

นันทวุฒิ คะอังกู : วิธีแบ่งช่วงข้อมูลสำหรับการหากฎความสัมพันธ์

(A DISCRETIZATION METHOD FOR ASSOCIATION RULE MINING)

อาจารย์ที่ปรึกษา: รองศาสตราจารย์ ดร. นิตยา เกิดประสพ, 88 หน้า.

งานวิจัยนี้ได้ศึกษาปัญหาการแบ่งช่วงข้อมูลสำหรับการหากฎความสัมพันธ์ ซึ่งการแบ่งช่วงข้อมูลเป็นการจัดการกับข้อมูลที่เป็นตัวเลขต่อเนื่องด้วยการจัดเป็นช่วงของค่า การแบ่งช่วงนี้เป็นขั้นตอนก่อนนำข้อมูลไปเข้ากระบวนการหากฎความสัมพันธ์เพื่อช่วยลดจำนวนกฎความสัมพันธ์ที่ได้ออกมาและช่วยให้ได้กฎที่มีค่าสนับสนุนที่สูงขึ้น ในอดีตได้มีหลากหลายงานวิจัยที่ได้เสนอเทคนิคในการแบ่งช่วงของข้อมูล แต่เทคนิคส่วนใหญ่ทำการแบ่งช่วงของข้อมูลสำหรับงานทำเหมืองข้อมูลประเภทการจำแนกข้อมูล ซึ่งวัดประสิทธิภาพของเทคนิคได้ชัดเจนจากการพิจารณาค่าความถูกต้องของโมเดล แต่การแบ่งช่วงข้อมูลเพื่อใช้กับการหากฎความสัมพันธ์จะวัดประสิทธิภาพได้ยากกว่า จึงทำให้มีงานวิจัยด้านนี้ปรากฏค่อนข้างน้อย ผู้วิจัยได้เห็นความสำคัญในจุดนี้จึงได้เสนอเทคนิคการแบ่งช่วงข้อมูลที่เป็นตัวเลขต่อเนื่องสำหรับการหากฎความสัมพันธ์ เพื่อเพิ่มประสิทธิภาพในการหากฎความสัมพันธ์ให้ดียิ่งขึ้น โดยจะทำการเปรียบเทียบประสิทธิภาพการแบ่งช่วงข้อมูลที่ได้เสนอในการวิจัยนี้กับเทคนิคการแบ่งช่วงข้อมูลอื่นด้วยวิธีอื่น ๆ โดยใช้ไลบรารีภาษาอาร์ในส่วนของข้องกับการหากฎความสัมพันธ์ งานวิจัยนี้จะใช้ค่าความถูกต้อง และมาตรวัด 4 มาตรวัดในการประเมินประสิทธิภาพกฎความสัมพันธ์

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ปีการศึกษา 2555

ลายมือชื่อนักศึกษา \_\_\_\_\_

ลายมือชื่ออาจารย์ที่ปรึกษา \_\_\_\_\_

NUNTAWUT KAOUNGKU : A DISCRETIZATION METHOD FOR  
ASSOCIATION RULE MINING. THESIS ADVISOR : ASSOC. PROF.  
NITTAYA KERDPRASOP, Ph.D., 88 PP.

#### DISCRETIZATION/ASSOCIATION RULE MINING/CHI SQUARE

In this research, we study the problem of discretization for association rule mining. The discretization method is a handling technique to cope with numerical or continuous data. It is the pre-processing step of association rule mining to reduce the number of rules, and to increase the support value. In the past, many researchers have proposed techniques for discretization, but most of them perform a discretization for classification in which the performance measure is obviously the accuracy of the model. The discretization method for association rule mining, on the contrary, has no clear evaluation metric. We thus propose in this research the discretization technique for a specific task of association rule mining, and the rule assessment method. The implementation and experimentation have been done with the R language and its libraries. We comparatively experiment our proposed discretization method with the existing techniques. Assessment metrics are accuracy and the other four metrics.

School of Computer Engineering

Academic Year 2012

Student's Signature \_\_\_\_\_

Advisor's Signature \_\_\_\_\_