

พงศกร คุรงค์คำรงค์ชัย : การค้นหาและจัดอันดับกฎความสัมพันธ์เพื่อวิเคราะห์ปัจจัยเสี่ยงที่นำไปสู่การเกิดโรค(ASSOCIATION RULE SEARCH AND RANKING FOR THE ANALYSIS OF FACTORS CONTRIBUTING TO DISEASES)อาจารย์ที่ปรึกษา:
รองศาสตราจารย์ ดร.นิตยา เกิดประสพ, 121หน้า

งานวิจัยนี้ได้ศึกษาปัญหาการค้นหากฎความสัมพันธ์กับข้อมูลทางการแพทย์ที่มีลักษณะของข้อมูลผสมระหว่างข้อมูลที่เป็นตัวเลขและข้อมูลที่เป็นข้อความ การค้นหาความสัมพันธ์ในงานวิจัยนี้ใช้อัลกอริทึมเอ็พไรเออร์ อัลกอริทึมพรีดิคชันเอ็พไรเออร์ และอัลกอริทึมเทอเซียสแต่เนื่องจากการค้นหาความสัมพันธ์แบบปกตินั้นไม่สามารถทำได้กับข้อมูลที่อยู่ในลักษณะของตัวเลข ฉะนั้นงานวิจัยนี้จึงได้ใช้การแบ่งช่วงข้อมูลเข้ามาช่วยเพื่อให้สามารถทำการค้นหาความสัมพันธ์ได้ในกรณีข้อมูลสูญหายหรือไม่สามารถระบุค่า ผู้วิจัยใช้วิธีลบข้อมูลทรานแซคชันที่ไม่สามารถระบุข้อมูลได้ชัดเจนทิ้งซึ่งเป็นข้อมูลส่วนน้อย การแบ่งช่วงข้อมูลและการลบข้อมูลที่ไม่สามารถระบุข้อมูลได้ชัดเจนนั้นจัดอยู่ในขั้นตอนการเตรียมข้อมูลก่อนการค้นหาความสัมพันธ์ ในอดีตได้มีงานวิจัยจำนวนมากที่ได้เสนอเทคนิคในการค้นหาความสัมพันธ์ในรูปแบบต่าง ๆ และทำการวัดประสิทธิภาพของกฎความสัมพันธ์ในแง่ของค่าความเชื่อมั่นเพียงอย่างเดียว แต่งานวิจัยในด้านการค้นหาความสัมพันธ์ในรูปแบบของการรวมกฎความสัมพันธ์และจัดอันดับกฎความสัมพันธ์รวมไปถึงการลดจำนวนกฎความสัมพันธ์ที่มีความซ้ำซ้อนนั้นค่อนข้างมีความซับซ้อนในกระบวนการทำงาน จึงทำให้งานวิจัยทางด้านนี้มีปรากฏค่อนข้างน้อย ผู้วิจัยได้เห็นถึงความสำคัญในจุดนี้จึงได้เสนอเทคนิคการเพิ่มกระบวนการทั้งก่อนและหลังการค้นหาความสัมพันธ์กับข้อมูลทางการแพทย์ที่ต้องอาศัยทั้งความละเอียดและรอบคอบในการวิเคราะห์ข้อมูล อีกทั้งผู้วิจัยยังได้เสนอมาตรวัดประสิทธิภาพใหม่ที่เรียกว่า Confidence and Accuracy (CAA) ซึ่งเป็นมาตรวัดที่เป็นการผสมผสานกันระหว่าง ค่าความเชื่อมั่นและค่าความแม่นยำ เพื่อสร้างเกณฑ์ในการค้นหาความสัมพันธ์ที่มีความครอบคลุมและสามารถวิเคราะห์ปัจจัยเสี่ยงได้อย่างชัดเจน

สาขาวิชาวิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2557

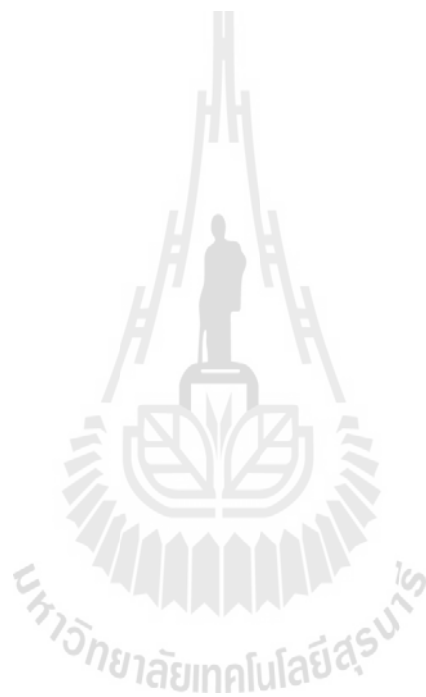
ลายมือชื่อนักศึกษา _____
ลายมือชื่ออาจารย์ที่ปรึกษา _____

PONGSAKORN DURONGDUMRONGCHAI: ASSOCIATION
RULESEARCH AND RANKING FOR THE ANALYSIS OF FACTORS
CONTRIBUTING TO DISEASES. THESIS ADVISOR:
ASSOC. PROF. NITTAYA KERDPRASOP, Ph.D., 121 PP.

ASSOCIATION RULES/MERGED RULE/RANKING RULE / REMOVED
REDUNDANT RULE

This research was studied the problem of association rule search for medical datasets with appearance numeric and nominal of data mixtures. Association rule search in this research use 3 algorithm including Apriori, Predictive Apriori, and Tertius. However, normal association rule search algorithm cannot deal with numeric data. Therefore, this research applies discretization technique to enable the association rule search. In the case of data with missing or unidentified values, we decide to remove such transactions, which rarely occur in our datasets. Research in the past offered several techniques for finding association rules in various forms, and measured the performance of the relationship in terms of confidence, which is a single measurement. However, there are a few research in the field of association rule search in the form of merged rule and ranking rules as well as removed redundant association rules, because these processes are rather complicated. We realize the importance of this point and thus propose the pre-processing and post-processing association rule mining techniques to work with the medical datasets that require detailed and careful in data analysis.

Moreover we also propose a new performance measure called Confidence and Accuracy(CAA) to integrate the confidence and accuracy metrics for the complete and accurate analysis of factors contributing to diseases.



School of Computer Engineering

Academic Year 2014

Student's Signature _____

Advisor's Signature _____