บทคัดย่อ

T163046

วิทยานิพนธ์ฉบับนี้นำเสนอวิธีการใหม่เพื่อเพิ่มประสิทธิภาพให้กับอัลกอริทึมไอเท็มเบสคอลลาบอราทีฟฟิลเตอริ่ง ที่ผ่านมาอัลกอริทึมนี้ออกแบบมาเพื่อแก้ปัญหาเรื่องปริมาณข้อมูล โดยพิจารณาความสัมพันธ์ระหว่างชิ้นข้อมูลที่แตกต่างกันก่อน แล้วจึงนำความสัมพันธ์เหล่านี้มาคำนวณ หาความคล้ายคลึงตามข้อมูลการให้เรตติ้งของชิ้นข้อมูลในสเปซที่ลดลง อย่างไรก็ตามอัลกอริทึมนี้ยังคงประสบกับปัญหาการให้เรตติ้งต่อชิ้นข้อมูลไม่ทั่วถึงและปัญหาชิ้นข้อมูลที่ยังไม่มีการให้เรตติ้งไว้

วิธีการที่นำเสนอเพื่อแก้ไขปัญหาทั้งสองดังกล่าว ประกอบด้วยสามขั้นตอน คือ ขั้นตอนแรกใช้เทคนิคการค้นหากฎความสัมพันธ์เพื่อค้นพบความคล้ายคลึงจากความสัมพันธ์กันระหว่างคุณสมบัติ ขั้นตอนที่สองนำความคล้ายคลึงที่ได้จากขั้นตอนแรกไปคำนวณหาชิ้นข้อมูลที่คล้ายคลึงตามคุณสมบัติ ขั้นตอนที่สามรวมค่าความคล้ายคลึงตามคุณสมบัติและค่าความคล้ายคลึงตามข้อมูลเรตติ้งเข้าด้วยกัน เพื่อใช้ค้นหาชิ้นข้อมูลที่ใกล้เคียงในอัลกอริทึมไอเท็มเบสคอลลาบอราทีฟ-ฟิลเตอริ่งและให้ผลการทำนายค่าความพึงพอใจออกมาโดยอาศัยค่าความคล้ายคลึงรวม

ผลการทดลองพบว่าวิธีการที่นำเสนอสามารถช่วยให้การทำนายถูกต้องดีกว่าอัลกอริทึมไอเท็มเบสคอลลาบอราทีฟฟิลเตอริ่งแบบเดิม

ABSTRACT

TE 163046

This thesis proposes a novel approach to enhance performance of item-based collaborative filtering (CF) algorithm. In the past, this algorithm has been designed to deal with scalability problems by first considering the relationships among different items and then use these relationships to perform similarity computations based on item's rating in a reduced space. However, this algorithm still generates sparsity and first-rater problems.

Therefore, this thesis concentrates on solving the two problems as mentioned. The method consists of three significant steps. The first step concentrates on using association rules to discover similarity relationships among attributes. The second step exploits these similarities during the calculation of item similar based on attributes. Finally, combines attributes similarity and rating similarity measures to find neighbor item in item-based CF algorithm and generating ratings predictions based on a combined similarity measure.

The experiments show that the proposed method can achieve better prediction accuracy than traditional item-based CF algorithm.