วารสารเทคโนโลยีสารสนเทศ
Information Technology Journal

ISSN 1685-8573

Vol. 12 No. 1 January - June 2016

บทความวิจัย

การสืบค้นผู้มีอิทธิพลและผู้ถูกครอบงำบนเฟสบุ๊ค
พนิดา ทรงรัมย์ 1

ระบบแคปช่าเสียงเชิงคำถามกับรูปแบบรหัสคำตอบสำหรับผู้พิการทางสายตา
เรืองไชย ไทรประสพสุข อมรชัย มโนปิยอนันต์ และ ณัฐธนนท์ หงส์วริทธิ์ธร 11

การพัฒนาชุดฝึกอบรมการเขียนโปรแกรมควบคุมหุ่นยนต์ด้วยภาษาซี สำหรับนักศึกษาครุศาสตร์อุตสาหกรรม
กิตติ เสือแพร และ พัฒพงษ์ อมรวงศ์ 18

การกำหนดแบนด์วิธอย่างยุติธรรมสำหรับควบคุมการเชื่อมต่ออินเตอร์เน็ตแบบทวีคูณด้วยนโยบายระดับกั้นค่าสูงสุดและต่ำสุด
อธิคม ศิริ เกศสุดา คำลือวงค์ ธนาวุฒิ ธนวาณิชย์ และกมล บุญล้อม 24

การย้ายถิ่นแบบปรับตัวสำหรับจีเนติกอัลกอริทึมแบบกระจาย
ชัยวิวัฒน์ จันต๊ะสาร และ พรเทพ โรจนวสุ 30

การทำเหมืองข้อมูลกฎความสัมพันธ์ด้วยวิธีค่าถ่วงน้ำหนักสูงสุดจากกราฟบริบูรณ์แบบสมมาตร
ประมูล สุขสกาวผ่อง และ พยุง มีสัจ 37

การกระจายภาระงานด้วยหลักโครงสร้างการรวมกลุ่มที่ดีที่สุด
พาณิชย์ สุดโคต ฉัตรตระกูล สมบัติธีระ และ สวนิต วัฒนศักดากุล 44

การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อย
สำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต
ภรัณยา ปาลวิสุทธิ์ 54

การพัฒนาระบบสารสนเทศสำหรับตรวจสอบสถานะเครื่องคอมพิวเตอร์ในเครือข่ายด้วยเทคนิคการเผยแพร่และการติดตาม
เอกรินทร์ วทัญญูเลิศสกุล 64

Research Paper

A Model Design for a Web-based Learning System using Project-based Learning of Imagineering to Enhance Creative Construction of Multimedia Skills and Cooperative Skills
Pinanta Chatwattana and Rattanakorn Phadungthin 74

การทำเหมืองข้อมูลกฎความสัมพันธ์ด้วยวิธีค่าถ่วงน้ำหนักสูงสุดจากกราฟบริบูรณ์แบบสมมาตร
Associations Rule Mining By Using Top Weight of Complete Symmetric Digraphs

ประมูล สุขสกาวผ่อง (Pramool Suksakaophong) และพยุง มีสัจ (Phayung Meesad)
คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
[1]pramool7@yahoo.com, [2]pym@kmutnb.ac.th

บทคัดย่อ

การทำเหมืองข้อมูลกฎความสัมพันธ์จากรายการซื้อขายเพื่อใช้ในการทำรายการส่งเสริมการขายแนะนำรายการสินค้าที่มักจะซื้อด้วยกันบ่อย แต่การหากฎความสัมพันธ์ต้องอาศัยการกำหนดค่าสนับสนุนขั้นต่ำที่เหมาะสม เพราะถ้ากำหนดค่าน้อยไป จำนวนกฎที่ได้ก็จะมากและถ้ากำหนดค่ามากไปจำนวนกฎที่ได้ก็จะน้อย ทำให้ผู้ใช้ยากจะหาค่าที่เหมาะสมได้ในแต่ละชุดข้อมูล จากปัญหาดังกล่าวผู้วิจัยนำเสนอวิธีการใหม่ในการเก็บความถี่การซื้อขายในรูปแบบกราฟบริบูรณ์สมมาตรแบบมีทิศทางเพื่อนำค่าถ่วงน้ำหนักสูงสุดไปคำนวณหากฎความสัมพันธ์ จากค่าสนับสนุนสูงสุดได้และใช้ได้กับทุกชุดข้อมูล

คำสำคัญ: กฎความสัมพันธ์ กราฟบริบูรณ์ เมทริกซ์ประชิด เหมืองข้อมูล

Abstract

Association rules mining from transaction data can be used to recommend the items that are often purchased together frequently. However, it is difficult to set minimum support threshold. If the minimum support threshold is set too high, then there may be only a small or even no result. If the threshold is set too low, it may generate many uninteresting associations. In addition, each supporting a different set of data, enabling users to find the optimal difficult. This paper presents a new approach to the collection frequency by using top weight of complete symmetric digraphs. Using the top weight, the association rule with the maximum support can be calculated and it works with any dataset.

Keyword: Association rule mining, Complete Symmetric Digraphs, Adjacency matrix, Data mining.

1. บทนำ

การทำเหมืองข้อมูลค้นหากฎความสัมพันธ์ (Association rule mining) จากรายการซื้อขายสินค้า (Marker Basket Transaction) โดยกฎความสัมพันธ์สามารถสร้างขึ้นจากเซตรายการความถี่ (Frequent Itemsets) นำมาเปรียบเทียบกับค่าสนับสนุนขั้นต่ำ (Minimum Support) และค่าความเชื่อมั่นขั้นต่ำ (Minimum Confidence) จากในชุดข้อมูลขนาดใหญ่ แต่การค้นหากฎความสัมพันธ์จากข้อมูลที่มีรายการสินค้ามากจะต้องใช้เวลาในการคำนวณมากและปัญหาที่สำคัญในการค้นหากฎความสัมพันธ์จากข้อมูลรายการซื้อขายสินค้าคือ ค่าที่เหมาะสมของค่าสนับสนุนขั้นต่ำ เพราะถ้ากำหนดค่าน้อยไป จำนวนกฎที่ได้ก็จะมากและถ้ากำหนดค่ามากไปจำนวนกฎที่ได้ก็จะน้อยหรือไม่สามารถหากฎความสัมพันธ์ได้

ขั้นตอนวิธีการค้นหากฎความสัมพันธ์ต้องอาศัยการกำหนดค่าสนับสนุนขั้นต่ำที่เหมาะสมเพื่อลดเวลาในการประมวลผลและตัดรายการที่มีความถี่น้อยๆออกไปทำให้ไม่สามารถหากฎความสัมพันธ์ในรายการสินค้าที่มีความถี่น้อยแต่มีมูลค่าสูงซึ่งเป็นที่ต้องการของฝ่ายการตลาดในการจัดทำรายการส่งเสริมการขาย ผู้วิจัยจึงคิดค้นขั้นตอนวิธีใหม่เพื่อให้เหมาะกับการใช้งานในพาณิชย์อิเล็กทรอนิกส์โดยใช้พื้นฐานจากทฤษฎีกราฟและกราฟบริบูรณ์สมมาตรแบบมีทิศทางเพื่อให้ลดการใช้หน่วยความจำและทำงานได้รวดเร็ว ข้อดีของ

การนำทฤษฎีกราฟมาใช้ทำให้การจัดเก็บความถี่ของแต่ละรายการสินค้าเป็นอิสระต่อกันทำให้สามารถค้นหากฎความสัมพันธ์ในทุกรายการสินค้าได้

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 กฎความสัมพันธ์

การค้นหากฎความสัมพันธ์ มีขั้นตอนหลักๆ 2 ขั้นตอนคือ ขั้นแรก การนับรายการความถี่ (Frequent Items Generation) ที่เกิดขึ้นบ่อยทั้งหมดก่อนโดยเซตรายการต้องมีค่าความถี่มากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำตามที่กำหนด ขั้นตอนที่สองคือ ค้นหากฎความสัมพันธ์ของกลุ่มข้อมูลที่ปรากฏร่วมกันบ่อย (Rule Generation) ที่มีค่าความเชื่อมั่นของกฎไม่น้อยกว่าค่าที่กำหนด และนำมาสร้างเป็นกฎความสัมพันธ์ ขั้นตอนการทำงานของการหากฎความสัมพันธ์ดังแสดงในภาพที่ 1

ภาพที่ 1: ขั้นตอนการทำงานของการหากฎความสัมพันธ์

2.2 นิยามเบื้องต้นของกฎความสัมพันธ์

คำนิยามที่ใช้เรียกเพื่อใช้แทนสิ่งต่างๆ ในการค้นหากฎความสัมพันธ์ โดยกำหนดให้ $I = \{i_1, i_2, \ldots, i_n\}$ เป็นเซตของชิ้นข้อมูลหรือรายการสินค้า (items) ที่ประกอบด้วยรายการสินค้าจำนวน n ชิ้นรายการสินค้า และให้ $T = \{t_1, t_2, \ldots, t_m\}$ คือ เซตข้อมูลรายการซื้อขายสินค้า (Transaction) ในฐานข้อมูล T ที่ประกอบด้วยรายการข้อมูลมีสมาชิกจำนวน m รายการข้อมูล โดยที่แต่ละรายการ t_i จะประกอบไปด้วยรายการย่อยของรายการสินค้าจากกลุ่มข้อมูล I

ค่าสนับสนุนขั้นต่ำ หรือเขียนแทนด้วย *minsup* [1] การหาค่าสนับสนุน (Support) จากสมการ (1) คำนวณจากข้อมูลที่ปรากฏร่วมกันของ X และ Y ต่อสัดส่วนจำนวนรายการทั้งหมด (N)

$$s(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{N} \quad (1)$$

ค่าความเชื่อมั่นขั้นต่ำของกฎหรือเขียนแทนด้วย *minconf* [1] สามารถคำนวณค่า Confidence จากสมการ (2) รายการ X และ Y ปรากฏร่วมกัน ต่อสัดส่วนรายการ X

$$c(X \rightarrow Y) = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (2)$$

จำนวนกฎที่เป็นไปได้จะคำนวณได้จากสมการที่ (3)

$$R = \sum_{k-1}^{d-1} \left[\binom{d}{k} \times \sum_{j=1}^{d-k} \binom{d-k}{j} \right] \quad (3)$$

เมื่อค่า d คือจำนวนรายการ สมมุติว่ามี 4 รายการ จำนวนกฎที่เป็นไปได้จะเท่ากับ 50 กฎ แต่ถ้ามี 6 รายการ จะได้ 602 กฎ : ซึ่งเป็นการเพิ่มขึ้นในลักษณะเอ็กซ์โพเน็นเชียล (Exponential) ยิ่งมีรายการสินค้ามาก จำนวนกฎจะมากตามไปด้วย

2.3 การใช้ทฤษฎีเซตในการหากฎความสัมพันธ์

ขั้นตอนวิธี Apriori (Apriori Algorithm) [1] เป็นขั้นตอนวิธีที่ได้รับการยอมรับและเป็นที่รู้จักในการค้นหากฎความสัมพันธ์โดยเซตที่มีความถี่มากกว่าค่าสนับสนุนขั้นต่ำเซตย่อยจะมีความถี่ที่มากด้วย และสามารถตัดเซตรายการที่มีความถี่ต่ำออก ขั้นตอนวิธีในการค้นหาเซตรายการที่ปรากฏร่วมกันบ่อยและนำมาสร้างกฎความสัมพันธ์ ข้อดีของขั้นตอนวิธี Apriori คือ สามารถทำงานได้ดีหากกำหนดค่าสนับสนุนขั้นต่ำมีค่ามากๆ มีขนาดของฐานข้อมูลขนาดเล็ก และมีจำนวนของเซตรายการน้อย ส่วนข้อเสีย คือต้องอ่านข้อมูลจากฐานข้อมูลหลายครั้ง จึงมีผู้คิดขั้นตอนวิธีใหม่โดยการนับข้อมูลในโครงสร้าง FP-Tree โดยใช้ชื่อขั้นตอนวิธี FP-Growth [2] ที่ใช้การอ่านข้อมูลจากฐานข้อมูลเพียง 2 ครั้ง ทำให้ทำงานได้เร็วขึ้น

2.4 การใช้ทฤษฎีกราฟในการหากฎความสัมพันธ์

การค้นหากฎความสัมพันธ์ด้วยวิธีกราฟ จะจำลองวิธีการเก็บความถี่ในรูปของโหนด (Node) และเส้นเชื่อม (Edge) และค้นหาของข้อมูลที่ปรากฏร่วมกันบ่อยในรูปแบบของกราฟย่อย ซึ่งมีปริมาณข้อมูลมาก ขั้นตอนวิธี FP-Growth Graph [3] จึงใช้วิธีของ FP-Growth ในการค้นหากฎความสัมพันธ์

การใช้เมทริกซ์ประชิด (Adjacency Matrix) [4] เป็นการจำลองข้อมูลกราฟในรูปของข้อมูลในเมริกซ์และใช้คุณสมบัติการบวกกันของเมทริกซ์เพื่อค้นหาความถี่ของข้อมูลที่ปรากฏร่วมกันบ่อย ขั้นตอนวิธี TDB [5] จะใช้วิถี (Path) ในการค้นหา นอกจากนี้ยังมีการใช้อาร์เรย์ (Array) [6] และการใช้กราฟบริบูรณ์ (Completed Graph) โดยขั้นตอนวิธี GARM [9] จะแปลงแต่ละรายการซื้อขายเป็นกราฟบริบูรณ์ ทำให้สามารถเพิ่มเติมรายการข้อมูล (Dynamic Update) ในภายหลังได้ และ

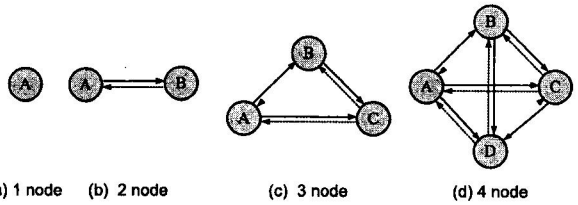
การใช้รูปแบบเสมือนกราฟบริบูรณ์ [10] โดยใช้เมทริกซ์ในการเก็บข้อมูล

2.5 ปัญหาการกำหนดค่าที่เหมาะสม สำหรับค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นของกฎ

ปัญหาสำคัญของการหากฎความสัมพันธ์คือการกำหนดค่าที่เหมาะสมเพื่อค้นหากฎความสัมพันธ์ ซึ่งมีค่าระหว่าง 0 กับ 1 ตัวอย่างค่าเริ่มในโปรแกรมด้านเหมืองข้อมูล เช่น WEKA กำหนดค่าเริ่มต้นของขั้นตอนวิธี Apriori ด้วยค่า *minsup* = 0.1 และ *minconf* = 0.9 ซึ่งไม่สามารถนำไปใช้ได้กับทุกชุดข้อมูล ต้องเลือกค่าหลายครั้งกว่าจะใช้ค้นหากฎความสัมพันธ์ได้ และได้มีจะมีผู้วิจัยได้ใช้ค่าถ่วงน้ำหนักเช่น WARM [8] และขั้นตอนวิธี Top-k [9] เพื่อแก้ปัญหาดังกล่าว

3. วิธีดำเนินงานวิจัย

จากปัญหาการหาค่าสนับสนุนขั้นต่ำ ผู้วิจัยจึงเลือกใช้ทฤษฎีกราฟเพราะมีความยืดหยุ่นและแสดงถึงการเชื่อมโยงของข้อมูลได้ดี ขั้นแรกจะทำการแปลงรายการซื้อขายในรูปแบบของกราฟบริบูรณ์แบบสมมาตร แทนสินค้าด้วยโหนดและความสัมพันธ์ระหว่างโหนดด้วยเส้นเชื่อม การเชื่อมต่อลักษณะนี้จะเป็น มีการเชื่อมต่อทั้งสองทิศทาง ซึ่งเหมาะกับคุณลักษณะของสินค้าที่การซื้อขายสินค้าไม่มีลำดับก่อนหลัง เนื่องจากในการซื้อขายกันจริงๆ ข้อมูลลำดับรายการซื้อขายอาจจะมีการสลับกันอยู่ จึงต้องเก็บข้อมูลทั้งสองทิศทางดังภาพที่ 2



ภาพที่ 2: กราฟบริบูรณ์แบบสมมาตร

ข้อมูลรายการซื้อขายจะแปลงข้อมูลในรูปเมทริกซ์ มีค่าแทนรายการสินค้าด้วย 1 และไม่มีรายการสินค้าด้วย 0

ตารางที่ 1: ข้อมูลรายการซื้อขายสินค้า

TID	Items Bought
01	{Apple, Bread, Cake}
02	{Apple, Bread, Cake, Diaper}
03	{Apple, Cake}
04	{Cake, Diaper}
05	{Bread, Cake, Diaper}
06	{Bread, Cake }
07	{Bread, Diaper}
08	{Bread, Cake, Diaper}

$$TID = \begin{Bmatrix} TID & A & B & C & D \\ 01 & 1 & 1 & 1 & 0 \\ 02 & 1 & 1 & 1 & 1 \\ 03 & 1 & 0 & 1 & 0 \\ 04 & 0 & 0 & 1 & 1 \\ 05 & 0 & 1 & 1 & 1 \\ 06 & 0 & 1 & 1 & 0 \\ 07 & 0 & 1 & 0 & 1 \\ 08 & 0 & 1 & 1 & 1 \end{Bmatrix}$$

3.1 รายการข้อมูลซื้อขายในตะกร้าสินค้า

ตัวอย่างรายการตะกร้าสินค้าในตารางที่ 1 ประกอบด้วยลำดับรายการซื้อ (Transaction ID) และ รายการสินค้า (Items) สามารถจัดอยู่ในรูปแบบเมทริกซ์ในตัวแปร TID ที่เก็บค่าแต่ละรายการซื้อขายสินค้าในรูปแบบ 0 กับ 1

3.2 การนับความถี่ของรายการซื้อขายสินค้า

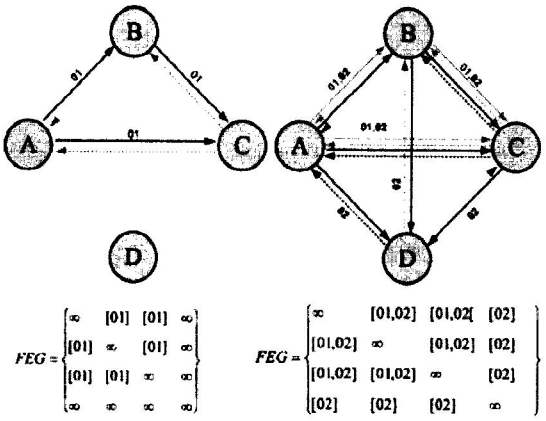
ขั้นตอนวิธีในการหากฎความสัมพันธ์จากรายการตะกร้าสินค้า จะทำการนับความถี่โดยในแต่ละลำดับรายการซื้อขายโดยจำลองข้อมูลรายการเป็นเมทริกซ์ประชิด (adjacency matrix) และทำการเก็บข้อมูลในเซลล์อาร์เรย์ (Cell Array, { }) โดย-ขั้นตอนวิธีการทำงานรวม ดังแสดงในภาพที่ 3

Algorithm1 FEG Frequency-Edge-Graph Construction

```
load data to transaction[]
[M, N ] = Transaction[]
for i=1 to M
    for j= 1 to N   /*count items in each TID */
            if transaction(i,j)==1
                Items[]= j
            end
            for k = 1 to size(Items)
                    FEG{}= [Items{M, N}, i]
            end
    end
end
```

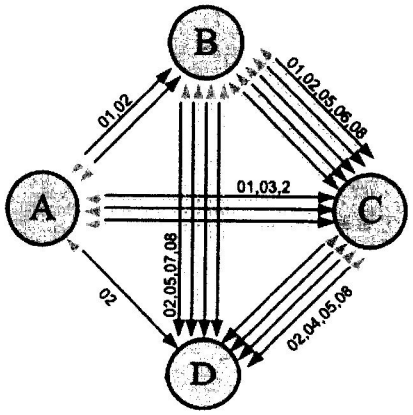
ภาพที่ 3: ขั้นตอนวิธีการนับความถี่

เริ่มแรกจะอ่านและเก็บข้อมูลรายการซื้อขายและเก็บในตัวแปร transaction[] จากนั้นเริ่มวนรอบในแต่ละรายการ เลือกเฉพาะรายการสินค้าที่มีค่าเป็น 1 นำมาสร้าง เมทริกซ์ประชิดในรูปของกราฟบริบูรณ์แบบสมมาตรเพื่อนับความถี่ของแต่ละรายการสินค้าที่มีการซื้อขายด้วยกันในตัวแปร FEG{}



ภาพที่ 4: การจำลองข้อมูลรายการซื้อขายที่ 01 และ 02

จากภาพที่ 4 เริ่มอ่านข้อมูลรายการซื้อขายที่ 01 มีรายการสินค้า items = {A, B, C} นำมาสร้างเมทริกซ์ประชิดแบบสมมาตรและเก็บค่าลำดับรายการซื้อขาย (TID = 01) ลงตัวแปร FEG {} ต่อด้วยรายการซื้อขายที่ 02 (TID = 02) มีรายการสินค้า items = {A, B, C, D} นำมาสร้างเป็นเมทริกซ์ประชิดแบบสมมาตร มาเก็บค่าลำดับรายการซื้อขายในเซลล์อาร์เรย์เดิม จำลองการทำงานเหมือน มัลติกราฟ (Multigraph)



ภาพที่ 5: จำลองการเก็บข้อมูล FEG{} เมื่อครบทุกรายการ

จากนั้นก็วนรอบเก็บข้อมูลลำดับรายการซื้อขายจนครบทุกรายการดังภาพที่ 5 จะได้ข้อมูลความถี่ในตัวแปร FEG{} ซึ่งจำลองการเก็บข้อมูลลำดับรายการซื้อขาย (TID) ในเส้นเชื่อมแบบมัลติกราฟ ข้อมูลที่เก็บในตัวแปรอาร์เรย์ดังตารางที่ 3

ตารางที่ 2: เมทริกซ์รายการข้อมูลความถี่ของสินค้า

	A	B	C	D	Top_W
A	∞	[01,04]	[01,03,04]	[04]	3
B	[01,04]	∞	[01,04,05,06,08]	[04,05,07,08]	5
C	[01,03,04]	[01,04,05,06,08]	∞	[02,04,05,08]	5
D	[04]	[04,05,07,08]	[02,04,05,08]	∞	4
Union	3	6	7	5	

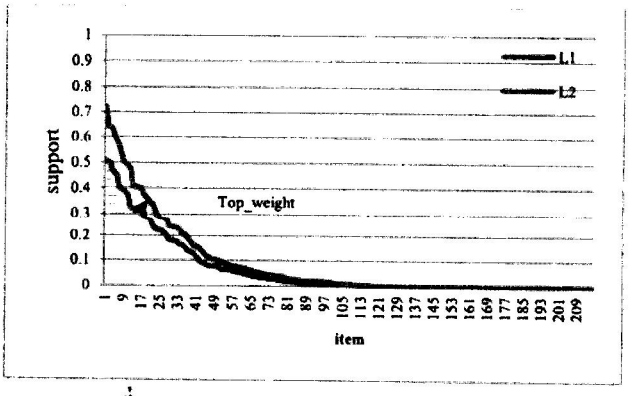
ข้อมูลที่จัดเก็บในตัวแปรเซลล์อาร์เรย์ FEG{} ดังกล่าวแสดงในตารางที่ 2 เป็นเซลล์อาร์เรย์เก็บลำดับรายการซื้อขายในเมทริกซ์ประชิด โดยการนับผลรวมแต่ละรายการสินค้าคือผลรวมการยูเนียน (Union)ในแต่ละคอลัมน์ ส่วนค่าถ่วงน้ำหนักสูงสุด (Top_W) คือการหาค่าผลรวมแต่ละคู่รายการที่มีค่าสูงสุดในแต่ละแถว เช่น แถวแรก {A,C} จะมีค่าสูงสุดคือ 3 เมื่อหาครบทุกแถวจะนำมาเรียงข้อมูลใหม่ จะได้ค่าถ่วงน้ำหนักจากมากไปหาน้อย จะได้ Top_Weight = [5 5 4 3]

3.3 การหาค่าสนับสนุนจากค่าถ่วงน้ำหนักสูงสุด (Top Weight)

ตามคุณสมบัติแอนตี้โมโนโทน (Anti-Monotone property) เซตที่มีจำนวนข้อมูลมากกว่าไม่สามารถมีค่าค่าสนับสนุนมากกว่าเซตย่อยได้ตามสมการที่ (4)

$$\forall X, Y : (X \subseteq Y) \Rightarrow s(X) \geq s(Y) \qquad (4)$$

จากคุณสมบัติดังกล่าวได้นำไปทดลองกับชุดข้อมูล supermarket เพื่อหาค่าสนับสนุนของชุดมูล 1 รายการและ 2 รายการ ได้ผลแสดงในภาพที่ 6



ภาพที่ 6: ชุดข้อมูล supermarket เรียงตามค่าถ่วงน้ำหนัก

จากผลการทดลองแสดงให้เห็นว่าเซตย่อยที่สุดคือ 1 รายการ L1 มีค่ามากที่สุด และ L2 รองลงมา โดยกฎความสัมพันธ์จะเริ่มต้นจากความสัมพันธ์ของ 2 รายการ ดังนั้นจึงเลือกค่าถ่วงน้ำหนักของ 2 รายการเพื่อนำไปคำนวณค่าถ่วงน้ำหนักสูงสุด เพื่อใช้ในกำหนดค่า Top_Weight

3.4 การหากฎความสัมพันธ์

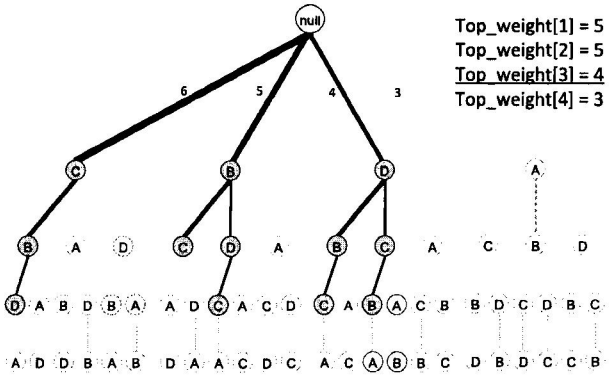
ผลจากการนับความถี่ในตารางที่ 2 แสดงความสัมพันธ์ของแต่ละคู่รายการสินค้าในรูปของเมทริกซ์ในตัวแปรเซลล์อาร์เรย์ FEG{} โดยเส้นเชื่อมจะเก็บค่ารายการความถี่ของคู่สินค้าที่ปรากฏร่วมกัน ขั้นตอนการค้นหากฎความสัมพันธ์ ดังแสดงในภาพที่ 7 เริ่มจากการหาค่าถ่วงน้ำหนักสูงสุด หาขนาดของตัวแปรเซลล์อาร์เรย์ FEG{} จากนั้นกำหนดการวนรอบข้อมูล โดยหาค่าสูงสุดในแต่ละแถว เมื่อได้ครบทุกแถวก็นำมาเรียงข้อมูลใหม่ จะได้ค่าถ่วงน้ำหนักที่เรียงจากมากไปหาน้อย ขั้นตอนที่ได้จะสามารถนำไปคำนวณค่าสนับสนุนตาม สมการที่ (1)

Algorithm2 Association rule Generation

```
[M, N ] = size(FEG)
for i=1 to M
   for j= 1 to N
      /*find Maximal Weight for each item */
         If ( sum(items(m,n)) > max_items
         max_items = items(m,n)
         end
   end
         Top_Weight[] = max_items
         Sort(Top_Weight)
end
[M, N ] = FEG{}
Top_Weight[] = d  /* define Top_Weight level */
minconf = d
for (L1 Items >= Top_Weight)
     for L1 to Ln Items
       sort(intersect(items(L1,L2, →Ln) ≥Top_Weitht
       and items(items(L1,L2, →Ln) ≥ minconf
       Asso_rule[] = (items(L1,L2, →Ln)
     end
end
```

ภาพที่ 7: ขั้นตอนการกฎความสัมพันธ์

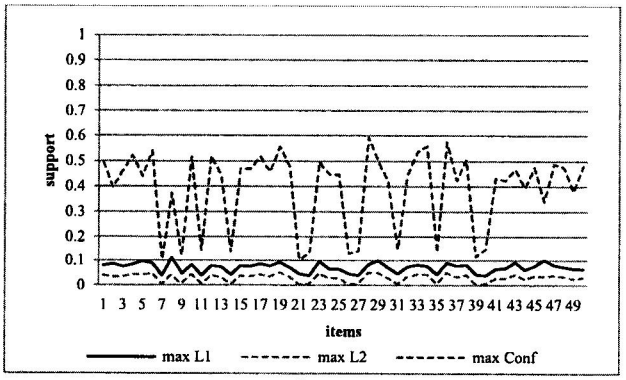
เมื่อได้ค่าสนับสนุนแต่ละรายการสินค้าแล้วนำมาเรียงข้อมูลและค้นหาลำดับถัดไป นำมาเปรียบเทียบกับค่าถ่วงน้ำหนักถ้ามีค่ามากกว่าก็นำไปเปรียบเทียบกับค่าความเชื่อมั่นของกฎถ้ามากกว่าก็จะนับเป็นกฎความสัมพันธ์ได้ ขั้นตอนวิธีหากฎความสัมพันธ์โดยการตัดรายการความถี่ที่น้อยกว่าค่าถ่วงน้ำหนักลำดับที่ 3 คือมีค่าความถี่เท่ากับ 4 ดังภาพที่ 8 ทำให้ได้รายการที่เหลือเฉพาะค่าสนับสนุนที่สูงนำไปสร้างกฎความสัมพันธ์โดยวนรอบแต่ละรายการมาคำนวณค่าความเชื่อมั่นตามที่กำหนด



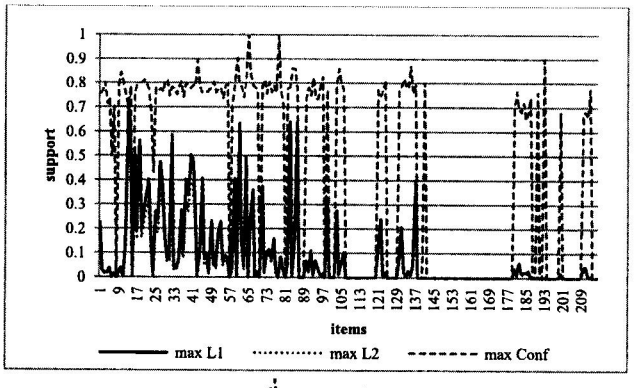
ภาพที่ 8: ผลการตัดรายการความถี่ที่ Top_weight = 4

4. ผลการดำเนินงาน

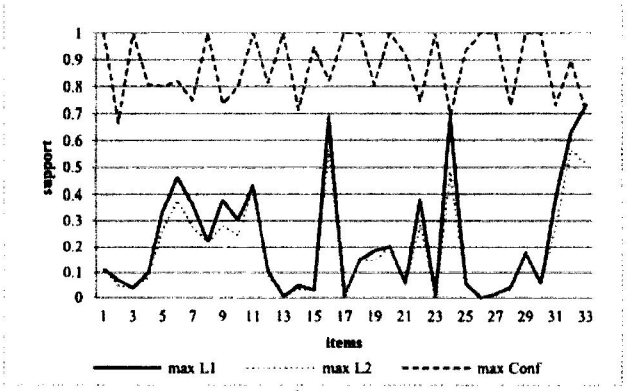
จากการออกแบบขั้นตอนวิธีที่ได้ นำมาเขียนโปรแกรมจำลองการทำงานบนโปรแกรม MATLAB R2012a 32BIT นำมาคำนวณหาค่าสนับสนุนตามสมการ (1) จะได้ค่าสนับสนุนสูงสุดของ 1 รายการ (max L1) และค่าสนับสนุน 2 รายการ (max L2) และค่าความเชื่อมั่นตามสมการ (2) จะได้ค่าความเชื่อมั่นสูงสุด 2 รายการ (max Conf) นำผลที่ได้มาสร้างเป็นกราฟได้ดังภาพที่ 9 ถึง 11



ภาพที่ 9: ชุดข้อมูล Bakery



ภาพที่ 10: ชุดข้อมูล supermarket



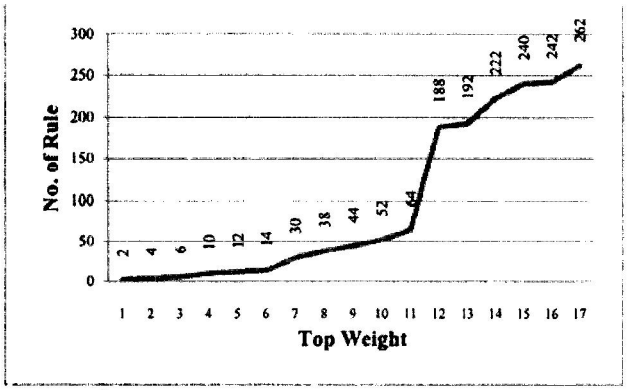
ภาพที่ 11: ชุดข้อมูล Chess

ภาพที่ 9 ชุดข้อมูล Bakery ประกอบด้วย 5,000 รายการซื้อขายสินค้าใน 50 สินค้า ค่าสนับสนุนแต่ละคู่สินค้า (max L2) จะมีค่า 0.0512 ส่วนภาพที่ 10 ชุดข้อมูล supermarket ประกอบไปด้วย 4,627 รายการซื้อขายสินค้าจาก 217 สินค้า ค่าสนับสนุนแต่ละคู่สินค้า (max L2) จะมีค่า 0.505079 และภาพที่ 11 ชุดข้อมูล Chess ประกอบไปด้วย 3,196 รายการซื้อขายสินค้าจาก 33 สินค้า ค่าสนับสนุนแต่ละคู่สินค้า (max L2) จะมีค่า 0.56383 จากการทดลองกับชุดข้อมูลต่างๆ ทำให้เป็นได้ว่า แต่ละชุดข้อมูลมีค่าสนับสนุนและค่าความเชื่อมั่นในระดับที่แตกต่างกัน สามารถนำมาเปรียบเทียบได้ดังตารางที่ 4 โดยเลือกมาเพียง 5 อันดับแรกเรียงจากค่าสูงสุด

ตารางที่ 3: ผลการคำนวณสนับสนุนสูงสุด 5 อันดับ

Dataset	Chess	Supermarket	Bakery
1	0.56383	0.505079	0.0512
2	0.56383	0.502485	0.0496
3	0.513454	0.49665	0.0472
4	0.480914	0.473525	0.044
5	0.431477	0.460125	0.0436

จากค่าสนับสนุนที่ได้นำไปหากฎความสัมพันธ์โดยทดลองกับชุดข้อมูล Supermarket เริ่มต้นนำสนับสนุนสูงสุดคือ 0.505079 จะได้ 2 กฎ และลำดับสอง 0.502485 ได้ 4 กฎ จนถึงอันดับที่ 5 คือ 0.460125 จะได้ 12 กฎ จำนวนกฎที่สัมพันธ์กับระดับค่าสนับสนุนจากสูงสุด แสดงในภาพที่ 12 ทำให้สมารถเลือกจำนวนกฎที่ต้องการอัตโนมัติโดยการกำหนดระดับของค่าสนับสนุนจากอันดับสูงสุด



ภาพที่ 12: จำนวนกฎต่อค่าสนับสนุนในชุดข้อมูล supermarket

5. สรุป

การจำลองข้อมูลในรูปแบบกราฟบริบูรณ์สามารถค้นหาค่าสนับสนุนสูงสุดและค่าความเชื่อมั่นสูงสุดได้ที่ใช้ได้กับทุกชุดข้อมูล และนำค่าดังกล่าวมาหากฎความสัมพันธ์โดยสามารถเลือกกฎความสัมพันธ์ที่มีค่าสนับสนุนและค่าความเชื่อมั่นที่จากระดับสูงสุดได้ และจากการทดลองค่าสนับสนุนแต่ละค่ามีความละเอียดถึงทศนิยม 6 ตำแหน่ง จึงเป็นการยากที่ผู้ใช้งานจะคาดคะเนเพื่อเลือกค่าที่เหมาะสมด้วยตนเองได้ เพราะบางชุดข้อมูลเช่น Bakery ระดับค่าสนับสนุนห่างกันแค่เพียง 0 ถึง 0.1 ทำให้การหาค่าที่เหมาะสมด้วยการกำหนดค่าเองเป็นไปได้ยาก การเลือกจากค่าสนับสนุนโดยกำหนดระดับจากสูงสุด ทำให้จำกัดปริมาณกฎความสัมพันธ์ และได้กฎความสัมพันธ์ที่มีค่าสนับสนุนและความเชื่อมั่นที่สูงได้

เอกสารอ้างอิง

[1] R. Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," *Proceedings of the 20th International Conference on Very Large Data Bases*, 672836: Morgan Kaufmann Publishers Inc.;. pp. 487-99, 1994

[2] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, ACM: Dallas, Texas, USA. pp. 1-12, 2000

[3] V. Tiwari et al., "Association rule mining: A graph based approach for mining frequent itemsets," *Networking and Information Technology (ICNIT)*, 2010

[4] A. Inokuchi,., T. Washio, and H. Motoda, "Complete Mining of Frequent Patterns from Graphs: Mining Graph Data," *Machine Learning*,. Vol 50(3): p. 321-354, 2003

[5] S. Arumugam,. and S. Sabeen, "Association Rule Mining using Path Systems in Directed Graphs," *International Journal of Computers, Communications & Control*, vol 8(6): pp. 791-799, 2013

[6] L. Naili, and M. Lei, "Discovering frequent itemsets an improved algorithm of directed graph and array," *4th IEEE International Conference on. 2013Software Engineering and Service Science (ICSESS)*, 2013

[7] P. Amal Dev, N.V. Sobhana, and P. Joseph, GARM: A Simple Graph Based Algorithm For Association Rule Mining. International Journal of Computer Applications,. Vol 76(1-16): pp. 1, 2013

[8] A. Salam, and M. Sikandar Hayat Khayal, "Mining top–k frequent patterns without minimum support threshold," *Knowledge and Information Systems*, Vol 30(1): pp. 57-86, 2012.

[9] R. Pears et al., "Weighted association rule mining via a graph based connectivity model," *Information Sciences*, Vol 218(0): pp. 61-84, 2013.

[10] P. Suksakaophong and P. Meesad, "Complete-Graph-Like Association Rule Mining,"*Joint Conference on ACTIS & NCOBA 2015"* Nakhon Phanom, Thailand, January 30-31, 2015 pp. 148-154, 2015