การรวมตัวจำแนกด้วยการลงคะแนนร่วมกับกฎใกล้เคียง

โดย

นายอิทธิ์ รมณียางกูร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิทยาการคอมพิวเตอร์ ภาควิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2550

การรวมตัวจำแนกด้วยวิธีลงคะแนนร่วมกับกฎใกล้เคียง

โดย

อิทธิ์ รมณียางกูร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิทยาการคอมพิวเตอร์ ภาควิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2550

Ensemble with Neighbor Rules Voting

By

Itt Romneeyangkurn

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิทยาศาสตรมหาบัณฑิต
สาขาวิทยาการคอมพิวเตอร์ ภาควิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2550

มหาวิทยาลัยธรรมศาสตร์
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

อิทธิ์ รมณียางกูร

เรื่อง

การรวมตัวจำแนกด้วยวิธีลงคะแนนร่วมกับกฎใกล้เคียง

ประธานกรรมการวิทยานิพนธ์ _____
(ดร.ชลวิช นัทรี)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์ _____
(ดร.สุกรี สินธุภิญโญ)

กรรมการวิทยานิพนธ์ _____
(ดร.วนิดา พฤทธิวิทยา)

กรรมการวิทยานิพนธ์ _____
(ดร.วนิดา พฤทธิวิทยา)

คณบดี _____
(ดร.เด่นดวง ประดับสุวรรณ)

บทคัดย่อ

ได้มีการพิสูจน์มาแล้วว่า ผลที่ได้จากการรวมคำตอบจากต้นไม้ตัดสินใจหลายต้นให้ความถูกต้องมากกว่าผลที่ได้จากต้นไม้ตัดสินใจต้นเดียว โดยหลายๆ วิธีของการผลิตต้นไม้ตัดสินใจหลายต้นเริ่มจากเตรียมชุดข้อมูลเรียนรู้สำหรับให้ต้นไม้ตัดสินใจแต่ละต้นเรียนรู้ ด้วยวิธีบูตสแตรปปิง โดยการสุ่มตัวอย่างจำนวนเท่ากับจำนวนตัวอย่างของชุดข้อมูลเรียนรู้ตั้งต้น ดังนั้นตัวอย่างบางตัวอย่างของข้อมูลเรียนรู้ตั้งต้น อาจจะปรากฏมากว่าหนึ่งครั้งในขณะที่บางตัวอย่างอาจไม่ปรากฏเลยในชุดข้อมูลเรียนรู้ใหม่นี้ เพื่อให้ได้ข้อมูลหลายๆ ชุดและนำมาใช้ในการฝึกต้นไม้ตัดสินใจแต่ละต้น วิทยานิพนธ์ฉบับนี้นำเสนอความคิดการนำกฎใกล้เคียงที่ได้จากแต่ละการจำแนกมาร่วมในการลงคะแนน แทนที่จะใช้กฎที่จำแนกตัวอย่างได้เท่านั้นเพราะจะเห็นได้ว่าบูตสแตรปปิงเกิดจากการสุ่มข้อมูลจากข้อมูลเรียนรู้ตั้งต้นเดียวกัน ดังนั้นกฎที่ได้จากแต่ละการจำแนกน่าจะมีความสัมพันธ์กัน จากการทดสอบพบว่าการที่เรานำกฎใกล้เคียงมาร่วมลงคะแนนด้วยนั้นให้ผลที่มีความถูกต้องที่มากขึ้นกว่าการลงคะแนนโดยวิธีปกติ นอกจากนี้ยังสังเกตหาค่าความใกล้เคียงที่น้อยที่สุดที่ทำให้ค่าความถูกต้องมากขึ้นอีกด้วย

Abstract

Ensembles of classifiers have been employed to improve accuracy over single classier. Various methods sequentially bootstrap data set and invoke a base classifier on these different bootstraps. In this paper, we propose an idea based on the use of “similar rules" or “neighbor rules" in voting for a given test example, instead of using only the rule that matches with the test example. From our experimental results, we can conclude that our method achieves comparable accuracy and is significantly better than a regular majority vote. We also empirically derive the least of value of a similarity between rules that gives more accurate result.

กิตติกรรมประกาศ

กราบขอบพระคุณอย่างสูงสุดในความเมตตาและความสามารถของ ผศ.ดร. สุกรี สินธุภิญโญ อาจารย์ที่ปรึกษา

กราบขอบพระคุณอย่างสูงสุดในมารดา
กราบขอบพระคุณอย่างสูงสุดในครูบาอาจารย์ทุกท่านในชีวิตข้าพเจ้า
กราบขอบพระคุณอย่างสูงสุดในญาติพี่น้องทุกท่าน
กราบขอบพระคุณอย่างสูงสุดในเพื่อนๆ พี่ๆ น้องๆ
กราบขอบพระคุณอย่างสูงสุดในบริษัท บีบีคอนเทนต์แอนด์มัลติมิเดียร์ จำกัด

ขอบคุณ พี่โต่ง, พี่เป้, พิ้งค์, น้องเต๋า และ กิ๊ก

นายอิทธิ์ รมณียางกูร
มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2550

สารบัญ

สารบัญตาราง

สารบัญภาพประกอบ