การสืบค้นผู้มีอิทธิพลและผู้ถูกครอบงำบนเฟสบุ๊ค
Mining of Influential Persons and Influenced Persons on Facebook

พนิดา ทรงรัมย์ (Panida Songram)*

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อค้นหากฎที่แสดงการมีอิทธิพลโดยใช้การสืบค้นกฎความสัมพันธ์แบบลำดับ กฎที่มีค่าความถี่และค่าความเชื่อมั่นเพียงพอจะถูกนำมาใช้ในการแสดงการมีอิทธิพลระหว่างผู้มีอิทธิพลและผู้ที่ถูกครอบงำ โดยกฎถูกสร้างขึ้นจากข้อมูลการโพสต์หรือการแสดงความคิดเห็นบนกลุ่มเฟสบุ๊คบนพื้นฐานแนวความคิดที่ว่าถ้าผู้มีอิทธิพลโพสต์หรือแสดงความคิดเห็นในหัวข้อใดๆ ก็ตาม ผู้ที่ถูกครอบงำจะแสดงความคิดเห็นในหัวข้อนั้นด้วย

คำสำคัญ: การมีอิทธิพล การถูกครอบงำ กฎความสัมพันธ์แบบลำดับ เครือข่ายสังคมออนไลน์

Abstract

This paper aims to discover influential rules by using sequential rule mining. The influential rules with sufficient frequency and confidence are exploited to show influential persons and their influenced persons. The rules are generated from post/comment dataset on Facebook groups based on the assumption that if influential persons post or commend any topics, the influenced persons always commend to the topics.

Keyword: Influence, Obsession, Sequential Rule Mining, Social Network.

1. บทนำ

การมีอิทธิพลเป็นแรงผลักดันหนึ่งที่ทำให้คนแสดงพฤติกรรมออกมา เช่น อิทธิพลของสังคมเป็นตัวผลักดันทำให้คนแสดงพฤติกรรมตามกัน [1] นอกจากนี้คนบางคนอาจมีอิทธิพลกับคนอื่นและทำให้คนอื่นกระทำหรือแสดงพฤติกรรมตาม งานวิจัยจำนวนมากจึงได้ทำการศึกษากลุ่มคนที่มีอิทธิพลในด้านต่างๆ เช่น การตรวจสอบการยอมรับของแพทย์ นวัตกรรมการเกษตร และการตลาด เป็นต้น โดยเฉพาะด้านการตลาดมีการศึกษาการมีอิทธิพลอย่างจริงจัง เช่น Domingos และ Richardon [2] นำเสนอโมเดลเครือข่ายลูกค้าเพื่อระบุกลุ่มลูกค้าที่มีอิทธิพลต่อผู้อื่นในการซื้อสินค้า โมเดลเครือข่ายลูกค้าที่นำเสนอถูกสร้างขึ้นโดยใช้ Markov random field เพื่อแสดงความสัมพันธ์ระหว่างลูกค้า จากนั้นพัฒนาวิธีการทางเหมืองข้อมูลเพื่อขุดค้นหาคนที่มีอิทธิพลต่อการซื้อสินค้าจากฐานข้อมูลที่เป็นตัวกรองเชิงร่วมมือ (Collaborative filtering) Kempe และคณะ [3] วิเคราะห์ผู้มีอิทธิพลในการซื้อสินค้าโดยใช้ Linear Threshold Model และ Independent Cascade Model เพื่อแสดงให้เห็นถึงพฤติกรรมของลูกค้าในการซื้อสินค้า จากการสังเกตพฤติกรรมการซื้อสินค้าของลูกค้าโดยใช้โมเดลดังกล่าวแสดงให้เห็นว่าคนใกล้เคียงมีอิทธิพลกับการซื้อสินค้า Leskovec และ คณะ [4] ทำการศึกษาและวิเคราะห์เครือข่ายการแนะนำจากคนไปสู่คน โดยเลือกใช้ข้อมูลการแนะนำหนังสือและการแนะนำวีดีโอมาเป็นกรณีศึกษา จากการ สังเกตเครือข่ายการแนะนำพบว่าลักษณะของเครือข่ายแนะนำมีอิทธิพลต่อรูปแบบการซื้อขายในกลุ่ม

นอกจากนี้นักวิจัยบางท่านยังได้ทำการศึกษาการมีอิทธิพลบนเครือข่ายสังคมออนไลน์ (Social Network) แทนที่จะทำการศึกษาจากแบบสอบถาม เนื่องจากเครือข่ายสังคมออนไลน์เข้ามามีบทบาทในชีวิตประจำวันมากขึ้น พฤติกรรมบางอย่างถูกแสดงออกมาบนเครือข่ายออนไลน์ เช่น Cha และคณะ [5] ทำการวิเคราะห์วิธีการกระจายรูปภาพที่ได้รับความนิยมบนเครือข่าย Flickr และแสดงให้เห็นว่า

*คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

ถ้าคนที่อัพโหลดรูปภาพเป็นคนที่มีอิทธิพลกับผู้อื่นจะทำให้การกระจายรูปภาพเป็นไปอย่างรวดเร็วและกว้างขวางขึ้น Trusove และคณะ[6] นำเสนอวิธีการระบุผู้ใช้ที่มีอิทธิพลมากที่สุดในการทำกิจกรรมบนเครือข่ายสังคมออนไลน์ โดยใช้ข้อมูลบันทึกกิจกรรม (Activity log) และประยุกต์ใช้ Bayesian เพื่อสร้างโมเดลสำหรับทำนายว่าใครเป็นผู้มีอิทธิพล Ballantine และคณะ [7] แสดงให้เห็นว่าการแสดงความคิดเห็นบนเครือข่ายสังคมออนไลน์เฟสบุ๊คมีอิทธิพลต่อแนวความคิดด้านบวกและด้านลบ Goyal และคณะ [8] ประยุกต์ใช้วิธีการสืบค้นเซตรายการความถี่ (Frequent itemset mining) เพื่อค้นหาผู้นำในเครือข่ายสังคมออนไลน์ โดยใช้ข้อมูลจาก Yahoo! Instant Messenger และพิจารณาการแท็ก การซื้อ การให้เรตติ้ง ของผู้ใช้ที่อยู่ในข้อมูล Yahoo! Movies นอกจากนี้ยังได้พัฒนาโปแกรมที่เรียกว่า GuruMine [9] เพื่อใช้ในการค้นหาผู้นำในการกระทำต่างๆ

จากงานวิจัยดังกล่าวจะเห็นได้ว่างานวิจัยส่วนใหญ่ทำการศึกษาการมีอิทธิพลโดยเน้นที่การค้นหาผู้มีอิทธิพลในกลุ่ม แต่อย่างไรก็ตามผู้มีอิทธิพลอาจจะมีอิทธิพลกับคนคนหนึ่งมาก แต่อาจจะมีอิทธิพลกับอีกคนหนึ่งน้อย ถึงแม้พวกเขาจะอยู่ในกลุ่มเดียวกันก็ตาม ดังนั้นงานวิจัยนี้จึงนำเสนอแนวความคิดในการสร้างกฎที่ใช้ในการแสดงการมีอิทธิพลโดยประยุกต์ใช้การสืบค้นกฎความสัมพันธ์แบบลำดับ (Sequential rule mining) โดยกฎที่ได้จะสามารถระบุได้ว่าใครมีอิทธิพลกับใครมากน้อยแค่ไหน โดยพิจารณาจากค่าความเชื่อมั่นของกฎ นอกจากนี้งานวิจัยนี้ยังค้นหากฎการมีอิทธิพลโดยใช้ข้อมูลการแสดงความคิดเห็นบนกลุ่มเฟสบุ๊ค ซึ่งเป็นข้อมูลที่เกิดจากพฤติกรรมของผู้ใช้จริงๆ และวิธีการที่นำเสนอในงานวิจัยนี้สามารถนำไปประยุกต์ใช้ในด้านต่างๆ ได้ เช่น ถ้าต้องการสืบค้นว่าใครมีอิทธิพลทางด้านการเมืองกับใครบ้าง สามารถนำแนวคิดในงานวิจัยนี้ไปประยุกต์ใช้กับข้อมูลแสดงความคิดเห็นบนกลุ่มการเมืองเป็นต้น

2. ทฤษฎีที่เกี่ยวข้อง

การสืบค้นกฎความสัมพันธ์ (Association rule mining) ถูกนำเสนอขึ้นครั้งแรกโดย Agrawal และคณะ [10] เพื่อหาความสัมพันธ์ของข้อมูลจากฐานข้อมูลขนาดใหญ่ โดยกฎที่ได้จะต้องผ่านค่าสนับสนุนขั้นต่ำ (Minimum support threshold: min_supp) และค่าความเชื่อมั่นขั้นต่ำ (Minimum confidence threshold: min_conf) การสร้างกฎความสัมพันธ์แบ่งออกเป็น 2 กลุ่ม คือ การสร้างกฎความสัมพันธ์จากเซตรายการ (Itemset Mining) และการสร้างกฎความสัมพันธ์จากรูปแบบลำดับเหตุการณ์ (Sequential Pattern Mining) เรียกว่า กฎความสัมพันธ์แบบลำดับ (Sequential Rules) การสร้างกฎความสัมพันธ์ด้วยเซตรายการเหมาะกับงานที่ไม่ต้องการพิจารณาถึงลำดับการเกิดของข้อมูล แต่การสร้างกฎความสัมพันธ์ด้วยรูปแบบลำดับเหตุการณ์เหมาะกับงานที่พิจารณาเรื่องลำดับการเกิดของข้อมูล

การสร้างกฎความสัมพันธ์แบบลำดับในปัจจุบันแบ่งออกเป็น 2 วิธีหลักๆ คือ 1)สืบค้นลำดับเหตุการณ์ความถี่ (Frequent Sequential Patterns) ก่อนแล้วนำไปสร้างกฎความสัมพันธ์แบบลำดับโดยลำดับเหตุการณ์ที่ฝั่งซ้าย (Antecedent) และฝั่งขวา (Consequent) ของกฎสร้างมาจากลำดับเหตุการณ์ความถี่ และ 2) สร้างกฎความสัมพันธ์แบบลำดับจากสองเซตรายการ X และ Y ในรูปแบบ $X \rightarrow Y$ โดยที่สมาชิกของเซตรายการ X จะต้องเกิดก่อนเซตรายการ Y กฎที่สร้างขึ้นด้วยวิธีที่สองเป็นกฎที่สามารถทำนายข้อมูลได้ถูกต้องกว่ากฎที่สร้างขึ้นจากวิธีแรก [11] ขั้นตอนวิธีการสร้างกฎความสัมพันธ์แบบลำดับด้วยวิธีที่สอง เช่น RuleGrowth, CMRules และ ERMiner เป็นต้น

ขั้นตอนวิธี RuleGrowth [11] สร้างกฎความสัมพันธ์แบบลำดับบนพื้นฐานวิธี Pattern-growth คือ ทำการขยายรูปแบบของข้อมูลไปเรื่อยๆ โดยเริ่มจากสร้างกฎที่มีขนาด 1*1 (ฝั่งซ้ายและฝั่งขวาของกฎมีแค่รายการ (Item) เดียว) ที่มีค่าสนับสนุนไม่น้อยกว่าค่าสนับสนุนขั้นต่ำ แล้วทำการขยายกฎไปเรื่อยๆ โดยการสแกนฐานข้อมูลเพื่อหารายการที่สามารถนำมาขยายฝั่งซ้ายและฝั่งขวาของกฎ ขั้นตอนวิธีนี้สามารถสร้างกฎความสัมพันธ์แบบลำดับโดยตรงโดยที่ไม่ต้องสร้างกฎความสัมพันธ์แบบลำดับคู่แข่ง (Candidate sequential rule) โดยจะพิจารณาจากเงื่อนไขที่ว่า ถ้ารายการ i ถูกเพิ่มไปทางฝั่งซ้ายหรือฝั่งขวาของกฎ ค่าสนับสนุนของกฎจะมีค่าน้อยกว่าหรือเท่ากับค่าสนับสนุนของกฎเดิม ดังนั้นถ้าพบว่ากฎที่เพิ่มรายการ i เข้าไปแล้วทำให้ค่าสนับสนุนน้อยกว่าค่าสนับสนุนขั้นต่ำ กฎดังกล่าวจะไม่ถูกขยายต่อไปอีก นอกจากนี้ RuleGrowth ยังไม่อนุญาตให้ทำการขยายเซตรายการที่อยู่ฝั่งซ้ายของกฎหลังจากการขยายเซตรายการที่อยู่ฝั่งขวา

ของกฎ เพื่อลดปัญหาการสร้างกฎที่ซ้ำซ้อน และเพื่อหลีกเลี่ยงการคำนวณที่นำไปสู่กฎเดียวกัน ขั้นตอนวิธี RuleGrowth สามารถแสดงได้ดังภาพที่ 1

Algorithm: The RuleGrowth algorithm

Input: *S*: a sequence database, *min_supp* and *min_conf*: The two user-specified thresholds
Output: The set of valid sequential rules

1. Scan *S* once. For each item c found, record the sids (transaction ids) of sequences that contains *c* in a variable *sids*(*c*)
2. Scan *S* a second time and remove each item *c* such that $|sids(c)|/|S| \leq min\_supp$
3. for each pair of items *i*, *j*
4. $sids(i \rightarrow j) := \varnothing$; $sids(i \rightarrow j) := \varnothing$;
5. for each *sid* $s \in (sids(i) \cap sids(j))$
6. if *i* occurs before *j* in *s*, $sids(i \rightarrow j)$: $sids(i \rightarrow j) \cup \{s\}$
7. if *j* occurs before *i* in *s*, $sids(j \rightarrow i)$: $sids(j \rightarrow i) \cup \{s\}$
8. if $(|sids(i \rightarrow j)|/|S|) \geq min\_supp$ then
9. EXPANDLEFT($\{i\} \rightarrow \{j\}$, *sids*(*i*), $sids(i \rightarrow j)$)
10. EXPANDRIGHT($\{i\} \rightarrow \{j\}$, *sids*(*i*), *sids*(*j*), $sids(i \rightarrow j)$)
11. if $(|sids(i \rightarrow j)|/|S|) \geq min\_supp$ then
12. EXPANDLEFT($\{j\} \rightarrow \{i\}$, *sids*(*j*), $sids(i \rightarrow j)$)
13. EXPANDRIGHT($\{j\} \rightarrow \{i\}$, *sids*(*j*), *sids*(*i*), $sids(j \rightarrow i)$)
14. if($|sids(j \rightarrow i)|/|sids(j)| \geq min\_conf$ then output rule $\{j\} \rightarrow \{i\}$

ภาพที่ 1 *ขั้นตอนวิธี RuleGrowth [11]*

ขั้นตอนวิธี CMRules [12] เป็นขั้นตอนหนึ่งที่มีประสิทธิภาพและไม่ซับซ้อน ขั้นตอนวิธีนี้ทำการค้นหากฎความสัมพันธ์แบบลำดับบนพื้นฐานการสร้างกฎความสัมพันธ์จากเซตรายการ ดังนั้นขั้นตอนวิธีนี้จึงทำการสร้างกฎคู่แข่งขึ้นมาก่อนแล้วค่อยมาพิจารณาว่ากฎใดเป็นกฎที่ยอมรับได้ จากภาพที่ 2 แสดงขั้นตอนวิธี CMRules โดยเริ่มจากการแปลงฐานข้อมูลลำดับเหตุการณ์เป็นฐานข้อมูลเซตรายการ จากนั้นประยุกต์ใช้ขั้นตอนวิธีการสร้างกฎความสัมพันธ์ที่ชื่อว่า Apriori [10] เพื่อสร้างกฎความสัมพันธ์ โดยขั้นตอนวิธี Apriori จะเริ่มจากการค้นหาเซตรายการความถี่ที่มีค่าสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำโดยการขยายเซตรายการไปเรื่อยๆ แล้วตรวจสอบว่าเซตรายการดังกล่าวผ่านค่าสนับสนุนขั้นต่ำหรือไม่ จากนั้นใช้เซตรายการความถี่ในการสร้างกฎความสัมพันธ์ เมื่อได้กฎความสัมพันธ์แล้ว CMRules จะทำการคำนวณหาค่าสนับสนุนและค่าความเชื่อมั่นของแต่ละกฎความสัมพันธ์โดยจะพิจารณาลำดับการเกิดของเหตุการณ์ด้วย จากนั้นจะทำการตัดกฎความสัมพันธ์ที่ไม่ผ่านค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำออก

Algorithm: The CMRules algorithm

Input: *S*: a sequence database, *min_supp* and *min_conf*: The two user-specified thresholds
Output: The set of valid sequential rules

1. Consider S as a transaction database.
2. Find all association rules from the transaction database by appling the Apriori algorithm. Select rules having support $\geq min\_supp$ and confidence $\geq min\_conf$
3. Scan S to calculate the sequential support and sequential confidence of each association rule found in the previous step. Eliminate each rule *r* having support$<min_supp$ and confidence$<min_conf$
4. return the set of rules

ภาพที่ 2 *ขั้นตอนวิธี CMRules [12]*

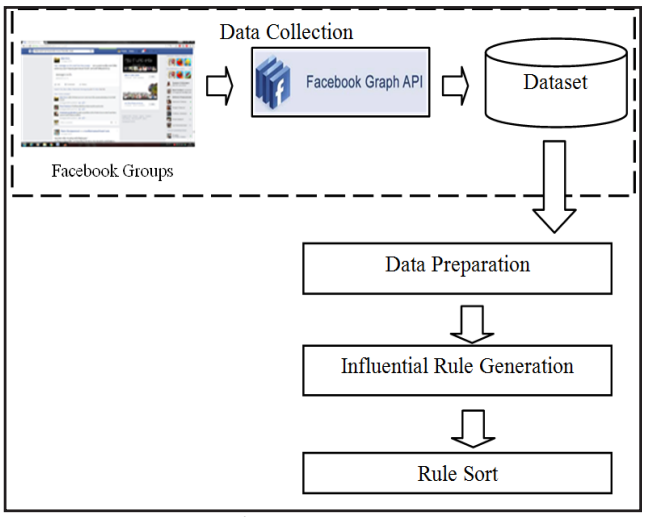
ขั้นตอนวิธี ERMiner [13] เป็นขั้นตอนวิธีที่มีประสิทธิภาพในเรื่องของความเร็วในการสืบค้นกฎความสัมพันธ์แบบลำดับ ขั้นตอนวิธีดังกล่าวทำงานบนพื้นฐานการค้นหาโดยใช้ Equivalence classes ซึ่งประกอบไปได้ Left equivalence class และ Right equivalence class กฎที่อยู่ใน Left equivalence class เดียวกัน คือ กฎที่มีเซตรายการฝั่งซ้ายเป็นเซตรายการเดียวกัน เช่น $\{a\} \rightarrow \{e,f\}$ และ $\{a\} \rightarrow \{b\}$ ถือว่าอยู่ใน Left equivalence class เดียวกัน ส่วนกฎที่อยู่ใน Right equivalence class เดียวกัน คือ กฎที่มีเซตรายการฝั่งขวาเป็นเซตรายการเดียวกัน เช่น กฎ $\{a\} \rightarrow \{e,f\}$ และ $\{b\} \rightarrow \{e,f\}$ อยู่ใน Right equivalence class เดียวกัน จากภาพที่ 3 แสดงขั้นตอน ERMiner โดยเริ่มจากการสแกนหากฎที่มีขนาด 1*1ในแต่ละ Equivalence classes จากนั้นทำการขยายกฎให้มีขนาดใหญ่ขึ้นโดยทำการยุบรวมกฎที่อยู่ Left equivalence class เดียวกันเข้าด้วยกันทีละคู่เพื่อขยายกฎฝั่งขวา เช่น กฎ $W \rightarrow X$ และ $W \rightarrow Y$ เมื่อยุบรวมจะได้ $W \rightarrow X \cup Y$ กฎทุกกฎที่สร้างขึ้นจะถูกนำไปคำนวณหาค่าสนับสนุน ถ้าค่าสนับสนุนของกฎมีค่ามากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำ กฎดังกล่าวจะถูกเพิ่มเข้าไปใน Left equivalence class อีกเพื่อทำการขยายกฎฝั่งขวาต่อไปและถ้ากฎผ่านค่าความเชื่อมั่นขั้นต่ำด้วย กฎจะถูกเลือกเป็นกฎความสัมพันธ์แบบลำดับที่ยอมรับได้ จากนั้นทำการยุบรวมกฎที่อยู่ Right equivalence class เดียวกันเข้าด้วยกันทีละคู่เพื่อขยายกฎฝั่งซ้าย เช่น กฎ $X \rightarrow W$ และ $Y \rightarrow W$ เมื่อยุบรวมจะได้ $X \cup Y \rightarrow W$ การยุบ

รวมกฎที่อยู่ Right equivalence class อาจจะทำให้เกิดกฎที่อยู่ใน Left equivalence classes เดียวกัน ดังนั้นจึงต้องทำการขยายกฎที่ฝั่งขวาอีกครั้ง

Algorithm: The ERMiner algorithm
Input: *SDB*: a sequence database, *min_supp* and *min_conf*: The two user-specified thresholds
Output: The set of valid sequential rules
1. *leftStore* ← ∅;
2. *rule* ← ∅ ;
3. Scan *SDB* once to calculate *EQ*, the set of all equivalence classes of rules of size 1*1
4. for each left equivalence class *H*∈*EQ* do
5. leftSearch(*H*, *rules*);
6. end
7. for each right equivalence class *J*∈*EQ* do
8. rightSearch(*J*, *rules*, *leftStore*);
9. end
10. for each left equivalence class *K*∈*leftStore* do
11. leftSearch(*K*, *rules*);
12. end
13. return *rules*;

ภาพที่ 3 *ขั้นตอนวิธี ERMiner [13]*

3. วิธีดำเนินการวิจัย



ภาพที่ 4 *ขั้นตอนการดำเนินการวิจัย*

ภาพที่ 4 แสดงขั้นตอนการดำเนินการวิจัยโดยรวมซึ่งประกอบไปด้วย การเก็บรวบรวมข้อมูล การเตรียมข้อมูล การสร้างกฎการมีอิทธิพล การเรียงกฎ ซึ่งแต่ละขั้นตอนมีรายละเอียดดังต่อไปนี้

3.1 การเก็บรวบรวมข้อมูล

งานวิจัยนี้ได้ทำการสร้างกฎเพื่อระบุผู้มีอิทธิพลและผู้ถูกครอบงำโดยใช้รหัสผู้ใช้งาน (User ids) จากการโพสต์หรือการแสดงความคิดเห็นในกลุ่มบนเฟสบุ๊คดังแสดงตัวอย่างข้อมูลในภาพที่ 5 โดยมีแนวคิดว่าถ้าคนที่มีอิทธิพลโพสต์หรือการแสดงความคิดเห็นหัวข้อใดๆ คนที่ถูกครอบงำมักจะเข้าไปแสดงความคิดเห็นหรือคอมเมนต์ในหัวข้อนั้นๆ ยกตัวอย่าง เช่น ถ้าผู้ใช้ u1 โพสต์หรือแสดงข้อความบนหัวข้อใดหัวข้อหนึ่งแล้ว u2, u3, และ u4 จะแสดงความคิดเห็นบนหัวข้อนั้นด้วย ซึ่งแสดงให้เห็นว่า u1 มีอิทธิพลกับ u2, u3, และ u4 นอกจากนี้การโพสต์หรือการแสดงความคิดเห็นบนเฟสบุ๊คเป็นข้อมูลที่ผู้ใช้แสดงออกโดยไม่ต้องมีการบังคับทำให้ข้อมูลที่ได้เป็นข้อมูลที่ได้จากพฤติกรรมการแสดงออกของผู้ใช้จริงๆ ในงานวิจัยนี้ทำการรวบรวมข้อมูลการแสดงความคิดเห็นจากกลุ่มเฟสบุ๊คจำนวนสามกลุ่ม โดยแต่ละกลุ่มมีจำนวนสมาชิกและความสัมพันธ์แตกต่างกันไปดังแสดงในตารางที่ 1 ข้อมูลที่ใช้ในงานวิจัยเป็นข้อมูลที่มีการโพสต์และแสดงความคิดเห็นในกลุ่มตั้งแต่วันที่ 6 มิถุนายน พ.ศ 2555 ถึง 23 เมษายน พ.ศ. 2558 การเก็บรวบรวมข้อมูลทำโดยการเขียนโปรแกรม PHP ร่วมกับ Facebook Graph API [14] เพื่อรวบรวมข้อมูลลงในฐานข้อมูล MySQL

id	post_id	user_id	post_date	action
496	289443467819333_787814331315575	849338825122598	2015-04-23 03:08:08	post
497	289443467819333_786420194788322	937318199651810	2015-04-20 13:44:17	post
498	289443467819333_786420194788322	937318199651810	2015-04-20 13:44:17	comment
499	289443467819333_786420194788322	10153252804829648	2015-04-20 13:44:17	comment
500	289443467819333_786420194788322	10204889349611782	2015-04-20 13:44:17	comment
501	289443467819333_786420194788322	937318199651810	2015-04-20 13:44:17	comment
502	289443467819333_786420194788322	4722138587555	2015-04-20 13:44:17	comment

ภาพที่ 5 *ตัวอย่างข้อมูลที่รวบรวมในฐานข้อมูล MySQL ของ Group1*

ตารางที่ 1 *ลักษณะข้อมูลแต่ละกลุ่ม*

กลุ่ม	#สมาชิก	คำอธิบาย
Group1	791	เป็นกลุ่มสำหรับแสดงความเห็นที่เกี่ยวข้องกับสถาบัน โดยมีสมาชิกอยู่ในสถาบันเดียวกัน
Group2	30,003	เป็นกลุ่มสำหรับแสดงความคิดเห็นเกี่ยวกับอาชีพ โดยมีสมาชิกทำอาชีพสายเดียวกัน
Group3	57,090	เป็นกลุ่มสำหรับแสดงความคิดเห็นเกี่ยวกับทางการเมืองในประเทศ มีสมาชิกหลากหลายทั่วประเทศ

3.2 การเตรียมข้อมูล

ในขั้นตอนนี้จะทำการแปลงข้อมูลที่รวบรวมได้ให้อยู่ในรูปแบบของรายการเปลี่ยนแปลง (Transaction) โดยหนึ่งกลุ่มคือ 1 ชุดข้อมูล แต่ละชุดข้อมูลประกอบไปด้วยรายการเปลี่ยนแปลง แต่ละรายการเปลี่ยนแปลงจะประกอบไปด้วยรหัสผู้ใช้ที่โพสต์และแสดงความคิดเห็นในหัวข้อเดียวกันตามลำดับ เช่น u2, u3, u1, u5 หมายความว่า u2 โพสต์ความคิดเห็นลงไปในเฟสบุ๊คแล้ว u3, u1 และ u5 ก็แสดงความคิดเห็นในหัวข้อดังกล่าวตามลำดับ ในหัวข้อใดที่ไม่มีการแสดงความคิดเห็นจะถูกตัดทิ้งไป จากตัวอย่างข้อมูลในตารางที่ 2 จะเห็นได้ว่า u3 โพสต์ข้อความบนเฟสบุ๊ค แต่ไม่มีใครแสดงความคิดเห็น ข้อมูลลักษณะแบบนี้จะถูกตัดออกไป ข้อมูลที่ผ่านการเตรียมข้อมูลแล้วจะแสดงได้ดังตารางที่ 3

ตารางที่ 2 *ตัวอย่างข้อมูลที่ดึงได้จากเฟสบุ๊ค*

User	Action
u1	post
u2	comment
u3	comment
u2	post
u1	comment
u3	comment
u2	post
u3	comment
u1	comment
u5	comment
u3	post
u1	post
u2	comment
u4	comment
u3	comment
u5	comment

ตารางที่ 3 *ข้อมูลที่ทำการแปลงเป็นรายการเปลี่ยนแปลง*

Transaction id	Sequence
1	u1, u2, u3
2	u2, u1, u3
3	u2, u3, u1, u5
4	u1, u2, u4, u3, u5

3.3 การสร้างกฎการมีอิทธิพล

งานวิจัยนี้สร้างกฎความสัมพันธ์แบบลำดับเพื่อแสดงความสัมพันธ์ของผู้มีอิทธิพลและผู้ที่ถูกครอบงำ โดยใช้ข้อมูลการโพสต์หรือการแสดงความคิดเห็นในกลุ่มบนเฟสบุ๊คซึ่งมีนิยามที่เกี่ยวข้องดังต่อไปนี้

กำหนดให้ $U = \{u_1, u_2,...,u_p\}$ เป็นเซตของรหัสผู้ใช้บนกลุ่มเฟสบุ๊ค และ ลำดับเหตุการณ์ $S = \{s_1, s_2,..., s_n\}$ คือ เซตรหัสผู้ใช้ที่มีการโพสต์หรือแสดงความคิดเห็นในหัวข้อเดียวกันตามลำดับโดยที่ $s_i \in U$ และ $1 \le i \le n$ ฐานข้อมูลประกอบไปด้วยลำดับเหตุการณ์หลายลำดับเหตุการณ์แทนด้วย $D = \{S_1, S_2,...,S_m\}$

นิยามที่ 1 กฎ $r: X \rightarrow Y$ คือ ความสัมพันธ์ระหว่างเซตรหัสผู้ใช้ X และเซตรหัสผู้ใช้ Y โดยที่ $X,Y \subseteq U$ ซึ่ง $X \cap Y = \varnothing$ และ $X,Y \neq \varnothing$ นอกจากนี้ X และ Y เป็นเซตรหัสผู้ใช้ที่ไม่คำนึงถึงลำดับการแสดงความคิดเห็น (Unordered set) แต่สมาชิกทุกตัวของ X จะต้องเกิดก่อนสมาชิกทุกตัวของ Y และเกิดในลำดับเหตุการณ์เดียวกัน เรียก X ว่า Antecedent และเรียก Y ว่า Consequent ในกฎ เช่น กฎ r: $u1, u2 \rightarrow u3$ หมายถึง เซตรหัสผู้ใช้ $u1, u2$ เป็นเซตที่เกิดในลำดับเหตุการณ์เดียวกันกับ $u3$ แต่ $u1,u2$ จะต้องเกิดก่อน $u3$

นิยามที่ 2 ค่าสนับสนุน (Support) ของกฎ $r: X \rightarrow Y$ คือ ความถี่ในการเกิด X แล้วตามด้วย Y ในฐานข้อมูล สามารถคำนวณได้ดังสมการที่ 1

$$supp(r) = \frac{|g(XY)|}{|g(D)|} \times 100 \qquad (1)$$

โดยที่ $|g(XY)|$ คือ จำนวนรายการเปลี่ยนแปลงที่เกิด X แล้วตามด้วย Y และ $|g(D)|$ คือ จำนวนรายการเปลี่ยนแปลงทั้งหมดในฐานข้อมูล

ตัวอย่างที่ 1 ค่าสนับสนุนของกฎ $r: u1, u2 \rightarrow u3$ คือ ความถี่หรือจำนวนรายการเปลี่ยนแปลงที่เกิด u1,u2 แล้วตามด้วย u3 ซึ่งมี 3 รายการเปลี่ยนแปลง คือ 1, 2 กับ 4 หารด้วยจำนวนรายการเปลี่ยนแปลงทั้งหมด ซึ่งมี 4 รายการ ดังนั้น $supp(r)$= (3/4)*100 = 75%

นิยามที่ 3 ค่าความเชื่อมั่น (Confidence) ของกฎ $r: X \rightarrow Y$ คือ ค่าความเชื่อมั่นในการเกิด X แล้วตามด้วย Y ซึ่งสามารถคำนวณได้ดังสมการที่ 2

$$conf(r) = \frac{|g(XY)|}{|g(D)|} \times 100 \qquad (2)$$

โดยที่ $|g(XY)|$ คือ จำนวนรายการเปลี่ยนแปลงที่เกิด X แล้วตามด้วย Y และ $|g(X)|$ คือ จำนวนรายการเปลี่ยนแปลงที่เกิด X

ตัวอย่างที่ 2 ค่าความเชื่อมั่นของกฎ r: $u1, u2 \rightarrow u3$ คือ ความถี่ในการเกิด u1,u2 แล้วตามด้วย u3 หารด้วยความถี่ของการเกิด $u1, u2$ ซึ่งเท่ากับ (3/4)*100 = 75%

นิยามที่ 4 กำหนดให้ค่าเชื่อมั่นขั้นต่ำคือ *min_supp* และค่าสนับสนุนขั้นต่ำ คือ *min_conf* กฎความสัมพันธ์แบบลำดับ $r: X \rightarrow Y$ เรียกว่า กฎแสดงอิทธิพล (Influential rule) ก็ต่อเมื่อ $supp(r) \geq min\_supp$ และ $conf(X \rightarrow Y) \geq min\_conf$

ค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำถูกนำมาใช้เป็นเกณฑ์ในการคัดเลือกกฎที่แสดงการมีอิทธิพล เนื่องมาจากค่าสนับสนุนขั้นต่ำจะเป็นตัวระบุความถี่ของการกระทำที่เพียงพอที่จะนำมาใช้ ส่วนค่าความเชื่อมั่นขั้นต่ำเป็นตัวระบุความเชื่อมั่นในการมีอิทธิพลครอบงำคนอื่น

ตารางที่ 4 *ตัวอย่างกฎที่ได้*

Influential rules	Supp (%)	Conf (%)	Influential rules	Supp (%)	Conf (%)
$u_2 \rightarrow u_1$	50	50	$u_1, u_2 \rightarrow u_5$	50	50
$u_1 \rightarrow u_2$	50	50	$u_1, u_3 \rightarrow u_5$	50	50
$u_1 \rightarrow u_3$	75	75	$u_1, u_2, u_3 \rightarrow u_5$	50	50
$u_2 \rightarrow u_3$	100	100	$u_2, u_3 \rightarrow u_5$	50	50
$u_1 \rightarrow u_5$	50	50	$u_1, u_2 \rightarrow u_3$	50	50
$u_2 \rightarrow u_5$	50	50	$u_2, u_1 \rightarrow u_3$	50	50
$u_3 \rightarrow u_5$	50	50	$u_2, u_3 \rightarrow u_5$	50	50
$u_1, u_2 \rightarrow u_3$	75	75			

ตารางที่ 4 เป็นกฎที่ได้จากการสืบค้นข้อมูลตัวอย่างในตารางที่ 3 เมื่อกำหนดให้ $min\_supp = 50\%$, $min\_conf = 50\%$ เช่น กฎ $u2 \rightarrow u3$ มีค่าสนับสนุนเท่ากับ (4/4)*100>*min_supp* ดังนั้นกฎ $u2 \rightarrow u3$ เป็นกฎที่มีค่าถี่เพียงพอ และค่าความเชื่อมั่น $conf(u2 \rightarrow u3)$ = (1/1)*100>*min_conf* ดังนั้น $u2 \rightarrow u3$ เป็นกฎที่มีค่าความเชื่อมั่น เพียงพอ ซึ่งแสดงให้เห็นว่า u2 มีอิทธิพลกับ u3 ด้วยค่าความเชื่อมั่นที่สูงถึง 100%

3.4 การเรียงกฎ

เมื่อได้กฎทั้งหมดแล้วกฎดังกล่าวจะต้องถูกเรียงก่อนจะนำไปใช้จริง ซึ่งในงานวิจัยนี้จะทำการเรียงกฎโดยประยุกต์ใช้แนวความคิดของขั้นตอนวิธี CBA [15] ซึ่งกฎจะถูกพิจารณานำไปใช้ก่อนเมื่อตรงกับเงื่อนไขดังต่อไปนี้

กำหนดให้กฎสองกฎ คือ r_i และ r_j กฎ r_i จะถูกเรียงเพื่อนำไปใช้ก่อน r_j เมื่อตรงกับเงื่อนไขดังต่อไปนี้ 1) $conf(r_i) > conf(r_j)$ หรือ 2) ถ้า $conf(r_i) = conf(r_j)$ แต่ $supp(r_i) > supp(r_j)$ หรือ 3) ถ้า $supp(r_i) = supp(r_j)$ แต่ $size(r_i) < size(r_j)$ ซึ่งหมายถึงความยาวของ Antecedent ของ r_i จะต้องสั้นกว่าความยาวของ Antecedent ของ r_j หรือ 4) ถ้า $size(r_i) = size(r_j)$ แต่ r_i มีลำดับการเกิดก่อน r_j จากตารางที่ 4 สามารถเรียงกฎที่ได้ดังตารางที่ 5

ตารางที่ 5 *ตัวอย่างกฎที่ถูกเรียงลำดับก่อนนำไปใช้*

Influential rules	Supp (%)	Conf (%)	Influential rules	Supp (%)	Conf (%)
$u_2 \rightarrow u_3$	100	100	$u_1 \rightarrow u_2, u_3$	50	50
$u_1 \rightarrow u_3$	75	75	$u_2 \rightarrow u_1, u_3$	50	50
$u_1, u_2 \rightarrow u_3$	75	75	$u_2 \rightarrow u_3, u_5$	50	50
$u_2 \rightarrow u_1$	50	50	$u_1, u_2 \rightarrow u_5$	50	50
$u_1 \rightarrow u_2$	50	50	$u_1, u_3 \rightarrow u_5$	50	50
$u_1 \rightarrow u_5$	50	50	$u_2, u_3 \rightarrow u_5$	50	50
$u_2 \rightarrow u_5$	50	50	$u_1, u_2, u_3 \rightarrow u_5$	50	50
$u_3 \rightarrow u_5$	50	50			

4. ผลการวิจัยและอภิปรายผล

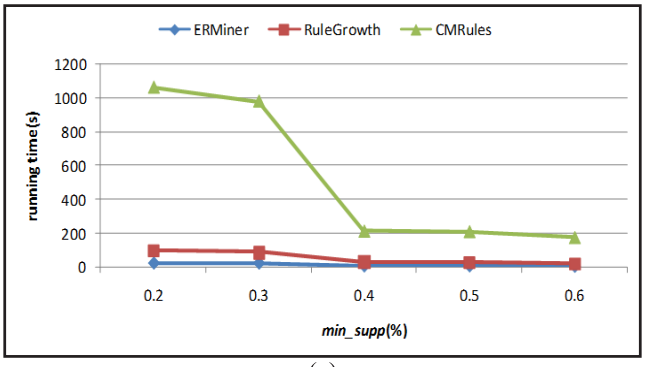
งานวิจัยนี้ได้ทำการทดลองบนคอมพิวเตอร์ 1.90GHz Core™ i3 PC หน่วยความจำ 4G บนระบบปฏิบัติการ Microsoft Windows 7 และใช้ขั้นตอนวิธี ERMiner, Rule-Growth และ CMRules ซึ่งเป็น Open-source data mining library ที่เขียนขึ้นโดยใช้ภาษาจาวา และสามารถดาวน์โหลดได้ที่ http://www.philippe-fournier-viger.com ทำการทดลองเพื่อทดสอบประสิทธิภาพของสามขั้นตอนวิธีกับข้อมูลโพสต์หรือแสดงความคิดเห็นบนกลุ่มเฟสบุ๊คที่ผ่านการเตรียมข้อมูลแล้วซึ่งมีรายละเอียดข้อมูลดังตารางที่ 6 ตารางที่ 7-8 และภาพที่ 6-7 แสดงถึงผลกระทบของค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำที่มีต่อจำนวนกฎ เวลาในการประมวลผลและการหน่วยความจำในการประมวลผลบนชุดข้อมูล Group1

ตารางที่ 6 *ลักษณะของข้อมูลที่ผ่านการเตรียมข้อมูล*

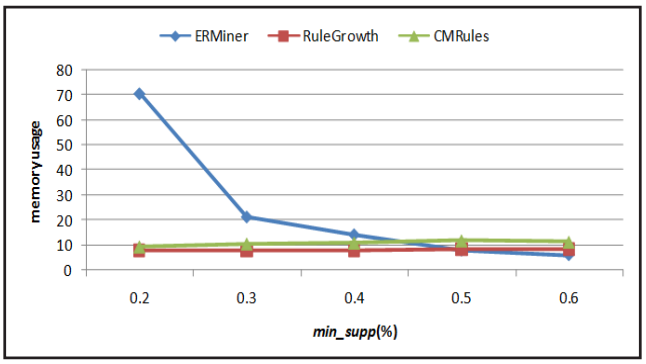
Dataset	#Distinct items	#Sequences	Max sequence
Group1	791	536	75
Group2	30,003	2329	99
Group3	57,090	4232	102

ตารางที่ 7 *จำนวนกฎที่ได้เมื่อค่าสนับสนุนขั้นต่ำเปลี่ยนแปลง*

Min_supp (%)	#Rules
0.2	916
0.3	916
0.4	33
0.5	33
0.6	2



(a)



(b)

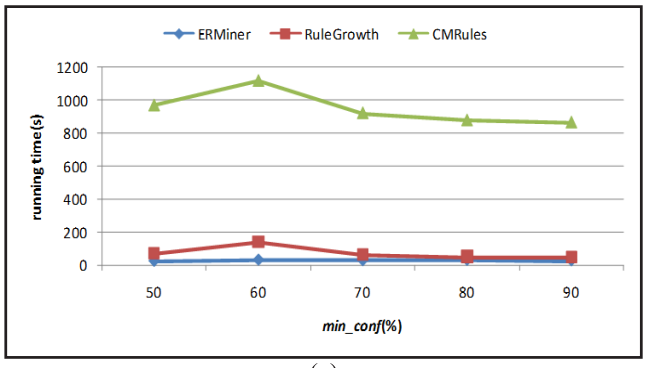
ภาพที่ 6 *เปรียบเทียบการใช้เวลา (a) และหน่วยความจำ (b) ในการประมวลผลเมื่อค่าสนับสนุนขั้นต่ำเปลี่ยนแปลง*

ตารางที่ 7 และภาพที่ 6 กำหนดค่าความเชื่อมั่นขั้นต่ำที่ 50% และกำหนดค่าสนับสนุนขั้นต่ำให้มีค่าเท่ากับ 0.2%, 0.3%, 0.4%, 0.5% และ 0.6% จากตารางที่ 7 จะเห็นว่าค่าสนับสนุนขั้นต่ำที่ 0.2% และ 0.3% จะให้จำนวนกฎที่เท่ากันและเป็นกฎเดียวกันโดยกฎที่ได้มีค่าสนับสนุนตั้งแต่ 0.3% ขึ้นไป แต่เมื่อกำหนดค่าสนับสนุนขั้นต่ำเพิ่มขึ้นเป็น 0.4% จำนวนกฎที่ได้ลดลงเป็นจำนวนมาก เนื่องจากกฎส่วนใหญ่ที่มีค่าสนับสนุนเท่ากับ 0.3% จะถูกตัดออก และเมื่อปรับค่าสนับสนุนขั้นต่ำเพิ่มขึ้นเป็น 0.5% ยังคงได้กฎจำนวนเท่าเดิม แสดงว่ากฎที่ได้จากการกำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.4% และ 0.5% เป็นกฎเดียวกันที่มีค่าสนับสนุนตั้งแต่ 0.5% ขึ้นไป และเมื่อกำหนดสนับสนุนขั้นต่ำที่ 0.6% กฎที่ได้มีจำนวนน้อยมากแค่ 2 กฎ ดังนั้นจะเห็นได้ว่าค่าสนับสนุนขั้นต่ำมีผลกระทบมากกับจำนวนกฎที่ได้ เมื่อค่าสนับสนุนขั้นต่ำเพิ่มขึ้นจำนวนกฎที่ได้จะลดลงมาก และจะเห็นว่าการสร้างกฎความสัมพันธ์แบบลำดับเพื่อหากฎแสดงอิทธิพลไม่สามารถกำหนดค่าสนับสนุนขั้นต่ำที่สูงได้ เนื่องมาจากความถี่ที่คนคนเดิมจะทำการแสดงความคิดเห็นมีค่าไม่สูง และจากภาพที่ 6 (a) แสดงให้เห็นว่าขั้นตอนวิธี CMRules ใช้เวลาในการประมวลผลมากกว่า ERMiner กับ RuleGrowth ทุกค่าสนับสนุนขั้นต่ำที่กำหนด ส่วนเวลาในการประมวลผลของ ERMiner กับ RuleGrowth เกือบเท่ากันและคงที่ แต่ ERMiner ใช้เวลาในการประมวลผลน้อยกว่า RuleGrowth เพียงเล็กน้อยสำหรับทุกค่าสนับสนุนขั้นต่ำ จากภาพที่ 6 (b) แสดงให้เห็นว่า ERMiner ใช้หน่วยความจำมากที่ค่าสนับสนุนขั้นต่ำเท่ากับ 0.2% และเมื่อค่าสนับสนุนขั้นต่ำเพิ่มขึ้นตั้งแต่ 0.3%-0.6% การใช้หน่วยความจำจะลงลงอย่างต่อเนื่องจนน้อยว่า RuleGrowth และ CMRules ตั้งแต่ค่าสนับสนุนขั้นต่ำที่ 0.5% เป็นต้นไป

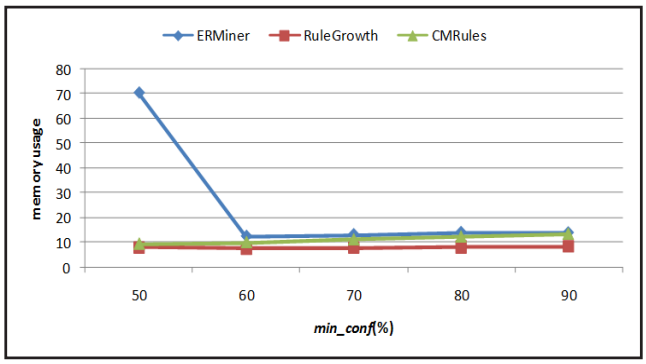
การกำหนดค่าสนับสนุนขั้นต่ำเพิ่มขึ้นใน ERMiner อาจจะทำให้ได้จำนวนกฎที่เท่ากัน แต่การใช้หน่วยความจำในการประมวลยังคงลดลง เช่น เมื่อค่าสนับสนุนขั้นต่ำเท่ากับ 0.2% และ 0.3% จำนวนกฎที่ได้เท่ากันดังตารางที่ 7 แต่ค่าสนับสนุนขั้นต่ำ 0.3% ใช้หน่วยความจำน้อยกว่ามาก เนื่องมาจากขั้นตอนวิธี ERMiner จะต้องสร้าง Left equivalence class และ Right equivalence class ที่ประกอบด้วยกฎที่มีสนับสนุนมากกว่าหรือเท่ากับค่าสนับสนุนขั้นต่ำ ถ้าค่าสนับสนุนเพิ่มขึ้นย่อมทำให้จำนวนกฎที่ได้ใน Left equivalence class และ Right equivalence class น้อยลงและทำให้การใช้หน่วยความจำลดลง จากนั้นกฎที่อยู่ใน Left equivalence class และ Right equivalence class จะถูกนำไปพิจาณาว่าผ่านค่าความเชื่อมั่นขั้นต่ำหรือไม่ ถ้าผ่านค่าความเชื่อมั่นขั้นต่ำ กฎดังกล่าวจะถูกเลือกเป็นผลลัพธ์ ดังนั้นจึงมีโอกาสเป็นไปได้ที่ค่าสนับสนุนขั้นต่ำต่างกันจะได้ผลลัพธ์จำนวนเท่ากัน แต่การใช้หน่วยความจำในการประมวลจะลดลงเมื่อกำหนดค่าสนับสนุนขั้นต่ำเพิ่มขึ้น ส่วนการใช้หน่วยความจำในการประมวลผลของ RuleGrowth และ CMRules มีแนวโน้มที่จะคงที่ถึงแม้ค่าสนับสนุนขั้นต่ำจะเพิ่มขึ้นก็ตาม

ตารางที่ 8 *จำนวนกฎที่ได้เมื่อค่าความเชื่อมั่นขั้นต่ำเปลี่ยนแปลง*

Min_conf (%)	#Rules
50	916
60	699
70	458
80	451
90	451



(a)



(b)

ภาพที่ 7 *เปรียบเทียบการใช้เวลา (a) และหน่วยความจำ (b) ในการประมวลผลเมื่อค่าความเชื่อมั่นขั้นต่ำเปลี่ยนแปลง*

ตารางที่ 8 และภาพที่ 7 กำหนดให้ค่าสนับสนุนขั้นต่ำเท่ากับ 0.2% และกำหนดค่าความเชื่อมั่นขั้นต่ำเท่ากับ 50%, 60%, 70%, 80% และ 90% จากตารางที่ 8 แสดงให้เห็นได้ว่าเมื่อค่าความเชื่อมั่นเปลี่ยนแปลงไป ไม่ได้ส่งผลกระทบกับจำนวนกฎที่ได้มากนัก และจากภาพที่ 7 (a) แสดงให้เห็นว่าเวลาในการประมวลผลของ ERMiner น้อยกว่า CMRules และ RuleGrowth และเกือบคงที่ ส่วนภาพที่ 7 (b) แสดงให้เห็นว่า ERMiner ใช้หน่วยความจำในการประมวลผลมากเมื่อค่าความเชื่อมั่นขั้นต่ำที่ 50% และลดลงอย่างรวดเร็วเมื่อค่าความเชื่อมั่นขั้นต่ำเท่ากับ 60% จากนั้นการใช้หน่วยความจำในการประมวลผลเกือบคงที่เมื่อค่าความเชื่อมั่นขั้นต่ำมีค่าเท่ากับ 60%-90% และทั้งสามขั้นตอนวิธีใช้หน่วยความจำในการประมวลผลต่างกันเล็กน้อยเมื่อค่าความเชื่อมั่นขั้นต่ำมีค่า 60%-90%

จากการทดลองในส่วนนี้จะเห็นได้ว่าขั้นตอนวิธี ERMiner เหมาะสำหรับประยุกต์ใช้ในการสร้างกฎเพื่อแสดงอิทธิพลโดยใช้ข้อมูลแสดงความคิดเห็นบนกลุ่มเฟสบุ๊ค เนื่องจาก ERMiner ใช้เวลาในการประมวลผลน้อยกว่า RuleGrowth และ CMRules และใช้เวลาเกือบคงที่ไม่ว่าจะกำหนดค่าสนับสนุนขั้นต่ำและค่าความเชื่อมั่นขั้นต่ำเป็นเท่าไรก็ตาม ส่วนการใช้หน่วยความจำในการประมวลผลของ ERMiner มีแนวโน้มที่จะลดลงจนน้อยกว่า RuleGrowth และ CMRules เมื่อค่าสนับสนุนขั้นต่ำเพิ่มขึ้น และใช้หน่วยความจำในการประมวลผลต่างกันเล็กน้อยกับขั้นตอนวิธี RuleGrowth และ CMRules เมื่อค่าความเชื่อมั่นขั้นต่ำสูงขึ้นตั้งแต่ 60% ถึง 90%

4.1 กฎที่ได้จากงานวิจัย

การทดลองส่วนนี้ทำการค้นหากฎด้วย ERMiner และพิจารณากฎ 10 อันดับแรกที่มีค่าความเชื่อมั่นสูงสุดในแต่ละกลุ่ม โดยพิจารณาความถี่ต่ำสุดของการแสดงความคิดเห็นร่วมกัน 3 ครั้ง เพื่อให้ได้กฎอย่างน้อย 10 กฎสำหรับทั้งสามกลุ่ม โดย Group 1 กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.4% ส่วน Group 2 กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.012% ส่วน Group 3 กำหนดค่าสนับสนุนขั้นต่ำเท่ากับ 0.07% และกำหนดค่าความเชื่อมั่นขั้นต่ำเท่ากับ 60% ทั้งสามกลุ่ม ผลลัพธ์ที่ได้แสดงในตารางที่ 9-11

ตารางที่ 9 *กฎ 10 อันดับแรกของ Group 1*

ID	Rules	Conf (%)	ID	Rules	Conf (%)
1	u214 → u98	100	6	u31, u221 → u32	50
2	u221 → u32	100	7	u61, u221 → u32	50
3	u9, u23 → u75	100	8	u61, u247 → u9	50
4	u14, u214 → u98	100	9	u206, u219 → u258	50
5	u23, u168 → u206	100	10	u5, u23, u168 → u206	50

ตารางที่ 10 *กฎ 10 อันดับแรกของ Group 2*

ID	Rules	Conf (%)	ID	Rules	Conf (%)
1	u3350 → u994	83	6	u64, u90 → u380	67
2	u2249 → u377	75	7	u2233 → u53	60
3	u4, u385 → u1	75	8	u3, u531 → u1	60
4	u53, u85 → u4	75	9	u3, u435 → u180	60
5	u380, u1024 → u31	75	10	u6, u134 → u285	60

ตารางที่ 11 *กฎ 10 อันดับแรกของ Group 3*

ID	Rules	Conf (%)	ID	Rules	Conf (%)
1	u1528 → u7	100	6	u344 → u39	100
2	u1528 → u398	100	7	u719, u1374 → u7	100
3	u1566 → u7	100	8	u178, u1511 → u7	100
4	u1721 → u470	100	9	u424, u1259 → u7	100
5	u1528 → u7, 398	100	10	u470, u2915 → u7	100

จากตารางที่ 9 จะเห็นว่ากฎ 10 อันแรกของ Group 1 มีค่าความเชื่อมั่นที่ 100% ซึ่งแสดงให้เห็นว่าทุกครั้งที่มีผู้มีอิทธิพลโพสต์ข้อความใดๆก็ตามผู้ที่ถูกครอบงำจะแสดงข้อคิดเห็นตาม เนื่องจากมีจำนวนสมาชิกในกลุ่มน้อย ดังนั้นโอกาสที่คนหนึ่งจะมีอิทธิพลครอบงำกับคนอีกคนหนึ่งจึงมีสูง แต่เมื่อขนาดของกลุ่มใหญ่ขึ้นการมีอิทธิพลและการครอบงำจะลดลง ดังจะเห็นในตารางที่ 10 ซึ่งกฎ 10 อันดับแรกของ Group 2 มีค่าความเชื่อมั่นอยู่ในช่วง 60-83% เนื่องมาจากความสัมพันธ์ของสมาชิกในกลุ่มไม่ใกล้ชิดเหมือนกลุ่ม Group 1 ส่วนตารางที่ 11 จะเห็นได้ว่ากฎ 10 อันดับแรกมีค่าความเชื่อมั่นสูงมากอยู่ที่ 100% ทุกกฎถึงแม้จะเป็นกลุ่มที่มีขนาดใหญ่ แต่เนื่องมาจาก Group 3 เป็นกลุ่มที่สร้างขึ้นเพื่อแสดงความคิดเห็นเกี่ยวกับการเมืองซึ่งมีสมาชิกที่มีแนวคิดเดียวกัน ดังนั้นคนที่มีอิทธิพลทางแนวความคิดทางการเมืองโพสต์ข้อความใดๆ ก็ตาม โอกาสที่คนที่ถูกครอบงำจะแสดงความคิดเห็นบนข้อความนั้นจึงมีโอกาสสูง จากตารางที่ 9-11 สามารถสรุปได้ว่า ความสัมพันธ์ของสมาชิกในกลุ่ม จะส่งเสริมให้คนหนึ่งมีอิทธิพลกับอีกคนหนึ่งสูง นอกจากนี้แล้วความคิดเห็นที่ตรงกันก็เป็นเหตุอย่างหนึ่งที่ส่งเสริมให้คนหนึ่งมีอิทธิพลกับอีกคนหนึ่งสูงเหมือนกัน

5. สรุปผลการวิจัย

งานวิจัยนี้นำเสนอแนวความคิดในการค้นหาผู้มีอิทธิพลและผู้ที่ถูกครอบงำบนเครือข่ายสังคมออนไลน์เฟสบุ๊คโดยประยุกต์ใช้การสืบค้นกฎความสัมพันธ์แบบลำดับกับข้อมูลการแสดงความคิดเห็นที่อยู่บนกลุ่มเฟสบุ๊ค โดยใช้แนวความคิดที่ว่าถ้าผู้มีอิทธิพลโพสต์หรือแสดงข้อคิดเห็นในหัวข้อใดๆ ก็ตามผู้ที่ถูกครอบงำมักจะแสดงความความคิดเห็นในหัวข้อนั้นด้วย ซึ่งกฎที่ได้จะทำให้ทราบกว่ากลุ่มคนไหนมีอิทธิพลกับกลุ่มคนไหนมากน้อยแค่ไหน งานวิจัยนี้ทดลองประยุกต์ใช้ 3 ขั้นตอนวิธีสำหรับค้นหากฎเพื่อแสดงอิทธิพล ซึ่งผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธี ERMiner เหมาะสำหรับประยุกต์ใช้ในการสร้างกฎเพื่อแสดงอิทธิพล เนื่องจาก ERMiner ใช้เวลาในการประมวลผลน้อยกว่า RuleGrowth และ CMRules และใช้เวลาเกือบคงที่ นอกจากนี้หน่วยความจำที่ใช้ใน ERMiner ก็ยังมีแนวโน้มที่ดีขึ้นเมื่อกำหนดค่าสนับสนุนขั้นต่ำสูงขึ้น จากนั้นได้ทำการทดลองค้นหากฎจากกลุ่มบนเฟสบุ๊คจำนวน 3 กลุ่มที่มีคุณลักษณะแตกต่างกันออกไป และพบว่าความสัมพันธ์ใกล้ชิดของสมาชิกในกลุ่มและสมาชิกที่มีแนวคิดเดียวกัน ทำให้การมีอิทธิพลซึ่งกันและกันสูงตามไปด้วย

6. เอกสารอ้างอิง

[1] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri. “Feedback effects between similarity and social influence in online communities.” *in Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Las Vegas, Nevada, pp. 160-168, USA, 2008.

[2] P. Domingos and M. Richardson. “Mining the network value of customers.” *in Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 57-66, San Francisco, California, 2001.

[3] D. Kempe, J. Kleinberg, and E. Tardos. “Maximizing the spread of influence through a social network.” *in Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, D.C., pp. 137-146, 2003.

[4] J. Leskovec, L. A. Adamic, and B. A. Huberman. “The dynamics of viral marketing.” *ACM Transactions Web*, Vol. 1, pp. 5, 2007.

[5] M. Cha, A. Mislove, and K. P. Gummadi. “A measurement-driven analysis of information propagation in the flickr social network.” *in Proceedings of the 18th International Conference on World Wide Web*, pp. 721-730, Madrid, Spain, 2009.

[6] M. Trusov, A. V. Bodapati, and R. E. Bucklin. “Determining Influential Users in Internet Social

Networks." *Journal of Marketing Research*, Vol. 47, pp. 643-658, August, 2010.

[7] P. W. Ballantine, Y. Lin, and E. Veer. "The Influence of User Comments on Perception of Facebook Relationship Status Updates." *Computers in Human Behaviour*, Vol. 49, pp. 50-55, 2015.

[8] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. "Discovering leaders from community actions." *in Proceedings of the 17th ACM Conference on Information and Knowledge Management*, Napa Valley, pp. 499-508, California, USA, 2008.

[9] A. Goyal, O. Byung-Won, F. Bonchi, and L. V. S. Lakshmanan. "GuruMine: A Pattern Mining System for Discovering Leaders and Tribes." *in Proceedings of the IEEE 25th International Conference in Data Engineering*, pp. 1471-1474, 2009.

[10] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases." *in Proceedings of the 1993 ACM SIGMOD Internatinal Conference on Mangament of Data*, pp. 207-216, USA, 1993.

[11] P. Fournier-Viger, R. Nkambou, and V. S.-M. Tseng. "RuleGrowth: mining sequential rules common to several sequences by pattern-growth." *presented at the 2011 ACM Symposium on Applied Computing*, pp. 956-961, TaiChung, Taiwan, 2011.

[12] P. Fournier-Viger, U. Faghihi, R. Nkambou, and E. M. Nguifo. "CMRules: Mining sequential rules common to several sequences." *Knowledge-Based Systems*, pp. 63-76, Vol. 25, 2012.

[13] P. Fournier-Viger, T. Gueniche, S. Zida, and V. Tseng. "ERMiner: Sequential Rule Mining Using Equivalence Classes." *Advances in Intelligent Data Analysis XIII*. Vol. 8819, H. Blockeel, M. van Leeuwen, and V. Vinciotti, Eds., ed: Springer International Publishing, pp. 108-119, 2014.

[14] Facebook. *The Graph API*, Available online at https://developers.facebook.com/docs/graph-api, accessed on 3 January 2015

[15] B. Liu. "Integrating Classification and Association Rule Mining." *in Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, pp. 80-86, New York, 1998.