



ระบบแนะนำส่วนบุคคลโดยใช้การจัดกลุ่มแบบลำดับชั้นร่วมกับวิธีการคัดกรองผู้ใช้และการคัดกรองสิ่งของร่วม

Personal Recommender System based on Agglomerative Clustering together with User-based and Item-based Collaborative Filtering Methods

รัฐวรรณ พันธุนิต และ นิเวศ จิระวิชิตชัย*

Ratawan Phantunin and Nivet Chirawichitchai*

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยศรีปทุม 2410/2 ถนน พหลโยธิน เขตจตุจักร กรุงเทพมหานคร 10900

Faculty of Information Technology, Sripatum University, 241/2 Phaholyothin Road, Jatjak, Bangkok 10900, Thailand

*Corresponding author e-mail: aor.tv5@gmail.com, nivet.ch@spu.ac.th

ARTICLE INFO

ABSTRACT

Article history:

Received: 30 May, 2020

Revised: 5 June, 2020

Accepted: 22 June, 2020

Available online: 25 June, 2020

DOI: 10.14456/rj-rmutt.2020.10

Keywords: recommender

system, agglomerative

clustering, user-based

collaborative filtering,

item-based collaborative

filtering

The objective of this study is to develop and increase efficiency of Personal Integrated Recommender System. The Recommender System plays an important role and is crucial to our everyday lives in online shopping and online services. We will find that the thing that comes with when shopping for products or using services is to recommend products or services. A good Recommender System helps generate more sales. In the meantime, various problems could be found with the system, e.g. scalable data, data sparsity, data accuracy, and having a lot of new users. Therefore, new techniques have been introduced and integrated with the recommender system in order to solve the problems and improve for greater recommender system efficiency. In this study, an Agglomerative Clustering together with a User-base and Item-base Collaborative Filtering Method is proposed. By combining the strengths of each method, we can improve the recommender system efficiency and accuracy. This combination helps to solve the problems of scalable data, data sparsity, and having a lot of new users. The results show that it reduces the

processing time and increases precision. Therefore, we can conclude that the Personal Recommender System developed based on Agglomerative Clustering together with User-based and Item-based Collaborative Filtering Method has the ability to increase system efficiency and is applicable. It also helped to solve the problems of scalable data, data sparsity, and having a lot of new users. When modern technology arrives in the future, we may be able to use cloud computing for data analysis in order to expand the capacity to process the information efficiently.

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาและเพิ่มประสิทธิภาพของแบบจำลองระบบแนะนำส่วนบุคคลด้วยเทคนิคการจัดกลุ่มแบบลำดับชั้นร่วมกับวิธีการคัดกรองผู้ใช้และการคัดกรองสิ่งของร่วม เนื่องมาจากในปัจจุบันระบบแนะนำได้มีบทบาทในชีวิตประจำวันในการเพิ่มยอดขายสินค้าและบริการออนไลน์เป็นอย่างมาก แต่ในขณะเดียวกันกลับพบปัญหาต่าง ๆ ที่เกิดขึ้นกับระบบแนะนำข้อมูล เช่น ปัญหาของขนาดข้อมูล (Scalability Problem) ความเบาบางของข้อมูล (Sparsity Problem) ความแม่นยำของข้อมูล และการมีผู้ใช้รายใหม่เพิ่มเข้ามาเป็นจำนวนมาก (First-rater Problem) ดังนั้นจึงได้มีการคิดค้นเทคนิคใหม่ ๆ มาผสมผสานกับระบบแนะนำเพื่อเป็นการแก้ปัญหาและปรับปรุงประสิทธิภาพของระบบแนะนำให้ดีและตรงกับความต้องการของผู้ซื้อยิ่งขึ้น โดยในงานวิจัยนี้ได้นำเสนอแบบจำลองวิธีการจัดกลุ่มแบบลำดับชั้น (Agglomerative Clustering) ทำงานร่วมกับวิธีการคัดกรองผู้ใช้และวิธีแบบคัดกรองสิ่งของร่วม (User-based and Item-based Collaborative Filtering Method) โดยเอาจุดเด่นของแต่ละวิธีมาผสมผสานกันเพื่อปรับปรุงประสิทธิภาพของระบบแนะนำให้มีความแม่นยำมากยิ่งขึ้น จากผลการทดลองพบว่าแบบจำลองที่พัฒนาขึ้นพิจารณาจากค่าพื้นที่ใต้เส้นกราฟ ค่าความแม่นยำ ค่า Normalized discounted cumulative gain และค่าเฉลี่ยของความแม่นยำ ให้ค่าที่สูงกว่าการใช้ User-

based Collaborative Filtering หรือ Item-based Collaborative Filtering เพียงอย่างเดียว ดังนั้นจึงสรุปได้ว่าแบบจำลองระบบแนะนำส่วนบุคคลที่พัฒนาขึ้นโดยใช้การจัดกลุ่มแบบลำดับชั้นร่วมกับวิธีการคัดกรองผู้ใช้และการคัดกรองสิ่งของร่วม มีประสิทธิภาพดีกว่าการใช้ User-based Collaborative Filtering หรือ Item-based Collaborative Filtering เพียงอย่างเดียว และยังสามารถนำไปประยุกต์ใช้งานได้จริง โดยในอนาคตเมื่อมีเทคโนโลยีที่ทันสมัยมากขึ้นอาจจะสามารถใช้คลาวด์คอมพิวเตอร์ในการวิเคราะห์ข้อมูล เพื่อขยายขีดความสามารถในการประมวลผลข้อมูลได้อย่างมีประสิทธิภาพ

คำสำคัญ: ระบบแนะนำ วิธีการจัดกลุ่มแบบลำดับชั้น วิธีแบบคัดกรองผู้ใช้ร่วม วิธีแบบคัดกรองสิ่งของร่วม

บทนำ

เนื่องจากในปัจจุบันโลกกำลังก้าวเข้าสู่ยุคของข้อมูลขนาดใหญ่ (Big Data) ซึ่งเราจะเห็นได้ว่า ข้อมูลต่าง ๆ ที่เข้ามาในระบบฐานข้อมูลมีจำนวนเพิ่มมากขึ้นแบบก้าวกระโดด จึงทำให้เกิดปัญหาขนาดของข้อมูล (Scalability Problem) ที่มีจำนวนมากและปัญหาการให้ Rating ต่อชิ้นข้อมูล (Sparsity Problem) ที่ผู้ใช้ให้ไว้ไม่เพียงพอต่อการคำนวณ ทำให้ผู้วิจัยมีความคิดที่จะพัฒนาระบบแนะนำข้อมูลส่วนบุคคล (Recommender System) โดยใช้การจัดกลุ่มแบบลำดับชั้นร่วมกับขั้นตอนวิธีแบบคัดกรองผู้ใช้ร่วมและการคัดกรองสิ่งของขึ้นเพื่อแก้ไขปัญหาดังกล่าวข้างต้น

ระบบแนะนำข้อมูล (Recommender System) (1) เป็นการประยุกต์ใช้เทคนิคการค้นหาคำแนะนำจากข้อมูลเพื่อนำความรู้เหล่านั้นไปประกอบการตัดสินใจให้กับผู้ใช้งาน เพื่อเพิ่มโอกาสให้กับผู้ซื้อหรือผู้ใช้งานที่ได้รับสินค้าหรือข้อมูลที่ตรงตามความต้องการ บริษัทที่ประสบความสำเร็จ เช่น Amazon Netflix และ Spotify ทุกบริษัทต่างใช้ ระบบที่เรียกว่า Recommender Systems โดยใช้เทคนิค Collaborative Filtering ซึ่งเป็นเทคนิคที่ประสบความสำเร็จและได้รับความนิยมอย่างสูงสำหรับการพัฒนาระบบแนะนำ (2) เช่น ระบบแนะนำภาพยนตร์ เนื่องจากผู้ใช้งานไม่สามารถศึกษารายละเอียดของภาพยนตร์ทุก ๆ เรื่องที่มีอยู่ในฐานข้อมูลได้ภายในระยะเวลาที่จำกัด ทั้งเป็นการสร้างประสบการณ์ที่ดีให้กับผู้ใช้ เพิ่มความพึงพอใจ รวมไปถึงเพิ่มยอดขายได้อย่างเห็นผลจริง

นอกจากนี้แล้ว จากการศึกษาระบบแนะนำข้อมูล พบว่าหากผู้ใช้มีจำนวนมากขึ้นจะทำให้เกิดปัญหาของอัลกอริทึม เช่น ขนาดของข้อมูลซึ่งมีวิธีแก้ปัญหาโดยการใช้การจัดกลุ่ม (Clustering) ในการแบ่งกลุ่มผู้ใช้งานก่อนเข้าสู่การทำงานของระบบเพื่อให้ได้คำแนะนำกับผู้ใช้ได้เร็วขึ้น และปัญหาต่อมาคือปัญหาการให้ค่าอันดับคะแนน (Rating) ต่อชิ้นข้อมูล เพราะเมื่อข้อมูลมีจำนวนมากขึ้นจะทำให้ผู้ใช้ไม่สามารถให้คะแนนได้อย่างทั่วถึง ทำให้มีค่าอันดับคะแนน (Rating) ไม่เพียงพอต่อการคำนวณ ทำให้มีผลกระทบต่อความพึงพอใจของผู้ใช้ ที่มีต่อระบบ (3) จากการศึกษาพบว่า มีงานวิจัยหลายชิ้นที่นำเสนอแนวทางในการแก้ปัญหาดังกล่าวโดยใช้เทคนิคทางการทำเหมืองข้อมูล (Data mining) เช่น เทคนิค Clustering เทคนิค Classification เป็นต้น

จากความเป็นมาข้างต้น ผู้วิจัยจึงพัฒนาแบบจำลองระบบแนะนำส่วนบุคคลโดยใช้โดยใช้การจัดกลุ่มแบบลำดับขั้น (Agglomerative Clustering) ร่วมกับวิธีการคัดกรองผู้ใช้กับการคัดกรองสิ่งของ (User-based and Item-based Collaborative Filtering Method) โดยนำจุดเด่นของแต่ละวิธีมาผสมผสานกันและนำมา

ทดลองใช้ในการแนะนำภาพยนตร์ต่อผู้ใช้งานระบบ เพื่อให้ผู้ใช้งานได้รับการแนะนำที่มีประสิทธิภาพตรงตามต้องการ ถูกต้อง และพึงพอใจที่สุด

วิธีดำเนินการวิจัย

ศึกษาและวิเคราะห์ปัญหา โดยแบบจำลองที่นำเสนอในงานวิจัยนี้มุ่งเน้นการพัฒนาแนะนำแบบผสมผสาน (Hybrid) โดยการใช้การจัดกลุ่มแบบลำดับขั้น (Agglomerative Clustering) ร่วมกับวิธีการคัดกรองผู้ใช้และการคัดกรองสิ่งของ โดยวิธีการคัดกรองผู้ใช้ (User-based Collaborative Filtering) มีจุดเด่นคือจะพิจารณาความเหมือนของผู้ใช้งานในระบบกับความชอบที่เคยมีมาก่อนและ วิธีการคัดกรองสิ่งของ (Item-based Collaborative Filtering) มีจุดเด่นคือจะพิจารณาความสัมพันธ์ระหว่างสิ่งของได้ดีกว่า User-based โดยปกติระบบแนะนำประเภทนี้จะมีขั้นตอนในการประมวลผล 3 ขั้นตอนใหญ่ ๆ ดังนี้ 1. สร้างประวัติผู้ใช้หรือข้อมูลที่ใช้เป็นพื้นฐานของระบบ 2. คัดเลือกผู้ใช้ที่เคยเลือกข้อมูลนั้น (co-rated item) หรือข้อมูลที่เคยถูกเลือกจากผู้ใช้งานดังกล่าว (co-user rate) ที่มีความใกล้เคียงหรือคล้ายคลึงกันขึ้นมาตามจำนวนที่กำหนดไว้ โดยการเปรียบเทียบประวัติผู้ใช้ หรือข้อมูลที่ใช้เป็นพื้นฐานของระบบ ซึ่งอาจใช้ค่าความเหมือนโคไซน์ (Cosine Similarity) (8) เพื่อคำนวณว่าผู้ใช้และข้อมูลนั้นมีความเหมาะสมกันมากน้อยเพียงใด

ข้อมูลในการทดลองงานวิจัยครั้งนี้ เป็นชุดข้อมูลพื้นฐานสำหรับใช้สร้างฐานข้อมูลของระบบและใช้ในการจัดกลุ่มผู้ใช้ โดยนำข้อมูลมาจาก MovieLens Project 1M Dataset มาใช้ ชุดข้อมูลประกอบด้วย

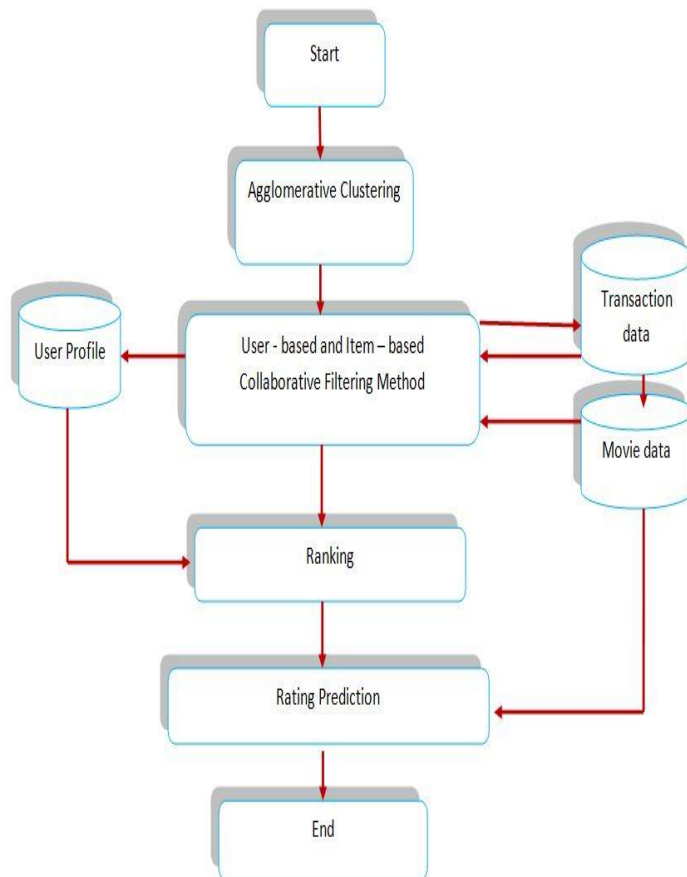
1. ชุดข้อมูลการให้คะแนนการจัดอันดับภาพยนตร์ (Rating) จำนวน 1,000,209 Records จากการเก็บรวบรวมข้อมูลผู้ใช้งาน จำนวน 6,040 คน และภาพยนตร์จำนวน 3,883 เรื่อง

2. ชุดข้อมูลผู้ใช้ ระบบ (User Profile) ประกอบด้วย เพศ อายุ อาชีพ (21 อาชีพ) รหัสไปรษณีย์

3. ประเภทของภาพยนตร์ (Item Feature) ประกอบไปด้วย ชื่อเรื่อง วันที่ฉาย วันที่เป็นวิดีโอ รหัสภาพยนตร์ และประเภทภาพยนตร์(19 ประเภท) เช่น แอ็คชั่น ผจญภัย ตลก ดราม่า สงคราม เป็นต้น

ซึ่งข้อมูลที่ใช้ในการทดสอบเป็นกลุ่มตัวอย่างที่เกิดจากการเก็บข้อมูลจริง โดยผู้ใช้งานได้เข้าไปให้ข้อมูลใน MovieLens Project เกี่ยวกับภาพยนตร์ที่เคยชมแล้ว โดยชุดข้อมูลนี้ประกอบด้วยไฟล์ข้อความซึ่งใช้การเข้ารหัสอักขรแบบ ANSI จำนวน 3 ไฟล์ได้แก่ Ratings.dat, Movies.dat และ Users.dat จากนั้นจึงทำการออกแบบและศึกษาการทำงานของอัลกอริทึม Collaborative Filtering ลำดับขั้นตอนการทำงาน การประยุกต์ใช้วิธีการการทำเหมืองข้อมูลและวิธีทางสถิติ มาช่วยในการหาความคล้ายคลึงกันของผู้ใช้แต่ละคน โดยใช้ขั้นตอนวิธี

แบบคัดกรองผู้ใช้ร่วมกับแบบคัดกรองสิ่งของร่วมกัน นอกจากนี้ยังได้ศึกษาปัญหาที่เกิดจาก Collaborative Filtering แบบดั้งเดิมว่ามีปัญหาใหญ่อยู่ที่เรื่องของขนาดของข้อมูลขนาดใหญ่ (Big Data) และ การท่วมท้นของข้อมูล (Information overload) ทำให้ประสิทธิภาพและความแม่นยำในการทำงานลดลง ผู้พัฒนาจึงได้ศึกษาอัลกอริทึม User-based and Item-based Collaborative Filtering เพิ่มเติมเพื่อมาช่วยแก้ปัญหานี้ ร่วมกับอัลกอริทึม Agglomerative Clustering แบบผสมผสานซึ่งจัดเป็นเทคนิคใหม่ในการนำเสนอระบบแนะนำข้อมูลส่วนบุคคลที่มีประสิทธิภาพสูงและมีความแม่นยำมากขึ้น จากการศึกษาปัญหาและรวบรวมข้อมูล ผู้พัฒนาได้ออกแบบขั้นตอนวิธีในการทำวิจัยดังรูปที่ 1



รูปที่ 1 ขั้นตอนวิธีระบบแนะนำส่วนบุคคลโดยใช้การจัดกลุ่มแบบลำดับขั้นร่วมกับวิธีการคัดกรองผู้ใช้กับการคัดกรองสิ่งของร่วมกัน

ผลการศึกษาและอภิปรายผล

งานวิจัยการพัฒนาระบบแนะนำส่วนบุคคลโดยใช้การจัดกลุ่มแบบลำดับขั้นร่วมกับวิธีการคัดกรองผู้ใช้กับการคัดกรองสิ่งของนี้ พิจารณาประสิทธิภาพของแบบจำลองโดยวัดจากค่า AUC (Area Under the Curve), prec (Precision), ค่า NDCG (Normalized Discounted Cumulative Gain), ค่า MAP (Mean Average Precision) ซึ่งจะอธิบายดังต่อไปนี้ (4-6)

1. ค่าพื้นที่ใต้เส้นกราฟ (Area under the Curve: AUC) เป็นการวัด ประสิทธิภาพการทำงานของแบบจำลองโดยที่เส้น ในแกนนอนจะเป็นอัตราบวกจริง (True Positive Rate) ส่วนในแกนตั้งจะเป็นอัตราบวกเท็จ (False positive rate) ค่าการวัดโดยใช้ AUC นี้จะเริ่มที่ 0 ถึง 1 โดยที่ 0 หมายถึง แบบจำลองนั้นมีประสิทธิภาพต่ำ ส่วน 1 หมายถึง แบบจำลอง นั้นมีประสิทธิภาพสูงสุด ซึ่งผลการวัดจะง่ายต่อความเข้าใจ

2. ค่าความแม่นยำ (Precision: prec) (6) กล่าวว่าเป็นการวัดความแม่นยำของระบบในการค้นหาเอกสารที่เกี่ยวข้องได้ถูกต้อง ได้ผลตามตารางที่ 1

ตารางที่ 1 ผลการสืบค้น

Doc	Action	
	Retrieved	Not Retrieved
Relevant	tp	fn
Not Relevant	fp	tn

$$\text{ค่า Precision} = \frac{tp}{(tp+fp)}$$

3. ค่า Normalized Discounted Cumulative Gain: NDCG เป็นทฤษฎีที่ใช้ในการประเมินผล Search Engine มีการให้ระดับคะแนนความเกี่ยวข้องของเอกสารรวมทั้งพิจารณาลำดับของผลการค้นหา (Ranking) และ DCG เป็นการวัดความเหมาะสมของเอกสารโดยสนใจตำแหน่งหรือลำดับของเอกสาร ค่าระดับคะแนนที่ได้รับจะสะสมจากลำดับบนของรายการผลลัพธ์การค้นหาไปยัง

ลำดับล่างของผลลัพธ์การค้นหา โดยค่าคะแนนจะลดลงเมื่อผลความพึงพอใจอยู่ในลำดับที่ต่ำ (7, 8) โดยสมการ NDCG เป็นดังนี้

$$NDCG_q = M_q \sum_{j=1}^k \frac{(2^{r(j)} - 1)}{\log(1+j)} \quad (1)$$

เมื่อ k คือ ระดับหรือเกณฑ์ที่ใช้

r(j) คือ ค่าลำดับความเกี่ยวข้องของเอกสารที่ได้จากคะแนนประเมินโดยผู้ใช้

M_q คือ ค่าคงที่ที่เกิดจากความสมบูรณ์ในการจัดลำดับโดยมีค่ามากที่สุด คือ 1

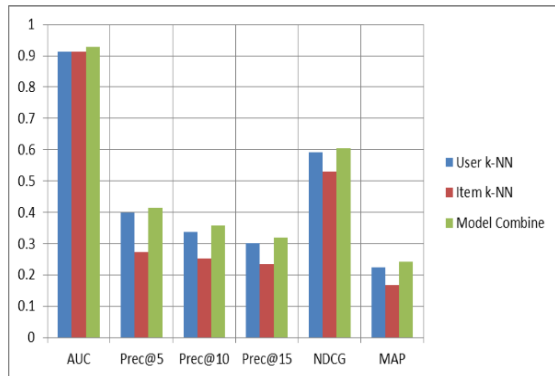
ทั้งนี้ NDCG จะให้รางวัลกับเอกสารที่เกี่ยวข้องที่ปรากฏในลำดับของการจัดอันดับผลการค้นหาและลงโทษเอกสารที่ไม่เกี่ยวข้องโดยการลดคะแนน NDCG

4. ค่าเฉลี่ยของความแม่นยำ (Mean Average Precision: MAP) เป็นค่าเฉลี่ยของความแม่นยำ ของคำค้นหาหลาย ๆ คำที่เกี่ยวข้องกัน (Relevance) และนำไปจัดอันดับโดยมีสูตรดังนี้

$$MAP = \frac{\sum_{q=1}^Q AveP(q)}{Q} \quad (2)$$

ตารางที่ 2 ผลการทดลองในกรณีใช้วิธีขั้นตอนวิธีการจัดกลุ่มแบบลำดับขั้นร่วมกับวิธีการคัดกรองผู้ใช้กับแบบคัดกรองสิ่งของ จำแนกแบบตาราง

Metrics	User	Item	Model
	k-NN	k-NN	Combiner
AUC	0.913	0.912	0.928
Prec@5	0.398	0.272	0.415
Prec@10	0.338	0.252	0.358
Prec@15	0.301	0.234	0.320
NDCG	0.591	0.530	0.604
MAP	0.223	0.167	0.241



รูปที่ 2 ผลการทดลองการจัดกลุ่มแบบลำดับชั้นร่วมกับวิธีการคัดกรองผู้ใช้กับการคัดกรองสิ่งของร่วม

จากการทดลองพบว่า วิธีผสมผสานระหว่างเทคนิคโดยใช้การจัดกลุ่มแบบลำดับชั้นร่วมกับวิธีการคัดกรองผู้ใช้กับการคัดกรองสิ่งของร่วม (Model Combiner) พบว่าได้ค่าพื้นที่ใต้เส้นกราฟ (Area Under the Curve: AUC) มีค่าเท่ากับ 0.928 ค่าความแม่นยำ (Precision:prec) ที่ k@5 มีค่าเท่ากับ 0.415 ที่ k@10 มีค่าเท่ากับ 0.358 ที่ k@15 มีค่าเท่ากับ 0.320 ค่า Normalized discounted cumulative gain: NDCG มีค่าเท่ากับ 0.604 ค่าเฉลี่ยของความแม่นยำ (Mean Average Precision : MAP) มีค่าเท่ากับ 0.241 ซึ่งทุกค่าที่ได้มาสูงกว่าค่าที่มาจากวิธีการคัดกรองผู้ใช้ร่วม (User-based Collaborative Filtering) และวิธีการคัดกรองสิ่งของร่วม (Item-based Collaborative Filtering) เพียงอย่างเดียวอย่างชัดเจน

สรุปผล

งานวิจัยนี้ได้พัฒนาระบบแนะนำโดยใช้เทคนิคการผสมผสานกันโดยใช้การจัดกลุ่มแบบลำดับชั้น (Agglomerative Clustering) ร่วมกับวิธีการคัดกรองผู้ใช้และการคัดกรองสิ่งของร่วม (User-based and Item-based Collaborative Filtering) เป็นการเอาจุดเด่นของแต่ละวิธีมาผสมผสานกัน ส่งผลให้ระบบแนะนำข้อมูลที่น่าเสนอมีประสิทธิภาพสูงขึ้นในทุกตัวชี้วัด

ดังนั้นจึงสรุปได้ว่าการจัดกลุ่มแบบลำดับชั้น (Agglomerative Clustering) ร่วมกับวิธีการคัดกรองผู้ใช้และการคัดกรองสิ่งของร่วม (User-based and Item-based Collaborative Filtering) สามารถสนับสนุนการตัดสินใจของผู้ใช้ระบบได้แม่นยำ สูงขึ้นกว่าระบบแนะนำส่วนบุคคลที่ใช้เทคนิคคัดกรองผู้ใช้ร่วมหรือคัดกรองสิ่งของร่วมเพียงอย่างเดียว และช่วยแก้ปัญหาขนาดของข้อมูล (Scalability Problem) ที่มีจำนวนมาก ปัญหาชั้นข้อมูลที่ไม่มีการให้ Rating ไว้ (First-rater Problem) ปัญหาการแยกแยะ Rating (Transparency Problem) และปัญหาการให้ Rating ต่อชั้นข้อมูล (Sparsity Problem) ซึ่งสามารถนำไปประยุกต์ใช้ได้ อย่างมีประสิทธิภาพอีกด้วย

เอกสารอ้างอิง

1. Adomavicius G, Tuzhilin A. Recommendation Technologies: Survey of Current Methods and Possible Extensions: Stern School of Business, New York University; 2004.IS-03-06.
2. Sarwar BM, Karypis G, Konstan JA, Riedl JT. Analysis of Recommendation Algorithms for E-commerce: ACM. In: Proceedings of the 2nd ACM; 2000 October; Minneapolis Minnesota USA, New York: Association for Computing Machinery; 2000. p. 158-67.
3. Shyu ML, Haruechaiyasak C, Chen SC, Zhao N. Collaborative filtering by mining association rules from user access sequences. In: Proceedings - International Workshop on Challenges in Web Information Retrieval and Integration, WIRI'05; 2005 April 8-9; Tokyo, Japan. 2005. p. 128-33.

4. Aljumily R. Agglomerative Hierarchical Clustering: An Introduction to Essentials (1) Proximity Coefficients and Creation of a Vector-Distance Matrix and (2) Construction of the Hierarchical Tree and a Selection of Methods. *Global Journal of Human Social Science Research*. 2016;16(3-G):23-50.
5. Janpla S, Wanapiron P. System framework for an intelligent question bank and examination system. *International Journal of Machine Learning and Computing*. 2018; 8(5):488-94.
6. Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett*. 2006;27(8):861–74.
7. Smoot BJ, Wong JF, Dodd MJ. Comparison of Diagnostic Accuracy of Clinical Measures of Breast Cancer–Related Lymphedema: Area Under the Curve. *Arch Phys Med Rehab*. 2011;92(4):603-10.
8. Chirawichitchai N. Developing term weighting scheme based on term occurrence ratio for sentiment analysis. In: Kim, Kuinam J, editor. *Information Science and Applications*. 1st ed. Springer eBook: Springer Nature; 2015. p. 737–44.