

KSUC-O-024

การลดขนาดมิติการวิเคราะห์ความหมายแฝงโดยใช้การเรียนรู้ของเครื่องจักร

สุขสันต์ พรหมบุญเรือง และ ธรรมรัตน์ บุญรอด

สาขาวิชาเทคโนโลยีดิจิทัล คณะบริหารศาสตร์ มหาวิทยาลัยกาฬสินธุ์

บทคัดย่อ

งานวิจัยนี้นำเสนอการลดขนาดมิติการวิเคราะห์ความหมายแฝงโดยใช้การเรียนรู้ของเครื่องจักร วัตถุประสงค์เพื่อคัดเลือกซบเซตดัชนีคำศัพท์ที่เหมาะสมที่สุดสำหรับขั้นตอนการวิเคราะห์ความหมายแฝงและเพื่อใช้เป็นตัวแทนชุดข้อมูลดัชนีคำศัพท์เดิมที่มีประสิทธิภาพมากยิ่งขึ้น แต่พบปัญหาว่าการจัดการประมาณการคำศัพท์ในส่วนที่ไม่เกี่ยวข้องกับเอกสารในขั้นตอนทางคณิตศาสตร์การแยกค่าแบบเดี่ยว (SVD) นั้นในบางกรณีคำศัพท์ที่ใช้ทำดัชนี (Index) จะถูกกำจัดออกไปโดยอาจจะเป็นเอกสารที่ผู้ใช้ต้องการก็เป็นได้ ดังนั้นจึงนำเสนอการคัดเลือกคำศัพท์ที่คลุมเครือในกระบวนการแยกค่าแบบเดี่ยวอยู่ในขั้นตอนวิธีที่ได้รับการคัดเลือกเฉพาะคำศัพท์ที่เหมาะสมที่สุดโดยใช้ขั้นตอนวิธี BPSO มาช่วยคัดเลือกซบเซตคำศัพท์ที่มีค่าน้ำหนักรวมทั้งหมดมีค่าความคล้ายที่มากที่สุด ผลการทดลองพบว่า จำนวนคำศัพท์จากขั้นตอนวิธี LSI สามารถลดมิติได้มากกว่าขั้นตอนวิธี LSIBPSO แต่ขั้นตอนวิธี NMF กับขั้นตอนวิธี LSI และขั้นตอนวิธี PCA มีวิธีการลดมิติแตกต่างกันทำให้ไม่สามารถวัดการลดขนาดมิติได้สำหรับขั้นตอนวิธี NMF และพบว่าสำหรับชุดข้อมูลทดสอบในงานวิจัยนี้ขั้นตอนวิธี PCA สามารถให้ผลค่าความแม่นยำได้มากกว่าขั้นตอนวิธี LSIBPSO และขั้นตอนวิธี LSI และขั้นตอนวิธี NMF

คำสำคัญ: การลดขนาดมิติ การวิเคราะห์ความหมายแฝง เครื่องจักรเรียนรู้