เทคนิคการแตกครึ่งตามสารสนเทศสำหรับซัพพอร์ตเวกเตอร์แมชชีนแบบหลายประเภท

นางสาวปทุมศิริ สงศิริ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2548

ISBN 974-53-2565-1

AN INFORMATION-BASED DICHOTOMIZATION TECHNIQUE FOR MULTICLASS SUPPORT VECTOR MACHINES

Miss Patoomsiri Songsiri

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2005

ISBN 974-53-2565-1

หัวข้อวิทยานิพนธ์	เทคนิคการแตกครึ่งตามสารสนเทศ สำหรับซัพพอร์ตเวกเตอร์แมชชีนแบบหลายประเภท
โดย	นางสาวปทุมศิริ สงศิริ
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล
อาจารย์ที่ปรึกษาร่วม	ดร.ฐิมาพร เพชรแก้ว

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

..... อาจารย์ที่ปรึกษา

(รองศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... อาจารย์ที่ปรึกษาร่วม

(อาจารย์ ดร.ฐิมาพร เพชรแก้ว)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์)

ปทุมศิริ สงศิริ : เทคนิคการแตกครึ่งตามสารสนเทศสำหรับซัพพอร์ตเวกเตอร์แมชชีนแบบหลายประเภท. (AN INFORMATION-BASED DICHOTOMIZATION TECHNIQUE FOR MULTICLASS SUPPORT VECTOR MACHINES) อ.ที่ปรึกษา: รศ.ดร.บุญเสริม กิจศิริกุล, อ. ที่ปรึกษาร่วม: อ.ดร.ฐิมาพร เพชรแก้ว, 74 หน้า. ISBN 974-53-2565-1.

การแก้ปัญหาการจำแนกแบบหลายประเภทด้วยซัพพอร์ตเวกเตอร์แมชชีนโดยส่วนใหญ่จะพิจารณาเป็นปัญหาของการนำตัวจำแนกแบบสองประเภทหลายตัวมาใช้ร่วมกัน วิธีการดังกล่าวยังมีข้อจำกัดในเรื่องของความถูกต้องและจำนวนครั้งในการจำแนก งานวิจัยนี้นำเสนอวิธีการใหม่ในการจำแนกข้อมูลด้วยเทคนิคการแตกครึ่งตามสารสนเทศโดยสร้างต้นไม้สำหรับการจำแนกแบบไบนารีจากตัวจำแนกแบบสองประเภท ซึ่งแต่ละโนดของต้นไม้จะเป็นตัวจำแนกแบบสองประเภทที่มีค่าเอนโทรปีต่ำที่สุด วิธีนี้สามารถลดจำนวนตัวจำแนกแบบสองประเภทที่ใช้ในการจำแนกลงได้เป็นลอการิทึมของจำนวนประเภทซึ่งต่ำกว่าวิธีอื่น จากผลการทดลองแสดงให้เห็นว่าวิธีนี้สามารถลดจำนวนครั้งของการจำแนกลงได้ และยังคงให้ค่าความถูกต้องใกล้เคียงกับวิธีอื่น

ภาควิชา..... วิศวกรรมคอมพิวเตอร์.....ลายมือชื่อนิสิต.....

สาขาวิชา.....วิทยาศาสตร์คอมพิวเตอร์...ลายมือชื่ออาจารย์ที่ปรึกษา.....

ปีการศึกษา ...2548.....ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

46703630 : MAJOR COMPUTER SCIENCE

KEY WORD: INFORMATION-BASED DICHOTOMIZATION / MULTICLASS SUPPORT VECTOR MACHINES

PATOOMSIRI SONGSIRI: AN INFORMATION-BASED DICHOTOMIZATION TECHNIQUE FOR MULTICLASS SUPPORT VECTOR MACHINES. THESIS ADVISOR: ASSOC. PROF. BOONSERM KIJSIRIKUL, Ph.D., THESIS COADVISOR : THIMAPORN PHETKAEW,Ph.D., 74 pp. ISBN 974-53-2565-1.

Approaches for solving a multiclass classification problem by Support Vector Machines (SVMs) are typically to consider the problem as combination of two-class classification problems. Previous approaches have some limitations in classification accuracy and the number of evaluations. This research proposes a novel method that employs information-based dichotomization for constructing a binary classification tree. Each node of the tree is a binary SVM with the minimum entropy. Our method can reduce the number of binary SVMs used in the classification to the logarithm of the number of classes which is lower than previous methods. The experimental results show that the proposed method takes lower the number of evaluations while it maintains accuracy compared to other methods.

Department..... Computer Engineering...Student's.....Patoomsiri Songsiri

Field of study.....Computer Science.....Advisor's.....

Academic year ...2005.....Coadvisor's.....Thimaporn Phetkaew

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความอนุเคราะห์อย่างยิ่งของอาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล และอาจารย์ที่ปรึกษาร่วม อาจารย์ ดร.ฐิมาพร เพชรแก้ว ซึ่งท่านได้ให้ความรู้ คำแนะนำ และข้อคิดเห็นอันเป็นประโยชน์ต่องานวิจัย จนทำให้การวิจัยในครั้งนี้สำเร็จออกมาด้วยดี

ขอขอบคุณรองศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา และผู้ช่วยศาสตราจารย์ ดร.ญาใจ ลิ่มปิยะกรณ์ กรรมการสอบวิทยานิพนธ์ที่กรุณาเสียสละเวลา ให้คำแนะนำ ตรวจสอบ และแก้ไขวิทยานิพนธ์ฉบับนี้

ขอบคุณสมาชิกห้องปฏิบัติการ Machine Intelligence & Knowledge Discovery (MIND-LAB) และห้องปฏิบัติการอื่นๆ ที่คอยช่วยเหลือให้คำแนะนำและให้กำลังใจด้วยดีเสมอมา

ท้ายที่สุด ผู้วิจัยใคร่ขอกราบขอบพระคุณบิดา มารดา รวมถึงทุกคนในครอบครัวที่ให้กำลังใจและให้การสนับสนุนเรื่อยมาจนสำเร็จการศึกษา

สารบัญ

หน้า

หน้า

สารบัญภาพ

หน้า

สารบัญตาราง

หน้า