



ระบบแนะนำวิดีโอเกมแบบไฮบริด

Hybrid Filtering based Video Game Recommendation System

จิรวัดน์ ไจรัก, ชนาภัทร ปราชญาวงศ์ และ อ.ดร.วิรัช เจริญเรืองกิจ
หลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

Abstract

Recommendation system is the system that recommends products and services to customers. The product rating of all customers and other features are used as information to predict what customers may like. This research proposes machine learning models to create a hybrid of Collaborative Filtering (CF) techniques and Content-based Filtering techniques for video game recommendation system. The game rating dataset from amazon.com in 2014 consists of 168,291 users and 23,177 items. The first model combines CF Matrix Factorization and the Content-based Filtering technique. The second combines CF Item-base and CF User-base combining with Content-based Filtering techniques. The 25% of data are holded out for evaluation of the models. The criteria Root Mean Square Error (RMSE) are used for evaluation and achieve the best accuracy with RMSE at 0.7.

บทคัดย่อ

ระบบแนะนำ เป็นระบบที่จะแนะนำสินค้า หรือบริการให้กับลูกค้า โดยจะใช้คะแนนที่ลูกค้าเคยให้คะแนนไว้ และคุณสมบัติต่างๆของสินค้า เป็นข้อมูลในการทำนายสิ่งที่ถูกอาจชอบ การทดลองนี้ได้เสนอโมเดล Machine Learning รูปแบบไฮบริดของเทคนิค Collaborative Filtering (CF) และ Content-based Filtering สำหรับการสร้างระบบแนะนำวิดีโอเกม โดยใช้ข้อมูลคะแนนที่ได้มาจาก amazon.com โดยมีจำนวนข้อมูลในปี 2014 ทั้งหมด 168,291 ผู้ใช้ และ 23,177 ไอเทม โมเดลแรกเกิดจากการรวมกันของเทคนิค CF Matrix Factorization กับ Content-based Filtering และ โมเดลที่สองเกิดจากการรวมกันของเทคนิค CF Item-base และ CF User-base กับ Content-based Filtering โดยร้อยละ 25 ของข้อมูลได้ถูกเก็บไว้ใช้ในการประเมินผล ความแม่นยำของโมเดล เราได้ใช้ค่า Root Mean Square Error (RMSE) ในการประเมินผลความแม่นยำ และผลการประเมินของโมเดลที่ดีที่สุดด้วยค่า RMSE เท่ากับ 0.7

องค์ความรู้ที่เกี่ยวข้อง

Recommendation Systems [1]

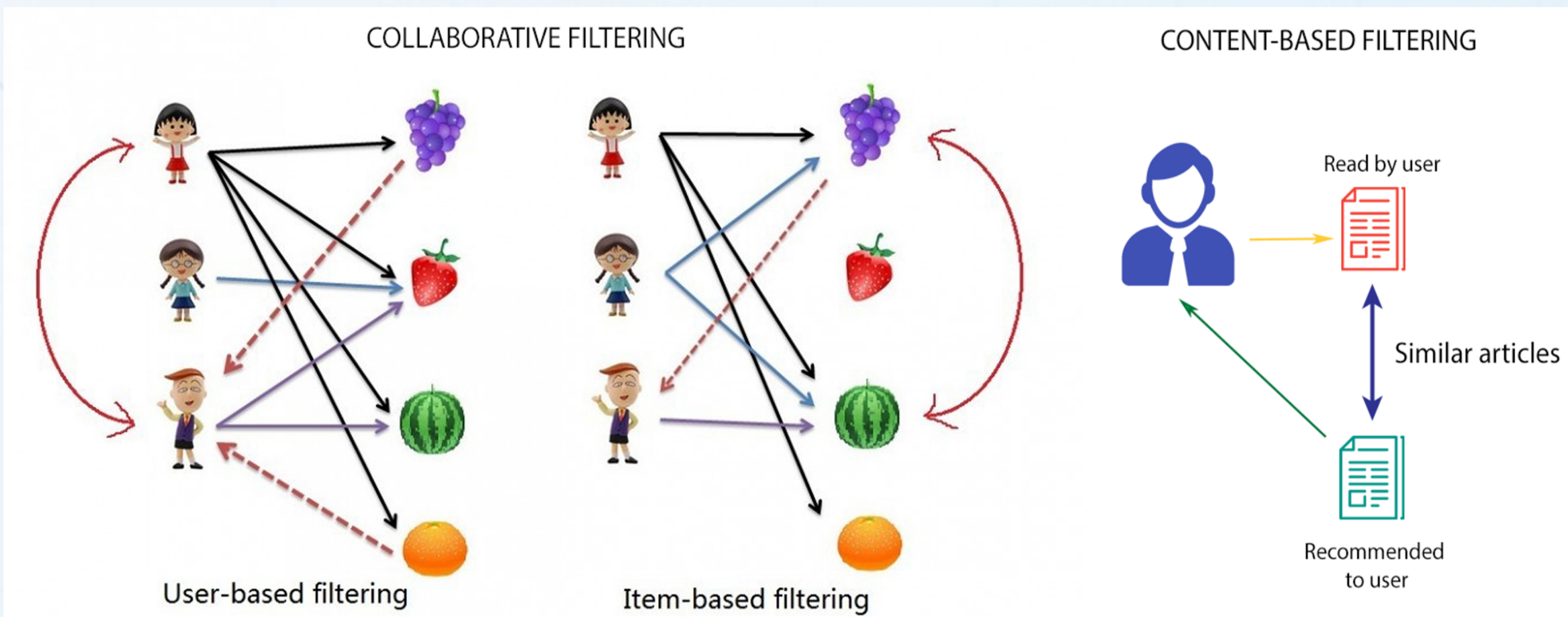
ระบบแนะนำเป็นโปรแกรมคอมพิวเตอร์ที่ทำหน้าที่ช่วยคัดกรองวัตถุที่ตรงกับความต้องการของผู้ใช้ จากวัตถุที่มีอยู่เป็นจำนวนมากให้กับผู้ใช้ โดยอ้างอิงจากสมมติฐานการเรียนรู้ข้อมูลความชอบหรือความต้องการ ณ ขณะนั้นของผู้ใช้ วัตถุในความหมายของระบบแนะนำหมายถึงสิ่งที่ต้องการจะแนะนำ เช่น หนังสือ ภาพยนตร์ บริการ หรือข้อมูลข่าวสาร เป็นต้น

Content-based Filtering [1]

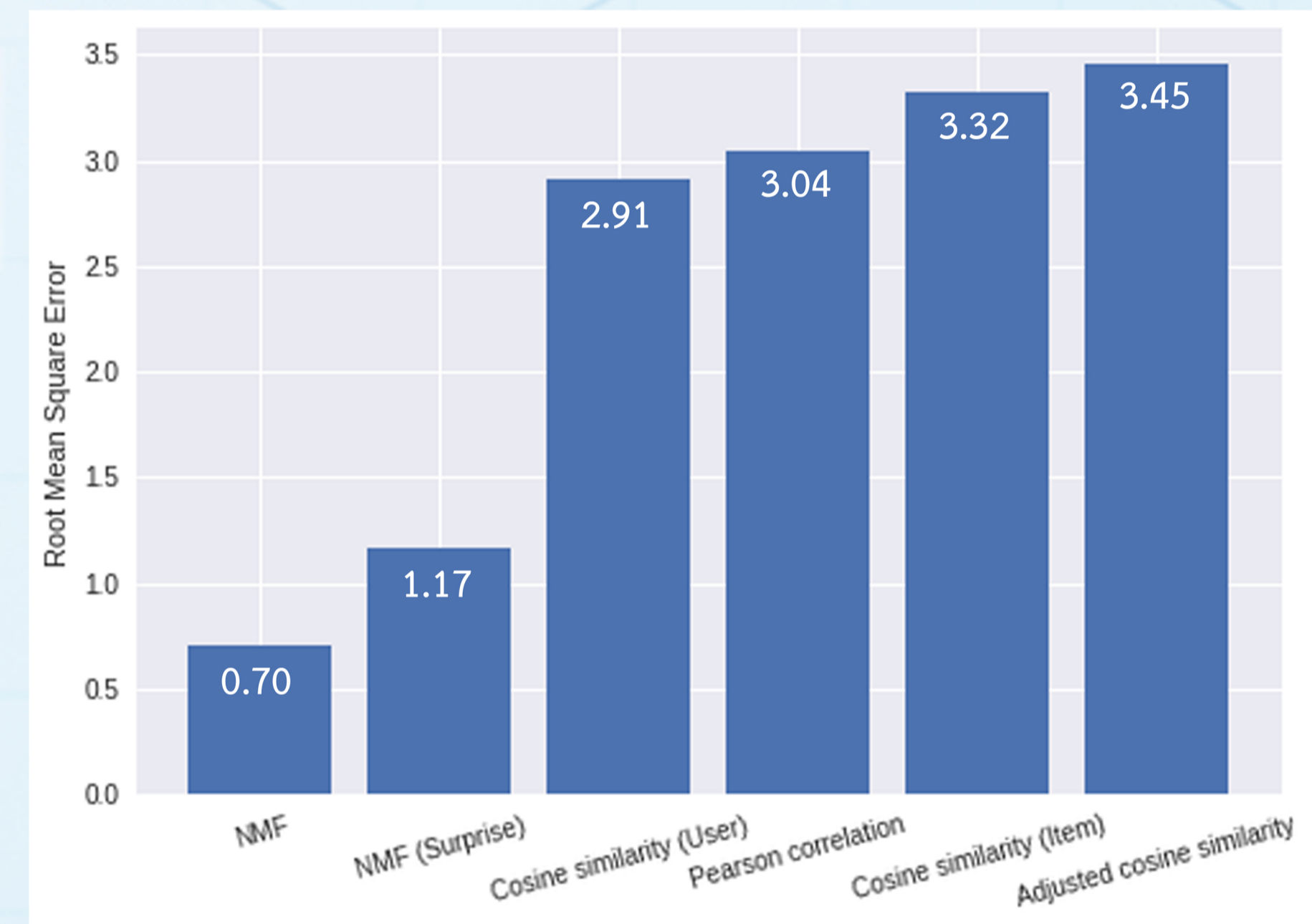
วิธีการที่มาจากคำอธิบายของวัตถุและโปรไฟล์ความชอบของผู้ใช้ โดยในระบบจะมี Keywords ที่ใช้ในการอธิบายวัตถุและโปรไฟล์ของผู้ใช้ที่ถูกสร้างขึ้นมาเพื่อระบุชนิดของ วัตถุที่ผู้ใช้ชอบหรือก็คือ อัลกอริทึมนี้จะพยายามที่จะแนะนำวัตถุที่ผู้ใช้ชอบมาก่อน (หรือกำลังสนใจอยู่ในปัจจุบัน) โดยจะมีวัตถุที่เป็น Candidate ต่างๆที่จะถูกนำมาเปรียบเทียบกับวัตถุที่ผู้ใช้เคยให้คะแนนมาก่อน แล้วจึงแนะนำวัตถุที่เหมาะสมที่สุดให้กับผู้ใช้

Collaborative Filtering [2]

ระบบแนะนำวัตถุโดยใช้ข้อมูลในระบบเป็นจำนวนมากที่มีความคล้ายคลึงหรือใกล้เคียงกันกับผู้ใช้เช่น ความชอบ ความเคลื่อนไหว กิจกรรมของผู้ใช้คนอื่นๆในระบบ ซึ่งจะเรียกเฉพาะเจาะจงได้ว่าระบบแนะนำแบบคัดกรองผู้ใช้ร่วม (User-based Filtering) แต่ถ้าใช้ข้อมูลลักษณะหรือคุณสมบัติของวัตถุอื่นๆในระบบก็จะเรียกอีกแบบหนึ่งว่า ระบบแนะนำแบบคัดกรองสิ่งของร่วม (Item-based Filtering) กล่าวคือใช้ข้อมูลของสิ่งอื่นๆในระบบมาช่วยในการคำนวณด้วย



ผลการดำเนินงาน

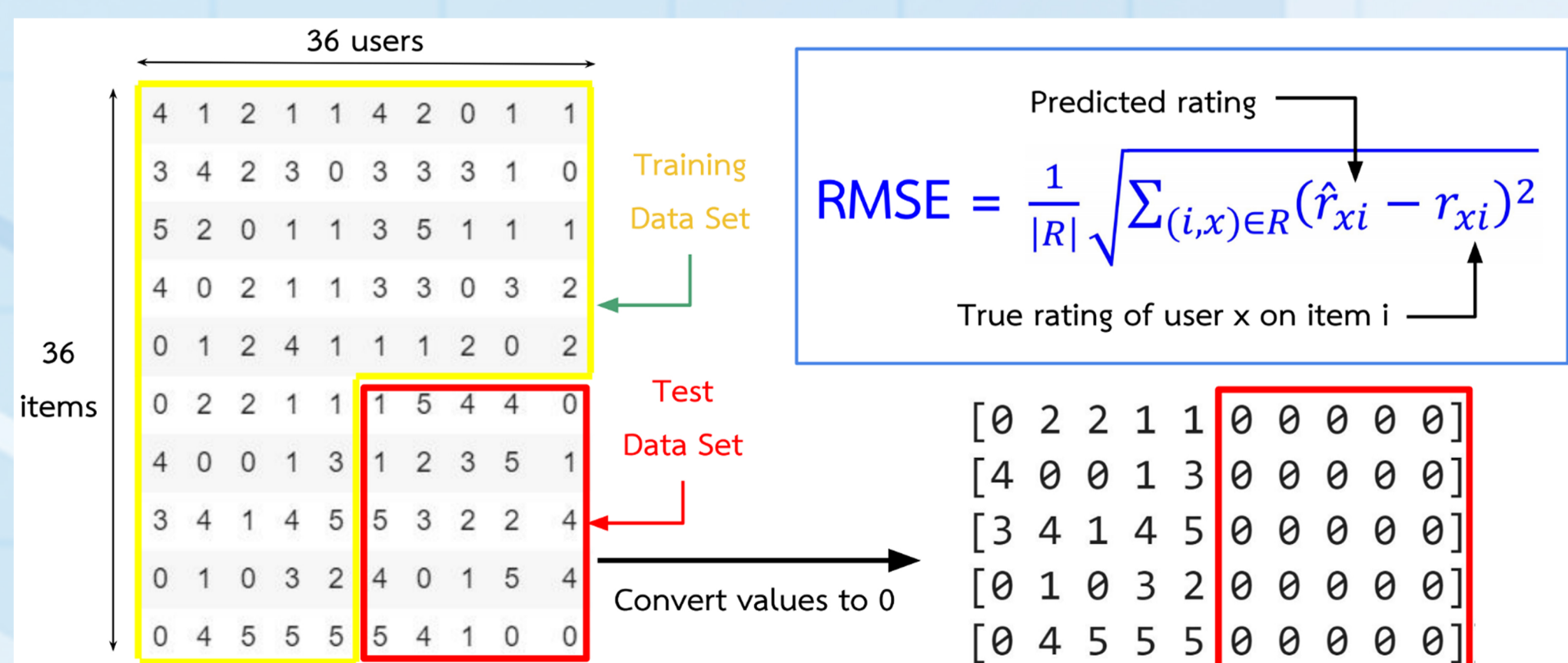


จากกราฟด้านบน คือ กราฟที่แสดงการเปรียบเทียบค่า Root Mean Square Error (RMSE) พบว่า

- โมเดลระบบแบบ Collaborative Filtering (CF) โดยใช้
 - Non-negative matrix factorization (NMF) ได้ค่า RMSE = 0.70
 - Surprise library (NMF) ได้ค่า RMSE = 1.17
- โมเดลระบบแบบ User-based CF โดยใช้
 - Cosine similarity ได้ค่า RMSE = 2.91
 - Pearson correlation similarity ได้ค่า RMSE = 3.04
- โมเดลระบบแบบ Item-based CF โดยใช้
 - Cosine similarity ได้ค่า RMSE = 3.32
 - Adjusted cosine similarity ได้ค่า RMSE = 3.45

การดำเนินงาน

- ★ เลือกข้อมูลผู้ใช้ 10 คนแรกที่ให้คะแนนเกมบอยที่ดีที่สุด ในปี 2014 ซึ่งจะได้เกม 37 เกมที่เกี่ยวข้อง
- ★ เลือกเกมมา 1 เกม แล้วใช้ Content-based Filtering ทำนายหาเกมที่คล้ายคลึงกัน 36 เกม
- ★ แปลงข้อมูลเป็นเมทริกซ์ แล้วแบ่งเมทริกซ์เป็นออก 4 คิวเตอร์ (Train data set 3 คิวเตอร์ และ Test data set 1 คิวเตอร์) และแปลงค่าใน Test data set ให้เท่ากับ 0
- ★ ใช้ Collaborative Filtering ทำนายคะแนน จากเมทริกซ์ที่มีข้อมูลผู้ใช้ 36 คน และเกม 36 เกม
- ★ ทดสอบความถูกต้องการทำนายคะแนน โดยใช้ค่า Root Mean Square error เปรียบเทียบคะแนนที่ทำนายได้กับคะแนนดั้งเดิม



สรุปผล

จากการทดสอบประสิทธิภาพในแต่ละโมเดล พบว่าโมเดลระบบแบบ Collaborative Filtering (CF) โดยใช้ Non-negative matrix factorization (NMF) ได้ค่า Root Mean Square Error เท่ากับ 0.70 ซึ่งเป็นผลลัพธ์ที่ดีที่สุด หมายความว่าความวิธีแบบ NMF สามารถทำนายคะแนนได้ใกล้เคียงกับคะแนนดั้งเดิมมากที่สุด

Actual Rating	[5.0, 4.0, 4.0, 5.0, 5.0, 5.0, 5.0, 4.0, 5.0, 5.0, 5.0, 5.0, 5.0, 5.0, 5.0]
Predicted Rating	[3.2, 4.7, 4.8, 5.0, 4.9, 5.1, 5.1, 4.6, 4.4, 5.1, 4.8, 4.8, 3.7, 4.0, 4.0, 4.5, 4.8]

ตารางแสดงคะแนนดั้งเดิมและคะแนนที่ทำนายได้ โดยใช้ Non-negative matrix factorization

บรรณานุกรม

- [1] Badaro, Gilbert & Hajj, Hazem & El-Hajj, Wassim & Nachman, Lama. (2556). A Hybrid Approach with Collaborative Filtering for Recommender Systems. 2556 9th International Wireless Communications and Mobile Computing Conference, IWCMC 2556.
- [2] John S. Breese; David Heckerman & Carl Kadie (2541). Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence (UAI'98).