do not contradict the rule. Its minimum weak support must be set higher than 0.5 since rules that admits more than 50% chance of contradictory events are useless. With the setting of minimum weak support and the minimum confidence, the WS algorithm can extract significant confidence rules. This cause a large number of rules to be generated. To reduce the number of these rules, this research proposes two filtering methods, the minimum contrapositive confidence filter and direction setting rule filter. The CCWS algorithm applies the minimum weak support and the minimum confidence together with the minimum contrapositive confidence while the DSCCWS algorithm additionally applies the direction setting rule concept to the CCWS algorithm. In order to compare among these algorithms, the sensitivity evaluation is selected. Our result shows that DSCCWS algorithm can effectively reduce a significant number of rules while maintains similar sensitivity to the rest of other algorithms.

Department : Mathematics Student's Signature .....

Field of Study : Computational Science Advisor's Signature .....

Academic Year : 2008

กิตติกรรมประกาศ

ผู้วิจัยขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. กรุง สินอภิรมย์สราญ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ท่านได้กรุณาให้ความรู้ คำแนะนำ และคำปรึกษาต่างๆ ที่ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. จารุโลจน์ จงสถิตย์วัฒนา ประธานกรรมการ รองศาสตราจารย์ ดร. พรชัย สาตรวาหา กรรมการ และอาจารย์ ดร. สิริพันธุ์ สงวนสินธุกุล กรรมการ ที่ได้ให้คำปรึกษาและการช่วยเหลือตลอดระยะเวลาในการทำงานวิจัยนี้ ซึ่งทำให้วิทยานิพนธ์ฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น

ขอกราบขอบพระคุณมารดา ตลอดจนพี่น้องในครอบครัวที่คอยเป็นกำลังใจ และช่วยเหลือผู้วิจัยมาโดยตลอด และขอบคุณเพื่อนๆ ทุกคนในความช่วยเหลือและเป็นกำลังใจให้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญภาพ..... ญ

บทที่ 1 บทนำ..... 1

บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 9

2.1 หลักเกณฑ์เชื่อมโยง..... 9

2.1.1 นิยาม..... 9

2.1.2 นิยามตัววัด..... 10

2.1.3 การสร้างหลักเกณฑ์เชื่อมโยง..... 11

2.2 ค่าสนับสนุนแบบอ่อน..... 12

2.2.1 ความสัมพันธ์ระหว่างค่าสนับสนุนแบบอ่อน ค่าสนับสนุน และค่าความเชื่อมั่น..... 13

2.2.2 ขั้นตอนวิธีค่าสนับสนุนแบบอ่อน..... 14

2.3 ค่าความเชื่อมั่นประพจน์แย้วสลับที่..... 16

2.3.1 ความสัมพันธ์ระหว่างค่าความเชื่อมั่นประพจน์แย้งสลับที่กับค่าความเชื่อมั่น..... 17

2.4 ขั้นตอนวิธีการหาหลักเกณฑ์กำหนดทิศทาง..... 20

2.4.1 ขั้นตอนวิธี P-DS..... 24

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

หน้า

บทที่ 3. การลดหลักเกณฑ์เชื่อมโยง..... 28

3.1 การสร้างหลักเกณฑ์เชื่อมโยง..... 28

3.2 การลดหลักเกณฑ์เชื่อมโยง..... 29

3.3 ขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยง..... 32

บทที่ 4 การทดลอง การเปรียบเทียบผลการทดลอง และผลการทดลอง..... 37

4.1 ข้อมูลที่ใช้ทดลอง..... 37

4.2 วิธีการทดลอง..... 39

4.3 เปรียบเทียบผลการทดลอง..... 39

4.4 ผลการทดลอง..... 41

บทที่ 5 สรุปผลการทดลอง และข้อเสนอแนะ..... 54

รายการอ้างอิง..... 56

ภาคผนวก..... 58

ประวัติผู้เขียนวิทยานิพนธ์..... 64

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

		หน้า
ตารางที่2.1	ข้อมูลที่แบ่งออกเป็นสี่กลุ่มตามตรรกะของ X, Y.....	13
ตารางที่2.2	ค่าสนับสนุนแบบอ่อนจากการคำนวณค่าสนับสนุนและค่าความเชื่อมั่น.....	14
ตารางที่2.3	ตารางค่าความจริงของ $p \to q$ และ $\neg q \to \neg p$	17
ตารางที่2.4	ตารางการจรณ์ (contingency table) ที่มีขนาด 2×2	21
ตารางที่2.5	ข้อมูลการขายสินค้าในร้านค้าปลีก.....	22
ตารางที่3.1	ตารางของข้อมูลที่เจาะจงการพบ x.....	31
ตารางที่3.2	ตารางของข้อมูลที่เจาะจงการไม่พบ x.....	31
ตารางที่4.1	แสดงข้อมูลสรุปที่ใช้ในการทดลอง.....	37
ตารางที่4.2	ตารางนิยาม TP, FP, FN, TN.....	40
ตารางที่ก.1	เป็นผลการทดลองจากข้อมูล breast-cancer.....	59
ตารางที่ก.2	เป็นผลการทดลองจากข้อมูล chess.....	60
ตารางที่ก.3	เป็นผลการทดลองจากข้อมูล connect.....	61
ตารางที่ก.4	เป็นผลการทดลองจากข้อมูล mushroom.....	62
ตารางที่ก.5	เป็นผลการทดลองจากข้อมูล zoo.....	63

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

		หน้า
รูปที่ 1.1	กระบวนการสืบค้นความรู้ในฐานข้อมูล.....	2
รูปที่ 2.1	ขั้นตอนวิธี WS.....	15
รูปที่ 2.2	ขั้นตอนการทำงานของขั้นตอนวิธีการกำหนดทิศทาง.....	20
รูปที่ 2.3	ขั้นตอนวิธี P – DS.....	25
รูปที่ 3.1	ขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยง.....	33
รูปที่ 4.1	จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล breast-cancer สำหรับขั้นตอนวิธี WS.	42
รูปที่ 4.2	จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล breast-cancer สำหรับขั้นตอนวิธี DSRCCWS.....	42
รูปที่ 4.3	ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงของข้อมูล breast-cancer สำหรับขั้นตอนวิธี WS.....	43
รูปที่ 4.4	ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล breast-cancer สำหรับขั้นตอนวิธี DSRCCWS.....	43
รูปที่ 4.5	จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล chess สำหรับขั้นตอนวิธี WS.....	44
รูปที่ 4.6	จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล chess สำหรับขั้นตอนวิธี DSRCCWS.	44
รูปที่ 4.7	ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงของข้อมูล chess สำหรับขั้นตอนวิธี WS..	45
รูปที่ 4.8	ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล chess สำหรับขั้นตอนวิธี DSRCCWS.....	45
รูปที่ 4.9	จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล connect สำหรับขั้นตอนวิธี WS.....	46
รูปที่ 4.10	จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล connect สำหรับขั้นตอนวิธี DSRCCWS.....	46
รูปที่ 4.11	ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล connect สำหรับขั้นตอนวิธี WS.....	47
รูปที่ 4.12	ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล connect สำหรับขั้นตอนวิธี DSRCCWS.....	47
รูปที่ 4.13	จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล mushroom สำหรับขั้นตอนวิธี WS.....	49

รูปที่ 4.14 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล mushroom สำหรับขั้นตอนวิธี DSRCCWS..... 49
รูปที่ 4.15 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล mushroom สำหรับขั้นตอนวิธี WS..... 50
รูปที่ 4.16 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล mushroom สำหรับขั้นตอนวิธี DSRCCWS..... 50
รูปที่ 4.17 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล zoo สำหรับขั้นตอนวิธี WS..... 51
รูปที่ 4.18 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล zoo สำหรับขั้นตอนวิธี DSRCCWS... 51
รูปที่ 4.19 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล zoo สำหรับขั้นตอนวิธี WS..... 52
รูปที่ 4.20 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล zoo สำหรับขั้นตอนวิธี DSRCCWS..... 52

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

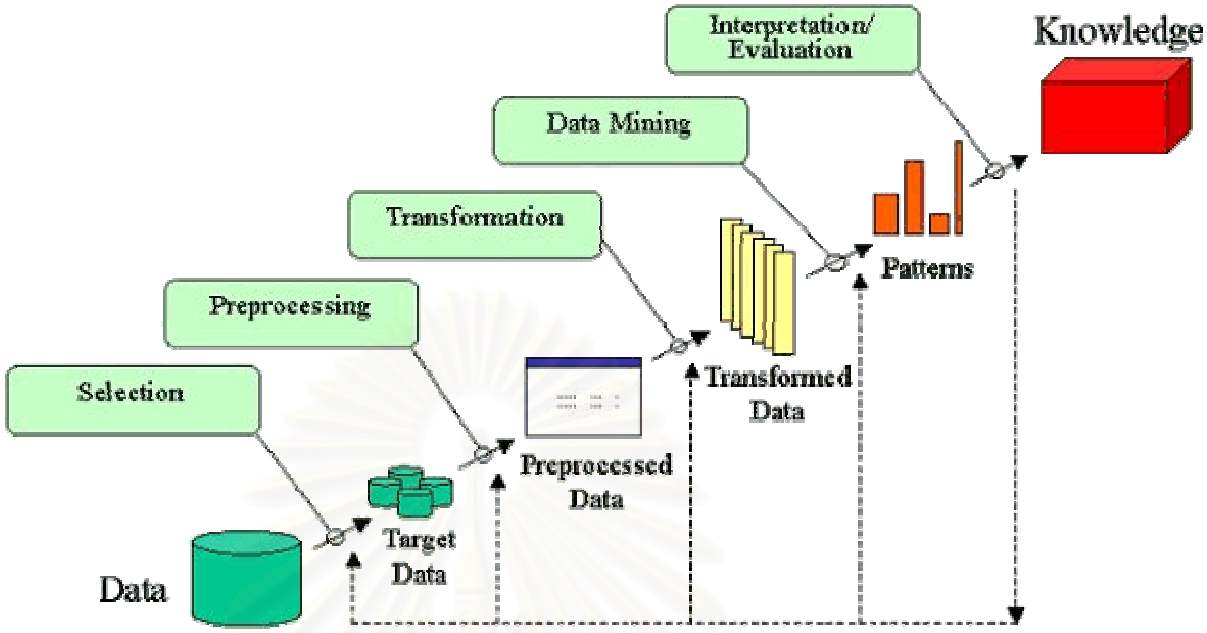
บทที่ 1

บทนำ

ในปัจจุบันทั้งหน่วยงานภาครัฐ และผู้ประกอบการภาคเอกชนได้ให้ความสำคัญกับข้อมูล ทำให้มีการพัฒนาการจัดเก็บข้อมูลขององค์กรที่มีประสิทธิภาพมากขึ้น เมื่อมีการจัดเก็บข้อมูลที่มีประสิทธิภาพทำให้ข้อมูลของแต่ละองค์กรมีปริมาณมาก และถ้าองค์กรขาดการจัดเรียงข้อมูลให้เป็นระบบก็จะทำให้เกิดความยุ่งยากในการเข้าถึงและค้นคืน แต่ละองค์กรอาจจะต้องเสียค่าใช้จ่ายจำนวนมหาศาลในการเก็บรักษาข้อมูลเหล่านั้นไว้โดยไม่จำเป็น เพื่อให้ได้รับประโยชน์จากข้อมูลจึงได้มีการนำวิธีการทำคลังข้อมูล (data warehousing) [1,9] มาช่วย แต่ในการที่จะนำข้อมูลที่น่าสนใจหรือสำคัญต่อองค์กรออกมาจะต้องใช้ผู้เชี่ยวชาญขององค์กรมาช่วยวิเคราะห์ข้อมูล แต่จากปริมาณข้อมูลที่มากจึงทำให้เกิดการพัฒนาการทำเหมืองข้อมูล (data mining) [1,7,8] ที่เป็นวิธีการวิเคราะห์ข้อมูลโดยอัตโนมัติ

การทำคลังข้อมูล (data warehousing) คือการปรับเปลี่ยนฐานข้อมูลขนาดใหญ่ที่รวบรวมข้อมูลทั้งจากแหล่งข้อมูลภายใน และภายนอกองค์กร โดยมีรูปแบบและวัตถุประสงค์ของการจัดเก็บข้อมูลแตกต่างจากฐานข้อมูลปฏิบัติการทั่วไป วิธีหนึ่งในการออกแบบระบบคลังข้อมูลเริ่มจากการศึกษาฐานข้อมูลจากระบบงานย่อย (data mart) ของแต่ละระบบงานในองค์กรก่อน แล้วจึงนำส่วนย่อยนั้นมารวมเป็นระบบคลังข้อมูลขององค์กรเพื่อนำข้อมูลมาทำการวิเคราะห์ต่อไป ในการวิเคราะห์ข้อมูลในคลังข้อมูลมีอยู่ 2 ประเภทใหญ่ คือ โอแลป (OnLine Analytical Processing) [1] วิธีนี้จะต้องใช้ผู้เชี่ยวชาญในการวิเคราะห์ และกำหนดทิศทางในการนำไปสู่ความรู้ที่ซ่อนอยู่ อีกประเภทหนึ่งคือ การทำเหมืองข้อมูล (data mining)

การทำเหมืองข้อมูล เป็นกระบวนการค้นหาและวิเคราะห์ข้อมูลปริมาณมากเพื่อให้ได้ความรู้ที่น่าสนใจ โดยความรู้ที่ค้นหานั้นอาจจะออกมาเป็น รูปแบบ แนวทาง หรือความสัมพันธ์ที่ซ่อนอยู่ในข้อมูล การทำเหมืองข้อมูลมีหลากหลายชื่อที่ใช้ ได้แก่ data / pattern analysis, knowledge extraction, knowledge discovery in databases (KDD) [7,8] เป็นต้น การที่จะทำเหมืองข้อมูลจะต้องมีการกำหนดวัตถุประสงค์ที่ชัดเจน เพื่อที่จะได้กำหนดขั้นตอนการทำงานให้ตรงตามเป้าหมายและได้ผลลัพธ์ที่ต้องการ โดยกระบวนการการทำเหมืองข้อมูลสามารถเลือกใช้กระบวนการในการสืบค้นความรู้ในฐานข้อมูล (knowledge discovery in database) ดังรูปที่ 1.1



รูปที่ 1.1 กระบวนการสืบค้นความรู้ในฐานข้อมูล

กระบวนการสืบค้นความรู้ในฐานข้อมูลมีอยู่ 5 ขั้นตอน ขั้นตอนแรกคือขั้นตอนการคัดเลือกข้อมูล (data selection) ในขั้นตอนนี้จะทำการคัดเลือกข้อมูลจากฐานข้อมูลเพื่อนำไปใช้ทำเหมืองข้อมูล ซึ่งข้อมูลที่เลือกจะสอดคล้องกับวัตถุประสงค์ของงาน เพื่อให้ได้ผลลัพธ์ดังที่ต้องการ แล้วทำการแบ่งข้อมูลออกเป็น 2 ส่วน ข้อมูลส่วนแรกเป็นข้อมูลที่นำไปสร้างตัวแบบหรือรูปแบบตามวัตถุประสงค์เรียกว่าข้อมูลทดลอง (train data) ข้อมูลอีกส่วนหนึ่งเก็บไว้สำหรับการประเมินเรียกว่า ข้อมูลทดสอบ (test data) ขั้นตอนที่สองคือ ขั้นตอนการเตรียมข้อมูล (preprocessing) ในขั้นตอนนี้จะต้องจัดการกับข้อมูลที่มีความผิดปรกติ ดังนั้นจะมีการทำความสะอาดข้อมูล (data cleansing) เพราะถ้าปล่อยข้อมูลที่มีความผิดปรกติผ่านกระบวนการทำเหมืองข้อมูลต่อไป ผลลัพธ์ที่ได้ก็จะเชื่อถือไม่ได้และไม่มีคุณภาพ ขั้นตอนที่สามคือ ขั้นตอนการแปลงข้อมูล (transformation) ในขั้นตอนนี้จะทำการแปลงข้อมูลให้อยู่ในรูปแบบที่ต้องการ โดยมีทั้งการสร้างลักษณะประจำ (attribute) ใหม่ การคำนวณค่ารวม (sum) การเปลี่ยนข้อมูลให้อยู่ในรูปแบบมาตรฐาน ซึ่งรวมถึงการลดรูปข้อมูล (data reduction) ขั้นตอนที่สี่คือ ขั้นตอนการทำเหมืองข้อมูล (data mining) ในขั้นตอนนี้จะทำการคัดเลือกเทคนิคการทำเหมืองข้อมูลและตัวแบบที่เหมาะสม เพื่อที่จะค้นหาความรู้ที่ซ่อนอยู่ในข้อมูล ขั้นตอนที่ห้าคือ ขั้นตอนการตีความหมายหรือการประเมินผล (interpretation / evaluation) ในขั้นตอนนี้จะนำผลลัพธ์ที่ได้

จากการทำเหมืองข้อมูลที่ออกมาเป็นความรู้ มาทำการแปลความหมายหรือทำการประเมินกับข้อมูลทดสอบที่เก็บไว้ที่ไม่ผ่านขั้นตอนการทำเหมืองข้อมูลมาทำการตรวจสอบความถูกต้อง

ในกระบวนการสืบค้นความรู้ แต่ละขั้นตอนนั้นสามารถที่จะย้อนกลับไปปรับเปลี่ยนขั้นตอนก่อนหน้าได้ เช่น ถ้าทำการประเมินผลแล้วได้ผลลัพธ์ที่ไม่สามารถตอบวัตถุประสงค์ที่ต้องการก็สามารถย้อนกลับไปเลือกข้อมูลที่เหมาะสมหรือทำการแปลงข้อมูลใหม่

การทำเหมืองข้อมูลเป็นการค้นหาความรู้ออกมาจากข้อมูล ซึ่งสามารถแบ่งประเภทการทำออกเป็น 2 ประเภท [1] คือ การทำเหมืองข้อมูลแบบอธิบาย (descriptive mining) เป็นการดึงความรู้ออกมาจากข้อมูล โดยเน้นการอธิบายแนวคิดหลักที่สั้นและกระชับ ประเภทที่สองคือ การทำเหมืองข้อมูลแบบทำนาย (predictive mining) เป็นการดึงความรู้ออกมาโดยการค้นหาตัวแบบ (model) ที่นำไปใช้ในการทำนายข้อมูล

เทคนิคการทำเหมืองข้อมูลมีมากมายหลายแบบ [1] ตัวอย่างเช่น การจำ แนกประเภท (classification) เป็นเทคนิคที่ข้อมูลมีการกำหนดลักษณะประจำที่ต้องการชัดเจน เรียกว่าลักษณะประจำเป้าหมาย เทคนิคนี้จะทำการแยกข้อมูลออกเป็นกลุ่มตามลักษณะประจำเป้าหมาย หลักการจำแนกประเภทจะมีอยู่สองขั้นตอน ขั้นแรก ทำการสร้างตัวแบบ ซึ่งตัวแบบก็คือผลลัพธ์ที่สามารถนำมาอธิบายเหตุการณ์ที่เกิดขึ้น ขั้นตอนที่สอง เป็นการนำตัวแบบที่สร้างมาจำแนกข้อมูลที่ยังไม่ทราบมาก่อน

เทคนิคจำแนกประเภทที่นิยมใช้ ได้แก่ ต้นไม้การตัดสินใจ (decision tree) วิธีนี้เป็นที่นิยมมากเนื่องจากมีลักษณะเหมือนแผนภูมิองค์กร คือเป็นกราฟต้นไม้ที่ประกอบไปด้วย จุดยอดภายใน แทนลักษณะประจำ กิ่ง แทน ผลของการทดสอบ ใบ แทน ค่าของลักษณะประจำเป้าหมาย

อีกเทคนิควิธีคือ การเกาะกลุ่ม (clustering) เป็นการที่ข้อมูลที่มีลักษณะเหมือนกันเกาะกลุ่มกัน เทคนิคนี้จะไม่มีการกำหนดลักษณะประจำเป้าหมายที่ต้องการ แต่ทำการเรียนรู้จากค่าลักษณะประจำของข้อมูล การวิเคราะห์การเกาะกลุ่ม คือ การหาวิธีการแบ่งข้อมูลออกเป็นกลุ่มตามลักษณะประจำที่มีของแต่ละข้อมูลโดยไม่มีการบอกลักษณะกลุ่ม ข้อมูลที่อยู่ในกลุ่มเดียวกันจะมีความคล้ายคลึงกัน และข้อมูลที่อยู่ต่างกลุ่มกันจะมีความแตกต่างกัน

เทคนิควิธีอีกอย่างคือ การสร้างหลักเกณฑ์เชื่อมโยง (association rule) [1,2,3,4,5] เป็นเทคนิคที่ค้นหาความสัมพันธ์ของข้อมูลจากข้อมูลขนาดใหญ่ที่มีอยู่เพื่อนำไปใช้ในการวิเคราะห์ หรือทำนายปรากฏการณ์ต่าง ๆ ตัวอย่างเช่น การวิเคราะห์การซื้อสินค้าของลูกค้า

เรียกว่า “Market Basket Analysis” [6] ซึ่งประเมินจากข้อมูลการซื้อที่รวบรวมไว้ ผลการวิเคราะห์ที่ได้จะออกมาเป็นหลักเกณฑ์ เป็นต้น

เทคนิคข้างต้นเป็นตัวอย่างเทคนิคในการทำเหมืองข้อมูล และยังมีเทคนิคอีกมากที่สามารถนำมาช่วยในการค้นหาความรู้ ซึ่งในการเลือกใช้เทคนิควิธีขึ้นอยู่กับข้อมูลและวัตถุประสงค์ของผู้บริหาร แล้วจึงคัดเลือกเทคนิควิธีที่เหมาะสม งานวิจัยนี้สนใจการวิเคราะห์หลักเกณฑ์เชื่อมโยงที่มีการใช้ตัววัดความน่าสนใจแบบใหม่

ความสัมพันธ์ที่เชื่อมโยงลักษณะประจำนั้นอยู่ในรูปแบบของหลักเกณฑ์ ที่เรียกว่า หลักเกณฑ์เชื่อมโยง (association rule) การทำเหมืองข้อมูลแบบหลักเกณฑ์เชื่อมโยงนั้นเพื่อหารูปแบบที่เกิดขึ้นบ่อยของลักษณะประจำ ซึ่งแนวคิดการสร้างหลักเกณฑ์เชื่อมโยงเกิดขึ้นจากความต้องการวิเคราะห์ข้อมูลการขายสินค้าหรือการวิเคราะห์มาเก็ตบาสเก็ต (market basket analysis)

ในธุรกิจการขาย ลูกค้าคือผู้ที่ซื้อสินค้าหรือบริการ ซึ่งลูกค้าแต่ละคนจะมีพฤติกรรมการซื้อที่แตกต่างกัน ทั้งลักษณะของสินค้าและจำนวนของสินค้า ดังนั้นการวิเคราะห์มาเก็ตบาสเก็ตจะพิจารณาจากสินค้าที่ลูกค้าซื้อแล้วหยิบสินค้าใส่ในรถเข็น หรือตระกร้าสินค้า (shopping basket) [6] กลุ่มสินค้าที่ลูกค้าแต่ละคนซื้อจะมีความแตกต่างกัน เมื่อนำข้อมูลการซื้อสินค้าของลูกค้ามารวมกัน แล้วจะได้ข้อมูลการซื้อสินค้าทั้งหมดของร้าน ในช่วงระยะเวลาหนึ่ง การวิเคราะห์มาเก็ตบาสเก็ต เป็นการวิเคราะห์ความสัมพันธ์ของข้อมูลการซื้อสินค้าของลูกค้าในลักษณะของภาพรวมไม่ใช่คนใดคนหนึ่ง และในงานวิจัยนี้จะไม่สนใจจำนวนสินค้าที่ลูกค้าซื้อเพราะจะพิจารณาเฉพาะการพบหรือไม่พบการซื้อสินค้าเท่านั้น ผลลัพธ์ที่ได้จากการวิเคราะห์มาเก็ตบาสเก็ตจะนำมาเขียนให้อยู่ในรูปของหลักเกณฑ์เชื่อมโยง

การวิเคราะห์ข้อมูลการขายสินค้าหรือการวิเคราะห์มาเก็ตบาสเก็ต นั้นเป็นการวิเคราะห์เพื่อดูความสัมพันธ์ของสินค้า ซึ่งเป็นตัวอธิบายพฤติกรรมการซื้อสินค้าของลูกค้า โดยพิจารณาจากการซื้อสินค้าในแต่ละครั้งของลูกค้า หรือในแต่ละใบเสร็จของลูกค้าว่ามีสินค้าชนิดใดถูกซื้อไปพร้อมกัน และได้รูปแบบความสัมพันธ์ของข้อมูลออกมา ข้อมูลจะถูกจัดเก็บอยู่ในรูปแบบของตารางข้อมูลเรียกแถวของตารางว่า “ระเบียน” และเรียกหลักของตารางว่า “ลักษณะประจำ” แล้วข้อมูลในหนึ่งระเบียนแทนการซื้อสินค้าหนึ่งครั้งหรือหนึ่งใบเสร็จ เรียกว่า ทรานแซคชัน (transaction) แต่ละระเบียนจะมีหมายเลขระบุการซื้อแต่ละครั้ง เรียกว่า ตัวระบุทรานแซคชัน (transaction identification : TID) [2,3,4,5] ส่วนข้อมูลในหนึ่งลักษณะประจำแทนสินค้าแต่ละสินค้า การจัดเก็บชื่อหรือรหัสของสินค้าให้เป็นตัวเลข แล้วทำการแปลงข้อมูลให้ชื่อหรือรหัส

เปลี่ยนไปอยู่ในรูปลักษณะประจำแบบทวิภาค (binary) หรือ เก็บข้อมูลเป็น 0 และ 1 นั้นก็คือ มีค่าเป็น 0 ก็ต่อเมื่อไม่พบการซื้อสินค้าชนิดนั้น และมีค่าเป็น 1 ก็ต่อเมื่อพบการซื้อสินค้าชนิดนั้น

การวิเคราะห์มาเก็ตบาสเก็ต จะถูกนำมาช่วยสร้างหลักเกณฑ์เชื่อมโยงของสินค้า เพื่อนำความรู้ที่ได้มาทำการวางแผนทางการตลาด และช่วยเพิ่มยอดขายหรือเพิ่มรายได้ให้กับร้านค้าหรือองค์กร ตัวอย่างเช่น การจัดรูปแบบการวางผังสินค้าภายในร้าน การนำข้อมูลไปจัดทำรายการพิเศษ (promotion) เพื่อทำให้ลูกค้าพึงพอใจ เป็นการดึงดูดลูกค้าให้มีการซื้อสินค้าเพิ่มมากขึ้น

การวิเคราะห์หลักเกณฑ์เชื่อมโยงถูกคิดค้นขึ้นในปี 1993 โดย Agrawal, Imielinski และ Swami [2] ต่อมาได้มีงานวิจัยที่พัฒนาขั้นตอนวิธีการสร้างหลักเกณฑ์เชื่อมโยงและตัววัดขึ้นอีกหลายงานวิจัย ตัวอย่างเช่น ขั้นตอนวิธีที่นิยมใช้ในการสร้างหลักเกณฑ์เชื่อมโยงก็คือ ขั้นตอนวิธีเอไพออรี (Apriori algorithm) [5] ซึ่งถูกพัฒนาขึ้นในปี 1994 โดย Agrawal และ Srikant ต่อมามีการพัฒนาการสร้างหลักเกณฑ์เชื่อมโยงไปหลากหลายทางเช่น การทำเหมืองเงื่อนไขบังคับพื้นฐาน (constraint-based mining) [10] เป็นงานวิจัยที่ศึกษาการสร้างหลักเกณฑ์เชื่อมโยงที่มีการกำหนดเงื่อนไขเพิ่มเติมเพื่อช่วยลดปริมาณหลักเกณฑ์ที่พิจารณา ในแนวทางเดียวกันยังมีขั้นตอนวิธี KORD (k-optimal rule discovery algorithm) [11] ขั้นตอนวิธีนี้จะทำการหาหลักเกณฑ์เชื่อมโยงออกมาจำนวน k หลักเกณฑ์ ต่อมามีงานวิจัยที่ทำการลดปริมาณหลักเกณฑ์เชื่องโยงโดยใช้การทดสอบความเป็นอิสระต่อกัน (chi-square test) และงานวิจัยที่ศึกษาการเลือกตัวแทนแบบกระชับของไอเทมเซตปรากฏบ่อย (mining concise representations of frequent itemset) [12] เป็นการค้นหาไอเทมเซตแบบปิด ซึ่งก็คือไอเทมเซตปรากฏบ่อยที่ใหญ่ที่สุดที่เกิดขึ้นในฐานข้อมูล นอกจากนี้ยังมีการนำการสร้างหลักเกณฑ์เชื่อมโยงไปใช้ในการแก้ปัญหาการจัดจำแนกประเภท (classification) เช่น ขั้นตอนวิธี CBA (classification based on association) เป็นขั้นตอนที่มีการสร้างตัวแบบจำแนกประเภท ซึ่งจะมีการกำหนดลักษณะประจำเป้าหมายให้เป็นพจน์หลังของหลักเกณฑ์

ในการคัดเลือกหลักเกณฑ์เชื่อมโยงที่น่าสนใจนั้นขึ้นอยู่กับตัววัด ตัววัดที่ใช้ในการหาหลักเกณฑ์เชื่อมโยงโดยทั่วไปนั้นนิยมใช้ตัววัด 2 ตัว คือค่าสนับสนุน (support) และค่าความเชื่อมั่น (confidence) หลักเกณฑ์เชื่อมโยงประกอบด้วย พจน์หน้า และพจน์หลัง ดังนั้นตัววัดทั้งสองจะแสดงความสัมพันธ์ของพจน์หน้ากับพจน์หลังของหลักเกณฑ์ ซึ่งหลักเกณฑ์เชื่อมโยงมีรูปแบบในการเขียนหลักเกณฑ์ $X \longrightarrow Y$ โดยที่ X และ Y เป็นเซตของข้อมูลที่เรียกว่า ไอเทม

เซต (itemset) [2] และ X แทนพจน์หน้าของหลักเกณฑ์ Y แทนพจน์หลังของหลักเกณฑ์ ซึ่งตัววัดทั้งค่าสนับสนุน และค่าความเชื่อมั่นสามารถอธิบายได้ดังนี้

ค่าสนับสนุนของหลักเกณฑ์ คือ ค่าความน่าจะเป็นที่จะพบ X และ Y

ค่าความเชื่อมั่นของหลักเกณฑ์ คือ ค่าความน่าจะเป็นแบบมีเงื่อนไขที่จะพบ Y ถ้ามีการพบ X

ซึ่งหลักเกณฑ์ที่น่าสนใจนั้น จะเป็นหลักเกณฑ์ที่มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนน้อยที่สุด (minimum support) และค่าความเชื่อมั่นมากกว่าหรือเท่ากับค่าความเชื่อมั่นน้อยที่สุด (minimum confidence)

ขั้นตอนวิธีเอไพออรี (Apriori algorithm) เป็นขั้นตอนวิธีการค้นหาหลักเกณฑ์เชื่อมโยงโดยเริ่มจากการหารูปแบบที่เกิดขึ้นบ่อยหรือ ไอเทมเซตปรากฏบ่อย (frequent itemset) [5] การเกิดขึ้นบ่อยพิจารณาจากไอเทมเซตที่มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนน้อยสุด แล้วนำไอเทมเซตปรากฏบ่อยไปสร้างหลักเกณฑ์เชื่อมโยงโดยใช้ค่าความเชื่อมั่นน้อยสุดได้มาเป็นตัววัดในการสร้างหลักเกณฑ์เชื่อมโยง ถ้ากำหนดค่าสนับสนุนน้อยที่สุดมีค่าเข้าใกล้หนึ่ง ทำให้หลักเกณฑ์บางหลักเกณฑ์ที่มีค่าความเชื่อมั่นสูงหายไป แต่ถ้าเรากำหนดค่าสนับสนุนน้อยที่สุดมีค่าเข้าใกล้ศูนย์ ทำให้ได้หลักเกณฑ์เชื่อมโยงออกมามากเกินความจำเป็น

ต่อมาได้มีการศึกษาตัววัดตัวใหม่ขึ้นโดย กรุง สินอภิรมย์สราญ และชรีย์วรรณ สิริศรีสัมฤทธิ์ ตัววัดดังกล่าวคือ ค่าสนับสนุนแบบอ่อน (weak support) [13] เป็นค่าความน่าจะเป็นที่ไม่เกิดเหตุการณ์ที่ขัดแย้งกับหลักเกณฑ์ คือพิจารณาทุกระเบียนหรือทรานแซคชันที่ไม่ขัดแย้งกับหลักเกณฑ์เทียบกับข้อมูลทั้งหมด

งานวิจัยดังกล่าวใช้ค่าสนับสนุนแบบอ่อนกับค่าความเชื่อมั่น มาเป็นตัววัดในการสร้างหลักเกณฑ์เชื่อมโยง และได้ทำการศึกษาหลักเกณฑ์ที่พจน์หลังมีสมาชิกเพียง 1 - ไอเทมเซตเท่านั้น ซึ่งค่าสนับสนุนแบบอ่อนนี้สามารถกำหนดค่าให้สูงขึ้นกว่าค่าสนับสนุนได้ แต่จะทำให้ได้หลักเกณฑ์ออกมามากกว่าหลักเกณฑ์ที่ได้จากการใช้ตัววัดค่าสนับสนุน

งานวิจัยของ กรุง สินอภิรมย์สราญ และอนันตพร ศรีสวัสดิ์ [14] ได้มีแนวความคิดเกี่ยวกับตัววัดของหลักเกณฑ์เชื่อมโยงใหม่ โดยพิจารณาประพจน์แย้งสลับที่ของหลักเกณฑ์ ซึ่งทำการศึกษาค่าความเชื่อมั่นประพจน์แย้งสลับที่เพื่อใช้ในการคัดเลือกหลักเกณฑ์เชื่อมโยงที่มีค่าความเชื่อมั่นที่เท่ากัน แล้วทำการคัดเลือกหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์

แย้งสลับที่ที่มีค่ามากที่สุด จากงานวิจัยของอนันตพรสามารถทำให้ปริมาณของหลักเกณฑ์เชื่อมโยงลดจำนวนลง

ในปัจจุบัน มีงานวิจัยเกี่ยวกับการลดหลักเกณฑ์เชื่อมโยงออกมาอย่างแพร่หลาย เช่น ขั้นตอนวิธีที่เป็นผลงานของ Bing Liu, Wynne Hsu และ Yiming ma [15] พิจารณาตัดหลักเกณฑ์ที่ไม่น่าสนใจแล้วทำการหาหลักเกณฑ์กำหนดทิศทาง (direction setting rule : DS rule) กับหลักเกณฑ์ที่ไม่ถูกตัดออกไป ซึ่งงานวิจัยนี้เป็นงานวิจัยที่ช่วยลดปริมาณหลักเกณฑ์เชื่อมโยง แต่ข้อมูลที่นำมาศึกษาเป็นข้อมูลแบบจำแนกประเภท (classification) คือข้อมูลมีลักษณะประจำเป้าหมาย (target class) ดังนั้นพจน์หลังของหลักเกณฑ์จะมีสมาชิก 1 - ไอเทมเซตเป็นลักษณะประจำเป้าหมาย

ในงานวิจัยเล่มนี้ ผู้วิจัยมีความสนใจศึกษาถึงขั้นตอนวิธีที่จะนำมาช่วยในการลดจำนวนหลักเกณฑ์จากการคัดเลือกหลักเกณฑ์เชื่อมโยง โดยใช้ค่าสนับสนุนแบบอ่อนกับค่าความเชื่อมั่น โดยนำตัววัดอีกตัวเข้ามาช่วยลดปริมาณหลักเกณฑ์ ได้แก่ ค่าความเชื่อมั่นประพจน์แย้งสลับที่ ซึ่งจะมีการกำหนดค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยที่สุดมาช่วยในการคัดเลือกหลักเกณฑ์ เราจะนำหลักเกณฑ์เชื่อมโยงที่เหลือมาทำการลดหลักเกณฑ์ต่อไปโดยใช้แนวคิดการหาหลักเกณฑ์กำหนดทิศทาง (direction setting rules หรือ DS rule) ซึ่งช่วยลดหลักเกณฑ์เชื่อมโยงที่ได้ออกมาจากขั้นตอนวิธีที่กล่าวมาข้างต้น งานวิจัยนี้จะทำการสร้างหลักเกณฑ์เชื่อมโยงที่พจน์หลังมีสมาชิกเพียง 1 - ไอเทมและเป็นลักษณะประจำเป้าหมาย ข้อมูลที่ศึกษาเป็นข้อมูลแบบจำแนกประเภท ในการทดลองจะทำการแบ่งข้อมูลออกเป็นสองส่วน ส่วนแรกใช้ในการสร้างหลักเกณฑ์เชื่อมโยงและในส่วนที่สองใช้ในการทดสอบหลักเกณฑ์ โดยใช้ค่าสภาพไว (sensitivity) มาประเมินความแตกต่างระหว่างกลุ่มของหลักเกณฑ์ที่ใช้ค่าสนับสนุนแบบอ่อนน้อยสุดกับค่าความเชื่อมั่นน้อยสุดเรียกขั้นตอนวิธีดับบลิวเอส และกลุ่มของหลักเกณฑ์ที่ใช้ค่าสนับสนุนแบบอ่อนน้อยสุด ค่าความเชื่อมั่นน้อยสุด กับค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุดเรียกขั้นตอนวิธีซีซีดับบลิวเอส และกลุ่มของหลักเกณฑ์ที่ใช้ค่าสนับสนุนแบบอ่อนน้อยสุด ค่าความเชื่อมั่นน้อยสุด กับค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุดกับแนวคิดการหาหลักเกณฑ์กำหนดทิศทางเรียกขั้นตอนวิธีดีเอสซีซีดับบลิวเอส

ในบทที่ 2 อธิบายถึงความรู้เบื้องต้นเกี่ยวกับการสร้างหลักเกณฑ์เชื่อมโยง ตัววัดที่นิยมใช้ และตัววัดอื่น ๆ ที่นำมาใช้ในงานวิจัยคือ ค่าสนับสนุนแบบอ่อน ค่าความเชื่อมั่นประพจน์แย้งสลับที่ แล้วอธิบายถึงขั้นตอนวิธีการหาหลักเกณฑ์กำหนดทิศทาง (DS rule) ที่นำมาช่วยลดหลักเกณฑ์เชื่อมโยง

ในบทที่ 3 อธิบายถึงที่มาของงานวิจัยนี้ อธิบายขั้นตอนวิธีที่ช่วยลดหลักเกณฑ์เชื่อมโยงที่ผ่านตัววัดค่าสนับสนุนแบบอ่อน กับค่าความเชื่อมั่น

ในบทที่ 4 อธิบายถึงวิธีการทดลอง ข้อมูลที่ใช้ในการทำการทดลอง ผลการทดลองที่นำมาเปรียบเทียบระหว่างหลักเกณฑ์เชื่อมโยงที่ผ่านค่าสนับสนุนแบบอ่อน กับค่าความเชื่อมั่น และหลักเกณฑ์ที่ผ่านค่าสนับสนุนแบบอ่อน ค่าความเชื่อมั่น ค่าความเชื่อมั่นประพจน์แย้งสลับที่กับขั้นตอนวิธีการหาหลักเกณฑ์กำหนดทิศทาง สุดท้ายจะพูดถึงตัววัดที่นำมาใช้ในการประเมินผลลัพธ์ของงานวิจัย

ในบทที่ 5 สรุปผลการทดลองของงานวิจัย ขั้นตอนวิธีที่ช่วยลดหลักเกณฑ์เชื่อมโยงที่ผ่านตัววัดค่าสนับสนุนแบบอ่อน กับค่าความเชื่อมั่น และงานวิจัยในอนาคต

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึง พื้นฐานของหลักเกณฑ์เชื่อมโยง การสร้างหลักเกณฑ์เชื่อมโยง ตัววัดที่นิยมใช้ และตัววัดที่นำมาใช้ในงานวิจัยนี้ได้แก่ ค่าสนับสนุนแบบอ่อน ค่าความเชื่อมั่นประพจน์แย้งสลับที่ และงานวิจัยที่นำมาช่วยลดหลักเกณฑ์เชื่อมโยง

การทำเหมืองข้อมูล เป็นกระบวนการค้นหาและวิเคราะห์ข้อมูลที่มีปริมาณมาก เพื่อให้ได้ความรู้ที่น่าสนใจ โดยความรู้ที่ค้นหานั้นอาจจะออกมาเป็น รูปแบบ แนวทาง หรือความสัมพันธ์ที่ซ่อนอยู่ในข้อมูล องค์กรต่าง ๆ สามารถนำการทำเหมืองข้อมูลไปประยุกต์ใช้กับสายงานขององค์กรได้ ซึ่งต้องทำการกำหนดเป้าหมายที่ชัดเจนว่าต้องการสิ่งใด แล้วนำเทคนิควิธีต่าง ๆ ของเหมืองข้อมูลไปใช้ให้เหมาะสม เพื่อให้ได้ผลลัพธ์ตามที่ต้องการ งานวิจัยนี้สนใจการสร้างหลักเกณฑ์เชื่อมโยง ซึ่งเป็นเทคนิคหนึ่งในการทำเหมืองข้อมูล

2.1 หลักเกณฑ์เชื่อมโยง

หลักเกณฑ์เชื่อมโยง (association rule) [2,3,4,5] เป็นหลักเกณฑ์ที่แสดงความสัมพันธ์ระหว่างกลุ่มของลักษณะประจำของข้อมูลที่เกิดขึ้นบ่อยและเกิดขึ้นพร้อมกัน โดยเขียนอยู่ในรูปแบบของหลักเกณฑ์ ถ้า – แล้ว ตัวอย่างเช่น $X \rightarrow Y$ ซึ่ง X และ Y เป็นเซตของกลุ่มลักษณะประจำของข้อมูลที่เรียกว่า ไอเทมเซต (itemset)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2.1.1 นิยาม

กำหนดให้ $I = \{I_1, \ldots, I_n\}$ เป็นเซตของไอเทม (item) หรือลักษณะประจำทั้งหมด ให้ $X \subseteq I$ เรียก X ว่า ไอเทมเซต (itemset) และเรียก k – ไอเทมเซต (k - itemset) เมื่อ X มีสมาชิกที่แตกต่างกันทั้งหมด k ตัว และให้ $D = \{t_1, \ldots, t_m\}$ เป็นเซตของระเบียนทั้งหมด โดยแต่ละระเบียน t_i เมื่อ $i \in \{1,2,\ldots,m\}$ เป็นไอเทมเซต กำหนดให้ T_X เป็นเซตของกลุ่มระเบียนที่สอดคล้องกับไอเทมเซต X ถ้า $X = \{I_1, I_2\}$ แล้วระเบียนที่สอดคล้องกับ X เขียน

แทนด้วย $T_{\{I_1,I_2\}}$ ให้ $T_{\{I_1\}}$ หมายถึงเซตของกลุ่มระเบียนที่พบการซื้อ I_1 และ $T_{\neg\{I_1\}}$ หมายถึงเซตของกลุ่มระเบียนที่ไม่พบการซื้อ I_1

หลักเกณฑ์เชื่อมโยงจะเขียนอยู่ในรูป $X \rightarrow Y$ เมื่อ $X, Y \subseteq I$ และ $X \cap Y = \varnothing$ โดย X เป็นไอเทมเซตของพจน์หน้า (antecedent) หรือพจน์ด้านซ้าย (LHS) และ Y เป็นไอเทมเซตของพจน์หลัง (consequent) หรือพจน์ด้านขวา (RHS) เนื่องจากหลักเกณฑ์เชื่อมโยงในรูปแบบ $X \rightarrow \{I_1, I_2, \ldots, I_n\}$ มีความหมายทางตรรกศาสตร์เหมือนกับ $X \rightarrow \{I_1\}$ และ $X \rightarrow \{I_2\}$ ไปเรื่อย ๆ จนถึง $X \rightarrow \{I_n\}$ ซึ่งงานวิจัยนี้จะสนใจเฉพาะไอเทมเซตในพจน์หลังที่มีสมาชิกเพียงหนึ่งตัว

ในปริภูมิของลักษณะประจำ หลักเกณฑ์ $X \rightarrow Y$ หมายถึงจะพบ X และ Y อยู่ในระเบียนเดียวกันเขียนแทนด้วย $T_{X \cup Y}$ การอ้างถึงนิเสธของเซต X, Y จะเชื่อมโยงกับระเบียนทั้งหมดใน D ที่ไม่มีการปรากฏของ X, Y กล่าวคือนิเสธของหลักเกณฑ์ $X \rightarrow Y$ และ $X \cap Y = \varnothing$ เขียนแทนด้วย $T_{X \cup \neg Y}$ คือเซตของระเบียนทั้งหมดใน D ที่พบ X แต่ไม่พบ Y ในทำนองเดียวกัน ประพจน์แย้งสลับที่ของหลักเกณฑ์ $X \rightarrow Y$ เขียนแทนด้วย $T_{\neg Y \cup \neg X}$ คือเซตของระเบียนทั้งหมดใน D ที่ไม่พบ X และ Y

2.1.2 นิยามตัววัด

การสร้างหลักเกณฑ์เชื่อมโยง $X \rightarrow Y$ จะเป็นการดูความสัมพันธ์ของพจน์หน้าและพจน์หลังของหลักเกณฑ์ จึงต้องมีการนำตัววัดมาช่วยในการพิจารณาความน่าเชื่อถือและความน่าสนใจของหลักเกณฑ์เชื่อมโยง ตัววัดที่นิยมใช้กันได้แก่ ค่าสนับสนุน (support) และค่าความเชื่อมั่น (confidence) ถูกนำเสนอโดย Argawal, Imielinski และ Swami ในปี 1993 [5]

ค่าสนับสนุน (support) คือค่าที่แสดงความถี่ของหลักเกณฑ์เทียบกับข้อมูลทั้งหมดหรือความน่าจะเป็นที่จะปรากฏหลักเกณฑ์ ซึ่งมีค่าตั้งแต่ 0 ไปถึง 1 เมื่อมีค่าสนับสนุนเป็น 0 หมายความว่า ไม่มีระเบียนที่สอดคล้องกับหลักเกณฑ์นั้น และเมื่อมีค่าสนับสนุนเป็น 1 หมายความว่าทุก ๆ ระเบียนจะสอดคล้องกับหลักเกณฑ์นั้น

$$\sup(X \rightarrow Y) = \sup(X \cup Y) = \frac{|T_{X \cup Y}|}{|D|}$$

ค่าความเชื่อมั่น (confidence) คือค่าที่แสดงการเกิดของหลักเกณฑ์ เทียบกับการเกิดเฉพาะพจน์หน้าของหลักเกณฑ์ ซึ่งมีค่าตั้งแต่ 0 ไปถึง 1 เมื่อหลักเกณฑ์มีค่าความเชื่อมั่นเป็น 0 หมายความว่า ไม่มีระเบียนที่สอดคล้องกับหลักเกณฑ์นั้น และเมื่อหลักเกณฑ์มีค่าความเชื่อมั่นเป็น 1 หมายความว่า ทุก ๆ ระเบียนที่พบพจน์หน้าจะพบพจน์หลังด้วย โดยมีเงื่อนไขว่าจำนวนระเบียนที่พบพจน์หน้าจะไม่เป็น 0

$$conf(X \rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)} = \frac{|T_{X \cup Y}|}{|T_X|}$$

หลักเกณฑ์เชื่อมโยง เป็นการบ่งบอกถึงความสัมพันธ์ระหว่างกลุ่มของลักษณะประจำของข้อมูล ดังนั้นในการสร้างหลักเกณฑ์เชื่อมโยงจะต้องระบุพจน์หน้าและพจน์หลังของหลักเกณฑ์ก่อน แล้วหาค่าของตัววัด โดยหลักเกณฑ์เชื่อมโยงที่ได้จะมีความน่าสนใจหรือไม่ขึ้นอยู่กับการกำหนดค่าน้อยสุดที่ยอมรับได้ของผู้ใช้

2.1.3 การสร้างหลักเกณฑ์เชื่อมโยง

หลักเกณฑ์เชื่อมโยงถูกสร้างขึ้นครั้งแรกโดย Argawal, Imielinski และ Swami ในปี 1993 [5] ซึ่งในปัจจุบันจะมีวิธีการสร้างหลักเกณฑ์เชื่อมโยงหลายวิธี แล้วแต่ผู้ใช้จะเลือกใช้งานตามความเหมาะสมของข้อมูลหรือแล้วแต่ความต้องการของผู้ใช้ โดยทั่วไปขั้นตอนวิธีการสร้างหลักเกณฑ์จะแบ่งออกเป็น 2 ขั้นตอนหลัก คือ

1. ขั้นตอนการสร้างไอเทมเซตปรากฏบ่อย (frequent itemset) ขั้นตอนนี้จะทำการสร้างไอเทมเซตปรากฏบ่อยออกมาจากข้อมูลทั้งหมด โดยที่ไอเทมเซตปรากฏบ่อย จะต้องมีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนน้อยสุดที่ยอมรับได้ ตามที่ผู้ใช้กำหนด
2. ขั้นตอนการสร้างหลักเกณฑ์ ขั้นตอนนี้จะนำไอเทมเซตปรากฏบ่อยจากขั้นตอนแรกมาทำการสร้างเป็นหลักเกณฑ์เชื่อมโยง โดยอาศัยตัววัดอื่น เช่น ความเชื่อมั่น มาเป็นตัวตัดสินความน่าเชื่อถือของหลักเกณฑ์

ขั้นตอนวิธีที่ใช้วิธีการดังกล่าว จะใช้ตัววัดสองตัว คือ ค่าสนับสนุน และค่าความเชื่อมั่น เรียกขั้นตอนวิธีนี้ว่า ขั้นตอนวิธีเอไพออรี (Apriori algorithm)

เนื่องจากตัววัดมีความสำคัญในการสร้างหลักเกณฑ์เชื่อมโยง ทั้งในส่วนของการหาไอเทมเซตปรากฏบ่อยที่ใช้ค่าสนับสนุนเป็นตัวคัดไอเทมเซตที่ไม่สำคัญออก แต่ผู้ใช้ซึ่งเป็นผู้กำหนดค่าสนับสนุนน้อยสุดก็ควรจะเป็นผู้ที่มีความเข้าใจในข้อมูลอย่างดี เพื่อที่จะกำหนดค่าสนับสนุนน้อยสุดได้อย่างเหมาะสม เพราะถ้ากำหนดมากเกินไปจะทำให้ตัดไอเทมเซตที่จะสร้างหลักเกณฑ์ที่มีค่าความเชื่อมั่นสูงออกไป หรือถ้ากำหนดน้อยเกินไปจะทำให้ได้ไอเทมเซตออกมามาก

ในปัจจุบันมีการพัฒนาการสร้างหลักเกณฑ์เชื่อมโยง เพื่อให้ได้หลักเกณฑ์ที่มีประสิทธิภาพเพิ่มมากขึ้น เช่นการพัฒนาตัววัดใหม่ขึ้นมาช่วยในการคัดเลือกหลักเกณฑ์ การพัฒนาการสร้างหลักเกณฑ์เชื่อมโยงให้มีปริมาณหลักเกณฑ์ที่ลดลง การสร้างขั้นตอนวิธีที่ช่วยลดเวลาในการทำงาน เป็นต้น งานวิจัยนี้จึงสนใจตัววัดความน่าสนใจสองตัวคือ ค่าสนับสนุนแบบอ่อน (weak support) [13] และค่าความเชื่อมั่นประพจน์แย้งสลับที่ (contrapositive confidence) [14]

2.2 ค่าสนับสนุนแบบอ่อน

ค่าสนับสนุนแบบอ่อน (weak support) [13] เป็นตัววัดตัวใหม่ที่ กรุง และ ชรีย์ วรรณ ศึกษาและวิจัย เป็นตัววัดตัวใหม่ที่สร้างขึ้นจากแนวคิดทางตรรกะซึ่งสามารถอธิบายได้ดังต่อไปนี้

พิจารณาประพจน์ ถ้า-แล้ว ทางตรรกศาสตร์ $p \rightarrow q$ พบว่านิเสธหรือข้อขัดแย้งของประพจน์ คือ $p \wedge \neg q$ เมื่อ p และ q คือประพจน์ของหลักเกณฑ์ และพิจารณาหลักเกณฑ์เชื่อมโยง $X \rightarrow Y$ เมื่อ X และ Y แทนเซตของกลุ่มลักษณะประจำของข้อมูล พบว่าสามารถแบ่งข้อมูลออกเป็น 4 กลุ่มดังนี้คือ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

TID	X	Y
t_i ; $\exists i \in \{1,...,n \mid t_i \subseteq \neg X \cup \neg Y\}$	0	0
t_i ; $\exists i \in \{1,...,n \mid t_i \subseteq \neg X \cup Y\}$	0	1
t_i ; $\exists i \in \{1,...,n \mid t_i \subseteq X \cup \neg Y\}$	1	0
t_i ; $\exists i \in \{1,...,n \mid t_i \subseteq X \cup Y\}$	1	1

ตารางที่ 2.1 ข้อมูลที่แบ่งออกเป็นสี่กลุ่มตามตรรกะของ X, Y

จากตารางที่ 2.1 พบว่าข้อมูลแบ่งออกเป็นกลุ่มตามตารางค่าความจริง ซึ่งกลุ่มของข้อมูลที่ขัดแย้งกับหลักเกณฑ์เชื่อมโยง $X \to Y$ คือกรณีที่ X เป็น 1 (คือการที่พบไอเทมเซต X ในระเบียนที่สนใจ) และ Y เป็น 0 (คือการที่ไม่พบไอเทมเซต Y ในระเบียนที่สนใจ) ซึ่งก็คือกลุ่มของระเบียนที่สอดคล้องกับไอเทมเซต $X \cup \neg Y$ ในปริภูมิลักษณะประจำหรือ $T_X \cap T_{\neg Y}$ ในปริภูมิระเบียน ดังนั้นจึงนิยามค่าสนับสนุนแบบอ่อน ให้เป็นตัววัดที่ไม่พิจารณาระเบียนหรือไอเทมเซตที่ขัดแย้งกับหลักเกณฑ์ โดยสามารถคำนวณได้จาก สัดส่วนระหว่างจำนวนระเบียนที่ไม่ขัดแย้งกับหลักเกณฑ์เชื่อมโยงต่อระเบียนทั้งหมด ซึ่งเขียนได้ดังนี้

$$\text{weaksup}(X \to Y) = \frac{|T_{X \cup Y}| + |T_{\neg X \cup Y}| + |T_{\neg X \cup \neg Y}|}{|D|}$$

หรือ $$\text{weaksup}(X \to Y) = 1 - \frac{|T_{X \cup \neg Y}|}{|D|}$$

2.2.1 ความสัมพันธ์ระหว่างค่าสนับสนุนแบบอ่อน ค่าสนับสนุน และค่าความเชื่อมั่น

ค่าสนับสนุนแบบอ่อนเขียนในรูปของความสัมพันธ์ระหว่างค่าสนับสนุนและค่าความเชื่อมั่นได้ดังสมการด้านล่าง [13]

$$\text{weaksup}(X \to Y) = \text{sup}(X \to Y) + 1 - \frac{\text{sup}(X \to Y)}{\text{conf}(X \to Y)}$$

ดังนั้นถ้าผู้ใช้ทราบค่าสนับสนุน และค่าความเชื่อมั่น ผู้ใช้สามารถคำนวณหาค่าสนับสนุนแบบอ่อนดังตารางที่ 2.2 [13]

conf \ *sup*	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.9	0.99	0.98	0.97	0.96	0.94	0.93	0.92	0.91	0.90
0.8	0.98	0.95	0.93	0.90	0.88	0.85	0.83	0.80	0.78
0.7	0.96	0.91	0.87	0.83	0.79	0.74	0.70	0.66	0.61
0.6	0.93	0.87	0.80	0.73	0.67	0.60	0.53	0.47	0.40
0.5	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10

ตารางที่ 2.2 ค่าสนับสนุนแบบอ่อนจากการคำนวณค่าสนับสนุนและค่าความเชื่อมั่น

จากสมการข้างต้นและตารางที่ 2.2 เป็นความสัมพันธ์ที่ได้จากงานวิจัยเรื่องค่าสนับสนุนแบบอ่อน [13] ซึ่งค่าสนับสนุนแบบอ่อนเป็นค่าความน่าจะเป็นที่ไม่เกิดเหตุการณ์ที่ขัดแย้งกับหลักเกณฑ์ โดยจะเห็นว่าค่าสนับสนุนแบบอ่อนนี้จะถูกกำหนดค่าให้สูงกว่าค่าสนับสนุนเสมอ และจากตารางที่ 2.2 การคำนวณค่าสนับสนุนแบบอ่อนจากค่าสนับสนุนและค่าความเชื่อมั่นจะเห็นว่าไม่ควรกำหนดค่าสนับสนุนให้มีค่ามากกว่าค่าความเชื่อมั่น ดังที่แรเงาไว้ในตาราง 2.2 เนื่องจากถ้าพิจารณาความสัมพันธ์ระหว่างค่าสนับสนุนกับค่าความเชื่อมั่นของหลักเกณฑ์ จะพบว่าค่าสนับสนุนของหลักเกณฑ์มีค่าน้อยกว่าเท่ากับค่าความเชื่อมั่นของหลักเกณฑ์ ฉะนั้นจึงไม่ควรกำหนดค่าสนับสนุนให้มากกว่าค่าความเชื่อมั่น เพราะจะทำให้ได้ค่าสนับสนุนแบบอ่อนที่ไม่สมเหตุสมผล ซึ่งไม่มีประโยชน์ต่อการนำไปใช้

2.2.2 ขั้นตอนวิธีค่าสนับสนุนแบบอ่อน

ขั้นตอนวิธีค่าสนับสนุนแบบอ่อนหรือขั้นตอนวิธี WS (weak support algorithm) จะคำนวณจากข้อมูลที่ขัดแย้งกับหลักเกณฑ์ ซึ่งจะพิจารณาเฉพาะหลักเกณฑ์ที่พจน์หลังเป็น 1 – ไอเทมเซต โดยกำหนดให้ x แทนพจน์หน้าของหลักเกณฑ์และ $x \in I$ และให้ y แทนพจน์หลังของหลักเกณฑ์โดยที่ $y \in I$ ซึ่งพจน์หลังจะต้องมีสมาชิกเพียง 1 ตัว กำหนดให้ *minWS* คือ ค่าสนับสนุนแบบอ่อนน้อยสุดที่ผู้ใช้กำหนด *minC* คือ ค่าความเชื่อมั่นน้อยสุดที่ผู้ใช้กำหนด *LHS* คือ เซตของพจน์หน้าของหลักเกณฑ์ที่มีค่าสนับสนุนแบบอ่อนน้อยสุดมากกว่าหรือเท่ากับ *minWS* และมีค่าความเชื่อมั่นที่คำนวณได้มากกว่าหรือเท่ากับ *minC* และ *CandidateLHS* คือ เซตของพจน์หน้าตัวเลือก

ขั้นตอนวิธี WS แสดงดังรูปที่ 2.1 และอธิบายได้ดังต่อไปนี้

```
1.  for each item y ∈ I
2.      LHS = { }
3.      for each item x ∈ I and
4.          if {x}→{y} pass minWS and minC
5.              LHS = LHS ∪ {x}
6.      for(k = 2; LHS ≠ {}; k++)
7.          CandidateLHS = Generate(LHS, k)
8.          LHS = {}
9.          for each x ∈ CandidateLHS
10.             if {x}→{y} pass minWS and minC
11.                 LHS = LHS ∪ {x}
```

รูปที่ 2.1: ขั้นตอนวิธี WS

- พิจารณาแต่ละพจน์หลังของหลักเกณฑ์ ซึ่งกำหนดให้พจน์หลังที่สนใจต้องเป็น 1–ไอเทมเซต เนื่องจาก *LHS* เป็นเซตของพจน์หน้าของหลักเกณฑ์สำหรับแต่ละพจน์หลังของ y ดังนั้นจึงต้องกำหนดให้มีค่าเริ่มต้นเป็นเซตว่าง
- ในบรรทัดที่ 3-5 สำหรับแต่ละพจน์หน้าของหลักเกณฑ์ นำมาสร้างเป็นหลักเกณฑ์เชื่อมโยง ถ้าหลักเกณฑ์ที่สร้างมีค่าสนับสนุนแบบอ่อนมากกว่าหรือเท่ากับ *minWS* และมีค่าความเชื่อมั่นมากกว่าหรือเท่ากับ *minC* จะกำหนดให้ $LHS = LHS \cup \{x\}$ ในขั้นตอนนี้จะได้ *LHS* เป็นเซตของพจน์หน้าที่เล็กที่สุด ที่ทำให้เกิดหลักเกณฑ์ที่ผ่านค่าสนับสนุนแบบอ่อนน้อยสุดที่ยอมรับได้และผ่านค่าความเชื่อมั่นน้อยสุดที่ยอมรับได้ ซึ่งเมื่อนำไปสร้างหลักเกณฑ์ที่มีเซตของพจน์หน้าให้มีขนาดใหญ่ขึ้นจากเซตเดิมที่เก็บใน *LHS* จะผ่านค่าสนับสนุนแบบอ่อนน้อยสุดที่ยอมรับได้

- ในบรรทัดที่ 6-11 แต่ละรอบของการทำงานที่ k จะสร้างเซตของพจน์หน้าตัวเลือก *CandidateLHS* ที่มีขนาด k จาก *LHS* แล้วจึงกำหนดให้ $LHS = \{\}$ เพื่อให้ *LHS* เป็นเซตของพจน์หน้าที่มีขนาด k ที่สร้างหลักเกณฑ์ที่มีค่าสนับสนุนแบบอ่อนมากกว่าหรือเท่ากับ *minWS* และมีค่าความเชื่อมั่นมากกว่าหรือเท่ากับ *minC* โดยกำหนดให้ค่า k เริ่มต้นมีค่าเป็น 2

- ในบรรทัดที่ 9-11 พิจารณาแต่ละพจน์หน้าที่อยู่ใน *CandidateLHS* มาทดสอบเพื่อสร้าง *LHS* โดยที่ $LHS = LHS \cup \{x\}$ เมื่อ $x \in CandidateLHS$ และหลักเกณฑ์ $\{x\} \rightarrow \{y\}$ เป็นหลักเกณฑ์ที่มีค่าสนับสนุนแบบอ่อนมากกว่าหรือเท่ากับ *minWS* และมีค่าความเชื่อมั่นมากกว่าหรือเท่ากับ *minC* แล้วกำหนดให้ขนาดของพจน์หน้ามีค่าเพิ่มขึ้นเป็น $k = k+1$

การค้นหาหลักเกณฑ์โดยใช้ค่าสนับสนุนแบบอ่อน ผู้ใช้ทั่วไปสามารถกำหนดค่าสนับสนุนแบบอ่อนน้อยสุดได้เอง โดยให้ค่าที่มากกว่า 0.5 เนื่องจากค่าสนับสนุนแบบอ่อนเป็นตัววัดที่มีแนวคิดที่สนใจข้อมูลที่ไม่ขัดแย้งกับหลักเกณฑ์ และหลักเกณฑ์ที่ผ่านการคัดเลือกควรมีความน่าเชื่อถือ ดังนั้นความน่าจะเป็นของการหลักเกณฑ์ควรจะมีค่ามากกว่า 0.5 ในขณะที่การใช้ค่าสนับสนุนในการคัดเลือกหลักเกณฑ์ ไม่สามารถแนะนำได้ว่าควรกำหนดค่าสนับสนุนน้อยสุดให้มีค่าเท่าไรซึ่งจะขึ้นอยู่กับผู้ใช้ที่ต้องมีความเข้าใจในข้อมูล แต่การใช้ค่าสนับสนุนแบบอ่อนในการคัดเลือกหลักเกณฑ์มักจะทำให้ได้หลักเกณฑ์ออกมามากกว่าหลักเกณฑ์ที่ได้จากการใช้ตัววัดค่าสนับสนุนและค่าความเชื่อมั่น

2.3 ค่าความเชื่อมั่นประพจน์แย้งสลับที่

ค่าความเชื่อมั่นประพจน์แย้งสลับที่ (contrapositive confidence) [14] เป็นตัววัดที่กรุง และ อนันตพร ศึกษาและวิจัย โดยใช้แนวคิดทางตรรกศาสตร์และเซตมาประยุกต์สามารถอธิบายได้ดังนี้

พิจารณาประพจน์ถ้า-แล้วทางตรรกศาสตร์ $p \rightarrow q$ พบว่าประพจน์แย้งสลับที่ของประพจน์ คือ $\neg q \rightarrow \neg p$ ซึ่งทั้งสองประพจน์ให้ค่าความจริงเหมือนกันดังตารางที่ 2.3 และค่าความเชื่อมั่นของหลักเกณฑ์ $\neg q \rightarrow \neg p$ มีค่าเหมือนกับความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q$ เมื่อ p และ q คือประพจน์ที่กำหนดค่าความจริงว่าเป็นจริงหรือเท็จเท่านั้น

p	$\neg p$	q	$\neg q$	$p \rightarrow q$	$\neg q \rightarrow \neg p$
T	F	T	F	T	T
T	F	F	T	F	F
F	T	T	F	T	T
F	T	F	T	T	T

ตารางที่ 2.3 ตารางค่าความจริงของ $p \rightarrow q$ และ $\neg q \rightarrow \neg p$

การคำนวณค่าความเชื่อมั่นประพจน์แย้งสลับที่ จะคำนวณจากค่าความเชื่อมั่นของพจน์ที่เป็นประพจน์แย้งสลับที่ พิจารณาหลักเกณฑ์เชื่อมโยง $X \rightarrow Y$ โดยที่ X และ Y แทนเซตลักษณะประจำของข้อมูล ค่าความเชื่อมั่นประพจน์แย้งสลับที่หาได้ดังนี้

$$conf(\neg Y \rightarrow \neg X) = \frac{\sup(\neg Y \rightarrow \neg X)}{\sup(\neg Y)}$$

$$= \frac{\frac{|T_{\neg Y \rightarrow \neg X}|}{|D|}}{\frac{|T_{\neg Y}|}{|D|}} = \frac{|T_{\neg Y \rightarrow \neg X}|}{|T_{\neg Y}|}$$

$$= \frac{|T_{\neg Y \cup \neg X}|}{|T_{\neg Y}|}$$

2.3.1 ความสัมพันธ์ระหว่างค่าความเชื่อมั่นประพจน์แย้งสลับที่กับค่าความเชื่อมั่น

จากสมการค่าความเชื่อมั่นประพจน์แย้งสลับที่ กับสมการค่าความเชื่อมั่นของหลักเกณฑ์ $X \rightarrow Y$ ซึ่งประพจน์แย้งสลับที่ของหลักเกณฑ์ได้แก่ $\neg Y \rightarrow \neg X$

พิจารณาค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $X \rightarrow Y$ คือ

$$conf(\neg Y \rightarrow \neg X) = \frac{\sup(\neg Y \rightarrow \neg X)}{\sup(\neg Y)}$$

$$= \frac{|T_{\neg Y \cup \neg X}|}{|T_{\neg Y}|}$$

เนื่องจาก $T_{\neg Y} = T_{\neg Y \cup X} + T_{\neg Y \cup \neg X}$

ดังนั้น

$$conf(\neg Y \rightarrow \neg X) = \frac{|T_{\neg Y \cup \neg X}|}{|T_{\neg Y}|}$$

$$= \frac{|T_{\neg Y \cup \neg X}|}{|T_{\neg Y \cup \neg X}| + |T_{\neg Y \cup X}|}$$

$$= \frac{\frac{|T_{\neg Y \cup \neg X}|}{|T_{\neg Y \cup \neg X}|}}{\frac{|T_{\neg Y \cup \neg X}|}{|T_{\neg Y \cup \neg X}|} + \frac{|T_{\neg Y \cup X}|}{|T_{\neg Y \cup \neg X}|}} = \frac{1}{1 + \frac{|T_{\neg Y \cup X}|}{|T_{\neg Y \cup \neg X}|}}$$

พิจารณาค่าความเชื่อมั่นของหลักเกณฑ์ $X \rightarrow Y$ คือ

$$conf(X \rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)}$$

$$= \frac{|T_{X \cup Y}|}{|T_X|}$$

เนื่องจาก $T_X = T_{X \cup Y} + T_{X \cup \neg Y}$

ดังนั้น

$$conf(X \rightarrow Y) = \frac{|T_{X \cup Y}|}{|T_X|}$$

$$= \frac{|T_{X \cup Y}|}{|T_{X \cup Y}| + |T_{X \cup \neg Y}|}$$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$$= \frac{\frac{|T_{X\cup Y}|}{|T_{X\cup Y}|}}{\frac{|T_{X\cup Y}|}{|T_{X\cup Y}|} + \frac{|T_{X\cup\neg Y}|}{|T_{X\cup Y}|}} = \frac{1}{1+\frac{|T_{X\cup\neg Y}|}{|T_{X\cup Y}|}}$$

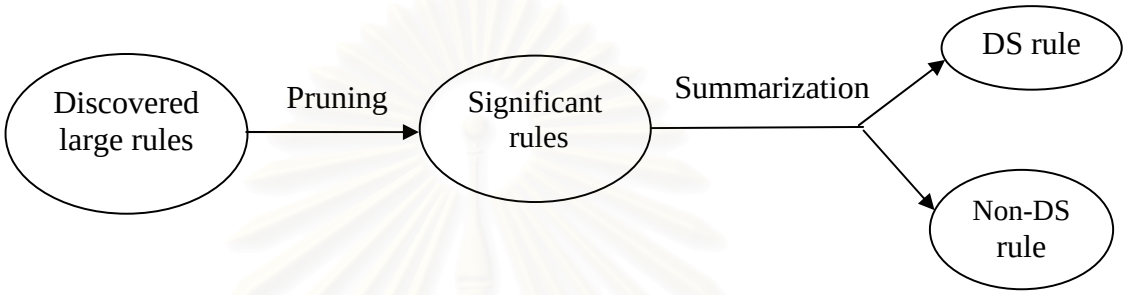
จะเห็นว่าจากสมการในการคำนวณตัวแปรที่แตกต่างกันคือ $T_{X\cup Y}$ กับ $T_{\neg X\cup\neg Y}$ ซึ่งสามารถสรุปความสัมพันธ์ระหว่างค่าความเชื่อมั่นทั้งสองได้ดังนี้

- กรณีที่ $T_{X\cup Y} = T_{\neg X\cup\neg Y}$ คือ จำนวนข้อมูลที่พบการซื้อ X และ Y ในระเบียนเดียวกันเท่ากับจำนวนข้อมูลที่ไม่พบการซื้อ X หรือ Y ในระเบียนเดียวกัน จะได้ว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่มีค่าเท่ากันกับค่าความเชื่อมั่นของหลักเกณฑ์
- กรณีที่ $T_{X\cup Y} > T_{\neg X\cup\neg Y}$ คือ จำนวนข้อมูลที่พบการซื้อ X และ Y ในระเบียนเดียวกันมีมากกว่าจำนวนข้อมูลที่ไม่พบการซื้อ X หรือ Y ในระเบียนเดียวกัน จะได้ว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่มีค่าน้อยกว่าค่าความเชื่อมั่นของหลักเกณฑ์
- กรณีที่ $T_{X\cup Y} < T_{\neg X\cup\neg Y}$ คือ จำนวนข้อมูลที่พบการซื้อ X และ Y ในระเบียนเดียวกันมีน้อยกว่าจำนวนข้อมูลที่ไม่พบการซื้อ X หรือ Y ในระเบียนเดียวกัน จะได้ว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่มีค่ามากกว่าค่าความเชื่อมั่นของหลักเกณฑ์ ซึ่งในกรณีนี้จะพบได้บ่อยในปัญหาจริง เช่นในปัญหาการวิเคราะห์ข้อมูลการขายสินค้า แต่ละร้านค้าจะมีสินค้าหลายชนิด ดังนั้นการที่จะไม่พบการซื้อสินค้าที่เราสนใจจะมีปริมาณมากกว่าการพบการซื้อสินค้าที่สนใจ

ในงานวิจัยนี้ผู้วิจัยได้นำตัววัดค่าความเชื่อมั่นประพจน์แย้งสลับที่ มาช่วยในการลดปริมาณหลักเกณฑ์เชื่อมโยง โดยกำหนดค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุดที่ยอมรับได้ให้มีค่าเท่ากับค่าความเชื่อมั่นน้อยสุดที่ยอมรับได้ของหลักเกณฑ์เชื่อมโยง

2.4 ขั้นตอนวิธีการกำหนดทิศทาง

ขั้นตอนวิธีการกำหนดทิศทางเป็นขั้นตอนวิธีที่ช่วยลดหลักเกณฑ์เชื่อมโยง โดย Bing Liu, Wynne Hsu และ Yiming Ma [15] เริ่มด้วยการพิจารณาคัดหลักเกณฑ์ที่ไม่น่าสนใจออกไป แล้วคัดเลือกหลักเกณฑ์ที่ผ่านขั้นตอนวิธีการกำหนดทิศทาง หลักเกณฑ์ที่ถูกคัดเลือกนั้นเรียกว่า หลักเกณฑ์กำหนดทิศทาง (Direction setting rule : DS rule) ซึ่งมีขั้นตอนการทำงานดังรูป 2.2



รูปที่ 2.2 ขั้นตอนการทำงานของขั้นตอนวิธีการกำหนดทิศทาง

การหาหลักเกณฑ์กำหนดทิศทางจะใช้ตัววัดทางสถิติที่สำคัญคือ ค่าไคกำลังสอง (chi-square : χ^2) ซึ่งจะนำมาทดสอบความเป็นอิสระต่อกันหรือพิจารณาความสัมพันธ์ของไอเทมเซตของหลักเกณฑ์ คือนำมาทดสอบความสัมพันธ์ของพจน์หน้ากับพจน์หลังของหลักเกณฑ์เชื่อมโยง เริ่มแรกในการทดสอบความสัมพันธ์จะต้องมีการกำหนดสมมติฐาน (Hypothesis) โดยกำหนดให้

H_0: พจน์หน้าและพจน์หลังของหลักเกณฑ์เป็นอิสระต่อกัน

H_1: พจน์หน้าและพจน์หลังของหลักเกณฑ์ไม่เป็นอิสระต่อกัน

ถ้าผลทดสอบออกมาแล้วสมมติฐานเป็นจริงคือ H_0 เป็นจริง ดังนั้นพจน์หน้าและพจน์หลังของหลักเกณฑ์เป็นอิสระต่อกัน แต่ถ้าทดสอบออกมาแล้วสมมติฐานเป็นเท็จคือ H_0 เป็นเท็จ ดังนั้นพจน์หน้าและพจน์หลังของหลักเกณฑ์ไม่เป็นอิสระต่อกันตามสมมติฐานของ H_1

ซึ่งในงานวิจัยนี้จะมีการคำนวณค่าไคกำลังสอง จากข้อมูลของตารางการจรณ์ (contingency table) ที่มีขนาด 2×2 ดังตารางที่ 2.4 เมื่อหลักเกณฑ์คือ $X \rightarrow Y$

	พบ Y	ไม่พบ Y	
พบ X	**ข้อมูลที่พบ X และ Y**	**ข้อมูลที่พบ X แต่ไม่พบ Y**	จำนวนข้อมูลที่พบ X
ไม่พบ X	**ข้อมูลที่ไม่พบ X แต่พบ Y**	**ข้อมูลที่ไม่พบ X และ Y**	จำนวนข้อมูลที่ไม่พบ X
	จำนวนข้อมูลที่พบ Y	จำนวนข้อมูลที่ไม่พบ Y	จำนวนข้อมูลทั้งหมด

ตารางที่ 2.4 ตารางการจรณ์ (contingency table) ที่มีขนาด 2×2

จากตารางที่ 2.4 ตารางการจรณ์ที่มีขนาด 2×2 นั้นจะเป็นข้อมูลของแต่ละหลักเกณฑ์เชื่อมโยง ขนาดของตารางเป็น 2×2 เพื่อแสดงข้อมูลพจน์หน้าของหลักเกณฑ์ทั้งการพบและไม่พบข้อมูล และแสดงข้อมูลในส่วนพจน์หลังของหลักเกณฑ์ทั้งการพบและไม่พบ กำหนดให้ $i=\{1,2\}$ แทนตำแหน่งของข้อมูลในตารางการจรณ์ทางด้านระเบียนของตาราง โดย i = 1 แทนตำแหน่งของข้อมูลการพบ X ส่วน i = 2 แทนตำแหน่งของข้อมูลการไม่พบ X และ $j=\{1,2\}$ แทนตำแหน่งของข้อมูลในตารางการจรณ์ทางด้านลักษณะประจำของตาราง โดย j=1 แทนตำแหน่งของข้อมูลการพบ Y ส่วน j = 2 แทนตำแหน่งของข้อมูลการไม่พบ Y

ในการคำนวณค่าไคกำลังสองสามารถคำนวณได้ดังนี้

$$\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}$$

โดยที่

$$E_{ij}=\frac{r_i c_j}{N}$$

กำหนดให้

O_{ij} แทน ความถี่ที่ได้จากการสังเกต (ข้อมูลจริง) ในตำแหน่งที่ i และตำแหน่งที่ j

E_{ij} แทน ความถี่ที่คาดหวัง (ข้อมูลสมมติ) ในตำแหน่งที่ i และตำแหน่งที่ j

r_i แทน ความถี่ (ข้อมูลจริง) รวมในตำแหน่งที่ i หรือแทนจำนวนข้อมูลที่พบ X เมื่อ i=1 และแทนจำนวนข้อมูลที่ไม่พบ X เมื่อ i=2

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

c_j แทน ความถี่ (ข้อมูลจริง) รวมในลักษณะประจำที่ j หรือแทนจำนวนข้อมูลที่พบ Y เมื่อ j=1 และแทนจำนวนข้อมูลที่ไม่พบ Y เมื่อ j=2

N แทน จำนวนข้อมูลจริงทั้งหมด

การคำนวณค่าไคกำลังสอง จะต้องมีการคำนวณค่า O_{ij} และ E_{ij} ซึ่งค่า O_{11} แทนจำนวนข้อมูลที่พบทั้ง X และ Y ค่า O_{12} แทนจำนวนข้อมูลที่พบ X แต่ไม่พบ Y ค่า O_{21} แทนจำนวนข้อมูลที่ไม่พบ X แต่พบ Y ค่า O_{22} แทนจำนวนข้อมูลที่ไม่พบทั้ง X และ Y ในการคำนวณค่า E_{ij} นั้นคำนวณจากสมการข้างต้น ซึ่งค่า E_{11} คำนวณจาก r_1 ที่แทนจำนวนข้อมูลที่พบ X และ c_1 แทนจำนวนข้อมูลที่พบ Y แล้วคำนวณค่า E_{12}, E_{21}, E_{22} ในทำนองเดียวกัน

ตัวอย่างการคำนวณค่าไคกำลังสอง จากข้อมูลการขายสินค้าปลีกของตารางที่ 2.5 โดยสนใจหลักเกณฑ์เชื่อมโยง ถ้ามีการซื้อขนมปังแล้วจะมีการซื้อนม (ขนมปัง → นม [sup=0.4, conf=0.67])

กำหนดสมมติฐาน (Hypothesis) โดยกำหนดให้

H_0: ขนมปังและนมไม่มีความสัมพันธ์กันหรือมีการซื้อขนมปังแล้วไม่จำเป็นต้องมีการซื้อนม

H_1: ขนมปังและนมมีความสัมพันธ์กันหรือมีการซื้อขนมปังแล้วอาจมีการซื้อนมด้วย

	พบการซื้อนม	ไม่พบการซื้อนม	รวม
พบการซื้อขนมปัง	200	100	300
ไม่พบการซื้อขนมปัง	80	120	200
รวม	280	220	500

ตารางที่ 2.5 ข้อมูลการขายสินค้าในร้านค้าปลีก

ค่าไคกำลังสองของหลักเกณฑ์ ขนมปัง → นม คำนวณได้ดังนี้

เริ่มจากหาค่า $E_{11} = (\frac{300 \times 280}{500}) = 168$

$$E_{12} = (\frac{300 \times 220}{500}) = 132$$

$$E_{11} = (\frac{200 \times 280}{500}) = 112$$

$$E_{11} = (\frac{200 \times 220}{500}) = 88$$

ดังนั้น $\chi^2 = \frac{(200-168)^2}{168} + \frac{(100-132)^2}{132} + \frac{(80-112)^2}{112} + \frac{(120-88)^2}{88}$

$$\chi^2 = 34.63$$

ในงานวิจัยนี้จะกำหนดค่าระดับนัยสำคัญเท่ากับ 0.05 และสามารถคำนวณองศาความเสรี (degree of freedom) จาก $df = (C\text{-}1)(R\text{-}1)$

เมื่อ C คือจำนวน Column

R คือจำนวน Row

ดังนั้น $df = (2\text{-}1)(2\text{-}1) = 1*1 = 1$

แล้วสามารถหาค่า $\chi^2_{0.05}$ จากตารางค่าไคกำลังสองได้เท่ากับ 3.84 ซึ่งค่าไคกำลังสองของหลักเกณฑ์ ขนมปัง → นม จากตาราง 2.5 สามารถคำนวณได้ค่าเท่ากับ 34.63 แสดงว่าหลักเกณฑ์ดังกล่าวมีความสัมพันธ์กันระหว่างพจน์หน้าและพจน์หลังของหลักเกณฑ์ หรือปฏิเสธสมมติฐาน H_0 นั่นเอง

เมื่อทดสอบความสัมพันธ์ของหลักเกณฑ์แล้ว จะมีการกำหนดทิศทาง (direction) ให้หลักเกณฑ์ สามแบบคือ 0, 1 และ -1 ถ้าทำการทดสอบแล้วหลักเกณฑ์มีความเป็นอิสระต่อกันก็จะกำหนดให้ค่ากำหนดทิศทางเป็น 0 แล้วทำการตัดหลักเกณฑ์นั้น แต่ถ้าทดสอบแล้วได้ว่าไม่เป็นอิสระต่อกัน ก็ให้นำหลักเกณฑ์ที่ผ่านการคัดเลือกมาหาค่ากำหนดทิศทางโดยคำนวณจากการนำค่า O_{ij} และ E_{ij} ของหลักเกณฑ์มาเปรียบเทียบเพื่อหาค่ากำหนดทิศทางตามงานวิจัย [15] ดังนี้

ถ้า $\frac{O_{11}}{E_{11}} > 1$ แล้ว กำหนดให้ค่ากำหนดทิศทางของหลักเกณฑ์เป็น 1

แต่ถ้า $\frac{O_{11}}{E_{11}} < 1$ แล้ว กำหนดให้ค่ากำหนดทิศทางของหลักเกณฑ์เป็น -1

จากการเปรียบเทียบข้างต้น เป็นการวัดความสัมพันธ์กันของข้อมูลจริงกับข้อมูลที่คาดหวัง เมื่อพิจารณาค่ากำหนดทิศทางของหลักเกณฑ์ที่มีค่าเท่ากับ 1 นั้นจะเห็นว่าหลักเกณฑ์เชื่อมโยงมีความสัมพันธ์ไปในทิศทางเดียวกันก็คือถ้า O_{11} มีค่ามากแล้ว E_{11} จะมีแนวโน้มที่จะมีค่ามากด้วย (มีการพบ X และ Y มากในข้อมูลจริงแล้วค่าคาดหวังที่จะพบ X และ Y มากตามไปด้วย) แต่ถ้าค่ากำหนดทิศทางของหลักเกณฑ์มีค่าเท่ากับ -1 จะเห็นว่าหลักเกณฑ์เชื่อมโยงมีความสัมพันธ์ไปในทิศทางตรงข้ามกันก็คือถ้า O_{11} มีค่ามากแล้ว E_{11} จะมีแนวโน้มที่จะมีค่าน้อยลง (มีการพบ X และ Y มากในข้อมูลจริงแต่ค่าคาดหวังที่จะพบ X และ Y มีค่าน้อย) และจากตัวอย่างการขายสินค้าข้างต้นสามารถหาค่ากำหนดทิศทางได้จาก $\frac{200}{168} = 1.19 > 1$ ดังนั้นค่ากำหนดทิศทางของหลักเกณฑ์มีค่าเท่ากับ 1 แสดงให้เห็นว่าจากข้อมูลการซื้อขนมปังกับนมมีความสัมพันธ์ไปในทิศทางเดียวกันหรือก็คือ เมื่อมีการซื้อขนมปังสูงขึ้นแล้วคาดว่าการซื้อนมจะสูงขึ้นตาม ขั้นตอนการหาหลักเกณฑ์กำหนดทิศทางจะมีขั้นตอนดังต่อไปนี้

2.4.1 **ขั้นตอนวิธี P – DS**

ขั้นตอนวิธี P-DS เป็นขั้นตอนวิธีในการหาหลักเกณฑ์กำหนดทิศทาง (DS rule) ที่ประกอบด้วยขั้นตอนการตัดหลักเกณฑ์ที่ไม่น่าสนใจออก และนำหลักเกณฑ์ที่เหลือไปหาหลักเกณฑ์กำหนดทิศทางดังนี้

กำหนดให้	*LHS*	แทนเซตของพจน์หน้าของหลักเกณฑ์
	RHS	แทนพจน์หลังของหลักเกณฑ์
	maxsizeLHS	แทนขนาดของพจน์หน้าที่มากที่สุด
	r.d*ir*	แทนค่ากำหนดทิศทางของหลักเกณฑ์
	r.prune	เซตว่าหลักเกณฑ์ถูกตัด
	r.justify	เซตเพื่อนำไปตรวจสอบว่าเป็น DSR
	DSR	แทนเซตของหลักเกณฑ์กำหนดทิศทาง
	NON-DSR	แทนเซตของหลักเกณฑ์ที่ไม่เป็นหลักเกณฑ์กำหนดทิศทาง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1. for each $X \in LHS$ and $Y \in RHS$ from level-1 to level-n
2. r.dir = *compDir(X, Y)*
3. if X is level-1
4. if $r.dir = 1$ then $DSR = DSR \cup X$
5. else r.prune = X
6. else for each pair $r_1(x \rightarrow y)$ and $r_{rest}(x_{rest} \rightarrow y)$ of *r*, where $x_1 \in X$ and $x_{rest} = X - \{x_1\}$
7. if r.prune = X then exit-for
8. else if r.dir ≠ 0 then r.prune = r_{rest}
9. elseif compDir(r, r_{rest}) ≠ 1 then
10. r.prune = r_{rest}
11. if r_1.dir = r_{rest}.dir = 1then
12. if r.dir = 1 then r.justify = Ø
13. else r.dir = undefined
14. elseif (r_1.dir = 1and r_{rest}.dir = 0) or (r_1.dir = 0 and r_{rest}.dir = 1) then
15. if r.dir = 1 then r.justify = Ø
16. else r.dir = undefined
17. elseif r_1.dir = r_{rest}.dir = 0 then
18. if r.dir = 0 then r.justify = Ø
19. elseif r.dir = 1 then r.justify = r.justify ∪ r
20. else r.dir = undefined
21. else if r.dir = 1 then r.justify = r.justify ∪ r

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

22. else r.dir = undefined
23. endfor
24. if r.justify ≠ Ø then
25. if r is prune then r.dir = undefined
26. else $DSR = DSR \cup X$
27. endif
28. endfor
29. unprune = { $X \in LHS$ | r.dir =1, r.prune ≠ r }
30. *NON-DSR* = unprune - *DSR*

รูปที่ 2.3 ขั้นตอนวิธี P - DS

ขั้นตอนวิธี P-DS ที่แสดงดังรูป 2.3 อธิบายได้ดังนี้

- พิจารณาแต่ละพจน์หน้าของหลักเกณฑ์หรือ *LHS* ที่มีขนาดตั้งแต่ 1 - ไอเทม ไปจนถึง *n* ไอเทม
- คำนวณค่ากำหนดทิศทางหรือ Direction ถ้าค่าเป็น 0 คือจะเป็นอิสระต่อกัน ค่าเป็น 1 คือ มีความสัมพันธ์ไปในทิศทางเดียวกัน และค่าเป็น -1 คือ มีความสัมพันธ์กันในทิศทางตรงข้ามกัน ดังที่ได้กล่าวมาข้างต้น
- ในบรรทัดที่ 3-5 พิจารณาที่ *LHS* มีจำนวนสมาชิก 1 ตัว หรือ 1 - ไอเทม ถ้าค่ากำหนดทิศทางมีค่าเท่ากับ 1 หลักเกณฑ์นั้นจะเป็นหลักเกณฑ์กำหนดทิศทาง แต่ถ้าไม่เท่ากับ 1 หลักเกณฑ์นั้นจะถูกคัดออก
- ในบรรทัดที่ 6-23 จะพิจารณา *LHS* ที่มีสมาชิกมากกว่า 1 ตัวขึ้นไป โดยทำการแยกสมาชิกออกเป็น 2 เซต คือ r_1 จะเก็บสมาชิกที่มีจำนวน 1 ตัว ส่วนสมาชิกที่เหลือเก็บไว้ใน r_{rest} และจะทำการเปลี่ยน r_1 ให้เป็นสมาชิกทุกตัวใน *LHS* จนครบ

- ในบรรทัดที่ 7-10 พิจารณาว่าหลักเกณฑ์ถูกคัดออกหรือยัง ถ้ายังให้ทำการทดสอบความเป็นอิสระต่อกันของหลักเกณฑ์ กับ r_{rest} เพราะถ้าหลักเกณฑ์กับ r_{rest} มีความเป็นอิสระต่อกันจะแสดงให้เห็นว่าข้อมูลที่อยู่ใน r_1 กับข้อมูลใน r_{rest} ไม่มีความสัมพันธ์กันดังนั้นหลักเกณฑ์จะถูกคัดออก แต่ถ้ามีความสัมพันธ์กันก็ทำการวัดความสัมพันธ์ของหลักเกณฑ์

- บรรทัดที่ 11-22 จะเป็นการหาหลักเกณฑ์กำหนดทิศทางที่แฝงอยู่ ส่วนในบรรทัดที่ 24-26 จะตรวจสอบว่าเป็นหลักเกณฑ์ที่แฝงอยู่หรือไม่ ถ้าใช่ให้กำหนดเป็นหลักเกณฑ์กำหนดทิศทาง

ในงานวิจัยนี้ผู้วิจัยมีความสนใจศึกษาถึง ขั้นตอนวิธีที่จะนำมาช่วยในการลดหลักเกณฑ์ จากการคัดเลือกหลักเกณฑ์เชื่อมโยงโดยใช้ค่าสนับสนุนแบบอ่อนน้อยสุดที่ยอมรับได้กับค่าความเชื่อมั่นน้อยสุดที่ยอมรับได้ ค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุดของหลักเกณฑ์ที่ได้ ซึ่งจะมีการกำหนดค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยที่สุดที่ยอมรับได้มาช่วยในการคัดเลือกหลักเกณฑ์ให้มีจำนวนลดลง แล้วนำแนวคิดการหาหลักเกณฑ์กำหนดทิศทางมาช่วยในการลดหลักเกณฑ์ ซึ่งจะอธิบายเพิ่มเติมในบทที่ 3 ต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

การลดหลักเกณฑ์เชื่อมโยง

ในบทนี้จะอธิบายที่มา และแนวคิดในการลดจำนวนหลักเกณฑ์เชื่อมโยงที่ผ่านตัววัดค่าสนับสนุนแบบอ่อนกับค่าความเชื่อมั่น โดยที่ตัววัดที่นำมาช่วยคือค่าความเชื่อมั่นประพจน์แย้งสลับที่ และขั้นตอนวิธีที่นำมาช่วยคือ ขั้นตอนวิธีการสร้างหลักเกณฑ์กำหนดทิศทางที่ปรับปรุงให้เข้ากับตัววัดค่าสนับสนุนแบบอ่อน

3.1 การสร้างหลักเกณฑ์เชื่อมโยง

การคัดเลือกหลักเกณฑ์เชื่อมโยง จะใช้ตัววัดมาเป็นตัวช่วยในการคัดเลือกหลักเกณฑ์ที่มีความน่าสนใจ ขั้นตอนวิธีที่นิยมได้แก่ขั้นตอนวิธี เอไพออรี ที่มีการใช้ค่าสนับสนุนกับค่าความเชื่อมั่น มาเป็นตัววัดในการคัดเลือกหลักเกณฑ์ แต่การที่จะกำหนดค่าสนับสนุนน้อยสุดที่ยอมรับได้ให้เหมาะสม จำเป็นจะต้องมีความรู้ความเข้าใจในข้อมูล ดังนั้นงานวิจัยนี้จึงสนใจการคัดเลือกหลักเกณฑ์เชื่อมโยงโดยใช้ตัววัด ค่าสนับสนุนแบบอ่อนกับค่าความเชื่อมั่น ซึ่งค่าสนับสนุนแบบอ่อน เป็นตัววัดที่มีแนวคิดเกิดมาจากหลักตรรกศาสตร์ ซึ่งไม่สนใจข้อมูลที่ขัดแย้งกับหลักเกณฑ์

พิจารณาประพจน์ p และ q และนิยามประพจน์ ถ้า - แล้ว ทางตรรกศาสตร์ $p \rightarrow q$ ข้อขัดแย้งของประพจน์ คือ $p \wedge \neg q$ เมื่อ p และ q คือประพจน์ของหลักเกณฑ์ ในการศึกษาค่าสนับสนุนแบบอ่อนจะทำการศึกษาหลักเกณฑ์ที่พจน์หลังมีสมาชิก 1 ตัว เท่านั้น ซึ่งในงานวิจัยนี้จะสนใจหลักเกณฑ์ที่พจน์หลังมีสมาชิก 1 - ไอเทมที่เฉพาะเจาะจงข้อมูล หรือเป็นลักษณะประจำเป้าหมาย ขั้นตอนวิธีในการสร้างหลักเกณฑ์เชื่อมโยงจะใช้แนวคิดของขั้นตอนวิธีเอไพออรี แต่ใช้ตัววัดค่าสนับสนุนแบบอ่อนกับค่าความเชื่อมั่น มาเป็นตัววัดแทน ซึ่งขั้นตอนวิธีของงานวิจัยนี้จะปรับปรุงจากขั้นตอนวิธี WS โดย ชรีย์วรรณ แล้วทำการลดหลักเกณฑ์เชื่อมโยงให้มีจำนวนหลักเกณฑ์น้อยลง

3.2 การลดหลักเกณฑ์เชื่อมโยง

หลักเกณฑ์เชื่อมโยงที่ผ่านขั้นตอนวิธี WS ที่ใช้ตัววัดค่าสนับสนุนแบบอ่อนจะมีจำนวนหลักเกณฑ์ที่ออกมามาก ผู้วิจัยจึงนำตัววัดอีกหนึ่งตัวเข้ามาช่วยในการคัดเลือกหลักเกณฑ์เชื่อมโยงคือ ค่าความเชื่อมั่นประพจน์แย้งสลับที่ ซึ่งใช้แนวความคิดจากหลักตรรกศาสตร์ ภายใต้สมมติฐานที่ว่า ค่าความเชื่อมั่นของหลักเกณฑ์จะมีค่าไปในทิศทางเดียวกันกับค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ ดังนั้นผู้วิจัยกำหนดค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุดที่ยอมรับได้ให้เท่ากับค่าความเชื่อมั่นน้อยสุดที่ยอมรับได้ ซึ่งจะช่วยในการคัดเลือกหลักเกณฑ์เชื่อมโยงออกบางส่วน แล้วจะนำแนวความคิดในการหาหลักเกณฑ์กำหนดทิศทางเข้ามาช่วยในการลดจำนวนหลักเกณฑ์เชื่อมโยงในขั้นต่อไป

ในการหาหลักเกณฑ์กำหนดทิศทาง จะมีการใช้ค่าไคกำลังสองเพื่อพิจารณาความสัมพันธ์ของลักษณะประจำกับลักษณะประจำเป้าหมาย แล้วคัดเลือกหลักเกณฑ์ที่ไม่น่าสนใจออก นำหลักเกณฑ์เชื่อมโยงที่เหลือมาเลือกหลักเกณฑ์ที่อธิบายข้อสรุปของข้อมูล ตัวอย่างเช่น จากข้อมูลในการขายสินค้าปลีกพบหลักเกณฑ์ว่า ถ้ามีการซื้อขนมปังแล้ว จะมีการซื้อนม มีค่าสนับสนุนเท่ากับ 40% ของข้อมูลทั้งหมดและมีค่าความเชื่อมั่นเท่ากับ 80% (ขนมปัง → นม [sup = 40%, conf = 80%]) กับหลักเกณฑ์ที่ว่า ถ้ามีการซื้อขนมปังกับเนยแล้ว จะมีการซื้อนม มีค่าสนับสนุนเท่ากับ 20% ของข้อมูลทั้งหมดและมีค่าความเชื่อมั่นเท่ากับ 81% (ขนมปัง, เนย → นม [sup = 20%, conf = 81%]) พิจารณาจากหลักเกณฑ์ที่ได้จะเห็นว่าเมื่อมีการซื้อขนมปังและมีการซื้อนม มีค่าความเชื่อมั่นใกล้เคียงกันกับข้อมูลที่ว่ามีการซื้อขนมปังกับเนยและมีการซื้อนมด้วย ซึ่งค่าความเชื่อมั่นเป็นตัววัดที่แสดงให้เห็นสัดส่วนของหลักเกณฑ์คือจำนวนข้อมูลของหลักเกณฑ์เทียบกับจำนวนข้อมูลของพจน์หน้าของหลักเกณฑ์ ดังนั้นหลักเกณฑ์ที่แสดงภาพสรุปที่ดีกว่าคือหลักเกณฑ์ที่ถ้ามีการซื้อขนมปังแล้วจะมีการซื้อนม (ขนมปัง → นม)

จากการที่ขั้นตอนวิธีการหาหลักเกณฑ์กำหนดทิศทาง จะมีการนำค่าไคกำลังสองมาพิจารณาความสัมพันธ์ที่เกิดขึ้นในหลักเกณฑ์ ข้อมูลที่นำมาคำนวณจะอยู่ในรูปตารางการจรณ์ (contingency table) ดังตารางที่ 2.4 ซึ่งจะพบปัญหาในการคำนวณค่าไคกำลังสองคือจำนวนข้อมูลในบางช่องของตารางมีปริมาณข้อมูลน้อย หรือบางหลักเกณฑ์เชื่อมโยงที่ผ่านการคัดเลือกจากขั้นตอนวิธี WS จำนวนข้อมูลในตารางบางค่าไม่เกิดขึ้นเลย เช่น จำนวนข้อมูลในตารางที่เก็บค่าการพบพจน์หน้าแต่ไม่พบพจน์หลังของหลักเกณฑ์ (ข้อมูลที่ขัดแย้งกับหลักเกณฑ์) มีค่าใกล้ศูนย์หรือไม่พบข้อมูลในส่วนนี้เกิดขึ้นเลย สาเหตุหนึ่งเกิดจากการใช้ค่าสนับสนุนแบบอ่อนในการ

คัดเลือกหลักเกณฑ์เชื่อมโยง จึงทำให้ข้อมูลในตารางที่เก็บค่านี้ไว้มีจำนวนน้อยหรือไม่เกิดขึ้นเลย เมื่อค่าสนับสนุนแบบอ่อนมีค่าเป็น 1

จากปัญหาที่เกิดขึ้นผู้วิจัยจึงแบ่งการพิจารณาเป็นกรณีย่อยดังนี้

- กรณีที่พจน์หน้ามีสมาชิก 1 ตัว

  ในกรณีนี้จะทำตามขั้นตอนวิธีการหาหลักเกณฑ์กำหนดทิศทาง คือ ทดสอบความสัมพันธ์ของพจน์หน้า กับพจน์หลังของหลักเกณฑ์เชื่อมโยง โดยการคำนวณค่าไคกำลังสอง แล้วเปรียบเทียบค่าว่าผ่านการทดสอบหรือไม่ ถ้าออกมาได้ว่ามีความเป็นอิสระต่อกัน ก็กำหนดให้ค่ากำหนดทิศทางมีค่าเป็น 0 แต่ถ้าทดสอบออกมาแล้วได้ว่าไม่เป็นอิสระต่อกันแล้ว ก็นำไปทำการคำนวณค่ากำหนดทิศทางว่าออกมามีค่าเป็น 1 หรือ -1 ตามบทที่ 2 ที่ได้กล่าวมาแล้ว ซึ่งถ้าค่าออกมาเป็น 1 จะได้ว่าหลักเกณฑ์เชื่อมโยงนั้นเป็นหลักเกณฑ์กำหนดทิศทาง หรือ DS rule แต่ถ้ามีค่าเป็น -1 แล้วจะได้ว่าหลักเกณฑ์เชื่อมโยงนั้นไม่เป็นหลักเกณฑ์กำหนดทิศทาง หรือเป็น NON-DS rule

- กรณีที่พจน์หน้ามีสมาชิกตั้งแต่ 2 ตัวขึ้นไป

  กรณีนี้จะใช้แนวคิดของขั้นตอนวิธีการหาหลักเกณฑ์กำหนดทิศทาง คือ ในกรณีที่พจน์หน้าของหลักเกณฑ์มีสมาชิกมากกว่า 1 ตัว จะแยกสมาชิกของพจน์หน้าออกเป็น 2 ส่วน ส่วนแรกจะทำการแยกออกมา 1 ตัว ใส่ไว้ในเซต r_1 ส่วนสมาชิกที่เหลือจะใส่ไว้ในเซต r_{rest} เพื่อพิจารณาว่าหลักเกณฑ์ที่ไม่มี r_1 ก็สามารถอธิบายข้อสรุปของข้อมูลได้กล่าวคือหลักเกณฑ์ที่มี r_{rest} เป็นพจน์หน้าของหลักเกณฑ์เพียงพอแล้วที่จะอธิบายข้อสรุปของข้อมูล ผู้วิจัยคัดเลือกหลักเกณฑ์ที่พิจารณาดังนี้

  สมมติหลักเกณฑ์ที่ได้คือ $X \rightarrow Y$ พจน์หน้าของหลักเกณฑ์มีสมาชิกมากกว่า 1 ตัว โดยที่ $X = \{x, z\}$ และ $Y = \{y\}$ ดังนั้นจะทำการแยกพจน์หน้าออกเป็นสองส่วนได้ดังนี้

  ให้ r_1 เก็บ x และให้ r_{rest} เก็บ z ข้อมูลที่จะนำมาเปรียบเทียบจะอยู่ในตารางที่ 3.1 และ ตารางที่ 3.2

$\{x\}$	$\{y\}$	$\neg\{y\}$
$\{z\}$	$T_{\{x,y,z\}}$	$T_{\{x,\neg y,z\}}$
$\neg\{z\}$	$T_{\{x,y,\neg z\}}$	$T_{\{x,\neg y,\neg z\}}$

ตารางที่ 3.1 ตารางของข้อมูลที่เจาะจงการพบ x

$\neg\{x\}$	$\{y\}$	$\neg\{y\}$
$\{z\}$	$T_{\{\neg x,y,z\}}$	$T_{\{\neg x,\neg y,z\}}$
$\neg\{z\}$	$T_{\{\neg x,y,\neg z\}}$	$T_{\{\neg x,\neg y,\neg z\}}$

ตารางที่ 3.2 ตารางของข้อมูลที่เจาะจงการไม่พบ x

ข้อมูลในตารางที่ 3.1 จะเป็นข้อมูลของหลักเกณฑ์ที่เจาะจงการพบ x ส่วนข้อมูลในตาราง 3.2 จะเป็นข้อมูลของหลักเกณฑ์ที่เจาะจงการไม่พบ x เนื่องจาก $T_{\{x,y,z\}}$ เป็นเซตของระเบียนที่มีการพบทั้ง x, y, z ดังนั้น $T_{\{\neg x,y,z\}}$ เป็นเซตของระเบียนที่มีการพบ y, z แต่ไม่พบ x เริ่มแรกจะทำการตรวจสอบข้อมูลในส่วน $\{x,\neg y,z\}$ ว่าเท่ากับศูนย์หรือไม่ ถ้าเท่ากับศูนย์ก็ให้ไปตรวจสอบว่านิเสธของหลักเกณฑ์ $\{x\}\rightarrow\{y\}$ ($T_{\{x,\neg y\}}$) หรือนิเสธของหลักเกณฑ์ $\{z\}\rightarrow\{y\}$ ($T_{\{z,\neg y\}}$) มีค่าเท่ากับศูนย์ด้วยหรือไม่ ถ้ามีค่าเป็นศูนย์จะสรุปว่าหลักเกณฑ์ $\{x,z\}\rightarrow\{y\}$ ไม่เป็นหลักเกณฑ์กำหนดทิศทาง แต่ถ้าไม่เท่ากับศูนย์จะกลับไปพิจารณา $T_{\{x,\neg y,z\}}$ โดยทำการเปรียบเทียบค่าความเชื่อมั่นว่าจากข้อมูลทั้งสองตาราง โดยคำนวณค่าความเชื่อมั่นของตารางที่ 3.1 จาก $\frac{T_{\{x,y,z\}}}{T_{\{x,y,z\}}+T_{\{x,\neg y,z\}}}$ และค่าความเชื่อมั่นของตารางที่ 3.2 จาก $\frac{T_{\{\neg x,y,z\}}}{T_{\{\neg x,y,z\}}+T_{\{\neg x,\neg y,z\}}}$ สมมติให้มีค่าใกล้เคียงกันจะได้ว่า

$$\frac{T_{\{x,y,z\}}}{T_{\{x,y,z\}}+T_{\{x,\neg y,z\}}}\approx\frac{T_{\{\neg x,y,z\}}}{T_{\{\neg x,y,z\}}+T_{\{\neg x,\neg y,z\}}}$$

$$\frac{T_{\{x,y,z\}}}{T_{\{x,\neg y,z\}}} \approx \frac{T_{\{\neg x,y,z\}}}{T_{\{\neg x,\neg y,z\}}}$$

ดังนั้นค่าระหว่าง $\frac{T_{\{x,y,z\}}}{T_{\{x,\neg y,z\}}}$ ในตารางที่ 3.1 กับ $\frac{T_{\{\neg x,y,z\}}}{T_{\{\neg x,\neg y,z\}}}$ ในตารางที่ 3.2 ถ้ามีค่าใกล้เคียงกันจะสรุปว่า $\{x,z\} \rightarrow \{y\}$ ไม่เป็นหลักเกณฑ์กำหนดทิศทาง แต่ถ้าไม่ใกล้เคียงกันจะสรุปว่า $\{x,z\} \rightarrow \{y\}$ เป็นหลักเกณฑ์กำหนดทิศทาง เนื่องจากค่าความเชื่อมั่นเป็นค่าของจำนวนข้อมูลที่พบหลักเกณฑ์ เทียบกับจำนวนข้อมูลที่พบพจน์หน้าของหลักเกณฑ์ ซึ่งถ้าหลักเกณฑ์ทั้งสองมีค่าใกล้เคียงกัน แสดงว่าแต่ละทรานเซคชันของหลักเกณฑ์ $\{x,z\} \rightarrow \{y\}$ กับ $\{z\} \rightarrow \{y\}$ จะพบ x และ z อยู่ในทรานเซคชันเดียวกันเป็นส่วนมาก ดังนั้นถ้าค่าทั้งสองมีค่าใกล้เคียงกันก็แสดงว่าหลักเกณฑ์ที่มี $r_{rest}(\{z\} \rightarrow \{y\})$ เป็นพจน์หน้าของหลักเกณฑ์สามารถอธิบายข้อสรุปของข้อมูลได้เพียงพอแล้ว จึงไม่มีความจำเป็นที่ต้องมีหลักเกณฑ์ $r(\{x,z\} \rightarrow \{y\})$ เกิดขึ้น ส่วนในกรณีที่พจน์หน้ามีสมาชิกมากกว่า 2 ตัว ก็ทำในทำนองเดียวกันคือ แบ่งข้อมูล 1 ตัวใส่เซต r_1 ข้อมูลที่เหลือเก็บไว้ใน r_{rest} แล้วพิจารณาเหมือนเดิม ดังขั้นตอนวิธีที่จะกล่าวถึงต่อไป

3.3 ขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยง

เนื่องจากขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยง จะศึกษาหลักเกณฑ์ที่สร้างจากขั้นตอนวิธี WS ที่ใช้ตัววัดค่าสนับสนุนแบบอ่อนและพิจารณาพจน์หลังของหลักเกณฑ์เป็น 1-ไอเทมเซต ซึ่งขั้นตอนวิธีในการลดหลักเกณฑ์นั้นจะพิจารณาพจน์หลังของหลักเกณฑ์เป็น 1-ไอเทมที่เป็นลักษณะประจำเป้าหมายเท่านั้น

กำหนดให้ *minWS* แทนค่าสนับสนุนแบบอ่อนน้อยสุดที่กำหนดโดยผู้ใช้

minC แทนค่าความเชื่อมั่นน้อยสุดที่กำหนดโดยผู้ใช้

CC แทนค่าความเชื่อมั่นประพจน์แย้งสลับที่

minCC แทนค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุด

LHS แทนเซตของพจน์หน้าของหลักเกณฑ์

RHS แทนพจน์หลังของหลักเกณฑ์

CandidateLHS แทนเซตของพจน์หน้าตัวเลือก

maxsizeLHS แทนขนาดของพจน์หน้าที่มากที่สุด

Dir แทนค่ากำหนดทิศทาง

DSR แทนเซตของหลักเกณฑ์กำหนดทิศทาง

NON-DSR แทนเซตของหลักเกณฑ์ที่ไม่เป็นหลักเกณฑ์กำหนดทิศทาง

ขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยง แสดงดังรูปที่ 3.1 และอธิบายได้ดังต่อไปนี้

```
1.  y = RHS
2.  LHS = { }
3.  for each item X ∈ I and X ∩ Y = ∅
4.      if X ⟶ Y pass minWS and minC
5.          LHS = LHS ∪ X
6.  for(k = 2; k < LHS.size; k++)
7.      CandidateLHS = Generate(LHS, k)
8.      for each X ∈ CandidateLHS
9.          if X ⟶ Y pass minWS and minC
10.             CC = computed contrapositive confidence
11.             if  CC >= minCC
```

12. $LHS = LHS \cup X$

13. for each $X \in LHS$ from level-1 to level-n

14. if X is level-1

15. $Dir = compDir(LHS)$

16. if $Dir = 1$

17. $DSR = DSR \cup X$

18. else for each pair $r_1(x \rightarrow y)$ and $r_{rest}(x_{rest} \rightarrow y)$ of r, where $x_1 \in X$ and $x_{rest} = X - \{x_1\}$

19. if $compDir2(LHS) = 1$

20. $DSR = DSR \cup X$

21. else

22. $NON\text{-}DSR = NON\text{-}DSR \cup X$

รูปที่ 3.1: ขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยง

1. กำหนดพจน์หลังของหลักเกณฑ์ ให้เป็นลักษณะประจำเป้าหมายของข้อมูล โดยที่ $Y \in I$, $Y = \{y\}$ ซึ่ง y คือข้อมูลของลักษณะประจำเป้าหมาย และเซตให้ *LHS* มีค่าเริ่มต้นเป็นเซตว่าง แล้วพิจารณาแต่ละพจน์หน้าของหลักเกณฑ์ $X \in I$ และ $X \cap Y = \varnothing$ ที่มีขนาด 1- ไอเทม ซึ่งจะนำมาสร้างเป็นหลักเกณฑ์เชื่อมโยง $X \rightarrow Y$ โดยที่หลักเกณฑ์มีค่ามากกว่าเท่ากับ *minWS* และ *minC* ก็จะกำหนดให้ $LHS = LHS \cup X$ และ *LHS* ที่ได้จากขั้นตอนนี้จะเป็นเซตของพจน์หน้าที่เล็กที่สุดที่ทำให้เกิดหลักเกณฑ์ที่ผ่านค่าสนับสนุนแบบอ่อนน้อยสุด กับผ่านค่าความเชื่อมั่นน้อยสุด เพื่อนำไปสร้างหลักเกณฑ์ที่เซตของพจน์หน้ามีขนาดใหญ่ขึ้นจาก *LHS* ในขั้นตอนนี้

2. ในแต่ละรอบการทำงานที่ k จะสร้างเซตของพจน์หน้าตัวเลือก *CandidateLHS* ที่มีขนาด k จาก *LHS* ของขั้นตอนที่ 1 ซึ่งมีขั้นตอนวิธี

ในการสร้างเซตของพจน์หน้าตัวเลือกเหมือนกับขั้นตอนวิธีเอไพออรี แล้วทำการพิจารณาพจน์หน้าที่อยู่ใน *CandidateLHS* ว่าหลักเกณฑ์ $X \rightarrow Y$ เป็นหลักเกณฑ์ที่มีค่าสนับสนุนแบบอ่อนและค่าความเชื่อมั่นมากกว่าเท่ากับ *minWS* และ *minC* ตามลำดับ แล้วนำหลักเกณฑ์ที่ผ่านออกมาได้ มาคำนวณค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ว่ามากกว่าเท่ากับ *minCC* หรือไม่ ถ้ามากกว่าแล้วจะทำการกำหนดให้ $LHS = LHS \cup X$

3. หลังจากได้หลักเกณฑ์เชื่อมโยงผ่านออกมาแล้ว จะเข้าสู่ขั้นตอนการหาหลักเกณฑ์กำหนดทิศทาง ในแต่ละพจน์หน้าของหลักเกณฑ์ตั้งแต่ 1-ไอเทมจนถึง n – ไอเทม พิจารณาพจน์หน้าที่มีขนาด 1 - ไอเทม ทำการคำนวณค่ากำหนดทิศทางตามที่ได้อธิบายในข้างต้น และถ้าค่ากำหนดทิศทางที่คำนวณได้มีค่าเท่ากับ 1 แล้วจะกำหนดให้ $DSR = DSR \cup X$

4. พิจารณาพจน์หน้าที่มีขนาดมากกว่า 1 – ไอเทม แล้วทำการแยกออกเป็น 2 เซต คือ r_1 จะเก็บสมาชิกที่มีขนาด 1 - ไอเทม ส่วนสมาชิกที่เหลือจะเก็บไว้ใน r_{rest}

5. เมื่อได้ r_1 กับ r_{rest} แล้วก็ทำการคำนวณเพื่อให้ได้ค่ากำหนดทิศทางของหลักเกณฑ์ ตามที่ได้อธิบายไว้แล้วในข้างต้น และถ้าค่ากำหนดทิศทางมีค่าเท่ากับ 1 จะกำหนดให้ $DSR = DSR \cup X$ คือจะได้ว่าหลักเกณฑ์นั้นเป็นหลักเกณฑ์กำหนดทิศทาง แต่ถ้าไม่เท่ากับ 1 แล้วกำหนดให้ $NON\text{-}DSR = NON - DSR \cup X$ คือจะได้ว่าหลักเกณฑ์นั้นไม่เป็นหลักเกณฑ์กำหนดทิศทาง

จากขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยงจะใช้ระยะเวลาในการทำงานดังต่อไปนี้

กำหนดให้ $|D|$ แทนจำนวนระเบียนทั้งหมด

$|L_k|$ แทนจำนวนไอเทมเซตใน LHS ที่มีขนาด k

t แทนเวลาที่ใช้ในการเข้าถึงและดึงข้อมูลจากฐานข้อมูลในแต่ละรอบการเข้าฐานข้อมูล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

$maxLHS$ แทนขนาดของพจน์หน้าที่มากที่สุดที่สร้าง

ระยะเวลาที่ใช้ในการเข้าถึงฐานข้อมูลคือ $t|D|$ ดังนั้นในแต่ละรอบการทำงานที่ k การสร้างพจน์หน้าของหลักเกณฑ์จะใช้เวลาในการทำงานเป็น $O(|L_k| \times t|D|)$

และจากขั้นตอนวิธี WS ที่ใช้ในการสร้างหลักเกณฑ์เชื่อมโยง (ในบรรทัดที่ 3-12) ใช้เวลาในการทำงานไม่เกิน $O(|L_{\max LHS-1}|^2 \times t|D|)$ แล้วจากการหาหลักเกณฑ์กำหนดทิศทางในบรรทัดที่ 13-22 จะใช้เวลาในการทำงานประมาณ $O(|L_{\max LHS}|^2 \times t|D|)$ ดังนั้นเวลาที่ใช้ในขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยง คือ

$$O(|L_{\max LHS-1}|^2 \times t|D|) + O(|L_{\max LHS}|^2 \times t|D|)$$

เพราะฉะนั้นเวลาในการทำงานจะไม่เกิน

$$O(|L_{\max LHS}|^2 \times t|D|)$$

งานวิจัยนี้ต้องการเปรียบเทียบจำนวนหลักเกณฑ์เชื่อมโยง ระหว่างขั้นตอนวิธี WS กับขั้นตอนวิธีการลดหลักเกณฑ์เชื่อมโยง และเปรียบเทียบประสิทธิภาพของหลักเกณฑ์เชื่อมโยงที่ได้จากทั้งสองขั้นตอนวิธี ซึ่งการทดลองเพื่อเปรียบเทียบขั้นตอนวิธีทั้งสองจะอธิบายในบทต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

การทดลอง และการวิเคราะห์ผลการทดลอง

ในบทนี้จะกล่าวถึงวิธีการทดลอง ข้อมูลที่ใช้ในการทำการทดลอง ผลการทดลองที่นำมาเปรียบเทียบระหว่างขั้นตอนวิธี WS ที่สร้างหลักเกณฑ์เชื่อมโยงที่ผ่านค่าสนับสนุนแบบอ่อนและค่าความเชื่อมั่น กับหลักเกณฑ์เชื่อมโยงที่ผ่านค่าสนับสนุนแบบอ่อน ค่าความเชื่อมั่น ค่าความเชื่อมั่นประพจน์แย้งสลับที่ และการหาหลักเกณฑ์กำหนดทิศทาง โดยข้อมูลที่ใช้ในการทดลองได้มาจาก UCI repository

4.1 ข้อมูลที่ใช้ทดลอง

ข้อมูลที่นำมาใช้ในการทดลองงานวิจัยนี้ได้มาจากฐานข้อมูล UCI [17,18] มีชุดข้อมูล chess ชุดข้อมูล mushroom และชุดข้อมูล zoo ได้มีการสรุปข้อมูลบางส่วนของแต่ละชุดข้อมูลดังตารางที่ 4.1 ซึ่งข้อมูลจะต้องผ่านขั้นตอนการเตรียมข้อมูลก่อนนำมาทดลอง

ชื่อ	จำนวนระเบียน	จำนวนลักษณะประจำ	จำนวนค่าที่แตกต่าง
Breast-cancer	699	10	92
Chess	3196	37	75
Connect	67557	43	129
mushroom	8124	23	119
zoo	101	18	28

ตารางที่ 4.1 แสดงข้อมูลสรุปที่ใช้ในการทดลอง

ข้อมูล breast-cancer เป็นข้อมูลทางการแพทย์ที่ศึกษาเกี่ยวกับโรคมะเร็งเต้านม จาก University of Wisconsin Hospitals ซึ่งเก็บข้อมูลเกี่ยวกับลักษณะของเซลล์ที่พิจารณา เช่น การจับตัวเป็นกลุ่มของเซลล์ ขนาดของเซลล์ ลักษณะของเซลล์ เป็นต้น ข้อมูลเป็นแบบจำแนกประเภท ที่มีลักษณะประจำเป้าหมาย 2 ค่า คือ เป็นเนื้องอกธรรมดากับเป็นเนื้องอกที่อันตราย (เป็นมะเร็ง)

ข้อมูล chess เป็นข้อมูลที่ได้จากการเดินหมากรุก ของ 2 ฝ่าย ได้แก่ ฝ่ายหมากขาว กับฝ่ายหมากดำ เป็นข้อมูลแบบจำแนกประเภทลักษณะประจำเป้าหมาย 2 ค่า คือ ฝ่ายหมากขาวเดินชนะ กับฝ่ายหมากขาวเดินไม่ชนะ โดยจะเริ่มเกมที่ฝ่ายขาวเป็นฝ่ายเริ่มเล่นก่อน เนื่องจากข้อมูลนี้เป็นข้อมูลของการเล่นหมากรุก ดังนั้นข้อมูลจะมีรูปแบบที่แน่นอนเพราะการเดินหมากก็จะอยู่บนกระดานหมากและจะเป็นการสลับกันเดินของฝ่ายหมากขาวกับหมากดำจนรู้ผลแพ้ชนะ

ข้อมูล connect เป็นข้อมูลเกี่ยวกับการเล่นเกมส์ connect-4 กับคอมพิวเตอร์ (แข่งกับโปรแกรมคอมพิวเตอร์) เป็นเกมส์ที่คล้ายกับการเล่น OX คือจะมีการวางให้ได้ 4 ตำแหน่งก่อนก็จะชนะ ลักษณะข้อมูลจะเป็นตำแหน่งที่กำหนดไว้ว่ามีการวางหรือไม่ เช่น a_1 = X คือ ตำแหน่งที่ a_1 มีการวางโดย X เป็นต้น และเป็นข้อมูลแบบจำแนกประเภท ที่มีลักษณะประจำเป้าหมาย 3 ค่า คือ เล่นชนะ เล่นแพ้ และเสมอ

ข้อมูล mushroom เป็นข้อมูลที่เก็บรายละเอียดของเห็ด มีการเก็บข้อมูลลักษณะรูปร่าง และข้อมูลทั่ว ๆ ไปของเห็ดตามรูปแบบที่นักพฤกษศาสตร์กำหนด เช่น ลักษณะของลำต้นของเห็ดว่ามีขนาด สี และผิวสัมผัสลักษณะเป็นอย่างไร และเป็นข้อมูลแบบจำแนกประเภท ที่มีลักษณะประจำเป้าหมาย 2 ค่า คือ เห็ดมีพิษ กับเห็ดที่ไม่มีพิษ

ข้อมูล zoo เป็นข้อมูลรายละเอียดของสัตว์ในสวนสัตว์ จะมีการเก็บข้อมูลรูปร่างลักษณะของสัตว์ เช่น มีขน มีกระดูกสันหลัง เป็นต้น ข้อมูลเป็นแบบจำแนกประเภท ซึ่งข้อมูลในลักษณะประจำเป้าหมายจะเป็นชนิดของสัตว์ แบ่งออกเป็น 7 ชนิด ตามการแบ่งของทางสวนสัตว์

4.2 วิธีการทดลอง

สำหรับงานวิจัยนี้สนใจการลดจำนวนหลักเกณฑ์ และเปรียบเทียบประสิทธิภาพของหลักเกณฑ์เชื่อมโยงที่ได้จากทั้งขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS ผู้วิจัยแบ่งข้อมูลออกเป็น 2 ส่วน ส่วนละ 50% ข้อมูลในส่วนแรกเรียกว่า ข้อมูลทดลอง (train data) เป็นส่วนที่จะนำไปทำการสร้างหลักเกณฑ์เชื่อมโยง ส่วนข้อมูลที่เหลือเรียกว่า ข้อมูลทดสอบ (test data) เป็นข้อมูลในส่วนที่จะนำไปใช้ในการเปรียบเทียบประสิทธิภาพของหลักเกณฑ์เชื่อมโยง

โดยนำข้อมูลในส่วนของข้อมูลทดลองมาทำการสร้างหลักเกณฑ์เชื่อมโยง จากชุดข้อมูลทั้ง 5 ที่ได้กล่าวไว้ในข้างต้น เมื่อได้หลักเกณฑ์ที่ผ่านออกมาจากทั้งขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS แล้ว ก็นำหลักเกณฑ์ที่ได้มาทำการหาค่าสภาพไวจากข้อมูลทดสอบที่เก็บไว้ซึ่งข้อมูลส่วนนี้ไม่ได้นำมาใช้ในการสร้างหลักเกณฑ์เชื่อมโยง

เนื่องจากขั้นตอนวิธีของงานวิจัยนี้ทำการปรับปรุงจากขั้นตอนวิธี WS ดังนั้นในการทดลองจะทำการคำนวณค่าสนับสนุนแบบอ่อนจากค่าสนับสนุนกับค่าความเชื่อมั่น ดังตารางที่ 2.2 ในบทที่ 2 ส่วนในการทดลองนี้จะมีการกำหนดค่าสนับสนุนน้อยสุดมีค่าตั้งแต่ 0.1 ไปจนถึง 0.5 เพิ่มค่าทีละ 0.1 และกำหนดค่าความเชื่อมั่นน้อยสุดมีค่าตั้งแต่ 0.6 ไปจนถึง 1.0 โดยเพิ่มค่าทีละ 0.1 แล้วคำนวณค่าสนับสนุนแบบอ่อนน้อยสุด และนำค่าสนับสนุนแบบอ่อนน้อยสุดที่ได้กับค่าความเชื่อมั่นน้อยสุดไปสร้างหลักเกณฑ์เชื่อมโยง

4.3 การเปรียบเทียบผลการทดลอง

งานวิจัยนี้สนใจการลดจำนวนหลักเกณฑ์เชื่อมโยง และวัดประสิทธิภาพของหลักเกณฑ์ที่เก็บไว้ ดังนั้นจะมีการเปรียบเทียบจำนวนหลักเกณฑ์กับ ค่าสภาพไว (sensitivity) [16] เพื่อใช้ประเมินประสิทธิภาพของหลักเกณฑ์ จากทั้งขั้นตอนวิธี WS และขั้นตอน DSRCCWS

จากการที่นำ สภาพไว มาเป็นตัววัดประสิทธิภาพข้อมูลที่เป็นแบบจำแนกประเภท ซึ่งจะทำการเปรียบเทียบข้อมูลจริงกับข้อมูลที่สรุปจากหลักเกณฑ์ทั้งหมดที่เก็บไว้ สามารถแสดงได้ดังตารางที่ 4.2

ข้อมูลสรุป \ ความจริง	TRUE	FALSE
Positive	TP	FP
Negative	FN	TN

ตารางที่ 4.2 ตารางนิยาม TP, FP, FN, TN

กำหนดให้ ข้อมูลที่สนใจที่มีอยู่จริงเป็นบวก (TRUE)

ไม่มีข้อมูลที่สนใจเป็นลบ (FALSE)

และ มีข้อมูลที่สรุปได้จากหลักเกณฑ์อยู่จริงเป็น Positive

ไม่มีข้อมูลที่สรุปได้จากหลักเกณฑ์เป็น Negative

- TP (True Positive) แทนจำนวนข้อมูลที่สรุปได้จากหลักเกณฑ์ที่ทายว่าเป็นบวก (Positive) แล้วข้อมูลจริงเป็นบวก
- FP (False Positive) แทนจำนวนข้อมูลที่สรุปได้จากหลักเกณฑ์ที่ทายว่าเป็นบวก (Positive) แต่ข้อมูลจริงเป็นลบ
- FN (False Negative) แทนจำนวนข้อมูลที่สรุปได้จากหลักเกณฑ์ที่ทายว่าเป็นลบ (Negative) แต่ข้อมูลจริงเป็นบวก
- TN (True Negative) แทนจำนวนข้อมูลที่สรุปได้จากหลักเกณฑ์ที่ทายว่าเป็นลบ (Negative) และข้อมูลจริงเป็นลบ

สภาพไว หรือเรียกอีกอย่างว่า อัตราบวกถูก (true positive rate) เป็นอัตราส่วนระหว่างจำนวนข้อมูลที่สรุปได้จากหลักเกณฑ์ที่ทายว่าเป็นบวก (Positive) และข้อมูลจริงเป็นบวก เทียบกับจำนวนข้อมูลจริงที่เป็นบวกทั้งหมด ถ้าค่าสภาพไวมีค่าสูงจะได้ว่าตัวแบบที่นำมาทำการทดลองมีประสิทธิภาพสูงในการทำนายเหตุการณ์ที่เป็นบวก

$$sensitivity = \frac{TP}{TP + FN}$$

ข้อมูลที่ใช้ในการทดลองจะเป็นข้อมูลที่เกิดขึ้นในปัญหาจริง และงานวิจัยนี้จะทำการแปลงข้อมูลให้เป็นข้อมูลแบบทวิภาค ดังนั้นเมื่อทำการทดลองแล้วข้อมูลที่เป็นบวกจะหมายความว่าพบการซื้อ แต่ในส่วนข้อมูลที่เป็นลบหรือหมายความถึงการไม่ซื้อ ซึ่งโดยปรกติแล้วจะไม่ทำการบันทึกข้อมูลในส่วนของการไม่พบการซื้อสินค้าไว้ ฉะนั้นงานวิจัยนี้จะไม่สนใจข้อมูลที่เป็นลบจะสนใจเฉพาะข้อมูลที่เป็นบวกเท่านั้นหรือสนใจแต่อัตราบวกถูก

ในการทดสอบประสิทธิภาพของหลักเกณฑ์ จะนำหลักเกณฑ์เชื่อมโยงที่ได้จากข้อมูลทดลองมาทำการเปรียบเทียบกับข้อมูลทดสอบโดยใช้ค่าสภาพไว ซึ่งจะพิจารณาข้อมูลแต่ละระเบียนแล้วนำหลักเกณฑ์เชื่อมโยงที่ผ่านออกมาจากขั้นตอนวิธีมาหาค่าสภาพไว เมื่อคำนวณค่าสภาพไวของแต่ละระเบียนแล้วก็นำค่าสภาพไวมาทำการหาค่าเฉลี่ย ก็จะได้ค่าสภาพไวเฉลี่ยของกลุ่มหลักเกณฑ์

การพิจารณาหลักเกณฑ์ที่สอดคล้อง จะพิจารณาว่าทั้งพจน์หน้าและพจน์หลังของหลักเกณฑ์สอดคล้องกับข้อมูลที่ทำการเปรียบเทียบเช่น หลักเกณฑ์ถ้ามีการซื้อขนมปังแล้วมีการซื้อนม จะเป็นหลักเกณฑ์ที่สอดคล้องเมื่อทำการเปรียบเทียบกับข้อมูล มีการซื้อขนมปังกับการซื้อนมอยู่ในระเบียนหรือบิลใบเสร็จเดียวกัน

4.4 ผลการทดลอง

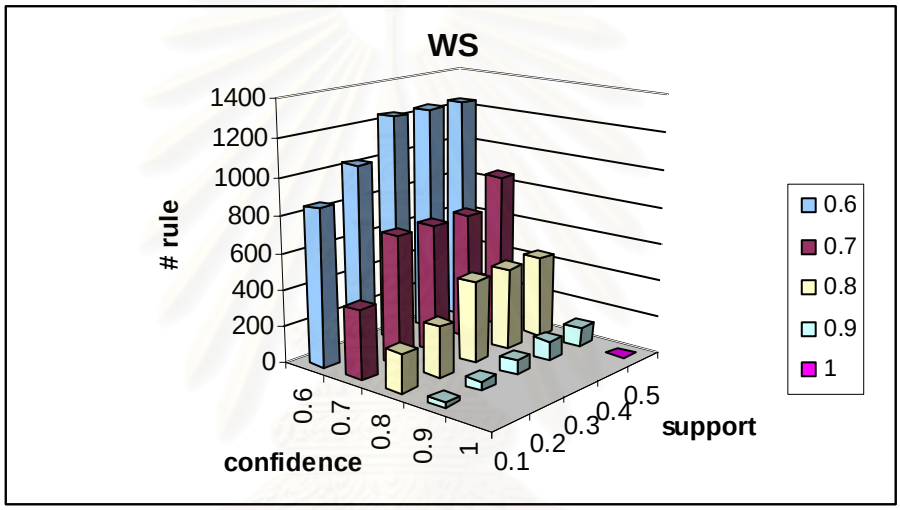
ผลการทดลองสำหรับงานวิจัยนี้จะเน้นที่การลดจำนวนหลักเกณฑ์เชื่อมโยง แล้วทำการเปรียบเทียบประสิทธิภาพของหลักเกณฑ์ที่ได้จากขั้นตอนวิธี WS กับหลักเกณฑ์ที่ได้จากขั้นตอนวิธี DSRCCWS โดยใช้ค่าสภาพไว

ในการทดลอง จะทำการเปรียบเทียบหลักเกณฑ์ที่จะผ่านออกมาให้เหมือนกันทั้งสองขั้นตอนวิธี คือ มีการกำหนดพจน์หลังของหลักเกณฑ์เป็นลักษณะประจำเป้าหมาย และเปรียบเทียบหลักเกณฑ์ที่มีขนาดของพจน์หน้าที่เท่ากัน ส่วนโปรแกรมที่ใช้ในการทดลองจะเขียนจากภาษาจาวา ซอฟต์แวร์ที่ใช้ในการจัดการฐานข้อมูลคือ MySQL 5.0 และเครื่องคอมพิวเตอร์ที่ใช้ทดลองคือ Pentium 4 2.0 GHz หน่วยความจำหลัก 1 GB

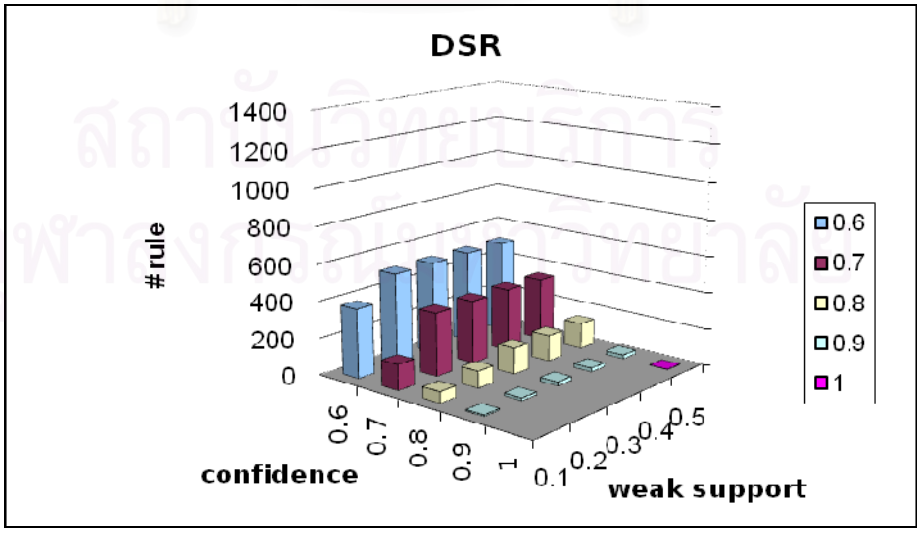
การทดลองนี้จะปรับเปลี่ยนค่าสนับสนุนน้อยสุด จาก 0.1 ไปถึง 0.5 และค่าความเชื่อมั่นน้อยสุด จาก 0.6 ไปถึง 1.0 แล้วทำการคำนวณค่าสนับสนุนแบบอ่อนน้อยสุดดังตารางที่ 2.2 ในบทที่ 2

นำข้อมูลในส่วนของข้อมูลทดลองมาทำการสร้างหลักเกณฑ์เชื่อมโยง โดยใช้ขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS แล้วนำข้อมูลในส่วนข้อมูลทดสอบมาทำการเปรียบเทียบให้ได้ค่าสภาพไวออกมา ผลการทดลองที่ได้จะทำการเปรียบเทียบจำนวนหลักเกณฑ์เชื่อมโยง และค่าสภาพไว ดังรูปที่ 4.1 ถึงรูปที่ 4.20 โดยเปรียบเทียบตามแต่ละชุดข้อมูล กราฟจะแสดงแกน X คือค่าสนับสนุนน้อยสุด แกน Y คือค่าความเชื่อมั่นน้อยสุด และแกน Z คือจำนวนหลักเกณฑ์เชื่อมโยง จากแกน X ที่เป็นค่าสนับสนุนนั้นจะนำไปเปรียบเทียบกับตารางที่ 2.2 ในบทที่ 2 ที่เป็นการคำนวณค่าสนับสนุนกับค่าความเชื่อมั่นเพื่อให้ได้ค่าสนับสนุนแบบอ่อนออกมา

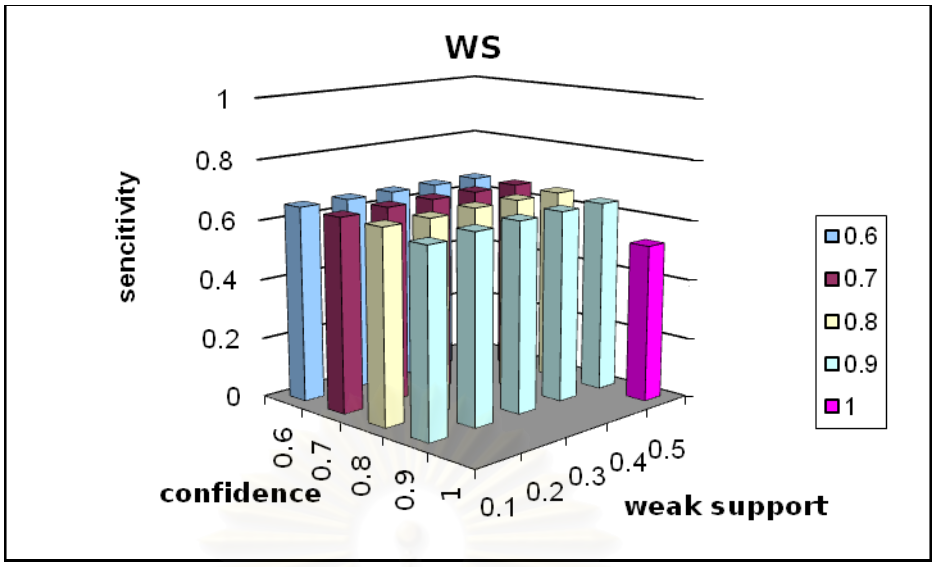
รูปที่ 4.1 ถึง รูปที่ 4.4 แสดงผลการทดลองจากข้อมูล breast-cancer



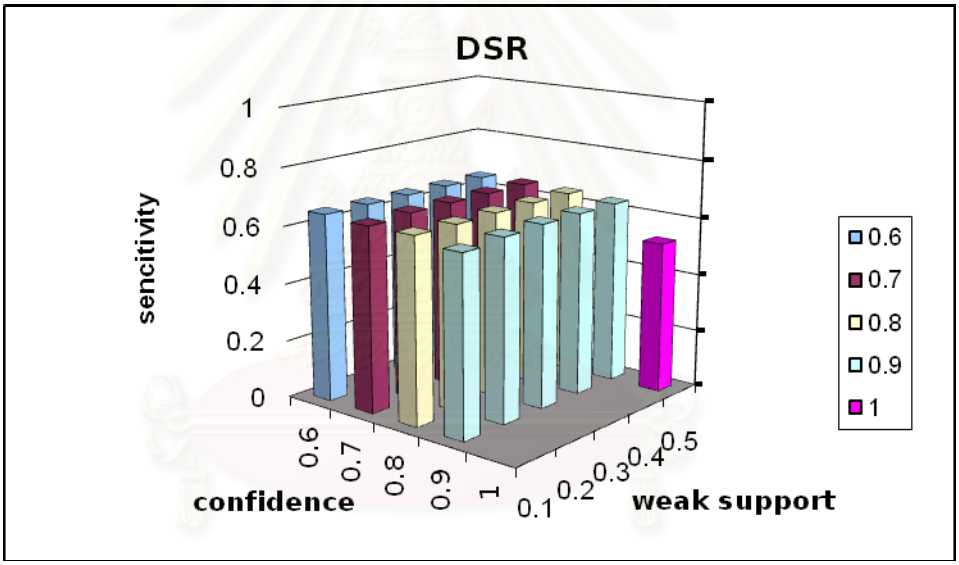
รูปที่ 4.1 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล breast-cancer สำหรับขั้นตอนวิธี WS



รูปที่ 4.2 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล breast-cancer สำหรับขั้นตอนวิธี DSRCCWS



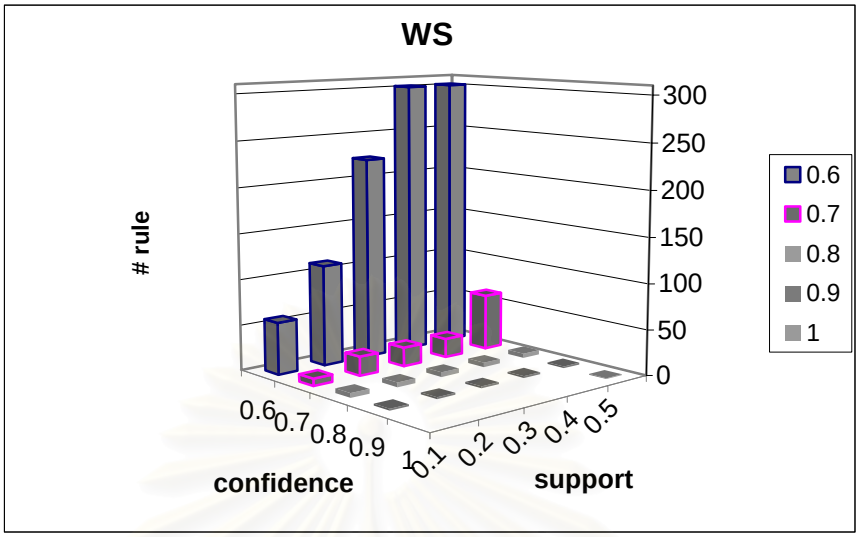
รูปที่ 4.3 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงของข้อมูล breast-cancer สำหรับขั้นตอนวิธี WS



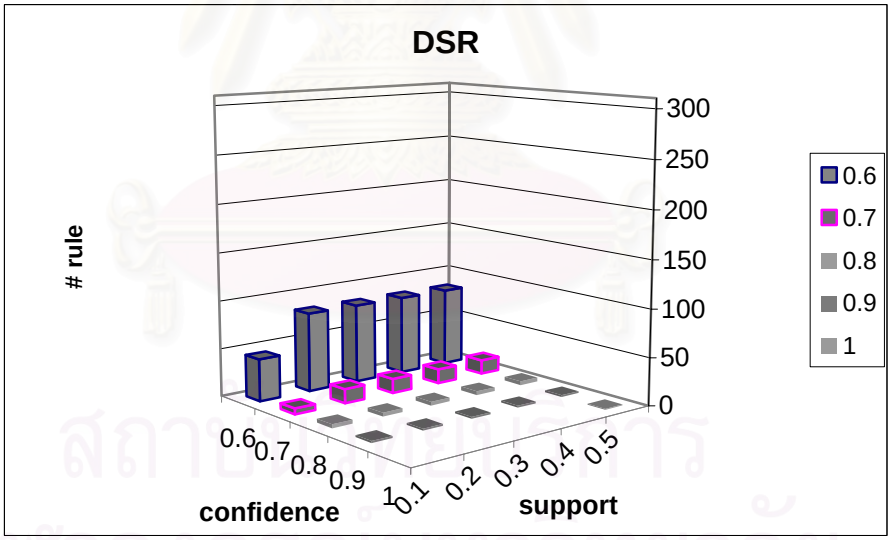
รูปที่ 4.4 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล breast-cancer สำหรับขั้นตอนวิธี DSRCCWS

จากรูปที่ 4.1 และ รูปที่ 4.2 เป็นข้อมูลของจำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล breast-cancer สำหรับขั้นตอนวิธี WS และ ขั้นตอนวิธี DSRCCWS ตามลำดับ จะเห็นว่าจำนวนหลักเกณฑ์เชื่อมโยงจากขั้นตอนวิธี WS จะมีปริมาณหลักเกณฑ์ที่มากกว่าจำนวนหลักเกณฑ์จากขั้นตอนวิธี DSRCCWS ซึ่งถ้าคิดเป็นสัดส่วนแล้วส่วนใหญ่จะมากกว่า 50 เปอร์เซ็นต์ โดยที่ในส่วนของค่าสภาพไวจากรูปที่ 4.3 และรูปที่ 4.4 มีค่าสภาพไวที่เท่ากันจากทั้งขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS

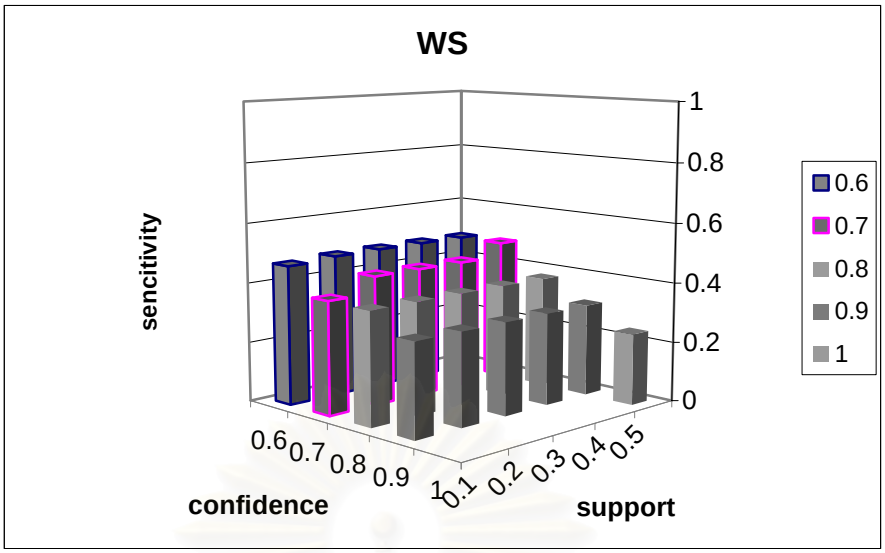
รูปที่ 4.5 ถึง รูปที่ 4.8 แสดงผลการทดลองจากข้อมูล chess



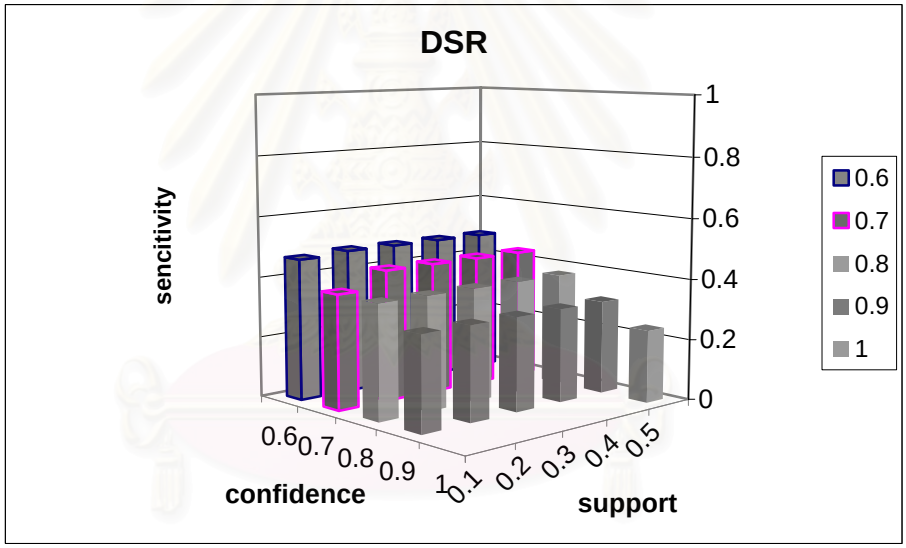
รูปที่ 4.5 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล chess สำหรับขั้นตอนวิธี WS



รูปที่ 4.6 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล chess สำหรับขั้นตอนวิธี DSRCCWS



รูปที่ 4.7 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงของข้อมูล chess สำหรับขั้นตอนวิธี WS

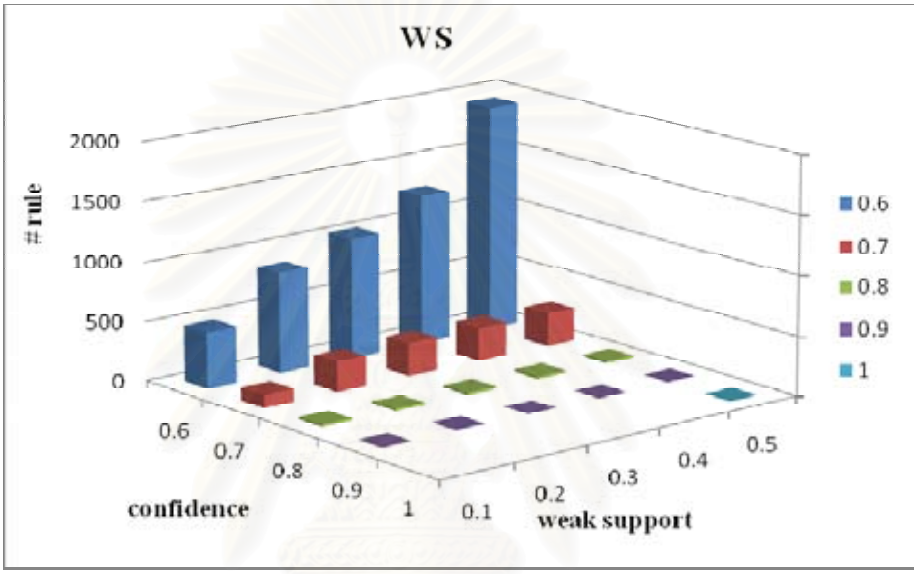


รูปที่ 4.8 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล chess สำหรับขั้นตอนวิธี DSRCCWS

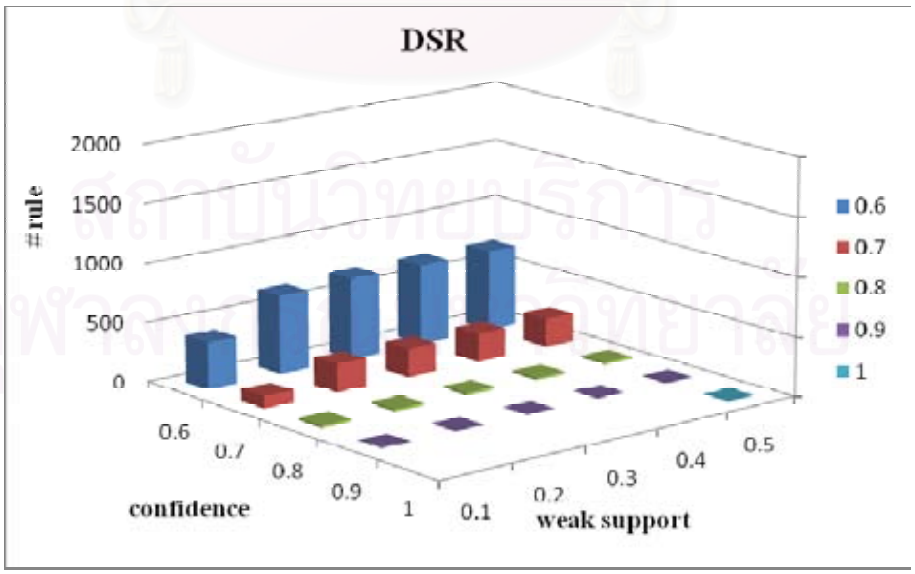
จากรูปที่ 4.5 และ รูปที่ 4.6 เป็นข้อมูลของจำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล chess สำหรับขั้นตอนวิธี WS และ ขั้นตอนวิธี DSRCCWS ตามลำดับ จะพบว่าจำนวนหลักเกณฑ์เชื่อมโยงจากขั้นตอนวิธี WS ในช่วงที่ค่าความเชื่อมั่นน้อยสุดของหลักเกณฑ์มีค่าอยู่ระหว่าง 0.6 ถึง 0.7 นั้นมีจำนวนหลักเกณฑ์มากกว่าขั้นตอนวิธี DSRCCWS ยิ่งในช่วงที่ค่าความเชื่อมั่นมีค่าประมาณ 0.6 จะมีจำนวนหลักเกณฑ์ที่แตกต่างกันเป็นหลักร้อย ซึ่งถ้าคิดเป็นสัดส่วนแล้วส่วนใหญ่จะมากกว่า 50 เปอร์เซ็นต์ โดยที่ในส่วนของค่าสภาพไวจากรูปที่ 4.3 และรูปที่ 4.4 มี

ค่าสภาพไวที่ใกล้เคียงกันมากแทบจะไม่ต่างกัน โดยช่วงที่ค่าสภาพไวมีค่าแตกต่างกันมากที่สุดประมาณ 0.04 โดยอยู่ในช่วงที่ค่าความเชื่อมั่นมีค่า 0.7 และค่าสนับสนุนแบบอ่อนมีค่า 0.78 ซึ่งจำนวนหลักเกณฑ์ก็มีจำนวนแตกต่างกันประมาณ 50 หลักเกณฑ์ ในส่วนที่ค่าความเชื่อมั่นมีค่ามากกว่า 0.7 จำนวนหลักเกณฑ์ของทั้งสองขั้นตอนวิธีมีค่าไม่ใกล้เคียงกัน เช่นเดียวกับค่าสภาพไวของทั้งขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS

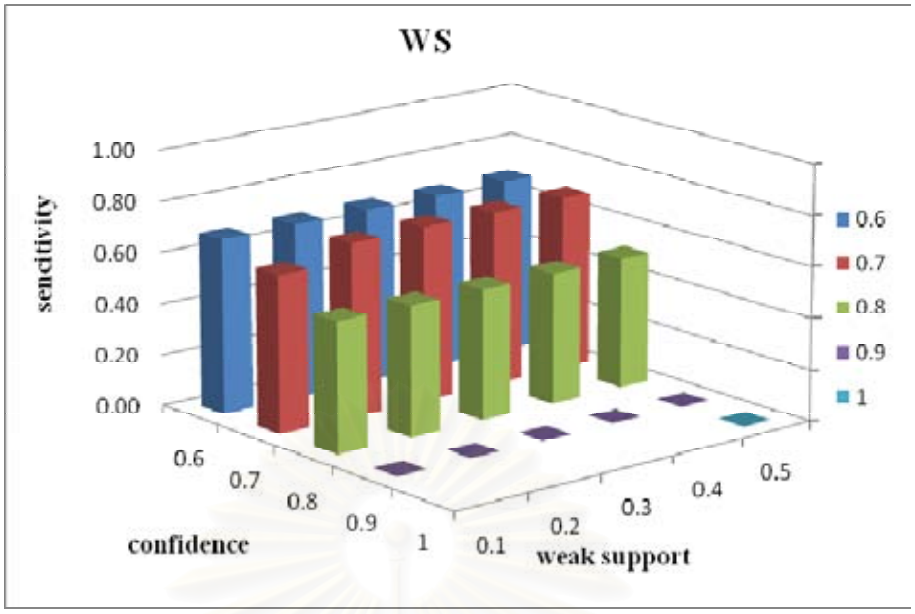
รูปที่ 4.9 ถึง รูปที่ 4.12 แสดงผลการทดลองจากข้อมูล connect



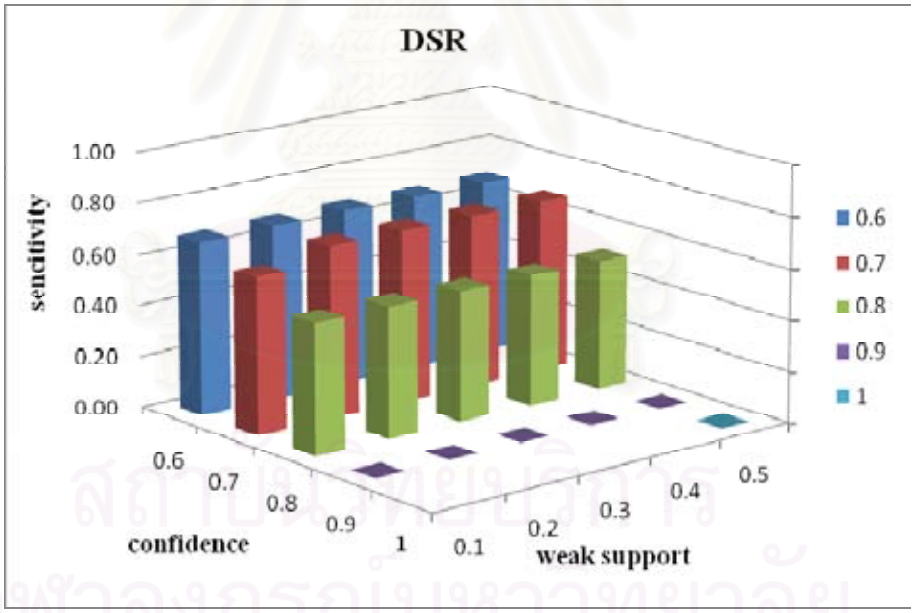
รูปที่ 4.9 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล connect สำหรับขั้นตอนวิธี WS



รูปที่ 4.10 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล connect สำหรับขั้นตอนวิธี DSRCCWS



รูปที่ 4.11 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล connect สำหรับขั้นตอนวิธี WS

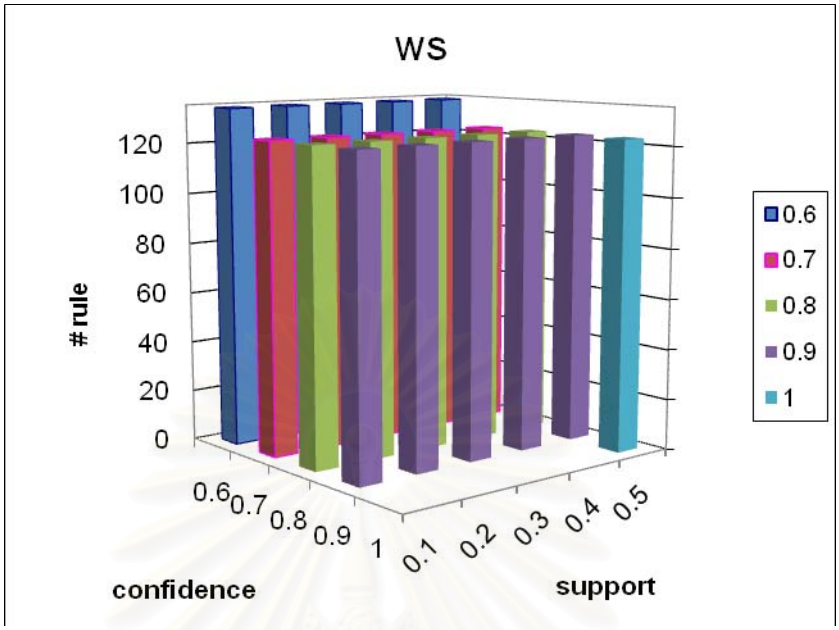


รูปที่ 4.12 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล connect สำหรับขั้นตอนวิธี DSRCCWS

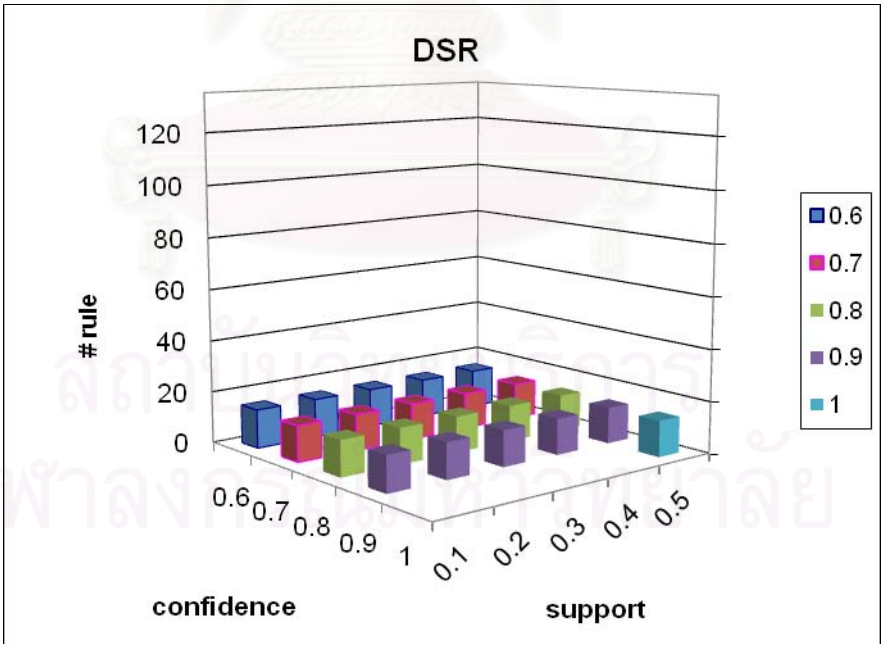
จากรูปที่ 4.9 และ รูปที่ 4.10 เป็นข้อมูลของจำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล connect สำหรับขั้นตอนวิธี WS และ ขั้นตอนวิธี DSRCCWS ตามลำดับ จะพบว่าจำนวนหลักเกณฑ์เชื่อมโยงจากขั้นตอนวิธี WS จะมีปริมาณหลักเกณฑ์ที่มากกว่าจำนวนหลักเกณฑ์จากขั้นตอนวิธี DSRCCWS จะมีจำนวนหลักเกณฑ์เชื่อมโยงแตกต่างกันมากเมื่อค่าความเชื่อมั่นน้อยสุดมีค่าเท่ากับ 0.6 และสัดส่วนความแตกต่างกันจะลดลงเรื่อย ๆ ตามการเพิ่มขึ้นของค่าความเชื่อมั่นและค่าสนับสนุนแบบอ่อน จนมีจำนวนหลักเกณฑ์เชื่อมโยงเท่ากันที่ค่าความเชื่อมั่นเท่ากับ 0.8 ซึ่งที่ค่าความเชื่อมั่นมากกว่า 0.8 จะพบว่าไม่มีหลักเกณฑ์เชื่อมโยงปรากฏออกมา โดยที่ในส่วนของค่าสภาพไวจากรูปที่ 4.11 และรูปที่ 4.12 มีค่าสภาพไวที่เท่ากันจากทั้งขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

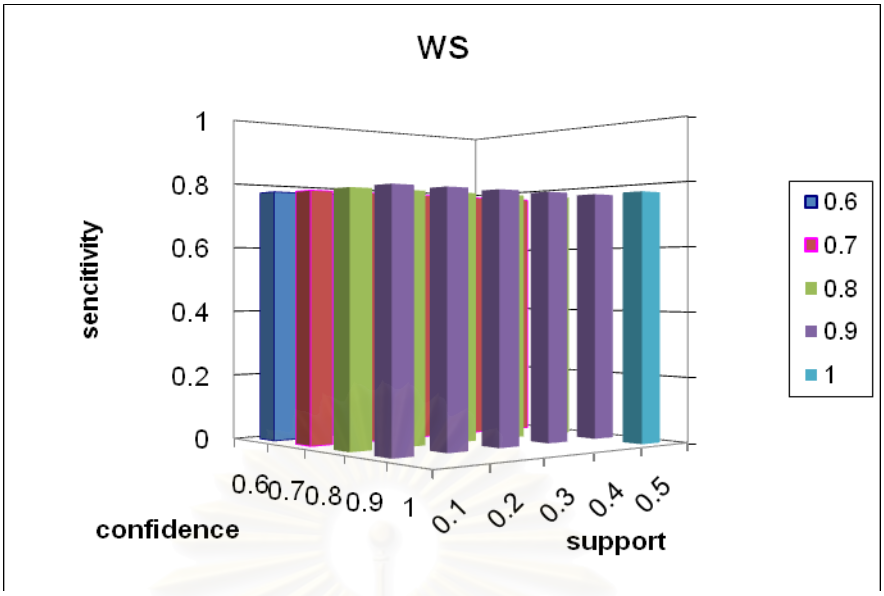
รูปที่ 4.13 ถึง รูปที่ 4.16 แสดงผลการทดลองจากข้อมูล mushroom



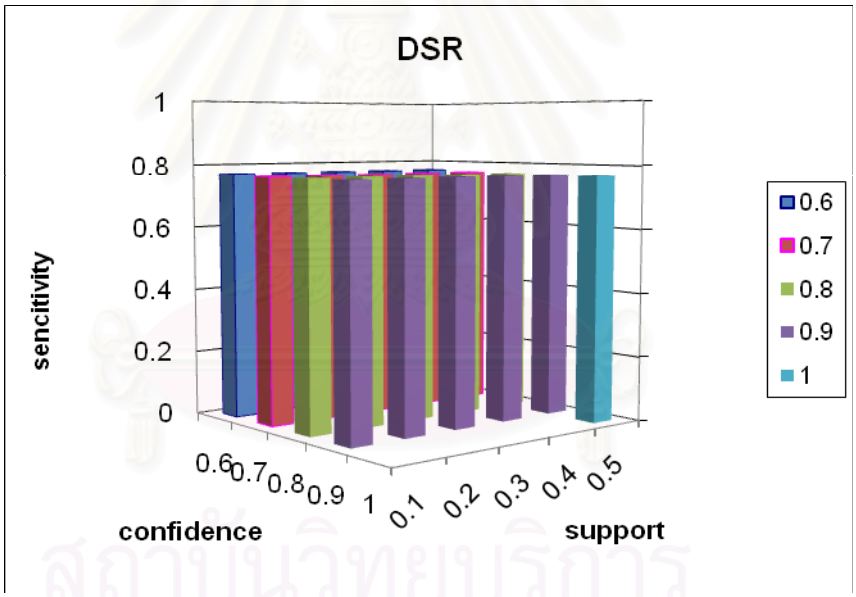
รูปที่ 4.13 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล mushroom สำหรับขั้นตอนวิธี WS



รูปที่ 4.14 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล mushroom สำหรับขั้นตอนวิธี DSRCCWS



รูปที่ 4.15 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล mushroom สำหรับขั้นตอนวิธี WS

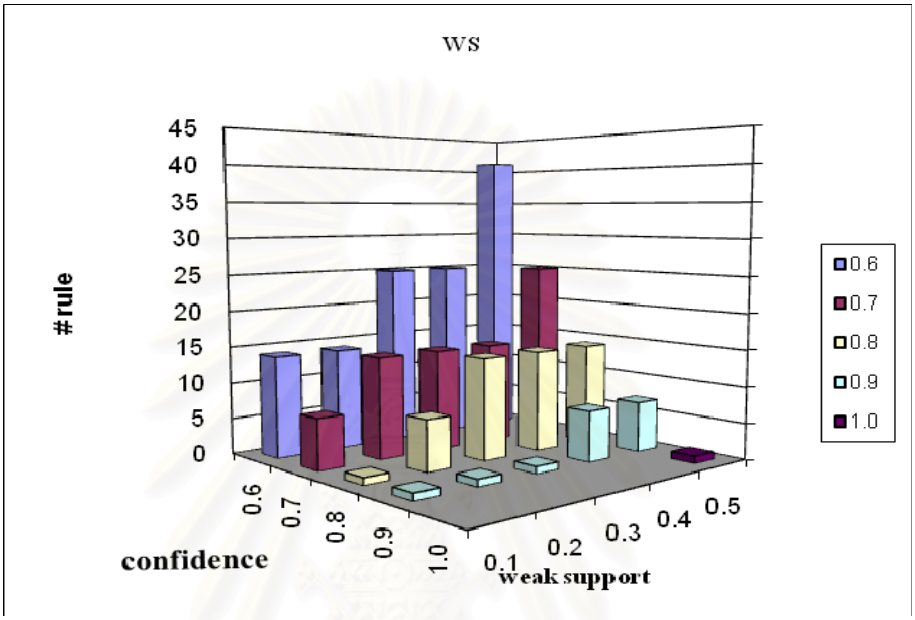


รูปที่ 4.16 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล mushroom สำหรับขั้นตอนวิธี DSRCCWS

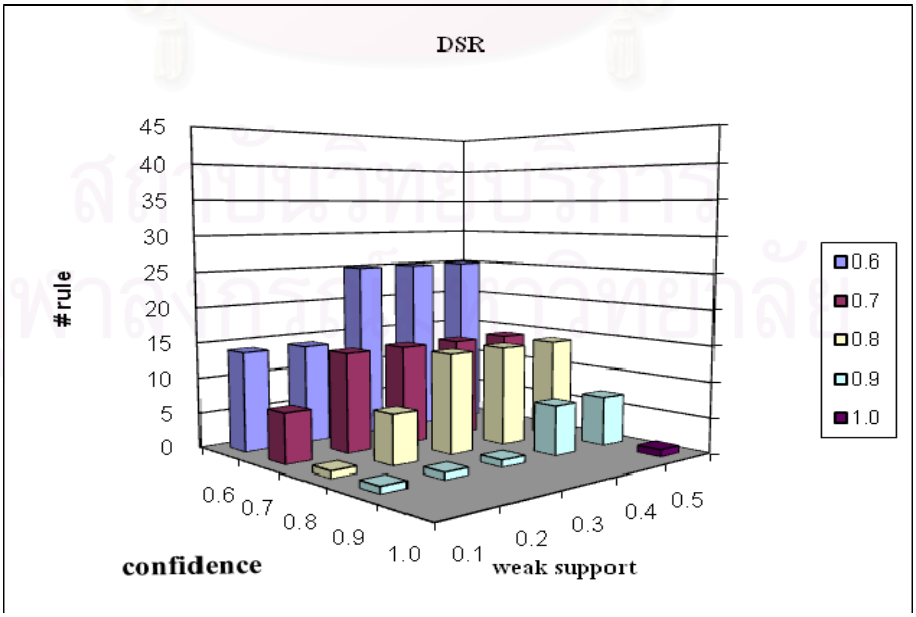
จากรูปที่ 4.13 และ รูปที่ 4.14 เป็นข้อมูลของจำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล mushroom สำหรับขั้นตอนวิธี WS และ ขั้นตอนวิธี DSRCCWS ตามลำดับ จะเห็นว่าจำนวนหลักเกณฑ์เชื่อมโยงจากขั้นตอนวิธี WS จะมีปริมาณหลักเกณฑ์ที่มากกว่าจำนวนหลักเกณฑ์จากขั้นตอนวิธี DSRCCWS ซึ่งถ้าคิดเป็นสัดส่วนแล้วจะมีจำนวนมากกว่าประมาณ 88 เปอร์เซ็นต์ โดยที่ในส่วนของค่าสภาพไวจากรูปที่ 4.15 และรูปที่ 4.16 มีค่าสภาพไวที่เท่ากัน

จากทั้งขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS แต่จะพบว่าเมื่อค่าความเชื่อมั่นน้อยสุดมีการเปรียบค่าจาก 0.6 เป็น 0.7 นั้นกราฟจะมีการลดลงของทั้งจำนวนหลักเกณฑ์เชื่อมโยงและค่าสภาพไว แต่มีการเปลี่ยนแปลงไม่มาก

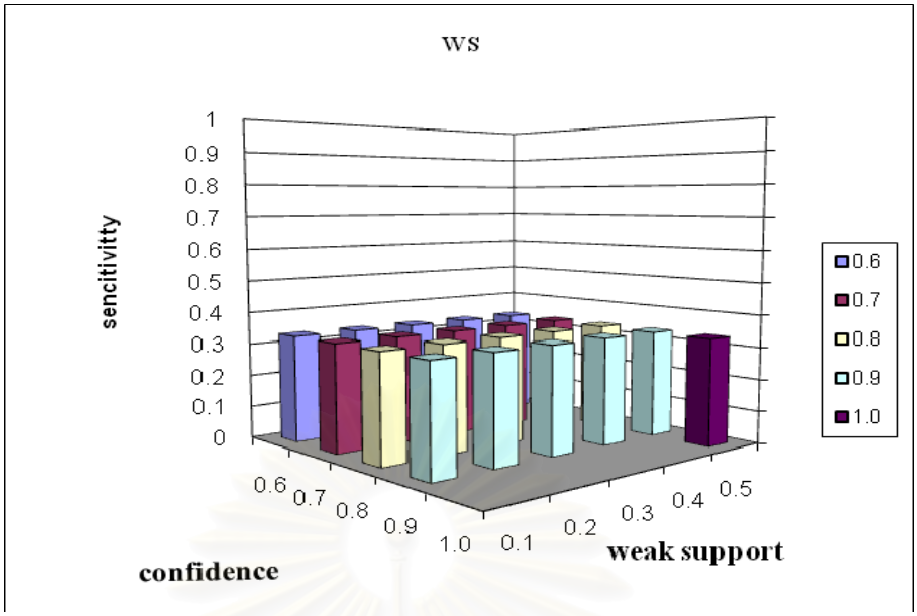
รูปที่ 4.17 ถึง รูปที่ 4.20 แสดงผลการทดลองจากข้อมูล zoo



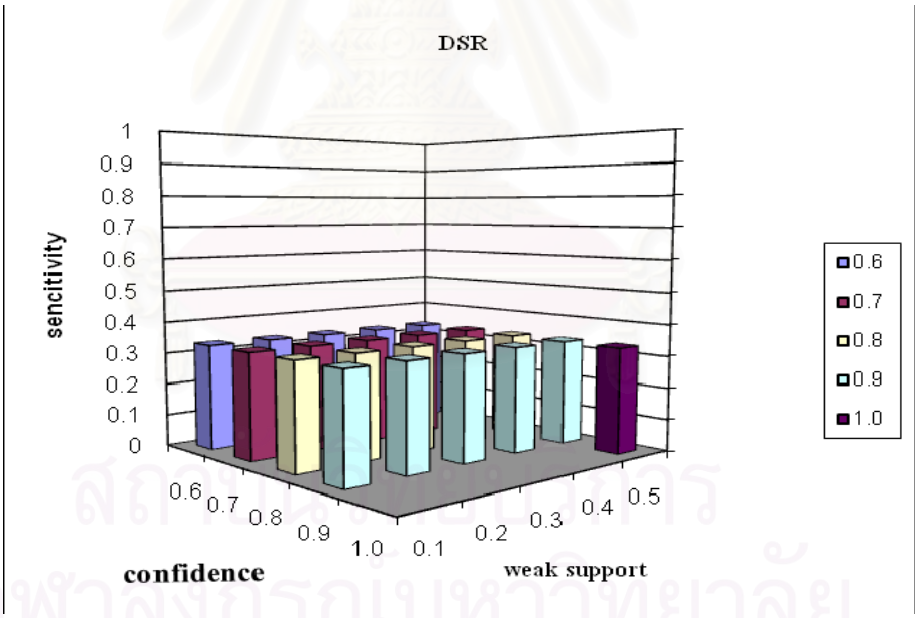
รูปที่ 4.17 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล zoo สำหรับขั้นตอนวิธี WS



รูปที่ 4.18 จำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล zoo สำหรับขั้นตอนวิธี DSRCCWS



รูปที่ 4.19 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล zoo สำหรับขั้นตอนวิธี WS



รูปที่ 4.20 ค่าสภาพไวของหลักเกณฑ์เชื่อมโยงข้อมูลของข้อมูล zoo สำหรับขั้นตอนวิธี DSRCCWS

จากรูปที่ 4.17 และ รูปที่ 4.18 เป็นข้อมูลของจำนวนหลักเกณฑ์เชื่อมโยงของข้อมูล zoo สำหรับขั้นตอนวิธี WS และ ขั้นตอนวิธี DSRCCWS ตามลำดับ จะเห็นว่าจำนวน

หลักเกณฑ์เชื่อมโยงจากขั้นตอนวิธี WS จะมีปริมาณหลักเกณฑ์ที่มากกว่าจำนวนหลักเกณฑ์จากขั้นตอนวิธี DSRCCWS แค่บางค่า โดยจะมีจำนวนหลักเกณฑ์เชื่อมโยงมากกว่าเมื่อค่าสนับสนุนแบบอ่อนน้อยสุดมีค่าน้อยกว่า 0.8 โดยที่ในส่วนของค่าสภาพไวจากรูปที่ 4.19 และรูปที่ 4.20 มีค่าสภาพไวที่เท่ากันจากทั้งขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS

จากผลการทดลองข้างต้น จะเห็นว่าจำนวนหลักเกณฑ์เชื่อมโยงที่ผ่านขั้นตอนวิธี DSRCCWS มีจำนวนหลักเกณฑ์ที่น้อยกว่าขั้นตอนวิธี WS แต่ค่าสภาพไวของทั้งขั้นตอนวิธี DSRCCWS และขั้นตอนวิธี WS นั้นมีค่าที่ใกล้เคียงกันมาก

จากผลการทดลอง พิจารณาค่าความเชื่อมั่นน้อยสุดและค่าสนับสนุนแบบอ่อนน้อยสุดที่ค่ามากเช่นค่าความเชื่อมั่นน้อยสุดเท่ากับ 1 และค่าสนับสนุนแบบอ่อนมีค่าเท่ากับ 1 เมื่อกำหนดค่าตัววัดมากแล้วก็จะพบว่าจำนวนหลักเกณฑ์เชื่อมโยงที่ได้จะมีปริมาณน้อย แต่หลักเกณฑ์ที่ผ่านขั้นตอนวิธีออกมาอาจจะมีปริมาณที่น้อยจนเกินไป สำหรับที่จะนำมาวิเคราะห์ข้อมูลทั้งหมด ดังนั้นจึงจะต้องมีการกำหนดค่าความเชื่อมั่นน้อยสุดและค่าสนับสนุนแบบอ่อนน้อยสุดให้มีค่าน้อยลง ซึ่งจะทำให้หลักเกณฑ์เชื่อมโยงที่ออกมามีปริมาณมากไม่สะดวกต่อผู้ใช้ ดังนั้นผู้วิจัยจึงได้สนใจการลดจำนวนหลักเกณฑ์เชื่อมโยงเพื่อให้ผู้ใช้สามารถวิเคราะห์ได้สะดวกยิ่งขึ้น

ขั้นตอนวิธี DSRCCWS สามารถลดจำนวนหลักเกณฑ์เชื่อมโยงลงได้จากขั้นตอนวิธี WS โดยหลักเกณฑ์ที่ได้จากขั้นตอนวิธี DSRCCWS นั้นยังคงเป็นสับเซตของหลักเกณฑ์เชื่อมโยงที่ได้มาจากขั้นตอนวิธี WS กล่าวคือหลักเกณฑ์จากขั้นตอนวิธี DSRCCWS เป็นหลักเกณฑ์ที่เกิดมาจากขั้นตอนวิธี WS แต่มีการตัดหลักเกณฑ์ที่ไม่น่าสนใจออกไปโดยใช้การทดสอบความเป็นอิสระต่อกันหรือการทดสอบค่าไคกำลังสอง และมีการพิจารณาหลักเกณฑ์ที่สามารถอธิบายข้อมูลโดยรวมเรียกหลักเกณฑ์นี้ว่า DS rule แล้วทำการคัดหลักเกณฑ์ที่ให้ข้อมูลที่ละเอียดเกินไปออกเก็บไว้เป็นหลักเกณฑ์ที่จะนำมาดูเพิ่มเติมได้ในการวิเคราะห์ข้อมูล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลงานวิจัย และข้อเสนอแนะ

ในงานวิจัยนี้ ผู้วิจัยมีความสนใจศึกษาถึงขั้นตอนวิธีที่จะนำมาช่วยในการลดปริมาณหลักเกณฑ์ จากการคัดเลือกหลักเกณฑ์เชื่อมโยงโดยใช้ค่าสนับสนุนแบบอ่อนน้อยสุดกับค่าความเชื่อมั่นน้อยสุด แล้วจะมีการกำหนดค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยที่สุด และนำแนวคิดการหาหลักเกณฑ์กำหนดทิศทางมาช่วยในการคัดเลือกหลักเกณฑ์ให้มีจำนวนลดลง แล้วใช้ค่าสภาพไว (Sensitivity) มาประเมินความแตกต่างระหว่างกลุ่มของหลักเกณฑ์ที่ได้ ระหว่างขั้นตอนวิธี WS ที่ใช้ค่าสนับสนุนแบบอ่อนน้อยสุดกับค่าความเชื่อมั่นน้อยสุด กับขั้นตอนวิธี DSRCCWS ที่ใช้ค่าสนับสนุนแบบอ่อนน้อยสุด ค่าความเชื่อมั่นน้อยสุด ค่าความเชื่อมั่นประพจน์แย้งสลับที่น้อยสุด และแนวคิดการหาหลักเกณฑ์กำหนดทิศทาง

การนำค่าความเชื่อมั่นประพจน์แย้งสลับที่ และแนวคิดการหาหลักเกณฑ์กำหนดทิศทางมาช่วยในการลดหลักเกณฑ์เชื่อมโยงที่ใช้ตัววัดในการสร้างหลักเกณฑ์คือ ค่าสนับสนุนแบบอ่อนน้อยสุดกับค่าความเชื่อมั่นน้อยสุดนั้นทำให้ได้หลักเกณฑ์ที่มีปริมาณลดลง ดังนั้นจึงช่วยให้ทำการวิเคราะห์หลักเกณฑ์เชื่อมโยงได้ง่ายขึ้น และเมื่อทำการประเมินผลการทดลองก็ให้ค่าสภาพไวที่ไม่ด้อยกว่าการไม่ได้ใช้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ กับแนวคิดการหาหลักเกณฑ์กำหนดทิศทางเข้ามาช่วย

เนื่องจากถ้าทำการกำหนดค่าสนับสนุนแบบอ่อนน้อยสุด กับค่าความเชื่อมั่นน้อยสุดให้มีค่าสูงก็สามารถลดปริมาณหลักเกณฑ์เชื่อมโยงให้น้อยลงได้ แต่ในการกำหนดค่าสนับสนุนแบบอ่อนน้อยสุดกับค่าความเชื่อมั่นน้อยสุดให้มีค่าที่เหมาะสม เพื่อให้ได้หลักเกณฑ์เชื่อมโยงที่น่าสนใจออกมานั้น จำเป็นที่จะต้องมีความเข้าใจในข้อมูลและต้องประกอบกับความสามารถในการวิเคราะห์อย่างมาก ซึ่งก็เป็นปัญหาในการทำงานของผู้ใช้

อนึ่งเนื่องจากขั้นตอนวิธีเอไพออรี (Apriori algorithm) ที่ใช้ตัววัดค่าสนับสนุนกับค่าความเชื่อมั่นในการคัดเลือกหลักเกณฑ์เชื่อมโยง ซึ่งถ้าทำการกำหนดค่าสนับสนุนให้มีค่าต่ำ ๆ ก็จะทำให้จำนวนหลักเกณฑ์ที่ได้มีปริมาณมาก เช่นเดียวกันกับขั้นตอนวิธี WS ที่ใช้ค่าสนับสนุนแบบอ่อนกับค่าความเชื่อมั่นในการคัดเลือกหลักเกณฑ์เชื่อมโยง แต่งานวิจัยนี้สนใจและเลือกใช้ขั้นตอนวิธี WS นั้นเพราะว่าหลักเกณฑ์เชื่อมโยงที่ได้จากการคัดเลือกนั้นจะมีค่าความเชื่อมั่นที่สูงจากทฤษฎีในงานวิจัยของชรีย์วรรณ [13]

เพราะฉะนั้นงานวิจัยนี้จึงมาช่วยแก้ปัญหาจำนวนหลักเกณฑ์ที่ออกมามาก โดยการลดปริมาณหลักเกณฑ์เชื่อมโยงให้มีจำนวนน้อยลงเพื่อสะดวกในการวิเคราะห์ข้อมูล โดยไม่ต้องใช้ความชำนาญเฉพาะตัวในการกำหนดค่าสนับสนุนแบบอ่อนน้อยสุด กับค่าความเชื่อมั่นน้อยสุด

จากการทดลองที่ผ่านมาจะเห็นว่าเมื่อค่าสนับสนุนแบบอ่อนกับค่าความเชื่อมั่นมีค่าเป็น 1 หลักเกณฑ์เชื่อมโยงที่ผ่านออกมาจะมีจำนวนหลักเกณฑ์ที่น้อย และมีค่าของตัววัดที่สูงอยู่แล้วจึงไม่ถูกตัดออกเมื่อผ่านขั้นตอนวิธี DSRCCWS แต่หลักเกณฑ์ที่ผ่านออกมาก็มีจำนวนน้อยไป สำหรับการนำข้อมูลมาวิเคราะห์ต่อไป ดังนั้นสำหรับงานวิจัยนี้จะมีความเหมาะสมกับช่วงค่าสนับสนุนแบบอ่อนกับค่าความเชื่อมั่นที่มีค่าน้อยกว่า 1

จากงานวิจัยนี้ไม่ได้ทำการเปรียบเทียบเวลาที่ใช้ในการทำการทดลอง เนื่องจากเป็นการนำขั้นตอนวิธี WS มาทำการปรับปรุงโดยการเพิ่มเงื่อนไขค่าความเชื่อมั่นประพจน์แย้งสลับที่และเพิ่มขั้นตอนวิธีในการหาหลักเกณฑ์กำหนดทิศทางเข้าไปในขั้นตอนวิธี ดังนั้นเวลาที่ใช้ในการทำงานจะมากกว่าเวลาที่ใช้ของขั้นตอนวิธี WS และเนื่องจากผลการทดลองได้ค่าสภาพไวของข้อมูลมีค่าที่น้อยกว่า 0.5 ดังนั้นงานในอนาคตอาจจะทำการแก้ไขขั้นตอนวิธีให้ใช้เวลาในการทำงานที่น้อยกว่า และทำการหาตัววัดตัวใหม่ที่จะนำมาเปรียบเทียบผลการทดลอง หรืออาจจะมีการเปลี่ยนวิธีการที่จะช่วยลดจำนวนหลักเกณฑ์เชื่อมโยงให้มีจำนวนหลักเกณฑ์ที่น้อยลงไปกว่าขั้นตอนวิธีของงานวิจัยนี้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

[1] Han, J., and Kamber, M. *Data mining : Concepts and techniques*. San Fransisco: Morgan Kaufmann Publishers, 2001.

[2] Agrawal, R., Imielinski, T. and Swami, A. Mining association rules between sets of items in large databases. *Proceeding of the ACM SIGMOD International Conference on Data* (1993) : 207-216.

[3] Agrawal, R. and Srikant, R. Fast algorithms for mining association rules. *Proceeding of the 20th International Conference on Very Large Data* (1994) : 487-499.

[4] Han, J., Pei, J., and Yin, Y. Mining frequent patterns without candidate generation. *In 2000 ACM SIGMOD International Conference on Management of Data* (2000) : 1-12.

[5] Zhang, C., and Zhang, S. *Association rule mining: Models and algorithms*. Berlin: Springer, 2002.

[6] Brin, S., Motwani, R., and Silverstein, C. Beyond market baskets: Generalizing association rules to correlations. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (1997) : 265-276.

[7] Fayyad, U., Shapiro, G., P., and Smyth, P. The kdd process for extraction useful knowledge from volumes data. *Communications of the ACM* 11 (November 1996): 27-31.

[8] W. Frawley and G. Piatetsky-Shapiro and C. Matheus Knowledge Discovery in Databases: An Overview. *AI Magazine* (1992): 213-228.

[9] Wikipeida. *Data warehouse* [online] . (n.d.). Available from : http://en. wikipedia.org/ wiki/Data warehouse , [2 February 2008].

[10] Hahsler, M. A comparison of commonly used interest measures for association *rules* [online]. Available from : http://wwwai.wu-wien.ac.at/~hahsler/research/association_rules [19 July 2007].

[11] Webb, G., I., and Zhang, S. K-optimal rule discovery. *Data Mining and Knowledge Discovery* (2005): 39-79.

[12] Pasquier, N., Bastide, Y., Taouil, R., and Lakhal, L. Efficient mining of associaiton rules using closed itemset lattices. *Information Systems* (1999): 25-46.

[13] Sinapiromsaran, K., and Sirisrisumrit, C. *Association* rule extraction based on the weak support. *The Second Graduate Congress of Mathematics and Physical Science*, 2006.

[14] Anantaporn Sritsawat and Krung Sinapiromsaran. *Effect of the Contrapositive Confidence for the Association Rule Selection*. Master thesis, Faculty of Science, Chulalongkorn University, 2002 ISBN 974-17-1841-1

[15] Liu, B., Hsu, W. and Ma, Y. 1999. Pruning and Summarizing the Discovered Associations. *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (KDD-99)

[16] Beck, J.R. and Schultz, E.K.1986. The use of ROC curves in test performance evaluation. *Arch Pathol Lab Med* (1986) :13-20.

[17] Blake, C.L. and Merz, C.L. *UCI repository of machine learning databases*. [online]. Available from : www.ics.uci.edu/~mlearn/MLRepository.html, [10 October 2006].

[18] FIMI. *Frequent Itemset Mining Dataset Repository* [Online]. (2003-2004). Available from : http://fimi.cs.helsinki.fi/data/ [10 November 2006].

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

ผลการทดลอง

การทดลองนี้จะปรับเปลี่ยนค่าสนับสนุนน้อยสุด จาก 0.1 ไปถึง 0.5 และค่าความเชื่อมั่นน้อยสุด จาก 0.6 ไปถึง 1.0 แล้วทำการคำนวณค่าสนับสนุนแบบอ่อนน้อยสุดดังตารางที่ 2.2 ในบทที่ 2 แล้วนำข้อมูลในส่วนของข้อมูลทดลองมาทำการสร้างหลักเกณฑ์เชื่อมโยง โดยใช้ขั้นตอนวิธี WS และขั้นตอนวิธี DSRCCWS แล้วนำข้อมูลในส่วนข้อมูลทดสอบมาทำการเปรียบเทียบให้ได้ค่าสภาพไวออกมา ผลการทดลองที่ได้จะทำการเปรียบเทียบเวลาในการทดลองเป็นวินาที จำนวนหลักเกณฑ์เชื่อมโยง และค่าสภาพไว โดยผลการทดลองจะแสดงในรูปแบบตารางดังต่อไปนี้

ตารางที่ ก.1 เป็นผลการทดลองจากข้อมูล breast-cancer

			ขั้นตอนวิธี WS			ขั้นตอนวิธี DSRCCWS		
minC	*minS*	*minWS*	เวลา	#หลักเกณฑ์	สภาพไว	เวลา	#หลักเกณฑ์	สภาพไว
	0.1	0.93	3.672	1246	0.65330	3.152	508	0.65043
	0.2	0.87	3.399	1246	0.65330	3.168	508	0.65043
	0.3	0.8	3.657	1246	0.65330	3.184	508	0.65043
	0.4	0.73	3.223	1028	0.65330	3.612	508	0.65043
0.6	0.5	0.67	4.124	851	0.65330	3.958	378	0.65043
	0.1	0.96	2.576	861	0.65330	2.285	346	0.65043
	0.2	0.91	2.171	696	0.65330	2.323	346	0.65043
	0.3	0.87	2.199	696	0.65330	2.33	346	0.65043
	0.4	0.83	2.181	696	0.65330	2.353	346	0.65043
0.7	0.5	0.79	1.429	373	0.64756	1.503	140	0.64183
	0.1	0.98	1.565	447	0.65043	1.484	143	0.64470
	0.2	0.95	1.547	447	0.65043	1.456	143	0.64470
	0.3	0.93	1.57	447	0.65043	1.445	143	0.64470
	0.4	0.9	1.103	272	0.64470	1.143	92	0.64183
0.8	0.5	0.88	0.929	209	0.64470	0.932	61	0.64183
	0.1	0.99	0.576	104	0.63897	0.463	21	0.63610
	0.2	0.98	0.578	104	0.63897	0.488	21	0.63610
	0.3	0.97	0.496	77	0.63610	0.507	21	0.63610
	0.4	0.96	0.349	50	0.63037	0.395	15	0.63037
0.9	0.5	0.94	0.229	25	0.61891	0.235	8	0.61891
1	0.1	1	0.147	3	0.52436	0.163	2	0.52436

ตารางที่ ก.2 เป็นผลการทดลองจากข้อมูล chess

			ขั้นตอนวิธี WS			ขั้นตอนวิธี DSRCCWS		
minC	*minS*	*minWS*	เวลา	#หลักเกณฑ์	สภาพไว	เวลา	#หลักเกณฑ์	สภาพไว
	0.1	0.93	0.508	301	0.478723	0.743	83	0.47622
	0.2	0.87	0.653	301	0.478723	0.755	83	0.47622
	0.3	0.8	0.546	223	0.478723	0.731	83	0.47622
	0.4	0.73	0.358	111	0.47622	0.1047	83	0.47622
0.6	0.5	0.67	0.278	57	0.466834	0.619	44	0.466834
	0.1	0.96	0.198	62	0.478723	0.255	15	0.433041
	0.2	0.91	0.157	21	0.433041	0.258	15	0.433041
	0.3	0.87	0.154	21	0.433041	0.247	15	0.433041
	0.4	0.83	0.122	21	0.433041	0.247	15	0.433041
0.7	0.5	0.79	0.107	8	0.377347	0.168	6	0.377347
	0.1	0.98	0.134	5	0.374844	0.126	4	0.374844
	0.2	0.95	0.129	5	0.374844	0.123	4	0.374844
	0.3	0.93	0.105	5	0.374844	0.166	4	0.374844
	0.4	0.9	0.13	5	0.374844	0.167	4	0.374844
0.8	0.5	0.88	0.103	5	0.374844	0.119	4	0.374844
	0.1	0.99	0.113	2	0.308511	0.145	2	0.308511
	0.2	0.98	0.117	2	0.308511	0.149	2	0.308511
	0.3	0.97	0.118	2	0.308511	0.119	2	0.308511
	0.4	0.96	0.118	2	0.308511	0.15	2	0.308511
0.9	0.5	0.94	0.099	2	0.308511	0.111	2	0.308511
1	0.1	1	0.116	1	0.239049	0.12	1	0.239049

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก.3 เป็นผลการทดลองจากข้อมูล connect

minC	*minS*	*minWS*	ขั้นตอนวิธี WS			ขั้นตอนวิธี DSRCCWS		
			เวลา	#หลักเกณฑ์	สภาพไว	เวลา	#หลักเกณฑ์	สภาพไว
	0.1	0.93	67.969	1919	0.68394	135.36	687	0.68394
	0.2	0.87	26.53	1267	0.68394	73.23	687	0.68394
	0.3	0.8	19.541	1023	0.68394	73.6	687	0.68394
	0.4	0.73	35.452	846	0.68394	69.64	663	0.68394
0.6	0.5	0.67	42.332	476	0.68127	41.56	398	0.67852
	0.1	0.96	1.263	281	0.68394	20.212	240	0.67556
	0.2	0.91	1.248	281	0.68394	50.954	240	0.67556
	0.3	0.87	1.261	281	0.68394	2.759	240	0.67556
	0.4	0.83	1.118	253	0.67568	2.764	240	0.67556
0.7	0.5	0.79	0.684	110	0.61309	2.759	103	0.61291
	0.1	0.98	0.25	20	0.51051	0.275	20	0.51051
	0.2	0.95	0.325	20	0.51051	0.317	20	0.51051
	0.3	0.93	0.315	20	0.51051	0.35	20	0.51051
	0.4	0.9	0.321	20	0.51051	0.389	20	0.51051
0.8	0.5	0.88	0.296	20	0.51051	0.299	20	0.51051
	0.1	0.99	0.175	0	0	0.179	0	0
	0.2	0.98	0.177	0	0	0.181	0	0
	0.3	0.97	0.179	0	0	0.179	0	0
	0.4	0.96	0.169	0	0	0.176	0	0
0.9	0.5	0.94	0.149	0	0	0.175	0	0
1	0.1	1	0.168	0	0	0.18	0	0

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก.4 เป็นผลการทดลองจากข้อมูล mushroom

minC	*minS*	*minWS*	ขั้นตอนวิธี WS			ขั้นตอนวิธี DSRCCWS		
			เวลา	#หลักเกณฑ์	สภาพไว	เวลา	#หลักเกณฑ์	สภาพไว
	0.1	0.93	0.586	134	0.771049	0.688	15	0.771049
	0.2	0.87	0.301	134	0.771049	0.697	15	0.771049
	0.3	0.8	0.308	134	0.771049	0.712	15	0.771049
	0.4	0.73	0.289	134	0.771049	0.71	15	0.771049
0.6	0.5	0.67	0.307	134	0.771049	0.968	15	0.771049
	0.1	0.96	0.275	123	0.76711	0.657	14	0.76711
	0.2	0.91	0.294	123	0.76711	0.652	14	0.76711
	0.3	0.87	0.303	123	0.76711	0.665	14	0.76711
	0.4	0.83	0.3	123	0.76711	0.66	14	0.76711
0.7	0.5	0.79	0.303	123	0.76711	0.664	14	0.76711
	0.1	0.98	0.302	123	0.76711	0.662	14	0.76711
	0.2	0.95	0.283	123	0.76711	0.66	14	0.76711
	0.3	0.93	0.32	123	0.76711	0.656	14	0.76711
	0.4	0.9	0.261	123	0.76711	0.66	14	0.76711
0.8	0.5	0.88	0.281	123	0.76711	0.612	14	0.76711
	0.1	0.99	0.253	123	0.76711	0.661	14	0.76711
	0.2	0.98	0.291	123	0.76711	0.67	14	0.76711
	0.3	0.97	0.0284	123	0.76711	0.651	14	0.76711
	0.4	0.96	0.265	123	0.76711	0.647	14	0.76711
0.9	0.5	0.94	0.308	123	0.76711	0.649	14	0.76711
1	0.1	1	0.275	123	0.76711	0.614	14	0.76711

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ก.5 เป็นผลการทดลองจากข้อมูล zoo

minC	*minS*	*minWS*	ขั้นตอนวิธี WS			ขั้นตอนวิธี DSRCCWS		
			เวลา	#หลักเกณฑ์	สภาพไว	เวลา	#หลักเกณฑ์	สภาพไว
	0.1	0.93	0.116	41	0.333	0.233	25	0.333
	0.2	0.87	0.093	25	0.333	0.227	25	0.333
	0.3	0.8	0.098	25	0.333	0.243	25	0.333
	0.4	0.73	0.101	14	0.333	0.137	14	0.333
0.6	0.5	0.67	0.088	14	0.333	0.189	14	0.333
	0.1	0.96	0.114	25	0.333	0.189	14	0.333
	0.2	0.91	0.096	14	0.333	0.182	14	0.333
	0.3	0.87	0.087	14	0.333	0.183	14	0.333
	0.4	0.83	0.097	14	0.333	0.183	14	0.333
0.7	0.5	0.79	0.079	7	0.333	0.156	7	0.333
	0.1	0.98	0.099	14	0.333	0.14	14	0.333
	0.2	0.95	0.085	14	0.333	0.182	14	0.333
	0.3	0.93	0.078	14	0.333	0.185	14	0.333
	0.4	0.9	0.088	7	0.333	0.155	7	0.333
0.8	0.5	0.88	0.075	1	0.333	0.079	1	0.333
	0.1	0.99	0.09	7	0.333	0.108	7	0.333
	0.2	0.98	0.091	7	0.333	0.155	7	0.333
	0.3	0.97	0.083	1	0.333	0.079	1	0.333
	0.4	0.96	0.078	1	0.333	0.079	1	0.333
0.9	0.5	0.94	0.068	1	0.333	0.079	1	0.333
1	0.1	1	0.067	1	0.333	0.083	1	0.333

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวธนาพร ธัญญะเศรษฐ์ เกิดเมื่อวันที่ 21 สิงหาคม พุทธศักราช 2526 สำเร็จการศึกษาระดับปริญญาวิทยาศาสตร์บัณฑิต สาขาคณิตศาสตร์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น เมื่อปีการศึกษา 2547 และเข้าศึกษาต่อในหลักสูตรปริญญามหาบัณฑิต สาขาวิทยาการคณนา ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2548

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย