ระบบสืบค้นพฤติกรรมการให้บริการของโรงพยาบาล ด้วยเทคนิคอพริโอริและเอนโทรปีพรุนอัลกอริทึม

นัฐพรรณ ฤทธิ์ขจร, อนุชา จันทเสรีนนท์, ปราณปรียา ศรีสุระ และ ทวีชัย อวยพรกชกร

ห้องปฏิบัติการระบบอัจฉริยะเพื่อภาคธุรกิจ คณะวิศวกรรมศาสตร์ศรีราชา มหาวิทยาลัยเกษตรศาสตร์วิทยาเขตศรีราชา
199 ถนนสุขุมวิท ตำบลทุ่งสุขลา อำเภอศรีราชา จังหวัดชลบุรี 20230 โทรศัพท์: 0-3835-4580-6
E-mail: pookka009@hotmail.com, sang-usa@hotmail.com, pranpreeya_flighting@hotmail.com, sfengtwo@src.ku.ac.th

บทคัดย่อ

โรงพยาบาลเป็นหน่วยงานด้านสุขภาพหนึ่ง ที่มีความสำคัญแก่ชุมชนเป็นอย่างสูง การให้บริการของโรงพยาบาลมีความสำคัญอย่างยิ่งยวดที่จะก่อให้เกิดความมั่นใจต่อผู้ป่วยที่มาใช้บริการ รวมถึงลักษณะของการใช้บริการโรงพยาบาลจะสะท้อนถึงข้อมูลสุขภาพของชุมชนอีกด้วย งานวิจัยนี้ได้นำเสนอกระบวนการสืบค้นความรู้ จากฐานข้อมูลการให้บริการของโรงพยาบาลด้วยการวิเคราะห์ค่าเอนโทรปี (Entropy) ร่วมกับการคัดเลือกกฎความสัมพันธ์ที่ได้จากเทคนิคอพริโอริ (Apriori Algorithm) ซึ่งจากผลการทดลองบนข้อมูลทดสอบจริงพบว่า กระบวนการที่นำเสนอให้ผลลัพธ์เป็นที่น่าพอใจมีค่าเอฟเมเชอร์ (F-Measure) เฉลี่ย 17.78% สามารถปรับปรุงคุณภาพเมื่อเทียบกับกรณีไม่ใช้ได้เพิ่มขึ้นถึง 10.71% บนข้อมูลทดสอบ โดยที่ความรู้ที่สืบค้นได้บนฐานข้อมูลจริงของโรงพยาบาล เป็นที่พอใจแก่ผู้ใช้งานอีกด้วย

คำสำคัญ-- *เหมืองความรู้; โรงพยาบาล; การให้บริการ; อพริโอริ*

1.บทนำ

การสืบค้นความรู้ในฐานข้อมูลขนาดใหญ่ (Knowledge Discovery on Large Database: KDD) เป็นเทคนิคที่มุ่งเน้นในการค้นหาความรู้ความสัมพันธ์ที่ซ่อนอยู่ภายในข้อมูลที่อาจจะยังไม่ถูกนำไปใช้ กฎความสัมพันธ์ (Association Rule) เป็นกระบวนการหนึ่งที่มีเป้าหมายในการหาความเกี่ยวเนื่องและความสัมพันธ์ระหว่างข้อมูลหนึ่งกับอีกข้อมูลหนึ่ง การนำกฎความสัมพันธ์ไปใช้งานจริงในปัจจุบัน ได้ถูกนำไปประยุกต์ใช้งานอย่างกว้างขวางเพื่อค้นหาพฤติกรรมของลูกค้าที่สามารถนำไปใช้ปรับปรุงการให้บริการ หรือวางแผนทางการตลาดต่อไปได้

โรงพยาบาลเป็นองค์กรด้านสุขภาพหนึ่ง ที่มีข้อมูลการให้บริการเป็นจำนวนมาก ข้อมูลในระบบงานโรงพยาบาลสะท้อนถึงคุณภาพการให้บริการ ขีดความสามารถในการให้บริการ รวมถึงข้อมูลสุขภาพและข้อมูลทางระบาดวิทยาอีกด้วย การนำเทคนิคการสืบค้นความรู้โดยเฉพาะกฎความสัมพันธ์มาใช้กับฐานข้อมูลการให้บริการของโรงพยาบาล จะสะท้อนไปถึงความพร้อมในการให้บริการด้านสุขภาพของโรงพยาบาล รวมถึงการวิเคราะห์องค์ประกอบด้านข้อมูลสุขภาพที่มีอยู่เป็นจำนวนมาก เช่น โรคที่เป็น แผนกที่ใช้บริการ ค่าบริการ ที่เกิด แพทย์ที่ให้การรักษา ความเกี่ยวเนื่องกับข้อมูลส่วนตัวของคนไข้ เช่น ที่อยู่ของคนไข้ อาชีพ วันที่ที่เข้ารับการรักษา ฯลฯ จะเป็นข้อมูลสำคัญทั้งสำหรับผู้บริหารของโรงพยาบาล และบุคลากรด้านการแพทย์ของโรงพยาบาลที่สามารถนำไปประยุกต์ใช้งานในการปรับปรุงหรือเตรียมการให้บริการได้

อย่างไรก็ดี แม้ว่ากฎความสัมพันธ์จะเป็นเครื่องมือที่มีประสิทธิภาพในการสืบค้นความรู้ แต่ด้วยทฤษฎีการหากฎความสัมพันธ์ที่ เป็นนิยมอยู่ในปัจจุบันคือการสร้างด้วยอัลกอริทึมอพริโอริ (Apriori Algorithm) [1][2] [11] นั้นจะให้จำนวนผลลัพธ์ของกฎความสัมพันธ์เป็นจำนวนมาก การนำไปใช้งานต่อจึงทำได้อย่างยากลำบาก งานวิจัยนี้ได้นำเสนอเทคนิคในการคัดกรองความรู้ที่ได้จาก Apriori Algorithm ด้วยอัลกอริทึมเอนโทรปีพรุน (Entropy Prune Algorithm) ที่จะช่วยคัดเลือกความรู้ที่น่าสนใจ โดยพิจารณาจากความเกี่ยวเนื่องและความอิสระในเนื้อข้อมูล ระหว่างความรู้ที่สนใจและรูปแบบการกระจายของข้อมูลที่ต้องการทดสอบ ซึ่งจากการทดลองกับข้อมูลทดสอบ อ้างอิงโครงสร้างฐานข้อมูลโรงพยาบาลจริง พบว่า สามารถคัดเลือกความรู้ที่มีความน่าสนใจได้ดีมากขึ้นได้

ลำดับการนำเสนอต่อไป ประกอบด้วยทฤษฎีที่เกี่ยวข้องในหัวข้อที่ 2 กระบวนการในการพัฒนาและผลการทดลองในหัวข้อที่ 3, 4 และสรุปผลในลำดับสุดท้าย

2. ทฤษฎีที่เกี่ยวข้อง

การทำเหมืองข้อมูล (Data Mining) เป็นกระบวนการสืบค้นความรู้โดยจะค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูล ที่อาจมีอยู่เป็นจำนวนมาก [11] ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลากหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์ เศรษฐกิจและสังคม รวมถึงด้านการแพทย์และสาธารณสุข [1][11] เทคนิคทางด้าน Data Mining หลายๆ เทคนิคได้ถูกพัฒนาขึ้นเพื่อการสืบค้นความรู้ในรูปแบบต่างๆ เช่น การหากฎความสัมพันธ์ (Association Rule) [1][3] การจัดกลุ่มและแบ่งกลุ่มข้อมูล (Classification/ Clustering) [1][2] เป็นต้น

2.1. Association Rule[5]

เป็นการค้นหากฎความสัมพันธ์ (Association Rule) ของข้อมูล โดยจะค้นหาความสัมพันธ์ที่เกิดขึ้นได้ทั้งหมดระหว่างข้อมูลหรือกลุ่มข้อมูล การวัดความสำคัญของกฎความสัมพันธ์จะใช้ค่าสนับสนุน (Support) และค่าความมั่นใจ (Confidence) เป็นเกณฑ์ในการคัดเลือก รูปแบบทั่วไปของการค้นหากฎความสัมพันธ์ จะอยู่ในรูปแบบ $A \Rightarrow B$ โดยที่ A, B เป็นเซตของไอเทม (Itemset) ที่ประกอบอยู่ในฐานข้อมูล ค่าสนับสนุน (Support) และค่าความมั่นใจ (Confidence) จะคำนวณได้จากสมการที่ 1 และ 2

$$\text{support}(A \Rightarrow B) = P(A \cup B) = \frac{|A \cup B|}{|All|} \quad (1)$$

$$\text{confidence}(A \Rightarrow B) = P(B|A) = \frac{|A \cup B|}{|A|} \quad (2)$$

ค่าสนับสนุน (Support) จะวัดความน่าจะเป็นของจำนวนรายการของข้อมูลที่เกิดร่วมกันเทียบกับจำนวนรายการทั้งหมด ส่วนค่าความมั่นใจ (Confidence) จะวัดความน่าจะเป็นเมื่อเกิดเหตุการณ์หนึ่ง (A) แล้วจะเกิดอีกเหตุการณ์หนึ่งตามมา (B) ในการหากฎความสัมพันธ์นั้นจะมีขั้นตอนวิธีการหาหลายวิธีด้วยกัน ขั้นตอนวิธีที่เป็นที่รู้จักและถูกใช้อย่างแพร่หลาย คือ อัลกอริทึมอพริโอริ (Apriori Algorithm)

2.2. Apriori Algorithm

อพริโอริเป็นอัลกอริทึมในการสืบค้นกฎความสัมพันธ์ที่เกิดขึ้นได้ทั้งหมดในฐานข้อมูล โดยเทียบกับค่าสนับสนุนขั้นต่ำ (Minimum Support) ที่กำหนดโดยผู้ใช้งาน ผลลัพธ์ในการสืบค้นจะอยู่ในรูปแบบ Lattice ที่แสดงถึงความสัมพันธ์ของไอเทมเซต (Item set) หรือเซตของหน่วยข้อมูลที่ปรากฏในฐานข้อมูล อัลกอริทึมอพริโอริแสดงได้ดังรูปที่ 1 (ฟังก์ชัน “apriori”) เริ่มด้วยบรรทัดที่ 1 ซึ่งเป็นการหาไอเทมเซตขนาด 1 ไอเทมที่มีจำนวนรายการในฐานข้อมูล (Transaction) ที่มีไอเทมนั้นๆ ประกอบอยู่ (หรือเรียกว่าค่าสนับสนุนของไอเทมเซต (Support)) ผ่านค่าสนับสนุนขั้นต่ำ ไอเทมเซตที่ผ่านค่าสนับสนุนขั้นต่ำจะเรียกว่าฟรีเควันไอเทมเซต (Frequent Itemset) หรือ L จากนั้นฟรีเควันไอเทมเซตขนาด 1 ไอเทมจะถูกนำไปสร้างไอเทมเซตขนาดที่สูงกว่า 1 ขั้นที่เป็นไปได้ทั้งหมดหรือแคนดิเดตไอเทมเซต (Candidate Itemset) C_k ในบรรทัดที่ 2-3 (และแสดงในฟังก์ชัน “apriori_gen” บรรทัดที่ 1-4) ซึ่งไอเทมเซตนั้นๆ จะต้องมีซับเซตที่เป็นฟรีเควันไอเทมเซตทั้งหมดด้วย แสดงดังฟังก์ชัน “apriori_gen” บรรทัดที่ 5-9 และฟังก์ชัน “has_infrequent_subset”

แต่ละแคนดิเดตไอเทมเซตจะถูกนำไปตรวจสอบกับรายการในฐานข้อมูล เพื่อนับจำนวนรายการที่พบว่าแคนดิเดตไอเทมเซตนั้นๆ เป็นซับเซต (จะได้ค่าสนับสนุนของไอเทมเซตนั้นๆ) ดังแสดงในบรรทัดที่ 4-9 ค่าสนับสนุนที่ได้จะถูกนำไปเทียบกับค่าสนับสนุนขั้นต่ำอีกครั้ง เพื่อคัดเลือกแคนดิเดตไอเทมเซตที่เป็นฟรีเควันไอเทมเซต ดังแสดงในบรรทัดที่ 10-11 จากนั้นฟรีเควันไอเทมเซตจะถูกนำไปใช้ในการสร้างแคนดิเดตไอเทมเซตในลำดับชั้นสูงขึ้นต่อไปและทำซ้ำไปเรื่อยๆ จนถึงลำดับชั้นที่ k ซึ่งเป็นขนาดของไอเทมเซตสูงสุดที่พบในฐานข้อมูล

```
apriori(D,min_sup)
(1)   L1=find_frequent_1-itemsets(D);
(2)   for(k=2;Lk-1 ≠∅;k++ ) {
(3)       Ck =apriori_gen(Lk-1);
(4)       for each transaction t∈D
(5)       {
(6)           Ct=subset(Ck,t );
(7)           for each candidate c ∈ Ct
(8)               c.count++;
(9)       }
(10)      Lk={c∈ Ck | c.count ≥ min_sup};
(11)  }
(12)  return L= ∪k Lk;
```

```
apriori_gen(Lk-1)
(1)   for each itemset l1∈ Lk-1
(2)       for each itemset l2∈ Lk-1
(3)           if((l1[1]= l2[1])∧(l1[2]=l2[2])
                ∧…∧(l1[k-2]=l2[k-2])
                ∧(l1[k-1]) = l2[k-1])) {
(4)               c=l1⊳⊲ l2;
(5)               if(has_infrequent_subset(c,Lk-1))
(6)                   delete c;
(7)               else add c to Ck ;
(8)           }
(9)   return Ck;
```

```
has_infrequent_subset(c,Lk-1)
(1)   for each (k-1)-subset s of c
(2)       if (s ∉ Lk-1)
(3)           return TRUE;
(4)   return FALSE;
```

```
Notation:
D=Database
Lk=Frequent itemset with k-items
Ck=Candidate itemset with k-items
```

รูปที่ 1. Apriori Algorithm

2.3. Entropy

ค่าเอนโทรปี (Entropy) เป็นค่าความเกี่ยวข้องระหว่างข้อมูลที่สนใจหรือคลาส (Class) กับข้อมูลปัจจัยที่ต้องการเทียบหรือแอททิบิว (Attribute) การวัดค่า Entropy จะวัดลักษณะของการกระจายของข้อมูลบนแต่ละค่าของ Class ที่เกิดขึ้นว่ามีลักษณะการกระจายที่สามารถสรุปได้ว่า Attribute นั้นมี

ความเกี่ยวข้องกับ Class ที่สนใจหรือไม่และมากน้อยเพียงไร การหาค่า Entropy จะหาจากค่า Information Gain (สมการที่ 3) ซึ่งเป็นค่า การกระจายรวม เทียบกับค่า Entropy ของแต่ละ Attribute (สมการที่ 4) การพิจารณาความเกี่ยวข้องของ Attribute ที่มีผลต่อ Class จะพิจารณาจากค่า Entropy Gain (สมการที่ 5)

$$I(s_1,s_2,\ldots,s_m)=-\sum_{i=1}^{m}\frac{s_i}{s}\log_2\frac{s_i}{s} \quad (3)$$

$$E(A)=\sum_{j=1}^{v}\frac{s_{1j}+\ldots+s_{mj}}{s}\,I\left(s_{1j},\ldots,s_{mj}\right) \quad (4)$$

$$\text{Entropy Gain}(A)=I(s_1,s_2,\ldots,s_m)-E(A) \quad (5)$$

s = Set of training sample
m = Number of class in training sample
s_i = Total training sample of class C_i
I = Expected Information of class C
A = Attribute with value $\{a_1,a_2,\ldots,a_v\}$ that can be used to partition S into subset $\{S_1,S_2,\ldots,S_v\}$where S_j have value a_j
S_j = Total training sample of S_j in Class C_i
E(A) = Entropy of A
Entropy Gain(A) = Information Gain

2.4. F-Measure

เอฟเมเชอร์ (F-Measure) เป็นเครื่องมือที่ใช้ในการวัดผลคุณภาพการค้นหาข้อมูล (Information Retrieval) โดยพิจารณาจากปัจจัยหลายส่วนประกอบกัน ทั้งค่าความแม่นยำ (Precision) ดังแสดงในสมการที่ 6 ที่พิจารณาจากความถูกต้องของข้อมูลที่สืบค้นได้ และค่าความครบถ้วน (Recall) ในสมการที่ 7 ที่พิจารณาจากความครบถ้วนเมื่อเทียบกับข้อมูลที่ควรได้ทั้งหมด การสืบค้นที่ดีจะต้องพิจารณาทั้ง Precision และ Recall ซึ่งผลการสืบค้นที่ดีที่สุดอาจจะไม่ได้ค่า Precision สูงสุดหรือให้ค่า Recall สูงสุดก็เป็นได้ ค่า F-Measure ได้ถูกออกแบบขึ้นเพื่อพิจารณาทั้ง Precision และ Recall ประกอบกัน ดังแสดงในสมการที่ 8 การเปรียบเทียบตัวสืบค้นเฉลี่ยที่ดีจึงควรพิจารณาจากค่า F-Measure เป็นหลัก

$$\text{Precision}(P)=\frac{w}{n_2}=\frac{|\{\text{relevant docs}\}\cap\{\text{retrieved docs}\}|}{|\{\text{retrieved docs}\}|} \quad (6)$$

$$\text{Recall}(R)=\frac{w}{n_1}=\frac{|\{\text{relevant docs}\}\cap\{\text{retrieved docs}\}|}{|\{\text{relevant docs}\}|} \quad (7)$$

$$\text{F-Measure}=\frac{2\times\text{Precision}\times\text{Recall}}{(\text{Precision}+\text{Recall})} \quad (8)$$

```
apriori_new(D,min_sup,minEG)
(1)  L1=find_frequent_1- itemset(D);
(2)  for (k=2;Lk-1≠Ø;k++){
(3)     Ck=apriori_gen(Lk-1);
(4)       for each transaction t∈D {
(5)          Ct=subset(Ck,t);
(6)          for each candidate  c∈Ct
(7)             c.count++;
(8)       }
(9)       Lk={c∈Ck|c.count≥min_sup};
(10)      for each itemset lk∈Lk
(11)         entropy_prune(lk,minEG);
(12) }
(13)    return  L=∪kLk;
```

```
entropy_prune(lk,minEG);
(1)  for each ck =subset(lk)
(2)       for each cj=subset(lk)| cj∩ ck=Ø
(3)          if(EntropyGain(ck,cj)≤minEG)
(4)             delete lk;
```

รูปที่ 2. Apriori Algorithm with Entropy Prune

3. กระบวนการในการพัฒนา

จากทฤษฎีการสืบค้นความรู้ข้างต้น แม้ว่าแนวทางการสืบค้นด้วยอัลกอริทึมอพริโอริจะสามารถสืบค้นความสัมพันธ์ที่เกิดขึ้นทั้งหมดได้ แต่อัลกอริทึมอพริโอริ ก็ยังมีข้อจำกัดที่ไม่สามารถคัดกรองข้อมูลที่อาจไม่เกี่ยวข้องกัน แต่ด้วยลักษณะของข้อมูลทำให้ข้อมูลปรากฏพร้อมกัน จึงทำให้ได้กฎความสัมพันธ์ที่ไม่เกี่ยวข้องเกิดขึ้นเป็นจำนวนมาก แนวทางในการวิจัยในครั้งนี้จึงนำเสนอการนำค่า Entropy Gain มาช่วยในการคัดกรองความสัมพันธ์ ซึ่งจะช่วยให้ความรู้ที่สืบค้นได้มีค่าประสิทธิภาพการสืบค้นดีขึ้นบนปริมาณของกฎที่ไม่มากจนเกินไป ซึ่งในงานวิจัยนี้ได้นำเสนอค่าคัดกรอง “ค่าความเกี่ยวข้องต่ำสุด” (Minimum Entropy Gain) ดังแสดงในคำจำกัดความ 3.1 เพื่อคัดเลือกกฎความสัมพันธ์ที่น่าสนใจ โดยปรับปรุงอัลกอริทึมอพริโอริ แสดงดังรูปที่ 2

คำจำกัดความ 3.1 ค่าความเกี่ยวข้องต่ำสุด (Minimum Entropy Gain: MinEG) เป็นค่าความเกี่ยวข้องระหว่างคุณสมบัติ (Attribute) ของข้อมูล หรือ Entropy Gain ที่กำหนดโดยผู้ใช้งาน โดยคำนวณจากค่า Entropy ของ Attribute หนึ่งเทียบอีกกับ Attribute หนึ่ง ในกรณีที่เป็นความสัมพันธ์ของไอเทมเซตที่ประกอบด้วยไอเทมมากกว่า 1 ไอเทม หากค่า Entropy Gain ของคู่ Attribute ของแต่ละไอเทมมีค่ามากกว่า Minimum Entropy Gain จะถือว่าไอเทมเซตนั้นๆ ผ่านการคัดเลือกเป็น ฟรีเควันไอเทมเซต (Frequent Itemset)

จากอัลกอริทึมอพริโอริที่ได้ปรับปรุงใหม่ (ฟังก์ชัน “apriori_new”) ในรูปที่ 3 ได้เพิ่มเติมค่าคัดกรองค่าความเกี่ยวข้องต่ำสุด (“MinEG”) ในบรรทัดที่ 1, 11 และเพิ่มเติมฟังก์ชันในการคัดกรองฟรีเคว้นไอเทมเซตเพิ่มเติมจากการคัดเลือกด้วยค่าสนับสนุนขั้นต่ำ ดังแสดงในบรรทัด 10-11 และในฟังก์ชัน “entropy_prune” ในการคัดกรองนั้นจะตรวจสอบว่าแต่ละคู่ไอเทมที่เป็นซับเซตของฟรีเคว้นไอเทมเซตที่ได้จากการคัดกรองด้วยค่าสนับสนุนขั้นต่ำ มีค่า Entropy Gain ของคู่ Attribute ของแต่ละไอเทมผ่านค่าความเกี่ยวข้องต่ำสุดหรือไม่ หากผ่านจะถูกคัดเลือก แต่หากไม่ผ่านจะถูกคัดออกและไม่ถูกนำไปใช้สร้างแคนดิเดตไอเทมเซตในลำดับขั้นต่อไป

จากรูปแบบการสืบค้นที่นำเสนอข้างต้น กฎความสัมพันธ์ที่ได้จึง ผ่านการคัดกรองด้วยค่าสนับสนุนขั้นต่ำสุด (Minimum Support) และผ่านค่าความเกี่ยวข้องขั้นต่ำ (MinEG) ไปพร้อมกัน ทำให้กฎความสัมพันธ์ที่อาจมีค่าสนับสนุนมากพอ แต่ไม่มีความเกี่ยวข้องกับข้อมูลไม่ผ่า นการคัดเลือก ประสิทธิภาพในการสืบค้นโดยรวมจึงดีขึ้น

4. ผลการทดลองกับข้อมูลทดสอบ

จากข้อมูลทดสอบ 40 รายการ ซึ่งเป็นข้อมูลการใช้บริการของโรงพยาบาลจริงของโรงพยาบาลบ้านแพ้วฯ โดย เลือกเฉพาะ Attribute ดังต่อไปนี้ 1) ความสามารถในการให้บริการของแผนก 2) วันที่ให้บริการ 3) แพทย์ 4) อาชีพของคนไข้ 5) อายุของคนไข้ กำหนดชื่อแทนดังแสดงรายละเอียดในตารางที่ 1 นำข้อมูลตัวอย่างไปทดสอบด้วยอัลกอริทึมอพริโอริ ก่อนนำไปทดสอบเทียบกับอพริโอริที่ได้ปรับปรุงใหม่ที่ได้นำเสนอในงานวิจัยนี้ โดยใช้ค่า F-Measure เป็นค่าวัด ทั้งนี้เกณฑ์การคัดเลือกข้อมูลตัวอย่างนี้ จะคัดเลือกจากกฎความสัมพันธ์ที่อาจได้และจำนวนของกฎที่อาจเกิด ที่สามารถให้เจ้าหน้าที่ผู้ทำรายงานสรุปของโรงพยาบาลบ้านแพ้ว (องค์การมหาชน) สามารถตรวจสอบผลลัพธ์ได้ จากการ ทดสอบบนค่าสนับสนุนขั้นต่ำ (Minimum Support) เท่ากับ 3, 5 และ 7 รายการ และค่าความเกี่ยวข้องขั้นต่ำ (Minimum Entropy Gain) เท่ากับ 18% ตัวอย่างผลที่ได้แสดงดังรูปที่ 3 และรูปที่ 4 ซึ่งจากรูปที่ 4 สามารถนำไปสร้างเป็นกฎความสัมพันธ์ให้กับผู้ใช้งานแสดงดังตารางที่ 2 โดยจะเห็นว่าด้วยเทคนิคที่นำเสนอ จะลดจำนวนกฎความสัมพันธ์ลงจาก 22 กฎในกระบวนการสร้างด้วยอัลกอริทึมอพริโอริเพียงอย่างเดียว เหลือเพียง 16 กฎตามเทคนิคที่นำเสนอในงานวิจัย ซึ่งลดลง 27.27% ซึ่งแม้จะยังคงมีกฎความสัมพันธ์ที่อาจไม่มีประโยชน์อยู่บ้าง แต่เทคนิคที่นำเสนอสามารถลดจำนวนกฎที่ไม่เกี่ยวข้องออกได้เป็นจำนวนมาก จึงทำให้สามารถค้นหากฎความสัมพันธ์ที่ต้องการได้มีประสิทธิภาพดีกว่าเดิม รายละเอียดการวัดผลประสิทธิภาพการค้นหาสามารถตรวจสอบได้จากค่า F-Measure ซึ่งจะได้กล่าวถึงต่อไป

จากการทดลองกับข้อมูลทดสอบข้างต้น จะได้ผลการทดสอบประสิทธิภาพแสดงดังตารางที่ 3-5 และรูปที่ 5 สรุปได้ว่าแนวทางที่นำเสนอในงานวิจัยสามารถปรับปรุงคุณภาพการสืบค้น (วัดจากค่า F-Measure) ได้เฉลี่ย 17.78% เพิ่มขึ้นจากกรณีที่ใช้อัลกอริทึมอพริโอริเพียงอย่างเดียวที่มีค่า F-Measure เฉลี่ยที่ 16.06% ถึง 10.71% (ตารางที่ 3) โดยให้ผลการสืบค้นที่ดีในค่าสนับสนุนต่ำ (แสดงได้จากกราฟในรูปที่ 5) ซึ่งผลการทดลองสนับสนุนกับลักษณะการสืบค้นตามอัลกอริทึมอพริโอริ ที่จะได้ความสัมพันธ์จำนวนมากบนค่าสนับสนุนขั้นต่ำที่น้อย โดยความสัมพันธ์ส่วนใหญ่ที่ได้จะไม่มีความน่าสนใจ เมื่อพิจารณาค่า Precision จะพบว่าแนวทางที่นำเสนอให้ผลการสืบค้นที่ดีกว่าด้วยค่าเฉลี่ย 43.45% ปรับปรุงเพิ่มขึ้น 67.89% (ตารางที่ 4) ในส่วนของ Recall นั้นพบว่าแนวทางที่นำเสนอให้ผลการสืบค้นที่แย่กว่าได้ค่าเฉลี่ย 11.34% ลดลง 12.50% (ตารางที่ 5) อันเนื่องมาจากกระบวนการที่นำเสนอเป็นกระบวนคัดออกเพิ่มเติมเท่านั้น ไม่ได้เป็นการปรับปรุงกระบวน การสร้างแต่อย่างไร

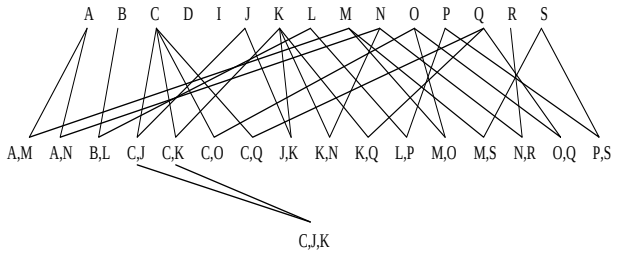
อนึ่งเป็นที่น่าสังเกตที่ค่า F-Measure และค่า Recall ในการทดลองมีค่าต่ำมากทั้งสองเทคนิค สาเหตุเนื่องจากเป็นข้อมูลจริงที่คัดมาทำเป็นข้อมูลทดสอบ มีจำนวนไอเทมถึง 19 ไอเทมในฐานข้อมูลจำนวน 40 รายการ บนขนาดไอเทมเซตสูงสุด 5 ไอเทมและ 5 Attribute ซึ่งสามารถสร้างความสัมพันธ์ได้สูงสุดถึง 2,047 รูปแบบ จากการตรวจสอบโดยผู้ใช้งานพบว่ามีความสัมพันธ์ที่ถูกคัดเลือกว่าน่าสนใจจำนวนถึง 144 ความสัมพันธ์ และจากผลการทดลองด้วยเทคนิคอพริโอริและเทคนิคที่นำเสนอสามารถค้นหาฟรีเคว้นไอเทมเซตได้อยู่ในช่วง 17-115 และ 11-34 ไอเทมเซตตามลำดับ (เรียงลำดับจากค่าสนับสนุนมากไปน้อย) ซึ่งหมายถึงมีค่า Recall อยู่ในช่วง 11.81%-79.86% และ 7.64%-23.6% หรือค่า F-Measure สูงสุดอยู่ในช่วง 21.13%-88.80% และ 14.20%-38.19% เท่านั้น ซึ่งจะเห็นว่าเทคนิคที่นำเสนอสามารถให้ค่า F-Measure ที่ใกล้เคียงค่าสูงสุดได้มากกว่า

สำหรับข้อมูลทดสอบส่วนที่ 2 เป็นการทดสอบกับฐานข้อมูลการใช้งานจริงของโรงพยาบาลจำนวน 10,000 รายการ บนเทคนิคที่นำเสนอเทียบจากผลการสืบค้นของอัลกอริทึมอพริโอริปกติ ผลการทดสอบพบว่ามีกฎความสัมพันธ์ที่ไม่น่าสนใจจำนวนมากที่ถูกคัดออกด้วยแนวทางที่นำเสนอ โดยกฎเหล่านั้นเป็นกฎที่มีค่าสนับสนุนสูงที่ผ่านการคัดเลือกด้วยอัลกอริทึมอพริโอริ แต่ไม่มีความเกี่ยวเนื่องระหว่างคู่ Attribute ของไอเทมที่เป็นซับเซต (ค่า Entropy Gain น้อยเกินไป) ตัวอย่างของกฎความสัมพันธ์ที่ถูกคัดออกแสดงดังตารางที่ 6

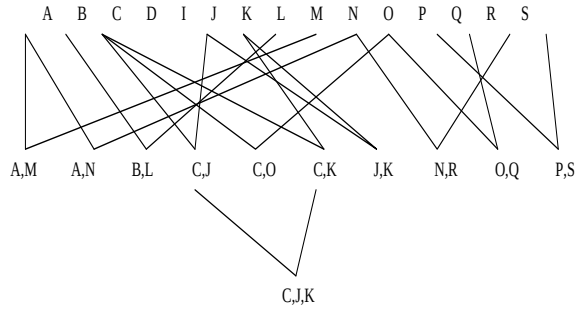
อนึ่งในส่วนของกฎความสัมพันธ์ที่สืบค้นได้ เนื่องจากข้อมูลที่นำมาทดสอบเป็นข้อมูลด้านการให้บริการของแผนกในโรงพยาบาล ตัวอย่างของกลุ่มของกฎความสัมพันธ์ที่เป็นประโยชน์ที่สืบค้นได้แสดงดังตารางที่ 7 ซึ่งผู้บริหารของโรงพยาบาลสามารถนำความรู้ที่สืบค้นได้ไปใช้ปรับปรุงการให้บริการของโรงพยาบาลได้ในอนาคต ซึ่งสามารถนำไปใช้เป็นข้อมูลในด้านการประกอบกิจการเพื่อแข่งขันกับโรงพยาบาลอื่นได้อีกด้วย

ตารางที่ 1. ตารางแทนชื่อข้อมูลด้วยตัวอักษร

Attribute Item	Symbol	Attribute Item	Symbol
Capacity 51-60%	A	Doctor A	K
Capacity 61-70%	B	Doctor B	L
Capacity 71-80%	C	Doctor C	M
Monday	D	Teacher	N
Tuesday	E	Merchant	O
Wednesday	F	Farmer	P
Thursday	G	Age <20	Q
Friday	H	Age 20-40	R
Saturday	I	Age >40	S
Sunday	J		



รูปที่ 3. ผลการทดสอบด้วย Apriori Algorithm เพียงอย่างเดียวด้วยค่าสนับสนุนขั้นต่ำเท่ากับ 7



รูปที่ 4. ผลการทดสอบด้วย Apriori Algorithm ร่วมกับการใช้ Entropy Prune ด้วยค่าสนับสนุนขั้นต่ำเท่ากับ 7 และค่าความเกี่ยวข้องขั้นต่ำ 18%

ตารางที่ 2. กฎความสัมพันธ์ที่ได้จากการทำ Apriori และกฎความสัมพันธ์ที่ได้จากการทำ Apriori with minimum entropy gain ที่สนับสนุนขั้นต่ำเท่ากับ 7 และค่าความเกี่ยวข้องขั้นต่ำ 18%

Association Rule of Apriori (Only)		Asso. Rule of Apriori with minimum entropy gain	
A⟹M	M⟹O	A⟹M	O⟹Q
A⟹N	M⟹S	A⟹N	P⟹S
B⟹L	N⟹R	B⟹L	C∧J⟹K
C⟹J	O⟹Q	C⟹J	C∧K⟹J
C⟹K	P⟹R	C⟹K	J∧K⟹C
C⟹O	C∧J⟹K	C⟹O	C⟹J∧K
C⟹Q	C∧K⟹J	J⟹K	J⟹C∧K
J⟹K	J∧K⟹C	N⟹R	K⟹C∧J
K⟹N	C⟹J∧K		
K⟹Q	J⟹C∧K		
L⟹P	K⟹C∧J		

ตารางที่ 3. ผลการทดสอบกับค่า F-Measure

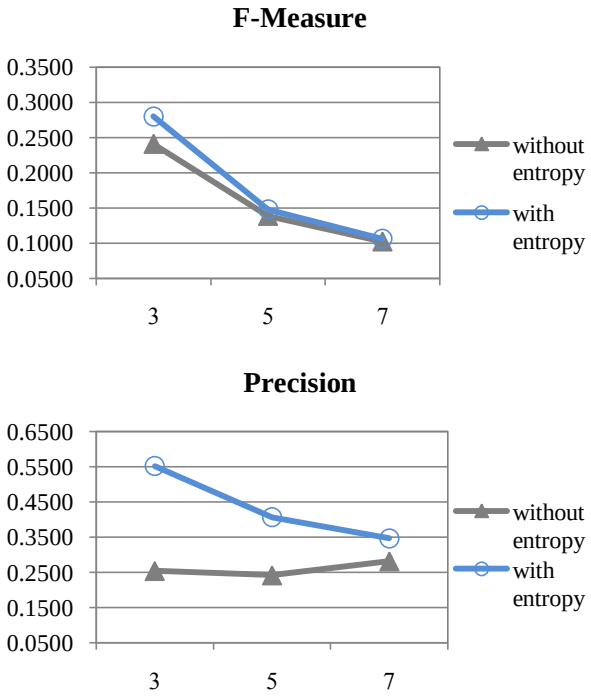
Apriori Algo. minimum support	F-Measure	
	Without minimum entropy gain	With minimum entropy gain
3	0.2409	0.2798
5	0.1386	0.1477
7	0.1023	0.1059
Average	0.1606	0.1778

ตารางที่ 4. ผลการทดสอบกับค่า Precision

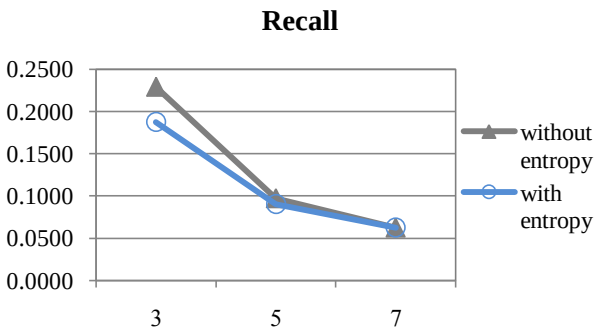
Apriori Algo. minimum support	Precision	
	Without minimum entropy gain	With minimum entropy gain
3	0.2538	0.5510
5	0.2414	0.4063
7	0.2813	0.3462
Average	0.2588	0.4345

ตารางที่ 5. ผลการทดสอบกับค่า Recall

Apriori Algo. minimum support	Recall	
	Without minimum entropy gain	With minimum entropy gain
3	0.2292	0.1875
5	0.0972	0.0903
7	0.0625	0.0625
Average	0.1296	0.1134



รูปที่ 5. กราฟผลการทดสอบกับค่า F-Measure, Precision และ Recall



รูปที่ 5. กราฟผลการทดสอบกับค่า F-Measure, Precision และ Recall (ต่อ)

ตารางที่ 6. ตัวอย่างความสัมพันธ์ที่ถูกคัดออกในกระบวนการวิจัย

ความสัมพันธ์ที่คัดออก	เหตุผลประกอบ
(หมอ, จังหวัดของคนไข้ที่อยู่ในสมุทรสาคร)	เป็นโรงพยาบาลประจำอำเภอ คนไข้ส่วนใหญ่อยู่ในจังหวัดสมุทรสาครอยู่แล้ว Attribute หมอและจังหวัดจึงไม่มีความเกี่ยวข้องกันในเชิงความสัมพันธ์ของข้อมูล
(คลินิกประกันสังคม, ค่าบริการ 0-100 บาท)	ค่าบริการส่วนใหญ่เป็น 30 บาทอยู่แล้ว คลินิกประกันสังคมจึงไม่มีความเกี่ยวข้องในเชิงความสัมพันธ์ของข้อมูลกับค่าบริการ

ตารางที่ 7. ตัวอย่างความสัมพันธ์ที่ถูกสืบค้นได้ในกระบวนการวิจัย

กลุ่มความรู้	ความสัมพันธ์ที่ค้นได้
อาชีพ, การใช้สิทธิ์	ความสัมพันธ์ของการใช้สิทธิ์ของคนไข้ในแต่ละกลุ่มอาชีพ
อาชีพ, ช่วงเวลาการรับบริการ	ความสัมพันธ์ของอาชีพต่อช่วงเวลาการให้บริการต่อการรับเข้าบริการในแต่แผนก
ที่อยู่คนไข้, แผนก	ความสัมพันธ์ระหว่างตำบลที่คนไข้พักและแผนกที่เข้ารับบริการ (สำหรับวางแผนการขยายคลินิกเคลื่อนที่ในการให้บริการ)
ที่อยู่คนไข้, ช่วงเวลา, แผนก	ความสัมพันธ์ระหว่างที่อยู่ของคนไข้ วัน/เวลาการเข้ารับบริการของแต่ละแผนก (สำหรับวางแผนเวลาการให้บริการคลินิกพิเศษ)

5. สรุป

การนำเทคนิคการสืบค้นความรู้ โดยเฉพาะ การนำกฎความสัมพันธ์มาใช้กับฐานข้อมูลการให้บริการของโรงพยาบาล จะทำให้โรงพยาบาลทราบถึงความสัมพันธ์ของรูปแบบการให้บริการที่ซ่อนอยู่ในฐานข้อมูล อันสามารถนำไปใช้ในการวางแผนและปรับปรุงการให้บริการในปัจจุบัน รวมถึงการเตรียมการในอนาคตให้กับผู้บริหาร หรือบุคลากรของโรงพยาบาลได้ งานวิจัยนี้จึงนำเสนอกระบวนปรับปรุงการค้นหากฎความสัมพันธ์ (Association Rule) ให้มีประสิทธิภาพมากยิ่งขึ้น โดยพิจารณาความเกี่ยวข้องระหว่าง Attribute ของไอเทมของกฎความสัมพันธ์ที่สืบค้นจากอัลกอริทึมอพริโอริเพิ่มเติม ผ่านทางค่า Entropy Gain

จากผลการทดลองด้วยกระบวนการที่นำเสนอ ผลปรากฏว่าสามารถคัดเลือกกฎความสัมพันธ์ที่มีความน่าสนใจได้มีประสิทธิภาพมากขึ้นและมีจำนวนกฎความสัมพันธ์ที่ไม่เกี่ยวข้องน้อยลง ทำให้เหมาะสมกับการนำไปใช้งานมากยิ่งขึ้นได้

อย่างไรก็ดี งานวิจัยนี้ไม่ได้มุ่งเน้นในเรื่องของความเร็วในการประมวลผลของอัลกอริทึมที่ใช้รวมถึงการใช้ทรัพยากรของเครื่องคอมพิวเตอร์ ดังนั้นแนวทางในการพัฒนาในอนาคต จึงจะปรับปรุงความเร็วในการประมวลผลให้เร็วมากขึ้น และลดการใช้ทรัพยากรบนเครื่องคอมพิวเตอร์ เพื่อจะทำให้สามารถนำไปใช้กับฐานข้อมูลที่มีขนาดใหญ่ได้

เอกสารอ้างอิง

[1] A.Taleb, C.Jutten, “Entropy optimization, application to blind source separation”, ICANN, Lausanne, Switzerland, oct 1997, pp529 – 534.

[2] J. Han and M. Kamber, “Data Mining Concepts and Techniques” 2nd Edition, Margan Kaufman Publishers, 2006.

[3] L. Kaufman and P. J. Rousseeuw, (1990), “Finding Groups in Data: an Introduction to Cluster Analysis”, John Wiley and Sons.

[4] M. Berry and G. Linoff, “Data Mining Techniques: For Marketing Sales and Customer Support”, John Wiley & Sons, 1997.

[5] M. S. Chen, J. Han, and P. S. Yu., “Data Mining: An overview from a database perspective”, IEEE Trans. Knowledge and Data Engineering, 8:866-883, 1996.

[6] R. Agrawal and R. Srikant, “Fast Algorithms for Mining Association Rules”, Proc. of the 20th Int. Conf. on VLDB, Santiago, Chile, 1994.

[7] R. Agrawal, T. Imielinski, and A. Sawami. “Mining Association Rules between sets of items in large database”, SIGMOD’93, pages 207-216, Washington,D.C.

[8] Richard C. Dubes and Anil K. Jain, (1988), “Algorithms for Clustering Data”, Prentice Hall.

[9] T. Mitchell. “Machine Learning”, McGraw-Hill, 1977.

[10] T. Ouypornkochagorn and K. Waiyamai, "Formal Concept Mining: A Statistic-Based Approach for Pertinent Concept Lattice Construction", in Proc. ASIAN, 2004.

[11] กฤษณะ ไวยมัย, ชิดชนก ส่งศิริ และธนาวินท์ รักธรรมานนท์ ., “การใช้เทคนิคดาต้าไมน์นิงเพื่อพัฒนาคุณภาพการศึกษานิสิตคณะวิศวกรรมศาสตร์”.NECTEC Technical Journal. Vol. III, No. 11, 2001.