

การรวมตัวจำแนกด้วยการลงคะแนนร่วมกับกฎใกล้เคียง

Ensemble with Neighbor Rules Voting

อิทธิ รมณียางกูร, สุกรี สินธุภิญโญ

บทคัดย่อ

ได้มีการพิสูจน์มาแล้วว่า ผลที่ได้จากการรวมคำตอบจากต้นไม้ตัดสินใจหลายต้นให้ความถูกต้องมากกว่าผลที่ได้จากต้นไม้ตัดสินใจต้นเดียว โดยหลายๆ วิธีของการผลิตต้นไม้ตัดสินใจหลายต้นเริ่มจากเตรียมชุดข้อมูลเรียนรู้สำหรับให้ต้นไม้ตัดสินใจแต่ละต้นเรียนรู้ ด้วยวิธีบูตสแตรปป์ง โดยการสุ่มตัวอย่างจำนวนเท่ากับจำนวนตัวอย่างของชุดข้อมูลเรียนรู้ตั้งต้น ดังนั้นตัวอย่างบางตัวอย่างของข้อมูลเรียนรู้ตั้งต้นอาจจะปรากฏมากกว่าหนึ่งครั้งในขณะที่บางตัวอย่างอาจไม่ปรากฏเลยในชุดข้อมูลเรียนรู้ใหม่นี้ เพื่อให้ได้ข้อมูลหลายๆ ชุดและนำมาใช้ในการฝึกต้นไม้ตัดสินใจแต่ละต้น วิทยานิพนธ์ฉบับนี้นำเสนอความคิดการนำกฎใกล้เคียงที่ได้จากแต่ละการจำแนกมาช่วยในการลงคะแนน แทนที่จะใช้กฎที่จำแนกตัวอย่างได้เท่านั้นเพราะจะเห็นได้ว่าบูตสแตรปป์งเกิดจากการสุ่มข้อมูลจากข้อมูลเรียนรู้ตั้งต้นเดียวกัน ดังนั้นกฎที่ได้จากแต่ละการจำแนกน่าจะมีความสัมพันธ์กัน จากการทดสอบพบว่าการที่เราใช้กฎใกล้เคียงมาช่วยลงคะแนนด้วยนั้นให้ผลที่มีความถูกต้องที่มากขึ้นกว่าการลงคะแนนโดยวิธีปกติ นอกจากนี้ยังสังเกตหาค่าความใกล้เคียงที่น้อยที่สุดที่ทำให้ค่าความถูกต้องมากขึ้นอีกด้วย

Abstract

Ensembles of classifiers have been employed to improve accuracy over single classifier. 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.