การลดผลกระทบของปัญหาความเบาบางของข้อมูลในวิธีการแบบการกรองผู้ใช้ร่วมด้วยข้อมูลความสัมพันธ์ทางสังคม

REDUCING IMPACT OF DATA SPARSITY IN COLLABORATIVE FILTERING USING SOCIAL RELATIONSHIP

กวินทร์ พิพัฒน์กุล

KAWIN PIPATKUL

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2556

KMITL-2013-SC-M-002-030

การลดผลกระทบของปัญหาความเบาบางของข้อมูลในวิธีการแบบการกรองผู้ใช้ร่วมด้วยข้อมูลความสัมพันธ์ทางสังคม

REDUCING IMPACT OF DATA SPARSITY IN COLLABORATIVE FILTERING USING SOCIAL RELATIONSHIP

กวินทร์ พิพัฒน์กุล

KAWIN PIPATKUL

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

สาขาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2556

KMITL-2013-SC-M-002-030

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

REDUCING IMPACT OF SPARSITY IN COLLABORATIVE FILTERING USING SOCIAL RELATIONSHIP

KAWIN PIPATKUL

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF SCIENCE IN COMPUTER SCIENCE

FACULTY OF SCIENCE

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2013

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2013

FACULTY OF SCIENCE

KING MONGK UT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะวิทยาศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ การลดผลกระทบของปัญหาความเบาบางของข้อมูลในวิธีการแบบการกรองผู้ใช้ร่วมด้วยข้อมูลความสัมพันธ์ทางสังคม
Reducing Impact of Data Sparsity in Collaborative Filtering Using Social Relationship

นักศึกษา นายกวินทร์ พิพัฒน์กุล
รหัสประจำตัว 51067501
ปริญญา วิทยาศาสตรมหาบัณฑิต
สาขาวิชา วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์ รศ.ดร.วีระ บุญจริง

คณะกรรมการสอบวิทยานิพนธ์		ลายมือชื่อ
ดร.วรางคณา	กิ้มปาน	
ผศ.ดร.ศรัณย์	อินทโกสุม	
ดร.เฉลิมศักดิ์	เลิศวงศ์เสถียร	
รศ.ดร.วีระ	บุญจริง	

วัน / เดือน/ ปี ที่สอบ 17 พฤษภาคม พ.ศ. 2556 เวลา 09.00-12.00 น.
สถานที่สอบ ณ ห้อง 304 ชั้น 3 อาคารจุฬาภรณวลัยลักษณ์ 1

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

คณะวิทยาศาสตร์รับรองแล้ว

(รองศาสตราจารย์ ดร.ดุษณี ธนะบริพัฒน์)
คณบดีคณะวิทยาศาสตร์
วันที่ 29 **เดือน** พ.ค. **พ.ศ.** 56

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การลดผลกระทบของปัญหาความเบาบางของข้อมูลในวิธีการแบบการกรองผู้ใช้ร่วมด้วยข้อมูลความสัมพันธ์ทางสังคม
นักศึกษา	นายกวินทร์ พิพัฒน์กุล
รหัสประจำตัว	51067501
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2556
อาจารย์ที่ปรึกษา	รศ.ดร.วีระ บุญจริง

บทคัดย่อ

ปัญหาความเบาบางของข้อมูล (Data Sparsity Problem) เป็นปัญหาที่สำคัญประการหนึ่งของวิธีการแบบการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering Method) ซึ่งเป็นการยากสำหรับการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมที่จะหาความเหมือนของผู้ใช้งานได้อย่างแม่นยำจากชุดข้อมูลที่มีการขาดหายของข้อมูลจำนวนมาก โดยในงานวิจัยฉบับนี้ได้นำเสนอวิธีการลดผลกระทบจากปัญหาดังกล่าวที่มีต่อประสิทธิภาพของวิธีการแบบการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม ด้วยการนำข้อมูลความสัมพันธ์ของผู้ใช้งานในเครือข่ายสังคมออนไลน์เข้ามาเพิ่มประสิทธิภาพของวิธีการแบบการกรองข้อมูลผู้ใช้ร่วม ซึ่งได้ทำการทดลองกับชุดข้อมูลบนเครือข่ายสังคมออนไลน์ที่ใช้งานอยู่จริง จากผลการทดลอง วิธีการที่นำเสนอนอกจากจะได้ค่าความผิดพลาดที่น้อยลงแล้ว ยังมีความทนทานต่อความเบาบางของข้อมูลมากขึ้นด้วย

คำสำคัญ: ระบบแนะนำ, การกรองข้อมูลผู้ใช้ร่วม, ความเบาบางของข้อมูล, ข้อมูลผู้ใช้เครือข่ายสังคมออนไลน์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Reducing Impact of Sparsity in Collaborative Filtering Using Social Relationship
Student	Kawin Pipatkul
Student ID	51067501
Degree	Master of Science
Program	Computer Science
Year	2013
Thesis Advisor	Assoc.Prof.Dr.Veera Boonjing

ABSTRACT

Data Sparsity Problem is known as a major problem of Collaborative Filtering Method, that is, it is difficult for collaborative filtering to accurately measure user similarities from the data that contain many missing value. In this paper we propose the method for reducing the impact of mentioned problem by incorporating social friend information to improve the performance of collaborative filtering method. Experimental result reveals that the proposed method not only yields lower error of recommendation but also increases the resistant to data sparsity.

Keywords: Recommender System, Collaborative Filtering, Data Sparsity, Social friends

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้มิอาจจะสำเร็จลุลวงไปด้วยดี หากมิได้รับคำแนะนำ คำชี้แจ้ง ความรู้ และความเอาใจใส่จาก รศ.ดร.วีระ บุญจริง ผู้เป็นอาจารย์ที่ปรึกษา ที่ปลูกฝังและสั่งสอนแนวทางการทำวิจัยมาตั้งแต่วันแรกที่ได้มาเป็นนักศึกษาของที่นี่ และได้สละเวลาให้อย่างเต็มที่ให้คำแนะนำให้คำปรึกษาอย่างใกล้ชิด และเสนอแนะแนวทางแก้ปัญหา รวมทั้งตรวจแก้วิทยานิพนธ์ฉบับนี้หลายรอบมาก ให้มีความสมบูรณ์เพิ่มขึ้น จึงใคร่ขอขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ ผศ.ดร. ศรัณย์ อินทโกสุม ดร. วรางคณา กิ้มปาน และดร. เฉลิมศักดิ์ เลิศวงศ์เสถียร คณะกรรมการสอบวิทยานิพนธ์ ที่ให้คำแนะนำ ให้คำปรึกษา และเสนอแนะแนวทางแก้ปัญหาจนในที่สุดทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลงได้

ขอขอบคุณเจ้าหน้าที่ประจำสาขาวิทยาการคอมพิวเตอร์ รวมทั้งเจ้าหน้าที่ประจำภาคบัณฑิต คณะวิทยาศาสตร์ ที่ให้ความร่วมมือ และอำนวยความสะดวก ในการทำวิทยานิพนธ์ให้สำเร็จลุลวงไปด้วยดี

สุดท้ายนี้ผู้จัดทำ ขอขอบพระคุณ บิดา มารดา และบุคคลในครอบครัว ที่ได้ให้ความช่วยเหลือทุกๆด้าน รวมทั้งเพื่อนๆ ทุกคน ที่ให้กำลังใจและช่วยสนับสนุนตลอดระยะเวลาในการทำวิทยานิพนธ์

กวินทร์ พิพัฒน์กุล

พฤษภาคม 2556

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ(ต่อ)

หน้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของปัญหา

ระบบผู้แนะนำ [3] (Recommender System) เป็นเครื่องมือและเทคนิควิธีการให้คำแนะนำข้อมูลที่น่าสนใจและมีความเกี่ยวข้องกับผู้ใช้ โดยมุ่งเน้นในการแก้ปัญหาที่ผู้ใช้ต้องเผชิญกับข้อมูลหรือทางเลือกจำนวนมหาศาล ซึ่งแทบเป็นไปไม่ได้ที่ผู้ใช้จะสามารถพิจารณาข้อมูลทั้งหมด จากที่กล่าวมาข้างต้น เรียกว่าปัญหาข้อมูลท่วมท้น (Information Overload) ระบบผู้แนะนำจะทำการเตรียมรายการข้อมูลที่น่าสนใจโดยเรียนรู้จากพฤติกรรมของผู้ใช้ในอดีต ซึ่งวิธีการที่ถูกใช้กันอย่างแพร่หลายและได้รับความนิยมมากที่สุดคือวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) [1] ด้วยแนวความคิดพื้นฐานที่ว่า บุคคลใดที่มีความนิยมคล้ายคลึงกันในอดีตจะมีแนวโน้มที่จะมีความนิยมคล้ายคลึงกันด้วยในอนาคต

ในทางกลับกันกับปัญหาข้อมูลท่วมท้น ผู้ใช้มักให้คะแนนชิ้นข้อมูล (สินค้า, บริการ, เอกสาร หรือข้อมูลอื่นๆ) เพียงไม่กี่รายการเท่านั้น ซึ่งเป็นการยากสำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมในการประเมินความคล้ายคลึงความนิยมของผู้ใช้ได้อย่างแม่นยำ และทำให้ประสิทธิภาพในการทำนายลดลง โดยปัญหาที่ระบบมีข้อมูลของผู้ใช้ไม่เพียงพอในการทำนายความนิยมได้อย่างมีประสิทธิภาพนั้น ถูกเรียกว่าปัญหาความเบาบางของข้อมูล (Data Sparsity) เพื่อแก้ปัญหาความเบาบางของข้อมูล วิธีการเพิ่มความหนาแน่นของข้อมูลเป็นวิธีการที่นำมาใช้กันอย่างแพร่หลาย ไม่ว่าจะเป็น (1) การเพิ่มความหนาแน่นของข้อมูลด้วยการตัดข้อมูลที่ไม่จำเป็นทิ้งจากผู้ใช้และชิ้นข้อมูลออกด้วยเทคนิคการลดมิติข้อมูล (Dimensional Reduction) [5,14] หรือ (2) การเพิ่มความหนาแน่นของข้อมูลด้วยการเติมค่าข้อมูลที่ขาดหายโดยอาศัยวิธีการต่างๆ เช่น การแทนค่าข้อมูลที่ขาดหายด้วยวิธีการทางสถิติ [17], การประมาณค่าความนิยมโดยดูจากระยะเวลาที่ผู้ใช้ใช้ในการพิจารณาชิ้นข้อมูล [6] ซึ่งเป็นการให้คะแนนความนิยมของผู้ใช้ทางอ้อม หรือแม้กระทั่งการนำข้อมูลส่วนบุคคลเข้ามาร่วมในการเติมค่าข้อมูลที่ขาดหาย โดยอาศัยเทคนิควิธีการกรองโดยดูที่เนื้อหา (Content-based Filtering) ร่วมด้วย

เนื่องจากวิธีการเติมค่าข้อมูลที่ขาดหายเป็นการเพิ่มความหนาแน่นของข้อมูลโดยไม่มีการตัดข้อมูลบางส่วนออกเช่นเดียวกับวิธีการลดมิติข้อมูล ทำให้สามารถใช้ข้อมูลที่ได้มาจากผู้ใช้ได้อย่าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เต็มที่ ประกอบกับความนิยมและการเติบโตในการใช้งานระบบเครือข่ายสังคมออนไลน์ (Online Social Network: OSN) ที่เพิ่มสูงขึ้น โดยระบบเครือข่ายสังคมออนไลน์จะมีคุณลักษณะสำคัญประการหนึ่งคือ ผู้ใช้สามารถสร้างความสัมพันธ์กับผู้ใช้คนอื่นได้ ทำให้เกิดความเชื่อมโยงกันระหว่างผู้ใช้ ซึ่งหากมองในเชิงปริมาณเพียงอย่างเดียวแล้วเครือข่ายสังคมออนไลน์สามารถเพิ่มปริมาณข้อมูลได้หลายเท่าตัว ดังนั้นเพื่อรับมือกับปัญหาความเบาบางของข้อมูลสำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม งานวิจัยนี้จึงมีแนวความคิดในการเติมค่าข้อมูลที่ขาดหายไปโดยอาศัยข้อมูลความสัมพันธ์ของผู้ใช้ในเครือข่ายสังคมออนไลน์ จากความเชื่อที่ว่าบุคคลจะมีแนวโน้มความนิยมเป็นไปตามกลุ่มเพื่อนของตนมากกว่าบุคคลทั่วไป

1.2 วัตถุประสงค์

เพื่อพัฒนาตัวแบบที่สามารถรับมือกับปัญหาความเบาบางของข้อมูลในวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม รวมไปถึงความทนทานต่อระดับความเบาบางของข้อมูลในระดับต่างๆ โดยอาศัยข้อมูลความสัมพันธ์ของผู้ใช้ในเครือข่ายสังคมออนไลน์ในการเติมค่าข้อมูลที่ขาดหาย

1.3 ขอบเขตงานวิจัย

งานวิทยานิพนธ์นี้มีขอบเขตของการวิจัย ดังนี้

1. งานวิทยานิพนธ์นี้พัฒนาตัวแบบการเติมค่าข้อมูลที่ขาดหายสำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมเพื่อลดผลกระทบของปัญหาความเบาบางของข้อมูล และเพิ่มประสิทธิภาพการทำนายของวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม

2. ข้อมูลทดสอบ สกัดมาจากเว็บไซต์ Yelp.com ซึ่งเป็นเว็บไซต์ที่มีชื่อเสียง และเป็นแหล่งรวบรวมบทวิจารณ์ของร้านค้าและมีคุณลักษณะที่เป็นเครือข่ายสังคม โดยสกัดข้อมูล ณ วันที่ 16 เมษายน พ.ศ. 2556 เป็นข้อมูลของร้านอาหารไทยในเมืองซอลต์เลกซิตี รัฐยูทาห์ ประเทศสหรัฐอเมริกา

3. ชุดข้อมูลทดสอบจะถูกนำไปทดสอบประสิทธิภาพด้วยค่าความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error: MAE) และทดสอบความคงทนต่อความเบาบางของข้อมูลด้วยการเปรียบเทียบประสิทธิภาพของตัวแบบเมื่อทดสอบกับชุดข้อมูลที่มีความเบาบางของข้อมูลระดับต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.4 ส่วนประกอบของวิทยานิพนธ์

ส่วนที่เหลือของวิทยานิพนธ์ประกอบด้วยบทต่างๆ ดังนี้

บทที่สอง อธิบายถึงระบบผู้แนะนำบนพื้นฐานของวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม ปัญหาและงานวิจัยที่เกี่ยวข้องกับความเบาบางของข้อมูลในวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม ตลอดจนการใช้ประโยชน์จากเครือข่ายสังคมในระบบผู้แนะนำ

บทที่สาม กล่าวถึงการลดผลกระทบของปัญหาความเบาบางของข้อมูลสำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมด้วยความสัมพันธ์ของผู้ใช้ โดยแบ่งออกเป็นสองส่วน ส่วนแรกกล่าวถึงตัวแบบและขั้นตอนการทำงานของตัวแบบ ส่วนที่สองจะกล่าวถึงตัวอย่างการทำงานของตัวแบบ

บทที่สี่ จะกล่าวถึงชุดข้อมูลที่ใช้ในการทดลอง วิธีการทดลอง และผลการทดลอง รวมไปถึงการวัดประสิทธิภาพของตัวแบบ ในการลดผลกระทบของปัญหาความเบาบางของข้อมูลในวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมด้วยการเติมค่าข้อมูลที่ขาดหายโดยอาศัยข้อมูลความสัมพันธ์ของผู้ใช้ในเครือข่ายสังคมออนไลน์ ทั้งในด้านประสิทธิภาพในการทำนาย และความทนทานของตัวแบบที่มีต่อระดับความเบาบางของข้อมูล

บทที่ห้า จะกล่าวสรุปขั้นตอนวิธีการเติมค่าข้อมูลที่ขาดหายสำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมโดยอาศัยข้อมูลความสัมพันธ์ของผู้ใช้ และข้อเสนอแนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการลดผลกระทบของปัญหาความเบาบางของข้อมูล (Data Sparsity) ในระบบผู้แนะนำ (Recommender System) ที่ใช้วิธีการแบบการกรองผู้ใช้ร่วม (Collaborative Filtering) โดยอาศัยข้อมูลความสัมพันธ์ของผู้ใช้

2.1 ระบบผู้แนะนำ

เนื่องจากการใช้เทคโนโลยีสารสนเทศในปัจจุบันช่วยให้การเข้าถึงและการกระจายข้อมูลสามารถทำได้อย่างสะดวกมากยิ่งขึ้น จนในบางครั้งผู้ใช้อาจตกอยู่ท่ามกลางข้อมูลหรือทางเลือกจำนวนมหาศาล ซึ่งปัญหานี้ทำให้การเข้าถึงและการได้รับสารสนเทศที่ตรงความต้องการเป็นเรื่องยากและใช้เวลานาน เราเรียกปัญหานี้ว่าปัญหาข้อมูลท่วมท้น (Information Overload) เพื่อที่จะแก้ปัญหาดังกล่าวระบบผู้แนะ (Recommender System) [3] ได้เข้ามามีบทบาทและได้รับความนิยมเป็นอย่างมากทั้งในการศึกษาการวิจัยและการใช้งานในเชิงพาณิชย์อิเล็กทรอนิกส์ ได้แก่ Amazon.com, Netflix.com, MovieLens.com และ IMDb.com โดยระบบผู้แนะนำจะช่วยแนะนำข้อมูลที่เกี่ยวข้องหรือน่าสนใจในฐานข้อมูลขนาดใหญ่ให้กับผู้ใช้โดยอาศัยหลักการและเทคนิควิธีการของศาสตร์ต่างๆมาประยุกต์ใช้ในการสร้างรายการแนะนำ เช่น การค้นคืนสารสนเทศ (Information Retrieval) [12],การทำเหมืองข้อมูล (Data Mining) [7]และ Machine Learning [2,13] เป็นต้น ซึ่งสามารถแบ่งประเภทของวิธีการสำหรับระบบผู้แนะนำได้เป็นสามประเภทหลักๆ ซึ่งจะกล่าวถึงในหัวข้อถัดไป

2.2 เทคนิควิธีการที่ใช้ในระบบผู้แนะนำ

เทคนิควิธีการที่ใช้ในระบบผู้แนะนำสามารถแบ่งออกได้เป็นสามประเภทหลักๆ ตามประเภทข้อมูลที่นำมาใช้ในการสร้างรายการแนะนำดังนี้

1. การกรองโดยดูที่เนื้อหา (Content-based Filtering) ซึ่งเป็นวิธีในการแนะนำชิ้นข้อมูลจากข้อมูลของชิ้นข้อมูลที่ผู้ใช้เคยมีประสบการณ์การบริโภคมาก่อน ตัวอย่างเช่นข้อมูลภาพยนต์ การกรองโดยดูที่เนื้อหาก็จะใช้ข้อมูลประเภทภาพยนต์, นักแสดงนำ, ผู้กำกับ หรือเนื้อเรื่องของภาพยนต์ที่ผู้ใช้ชอบ มาใช้ในการสร้างรายการแนะนำ เป็นต้น

2. การกรองแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) [1] เป็นวิธีการแนะนำชิ้นข้อมูลจากข้อมูลพฤติกรรมรสนิยมของผู้ใช้คนอื่นที่มีความนิยมคล้ายคลึงกับผู้ใช้ ซึ่งเป็นวิธีที่ได้รับความนิยม

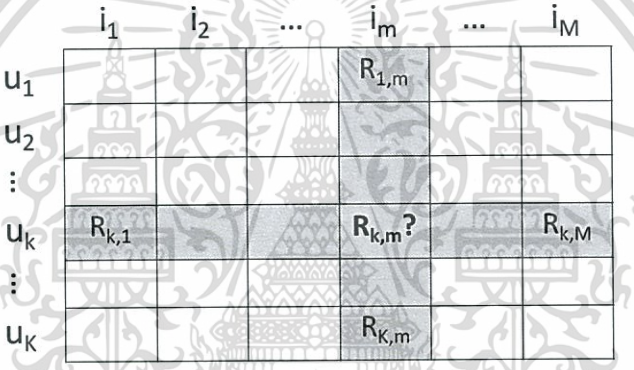
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และประสบความสำเร็จมากที่สุดวิธีหนึ่ง ดังนั้นในงานวิจัยนี้จึงมุ่งเน้นศึกษาเฉพาะการกรองแบบพึ่งพาผู้ใช้ร่วมเท่านั้น

3. แบบผสมผสาน เป็นวิธีการแนะนำชิ้นข้อมูลจากทั้งวิธีการกรองแบบดูที่เนื้อหา และการกรองแบบพึ่งพาผู้ใช้ร่วมมาใช้ร่วมกัน

2.3 การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม

การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมจะแนะนำชิ้นข้อมูลให้กับผู้ใช้โดยพิจารณาจากความนิยมของผู้ใช้ที่มีต่อชิ้นข้อมูล และพฤติกรรมความนิยมนั้น มาวิเคราะห์เปรียบเทียบกับความนิยมของผู้ใช้คนอื่นๆ โดยมีสมมติฐานที่ว่าบุคคลที่มีความนิยมคล้ายคลึงกันในอดีตจะมีแนวโน้มที่จะมีความนิยมที่คล้ายคลึงกันด้วยในอนาคต ซึ่งความนิยมของผู้ใช้มักถูกแสดงในรูปแบบของการให้คะแนน (Ratting) ดังรูปที่ 2.1



รูปที่ 2.1 ตัวอย่างตารางการให้คะแนนความนิยมของผู้ใช้ U_K ที่มีต่อสินค้า I_M ด้วยคะแนน $R_{K,M}$

การให้คะแนนมักจะเป็นการให้คะแนนความนิยมของผู้ใช้ในชิ้นข้อมูล โดยอาจอยู่ในช่วงคะแนนหนึ่งถึงห้าตามระดับความชอบและไม่ชอบ หรือจากพฤติกรรมการการบริโภค เช่น การซื้อหรือไม่ซื้อสินค้า โดยการทำงานของการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม ประกอบด้วยขั้นตอนการทำงานอยู่ 3 ส่วนดังนี้

2.3.1 วิธีการคำนวณความคล้ายคลึง

ในส่วนนี้จะกล่าวถึงตัวชี้วัดค่าความเหมือน (Similarity Computation) [2] แบบต่างๆ ที่ถูกนำมาใช้สำหรับวัดความคล้ายคลึงกันของผู้ใช้สำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม เราจะเริ่มกันที่การวัดระยะทางแบบยูคลิด (Euclidean Distance) ซึ่งเป็นวิธีการในการวัดระยะห่างสำหรับจุดสองจุด และเป็นวิธีที่นิยมนำมาใช้วัดค่าความเหมือนในหลายกรณี

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1.1 การวัดระยะทางแบบยูคลิด

การวัดระยะทางแบบยูคลิด ใช้สำหรับวัดระยะทางระหว่างจุด x และ จุด y ในระบบปริภูมิเวคเตอร์ (Vector Space) ซึ่งสามารถใช้หาระยะห่างในระบบ 1 มิติ, 2 มิติ ไปจนถึงการวัดระยะห่างในระบบหลายมิติ ดังสมการที่ 2.1

$$d(x,y) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2} \tag{2.1}$$

โดยที่

n คือ จำนวนของแอตทริบิวต์ หรือ มิติ

x_k และ y_k หมายถึงแต่ละแอตทริบิวต์ของจุด x และ y

2.3.1.2 การวัดความคล้ายเชิงมุม

การวัดความคล้ายด้วยวิธีการวัดความคล้ายเชิงมุม (Cosine Similarity) นั้นเป็นวิธีการเปรียบเทียบความคล้ายคลึงของเวคเตอร์ โดยพิจารณาจากมุมโคไซน์ (Cosine) ของมุมระหว่าง 2 เวคเตอร์ หากเวคเตอร์มีความคล้ายคลึงกันเวคเตอร์จะทับกันเกือบสนิท มุมจึงมีค่าน้อยค่าโคไซน์ที่ได้จะมีค่ามาก วิธีการคำนวณค่าความคล้ายเชิงมุมดังสมการ 2.2

$$\cos(x,y) = \frac{x \cdot y}{\|x\|\|y\|} \tag{2.2}$$

โดยที่ $\cdot$ หมายถึง vector dot product

$$x \cdot y = \sum_{k=1}^{n} x_k y_k \tag{2.3}$$

ขนาดของเวคเตอร์ x สามารถคำนวณได้จากสมการ 2.4

$$\|x\| = \sqrt{\sum_{k=1}^{n} x_k^2} \tag{2.4}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.3.1.3 สหสัมพันธ์แบบเพียร์สัน (Pearson Correlation Coefficient)

งานวิจัยของ Herlocker และคณะ [4] แสดงให้เห็นว่าวิธีการวัดค่าความคล้ายระหว่างผู้ใช้ที่เหมาะสมที่สุดกับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม คือค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Correlation Coefficient) ดังสมการ 2.5

$$corr(x,y) = \frac{covariance(x,y)}{standard\ deviation(x) * standard\ deviation(y)} \tag{2.5}$$

$$covariance(x,y) = \frac{1}{1-n}\sum_{k=1}^{n}(x_k - \bar{x})(y_k - \bar{y}) \tag{2.6}$$

$$standard\ deviation(x) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2} \tag{2.7}$$

$$standard\ deviation(y) = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(y_k - \bar{y})^2} \tag{2.8}$$

โดยที่ $\bar{x}$ และ $\bar{y}$ หมายถึงค่าเฉลี่ยของ x และ y ตามลำดับ

ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันจะเป็นการบอกระดับความสอดคล้องหรือขัดแย้งกันระหว่างผู้ใช้งาน ซึ่งจะมีค่าอยู่ระหว่าง -1 ถึง 1 โดยค่ามากกว่าศูนย์หมายถึงผู้ใช้มีความนิยมที่มีความสัมพันธ์ในเชิงสอดคล้อง ค่าน้อยกว่าศูนย์หมายถึงผู้ใช้มีความนิยมที่มีความสัมพันธ์ในเชิงขัดแย้ง และค่าศูนย์หมายถึงความชอบระหว่างผู้ใช้งานไม่ได้มีความสัมพันธ์กัน ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันยิ่งค่ามากกว่าศูนย์ยิ่งมีระดับความสอดคล้องกันมาก ในทางกลับกันค่าสัมประสิทธิ์สหสัมพันธ์ยิ่งค่าน้อยกว่าศูนย์ยิ่งมีระดับความขัดแย้งมาก

2.3.2 วิธีการทำนาย

การทำนาย (Prediction) เป็นการพยากรณ์ค่าความนิยมของผู้ใช้ที่มีต่อสินค้า โดยพิจารณาจากความนิยมของผู้ใช้คนอื่นๆ ไม่ว่าจะเป็นผู้ใช้ที่มีความนิยมสอดคล้องกันหรือผู้ใช้ที่มีความนิยมขัดแย้งกันดังสมการที่ 2.9 [8]

$$p_{a,i} = \bar{r_a} + \frac{\sum_{u=1}^{n} w_{a,u}(r_{u,i} - \bar{r_u})}{\sum_{u=1}^{n} w_{a,u}} \tag{2.9}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยที่ n คือจำนวนผู้ใช้

$p_{a,i}$ คือประมาณค่าความชอบของผู้ใช้ a สำหรับสินค้า i

$\bar{r_a}$ คือค่าเฉลี่ยความชอบของผู้ใช้ u

$r_{u,i}$ คือค่าความชอบที่ผู้ใช้ u ให้กับสินค้า i

$w_{a,u}$ คือค่าความคล้ายคลึงระหว่าผู้ใช้ a และผู้ใช้ u

2.3.3 วิธีการสร้างรายการแนะนำ (Recommendation)

การสร้างรายการแนะนำ เมื่อทำการทำนายค่าความนิยมของผู้ใช้ในชิ้นข้อมูลทุกๆรายการแล้ว จะนำค่าที่ได้มาทำการเรียงลำดับ โดยเริ่มจากชิ้นข้อมูลที่ได้คะแนนการทำนายค่าความนิยมที่สูงที่สุดไปจนถึงชิ้นข้อมูลที่มีค่าคะแนนการทำนายต่ำที่สุด ซึ่งการกำหนดว่าจะแสดงรายการแนะนำให้กับผู้ใช้จำนวนกี่รายการนั้นสามารถกำหนดได้ว่าต้องการให้แสดงรายการแนะนำกี่รายการ

2.4 การแก้ปัญหาความเบาบางของข้อมูลในการกรองแบบพึ่งพาผู้ใช้ร่วม

ในทางกลับกันกับปัญหาข้อมูลท่วมท้น (Information Overload) ผู้ใช้มักจะทำการให้คะแนนสินค้าเพียงไม่กี่รายการเท่านั้น ตัวอย่างเช่น ในร้านขายหนังสือออนไลน์ ที่มีหนังสือจำหน่ายมากกว่าสองล้านเล่ม แม้แต่ผู้ใช้ที่มีนิสัยรักการอ่านก็อาจให้คะแนนความนิยมกับหนังสือเพียงไม่กี่ร้อยเล่มเท่านั้น ซึ่งเมื่อเทียบกับจำนวนหนังสือทั้งหมดแล้วถือเป็นอัตราส่วนที่น้อยมาก ซึ่งจากความเบาบางของข้อมูล (Data Sparsity) ดังกล่าว ทำให้เป็นการยากสำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมในการประเมินความคล้ายคลึงความนิยมของผู้ใช้ได้อย่างแม่นยำ และทำให้ประสิทธิภาพในการทำนายลดลง โดยเทคนิคในการแก้ไขปัญหาดังกล่าวจะเป็นการเพิ่มความหนาแน่นให้กับข้อมูลเดิม ด้วยวิธีการต่างๆดังนี้

2.4.1 เทคนิคการลดมิติข้อมูล

เทคนิคการเพิ่มความหนาแน่นของข้อมูลการให้คะแนนความนิยมของผู้ใช้ ด้วยการกรองข้อมูลที่ไม่ได้เป็นปัจจัยสำคัญออกโดยอาศัยเทคนิคการลดมิติข้อมูล (Dimensionality Reduction) [14] ตัวอย่างเช่นเทคนิคการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis: PCA) เป็นต้น ซึ่งเมื่อทำการกรองข้อมูลที่ไม่ได้เป็นปัจจัยสำคัญออกจะทำให้ความหนาแน่นของข้อมูลการให้คะแนนความนิยมเพิ่มขึ้น

การแก้ปัญหาความเบาบางของข้อมูลด้วยการเพิ่มความหนาแน่นของข้อมูลใน [5] ทำโดยการใช้เทคนิคการวิเคราะห์องค์ประกอบหลัก (PCA) มาคัดผู้ใช้และชิ้นข้อมูลที่ไม่จำเป็นออก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

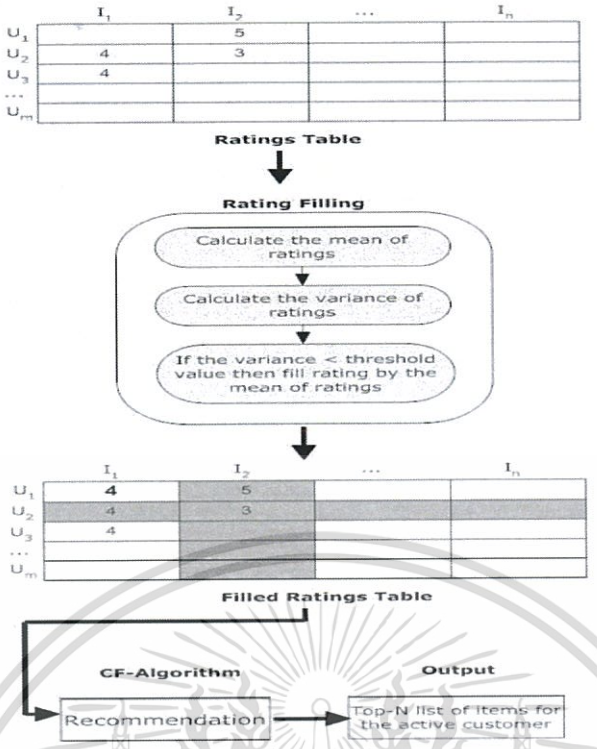
เพื่อให้ได้ข้อมูลที่มีความหนาแน่นมากขึ้น ซึ่งจากผลการทดลองวิธีการนี้สามารถเพิ่มความแม่นยำให้กับการกรองแบบพึ่งพาผู้ใช้ร่วมได้เพียงเล็กน้อย และในงานวิจัยอื่นๆที่ใช้วิธีการ Dimensional Reduction ก็ได้ผลออกมาใกล้เคียงกัน

2.4.2 เทคนิคการเติมค่าข้อมูลที่ขาดหาย

เพื่อลดผลกระทบของปัญหาความเบาบางของข้อมูลที่มีต่อวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม การเติมค่าข้อมูลที่ขาดหายด้วยวิธีการที่หลากหลาย ไม่ว่าจะเป็นการแทนค่าข้อมูลที่ขาดหายด้วยวิธีการทางสถิติ, การประมาณค่าโดยพิจารณาจากการให้คะแนนของผู้ใช้ทางอ้อม เช่น ระยะเวลาที่ผู้ใช้ใช้ในการอ่านบทความของสินค้านั้นๆ ดังที่ได้กล่าวไว้ในงานวิจัยของ [12] หรือแม้กระทั้งการนำเทคนิควิธีการกรองโดยพิจารณาที่เนื้อหา (Content-based Filtering) เข้ามาร่วมด้วย โดยกระบวนการในการเติมค่ามักจะเป็นกระบวนการทำ Data Preprocessing ก่อนที่จะเข้าสู่กระบวนการของการกรองแบบพึ่งพาผู้ใช้ร่วม

การเติมค่าข้อมูลที่ขาดหายด้วยวิธีการทางสถิติใน [17] ทำโดยพิจารณาจากค่าเฉลี่ยความนิยมในสินค้าแต่ละรายการ และค่าความแปรปรวนของข้อมูลหากค่าความแปรปรวนไม่เกินค่าที่กำหนดก็จะนำค่าเฉลี่ยความนิยมของสินค้านั้นมาใช้ในการเติมค่าดังรูปที่ 2.2 ซึ่งจากผลการทดลองพบว่าค่าความผิดพลาดของการทำนายเฉลี่ยที่ได้จากวิธีการที่มีการเติมค่าข้อมูลลงไปแล้วนั้น มีค่าต่ำกว่าค่าความผิดพลาดเฉลี่ยที่ได้จากวิธีการที่ยังไม่มีการเติมค่าข้อมูล ซึ่งส่งผลให้ประสิทธิภาพของระบบผู้แนะนำเพิ่มขึ้น

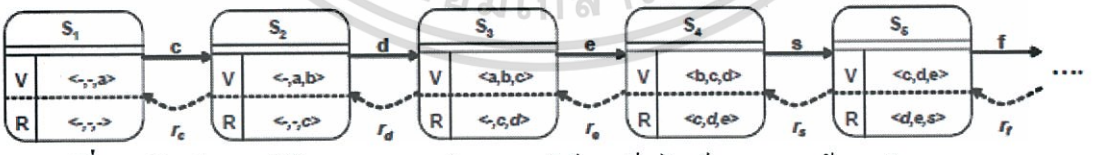
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.2 วิธีการเติมค่าข้อมูลที่ขาดหายของ วงกต ศรีอุไร

(ที่มา วารสารพระจอมเกล้าลาดกระบัง: ปีที่ 16 ฉบับที่ 1 เดือนเมษายน 2551)

การเติมค่าข้อมูลที่ขาดหายด้วยการให้คะแนนความชอบของผู้ใช้ทางอ้อม [6] ได้นำเสนอระบบผู้แนะนำสำหรับแนะนำเว็บไซต์ โดยอาศัยหลักการของการเรียนรู้แบบเสริมกำลัง [12] (Reinforcement Learning) ซึ่งเป็นวิธีการที่เลียนแบบพฤติกรรมการเรียนรู้ของสิ่งมีชีวิต โดยพิจารณาจากประสบการณ์ในการลองผิดลองถูก และได้รับรางวัล (Reward) เป็นผลลัพธ์จากการกระทำนั้นๆ ในงานชิ้นนี้อาศัยเรียนรู้จากพฤติกรรมและรูปแบบการเลือกชมเว็บไซต์ดังตัวอย่างจากรูปที่ 2.3



รูปที่ 2.3 ตัวอย่างการให้คะแนนจากรูปแบบการใช้งานเว็บไซต์ของระบบผู้แนะนำแบบ RL

(ที่มา เอกสารงานประชุมวิชาการ SAC'08 วันที่ 16-20 มีนาคม 2551)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

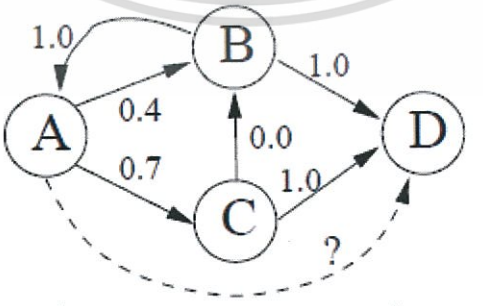
จากรูปที่ 2.3 ระบบจะได้รับรางวัลในการการแนะนำจากเวลาที่ผู้ใช้ใช้ในการรับชมเว็บไซต์ที่ระบบแนะนำ โดยมีหลักการที่ว่าผู้ใช้จะใช้เวลาในการรับชมเว็บไซต์ที่น่าสนใจนานกว่าเว็ปไซต์ที่ไม่น่าสนใจหรือไม่เกี่ยวข้อง ดังนั้นระบบจะพยายามแนะนำเว็บไซต์ที่ผู้อ่านใช้เวลาในการอ่านมากที่สุด

2.5 เครือข่ายสังคมกับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม

ถึงแม้ว่าเครือข่ายสังคมออนไลน์ (Online Social Network) และวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมจะได้รับความนิยมอย่างสูง ทั้งในการศึกษาวิจัย และใช้งานกันอย่างแพร่หลายทั้งโดยบุคคลทั่วไปและองค์กรธุรกิจ แต่การศึกษาการใช้ประโยชน์จากข้อมูลบนเครื่อข่ายสังคมออนไลน์สำหรับวิธีการกรองแบบพึ่งพาผู้ใช่ร่วมยังมีการศึกษากันอย่างจำกัด โดยในหัวข้อนี้จะกล่าวถึงงานวิจัยที่เกี่ยวข้องกับการทำงานร่วมกันระหว่างเครือข่ายสังคมและวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม

ใน [16] ได้นำระบบแนะนำที่จะทำการแนะนำบุคคล ให้กับผู้ใช้ในระบบเครือข่ายสังคมออนไลน์ โดยที่ผู้ใช้งานจะมีบทบาทเป็นทั้งผู้ใช้ (User) และชิ้นข้อมูล (Item) จึงทำให้วิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบปรกติไม่สามารถนำมาประยุกต์ใช้ได้กับการแนะนำบุคคลให้กับผู้ใช้ ดังนั้นจึงได้ทำการปรับปรุงวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมเพื่อให้สามารถที่จะแนะนำบุคคลให้กับผู้ใช้ได้ โดยวิธีการใหม่นี้มีชื่อว่า SocialColab ซึ่งเป็นวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบดูที่เพื่อนบ้าน (neighbor-based Collaborative Filtering) โดยพิจารณาจากความชอบและความน่าดึงดูดใจของผู้ใช้ จากผลการทดลองเปรียบเทียบกับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมปรกติ แสดงให้เห็นว่าวิธีการ SocialColab สามารถที่จะพัฒนาประสิทธิภาพของวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบปรกติได้

ใน [9] ได้นำเสนองานวิจัยวิธีการสำหรับพัฒนาวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมโดยที่ผู้ใช้สามารถที่จะให้คะแนนความน่าเชื่อถือให้กับผู้ใช้คนอื่นๆได้ ซึ่งการคำนวณความน่าเชื่อถือของผู้ใช้แต่ละคนจะใช้หลักการคล้ายกันกับการคิด Page Rank ของ Google ที่ใช้ในการให้ค่าความน่าเชื่อถือของเว็บไซต์ต่างๆ ดังตัวอย่างในรูปที่ 2.4



รูปที่ 2.4 ตัวอย่างเครือข่ายความน่าเชื่อถือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.4 จุดคือผู้ใช้แต่ละคน และด้านที่เชื่อมจุดแต่ละจุดคือคะแนนความน่าเชื่อถือที่ผู้ใช้ให้กับผู้ใช้คนอื่น จากผลการทดลองแสดงให้เห็นว่า Trust-aware CF มีประสิทธิภาพที่ดีในการทำนายถึงแม้ว่าจะเจอกับปัญหาความเบาบางของข้อมูลก็ตาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

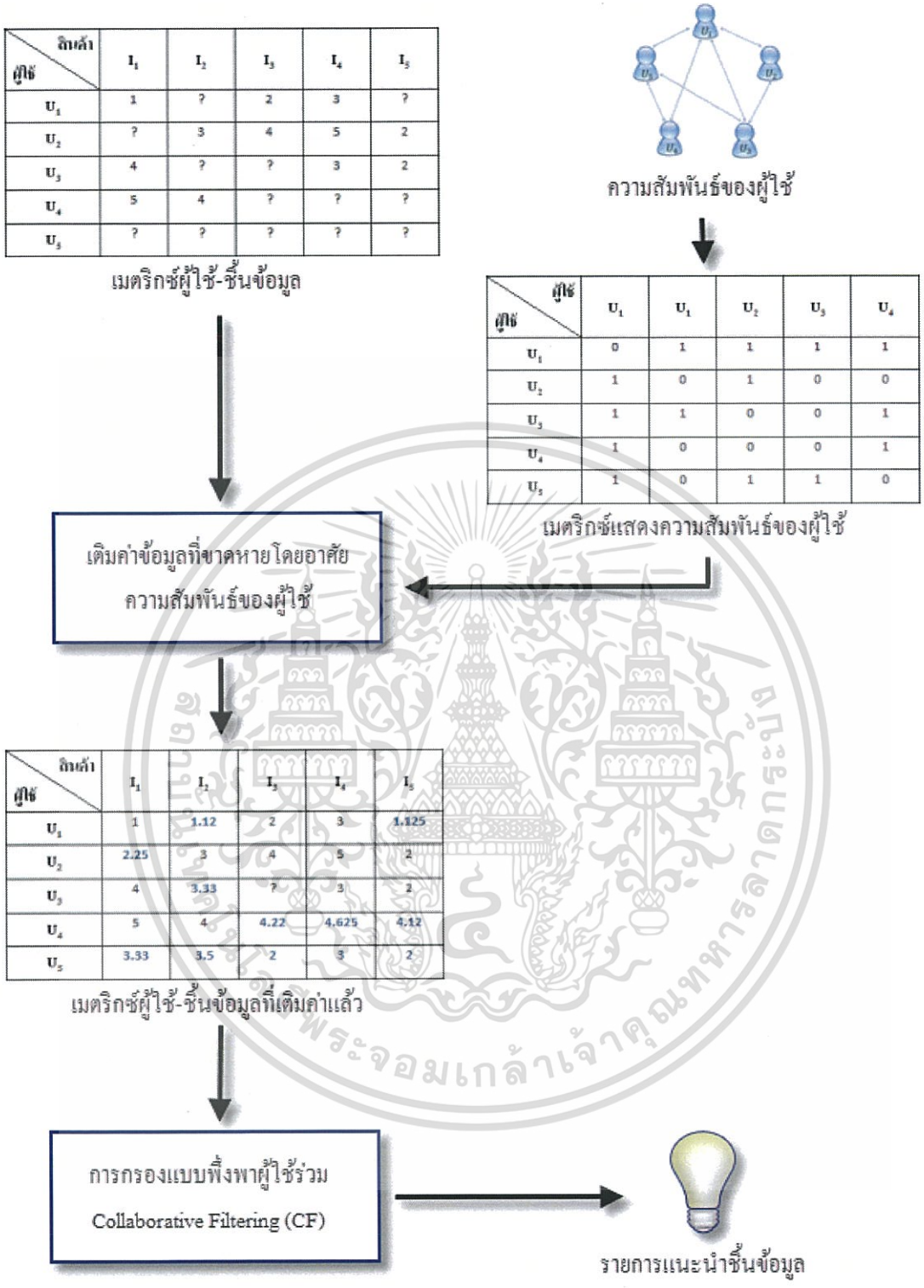
บทที่ 3
การลดผลกระทบของปัญหาความเบาบางของข้อมูลสำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมด้วยความสัมพันธ์ของผู้ใช้

ในบทนี้จะกล่าวถึงการลดผลกระทบของปัญหาความเบาบางของข้อมูลสำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมด้วยความสัมพันธ์ของผู้ใช้ โดยแบ่งออกเป็น 2 ส่วน ส่วนแรกกล่าวถึงตัวแบบและขั้นตอนการทำงานของตัวแบบ ส่วนที่สองจะกล่าวถึงตัวอย่างการทำงานของตัวแบบ

3.1 ตัวแบบ

ถึงแม้ว่าวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมจะเป็นที่รู้จักในอีกชื่อว่า Social-based Filtering [15] อันเนื่องมาจากการกรองแบบพึ่งพาผู้ใช้ร่วมนั้นมีแนวคิดที่คล้ายกับการขอคำแนะนำจากผู้อื่นที่มีพฤติกรรมความนิยมคล้ายคลึงกัน แต่กลับไม่ได้มีการนำข้อมูลความสัมพันธ์ของบุคคลมาใช้ประโยชน์เท่าที่ควร โดยเฉพาะอย่างยิ่งข้อมูลความสัมพันธ์ในเครือข่ายสังคมออนไลน์ (Online Social Network: OSN) ซึ่งกำลังได้รับความนิยมอย่างสูงในเวลานี้ ในระบบเครือข่ายสังคมออนไลน์ผู้ใช้สามารถสร้างความสัมพันธ์กับผู้ใช้คนอื่นได้ ทำให้ระหว่างผู้ใช้แต่ละคนเกิดจุดเชื่องโยงข้อมูลซึ่งกันและกัน และหากมองในเชิงปริมาณข้อมูลเพียงอย่างเดียว การใช้ข้อมูลความสัมพันธ์บนเครือข่ายสังคมออนไลน์สามารถเพิ่มปริมาณข้อมูลได้อีกหลายเท่าตัว ซึ่งน่าจะเป็นประโยชน์ในการเพิ่มความหนาแน่นให้กับข้อมูล ดังนั้นงานวิจัยนี้จึงมีแนวความคิดที่จะนำข้อมูลความสัมพันธ์ดังกล่าวมาช่วยลดผลกระทบของปัญหาความเบาบางของข้อมูลที่มีต่อการกรองแบบพึ่งพาผู้ใช้ร่วมด้วยการประมาณค่าข้อมูลที่ขาดหายจากค่าความนิยมของผู้ใช้ที่มีความสัมพันธ์กันในเครือข่ายสังคมออนไลน์ โดยมีสมมติฐานที่ว่า บุคคลจะมีแนวโน้มความชอบเป็นไปตามกลุ่มเพื่อนของตนมากกว่าบุคคลทั่วไป โดยแนวคิดการทำงานของตัวแบบแสดงดังรูปที่ 3.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.1 แนวคิดการทำงานของตัวแบบการเติมค่าข้อมูลด้วยความสัมพันธ์ของผู้ใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.1 ข้อมูลนำเข้า

จากรูปที่ 3.1 แสดงแนวคิดการทำงานหลักของตัวแบบการเติมค่าข้อมูลด้วยความสัมพันธ์ของผู้ใช้ โดยข้อมูลนำเข้าสำหรับตัวแบบจะประกอบด้วยข้อมูลสองประเภทคือ 1. ข้อมูลการให้คะแนนชิ้นข้อมูลของผู้ใช้ และ 2. ข้อมูลความสัมพันธ์ของผู้ใช้ จากนั้นตัวแบบจะทำการเติมค่าข้อมูลที่ขาดหายจากการคำนวณค่าประมาณความนิยมโดยอาศัยข้อมูลนำเข้าทั้งสองประเภท เพื่อเพิ่มความหนาแน่นของข้อมูลก่อนที่จะเข้าสู่กระบวนการกรองแบบพึ่งพาผู้ใช้ร่วม เพื่อสร้างรายการแนะนำต่อไป

โดยทั่วไปข้อมูลการให้คะแนนชิ้นข้อมูลของผู้ใช้มักจะอยู่ในรูปแบบของเมตริกซ์ผู้ใช้-ชิ้นข้อมูล (User-Item Matrix) เพื่อแสดงค่าความนิยมของผู้ใช้ U แต่ละคนที่มีต่อชิ้นข้อมูล I โดยที่ U คือเซตของผู้ใช้ $U = \{U_1, U_2, \dots, U_m\}$ I คือเซตของชิ้นข้อมูล $I = \{I_1, I_2, \dots, I_n\}$ ส่วน m และ n คือจำนวนของผู้ใช้และจำนวนของชิ้นข้อมูลตามลำดับ ซึ่งค่าคะแนนความนิยมมักถูกแสดงในรูปแบบของตัวเลข เช่น คะแนนความนิยมอยู่ในช่วงหนึ่งถึงห้าตามระดับความชอบและไม่ชอบทั่วไปแล้วคะแนนความนิยมหนึ่งหมายถึงไม่ชอบอย่างมากและคะแนนความนิยมห้าหมายถึงชอบมาก ดังตัวอย่างการให้คะแนนชิ้นข้อมูลของผู้ใช้ตามตารางที่ 3.1

ตารางที่ 3.1 ตัวอย่างข้อมูลการให้คะแนนชิ้นข้อมูลของผู้ใช้แต่ละคน

สินค้า / ผู้ใช้	I_1	I_2	I_3	I_4	I_5
U_1	1	?	2	3	?
U_2	?	3	4	5	2
U_3	4	?	?	3	2
U_4	5	4	?	?	?
U_5	?	?	?	?	?

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 3.1 เป็นการแสดงข้อมูลการให้คะแนนของผู้ใช้แต่ละคน โดยในแต่ละแถวจะหมายถึงผู้ใช้ได้ให้คะแนนความนิยมในสินค้าแต่ละรายการ ตัวอย่างเช่นในแถวแรกหมายถึง ผู้ใช้ U_1 ให้คะแนนชิ้นข้อมูล I_1, I_3 และ I_4 ด้วยคะแนน 1,2 และ 3 ตามลำดับ และมีข้อมูลบางส่วนที่ยังว่างอยู่ ซึ่งเกิดจากการที่ผู้ใช้ไม่ได้ให้คะแนนความนิยมกับชิ้นข้อมูลนั้น จะสังเกตเห็นว่าผู้ใช้ U_5 ไม่ได้ทำการให้คะแนนชิ้นข้อมูลใดๆแม้แต่รายการเดียว ซึ่งแทบเป็นไปไม่ได้ที่วิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบปรกติจะทำการเปรียบเทียบความนิยมกับผู้ใช้คนอื่นๆ

ข้อมูลนำเข้าอีกส่วนหนึ่งก็คือข้อมูลความสัมพันธ์ของผู้ใช้ ซึ่งสามารถแสดงในรูปแบบของกราฟโดยจุดยอด (Vertex) แต่ละจุดแทนผู้ใช้แต่ละคน และเส้นเชื่อม (Edge) แทนความสัมพันธ์ของผู้ใช้ และสามารถแปลงให้อยู่ในรูปของเมตริกซ์ประชิด (Adjacency Matrix) โดย 0 หมายถึงระหว่างผู้ใช้ไม่มีความสัมพันธ์กัน และ 1 หมายถึงระหว่างผู้ใช้มีความสัมพันธ์กัน ดังแสดงในรูปที่ 3.2

ผู้ใช้ / ผู้ใช้	U_1	U_1	U_2	U_3	U_4
U_1	0	1	1	1	1
U_2	1	0	1	0	0
U_3	1	1	0	0	1
U_4	1	0	0	0	1
U_5	1	0	1	1	0

รูปที่ 3.2 ความสัมพันธ์ระหว่างผู้ใช้ในรูปแบบของกราฟ และเมตริกซ์ประชิด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.1.2 การเติมค่าข้อมูลที่ขาดหาย

จากสมมติฐานที่ว่า บุคคลจะมีแนวโน้มความนิยมเป็นไปตามกลุ่มเพื่อนของตน (ผู้ใช้ที่มีความสัมพันธ์กัน) ในขั้นตอนการเติมค่าความนิยมที่ขาดหายจึงได้นำค่าเฉลี่ยความเบี่ยงเบนความนิยมของกลุ่มเพื่อนของผู้ใช้มาใช้ในการเติมค่า ซึ่งได้แนวคิดจากวิธีการทำนายค่าความนิยมของการกรองแบบพึ่งพาผู้ใช้ร่วม [17] โดยตัดค่าถ่วงน้ำหนักความคล้ายของผู้ใช้ออกและพิจารณาจากผู้ใช้ที่มีความสัมพันธ์แทน ดังสมการที่ 3.1

$$E_{U,I} = \overline{R_U} + \frac{\sum_{V \in \Psi}(R_{V,I} - \overline{R_V})}{n_{uf}} \tag{3.1}$$

โดยที่ $E_{U,I}$ คือ ค่าประมาณความนิยมของผู้ใช้เป้าหมาย U ที่มีต่อสินค้า I ที่ใช้ในการเติมค่า

$\overline{R_U}$ คือ ค่าเฉลี่ยของคะแนนความนิยมของผู้ใช้เป้าหมาย U

$\overline{R_V}$ คือ ค่าเฉลี่ยของคะแนนความนิยมของผู้ใช้อ้างอิง V ที่มีความสัมพันธ์กับผู้ใช้เป้าหมาย

$\Psi = \{U_1, U_2, \dots, U_m\}$ คือเซตของผู้ใช้ทั้งหมด

n_{uf} คือ จำนวนเพื่อนของผู้ใช้เป้าหมาย U

เนื่องจากเราต้องการเติมค่าความนิยมเฉพาะกับข้อมูลที่ขาดหาย จึงกำหนดเงื่อนไขในการสร้างเมตริกซ์ที่เติมค่าแล้วตามสมการที่ 3.2

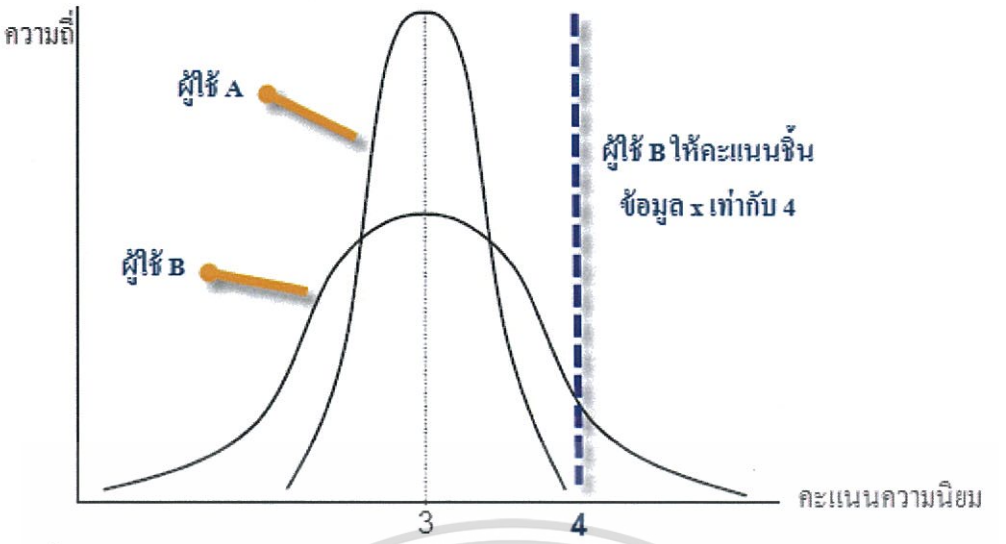
$$[x]_{U,I} \begin{cases} E_{U,I} & \text{if} \quad R_{U,I} = 0 \\ R_{U,I} & \text{if} \quad R_{U,I} > 0 \end{cases} \tag{3.2}$$

ในกรณีที่ไม่สามารถคำนวณค่าเฉลี่ยความนิยมของผู้ใช้ได้ เนื่องจากผู้ใช้ไม่ระบุคะแนนความนิยมในชิ้นข้อมูลใดๆ ให้ใช้ค่าเฉลี่ยความนิยมของผู้ใช้อ้างอิงที่มีความสัมพันธ์กับผู้ใช้เป้าหมายแทน ดังสมการที่ 3.3

$$E_{U,I} = \frac{\sum_{V \in \Psi}(R_{V,I} - \overline{R_V})}{n_{uf}} \tag{3.3}$$

จากสมการที่ 3.1 จะเป็นการใช้ค่าเฉลี่ยความเบี่ยงเบนความนิยมของกลุ่มเพื่อนมาใช้ในการประมาณความนิยมของผู้ใช้ โดยมิได้คำนึงถึงลักษณะการแจกแจงของการให้คะแนนของผู้ใช้แต่อย่างใด ซึ่งอาจทำให้เกิดความคลาดเคลื่อนในการทำนายคะแนนความนิยมได้หากผู้ใช้แต่ละคนมีการแจกแจงของการให้คะแนนแตกต่างกัน ดังตัวอย่างในรูปที่ 3.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 ตัวอย่างการแจกแจงการให้คะแนนของผู้ใช้ A และ B ที่มีการกระจายตัวแตกต่างกัน

ตัวอย่างในรูปที่ 3.3 เป็นตัวอย่างแสดงการแจกแจงข้อมูลของผู้ใช้เป้าหมาย A และผู้ใช้อ้างอิง B ที่มีการแจกแจงแบบปรกติ และมีค่าเฉลี่ยการให้คะแนนเท่ากับ 3 เท่ากัน แต่มีค่าส่วนเบี่ยงเบนมาตรฐานต่างกัน สมมติให้ผู้ใช้ A และ B มีความสัมพันธ์กัน และเราต้องการเติมค่าความชอบสำหรับชิ้นข้อมูล x ให้กับผู้ใช้ A โดยที่ผู้ใช้ B ให้คะแนนความนิยม 4 ให้กับชิ้นข้อมูล x ซึ่งใช้จากสมการที่ 3.1 การเติมค่าชิ้นข้อมูล x ให้แก่ผู้ใช้ A ค่าประมาณความนิยมจะเท่ากับ 4 แต่จากกราฟจะเห็นว่าผู้ใช้ A ให้คะแนนชิ้นข้อมูลด้วยคะแนน 4 เป็นอัตราส่วนที่น้อยมากเมื่อเทียบกับผู้ใช้ B ดังนั้นแทนที่จะใช้คะแนนดิบ เราจึงใช้ค่าคะแนนมาตรฐาน (Z-Score) ในการเปรียบเทียบคะแนนความนิยม โดยที่คะแนนมาตรฐานเป็นอัตราส่วนระหว่างความแตกต่างของคะแนนดิบกับค่าเฉลี่ยต่อความเบี่ยงเบนมาตรฐาน ดังสมการที่ 3.4

$$Z = \frac{X - \bar{X}}{SD} \tag{3.4}$$

โดยที่ Z คือ คะแนนมาตรฐาน

X คือ คะแนนที่ต้องการเปลี่ยนเป็นคะแนนมาตรฐาน

$\bar{X}$ คือ ค่าเฉลี่ยคะแนน

SD คือ ส่วนเบี่ยงเบนมาตรฐาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นเราจึงคำนวณคะแนนความนิยมของผู้ใช้เป้าหมายจากค่าคะแนนมาตรฐานของผู้ใช้เป้าหมายที่มีค่าเท่ากับคะแนนมาตรฐานของผู้ใช้อ้างอิง โดยใช้คะแนนความนิยมของผู้ใช้เป้าหมายเป็นคะแนนที่ต้องการเปลี่ยนเป็นคะแนนมาตรฐาน ดังสมการ 3.5

$$\frac{R_U - \overline{R_U}}{SD_U} = \frac{R_V - \overline{R_V}}{SD_V} \tag{3.5}$$

ดังนั้นคะแนนความนิยมของผู้ใช้เป้าหมาย U (R_U) ที่คำนวณจากผู้ใช้อ้างอิง V (R_V) คือ

$$R_U = \overline{R_U} + \left(\frac{R_{V,I} - \overline{R_V}}{SD_V}\right) \cdot SD_U \tag{3.6}$$

ค่าประมาณความนิยมของผู้ใช้เป้าหมาย U ที่มีต่อสินค้า I ที่ใช้ในการเติมค่า คำนวณจากค่าเฉลี่ยของคะแนนความนิยมในสินค้า I ของผู้ใช้อ้างอิงที่มีความสัมพันธ์กับผู้ใช้เป้าหมาย โดยปรับคะแนนความชอบตามคะแนนมาตรฐานของผู้ใช้งานเป้าหมาย ดังสมการที่ 3.7

$$\mathrm{E}_{\mathrm{U,I}} = \frac{\Sigma_{\mathrm{V}\in\Psi}\overline{R_U} + \left(\frac{R_{V,I} - \overline{R_V}}{SD_V}\right) \cdot SD_U}{\mathrm{n_{uf}}} \tag{3.7}$$

โดยที่ $\mathrm{E}_{\mathrm{U,I}}$ คือ ค่าประมาณความชอบของผู้ใช้ U ที่มีต่อสินค้า I ที่ใช้ในการเติมค่า

$\overline{\mathrm{R_U}}$ คือ ค่าเฉลี่ยของคะแนนความชอบในสินค้าของผู้ใช้ U

$\overline{\mathrm{R_V}}$ คือ ค่าเฉลี่ยของคะแนนความชอบในสินค้าของผู้ใช้ V คนอื่นๆ

Ψ คือ เซ็ตของผู้ใช้งานทั้งหมด

$\mathrm{n_{uf}}$ คือ จำนวนเพื่อนของผู้ใช้ U

SD_U และ SD_V คือ ส่วนเบี่ยงเบนมาตรฐานการให้คะแนนของผู้ใช้ U และผู้ใช้ V

เนื่องจากโดยปรกติแล้วผู้ใช้จะทำการให้คะแนนสินค้าเพียงไม่กี่รายการเท่านั้น ซึ่งการนำข้อมูลของผู้ใช้ที่ทำการให้คะแนนสิ้นค้าเพียงไม่กี่รายการไปใช้ อาจทำให้เกิดความคลาดเคลื่อนได้ ดังนั้นในการคำนวณจึงควรให้ความน่าเชื่อถือของผู้ใช้ที่ให้คะแนนสินค้าหลายรายการมากกว่าผู้ใช้ที่ให้คะแนนสินค้าเพียงไม่กี่รายการ จึงได้มีการปรับปรุงสมการ 3.8 โดยการเพิ่มค่าถ่วงน้ำหนักสำหรับผู้ใช้แต่ละคนดังสมการที่ 3.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$E_{U,I} = \frac{\Sigma_{V \in \Psi} W_V \cdot \left(\overline{R_U} + \left(\frac{R_{V,I} - \overline{R_V}}{SD_V} \right) \cdot SD_U \right) + (1 - W_V) \cdot \overline{R_I}}{\mathrm{n_{uf}}} \quad (3.8)$$

โดยที่ $\overline{R_I}$ คือค่าเฉลี่ยการให้คะแนนสำหรับสินค้า I และค่าถ่วงน้ำหนักความน่าเชื่อถือ (W_V) คำนวณได้จากสมการที่ 3.9

$$W_V = \begin{cases} \frac{c_V}{Threshold} & if \quad c_V < Threshold \\ 1 & if \quad c_V > Threshold \end{cases} \quad (3.9)$$

โดยที่ c_V คือจำนวนสินค้าที่ผู้ใช้อ้างอิง V ได้ทำการให้คะแนน

$Threshold$ คือค่าคงที่สำหรับกำหนดค่าความน่าเชื่อถือ

ขั้นตอนการทำงานของตัวแบบในการเติมค่าข้อมูลที่ขาดหายดังแสดงในรูปที่ 3.4

รูปที่ 3.4 ขั้นตอนการทำงานของตัวแบบในการเติมค่าข้อมูลที่ขาดหาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 ตัวอย่างการทำงานของตัวแบบ

ในหัวข้อนี้จะแสดงตัวอย่างการเติมค่าข้อมูลที่ขาดหายจากข้อมูลตัวอย่างการให้คะแนนในตารางที่ 3.2 และข้อมูลความสัมพันธ์ของผู้ใช้จากตารางที่ 3.3 โดยสามารถแบ่งตามกระบวนการทำงานได้เป็น 5 ขั้นตอนดังรูปที่ 3.4 ดังนี้

ตารางที่ 3.2 ตัวอย่างการให้คะแนนชิ้นข้อมูลของผู้ใช้สำหรับแสดงการเติมค่า

สินค้า / ผู้ใช้	I_1	I_2	I_3	I_4	I_5
U_1	1	?	2	3	?
U_2	?	3	4	5	2
U_3	4	?	?	3	2
U_4	5	4	?	?	?
U_5	?	?	?	?	?

ตารางที่ 3.3 ตัวอย่างข้อมูลความสัมพันธ์ของผู้ใช้สำหรับแสดงการเติมค่า

สินค้า / ผู้ใช้	U_1	U_2	U_3	U_4	U_5
U_1	0	1	1	1	1
U_2	1	0	1	0	0
U_3	1	1	0	0	1
U_4	1	0	0	0	1
U_5	1	0	1	1	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.1 คำนวณค่าเฉลี่ยการให้คะแนนของผู้ใช้แต่ละคน

ค่าเฉลี่ยการให้คะแนนของผู้ใช้แต่ละคนทำได้โดยใช้สมการที่ 3.10 ดังนี้

$$\overline{R_U} = \frac{\sum_{i=1}^{n} R_{U,I}}{n} \tag{3.10}$$

โดยที่ $\overline{R_U}$ คือค่าเฉลี่ยการให้คะแนนของผู้ใช้ U

$R_{U,I}$ คือคะแนนความนิยมของผู้ใช้ U ที่มีต่อชิ้นข้อมูล I

n คือจำนวนชิ้นข้อมูลที่ผู้ใช้ U ทำการให้คะแนนความนิยม

คำนวณได้ดังนี้

$\overline{R_{U_1}} = \frac{1+2+3}{3} = 2$

$\overline{R_{U_2}} = \frac{2+4+5+2}{4} = 3.25$

$\overline{R_{U_3}} = \frac{4+3+2}{3} = 3$

$\overline{R_{U_4}} = \frac{5+4}{2} = 4.5$

$\overline{R_{U_5}}$ ไม่สามารถคำนวณได้ เนื่องจากผู้ใช้คนที่ 5 ไม่มีการให้คะแนนชิ้นข้อมูล

3.2.2 คำนวณค่าส่วนเบี่ยงเบนมาตรฐานของผู้ใช้แต่ละคน

ค่าส่วนเบี่ยงเบนมาตรฐานคำนวณจากสมการที่ 3.11 ดังนี้

$$SD_U = \sqrt{\frac{\sum_{I \in \check{I}} \left(R_{U,I} - \overline{R_U}\right)^2}{n}} \tag{3.11}$$

โดยที่ SD_U คือส่วนเบี่ยงเบนมาตรฐานและ $\check{I}$ คือเซตของชิ้นข้อมูลที่ผู้ใช้ U ทำการให้คะแนนคำนวณได้ดังนี้

$SD_1 = \sqrt{\frac{(1-2)^2+(2-2)^2+(3-2)^2}{3}} = 0.82$

$SD_2 = \sqrt{\frac{(2-3.25)^2+(4-3.25)^2+(5-3.25)^2+(2-3.25)^2}{4}} = 1.28$

$SD_3 = \sqrt{\frac{(4-3)^2+(3-3)^2+(2-3)^2}{3}} = 0.82$

$SD_4 = \sqrt{\frac{(4-4.5)^2+(5-4.5)^2}{2}} = 0.5$

$SD_5 =$ ไม่สามารถคำนวณได้ เนื่องจากผู้ใช้คนที่ 5 ไม่มีการให้คะแนนชิ้นข้อมูล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2.3 คำนวณค่าเฉลี่ยคะแนนความนิยมของชิ้นข้อมูล

ค่าเฉลี่ยคะแนนความนิยมของชิ้นข้อมูลคำนวณได้ดังสมการที่ 3.12 ดังนี้

$$\overline{R_I} = \frac{\Sigma_{U \in \Psi} R_{U,I}}{n} \tag{3.12}$$

โดยที่ $\overline{R_I}$ คือค่าเฉลี่ยความนิยมของชิ้นข้อมูล I

Ψ คือเซตของผู้ใช้ที่ทำการให้คะแนนชิ้นข้อมูล I

$R_{U,I}$ คือคะแนนความนิยมของผู้ใช้ U ที่มีต่อชิ้นข้อมูล I

n คือจำนวนผู้ใช้ที่ทำการให้คะแนนชิ้นข้อมูล I

คำนวณได้ดังนี้

$$\overline{R_{I_1}} = \frac{1+4+5}{3} = 3.33$$
$$\overline{R_{I_2}} = \frac{3+4}{2} = 3.5$$
$$\overline{R_{I_3}} = \frac{2+4}{2} = 3$$
$$\overline{R_{I_4}} = \frac{3+5+3}{3} = 3.67$$
$$\overline{R_{I_5}} = \frac{2+2}{2} = 2$$

3.2.4 คำนวณค่าถ่วงน้ำหนักความน่าเชื่อถือของผู้ใช้

ในงานวิจัยนี้ใช้สมการที่ 3.8 ในการคำนวณค่าถ่วงน้ำหนักความน่าเชื่อถือ และกำหนดค่าคงที่ Threshold เท่ากับ 3 สำหรับตัวอย่างการคำนวณค่าถ่วงน้ำหนักความน่าเชื่อถือ

คำนวณได้ดังนี้

$$W_{U_1} = \frac{3}{3} = 1$$
$$W_{U_2} = 1$$
$$W_{U_3} = \frac{3}{3} = 1$$
$$W_{U_4} = \frac{2}{3} = 0.67$$
$$W_{U_5} = \frac{0}{3} = 0$$

3.2.5 สร้างเมตริกซ์การให้คะแนนที่เติมค่าแล้ว

การเติมค่าข้อมูลที่ขาดหายด้วยความสัมพันธ์ของผู้ใช้เพื่อสร้างเมตริกซ์การให้คะแนนที่เติมค่าแล้วนั้นอาศัยสมการที่ 3.2 และ 3.8 ในการสร้างเมตริกซ์ คำนวณได้ดังนี้

สำหรับผู้ใช้คนที่ 1 ชิ้นข้อมูลที่ 1

$$[x]_{1,1} = R_{1,1}$$
$$[x]_{1,1} = 1$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับผู้ใช้คนที่ 1 ชิ้นข้อมูลที่ 2

$$[x]_{1,2} = E_{1,2}$$

$$[x]_{1,2} = \frac{\left(1\left(2+\left(\frac{3-3.25}{1.28}\right)0.82\right)+(0)(3.5)\right)+\left(0.67\left(2+\left(\frac{4-4.5}{0.82}\right)0.82\right)+(0.33)(3.5)\right)}{2}$$

$$[x]_{1,2} = 2$$

สำหรับผู้ใช้คนที่ 1 ชิ้นข้อมูลที่ 3

$$[x]_{1,3} = R_{1,3}$$

$$[x]_{1,3} = 2$$

สำหรับผู้ใช้คนที่ 1 ชิ้นข้อมูลที่ 4

$$[x]_{1,4} = R_{1,4}$$

$$[x]_{1,4} = 3$$

สำหรับผู้ใช้คนที่ 1 ชิ้นข้อมูลที่ 5

$$[x]_{1,5} = E_{1,5}$$

$$[x]_{1,5} = \frac{\left(1\left(2+\left(\frac{2-3.25}{1.28}\right)0.82\right)+(0)(2)\right)+\left(1\left(2+\left(\frac{2-3}{0.82}\right)0.82\right)+(0)(2)\right)}{2}$$

$$[x]_{1,5} = 1.1$$

สำหรับผู้ใช้คนที่ 2 ชิ้นข้อมูลที่ 1

$$[x]_{2,1} = E_{2,1}$$

$$[x]_{2,1} = \frac{\left(1\left(3.25+\left(\frac{1-2}{0.82}\right)1.28\right)+(0)(3.33)\right)+\left(1\left(3.25+\left(\frac{4-3}{0.82}\right)1.28\right)+(0)(3.33)\right)}{2}$$

$$[x]_{2,1} = 3.25$$

สำหรับผู้ใช้คนที่ 2 ชิ้นข้อมูลที่ 2

$$[x]_{2,2} = R_{2,2}$$

$$[x]_{2,2} = 3$$

สำหรับผู้ใช้คนที่ 2 ชิ้นข้อมูลที่ 3

$$[x]_{2,3} = R_{2,3}$$

$$[x]_{2,3} = 4$$

สำหรับผู้ใช้คนที่ 2 ชิ้นข้อมูลที่ 4

$$[x]_{2,4} = R_{2,4}$$

$$[x]_{2,4} = 5$$

สำหรับผู้ใช้คนที่ 2 ชิ้นข้อมูลที่ 5

$$[x]_{2,5} = R_{2,5}$$

$$[x]_{2,5} = 2$$

สำหรับผู้ใช้คนที่ 3 ชิ้นข้อมูลที่ 1

$$[x]_{3,1} = R_{3,1}$$

$$[x]_{3,1} = 4$$

สำหรับผู้ใช้คนที่ 3 ชิ้นข้อมูลที่ 2

$$[x]_{3,2} = E_{3,2}$$

$$[x]_{3,2} = \left(1\left(3+\left(\frac{3-3.25}{1.28}\right)0.82\right)+(0)(3.5)\right)$$

$$[x]_{3,2} = 3.16$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับผู้ใช้คนที่ 3 ชิ้นข้อมูลที่ 3

$$[x]_{3,3} = E_{3,3}$$

$$[x]_{3,3} = \frac{\left(1\left(3+\left(\frac{2-2}{0.82}\right)0.82\right)+(0)(3)\right)+\left(1\left(3+\left(\frac{4-3.25}{1.28}\right)0.82\right)+(0)(3)\right)}{2}$$

$$[x]_{3,3} = 3.24$$

สำหรับผู้ใช้คนที่ 3 ชิ้นข้อมูลที่ 4

$$[x]_{3,4} = R_{3,4}$$

$$[x]_{3,4} = 3$$

สำหรับผู้ใช้คนที่ 3 ชิ้นข้อมูลที่ 5

$$[x]_{3,5} = R_{3,5}$$

$$[x]_{3,5} = 2$$

สำหรับผู้ใช้คนที่ 4 ชิ้นข้อมูลที่ 1

$$[x]_{4,1} = R_{4,1}$$

$$[x]_{4,1} = 5$$

สำหรับผู้ใช้คนที่ 4 ชิ้นข้อมูลที่ 2

$$[x]_{4,2} = R_{4,2}$$

$$[x]_{4,2} = 4$$

สำหรับผู้ใช้คนที่ 4 ชิ้นข้อมูลที่ 3

$$[x]_{4,3} = E_{4,3}$$

$$[x]_{4,3} = \left(1\left(4.5+\left(\frac{2-2}{0.82}\right)0.5\right)+(0)(3)\right)$$

$$[x]_{4,3} = 4.5$$

สำหรับผู้ใช้คนที่ 4 ชิ้นข้อมูลที่ 4

$$[x]_{4,4} = E_{4,4}$$

$$[x]_{4,4} = \left(1\left(4.5+\left(\frac{3-2}{0.82}\right)0.5\right)+(0)(3)\right)$$

$$[x]_{4,4} = 5.11$$

สำหรับผู้ใช้คนที่ 4 ชิ้นข้อมูลที่ 5

$$[x]_{4,5} = \overline{R_{I_5}}$$

$$[x]_{4,5} = 2$$

สำหรับผู้ใช้คนที่ 5 ชิ้นข้อมูลที่ 1

$$[x]_{5,1} = E_{5,1}$$

$$[x]_{5,1} = \frac{1+4+5}{3}$$

$$[x]_{5,1} = 3.33$$

สำหรับผู้ใช้คนที่ 5 ชิ้นข้อมูลที่ 2

$$[x]_{5,2} = E_{5,2}$$

$$[x]_{5,2} = 4$$

สำหรับผู้ใช้คนที่ 5 ชิ้นข้อมูลที่ 3

$$[x]_{5,3} = E_{5,3}$$

$$[x]_{5,3} = 2$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำหรับผู้ใช้คนที่ 5 ชิ้นข้อมูลที่ 4

$$[x]_{5,4} = E_{5,4}$$
$$[x]_{5,4} = \frac{3+3}{2}$$
$$[x]_{5,4} = 3$$

สำหรับผู้ใช้คนที่ 5 ชิ้นข้อมูลที่ 5

$$[x]_{5,5} = E_{5,5}$$
$$[x]_{5,5} = 2$$

ดังนั้นเมตริกซ์ใหม่ที่เติมค่าด้วยวิธีการการเติมค่าข้อมูลที่ขาดหายโดยอาศัยข้อมูลความสัมพันธ์ระหว่างผู้ใช้ จากข้อมูลตัวอย่างแสดงดังตารางที่ 3.4

ตารางที่ 3.4 ผลการเติมค่าความนิยมที่ขาดหายโดยอาศัยความสัมพันธ์ของผู้ใช้

สินค้า / ผู้ใช้	I_1	I_2	I_3	I_4	I_5
U_1	1	**2**	2	3	**1.1**
U_2	**3.25**	3	4	5	2
U_3	4	**3.16**	**3.24**	3	2
U_4	5	4	**4.5**	**5.11**	?
U_5	**3.33**	**4**	**2**	**3**	**2**

จากนั้นนำเมตริกซ์ที่ผ่านการเติมค่าข้อมูลที่ขาดหายโดยอาศัยข้อมูลความสัมพันธ์ของผู้ใช้เข้าสู่กระบวนการกรองแบบพึ่งพาผู้ใช้ร่วมตามปรกติต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
ผลการทดลอง

ในบทนี้จะกล่าวถึงชุดข้อมูลที่ใช้ในการทดลอง วิธีการทดลอง และผลการทดลอง ที่ใช้ในการทดสอบเพื่อพิสูจน์สมมติฐานที่ได้กล่าวไว้ในบทก่อนหน้า และวัดประสิทธิภาพของตัวแบบ ในการลดผลกระทบของปัญหาความเบาบางของข้อมูลในวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering: CF) ด้วยการเติมค่าข้อมูลที่ขาดหายโดยอาศัยข้อมูลความสัมพันธ์ของผู้ใช้ในเครือข่ายสังคมออนไลน์ ทั้งในด้านประสิทธิภาพในการทำนาย และความทนทานของตัวแบบที่มีต่อระดับความเบาบางของข้อมูล

4.1 ชุดข้อมูลที่ใช้ในการทดลอง

ชุดข้อมูลที่ใช้ในการทดลองเป็นข้อมูลที่สกัดมาจาก เว็บไซต์เครือข่ายสังคมออนไลน์ Yelp.com ซึ่งเป็นเว็บไซต์ที่มีชื่อเสียง และเป็นแหล่งรวบรวมบทวิจารณ์ของร้านค้าและบริการต่างๆ ที่หลากหลาย เช่น ร้านอาหาร, แหล่งจับจ่ายใช้สอย, โรงแรม, สปาร์ และบริการทางการเงิน เป็นต้น นอกจากผู้ใช้สามารถเข้ามาหาข้อมูลอ่านบทวิจารณ์ของร้านค้าแล้ว ยังสามารถที่จะเขียนคำวิจารณ์รวมไปถึงให้คะแนนสำหรับร้านค้าที่ตนเองเคยเข้าไปใช้บริการได้อีกด้วย ที่สำคัญผู้ใช้งานที่เป็นสมาชิกของ Yelp.com สามารถที่จะสร้างความสัมพันธ์กับสมาชิกอื่นด้วยการยื่นคำขอเป็นเพื่อน หรือแม้กระทั่งสามารถเชิญบุคคลภายนอกให้เข้ามาเป็นสมาชิก และเป็นเพื่อนของตนได้ด้วยการส่งอีเมล์ ซึ่งทำให้ Yelp.com มีลักษณะเป็นเครือข่ายสังคมอีกด้วย

ข้อมูลที่สกัดจาก Yelp.com สกัดมา ณ วันที่ 16 เมษายน พ.ศ. 2556 เป็นข้อมูลของร้านอาหารไทย ในเมืองซอลต์เลกซิตี รัฐยูทาห์ ประเทศสหรัฐอเมริกาจำนวน 20 ร้าน ผู้เขียนบทวิจารณ์และให้คะแนนจำนวน 632 คน บทวิจารณ์จำนวน 1,355 บทวิจารณ์ มีข้อมูลความสัมพันธ์ระหว่างสมาชิกจำนวน 2,356 ความสัมพันธ์ โดยผู้เขียนบทวิจารณ์จะเขียนบทวิจารณ์ร้านอาหารเฉลี่ย 2.14 บทวิจารณ์ต่อคน 67.75 บทวิจารณ์ต่อร้านค้า และผู้ให้บทวิจารณ์แต่ละคนมีความสัมพันธ์กับผู้ให้บทวิจารณ์คนอื่นโดยเฉลี่ย 3.73 ความสัมพันธ์ต่อคน

นอกจากนี้เพื่อเป็นการตอบคำถามที่ว่า บุคคลที่มีความสัมพันธ์กันจะมีแนวโน้มความนิยมไปในทางเดียวกัน มากกว่าบุคคลที่ไม่มีความสัมพันธ์กันหรือไม่ จึงได้ทำการเปรียบเทียบผลต่างของค่าเฉลี่ยของการให้คะแนนร้านอาหารแต่ละร้าน ระหว่างผู้ให้คะแนนที่มีความสัมพันธ์กันและผู้ให้คะแนนที่ไม่มีความสัมพันธ์กันพบว่า สำหรับผู้ให้คะแนนที่มีความสัมพันธ์กันจะให้คะแนนผลต่างกันเฉลี่ย 0.88 คะแนน และค่าความเบี่ยงเบนอยู่ที่ 0.89 สำหรับผู้ให้คะแนนที่ไม่มีความสัมพันธ์กันจะให้คะแนนผลต่างเฉลี่ย 1.13 และมีความความเบี่ยงเบน 1.01 และคะแนนผลต่างเฉลี่ย 1.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คะแนน ที่ค่าความเบี่ยงเบน 0.98 เมื่อพิจารณาจากผู้ใช้ทั้งหมดโดยไม่คำนึงว่าผู้ให้คะแนนมีความสัมพันธ์กันหรือไม่ ดังแสดงในตารางที่ 4.1 ซึ่งจากตัวเลขดังกล่าวจะเห็นว่า โดยเฉลี่ยแล้วผู้ให้คะแนนที่มีความสัมพันธ์กันมีความใกล้เคียงกันมากกว่า

ตารางที่ 4.1: ค่าเฉลี่ยผลต่างการให้คะแนนและค่าความเบี่ยงเบนสำหรับความสัมพันธ์แบบต่างๆ

เงื่อนไข	ค่าเฉลี่ยผลต่าง	ค่าความเบี่ยงเบน
จากผู้ให้คะแนนทั้งหมด	1.10	0.98
ผู้ให้คะแนนที่มีความสัมพันธ์กัน	**0.88**	**0.89**
ผู้ให้คะแนนที่ไม่มีความสัมพันธ์กัน	1.13	1.01

จากบทที่ 3 วิธีการเติมค่าที่ขาดหายของตัวแบบที่นำเสนอ มีสมมติฐานว่าการกระจายตัวของการให้คะแนน มีการกระจายตัวแบบปรกติ (Normal Distribution) จึงได้ทำการทดสอบการกระจายตัวของชุดข้อมูลด้วยวิธีการทดสอบ โคลโมโกรอฟ-สเมียร์นอฟ (Kolmogorov – Smirnov Test) ด้วยโปรแกรม SPSS โดยทำการทดสอบที่ความเชื่อมั่น 95% พบว่าชุดข้อมูลมีการแจกแจงแบบปรกติ

4.2 วิธีการทดลองและการวัดประสิทธิภาพ

ในการทดลองสำหรับการวัดประสิทธิภาพของตัวแบบที่นำเสนอ ใช้วิธีการแบบ Leave-One-Out Cross-Validation (LOOCV) ในการทดสอบโดยจะทำการดึงข้อมูลการให้คะแนนความนิยมของผู้ใช้ออก 1 รายการให้เป็นกลุ่มข้อมูลทดสอบ (Testing Set) ซึ่งเป็นข้อมูลที่จะให้ตัวแบบทำนาย ส่วนข้อมูลที่เหลือทั้งหมดจะใช้เป็นกลุ่มข้อมูลการเรียนรู้ (Training Set) จากนั้นจะการทดสอบซ้ำโดยเปลี่ยนกลุ่มข้อมูลทดสอบจนครบทุกรายการ และใช้ตัวชี้วัดด้วยค่าความคลาดเคลื่อนสัมบูรณ์ [19] (Mean Absolute Error: MAE) ในการประเมินประสิทธิภาพความแม่นยำในการทำนาย ซึ่งวิธีการนี้จะทำการหาค่าเฉลี่ยความผิดพลาดในการทำนายแต่ละรายการ ดังสมการที่ 4.1

$$MAE = \frac{\Sigma_{U,I}|R_{U,I} - R'_{U,I}|}{L} \tag{4.1}$$

โดยที่ $R_{U,I}$ คือคะแนนความนิยมของผู้ใช้ U ที่มีต่อสินค้า I , $R'_{U,I}$ คือค่าประมาณคะแนนความนิยมของผู้ใช้ U ที่มีต่อสินค้า U และ L คือจำนวนรายการที่ใช้ในการทดสอบ ซึ่งค่าความคลาดเคลื่อนสัมบูรณ์ยิ่งน้อยค่าความแม่นยำในการทำนายยิ่งสูง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เนื่องจากวิธีการเติมค่าข้อมูลที่ขาดหายของตัวแบบ ได้นำค่าคะแนนมาตรฐาน (Z-Score) ในการเปรียบเทียบคะแนนความนิยมแทนการใช้คะแนนดิบ จึงได้เปรียบเทียบประสิทธิภาพของตัวแบบเมื่อใช้คะแนนมาตรฐานและคะแนนดิบในการเติมค่าข้อมูลที่ขาดหายดังตารางที่ 4.3 และกำหนดค่าคงที่ *Threshold* ซึ่งใช้สำหรับคำนวณค่าความน่าเชื่อถือของผู้ใช้ โดยกำหนดค่าคงที่ *Threshold* เท่ากับ 7 สำหรับตัวแบบที่ใช้ในการทดลอง

ตารางที่ 4.2: เปรียบเทียบประสิทธิภาพของตัวแบบเมื่อใช้คะแนนมาตรฐานและคะแนนดิบ

วิธีการ	**MAE**
เมื่อใช้คะแนนมาตรฐานในการคำนวณสำหรับการเติมค่าข้อมูลที่ขาดหาย	**0.689**
เมื่อใช้คะแนนดิบในการคำนวณสำหรับเติมค่าข้อมูลที่ขาดหาย	0.697

จากผลการเปรียบเทียบประสิทธิภาพของตัวแบบเมื่อใช้คะแนนมาตรฐานและค่าคะแนนดิบในการคำนวณสำหรับเติมค่าข้อมูลที่ขาดหายในตารางที่ 4.2 แสดงให้เห็นว่า การใช้ค่าคะแนนมาตรฐานในการคำนวณจะให้ค่าความแม่นยำมากกว่าการใช้ค่าคะแนนดิบ

สำหรับการทดสอบประสิทธิภาพของตัวแบบ ได้ทำการเปรียบเทียบประสิทธิภาพของตัวแบบซึ่งจะมีขั้นตอนการเติมค่าข้อมูลที่ขาดหายไปด้วยข้อมูลความสัมพันธ์ของผู้ใช้ ก่อนที่จะเข้ากระบวนการวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม กับการเติมค่าข้อมูลที่ขาดหายโดยใช้ค่าเฉลี่ยของสินค้า และ วิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบปรกติ ดังแสดงในตารางที่ 4.3

ตารางที่ 4.3: ค่าความคลาดเคลื่อนสัมบูรณ์ระหว่างวิธีการต่างๆที่ใช้ในการทดลอง

วิธีการ	**MAE**
วิธีที่นำเสนอ	**0.689**
CF แบบเติมค่าข้อมูลที่ขาดหายด้วยค่าเฉลี่ยของชิ้นข้อมูล	0.714
CF แบบปรกติ	0.769

นอกจากนี้ทางคณะผู้วิจัยได้ทำการทดลองเปรียบเทียบประสิทธิภาพในการเติมค่าข้อมูลที่ขาดหายด้วยความสัมพันธ์ของผู้ใช้ดังนี้

1) เติมค่าโดยประมาณค่าความนิยมจากผู้ใช้ที่มีความสัมพันธ์กัน

2) เติมค่าที่ขาดหายโดยประมาณค่าความนิยมจากผู้ใช้ที่ไม่มีความสัมพันธ์กัน

3) เติมค่าข้อมูลที่ขาดหายโดยประมาณค่าความนิยมจากผู้ใช้คนอื่นโดยไม่คำนึงว่ามีความสัมพันธ์กันหรือไม่ ดังตารางที่ 4.4

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

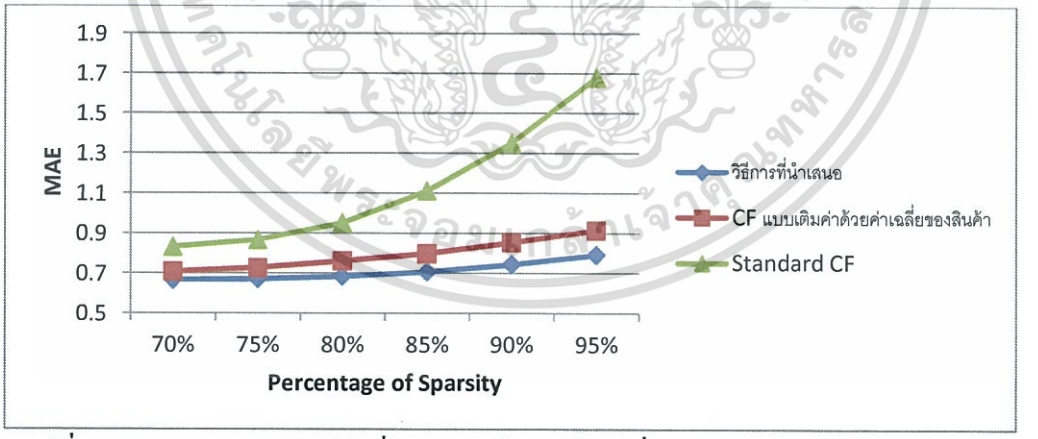
ตารางที่ 4.4: ค่าความคลาดเคลื่อนสัมบูรณ์จากการทดลองด้วยความสัมพันธ์แบบต่างๆ

วิธีการ	MAE
เติมค่าโดยประมาณค่าความนิยมจากผู้ใช้ที่มีความสัมพันธ์กัน	**0.689**
เติมค่าที่ขาดหายโดยประมาณค่าความนิยมจากผู้ใช้ที่ไม่มีความสัมพันธ์กัน	0.716
เติมค่าข้อมูลที่ขาดหายโดยประมาณค่าความนิยมจากผู้ใช้ทั้งหมด	0.714

จากผลการทดลองในตารางที่ 4.4 แสดงให้เห็นว่าการเติมค่าข้อมูลที่ขาดหายด้วยข้อมูลความสัมพันธ์ของผู้ใช้ที่มีความสัมพันธ์กันให้ความแม่นยำมากที่สุด เมื่อเปรียบเทียบกับการเติมข้อมูลที่ขาดหายด้วยข้อมูลความสัมพันธ์ของผู้ใช้ที่มีความสัมพันธ์กัน และ การเติมค่าข้อมูลที่ขาดหายด้วยข้อมูลของผู้ใช้ทั่วไป (ไม่คำนึงว่าผู้ใช้มีความสัมพันธ์กันหรือไม่) ซึ่งเป็นอีกข้อพิสูจน์หนึ่งสำหรับสมมติฐานของเราที่ว่า บุคคลจะมีแนวโน้มความนิยมเป็นไปตามกลุ่มเพื่อนของตนมากกว่าบุคคลทั่วไป

4.3 ความทนทานต่อความเบาบางของข้อมูล

นอกจากความแม่นยำในการทำนายแล้ว ความทนทานต่อความเบาบางของข้อมูลก็เป็นอีกประเด็นหนึ่งที่มีความสำคัญ ในส่วนนี้จะกล่าวถึงผลการทดสอบประสิทธิภาพของวิธีการต่างๆ กับระดับความเบาบางของข้อมูลในหลายๆระดับ



รูปที่ 4.1 ค่า MAE ของแต่ละวิธี เมื่อทดสอบด้วยชุดข้อมูลที่ระดับความเบาบางในระดับต่างๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากกราฟในรูปที่ 4.1 แสดงให้เห็นว่า การเปลี่ยนแปลงระดับความเบาบางของข้อมูล มีผลกระทบกับวิธีการที่นำเสนอน้อยที่สุด ซึ่งนั่นหมายความว่าวิธีการที่นำเสนอมีความทนทานต่อความเบาบางของข้อมูลมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปและข้อเสนอแนะ

5.1 สรุป

งานวิจัยนี้ต้องการที่จะลดผลกระทบของปัญหาความเบาบางของข้อมูลที่มีต่อวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม ทั้งในด้านของความแม่นยำในการทำนาย และความทนทานต่อความเบาบางของข้อมูล โดยการใช้ประโยชน์จากข้อมูลความสัมพันธ์ของผู้ใช้ด้วยการเติมค่าข้อมูลที่ขาดหายโดยอาศัยข้อมูลประวัติพฤติกรรมการบริโภคร่วมกับข้อมูลความสัมพันธ์ของผู้ใช้ จากสมมติฐานของผู้วิจัยที่ว่าบุคคลมักจะมีแนวโน้มความชอบเป็นไปตามกลุ่มเพื่อนของตน

เพื่อที่จะพิสูจน์สมมติฐานดังกล่าว จึงได้ทำการสกัดข้อมูลจากเว็บไซต์ Yelp.com ซึ่งเป็นเว็บไซต์ที่รวบรวมบทวิจารณ์เพื่อแลกเปลี่ยนความคิดเห็นสำหรับร้านค้าต่างๆ โดยเฉพาะอย่างยิ่งร้านอาหารซึ่งเป็นประเภทร้านค้าที่ได้รับความนิยมมากที่สุด โดยเว็บไซต์ Yelp.com มีคุณลักษณะพิเศษอีกประการหนึ่งคือ มีลักษณะเป็นเครือข่ายสังคมออนไลน์ ซึ่งสามารถบอกถึงความสัมพันธ์ของผู้ใช้ได้ และจากการวิเคราะห์ข้อมูลที่ได้สกัดออกมาด้วยค่าเฉลี่ยผลต่างการให้คะแนนของผู้ใช้ พบว่า ผู้ใช้ที่มีความสัมพันธ์กันมีค่าเฉลี่ยผลต่างเท่ากับ 0.88 ซึ่งน้อยกว่าผู้ใช้ที่ไม่มีความสัมพันธ์กันโดยมีค่าเฉลี่ยผลต่างเท่ากับ 1.13 และ 1.10 สำหรับผู้ใช้ทั่วไป

จากการทดสอบเปรียบเทียบประสิทธิภาพของตัวแบบกับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบปรกติและวิธีการเติมค่าข้อมูลที่ขาดหายด้วยค่าเฉลี่ย โดยทดสอบด้วยค่าความคลาดเคลื่อนสัมบูรณ์ พบว่าตัวแบบที่นำเสนอมีค่าความคลาดเคลื่อนสัมบูรณ์ต่ำที่สุด และสามารถเพิ่มประสิทธิภาพของวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมปรกติด้วยการลดค่าความคลาดเคลื่อนสมบูรณ์ได้ถึง 10.4% นอกจากนั้นในการทดสอบประสิทธิภาพการทำนายด้วยการทดสอบกับชุดข้อมูลที่มีระดับความเบาบางของข้อมูลในระดับต่างๆพบว่า ตัวแบบที่นำเสนอมีความทนทานต่อความเบาบางของข้อมูลดีที่สุดเมื่อเปรียบเทียบกับวิธีการอื่นๆที่ใช้ในการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ข้อเสนอแนะ

จะเห็นได้ว่าวิธีการเติมข้อมูลที่นำเสนอไปนั้นอาศัยเพียงข้อมูลความสัมพันธ์บนเครือข่ายสังคมออนไลน์เพียงอย่างเดียว ทั้งๆที่ในความเป็นจริงแล้วยังมีข้อมูลอื่นๆบนเครือข่ายสังคมออนไลน์ที่น่าสนใจอีกมาก ทั้งข้อมูลส่วนตัว เช่น เพศ, อายุ, ระดับการศึกษา และอาชีพ เป็นต้น หรือแม้กระทั่งข้อมูลความสัมพันธ์ที่เรียกว่า “เพื่อนของเพื่อน” อีกด้วย ซึ่งข้อมูลเหล่านี้น่าจะเป็นข้อมูลที่ทำให้เราเข้าใจผู้ใช้ได้มากขึ้น รวมไปถึงศึกษาการประยุกต์ใช้ตัวแบบสำหรับให้คำแนะนำในชุดข้อมูลที่ใหญ่ขึ้น และประเภทข้อมูลหลากหลายมากขึ้น เช่นการแนะนำสำหรับภาพยนตร์หรือหนังสือ เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] D. Billsus and M. Pazzani. Learning Collaborative Information Filters. **In Proceedings of International Conference on Machine Learning**, 1998

[2] Ethem Alpaydin. Introduction to Machine Learning. 2nd ED. 2009

[3] F. Ricci et al. (eds), Recommender Systems Handbook. Springer Science+Business Media, LLC, 2011

[4] Herlocker, J. L., Konstan, J. A., Borchers, A. and Riedl, J. An Algorithmic Framework for Performing Collaborative Filtering, In SIGIR '99: **proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**, 1999, 230-237

[5] Katsuhiro Honda, Nobukazu Sugiura, Hidetomo Ichihashi, and Shoichi Araki. Collaborative Filtering Using Principal Component Analysis and Fuzzy Clustering. **First Asia-Pacific Conference**, October 23-26 2001. pp 394-402

[6] Nima Taghipour and Ahmad Kardan. A Hybrid Web Recommender System Based on Q-Learning. **In Proceeding of the 2008 ACM symposium on Applied computing**, 2008. pp 1164-1168

[7] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. Introduction to Data Mining. Addison-Wesley. 2005

[8] Paul R., Neophytos I., Mitesh S., Peter B. and Jonh R. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. **1994 Computer Supported Cooperative Work Conference**, 1994

[9] Paolo Massa and Paolo Avesani. Trust-aware Collaborative Filtering for Recommender System. **In Proc. of Federated Int. Conference on the Move to Meaningful Internet: CoopIS**, 2004

[10] P. Brusilovsky, A. Kobsa and W. Nejdl (Eds.): The Adaptive Web. **LNCS** 4321, pp. 325-341, 2009

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม (ต่อ)

[11] Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning). A Bradford Book, 1998

[12] Stefan Buettcher, Charles L. A. Clarke, and Gordon V. Cormack. Information Retrieval: Implementing and Evaluating Search Engines. The MIT Press. 2010

[13] Stephen Marsland. Machine Learning: An Algorithmic Perspective. Champman and Hall/CRC. 2009

[14] Sarwar, G. Karypis, J. Konstain and J. Riedl. Application of Dimensionality Reduction in Recommender System – A Case Study. **ACM WebKDD Workshop**, 2000

[15] Upendra S. and Pattie M. Social information filtering: Algorithm for automating "Word of Mouth". **Proceeding of the SIGCHI Conference on Human Factors in Computing System**, 1995. Page 210-217

[16] Xiongcai C., Michael B., Alfred K., Wayne W., Yang Kim, Paul C., and Ashesh M. Collaborative Filtering for People to People Recommendation in Social Networks. AI 2010: Advance in Artificial Intelligence, 2011. pp 476-485

[17] วงกต ศรีอุไร, ชูชาติ หฤไชยศักดิ์ และ จิรารัตน์ สิทธิวรชาติ. การแทนค่าข้อมูลที่ขาดหายไปเพื่อแก้ไขปัญหาความเบาบางของข้อมูลในการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม. วารสารพระจอมเกล้าลาดกระบัง. 16(1). 2551

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก
ผลงานที่ได้รับตีพิมพ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การลดผลกระทบของปัญหาความเบาบางของข้อมูลด้วยความสัมพันธ์ของผู้ใช้สำหรับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม

Social Enhancement to Reduce The Impact of Sparsity in Collaborative Filtering Method

กวินทร์ พิพัฒน์กุล[1,*] และ วีระ บุญจริง[1]
Kawin Pipatkul[1,*] and Veera Boonjing[1]

[1]สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ 10520
โทร 0-2326-4339-53 โทรสาร 0-2329-8412 E-mail: kawin.pipatkul@gmail.com

บทคัดย่อ

ปัญหาความเบาบางของข้อมูล (Data Sparsity Problem) เป็นปัญหาที่สำคัญประการหนึ่งของวิธีการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering Method) ซึ่งเป็นการยากสำหรับการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมที่จะหาความเหมือนของผู้ใช้งานได้อย่างแม่นยำจากชุดข้อมูลที่มีการขาดหายของข้อมูลจำนวนมาก โดยในงานวิจัยฉบับนี้ได้นำเสนอวิธีการลดผลกระทบจากปัญหาดังกล่าวที่มีต่อประสิทธิภาพของวิธีการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมด้วยการนำข้อมูลความสัมพันธ์ของผู้ใช้งานในเครือข่ายสังคมออนไลน์เข้ามาเพิ่มประสิทธิภาพของวิธีการกรองข้อมูลผู้ใช้ร่วม ซึ่งได้ทำการทดลองกับชุดข้อมูลบนเครือข่ายสังคมออนไลน์ที่ใช้งานอยู่จริง จากผลการทดลองวิธีการที่นำเสนอนอกจากจะได้ค่าความผิดพลาดที่น้อยลงแล้ว ยังมีความทนทานต่อความเบาบางของข้อมูลมากขึ้นด้วย

คำสำคัญ: ระบบแนะนำ/ การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม/ ความเบาบางของข้อมูล/ ความสัมพันธ์ในเครือข่ายสังคมออนไลน์

Abstract

Data Sparsity Problem is known as a major problem of Collaborative Filtering Method. That is, it is difficult for collaborative filtering to accurately measure user similarities from the data that contain many missing value. In this paper we propose the method for reducing the impact of mentioned problem by incorporating social friend information to improve the performance of collaborative filtering method. Experiment test with the real online social network web application and the result reveals that the proposed method not only yielded lower error of recommendation but also increase the resistant to data sparsity.

Keywords: Recommender system/ Collaborative filtering/ Data sparsity/ Social friends/ Social network

บทนำ

เพื่อที่จะรับมือกับปัญหา Information Overload ระบบผู้แนะนำ (Recommender System) ได้เข้ามามีบทบาทและได้รับความนิยมเป็นอย่างมาก ทั้งในการศึกษาการวิจัยและการใช้งานในเชิงพาณิชย์อิเล็คทรอนิค ได้แก่ Amazon.com, Netflix.com, MovieLens.com และ IMDb.com โดยระบบผู้แนะนำจะช่วยแนะนำสินค้าที่เกี่ยวข้องหรือน่าสนใจท่ามกลางสินค้าจำนวนมหาศาลให้กับผู้ใช้งาน ซึ่งวิธีการที่นิยมนำมาใช้ในระบบผู้แนะนำมีอยู่สองประเภทหลักๆ คือ (1) วิธีการกรองโดยดูที่เนื้อหา (Content-based Filtering) [1,2] และ (2) วิธีการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) [3,4] ในบทความฉบับนี้ศึกษาเกี่ยวกับวิธีการที่จะลดผลกระทบของปัญหาความเบาบางของข้อมูลที่มีผลกระทบต่อวิธีการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม

ในทางกลับกันกับจำนวนสินค้าที่มากมายมหาศาล ผู้ใช้งานมักจะทำการให้คะแนนสินค้าเพียงไม่กี่รายการ ซึ่งจะมีผลทำให้ประสิทธิภาพของวิธีการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมลดลง เราเรียกปัญหานี้ว่าปัญหาความเบาบางของข้อมูล (Data Sparsity) เพื่อที่จะแก้ไขปัญหาดังกล่าวได้มีผู้ที่เสนอวิธีในการเติมค่า (Enrich) ข้อมูลที่ขาดหายไปด้วยการประมาณค่าที่ขาดด้วยวิธีต่างๆ เช่น การแทนค่า

*Corresponding author. E-mail: kawin.pipatkul@gmail.com

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ด้วยวิธีการทางสถิติ [5], การประมาณค่าโดยดูจากระยะเวลาที่ผู้ใช้งานใช้ในการอ่านซึ่งเป็นการให้คะแนนทางอ้อม [6], หรือแม้กระทั่งการนำเทคนิควิธีการกรองโดยดูที่เนื้อหาเข้ามาร่วมด้วย [7,8]

ถึงแม้ว่าวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมจะเป็นที่รู้จักในอีกชื่อว่า Social-based Filtering [9] อันเนื่องมาจากการกรองแบบพึ่งพาผู้ใช้ร่วมนั้นเป็นวิธีการที่คล้ายกับการขอคำแนะนำจากผู้อื่นโดยมีความสัมพันธ์กันด้วยความชอบที่สอดคล้องหรือแตกต่างกัน แต่ก็ไม่ได้มีการใช้ประโยชน์จากข้อมูลความสัมพันธ์จริง ๆ ของบุคคลมาใช้ประโยชน์แต่อย่างใด โดยเฉพาะข้อมูลความสัมพันธ์ในเครือข่ายสังคมออนไลน์ซึ่งกำลังได้รับความนิยมอย่างสูงในเวลานี้ ทางคณะผู้วิจัยจึงมีแนวความคิดที่จะลดผลกระทบของปัญหาความเบาบางของข้อมูลที่มีต่อการกรองแบบพึ่งพาผู้ใช้ร่วม ด้วยการประมาณค่าข้อมูลที่ขาดหายด้วยค่าความชอบของเพื่อนของผู้ใช้งานในเครือข่ายสังคมออนไลน์ โดยมีสมมติฐานที่ว่าบุคคลจะมีแนวโน้มความชอบเป็นไปตามกลุ่มเพื่อนของตนมากกว่าบุคคลทั่วไป

การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering)

การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม (Collaborative Filtering) [5] จะแนะนำสินค้าให้กับผู้ใช้โดยพิจารณาจากความชอบของผู้ใช้กับสินค้า และนำความชอบของผู้ใช้มาวิเคราะห์เปรียบเทียบกับความชอบของผู้ใช้อื่น ๆ โดยมีสมมติฐานที่ว่าบุคคลที่มีความชอบคล้ายคลึงกันในอดีตจะมีแนวโน้มที่จะมีความชอบที่คล้ายคลึงกันด้วยในอนาคต ซึ่งความชอบของผู้ใช้มักถูกแสดงในรูปแบบของการให้คะแนน (Ratting) เช่น คะแนนอยู่ในช่วงหนึ่งถึงห้าตามระดับความชอบหรือไม่ชอบ ทั่วไปแล้วคะแนนหนึ่ง จะหมายถึง ไม่ชอบอย่างมาก และคะแนนห้า หมายถึง ชอบมาก ให้ $R_{U,I}$ เป็นค่าประมาณคะแนนความชอบของผู้ใช้ U ที่มีต่อสินค้า I ดังสมการ

$$R_{U,I} = \bar{R}_U + z\sum_{V\in\Psi} w(U,V) \times (R_{V,I} - \bar{R}_V) \quad (1)$$

โดยที่

$\bar{R}_U$ คือ ค่าเฉลี่ยของคะแนนความชอบในสินค้าของผู้ใช้ U

$\bar{R}_V$ คือ ค่าเฉลี่ยของคะแนนความชอบในสินค้าของผู้ใช้ V คนอื่น ๆ

Ψ คือ เซ็ตของผู้ใช้งานทั้งหมด

$w(U,V)$ คือ ค่าถ่วงน้ำหนัก หรือค่าความคล้ายคลึงระหว่างผู้ใช้ U และผู้ใช้ V

z คือ ค่า คงที่สำหรับการนอร์มอลไลซ์ ซึ่งมีค่าเท่ากับ $\frac{1}{\sum_V w(U,V)}$

งานวิจัยฉบับนี้ใช้ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson Correlation Coefficient–PCC) [10] ในการคำนวณค่าความคล้ายคลึง $w(U,V)$ ซึ่งเป็นค่าที่บ่งบอกความชอบของผู้ใช้งานว่า มีความสัมพันธ์กันมากน้อยเพียงไร นั่นหมายถึง หากผู้ใช้ที่มักมีความชอบในสินค้าที่ใกล้เคียงกับเราชอบในสินค้าตัวหนึ่ง เราก็มีแนวโน้มสูงที่จะชอบในสินค้าตัวนั้นด้วยเช่นกัน

$$w(U,V) = \frac{\sum_I (R_{U,I} - \bar{R}_U)(R_{V,I} - \bar{R}_V)}{\sqrt{\sum (R_{U,I} - \bar{R}_U)^2 \sum_I (R_{V,I} - \bar{R}_V)^2}} \quad (2)$$

ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันจะเป็นการบอกระดับความสอดคล้องหรือขัดแย้งกันระหว่างผู้ใช้งาน ซึ่งจะมีค่าอยู่ระหว่าง -1 ถึง 1 โดยที่ค่าที่มากกว่าศูนย์ หมายถึง ผู้ใช้มีความชอบที่มีความสัมพันธ์ในเชิงสอดคล้อง ค่าที่น้อยกว่าศูนย์ หมายถึง ผู้ใช้มีความชอบที่มีความสัมพันธ์ในเชิงขัดแย้ง และค่าศูนย์ หมายถึง ความชอบระหว่างผู้ใช้งานไม่ได้มีความสัมพันธ์กัน ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สันยิ่งมากกว่าศูนย์ ยิ่งมีระดับความสอดคล้องกันมาก ในทางกลับกัน ค่าสัมประสิทธิ์สหสัมพันธ์ยิ่งน้อยกว่าศูนย์ยิ่งมีระดับความขัดแย้งมาก

วิธีการ

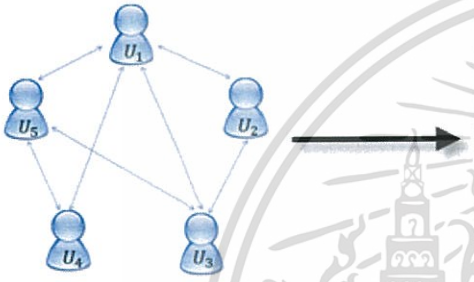
เพื่อที่จะลดผลกระทบของปัญหาความเบาบางของข้อมูล คณะผู้วิจัยได้เสนอวิธีการในการนำข้อมูลความสัมพันธ์บนเครือข่ายสังคมออนไลน์มาพัฒนาประสิทธิภาพของวิธีการกรองแบบพึ่งพาผู้ใช้ร่วม โดยการเพิ่มขั้นตอนการเติมค่าความชอบที่ขาดหายไปจากข้อมูลความชอบของผู้ใช้คนอื่น ๆ ที่มีความสัมพันธ์กับผู้ใช้งานแต่ละคน ก่อนที่จะเข้าสู่กระบวนการของวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมตามปรกติ เพื่อให้ง่ายต่อความเข้าใจจะขอยกตัวอย่างข้อมูลการให้คะแนน และข้อมูลความสัมพันธ์ดังที่แสดงในตารางที่ 1 และรูปที่ 1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 1 แสดงข้อมูลการให้คะแนนสินค้าของผู้ใช้แต่ละคน

	Item1	Item2	Item3	Item4	Item5
User1	1	-	2	3	-
User2	-	3	4	5	2
User3	4	-	-	3	2
User4	5	4	-	-	-
User5	-	-	-	-	-

จากตารางที่ 1 เป็นการแสดงข้อมูลการให้คะแนนของผู้ใช้แต่ละคน โดยในแต่ละแถวจะหมายถึงผู้ใช้ได้ให้คะแนนความชอบในสินค้าแต่ละรายการ ตัวอย่างเช่นในแถวแรกจะหมายถึง User1 ได้ให้คะแนนสินค้า Item1, Item3 และ Item4 ด้วยคะแนน 1, 2 และ 3 ตามลำดับ และมีข้อมูลบางส่วนที่ยังว่างอยู่ ซึ่งเกิดจากการที่ผู้ใช้ไม่ได้ให้คะแนนความชอบกับสินค้านั้น จะสังเกตเห็นว่า User5 ไม่ได้ทำการให้คะแนนสินค้าเลยแม้แต่รายการเดียว ซึ่งแทบเป็นไปไม่ได้เลยที่วิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบปรกติจะทำการเปรียบเทียบความชอบกับผู้ใช้คนอื่นๆ



	u_1	u_2	u_3	u_4	u_5
u_1	0	1	1	1	1
u_2	1	0	1	0	0
u_3	1	1	0	0	1
u_4	1	0	0	0	1
u_5	1	0	1	1	0

รูปที่ 1 แสดงความสัมพันธ์ระหว่างผู้ใช้แต่ละคนในรูปแบบของกราฟ และในรูปแบบของเมตริกซ์ประชิด (Adjacency Matrix)

ในรูปที่ 1 เป็นตัวอย่างแสดงความสัมพันธ์ระหว่างผู้ใช้ในรูปแบบกราฟโดยมีเส้นเชื่อม (edge) ในการระบุความสัมพันธ์กันระหว่างผู้ใช้ ซึ่งสามารถอธิบายในรูปแบบเมตริกซ์ โดย 0 หมายถึงระหว่างผู้ใช้ไม่มีความสัมพันธ์กันและ 1 หมายถึง ระหว่างผู้ใช้มีความสัมพันธ์กัน

จากสมมติฐานที่ว่า บุคคลจะมีแนวโน้มความชอบเป็นไปตามกลุ่มเพื่อนของตน (ผู้ใช้ที่มีความสัมพันธ์กัน) ในขั้นตอนการเติมค่าความชอบที่ขาดหาย จึงได้นำค่าเฉลี่ยคะแนนความชอบของกลุ่มเพื่อนของผู้ใช้มาใช้ในการเติมค่า

$$\overline{R_{UFI}} = \frac{\sum_{UF} R_{UFI}}{n_{UF}} \quad (3)$$

โดยที่

$\overline{R_{UFI}}$ คือ ค่าเฉลี่ยคะแนนความชอบของเพื่อนของผู้ใช้ U ที่มีต่อสินค้า I

UF คือ ผู้ใช้ที่มีความสัมพันธ์เป็นเพื่อนกับผู้ใช้ U

R_{UFI} คือ คะแนนความชอบของเพื่อนของผู้ใช้ U ที่มีต่อสินค้า I

n_{UF} คือ จำนวนเพื่อนของผู้ใช้ U

เนื่องจากเราต้องการเติมค่าความชอบเฉพาะกับข้อมูลที่ขาดหาย จึงกำหนดเงื่อนไขในการสร้างเมตริกซ์ที่เติมค่าแล้วตามสมการ

$$[x]_{i,j} = \begin{cases} R_{ui} & if \quad R_{ui} = 0 \\ \overline{R_{UFI}} & if \quad R_{ui} > 0 \end{cases} \quad (4)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2 แสดงผลการเติมค่าความชอบที่ขาดหายโดยให้ค่าเฉลี่ยคะแนนความชอบของกลุ่มเพื่อน

	Item1	Item2	Item3	Item4	Item5
User1	1	3.5	2	3	2
User2	2.5	3	4	5	2
User3	4	3	3	3	2
User4	5	4	2	3	-
User5	3.33	3.5	2	3	2

ผลและอภิปราย

ในการทดลองเราได้ทำการเปรียบเทียบวิธีการที่นำเสนอ กับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบปรกติ และวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมแบบเติมค่าด้วยค่าเฉลี่ยความชอบของสินค้าแต่ละรายการ เพื่อพิสูจน์สมมติฐานของเราที่ได้กล่าวไว้ก่อนหน้านี้ และทำการทดสอบวิธีการเหล่านี้กับข้อมูลที่มีระดับความเบาบางในระดับต่างๆ เพื่อวัดระดับความทนทานต่อระดับความเบาบางของข้อมูล

1. **ชุดข้อมูล**

ชุดข้อมูลที่ใช้ในการทดลองเป็นข้อมูลที่สกัดมาจาก เวปไซต์เครือข่ายสังคมออนไลน์ Yelp.com ซึ่งเป็นเวปไซต์ที่มีชื่อเสียงและเป็นแหล่งรวบรวมบทวิจารณ์ของร้านค้าและบริการต่างๆ มากมาย เช่น ร้านอาหาร, แหล่งจับจ่ายใช้สอย, โรงแรม, สปาร์ และบริการทางการเงินเป็นต้น นอกจากผู้ใช้สามารถเข้ามาหาข้อมูลอ่านบทวิจารณ์ของร้านค้าแล้ว ยังสามารถที่จะเขียนคำวิจารณ์รวมไปถึง ให้คะแนนสำหรับร้านค้าที่ตนเองเคยเข้าไปใช้บริการได้อีกด้วย ที่สำคัญผู้ใช้งานที่เป็นสมาชิกของ Yelp.com สามารถที่จะสร้างความสัมพันธ์กับสมาชิกคนอื่นด้วยการยื่นคำขอเป็นเพื่อน หรือแม้กระทั่งสามารถเชิญบุคคลภายนอกให้เข้ามาเป็นสมาชิกและเป็นเพื่อนของตนได้ด้วยการส่งอีเมลล์ ซึ่งทำให้ Yelp.com มีลักษณะเป็นเครือข่ายสังคมอีกด้วย

ข้อมูลที่สกัดจาก Yelp.com สกัดมา ณ วันที่ 30 มิถุนายน พ.ศ. 2555 เป็นข้อมูลของร้านอาหารไทย ในเมืองซอลต์เลกซิตี รัฐยูทาห์ ประเทศสหรัฐอเมริกาจำนวน 10 ร้าน ผู้เขียนบทวิจารณ์และให้คะแนนจำนวน 532 คน บทวิจารณ์จำนวน 997 บทวิจารณ์ และข้อมูลความสัมพันธ์ของผู้เขียนบทวิจารณ์จำนวน 1,505 ความสัมพันธ์

2. **การวัดประสิทธิภาพ**

ในการทดสอบเราใช้วิธีการแบบ Leave-One-Out Cross-Validation ในการทดสอบ และใช้ตัวชี้วัด Mean Absolute error (MAE) ในการประเมินประสิทธิภาพความแม่นยำในการทำนาย ซึ่งวิธีการนี้จะทำการหาค่าเฉลี่ยความผิดพลาดในการทำนายแต่ละรายการ

$$MAE = \frac{\sum_{U,I}|r_{UI} - r'_{UI}|}{L} \quad (5)$$

โดยที่ r_{UI} คือคะแนนความชอบของผู้ใช้ U ที่มีต่อสินค้า I, r'_{UI} คือค่าประมาณคะแนนความชอบของผู้ใช้ U ที่มีต่อสินค้า I และ L คือจำนวนรายการที่ใช้ในการทดสอบ ซึ่งค่า MAE ยิ่งน้อยค่าความแม่นยำในการทำนายยิ่งสูง

3. **การอภิปรายผล**

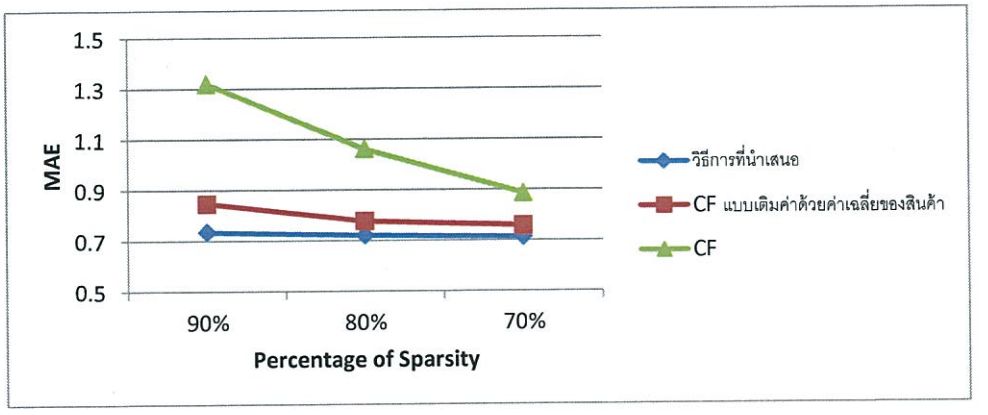
จากผลการทดลองในตารางที่ 3 และรูปที่ 2 แสดงให้เห็นว่าวิธีการที่นำเสนอมีประสิทธิ์ภาพดีที่สุดทั้งความแม่นยำในการทำนาย และความทนทานต่อความเบาบางของข้อมูล

ตารางที่ 3 แสดงค่า MAE ระหว่างวิธีการต่าง ๆ ที่ใช้ในการทดลอง

	MAE
วิธีที่นำเสนอ	0.692
CF แบบเติมค่าด้วยค่าเฉลี่ยของสินค้า	0.716
CF	0.771

การเปรียบเทียบระหว่างวิธีการที่นำเสนอซึ่งจะเป็นการเติมค่าด้วยค่าเฉลี่ยคะแนนความชอบของกลุ่มเพื่อน กับวิธีการเติมค่าด้วยค่าเฉลี่ยความชอบของสินค้าเป็นการพิสูจน์สมมติฐานของเราที่ว่า บุคคลจะมีแนวโน้มความชอบเป็นไปตามกลุ่มเพื่อนของตนมากกว่าบุคคลทั่วไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2 แสดงค่า MAE ของแต่ละวิธีการ เมื่อทดสอบด้วยชุดข้อมูลที่มีระดับความเบาบางในระดับต่าง ๆ

จากรูปที่ 2 แสดงให้เห็นว่า การเปลี่ยนแปลงระดับความเบาบางของข้อมูล มีผลกระทบกับวิธีการที่นำเสนอน้อยที่สุด ซึ่งนั่นหมายความว่า วิธีการที่นำเสนอมีความทนทานต่อความเบาบางของข้อมูลมากที่สุด

บทสรุป

ในบทความชิ้นนี้ ได้นำเสนอวิธีการในการลดผลกระทบจากปัญหาความเบาบางของข้อมูลที่มีต่อประสิทธิภาพของวิธีการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม ด้วยการเติมค่าความชอบที่ขาดหายไปด้วยค่าความชอบของกลุ่มเพื่อน ซึ่งจากผลการทดลองพบว่า วิธีการที่นำเสนอสามารถเพิ่มประสิทธิภาพให้กับวิธีการกรองแบบพึ่งพาผู้ใช้ร่วมได้ทั้งความแม่นยำและความทนทานต่อความเบาบางของข้อมูล เนื่องจากวิธีการที่นำเสนอนี้เป็นการเพิ่มความหนาแน่นของข้อมูล

จะเห็นได้ว่าวิธีการเติมข้อมูลที่นำเสนอไปนั้นอาศัยเพียงข้อมูลความสัมพันธ์บนเครือข่ายสังคมออนไลน์เพียงอย่างเดียวทั้งๆ ที่ในความเป็นจริงแล้วยังมีข้อมูลที่น่าสนใจบนเครือข่ายสังคมออนไลน์ที่น่าสนใจอีกมาก ทั้งข้อมูลส่วนตัว เช่น เพศ, อายุ, ระดับการศึกษา และอาชีพ เป็นต้น รวมไปถึงข้อมูลความสัมพันธ์ที่เรียกว่า "เพื่อนของเพื่อน" อีกด้วย ซึ่งข้อมูลเหล่านี้น่าจะเป็นข้อมูลที่ทำให้เราเข้าใจผู้ใช้ได้มากขึ้น

เอกสารอ้างอิง

วงกต ศรีอุไร, ชูชาติ หฤไชยศักดิ์ และจิรารัตน์ สิทธิวรชาติ. (2551). การแทนค่าข้อมูลที่ขาดหายไปเพื่อแก้ไขปัญหาความเบาบางของข้อมูลในการกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วม. **วารสารพระจอมเกล้าลาดกระบัง**, 16(1).

อุรีรัฐ สุขสวัสดิ์ชน และจักริน สุขสวัสดิ์ชน. (2010). **ระบบแนะนำภาพยนตร์ที่ใช้การกรองข้อมูลแบบพึ่งพาผู้ใช้ร่วมและข้อมูลส่วนบุคคล Movie recommender system using collaborative filtering and contextual information.** Knowledge and Smart Technologies.

Bogdanov, D., Haro, M., Fuhrmann, F., Gomez, E. and Herera, P. (2010). Content-based music recommendation based on user preference examples. **ACM Conference on Recommender System.** Workshop on Music Recommendation and Discovery.

Brusilovsky, P., Kobsa, A. and Nejdl, W. (2007). The adaptive web. **LNCS,** 4321, 325-341.

Canny, J. (2002). Collaborative filtering with privacy via factor analysis. **SIGIR,** 238-245.

Jin, R., Chai, J.Y. and Si, L. (2004). An automantic weighting scheme for collaborative filtering. **SIGIR,** 337-344.

Morita, M. and Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. **ACM SIGIR Conference on research and development in information retrieval,** 7, 272-281.

Paul, R., Neophytos, I., Mitesh, S., Peter, B. and Jonh, R. (1994). **GroupLens: An open architecture for collaborative filtering of netnews.** Computer supported cooperative work conference.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Prem Melville, Raymond J. Mooney and Ramadass Nagarajan. (2002). Content-boosted collaborative filtering for improve recommendation. **Conference on Artificial Intelligence AAAI,** 18, 187-192.

Upendra S. and Pattie M. (1995). Social information filtering: algorithm for automating "word of mouth". **Computer human interaction CHI,** 210-217.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ – สกุล	นาย กวินทร์ พิพัฒน์กุล
วัน เดือน ปี เกิด	23 ธันวาคม 2526
ที่อยู่	105 ซ.พัฒนาการ 52 แขวงสวนหลวง เขตสวนหลวง
	จังหวัดกรุงเทพเทพมหานคร 10250

ประวัติการศึกษา

2549	จบการศึกษาปริญญาวิทยาศาสตรบัณฑิต
	สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์
	สถาบันเทคโนโลยีพระจอมเกล้าลาดกระบัง

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้