

เว็บล็อกทรานเซกชันประกอบไปด้วยข้อมูลที่บันทึกการเรียกใช้งานของผู้เรียกใช้ งานวิจัยนี้จะนำเสนออัลกอริทึมซึ่งมีหน้าที่หลักในการค้นหาเพื่อใช้ในการทำนาย ข้อมูลที่เป็นลักษณะเว็บเพจนั้นแตกต่างจากข้อมูลทั่วไปตรงที่ลำดับการเรียกใช้เว็บเพจ (user's next requests) นั้นจะมีความสำคัญมาก ดังจะเห็นว่าเมื่อนำวิธีการ Sequential Pattern กับข้อมูลที่เป็นเว็บเพจนั้นจะมีการกระโดดของข้อมูลเกิดขึ้น เมื่อนำผลที่ได้นั้นเมื่อนำไปใช้งานในการทำนายจะทำให้ผลลัพธ์ที่มีประสิทธิภาพต่ำ

จากการศึกษาทำให้ทราบถึงปัญหาว่าการค้นพบกฎและการลดกฎความสัมพันธ์ดังกล่าว ขนาดของการสับกฎ (slide window) จะถูกกำหนดไว้ตายตัว ซึ่งจะส่งผลโดยตรงต่อการค้นหา กฎ ดังเช่น เมื่อมีการกำหนดการสับกฎที่มีขนาดเล็กเกินไป จะทำให้กฎที่มีขนาดยาวไม่สามารถค้นพบได้ ในทางกลับกันหากกำหนดการสับกฎมากเกินไป ก็จะทำให้เกิดการสูญเสียกฎที่มีขนาดเล็กเพราะไม่สามารถค้นพบได้เช่นกัน โดยการแก้ปัญหาจะให้วิธีการแบบ Adaptive Window นั้นหมายถึงว่าจะมีการสับกฎได้ทั้งกฎที่มีขนาดเล็กและใหญ่ ซึ่งผลลัพธ์ที่ได้จะทำให้สามารถค้นพบทั้งกฎที่มีรูปแบบสั้นและยาวนั่นเอง ทำให้ความแม่นยำในการทำนายสูงขึ้นด้วย

หลักการในการพิจารณากฎจะให้ Path-based Model โดยการสร้างกฎที่มีความยาวของเว็บเพจมากขึ้นจะใช้วิธีการ Join Operation เพื่อสร้าง A New Candidate Generations (Ck) โดยวิธีการนี้จะทำให้เวลาในการประมวลผลลดน้อยลงไปด้วย เพราะเหตุว่ากฎหรือรูปแบบการเรียกใช้ บางรูปแบบที่มีความน่าจะเป็นน้อยมากก็จะถูกลดทอนออกไปนั่นเอง

ในการทดลองได้มีการกำหนดกฎหลักขึ้นมา 5 กฎ จากนั้นได้มีการสร้างข้อมูลประกอบเพิ่มเติมดังตารางเว็บล็อก แล้วนำข้อมูลผ่านขั้นตอนของอัลกอริทึมที่ได้ออกแบบไว้ หลังจากนั้นจะได้ข้อมูลที่เป็นลักษณะต้นไม้ (Tree) โดยผลลัพธ์ที่ได้จะถูกลดทอนบางกฎด้วยค่า Confidence และ Page Prediction โดยขั้นตอนสุดท้ายจะมีการพรมนึ่งบางโหนดทิ้งเพื่อให้เหลือเฉพาะกฎที่มีความแม่นยำสูงอีกทั้งผลลัพธ์ที่ได้ก็จะปรากฏกฎที่เราได้กำหนดไว้ด้วยเช่นกัน

Web Log Transaction composed of data that records web page usage of users. This research will purpose algorithm that has major responsibility in searching rules for prediction. Data in form of webpage is different from any other data in sequence of webpage usage as user's next requests that is very important. Sequential Pattern applies to data of webpage will occur the skip in data error and this will create lower effective result.

Referring to the problem, we could learn that relative rule incurred and decreased affects from slide window that is fixed. This result in rules searching for example when slide window is too small it will result in long rule cannot be found. Vice versa, in case of too much slide window, the lost of small rule will be incurred because it could not be found neither. The solution is Adaptive Window that is switching both small and large rules that result in the foundation of both short and long rules. Also the correction of prediction increases.

The principle in rules consideration will create Path-based Model by long rule establishment more and apply the Join Operation method in order to create A New Candidate Generation (Ck). By this method, the time consumption in processing will be reduced because some rules or usage patterns that are lower possibility would be tare off.

In part of experiment of this research, 5-principle rules then the additional data used in web log table will be established. After that the data is processed by algorithm would be designed then the data as Tree result in decreased some rules by Confidence and Page Prediction. The last process is pruning some nodes in order to leave only rule that is very concise and result in what we specify.