การเพิ่มค่าความเชื่อมั่นประพจน์แย้งสลับที่ในการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน

นางสาวอนันตพร ศรีสวัสดิ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคณนา ภาควิชาคณิตศาสตร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2545

ISBN 974-17-1841-1

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

ADDING CONTRAPOSITIVE PROPOSITION CONFIDENCE IN THE ASSOCIATION RULE SELECTION

Miss Anantaporn Srisawat

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computational Science

Department of Mathematics
Faculty of Science
Chulalongkorn University

Academic Year 2002

ISBN 974-17-1841-1

หัวข้อวิทยานิพนธ์	การเพิ่มค่าความเชื่อมั่นประพจน์แย้งสลับที่ในการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน
โดย	นางสาว อนันตพร ศรีสวัสดิ์
สาขาวิชา	วิทยาการคณนา
อาจารย์ที่ปรึกษา	อาจารย์ ดร. กรุง สินอภิรมย์สราญ

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิทยาศาสตร์
(รองศาสตราจารย์ ดร.วันชัย โพธิ์พิจิตร)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.ชิดชนก เหลือสินทรัพย์)

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร. กรุง สินอภิรมย์สราญ)

..... กรรมการ
(อาจารย์ ดร. จารุโลจน์ จงสถิตย์วัฒนา)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

อนันตพร ศรีสวัสดิ์ : การเพิ่มค่าความเชื่อมั่นประพจน์แย้งสลับที่ในการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน. (ADDING CONTRAPOSITIVE PROPOSITION CONFIDENCE IN THE ASSOCIATION RULE SELECTION) อ. ที่ปรึกษา : อาจารย์ ดร. กรุง สินอภิรมย์สราญ, 83 หน้า. ISBN 974-17-1841-1.

งานวิจัยนี้ได้เสนอแนวความคิดในการคัดเลือกหลักเกณฑ์แอสโซซิเอชันจากหลักเกณฑ์ทั้งหมด ที่มีค่าความเชื่อมั่นเท่ากัน ซึ่งหลักเกณฑ์ที่นำมาพิจารณานั้นมีรูปแบบ $P \rightarrow Q$ โดยที่ P และ Q คือเซตของสินค้าที่มีสมาชิกเพียงตัวเดียว แนวคิดนี้ใช้หลักเกณฑ์คณิตตรรกศาสตร์และเซตมาประยุกต์โดยกำหนดค่าความเชื่อมั่นประพจน์แย้งสลับที่ขึ้นมาเพื่อใช้ในการคัดเลือกหลักเกณฑ์แอสโซซิเอชันที่มีค่าความเชื่อมั่นเท่ากัน และทำการคัดเลือกหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่ที่มากที่สุด ผู้วิจัยได้ทำการทดลองโดยทำการจำลองฐานข้อมูลขึ้นจากโปรแกรมจำลองข้อมูล จากการทดลองพบว่าในกรณีที่ค่าสนับสนุนประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ไม่เท่ากัน เมื่อค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์มีค่ามากขึ้นทำให้หลักเกณฑ์นั้นมีค่าความเชื่อมั่นประพจน์แย้งสลับที่และค่าความเชื่อมั่นมากขึ้นด้วย สุดท้ายผู้วิจัยได้ทำการวิเคราะห์กรณีที่หลายหลักเกณฑ์มีค่าความเชื่อมั่นเท่ากัน โดยพิจารณาแนวโน้มของค่าความเชื่อมั่นประพจน์แย้งสลับที่ในแต่ละกรณี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา **คณิตศาสตร์** ลายมือชื่อนิสิต.....
สาขาวิชา **วิทยาการคณนา** ลายมือชื่ออาจารย์ที่ปรึกษา.....
ปีการศึกษา 2545

4372472823 : MAJOR COMPUTATIONAL SCIENCE

KEY WORD: ASSOCIATION RULE / CONTRAPOSITIVE CONFIDENCE / SELECTION

ANANTAPORN SRISAWAT : ADDING CONTRAPOSITIVE PROPOSITION CONFIDENCE IN THE ASSOCIATION RULE SELECTION. THESIS ADVISOR : KRUNG SINAPIROMSARAN, Ph.D., [83] pp. ISBN 974-17-1841-1.

This thesis proposes a selection criteria for association rules with the same confidence. We concentrate on the association rule $P \longrightarrow Q$, where P and Q are sets of distinct single item. A new criteria value is called the contrapositive proposition confidence by computing the confidence of its contrapositive rule. In our experiments, we simulated data from our synthetic data generation program. Experiments with simulated data show that in the case of distinct supports of the contrapositive rules, the increasing support of the contrapositive rule causes both contrapositive proposition confidence and confidence value increase. Finally, we analyzed the cause for the equivalent confidence rules to identify the trend of contrapositive proposition confidence for all cases.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department **Mathematics** Student's signature.....

Field of study **Computational Science** Advisor's signature.....

Academic year **2002**

กิตติกรรมประกาศ

ในการทำวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าได้รับความอนุเคราะห์ และความช่วยเหลือจากบุคคลหลายท่าน จึงขอแสดงความขอบคุณไว้ ณ โอกาสนี้ ขอขอบคุณอาจารย์ ดร.กรุง สินอภิรมย์สราญ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้กรุณาถ่ายทอดความรู้และให้คำแนะนำที่เป็นประโยชน์ยิ่งทำให้วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยดี ขอขอบคุณศาสตราจารย์ ดร.ชิดชนก เหลือสินทรัพย์ ประธานกรรมการและอาจารย์ ดร.จารุโลจน์ จงสถิตย์วัฒนา กรรมการสอบวิทยานิพนธ์ ที่ได้ให้ข้อแนะนำอันเป็นประโยชน์ยิ่ง

สุดท้ายนี้ข้าพเจ้าขอขอบพระคุณครอบครัวของข้าพเจ้าที่เป็นกำลังใจและกำลังทรัพย์ให้ข้าพเจ้ามาโดยตลอด และขอขอบคุณเพื่อนๆ สาขาวิทยาการคณนาทุกคนที่เป็นกำลังใจและพี่ธวัช หรรษคุณาฒัย ที่ให้ความช่วยเหลือในด้าน hardware และ software ระหว่างการทำวิทยานิพนธ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ซ

สารบัญภาพ ..... ฌ

บทที่

1. บทนำ 1
2. การทำเหมืองข้อมูล 5
3. การคัดเลือกหลักเกณฑ์แอสโซซิเอชันด้วยค่าความเชื่อมั่นประพจน์แย้งสลับที่ 18
4. การทดลองและผลการทดลอง 31
5. การวิเคราะห์ค่าความเชื่อมั่นประพจน์แย้งสลับที่ 46
6. สรุปผล 64

รายการอ้างอิง 66

ภาคผนวก 68

ประวัติผู้เขียนวิทยานิพนธ์ 83

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

หน้า

ตารางที่ 2.1: เทคนิคต่างๆ ในการทำเหมืองข้อมูล 6
ตารางที่ 2.2: กลุ่มตัวอย่างของฐานข้อมูลการซื้อ 10
ตารางที่ 3.1: ตาราง Tran ในฐานข้อมูล 18
ตารางที่ 3.2: ค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ในกรณีต่างๆ 25
ตารางที่ 3.3: ค่าความจริงของ $(\forall x)(P(x) \rightarrow Q(x))$, $(\forall x)(Q(x) \rightarrow P(x))$ และ $(\forall x)(\neg Q(x) \rightarrow \neg P(x))$ 27
ตารางที่ 3.4: ตัวอย่างฐานข้อมูลการซื้อ 28
ตารางที่ 4.1: รูปแบบของข้อมูลในแต่ละกลุ่ม 34
ตารางที่ 5.1: เปรียบเทียบค่าต่างๆ ระหว่างหลักเกณฑ์ $p_k \rightarrow q_k$ และหลักเกณฑ์ $p_k' \rightarrow q_k'$. 60
ตารางที่ ก.1: ตารางแสดงชื่อ ชนิด และความหมายของตัวแปรที่สำคัญในโปรแกรมจำลองข้อมูล 71
ตารางที่ ก.2: โครงสร้างข้อมูลของตาราง data 72
ตารางที่ ก.3: ตารางแสดงชื่อ ชนิด และความหมายของตัวแปรที่สำคัญในโปรแกรมสร้างหลักเกณฑ์แอสโซซิเอชัน 72
ตารางที่ ก.4: ตารางฐานข้อมูลตัวอย่าง 74

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

หน้า

รูปที่ 2.1: กระบวนการทำงานของการทำเหมืองข้อมูล 5
รูปที่ 2.2: สินค้าทั้งหมดที่ลูกค้าซื้อในครั้งเดียวกัน 8
รูปที่ 2.3: แสดงขั้นตอนวิธีของการหา frequent itemset 13
รูปที่ 2.4: ตัวอย่างขั้นตอนการหา frequent itemset 14
รูปที่ 3.1: แสดงความสัมพันธ์ระหว่างเซต A และเซต B 19
รูปที่ 3.2: SQL query ในการหาสมาชิกของเซต T 20
รูปที่ 3.3: SQL query ในการหาสมาชิกของเซต A 20
รูปที่ 3.4: SQL query ในการหาสมาชิกของเซต B 20
รูปที่ 3.5: SQL query ในการหาสมาชิกของเซต A$\cap$B 21
รูปที่ 3.6: เงื่อนไขของสมาชิกในเซต S_1, S_2, S_3 และ S_4 22
รูปที่ 4.1: กระบวนการทำงานของโปรแกรมจำลองข้อมูล 33
รูปที่ 4.2: การกำหนดกลุ่มและรูปแบบของชุดข้อมูล 34
รูปที่ 4.3: ผลลัพธ์ที่ได้จากการทำงานของโปรแกรมของข้อมูล 1 ชุด 35
รูปที่ 4.4: แผนภูมิแสดงจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดของข้อมูลในกลุ่มที่ 1 36
รูปที่ 4.5: แผนภูมิแสดงจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดของข้อมูลในกลุ่มที่ 2 36
รูปที่ 4.6: แผนภูมิแสดงจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดของข้อมูลในกลุ่มที่ 3 37
รูปที่ 4.7: แผนภูมิแสดงจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดของข้อมูลในกลุ่มที่ 4 37
รูปที่ 4.8: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 1 40
รูปที่ 4.9: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 1 40

สารบัญภาพ (ต่อ)

หน้า

รูปที่ 4.10: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับ
ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 2 41
รูปที่ 4.11: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับ
ค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 2 41
รูปที่ 4.12: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับ
ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 3..... 42
รูปที่ 4.13: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับ
ค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 3 42
รูปที่ 4.14: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับ
ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 4 43
รูปที่ 4.15: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับ
ค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 4 43
รูปที่ 4.16: กราฟ 3 มิติแสดงความสัมพันธ์ระหว่างค่าความเชื่อมั่นและ
ค่าความเชื่อมั่นประพจน์แย้งสลับที่ 44
รูปที่ 5.1: กรณีที่ $x_2^{k'} > x_2^k$ 48
รูปที่ 5.2: กรณีที่ $x_2^{k'} < x_2^k$ 49
รูปที่ ก.1: แผนภาพแสดงขั้นตอนการทำงานของโปรแกรมจำลองข้อมูล 69
รูปที่ ก.2: แผนภาพแสดงขั้นตอนการทำงานของโปรแกรมจำลองข้อมูล (ต่อ) 70
รูปที่ ก.3: ผลลัพธ์ที่ได้จากโปรแกรม Create_rule.java 77

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

ในปัจจุบันธุรกิจการขายเติบโตอย่างรวดเร็วและมีการแข่งขันสูง บริษัทร้านค้าต่างใช้กลยุทธ์ในการจัดทำรายการพิเศษหรือเพิ่มยอดขายของสินค้าไม่ว่าจะเป็นสินค้าอุปโภคบริโภคหรือการบริการต่างๆ ตัวอย่างเช่น ในซุปเปอร์มาเก็ตมีการขายสินค้าชนิดหนึ่งและมีการแถมสินค้าอีกชนิดหนึ่ง เช่น ขายยาสีฟันพร้อมแถมแปรงสีฟัน ขายบะหมี่กึ่งสำเร็จรูปแถมชาม เป็นต้น ส่วนประเภทร้านขายอาหารจานด่วน เช่น KFC ใช้การขายเป็นชุดอาหารสุดประหยัดเพื่อให้ลูกค้าเลือกสั่งได้สะดวก ซึ่งการขายอาหารเป็นชุดนี้ทำกำไรให้กับทางร้านและทำให้ลูกค้าพึงพอใจในการเลือกซื้อ ในการวางแผนว่าจะให้อาหารแต่ละชุดมีอะไรบ้างทางร้านค้าจะต้องพิจารณาว่าลูกค้านิยมรับประทานอาหารใดร่วมกันบ้างโดยการดูจากยอดขายของอาหารแต่ละอย่างจากฐานข้อมูลการซื้อ

จากเหตุผลดังกล่าว จะเห็นว่าฐานข้อมูลการซื้อนั้นมีความสำคัญกับการวางแผนการขายสินค้า ผู้วิจัยจึงสนใจนำทฤษฎีทางด้านการทำเหมืองข้อมูล (data mining) [1,2] มาประยุกต์ใช้กับการตัดสินใจด้านธุรกิจที่กล่าวข้างต้น และในงานวิจัยนี้ผู้วิจัยใช้เทคนิคหลักเกณฑ์แอสโซซิเอชัน (association rule) [1] เพื่อหารูปแบบความสัมพันธ์ของข้อมูล เนื่องจากหลักเกณฑ์แอสโซซิเอชันอยู่ในรูปแบบที่ชัดเจนและเข้าใจง่าย โดยหลักเกณฑ์นี้แสดงถึงรูปแบบการซื้อสินค้าของลูกค้าว่าเมื่อลูกค้าซื้อสินค้าชนิดหนึ่งแล้วจะมีการซื้อสินค้าอีกชนิดหนึ่งพร้อมกันในการซื้อสินค้าครั้งเดียวกันหรือไม่ และผู้ใช้สามารถนำหลักเกณฑ์แอสโซซิเอชันที่ได้มาทำการวางแผนการขายเพื่อจัดทำรายการพิเศษหรือเพิ่มยอดขายตามที่ต้องการ

เทคนิคหลักเกณฑ์แอสโซซิเอชันนี้ได้มีการคิดค้นขึ้นครั้งแรกในปี 1993 โดย Agrawal, Imielinski และ Swami [3] หลังจากนั้นได้มีงานวิจัยเกี่ยวกับการพัฒนากระบวนการสร้างหลักเกณฑ์แอสโซซิเอชันหลายงานวิจัย [4,5,6,7,8,9] โดยทำการสร้างเซตสินค้าที่มีค่าสนับสนุนของเซตสินค้ามากกว่าหรือเท่ากับค่าที่กำหนดไว้ จากนั้นทำการสร้างหลักเกณฑ์จากเซตสินค้าเหล่านั้น โดยหลักเกณฑ์ที่จะสร้างนั้นต้องมีค่าความเชื่อมั่นของหลักเกณฑ์มากกว่าหรือเท่ากับค่าที่กำหนดไว้ ซึ่งค่าความเชื่อมั่นของหลักเกณฑ์คำนวณได้จากจำนวนรายการที่ปรากฏสินค้าในหลักเกณฑ์เท่านั้น และเนื่องจากในร้านค้าใหญ่ๆ เช่น บิ๊กซีหรือโลตัสมีสินค้าจำหน่ายอยู่หลายหมื่นชนิดทำให้หลักเกณฑ์แอสโซซิเอชันโซซิเอชัน $P \rightarrow Q$ โดยที่ P และ Q เป็นเซตที่มี

สมาชิกเพียง 1 ตัวมีจำนวนมากและอาจมีได้สูงสุดถึง n×(n-1) หลักเกณฑ์ เมื่อ n เป็นจำนวนชนิดสินค้าทั้งหมด ถ้าจำนวนสินค้าทั้งหมดมี 10,000 ชนิดหรือ n=10,000 จะได้จำนวนหลักเกณฑ์ทั้งหมดที่เป็นไปได้เท่ากับ 10,000×9,999 ซึ่งเท่ากับ 99,990,000 หลักเกณฑ์ ดังนั้นจึงมีโอกาสสูงที่หลักเกณฑ์จะมีค่าความเชื่อมั่นเท่ากันเกิดขึ้นเป็นจำนวนมาก ซึ่งทำให้ผู้ใช้เกิดความไม่แน่ใจว่าหลักเกณฑ์ใดที่นำมาใช้แล้วจะให้ผลตามที่ผู้ใช้คาดหวังไว้

จากปัญหาที่กล่าวมานี้ ผู้วิจัยจึงได้เสนอแนวความคิดในการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน จากหลักเกณฑ์ทั้งหมดที่มีค่าความเชื่อมั่นเท่ากันที่ได้ทำการสร้างและหาค่าความเชื่อมั่นไว้แล้วจากกระบวนการสร้างหลักเกณฑ์แอสโซซิเอชันเช่นเดียวกับงานวิจัย [4,5,6,7,8,9] โดยทำการสร้างค่าที่ใช้ในการคัดเลือกหลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากันเรียกว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ [10] ซึ่งค่าความเชื่อมั่นประพจน์แย้งสลับที่ใช้ข้อมูลในส่วนของรายการที่ไม่ปรากฏสินค้าในหลักเกณฑ์นั้นและทำการคัดเลือกหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่ที่มากที่สุด

หลักเกณฑ์แอสโซซิเอชัน $P \rightarrow Q$ ที่นำมาพิจารณาในการคัดเลือกจะพิจารณาเฉพาะหลักเกณฑ์ที่เซต P และ Q มีสมาชิกเพียงเซตละ 1 ตัว และทำการวิเคราะห์ความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ โดยทำการทดลองจากฐานข้อมูลที่ได้จากการเขียนโปรแกรมจำลองข้อมูล วิเคราะห์ความสัมพันธ์ระหว่างฐานข้อมูลกับค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่และวิเคราะห์ข้อมูลที่สามารถทำการคัดเลือกหลักเกณฑ์โดยใช้ค่าความเชื่อมั่นประพจน์แย้งสลับที่เป็นตัวคัดเลือกหลักเกณฑ์

เนื่องจากงานวิจัยนี้เล็งเห็นถึงประโยชน์ของหลักเกณฑ์แอสโซซิเอชันที่นำมาใช้ในอุตสาหกรรมการขายปลีก ดังนั้นข้อมูลที่งานวิจัยนี้สนใจคือข้อมูลการซื้อขายในร้านค้าปลีกที่เป็นลักษณะแบบ basket data [4] ซึ่งเป็นข้อมูลการซื้อสินค้าของลูกค้าในแต่ละครั้ง และนำหลักเกณฑ์ที่ได้มาทำการวางแผนการขายสินค้าและจัดทำรายการพิเศษ (promotion) เพื่อให้เป็นที่ถูกใจและดึงดูดลูกค้าให้มาซื้อสินค้ามากขึ้น ในการที่จะนำข้อมูลเหล่านี้มาวิเคราะห์ความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ผู้วิจัยจึงทำการติดต่อขอข้อมูลเหล่านี้กับผู้ค้า แต่เนื่องจากข้อมูลเหล่านี้เป็นข้อมูลทางด้านธุรกิจการค้าปลีกซึ่งเป็นความลับของทางบริษัท ดังนั้นบริษัทจึงไม่เปิดเผยข้อมูลให้ได้ ผู้วิจัยจึงทำการเก็บข้อมูลด้วยตนเองโดยการเก็บใบเสร็จเป็นเวลา 1 เดือน หลังจากนำข้อมูลที่ได้มาสร้างฐานข้อมูล

และทำการสร้างหลักเกณฑ์แอสโซซิเอชัน พบว่าหลักเกณฑ์ต่างๆที่ได้มีค่าสนับสนุนของหลักเกณฑ์น้อยมาก ที่เป็นเช่นนี้เพราะลูกค้าบางรายต้องใช้ใบเสร็จจึงไม่สามารถให้ใบเสร็จแก่ผู้วิจัยได้และผู้วิจัยไม่ได้เก็บใบเสร็จทุกวันทำให้ข้อมูลที่ได้ไม่สมบูรณ์ ซึ่งข้อมูลที่สมบูรณ์นั้นจะต้องเป็นข้อมูลทั้งหมดทุกรายการที่มีการซื้อใน 1 เดือน

จากปัญหาในเรื่องข้อมูลนี้ ผู้วิจัยจึงได้ทำการเขียนโปรแกรมจำลองข้อมูลขึ้นมาเพื่อที่จะทำการวิเคราะห์ความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ ซึ่งขั้นตอนการจำลองข้อมูล 1 ชุด ประกอบด้วยข้อมูลที่เป็นรหัสรายการและสินค้าที่มีการซื้อในแต่ละรายการ โดยกำหนดให้ตัวเลข 1-9 แทนชนิดของสินค้าที่ปรากฏในแต่ละรายการ ในการสร้างข้อมูลแต่ละชุดผู้วิจัยกำหนดจำนวนรายการทั้งหมด หลักเกณฑ์ที่มีค่าสนับสนุนใกล้เคียงกัน หลักเกณฑ์ที่มีค่าสนับสนุนประพจน์แย้งสลับที่มากที่สุดในแต่ละกลุ่มข้อมูล และในแต่ละรายการโปรแกรมจะทำการสุ่มตัวเลขเพื่อกำหนดว่าในรายการนั้นจะมีการปรากฏสินค้ากี่ชนิด ซึ่งเปรียบได้กับการซื้อสินค้าของลูกค้าในแต่ละครั้งซึ่งพบว่าจำนวนชนิดสินค้านั้นไม่จำเป็นต้องเท่ากันในทุกครั้งที่มีการซื้อ ดังนั้นผู้วิจัยจึงได้ทำการสุ่มค่านี้ขึ้นมาว่าแต่ละรายการจะมีการซื้อสินค้ากี่ชนิดซึ่งเก็บค่านี้ไว้ในตัวแปร L และโปรแกรมก็จะทำการสุ่มตัวเลข 1-9 ทั้งหมด L ครั้งโปรแกรมจะทำเช่นนี้ไปจนกว่าจะครบทุกรายการตามที่กำหนด

ในบทที่ 2 อธิบายถึงความหมายและขั้นตอนการทำเหมืองข้อมูล (data mining) ประโยชน์ของการทำเหมืองข้อมูลที่นำมาใช้ในด้านธุรกิจและองค์กรต่างๆ ความหมายของการวิเคราะห์มาเก็ตบาสเก็ต และกล่าวถึงประโยชน์ที่ได้จากการนำผลของการวิเคราะห์มาเก็ตบาสเก็ตมาใช้ประโยชน์ในทางธุรกิจ จากนั้นอธิบายถึงความรู้เบื้องต้นเกี่ยวกับหลักเกณฑ์แอสโซซิเอชัน และขั้นตอนการสร้างหลักเกณฑ์แอสโซซิเอชัน

ในบทที่ 3 อธิบายถึงโครงสร้างของฐานข้อมูลที่ใช้ในงานวิจัยนี้ และพิจารณาหลักเกณฑ์แอสโซซิเอชันในรูปแบบเซตและในรูปแบบคณิตตรรกศาสตร์ ซึ่งจะอธิบายถึงแนวความคิดในการนำค่าความเชื่อมั่นของประพจน์แย้งสลับที่หรือที่เรียกว่า ค่าความเชื่อมั่นประพจน์แย้งสลับที่ (Contrapositive confidence) มาใช้เป็นเกณฑ์การคัดเลือกหลักเกณฑ์แอสโซซิเอชันในกรณีที่ค่าความเชื่อมั่นของหลักเกณฑ์มีค่าเท่ากัน และทำการวิเคราะห์ค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ และในส่วนท้ายของบทนี้เป็นการแสดง

ขั้นตอนการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน โดยใช้ค่าความเชื่อมั่นประพจน์แย้งสลับที่เป็นเกณฑ์ในการคัดเลือก

ในบทที่ 4 อธิบายถึงการทำการทดลองเพื่อแสดงความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่และค่าความเชื่อมั่นของหลักเกณฑ์ แนวโน้มระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ รูปแบบของชุดข้อมูลที่นำมาใช้ในการทดลอง การทำงานของโปรแกรมจำลองข้อมูล แสดงผลการทดลองที่ได้ในรูปแบบของแผนภูมิแท่งเพื่อแสดงให้เห็นว่าค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์นั้นมีส่วนช่วยทำให้หลักเกณฑ์นั้นมีค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่มากขึ้น อีกส่วนหนึ่งจะเป็นการแสดงผลลัพธ์ในรูปแบบของกราฟ ซึ่งแสดงความสัมพันธ์ระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่กับค่าสนับสนุนประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ และในส่วนท้ายของบทแสดงกราฟ 3 มิติที่บอกสัมพันธ์ระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่

ในบทที่ 5 วิเคราะห์สาเหตุที่ทำให้ค่าความเชื่อมั่นของหลักเกณฑ์แอสโซซิเอชันของแต่ละหลักเกณฑ์มีค่าเท่ากันและเปรียบเทียบค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากัน เมื่อจำนวนรายการ x_1, x_2, x_3 , x_4 ของแต่ละหลักเกณฑ์มีค่าต่างกัน

ในบทที่ 6 สรุปผลของงานวิจัย แสดงลักษณะของข้อมูลที่สามารถใช้ค่าความเชื่อมั่นประพจน์แย้งสลับที่มาทำการคัดเลือกหลักเกณฑ์แอสโซซิเอชันที่มีค่าความเชื่อมั่นเท่ากันได้ แสดงจุดเด่นและจุดด้อยของงานวิจัยนี้ และเสนอแนะแนวทางที่จะนำแนวคิดของงานวิจัยนี้ไปทำการวิจัยต่อไปในอนาคต

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

การทำเหมืองข้อมูล (Data Mining)

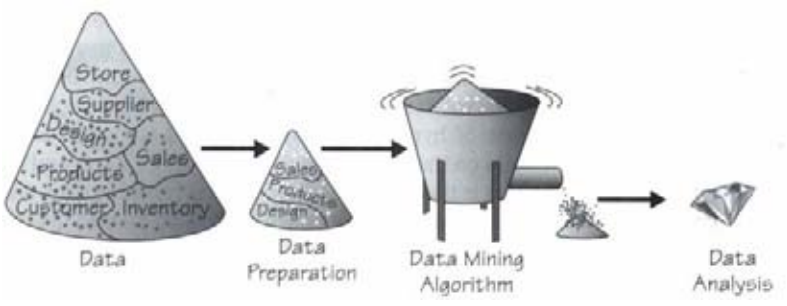
ในบทนี้จะกล่าวถึงความหมายและขั้นตอนการทำเหมืองข้อมูล (data mining) ประโยชน์ของการทำเหมืองข้อมูลที่นำมาใช้ในด้านธุรกิจและองค์กรต่างๆ ความหมายของการวิเคราะห์มาเก็ตบาสเก็ต และกล่าวถึงประโยชน์ที่ได้จากการนำผลของการวิเคราะห์มาเก็ตบาสเก็ตมาใช้ประโยชน์ในทางธุรกิจ จากนั้นอธิบายถึงความรู้เบื้องต้นเกี่ยวกับหลักเกณฑ์แอสโซซิเอชัน และขั้นตอนการสร้างหลักเกณฑ์แอสโซซิเอชัน

ในปัจจุบันองค์กรทางธุรกิจได้เติบโตอย่างรวดเร็วและมีการแข่งขันสูง สิ่งที่สำคัญที่สุดสำหรับการดำเนินการในองค์กรคือข้อมูลซึ่งมีขนาดใหญ่มาก เนื่องจากได้มีการนำข้อมูลเหล่านี้ไปใช้ประโยชน์ในการวางนโยบายขององค์กร จึงทำให้เกิดความต้องการในการที่จะดึงข้อมูลทางธุรกิจที่ได้เก็บรวบรวมไว้ออกมาเพื่อใช้ประโยชน์ เช่น ในการวิเคราะห์เพื่อวางนโยบายทางการตลาดของผู้บริหารในระดับสูงโดยใช้ข้อมูลการซื้อสินค้าของลูกค้า เนื่องจากข้อมูลที่รวบรวมไว้มีขนาดใหญ่มาก ทำให้ไม่สามารถทราบความสัมพันธ์ระหว่างข้อมูล ดังนั้นจึงมีความต้องการกระบวนการวินิจฉัยข้อมูลโดยอัตโนมัติ เพื่อที่จะทำการค้นหาความรู้ที่แอบซ่อนอยู่ในข้อมูล ซึ่งกระบวนการเหล่านี้เรียกว่าการทำเหมืองข้อมูล (data mining)

การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูลหรือเรียกอีกชื่อหนึ่งว่า การสืบค้นความรู้บนฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Database:KDD) [1] เป็นวิทยาการคอมพิวเตอร์ศาสตร์หนึ่งที่ได้รับความสนใจอย่างสูงในปัจจุบัน การสืบค้นความรู้นี้สามารถวิเคราะห์ข้อมูลขนาดใหญ่ออกมาเป็นข้อสรุปหรือความรู้ที่สำคัญที่เป็นสมบัติหรือลักษณะเฉพาะของข้อมูลในฐานข้อมูล โดยมีกระบวนการทำงาน 3 ขั้นตอน [11] ดังรูปที่ 2.1 คือ

รูปที่ 2.1 : กระบวนการทำงานของการทำเหมืองข้อมูล



1. Data Preparation : ขั้นตอนนี้ทำการกำหนดชุดข้อมูล (data set) และสมบัติของข้อมูลที่สำคัญที่นำมาใช้ในการหารูปแบบจากข้อมูลดิบทั้งหมดที่รวบรวมได้ หลังจากนั้นทำการคัดเลือกเฉพาะชุดข้อมูลและสมบัติของข้อมูลที่ต้องการออกมาจากข้อมูลทั้งหมดแล้วนำมาสร้างเป็นฐานข้อมูล โดยในขั้นตอนการสร้างฐานข้อมูลนี้ต้องอาศัยความรู้ทางด้านการจัดการระบบฐานข้อมูลมาช่วยสร้างฐานข้อมูลที่มีประสิทธิภาพ
2. Data Mining Algorithm : ขั้นตอนนี้ทำการสร้างรูปแบบจากข้อมูลที่ผ่านขั้นตอนที่ 1 มาแล้วซึ่งมีหลายเทคนิค โดยที่แต่ละเทคนิคมีขั้นตอนวิธีในการทำงานแตกต่างกันและสามารถนำไปประยุกต์ใช้กับงานต่างๆ ได้ตามความเหมาะสม เช่นงานวิจัยของ Grudnitski และ Osburn [12] Komo, C. Chang และ A. Beltratti [13] Margarita และ A. Beltratti [14] Udo [15] Wong, T.A. Bodnovich และ Y. Selvi [16] สามารถสรุปได้ดังตารางที่ 2.1
3. Data Analysis : ขั้นตอนนี้เป็นการนำเสนอรูปแบบผลลัพธ์ที่ได้มาจากขั้นตอนที่ 2 ให้ผู้คนทั่วไปเห็นภาพและเข้าใจผลลัพธ์ที่ได้อย่างชัดเจน เช่น อาจจะนำเสนอในรูปของการสร้างภาพจำลองหรือแผนภูมิ

ตารางที่ 2.1: เทคนิคต่าง ๆ ในการทำเหมืองข้อมูล

Data mining technic	Algorithm	Application examples
Association	Statistics, set	Market basket analysis
Classification	Decision trees,neural networks	Target marketing, quality control, risk assessment
Clustering	Neural networks, statistics	Market segmentation, design reuse
Modeling	Linear and nonlinear regression, curve fitting, neural networks	Ranking/scoring customers, pricing models, process control
Times-series forecasting	Statistics ARMA models, Box-Jenkins, neural networks	Sale forecasting, interest rate prediction, inventory control
Sequential patterns	Statistics, set	Market basket analysis over time

ประโยชน์ของการทำเหมืองข้อมูล

การทำเหมืองข้อมูลนั้นมีประโยชน์อย่างมากในองค์กรบริษัทต่างๆ เพราะผลที่ได้จากการทำเหมืองข้อมูลคือความรู้ที่ซ่อนอยู่ในข้อมูลที่ได้เก็บรวบรวมไว้ (historical data) และสามารถทำนายผลลัพธ์ของสถานการณ์ที่จะเกิดขึ้นในอนาคตได้ ซึ่งช่วยในการตัดสินใจทางด้านธุรกิจเพื่อให้ได้ผลที่ดีที่สุด ช่วยเพิ่มจำนวนลูกค้าและการติดต่อสื่อสารและช่วยปรับปรุงการบริการให้เป็นที่พึงพอใจของลูกค้า ข้อมูลที่ใช้วิเคราะห์ในการทำเหมืองข้อมูลนั้นแตกต่างกันโดยขึ้นอยู่กับธุรกิจการค้าของแต่ละองค์กรนั้น ซึ่งในงานวิจัยของ Kleissner [17] กล่าวถึงประโยชน์ของการทำเหมืองข้อมูลที่มีต่อองค์กรต่างๆ ดังนี้

- ธุรกิจการขายปลีก : ใช้ในการวิเคราะห์ข้อมูลมาเก็ตบาสเก็ต (market basket analysis) และกำหนดเป้าหมายยุทธการทางการตลาด
- การแพทย์ : ใช้ในการวิเคราะห์พฤติกรรมของคนไข้และวิเคราะห์การรักษาโรค
- ธนาคารและระบบความปลอดภัย : ใช้ในการวิเคราะห์การให้สินเชื่อและการตรวจสอบบัตรเครดิตปลอม
- ธุรกิจการประกันภัย : ใช้ในการวิเคราะห์การอ้างสิทธิ์ (claim analysis) และการวิเคราะห์การทุจริต (fraud analysis)
- ธุรกิจการขนส่ง : ใช้ในการวิเคราะห์รูปแบบของการบรรทุกสิ่งของ

จากองค์กรต่างๆ ที่กล่าวมาทั้งหมดนี้ใช้การทำเหมืองข้อมูลช่วยในการดึงความรู้ที่ซ่อนอยู่ในข้อมูลเพื่อนำความรู้ที่ได้นี้ไปใช้ประโยชน์ในการออกแบบสินค้าและบริการให้ผู้บริโภคพอใจและวางกลยุทธ์ในการแข่งขันกับสินค้าของคู่ต่อสู้

เนื่องจากในงานวิจัยนี้สนใจเทคนิคหลักเกณฑ์แอสโซซิเอชัน (association rule) ซึ่งเป็นเทคนิคหนึ่งในการทำเหมืองข้อมูล ซึ่งเทคนิคนี้ไปใช้กับงานทางด้านการวิเคราะห์มาเก็ตบาสเก็ต (market basket analysis) ดังนั้นในบทนี้ผู้วิจัยจะอธิบายเฉพาะการวิเคราะห์การขายและหลักเกณฑ์แอสโซซิเอชัน

การวิเคราะห์มาเก็ตบาสเก็ต (market basket analysis)

การวิเคราะห์มาเก็ตบาสเก็ต เป็นการวิเคราะห์ข้อมูลทางการตลาดวิธีหนึ่งที่เป็นที่แพร่หลายและมีประโยชน์ โดยจุดมุ่งหมายของการวิเคราะห์มาเก็ตบาสเก็ตคือ เพื่อปรับปรุงประสิทธิภาพทางการตลาดและเพิ่มกลยุทธ์ในการขายโดยใช้ข้อมูลการซื้อสินค้าแต่ละครั้งของลูกค้า ผลที่ได้จากการวิเคราะห์นี้คือการทราบว่าสินค้าชนิดใดมีการถูกซื้อพร้อมกันในการซื้อสินค้าครั้งเดียวกัน

Megaputer [2] ได้อธิบายว่ามาเก็ตบาสเก็ต (market basket) มาจากการกระทำของลูกค้าในระหว่างการเลือกซื้อสินค้าเลือกซื้อของในร้านค้าคือหยิบสินค้าที่จะซื้อใส่ในรถเข็นหรือตะกร้า (market basket) ในรูปที่ 2.2 Berry และ Linoff [18] ได้แสดงสินค้าต่างๆ ที่ลูกค้าคนหนึ่งซื้อในซุปเปอร์มาเก็ตใส่ในรถเข็น ซึ่งมีสินค้าหลายชนิดด้วยกัน คือ น้ำส้ม กล้วย โซดา น้ำยาล้างจานและน้ำยาเช็ดกระจก ซึ่งรายการสินค้าเหล่านี้บอกให้รู้ว่าลูกค้าคนหนึ่งซื้ออะไรบ้าง

รูปที่ 2.2 : **สินค้าทั้งหมดที่ลูกค้าซื้อในครั้งเดียวกัน**

ข้อดีของการวิเคราะห์มาเก็ตบาสเก็ตเมื่อเทียบกับการวิเคราะห์ด้วยวิธีอื่นๆ ของการทำเหมืองข้อมูล คือในการวิเคราะห์ข้อมูลไม่จำเป็นต้องเลือกสินค้าที่ต้องการพิจารณาก่อน แต่จะทำการพิจารณาสินค้าทั้งหมดแล้วโปรแกรมจะทำการดึงสินค้าที่สำคัญออกมาให้วิเคราะห์เอง นอกจากนี้ผลลัพธ์ที่ได้จากการวิเคราะห์มาเก็ตบาสเก็ตนี้มีความ ชัดเจนเข้าใจได้ง่าย ซึ่งผลลัพธ์นี้แสดงได้ในรูปแบบที่เรียกว่าหลักเกณฑ์แอสโซซิเอชัน (association rule)

ประโยชน์ของการวิเคราะห์มาเก็ตบาสเก็ต

ผลที่ได้จากการวิเคราะห์มาเก็ตบาสเก็ตหรือหลักเกณฑ์แอสโซซิเอชันทำให้ทราบว่าสินค้าใดที่ขายได้พร้อมกัน ซึ่งก่อให้เกิดประโยชน์อย่างมากในด้านธุรกิจ ที่เห็นได้ชัดคือยอดขายที่เพิ่มขึ้นในร้านค้าปลีกหลังจากได้มีการปรับโครงสร้างการวางตำแหน่งของสินค้าต่างๆ ภายในร้าน โดยจัดวางสินค้าที่ขายได้พร้อมกันไว้ใกล้ๆกัน เพื่อช่วยให้ลูกค้าที่ต้องการซื้อสินค้าอย่างหนึ่งแล้วไม่ลืมที่จะซื้อสินค้าอีกอย่างเนื่องจากมองไม่เห็นสินค้านั้นในบริเวณเดียวกัน

นอกจากการวิเคราะห์มาเก็ตบาสเก็ตจะให้ประโยชน์ในการจัดตำแหน่งการวางสินค้าของร้านแล้ว ในส่วนของผู้ขายตรง (direct marketer) ก็สามารถนำผลการวิเคราะห์มาใช้ประโยชน์ได้ด้วย โดยอาศัยข้อมูลประวัติการซื้อของลูกค้าประจำแต่ละคนว่าเคยซื้อสินค้าใดบ้าง หลังจากทำการวิเคราะห์ข้อมูลเหล่านี้แล้ว เมื่อมีการวางจำหน่ายสินค้าใหม่ ผู้ขายตรงก็จะทำการติดต่อกับกลุ่มลูกค้าที่คาดว่าจะมีความสนใจสินค้าชนิดใหม่แทนที่จะทำการติดต่อกับลูกค้าทุกคน ซึ่งเป็นการช่วยลดค่าใช้จ่ายในการติดต่อลูกค้าและได้รับการตอบสนองของลูกค้าไม่ต่างกัน เนื่องจากลูกค้าบางกลุ่มอาจไม่สนใจสินค้าชนิดใหม่ก็ได้

ส่วนบริษัทที่ขายสินค้าโดยที่ไม่มีร้านค้าแสดงสินค้าให้ลูกค้าได้เห็นตัวสินค้าจริงก่อนซื้อ เช่น ธุรกิจการค้าอิเล็กทรอนิกส์ (E-commerce) บริษัทขายสินค้าทางไปรษณีย์และบริษัทขายเครื่องสำอาง ลูกค้าต้องสั่งจองสินค้าทางใบสั่งและดูรายละเอียดสินค้าและราคาสินค้าจากหนังสือรายการสินค้าเพื่อจำหน่าย (catalog) บริษัทเหล่านั้นสามารถนำผลที่ได้จากการวิเคราะห์มาเก็ตบาสเก็ตมาใช้ในการกำหนดการส่งเสริมการขาย (promotion) เช่น ถ้าลูกค้าซื้อสินค้าที่กำหนดไว้ก็จะสามารถซื้อสินค้าอีกชนิดหนึ่งในราคาลดพิเศษ ซึ่งสินค้าสองชนิดนี้ได้มาจากผลการวิเคราะห์ของมาเก็ตบาสเก็ตว่าเป็นสินค้าที่ขายได้คู่กัน นอกจากนั้นสามารถนำผลการ

วิเคราะห์ที่ได้มาทำการจัดวางตำแหน่งของสินค้าที่ปรากฏในหนังสือรายการสินค้าเพื่อจำหน่าย (catalog) โดยจัดให้สินค้าที่ขายได้คู่กันมาแสดงไว้ใกล้ๆกัน หรือภายในหน้าเดียวกัน

หลักเกณฑ์แอสโซซิเอชัน

หลักเกณฑ์แอสโซซิเอชัน คือผลลัพธ์ที่ได้จากการวิเคราะห์มาสเก็ตบาสเก็ต โดยข้อมูลที่ใช้สำหรับพิจารณานี้เรียกว่าบาสเก็ตดาต้า (basket data) ประกอบด้วยรายการสินค้าที่ลูกค้าซื้อและรหัสรายการซื้อสินค้า ดังตารางที่ 2.2

ตารางที่ 2.2 : กลุ่มตัวอย่างของฐานข้อมูลการซื้อ

รายการ	สินค้า
111	ปากกา หมึก นม น้ำผลไม้
112	น้ำผลไม้ นม
113	ปากกา นม
114	ปากกา หมึก

หลักเกณฑ์แอสโซซิเอชันที่มีค่าความเชื่อมั่น 80% แสดงให้เห็นถึงความน่าจะเป็นของการซื้อสินค้าชนิดหนึ่งแล้วแสดงเปอร์เซ็นต์ของการซื้อสินค้าอีกชนิดหนึ่งในการซื้อสินค้าครั้งเดียวกัน เช่น จากหลักเกณฑ์แอสโซซิเอชันบอกได้ว่า “80 เปอร์เซ็นต์ของลูกค้าที่ซื้อขนมปังจะซื้อนมด้วยในการซื้อครั้งเดียวกัน”

กำหนดให้เซต $I=\{i_1,i_2,\ldots,i_m\}$ คือเซตของสินค้าทั้งหมดที่แตกต่างกัน m ตัว $T=\{t_1,t_2,\ldots,t_n\}$ เป็นเซตของรายการ (transaction) ทั้งหมด n รายการซึ่งอยู่ในฐานข้อมูล กำหนดให้ t คือเซตของสินค้า โดยที่ $t \subseteq I$ สัญลักษณ์ของหลักเกณฑ์แอสโซซิเอชันคือ $P \rightarrow Q$ โดยที่ P และ Q เป็นเซตของสินค้าซึ่ง P และ Q มีสมบัติดังนี้

1. $P \subseteq I$
2. $Q \subseteq I$
3. $P \cap Q = \varnothing$

จากรูปแบบ $P \rightarrow Q$ เรียก P ว่า LHS หรือพจน์แรก (antecedent) ส่วน Q เรียกว่า RHS หรือพจน์หลัง (consequent) ความหมายของ $P \rightarrow Q$ คือ “ถ้ามีการซื้อสินค้าในเซต P แล้วสินค้าในเซต Q จะถูกซื้อด้วย” เช่นหลักเกณฑ์ {ชา,กาแฟ} $\rightarrow$ {น้ำตาลทราย} กล่าวได้ว่า “ถ้าลูกค้าซื้อชาและกาแฟแล้วจะมีการซื้อน้ำตาลทรายด้วย”

จากหลักเกณฑ์แอสโซซิเอชัน $P \rightarrow Q$ กำหนดให้เซต $Z = P \cup Q$ และกำหนดให้ค่าสนับสนุนของเซตสินค้าคือ เปอร์เซ็นต์ของอัตราส่วนของจำนวนรายการที่มีเซตสินค้านั้นอยู่กับจำนวนรายการทั้งหมด ในการสร้างหลักเกณฑ์แอสโซซิเอชันนั้นมีปัจจัยที่จำเป็นสองค่า คือ ค่าสนับสนุนและค่าความเชื่อมั่นของหลักเกณฑ์

1. ค่าสนับสนุนของหลักเกณฑ์ $P \rightarrow Q$ คือ ค่าสนับสนุนของเซตสินค้า Z
2. ค่าความเชื่อมั่นของหลักเกณฑ์ $P \rightarrow Q$ คือเปอร์เซ็นต์ของอัตราส่วนระหว่างจำนวนรายการที่มีสินค้าในเซต Z กับจำนวนรายการที่มีสินค้าในเซต P หรือกล่าวได้ว่าเปอร์เซ็นต์ของอัตราส่วนระหว่างค่าสนับสนุนของหลักเกณฑ์กับค่าสนับสนุนของเซตสินค้า P

พิจารณาหลักเกณฑ์ {ปากกา}$\rightarrow${หมึก} จากตารางที่ 2.2 ค่าสนับสนุนของหลักเกณฑ์นี้คือ

$$\text{ค่าสนับสนุนของเซตสินค้า \{ปากกา,หมึก\}} = \frac{\text{จำนวนรายการที่มีสินค้า \{ปากกา,หมึก\}}}{\text{จำนวนรายการทั้งหมด}} * 100$$

$$= (2/4) * 100 = 50\%$$

$$\text{ค่าความเชื่อมั่นของหลักเกณฑ์นี้} = \frac{\text{ค่าสนับสนุนของเซตสินค้า \{ปากกา,หมึก\}}}{\text{ค่าสนับสนุนของสินค้า \{ปากกา\}}} * 100$$

$$= (2/4) * (4/3) * 100 = 66.7\%$$

จากหลักเกณฑ์นี้หมายความว่า “มีลูกค้าที่ซื้อปากกาและหมึกด้วยกันในการซื้อสินค้าครั้งเดียวกันเป็นจำนวน 50% ของการซื้อทั้งหมดและถ้าลูกค้ามีการซื้อปากกาแล้วจะมีการซื้อหมึกด้วยเป็นจำนวน 66.7% ของลูกค้าทั้งหมดที่ซื้อปากกาในครั้งเดียวกัน”

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

การสร้างหลักเกณฑ์แอสโซซิเอชัน

หลักเกณฑ์แอสโซซิเอชันได้ถูกคิดค้นครั้งแรกโดย Agrawal, Imielinski และ Swami [3] ในปี 1993 ซึ่งในงานวิจัยนี้ทำการสร้างหลักเกณฑ์แอสโซซิเอชันจากเซตของสินค้าในฐานข้อมูลการซื้อขนาดใหญ่หรือเรียกว่าบาสเก็ตดาต้า ซึ่งแต่ละรายการในฐานข้อมูลประกอบด้วยสินค้าที่ลูกค้าซื้อในครั้งหนึ่งๆ โดยหลักเกณฑ์ที่ได้ต้องมีค่าสนับสนุนของหลักเกณฑ์มากกว่าหรือเท่ากับค่าสนับสนุนที่กำหนดและมีค่าความเชื่อมั่นของหลักเกณฑ์มากกว่าหรือเท่ากับค่าความเชื่อมั่นที่กำหนดไว้ และหลักเกณฑ์ที่ได้จากกระบวนการสร้างหลักเกณฑ์ในงานวิจัยนี้มีสมาชิกในเซต Q จากหลักเกณฑ์ $P \rightarrow Q$ เพียงตัวเดียว ส่วนสมาชิกในเซต P นั้นไม่จำกัดจำนวนสมาชิก ในงานวิจัยนี้ได้แบ่งขั้นตอนในการสร้างหลักเกณฑ์แอสโซซิเอชันเป็น 2 ขั้นตอนดังนี้

1. สร้างเซตของสินค้าที่มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่ผู้ใช้กำหนด (frequent itemset)
2. สร้างหลักเกณฑ์แอสโซซิเอชันจาก frequent itemset ที่ได้จากขั้นตอนที่ 1

ต่อมาได้มีงานวิจัยของ Agrawal และ Srikant [4] ซึ่งพัฒนามาจากงานวิจัยของ Agrawal, Imielinski และ Swami [3] โดยงานวิจัยนี้ได้มีการพัฒนาขั้นตอนวิธีในการสร้างหลักเกณฑ์แอสโซซิเอชันจาก frequent itemset ให้สามารถสร้างหลักเกณฑ์แอสโซซิเอชันที่มีสมาชิกในเซต Q ของหลักเกณฑ์มากกว่า 1 ตัว และได้มีการพัฒนาขั้นตอนวิธีในการหา frequent itemset ทั้งหมดให้มีประสิทธิภาพขึ้น คือใช้เวลาในการหา frequent itemset น้อยกว่าขั้นตอนวิธีในงานวิจัยของ Agrawal, Imielinski และ Swami [3] โดยขั้นตอนวิธีที่ใช้ในงานวิจัยนี้ชื่อว่า Apriori ซึ่งเป็นขั้นตอนวิธีที่รู้จักและนิยมใช้กันอย่างแพร่หลายในปัจจุบัน

ขั้นตอนการสร้างหลักเกณฑ์แอสโซซิเอชัน

ประกอบด้วย 2 ขั้นตอนดังนี้

1. สร้างเซตของสินค้าที่มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่ผู้ใช้กำหนด (frequent itemset)
2. สร้างหลักเกณฑ์แอสโซซิเอชัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1. **ขั้นตอนการสร้างเซตของสินค้าที่มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำที่ผู้ใช้กำหนด** (frequent itemset)

ขั้นตอนนี้ทำการสร้างเซตของสินค้าที่มีค่าสนับสนุนมากกว่าค่าสนันสนุนขั้นต่ำที่ผู้ใช้กำหนด (minimum support) โดยทำการสร้าง frequent itemset ที่มีสมาชิก 1 ตัว และมีสมาชิกเพิ่มทีละตัวไปเรื่อยๆ จนไม่สามารถจะสร้าง frequent itemset ได้อีก ซึ่ง Ramakrishnan และ Gehrke [1] แสดง ขั้นตอนวิธีสร้าง frequent itemset ดังรูปที่ 2.3

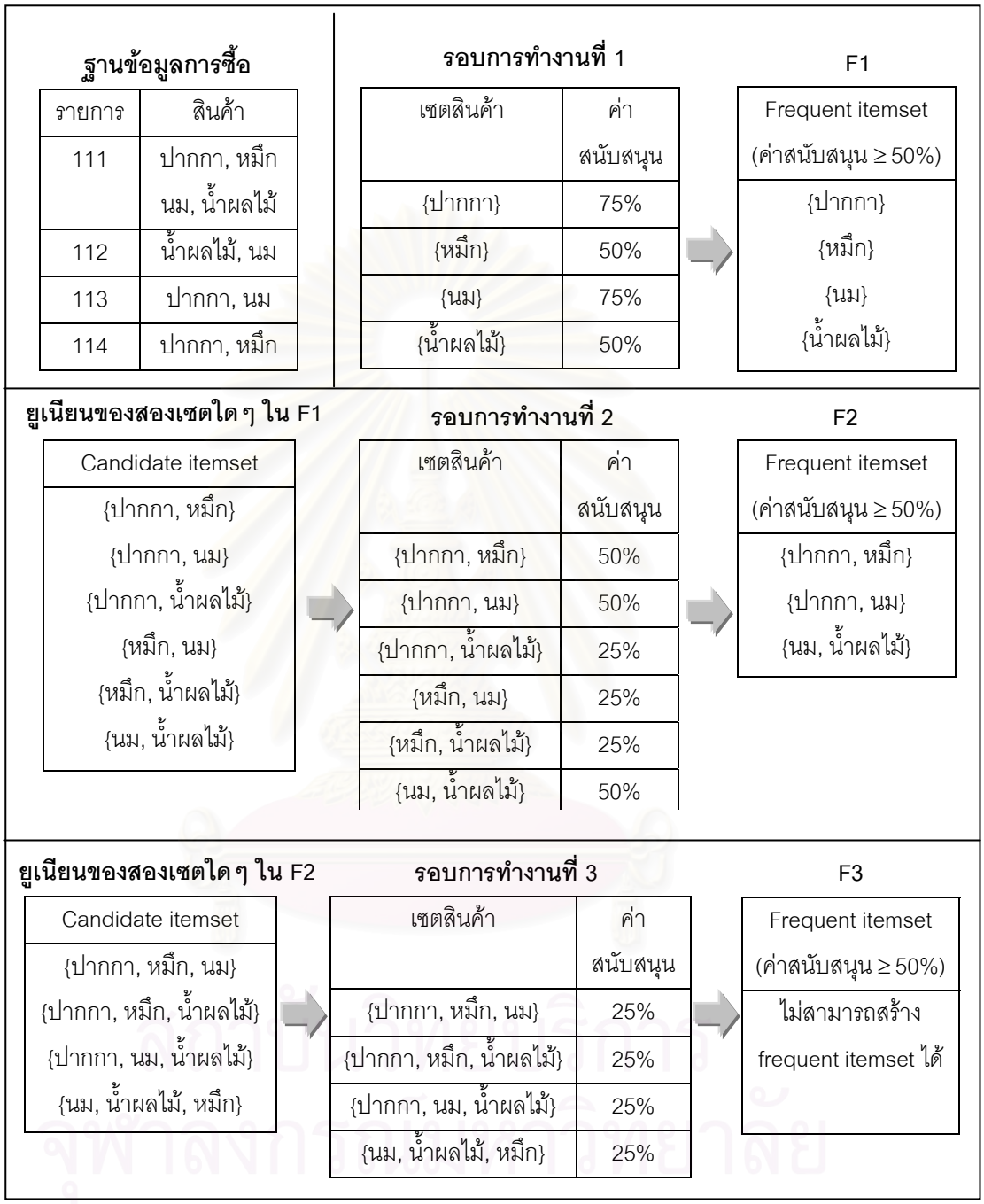
รูปที่ 2.3 : **แสดงขั้นตอนวิธีของการหา** frequent itemset

```
foreach item,                                   // level 1
        check if it is a frequent itemset       // appears in > minimum support transaction
k=1
repeat          // iterative,level-wise identification of frequent itemsets
        foreach new frequent itemset l_k with k items          // level k+1
                generate all itemsets l_{k+1} with k+1 items, l_k ⊂ l_{k+1}
                scan all transactions once and check if the generated k+1-itemsets are
                frequent
        k=k+1
until no new frequent itemsets are identified
```

จากรูปที่ 2.3 แสดงขั้นตอนการสร้าง frequent itemset ได้ดังนี้

1. หาค่าสนับสนุนของสินค้าแต่ละชนิดจากฐานข้อมูลรายการสินค้า ถ้าสินค้าชนิดใดมีค่าสนับสนุนมากกว่าหรือเท่ากับค่า minimum support สินค้าชนิดนั้นจะถูกกำหนดให้เป็น frequent itemset ที่มีสมาชิก 1 ตัว
2. ในแต่ละรอบของการทำงาน frequent itemset ที่ได้ในรอบก่อนหน้านี้จะถูกนำมาเพิ่มสมาชิกทีละ 1 ตัว ซึ่งเซตใหม่ที่มีสมาชิกเพิ่มขึ้นอีก 1 ตัว เรียกว่า candidate itemset
3. ทำการหาค่าสนับสนุนของ candidate itemset ทั้งหมดที่ได้จากขั้นตอนที่ 2 จากฐานข้อมูลรายการสินค้า ถ้าเซตใดมีค่าสนับสนุนมากกว่าหรือเท่ากับค่า minimum support จะกำหนดให้เซตนั้นเป็น frequent itemset
4. ทำซ้ำในขั้นตอนที่ 2 และ 3 ไปเรื่อยๆ จนกระทั่งไม่สามารถสร้าง frequent itemset ได้อีก

รูปที่ 2.4 : **ตัวอย่างขั้นตอนการหา** frequent itemset



รูปที่ 2.4 แสดงให้เห็นถึง frequent itemset และ candidate itemset ที่ได้จากการทำงานในแต่ละรอบตามขั้นตอนวิธีในรูปที่ 2.3 โดยอาศัยข้อมูลในตารางที่ 2.2 เมื่อกำหนดให้ minimum support มีค่า 50% จะได้ frequent itemset และ candidate itemset ดังนี้

- รอบการทำงานที่ 1

  frequent itemset = {ปากกา}, {หมึก}, {นม}, {น้ำผลไม้}

- รอบการทำงานที่ 2

  candidate itemset = {ปากกา,หมึก}, {ปากกา,นม}, {ปากกา,น้ำผลไม้}, {หมึก,นม}, {หมึก,น้ำผลไม้}, {นม,น้ำผลไม้}

  frequent itemset = {ปากกา,หมึก}, {ปากกา,นม}, {นม,น้ำผลไม้}

- รอบการทำงานที่ 3

  candidate itemset = {ปากกา,หมึก,นม}, {ปากกา,หมึก,น้ำผลไม้}, {ปากกา,นม,น้ำผลไม้}, {หมึก,นม,น้ำผลไม้}

  frequent itemset ไม่สามารถสร้างได้อีก

ดังนั้น frequent itemset ทั้งหมด คือ {ปากกา}, {หมึก}, {นม}, {น้ำผลไม้}, {ปากกา,หมึก}, {ปากกา,นม} และ {นม,น้ำผลไม้}

ในรอบการทำงานที่ 3 พบว่า candidate itemset ที่ถูกสร้างขึ้นจาก frequent itemset ที่ได้จากรอบการทำงานที่ 2 มีถึง 3 เซตด้วยกัน แต่เมื่อมีการหาค่าสนับสนุนของแต่ละเซตสินค้าแล้วพบว่าทุกเซตมีค่าสนับสนุนน้อยกว่า 50% ดังรูปที่ 2.4 ซึ่งทำให้เปลืองเวลาในการติดต่อกับฐานข้อมูลเพื่อที่จะทำการหาค่าสนับสนุนของเซตสินค้า ซึ่งถ้า candidate itemset มีจำนวนมากก็จะทำให้ใช้เวลาในการทำงานมากขึ้น ดังนั้นเราจึงใช้สมบัติที่ช่วยลดจำนวน candidate itemset ซึ่งสมบัตินี้มีชื่อว่า สมบัติเอไพรออรี่ (a priori property) [1]

สมบัติเอไพรออรี่ (a priori property) : ทุกๆ สับเซตของเซตที่เป็น frequent itemset ยกเว้นเซตว่างจะต้องเป็น frequent itemset

ดังนั้นเมื่ออาศัยสมบัติเอไพรออรี่ในการกำจัด candidate itemset ก็จะช่วยลดจำนวนเซตสินค้าที่ต้องติดต่อกับฐานข้อมูลเพื่อทำการหาค่าสนับสนุนของเซตสินค้า โดยที่ candidate itemset เซตใดไม่มีสมบัตินี้ก็จะถูกกำจัดออกไปจากการเป็น candidate itemset

พิจารณา candidate itemset ในรอบการทำงานที่ 3 พบว่าทุกเซตไม่มีสมบัติเอไพรออรี่เลย เนื่องจากมีบางสับเซตของเซต {ปากกา, หมึก, นม} และ {นม, น้ำผลไม้, หมึก} ไม่ได้เป็น frequent itemset ซึ่งก็คือ {หมึก, นม} ดังนั้นเซต {ปากกา, หมึก, นม} และ

{นม, น้ำผลไม้, หมึก} จึงถูกกำจัดออก ส่วนเซต {หมึก, น้ำผลไม้} ซึ่งเป็นสับเซตของเซต {ปากกา, หมึก, น้ำผลไม้} ไม่ได้เป็น frequent itemset ดังนั้นเซต {ปากกา, หมึก, น้ำผลไม้} จึงถูกกำจัดออก ในกรณีเดียวกันเซต {ปากกา, น้ำผลไม้} ซึ่งเป็นสับเซตของเซต {ปากกา, นม, น้ำผลไม้} ก็ไม่ได้เป็น frequent itemset ดังนั้นเซต {ปากกา, นม, น้ำผลไม้} จึงถูกกำจัด

เห็นได้ว่าเมื่อใช้สมบัติเอไพรออรี่มาช่วยกำจัด candidate itemset แล้วพบว่าในรอบการทำงานที่ 3 จะไม่มี candidate itemset เลย โปรแกรมก็จะหยุดทำการหา frequent itemset ในรอบที่ 3 ซึ่ง frequent itemset ที่ได้ทั้งหมดก็เหมือนกับในรูปที่ 2.4 โดยที่ไม่ต้องเสียเวลาไปติดต่อกับฐานข้อมูล

2. ขั้นตอนการสร้างหลักเกณฑ์แอสโซซิเอชัน

หลักเกณฑ์แอสโซซิเอชันถูกสร้างจาก frequent itemset ที่ได้จากขั้นตอนการสร้าง frequent itemset โดยมีขั้นตอนดังนี้คือ

1. นำ frequent itemset แต่ละเซตที่มีสมาชิกมากกว่าหรือเท่ากับ 2 ตัว กำหนดให้เป็นเซต S มาทำการแบ่งเป็นสองเซต คือ LHS และ RHS
2. คำนวณหาค่าความเชื่อมั่นของหลักเกณฑ์ LHS→RHS คือ อัตราส่วนระหว่างค่าสนับสนุนของเซต S กับค่าสนับสนุนของเซต LHS ถ้าค่าที่คำนวณได้มีค่ามากกว่าหรือเท่ากับค่าความเชื่อมั่นขั้นต่ำที่ผู้ใช้กำหนด (minimum confidence) ก็จะทำการสร้างหลักเกณฑ์ LHS→RHS จาก frequent itemset S

พิจารณา frequent itemset ทั้งหมดที่มีสมาชิกมากกว่าหรือเท่ากับ 2 ที่ได้จากขั้นตอนการสร้าง frequent itemset พบว่ามี {ปากกา,นม}, {ปากกา,หมึก}, {นม,น้ำผลไม้} จากขั้นตอนการสร้างหลักเกณฑ์แอสโซซิเอชัน กำหนดให้ minimum confidence เท่ากับ 60% ในการสร้างหลักเกณฑ์ทำได้ดังนี้

1. นำเซต {ปากกา,นม} มาแบ่งเป็น 2 เซตคือ LHS={ปากกา} และ RHS={นม}

2. คำนวณหาค่าความเชื่อมั่น $= \frac{\text{ค่าสนับสนุนของเซต } \{\text{ปากกา,นม}\}}{\text{ค่าสนับสนุนของเซต } \{\text{ปากกา}\}} * 100$

= (2/4) *(4/3)*100 = 66.7%

จากค่าความเชื่อมั่นที่คำนวณได้มีค่ามากกว่าค่า minimum confidence ดังนั้นจึงทำการสร้างหลักเกณฑ์ {ปากกา} → {นม}

เนื่องจากในขั้นตอนการสร้าง frequent itemset นั้นใช้เวลาในการทำงานมากเมื่อเทียบกับขั้นตอนการสร้างหลักเกณฑ์แอสโซซิเอชันจาก frequent itemset และทำให้สิ้นเปลืองหน่วยความจำในคอมพิวเตอร์มากเพราะฐานข้อมูลมีขนาดใหญ่และมีสินค้าเป็นจำนวนหลายหมื่นชนิด ดังนั้นในงานวิจัยของ Agrawal และ Srikant [4] Liu, Hsu และ Ma [5] Savasere, Omiecinski และ Navathe [6] Yen และ Chen [7] Zaki, Parthasarathy, Ogihara และ Li [8] เชาวนี ศรีวิศาล และวิเชียร เปรมชัยสวัสดิ์ [9] จึงสนใจที่จะพัฒนาขั้นตอนวิธีในส่วนของการสร้าง frequent itemset เพื่อเพิ่มประสิทธิภาพในการทำงานในเรื่องของเวลาที่ใช้ในการหา frequent itemset ขนาดและจำนวนของ frequent itemset ที่ใช้สร้างหลักเกณฑ์แอสโซซิเอชันที่มีค่าความเชื่อมั่นของหลักเกณฑ์มากกว่าหรือเท่ากับระดับที่กำหนด

แต่เนื่องจากในชีวิตจริงสินค้าในร้านค้าเช่น BigC, Lotus นั้นมีจำนวนหลายหมื่นชนิด ดังนั้นจึงมีโอกาสสูงที่หลักเกณฑ์จะมีค่าความเชื่อมั่นเท่ากันเกิดขึ้นเป็นจำนวนมาก ซึ่งทำให้ผู้ใช้เกิดความไม่แน่ใจว่าหลักเกณฑ์ใดที่นำมาใช้แล้วจะให้ผลตามที่ผู้ใช้คาดหวังไว้

พิจารณาฐานข้อมูลจากตารางที่ 1 เมื่อนำมาสร้างหลักเกณฑ์แอสโซซิเอชัน โดยกำหนดค่าสนับสนุน 50% และค่าความเชื่อมั่น 60% จะได้หลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากัน 4 หลักเกณฑ์ดังนี้ {ปากกา}→{หมึก}, {ปากกา}→{นม}, {นม}→{ปากกา}, {นม}→{น้ำผลไม้} สมมติว่าผู้ใช้เลือกหลักเกณฑ์ {ปากกา}→{นม} มาทำการวางแผนการขาย เช่นต้องการเพิ่มยอดขายของนมโดยทำการลดราคาปากกา เพื่อที่จะให้ปากกาขายได้มากขึ้นและทำให้ขายนมได้มากขึ้นด้วย ในความเป็นจริงแล้วปากกาจะนำมาใช้ได้ก็ต่อเมื่อมีหมึก ถ้าไม่มีหมึกปากกาก็จะนำมาใช้ไม่ได้ ดังนั้นโอกาสที่ลูกค้าซื้อปากกาแล้วจะซื้อหมึกจึงมีมากกว่าซื้อนม เนื่องจากปากกาและหมึกมีความสัมพันธ์กันในแง่ของการใช้งาน ด้วยเหตุนี้เมื่อมีการซื้อปากกามากขึ้นแทนที่จะเป็นการเพิ่มยอดขายของนมจึงกลับไปเพิ่มยอดขายของหมึกแทน ซึ่งไม่สอดคล้องกับหลักเกณฑ์ {ปากกา}→{นม} ที่เลือก ด้วยเหตุนี้งานวิจัยนี้จึงสนใจการพัฒนาการคัดเลือกหลักเกณฑ์แอสโซซิเอชันที่มีค่าความเชื่อมั่นเท่ากัน ซึ่งวิธีการคัดเลือกหลักเกณฑ์แอสโซซิเอชันจะกล่าวถึงอย่างละเอียดในบทที่ 3

บทที่ 3

การคัดเลือกหลักเกณฑ์แอสโซซิเอชันด้วยค่าความเชื่อมั่นประพจน์แย้งสลับที่

ในบทนี้จะกล่าวถึงโครงสร้างของฐานข้อมูลที่ใช้ในงานวิจัยนี้ และพิจารณาหลักเกณฑ์แอสโซซิเอชันในรูปแบบเซตและในรูปแบบคณิตตรรกศาสตร์ [19] ซึ่งจะอธิบายถึงแนวความคิดในการนำค่าความเชื่อมั่นของประพจน์แย้งสลับที่หรือที่เรียกว่า ค่าความเชื่อมั่นประพจน์แย้งสลับที่ (Contrapositive confidence) มาใช้เป็นเกณฑ์การคัดเลือกหลักเกณฑ์แอสโซซิเอชันในกรณีที่ค่าความเชื่อมั่นของหลักเกณฑ์มีค่าเท่ากัน และทำการวิเคราะห์ค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ และในส่วนท้ายของบทนี้เป็นการแสดงขั้นตอนการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน โดยใช้ค่าความเชื่อมั่นประพจน์แย้งสลับที่เป็นเกณฑ์ในการคัดเลือก

โครงสร้างฐานข้อมูล

เนื่องจากในการสร้างหลักเกณฑ์แอสโซซิเอชันต้องอาศัยข้อมูลบาสเก็ตดาต้าเป็นส่วนประกอบในการสร้าง เพราะหลักเกณฑ์ที่ได้จะต้องมีค่าสนับสนุนและค่าความเชื่อมั่นของหลักเกณฑ์มากกว่าหรือเท่ากับค่าที่กำหนด จึงต้องมีการดึงข้อมูล (query) ที่อยู่ในฐานข้อมูลออกมา เช่น จำนวนรายการที่มีการซื้อสินค้าแต่ละชนิด เพื่อทำการหาค่าสนับสนุนและค่าความเชื่อมั่น ดังนั้นการออกแบบโครงสร้างของฐานข้อมูลให้สามารถทำการดึงข้อมูลได้เร็วและสะดวกจึงมีความจำเป็นมาก

ในตารางที่ 3.1 แสดงโครงสร้างและข้อมูลบาสเกตดาต้าของตาราง Tran ในฐานข้อมูล โดยข้อมูลที่อยู่ในตารางนี้ได้มาจากข้อมูลตัวอย่างในบทที่ 2 (ตารางที่ 2.2)

ตารางที่ 3.1 : ตาราง Tran ในฐานข้อมูล

Tid	Pen	Ink	Milk	Juice
111	1	1	1	1
112	0	0	1	1
113	1	0	1	0
114	1	1	0	0

จากตาราง Tran ประกอบด้วย 5 สดมภ์ คือ Tid ปากกา หมึก นม และน้ำผลไม้ โดยที่

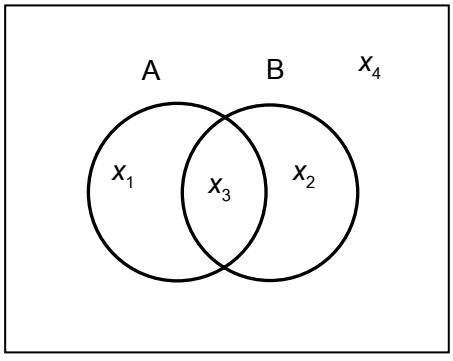
Tid	คือ รหัสรายการ มีค่าเป็นตัวอักษร
Pen	มีค่าเป็นเลข 0 หรือ 1
Ink	มีค่าเป็นเลข 0 หรือ 1
Milk	มีค่าเป็นเลข 0 หรือ 1
Juice	มีค่าเป็นเลข 0 หรือ 1

จากฐานข้อมูลให้แต่ละแถวแทนการซื้อสินค้า 1 รายการ ซึ่งมีทั้งหมด n แถว เมื่อ n เป็นจำนวนรายการทั้งหมด และมีทั้งหมด m+1 สดมภ์ เมื่อ m เป็นจำนวนชนิดสินค้าทั้งหมดที่มีการซื้อ ถ้ามีการซื้อสินค้าชนิดใดในรายการก็จะกำหนดสดมภ์สินค้าชนิดนั้นเป็น 1 ในทางกลับกันถ้าไม่มีการซื้อสินค้าชนิดใดในรายการก็จะกำหนดให้สดมภ์สินค้าชนิดนั้นเป็น 0 ตารางที่ 3.1 เป็นฐานข้อมูลตัวอย่างซึ่งประกอบด้วย 4 รายการและมีชนิดสินค้าทั้งหมด 4 ชนิดคือ ปากกา หมึก นม และน้ำผลไม้

หลักเกณฑ์แอสโซซิเอชันในรูปแบบเซต

กำหนดให้ p และ q คือสินค้า ซึ่งเป็นสมาชิกของเซต P และ Q ตามลำดับ ดังนั้นผู้วิจัยจะเขียนหลักเกณฑ์โดยใช้ $p \rightarrow q$ แทน กำหนดให้เอกภพสัมพัทธ์ (U) เป็นจำนวนรายการทั้งหมดในฐานข้อมูล ผู้วิจัยสนใจเซตที่มี p และ q คือ เซตของรายการที่ปรากฏสินค้า p (เซต A) และเซตของรายการที่ปรากฏสินค้า q (เซต B) โดยในรูปที่ 3.1 แสดงแผนภาพเวนน์-ออยเลอร์และความสัมพันธ์ระหว่างเซตต่างๆ [20,21]

รูปที่ 3.1 : แสดงความสัมพันธ์ระหว่างเซต A และ เซต B



จากรูปที่ 3.1 กำหนดให้

A คือ เซตของรายการที่ปรากฏสินค้า p

B คือ เซตของรายการที่ปรากฏสินค้า q

x_1 คือ จำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q

x_2 คือ จำนวนรายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p

x_3 คือ จำนวนรายการที่ปรากฏสินค้า p และ q

x_4 คือ จำนวนรายการที่ไม่ปรากฏสินค้า p หรือสินค้า q

กำหนดให้ T เป็นเซตของของรหัสรายการ (Tid) ทั้งหมดในฐานข้อมูล ดังนั้นจากตารางที่ 4.1 T = {111,112,113,114} โดยที่สมาชิกในเซต T ได้มาจาก SQL query ดังรูปที่ 3.2

รูปที่ 3.2 : SQL query **ในการหาสมาชิกของเซต** T

```
Select Tid From Tran
```

และกำหนดให้ I เป็นเซตของชนิดสินค้าทั้งหมดในฐานข้อมูล ดังนั้นจากตารางที่ 3.1 I = {ปากกา, หมึก, นม, น้ำผลไม้}

จากหลักเกณฑ์ p $\rightarrow$ q ในการสร้างเซต A ซึ่งเป็นเซตของรหัสรายการที่มีสินค้า p ปรากฏอยู่ในรายการ และเซต B ซึ่งเป็นเซตของรหัสรายการที่มีสินค้า q ปรากฏอยู่ในรายการ โดยที่ p และ q $\in$ I ทำได้โดยใช้คำสั่ง SQL query ดึงรหัสรายการจากฐานข้อมูลในตาราง Tran ได้ดังรูปที่ 3.3 และรูปที่ 3.4

รูปที่ 3.3 : SQL query **ในการหาสมาชิกของเซต** A

```
Select Tid From Tran
        Where p=1
```

รูปที่ 3.4 : SQL query **ในการหาสมาชิกของเซต** B

```
Select Tid From Tran
        Where q=1
```

พิจารณาอย่างหลักเกณฑ์ ปากกา $\rightarrow$ หมึก ในการหาสมาชิกของเซต A ทำได้โดยการแทน p ด้วย Pen ในคำสั่ง SQL ในรูปที่ 4.3 และแทน q ด้วย Ink ในคำสั่ง SQL ในรูปที่ 3.4 จะได้เซต A = {111,113,114} และเซต B = {111,114}

ส่วนการหาค่า x_1, x_2, x_3 และ x_4 ของหลักเกณฑ์ p $\rightarrow$ q ทำได้ดังนี้

x_1 = |A-B| = |A| - |A$\cap$B|

x_2 = |B-A| = |B| - |A$\cap$B|

x_3 = |A$\cap$B|

x_4 = $\overline{|A \cup B|}$ = |U| - |A$\cup$B| = |U| - (|A| + |B| - |A$\cap$B|)

กำหนดให้ |U| คือจำนวนรายการทั้งหมดในฐานข้อมูล และเนื่องจากได้ทำการหาสมาชิกของเซต A และเซต B ไปแล้วข้างต้น จึงทำให้ทราบค่าของ |A| และ |B| ดังนั้นเพื่อที่จะทำการหาค่า |A$\cap$B| จึงต้องทำการหาสมาชิกที่อยู่ในเซต A$\cap$B โดยใช้ SQL query ดังรูปที่ 3.5

รูปที่ 3.5 : SQL query ในการหาสมาชิกของเซต A$\cap$B

```
Select Tid From Tran
        Where p=1 and q=1
```

หลักเกณฑ์แอสโซซิเอชันในรูปแบบคณิตตรรกศาสตร์

พิจารณารูปที่ 3.1 พบว่าจำนวนรายการทั้งหมดในฐานข้อมูล (U) สามารถแบ่งได้ 4 ส่วน คือจำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) จำนวนรายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p (x_2) จำนวนรายการที่ปรากฏสินค้า p และ q (x_3) และจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) ซึ่ง |U|= $x_1+x_2+x_3+x_4$ และพิจารณาฐานข้อมูลในตารางที่ 3.1 พบว่าเมื่อกำหนดหลักเกณฑ์ p $\rightarrow$ q จะมีเหตุการณ์ที่เป็นไปได้ 4 เหตุการณ์ดังนี้

1. เหตุการณ์ที่ปรากฏสินค้า p และไม่ปรากฏสินค้า q ในรายการเดียวกัน ($p \wedge \neg q$)
2. เหตุการณ์ที่ไม่ปรากฏสินค้า p และปรากฏสินค้า q ในรายการเดียวกัน ($\neg p \wedge q$)

3. เหตุการณ์ที่ปรากฏสินค้า p และสินค้า q ในรายการเดียวกัน ($p \wedge q$)
4. เหตุการณ์ที่ไม่ปรากฏสินค้า p และไม่ปรากฏสินค้า q ในรายการเดียวกัน ($\neg p \wedge \neg q$)

โดยที่ในแต่ละรายการในฐานข้อมูลสามารถเกิดเหตุการณ์เหล่านี้ได้เพียงเหตการณ์เดียวเท่านั้น และกำหนดให้เซต S_1, S_2, S_3 และ S_4 คือเซตของรหัสรายการที่เกิดเหตุการณ์ ดังนี้

S_1 คือเซตของรหัสรายการที่เกิดเหตุการณ์ที่ 1 มีจำนวนสมาชิกในเซตเท่ากับ x_1

S_2 คือเซตของรหัสรายการที่เกิดเหตุการณ์ที่ 2 มีจำนวนสมาชิกในเซตเท่ากับ x_2

S_3 คือเซตของรหัสรายการที่เกิดเหตุการณ์ที่ 3 มีจำนวนสมาชิกในเซตเท่ากับ x_3

S_4 คือเซตของรหัสรายการที่เกิดเหตุการณ์ที่ 4 มีจำนวนสมาชิกในเซตเท่ากับ x_4

รูปที่ 3.6 แสดงให้เห็นถึงเงื่อนไขของสมาชิกในเซต S_1, S_2, S_3 และ S_4 โดยที่ p และ q คือสินค้าที่ได้จากหลักเกณฑ์ $p \rightarrow q$ ในตารางฐานข้อมูล Tran ซึ่งมีค่าเป็น 0 หรือ 1

รูปที่ 3.6 : เงื่อนไขของสมาชิกในเซต S_1, S_2, S_3 และ S_4

$S_1=\{Tid \mid (p \wedge \neg q) = 1\}$	x_1
$S_2=\{Tid \mid (\neg p \wedge q) = 1\}$	x_2
$S_3=\{Tid \mid (p \wedge q) = 1\}$	x_3
$S_4=\{Tid \mid (\neg p \wedge \neg q) = 1\}$	x_4

จำนวนรายการทั้งหมด $|U| = x_1+x_2+x_3+x_4$

จากทฤษฎีทางคณิตตรรกศาสตร์ในส่วนของตรรกศาสตร์ภาคแสดง (First-order Logic) [13] กำหนดให้

x หมายถึง รายการแต่ละรายการในฐานข้อมูล

P(x) หมายถึง มีสินค้า p ปรากฏอยู่ในรายการ x

Q(x) หมายถึง มีสินค้า q ปรากฏอยู่ในรายการ x

ซึ่ง P(x) และ Q(x) เรียกว่าฟังก์ชันเชิงประพจน์ (proposition function) หรือประโยคเปิด (open sentence) คือเป็นข้อความที่ไม่ทราบค่าความจริงจนกว่าจะทราบค่าของ

ตัวแปรที่ปรากฏ ในการที่จะทำให้ประโยคเปิดกลายเป็นประพจน์ (proposition) หรือประโยคปิด (closed sentence) ทำได้ 2 วิธีคือ แทนตัวแปรด้วยค่าคงที่และการใช้ตัวบ่งปริมาณ

ตัวบ่งปริมาณ (quantifier) มีอยู่ 2 แบบ คือ

1. ตัวบ่งปริมาณสำหรับทุกตัว (universal quantifier) ใช้สัญลักษณ์ว่า $(\forall x)$ อ่านว่า x ทุกตัว หรือ x แต่ละตัว
2. ตัวบ่งปริมาณสำหรับตัวมีจริง (existential quantifier) ใช้สัญลักษณ์ว่า $(\exists x)$ อ่านว่า มี x บางตัว หรือ มี x อย่างน้อย 1 ตัว

การบ่งปริมาณจะทำให้ฟังก์ชันเชิงประพจน์เป็นประพจน์ที่บอกค่าความจริงได้

เช่น $(\forall x)P(x)$ หมายความว่า ทุกรายการมีสินค้า p ปรากฏอยู่

$(\exists x)P(x)$ หมายความว่า บางรายการมีสินค้า p ปรากฏอยู่

ค่าความจริงของประพจน์ที่มีการบ่งปริมาณ พิจารณาดังนี้

$(\forall x)P(x)$ จริงก็ต่อเมื่อ $P(x)$ จริงเสมอไม่ว่าจะแทน x ด้วยสมาชิกตัวใดในโดเมน

$(\exists x)P(x)$ จริงก็ต่อเมื่อมี x ในโดเมนอย่างน้อย 1 ตัวที่ทำให้ $P(x)$ เป็นจริง

จากเหตุการณ์ทั้ง 4 ที่เกิดขึ้น สามารถเขียนในรูปแบบตรรกศาสตร์ภาคแสดงได้ดังนี้

1. $(\forall x)(P(x) \rightarrow \neg Q(x))$ และ $(\forall x)(\neg Q(x) \rightarrow P(x))$ มีค่าเป็นจริงเมื่อโดเมนของ x คือ S_1
2. $(\forall x)(\neg P(x) \rightarrow Q(x))$ และ $(\forall x)(Q(x) \rightarrow \neg P(x))$ มีค่าเป็นจริงเมื่อโดเมนของ x คือ S_2
3. $(\forall x)(P(x) \rightarrow Q(x))$ และ $(\forall x)(Q(x) \rightarrow P(x))$ มีค่าเป็นจริงเมื่อโดเมนของ x คือ S_3
4. $(\forall x)(\neg P(x) \rightarrow \neg Q(x))$ และ $(\forall x)(\neg Q(x) \rightarrow \neg P(x))$ มีค่าเป็นจริงเมื่อโดเมนของ x คือ S_4

เมื่อนำหลักเกณฑ์ทั้ง 8 หลักเกณฑ์ที่ได้จากข้างต้นมาหาค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เพื่อดูความสัมพันธ์ของค่าความเชื่อมั่นเหล่านี้ว่ามีสัมพันธ์กันอย่างไรกับค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q$ โดยแบ่งเป็น 4 กรณีดังนี้

1. กรณีที่ x_1 มีค่าเพิ่มขึ้น แต่ x_2, x_3 และ x_4 มีค่าคงที่ หมายความว่ารายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q มีจำนวนเพิ่มขึ้นในขณะที่รายการอื่นมีจำนวนคงที่

2. กรณีที่ x_2 มีค่าเพิ่มขึ้น แต่ x_1, x_3 และ x_4 มีค่าคงที่ หมายความว่ารายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p มีจำนวนเพิ่มขึ้นในขณะที่รายการอื่นมีจำนวนคงที่

3. กรณีที่ x_3 มีค่าเพิ่มขึ้น แต่ x_1, x_2 และ x_4 มีค่าคงที่ หมายความว่ารายการที่ปรากฏสินค้า p และ q มีจำนวนเพิ่มขึ้นในขณะที่รายการอื่นมีจำนวนคงที่

4. กรณีที่ x_4 มีค่าเพิ่มขึ้น แต่ x_1, x_2 และ x_3 มีค่าคงที่ หมายความว่ารายการที่ไม่ปรากฏสินค้า p หรือ q มีจำนวนเพิ่มขึ้นในขณะที่รายการอื่นมีจำนวนคงที่

ผู้วิจัยอธิบายเฉพาะกรณีที่ค่า x_1, x_2, x_3 และ x_4 มีค่าเพิ่มขึ้น เนื่องจากกรณีที่ค่า x_1, x_2, x_3 และ x_4 มีค่าลดลงนั้น ค่าความเชื่อมั่นที่ได้ในแต่ละหลักเกณฑ์จะตรงข้ามกับกรณีที่ค่า x_1, x_2, x_3 และ x_4 มีค่าเพิ่มขึ้น

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 3.2 : **ค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ในกรณีต่าง ๆ**

หลักเกณฑ์	ค่าความเชื่อมั่น			
	เพิ่มค่า x_1	เพิ่มค่า x_2	เพิ่มค่า x_3	เพิ่มค่า x_4
$p \rightarrow \neg q$	$\uparrow$	=	$\downarrow$	=
$\neg q \rightarrow p$	$\uparrow$	=	=	$\downarrow$
$\neg p \rightarrow q$	=	$\uparrow$	=	$\downarrow$
$q \rightarrow \neg p$	=	$\uparrow$	$\downarrow$	=
$p \rightarrow q$	$\downarrow$	=	$\uparrow$	=
$q \rightarrow p$	=	$\downarrow$	$\uparrow$	=
$\neg p \rightarrow \neg q$	=	$\downarrow$	=	$\uparrow$
$\neg q \rightarrow \neg p$	$\downarrow$	=	=	$\uparrow$

จากตารางที่ 3.2 แสดงให้เห็นถึงค่าความเชื่อมั่นในกรณีทั้ง 4 ของหลักเกณฑ์ทั้ง 8 หลักเกณฑ์ โดยที่

เครื่องหมาย = หมายความว่า ค่าความเชื่อมั่นมีค่าคงที่

เครื่องหมาย $\uparrow$ หมายความว่า ค่าความเชื่อมั่นมีค่าเพิ่มขึ้น

เครื่องหมาย $\downarrow$ หมายความว่า ค่าความเชื่อมั่นมีค่าลดลง

ซึ่งค่าความเชื่อมั่นของแต่ละหลักเกณฑ์แสดงได้ดังนี้

1. ค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow \neg q = \frac{|A \cap \overline{B}|}{|A|} = \frac{x_1}{x_1 + x_3} = \frac{1}{1 + \frac{x_3}{x_1}}$

2. ค่าความเชื่อมั่นของหลักเกณฑ์ $\neg q \rightarrow p = \frac{|A \cap \overline{B}|}{|\overline{B}|} = \frac{x_1}{x_1 + x_4} = \frac{1}{1 + \frac{x_4}{x_1}}$

3. ค่าความเชื่อมั่นของหลักเกณฑ์ $\neg p \rightarrow q = \frac{|\overline{A} \cap B|}{|\overline{A}|} = \frac{x_2}{x_2 + x_4} = \frac{1}{1 + \frac{x_4}{x_2}}$

4. ค่าความเชื่อมั่นของหลักเกณฑ์ $q \rightarrow \neg p = \frac{|\overline{A} \cap B|}{|B|} = \frac{x_2}{x_2 + x_3} = \frac{1}{1 + \frac{x_3}{x_2}}$

5. ค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q = \frac{|A \cap B|}{|A|} = \frac{x_3}{x_1 + x_3} = \frac{1}{1 + \frac{x_1}{x_3}}$

6. ค่าความเชื่อมั่นของหลักเกณฑ์ $q \rightarrow p = \frac{|A \cap B|}{|B|} = \frac{x_3}{x_2 + x_3} = \frac{1}{1 + \frac{x_2}{x_3}}$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

7. ค่าความเชื่อมั่นของหลักเกณฑ์ $\neg p \rightarrow \neg q = \frac{|\overline{A} \cap \overline{B}|}{|\overline{A}|} = \frac{x_4}{x_2 + x_4} = \frac{1}{1 + \frac{x_2}{x_4}}$

8. ค่าความเชื่อมั่นของหลักเกณฑ์ $\neg q \rightarrow \neg p = \frac{|\overline{A} \cap \overline{B}|}{|\overline{B}|} = \frac{x_4}{x_1 + x_4} = \frac{1}{1 + \frac{x_1}{x_4}}$

จากตารางที่ 3.2 จะเห็นว่าหลักเกณฑ์ที่ให้ค่าความเชื่อมั่นใกล้เคียงกับหลักเกณฑ์ $p \rightarrow q$ มากที่สุดมี 3 หลักเกณฑ์ คือ $p \rightarrow \neg q$, $q \rightarrow p$ และ $\neg q \rightarrow \neg p$ ซึ่งต่างก็มีค่าความเชื่อมั่นเหมือนกับหลักเกณฑ์ $p \rightarrow q$ 2 กรณี ดังนี้

$p \rightarrow \neg q$ มีค่าความเชื่อมั่นเหมือนกับ $p \rightarrow q$ ในกรณีที่รายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_2) และรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) มีจำนวนเพิ่มขึ้น

$q \rightarrow p$ มีค่าความเชื่อมั่นเหมือนกับ $p \rightarrow q$ ในกรณีที่รายการที่ปรากฏสินค้า p และ q (x_3) และรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) มีจำนวนเพิ่มขึ้น

$\neg q \rightarrow \neg p$ มีค่าความเชื่อมั่นเหมือนกับ $p \rightarrow q$ ในกรณีที่รายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) และรายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p (x_2) มีจำนวนเพิ่มขึ้น

แต่เนื่องจากค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow \neg q$ ในกรณีที่รายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) มีจำนวนเพิ่มขึ้น จะทำให้ค่าความเชื่อมั่นมีค่าเพิ่มขึ้น ซึ่งตรงกันข้ามกับค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q$ กรณีที่รายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) มีจำนวนเพิ่มขึ้น จะทำให้ค่าความเชื่อมั่นของหลักเกณฑ์มีค่าลดลง และในกรณีที่รายการที่ปรากฏสินค้า p และ q (x_3) มีจำนวนเพิ่มขึ้น ค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow \neg q$ มีค่าลดลงแต่ค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q$ กลับมีค่าเพิ่มขึ้น นั่นคือค่าความเชื่อมั่นของทั้งสองหลักเกณฑ์ขัดแย้งกัน จึงสรุปได้ว่าค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow \neg q$ ไม่สนับสนุนค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q$ ในทุกกรณี

พิจารณาหลักเกณฑ์ $p \rightarrow q$, $q \rightarrow p$ และ $\neg q \rightarrow \neg p$ ในทางตรรกศาสตร์ภาคแสดง ซึ่งมีสัญลักษณ์ $(\forall x)(P(x) \rightarrow Q(x))$, $(\forall x)(Q(x) \rightarrow P(x))$ และ $(\forall x)(\neg Q(x) \rightarrow \neg P(x))$ ตามลำดับ

ตารางที่ 3.3 : **ค่าความจริงของ** $(\forall x)(P(x) \rightarrow Q(x))$, $(\forall x)(Q(x) \rightarrow P(x))$ **และ** $(\forall x)(\neg Q(x) \rightarrow \neg P(x))$

โดเมนของ x	P(x)	¬P(x)	Q(x)	¬Q(x)	P(x)→Q(x)	Q(x)→P(x)	¬Q(x)→¬P(x)
เซต S_1	T	F	T	F	T	T	T
เซต S_2	T	F	F	T	F	T	F
เซต S_3	F	T	T	F	T	F	T
เซต S_4	F	T	F	T	T	T	T

พิจารณาตารางที่ 3.3 พบว่าหลักเกณฑ์ $p \rightarrow q$ มีค่าความจริงเหมือนกับหลักเกณฑ์ $\neg q \rightarrow \neg p$ ในทุกโดเมนของ x ดังนั้นจึงเรียกประพจน์ $\neg q \rightarrow \neg p$ ว่าเป็นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p \rightarrow q$ และค่าความเชื่อมั่นของหลักเกณฑ์ $\neg q \rightarrow \neg p$ มีค่าเหมือนกับค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q$ ในกรณีที่ x_1 และ x_2 มีค่าเพิ่มขึ้น จึงสรุปได้ว่าค่าความเชื่อมั่นของหลักเกณฑ์ $\neg q \rightarrow \neg p$ สนับสนุนค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q$ ดังนั้นค่าที่จะมาใช้ในการคัดเลือกหลักเกณฑ์แอสโซซิเอชันที่มีค่าความเชื่อมั่นของหลักเกณฑ์เท่ากันคือค่าความเชื่อมั่นของหลักเกณฑ์ $\neg q \rightarrow \neg p$ ซึ่งเรียกอีกชื่อหนึ่งว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ (Contrapositive confidence) ซึ่งมีค่าเท่ากับเปอร์เซ็นต์ของอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p หรือสินค้า q กับจำนวนรายการที่ไม่ปรากฏสินค้า q

เนื่องจากหลักเกณฑ์ $p \rightarrow q$ หมายความว่า ถ้ามีการซื้อสินค้า p แล้วจะมีการซื้อสินค้า q แสดงว่าสินค้า p และ q ถูกซื้อพร้อมกัน เมื่อพิจารณารายการที่ไม่มีการซื้อสินค้า p และ q ด้วยกัน (รายการที่ไม่อยู่ในเหตุการณ์ที่ 3) พบว่าหลักเกณฑ์ $p \rightarrow q$ ที่น่าเชื่อถือจะต้องมีจำนวนรายการที่มีการซื้อสินค้า p กับสินค้าชนิดอื่นที่ไม่ใช่สินค้า q (รายการที่อยู่ในเหตุการณ์ที่ 1) หรือมีจำนวนรายการที่มีการซื้อสินค้า q กับสินค้าชนิดอื่นที่ไม่ใช่สินค้า p (รายการที่อยู่ในเหตุการณ์ที่ 2) น้อยที่สุด

ตารางที่ 3.4 : ตัวอย่างฐานข้อมูลการซื้อ

Tid	1	2	3	4	5	6	7	8	9
111	1	1	1	1	1	0	0	0	0
112	1	1	0	0	0	1	1	0	0
113	1	0	1	1	0	1	0	0	0
114	1	1	0	1	0	0	0	0	0
115	0	0	1	0	1	0	1	1	0
116	0	0	1	0	0	1	0	0	1
117	1	0	1	1	0	0	0	0	0

ตัวอย่างเช่น พิจารณาหลักเกณฑ์ 1 $\longrightarrow$ 2 และ 3 $\longrightarrow$ 4 จากฐานข้อมูลตัวอย่างในตารางที่ 3.4 พบว่าหลักเกณฑ์ทั้งสองมีค่าสนับสนุนของหลักเกณฑ์เท่ากันคือ 42.86% และมีค่าความเชื่อมั่นเท่ากัน คือเท่ากับ 60% และหลักเกณฑ์ทั้งสองมีจำนวนรายการที่ปรากฏสินค้า 1 และไม่ปรากฏสินค้า 2 เท่ากับจำนวนรายการที่ปรากฏสินค้า 3 และไม่ปรากฏสินค้า 4 (x_1 ของแต่ละหลักเกณฑ์) เมื่อพิจารณาจำนวนรายการที่ปรากฏสินค้า 2 และไม่ปรากฏสินค้า 1 (ค่า x_2 ของหลักเกณฑ์ 1 $\longrightarrow$ 2) มีค่าเท่ากับศูนย์ ในขณะที่จำนวนรายการที่ปรากฏสินค้า 4 และไม่ปรากฏสินค้า 3 (ค่า x_2 ของหลักเกณฑ์ 3 $\longrightarrow$ 4) มีค่าเท่ากับ 1 ซึ่งพบว่าค่า x_2 ของหลักเกณฑ์ 1 $\longrightarrow$ 2 มีค่าน้อยกว่าค่า x_2 ของหลักเกณฑ์ 3 $\longrightarrow$ 4 ดังนั้นหลักเกณฑ์ 1 $\longrightarrow$ 2 จึงเป็นหลักเกณฑ์ที่น่าเชื่อถือกว่าหลักเกณฑ์ 3 $\longrightarrow$ 4 และเมื่อพิจารณาจำนวนรายการที่ไม่ปรากฏสินค้า 1 หรือ 2 (ค่า x_4 ของหลักเกณฑ์ 1 $\longrightarrow$ 2) มีค่าเท่ากับ 2 ซึ่งมีค่ามากกว่าจำนวนรายการที่ไม่ปรากฏสินค้า 3 หรือ 4 (ค่า x_4 ของหลักเกณฑ์ 3 $\longrightarrow$ 4) ซึ่งมีค่าเท่ากับ 1 เมื่อนำหลักเกณฑ์ทั้งสองมาหาค่าความเชื่อมั่นประพจน์แย้งสลับที่ พบว่าหลักเกณฑ์ 1 $\longrightarrow$ 2 มีค่าเท่ากับ 50% ส่วนหลักเกณฑ์ 3 $\longrightarrow$ 4 มีค่าเท่ากับ 33.33% เนื่องจากหลักเกณฑ์ที่น่าเชื่อถือคือหลักเกณฑ์ 1 $\longrightarrow$ 2 นั้นมีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงกว่าหลักเกณฑ์ 3 $\longrightarrow$ 4 ดังนั้นในการคัดเลือกหลักเกณฑ์จึงทำการคัดเลือกหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุด

ค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์

$$\text{ค่าความเชื่อมั่นของหลักเกณฑ์ } p \rightarrow q = \frac{1}{1+\frac{x_1}{x_3}} \quad (1)$$

$$\text{ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของ } p \rightarrow q = \frac{1}{1+\frac{x_1}{x_4}} \quad (2)$$

พิจารณาสมการที่ 1 และ 2 พบว่าค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่มีรูปแบบคล้ายกัน แตกต่างกันที่ตัวแปร x_3 และ x_4 เท่านั้น ดังนั้นสามารถสรุปความสัมพันธ์ระหว่างค่าความเชื่อมั่นทั้งสองของหลักเกณฑ์ $p \rightarrow q$ ได้ดังนี้

1. กรณีที่ $x_3 = x_4$ หมายความว่า เมื่อจำนวนรายการที่ปรากฏสินค้า p และ q ในรายการเดียวกันเท่ากับจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q ในรายการเดียวกัน จะทำให้ค่าความเชื่อมั่นทั้งสองจะมีค่าเท่ากัน

2. กรณีที่ $x_3 > x_4$ หมายความว่า เมื่อจำนวนรายการที่ปรากฏสินค้า p และ q ในรายการเดียวกันมีค่ามากกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q ในรายการเดียวกันจะทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่มีค่าน้อยกว่าค่าความเชื่อมั่นของหลักเกณฑ์นั้น

3. กรณีที่ $x_3 < x_4$ หมายความว่า เมื่อจำนวนรายการที่ปรากฏสินค้า p และ q ในรายการเดียวกันมีค่าน้อยกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q ในรายการเดียวกันจะทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่มีค่ามากกว่าค่าความเชื่อมั่นของหลักเกณฑ์นั้น ซึ่งในกรณีนี้พบได้บ่อยในฐานข้อมูลการซื้อขายในชีวิตประจำวัน เช่น ฐานข้อมูลการซื้อของบิ๊กซี เนื่องจากศูนย์การค้าเหล่านี้มีจำนวนชนิดสินค้าหลายพันชนิด และในการซื้อสินค้าของลูกค้าในแต่ละครั้งส่วนใหญ่มีการซื้อสินค้าไม่เกิน 100 ชนิด ดังนั้นจึงทำให้มีจำนวนรายการที่ไม่มีการซื้อสินค้า p หรือสินค้า q สูงกว่าจำนวนรายการที่มีการซื้อสินค้า p และ q ด้วยกันในการซื้อสินค้าครั้งเดียวกัน

ขั้นตอนวิธีในการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน

การคัดเลือกหลักเกณฑ์แอสโซซิเอชัน มีขั้นตอนดังนี้

1. สร้างเซตของสินค้าที่มีสมาชิก 1 หรือ 2 ตัว ที่มีค่าสนับสนุนมากกว่าหรือเท่ากับค่า minimum support และเก็บค่าจำนวนรายการที่มีเซตนั้นปรากฏในรายการสินค้าในฐานข้อมูลของแต่ละเซต
2. นำเซตที่มีสมาชิก 2 ตัวทั้งหมดที่ได้ในข้อ 1 มาสร้างหลักเกณฑ์ $p \rightarrow q$ โดยที่หลักเกณฑ์ที่ได้ต้องมีค่าความเชื่อมั่นมากกว่าหรือเท่ากับค่าความเชื่อมั่นที่ผู้ใช้กำหนด
3. นำหลักเกณฑ์จากข้อ 2 ที่มีค่าความเชื่อมั่นของหลักเกณฑ์เท่ากันมาหาค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ โดยคำนวณได้จาก

$$\frac{|\overline{A \cup B}|}{|\overline{B}|} = \frac{\text{จำนวนรายการทั้งหมด} - |A \cup B|}{\text{จำนวนรายการทั้งหมด} - |B|}$$

$$= \frac{\text{จำนวนรายการทั้งหมด} - (|A| + |B| - |A \cap B|)}{\text{จำนวนรายการทั้งหมด} - |B|}$$

เนื่องจาก $|A|$ $|B|$ และ $|A \cap B|$ ได้ทำการหาค่าไว้แล้วในข้อ 1 จึงนำค่าเหล่านี้มาใช้คำนวณค่าความเชื่อมั่นประพจน์แย้งสลับที่

4. คัดเลือกหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุด

ขั้นตอนวิธีในการคัดเลือกหลักเกณฑ์แอสโซซิเอชันในส่วนของการสร้างหลักเกณฑ์แอสโซซิเอชัน (ขั้นตอนที่ 1-2) ที่กล่าวมานี้ได้นำมาใช้ในโปรแกรมการสร้างหลักเกณฑ์แอสโซซิเอชันและคำนวณค่าความเชื่อมั่นประพจน์แย้งสลับที่จากสูตรในขั้นตอนที่ 3 ของการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน เพื่อทำการสร้างหลักเกณฑ์แอสโซซิเอชันจากข้อมูลที่ได้จากโปรแกรมจำลองข้อมูล ซึ่งรูปแบบของข้อมูลและวิธีการทดลองจะกล่าวอย่างละเอียดในบทที่ 4

บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงการทำการทดลองเพื่อแสดงความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่และค่าความเชื่อมั่นของหลักเกณฑ์แนวโน้มระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ โดยข้อมูลที่ใช้ในการทดลองนี้ได้มาจากโปรแกรมจำลองข้อมูล

ผลลัพธ์ที่ได้จากการทดลองจะแสดงหลักเกณฑ์แอสโซซิเอชันทุกหลักเกณฑ์ที่มีค่าสนับสนุนของหลักเกณฑ์มากกว่าหรือเท่ากับ 20% และแสดงค่าความเชื่อมั่น ค่าความเชื่อมั่นประพจน์แย้งสลับที่ จำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) จำนวนรายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p (x_2) จำนวนรายการที่ปรากฏสินค้า p และ q (x_3) และจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) ของแต่ละหลักเกณฑ์ โดยเรียงลำดับจากหลักเกณฑ์ที่มีค่าความเชื่อมั่นน้อยไปหาหลักเกณฑ์ที่มีค่าความเชื่อมั่นมาก และนำผลลัพธ์ที่ได้มาแสดงในรูปแบบของแผนภูมิแท่งเพื่อแสดงให้เห็นว่าค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์นั้นมีส่วนช่วยทำให้หลักเกณฑ์มีค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่มากขึ้น อีกส่วนหนึ่งจะเป็นการแสดงผลลัพธ์ในรูปแบบของกราฟซึ่งแสดงความสัมพันธ์ระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่กับค่าสนับสนุนประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์

ในส่วนท้ายของบทได้แสดงกราฟ 3 มิติแสดงความสัมพันธ์ระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ โดยจำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q จำนวนรายการที่ปรากฏสินค้า p และ q จำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q มีค่าตั้งแต่ 1 ถึง 10,000 รายการ

การทดลอง

ในการทดลองนี้ใช้โปรแกรมที่เขียนด้วยภาษาจาวาทำการดึงข้อมูลจากฐานข้อมูลที่อยู่ในโปรแกรม Microsoft Access ด้วยคำสั่ง SQL language เพื่อใช้ในการหาค่าสนับสนุนของเซตสินค้า และกำหนดให้ minimum support มีค่า 20% ซึ่งข้อมูลที่ใช้ทดสอบกับโปรแกรมนี้ได้มา

จากโปรแกรมจำลองข้อมูลที่ผู้วิจัยทำการเขียนโปรแกรมขึ้นมา โดยแทนสินค้าแต่ละชนิดด้วยตัวเลข 1-9 ผู้วิจัยได้กำหนดให้ข้อมูลที่ใช้ทดสอบนี้มีค่าสนับสนุนของเซตสินค้า {1,2}, {3,4}, {5,6} และ {7,8} มีค่าใกล้เคียงกันและมีค่ามากกว่าหรือเท่ากับ 20% และกำหนดให้จำนวนชนิดสินค้าที่ปรากฏในแต่ละรายการมีค่ามากที่สุด 3 ชนิด

พิจารณาเซตสินค้า {p,q} พบว่าสมาชิกในเซตนี้สามารถนำมาสร้างหลักเกณฑ์ได้ 2 หลักเกณฑ์ คือ $p \rightarrow q$ และ $q \rightarrow p$ ซึ่งมี $\neg q \rightarrow \neg p$ เป็นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p \rightarrow q$ และ $\neg p \rightarrow \neg q$ เป็นประพจน์แย้งสลับที่ของหลักเกณฑ์ $q \rightarrow p$ ดังนั้นค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์ $p \rightarrow q$ และ $q \rightarrow p$ คือเปอร์เซ็นต์ของอัตราส่วนระหว่างจำนวนรายการที่ไม่มีสมาชิกในเซตสินค้า {p,q} ปรากฏอยู่กับจำนวนรายการทั้งหมด ซึ่งพบว่าหลักเกณฑ์ทั้งสองมีค่าสนับสนุนประพจน์แย้งสลับที่เท่ากัน ดังนั้นเพื่อความสะดวกในการอธิบายถึงกลุ่มของข้อมูล ผู้วิจัยจึงใช้คำว่าค่าสนับสนุนประพจน์แย้งสลับที่ของเซตสินค้า {p,q} แทนคำว่าค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์ $p \rightarrow q$ และ $q \rightarrow p$ เพราะหลักเกณฑ์ทั้งสองมีค่าสนับสนุนประพจน์แย้งสลับที่เท่ากัน

เนื่องจากการทดลองนี้ต้องการแสดงให้เห็นถึงความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ ดังนั้นจึงทำการแบ่งข้อมูลเป็น 4 กลุ่ม โดยแต่ละกลุ่มจะมีการกำหนดให้มีรายการที่ไม่ปรากฏเซตสินค้า {1,2}, {3,4}, {5,6} และ {7,8} เกิดขึ้นตามลำดับในแต่ละกลุ่ม เพื่อเป็นการเพิ่มค่าสนับสนุนประพจน์แย้งสลับที่ของเซตสินค้า {p,q} ในแต่ละกลุ่ม สำหรับตัวเลขอื่นๆ ที่ไม่อยู่ในเงื่อนไขที่ปรากฏในแต่ละรายการนั้นจะทำการสุ่มตัวเลขขึ้นมาด้วยฟังก์ชันการสุ่มตัวเลข ซึ่งกระบวนการทำงานของโปรแกรมจำลองข้อมูลมีขั้นตอนดังรูปที่ 4.1

จากรูปที่ 4.1 จะเห็นว่าขั้นตอนการจำลองข้อมูลมี 2 ขั้นตอนหลักๆ คือ ขั้นตอนการสร้างรายการแต่ละรายการ และขั้นตอนการสร้างชนิดสินค้าภายในรายการแต่ละรายการ โดยในบรรทัดที่ 1 ทำการกำหนดจำนวนรายการทั้งหมดในข้อมูล 1 ชุด และทำการสร้างรายการแต่ละรายการในบรรทัดที่ 2-34 ซึ่งภายในแต่ละรอบของการสร้างรายการจะทำการสุ่มตัวเลข 1-3 เพื่อกำหนดจำนวนสินค้าในแต่ละรายการในบรรทัดที่ 4 จากนั้นจะทำการสร้างสินค้าแต่ละชนิดใน 1 รายการในบรรทัดที่ 5-33 โดยเงื่อนไขในบรรทัดที่ 7-30 จะทำให้จำนวนรายการที่มีเซตสินค้า {1,2}, {3,4}, {5,6} และ {7,8} มีค่าใกล้เคียงกัน ดังนั้นทุกกลุ่มข้อมูลจึงมีจำนวนเซตสินค้าเหล่านี้ใกล้เคียงกันเพราะมีเงื่อนไขตามบรรทัดที่ 7-30 เหมือนกันทั้งหมด แต่ในแต่ละกลุ่มข้อมูล

จะมีเงื่อนไขแตกต่างกันในบรรทัดที่ 31 ซึ่งตัวแปร n1 และ n2 คือชนิดสินค้าที่อยู่ในเซตสินค้าของแต่ละกลุ่ม เช่น ถ้าต้องการสร้างข้อมูลในกลุ่มที่ 1 โดยกำหนดให้กลุ่มที่ 1 คือเซตสินค้า {1,2} ดังนั้นค่าของ n1 คือ 1 และค่าของ n2 คือ 2

รูปที่ 4.1: กระบวนการทำงานของโปรแกรมจำลองข้อมูล

```
1   n=total transaction                          // define total transactions
2   for each transaction (i=0;i<n;i++)           // generate each transaction
3   begin
4         l=random(max_length)+1                 // define total items in each transaction
5         for each item in ith transaction (j=0;j<l;j++)  //  generate item in each transaction
6         begin
7               if(i mod 5=1)
8               begin
9                     if(j=0)  generate item 1 in ith transaction
10                    else if(j=1)  generate item 2 in ith transaction
11                    else  random another number in ith transaction
12              end
13              if(i mod 5=2)
14              begin
15                    if(j=0)  generate item 3 in ith transaction
16                    else if(j=1)  generate item 4 in ith transaction
17                    else  random another number in ith transaction
18              end
19              if(i mod 5=3)
20              begin
21                    if(j=0)  generate item 5 in ith transaction
22                    else if(j=1)  generate item 6 in ith transaction
23                    else  random another number in ith transaction
24              end
25              if(i mod 5=4)
26              begin
27                    if(j=0)  generate item 7 in ith transaction
28                    else if(j=1)  generate item 8 in ith transaction
29                    else  random another number in ith transaction
30              end
31              if(i mod 10=0)  random another number except n1 and n2
32              if(i mod 5=0 and i mod 10≠0)  random number 1-9
33        end
34  end
```

รูปแบบของชุดข้อมูล

เนื่องจากผู้วิจัยต้องการหาความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ ดังนั้นเพื่อให้ได้ผลที่แน่นอนผู้วิจัยจึงทำการทดสอบกับข้อมูลหลายๆ ชุดที่มีรูปแบบแตกต่างกันไปเพื่อดูว่าผลสรุปที่ได้ในแต่ละชุดข้อมูลที่นำมาทดสอบนั้นมีแนวโน้มอย่างเดียวกันหรือไม่ จึงได้ทำการแบ่งกลุ่มข้อมูลออกเป็น 4 กลุ่มตามเซตสินค้าที่

ต้องการให้มีค่าสนับสนุนประพจน์แย้งสลับที่ของเซตสินค้า {p,q} มากที่สุด โดยกำหนดเซตสินค้า {p,q} ในแต่ละกลุ่มดังนี้

กลุ่มที่ 1 กำหนดให้ p=1 q=2

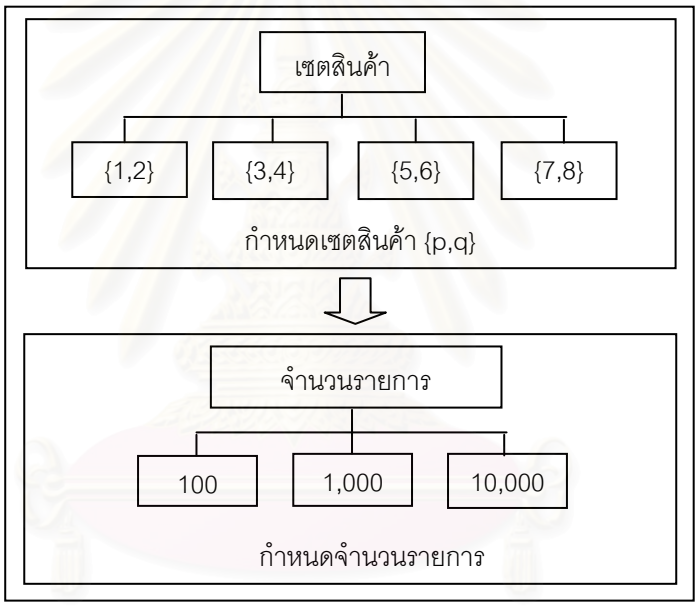
กลุ่มที่ 2 กำหนดให้ p=3 q=4

กลุ่มที่ 3 กำหนดให้ p=5 q=6

กลุ่มที่ 4 กำหนดให้ p=7 q=8

ในแต่ละกลุ่มทำการสร้างข้อมูลที่มีจำนวนรายการทั้งหมดเป็น 100, 1,000 และ 10,000 รายการ ในแต่ละรูปแบบดังรูปที่ 4.2

รูปที่ 4.2 การกำหนดกลุ่มและรูปแบบของชุดข้อมูล



ตารางที่ 4.1: รูปแบบของข้อมูลในแต่ละกลุ่ม

กลุ่มข้อมูล	เซตสินค้า	จำนวนรายการ	จำนวนชุดข้อมูล
1	{1,2}	100	10
		1000	10
		10000	10
2	{3,4}	100	10
		1000	10
		10000	10
3	{5,6}	100	10
		1000	10
		10000	10
4	{7,8}	100	10
		1000	10
		10000	10

จากตารางที่ 4.1 แสดงรูปแบบทั้งหมดของข้อมูลทั้ง 4 กลุ่มมีรูปแบบที่แตกต่างกันทั้งหมด 12 รูปแบบ ซึ่งในแต่ละรูปแบบนั้นผู้วิจัยได้จำลองข้อมูลรูปแบบละ 10 ชุด โดยกำหนดค่าเริ่มต้นของเลขสุ่มในฟังก์ชัน srand เป็น 1-10 จึงทำให้มีข้อมูลทั้งหมดที่ใช้ในการทดลองนี้ 120 ชุดข้อมูล

ผลการทดลอง

เมื่อนำชุดข้อมูลทั้งหมดที่ได้จากการจำลองข้อมูลมาทดสอบกับโปรแกรมสร้างหลักเกณฑ์แอสโซซิเอชัน โดยกำหนดให้ค่า minimum support มีค่าเท่ากับ 20% จะได้หลักเกณฑ์แอสโซซิเอชันทั้งหมดที่มีค่าสนับสนุนของหลักเกณฑ์มากกว่าหรือเท่ากับ 20% และแสดงค่าความเชื่อมั่น ค่าความเชื่อมั่นประพจน์แย้งสลับที่ จำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) จำนวนรายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p (x_2) จำนวนรายการที่ปรากฏสินค้า p และ q (x_3) และจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) ของแต่ละหลักเกณฑ์และแสดงผลเรียงลำดับตามค่าความเชื่อมั่นของหลักเกณฑ์คือเรียงจากหลักเกณฑ์ที่มีค่าความเชื่อมั่นน้อยไปหาหลักเกณฑ์ที่มีค่าความเชื่อมั่นมาก ตัวอย่างของผลลัพธ์ที่ได้จากการทำงานของโปรแกรมของข้อมูล 1 ชุด แสดงได้ดังรูปที่ 4.3

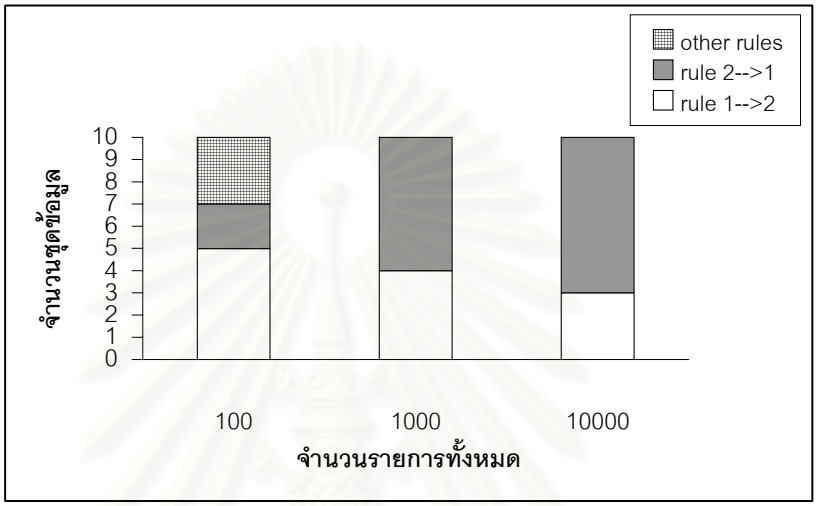
รูปที่ 4.3: ผลลัพธ์ที่ได้จากการทำงานของโปรแกรมของข้อมูล 1 ชุด

rule	conf	contra_conf	x_1	x_2	x_3	x_4
3-->4	69.06	85.94	967	963	2158	5912
8-->7	69.10	86.03	964	943	2156	5937
4-->3	69.14	85.99	963	967	2158	5912
6-->5	69.16	86.13	961	914	2156	5969
7-->8	69.57	86.29	943	964	2156	5937
5-->6	70.22	86.72	914	961	2156	5969
1-->2	73.68	89.75	738	727	2067	6468

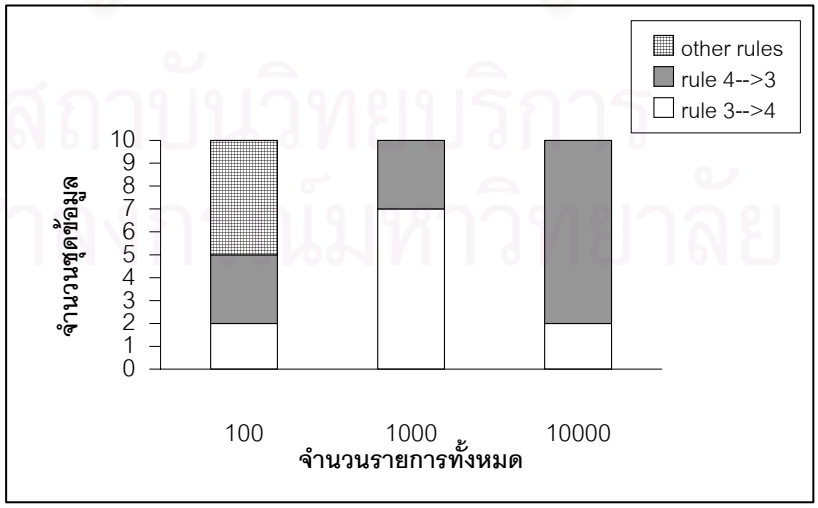
จากผลการทดลองที่ได้นำมาสร้างแผนภูมิแท่ง เพื่อแสดงให้เห็นว่าการเพิ่มค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์นั้นมีส่วนช่วยในการทำให้หลักเกณฑ์นั้นมีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดหรือไม่เมื่อเทียบกับหลักเกณฑ์อื่นๆ ที่ไม่ได้ทำการเพิ่มค่า

สนับสนุนประพจน์แย้งสลับที่จากเงื่อนไขในบรรทัดที่ 31 ของโปรแกรมในรูปที่ 4.1 โดยผลที่ได้แสดงในรูปของแผนภูมิแท่งในรูปที่ 4.4 - 4.7

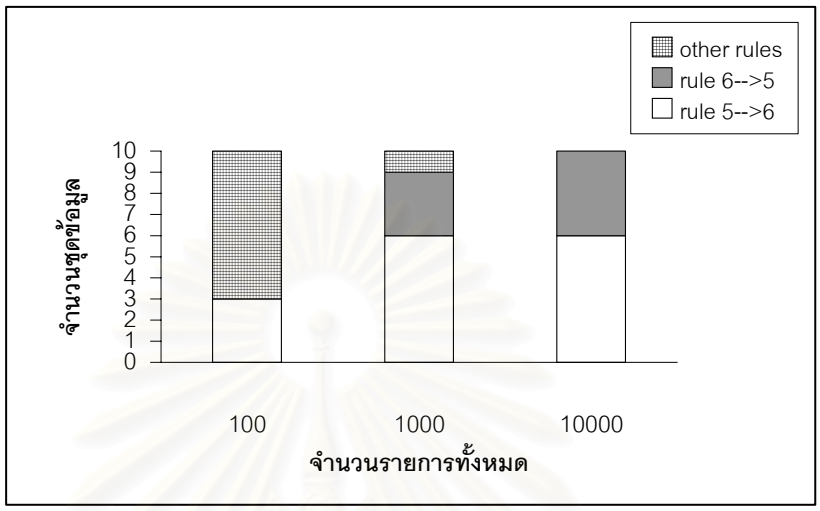
รูปที่ 4.4: แผนภูมิแสดงจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดของข้อมูลในกลุ่มที่ 1



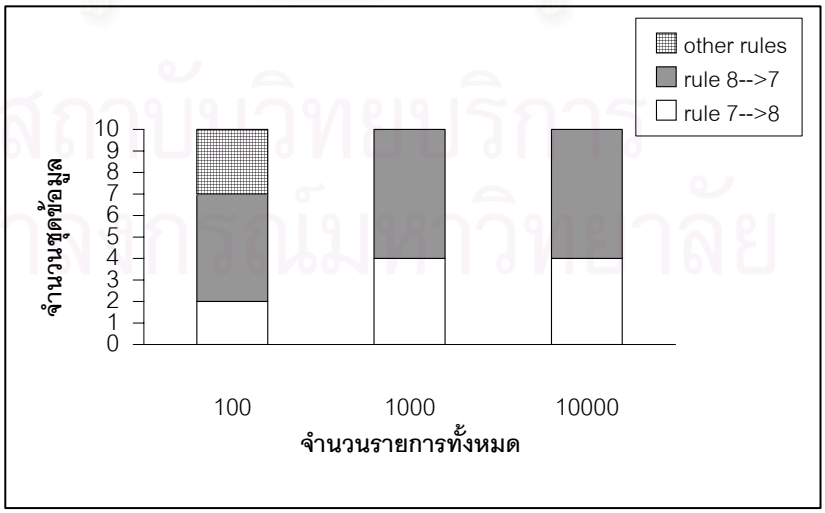
รูปที่ 4.5: แผนภูมิแสดงจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดของข้อมูลในกลุ่มที่ 2



รูปที่ 4.6: แผนภูมิแสดงจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดของข้อมูลในกลุ่มที่ 3



รูปที่ 4.7: แผนภูมิแสดงจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุดของข้อมูลในกลุ่มที่ 4



แผนภูมิในรูปที่ 4.4 – 4.7 แสดงจำนวนชุดข้อมูลของแต่ละหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์สูงสุดในแต่ละชุดข้อมูล ซึ่งแต่ละแผนภูมิแสดงผลที่ได้เปรียบเทียบกันตามจำนวนรายการทั้งหมดจากรูปแบบละ 10 ชุดข้อมูล

ผลการทดลองพบว่าเมื่อจำนวนรายการทั้งหมดมี 100 รายการ จำนวนชุดข้อมูลที่มีหลักเกณฑ์ 1 $\rightarrow$ 2 เป็นหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดมีจำนวน 5 ชุดข้อมูล จำนวนชุดข้อมูลที่มีหลักเกณฑ์ 2 $\rightarrow$ 1 เป็นหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดมีจำนวน 2 ชุดข้อมูล และจำนวนชุดข้อมูลที่มีหลักเกณฑ์อื่นๆ ที่ไม่ใช่หลักเกณฑ์ 1 $\rightarrow$ 2 หรือ 2 $\rightarrow$ 1 เป็นหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดมีจำนวน 3 ชุดข้อมูล เมื่อจำนวนรายการทั้งหมดมี 1,000 และ 10,000 รายการ พบว่าหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดในแต่ละรูปแบบมี 2 หลักเกณฑ์ คือ หลักเกณฑ์ 1 $\rightarrow$ 2 หรือ 2 $\rightarrow$ 1 โดยรูปแบบที่จำนวนรายการทั้งหมดมี 1,000 รายการ หลักเกณฑ์ 1 $\rightarrow$ 2 และ 2 $\rightarrow$ 1 มีจำนวน 4 ชุดข้อมูลและ 6 ชุดข้อมูลตามลำดับ และรูปแบบที่จำนวนรายการทั้งหมดมี 10,000 รายการ หลักเกณฑ์ 1 $\rightarrow$ 2 และ 2 $\rightarrow$ 1 มีจำนวน 3 ชุดข้อมูลและ 7 ชุดข้อมูลตามลำดับ

เมื่อพิจารณาหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดในทุกชุดข้อมูล พบว่าหลักเกณฑ์นั้นมีค่าความเชื่อมั่นและมีจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) สูงสุดด้วยเมื่อเทียบกับหลักเกณฑ์อื่นๆ ที่ได้จากการทดสอบข้อมูล 1 ชุด ซึ่งค่า x_4 เมื่อนำมาหารด้วยจำนวนรายการทั้งหมดก็คือค่าสนับสนุนประพจน์แย้งสลับที่ เนื่องจากในข้อมูลกลุ่มที่ 1 ผู้วิจัยกำหนดให้เซตสินค้า {1,2} มีค่าสนับสนุนประพจน์แย้งสลับที่ของเซตสินค้ามากที่สุด ดังนั้นจึงทำให้ผลที่ได้ในการทดสอบกับข้อมูลในกลุ่มที่ 1 ที่แสดงในแผนภูมิรูปที่ 4.4 - 4.5 พบว่าหลักเกณฑ์ 1 $\rightarrow$ 2 หรือ 2 $\rightarrow$ 1 มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดเป็นส่วนมากจากจำนวนข้อมูลทั้งหมด 10 ชุด โดยเฉพาะเมื่อทดสอบกับข้อมูลที่มีจำนวนรายการทั้งหมดเป็น 1,000 และ 10,000 รายการ ทั้งนี้เนื่องจากค่าสนับสนุนประพจน์แย้งสลับที่ (x_4) ของหลักเกณฑ์ 1 $\rightarrow$ 2 และ 2 $\rightarrow$ 1 มีค่าเท่ากัน ดังนั้นค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเณฑ์ในกรณีที่หลักเกณฑ์ทั้งสองมีค่าเท่ากันจึงขึ้นอยู่กับค่า x_1 ของแต่ละหลักเกณฑ์ตามสมการที่ 2 ในบทที่ 3 คือถ้าหลักเกณฑ์ใดมีค่า x_1 น้อยหลักเกณฑ์นั้นจะมีค่าความเชื่อมั่นประพจน์แย้งสลับที่มาก ดังนั้น

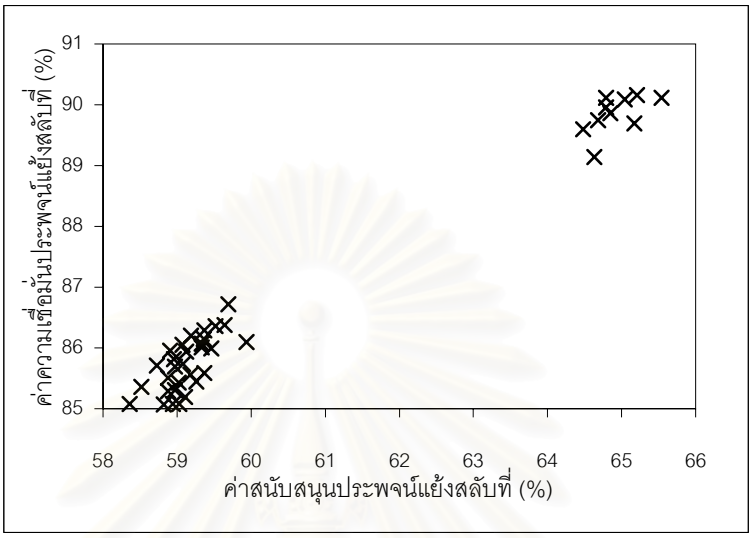
ถ้าหลักเกณฑ์ 1 $\longrightarrow$ 2 มีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากกว่าหลักเกณฑ์ 2 $\longrightarrow$ 1 แสดงว่าค่า x_1 ของหลักเกณฑ์ 1 $\longrightarrow$ 2 (จำนวนรายการที่ปรากฏสินค้า 1 และไม่ปรากฏสินค้า 2) มีค่าน้อยกว่า x_1 ของหลักเกณฑ์ 2 $\longrightarrow$ 1 (จำนวนรายการที่ปรากฏสินค้า 2 และไม่ปรากฏสินค้า 1)

พิจารณาหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดในข้อมูลกลุ่มที่ 2 เมื่อจำนวนรายการทั้งหมดมี 1,000 และ 10,000 รายการ พบว่าหลักเกณฑ์ที่ให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดคือหลักเกณฑ์ 3 $\longrightarrow$ 4 หรือ 4 $\longrightarrow$ 3 ในข้อมูลกลุ่มที่ 3 และกลุ่มที่ 4 เมื่อจำนวนรายการทั้งหมดมี 1,000 และ 10,000 รายการ พบว่าหลักเกณฑ์ที่ให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดคือหลักเกณฑ์ 5 $\longrightarrow$ 6 หรือ 6 $\longrightarrow$ 5 และ 7 $\longrightarrow$ 8 หรือ 8 $\longrightarrow$ 7 ตามลำดับ

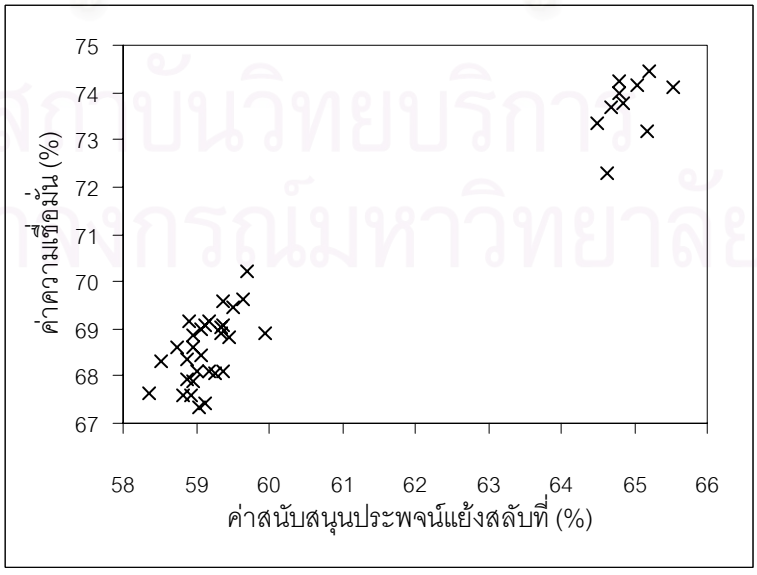
ผลที่ได้จากข้อมูลทั้ง 4 กลุ่ม พบว่าหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดของข้อมูลแต่ละกลุ่มคือ p $\longrightarrow$ q หรือ q $\longrightarrow$ p เมื่อ p และ q เป็นสมาชิกของเซตสินค้าที่กำหนดไว้ในแต่ละกลุ่ม ที่เป็นเช่นนี้เพราะเซตสินค้าที่กำหนดในแต่ละกลุ่มมีค่าสนับสนุนประพจน์แย้งสลับที่มากกว่าเซตสินค้าอื่นๆที่เกิดจากการสุ่มตัวเลขในโปรแกรมการจำลองข้อมูล ดังนั้นจึงสรุปได้ว่า ค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์มีส่วนช่วยทำให้หลักเกณฑ์นั้นมีค่าความเชื่อมั่นประพจน์แย้งสลับที่และค่าความเชื่อมั่นของหลักเกณฑ์สูงขึ้น แต่เนื่องจากหลักเกณฑ์ p $\longrightarrow$ q และ q $\longrightarrow$ p มีค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์เท่ากัน ดังนั้นค่าความเชื่อมั่นประพจน์แย้งสลับที่จึงขึ้นอยู่กับจำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) ของแต่ละหลักเกณฑ์ โดยหลักเกณฑ์ที่มีค่า x_1 น้อยที่สุดหลักเกณฑ์นั้นจะมีค่าความเชื่อมั่นประพจน์แย้งสลับที่มากที่สุด

เมื่อนำผลการทดลองที่ได้มาแสดงแนวโน้มความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่และค่าความเชื่อมั่นของหลักเกณฑ์จะได้กราฟดังรูปที่ 4.8 - 4.15 โดยพิกัดในกราฟคือค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่และค่าความเชื่อมั่นของหลักเกณฑ์ 1 $\longrightarrow$ 2, 3 $\longrightarrow$ 4, 5 $\longrightarrow$ 6 และ 7 $\longrightarrow$ 8 จากข้อมูลทั้งหมด 10 ชุด ดังนั้นจึงมีพิกัดทั้งหมด 40 พิกัด และพิจารณาเฉพาะรูปแบบที่มีจำนวนรายการทั้งหมด 10,000 รายการ

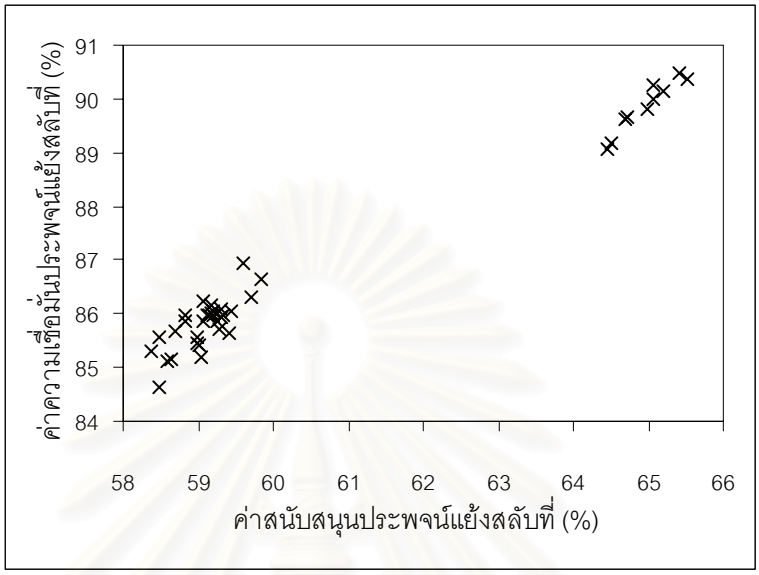
รูปที่ 4.8: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 1



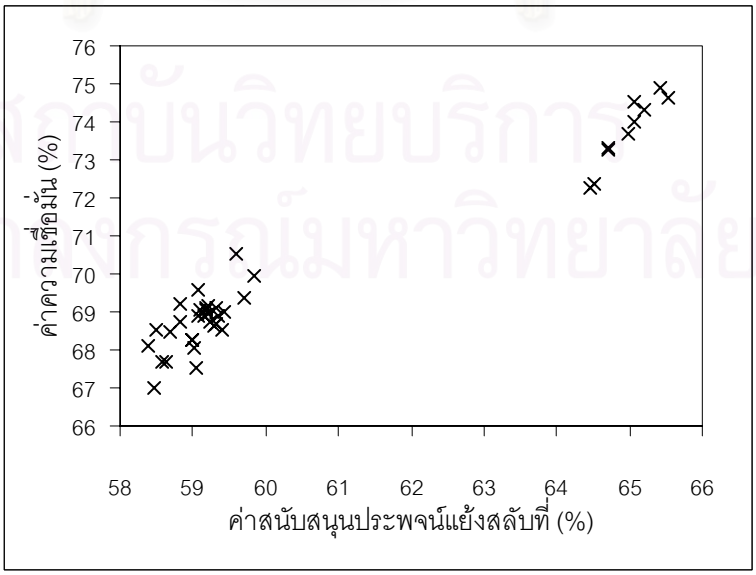
รูปที่ 4.9: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 1



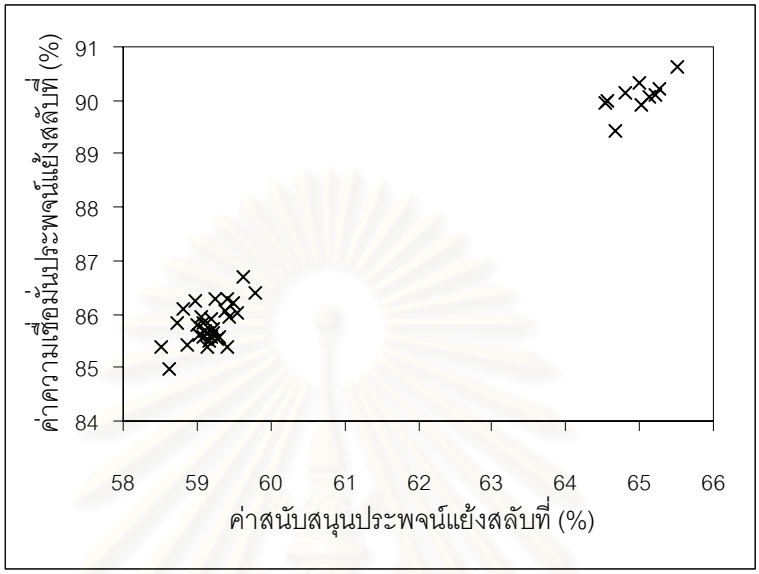
รูปที่ 4.10: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 2



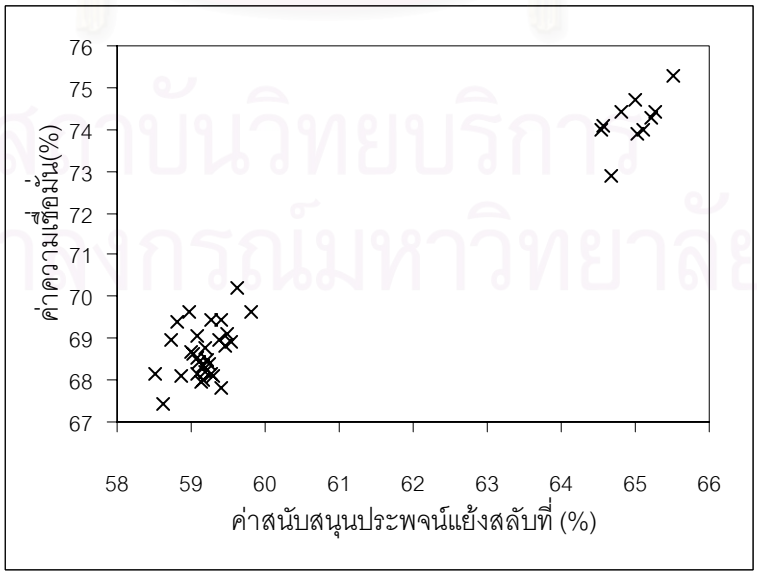
รูปที่ 4.11: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 2



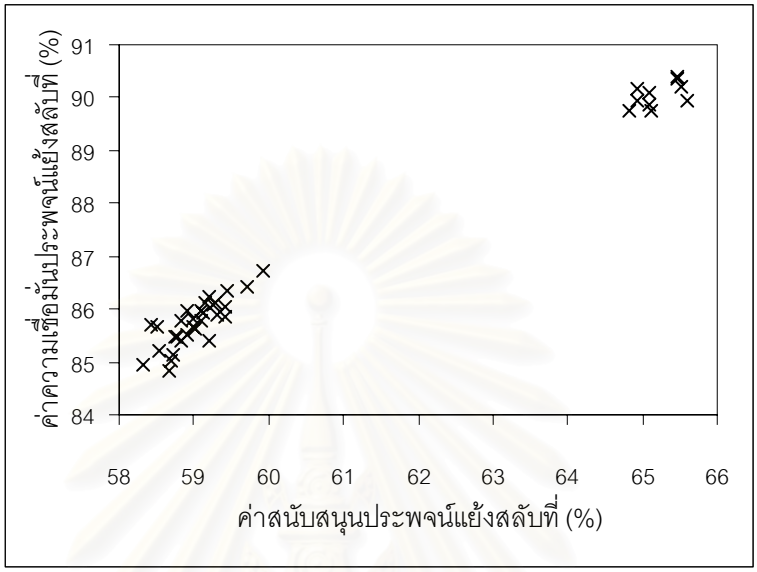
รูปที่ 4.12: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 3



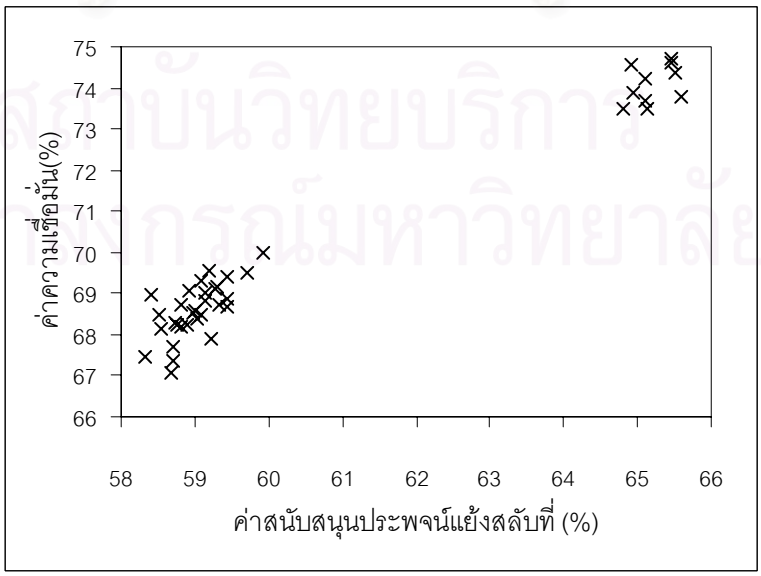
รูปที่ 4.13: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 3



รูปที่ 4.14: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 4



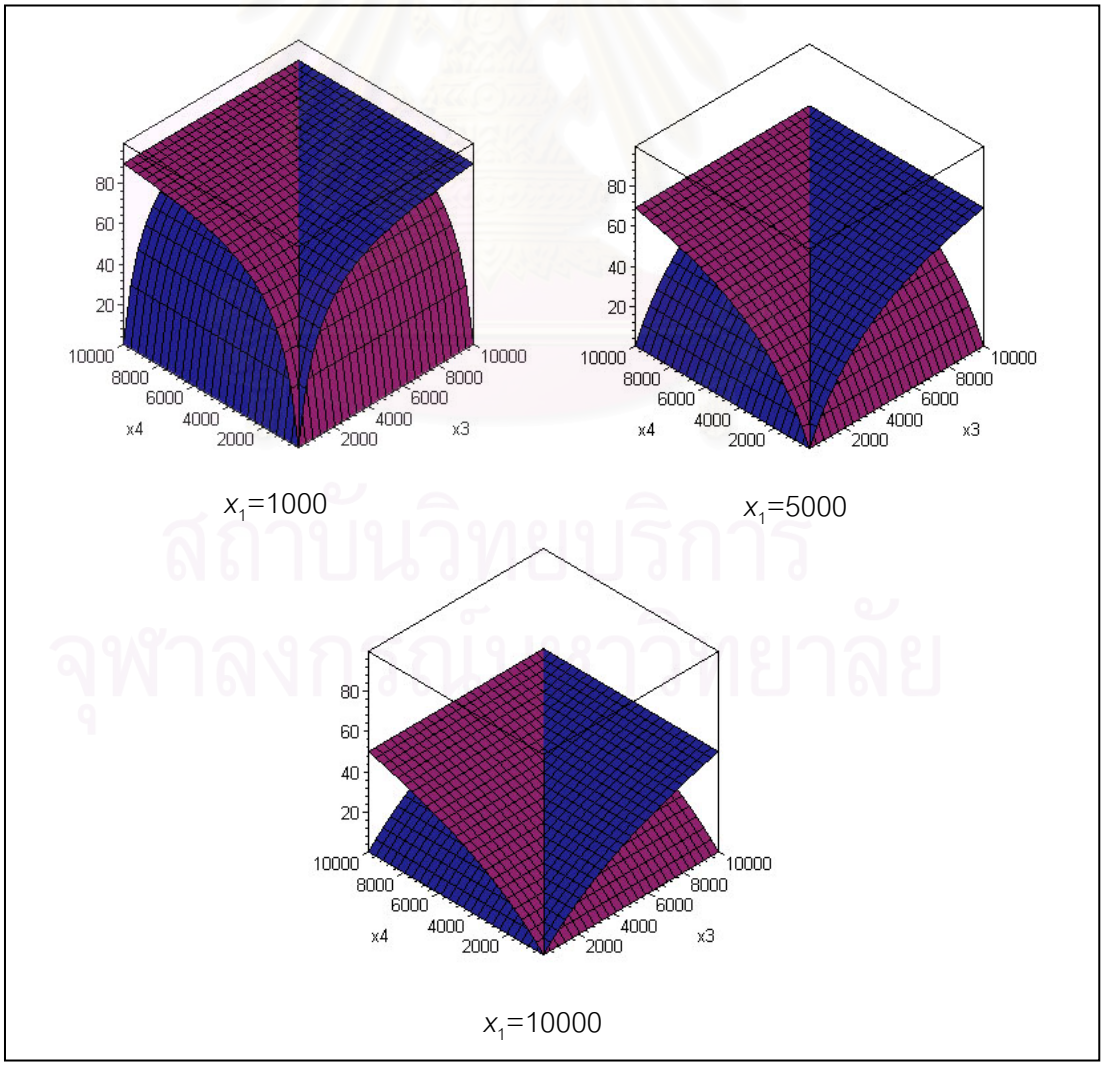
รูปที่ 4.15: กราฟแสดงแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นของแต่ละหลักเกณฑ์ เมื่อใช้ข้อมูลกลุ่มที่ 4



สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

จากกราฟแสดงความสัมพันธ์ระหว่างค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นประพจน์แย้งสลับที่ พบว่าแนวโน้มของค่าทั้งสองมีความสัมพันธ์กันแบบแปรผันตรง คือ เมื่อค่าสนับสนุนประพจน์แย้งสลับที่มีค่าเพิ่มขึ้นค่าความเชื่อมั่นประพจน์แย้งสลับที่ก็จะมีแนวโน้มเพิ่มขึ้นด้วย และแนวโน้มของค่าสนับสนุนประพจน์แย้งสลับที่กับค่าความเชื่อมั่นมีความสัมพันธ์กันแบบแปรผันตรงคือ เมื่อค่าสนับสนุนประพจน์แย้งสลับที่มีค่าเพิ่มขึ้นค่าความเชื่อมั่นก็จะมีแนวโน้มเพิ่มขึ้นด้วย และเมื่อเปรียบเทียบความสัมพันธ์ระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ พบว่าเมื่อหลักเกณฑ์ที่มีค่าความเชื่อมั่นเพิ่มขึ้นหลักเกณฑ์นั้นจะมีค่าความเชื่อมั่นประพจน์แย้งสลับที่เพิ่มขึ้นด้วย และจากหลักเกณฑ์ทั้งหมดพบว่าในหลักเกณฑ์เดียวกันค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์มีค่ามากกว่าค่าความเชื่อมั่นของหลักเกณฑ์

รูปที่ 4.16: **กราฟ 3 มิติแสดงความสัมพันธ์ระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่**



รูปที่ 4.16 คือกราฟ 3 มิติแสดงความสัมพันธ์ระหว่างค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่เมื่อจำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) มีค่าต่างกันคือมีค่า 1,000 5,000 และ 10,000 รายการตามลำดับ โดยบริเวณพื้นที่สีน้ำเงินคือค่าความเชื่อมั่นและบริเวณพื้นที่สีแดงคือค่าความเชื่อมั่นประพจน์แย้งสลับที่ซึ่งค่าทั้งสองมีหน่วยเป็นเปอร์เซ็นต์

จากกราฟพบว่าเมื่อจำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) มีค่าเพิ่มขึ้นค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่มีค่าลดลง โดยค่าความเชื่อมั่นทั้งสองนี้ขึ้นอยู่กับจำนวนรายการที่ปรากฏสินค้า p และ q (x_3) และจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) คือถ้าจำนวนรายการที่ปรากฏสินค้า p และ q (x_3) มีค่าเพิ่มขึ้นค่าความเชื่อมั่นจะมีค่าเพิ่มขึ้น ในกรณีเดียวกันเมื่อจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) มีค่าเพิ่มขึ้นค่าความเชื่อมั่นประพจน์แย้งสลับที่จะมีค่าเพิ่มขึ้นด้วย เมื่อพิจารณาแนวโน้มของค่าความเชื่อมั่นทั้งสองจากกราฟพบว่าถ้าจำนวนรายการที่ปรากฏสินค้า p และ q (x_3) มีค่ามากกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) ค่าความเชื่อมั่นจะมีค่าสูงกว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ แต่ถ้าจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) มีค่ามากกว่าจำนวนรายการที่ปรากฏสินค้า p และ q (x_3) ค่าความเชื่อมั่นประพจน์แย้งสลับที่จะมีค่าสูงกว่าค่าความเชื่อมั่นและถ้าจำนวนรายการที่ปรากฏสินค้า p และ q (x_3) มีค่าเท่ากับจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) พบว่าค่าความเชื่อมั่นมีค่าเท่ากับค่าความเชื่อมั่นประพจน์แย้งสลับที่

ในบทที่ 5 จะวิเคราะห์สาเหตุที่ทำให้ค่าความเชื่อมั่นของหลักเกณฑ์แอสโซซิเอชันของแต่ละหลักเกณฑ์มีค่าเท่ากันและเปรียบเทียบค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากัน เมื่อจำนวนรายการ x_1, x_2, x_3 , x_4 ของแต่ละหลักเกณฑ์มีค่าต่างกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

การวิเคราะห์ค่าความเชื่อมั่นประพจน์แย้งสลับที่

ในบทนี้จะวิเคราะห์สาเหตุที่ทำให้ค่าความเชื่อมั่นของหลักเกณฑ์แอสโซซิเอชันของแต่ละหลักเกณฑ์มีค่าเท่ากันและเปรียบเทียบค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากัน เมื่อจำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) จำนวนรายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p (x_2) จำนวนรายการที่ปรากฏสินค้า p และ q (x_3) และจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) ของแต่ละหลักเกณฑ์มีค่าต่างกัน

สาเหตุที่ทำให้หลักเกณฑ์แอสโซซิเอชันมีค่าความเชื่อมั่นเท่ากัน

พิจารณาค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q = \dfrac{1}{1+\frac{x_1}{x_3}}$ (1)

พิจารณาค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p \rightarrow q = \dfrac{1}{1+\frac{x_1}{x_4}}$ (2)

พิจารณาค่าความเชื่อมั่นของหลักเกณฑ์ $q \rightarrow p = \dfrac{1}{1+\frac{x_2}{x_3}}$ (3)

พิจารณาค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $q \rightarrow p = \dfrac{1}{1+\frac{x_2}{x_4}}$ (4)

จากสมการ (1) และ (3) พบว่าถ้าค่าความเชื่อมั่นของหลักเกณฑ์ $p \rightarrow q$ มีค่าเท่ากับค่าความเชื่อมั่นของหลักเกณฑ์ $q \rightarrow p$ ทำให้จำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) มีค่าเท่ากับจำนวนรายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p (x_2) นั่นคือ $x_1 = x_2$ ดังนั้นเมื่อพิจารณาค่าความเชื่อมั่นประพจน์แย้งสลับที่ของทั้งสองหลักเกณฑ์จากสมการ (2) และ (4) พบว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ของทั้งสองหลักเกณฑ์มีค่าเท่ากัน ดังนั้นในบทนี้จึงไม่พิจารณาหลักเกณฑ์ $q \rightarrow p$

เนื่องจากเป้าหมายของงานวิจัยนี้ คือการคัดเลือกหลักเกณฑ์แอสโซซิเอชันจากหลักเกณฑ์ทั้งหมดที่มีค่าความเชื่อมั่นของหลักเกณฑ์เท่ากัน ดังนั้นในส่วนนี้จึงวิเคราะห์ถึงสาเหตุที่ทำให้ค่าความเชื่อมั่นของหลักเกณฑ์มีค่าเท่ากัน

กำหนดให้หลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากันมีจำนวน n หลักเกณฑ์และแต่ละหลักเกณฑ์มีรูปแบบดังนี้ $p_1 \rightarrow q_1$, $p_2 \rightarrow q_2$,..., $p_n \rightarrow q_n$ เมื่อ p_1, p_2,..., p_n แทนสินค้าที่อยู่ทางด้านซ้ายของหลักเกณฑ์ q_1, q_2,..., q_n แทนสินค้าที่อยู่ทางด้านขวาของหลักเกณฑ์

เนื่องจากในการเปรียบเทียบค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากันจะต้องนำหลักเกณฑ์ 2 หลักเกณฑ์มาเปรียบเทียบกัน ดังนั้นในบทนี้จะทำการวิเคราะห์ระหว่างหลักเกณฑ์ $p_1 \rightarrow q_1$ และ $p_2 \rightarrow q_2$ โดยกำหนดให้

x_1^1 แทนจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1

x_1^2 แทนจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2

x_2^1 แทนจำนวนรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

x_2^2 แทนจำนวนรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2

x_3^1 แทนจำนวนรายการที่ปรากฏสินค้า p_1 และ q_1

x_3^2 แทนจำนวนรายการที่ปรากฏสินค้า p_2 และ q_2

x_4^1 แทนจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1

x_4^2 แทนจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2

พิจารณาค่าความเชื่อมั่นของหลักเกณฑ์ $p_1 \rightarrow q_1 = \dfrac{1}{1+\dfrac{x_1^1}{x_3^1}}$ (5)

และค่าความเชื่อมั่นของหลักเกณฑ์ $p_2 \rightarrow q_2 = \dfrac{1}{1+\dfrac{x_1^2}{x_3^2}}$ (6)

และค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_1 \rightarrow q_1 = \dfrac{1}{1+\dfrac{x_1^1}{x_4^1}}$ (7)

และค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2 = \dfrac{1}{1+\dfrac{x_1^2}{x_4^2}}$ (8)

จากสมการ (5) และ (6) พบว่ามี 3 กรณีที่ทำให้ค่า ความเชื่อมั่นของหลักเกณฑ์มีค่าเท่ากันดังนี้

1. $x_1^2 = x_1^1$ และ $x_3^2 = x_3^1$

2. $x_1^2 > x_1^1$, $x_3^2 > x_3^1$ และ $\dfrac{x_1^1}{x_3^1} = \dfrac{x_1^2}{x_3^2}$

3. $x_1^2 < x_1^1$, $x_3^2 < x_3^1$ และ $\dfrac{x_1^1}{x_3^1} = \dfrac{x_1^2}{x_3^2}$

กรณีที่ 1: $x_1^2 = x_1^1$ **และ** $x_3^2 = x_3^1$

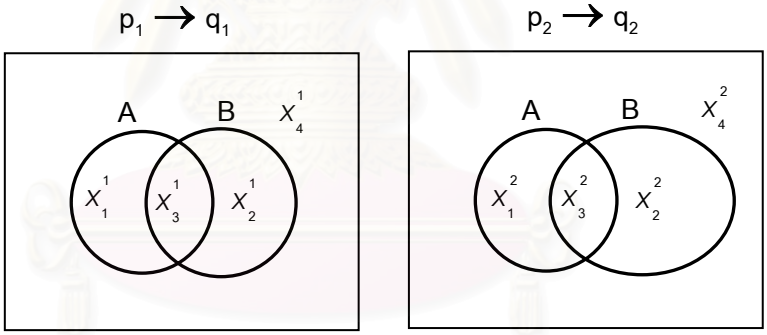
ในกรณีหมายถึงจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 มีค่าเท่ากับจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 และจำนวนรายการที่ปรากฏสินค้า p_1 และ q_1 มีค่าเท่ากับจำนวนรายการที่ปรากฏสินค้า p_2 และ q_2

เนื่องจากกรณีนี้ในแต่ละหลักเกณฑ์ $x_1^2 = x_1^1$ และ $x_3^2 = x_3^1$ ดังนั้นจะทำการวิเคราะห์ค่าความเชื่อมั่นประพจน์แย้งสลับที่จากความสัมพันธ์ระหว่างค่า x_2^1 และ x_2^2 ดังนี้

1.1 **เมื่อ** $x_2^2 > x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

รูปที่ 5.1: **กรณีที่** $x_2^2 > x_2^1$



จากรูปที่ 5.1 |U| ของทั้ง 2 หลักเกณฑ์มีค่าเท่ากัน $x_1^1 = x_1^2$ และ $x_3^1 = x_3^2$ ดังนั้น

$$x_2^1 + x_4^1 = x_2^2 + x_4^2$$
$$x_2^2 - x_2^1 = x_4^1 - x_4^2 \qquad (9)$$

เนื่องจาก $x_2^2 > x_2^1$ ดังนั้น $x_2^2 - x_2^1 > 0$

จากสมการ (9) ทำให้ $x_4^1 > x_4^2$

จากสมการ (7) และ (8) เมื่อ $x_1^1 = x_1^2$ และ $x_4^1 > x_4^2$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

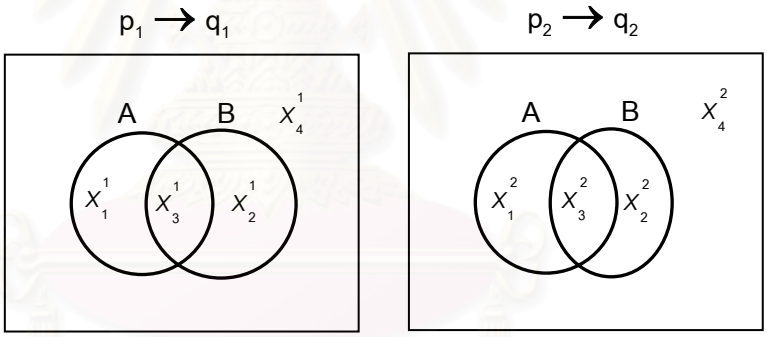
ดังนั้นเมื่อเปรียบเทียบค่าต่างๆ ของหลักเกณฑ์ $p_2 \rightarrow q_2$ กับหลักเกณฑ์ $p_1 \rightarrow q_1$ เมื่อรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1 ($x_2^2 > x_2^1$) พบว่า

- รายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนน้อยกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($x_4^2 < x_4^1$)
- ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

1.2 **เมื่อ** $x_2^2 < x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

รูปที่ 5.2: **กรณีที่** $x_2^2 < x_2^1$



เมื่อ $x_2^2 < x_2^1$ ดังนั้น $x_2^2 - x_2^1 < 0$

จากรูปที่ 5.2 และสมการ (9) ทำให้ $x_4^1 < x_4^2$

จากสมการ (7) และ (8) เมื่อ $x_1^1 = x_1^2$ และ $x_4^1 < x_4^2$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ดังนั้นเมื่อเปรียบเทียบค่าต่างๆ ของหลักเกณฑ์ $p_2 \rightarrow q_2$ กับหลักเกณฑ์ $p_1 \rightarrow q_1$ เมื่อรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1 ($x_2^2 < x_2^1$) พบว่า

- รายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนมากกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($x_4^2 > x_4^1$)
- ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

1.3 **เมื่อ** $x_2^2 = x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

เมื่อ $x_2^2 = x_2^1$ ดังนั้น $x_2^2 - x_2^1 = 0$

จากสมการ (9) ทำให้ $x_4^1 = x_4^2$

จากสมการ (7) และ (8) เมื่อ $x_1^1 = x_1^2$ และ $x_4^1 = x_4^2$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าเท่ากับหลักเกณฑ์ $p_1 \rightarrow q_1$

ดังนั้นเมื่อเปรียบเทียบค่าต่างๆ ของหลักเกณฑ์ $p_2 \rightarrow q_2$ กับหลักเกณฑ์ $p_1 \rightarrow q_1$ เมื่อรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1 ($x_2^2 = x_2^1$) พบว่า

- รายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนมากกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($x_4^1 = x_4^2$)
- ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าเท่ากับหลักเกณฑ์ $p_1 \rightarrow q_1$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

กรณีที่ 2: $x_1^2 > x_1^1$, $x_3^2 > x_3^1$ **และ** $\frac{x_1^1}{x_3^1} = \frac{x_1^2}{x_3^2}$

ในกรณีหมายถึงรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 ($x_1^2 > x_1^1$) และรายการที่ปรากฏสินค้า p_2 และ q_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า p_1 และ q_1 ($x_3^2 > x_3^1$) โดยที่อัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_2 และ q_2 มีค่าเท่ากับอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 กับจำนวนรายการที่ปรากฏสินค้า p_1 และ q_1 ($\frac{x_1^1}{x_3^1} = \frac{x_1^2}{x_3^2}$)

กำหนดให้ $y_1 = x_1^1 + x_3^1$ และ $y_2 = x_1^2 + x_3^2$ เนื่องจาก

$|U| = x_1^1 + x_2^1 + x_3^1 + x_4^1$

$= x_1^2 + x_2^2 + x_3^2 + x_4^2$ จะได้ว่า

$y_1 + x_2^1 + x_4^1 = y_2 + x_2^2 + x_4^2$

$x_4^1 - x_4^2 = (x_2^2 - x_2^1) + (y_2 - y_1)$ (10)

กำหนดให้ x_1^2 มีค่าเท่ากับ $m_1 x_1^1$ โดยที่ $m_1 > 1$ เพราะว่า $x_1^2 > x_1^1$ ดังนั้นจะทำการวิเคราะห์ค่าความเชื่อมั่นประพจน์แย้งสลับที่จากความสัมพันธ์ระหว่างค่า x_2^1 และ x_2^2 ดังนี้

2.1 **เมื่อ** $x_2^2 > x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

เนื่องจาก $x_1^2 > x_1^1$ และ $x_3^2 > x_3^1$ ทำให้ $y_2 > y_1$ ดังนั้น $y_2 - y_1 > 0$

และเนื่องจาก $x_2^2 > x_2^1$ ดังนั้น $x_2^2 - x_2^1 > 0$

จากสมการ (10) ทำให้ $x_4^1 > x_4^2$

จากสมการ (7) และ (8) เมื่อ $x_1^2 > x_1^1$ และ $x_4^1 > x_4^2$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ดังนั้นเมื่อเปรียบเทียบค่าต่างๆ ของหลักเกณฑ์ $p_2 \rightarrow q_2$ กับหลักเกณฑ์ $p_1 \rightarrow q_1$ เมื่อรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า x_{21}^2 (x_2^1 >) พบว่า

- รายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนน้อยกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($x_4^2 < x_4^1$)
- ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

2.2 **เมื่อ** $x_2^2 < x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

กำหนดให้ $a_1 = y_2 - y_1$ ซึ่งหมายความว่า a_1 = (จำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 + จำนวนรายการที่ปรากฏสินค้า p_2 และ q_2) – (จำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 + จำนวนรายการที่ปรากฏสินค้า p_1 และ q_1)

กำหนดให้ $b_1 = x_2^1 - x_2^2$ ซึ่งหมายความว่า b_1 = จำนวนรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1 - จำนวนรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2

จาก (10) จะได้ $x_4^1 - x_4^2 = a_1 - b_1$ (11)

ถ้า $a_1 = b_1$ จากสมการ (11) ทำให้ $x_4^2 = x_4^1$ โดยสมการ (7) และ (8) เมื่อ $x_1^2 > x_1^1$ และ $x_4^2 = x_4^1$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตัวอย่างเช่น กำหนดให้

จำนวนรายการทั้งหมดในฐานข้อมูล (|U|) =100

จำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า $q_1(x_1^1)$ =10

จำนวนรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า $p_1(x_2^1)$ =30

จำนวนรายการที่ปรากฏสินค้า p_1 และ q_1 (x_3^1) =20

จำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 (x_4^1) =40

จะได้ 100=10+30+20+40 (12)

ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_1 \rightarrow q_1$ = 80% กำหนดให้ $a_1 = b_1$ =15 และ m_1 = 3/2 ดังนั้น x_1^2 =15 x_2^2 =15 x_3^2=30 และ x_4^2 =40 จะได้ 100=15+15+30+40 ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ = 72.72%

ถ้า $a_1 < b_1$ ดังนั้น $a_1 - b_1 < 0$ จากสมการ (12) ทำให้ $x_4^2 > x_4^1$ จากสมการ (7) และ (8) เมื่อ $x_1^2 > x_1^1$ และ $x_4^2 > x_4^1$ ทำให้ไม่สามารถเปรียบเทียบค่าความเชื่อมั่นประพจน์แย้งสลับที่ได้เนื่องจากอัตราส่วนระหว่าง x_1^2 กับ x_1^1 เมื่อเทียบกับอัตราส่วนระหว่าง x_4^2 กับ x_4^1 มีหลายกรณีและในแต่ละกรณีค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าต่างกัน

กำหนดให้ x_4^2 มีค่าเท่ากับ $m_2 x_4^1$ โดยที่ $m_2 > 1$ เพราะว่า $x_4^2 > x_4^1$ ดังนั้นความสัมพันธ์ระหว่างอัตราส่วนของ x_1^2 กับ x_1^1 และอัตราส่วนของ x_4^2 กับ x_4^1 มีดังนี้

- $m_1 > m_2$ หมายความว่าอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 มีค่ามากกว่าอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 กับจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(\frac{x_1^2}{x_1^1} > \frac{x_4^2}{x_4^1})$ จากสมการ (7) และ (8) ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่า น้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

  ตัวอย่างเช่นจากสมการ (13) กำหนดให้ a_1=15 b_1=25 m_1=3/2 m_2=5/4 ดังนั้น x_1^2=15 x_2^2=5 x_3^2 = 30 และ x_4^2 =50 จะได้

100 =15+5+30+50 ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ =76.92%

- $m_2 > m_1$ หมายความว่าอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 มีค่าน้อยกว่าอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 กับจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($\frac{x_1^2}{x_1^1} < \frac{x_4^2}{x_4^1}$) จากสมการ (7) และ (8) ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่า มากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

  ตัวอย่างเช่นจากสมการ (12) กำหนดให้ a_1=6 b_1=22 m_1=6/5 m_2=7/5 ดังนั้น $x_1^2 = 12$ $x_2^2 = 8$ $x_3^2 = 24$ และ $x_4^2 = 56$ จะได้ 100=12+8+24+56 มีค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ =82.35%

- $m_1 = m_2$ หมายความว่าอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 มีค่าเท่ากับอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 กับจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($\frac{x_1^2}{x_1^1} = \frac{x_4^2}{x_4^1}$) จากสมการ (7) และ (8) ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่า เท่ากับหลักเกณฑ์ $p_1 \rightarrow q_1$

  ตัวอย่างเช่นจากสมการ (12) กำหนดให้ a_1=6 b_1=14 m_1=m_2=6/5 ดังนั้น x_1^2=12 x_2^2= 16 x_3^2 = 24 และ x_4^2 = 48 จะได้ 100=12+16+24+48 ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ =80%

ถ้า $a_1 > b_1$ จากสมการ (11) ทำให้ $x_4^2 < x_4^1$ จากสมการ (7) และ (8) เมื่อ $x_1^2 > x_1^1$ และ $x_4^2 < x_4^1$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ตัวอย่างเช่นจากสมการ (12) กำหนด a_1=15 b_1=10 m_1=3/2 ดังนั้น x_1^2= 15 x_2^2= 20 x_3^2= 30 และ x_4^2 = 35 จะได้ 100=15+20+30+35 มีค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$=70%

2.3 **เมื่อ** $x_2^2 = x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

เนื่องจาก $x_1^2 > x_1^1$ และ $x_3^2 > x_3^1$ ทำให้ $y_2 > y_1$ ดังนั้น $y_2 - y_1 > 0$

และเนื่องจาก $x_2^2 = x_2^1$ ดังนั้น $x_2^2 - x_2^1 = 0$

จากสมการ (10) ทำให้ $x_4^1 > x_4^2$

จากสมการ (7) และ (8) เมื่อ $x_1^2 > x_1^1$ และ $x_4^1 > x_4^2$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ดังนั้นเมื่อเปรียบเทียบค่าต่างๆ ของหลักเกณฑ์ $p_2 \rightarrow q_2$ กับหลักเกณฑ์ $p_1 \rightarrow q_1$ เมื่อรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1 ($x_2^2 = x_2^1$) พบว่า

- รายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนน้อยกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($x_4^2 < x_4^1$)
- ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

กรณีที่ 3: $x_1^2 < x_1^1$, $x_3^2 < x_3^1$ **และ** $\frac{x_1^1}{x_3^1} = \frac{x_1^2}{x_3^2}$

ในกรณีหมายถึงรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 ($x_1^2 < x_1^1$) และรายการที่ปรากฏสินค้า p_2 และ q_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า p_1 และ q_1 ($x_3^2 < x_3^1$) โดยที่อัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_2 และ q_2 มีค่าเท่ากับอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 กับจำนวนรายการที่ปรากฏสินค้า p_1 และ q_1 ($\frac{x_1^1}{x_3^1} = \frac{x_1^2}{x_3^2}$)

กำหนดให้ x_1^2 มีค่าเท่ากับ $m_1 x_1^1$ โดยที่ $0 < m_1 < 1$ เพราะว่า $x_1^2 < x_1^1$ ดังนั้นในกรณีนี้ทำการวิเคราะห์ค่าความเชื่อมั่นประพจน์แย้งสลับที่จากความสัมพันธ์ระหว่างค่า x_2^1 และ x_2^2 ดังนี้

3.1 **เมื่อ** $x_2^2 > x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

กำหนดให้ $a_2 = y_1 - y_2$ ซึ่งหมายความว่า a_2 = (จำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 + จำนวนรายการที่ปรากฏสินค้า p_1 และ q_1) – (จำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 + จำนวนรายการที่ปรากฏสินค้า p_2 และ q_2)

กำหนดให้ $b_2 = x_2^2 - x_2^1$ ซึ่งหมายความว่า b_2 = จำนวนรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 - จำนวนรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

จาก (10) จะได้ $x_4^1 - x_4^2 = b_2 - a_2$ (13)

ถ้า $a_2 = b_2$ จากสมการ (13) ทำให้ $x_4^2 = x_4^1$ จากสมการ (7) และ (8) เมื่อ $x_1^2 < x_1^1$ และ $x_4^2 = x_4^1$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ตัวอย่างเช่น กำหนดให้

จำนวนรายการทั้งหมดในฐานข้อมูล (|U|) =100

จำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า $q_1(x_1^1)$ =20

จำนวนรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า $p_1(x_2^1)$ =10

จำนวนรายการที่ปรากฏสินค้า p_1 และ q_1 (x_3^1) =40

จำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 (x_4^1) =30

จะได้ 100=20+10+40+30 (14)

ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_1 \rightarrow q_1 = 60\%$ กำหนดให้ $a_2 = b_2 = 15$ และ $m_1 = 3/4$ ดังนั้น $x_1^2 = 15$ $x_2^2 = 25$ $x_3^2 = 30$ และ $x_4^2 = 30$ จะได้ 100=15+25+30+30 ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2 = 66.67\%$

ถ้า $a_2 < b_2$ ดังนั้น $b_2 - a_2 < 0$ จากสมการ (13) ทำให้ $x_4^2 < x_4^1$ จากสมการ (7) และ (8) เมื่อ $x_1^2 < x_1^1$ และ $x_4^2 < x_4^1$ ทำให้ไม่สามารถเปรียบเทียบค่าความเชื่อมั่นประพจน์แย้งสลับที่ได้เนื่องจากอัตราส่วนระหว่าง x_1^2 กับ x_1^1 เมื่อเทียบกับอัตราส่วนระหว่าง x_4^2 กับ x_4^1 มีหลายกรณีและในแต่ละกรณีค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าต่างกัน

กำหนดให้ x_4^2 มีค่าเท่ากับ $m_2 x_4^1$ โดยที่ $0 < m_2 < 1$ เพราะว่า $x_4^2 < x_4^1$ ดังนั้นความสัมพันธ์ระหว่างอัตราส่วนของ x_1^2 กับ x_1^1 และอัตราส่วนของ x_4^2 กับ x_4^1 มีดังนี้

- $m_1 > m_2$ หมายความว่าอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 มีค่ามากกว่าอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 กับจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($\frac{x_1^2}{x_1^1} > \frac{x_4^2}{x_4^1}$) จากสมการ (7) และ (8) ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่า น้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$
  ตัวอย่างเช่นจากสมการ (14) กำหนดให้ $a_2=15$ $b_2=25$ $m_1=3/4$ $m_2=2/3$ ดังนั้น $x_1^2=15$ $x_2^2=35$ $x_3^2=30$ และ $x_4^2=20$ จะได้ 100 =15+35+30+20 ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2 = 57.14\%$

- $m_2 > m_1$ หมายความว่าอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 มีค่าน้อยกว่าอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 กับจำนวนรายการที่ไม่ปรากฏ

สินค้า p_1 หรือ q_1 ($\frac{x_1^2}{x_1^1} < \frac{x_4^2}{x_4^1}$) จากสมการ (7) และ (8) ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ตัวอย่างเช่นจากสมการ(14) กำหนดให้ a_2=18 b_2=20 m_1=7/10 m_2=14/15 ดังนั้น x_1^2 =14 x_2^2 = 30 x_3^2 = 28 และ x_4^2 = 28 จะได้ 100=14+30+28+28 มีค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ =66.67%

- $m_1 = m_2$ หมายความว่าอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 มีค่าเท่ากับอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 กับจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1($\frac{x_1^2}{x_1^1} = \frac{x_4^2}{x_4^1}$) จากสมการ (7) และ (8) ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าเท่ากับหลักเกณฑ์ $p_1 \rightarrow q_1$

  ตัวอย่างเช่นจากสมการ (14) กำหนดให้ a_2=12 b_2=18 m_1=m_2=4/5 ดังนั้น x_1^2=16 x_2^2 = 28 x_3^2 = 32 และ x_4^2 = 24 จะได้ 100=16+28+32+24 ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ =60%

ถ้า $a_2 > b_2$ จากสมการ (13) ทำให้ $x_4^2 > x_4^1$ จากสมการ (7) และ (8) เมื่อ $x_1^2 < x_1^1$ และ $x_4^2 > x_4^1$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ตัวอย่างเช่นจากสมการ (14) กำหนด a_2=15 b_2=10 m_1=3/4 ดังนั้น x_1^2 = 15 x_2^2 = 20 x_3^2 = 30 และ x_4^2= 35 จะได้ 100=15+20+30+35 มีค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$=70%

3.2 เมื่อ $x_2^2 < x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

เนื่องจาก $x_1^2 < x_1^1$ และ $x_3^2 < x_3^1$ ทำให้ $y_2 < y_1$ ดังนั้น $y_2 - y_1 < 0$ และเนื่องจาก $x_2^2 < x_2^1$ ดังนั้น $x_2^2 - x_2^1 < 0$

จากสมการ (10) ทำให้ $x_4^2 > x_4^1$

จากสมการ (7) และ (8) เมื่อ $x_1^2 < x_1^1$ และ $x_4^2 > x_4^1$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ดังนั้นเมื่อเปรียบเทียบค่าต่างๆ ของหลักเกณฑ์ $p_2 \rightarrow q_2$ กับหลักเกณฑ์ $p_1 \rightarrow q_1$ เมื่อรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1 ($x_2^2 < x_2^1$) พบว่า

- รายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนมากกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($x_4^2 > x_4^1$)
- ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

3.3 **เมื่อ** $x_2^2 = x_2^1$

ในกรณีนี้หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

เนื่องจาก $x_1^2 < x_1^1$ และ $x_3^2 < x_3^1$ ทำให้ $y_2 < y_1$ ดังนั้น $y_2 - y_1 < 0$ และเนื่องจาก $x_2^2 = x_2^1$ ดังนั้น $x_2^2 - x_2^1 = 0$

จากสมการ (10) ทำให้ $x_4^1 < x_4^2$

จากสมการ (7) และ (8) เมื่อ $x_1^2 < x_1^1$ และ $x_4^1 < x_4^2$ ทำให้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ดังนั้นเมื่อเปรียบเทียบค่าต่างๆ ของหลักเกณฑ์ $p_2 \rightarrow q_2$ กับหลักเกณฑ์ $p_1 \rightarrow q_1$ เมื่อรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1 ($x_2^2 = x_2^1$) พบว่า

- รายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนมากกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 ($x_4^2 > x_4^1$)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

- ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

ตารางที่ 5.1 :เปรียบเทียบค่าต่าง ๆ ระหว่างหลักเกณฑ์ $p_1 \rightarrow q_1$ และหลักเกณฑ์ $p_2 \rightarrow q_2$

กรณี	x_1^2 กับ x_1^1	x_2^2 กับ x_2^1	x_3^2 กับ x_3^1	x_4^2 กับ x_4^1	การเปรียบเทียบค่าความเชื่อมั่นประพจน์แย้งสลับที่	หมายเหตุ
1. $x_1^2 = x_1^1$ และ $x_3^2 = x_3^1$	=	>	=	<	<	
	=	<	=	>	>	
	=	=	=	=	=	
2. $x_1^2 > x_1^1$, $x_3^2 > x_3^1$ และ $\frac{x_1^1}{x_3^1} = \frac{x_1^2}{x_3^2}$	>	>	>	<	<	
	>	<	>	=	<	a=b
	>	<	>	>	<	a<b และ $m_1>m_2$
	>	<	>	>	>	a<b และ $m_2>m_1$
	>	<	>	>	=	a<b และ $m_1=m_2$
	>	<	>	<	<	a>b
	>	=	>	<	<	
3. $x_1^2 < x_1^1$, $x_3^2 < x_3^1$ และ $\frac{x_1^1}{x_3^1} = \frac{x_1^2}{x_3^2}$	<	>	<	=	>	a=b
	<	>	<	>	>	a>b
	<	>	<	<	=	a<b และ $m_1=m_2$
	<	>	<	<	>	a<b และ $m_2>m_1$
	<	>	<	<	<	a<b และ $m_1>m_2$
	<	<	<	>	>	
	<	=	<	>	>	

จากตารางที่ 5.1 แสดงการเปรียบเทียบจำนวนรายการต่างๆ และค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ กับหลักเกณฑ์ $p_1 \rightarrow q_1$ ในแต่ละกรณีที่ทำให้ค่าความเชื่อมั่นของแต่ละหลักเกณฑ์มีค่าเท่ากันซึ่งสัญลักษณ์ที่ปรากฏในตารางมีความหมายดังนี้

1. ในสดมภ์ x_1^2 กับ x_1^1

= หมายถึงรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1

> หมายถึงรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1

< หมายถึงรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1

2. ในสดมภ์ x^2_2 กับ x^1_2

= หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

> หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

< หมายถึงรายการที่ปรากฏสินค้า q_2 แต่ไม่ปรากฏสินค้า p_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า q_1 แต่ไม่ปรากฏสินค้า p_1

3. ในสดมภ์ x^2_3 กับ x^1_3

= หมายถึงรายการที่ปรากฏสินค้า p_2 และ q_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า p_1 และ q_1

> หมายถึงรายการที่ปรากฏสินค้า p_2 และ q_2 มีจำนวนมากกว่ารายการที่ปรากฏสินค้า p_1 และ q_1

< หมายถึงรายการที่ปรากฏสินค้า p_2 และ q_2 มีจำนวนน้อยกว่ารายการที่ปรากฏสินค้า p_1 และ q_1

4. ในสดมภ์ x^2_4 กับ x^1_4

= หมายถึงรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนเท่ากับรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1

> หมายถึงรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนมากกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1

< หมายถึงรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีจำนวนน้อยกว่ารายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5. ในสดมภ์การเปรียบเทียบค่าความเชื่อมั่นประพจน์แย้งสลับที่

= หมายถึงค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \longrightarrow q_2$ มีค่าเท่ากับหลักเกณฑ์ $p_1 \longrightarrow q_1$

> หมายถึงค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \longrightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \longrightarrow q_1$

< หมายถึงค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \longrightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \longrightarrow q_1$

ซึ่งสามารถสรุปแนวโน้มของค่าความเชื่อมั่นประพจน์แย้งสลับที่ได้ดังนี้

1. เมื่อรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีจำนวนเท่ากับรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 $(x_1^2 = x_1^1)$ ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \longrightarrow q_2$ จะมีแนวโน้มเหมือนกับจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 (x_4^2) คือถ้าจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่ามากกว่าถ้าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(x_4^2 > x_4^1)$ พบว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \longrightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \longrightarrow q_1$ แต่ถ้าจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่าน้อยกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(x_4^2 < x_4^1)$ ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \longrightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \longrightarrow q_1$ ด้วย
2. เมื่อจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีค่ามากกว่าจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 $(x_1^2 > x_1^1)$ และจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่าน้อยกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(x_4^2 < x_4^1)$ เช่นในกรณีที่ 2.1 และ 2.3 หรือเมื่อจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีค่าน้อยกว่าจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 $(x_1^2 < x_1^1)$ และจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่ามากกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(x_4^2 > x_4^1)$ เช่นในกรณีที่ 3.2 และ 3.3 พบว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \longrightarrow q_2$ จะมีแนวโน้มเหมือนกับจำนวนรายการที่ไม่ปรากฏ สินค้า p_2 หรือ q_2 (x_4^2) เช่นเมื่อจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีค่าน้อยกว่าจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 $(x_1^2 < x_1^1)$ และ

จำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่ามากกว่าถ้าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(x_4^2 > x_4^1)$ พบว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$

3. เมื่อจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีค่ามากกว่าจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 $(x_1^2 > x_1^1)$ และจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่ามากกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(x_4^2 > x_4^1)$ หรือเมื่อจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 มีค่าน้อยกว่าจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 $(x_1^2 < x_1^1)$ และจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่าน้อยกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(x_4^2 < x_4^1)$ พบว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ เมื่อเปรียบเทียบกับหลักเกณฑ์ $p_1 \rightarrow q_1$ จะมีแนวโน้มตรงข้ามกับค่า x_4^2 เมื่อเปรียบเทียบกับค่า x_4^1 คือเมื่อจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่ามากกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ $q_1 (x_4^2 > x_4^1)$ ความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่าน้อยกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$ แต่ถ้าจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 มีค่าน้อยกว่าจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(x_4^2 < x_4^1)$ ความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ $p_2 \rightarrow q_2$ มีค่ามากกว่าหลักเกณฑ์ $p_1 \rightarrow q_1$ ซึ่งมี 2 กรณีดังนี้

   3.1 กรณีอัตราส่วนระหว่างจำนวนรายการที่ปรากฏ สินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 มีค่ามากกว่าอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 กับจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 $(\frac{x_1^2}{x_1^1} > \frac{x_4^2}{x_4^1})$ เมื่อ $x_1^2 > x_1^1$ และ $x_4^2 > x_4^1$

   3.2 กรณีอัตราส่วนระหว่างจำนวนรายการที่ไม่ปรากฏสินค้า p_2 หรือ q_2 กับจำนวนรายการที่ไม่ปรากฏสินค้า p_1 หรือ q_1 มีค่ามากกว่าอัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_2 แต่ไม่ปรากฏสินค้า q_2 กับจำนวนรายการที่ปรากฏสินค้า p_1 แต่ไม่ปรากฏสินค้า q_1 $(\frac{x_1^2}{x_1^1} < \frac{x_4^2}{x_4^1})$ เมื่อ $x_1^2 < x_1^1$ และ $x_4^2 < x_4^1$

บทที่ 6

สรุปผล

เนื่องจากเป้าหมายของงานวิจัยนี้คือทำการคัดเลือกหลักเกณฑ์แอสโซซิเอชันจากหลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากัน ดังนั้นงานวิจัยนี้จึงทำการเสนอค่าใหม่ขึ้นมาเพื่อใช้ในการคัดเลือกหลักเกณฑ์ ซึ่งค่านี้เรียกว่าค่าความเชื่อมั่นประพจน์แย้งสลับที่ และทำการคัดเลือกหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุดจากหลักเกณฑ์ที่มีค่าความเชื่อมั่นของหลักเกณฑ์เท่ากันจำนวน n หลักเกณฑ์ซึ่งแต่ละหลักเกณฑ์มีรูปแบบดังนี้ $p_1 \rightarrow q_1$, $p_2 \rightarrow q_2$,..., $p_n \rightarrow q_n$ จากการทดลองในบทที่ 4 พบว่าใน n หลักเกณฑ์ ถ้าหลักเกณฑ์ใดมีค่าสนับสนุนประพจน์แย้งสลับที่สูงสุดหลักเกณฑ์นั้นจะมีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุด แต่ถ้ามีหลักเกณฑ์ที่มีค่าสนับสนุนประพจน์แย้งสลับที่ของหลักเกณฑ์เท่ากัน หลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่สูงสุด คือหลักเกณฑ์ $p_k \rightarrow q_k$ ที่มีจำนวนรายการที่ปรากฏสินค้า p_k แต่ไม่ปรากฏสินค้า q_k ต่ำสุดเมื่อเทียบกับหลักเกณฑ์ทั้งหมดที่มีค่าความเชื่อมั่นเท่ากัน

ดังนั้นในบทที่ 5 จึงทำการวิเคราะห์แนวโน้มของค่าความเชื่อมั่นประพจน์แย้งสลับที่ในกรณีต่างๆ ที่ทำให้หลักเกณฑ์มีค่าความเชื่อมั่นเท่ากัน ผลการวิเคราะห์พบว่าเมื่อหลักเกณฑ์มีค่าความเชื่อมั่นเท่ากัน ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์มีค่าไม่เท่ากัน จึงทำให้สามารถคัดเลือกหลักเกณฑ์โดยใช้ค่าความเชื่อมั่นประพจน์แย้งสลับที่ได้ ซึ่งลักษณะของข้อมูลที่สามารถนำค่าความเชื่อมั่นประพจน์แย้งสลับที่มาใช้ในการคัดเลือกหลักเกณฑ์ มีดังนี้

1. กรณีที่จำนวนรายการที่ปรากฏสินค้า p_k แต่ไม่ปรากฏสินค้า q_k มีค่าเท่ากันทุกหลักเกณฑ์ $(x_1^1 = x_1^2 = ... = x_1^n)$ และจำนวนรายการที่ปรากฏสินค้า p_k และ q_k มีค่าเท่ากันทุกหลักเกณฑ์ $(x_3^1 = x_3^2 = ... = x_3^n)$
2. กรณีที่อัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_k แต่ไม่ปรากฏสินค้า q_k กับจำนวนรายการที่ปรากฏสินค้า p_k และ q_k มีค่าเท่ากันในแต่ละหลักเกณฑ์ $(\frac{x_1^1}{x_3^1} = \frac{x_1^2}{x_3^2} ... = \frac{x_1^n}{x_3^n})$

เมื่อพิจารณาหลักเกณฑ์แอสโซซิเอชันทั้งหมดที่ได้จากกระบวนการสร้างหลักเกณฑ์แอสโซซิเอชันของ Agrawal, Imielinski และ Swami [4] Agrawal และ Srikant [13] Liu, Hsu และ Ma [14] Savasere, Omiecinski และ Navathe [15] Yen และ Chen [16] Zaki, Parthasarathy, Ogihara และ Li [17] เชาวนี ศรีวิศาล และวิเชียร เปรมชัยสวัสดิ์ [18] พบว่าหลักเกณฑ์ที่ได้มีจำนวนมาก แต่เมื่อนำหลักการคัดเลือกหลักเกณฑ์แอสโซซิเอชันที่มีค่าความเชื่อมั่นเท่ากันด้วยค่าความเชื่อมั่นประพจน์แย้งสลับที่ พบว่าช่วยลดจำนวนหลักเกณฑ์ที่มีค่าความเชื่อมั่นเท่ากันเพื่อนำหลักเกณฑ์มาวิเคราะห์วางแผนทางการตลาด

ข้อจำกัดของงานวิจัยนี้คือ จากหลักเกณฑ์ทั้งหมดที่มีค่าความเชื่อมั่นเท่ากันอาจมีหลักเกณฑ์ที่มีค่าความเชื่อมั่นประพจน์แย้งสลับที่เท่ากันได้ ซึ่งหลักเกณฑ์เหล่านั้นมีลักษณะดังนี้

1. จำนวนรายการที่ปรากฏสินค้า p_k แต่ไม่ปรากฏสินค้า q_k ของแต่ละหลักเกณฑ์มีค่าเท่ากัน $(x_1^1 = x_1^2 = ... = x_1^n)$ จำนวนรายการที่ปรากฏสินค้า q_k แต่ไม่ปรากฏสินค้า p_k ของแต่ละหลักเกณฑ์มีค่าเท่ากัน $(x_2^1 = x_2^2 = ... = x_2^n)$ จำนวนรายการที่ปรากฏสินค้า p_k และ q_k มีค่าเท่ากันในแต่ละหลักเกณฑ์ $(x_3^1 = x_3^2 = ... = x_3^n)$ และจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q ของแต่ละหลักเกณฑ์มีค่าเท่ากัน $(x_4^1 = x_4^2 = ... = x_4^n)$

2. อัตราส่วนระหว่างจำนวนรายการที่ปรากฏสินค้า p_k แต่ไม่ปรากฏสินค้า q_k กับจำนวนรายการที่ไม่ปรากฏสินค้า p_k หรือ q_k มีค่าเท่ากันทุกหลักเกณฑ์ $(\frac{x_1^1}{x_4^1} = \frac{x_1^2}{x_4^2} = ... = \frac{x_1^n}{x_4^n})$

ดังนั้นเมื่อข้อมูลมีลักษณะดังนี้ จะทำให้ไม่สามารถใช้ค่าความเชื่อมั่นประพจน์แย้งสลับที่มาทำการคัดเลือกหลักเกณฑ์ได้ และในงานวิจัยนี้ได้การคัดเลือกหลักเกณฑ์ $P \rightarrow Q$ โดยพิจารณาเฉพาะ เซต P และ Q ที่มีสมาชิกเพียงตัวเดียว ซึ่งในบางกรณีอาจมีหลักเกณฑ์ที่มีสมาชิกของเซต P หรือ Q มากกว่า 1 ตัวที่มีค่าความเชื่อมั่นเท่ากัน ดังนั้นงานวิจัยที่น่าสนใจคือใช้ทฤษฎีทางด้านตรรกศาสตร์วิภัชนัย (fuzzy logic) มาประยุกต์ในการสร้างค่าเพื่อทำการคัดเลือกหลักเกณฑ์ที่มีจำนวนสมาชิกในเซต P หรือ Q มากกว่าหนึ่ง

รายการอ้างอิง

1. Ramakrishan, R. and Gehrke, J. **Database Management Systems**. Singapore: McGraw-Hill, 2000.
2. Megaputer. **Data Mining**[online].(n.d.). Available from: http://www.megaputer.com/dm/dm101.php3 [2002, December 16]
3. Agrawal, R., Imielinski, T. and Swami, A. Mining Association Rules Between Sets of Items in Large Databases. **Proceedings of the 1993 ACM SIGMOD Conference** (May 1993): 207-216.
4. Agrawal, R. and Srikant, R. Fast Algorithms for Mining Association Rules. **Proceedings of the VLDB Conference** 20 (September 1994): 487-499.
5. Liu, B., Hsu, W. and Ma, Y. Mining Association Rules with Multiple Minimum Supports. **Proceedings of the International Conference on Knowledge Discovery and Data Mining** 5 (August 1999): 337-341.
6. Savasere, A. Omiecinski, E. and Navathe, S. An Efficient Algorithms for Mining Association Rules in Large Database. **Proceedings of the VLDB Conference** 21 (September 1995): 432-444.
7. Yen, S. J. and Chen, A. L. P. An Efficient Data Mining Technique for Discovery Interesting Association Rules. **Proceedings of Database and Expert Systems Application** 8 (September 1997): 664-669.
8. Zaki, M. J., Parthasarathy, S., Ogihara, M. and Li, W. New Algorithms for Fast Discovery of Association Rules. **Proceedings of the International Conference on Knowledge Discovery and Data Mining** 3 (August 1997): 283-286.
9. เชาวนี ศรีวิศาล และวิเชียร เปรมชัยสวัสดิ์. วิธีการแอสโซซิเอชันรูลย์สำหรับงานทางการตลาด. **Proceedings of the National Computer Science and Engineering** 5 (November 2001): 223-231.
10. อนันตพร ศรีสวัสดิ์ และกรุง สินอภิรมย์สราญ. ผลกระทบของค่าความเชื่อมั่นประพจน์แย้งสลับที่ที่มีต่อการคัดเลือกหลักเกณฑ์แอสโซซิเอชัน. **Proceedings of the National Computer Science and Engineering** 6 (October 2002): 164-173.

11. Bigus, J. P. **Data Mining with Neural Networks, Solving Business Problems from Application Development to Decision Support**. New York: McGraw-Hill, 1996.

12. Grudnitski, G. and Osburn, L. Forecasting S&P and Gold Futures Prices: An Application of Neural Networks. **Journal of Futures Markets** 13 (1993): 631-643.

13. Komo, D. Chang, C.I. and Ko, H. Stock Market Index Prediction Using Neural Networks. **Applications of Artificial Neural Networks V, SPIE** 2243 (April 1994): 516-526.

14. Margarita, S. and Beltratti, A. Credit Risk and Lending in an Artificial Adaptive Banking System, Adaptive Intelligent Systems. **Proceeding of the BANKAI Workshop** (1992): 161-176.

15. Udo, G. Neural Network Performance on the Bankruptcy Classification Problem. **Computers and Industrial Engineering** 25 (1993): 377-380.

16. Wong, B.K., Bodnovich, T.A. and Selvi, Y. A Bibliography of Neural Network Business Applications Research. **Expert Systems: The International Journal of Knowledge Engineering and Neural Networks** (August 1995): 253-262.

17. Kleissner, C. Data Mining for the Enterprise. **Proceedings of the Annual International Conference on System Sciences-Volume 7** 31 (January 1998): 295-304.

18. Berry, M.J.A. and Linoff, G. **Data Mining Techniques for Marketing, Sales and Customer Support**. United States of America: Wiley Computer Publishing, 1997.

19. ภัทรา เตชาภิวาทย์. **คณิตตรรกศาสตร์**. พิมพ์ครั้งที่ 3.กรุงเทพฯ: มหาวิทยาลัยเกษตรศาสตร์, 2540.

20. Klir, G.J., Clair, U.St. and Yuan, B. **Fuzzy Set Theory Foundations and Applications**. America: Prentice-Hall, 1997.

21. Vince, A.J. and Morris, C.A.N. **Discrete Mathematics for Computing**. England: Ellis Horwood, 1990.

22. Horstmann, C.S. and Cornell, G. **Core JAVA 1.1 Volume II-Advanced Features**. United States of America: Sun Microsystems, 1998.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

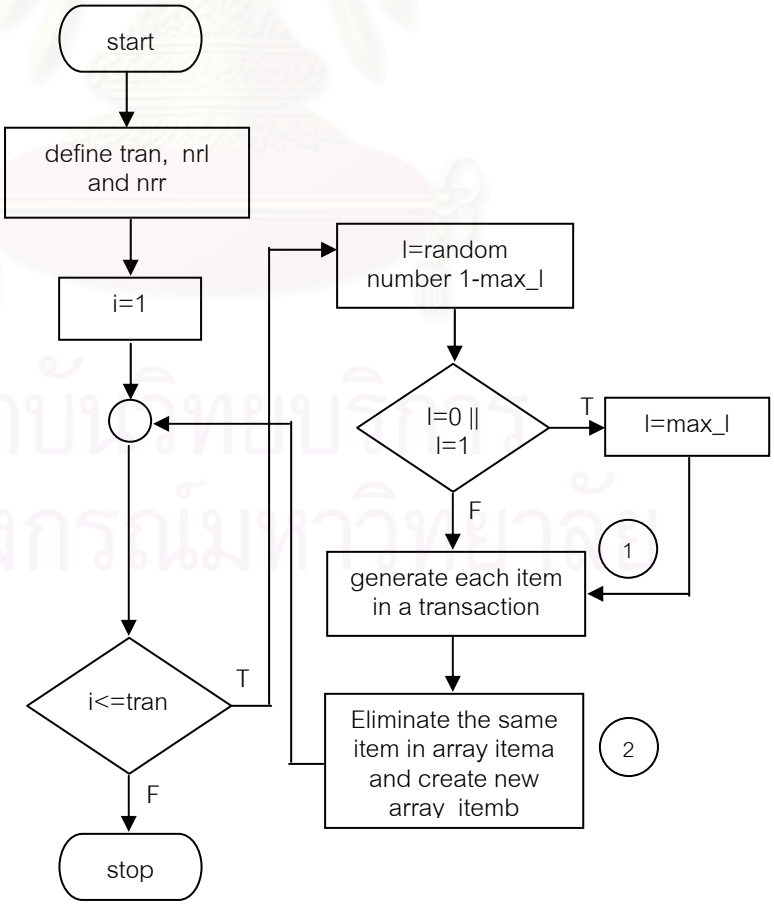
สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

โปรแกรมจำลองข้อมูล

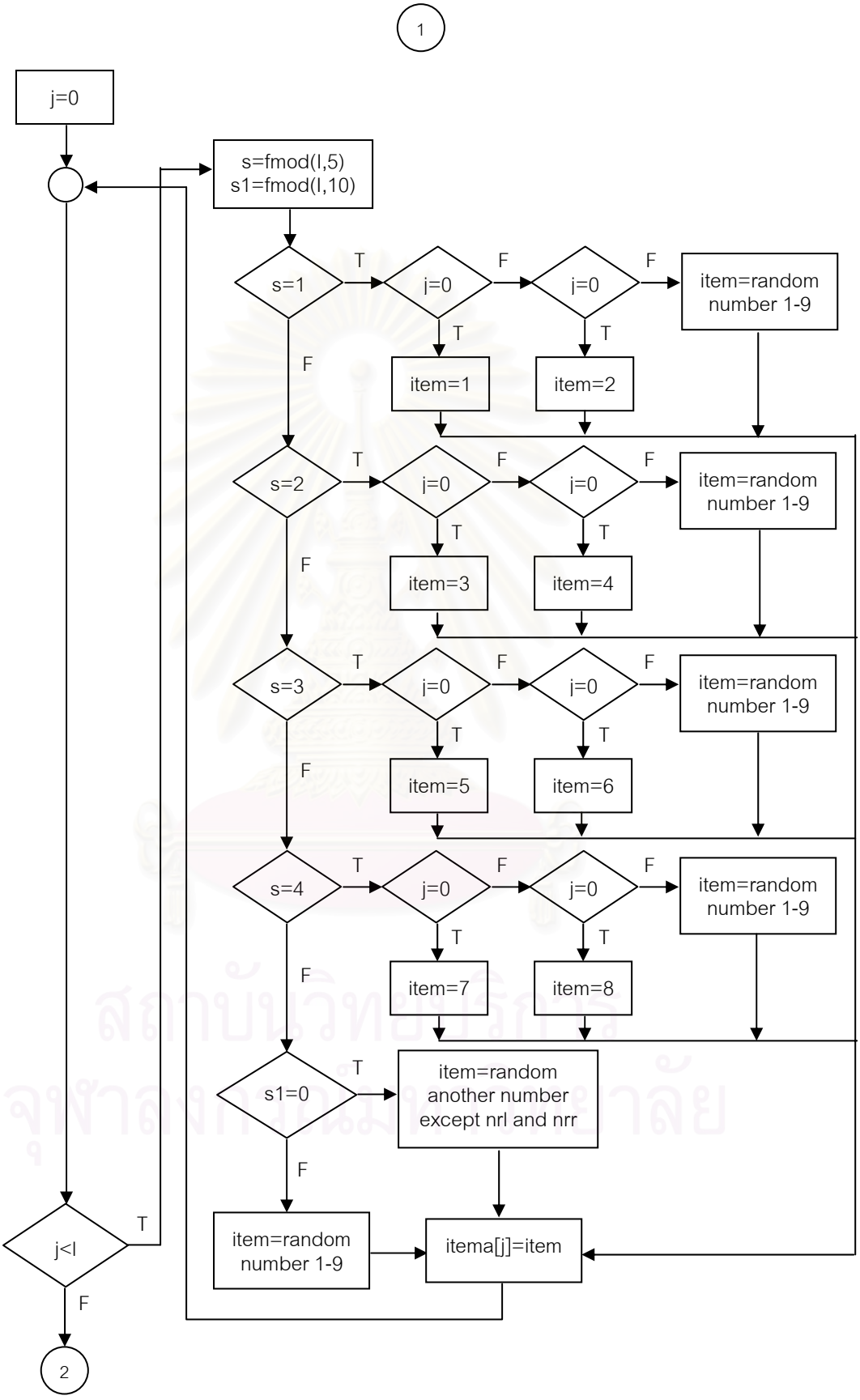
ผลลัพธ์ที่ได้จากโปรแกรมนี้คือข้อมูลรายการสินค้าโดยแทนชนิดสินค้าด้วยตัวเลข 1-9 โดยในโปรแกรมมีการกำหนดให้ชนิดสินค้าทั้งหมด (ti) มี 9 ชนิดและจำนวนชนิดสินค้าที่มากที่สุดในแต่ละรายการ (max_l) เท่ากับ 3 ซึ่งโปรแกรมนี้แบ่งการทำงานได้ 4 ส่วนดังนี้

1. ผู้ใช้กำหนดจำนวนรายการทั้งหมด (tran) และกำหนดเซตสินค้าที่ต้องการให้มีค่าสนับสนุนประพจน์แย้งสลับที่ของเซตสินค้านั้นมากที่สุด (nrl และ nrr)
2. สร้างรายการแต่ละรายการ (transaction) กำหนดให้จำนวนชนิดสินค้าในแต่ละรายการด้วยการสุ่มตัวเลข (l)
3. สร้างสินค้าแต่ละชนิดในแต่ละรายการจากเงื่อนไขที่กำหนดและทำการสุ่มตัวเลขเพื่อแทนสินค้าแต่ละชนิด
4. ในแต่ละรายการ กำจัดตัวเลขที่มีค่าซ้ำกัน ดังนั้นผลที่ได้จึงไม่มีสินค้าชนิดเดียวกันปรากฏอยู่ในรายการเดียวกัน

รูปที่ ก.1: แผนภาพแสดงขั้นตอนการทำงานของโปรแกรมจำลองข้อมูล



รูปที่ ก.2: แผนภาพแสดงขั้นตอนการทำงานของโปรแกรมจำลองข้อมูล(ต่อ)



รูปที่ ก.1 และรูปที่ ก.2 แสดงการทำงานของโปรแกรมจำลองข้อมูลอย่างละเอียด ซึ่งชื่อ ชนิด และความหมายของตัวแปรต่างๆ ที่แสดงไว้ในรูปที่ ก.1 และ รูปที่ ก.2 อธิบายได้ดังตารางที่ ก.1

ตารางที่ ก.1: ตารางแสดงชื่อ ชนิด และความหมายของตัวแปรที่สำคัญในโปรแกรมจำลองข้อมูล

ชื่อ	ชนิด	ความหมาย
tran	integer	จำนวนรายการทั้งหมด
max_l	integer	จำนวนชนิดสินค้าที่มากที่สุดในแต่ละรายการ
ti	integer	จำนวนชนิดสินค้าทั้งหมด
l	integer	จำนวนชนิดสินค้าในแต่ละรายการ
nrl	integer	สินค้าตัวแรกในเซตสินค้าที่ต้องการให้มีค่าสนับสนุนประพจน์แย้งสลับที่มากที่สุด
nrr	integer	สินค้าตัวที่สองในเซตสินค้าที่ต้องการให้มีค่าสนับสนุนประพจน์แย้งสลับที่มากที่สุด
item	integer	ชนิดสินค้าที่ได้จากการกำหนดหรือการสุ่มตัวเลข
itema	array of integer	เก็บสินค้าชนิดต่างๆในแต่ละรายการ
itemb	array of integer	เก็บสินค้าชนิดต่างๆในแต่ละรายการ หลังจากกำจัดชนิดสินค้าที่ซ้ำออกไป

เนื่องจากโปรแกรมสร้างหลักเกณฑ์แอสโซซิเอชันของงานวิจัยนี้สร้างจากภาษาจาวา โดยดึงข้อมูลจากฐานข้อมูลในโปรแกรม Microsoft Access ด้วยคำสั่ง sql แต่ข้อมูลที่ได้จากโปรแกรมจำลองข้อมูลนี้อยู่ในรูปแฟ้มเก็บข้อความ (text file) ดังนั้นผู้วิจัยต้องทำการแปลงข้อมูลนี้ให้อยู่ในรูปฐานข้อมูลในโปรแกรม Microsoft Access โดยอาศัยโปรแกรม Makedata.java [22] ซึ่งการทำงานของโปรแกรมนี้จะทำการนำข้อมูลที่อยู่ในรูปแฟ้มเก็บข้อความไปสร้างเป็นฐานข้อมูล

กำหนดให้ตารางฐานข้อมูลที่เก็บข้อมูลที่ได้จากโปรแกรมจำลองข้อมูลชื่อว่า data โครงสร้างของฐานข้อมูลในตาราง data แสดงได้ดังตารางที่ ก.2

ตารางที่ ก.2: โครงสร้างข้อมูลของตาราง data

ชื่อฟิลด์	ชนิด	ขนาด	ความหมาย
transid	text	4	รหัสรายการ
item	text	1	ชนิดสินค้า

โปรแกรมสร้างหลักเกณฑ์แอสโซซิเอชัน

โปรแกรมสร้างหลักเกณฑ์แอสโซซิเอชันที่ใช้ในงานวิจัยนี้สร้างจากภาษาจาวา มีชื่อว่า Create_rule.java โดยทำการดึงข้อมูลจากฐานข้อมูลตาราง data ที่ได้จากการจำลองข้อมูล และทำการสร้างหลักเกณฑ์แอสโซซิเอชันที่มีค่าสนับสนุนของหลักเกณฑ์มากกว่าหรือเท่ากับ 20%

ตารางที่ ก.3: ตารางแสดงชื่อ ชนิด และความหมายของตัวแปรที่สำคัญในโปรแกรมสร้างหลักเกณฑ์แอสโซซิเอชัน

ชื่อ	ชนิด	ความหมาย
si	Array of integer	เก็บจำนวนรายการของสินค้าแต่ละชนิด โดยแทนชนิดสินค้าด้วย index ของ array มีขนาด 0-n เมื่อ n คือจำนวนชนิดสินค้าทั้งหมด
item	Array of string	เก็บชนิดสินค้าที่เป็น frequent itemset ที่มีสมาชิก 1 ตัว
item21	Array of string	เก็บชนิดสินค้าตัวแรกใน frequent itemset ที่มีสมาชิก 2 ตัว
item22	Array of string	เก็บชนิดสินค้าตัวที่สองใน frequent itemset ที่มีสมาชิก 2 ตัว
si2	Array of string	เก็บจำนวนรายการของแต่ละเซตสินค้า
RuleLeft	Array of string	เก็บชนิดสินค้าที่อยู่ทาง LHS ของหลักเกณฑ์ ซึ่ง index ของ array คือลำดับของหลักเกณฑ์
RuleRight	Array of string	เก็บชนิดสินค้าที่อยู่ทาง RHS ของหลักเกณฑ์ ซึ่ง index ของ array คือลำดับของหลักเกณฑ์
conf	Array of string	เก็บค่าความเชื่อมั่นของแต่ละหลักเกณฑ์
supRule	Array of string	เก็บจำนวนรายการของเซตสินค้าในหลักเกณฑ์
Pcon	Array of double	เก็บค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์
order1	Array of string	เก็บลำดับของหลักเกณฑ์ โดยเรียงลำดับจากหลักเกณฑ์ที่มีค่าความเชื่อมั่นน้อยไปหาหลักเกณฑ์ที่มีค่าความเชื่อมั่นมาก

ขั้นตอนการทำงานของโปรแกรมสร้างหลักเกณฑ์แอสโซซิเอชัน

1. **สร้าง frequent itemset ที่มีสมาชิก 1 ตัว**

นับจำนวนรายการที่มีสินค้าแต่ละชนิดปรากฏอยู่ นำค่าที่ได้เก็บใน array si โดยที่ index ของ array คือชนิดของสินค้า และถ้าสินค้าชนิดใดมีจำนวนรายการที่มีสินค้านั้นอยู่มากกว่าหรือเท่ากับค่า support ที่กำหนดไว้ จะทำการเก็บสินค้าชนิดนั้นไว้ใน array item

2. **สร้าง frequent itemset ที่มีสมาชิก 2 ตัว**

นำสมาชิกใน array item มาจับคู่กันทั้งหมด ซึ่งมีทั้งหมด $\frac{n(n-1)}{2}$ คู่ และนำแต่ละคู่มาหาจำนวนรายการที่มีสินค้าคู่นั้นปรากฏอยู่ ถ้าสินค้าคู่ใดมีจำนวนรายการมากกว่าหรือเท่ากับค่า support ที่กำหนดจึงทำการเก็บสินค้าตัวแรกไว้ใน array item21 และสินค้าตัวที่สองไว้ใน array item22 โดยใช้ index ตัวเดียวกันในการเก็บค่า และเก็บจำนวนรายการที่ได้ไว้ใน array si2 โดยใช้ index ตัวเดียวกัน

3. **สร้างหลักเกณฑ์แอสโซซิเอชัน**

นำสมาชิกใน array item21 และ item22 มาสร้างหลักเกณฑ์ โดยเก็บสินค้าที่เป็น LHS ของหลักเกณฑ์ไว้ที่ array RuleLeft และเก็บสินค้าที่เป็น RHS ของหลักเกณฑ์ไว้ที่ array RuleRight ซึ่งตำแหน่ง index เดียวกันแสดงถึงหลักเกณฑ์เดียวกัน และคำนวณหาค่าความเชื่อมั่นของหลักเกณฑ์ ค่าความเชื่อมั่นประพจน์แย้งสลับที่ของหลักเกณฑ์ ค่า x_1, x_2, x_3, และ x_4 ของแต่ละหลักเกณฑ์

ตัวอย่างการทำงานของโปรแกรมสร้างหลักเกณฑ์แอสโซซิเอชัน

ตารางที่ ก.4 แสดงฐานข้อมูลตัวอย่าง โดยกำหนดให้ตัวเลขในฟิลด์ item แทนชนิดสินค้าดังนี้ 1 แทนสินค้าปากกา 2 แทนสินค้าหมึก 3 แทนสินค้านม 4 แทนสินค้าน้ำผลไม้

ตารางที่ ก.4: ตารางฐานข้อมูลตัวอย่าง

transid	item
111	1
111	2
111	3
111	4
112	3
112	4
113	1
113	3
114	1
114	2

เมื่อนำข้อมูลจากฐานข้อมูลในตารางที่ ก.4 มาใช้ในการสร้างหลักเกณฑ์แอสโซซิเอชันตามขั้นตอนการทำงานของโปรแกรม Create_rule.java แสดงได้ดังนี้

ขั้นตอนที่ 1: สร้าง frequent itemset ที่มีสมาชิก 1 ตัว

สินค้า	0	1	2	3	4
si	0	3	2	3	2

item	1	2	3	4

การทำงานของโปรแกรมในขั้นตอนนี้จะทำการ query ฐานข้อมูลในตารางที่ 4 เพื่อหาจำนวนรายการของสินค้าแต่ละชนิด จากตัวอย่างจะเห็นว่าใน array si เก็บจำนวนรายการของสินค้าแต่ละชนิด เช่น สินค้า 1 มีทั้งหมด 3 รายการ สินค้า 2 มีทั้งหมด 2 รายการ เป็นต้น และเมื่อทำการ query ครบทุกสินค้าแล้วทำการพิจารณาว่าสินค้าชนิดใดมีจำนวนรายการน้อยกว่า 20% ของจำนวนรายการทั้งหมด ซึ่งคือจำนวน 2 รายการ พบว่าสินค้าทั้ง 4 ชนิดมีจำนวนรายการมากกว่าหรือเท่ากับ 2 ดังนั้นจึงเก็บสินค้าทั้ง 4 ชนิดไว้ใน array item

ขั้นตอนที่ 2: สร้าง frequent itemset ที่มีสมาชิก 2 ตัว

index	0	1	2
item21	1	1	3

index	0	1	2
item22	2	3	4

index	0	1	2
si2	2	2	2

การทำงานของโปรแกรมในขั้นตอนนี้นำสมาชิกใน array item มาจับคู่กันเป็นเซตสินค้าที่มีสมาชิก 2 ตัว และทำการ query หาจำนวนรายการของเซตสินค้าแต่ละเซต ดังนั้นจากขั้นตอนนี้ เซตสินค้าทั้งหมดที่ได้คือ {1,2}, {1,3} และ {3,4} ซึ่งสมาชิกตัวแรกของเซตจะเก็บไว้ใน array item21 และสมาชิกตัวหลังเก็บไว้ใน array item22 และทำการเก็บค่าที่ได้จากการ query ไว้ใน array si2 โดยมี index เป็นตัวกำหนดลำดับเซตสินค้า เช่น เซตสินค้าที่ index = 0 หมายถึงเซตสินค้าลำดับ 0 คือ {1,2} ซึ่งได้มาจากค่าใน array item21[0] และ item22[0] และจำนวนรายการที่มีเซตสินค้านี้เท่ากับ 2 ซึ่งได้มาจาก array si2[0]

ขั้นตอนที่ 3: สร้างหลักเกณฑ์แอสโซซิเอชัน

หลักเกณฑ์	0	1	2	3	4	5
RuleLeft	1	2	1	3	3	4

หลักเกณฑ์	0	1	2	3	4	5
RuleRight	2	1	3	1	4	3

หลักเกณฑ์	0	1	2	3	4	5
conf	66.66	100.0	66.66	66.66	66.66	100.0

หลักเกณฑ์	0	1	2	3	4	5
supRule	2	2	2	2	2	2

หลักเกณฑ์	0	1	2	3	4	5
Pcon	50.0	100.0	0.0	0.0	50.0	100.0

index	0	1	2	3	4	5
order1	0	2	3	4	1	5

การทำงานของโปรแกรมในขั้นตอนนี้นำเซตสินค้าที่ได้จากขั้นตอนที่ 2 มาสร้างเป็นหลักเกณฑ์แอสโซซิเอชัน โดยเก็บหลักเกณฑ์ที่สร้างได้ไว้ใน array RuleLeft และ array RuleRight ซึ่งเก็บสินค้าที่เป็น LHS และ RHS ตามลำดับ และมี index แสดงลำดับของหลักเกณฑ์ จากนั้นเก็บจำนวนรายการที่มีเซตสินค้าในหลักเกณฑ์ไว้ใน array supRule ดังนั้นหลักเกณฑ์ทั้งหมดที่ได้มีดังนี้

ลำดับที่ 0 หลักเกณฑ์ 1 $\rightarrow$ 2

ลำดับที่ 1 หลักเกณฑ์ 2 $\rightarrow$ 1

ลำดับที่ 2 หลักเกณฑ์ 1 $\rightarrow$ 3

ลำดับที่ 3 หลักเกณฑ์ 3 $\rightarrow$ 1

ลำดับที่ 4 หลักเกณฑ์ 3 $\rightarrow$ 4

ลำดับที่ 5 หลักเกณฑ์ 4 $\rightarrow$ 3

เมื่อได้หลักเกณฑ์ทั้งหมดแล้วทำการหาค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่ของแต่ละหลักเกณฑ์และเก็บค่านี้ไว้ใน array conf และ array Pcon ตามลำดับ ซึ่งมี index เป็นค่าบอกลำดับของหลักเกณฑ์ เช่นลำดับที่ 0 หมายถึงหลักเกณฑ์ 1 $\rightarrow$ 2 ซึ่งค่าความเชื่อมั่นและค่าความเชื่อมั่นประพจน์แย้งสลับที่นั้นคำนวณได้ดังสมการที่ (1) และ (2) ดังนี้

$$\text{ค่าความเชื่อมั่น} = \frac{\text{supRule[i]}}{\text{si[RuleLeft[i]]}} * 100 \qquad (1)$$

$$\text{ค่าความเชื่อมั่นประพจน์แย้งสลับที่} = \frac{\text{n-(si[RuleLeft[i]]-supRule[i]+si[RuleRight[i]])}}{\text{n-si[RuleRight[i]]}} * 100 \qquad (2)$$

จากสมการ (1) และ (2) ค่า i คือลำดับของหลักเกณฑ์ ในตัวอย่างนี้มีค่าจาก 0 ถึง 5 และ n คือจำนวนรายการทั้งหมด ซึ่งในตัวอย่างนี้มีค่าเท่ากับ 4 จากนั้นทำการเรียงลำดับหลักเกณฑ์ทั้งหมดที่ได้ โดยเรียงลำดับจากหลักเกณฑ์ที่มีค่าความเชื่อมั่นน้อยไปหาหลักเกณฑ์ที่มีค่าความเชื่อมั่นมาก ดังนั้นลำดับของหลักเกณฑ์ที่มีค่าความเชื่อมั่นที่เรียงจากน้อยไปมากจึงเก็บไว้ใน array order1 สุดท้ายทำการคำนวณหาจำนวนรายการที่ปรากฏสินค้า p แต่ไม่ปรากฏสินค้า q (x_1) จำนวนรายการที่ปรากฏสินค้า q แต่ไม่ปรากฏสินค้า p (x_2) จำนวนรายการที่ปรากฏสินค้า p และ q (x_3) และจำนวนรายการที่ไม่ปรากฏสินค้า p หรือ q (x_4) ของแต่ละหลักเกณฑ์ดังนี้

x_1 = si[RuleLeft[order1[i]]]-supRule[order1[i]] (3)

x_2 = si[RuleRight[order1[i]]]-supRule[order1[i]] (4)

x_3 = supRule[order1[i]] (5)

x_4 = n-(si[RuleLeft[order1[i]]]+si[RuleRight[order1[i]]]-supRule[order1[i]]) (6)

จากสมการที่ (3)-(6) ค่า i คือ index ของ array order1 ซึ่งผลลัพธ์ทั้งหมดที่ได้จากโปรแกรม Create_rule.java แสดงได้ดังรูปที่ ก.3

รูปที่ ก.3: ผลลัพธ์ที่ได้จากโปรแกรม Create_rule.java

rule	conf	contra_conf	x1	x2	x3	x4
1-->2	66.66	50.0	1	0	2	1
1-->3	66.66	0.0	1	1	2	0
3-->1	66.66	0.0	1	1	2	0
3-->4	66.66	50.0	1	0	2	1
2-->1	100.0	100.0	0	1	2	1
4-->3	100.0	100.0	0	1	2	1

total rule=6

source code โปรแกรม Create_rule.java

```
import java.net.*;
import java.sql.*;
import java.io.*;
import java.util.*;
import java.math.*;
class Create_rule{
public static void main (String args[]) {
Vector vitem=new Vector();
Vector CPcon=new Vector();
Vector vorder=new Vector();
Vector vitem21=new Vector();
Vector vitem22=new Vector();
Vector vsi2=new Vector();
Vector vRuleRight=new Vector();
Vector vRuleLeft=new Vector();
Vector vcon=new Vector();
Vector vsupRule=new Vector();
Vector vConValue=new Vector();
int support=2;
int si [];
```

```
long n=4;
int item1=4;
int i,sum,j,k,rhs,lhs,supr,rep,l,r;
long x1,x2,x4,s,sl,sr;
int num_si2,numi2,num_item;
double conf1;
try
{
        si=new int[item1+1];
        for(i=0;i<item1+1;i++)
                si[i]=0;
        Class.forName("sun.jdbc.odbc.JdbcOdbcDriver");
        // force loading of driver
        Connection con = DriverManager.getConnection("jdbc:odbc:corejava2","","");
        Statement stmt = con.createStatement();
        // ----------step 1 query the number of transaction for each item-------//
        for(i=1;i<=item1;i++)
        {
                String query = "select transid from tran where item="+"'"+i+"' group by
                transid";
                ResultSet rs = stmt.executeQuery(query);
                sum=0;
                while(rs.next())
                        sum++;
                si[i]=sum;  // store the number of transaction for each item i in array si
                rs.close();
                if(sum>=support)
                        vitem.addElement(Integer.toString(i)); // store frequent item in
                        vector vitem
        }
        num_item=vitem.size();
        // transfer members in vector vitem to array item
        String item[];
        item=new String[num_item];
        vitem.copyInto(item);
        //------- end step1 ------------------
        //-------  step 2 generate 2-frequent itemset ----------//
        k=0;
        for(i=0;i<num_item;i++)
                for(j=i+1;j<num_item;j++)
                {   // query number of transaction for each pair of items
                        String query1 = "select count(*) from tran where item
                        in("+"'"+item[i]+"','"+item[j]+"') group by transid having
                        count(*)=2";
                        ResultSet rs1 = stmt.executeQuery(query1);
                        sum=0;
                        while(rs1.next())
                                sum++;
                        rs1.close();
```

```
                    if(sum>=support)
                    {
                              vsi2.addElement(Integer.toString(sum));  // store the
                              number of transaction for each itemset in vector vsi2
                            vitem21.addElement(item[i]);  // store the first item of
                            itemset in vector vitem21
                            vitem22.addElement(item[j]);  // store the last item of
                            itemset in vector vitem22
                    }
          }
 numi2=vitem21.size();
 // transfer members in vector vitem21 to array item21
 String item21[];
 item21=new String[numi2];
vitem21.copyInto(item21);
 // transfer members in vector vitem22 to array item22
 String item22[];
 item22=new String[numi2];
 vitem22.copyInto(item22);
 // transfer members in vector vsi2 to array si2
 String si2[];
 si2=new String[numi2];
 vsi2.copyInto(si2);
 vitem.removeAllElements();
 vitem21.removeAllElements();
 vitem22.removeAllElements();
 vsi2.removeAllElements();
 //--------- end step2 ------------
 //--------- step 3 generate association rule----------//
 // generate all rules in vector vRuleLeft and vRuleRight
 for(i=0;i<numi2;i++)
 {
          vRuleLeft.addElement(item21[i]); // store lhs item in vector vRuleLeft
          vRuleRight.addElement(item22[i]);// store rhs item in vector vRuleRight
          vsupRule.addElement(si2[i]);     // store the number of transaction
          of lhs and rhs
          vRuleLeft.addElement(item22[i]);
          vRuleRight.addElement(item21[i]);
          vsupRule.addElement(si2[i]);
 }
 int total_rule=vsupRule.size();
 // transfer members in vector vRuleLeft to array RuleLeft
 String RuleLeft[];
 RuleLeft=new String[total_rule];
 vRuleLeft.copyInto(RuleLeft);
 // transfer members in vector vRuleRight to array RuleRight
 String RuleRight[];
 RuleRight=new String[total_rule];
 vRuleRight.copyInto(RuleRight);
```

```
// transfer members in vector vsupRule to array supRule
String supRule[];
supRule=new String[total_rule];
vsupRule.copyInto(supRule);
// compute confidence for each rule
for(i=0;i<total_rule;i++)
{
        supr=Integer.parseInt(supRule[i]);
        lhs=Integer.parseInt(RuleLeft[i]);
        conf1=((supr+0.0)/si[lhs])*100;
        Double con1=new Double(conf1);
        vcon.addElement(con1.toString());
}
// transfer members in vector vcon to array conf
String conf[];
conf=new String[total_rule];
vcon.copyInto(conf);
// copy array conf into array conf11
String conf11[];
conf11=new String[total_rule];
for(i=0;i<total_rule;i++)
conf11[i]=conf[i];
vcon.removeAllElements();
vRuleLeft.removeAllElements();
vRuleRight.removeAllElements();
vsupRule.removeAllElements();
// compute contrapositive confidence for each rule
for(i=0;i<total_rule;i++)
{
        String p=RuleLeft[i];
        String q=RuleRight[i];
        double conp1=(n-(si[Integer.parseInt(p)]-Integer.parseInt(supRule[i])+si
        [Integer.parseInt(q)]))/(0.0+n-si[Integer.parseInt(q)])*100;
        System.out.println(n+"-("+si[Integer.parseInt(p)]+"-"+Integer.parseInt
        (supRule[i])+"+"+si[Integer.parseInt(q)]+"))/(0.0+"+n+"-
        "+si[Integer.parseInt(q)]+")*100");
        Double conp11=new Double(conp1);
        String c1=conp11.toString();
        int len1=c1.length();
        if(len1>6)      c1=c1.substring(0,5);
        CPcon.addElement(c1);
}
// transfer members in vector CPcon to array Pcon
String Pcon[];
Pcon=new String[total_rule];
CPcon.copyInto(Pcon);
for(i=0;i<total_rule;i++)
{
        int len=conf[i].length();
```

```
        if (len>6)          conf[i]=conf[i].substring(0,5);
}
// sorting all rule by ascending confidence

for(j=0;j<total_rule;j++)
{
        int min=0;
        for(i=1;i<total_rule;i++)
        {
                Double a1=Double.valueOf(conf11[min]);
                Double a2=Double.valueOf(conf11[i]);
                double b1=a1.doubleValue();
                double b2=a2.doubleValue();
                if(b1>b2) min=i;
        }
        vorder.addElement(Integer.toString(min)); // store index of rule in vector
        vorder
        conf11[min]="999";
}
//----- end sorting-------//
// transfer members in vector vorder to array order1
String order1[];
order1=new String[total_rule];
vorder.copyInto(order1);
// display association rule, confidence, contrapositive confidence, x1,
x2,x3 and x4 value
FileWriter fout = new FileWriter("Order_AllRule.dat");
fout.write("rule  conf  contra_conf  x1   x2   x3   x4");
fout.write("\n");
for(i=0;i<total_rule;i++)
{
        l=Integer.parseInt(RuleLeft[Integer.parseInt(order1[i])]);
        System.out.print(l+"-->");
        fout.write(l+"-->");
        r=Integer.parseInt(RuleRight[Integer.parseInt(order1[i])]);
        System.out.print(r);
        fout.write(r+"");
        System.out.print("  "+conf[Integer.parseInt(order1[i])]);
        fout.write("  "+conf[Integer.parseInt(order1[i])]);
        System.out.print("  "+Pcon[Integer.parseInt(order1[i])]);
        fout.write("  "+Pcon[Integer.parseInt(order1[i])]);
        s=Integer.parseInt(supRule[Integer.parseInt(order1[i])]);
        sl=si[l];
        sr=si[r];
        x1=sl-s;                    // compute x1
        x2=sr-s;                    // compute x2
        x4=n-(sl+sr-s);             // compute x4
        System.out.print("  "+x1);
        fout.write("  "+x1);
```

```
                System.out.print("  "+x2);
                fout.write("  "+x2);
                System.out.print("  "+s);
                fout.write("  "+s);
                System.out.print("  "+x4);
                fout.write("  "+x4);
                System.out.print("\n");
                fout.write("\n");
            }
            System.out.print("total rule="+total_rule);
            fout.write("total rule="+total_rule);
            fout.close();
            //--------- end step3----------
            stmt.close();
            con.close();
        } //end try
        catch (SQLException ex)
        {
            System.out.println ("SQLException:");
            while (ex != null)
            {
                 System.out.println ("SQLState: "
                + ex.getSQLState());
                System.out.println ("Message:  "
                + ex.getMessage());
                System.out.println ("Vendor:   "
                + ex.getErrorCode());
                ex = ex.getNextException();
                System.out.println ("");
            }
        } //end catch
       catch (java.lang.Exception ex)
       {
            System.out.println("Exception: " + ex);
            ex.printStackTrace ();
        } // end catch
      }
    }
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวอนันตพร ศรีสวัสดิ์ เกิดเมื่อวันที่ 29 ตุลาคม พ.ศ. 2522 ที่จังหวัดราชบุรี สำเร็จการศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ สาขาวิทยาการคอมพิวเตอร์ เกียรตินิยมอันดับ 2 จากมหาวิทยาลัยศิลปากร ในปีการศึกษา 2542 และในปีการศึกษา 2543 ได้เข้าศึกษาในระดับปริญญาโท คณะวิทยาศาสตร์ สาขาวิทยาการคณนา ที่จุฬาลงกรณ์มหาวิทยาลัย และได้รับทุนอุดหนุนการศึกษาโครงการพัฒนาอาจารย์ สาขาขาดแคลนเพื่อศึกษาในประเทศ สาขาวิชาคอมพิวเตอร์ ปีการศึกษา 2544

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย