การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวโดยใช้ความสัมพันธ์เชิงสังคมออนไลน์

นายสิรวิชญ์ สพโชค

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the Graduate School.

ONE-CLASS COLLABORATIVE FILTERING USING ONLINE SOCIAL RELATIONSHIPS

Mr. Sirawit Sopchoke

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2011

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวโดยใช้ความสัมพันธ์เชิงสังคมออนไลน์
โดย	นายสิรวิชญ์ สพโชค
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

…….. คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

…… ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ นครทิพย์ พร้อมพูล)

…….. อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

…… กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สินธุภิญโญ)

…… กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.ชลวิช นัทธี)

สิรวิชญ์ สพโชค : การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวโดยใช้ความสัมพันธ์เชิงสังคมออนไลน์. (ONE-CLASS COLLABORATIVE FILTERING USING ONLINE SOCIAL RELATIONSHIPS) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ศ.ดร.บุญเสริม กิจศิริกุล, 56 หน้า.

ระบบแนะนำ (Recommender Systems) ที่มีในปัจจุบันส่วนใหญ่ ใช้เทคนิคการกรองแบบร่วมมือกัน (Collaborative Filtering) เพื่อทำนายความชอบของผู้ใช้เป้าหมายและให้คำแนะนำสำหรับช่วยเหลือกระบวนการตัดสินใจของผู้ใช้ เทคนิคการกรองแบบร่วมมือกันจะเรียนรู้ลักษณะความชอบของผู้ใช้ ด้วยข้อมูลพฤติกรรมต่อระบบของผู้ใช้ทั้งหมด อย่างไรก็ตาม เทคนิคการกรองแบบร่วมมือกันแบบพื้นฐานไม่สามารถให้คำแนะนำที่ถูกต้องได้เช่นเดิมในสถานการณ์ของระบบที่ขาดแคลนข้อมูลป้อนกลับที่ชัดแจ้งหรือค่าคะแนนความชอบจากผู้ใช้ ปัญหาดังกล่าวส่งผลให้เกิดความสนใจในเทคนิคที่เรียกว่า การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering) ซึ่งการประมวลผลของระบบแนะนำจะใช้การพิจารณาเฉพาะข้อมูลที่เป็นตัวอย่างบวกหรือข้อมูลป้อนกลับโดยปริยายเท่านั้น นอกจากนี้ การปรากฏตัวและเติบโตอย่างรวดเร็วของบริการเครือข่ายสังคมออนไลน์ ซึ่งมุ่งเน้นการสร้างแล้วสะท้อนปฏิสัมพันธ์เชิงสังคมระว่างผู้ใช้ด้วยกัน อีกทั้งยังอุดมไปด้วยข้อมูลบ่งบอกตัวตน ความสนใจ และกิจกรรมของผู้ใช้ทุกคน ทำให้มีโอกาสที่ใช้ในการปรับปรุงคุณภาพของการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวได้ ดังนั้นการวิจัยนี้ นำเสนอวิธีการใหม่สำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวโดยใช้ความสัมพันธ์เชิงสังคมออนไลน์เพื่อเพิ่มความแม่นในการทำนายของระบบแนะนำ ข้อมูลความสัมพันธ์เชิงสังคมสามารถสะท้อนให้เห็นการจูงใจเชิงสังคม ซึ่งผู้ใช้มีแนวโน้มที่จะมีความเชื่อถือโดยปริยายต่อรสนิยมความชอบของเพื่อนบนเครือข่ายสังคม สิ่งนี้ตรงกับวัตถุประสงค์เทคนิคการกรองแบบร่วมมือกัน ในการวิจัยนี้ประยุกต์ใช้การจูงใจเชิงสังคมกับเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก เพื่อแก้ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว จากการทดลองเพื่อทดสอบประสิทธิผลด้วยตัววัดเชิงสนับสนุนการตัดสินใจ 2 ตัววัด พบว่าวิธีการที่นำเสนอให้ผลลัพธ์ที่พิสูจน์ว่าได้การประมวลผลของระบบแนะนำที่มีคุณภาพสูงกว่าเมื่อเทียบกับวิธีการพื้นฐานแบบอื่น

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต.....

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.....

ปีการศึกษา 2554

5270542921 : MAJOR COMPUTER SCIENCE
KEYWORDS : RECOMMENDER SYSTEMS / COLLABORATIVE FILTERING / MATRIX FACTORIZATION / SOCIAL RELATIONSHIPS / SOCIAL BOOKMARKING

SIRAWIT SOPCHOKE : ONE-CLASS COLLABORATIVE FILTERING USING ONLINE SOCIAL RELATIONSHIPS. ADVISOR : PROF. BOONSERM KIJSIRIKUL, Ph.D., 56 pp.

Most existing recommender systems use a traditional method called Collaborative Filtering (CF) to predict the preferences of target users in order to provide recommendations to help them with their decision making process. CF learns user preferences based on user community's past behaviors. However, Collaborative Filtering may provide an invalid recommendation if a system does not have explicit feedback or item rating from users. As a result, One-class Collaborative Filtering (OCCF) came into existence since it took only positive examples or implicit feedback into consideration to provide recommendations. With the rapid growing trend of the online social networking services which generate and reflect social relationships among their users as well as consist of representation, interests and activities of each user, there is an opportunity to improve OCCF. Therefore, this research aims to propose a novel method for OCCF using online social relationships to increase a prediction accuracy of the recommendation. It is believed that social relationship data can reflect the social influence as people tend to have a default trust in their friends' tastes in an online social network. This is also what the CF algorithms try to reveal. In this research, Non-negative Matrix Factorization method was applied with social influence weighting scheme to the one-class problem. Based on the experimental evaluation and two decision-support measures, the proposed method was proved to provide higher quality of recommendation results than the other baseline methods.

Department : Computer Engineering Student's Signature .....
Field of Study : Computer Science Advisor's Signature .....
Academic Year : 2011

กิตติกรรมประกาศ

ในความสำเร็จลุล่วงของวิทยานิพนธ์นี้ ข้าพเจ้าขอกราบขอบพระคุณท่าน ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล ผู้เป็นอาจารย์ที่ปรึกษา ผู้ให้ความรู้ ผู้คอยแนะนำชี้แนะแนวทางอันเป็นประโยชน์ และผู้ช่วยเหลือดูแลตลอดมา

ขอขอบคุณภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ขอขอบคุณสมาชิกห้องปฏิบัติการอัจฉริยะภาพเครื่องจักรและการค้นพบความรู้ (MIND LAB) ทุกคนที่ได้ให้ความรู้และคำแนะนำซึ่งมีส่วนช่วยในการทำวิจัยเป็นอย่างมาก

ขอกราบขอบพระคุณอาจารย์ทุกท่าน ผู้เคยประสิทธิ์ประสาทวิชาความรู้แก่ข้าพเจ้า

ข้าพเจ้าขอกราบขอบพระคุณบิดามารดาและคุณยายผู้มีพระคุณ ผู้ให้การสนับสนุนและเป็นกำลังใจแก่ข้าพเจ้าเสมอมา

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง
บทคัดย่อภาษาอังกฤษ..... จ
กิตติกรรมประกาศ..... ฉ
สารบัญ..... ช
สารบัญตาราง..... ฌ
สารบัญภาพ..... ญ
บทที่ 1 บทนำ..... 1
ความเป็นมาและความสำคัญของปัญหา..... 1
วัตถุประสงค์ของการวิจัย..... 3
ขอบเขตของการวิจัย..... 3
ขั้นตอนการดำเนินงาน..... 4
ประโยชน์ที่คาดว่าจะได้รับ..... 4
ผลงานตีพิมพ์จากวิทยานิพนธ์..... 5
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง..... 6
ระบบแนะนำ..... 6
การกรองแบบร่วมมือกัน (Collaborative Filtering)..... 10
ปัญหาของระบบแนะนำในปัจจุบัน..... 13
การจำแนกประเภทแบบคลาสเดี่ยว (One-class Classification)..... 15
การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering)..... 16
บทที่ 3 งานวิจัยที่เกี่ยวข้อง..... 19
กรองแบบร่วมมือกันสำหรับคลาสเดี่ยว..... 19
การแยกตัวประกอบเมทริกซ์สำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (Matrix Factorization for One-class Collaborative Filtering)..... 22
การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization)..... 24
แนวทางการแก้ปัญหาด้วยการพิจารณาค่าน้ำหนัก (Weighting based Approaches)..... 26

หน้า

ความสัมพันธ์เชิงสังคมออนไลน์..... 29
บทที่ 4 การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวโดยใช้ความสัมพันธ์เชิงสังคมออนไลน์ 32
แรงจูงใจ..... 32
แนวคิด..... 36
การใช้ความสัมพันธ์เชิงสังคมออนไลน์..... 38
บทที่ 5 ผลการทดลอง..... 41
ชุดข้อมูล..... 41
การทดลองและผลการทดลอง..... 44
บทที่ 6 สรุปผลการวิจัยและข้อเสนอแนะ..... 51
สรุปผลการวิจัย..... 51
ข้อเสนอแนะ..... 52
รายการอ้างอิง..... 53
ประวัติผู้เขียนวิทยานิพนธ์..... 56

สารบัญตาราง

ตารางที่		หน้า
1	ส่วนหนึ่งของเมทริกซ์คะแนนความชอบสำหรับระบบแนะนำภาพยนตร์.....	8
2	ข้อแตกต่างระหว่างงานวิจัยที่เกี่ยวข้องกับแนวคิดทางการวิจัยที่นำเสนอ.....	21
3	การกำหนดเมทริกซ์ค่าน้ำหนัก W ในงานวิจัยของ Pan และคณะ.....	27
4	สมการที่ใช้ในขั้นตอนวิธีการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ เปรียบเทียบแบบที่ไม่มีการพิจารณาค่าน้ำหนักกับแบบที่มีการพิจารณาค่าน้ำหนัก.....	29
5	การระบุค่าน้ำหนักโดยใช้ความสัมพันธ์เชิงสังคมออนไลน์.....	39
6	ผลลัพธ์ค่าประสิทธิผลของระบบแนะนำที่ใช้เทคนิค K-Nearest-Neighbor.....	45
7	ผลลัพธ์ค่าประสิทธิผลของระบบแนะนำที่ใช้เทคนิคการแยกตัวประกอบเมทริกซ์.....	45
8	ผลลัพธ์ค่าประสิทธิผลซึ่งเปรียบเทียบเทคนิคที่นำเสนอกับเทคนิคพื้นฐานแบบอื่น.....	49

สารบัญภาพ

ภาพที่		หน้า
1	แผนภาพขั้นตอนการดำเนินงาน.....	4
2	การแนะนำแบบอาศัยเนื้อหา.....	9
3	การกรองแบบร่วมมือกันประเภท Memory-based.....	12
4	รูปแบบการจำแนกคลาสด้วยเครื่อง.....	16
5	การแยกตัวประกอบเมทริกซ์แบบพื้นฐาน.....	23
6	ขั้นตอนวิธี Alternating Least Squares (ALS).....	24
7	ขั้นตอนวิธี MUR สำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ.....	25
8	ขั้นตอนวิธี ALS สำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ.....	26
9	การพิจารณากลุ่มผู้ใช้อื่นที่ผู้ใช้เป้าหมายรู้จักและเชื่อถือในงานวิจัยของ Walter และคณะ.....	30
10	ปริมาณการระบุค่าคะแนนความชอบแต่ละค่าของชุดข้อมูล Netflix.....	33
11	ความสัมพันธ์เชิงสังคมออนไลน์และการจูงใจเชิงสังคม (social influence)....	35
12	ขั้นตอนวิธีสำหรับการให้คำแนะนำของระบบแนะนำ โดยใช้การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก.....	40
13	ตัวอย่างข้อมูลการกดปุ่ม Like ต่อ Facebook Pageบางส่วนภายในชุดข้อมูล	42
14	ความสัมพันธ์เชิงสังคมหรือความเป็นเพื่อนของชุดข้อมูล Facebook ในการทดลอง.....	43
15	การแบ่งพับของชุดข้อมูลในการทดลอง.....	44
16	จำนวนปัจจัยแฝงส่งผลที่แตกต่างกันต่อค่า Mean Average Precision.....	48

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

บ่อยครั้งการพิจารณาเลือกสิ่งต่าง ๆ ในชีวิตประจำวัน ตัวอย่างเช่น ข้อมูลข่าวสาร เว็บเพจ หรือสื่อบันเทิงต่าง ๆ ซึ่งมีความหลากหลายและมีปริมาณมหาศาลในปัจจุบัน อาจทำให้ไม่ทราบว่าจะเริ่มต้นค้นหาและตัดสินใจเลือกอย่างไร ซึ่งโดยทั่วไป เราอาจเริ่มต้นค้นหาสิ่งที่ต้องการโดยอ้างอิงตามประสบการณ์ หรือความชอบส่วนตัวเกี่ยวกับสิ่งที่เคยสัมผัสรับรู้มาก่อน ตัวอย่างเช่น การเลือกชมภาพยนตร์โดยยึดติดกับประเภทของเนื้อหาภาพยนตร์หรือนักแสดงนำที่ชอบ หรือหากในกรณีที่ปราศจากความรู้หรือประสบการณ์ที่จะใช้ในการพิจารณา การพิจารณาเลือกอาจกระทำโดยอาศัยจากคำแนะนำของผู้อื่น โฆษณา หรือคำวิจารณ์ที่กล่าวถึงในสื่อต่าง ๆ ซึ่งจากการพิจารณาสิ่งเหล่านี้ทำให้ทราบถึงข้อมูลของสิ่งที่ต้องการได้ระดับหนึ่ง อย่างไรก็ตาม ในความเป็นจริงแล้ว การพิจารณาเลือกแต่ละครั้งนั้น อาจต้องมีการพิจารณาข้อมูลสำหรับใช้ประกอบการเลือกจากแหล่งต่าง ๆ มากมาย ทำให้ต้องใช้เวลาอย่างมากในการวิเคราะห์ข้อมูลที่ได้มาเพื่อตัดสินใจเลือก และเนื่องด้วยความหลากหลายและปริมาณที่เพิ่มขึ้นตลอดเวลา จึงทำให้ไม่สามารถรับทราบถึงข้อมูลของสิ่งต่าง ๆ ที่มีเหล่านั้นได้อย่างครบถ้วน การเลือกจึงถูกจำกัดอยู่ในกรอบที่สามารถเห็น รับรู้ และหาข้อมูลได้เท่านั้น ส่งผลให้อาจสูญเสียโอกาสในการเข้าถึงตัวเลือกที่ดีกว่าหรือค้นพบตัวเลือกอื่น ๆ เพิ่มเติม

ด้วยความก้าวหน้าด้านเทคโนโลยีอินเทอร์เน็ต รวมถึงปริมาณและพฤติกรรมการใช้งานของผู้ใช้ในปัจจุบัน ก่อให้เกิดการเพิ่มขึ้นและการเปลี่ยนแปลงของข้อมูลสารสนเทศผ่านทางเว็บไซต์รูปแบบต่าง ๆ ในปริมาณมหาศาลอย่างต่อเนื่อง ผู้ให้บริการจึงให้ความสนใจเป็นอย่างยิ่งถึงเรื่องความสามารถของระบบที่จะช่วยแนะนำตัวเลือกเข้าสู่กระบวนการตัดสินใจของผู้ใช้แต่ละคน เพื่อให้ผู้ใช้เข้าถึงสิ่งที่ผู้ใช้ต้องการได้อย่างถูกต้องและสะดวกรวดเร็ว ด้วยเหตุนี้เว็บไซต์หลายแห่ง ซึ่งให้บริการเกี่ยวกับข้อมูลสารสนเทศหรือจำหน่ายสินค้าที่มีผู้เข้าใช้งานเป็นจำนวนมาก เช่น Amazon.com หรือ YouTube จึงมีการพัฒนาและใช้งานเครื่องมือหรือซอฟต์แวร์ที่เรียกว่า ระบบแนะนำ (Recommender Systems) ซึ่งเป็นระบบมีวัตถุประสงค์เพื่อใช้คอมพิวเตอร์ทำงานแทนมนุษย์ในการแนะนำสิ่งต่าง ๆ ให้แก่ผู้ใช้แต่ละคน โดยพิจารณาจากข้อมูลความชอบและพฤติกรรมที่ผ่านมาของผู้ใช้เหล่านั้น การทำงานระบบแนะนำส่วนใหญ่ในปัจจุบัน ประมวลผลด้วย

การรวบรวมข้อมูลความชอบของผู้ใช้ทั้งหมด จากนั้นค้นหาผู้ใช้อื่น ๆ ในระบบที่มีลักษณะความชอบเช่นเดียวกันหรือใกล้เคียงกัน กับลักษณะความชอบของผู้ใช้เป้าหมายที่ต้องการคำแนะนำ เพื่อที่จะนำข้อมูลความชอบของผู้ใช้อื่นที่มีลักษณะใกล้เคียงกันนั้น มาใช้ในการแนะนำ วิธีการดังกล่าวนี้ ถูกเรียกว่า การกรองแบบร่วมมือกัน (Collaborative Filtering) [1]

อย่างไรก็ตาม วิธีการของการกรองแบบร่วมมือกันส่วนใหญ่ในปัจจุบัน ใช้การค้นหากลุ่มผู้ใช้ที่มีความชอบใกล้เคียงกันกับผู้ใช้ที่ต้องการคำแนะนำ จากการพิจารณาค่าคะแนนความชอบ หรืออัตราเชิงตัวเลข (numerical rating) ที่ผู้ใช้กำหนดให้แก่สิ่งต่าง ๆ ซึ่งบ่งบอกถึงระดับความรู้สึกชอบที่มีต่อสิ่งที่ระบุค่าเหล่านั้นได้อย่างชัดเจน จึงเป็นประเด็นที่ต้องพิจารณาเมื่อต้องการนำวิธีการของการกรองแบบร่วมมือกันในปัจจุบัน มาใช้กับระบบที่ไม่มีการเก็บข้อมูลค่าคะแนนความชอบหรืออัตราเชิงตัวเลขดังกล่าว นอกจากนี้ หากพิจารณาถึงพฤติกรรมการใช้งานเว็บไซต์ของผู้ใช้โดยทั่วไปตามความเป็นจริงแล้ว พบว่าหากผู้ใช้รู้สึกไม่ชอบหรือไม่สนใจสิ่งใด ก็อาจละเลยการให้ค่าคะแนนความชอบต่ำ เพื่อระบุถึงความไม่พอใจที่มีต่อสิ่งนั้น ยังผลให้เกิดความไม่สมดุลกันของตัวอย่างข้อมูลที่แสดงความชอบกับตัวอย่างข้อมูลที่แสดงความไม่ชอบของผู้ใช้ ซึ่งระบบต้องใช้ข้อมูลเหล่านี้ ในการเรียนรู้ลักษณะความชอบเพื่อค้นหาคำแนะนำให้แก่ผู้ใช้แต่ละคน ดังนั้น หลายสถานการณ์ของระบบแนะนำที่ใช้งานจริงในปัจจุบัน ระบบแนะนำจำเป็นต้องเรียนรู้ลักษณะความชอบของผู้ใช้ ด้วยข้อมูลความชอบของผู้ใช้ที่เป็นตัวอย่างบวก (positive example) หรือด้วยข้อมูลการใช้งานตอบสนองระบบของผู้ใช้ ตัวอย่างเช่น การติดตามการสั่งซื้อสินค้าในเว็บไซต์ การใช้งานเว็บไซต์ให้บริการประเภทการคั่นหน้าเชิงสังคม (Social Bookmarking) เพื่อคั่นหน้าเว็บเพจที่สนใจ การติดตามประวัติการคลิกเชื่อมโยง หรือการ “Like” สิ่งที่สนใจบน Facebook เป็นต้น ซึ่งการเรียนรู้ของเครื่องด้วยตัวอย่างบวกเพียงด้านเดียวนี้ ถูกระบุให้เป็นประเด็นปัญหาของระบบแนะนำที่เรียกว่า การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering) [2]

งานวิจัยในการแก้ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวที่ผ่านมา ได้นำเสนอรูปแบบในการแก้ไขปัญหา โดยใช้เทคนิคการแยกตัวประกอบเมทริกซ์ (Matrix Factorization) กับขั้นตอนวิธี Weighted Low Rank Approximation (wLRA) [3] ในงานวิจัยของ Pan และคณะ [2] หรือใช้เทคนิคแบบการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization) ในงานวิจัยของ Sindhwani และคณะ [4] เพื่อแก้ปัญหาการเรียนรู้ของเครื่องด้วยตัวอย่างบวกเพียงด้านเดียวของระบบแนะนำรูปแบบการกรองแบบร่วมมือกัน เพื่อแนะนำผู้ใช้แต่ละคน แต่ทว่างานวิจัยดังกล่าว ไม่มีการคำนึงข้อมูลคุณลักษณะของผู้ใช้หรือข้อมูล

คุณลักษณะของสิ่งต่าง ๆ ที่มีไว้ให้ผู้ใช้พิจารณา ทำให้อาจสามารถกำหนดวิธีการที่เหมาะสมกว่าสำหรับการระบุค่าน้ำหนักของข้อมูลที่ไม่ทราบค่าหรือตัวอย่างที่ไม่ระบุป้ายกำกับคลาส (unlabeled example) เพื่อการประมวลผลข้อมูลด้วยเทคนิคการแยกตัวประกอบเมทริกซ์กับขั้นตอนวิธี wLRA จึงทำให้ผู้ดำเนินการวิจัยเล็งเห็นแนวทางในการนำมาวิจัยพัฒนาต่อ เพื่อให้ระบบสามารถแนะนำสิ่งที่เหมาะสมแก่ผู้ใช้ระบบยิ่งขึ้น นอกจากนี้ สถานการณ์ของระบบแนะนำที่ต้องเรียนรู้ลักษณะความชอบของผู้ใช้ด้วยข้อมูลที่เป็นตัวอย่างบวก มักปรากฏให้เห็นได้อย่างชัดเจนบนเว็บไซต์ประเภทที่มีให้บริการเครือข่ายสังคม (Social Networking Service) ซึ่งผู้ใช้สามารถสื่อสาร ได้รับข้อมูล และแบ่งปันข้อมูล ในปริมาณมหาศาลได้อย่างสะดวกและมีประสิทธิภาพ จากการวิเคราะห์พิจารณาความสัมพันธ์ระหว่างกันของผู้ใช้เว็บไซต์เครือข่ายสังคมออนไลน์ดังกล่าว และนำมาใช้ในการเรียนรู้ลักษณะความชอบของผู้ใช้ระบบ อาจส่งผลระบบสามารถแนะนำผู้ใช้แต่ละคนได้อย่างเหมาะสมยิ่งขึ้น

ดังนั้น เพื่อให้มั่นใจว่าการทำงานของระบบ จะสามารถแนะนำสิ่งที่เหมาะสมยิ่งขึ้นแก่ผู้ใช้ระบบ ผู้ดำเนินการวิจัยจึงต้องการนำเสนอวิธีการใหม่สำหรับการกรองแบบร่วมมือกัน ซึ่งเหมาะสมกว่าสำหรับระบบแนะนำที่เรียนรู้ลักษณะความชอบของผู้ใช้ ด้วยการพิจารณาเฉพาะข้อมูลความชอบของผู้ใช้ที่เป็นตัวอย่างบวกในการเรียนรู้ของเครื่อง

วัตถุประสงค์ของการวิจัย

เพื่อนำเสนอขั้นตอนวิธีใหม่สำหรับระบบแนะนำแบบการกรองแบบร่วมมือกัน ซึ่งระบบแนะนำสามารถเรียนรู้ลักษณะความชอบของผู้ใช้ โดยใช้เฉพาะข้อมูลความชอบของผู้ใช้ที่เป็นตัวอย่างบวก และข้อมูลความสัมพันธ์ระหว่างกันของผู้ใช้บนเครือข่ายสังคม ช่วยให้ระบบสามารถแนะนำตัวเลือกแก่ผู้ใช้แต่ละคนได้อย่างเหมาะสมยิ่งขึ้น

ขอบเขตของการวิจัย

1. ข้อมูลที่จะนำมาใช้ในการทดสอบ เป็นข้อมูลซึ่งบ่งบอกความชอบของผู้ใช้จริง โดยเก็บรวบรวมจากเว็บไซต์เครือข่ายสังคมที่มีข้อมูลความสนใจต่าง ๆ ของผู้ใช้ ทั้งนี้การเก็บรวบรวมข้อมูล เป็นการเก็บข้อมูลแบบนิรนาม ไม่บันทึกชื่อนามสกุลจริงของผู้ใช้

2. ระบบแนะนำเรียนรู้ลักษณะความชอบของผู้ใช้ ด้วยข้อมูลที่เป็นตัวอย่างบวกเท่านั้น

ขั้นตอนการดำเนินงาน

1. ศึกษางานวิจัยที่เกี่ยวข้องกับระบบแนะนำที่ผ่านมา

2. วิเคราะห์รูปแบบสถานการณ์และวิธีการทำงานของระบบแนะนำในปัจจุบัน

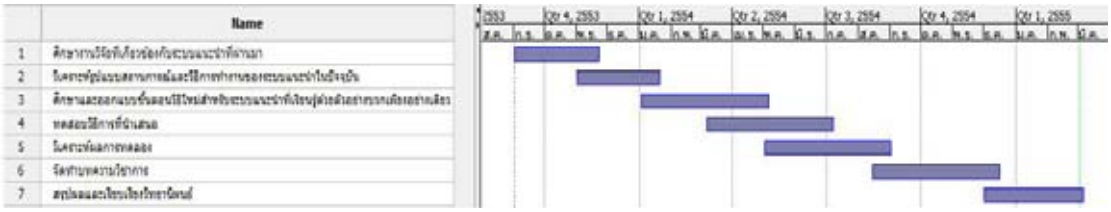
3. ศึกษาและออกแบบขั้นตอนวิธีใหม่ ซึ่งเหมาะสมสำหรับระบบแนะนำที่เรียนรู้ด้วยตัวอย่างบวกเพียงอย่างเดียว

4. ทดสอบวิธีการที่นำเสนอ

5. วิเคราะห์ผลการทดลอง

6. จัดทำบทความวิชาการ

7. สรุปผลและเรียบเรียงวิทยานิพนธ์



ภาพที่ 1 แผนภาพขั้นตอนการดำเนินงาน

ประโยชน์ที่คาดว่าจะได้รับ

สามารถพัฒนาระบบแนะนำ ซึ่งสามารถแนะนำตัวเลือกที่เหมาะสมแก่ผู้ใช้ระบบแต่ละคนยิ่งขึ้น โดยอาศัยวิธีใหม่สำหรับระบบแนะนำแบบการกรองแบบร่วมมือกันที่นำเสนอ และเหมาะสมกว่าในการเรียนรู้ลักษณะความชอบของผู้ใช้ด้วยข้อมูลความชอบของผู้ใช้ที่เป็นตัวอย่างบวก ดังพบได้ในสถานการณ์ของระบบแนะนำในปัจจุบัน ส่งผลให้เป็นประโยชน์ต่อผู้พัฒนาระบบ

ที่ต้องการตอบสนองความต้องการของผู้ใช้ระบบหรือลูกค้าได้อย่างเต็มที่ เพื่อช่วยเพิ่มโอกาสทางธุรกิจ โดยการแนะนำตัวเลือกในการเข้าถึงข้อมูล ข่าวสาร สินค้าและบริการต่าง ๆ ให้ตรงตามความชอบของผู้ใช้แต่ละคน

ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้รับการตีพิมพ์เป็นบทความทางวิชาการจำนวน 1 บทความ ดังนี้

- บทความในงานประชุมวิชาการระดับนานาชาติ

Sopchoke, S., and Kijsirikul, B. A Step Towards High Quality One-class Collaborative Filtering using Online Social Relationships. In Proc. 2011 International Conference on Advanced Computer Science and Information System, pp. 243-248. 2011.

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

การวิจัยนี้นำเสนอวิธีการใหม่สำหรับการกรองแบบร่วมมือกัน (Collaborative Filtering) ซึ่งเป็นรูปแบบวิธีการหนึ่งสำหรับระบบแนะนำ (Recommender Systems) เนื้อหาในบทนี้จะกล่าวถึงแนวคิดและทฤษฎีที่เกี่ยวข้อง อันได้แก่ ความหมายของระบบแนะนำ วิธีการของระบบแนะนำที่ถูกใช้ในปัจจุบัน ระบุข้อดีข้อเสียของแต่ละวิธีการที่ถูกใช้กับระบบแนะนำในปัจจุบัน พร้อมทั้งระบุสถานการณ์ที่เป็นประเด็นปัญหาและเกิดขึ้นจริงในการทำงานของระบบแนะนำ ซึ่งนำไปสู่เทคนิคการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering) โดยมีรายละเอียดเนื้อหาทั้งหมดดังต่อไปนี้

Adomavicius และ Tuzhilin [1] ได้รวบรวมนำเสนอเกี่ยวกับวิธีการที่ใช้กับระบบแนะนำทั้งหมดที่มีในปัจจุบัน รวมถึงแนวทางซึ่งเป็นความพยายามที่จะพัฒนาขีดความสามารถของระบบแนะนำ จากบทความดังกล่าว สามารถกล่าวสรุปเนื้อหาได้ดังนี้

ระบบแนะนำ

ระบบแนะนำ คือ ระบบที่จะช่วยแนะนำสิ่งต่าง ๆ แก่ผู้ใช้งานระบบแต่ละคน โดยการเสนอหรือแนะนำบางสิ่งจากทั้งหมดที่มีให้เลือกในระบบ เข้าสู่กระบวนการตัดสินใจของผู้ใช้ โดยคาดว่าผู้ใช้จะรู้สึกชอบหรือสนใจสิ่งที่แนะนำให้ ระบบแนะนำเป็นหัวข้องานวิจัยที่ได้รับความสนใจ เริ่มตั้งแต่การปรากฏตัวของงานวิจัยหัวข้อการกรองแบบร่วมมือกัน ในช่วงกลางทศวรรษ 1990 จากนั้นมาก็มีงานวิจัยอื่น ๆ ที่เกี่ยวข้องเกิดขึ้นอย่างต่อเนื่อง เพื่อช่วยส่งเสริมการพัฒนาทางด้านธุรกิจ และเพื่อศึกษาพัฒนาหาวิธีการใหม่ ๆ สำหรับระบบแนะนำ เนื่องจากยังคงมีปัญหาสถานการณ์ต่าง ๆ มากมาย กระตุ้นให้เกิดการวิจัยพัฒนาต่อ นอกจากนี้ ระบบแนะนำยังช่วยจัดการข้อมูลที่มีปริมาณมาก ทำให้ประยุกต์ใช้ได้จริง ในการให้คำแนะนำเชิงบุคคล เชิงข้อมูล หรือสินค้าและบริการต่าง ๆ ตัวอย่างของงานประยุกต์ต่าง ๆ ของระบบแนะนำซึ่งเป็นที่รู้จักแพร่หลายในปัจจุบัน ได้แก่ ระบบแนะนำสินค้าของ Amazon.com [5] ระบบแนะนำภาพยนตร์ของ MovieLens [6] และ Netflix [7] หรือระบบแนะนำเพลงของ Last.fm

ประเด็นปัญหาการทำงานของระบบแนะนำ คือ การทำนาย (predict) ค่าคะแนนความชอบ หรืออัตราเชิงตัวเลข (numerical rating) ของตัวเลือกที่ผู้ใช้ไม่เคยพบเจอมาก่อน ซึ่งการทำนายค่าคะแนนความชอบนั้น พิจารณาจากการให้คะแนนความชอบแก่ตัวเลือกอื่น ๆ ของผู้ใช้งานระบบ ประกอบกับข้อมูลคุณลักษณะอื่น ๆ ที่เกี่ยวข้อง ซึ่งเมื่อระบบสามารถทำนายค่าคะแนนความชอบได้แล้ว ก็สามารถใช้พิจารณาแนะนำตัวเลือกที่ผู้ใช้ไม่เคยพบเจอมาก่อนและระบบทำนายได้ค่าคะแนนความชอบที่สูง ให้แก่ผู้ใช้แต่ละคนได้

หากต้องการอธิบายในรูปสมการ สามารถอธิบายได้ดังนี้ สมมติให้ C แทนเซตของผู้ใช้งานระบบแนะนำทั้งหมด ให้ S แทนเซตของตัวเลือกที่เป็นไปได้ซึ่งสามารถนำมาใช้แนะนำ เช่น ภาพยนตร์ หนังสือ หรือเพลง โดยปริภูมิของ S ที่สามารถเป็นไปได้ในบางระบบนั้น อาจมีค่ามหาศาล ตัวอย่างเช่น การแนะนำภาพยนตร์ หรือ หนังสือ ซึ่งมีตัวเลือกที่เป็นไปได้เป็นจำนวนในหลักล้านตัวเลือก เช่นเดียวกันปริภูมิของ C หรือผู้ใช้ในบางระบบก็อาจมีค่ามากเช่นกัน กำหนดให้ u เป็นฟังก์ชันอรรถประโยชน์แสดงความมีประโยชน์ของตัวเลือก s ต่อผู้ใช้ c หรือ $u : C \times S \rightarrow R$ โดยที่ R คือ ค่าระดับความมีประโยชน์ต่อผู้ใช้ที่อยู่ในช่วงที่เป็นไปได้ทั้งหมด สำหรับแต่ละผู้ใช้ $c \in C$ ระบบต้องเลือกตัวเลือก $s' \in S$ ที่มีประโยชน์ต่อผู้ใช้มากที่สุด จะได้ดังสมการที่ 2.1

$$\forall c \in C,\ s'_c = \arg\max_{s \in S} u(c, s) \qquad (2.1)$$

ในระบบแนะนำโดยทั่วไปตัวเลือกที่มีประโยชน์นั้น ถูกแสดงให้เห็นด้วยการให้ค่าคะแนนความชอบ ซึ่งบ่งชี้ถึงความชอบของผู้ใช้หนึ่งที่มีต่อตัวเลือกหนึ่ง เช่น ผู้ใช้ ก. ให้คะแนนความชอบแก่ภาพยนตร์เรื่อง “Harry Potter” 7 คะแนน จาก 10 คะแนน ระบบสามารถกำหนดค่าฟังก์ชันอรรถประโยชน์ u แต่ละฟังก์ชัน ได้จากตัวเลือกทั้งหมดที่ถูกผู้ใช้ระบุค่าคะแนนความชอบแล้ว หรือจากการคำนวณด้วยการทำงานตัวระบบแนะนำ นอกจากนี้ ระบบสามารถใช้ประโยชน์จากคุณลักษณะบางอย่างของผู้ใช้ เช่น อายุ เพศ รายได้ หรือของตัวเลือก เช่น ประเภทของเนื้อหาในกรณีของภาพยนตร์หรือหนังสือ นำมาใช้ในการพิจารณาได้เช่นกัน

หัวใจหลักของปัญหาที่เกิดขึ้นสำหรับการทำงานของระบบแนะนำ คือ การที่ค่าฟังก์ชันอรรถประโยชน์ u ไม่ได้อธิบายถึงปริภูมิ C x S ทั้งหมด หากอธิบายเพียงบางส่วน ซึ่งหมายถึง ระบบต้องทำนายหาค่าฟังก์ชันอรรถประโยชน์ u ที่เหลือในการอธิบายปริภูมิ C x S ทั้งหมด ในระบบแนะนำ ฟังก์ชันอรรถประโยชน์ถูกแสดงโดยค่าคะแนนความชอบ ซึ่งผู้ให้ระบุให้แก่ตัวเลือก ตัวอย่างเช่น ในระบบแนะนำภาพยนตร์ ผู้ใช้จะต้องทำการให้คะแนนความชอบ

ภาพยนตร์ที่เคยชม ซึ่งได้เป็นชุดข้อมูลตัวอย่างสำหรับการเรียนรู้ของระบบแนะนำ ดังตัวอย่างส่วนของเมทริกซ์คะแนนความชอบในระบบแนะนำภาพยนตร์ ตามตารางที่ 1 [1]

ตารางที่ 1 ส่วนหนึ่งของเมทริกซ์คะแนนความชอบสำหรับระบบแนะนำภาพยนตร์

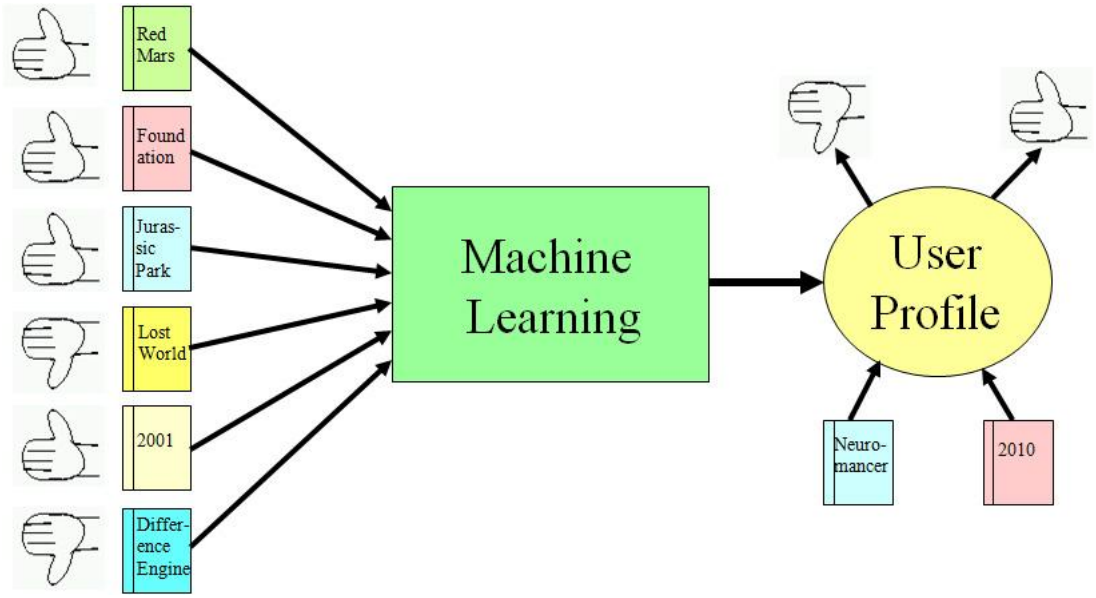
	K-PAX	Life of Brian	Memento	Notorious
Alice	4	3	2	4
Bob	∅	4	5	5
Cindy	2	2	4	∅
David	3	∅	5	2

เมื่อค่าคะแนนความชอบอยู่ในช่วง 1 ถึง 5 และ “∅” เป็นเครื่องหมายแสดงถึงตัวเลือกที่ยังไม่ถูกให้ค่าคะแนนความชอบโดยผู้ใช้ จากนั้นระบบจะทำนายค่าคะแนนความชอบของภาพยนตร์ที่ยังไม่ได้ถูกให้ค่าคะแนนความชอบ และสามารถเลือกแนะนำภาพยนตร์ซึ่งเหมาะสมแก่ผู้ใช้แต่ละคน

ระบบแนะนำสามารถแบ่งประเภทได้ ตามลักษณะวิธีการที่ใช้ในการทำนายค่าคะแนนความชอบที่ยังไม่ได้ถูกระบุ ซึ่งสามารถแบ่งออกเป็น 3 รูปแบบ [1] ดังต่อไปนี้

1. การแนะนำแบบอาศัยเนื้อหา (Content-based Recommendations)

วิธีการรูปแบบนี้ ระบบแนะนำเสนอคำแนะนำแก่ผู้ใช้ โดยพยายามค้นหาตัวเลือกซึ่งมีความคล้ายคลึงกับสิ่งที่ผู้ใช้แต่ละคนรู้สึกชอบหรือสนใจในอดีต โดยพิจารณาจากคุณลักษณะของตัวเลือก คำวิจารณ์ของผู้ใช้ และข้อมูลอื่น ๆ ที่เกี่ยวข้อง เพื่อการคำนวณค่าความคล้ายคลึงกัน (similarity) ในการทำนายให้การแนะนำของระบบ การแนะนำแบบอาศัยเนื้อหา สามารถอธิบายวิธีการทำงานอย่างง่ายได้ดังภาพที่ 2



ภาพที่ 2 การแนะนำแบบอาศัยเนื้อหา

2. การกรองแบบร่วมมือกัน (Collaborative Filtering)

วิธีการรูปแบบนี้ ระบบแนะนำเสนอคำแนะนำแก่ผู้ใช้ โดยพยายามค้นหาผู้ใช้อื่นภายในระบบที่มีความชอบเหมือนหรือคล้ายคลึงกันกับความชอบของผู้ใช้ที่ต้องการคำแนะนำ เพื่อที่จะนำข้อมูลความชอบของผู้ใช้อื่นที่มีลักษณะความชอบคล้ายคลึงกันนั้น มาใช้ในการแนะนำ โดยวิธีการรูปแบบนี้ เป็นวิธีการที่ได้รับความนิยมและถูกนำไปใช้กับระบบแนะนำหลายระบบในปัจจุบัน ดังจะกล่าวถึงรายละเอียดในเนื้อหาส่วนถัดไป

3. วิธีลูกผสม (Hybrid Methods)

วิธีการรูปแบบลูกผสม เป็นการประยุกต์รวมกันของวิธีการรูปแบบการกรองแบบร่วมมือกัน และแบบการแนะนำแบบอาศัยเนื้อหา ซึ่งช่วยแก้ไขข้อจำกัดบางประการที่พบในวิธีการทั้ง 2 รูปแบบแรก โดยสามารถวิธีการรูปแบบลูกผสมได้ 4 รูปแบบหลัก ได้แก่ รูปแบบแรก คือ การใช้การแนะนำแบบอาศัยเนื้อหาและการกรองแบบร่วมมือกันทั้ง 2 วิธีการแยกกัน จากนั้นนำผลลัพธ์ที่ทำนายได้มาใช้พิจารณารวมกัน รูปแบบที่ 2 คือ การประสานลักษณะเฉพาะบางประการของเทคนิคการแนะนำแบบอาศัยเนื้อหาเข้ากับการใช้การกรองแบบร่วมมือกัน รูปแบบที่ 3 คือ การประสานลักษณะเฉพาะบางประการของเทคนิคการใช้การกรองแบบร่วมมือกันเข้ากับการแนะนำแบบอาศัยเนื้อหา และรูปแบบที่ 4 คือ การสร้างโมเดลร่วมเพื่อการทำนายจากวิธีการทั้งสอง อย่างไรก็ตาม ในปัจจุบันระบบแนะนำส่วนใหญ่ ใช้การกรองแบบร่วมมือกันเป็นวิธีการหลัก จึงอาจ

กล่าวได้ว่า วิธีลูกผสม เป็นส่วนเพิ่มเติมเพื่อเพิ่มความสามารถและแก้ไขข้อจำกัดบางประการของการกรองแบบร่วมมือกัน

การกรองแบบร่วมมือกัน (Collaborative Filtering)

ระบบแนะนำที่ใช้วิธีการรูปแบบการกรองแบบร่วมมือกัน [1] เป็นการใช้ข้อมูลความชอบและพฤติกรรมของผู้ใช้ระบบทุกคน โดยระบบจะพยายามค้นหาผู้ใช้อื่นภายในระบบที่มีความชอบเหมือนหรือคล้ายคลึงกันกับความชอบของผู้ใช้ที่ต้องการได้รับคำแนะนำ เพื่อที่จะนำข้อมูลความชอบของผู้ใช้อื่นที่มีลักษณะความชอบคล้ายคลึงกันนั้น มาใช้ในการแนะนำตัวเลือกของระบบ ซึ่งถ้ากำหนดให้ u(c, s) เป็นฟังก์ชันแสดงประโยชน์ของตัวเลือก s ที่มีต่อผู้ใช้ c โดยการทำนายอาศัยการพิจารณา $u(c_j, s)$ ซึ่งแทนประโยชน์ของตัวเลือก s ที่มีต่อผู้ใช้ $c_j \in C$ ผู้ซึ่งเป็นผู้ใช้ที่มีลักษณะความชอบคล้ายคลึงกับผู้ใช้ c ตัวอย่างเช่น ในระบบแนะนำภาพยนตร์ การที่ผู้ใช้ c ขอคำแนะนำจากระบบแนะนำนั้น ตามรูปแบบวิธีการกรองแบบร่วมมือกัน ระบบจะพยายามค้นหา “เพื่อน” ของผู้ใช้ c ซึ่งหมายถึงผู้ใช้อื่น ๆ ภายในระบบที่ชอบภาพยนตร์ลักษณะเดียวกันกับผู้ใช้ c จากนั้นภาพยนตร์ซึ่ง “เพื่อน” ของผู้ใช้ c ชอบและผู้ใช้ c ยังไม่เคยรู้จักมาก่อน จะถูกนำมาแนะนำให้แก่ผู้ใช้ c

ระบบแนะนำมากมายใช้การกรองแบบร่วมมือกัน เป็นพื้นฐานในการประมวลผลของระบบ ซึ่งสามารถพบได้ทั้งในระบบแนะนำที่ใช้เพื่อการศึกษาวิจัยและใช้เพื่ออุตสาหกรรม สำหรับภาคอุตสาหกรรม สามารถเห็นตัวอย่างได้จากระบบแนะนำที่ประสบความสำเร็จเป็นอย่างยิ่งของ Amazon.com [5] ซึ่งช่วยค้นหาสินค้าที่ใกล้เคียงกับสินค้าที่ผู้ใช้สั่งซื้อ นอกจาก Amazon.com แล้ว ตัวอย่างของระบบที่ใช้การกรองแบบร่วมมือกัน ซึ่งรู้จักกันอย่างแพร่หลาย ได้แก่ Netflix [8] ซึ่งช่วยแนะนำในการเลือกชมภาพยนตร์ของลูกค้าพร้อมทั้งได้จัดแข่งขัน Netflix’s Prize เพื่อเปิดโอกาสในการพัฒนาขีดความสามารถของระบบแนะนำ Google News [9] ก็ใช้การกรองแบบร่วมมือกัน ในการเลือกแนะนำข่าวสารแก่ผู้ใช้ และถึงแม้ว่าจะไม่พบรายละเอียดการตีพิมพ์ของเป็นผลงานวิจัย แต่บางเว็บไซต์ที่เป็นของบริษัทซึ่งดำเนินแนวทางเว็บไซต์ในรูปแบบ Web 2.0 และให้บริการเครือข่ายสังคม เช่น Del.icio.us หรือ Last.fm ต่างก็มีระบบแนะนำที่อาศัยรูปแบบวิธีการกรองแบบร่วมมือกันเช่นกัน สำหรับระบบแนะนำเพื่อการศึกษาวิจัยนั้น Grundy เป็นระบบแนะนำระบบแรก ซึ่งมีจุดมุ่งหมายเพื่อสร้างแม่แบบโมเดลการใช้งานผู้ใช้แต่ละคน โดยอาศัยข้อมูลของผู้ใช้ระบบที่จำกัด และสามารถแนะนำหนังสือสำหรับผู้ใช้แต่ละคนได้ ภายหลัง

จากนั้น สามารถพบเห็นระบบแนะนำเพื่อการศึกษาวิจัยได้อย่างมากมาย เช่น Tapestry ซึ่งทำงานโดยอาศัยการระบุกลุ่มผู้ใช้อื่นที่มีความชอบคล้ายคลึงกันด้วยตนเอง GroupLens [6] ซึ่งศึกษาวิจัยเกี่ยวกับระบบแนะนำภาพยนตร์ ด้วยการใช้วิธีการกรองแบบร่วมมือกัน ทำงานโดยอัตโนมัติในการให้คำแนะนำ นอกจากนี้ ตัวอย่างอื่นของระบบแนะนำแบบการกรองแบบร่วมมือกัน ได้แก่ การแนะนำหนังสือของ Amazon.com หรือ Jester ซึ่งแนะนำเรื่องขำขันแก่ผู้ใช้

ใน [1] ผู้เขียนได้กล่าวถึง การแบ่งประเภทของการกรองแบบร่วมมือกัน ออกเป็น 2 กลุ่มใหญ่ ๆ ได้แก่ Memory-based (หรือ Heuristic-based) และ Model-based

ขั้นตอนวิธี Memory-based หรือ Heuristic-based การทำนายค่าคะแนนความชอบ กระทำโดยใช้ค่าคะแนนความชอบของผู้ใช้ทั้งหมดที่ระบบรวมรวมได้ เพื่อค้นหากลุ่มผู้ใช้อื่นที่มีลักษณะความชอบใกล้เคียงกับผู้ใช้เป้าหมาย (target user) มากที่สุด และใช้ข้อมูลความชอบของกลุ่มผู้ใช้ดังกล่าวเพื่อการแนะนำแก่ผู้ใช้เป้าหมายแต่ละคน กล่าวคือ การทำนายค่าคะแนนความชอบที่ไม่ทราบค่า $r_{c,s}$ สำหรับผู้ใช้ c ที่มีต่อตัวเลือก s จะถูกคำนวณจากกลุ่มค่าคะแนนความชอบของผู้ใช้อื่น ๆ ที่มีต่อตัวเลือก s ดังสมการที่ 2.2

$$r_{c,s} = \underset{c' \in \hat{C}}{\text{aggr}}\, r_{c',s} \tag{2.2}$$

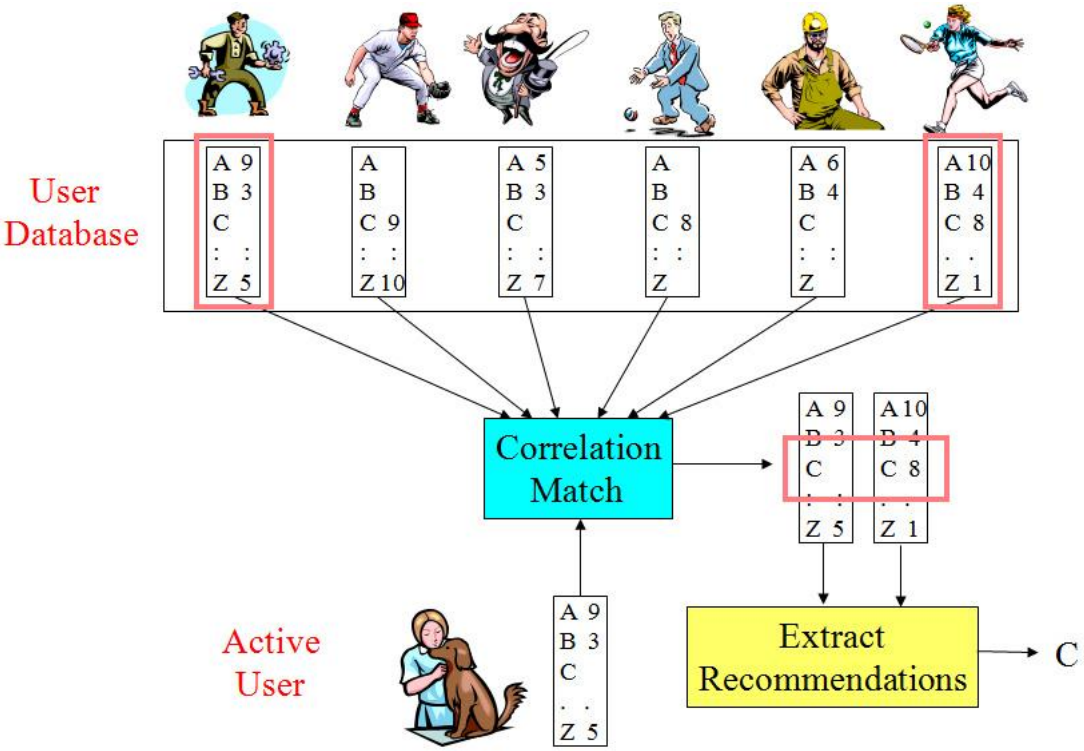
ซึ่ง $\hat{C}$ คือ เซตของผู้ใช้ N คนที่มีความชอบคล้ายคลึงกับผู้ใช้ c มากที่สุด และมีการให้คะแนนตัวเลือก s ทุกคน (N สามารถมีค่าได้ตั้งแต่ 1 ถึง จำนวนผู้ใช้ทั้งหมด) นอกจากนี้โดยส่วนใหญ่แล้วให้การคำนวณจากค่าคะแนนความชอบของผู้ใช้ทั้งหมด มักต้องกำหนดค่าน้ำหนักซึ่งแสดงผลกระทบของผู้ใช้อื่นแต่ละคนที่มีต่อค่าทำนายคะแนนความชอบสุดท้ายด้วย ดังเช่นสมการที่ 2.3

$$r_{c,s} = \sum_{c' \in \hat{C}} \text{sim}(c,c') \times r_{c',s} \tag{2.3}$$

โดยฟังก์ชัน $\text{sim}(c,c')$ แสดงค่าคล้ายคลึงกันระหว่างผู้ใช้ c และ c' ซึ่งอาจให้แทนค่าน้ำหนักได้

ในการทำความเข้าใจแนวคิดเบื้องหลังของขั้นตอนวิธี Memory-based ได้ดียิ่งขึ้น อาจต้องใช้การพิจารณาขั้นตอนวิธี K-Nearest-Neighbor (k-NN) ซึ่งเป็นขั้นตอนวิธีอย่างง่ายที่สุดในการทำการกรองแบบร่วมมือกัน แบบ Memory-based ซึ่งทำการหาผู้ใช้อื่น K คนแรก ที่มีความชอบคล้ายคลึงกับผู้ใช้แต่ละคน ด้วยวิธีการคำนวณค่าฟังก์ชันความคล้ายคลึง (Similarity Function) $\text{sim}(c,c')$ ซึ่งเป็นที่รู้จักและใช้กันอย่างแพร่หลาย เช่น วิธี Minkowski Distance วิธี

Jaccard Similarity หรือวิธี Cosine Similarity เป็นต้น การกรองแบบร่วมมือกันประเภท Memory-based สามารถอธิบายวิธีการทำงานอย่างง่ายได้ดังภาพที่ 3



ภาพที่ 3 การกรองแบบร่วมมือกันประเภท Memory-based

ในทางกลับกัน สำหรับขั้นตอนวิธี Model-based การทำนายค่าคะแนนความชอบ กระทำโดยการใช้ข้อมูลคะแนนความชอบของผู้ใช้ในฐานข้อมูล ในการเรียนรู้ของเครื่องเพื่อสร้างโมเดลในการทำนายค่าคะแนนความชอบของตัวเลือกที่ผู้ใช้ไม่รู้จักมาก่อน ซึ่งใน [1] ได้กล่าวถึงงานวิจัยที่นำเสนอวิธีการใช้งานขั้นตอนวิธี Model-based ไว้อย่างมากมาย ตัวอย่างเช่นใน [10] ซึ่งนำเสนอการใช้ Cluster Models และ Bayesian Network Model หรืองานวิจัยในระยะหลังซึ่งสนใจในการพิจาณาลักษณะความชอบหลากหลายรูปแบบของผู้ใช้ [9] ซึ่งใช้ Probabilistic Latent Semantic Indexing (PLSI) และ MinHash ซึ่งค่อนข้างแตกต่างจากวิธีการที่ถูกนำเสนอใน [10] สำหรับเทคนิคการแยกตัวประกอบเมทริกซ์ (Matrix Factorization) [11] ซึ่งถูกพิจารณาใช้การวิจัยนี้ ก็ถือได้ว่าเป็นเทคนิคหนึ่งในรูปแบบขั้นตอนวิธี Model-based เช่นกัน โดยจะกล่าวถึงในเนื้อหาบทต่อไป

ปัญหาของระบบแนะนำในปัจจุบัน

จากการวิเคราะห์รูปแบบสถานการณ์และวิธีการทำงานของระบบแนะนำในปัจจุบัน พบว่าวิธีการแนะนำแบบอาศัยเนื้อหา และการกรองแบบร่วมมือกัน ต่างก็มีข้อด้อยทั้ง 2 วิธี [1] รวมถึงพบปัญหาจากสถานการณ์การใช้จริงของระบบแนะนำในปัจจุบัน เนื้อหาส่วนนี้จะกล่าวถึงข้อด้อยของรูปแบบวิธีการ และปัญหาที่พบในสถานการณ์ต่าง ๆ ดังนี้

รูปแบบของคุณลักษณะที่ถูกจำกัด (Limited Content Analysis)

เป็นข้อด้อยของวิธีการแนะนำแบบอาศัยเนื้อหา ซึ่งเป็นข้อจำกัดของเครื่องในการพิจารณาคุณลักษณะบางอย่าง ที่เครื่องไม่สามารถวิเคราะห์แยกแยะความแตกต่างเพื่อการคำนวณได้โดยอัตโนมัติ ตัวอย่างเช่น เสียง ภาพ หรือวีดิทัศน์ ข้อจำกัดของเครื่องในอีกกรณีหนึ่งคือ เครื่องอาจไม่สามารถจำแนกความแตกต่างทางด้านมุมมองได้ ตัวอย่างเช่น ไม่สามารถแยกแยะความแตกต่างของหนังสือที่เขียนเนื้อหาได้ดีหรือไม่ดีได้

ระบบแนะนำในกรอบเฉพาะเจาะจง (Overspecialization)

เป็นข้อด้อยของวิธีการแนะนำแบบอาศัยเนื้อหา ซึ่งจากการที่ระบบทำนายตัวเลือกที่จะแนะนำ ด้วยการพิจารณาเฉพาะข้อมูลของสิ่งที่ผู้ใช้แต่ละคนรู้สึกชอบหรือสนใจในอดีต ส่งผลให้ระบบให้คำแนะนำได้ในกรอบที่เฉพาะเจาะจงกับผู้ใช้ และระบบไม่สามารถแนะนำในสิ่งที่แตกต่างออกไปได้จากกลุ่มลักษณะความชอบเดิมของผู้ใช้แต่ละคนได้ ตัวอย่างเช่น ผู้ใช้ซึ่งไม่เคยมีประสบการณ์เกี่ยวกับร้านอาหารจีน แต่ต้องการคำแนะนำในการเลือกร้านอาหารจีน ระบบแนะนำแบบการแนะนำแบบอาศัยเนื้อหา อาจไม่สามารถให้คำแนะนำได้ สำหรับการแก้ไขปัญหานั้น ในบางระบบอาจใช้วิธีการสุ่มตัวเลือกในกลุ่มที่แตกต่างให้แก่ผู้ใช้ เพื่อช่วยปรับปรุงให้การแนะนำในภายหลังสามารถแนะนำได้หลากหลายยิ่งขึ้น นอกจากนี้ในบางระบบ เช่น ระบบแนะนำข่าวสาร อาจมีข้อกำหนดในการเลือกไม่แนะนำสิ่งที่พิจารณาแล้วมีความใกล้เคียงกันมากเกินไปด้วย เนื่องจากต้องการป้องกันตัวเลือกที่มีความซ้ำซ้อนกัน

ปัญหาผู้เข้าใช้งานใหม่ในระบบ (New User Problem)

ผู้เข้าใช้งานใหม่ในระบบ ซึ่งยังไม่ได้ให้ข้อมูลความชอบแก่ระบบแนะนำอย่างเพียงพอ เพื่อนำมาใช้ในการพิจารณาลักษณะความชอบของผู้ใช้ อาจยังผลให้ระบบไม่สามารถให้คำแนะนำที่แม่นยำแก่ผู้เข้าใช้ใหม่รายนั้น

ปัญหาการเพิ่มสิ่งแนะนำใหม่ในระบบ (New Item Problem)

สิ่งแนะนำใหม่ที่ถูกเพิ่มเข้ามาภายในระบบแนะนำแบบการกรองแบบร่วมมือกันนั้น อาจไม่ได้รับการถูกแนะนำให้แก่ผู้ใช้ เนื่องจากสิ่งแนะนำใหม่มักยังไม่เคยถูกเลือกให้ค่าคะแนนความชอบโดยผู้ใช้ ทำให้ระบบไม่พิจารณาเลือกแนะนำสิ่งแนะนำใหม่เหล่านั้นแก่ผู้ใช้

ความต้องการใช้งานทรัพยากรระบบ (Resource Requirements)

จากสถานการณ์การใช้งานจริงที่เกิดขึ้นในปัจจุบัน ระบบแนะนำอาจต้องรองรับการแนะนำสินค้าที่มีให้เลือกในปริมาณมหาศาลและเพิ่มขึ้นเปลี่ยนแปลงอยู่ตลอดเวลา แก่ผู้ใช้ระบบที่มีปริมาณเพิ่มขึ้นในแนวโน้มเดียวกัน ส่งผลให้ระบบแนะนำต้องการใช้งานทรัพยากรระบบเพื่อการคำนวณมากขึ้น ซึ่งให้ความสนใจกับการพิจารณารูปแบบการประมวลผลข้อมูลในสเกลขนาดใหญ่ (large-scale data processing) เช่นกัน

ข้อมูลในระบบมีปริมาณเบาบาง (Data Sparsity)

ระบบแนะนำส่วนใหญ่ประสบปัญหาในเรื่องปริมาณของข้อมูลความชอบที่ระบบรวบรวมได้จากผู้ใช้ มีปริมาณน้อยกว่ามากเมื่อเทียบกับปริมาณข้อมูลที่ระบบต้องทำการทำนายเพื่อพิจารณาเลือกแนะนำ ซึ่งเรียกปัญหาดังกล่าวนี้ว่า Data Sparsity และสามารถพบเห็นตัวอย่างของปัญหาของนี้ได้จากการพิจารณาชุดข้อมูลทดสอบ (datasets) ซึ่งเป็นที่รู้จักสำหรับระบบแนะนำ เช่น MovieLens Jester EachMovie หรือ Netflix จากการพิจารณาชุดข้อมูลทดสอบเหล่านี้ ต่างก็พบลักษณะของปัญหา Data Sparsity เนื่องจากมีความหนาแน่นของข้อมูลในระดับต่ำ จากปัญหาดังกล่าวจึงเป็นเรื่องค่อนข้างยากที่ระบบแนะนำ จะสามารถเรียนรู้ลักษณะความชอบของผู้ใช้แต่ละคนเพื่อแนะนำได้อย่างมีประสิทธิภาพ อย่างไรก็ตาม ในความเป็นจริงแล้ว Data Sparsity คือ ประเด็นสำคัญของปัญหาระบบแนะนำ เนื่องจากถ้าหากระบบทราบค่าคะแนนความชอบของผู้ใช้ที่มีต่อตัวเลือกทั้งหมดโดยสมบูรณ์แล้ว ระบบอาจไม่มีความจำเป็นที่จะต้องทำนายเพื่อพิจารณาให้คำแนะนำใด ๆ อีก

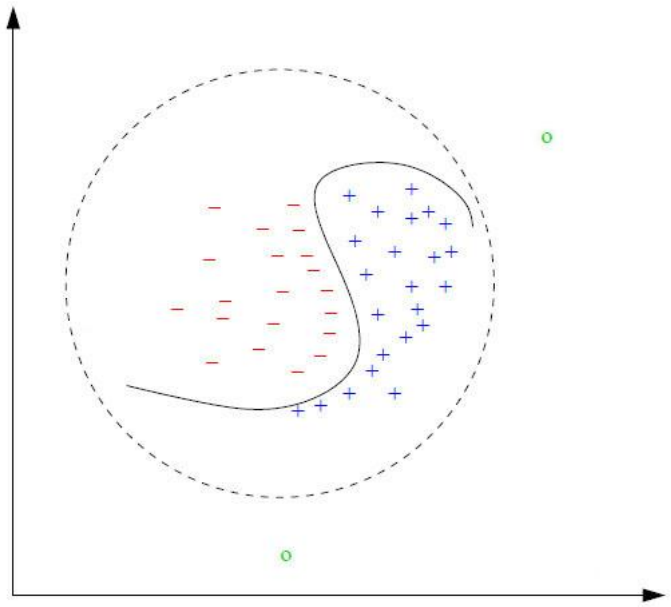
สำหรับปัญหา Data Sparsity ได้มีการนำเสนอวิธีในการแก้ปัญหาเบื้องต้นหลายรูปแบบ เช่น การข้อมูลคุณลักษณะของผู้ใช้ (user profile) เช่น อายุ เพศ สถานะการเงิน ในการพิจารณาร่วมด้วยซึ่งเรียกวิธีการนี้ว่า การกรองแบบอาศัยประชากรศาสตร์ (Demographic Filtering) [1] หรือการใช้เทคนิค Singular Value Decomposition (SVD) หรือการใช้เทคนิคการแยกตัวประกอบเมทริกซ์ (Matrix Factorization) [11] เพื่อลดขนาดมิติในการพิจารณาข้อมูล

ความชอบที่เป็นการคำนวณแบบเวกเตอร์ และหากพิจารณาต่อเนื่องจากปัญหา Data Sparsity อาจกล่าวได้ว่า เกิดจากระบบไม่สามารถเก็บข้อมูลความชอบของผู้ใช้ได้หรือเก็บได้น้อย ซึ่งได้มีแนวทางที่จะใช้ข้อมูลอื่นมาทดแทนข้อมูลความชอบที่ระบุโดยชัดแจ้ง เช่น ข้อมูลการสั่งซื้อสินค้า ข้อมูลเวลาเข้าใช้งานในแต่ละเว็บเพจ ข้อมูลการคลิกเชื่อมโยง หรือข้อมูลการค้นหน้าเว็บเพจที่สนใจ ซึ่งข้อมูลเหล่าสามารถเก็บรวบรวมได้ง่ายจากพฤติกรรมการใช้งานของผู้ใช้ระบบแต่ละคน แต่อย่างไรก็ตาม ข้อมูลพฤติกรรมหรือข้อมูลป้อนกลับโดยปริยาย (implicit feedback) เหล่านี้สามารถแสดงได้เพียงลักษณะเชิงบวก หรือตัวอย่างบวก (positive example) ในการเรียนรู้ของเครื่องเท่านั้น

การจำแนกประเภทแบบคลาสเดี่ยว (One-class Classification)

การเรียนรู้แบบมีผู้สอน (Supervised Learning) มีวัตถุประสงค์เพื่อการสร้างโมเดลจากชุดข้อมูลตัวอย่างที่ทราบป้ายกำกับคลาส (class label) และด้วยการใช้ฟังก์ชันหรือโมเดลที่เรียนรู้ได้ดังกล่าว เพื่อการระบุป้ายกำกับคลาสให้แก่ตัวอย่าง (instance) ใหม่ หรือกล่าวได้ว่างานของการกำหนดป้ายกำกับคลาสให้แก่ตัวอย่างใหม่ดังกล่าว เป็นงานของการเรียนรู้ด้วยเครื่องเพื่อการจำแนกประเภท (Classification)

การจำแนกประเภทด้วยเครื่องโดยทั่วไปนั้น เป็นการจำแนกประเภทแบบหลายคลาส (Multi-class Classification) ซึ่งปัญหาที่พบได้ส่วนใหญ่ในงานของการเรียนรู้ด้วยเครื่องพื้นฐานทั่วไป คือ การจำแนกประเภทแบบสองคลาส (Two-class Classification) ตัวอย่างเช่นการระบุว่า ใช่หรือไม่ใช่ เกี่ยวข้องหรือไม่เกี่ยวข้อง ซึ่งเครื่องจะสร้างฟังก์ชันหรือโมเดลที่ใช้ในการทำนายจากการเรียนรู้ชุดข้อมูลฝึกที่ประกอบด้วย ตัวอย่างบวกและตัวอย่างลบ (negative example) อย่างไรก็ตาม ในบางงานประยุกต์ที่เป็นงานของการค้นหาสิ่งแปลกแยก (outlier) ตัวจำแนกด้วยเครื่องอาจต้องถูกใช้ในการพิจารณาตรวจจับสิ่งแปลกแยกนั้น ซึ่งลักษณะการทำงานนี้อาจกล่าวได้ว่าเป็นการจำแนกประเภทแบบคลาสเดี่ยว (One-class Classification) ดังอธิบายให้สามารถเข้าใจได้โดยง่าย ด้วยภาพที่ 4 นอกจากนี้ สำหรับระบบแนะนำ ปัญหาการจำแนกประเภทแบบคลาสเดี่ยว อาจเกิดขึ้นในสถานการณ์ที่ต้องเรียนรู้ลักษณะความชอบของผู้ใช้ จากข้อมูลความชอบของผู้ใช้ที่เป็นตัวอย่างบวก หรือจากการข้อมูลการใช้งานตอบสนองระบบของผู้ใช้เพียงอย่างเดียวอีกด้วย



ภาพที่ 4 รูปแบบการจำแนกคลาสด้วยเครื่อง โดยเครื่องหมาย + แทนตัวอย่างบวก เครื่องหมาย – แทนตัวอย่างลบ ซึ่งถูกจำแนกคลาสด้วยเส้นทึบ และเครื่องหมาย o แทนสิ่งแปลกแยก ซึ่งถูกจำแนกคลาสด้วยเส้นประ

เพื่อแก้ปัญหาลักษณะการจำแนกประเภทแบบคลาสเดี่ยว วิธีการหลากหลายแบบถูกนำมาใช้ ตัวอย่างเช่น One-class Support Vector Machines (One-class SVMs) [12] ซึ่งนำเสนอวิธีการที่พัฒนาต่อยอดจาก SVMs เพื่อตรวจจับสิ่งแปลกแยก การใช้การประมาณค่าความหนาแน่น [13] หรือการใช้ขั้นตอนวิธีที่มีลักษณะคล้ายคลึงกับขั้นตอนวิธี Expectation-maximization (EM) ในการจัดการกับปัญหาการจำแนกประเภทแบบคลาสเดี่ยว [14]

การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering)

การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว หรือ OCCF ถูกนำเสนอครั้งแรกโดย Pan และคณะ [2] ซึ่งระบุถึงปัญหาของระบบแนะนำที่ใช้การเรียนรู้ของเครื่องด้วยตัวอย่างบวก (positive example) และตัวอย่างที่ไม่ระบุป้ายกำกับคลาส (unlabeled example) โดยตัวอย่างที่ไม่ระบุป้ายกำกับคลาสดังกล่าวประกอบด้วยข้อมูลส่วนที่เป็นตัวอย่างลบ และข้อมูลส่วนที่คาดว่าจะถูกพิจารณาเป็นตัวอย่างบวก ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวเกิดขึ้นกับระบบรูปแบบต่าง ๆ ตัวอย่างเช่น การคั่นหน้าเชิงสังคม (Social Bookmarking) เพื่อคั่นหน้าเว็บเพจที่สนใจ การติดตามประวัติการคลิกเชื่อมโยง และการติดตามประวัติการซื้อสินค้า ซึ่งตัวอย่าง

ระบบแบบการคั่นหน้าเชิงสังคมนั้น ค่อนข้างได้รับความสนใจเป็นอย่างมากในปัจจุบันและมีเว็บไซต์ที่มีส่วนบริการในลักษณะการคั่นหน้าเชิงสังคมมากมาย เช่น Facebook หรือ Del.icio.us เป็นต้น การคั่นหน้าเชิงสังคม คือ ระบบที่เปิดให้ผู้ใช้พิจารณาคั่นหน้าหรือเก็บยูอาร์แอล (url) สำหรับเชื่อมโยงไปไอเทม (item) ที่ผู้ใช้สนใจหรือชอบบนอินเทอร์เน็ต เช่น เว็บเพจ ข่าวสาร เอกสาร ภาพยนตร์ หนังสือ เป็นต้น นอกจากนี้ผู้ใช้สามารถแบ่งปันประสบการณ์หรือข้อมูลการคั่นหน้าดังกล่าวแก่ผู้ใช้อื่นในระบบได้ อย่างไรก็ตาม สำหรับพฤติกรรมของผู้ใช้ระบบการคั่นหน้าเชิงสังคม มีคำอธิบายสาเหตุพฤติกรรมที่ผู้ใช้ไม่ทำการคั่นหน้าไอเทม i ใด ๆ ในระบบอยู่ 2 สาเหตุ โดยสาเหตุแรก คือ ผู้ใช้ควรที่จะสนใจและทำการคั่นหน้าไอเทม i นั้นแต่ไอเทม i ไม่เคยผ่านเข้าสู่การพิพิจารณาของผู้ใช้เลย ส่วนอีกสาเหตุ คือ ไอเทม i ดังกล่าวผ่านการพิจารณาของผู้ใช้แล้วแต่ผู้ใช้ไม่สนใจ จึงไม่ทำการคั่นหน้าไอเทม i นั้น

งานวิจัยแรก ๆ เกี่ยวกับปัญหาในลักษณะที่ใกล้เคียงกับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวนั้น มักแก้ปัญหาด้วยเทคนิคการจำแนกประเภทแบบคลาสเดี่ยว ดังอธิบายในเนื้อหาส่วนที่ผ่านมา ซึ่งไม่ได้ใช้ประโยชน์อย่างเต็มที่จากตัวอย่างที่ไม่ระบุป้ายกำกับคลาส อย่างไรก็ตาม Pan และคณะ [2] ได้นำเสนอถึงปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว และวิธีแก้ปัญหาที่ใช้ประโยชน์จากข้อมูลตัวอย่างที่ไม่ระบุป้ายกำกับคลาสดังกล่าว ด้วยเทคนิคการแยกตัวประกอบเมทริกซ์ที่มีการพิจารณาค่าน้ำหนัก แนวคิดหลักของเทคนิคดังกล่าว คือ การปฏิบัติเสมือนว่าตัวอย่างที่ไม่ระบุป้ายกำกับคลาสทั้งหมดเป็นตัวอย่างลบ จากนั้นกำหนดค่าน้ำหนักซึ่งสะท้อนระดับความสำคัญของตัวอย่างที่ไม่ระบุป้ายกำกับคลาสแต่ละตัวอย่างในรูปของเมทริกซ์ค่าน้ำหนัก และใช้ขั้นตอนวิธี Weighted Alternating Least Squares (wALS) ในการแก้ปัญหาการแยกตัวประกอบเมทริกซ์ที่มีการพิจารณาค่าน้ำหนักดังกล่าว

นอกจากงานวิจัยของ Pan และคณะ [2] แล้ว ภายหลังได้มีผู้นำเสนอเทคนิคสำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ซึ่งใช้การแยกตัวประกอบเมทริกซ์ในรูปแบบอื่นและให้ประสิทธิผลสำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวที่ดีกว่าแก้ปัญหาการแยกตัวประกอบเมทริกซ์แบบพื้นฐาน งานวิจัยดังกล่าวถูกนำเสนอโดย Sindhwani และคณะ [4] ซึ่งนำเสนอการใช้การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization) สำหรับปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว

อย่างไรก็ตาม เทคนิคการแยกตัวประกอบเมทริกซ์ สำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ในงานวิจัยต่าง ๆ ที่กล่าวมาข้างต้นนั้น พิจารณาเฉพาะข้อมูลตัวอย่างบวกที่

เป็นความชอบความสนใจของผู้ใช้เท่านั้น โดยไม่ได้ใช้ข้อมูลคุณลักษณะอื่น ๆ ของผู้ใช้เป็นข้อมูลนำเข้าเพิ่มเติมสำหรับการเรียนรู้ด้วยเครื่องเลย ดังนั้นการใช้ประโยชน์จากข้อมูลคุณลักษณะของผู้ใช้ ตัวอย่างเช่น ข้อมูลความสัมพันธ์ระหว่างกันของผู้ใช้ทุกคนในระบบ ซึ่งโดยทั่วไปมีอยู่แล้วในระบบปัจจุบัน เช่น บริการเครือข่ายสังคมออนไลน์ จึงเป็นทางเลือกหนึ่งสำหรับการปรับปรุงคุณภาพของระบบแนะนำที่ใช้การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว

บทที่ 3
งานวิจัยที่เกี่ยวข้อง

กรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering)

งานวิจัยจำนวนมากในหลายปีที่ผ่านมา ได้นำเสนอวิธีการสำหรับระบบแนะนำแบบการกรองแบบร่วมมือกัน เพื่อการแนะนำสิ่งต่าง ๆ แก่ผู้ใช้ระบบ ซึ่งลักษณะการทำงานของระบบแนะนำส่วนใหญ่ในปัจจุบันจะเก็บข้อมูลความชอบของผู้ใช้ ในลักษณะการเก็บค่าคะแนนความชอบที่ผู้ใช้กำหนดให้แก่สิ่งต่าง ๆ ซึ่งระบบสามารถนำมาใช้ในการพิจารณาได้ทันที หรืออาจใช้วิธีพิจารณาค่าคะแนนความชอบสูงเป็นตัวอย่างบวก และค่าคะแนนความชอบต่ำเป็นตัวอย่างลบ ในการเรียนรู้ของเพื่อการจำแนกคลาสของระบบแนะนำได้เช่นกัน ซึ่ง Adomavicius และ Tuzhilin [1] ได้นำเสนอข้อมูลสรุปเกี่ยวกับตัวอย่างระบบแนะนำเหล่านั้น นอกจากนี้ ในกรณีของปัญหาการกรองแบบร่วมมือกัน ในระบบที่มีปริมาณข้อมูลสูญหาย (missing data) เป็นจำนวนมาก ก็มีอธิบายถึงในงานวิจัย เช่น Marlin และคณะ [15] ได้นำเสนอเกี่ยวกับโมเดลการแจกแจงของข้อมูลสูญหายในปัญหาการกรองแบบร่วมมือกัน แต่ก็ไม่ได้อธิบายถึงกรณีที่ไม่มีตัวอย่างลบ

Das และคณะ [9] นำเสนอระบบแนะนำแบบการกรองแบบร่วมมือกัน ซึ่งแนะนำด้วยการเรียนรู้จากข้อมูลทั้งตัวอย่างบวกและตัวอย่างลบ ดังเช่นลักษณะการเก็บค่าคะแนนความชอบที่ได้กล่าวมาแล้ว รวมถึงในกรณีของระบบแนะนำที่ต้องมีการประมวลผลข้อมูลในสเกลขนาดใหญ่ เช่น ระบบแนะนำของ Netflix [7] ทีมผู้ชนะจากการแข่งขัน KDD Cup 2007 ในการแก้โจทย์ปัญหาแบบการกรองแบบร่วมมือกัน ได้แก่ Kurucz และคณะ [16] ได้นำเสนอการใช้ Singular Value Decomposition (SVD) ร่วมกับการเรียนรู้ด้วยข้อมูลฝึกแบบทวิภาค (ประกอบด้วยตัวอย่างบวกและตัวอย่างลบ)

Koren และคณะ [17] นำเสนอข้อมูลสรุปเกี่ยวกับการใช้เทคนิคการแยกตัวประกอบเมทริกซ์ [11] กับระบบแนะนำแบบการกรองแบบร่วมมือกัน และเก็บข้อมูลค่าคะแนนความชอบของผู้ใช้ ซึ่งเทคนิคการแยกตัวประกอบเมทริกซ์ ช่วยแก้ปัญหา Data Sparsity โดยใน [17] กล่าวถึงโมเดลพื้นฐานสำหรับเทคนิคการแยกตัวประกอบเมทริกซ์ และวิธีการหาผลเฉลยด้วยขั้นตอนวิธี Low Rank Approximation ตามรูปแบบที่นำเสนอโดย Srebro [11] พร้อมทั้งอธิบายถึงแนวทางเพิ่มขีดความสามารถของระบบแนะนำที่ใช้เทคนิคการแยกตัวประกอบเมทริกซ์ ได้แก่ การเพิ่มความลำเอียง การเพิ่มคุณลักษณะของผู้ใช้ที่เกี่ยวข้อง การเปลี่ยนแปลงลักษณะ

ความชอบของผู้ใช้ตามช่วงเวลา และการหาพิจารณาค่าน้ำหนักหรือค่าความเชื่อมั่นร่วมด้วยในการผลเฉลยของเทคนิคการแยกตัวประกอบเมทริกซ์ ซึ่งอาจเรียกขั้นตอนวิธีดังกล่าวนี้ได้ว่า Weighted Low Rank Approximation (wLRA) [3]

Pan และคณะ [2] ได้นำเสนอระบบแนะนำรูปแบบการกรองแบบร่วมมือกัน ซึ่งเรียนรู้ลักษณะความชอบของผู้ใช้ ด้วยข้อมูลความชอบของผู้ใช้ที่เป็นตัวอย่างบวกเพียงอย่างเดียว หรือถูกระบุให้เป็นปัญหาที่เรียกว่า One-class Collaborative Filtering (OCCF) โดยการใช้เทคนิคการแยกตัวประกอบเมทริกซ์ และวิธีการหาผลเฉลยด้วยขั้นตอนวิธี wLRA ซึ่ง Pan และคณะ ได้ทำการกำหนดรูปแบบการระบุค่าน้ำหนักของข้อมูลที่ไม่ทราบค่าในขั้นตอนการหาผลเฉลยของเทคนิคการแยกตัวประกอบเมทริกซ์ ด้วยการพิจารณาปริมาณตัวอย่างบวกที่ระบบทราบค่าและกำหนดให้สัมพันธ์กับการระบุค่าน้ำหนักของข้อมูลที่ไม่ทราบค่า ซึ่งเป็นค่าคงที่แต่ละตำแหน่งภายในเมทริกซ์ โดยรายละเอียดของการกำหนดรูปแบบการระบุค่าน้ำหนักดังกล่าว จะอธิบายเพิ่มเติมในเนื้อหาหัวข้อแนวทางการแก้ปัญหาด้วยการพิจารณาค่าน้ำหนัก (Weighting based Approaches)

Sindhwani และคณะ [4] เป็นอีกกลุ่มหนึ่งที่ได้นำเสนอเทคนิคสำหรับระบบแนะนำรูปแบบการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering) ในภายหลังจากงานวิจัยของ Pan และคณะ [2] โดย Sindhwani และคณะ พบว่าเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization) ซึ่งเป็นรูปแบบหนึ่งในกลุ่มโมเดลการแยกตัวประกอบเมทริกซ์ ให้ผลลัพธ์ในการประมวลผลของระบบแนะนำที่ดีกว่าเมื่อเทียบกับการใช้การแยกตัวประกอบเมทริกซ์แบบพื้นฐาน โดยจะอธิบายเพิ่มเติมในเนื้อหาหัวข้อการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ แต่ทั้งนี้เทคนิคสำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวที่ Sindhwani และคณะ นำเสนอ ไม่ได้กล่าวถึงการพิจารณาค่าน้ำหนักในขั้นตอนการหาผลเฉลยของเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ

จากรูปแบบการหาผลเฉลยของเทคนิคการแยกตัวประกอบเมทริกซ์ ใน [2] และ [4] ซึ่งไม่มีการคำนึงข้อมูลคุณลักษณะของผู้ใช้หรือข้อมูลคุณลักษณะของตัวเลือกต่าง ๆ รวมถึงอาจสามารถหาวิธีการที่เหมาะสมกว่า สำหรับการกำหนดค่าน้ำหนักของข้อมูลที่ไม่ทราบค่า ในขั้นตอนการหาผลเฉลย ผู้ดำเนินการวิจัยจึงเล็งเห็นแนวทางในการนำมาวิจัยพัฒนาต่อ นอกจากนี้สถานการณ์ของระบบที่ต้องเรียนรู้ลักษณะความชอบของผู้ใช้ด้วยข้อมูลที่เป็นตัวอย่างบวก มักปรากฏให้เห็นได้อย่างชัดเจนบนเว็บไซต์ที่ให้บริการเครือข่ายสังคม ผู้ดำเนินการวิจัยจึงเกิดแนวคิด

ที่จะใช้ข้อมูลความสัมพันธ์บนเครือข่ายสังคม ช่วยในการประมวลผลข้อมูลของระบบแนะนำ มีงานวิจัยจำนวนหนึ่ง นำเสนอการประยุกต์ใช้งานข้อมูลเครือข่ายสังคม ในการเรื่องการกรองแบบร่วมมือกัน ซึ่งสามารถยกตัวอย่างงานวิจัย ที่อธิบายรูปแบบการประยุกต์ใช้ได้อย่างชัดเจน เช่น Walter และคณะ [18] ได้นำเสนอโมเดลของระบบแนะนำ โดยอาศัยค่าความเชื่อถือจากข้อมูลบนเครือข่ายสังคม หรือ Trust-based Recommendation ซึ่งเกิดจากแนวคิดบนสมมติฐานเรื่อง การแนะนำโดยอ้างอิงข้อมูลของผู้ใช้ที่เรารู้จักเชื่อถือกันในชีวิตจริง และการพิจารณาข้อมูลกลุ่มผู้ใช้ที่แตกต่างกันในการแนะนำสิ่งที่อยู่ในโดเมนที่แตกต่างกัน

ทั้งนี้เพื่อเป็นการสรุปเปรียบเทียบข้อแตกต่างระหว่างวิธีการของงานวิจัยที่เกี่ยวข้องกับแนวคิดทางการวิจัยที่นำเสนอในวิทยานิพนธ์นี้ ผู้ดำเนินการวิจัยจึงสรุปเป็นตารางเปรียบเทียบ ดังตารางที่ 2

ตารางที่ 2 ข้อแตกต่างระหว่างงานวิจัยที่เกี่ยวข้องกับแนวคิดทางการวิจัยที่นำเสนอ

งานวิจัยที่เกี่ยวข้อง	**แนวคิดทางการวิจัยที่นำเสนอ**
1. ข้อมูลแสดงความชอบของผู้ใช้เป็นค่าคะแนนความชอบหรืออัตราเชิงตัวเลข [5], [6], [7], [8], [17], [19]	1. ข้อมูลแสดงความชอบของผู้ใช้เป็นข้อมูลประเภทตัวอย่างบวกเท่านั้น
2. ไม่ใช้ข้อมูลความสัมพันธ์ระหว่างกันของผู้ใช้บนเครือข่ายสังคม [2], [5], [6], [7], [17], [19], [20]	2. ใช้ข้อมูลความสัมพันธ์ระหว่างกันของผู้ใช้บนเครือข่ายสังคม
3. ระบบแนะนำพิจารณาแนะนำตัวเลือก เพียงโดเมนเดียว [5], [6], [7], [17], [19]	3. ระบบแนะนำพิจารณาแนะนำตัวเลือกหลากหลายโดเมน
4. ไม่ใช้ข้อมูลคุณลักษณะของผู้ใช้ [2], [4], [5]	4. ใช้ข้อมูลคุณลักษณะของผู้ใช้

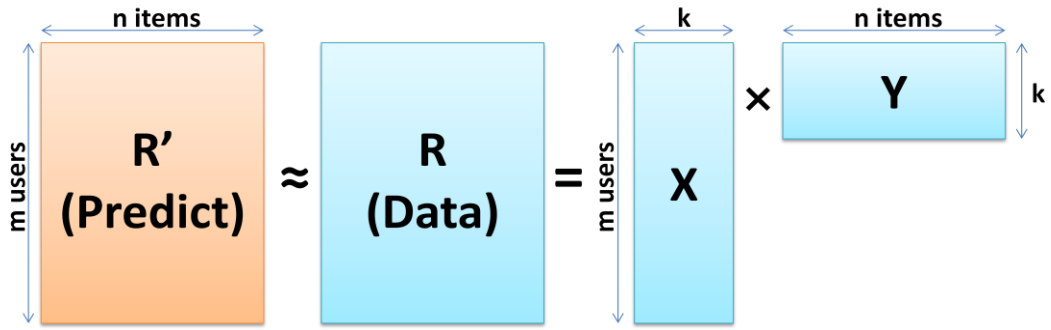
การแยกตัวประกอบเมทริกซ์สำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว

(Matrix Factorization for One-class Collaborative Filtering)

การแยกตัวประกอบเมทริกซ์ (Matrix Factorization) เป็นวิธีการที่ประสบความสำเร็จเป็นอย่างมาก ในการจัดการกับปัญหาข้อมูลในระบบมีปริมาณเบาบาง (Data Sparsity) ดังแสดงให้เห็นประสิทธิผลจากการถูกนำไปใช้และกล่าวถึงในบทความของกลุ่มผู้ชนะการแข่งขันปรับปรุงคุณภาพของระบบแนะนำภาพยนตร์ของ Netflix [7] ในปี ค.ศ. 2009 รวมถึงได้รับการยอมรับว่าเป็นวิธีการที่ดีที่สุดในปัจจุบันสำหรับปัญหาการกรองแบบร่วมมือกัน (Collaborative Filtering) และปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (OCCF) ในเทคนิคการแยกตัวประกอบเมทริกซ์แบบพื้นฐาน กำหนดเมทริกซ์ผู้ใช้-ไอเทม (user-item) R ซึ่งเป็นเมทริกซ์ที่เก็บข้อมูลความสนใจหรือความชอบของผู้ใช้ ตัวอย่างเช่น ค่าคะแนนความชอบที่ชัดแจ้ง (explicit rating) หรือ ข้อมูลคั่นหน้าเชิงสังคม (social bookmark) โมเดลของเทคนิคการแยกตัวประกอบเมทริกซ์นี้ มีคุณลักษณะคล้ายคลึงกับเทคนิค Singular Value Decomposition (SVD) ซึ่งเป็นที่รู้จักกันอย่างแพร่หลาย สำหรับแนวคิดพื้นฐานของการแยกตัวประกอบเมทริกซ์ คือ เมทริกซ์ผู้ใช้-ไอเทม R สามารถถูกประมาณค่าได้โดยการแยก (decomposition) เมทริกซ์ R ไปเป็นเมทริกซ์ที่มีค่าลำดับชั้นต่ำ (low rank matrix) 2 เมทริกซ์ เมทริกซ์ X และเมทริกซ์ Y ดังสมการที่ 3.1

$$R'_{ui} \approx X_u Y_i^T \tag{3.1}$$

เมื่อผลลัพธ์ที่ได้จากผลคูณจุด (dot product) ใช้ในการประมาณค่าปฏิสัมพันธ์กันระหว่างผู้ใช้ u และไอเทม i หากกำหนดให้ m เป็นจำนวนผู้ใช้ในระบบ และ n เป็นจำนวนไอเทมในระบบ มิติของเมทริกซ์ X และเมทริกซ์ Y จะเป็น m คูณ k และ n คูณ k ตามลำดับ โดย k แสดงถึงจำนวนปัจจัยแฝง (latent factor) ซึ่งอนุมานจากแบบรูปข้อมูลความสนใจหรือความชอบของผู้ใช้ การประมาณค่าปฏิสัมพันธ์กันระหว่างผู้ใช้ u และไอเทม i ด้วยการแยกตัวประกอบเมทริกซ์แบบพื้นฐานสามารถอธิบายให้เข้าได้ง่ายยิ่งขึ้นดังภาพที่ 5



ภาพที่ 5 การแยกตัวประกอบเมทริกซ์แบบพื้นฐาน

งานวิจัยที่ผ่านมา ดังเช่น [2] และ [17] เรียนรู้เมทริกซ์ตัวประกอบ (เมทริกซ์ X และเมทริกซ์ Y) รวมถึงป้องกันปัญหาการสร้างโมเดลของการเรียนรู้ที่ปรับเหมาะเกินไป (overfitting) ด้วยโมเดล Regularization ดังสมการที่ 3.2

$$\min_{X,Y} \sum_{ui} (R_{ui} - X_u Y_i^T)^2 + \lambda (\|X_u\|_F^2 + \|Y_i\|_F^2) \quad (3.2)$$

เมื่อ $\|.\|_F$ คือ นอร์มของ Frobenius และ λ คือ พารามิเตอร์สำหรับ Regularization ซึ่งเป็นตัวแปรเสริมสำหรับควบคุมค่าที่เปลี่ยนแปลงในการหาผลเฉลยของโมเดล Regularization สำหรับปัญหาการหาค่าเหมาะที่สุดเพื่อการลดให้เหลือน้อยที่สุด (minimizing) สมการที่ 3.2 กระทำได้โดยขั้นตอนวิธี เช่น Stochastic Gradient Descent หรือ Alternating Least Squares (ALS)

การแก้ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวที่นำเสนอในงานวิจัยของ Pan และคณะ [2] กระทำโดยขั้นตอนวิธี Alternating Least Squares ซึ่งสามารถประยุกต์ใช้กับปัญหาดังกล่าวได้และให้ประสิทธิผลของการแนะนำที่ดีกว่า ขั้นตอนวิธี Alternating Least Squares เป็นขั้นตอนวิธีแบบทำซ้ำ (iterative algorithm) โดยเริ่มต้นจากการสุ่มกำหนดค่าของเมทริกซ์ X หรือเมทริกซ์ Y จากนั้นทำการปรับค่าของเมทริกซ์ Y ด้วยการแก้ปัญหา Least Squares และกระทำในทางกลับกันเพื่อการปรับค่าของเมทริกซ์ X ทำซ้ำขั้นตอนเช่นนี้ไปเรื่อย ๆ จนกระทั่งค่าของเมทริกซ์ X และเมทริกซ์ Y ลู่เข้า ขั้นตอนวิธี Alternating Least Squares สามารถอธิบายได้ดังภาพที่ 6

ข้อมูลนำเข้า: เมทริกซ์ผู้ใช้-ไอเทม R ขนาด m คูณ n, จำนวนปัจจัยแฝง (พารามิเตอร์ k)

1. **กำหนด** เมทริกซ์ X **ขนาด** m **คูณ** k, เมทริกซ์ Y **ขนาด** n **คูณ** k
2. สุ่มกำหนดค่าของเมทริกซ์ Y
3. **ทำซ้ำ**

    ปรับค่าของเมทริกซ์ X ด้วยการแก้ปัญหา Least Squares ที่เหมาะสมกับเมทริกซ์ Y

    ปรับค่าของเมทริกซ์ Y ด้วยการแก้ปัญหา Least Squares ที่เหมาะสมกับเมทริกซ์ X

จนกระทั่ง ค่าของเมทริกซ์ X และเมทริกซ์ Y ลู่เข้าสู่ค่าคงที่

4. **ส่งออก** เมทริกซ์ X และเมทริกซ์ Y

ภาพที่ 6 ขั้นตอนวิธี Alternating Least Squares (ALS)

แนวคิดของการแยกตัวประกอบเมทริกซ์ สามารถขยายไปสู่โมเดลการแยกตัวประกอบเมทริกซ์ได้หลากหลายรูปแบบ ซึ่งมีคุณลักษณะและคุณสมบัติที่แตกต่างกันเล็กน้อย เพื่อให้เหมาะสมกับการนำไปใช้งานกับการแก้ปัญหารูปแบบต่าง ๆ รูปแบบหนึ่งของโมเดลการแยกตัวประกอบเมทริกซ์ที่สามารถนำมาใช้กับปัญหาการกรองแบบร่วมมือกันได้ และจะกล่าวถึงในเนื้อหาหัวข้อถัดไปได้แก่ การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization)

การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization)

การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ เป็นรูปแบบหนึ่งของโมเดลการแยกตัวประกอบเมทริกซ์ ซึ่งถูกคิดค้นและมีประสิทธิภาพอย่างมากในการนำไปใช้กับปัญหาทางด้านการประมวลผลภาพ (Image Processing) อย่างไรก็ตาม Sindhwani และคณะ [4] ได้ค้นพบว่าการประยุกต์ใช้เทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบกับปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว สามารถช่วยปรับปรุงประสิทธิผลของระบบแนะนำได้ ระหว่างทำการทดลองเพื่อค้นหาวิธีการที่ดีที่สุดสำหรับจัดการกับข้อมูลตัวอย่างที่ไม่ระบุป้ายกำกับคลาส (unlabeled example) แนวคิดการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ คือ การหาเมทริกซ์ที่มีค่าลำดับชั้นต่ำที่มีสมาชิกไม่เป็นลบ 2 เมทริกซ์ เมทริกซ์ X และเมทริกซ์ Y ซึ่งจะถูกนำมาใช้ในการประมาณค่าปฏิสัมพันธ์กันระหว่างผู้ใช้ u และไอเทม i หรือเมทริกซ์ผู้ใช้-ไอเทม R

แม้ว่าการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบถูกนำเสนอครั้งแรกในชื่อ Positive Matrix Factorization โดย Paatero และ Tapper ในบทความวิจัยของพวกเขา [21] แต่รูปแบบการแยกตัวประกอบเมทริกซ์ดังกล่าวเริ่มเป็นที่รู้จักและถูกยอมรับมากขึ้นจากการใช้เทคนิคดังกล่าวในงานวิจัยของ Lee และ Seung [22] ซึ่งนำเสนอในชื่อการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization) เพื่อประยุกต์ใช้กับปัญหาการรู้จำใบหน้าบุคคล (Face Recognition) สำหรับขั้นตอนวิธีที่ถูกใช้ในงานวิจัยดังกล่าวและเป็นที่รู้จักอย่างแพร่หลายสำหรับการหาผลเฉลยของเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ คือ ขั้นตอนวิธี Multiplicative Update Rules (MUR) ซึ่งขั้นตอนวิธี Multiplicative Update Rules นั้น ทำโดยเบื้องต้นเลือกกำหนดให้เมทริกซ์หนึ่งในสองเมทริกซ์ปัจจัยแฝง ได้แก่ เมทริกซ์ X หรือเมทริกซ์ Y เป็นเมทริกซ์ที่มีค่าคงที่ จากนั้นพยายามทำการลดให้เหลือน้อยที่สุด (minimizing) ฟังก์ชันต้นทุน (cost function) ของเมทริกซ์นั้น ด้วยค่าของเมทริกซ์ปัจจัยแฝงที่เหลืออีกเมทริกซ์หนึ่ง ขั้นตอนวิธี Multiplicative Update Rules สามารถอธิบายได้ดังภาพที่ 7

ข้อมูลนำเข้า: เมทริกซ์ผู้ใช้-ไอเทม R ขนาด m คูณ n, จำนวนปัจจัยแฝง (พารามิเตอร์ k)

1. **กำหนด** เมทริกซ์**จำนวนจริง**ไม่เป็นลบ X **ขนาด** m **คูณ** k, เมทริกซ์**จำนวนจริง**ไม่เป็นลบ Y **ขนาด** n **คูณ** k
2. สุ่มกำหนดค่าของเมทริกซ์ X และเมทริกซ์ Y
3. เลือกกำหนดให้เมทริกซ์ X หรือเมทริกซ์ Y **เป็น**เป็นเมทริกซ์ที่มีค่าคงที่
4. **กำหนด** i = 0
5. **ทำซ้ำ**

$$X_{i+1} = X_i \circ \frac{[RY_i]}{[X_i Y_i^T Y_i]}$$

$$Y_{i+1} = Y_i \circ \frac{[R^T X_{i+1}]}{[Y_i X_{i+1}^T X_{i+1}]}$$

จนกระทั่ง สำเร็จเงื่อนไขหยุดที่กำหนด

6. **ส่งออก** เมทริกซ์ X และเมทริกซ์ Y

ภาพที่ 7 ขั้นตอนวิธี MUR สำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ

นอกจาก Multiplicative Update Rules แล้ว การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบยังสามารถแก้ได้ด้วยขั้นตอนวิธี Alternating Least Squares ซึ่งแนวคิดของขั้นตอนวิธี Alternating Least Squares สำหรับโมเดลการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ โดยทั่วไป

แล้วมีคุณลักษณะแบบเดียวกันกับขั้นตอนวิธีของโมเดลการแยกตัวประกอบเมทริกซ์แบบพื้นฐาน ดังได้อธิบายรายละเอียดไปในเนื้อหาส่วนที่ผ่านมา หากแต่มีความแตกต่างกันเล็กน้อยในรายละเอียดบางประการ ทั้งนี้ขั้นตอนวิธี Alternating Least Squares ในการหาเมทริกซ์ที่มีค่าลำดับชั้นต่ำที่มีสมาชิกไม่เป็นลบ เมทริกซ์ X และเมทริกซ์ Y สามารถพบได้ในงานวิจัยของ Berry และคณะ [23] และอธิบายได้ดังภาพที่ 8

ข้อมูลนำเข้า: เมทริกซ์ผู้ใช้-ไอเทม R ขนาด m คูณ n, จำนวนปัจจัยแฝง (พารามิเตอร์ k)

1. **กำหนด** เมทริกซ์**จำนวนจริงไม่เป็นลบ** X **ขนาด** m **คูณ** k, เมทริกซ์**จำนวนจริงไม่เป็นลบ** Y **ขนาด** n **คูณ** k
2. สุ่มกำหนดค่าของเมทริกซ์ Y
3. **ทำซ้ำ**

$$X \leftarrow [RY^T(YY^T)^{-1}]_+ = [RY^{MP}]_+$$

$$Y \leftarrow [(X^TX)^{-1}X^TR]_+ = [X^{MP}R]_+$$

(เมื่อ A^{MP} **คือ** Moore-Penrose pseudo inverse)

จนกระทั่ง สำเร็จเงื่อนไขหยุดที่กำหนด

4. **ส่งออก** เมทริกซ์ X และเมทริกซ์ Y

ภาพที่ 8 ขั้นตอนวิธี ALS สำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ

ขั้นตอนวิธีสำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบทั้ง 2 แบบ สามารถขยายขอบเขตไปสู่การใช้งานกับเทคนิคการแยกตัวประกอบเมทริกซ์ที่เพิ่มการพิจารณาค่าน้ำหนักประกอบด้วยได้ ดังนั้น จากการพิจารณาบทความงานวิจัยของ Sindhwani และคณะ [4] และการพิจารณาคุณลักษณะความสามารถของการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ ผู้ดำเนินการวิจัยจึงเลือกการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ สำหรับการแก้ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวในการวิจัยนี้

แนวทางการแก้ปัญหาด้วยการพิจารณาค่าน้ำหนัก (Weighting based Approaches)

การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวในงานวิจัยที่ผ่านมานั้น มีกลยุทธ์เบื้องต้น 2 แบบในการจัดการ แบบแรก คือ การปฏิบัติเสมือนว่าตัวอย่างที่ไม่ระบุป้ายกำกับคลาส

(unlabeled example) ทั้งหมดเป็นตัวอย่างลบ ส่วนอีกแบบ คือ การปฏิบัติเสมือนว่าตัวอย่างที่ไม่ระบุป้ายกำกับคลาสเป็นตัวอย่างที่ไม่รู้ค่า (unknown) ซึ่ง Pan และคณะ [2] รับรองในบทความงานวิจัยของพวกเขาว่ากลยุทธ์แบบแรกดีกว่า สำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว โดยในงานวิจัยดังกล่าว พวกเขานิยามเมทริกซ์ค่าน้ำหนัก (weight matrix) W เพื่อช่วยในการจัดการกับข้อมูลที่ไม่รู้ค่าภายในเมทริกซ์ผู้ใช้-ไอเทม R ซึ่งรายละเอียดโดยย่อของวิธีการที่ Pan และคณะ ใช้ในงานวิจัยดังกล่าว คือ การหาเมทริกซ์ที่มีค่าลำดับชั้นต่ำ เมทริกซ์ X และเมทริกซ์ Y เพื่อประมาณค่าปฏิสัมพันธ์กันระหว่างผู้ใช้และไอเทมในเมทริกซ์ R ด้วยขั้นตอนวิธีที่มีชื่อว่า Weighted Low Rank Approximation (wLRA) ซึ่งเป็นขั้นตอนวิธีที่เพิ่มการพิจารณาค่าน้ำหนักเข้าไปร่วมด้วย ทำให้มีการเพิ่มเติมส่วนตัวแปรที่แสดงค่าน้ำหนักเข้าไปในโมเดล Regularization ดังสมการที่ 3.2 จากนั้นจึงใช้โมเดล Regularization ที่เพิ่มเติมการพิจารณาค่าน้ำหนักแล้ว กับการหาผลเฉลยของการแยกตัวประกอบเมทริกซ์ด้วยขั้นตอนวิธี Alternating Least Squares เพื่อหาเมทริกซ์ที่มีค่าลำดับชั้นต่ำ เมทริกซ์ X และเมทริกซ์ Y ในการประมาณค่าปฏิสัมพันธ์กันระหว่างผู้ใช้และไอเทมต่อไป ทั้งนี้ในกรณีของงานวิจัยของ Pan และคณะดังกล่าว ตัวแปรที่แสดงค่าน้ำหนักจะอยู่ในรูปของเมทริกซ์ค่าน้ำหนัก W ซึ่งค่าของสมาชิกแต่ละตำแหน่งภายในเมทริกซ์ W นั้น กำหนดโดยอาศัยการพิจารณาดังตารางที่ 3 [2]

ตารางที่ 3 การกำหนดเมทริกซ์ค่าน้ำหนัก W ในงานวิจัยของ Pan และคณะ

	Positive Examples	“Negative” Examples
Uniform	$W_{ij} = 1$	$W_{ij} = \delta$
User-oriented	$W_{ij} = 1$	$W_{ij} \propto \sum_j R_{ij}$
Item-oriented	$W_{ij} = 1$	$W_{ij} \propto m - \sum_j R_{ij}$

จากตารางที่ 3 การกำหนดเมทริกซ์ค่าน้ำหนัก W กระทำโดย เมื่อเป็นการพิจารณาตำแหน่งที่เป็นตัวอย่างบวก ค่าของสมาชิกตำแหน่งนั้นภายในเมทริกซ์ W จะถูกระบุให้มีค่าเท่ากับ 1 เนื่องจากข้อมูลที่เป็นตัวอย่างบวกเป็นข้อมูลที่มีความสำคัญสูงสุดต่อการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวด้วยเทคนิคการแยกตัวประกอบเมทริกซ์ ส่วนในกรณีเมื่อเป็นการพิจารณาตำแหน่งที่เป็นตัวอย่างที่ไม่ระบุป้ายกำกับคลาส และใช้กลยุทธ์การปฏิบัติเสมือนว่าตัวอย่างที่ไม่ระบุป้ายกำกับ

คลาสทั้งหมดเป็นตัวอย่างลบ จะมีวิธีทั้งหมด 3 แบบในการกำหนดค่าของสมาชิกตำแหน่งนั้นภายในเมทริกซ์ W ได้แก่ แบบแรก แบบเอกรูป (uniform) ซึ่งจะกำหนดค่าคงที่ δ ซึ่งเป็นจำนวนจริงในช่วง 0 ถึง 1 ให้แก่สมาชิกทุกตำแหน่งที่เป็นตัวอย่างที่ไม่ระบุป้ายกำกับคลาสภายในเมทริกซ์ W แบบที่ 2 คือ แบบเชิงผู้ใช้ (user-oriented) ซึ่งจะกำหนดค่าของสมาชิกตำแหน่งที่เป็นตัวอย่างที่ไม่ระบุป้ายกำกับคลาสภายในเมทริกซ์ W โดยอ้างอิงสมมติฐานที่ว่า หากผู้ใช้ u ใด ๆ มีปริมาณตัวอย่างบวกมากแล้ว ผู้ใช้ u ดังกล่าวจะมีแนวโน้มไม่ชอบไอเทมอื่น ๆ ที่เหลืออีก ส่งผลให้ตัวอย่างที่ไม่ระบุป้ายกำกับคลาสของผู้ใช้ u มีความน่าจะเป็นสูงที่จะเป็นตัวอย่างลบ และแบบสุดท้าย คือ แบบเชิงไอเทม (Item-oriented) ซึ่งจะกำหนดค่าของสมาชิกตำแหน่งที่เป็นตัวอย่างที่ไม่ระบุป้ายกำกับคลาสภายในเมทริกซ์ W โดยอ้างอิงสมมติฐานที่ว่า หากไอเทม i ใด ๆ ถูกเลือกระบุให้เป็นตัวอย่างบวกในปริมาณน้อย ไอเทม i ดังกล่าวจะมีความน่าจะเป็นสูงที่จะเป็นตัวอย่างลบ จากผลการทดลองในงานวิจัยของ Pan และคณะ [2] สรุปได้ว่า วิธีแบบที่ 3 ให้ประสิทธิผลดีที่สุด

จากบทความของ Koren และคณะ [17] ค่าน้ำหนักสามารถใช้ชี้วัดระดับความเชื่อมั่นของข้อมูลตัวอย่างความชอบของผู้ใช้แต่ละตัวที่เก็บข้อมูลมาได้ ดังนั้น การพิจารณาค่าน้ำหนักช่วยทำให้โมเดลการแยกตัวประกอบเมทริกซ์ สามารถพิจารณาความสำคัญของตัวอย่างที่ไม่ระบุป้ายกำกับคลาสแต่ละตัวในการแก้ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวได้ และสามารถทำได้โดยการเพิ่มเติมส่วนตัวแปรค่าน้ำหนัก W เข้าไปในฟังก์ชันต้นทุน (cost function) ของโมเดล Regularization ในสมการที่ 3.2 ทั้งนี้สมการของโมเดล Regularization ที่เพิ่มเติมส่วนตัวแปรที่แสดงค่าน้ำหนัก ซึ่งถูกกล่าวถึงใน [2] และ [17] สามารถแสดงได้สมการที่ 3.3

$$\min_{X,Y} \sum_{ui} W_{ui}(R_{ui}-X_u Y_i^T)^2 + (\|X_u\|_F^2+\|Y_i\|_F^2)) \qquad (3.3)$$

สำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ สมการที่ใช้ในขั้นตอนวิธี Multiplicative Update Rules และขั้นตอนวิธี Alternating Least Squares แบบพื้นฐาน สามารถปรับปรุงเพื่อให้รองรับกับแนวทางการแก้ปัญหาด้วยการพิจารณาค่าน้ำหนักได้ และขยายขอบเขตไปสู่เทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก (Weighted Non-negative Matrix Factorization) สมการสำหรับขั้นตอนวิธี Multiplicative Update Rules แบบมีการพิจารณาค่าน้ำหนัก สามารถพบได้ในบทความของ Paatero [24] ส่วนสมการสำหรับขั้นตอนวิธี Alternating Least Squares แบบมีการพิจารณาค่าน้ำหนัก สามารถพบได้ในหนังสือของ Cichocki และคณะ [25] ดังนั้นเพื่อให้สามารถอธิบายได้โดยง่าย สมการที่ใช้ในขั้นตอนวิธี

สำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่ไม่มีการพิจารณาค่าน้ำหนัก สามารถพิจารณาเป็นกรณีเฉพาะของขั้นตอนวิธีสำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนักได้ ดังแสดงในตารางที่ 4 โดย พารามิเตอร์ W คือเมทริกซ์ค่าน้ำหนักที่มีสมาชิกเป็นจำนวนจริงไม่เป็นลบ และเครื่องหมาย $\circ$ คือ ผลคูณ Hadamard (entry-wise product)

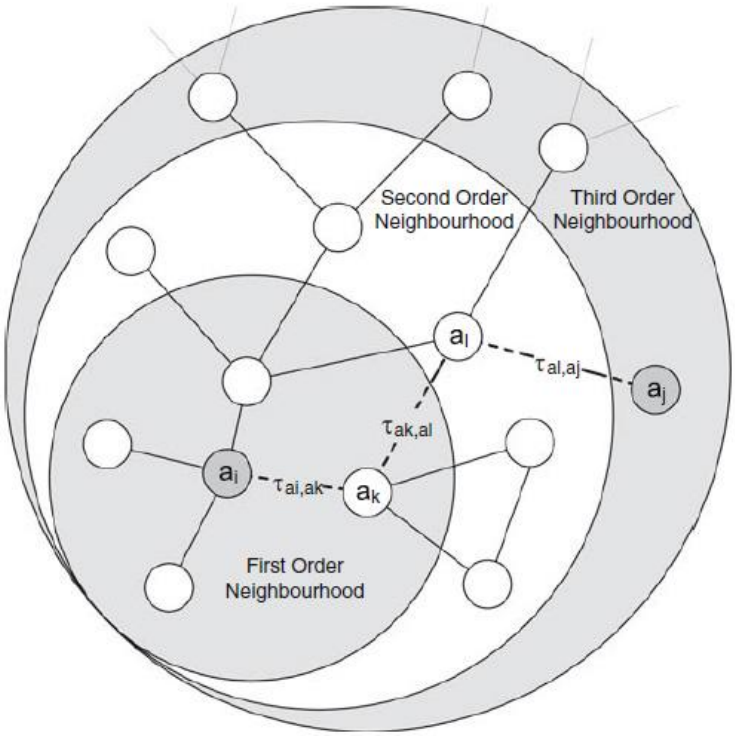
ตารางที่ 4 สมการที่ใช้ในขั้นตอนวิธีการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ เปรียบเทียบแบบที่ไม่มีการพิจารณาค่าน้ำหนักกับแบบที่มีการพิจารณาค่าน้ำหนัก

ขั้นตอนวิธี	NMF	WNMF
MUR	$Y \leftarrow Y^{\circ} \frac{[(1_{u\times i} \circ R)^T X]}{[(1_{u\times i} \circ (XY^T))^T X]}$	$Y \leftarrow Y^{\circ} \frac{[(W_1 \circ R)^T X]}{[(W_1 \circ (XY^T))^T X]}$
ALS	$Y \leftarrow [(X^T X)^{-1} X^T R]_+$	$Y \leftarrow [(X^T WX)^{-1} X^T WR]_+$

ความสัมพันธ์เชิงสังคมออนไลน์

บริการเครือข่ายสังคมออนไลน์ (Social Networking Service) ซึ่งเปิดให้บริการในรูปแบบเว็บไซต์ ตัวอย่างที่โด่งดัง เช่น Facebook หรือ Twitter กำลังได้รับความสนใจอย่างมากจากผู้คนทั่วโลก และสามารถเติบโตได้อย่างรวดเร็วในช่วงเวลาไม่กี่ปีที่ผ่านมา ด้วยปริมาณผู้ใช้บริการเครือข่ายสังคมจำนวนมหาศาล ส่งผลให้เกิดการสร้างและแบ่งปันเนื้อหาระหว่างกันเพิ่มขึ้นในอัตราที่รวดเร็ว ซึ่งเนื้อหาที่ปรากฏขึ้นภายในเครือข่ายสังคมดังกล่าวมีขอบเขตครอบคลุมข้อมูลคุณลักษณะของผู้ใช้ ข้อมูลความสนใจต่อสิ่งต่าง ๆ ของผู้ใช้ และข้อมูลความเห็นส่วนตัวของผู้ใช้ ข้อมูลเหล่านี้เป็นข้อมูลที่มีค่าและสามารถนำมาใช้เป็นข้อมูลป้อนกลับโดยปริยาย (implicit feedback) สำหรับการเรียนรู้โมเดลลักษณะความชอบของผู้ใช้ในปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวได้เป็นอย่างดี นอกจากนี้ บริการเครือข่ายสังคมออนไลน์ สามารถให้ข้อมูลความสัมพันธ์เชิงสังคมออนไลน์ ซึ่งแสดงความสัมพันธ์หรือความเป็นเพื่อนระหว่างผู้ใช้หนึ่งกับผู้ใช้อื่นบนเครือข่ายสังคมที่ผู้ใช้ทั้งสองใช้บริการร่วมกัน โดยปัจจุบันกลุ่มเพื่อนของผู้ใช้ภายในเครือข่ายสังคมออนไลน์ ส่วนมากมักอ้างอิงจากบุคคลที่รู้จักหรือเป็นเพื่อนกับผู้ใช้ในชีวิตจริง งานวิจัยจำนวนหนึ่งได้กล่าวถึงการประยุกต์ใช้งานข้อมูลเครือข่ายสังคมกับระบบแนะนำที่ใช้การกรองแบบร่วมมือกัน ซึ่งสามารถยกตัวอย่างงานวิจัยที่อธิบายรูปแบบการประยุกต์ใช้ได้อย่างชัดเจน เช่น

Walter และคณะ [18] ได้นำเสนอโมเดลของระบบแนะนำ โดยอาศัยค่าความเชื่อถือจากข้อมูลบนเครือข่ายสังคม หรือ Trust-based Recommendation ซึ่งเกิดจากแนวคิดบนสมมติฐานเรื่อง การแนะนำโดยอ้างอิงข้อมูลของผู้ใช้ที่รู้จักและเชื่อถือกันในชีวิตจริง และการพิจารณาข้อมูลกลุ่มผู้ใช้ที่แตกต่างกันในการแนะนำสิ่งที่อยู่ในโดเมนที่แตกต่างกัน ด้วยเครือข่ายสังคมออนไลน์ ระบบแนะนำสามารถให้การแนะนำสิ่งต่าง ๆ ที่หลากหลายขึ้น ซึ่งถูกเข้าถึงได้จากข้อมูลของผู้ใช้ในระดับความสัมพันธ์เชิงสังคมออนไลน์ที่ไกลออกไปจากผู้ใช้ที่ต้องการคำแนะนำ โดยอาศัยกลุ่มผู้ใช้อื่นที่ผู้ใช้ที่ต้องการคำแนะนำรู้จักและเชื่อถือ (trust relationships) ดังแสดงด้วยภาพที่ 9 [18]



ภาพที่ 9 การพิจารณากลุ่มผู้ใช้อื่นที่ผู้ใช้เป้าหมายรู้จักและเชื่อถือในงานวิจัยของ Walter และคณะ

ดังที่ได้กล่าวมาข้างต้น ข้อมูลที่ถูกแบ่งปันกันบนบริการเครือข่ายสังคมออนไลน์สามารถแสดงถึงข้อมูลป้อนกลับโดยปริยายหรือตัวอย่างบวกของผู้ใช้แต่ละคน เพื่อเรียนรู้โมเดลลักษณะความชอบของผู้ใช้ และอาจสามารถกล่าวได้เช่นกันว่า เว็บไซต์เครือข่ายสังคมอุดมไปด้วยตัวอย่างบวกของข้อมูลลักษณะความชอบของผู้ใช้ และระบบแนะนำที่ให้การแนะนำบนเว็บไซต์ประเภทเครือข่ายสังคมนี้ จึงสามารถถูกระบุได้ว่าเป็นระบบแนะนำที่ใช้การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ดังนั้นผู้ดำเนินการวิจัยจึงมีความคาดหวังที่จะใช้ประโยชน์จากคุณสมบัติของบริการเครือข่ายสังคมออนไลน์ เพื่อค้นหาวิธีการที่เหมาะสมกว่าสำหรับระบบแนะนำที่ประสบปัญหาคลาสเดี่ยว หรือต้องเรียนรู้ด้วยตัวอย่างบวกเพียงอย่างเดียว

การวิจัยนี้ได้นำเสนอวิธีการใช้ข้อมูลความสัมพันธ์เชิงสังคมออนไลน์กับปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว โดยตั้งอยู่บนสมมติฐานที่ว่าข้อมูลความสัมพันธ์เชิงสังคมออนไลน์นี้จะมีบทบาทสำคัญในการแก้ปัญหา เนื่องจากในชีวิตประจำวันคนส่วนใหญ่ มักจะมีความเชื่อถือโดยปริยาย (default trust) ต่อรสนิยมความชอบของเพื่อนหรือคนรู้จักของตนเอง ซึ่งสิ่งนี้เป็นคำตอบหลักที่การกรองแบบร่วมมือกันต้องค้นหา ดังนั้นหากมีข้อมูลของกลุ่มบุคคลที่ผู้ใช้ทุกคนเชื่อถือไว้สำหรับใช้ในการพัฒนาระบบแนะนำแล้ว คุณภาพของการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวก็อาจสามารถพัฒนาขึ้นอีกได้

บทที่ 4
การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว โดยใช้ความสัมพันธ์เชิงสังคมออนไลน์

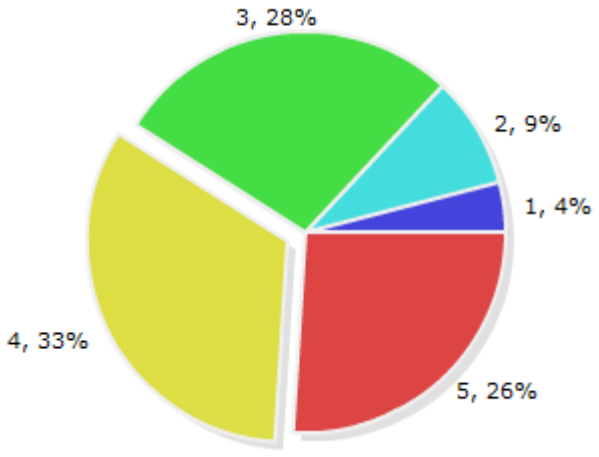
แรงจูงใจ

ปัจจุบันทั้งระบบแนะนำและเครือข่ายสังคมออนไลน์ เป็นสิ่งที่กำลังอยู่ในความสนใจเป็นอย่างมากบนโลกธุรกิจและอินเตอร์เน็ต สำหรับหัวข้อระบบแนะนำนั้น ผู้ประกอบกิจการหลายแห่งในระดับสากล ตัวอย่างที่โดดเด่นเช่น Amazon.com ให้ความสนใจและนำไปใช้ส่งเสริมกิจการของตน เนื่องจากพบว่าระบบแนะนำสามารถช่วยแนะนำสิ่งที่คาดว่าผู้ใช้หรือลูกค้าแต่ละจะสนใจได้อย่างถูกต้องแม่นยำ ทำให้ผู้ใช้เข้าถึงสินค้าของตนได้ง่ายขึ้น ส่งผลให้ผู้ประกอบกิจการดังกล่าวมียอดขายที่เพิ่มขึ้นอย่างมีนัยสำคัญ นอกจากตัวอย่างการใช้เพื่อส่งเสริมยอดขายแล้ว เว็บไซต์ต่าง ๆ ยังนำระบบแนะนำมาประยุกต์ใช้ระบบแนะนำ เพื่อตอบสนองความต้องการในการเข้าข้อมูลของผู้ใช้ภายในเว็บไซต์ของตนเช่นกัน และหากพิจารณาในมุมมองของผู้ใช้นั้น ผู้ใช้ก็ต้องการที่จะค้นพบสิ่งต่าง ๆ ภายในระบบซึ่งตนเองจะชอบหรือสนใจได้อย่างสะดวกรวดเร็ว เพื่อให้เข้าถึงข้อมูลได้มากขึ้น ค้นพบสิ่งใหม่ ๆ ได้มากขึ้น ด้วยสาเหตุเหล่านี้ ระบบแนะนำจึงเริ่มเป็นความต้องการหนึ่งของระบบต่าง ๆ ในยุคของอินเตอร์เน็ตซึ่งปริมาณมหาศาลของข้อมูลข่าวสารและการแข่งขันในเชิงธุรกิจที่สูงดังเช่นปัจจุบัน และเกิดเป็นแรงจูงใจต่อผู้ดำเนินการวิจัยที่ต้องการค้นหาวิธีการใหม่สำหรับปรับปรุงคุณภาพของเทคนิคและวิธีการของระบบแนะนำที่มีอยู่ในปัจจุบัน ให้สามารถรองรับสถานการณ์บางกรณีของระบบแนะนำ และสามารถประมวลผลเพื่อทำนายสิ่งที่ผู้ใช้แต่ละคนจะชอบได้ รวมถึงสามารถทำนายได้ถูกต้องแม่นยำยิ่งขึ้น

สำหรับหัวข้อเครือข่ายสังคมออนไลน์นั้น เป็นหัวข้อที่กำลังได้รับความนิยมและถูกกล่าวถึงมากที่สุดหัวข้อหนึ่งในปัจจุบัน ตั้งแต่การปรากฏตัวของเว็บไซต์บริการเครือข่ายสังคมออนไลน์ ชื่อ Facebook ซึ่งปัจจุบันมีจำนวนผู้ใช้ถึง 800 ล้านคนทั่วโลก ดังเนื้อหาที่ได้กล่าวถึงในบทที่แล้ว ภายในเครือข่ายสังคมออนไลน์ ประกอบด้วยข้อมูลมหาศาลที่ผู้ใช้แบ่งปันระหว่างกันสำหรับการศึกษาวิจัยต่าง ๆ แล้ว ข้อมูลเหล่านั้นมีประโยชน์เป็นอย่างมากในการศึกษาวิจัยในหัวข้อที่เกี่ยวข้องกับผู้ใช้ภายในเครือข่ายสังคม ไม่ว่าจะเป็นทางด้านคุณลักษณะ ด้านพฤติกรรม ด้านความสนใจ หรือด้านปฏิสัมพันธ์ ดังนั้นหากพิจารณาคุณสมบัติและความสามารถของเครือข่ายสังคมออนไลน์แล้ว จะพบว่ามีความเข้ากันได้และเป็นประโยชน์อย่างยิ่งต่อเทคนิควิธีการ

ของระบบแนะนำ จึงเหมาะสมที่จะพัฒนาต่อยอดเทคนิควิธีการของระบบแนะนำที่มีในปัจจุบัน ให้สามารถรองรับการใช้งานบนเครือข่ายสังคมออนไลน์ หรือสามารถใช้ประโยชน์จากข้อมูลที่มีบนเครือข่ายสังคมออนไลน์ เพื่อขยายขอบเขตความสามารถของระบบแนะนำให้มีประสิทธิผลและประสิทธิภาพมากยิ่งขึ้น

วิทยานิพนธ์นี้ ผู้ดำเนินการวิจัยให้ความสนใจในเรื่องการประมวลผลของเทคนิคสำหรับระบบแนะนำ ภายใต้สถานการณ์ที่ระบบแนะนำจำเป็นต้องเรียนรู้ลักษณะความชอบของผู้ใช้ด้วยข้อมูลตัวอย่างบวกเพียงอย่างเดียว ซึ่งเป็นสถานการณ์หนึ่งที่เกิดขึ้นได้บ่อยกับระบบแนะนำในปัจจุบัน เนื่องจากเทคนิคการกรองแบบร่วมมือกัน (Collaborative Filtering) ซึ่งเป็นเทคนิคที่นิยมใช้กับระบบแนะนำในปัจจุบัน มีข้อจำกัดหลักอยู่ 2 ประการ ได้แก่ ข้อแรก วิธีการของการกรองแบบร่วมมือกัน เรียนรู้เพื่อค้นหาคำแนะนำแก่ผู้ใช้ด้วยการใช้ค่าคะแนนความชอบที่เป็นอัตราเชิงตัวเลข (numerical rating) จากผู้ใช้เท่านั้น จึงเป็นไปไม่ได้ที่จะนำเทคนิคการกรองแบบร่วมมือกันแบบพื้นฐานดังกล่าว มาใช้กับระบบแนะนำที่ไม่ได้มีการเก็บข้อมูลค่าคะแนนความชอบที่เป็นอัตราเชิงตัวเลข ส่วนข้อจำกัดอีกประการ คือ ผู้ใช้ระบบจริงส่วนใหญ่แล้ว มักมีพฤติกรรมที่จะเพิกเฉยต่อการระบุค่าคะแนนความชอบให้แก่สิ่งที่ตนเองไม่สนใจ ดังนั้นการเก็บข้อมูลตัวอย่างลบของผู้ใช้จึงทำได้ยากหรือไม่สามารถทำได้ ภาพที่ 10 แสดงปริมาณการระบุค่าคะแนนความชอบแต่ละค่า (1 ถึง 5) ของผู้ใช้ให้แก่ภาพยนตร์ภายในเว็บไซต์บริการให้เช่าภาพยนตร์ออนไลน์ Netflix โดยข้อมูลดังกล่าวนำเสนอโดยเว็บไซต์ TimelyDevelopment.com และแสดงให้เห็นว่าปริมาณข้อมูลค่าคะแนนความชอบ 1 และ 2 ซึ่งหมายถึงผู้ใช้ไม่ชอบภาพยนตร์นั้น ๆ และเป็นตัวอย่างลบ มีปริมาณตัวอย่างข้อมูลน้อยมากเมื่อเทียบกับปริมาณข้อมูลตัวอย่างทั้งหมด



ภาพที่ 10 ปริมาณการระบุค่าคะแนนความชอบแต่ละค่าของชุดข้อมูล Netflix

ดังนั้น เพื่อรับมือกับข้อจำกัดทั้ง 2 ประการนี้ Pan และคณะ [2] จึงได้นำเสนอเทคนิคการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering) และด้วยเทคนิคดังกล่าว ระบบแนะนำไม่จำเป็นต้องใช้ค่าคะแนนความชอบที่เป็นอัตราเชิงตัวเลขหรือข้อมูลป้อนกลับที่ชัดแจ้ง (explicit feedback) แต่ระบบแนะนำจะใช้ข้อมูลป้อนกลับโดยปริยาย (implicit feedback) มาพิจารณาเป็นตัวอย่างบวก (positive example) สำหรับการเรียนรู้ด้วยเครื่องของระบบแนะนำแทน ซึ่งข้อมูลป้อนกลับโดยปริยายดังกล่าว สามารถเก็บได้ง่ายกว่าหรือมีอยู่แล้วภายในระบบ รวมถึงมีปริมาณที่มากกว่าด้วย

งานวิจัยเกี่ยวกับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวที่ผ่านมา ไม่ว่าจะเป็น Pan และคณะ [2] Sindhwani และคณะ [4] และ Li และคณะ [20] ต่างแก้ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวด้วยเทคนิคการแยกตัวประกอบเมทริกซ์ ซึ่งเป็นเทคนิคหนึ่งที่ได้รับการยอมรับว่าเหมาะสำหรับระบบแนะนำที่ใช้การกรองแบบร่วมมือกัน โดยการแยกตัวประกอบเมทริกซ์สำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว จะมีคุณลักษณะแตกต่างจากการแยกตัวประกอบเมทริกซ์สำหรับระบบแนะนำที่ใช้ข้อมูลค่าคะแนนความชอบที่เป็นอัตราเชิงตัวเลขเพียงเล็กน้อย คือ จะใช้เมทริกซ์ทวิภาค (binary matrix) ในการเก็บข้อมูลปฏิสัมพันธ์กันระหว่างผู้ใช้และไอเทม ด้วยเทคนิคการแยกตัวประกอบเมทริกซ์ Pan และคณะได้เพิ่มการพิจารณาค่าน้ำหนัก ประกอบเข้าไปเพื่อพิจารณาข้อมูลส่วนที่ไม่ระบุป้ายกำกับคลาส ทำให้ได้เทคนิคที่มีประสิทธิภาพสำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ซึ่งสามารถจำแนกข้อมูลที่คาดว่าจะเป็นตัวอย่างบวกของผู้ใช้ออกมาจากข้อมูลส่วนที่ไม่ระบุป้ายกำกับคลาสได้

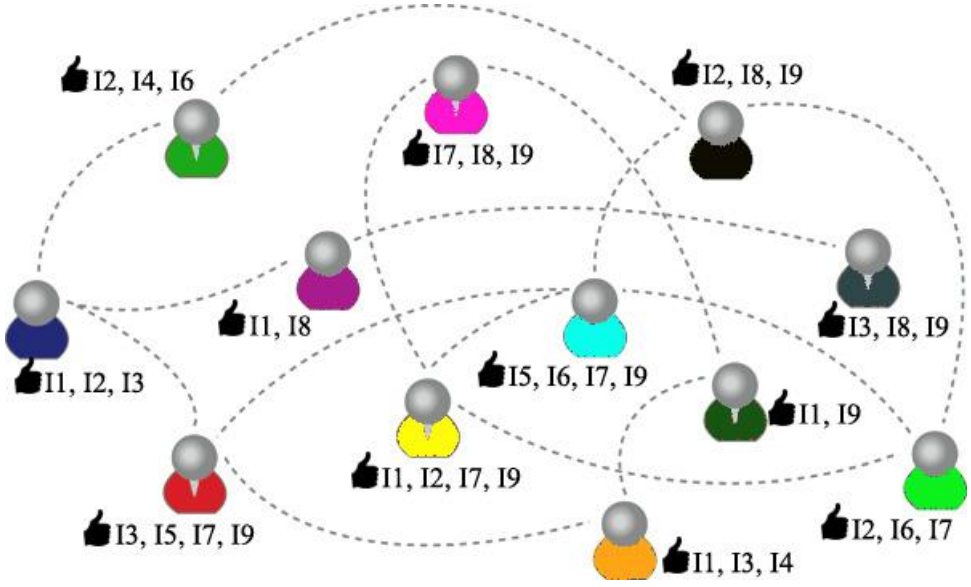
Sindhwani และคณะ [4] นำเสนอเทคนิคสำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ซึ่งใช้การแยกตัวประกอบเมทริกซ์รูปแบบหนึ่งที่มีคุณลักษณะแตกต่างจากการแยกตัวประกอบเมทริกซ์แบบพื้นฐาน และให้ประสิทธิผลสำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวที่ดีกว่าแก้ปัญหาการแยกตัวประกอบเมทริกซ์แบบพื้นฐาน เทคนิคดังกล่าว คือ การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization) แต่ทว่างานวิจัยดังกล่าวไม่ได้ใช้การพิจารณาค่าน้ำหนักใด ๆ ร่วมด้วย ในการหาผลเฉลยของเทคนิคการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว

อย่างไรก็ตาม เทคนิคการแยกตัวประกอบเมทริกซ์ สำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ในงานวิจัยที่กล่าวมาข้างต้น พิจารณาเฉพาะข้อมูลตัวอย่างบวกที่เป็นความชอบความสนใจของผู้ใช้เท่านั้น โดยไม่ได้ใช้ข้อมูลคุณลักษณะอื่น ๆ ของผู้ใช้ ตัวอย่างเช่น

ข้อมูลทางประชากรศาสตร์ หรือข้อมูลความสัมพันธ์ระหว่างกันของผู้ใช้บนเครือข่ายสังคม เป็นข้อมูลนำเข้าเพิ่มเติมสำหรับการเรียนรู้ของระบบแนะนำเลย ดังนั้นการใช้ประโยชน์จากข้อมูลคุณลักษณะของผู้ใช้ เช่น ข้อมูลความสัมพันธ์เชิงสังคมออนไลน์ จึงเป็นทางเลือกหนึ่งสำหรับการปรับปรุงคุณภาพของระบบแนะนำที่ใช้การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว

นอกจากนี้ ระบบแนะนำที่มีปัญหาเรื่องระบบจำเป็นต้องประมวลผลโดยใช้ข้อมูลที่เป็นตัวอย่างบวกเพียงอย่างเดียว หรือปัญหาที่เรียกว่าการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว สามารถพบได้เมื่อต้องการพัฒนาระบบแนะนำสำหรับใช้บนเว็บไซต์บริการเครือข่ายสังคมออนไลน์ ซึ่งเป็นบริการที่มุ่งเน้นการสร้างและสะท้อนความสัมพันธ์ระหว่างกันของผู้ใช้ทุกคน และเต็มไปด้วยข้อมูลที่เหมาะสมและจำเป็นสำหรับระบบแนะนำ

ดังนั้นผู้ดำเนินการวิจัย จึงมีแรงจูงใจที่จะศึกษาการประยุกต์ใช้ข้อมูลความสัมพันธ์ระหว่างกันของผู้ใช้บนเครือข่ายสังคมออนไลน์ หรือความสัมพันธ์เชิงสังคมออนไลน์เพื่อปรังปรุงคุณภาพของระบบแนะนำที่ใช้การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ผู้ดำเนินการวิจัยเชื่อว่าความสัมพันธ์เชิงสังคมออนไลน์บนบริการเครือข่ายสังคมที่กำลังจะได้รับความนิยมนั้น จะมีความสำคัญในการแก้ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว และนำไปสู่การพิสูจน์เรื่องการจูงใจเชิงสังคม (social influence) ว่าเป็นสิ่งที่มีผลต่อคุณภาพของระบบแนะนำที่ใช้การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ดังแสดงตัวอย่างความสัมพันธ์เชิงสังคมออนไลน์และการจูงใจเชิงสังคมด้วยภาพที่ 11



ภาพที่ 11 ความสัมพันธ์เชิงสังคมออนไลน์และการจูงใจเชิงสังคม (social influence)

แนวคิด

จากแรงจูงใจที่กล่าวมาข้างต้น การวิจัยนี้จึงมีแนวคิดที่จะนำเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ [22] มาใช้กับปัญหาของระบบแนะนำแบบการกรองแบบร่วมมือกัน ซึ่งเรียนรู้ลักษณะความชอบของผู้ใช้ ด้วยข้อมูลความชอบของผู้ใช้ที่เป็นตัวอย่างบวก หรือด้วยข้อมูลการใช้งานตอบสนองระบบของผู้ใช้ โดยระบบแนะนำที่ต้องทำงานกับข้อมูลลักษณะนี้สามารถพบได้ในสถานการณ์การใช้งานเว็บไซต์ในลักษณะต่าง ๆ เช่น การติดตามการสั่งซื้อสินค้าในเว็บไซต์ การใช้งานเว็บไซต์ให้บริการประเภทการคั่นหน้าเชิงสังคม (Social Bookmarking) เพื่อคั่นหน้าเว็บเพจที่สนใจ การติดตามประวัติการคลิกเชื่อมโยง หรือการ “Like” สิ่งที่สนใจบน Facebook เป็นต้น จากตัวอย่างเว็บไซต์ดังกล่าว พบว่าเว็บไซต์ลักษณะประเภทนี้ มักจะมีบริการเครือข่ายสังคมร่วมด้วย ผู้ดำเนินการวิจัยจึงเกิดแนวคิด ที่จะใช้ข้อมูลความสัมพันธ์บนเครือข่ายสังคม ช่วยการประมวลผลข้อมูลของระบบแนะนำแบบการกรองแบบร่วมมือกัน เพื่อให้สามารถแนะนำตัวเลือกแก่ผู้ใช้แต่ละคนได้อย่างเหมาะสมยิ่งขึ้น โดยในการวิจัยนี้ ผู้ดำเนินการต้องการทดลองแก้ไขปัญหาดังกล่าว ด้วยการใช้ชุดข้อมูลการ “Like” สิ่งที่สนใจบน Facebook โดยในกรณีของชุดข้อมูลนี้ คือ การ “Like” ต่อ Facebook Pages ของผู้ใช้ ซึ่งได้ถูกนำเสนอและรับรองด้วยงานวิจัยของ Lewis และคณะ [26] ว่า ชุดข้อมูลของเครือข่ายสังคม Facebook มีแนวโน้มที่จะสามารถแสดงความสัมพันธ์ระหว่างกันและรสนิยมความชอบของผู้ใช้ได้เป็นอย่างดี

จากแนวคิดที่ถูกนำเสนอในงานวิจัยของ Pan และคณะ [2] แก้ไขปัญหาเรียนรู้ลักษณะความชอบของผู้ใช้ ด้วยข้อมูลความชอบของผู้ใช้ซึ่งเป็นตัวอย่างบวกเพียงด้านเดียว โดยการใช้เทคนิคการแยกตัวประกอบเมทริกซ์ ทำให้ระบุได้ว่าสามารถแก้ไขปัญหาในลักษณะเดียวกันนี้ของผู้ดำเนินการวิจัย ด้วยรูปแบบวิธีการแยกตัวประกอบเมทริกซ์ที่มีการพิจารณาค่าน้ำหนัก และด้วยการศึกษางานวิจัยของ Sindhwani และคณะ [4] ทำให้พบว่าการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ (Non-negative Matrix Factorization) อาจเป็นรูปแบบหนึ่งของการแยกตัวประกอบเมทริกซ์ที่เหมาะสมกว่า สำหรับปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว อย่างไรก็ตาม ขั้นตอนการหาผลเฉลยของการแยกตัวประกอบเมทริกซ์ ใน [2] และ [4] ไม่มีการคำนึงข้อมูลคุณลักษณะของผู้ใช้หรือข้อมูลคุณลักษณะของตัวเลือกต่าง ๆ ซึ่งเป็นข้อมูลที่มีประโยชน์ในการพิจารณาของระบบแนะนำ และสามารถเก็บรวบรวมได้โดยง่ายจากผู้ใช้บนเครือข่ายสังคมออนไลน์

การวิจัยนี้ต้องการใช้ประโยชน์จากข้อมูลผู้ใช้บนเครือข่ายสังคม เพื่อการทำงานของระบบแนะนำที่ใช้การกรองแบบร่วมมือกัน ซึ่งเป็นระบบที่ให้คำแนะนำจากการเรียนรู้ลักษณะ

ความชอบของผู้ใช้ด้วยข้อมูลความชอบของผู้ใช้ซึ่งเป็นตัวอย่างบวกเพียงด้านเดียว โดยการใช้ข้อมูลคุณลักษณะของผู้ใช้ ได้แก่ ข้อมูลความสัมพันธ์กับบุคคลอื่นที่เป็นผู้ใช้เว็บไซต์เครือข่ายสังคมเดียวกัน หรือเรียกว่าความสัมพันธ์เชิงสังคมออนไลน์ ซึ่งในกรณีของ Facebook ตั้งอยู่บนสมมติฐานของบุคคลที่รู้จักกันในสังคมจริง ทำให้เหมาะสมเป็นอย่างยิ่งในการใช้พิจารณาความเชื่อมั่นที่มีต่อลักษณะความชอบของบุคคลนั้น ๆ

ระเบียบวิธีสำหรับระบบแนะนำแบบการกรองแบบร่วมมือกัน ซึ่งใช้ขั้นตอนวิธีตามเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ เพื่อแก้ปัญหาของชุดข้อมูลความชอบของผู้ใช้บริการเครือข่ายสังคมออนไลน์ Facebook ดังกล่าว มีขั้นตอนการทำงานดังนี้

1. เก็บข้อมูลความชอบของผู้ใช้ m คน ที่มีต่อตัวเลือกภายในระบบ n สิ่ง ซึ่งในกรณีของชุดข้อมูลของบริการเครือข่ายสังคม Facebook แห่งนี้ คือ การที่ผู้ใช้ “Like” ต่อ Facebook Pages ซึ่งต่อไปนี้จะขอเรียกแทน Facebook Pages ว่าไอเทม (item) โดยเก็บอยู่ในรูปของเมทริกซ์ R ซึ่ง $R = (R_{ui})_{mxn} \in \{0, 1\}^{mxn}$ สมาชิกของ R แต่ละตำแหน่งแสดงการ “Like” ผู้ใช้ u ต่อไอเทม i แทนด้วยค่า 1 ในกรณีที่มีการ “Like” และเติมค่า 0 ให้แก่ตำแหน่งที่ไม่ทราบค่า

2. จากวิธีการของเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ ระบบต้องการประมาณค่าปฏิสัมพันธ์กันระหว่างผู้ใช้ u และไอเทม i ภายในเมทริกซ์ R ด้วยเมทริกซ์ R’ ซึ่งเมทริกซ์ R’ พิจารณาจากการแยก (decomposition) $R' \approx XY^T$ เมื่อเมทริกซ์ X และเมทริกซ์ Y เป็นเมทริกซ์ตัวประกอบที่มีค่าลำดับชั้นต่ำ มีสมาชิกเป็นเซตย่อยของเซตจำนวนจริงไม่เป็นลบ และแสดงถึงปัจจัยแฝง (latent factor) ของผู้ใช้และไอเทม

3. พิจารณาข้อมูลความสัมพันธ์เชิงสังคมออนไลน์ ของผู้ใช้บริการเครือข่ายสังคมแห่งเดียวกัน ซึ่งในกรณีของ Facebook ตั้งอยู่บนสมมติฐานของบุคคลที่รู้จักกันในสังคมจริง และสะท้อนในเรื่องการจูงใจเชิงสังคม (social influence) เป็นค่าน้ำหนักในการประมาณค่าเมทริกซ์ผู้ใช้-ไอเทม R ด้วยเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ ซึ่งรายละเอียดการใช้ข้อมูลความสัมพันธ์เชิงสังคมออนไลน์กับปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวนั้น จะถูกกล่าวถึงเนื้อหาหัวข้อถัดไป

4. เรียนรู้หาผลเฉลยด้วยการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก เพื่อเป็นการหาเมทริกซ์ X และเมทริกซ์ Y ซึ่งเป็นเมทริกซ์ตัวประกอบที่มีค่าลำดับชั้นต่ำ ด้วยขั้นตอนวิธีที่เหมาะสมกับเทคนิคการแยกตัวประกอบเมทริกซ์ โดยตัวอย่าง

ขั้นตอนวิธีที่ใช้กันอย่างแพร่หลาย ได้แก่ ขั้นตอนวิธี Alternating Least Square ขั้นตอนวิธี Multiplicative Update Rules หรือขั้นตอนวิธี Rank-one Residue Iteration ซึ่งแต่ละขั้นตอนวิธีนั้นมีข้อดีข้อด้อยที่แตกต่างกัน รวมถึงสามารถปรับรูปแบบขั้นตอนวิธีดังกล่าวให้รองรับการพิจารณาค่าน้ำหนักได้ด้วย

5. พิจารณาข้อมูลภายในเมทริกซ์ค่าประมาณ R' ที่ได้ ซึ่งปรากฏแทนที่ข้อมูลที่ไม่ทราบค่าในเมทริกซ์ผู้ใช้-ไอเทม R เพื่อใช้ค่าประมาณที่ได้ในการพิจารณาแนะนำไอเทมที่ผู้ใช้เป้าหมายแต่ละคนยังไม่เคยพบและระบบทำนายว่าผู้ใช้เป้าหมายอาจจะชอบ

6. ประเมินค่าประสิทธิผลของระบบแนะนำแต่ละระบบ ด้วยตัววัดมาตรฐานสำหรับการวัดค่าความแม่นในการทำนายที่มีอยู่ในปัจจุบัน

การใช้ความสัมพันธ์เชิงสังคมออนไลน์

การกำหนดรูปแบบการใช้ข้อมูลความสัมพันธ์เชิงสังคมออนไลน์ สำหรับปรับปรุงคุณภาพของระบบแนะนำที่เป็นการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว และใช้เทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก สามารถอธิบายได้โดย กำหนดเมทริกซ์ทวิภาค $R = (R_{ui})_{mxn} \in \{0, 1\}^{mxn}$ ซึ่งสมาชิกของเมทริกซ์ผู้ใช้-ไอเทม R แต่ละตำแหน่งแสดงปฏิสัมพันธ์ระหว่างผู้ใช้และไอเทมทั้งหมดในระบบ พร้อมทั้งกำหนดเมทริกซ์ค่าน้ำหนักที่มีสมาชิกเป็นจำนวนจริงไม่เป็นลบ $W = (W_{ui})_{mxn}$ โดยระบุให้สมาชิกของเมทริกซ์ค่าน้ำหนัก W ตำแหน่งใด ๆ คือ W_{ui} มีค่าเป็น 1 เมื่อสมาชิกของเมทริกซ์ผู้ใช้-ไอเทม R ในตำแหน่งเดียวกัน คือ R_{ui} มีค่าเป็น 1 ซึ่งหมายถึงข้อมูลปฏิสัมพันธ์ระหว่างผู้ใช้และไอเทมตำแหน่งดังกล่าวเป็นตัวอย่างบวก และเป็นตัวอย่างที่มีความหมายสูงสุดต่อการเรียนรู้ลักษณะความชอบของผู้ใช้ ส่วนสมาชิกของเมทริกซ์ R ตำแหน่งอื่น ๆ ที่เป็นตัวอย่างไม่ระบุป้ายกำกับคลาสและถูกแทนค่าด้วย 0 นั้น การพิจารณาระบุค่าของสมาชิกในตำแหน่งเดียวกันสำหรับเมทริกซ์ค่าน้ำหนัก W สามารถระบุค่าน้ำหนักที่แตกต่างกันได้ โดยการพิจารณาความสัมพันธ์เชิงสังคมออนไลน์ ซึ่งค่าน้ำหนักดังกล่าวจะวัดความระดับความสำคัญและความหมายต่อการเรียนรู้ลักษณะความชอบของผู้ใช้ ด้วยตัวอย่างไม่ระบุป้ายกำกับคลาสเหล่านั้น โดยค่าน้ำหนักที่สูงหมายถึง ตัวอย่างไม่ระบุป้ายกำกับคลาสตำแหน่งดังกล่าวมีแนวโน้มที่จะเป็นตัวอย่างบวก ในทางกลับกัน ค่าน้ำหนักที่ต่ำ หมายถึงตัวอย่างไม่ระบุป้ายกำกับคลาสตำแหน่งดังกล่าวมีแนวโน้มที่จะเป็นตัวอย่างลบ

การระบุค่าน้ำหนักแต่ละตำแหน่ง ด้วยการใช้ความสัมพันธ์เชิงสังคมออนไลน์ ในกรณีของการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว กำหนดโดยสมมติฐานที่ว่า หากผู้ใช้เป้าหมาย u ที่ระบบกำลังพิจารณา (target user) ใด ๆ มีจำนวนเพื่อนที่ชอบหรือสนใจในไอเทม i เดียวกันเป็นจำนวนมาก ผู้ใช้เป้าหมาย u ที่ระบบกำลังพิจารณานั้นก็จะมีแนวโน้มที่จะสนใจหรือชอบไอเทม i ดังกล่าวนั้นด้วย หรือเรียกได้ว่าเกิดจากการจูงใจเชิงสังคม (social influence) ดังที่ได้กล่าวไปแล้วในเนื้อหาส่วนที่ผ่านมา ดังนั้นตัวอย่างไม่ระบุป้ายกำกับคลาสตำแหน่งของผู้ใช้เป้าหมาย u และไอเทม i ดังกล่าว มีความน่าเป็นสูงที่จะเป็นตัวอย่างบวก สำหรับการระบุค่าน้ำหนักแต่ละตำแหน่งในเมทริกซ์ค่าน้ำหนัก W ด้วยการใช้ความสัมพันธ์เชิงสังคมออนไลน์ สามารถสรุปได้ดังตารางที่ 5

ตารางที่ 5 การระบุค่าน้ำหนักโดยใช้ความสัมพันธ์เชิงสังคมออนไลน์

ตัวอย่างบวก (ใน R_{ui})	**ตัวอย่างที่ไม่ระบุป้ายกำกับคลาส** (ใน R_{ui})
$W_{ui}=1$	$W_{ui}=\frac{\text{จำนวนเพื่อนของผู้ใช้เป้าหมายที่ชอบไอเทม i}}{\text{จำนวนเพื่อนทั้งหมดของผู้ใช้เป้าหมาย}}$

ดังนั้น เมื่อได้รูปแบบการระบุค่าน้ำหนักโดยใช้ความสัมพันธ์เชิงสังคมออนไลน์แล้ว กระบวนการประมวลผลของระบบแนะนำเพื่อให้คำแนะนำในการวิจัยนี้ สามารถอธิบายขั้นตอนวิธีได้ดังภาพที่ 12

ข้อมูลนำเข้า: เมทริกซ์ผู้ใช้-ไอเทม R ขนาด m ผู้ใช้ คูณ n ไอเทม, เมทริกซ์ค่าน้ำหนักที่มีสมาชิกเป็นจำนวนจริงไม่เป็นลบ W, จำนวนปัจจัยแฝง (พารามิเตอร์ k)

ข้อมูลส่งออก: เมทริกซ์สำหรับให้คำแนะนำ A

1. กำหนด เมทริกซ์จำนวนจริงไม่เป็นลบ X ขนาด m คูณ k, เมทริกซ์จำนวนจริงไม่เป็นลบ Y ขนาด n คูณ k
2. ใช้การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก (WNMF) เพื่อหาผลเฉลยของเมทริกซ์ X และเมทริกซ์ Y ด้วยการหาค่าเหมาะที่สุดเพื่อการลดให้เหลือน้อยที่สุด (minimizing) สมการที่ 3.3
3. ประมาณค่าปฏิสัมพันธ์ระหว่างผู้ใช้และไอเทมในรูปเมทริกซ์ R' ด้วยเมทริกซ์ที่มีค่าลำดับชั้นต่ำ (low rank matrix) เมทริกซ์ X และเมทริกซ์ Y
4. สร้างเมทริกซ์สำหรับให้คำแนะนำ A จากเมทริกซ์ R' โดย
   - เมื่อสมาชิกเมทริกซ์ใน R' มีค่ามากกว่าขีดแบ่งที่กำหนดสมาชิกตำแหน่งนั้นในเมทริกซ์ A จะมีค่า 1 มิฉะนั้นมีค่า 0
   - แนะนำไอเทม i ให้แก่ผู้ใช้ u เมื่อ $A_{ui} = 1$ และ $R_{ui} = 0$

ภาพที่ 12 ขั้นตอนวิธีสำหรับการให้คำแนะนำของระบบแนะนำ โดยใช้การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก

บทที่ 5

ผลการทดลอง

การวิจัยได้ดำเนินการทดลองเพื่อทดสอบประสิทธิผลของเทคนิคและสมมติฐานต่าง ๆ บนโปรแกรม MATLAB เป็นหลัก ซึ่งเป็นโปรแกรมที่มีชุดคำสั่งสำเร็จสำหรับการคำนวณทางด้านคณิตศาสตร์มากมาย รวมทั้งชุดคำสั่งสำเร็จสำหรับกลุ่มขั้นตอนวิธีการแยกตัวประกอบเมทริกซ์ด้วย พร้อมกันนี้ผู้ดำเนินการวิจัยได้พัฒนาโปรแกรมด้วยภาษา Python เพื่อใช้สำหรับการเก็บรวบรวมชุดข้อมูลและการประเมินค่าประสิทธิผลของระบบแนะนำ การทดลองนี้พิจารณาทดสอบเปรียบเทียบความแม่นในการทำนายของระบบแนะนำเป็นสำคัญ

ชุดข้อมูล

ชุดข้อมูลที่ผู้ดำเนินการวิจัยใช้ในการทดลองเพื่อทดสอบประสิทธิผลของเทคนิคและสมมติฐานต่าง ๆ เป็นชุดข้อมูลที่แสดงถึงความชอบหรือความสนใจของผู้ใช้จริง บนเว็บไซต์บริการเครือข่ายสังคมออนไลน์ ซึ่งในการวิจัยนี้ คือ Facebook (http://www.facebook.com) โดยผู้ดำเนินการวิจัยได้พัฒนาชุดคำสั่งเพื่อเก็บรวบรวมข้อมูลการกดปุ่ม “ถูกใจ” หรือ “Like” ของผู้ใช้ต่อ “Facebook Page” หน้าหนึ่ง ๆ ซึ่งลักษณะชอบพฤติกรรมดังกล่าวสามารถแสดงถึงความชอบหรือความสนใจของผู้ใช้ได้ โดยการกดปุ่ม Like ต่อ Facebook Page นี้อาจถือได้ว่าเป็นการทำการคั่นหน้าเชิงสังคม (Social Bookmarking) และเป็นข้อมูลที่เป็นตัวอย่างบวกเพียงอย่างเดียวได้ อนึ่ง Facebook เป็นเว็บไซต์บริการเครือข่ายสังคมออนไลน์ที่ขยายใหญ่และเป็นที่รู้จักอย่างแพร่หลายมากที่สุดในโลก ด้วยปริมาณสมาชิกในปัจจุบันที่ 800 ล้านคน ภายใน Facebook ผู้ใช้สามารถสร้างข้อมูลคุณลักษณะของผู้ใช้ (user profile) เพิ่มผู้ใช้อื่นเป็นเพื่อนในเครือข่ายสังคม และแบ่งปันสิ่งต่าง ๆ ที่ผู้ใช้สนใจ จากงานวิจัยของ Lewis และคณะ [26] ชุดข้อมูลที่ได้จาก Facebook มีแนวโน้มที่จะสะท้อนถึงความชอบหรือรสนิยมของผู้ใช้และความสัมพันธ์เชิงสังคมระหว่างกันของผู้ใช้ได้เป็นอย่างดี

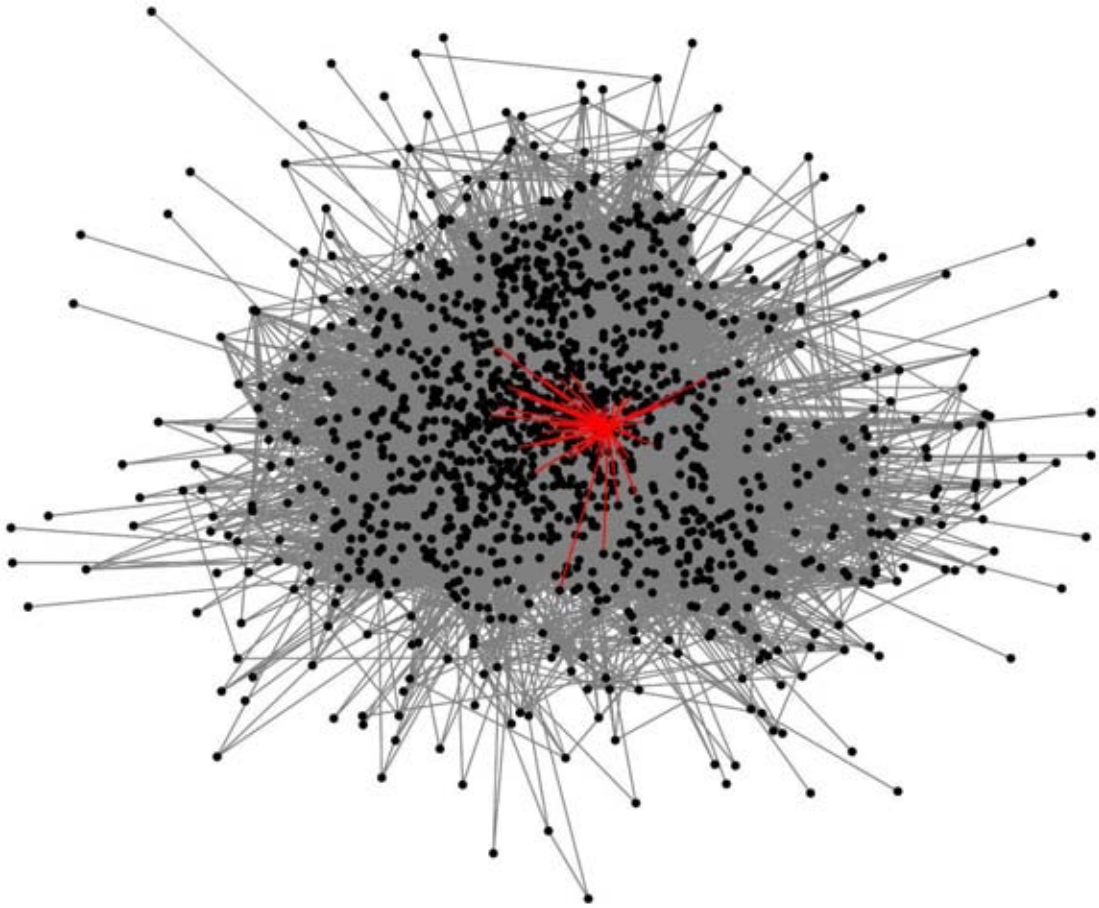
ชุดข้อมูลในการวิจัยนี้ประกอบด้วยข้อมูลการกดปุ่ม Like ต่อ Facebook Page จำนวน 39,421 ตัวอย่าง ซึ่งเป็นข้อมูลที่เกิดจากผู้ใช้ 394 คน และ Facebook Page 500 หน้า ทั้งนี้ในการวิจัยจำเป็นต้องให้ความสำคัญกับเรื่องภาวะส่วนตัว (privacy) ของผู้ใช้ทุกคน ชุดข้อมูลดังกล่าวจึงไม่มีการเก็บข้อมูลชื่อและนามสกุลจริงของผู้ใช้ใด ๆ ในชุดข้อมูล รวมถึงไม่ได้ใช้ข้อมูลชื่อและนามสกุลดังกล่าวในส่วนใด ๆ ของการทดลองด้วย ชุดข้อมูลดังกล่าว เบื้องต้นถูกบรรจุลง

ในรูปแบบ JSON จากนั้นจึงถูกจัดรูปแบบใหม่ให้สามารถอ่านและนำไปประมวลผลต่อได้ง่ายยิ่งขึ้น ตัวอย่างข้อมูลบางส่วนภายในชุดข้อมูลของการวิจัยนี้สามารถแสดงได้ดังภาพที่ 13 ซึ่งข้อมูลแต่ละบรรทัด คือ ตัวอย่างข้อมูลการกดปุ่ม Like ต่อ Facebook Page 1 ตัวอย่าง โดยระบุถึงรหัสแทนผู้ใช้ (ปกปิดข้อมูลรหัสแทนผู้ใช้จริงบน Facebook ด้วยจำนวนนับ) รหัสแทน Facebook Page ประเภทของ Facebook Page และวันเวลาที่ผู้ใช้ดังกล่าวกดปุ่ม Like ต่อ Facebook Page หน้านั้น ๆ ตามลำดับ

```
1,159311710749765,Local business,2010-09-18T22:57:34+0000
1,146764688688623,Community,2010-08-31T14:51:57+0000
1,17430971715,Company,2010-05-04T20:46:47+0000
1,220613901897,Local business,2010-05-04T20:46:47+0000
1,298407696340,Appliances,2010-04-24T07:52:59+0000
2,126725424056960,Community,2011-02-23T09:08:00+0000
2,102389789801838,Club,2011-02-15T04:49:45+0000
2,230133619459,Company,2011-02-01T08:05:30+0000
2,54651330333,Local business,2011-01-25T08:42:30+0000
2,154506381255384,Website,2011-01-15T15:34:22+0000
2,244317796730,Website,2011-01-06T08:02:09+0000
2,291549705119,Application,2011-01-04T06:29:11+0000
2,115930295140675,Product/service,2010-12-07T14:29:25+0000
2,141930619169289,News/media,2010-12-02T10:34:13+0000
2,223973546874,Website,2010-11-11T13:35:15+0000
2,123413162318,Product/service,2010-10-15T02:04:58+0000
2,149364350027,Local business,2010-10-04T04:33:26+0000
2,142765199078274,Public figure,2010-10-04T04:31:45+0000
2,125236617497014,Local business,2010-09-30T10:27:11+0000
2,131393086891220,Company,2010-09-30T06:22:12+0000
2,104779379576831,Product/service,2010-09-28T13:23:13+0000
```

ภาพที่ 13 ตัวอย่างข้อมูลการกดปุ่ม Like ต่อ Facebook Pageบางส่วนภายในชุดข้อมูล

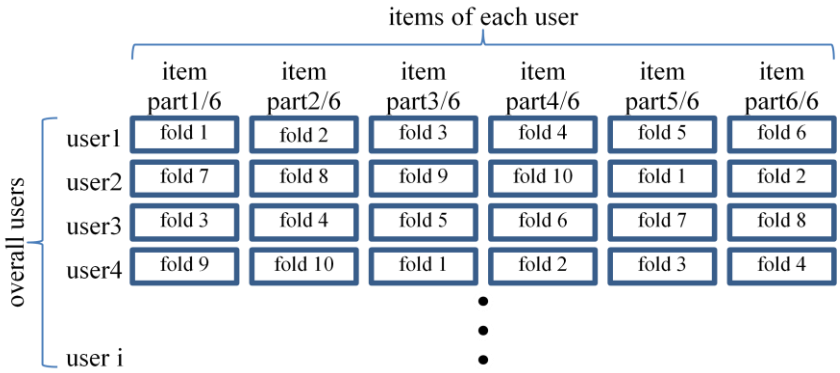
นอกจากข้อมูลการกดปุ่ม Like ต่อ Facebook Page ของผู้ใช้แล้ว ชุดข้อมูลยังประกอบด้วยข้อมูลความสัมพันธ์เชิงสังคมหรือข้อมูลความเป็นเพื่อนของผู้ใช้ดังกล่าวเหล่านั้นด้วย ซึ่งข้อมูลความสัมพันธ์เชิงสังคมออนไลน์ของผู้ใช้ภายในชุดข้อมูลมีทั้งสิ้น 49,739 ความสัมพันธ์ และสามารถแสดงความสัมพันธ์เชิงสังคมดังกล่าวในรูปแบบกราฟเครือข่าย ซึ่งวาดด้วยโปรแกรม NodeXL (http://nodexl.codeplex.com) และขั้นตอนวิธี Fruchterman-Reingold ได้ดังภาพที่ 14



ภาพที่ 14 ความสัมพันธ์เชิงสังคมหรือความเป็นเพื่อนของชุดข้อมูล Facebook ในการทดลอง

การทดลองเพื่อทดสอบประสิทธิผลของเทคนิคและสมมติฐานต่าง ๆ จะใช้วิธีการตรวจสอบไขว้ 10 พับ (10-fold Cross Validation) ในการทดสอบเปรียบเทียบความแม่นในการทำนายเพื่อการแนะนำผู้ใช้แต่ละคน เนื่องด้วยปริมาณตัวอย่างของผู้ใช้แต่ละคนมีจำนวนมากน้อยไม่เท่ากันและด้วยรูปแบบของการทดสอบประสิทธิผลของระบบแนะนำดังกล่าว จึงต้องทำการแบ่งข้อมูลตัวอย่างของผู้ใช้แต่ละคนในชุดข้อมูลออกเป็นส่วน ๆ เท่า ๆ กัน และกระจายไปสู่แต่พับ (fold) ซึ่งในการทดลองนี้แบ่งเป็น 6 ส่วน ทำให้เกิดการทดสอบเปรียบเทียบความแม่นในการ

ทำนายเพื่อการแนะนำผู้ใช้แต่ละคนที่เหมาะสมกว่า สามารถวัดประสิทธิผลของเทคนิคและสมมติฐานต่าง ๆ ได้ดียิ่งขึ้น ทั้งนี้สามารถอธิบายรูปแบบการแบ่งพับของชุดข้อมูลได้ดังภาพที่ 15



ภาพที่ 15 การแบ่งพับของชุดข้อมูลในการทดลอง

การทดลองและผลการทดลอง

การทดลองส่วนสำคัญดำเนินการโดยทดสอบแนวคิดที่นำเสนอกับเทคนิคการแยกตัวประกอบเมทริกซ์ 3 เทคนิค ได้แก่ การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบแบบพื้นฐานโดยใช้ขั้นตอนวิธี Multiplicative Update Rules (MUR) การแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบแบบพื้นฐานโดยใช้ขั้นตอนวิธี Alternating Least Squares (ALS) และขั้นตอนวิธี Weighted Low Rank Approximation [3] ซึ่งเทคนิคทั้ง 3 ไม่ได้นำความสัมพันธ์เชิงสังคมออนไลน์มาใช้ช่วยในการพิจารณา ส่วนเทคนิคที่นำเสนอในการวิจัยนี้ ซึ่งใช้ความสัมพันธ์เชิงสังคมออนไลน์ จะถูกทดสอบด้วยขั้นตอนวิธีทั้ง MUR และ ALS เพื่อเปรียบเทียบเช่นกัน สำหรับการทดลองย่อยอื่น ๆ นั้นกระทำเพื่อเปรียบเทียบแนวคิดที่นำเสนอกับเทคนิคสำหรับการกรองแบบร่วมมือกัน และการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวโดยทั่วไป ตัวอย่างเช่น ขั้นตอนวิธี K-Nearest-Neighbor (k-NN) หรือ การแยกตัวประกอบเมทริกซ์รูปแบบพื้นฐานอื่น ๆ เพื่อให้เกิดแนวคิดในการเลือกหาเทคนิคที่เหมาะสมกับปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวบนเครือข่ายสังคมออนไลน์ ขั้นตอนวิธี k-NN เป็นขั้นตอนวิธีพื้นฐานที่สุดสำหรับการกรองแบบร่วมมือกันด้วยเทคนิคประเภท Memory-based ในการทดลองย่อยนี้ ขั้นตอนวิธี k-NN ใช้การวัดระยะทางแบบยุคลิด (Euclidean distance) กับคุณลักษณะของผู้ใช้และคุณลักษณะของไอเทม เพื่อพิจารณาหาผู้ใช้อื่น k คน ที่มีลักษณะความชอบเหมือนหรือใกล้เคียงกับผู้ใช้เป้าหมาย ผลลัพธ์จากการทดลองย่อยดังกล่าวสามารถแสดงได้ดังตารางที่ 6 และตารางที่ 7

ตารางที่ 6 ผลลัพธ์ค่าประสิทธิผลของระบบแนะนำที่ใช้เทคนิค K-Nearest-Neighbor

k (จำนวนผู้ใช้ที่ลักษณะความชอบใกล้เคียงที่สุด k คนแรก)	ค่าความเที่ยง (Precision)
1	0.45
5	0.54
10	0.53

ตารางที่ 7 ผลลัพธ์ค่าประสิทธิผลของระบบแนะนำที่ใช้เทคนิคการแยกตัวประกอบเมทริกซ์

		จำนวนปัจจัยแฝง (latent factor)				
		10	20	30	40	50
ค่าความเที่ยง (Precision)	NMF	0.19	0.41	0.54	0.60	0.61
	WNMF	0.20	0.38	0.59	0.67	0.70

จากผลการทดลองในตารางที่ 6 และตารางที่ 7 ซึ่งใช้การพิจารณาด้วยตัววัดเชิงสนับสนุนการตัดสินใจพื้นฐาน คือ ค่าความเที่ยง (Precision) จะเห็นได้ว่าเทคนิคการแยกตัวประกอบเมทริกซ์มีความเหมาะสมกว่าเทคนิคขั้นตอนวิธี k-NN ในการแก้ปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว โดยสาเหตุที่ทำให้ขั้นตอนวิธี k-NN ไม่เหมาะกับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวและชุดข้อมูลในการวิจัยนี้ สามารถสรุปได้ดังต่อไปนี้

1. ชุดข้อมูลในการวิจัยนี้มีปัญหาข้อมูลในระบบมีปริมาณเบาบาง (Data Sparsity) ซึ่งมีตัวอย่างการกดปุ่ม Like ต่อ Facebook Page ของผู้ใช้บางคนน้อย หรือไอเทม Facebook Page บางหน้ามีผู้ใช้กดปุ่ม Like น้อยมาก

2. ระบบแนะนำจะประมวลผลและทำนายได้ดี เฉพาะเมื่อเป็นไอเทม Facebook Page ที่ได้รับความนิยมสูง ซึ่งมีปริมาณข้อมูลตัวอย่างมาก

3. การกำหนดวิธีพิจารณาคุณลักษณะของไอเทม ยังไม่สามารถทำได้อย่างเหมาะสมพอ ตัวอย่างเช่น ประเภทของ Facebook Page ซึ่งมีหน้าที่ถูกระบุประเภทมากกว่า 1 ประเภท หรือความขัดแย้งในการเลือกระบุประเภท

4. ชุดข้อมูลในการวิจัยนี้ เป็นข้อมูลตัวอย่างบวกเท่านั้น

5. ไม่มีการใช้ประโยชน์จากข้อมูลความสัมพันธ์เชิงสังคมออนไลน์

การประเมินค่าประสิทธิผลของระบบแนะนำแต่ละระบบ สามารถพิจารณาโดยใช้ตัววัดมาตรฐานสำหรับการวัดค่าความแม่นในการทำนายที่มีอยู่ในปัจจุบัน ในกรณีของระบบแนะนำแบบการกรองแบบร่วมมือกันโดยทั่วไป ตัววัดที่ใช้จะมี 2 ประเภท ประเภทแรก คือ ตัววัดเชิงสถิติ ตัวอย่างเช่น Root Mean Square Error (RMSE) และ Mean Absolute Error (MAE) ส่วนประเภทที่สอง คือ ตัววัดเชิงสนับสนุนการตัดสินใจ ตัวอย่างเช่น Precision และ Recall สำหรับกรณีของระบบแนะนำที่เรียนรู้ลักษณะความชอบของผู้ใช้ โดยใช้เฉพาะข้อมูลความชอบของผู้ใช้ที่เป็นตัวอย่างบวก ตัววัดประเภทแรกจะไม่สามารถใช้ในการประเมินค่าประสิทธิผลของระบบแนะนำได้ การวิจัยนี้การประเมินค่าประสิทธิผลของระบบ จึงสามารถทำได้โดยใช้ตัววัดมาตรฐาน 2 ตัววัด ได้แก่ Mean Average Precision และ Mean Percentage Ranking

Mean Average Precision (MAP) เป็นตัววัดที่ประเมินประสิทธิผลของระบบทั้งหมดโดยอาศัยค่าความเที่ยง (Precision) ที่ค่า Recall ระดับต่าง ๆ จากการประเมินบนเซตทดสอบ การคำนวณค่า MAP กระทำโดยการหาค่า Average Precision (AP) ของผู้ใช้ทุกคนที่อยู่ในเซตทดสอบ ซึ่งค่า AP ของผู้ใช้ u ใด ๆ จะถูกคำนวณจากทุกตำแหน่งของไอเทมที่ i ซึ่งถูกระบบแนะนำให้แก่ผู้ใช้ u และผู้ใช้ u ระบุว่าชอบ การคำนวณ MAP สามารถแสดงได้ดังสมการที่ 5.1

$$MAP = \frac{1}{\text{จำนวนผู้ใช้ทั้งหมด}} \times \sum_{u=1}^{U} \left(\frac{\sum_{i=1}^{N} prec(i) \times pref(i)}{N_p} \right) \qquad (5.1)$$

เมื่อ prec(i) แทน ค่าความเที่ยง (Precision) (เศษส่วนของไอเทมที่แนะนำโดยระบบซึ่งถูกผู้ใช้ระบุว่าชอบ) ของไอเทมที่ i ซึ่งถูกระบบแนะนำให้แก่ผู้ใช้ u

pref(i) แทน ค่าฟังก์ชันทวิภาค (มีค่า 1 เมื่อไอเทมที่ i ถูกผู้ใช้ u เลือกระบุว่าชอบ มิฉะนั้นมีค่า 0)

U แทน จำนวนผู้ใช้ทั้งหมด

N แทน จำนวนไอเทมที่ระบบแนะนำทั้งหมด

N_p แทน จำนวนไอเทมที่ผู้ใช้เลือกระบุว่าชอบ

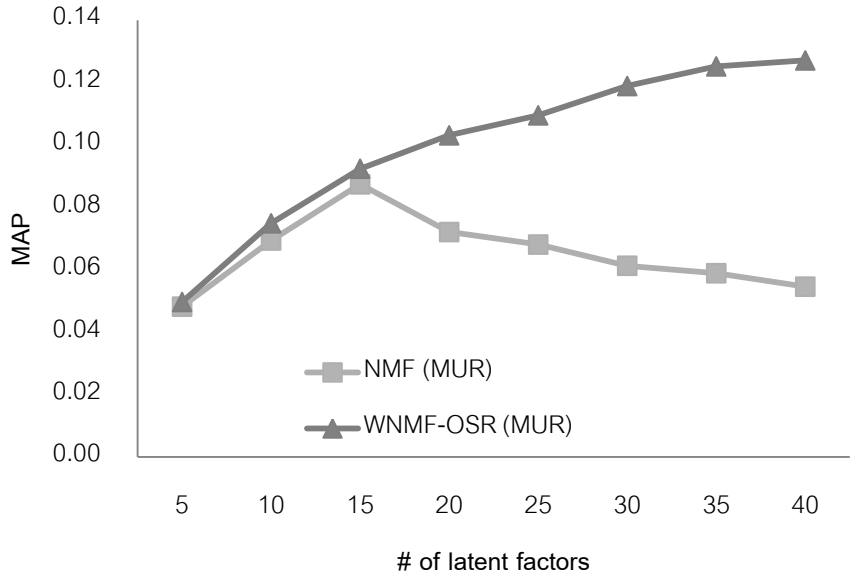
Mean Percentage Ranking (MPR) เป็นตัววัดแบบ Recall-oriented แบบหนึ่ง ซึ่งถูกใช้ในงานวิจัยของ Koren และคณะ [19] และงานวิจัยของ Li และคณะ [20] โดยเนื่องจากคุณลักษณะของปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ที่ไม่สามารถเชื่อมั่นในข้อมูลตัวอย่างบวกของความชอบของผู้ใช้ได้อย่างเต็มที่ กล่าวคือไอเทม i ใด ๆ ที่ไม่ได้เลือกระบุว่าชอบอาจไม่ได้หมายถึงผู้ใช้ไม่ได้สนใจไอเทม i นั้น หากแต่เกิดจากการที่ผู้ใช้ไม่เคยรู้จักพบเห็นไอเทม i มาก่อน ส่งผลให้ไม่สามารถหาค่า False Negative ที่ถูกต้อง ซึ่งต้องใช้ในการคำนวณของตัววัด Recall แบบพื้นฐานได้ ดังนั้นการวิจัยนี้จึงเลือกใช้ตัววัด MPR แทนตัววัด Recall แบบพื้นฐาน MPR จะประเมินความพอใจต่อตัวเลือกไอเทมที่ถูกเรียงลำดับ วัตถุประสงค์ของตัววัดชนิดนี้ คือ การค้นหาลำดับของตัวเลือกที่ระบบแนะนำให้แก่ผู้ใช้ได้อย่างถูกต้อง ภายในตัวเลือกไอเทมทั้งหมดที่ถูกเรียงลำดับด้วยค่าแสดงระดับความชอบที่ระบบทำนายแก่ผู้ใช้แต่ละคน สำหรับวิธีการคำนวณค่า MPR ในทางปฏิบัติเพื่อประเมินประสิทธิผลของระบบนั้น กระทำโดยสุ่มไอเทมที่มาจำนวนหนึ่ง ตัวอย่างเช่น 1000 ไอเทม และทำการทำนายค่าระดับความชอบของผู้ใช้ต่อไอเทมเหล่านั้น รวมกับคู่ปฏิสัมพันธ์ระหว่างผู้ใช้กับตัวเลือกไอเทมจริง i ที่สนใจ จะได้เป็นตัวเลือกไอเทมที่ถูกเรียงลำดับที่เรียงค่าทำนายระดับความชอบจากมากไปน้อยทั้งหมด 1001 ตัวเลือก และทำให้มีลำดับที่เป็นไปได้สำหรับตัวเลือกไอเทมจริง i ที่สนใจ ทั้งหมด 1001 ลำดับ ในการพิจารณา 0% เป็นกรณีซึ่งดีที่สุด หมายถึง ตัวเลือกไอเทม i ถูกผู้ใช้ u ให้ความสนใจมากที่สุด หรือกล่าวได้ว่าไม่มีไอเทมอื่นใดถูกระบบแนะนำในลำดับก่อนตัวเลือกไอเทม i ขณะที่ 100% เป็นกรณีตรงกันข้ามซึ่งแย่ที่สุด การคำนวณ MPR สามารถแสดงได้ดังสมการที่ 5.2 MPR ควรมีค่าน้อยกว่า 50% เนื่องจากหาก MPR มีค่ามากกว่าหรือเท่ากับ 50% จะหมายถึงการแนะนำตัวเลือกดังกล่าวเกิดจากการสุ่ม

$$MPR = \frac{\sum_{u,i} r_{ui} \times rank_{ui}}{\sum_{u,i} r_{ui}} \qquad (5.2)$$

เมื่อ r_{ui} แทน ค่าฟังก์ชันทวิภาค (มีค่า 1 เมื่อไอเทมที่ i ถูกผู้ใช้ u เลือกระบุว่าชอบ มิฉะนั้นมีค่า 0)

$rank_{ui}$ แทน ค่าลำดับชั้นเปอร์เซ็นไทล์ของไอเทมที่ i ซึ่งระบบแนะนำให้แก่ผู้ใช้ u (หากมีค่า 0% หมายถึง ไอเทมที่ i ถูกผู้ใช้ u ให้ความสนใจมากที่สุด หรือกล่าวได้ว่า ไม่มีไอเทมอื่นใดถูกระบบแนะนำในลำดับก่อนไอเทม i)

พารามิเตอร์หนึ่งที่จำเป็นต้องพิจารณาของเทคนิคการแยกตัวประกอบเมทริกซ์ คือ จำนวนปัจจัยแฝง (latent factor) หรือ พารามิเตอร์ k ซึ่งเป็นค่าแสดงมิติของเมทริกซ์ที่มีค่าลำดับชั้นต่ำในการประมาณค่าปฏิสัมพันธ์กันระหว่างผู้ใช้และไอเทม ดังรายละเอียดในเนื้อหาบทที่ 3 จากการทดลองเบื้องต้นเพื่อหาค่าพารามิเตอร์ k ที่เหมาะสมสำหรับการใช้ในการทดสอบแนวคิด ดังแสดงด้วยกราฟในภาพที่ 16 ซึ่งแสดงการใช้จำนวนปัจจัยแฝงแต่ละค่าส่งผลที่แตกต่างกันต่อค่า MAP บนการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบ และการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก โดยการทดลองเบื้องต้นเพื่อหาค่าพารามิเตอร์ k ที่เหมาะสมนั้น ทดสอบด้วยขั้นตอนวิธี MUR และทดสอบกับค่าพารามิเตอร์ k ในช่วงตั้งแต่ 5 ถึง 40



ภาพที่ 16 จำนวนปัจจัยแฝงส่งผลที่แตกต่างกันต่อค่า Mean Average Precision

จากการทดลองเบื้องต้นเพื่อพิจารณาเลือกค่าพารามิเตอร์ k ที่เหมาะสม พบว่าประสิทธิผลของระบบแนะนำเพิ่มขึ้นเมื่อจำนวนปัจจัยแฝงเพิ่มขึ้น หรือกล่าวคือเมื่อโมเดลปัจจัยแฝงมีความซับซ้อนขึ้นจะส่งผลให้ความแม่นในการทำนายของระบบแนะนำเพิ่มขึ้น อย่างไรก็ตาม เมื่อทดสอบค่าพารามิเตอร์ k ไปถึงจุดหนึ่ง ค่า MAP จะมีแนวโน้มลู่เข้าหรือปรับลดลงขึ้นกับขั้นตอนวิธีที่ใช้ส่งผลให้ต้องพิจารณาเลือกค่าพารามิเตอร์ k ที่เหมาะสม ซึ่งอยู่ในช่วงที่ทำให้ประสิทธิผลของระบบแนะนำเพิ่มขึ้นอย่างมีนัยสำคัญและเริ่มลู่เข้าสู่ค่าคงที่ ทั้งนี้การเลือกค่าพารามิเตอร์ k ที่เหมาะสม จะช่วยให้ประสิทธิภาพของระบบแนะนำมีแนวโน้มที่ดี เนื่องจากระบบจะใช้เวลาในการคำนวณเพิ่มขึ้นในขั้นตอนของเทคนิคการแยกตัวประกอบเมทริกซ์ เมื่อกำหนดค่าพารามิเตอร์ k มากขึ้น การวิจัยนี้พิจารณาใช้ค่าพารามิเตอร์ k หรือจำนวนปัจจัยแฝง เท่ากับ 15 สำหรับการ

แยกตัวประกอบเมทริกซ์แบบไม่เป็นลบพื้นฐาน และ 40 สำหรับการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก

การทดลองส่วนสำคัญดำเนินการโดยทดสอบเทคนิคที่นำเสนอกับเทคนิคการแยกตัวประกอบเมทริกซ์ 3 เทคนิค ทั้งนี้เพื่อควบคุมประสิทธิภาพของระบบแนะนำ จึงต้องจำกัดเวลาในการคำนวณของระบบ เพื่อไม่ให้ใช้เวลาประมวลผลนานจนเกินไป ด้วยการกำหนดจำนวนรอบของการทำซ้ำขั้นตอนในขั้นตอนวิธีของเทคนิคการแยกตัวประกอบเมทริกซ์ โดยในการทดลองนี้กำหนดจำนวนรอบของการทำซ้ำสำหรับทุก ๆ เทคนิคไว้ที่ 100,000 รอบ ซึ่งจำนวนรอบของการทำซ้ำดังกล่าวให้ผลลัพธ์ที่ไม่แตกต่างจากการปล่อยให้ทำซ้ำจนกระทั่งเมทริกซ์ตัวประกอบที่ต้องการลู่เข้าสู่ค่าคงที่ ผลลัพธ์จากการทดลองสามารถแสดงได้ดังตารางที่ 8

ตารางที่ 8 ผลลัพธ์ค่าประสิทธิผลซึ่งเปรียบเทียบเทคนิคที่นำเสนอกับเทคนิคพื้นฐานแบบอื่น

เทคนิคที่ใช้ทดสอบ	Mean Average Precision	Mean Percentage Ranking (%)
NMF แบบพื้นฐาน (MUR)	0.0874	26.20
NMF แบบพื้นฐาน (ALS)	0.0943	24.69
WNMF-OSR (MUR)	0.1295	15.03
WNMF-OSR (ALS)	0.1398	13.92
wALS	0.1179	18.79

จากผลการทดลองในตารางที่ 8 จะเห็นได้ว่าเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนักด้วยความสัมพันธ์เชิงสังคมออนไลน์ (WNMF-OSR) ที่นำเสนอ ช่วยเพิ่มคุณภาพของระบบแนะนำที่ใช้การกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวได้ดีกว่าเทคนิคอื่น ๆ โดย WNMF-OSR ที่ใช้ขั้นตอนวิธี ALS ช่วยปรับปรุง ค่า MAP และค่า MPR ให้ดีขึ้นจากผลลัพธ์ที่ได้จากเทคนิค wALS ซึ่งไม่ได้ใช้การพิจารณาค่าน้ำหนักด้วยความสัมพันธ์เชิงสังคมออนไลน์ ด้วยค่าที่แตกต่างกัน 0.0219 และ 4.87% ตามลำดับ พร้อมกันนั้น ผลการทดลองแสดงให้เห็นว่าการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่ไม่มีการพิจารณาค่าน้ำหนัก จะให้ผลลัพธ์ที่ไม่ดีสำหรับชุดข้อมูลที่เป็นปัญหาการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ดังนั้นการพิจารณาค่าน้ำหนักจึงเป็นสิ่งที่ควรกระทำ เมื่อต้องการใช้เทคนิคการแยกตัวประกอบเมทริกซ์ เพื่อจัดการปัญหาที่จำเป็นต้องพิจารณาตัวอย่างที่ไม่ระบุป้ายกำกับคลาส

นอกจากนี้ จากการทดสอบด้วยขั้นตอนวิธีทั้ง MUR และขั้นตอนวิธี ALS ทำให้พบว่าการใช้ขั้นตอนวิธี ALS ค่อนข้างให้ผลลัพธ์ที่ดีกว่าเล็กน้อย เมื่อเทียบกับขั้นตอนวิธี MUR ไม่ว่าจะใช้ขั้นตอนวิธีดังกล่าวกับเทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนัก หรือไม่มีการพิจารณาค่าน้ำหนัก

ดังนั้นจากการทดลองส่วนสำคัญ สรุปได้ว่าเทคนิคใหม่ที่นำเสนอ ซึ่งใช้ความสัมพันธ์เชิงสังคมออนไลน์ สามารถทำงานได้อย่างมีประสิทธิภาพและให้ผลลัพธ์ที่ค่อนข้างดีเมื่อทดสอบกับชุดข้อมูลของผู้ใช้จริงบนบริการเครือข่ายสังคม โดยสาเหตุที่เป็นเช่นนี้ เนื่องจากพฤติกรรมของผู้ใช้ในการพิจารณาทางด้านความชอบความสนใจส่วนใหญ่แล้ว หากไม่เป็นการพิจารณาที่เกิดจากความสนใจส่วนบุคคล ผู้ใช้ก็มักจะพิจารณาโดยอ้างอิงจากข้อมูลที่ได้รับจากเพื่อนของตนนั่นเอง ซึ่งพฤติกรรมดังกล่าวนี้อาจเรียกได้ว่าเป็น การจูงใจเชิงสังคม (social influence) และสามารถบ่งชี้ความน่าจะเป็นที่ตัวอย่างที่ไม่ระบุป้ายกำกับคลาสตำแหน่งนั้น ๆ จะเป็นตัวอย่างบวกสำหรับการเรียนรู้ลักษณะความชอบของผู้ใช้ระบบแนะนำ

บทที่ 6
สรุปผลการวิจัยและข้อเสนอแนะ

สรุปผลการวิจัย

วิทยานิพนธ์นี้ได้กล่าวถึงการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว (One-class Collaborative Filtering) ซึ่งเป็นปัญหาของระบบแนะนำที่กำลังอยู่ในความสนใจ เนื่องจากระบบขาดแคลนตัวอย่างข้อมูลป้อนกลับที่ชัดแจ้ง (explicit feedback) หรือค่าคะแนนความชอบที่เป็นอัตราเชิงตัวเลข (numerical rating) จากผู้ใช้ การวิจัยนี้มีแรงจูงใจจากข้อเท็จจริงที่ว่า ผู้ใช้แต่ละคนจะรับเอาพฤติกรรมของเพื่อนตนเองบนเครือข่ายสังคมออนไลน์ในการตัดสินใจหรือพิจารณาเลือกไอเทมต่าง ๆ ที่ตนเองชอบ บริการเครือข่ายสังคมออนไลน์เป็นสิ่งที่กำลังได้รับความนิยม มีอัตราการเติมโตที่สูงมาก และอุดมไปด้วยข้อมูลคุณลักษณะ ปฏิสัมพันธ์ และความสนใจต่าง ๆ ของผู้ใช้ทุกคนในเครือข่ายสังคม ข้อมูลผู้ใช้จริงบนเครือข่ายเหล่านี้ สะท้อนมาจริงชีวิตในโลกแห่งความจริงของตัวผู้ใช้เอง ซึ่งสามารถใช้ประโยชน์ในปรับปรุงประสิทธิผลของระบบแนะนำที่เป็นการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยวได้ ผู้ดำเนินการวิจัยจึงเสนอแนวคิดวิธีการใหม่สำหรับการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ซึ่งใช้ประโยชน์จากข้อมูลความสัมพันธ์ของผู้ใช้บนเครือข่ายสังคมออนไลน์ วิธีการที่นำเสนอใช้โมเดลการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนักในการเรียนรู้ลักษณะความชอบของผู้ใช้เพื่อการแนะนำผู้ใช้แต่ละคน โดยการพิจารณาค่าน้ำหนักดังกล่าวอ้างอิงจากการจูงใจเชิงสังคม (social influence) บนเครือข่ายสังคมออนไลน์ ซึ่งเชื่อว่าหากผู้ใช้เป้าหมายที่ระบบต้องการแนะนำ มีเพื่อนในสังคมออนไลน์เป็นจำนวนมากที่สนใจในไอเทม i ใด ๆ ก็จะมีแนวโน้มสูงที่ผู้ใช้เป้าหมายนั้นสนใจชอบไอเทม i นั้น

วิธีการที่นำเสนอในวิทยานิพนธ์นี้ ปรับปรุงคุณภาพของระบบแนะนำสำหรับผู้ใช้บนเครือข่ายสังคมออนไลน์ การทดลองเพื่อทดสอบประสิทธิผลของเทคนิคและสมมติฐานต่าง ๆ กระทำบนชุดข้อมูลที่แสดงถึงความชอบหรือความสนใจของผู้ใช้จริงในบริการเครือข่ายสังคม Facebook จากการทดลองพบว่า เทคนิคการแยกตัวประกอบเมทริกซ์แบบไม่เป็นลบที่มีการพิจารณาค่าน้ำหนักและใช้ข้อมูลความสัมพันธ์เชิงสังคม ให้ผลดีกว่าสำหรับระบบแนะนำที่ใช้เฉพาะข้อมูลปฏิสัมพันธ์ระหว่างผู้ใช้และไอเทม หรือตัวอย่างบวกเพียงอย่างเดียว ดังเช่นการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว

นอกจากนี้ วิทยานิพนธ์นี้ยังได้สนับสนุนแนวคิดการประยุกต์ใช้ระบบแนะนำกับบริการเครือข่ายสังคมออนไลน์ ด้วยการใช้ประโยชน์จากข้อมูลที่มีอยู่แล้วภายในระบบเครือข่ายสังคมออนไลน์ ทำให้เกิดการปรับปรุงคุณภาพของระบบแนะนำพร้อมกันกับบริการเครือข่ายสังคมออนไลน์สามารถตอบสนองความต้องการของผู้ใช้ได้อย่างเต็มที่ รวมถึงแนวคิดดังกล่าวยังมีความยืดหยุ่นในการนำไปใช้งานจริงและการคิดค้นเทคนิคใหม่ ๆ ต่อไป

ข้อเสนอแนะ

อย่างไรก็ตามในประเด็นของระบบแนะนำและการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว สามารถปรับปรุงด้วยแนวทางการใช้ความสัมพันธ์เชิงสังคมออนไลน์ ในรูปแบบอื่น ๆ ตัวอย่างเช่น การปรับวิธีประยุกต์ใช้ความสัมพันธ์เชิงสังคมออนไลน์ดังกล่าว การใช้เทคนิคการแยกตัวประกอบเมทริกซ์และขั้นตอนวิธีแบบอื่น เช่น Rank-one Residue Iteration (RRI) รวมถึงสามารถใช้ข้อมูลคุณลักษณะพื้นฐานของผู้ใช้ เช่น เพศ ช่วงอายุ ตำแหน่งพื้นที่อาศัย เพื่อปรับปรุงคุณภาพของระบบแนะนำที่เป็นการกรองแบบร่วมมือกันสำหรับคลาสเดี่ยว ทั้งหมดนี้ยังคงเป็นประเด็นที่ต้องมีการศึกษาวิจัยต่อไป

รายการอ้างอิง

[1] Adomavicius, G., and Tuzhilin, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. IEEE Transactions on Knowledge and Data Engineering 17, 6 (2005) : 734-749.

[2] Pan, R., and others. One-Class Collaborative Filtering. In Proc. 8th IEEE International Conference on Data Mining, pp. 502-511. 2008.

[3] Nati N.S., and Jaakkola, T. Weighted Low-Rank Approximations. In Proc. 20th International Conference on Machine Learning, pp. 720-727. 2003.

[4] Sindhwani, V., Bucak, S.S., Hu, J., and Mojsilovic, A. A Family of Non-negative Matrix Factorizations for One-Class Collaborative Filtering Problems. In Proc. Recommender Based Industrial Applications Workshop, 2009.

[5] Linden, G., Smith, B., and York, J. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. IEEE Internet Computing 7, 1 (2003) : 76-80.

[6] Miller, B., Albert, I., Lam, S., Konstan, J., and Riedl, J. MovieLens Unplugged: Experiences with an Occasionally Connected Recommender System. In Proc. ACM Conference on Intelligent User Interfaces, pp. 263-266. 2003.

[7] Koren Y. The BellKor Solution to the Netflix Grand Prize [Online]. 2009. Available from : http://www.netflixprize.com/assets/GrandPrize2009_BPC_BellKor.pdf [15 January 2011]

[8] Zhou, Y., Wilkinson, D., Schreiber, R., and Pan, R. Large-Scale Parallel Collaborative Filtering for the Netflix Prize. In Proc. 4th International Conference on Algorithmic Aspects in Information and Management, pp. 337-348. 2008.

[9] Das, A.S., Datar, M., Garg, A., and Rajaram, S. Google news personalization: scalable online collaborative filtering. In Proc. 16th International Conference on World Wide Web, pp. 271-280. 2007.

[10] Breese, J., Heckerman, D., and Kadie, C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proc. 14th Annual Conference on Uncertainty in Artificial Intelligence, pp. 43-52. 1998.

[11] Srebro, N. Learning with Matrix Factorizations. Doctoral dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2004.

[12] Schölkopf, B., Platt, J.C., Shawe-Taylor, J.C., Smola, A.J., and Williamson, R.C. Estimating the Support of a High-Dimensional Distribution. Neural Computation 13 (July 2001) : 1443-1471.

[13] Ben-David, S., and Lindenbaum, M. Learning Distributions by Their Density Levels: A Paradigm for Learning without a Teacher. Journal of Computer and System Sciences 55 (August 1997) : 171-182.

[14] Ward, G., Hastie, T., Barry, S., Elith, J., and Leathwick. J.R. Presence-Only Data and the EM Algorithm. Biometrics 65 (2009) : 554-563.

[15] Marlin, B.M., and Zemel, R.S. Collaborative filtering and the missing at random assumption. In Proc. 23rd Conference on Uncertainty in Artificial Intelligence, pp. 267-275. 2007.

[16] Kurucz, M., and others. Who rated what: a combination of SVD, correlation and frequent sequence mining. In Proc. KDD Cup and Workshop, 2007.

[17] Koren, Y., Bell, R., and Volinsky, C. Matrix Factorization Techniques for Recommender Systems. Computer 42, 8 (August 2009) : 30-37.

[18] Walter, F.E., Battiston, S., and Schweitzer, F. A model of a trust-based recommendation system on a social network. Autonomous Agents and Multi-Agent Systems 16, 1 (October 2007) : 57-74.

[19] Koren, Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 426-434. 2008.

[20] Li, Y., Hu, J., Zhai, C., and Chen, Y. Improving one-class collaborative filtering by incorporating rich user information. In Proc. 19th ACM international conference on Information and knowledge management, pp. 959-968. 2010.

[21] Paatero, P., and Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics 5, 2 (June 1994) : 111-126.

[22] Lee, D.D., and Seung, H.S. Learning the parts of objects by non-negative matrix factorization. Nature 401, 6755 (October 1999) : 788-791.

[23] Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., and Plemmons, R.J. Algorithms and applications for approximate nonnegative matrix factorization. Computational Statistics & Data Analysis 52, 1 (September 2007) : 155-173.

[24] Paatero, P. A weighted non-negative least squares algorithm for three-way `PARAFAC' factor analysis. Chemometrics and Intelligent Laboratory Systems 38, 2 (October 1997) : 223-242.

[25] Cichocki, A., Zdunek, R., Phan, A.H., and Amari, S. Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. Hoboken, NJ : John Wiley & Sons, 2009.

[26] Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., and Christakis, N. Tastes, ties, and time: A new social network dataset using Facebook.com. Social Networks 30, 4 (October 2008) : 330-342.

ประวัติผู้เขียนวิทยานิพนธ์

นายสิรวิชญ์ สพโชค สำเร็จการศึกษาวิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ จากภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2551 (เมษายน 2552) และได้เข้าศึกษาต่อในหลักสูตรวิทยาศาสตรมหาบัณฑิต ที่ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2552