

ขั้นตอนวิธีที่มีประสิทธิภาพสำหรับการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ

AN EFFICIENT ALGORITHM FOR MINING MULTIDIMENSIONAL
SEQUENTIAL PATTERNS

พนิดา พงษ์รัมย์
PANIDA SONGRAM

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของเกรตดิพลอมทรีศึกษาระดับปริญญาโท สาขาวิศวกรรมคอมพิวเตอร์

สาขาวิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2549

ISBN 974-15-2122-7

ขั้นตอนวิธีที่มีประสิทธิภาพสำหรับการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ

AN EFFICIENT ALGORITHM FOR MINING MULTIDIMENSIONAL
SEQUENTIAL PATTERNS

พนิดา ทรงรัมย์

PANIDA SONGRAM

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ.2549

ISBN 974-15-2122-7

**AN EFFICIENT ALGORITHM FOR MINING MULTIDIMENSIONAL
SEQUENTIAL PATTERNS**

PANIDA SONGRAM

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE IN COMPUTER SCIENCE
SCHOOL OF GRADUATE STUDIES
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

2006

ISBN 974-15-2122-7

COPYRIGHT 2006

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

หัวข้อวิทยานิพนธ์	ขั้นตอนวิธีที่มีประสิทธิภาพสำหรับการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ
นักศึกษา	นางสาวพนิดา ทรงรัมย์
รหัสประจำตัว	46063603
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2549
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร.วีระ บุญจริง

บทคัดย่อ

งานวิจัยนี้เสนอกรรมวิธีใหม่ในการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ ซึ่งเรียกว่า การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด โดยผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด ความถูกต้องของกรรมวิธีใหม่นี้สนับสนุนโดยทฤษฎีต่อไปนี้: (1) การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดให้จำนวนรูปแบบที่น้อยกว่าหรือเท่ากับจำนวนรูปแบบที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ, (2) สามารถสร้างรูปแบบลำดับเหตุการณ์หลายมิติทั้งหมดจากรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดได้, และ (3) การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดจะทำการสืบค้นกับสารสนเทศหลายมิติที่สัมพันธ์กับลำดับเหตุการณ์แบบปิดเท่านั้น ในกรณีที่ทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์ และการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดจะทำการสืบค้นกับลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูลแบบปิดเท่านั้น ในกรณีที่ทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดที่สัมพันธ์; กรรมวิธีใหม่นี้ประกอบด้วยขั้นตอนหลักสองขั้นตอนได้แก่ (1) การรวมการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดเข้ากับการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด และ (2) การจัดรูปแบบซ้ำซ้อน; ขั้นตอนที่สองเป็นขั้นตอนที่จำเป็นเนื่องจากขั้นตอนแรกให้ผลลัพธ์ที่เป็นรูปแบบคู่แข่งที่ซ้ำซ้อน ในขั้นตอนแรกการรวมทำได้สองวิธี ได้แก่ (1) การสืบค้นรูปแบบกลุ่มข้อมูลปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดที่สัมพันธ์ และ (2) การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์; จากนั้นได้ทำการทดลองโดยใช้วิธีการสองวิธีนี้เพื่อหาปัจจัยที่ส่งผลต่อจำนวนรูปแบบคู่แข่งโดยพิจารณาค่าสนับสนุนต่ำสุดที่ไม่มากกว่า 20 ผลจากการทดลองพบว่า วิธีการแรกให้จำนวนรูปแบบคู่แข่งน้อยกว่า เมื่อสารสนเทศหลายมิติมีจำนวนมิติไม่มากกว่า 3 มิติ มิฉะนั้นแล้ววิธีการที่สองจะให้จำนวนรูปแบบคู่แข่งน้อยกว่าวิธีการแรก

Thesis	An Efficient Algorithm for Mining Multidimensional Sequential Patterns
Student	Miss.Pandia Songram
Student ID	46063603
Degree	Master of Science
Programme	Computer Science
Year	2006
Thesis Advisor	Assoc.Prof.Dr.Veera Boonjing

ABSTRACT

This research proposes a new method, called closed multidimensional sequential pattern mining, for mining multidimensional sequential patterns. The new method is an integration of closed sequential pattern mining and closed itemset pattern mining. Its correctness is supported by the following theories: (1) the number of complete closed multidimensional sequential patterns is not larger than the number of complete multidimensional sequential patterns, (2) the set of complete multidimensional sequential patterns can be generated from the set of complete closed multidimensional sequential patterns, and (3) mining using closed itemset pattern mining on multidimensional information would mine only multidimensional information associated with mined closed sequential patterns, and mining using closed sequential pattern mining on sequences would mine only sequences associated with mined closed itemset patterns. This new method consists of two major steps: (1) combining closed itemset pattern mining with closed sequential pattern mining and (2) eliminating redundant patterns. The second step is needed because the first step gives redundant candidate patterns. In the first step, there are two approaches for combining: (1) mining using closed sequential pattern mining on sequences associated with mined closed itemset patterns and (2) mining using closed itemset pattern mining on multidimensional information associated with mined closed sequential patterns. We performed experiments using these two approaches to determine factors effecting number of candidate patterns at minimum support not exceed 20. The results show that the first approach gives less candidate patterns than the second when multidimensional information is not larger than 3 dimensions. Otherwise, the second gives less candidate patterns than the first.

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จได้ด้วยความกรุณาจากอาจารย์ที่ปรึกษา รศ.ดร.วีระ บุญจริง ที่ให้ความช่วยเหลือ ให้คำชี้แนะช่วยแก้ปัญหาตลอดจนให้ความรู้และประสบการณ์ที่ดีแก่ข้าพเจ้า

ขอขอบคุณ Xifeng Yan สำหรับโปรแกรม CloSpan และคำแนะนำ ขอขอบคุณ Simone Sheu สำหรับคำแนะนำและโปรแกรมสังเคราะห์ชุดข้อมูลลำดับเหตุการณ์

ขอขอบนายรุ่งโรจน์ วชิรจารุ และ นายอภิวัฒน์ ศรีเลิศมาลา ที่ช่วยแนะนำการเขียนโปรแกรม และเพื่อนๆ ที่คอยเป็นแรงผลักดันให้ข้าพเจ้ามีแรงในการทำวิทยานิพนธ์ให้สำเร็จ

สุดท้ายขอพระคุณบิดามารดาและครอบครัวของข้าพเจ้าที่เป็นกำลังใจ คอยผลักดันข้าพเจ้ามาโดยตลอด

สำหรับคุณงามความดีอันใดที่เกิดจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบให้กับบิดามารดาที่รักและเคารพยิ่ง ตลอดจนครูอาจารย์ที่เคารพทุกท่านที่ได้ประสิทธิ์ประสาทวิชาความรู้และถ่ายทอดประสบการณ์ที่ดีให้กับข้าพเจ้า

พนิดา ทรงรัมย์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา.....	2
1.3 สมมุติฐานของการศึกษา.....	2
1.4 ขอบเขตการวิจัย.....	2
1.5 ส่วนประกอบของวิทยานิพนธ์.....	3
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	4
2.1 การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด.....	4
2.1.1 นิยามพื้นฐาน.....	4
2.1.2 คุณสมบัติของรูปแบบกลุ่มข้อมูลแบบปิด.....	7
2.1.3 การพัฒนาขั้นตอนวิธีสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด.....	7
2.2 การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด.....	8
2.2.1 นิยามพื้นฐาน.....	9
2.2.2 การพัฒนาขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด.....	11
2.3 การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ.....	12
2.3.1 นิยามพื้นฐาน.....	12
2.3.2 การพัฒนาขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ.....	13
บทที่ 3 การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด.....	15
3.1 นิยามพื้นฐาน.....	15
3.2 คุณสมบัติของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด.....	18

สารบัญ(ต่อ)

	หน้า
3.3 ขั้นตอนวิธี.....	20
3.3.1 การผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้น รูปแบบลำดับเหตุการณ์แบบปิด	20
3.3.2 การจัดความซ้ำซ้อน.....	23
3.4 ความถูกต้อง.....	24
บทที่ 4 การทดลองและผลการทดลอง	28
4.1 การทดลอง.....	28
4.1.1 ข้อมูลที่ใช้ในการทดลอง	28
4.1.2 โปรแกรมที่ใช้ในการทดลอง.....	31
4.1.3 เครื่องมือที่ใช้ในการวิจัย.....	31
4.1.4 การออกแบบการทดลองและเหตุผล.....	31
4.2 ผลการทดลอง.....	32
4.2.1 ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 10,000 รายการ.....	32
4.2.2 ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 20,000 รายการ.....	34
4.2.3 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 10 มิติ	35
4.2.4 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 2 มิติ	36
4.2.5 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 4 มิติ	37
4.2.6 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 3 มิติ	38
4.2.7 สรุปผลการทดลอง.....	39
บทที่ 5 สรุปและข้อเสนอแนะ.....	40
5.1 สรุป.....	40
5.2 ข้อเสนอแนะ	41
เอกสารอ้างอิง.....	42

สารบัญ(ต่อ)

หน้า

ภาคผนวก ก โปรแกรมที่ใช้ในการสังเคราะห์ลำดับเหตุการณ์หลายมิติ.....45

ประวัติผู้เขียน.....50

สารบัญตาราง

ตารางที่	หน้า
2.1	ฐานข้อมูลกลุ่มข้อมูล..... 6
2.2	ฐานข้อมูลลำดับเหตุการณ์..... 10
2.3	ตัวอย่างฐานข้อมูลการซื้อสินค้า..... 13
3.1	ฐานข้อมูลลำดับเหตุการณ์หลายมิติ..... 16
3.2	สารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์ (bcb)..... 26
3.3	ลำดับเหตุการณ์ที่สัมพันธ์กับกลุ่มข้อมูล (2,6) 27
4.1	แสดงความหมายของพารามิเตอร์ที่ใช้ในการสร้างลำดับเหตุการณ์ 29
4.2	ค่าพารามิเตอร์ที่ป้อนเข้าไปในโปรแกรมสังเคราะห์ข้อมูล 29
4.3	แสดงค่าแต่ละมิติของสารสนเทศหลายมิติ..... 30
4.4	แสดงรายละเอียดของชุดข้อมูลลำดับเหตุการณ์หลายมิติ 30
4.5	แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้น ตามการทดลองในส่วนที่ 1 33
4.6	แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้น ตามการทดลองในส่วนที่ 2 34
4.7	แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้น ตามการทดลองในส่วนที่ 3 35
4.8	แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้น ตามการทดลองในส่วนที่ 4 36
4.9	แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้น ตามการทดลองในส่วนที่ 5 37
4.10	แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้น ตามการทดลองในส่วนที่ 6 38

สารบัญรูป

รูปที่	หน้า
4.1 ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 10,000 รายการ	33
4.2 ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 20,000 รายการ	34
4.3 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 10 มิติ.....	36
4.4 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 2 มิติ.....	37
4.5 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 4 มิติ.....	38
4.6 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 3 มิติ.....	39

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ข้อมูลเป็นสิ่งที่มีความสำคัญและเข้ามามีบทบาทในชีวิตประจำวัน เทคโนโลยีทางคอมพิวเตอร์ได้ถูกนำมาใช้ในการจัดเก็บข้อมูลต่างๆ เหล่านี้ในรูปแบบของฐานข้อมูล เพื่อนำข้อมูลที่จัดเก็บไปใช้ประโยชน์ในด้านต่างๆ นอกจากนี้ยังมีการนำเทคโนโลยีสารสนเทศการสืบค้นความรู้บนฐานข้อมูลขนาดใหญ่หรือเรียกว่าการทำเหมืองข้อมูลมาวิเคราะห์และกลั่นกรองข้อมูลที่มีปริมาณมหาศาล เพื่อให้ได้ผลลัพธ์ที่สามารถนำไปใช้ประโยชน์ได้

การสืบค้นรูปแบบลำดับเหตุการณ์ (Sequential Pattern Mining) เป็นเทคนิคที่สำคัญในการสืบค้นความรู้บนฐานข้อมูลขนาดใหญ่ ผลที่ได้จากการสืบค้นจะได้รูปแบบลำดับเหตุการณ์ที่สามารถนำมาใช้ประโยชน์ในด้านต่าง ๆ เช่น การวิเคราะห์พฤติกรรมการซื้อของลูกค้า ลำดับการเข้าไปยังหน้าต่างๆของเว็บไซต์ กระบวนการวิเคราะห์และการทดลองทางวิทยาศาสตร์ การทำนายเหตุการณ์ทางธรรมชาติ การวิเคราะห์ลำดับสายพันธุกรรม เป็นต้น ถ้ามีการสร้างความสัมพันธ์ระหว่างรูปแบบลำดับเหตุการณ์กับสารสนเทศหลายมิติ (Multidimensional Information) จะทำให้ได้รูปแบบลำดับเหตุการณ์ที่มีประโยชน์มากขึ้น เช่น รูปแบบลำดับเหตุการณ์พฤติกรรมการซื้อของลูกค้าในร้านขายอุปกรณ์คอมพิวเตอร์เป็นดังนี้ $P1 = (\text{computer, speaker}) \rightarrow \text{printer} \rightarrow \text{paper}$ หมายความว่าลูกค้าซื้อคอมพิวเตอร์พร้อมกับลำโพงแล้วต่อมาจะซื้อเครื่องพิมพ์และซื้อกระดาษตามลำดับ ซึ่งรูปแบบลำดับเหตุการณ์ที่ได้จากการสืบค้นไม่ได้ระบุว่ารูปแบบลำดับเหตุการณ์ดังกล่าวเหมาะกับกลุ่มลูกค้าวัยใด แต่ถ้ามีการนำข้อมูลอายุลูกค้าหรือสารสนเทศหลายมิติอื่นมาสร้างความสัมพันธ์กับรูปแบบลำดับเหตุการณ์ ก็จะสามารถรู้ความต้องการของลูกค้าแต่ละวัยและสามารถนำเสนอสินค้าให้เหมาะสมกับผู้ซื้อได้

การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ (Multidimensional Sequential Pattern Mining) คือ การสืบค้นรูปแบบที่เกิดจากความสัมพันธ์ระหว่างรูปแบบลำดับเหตุการณ์และสารสนเทศหลายมิติ วิธีการสืบค้นความสัมพันธ์ดังกล่าวเริ่มต้นด้วยการหารูปแบบที่มีขนาดเล็กและทำการขยายรูปแบบไปจนถึงรูปแบบขนาดใหญ่ ถ้าฐานข้อมูลมีขนาดใหญ่หรือค่าสนับสนุนต่ำสุดมีค่าต่ำ อาจทำให้เกิดรูปแบบที่เป็นไปได้จำนวนมาก ซึ่งต้องใช้เนื้อที่ในคอมพิวเตอร์จำนวนมากในการเก็บ

รูปแบบทั้งหมด งานวิจัยนี้จึงเสนอแนวทางใหม่ในการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ เพื่อให้ได้รูปแบบขนาดใหญ่ที่สามารถครอบคลุมคำตอบที่เป็นไปได้ทั้งหมด

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการใหม่ที่เรียกว่า การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ซึ่งจะทำการสืบค้นรูปแบบขนาดใหญ่ที่สามารถครอบคลุมคำตอบที่เป็นไปได้ทั้งหมด

1.3 สมมติฐานของการศึกษา

การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด เป็นการผสมผสานวิธีการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดเข้ากับวิธีการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด ซึ่งจะสามารถลดจำนวนรูปแบบได้

1.4 ขอบเขตการวิจัย

งานวิจัยนี้เป็นแนวทางในการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด (Closed Multidimensional Sequential Pattern Mining) โดยการผสมผสานขั้นตอนวิธีการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด (Closed Sequential Pattern Mining) และการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด (Closed Itemset Pattern Mining) นอกจากนี้ยังทำการเปรียบเทียบผลลัพธ์ที่ได้จากวิธีผสมผสานดังกล่าว 2 วิธี คือ วิธีการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดที่สัมพันธ์ และวิธีการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์ ซึ่งมีขอบเขตการวิจัยดังต่อไปนี้

1. ข้อมูลที่ใช้ในการทดสอบเป็นข้อมูลลำดับเหตุการณ์หลายมิติที่สร้างขึ้นมาจากการรวมลำดับเหตุการณ์และสารสนเทศหลายมิติเข้าด้วยกัน ซึ่งลำดับเหตุการณ์สร้างจากโปรแกรมสังเคราะห์ข้อมูลของไอบีเอ็ม และสารสนเทศหลายมิติสร้างขึ้นมาโดยการสุ่มตัวเลขในแต่ละมิติ
2. ข้อมูลที่ใช้ในการทดสอบเป็นข้อมูลที่เก็บในรูปแบบของไฟล์
3. ข้อมูลลำดับเหตุการณ์หลายมิติถูกนำไปทดสอบกับสองโปรแกรม ซึ่งมีลักษณะการทำงานที่ตรงข้ามกัน คือ โปรแกรมแรกทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อน แล้วสืบค้น

รูปแบบลำดับเหตุการณ์แบบปิดที่สัมพันธ์ โปรแกรมที่สอง ทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อน แล้วสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์

1.5 ส่วนประกอบของวิทยานิพนธ์

ส่วนที่เหลือของวิทยานิพนธ์ฉบับนี้ประกอบด้วยบทต่างๆ ดังนี้

บทที่ 2 กล่าวถึงนิยามและทฤษฎีบทพื้นฐานพร้อมทั้งนำเสนองานวิจัยที่เกี่ยวข้อง โดยแบ่งออกเป็น 3 ส่วน คือ ส่วนแรกเป็นการนำเสนอการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด ส่วนที่สองเป็นการนำเสนอการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด และปิดท้ายด้วยการนำเสนอการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ

บทที่ 3 กล่าวถึงนิยามพื้นฐานและคุณสมบัติที่ใช้ในการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด และส่วนสุดท้ายเป็นการแสดงความถูกต้องของการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด

บทที่ 4 กล่าวถึงการทดลองและผลการทดลองเพื่อเปรียบเทียบผลลัพธ์ที่ได้จากการผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด โดยการพัฒนาโปรแกรมสองโปรแกรมเพื่อทดสอบกับข้อมูลที่จัดเตรียมไว้ คือ โปรแกรมแรกทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดที่สัมพันธ์ และโปรแกรมที่สองทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์

บทที่ 5 สรุปวิธีการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด สรุปผลการทดลอง และข้อเสนอแนะเกี่ยวกับการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทนี้เป็นบทที่ว่าด้วยนิยามและทฤษฎีบทพื้นฐานพร้อมทั้งนำเสนองานวิจัยที่เกี่ยวข้อง โดยแบ่งออกเป็น 3 ส่วน คือ ส่วนแรกเป็นการนำเสนอการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด ส่วนที่สองเป็นการนำเสนอการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด และปิดท้ายด้วยการนำเสนอการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ

2.1 การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด

การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด คือ การสืบค้นกลุ่มข้อมูลขนาดใหญ่ที่มีความถี่สูงหรือกลุ่มข้อมูลขนาดใหญ่ที่เกิดขึ้นบ่อย โดยรูปแบบกลุ่มข้อมูลดังกล่าวสามารถครอบคลุมรูปแบบกลุ่มข้อมูลความถี่ขนาดย่อยที่มีค่าความถี่หรือค่าสนับสนุนเท่ากันได้ [1] [6] [7] [9] [11] [12] [13]

การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดถูกเสนอขึ้นเพื่อแก้ปัญหาการสืบค้นรูปแบบกลุ่มข้อมูล ซึ่งจะให้รูปแบบกลุ่มข้อมูลจำนวนมากถ้าฐานข้อมูลมีขนาดใหญ่และค่าสนับสนุนต่ำสุดมีค่าต่ำ ดังนั้นการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดจึงทำการขจัดรูปแบบกลุ่มข้อมูลที่ซ้ำซ้อนหรือรูปแบบกลุ่มข้อมูลที่ถูกครอบคลุมโดยรูปแบบกลุ่มข้อมูลอื่น ทำให้การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดได้จำนวนรูปแบบน้อยกว่าหรือเท่ากับจำนวนรูปแบบที่ได้จากการสืบค้นรูปแบบกลุ่มข้อมูล ปัจจุบันได้มีการพัฒนาขั้นตอนวิธีสืบค้นรูปแบบกลุ่มข้อมูลหลายขั้นตอนวิธี เช่น A-CLOSE [13], CLOSET [7], CLOSET+ [9], TFP [6], CHARM [12], GRG [11] และ DCI-CLOSE [1] เป็นต้น ขั้นตอนวิธีเหล่านี้ใช้นิยามพื้นฐานและคุณสมบัติเดียวกันเพื่อใช้ในการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดแต่มีขั้นตอนที่แตกต่างกัน ซึ่งจะกล่าวต่อไปในหัวข้อที่ 2.1.1 2.1.2 และ 2.1.3 ตามลำดับ

2.1.1 นิยามพื้นฐาน

กำหนดให้ $I = \{i_1, i_2, \dots, i_n\}$ คือ เซตของอินซันข้อมูล (Item)

นิยามที่ 2.1 กลุ่มข้อมูล (Itemset) คือ ซับเซตของ I ที่ไม่ใช่เซตว่างแทนด้วย $\delta = (x_1, x_2, \dots, x_k)$ ซึ่ง k คือ จำนวนอินซันข้อมูลหรือความยาวของกลุ่มข้อมูล δ แทนด้วย k -itemset

นิยามที่ 2.2 ค่าสนับสนุน (Support) คือ จำนวนรายการเปลี่ยนแปลงที่พบข้อมูลนั้นๆ ค่าสนับสนุนของ δ แทนด้วย $\text{supp}(\delta)$

นิยามที่ 2.3 กลุ่มข้อมูล $\delta = (x_1, x_2, \dots, x_k)$ เป็นกลุ่มข้อมูลย่อย (Sub-itemset) ของกลุ่มข้อมูล $\lambda = (y_1, y_2, \dots, y_l)$ ก็ต่อเมื่อ กลุ่มข้อมูล δ เป็นซับเซตของกลุ่มข้อมูล λ และเรียกกลุ่มข้อมูล λ ว่า ซุปเปอร์กลุ่มข้อมูล (Super-itemset) เขียนแทนด้วย

$$\delta \subseteq \lambda \quad (2.1)$$

ข้อสังเกต จากนิยามที่ 2.3 ถ้าค่าสนับสนุนกลุ่มข้อมูล δ และกลุ่มข้อมูล λ มีค่าเท่ากัน แสดงว่ากลุ่มข้อมูล δ ถูกครอบคลุมโดยกลุ่มข้อมูล λ

นิยามที่ 2.4 เซตของรูปแบบกลุ่มข้อมูล (FI) ประกอบด้วยกลุ่มข้อมูลที่มีความถี่หรือค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนต่ำสุด (min_support) เขียนแทนด้วย

$$\text{FI} = \{ \delta \mid \text{supp}(\delta) \geq \text{min_support} \} \quad (2.2)$$

นิยามที่ 2.5 เซตของรูปแบบกลุ่มข้อมูลแบบปิด (CI) ประกอบด้วยรูปแบบกลุ่มข้อมูลที่ไม่มี ซุปเปอร์กลุ่มข้อมูลที่มีค่าสนับสนุนเท่ากัน เขียนแทนด้วย

$$\text{CI} = \{ \delta \mid \delta \in \text{FI} \wedge (\nexists \lambda \in \text{FI} \mid \delta \subseteq \lambda \wedge \text{supp}(\delta) = \text{supp}(\lambda)) \} \quad (2.3)$$

ข้อสังเกต จากนิยามที่ 2.5 เนื่องจาก CI เกิดจากการจัดรูปแบบกลุ่มข้อมูลที่เป็นกลุ่มข้อมูลย่อยที่มีค่าสนับสนุนเท่ากับรูปแบบกลุ่มข้อมูลใดๆ ใน FI ดังนั้น เซตของรูปแบบกลุ่มข้อมูลแบบปิดเป็นซับเซตของรูปแบบกลุ่มข้อมูล เขียนแทนด้วย

$$\text{CI} \subseteq \text{FI} \quad (2.4)$$

นิยามที่ 2.6 เซตของรูปแบบกลุ่มข้อมูลขนาดใหญ่ (MI) คือ เซตของรูปแบบกลุ่มข้อมูลที่ไม่มี ซุปเปอร์กลุ่มข้อมูล เขียนแทนด้วย

$$\text{MI} = \{ \delta \in \text{FI} \mid \nexists \delta' \in \text{FI}, \delta \subset \delta' \} \quad (2.5)$$

นิยามที่ 2.7 เซตของรูปแบบกลุ่มข้อมูลแบบปิดขนาดใหญ่ (MCI) คือ เซตของรูปแบบกลุ่มข้อมูลแบบปิดที่ไม่มีซูปเปอร์กลุ่มข้อมูล เขียนแทนด้วย

$$MCI = \{ \lambda \in CI \mid \nexists \lambda' \in CI, \lambda \subset \lambda' \} \quad (2.6)$$

ตัวอย่างที่ 2.1 การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดจากฐานข้อมูลกลุ่มข้อมูลในตารางที่ 2.1 [11] โดยกำหนดตัวเลข 1, 2, 3, 4 และ 5 แทนชั้นข้อมูลและค่าสนับสนุนต่ำสุดเท่ากับ 2

ตารางที่ 2.1 ฐานข้อมูลกลุ่มข้อมูล

ลำดับที่	กลุ่มข้อมูล		
1.	2	4	5
2.	1	3	-
3.	2	4	5
4.	1	3	5
5.	1	4	5

จากตารางที่ 2.1 พบกลุ่มข้อมูล (2, 4, 5) ในรายการเปลี่ยนแปลงลำดับที่ 1 และ 3 แสดงว่าค่าสนับสนุนของกลุ่มข้อมูล (2, 4, 5) มีค่าเท่ากับ 2 ซึ่งค่าสนับสนุนเท่ากับค่าสนับสนุนต่ำสุด ดังนั้นกลุ่มข้อมูล (2, 4, 5) เป็นรูปแบบกลุ่มข้อมูล และพบกลุ่มข้อมูล (2, 4) ในรายการเปลี่ยนแปลงลำดับที่ 1 และ 3 แสดงว่าค่าสนับสนุนของกลุ่มข้อมูล (2, 4) มีค่าเท่ากับ 2 ดังนั้นกลุ่มข้อมูล (2, 4) จึงเป็นรูปแบบกลุ่มข้อมูลเช่นกัน เนื่องจากรูปแบบกลุ่มข้อมูล (2, 4) เป็นกลุ่มข้อมูลย่อยของรูปแบบกลุ่มข้อมูล (2, 4, 5) และมีค่าสนับสนุนเท่ากัน ดังนั้นรูปแบบกลุ่มข้อมูล (2, 4) จึงถูกรอบคลุมโดยรูปแบบกลุ่มข้อมูล (2, 4, 5) ซึ่งสามารถพิจารณาหารูปแบบกลุ่มข้อมูล (2, 4) ได้จากรูปแบบกลุ่มข้อมูล (2, 4, 5) ได้ การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดจึงทำการสืบค้นเฉพาะรูปแบบกลุ่มข้อมูลขนาดใหญ่หรือรูปแบบกลุ่มข้อมูลแบบปิดที่สามารถครอบคลุมกลุ่มข้อมูลย่อยได้ ผลลัพธ์ที่ได้จากการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดจากฐานข้อมูลในตารางที่ 2.1 มีทั้งหมด 6 รูปแบบ ซึ่งเขียนในรูปแบบ (รูปแบบกลุ่มข้อมูล): ค่าสนับสนุน ได้ดังนี้ $\{(1):3, (5):4, (1, 3):2, (1, 5):2, (4, 5):3, (2, 4, 5):2\}$ ส่วนผลลัพธ์ที่ได้จากการสืบค้นรูปแบบกลุ่มข้อมูลมีทั้งหมด 11 รูปแบบ คือ $\{(1):3, (2):2, (3):2, (4):3, (5):4, (1, 3):2, (1, 5):2, (2, 4):2, (2, 5):2, (4, 5):3, (2, 4, 5):2\}$ แสดงให้เห็นว่าการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดให้จำนวนรูปแบบน้อยกว่าการสืบค้นรูปแบบกลุ่มข้อมูลโดยที่ข้อมูลไม่สูญหาย

2.1.2 คุณสมบัติของรูปแบบกลุ่มข้อมูลแบบปิด

ขั้นตอนวิธีที่ใช้ในการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด พิจารณาคุณสมบัติต่อไปนี้ในการสืบค้นรูปแบบกลุ่มข้อมูลปิด [2] [13] [14] [15]

ข้อที่ 1 กลุ่มข้อมูลย่อยทุกตัวของรูปแบบกลุ่มข้อมูลเป็นรูปแบบกลุ่มข้อมูล

ข้อที่ 2 ซุปเปอร์กลุ่มข้อมูลทุกตัวของรูปแบบที่ไม่ใช่รูปแบบกลุ่มข้อมูลจะไม่ใช่รูปแบบกลุ่มข้อมูล

ข้อที่ 3 เซตของรูปแบบกลุ่มข้อมูลขนาดใหญ่เท่ากับเซตของรูปแบบกลุ่มข้อมูลแบบปิดขนาดใหญ่

ข้อที่ 4 ค่าสนับสนุนของรูปแบบกลุ่มข้อมูล I ใดๆ เท่ากับค่าสนับสนุนของรูปแบบกลุ่มข้อมูลแบบปิดขนาดเล็กที่สุดที่บรรจุรูปแบบกลุ่มข้อมูล I

จากคุณสมบัติข้อที่ 1 3 และ 4 แสดงให้เห็นว่าเซตของรูปแบบกลุ่มข้อมูลแบบปิดและค่าสนับสนุนของรูปแบบกลุ่มข้อมูลแบบปิด สามารถสร้างเซตของรูปแบบกลุ่มข้อมูลและค่าสนับสนุนของรูปแบบกลุ่มข้อมูลทั้งหมดได้ ซึ่งแสดงให้เห็นว่าการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดสามารถครอบคลุมการสืบค้นรูปแบบกลุ่มข้อมูล [15]

2.1.3 การพัฒนาขั้นตอนวิธีสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด

การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดเป็นงานวิจัยที่มีการศึกษาอย่างกว้างขวาง เริ่มโดย Nicolas Pasquier และคณะเสนอขั้นตอนวิธีที่ชื่อว่า A-CLOSE [13] ซึ่งมีขั้นตอนโดยสังเขปดังนี้ ขั้นตอนแรกจะทำการค้นหากกลุ่มข้อมูลความถี่โดยใช้พื้นฐานของ Apriori-like และตัดตัวที่ไม่ใช่กลุ่มข้อมูลขนาดเล็กของแต่ละคลาสสมมูลออก เพราะว่า k -itemset เป็นคำตอบก็ต่อเมื่อไม่มี $(k-1)$ -subset ที่มีค่าสนับสนุนเท่ากัน ขั้นตอนที่สอง ทำการคำนวณหาส่วนปิดคลุมของทุกกลุ่มข้อมูลขนาดเล็กที่ค้นพบก่อนหน้านั้น วิธีการนี้อาจเกิดความซ้ำซ้อนในการคำนวณหาส่วนปิดคลุม เพราะว่าในแต่ละคลาสสมมูลอาจจะมีกลุ่มข้อมูลขนาดเล็กมากกว่า 1 ตัว ทำให้เสียเวลาในการคำนวณหาส่วนปิดคลุมและค้นหาจำนวนซับเซตที่มีจำนวนมาก

Jiawei Han และคณะได้นำเสนอ CLOSET [7] และ CLOSET+ [9] เป็นการคำนวณหา รูปแบบกลุ่มข้อมูลแบบปิดที่เกิดขึ้นพร้อมกันบ่อย โดยจะทำการกราดตรวจฐานข้อมูลครั้งแรกเพื่อหากกลุ่มข้อมูลความถี่ความยาว 1 ชั้น และตัดชั้นข้อมูลที่ไม่ใช่กลุ่มข้อมูลความถี่ความยาว 1 ชั้นออก ซึ่งแสดงการสืบค้นในรูป FP-Tree ในหน่วยความจำ โดยการสืบค้นแนวลึก (Depth-first Search) และการเรียกซ้ำในการท่องไปตาม FP-Tree ขั้นตอนวิธี CLOSET เป็นขั้นตอนวิธีที่หากกลุ่มข้อมูลแบบปิดโดยการปีนป่ายส่วนปิดคลุม (Closure Climbing) และใช้การสืบค้นแนวลึกเป็นพื้นฐานเพื่อหยุดการขยายของกลุ่มข้อมูลเดิมนั้น (Prefix Item) โดยการพิจารณาชุดข้อมูลนั้นว่าถูกครอบคลุม

โดยรูปแบบกลุ่มข้อมูลแบบปิดหรือเปล่า ดังนั้นต้องเก็บรูปแบบกลุ่มข้อมูลแบบปิดทุกตัวลงในหน่วยความจำ ส่วนขั้นตอนวิธี CLOSET+ คล้ายกับ CLOSET แต่ใช้ดัดแปลงต้นไม้เป็น 2 ระดับเพื่อให้เหมาะกับชุดข้อมูลที่เป็นแบบกระจายกับแบบหนาแน่น

Petre Tzvetkov และคณะได้เสนอขั้นตอนวิธีที่เรียกว่า TFP [6] โดยพยายามที่จะลดพื้นที่ในการเก็บกลุ่มข้อมูลแบบปิดโดยเก็บอยู่ในรูปของต้นไม้กลุ่มข้อมูลเต็มหน้า และใช้ดัชนีแฮช 2 ระดับช่วยลดพื้นที่ในการค้นหา

Mohammed J. Zaki และคณะได้พัฒนาขั้นตอนวิธีที่ชื่อว่า CHARM [12] ซึ่งใช้การสืบค้นแนวลึกเป็นตัวแทนไปตามต้นไม้กลุ่มข้อมูลเต็มหน้าที่เกิดจากกลุ่มข้อมูลความถี่ แล้วจะทำการเปรียบเทียบลำดับที่ของกลุ่มข้อมูลความถี่ที่มีโหนดแม่เดียวกัน ถ้ามีลำดับที่เท่ากันจะทำการรวมโหนดทั้งสองเข้าด้วยกัน เพราะแน่ใจได้ว่ากลุ่มข้อมูลทั้งสองอยู่ในคลาสสมมูลเดียวกัน ซึ่งขั้นตอนวิธี CHARM ลำดับที่ของกลุ่มข้อมูลจะต้องถูกเก็บลงในแต่ละโหนด เพราะว่าแต่ละเส้นทางสามารถนำไปสู่รูปแบบกลุ่มข้อมูลแบบปิดเหมือนกันซึ่งทำให้เกิดความซ้ำซ้อน และขั้นตอนวิธีนี้ยังต้องเก็บดัชนีที่ชี้ไปยังรูปแบบกลุ่มข้อมูลแบบปิดลงในหน่วยความจำโดยใช้ตารางแฮช

LiLi และคณะเสนอขั้นตอนวิธีที่ชื่อว่า GRG [11] โดยใช้กราฟเพื่อแสดงความสัมพันธ์ระหว่างกลุ่มข้อมูล ซึ่งใช้บิตเวกเตอร์แทนกลุ่มข้อมูล ขั้นตอนวิธีดังกล่าวสามารถลดจำนวนรอบในการอ่านฐานข้อมูล หลีกเลี่ยงการสร้างกลุ่มข้อมูลแบบปิดคู่แข่ง การแทนกลุ่มข้อมูลด้วยบิตเวกเตอร์อาจทำให้เสียเวลาในการแทนค่า ถ้าในกรณีที่กลุ่มข้อมูลมีขนาดใหญ่

Claudio Lucchese และคณะได้เสนอ DCI-CLOSE [1] เป็นขั้นตอนวิธีสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดโดยปราศจากความซ้ำซ้อน ซึ่งขั้นตอนวิธีดังกล่าวสามารถค้นหารูปแบบกลุ่มข้อมูลแบบปิดและตัดกลุ่มข้อมูลที่น่าไปสู่กลุ่มข้อมูลแบบปิดตัวเดียวกันออก ไม่ต้องเก็บกลุ่มข้อมูลแบบปิดนั้นลงในหน่วยความจำ โดยอาศัยการสืบค้นแนวลึกและการเรียกซ้ำในการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด

2.2 การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด

การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด คือ การสืบค้นลำดับเหตุการณ์ขนาดใหญ่ที่มีความถี่สูงหรือลำดับเหตุการณ์ขนาดใหญ่ที่เกิดขึ้นบ่อย โดยรูปแบบลำดับเหตุการณ์ดังกล่าวสามารถครอบคลุมรูปแบบลำดับเหตุการณ์ความถี่ขนาดย่อยที่มีค่าความถี่หรือค่าสนับสนุนเท่ากัน

ได้ การสืบค้นลำดับเหตุการณ์แบบปิดจะพิจารณาถึงลำดับการเกิดของข้อมูลและการเกิดพร้อมกันของข้อมูลด้วย [8] [16] [17] [19]

การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดถูกเสนอขึ้นมาเพื่อแก้ปัญหการสืบค้นรูปแบบลำดับเหตุการณ์ ซึ่งจะสร้างรูปแบบลำดับเหตุการณ์จำนวนมากถ้าฐานข้อมูลมีขนาดใหญ่และค่าสนับสนุนมีค่าต่ำ ดังนั้นการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดจึงทำการขจัดรูปแบบลำดับเหตุการณ์ที่ซ้ำซ้อนหรือรูปแบบที่ถูกครอบคลุมโดยรูปแบบลำดับเหตุการณ์อื่นออกไป ซึ่งทำให้การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดให้จำนวนรูปแบบน้อยกว่าหรือเท่ากับจำนวนรูปแบบที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์ ดังนั้นจึงมีผู้พัฒนาขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดหลายขั้นตอนวิธี เช่น CloSpan [19], TSP [16] และ BIDE [8] เป็นต้น ขั้นตอนวิธีเหล่านี้มีนิยามพื้นฐานเดียวกันแต่มีขั้นตอนที่แตกต่างกัน ซึ่งจะกล่าวในหัวข้อที่ 2.2.1 และ 2.2.2 ตามลำดับ

2.2.1 นิยามพื้นฐาน

นิยามที่ 2.8 $\alpha = \langle a_1 a_2 \dots a_n \rangle$ คือ ลำดับเหตุการณ์ ซึ่ง $a_i \subseteq I$ โดยที่ I คือ เซตของชั้นข้อมูล และ n คือ จำนวนชั้นข้อมูลหรือความยาวของลำดับเหตุการณ์ α แทนด้วย n -sequence

นิยามที่ 2.9 ลำดับเหตุการณ์ $\alpha = (a_1 a_2 \dots a_n)$ เป็นลำดับเหตุการณ์ย่อย (Sub-sequence) ของลำดับเหตุการณ์ $\beta = (b_1 b_2 \dots b_m)$ ก็ต่อเมื่อมีจำนวนเต็มบางตัว i_1, i_2, \dots, i_n ซึ่ง $1 \leq i_1 < i_2 < \dots < i_n \leq m$ และ $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \dots, a_n \subseteq b_{i_n}$ เรียกลำดับเหตุการณ์ β ว่าซูปเปอร์ลำดับเหตุการณ์ (Super-sequence) เขียนแทนด้วย

$$\alpha \sqsubseteq \beta \quad (2.7)$$

ข้อสังเกต จากนิยามที่ 2.9 ถ้าค่าสนับสนุนของลำดับเหตุการณ์ α และลำดับเหตุการณ์ β มีค่าเท่ากัน แสดงว่าลำดับเหตุการณ์ α ถูกครอบคลุมโดยลำดับเหตุการณ์ β

นิยามที่ 2.10 เซตของรูปแบบลำดับเหตุการณ์ (FS) ประกอบด้วยลำดับเหตุการณ์ที่มีความถี่หรือค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนต่ำสุด เขียนแทนด้วย

$$FS = \{ \alpha \mid \text{supp}(\alpha) \geq \text{min_support} \} \quad (2.8)$$

นิยามที่ 2.11 เซตของรูปแบบลำดับเหตุการณ์แบบปิด (CS) ประกอบด้วยรูปแบบลำดับเหตุการณ์ที่ไม่มีซูปเปอร์ลำดับเหตุการณ์ที่มีค่าสนับสนุนเท่ากัน เขียนแทนด้วย

$$CS = \{ \alpha \mid \alpha \in FS \wedge (\nexists \beta \in FS \mid \alpha \sqsubseteq \beta \wedge \text{supp}(\alpha) = \text{supp}(\beta)) \} \quad (2.9)$$

ข้อสังเกต จากนิยามที่ 2.11 เนื่องจาก CS เกิดจากการขจัดลำดับเหตุการณ์ที่เป็นลำดับเหตุการณ์ย่อยที่มีค่าสนับสนุนเท่ากับรูปแบบลำดับเหตุการณ์ใดๆ ใน FS ดังนั้นเซตของรูปแบบลำดับเหตุการณ์แบบปิดเป็นซัพเซตของรูปแบบลำดับเหตุการณ์ เขียนแทนด้วย

$$CS \subseteq FS \quad (2.10)$$

ตัวอย่าง 2.2 การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดจากฐานข้อมูลลำดับเหตุการณ์ในตารางที่ 2.2 [19] กำหนดตัวอักษร a, b, d, e และ f แทนชิ้นข้อมูลในลำดับเหตุการณ์ และค่าสนับสนุนต่ำสุดเท่ากับ 2

ตารางที่ 2.2 ฐานข้อมูลลำดับเหตุการณ์

ลำดับที่	ลำดับเหตุการณ์
1.	< (af) d e a >
2.	< e a b >
3.	< e (abf) (bde) >

จากตารางที่ 2.2 ลำดับเหตุการณ์ <(af) d e a > หมายความว่า ลำดับเหตุการณ์ a และ f เกิดขึ้นพร้อมกัน จากนั้นเกิดลำดับเหตุการณ์ d e และ a ตามลำดับ เมื่อทำการสืบค้นจากฐานข้อมูลพบลำดับเหตุการณ์ <(af)d > ในรายการเปลี่ยนแปลงลำดับที่ 1 และ 3 ดังนั้นค่าสนับสนุนมีค่าเท่ากับ 2 ซึ่งมากกว่าหรือเท่ากับค่าสนับสนุนต่ำสุด แสดงว่าลำดับเหตุการณ์ <(af)d > เป็นรูปแบบลำดับเหตุการณ์ และพบลำดับเหตุการณ์ <(af)> ในรายการเปลี่ยนแปลงลำดับที่ 1 และ 3 ดังนั้นค่าสนับสนุนมีค่าเท่ากับ 2 แสดงว่าลำดับเหตุการณ์ <(af)> เป็นรูปแบบลำดับเหตุการณ์เช่นกัน เนื่องจากรูปแบบลำดับเหตุการณ์ <(af)> เป็นลำดับเหตุการณ์ย่อยของรูปแบบลำดับเหตุการณ์ <(af)d > และมีค่าสนับสนุนเท่ากัน แสดงว่ารูปแบบลำดับเหตุการณ์ <(af)> ถูกครอบคลุมโดยรูปแบบลำดับเหตุการณ์ <(af)d > ซึ่งสามารถพิจารณารูปแบบลำดับเหตุการณ์ <(af)> จากรูปแบบลำดับเหตุการณ์ <(af)d > ได้ การสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดจึงทำการสืบค้นเฉพาะรูปแบบลำดับเหตุการณ์ที่มีขนาดใหญ่หรือรูปแบบลำดับเหตุการณ์แบบปิดที่ครอบคลุมรูปแบบ

ลำดับเหตุการณ์ย่อยได้ ผลลัพธ์ที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดจากฐานข้อมูลในตารางที่ 2.2 มีทั้งหมด 4 รูปแบบ คือ $\{ \langle (af)d \rangle:2, \langle (af)e \rangle:2, \langle ea \rangle:3, \langle eab \rangle:2 \}$ แต่ผลลัพธ์ที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์มีทั้งหมด 16 รูปแบบ คือ $\{ \langle a \rangle:3, \langle b \rangle:2, \langle d \rangle:2, \langle e \rangle:3, \langle f \rangle:2, \langle ad \rangle:2, \langle ae \rangle:2, \langle ab \rangle:2, \langle fd \rangle:2, \langle fe \rangle:2, \langle ea \rangle:3, \langle eb \rangle:2, \langle af \rangle:2, \langle (af)d \rangle:2, \langle (af)e \rangle:2, \langle eab \rangle:2 \}$ จะเห็นได้ว่าการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดให้จำนวนรูปแบบน้อยกว่าการสืบค้นรูปแบบลำดับเหตุการณ์โดยข้อมูลไม่สูญหาย

2.2.2 การพัฒนาขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด

Xifeng Yan และคณะได้นำเสนอขั้นตอนวิธีที่เรียกว่า CloSpan [19] โดยสร้างเซตของรูปแบบลำดับเหตุการณ์แบบปิดคู่แข่ง (Candidate Closed Sequential Pattern) แล้วเก็บลงในโครงสร้างต้นไม้ดัชนีแฮช หลังจากนั้นจะตัดตัวที่ไม่ใช่ลำดับเหตุการณ์แบบปิดออก เพราะว่า CloSpan จำเป็นต้องเก็บเซตของลำดับเหตุการณ์แบบปิดคู่แข่งไว้ จึงอาจจะทำให้เปลืองหน่วยความจำมากและอาจทำให้เสียเวลาในการค้นหาส่วนปิดคลุม เมื่อมีลำดับเหตุการณ์แบบปิดคู่แข่งจำนวนมาก

Petre Tzvetkov และคณะเสนอขั้นตอนวิธี TSP [16] ทำการสืบค้นลำดับเหตุการณ์แบบปิดโดยไม่ต้องผ่านค่าสนับสนุน เป็นขั้นตอนวิธีที่ช่วยแก้ปัญหาในกรณีที่ค่าสนับสนุนมีค่าน้อยมากจนกระทั่งไม่สามารถสืบค้นลำดับเหตุการณ์ได้ โดยการเปลี่ยนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดเป็นการสืบค้นจำนวนรูปแบบที่มีความยาวไม่น้อยกว่าความยาวต่ำสุดที่กำหนดไว้ โดย TSP จะตัดแปลง เป็นการค้นหาหลายส่วนพร้อมกันเพื่อให้ง่ายต่อการสืบค้นรูปแบบและเพิ่มค่าสนับสนุน และใช้วิธีการที่เชื่อว่าการพิสูจน์รูปแบบแบบปิดในระหว่างที่ทำการสืบค้น เพื่อให้แน่ใจว่าการเพิ่มค่าสนับสนุนเป็นไปอย่างถูกต้องและสามารถสืบค้นคำตอบทั้งหมดได้ นอกจากนี้ยังใช้เทคนิคหลายเทคนิค เช่น การประยุกต์การกำหนดความยาวของรูปแบบและใช้เทคนิคในลักษณะเดียวกับ CloSpan

Jainyong Wang และคณะนำเสนอขั้นตอนวิธี BIDE [8] เป็นขั้นตอนวิธีแรกที่สืบค้นรูปแบบลำดับเหตุการณ์แบบปิดที่ไม่จำเป็นต้องเก็บรูปแบบลำดับเหตุการณ์แบบปิดคู่แข่ง โดยใช้เทคนิคการตรวจสอบส่วนปิดคลุมแบบสองทิศทางเพื่อหาลำดับเหตุการณ์แบบปิด และทำการตัดเส้นทางที่ไม่สามารถนำไปสู่ลำดับเหตุการณ์แบบปิดโดยการหยุดขยายลำดับเหตุการณ์เดิมนั้น นอกจากนี้ยังใช้เทคนิคข้ามการกราดตรวจเพื่อให้การคำนวณเร็วขึ้น ผลลัพธ์ทั้งหมดที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดโดยใช้ขั้นตอนวิธี BIDE ใช้หน่วยความจำน้อย และใช้เวลาในการคำนวณน้อยเมื่อค่าสนับสนุนมีค่าต่ำและลำดับเหตุการณ์มีขนาดยาวมากๆ

2.3 การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ

การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ คือ กระบวนการสืบค้นสารสนเทศที่มี 1 มิติ หรือมากกว่า 1 มิติที่ไม่คำนึงถึงลำดับของมิติ และสารสนเทศ 1 มิติที่คำนึงถึงลำดับ ซึ่งสารสนเทศที่มี 1 มิติหรือมากกว่า 1 มิติที่ไม่คำนึงถึงลำดับของมิติ เรียกว่า สารสนเทศหลายมิติ และสารสนเทศ 1 มิติที่คำนึงถึงลำดับ คือ ลำดับเหตุการณ์ การสร้างความสัมพันธ์ระหว่างการสืบค้นรูปแบบลำดับเหตุการณ์กับสารสนเทศหลายมิติในรูปของการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ ทำให้ได้รูปแบบที่มีประโยชน์มากกว่าการสืบค้นรูปแบบลำดับเหตุการณ์อย่างเดียว เช่นรูปแบบลำดับเหตุการณ์หลายมิติ (Business, Boston, middle, <(Arrow suit, Dapper tie) > Chap shoes) รูปแบบลำดับเหตุการณ์หลายมิติดังกล่าวทำให้รู้ว่า กลุ่มลูกค้าที่ทำการเกี่ยวกับธุรกิจ อาศัยอยู่ในเมือง Boston วิทยกลางคนจะชอบซื้อชุดสูทที่ห้อย Arrow และเนคไทที่ห้อย Dapper พร้อมกัน แล้วต่อมาจะซื้อรองเท้าที่ห้อย Chap ซึ่งทำให้ร้านสามารถแนะนำสินค้าให้ตรงกับความต้องการของลูกค้าได้

การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติถูกนำเสนอขึ้นครั้งแรกโดย Helen Pinto ซึ่งได้นำเสนอ 2 ขั้นตอนวิธีที่มีชื่อว่า PSFP และ HYBRID [3] ซึ่งจะกล่าวถึงนิยามพื้นฐานและขั้นตอนวิธีของทั้งสองขั้นตอนวิธีในหัวข้อที่ 2.3.1 และ 2.3.2 ตามลำดับ

2.3.1 นิยามพื้นฐาน

กำหนดให้ $D = (TID, X_1, X_2, \dots, X_m, S)$ เป็นฐานข้อมูลลำดับเหตุการณ์หลายมิติ ซึ่งในแต่ละแถว $t = \{tid, x_1, x_2, \dots, x_m, s_i\}$ ประกอบไปด้วย

1. tid : คือตัวระบุชุดข้อมูลหรือลำดับที่
2. x_1, x_2, \dots, x_m : คือสารสนเทศหลายมิติ
3. s_i : คือ ลำดับเหตุการณ์

นิยามที่ 2.12 ลำดับเหตุการณ์หลายมิติแสดงอยู่ในรูป $p = (y_1, y_2, \dots, y_m, s)$ ซึ่ง $y_i \in (X_i \cup \{*\})$ สำหรับ $1 \leq i \leq m$ s คือ ลำดับเหตุการณ์ และ $*$ แทนข้อมูลใดๆที่อยู่ใน X_1, \dots, X_m

นิยามที่ 2.13 ลำดับเหตุการณ์หลายมิติ $p = (a_1, a_2, \dots, a_m, s)$ ตรงกับแถว $t = (x_1, x_2, \dots, x_m, s_i)$ ในฐานข้อมูลลำดับเหตุการณ์หลายมิติ ก็ต่อเมื่อ $1 \leq i \leq m$ ซึ่ง $a_i = x_i$ หรือ $a_i = *$ และ $s \subseteq s_i$ จำนวนแถวในฐานข้อมูลที่ตรงกับลำดับเหตุการณ์หลายมิติ p เรียกว่า ค่าสนับสนุนของ p แทนด้วย $\text{supp}(p)$

นิยามที่ 2.14 เซตของรูปแบบลำดับเหตุการณ์หลายมิติ (MD) ประกอบด้วยลำดับเหตุการณ์หลายมิติที่มีความถี่หรือค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนต่ำสุด เขียนแทนด้วย

$$MD = \{p \mid \text{supp}(p) \geq \text{min_support}\} \quad (2.11)$$

ตัวอย่างที่ 2.3 การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติจากฐานข้อมูลการซื้อสินค้าในร้านแห่งหนึ่งดังแสดงในตารางที่ 2.3 [3] ซึ่งฐานข้อมูลประกอบด้วยสารสนเทศหลายมิติจำนวน 3 มิติ คือ กลุ่มลูกค้า เมือง กลุ่มอายุ และลำดับเหตุการณ์การซื้อสินค้า โดยกำหนดให้ a, b, c, d, e, f, g และ h แทนสินค้าแต่ละชิ้น และกำหนดค่าสนับสนุนต่ำสุดเท่ากับ 2

ตารางที่ 2.3 ตัวอย่างฐานข้อมูลการซื้อสินค้า

ลำดับที่	กลุ่มลูกค้า	เมือง	กลุ่มอายุ	ลำดับเหตุการณ์
1.	business	Boston	Middle	$\langle (b d) c b a c \rangle$
2.	professional	Chicago	Young	$\langle (b f) (c e) (f g) \rangle$
3.	business	Chicago	Middle	$\langle (a h) b a f \rangle$
4.	education	New York	Retired	$\langle (b e) (c e) \rangle$

จากตารางที่ 2.3 จะเห็นได้ว่าลำดับเหตุการณ์หลายมิติ (business, *, Middle, $\langle ba \rangle$) ตรงกับแถวลำดับที่ 1 และ 3 ค่าสนับสนุนของลำดับเหตุการณ์ดังกล่าวเท่ากับ 2 ซึ่งมีค่าเท่ากับค่าสนับสนุนต่ำสุด ดังนั้นลำดับเหตุการณ์หลายมิติ (business, *, Middle, $\langle ba \rangle$) เป็นรูปแบบลำดับเหตุการณ์หลายมิติ

2.3.2 การพัฒนาขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ

Helen Pinto และคณะได้เสนอขั้นตอนวิธีที่ใช้เพื่อแก้ปัญหการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติสองขั้นตอนวิธี คือ PSFP และ HYBRID ซึ่งมีลักษณะการทำงานที่ตรงข้ามกัน ดังนี้

1. ขั้นตอนวิธี PSFP เกิดจากการรวมสองขั้นตอนวิธีคือ PrefixSpan [4] และ FP-growth [5] ซึ่งขั้นตอนวิธี PrefixSpan เป็นขั้นตอนวิธีที่ใช้ในการสืบค้นรูปแบบลำดับเหตุการณ์ และขั้นตอนวิธี FP-growth เป็นขั้นตอนวิธีที่ใช้ในการสืบค้นรูปแบบกลุ่มข้อมูล ขั้นตอนวิธี PSFP จะทำการสืบค้นรูปแบบลำดับเหตุการณ์ก่อนโดยใช้ขั้นตอนวิธี PrefixSpan จากนั้นทำการสืบค้นสารสนเทศ

หลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์โดยใช้ขั้นตอนวิธี FP-growth ซึ่งวิธีการนี้จำนวนรอบในการเรียกขั้นตอนวิธี PrefixSpan น้อย แต่จำนวนรอบในการเรียกขั้นตอนวิธี FP-growth มาก วิธีการนี้ทำให้สิ้นเปลืองในการสร้างโครงข่ายข้อมูลมาก และไม่สามารถขยายเป็นโครงข่ายสองระดับได้ เพราะจำเป็นต้องมีการเก็บ FP-tree สำหรับแต่ละรูปแบบและจะไม่มีประสิทธิภาพถ้าจำนวนมิติสูง

2. ขั้นตอนวิธี HYBRID เกิดจากการรวมสองขั้นตอนวิธีคือ PrefixSpan และ BUC [10] ซึ่งขั้นตอนวิธี BUC เป็นขั้นตอนวิธีการท่องแบบลูกบาศก์ (Cube Traversal) ผ่านรูปแบบที่สัมพันธ์กัน ขั้นตอนวิธี HYBRID จะทำการสืบค้นสารสนเทศหลายมิติก่อนโดยใช้ขั้นตอนวิธี BUC จากนั้นทำการสืบค้นรูปแบบลำดับเหตุการณ์ที่สัมพันธ์กับสารสนเทศหลายมิติโดยใช้ขั้นตอนวิธี PrefixSpan วิธีการนี้จำนวนรอบในการเรียกขั้นตอนวิธี BUC น้อย แต่จำนวนรอบในการเรียกขั้นตอนวิธี PrefixSpan มาก

ขั้นตอนวิธีทั้ง 2 เริ่มต้นด้วยการหารูปแบบที่มีขนาดเล็กและทำการขยายรูปแบบไปจนถึงรูปแบบขนาดใหญ่ ถ้าฐานข้อมูลประกอบไปด้วยสารสนเทศหลายมิติที่มีจำนวนมิติสูงและลำดับเหตุการณ์มีขนาดยาว หรือค่าสนับสนุนต่ำสุดมีค่าต่ำ จะทำให้ได้รูปแบบลำดับเหตุการณ์หลายมิติจำนวนมาก ซึ่งรูปแบบเหล่านั้นเป็นรูปแบบที่ซ้ำซ้อน งานวิจัยนี้จึงได้เสนอการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดเพื่อลดจำนวนรูปแบบโดยที่ข้อมูลไม่สูญหาย

บทที่ 3

การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด

จากการศึกษางานวิจัยในบทที่ 2 ทำให้ทราบว่าสามารถผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลกับการสืบค้นรูปแบบลำดับเหตุการณ์เพื่อใช้ในการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ นอกจากนี้ยังทราบว่า การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดให้จำนวนรูปแบบน้อยกว่าหรือเท่ากับ การสืบค้นรูปแบบกลุ่มข้อมูล และการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดให้จำนวนรูปแบบน้อยกว่าหรือเท่ากับ การสืบค้นรูปแบบลำดับเหตุการณ์ ดังนั้นจึงได้นำเสนอการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดที่ผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดกับการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดเพื่อลดจำนวนรูปแบบลำดับเหตุการณ์หลายมิติ โดยในบทนี้จะแบ่งเนื้อหาออกเป็น 4 ส่วน คือ ส่วนแรกและส่วนที่ 2 กล่าวถึงนิยามพื้นฐานและคุณสมบัติที่ใช้ในการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ส่วนที่ 3 ขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด และส่วนสุดท้ายเป็นการแสดงความถูกต้องของการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด

3.1 นิยามพื้นฐาน

เพื่อให้เข้าใจการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด จึงได้สร้างนิยามพื้นฐานขึ้นมาดังนี้

กำหนดให้ I_m แทนสารสนเทศหลายมิติ x_1, x_2, \dots, x_m และ s แทนลำดับเหตุการณ์

นิยามที่ 3.1 ลำดับเหตุการณ์หลายมิติ $md = (I_m, s_1)$ เป็นลำดับเหตุการณ์หลายมิติย่อย (Sub-multidimensional Sequence) ของลำดับเหตุการณ์หลายมิติ $md' = (I_n, s_k)$ ถ้าสอดคล้องกับเงื่อนไขใดเงื่อนไขหนึ่งต่อไปนี้

$$1. s_1 \subseteq s_k \text{ และ } I_m = I_n$$

$$2. I_m \subseteq I_n \text{ และ } s_1 = s_k$$

$$3. s_1 \subseteq s_k \text{ และ } I_m \subseteq I_n$$

เรียก md' ว่า ซุปเปอร์ลำดับเหตุการณ์หลายมิติ (Super-multidimensional Sequence) ของ md สามารถเขียนแทนความสัมพันธ์ของลำดับเหตุการณ์หลายมิติ md และลำดับเหตุการณ์หลายมิติ md' ได้ดังนี้

$$md \subseteq md' \tag{3.1}$$

ข้อสังเกต ถ้าค่าสนับสนุนของลำดับเหตุการณ์หลายมิติ md และลำดับเหตุการณ์หลายมิติ md' มีค่าเท่ากัน แสดงว่าลำดับเหตุการณ์หลายมิติ md ถูกครอบคลุมโดยลำดับเหตุการณ์หลายมิติ md'

ตัวอย่างที่ 3.1 พิจารณาตารางที่ 3.1 [3] กำหนดให้ตัวอักษร $a - h$ แทนข้อมูลที่เป็นลำดับเหตุการณ์ ส่วนตัวเลข 1 - 3 ในมิติที่ 1 ตัวเลข 4 - 6 ในมิติที่ 2 และตัวเลข 7 - 9 ในมิติที่ 3 แทนสารสนเทศหลายมิติ และกำหนดให้ค่าสนับสนุนต่ำสุดมีค่าเท่ากับ 2

ตารางที่ 3.1 ฐานข้อมูลลำดับเหตุการณ์หลายมิติ

ลำดับที่	มิติที่ 1	มิติที่ 2	มิติที่ 3	ลำดับเหตุการณ์
1.	1	5	8	$\langle (bd)cb(ac) \rangle$
2.	2	6	7	$\langle (bf)(ce)b(fg) \rangle$
3.	1	6	8	$\langle (ah)(bf)abf \rangle$
4.	3	4	9	$\langle (be)(ce)d \rangle$
5.	2	6	8	$\langle a(bd)bc b(ade) \rangle$

จากตารางที่ 3.1 จะเห็นได้ว่าลำดับเหตุการณ์หลายมิติ $\langle aa \rangle, *, 6, *$ เป็นลำดับเหตุการณ์หลายมิตีย่อยของลำดับเหตุการณ์หลายมิติ $\langle aba \rangle, *, 6, 8$ เนื่องจากเงื่อนไขในข้อที่ 3 ในนิยามที่ 3.1 คือ $\langle aa \rangle \sqsubseteq \langle aba \rangle$ และ $(6) \subseteq (6, 8)$ และเรียกลำดับเหตุการณ์หลายมิติ $\langle aba \rangle, *, 6, 8$ ว่าเป็นซูปเปอร์ลำดับเหตุการณ์หลายมิติของลำดับเหตุการณ์หลายมิติ $\langle aa \rangle, *, 6, *$

นิยามที่ 3.2 ลำดับเหตุการณ์ย่อยทั้งหมดของลำดับเหตุการณ์ s เขียนแทนด้วย

$$P(s) = \{s' \mid s' \sqsubseteq s\} - \emptyset \quad (3.2)$$

ตัวอย่างที่ 3.2 ลำดับเหตุการณ์ย่อยทั้งหมดของลำดับเหตุการณ์ $\langle cba \rangle$ มีทั้งหมด 7 รูปแบบ คือ $\{\langle c \rangle, \langle b \rangle, \langle a \rangle, \langle cb \rangle, \langle ca \rangle, \langle ba \rangle, \langle cba \rangle\}$

นิยามที่ 3.3 สารสนเทศหลายมิตีย่อยทั้งหมดของสารสนเทศหลายมิติ I เขียนแทนด้วย

$$P(I) = \{I' \mid I' \subseteq I\} - \emptyset \quad (3.3)$$

ตัวอย่างที่ 3.3 สารสนเทศหลายมิติย่อยของสารสนเทศหลายมิติ $(1, 2, 8)$ มีทั้งหมด 7 รูปแบบ คือ $\{(1), (2), (8), (1, 2), (1, 8), (2, 8), (1, 2, 8)\}$

นิยามที่ 3.4 ลำดับเหตุการณ์หลายมิติย่อยทุกตัวของลำดับเหตุการณ์หลายมิติ md เขียนแทนด้วย

$$P(md) = \{P(s) \times P(I)\} \quad (3.4)$$

ตัวอย่างที่ 3.4 ลำดับเหตุการณ์หลายมิติย่อยของลำดับเหตุการณ์ $(\langle ba \rangle, 1, *, 8)$ มีทั้งหมด 9 รูปแบบ ดังนี้ $\{(\langle b \rangle, 1, *, *), (\langle b \rangle, *, *, 8), (\langle b \rangle, 1, *, 8), (\langle a \rangle, 1, *, *), (\langle a \rangle, *, *, 8), (\langle a \rangle, 1, *, 8), (\langle ba \rangle, 1, *, *), (\langle ba \rangle, *, *, 8), (\langle ba \rangle, 1, *, 8)\}$

นิยามที่ 3.5 เซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด (CMD) ประกอบด้วยรูปแบบลำดับเหตุการณ์หลายมิติ (MD) ที่ไม่มีซูปเปอร์ลำดับเหตุการณ์หลายมิติที่มีค่าสนับสนุนเท่ากัน เขียนแทนด้วย

$$CMD = \{md \mid md \in MD \wedge (\nexists md' \in MD \mid md \subseteq md' \wedge \text{supp}(md) = \text{supp}(md'))\} \quad (3.5)$$

ตัวอย่างที่ 3.5 จากตารางที่ 3.1 พบลำดับเหตุการณ์หลายมิติ $(\langle ab \rangle, *, 6, 8)$ และลำดับเหตุการณ์หลายมิติ $(\langle aba \rangle, *, 6, 8)$ ในรายการเปลี่ยนแปลงลำดับที่ 3 และ 5 แสดงว่าลำดับเหตุการณ์หลายมิติทั้งสองเป็นรูปแบบลำดับเหตุการณ์หลายมิติ เพราะมีค่าสนับสนุนเท่ากับค่าสนับสนุนต่ำสุด เมื่อทำการพิจารณารูปแบบลำดับเหตุการณ์หลายมิติ $(\langle ab \rangle, *, 6, 8)$ พบว่ารูปแบบดังกล่าวเป็นลำดับเหตุการณ์หลายมิติย่อยของรูปแบบลำดับเหตุการณ์หลายมิติ $(\langle aba \rangle, *, 6, 8)$ และทั้งสองรูปแบบมีค่าสนับสนุนเท่ากัน ดังนั้นรูปแบบลำดับเหตุการณ์หลายมิติ $(\langle ab \rangle, *, 6, 8)$ ไม่ใช่รูปแบบลำดับเหตุการณ์หลายมิติแบบปิด นอกจากนี้ยังพบว่าไม่มีรูปแบบลำดับเหตุการณ์หลายมิติใดที่เป็นซูปเปอร์ลำดับเหตุการณ์หลายมิติที่มีค่าสนับสนุนเท่ากับรูปแบบลำดับเหตุการณ์หลายมิติ $(\langle aba \rangle, *, 6, 8)$ แสดงให้เห็นว่ารูปแบบลำดับเหตุการณ์หลายมิติ $(\langle aba \rangle, *, 6, 8)$ เป็นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ผลลัพธ์ที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดจากฐานข้อมูลในตารางที่ 3.1 มีทั้งหมด 14 รูปแบบ คือ $\{(\langle aba \rangle, *, 6, 8):2, (\langle abb \rangle, *, 6, 8):2, (\langle (bd)bc \rangle, *, *, 8):2, (\langle (bd)ba \rangle, *, *, 8):2, (\langle (bf)bf \rangle, *, 6, *):2, (\langle cba \rangle, *, *, 8):2, (\langle ba \rangle, 1, *, 8):2, (\langle bcb \rangle, *, *, 8):2, (\langle bcb \rangle, 2, 6, *):2, (\langle be \rangle, 2, 6, *):2, (\langle bb \rangle, 1, *, 8):2, (\langle ba \rangle, *, *, 8):3, (\langle bb \rangle, *, 6, *):3, (\langle bb \rangle, *, *, 8):3\}$

นิยามที่ 3.6 เซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขนาดใหญ่ (MCMD) ประกอบด้วย รูปแบบลำดับเหตุการณ์หลายมิติแบบปิดที่ไม่มีซูปเปอร์ลำดับเหตุการณ์หลายมิติ เขียนแทนด้วย

$$\text{MCMD} = \{md \in \text{CMD} \mid \exists md' \in \text{CMD}, md \subset md'\} \quad (3.6)$$

ตัวอย่างที่ 3.6 จากตัวอย่างที่ 3.5 พิจารณารูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ($\langle ba \rangle, *, *, 8$) เป็นลำดับเหตุการณ์หลายมิติย่อยของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ($\langle aba \rangle, *, 6, 8$) แสดงว่ารูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ($\langle ba \rangle, *, *, 8$) ไม่ใช่รูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขนาดใหญ่ รูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ($\langle bb \rangle, *, 6, *$) และรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ($\langle bb \rangle, *, *, 8$) ไม่ใช่รูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขนาดใหญ่เช่นเดียวกัน เพราะทั้งสองรูปแบบเป็นลำดับเหตุการณ์หลายมิติย่อยของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดอื่น ดังนั้นเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขนาดใหญ่มีทั้งหมด 11 รูปแบบ คือ $\{(\langle aba \rangle, *, 6, 8), (\langle abb \rangle, *, 6, 8), (\langle bd \rangle bc, *, *, 8), (\langle bd \rangle ba, *, *, 8), (\langle bf \rangle bf, *, 6, *), (\langle cba \rangle, *, *, 8), (\langle ba \rangle, 1, *, 8), (\langle bcb \rangle, *, *, 8), (\langle bcb \rangle, 2, 6, *), (\langle be \rangle, 2, 6, *), (\langle bb \rangle, 1, *, 8)\}$

นิยามที่ 3.7 เซตของรูปแบบลำดับเหตุการณ์หลายมิติขนาดใหญ่ (MMD) ประกอบด้วย รูปแบบลำดับเหตุการณ์หลายมิติที่ไม่มีซูปเปอร์ลำดับเหตุการณ์หลายมิติ เขียนแทนด้วย

$$\text{MMD} = \{md \in \text{MD} \mid \nexists md' \in \text{MD}, md \subset md'\} \quad (3.7)$$

ตัวอย่างที่ 3.7 จากตารางที่ 3.1 เซตของรูปแบบลำดับเหตุการณ์หลายมิติขนาดใหญ่ประกอบด้วยรูปแบบ 11 รูปแบบ คือ $\{(\langle aba \rangle, *, 6, 8), (\langle abb \rangle, *, 6, 8), (\langle bd \rangle bc, *, *, 8), (\langle bd \rangle ba, *, *, 8), (\langle bf \rangle bf, *, 6, *), (\langle cba \rangle, *, *, 8), (\langle ba \rangle, 1, *, 8), (\langle bcb \rangle, *, *, 8), (\langle bcb \rangle, 2, 6, *), (\langle be \rangle, 2, 6, *), (\langle bb \rangle, 1, *, 8)\}$ ซึ่งรูปแบบเหล่านี้เป็นรูปแบบลำดับเหตุการณ์หลายมิติที่ไม่มีซูปเปอร์ลำดับเหตุการณ์หลายมิติ

3.2 คุณสมบัติของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด

จากคุณสมบัติที่ใช้ในการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดในหัวข้อ 2.1.2 เมื่อพิจารณารูปแบบลำดับเหตุการณ์หลายมิติแบบปิด $(y_1, y_2, \dots, y_m, s)$ มีลักษณะรูปแบบเช่นเดียวกับรูปแบบ

กลุ่มข้อมูลแบบปิด (i_1, i_2, \dots, i_n) โดยพิจารณารูปแบบลำดับเหตุการณ์ s เป็นชั้นข้อมูลตัวหนึ่ง เราจึงใช้คุณสมบัติเดียวกันในการพิจารณาหารูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ซึ่งสามารถเขียนคุณสมบัติดังกล่าวใหม่ได้ดังนี้

ข้อที่ 1 ลำดับเหตุการณ์หลายมิติย่อยทุกตัวของรูปแบบลำดับเหตุการณ์หลายมิติเป็นรูปแบบลำดับเหตุการณ์หลายมิติ

ตัวอย่างที่ 3.8 จากตารางที่ 3.1 รูปแบบลำดับเหตุการณ์หลายมิติ $\langle (bd)\rangle, *, *, 8\rangle:2$ ประกอบด้วยลำดับเหตุการณ์หลายมิติย่อย $\langle b\rangle, *, *, 8\rangle:3$, $\langle d\rangle, *, *, 8\rangle:2$ และ $\langle (bd)\rangle, *, *, 8\rangle:2$ ซึ่งลำดับเหตุการณ์หลายมิติย่อยทุกตัวมีความสนับสนุนมากกว่าหรือเท่ากับ 2 ดังนั้นลำดับเหตุการณ์หลายมิติย่อยทุกตัวเป็นรูปแบบลำดับเหตุการณ์หลายมิติ

ข้อที่ 2 ซุปเปอร์ลำดับเหตุการณ์หลายมิติทุกตัวของรูปแบบที่ไม่ใช่รูปแบบลำดับเหตุการณ์หลายมิติ จะไม่ใช่รูปแบบลำดับเหตุการณ์หลายมิติ

ตัวอย่างที่ 3.9 จากตารางที่ 3.1 ลำดับเหตุการณ์หลายมิติ $\langle (bd)bcb(ade)\rangle, 2, 6, *\rangle:1$ ไม่ใช่รูปแบบลำดับเหตุการณ์หลายมิติ เนื่องจากมีค่าสนับสนุนน้อยกว่าค่าสนับสนุนต่ำสุด รูปแบบดังกล่าวมีซุปเปอร์รูปแบบลำดับเหตุการณ์หลายมิติ 3 รูปแบบ คือ $\langle (bd)bcb(ade)\rangle, 2, 6, 8\rangle:1$, $\langle a(bd)bcb(ade)\rangle, 2, 6, *\rangle:1$ และ $\langle a(bd)bcb(ade)\rangle, 2, 8, 6\rangle$ ซึ่งทั้งสามรูปแบบมีค่าสนับสนุนน้อยกว่า 2 ดังนั้นไม่ใช่รูปแบบลำดับเหตุการณ์หลายมิติ

ข้อที่ 3 เซตของรูปแบบลำดับเหตุการณ์หลายมิติขนาดใหญ่เท่ากับเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขนาดใหญ่

ตัวอย่างที่ 3.10 พิจารณาจากตัวอย่างที่ 3.6 และ 3.7 จะเห็นได้ว่าเซตของรูปแบบลำดับเหตุการณ์หลายมิติขนาดใหญ่จะเท่ากับเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขนาดใหญ่

ข้อที่ 4 ค่าสนับสนุนของรูปแบบลำดับเหตุการณ์หลายมิติ md เท่ากับค่าสนับสนุนของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขนาดเล็กที่สุดที่บรรจุรูปแบบลำดับเหตุการณ์หลายมิติ md

ตัวอย่างที่ 3.11 ค่าสนับสนุนของรูปแบบลำดับเหตุการณ์หลายมิติ $\langle ba\rangle, *, 6, *\rangle$ เท่ากับ 2 มีค่าเท่ากับค่าสนับสนุนของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด $\langle aba\rangle, *, 6, 8\rangle$ ซึ่ง $\langle aba\rangle, *, 6, 8\rangle$ เป็นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขนาดเล็กที่สุดที่บรรจุรูปแบบลำดับเหตุการณ์หลายมิติ $\langle ba\rangle, *, 6, *\rangle$

3.3 ขั้นตอนวิธี

ขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดมีขั้นตอนหลัก 2 ขั้นตอน ดังนี้
 ขั้นตอนที่ 1 ทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดผสมผสานกับการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด

ขั้นตอนที่ 2 การจัดรูปแบบที่ซ้ำซ้อน

เนื่องจากผลลัพธ์ที่ได้จากขั้นตอนที่ 1 เป็นเพียงรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง (Candidate Closed Multidimensional Sequential Pattern) ซึ่งยังคงมีรูปแบบที่ซ้ำซ้อน ดังนั้นจึงต้องทำการจัดรูปแบบที่ซ้ำซ้อนดังกล่าวออกไปในขั้นตอนที่ 2

3.3.1 การผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด

การผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดสามารถทำได้ 2 วิธี ดังนี้

กำหนดให้ $CI = \{x_1, x_2, \dots, x_h\}$ แทนเซตของรูปแบบกลุ่มข้อมูลแบบปิด และกำหนดให้ $CS = \{y_1, y_2, \dots, y_k\}$ แทนเซตรูปแบบลำดับเหตุการณ์แบบปิด

วิธีที่ 1 : ทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด x บนสารสนเทศหลายมิติ แล้วทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด CS บนลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูลแบบปิด x ค่าสนับสนุนของแต่ละรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขึ้นอยู่กับค่าสนับสนุนของรูปแบบลำดับเหตุการณ์แบบปิด

$$\left. \begin{array}{l} CCMD_1 = x_1 \times CS_1 \\ CCMD_2 = x_2 \times CS_2 \\ \dots \\ \dots \\ CCMD_h = x_h \times CS_h \end{array} \right\} \text{แทนด้วย } CCMD_i$$

วิธีที่ 2 : ทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด y บนลำดับเหตุการณ์ แล้วทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิด CI บนสารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์แบบปิด y ค่าสนับสนุนของแต่ละรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดขึ้นอยู่กับค่าสนับสนุนของรูปแบบกลุ่มข้อมูลแบบปิด

$$\left. \begin{array}{l}
 \text{CCMD}_1 = y_1 \times \text{CI}_1 \\
 \text{CCMD}_2 = y_2 \times \text{CI}_2 \\
 \dots \\
 \dots \\
 \text{CCMD}_k = y_k \times \text{CI}_k
 \end{array} \right\} \text{แทนด้วย } \text{CCMD}_2$$

จากการผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดทั้งสองวิธี ผลลัพธ์ที่ได้เป็นเพียงเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง (CCMD) ซึ่งยังคงมีรูปแบบที่ซ้ำซ้อน เพราะมีบางรูปแบบเป็นลำดับเหตุการณ์หลายมิติย่อยที่มีค่าสนับสนุนเท่ากับรูปแบบอื่น ซึ่งจะแสดงให้เห็นในตัวอย่างที่ 3.12

ตัวอย่างที่ 3.12 จากตารางที่ 3.1 การผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดด้วยวิธีที่ 2 ได้ผลลัพธ์ดังนี้

<aba>:2	(6,8):2
<abb>:2	(6,8):2
<(bd)bc>:2	(8):2
<(bd)ba>:2	(8):2
<bcd>:2	-
<(bf)bf>:2	(6):2
<b(ce)>:2	-
<cba>:2	(8):2
<ba>: 3	(8):3, (1,8):2, (6,8):2
<bcb>:3	(8):2, (2,6):2
<be>:3	(2,6):2
<d>:3	(8):2
<bb>:4	(6):3, (8):3, (1,8):2, (2,6):2, (6,8):2
<bc> :4	(8):2, (2,6):2
:5	(6):3, (8):3, (1,8):2, (2,6):2, (6,8):2

ผลลัพธ์ที่ได้จากการสืบค้นสามารถเขียนอยู่ใน (รูปแบบลำดับเหตุการณ์แบบปิด, มิติที่1, มิติที่2, มิติที่3):ค่าสนับสนุน ได้ดังนี้

1. (<aba>, *, 6, 8):2

2. ($\langle abb \rangle$, *, 6, 8):2
3. ($\langle (bd)bc \rangle$, *, *, 8):2
4. ($\langle (bd)ba \rangle$, *, *, 8):2
5. ($\langle (bf)bf \rangle$, *, 6, *):2
6. ($\langle cba \rangle$, *, *, 8):2
7. ($\langle ba \rangle$, *, *, 8):3
8. ($\langle ba \rangle$, 1, *, 8):2
9. ($\langle ba \rangle$, *, 6, 8):2
10. ($\langle bcb \rangle$, *, *, 8):2
11. ($\langle bcb \rangle$, 2, 6, *):2
12. ($\langle be \rangle$, 2, 6, *):2
13. ($\langle d \rangle$, *, *, 8):2
14. ($\langle bb \rangle$, *, 6, *):3
15. ($\langle bb \rangle$, *, *, 8):3
16. ($\langle bb \rangle$, 1, *, 8):2
17. ($\langle bb \rangle$, 2, 6, *):2
18. ($\langle bb \rangle$, *, 6, 8):2
19. ($\langle bc \rangle$, *, *, 8):2
20. ($\langle bc \rangle$, 2, 6, *):2
21. ($\langle b \rangle$, *, 6, *):3
22. ($\langle b \rangle$, *, *, 8):3
23. ($\langle b \rangle$, 1, *, 8):2
24. ($\langle b \rangle$, 2, 6, *):2
25. ($\langle b \rangle$, *, 6, 8):2

พิจารณารูปแบบดังกล่าวทั้งหมดแล้ว พบว่าเป็นเพียงเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ยังมีรูปแบบที่ซ้ำซ้อน เช่น รูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle bc \rangle$, *, *, 8) ในลำดับที่ 19 ถูกครอบคลุมรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle (bd)bc \rangle$, *, *, 8) ในลำดับที่ 3 ได้ เนื่องจากจากรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle bc \rangle$, *, *, 8) เป็นลำดับเหตุการณ์หลายมิติย่อยของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle (bd)bc \rangle$, *, *, 8) และทั้งสองรูปแบบมีค่าสนับสนุนเท่ากัน คือ 2 ซึ่งรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle bc \rangle$, *, *, 8) สามารถพิจารณาได้จากรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด

คู่แข่ง ($\langle bd \rangle_{bc}$, *, *, 8) ดังนั้นจึงสามารถจัดรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle bc \rangle$, *, *, 8) ออกได้

3.3.2 การจัดความซ้ำซ้อน

เนื่องจากผลลัพธ์ที่ได้จากขั้นตอนแรกยังมีบางรูปแบบที่ซ้ำซ้อน ดังนั้นต้องทำการพิจารณาจัดรูปแบบที่ซ้ำซ้อนออกเพื่อให้ได้เซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดตามนิยามที่ 3.5 โดยใช้กฎดังนี้

กำหนดให้ c และ c' เป็นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ถ้า c' เป็นลำดับเหตุการณ์หลายมิติย่อยของ c และค่าสนับสนุนของ c' เท่ากับค่าสนับสนุนของ c แล้ว ให้ทำการตัด c' ออก เพราะ c' ถูกครอบคลุมโดย c

เมื่อทำการตัดรูปแบบที่มีความซ้ำซ้อนทั้งหมดออกไปจากเซตของรูปแบบลำดับเหตุการณ์แบบปิดคู่แข่ง (CCMD) ทำให้ได้เซตของรูปแบบที่ไม่มีซูปเปอร์ลำดับเหตุการณ์หลายมิติที่มีค่าสนับสนุนเท่ากัน ซึ่งจะได้เซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด (CMD) ทั้งหมด

ตัวอย่างที่ 3.13 จากตัวอย่างที่ 3.12 พิจารณารูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle bc \rangle$, 2, 6, *)₂ เป็นลำดับเหตุการณ์หลายมิติย่อยของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle bcb \rangle$, 2, 6, *)₂ และมีค่าสนับสนุนเท่ากัน ดังนั้นสามารถตัดรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ($\langle bc \rangle$, 2, 6, *)₂ ออกได้ ทำการพิจารณาจัดรูปแบบที่ซ้ำซ้อนออกทั้งหมด จะทำให้ได้เซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด (CMD) ทั้งหมด 14 รูปแบบดังนี้

1. ($\langle aba \rangle$, *, 6, 8)₂
2. ($\langle abb \rangle$, *, 6, 8)₂
3. ($\langle bd \rangle_{bc}$, *, *, 8)₂
4. ($\langle bd \rangle_{ba}$, *, *, 8)₂
5. ($\langle bf \rangle_{bf}$, *, 6, *)₂
6. ($\langle cba \rangle$, *, *, 8)₂
7. ($\langle ba \rangle$, 1, *, 8)₂
8. ($\langle bcb \rangle$, *, *, 8)₂
9. ($\langle bcb \rangle$, 2, 6, *)₂
10. ($\langle bc \rangle$, 2, 6, *)₂
11. ($\langle bb \rangle$, 1, *, 8)₂,
12. ($\langle ba \rangle$, *, *, 8)₃
13. ($\langle bb \rangle$, *, 6, *)₃

14. ($\langle \text{bb} \rangle, *, *, 8 \rangle : 3$)

3.4 ความถูกต้อง

ความถูกต้องของการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด แสดงในทฤษฎีบทต่างๆ ดังนี้

ทฤษฎีบทที่ 3.1 $\text{CMD} \subseteq \text{MD}$ ซึ่ง CMD แทนเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด และ MD แทนเซตของรูปแบบลำดับเหตุการณ์หลายมิติ

พิสูจน์ กำหนดให้ (y, x) เป็นลำดับเหตุการณ์หลายมิติใดๆ ในเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง CCMD

$(y, x) \in \text{CCMD}$ ซึ่ง $y \in \text{CS}, x \in \text{CI}$

เนื่องจาก $\text{CI} \subseteq \text{FI}$ และ $\text{CS} \subseteq \text{FS}$

จะได้ว่า $\forall y (y \in \text{CS} \rightarrow y \in \text{FS})$ และ $\forall x (x \in \text{CI} \rightarrow x \in \text{FI})$

ดังนั้น $y \in \text{FS}$ และ $x \in \text{FI}$ แสดงว่า $(y, x) \in \text{MD}$

แสดงให้เห็นว่าสมาชิก (y, x) ใดๆ ใน CCMD ย่อมต้องเป็นสมาชิกใน MD ด้วย

ดังนั้น $\text{CCMD} \subseteq \text{MD}$

จากขั้นตอนที่ 2 ในหัวข้อที่ 3.3.2 การขจัดความซ้ำซ้อนออกจากเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง CCMD ทำให้ได้เซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด CMD

ดังนั้น $\text{CMD} \subseteq \text{CCMD}$

แสดงว่า $\text{CMD} \subseteq \text{CCMD} \subseteq \text{MD}$

สรุปได้ว่า $\text{CMD} \subseteq \text{MD}$

จากทฤษฎีบทที่ 3.1 แสดงให้เห็นว่าเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดเป็นซับเซตของรูปแบบลำดับเหตุการณ์หลายมิติ ดังนั้นการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดให้จำนวนรูปแบบน้อยกว่าหรือเท่ากับจำนวนรูปแบบที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ ซึ่งก็คือ $|\text{CMD}| \leq |\text{MD}|$ เช่น พิสูจน์ตัวอย่างที่ 3.13 การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดจะให้รูปแบบทั้งหมดแค่ 14 รูปแบบ แต่การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติจะให้รูปแบบทั้งหมด 81 รูปแบบ [3] ซึ่งมากกว่ารูปแบบที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดมาก

ทฤษฎีบทที่ 3.2 เซตของรูปแบบลำดับเหตุการณ์หลายมิติสามารถสร้างได้จากเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด

พิสูจน์ จากคุณสมบัติของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดข้อที่ 3 และข้อที่ 1 รูปแบบลำดับเหตุการณ์หลายมิติทุกตัวสามารถสร้างได้จากรูปแบบลำดับเหตุการณ์แบบปิดขนาดใหญ่ จากคุณสมบัติข้อที่ 4 ค่าสนับสนุนของรูปแบบลำดับเหตุการณ์หลายมิติทุกตัวสามารถสร้างได้จากค่าสนับสนุนของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ดังนั้นเซตของรูปแบบลำดับเหตุการณ์หลายมิติสามารถสร้างได้จากเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด

ข้อสังเกต การพิสูจน์ว่าสารสนเทศหลายมิติเท่ากันหรือไม่ในทฤษฎีบทที่ 3.3 และการพิสูจน์ว่าลำดับเหตุการณ์เท่ากันหรือไม่ในทฤษฎีบทที่ 3.4 สามารถพิสูจน์จากเซตของลำดับที่ เพราะถ้าสารสนเทศหลายมิติอยู่ในลำดับที่เดียวกัน ก็แสดงว่า สารสนเทศหลายมิติเป็นตัวเดียวกัน ในทำนองเดียวกัน ถ้าลำดับเหตุการณ์อยู่ในลำดับที่เดียวกัน ก็แสดงว่า ลำดับเหตุการณ์เป็นตัวเดียวกัน

ทฤษฎีบทที่ 3.3 $(s \sqsubseteq s') \wedge (\text{supp}(s) = \text{supp}(s')) \rightarrow I_s = I_{s'}$ ซึ่ง I_s คือเซตของลำดับที่สารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์ s และ $I_{s'}$ คือเซตของลำดับที่สารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์ s'

พิสูจน์ กำหนดให้ $I_s \neq I_{s'}$ ซึ่งกรณีที่เซตไม่เท่ากันเกิดขึ้นสองกรณีคือ
กรณีแรก จำนวนสมาชิกไม่เท่ากัน แสดงว่าค่าสนับสนุนไม่เท่ากันด้วย

$$\text{ดังนั้น } \text{supp}(s) \neq \text{supp}(s') \quad (1)$$

กรณีที่สอง จำนวนสมาชิกเท่ากัน แต่มีสมาชิกบางตัวไม่เหมือนกัน

$$\text{แสดงว่าค่าสนับสนุนเท่ากัน } \text{supp}(s) = \text{supp}(s')$$

$$\text{แต่ } s \not\sqsubseteq s' \quad (2)$$

จาก (1) และ (2)

$$\text{สรุปได้ว่า } ((s \sqsubseteq s') \wedge (\text{supp}(s) = \text{supp}(s'))) \rightarrow I_s = I_{s'}$$

จากทฤษฎีบทที่ 3.3 แสดงให้เห็นว่าสามารถทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดจากสารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์แบบปิดเท่านั้น โดยที่ข้อมูลไม่สูญหาย ในกรณีที่สืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์ เพราะว่ารูปแบบลำดับเหตุการณ์ทุกตัวที่ถูกครอบคลุมโดยรูปแบบลำดับเหตุการณ์แบบปิด จะมีสารสนเทศหลายมิติเท่ากับสารสนเทศหลายมิติของรูปแบบลำดับเหตุการณ์แบบปิด ดังนั้นวิธีการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด จึงไม่จำเป็นต้องทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดกับสารสนเทศหลายมิติของทุกรูปแบบลำดับเหตุการณ์

ตัวอย่างที่ 3.14 จากตารางที่ 3.1 พบว่า $\langle cb \rangle \subseteq \langle bcb \rangle$ และค่าสนับสนุนของ $\langle cb \rangle$ และ $\langle bcb \rangle$ เท่ากัน คือ 3 พบสารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์ $\langle bcb \rangle$ ในลำดับที่ 1, 2 และ 5

ตารางที่ 3.2 สารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์ (bcb)

ลำดับที่	มิติที่ 1	มิติที่ 2	มิติที่ 3
1.	1	5	8
2.	2	6	7
5.	2	6	8

นอกจากนี้ ยังพบสารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์ $\langle cb \rangle$ ในลำดับ 1, 2 และ 5 ด้วย แสดงว่าสารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์ $\langle bcb \rangle$ และ $\langle cb \rangle$ เป็นสารสนเทศหลายมิติเดียวกัน

ทฤษฎีบทที่ 3.4 $(I \subseteq I') \wedge (\text{supp}(I) = \text{supp}(I')) \rightarrow S_I = S_{I'}$ ซึ่ง $S_{I'}$ คือ เซตลำดับที่ของลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูล I' และ S_I คือ เซตลำดับที่ของรูปแบบลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูล I

พิสูจน์ โดยกำหนดให้ $S_I \neq S_{I'}$ ซึ่งเกิดขึ้นสองกรณีคือ

กรณีแรก จำนวนสมาชิกไม่เท่ากัน แสดงว่าค่าสนับสนุนไม่เท่ากันด้วย

$$\text{ดังนั้น } \text{supp}(I) \neq \text{supp}(I') \quad (3)$$

กรณีที่สอง จำนวนสมาชิกเท่ากัน แต่มีสมาชิกไม่เหมือนกันทุกตัว

$$\text{แสดงว่าค่าสนับสนุนเท่ากัน } \text{supp}(I) = \text{supp}(I')$$

$$\text{แต่ } I \not\subseteq I' \quad (4)$$

จาก (3) และ (4)

$$\text{สรุปได้ว่า } ((I \subseteq I') \wedge (\text{supp}(I) = \text{supp}(I'))) \rightarrow S_I = S_{I'}$$

จากทฤษฎีบทที่ 3.4 แสดงให้เห็นว่าสามารถทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดจากลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูลแบบปิดเท่านั้น โดยที่ข้อมูลไม่สูญหาย ในกรณีที่ทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด เพราะรูปแบบกลุ่มข้อมูลทุกตัวที่ถูกครอบคลุมโดยรูปแบบกลุ่มข้อมูลแบบปิด จะมีลำดับเหตุการณ์เท่ากับลำดับเหตุการณ์ของรูปแบบกลุ่มข้อมูลแบบปิด ดังนั้นวิธีการสืบค้นรูปแบบลำดับเหตุการณ์หลาย

มิติแบบปิด จึงไม่จำเป็นต้องทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดกับลำดับเหตุการณ์ของทุกรูปแบบกลุ่มข้อมูล

ตัวอย่างที่ 3.15 จากตารางที่ 3.1 พบว่า $(2) \subseteq (2, 6)$ และค่าสนับสนุนของ (2) และ $(2, 6)$ เท่ากัน คือ 2 พบลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูล $(2, 6)$ ในลำดับที่ 2 และ 5

ตารางที่ 3.3 ลำดับเหตุการณ์ที่สัมพันธ์กับกลุ่มข้อมูล $(2,6)$

ลำดับที่	ลำดับเหตุการณ์
2.	$\langle (bf)(ce)b(fg) \rangle$
5.	$\langle a(bd)bc b(ade) \rangle$

นอกจากนี้พบลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูล (2) ในลำดับ 2 และ 5 เหมือนกัน แสดงว่าลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูล (2) กับ $(2, 6)$ เป็นลำดับเหตุการณ์เดียวกัน

กล่าวโดยสรุป ทฤษฎีบทที่ 3.1 รับประกันได้ว่าเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดเป็นซับเซตของรูปแบบลำดับเหตุการณ์หลายมิติ แสดงให้เห็นว่าการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดให้จำนวนรูปแบบน้อยกว่าหรือเท่ากับการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ ทฤษฎีบทที่ 3.2 รับประกันได้ว่าเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดสามารถสร้างเซตของรูปแบบลำดับเหตุการณ์หลายมิติ แสดงให้เห็นว่าเซตของรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดครอบคลุมรูปแบบลำดับเหตุการณ์หลายมิติทุกตัว ดังนั้นสามารถสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดได้โดยที่ข้อมูลไม่สูญหาย สำหรับทฤษฎีที่ 3.3 และ 3.4 รับประกันการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดโดยการผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดด้วยวิธีที่ 1 จะทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดบนลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูลแบบปิดเท่านั้น โดยที่ข้อมูลไม่สูญหาย ส่วนการผสมผสานด้วยวิธีที่ 2 จะทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดบนสารสนเทศหลายมิติที่สัมพันธ์กับรูปแบบลำดับเหตุการณ์แบบปิดเท่านั้น โดยที่ข้อมูลไม่สูญหาย

บทที่ 4

การทดลองและผลการทดลอง

จากขั้นตอนวิธีสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดทั้งสองขั้นตอน ได้แก่ (1) การผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด (2) การขจัดความซ้ำซ้อน จะเห็นว่าในขั้นตอนที่ 2 การขจัดความซ้ำซ้อนขึ้นอยู่กับผลลัพธ์จากขั้นตอนที่ 1 ซึ่งให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่ง ถ้าจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อย จะทำให้ลดเวลาในการจัดรูปแบบความซ้ำซ้อนในขั้นตอนที่ 2 ได้ ดังนั้นในบทนี้จึงนำเสนอการทดลองเพื่อเปรียบเทียบจำนวนรูปแบบที่ได้จากการผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดทั้งสองวิธี คือ วิธีแรกการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดที่สัมพันธ์ และวิธีที่สองการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์

4.1 การทดลอง

4.1.1 ข้อมูลที่ใช้ในการทดสอบ

ข้อมูลที่ใช้ในการทดลองเป็นชุดข้อมูลลำดับเหตุการณ์หลายมิติที่ถูกสร้างขึ้นโดยใช้วิธีการเดียวกับที่ปรากฏใน [3] วิธีการนี้สร้างชุดข้อมูลลำดับเหตุการณ์หลายมิติจากชุดข้อมูลลำดับเหตุการณ์และชุดข้อมูลสารสนเทศหลายมิติ

1. ชุดข้อมูลลำดับเหตุการณ์ ถูกสร้างขึ้นจากโปรแกรมสังเคราะห์ข้อมูลของไอบีเอ็ม (ที่มา: <http://www.almaden.ibm.com/software/quest/Resources/>) ซึ่งโปรแกรมหดกล่าวเป็นเครื่องมือมาตรฐานที่ใช้ในการสร้างชุดข้อมูลลำดับเหตุการณ์ เพื่อใช้ในการทดลองการสืบค้นรูปแบบลำดับเหตุการณ์และการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด ชุดข้อมูลลำดับเหตุการณ์ที่ถูกสร้างขึ้นเพื่อใช้ในการทดลองในงานวิจัยนี้ประกอบไปด้วยชุดข้อมูลลำดับเหตุการณ์ 2 ชุดข้อมูล คือ C10N10T2.5S8I1.25 และ C20N10T2.5S8I1.25 ซึ่งในตารางที่ 4.1 แสดงความหมายของพารามิเตอร์หลักแต่ละตัวที่ต้องป้อนเข้าไปในโปรแกรมสังเคราะห์ข้อมูลของไอบีเอ็ม และตารางที่ 4.2 แสดงค่าพารามิเตอร์แต่ละตัวที่ใช้ในการสร้างชุดข้อมูลลำดับเหตุการณ์ทั้งสองชุดข้อมูล

ตารางที่ 4.1 แสดงความหมายของพารามิเตอร์ที่ใช้ในการสร้างลำดับเหตุการณ์

เครื่องหมาย	ความหมาย
C	จำนวนรายการเปลี่ยนแปลง
N	จำนวนของชั้นข้อมูล
T	ค่าเฉลี่ยของชั้นข้อมูลในแต่ละกลุ่มข้อมูล
S	ค่าเฉลี่ยของกลุ่มข้อมูลในแต่ละรายการเปลี่ยนแปลง
I	ค่าเฉลี่ยของชั้นข้อมูลในแต่ละรายการเปลี่ยนแปลง

ตารางที่ 4.2 ค่าพารามิเตอร์ที่ป้อนเข้าไปในโปรแกรมสังเคราะห์ข้อมูล

ลำดับเหตุการณ์	C	N	T	S	I
C10N10T2.5S8I1.25	10,000	10,000	2.5	8	1.25
C20N10T2.5S8I1.25	20,000	10,000	2.5	8	1.25

2. ชุดข้อมูลสารสนเทศหลายมิติ ถูกสร้างขึ้นจากโปรแกรมที่ทำการสุ่มตัวเลขในแต่ละมิติ ซึ่งงานวิจัยนี้ได้ทำการสร้างชุดข้อมูลสารสนเทศหลายมิติเพื่อใช้ในการทดลองทั้งหมด 12 ชุดข้อมูล โดยที่ชุดข้อมูลสารสนเทศหลายมิติ 6 ชุดแรก คือ 2_2S1, 3_3S1, 4_4S1, 5_5S1, 8_8S1 และ 10_10S1 มีจำนวนรายการเปลี่ยนแปลงเท่ากับชุดข้อมูลลำดับเหตุการณ์ C10N10T2.5S8I1.25 และชุดข้อมูลสารสนเทศหลายมิติอีก 6 ชุด คือ 2_2S2, 3_3S2, 4_4S2, 5_5S2, 8_8S2 และ 10_10S2 มีจำนวนรายการเปลี่ยนแปลงเท่ากับชุดข้อมูลลำดับเหตุการณ์ C20N10T2.5S8I1.25

ชุดข้อมูลสารสนเทศหลายมิติแต่ละชุดมีจำนวนมิติและค่าตามชื่อที่ตั้งไว้ เช่น ชุดข้อมูลสารสนเทศหลายมิติ 3_3S1 หมายความว่า มีจำนวนมิติเท่ากับ 3 มิติ และในแต่ละมิติมีข้อมูล 3 ค่า โดยในมิติที่ 1 ประกอบไปด้วยค่า 1 ถึง 3 มิติที่ 2 ประกอบไปด้วยค่า 4 ถึง 6 และมิติที่ 3 ประกอบไปด้วยค่า 7 ถึง 9 ชุดข้อมูลสารสนเทศหลายมิติที่ใช้ในการทดลองมีจำนวนมิติเท่ากับ 2 มิติ, 3 มิติ, 4 มิติ, 5 มิติ, 8 มิติ และ 10 มิติ ซึ่งค่าในแต่ละมิติจะขึ้นอยู่กับจำนวนมิติ ซึ่งแสดงรายละเอียดในตารางที่ 4.3

ตารางที่ 4.3 แสดงค่าแต่ละมิติของสารสนเทศหลายมิติ

มิติที่ จำนวน	1	2	3	4	5	6	7	8	9	10
2 มิติ	1-2	3-4								
3 มิติ	1-3	4-6	7-9							
4 มิติ	1-4	5-8	9-12	13-16						
5 มิติ	1-5	6-10	11-15	16-20	21-25					
8 มิติ	1-8	9-16	17-24	25-32	33-40	41-48	49-56	57-64		
10 มิติ	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100

ชุดข้อมูลลำดับเหตุการณ์หลายมิติที่ใช้ในการทดลองมีทั้งหมด 12 ชุดข้อมูล ซึ่งแต่ละชุดข้อมูลเกิดจากการรวมกันของชุดข้อมูลลำดับเหตุการณ์และชุดข้อมูลสารสนเทศหลายมิติ ดังแสดงในตารางที่ 4.4

ตารางที่ 4.4 แสดงรายละเอียดของชุดข้อมูลลำดับเหตุการณ์หลายมิติ

ชุดข้อมูลลำดับเหตุการณ์หลายมิติ	สารสนเทศหลายมิติ	ลำดับเหตุการณ์
MDS1_2	C10N10T2.5S8I1.25	2_2S1
MDS1_3	C10N10T2.5S8I1.25	3_3 S1
MDS1_4	C10N10T2.5S8I1.25	4_4 S1
MDS1_5	C10N10T2.5S8I1.25	5_5 S1
MDS1_8	C10N10T2.5S8I1.25	8_8 S1
MDS1_10	C10N10T2.5S8I1.25	10_10 S1
MDS2_2	C20N10T2.5S8I1.25	2_2 S2
MDS2_3	C20N10T2.5S8I1.25	3_3 S2
MDS2_4	C20N10T2.5S8I1.25	4_4 S2
MDS2_5	C20N10T2.5S8I1.25	5_5 S2
MDS2_8	C20N10T2.5S8I1.25	8_8 S2
MDS2_10	C20N10T2.5S8I1.25	10_10 S2

4.1.2 โปรแกรมที่ใช้ในการทดลอง

โปรแกรมที่ใช้ในการทดสอบกับชุดข้อมูลลำดับเหตุการณ์หลายมิติมี 2 โปรแกรม คือ CIS (โปรแกรมที่ทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบเปิดที่สัมพันธ์) กับ CSI (โปรแกรมที่ทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบเปิดที่สัมพันธ์) ซึ่งทั้งสองโปรแกรมได้ผสมผสานโปรแกรมการสืบค้นรูปแบบกลุ่มข้อมูลแบบเปิดที่ชื่อว่า Closet [7] และโปรแกรมการสืบค้นรูปแบบลำดับเหตุการณ์แบบเปิดที่ชื่อว่า CloSpan [19] เข้าด้วยกัน

4.1.3 เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการทดสอบ มีคุณสมบัติดังต่อไปนี้

หน่วยประมวลผลกลาง (CPU)	: Intel Pentium III processor 1.00 GHz
หน่วยความจำหลัก (RAM)	: 256 MB SD-RAM
หน่วยความจำสำรอง (Hard Disk)	: 20 GB
ระบบปฏิบัติการ (OS)	: Window XP Professional Version 2003
โปรแกรมที่ใช้ในการพัฒนา	: Visual Studio C++ Version 6.0

4.1.4 การออกแบบการทดลองและเหตุผล

งานวิจัยนี้ได้ทำการทดลองเพื่อเปรียบเทียบจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นของโปรแกรม CIS และ CSI โดยการทดลองจะแบ่งออกเป็น 4 ส่วน ดังนี้

ส่วนที่ 1 ทำการทดลองชุดข้อมูลลำดับเหตุการณ์หลายมิติ MDS1_2, MDS1_3, MDS1_4, MDS1_5, MDS1_8 และ MDS1_10 กับโปรแกรม CIS และ CSI โดยกำหนดค่าสนับสนุนต่ำสุดเท่ากับ 10

ส่วนที่ 2 ทำการทดลองชุดข้อมูลลำดับเหตุการณ์หลายมิติ MDS2_2, MDS2_3, MDS2_4, MDS2_5, MDS2_8 และ MDS2_10 กับโปรแกรม CIS และ CSI โดยกำหนดค่าสนับสนุนต่ำสุดเท่ากับ 10

ส่วนที่ 3 ทำการทดลองชุดข้อมูลลำดับเหตุการณ์หลายมิติ MDS1_10 กับโปรแกรม CIS และ CSI โดยกำหนดค่าสนับสนุนต่ำสุด 3 ค่า คือ 5 10 และ 20

ส่วนที่ 4 ทำการทดลองชุดข้อมูลลำดับเหตุการณ์หลายมิติ MDS1_2 กับโปรแกรม CIS และ CSI โดยกำหนดค่าสนับสนุนต่ำสุด 3 ค่า คือ 5 10 และ 20

ส่วนที่ 5 ทำการทดลองชุดข้อมูลลำดับเหตุการณ์หลายมิติ MDS1_4 กับโปรแกรม CIS และ CSI โดยกำหนดค่าสนับสนุนต่ำสุด 3 ค่า คือ 5 10 และ 20

ส่วนที่ 6 ทำการทดลองชุดข้อมูลลำดับเหตุการณ์หลายมิติ MDS1_3 กับโปรแกรม CIS และ CSI โดยกำหนดค่าสนับสนุนต่ำสุด 3 ค่า คือ 5 10 และ 20

ส่วนที่ 1 กับส่วนที่ 2 ทำการทดลองเพื่อพิจารณาผลกระทบของจำนวนมิติของชุดข้อมูลสารสนเทศหลายมิติที่มีต่อการสืบค้นของโปรแกรม CIS และ CSI เมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 10,000 และ 20,000 รายการ ส่วนที่ 3 4 5 และ 6 ทำการทดลองเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดที่ไม่มากกว่า 20 ที่มีต่อการสืบค้นจากโปรแกรม CIS และ CSI เมื่อจำนวนมิติของสารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 2 มิติ, 3 มิติ, 4 มิติ และ 10 มิติ

4.2 ผลการทดลอง

ผลการทดลองโดยใช้โปรแกรม CIS และโปรแกรม CSI กับชุดข้อมูลลำดับเหตุการณ์หลายมิติ และค่าสนับสนุนต่ำสุดที่ออกแบบในหัวข้อที่ 4.1.4 แบ่งออกเป็น 6 ส่วน ดังนี้

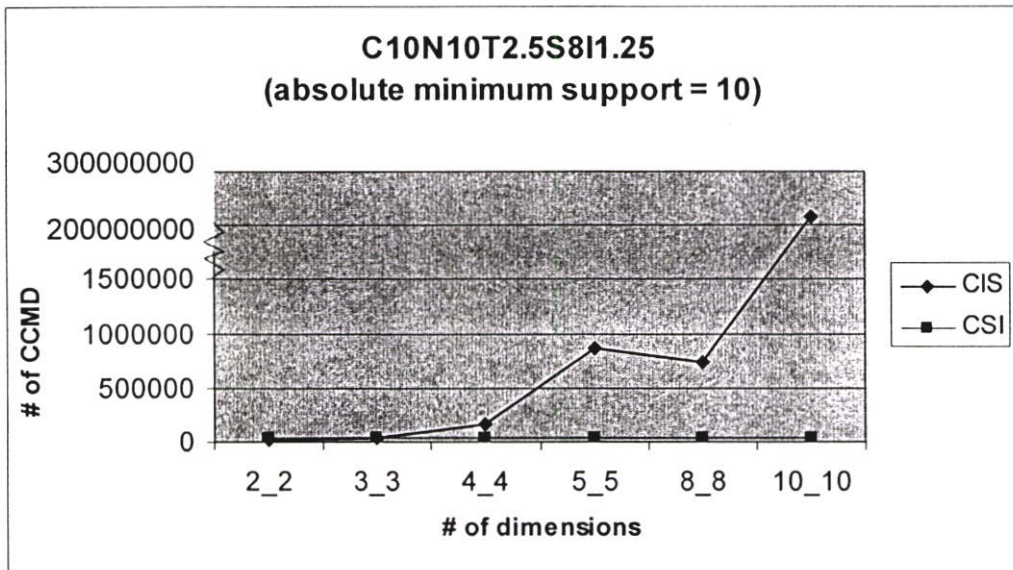
1. ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 10,000 รายการ
2. ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 20,000 รายการ
3. ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 10 มิติ
4. ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 2 มิติ
5. ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 4 มิติ
6. ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 3 มิติ

4.2.1 ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 10,000 รายการ

การทดลองเพื่อพิจารณาผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 10,000 รายการ ได้ทำการทดลองตามหัวข้อ 4.1.4 ส่วนที่ 1 ซึ่งผลการทดลองแสดงในตารางที่ 4 .5 ดังนี้

ตารางที่ 4.5 แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นตามการทดลองในส่วนที่ 1

ชุดข้อมูลลำดับเหตุการณ์หลายมิติ	CIS	CSI
MDS1_2	25,323	30,016
MDS1_3	32,366	34,940
MDS1_4	161,556	33,974
MDS1_5	867,430	32,957
MDS1_8	732,537	31,510
MDS1_10	207,181,593	28,489



รูปที่ 4.1 ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 10,000 รายการ

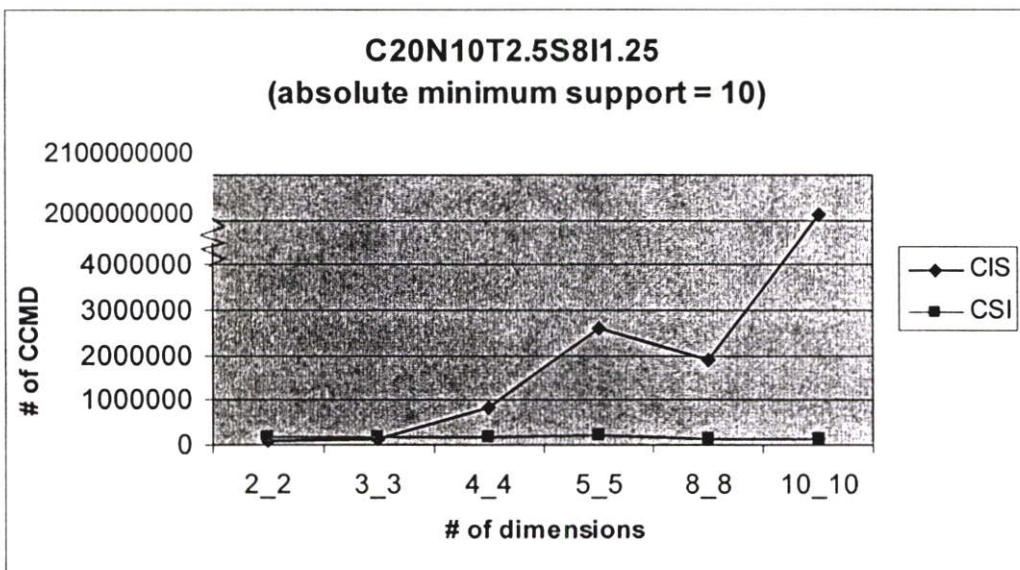
รูปที่ 4.1 แสดงการเปรียบเทียบจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นทั้งสองโปรแกรมเพื่อพิจารณาผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 10,000 รายการ จากรูปดังกล่าวแสดงให้เห็นว่า การสืบค้นโดยใช้โปรแกรม CIS จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นโดยใช้โปรแกรม CSI ในกรณีที่ชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 2 มิติและ 3 มิติ แต่เมื่อเพิ่มจำนวนมิติของสารสนเทศหลายมิติเป็น 4 มิติ, 5 มิติ, 8 มิติและ 10 มิติ การสืบค้นโดยใช้โปรแกรม CSI จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นโดยใช้โปรแกรม CIS

4.2.2 ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 20,000 รายการ

การทดลองเพื่อพิจารณาผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 20,000 รายการ ได้ทำการทดลองตามหัวข้อ 4.1.4 ส่วนที่ 2 ซึ่งผลการทดลองของทั้งสองโปรแกรม แสดงในตารางที่ 4.6 ดังนี้

ตารางที่ 4.6 แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นตามการทดลองในส่วนที่ 2

ชุดข้อมูลลำดับเหตุการณ์หลายมิติ	CIS	CSI
MDS2_2	101,762	173,781
MDS2_3	116,527	176,660
MDS2_4	835,284	160,290
MDS2_5	2,595,413	215,631
MDS2_8	1,876,119	115,498
MDS2_10	2,008,698,033	124,077



รูปที่ 4.2 ผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 20,000 รายการ

รูปที่ 4.2 แสดงการเปรียบเทียบจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากสืบค้นทั้งสองโปรแกรมเพื่อพิจารณาผลกระทบของจำนวนมิติเมื่อจำนวนรายการเปลี่ยนแปลงเท่ากับ 20,000 รายการ จากรูปดังกล่าวแสดงให้เห็นว่า การสืบค้นโดยใช้โปรแกรม CIS จะให้

จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นโดยใช้โปรแกรม CSI ในกรณีที่ชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 2 มิติและ 3 มิติ แต่เมื่อเพิ่มจำนวนมิติเป็น 4 มิติ, 5 มิติ, 8 มิติและ 10 มิติแล้ว การสืบค้นโดยใช้โปรแกรม CSI จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติคู่แข่งน้อยกว่าการสืบค้นโดยใช้โปรแกรม CIS

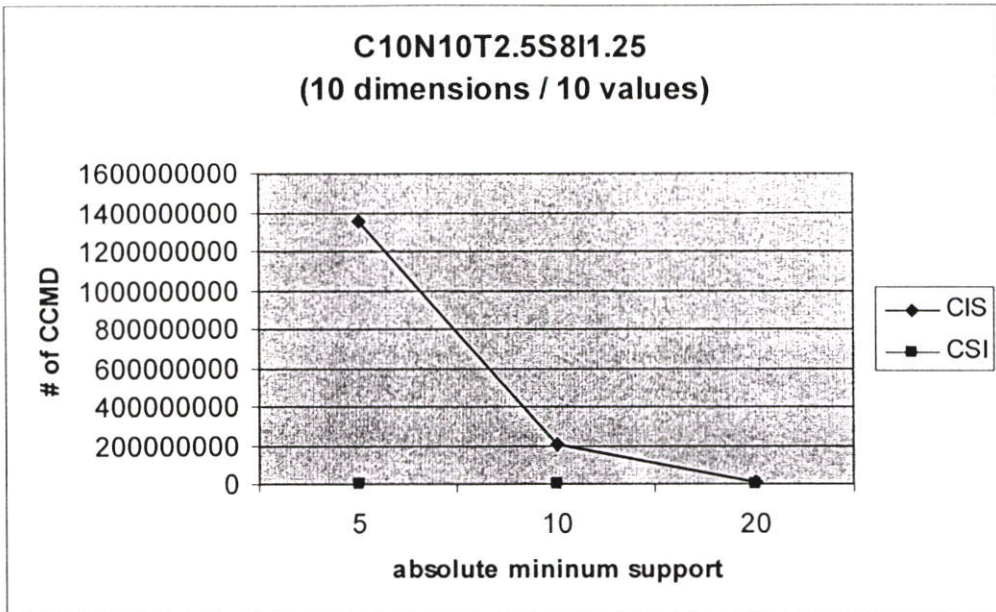
จากรูปที่ 4.1 และ 4.2 แสดงให้เห็นว่า ในกรณีที่ชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 2 มิติและ 3 มิติ การสืบค้นโดยใช้โปรแกรม CIS จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นโดยใช้โปรแกรม CSI และในกรณีที่สารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 4 มิติ, 5 มิติ, 8 มิติและ 10 มิติ การสืบค้นโดยใช้โปรแกรม CSI จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นโดยใช้โปรแกรม CIS ซึ่งแสดงให้เห็นแนวโน้มว่าเมื่อชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติมากขึ้น การสืบค้นโดยใช้โปรแกรม CSI ยังคงให้จำนวนรูปแบบน้อยกว่าการสืบค้นโดยใช้โปรแกรม CIS นอกจากนี้ยังพบว่าไม่ว่ารายการเปลี่ยนแปลงจะเป็นเท่าไรก็ยังคงได้ผลลัพธ์ลักษณะคล้ายกัน อย่างไรก็ตามข้อสรุปที่ได้เป็นข้อสรุปที่ค่าสนับสนุนต่ำสุดเป็น 10 เท่านั้น ดังนั้นการทดลองในหัวข้อที่ 4.2.3 4.2.4 4.2.5 และ 4.2.6 จึงเลือกที่จะทดลองกับชุดข้อมูลลำดับเหตุการณ์หลายมิติ MDS1_2, MDS1_3, MDS1_4, และ MDS1_10 เพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดที่มีต่อการสืบค้นของโปรแกรม CIS และ CSI

4.2.3 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 10 มิติ

การทดลองเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 10 มิติ ได้ทำการทดลองตามหัวข้อ 4.1.4 ส่วนที่ 3 ซึ่งผลการทดลองของทั้งสองโปรแกรมแสดงในตารางที่ 4.7 ดังนี้

ตารางที่ 4.7 แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นตามการทดลองในส่วนที่ 3

ค่าสนับสนุนต่ำสุด	CIS	CSI
5	1,348,736,445	174,411
10	207,181,593	28,489
20	11,502,198	1,442



รูปที่ 4.3 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 10 มิติ

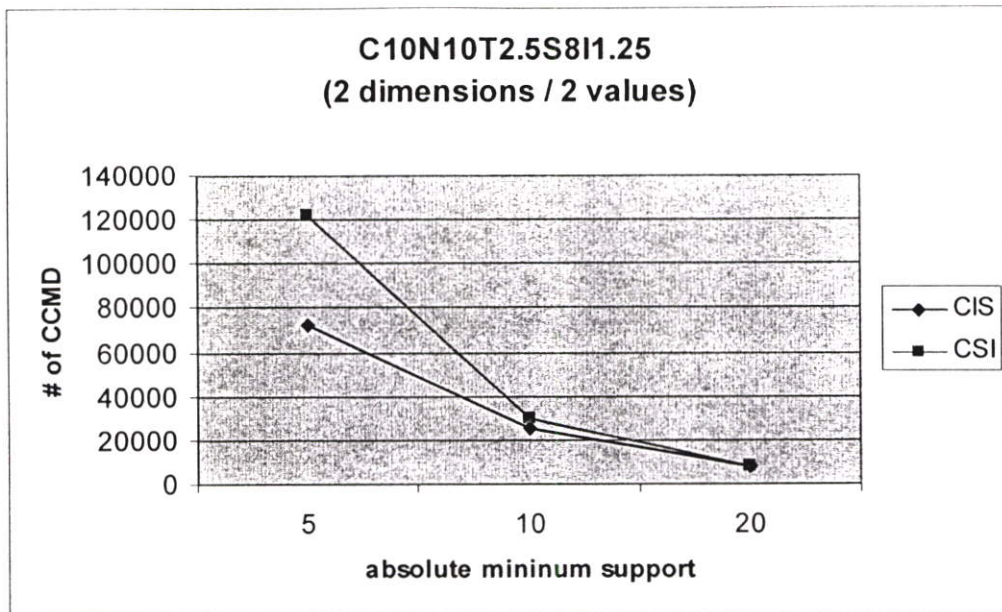
รูปที่ 4.3 แสดงการเปรียบเทียบจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นทั้งสองโปรแกรมเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดเมื่อชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 10 มิติ จากรูปดังกล่าวแสดงให้เห็นว่า ในกรณีที่ค่าสนับสนุนต่ำสุดมีค่าเท่ากับ 5 10 และ 20 การสืบค้นจากโปรแกรม CSI จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นจากโปรแกรม CIS

4.2.4 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 2 มิติ

การทดลองเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 2 มิติ ได้ทำการทดลองตามหัวข้อ 4.1.4 ส่วนที่ 4 ซึ่งผลการทดลองของทั้งสองโปรแกรมแสดงในตารางที่ 4.8 ดังนี้

ตารางที่ 4.8 แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นตามการทดลองในส่วนที่ 4

ค่าสนับสนุนต่ำสุด	CIS	CSI
5	72,633	121,597
10	25,323	30,016
20	7,971	8,024



รูปที่ 4.4 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 2 มิติ

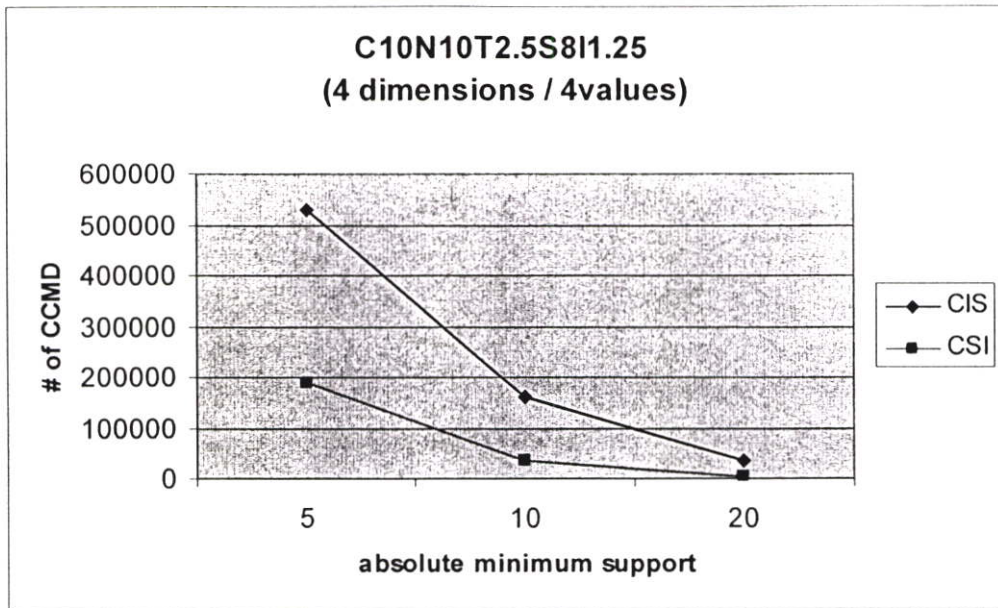
รูปที่ 4.4 แสดงการเปรียบเทียบจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นทั้งสองโปรแกรมเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดเมื่อชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 2 มิติ จากรูปดังกล่าวแสดงให้เห็นว่า ในกรณีที่ค่าสนับสนุนต่ำสุดมีค่าเท่ากับ 5 10 และ 20 การสืบค้นจากโปรแกรม CIS จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นจากโปรแกรม CSI

4.2.5 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 4 มิติ

การทดลองเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 4 มิติ ได้ทำการทดลองตามหัวข้อ 4.1.4 ส่วนที่ 5 ซึ่งผลการทดลองของทั้งสองโปรแกรมแสดงในตารางที่ 4.9 ดังนี้

ตารางที่ 4.9 แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นตามการทดลองในส่วนที่ 5

ค่าสนับสนุนต่ำสุด	CIS	CSI
5	530,292	187,415
10	161,556	33,974
20	37,164	3,398



รูปที่ 4.5 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 4 มิติ

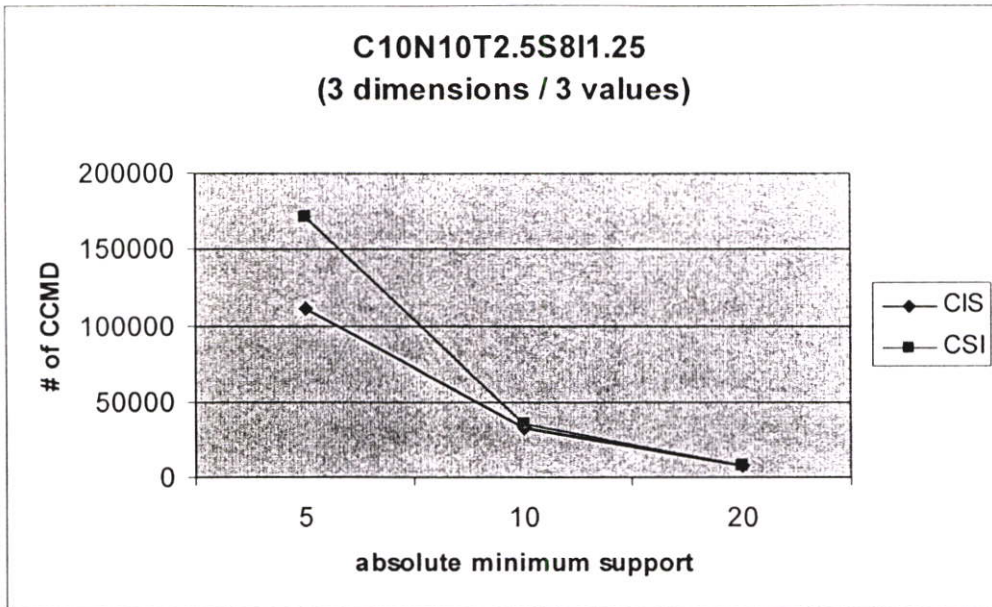
รูปที่ 4.5 แสดงการเปรียบเทียบจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นทั้งสองโปรแกรมเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดเมื่อชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 4 มิติ จากรูปดังกล่าวแสดงให้เห็นว่า ในกรณีที่ค่าสนับสนุนต่ำสุดมีค่าเท่ากับ 5 10 และ 20 การสืบค้นจากโปรแกรม CSI จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นจากโปรแกรม CIS

4.2.6 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 3 มิติ

การทดลองเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติเท่ากับ 3 มิติ ได้ทำการทดลองตามหัวข้อ 4.1.4 ส่วนที่ 6 ซึ่งผลการทดลองของทั้งสองโปรแกรมแสดงในตารางที่ 4.10 ดังนี้

ตารางที่ 4.10 แสดงจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นตามการทดลองในส่วนที่ 6

ค่าสนับสนุนต่ำสุด	CIS	CSI
5	111,308	171,845
10	32,366	34,940
20	7807	7,813



รูปที่ 4.6 ผลกระทบของค่าสนับสนุนต่ำสุดเมื่อจำนวนมิติมีเท่ากับ 3 มิติ

รูปที่ 4.6 แสดงการเปรียบเทียบจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นทั้งสองโปรแกรมเพื่อพิจารณาผลกระทบของค่าสนับสนุนต่ำสุดเมื่อชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติเท่ากับ 3 มิติ จากรูปดังกล่าวแสดงให้เห็นว่า ในกรณีที่ค่าสนับสนุนต่ำสุดมีค่าเท่ากับ 5 10 และ 20 การสืบค้นจากโปรแกรม CIS จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นจากโปรแกรม CSI

4.2.7 สรุปผลการทดลอง

จากรูปที่ 4.3 4.4 4.5 และ 4.6 จะเห็นว่าที่ค่าสนับสนุนต่ำสุดไม่มากกว่า 20 และชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติอย่างมาก 3 มิติ การสืบค้นโดยใช้โปรแกรม CIS จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นโดยใช้โปรแกรม CSI แต่ถ้าชุดข้อมูลสารสนเทศหลายมิติมีจำนวนมิติมากกว่า 3 มิติ การสืบค้นโดยใช้โปรแกรม CSI จะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นโดยใช้โปรแกรม CIS ผลการทดลองนี้สนับสนุนผลการทดลองในรูปที่ 4.1 และ 4.2 ดังนั้นจึงสามารถสรุปได้ว่าที่ค่าสนับสนุนต่ำสุดไม่มากกว่า 20 จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากการสืบค้นโดยโปรแกรม CIS จะน้อยกว่าจำนวนรูปแบบที่ได้จากการสืบค้นโดยโปรแกรม CSI เมื่อสารสนเทศหลายมิติมีจำนวนมิติน้อยกว่าหรือเท่ากับ 3 มิติ แต่ถ้าจำนวนมิติมากกว่า 3 มิติ จำนวนรูปแบบที่ได้จากการสืบค้นโดยโปรแกรม CSI จะน้อยกว่าจำนวนรูปแบบที่ได้จากโปรแกรม CIS

บทที่ 5

สรุปและข้อเสนอแนะ

5.1 สรุป

งานวิจัยนี้ได้นำเสนอการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด ซึ่งประกอบด้วย ขั้นตอนหลัก 2 ขั้นตอน คือ ขั้นตอนแรก การผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิด ขั้นตอนที่สอง การขจัดความซ้ำซ้อน ซึ่งขั้นตอนที่สองขึ้นอยู่กับผลลัพธ์ของขั้นตอนแรก ถ้าจำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งที่ได้จากขั้นตอนแรกมีจำนวนน้อย ก็จะช่วยลดเวลาในการจัดรูปแบบที่ซ้ำซ้อนในขั้นตอนที่สองได้

จากผลการทดลองเพื่อเปรียบเทียบจำนวนรูปแบบที่ได้จากการผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดทั้งสองวิธี คือ การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดที่สัมพันธ์ และการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์ แสดงให้เห็นว่า ที่ค่าสนับสนุนต่ำสุดไม่มากกว่า 20 การสืบค้นโดยวิธีแรกจะให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยกว่าการสืบค้นโดยวิธีที่สอง เมื่อสารสนเทศหลายมิตินี้มีจำนวนมิติไม่มากกว่า 3 มิติ แต่ถ้าจำนวนมิติมากกว่า 3 มิติ การสืบค้นโดยวิธีที่สองจะให้จำนวนรูปแบบน้อยกว่าการสืบค้นโดยวิธีแรก

นอกจากนี้งานวิจัยนี้ยังแสดงการพิสูจน์ทางคณิตศาสตร์ให้เห็นว่า: (1) การสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดให้จำนวนรูปแบบที่น้อยกว่าหรือเท่ากับจำนวนรูปแบบที่ได้จากการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติ ซึ่งทำให้ประหยัดพื้นที่ในการเก็บรูปแบบ, (2) สามารถสร้างรูปแบบลำดับเหตุการณ์หลายมิติทั้งหมดจากรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดได้ , และ (3) การสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดจะทำการสืบค้นกับสารสนเทศหลายมิติที่สัมพันธ์กับลำดับเหตุการณ์แบบปิดเท่านั้น โดยที่ข้อมูลไม่สูญหาย ในกรณีที่ทำการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดก่อนการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดที่สัมพันธ์ และการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดจะทำการสืบค้นกับลำดับเหตุการณ์ที่สัมพันธ์กับรูปแบบกลุ่มข้อมูลแบบปิดเท่านั้น โดยที่ข้อมูลไม่สูญหาย ในกรณีที่ทำการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดก่อนการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดที่สัมพันธ์

5.2 ข้อเสนอแนะ

การผสมผสานการสืบค้นรูปแบบกลุ่มข้อมูลแบบปิดและการสืบค้นรูปแบบลำดับเหตุการณ์แบบปิดเพื่อใช้ในการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด เป็นการผสมผสานที่มีได้คำนึงถึงจำนวนรูปแบบที่เป็นผลลัพธ์จากการผสมผสาน ดังนั้นแนวทางหนึ่งในการพิจารณาขั้นตอนวิธีผสมผสาน คือ การพัฒนาขั้นตอนวิธีที่ให้จำนวนรูปแบบลำดับเหตุการณ์หลายมิติแบบปิดคู่แข่งน้อยที่สุด หากมีการพัฒนาขั้นตอนวิธีนี้ร่วมกับขั้นตอนวิธีที่มีประสิทธิภาพที่ใช้ในการจัดรูปแบบซ้ำซ้อน จะทำให้ได้ขั้นตอนวิธีที่มีประสิทธิภาพในการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด นอกจากนี้แนวทางอีกแนวทางหนึ่งในการพัฒนาขั้นตอนวิธีเพื่อการสืบค้นรูปแบบลำดับเหตุการณ์หลายมิติแบบปิด คือ การพัฒนาขั้นตอนวิธีที่ไม่มีการสร้างรูปแบบซ้ำซ้อนแนวทางสุดท้ายในการวิจัยในอนาคต ได้แก่ การเตรียมข้อมูลเพื่อการสืบค้นโดยการเตรียมข้อมูลซึ่งรวมสารสนเทศหลายมิติและลำดับเหตุการณ์ให้อยู่ในรูปแบบเดียวกันก่อนการสืบค้น เพื่อให้การสืบค้นทำได้ง่ายขึ้น

เอกสารอ้างอิง

- [1] C. Lucchese, S. Orlando and R. Perego, “**DCI-CLOSED: A Fast and Memory Efficient Algorithm to Mine Frequent Closed Itemsets,**” In Proc. of IEEE ICDM workshop on Frequent Itemset Mining Implementation (FIMI'04), Vol 126, Brighton, UK, November 2004.
- [2] D. Lin and Z. Kedem, “**Pincer-Search: A New Algorithm for Discovering the Maximum Frequent Set,**” In Proc. of the 6th Int'l Conference on Extending Database Technology (EDBT'98), pp. 105-119, Valencia, Spain, March 1998.
- [3] H. Pinto, J. Han, J. Pei, K. Wang, Q. Chen and U. Dayal, **Multi-dimensional Sequential Pattern Mining.** M.Sc. Thesis, Simon Fraser University, Canada, April 2001.
- [4] J. Han, J. Pei, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal and M-C. Hsu, “**PrefixSpan: Mining Sequential Patterns Efficiently by Prefix-Projected Pattern Growth,**” In Proc. 2001 Int. Conf. Data Engineering (ICDE'01), pp. 215-224, Heidelberg, Germany, April 2001.
- [5] J. Han, J. Pei and Y. Yin, “**Mining Frequent Patterns without Candidate Generation,**” In Proc. 2000 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'00), pp.1-12, Dallas, TX, May 2000.
- [6] J. Han, J. Wang, Y. Lu and P. Tzvetkov, “**Mining Top-k Frequent Closed Patterns without Minimum Support,**” In Proc. 2002 Int. Conf. on Data Mining (ICDM'02), pp. 211-218, Maebahi, Japan, December 2002.
- [7] J. Pei, J. Han and R. Mao, “**CLOSET: An Efficient Algorithm for Mining Frequent Closed Itemsets,**” In Proc. 2000 ACM-SIGMOD Int. Workshop Data Mining and Knowledge Discovery (DMKD'00), pp. 21-30, May 2000.
- [8] J. Wang and J. Han, “**BIDE: Efficient Mining of Frequent Closed Sequences,**” In Proc. of the 20th Int. Conf. on Data Engineering (ICDE'04), pp. 79-90, Boston, MA, USA, April 2004.
- [9] J. Wang, J. Han and J. Pei, “**CLOSET+: Searching for the Best Strategies for Mining Frequent Closed Itemsets,**” In Proc. of the 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'03), pp. 236-245, Washington, DC, August 2003.

- [10] K. Beyer and R. Ramakrishnan, "**Bottom-up computation of sparse and iceberg cubes,**" In Proc. 1999 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'99), pp. 359-370, Philadelphia, Pennsylvania, USA, June 1999.
- [11] L. Li, D. Zhai and F. Jin, "**GRG: An Efficient Method for Association Rules Mining on Frequent Closed Itemsets,**" In Proc. of the 2003 IEEE Int. Sym. on Intelligent Control (ISIC'03), pp. 854-859, Houston, Texas, October 2003.
- [12] M.J. Zaki and C.J. Hsiao, "**CHARM: An Efficient Algorithm for Closed Itemsets Mining,**" In Proc. 2002 SAIM Int. Conf. Data Mining (SDM'02), pp. 457-473, Arlington, VA, April 2002.
- [13] N. Pasquire, Y. Bastide, R. Taouil and L. Lakhal, "**Discovering Frequent Closed Itemsets for Association Rules,**" In Proc. of the 7th Int. Conf. on Database Theory(ICDT'99), pp. 398-416, Jerusalem, Israel, January 1999.
- [14] N. Pasquier, Y. Bastide, R. Taouil and L. Lakhal, "**Efficient Mining of Association Rules Using Closed Itemset Lattices,**" Journal of Information Systems, Vol 24, No.1, pp. 25-46, 1999.
- [15] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "**Mining Based for Associate Rules using Galois Closed Sets,**" In Research Report, Computer Science Dept., Nice Sophia Antipolis University, 2000.
- [16] P. Tzvetkov, X. Yan and J. Han, "**TSP: Mining Top-k Closed Sequential Patterns,**" In Proc. of the 3rd IEEE Int. Conf. on Data Mining (ICDM'03), pp. 347-354, Melbourne, Florida, USA, December 2003.
- [17] R. Afshar, **Mining Frequent Max and Closed Sequential Patterns.** M.Sc. Thesis, University of Alberta, Alberta, 2000.
- [18] V. Pudi and R. Haritsa, "**Generalized Closed Itemsets for Association Rule Mining,**" In Proc. of the 19th Int. Conf. on Data Engineering (ICDE'03), pp. 714-716, Bangalore, India, March 2003.
- [19] X. Yan, J. Han and R. Afshar, "**Clospan: Mining Closed Sequential Patterns in Large Database,**" In Proc. 2003 SIAM Int. Conf. Data Mining (SDM'03), pp. 166-177, San Francisco, CA, May 2003.

ภาคผนวก

ภาคผนวก ก.

โปรแกรมที่ใช้ในสร้างลำดับเหตุการณ์หลายมิติ

โปรแกรมสำหรับสร้างสารสนเทศหลายมิติ

```

/*-----
Program : generate multidimensional information
Input : number of transaction
        number of dimension
        name of file
output : text file

-----*/

#include <fstream>
#include <iostream>
#include <cstdlib>

using namespace std;

int main()
{
    int numTrans;
    int numDimen;
    char fileName[81];

    cout<<"input the number of transaction = "; //input data
    cin>>numTrans;
    cout<<"input the number of dimension = ";
    cin>>numDimen;
    cout<<"input the name of file = ";
    cin>>fileName;

    ofstream dataFile(fileName,ios::out);
    if(dataFile.fail()) //check file error

```

```

    {
        cout<<"The file could not be created";
    }
else
{

    int d;
    int i,j;
    dataFile<<numDimen<<" ";
    for(i=0;i<numTrans;i++)
    {
        for(j=0;j<numDimen;j++)
        {
            d = rand() % numDimen + (numDimen*j)+1;//random
number

            dataFile<<d; //write to file
            dataFile<<" ";
        }
        if(i<numTrans-1)
        {
            dataFile<<"\n"; //end of line in file
            dataFile<<numDimen<<" ";
        }
    }

}

dataFile.close();
return 0;
}

```

โปรแกรมสำหรับสร้างลำดับเหตุการณ์

1. ดาวน์โหลดโปรแกรมสังเคราะห์ข้อมูลของไอบีเอ็มเพื่อทำการสร้างลำดับเหตุการณ์จากเว็บไซต์ <http://www.almaden.ibm.com/software/quest/Resources/> ซึ่งมีไฟล์ทั้งหมด 14 ไฟล์

- Makefile
- command.C
- dist.C
- dist.h
- expdev.C
- gammln.C
- gasdev.C
- gen.C
- gen.h
- glob.h
- main.C
- poidev.C
- ran0.C
- ran1.C

2. ทำการคอมไพล์โปรแกรมบน g++ (โปรแกรม cygwin) โดยใช้คำสั่ง make

3. ใช้คำสั่ง `gen lit|tax|seq -help` เพื่อตรวจสอบละเอียดในการสร้างข้อมูล เช่น คำสั่ง `gen seq -help` เป็นคำสั่งตรวจสอบละเอียดของพารามิเตอร์ที่ใช้ในการสร้างลำดับเหตุการณ์ ซึ่งจะแสดงรายละเอียดดังนี้

Command Line Options:

- ncust number_of_customers_in_000s (default: 100)
- slen avg_trans_per_customer (default: 10)
- tlen avg_items_per_transaction (default: 2.5)

-nitems number_of_different_items_in_000s (default: 10)
 -rept repetition-level (default: 0)
 -seq.npats number_of_seq_patterns (default: 5000)
 -seq.patlen avg_length_of_maximal_pattern (default: 4)
 -seq.corr correlation_between_patterns (default: 0.25)
 -seq.conf avg_confidence_in_a_rule (default: 0.75)

-lit.npats number_of_patterns (default: 25000)
 -lit.patlen avg_length_of_maximal_pattern (default: 1.25)
 -lit.corr correlation_between_patterns (default: 0.25)
 -lit.conf avg_confidence_in_a_rule (default: 0.75)

-fname <filename> (write to filename.data and filename.pat)
 -ascii (Write data in ASCII format; default: False)
 -version (to print out version info)

4. ใช้คำสั่ง `gen lit|tax|seq [options]` ใช้สำหรับการสร้างลำดับเหตุการณ์และกลุ่มข้อมูลโดยดูรายละเอียดของพารามิเตอร์จากการใช้คำสั่งในข้อ 3

5. ตัวอย่างคำสั่งที่ใช้ในการสร้างลำดับเหตุการณ์

```
gen seq -ncust 10 -tlen 2.5 -nitems 10 -seq.patlen 8 -lit.patlen 1.25 -fname
C10N10T2.5S8I1.25
```

ประวัติผู้เขียน

ชื่อ – นามสกุล	นางสาวพนิดา ทรงรัมย์
วัน เดือน ปี เกิด	วันที่ 25 เมษายน พ.ศ. 2521 ที่จังหวัดบุรีรัมย์
ที่อยู่	21 หมู่ 2 ตำบลเสม็ด อำเภอเมือง จังหวัดบุรีรัมย์ 31000
ประวัติการศึกษา	2543 วิทยาศาสตรบัณฑิต เกียรตินิยมอันดับ 2 สาขาวิทยาการคอมพิวเตอร์ จากภาควิชาวิทยาการคอมพิวเตอร์ คณะวิศวกรรมศาสตร์และวิทยาการคอมพิวเตอร์ มหาวิทยาลัยมหาสารคาม
ประสบการณ์ทำงาน	
ปัจจุบัน	อาจารย์พิเศษ สังกัดคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัย กรุงเทพ ผู้ช่วยสอน สถาบันศูนย์ศึกษาต่อเนื่องจุฬาฯ
พ.ศ. 2545-พ.ศ 2546	อาจารย์ผู้ช่วยสอน ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาการ สารสนเทศ มหาวิทยาลัยมหาสารคาม
ทุนการศึกษา	
พ.ศ. 2547-2548	ทุนแลกเปลี่ยนสารสนเทศไทย-อินเดีย จากกระทรวงวิทยาศาสตร์และ เทคโนโลยี
พ.ศ. 2546-2548	ทุนพัฒนาอาจารย์สาขาขาดแคลน สาขาวิทยาการคอมพิวเตอร์ สังกัด มหาวิทยาลัยมหาสารคาม