

ลักษมี โชมโนทัย : การค้นหาความสัมพันธ์ที่แข็งแกร่ง (MINING STRONG ASSOCIATION RULES) อาจารย์ที่ปรึกษา : ผศ. ดร.กิตติศักดิ์ เกิดประสพ, 133 หน้า.
ISBN 974-533-518-5

กฎความสัมพันธ์ คือ กฎในลักษณะ ถ้า-แล้ว ใช้อธิบายความสัมพันธ์ในกลุ่มข้อมูลว่า ถ้าเกิดลักษณะหนึ่งแล้วจะเป็นผลให้เกิดอีกลักษณะหนึ่งตามมา กฎความสัมพันธ์ที่มีความแข็งแกร่ง หมายถึง กฎความสัมพันธ์ที่มีจำนวนข้อมูลสนับสนุนมาก และมีค่าความถูกต้องสูง การทำเหมืองข้อมูลในลักษณะการค้นหาความสัมพันธ์ภายในกลุ่มข้อมูลมีประโยชน์ เพื่อให้ได้กฎที่มีความสัมพันธ์แข็งแกร่งจะต้องทำการค้นหาจากข้อมูลที่มีปริมาณมาก ต้องมีจำนวนข้อมูลสนับสนุนและความถูกต้องสูง เพื่อความน่าเชื่อถือของกฎที่ได้ และสามารถนำผลที่ได้ไปใช้ได้อย่างมีประสิทธิภาพ แต่อุปสรรคในการทำเหมืองข้อมูลกับข้อมูลทรานแซคชันคือปริมาณของข้อมูลที่มีมาก ทำให้การประมวลผลเพื่อค้นหาความสัมพันธ์ภายในข้อมูลต้องใช้เวลามาก งานวิจัยนี้จึงมีจุดมุ่งหมายเพื่อเพิ่มความเร็วในการค้นหาความสัมพันธ์ที่มีความแข็งแกร่งจากข้อมูลปริมาณมาก ด้วยเทคนิคการสุ่มข้อมูลโดยผลลัพธ์ที่ได้จะต้องไม่แตกต่างจากการค้นหาจากข้อมูลทั้งหมด

จากการศึกษาพบว่าข้อมูลสุ่มเพียง 1 เปอร์เซ็นต์ สามารถให้ผลการค้นพบกฎความสัมพันธ์ที่แข็งแกร่งได้เทียบเท่ากับการใช้ข้อมูลทั้งหมด และสามารถลดเวลาที่ใช้ในการค้นหาความสัมพันธ์ที่แข็งแกร่งได้มากกว่า 95 เปอร์เซ็นต์

จากผลการศึกษาเบื้องต้นนี้ผู้วิจัยได้พัฒนาโปรแกรม เพื่อค้นหาความสัมพันธ์ที่แข็งแกร่ง โดยเพิ่มส่วนการสุ่มข้อมูลในอัลกอริทึมเอไพรออริด้วยภาษา SQL และปรับปรุงการทำงานของโปรแกรมให้สามารถรองรับการทำงานกับข้อมูลสตรีมมิง

สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2548

ลายมือชื่อนักศึกษา สกลพล โชมโนทัย
ลายมือชื่ออาจารย์ที่ปรึกษา กิตติศักดิ์ เกิดประสพ
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม วิมล น.

LAKSAMEE KHOMNOTAI : MINING STRONG ASSOCIATION RULES.

THESIS ADVISOR : ASST. PROF. KITTISAK KERDPRASOP, Ph.D.,

133 PP. ISBN 974-533-518-5

APRIORI/APRIORI IN SQL/STRONG ASSOCIATION RULES/RANDOM
SAMPLING/STREAMING

Association rules are if-then rules that are used to express the causes and effects among items in a dataset. Strong association rules are association rules with high support and high confidence. Association rule mining for discovering strong association rules requires a very large dataset to assure high data support and high confidence. Time usage is an important problem in association rule mining for strong rules, especially when mining from large transactional data. We are interested in studying the technique to decrease time usage in association rule mining from a very large dataset. We use random sampling technique to mine strong association rules from a sampled transactional data with the constraint that the discovered rules have to be identical to these discovered from the original large transactions.

From the experimental results, we have found that a 1% random sample of the data can yield exactly the same set of strong rules as those obtained from the large original transaction data set. By means of sampling, we can save as much as 95% of association mining time.

The results of the preliminary study lead to the development of strong association mining program. The program is an extension of APRIORI algorithm to include the sampling phase. We also design our program to handle streaming data.

School of Computer Engineering

Academic Year 2005

Student's Signature Wassamee Khomnoi

Advisor's Signature Kittisak Kungpoo

Co-advisor's Signature Nittayong Kungpoo