



การหาความสัมพันธ์ของการจราจรด้วยกฎการเชื่อมโยงโดยใช้ทฤษฎีกราฟเซต

โดย

นายปณณวัฒน์ ศรีรัตน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตร์มหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2558

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์

การหาความสัมพันธ์ของการจราจรด้วยกฎการเชื่อมโยงโดยใช้ทฤษฎีกราฟเซต

โดย

นายปณณวัฒน์ ศรีรัตน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตร์มหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

ปีการศึกษา 2558

ลิขสิทธิ์ของมหาวิทยาลัยธรรมศาสตร์



Association Mining based on Traffic Data by Using Rough Set Theory

BY

Mr. Pannawat Sriratna

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

FACULTY OF SCIENCE AND TECHNOLOGY

THAMMASAT UNIVERSITY

ACADEMIC YEAR 2015

COPYRIGHT OF THAMMASAT UNIVERSITY

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

นายปณณวัฒน์ ศรีรัตน์

เรื่อง

การหาความสัมพันธ์ของการจรรยาบรรณการเชื่อมโยงโดยใช้ทฤษฎีกราฟเซต

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)

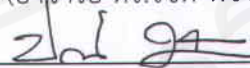
เมื่อ วันที่ 7 มกราคม พ.ศ. 2559

ประธานกรรมการสอบวิทยานิพนธ์



(อาจารย์ ดร.รัชต พิษวณิชย์)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์



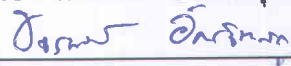
(อาจารย์ ดร.ปกรณ์ ลีสุทธิพรชัย)

กรรมการสอบวิทยานิพนธ์



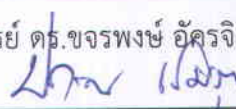
(อาจารย์ ดร.เด่นดวง ประดับสุวรรณ)

กรรมการสอบวิทยานิพนธ์



(อาจารย์ ดร.ชจรพงษ์ อัครจิตสกุล)

คณบดี



(รองศาสตราจารย์ ปกรณ์ เสริมสุข)

หัวข้อวิทยานิพนธ์	การหาความสัมพันธ์ของการจราจรด้วยกฎการเชื่อมโยงโดยใช้ทฤษฎีกราฟเซต
ชื่อผู้เขียน	นายปณณวัฒน์ ศรีรัตน์
ชื่อปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา/คณะ/มหาวิทยาลัย	สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์	อาจารย์ ดร.ปกรณ์ ลีสุทธิพรชัย
ปีการศึกษา	2558

บทคัดย่อ

ข้อมูลการจราจรบนทางหลวงพิเศษระหว่างเมืองหมายเลข 7 ได้รับความอนุเคราะห์จากกองทางหลวงพิเศษระหว่างเมือง กรมทางหลวง ข้อมูลการจราจรอยู่ในช่วงตั้งแต่วันที่ 1 มกราคม 2557 ถึง 30 เมษายน 2558 รายละเอียดประกอบด้วย วัน วันหยุด ช่วงเวลา ด่านขาเข้า ด่านขาออก จำนวนอุบัติเหตุ จำนวนรถยนต์ส่วนบุคคล จำนวนรถบรรทุก และจำนวน รถยนต์ทั้งหมด ชุดข้อมูลถูกนำมาวิเคราะห์เพื่อหาความสัมพันธ์ของการจราจร งานวิจัยนี้ได้ประยุกต์ใช้ทฤษฎีกราฟเซต (Rough Set Theory) เพื่อหาความสัมพันธ์ของการจราจรเปรียบเทียบกับผลที่ได้จากอัลกอริทึม Apriori ผลการทดลองแสดงให้เห็นว่าทฤษฎีกราฟเซตสามารถค้นหากฎการเชื่อมโยงที่เงื่อนไขของกฎไม่เป็นเซตย่อยของเงื่อนไขของกฎข้ออื่น โดยใช้เวลาในการประมวลผลน้อยกว่าอัลกอริทึม Apriori นอกจากนี้กฎการเชื่อมโยงที่มีจำนวนมากทำให้การอธิบายและการตัดสินใจเลือกกฎที่น่าสนใจทำได้ลำบาก งานวิจัยนี้นำเสนอการประยุกต์ใช้เครื่องมือเพื่อใช้ในการคัดกรองกฎการเชื่อมโยงที่น่าสนใจด้วยค่าคะแนนตัวชี้วัดความน่าสนใจ

คำสำคัญ: กฎการเชื่อมโยง, ทฤษฎีกราฟเซต, การจราจรของประเทศไทย

Thesis Title	Association Mining based on Traffic Data by Using Rough Set Theory
Author	Mr. Pannawat Sriratna
Degree	Master of Science
Major Field/Faculty/University	Science in Computer Science Department of Computer Science Faculty of Science and Technology Thammasat University
Thesis Advisor	Dr. Pakorn Leesutthipornchai
Academic Years	2015

ABSTRACT

The traffic data of highway no.7 which gathered from Division of Inter City Motorway, Department of Highways in the period from 1 January 2014 to 30 April 2015 including date, time, entering station, exit station, accident, number of private cars, number of trucks and vehicular traffic have been analyzed to find out how they related. This research applies theory of rough set to find relationship of traffic data comparing with the result from Apriori algorithm. The results show that rough set theory finds association rules which the rule condition is not subset of others with the less computational time than Apriori algorithm. Moreover, the obtained association results are numerous. This is difficult to describe and select. This research proposed the mechanism for selecting the set of interesting association rules by using interesting indicators.

Keywords: Association Rule, Rough Set Theory, Thai Highway Traffic

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความรู้และความช่วยเหลือเป็นอย่างดีจากอาจารย์ที่ปรึกษาวิทยานิพนธ์ อาจารย์ ดร.ปกรณ์ ลีสุทธิพรชัย ที่กรุณาเสียสละเวลาให้คำแนะนำ คำปรึกษา ตรวจสอบแก้ไขข้อบกพร่องด้วยความเอาใจใส่ ผู้วิจัยรู้สึกซาบซึ้งในความกรุณาของอาจารย์และขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอกราบขอบพระคุณ อาจารย์ ดร.รัชต์ พิษฉนิชย์ อาจารย์ ดร.เด่นดวง ประดับสุวรรณ และอาจารย์ ดร.ขจรพงษ์ อัครจิตสกุล กรรมการสอบวิทยานิพนธ์ ที่ให้คำแนะนำและเสียสละเวลาในการตรวจสอบวิทยานิพนธ์ และขอขอบคุณเจ้าหน้าที่ภาควิชาวิทยาการคอมพิวเตอร์ที่ช่วยอำนวยความสะดวกในการศึกษาตลอดหลักสูตรให้เป็นไปด้วยความเรียบร้อย

ขอขอบพระคุณ นายเผชฌิม หุนตระณี ผู้อำนวยการกลุ่มอำนวยความสะดวกแก่ค่าธรรมเนียมผ่านทาง กองทางหลวงพิเศษระหว่างเมือง กรมทางหลวง และเจ้าหน้าที่ทุกท่าน ที่ได้อำนวยความสะดวกในการดำเนินงานวิจัยและให้ความรู้เกี่ยวกับการจราจรบนทางหลวงพิเศษระหว่างเมืองหมายเลข 7

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณครอบครัวศรีรัตน์ มารดา และน้องสาวที่เป็นขวัญ กำลังใจ สนับสนุน และอยู่เคียงข้างมาโดยตลอด และขอบคุณเพื่อน ๆ ทุกคนที่ให้ความช่วยเหลือในการจัดทำวิทยานิพนธ์นี้สำเร็จลุล่วงไปด้วยดี

นายปณณวัฒน์ ศรีรัตน์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	(1)
บทคัดย่อภาษาอังกฤษ.....	(2)
กิตติกรรมประกาศ.....	(3)
สารบัญตาราง.....	(7)
สารบัญภาพ.....	(8)
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของงานวิจัย.....	1
1.2 วัตถุประสงค์การทำวิจัย.....	3
1.3 ขอบเขตของการวิจัย.....	3
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	3
1.5 นิยามศัพท์เฉพาะ.....	4
1.6 รายละเอียดของเค้าโครงวิทยานิพนธ์.....	4
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	5
2.1 กฎการเชื่อมโยง (Association Rules).....	5
2.1.1 นิยามเบื้องต้นของกฎการเชื่อมโยง.....	5
2.1.2 อัลกอริทึม Apriori.....	7
2.1.3 ทฤษฎีเซตหยาบ (Rough Set Theory).....	10

2.1.3.1 ระบบสารสนเทศ (Information System).....	10
2.1.3.2 ความสัมพันธ์ที่ไม่สามารถมองเห็นได้ (Indiscernibility Relation).....	11
2.1.3.3 ค่าประมาณขอบเขตล่างและขอบเขตบน (Lower and Upper Approximations)	12
2.1.2.4 กฎการตัดสินใจ (Decision Rule).....	14
2.2 ตัวชี้วัดความน่าสนใจ (Interesting Indicators).....	15
2.2.1 Confidence.....	15
2.2.1 Coverage.....	15
2.2.2 Leverage.....	15
2.2.3 Lift.....	16
2.2.4 Generalization.....	16
2.3 งานวิจัยที่เกี่ยวข้อง.....	18
บทที่ 3 วิธีการดำเนินงานวิจัย.....	22
3.1 การเตรียมข้อมูล (Data Pre-processing) และการวิเคราะห์ข้อมูล.....	23
3.1.1 แหล่งที่มาของข้อมูล.....	24
3.1.2 การวิเคราะห์ข้อมูล.....	24
3.1.3 การแปลงข้อมูล.....	26
3.2 การแบ่งช่วงข้อมูล (Discretization).....	28
3.2.1 การแบ่งช่วงคุณลักษณะ.....	28
3.2.2 การแปลงชุดข้อมูลตามช่วงของข้อมูล.....	30
3.3 ขั้นตอนการทดลอง.....	31
3.4 วิธีการวัดผลการทดลอง.....	31
3.5 เครื่องมือและซอฟต์แวร์ที่ใช้ในการทดลอง.....	32
บทที่ 4 ผลการทดลอง.....	33
4.1 ข้อมูลทดลอง.....	33
4.2 เปรียบเทียบจำนวนกฎการเชื่อมโยง.....	34

4.3 เปรียบเทียบเวลาที่ใช้ในการประมวลผล.....	35
4.4 เปรียบเทียบกฎการเชื่อมโยงจากตัวชี้วัดความน่าสนใจ (Interesting Indicators).....	35
4.4.1 กฎการเชื่อมโยงที่ปรากฏไม่บ่อย.....	36
4.4.2 กฎการเชื่อมโยงที่ปรากฏบ่อย.....	38
บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ.....	41
5.1 สรุปผลการวิจัย.....	41
5.2 ข้อเสนอแนะ.....	42
5.3 งานวิจัยในอนาคต.....	42
รายการอ้างอิง.....	43
ประวัติผู้เขียน.....	45

สารบัญตาราง

ตารางที่	หน้า
2.1 ตัวอย่างข้อมูลรายการซื้อสินค้า.....	6
2.2 ตัวอย่างค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence).....	7
2.3 กฎการเชื่อมโยงทั้งหมดที่ถูกสร้างโดย อัลกอริทึม Apriori.....	9
2.4 ตัวอย่างระบบสารสนเทศ.....	11
2.5 แสดงคะแนนจากตัวชี้วัดความน่าสนใจ (Interesting Indicators).....	17
2.6 ตัวอย่างคุณลักษณะของข้อมูลที่ได้จากการเกิดอุบัติเหตุ.....	19
2.7 แสดงช่วงข้อมูล.....	20
2.8 ตัวอย่างข้อมูลหลังจากทำการแปลงค่าตามช่วงข้อมูล.....	21
3.1 แสดงรายละเอียดและคำอธิบายของบัตรผ่านทาง.....	25
3.2 โครงสร้างข้อมูลส่วนขาออก.....	26
3.3 คุณลักษณะ (Attributes) และคำอธิบาย.....	27
3.4 แสดงข้อมูลจราจรในรูปแบบเรคคอร์ด (Record-data Format).....	27
3.5 ช่วงกำหนดค่าของคุณลักษณะ.....	28
3.6 ตัวอย่างข้อมูลที่ใช้คำนวณหาปริมาณความหนาแน่นการจราจร.....	29
3.7 ตัวอย่างข้อมูลหลังจากทำการแปลงค่าตามช่วงของข้อมูล.....	30
4.1 แสดงจำนวนแถว (Rows) ในแต่ละช่วงของคุณลักษณะที่ตติสนใจ.....	33
4.2 แสดงจำนวนกฎการเชื่อมโยง.....	34
4.3 แสดงกฎการเชื่อมโยงที่ปรากฏไม่บ่อยของทฤษฎีเซต (Rough Set Theory).....	37
4.4 แสดงกฎการเชื่อมโยงที่ปรากฏไม่บ่อยของอัลกอริทึม Apriori.....	38
4.5 แสดงกฎการเชื่อมโยงที่ปรากฏบ่อยของทฤษฎีเซต (Rough Set Theory).....	40
4.6 แสดงกฎการเชื่อมโยงที่ปรากฏบ่อยของอัลกอริทึม Apriori.....	40

สารบัญภาพ

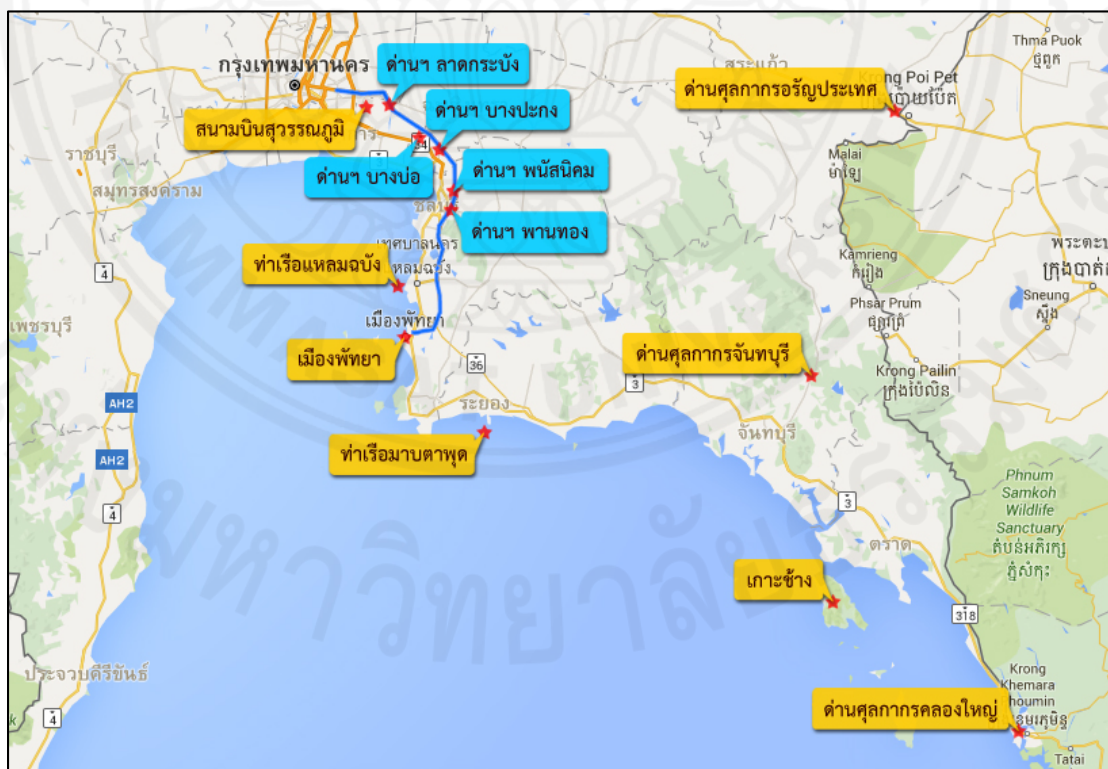
ภาพที่	หน้า
1.1 แสดงเส้นทางหลวงพิเศษหมายเลข 7 และสถานที่สำคัญของภาคตะวันออก.....	1
1.2 แสดงปริมาณจราจรบนทางหลวงพิเศษหมายเลข 7.....	2
2.1 การค้นหา Frequent Itemsets รอบที่ 1.....	8
2.2 การค้นหา Frequent Itemsets รอบที่ 2.....	8
2.3 การค้นหา Frequent Itemsets รอบที่ 3.....	9
2.4 ค่าประมาณของเซตคุณลักษณะ Season.....	13
2.5 ตัวอย่างข้อมูลที่ใช้ในทฤษฎีร์ฟเซต (Rough Set Theory).....	14
3.1 ภาพรวมขั้นตอนการทดลอง.....	22
3.2 ตัวอย่างรายงานข้อมูลการจราจร.....	23
3.3 ขั้นตอนการค้นหากฎการเชื่อมโยง.....	31
4.1 ร้อยละเปรียบเทียบจำนวนแถวข้อมูลในแต่ละช่วงของคุณลักษณะตัดสินใจ.....	33
4.2 ร้อยละเปรียบเทียบจำนวนกฎการเชื่อมโยง.....	34
4.3 แสดงเวลาในการประมวลผล.....	35
4.4 แสดงคะแนนกฎการเชื่อมโยงที่ปรากฏไม่บ่อยของทฤษฎีร์ฟเซต (Rough Set Theory).....	36
4.5 แสดงคะแนนกฎการเชื่อมโยงที่ปรากฏไม่บ่อยของอัลกอริทึม Apriori.....	37
4.6 แสดงคะแนนกฎการเชื่อมโยงที่ปรากฏบ่อยของทฤษฎีร์ฟเซต (Rough Set Theory).....	39
4.7 แสดงคะแนนกฎการเชื่อมโยงที่ปรากฏบ่อยของอัลกอริทึม Apriori.....	39

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของงานวิจัย

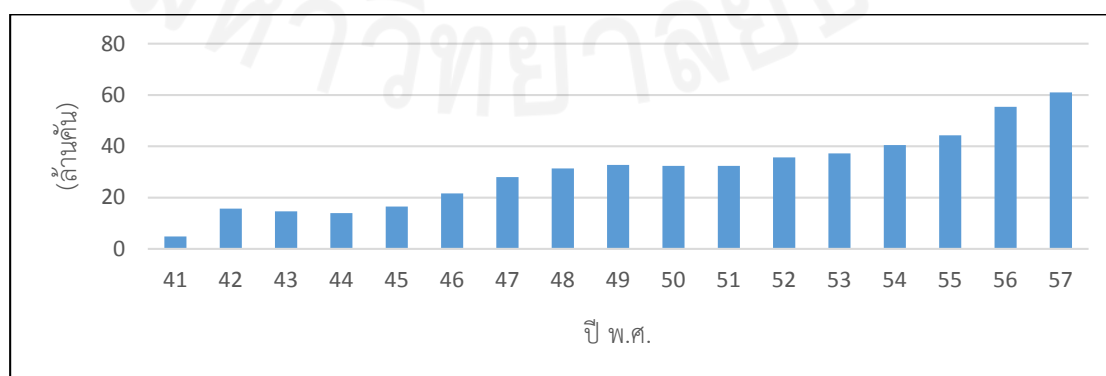
ทางหลวงพิเศษระหว่างเมืองหมายเลข 7 เป็นทางหลวงพิเศษระหว่างเมืองสายแรกของประเทศไทย เริ่มต้นที่ปลายทางพิเศษศรีรัชและถนนพระราม 9 บริเวณจุดตัดกับถนนศรีนครินทร์ ที่ทางแยกต่างระดับศรีนครินทร์ เขตสวนหลวง กรุงเทพมหานคร ปัจจุบันสิ้นสุดที่ถนนสุขุมวิท เมืองพัทยา จังหวัดชลบุรี มีระยะทางยาว 125 กิโลเมตร ทางหลวงพิเศษระหว่างเมืองหมายเลข 7 เป็นเส้นทางหลักในการคมนาคมขนส่งระหว่างกรุงเทพมหานครกับภาคตะวันออกและสามารถเชื่อมโยงเข้าสู่ท่าอากาศยานสุวรรณภูมิ นอกจากนี้ยังได้รับการกำหนดให้เป็นส่วนหนึ่งของทางหลวงเอเชียสาย 19 (AH19) และทางหลวงเอเชียสาย 123 (AH123) ดังแสดงในภาพที่ 1.1



ภาพที่ 1.1 แสดงเส้นทางหลวงพิเศษหมายเลข 7 และสถานที่สำคัญของภาคตะวันออก

จากภาพที่ 1.1 ภาคตะวันออกมีลักษณะภูมิประเทศเป็นที่ราบระหว่างภูเขา ที่ราบลูกฟูก ที่ราบลุ่มแม่น้ำ และที่ราบชายฝั่งทะเล อาณาเขตทิศเหนือจรดภาคกลางและภาคตะวันออกเฉียงเหนือ ทิศตะวันออกมีพรมแดนติดกับประเทศกัมพูชาประชาธิปไตย ทิศใต้และทิศตะวันตกเป็นพื้นที่ชายฝั่งทะเลอ่าวไทย จากสภาพภูมิศาสตร์ดังกล่าวทำให้ภาคตะวันออกมีปัจจัยพื้นฐานค่อนข้างที่จะสมบูรณ์ ทั้งด้านเกษตรกรรม แหล่งท่องเที่ยว ทรัพยากรธรรมชาติ อัญมณี และโครงข่ายการคมนาคมขนส่ง ทั้งทางบกและทางน้ำ ภาคตะวันออกจึงมีบทบาทและความสำคัญต่อเศรษฐกิจของประเทศเป็นอย่างมาก เป็นที่ตั้งของนิคมอุตสาหกรรมขนาดใหญ่ เช่น นิคมอุตสาหกรรมอมตะนคร นิคมอุตสาหกรรมมาบตาพุด และนิคมอุตสาหกรรมแหลมฉบัง เป็นต้น พื้นที่ชายฝั่งทะเลเป็นที่ตั้งท่าเรือน้ำลึก โดยมีท่าเรือแหลมฉบัง เป็นประตูการนำเข้าและส่งออกสินค้าหลักของประเทศ อีกบทบาทหนึ่งที่สำคัญมากคือการเป็นประตูเชื่อมโยงทางการค้าระหว่างไทย-กัมพูชา โดยมีด่านการค้าชายแดนที่สำคัญ 3 ด่าน ได้แก่ ด่านศุลกากร อรัญประเทศ ด่านศุลกากรคลองใหญ่ และด่านศุลกากรจันทบุรี ขณะที่อุตสาหกรรมการท่องเที่ยวนั้น พื้นที่ชายฝั่งทะเลมีลักษณะโค้งเว้าและเรียวยาวจึงทำให้มีสถานที่ท่องเที่ยวต่างๆ เช่น ชายหาดพัทยา ชายหาดบางแสน เกาะล้าน เกาะกูด เกาะช้าง เป็นต้น

ทางหลวงพิเศษระหว่างเมืองหมายเลข 7 ถูกสร้างเพื่อกระจายความหนาแน่นการจราจร ของถนนสุขุมวิทและถูกกำหนดให้เป็นเส้นทางหลักในการขนส่งสินค้าระหว่างกรุงเทพมหานครกับ ภาคตะวันออก จากสถิติการจราจรตั้งแต่ปี พ.ศ. 2541 จนถึงปี พ.ศ. 2557 ดังแสดงในภาพที่ 1.2 แสดงให้เห็นถึงปัญหาสภาพการจราจรที่หนาแน่นและมีแนวโน้มที่เพิ่มปริมาณการจราจรในอัตราที่ค่อนข้างสูงขึ้นในอนาคตจากอัตราการเจริญเติบโตอย่างรวดเร็วของอุตสาหกรรมการท่องเที่ยวและ นิคมอุตสาหกรรม หลังการเปิดประชาคมเศรษฐกิจอาเซียน (ASEAN Economic Community : AEC) ปัจจุบันทางหลวงพิเศษระหว่างเมืองหมายเลข 7 มีด่านเก็บค่าธรรมเนียม จำนวน 5 ด่าน ได้แก่ ด่านลาดกระบัง ด่านบางบ่อ ด่านบางปะกง ด่านพนัสนิคม และด่านพานทอง ซึ่งทุกด่านสามารถเชื่อมโยง หักกันได้และเป็นเส้นทางเชื่อมโยงไปยังสถานที่สำคัญ ๆ ของภาคตะวันออก



ภาพที่ 1.2 แสดงปริมาณจราจรบนทางหลวงพิเศษหมายเลข 7

งานวิจัยนี้ได้นำข้อมูลการจราจรบนทางหลวงพิเศษระหว่างเมืองหมายเลข 7 จากกองทางหลวงพิเศษระหว่างเมือง กรมทางหลวง ตั้งแต่วันที่ 1 มกราคม 2557 ถึง 30 เมษายน 2558 มีข้อมูลทั้งหมด 242,066 แถว และมีจำนวนรถทั้งหมด 60,766,872 คัน มาวิเคราะห์หาความสัมพันธ์ระหว่าง วัน วันหยุด ช่วงเวลา ด้านขาเข้า ด้านขาออก จำนวนอุบัติเหตุ จำนวนรถยนต์ส่วนบุคคล จำนวนรถบรรทุก กับปริมาณความหนาแน่นของการจราจร โดยใช้ทฤษฎีเซต (Rough Set Theory) เพื่อหาความสัมพันธ์ของการจราจรเปรียบเทียบกับผลที่ได้จากอัลกอริทึม Apriori ทฤษฎีเซต (Rough Set Theory) สามารถค้นหากฎการเชื่อมโยงที่มีความชัดเจนได้โดยใช้เวลาในการประมวลผลน้อยกว่า นอกจากนี้กฎการเชื่อมโยงที่ได้จากอัลกอริทึม Apriori มีจำนวนมากทำให้การอธิบายและการตัดสินใจเลือกกฎทำได้ด้วยความยากลำบาก งานวิจัยนี้จึงนำเสนอเครื่องมือเพื่อช่วยเปรียบเทียบคุณภาพข้อมูลและใช้ในการคัดกรองกฎการเชื่อมโยงที่น่าสนใจจากกฎการเชื่อมโยงที่ได้ทั้งหมดด้วยค่าคะแนนตัวชี้วัดความน่าสนใจ

1.2 วัตถุประสงค์การทำวิจัย

1.2.1 เพื่อศึกษาความสัมพันธ์ของการจราจร โดยใช้ทฤษฎีเซต (Rough Set Theory)

1.3 ขอบเขตของการวิจัย

1.3.1 ข้อมูลที่ใช้ในการทดลอง คือ ปริมาณการจราจรบนทางหลวงพิเศษหมายเลข 7 ตั้งแต่วันที่ 1 มกราคม 2557 ถึงวันที่ 30 เมษายน 2558

1.3.2 การทดลองจะวิเคราะห์และเปรียบเทียบข้อมูลเชิงคุณภาพระหว่างทฤษฎีเซต (Rough Set Theory) และอัลกอริทึม Apriori

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 สามารถแสดงความสัมพันธ์ของการจราจรออกมาในรูปของกฎการเชื่อมโยงได้

1.5 นิยามศัพท์เฉพาะ

1.5.1 กฎการเชื่อมโยง คือ การค้นหาความสัมพันธ์ของข้อมูลที่ซ่อนอยู่ในชุดข้อมูล และนำรูปแบบที่ค้นหาได้มาสร้างเป็นกฎซึ่งประกอบด้วยคุณลักษณะที่เป็นเงื่อนไขและผลสรุป

1.5.2 ทฤษฎีรีฟเซต (Rough Set Theory) เป็นทฤษฎีที่แบ่งข้อมูลออกเป็นเซตของคำตอบหลายเซต และหาความสัมพันธ์ของข้อมูลภายในเซต ทฤษฎีนี้สนใจคุณภาพของข้อมูลมากกว่าปริมาณข้อมูล หมายความว่า กฎการเชื่อมโยงที่ได้อาจจะปรากฏไม่บ่อย แต่ทุกครั้งที่ปรากฏจะเป็นไปตามเงื่อนไข

1.6 รายละเอียดของเค้าโครงวิทยานิพนธ์

รายละเอียดของเค้าโครงวิทยานิพนธ์ แบ่งออกเป็น 5 บท ได้แก่

บทที่ 1 กล่าวถึงความเป็นมาและความสำคัญของงานวิจัย วัตถุประสงค์การทำวิจัย ขอบเขตของการวิจัย ประโยชน์ที่คาดว่าจะได้รับ นิยามศัพท์เฉพาะ

บทที่ 2 กล่าวถึงทฤษฎีที่เกี่ยวข้องกับงานวิจัยและงานวิจัยที่เกี่ยวข้อง

บทที่ 3 กล่าวถึงวิธีการวิจัย ประกอบด้วย ขั้นตอนการดำเนินงานวิจัย การเตรียมชุดข้อมูลที่ใช้ในงานวิจัย เครื่องมือที่ใช้ในการทดลอง และวิธีการวัดผลการวิจัย

บทที่ 4 กล่าวถึงผลการดำเนินงานวิจัย ประกอบด้วย การวิเคราะห์ข้อมูลที่ใช้ในการทดลอง การเปรียบเทียบจำนวนกฎการเชื่อมโยง การเปรียบเทียบเวลาที่ใช้ในการประมวลผล และการเปรียบเทียบกฎการเชื่อมโยงที่น่าสนใจ

บทที่ 5 กล่าวถึงผลสรุปการดำเนินงานวิจัย ข้อเสนอแนะ และงานวิจัยในอนาคต

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 กฎการเชื่อมโยง (Association Rules)

กฎการเชื่อมโยง (Association Rules) เป็นผลที่ได้จากระบวนการหาความสัมพันธ์ของข้อมูล กระบวนการหาความสัมพันธ์นี้สามารถค้นพบรูปแบบ (Pattern) ที่น่าสนใจสามารถบ่งบอกถึงลักษณะเฉพาะหรือคุณสมบัติเด่นของชุดข้อมูล กฎการเชื่อมโยงถูกนำไปใช้อย่างแพร่หลายในด้านต่าง ๆ เช่น การวิเคราะห์พฤติกรรมของลูกค้า การทดลองทางวิทยาศาสตร์ การวินิจฉัยโรคทางการแพทย์ และการทำนายปรากฏการณ์ธรรมชาติ เป็นต้น

2.1.1 นิยามเบื้องต้นของกฎการเชื่อมโยง

กฎการเชื่อมโยง คือ กฎที่อธิบายความสัมพันธ์ของข้อมูลในฐานข้อมูล จะอยู่ในรูปของการทำนายว่า เมื่อมีเหตุการณ์เกิดขึ้นแล้วจะมีเหตุการณ์ใดเกิดขึ้นตามมาด้วย ด้วยค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) ตามที่กำหนด

นิยามกฎการเชื่อมโยง กำหนดให้

$D = \{T_1, T_2, \dots, T_n\}$ คือ ฐานข้อมูลซึ่งประกอบด้วยเซตของข้อมูล n ตัว

$I = \{I_1, I_2, \dots, I_n\}$ คือ ข้อมูลที่ใช้ในการหากฎการเชื่อมโยง เรียกว่า ไอเทม

(Items)

T คือ ทรานแซคชัน (Transactions) ซึ่งเป็นเซตของข้อมูล (Item) โดยที่

$T \subseteq I$

กฎการเชื่อมโยงสามารถเขียนในรูปแบบ $X \rightarrow Y$

โดยที่ $X, Y \subseteq I$ และ $X \cap Y = \emptyset$

ค่าสนับสนุน (Support) คือ สัดส่วนของจำนวนทรานแซคชันที่มีไอเทมของกฎเป็นสมาชิกต่อจำนวนทรานแซคชันทั้งหมด ดังนั้น ค่า $\text{Support}(X \rightarrow Y)$ คำนวณได้ตามสมการที่ 1

$$\text{Support}(X \rightarrow Y) = P(X \cup Y) = \frac{\text{no. of transactions containing both } X \text{ and } Y}{\text{total no. of transactions}} \quad (1)$$

ค่าความเชื่อมั่น (Confidence) คือ สัดส่วนของจำนวนทรานแซคชันที่สอดคล้องตามกฎต่อจำนวนทรานแซคชันทั้งหมดที่มีไอเทมเงื่อนไขของกฎเป็นสมาชิก ดังนั้น $\text{Confidence}(X \rightarrow Y)$ คำนวณได้ตามสมการที่ 2

$$\text{Confidence}(X \rightarrow Y) = P(X | Y) = \frac{\text{no. of transactions containing both X and Y}}{\text{total no. of transactions containing X}} \quad (2)$$

ตัวอย่างวิธีการสืบค้นหากฎการเชื่อมโยงจากฐานข้อมูลการซื้อสินค้า โดยรายการซื้อสินค้า แสดงตามตารางที่ 2.1 กำหนดให้

Tid คือ รายการซื้อสินค้า (Transaction)

Items คือ สินค้าที่ปรากฏในรายการซื้อสินค้า

A, B, C, D และ E คือ ซื้อสินค้าแต่ละชนิด

ตารางที่ 2.1

ตัวอย่างข้อมูลรายการซื้อสินค้า

Tid	Items		
1	A	C	D
2	B	C	E
3	A	B	E
4	B		E

จากตัวอย่างข้อมูลรายการซื้อสินค้าในตารางที่ 2.1 สามารถทำการคำนวณหาค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) ของกฎการเชื่อมโยง จะได้ค่าสนับสนุนและค่าความเชื่อมั่น ตามตารางที่ 2.2

ตัวอย่างเช่น กฎข้อที่ 4 คือ $\{B, C \rightarrow E\}$ มีค่าสนับสนุน เท่ากับ 50% และค่าความเชื่อมั่น เท่ากับ 100% หมายความว่า มีการซื้อสินค้า B, C และ E พร้อมกันปรากฏอยู่ครึ่งหนึ่งของจำนวนทรานแซคชันทั้งหมด และร้อยละ 100 ของลูกค้าที่ซื้อสินค้า B และ C จะซื้อสินค้า E ไปด้วย

กฎการเชื่อมโยงเป็นการอธิบายความสัมพันธ์ภายในเซตย่อย $\{B, C, E\}$ ของข้อมูลว่า ร้อยละ 100 ของลูกค้าที่ซื้อสินค้า B และ C จะซื้อสินค้า E ไปด้วย แต่ไม่ได้หมายความว่าทุกทรานแซคชันต้องมีการซื้อสินค้า B และ C

ตารางที่ 2.2

ตัวอย่างค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence)

ค่าสนับสนุน (Support)	ค่าความเชื่อมั่น (Confidence)
1) $\{A \rightarrow B\} = 1/4 = 25\%$	1) $\{A \rightarrow B\} = 1/2 = 50\%$
2) $\{B \rightarrow C\} = 2/4 = 50\%$	2) $\{B \rightarrow C\} = 2/3 = 67\%$
3) $\{B \rightarrow E\} = 3/4 = 75\%$	3) $\{B \rightarrow E\} = 3/3 = 100\%$
4) $\{B, C \rightarrow E\} = 2/4 = 50\%$	4) $\{B, C \rightarrow E\} = 2/2 = 100\%$
...	...

หมายเหตุ ค่าสนับสนุน (Support) ทั้งหมด เท่ากับ 4

2.1.2 อัลกอริทึม Apriori

อัลกอริทึม Apriori เป็นอัลกอริทึมที่ได้รับความนิยมและเป็นที่ยอมรับกันดี ใช้สำหรับค้นหาความสัมพันธ์ของไอเทมเซต (Itemsets) ที่เกิดขึ้นบ่อย (Frequent Itemsets) ในฐานข้อมูล โดยความสัมพันธ์ที่ได้จากไอเทมเซต นั้นเรียกว่า กฎการเชื่อมโยง (Association Rules) กฎการเชื่อมโยงที่ได้จะพิจารณาจากค่าสนับสนุนขั้นต่ำ (Minimum Support) และค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) ซึ่งจะถูกระบุโดยผู้ใช้เอง (User) หรือผู้เชี่ยวชาญ (Expert User) ขั้นตอนการทำงานของอัลกอริทึม Apriori อธิบายได้ดังนี้

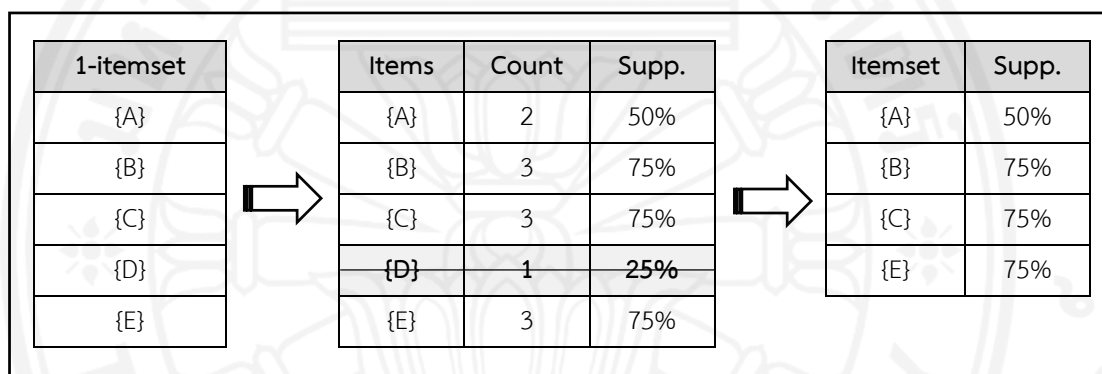
1. ตรวจสอบไอเทมเซต (Itemsets) ทั้งหมดที่มีจำนวนสมาชิกเท่ากับ 1 และตัดไอเทมเซต (Itemsets) ที่มีค่าสนับสนุนน้อยกว่าค่าสนับสนุนที่กำหนด
2. สร้างไอเทมเซต (Itemsets) ในระดับขั้นถัดไปโดยการเพิ่มจำนวนสมาชิกแล้วทำการตรวจสอบและตัดไอเทมเซต (Itemsets) ที่มีค่าสนับสนุนน้อยกว่าค่าสนับสนุนที่กำหนด
3. การทำงานของอัลกอริทึม Apriori จะทำงานวนตามขั้นตอนที่ 2 ไปเรื่อย ๆ จนกระทั่งได้ครบทุกระดับขั้นหรือไม่เหลือไอเทมเซต (Itemsets) ในลำดับขั้นต่อไป
4. เมื่อได้ครบทุกระดับขั้นแล้วจะนำไอเทมเซต (Itemsets) ที่ได้มาสร้างเป็นกฎการเชื่อมโยง เขียนแทนด้วย $LHS \rightarrow RHS$ โดยที่ LHS (Left Hand Side) คือ รูปแบบไอเทมเซต (Itemsets) ด้านซ้ายของกฎการเชื่อมโยง และ RHS (Right Hand Side) คือ รูปแบบไอเทมเซต (Itemsets) ด้านขวาของกฎการเชื่อมโยง

ค่าความเชื่อมั่น (Confidence) คำนวณได้ตามสมการที่ 3

$$\text{Confidence}(LHS \rightarrow RHS) = \frac{\text{Support}(LHS, RHS)}{\text{Support}(LHS)} \quad (3)$$

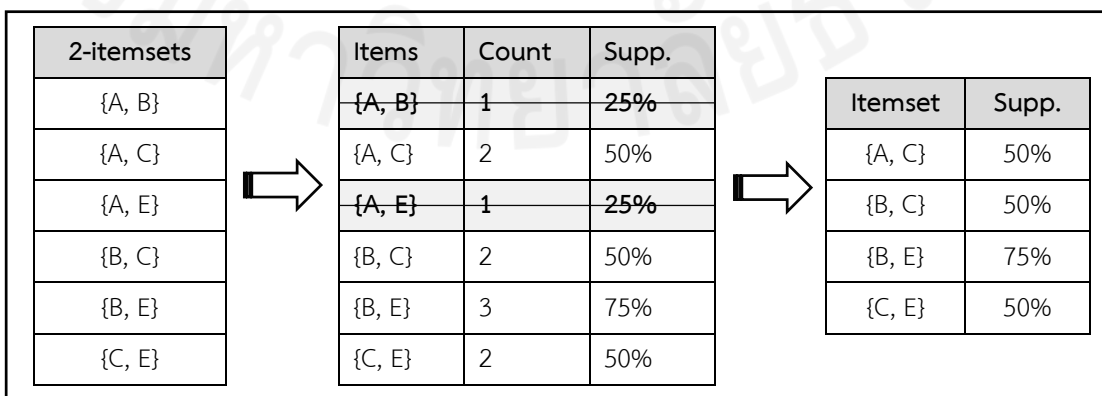
จากตารางที่ 2.1 สามารถประยุกต์ใช้อัลกอริทึม Apriori ค้นหากฎการเชื่อมโยงข้อมูลการซื้อสินค้า โดยกำหนดให้ ค่าสนับสนุนขั้นต่ำเท่ากับร้อยละ 50 และค่าความเชื่อมั่นเท่ากับร้อยละ 70 อธิบายการทำงานของอัลกอริทึม Apriori ได้ดังนี้

รอบที่ 1 อัลกอริทึม Apriori ทำการตรวจสอบไอเทมเซต (Itemsets) ทั้งหมดที่มีจำนวนสมาชิกเท่ากับหนึ่ง ได้แก่ {A}, {B}, {C}, {D} และ {E} จากนั้นคำนวณหาค่าสนับสนุนของแต่ละไอเทมเซต สุดท้ายอัลกอริทึม Apriori จะตัดไอเทมเซตที่มีค่าสนับสนุนน้อยกว่าค่าสนับสนุนที่กำหนด ซึ่งไอเทมเซตที่มีค่าสนับสนุนน้อยกว่าค่าที่กำหนด คือ {D} มีค่าเท่ากับร้อยละ 25 ดังแสดงในภาพที่ 2.1



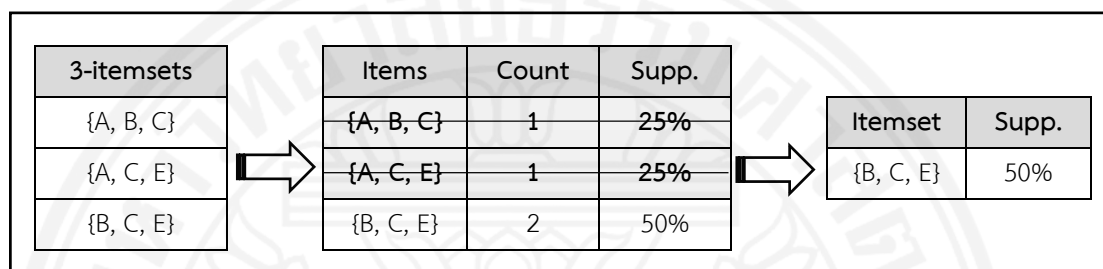
ภาพที่ 2.1 การค้นหา Frequent Itemsets รอบที่ 1

รอบที่ 2 อัลกอริทึม Apriori ทำการสร้างไอเทมเซต (Itemsets) ในระดับขั้นถัดไปโดยการเพิ่มจำนวนสมาชิก ได้แก่ {A, B}, {A, C}, {A, E}, {B, C}, {B, E} และ {C, E} จากนั้นคำนวณหาค่าสนับสนุนของแต่ละไอเทมเซต สุดท้ายอัลกอริทึม Apriori จะตัดไอเทมเซตที่มีค่าสนับสนุนน้อยกว่าค่าสนับสนุนที่กำหนด ซึ่งไอเทมเซตที่มีค่าสนับสนุนน้อยกว่าค่าที่กำหนด คือ {A, B} มีค่าเท่ากับร้อยละ 25 และ {A, E} มีค่าเท่ากับร้อยละ 25 ดังแสดงในภาพที่ 2.2



ภาพที่ 2.2 การค้นหา Frequent Itemsets รอบที่ 2

รอบที่ 3 อัลกอริทึม Apriori ทำการสร้างไอเทมเซต (Itemsets) ในระดับชั้นถัดไปโดยการเพิ่มจำนวนสมาชิก ได้แก่ {A, B, C}, {A, C, E} และ {B, C, E} จากนั้นคำนวณหาค่าสนับสนุนของแต่ละไอเทมเซต สุดท้ายทำการตัดไอเทมเซตที่มีค่าสนับสนุนน้อยกว่า ค่าสนับสนุนที่กำหนด ซึ่งไอเทมเซตที่มีค่าสนับสนุนน้อยกว่าค่าที่กำหนด คือ {A, B, C} มีค่าเท่ากับร้อยละ 25 และ {A, C, E} มีค่าเท่ากับร้อยละ 25 ดังแสดงในภาพที่ 2.3



ภาพที่ 2.3 การค้นหา Frequent Itemsets รอบที่ 3

รอบที่ 4 เมื่ออัลกอริทึม Apriori ทำการตรวจสอบไอเทมเซต (Itemsets) จนครบทุกระดับชั้นแล้วจะนำไอเทมเซตที่ได้ คือ {A, C}, {B, C}, {B, E}, {C, E} และ {B, C, E} มาสร้างเป็นกฎการเชื่อมโยง จากนั้นคำนวณหาค่าความเชื่อมั่น สุดท้ายทำการตัดกฎการเชื่อมโยงที่มีค่าความเชื่อมั่นน้อยกว่าค่าที่กำหนด กฎการเชื่อมโยงที่ได้ทั้งหมดแสดงในตารางที่ 2.3 อธิบายตัวอย่างกฎข้อที่ 14 คือ $\{C, E\} \rightarrow \{B\}$ หมายความว่า ความน่าจะเป็นเมื่อมีการซื้อสินค้า C และ E พร้อมกันแล้วจะซื้อสินค้า B ด้วยคิดเป็นร้อยละ 100

ตารางที่ 2.3

กฎการเชื่อมโยงทั้งหมดที่ถูกสร้างโดย อัลกอริทึม Apriori

ลำดับ	กฎการเชื่อมโยง	ค่าสนับสนุน (Support)	ค่าความเชื่อมั่น (Confidence)
1	{A} → {C}	50%	100%
2	{C} → {A}	50%	67%
3	{B} → {C}	50%	67%
4	{C} → {B}	50%	67%
5	{B} → {E}	75%	100%
6	{E} → {B}	75%	100%
7	{C} → {E}	50%	67%
8	{E} → {C}	50%	67%

ตารางที่ 2.3

กฎการเชื่อมโยงทั้งหมดที่ถูกสร้างโดย อัลกอริทึม Apriori (ต่อ)

ลำดับ	กฎการเชื่อมโยง	ค่านับสนับสนุน (Support)	ค่าความเชื่อมั่น (Confidence)
9	{B} → {C, E}	50%	67%
10	{C} → {B, E}	50%	67%
11	{E} → {B, C}	50%	67%
12	{B, C} → {E}	50%	100%
13	{B, E} → {C}	50%	67%
14	{C, E} → {B}	50%	100%

2.1.3 ทฤษฎีร์ฟเซต (Rough Set Theory)

ในปี ค.ศ. 1980 Zdzislaw Pawlak ได้เสนอทฤษฎีร์ฟเซต (Rough Set Theory) ซึ่งเป็นระบบการค้นพบความรู้แบบใหม่และชี้ให้เห็นว่าวิธีการร์ฟเซตสามารถนำมาใช้เพื่อเพิ่มโอกาสการคาดการณ์ที่ถูกต้องและแม่นยำโดยการระบุและลบตัวแปรที่ซ้ำซ้อน การประยุกต์ใช้ทฤษฎีร์ฟเซตช่วยเพิ่มประสิทธิภาพในการวิเคราะห์ข้อมูล การตัดสินใจ การค้นพบลักษณะการเชื่อมโยงระหว่างข้อมูลและการค้นพบรูปแบบที่น่าสนใจ

ความสัมพันธ์ที่พบไม่บ่อยสามารถถูกค้นพบโดยใช้วิธีการร์ฟเซต เช่น จากข้อมูลสภาพอากาศ 100 วัน แบ่งเป็นวันที่มีแดด 80 วัน วันที่ท้องฟ้ามีเมฆ 10 วัน และวันที่ฝนตก 10 วัน แต่ทุกครั้งที่ฝนตกมักจะจะมีลมแรงเสมอ การปรากฏข้อมูลในวันที่ฝนตกเกิดขึ้นเพียงร้อยละ 10 แต่ทุกครั้งที่ฝนตกจะเกิดลมแรงเสมอ เป็นความสัมพันธ์ที่เกิดขึ้นกับเซตย่อย (Subset) ของข้อมูลซึ่งจะปรากฏร่วมกันเสมอ

ทฤษฎีร์ฟเซต (Rough Set Theory) เป็นทฤษฎีทางคณิตศาสตร์ที่ใช้ในการจัดการเกี่ยวกับความคลุมเครือและความไม่แน่นอนของข้อมูล แนวคิดพื้นฐานอธิบายดังต่อไปนี้

2.1.3.1 ระบบสารสนเทศ (Information System)

ระบบสารสนเทศหรือตารางข้อมูลที่ประกอบด้วยวัตถุตามแนวระเบียน (Rows) และคุณลักษณะตามแนวคอลัมน์ (Columns) ซึ่งจะถูกใช้เป็นตัวแทนข้อมูลที่จะนำไปใช้โดยวิธีการร์ฟเซต (Rough Set) ระบบสารสนเทศ คือ คู่ของ (U, A) โดยที่

- U แทนเซตของเอกภพสัมพัทธ์ (Universe) ของเซตของวัตถุ

- A คือ เซตของคุณลักษณะ (Attributes) ของวัตถุที่ไม่มีค่าว่าง ทุกสมาชิกของ A เขียนแทนด้วย $a \in A$; $a : U \rightarrow V_a$

- V_a คือ ค่าของเซตที่เป็นไปได้ของคุณลักษณะ a

ตัวอย่างระบบสารสนเทศเกี่ยวกับข้อมูลสภาพอากาศ ประกอบด้วย 3 คุณลักษณะ (Attributes) รายละเอียดแสดงตามตารางที่ 2.4 อธิบายได้ดังนี้

1) คุณลักษณะ ฤดู (Season) ประกอบด้วยค่า ฤดูฝน (Rains) ฤดูร้อน (Summer) และ ฤดูหนาว (Winters)

2) คุณลักษณะ เมฆ (Cloud) ประกอบด้วยค่า มีดครึ้ม (Overcast) ปานกลาง (Medium) และ น้อย (Few)

3) คุณลักษณะ ฝน (Rain) ประกอบด้วยค่า มีฝน (Yes) และ ไม่มีฝน (No)

ตารางที่ 2.4

ตัวอย่างระบบสารสนเทศ

No.	Attributes		
	Season	Cloud	Rain
1.	Rains	Overcast	Yes
2.	Rains	Cloudless	No
3.	Summer	Medium	No
4.	Summer	Medium	Yes
5.	Winters	Few	No
6.	Rains	Few	Yes

2.1.3.2 ความสัมพันธ์ที่ไม่สามารถมองเห็นได้ (Indiscernibility Relation)

ความสัมพันธ์ที่ไม่สามารถมองเห็นได้ คือ ความสัมพันธ์ระหว่างวัตถุ 2 อย่าง หรือมากกว่า 2 อย่างขึ้นไป ซึ่งความสัมพันธ์นี้จะเหมือนกันในทุกทราจแซคชันที่เป็นสมาชิกของเซตย่อยของคุณลักษณะตัดสินใจ ความสัมพันธ์ที่ไม่สามารถมองเห็นได้ นิยามได้ตามสมการที่ 4

$$IND(B) = \{(x, y) \in U \times U : \text{for all } a \in B, a(x) = a(y)\} \text{ where } a \in A \text{ and } B \subseteq A \quad (4)$$

กำหนดให้

- B เป็นเซตย่อยของ A
- $IND(B)$ คือ ความสัมพันธ์ของเซตย่อย B บนเอกภพสัมพัทธ์ U เป็นความสัมพันธ์ที่ไม่สามารถมองเห็นได้

- $a(x)$ คือ ค่าคุณสมบัติ a ของวัตถุ x
- $a(y)$ คือ ค่าคุณสมบัติ a ของวัตถุ y
- ถ้า (x, y) ใด ๆ มีความสัมพันธ์กับ $IND(B)$ แล้ว x และ y จะไม่สามารถแยกความแตกต่างในกลุ่มคุณลักษณะ B (B-indiscernible) ได้ นั่นคือ x และ y มีค่าเท่ากันหมดสำหรับทุกคุณลักษณะใน B

ยกตัวอย่างความสัมพันธ์ที่ไม่สามารถมองเห็นได้ของเซตย่อยจากตารางที่ 2.4 โดยกำหนดให้คุณลักษณะเงื่อนไข คือ {Season}, {Cloud}, {Season, Cloud} แสดงได้ดังนี้

$$1. IND(\{Season\}) = \{\{1, 2, 6\}, \{3, 4\}, \{5\}\}$$

$$2. IND(\{Cloud\}) = \{\{1\}, \{2\}, \{3, 4\}, \{5, 6\}\}$$

$$3. IND(\{Season, Cloud\}) = \{\{1\}, \{2\}, \{3, 4\}, \{5\}, \{6\}\}$$

2.1.3.3 ค่าประมาณขอบเขตล่างและขอบเขตบน (Lower and Upper Approximations)

กำหนดให้ ตารางข้อมูลประกอบด้วยเอกภพและเซตของคุณลักษณะที่ไม่มีค่าว่าง เขียนแทนด้วย $T = (U, A)$ และให้ B เป็นเซตย่อยของคุณลักษณะ A และ x เป็นเซตย่อยของเอกภพ เขียนแทนด้วย $X \rightarrow U$ เราสามารถนิยาม B-lower และ B-upper ได้ตามสมการที่ 5 และ 6

$$\underline{BX} = \{X \mid [X]_B \subseteq X\} \quad (5)$$

$$\overline{BX} = \{X \mid [X]_B \cap X \neq \emptyset\} \quad (6)$$

กำหนดให้

- B-lower (\underline{B}) คือ การรวมกันของทุก ๆ กลุ่มข้อมูลที่แบ่งด้วยระดับเท่ากันโดยความสัมพันธ์ $IND(B)$ ที่มี x เป็นเซตย่อยของ X

- B-upper (\overline{B}) คือ การรวมกันของทุก ๆ กลุ่มข้อมูลที่แบ่งด้วยระดับเท่ากันโดยความสัมพันธ์ $IND(B)$ ที่อินเตอร์เซกกับ X แล้วไม่เท่ากับเซตว่าง

- Boundary Region คือ ขอบเขตภูมิภาค ซึ่งจะอยู่ในขอบเขตบน (Upper) แต่ไม่อยู่ในขอบเขตล่าง (Lower)

จากที่กำหนดให้ B เป็นเซตย่อยของคุณลักษณะ A และ x เป็นเซตย่อยของเอกภพนั้น สามารถหาค่าขอบเขตภูมิภาค B ของ x (B-boundary Region of x) ได้ตามสมการที่ 7

$$BN_B(x) = \overline{BX} - \underline{BX} \quad (7)$$

และหาขอบเขตนอกภูมิภาค B ของ x (B-outside region of x) ได้ตามสมการที่ 8

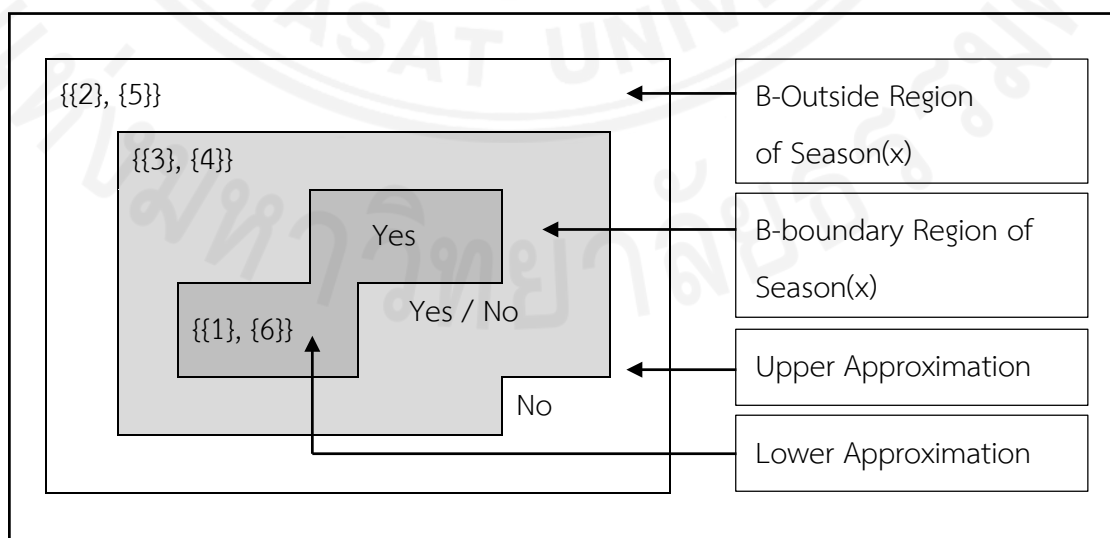
$$BO_B(x) = U - \overline{BX} \quad (8)$$

ถ้าขอบเขตภูมิภาค (Boundary Region) เป็นเซตว่าง แสดงว่า x เป็นเซตที่ชัดเจน (Crisp Set) บนเซตคุณลักษณะ B และถ้าขอบเขตภูมิภาค (Boundary Region) ไม่เป็นเซตว่าง แสดงว่า x เป็นรัฟเซตของคุณลักษณะ B

จากตารางที่ 2.4 สามารถหาความสัมพันธ์ที่ไม่สามารถมองเห็นได้ของคุณลักษณะ Season ดังที่แสดงในภาพที่ 2.4

กำหนดให้

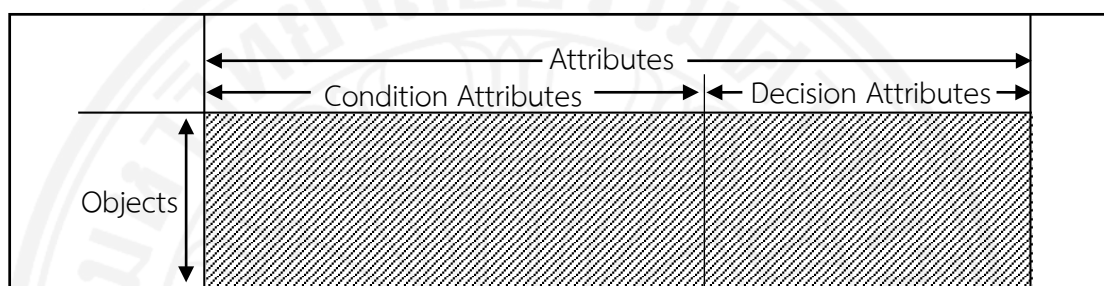
- $S = \{x \mid \text{Season}(x) = \text{yes}\}$
- $\underline{BS} = \{1, 6\}$
- $\overline{BS} = \{1, 3, 4, 6\}$
- $BN_B(S) = \{3, 4\}$
- $BO_B(S) = \{2, 5\}$



ภาพที่ 2.4 ค่าประมาณของเซตคุณลักษณะ Season

2.1.2.4 กฎการตัดสินใจ (Decision Rule)

กฎการตัดสินใจถูกสร้างขึ้นโดยการรวมชุดของคุณลักษณะ ข้อมูลหลายแถว (Records) คือ ตัวกำหนดกฎการตัดสินใจ การตัดสินใจจะทำได้ก็ต่อเมื่อคุณสมบัติที่ระบุตามเงื่อนไขนั้นเป็นจริง กฎการตัดสินใจจะแสดงความสัมพันธ์ออกมาในรูปแบบ ถ้า...แล้ว... (if...then...) รูปแบบข้อมูลที่ใช้ในทฤษฎีเซต (Rough Set Theory) ดังแสดงในภาพที่ 2.5



ภาพที่ 2.5 รูปแบบข้อมูลที่ใช้ในทฤษฎีเซต (Rough Set Theory)

จากตารางที่ 2.4 สามารถนำมาสร้างเป็นตารางตัดสินใจได้ โดยต้องจัดรูปแบบข้อมูลให้อยู่ในรูปแบบตารางตัดสินใจตามภาพที่ 2.5 กำหนดให้คุณลักษณะ Season และ Cloud เป็นคุณลักษณะเงื่อนไข (Condition Attributes) และกำหนดให้คุณลักษณะ Rain เป็นคุณลักษณะตัดสินใจ (Decision Attributes) ตัวอย่าง กฎการตัดสินใจ เช่น “If Season = Rains and Cloud = Overcast Then Rain = Yes” หมายความว่า “ถ้า เป็นฤดูฝนและมีเมฆมืดครึ้ม แล้ว จะทำให้ฝนตก”

ในปัจจุบัน ข้อมูลในฐานข้อมูลนั้นมีขนาดใหญ่และมีความซับซ้อนมากยิ่งขึ้น ทำให้ใช้เวลาในการประมวลผลนานและความสัมพันธ์หรือกฎการเชื่อมโยงที่ได้มีจำนวนมาก การตัดสินใจเลือกกฎการเชื่อมโยงจึงทำได้ด้วยความยากลำบาก หากจะต้องพิจารณาจากกฎทุกกฎที่ได้ งานวิจัยนี้นำเสนอเครื่องมือที่เรียกว่า ตัวชี้วัดความน่าสนใจ (Interesting Indicators) เพื่อเปรียบเทียบคุณภาพและคัดกรองกฎการเชื่อมโยงที่น่าสนใจจากกฎการเชื่อมโยงทั้งหมด

2.2 ตัวชี้วัดความน่าสนใจ (Interesting Indicators)

งานวิจัยนี้นำเสนอการประยุกต์ใช้เครื่องมือตัวชี้วัดความน่าสนใจของกฎการเชื่อมโยงจากบทความวิจัยเรื่อง Association Rules Selection Approach Based on Interesting Measures ของ ปกรณ์ ลีสุทธิพรชัย ค่าตัวชี้วัดความน่าสนใจใช้เป็นคะแนนในการคัดกรองเลือกกฎส่วนหนึ่งจากกฎทั้งหมด กฎการเชื่อมโยงที่มีค่าคะแนนมากถือเป็นกฎการเชื่อมโยงที่น่าสนใจ คำอธิบายสมการตัวชี้วัดความน่าสนใจ มีรายละเอียดดังนี้

กำหนดให้

$T = \{t_1, t_2, t_3, \dots, t_n\}$ คือ เซตของทรานแซคชัน (Transaction)

$X \rightarrow Y$ คือ รูปแบบกฎการเชื่อมโยง โดยที่ X และ Y คือ รูปแบบไอเทมเซต (Item Sets) โดยที่ $X \cap Y = \phi$

X คือ เงื่อนไขของกฎหรือพจน์ทางซ้ายของกฎ (Left Hand Side: LHS)

Y คือ ผลการตัดสินใจหรือพจน์ทางขวาของกฎ (Right Hand Side: RHS)

$\text{sup}(X)$ คือ ค่าสนับสนุน (Support) ของเซต X

$\text{sup}(Y)$ คือ ค่าสนับสนุน (Support) ของเซต Y

2.2.1 Confidence

Confidence คือ ค่าความน่าจะเป็นที่จะปรากฏเซต Y เมื่อมีการปรากฏของเซต X ค่า Confidence สามารถคำนวณได้ตามสมการที่ 9

$$\text{con}(X \rightarrow Y) = \frac{\text{sup}(X \rightarrow Y)}{\text{sup}(X)} = \frac{\text{su}(X \cup Y)}{\text{sup}(X)} \quad (9)$$

2.2.2 Coverage

Coverage เป็นตัวชี้วัดที่แสดงถึงความถี่ของกฎการเชื่อมโยง $X \rightarrow Y$ ที่ตรงตามคุณลักษณะเงื่อนไขต่อจำนวนรายการทั้งหมด ค่า Coverage สามารถคำนวณได้ตามสมการที่ 10

$$\text{cov}(X \rightarrow Y) = \text{sup}(X) \quad (10)$$

2.2.3 Leverage

Leverage เป็นตัวชี้วัดความแตกต่างของเซต X และเซต Y ที่ปรากฏร่วมกันในชุดข้อมูลเทียบกับการปรากฏของเซต X และเซต Y ที่เป็นอิสระต่อกัน ค่า Leverage สามารถคำนวณได้ตามสมการที่ 11

$$\begin{aligned} \text{lev}(X \rightarrow Y) &= \text{sup}(X \rightarrow Y) - \text{sup}(\text{sup}(X) \cdot \text{sup}(Y)) \\ &= \text{sup}(X \cup Y) - (\text{sup}(X) \cdot \text{sup}(Y)) \end{aligned} \quad (11)$$

2.2.4 Lift

Lift เป็นตัวชี้วัดสัดส่วนของเซต X และเซต Y ที่ปรากฏร่วมกันต่อการปรากฏของเซต X และเซต Y ที่เป็นอิสระต่อกัน ค่า Lift สามารถคำนวณได้ตามสมการที่ 12

$$\text{lift}(X \rightarrow Y) = \frac{\text{con}(X \rightarrow Y)}{\text{sup}(Y)} = \frac{\text{sup}(X \cup Y)}{\text{sup}(X) \cdot \text{sup}(Y)} \quad (12)$$

2.2.5 Generalization

Generalization เป็นตัวชี้วัดที่แสดงถึงจำนวนคุณลักษณะเงื่อนไขของกฎการเชื่อมโยง ถ้าค่า Generalization มีค่ามากแสดงว่าจำนวนคุณลักษณะเงื่อนไขในกฎการเชื่อมโยงมีจำนวนน้อย ค่า Generalization สามารถคำนวณได้ตามสมการที่ 13

$$\text{gen}(X \rightarrow Y) = 1 - \frac{|X|}{N} \quad (13)$$

กำหนดให้

- N คือ จำนวนทั้งหมดของคุณลักษณะเงื่อนไข
- |X| คือ จำนวนคุณลักษณะเงื่อนไข X ที่ปรากฏในกฎการเชื่อมโยง

ตัวอย่างการคำนวณคะแนนตัวชี้วัดความน่าสนใจ (Interesting Indicators) โดยใช้ข้อมูลจากตารางที่ 2.4 กำหนดให้ คุณลักษณะเงื่อนไข คือ Season และ Cloud คุณลักษณะตัดสินใจ คือ Rain นำข้อมูลทั้งหมดมาสร้างเป็นกฎการเชื่อมโยงได้ตามตารางที่ 2.5 จากกฎข้อที่ 1 คือ {Season=Rains and Cloud=Overcast \rightarrow Rain=Yes} เซต X มีค่าเป็น {Season=Rains and Cloud=Overcast} และเซต Y มีค่าเป็น {Rain=Yes} สามารถหาคะแนนตัวชี้วัด ได้ดังนี้

ค่า Sup(X) มีค่าเท่ากับ 0.1667 มีเซต X ที่ทำให้เกิดเซต Y อยู่จำนวน 1 รายการ จากทั้งหมด 6 รายการ

ค่า Sup(Y) มีค่าเท่ากับ 0.5 มีเซต Y จำนวน 3 รายการ จากทั้งหมด 6 รายการ

ค่า Sup(X \rightarrow Y) มีค่าเท่ากับ 0.1667 หมายความว่า ข้อมูลในตารางที่ 2.5 มีจำนวนเซต X และเซต Y อยู่ 1 รายการ จากทั้งหมด 6 รายการ

ค่า Confidence คำนวณตามสมการที่ 9 จะได้

$$\text{Con}(X \rightarrow Y) = 0.1667/0.1667 = 1$$

ค่า Coverage คำนวณตามสมการที่ 10 จะได้

$$\text{Cov}(X \rightarrow Y) = 0.1667$$

ค่า Leverage คำนวณตามสมการที่ 11 จะได้

$$\text{Lev}(X \rightarrow Y) = 0.1667 - (0.1667 * 0.5) = 0.0834$$

ค่า Lift คำนวณตามสมการที่ 12 จะได้

$$\text{Lift}(X \rightarrow Y) = 0.1667/(0.1667 * 0.5) = 2$$

ค่า Generalization คำนวณตามสมการที่ 13 จะได้

$$\text{Gen}(X \rightarrow Y) = 1 - (2/2) = 0$$

ตารางที่ 2.5

แสดงคะแนนจากตัวชี้วัดความน่าสนใจ (Interesting Indicators)

No.	Attributes		Interesting Indicators					
	Condition	Decision	Con	Cov	Lev	Lift	Normalized Lift	Gen
1.	Season=Rains and Cloud=Overcast	Rain=Yes	1.0000	0.1667	0.0834	2.0000	1.0000	0.0000
2.	Season=Rains and Cloud=Cloudless	Rain=No	1.0000	0.1667	0.0834	2.0000	1.0000	0.0000
3.	Season=Summer and Cloud=Medium	Rain=No	0.5000	0.3334	0.0000	1.0000	0.5000	0.0000
4.	Season=Summer and Cloud=Medium	Rain=Yes	0.5000	0.3334	0.0000	1.0000	0.5000	0.0000
5.	Season=Winters and Cloud=Few	Rain=No	1.0000	0.1667	0.0834	2.0000	1.0000	0.0000
6.	Season=Rains and Cloud=Few	Rain=Yes	1.0000	0.1667	0.0834	2.0000	1.0000	0.0000

2.3 งานวิจัยที่เกี่ยวข้อง

ในปี ค.ศ. 2010 Rui Tian และ Zhaosheng Yang ได้เสนอวิธีการวิเคราะห์สาเหตุการเกิดอุบัติเหตุการจราจรจากการทำเหมืองข้อมูล โดยได้ประยุกต์ใช้ทฤษฎีเซต (Rough Set Theory) หาความสัมพันธ์ระหว่างสาเหตุการเกิดอุบัติเหตุกับประเภทอุบัติเหตุ ชุดข้อมูลได้จากการสอบถามผู้ที่เกี่ยวข้องในเหตุการณ์ การวิเคราะห์สัญลักษณ์ที่เกิดเหตุ และอื่น ๆ ที่เกี่ยวข้องกับที่เกิดอุบัติเหตุ เช่น เวลา สภาพอากาศ สิ่งแวดล้อม ลักษณะพื้นผิว ยานพาหนะ และการบาดเจ็บ

ในปี ค.ศ. 2011 Sibar Kaan Manga ได้ศึกษาวิธีการทำนายโอกาสในการเกิดวิกฤตสกุลเงิน โดยประยุกต์ใช้ทฤษฎีเซต (Rough Set Theory) เพื่อหาความสัมพันธ์ที่เกิดขึ้นในเศรษฐกิจมหภาค ชุดข้อมูลเก็บรวบรวมจากเว็บไซต์ของธนาคารกลาง ประเทศตุรกี คุณลักษณะเงื่อนไข ได้แก่ ราคาทองคำ ราคาหุ้น ค่าสกุลเงิน และอัตราดอกเบี้ย

ในปี ค.ศ. 2012 Ogwueleka ได้เสนอการรวบรวมข้อมูลอุบัติเหตุและการจัดกลุ่มอุบัติเหตุการจราจร โดยใช้กระบวนการทำเหมืองข้อมูลเพื่อหาความสัมพันธ์ระหว่างสาเหตุการเกิดอุบัติเหตุกับระดับความรุนแรงของการบาดเจ็บ ชุดข้อมูลรวบรวมจากคณะกรรมการความปลอดภัยทางถนนและสำนักงานตำรวจแห่งชาติ คุณลักษณะของสาเหตุการเกิดอุบัติเหตุ แบ่งเป็น ประเภทยานพาหนะ อายุผู้ขับขี่ ระดับการศึกษา อายุยานพาหนะ พื้นที่เกิดอุบัติเหตุ สภาพผิวจราจร ความสว่าง และสภาพอากาศ คุณลักษณะการตัดสินใจ แบ่งระดับความรุนแรงของการบาดเจ็บเป็น 4 ระดับ คือ 1) ไม่ได้รับบาดเจ็บ 2) ได้รับบาดเจ็บเล็กน้อย 3) ได้รับบาดเจ็บปานกลาง และ 4) ได้รับบาดเจ็บสาหัส

ในปี ค.ศ. 2012 Xiaofeng Zheng และ Jianmin Xu ได้ศึกษาการประยุกต์ใช้วิธีเซต (Rough Set) กับการทำเหมืองข้อมูลเพื่อวิเคราะห์หาความสัมพันธ์ที่เกิดขึ้นระหว่างคุณลักษณะการขนส่งกับคะแนนการประเมินคุณภาพ ชุดข้อมูลเก็บรวบรวมจากระบบจัดการสารสนเทศ กรมทางหลวงมณฑลกว่างตุง คุณลักษณะการขนส่ง ได้แก่ ความปลอดภัยในการขนส่ง ลักษณะธุรกิจขนส่ง คุณภาพการให้บริการ และความรับผิดชอบต่อสังคม

ในปี ค.ศ. 2013 Muhammad Durairaj ได้ประยุกต์ใช้ทฤษฎีเซต (Rough Set Theory) วิเคราะห์อิทธิพลที่มีผลต่อการปฏิสนธิในหลอดทดลองและการรักษาผู้มีบุตรยาก เพื่อให้สูตินารีแพทย์มีความแม่นยำในการวิเคราะห์และให้คำแนะนำสำหรับผู้ที่มีบุตรยาก ชุดข้อมูลได้จากการเก็บรวบรวมข้อมูลผลการรักษาผู้มีบุตรยากจากโรงพยาบาลและสถานที่ต่าง ๆ ที่ให้บริการรักษาผู้มีบุตรยาก มีข้อมูลทั้งหมด 114 ตัวอย่าง คุณลักษณะเงื่อนไขมีทั้งหมด 23 ตัว เช่น อายุของผู้หญิง การตกไข่ ภาวะท่อหน้าไข่ ฮอโมน ปากมดลูก ปริมาณน้ำเชื้อ ความเข้มข้นของอสุจิ ฯลฯ

ในปี ค.ศ. 2014 Caner Erden และ Fatih Tüysüz ได้ประยุกต์ใช้ทฤษฎีเซต (Rough Set Theory) เพื่อหาความสัมพันธ์ของสาเหตุการเสียชีวิตจากการเกิดอุบัติเหตุ โดยชุดข้อมูลได้จากการสอบถามบุคคลที่เกี่ยวข้องในเหตุการณ์ขณะเกิดอุบัติเหตุรวมถึงข้อมูลอื่น ๆ ที่เกี่ยวข้องกับการเกิดอุบัติเหตุ ได้แก่ รหัสบัตรประชาชน (ID) อายุ (Age) เพศ (Gender) ประเภทบุคคล (Person Type) เชื้อชาติ (Race) รัฐ (State) ผลตรวจแอลกอฮอล์ (Alcohol Results) สภาพอากาศ (Atmospheric Condition) วันที่เกิดเหตุ (Crash Date) สารเสพติด (Drug Involvement) เหตุการณ์การเกิดอุบัติเหตุ (First Harmful Event) จำนวนผู้เสียชีวิต (Fatalities in Crash) การบาดเจ็บรุนแรง (Injury Severity) สภาพถนน (Roadway) ดังแสดงในตารางที่ 2.6

ตารางที่ 2.6

ตัวอย่างคุณลักษณะของข้อมูลที่ได้จากการเกิดอุบัติเหตุ

Age	Alcohol Results	Atm. Con.	Drug Involvement	First Harmful Event	Gender	Injury Severity	Race	Roadway	Fatalities
20	0.00	Snow	Not Reported	Motor Vehicle In Transport	Female	Fatal Injury (K)	White	Rural Principal Arterial Other	1
41		Clear	Not Reported	Motor Vehicle In Transport	Female	No Injury (0)		Rural Minor Arterial Other	1
15		Clear	No	Motor Vehicle In Transport	Female	Incapacitating Injury (A)		Rural Principal Arterial Other	2
66	0.00	Cloudy	Reported	Motor Vehicle In Transport	Male	No Injury (0)		Rural Principal Arterial Other	1
44		Clear	No	Motor Vehicle In Transport	Male	Incapacitating Injury (A)	Black	Urban Minor Arterial	1
53	0.04	Clear	Unknown	Pedestrian	Male	Fatal Injury (K)		Urban Principal Arterial Interstate	1
79		Unknown	No	Pedal cyclist	Male	Fatal Injury (K)	White	Urban Other Principal Arterial	1
43		Clear	Unknown	Motor Vehicle In Transport	Male	No Injury (0)		Rural Minor Collector	1
84		Clear	Unknown	Tree (Standing Only)	Male	Fatal Injury (K)	White	Urban Other Principal Arterial	1

จากนั้นได้ทำการแบ่งช่วงข้อมูล (Discretization) โดยกำหนดค่าแทนคุณลักษณะเงื่อนไขด้วยค่า a1, a2, a3, ..., a9 คุณลักษณะตัดสินใจคือ d และแทนค่าข้อมูลด้วยตัวเลข 1, 2, 3, 4 และ 5 ดังแสดงในตารางที่ 2.7

ตารางที่ 2.7

แสดงช่วงข้อมูล

Attributes		Values				
a1	Gender	Male	Female			
	Domain	1	2			
a2	Age	Young (0-25)	Average (25-50)	Old (50-50+)		
	Domain	1	2	3		
a3	Race	White	Black	All other		
	Domain	1	2	3		
a4	Alcohol	0-0,1	0,1-0,2	0,2-0,2+		
	Domain	1	2	3		
a5	Drug Involvement	Yes	No			
	Domain	1	2			
a6	Atmospheric Condition	Clear	Cloudy	Rain	Snow	Other
	Domain	1	2	3	4	5
a7	First Harmful Event	Vehicle	Pedestrian	Other		
	Domain	1	2	3		
a8	Injury Severity	Fatal	No Injury	Possible	Unknown	
	Domain	1	2	3	?	
a9	Roadway	Urban	Rural			
	Domain	1	2			
Decision Attribute	d	Fatalities in crash	1	2-2+		
		Domain	1	2		

หลังจากนั้นได้ทำการแปลงชุดข้อมูลตามช่วงข้อมูลโดยใช้หลักเกณฑ์ตามตารางที่ 2.7 เมื่อแปลงเสร็จแล้วได้ข้อมูลดังแสดงในตารางที่ 2.8 จากนั้นจึงนำข้อมูลที่ได้ไปทำการทดลอง

ตารางที่ 2.8

ตัวอย่างข้อมูลหลังจากทำการแปลงค่าตามช่วงข้อมูล

No.	a1	a2	a3	a4	a5	a6	a7	a8	a9	d
X1	2	1	1	1	2	4	1	1	1	1
X2	2	2	2	1	2	1	1	2	3	1
X3	2	1	1	1	2	1	1	3	1	2
...

จากตารางที่ 2.8 เมื่อนำมาทดลองโดยใช้ทฤษฎีรฟเซต (Rough Set Theory) สามารถวิเคราะห์และแสดงกฎการเชื่อมโยงที่สำคัญได้ดังต่อไปนี้

กฎข้อที่ 1 If $a_1=1$ and $a_5=1$ then $d=1$ แสดงว่า เมื่อผู้ชายเป็นผู้ขับรถและเสพยาเสพติดก่อนออกเดินทางเป็นไปได้ที่การเกิดอุบัติเหตุจะมีผู้เสียชีวิต จำนวน 1 คน

กฎข้อที่ 2 If $a_7=4$ then $d=1$ or 2 แสดงว่า เมื่อสภาพอากาศมีหิมะตกเป็นไปได้ที่การเกิดอุบัติเหตุจะมีผู้เสียชีวิตจำนวน 1 คน หรือมากกว่า 1 คน

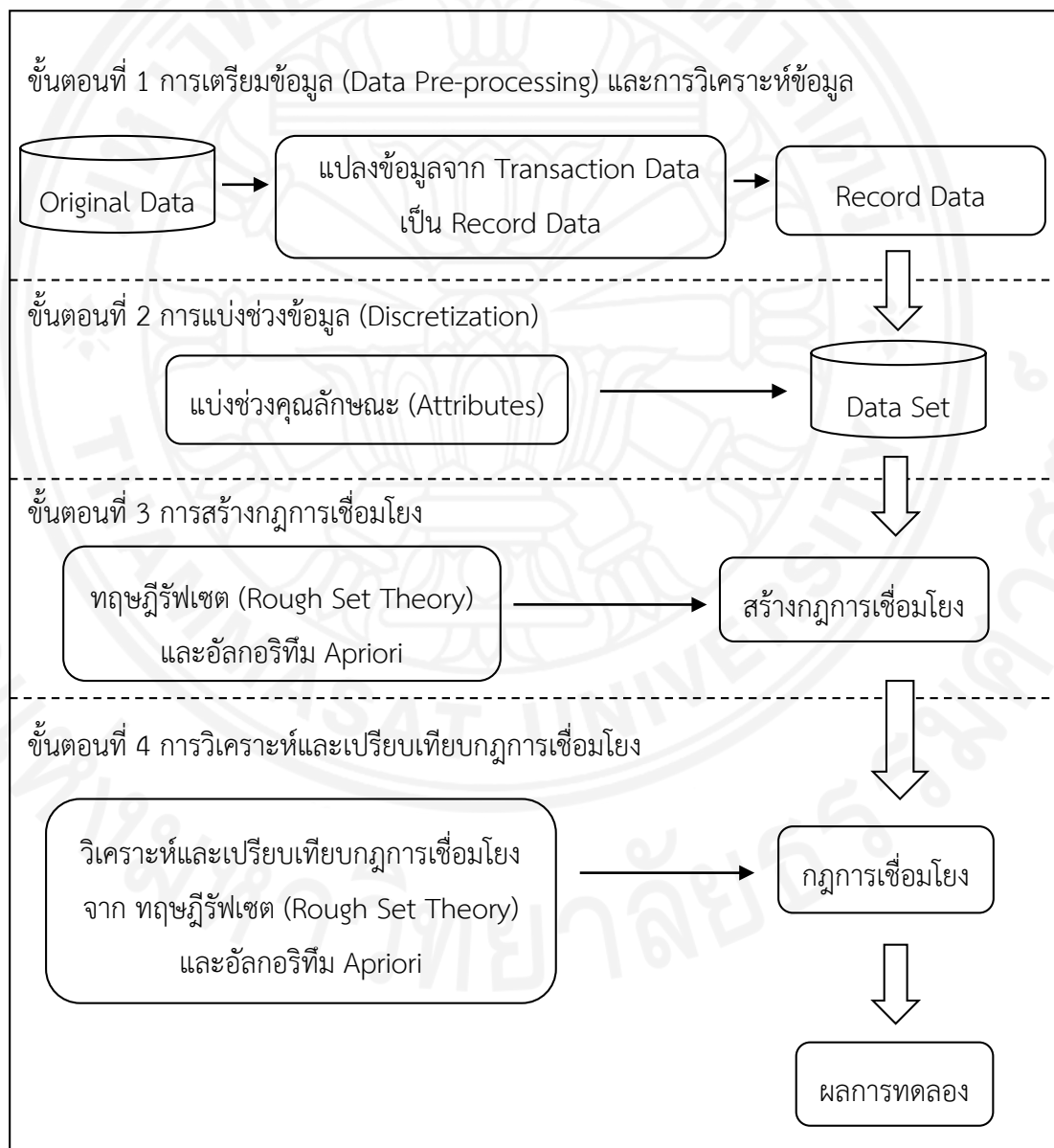
กฎข้อที่ 3 If $a_2=1$ and $a_5=1$ and $a_7=3$ then $d=1$ or 2 แสดงว่า เมื่อผู้ขับรถมีอายุระหว่าง 25 ถึง 50 ปี และเสพยาเสพติดก่อนออกเดินทางและเกิดอุบัติเหตุขึ้นครั้งแรกในการใช้รถยนต์เป็นไปได้ที่การเกิดอุบัติเหตุจะมีผู้เสียชีวิตจำนวน 1 คน หรือมากกว่า 1 คน

จากงานวิจัยที่ได้นำเสนอในข้างต้น แสดงให้เห็นว่า ทฤษฎีรฟเซต (Rough Set Theory) เป็นเทคนิคการทำเหมืองข้อมูลอีกวิธีหนึ่งที่ถูกนำมาประยุกต์ใช้วิเคราะห์ข้อมูลที่หลากหลาย เช่น วิเคราะห์การขนส่ง วิเคราะห์ทางการแพทย์ วิเคราะห์เศรษฐกิจ วิเคราะห์การตลาด และวิเคราะห์อุบัติเหตุการจราจรบนถนน เป็นต้น

บทที่ 3

วิธีการดำเนินงานวิจัย

งานวิจัยนี้ได้ประยุกต์ใช้ทฤษฎีเซตหยาบ (Rough Set Theory) และอัลกอริทึม Apriori เพื่อเปรียบเทียบและวิเคราะห์หาความสัมพันธ์ของการจราจร โดยได้กำหนดขั้นตอนการทดลองและดำเนินงานวิจัย ดังแสดงในภาพที่ 3.1



ภาพที่ 3.1 ภาพรวมขั้นตอนการทดลอง

3.1 การเตรียมข้อมูล (Data Pre-processing) และการวิเคราะห์ข้อมูล

กรมทางหลวง กองทางหลวงพิเศษระหว่างเมือง						
PCS COMPUTER SYSTEM						
ปริมาณจราจรรายชั่วโมง แยกตามประเภทรถ						
วันที่ปฏิทิน 01/08/2557	ด่านผ่านทาง 701-ด่านลาดกระบัง		ประเภทช่องทาง ทั้งหมด			
ช่วงชั่วโมง 1 - 24	ด่านต้นทาง ทั้งหมด		กลุ่มช่องทาง ทั้งหมด			
ประเภทข้อมูล ทั้งหมด			หมายเลขช่องทาง ทั้งหมด			
ด้าน/ชั่วโมง/ด่านต้นทาง	ประเภท 0	ประเภท 1	ประเภท 2	ประเภท 3	รวมทุกประเภท	
ขาเข้า (Entry)						
701 - ด่านลาดกระบัง						
00:00-01:00	0	319	24	139	482	
ผู้จ่ายบัตร	0	318	24	139	481	
ฝาด่าน	0	1	0	0	1	
01:00-02:00	0	239	27	86	352	
...	
23:00-00:00	1	726	43	134	904	
ผู้จ่ายบัตร	1	726	43	134	904	
รวมทุกชั่วโมง	100	40,579	3,023	4,470	48,172	
รวมขาเข้า	100	40,579	3,023	4,470	48,172	
ขาออก (Exit)						
701 - ด่านลาดกระบัง						
00:00-01:00	0	379	46	148	573	
701 - ด่านลาดกระบัง	0	0	1	0	1	
รถยกเว้น	0	0	1	0	1	
702 - ด้านบางบ่อ	0	15	3	23	41	
เงินสด	0	15	3	23	41	
703 - ด้านบางปะกง	0	43	6	8	57	
...	
23:00-00:00	0	952	54	152	1,158	
701 - ด่านลาดกระบัง	0	1	0	0	1	
เงินสด	0	1	0	0	1	
702 - ด้านบางบ่อ	0	22	7	35	64	
เงินสด	0	22	7	35	64	
703 - ด้านบางปะกง	0	72	5	8	85	
เงินสด	0	72	5	8	85	
704 - ด้านพนัสนิคม	0	140	13	17	170	
เงินสด	0	140	11	17	168	
รถยกเว้น	0	0	2	0	2	
705 - ด้านพานทอง	0	717	29	92	838	
เงินสด	0	717	22	92	831	
รถยกเว้น	0	0	7	0	7	
รวมทุกชั่วโมง	0	40,898	3,008	4,655	48,561	
รวมขาออก	0	40,898	3,008	4,655	48,561	
รวมทั้งหมด	100	81,477	6,031	9,125	96,733	

ภาพที่ 3.2 ตัวอย่างรายงานข้อมูลการจราจร

3.1.1 แหล่งที่มาของข้อมูล

ในงานวิจัยนี้ใช้ข้อมูลการจราจรบนทางหลวงพิเศษระหว่างเมืองหมายเลข 7 ได้รับความอนุเคราะห์ข้อมูลจากกองทางหลวงพิเศษระหว่างเมือง กรมทางหลวง ข้อมูลอยู่ในช่วงตั้งแต่วันที่ 1 มกราคม 2557 ถึงวันที่ 30 เมษายน 2558 มีข้อมูลทั้งหมด 242,066 แถว และมีจำนวนรถทั้งหมด 60,766,872 คัน ตัวอย่างรายงานข้อมูลการจราจรแสดงในภาพที่ 3.2

3.1.2 การวิเคราะห์ข้อมูล

ทางหลวงพิเศษระหว่างเมืองหมายเลข 7 ใช้ระบบจัดเก็บค่าธรรมเนียมผ่านทางระบบปิด ด้านผ่านทางทุกด้านสามารถเชื่อมโยงข้อมูลหากันได้ ขั้นตอนการใช้บริการทางหลวงพิเศษระหว่างเมืองหมายเลข 7 นั้น กำหนดให้ผู้ใช้รับบัตรผ่านทางที่ด่านขาเข้าและชำระค่าธรรมเนียมผ่านทางที่ด่านขาออก ปัจจุบันมีด่านผ่านทาง ทั้งหมด 5 ด้าน ได้แก่ ด่านลาดกระบัง ด่านบางบ่อ ด่านบางปะกง ด่านพนัสนิคม และด่านพานทอง จากภาพที่ 3.2 ข้อมูลการจราจรแบ่งเป็น 2 ส่วน ได้แก่

1) ส่วนหัวรายงาน คือ ส่วนที่แสดงรายละเอียดแหล่งที่มาของรายงานการจราจร ประกอบด้วย สัญลักษณ์ส่วนงาน ชื่อส่วนงาน ชื่อรายงาน และอื่น ๆ ที่สำคัญ ได้แก่

- 1) วันที่ปฏิทิน คือ วันที่จัดเก็บข้อมูล
- 2) ช่วงชั่วโมง คือ ช่วงระยะเวลาในการรวมสรุปจำนวนรถยนต์
- 3) ด้านผ่านทาง คือ ด้านที่รถวิ่งเข้า-ออก ด้านนั้น ๆ

2) ส่วนข้อมูล คือ ส่วนที่แสดงจำนวนรถยนต์ โดยแบ่งคอลัมน์ (Columns) ตามประเภทรถยนต์ ดังนี้ 1) ประเภท 0 คือ รถยนต์ประเภทอื่น ๆ ที่ไม่จัดอยู่ในประเภท 1, 2 และประเภท 3 2) ประเภท 1 คือ รถยนต์ 4 ล้อ 3) ประเภท 2 คือ รถบรรทุกตั้งแต่ 6-10 ล้อ 4) ประเภท 3 คือ รถบรรทุกมากกว่า 10 ล้อ และ 5) จำนวนรวมทั้งหมดของรถยนต์ ในส่วนของข้อมูลยังแยกสรุปข้อมูลออกเป็น 2 ส่วน คือ ข้อมูลขาเข้า (Entry) และข้อมูลขาออก (Exit)

2.1) ข้อมูลขาเข้า (Entry) คือ จำนวนรถยนต์ที่วิ่งเข้าด้านนั้น ๆ โดยแยกข้อมูลสรุปเป็นช่วงเวลาและประเภทการรับบัตรผ่านด่าน จากภาพที่ 3.2 สามารถอธิบายได้ดังนี้

ที่ด่านขาเข้าลาดกระบัง ในช่วงเวลา 00:00 - 01:00 มีรถยนต์รับบัตรจากตู้จ่ายบัตร แยกประเภทได้ดังนี้

- 1) รถยนต์ประเภท 0 จำนวน 0 คัน
- 2) รถยนต์ประเภท 1 จำนวน 318 คัน
- 3) รถยนต์ประเภท 2 จำนวน 24 คัน
- 4) รถยนต์ประเภท 3 จำนวน 139 คัน
- 4) รวมทุกประเภท มีจำนวน 481 คัน

และมีรถยนต์ฝาด่าน แบ่งออกเป็น

- 1) รถยนต์ประเภท 0 จำนวน 0 คัน
- 2) รถยนต์ประเภท 1 จำนวน 1 คัน
- 3) รถยนต์ประเภท 2 จำนวน 0 คัน
- 4) รถยนต์ประเภท 3 จำนวน 0 คัน
- 5) รวมทุกประเภท มีจำนวน 1 คัน

2.2) ข้อมูลขาออก (Exit) คือ จำนวนรถยนต์ที่วิ่งเข้า - ออก ด้านนั้น ๆ โดยแยกข้อมูลสรุปเป็นช่วงเวลา ด้านขาเข้า และประเภทการชำระค่าธรรมเนียมผ่านทาง จากภาพที่ 3.2 สามารถอธิบายได้ดังนี้ ในช่วงเวลา 23:00 - 00:00 มีรถยนต์วิ่งมาจากด้านพานทองและออกที่ด้านลาดกระบัง แบ่งเป็นชำระค่าธรรมเนียมผ่านทางด้วยเงินสด แยกประเภทได้ดังนี้ 1) รถยนต์ประเภท 0 จำนวน 0 คัน 2) รถยนต์ประเภท 1 จำนวน 717 คัน 3) รถยนต์ประเภท 2 จำนวน 22 คัน 4) รถยนต์ประเภท 3 จำนวน 92 คัน และ 5) รวมทุกประเภท มีจำนวน 831 คัน และมีรถยนต์ได้รับการยกเว้นค่าธรรมเนียมผ่านทาง แยกประเภทได้ดังนี้ 1) รถยนต์ประเภท 0 จำนวน 0 คัน 2) รถยนต์ประเภท 1 จำนวน 0 คัน 3) รถยนต์ประเภท 2 จำนวน 7 คัน 4) รถยนต์ประเภท 3 จำนวน 0 คัน และ 5) รวมทุกประเภท มีจำนวน 7 คัน

ประเภทบัตรขาเข้า-ออก ด้านชำระค่าธรรมเนียมผ่านทาง สามารถอธิบายได้ตามตารางที่ 3.1

ตารางที่ 3.1

แสดงรายละเอียดและคำอธิบายของบัตรผ่านทาง

การจราจร	ประเภทบัตร	คำอธิบาย
ขาเข้า (Entry)	1) ผู้จ่ายบัตร	รับบัตรที่ด้านขาเข้า พร้อมบันทึกข้อมูลวัน เวลา ด้านขาเข้า
	2) ฝาด่าน	1) เกิดขึ้นในกรณีที่ไม่ได้รับบัตรที่ด้านขาเข้า 2) เกิดขึ้นจากความผิดพลาดของระบบที่ไม่สามารถบันทึกข้อมูล วัน เวลา และด้านขาเข้า ลงในบัตรได้
ขาออก (Exit)	1) EMR	รถที่คอยให้บริการในกรณีเกิดอุบัติเหตุ
	2) คู่มือ	ส่วนลดในการชำระค่าธรรมเนียมผ่านทาง
	3) ฝาด่าน	1) เกิดขึ้นในกรณีที่ไม่ได้รับบัตรที่ด้านขาเข้า 2) เกิดขึ้นจากความผิดพลาดของระบบที่ไม่สามารถบันทึกข้อมูล วัน เวลา และด้านขาออก ลงในบัตรได้
	4) ยกเลิกการเก็บเงิน	ไม่มีการเรียกเก็บค่าธรรมเนียมผ่านทาง
	5) รถขบวน	รถเฉพาะกิจในราชการ
	6) รถบริการ	รถให้บริการซ่อมบำรุงถนน
	7) รถยกเว้น	ไม่มีการเรียกเก็บค่าธรรมเนียมผ่านทาง
	8) เงินสด	ชำระค่าธรรมเนียมผ่านทางปกติ

3.1.3 การแปลงข้อมูล

จากการวิเคราะห์ข้อมูลภาพที่ 3.2 งานวิจัยนี้ได้เลือกใช้ข้อมูลส่วนขาออก (Exit) เนื่องจากสามารถระบุได้ว่ามีรถมาจากด้านขาเข้าใดบ้างและมีจำนวนเท่าใด คุณลักษณะ (Attributes) ของข้อมูลการจราจรประกอบด้วย วัน วันหยุด ช่วงเวลา ด้านขาเข้า ด้านขาออก จำนวนอุบัติเหตุ จำนวนรถยนต์ส่วนบุคคล จำนวนรถบรรทุก และจำนวนรถยนต์ทั้งหมด โครงสร้างข้อมูลส่วนขาออก จากภาพที่ 3.2 แสดงตามตารางที่ 3.2

ตารางที่ 3.2

โครงสร้างข้อมูลส่วนขาออก

ด้าน/ชั่วโมง/ด้านต้นทาง	ประเภท 1	ประเภท 2	ประเภท 3	รวมทุกประเภท
ด้านขาออก - ลาดกระบัง	40,898	3,008	4,655	48,561
ช่วงเวลา 23:00 - 00:00	952	54	152	1,118
ด้านขาเข้า - พานทอง	717	29	92	838
EMR	0	0	0	0
คูปอง	0	0	0	0
ยกเลิกการเก็บเงิน	0	0	0	0
รถบริการ	0	0	0	0
รถยกเว้น	0	7	0	7
เงินสด	717	22	92	831

เนื่องจากรูปแบบข้อมูลที่ได้มายังไม่อยู่ในรูปแบบที่สามารถนำไปใช้ในการทดลอง และวิเคราะห์การจราจรได้ จึงต้องทำการแปลงข้อมูลเพื่อให้ได้คุณลักษณะ (Attributes) ตามที่ต้องการ โดยดำเนินการดังนี้

1) รวมข้อมูลที่ซ้ำซ้อนกัน เนื่องจากข้อมูลมีลักษณะคล้ายกันและสามารถจัดให้อยู่ในคุณลักษณะ (Attributes) เดียวกันได้ คือ จำนวนรถยนต์ประเภท 2 (รถบรรทุก 6 - 10 ล้อ) และจำนวนรถยนต์ประเภท 3 (รถบรรทุกมากกว่า 10 ล้อขึ้นไป) ซึ่งจำนวนรถยนต์ทั้ง 2 ประเภท นั้นสามารถจัดรวมให้อยู่ในกลุ่มคุณลักษณะ (Attribute) จำนวนรถบรรทุกได้

2) แยกประเภทการชำระค่าธรรมเนียมผ่านทาง งานวิจัยนี้ให้ความสำคัญกับจำนวนอุบัติเหตุหรือเหตุฉุกเฉินในระหว่างเดินทาง จึงแยกประเภทการชำระค่าธรรมเนียมผ่านทาง ออกเป็น 2 กลุ่ม คือ EMR และทั่วไป กลุ่มทั่วไปจะประกอบด้วย การชำระค่าธรรมเนียมผ่านทาง ประเภท คูปอง ฝาด่าน ยกเลิกการเก็บเงิน รถขบวน รถบริการ รถยกเว้น และเงินสด

3) แปลงข้อมูล หลังจากทำการรวมข้อมูลและแยกประเภทการชำระค่าธรรมเนียมผ่านทางแล้วจะได้คุณลักษณะ (Attributes) แสดงในตารางที่ 3.3 จากนั้นข้อมูลตามภาพที่ 3.2 จะถูกแปลงข้อมูลให้อยู่ในรูปแบบเรคคอร์ด ตามตารางที่ 3.4

ตารางที่ 3.3

คุณลักษณะ (Attributes) และคำอธิบาย

คุณลักษณะ	คำอธิบาย
วัน (Date)	จันทร์ อังคาร พุธ พฤหัสบดี ศุกร์ เสาร์ และอาทิตย์
วันหยุด (Holiday)	วันหยุดนักขัตฤกษ์ประจำปี
ช่วงเวลา (Time Period)	ช่วงเวลาในการสรุปจำนวนรถยนต์
ด่านขาเข้า (Entry Toll)	ด่านที่รับบัตรจากผู้จ่ายบัตร
ด่านขาออก (Exit Toll)	ด่านที่คืนบัตรและชำระค่าธรรมเนียมผ่านทาง
จำนวนอุบัติเหตุหรือเหตุฉุกเฉิน (No. of EMR Cars)	จำนวนอุบัติเหตุทั้งหมด
จำนวนรถยนต์ส่วนบุคคล (No. of Private Cars)	จำนวนรถยนต์ประเภท 1
จำนวนรถบรรทุก (No. of Trucks)	จำนวนรถยนต์ประเภท 2 และประเภท 3
จำนวนรถยนต์ทั้งหมด (No. of Vehicles)	จำนวนรถยนต์ทั้งหมด

ตารางที่ 3.4

แสดงข้อมูลจราจรในรูปแบบเรคคอร์ด (Record-data Format)

Condition Attributes							Decision Attributes
Date	Time Period	Entry Toll	Exit Toll	No. of EMR Cars	No. of Private Cars	No. of Trucks	No. of Vehicles
1/8/2557	00:00-01:00	บางปะอ	ลาดกระบัง	0	15	26	41
1/8/2557	00:00-01:00	บางปะกง	ลาดกระบัง	0	43	14	57
1/8/2557	00:00-01:00	พนัสนิคม	ลาดกระบัง	0	53	35	88
...
1/8/2557	23:00-00:00	บางปะกง	ลาดกระบัง	0	72	13	85
1/8/2557	23:00-00:00	พนัสนิคม	ลาดกระบัง	0	140	30	170
1/8/2557	23:00-00:00	พานทอง	ลาดกระบัง	0	717	121	838

3.2 การแบ่งช่วงข้อมูล (Discretization)

3.2.1 การแบ่งช่วงคุณลักษณะ

ค่าของข้อมูลแต่ละคุณลักษณะมีค่าที่เป็นไปได้มากเกินไป ดังนั้น งานวิจัยนี้จึงใช้เทคนิคมาตรฐานของการทำเหมืองข้อมูลคือการแบ่งช่วงของข้อมูล (Discretization) เพื่อแปลงข้อมูลต่อเนื่องให้เป็นข้อมูลที่ไม่ต่อเนื่อง โดยแต่ละคุณลักษณะแบ่งได้ตามตารางที่ 3.5 ดังนี้

ตารางที่ 3.5

ช่วงกำหนดค่าของคุณลักษณะ

Attributes		Possible Ranges							
a1	Day	จันทร์	อังคาร	พุธ	พฤหัสบดี	ศุกร์	เสาร์	อาทิตย์	
	Name	Mon	Tue	Wed	Thu	Fri	Sat	Sun	
a2	Holiday	ใช่	ไม่ใช่						
	Name	Yes	No						
a3	Time	00:00-02:59	03:00-05:59	06:00-08:59	09:00-11:59	12:00-14:59	15:00-17:59	18:00-20:59	21:00-23:59
	Name	T1	T2	T3	T4	T5	T6	T7	T8
a4	Entry	ลาดกระบัง	บางบ่อ	บางปะกง	พนัสนิคม	พานทอง			
	Name	LB	BB	BK	PN	PT			
a5	Exit	ลาดกระบัง	บางบ่อ	บางปะกง	พนัสนิคม	พานทอง			
	Name	LB	BB	BK	PN	PT			
a6	EMR	มี	ไม่มี						
	Name	Yes	No						
a7	Private Car	$\leq 1,953$	1,954-3,906	3,907-5,859	5,860-7,812	$\geq 7,813$			
	Name	R1	R2	R3	R4	R5			
a8	Truck	≤ 172	173-344	345-516	517-688	≥ 689			
	Name	R1	R2	R3	R4	R5			
a9	No. of Vehicles	$\leq 2,062$	2,063-4,124	4,125-6,186	6,187-8,248	$\geq 8,249$			
	Name	R1	R2	R3	R4	R5			

ในงานวิจัยนี้ ผู้วิจัยได้กำหนดนิยามคำว่า “ปริมาณความหนาแน่นการจราจรมากที่สุด” ของทางหลวงพิเศษหมายเลข 7 คือ ค่าเฉลี่ยที่มากที่สุดของจำนวนรถยนต์ ณ วัน ด้านขาเข้า ด้านขาออก และช่วงเวลานั้น ๆ ซึ่งไม่ใช่จำนวนรถยนต์ที่มากที่สุดในชุดข้อมูล ตัวอย่างข้อมูลที่ใช้คำนวณหาปริมาณความหนาแน่นของการจราจรแสดงในตารางที่ 3.6

ตารางที่ 3.6

ตัวอย่างข้อมูลที่ใช้คำนวณหาปริมาณความหนาแน่นการจราจร

ลำดับ	วัน/เดือน/ปี	วัน	ด้านขาเข้า	ด้านขาออก	ช่วงเวลา	จำนวนรถยนต์
1	5/1/2558	จันทร์	ลาดกระบัง	พานทอง	06:00-08:59	8,000
2	5/1/2558	จันทร์	ลาดกระบัง	พานทอง	09:00-11:59	6,000
3	5/1/2558	จันทร์	ลาดกระบัง	บางปะกง	06:00-08:59	5,000
4	12/1/2558	จันทร์	ลาดกระบัง	พานทอง	06:00-08:59	9,000
5	12/1/2558	จันทร์	ลาดกระบัง	พานทอง	09:00-11:59	7,000
6	12/1/2558	จันทร์	ลาดกระบัง	บางปะกง	06:00-08:59	4,000
7	19/1/2558	จันทร์	ลาดกระบัง	พานทอง	06:00-08:59	8,500
8	19/1/2558	จันทร์	ลาดกระบัง	พานทอง	09:00-11:59	6,000
9	19/1/2558	จันทร์	ลาดกระบัง	บางปะกง	06:00-08:59	5,500

จากตารางที่ 3.6 คำนวณหาปริมาณความหนาแน่นการจราจรมากที่สุด ได้ดังนี้

ค่าเฉลี่ยที่ 1 คือ ข้อมูลลำดับที่ 1, 4 และ 7 เป็นข้อมูลการจราจรในวันจันทร์ ด้านขาเข้า คือ ด้านลาดกระบัง ด้านขาออก คือ ด้านพานทอง และอยู่ในช่วงเวลา 06:00 – 08:59 คำนวณหาค่าเฉลี่ยได้ $(8,000 + 9,000 + 8,500)/3$ เท่ากับ 8,500

ค่าเฉลี่ยที่ 2 คือ ข้อมูลลำดับที่ 2, 5 และ 8 เป็นข้อมูลการจราจรในวันจันทร์ ด้านขาเข้า คือ ด้านลาดกระบัง ด้านขาออก คือ ด้านพานทอง และอยู่ในช่วงเวลา 09:00 – 11:59 คำนวณหาค่าเฉลี่ยได้ $(6,000 + 7,000 + 6,000)/3$ เท่ากับ 6,333

ค่าเฉลี่ยที่ 3 คือ ข้อมูลลำดับที่ 3, 6 และ 9 เป็นข้อมูลการจราจรในวันจันทร์ ด้านขาเข้า คือ ด้านลาดกระบัง ด้านขาออก คือ ด้านบางปะกง และอยู่ในช่วงเวลา 06:00 – 08:59 คำนวณหาค่าเฉลี่ยได้ $(5,000 + 4,000 + 5,500)/3$ เท่ากับ 4,833

จากตัวอย่างการคำนวณ ค่าเฉลี่ยที่ 1 มีค่าที่มากที่สุด ดังนั้น งานวิจัยนี้จึงใช้ค่าเฉลี่ยนี้เป็นเกณฑ์กำหนดปริมาณความหนาแน่นการจราจรมากที่สุด โดยแบ่งค่าเฉลี่ยที่ได้ออกเป็นช่วงเท่า ๆ กัน 5 ช่วง ได้แก่ น้อยมาก (R1) น้อย (R2) ปานกลาง (R3) มาก (R4) และมากที่สุด (R5)

3.2.2 การแปลงชุดข้อมูลตามช่วงของข้อมูล

ข้อมูลการจราจรจากตารางที่ 3.4 จะถูกแปลงค่าตามช่วงข้อมูลให้อยู่ในรูปแบบตารางที่ 3.7 โดยใช้หลักเกณฑ์ตามตารางที่ 3.5 เพื่อให้เป็นข้อมูลที่สมบูรณ์และพร้อมนำไปทดลองยกตัวอย่างข้อมูลในแถวที่ 1 อธิบายได้ ดังนี้

- Date มีค่าเป็น 1/8/2557 ซึ่งเป็นวันศุกร์ เทียบค่าได้เท่ากับ Fri
- Holiday เนื่องจาก 1/8/2557 ไม่ใช่วันหยุดนักขัตฤกษ์ เทียบค่าได้เท่ากับ No
- Time มีค่าเป็น 00:00-01:00 อยู่ในช่วงเวลา 00:00-02:59 เทียบค่าได้เท่ากับ T1

- Entry มีค่าเป็น บางบ่อ เทียบค่าได้เท่ากับ BB
- Exit มีค่าเป็น ลาดกระบัง เทียบค่าได้เท่ากับ LB
- EMR มีค่าเป็น 0 แสดงว่าไม่มีอุบัติเหตุ เทียบค่าได้เท่ากับ No
- Private Car มีค่าเป็น 15 อยู่ในชวงน้อยกว่า 1,953 เทียบค่าได้เท่ากับ R1
- Truck มีค่าเป็น 26 อยู่ในชวงน้อยกว่า 172 เทียบค่าได้เท่ากับ R1
- No. of Vehicles มีค่าเป็น 41 อยู่ในชวงน้อยกว่า 2,062 เทียบค่าได้เท่ากับ R1

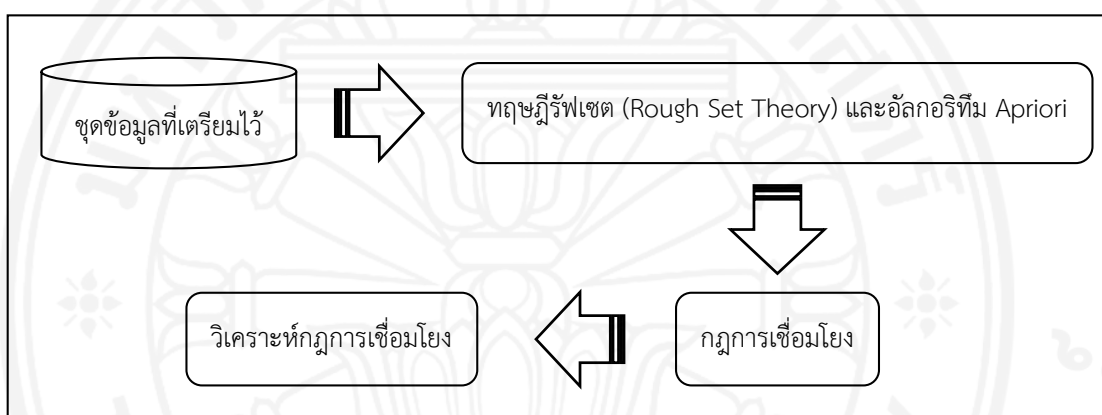
ตารางที่ 3.7

ตัวอย่างข้อมูลหลังจากทำการแปลงค่าตามช่วงของข้อมูล

Condition Attributes								Decision Attributes
Date	Holiday	Time	Entry	Exit	EMR	Private Car	Truck	No. of Vehicles
Fri	No	T1	BB	LB	No	R1	R1	R1
Fri	No	T1	BK	LB	No	R1	R1	R1
Fri	No	T1	PN	LB	No	R1	R1	R1
...
Fri	No	T8	BK	LB	No	R1	R1	R1
Fri	No	T8	PN	LB	No	R1	R1	R1
Fri	No	T8	PT	LB	No	R1	R1	R1

3.3 ขั้นตอนการทดลอง

งานวิจัยนี้ใช้ข้อมูลในการทดลอง ตั้งแต่วันที่ 1 มกราคม 2557 ถึงวันที่ 30 เมษายน 2558 โดยใช้ชุดข้อมูลที่ถูกเตรียมและแบ่งช่วงแล้วเป็นข้อมูลรับเข้า มาทำการทดลองเพื่อหาความสัมพันธ์ระหว่าง วัน วันหยุด ช่วงเวลา ด้านขาเข้า ด้านขาออก จำนวนอุบัติเหตุ จำนวนรถยนต์ส่วนบุคคล จำนวนรถบรรทุก และจำนวนรถยนต์ทั้งหมด ทำการทดลองโดยนำชุดข้อมูลมาผ่านกระบวนการค้นหากฎการเชื่อมโยงด้วยทฤษฎีเซต (Rough Set Theory) และอัลกอริทึม Apriori สุดท้ายนำผลที่ได้มาวิเคราะห์และเปรียบเทียบกฎการเชื่อมโยง ขั้นตอนการทดลองอธิบายตามภาพที่ 3.3



ภาพที่ 3.3 ขั้นตอนการค้นหากฎการเชื่อมโยง

3.4 วิธีวัดผลการทดลอง

การเปรียบเทียบและวัดผลการทดลองที่ได้จากทฤษฎีเซต (Rough Set Theory) และอัลกอริทึม Apriori ใช้การวิเคราะห์และเปรียบเทียบข้อมูลเชิงคุณภาพ เปรียบเทียบความชัดเจนของเงื่อนไขกฎการเชื่อมโยงที่ได้จากทั้ง 2 อัลกอริทึม และเปรียบเทียบข้อมูลเชิงปริมาณในเรื่องจำนวนกฎที่ได้และเวลาที่ใช้ในการประมวลผล

งานวิจัยนี้นำเสนอเครื่องมือสำหรับใช้คัดกรองกฎหรือความสัมพันธ์ที่มีจำนวนมาก โดยใช้เครื่องมือที่เรียกว่า ตัวชี้วัดความน่าสนใจ (Interesting Indicators) กฎการเชื่อมโยงที่ได้คะแนนมากจะเป็นกฎการเชื่อมโยงที่น่าสนใจ ตัวชี้วัดความน่าสนใจถูกนำไปใช้คัดกรองกฎการเชื่อมโยงที่ได้จากทั้ง 2 อัลกอริทึมเพื่อเปรียบเทียบกัน

3.5 เครื่องมือและซอฟต์แวร์ที่ใช้ในการทดลอง

ฮาร์ดแวร์	- หน่วยประมวลผล Intel Core i5 2.67 GHz
	- หน่วยความจำ 4 GB
	- หน่วยเก็บ 500 GB
ระบบปฏิบัติการ	- Windows XP Service Pack 3
ซอฟต์แวร์	- Weka 3.6.11
	- Rosetta Rough set Toolkit
	- MySQL
	- Microsoft Office Excel 2007

3.5.1 ซอฟต์แวร์ Weka (Waikato Environment for Knowledge Analysis) เป็นโปรแกรมที่ได้พัฒนาขึ้นบนพื้นฐานของภาษาจาวา (Java) สามารถประมวลผลได้หลายระบบปฏิบัติการ ซอฟต์แวร์ Weka ได้รับการพัฒนาจากมหาวิทยาลัย Waikato ประเทศนิวซีแลนด์ ถูกพัฒนาสำหรับใช้ในงานวิจัยด้านการเรียนรู้ของเครื่องจักร (Machine Learning) และการทำเหมืองข้อมูล (Data Mining) อัลกอริทึมสามารถเลือกใช้งานได้จาก 2 ทาง คือ อัลกอริทึมที่มีมาพร้อมกับซอฟต์แวร์ หรือผู้ใช้เขียนอัลกอริทึมเป็นโปรแกรมลงไปในชุดเครื่องมือเพิ่มเติม

3.5.2 ซอฟต์แวร์ Rosetta Rough Set Toolkit เป็นเครื่องมือสำหรับวิเคราะห์ข้อมูลในรูปแบบตาราง ออกแบบมาเพื่อการทำเหมืองข้อมูลและการค้นพบความรู้ใหม่ โดยผ่านการคำนวณด้วยชุดคุณลักษณะ ซอฟต์แวร์ Rosetta Rough Set Toolkit รองรับระบบปฏิบัติการ Windows 2000 หรือระบบปฏิบัติการ Windows XP แบบ 32 บิต การพัฒนาถูกแบ่งออกเป็น 2 ส่วน คือ 1) ส่วนโปรแกรม ออกแบบและพัฒนาโดยภาควิชาวิทยาการคอมพิวเตอร์และสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยนอร์เวย์ เมืองทรอนไฮม์ ประเทศนอร์เวย์ และ 2) ส่วนเคอร์เนลประมวลผล (RSES) พัฒนาโดยสถาบันคณิตศาสตร์ มหาวิทยาลัยวอร์ซอ ประเทศโปแลนด์

บทที่ 4

ผลการทดลอง

ในการดำเนินงานวิจัยนี้ ผู้วิจัยได้นำข้อมูลมาทำการแปลงข้อมูลและแบ่งช่วงของข้อมูล (Discretization) ให้อยู่ในช่วงข้อมูลตามตารางที่ 3.5 เพื่อให้ได้ชุดข้อมูลที่พร้อมสำหรับทำการทดลอง ผู้วิจัยได้ใช้ซอฟต์แวร์ Rosetta Rough Set Toolkit เพื่อประมวลผลข้อมูลตามทฤษฎีรีฟเซต (Rough Set Theory) และใช้ซอฟต์แวร์ Weka เพื่อประมวลผลข้อมูลตามอัลกอริทึม Apriori โดยได้ทำการทดลองบนเครื่องคอมพิวเตอร์เครื่องเดียวกัน

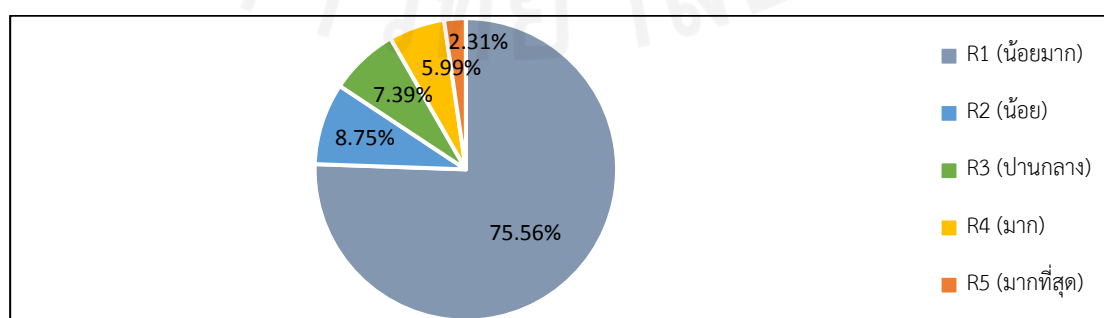
4.1 ข้อมูลทดลอง

ข้อมูลที่ใช้ทดลองในงานวิจัยนี้ หลังจากผ่านกระบวนการเตรียมและแบ่งช่วงข้อมูล มีจำนวนทั้งหมด 4,504 แถว (Rows) โดยมีคุณลักษณะจำนวนรถยนต์ทั้งหมด (No. of Vehicles) เป็นคุณลักษณะตัดสินใจ (Decision Attributes) จำนวนรถยนต์ทั้งหมดจะถูกแปลงค่าให้อยู่ในช่วงข้อมูล 5 ช่วง คือ R1 (น้อยมาก) R2 (น้อย) R3 (ปานกลาง) R4 (มาก) และ R5 (มากที่สุด) จำนวนแถว (Rows) ในแต่ละช่วงคุณลักษณะตัดสินใจ แสดงในตารางที่ 4.1 และร้อยละเปรียบเทียบจำนวนแถว (Rows) ในแต่ละช่วงคุณลักษณะตัดสินใจ แสดงในภาพที่ 4.1

ตารางที่ 4.1

แสดงจำนวนแถว (Rows) ในแต่ละช่วงของคุณลักษณะตัดสินใจ

R1 (น้อยมาก)	R2 (น้อย)	R3 (ปานกลาง)	R4 (มาก)	R5 (มากที่สุด)	จำนวนแถวทั้งหมด
3,403	394	333	270	104	4,504



ภาพที่ 4.1 ร้อยละเปรียบเทียบจำนวนแถว (Rows) ในแต่ละช่วงของคุณลักษณะตัดสินใจ

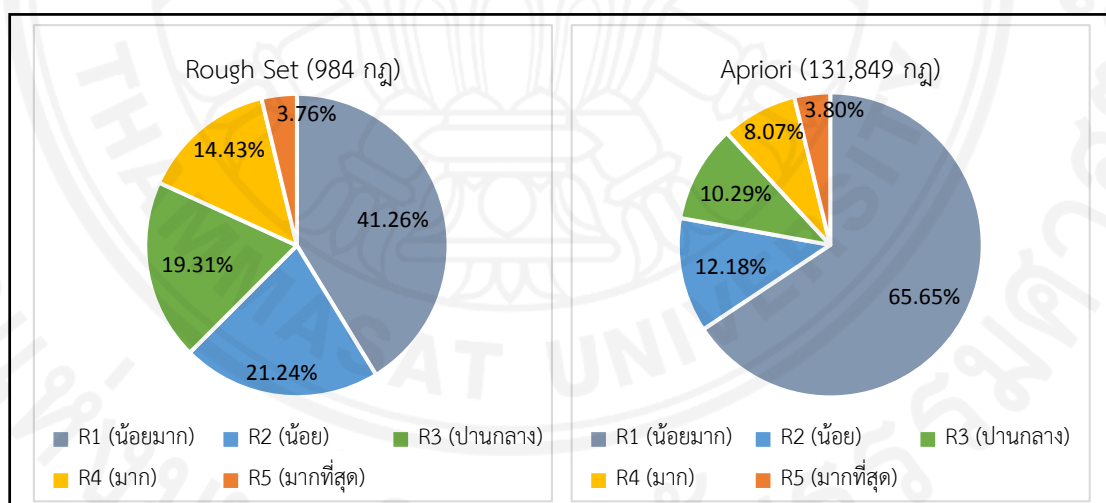
4.2 เปรียบเทียบจำนวนกฎการเชื่อมโยง

จำนวนกฎการเชื่อมโยงที่ได้จากการทดลอง ผู้วิจัยได้สรุปผลโดยแยกตามทฤษฎีและช่วงคุณลักษณะตัดสินใจ (Decision Attributes) แสดงตามตารางที่ 4.2 และร้อยละเปรียบเทียบจำนวนกฎการเชื่อมโยงแสดงในภาพที่ 4.2

ตารางที่ 4.2

แสดงจำนวนกฎการเชื่อมโยง

ทฤษฎี	จำนวนกฎในแต่ละช่วงข้อมูลของคุณลักษณะตัดสินใจ					จำนวนกฎทั้งหมด
	R1 (น้อยมาก)	R2 (น้อย)	R3 (ปานกลาง)	R4 (มาก)	R5 (มากที่สุด)	
Rough Set	406	209	190	142	37	984
Apriori	86,564	16,057	13,568	10,644	5,016	131,849

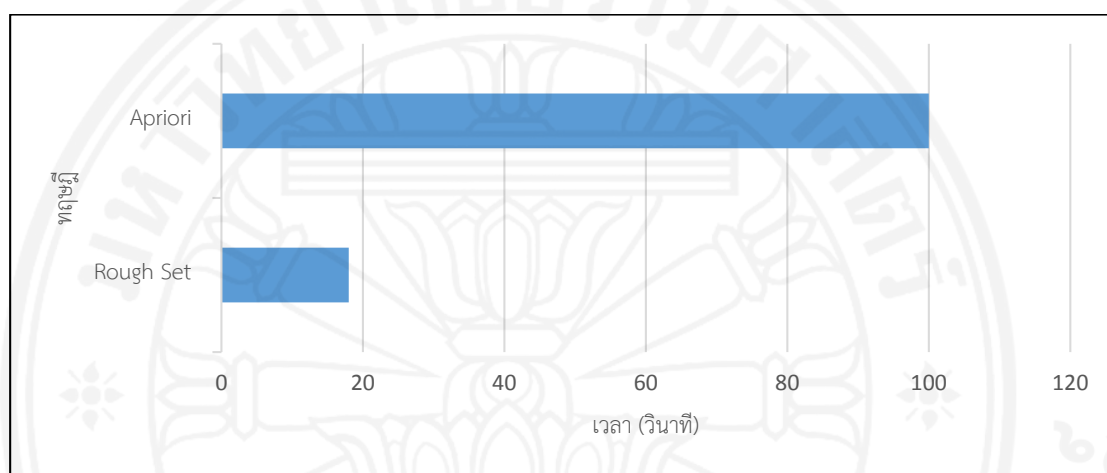


ภาพที่ 4.2 ร้อยละเปรียบเทียบจำนวนกฎการเชื่อมโยง

จากตารางที่ 4.2 แสดงให้เห็นว่าจำนวนกฎการเชื่อมโยงที่ได้จากอัลกอริทึม Apriori นั้นมีจำนวนมากกว่า เนื่องจากกฎการเชื่อมโยงส่วนใหญ่มีเซตย่อย (Subset) เกิดขึ้น ตัวอย่างเช่น กฎที่ 1 มีค่าเป็น {Day=Wed and Private=R5 => R5} และกฎที่ 2 มีค่าเป็น {Day=Wed and Private=R5 and Truck=R4 => R5} หมายความว่า กฎที่ 2 เป็นเซตย่อยของกฎที่ 1 หรือเรียกได้ว่า กฎที่ 1 เป็นกฎที่ครอบคลุมกฎที่ 2 จากตัวอย่างข้างต้นทฤษฎีรัพเซต (Rough Set Theory) จะตัดกฎที่ 2ทิ้ง เพราะกฎการเชื่อมโยงที่ไม่เป็นเซตย่อยของกฎข้ออื่นจะมีความสัมพันธ์ที่ชัดเจนและง่ายต่อความเข้าใจ

4.3 เปรียบเทียบเวลาที่ใช้ในการประมวลผล

จากผลการทดลองพบว่า ทฤษฎีเซต (Rough Set Theory) ใช้เวลาทั้งหมด 18 วินาที ประมวลผลได้จำนวนกฎการเชื่อมโยงทั้งหมด 984 กฎ อัลกอริทึม Apriori ใช้เวลาทั้งหมด 100 วินาที ประมวลผลได้จำนวนกฎการเชื่อมโยงทั้งหมด 131,849 กฎ เวลาในการประมวลผลแสดงในภาพที่ 4.3



ภาพที่ 4.3 แสดงเวลาในการประมวลผล

4.4 เปรียบเทียบกฎการเชื่อมโยงจากตัวชี้วัดความน่าสนใจ (Interesting Indicators)

งานวิจัยนี้ได้เลือกใช้ตัวชี้วัดความน่าสนใจ (Interesting Indicators) จำนวน 5 ตัว ได้แก่ Confidence, Coverage, Leverage, Lift และ Generalization ซึ่งมีค่าคะแนนอยู่ในช่วง 0 ถึง 1 ผลคะแนนที่ได้จากตัวชี้วัดความน่าสนใจ (Interesting Indicators) ถูกใช้เป็นคะแนนในการคัดกรองกฎการเชื่อมโยงที่น่าสนใจจากกฎการเชื่อมโยงทั้งหมด

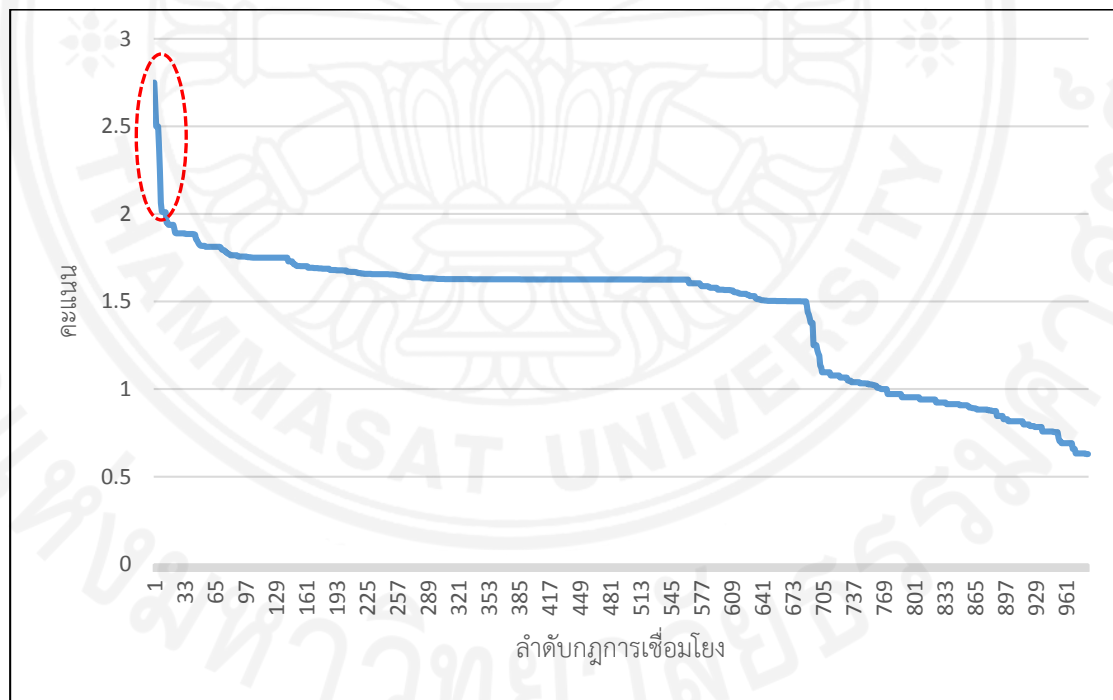
จากการทดลอง กฎการเชื่อมโยงที่ค้นพบจากทั้งสองอัลกอริทึม มีลักษณะเหมือนกัน คือ ประกอบไปด้วยชุดของกฎ 2 ประเภท ได้แก่ กฎที่เป็นความสัมพันธ์ของเซตย่อยที่มีจำนวนรายการสนับสนุนน้อย และกฎที่เป็นความสัมพันธ์ของเซตย่อยที่มีจำนวนรายการสนับสนุนมากซึ่งความสัมพันธ์จากเซตย่อยทั้งสองประเภทมีความเชื่อมั่นเท่ากับ ร้อยละ 100 ทั้งคู่ งานวิจัยนี้นำเสนอมุมมองในการคัดกรองเลือกกฎจากลักษณะของเซตย่อยทั้ง 2 ประเภท เรียกว่า กฎการเชื่อมโยงที่ปรากฏไม่บ่อย และกฎการเชื่อมโยงที่ปรากฏบ่อย ตามลำดับ

4.4.1 กฎการเชื่อมโยงที่ปรากฏไม่บ่อย

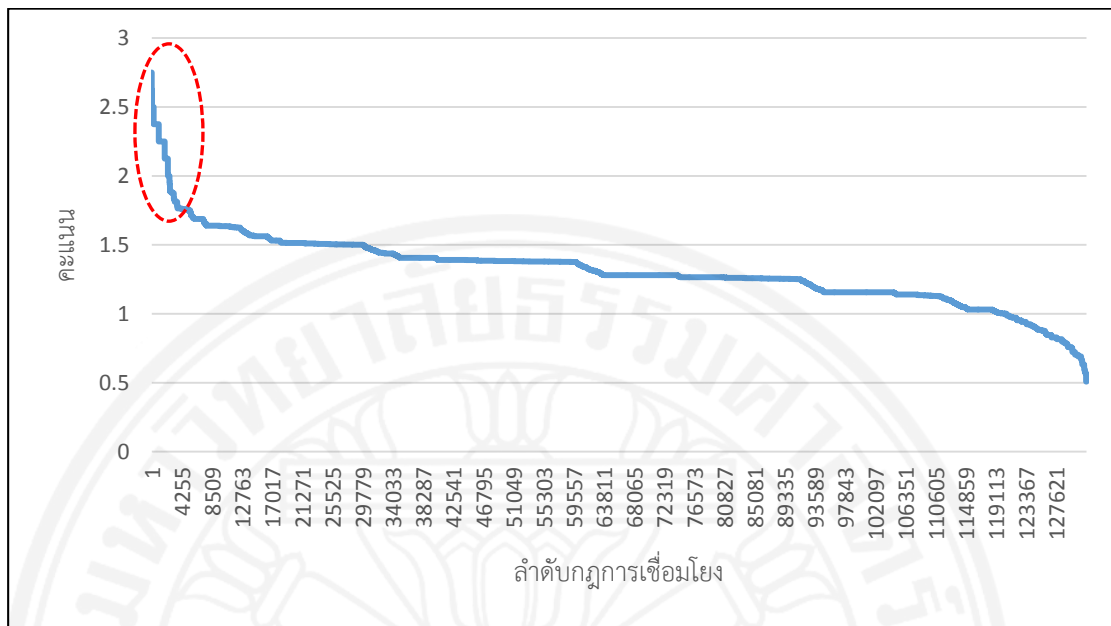
กฎการเชื่อมโยงที่ปรากฏไม่บ่อย คือ กฎการเชื่อมโยงที่ได้คะแนนรวมจากตัวชี้วัด Confidence (Con), Generalization (Gen) และ Lift มากที่สุด

จากข้อมูลในหัวข้อ 4.1 กฎการเชื่อมโยงที่มีคุณลักษณะตัดสินใจเท่ากับ R1 มีจำนวนมากที่สุด รองลงมาคือ R2, R3 และ R4 ตามลำดับ กฎการเชื่อมโยงที่มีคุณลักษณะตัดสินใจเท่ากับ R5 มีจำนวนน้อยที่สุด

ผู้วิจัยได้นำคะแนนทั้งหมดที่ได้จากตัวชี้วัดความน่าสนใจ (Interesting Indicators) มาสร้างเป็นกราฟแสดงในภาพที่ 4.4 และ 4.5 โดยกำหนดให้แกนแนวตั้ง (Y) คือ ระดับคะแนนความน่าสนใจ และแกนแนวนอน (X) คือ ลำดับกฎการเชื่อมโยง จากภาพแสดงให้เห็นว่า สัดส่วนของกฎการเชื่อมโยงที่ได้คะแนนมากนั้นมีจำนวนน้อยเมื่อเปรียบเทียบกับสัดส่วนของกฎการเชื่อมโยงทั้งหมด ซึ่งเกิดขึ้นทั้งในทฤษฎีเซต (Rough Set Theory) และอัลกอริทึม Apriori



ภาพที่ 4.4 แสดงคะแนนกฎการเชื่อมโยงที่ปรากฏไม่บ่อยของทฤษฎีเซต (Rough Set Theory)



ภาพที่ 4.5 แสดงคะแนนกฎการเชื่อมโยงที่ปรากฏไม่บ่อยของอัลกอริทึม Apriori

ตารางที่ 4.3 และ 4.4 แสดงกฎการเชื่อมโยงที่ปรากฏไม่บ่อยและมีคะแนนรวมมากที่สุด 10 อันดับแรกจากทฤษฎีเซต (Rough Set Theory) และอัลกอริทึม Apriori

ตารางที่ 4.3

แสดงกฎการเชื่อมโยงที่ปรากฏไม่บ่อยของทฤษฎีเซต (Rough Set Theory)

No.	Attributes		Interesting Indicators			Interesting Score
	Condition	Decision	Con	Gen	Lift	
1.	Day=Wed and Private=R5	R5	1.0000	0.7500	1.0000	2.7500
2.	Day=Thu and Holiday=Yes and Truck=R5	R5	1.0000	0.6250	1.0000	2.6250
3.	Day=Sat and Time=T3 and Private=R4 and EMR=Yes	R5	1.0000	0.5000	1.0000	2.5000
4.	Day=Fri and Holiday=Yes and Time=T6 and Truck=R5	R5	1.0000	0.5000	1.0000	2.5000
5.	Holiday=Yes and Time=T6 and Truck=R4 and EMR=Yes	R5	1.0000	0.5000	1.0000	2.5000
6.	Day=Sat and Time=T4 and Entry=LB and Truck=R3 and EMR=Yes	R5	1.0000	0.3750	1.0000	2.3750
7.	Day=Fri and Holiday=Yes and Time=T6 and Entry=LB and Private=R4 and Truck=R4	R5	1.0000	0.2500	1.0000	2.2500
8.	Entry=BK and Private=R3	R3	1.0000	0.7500	0.3123	2.0623
9.	Time=T1 and Truck=R5	R2	1.0000	0.7500	0.2640	2.0140
10.	Day=Sat and Time=T8 and Truck=R4	R4	1.0000	0.6250	0.3852	2.0102

ตารางที่ 4.4

แสดงกฎการเชื่อมโยงที่ปรากฏไม่บ่อยของอัลกอริทึม Apriori

No.	Attributes		Interesting Indicators			Interesting Score
	Condition	Decision	Con	Gen	Lift	
1.	Day=Wed and Private=R5	R5	1.0000	0.7500	1.0000	2.7500
2.	Day=Mon and Holiday=No and Private=R5	R5	1.0000	0.6250	1.0000	2.6250
3.	Day=Mon and Time=T4 and Private=R5	R5	1.0000	0.6250	1.0000	2.6250
4.	Day=Mon and Private=R5 and Truck=R5	R5	1.0000	0.6250	1.0000	2.6250
5.	Day=Thu and Holiday=Yes and Truck=R5	R5	1.0000	0.6250	1.0000	2.6250
6.	Day=Thu and Time=T4 and Private=R5	R5	1.0000	0.6250	1.0000	2.6250
7.	Day=Thu and Entry=LB and Private=R5	R5	1.0000	0.6250	1.0000	2.6250
8.	Day=Thu and Exit=PT and Private=R5	R5	1.0000	0.6250	1.0000	2.6250
9.	Day=Thu and Private=R5 and Truck=R2	R5	1.0000	0.6250	1.0000	2.6250
10.	Day=Thu and Private=R5 and Truck=R5	R5	1.0000	0.6250	1.0000	2.6250

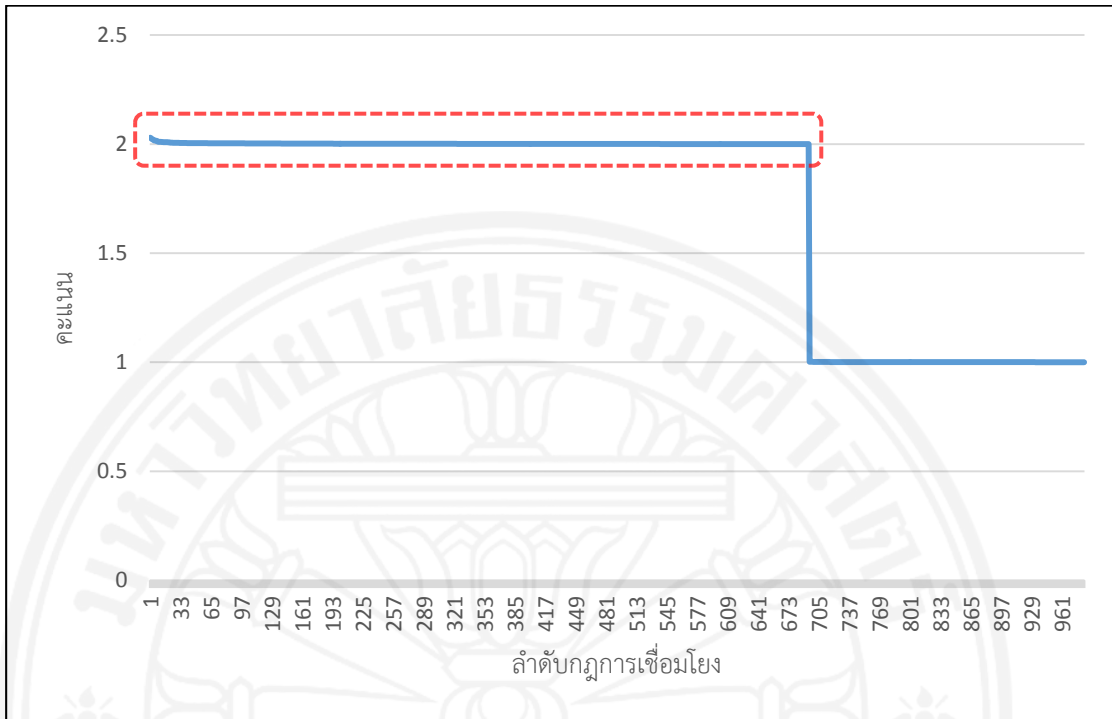
จากตารางที่ 4.3 และ 4.4 พบว่ากฎการเชื่อมโยงที่ได้คะแนนมากที่สุดจากทั้ง 2 ทฤษฎีคล้ายกันและมีคุณลักษณะตัดสินใจส่วนใหญ่ คือ R5 ซึ่งมีสัดส่วนที่น้อยเมื่อเทียบกับจำนวนกฎที่พบทั้งหมด

4.4.2 กฎการเชื่อมโยงที่ปรากฏบ่อย

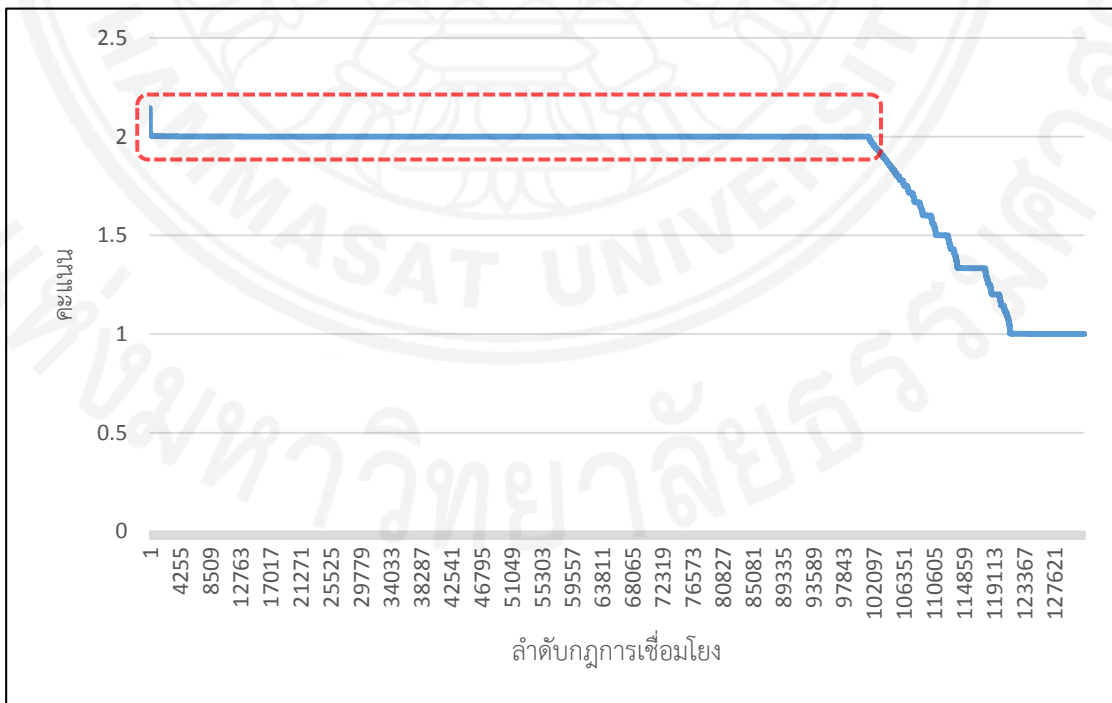
กฎการเชื่อมโยงที่ปรากฏบ่อย คือ กฎการเชื่อมโยงที่ได้คะแนนรวมจากตัวชี้วัด Confidence (Con), Leverage (Lev) และ Coverage (Cov) มากที่สุด

จากข้อมูลในหัวข้อ 4.1 กฎการเชื่อมโยงที่มีคุณลักษณะตัดสินใจเท่ากับ R1 มีจำนวนมากที่สุด รองลงมาคือ R2, R3 และ R4 ตามลำดับ กฎการเชื่อมโยงที่มีคุณลักษณะตัดสินใจเท่ากับ R5 มีจำนวนน้อยที่สุด

ผู้วิจัยได้นำคะแนนทั้งหมดที่ได้จากตัวชี้วัดความน่าสนใจ (Interesting Indicators) มาสร้างเป็นกราฟ แสดงในภาพที่ 4.6 และ 4.7 โดยกำหนดให้แกนแนวตั้ง (Y) คือ ระดับคะแนน และแกนแนวนอน (X) คือ ลำดับกฎการเชื่อมโยง จากภาพแสดงให้เห็นว่า สัดส่วนของกฎการเชื่อมโยงที่ได้คะแนนมากนั้นมีจำนวนมากเมื่อเปรียบเทียบกับสัดส่วนของกฎการเชื่อมโยงทั้งหมด ซึ่งเกิดขึ้นทั้งในทฤษฎีรีฟเซต (Rough Set Theory) และอัลกอริทึม Apriori



ภาพที่ 4.6 แสดงคะแนนกฎการเชื่อมโยงที่ปรากฏบ่อยของทฤษฎีเซต (Rough Set Theory)



ภาพที่ 4.7 แสดงคะแนนกฎการเชื่อมโยงที่ปรากฏบ่อยของอัลกอริทึม Apriori

ตารางที่ 4.5 และ 4.6 แสดงกฎการเชื่อมโยงที่ปรากฏบ่อยและมีคะแนนรวมมากที่สุด 10 อันดับแรกจากทฤษฎีเซต (Rough Set Theory) และอัลกอริทึม Apriori

ตารางที่ 4.5

แสดงกฎการเชื่อมโยงที่ปรากฏบ่อยของทฤษฎีเซต (Rough Set Theory)

No.	Attributes		Interesting Indicators			Interesting Score
	Condition	Decision	Con	Cov	Lev	
1.	Entry=BK and Private=R1 and Truck=R1	R1	1.0000	0.1197	0.9096	2.0293
2.	Entry=PN and Private=R1 and Truck=R1	R1	1.0000	0.1083	0.9181	2.0265
3.	Day=Sun and Private=R1	R1	1.0000	0.0926	0.9300	2.0226
4.	Private=R1 and Truck=R1 and EMR=Yes	R1	1.0000	0.0846	0.9361	2.0207
5.	Time=T7 and Private=R1 and Truck=R1	R1	1.0000	0.0702	0.9470	2.0172
6.	Holiday=Yes and Exit=BB and Private=R1	R1	1.0000	0.0653	0.9507	2.0160
7.	Time=T2 and Truck=R1	R1	1.0000	0.0633	0.9522	2.0155
8.	Holiday=Yes and Entry=BK and Private=R1	R1	1.0000	0.0599	0.9547	2.0147
9.	Entry=LB and Exit=BB and Private=R1	R1	1.0000	0.0466	0.9648	2.0114
10.	Entry=PT and Exit=BK and Private=R1	R1	1.0000	0.0435	0.9671	2.0106

ตารางที่ 4.6

แสดงกฎการเชื่อมโยงที่ปรากฏบ่อยของอัลกอริทึม Apriori

No.	Attributes		Interesting Indicators			Interesting Score
	Condition	Decision	Con	Cov	Lev	
1.	Private=R1	R1	0.9781	0.7693	0.3968	2.1442
2.	Private=R1 and Truck=R1	R1	0.9954	0.5273	0.5970	2.1196
3.	Private=R1 and EMR=No	R1	0.9764	0.6485	0.4864	2.1113
4.	Private=R1 and Truck=R1 and EMR=No	R1	0.9945	0.4427	0.6600	2.0972
5.	Holiday=No and Private=R1 and Truck=R1	R1	0.9939	0.2924	0.7730	2.0593
6.	Truck=R1	R1	0.9620	0.5488	0.5473	2.0581
7.	Holiday=Yes and Private=R1	R1	0.9931	0.2913	0.7731	2.0575
8.	Holiday=No and Private=R1	R1	0.9689	0.4780	0.6077	2.0546
9.	Holiday=Yes and Private=R1 and EMR=No	R1	0.9928	0.2782	0.7826	2.0536
10.	Holiday=Yes and Private=R1 and Truck=R1	R1	0.9972	0.2349	0.8197	2.0518

จากตารางที่ 4.5 และ 4.6 พบว่ากฎการเชื่อมโยง 2 ทฤษฎีคล้ายกันและมีคุณลักษณะที่ตัดสินใจ คือ R1 ซึ่งมีสัดส่วนมากเมื่อเปรียบเทียบกับจำนวนกฎที่ค้นพบทั้งหมด

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาความสัมพันธ์ของการจรรยาบรรณบนถนนทางหลวงพิเศษ หมายเลข 7 ข้อมูลการจราจรถูกนำมาแปลงข้อมูลให้อยู่ในรูปแบบเรคคอร์ด (Record-data Format) และแบ่งช่วงของข้อมูล (Discretization) เพื่อให้เป็นข้อมูลที่สมบูรณ์ จากนั้นนำชุดข้อมูลไปวิเคราะห์หาความสัมพันธ์ด้วยทฤษฎีเซต (Rough Set Theory) เปรียบเทียบกับอัลกอริทึม Apriori โดยใช้ซอฟต์แวร์ Rosetta Rough Set Toolkit และซอฟต์แวร์ Weka ตามลำดับ ความสัมพันธ์ที่ได้จะอยู่ในรูปแบบของกฎการเชื่อมโยง กฎการเชื่อมโยงที่ได้จากการทดลองมีจำนวนมากทำให้การตัดสินใจเลือกกฎการเชื่อมโยงที่น่าสนใจนั้นมีความยากลำบากหากจะต้องพิจารณาจากกฎทุกกฎ งานวิจัยนี้จึงใช้เครื่องมือที่เรียกว่า ตัวชี้วัดความน่าสนใจ (Interesting Indicators) ได้แก่ Confidence, Coverage, Leverage, Lift และ Generalization มาเปรียบเทียบคุณภาพและคัดกรองกฎการเชื่อมโยงที่น่าสนใจได้

5.1 สรุปผลการวิจัย

จำนวนกฎการเชื่อมโยงที่ได้จากทฤษฎีเซต (Rough Set Theory) มีจำนวน 984 กฎ และจากอัลกอริทึม Apriori มีจำนวน 131,849 กฎ ซึ่งมีความแตกต่างกัน เนื่องจากกฎการเชื่อมโยงที่ได้จากอัลกอริทึม Apriori มีเกิดเซตย่อยของคุณลักษณะเงื่อนไขขึ้นจำนวนมาก แต่กฎการเชื่อมโยงที่ได้จากทฤษฎีเซต (Rough Set Theory) ไม่ปรากฏเซตย่อยของคุณลักษณะเงื่อนไข เป็นผลทำให้กฎการเชื่อมโยงที่ได้มีจำนวนน้อย แต่กฎการเชื่อมโยงที่ได้มีความสัมพันธ์ที่ชัดเจนและสามารถทำความเข้าใจได้ง่าย

เวลาที่ใช้ในการประมวลผล ทฤษฎีเซต (Rough Set Theory) ใช้เวลาในการประมวลผลน้อยกว่าอัลกอริทึม Apriori เมื่อข้อมูลรับเข้าเป็นชุดข้อมูลเดียวกัน

ค่าคะแนนตัวชี้วัดความน่าสนใจ (Interesting Indicators) ถูกนำเสนอเพื่อใช้คัดกรองกฎการเชื่อมโยงส่วนหนึ่งจากจำนวนกฎที่มีทั้งหมด งานวิจัยนี้นำเสนอมุมมองในการคัดกรอง 2 มุมมอง คือ 1) มุมมองที่ต้องการเลือกกฎการเชื่อมโยงที่มีสัดส่วนน้อยและปรากฏไม่บ่อย โดยใช้ตัวชี้วัดความน่าสนใจ Confidence, Generalization และ Lift และ 2) มุมมองที่ต้องการเลือกกฎการเชื่อมโยงที่มีสัดส่วนมากและปรากฏบ่อยโดยใช้ตัวชี้วัดความน่าสนใจ Confidence, Leverage และ Coverage

5.2 ข้อเสนอแนะ

ความสัมพันธ์ในรูปแบบกฎการเชื่อมโยงที่ได้จากทั้งทฤษฎีรีฟเซต (Rough Set Theory) และอัลกอริทึม Apriori เป็นการอธิบายธรรมชาติของข้อมูลในแต่ละเซตย่อยเท่านั้น ควรให้ความระมัดระวังในการแปลผลของกฎการเชื่อมโยง ความสัมพันธ์ที่ค้นพบไม่สามารถใช้ทำนายชุดข้อมูลใหม่ หรือไม่สามารถนำไปใช้กับลักษณะของเซตหรือชุดข้อมูลอื่นได้

5.3 งานวิจัยในอนาคต

งานวิจัยนี้ผู้วิจัยได้เลือกใช้ข้อมูลการจราจรบนทางหลวงพิเศษหมายเลข 7 เพียงอย่างเดียว แต่ยังมีข้อมูลการจราจรบนทางด่วนบูรพาวิถี (บางนา-ตราด) ซึ่งเป็นอีกหนึ่งเส้นทางในการคมนาคมสู่ภาคตะวันออก

งานวิจัยในอนาคต นักวิจัยสามารถเลือกข้อมูลการจราจรเฉพาะช่วงเวลาที่มีการจราจรหนาแน่นปานกลางถึงมากที่สุดมาดำเนินการวิจัยได้ นอกจากนี้ ยังสามารถเพิ่มคุณลักษณะสภาพอากาศ สาเหตุการเกิดอุบัติเหตุ หรือแบ่งประเภทรถยนต์ได้อีก เพื่อให้สามารถอธิบายความสัมพันธ์ที่ซ่อนอยู่ได้ชัดเจนมากยิ่งขึ้น

รายการอ้างอิง

บทความวารสาร

- R. Tian and Z. Yang, "Method of Road Traffic Accidents Causes Analysis Based on Data Mining", Proc. of International Conference on Computational Intelligence and Software Engineering (CiSE), 2010, pp. 1-4.
- S. K. Manga, "Currency Crises Prediction with Rough Set Theory", International Journal of Computer Applications, vol. 32, 2011, pp. 48-52.
- F. N. Ogwueleka and T. C. Ogwueleka, "Traffic Accident Data Profiling and Clustering with Data Mining Process", IOSR Journal of Computer Engineering, vol.6, 2012, pp. 14-22.
- X. Zheng and J. Xu, "Studies on the Application of Rough set Analysis in Mining of Association Rules and the Realization in Provincial Road Transportation Management Information System", International Conference on Industrial Control and Electronics Engineering, 2012.
- C. Erden and F. Tüysüz, "An Application of Rough Sets Theory on Traffic Accidents", Proc. of International Conference on Engineering and Applied Sciences Optimization, 2014.
- Z. Pawlak, Rough Sets, "In Theoretical Aspects of Reasoning about Data", Kluwer, Netherlands, 1991.
- M N. A Rahman and Y. M Lazim and F. Mohamed, "Applying Rough Set Theory in Multimedia Data Classification", The Society of Digital Information and Wireless Communications, 2011.
- L. Geng and HJ. Hamilton, "Interestingness Measures for Data Mining: A Survey", ACM Computing Surveys, vol. 38, 2006.
- Z. Pawlak and J. Grzymala-Busse and R. Slowinski and W. Ziarko, "Rough Sets", Communications of the ACM, vol. 38, 1995, pp. 88-95.

- J. Hong-bo and Y. De-li, "Application Research on Fast Discovery of Association Rules Based on Air Transportation", Proc. of International Conference on Service Systems and Service Management, 2007, pp. 1-6.
- M H. F. Zarandi, "Application of Rough Set Theory in Data Mining for Decision Support Systems (DSSs)", Journal of Industrial Engineering, vol.1, 2008, pp.25-34.
- Y. Zhao and H. Jiang and R. Geng and X. Dong, "Mining Weighted Negative Association Rules Based on Correlation from Infrequent Items", Advanced Computer Control, 2009. ICACC '09. International Conference on, 2009.
- Pakorn Leesutthipornchai, "Association Rules Selection Approach Based on Interesting Measures", Proc. of the Fourth International Conference on Informatics & Applications (ICIA2015), 2015, pp.62-67.

สื่ออิเล็กทรอนิกส์

- พิชญ์ แก้วกสิ, "การขุดค้นข้อมูล (Data Mining)", สืบค้นเมื่อวันที่ 11 ตุลาคม 2557,
<http://www.sut.ac.th/ist/Courses/204406/DataMining1.ppt>
- รัตนาวดี พานทอง, "การทำเหมืองข้อมูล (Data Mining)", สืบค้นเมื่อวันที่ 11 ตุลาคม 2557,
<http://www.ict.up.ac.th/rattanawadeep/dw/chapter8.pdf>
- ผศ.วิภาวรรณ บัวทอง, "Association Rule (เหมืองข้อมูลแบบกฎความสัมพันธ์)", สืบค้นเมื่อวันที่ 12 ตุลาคม 2557,
<https://wipawanblog.files.wordpress.com/2014/06/chapter-4-association-rule.pdf>
- Agnieszka Nowak, "Rough Set Theory in Decision Support Systems", สืบค้นเมื่อวันที่ 14 ตุลาคม 2557,
<http://zsi.tech.us.edu.pl/~nowak/bien/w2.pdf>
- Rosetta Rough set Toolkit, สืบค้นเมื่อวันที่ 18 ตุลาคม 2557,
<http://www.lcb.uu.se/tools/rosetta/>
- WEKA 3.6.11 (2014), สืบค้นเมื่อวันที่ 18 ตุลาคม 2557,
<http://www.cs.waikato.ac.nz/ml/weka/>

ประวัติผู้เขียน

ชื่อ	นายปณณวัฒน์ ศรีรัตน์
วันเดือนปีเกิด	2 พฤษภาคม 2526
วุฒิการศึกษา	วิทยาศาสตร์บัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์ พ.ศ. 2550
ผลงานทางวิชาการ	1) Pannawat Sriratnat and Pakorn Leesutthipornchai “Association Mining based on Thai Highway Traffic Data by Using Rough Set Theory” The 19 th International Annual Symposium on Computational Science and Engineering (ANSCSE19), June 2015, Ubon Ratchathani, Thailand. 2) Pannawat Sriratna and Pakorn Leesutthipornchai “Interesting-based Association Rules for Highway Traffic Data” The 19 th International Computer Science and Engineering Conference (ICSEC 2015), November 2015, Chiang Mai, Thailand.
ประสบการณ์ทำงาน	1) มกราคม 2552 - ปัจจุบัน : นักวิชาการคอมพิวเตอร์ บริษัท ทีโอที จำกัด (มหาชน) 2) พฤศจิกายน 2549 – ตุลาคม 2551 : ทหารกองประจำการ สังกัด กองพันทหารอากาศโยธิน กองบิน 21 (อุบลราชธานี) 3) มีนาคม 2549 – ตุลาคม 2549 : โปรแกรมเมอร์ บริษัท เอเออี ซีสเต็ม จำกัด