สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน

ITEM-BASED TOP-N RECOMMENDATION USING MODIFIED FUZZY C-MEAN

T110608

ปนิดา มักประเสริฐ

PANITA MAKPRASERT

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

สบหมู่.....
เลขทะเบียน 110608
วัน,เดือน,ปี - 9 พ.ย. 2553

b.....
i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
พ.ศ. 2553

KMITL- 2010-IT-M-001-005

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ITEM-BASED TOP-N RECOMMENDATION USING MODIFIED FUZZY C-MEAN

PANITA MAKPRASERT

A THESIS SUBMITTED IN PATIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEEGREE OF

MASTER OF SCIENCE

FUCULTY OF INFORMATION TECNOLOGY

KING MONGKUT' INSTITUTE OF TECHNOLOGY LADKRABANG

2010

KMITL-2010-IT-M-001-005

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2010

FUCULTY OF INFORMATION TECNOLOGY

KING MONGKUT' INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ใบรับรองวิทยานิพนธ์

หัวข้อวิทยานิพนธ์ การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น โดยใช้โมดิฟายฟัซซีซีมีน

Item-Based Top-N Recommendation Using Modified Fuzzy C-Mean

นักศึกษา นางสาวปนิตา มักประเสริฐ

รหัสประจำตัว 48066433

ปริญญา วิทยาศาสตรมหาบัณฑิต

สาขาวิชา เทคโนโลยีสารสนเทศ

อาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.วรพจน์ กรีสุระเดช

คณะกรรมการสอบวิทยานิพนธ์	ลายมือชื่อ
รองศาสตราจารย์ ดร.พีระพนธ์ โสพัศสถิตย์ รองศาสตราจารย์ ดร.วรพจน์ กรีสุระเดช รองศาสตราจารย์ ดร.อาริต ธรรมโน รองศาสตราจารย์ ดร.บุญธีร์ เครือตราชู ดร. ปานวิทย์ ธุวะนุติ	

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

วัน / เดือน/ ปี ที่สอบ วันพุธที่ 2 มิถุนายน 2553 เวลา 10.30 น.

สถานที่สอบ ณ ห้อง M 22 ชั้น M คณะเทคโนโลยีสารสนเทศ

คณะเทคโนโลยีสารสนเทศรับรองแล้ว

(รองศาสตราจารย์ ดร.จันทร์บูรณ์ สถิตวิริยวงศ์)

คณบดีคณะเทคโนโลยีสารสนเทศ

วันที่ 7 **เดือน** มิถุนายน **พ.ศ.** 2553

สำนักทะเบียนและประมวลผล สจล.
วันที่ส่งเล่มวิทยานิพนธ์ฉบับสมบูรณ์
วันที่ 11 เดือน มิ.ย. พ.ศ. 53
ลงชื่อ.....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การแนะนำสินค้าแบบไอเทมเบสท็อปเอ็นโดยใช้โมดิฟายฟ์ซซีซี-มีน
นักศึกษา	นางสาวปนิตา มักประเสริฐ
รหัสนักศึกษา	48066433
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
พ.ศ.	2553
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.ดร.วรพจน์ กรีสุระเดช

บทคัดย่อ

ระบบแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นจะใช้การวิเคราะห์ข้อมูลเมตริกซ์ผู้ใช้-ไอเท็ม เพื่อระบุความสัมพันธ์ระหว่างไอเท็มที่แตกต่างและหลากหลายในระบบ แล้วใช้ความสัมพันธ์เหล่านั้นคำนวณหารายการไอเท็มสำหรับแนะนำ ทว่าความซับซ้อนของการคำนวณนั้นจะเพิ่มขึ้นตามจำนวนไอเท็ม ซึ่งโดยปกติแล้วระบบในเชิงพาณิชย์ส่วนใหญ่ก็มักจะมีไอเท็มอยู่เป็นจำนวนมหาศาล วิทยานิพนธ์ฉบับนี้จึงนำเสนอวิธีการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟ์ซซีซีมีนเพื่อแก้ไขปัญหาที่กล่าวมาข้างต้น ภายในวิทยานิพนธ์ฉบับนี้จะกล่าวถึงวิธีการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟ์ซซีซีมีน อันดับแรกคือการจัดกลุ่มไอเท็มโดยใช้อัลกอริทึมโมดิฟายฟ์ซซีซีมีน ซึ่งนำการหาค่าความคล้ายคลึงด้วยค่าสหสัมพันธ์แบบเพียร์สันเข้ามาประยุกต์ใช้ จากนั้นจึงคำนวณหาความคล้ายคลึงระหว่างไอเท็มภายในกลุ่ม และสุดท้ายคือการคำนวณหากลุ่มของไอเท็มที่จะใช้แนะนำ ซึ่งจากผลการทดลองโดยใช้ข้อมูลชุดมูฟวี่เลนส์แสดงให้เห็นว่าเวลาในการประมวลผลเพื่อหารายการแนะนำนั้นลดลงตามวัตถุประสงค์ เมื่อเปรียบเทียบระหว่างวิธีการแนะนำแบบไอเท็มเบสโดยใช้โมดิฟายฟ์ซซีซีมีน กับการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis Title	Item-Based Top-N Recommendation Using Modified Fuzzy C-Mean
Student	Ms. Panita Makprasert
Student ID.	48066433
Degree	Master of Science
Major	Information Science
Year	2010
Thesis Advisor	Assoc. Prof. Dr. Worapoj Kreesuradej

ABSTRACT

An item based top-N recommendation system is a recommendation system that analyzes the user-item matrix to identify relations between the different items, and uses these relations to compute the list of recommendations. Unfortunately, the computational complexity of the system grows with the number of items that in typical commercial applications can grow to be several hundred. To solve this problem, this thesis proposes an item based top-N recommendation system using modified fuzzy C-mean clustering algorithm. First, the proposed method groups similar items together using modified fuzzy C-mean algorithm that based on Pearson correlation similarity. Then, items in a cluster are computed the similarity between the items. Finally, the method used to combine these similarities in order to compute the similarity between a basket of items and a candidate recommender item. Our experimental evaluating on Movielens dataset shows that the execution time of the proposed item-based algorithm is faster than that of the traditional proposed item-based algorithms.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความกรุณาจาก รศ.ดร. วรพจน์ กรีสุระเดช ซึ่งเป็นอาจารย์ผู้ควบคุมวิทยานิพนธ์ ที่กรุณาให้คำแนะนำพร้อมทั้งชี้แนะแนวทางแก้ไขปัญหาของงานวิจัย ข้าพเจ้ารู้สึกซาบซึ้งในความอนุเคราะห์จากท่านเป็นอย่างยิ่งและขอขอบพระคุณเป็นอย่างสูง

ขอกราบพระคุณคณาจารย์คณะวิทยาศาสตร์แห่งมหาวิทยาลัยบูรพา และคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่าน ที่ได้ถ่ายทอดความรู้ทางวิชาการให้กับข้าพเจ้า

ขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่ให้การสนับสนุนในทุกเรื่อง รวมถึงกำลังใจอันเปี่ยมล้น ทำให้ข้าพเจ้าสามารถดำเนินงานวิทยานิพนธ์ฉบับนี้จนสำเร็จ

ขอขอบคุณคุณชยานนท์ ทรัพย์อากาซึ่งคอยช่วยเหลือให้คำปรึกษา และชี้แนะเกี่ยวกับการเขียนโปรแกรมภาษาไพธอน

ขอขอบคุณคุณเทวัญ นาคเทวัญ ซึ่งคอยอำนวยความสะดวกในการรันโปรแกรมทดสอบและจัดทำเอกสาร

สุดท้ายนี้ขอขอบคุณเพื่อนๆ ร่วมห้องวิจัยและต่างห้องวิจัย ในคณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่าน ที่ช่วยอยู่เป็นเพื่อนกันเสมอทำให้ข้าพเจ้าไม่รู้สึกโดดเดี่ยวตลอดการจัดทำวิทยานิพนธ์

ปนิตา มักประเสริฐ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VI

สารบัญรูป VIII

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.3 ขอบเขตงานวิจัย 2

1.4 วิธีดำเนินงานวิจัย 3

1.5 การจัดเรียงหัวข้อในการเรียบเรียงวิทยานิพนธ์ 3

บทที่ 2 ทฤษฎีพื้นฐานที่ใช้ในการวิจัย และงานวิจัยที่เกี่ยวข้อง 4

2.1 ฟัซซีเซ็ต (Fuzzy set) 4

2.1.1 นิยามของฟัซซีเซ็ต 4

2.1.2 โอเปอร์ชั่นของฟัซซีเซ็ต (Operation on Fuzzy set) 5

2.2 ฟัซซีซีมีน (Fuzzy C-Mean) 5

2.3 การวัดค่าความคล้ายคลึง 7

2.3.1 การวัดค่าความคล้ายคลึงแบบทั่วไป 7

2.3.2 การวัดค่าความคล้ายคลึงโดยใช้การหาค่าสหสัมพันธ์ 8

2.4 วิธีการเคโฟล์ครอสวาลิเดชั่น 8

2.5 งานวิจัยที่เกี่ยวข้อง 9

2.5.1 งานวิจัยด้านการแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็น 9

2.5.2 งานวิจัยด้านการจัดกลุ่มข้อมูลเพื่อการทำนายค่าคะแนนและการแนะนำ 10

บทที่ 3 การแนะนำสินค้าแบบไอเทมเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน 14

3.1 โมดิฟายฟัซซีซีมีน 14

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.2 ขั้นตอนการดำเนินงาน 15
3.2.1 การจัดกลุ่มด้วยโมดิฟายฟัซซีซีมีน 16
3.2.2 คำนวณหากลุ่มที่เหมาะสมสำหรับแอคทีฟยูสเซอร์ 21
3.2.3 คำนวณหาไอเท็มภายในกลุ่มสำหรับสร้างรายการแนะนำ 21
3.2.4 สร้างรายการแนะนำ 23
3.2.5 แนะนำรายการไอเท็มแก่แอคทีฟยูสเซอร์ 23

บทที่ 4 การทดลองและผลการทดลอง 24
4.1 ชุดข้อมูลที่ใช้ในการทดลอง 24
4.1.1 ชุดข้อมูลมูฟวี่เลนส์ 24
4.1.2 การเตรียมข้อมูล 25
4.2 การประเมินประสิทธิภาพของอัลกอริทึม 26
4.3 ผลการทดลองการแนะนำแบบไอเท็มเบสท็อปเอ็นและผลการทดลอง 26
4.3.1 การแนะนำแบบไอเท็มเบสท็อปเอ็น 27
4.3.2 การแนะนำแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนและฟัซซีซีมีน 33

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ 48
5.1 สรุปผลการวิจัย 48
5.2 ประสิทธิภาพของอัลกอริทึม 50
5.3 ข้อเสนอแนะ 50

เอกสารอ้างอิง 51

ภาคผนวก ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่ 53

ประวัติผู้เขียน 60

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

2.1 แสดงอัลกอริทึมของฟัซซีซีมีน 7

2.2 แสดงชุดข้อมูลสินค้าในรูปผู้ใช้ – ไอเท็ม เมตริกซ์ 10

2.3 แสดงไอเท็ม – ไอเท็มเมตริกซ์ (ชุดข้อมูลสินค้า) 11

2.4 แสดงเมตริกซ์ 1 * *m* ในการแนะนำแบบไอเท็มเบสท็อปเอ็น 12

2.5 แสดงเมตริกซ์ผลลัพธ์ 12

3.1 แสดงชุดข้อมูลเรตติ้งสินค้าในรูปผู้ใช้ – ไอเท็ม เมตริกซ์ 16

3.2 แสดงชุดข้อมูลเรตติ้งสินค้าในรูปไอเท็ม – ผู้ใช้ เมตริกซ์ 17

3.3 แสดงค่าความเป็นสมาชิกของไอเท็มต่อแต่ละกลุ่มไอเท็ม 17

3.4 แสดงการคำนวณหาค่าศูนย์กลางของกลุ่มไอเท็ม 1 18

3.5 แสดงค่าศูนย์กลางของแต่ละกลุ่มไอเท็ม 19

3.6 แสดงค่าความเป็นสมาชิกของไอเท็มต่อแต่ละกลุ่มไอเท็ม 19

3.7 แสดงค่าศูนย์กลางใหม่ของแต่ละกลุ่มไอเท็มเมื่อเงื่อนไขการหยุดคำนวณเป็นจริง 20

3.8 แสดงค่าความเป็นสมาชิกของไอเท็มต่อแต่ละกลุ่มไอเท็ม เทียบจากค่าศูนย์กลางสุดท้าย 20

3.9 แสดงกลุ่มไอเท็มและไอเท็มสมาชิก 21

3.10 แสดงการให้ค่าเรตติ้งของ U_7 21

3.11 แสดงชุดข้อมูลเรตติ้งสินค้าของไอเท็มภายในกลุ่ม C_2 22

3.12 แสดงไอเท็ม-ไอเท็มเมตริกซ์ของกลุ่ม C_2 22

3.13 แสดง 1 * *m* เมตริกซ์ของผู้ใช้รหัส U_7 22

3.14 แสดงการคูณเมตริกซ์เพื่อหาเมตริกซ์ผลลัพธ์ 23

4.1 แสดงตัวอย่างชุดข้อมูลมูฟวี่เลนส์ 24

4.2 แสดงตัวอย่างชุดข้อมูลที่1 25

4.3 แสดงตัวอย่างชุดข้อมูลที่2 25

4.4 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 1 28

4.5 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 2 29

4.6 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 3 30

4.7 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 4 31

4.8 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 5 32

4.9 แสดงเวลาการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น 33

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4.10 แสดงค่าฮิตเรตการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น 33

4.11 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟาย-ฟ็ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 2 กลุ่ม) 35

4.12 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟาย-ฟ็ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 3 กลุ่ม) 36

4.13 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟาย-ฟ็ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 4 กลุ่ม) 37

4.14 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟาย-ฟ็ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 5 กลุ่ม) 38

4.15 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้ฟ็ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 2 กลุ่ม) 39

4.16 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้ฟ็ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 3 กลุ่ม) 40

4.17 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้ฟ็ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 4 กลุ่ม) 41

4.18 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้ฟ็ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 5 กลุ่ม) 42

4.19 แสดงเวลาที่ใช้ในการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น โดยใช้โมดิฟายฟ็ซซีซีมีน เปรียบเทียบกับฟ็ซซีซีมีน 43

4.20 แสดงค่าความถูกต้องของการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น โดยใช้โมดิฟายฟ็ซซีซีมีน เปรียบเทียบกับฟ็ซซีซีมีน 44

4.21 แสดงการเปรียบเทียบเวลาที่ใช้ในการแนะนำแบบต่างๆ 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

4.1 แสดงแผนภูมิแท่งเปรียบเทียบด้านเวลา 47

4.2 แสดงแผนภูมิแท่งเปรียบเทียบค่าฮิตเรต 47

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเวิลด์ ไวด์ เว็บ (World Wide Web: WWW) ซึ่งเป็นบริการอย่างหนึ่งบนอินเตอร์เน็ตได้เข้ามามีบทบาทต่อชีวิตประจำวันของมนุษย์เป็นอย่างมาก เพราะระบบอินเทอร์เน็ตช่วยให้เข้าถึงสารสนเทศได้อย่างสะดวกและประหยัดเวลาในการสืบค้นข้อมูลที่ต้องการ ทำให้เป็นที่นิยมรวมถึงรู้จักกันในวงกว้างทั้งในหมู่นักเรียนนักศึกษาและคนทำงาน

หลายองค์กรจึงให้ความสนใจกับประโยชน์ของอินเทอร์เน็ต และได้นำมาประยุกต์ใช้เพื่อวัตถุประสงค์ที่แตกต่างกันไป ไม่ว่าจะเป็นสำหรับเผยแพร่ข้อมูลหรือเสนอขายสินค้า/บริการ ส่งผลให้มีการปรับปรุงและพัฒนาแอปพลิเคชั่นบนเว็บอย่างต่อเนื่องเพื่อตอบโจทย์ความต้องการของผู้ใช้ อาทิ ระบบอีเลิร์นนิ่ง (e-learning) สำหรับด้านการศึกษา, การให้บริการด้านการเงินของธนาคาร, ระบบพาณิชย์อิเล็กทรอนิกส์หรือ 'อีคอมเมิซ' (e-commerce)

อีคอมเมิซเป็นเทคโนโลยีที่ยังคงได้รับความนิยม เนื่องด้วยการทำธุรกรรมผ่านทางเว็บไซต์นั้นก่อเกิดประโยชน์สองทางแก่ทั้งผู้ซื้อและผู้เสนอขาย ระบบงานจะช่วยลดต้นทุนของฝ่ายเจ้าของกิจการและอำนวยความสะดวกแก่ผู้บริโภค เพียงแค่ใช้การตัดสินใจกับปลายนิ้วคลิก ลูกค้าก็ได้เป็นเจ้าของสินค้าหรือบริการที่ต้องการ

ด้วยเหตุนี้ระบบแนะนำ [1] (Recommender system) จึงถูกพัฒนาขึ้น เพื่อที่จะนำเสนอสินค้าหรือบริการซึ่งคล้ายคลึงกับสินค้าหรือบริการที่ลูกค้าเคยเลือกซื้อไปแล้วในอดีต นับว่าเป็นตัวช่วยที่ดีในการตัดสินใจแก่ลูกค้า หรือระบบจะช่วยแนะนำสินค้าที่คล้ายคลึงกันให้ในกรณีที่สินค้าซึ่งลูกค้าต้องการหมดหรือไม่มีอยู่ในระบบ และแน่นอนว่าระบบแนะนำจะช่วยขยายโอกาสในการขายสินค้าได้เพิ่มขึ้นแก่เจ้าของกิจการด้วยเช่นเดียวกัน

ระบบแนะนำแบ่งได้เป็น 2 ประเภท [2] คือ ระบบแนะนำแบบยูสเซอร์เบส (User-based recommendation) และระบบแนะนำแบบไอเท็มเบส (Item-Based recommendation) ซึ่งระบบแนะนำแบบยูสเซอร์เบสนั้นจะใช้ข้อมูลทางด้านประชากรศาสตร์ ยกตัวอย่างเช่น เพศ อายุ หรือถิ่นที่อยู่อาศัยของผู้ใช้ที่ระบบกำลังจะแนะนำหรือแอคทีฟยูสเซอร์ (Active User) มาวิเคราะห์ เพื่อหาผู้ใช้รายอื่นๆ ที่มีความคล้ายคลึงกัน และแนะนำสินค้าที่ผู้ใช้ซึ่งมีความคล้ายคลึงกันได้เลือกซื้อแก่แอคทีฟยูสเซอร์ สำหรับระบบแนะนำแบบไอเท็มเบสจะใช้ข้อมูลการซื้อขายของสินค้าในอดีตมาวิเคราะห์ความสัมพันธ์ระหว่างสินค้าที่อยู่ในระบบ วิธีการนี้จะได้รายการสินค้าสำหรับแนะนำแก่แอคทีฟยูสเซอร์จากสินค้าที่มีความคล้ายคลึงกับสินค้าที่แอคทีฟยูสเซอร์เคยเลือกซื้อ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สินค้าหรือไอเท็ม (Item) ในระบบแนะนำแบบไอเท็มเบสที่จะแนะนำแก่ลูกค้านั้น ได้มาจากวิธีการเปรียบเทียบข้อมูลซื้อขายในอดีต กล่าวคือ ระบบจะทำการเทียบรายการไอเท็มภายในระบบทั้งหมด กับสินค้าที่ผู้ใช้หรือลูกค้าได้เคยเลือกซื้อ โดยมีอัลกอริทึมสำหรับวัดค่าความคล้ายคลึง (Similarity) [3] เช่น วิธีการวัดค่ามุมโคซายน์ของเวคเตอร์ (Cosine-based) [4] ซึ่งจะมองความเหมือนของสองไอเท็มจากมุมระหว่างเวคเตอร์ของสองไอเท็มนั้น หากค่าที่ได้หลังจากการคำนวณเป็น 1 หมายถึง สองไอเท็มที่นำมาเปรียบเทียบเหมือนหรือคล้ายคลึงกัน กรณีเป็น 0 หมายถึง สองไอเท็มนั้นไม่เหมือนหรือแตกต่างกัน หลังจากทำการเปรียบเทียบไอเท็มที่ถูกเลือกซื้อโดยแอคทีฟยูสเซอร์กับข้อมูลของไอเท็มที่ถูกเลือกซื้อทั้งระบบแล้ว จึงนำไอเท็มที่คล้ายคลึงกันมาสร้างเป็นรายการแนะนำ เพื่อจะได้ทำการแนะนำให้กับผู้ใช้หรือลูกค้าต่อไป

แต่เวลาการคำนวณหาไอเท็มสำหรับแนะนำแก่แอคทีฟยูสเซอร์นั้น จะเพิ่มขึ้นตามจำนวนไอเท็มที่อยู่ในระบบ ยิ่งมีจำนวนไอเท็มมากเวลาที่ผู้ใช้ต้องรอคอยการแนะนำจากระบบก็นานขึ้นตามไปด้วย นักวิจัยทางด้านระบบแนะนำจึงเกิดแนวคิดนำอัลกอริทึมสำหรับจัดกลุ่มข้อมูลเข้ามาใช้ร่วม เพื่อที่จะลดเวลาในการแนะนำไอเท็มแก่ผู้ใช้

สำหรับงานวิจัยเรื่องการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนนี้ ผู้วิจัยได้นำเสนออีกแนวทางหนึ่งสำหรับการแนะนำไอเท็มแบบท็อปเอ็น ด้วยการนำข้อมูลรายละเอียดไอเท็ม [5] อันประกอบด้วย รหัสลูกค้า (User/Customer ID), รหัสไอเท็ม (Item ID), และค่าเรตติ้ง (Rating) ไปทำการจัดกลุ่มด้วยโมดิฟายฟัซซีซีมีนก่อน เมื่อต้องการแนะนำไอเท็มแก่แอคทีฟยูสเซอร์ก็จะหากลุ่มไอเท็มที่เหมาะสม โดยการหาค่าเฉลี่ยเรตติ้งของไอเท็มที่แอคทีฟ-ยูสเซอร์เคยให้เรตติ้งเอาไว้ แล้วจึงเปรียบเทียบไอเท็มดังกล่าวกับไอเท็มอื่นๆ ในกลุ่มเดียวกันตามหลักการของการแนะนำแบบไอเท็มเบส เพื่อนำมาสร้างรายการแนะนำเพื่อแนะนำไอเท็มให้แก่แอคทีฟยูสเซอร์โดยเรียงลำดับค่าผลลัพธ์จากมากไปน้อย

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาวิธีการจัดกลุ่มไอเท็มด้วยการใช้โมดิฟายฟัซซีซีมีน
2. เพื่อศึกษาวิธีการแนะนำไอเท็มแบบไอเท็มเบสท็อปเอ็น โดยนำข้อมูลไอเท็มไปจัดกลุ่มด้วยโมดิฟายฟัซซีซีมีนก่อน

1.3 ขอบเขตการวิจัย

1. การทดลองทำบนระบบแบบปิด (Off-line)
2. ข้อมูลที่นำมาทดลองเป็นแบบคงที่ไม่ใช่แบบเพิ่มขึ้นอยู่เสมอ (Incremental)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. ไม่ครอบคลุมเรื่องปัญหาข้อมูลเบาบาง (Data sparse)

1.4 วิธีการดำเนินงานวิจัย

1. ศึกษาแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับการจัดกลุ่มไอเท็มเพื่อกรองข้อมูลหรือระบบแนะนำ [7, 8, 16] วิธีการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น [9]

2. ปรับอัลกอริทึมการจัดกลุ่มแบบฟัซซีซีมีน ในส่วนของการหาค่าสมาชิกของไอเท็มโดยนำวิธีวัดค่าความคล้ายคลึงด้วยวิธีการหาค่าสหสัมพันธ์ มาประยุกต์ใช้แทนการคำนวณระยะ ห่างระหว่างไอเท็ม

3. ออกแบบรวมถึงดำเนินการทดลอง การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน

4. สรุปผลการทดลอง

5. เรียบเรียงเอกสารประกอบวิทยานิพนธ์

1.5 การจัดเรียงหัวข้อในการเรียบเรียงวิทยานิพนธ์

วิทยานิพนธ์ฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทด้วยกันคือ

บทที่ 1 กล่าวถึงความเป็นมาของงานวิจัย ความมุ่งหมายและวัตถุประสงค์ของการศึกษา ขอบเขตของการวิจัย และขั้นตอนการศึกษา

บทที่ 2 กล่าวถึงทฤษฎีพื้นฐานและงานวิจัยที่เกี่ยวข้อง ซึ่งมีเนื้อหาเกี่ยวข้องกับฟัซซีเซ็ต การจัดกลุ่มข้อมูลแบบฟัซซีซีมีน วิธีการหาค่าสหสัมพันธ์ วิธีการเคโฟล์วาลิเดชั่น และงานวิจัยด้านระบบแนะนำ งานวิจัยด้านการจัดกลุ่มข้อมูลสำหรับการทำนายค่าคะแนน

บทที่ 3 กล่าวถึงรูปแบบข้อมูล ขั้นตอน การการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน

บทที่ 4 กล่าวถึงผลทดลองและสรุปผล

บทที่ 5 กล่าวถึงบทสรุปภาพรวมของวิทยานิพนธ์ ผลการวิจัยและข้อเสนอแนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีพื้นฐานและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีพื้นฐาน งานวิจัยด้านการกรองข้อมูลแบบไอเท็มเบส การจัดกลุ่มข้อมูลเพื่อทำนายค่าคะแนน และงานวิจัยด้านระบบแนะนำที่เกี่ยวข้องกับการทดลองเรื่องการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน

2.1 ฟัซซีเซต (Fuzzy set)

แนวคิดแบบฟัซซี่เซ็ต (Fuzzy sets) [11] ถูกนำเสนอครั้งแรกในปี 1965 โดย Lotfi Zadeh เป็นเครื่องมือทางภาษาคณิตศาสตร์ที่แสดงถึงความคลุมเครือ แนวคิดพื้นฐานของฟัซซีใช้หลักการเหตุผลซึ่งคล้ายการเลียนแบบวิธีความคิดที่ซับซ้อนของมนุษย์ กล่าวคือค่าความเป็นสมาชิกในเซ็ตมีไม่ได้จำกัดแค่ 0 หรือ 1 [0,1] ทว่ามีค่าความเป็นสมาชิกได้หลากหลายอยู่ในช่วงระหว่าง 0 ถึง 1

2.1.1 นิยามของฟัซซีเซ็ต

สำหรับเซ็ตแบบฉบับ (classical set) หรือเซ็ตทวินัย (crisp set) เป็นเซ็ตที่มีค่าความเป็นสมาชิกเพียง 2 ค่า คือ 0 กับ 1 หรือ {0,1} ซึ่งหมายถึงเป็นสมาชิกของเซ็ต หรือไม่เป็นสมาชิกของเซ็ต จริงหรือเท็จเท่านั้น ซึ่งแตกต่างจากฟัซซีเซ็ตคือเป็นเซ็ตแบบไม่ยืดหยุ่น แยกจากกันโดยเด็ดขาด

ให้ U เป็นเอกภพสัมพัทธ์ ดังนั้นฟังก์ชันความเป็นสมาชิก $\mu_A(x)$ ของเซ็ตทวินัย A ใน U จะมีค่าแค่ {0, 1} หาก $\mu_A(x)=1$ ถ้า x เป็นสมาชิกของ A และ $\mu_A(x)=0$ ถ้า x ไม่เป็นสมาชิกของ A เขียนเป็นรูปแบบทางคณิตศาสตร์ได้ดังนี้

$$\mu_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \tag{2.1}$$

ฟัซซีเซตจะมีขอบเขตแบบที่ไม่ได้เปลี่ยนแปลงทันทีทันใด ยกตัวอย่างเช่น เซตของคู่หย่าร้างที่ไม่มีความสุข สมาชิกของเซตนี้จะเป็นคู่หย่าร้างที่ไม่ได้รู้สึกถึงความทุกข์ในระดับเดียวกันทั้งหมด บางคู่อาจจะมีความทุกข์มาก บางคู่มีความทุกข์น้อย ลดหลั่นแตกต่างกันไป การใช้เซตแบบดั้งเดิมจึงไม่สามารถแสดงสมาชิกภาพได้ทั้งหมด

ให้ X เป็นเซ็ตที่ไม่ใช่เซ็ตว่าง ดังนั้นจะได้ฟัซซีเซ็ต A คือ

$$\mu_{\tilde{A}}(x): X \rightarrow [0,1] \tag{2.2}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อ $\mu_{\tilde{A}}$ หมายถึงค่าความเป็นสมาชิกภาพของตัวประกอบ x ในเซต $\tilde{A}$ สำหรับแต่ละ $x \in X$ ดังนั้น ฟัซซีเซ็ต A จะสามารถเขียนในรูปแบบของเซตคู่ลำดับได้เป็น

$$\tilde{A} = \{(x, \mu_{\tilde{A}}(x)) \quad | \quad x \in X \tag{2.3}$$

เมื่อ $\tilde{A}$ หมายถึง ฟัซซีเซต A

x หมายถึง สมาชิกของเซ็ต

$\mu_{\tilde{A}}$ หมายถึง ฟังก์ชันความเป็นสมาชิกของฟัซซีเซ็ต อาจจะเขียนแทนด้วย $\tilde{A}(x)$ ซึ่งสามารถมีค่าต่อเนื่องกันได้ตั้งแต่ 0 ถึง 1

X หมายถึง เอกภพสัมพัทธ์

2.1.2 โอเปอร์เรชั่นของฟัซซีเซ็ต (Operation on Fuzzy set)

- ยูเนียน (Union) ของฟัซซีเซ็ตมีสมการดังนี้

$$\mu_{\tilde{A} \cup \tilde{B}}(x) = \mu_{\tilde{A}}(x) \vee \mu_{\tilde{B}}(x) \tag{2.4}$$

หรือ $$= \max \mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x) \tag{2.5}$$

- อินเตอร์เซกชัน (Intersection) ของฟัซซีเซ็ตมีสมการดังนี้

$$\mu_{\tilde{A} \cap \tilde{B}}(x) = \mu_{\tilde{A}}(x) \wedge \mu_{\tilde{B}}(x) \tag{2.6}$$

หรือ $$= \min \mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x) \tag{2.7}$$

- คอมพลีเมนต์ (Complement) หรือส่วนเติมเต็มของฟัซซีเซ็ตมีสมการดังนี้

$$\mu_{\tilde{A}}(x) = 1 - \mu_{\tilde{A}}(x) \tag{2.8}$$

2.2 ฟัซซีซีมีน (Fuzzy C-Mean)

ฟัซซีซีมีนเป็นกรรมวิธีที่ใช้เพื่อจัดกลุ่มข้อมูล (Cluster) ซึ่งถูกพัฒนาขึ้นโดย Dunn [12] และปรับปรุงโดย Bezdek [13] ได้พัฒนาและปรับปรุงมาจากอัลกอริทึมเคมีน (K-Mean) [14] โดยที่ฟัซซีซีมีนยินยอมให้ข้อมูลหนึ่งๆ ที่กำลังพิจารณาสามารถเป็นสมาชิกของหลายกลุ่มข้อมูลได้พร้อมๆ กัน ในขณะที่เคมีนนั้นกำหนดให้ข้อมูลเป็นสมาชิกเพียงกลุ่มใดกลุ่มหนึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเป็นสมาชิกในหลายกลุ่มข้อมูลของฟัซซีซีมีนนั้น กล่าวคือ ข้อมูลหนึ่งๆ จะมีค่าความเป็นสมาชิกของแต่ละกลุ่มที่แตกต่างกันออกไป ผลจากการใช้ฟัซซีเข้ามาช่วยในการจัดกลุ่มข้อมูลนี้ทำให้การจัดกลุ่มข้อมูลที่มีความคลุมเครือได้อย่างมีประสิทธิภาพ

ขั้นตอนการทำงานของฟัซซีซีมีนเป็นแบบวนซ้ำ (Iterative) ในแต่ละรอบจะมีการปรับค่าความเป็นสมาชิกและเวกเตอร์ศูนย์กลางของกลุ่ม จนกว่าค่าทั้งสองไม่มีการเปลี่ยนแปลงใดๆ ทั้งนี้ก็เพื่อหาค่าเหมาะที่สุดของฟังก์ชันวัตถุประสงค์ (Objective function)

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} \mu_{ij} \left\| x_i - c_j \right\|^2, 1 \le m \le \infty \qquad (2.9)$$

โดยที่

m หมายถึง ค่าสัมประสิทธิ์ของความเป็นฟัซซีซึ่งมีค่าอยู่ระหว่าง $1 \le m \le \infty$ โดยทั่วไปแล้ว มักจะให้ค่า m เท่ากับ 2 (m เข้าใกล้ 1 เวกเตอร์ศูนย์กลางของกลุ่มข้อมูลก็จะยิ่งมีน้ำหนักมากกว่าค่าอื่นๆ ในกลุ่ม)

x_i หมายถึง เวกเตอร์ข้อมูลลำดับที่ i ขนาด d มิติ

μ_{ij} หมายถึง ค่าความเป็นสมาชิกของข้อมูลตัวที่ x_i ในกลุ่มที่ j อยู่ระหว่าง 0 ถึง 1

c_j หมายถึง เวกเตอร์ขนาด d มิติที่เป็นศูนย์กลางของกลุ่ม

N หมายถึง จำนวนข้อมูลในกลุ่ม

C หมายถึง จำนวนกลุ่มข้อมูลทั้งหมด

d_{ij} หมายถึงระยะห่างยูคลิเดียนระหว่าง i_{th} ศูนย์กลางของกลุ่ม C_i กับจุดของข้อมูล j_{th} หรือกลุ่มข้อมูลที่กำลังสนใจ

d_{ik} หมายถึงระยะห่างยูคลิเดียนระหว่าง i_{th} ศูนย์กลางของกลุ่ม C_k กับจุดของข้อมูล j_{th} หรือกลุ่มข้อมูลอื่นๆ

k หมายถึง ลำดับรอบที่กระทำซ้ำ

กำหนดค่าเริ่มต้นให้กับเมตริกซ์ (U) ซึ่งเป็นเมตริกซ์แสดงถึงความเป็นสมาชิกระหว่างข้อมูลกับกลุ่มข้อมูลตามสมการ 2.10

$$\sum_{i=1}^{c} \mu_{ij} = 1, \forall_j = 1, \ldots, n \qquad (2.10)$$

ศูนย์กลางของกลุ่ม (Centroid: Center of Cluster) คำนวณได้จากสมการ 2.11

$$C_i = \frac{\sum_{j=1}^{n} \mu_{ij}^{\ m} x_j}{\sum_{j=1}^{n} \mu_{ij}^{\ m}} \qquad (2.11)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าความเป็นสมาชิกของข้อมูลที่มีต่อกลุ่มข้อมูล คำนวณได้จากสมการ 2.12

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{d_{ij}}{d_{kj}}\right)^{1/(m-1)}} \qquad (2.12)$$

รายละเอียดของอัลกอริทึมสำหรับคำนวณหาการจัดกลุ่มแบบฟัซซีซีมีน [15] นำเสนอโดย Bezdek ในปี 1973 มีขั้นตอนดัง Algorithm 2.1

ตารางที่ 2.1 แสดงอัลกอริทึมของฟัซซีซีมีน

Algorithm 2.1 : Fuzzy C-Mean Clustering (data)

1. สร้างเมตริกซ์ $U = [\mu_{ij}]$ ตามสมการ 2.10 สำหรับจัดเก็บค่าความเป็นสมาชิกระหว่างไอเท็มกับกลุ่มไอเท็ม
2. คำนวณหาเวกเตอร์ศูนย์กลางของกลุ่มในรอบที่ k ตามสมการ 2.11
3. ปรับค่าความเป็นสมาชิกตามสมการ 2.12
4. ตรวจสอบเงื่อนไขหาก $\left\|U^{k+1} - U^{k}\right\| < \varepsilon$ หยุดการคำนวณ หากยังไม่ตรงกับเงื่อนไขให้กลับขึ้นวนซ้ำที่ข้อ 2 ใหม่

2.3 การวัดค่าความคล้ายคลึง (Similarity Measurement)

ความคล้ายคลึงหรือความแตกต่างกันเพียงเล็กน้อยระหว่างสิ่งสองสิ่ง คือการหาปริมาณที่สะท้อนถึงความสัมพันธ์ระหว่างสิ่งสองสิ่งนักวิจัยหลายต่อหลายท่านได้คิดหาสมการคณิตศาสตร์สำหรับวัดค่าความคล้ายคลึงระหว่างวัตถุ ทั้งนี้ขึ้นอยู่กับลักษณะและขนาดของสิ่งที่ต้องการวัด ผู้วิจัยขอกล่าวถึงวิธีการวัดค่าความคล้ายคลึงเพียง 2 แบบที่เกี่ยวข้องกับงานวิจัยฉบับนี้

2.3.1 วิธีวัดค่าความคล้ายคลึงแบบทั่วไป

2.3.1.1 วิธีการวัดค่ามุมโคซายน์ของเวคเตอร์ (Cosine-Based)

วิธีการวัดค่าความคล้ายคลึงโดยอาศัยการหาค่ามุมโคโซน์ระหว่างสองเวคเตอร์ ผลลัพธ์จะได้จากการการดอทโปรดัค (DOT Product) ของสองเวคเตอร์ที่ต้องการวัดค่าความคล้ายคลึง ตามสมการ 2.13

$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} \qquad (2.13)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าความคล้ายคลึงที่ได้จากสมการ 2.13 จะอยู่ระหว่าง 0 ถึง 1 ยิ่งค่าเข้าใกล้ 1 แสดงว่าสองสิ่งที่ต้องการวัดนั้นมีความคล้ายคลึงกันมาก หากค่ายิ่งเข้าใกล้ 0 หมายถึงสิ่งที่ต้องการวัดนั้นยิ่งมีความคล้ายคลึงกันน้อย หรือแตกต่างกันมาก

2.3.1.2 วิธีการวัดระยะห่างยูคลิเดียน (Euclidean Distance)

เป็นการวัดค่าความคล้ายคลึงด้วยวิธีการหาระยะห่าง ผลลัพธ์จะได้จากรากที่สองของความแตกต่างระหว่างพิกัดของคู่ของวัตถุ ตามสมการ 2.13

$$d_{ij} = \sqrt{\sum_{k=1}^{n} (x_{ik} - x_{jk})^2} \qquad (2.14)$$

การแปรความหมายจากผลลัพธ์ของสมการ 2.14 จะเป็นแบบผกผัน คือหากค่าที่คำนวณได้ยิ่งน้อยนั่นหมายถึงทั้งสองสิ่งที่นำมาวัดคล้ายคลึงกันมาก หากว่าค่ายิ่งมากแสดงว่าสองสิ่งมีความคล้ายคลึงกันน้อย

2.3.2 วิธีวัดค่าความคล้ายคลึงโดยใช้การหาค่าสหสัมพันธ์

สัมประสิทธิ์สหสัมพันธ์เป็นค่าที่ได้จากการหาความสัมพันธ์ระหว่างสองตัวแปร เช่น คำนวณจากค่าเรตติ้งที่ผู้ใช้มีต่อสินค้าชิ้นเดียวกัน (Co-rating) ค่าสัมประสิทธิ์สหสัมพันธ์จะอยู่ระหว่าง -1 ถึง +1 ซึ่งแปลความหมายได้ดังนี้

เข้าใกล้ +1 คือ ตัวแปรสองตัวมีความสัมพันธ์หรือคล้ายคลึงกันมากในทิศทางเดียวกัน

เข้าใกล้ -1 คือ ตัวแปรทั้งสองมีความสัมพันธ์กันแบบผกผัน

ค่า 0 นั่นหมายถึงสองตัวแปรไม่มีความสัมพันธ์กันหรือไม่มีความคล้ายคลึงกันเลย

2.3.3.1 วิธีการหาสหสัมพันธ์แบบเพียร์สัน [6] (Correlation Coefficient)

คำนวณได้จากสมการ 2.16

$$Sim(t,c) = \frac{\sum_{i \in I} (R_{i,t} - \overline{R}_t)(R_{i,c} - \overline{R}_c)}{\sqrt{\sum_{i \in I} (R_{i,t} - \overline{R}_t)^2} \sqrt{\sum_{i \in I} (R_{i,c} - \overline{R}_c)^2}} \qquad (2.15)$$

2.4 วิธีการเคโฟล์ครอสวาลิเดชั่น (K-Fold Cross Validation)

พื้นฐานของวิธีการแบบเคโฟล์ครอสวาลิเดชั่น [17] คือการสุ่มตัวอย่างข้อมูลลงในแต่ละชุดข้อมูลที่ถูกแบ่งออกเป็นจำนวน K ชุด โดยทำการคำนวณทั้งหมด K รอบ แต่ละรอบการคำนวณจะเลือกข้อมูลชุดหนึ่งมาเป็นชุดข้อมูลทดสอบและชุดข้อมูลที่เหลือใช้เป็นข้อมูลสำหรับเทรนนิ่ง แล้วจึงนำผลลัพธ์จากการคำนวณในแต่ละรอบมาหาค่าเฉลี่ย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ยกตัวอย่างเช่น กำหนดให้ค่า K = 5 นั่นคือแบ่งชุดข้อมูลใหญ่ออกเป็น 5 ชุด แต่ละชุดจำมีข้อมูลในจำนวนเท่าๆ กัน

ข้อมูลชุดที่1	ข้อมูลชุดที่2	ข้อมูลชุดที่3	ข้อมูลชุดที่4	ข้อมูลชุดที่5

รอบการคำนวณที่ 1 :

ข้อมูลชุดที่1	ข้อมูลชุดที่2	ข้อมูลชุดที่3	ข้อมูลชุดที่4	ข้อมูลชุดที่5

รอบการคำนวณที่ 2 :

ข้อมูลชุดที่1	ข้อมูลชุดที่2	ข้อมูลชุดที่3	ข้อมูลชุดที่4	ข้อมูลชุดที่5

รอบการคำนวณที่ 3 :

ข้อมูลชุดที่1	ข้อมูลชุดที่2	ข้อมูลชุดที่3	ข้อมูลชุดที่4	ข้อมูลชุดที่5

รอบการคำนวณที่ 4 :

ข้อมูลชุดที่1	ข้อมูลชุดที่2	ข้อมูลชุดที่3	ข้อมูลชุดที่4	ข้อมูลชุดที่5

รอบการคำนวณที่ 5 :

ข้อมูลชุดที่1	ข้อมูลชุดที่2	ข้อมูลชุดที่3	ข้อมูลชุดที่4	ข้อมูลชุดที่5

คำตอบ = ผลการคำนวณรอบที่ 1 + ผลการคำนวณรอบที่ 2 +ผลการคำนวณรอบที่ 3 + ผลการคำนวณรอบที่ 4 + ผลการคำนวณรอบที่ 5 / 5 (จำนวน K)

2.5 งานวิจัยที่เกี่ยวข้อง

หลายปีที่ผ่านมามีงานวิจัยด้านเทคโนโลยีการกรองข้อมูลเพื่อนำสารสนเทศที่ได้ไปใช้ในการทำนาย หรือสร้างระบบแนะนำมากมาย โดยมีจุดมุ่งหมายที่จะปรับปรุงหรือพัฒนาแนวคิดให้การทำงานของระบบมีประสิทธิภาพมากยิ่งขึ้น

งานวิจัยด้านระบบแนะนำนั้นถูกแบ่งออกเป็น 2 รูปแบบ นั่นคือ 1. ไอเท็มหรือโมเดลเบส (Item/Model -Based) และ 2. แบบคอนเท็นหรือยูสเซอร์เบส (Content / User-Based) ทั้งสองแบบจะแตกต่างกันตรงสิ่งที่นำมาเปรียบเทียบ แบบไอเท็มเบสจะนำคุณสมบัติของไอเท็มมาใช้อ้างอิง ส่วนยูสเซอร์เบสนั้นจะใช้คุณสมบัติของผู้ใช้มาพิจารณา ในที่นี้ขอกล่าวถึงงานวิจัยด้านไอเท็มเบสที่เกี่ยวข้องกับงานวิจัยฉบับนี้

2.5.1 งานวิจัยด้านการแนะนำข้อมูลแบบไอเท็มเบสท๊อปเอ็น

ผลงานวิจัยของ Mukmund Deshpande และ George Karypis แห่งมหาวิทยาลัยมินเนโซต้า เรื่อง Item-Based Top-*N* Recommendation Algorithms [9] นั้นเป็นอีกงานหนึ่งที่ศึกษาหลักการสร้างระบบแนะนำข้อมูลแบบไอเท็มเบส โดยใช้การวิเคราะห์เมตริกซ์ยูสเซอร์-ไอเท็ม (User-Item)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

(หลัก * แถว) แสดงดังตาราง 2.1 มาระบุความสัมพันธ์ระหว่างไอเท็มที่หลากหลายในระบบ ด้วยวิธีการหาค่าความคล้ายคลึงของไอเท็ม จากนั้นนำความสัมพันธ์ระหว่างไอเท็มที่ได้มาสร้างรายการไอเท็มหรือรายการแนะนำแก่ลูกค้าหรือที่เรียกว่าท๊อปเอ็น (Top-*N*)

วิธีการแนะนำสินค้าแบบไอเท็มเบสท๊อปเอ็นนั้นแบ่งขั้นตอนหลักได้เป็น 4 ขั้นตอนดังนี้

1. การสร้างโมเดล หมายถึงการสร้างโมเดลของไอเท็มในระบบ โดยการเปรียบเทียบความคล้ายคลึงระหว่างไอเท็มทั้งหมด จะได้เป็นข้อมูลในรูปแบบไอเท็ม – ไอเท็มเมตริกซ์

2. การประยุกต์ใช้โมเดล หมายถึงการนำไอเท็ม-ไอเท็มเมตริกซ์ไปใช้ในการแนะนำไอเท็มให้แก่แอคทีฟยูสเซอร์ ด้วยการคูณกับ 1 * m เมตริกซ์ ซึ่งเป็นเมตริกซ์ที่แสดงถึงความสัมพันธ์ระหว่างแอคทีฟยูสเซอร์กับไอเท็มภายในระบบ

3. สร้างรายการแนะนำ หมายถึงการนำค่าในเมตริกซ์ผลลัพธ์ที่ได้จากข้อ 2 มาเรียงลำดับจากมากไปหาน้อย ตัดไอเท็มตัวที่ผู้ใช้เคยเลือกและมีค่าเป็นศูนย์ออกไป

4. การแนะนำแก่แอคทีฟยูสเซอร์ หมายถึงการนำรายการแนะนำที่สร้างจากในข้อ 3 ไปทำการแนะนำแก่แอคทีฟยูสเซอร์ตามจำนวนรายการที่กำหนด

ยกตัวอย่างเช่น ระบบจะทำการแนะนำสินค้าให้แก่แอคทีฟยูสเซอร์รหัส U_7 ซึ่งมีข้อมูลการเลือกไอเท็มแสดงดังตาราง 2.3 กล่าวคือผู้ใช้รหัส U_7 เคยเลือกไอเท็ม 1 ไอเท็ม 2 และไอเท็ม 7 มาในอดีต ทำการสร้าง 1 * m เมตริกซ์ของ U_7 ขึ้นมา และนำไปคูณกับเมตริกซ์ไอเท็ม-ไอเท็มแสดงในตาราง 2.4 เมื่อได้เมตริกซ์ผลลัพธ์เป็นที่เรียบร้อยแล้ว จึงสร้างรายการแนะนำจากการนำค่าในเมตริกซ์ผลลัพธ์ไปเรียงลำดับจากมากที่สุดไปน้อยที่สุด และตัดไอเท็มที่ U_7 เคยเลือกกับไอเท็มที่มีค่าเป็นศูนย์ออกจากรายการแนะนำ จากข้อมูลตัวอย่างนี้ ไอเท็มแนะนำสำหรับ U_7 ได้แก่ I_7, I_3, I_6, I_8, I_9, I_{10} และ I_5

ตารางที่ 2.2 แสดงชุดข้อมูลสินค้าในรูปผู้ใช้-ไอเท็ม เมตริกซ์

รหัสผู้ใช้	รหัสไอเท็ม									
	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
U_1	1	1	1	1	1	1	1	1	1	1
U_2	1	0	1	1	0	1	1	1	1	1
U_3	1	1	1	1	1	1	1	1	1	1
U_4	0	0	1	1	0	1	0	1	1	1
U_5	1	1	1	1	1	1	1	1	1	1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.3 แสดงไอเท็ม-ไอเท็มเมตริกซ์ (ชุดข้อมูลสินค้า)

รหัสไอเท็ม	รหัสไอเท็ม									
	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
I_1	1	0.866025	0.894427	0.894427	0.866025	0.894427	1	0.894427	0.894427	0.894427
I_2	0.866025	1	0.774597	0.774597	1	0.774597	0.866025	0.774597	0.774597	0.774597
I_3	0.894427	0.774597	1	1	0.774597	1	0.894427	1	1	1
I_4	0.894427	0.774597	1	1	0.774597	1	0.894427	1	1	1
I_5	0.866025	1	0.774597	0.774597	1	0.774597	0.866025	0.774597	0.774597	0.774597
I_6	0.894427	0.774597	1	1	0.774597	1	0.894427	1	1	1
I_7	1	0.866025	0.894427	0.894427	0.866025	0.894427	1	0.894427	0.894427	0.894427
I_8	0.894427	0.774597	1	1	0.774597	1	0.894427	1	1	1
I_9	0.894427	0.774597	1	1	0.774597	1	0.894427	1	1	1
I_{10}	0.894427	0.774597	1	1	0.774597	1	0.894427	1	1	1

ตาราง 2.4 แสดงเมตริกซ์ 1 * *m* ในการแนะนำแบบไอเท็มเบสท็อปเอ็น

Item	U_7
I_1	1
I_2	1
I_3	0
I_4	1
I_5	0
I_6	0
I_7	0
I_8	0
I_9	0
I_{10}	0

ตารางที่ 2.5 แสดงเมตริกซ์ผลลัพธ์

$$\begin{bmatrix} 2.760452 & I_1 \\ 2.640622 & I_2 \\ 2.669024 & I_3 \\ 2.669024 & I_4 \\ 2.634847 & I_5 \\ 2.669024 & I_6 \\ 2.760452 & I_7 \\ 2.669024 & I_8 \\ 2.669024 & I_9 \\ 2.669024 & I_{10} \end{bmatrix}$$

จะเห็นได้ว่าขนาดเมตริกซ์ของไอเท็ม-ไอเท็มนั้น มีขนาดเท่ากับจำนวนไอเท็มที่อยู่ในระบบ ยิ่งจำนวนไอเท็มมีมากขึ้น เวลาในการคำนวณหาไอเท็มเมตริกซ์ก็เพิ่มตามไปด้วย (ไม่นับรวมเวลาในการสร้างไอเท็ม-ไอเท็มเมตริกซ์) ซึ่งแน่นอนว่าเวลาที่ใช้ในการแนะนำนั้นต้องเพิ่มตามจำนวนของไอเท็ม

เพื่อที่จะลดเวลาในการคำนวณหาเมตริกซ์ผลลัพธ์ไปสร้างรายการแนะนำลง นักวิจัยด้านการสร้างระบบแนะนำจึงใช้อัลกอริทึมการจัดกลุ่มข้อมูลเข้ามาประยุกต์ใช้ แล้วจึงค่อยเปรียบเทียบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หากลุ่มไอเท็มที่เหมาะสมแก่แอคทีฟยูสเซอร์ แล้วจึงค่อยสร้างรายการไอเท็มเพื่อทำการแนะนำจากกลุ่มไอเท็มนั้น

2.5.2 งานวิจัยการจัดกลุ่มข้อมูลไอเท็มเพื่อทำนายค่าคะแนนและการแนะนำ

การนำไอเท็มมาจัดกลุ่มก่อนแทนการเปรียบเทียบหาค่าความคล้ายคลึงไอเท็มทั้งหมดในระบบจะช่วยให้การคำนวณนั้นลดลง เนื่องจากขอบเขตไอเท็มที่จะนำมาเปรียบเทียบแคบลง งานวิจัยเรื่อง An Item Based Collaborative Filtering Using Item Clustering Prediction [7] ได้นำอัลกอริทึมเคมีนมาใช้ในการจัดกลุ่มข้อมูล ข้อดีของเคมีนคือเป็นอัลกอริทึมที่ไม่ซับซ้อนและประยุกต์ใช้งานได้ง่าย แต่ตัวอัลกอริทึมเองก็มีข้อจำกัดดังที่ได้กล่าวไปแล้วในหัวข้อ 2.2 งานวิจัยเรื่อง A Hybrid Clustering Based Filtering Approach with Efficient Sequencing [8] จึงได้นำอัลกอริทึมฟัซซีซีมีนเข้ามาช่วยในการจัดกลุ่มไอเท็ม

สำหรับการจัดกลุ่มไอเท็มเพื่อนำไปใช้หารายการไอเท็มเพื่อทำการแนะนำนั้น จะหากลุ่มไอเทมที่เหมาะสมแก่แอคทีฟยูสเซอร์ และคำนวณหารายการไอเท็มสำหรับแนะนำจากไอเท็มที่อยู่ภายในกลุ่ม โดยตัดไอเท็มตัวที่แอคทีฟยูสเซอร์ได้เลือกหรือให้ค่าคะแนนไว้แล้วออกจากรายการ

เนื่องจากข้อมูลที่นำมาใช้ในงานวิจัยการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนนี้มีเซตข้อมูลของค่าคะแนนเรตติ้งจากผู้ใช้ต่อแต่ละไอเท็มประกอบอยู่ด้วย ผู้วิจัยจึงได้ปรับปรุงอัลกอริทึมฟัซซีซีมีนสำหรับจัดกลุ่มไอเท็ม โดยนำวิธีการหาค่าสหสัมพันธ์เข้ามาประยุกต์ใช้ ในส่วนของการหาค่าความเป็นสมาชิกของไอเท็มที่มีต่อกลุ่มข้อมูล ซึ่งจะกล่าวถึงในบทถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การแนะนำสินค้าแบบไอเท็มเบสท๊อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน

การนำไอเท็มทั้งหมดในระบบไปจัดกลุ่มไอเท็มก่อนดำเนินการหารายการไอเท็มที่จะแนะนำแก่แอคทีฟยูสเซอร์ จะช่วยลดเวลาการคำนวณเพื่อสร้างรายการไอเท็ม เนื่องด้วยการคำนวณนี้จะเปรียบเทียบเฉพาะไอเท็มภายในกลุ่มไอเท็มแต่ละกลุ่ม ไม่ได้ทำการเปรียบเทียบกับทุกไอเท็มที่มีอยู่ในระบบ

เคมีนเป็นอัลกอริทึมที่แพร่หลาย ไม่ซับซ้อนและประยุกต์ใช้งานได้ง่าย ทว่าก็มีข้อจำกัดในเรื่องของความไม่ยืดหยุ่น กล่าวคือเคมีนนั้นกำหนดให้ข้อมูลที่กำลังพิจารณาเป็นสมาชิกเพียงกลุ่มใดกลุ่มหนึ่ง ผู้วิจัยจึงได้เลือกใช้ฟัซซีซีมีนซึ่งเป็นอัลกอริทึมที่พัฒนามาจากเคมีนแต่แตกต่างตรงที่ยืดหยุ่นกว่า ฟัซซีซีมีนยินยอมให้ข้อมูลหนึ่งๆ ที่กำลังพิจารณาสามารถเป็นสมาชิกของหลายกลุ่มข้อมูลได้พร้อมๆ กัน โดยที่ข้อมูลหนึ่งๆ จะมีค่าความเป็นสมาชิกของแต่ละกลุ่มที่แตกต่างกันออกไป ผลจากการใช้ฟัซซีเข้ามาช่วยในการจัดกลุ่มข้อมูลนี้ทำให้จัดกลุ่มข้อมูลที่มีความคลุมเครือได้อย่างมีประสิทธิภาพ

และเนื่องจากการแนะนำสินค้าแบบไอเท็มเบสนั้นอยู่บนแนวคิดพื้นฐานที่ว่า ไอเท็มที่จะแนะนำแก่แอคทีฟยูสเซอร์คล้ายคลึงกับไอเท็มที่แอคทีฟยูสเซอร์เคยเลือกซื้อไปแล้วในอดีต ดังนั้นผู้วิจัยจึงได้ทำการปรับสมการสำหรับหาค่าความเป็นสมาชิกของไอเท็มที่มีต่อกลุ่มไอเท็ม ด้วยการหาความคล้ายคลึงด้วยค่าสหสัมพันธ์จากข้อมูลที่เกิดร่วมกัน (Co-Rate) มาประยุกต์ใช้แทนการคำนวณหาค่าระยะห่างระหว่างไอเท็มกับจุดศูนย์กลางของกลุ่มไอเท็ม

3.1 โมดิฟายฟัซซีซีมีน

โดยทั่วไปแล้ว สมการหาค่าความเป็นสมาชิกแบบฟัซซีซีมีนที่งานวิจัย [15] นำมาใช้ จะมีรูปแบบดังสมการ 3.1 คือ

$$\mu_{c_i}(x) = \frac{1}{\sum_{k=1}^{C}\left(\frac{\|x-v_i\|^2}{\|x-v_k\|^2}\right)^{1/(m-1)}} \tag{3.1}$$

v_i คือ ศูนย์กลางของกลุ่มไอเท็มที่กำลังพิจารณาความเป็นสมาชิก

v_k คือ ศูนย์กลางของกลุ่มไอเท็มอื่นๆ ที่เหลือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมื่อนำการหาค่าสหสัมพันธ์สมการ 2.15 มาใช้แทนระยะห่างระหว่างไอเท็ม จะได้สมการสำหรับหาค่าความเป็นสมาชิกที่มีรูปแบบดัง 3.2

$$\mu_{sim\,c_i}(x) = \frac{1}{\sum_{k=1}^{C}\left(\frac{[1 - Sim(x, v_i)]}{[1 - Sim(x, v_k)]}\right)^{1/(m-1)}} \tag{3.2}$$

โดยที่

$$Sim\ (x, v_i) = RSim\ (t, c) = \frac{\sum_{i \in I}(R_{i,t} - \overline{R_t})(R_c - \overline{R_c})}{\sqrt{\sum_{i \in I}(R_{i,t} - \overline{R_t})^2}\sqrt{\sum_{i \in I}(R_c - \overline{R_c})^2}} \tag{3.3}$$

และ

$$Sim\ (x, v_k) = RSim\ (t, k) = \frac{\sum_{i \in I}(R_{i,t} - \overline{R_t})(R_k - \overline{R_k})}{\sqrt{\sum_{i \in I}(R_{i,t} - \overline{R_t})^2}\sqrt{\sum_{i \in I}(R_k - \overline{R_k})^2}} \tag{3.4}$$

เมื่อ

$R_{i,t}$ คือ เรตติ้งของไอเท็มเป้าหมายโดยผู้ใช้คนที่ i

$\overline{R_t}$ คือ ค่าเฉลี่ยเรตติ้งของไอเท็มเป้าหมายจากผู้ใช้ทุกคนที่ให้คะแนนแก่ไอเท็มเป้าหมายร่วมกัน

R_c คือ เรตติ้งของศูนย์กลางของกลุ่มที่กำลังพิจารณา

$\overline{R_c}$ คือ ค่าเรตติ้งเฉลี่ยของศูนย์กลางของกลุ่มไอเท็มที่กำลังพิจารณา

R_k คือ เรตติ้งของศูนย์กลางของกลุ่มที่เหลือ

$\overline{R_k}$ คือ ค่าเรตติ้งเฉลี่ยของศูนย์กลางของกลุ่มไอเท็มที่เหลือ

ดังนั้น จะได้สมการสำหรับคำนวณหาค่าความเป็นสมาชิกของไอเท็มในแต่ละกลุ่มไอเท็มดังสมการ 3.5

$$\mu_{simc_i}(x) = \frac{1}{\sum_{k=1}^{C}\left(\frac{1 - \left[\frac{\sum_{i \in I}(R_{i,t} - \overline{R_t})(R_c - \overline{R_c})}{\sqrt{\sum_{i \in I}(R_{i,t} - \overline{R_t})^2}\sqrt{\sum_{i \in I}(R_c - \overline{R_c})^2}}\right]}{1 - \left[\frac{\sum_{i \in I}(R_{i,t} - \overline{R_t})(R_k - \overline{R_k})}{\sqrt{\sum_{i \in I}(R_{i,t} - \overline{R_t})^2}\sqrt{\sum_{i \in I}(R_k - \overline{R_k})^2}}\right]}\right)^{1/(m-1)}} \tag{3.5}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 ขั้นตอนการดำเนินงาน

งานวิจัยการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน แบ่งขั้นตอนการทำงานออกเป็น 5 ขั้นตอนหลักดังต่อไปนี้

1. จัดกลุ่มไอเท็มด้วยโมดิฟายฟัซซีซีมีน (Cluster)
2. คำนวณหากลุ่มที่เหมาะสมสำหรับแอคทีฟยูสเซอร์
3. คำนวณหาไอเท็มภายในกลุ่มสำหรับสร้างรายการแนะนำ
3. สร้างรายการไอเท็มแนะนำ
4. แนะนำรายการไอเท็มแก่แอคทีฟยูสเซอร์

3.2.1 จัดกลุ่มไอเท็มด้วยโมดิฟายฟัซซีซีมีน (Cluster)

นำข้อมูลไอเท็มทั้งหมดมาจัดกลุ่มด้วยวิธีการโมดิฟายฟัซซีซีมีน ซึ่งยังคงมีลำดับขั้นตอนการทำงานหลักเหมือนฟัซซีซีมีนอัลกอริทึม (*Algorithm 2.1*) แต่ปรับปรุงสมการที่ใช้คำนวณหาค่าความเป็นสมาชิกระหว่างไอเท็มและกลุ่มไอเท็ม โดยนำวิธีการหาค่าสหสัมพันธ์มาประยุกต์ใช้แทนวิธีการหาระยะห่างแบบยูคลิเดียนระหว่างไอเท็มกับจุดศูนย์กลางของแต่ละกลุ่มไอเท็ม

สมมติชุดข้อมูลเรตติ้งสินค้าในระบบมีค่าดังตาราง 3.1 อันประกอบด้วยรหัสผู้ใช้ของผู้ใช้จำนวน 5 คน รหัสสินค้า 10 ชนิดและค่าเรตติ้งที่ผู้ใช้แต่ละรายให้แก่สินค้าแต่ละชนิด จัดไอเท็มเหล่านี้แบ่งออกเป็น 3 กลุ่ม

ตารางที่ 3.1 แสดงชุดข้อมูลค่าเรตติ้งสินค้าในรูปผู้ใช้-ไอเท็ม เมตริกซ์

รหัสผู้ใช้	รหัสไอเท็ม									
	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
U_1	3	3	1	2	1	4	2	5	3	3
U_2	2	0	5	0	3	5	2	4	2	2
U_3	5	4	2	4	5	2	4	5	3	5
U_4	0	0	3	0	4	4	1	3	4	2
U_5	2	1	5	2	5	3	5	2	1	4

นำข้อมูลในเมตริกซ์จากตาราง 3.1 มาทำการทรานโพส (Transpose) ได้เป็นเมตริกซ์ระหว่างไอเท็ม-ผู้ใช้ แสดงดังตาราง 3.2 เริ่มต้นการจัดกลุ่มไอเท็มด้วยโมดิฟายฟัซซีซีมีนด้วยการสร้างเมตริกซ์ (U) ซึ่งเป็นเมตริกซ์เก็บค่าความเป็นสมาชิกของไอเท็มทั้ง 10 ไอเท็มที่มีต่อกลุ่มไอเท็มทั้ง 3 กลุ่มไอเท็ม ด้วยวิธีการสุ่มเลือก (Random) ผลรวมค่าความเป็นสมาชิกของแต่ละไอเท็มที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

มีต่อกลุ่มไอเท็มทั้ง 3 จะต้องเท่ากับ 1 ตามสมการ 2.10 แสดงผลการสุ่มเลือกค่าความเป็นสมาชิกของแต่ละไอเท็มดังตาราง 3.3

ตารางที่ 3.2 แสดงชุดข้อมูลค่าเรตติ้งสินค้าในรูปไอเท็ม-ผู้ใช้ เมตริกซ์

รหัสไอเท็ม	รหัสผู้ใช้				
	U_1	U_2	U_3	U_4	U_5
I_1	3	2	5	0	2
I_2	3	0	4	0	1
I_3	1	5	2	3	5
I_4	2	0	4	0	2
I_5	1	3	5	4	5
I_6	4	5	2	4	3
I_7	2	2	4	1	5
I_8	5	4	5	3	2
I_9	3	2	3	4	1
I_{10}	3	2	5	2	4

ตารางที่ 3.3 แสดงค่าความเป็นสมาชิกของไอเท็มต่อแต่ละกลุ่มไอเท็ม

รหัสไอเท็ม	กลุ่มไอเท็ม		
	1	2	3
I_1	0.217549	0.680165	0.102286
I_2	0.115873	0.785547	0.09858
I_3	0.114236	0.093237	0.792527
I_4	0.600000	0.000000	0.400000
I_5	0.307721	0.334974	0.357305
I_6	0.237958	0.283906	0.478136
I_7	0.909608	0.039346	0.051046
I_8	0.260906	0.590029	0.149065
I_9	0.234984	0.520741	0.244275
I_{10}	0.722047	0.185740	0.092213

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
110608

คำนวณหาค่าศูนย์กลางของกลุ่มไอเท็มตามอัลกอริทึม 2.1 แสดงวิธีคำนวณค่าศูนย์กลางของกลุ่มไอเท็มที่ 1 ดังตาราง 3.4 และผลลัพธ์ค่าศูนย์กลางของแต่ละกลุ่มไอเทมแสดงทั้ง 3 ดังตาราง 3.5

ตารางที่ 3.4 แสดงการคำนวณหาค่าศูนย์กลางของกลุ่มไอเท็ม 1

Item	$(\mu_{c_1}(I))^2$	$(\mu_{c_1}(I))^2 \times (x_1)$	$(\mu_{c_1}(I))^2 \times (x_2)$	$(\mu_{c_1}(I))^2 \times (x_3)$	$(\mu_{c_1}(I))^2 \times (x_4)$	$(\mu_{c_1}(I))^2 \times (x_5)$
I_1	$(0.217549)^2$	$(0.217549)^2$ *3	$(0.217549)^2$ *2	$(0.217549)^2$ *5	$(0.217549)^2$ *0	$(0.217549)^2$ *2
I_2	$(0.115873)^2$	$(0.115873)^2$*3	$(0.115873)^2$*0	$(0.115873)^2$*4	$(0.115873)^2$*0	$(0.115873)^2$*1
I_3	$(0.114236)^2$	$(0.114236)^2$*1	$(0.114236)^2$*5	$(0.114236)^2$*2	$(0.114236)^2$*3	$(0.114236)^2$*5
I_4	$(0.600000)^2$	$(0.600000)^2$*2	$(0.600000)^2$*0	$(0.600000)^2$*4	$(0.600000)^2$*0	$(0.600000)^2$*2
I_5	$(0.307721)^2$	$(0.307721)^2$*1	$(0.307721)^2$*3	$(0.307721)^2$*5	$(0.307721)^2$*4	$(0.307721)^2$*5
I_6	$(0.237958)^2$	$(0.237958)^2$*4	$(0.237958)^2$*5	$(0.237958)^2$*2	$(0.237958)^2$*5	$(0.237958)^2$*3
I_7	$(0.909608)^2$	$(0.909608)^2$*2	$(0.909608)^2$*2	$(0.909608)^2$*4	$(0.909608)^2$*1	$(0.909608)^2$*5
I_8	$(0.260906)^2$	$(0.260906)^2$*5	$(0.260906)^2$*4	$(0.260906)^2$*5	$(0.260906)^2$*3	$(0.260906)^2$*2
I_9	$(0.234984)^2$	$(0.234984)^2$*3	$(0.234984)^2$*2	$(0.234984)^2$*3	$(0.234984)^2$*4	$(0.234984)^2$*1
I_{10}	$(0.722047)^2$	$(0.722047)^2$*3	$(0.722047)^2$*2	$(0.722047)^2$*5	$(0.722047)^2$*2	$(0.722047)^2$*4

ตารางที่ 3.5 แสดงค่าศูนย์กลางของแต่ละกลุ่มไอเท็ม

กลุ่มไอเท็ม	ค่าศูนย์กลาง				
C_1	2.411757	1.850767	4.260982	1.428964	3.864752
C_2	3.275499	1.918512	4.262166	1.547561	1.808264
C_3	1.885871	3.910692	2.724321	2.890021	3.949289

นำค่าศูนย์กลางของแต่ละกลุ่มไอเท็มไปทำการคำนวณหาค่าสมาชิกของแต่ละไอเท็ม ที่มีต่อแต่ละกลุ่มไอเท็มตามสมการ 3.2 นั่นคือคำนวณค่าความเป็นสมาชิกของไอเท็มตั้งแต่ไอเท็ม 1 ในกลุ่มไอเท็ม 1 กลุ่มไอเท็ม 2 และกลุ่มไอเทม 3 ไปจนถึงไอเท็ม 10 ในกลุ่มไอเท็ม 1 กลุ่มไอเท็ม 2 และกลุ่มไอเท็ม 3 ได้ผลลัพธ์แสดงดังตาราง 3.6

ตารางที่ 3.6 แสดงค่าความเป็นสมาชิกของไอเท็มต่อแต่ละกลุ่มไอเท็ม

รหัสไอเท็ม	กลุ่มไอเท็ม		
	1	2	3
I_1	0.075526	0.904029	0.024455
I_2	0.003035	0.995239	0.001726
I_3	0.012006	0.007343	0.980651
I_4	0.332192	0.567555	0.100253
I_5	0.4049421	0.17273	0.417849
I_6	0.226358	0.263045	0.510597
I_7	0.834398	0.077363	0.088239
I_8	0.14909	0.766161	0.084749
I_9	0.268633	0.500091	0.231276
I_{10}	0.910959	0.070502	0.018538

เมื่อคำนวณค่าความเป็นสมาชิกจนครบทุกไอเท็มในกลุ่มไอเท็มแล้ว จึงนำผลลัพธ์มาตรวจสอบว่าหยุดตรงตามเงื่อนไข $\|U^{k+1} - U^k\| < \varepsilon$ หรือยัง หากเงื่อนไขเป็นจริง การจัดกลุ่มไอเท็มด้วยโมดิฟายฟัซซีซีมีนถือว่าสิ้นสุด แต่หากยังไม่ตรงตามเงื่อนไขที่ตั้งเอาไว้ การทำงานก็จะวนกลับไปคำนวณหาค่าศูนย์กลางของกลุ่มไอเท็มอีกครั้งตามอัลกอริทึม 2.1 และจำกระทำซ้ำจนกว่าเงื่อนไขที่ใช้ตรวจสอบจะเป็นจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากผลลัพธ์ในตาราง 3.6 และกำหนดค่า ε เป็น 0.1 หลังตรวจสอบแล้วพบว่าเงื่อนไขการหยุดคำนวณยังเป็นเท็จ ดังนั้นจึงต้องวนกลับไปคำนวณหาค่าศูนย์กลางของแต่ละกลุ่มไอเท็มซ้ำและนำค่าที่คำนวณได้ไปอัพเดตเมตริกซ์ U เพื่อปรับค่าความเป็นสมาชิกของไอเท็มในแต่ละกลุ่มไอเท็มให้เป็นปัจจุบัน

เมื่อนำข้อมูลเรตติ้งจากชุดข้อมูลตัวอย่างและค่าความเป็นสมาชิกของไอเท็มตาราง 3.6 มาปรับค่าศูนย์กลางตามสมการของฟัซซีซีมีนแล้ว จะได้ผลลัพธ์เป็นค่าศูนย์กลางใหม่ของแต่ละกลุ่มไอเท็ม หลังจากนั้นให้นำค่าศูนย์กลางใหม่ที่คำนวณได้ไปใช้เทียบกับข้อมูลไอเท็มเพื่อคำนวณหาค่าความเป็นสมาชิกของแต่ละไอเท็มอีกครั้ง

ผลการคำนวณการปรับค่าศูนย์กลางและปรับค่าความเป็นสมาชิกของแต่ละไอเท็ม แสดงดังตาราง 3.7 และ 3.8 ตามลำดับ ซึ่งค่าความเป็นสมาชิกใหม่นี้ทำให้เงื่อนไขการหยุดคำนวณเป็นจริง

ตารางที่ 3.7 แสดงค่าศูนย์กลางของแต่ละกลุ่มไอเท็มเมื่อเงื่อนไขการหยุดคำนวณเป็นจริง

กลุ่มไอเท็ม	ค่าศูนย์กลาง				
C_1	2.376105	2.114752	4.434904	1.897466	4.239763
C_2	3.250825	1.550966	4.356064	0.968128	1.662523
C_3	1.744357	4.676744	2.307017	3.302679	4.377989

ตารางที่ 3.8 แสดงค่าความเป็นสมาชิกของไอเท็มต่อแต่ละกลุ่มไอเท็ม เทียบจากค่าศูนย์กลางสุดท้าย

รหัสไอเท็ม	กลุ่มไอเท็ม		
	1	2	3
I_1	0.075452	0.902954	0.021594
I_2	0.002353	0.996272	0.001374
I_3	0.003936	0.001917	0.994808
I_4	0.303936	0.59987	0.096195
I_5	0.491146	0.169854	0.339
I_6	0.202813	0.235699	0.561487
I_7	0.86568	0.073016	0.061304
I_8	0.153938	0.753323	0.09274
I_9	0.277529	0.472434	0.250037
I_{10}	0.866079	0.111304	0.022617

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากค่าความเป็นสมาชิกของไอเท็มต่อแต่ละกลุ่มไอเท็มในตาราง 3.8 ทำให้ได้กลุ่มไอเท็มและไอเท็มสมาชิกแสดงดังตาราง 3.9

ตารางที่ 3.9 แสดงกลุ่มไอเท็มและไอเท็มสมาชิก

กลุ่มไอเท็ม	ไอเท็มสมาชิก
C_1	I_5, I_7, I_{10}
C_2	I_1, I_2, I_4, I_8, I_9
C_3	I_3, I_6

3.2.2 คำนวณหากลุ่มที่เหมาะสมสำหรับแอคทีฟยูสเซอร์

สำหรับขั้นตอนการแนะนำสินค้านั้น จะเริ่มต้นด้วยการหากลุ่มไอเทมที่เหมาะสมแอคทีฟยูสเซอร์ ด้วยการเทียบจากข้อมูลการให้เรตติ้งของแอคทีฟยูสเซอร์ต่อไอเท็มในอดีต ดูว่าไอเท็มที่ผู้ใช้เคยให้ค่าเรตติ้งนั้นอยู่กลุ่มข้อมูลใดบ้าง และหาค่าเฉลี่ยเรตติ้งของไอเท็มที่แอคทีฟยูสเซอร์เคยให้เรตติ้งไว้ในแต่ละกลุ่ม เพื่อระบุกลุ่มที่เหมาะสมสำหรับการแนะนำสำหรับแอคทีฟยูสเซอร์โดยเลือกจากกลุ่มที่มีค่าเรตติ้งเฉลี่ยสูงสุด

เช่น ต้องการแนะนำไอเท็มแก่แอคทีฟยูสเซอร์รหัส U_7 ซึ่งเคยให้ค่าเรตติ้งแก่ไอเท็ม I_1, I_2 และ I_4เป็นค่าเรตติ้ง 3, 1 และ 5 ตามลำดับ แสดงข้อมูลการให้ค่าเรตติ้งดังกล่าวในตาราง 3.10

ตารางที่ 3.10 แสดงการให้ค่าเรตติ้งของ U_7

รหัสผู้ใช้	รหัสไอเท็ม									
	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
7	3	1	0	5	0	0	0	0	0	0

ค่าเฉลี่ยเรตติ้งของ U_7 ที่มีต่อไอเท็มในกลุ่ม C_2 คือ 3+1+5 / 2 = 5.5 ดังนั้นกลุ่มไอเท็มที่เหมาะสมสำหรับแนะนำแก่ U_7 คือกลุ่ม C_2

3.2.3 คำนวณหาไอเท็มภายในกลุ่มสำหรับสร้างรายการแนะนำ

ดำเนินการสร้างไอเท็ม-ไอเท็มเมตริกซ์ จากไอเท็มที่อยู่เฉพาะในกลุ่มที่เหมาะสมสำหรับการแนะนำ ในกรณีตัวอย่างคือกลุ่ม C_2 ดังตาราง 3.11 ดังนั้นให้นำข้อมูลไอเท็มจากตาราง 3.11 ไปทำการเปรียบเทียบระหว่างกันด้วยสมการสำหรับหาค่าความคล้ายคลึง และนำผลลัพธ์ไปสร้างเมตริกซ์ไอเท็ม-ไอเท็ม แสดงดังตาราง 3.12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.11 แสดงชุดข้อมูลเรตติ้งสินค้าของไอเท็มภายในกลุ่ม C_2

รหัสไอเท็ม	รหัสผู้ใช้				
	U_1	U_2	U_3	U_4	U_5
I_1	3	2	5	0	2
I_2	3	0	4	0	1
I_4	2	0	4	0	2
I_8	5	4	5	3	2
I_9	3	2	3	4	1

ตารางที่ 3.12 แสดงไอเท็ม-ไอเท็มเมตริกซ์ของกลุ่ม C_2

รหัสไอเท็ม	I_1	I_2	I_4	I_8	I_9
I_1	1	0.928571	0.944911	0.666666	0.730296
I_2	0.928571	1	0.755928	0.944911	0.327326
I_4	0.944911	0.755928	1	0.500000	0.866025
I_8	0.666666	0.944911	0.500000	1	0.176470
I_9	0.730296	0.327326	0.866025	0.176470	1

จากนั้นทำการสร้างเมตริกซ์ขนาด 1* m ซึ่งเป็นเมตริกซ์ที่แสดงถึงความสัมพันธ์ระหว่างแอคทีฟยูสเซอร์และค่าเรตติ้งของไอเท็มภายในกลุ่มไอเท็มที่เหมาะสมแก่การแนะนำ แสดงดังตาราง 3.13

นำทั้งสองเมตริกซ์การคูณกันตามขั้นตอนของวิธีการแนะนำสินค้าด้วยไอเท็มเบสท็อปเอ็นเพื่อหาเมตริกซ์ผลลัพธ์ที่จะนำไปสร้างรายการแนะนำ แสดงดังตาราง 3.13

ตาราง 3.13 แสดง 1* m เมตริกซ์ของผู้ใช้รหัส U_7 กับไอเท็มในกลุ่ม C_2

รหัสไอเท็ม	U_7
I_1	3
I_2	1
I_4	5
I_8	0
I_9	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตาราง 3.14 แสดงการคูณเมตริกซ์เพื่อหาเมตริกซ์ผลลัพธ์

$$\begin{bmatrix} & I_1 & I_2 & I_4 & I_8 & I_9 \\ I_1 & 1 & 0.928571 & 0.944911 & 0.666666 & 0.730296 \\ I_2 & 0.928571 & 1 & 0.755928 & 0.944911 & 0.327326 \\ I_4 & 0.944911 & 0.75928 & 1 & 0.500000 & 0.866025 \\ I_8 & 0.666666 & 0.944911 & 0.50000 & 1 & 0.176470 \\ I_9 & 0.730296 & 0.327326 & 0.866025 & 0.176470 & 1 \end{bmatrix} \times \begin{bmatrix} U_7 \\ 3 \\ 1 \\ 5 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} & \\ 8.623126 & I_1 \\ 7.655353 & I_2 \\ 8.594013 & I_4 \\ 5.444909 & I_8 \\ 6.848333 & I_9 \end{bmatrix}$$

3.2.4 สร้างรายการไอเท็มแนะนำ

นำผลลัพธ์จากจากหัวข้อ 3.2.3 มาสร้างรายการแนะนำหรือ Top-*N* ตามจำนวนไอเท็มที่ต้องการแนะนำ ด้วยการเรียงลำดับผลลัพธ์จากมากไปน้อย และตัดไอเท็มตัวที่ผู้ใช้เคยให้ค่าคะแนนและเป็นศูนย์ออกจากรายการแนะนำ ดังนั้นไอเท็มที่อยู่รายการแนะนำได้แก่ I_9, I_8

3.2.5 แนะนำรายการไอเท็มแก่แอคทีฟยูสเซอร์

การแนะนำถือเป็นขั้นตอนสุดท้ายของการทำงาน กล่าวคือนำรายการแนะนำที่สร้างเอาไว้สำหรับแอคทีฟยูสเซอร์จากหัวข้อ 3.2.4 ไปทำการแนะนำให้แก่แอคทีฟยูสเซอร์ ตามจำนวนรายการแนะนำที่กำหนด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การทดลองและผลการทดลอง

ในบทนี้จะอธิบายถึงแผนการทดลอง ชุดข้อมูลที่ใช้ในการทดลอง การทดลอง และผลการทดลอง สำหรับโปรแกรมประยุกต์ที่นำมาใช้ทดลองแนวคิดของวิจัยนั้นพัฒนาขึ้นจากภาษาไพธอน (Python)

4.1 ชุดข้อมูลที่ใช้ในการทดลอง

เนื่องจากงานวิจัยการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟ์ซซีซีมีน ใช้ค่าเรตติ้งเป็นข้อมูลสำหรับจัดกลุ่มตามไอเท็ม ผู้วิจัยจึงนำชุดข้อมูลมูฟวี่เลนส์มาใช้ในการทดลอง

4.1.1 ชุดข้อมูลมูฟวี่เลนส์

ชุดข้อมูลมูฟวี่เลนส์เป็นชุดข้อมูลเรตติ้งที่ถูกรวบรวมโดยกลุ่มนักวิจัย ในโครงการวิจัยกรุ๊ปเลนส์แห่งมหาวิทยาลัยมินเนโซต้า ข้อมูลเรตติ้งประกอบไปด้วย 100,000 เรตติ้งจากผู้ใช้ 943 คนที่มีต่อไอเท็มภาพยนตร์จำนวน 1,682 ไอเท็ม ทุกทรานแซคชั่นจะมีการเก็บเวลา (Timestamp) ที่ทำรายการแนบท้าย โดยผู้ใช้แต่ละรายจะให้ค่าคะแนนเรตติ้งต่อภาพยนตร์ไม่ต่ำกว่า 20 ไอเท็ม แสดงตัวอย่างชุดข้อมูลดังตาราง 4.1

ตารางที่ 4.1 แสดงตัวอย่างชุดข้อมูลมูฟวี่เลนส์

รหัสผู้ใช้ (user id)	รหัสไอเท็ม (item id)	ค่าคะแนนเรตติ้ง(rating)	เวลา (timestamp)
196	242	3	881250949
186	302	3	891717742
22	377	1	878887116
244	51	2	880606923
166	346	1	886397596
298	474	4	884182806
115	265	2	881171488
253	465	5	891628467
305	451	3	886324817

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.2 การเตรียมข้อมูล

เนื่องด้วยการทดลองของงานวิจัยนี้ใช้เพียงรหัสผู้ใช้ รหัสไอเท็มและค่าคะแนนเรตติ้งที่ผู้ใช้มีต่อไอเท็ม แต่ในตัวอย่างข้อมูลตาราง 4.1 จะพบว่ามีข้อมูลเวลาแนบท้ายมาในแต่ละทรานแซคชั่นข้อมูล จึงต้องทำการขจัดข้อมูลส่วนเกินออกจากชุดข้อมูลเพื่อนำไปจัดรูปแบบข้อมูลสำหรับใช้ในการทดลอง โดยคงเหลือรหัสผู้ใช้ รหัสไอเท็มภาพยนตร์และค่าคะแนนเรตติ้ง

ชุดข้อมูลในการทดลองจะถูกแบ่งออกเป็น 2 แบบ คือ

1. ชุดข้อมูลที่ประกอบไปด้วยรหัสผู้ใช้ รหัสไอเท็มและค่าคะแนนเรตติ้งไว้สำหรับจัดกลุ่มข้อมูลไอเท็ม แสดงตัวอย่างดังตาราง 4.2

2. ชุดข้อมูลไอเท็มซึ่งประกอบไปด้วยรหัสผู้ใช้ รหัสไอเท็มโดยที่ไม่สนใจค่าคะแนนเรตติ้ง ข้อมูลในชุดข้อมูลจะมีเพียงแค่ 0 กับ 1 หากผู้ใช้ให้คะแนนเรตติ้งแก่สินค้าถือว่าได้เลือกไอเท็มนั้น กำหนดค่าเป็น 1 และเป็น 0 หากไม่มีค่าคะแนนเรตติ้ง แสดงตัวอย่างดังตาราง 4.3

ตารางที่ 4.2 แสดงตัวอย่างข้อมูลชุดที่ 1

รหัสไอเท็ม	รหัสผู้ใช้							
	U_1	U_2	U_3	U_4	U_4	U_6	U_7	U_8
I_1	5	0	0	0	4	0	5	5
I_2	0	0	0	0	0	0	0	5
I_3	0	0	0	0	0	0	0	0
I_4	0	0	0	0	0	0	0	0
I_5	0	0	0	0	0	0	0	0

ตารางที่ 4.3 แสดงตัวอย่างข้อมูลชุดที่ 2

รหัสไอเท็ม	รหัสผู้ใช้							
	U_1	U_2	U_3	U_4	U_4	U_6	U_7	U_8
I_1	1	0	0	0	1	0	1	1
I_2	0	0	0	0	0	0	0	1
I_3	0	0	0	0	0	0	0	0
I_4	0	0	0	0	0	0	0	0
I_5	0	0	0	0	0	0	0	0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นำชุดข้อมูลทั้ง 2 มาแบ่งเป็น 5 ชุดย่อยด้วยวิธีสุ่มเลือก แต่ละชุดย่อยจะมีจำนวนข้อมูลเท่าๆ กัน เพื่อใช้เป็นข้อมูลเทรนนิ่ง (Training) สำหรับประมวลผลของอัลกอริทึมโดยนำผลลัพธ์ไปตรวจสอบความถูกต้องของการทำงานของอัลกอริทึมในการแนะนำไอเท็มกับข้อมูลเทส (Test) ตามหลักการเคโฟล์ครอสวาลิเดชั่น (กำหนด K = 5)

4.2 การประเมินประสิทธิภาพของอัลกอริทึม

ด้วยวัตถุประสงค์ของงานวิจัย ที่มุ่งจะลดเวลาที่ใช้ในแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น ผู้วิจัยจึงนำอัลกอริทึมฟัซซีซีมีนเข้ามาใช้ โดยปรับเปลี่ยนสมการสำหรับหาค่าความเป็นสมาชิกด้วยประยุกต์การหาค่าความคล้ายคลึงด้วยค่าสหสัมพันธ์ของข้อมูลที่เป็นโคเรตกัน เข้ามาใช้แทนการหาค่าระยะห่างระหว่างไอเท็ม เพื่อให้ได้ผลการจัดกลุ่มสำหรับแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นที่ถูกต้องยิ่งขึ้น

ดังนั้นจึงแบ่งการประเมินผลการทดลองของงานวิจัยนี้ออกเป็น 2 ด้าน คือ เวลาและความถูกต้อง โดยทำการเปรียบเทียบในเรื่องของเวลาระหว่างการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นกับการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน และเปรียบเทียบเรื่องของความถูกต้องด้วยการคำนวณค่าฮิตเรต ระหว่างการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยการจัดกลุ่มข้อมูลไอเท็มด้วยโมดิฟายฟัซซีซีมีน กับการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยการจัดกลุ่มข้อมูลไอเท็มด้วยฟัซซีซีมีน

ในส่วนของการวัดเวลาที่ใช้นั้น เริ่มต้นนับตรงขั้นตอนสร้าง $1 * m$ เมตริกซ์ไปจนกระทั่งสิ้นสุดการแนะนำให้แก่แอคทีฟยูสเซอร์ สำหรับค่าฮิตเรตคำนวณจากสมการ (4.1)

$$\text{Hit-Rate (HR)} = \text{Number Of Hits} / n \qquad (4.1)$$

เมื่อ Number Of Hits หมายถึง จำนวนครั้งในการแนะนำถูกต้อง

และ n หมายถึง จำนวนของแอคทีฟยูสเซอร์ทั้งหมดที่ได้รับการแนะนำ

4.3 การทดลองการแนะนำสินค้าแบบไอเท็มเบสและผลการทดลอง

ด้วยวิธีการเคโฟล์ครอสวาลิเดชั่น ชุดข้อมูลที่ถูกแบ่งออกเป็นจำนวน 5 ชุด จะถูกนำมาทำการคำนวณทั้งหมด 5 รอบ แต่ละรอบการคำนวณก็จะใช้ข้อมูลชุดหนึ่งมาเป็นชุดข้อมูลทดสอบ และอีก 4 ชุดข้อมูลที่เหลือจะถูกใช้เป็นข้อมูลสำหรับเทรนนิ่ง ดังแผนภาพประกอบด้านล่าง

รอบการคำนวณที่ 1 :

ชุดข้อมูลเทรนนิ่ง :: ชุดข้อมูลที่ 2, 3, 4 และ 5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดข้อมูลทดสอบ :: ชุดข้อมูลที่ 1

รอบการคำนวณที่ 2 :

ชุดข้อมูลเทรนนิ่ง :: ชุดข้อมูลที่ 1, 3, 4 และ 5

ชุดข้อมูลทดสอบ :: ชุดข้อมูลที่ 2

รอบการคำนวณที่ 3 :

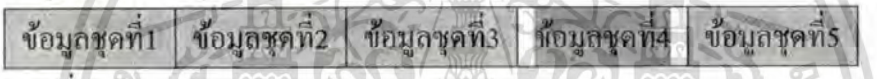
ชุดข้อมูลเทรนนิ่ง :: ชุดข้อมูลที่ 1, 2, 4 และ 5

ชุดข้อมูลทดสอบ :: ชุดข้อมูลที่ 3

รอบการคำนวณที่ 4 :

ชุดข้อมูลเทรนนิ่ง :: ชุดข้อมูลที่ 1, 3, 3 และ 5

ชุดข้อมูลทดสอบ :: ชุดข้อมูลที่ 4



รอบการคำนวณที่ 5 :

ชุดข้อมูลเทรนนิ่ง :: ชุดข้อมูลที่ 1, 2, 3 และ 4

ชุดข้อมูลทดสอบ :: ชุดข้อมูลที่ 5

ทั้งนี้ได้ออกแบบการทดลองแบ่งตามลักษณะวิธีการแนะนำไอเท็ม เพื่อเปรียบเทียบและวัดประสิทธิภาพของอัลกอริทึมตามหัวข้อ 4.2 จำแนกเป็น 2 ลักษณะดังนี้

- การแนะนำแบบไอเท็มเบสท็อปเอ็น
- การแนะนำแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟังชั่นซีซีมีนและฟังชั่นซีซีมีน

4.3.1 การแนะนำแบบไอเท็มเบสท็อปเอ็น

ในแต่ละรอบการคำนวณจะนำข้อมูลสำหรับเทรนนิ่งมาสร้างเมตริกซ์ไอเท็ม-ไอเท็ม ด้วยการเปรียบเทียบค่าความคล้ายคลึงแบบโคซายน์ระหว่างไอเท็มในส่วนข้อมูลเทรนนิ่ง ผลลัพธ์ที่ได้คือความสัมพันธ์ระหว่างไอเท็มไอเท็มทั้งหมดในชุดข้อมูลเทรนนิ่ง และสร้างเมตริกซ์ 1* m ของแอคทีฟยูสเซอร์ หลังจากทำการคูณเมตริกซ์ระหว่างไอเท็ม-ไอเท็มเมตริกซ์และ 1* m เมตริกซ์แล้วจึงนำค่าในเมตริกซ์ผลลัพธ์ที่ได้ไปสร้างรายการแนะนำ เพื่อแนะนำแก่แอคทีฟยูสเซอร์ทั้งหมด 10 รายการ (Top-N = 10) เรียงตามค่าไอเท็มในตารางผลลัพธ์จากมากไปน้อย และตัดไอเท็มที่แอคทีฟยูส-เซอร์เคยให้ค่าเรตติ้งไว้แล้วกับไอเท็มที่เป็นศูนย์ออกไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.4 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 1

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.432000	608,296,1265,593,318,1617,50,1213,1094,300
U8	0.426000	608,296,318,1580,593,377,1265,1393,589,1573
U17	0.434000	589,1196,1240,260,1200,1580,32,608,1214,1210
U26	0.428000	377,736,1393,380,356,1265,587,1517,339,1580
U35	0.435000	1247,919,912,608,1265,1136,1307,1225,1198,1196
U56	0.432000	608,318,296,593,1265,527,50,457,1617,1213
U60	0.429000	780,380,736,1552,1580,1370,377,733,1573,1608
U61	0.369000	762,786,208,95,89,466,832,494,996,165
U63	0.431000	1200,1196,32,608,1214,589,296,1240,1580,541
U69	0.432000	608,296,1094,300,1358,593,318,1095,337,1183
U73	0.431000	1580,780,589,1527,1573,1240,1370,1291,316,380
U75	0.429000	364,588,586,367,1,2,595,153,1073,780
U76	0.431000	608,296,318,593,1213,50,1617,1358,1265,1089
U79	0.425000	1617,800,608,648,1270,1597,589,296,593,457
U80	0.431000	1265,589,1580,356,480,593,608,1196,110,1210
U82	0.430000	589,1580,377,457,780,1610,1573,608,593,480
U88	0.437000	608,1247,1225,858,318,1198,1196,919,527,296
U89	0.434000	527,1193,733,608,318,1198,1196,593,296,260
U93	0.432000	780,1580,589,1527,1240,1291,1370,1573,316,1552
U94	0.428000	1258,551,1261,1321,1130,1219,1345,1387,1214,908
U102	0.440000	1247,919,1207,923,912,608,1196,924,750,1214
U107	0.437000	919,1247,912,608,1196,1198,858,260,923,924
U111	0.429000	1265,608,318,356,1393,296,593,357,1517,1580
U118	0.433000	1580,1196,589,1198,1240,1270,1291,1210,1265,260
U123	0.434000	608,1196,318,1198,1270,1240,296,1265,1036,593
U126	0.440000	1196,260,1265,1210,608,1307,858,1198,1247,1270
U127	0.434000	589,1580,1196,1240,260,480,1198,780,1210,457
U139	0.436000	1196,260,1198,1240,1200,1270,1214,589,1210,1097

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 2

ไอเท็มเป้าหมาย :: 34, 39, 157, 318, 592, 670, 1131, 1200, 1407, 1632		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U2	0.430000	1196,589,1580,480,1198,318,1210,592,1610,457
U10	0.430000	1196,1270,1097,1210,260,592,1580,1198,1200,919
U13	0.436000	1580,1196,589,480,1198,260,780,1210,1291,1240
U14	0.407000	296,1225,1263,608,318,1213,527,1208,593,1090
U16	0.387000	1269,266,1294,1234,1136,1259,357,951,1035,539
U22	0.432000	1580,589,1196,592,1200,380,1036,1240,780,1270
U27	0.439000	858,912,1247,1196,919,260,1136,1198,608,1252
U46	0.414000	1407,1347,1644,366,253,1327,891,1339,799,1261
U47	0.431000	608,318,296,527,593,1213,858,1247,50,1193
U49	0.434000	589,1580,1196,480,457,1200,1240,1210,1036,260
U58	0.433000	1196,1200,260,1580,592,589,1210,1198,1240,1270
U59	0.437000	1247,608,912,919,1196,318,1207,296,924,858
U66	0.429000	1196,480,110,356,1580,589,1259,1210,260,1270
U67	0.436000	1247,608,318,1207,1208,296,527,593,858,923
U72	0.452000	356,1265,597,593,1210,587,608,1580,318,1270
U77	0.342000	536,1131,1132,75,304,462,1170,1494,800,241
U78	0.433000	1265,357,1,588,39,919,1307,34,1517,356
U81	0.437000	1196,858,260,924,608,1198,296,1247,1208,1617
U90	0.433000	1580,356,1265,380,318,608,1517,589,592,377
U91	0.435000	1307,1265,1197,1136,1079,1270,1196,912,1394,1198
U97	0.439000	912,1247,919,858,608,1252,904,1196,1617,1208
U98	0.432000	318,527,593,296,608,1196,1210,1580,648,480
U105	0.434000	1196,1240,260,1200,1198,1036,1210,1291,589,457
U108	0.434000	480,1580,648,589,1196,1270,260,1544,593,1259
U114	0.439000	364,588,919,595,1028,1029,1073,1022,594,1282
U128	0.425000	480,1210,1580,1196,1282,589,260,648,1249,1250
U129	0.434000	1036,1580,1240,589,1196,1200,592,1198,1610,260
U132	0.432000	1265,1517,1,1270,1197,1136,1580,1073,1307,1220

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.6 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 3

ไอเท็มเป้าหมาย :: 3, 7, 88, 413, 772, 963, 1198, 1220, 1580, 1610		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U3	0.436000	1196,1198,1580,1291,1270,1197,260,480,1210,589
U9	0.435000	608,593,318,296,1265,50,527,1580,1213,1617
U15	0.435000	1580,589,356,480,457,1370,780,1610,380,377
U24	0.441000	608,1265,318,593,1196,1270,1198,296,919,1307
U29	0.440000	1196,1240,1198,1200,589,260,1036,1580,1291,457
U32	0.440000	608,593,296,589,318,50,1617,480,32,457
U37	0.438000	356,597,1265,587,1307,357,339,1569,708,539
U40	0.442000	1196,1198,260,592,1291,1210,1580,919,1097,780
U41	0.438000	1196,260,1210,1097,1580,589,480,1200,1198,1240
U42	0.442000	1580,589,1240,1196,1200,780,260,480,1210,1270
U50	0.437000	480,1544,648,1580,266,589,524,1466,1259,380
U51	0.436000	1265,1,34,39,588,1517,318,356,1580,608
U52	0.436000	380,733,1552,1370,1610,780,377,1608,165,1573
U62	0.442000	608,919,1247,1196,1265,318,593,1097,1270,296
U65	0.440000	1580,592,1196,589,480,1291,1198,780,1270,377
U96	0.443000	608,318,593,1196,589,296,1265,527,1270,260
U100	0.439000	1196,1580,1198,589,260,1210,480,1240,1291,1200
U113	0.438000	780,1580,1527,589,480,1573,316,1544,1676,32
U115	0.433000	1544,480,70,648,1580,589,780,1527,1407,1129
U120	0.444000	608,318,593,1196,260,1198,296,858,527,919
U124	0.444000	608,593,318,527,1213,1358,296,34,1094,457
U130	0.440000	589,1580,1240,1196,1200,1036,592,480,1198,1291
U136	0.444000	608,1265,919,1097,1196,1270,1259,296,356,593
U142	0.439000	1196,260,593,296,318,608,1097,1198,589,919
U144	0.438000	1265,356,1580,593,1396,527,780,1270,589,924
U149	0.442000	608,593,1580,1196,1265,589,296,1200,318,592
U155	0.441000	1265,1580,356,587,367,1196,593,608,1270,1517
U157	0.442000	1196,1580,608,1265,589,1270,1200,260,1198,1240

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 4

ไอเท็มเป้าหมาย :: 1, 11, 110, 318, 339, 457, 608, 708, 1197, 1513		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U1	0.436000	919,588,364,595,1097,1028,1,260,1196,1270
U4	0.440000	260,1196,1240,1198,1210,1214,1036,1200,1387,1291
U11	0.439000	608,1265,318,356,593,296,1270,1393,1580,1
U12	0.446000	858,1198,919,1221,608,260,593,1196,1247,923
U18	0.441000	1196,1580,260,1210,1270,589,592,1097,480,780
U20	0.438000	589,1580,1240,1527,457,480,110,780,648,1200
U21	0.438000	1210,1,588,1196,1270,1580,260,364,589,780
U23	0.445000	1580,589,1196,260,1240,1200,480,1270,1210,780
U25	0.440000	1580,780,589,1196,480,1356,1210,260,1270,1374
U28	0.447000	1196,260,608,1198,318,919,1270,1265,593,1097
U31	0.446000	1270,1265,1307,1197,1136,260,1234,1196,1394,356
U36	0.443000	1580,1270,589,1196,480,1240,780,1210,457,377
U45	0.441000	339,357,1265,440,39,1307,539,597,708,11
U48	0.441000	1580,1270,380,780,1196,356,377,589,457,1240
U54	0.449000	1234,1136,1394,1079,1278,919,1270,1230,260,1294
U64	0.441000	1210,1291,1198,1196,589,260,1580,110,480,1240
U70	0.441000	1127,223,260,1196,1580,480,1270,589,1210,1513
U74	0.444000	608,1617,296,318,50,593,858,1213,1252,527
U83	0.442000	260,1196,1097,1270,1210,1580,589,1240,480,1198
U85	0.432000	648,480,589,47,380,1101,1544,1580,457,377
U86	0.440000	260,1580,480,1196,1210,364,1270,589,588,919
U87	0.444000	589,1196,1240,260,1198,457,1291,480,1580,1036
U95	0.440000	1580,608,318,593,589,1265,457,480,296,110
U103	0.439000	780,380,316,1370,1573,1552,733,1580,1527,208
U109	0.447000	260,1196,1270,1198,1097,919,1210,1580,1197,480
U116	0.441000	1240,260,1196,1200,1097,1214,1036,1270,589,1210
U117	0.441000	780,1580,589,1240,1200,1196,1270,316,260,1210
U119	0.437000	588,364,595,1028,1,1022,1029,586,2,1032

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.8 แสดงผลการทดลองบางส่วน การแนะนำข้อมูลแบบไอเท็มเบสท็อปเอ็นรอบที่ 5

ไอเท็มเป้าหมาย :: 18, 33, 223, 1073, 1196, 1265, 1270, 1319, 1455, 1501		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U6	0.441000	588,597,364,1265,595,1380,1,1028,919,356
U7	0.446000	457,1580,589,480,1610,733,377,380,1573,110
U19	0.447000	1196,1580,260,589,1198,1270,1200,1210,1240,480
U30	0.448000	608,318,593,296,527,1617,1213,50,1265,1196
U33	0.451000	1196,608,1198,260,1270,919,1097,1200,318,296
U34	0.448000	1265,356,1307,1393,1270,1197,39,357,587,1
U38	0.449000	357,539,1393,339,1265,587,356,597,39,440
U39	0.442000	1617,318,223,296,608,1193,593,50,1196,527
U43	0.426000	1307,1197,356,1265,1270,1580,1393,1196,1210,597
U44	0.448000	1196,260,1270,1580,1198,1210,1197,1097,480,589
U53	0.448000	608,1580,296,1196,1200,1265,1573,32,457,1270
U55	0.457000	318,593,527,608,296,457,589,110,1213,356
U57	0.450000	1196,1198,1200,480,260,1580,1259,589,608,1090
U68	0.446000	1265,1307,1197,1136,1270,1394,1073,1,1517,1278
U71	0.444000	480,1198,1580,648,589,260,1196,457,110,1210
U84	0.450000	1196,260,1200,1198,1240,1210,1291,589,1214,1270
U92	0.454000	1580,588,780,1270,367,364,380,1196,1197,480
U99	0.448000	1196,1265,296,318,608,1198,1270,260,1580,457
U101	0.448000	780,1580,589,316,380,1573,1527,1240,480,736
U104	0.432000	858,223,785,1221,1193,608,296,1196,260,1198
U106	0.448000	1265,357,1307,1247,899,920,912,914,915,1094
U110	0.450000	296,608,50,1089,1617,47,1580,593,1213,480
U112	0.452000	919,1282,1617,588,1,1252,1196,594,541,260
U122	0.446000	1291,1196,1198,1200,1240,1580,260,457,1210,1036
U125	0.448000	457,1240,1291,480,1036,1196,589,1200,1580,1198
U135	0.442000	1210,1197,1127,1196,260,223,1270,1580,1200,1240
U150	0.455000	608,1265,318,1196,296,1270,593,1198,1307,527
U151	0.446000	1265,1580,356,1270,608,1196,296,357,1197,589

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากตารางที่ 4.4 – 4.8 ไอเท็มเป้าหมาย คือ ไอเท็มที่ถูกซ่อนค่าเรตติ้งไว้เพื่อตรวจสอบความถูกต้อง เวลาที่ใช้เริ่มตั้งแต่การหาเมตริกซ์ 1 * m ไปจนถึงเสร็จสิ้นการแนะนำ หน่วยเวลาเป็นวินาที ในส่วนของรหัสไอเท็มนั้น คือรหัสของไอเท็มที่แนะนำให้กับแอคทีฟยูสเซอร์ แสดงผลตามการเรียงค่าในเมตริกซ์ผลลัพธ์จากมากไปสู่น้อย แล้วจึงนำผลการทดลองในส่วนของเวลาและค่าฮิตเรตในแต่ละรอบการคำนวณมาหาค่าเฉลี่ย ได้ผลลัพธ์แสดงดังตาราง 4.9 และ 4.10 ตามลำดับ

ตารางที่ 4.9 แสดงเวลาการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น

เวลา (s) รอบที่1	เวลา (s) รอบที่ 2	เวลา (s) รอบที่ 3	เวลา (s) รอบที่ 4	เวลา (s) รอบที่ 5	เวลา (s) เฉลี่ย
78.859000	62.682000	81.213001	79.894000	81.104000	76.750400

ตารางที่ 4.10 แสดงค่าฮิตเรตการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น

Hit-Rateรอบที่1	Hit-Rateรอบที่ 2	Hit-Rateรอบที่ 3	Hit-Rateรอบที่ 4	Hit-Rateรอบที่ 5	Hit-Rate เฉลี่ย
5.093474	5.266314	5.470899	5.372340	5.157801	5.272166

จากตาราง 4.9 แสดงให้เห็นว่าค่าความถูกต้องในการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นจากการแนะนำสินค้า 10 อันดับรายการนั้นมีค่าเฉลี่ยอยู่ที่ 5.272166 และจากตาราง 4.10 แสดงให้เห็นว่าใช้เวลารวมในการแนะนำสินค้าแก่ผู้ใช้เฉลี่ย 76.750400 วินาที

4.3.2 การแนะนำแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนและฟัซซีซีมีน

สำหรับการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนนั้น จะประเมินประสิทธิภาพในด้านความถูกต้องเทียบกับการใช้ฟัซซีซีมีนอัลกอริทึม ดังนั้นแต่ละรอบการคำนวณจึงต้องทำการจัดกลุ่มขนาดต่างๆ ตั้งแต่ 2, 3, 4 และ 5 ด้วยอัลกอริทึมโมดิฟายฟัซซีซีมีนและฟัซซีซี-มีน กำหนดให้ค่า $\varepsilon = 0.1$

จากนั้นจึงสร้างเมตริกซ์ไอเท็ม-ไอเท็มของแต่ละกลุ่มข้อมูลไอเท็ม ซึ่งเป็นเมตริกซ์แสดงถึงค่าความสัมพันธ์ระหว่างไอเท็มในกลุ่มข้อมูลเดียวกัน ใช้การเปรียบเทียบค่าความคล้ายคลึงระหว่างไอเท็มจากข้อมูลเรตติ้งของไอเท็มโดยการหาค่าสหสัมพันธ์

ดำเนินการเลือกกลุ่มที่เหมาะสมแก่แอคทีฟยูสเซอร์ ซึ่งคำนวณมาจากค่าเฉลี่ยของไอเท็มที่แอคทีฟยูสเซอร์เคยให้ค่าเรตติ้งไว้ในแต่ละกลุ่มไอเท็มโดยเลือกกลุ่มที่มีค่าเฉลี่ยสูงที่สุด หลังจากได้กลุ่มที่เหมาะสม ก็นำข้อมูลในอดีตของผู้ใช้มาทำการสร้างเมตริกซ์ 1* m ของแอคทีฟยูสเซอร์ ซึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เมตริกซ์ 1* m นี้แสดงถึงความสัมพันธ์ระหว่างแอคทีฟยูสเซอร์กับไอเท็มในกลุ่มสำหรับแนะนำที่เคยให้เรตติ้งเอาไว้

หลังจากที่คูณเมตริกซ์ไอเท็ม-ไอเท็มเมตริกซ์ของกลุ่มไอเท็มที่เหมาะสมสำหรับแนะนำ กับ 1 * m เมตริกซ์แล้ว นำค่าในเมตริกซ์ผลลัพธ์ที่ได้ไปสร้างรายการแนะนำ สำหรับแนะนำแก่แอคทีฟยูสเซอร์เป็นจำนวน 10 รายการ (Top-N = 10) เรียงตามค่าไอเท็มในตารางผลลัพธ์จากมากไปน้อย และตัดไอเท็มที่แอคทีฟยูสเซอร์เคยให้ค่าเรตติ้งไว้แล้วกับไอเท็มที่เป็นศูนย์ออกไป

แสดงผลการทดลองบางส่วนของการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนในการจัดกลุ่มขนาดตั้งแต่ 2-5 กลุ่ม ดังตารางที่ 4.11 – 4.14 และผลการทดลองบางส่วนของการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นแบบฟัซซีซีมีน ดังตารางที่ 4.15 – 4.18 ซึ่งในรายงานฉบับนี้ขอแสดงผลการทดลองบางส่วนในรอบแรก (K = 1) ของแต่ละอัลกอริทึมการจัดกลุ่ม โดยที่ไอเท็มเป้าหมายของการจัดกลุ่มด้วยโมดิฟายฟัซซีซีมีนและฟัซซีซีมีน จะเหมือนกับไอเท็มเป้าหมายในการทดลองการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น

สำหรับช่องตารางที่แสดงผลรหัสไอเท็ม คือไอเท็มที่อยู่ในรายการแนะนำจำนวน 10 อันดับ โดยเรียงลำดับจากค่าในเมตริกซ์ผลลัพธ์มากไปน้อย กลุ่มที่อยู่ส่วนท้ายหลังรหัสไอเท็ม คือกลุ่มไอเท็มที่เหมาะสมสำหรับการแนะนำแก่แอคทีฟยูสเซอร์

จากตาราง 4.11 – 4.19 หากแถวใดว่างเปล่า หมายถึง ไม่มีไอเท็มใดๆ ที่เหมาะสมแก่การแนะนำให้กับแอคทีฟยูสเซอร์รหัสนั้นๆ ซึ่งอาจจะแปลความหมายได้ 2 กรณี คือ

1. แอคทีฟยูสเซอร์ไม่เคยเลือกหรือให้ค่าคะแนนเรตติ้งใดๆ แก่ไอเท็มเลย

2. ไอเท็มที่คำนวณและสมควรจะอยู่ในรายการแนะนำนั้นได้ถูกแอคทีฟยูสเซอร์ให้ค่าคะแนนไปแล้ว

ลำดับสุดท้ายของการทดลองการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนและอัลกอริทึมฟัซซีซีมีน ก็คือการนำผลการทดลองด้านเวลาและค่าฮิตเรตจากแต่ละรอบการคำนวณของทั้งสองอัลกอริทึมสำหรับการจัดกลุ่มมาหาค่าเฉลี่ย เปรียบเทียบทั้งสองอัลกอริทึมเรียงลำดับตามขนาดของการจัดกลุ่มซึ่งเริ่มตั้งแต่ 2 กลุ่มไอเท็มไปจนถึง 5 กลุ่มไอเท็ม ซึ่งได้ผลการทดลองด้านเวลาแสดงดังตาราง 4.19 ตาราง และผลการทดลองเรื่องความถูกต้อง (Hit-Rate) แสดงดังตาราง 4.20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.11 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิ-ฟายฟ์ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 2 กลุ่ม)

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.013000	293,350,628,141,377,608,249,543,164,224 (กลุ่ม 0)
U8	0.013000	11,497,249,377,293,246,173,608,266,288 (กลุ่ม 0)
U17	0.013000	224,720,231,628,293,296,653,350,164,222 (กลุ่ม 0)
U26	0.013000	249,543,720,11,162,628,164,357,246,231 (กลุ่ม 0)
U35	0.013000	21,47,11,293,150,249,741,173,350,543 (กลุ่ม 0)
U56	0.014000	350,150,21,249,11,357,653,500,7,628 (กลุ่ม 0)
U60	0.014000	543,500,231,208,608,246,11,21,224,44 (กลุ่ม 0)
U61	0.024000	(กลุ่ม 0)
U63	0.014000	163,357,173,594,367,223,736,442,541,172 (กลุ่ม 0)
U69	0.014000	208,500,551,543,293,222,431,11,736,25 (กลุ่ม 0)
U73	0.014000	500,21,594,173,653,551,720,52,249,25 (กลุ่ม 0)
U75	0.140000	632,658,463,200,634,561,576,347,369,575 (กลุ่ม 1)
U76	0.013000	543,150,364,653,720,21,293,249,377,11 (กลุ่ม 0)
U79	0.132000	40,82,89,145,207,212,263,322,437,453 (กลุ่ม 1)
U80	0.014000	543,608,350,231,44,7,288,500,377,524 (กลุ่ม 0)
U82	0.014000	543,21,296,350,377,500,720,24,11,608 (กลุ่ม 0)
U88	0.162000	(กลุ่ม 1)
U89	0.013000	281,527,733,303,500,6,648,173,296,372 (กลุ่ม 0)
U93	0.014000	296,500,293,224,208,162,249,164,497,39 (กลุ่ม 0)
U94	0.012000	3,7,25,111,260,293,306,317,337,342 (กลุ่ม 0)
U102	0.013000	246,586,543,350,314,551,24,457,306,653 (กลุ่ม 0)
U107	0.013000	543,314,357,293,231,741,350,163,162,224 (กลุ่ม 0)
U111	0.013000	628,720,21,249,150,377,653,208,11,733 (กลุ่ม 0)
U118	0.013000	500,21,208,543,720,47,249,164,112,350 (กลุ่ม 0)
U123	0.014000	608,500,11,249,543,112,164,720,21,266 (กลุ่ม 0)
U126	0.014000	509,515,314,551,367,589,457,223,741,162 (กลุ่ม 0)
U127	0.014000	224,500,543,720,21,293,342,350,272,173 (กลุ่ม 0)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.12 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิ-ฟายฟ์ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 3 กลุ่ม)

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.014000	52,16,377,29,39,628,356,597,515,647 (กลุ่ม 0)
U8	0.014000	628,52,377,16,420,345,597,592,39,288 (กลุ่ม 0)
U17	0.002000	428,446,492,81,605,105,272,448,431,222 (กลุ่ม 1)
U26	0.002000	412,605,125,361,272,448,222,290,236,94 (กลุ่ม 1)
U35	0.002000	58,194,215,448,236,412,101,290,105,538 (กลุ่ม 1)
U56	0.015000	628,52,39,16,364,493,592,420,356,150 (กลุ่ม 0)
U60	0.014000	628,356,164,350,647,52,2,288,377,16 (กลุ่ม 0)
U61	0.025000	(กลุ่ม 0)
U63	0.014000	223,6,70,628,541,592,16,288,377,555 (กลุ่ม 0)
U69	0.002000	272,446,431,538,448,232,492,605,222,290 (กลุ่ม 1)
U73	0.015000	597,52,628,592,377,356,16,420,164,39 (กลุ่ม 0)
U75	0.124000	136,575,616,354,396,214,154,43,625,560 (กลุ่ม 2)
U76	0.004000	(กลุ่ม 1)
U79	0.112000	40,617,175,433,436,728,752,425,758,120 (กลุ่ม 2)
U80	0.015000	597,420,110,52,48,628,164,345,356,249 (กลุ่ม 0)
U82	0.014000	628,420,345,52,16,597,377,356,508,164 (กลุ่ม 0)
U88	0.146000	(กลุ่ม 2)
U89	0.004000	(กลุ่ม 1)
U93	0.014000	628,52,592,377,420,597,164,345,647,356 (กลุ่ม 0)
U94	0.014000	551,60,353,34,225,357,508,529,150,237 (กลุ่ม 0)
U102	0.004000	(กลุ่ม 1)
U107	0.015000	260,493,593,592,356,628,52,597,594,608 (กลุ่ม 0)
U111	0.015000	52,377,628,356,16,597,420,592,493,164 (กลุ่ม 0)
U118	0.004000	(กลุ่ม 1)
U123	0.014000	628,52,592,16,356,597,164,377,420,555 (กลุ่ม 0)
U126	0.002000	412,101,281,534,215,535,492,308,222,125 (กลุ่ม 1)
U127	0.015000	420,628,52,377,356,288,345,597,164,745 (กลุ่ม 0)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.13 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิ-ฟายฟัซซีซีมีน (กลุ่มข้อมูลไอเท็ม 4 กลุ่ม)

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.016000	52,16,377,29,39,222,597,356,647,288 (กลุ่ม 0)
U8	0.016000	52,377,345,16,222,597,508,288,39,592 (กลุ่ม 0)
U17	0.025000	428,628,252,44,585,431,493,196,524,225 (กลุ่ม 2)
U26	0.016000	412,125,307,73,314,306,678 (กลุ่ม 3)
U35	0.016000	(กลุ่ม 3)
U56	0.015000	39,52,16,364,150,222,592,492,345,356 (กลุ่ม 0)
U60	0.015000	356,164,2,647,350,52,288,377,16,222 (กลุ่ม 0)
U61	0.141000	762,60,704,546,115,439,633,660,109,123 (กลุ่ม 1)
U63	0.016000	223,6,70,541,592,16,288,377,555,597 (กลุ่ม 0)
U69	0.025000	431,428,538,628,237,208,44,442,466,196 (กลุ่ม 2)
U73	0.015000	597,52,222,377,356,592,164,16,223,249 (กลุ่ม 0)
U75	0.187000	575,136,616,625,354,154,396,43,560,214 (กลุ่ม 1)
U76	0.015000	345,52,745,222,597,293,720,647,377,164 (กลุ่ม 0)
U79	0.156000	40,481,617,175,433,436,728,752,425,361 (กลุ่ม 1)
U80	0.015000	597,110,52,48,222,164,345,272,356,249 (กลุ่ม 0)
U82	0.015000	345,16,52,597,377,356,508,222,164,344 (กลุ่ม 0)
U88	0.219000	(กลุ่ม 1)
U89	0.015000	527,733,52,745,720,475,272,345,151,377 (กลุ่ม 0)
U93	0.015000	52,377,592,222,345,597,164,647,720,356 (กลุ่ม 0)
U94	0.015000	551,353,34,357,508,529,150,468,708,317 (กลุ่ม 0)
U102	0.015000	377,24,111,555,222,595,52,592,597,249 (กลุ่ม 0)
U107	0.015000	260,593,592,356,492,52,222,597,594,608 (กลุ่ม 0)
U111	0.120000	(กลุ่ม 3)
U118	0.120000	(กลุ่ม 3)
U123	0.016000	592,52,356,16,164,597,377,222,555,25 (กลุ่ม 0)
U126	0.120000	412,125,307,314,73,306 (กลุ่ม 3)
U127	0.016000	356,52,377,288,597,164,222,345,745,16 (กลุ่ม 0)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.14 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิ-ฟายฟ์ซซีซีมีน (กลุ่มข้อมูลไอเท็ม 5 กลุ่ม)

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.002000	176,101,125,464,338,556,663,157,81 (กลุ่ม 0)
U8	0.021000	628,52,377,420,16,345,597,592,39,288 (กลุ่ม 2)
U17	0.020000	628,52,164,597,420,327,356,162,377,293 (กลุ่ม 2)
U26	0.002000	125,556,176,663,101,338,157,81,574,464 (กลุ่ม 0)
U35	0.002000	(กลุ่ม 0)
U56	0.019000	628,52,39,16,364,493,592,420,356,150 (กลุ่ม 2)
U60	0.001000	60,546,616,19 (กลุ่ม 4)
U61	0.152000	762,704,115,439,633,660,123,109,751,417 (กลุ่ม 1)
U63	0.020000	223,6,70,628,541,592,16,288,377,555 (กลุ่ม 2)
U69	0.020000	25,356,592,52,104,628,272,431,515,16 (กลุ่ม 2)
U73	0.021000	597,52,628,592,377,356,16,420,164,327 (กลุ่ม 2)
U75	0.001000	616,19,546,107,60 (กลุ่ม 4)
U76	0.026000	345,52,745,597,293,327,720,628,420,647 (กลุ่ม 2)
U79	0.168000	40,481,617,175,433,436,728,752,425,361 (กลุ่ม 1)
U80	0.021000	597,420,110,52,48,628,164,345,272,356 (กลุ่ม 2)
U82	0.020000	628,52,420,16,345,597,377,356,164,508 (กลุ่ม 2)
U88	0.218000	(กลุ่ม 1)
U89	0.020000	527,733,475,52,745,720,345,272,493,420 (กลุ่ม 2)
U93	0.020000	628,52,592,377,420,597,327,164,345,647 (กลุ่ม 2)
U94	0.018000	551,353,34,225,357,508,529,150,517,468 (กลุ่ม 2)
U102	0.168000	203,237,68,652,4,492,71,84,309,243 (กลุ่ม 1)
U107	0.020000	260,493,593,592,356,628,52,597,594,608 (กลุ่ม 2)
U111	0.002000	(กลุ่ม 0)
U118	0.001000	(กลุ่ม 4)
U123	0.001000	101,81,663,556,176,464,125,338,157,574 (กลุ่ม 0)
U126	0.182000	412,390,171,579,625,581,523,495,200,396 (กลุ่ม 1)
U127	0.020000	420,628,52,377,356,288,345,597,164,745 (กลุ่ม 2)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.15 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้ฟังชั่น-ซีมีน (กลุ่มข้อมูลไอเท็ม 2 กลุ่ม)

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.113000	136,52,16,354,682,269,728,625,8,642 (กลุ่ม 0)
U8	0.023000	110,377,737,420,154,650,475,344,597,733 (กลุ่ม 1)
U17	0.023000	213,597,556,154,319,720,43,214,249,110 (กลุ่ม 1)
U26	0.114000	136,628,52,118,31,682,563,16,354,438 (กลุ่ม 0)
U35	0.115000	28,708,7,155,608,136,545,628,269,749 (กลุ่ม 0)
U56	0.132000	628,136,682,747,118,146,243,288,52,563 (กลุ่ม 0)
U60	0.115000	136,350,518,356,513,164,2,269,200,52 (กลุ่ม 0)
U61	0.099000	762,60,704,546,115,439,160,109,551,633 (กลุ่ม 0)
U63	0.115000	223,750,6,642,541,136,628,601,248,16 (กลุ่ม 0)
U69	0.023000	515,593,538,296,360,26,17,213,492,592 (กลุ่ม 1)
U73	0.023000	597,214,43,592,420,39,577,154,377,159 (กลุ่ม 1)
U75	0.114000	575,136,52,396,354,682,651,560,222,269 (กลุ่ม 0)
U76	0.024000	296,593,47,492,592,497,493,720,577,360 (กลุ่ม 1)
U79	0.114000	40,164,350,354,436,448,175,433,584,625 (กลุ่ม 0)
U80	0.115000	527,356,136,682,48,62,590,118,52,448 (กลุ่ม 0)
U82	0.023000	110,420,556,587,43,597,154,377,344,249 (กลุ่ม 1)
U88	0.115000	527,111,608,136,682,448,611,118,120,52 (กลุ่ม 0)
U89	0.023000	733,493,296,454,360,714,137,592,492,435 (กลุ่ม 1)
U93	0.023000	420,720,592,377,43,154,214,737,597,249 (กลุ่ม 1)
U94	0.105000	551,109,60,640,297,34,633,353,660,723 (กลุ่ม 0)
U102	0.116000	111,750,555,136,518,642,203,628,164,31 (กลุ่ม 0)
U107	0.023000	593,594,318,413,335,309,358,533,343,313 (กลุ่ม 1)
U111	0.115000	222,136,682,52,438,118,354,254,611,625 (กลุ่ม 0)
U118	0.024000	39,377,592,214,597,420,43,159,577,110 (กลุ่ม 1)
U123	0.115000	136,628,518,356,52,438,16,523,31,354 (กลุ่ม 0)
U126	0.023000	589,648,171,587,85,556,344,30,558,262 (กลุ่ม 1)
U127	0.023000	110,43,420,737,377,154,214,597,556,647 (กลุ่ม 1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.16 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้ฟังชั่น-ซีมีน (กลุ่มข้อมูลไอเท็ม 3 กลุ่ม)

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.006000	377,29,272,269,39,31,523,747,76,50 (กลุ่ม 2)
U8	0.009000	110,650,105,24,25,647,438,597,278,253 (กลุ่ม 1)
U17	0.006000	76,235,293,30,747,50,269,436,31,32 (กลุ่ม 2)
U26	0.008000	597,278,605,438,154,577,249,381,547,556 (กลุ่ม 1)
U35	0.005000	39,28,58,29,269,709,377,747,31,523 (กลุ่ม 2)
U56	0.111000	628,682,136,563,52,222,16,364,118,150 (กลุ่ม 0)
U60	0.111000	350,136,737,356,628,518,368,164,43,682 (กลุ่ม 0)
U61	0.095000	762,60,546,115,160,439,353,34,123,551 (กลุ่ม 0)
U63	0.006000	541,671,528,375,280,377,459,29,536,425 (กลุ่ม 2)
U69	0.009000	515,25,321,17,141,577,137,647,358,438 (กลุ่ม 1)
U73	0.111000	136,592,628,356,682,52,354,222,625,200 (กลุ่ม 0)
U75	0.009000	19,631,577,154,597,438,249,365,2,278 (กลุ่ม 1)
U76	0.005000	50,293,593,536,441,76,377,747,523,269 (กลุ่ม 2)
U79	0.111000	40,164,350,354,448,43,214,175,737,433 (กลุ่ม 0)
U80	0.114000	136,682,118,48,611,527,448,560,354,420 (กลุ่ม 0)
U82	0.008000	587,589,110,648,278,556,24,154,547,249 (กลุ่ม 1)
U88	0.008000	110,318,443,597,233,438,249,365,154,647 (กลุ่ม 1)
U89	0.110000	527,475,448,120,611,720,345,118,682,344 (กลุ่ม 0)
U93	0.009000	589,741,648,278,316,556,653,249,182,438 (กลุ่ม 1)
U94	0.017000	(กลุ่ม 1)
U102	0.008000	24,105,515,542,435,647,505,650,137,319 (กลุ่ม 1)
U107	0.008000	318,458,71,216,309,540,141,357,233,80 (กลุ่ม 1)
U111	0.008000	318,357,36,246,281,24,80,505,137,577 (กลุ่ม 1)
U118	0.005000	377,39,95,523,165,733,269,493,333,272 (กลุ่ม 2)
U123	0.006000	377,523,333,269,436,31,747,273,95,272 (กลุ่ม 2)
U126	0.008000	648,412,589,581,495,587,85,278,556,316 (กลุ่ม 1)
U127	0.009000	110,589,587,253,648,647,443,278,2,154 (กลุ่ม 1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.17 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้ฟัซซี-ซีมีน (กลุ่มข้อมูลไอเท็ม 4 กลุ่ม)

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.031000	16,728,39,162,420,272,222,448,549,575 (กลุ่ม 3)
U8	0.047000	136,337,253,288,475,682,527,396,369,754 (กลุ่ม 0)
U17	0.031000	162,448,164,420,293,118,575,720,728,214 (กลุ่ม 3)
U26	0.015000	16,118,356,222,420,565,39,448,214,164 (กลุ่ม 3)
U35	0.016000	(กลุ่ม 2)
U56	0.047000	527,475,120,682,136,369,396,337,563,331 (กลุ่ม 0)
U60	0.016000	350,2,10,329,647,737,496,97,492,518 (กลุ่ม 1)
U61	0.031000	(กลุ่ม 1)
U63	0.032000	750,223,541,214,577,448,427,420,16,164 (กลุ่ม 3)
U69	0.016000	593,335,71,583,709,313,628,141,31,343 (กลุ่ม 2)
U73	0.016000	597,377,110,159,154,692,438,354,52,710 (กลุ่ม 1)
U75	0.016000	19,628,709,13,580,31,547,249,500,609 (กลุ่ม 2)
U76	0.016000	593,335,71,583,709,313,628,141,31,343 (กลุ่ม 2)
U79	0.047000	40,648,125,425,433,120,33,752,359,475 (กลุ่ม 0)
U80	0.016000	231,24,593,709,628,31,313,269,343,71 (กลุ่ม 2)
U82	0.062000	32,648,527,256,120,749,395,411,425,475 (กลุ่ม 0)
U88	0.062000	527,475,120,682,136,369,396,337,563,331 (กลุ่ม 0)
U89	0.047000	527,475,120,682,136,369,396,337,563,331 (กลุ่ม 0)
U93	0.016000	377,741,6,492,438,154,116,52,710,354 (กลุ่ม 1)
U94	0.031000	551,353,34,417,68,508,483,289,309,708 (กลุ่ม 3)
U102	0.031000	111,750,555,577,214,728,420,164,549,272 (กลุ่ม 3)
U107	0.016000	260,593,714,419,454,667,709,141,408,605 (กลุ่ม 2)
U111	0.047000	357,32,497,342,529,369,682,468,136,120 (กลุ่ม 0)
U118	0.016000	260,95,380,454,605,419,714,165,667,137 (กลุ่ม 2)
U123	0.047000	592,527,416,81,642,37,136,682,523,126 (กลุ่ม 0)
U126	0.016000	260,714,454,419,667,408,137,145,473,80 (กลุ่ม 2)
U127	0.047000	527,288,136,32,648,253,396,475,682,369 (กลุ่ม 0)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.18 แสดงผลการทดลองบางส่วน การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้ฟัซซี-ซีมีน (กลุ่มข้อมูลไอเท็ม 5 กลุ่ม)

ไอเท็มเป้าหมาย :: 50, 296, 364, 377, 589, 733, 858, 925, 1393, 1617		
รหัสผู้ใช้	**เวลาที่ใช้ (s)**	**รหัสไอเท็ม**
U5	0.010000	515,39,162,565,76,438,523,43,747,647 (กลุ่ม 3)
U8	0.057000	253,73,475,140,375,53,756,136,699,274 (กลุ่ม 0)
U17	0.023000	164,223,319,235,222,327,436,616,518,345 (กลุ่ม 1)
U26	0.011000	52,16,592,356,377,467,354,710,19,611 (กลุ่ม 4)
U35	0.011000	154,39,28,17,438,597,357,43,249,565 (กลุ่ม 3)
U56	0.008000	246,608,34,609,360,493,102,166,371,263 (กลุ่ม 1)
U60	0.026000	2,380,113,357,647,76,515,39,154,577 (กลุ่ม 3)
U61	0.047000	(กลุ่ม 2)
U63	0.114000	(กลุ่ม 0)
U69	0.011000	446,356,431,592,492,296,467,52,232,16 (กลุ่ม 4)
U73	0.026000	597,692,39,577,43,154,671,278,523,438 (กลุ่ม 3)
U75	0.008000	87,616,609,60,546,269,222,327,393,493 (กลุ่ม 1)
U76	0.011000	47,296,488,329,492,356,467,592,19,16 (กลุ่ม 4)
U79	0.024000	164,518,532,616,222,365,182,436,272,327 (กลุ่ม 1)
U80	0.008000	608,609,435,493,360,487,203,368,141,468 (กลุ่ม 1)
U82	0.026000	587,589,457,24,318,278,154,747,43,50 (กลุ่ม 3)
U88	0.008000	608,609,435,493,360,487,203,368,141,468 (กลุ่ม 1)
U89	0.039000	(กลุ่ม 3)
U93	0.023000	741,368,172,292,518,203,271,182,556,497 (กลุ่ม 1)
U94	0.038000	(กลุ่ม 1)
U102	0.026000	24,105,515,542,747,259,647,650,357,181 (กลุ่ม 3)
U107	0.031000	593,260,594,628,111,709,733,288,80,580 (กลุ่ม 2)
U111	0.009000	222,497,246,32,327,269,342,345,427,348 (กลุ่ม 1)
U118	0.026000	39,357,589,443,577,367,380,154,43,333 (กลุ่ม 3)
U123	0.009000	164,223,436,608,616,518,327,222,368,365 (กลุ่ม 1)
U126	0.024000	(กลุ่ม 4)
U127	0.009000	608,32,158,41,203,493,691,393,360,319 (กลุ่ม 1)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.19 แสดงเวลาที่ใช้ในการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น โดยใช้โมดิฟายฟัซซีซีมีนเปรียบเทียบกับฟัซซีซีมีน

จำนวนกลุ่ม	การจัดกลุ่ม	เวลา (s) รอบที่ 1	เวลา (s) รอบที่ 2	เวลา (s) รอบที่ 3	เวลา (s) รอบที่ 4	เวลา (s) รอบที่ 5	เวลา (s) เฉลี่ย
2C	MFCM	6.199997	5.547000	6.727998	5.771000	5.490001	5.947200
	FCM	14.348001	9.017998	13.100999	12.319999	13.014003	12.360200
3C	MFCM	4.079998	2.662999	4.733999	3.730000	3.287001	3.698800
	FCM	9.040999	4.020000	6.322998	5.954999	6.377000	6.343199
4C	MFCM	4.839999	4.835000	5.392000	5.492001	3.651000	3.698799
	FCM	4.407000	3.848998	4.287000	4.777001	4.005002	4.265000
5C	MFCM	4.377000	4.827001	5.431001	4.870001	5.720001	5.045000
	FCM	4.540998	3.141002	2.942001	3.607998	3.945003	3.635400

ตารางที่ 4.20 แสดงค่าความถูกต้องของการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็น โดยใช้โมดิฟายฟัซซีซีมีนเปรียบเทียบกับฟัซซีซีมีน

จำนวนกลุ่ม	การจัดกลุ่ม	Hit-Rate รอบที่ 1	Hit-Rate รอบที่ 2	Hit-Rate รอบที่ 3	Hit-Rate รอบที่ 4	Hit-Rate รอบที่ 5	Hit-Rate เฉลี่ย
2C	MFCM	3.510638	3.468085	3.687831	3.497354	3.386243	3.510030
	FCM	3.015957	2.898936	2.698413	2.978836	2.423280	2.803084
3C	MFCM	2.882979	3.122340	3.185185	2.693122	2.693122	2.915349
	FCM	2.882379	3.053191	2.851852	2.608466	2.539683	2.787114
4C	MFCM	2.606383	3.234043	2.888889	2.719577	3.074074	2.908993
	FCM	2.579787	2.696809	2.862434	2.751323	20264550	2.630981
5C	MFCM	2.505319	2.186170	2.904762	2.455026	2.798942	2.570044
	FCM	2.675532	2.335106	2.481481	2.074074	1.978836	2.390060

จากผลลัพธ์ในตาราง 4.19 ชี้ว่าเวลาที่ใช้ในการแนะนำนั้นจะค่อยๆ ลดลงเมื่อการแบ่งกลุ่มข้อมูลไอเท็มมีจำนวนมากขึ้น แต่ทั้งนี้จำนวนไอเท็มสมาชิกของแต่ละกลุ่มข้อมูล ก็มีผลต่อเวลาที่ใช้ในการแนะนำเช่นเดียวกัน ในการแบ่งกลุ่มข้อมูลออกเป็น 5 กลุ่มข้อมูล เวลาที่ใช้ในการแนะนำโดยใช้โมดิฟายฟังชั่นซีมีนจะสูงกว่าการใช้ฟังชั่นซีมีน เนื่องจากไอเท็มสมาชิกภายในกลุ่มที่เหมาะสมกับแอคทีพยูสเซอร์ส่วนใหญ่ของการจัดกลุ่มโดยใช้โมดิฟายฟังชั่นซีมีนนั้น มีจำนวนไอเท็มสมาชิกของกลุ่มมากกว่าการจัดกลุ่มด้วยฟังชั่นซีมีน แต่การจัดกลุ่มข้อมูลไอเท็มด้วยอัลกอริทึมทั้งสองแบบช่วยให้เวลาในการแนะนำนั้นดีขึ้น หรืออีกนัยหนึ่งคือใช้เวลาน้อยกว่าการแนะนำสินค้าแบบไอเท็ม-เบสท็อปเอ็นนั่นเอง สามารถดูการเปรียบเทียบประสิทธิภาพด้านเวลาจากตารางที่ 4.21 และแผนภูมิแท่งตามรูปที่ 4.1

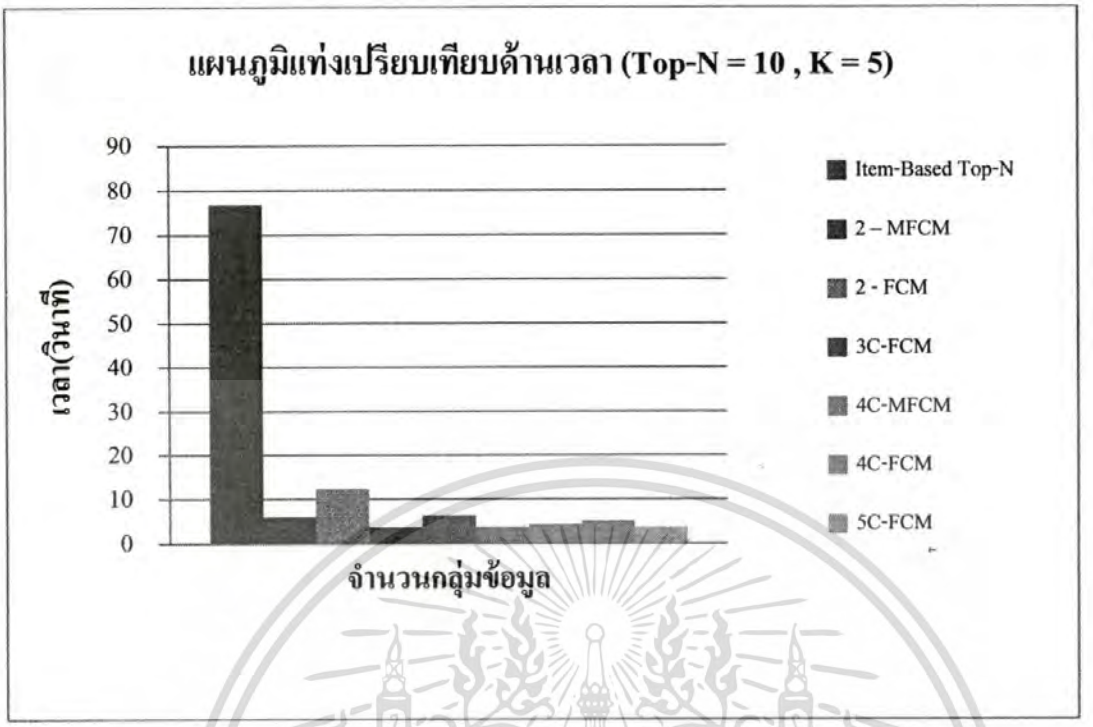
และจากผลลัพธ์ในตาราง 4.20 แสดงให้เห็นว่าค่าความถูกต้องในการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟังชั่นซีมีนในการแนะนำสินค้าจำนวน 10 อันดับรายการ มีค่าเฉลี่ยสูงกว่าการใช้การจัดกลุ่มด้วยฟังชั่นซีมีน ซึ่งค่าความถูกต้องนี้ขึ้นอยู่กับจำนวนกลุ่มข้อมูลไอเท็มด้วย การแบ่งกลุ่มข้อมูลออกเป็น 2 กลุ่มจะมีค่าความถูกต้องเฉลี่ยสูงกว่า การแบ่งกลุ่มข้อมูลออกเป็น 3 กลุ่ม 4 กลุ่มและ 5 กลุ่ม เนื่องจากจำนวนไอเท็มที่สามารถแนะนำได้ในการแบ่งกลุ่มข้อมูลไอเท็ม 2 กลุ่มจะมีมากกว่า 3 กลุ่ม 4 กลุ่ม และ 5 กลุ่ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

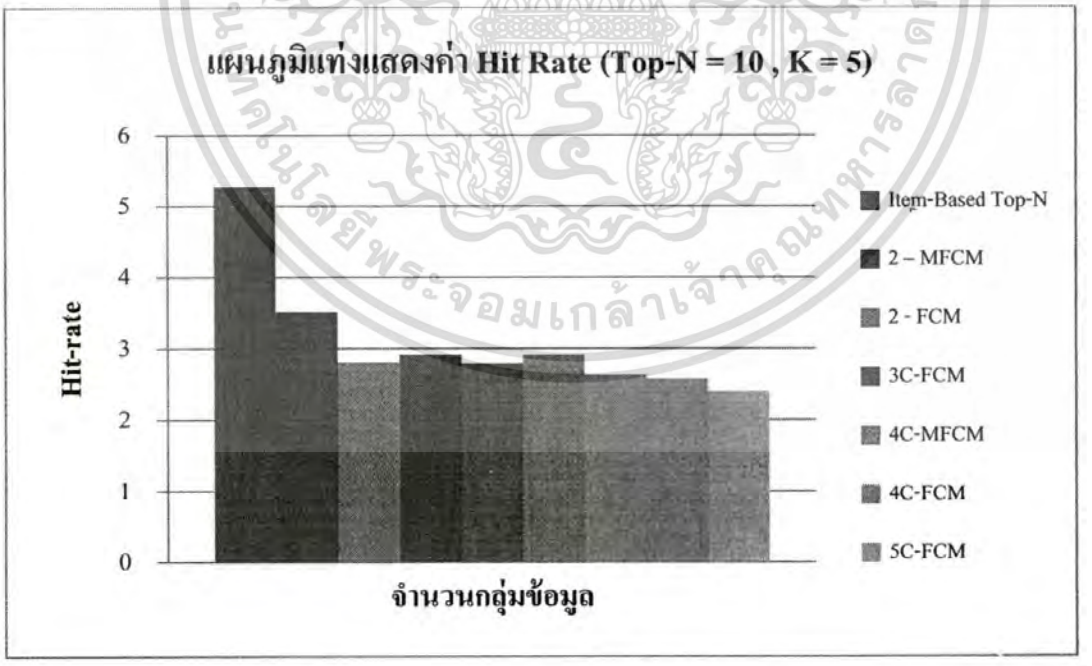
ตารางที่ 4.21 แสดงการเปรียบเทียบเวลาที่ใช้ในการแนะนำแบบต่างๆ

จำนวนกลุ่ม	การจัดกลุ่ม	เวลา (s) รอบที่ 1	เวลา (s) รอบที่ 2	เวลา (s) รอบที่ 3	เวลา (s) รอบที่ 4	เวลา (s) รอบที่ 5	เวลา (s) เฉลี่ย
1C	Item-Based	78.859000	62.682000	81.213001	79.894000	81.104000	76.750400
2C	MFCM	6.199997	5.547000	6.727998	5.771000	5.490001	5.947200
	FCM	14.348001	9.017998	13.100999	12.319999	13.014003	12.360200
3C	MFCM	4.079998	2.662999	4.733999	3.730000	3.287001	3.698800
	FCM	9.040999	4.020000	6.322998	5.954999	6.377000	6.343199
4C	MFCM	4.839999	4.835000	5.392000	5.492001	3.651000	3.698799
	FCM	4.407000	3.848998	4.287000	4.777001	4.005002	4.265000
5C	MFCM	4.377000	4.827001	5.431001	4.870001	5.720001	5.045000
	FCM	4.540998	3.141002	2.942001	3.607998	3.945003	3.635400

จากตาราง 4.21 แสดงว่าเวลาที่ใช้ในการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนนั้น เร็วกว่าการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นทุกการแบ่งกลุ่ม ซึ่งจากผลการทดลองภายในบทที่ 4 ชี้ให้เห็นว่าการนำข้อมูลไอเท็มไปจัดกลุ่มก่อนทำการแนะนำ จะช่วยให้เวลาที่ใช้สำหรับสร้างรายการแนะนำลดลงได้จริงตรงตามวัตถุประสงค์หลักของงานวิจัยเรื่องการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนฉบับนี้



รูปที่ 4.1 แสดงแผนภูมิแท่งเปรียบเทียบด้านเวลา



รูปที่ 4.2 แสดงแผนภูมิแท่งเปรียบเทียบค่าฮิตเรต

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
สรุปผลการวิจัย และข้อเสนอแนะ

ในบทนี้จะกล่าวถึงผลสรุปของงานวิจัยเรื่อง การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน และข้อเสนอแนะ

5.1 สรุปผลการวิจัย

ระบบแนะนำพัฒนามาจากเทคโนโลยีการกรองข้อมูลส่วนบุคคล เพื่อการทำนายหรือระบุกลุ่มของไอเท็มที่น่าสนใจแก่ผู้ใช้ แนวคิดเรื่องการทำระบบแนะนำถูกนำไปใช้กับหลายโปรแกรมประยุกต์ หลายวัตถุประสงค์ทั้งในด้านการนำเสนอข้อมูลหรือสินค้าในแง่ของธุรกิจ ประเภทของระบบแนะนำนั้นแบ่งออกเป็น 2 ประเภทหลัก คือ

1. User-Based Recommend จะอิงข้อมูลประชากรศาสตร์ของผู้ใช้ในระบบนำมาวิเคราะห์ความสัมพันธ์ระหว่างผู้ใช้ ค้นหาตัวผู้ใช้ที่มีความคล้ายคลึงกันเพื่อแนะนำสินค้าของคนหนึ่งให้อีกคนหนึ่ง ทว่าการคำนวณเพื่อสร้างรายการไอเท็มสินค้าแบบอิงข้อมูลผู้ใช้นี้ซับซ้อน เติบโตตามจำนวนของผู้ใช้ในระบบและจำนวนด้านของข้อมูลส่วนตัวของผู้ใช้ ซึ่งการที่จะแนะนำข้อมูลให้มีประสิทธิภาพอาจจะต้องใช้ข้อมูลประชากรศาสตร์หลายๆ ด้านมาประกอบกัน จึงเกิดเป็นแนวคิดการแนะนำประเภทที่ 2 โดยใช้ข้อมูลไอเท็มในระบบขึ้น

2. Item-Based Recommend สร้างรายการแนะนำโดยอิงมาจากข้อมูลไอเท็มภายในระบบเปรียบเทียบความคล้ายคลึงระหว่างไอเท็ม สร้างเป็นเมตริกซ์หว่างไอเท็ม-ไอเท็มเพื่อค่าความสัมพันธ์ระหว่างไอเท็มสำหรับนำไปสร้างรายการแนะนำ โดยการคูณเข้ากับเมตริกซ์ 1 * m ซึ่งเป็นเมตริกซ์แสดงความสัมพันธ์ระหว่างแอคทีฟยูสเซอร์และไอเท็มในระบบ ทว่าเมื่อปริมาณไอเท็มมีจำนวนมากขึ้นการสร้างรายการแนะนำจากเมตริกซ์ผลลัพธ์ก็ใช้เวลานานตามไปโดยปริยาย

นักวิจัยทางด้านเทคโนโลยีการกรองข้อมูลและระบบแนะนำ จึงได้นำแนวคิดการจัดกลุ่มเข้ามาประยุกต์ใช้ โดยให้นำข้อมูลไอเท็มไปจัดกลุ่มเสียก่อนแล้วจึงเข้าสู่กระบวนการสร้างโมเดลของไอเท็ม ซึ่งอัลกอริทึมสำหรับจัดกลุ่มข้อมูลนั้นมีอยู่หลายอัลกอริทึม เคมีนเป็นหนึ่งอัลกอริทึมสำหรับจัดกลุ่มที่เป็นที่รู้จักและเป็นที่นิยม เพราะประยุกต์ใช้งานได้ง่าย/ไม่ซับซ้อน แต่เคมีนก็มีข้อจำกัดในเรื่องของการที่ตัวอัลกอริทึมจะยอมให้ข้อมูลตัวที่กำลังพิจารณาอยู่นั้น เป็นสมาชิกของกลุ่มใดกลุ่มหนึ่งได้เพียงกลุ่มเดียว ผู้วิจัยจึงเลือกใช้ฟัซซีซีมีนซึ่งเป็นอัลกอริทึมที่พัฒนามาจากเคมีนแต่แตกต่างตรงที่ยืดหยุ่นกว่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฟัซซีซีมีนยินยอมให้ข้อมูลหนึ่งๆ ที่กำลังพิจารณาสามารถที่จะเป็นสมาชิกของหลายกลุ่มข้อมูลได้พร้อมกัน โดยที่ข้อมูลตัวหนึ่งๆ จะมีค่าความเป็นสมาชิกของแต่ละกลุ่มที่แตกต่างกันออกไป ผลจากการใช้ฟัซซีเข้ามาช่วยในการจัดกลุ่มข้อมูลนี้ทำให้จัดกลุ่มข้อมูลที่มีความคลุมเครือได้อย่างมีประสิทธิภาพ และเนื่องจากการแนะนำสินค้าแบบไอเท็มเบสอยู่บนแนวคิดพื้นฐานที่ว่าไอเท็มที่สมควรอยู่ในรายการแนะนำแก่แอคทีฟยูสเซอร์ จะคล้ายคลึงกับไอเท็มที่แอคทีฟยูสเซอร์เคยเลือกซื้อไปแล้วในอดีต ดังนั้นผู้วิจัยจึงได้นำการหาค่าความคล้ายคลึงด้วยค่าสหสัมพันธ์จากข้อมูลที่เกิดร่วมกันมาประยุกต์ใช้ในสมการสำหรับหาค่าความเป็นสมาชิกในฟัซซีซีมีนอัลกอริทึมแทนการคำนวณหาค่าระยะห่างระหว่างไอเท็มกับจุดศูนย์กลางของกลุ่มไอเท็ม

งานวิจัยฉบับนี้ประกอบไปด้วย 3 ส่วนหลัก คือ

1. การจัดกลุ่มไอเท็มด้วยโมดิฟายฟัซซีซีมีนอัลกอริธึม
2. การแนะนำไอเท็มแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน
3. การเปรียบเทียบประสิทธิภาพด้านเวลาและความถูกต้อง

ในส่วนของการทดลองเพื่อพิสูจน์แนวคิดของงานวิจัย ผู้วิจัยได้ใช้หลักการเคโฟล์ครอส-วาลิเดชั่น โดยนำชุดข้อมูลซึ่งในที่นี้ใช้ชุดข้อมูลเรตติ้งมูฟวี่เลนส์มาทำการแบ่งย่อยออกเป็น 5 ชุด ด้วยการสุ่มเลือกสำหรับเป็นชุดข้อมูลเทรนนิ่งกับเทส เพื่อประมวลผลและทดสอบความถูกต้องในการแนะนำไอเท็ม แล้วจึงจัดกลุ่มไอเท็มจากข้อมูลเทรนนิ่งแต่ละรอบการคำนวณด้วยอัลกอริทึมโมดิฟายฟัซซีซีมีนและฟัซซีซีมีน ในจำนวนกลุ่ม 2, 3, 4 และ 5 กำหนดให้ค่า $\varepsilon = 0.1$

สร้างเมตริกซ์ไอเท็ม-ไอเท็มของแต่ละกลุ่มข้อมูลไอเท็มที่อยู่ในแต่ละรอบการคำนวณ ด้วยการเปรียบเทียบค่าความคล้ายคลึงระหว่างไอเท็ม ซึ่งเมตริกซ์ดังกล่าวเป็นสิ่งที่บอกถึงความสัมพันธ์ระหว่างไอเท็มในกลุ่มข้อมูลเดียวกัน

เริ่มต้นกระบวนการแนะนำสำหรับแอคทีฟยูสเซอร์ด้วยการเลือกกลุ่มที่เหมาะสม ใช้วิธีการคำนวณจากค่าเฉลี่ยเรตติ้งของไอเท็มที่แอคทีฟยูสเซอร์เคยให้เอาไว้ในแต่ละกลุ่มข้อมูล ทำการเลือกกลุ่มไอเท็มที่มีค่าเฉลี่ยสูงที่สุด แล้วจึงสร้างเมตริกซ์ $1^* m$ ของแอคทีฟยูสเซอร์ เมตริกซ์ดังกล่าวนี้เป็นตัวแทนที่แสดงถึงความสัมพันธ์ระหว่างแอคทีฟยูสเซอร์กับไอเท็มในกลุ่มที่เหมาะสม

นำไอเท็ม-ไอเท็มเมตริกซ์มาคูณกับ $1^* m$ เมตริกซ์ของแอคทีฟยูสเซอร์ หลังจากดำเนินการเสร็จก็นำเมตริกซ์ผลลัพธ์ที่ได้ไปสร้างรายการแนะนำ เพื่อแนะนำแก่แอคทีฟยูสเซอร์ทั้งหมด 10 รายการ (Top-N = 10) เรียงตามค่าไอเท็มในเมตริกซ์ผลลัพธ์จากมากไปน้อย และตัดไอเท็มตัวที่แอคทีฟยูสเซอร์เคยให้ค่าเรตติ้งไว้แล้วกับไอเท็มที่เป็นศูนย์ออกไป ด้วยดำเนินการทดลองตามหลักการของเคโฟล์ครอสวาลิเดชั่น ขั้นตอนสุดท้ายก็คือการนำผลการทดลองด้านเวลาและค่าฮิตเรตในแต่ละรอบการคำนวณมาหาค่าเฉลี่ย

สำหรับผลของการแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นเพื่อนำมาเปรียบเทียบด้านเวลา ที่ใช้ในการแนะนำไอเท็มให้เป็นไปตามงานวิจัยเดิม [10]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 ประสิทธิภาพของอัลกอริธึม

ในงานวิจัยฉบับนี้มุ่งเน้นที่จะปรับปรุงเรื่องความเร็วในการแนะนำไอเท็มเป็นหลัก ดังนั้นจึงได้ทำการจับเวลาที่ใช้ในการแนะนำสินค้าด้วยไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีน ของแต่ละแอคทีฟยูสเซอร์ นำผลรวมของทั้ง 5 รอบการคำนวณเปรียบเทียบกับวิธีการแนะนำสินค้าด้วยไอ-เท็มเบสท็อปเอ็น

คำนวณหาค่าฮิตเรตเพื่อเปรียบเทียบประสิทธิภาพด้านความถูกต้องในการจัดกลุ่ม ระหว่างการแนะนำสินค้าด้วยไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนกับการใช้ฟัซซีซีมีน

เมื่อนำข้อมูลชุดมูฟวี่เลนส์ซึ่งเป็นชุดเดียวกันกับที่ใช้ในการแนะนำสินค้าแบบไอเท็มเบสมาผ่านกระบวนการแนะนำสินค้าด้วยโมดิฟายฟัซซีซีมีนแล้ว จะพบว่าเวลาที่ใช้สำหรับสร้างรายการไอเท็มจนถึงแนะนำไอเท็มแก่ผู้ใช้ของแต่ละแอคทีฟยูสเซอร์และผลรวมเวลาทั้งหมดลดลง

ในด้านการประเมินประสิทธิภาพของอัลกอริทึมโมดิฟายฟัซซีซีมีน ผู้วิจัยได้เทียบเคียงจากผลการแนะนำที่เกิดจากการจัดกลุ่มข้อมูลด้วยฟัซซีซีมีนอัลกอริทึม ซึ่งคำนวณในรูปของฮิตเรตและจากผลการทดลองพบว่า ค่าฮิตเรตของการแนะนำโดยใช้โมดิฟายฟัซซีซีมีนสูงกว่าการใช้ฟัซซีซีมีน

ทั้งนี้จำนวนกลุ่มของไอเท็มและจำนวนสมาชิกของไอเท็มภายในกลุ่ม ก็ส่งผลต่อเวลาที่ใช้ในการแนะนำด้วยเช่นเดียวกัน ถ้าจำนวนกลุ่มข้อมูลยิ่งมากขึ้น เวลาที่สูญเสียการในการสร้างการแนะนำก็ยิ่งลดลง

5.3 ข้อเสนอแนะ

ควรทำการทดลองตามแนวคิดของงานวิจัยเรื่อง การแนะนำสินค้าแบบไอเท็มเบสท็อปเอ็นโดยใช้โมดิฟายฟัซซีซีมีนกับชุดข้อมูลอื่นๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] P. Resnick and H. R. Varian, "Recommender systems," Communications of the ACM, 40(3):77-87, 1997.

[2] G. Karypis, "Evaluation of Item-Based Top-N Recommendation algorithm,": In Proceedings of the ACM Conference on Information and Knowledge Mangement. ACM, New York, 2001.

[3] K. Teknomo, "Similarity Measurement," [Online]. Available: http://people.revoledu.com/kardi/tutorial/Similarity/index.html.

[4] E. Garcia, "Cosine Similarity and Term Weight Tutorial" [Online]. Available: http://www.miislita.com/information-retrieval-tutorial/cosine-similarity -tutorial.html#Cosim.

[5] MovieLens data, http://www.grouplens.org/.

[6] J. L. Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," The American Statistician, 42(1):59–66, Febuary 1988.

[7] "Fuzzy C-Means Clustering," [Online]. Available: http://home.dei.polimi.it/matteucc/Clustering/ tutorial_html/cmeans.html.

[8] Y. Huang, "An Item Based Collaborative Filtering Using Item Clustering Prediction", ISECS International Colloquium on Computing, Commnication, Control, and Management, 2009.

[9] M. O'Connor and J. Herlocker, "Clustering Items for Collaborative Filtering," In Proceedings of the ACM SIGIR Workshop on Recommender Systems, Berkeley, CA, 1999.

[10] M. Deshpande and G. Karypis, "Item-Based Top-N Recommendation Algorithms," ACM Transaction on ACM Transactions on Information Systems. Volume 22, Issue 1, pp. 143 - 177, 2004.

[11] T. Motzkin, "The Euclidean Algorithm." Bull. Amer. Math. Soc. 55, 1142-1146, 1949.

[12] L. A. Zadeh, "Fuzzy sets," Information and Control 8 (3) 338–353, 1965.

[13] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics 3: 32-57, 1973.

[14] J. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," Plenum Press, New York, 1981.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[15] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations, In Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297, 1967.

[16] N. Mittal, M.C. Govil, R. Nayak, R. Kumar, H. Gothwal and D. Das, "A Hybrid Clustering Based Filtering Approach with Efficient Sequencing," In Proceeding of the International MultiConference of Engineering and Computer Scientists 2008 Vol I, March 2008.

[17] R. Kohavi, (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection". Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence 2 (12): 1137–1143, 1995

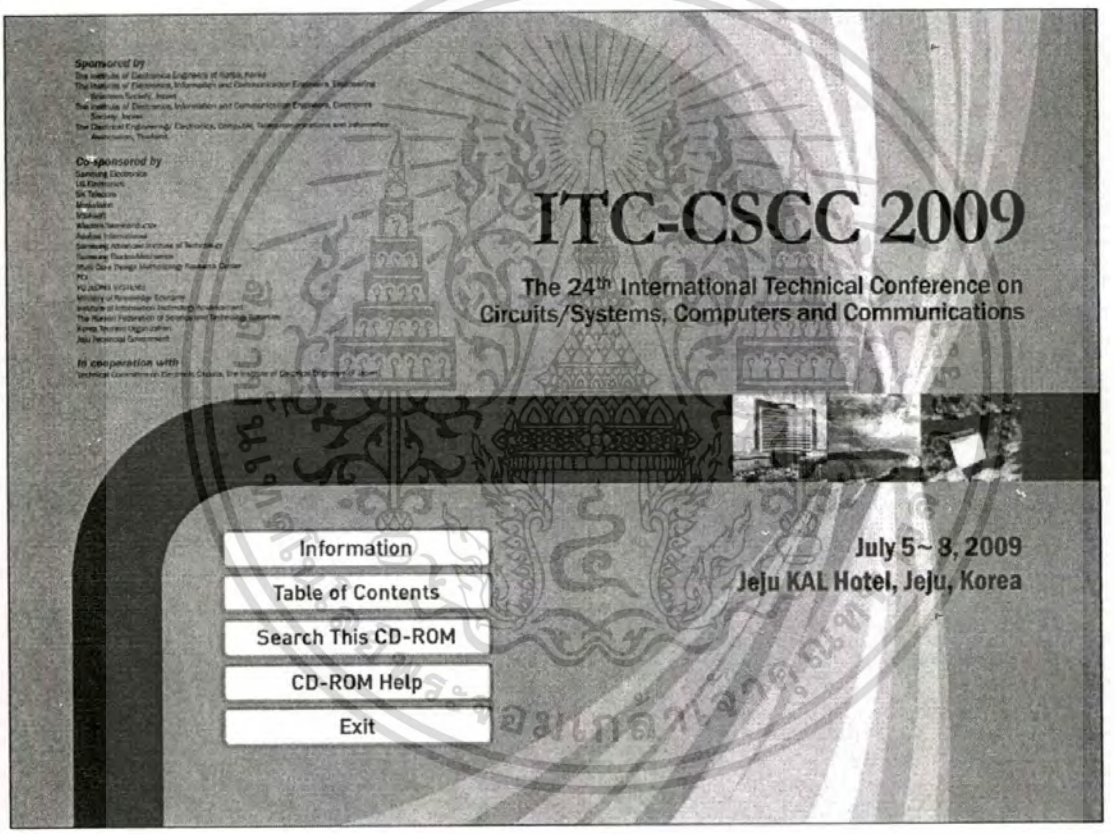
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

ผลงานวิจัยที่ได้รับการตีพิมพ์เผยแพร่

1. Panita Makprasert, Worapoj Kreesuradej, "**Fuzzy Similarity for Item-Based Top-*N* Recommendation**," The 24th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2009), pp. 883-885, Jeju, Korea, July 5-8, 2009.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Table of Contents

Plenary Talks

Plenary Talk 1

Characterization of the Best Path in Wireless Multi-hop Networks: Towards a Theoretical Framework
Prof. Masakazu Sengoku (Niigata University, Japan)

Plenary Talk 2

Start-up of Next-Generation Communication
Dr. Youngky Kim (Samsung Electronics Co. , Ltd., Korea)

Plenary Talk 3

Harmonic Detection Methods in Power Quality Issues
Sarawut Sujitjorn (Suranaree University of Technology, Thailand)

Special Sessions

Special Session 1 : Energy Technology Systems

1. The Biodiesel Production from Roast Thai Sausage Oil by Transesterification Reaction 1
Wasin Theansuwan, Kaitisak Sriprateep, Anirut Pattarateera (Mahasarakham University, Thailand)
2. Biogas Production from Waste Water in Cassava Factory 5
Worapot Khamjanla, Tanakorn Wongwuttanasatian (Khon Kaen University, Thailnad)
3. The Effect of Moisture Content on Bio-Fuel Management of Steam Generator in Sugar Plant 8
Purimpat Sujumnongtokul, Amnat Yospanya (Khon Kaen University, Thailand)
4. Increasing Venturi-assisted inlet flow of biogas into dual-fuel engine 12
Kitti Wichettapong, Tanakorn Wongwuttanasatian (Khon Kaen University, Thailand)
5. Efficiency Investigation for a Coal-Fired Circulating Fluidized Bed Boiler of Power Generation Plant 15
Purimpat Sujumnongtokul, Tanakit Chunmontean (Khon Kaen University, Thailand)

Special Session 2 : Information and Communication Networks for Safe and Secure Life

1. Achievement of a Simple Impedance Transform Structure by Setting the Feed Point to the End of a Very Small Normal-Mode Helical Antenna 22
Won Gook Hong, Naobumi Michishita, Yoshihide Yamada (National Defense Academy, Japan)
2. An Epidemic P2P Computing System and Its Application in High-Resolution Video Processing 26
Guowei Chen, Yanru Chen (Waseda University, Japan), Koji Kataoka (M-Soft Co., Ltd, Japan),
Md. Humayun Kabir, Takuro Sato (Waseda University, Japan)
3. Analysis of Mean Waiting Time for Message Delivery in Street Ad Hoc Networks 30
Keisuke Nakano (Niigata University, Japan), Akira Otsuka (Mitsubishi Electric Corporation, Japan),
Kazuyuki Miyakita, Masakazu Sengoku (Niigata University, Japan), Shoji Shinoda (Chuo University, Japan)
4. Analysis of Relaying Probability in Wireless Multi-hop Networks with QoS Routing 34
Kazuyuki Miyakita, Keisuke Nakano, Tetsuya Hasegawa, Masakazu Sengoku (Niigata University, Japan),
Shoji Shinoda (Chuo University, Japan)
5. The Shortest Step Algorithm with Network Coding on Multihop Wireless Cyclic Networks 38
Hiroshi Tamura, Yuki Shimizu (Niigata Institute of Technology, Japan), Masakazu Sengoku (Niigata University, Japan),
Shoji Shinoda (Chuo University, Japan)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. Confined-Chalcogenide Based Double-Heater Phase Change Memory with Metal Oxide Capped for Low-Current Consumption 847
Sanchai Harnsoongnoen, Chiranut Sa-ngiamsak (Khon Kaen University, Thailand)
6. A Parallel Simulated Annealing Algorithm for LSI Floorplanning Running on Multicore Processors 851
Tomoaki SATO, Masato INAGI, Shinobu NAGAYAMA, Shin'ichi WAKABAYASHI (Hiroshima City University, Japan)

S25-B5 : Graph Theory & Petri Nets

1. On some analysis problem of extended batch Petri nets 855
Atsushi Ohta, Kohkichi Tsuji (Aichi Prefectural University, Japan)
2. Two Step Framework to Extend Workflow Nets and Soundizability Problem 859
Shingo Yamaguchi, Ryo Ikeda, Minoru Tanaka, (Yamaguchi University, Japan)
3. A Method of Handwritten Character Recognition by Feature Graphs 863
Shuichi Nishida, Mitsuru Nakata, Qi-Wei Ge, Makoto Yoshimura (Yamaguchi University, Japan)
4. A Multi-Depot Vehicle Routing Algorithm with Decomposition based on Minimum Spanning Tree 867
Hiroki Narita, Morikazu Nakamura (University of the Ryukyus Okinawa, Japan)
5. Flexible Online Scheduling for Multi-Car Elevator based on Call Distribution 871
Yoshinori Miyagi, Morikazu Nakamura (University of the Ryukyus Okinawa, Japan)
6. A Proposal of Refactoring to Well-Structured Workflow Nets 875
Yuki Kuroda, Shingo Yamaguchi, Minoru Tanaka (Yamaguchi University, Japan)

S26-C5 : DB & Software

1. Hybrid-Dimension Fast Update Algorithm 879
S.Chatsettakul, W.Kreesuradej (King Mongkut's Institute of Technology Ladkrabang, Thailand)
2. Fuzzy Similarity for Item-Based Top-N Recommendation 883
Panita Makprasert, Worapoj Kreesurad (King Mongkut's Institute of Technology Ladkrabang, Thailand)
3. A Block-Based Graph Model for Effective Wrapper Maintenance 886
Juyoung Park, Joongmin Choi (Hanyang University, Korea)
4. A Simple Case Study of Program Verification with ESC/Java2 889
Hiroshi ISHIKAWA (Niigata University, Japan)
5. Web-Based Analytical Tool for Visual Depiction, Exploration and Mapping of Gross National Happiness Spatial Data 893
Sonam TSHERING, Takeo OKAZAKI (University of the Ryukyus, Japan)
6. A tool support for verifying consistency between UML diagrams by SMV 897
Shinji HARADA, Tomoyuki YOKOGAWA, Hisashi MIYAZAKI, Yoichiro SATO, Michiyoshi HAYASE
(Okayama Prefectural University, Japan)

S27-D5 : Mobile & Wireless Communications (1)

1. A convolutional interleaver for LTE based mobile satellite systems 901
Sungmoon Yeo, Jongsoo Lee, Sooyoung Kim, Sang Seob Song (Chonbuk National University, Korea), Doseob Ahn (ETRI, Korea)
2. An Improved Integer Frequency Offset Estimation Scheme for OFDM systems 905
Eu-Suk Shim, Young-Hwan You (Sejong University, Korea)
3. System Capacity of DVB-S2 by Adaptive Coding and Modulation for Ka-band Mobile Satellite channel 908
Seon Ae Kim, Heung Gyoon Ryu (Chungbuk National University, Korea),
Byung Gak Jo, Gwang Hoon Baek (Agency for Defense Development, Korea)
4. Low-Complexity Channel Estimation for Multi-Band OFDM Ultra-Wideband Communications 912
Chia-Chang Hu, Shih-Chang Lee, Ching-Wen Tang (National Chung Cheng University, Taiwan)
5. Performance Comparison of Bit Metric Derivation Method for OFDM Adaptive Array Antenna 916
Yunhan WANG, Fumiaki MAEHARA (Waseda University, Japan)

S29-F5 : RFID & WSN

1. Performance of Tag Cognizance Scheme Using Tag Separation in RFID Networks 920
Jin Kyung Park (University of California at Los Angeles, USA), Jun Ha, Seokhyun Yoon, Cheon Won Choi (Dankook University, Korea)
2. Tag Cognizance Performance of Tag Purification in RFID Networks 922
Jin Kyung Park (University of California at Los Angeles, USA), Woo Cheol Shin, Jun Ha, Cheon Won Choi (Dankook University, Korea)
3. Efficient Path Rebuilding Algorithm for Grid Based WSN 924
Kibeom Seong (Qualcomm Inc., CA), Young Sub Yoon (Hanyang University, Korea), Sung Hyun Cho (Hansei University, Korea),
Sung Han Park (Hanyang University, Korea)
4. Bayesian Cognizance of RFID Tag with Tag Separation 928
Jin Kyung Park (University of California at Los Angeles, USA), Jun Ha, Joonmo Kim, Cheon Won Choi (Dankook University, Korea)
5. Increasing Network Lifetime and Data Transfer through Node Vulnerability Aware Routing in Wireless Sensor Networks 930
Rahim Khan, Shah Nawaz Khan, Mushtaq Ahmad (Ghulam Ishaq Khan Institute of Engineering, Pakistan)
6. Performance Study of Clustering Self-Reconfiguring Algorithm for Wireless Sensor Networks 934
Dujdow Buranapanichkit, Fatima Chopngam, Thradon Wattananavin, Nattha Jindapetch (Prince of Songkla University, Thailand)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Fuzzy Similarity for Item-Based Top-*N* Recommendation

Panita Makprasert, Worapoj Kreesuradej
Faculty of Information Technology,
King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand
E-mail: mear_jomza@hotmail.com, worapoj@it.kmitl.ac.th

Abstract

This article proposes a new method to predict a list of item for active user by applying fuzzy relation and fuzzy projection for recommending a system. First, we transform transaction data to user-to-item fuzzy set and item-to-user fuzzy set. Then, we make fuzzy relation between user-to-item fuzzy set and item-to-item fuzzy set by using Lukasiewicz's implication. Compose two fuzzy relations by composite operation. Then, the composition of both fuzzy relations is used to find Top-N recommendation item for active user.

Keywords: Fuzzy Relation, Similarity, Item-based, Top-*N* Recommendation.

1. Introduction

Collaborative filtering [8] is a process for filtering information. Collaborative filtering techniques can be categorized into two approach: User-based Collaborative filtering and Model-based Collaborative filtering. User-based collaborative filtering is based on uses information from user profile. User-based collaborative filtering has shown the most successful recommend system.

However, user-based recommendation techniques had been develop for address scalability concerned with user-based recommendations. These techniques analyze use-item matrix to identify relations between various items from similarity method. Use these relationships to compute recommendations for certain users.

Over the years, many researchers have researched and developed many recommended system and used in different applications.

1. Recommending product to customer that will be interested, i.e. Movie, TV program or songs. This system includes web pages and search engines for finding information. [2,4,10-12]
2. Use useful information about demographic, content, or historical information for building recommender systems [10,12]
3. Use historical information in collaborative filtering (CF) [9] with successfully techniques and extensive use.
4. Use collaborative filtering in Tapestry system [8]. Tapestry is a filter e-mail system. Architecture of system is filter e-mail that users receive from mailing lists and newsgroups. Users can write comments about e-mail and share annotation in Tapestry.
5. Building first automatic recommender system that name GroupLens [9] which personalized recommend on Usenet posting.

   The recommendations for each user are indentified by a neighborhood of similar users and recommend articles for group of users.
6. Item-based *top-N* recommendation algorithm [5] is a technique that improved from Item-based collaborative filtering Recommendations of same author in 2001.

For developing a collaborative filtering *top-N* recommendation has 2 approaches. Firstly, cluster each user in groups of similar behaving individuals. Results from calculation are frequent of items (product, movie, book) in like/buy from information of member in group which can use for base when recommend items [1,10]. Secondly, use analyze historical information to find relations between various items like purchase items. Then use these relations for identify recommend items [10].

In model-based *top-N* Recommend have 2 key steps: (1) Method for compute the similarity between items and (2) Method for combine similarities in order to compute similarity between basket of items and candidate recommend item.

This paper presents a new method for identify item-to-item similarity and user-to-user similarity. The new method uses fuzzy relation for finding composite

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

relation between user-to-item relation and item-to-item relation. Then, project active user is projected on composite relation for predict recommended item.

2. Preliminaries

2.1 Fuzzy set [7] is mathematically represent uncertainly and vagueness. Fuzzy set theory permits the gradual assessment of the membership of elements in a set.

Definition 1: Let X is not empty set. Fuzzy set A is

$$\mu_{\tilde{A}}(x) : X \rightarrow [0,1] \quad (2\text{-}1)$$

$\mu_{\tilde{A}}(x)$ is member value of x in fuzzy set $\tilde{A}$

for each $x \in X$ fuzzy set $\tilde{A}$ can show in tuple set

$$\tilde{A} = \{x, \mu_A(x) \mid x \in X\} \quad (2\text{-}2)$$

2.2 Fuzzy relation R from X to Y is relation fuzzy set in X*Y. Denoted by R(x,y); $x \in X$ and $y \in Y$

A binary fuzzy relation is special fuzzy relation between two sets X and Y, denoted by R(X,Y). R(X,Y) is referred to as bipartite graphs. When X=Y, R is a binary fuzzy relation on a single or digraph, denoted by R(x,x) or R(X^2)

Let X = $\{x_1, x_2, \ldots, x_n\}$ and Y = $\{y_1, y_2, \ldots, y_m\}$ fuzzy relation. R(X,Y) can be shown in n x m matrix as

$$R(X,Y) = \begin{bmatrix} \mu_R(x_1,y_1) & \mu_R(x_1,y_2) & \cdots & \mu_R(x_1,y_m) \\ \mu_R(x_2,y_1) & \mu_R(x_2,y_2) & \cdots & \mu_R(x_2,y_m) \\ \vdots & \vdots & & \vdots \\ \mu_R(x_n,y_1) & \mu_R(x_n,y_2) & \cdots & \mu_R(x_n,y_m) \end{bmatrix}$$

2.3 Fuzzy implication is a technique for finding fuzzy relation by mapping membership function of relation R to Cartesian product between 2 fuzzy sets. the value of membership function is between 0 and 1

2.4 Composition operations

Let R(X,Y) and S(Y,Z) denoted as R(X,Y) ∘ S(Y,Z) "∘" mean max-min operation is defined by

$$\mu_{R \circ S}(x,z) \triangleq \max_{y \in Y} \min[\mu_R(x,y), \mu_S(y,z)] \quad \forall \ x \in X, z \in Z \quad (2\text{-}3)$$

2.5 Projection operation on fuzzy relation

Let a fuzzy relation R(X,Y). [R↓Y] denote the projection of R onto Y. Then, [R↓Y] is a fuzzy relation in Y whose membership function is defined by $\mu_{[R \downarrow Y]}(y) = \max_x \mu_R(x,y)$ This concept can be extend to n-ary relation $R(\{x_1, x_2, \ldots, x_n\})$

3. Fuzzy Similarity for Item-based Top-N Recommendation Algorithms

From original Item-Based Top-N Recommendation algorithms [5] analyzing historical purchasing information can know about user (customer) that is interested to purchase similar items to the items that user has already purchase represent in user-item matrix

Using similarity algorithms are similar to previously developed item-based schemes [10-11]. However, the proposed method is different in a number of key aspects related to how the similarity between the different items is computed.

Fuzzy Similarity for Item-Based Top-N Recommendation Algorithms uses information of user and items to identify fuzzy relations.

3.1 Fuzzy set to fuzzy relation

Major data of these transactions are buying product, buying quantity, transaction date/time and month. For the Top-N recommender system, we only use data about customer, product and buying quantity.

User	Item						
	I_1	I_2	I_3	I_4	I_5	....	I_n
U_1	x	x	x	x	x	....	x
U_2	x	x	x	x	x	....	x
U_n	x	x	x	x	x	....	x

Table 1 : User – Item Transaction

From Table 1, U represent each user or customer who has transactions with the system. we represent each product in the system. The number in the table is a quality of brought product by that customer. Customer data are made to fuzzy set U and product data are made to fuzzy set I. Then, U and I are taken to fuzzy relation by Lukasiewicz's Implication(3-1) for getting relation mapping value from U to I denoted by $R = \tilde{U} \times \tilde{I}$

$$\mu_R(x,y) = \min[1, (1 - \mu_U(x) + \mu_I(y))] \quad (3\text{-}1)$$

As an example,

$\tilde{U}$ = [0.7 0.1] $\tilde{I}$ = [0.9 0.2 0.8]

$R = \tilde{U} \times \tilde{I}$ = [0.6 0.1 0.5 0.0 0.7 0.1]

Let, fuzzy relation S be a binary fuzzy relation of fuzzy set of items denoted by $S = \tilde{I} \times \tilde{I}$ .When getting two fuzzy relations, fuzzy relation of user-item relation and fuzzy relation of item-item relation, compose these relations by composition operation(3-2) for finding the relation between user and recommended item.

$$\mu_{R \circ S}(u,i) = \max_{j \in I} \min[\mu_R(u,i), \mu_T(i,i)] \quad (3\text{-}2)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

When we recommend item, we use projection on to active user in composite relation $T = R \circ S$ for reach item group that relate to user and provide top-N items to user.

4. Conclusion

In this work, we have proposed fuzzy relation for creating relation between user-item relation and item-item relation. In doing so, we have transformed transaction that have information of user and item into simple structure type then create fuzzy sets. With a fuzzy set, we create fuzzy relation using Lukasiewicz's implication method.

As a result, fuzzy relation of user-item and item-item can be used for predicting which product that recommender system should provide to user next time. In the future, further researches and experiments on the proposed algorithm will be presented.

References

[1] B. Sarwar, G.Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms", *In WWW10*. 2001.

[2] B. Mobasher, R. Cooley, and J. Srivastava "Automatic Personalization based on web usage mining", *Commun ACM*, pp.142-151. 2000.

[3] G. Karypis, "Experimental evaluation of item-based top-n recommendation algorithms", *Proceedings of the ACM Conference on Information and Knowledge Management*, 2001.

[4] J. Konstan, B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying collaborative filtering to Usenet news", *Commun ACM*, pp77-87. 1997.

[5] M. Deshpande and G. Karypis, "Item-Based Top-N Recommendation Algorithm", *ACM Transactions on Information Systems*, pp. 143-177, January 2004.

[6] L. Chin-Teng and C.S. L. George, "A Neuro-Fuzzy Synergism to Intelligent Systems" *Neural Fuzzy System*, Prentice Hall PTR, pp. 10-62. 1996.

[7] L.A. Zadeh, "Fuzzy sets" *Information and Control*, pp 338-353. 1965.

[8] D. Goldberg, D. Nichols, B. Oki, and D. Terry. "Using collaborative filtering to weave an information tapestry", *Commun ACM*, pp.61-72. 1992.

[9] P. Resnick, N. Iacovou, M. Suchak, P.Bergstrom, and J. Riedl. "GroupLens: An open architecture for collaborative filtering of netnews", *Proceeding of CSCW*, 1994.

[10] U. Shardanand, and P. Maes. "Social information filtering: Algorithms for automating", *Proceeding of ACM CHI'95 Conference on Human Factors in Computing System*, pp 210-217, 1995.

[11] B. Kitts, D. Freed, and M. Vrieze. "Cross-sell: A fast promotion-tunable customer-item recommendation method based on conditional independent probabilities", *Proceeding of ACM SIGKDD International Conference*. pp.437-446. 2000.

[12] W. Hill, L. Stead, M. Rosenstein, and D. Terry. "Recommending and evaluating choices in a virtual community of use", *Proceeding of ACM CHI*,1995.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

นางสาวปนิตา มักประเสริฐ สำเร็จการศึกษาระดับปริญญาตรีวิทยาศาตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ จากคณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา ในปีการศึกษา 2547 จากนั้นเข้าศึกษาต่อในระดับปริญญาโท หลักสูตรวิทยาศาสตรมหาบัณฑิต (วท.ม) สาขาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ในปีการศึกษา 2548

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้