30 DEC 1999

LIBRARY AND INFORMATION CENTER
MAHIDOL UNIVERSITY

KNOWLEDGE DISCOVERY IN DATABASE BASED ON KOHONEN SELF-ORGANIZING MAP ALGORITHM

SURAJIT NUDDAMONGKOL

With compliments
of
บัณฑิตวิทยาลัย มหิดล

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(COMPUTER SCIENCE)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
1999

ISBN 974-662-915-8

COPYRIGHT OF MAHIDOL UNIVERSITY

Copyright © by Mahidol University

TH
S961k
1999

.43303 c.2

4037540 SCCS/M : MAJOR : COMPUTER SCIENCE ; M.Sc. (COMPUTER SCIENCE)
KEY WORDS : KNOWLEDGE DISCOVERY IN DATABASE / DISCRETIZATION / KOHONEN SELF-ORGANIZING MAP
SURAJIT NUDDAMONGKOL : KNOWLEDGE DISCOVERY IN DATABASE BASED ON KOHONEN SELF-ORGANIZING MAP ALGORITHM. THESIS ADVISORS : SUPACHAI TANGWONGSAN Ph.D., JARERNSRI L. MITRPANONT Ph.D. 85 p. ISBN 974-662-915-8

Knowledge Discovery in Database (KDD) is a new methodology for searching and extracting potential knowledge from database. Data mining is a key element of KDD process to seek knowledge. However, some of the data mining algorithms are unable to handle numerical value. This problem recently can be solved by adding a process to reduce details of data before data mining, so-called discretization. Discretization is data transformation so that it has less detail by dividing the data value into a set of sub-range in order to decrease processing time of data mining step. However, conventional methods are not quite efficient to manage the data with numerical value because they consider only the value, not others, resulting in inefficient knowledge produced.

This research presents a new approach of discretization on numerical data by changing from value-based to content-based approach. Content-based discretization is based on data content and conforms to Kohonen Self-Organizing Map (KSOM).

KSOM is applied for solving the above problem. It is able to learn structure and relationship of data by self learning on sole input data and to perform data discretization afterward. From research and test, we observe that the result is more accurate and reliable and also takes less processing time than other methods. This research can be further applied for KDD, which gives more meaningful and accurate knowledge.

Copyright by Mahidol University

4037540 SCCS/M : สาขาวิชา : วิทยาการคอมพิวเตอร์; วท.ม. (วิทยาการคอมพิวเตอร์)

สุรจิตต์ นัดดามงคล : การค้นหาคลังความรู้โดยวิธีการจัดผังด้วยตนเองของโคโฮเนน (KNOWLEDGE DISCOVERY IN DATABASE BASED ON KOHONEN SELF-ORGANIZING MAP ALGORITHM). คณะกรรมการควบคุมวิทยานิพนธ์ : ศุภชัย ตั้งวงศ์ศานต์, Ph.D., เจริญศรี มิตรภานนท์, Ph.D. 85 หน้า. ISBN 974-662-915-8

Knowledge Discovery in Database (KDD) เป็นวิธีการใหม่ในการค้นหา และแยกเอาองค์ความรู้ที่มีศักยภาพจากฐานข้อมูล โดย Data Mining เป็นขั้นตอนสำคัญของการทำ KDD เพื่อค้นหาองค์ความรู้ แต่เนื่องจากบาง Data Mining algorithm ไม่สามารถจัดการกับ Numerical value ซึ่งการแก้ไขปัญหาดังกล่าวสามารถทำได้โดยเพิ่มขั้นตอนการทำ Discretization ก่อนการทำ Data Mining โดย Discretization คือการแปลงค่าของข้อมูลให้มีรายละเอียดต่ำลง โดยแบ่งค่าของข้อมูลออกเป็นช่วงย่อย ๆ ซึ่งเป็นการช่วยลดการประมวลผลในการทำ Data Mining ได้อย่างมาก แต่เนื่องจากวิธีการเดิมไม่มีศักยภาพเพียงพอ เพราะพิจารณาจากค่าของข้อมูลเพียงอย่างเดียว จึงส่งผลให้องค์ความรู้ที่ได้มีประสิทธิภาพต่ำ

โครงงานวิจัยนี้ได้เสนอวิธีการใหม่ในการทำ Discretization บน Numerical data จากวิธีการเดิมที่เป็น Value-based ไปสู่ Content-based ซึ่งเป็นการทำ Discretization ที่อยู่บนพื้นฐานของเนื้อหาสาระของข้อมูล โดย Content-based มีแนวความคิดที่สอดคล้องกับ Kohonen Self-Organizing Map (KSOM)

KSOM ได้ถูกนำมาประยุกต์ใช้เพื่อแก้ไขปัญหาดังกล่าว โดยมีคุณสมบัติในการเรียนรู้โครงสร้าง ตลอดจนความสัมพันธ์ของข้อมูล ด้วยการเรียนรู้ของตนเองจากข้อมูลเพียงอย่างเดียว จนสามารถทำ Discretization ข้อมูลได้ จากการศึกษาวิจัยและทดลอง พบว่าผลลัพธ์ที่ได้มีความถูกต้องและมี Reliability มากกว่าวิธีเดิม อีกทั้งโดยเฉลี่ยแล้ว KSOM ใช้เวลาในการทำ Discretization น้อยกว่าวิธีเดิม โดยงานวิจัยนี้สามารถนำไปใช้ในการทำ KDD ซึ่งจะส่งผลให้องค์ความรู้ที่ได้มีความหมาย และมีความถูกต้องมากขึ้น

Copyright by Mahidol University