ชื่อเรื่องวิทยานิพนธ์ วิธีคลัสเตอริงแบบใหม่และการประยุกต์ใช้ในระบบแนะนำพาณิชย์อิเล็กทรอนิกส์

ผู้เขียน นายอุกฤษณ์ มารังค์

ปริญญา วิศวกรรมศาสตรดุษฎีบัณฑิต (วิศวกรรมไฟฟ้า)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์

รศ. ดร. นิพนธ์ ธีรอำพน	อาจารย์ที่ปรึกษาหลัก
รศ. ดร. เสริมศักดิ์ เอื้อตรงจิตต์	อาจารย์ที่ปรึกษาร่วม
ผศ. ดร. ศันสนีย์ เอื้อพันธ์วิริยะกุล	อาจารย์ที่ปรึกษาร่วม

บทคัดย่อ

วิทยานิพนธ์นี้นำเสนอวิธีการคลัสเตอริงแบบใหม่สำหรับระบบแนะนำพาณิชย์เล็กทรอนิกส์ภายใต้ปัญหาการไม่มีข้อมูลผู้ใช้ใหม่หรือสินค้าใหม่เพียงพอ และปัญหาการกระจายข้อมูลในกรณีสินค้าถูกจำหน่ายจำนวนน้อย ในกรณีที่สองปัญหานี้เกิดขึ้นพร้อมกันจะมีผลทำให้ระบบแนะนำมีประสิทธิภาพที่ต่ำ ในวิทยานิพนธ์นี้ได้เสนอวิธีการคลัสเตอริงแบบใหม่โดยใช้พื้นฐานเจเนติกอัลกอริทึม และเมเมติกอัลกอริทึม มีฟังก์ชันฟิตเนส (Fitness function) 2 ฟังก์ชันที่ใช้งานในเจเนติกอัลกอริทึม และเมเมติกอัลกอริทึม เรายังได้เสนอวิธีการรวมระบบแนะนำระหว่างวิธีการคลัสเตอริงที่มองเห็นได้โดยใช้พื้นฐานเจเนติกและเมเมติกอัลกอริทึมกับวิธีการผู้ใช้เป็นฐานระบบแนะนำระหว่างวิธีการคลัส
เตอริงที่มองเห็นโดยใช้พื้นฐานเจเนติกและเมเมติก อัลกอริทึมกับวิธีการสินค้าเป็นฐาน

ในการทดลอง เราได้ทดสอบประสิทธิภาพของการคลัสเตอริงกับข้อมูลสังเคราะห์จำนวนสามชุดและข้อมูลจริงจำนวนห้าชุด ข้อมูลชุดแรกเป็นข้อมูลการสั่งซื้อจาก Gazella.com ข้อมูลชุด

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ที่สองเป็นข้อมูลที่เก็บจากการสั่งซื้อสินค้าผ่านระบบออนไลน์ Thaiherbs-Thaimassage.com ข้อมูลชุดที่สามเป็นข้อมูลการให้คะแนนร้านอาหาร (Restaurant & Consumer) ข้อมูลสามชุดนี้เราแบ่งข้อมูลเพื่อเรียนรู้ และทดสอบโดยใช้ 10-fold cross validation ข้อมูลชุดที่สี่เป็นข้อมูลการเข้าร้านอาหาร Entree Chicago restaurant เราแบ่งข้อมูล 80% สำหรับการเรียนรู้ และ 20% สำหรับการทดสอบประสิทธิภาพระบบแนะนำ ข้อมูลชุดที่ห้าเป็นข้อมูลการให้คะแนนภาพยนต์ MovieLens เราแบ่งข้อมูล 80% สำหรับการเรียนรู้ และ 20% สำหรับการทดสอบประสิทธิภาพระบบแนะนำ เราใช้วิธี F-measure ในการวัดผล วิธีการที่นำเสนอถูกนำไปเปรียบเทียบกับวิธีการทั่วไปคือ วิธีการคลัสเตอริงลูกค้าและสินค้าโดยใช้วิธี *k*-means วิธีใช้ความถี่เป็นฐาน วิธีใช้ผู้ใช้เป็นฐาน และวิธีใช้สินค้าเป็นฐาน ผลการทดลองพบว่าวิธีการคลัสเตอริงสามารถแบ่งกลุ่มผู้ใช้และสินค้าได้อย่างมีประสิทธิภาพ ผลการทดลองยังแสดงให้เห็นอีกว่าระบบแนะนำโดยใช้พื้นฐานการคลัสเตอริงมีประสิทธิภาพมากกว่าวิธีการทั่วไปภายใต้ปัญหาการไม่มีข้อมูลผู้ใช้ใหม่หรือสินค้าใหม่เพียงพอ และปัญหาการกระจายข้อมูลในกรณีสินค้าถูกจำหน่ายจำนวนน้อย

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Thesis Title A Novel Clustering Method and Its Application to E-commerce Recommendation Systems

Author Mr. Ukrit Marung

Degree Doctor of Philosophy (Electrical Engineering)

Thesis Advisory Committee

Assoc. Prof. Dr. Nipon Theera-Umpon	Advisor
Assoc. Prof. Dr. Sermsak Uatrongjit	Co-advisor
Asst. Prof. Dr. Sansanee Auephanwiriyakul	Co-advisor

ABSTRACT

This thesis proposes a novel clustering method for E-commerce recommendation systems under the cold-start and sparsity problems. The cold-start and sparsity problems lead to low performance of the recommendation systems, especially when these problems occur together. In this thesis, to solve the cold-start and sparsity problems, we proposed the recommendation methods including the visual clustering recommendation based on the genetic algorithm (VCR-GAs) and the visual clustering recommendation based on the memetic algorithm (VCR-MAs). There are two fitness functions in each of the process of the genetic algorithm and the memetic algorithm. We also proposed the hybrid between the VCR-GAs and user-based methods, the hybrid between the VCR-GAs and item-based methods, the hybrid between the VCR-MAs and user-based methods, and the hybrid between the VCR-MAs and item-based methods.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

In the experiments, we tested the clustering performance on three synthetic and five real-world data sets. We also tested the performance of the recommendation systems on the five real-world data sets. The first data set is the transaction of purchasing from Gazelle.com, leg wear and leg care e-tailer, collected by Blue Martini Software on KDD (KDD-CUP2000). The second data set is our private data set from Thaiherbs-Thaimassage.com (TTS). The third data set is transaction of the rated restaurants collected by the Department of Computer Science, National Center for Research and Technological Development in Mexico. In these data sets, we used ten-fold cross validation to ensure the generalization of results. The fourth data set is the transaction of the entree Chicago restaurant (ECR) collected by the University of California on the UCI machine learning repository. The fifth data set is MovieLens from the GroupLens Research Project. We divided the ECR and MovieLens data sets into 80% and 20% for the training set and test set, respectively. The F-measure was chosen as the evaluation measure. Our proposed methods were compared with four common methods, i.e., the visual clustering method using *k*-means (VCR-KM), the frequency-based (FB), the user-based (UB), and the item-based methods (IB). The results show that the visual clustering was able to properly cluster users-items binary images. The results also show that the proposed recommendation systems were more efficient under the cold-start and sparsity problems than the common methods.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved