

ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล.

ระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์

PRODUCT AND SERVICE RECOMMENDER SYSTEM IN
E-COMMERCE



H003455



๒๕๕๑๑๔๖
11/1/2549

วัน เดือน ปี 04 S.A. 2550
เลขทะเบียน..... H003455
เลขเรียกหนังสือ... อพ. ศี 684 ธ 2549
"ห้องสมุดคณะเทคโนโลยีสารสนเทศ สจล."

รายงานนี้เป็นส่วนหนึ่งของวิชาโครงการพัฒนาระบบงาน
หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ
คณะเทคโนโลยีสารสนเทศ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ภาคเรียนที่ 2 ปีการศึกษา 2549
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปะสิ่งนี้ที่ และต้องอ้างอิงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

**PRODUCT AND SERVICE RECOMMENDER SYSTEM IN
E-COMMERCE**



**A SYSTEM DEVELOPMENT PROJECT
OF THE REQUIREMENT FOR THE DEGREE OF
MASTER OF SCIENCE PROGRAM IN INFORMATION TECHNOLOGY
FACULTY OF INFORMATION TECNOLOGY
KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG**

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
2/ 2006
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



COPYRIGHT 2007

FACULTY OF INFORMATION TECHNOLOGY

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อใช้ในการศึกษาเท่านั้น การนำเอกสารนี้ไปใช้โดยไม่ได้รับอนุญาตถือว่าผิดกฎหมาย
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อ	ระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์
นักศึกษา	นายศุภชัย อัสวชินเทพกุล
รหัสนักศึกษา	46066708
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	เทคโนโลยีสารสนเทศ
แขนงวิชา	วิทยาการสารสนเทศ
ปีการศึกษา	2549
อาจารย์ที่ปรึกษา	รศ.ดร.วรพจน์ กรีสระเดช

บทคัดย่อ

ธุรกิจสมัยใหม่ได้นำระบบสารสนเทศมาช่วยในการดำเนินธุรกิจให้มีประสิทธิภาพ โดยธุรกิจ E-commerce เป็นธุรกิจรูปแบบหนึ่งบนเครือข่ายอินเทอร์เน็ตที่ได้รับความนิยมมากในปัจจุบัน ซึ่งเพิ่มความสะดวกสบายให้กับลูกค้าในการเลือกซื้อสินค้าและบริการผ่านอินเทอร์เน็ต และสิ่งที่ทำให้ธุรกิจการซื้อขายสินค้าและบริการผ่านอินเทอร์เน็ตนั้นเป็นที่น่าสนใจ คือ การนำเสนอและแนะนำทางเลือกสินค้าและบริการแก่ผู้ซื้อ ดังนั้นจึงมีการพัฒนาระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์ ซึ่งระบบจะแนะนำสินค้าและบริการที่เกี่ยวข้อง ที่มีความสัมพันธ์กันกับผู้ซื้อสินค้าและบริการ หรือสินค้าและบริการที่มักจะถูกซื้อควบคู่กัน จะช่วยให้ผู้ซื้อลดขั้นตอนและเวลาในการค้นหาสินค้าและบริการที่ต้องการ ทำให้ผู้ซื้อได้รับความสะดวกสบายเพิ่มมากขึ้น ในโครงการนี้จะนำเสนอการทำงานของระบบแนะนำสินค้าและบริการ โดยใช้เทคนิค Collaborative filtering

Title	Product and Service Recommender System in E-Commerce
Student	Mr. Supachai Asawachintheikul
Student ID.	46066708
Degree	Master of Science
Programme	Information Science
Academic Year	2006
Advisor	Assoc. Prof. Dr. Worapoj Kreesuradej

ABSTRACT

There is a use of IT system to help improve modern business efficiency. At the moment , E-commerce is one of the most popular way to do business. IT helps customers to get more comfortable way of buying products and services on the internet. The things which make doing business on the internet network interesting are Product presentation and also various choices for customers to select their needed products and services so here comes. The Product and Services Recommendation System. The system will recommend the related products and services to the customers by helping customers reducing the process and time to search for their wanted product and services which resulted in more comfortable for the customers. In this project present recommender system based on Collaborative Filtering.

กิตติกรรมประกาศ

โครงการพัฒนาระบบงานนี้สำเร็จได้อย่างดี ด้วยคำแนะนำ และคำปรึกษาจาก รศ.ดร. วรพจน์ กรีสุระเดช ซึ่งเป็นอาจารย์ที่ปรึกษาโครงการพัฒนาระบบงาน ข้าพเจ้ารู้สึกทราบบซึ่งในความอนุเคราะห์จากท่านอาจารย์ และขอขอบพระคุณเป็นอย่างสูง

ขอกราบพระคุณคณาจารย์ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุก ๆ ท่านที่ได้ประสิทธิ์ประสาทวิชาให้กับข้าพเจ้า ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ในภาควิชาเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยี พระจอมเกล้าเจ้าคุณทหารลาดกระบัง ทุกคนที่ให้คำแนะนำต่างๆ และคอยให้กำลังใจเสมอมา

ขอขอบคุณบัณฑิตศึกษาและบัณฑิตวิทยาลัย คณะเทคโนโลยีสารสนเทศ ที่ให้ความช่วยเหลือ ในเรื่องต่างๆ

สุดท้ายนี้ข้าพเจ้าขอกราบขอบพระคุณ บิดา มารดา และครอบครัวของข้าพเจ้าที่เป็นกำลังใจ และให้การสนับสนุนในทุกเรื่องๆ ทำให้ข้าพเจ้าสามารถทำโครงการพัฒนาระบบฉบับนี้สำเร็จลุล่วง ด้วยดี

คุณค่าและประโยชน์อันพึงมาจาก โครงการพัฒนาระบบฉบับนี้ ข้าพเจ้าขอบอบแด่ผู้มีพระคุณทุกท่าน

ศุภชัย อัสวชินเทพกุล

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
กิตติกรรมประกาศ.....	III
สารบัญ.....	IV
สารบัญตาราง.....	VII
สารบัญรูป.....	VIII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาของปัญหา.....	1
1.2 วัตถุประสงค์ของการพัฒนาระบบงาน.....	2
1.3 ขอบเขตของการพัฒนาระบบงาน.....	2
1.4 ขั้นตอนการดำเนินการพัฒนาระบบงาน.....	2
1.5 ประโยชน์ที่คาดว่าจะได้รับการพัฒนาระบบงาน.....	3
บทที่ 2 ความรู้เบื้องต้นเกี่ยวกับเว็บ ไมนิ่งและระบบแนะนำสินค้าและบริการในธุรกิจ e-commerce.....	4
2.1 บทนำ.....	4
2.2 การประยุกต์ใช้เว็บ ไมนิ่งเพื่อทำให้บริการเว็บไซต์.....	4
2.3 การใช้เว็บ ไมนิ่งทางด้านการตลาดของธุรกิจ e-commerce.....	5
2.4 เทคนิคในการนำเว็บ ไมนิ่งมาทำการวิเคราะห์ Web log records	6
2.5 ข้อดีของการใช้เว็บ ไมนิ่ง.....	8
2.5.1 ทำให้สามารถจัดสรรทรัพยากรที่มีให้ตรงกับสิ่งที่ผู้เยี่ยมชมสนใจ.....	8
2.5.2 เพิ่มคุณค่าของผู้เยี่ยมชมแต่ละคน	8
2.5.3 ทำให้ประสบการณ์ของผู้เยี่ยมชมเว็บไซต์ดีขึ้น.....	9
2.5.4 จัดสรรทรัพยากรให้กับกลุ่มเป้าหมายได้อย่างถูกต้อง.....	9
2.5.5 เป็นแนวทางในการจัดเก็บข้อมูลแบบใหม่.....	9
2.5.6 สามารถทดสอบความสัมพันธ์ของเนื้อหากับ โครงสร้างของเว็บไซต์ได้.....	10
2.6 ระบบแนะนำสินค้าและบริการในธุรกิจ e-commerce.....	10
2.7 การทำงานของระบบแนะนำสินค้า.....	12

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

สารบัญ (ต่อ)

	หน้า
2.7.1 ระบบแนะนำสินค้าช่วยธุรกิจ e-commerce ได้อย่างไร.....	12
2.7.2 การทำงานของระบบแนะนำสินค้า	12
2.7.3 เทคนิคในการสร้างการแนะนำสินค้า (Recommendation Method)	13
2.8 Recommender Systems Based on Collaborative Filtering	14
2.9 Selecting Relevant Instances	15
2.10 Metrics and Methodology	16
บทที่ 3 การออกแบบระบบแนะนำภาพยนตร์.....	18
3.1 Use Case Diagram.....	18
3.2 Activity Diagram	25
3.3 Class Diagram.....	29
3.4 การเตรียมข้อมูล.....	30
3.5 การออกแบบฐานข้อมูล.....	30
บทที่ 4 การประยุกต์เทคนิค Collaborative filtering กับระบบแนะนำภาพยนตร์.....	34
4.1 กำหนดวัตถุประสงค์.....	34
4.2 เครื่องมือที่ใช้ในการพัฒนาระบบ.....	34
4.3 ขั้นตอนและรายละเอียดการใช้งาน.....	35
4.3.1 ขั้นตอนและรายละเอียดการใช้งานในส่วนของผู้ดูแลระบบการแนะนำ ภาพยนตร์.....	35
4.3.2 ขั้นตอนและรายละเอียดการใช้งานในส่วนของผู้ใช้.....	43
บทที่ 5 สรุปผลการศึกษาและข้อเสนอแนะ.....	56
บรรณานุกรม.....	58
ภาคผนวก.....	59
ภาคผนวก ก. ผลงานวิจัยที่เกี่ยวข้องกับการพัฒนาโครงการ.....	60

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

สารบัญ (ต่อ)

หน้า

ประวัติผู้เขียน.....67



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่	หน้า
3.1 รายละเอียด Use Case Diagram ของ Login.....	20
3.2 รายละเอียด Use Case Diagram ของ Register.....	21
3.3 รายละเอียด Use Case Diagram ของ Vote.....	21
3.4 รายละเอียด Use Case Diagram ของ Edit Vote.....	22
3.5 รายละเอียด Use Case Diagram ของ Edit Profile..	22
3.6 รายละเอียด Use Case Diagram ของ Recommend Movie.	23
3.7 รายละเอียด Use Case Diagram ของ Select Target.....	23
3.8 รายละเอียด Use Case Diagram ของ Training Data.....	24
3.9 รายละเอียด Use Case Diagram ของ Evaluation.....	24
3.10 ตาราง tbl_Rating ข้อมูลรายละเอียดของ Entity Rate.....	32
3.11 ตาราง tbl_User ข้อมูลรายละเอียดของ Entity User.....	32
3.12 ตาราง tbl_Movie ข้อมูลรายละเอียดของ Entity Movie.....	32
3.13 ตาราง tbl_Movietype ข้อมูลรายละเอียดของ Entity Movie Type.....	33
3.14 ตาราง tbl_TypeOfMovie ข้อมูลรายละเอียดของ Entity Type of Movie.....	33
3.15 ตาราง tbl_Training ข้อมูลรายละเอียดของ Entity Data Training Set.....	33
3.16 ตาราง tbl_TopRelevant ข้อมูลรายละเอียดของ Entity User Relevant.....	33

สารบัญรูป

รูปที่	หน้า
2.1 แสดงการทำงานของระบบแนะนำสินค้า.....	12
3.1 Use Case Diagram ระบบแนะนำภาพยนตร์.....	19
3.2 Activity Diagram ของ Use Case UC1.....	25
3.3 Activity Diagram ของ Use Case UC2.....	25
3.4 Activity Diagram ของ Use Case UC3.....	26
3.5 Activity Diagram ของ Use Case UC4.....	26
3.6 Activity Diagram ของ Use Case UC5.....	27
3.7 Activity Diagram ของ Use Case UC6.....	27
3.8 Activity Diagram ของ Use Case UC7.....	28
3.9 Activity Diagram ของ Use Case UC8.....	28
3.10 Activity Diagram ของ Use Case UC9.....	29
3.11 Class Diagram ของระบบแนะนำภาพยนตร์.....	30
3.12 แสดงความสัมพันธ์ระหว่างข้อมูลเชิงสัมพันธ์ของระบบแนะนำภาพยนตร์.....	31
4.1 แสดงหน้าจอหลักของระบบการแนะนำภาพยนตร์.....	35
4.2 แสดงหน้าจอของการติดต่อฐานข้อมูลของระบบการแนะนำภาพยนตร์.....	36
4.3 แสดงหน้าจอรายละเอียดข้อมูลการให้คะแนนความชอบของผู้เข้าเยี่ยมชม.....	37
4.4 แสดงหน้าจอรายละเอียดการเลือกกลุ่มของภาพยนตร์ที่จะนำมาแนะนำให้ผู้เยี่ยมชม.....	38
4.5 แสดงหน้าจอรายละเอียดการ Training Data ทาค่า Relevant ระหว่างภาพยนตร์กับลูกค้า.....	39
4.6 แสดงหน้าจอรายละเอียดค่า Relevant ระหว่างภาพยนตร์กับลูกค้าแต่ละคน.....	40
4.7 แสดงหน้าจอรายละเอียดการวัดประสิทธิภาพของระบบแนะนำภาพยนตร์.....	41
4.8 แสดงหน้าจอรายละเอียดการบันทึกการสร้างการแนะนำภาพยนตร์ที่จะแนะนำให้ผู้เข้ามา เยี่ยมชม.....	42
4.9 แสดงหน้าจอการ Login เข้าสู่ระบบการแนะนำภาพยนตร์.....	43
4.10 แสดงหน้าจอลงทะเบียนสำหรับสมาชิกใหม่.....	44
4.11 แสดงหน้าจอการลงคะแนนความชอบ (Rating) ของผู้เข้ามาเยี่ยมชม.....	45
4.12 แสดงหน้าจอการแนะนำภาพยนตร์ของผู้เข้ามาเยี่ยมชม.....	46
4.13 แสดงหน้าจอการเพิ่มการให้คะแนนความชอบของผู้เข้ามาเยี่ยมชม.....	47
4.14 แสดงหน้าจอการแก้ไขให้คะแนนความชอบของผู้เข้ามาเยี่ยมชม.....	48

สารบัญรูป (ต่อ)

รูปที่	หน้า
4.15 แสดงหน้าจอการแก้ไขข้อมูลของผู้เข้ามาเยี่ยมชม.....	49
4.16 แสดงหน้าจอข้อมูลภาพยนตร์ที่เพิ่งออกฉายขณะนี้.....	50
4.17 แสดงหน้าจอข้อมูลภาพยนตร์ที่เป็นที่นิยมในขณะนี้.....	51
4.18 แสดงหน้าจอการเลือกข้อมูลภาพยนตร์ตามประเภทของภาพยนตร์.....	52
4.19 แสดงหน้าจอเลือกข้อมูลภาพยนตร์ตามประเภทของภาพยนตร์.....	53
4.20 แสดงหน้าจอการเลือกข้อมูลภาพยนตร์ตามปีที่ภาพยนตร์นั้นออกฉาย.....	54
4.21 แสดงหน้าจอข้อมูลของภาพยนตร์ตามปีที่ภาพยนตร์นั้นออกฉาย.....	55



บทที่ 1

บทนำ

1.1 ความเป็นมาของปัญหา

ในปัจจุบันระบบอินเทอร์เน็ตได้มีการใช้กันอย่างแพร่หลาย และมีการเจริญเติบโตขึ้นอย่างรวดเร็วภายในระยะเวลาไม่กี่ปีที่ผ่านมา โดยในแต่ละบริษัทและธุรกิจต่างๆ ต่างก็มีเว็บไซต์เป็นของตัวเองเพื่อช่วยอำนวยความสะดวกรวดเร็วในการดำเนินธุรกิจ โดยได้มีการสร้างธุรกิจในรูปแบบใหม่ซึ่งเป็นการซื้อขายสินค้าและบริการผ่านทางอินเทอร์เน็ตเป็นธุรกิจหนึ่งที่ได้รับคามนิยมเป็นอย่างมากในปัจจุบันคือ “พาณิชย์อิเล็กทรอนิกส์” (e-commerce)

ในปัจจุบันการทำธุรกิจ e-commerce ได้เกิดขึ้นอย่างมากมาและมีการแข่งขันกันสูงเนื่องจากธุรกิจประเภทนี้เป็นธุรกิจที่ทำการซื้อขายสินค้าและบริการผ่านอินเทอร์เน็ต ซึ่งเป็นการเพิ่มความสะดวกให้กับผู้ซื้อเป็นอย่างมาก จึงทำให้เกิดการศึกษาและการคิดหาหนทางใหม่ๆ ในการบริการลูกค้าให้ลูกค้าเกิดความพึงพอใจในการใช้บริการซื้อสินค้าและบริการผ่านเว็บไซต์ โดยหนทางหนึ่งที่จะทำให้ธุรกิจการซื้อสินค้าและบริการเป็นที่น่าสนใจก็คือ การนำเสนอและแนะนำทางเลือกสินค้าให้แก่ผู้ซื้อซึ่งจัดเป็นบริการเสริมอย่างหนึ่งที่จะช่วยให้ผู้ซื้อสินค้าและบริการลดขั้นตอนและเวลาในการค้นหา และยังช่วยอำนวยความสะดวกในการเลือกซื้อสินค้าและบริการของผู้ซื้อได้อย่างมาก ซึ่งในปัจจุบันได้มีสินค้าและบริการให้เลือกอย่างมากมาจากการที่ผู้ซื้อจะตัดสินใจเลือกซื้อสินค้าและบริการต้องอาศัยคำแนะนำจากแหล่งอื่นๆ เช่น สื่อโฆษณาต่างๆ คำแนะนำรายชื่อที่ขายดีภายในร้าน ดังนั้นจึงได้มีการพัฒนาระบบที่เรียกว่า “ระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์” ขึ้นมา ซึ่งระบบจะช่วยแนะนำสินค้าและบริการที่เกี่ยวข้อง มีความสัมพันธ์กันกับสินค้าที่ผู้ซื้อได้เลือกซื้อไปหรือสินค้าที่มักจะถูกซื้อควบคู่กันไปหรือแนะนำจากกลุ่มผู้ที่มีลักษณะพฤติกรรมในการซื้อสินค้าที่เหมือนกันหรือมีความชอบแบบเดียวกัน โดยที่ระบบแนะนำสินค้าและบริการนั้นจะช่วยให้ผู้ซื้อลดขั้นตอนและเวลาในการซื้อสินค้าเป็นอย่างมาก ทำให้ผู้ซื้อได้รับความสะดวกเพิ่มมากขึ้นในการเลือกซื้อสินค้าผ่านเครือข่ายอินเทอร์เน็ตในธุรกิจ e-commerce

โดยที่ในโครงการพัฒนาระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์ได้ทำการศึกษาและพัฒนาระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์ขึ้นมา โดยใช้เทคนิค Collaborative filtering ซึ่งเป็นเทคนิคหนึ่งที่ได้รับคามนิยมในระบบแนะนำสินค้าในปัจจุบันนี้ ซึ่งจะเป็นการสร้างสารสนเทศสำหรับลูกค้าที่จะช่วยให้พวกเขาสามารถค้นหาและเลือกซื้อสินค้าและบริการตามที่พวกเขาสนใจ ซึ่งการแนะนำสินค้าเป็นการสร้างการบริการสำหรับลูกค้า

แต่ละคน โดยที่เว็บไซต์ของธุรกิจสามารถแนะนำ และนำเสนอสินค้าที่คาดว่าจะตรงความต้องการ
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ของลูกค้าแต่ละคนให้มากที่สุด ซึ่งจะช่วยสร้างโอกาสในการขายสินค้าและบริการของรัฐกิจ โดยที่จะเป็นวิธีการหนึ่งที่จะช่วยให้ผู้ซื้อเกิดความสะดวกสบายในการเลือกซื้อสินค้าและบริการทำให้ลูกค้าเกิดความพึงพอใจและเกิดความจงรักภักดีต่อรัฐกิจ ซึ่งจะนำมาซึ่งผลกำไรและความสำเร็จในรัฐกิจพาณิชย์อิเล็กทรอนิกส์ที่มีการแข่งขันกันสูงในปัจจุบัน

1.2 วัตถุประสงค์ของการพัฒนาระบบงาน

การศึกษาโครงการพัฒนาระบบแนะนำสินค้าและบริการในรัฐกิจพาณิชย์อิเล็กทรอนิกส์นี้มีวัตถุประสงค์ดังนี้

1. ทำการศึกษาความรู้เบื้องต้นของ Web mining ซึ่งเป็นการนำเทคนิคของ Data mining มาประยุกต์ใช้กับข้อมูลบนเว็บไซต์
2. ศึกษาหลักการทำงานของระบบแนะนำสินค้าในรัฐกิจ e-commerce
3. ศึกษาเทคนิคในการสร้างการแนะนำสินค้าของ ระบบแนะนำสินค้าและบริการในรัฐกิจพาณิชย์อิเล็กทรอนิกส์
4. นำเทคนิคการแนะนำสินค้าที่ได้จากการศึกษา มาใช้กับการซื้อขายสินค้าบนอินเทอร์เน็ตในรัฐกิจ e-commerce

1.3 ขอบเขตของการพัฒนาระบบงาน

ในการศึกษาโครงการพัฒนาระบบแนะนำสินค้าและบริการในรัฐกิจพาณิชย์อิเล็กทรอนิกส์นี้ได้กำหนดขอบเขตการศึกษาไว้ดังนี้

1. การวิเคราะห์และออกแบบระบบการแนะนำสินค้าและบริการในรัฐกิจพาณิชย์อิเล็กทรอนิกส์ เลือกใช้สำหรับการพัฒนาเว็บไซต์ ของการขายภาพยนตร์
2. การศึกษาเทคนิคในการสร้าง การแนะนำสินค้าและบริการในรัฐกิจพาณิชย์อิเล็กทรอนิกส์ นั้นจะทำการศึกษาเฉพาะเทคนิคที่เรียกว่า Collaborative filtering
3. ทำการสร้างเว็บไซต์สำหรับการแนะนำภาพยนตร์ เพื่อให้บริการแนะนำภาพยนตร์โดยมีพื้นฐานมาจากการพิจารณาพฤติกรรมทำให้คะแนนความชอบของลูกค้า ซึ่งการนำไปสู่การแนะนำสินค้าที่คาดว่าจะตรงตามความต้องการของลูกค้าไปให้กับลูกค้าได้

1.4 ขั้นตอนการดำเนินการพัฒนาระบบงาน

การทำการศึกษาแนะนำสินค้าและบริการในรัฐกิจพาณิชย์อิเล็กทรอนิกส์นี้มีขั้นตอนในการดำเนินการ ดังนี้

1. ศึกษาหลักการทำงานของระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์ จากตัวอย่างของเว็บไซต์ อย่างเช่น เว็บไซต์ Amazon.com , Cdnw.com , Movieslen.com เป็นต้น
2. ศึกษาเทคนิคที่ใช้กับระบบแนะนำสินค้าในแต่ละวิธี
3. คัดเลือกเทคนิคที่จะมาใช้กับระบบแนะนำสินค้าและบริการของโครงการนี้ ซึ่งในโครงการฉบับนี้ได้เลือกเทคนิค Collaborative filtering มาใช้กับระบบแนะนำสินค้าและบริการ
4. นำเทคนิคที่จากการศึกษามาประยุกต์ใช้กับข้อมูลของลูกค้าและข้อมูลของการให้คะแนนความชอบในรูปภาพยนตร์ของลูกค้าแต่ละคน
5. สร้างการแนะนำสินค้าให้กับลูกค้าแต่ละคน โดยมีพื้นฐานมาจากกลุ่มลูกค้าที่มีพฤติกรรม การซื้อสินค้าที่เหมือนกัน หรือมีความชอบในตัวสินค้าที่เหมือนกัน
6. ประเมินผลและวิเคราะห์ผลที่ได้รับ
7. สรุปผลจากการศึกษา

1.5 ประโยชน์ที่คาดว่าจะได้รับจากการพัฒนาระบบงาน

ประโยชน์ที่คาดว่าจะได้รับจากการศึกษาโครงการ มีดังนี้

1. สามารถเข้าใจหลักการทำงานของระบบแนะนำสินค้าที่ใช้ในปัจจุบัน
2. สามารถเข้าใจเทคนิคการสร้างคำแนะนำของระบบแนะนำสินค้าได้
3. สามารถนำเทคนิคการแนะนำสินค้าที่ได้จากการศึกษามาใช้กับการซื้อขายสินค้าและบริการบนอินเทอร์เน็ต หรือที่เรียกว่า e-commerce ได้อย่างมีประสิทธิภาพ
4. สามารถนำเอาแนวคิดและทฤษฎีต่างๆ ที่ได้จากการศึกษาในการทำโครงการไปประยุกต์ใช้เพื่อเป็นประโยชน์ในการพัฒนาระบบงานจริงต่อไป
5. ได้เรียนรู้และแนวทางในการแก้ไขปัญหาต่างๆ ที่เกิดขึ้นจากการพัฒนาโครงการ เพื่อเป็นประโยชน์ในการทำงานและการจัดการงานต่างๆ ในอนาคตต่อไป

บทที่ 2

ความรู้เบื้องต้นเกี่ยวกับเว็บไมนิ่งและระบบแนะนำสินค้าและบริการในธุรกิจ e-commerce

2.1 บทนำ

การค้าไมนิ่งเป็นกระบวนการสืบค้นข้อมูลสำคัญที่เป็นประโยชน์ต่อการดำเนินงานของธุรกิจออกจากกองข้อมูลขนาดมหาศาลที่จัดเก็บอยู่ภายในแหล่งฐานข้อมูลของแต่ละองค์กร และนำข้อมูลที่ได้นั้นมาช่วยในการตัดสินใจและการดำเนินธุรกิจของบริษัท ทำให้ข้อมูลที่มีอยู่นั้นกลายมาเป็นข้อมูลที่มีค่ามาได้

โดยที่การขุดค้นข้อมูลด้วยเทคโนโลยีการค้าไมนิ่ง ประกอบไปด้วยขั้นตอนเทคโนโลยีอันสลับซับซ้อนมากมาย ไม่ว่าจะเป็นการเรียนรู้จดจำแบบแผนมาตรฐานของข้อมูลเฉพาะกลุ่ม (Pattern recognition technology) การใช้ประโยชน์จากอัลกอริทึม ซึ่งออกแบบมาเพื่อให้เครื่องคอมพิวเตอร์สามารถเรียนรู้และพัฒนาตัวมันเองได้ (Machine learning & Genetic algorithms) หรือการเชื่อมโยงแลกเปลี่ยนข้อมูลภายในเครือข่ายเน็ตเวิร์คแบบโพลีโนเมียลซึ่งเลียนแบบการทำงานของระบบประสาทของมนุษย์ (Netural & Polynomial networks) เป็นต้น

อย่างไรก็ตามไม่ว่ามีการใช้เทคโนโลยีทางคอมพิวเตอร์ที่สลับซับซ้อนเพียงไร เป้าหมายสำคัญของการค้าไมนิ่ง ก็คือทำอะไรถึงจะเอาชนะคู่แข่งทางธุรกิจได้ ทำอย่างไรถึงจะรู้ได้ว่าประชากรกลุ่มไหนคือ กลุ่มลูกค้าเป้าหมาย (Who) และประชากรกลุ่มนี้ต้องการซื้ออะไร (What) ซื้อที่ไหน (Where) ซื้อเท่าไร (When) และที่สำคัญที่สุดก็คือ อะไรคือแรงขับเคลื่อนให้ลูกค้าเป้าหมายต้องการสินค้านั้น (Why) ซึ่งคำถามหลักๆเหล่านี้ เมื่อมาผนวกเข้ากับรูปแบบการดำเนินธุรกิจยุคดิจิทัลผ่านเครือข่ายอินเทอร์เน็ตที่เรียกว่า การทำธุรกิจ e-commerce ก็ส่งผลให้มีการนำเทคนิคของ “เว็บ ไมนิ่ง” ซึ่งเป็นลักษณะเฉพาะอย่างหนึ่งของการนำหลักการการค้าไมนิ่งมาใช้กับข้อมูลบนอินเทอร์เน็ต เข้ามาประยุกต์ใช้กับการตลาด รวมถึงการกำหนดกลยุทธ์ทางการตลาดของธุรกิจประเภท e-commerce เพื่อให้เข้าถึงกลุ่มลูกค้าได้ดี และสามารถเอาชนะคู่แข่งในธุรกิจ e-commerce โดยที่ในบทที่ 2 นี้จะนำเสนอความรู้เบื้องต้นเกี่ยวกับเว็บ ไมนิ่ง การประยุกต์ใช้เว็บ ไมนิ่งเพื่อการให้บริการเว็บไซต์ และระบบแนะนำสินค้าและบริการ ในธุรกิจ e-commerce

2.2 การประยุกต์เว็บไมนิ่งเพื่อการให้บริการเว็บไซต์

ปัจจุบันได้มีการใช้เว็บไมนิ่งเข้ามาช่วยในงานบริการ เว็บไซต์เพื่อนำไปพัฒนาการให้บริการเว็บไซต์ที่ดี และน่าสนใจขึ้น ส่วนใหญ่จะเป็นการใช้ข้อมูลประเภทตัวอักษร (Text) มาใช้

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์การใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิเคราะห์ร่วมกับข้อมูลชนิดอื่นๆ โดยที่ข้อมูลเหล่านี้จะได้มาจากการบันทึกรายละเอียดของผู้ใช้เมื่อผู้ใช้เข้ามาเยี่ยมชมเว็บไซต์แต่ละแห่งนั้น ซึ่งการนำเว็บไมนิ่งมาประยุกต์ใช้ในลักษณะนี้ได้แก่

- การรวมข้อมูลตัวอักษรเข้ากับลิงค์บนเว็บไซต์เพื่อสรุปหาเว็บเพจที่จะต้องได้รับอนุญาตก่อน จึงจะสามารถเข้าใช้เว็บไซต์นั้นได้ ซึ่งการเข้าใช้ได้นั้นจะต้องอยู่ในขอบเขตที่กำหนดเอาไว้ โดยที่การทำแบบนี้จะมีคุณภาพดีกว่าการใช้ Search Engines ที่มีอยู่ในปัจจุบัน
- การรวบรวมข้อมูลเกี่ยวกับลักษณะการใช้งานของผู้เข้าเยี่ยมชมเว็บไซต์, ช่วงระยะเวลาและลิงค์ที่มีผู้เข้ามาใช้บริการ เพื่อที่จะนำข้อมูลเหล่านั้นมาพิจารณาปรับปรุงแก้ไข โดยอาจจะเพิ่มหรือลดบริการบางชนิดเพื่อให้เหมาะสมกับกลุ่มผู้ใช้ในแต่ละสภาพแวดล้อมซึ่งอาจจะมีความสนใจที่แตกต่างกันไป
- การนำข้อมูลประเภทตัวอักษร (Text) และข้อมูลประเภทรูปภาพ (Image) มาผสมผสานกันเพื่อประโยชน์ในการค้นหาของ Search Engines ในกรณีที่ต้องการค้นหาข้อมูลประเภทรูปภาพ

2.3 การใช้เว็บไมนิ่งทางการตลาดของธุรกิจ e-commerce

ปัจจุบันการเกิดขึ้นของธุรกิจ e-commerce ทำให้เกิดการเปลี่ยนแปลงของการทำธุรกิจค้าขายแบบเดิมเป็นอย่างมาก จากการทำการค้าขายแบบ การเร่ขาย ห้างร้านหรือร้านค้า เวลาที่ลูกค้าต้องการสินค้าก็จะต้องเข้าไปซื้อในห้างร้านหรือร้านค้าเหล่านั้น แต่เมื่อมีธุรกิจ e-commerce เข้ามาทำให้รูปแบบการค้าขายเปลี่ยนไป จากการค้าขายกันแบบเดิมก็จะเปลี่ยนมาเป็นการค้าขายในรูปแบบอิเล็กทรอนิกส์โดยผ่านทางระบบเครือข่าย ซึ่งลูกค้าจะสามารถซื้อสินค้าได้โดยไม่ต้องไปซื้อที่ร้านค้าซึ่งจะช่วยลดเวลาและค่าใช้จ่ายในการเดินทาง, สามารถทำการซื้อสินค้าได้ตลอด 24 ชั่วโมง และสามารถซื้อสินค้าตามที่ต้องการซึ่งมีอยู่มากมายหลากหลายประเภทให้เลือกบนอินเทอร์เน็ตได้ ซึ่งเป็นการเพิ่มความสะดวกสบายให้กับลูกค้าเป็นอย่างมาก

แต่ในการทำธุรกิจค้าขายใดๆ นั้นไม่ว่าจะอยู่ในรูปของห้างร้านหรือร้านค้า, การให้บริการส่งของทางไปรษณีย์ หรือการทำธุรกิจแบบอิเล็กทรอนิกส์นั้น ปัจจัยที่มีความสำคัญอย่างมากต่อความสำเร็จของธุรกิจก็คือ ความเข้าใจในตัวลูกค้า หรือกลุ่มลูกค้า ยิ่งรู้ข้อมูลเกี่ยวกับลูกค้ามากเท่าไรก็จะยิ่งทำให้เข้าใจลูกค้าอย่างแท้จริงมากขึ้นเท่านั้น ซึ่งจะเป็นการเพิ่มโอกาสทางธุรกิจและจำทำให้ธุรกิจตรงกับความต้องการของตลาดมากขึ้นด้วย

จะเห็นได้ว่าข้อมูลเกี่ยวกับลูกค้าเป็นสิ่งที่สำคัญมากต่อธุรกิจ แต่อาจจะอยู่ในรูปแบบที่ไม่ชัดเจน อันได้แก่ ข้อมูลที่รวบรวมได้จากการบันทึกใน log file ของการใช้บริการ หรือข้อมูลจากการสมัครสมาชิกในรูปแบบต่างๆ ข้อมูลเหล่านี้สามารถอำนวยความสะดวกในการติดตามผู้ใช้ (user tracking) ได้เป็นอย่างดี ยิ่งถ้าผู้ใช้เข้าใช้บริการเว็บไซต์บ่อยและนานเท่าใด ก็จะทำให้มีโอกาส

ทราบและรู้จักกับผู้ใช้มากขึ้นเท่านั้น สำหรับข้อมูลเกี่ยวกับผู้ใช้นั้น ได้มีการวิเคราะห์ออกมาได้ในลักษณะดังต่อไปนี้

- Geographic เป็นข้อมูลเกี่ยวกับที่อยู่ หรือสถานที่ของผู้ใช้ในขณะที่เข้าใช้บริการเว็บไซต์ซึ่งจะสามารถประมวลผลเป็นสถิติบริเวณที่อยู่อาศัยของกลุ่มผู้ใช้ได้
- Demographic เป็นข้อมูลเกี่ยวกับ ลักษณะที่เกี่ยวกับผู้ใช้ เช่น อายุ เพศ เป็นต้น
- Psychographics เป็นข้อมูลทางด้านจิตวิทยา ซึ่งแสดงถึงพฤติกรรม หรือค่านิยมในด้านต่างๆ ของผู้ใช้ ซึ่งจะสามารถแบ่งแยกกลุ่มผู้ใช้ได้ตามข้อมูลการเข้าใช้บริการเว็บไซต์ ทั้งในแง่ของเวลาและเนื้อหา
- Technographics เป็นข้อมูลที่แสดงถึงระดับความรู้และความสนใจเทคโนโลยีด้านต่างๆ ของผู้ใช้ รวมถึงเครื่องคอมพิวเตอร์ที่ติดต่อเข้ามาด้วย ซึ่งจะช่วยในการพัฒนาสินค้าและบริการให้น่าสนใจและเหมาะสมในแง่ของเทคโนโลยีได้ดียิ่งขึ้น

เมื่อนำข้อมูลที่วิเคราะห์ทั้ง 4 ลักษณะนี้มาพิจารณาโดยละเอียด จะเกิดประโยชน์อย่างมากในการศึกษาเกี่ยวกับพฤติกรรมโดยรวมของผู้ใช้ ซึ่งจำนวนข้อมูลที่นำมาใช้วิเคราะห์นั้นมักจะมีจำนวนมาก ซึ่งจะทำให้ผลการวิเคราะห์ที่ได้นั้นมีความแม่นยำสูง

2.4 เทคนิคในการนำเว็บไมนิ่งมาทำการวิเคราะห์ Web log records

การทำงานที่สำคัญอย่างหนึ่งสำหรับเว็บไมนิ่งก็คือ การทำการวิเคราะห์ Web log records เพื่อที่จะได้นำข้อมูลมาวิเคราะห์รูปแบบการเข้าถึง Web page ของผู้ใช้ ซึ่งการวิเคราะห์และการตรวจสอบอย่างสม่ำเสมอใน Web log records นี้จะสามารถที่จะระบุลูกค้าที่เป็นไปได้สำหรับธุรกิจ e-commerce , การเพิ่มคุณภาพในการบริการ , การบริการส่งข้อมูลอินเทอร์เน็ตไปให้กับผู้ใช้และการปรับปรุงประสิทธิภาพของระบบ Web server ได้อีกด้วย

โดยปกติ Web server จะทำการเก็บ log entry หรือ Web log entry สำหรับทุกๆ การเข้าถึง Web page ของผู้ใช้ซึ่งจะรวมถึงการร้องขอ URL , IP Address จากเครื่องที่ร้องขอ URL และเวลาที่ทำการร้องขอ URL นั้นๆ สำหรับ Web server ของธุรกิจนั้น ได้มีการจัดเก็บ Web log records จำนวนมากเอาไว้ โดย Web site ที่ได้รับความนิยมอาจจะต้องเก็บ Web log records ในปริมาณที่เป็นจำนวนมากหลายๆ ร้อยเมกะไบต์ในทุกๆ วัน ซึ่งจะทำให้ฐานข้อมูลที่จัดเก็บ Web log จะต้องจัดเตรียมข้อมูลเป็นจำนวนมากที่เกี่ยวกับการทำ Web dynamic ดังนั้นจึงมีการนำเอาเทคนิคของ Web mining มาใช้ในการวิเคราะห์ Web log records ขึ้น

โดยที่การพัฒนาเทคนิคของ Web mining นั้นอาจจะพิจารณาได้ดังต่อไปนี้

1. แม้ว่าจะมีการส่งเสริมการพัฒนาแอปพลิเคชันหลายๆ แอปพลิเคชันที่เป็นไปได้ที่ใช้ในการวิเคราะห์ไฟล์ Web log แต่สิ่งที่สำคัญที่จะทราบถึงความสำเร็จในการพัฒนาของแอปพลิเคชัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

นั้นๆ จะขึ้นอยู่กับว่าข้อมูลที่จะใช้คืออะไร มีความถูกต้องและความน่าเชื่อถือมากน้อยเพียงใด ซึ่งข้อมูลนั้นจะสามารถค้นหาได้จาก Web log ไปใช้ได้นั้นจะต้องนำข้อมูลเหล่านั้นไปทำการคัดสรร (cleaned) , พิจารณา (condensed) และเปลี่ยนแปลงรูป (transformed) ก่อนที่จะนำไปแก้ไข วิเคราะห์สิ่งที่มีความหมาย และสามารถนำข้อมูลที่มีไปใช้ประโยชน์ได้ ซึ่งวิธีการเหล่านี้จะเป็นสิ่งที่จำเป็นก่อนที่จะนำ Web log ไปทำการประมวลผล

2. เกี่ยวกับข้อมูลของการร้องขอ URL , เวลา , IP Address และข้อมูล Web page ที่ผู้ใช้เข้าถึง ที่อยู่ใน Web log นั้นจะมีการสร้าง multidimensional view บนฐานข้อมูลของ Web log และการวิเคราะห์ multidimensional OLAP จะถูกดำเนินการบนฐานข้อมูลของ Web log นั้น เพื่อหา top-N users , top N ที่เข้าถึง Web page , จำนวนการเข้าถึง Web page ที่มากที่สุดในช่วงระยะเวลาหนึ่งๆ ซึ่งเหล่านี้จะช่วยกันค้นหาลูกค้าที่เป็นไปได้ของธุรกิจ , ที่ผู้ใช้ และตลาดสำหรับธุรกิจได้

3. มีการใช้เทคนิคดาต้าไมนิ่งกับ Web log records ซึ่งเทคนิคหนึ่งที่น่าสนใจในการวิเคราะห์ Web log records ก็คือ “association rule discovery” ซึ่งเป็นเทคนิคของดาต้าไมนิ่งที่สามารถนำไปประยุกต์ใช้ได้กับงานต่างๆ โดยที่หลักการทํางานของวิธีนี้ก็คื การค้นหาความสัมพันธ์จากข้อมูล Web log records ที่มีเป็นจำนวนมากอยู่ในฐานข้อมูล เพื่อนำไปใช้ในการวิเคราะห์ความสัมพันธ์ของสินค้าของธุรกิจที่มีอยู่ว่ามีความสัมพันธ์กับสินค้าตัวอื่นๆอย่างไร

ซึ่งการใช้ association rule discovery ในการค้นหาความสัมพันธ์แบบไม่เป็นลำดับนี้จะอยู่ในรูปของกลุ่มของ Web page ที่ถูกเข้าถึงของผู้เยี่ยมชม โดยที่จะค้นหาความสัมพันธ์ในการเข้าถึง Web page ของผู้เยี่ยมชมเว็บไซต์

โดยที่ตัวอย่างของเว็บไซต์ที่มีการใช้เทคนิค association rule discovery มาใช้ในการวิเคราะห์การเข้าถึง Web page ของผู้เข้ามาเยี่ยมชมจะมีดังนี้จากเว็บไซต์อย่างเป็นทางการของโอลิมปิกปี 1996 พบว่า 45% ของผู้เยี่ยมชมเว็บไซต์ที่เข้าเยี่ยมชมหน้า Web page เกี่ยวกับกีฬา วอลเลย์บอลจะเข้าชมหน้า Web page ที่เกี่ยวกับ กีฬาแฮนด์บอลด้วยหรือ 59.7% ของผู้ที่เข้าเยี่ยมชมเว็บไซต์ที่เข้าเยี่ยมชมหน้า Web page เกี่ยวกับกีฬาเบตมินตันและกระโดดน้ำจะเยี่ยมชม Web page ที่เกี่ยวกับกีฬาปีโปงด้วยหรือมีการใช้เทคนิค association rule discovery มาใช้ในการขายสินค้าของร้านค้า Amazon ซึ่งร้านค้า Amazon จะมีการใช้เทคนิคนี้ไปประยุกต์ใช้กับระบบแนะนำหนังสือให้กับลูกค้าแบบอัตโนมัติ โดยลูกค้าที่ซื้อหนังสือเล่มหนึ่งๆ มักจะซื้อหนังสือเล่มใดบ้างไปพร้อมกัน ซึ่งข้อมูลที่ได้จากการวิเคราะห์ความสัมพันธ์จากกระบวนการนี้จะสามารถนำไปใช้ในการคาดเดาได้ว่าควรจะแนะนำหนังสือเล่มใดเพิ่มเติมให้กับลูกค้าที่ทำการซื้อสินค้าไป ตัวอย่าง 60% ของลูกค้าเมื่อซื้อหนังสือ Database แล้วมีโอกาที่จะซื้อเกี่ยวกับ Data mining ด้วยเป็นต้น

ซึ่งการใช้เทคนิคนี้เข้ามาช่วย จะทำให้สามารถกำหนดได้ว่าสินค้าประเภทใดมักจะถูกซื้อควบคู่กันในการซื้อแต่ละครั้ง ทำให้ธุรกิจสามารถกำหนดได้ว่าควรจัดเตรียมการแสดงสินค้าที่มี

ความสัมพันธ์กันมาแสดงไว้ในส่วนเดียวกัน หรือในหน้า Web page เดียวกัน หรือการวิเคราะห์

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพื่อให้ทราบถึงรูปแบบการซื้อสินค้าของลูกค้าว่าจะมีรูปแบบเป็นอย่างไร เพื่อที่จะได้นำไปเป็นเกณฑ์ในการจัดรูปแบบการแนะนำสินค้าให้กับลูกค้าได้

การทำวิเคราะห์ Web log file นั้นจะสามารถวิเคราะห์ประสิทธิภาพของระบบได้และจะสามารถเข้าใจถึงการตอบสนอง แรงจูงใจของผู้ใช้ได้อีกด้วย ตัวอย่างเช่น หลายๆการศึกษาเสนอแนะว่าการคัดแปลงเว็บไซต์ Web site โดยจะปรับปรุงตัวเองโดยเรียนรู้จากรูปแบบการเข้าถึงของผู้ใช้ ว่าผู้ใช้ประเภทใดจะเข้าถึง Web page ประเภทใด ซึ่งเหล่านี้จะต้องเข้าใจถึงการตอบสนองและแรงจูงใจของผู้ใช้ด้วย และการวิเคราะห์ Web log นี้จะช่วยสร้างการให้บริการเว็บไซต์สำหรับผู้ใช้ โดยได้อีกด้วย

2.5 ข้อดีของการใช้เว็บไมนิ่ง

เว็บไมนิ่งถูกนำมาใช้สนับสนุนการตัดสินใจของธุรกิจ โดยการสร้างมูลค่าเพิ่มให้กับข้อมูลที่มีอยู่ ซึ่งประโยชน์ที่ธุรกิจได้รับจากการใช้เว็บไมนิ่งสามารถสรุปเป็นหัวข้อย่อได้ดังนี้

2.5.1 ทำให้สามารถจัดสรรทรัพยากรที่มีให้ตรงกับสิ่งที่ผู้เยี่ยมชมสนใจ

ทรัพยากรที่ใช้เพื่อการขายสินค้าของธุรกิจ ซึ่งจะเป็นข้อมูลถูกแบ่งแยกเพื่อที่จะสามารถกระจายข้อมูลผ่านเน็ตเวิร์คไปได้, แผ่นป้ายโฆษณาจากลูกค้าที่เป็นผู้โฆษณา, อีเมลล์ที่ถูกแยกมาจากในเมลล์ลิสต์ หรือทุกๆสิ่งที่จะกระจายออนไลน์ออกไป ข้อมูลจากทรัพยากรเหล่านี้จะถูกเก็บไว้ในฐานข้อมูล ซึ่งการใช้เว็บไมนิ่งจะช่วยในการเรียนรู้สิ่งที่ผู้เยี่ยมชมสนใจ โดยการรวบรวมและการวิเคราะห์ข้อมูลที่เกิดขึ้นจากการเข้ามาติดต่อเว็บไซต์ เช่นเดียวกันกับข้อมูลการ clickstream, ความต้องการในการค้นหา และข้อมูลจากไฟล์คูกี้ โดยที่เว็บไมนิ่งสามารถใช้เพื่อจัดระเบียบทรัพยากรของธุรกิจ โดยจะจัดทรัพยากรเหล่านั้นให้ตรงกับความต้องการของผู้ใช้ การให้บริการจัดสรรทรัพยากรที่ดีที่สุดให้ตรงกับกรร้องขอของผู้ใช้ให้มากที่สุดซึ่งส่งผลให้การเปลี่ยนจากผู้เยี่ยมชมไปเป็นลูกค้าของธุรกิจมีอัตราที่สูงขึ้นสำหรับธุรกิจ

2.5.2 เพิ่มคุณค่าของผู้เยี่ยมชมแต่ละคน

เมื่อมีการใช้เว็บไมนิ่งก็สามารถที่จะคาดคะเนได้ว่าอะไรที่เป็นข้อมูลที่ผู้เยี่ยมชมอาจจะสนใจ และสินค้าอะไรที่ผู้เยี่ยมชมอาจจะพิจารณาสั่งซื้อ การคาดคะเนเหล่านี้ถูกใช้เพื่อแสดงความสัมพันธ์ระหว่างผู้เยี่ยมชมกับสินค้าและทรัพยากร ความรู้ที่มีความหมายเหล่านี้จะเพิ่มคุณค่าให้กับลูกค้าสำหรับธุรกิจ และเมื่อนำมาใช้กับการส่งเสริมการขายเฉพาะตัวแบบ cross-selling และ up-selling ก็จะทำให้รายรับของธุรกิจเพิ่มขึ้นด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.3 ทำให้ประสบการณ์ของผู้เยี่ยมชมเว็บไซต์ดีขึ้น

การรวมเสียงเข้ากับข้อมูลและการใช้เทคนิคของเท็กซ์ไมนิ่งสามารถช่วยตัดติณใจความสนใจของผู้ใช้ และสามารถส่งทรัพยากรที่ตรงประเด็นมากที่สุดให้กับผู้เยี่ยมชมได้ ในโลกของการค้าขายบนอินเทอร์เน็ตที่ซึ่งการปรับเปลี่ยนของราคาของผู้ขายแต่ละรายจะไม่มี ความแตกต่างกันง่ายต่อการเข้าถึงข้อมูลที่ตรงประเด็นอาจจะกลายเป็นความแตกต่างระหว่างผลกำไรจากลูกค้ากับโอกาสที่สูญเสียไป ซึ่งการที่จะทำให้ความพึงพอใจของลูกค้าเพิ่มขึ้นนั้น ธุรกิจก็จะต้องเพิ่มความพึงพอใจให้กับลูกค้าและจะต้องสร้าง brand loyalty เพิ่มเติมด้วย

2.5.4 จัดสรรทรัพยากรให้กับกลุ่มเป้าหมายได้อย่างถูกต้อง

ผู้ที่เยี่ยมชมเว็บไซต์นั้น ไม่ได้หมายความว่าผู้เยี่ยมชมทั้งหมดจะเป็นลูกค้าของธุรกิจ จะมีเพียงบางคนเท่านั้นที่จะเป็นลูกค้าที่เป็นไปได้ที่พร้อมจะคลิกและซื้อสินค้า ความเป็นไปได้สำหรับข้อมูลพร้อมกันกับให้ผู้เยี่ยมชมเหล่านั้นรู้จัก คุณเคยตราสินค้าของธุรกิจ ซึ่งสิ่งเหล่านั้นจะกลายมาเป็นสิ่งสำคัญมากและจะกลายมาเป็นลูกค้าที่จะทำกำไรให้กับธุรกิจได้ในอนาคต แต่ก็จะมีกลุ่มของผู้เยี่ยมชมกลุ่มอื่นๆ ที่สนุกกับของฟรีเพียงอย่างเดียว กลุ่มคนเหล่านั้นจะให้ทรัพยากรที่ใช้สำหรับส่งเสริมการขายอย่างเต็มที่แต่ไม่เคยที่จะซื้อสินค้าของธุรกิจเลย ผู้เยี่ยมชมเหล่านี้จะเข้ามาชมเว็บไซต์และอยู่ในใจที่จะใช้ทรัพยากรในเว็บไซต์ของธุรกิจ แต่เมื่อการนำเว็บไมนิ่งเข้ามาใช้สามารถที่จะกำหนดขอบเขตการใช้งานประสิทธิภาพของเว็บไซต์ได้ และจะต้องการเพียงการร้องขอที่สำคัญจากลูกค้าที่เป็นไปได้มากที่สุดเท่านั้น ถ้ามีการกระจายทรัพยากรที่ใช้การส่งเสริมการขายในปริมาณสูงก็จะต้องใช้งบประมาณสำหรับการโปรโมชันอย่างฉลาด โดยเสมอที่สุดให้ ไม่ใช่ส่งโปรโมชันเหล่านั้นให้กับทุกๆ คนที่เข้ามาชมเว็บไซต์

2.5.5 เป็นแนวทางในการจัดเก็บข้อมูลแบบใหม่

งานหลักๆ ของผู้ขายทางอินเทอร์เน็ตเป็นงานที่เกี่ยวกับการจัดเก็บข้อมูล ซึ่งเป็นสิ่งที่จำเป็นสำหรับตลาดเป้าหมาย สำหรับเป้าหมายอื่นๆ การจัดเก็บข้อมูลนี้ก็จะสร้างแรงจูงใจสำหรับการสร้างเว็บไซต์ของธุรกิจได้เช่นกัน โดยวิธีการจัดเก็บข้อมูลแบบดั้งเดิมก็จะมี การส่งเสริมการขาย การสำรวจ , การโฟกัสกลุ่มเป้าหมาย ซึ่งเหล่านี้จะมีปัญหาต่างๆ มากมายเกิดขึ้นรวมถึงอัตราการตอบสนองของผู้เยี่ยมชมมีอัตราที่ต่ำ , การขาดความถูกต้องของข้อมูล และมีสารบัญชของเว็บไซต์ ซึ่งจะทำให้ผู้เยี่ยมชมสามารถที่จะค้นหาข้อมูลที่ใช่ประโยชน์ และสามารถบอกความต้องการของการส่งเสริมการขายได้อีกด้วย การใช้เว็บไมนิ่งนี้โดยที่จะยึดถือตามเนื้อหาที่ผู้ใช้เคยเข้าถึง Web page นั้นๆ

2.5.6 สามารถทดสอบความสัมพันธ์ของเนื้อหากับโครงสร้างของเว็บไซต์ได้

การที่จะมองคุณลักษณะเฉพาะของเนื้อหาและโครงสร้างของเว็บไซต์นั้น บางครั้งก็อาจต้องการที่จะเพิ่มประโยชน์จากการใช้งาน หรือจัดสรรเว็บไซต์สำหรับการมองที่ดีที่สุดสำหรับเว็บไซต์ของธุรกิจ ซึ่งการใช้เว็บไมนิ่งมาวิเคราะห์ log file ก็สามารถช่วยให้มองเห็นเส้นทางการเดินทางที่ผ่านเว็บไซต์ของธุรกิจ ซึ่งจะเฉลี่ยมาจากผู้เยี่ยมชมทั้งหมด แต่ถึงนั้นก็ไม่ใช่สิ่งที่ธุรกิจกำลังมองหาอยู่ และเมื่อทำการจัดสรรโครงสร้างเว็บไซต์ของธุรกิจ ก็ควรที่จะจัดสรรให้กับโอกาสที่ดีที่สุด และก็ไม่ต้องทำกับทุกคน ประมาณ 15% ของผู้เยี่ยมชมเว็บไซต์จะครอบคลุมไปถึงลูกค้าที่เป็นไปได้ของธุรกิจจริงๆ ด้วย และอีก 85% ที่เหลือก็จะเป็นผู้เยี่ยมชมโดยปกติที่ไม่ค่อยมีค่าต่อธุรกิจเท่าไรนัก ดังนั้นก็ควรที่จะแยกโอกาสที่เป็นไปได้น้อยๆ ออก และเอาสิ่งที่สนับสนุนสิ่งเหล่านั้นออกตามไปด้วย ซึ่งก็จะเป็นสิ่งที่เหลือไว้แสดงคุณภาพให้กับลูกค้าที่เป็นไปได้ของเว็บไซต์ของธุรกิจจริงๆ และสิ่งนี้จะเป็นภาพรวมที่สามารถช่วยปรับปรุงประสิทธิภาพในการขายได้

ในปัจจุบันการค้าขายบนอินเทอร์เน็ตหรือที่เรียกกันว่าการทำธุรกิจ e-commerce ได้รับความนิยมจากองค์กรธุรกิจเป็นจำนวนมาก เนื่องจากการทำธุรกิจ e-commerce เป็นธุรกิจที่มีลูกค้าเป็นจำนวนมากและเป็นช่องทางหนึ่งที่จะทำกำไรให้กับธุรกิจ และทำให้ธุรกิจประสบความสำเร็จในการดำเนินการธุรกิจได้ด้วย การเข้าใจในกระบวนการการทำงานและนำเว็บไมนิ่งมาใช้กับธุรกิจจะเป็นสิ่งที่ช่วยในการตัดสินใจในการดำเนินงานของธุรกิจ ซึ่งจะทำให้ธุรกิจประสบความสำเร็จเหนือคู่แข่ง

2.6 ระบบแนะนำสินค้าและบริการในธุรกิจ e-commerce

ปัจจุบันธุรกิจการซื้อขายสินค้าบนอินเทอร์เน็ตหรือที่เรียกกันว่า e-commerce ได้รับความนิยมมากขึ้นเรื่อยๆ ซึ่งหลายๆ บริษัทได้จัดทำเว็บไซต์ของตนขึ้นมาเพื่อทำการแข่งขันในธุรกิจ e-commerce ซึ่งเป็นหนทางหนึ่งในการนำเสนอสินค้าของตนเองออกสู่ผู้บริโภค โดยจุดประสงค์หลักของบริษัทคือ ต้องการจะนำเสนอสินค้าที่ตนมีเป็นจำนวนมากนี้ไปให้กับความต้องการหลายๆ ความต้องการจากผู้บริโภคหลายๆ คน แต่ในขณะที่ธุรกิจ e-commerce ก็ไม่ได้จำเป็นธุรกิจจะต้องผลิตสินค้าออกมาเสนอขายเป็นจำนวนมาก ธุรกิจ e-commerce จะให้ลูกค้ามีทางเลือกในสินค้ามาก ตัวอย่างเช่น หนังสือจำนวน 10,000 เล่มในร้านหนังสือ ลูกค้าจะเลือกเป็นจำนวนหลายเล่มในการซื้อจากร้านทางอินเทอร์เน็ต การเพิ่มจำนวนการเลือกซื้อเหล่านี้จำเป็นที่จะต้องเพิ่มให้ข้อมูลของสินค้าให้กับลูกค้าก่อนที่ลูกค้าจะทำการเลือกสินค้า ซึ่งลูกค้าจะต้องใช้ระยะเวลาในการเลือกก่อนที่พวกเขาจะสามารถเลือกสินค้าที่ตรงกับความต้องการ โดยที่ข้อมูลที่นำเสนอเหล่านั้นมีเป็นจำนวนมาก ซึ่งในธุรกิจ e-commerce ได้เก็บข้อมูลเบื้องต้นเอาไว้ ทางเลือกหนึ่งที่จะนำข้อมูลที่มีอยู่ในธุรกิจ e-commerce นี้ไปใช้ประโยชน์ได้แก่ “ระบบการแนะนำสินค้า”

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบแนะนำสินค้าเป็นของแปลกใหม่ในธุรกิจ e-commerce ซึ่งได้เข้ามาเป็นเครื่องมือทางธุรกิจในการเปลี่ยนแปลงรูปร่างในโลกของธุรกิจ e-commerce เว็บไซต์ของธุรกิจ e-commerce ขนาดใหญ่จะมีสินค้าสำหรับจำหน่ายเป็นจำนวนหลายล้านชิ้น ซึ่งเป็นทางเลือกใหม่ให้กับลูกค้าเป็นจำนวนมาก ระบบแนะนำสินค้าได้ออกมาเพื่อตอบสนองทางแก้ไขของปัญหา ปัจจุบันนี้เว็บไซต์ของธุรกิจ e-commerce ส่วนมากใช้ระบบแนะนำสินค้ากันเป็นจำนวนมาก เพื่อช่วยอำนวยความสะดวกให้กับลูกค้าในการช่วยหาสินค้าที่ลูกค้าต้องการซื้อ ระบบแนะนำสินค้าสามารถเรียนรู้ลูกค้าและสามารถแนะนำสินค้าที่ลูกค้าต้องการที่มีอยู่ และจัดหาข้อมูลในการอธิบายสินค้าที่ลูกค้าต้องการซื้อ โดยการแนะนำสินค้าโดยสามารถแนะนำได้จาก สินค้าที่ขายดีที่สุดในเว็บไซต์ แนะนำตามรสนิยมของกลุ่มประชากรของลูกค้า หรือแนะนำจากการวิเคราะห์พฤติกรรมการซื้อสินค้าในอดีตของลูกค้า และคาดการณ์ว่าในอนาคตของพฤติกรรมการซื้อของลูกค้าจะเป็นเช่นไรรูปแบบของการแนะนำสินค้าจะรวมถึงการแนะนำสินค้าให้กับลูกค้า การจัดเตรียมข้อมูลสินค้าเฉพาะบุคคล และจัดเตรียมการติชมของลูกค้า เทคนิคการแนะนำนี้เป็นส่วนหนึ่งของการทำการแนะนำเฉพาะบุคคลบนเว็บไซต์เพราะว่าระบบจะสามารถช่วยเว็บไซต์ในการปรับเปลี่ยนตัวมันเองสำหรับลูกค้าแต่ละคนได้

ระบบแนะนำสินค้านี้มีส่วนที่คล้ายและแตกต่างจากระบบการตลาดและระบบ supply-chain decision-support ด้วยระบบการตลาดจะช่วยนักการตลาดในการตัดสินใจเกี่ยวกับ การตลาดให้กับกลุ่มลูกค้าได้อย่างไร ซึ่งสามารถจัดกลุ่มลูกค้าที่มีลักษณะเหมือนกันไว้ในกลุ่มการตลาดเดียวกันและจัดกลุ่มประเภทของสินค้าให้ตรงกับกลุ่มการตลาดนั้นๆ และทำการจัด โปรโมชันต่างๆ ไปให้กับกลุ่มลูกค้าในแต่ละส่วน โดยที่สินค้าในแต่ละประเภทจะถูกเลือกโดยนักการตลาดในทางตรงกันข้ามระบบแนะนำสินค้าจะติดต่อโดยตรงกับลูกค้า และจะช่วยลูกค้าค้นหาสินค้าที่ลูกค้าพอใจที่จะซื้อ ส่วนระบบ supply-chain decision-support จะช่วยนักการตลาดทำการตัดสินใจเกี่ยวกับจำนวนสินค้าที่จะสั่งผลิต และจะทำการจัดเก็บสินค้าไว้ในโกดัง ซึ่งระบบสามารถช่วยเก็บข้อมูลในการวิเคราะห์เพื่อทำนายว่าในแต่ละสถานที่จะมีการซื้อสินค้าชนิดใดและเป็นจำนวนเท่าไร ดังนั้นจึงเป็นการนำสินค้าที่ถูกต้องไปให้กับลูกค้าที่ต้องการจะซื้อ ระบบ supply-chain decision-support ส่วนระบบจะตอบคำถามกับข้อสรุป ตัวอย่างเช่น จำนวนลูกค้าทั้งหมดใน Minneapolis จำนวนการซื้อยาตีฟีนในเดือนกุมภาพันธ์ เป็นต้น แต่ระบบแนะนำสินค้าจะตอบคำถามเกี่ยวกับลูกค้าเฉพาะบุคคล ตัวอย่าง สินค้าชนิดไหนที่ลูกค้าคนนี้จะชอบที่จะซื้อในขณะนี้ เป็นต้น

ระบบแนะนำสินค้าประกอบด้วยกระบวนการที่กระทำด้วยมือ ตัวอย่างเช่น การสร้าง cross-sell list และกระบวนการที่กระทำด้วยคอมพิวเตอร์ ตัวอย่างเช่น collaborative filtering ซึ่งจะอ้างถึงระบบแนะนำสินค้าอัตโนมัติ ระบบแนะนำสินค้าอัตโนมัติเป็นระบบที่ใช้ระบบคาด้าไมนิ่งซึ่งได้ผลสำหรับการติดต่อลูกค้ากับกลุ่มลูกค้ามากกว่านักการตลาด โดยถูกออกแบบมาอย่างชัดเจนซึ่งจะดีสำหรับการแนะนำเฉพาะบุคคลแบบ real-time ซึ่งจะเหมาะสมกับ interactive e-commerce

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งยังมีให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7 การทำงานของระบบแนะนำสินค้า

2.7.1 ระบบแนะนำสินค้าช่วยธุรกิจ e-commerce ได้อย่างไร

- การแปลงผู้เข้าชมเว็บไซต์ให้กลายเป็นผู้ซื้อ (Converting Browsers into buyers)
ผู้เข้ามาเยี่ยมชมเว็บไซต์เป็นผู้ที่เข้ามาเยี่ยมชมเว็บไซต์ของธุรกิจเพียงอย่างเดียว ไม่ได้ทำการซื้อสินค้าใดๆ จากธุรกิจเลย ระบบแนะนำสินค้าสามารถช่วยผู้เข้าชมมาเยี่ยมชมเหล่านี้หาสินค้าที่ตนเองปรารถนาที่จะซื้อได้

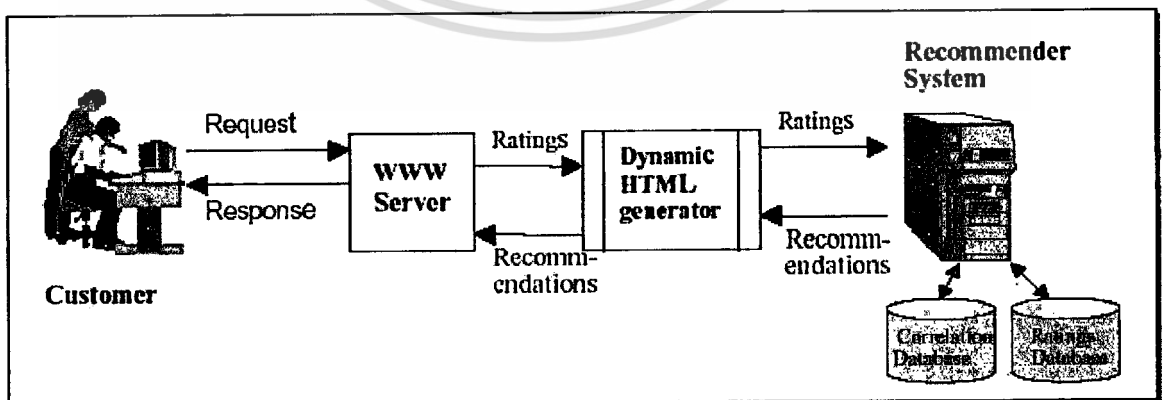
- เพิ่มการขายแบบ Cross-sell (Increasing Cross-sell)

ระบบแนะนำสินค้าส่งเสริมการขายแบบ Cross-sell ให้ดีขึ้นโดยแนะนำสินค้าเพิ่มเติมให้กับลูกค้าทำการซื้อสินค้า ถ้าระบบแนะนำสินค้ามีประสิทธิภาพ ค่าเฉลี่ยของขนาดคำสั่งซื้อที่ควรที่จะเพิ่มขึ้น ตัวอย่างเช่น เว็บไซต์อาจจะแนะนำสินค้าเพิ่มเติมในกระบวนการ checkout ของลูกค้า โดยมีพื้นฐานมาจากสินค้าที่อยู่ใน shopping cart เรียบร้อยแล้ว

- สร้างความจงรักภักดีให้กับลูกค้า (Building Loyalty)

ในโลกปัจจุบันธุรกิจ e-commerce มีการแข่งขันกันสูงมาก การได้รับความยอมรับจากลูกค้าถือว่าเป็นหัวใจของกลยุทธ์ของธุรกิจ ระบบแนะนำสินค้าจะเพิ่มรายละเอียดของสินค้า ซึ่งสามารถเพิ่มความสัมพันธ์ระหว่างเว็บไซต์กับตัวลูกค้า เว็บไซต์สร้างไว้สำหรับการเรียนรู้เกี่ยวกับตัวลูกค้า โดยใช้ระบบแนะนำสินค้าเป็นตัวทำการเรียนรู้ และนำเสนอให้ตรงกับความต้องการของลูกค้า ซึ่งลูกค้าเองสามารถซื้อสินค้าตรงกับความต้องการเช่นกัน นอกจากนี้ สิ่งอื่นๆ ที่ลูกค้าให้กับระบบแนะนำสินค้าก็คือ การสามารถใช้ระบบแนะนำสินค้า และยังสามารถแนะนำลูกค้าอื่นเข้ามาเยี่ยมชมเว็บไซต์ได้อีกด้วย

2.7.2 การทำงานของระบบแนะนำสินค้า



รูปที่ 2.1 แสดงการทำงานของระบบแนะนำสินค้า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การทำงานของระบบแนะนำสินค้า ดังรูปที่ 2.1 ลูกค้าจะเข้ามาติดต่อเว็บไซต์ของธุรกิจโดยผ่านทาง Web Interface ผ่านทาง Web Browser แล้ว Web Browser จะร้องขอไปยัง web server ซึ่งจะมี Web server software ติดต่อกับระบบแนะนำสินค้า เพื่อทำหน้าที่เลือกสินค้าที่จะแนะนำให้กับลูกค้า โดยที่ระบบแนะนำสินค้าจะใช้ Rating Database มาทำ rating ของสินค้าและใช้ข้อมูลจาก Correlation Database ฐานข้อมูลของสินค้าที่มีความเกี่ยวข้องกัน มาทำการแนะนำ และจะส่งการแนะนำไปให้กับ Web server ซึ่งจะนำการแนะนำสินค้านั้นไปแสดงกับลูกค้าต่อไป

2.7.3 เทคนิคในการสร้างการแนะนำสินค้า (Recommendation Method)

เทคนิคที่ใช้กระบวนการ การสร้างการแนะนำสินค้าผ่านเครือข่ายอินเทอร์เน็ต อาจจะมีการเชื่อมโยงหลายเทคนิคเพื่อสร้างการแนะนำสินค้า ซึ่งการสร้างการแนะนำสินค้านั้นมีอยู่หลายวิธีด้วยกัน โดยที่สามารถสรุปเทคนิคที่เป็นที่นิยมใช้ในกระบวนการสร้างการแนะนำสินค้าได้ดังต่อไปนี้

1. การดึงข้อมูล (Raw retrieval) เป็นลักษณะของระบบที่ไม่มี การแนะนำสินค้าคือจะเป็นการร้องขอเพื่อค้นข้อมูลสินค้าที่ต้องการจากฐานข้อมูล ซึ่งเป็นการตอบสนองเมื่อลูกค้าร้องขอตัวอย่างเช่น เมื่อลูกค้าร้องขอข้อมูลเกี่ยวกับภาพยนตร์เรื่อง Terminator ระบบก็จะส่งรายการสินค้าที่เกี่ยวกับ Terminator ออกมาทั้งหมด ซึ่งอาจจะเป็นประโยชน์ในการแนะนำภาพยนตร์ Terminator ภาคอื่นๆ ให้กับลูกค้าได้

2. การเลือกด้วยคน (Manual selected) เป็นการเลือกสินค้าที่จะแนะนำให้กับลูกค้าด้วยคน ซึ่งอาจจะเป็นบรรณาธิการ หรือผู้เชี่ยวชาญเฉพาะทาง โดยที่การแนะนำสินค้านั้นจะแนะนำบนพื้นฐานในเรื่องรสนิยม ความสนใจของลูกค้า ซึ่งการแนะนำในลักษณะนี้มักจะมิตัววิจารณ์ประกอบอยู่ด้วยเพื่อให้ลูกค้าอื่นๆ สามารถที่จะประเมินและสร้างความเข้าใจถึงเหตุผลในการแนะนำสินค้านั้นได้ด้วย

3. การสรุปผลทางสถิติ (Statistical summaries) ซึ่งวิธีนี้จะเป็นการสร้างการแนะนำจากผลสรุปทางสถิติที่แสดงออกถึงความคิดเห็นต่างๆ ในสังคมเป็นตัววัดความนิยมในสินค้า อย่างเช่น ค่าเฉลี่ยของคะแนนความนิยมของสินค้าแต่ละอย่างเป็นต้น

4. ความสัมพันธ์ระหว่างสินค้าต่อสินค้า (Item-to-Item correlation) เป็นการกำหนดความสัมพันธ์ระหว่างสินค้าที่ลูกค้าแสดงความสนใจร่วมกัน โดยอาศัยข้อมูลที่เกิดขึ้นด้วยกัน ความสัมพันธ์นี้อาจจะเป็นพื้นฐานมาจากสินค้าที่มักจะถูกซื้อด้วยกัน โดยที่ความสัมพันธ์กันระหว่างตัวสินค้าอาจจะใช้การเข้าคู่กัน (Match item) สำหรับสินค้า 1 อย่าง อย่างเช่น สินค้าเสื้อคลุม มักจะถูกซื้อด้วยกันกับกางเกงขาสั้น เป็นต้น

5. ความสัมพันธ์ระหว่างผู้ใช้งานกับผู้ใช้งาน (User-to-User correlation) โดยที่วิธีนี้จะอาศัย

ความสัมพันธ์ของลูกค้าที่ต้องการแนะนำสินค้าคนอื่นๆ ที่มีรายการข้อมูลเกี่ยวกับสินค้าที่แสดงถึง
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความสัมพันธ์บางอย่างร่วมกัน เป็นการกำหนดกลุ่มทางสังคมที่มีพฤติกรรมการซื้อสินค้าและความชอบในลักษณะเดียวกันหรือที่เรียกว่า (Nearest Neighborhood) เพื่อนำไปสู่การสร้างการแนะนำ

2.8 Recommender Systems Based on Collaborative Filtering

จากเทคนิคในข้างต้น ในระบบงานนี้ได้เลือกใช้เทคนิคการสร้างการแนะนำสินค้าโดยใช้เทคนิคที่เรียกว่า Collaborative Filtering (CF) ซึ่งเป็นเทคนิคที่ใช้ในระบบแนะนำสินค้าที่ประสบความสำเร็จมากที่สุดในปัจจุบัน และถูกใช้กับหลายๆระบบแนะนำสินค้าบนเว็บไซต์ที่ประสบความสำเร็จ ระบบแนะนำสินค้าที่ใช้เทคนิค CF จะแนะนำสินค้าให้กับกลุ่มลูกค้าเป้าหมายโดยใช้ฐานความรู้จากความคิดเห็นจากลูกค้าคนอื่นๆระบบนี้จะใช้เทคนิคทางสถิติเพื่อค้นหาเซตของลูกค้าซึ่งจะมีข้อมูลความคิดเห็นของลูกค้ากลุ่มเป้าหมายที่ผ่านมา อย่างเช่น อัตราการซื้อสินค้าของลูกค้าที่ต่างกันในกลุ่มเดียวกัน หรือลักษณะนิสัยของลูกค้าที่ซื้อสินค้าในประเภทเดียวกัน ถ้ามีลูกค้าคนหนึ่งที่อยู่ในกลุ่มทำการซื้อสินค้าขึ้นมา เทคนิคนี้ก็จะใช้อัลกอริทึมหลายๆอัลกอริทึมเพื่อที่จะใช้นำมาสร้างการแนะนำ

ซึ่งการทำงานของเทคนิค CF แบ่งเป็นการทำงานคือ เทคนิคการทำนายใน CF โดยนำแนวความคิดของการคำนวณจากการให้คะแนนความชอบต่อสินค้าของลูกค้าที่เข้ามาเยี่ยมชมสินค้าเทียบกับค่าเฉลี่ยน้ำหนักของการให้คะแนนความชอบต่อสินค้าของกลุ่มลูกค้าคนอื่นเพื่อใช้ในการทำนายซึ่งจะมีสูตรการคำนวณดังนี้

$$P_{a,j} = \bar{v}_a + k \sum_{i=1}^n w(\alpha, i)(v_{i,j} - \bar{v}_i) \quad (2.1)$$

$P_{a,j}$ = ค่าของการทำนายความชอบของลูกค้า α ที่เยี่ยมชมต่อสินค้า j

\bar{v}_a = ค่าเฉลี่ยของการให้คะแนนความชอบของลูกค้าที่เยี่ยมชม

n = จำนวนลูกค้าที่ให้คะแนนความชอบต่อสินค้าที่ต้องการทำนาย

\bar{v}_i = ค่าเฉลี่ยของการให้คะแนนความชอบของลูกค้าแต่ละคนให้คะแนนความชอบต่อสินค้าที่ต้องการทำนาย

$v_{i,j}$ = ระดับคะแนนความชอบที่ลูกค้าแต่ละคนให้คะแนนต่อสินค้าที่ต้องการทำนาย

$w(\alpha, i)$ = ค่าถ่วงน้ำหนักวัดความเกี่ยวข้องกันระหว่างลูกค้าที่เยี่ยมชมกับลูกค้าแต่ละคน

การวัดความเกี่ยวข้องกันระหว่างลูกค้า α และลูกค้า i สามารถวัดได้จากการคำนวณความ

เกี่ยวข้องกันของบุคคล(Person Correlation $w(\alpha, i)$)

$$w(a, i) = \frac{\sum_j w_j^2 (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j w_j^2 (v_{a,j} - \bar{v}_a)^2 \sum_j w_j^2 (v_{i,j} - \bar{v}_i)^2}} \quad (2.2)$$

w_j = ค่าถ่วงน้ำหนักวัดความเกี่ยวพันระหว่างสินค้า j

จากสมการ 2.2 จะเห็นว่ามีการปรับปรุงสมการเพื่อเพิ่มประสิทธิภาพให้มีความแม่นยำมากขึ้นโดยเพิ่มค่าถ่วงน้ำหนัก w_j เข้าไปในสมการ 2.2 ซึ่งเป็นค่าถ่วงน้ำหนักที่ใช้วัดความเกี่ยวพันระหว่างสินค้า (Mutual Information) ที่ต้องการแนะนำให้กับลูกค้ากับกลุ่มของสินค้าที่ถูกค้าคนอื่นๆ มีความชอบร่วมกันซึ่งค่าถ่วงน้ำหนัก w_j หาได้จากสูตรการคำนวณดังนี้

$$w_j = I(V_j; V_t) \quad (2.3)$$

โดยที่

$$I(V_j; V_t) = H(V_j) + H(V_t) - H(V_j, V_t) \quad (2.4)$$

$H(V_j)$ คือ Entropy ของการให้คะแนนความชอบของ Product j และ $H(V_j, V_t)$ คือ Entropy ของการให้คะแนนความชอบร่วมกันระหว่าง Product j และ Product t ซึ่ง Entropy หาได้จากสูตรการคำนวณดังนี้

$$H_j = -\sum p_{i,j} \cdot \log_2 p_{i,j} \quad (2.5)$$

$p_{i,j}$ = ค่าความน่าจะเป็นที่ถูกค้าให้คะแนนแต่ละระดับคะแนนต่อสินค้า

2.9 Selecting Relevant Instances

อย่างไรก็ตามอุปสรรคพื้นฐานที่สำคัญของเทคนิค CF ซึ่งอุปสรรคแรก คือการขยายตัวเพิ่มขึ้นของจำนวนข้อมูลลูกค้าที่มีมากขึ้นซึ่งปัจจุบันเทคนิค CF สามารถประมวลผลจำนวนข้อมูลลูกค้าได้ประมาณพันคนในเวลาการประมวลผลระดับหนึ่ง แต่ระบบ e-commerce สมัยใหม่จะต้องรองรับข้อมูลของลูกค้าได้เป็นจำนวนหลายสิบล้านอาจจะทำให้เกิดปัญหาใช้เวลาในการประมวลผลนานเกินไป

อุปสรรคที่สองการปรับปรุงคุณภาพการแนะนำสินค้าให้กับลูกค้า โดยลูกค้าสามารถเชื่อมโยงได้กับการแนะนำสินค้าจากระบบที่จะช่วยลูกค้าหาสินค้าที่ลูกค้าชอบ ซึ่งถ้าลูกค้าเชื่อมโยงระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แนะนำสินค้าแล้วซื้อสินค้านั้นแต่กลับเป็นสินค้าที่ถูกค้าไม่มีความชอบเลยลูกค้าก็จะไม่ใช่ระบบแนะนำสินค้านั้นอีกต่อไป

จากอุปสรรคที่กล่าวถึงข้างต้นในเรื่องของปัญหาของเทคนิค CF ถ้าเกิดมีจำนวนข้อมูลลูกค้าเพิ่มขึ้นเป็นจำนวนมากจะเพิ่มความเร็วในการทำนายความชอบของลูกค้าต่อสินค้านั้นได้โดยการเตรียมข้อมูลด้วยวิธีการลดจำนวนข้อมูลลูกค้าที่จะใช้ในการทำนายลงซึ่งจะเลือกกลุ่มของลูกค้าที่มีความเกี่ยวพันเหมาะสมที่สุดกับสินค้าที่ต้องการแนะนำให้กับลูกค้าที่เยี่ยมชมเป็นจำนวน n คน โดยหาได้จากสูตรคำนวณดังนี้

$$R_{i,t} = \log n_i \cdot \frac{\sum_{j \in M_i, j \in t} I(V_j; V_t)}{n_i - 1} \quad (2.6)$$

$R_{i,t}$ = ค่าความเกี่ยวพันที่เหมาะสมกับสินค้าที่ต้องการแนะนำให้กับลูกค้าที่เยี่ยมชม

n_i = จำนวนคะแนนที่ลูกค้าแต่ละคนให้คะแนน

M_i = กลุ่มของสินค้าที่ให้คะแนนโดยลูกค้าแต่ละคน

โดยค่าที่หาได้นำมาเรียงตามลำดับจากค่าที่มีความเหมาะสมมากที่สุดระหว่างลูกค้าต่อสินค้าที่ต้องการแนะนำให้กับลูกค้าที่เยี่ยมชมเป็นจำนวน $k\%$ คนแล้วนำไปใช้ในการทำนายความชอบของลูกค้าต่อสินค้าที่ต้องการแนะนำ ส่วนจำนวนที่เหลือ $(1-k)\%$ คนที่มีความเหมาะสมน้อยระหว่างลูกค้าต่อสินค้าที่ต้องการแนะนำให้กับลูกค้าที่เยี่ยมชมนำออกจากกลุ่มลูกค้าที่นำไปใช้ในการทำนาย

2.10 Metrics and Methodology

เทคนิควิธีการวัดระบบว่ามีความถูกต้องแม่นยำในการทำนายความชอบของลูกค้าที่เยี่ยมชมต่อสินค้าซึ่งจะใช้เทคนิค(Mean Absolute Error: MAE) ที่หาจากค่าผลต่างระหว่างค่าที่ทำนายได้กับค่าจริงที่ถูกค้าให้คะแนนความชอบเอาไว้เพื่อใช้หาคุณภาพของการทำนาย หาได้จากสูตรคำนวณดังนี้

$$MAE = \frac{\sum_{a \in T} |v_{a,j} - p_{a,j}|}{|T|} \quad (2.7)$$

$v_{a,j}$ = ค่าคะแนนความชอบที่ถูกค้าในกลุ่ม Test Set ให้ไว้

$p_{a,j}$ = ค่าคะแนนความชอบที่ทำนายไว้

T = กลุ่มลูกค้าใน Test Set

$|T|$ = จำนวนลูกค้าใน Test Set

วิธีการในการวัด MAE คือทำการแบ่งกลุ่มข้อมูลของลูกค้าที่ได้ให้คะแนนความชอบไว้ โดยแบ่งกลุ่มข้อมูลลูกค้าจำนวน 80% ไว้เป็น Training Set และ 20% ไว้เป็น Test Set และทำการสุ่มเลือกคะแนนความชอบขึ้นมาจำนวนหนึ่งคะแนนจากลูกค้าแต่ละคนใน Test Set และพยายามทำนายคะแนนความชอบจากคะแนนที่สุ่มขึ้นมาจากลูกค้าแต่ละคนแล้วหาค่า MAE ของทุกคนส่วนคะแนนที่เหลือที่ไม่ได้ทำการสุ่มขึ้นมาเพื่อทำนายก็นำไปใช้เป็นส่วนคำนวณในการทำนายค่า



บทที่ 3

การออกแบบระบบแนะนำภาพยนตร์

โครงการนี้สนับสนุนการพัฒนากระบบแนะนำภาพยนตร์โดยใช้เทคนิค Collaborative filtering การทำงานของระบบแนะนำภาพยนตร์จะทำการเลือกกลุ่มภาพยนตร์ที่ต้องการแนะนำให้กับลูกค้าเป้าหมายแล้วพิจารณาว่ากลุ่มภาพยนตร์นี้มีลูกค้าคนอื่น ๆ เคยชมและให้คะแนนความชอบไว้และนำไปใช้เป็นพื้นฐานในการทำนายความชอบของลูกค้าเป้าหมายเพื่อที่จะทำการแนะนำภาพยนตร์ผ่านทางเว็บไซต์ต่อไป และแก้ปัญหาจำนวนของลูกค้าที่เพิ่มขึ้นอันมีผลทำให้ประสิทธิภาพการทำงานของระบบแนะนำภาพยนตร์ช้าลง โดยทำการปรับปรุงคุณภาพในการทำงานคือลดจำนวนลูกค้าลงด้วยวิธี Selecting Relevant Instances ซึ่งทำให้การทำงานของระบบแนะนำภาพยนตร์เร็วขึ้นและปรับปรุงคุณภาพในการแนะนำภาพยนตร์ให้เพิ่มขึ้นอีกด้วย

3.1 Use Case Diagram

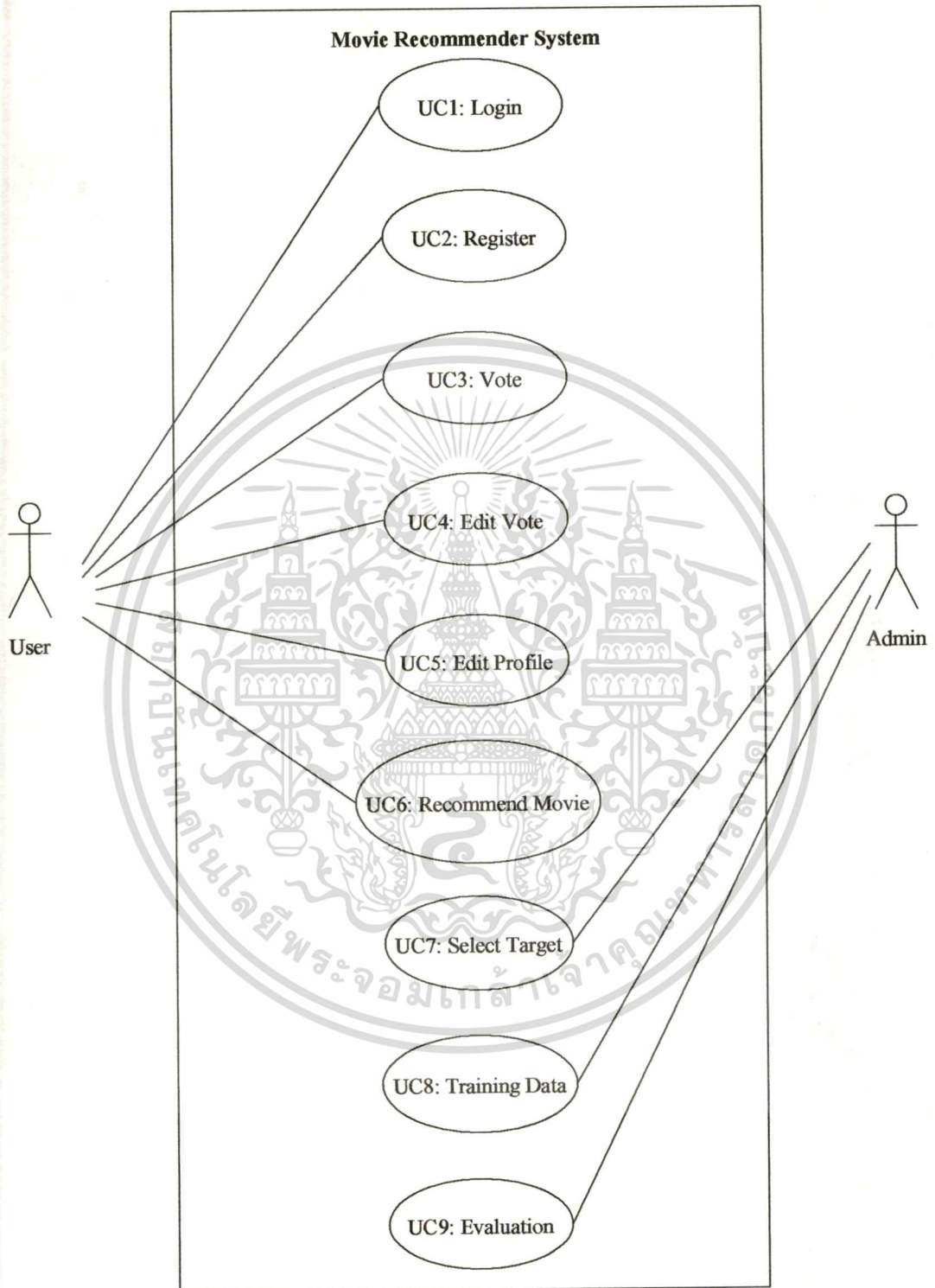
Use Case Diagram เป็น Diagram ที่แสดงภาพรวมการทำงานของระบบ ในระบบแนะนำภาพยนตร์นี้ จะให้บริการแนะนำภาพยนตร์แก่ผู้เยี่ยมชมเว็บไซต์โดยอาศัยคะแนนความชอบที่ผู้เยี่ยมชมเว็บไซต์ได้ให้ไว้กับระบบ ซึ่งมี Actor ที่เกี่ยวข้องดังต่อไปนี้

1. ผู้ดูแลระบบ (Admin) คือผู้ที่ทำการเลือกกลุ่มภาพยนตร์และเตรียมข้อมูลของกลุ่มภาพยนตร์เพื่อที่จะทำการแนะนำให้กับผู้เยี่ยมชมและวัดประสิทธิภาพของระบบแนะนำโดยการทำนายคะแนนความชอบของกลุ่มลูกค้าทดลองเทียบกับค่าคะแนนความชอบที่กลุ่มลูกค้าทดลองให้ไว้หาเป็นค่า Error ออกมา
2. ผู้เยี่ยมชม (User) คือผู้ที่เข้ามาเยี่ยมชมและร้องขอการแนะนำภาพยนตร์จากระบบผ่านเว็บไซต์ซึ่งผู้เยี่ยมชมจะเข้ามาใช้ระบบได้ต้องทำการ Register กับระบบและทำการ Login เข้าใช้งานโดยผู้เยี่ยมชมสามารถให้คะแนนความชอบกับภาพยนตร์แต่ละเรื่อง แก่ใจคะแนนความชอบที่เคยให้ไว้ และแก้ไขข้อมูลส่วนตัวของผู้เยี่ยมชมเอง โดยระบบจะทำการแนะนำภาพยนตร์ให้กับผู้เยี่ยมชม

จากการศึกษาระบบแนะนำภาพยนตร์ นำมาออกแบบระบบและเขียนเป็น Use Case Diagram ได้ดังรูปที่ 3.1 โดยระบบจะแยกออกเป็นสองส่วนหลัก คือส่วนที่หนึ่งส่วนของเว็บไซต์ที่ทำหน้าที่ติดต่อผู้เยี่ยมชมซึ่งให้ผู้เยี่ยมชมสามารถให้คะแนนความชอบต่อภาพยนตร์แต่ละเรื่องและแนะนำภาพยนตร์ให้กับผู้เยี่ยมชมที่เข้าไปให้คะแนนความชอบ ส่วนที่สองเป็นส่วนของการเตรียมข้อมูลภาพยนตร์ที่นำไปแนะนำให้กับผู้เยี่ยมชม เพื่อใช้ในการสร้างการแนะนำให้กับผู้เยี่ยมชม ทั้ง

เอกสารนี้เป็นเอกสารของบริษัทเอกชน ไม่สามารถเผยแพร่ได้โดยไม่ได้รับอนุญาต
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Register, Vote, Edit Vote, Edit Profile, Recommend Movie, Select Target, Training Data และ Evaluation



รูปที่ 3.1 Use Case Diagram ระบบแนะนำภาพยนตร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 3.1 แสดงการทำงาน โดยผู้เยี่ยมชมจะเข้ามาใช้ระบบแนะนำภาพยนตร์ผ่านเว็บไซต์จะต้องทำการ Login เข้าสู่ระบบก่อนซึ่งการที่จะ Login ได้นั้นผู้เยี่ยมชมต้องทำการลงทะเบียนกับระบบเพื่อให้มีสิทธิ์เข้าใช้ระบบ เมื่อผู้เยี่ยมชม Login เข้ามาในระบบแล้วระบบจะแสดงรายการภาพยนตร์ให้ผู้เยี่ยมชม ภาพยนตร์เรื่องไหนที่ผู้เยี่ยมชมเคยชมสามารถให้คะแนนความชอบโดยระดับคะแนนความชอบจะมีค่า 1-5 ตามระดับความชอบน้อยไปมากตามลำดับ ซึ่งผู้เยี่ยมชมสามารถแก้ไขคะแนนความชอบที่เคยให้ไว้แล้วได้ และสามารถแก้ไขประวัติของผู้เยี่ยมชมเปลี่ยนรหัสผ่าน เมื่อผู้เยี่ยมชมให้คะแนนความชอบเรียบร้อยแล้วระบบก็จะทำการแนะนำภาพยนตร์ให้กับผู้เยี่ยมชมโดยอาศัยคะแนนความชอบที่ผู้ชมให้ไว้นำไปสร้างการแนะนำ

การทำงานในส่วนของผู้ดูแลระบบจะทำการเลือกกลุ่มภาพยนตร์ที่จะแนะนำนำไปเตรียมข้อมูลเพื่อหาค่า High Relevant ของผู้ที่มีความเกี่ยวข้องกับภาพยนตร์ที่แนะนำและเลือก คนที่มีค่า Relevant สูงสุดจำนวน K เพื่อใช้ในการสร้างการแนะนำให้กับผู้เยี่ยมชม

ตารางที่ 3.1 แสดงคำอธิบาย Use Case Diagram ของ Login

Use Case Name	UC1 Login
Use Case Description	การ Login เข้าใช้งานระบบแนะนำภาพยนตร์
Actor	ผู้เยี่ยมชมเว็บไซต์
Trigger	ผู้เยี่ยมชมเว็บไซต์เลือกเมนู Login
Pre-Condition	ผู้เยี่ยมชมเว็บไซต์ทำการลงทะเบียนเพื่อสร้างสิทธิ์ในการใช้งานระบบแนะนำภาพยนตร์และพิสูจน์ตัวตนจริง
Post-Condition	ผู้เยี่ยมชมเว็บไซต์สามารถเข้าใช้งานระบบแนะนำภาพยนตร์ได้
Basic Flow	1) ผู้เยี่ยมชมกรอก Username และ Password 2) ระบบตรวจสอบ Username และ Password ถูกต้อง 3) ระบบอนุญาตให้เข้าใช้งานระบบแนะนำภาพยนตร์
Alternative Flow	2a) ระบบตรวจสอบ Username และ Password ไม่ถูกต้อง : ระบบแสดงข้อความเตือน

ตารางที่ 3.2 แสดงคำอธิบาย Use Case Diagram ของ Register

Use Case Name	UC2 Register
Use Case Description	การลงทะเบียนเพื่อสร้างสิทธิ์ในการใช้งานระบบและพิสูจน์ตัวตนจริง
Actor	ผู้เยี่ยมชมเว็บไซต์
Trigger	ผู้เยี่ยมชมเว็บไซต์เลือกเมนูลงทะเบียน
Pre-Condition	-
Post-Condition	ผู้เยี่ยมชมเว็บไซต์ลงทะเบียนได้
Basic Flow	<ol style="list-style-type: none"> 1) กรอกข้อมูลผู้เยี่ยมชม 2) ระบบตรวจสอบความถูกต้องของข้อมูลผู้เยี่ยมชม เช่น ชื่อ-สกุล เพศ อายุ อี-เมล์ รหัสไปรษณีย์ Username และ Password ครบถ้วน 3) ระบบตรวจสอบว่า Username ที่ผู้เยี่ยมชมกรอกไม่มีใครใช้ 4) บันทึกลงฐานข้อมูล
Alternative Flow	<p>2a) ระบบตรวจสอบข้อมูลของผู้เยี่ยมชมกรอกไม่ถูกต้องครบถ้วน: ระบบแสดงข้อความเตือน</p> <p>3a) ระบบตรวจสอบว่า Username ที่ผู้เยี่ยมชมกรอกมีคนอื่นใช้อยู่ แล้ว: ระบบแสดงข้อความเตือน</p>

ตารางที่ 3.3 แสดงคำอธิบาย Use Case Diagram ของ Vote

Use Case Name	UC3 Vote
Use Case Description	ลงคะแนนความชอบให้กับภาพยนตร์แต่ละเรื่อง
Actor	ผู้เยี่ยมชมเว็บไซต์
Trigger	ผู้เยี่ยมชมเว็บไซต์เลือกเมนู Add Vote
Pre-Condition	ผู้เยี่ยมชมเว็บไซต์ทำการ Login เข้าใช้งานระบบ
Post-Condition	ผู้เยี่ยมชมเว็บไซต์ลงคะแนนความชอบภาพยนตร์
Basic Flow	<ol style="list-style-type: none"> 1) ผู้เยี่ยมชมเลือกรายการภาพยนตร์ที่ต้องการลงคะแนนให้ 2) ผู้เยี่ยมชมเลือกระดับคะแนนความชอบต่อภาพยนตร์ที่ต้องการลงคะแนนให้ 3) ผู้เยี่ยมชมยืนยันการลงคะแนนต่อภาพยนตร์ที่ต้องการลงคะแนน 4) ผู้เยี่ยมชมกดปุ่ม Finish เพื่อบันทึกคะแนนทั้งที่ได้ลงคะแนนไว้ลงฐานข้อมูล
Alternative Flow	3a) ผู้เยี่ยมชมยกเลิกการลงคะแนน: เลือกรายการภาพยนตร์ใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 3.4 แสดงคำอธิบาย Use Case Diagram ของ Edit Vote

Use Case Name	UC4 Edit Vote
Use Case Description	แก้ไขคะแนนความชอบที่เคยให้กับภาพยนตร์แต่ละเรื่อง
Actor	ผู้เยี่ยมชมเว็บไซต์
Trigger	ผู้เยี่ยมชมเว็บไซต์เลือกเมนู Edit Vote
Pre-Condition	ผู้เยี่ยมชมเว็บไซต์ทำการ Login เข้าใช้งานระบบ
Post-Condition	ผู้เยี่ยมชมเว็บไซต์แก้ไขคะแนนความชอบภาพยนตร์ได้
Basic Flow	<ol style="list-style-type: none"> 1) ผู้เยี่ยมชมเลือกรายการภาพยนตร์ที่ต้องการแก้ไขคะแนน 2) ผู้เยี่ยมชมเลือกระดับคะแนนความชอบที่ต้องการแก้ไขกับภาพยนตร์ที่เคยให้คะแนนไว้ 3) ผู้เยี่ยมชมยืนยันการแก้ไขคะแนนต่อภาพยนตร์ที่เคยให้คะแนนไว้
Alternative Flow	3a) ผู้เยี่ยมชมยกเลิกการแก้ไขคะแนน: เลือกรายการภาพยนตร์ใหม่

ตารางที่ 3.5 แสดงคำอธิบาย Use Case Diagram ของ Edit Profile

Use Case Name	UC5 Edit Profile
Use Case Description	ผู้เยี่ยมชมเว็บไซต์แก้ไขข้อมูลส่วนตัวและ password ของตนเอง
Actor	ผู้เยี่ยมชมเว็บไซต์
Trigger	ผู้เยี่ยมชมเว็บไซต์เลือกเมนู Edit Profile
Pre-Condition	ผู้เยี่ยมชมเว็บไซต์ทำการ Login เข้าใช้งานระบบ
Post-Condition	ผู้เยี่ยมชมเว็บไซต์แก้ไขข้อมูลส่วนตัวได้สำเร็จ
Basic Flow	<ol style="list-style-type: none"> 1) ผู้เยี่ยมชมแก้ไขข้อมูลส่วนตัวเช่น ชื่อ-สกุล เพศ อายุ อี-เมลล์ รหัสไปรษณีย์ 2) ระบบตรวจสอบความถูกต้องของข้อมูลผู้เยี่ยมชม เช่น ชื่อ-สกุล เพศ อายุ อี-เมลล์ รหัสไปรษณีย์ ครบถ้วน 3) ผู้เยี่ยมชมกรอก password เก่า 4) ผู้เยี่ยมชมกรอก password ใหม่ 5) ระบบตรวจสอบ password เก่าถูกต้อง 6) ระบบทำการบันทึกข้อมูลที่เปลี่ยนแปลง
Alternative Flow	<p>2a) ระบบตรวจสอบข้อมูลที่ผู้เยี่ยมชมกรอกไม่ถูกต้องครบถ้วน: ระบบแสดงข้อความเตือน</p> <p>5a) ระบบตรวจสอบ password ไม่ถูกต้อง: แสดงข้อความแจ้งเตือน</p>

ตารางที่ 3.6 แสดงคำอธิบาย Use Case Diagram ของ Recommend Movie

Use Case Name	UC6 Recommend Movie
Use Case Description	ระบบทำการแนะนำภาพยนตร์ให้กับผู้เยี่ยมชม
Actor	ผู้เยี่ยมชมเว็บไซต์
Trigger	ผู้เยี่ยมชมเว็บไซต์เลือกเมนู Recommend
Pre-Condition	ผู้เยี่ยมชมเว็บไซต์ทำการ Login เข้าใช้งานระบบ
Post-Condition	ระบบแนะนำภาพยนตร์ให้กับผู้เยี่ยมชม
Basic Flow	1) ระบบแนะนำภาพยนตร์ให้ผู้เยี่ยมชมตามลำดับความชอบมากที่สุด ไปน้อยสุด
Alternative Flow	-

ตารางที่ 3.7 แสดงคำอธิบาย Use Case Diagram ของ Select Target

Use Case Name	UC7 Select Target
Use Case Description	ผู้ดูแลระบบทำการเลือกกลุ่มภาพยนตร์ที่ต้องแนะนำให้กับผู้เยี่ยมชม เว็บไซต์
Actor	ผู้ดูแลระบบ
Trigger	ผู้ดูแลระบบเลือกเมนู Select Target
Pre-Condition	ระบบทำการเชื่อมต่อนฐานข้อมูล
Post-Condition	ได้กลุ่มภาพยนตร์ที่นำไปแนะนำให้กับผู้เยี่ยมชมเว็บไซต์
Basic Flow	1) ผู้ดูแลระบบเลือกประเภทภาพยนตร์ 2) ผู้ดูแลระบบเลือกช่วงปีที่ภาพยนตร์ออกฉาย 3) ระบบแสดงรายการภาพยนตร์ตามประเภทและปีที่ออกฉาย 4) ผู้ดูแลระบบเลือกภาพยนตร์จากรายการภาพยนตร์ที่ระบบแสดง นำไปใช้แนะนำให้กับผู้เยี่ยมชมเว็บไซต์ 5) ระบบบันทึกกลุ่มภาพยนตร์ที่ได้เลือกไว้
Alternative Flow	1a) ผู้ดูแลระบบไม่ได้เลือกประเภทภาพยนตร์: ระบบแสดงข้อความ เตือน

ตารางที่ 3.8 แสดงคำอธิบาย Use Case Diagram ของ Training Data

Use Case Name	UC8 Training Data
Use Case Description	ผู้ดูแลระบบทำการ Training Data
Actor	ผู้ดูแลระบบ
Trigger	ผู้ดูแลระบบเลือกเมนู Training Data
Pre-Condition	ผู้ดูแลระบบเลือกกลุ่มภาพยนตร์ที่ต้องการแนะนำให้ผู้เยี่ยมชม
Post-Condition	ได้ Training Data
Basic Flow	<ol style="list-style-type: none"> 1) ผู้ดูแลระบบทำการ Training Data 2) ระบบแสดงกลุ่ม user ที่ high relevant กับกลุ่มภาพยนตร์แต่ละเรื่องที่น่าไปแนะนำให้กับผู้เยี่ยมชม 3) ผู้ดูแลระบบเลือก user ที่ high relevant กับกลุ่มภาพยนตร์แต่ละเรื่องที่น่าไปแนะนำให้กับผู้เยี่ยมชมเป็นจำนวน Top-K เพื่อที่จะไปใช้ในการทำนายความชอบ
Alternative Flow	-

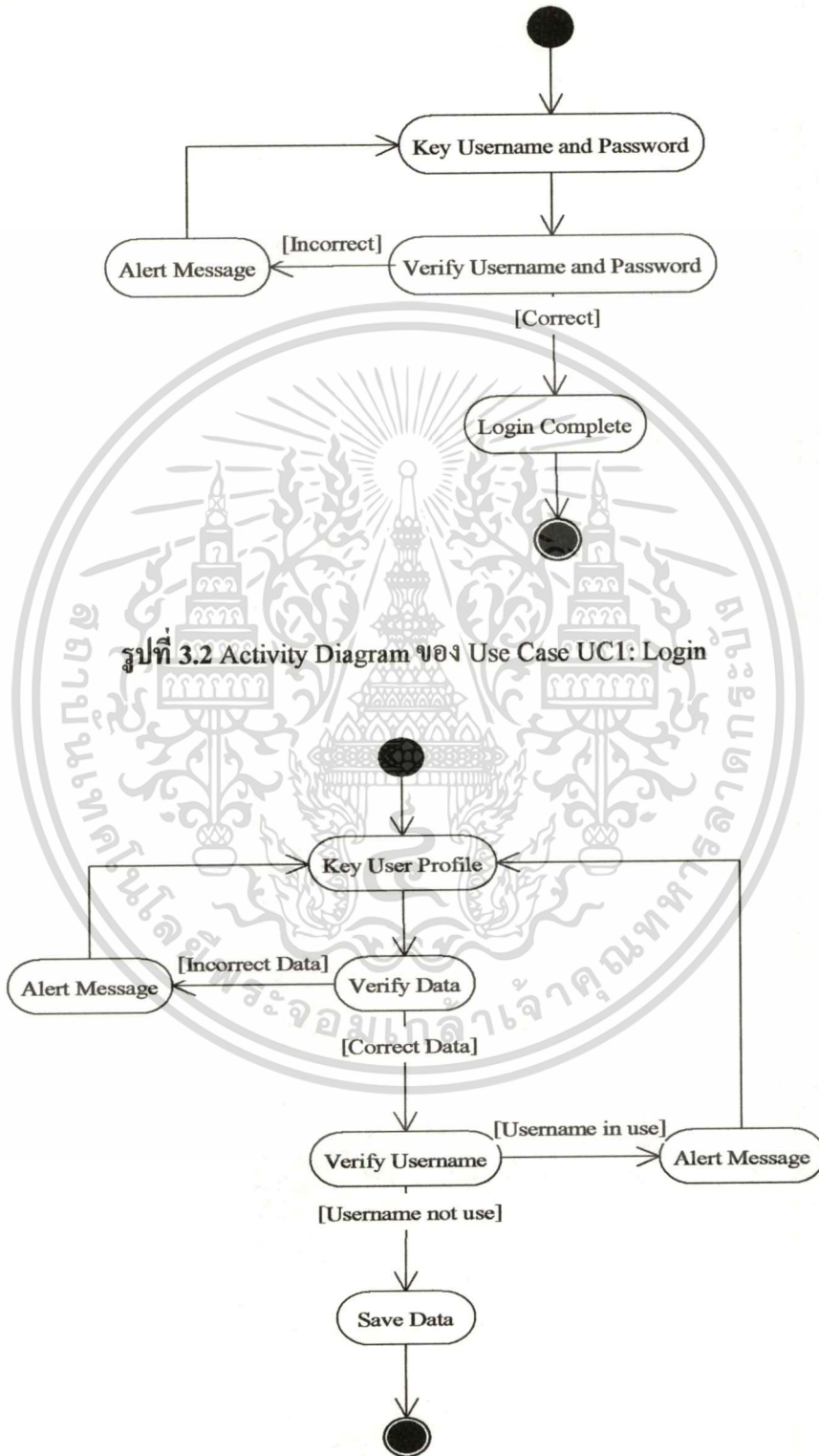
ตารางที่ 3.9 แสดงคำอธิบาย Use Case Diagram ของ Evaluation

Use Case Name	UC9 Evaluation
Use Case Description	ระบบทำการวัดประสิทธิภาพในการทำนายความชอบ
Actor	ผู้ดูแลระบบ
Trigger	ผู้ดูแลระบบเลือกเมนู MAE(Mean Absolute Error)
Pre-Condition	ระบบทำการ Training Data
Post-Condition	ระบบแสดงผลค่า MAE ที่ได้จากการวัดประสิทธิภาพ
Basic Flow	<ol style="list-style-type: none"> 1) ระบบทำการสุ่มเลือกคะแนนความชอบขึ้นมาหนึ่งคะแนนจากกลุ่มลูกค้าทดสอบแต่ละคน 2) ระบบทำการทำนายคะแนนความชอบที่สุ่มขึ้นมาจากกลุ่มลูกค้าทดลองแต่ละคน 3) ระบบคำนวณหาค่าเฉลี่ย Error แสดงออกมาหน้าจอ
Alternative Flow	-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 Activity Diagram

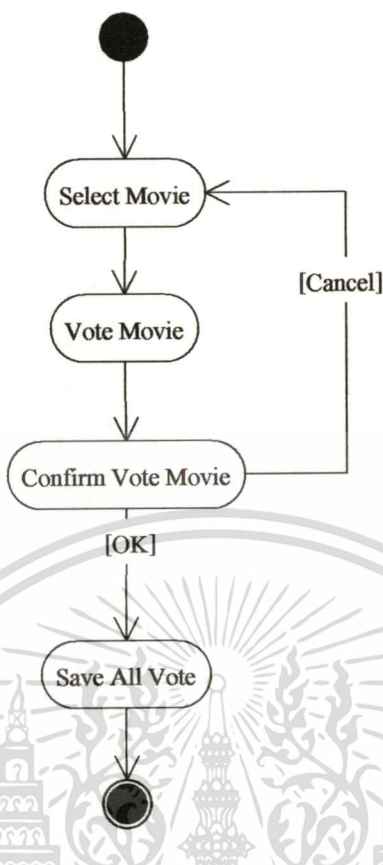
Activity Diagram นำมาใช้ในการอธิบายขั้นตอนการทำงานของ Use Case ต่างๆ โดยได้ออกแบบ Activity Diagram ของ Use Case ระบบแนะนำภาพยนตร์ดังรูปที่ 3.2 ถึง 3.10



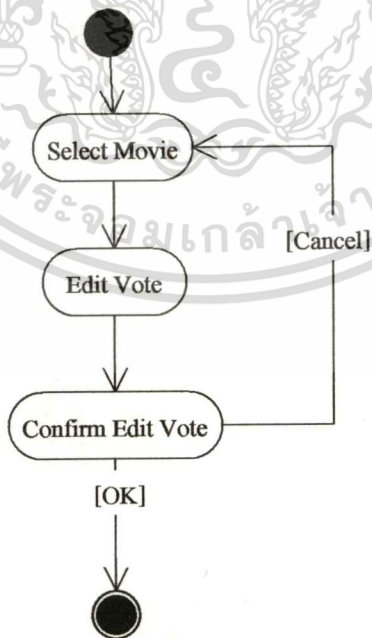
รูปที่ 3.2 Activity Diagram ของ Use Case UC1: Login

รูปที่ 3.3 Activity Diagram ของ Use Case UC2: Register

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นอนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

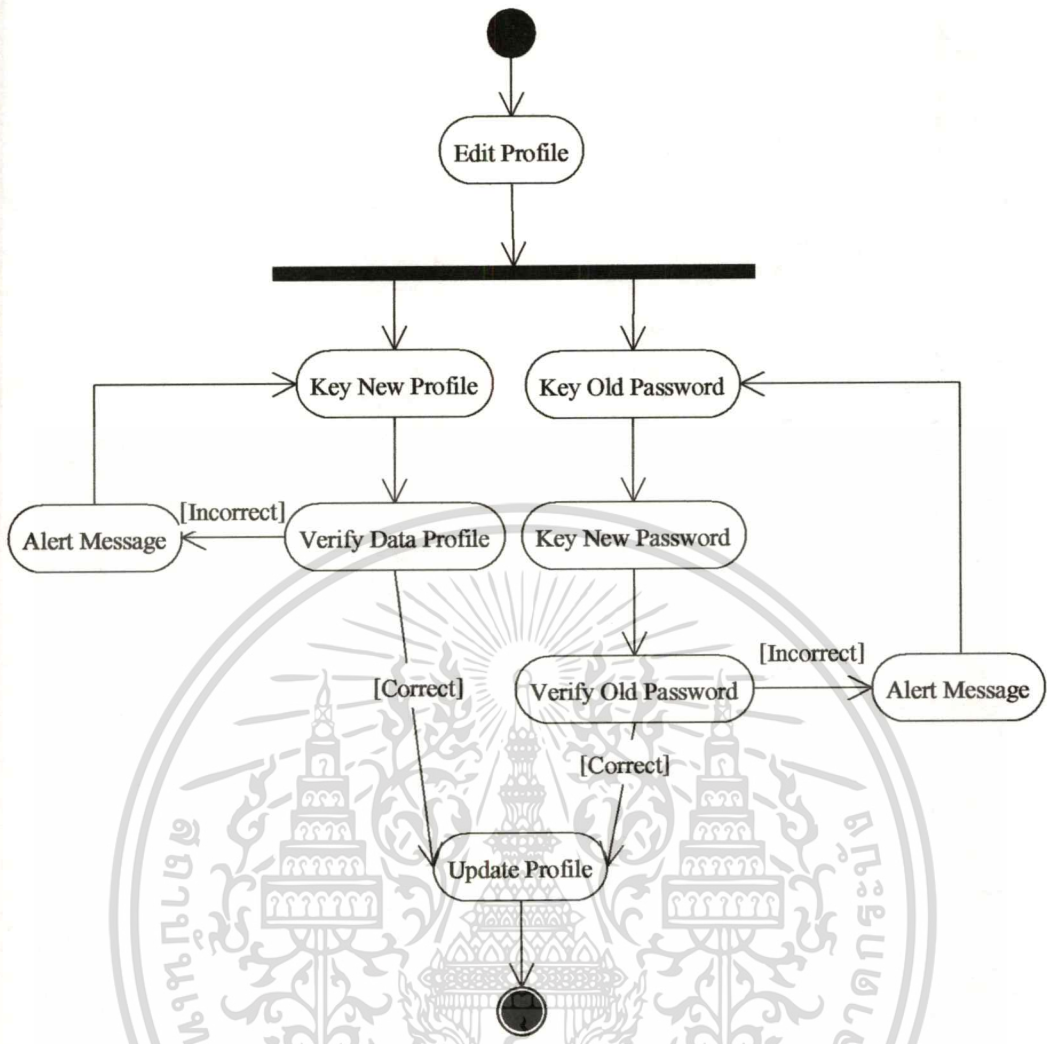


รูปที่ 3.4 Activity Diagram ของ Use Case UC3: Vote

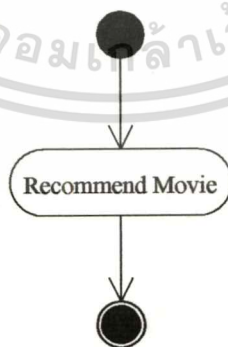


รูปที่ 3.5 Activity Diagram ของ Use Case UC4: Edit Vote

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

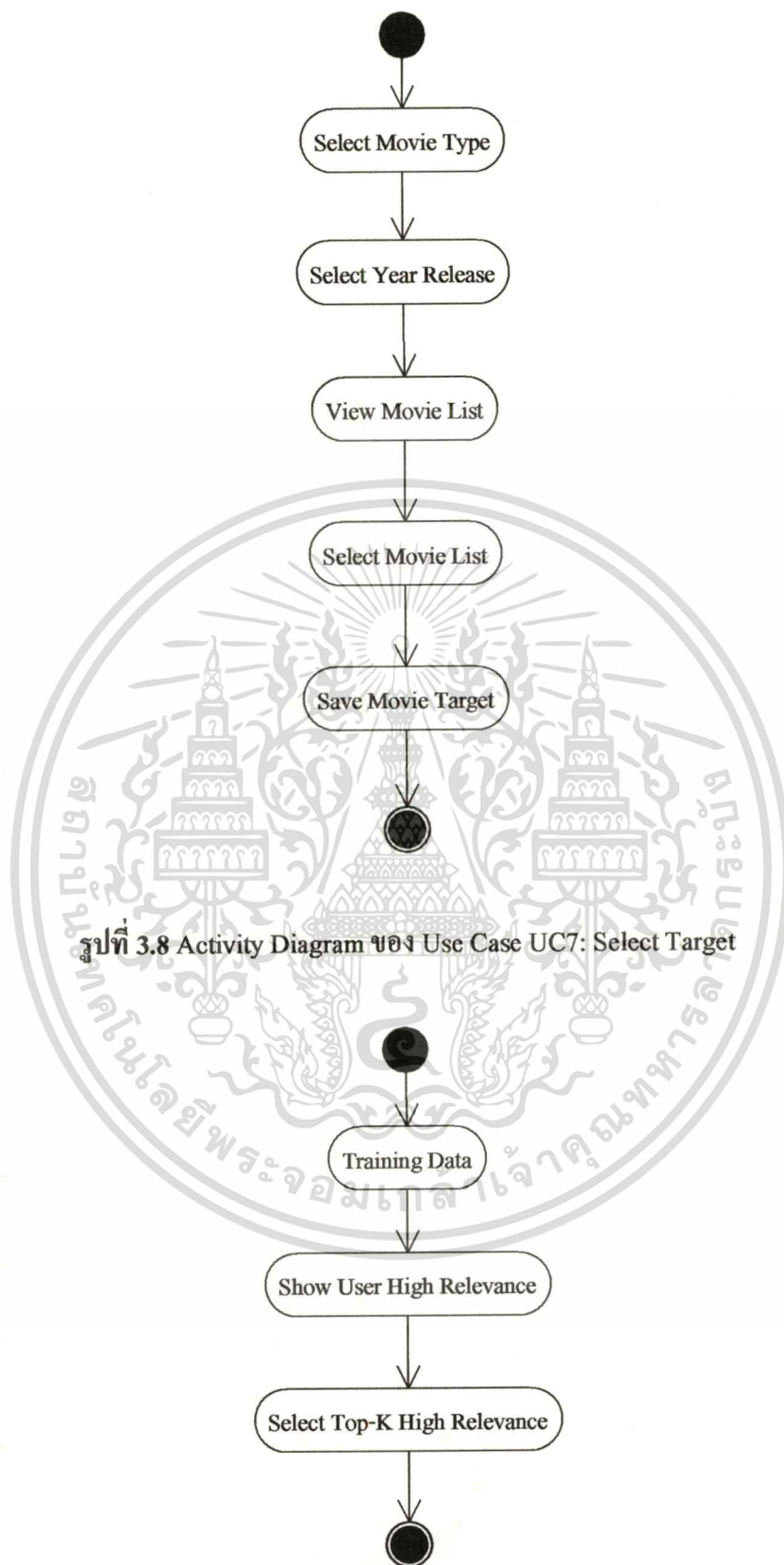


รูปที่ 3.6 Activity Diagram ของ Use Case UC5: Edit Profile



รูปที่ 3.7 Activity Diagram ของ Use Case UC6: Recommend Movie

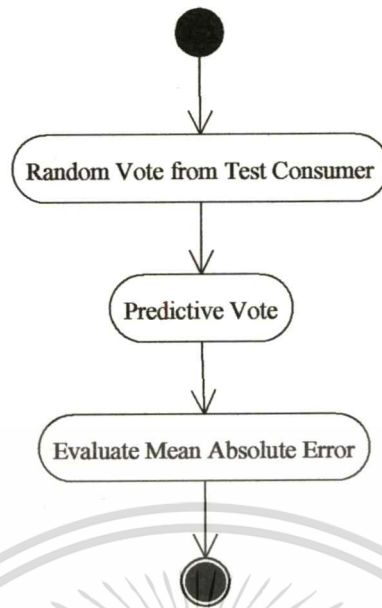
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.8 Activity Diagram ของ Use Case UC7: Select Target

รูปที่ 3.9 Activity Diagram ของ Use Case UC8: Training Data

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.10 Activity Diagram ของ Use Case UC9: Evaluate

3.3 Class Diagram

รายละเอียดของ Class Diagram ของระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์ ประกอบด้วย Class ทั้งหมด 5 Class ดังรูปที่ 3.11 ดังนี้

1. Movie คือ ภาพยนตร์ที่แสดงให้กับผู้เยี่ยมชม
2. User คือ ผู้เยี่ยมชมเว็บไซต์แนะนำภาพยนตร์
3. Rate คือ คะแนนความชอบที่ผู้เยี่ยมชมให้กับภาพยนตร์แต่ละเรื่อง
4. Recommendation คือ การสร้างการแนะนำภาพยนตร์
5. MeanError คือค่าเฉลี่ย Error ที่ได้จากการวัดประสิทธิภาพของการแนะนำภาพยนตร์ โดยแต่ละ Class มีความสัมพันธ์ระหว่าง Class ดังนี้

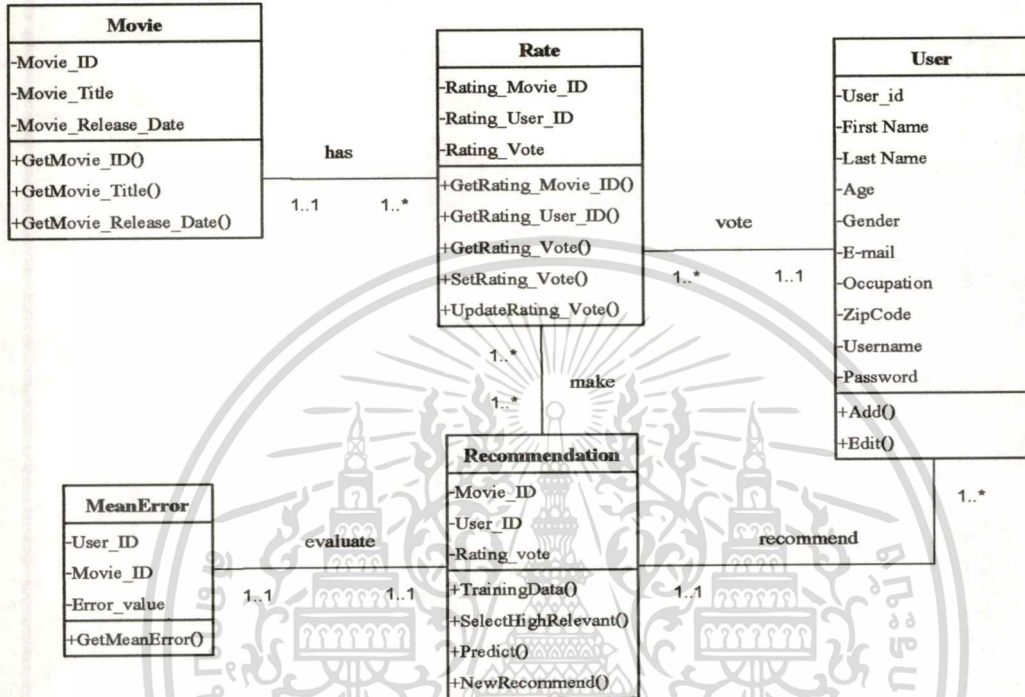
1. Class User สัมพันธ์กับ Class Rate โดยผู้เยี่ยมชมหนึ่งคนสามารถให้คะแนนความชอบได้หลายคะแนนแต่คะแนนหนึ่งคะแนนสามารถให้ผู้เยี่ยมชมลงคะแนนได้หนึ่งคน

2. Class Movie สัมพันธ์กับ Class Rate โดยภาพยนตร์หนึ่งเรื่องสามารถได้รับคะแนนความชอบหลายคะแนนแต่คะแนนหนึ่งคะแนนสามารถลงให้กับภาพยนตร์ได้หนึ่งเรื่อง

3. Class Rate สัมพันธ์กับ Class Recommendation โดยคะแนนความชอบหลายคะแนนสามารถนำไปสร้างรายการภาพยนตร์แนะนำได้หลายรายการ

4. Class Recommendation สัมพันธ์กับ Class MeanError โดยรายการภาพยนตร์แนะนำหนึ่งรายการสามารถนำไปหาค่าเฉลี่ย Error ได้หนึ่งค่า

5. Class Recommendation สัมพันธ์กับ Class User โดยรายการภาพยนตร์แนะนำหนึ่งรายการสามารถนำไปแนะนำให้กับผู้เยี่ยมชมได้หลายคนแต่ผู้เยี่ยมชมหนึ่งคนสามารถได้รับการแนะนำภาพยนตร์ได้หนึ่งรายการ



รูปที่ 3.11 Class Diagram ของระบบแนะนำภาพยนตร์

3.4 การเตรียมข้อมูล

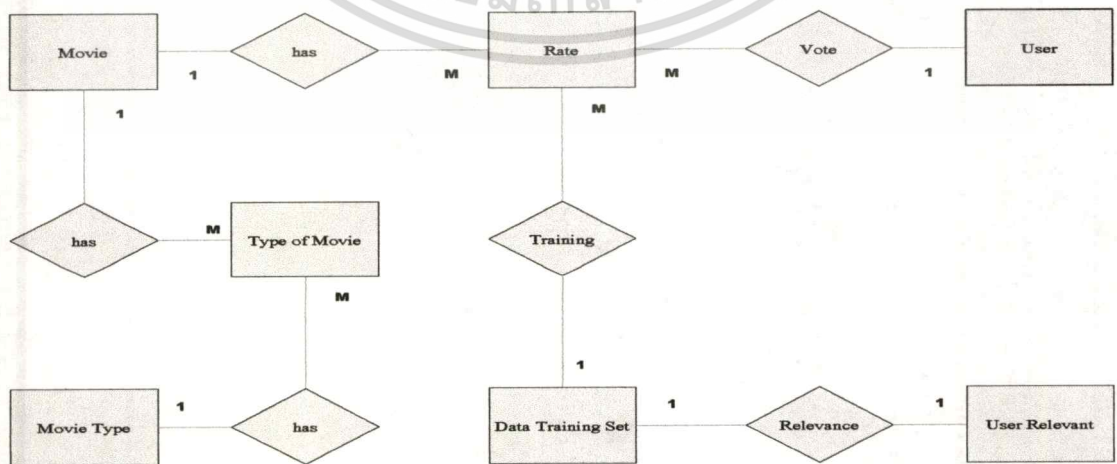
ข้อมูลที่ใช้ในการพัฒนาระบบงานในครั้งนี้ดาวน์โหลดมาจาก <http://www.grouplens.org/> ซึ่งเก็บรวบรวมจากเว็บไซต์ <http://www.movieLens.unn.edu> ซึ่งเป็นเว็บไซต์แนะนำภาพยนตร์โดยใช้เทคนิค Collaborative filtering โดย GroupLens Research Project ที่ University of Minnesota โดยที่ข้อมูลที่นำมาใช้ในการพัฒนาระบบงานในครั้งนี้เก็บรวบรวมมาจากผู้เข้ามาเยี่ยมชมจำนวน 943 คน ทำการ rating ภาพยนตร์จำนวน 1,682 เรื่อง ซึ่งมีจำนวนการ rating ทั้งหมด 100,000 rating ที่เข้าเยี่ยมชมเว็บไซต์ <http://www.movieLens.unn.edu> ในช่วงปี ค.ศ. 2000

3.5 การออกแบบฐานข้อมูล

ในขั้นตอนของการออกแบบฐานข้อมูลนั้นวัตถุประสงค์ของการออกแบบก็เพื่อให้ได้ฐานข้อมูลที่ถูกต้องและสามารถเปลี่ยนแปลงได้ง่าย โดยสร้างแบบจำลองขึ้นมาเพื่อใช้เป็นเครื่องมือในการแสดงโครงสร้างภายในระบบฐานข้อมูลโดยใช้รูปภาพเป็นสื่อ ทำให้ง่ายต่อการเข้าใจและไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การเรียนรู้โครงสร้างพื้นฐานของระบบ ซึ่งแบบจำลองดังกล่าวก็คือ ER-Diagram ซึ่งใช้แสดงความสัมพันธ์ระหว่างข้อมูลในฐานข้อมูล ดังนั้นจึงได้แสดงความสัมพันธ์ระหว่างข้อมูลในระบบแนะนำภาพยนตร์ดังรูปที่ 3.12 ซึ่งจากแบบจำลองสามารถแปลงให้เป็นตารางในฐานข้อมูลเชิงสัมพันธ์ได้ดังนี้

1. Movie หมายถึง ภาพยนตร์ โดยมีรายละเอียดรหัสภาพยนตร์ ชื่อภาพยนตร์ และวันที่ภาพยนตร์ออกฉาย
2. Movie Type หมายถึง ประเภทภาพยนตร์ โดยมีรายละเอียดรหัสประเภทภาพยนตร์ ชื่อประเภทภาพยนตร์
3. Type of Movie หมายถึง ประเภทของภาพยนตร์ โดยมีรายละเอียดรหัสภาพยนตร์ รหัสประเภทภาพยนตร์
4. Rate หมายถึง คะแนนความชอบ ที่ผู้เยี่ยมชมเว็บไซต์ให้กับภาพยนตร์ โดยมีระดับคะแนน 1 ถึง 5 โดยมีรายละเอียดรหัสผู้เยี่ยมชม รหัสภาพยนตร์ คะแนนความชอบ
5. User หมายถึง ผู้เยี่ยมชมเว็บไซต์ โดยมีรายละเอียดรหัสผู้เยี่ยมชม ชื่อผู้เยี่ยมชม นามสกุล ผู้เยี่ยมชม ชื่อบัญชี รหัสผ่าน อายุ เพศ จดหมายอิเล็กทรอนิกส์ อาชีพ รหัสประเทศ
6. Data Training Set หมายถึง กลุ่มข้อมูลที่ได้รับการ Training จากคะแนนความชอบที่ผู้เยี่ยมชมเว็บไซต์ให้กับภาพยนตร์ โดยมีรายละเอียดเก็บค่าตัวเลขจากการคำนวณค่า ความเกี่ยวข้องระหว่างภาพยนตร์ที่ต้องการแนะนำให้กับลูกค้ากับกลุ่มของภาพยนตร์ที่ถูกค่าคนอื่น ๆ มีความชอบร่วมกัน (Mutual Information) เพื่อใช้ในการคำนวณหา User Relevant ต่อไป
7. User Relevant หมายถึง ผู้เยี่ยมชมเว็บไซต์คนอื่นๆที่เคยให้คะแนนความต่อภาพยนตร์แนะนำ โดยมีรายละเอียดเก็บค่าตัวเลขจากการคำนวณหาความเหมาะสมตรงประเด็นระหว่างภาพยนตร์แนะนำกับผู้เยี่ยมชมเว็บไซต์คนอื่นๆที่เคยให้คะแนนความชอบต่อภาพยนตร์แนะนำ



รูปที่ 3.12 แสดงความสัมพันธ์ระหว่างข้อมูลเชิงสัมพันธ์ของระบบแนะนำภาพยนตร์
 เอกสารนี้เป็นทรัพย์สินทางปัญญาของสถาบันวิจัยและพัฒนาพื้นที่สูง (องค์การมหาชน) ห้ามเผยแพร่โดยไม่ได้รับอนุญาต
 ไม่ว่าจะโดยวิธีใด ทั้งสิ้น อีกทั้งห้ามมิให้คัดลอกเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ชุดของข้อมูลที่ใช้ในการพัฒนาในระบบงานนี้จะประกอบด้วย

1. ข้อมูลรายละเอียดการ rating ของผู้เข้ามาเยี่ยมชม

ตารางที่ 3.10 tbl_Rating แสดงข้อมูลรายละเอียดของ Entity Rate

ชื่อ	คำอธิบาย	ชนิดข้อมูล	คีย์
Rating_user_id	รหัสผู้เยี่ยมชม	int	PK
Rating_movie_id	รหัสภาพยนตร์	int	PK
Rating_vote	คะแนนความชอบ	float	

2. ข้อมูลรายละเอียดของผู้เข้าเยี่ยมชม

ตารางที่ 3.11 tbl_User แสดงข้อมูลรายละเอียดของ Entity User

ชื่อ	คำอธิบาย	ชนิดข้อมูล	คีย์
User_id	รหัสผู้เยี่ยมชม	int	PK
User_Fname	ชื่อ	varchar	
User_Lname	นามสกุล	varchar	
User_Username	ชื่อบัญชี	varchar	
User_Password	รหัสผ่าน	varchar	
User_age	อายุ	int	
User_gender	เพศ	varchar	
User_Email	จดหมายอิเล็กทรอนิกส์	varchar	
User_Occupation	อาชีพ	varchar	
User_Zipcode	รหัสประเทศ	varchar	

3. ข้อมูลรายละเอียดของภาพยนตร์

ตารางที่ 3.12 tbl_Movie แสดงข้อมูลรายละเอียดของ Entity Movie

ชื่อ	คำอธิบาย	ชนิดข้อมูล	คีย์
Movie_id	รหัสภาพยนตร์	int	PK
Movie_title	ชื่อภาพยนตร์	varchar	
Movie_release_date	ปีที่ออกฉาย	datetime	

ตารางที่ 3.13 tbl_Movietype แสดงข้อมูลรายละเอียดของ Entity Movie Type

ชื่อ	คำอธิบาย	ชนิดข้อมูล	คีย์
Movietype_Type_id	รหัสประเภทภาพยนตร์	int	PK
Movietype_Type_des	ชื่อประเภทภาพยนตร์	varchar	

ตารางที่ 3.14 tbl_TypeOfMovie แสดงข้อมูลรายละเอียดของ Entity Type of Movie

ชื่อ	คำอธิบาย	ชนิดข้อมูล	คีย์
Movie_id	รหัสภาพยนตร์	int	PK
Movie_type	รหัสประเภทภาพยนตร์	int	PK

4. ข้อมูลผลลัพธ์จากการ Training Data และหาค่า User Relevant

ตารางที่ 3.15 tbl_Training แสดงข้อมูลรายละเอียดของ Entity Data Training Set

ชื่อ	คำอธิบาย	ชนิดข้อมูล	คีย์
Training_Movie_T	รหัสภาพยนตร์แนะนำ	int	PK
Training_User_Movie	รหัสผู้เยี่ยมชม	int	PK
Training_Movie_J	รหัสภาพยนตร์ที่ผู้เยี่ยมชมให้คะแนนความชอบไว้	int	PK
Training_Mutual	ค่าความเกี่ยวพันระหว่างภาพยนตร์แนะนำกับภาพยนตร์ที่ผู้เยี่ยมชมให้คะแนนความชอบ	float	

ตารางที่ 3.16 tbl_TopRelevant แสดงข้อมูลรายละเอียดของ Entity User Relevant

ชื่อ	คำอธิบาย	ชนิดข้อมูล	คีย์
Target_Movie	รหัสภาพยนตร์แนะนำ	int	PK
User_i	รหัสผู้เยี่ยมชม	int	PK
relevant	ค่าความเหมาะสมตรงประเด็นระหว่างภาพยนตร์แนะนำกับผู้เยี่ยมชม	float	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ประยุกต์เทคนิค Collaborative filtering กับระบบแนะนำ ภาพยนตร์

ในบทที่ 4 นี้จะกล่าวถึงการพัฒนาาระบบแนะนำภาพยนตร์โดยการนำเทคนิค Collaborative filtering มาประยุกต์ใช้ ซึ่งภายในบทนี้จะอธิบายถึงการทำงานของระบบแนะนำภาพยนตร์ และแสดงผลการแนะนำภาพยนตร์ให้กับผู้ที่เข้ามาเยี่ยมชมเว็บไซต์

4.1 กำหนดวัตถุประสงค์

จากการที่ในปัจจุบันการทำธุรกิจ e-commerce บนเครือข่ายอินเทอร์เน็ตได้เกิดขึ้นอย่างมากมาย ซึ่งทางเลือกหนึ่งที่ทำให้ธุรกิจบนเครือข่ายอินเทอร์เน็ตนั้นเป็นที่น่าสนใจก็คือ การนำเสนอและแนะนำทางเลือกสินค้าให้แก่ผู้ซื้อซึ่งจัดเป็นบริการเสริมอย่างหนึ่งที่จะช่วยให้ผู้ซื้อสินค้าลดขั้นตอนและเวลาในการค้นหา จดจำและอำนวยความสะดวกในเรื่องการเลือกสินค้าของผู้ซื้อสินค้าได้เป็นอย่างมาก โดยการพัฒนาะบบงานขึ้นนี้มีวัตถุประสงค์หลักก็เพื่อที่จะช่วยแนะนำสินค้าให้กับผู้ที่เข้ามาเยี่ยมชมเว็บไซต์ โดยการแนะนำนั้นจะมีพื้นฐานมาจากกลุ่มผู้ที่มีลักษณะพฤติกรรมการซื้อสินค้าที่เหมือนกันหรือมีความชอบในสินค้าแบบเดียวกัน โดยที่ในระบบงานที่ได้ทำการพัฒนานี้ใช้ข้อมูลเกี่ยวกับภาพยนตร์มาเป็นสินค้าที่ใช้ทำการศึกษาและพัฒนาะบบแนะนำสินค้าผ่านเว็บไซต์

4.2 เครื่องมือที่ใช้ในการพัฒนาะบบ

การพัฒนาะบบแนะนำสินค้าผ่านเครือข่ายอินเทอร์เน็ตนี้ได้แบ่งการพัฒนาออกเป็น 2 ส่วน ดังต่อไปนี้

- ส่วนแรกจะเป็นเว็บไซต์ที่ทำหน้าที่ในการรับข้อมูลและแสดงผลการแนะนำภาพยนตร์ให้กับผู้ที่เข้ามาเยี่ยมชม โดยในส่วนนี้จะพัฒนาโดยใช้ภาษา ASP.NET (Active Server Page .Net) ซึ่งเป็นภาษาที่ใช้ในการพัฒนาเว็บไซต์
- ส่วนที่สองจะเป็นส่วนสำหรับผู้ดูแลระบบการแนะนำสินค้า โดยในส่วนนี้จะพัฒนาโดยใช้ภาษา VB.NET (Visual Basic .Net) ในการพัฒนาะบบงานเนื่องมาจาก VB.NET เป็นเครื่องมือที่สามารถพัฒนาะบบงานบนระบบปฏิบัติการ Windows ได้ และอีกทั้งยังมีความเข้ากันได้กับภาษา ASP.NET และระบบฐานข้อมูลที่ใช้ในการพัฒนาะบบงานในครั้งนี้อีกด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- ส่วนระบบฐานข้อมูล ในการพัฒนาระบบงานชิ้นนี้ได้พัฒนาโดยใช้ระบบฐานข้อมูล Microsoft SQL Server 2000 เป็นฐานข้อมูล

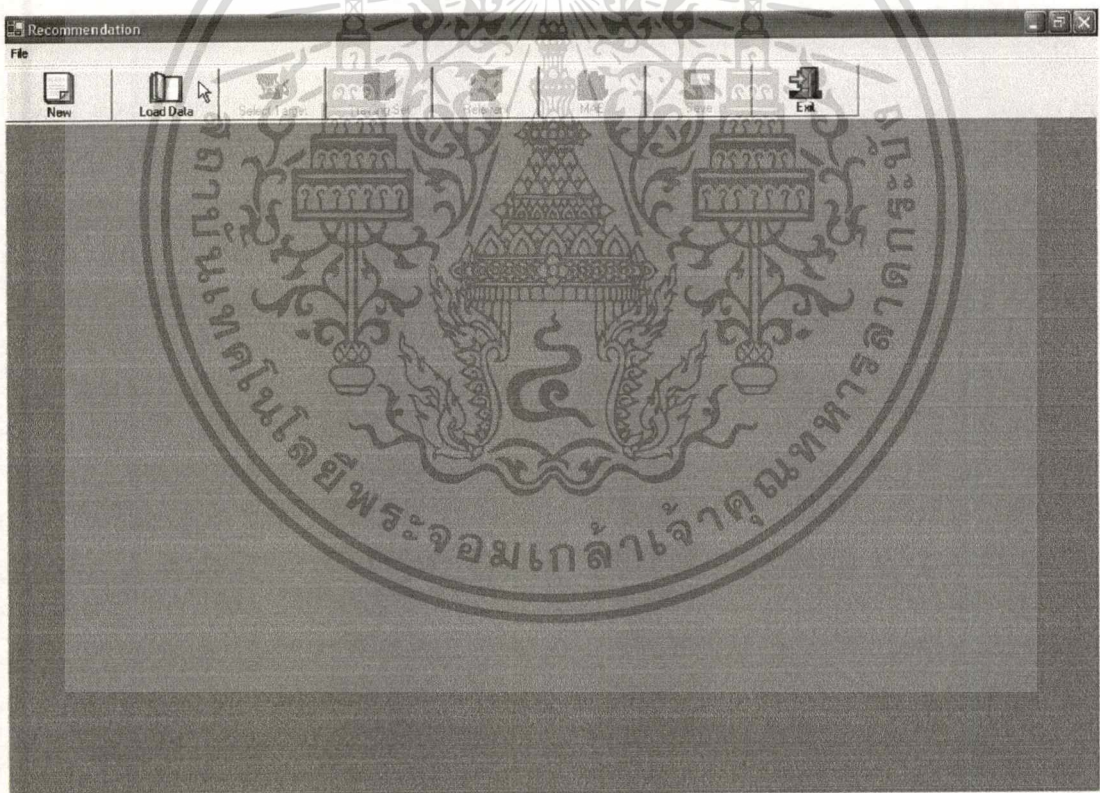
4.3 ขั้นตอนและรายละเอียดการใช้งาน

รายละเอียดการใช้งานจะแบ่งออกเป็น 2 ส่วนก็คือส่วนของผู้ดูแลระบบการแนะนำภาพยนตร์ ซึ่งทำหน้าที่สร้างการแนะนำสินค้าให้กับลูกค้า และในส่วนของผู้ใช้เว็บไซต์แนะนำภาพยนตร์ที่ใช้ติดต่อกับผู้เข้าชม

4.3.1 ขั้นตอนและรายละเอียดการใช้งานในส่วนของผู้ดูแลระบบการแนะนำภาพยนตร์

1. หน้าจอหลักของระบบการแนะนำภาพยนตร์

หน้าจอนี้เป็นหน้าจอแรกของระบบการแนะนำภาพยนตร์ ซึ่งผู้ใช้งานจะต้องกดปุ่ม Load Data เพื่อทำการนำข้อมูลเข้าสู่ระบบ

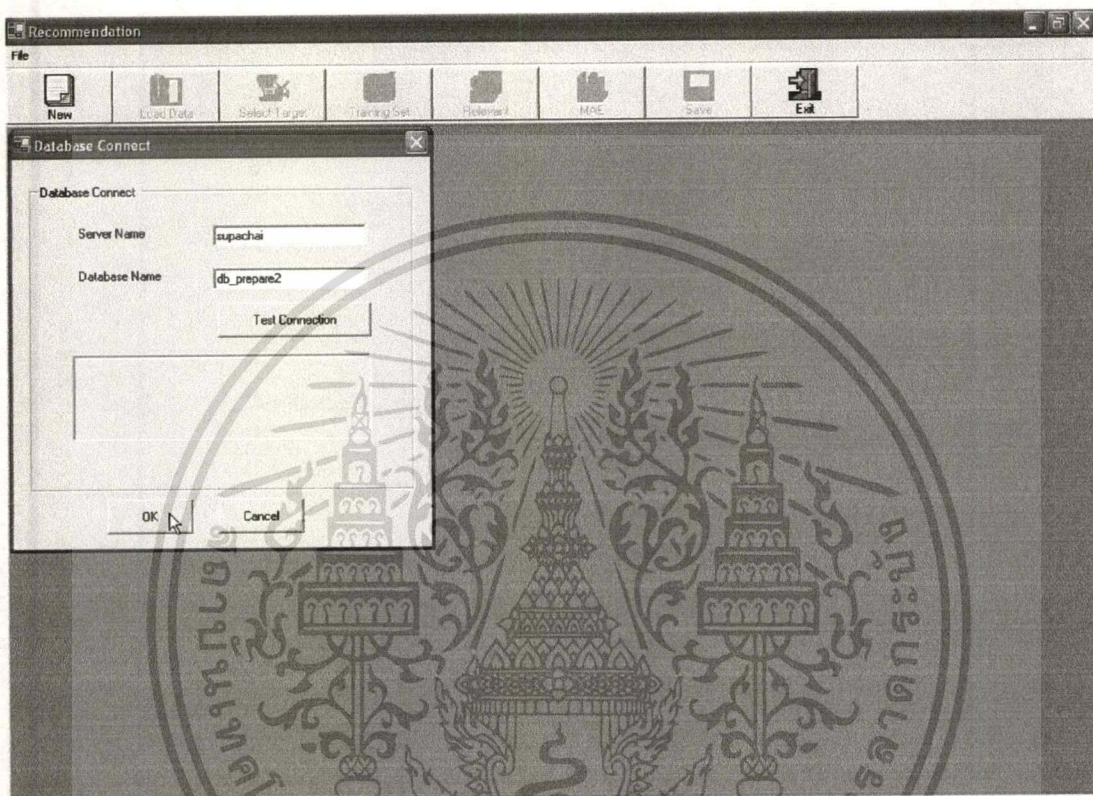


รูปที่ 4.1 แสดงหน้าจอหลักของระบบการแนะนำภาพยนตร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. หน้าจอของการติดตั้งฐานข้อมูลของระบบการแนะนำภาพยนตร์

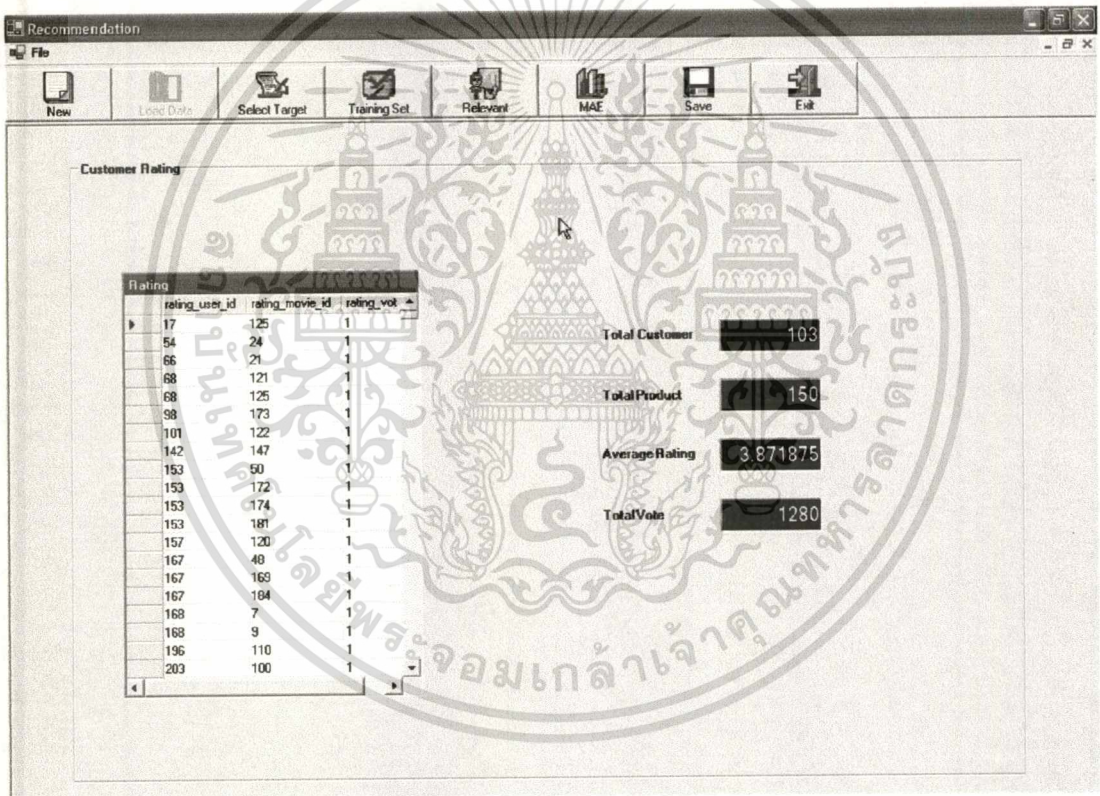
หน้าจอนี้จะทำหน้าที่ติดตั้งฐานข้อมูลซึ่งในระบบการแนะนำภาพยนตร์นี้ใช้ Microsoft SQL Server เป็นตัวจัดการฐานข้อมูล โดยที่ผู้ใช้จะต้องเลือก Server Name, Database Name เพื่อทำการติดตั้งกับฐานข้อมูล



รูปที่ 4.2 แสดงหน้าจอของการติดตั้งฐานข้อมูลของระบบการแนะนำภาพยนตร์

3. หน้าจอแสดงรายละเอียดข้อมูลการให้คะแนนความชอบของผู้เข้ามาเยี่ยมชม หน้าจอนี้จะแสดงข้อมูลรายละเอียดการให้คะแนนความชอบของผู้เข้ามาเยี่ยมชม ซึ่งผู้ใช้สามารถที่จะเลือกดูรายละเอียดการให้คะแนนของลูกค้าแต่ละคนได้ โดยที่ค่าคะแนนความชอบมีค่าได้ตั้งแต่ 1-5 และในหน้าจอนี้จะแสดงค่าต่างๆ ที่ได้จากข้อมูลรายละเอียดการให้คะแนนความชอบของผู้เยี่ยมชมซึ่งประกอบด้วย

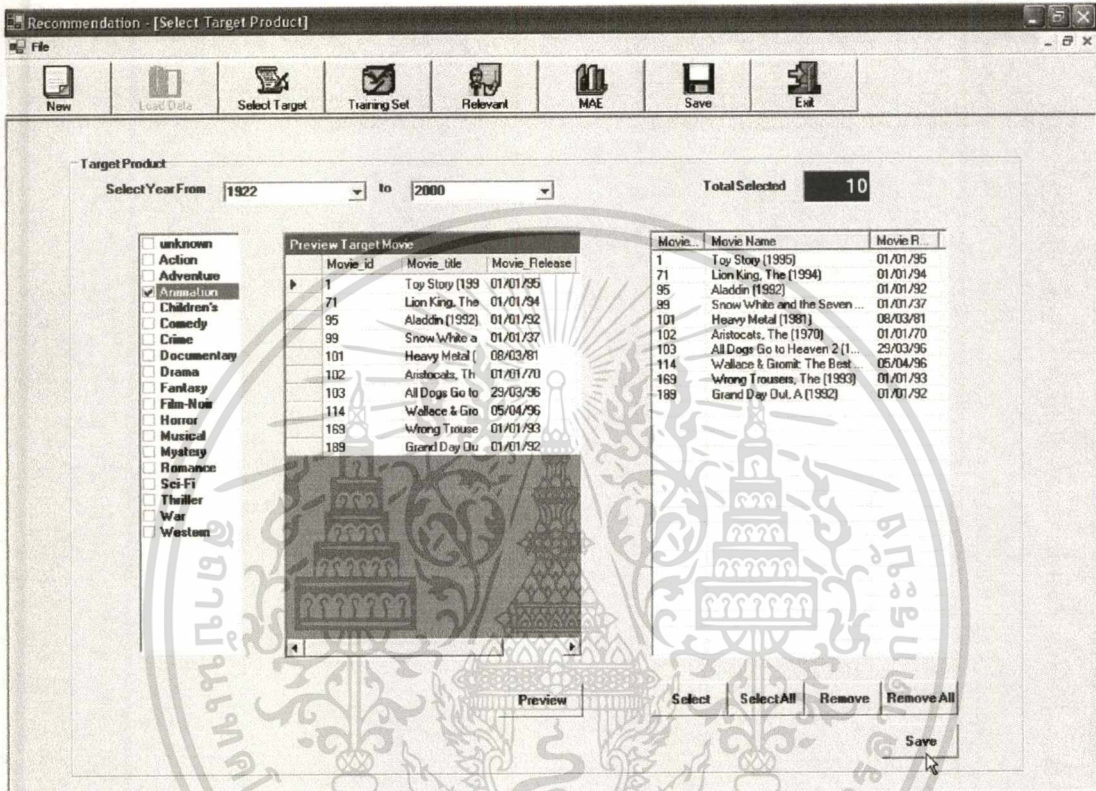
- Total Customer คือจำนวนของผู้เข้ามาเยี่ยมชมทั้งหมดที่ได้ให้คะแนนความชอบไว้
- Total Product คือจำนวนภาพยนตร์ที่ผู้เข้ามาเยี่ยมชมได้ให้คะแนนชอบเอาไว้
- Average Rating คือค่าเฉลี่ยของคะแนนความชอบของผู้เข้ามาเยี่ยมชม
- Total Vote คือจำนวนคะแนนชอบทั้งหมดที่ผู้เข้ามาเยี่ยมชมทั้งหมดให้คะแนนชอบเอาไว้



รูปที่ 4.3 หน้าจอแสดงรายละเอียดข้อมูลการให้คะแนนความชอบของผู้เยี่ยมชม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. หน้าจอแสดงรายละเอียดการเลือกกลุ่มของภาพยนตร์ที่จะนำมาแนะนำให้ผู้เยี่ยมชม เมื่อผู้ใช้กดปุ่ม Select Target ก็จะแสดงหน้าการเลือกกลุ่มของภาพยนตร์เพื่อที่จะไปใช้ในการแนะนำให้กับผู้เยี่ยมชม โดยสามารถเลือกกลุ่มของภาพยนตร์ได้ตามปีที่ออกฉายและตามประเภทภาพยนตร์



รูปที่ 4.4 หน้าจอแสดงรายละเอียดการเลือกกลุ่มของภาพยนตร์ที่จะนำมาแนะนำให้ผู้เยี่ยมชม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. หน้าจอแสดงรายละเอียดการ Training Data

เมื่อผู้ใช้คลิกปุ่ม Training Data ก็จะแสดงหน้าจอรายละเอียดการ Training Data การทำ Training Data จะนำกลุ่มของภาพยนตร์ที่ได้ทำการเลือกเอาไว้เพื่อที่จะทำการแนะนำให้กับผู้เข้าชม โดยคิดว่าภาพยนตร์แต่ละเรื่องมีลูกค้ายคนไหนเคยดูภาพยนตร์เรื่องนี้บ้างแล้วนำมาหาค่าความเกี่ยวข้องกับกลุ่มของภาพยนตร์ที่ถูกค้ายคนนั้นเคยดู เพื่อที่จะไปคำนวณหาค่า relevant ระหว่างภาพยนตร์ที่จะทำการแนะนำให้กับผู้เข้าชมกับลูกค้ายแต่ละคนที่เคยดูภาพยนตร์เรื่องนี้

The screenshot shows a software window titled "Recommendation" with a menu bar containing: New, Load Data, Select Target, Training Set, Relevant, MAE, Save, and Exit. The main content area is titled "Training Data" and is split into two panes:

- Target Movie:** A table listing movies with their IDs, titles, and release dates.

Movie_id	Movie_title	Movie_released
1	Top Story (19	01/01/95
71	Lion King, Th	01/01/94
95	Aladdin (1992	01/01/92
98	Snow White a	01/01/37
101	Heavy Metal (08/03/81
102	Aristocats, Th	01/01/70
103	All Dogs Got	23/03/96
114	Wallace & Gr	05/04/96
169	Wrong Trous	01/01/93
189	Grand Day O	01/01/82
- Information:** A list of target movie IDs and their corresponding relevant movie IDs.


```

      ***** start in training data *****
      Target TWith 10->1.36357975006104
      Target TWith 13->1.86417865753174
      Target TWith 14->1.3295922672005
      Target TWith 19->1.66959119577630
      Target TWith 25->0.7980592716217
      Target TWith 50->0.491546154022217
      Target TWith 100->0.295873165130615
      Target TWith 111->0.579072594642639
      Target TWith 127->0.321203708649682
      Target TWith 7->0.124731593663221
      Target TWith 9->0.116262435313006
      Target TWith 13->1.86417865753174
      Target TWith 100->0.295873165130615
      Target TWith 111->0.579072594642639
      Target TWith 117->0.0891544818878174
      Target TWith 125->0.340008735656738
      Target TWith 126->1.39068675041199
      Target TWith 137->0.63382689526583
      
```

At the bottom of the interface, there is a "Please Wait" dialog box, a progress bar, and a "Total Target Movie" indicator showing 10.

รูปที่ 4.5 หน้าจอแสดงรายละเอียดการ Training Data หาค่า relevant ระหว่างภาพยนตร์กับลูกค้าย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. หน้าจอแสดงรายละเอียดค่า relevant ระหว่างภาพยนตร์กับลูกค้าแต่ละคน

เมื่อผู้ใช้คลิกปุ่ม Relevant ก็จะแสดงหน้าจอรายละเอียดของค่า relevant ระหว่างภาพยนตร์แต่ละเรื่องกับลูกค้าแต่ละคนซึ่งค่า relevant ที่คำนวณได้มาถ้าค่ายิ่งมากจะหมายถึงภาพยนตร์กับลูกค้ามีค่าที่เป็น Relevant สูง ซึ่งผู้ใช้กำหนดค่า K เพื่อกำหนดจำนวนลูกค้าที่มีค่า Relevant มากสุดจำนวน K คนในภาพยนตร์แต่ละเรื่องเพื่อที่จะนำไปใช้ในการสร้างการแนะนำภาพยนตร์ให้กับผู้เข้ามาเยี่ยมชมต่อไป

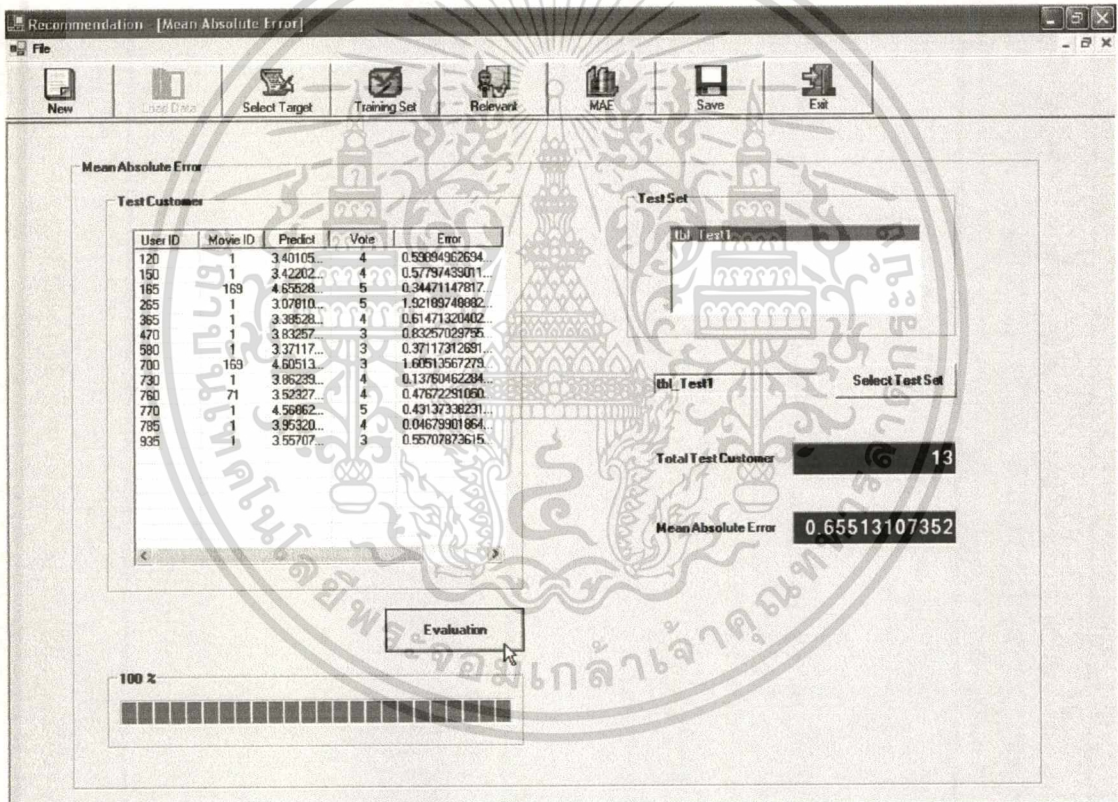
Target_Movie	User_i	relevant
1	767	4.502923429
1	512	4.377309440
1	419	4.267545653
1	777	4.160394865
1	681	4.044145525
1	388	3.782069426
1	471	3.509449526
1	723	3.373810902
1	902	3.347666631
1	613	2.997622098
1	79	2.923773692
1	794	2.820249346
1	182	2.589460325
1	677	2.255706568
1	108	2.244707494
1	2	2.234439406
1	247	2.205060067
1	923	2.155682404
1	839	2.135821626
1	576	2.031179366

Target_Movie	User_i	relevant
1	767	4.502923429
1	512	4.377309440
1	419	4.267545653
1	777	4.160394865
1	681	4.044145525
1	388	3.782069426
1	471	3.509449526
1	723	3.373810902
1	902	3.347666631
1	613	2.997622098
1	79	2.923773692
1	794	2.820249346
1	182	2.589460325
1	677	2.255706568
1	108	2.244707494
1	2	2.234439406
1	247	2.205060067
1	71	6.553645524
95	779	5.934457505
89	167	5.059577551

รูปที่ 4.6 หน้าจอแสดงรายละเอียดค่า relevant ระหว่างภาพยนตร์กับลูกค้าแต่ละคน

7. หน้าจอแสดงรายละเอียดการวัดค่า Error เพื่อวัดประสิทธิภาพของระบบแนะนำสินค้า โดยแบ่งข้อมูลกลุ่มลูกค้าส่วนหนึ่งเพื่อให้ระบบทดลองแนะนำภาพยนตร์ให้ลูกค้ากลุ่มนี้ ซึ่งกลุ่มลูกค้าแต่ละคนที่นำมาทดลองนี้จะมีกลุ่มภาพยนตร์จำนวนหนึ่งที่ลูกค้าเคยดู ระบบก็จะสุ่มภาพยนตร์หนึ่งเรื่องนำมาทำนายความชอบว่าลูกค้าคนนี้จะชอบภาพยนตร์เรื่องนี้มากน้อยแค่ไหน ส่วนภาพยนตร์ที่เหลือที่ไม่ได้นำมาทำนายก็นำไปใช้ในการทำนายความชอบ ซึ่งค่าที่ทำนายได้นำไปเทียบกับค่าจริงเพื่อหาค่า Error ออกมาเมื่อได้ค่า Error ของลูกค้าแต่ละคนในกลุ่มลูกค้าทดลองมาแล้วนำทั้งหมดมาหาค่าเฉลี่ย Error ได้ค่าๆ หนึ่งออกมา

- Total Test Customer คือจำนวนกลุ่มลูกค้าที่นำมาทดลอง
- Mean Absolute Error คือค่าเฉลี่ย Error ที่วัดได้ออกมา

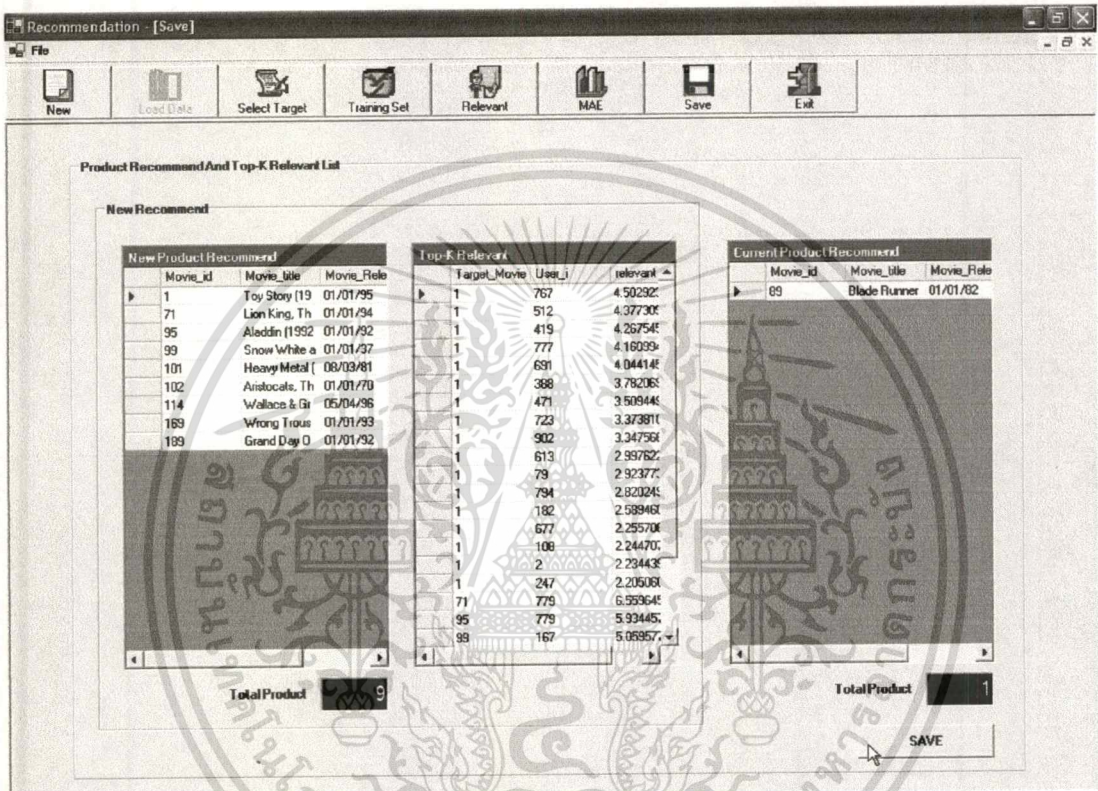


รูปที่ 4.7 หน้าจอแสดงรายละเอียดการวัดประสิทธิภาพของระบบแนะนำภาพยนตร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8. หน้าจอแสดงการบันทึกการสร้างการแนะนำภาพยนตร์ที่จะแนะนำให้กับผู้ที่เข้ามาเยี่ยมชม

เมื่อผู้ใช้กดปุ่ม Save ก็จะปรากฏหน้าจอที่แสดงข้อมูลที่ได้กำหนดเอาไว้ซึ่งก็คือ รายการภาพยนตร์ที่ใช้ในการแนะนำให้กับผู้เยี่ยมชม และ ค่าของ K-relevant คือจำนวนลูกค้ำที่มีความเกี่ยวข้องกับกลุ่มของภาพยนตร์แต่ละเรื่องที่ใช้ในการแนะนำให้กับผู้เยี่ยมชม



รูปที่ 4.8 หน้าจอแสดงการบันทึกการสร้างการแนะนำภาพยนตร์ที่จะแนะนำให้ผู้เยี่ยมชม

4.3.2 ขั้นตอนและรายละเอียดการใช้งานในส่วนของเว็บไซต์

1. หน้าจอแรกแสดงการ Login เข้าสู่ระบบการแนะนำภาพยนตร์

เมื่อผู้เข้าเยี่ยมชมเข้ามาที่จะรับการแนะนำภาพยนตร์อีกครั้งก็จะมีหน้าจอ Login สำหรับสมาชิกเพื่อเข้าสู่ระบบการแนะนำภาพยนตร์ ถ้าหากผู้เข้ามาเยี่ยมชมยังไม่ได้ลงทะเบียนก็จะต้องไปยังหน้าจอลงทะเบียนก่อน โดยที่หน้าจอนี้จะทำหน้าที่รับ Username และ Password ถ้าหากเป็นผู้ใช้ที่มีสิทธิ์เข้าใช้งานก็จะสามารถไปยังส่วนของเว็บไซต์ต่อไปได้



รูปที่ 4.9 หน้าจอแสดงการ Login เข้าสู่ระบบการแนะนำภาพยนตร์

2. หน้าจอการลงทะเบียนสำหรับสมาชิกใหม่

หน้าจอแบบฟอร์มรับการลงทะเบียนสำหรับสมาชิกใหม่ โดยที่หน้าจอนี้จะทำหน้าที่รับข้อมูลจากผู้เข้ามาเยี่ยมชมซึ่งจะประกอบไปด้วย

- First Name ชื่อ
- Last Name นามสกุล
- Age อายุ
- Gender เพศ
- Occupation อาชีพ
- E-mail อีเมล
- Zip code รหัสไปรษณีย์
- Username ชื่อบัญชี
- Password รหัสผ่าน

★★★★ = Must See
 ★★★ = Will Enjoy
 ★★ = It's OK
 ★ = Fairly Bad
 ☆ = Awful

Recommendation Register Genres Year of Release About

Welcome

MEMBER LOG-IN

WELCOME TO MOVIE RECOMMENDATION
 Please register for receive movies recommendation that helping you find the right movies

Please to Register

First Name :** Supachei
 Last Name :** Asewachinthepkul
 Age :** 25-34
 Gender :** Male Female
 Occupation :** programmer
 E-mail : supacheia@gmail.com
 Zipcode : 10270

Username :** u1000
 Password :**
 RePassword :**

Submit

Copyright © 2006-2007 AMIT., All right reserved.

รูปที่ 4.10 แสดงหน้าจอลงทะเบียนสำหรับสมาชิกใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. หน้าจอการลงคะแนนความชอบ (Rating) ของผู้เข้ามาเยี่ยมชม

เมื่อผู้เข้ามาเยี่ยมชมได้ลงทะเบียนเรียบร้อยแล้วและทำการ Login แล้วก็จะเข้าสู่หน้าจอการลงคะแนนความชอบต่อภาพยนตร์ที่ผู้เข้ามาเยี่ยมชมเคยชมหรือชอบ เพื่อนำคะแนนตัวอย่างที่ได้รับนี้ไปเป็นพื้นฐานในการแนะนำภาพยนตร์กับผู้เข้าเยี่ยมชมต่อไป โดยที่ผู้เข้ามาเยี่ยมชมสามารถที่จะทำการค้นหาภาพยนตร์ที่ต้องการได้ ซึ่งการค้นหาสามารถค้นหาได้จาก ชื่อภาพยนตร์และประเภทภาพยนตร์

Movie Recommendation

★★★★ = Must See
 ★★★★★ = Will Enjoy
 ★★★★★ = It's OK
 ★★★★★ = Fairly Bad
 ★★★★★ = Awful

Recommendation Register Genres Year of Release About

Welcome Recommend Add Vote Edit Vote Edit Profile

Welcome u1000
 History Voted Is 0
 Log Out

Your Current Voted is 10
 Prev Next

Title	Voted	Star
Movie Title : Toy Story (1995) Add Type of Movie : Animation Children's Comedy	5	★★★★★
Movie Title : GoldenEye (1995) Add Type of Movie : Action Adventure Thriller	4	★★★★
Movie Title : Four Rooms (1995) Add Type of Movie : Thriller	3	★★★
Movie Title : Get Shorty (1995) Add Type of Movie : Action Comedy Drama	3	★★★
Movie Title : Copycat (1995) Add Type of Movie : Crime Drama Thriller	2	★★
Movie Title : Shanghai Triad (Yao a yao dao dai waipo qiao) (1995) Add Type of Movie : Drama	0	???
Movie Title : Twelve Monkeys (1995) Add Type of Movie : Drama Sci-Fi	3	★★★

Top 20

- 1 Star Wars (1977)
- 2 Fargo (1996)
- 3 Toy Story (1995)
- 4 Return of the Jedi (1983)
- 5 Twelve Monkeys (1995)

more...

20 New Release

- 1 Return of the Jedi (1983)
- 2 Fanny (1996)

รูปที่ 4.11 หน้าจอการลงคะแนนความชอบ (Rating) ของผู้เข้ามาเยี่ยมชม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. หน้าจอแสดงการแนะนำภาพยนตร์ของผู้เข้ามาเยี่ยมชม

หน้าจอนี้เป็นหน้าจอหลักของการทำงานของระบบการแนะนำภาพยนตร์ โดยที่หน้าจอนี้จะแสดงการแนะนำภาพยนตร์ให้กับผู้เข้ามาเยี่ยมชม โดยจะอาศัยข้อมูลพื้นฐานที่ผู้เข้ามาเยี่ยมชมได้ให้ไว้กับระบบ ซึ่งข้อมูลที่จะนำมาใช้ก็คือข้อมูลการให้คะแนนความชอบของลูกค้า (rating) โดยที่ระบบจะนำเอาข้อมูลเหล่านี้ไปคำนวณหา Similarity จากฐานข้อมูลของระบบแนะนำสินค้าและระบบจะทำการสร้างการแนะนำภาพยนตร์ให้กับผู้เข้ามาเยี่ยมชมคนนั้น โดยใช้เทคนิค Collaborative filtering มาทำการแนะนำภาพยนตร์ให้กับผู้เข้ามาเยี่ยมชมได้ โดยที่หน้าจอนี้จะแสดงการให้คะแนนของระบบซึ่งจะแสดงด้วยจำนวนรูปภาพดาวซึ่งถ้ามีจำนวนมากก็จะแสดงว่าระบบคิดว่าน่าจะชอบภาพยนตร์เรื่องนี้มาก

★★★★★ = Must See
 ★★★★★ = Will Enjoy
 ★★★★★ = It's OK
 ★★★★★ = Fairly Bad
 ★★★★★ = Awful

Recommendation Register Games Year of Release About

Welcome u1000
 History Voted is 10
 Log Out

Search

All Movies GO

More to Explore

Top 20

- 1 Star Wars (1977)
- 2 Fargo (1996)
- 3 Toy Story (1995)
- 4 Return of the Jedi (1983)
- 5 Twelve Monkeys (1995)

more...

20 New Release

- 1 Return of the Jedi (1983)
- 2 Faran (1996)

Previous Rating for You	Movie Name	Type of Movie
★★★★★	Heavy Metal (1981)	Action Adventure Animation Horror Sci-Fi
★★★★★	Aristocats, The (1970)	Animation Children's
★★★★★	Aladdin (1992)	Animation Children's Comedy Musical
★★★★★	Grand Day Out, A (1992)	Animation Comedy
★★★★	Lion King, The (1994)	Animation Children's Musical
★★★★	Wallece S. Gromit: The Best of Aardman Animation (1999)	Animation
★★★	Wrong Trousers, The (1993)	Animation Comedy

Prev-Next

รูปที่ 4.12 หน้าจอแสดงการแนะนำภาพยนตร์ของผู้เข้ามาเยี่ยมชม

5. หน้าจอแสดงการเพิ่มการให้คะแนนความชอบของผู้เข้ามาเยี่ยมชม

หน้าจอนี้จะทำหน้าที่รับการให้คะแนนความชอบเพิ่มจากผู้เข้ามาเยี่ยมชม โดยที่ผู้เข้ามาเยี่ยมชมสามารถที่จะให้คะแนนความชอบเพิ่มต่อภาพยนตร์ที่มีอยู่ในระบบแนะนำภาพยนตร์ได้ซึ่งโดยปกติระบบแนะนำภาพยนตร์จะกำหนดการให้คะแนนความชอบเอาไว้แล้วแต่ผู้เข้ามาเยี่ยมชมสามารถที่จะเพิ่มการให้คะแนนความชอบกับระบบได้ เพื่อที่ระบบจะสามารถสร้างการแนะนำภาพยนตร์ที่มีประสิทธิภาพมากขึ้นให้กับผู้เข้ามาเยี่ยมชมคนนั้น ได้จากนั้นระบบแนะนำภาพยนตร์ก็จะคำนวณหา Similarity ใหม่และจะทำการสร้างการแนะนำภาพยนตร์ให้กับผู้เข้ามาเยี่ยมชมคนนั้นใหม่ต่อไป

Movie Recommendation

★★★★ = Must See
 ★★★★★ = Will Enjoy
 ★★★☆☆ = It's OK
 ★★☆☆☆ = Fairly Bad
 ★☆☆☆☆ = Awful

Recommendation Register Genre Year of Release About

Welcome u1000
 History Voted Is 10
 Log Out

Recommend Add Vote Edit Vote Edit Profile

Your Current Voted Is 4
 Prev Next

Title	Voted	Star
Movie Title :Shanghai Triad (Yao a yao dao waipo qiao) (1997) Add Type of Movie :Drama	3	★★★
Movie Title :Babe (1995) Add Type of Movie :Children's Comedy Drama	4	★★★★★
Movie Title :Mighty Aphrodite (1995) Add Type of Movie :Comedy	0	???
Movie Title :Postino, Il (1994) Add Type of Movie :Drama Romance	0	???
Movie Title :Mr. Holland's Opus (1995) Add Type of Movie :Drama	2	★★
Movie Title :French Twist (Gazon maudit) (1995) Add Type of Movie :Comedy Romance	4	★★★★★
Movie Title :From Dusk Till Dawn (1996) Add Type of Movie :Action Comedy Crime Horror Thriller	0	???

Search
 All Movies GO

More to Explore
 Top 20
 1 Star Wars (1977)
 2 Fargo (1996)
 3 Toy Story (1995)
 4 Return of the Jedi (1983)
 5 Twelve Monkeys (1995)
 more...

20 New Release
 1 Return of the Jedi (1983)
 2 Farin (1996)

รูปที่ 4.13 หน้าจอแสดงการเพิ่มการให้คะแนนความชอบของผู้เข้ามาเยี่ยมชม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6. หน้าจอแสดงการแก้ไขการให้คะแนนความชอบของผู้เข้ามาเยี่ยมชม

ผู้เข้ามาเยี่ยมชมสามารถที่จะแก้ไขการให้คะแนนความชอบของตนเองได้ โดยที่หน้าจอนี้จะทำหน้าที่รับการแก้ไขการให้คะแนนความชอบในภาพยนตร์ที่ผู้เข้ามาเยี่ยมชมได้ให้ไว้กับระบบในตอนแรก เมื่อผู้เข้ามาเยี่ยมชมแก้ไขการให้คะแนนความชอบในภาพยนตร์แล้วระบบแนะนำภาพยนตร์ก็จะคำนวณหา Similarity ใหม่และจะทำการสร้างการแนะนำภาพยนตร์ให้กับผู้เข้ามาเยี่ยมชมคนนั้นใหม่ต่อไป



The screenshot shows a web interface for a movie recommendation system. At the top, there's a navigation bar with links for Recommendation, Register, Genres, Year of Release, and About. Below this is a user profile section for 'u1000' with options to Recommend, Add Vote, Edit Vote, and Edit Profile. The main content area displays a list of movies with their titles, edit types, voted counts, and star ratings. A large watermark of a university seal is overlaid on the page.

Movie Title	Edit Type of Movie	Voted	Star
Toy Story (1995)	Animation Children's Comedy	5	★★★★★
GoldenEye (1995)	Action Adventure Thriller	4	★★★★
Four Rooms (1995)	Thriller	3	★★★
Get Shorty (1995)	Action Comedy Drama	3	★★★
Copycat (1995)	Crime Drama Thriller	2	★★
Twelve Monkeys (1995)	Drama Sci-Fi	3	★★★
Dead Man Walking (1995)	Drama	2	★★

รูปที่ 4.14 หน้าจอแสดงการแก้ไขการให้คะแนนความชอบของผู้เข้ามาเยี่ยมชม

7. หน้าจอแสดงการแก้ไขข้อมูลของผู้เข้ามาเยี่ยมชม

ผู้เข้ามาเยี่ยมชมสามารถที่จะทำการแก้ไขข้อมูลส่วนตัวของตนเองได้ โดยทำการแก้ไขผ่านทางหน้าจอนี้ โดยที่ผู้เข้ามาเยี่ยมชมสามารถแก้ไขข้อมูลได้ทั้งหมดยกเว้น Username ซึ่งระบบจะไม่ยอมให้สมาชิกสามารถแก้ไขในส่วนนี้ได้

Recommendation Register Genres Year of Release About

Welcome u1000
History Voted is 10
Log Out

Search

All Movies GO

More To Explore

Top 20

- 1 Star Wars (1977)
- 2 Fargo (1996)
- 3 Toy Story (1995)
- 4 Return of the Jedi (1983)
- 5 Twelve Monkeys (1995)

more...

20 New Release

- 1 Return of the Jedi (1983)
- 2 Fargo (1996)
- 3 Jude (1996)
- 4 Unhook the Stars (1996)
- 5 Swingers (1996)

more

Edit Profile

First Name :** Supachai

Last Name :** Asawachinthepkul

Age :** 25-34

Gender :** Male Female

Occupation :** programmer

E-mail : supacheis@gmail.com

Zipcode : 10270

Update

Change Password

Username :** u1000

Old Password :**

Password :**

RePassword :**

Change Password

รูปที่ 4.15 หน้าจอแสดงการแก้ไขข้อมูลของผู้เข้ามาเยี่ยมชม

8. หน้าจอแสดงข้อมูลภาพยนตร์ที่เพิ่งออกฉายในขณะนี้

หน้าจอนี้แสดงข้อมูลของภาพยนตร์ที่เพิ่งเข้ามาฉายในขณะนี้ซึ่งข้อมูลที่แสดงก็คือ ชื่อภาพยนตร์ ประเภทของภาพยนตร์ และปีที่ภาพยนตร์เรื่องนั้นเข้ามาฉาย

Movie Recommendation

★★★★ = Must See
 ★★★ = Will Enjoy
 ★★ = It's OK
 ★ = Fairly Bad
 ☆ = Awful

Recommendation Register Genres Year of Release About

Welcome Recommend Add Vote Edit Vote Edit Profile

Welcome u1000
 History Voted Is 10
 Log Out

20 Movie New Release :

No.	Movie Name	Type of Movie	Release Year
1	Return of the Jedi (1983)	Action Adventure Romance Sci-Fi War	1997
2	Fargo (1996)	Crime Drama Thriller	1997
3	Jude (1996)	Drama	1996
4	Unhook the Stars (1996)	Drama	1996
5	Swingers (1996)	Comedy Drama	1996
6	Ghost and the Darkness, The (1996)	Action Adventure	1996
7	Long Kiss Goodnight, The (1996)	Action Thriller	1996
8	Bound (1996)	Crime Drama Romance Thriller	1996
9	D3: The Mighty Ducks (1996)	Children's Comedy	1996
10	Big Night (1996)	Drama	1996
11	Spitfire Grill, The (1996)	Drama	1996
12	Kansas City (1996)	Crime	1996
13	Supercop (1992)	Action Thriller	1996
14	Frighteners, The (1996)	Comedy Horror	1996
15	Independence Day (ID4) (1996)	Action Sci-Fi War	1996
16	Phenomenon (1996)	Drama Romance	1996
17	Strip tease (1996)	Comedy Crime	1996
18	Lone Star (1996)	Drama Mystery	1996
19	Moll Flanders (1996)	Drama	1996
20	Cable Guy, The (1996)	Comedy	1996

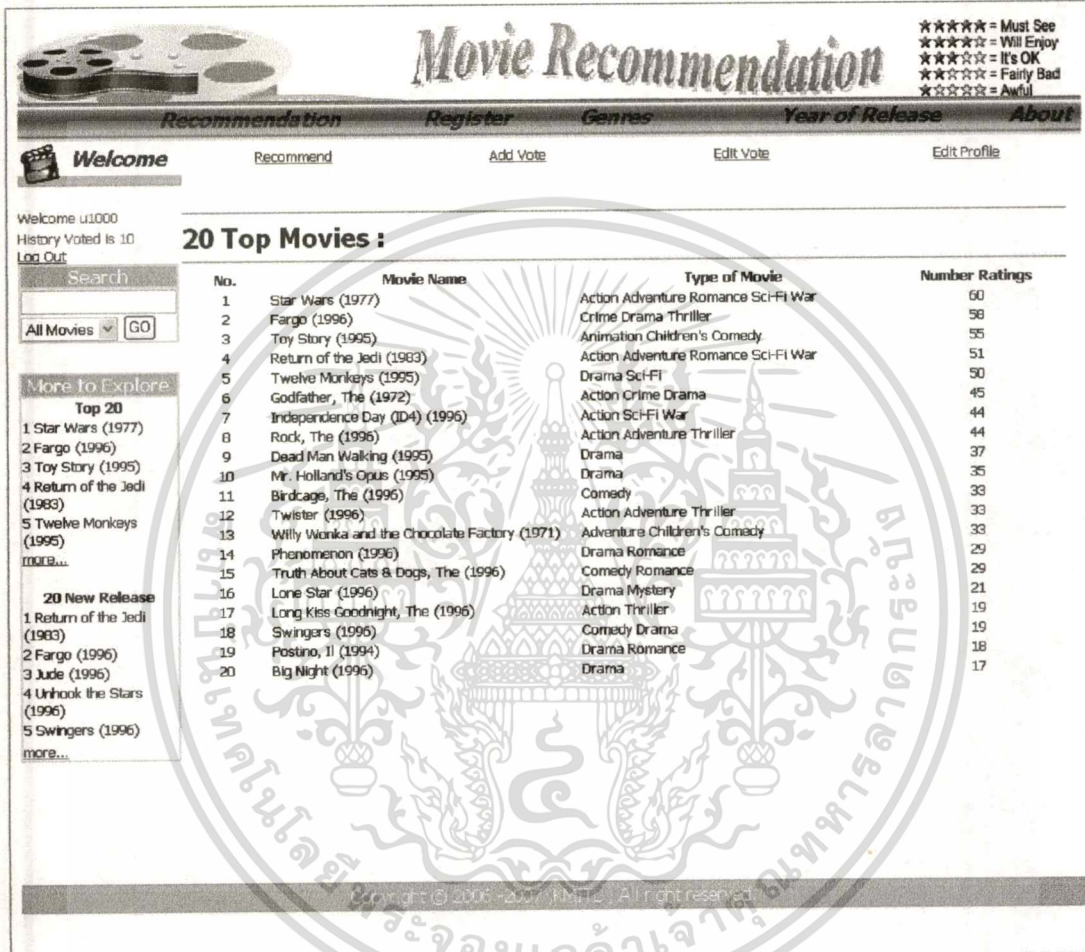
Copyright © 1997 MIT. All rights reserved.

รูปที่ 4.16 หน้าจอแสดงข้อมูลภาพยนตร์ที่เพิ่งออกฉายในขณะนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9. หน้าจอแสดงข้อมูลภาพยนตร์ที่เป็นที่นิยมในขณะนี้

หน้าจอนี้แสดงข้อมูลของภาพยนตร์ที่เป็นที่นิยม 20 เรื่องแรก ซึ่งจะแนะนำความนิยมนี้ได้มาจากคนที่ผู้เข้ามาเยี่ยมชมได้ทำการให้คะแนนเอาไว้ ซึ่งข้อมูลที่แสดงก็คือ ชื่อภาพยนตร์ ประเภทภาพยนตร์ จำนวนคนที่ให้คะแนนความชอบต่อภาพยนตร์แต่ละเรื่อง



Movie Recommendation

★★★★ = Must See
★★★★ = Will Enjoy
★★★☆☆ = It's OK
★★☆☆☆ = Fairly Bad
★☆☆☆☆ = Awful

Recommendation Register Genres Year of Release About

Welcome Recommend Add Vote Edit Vote Edit Profile

Welcome u1000
History Voted is 10
Log Out

Search

All Movies GO

More to Explore

Top 20

1 Star Wars (1977)
2 Fargo (1996)
3 Toy Story (1995)
4 Return of the Jedi (1983)
5 Twelve Monkeys (1995)
more...

20 New Release

1 Return of the Jedi (1983)
2 Fargo (1996)
3 Jude (1996)
4 Unhook the Stars (1996)
5 Swingers (1996)
more...

No.	Movie Name	Type of Movie	Number Ratings
1	Star Wars (1977)	Action Adventure Romance Sci-Fi War	60
2	Fargo (1996)	Crime Drama Thriller	58
3	Toy Story (1995)	Animation Children's Comedy	55
4	Return of the Jedi (1983)	Action Adventure Romance Sci-Fi War	51
5	Twelve Monkeys (1995)	Drama Sci-Fi	50
6	Godfather, The (1972)	Action Crime Drama	45
7	Independence Day (ID4) (1996)	Action Sci-Fi War	44
8	Rock, The (1996)	Action Adventure Thriller	44
9	Dead Man Walking (1995)	Drama	37
10	Mr. Holland's Opus (1995)	Drama	35
11	Birdcage, The (1996)	Comedy	33
12	Twister (1996)	Action Adventure Thriller	33
13	Willy Wonka and the Chocolate Factory (1971)	Adventure Children's Comedy	33
14	Phenomenon (1996)	Drama Romance	29
15	Truth About Cats & Dogs, The (1996)	Comedy Romance	29
16	Lone Star (1996)	Drama Mystery	21
17	Long Kiss Goodnight, The (1996)	Action Thriller	19
18	Swingers (1996)	Comedy Drama	19
19	Postino, Il (1994)	Drama Romance	18
20	Big Night (1996)	Drama	17

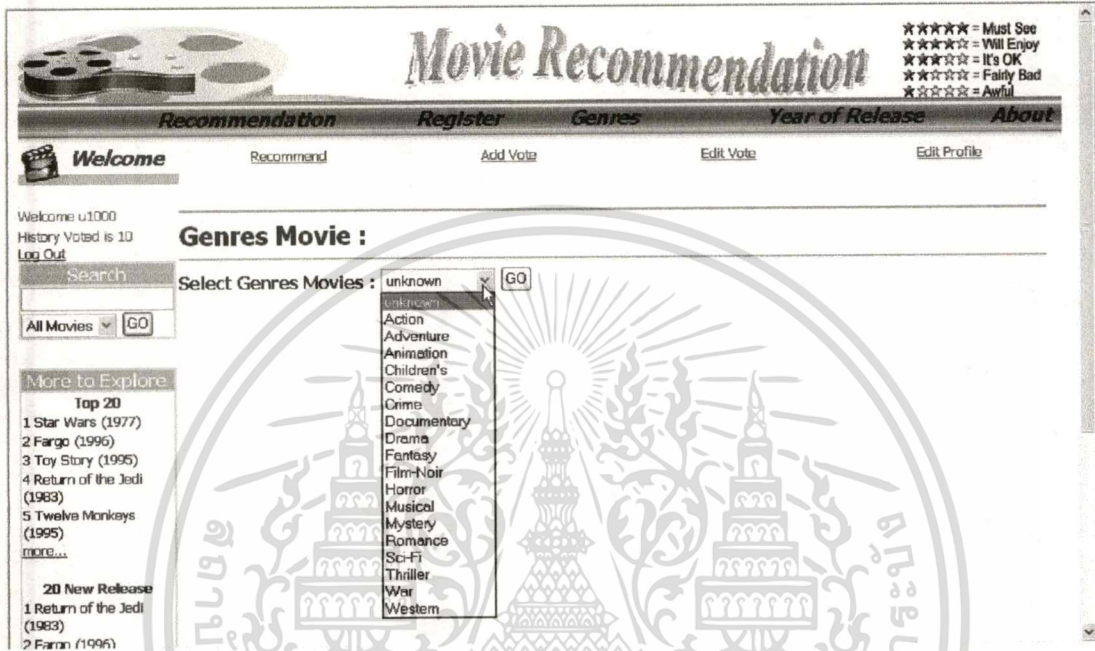
Copyright © 2006-2007 JIMPLY. All right reserved.

รูปที่ 4.17 หน้าจอแสดงข้อมูลภาพยนตร์ที่เป็นที่นิยมในขณะนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

10. หน้าจอแสดงการเลือกข้อมูลของภาพยนตร์ตามประเภทของภาพยนตร์

หน้าจอนี้แสดงการเลือกข้อมูลของภาพยนตร์ตามประเภทของภาพยนตร์ ซึ่งประเภทของภาพยนตร์นั้นก็จะเป็นประเภทที่ได้แสดงไว้ในข้างต้น โดยที่ข้อมูลที่แสดงก็คือ ชื่อภาพยนตร์ ประเภทของภาพยนตร์ และปีที่ภาพยนตร์เข้าฉาย



รูปที่ 4.18 หน้าจอแสดงการเลือกข้อมูลของภาพยนตร์ตามประเภทของภาพยนตร์

11. หน้าจอแสดงข้อมูลของภาพยนตร์ตามประเภทของภาพยนตร์

หน้านี้แสดงข้อมูลของภาพยนตร์ตามประเภทของภาพยนตร์ที่ได้เลือกเอาไว้ โดยที่ข้อมูลที่แสดงก็คือ ชื่อภาพยนตร์ ประเภทของภาพยนตร์ และปีที่ภาพยนตร์เรื่องนั้นเข้าฉาย

Welcome u1000
History Voted Is 10
Log Out

Search

All Movies GO

More to Explore

Top 20

- 1 Star Wars (1977)
- 2 Fargo (1996)
- 3 Toy Story (1995)
- 4 Return of the Jedi (1983)
- 5 Twelve Monkeys (1995)

more...

20 New Release

- 1 Return of the Jedi (1983)
- 2 Fargo (1996)
- 3 Jude (1996)
- 4 Unhook the Stars (1996)
- 5 Swingers (1996)

more...

Genres Movie :

Select Genres Movies : Action GO

Prev Next

Title	Type of Movie	Release Year
Apollo 13 (1995)	Action Drama Thriller	1995
Batman Forever (1995)	Action Adventure Comedy Crime	1995
Desperado (1995)	Action Romance Thriller	1995
Strange Days (1995)	Action Crime Sci-Fi	1995
Star Wars (1977)	Action Adventure Romance Sci-Fi War	1977
Natural Born Killers (1994)	Action Thriller	1994
Outbreak (1995)	Action Drama Thriller	1995
Stargate (1994)	Action Adventure Sci-Fi	1994
Crow, The (1994)	Action Romance Thriller	1994
Maverick (1994)	Action Comedy Western	1994
Faster Pussycat! Kill! Kill! (1965)	Action Comedy Drama	1965
Operation Dumbo Drop (1995)	Action Adventure Comedy War	1995
Rock, The (1996)	Action Adventure Thriller	1996
Twister (1996)	Action Adventure Thriller	1996
Independence Day (ID4) (1996)	Action Sci-Fi War	1996

Prev Next

รูปที่ 4.19 หน้าจอแสดงเลือกข้อมูลของภาพยนตร์ตามประเภทของภาพยนตร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

12. หน้าจอแสดงการเลือกข้อมูลของภาพยนตร์ตามปีที่ภาพยนตร์นั้นออกฉาย
 หน้าจอนี้แสดงการเลือกข้อมูลของภาพยนตร์ตามปีที่ภาพยนตร์นั้นออกฉาย ซึ่งปีของ
 ภาพยนตร์นั้นจะตั้งแต่ปี ค.ศ. 1922-2000 โดยที่ข้อมูลที่แสดงก็คือ ชื่อภาพยนตร์ ประเภทของ
 ภาพยนตร์ และปีที่ภาพยนตร์เรื่องนั้นเข้าฉาย



รูปที่ 4.20 หน้าจอแสดงการเลือกข้อมูลของภาพยนตร์ตามปีที่ภาพยนตร์นั้นออกฉาย

13. หน้าจอแสดงข้อมูลของภาพยนตร์ตามปีที่ภาพยนตร์นั้นออกฉาย
 หน้าจอนี้แสดงข้อมูลของภาพยนตร์ตามปีที่ภาพยนตร์นั้นออกฉายที่ได้เลือกเอาไว้ โดยที่
 ข้อมูลที่แสดงก็คือ ชื่อภาพยนตร์ ประเภทภาพยนตร์ และปีภาพยนตร์เรื่องนั้นเข้าฉาย

The screenshot shows a web interface with a navigation bar at the top containing links: Recommendation, Register, Genres, Year of Release, and About. Below the navigation bar is a 'Welcome' section with a user ID 'u1000' and a 'History Voted Is 10' indicator. A search bar is present with a 'GO' button. The main content area is titled 'Year of Release :'. Below this title, there is a dropdown menu for 'Select Release Year' set to '1995' and a 'GO' button. A table lists movies from 1995 with columns for 'title', 'Type of Movie', and 'Release Year'. The table contains 15 rows of movie data. On the left side, there are two sections: 'More to Explore' with a 'Top 20' list of movies and '20 New Release' list of movies.

title	Type of Movie	Release Year
Apollo 13 (1995)	Action Drama Thriller	1995
Batman Forever (1995)	Action Adventure Comedy Crime	1995
Crimson Tide (1995)	Drama Thriller War	1995
Desperado (1995)	Action Romance Thriller	1995
Doom Generation, The (1995)	Comedy Drama	1995
Free Willy 2: The Adventure Home (1995)	Adventure Children's Drama	1995
Mad Love (1995)	Drama Romance	1995
Net, The (1995)	Sci-Fi Thriller	1995
Strange Days (1995)	Action Crime Sci-Fi	1995
To Wong Foo, Thanks for Everything! Julie Newmar (1995)	Comedy	1995
Billy Madison (1995)	Comedy	1995
Outbreak (1995)	Action Drama Thriller	1995
While You Were Sleeping (1995)	Comedy Romance	1995
Operation Dumbo Drop (1995)	Action Adventure Comedy War	1995
Toy Story (1995)	Animation Children's Comedy	1995

รูปที่ 4.21 หน้าจอแสดงข้อมูลของภาพยนตร์ตามปีที่ภาพยนตร์นั้นออกฉาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5

สรุปผลการศึกษาและข้อเสนอแนะ

สรุปผลการศึกษาระบบการแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์ได้มีการนำเทคนิคการสร้างการแนะนำที่เรียกว่าเทคนิค Collaborative filtering ซึ่งเป็นเทคนิคที่ใช้สร้างการแนะนำสินค้าโดยใช้พื้นฐานมาจากพฤติกรรมความชอบของกลุ่มลูกค้าที่มีความชอบเหมือนกัน โดยที่ระบบแนะนำสินค้าและบริการที่ใช้เทคนิค Collaborative filtering จะแนะนำสินค้าให้กับกลุ่มลูกค้าเป้าหมาย โดยใช้ฐานความรู้จากความคิดเห็นจากลูกค้าคนอื่นๆ ซึ่งเทคนิคนี้จะใช้เทคนิคทางสถิติเพื่อค้นหาเซตของลูกค้าซึ่งเรียกว่า “Neighbors” ซึ่งจะมีข้อมูลความคิดเห็นของลูกค้ากลุ่มเป้าหมายที่ผ่านมาก็คือการลงคะแนนความชอบของลูกค้ากลุ่มเป้าหมายมาเป็นพื้นฐานในการสร้างการแนะนำให้กับลูกค้าต่อไป

เนื่องจากระบบแนะนำสินค้าที่ใช้พื้นฐานมาจากพฤติกรรมความชอบของกลุ่มลูกค้าที่เหมือนกัน พบปัญหาว่าถ้าเกิดจำนวนของลูกค้ามีจำนวนเพิ่มขึ้นเป็นจำนวนมากจะทำให้ประสิทธิภาพของระบบแนะนำสินค้าลดลงคือ ต้องใช้เวลาเพิ่มขึ้นเพื่อสร้างการแนะนำสินค้าให้กับลูกค้า จึงนำเทคนิคที่เรียกว่า “Selecting Relevant Instances” มาใช้โดยจะหาความเกี่ยวพันระหว่างกลุ่มลูกค้ากับสินค้าแนะนำซึ่งจะเลือกลูกค้าที่มีความเกี่ยวพันกับสินค้าแนะนำแต่ละเรื่องมากที่สุด “High Relevance” ตามลำดับมาจำนวน K ทำให้สามารถลดจำนวนลูกค้าลงได้ $(1-K)\%$ ซึ่งช่วยลดเวลาในการสร้างการแนะนำสินค้าและอีกทั้งยังช่วยเพิ่มประสิทธิภาพระบบให้มีความแม่นยำในการแนะนำสินค้าให้ตรงกับความชอบของลูกค้าอีกด้วย

ระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์เป็นเทคโนโลยีที่เพิ่มค่าให้กับธุรกิจโดยการใช้อ้างอิงข้อมูลคะแนนความชอบของลูกค้ามาเป็นพื้นฐานในการสร้างการแนะนำ โดยที่ระบบนี้จะช่วยให้ลูกค้าสามารถหาผลิตภัณฑ์ที่ชอบได้ ในทางตรงกันข้ามระบบนี้ยังช่วยให้ธุรกิจมียอดขายเพิ่มขึ้น ซึ่งระบบการนำเสนอที่กำลังจะเป็นเครื่องมือที่สำคัญในธุรกิจพาณิชย์อิเล็กทรอนิกส์ บนเว็บไซต์ และจะถูกนำไปใช้กับข้อมูลลูกค้าที่ปริมาณมากในฐานข้อมูล และจะต้องนำไปใช้กับข้อมูลลูกค้าบนเว็บที่กำลังเพิ่มขึ้นอย่างมาก ซึ่งเทคโนโลยีระบบแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์นี้จะเป็นที่ต้องการเนื่องจากเป็นระบบที่สามารถทำกำไรให้กับธุรกิจได้เป็นอย่างมากอีกด้วย

ข้อเสนอแนะในการพัฒนาระบบมีรายละเอียดดังต่อไปนี้

1. ในการพัฒนาระบบการแนะนำสินค้าและบริการในธุรกิจพาณิชย์อิเล็กทรอนิกส์นี้ เป็นการพัฒนาเพื่อสร้างการแนะนำสินค้าให้กับลูกค้าที่เข้ามาเยี่ยมชมเว็บไซต์ของธุรกิจ ดังนั้นจึงต้องมี

เอกสารนี้เป็นเอกสารลิขสิทธิ์ของมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี ผู้ใช้เอกสารนี้ต้องรับผิดชอบต่อการใช้งาน และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การนำระบบงานนี้ไปประยุกต์ใช้กับการทำพาณิชย์อิเล็กทรอนิกส์ คือให้มีการซื้อขายสินค้า online ผ่านอินเทอร์เน็ต เพื่อที่จะทำให้เว็บไซต์นี้สามารถทำการเสนอขายและทำการแนะนำสินค้าบนอินเทอร์เน็ตให้กับธุรกิจ ได้อย่างแท้จริง

2. นอกจากข้อมูลคะแนนความชอบต่อภาพยนตร์ในเว็บไซต์ของผู้เข้ามาเยี่ยมชมที่ระบบใช้ เป็นพื้นฐานในการสร้างการแนะนำภาพยนตร์ให้กับผู้เข้ามาเยี่ยมชมคนนั้นแล้ว อาจจะใช้ ข้อมูลอื่นๆ มาเป็นพื้นฐานในการสร้างการแนะนำภาพยนตร์ให้กับผู้เข้ามาเยี่ยมชมได้อีกด้วย อย่างเช่นประวัติการซื้อสินค้าของผู้เข้ามาเยี่ยมชม ว่าผู้เข้ามาได้เคยทำการสั่งซื้อสินค้าอะไรจากทาง เว็บไซต์ไปบ้าง ระบบก็จะนำข้อมูลเหล่านี้มาเป็นพื้นฐานร่วมกับคะแนนความชอบของลูกค้า เพื่อทำการสร้างการแนะนำสินค้าที่มีประสิทธิภาพ และตรงกับความต้องการของลูกค้าให้มากที่สุด ได้อีกด้วย




บรรณานุกรม

- ไพฑูริย์ นพเกื้อ. 2546. “ระบบแนะนำสินค้าผ่านเครือข่ายอินเทอร์เน็ต. ”โครงการพัฒนาระบบงาน
วิทยาศาสตร์มหาบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ,
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง.
- Kai Yu and Zhong Wen. ET.AL. 2001. “Feature Weighting and Instance Selection for
Collaborative Filtering,” IEEE, pp. 285-290.
- Kai Yu and Zhong Wen. ET.AL. 2001. “Selecting Relevant Instances for Efficient and Accurate
Collaborative,” IEEE, pp. 239-246.
- Matthew R. McLaughlin and Jonathan L. Herlocker. 2001. “A Collaborative Filtering Algorithm
and Evaluation Metric that Accurately Model the User Experience,” IEEE.
- Stuart J. Russell and Peter Norvig. 1995. **Artificial Intelligence a Modern Approach.** New
Jersey: Prentice Hall PTR.



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาคผนวก ก.

ผลงานวิจัยที่เกี่ยวข้องกับการพัฒนาโครงการ

1. Kai Yu, Zhong Wen, Xiaowei Xu and Martin Ester, “**Feature Weighting and Instance Selection for Collaborative Filtering**,” Information and Communications, Corporate Techonology, Siemens AG Institute for Computer Science, University of Munich, pp. 285-290, 2001

Feature Weighting and Instance Selection for Collaborative Filtering*

Kai Yu², Zhong Wen², Xiaowei Xu¹, Martin Ester²

¹ Information and Communications, Corporate Technology, Siemens AG

² Institute for Computer Science, University of Munich

Xiaowei.Xu@mchp.siemens.de, {yu_k, wen, ester}@dbs.informatik.uni-muenchen.de

Abstract

Collaborative filtering uses a database about consumers' preferences to make personal product recommendations and is achieving widespread success in E-Commerce nowadays. In this paper, we present several feature-weighting methods to improve the accuracy of collaborative filtering algorithms. Furthermore, we propose to reduce the training data set by selecting only highly relevant instances. We evaluate various methods on the well-known EachMovie data set. Our experimental results show that mutual information achieves the largest accuracy gain among all feature-weighting methods. The most interesting fact is that our data reduction method even achieves an improvement of the accuracy of about 6% while speeding up the collaborative filtering algorithm by a factor of 15.

1. Introduction

The Internet is increasingly used as a channel for sales and marketing. More and more people purchase products through the Internet. One main problem that the customers face is how to find the product they like from millions of products. For the vendor, again, it is crucial to find out about the customers' preferences for products. Collaborative filtering or recommender systems have emerged in response to these problems [1] [6] [10].

Collaborative filtering accumulates a database of consumers' product preferences, and then uses them to make customer-tailored recommendations for products such as clothing, music, books, furniture, and movies. The consumer's preference can be either explicit votes or implicit usage/purchase history. Collaborative filtering can help E-commerce in converting web surfers into buyers by personalization of the web interface. It can also improve cross-sell by suggesting other products the consumer might be interested in. In a world where an E-commerce site's competitors are only a click or two away, gaining customer loyalty is an essential business strategy. Collaborative filtering can improve the loyalty by creating a value-added relationship between supplier and

consumer.

Collaborative filtering has been very successful in both research and practice. However, there still remain important research issues in overcoming two fundamental challenges for collaborative filtering [8].

The first challenge is to improve the scalability of the collaborative filtering algorithms. Existing collaborative filtering algorithms can deal with thousands of consumers within a reasonable time, but the demand of modern E-Commerce systems is to handle tens of millions of consumers.

The second challenge is to improve the quality of the recommendations for the consumers. Consumers need recommendations they can trust to help them find products they will like. If a consumer trusts a recommender system, purchases a product, but finds out he does not like the product, the consumer will be unlikely to use the recommender systems again.

In this paper, we present different feature weighting methods to improve the accuracy of collaborative filtering algorithm. Furthermore, we introduce a relevance measure of an instance to the target and propose to reduce the training data set by selecting only highly relevant instances.

In section 2, we briefly introduce collaborative filtering algorithms. We present different feature weighting methods including inverse user frequency, entropy and mutual information in section 3. We propose a mutual information based data reduction method for collaborative filtering in section 4. The empirical evaluation of these methods and results are reported in section 5. The paper ends with a summary and some interesting future work.

2. Collaborative Filtering

The task in collaborative filtering is to predict the preference of an active consumer to a given product based on a database of consumer' product preferences. There are two general classes of collaborative filtering algorithms: memory-based methods and model-based methods.

Memory-based algorithm [6][10] is the most popular

* The work was performed in Corporate Technology, Siemens AG. The contact author is Xiaowei Xu: Xiaowei.Xu@mchp.siemens.de

prediction technique in collaborative filtering applications. The basic idea is to compute the active consumer's predicted vote of a product as a weighted average of the votes given to that product by other consumers. Specifically, the prediction $p_{a,j}$ of the active consumer a on product j is given by:

$$p_{a,j} = \bar{v}_a + k \sum_{i=1}^n w(a,i)(v_{i,j} - \bar{v}_i) \quad (2.1)$$

where n is the number of the consumers who rated product j . \bar{v}_i is the mean vote for consumer i . $v_{i,j}$ is the vote cast by i on j . $w(a,i)$ is the similarity measure between a and i . k is a normalizing factor such that the absolute values of the weights sum to unity. There are two popular similarity measures: a person correlation coefficient and cosine vector similarity. Since the correlation-based algorithm outperforms the cosine vector based algorithm [1], we use the former one as the similarity measure. The person correlation coefficient [6] between consumer a and i is defined as:

$$w(a,i) = \frac{\sum_j (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j (v_{a,j} - \bar{v}_a)^2 \sum_j (v_{i,j} - \bar{v}_i)^2}} \quad (2.2)$$

Memory-based methods have the advantages of being able to rapidly incorporate the most up-to-date information and relatively accurate prediction [1], but they suffer from poor scalability for large numbers of consumers. This is because the search for all similar consumers is slow in large databases.

Model-based collaborative filtering, in contrast, uses the consumers' preference database to learn a model, which is then used for predications. The model can be built off-line over several hours or days. The resulting model is very small, very fast, and essentially as accurate as memory-based methods [1]. Model-based methods may prove practical for environments in which consumer preferences change slowly with respect to the time needed to build the model. Model-based methods, however, are not suitable for environments in which consumer preference models must be updated rapidly or frequently.

In this paper, we will focus on memory-based algorithms and present some new methods to improve the scalability and the accuracy.

3. Feature Weighting Methods

As indicated before, collaborative filtering is built on the assumption that a good way to predict the preference of the active consumer for a target product is to find other consumers who have similar preferences, and then use those similar consumers' preferences for that product to make a prediction. The similarity measure is based on preference patterns of consumers. Therefore, a

consumer's votes on the product set except the target product can be regarded as features of this consumer. Hence, introduction of some feature weighting methods may be useful to improve the accuracy of prediction. Through weighting, we can focus on the *good* products while removing *bad* ones or reducing their impacts. Votes on a 'good product' are highly relevant to the preference for the target product, while a 'bad product' is irrelevant or noisy in prediction for the target product. Such weighting methods can be derived from psychological and statistical observations. When using weight the similarity measures between consumers are modified as follows:

$$w(a,i) = \frac{\sum_j w_j^2 (v_{a,j} - \bar{v}_a)(v_{i,j} - \bar{v}_i)}{\sqrt{\sum_j w_j^2 (v_{a,j} - \bar{v}_a)^2 \sum_j w_j^2 (v_{i,j} - \bar{v}_i)^2}} \quad (3.1)$$

where w_j represent the weight of product j with respect to the target product.

3.1 Inverse User Frequency

In applications of vector similarity in information retrieval, word frequencies are typically modified by the *inverse document frequency* [7]. The idea is to reduce weights for commonly occurring words, capturing the intuition that they are not useful in identifying the topic of a document, while words that occur less frequently are more indicative of the topic. Bresse et al [1] applied an analogous transformation to votes in a collaborative filtering database, which is termed inverse user frequency. The idea is that universally liked products are not as useful in capturing similarity as less common products. So inverse user frequency weight is defined as follows:

$$w_j = \log \frac{n}{n_j} \quad (3.2)$$

where n_j is the number of consumers who have voted for product j , and n is the total number of consumers in the database. Note that if everyone has voted on product j , then the weight is zero. However, if in a database every product received about the same number of votes, this weighting method can not make sense.

3.2 Entropy

The concept of entropy was introduced as a measure of uncertainty of a random variable [9]. The diversity (or distribution) of consumer votes to a specific product will be apparently meaningful in collaborative filtering. Let's consider a special case, if all the consumers give a very high vote on a product, then it is indicated that all the votes on this product will make no sense in computing similarity between consumers, because it can't tell any distinction among consumers. But if all the votes on a

item are very diverse, almost identically distributed over the range of the vote, then all the votes on this product will be very indicative in capturing the bias of consumers. Based on the above intuition, we propose an entropy-based weighting method in collaborative filtering:

$$w_j = H_j / H_{j,max} \quad (3.3)$$

where $H_j = -\sum_i p_{i,j} \cdot \log_2 p_{i,j}$

In eq.(3.3) H_j is the entropy of product j , p_{ij} is the probability of votes on product j valued i , and $H_{j,max}$ represents the maximum entropy which assumes the distributions over all classes of vote are identical. This term is introduced to avoid the impact of different discrete vote ranges for different products. Thus, a large value of w_j means diverse preference for product j , and hence more emphases should be put on those votes for j in prediction. However, the proposed entropy-based weighting scheme might encounter the risk that there is no significant difference of the entropy from product to product. In the case of movie recommendation it is quite possible that the people's tastes towards every specific movie are all very diverse. In such a case, w_j is close to 1 for most of the movies and the entropy-based feature weighting can't result in any impressive improvement.

3.3 Mutual Information

The two feature-weighting methods mentioned above are derived from the characteristics of single products. But our task is from some knowledge about other features to make a prediction for the target. So a better way should be to explore some kind of internal connection between features and the target. If the votes of the target product are found to be highly dependent on the votes of product j , we should assign a large weight to j .

Example 1. If 50 consumers/users give votes for two movies, i and j , and vote takes the value from 0 to 1, let us consider two different situations, case 1 and case 2 respectively, as shown in fig.1. In case 1, we find consumers are nearly uniformly distributed in the movie-movie vote space. If A and B are two arbitrary users who both have close interests to movie i , it does not necessarily indicate that they have also similar preferences for movie j . But in case 2, the situation is quite different. We can find for those consumers who dislike movie i , movie j always is their favorite. While those consumers who like movie i always rate the other one just above the average. In summary, in the second case movie i should play an important role in inferring some consumer's preference for movie j , while in case 1 it is not so useful.

Formally, the dependence of product j on i may be defined by a conditional probability:

$$p(v_{j,u1} - v_{j,u2} < e | v_{i,u1} - v_{i,u2} < e) \quad (3.4)$$

where $u1$ and $u2$ represent two arbitrary consumers and e is a threshold. If the difference between two votes is below e , those two votes are regarded to be 'close'. The above conditional probability indicates the probability of the case that two arbitrary consumers have close preference for product j given the condition that those two consumers have close preference for product i .

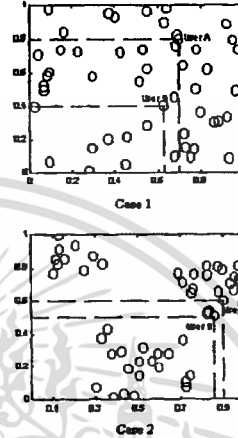


Figure 1. Consumer in example 1

To apply dependence as a weighting scheme in collaborative filtering, we could calculate it according to formula (3.4). However, this would be very expensive since its runtime complexity is $O(n^2m^2)$ if n is the number of consumers and m the number of products. Instead, we will approximate dependence by the mutual information between a feature and the target. We will see below that this approximation behaves well and is significantly more efficient to calculate.

In information theory, mutual information represents a measure of statistic dependence between two random variables X and Y with associated probability distributions $p(x)$ and $p(y)$ respectively. Following Shannon [9], the mutual information between X and Y is defined as:

$$I(X;Y) = \sum_x \sum_y p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \quad (3.5)$$

Furthermore, mutual information can be equivalently transformed into the following formulas:

$$I(X;Y) = H(X) - H(X|Y) \tag{3.6}$$

$$I(X;Y) = H(Y) - H(Y|X) \tag{3.7}$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{3.8}$$

where $H(X)$ is the entropy of X , $H(X|Y)$ is the conditional entropy of X given Y and $H(X,Y)$ is the joint entropy of two random variables. The definition of the conditional entropy, the joint entropy and the proof of the above equations can be found in [2]. The equations above indicate that mutual information also represents the reduction of entropy (uncertainty) of one variable given information of the other variable.

Theorem: Given two products i and j , as well as distributions of votes on them, $P(V_i)$ and $P(V_j)$. And ϵ is the interval of discrete value for vote. If u_1 and u_2 are two arbitrary consumers who have voted for both products, then

$$\frac{d [p(|v_{j,u_1} - v_{j,u_2}| < \epsilon | |v_{i,u_1} - v_{i,u_2}| < \epsilon)]}{d [I(V_j;V_i)]} > 0 \tag{3.9}$$

Proof:

Since $P(V_i)$ and $P(V_j)$ are given, we have:

$$d [I(V_j;V_i)] = d [H(V_j) - H(V_j|V_i)] \tag{3.10}$$

$$= -d [H(V_j|V_i)]$$

Inequality (3.8) can be written as:

$$\frac{d [p(|v_{j,u_1} - v_{j,u_2}| < \epsilon | |v_{i,u_1} - v_{i,u_2}| < \epsilon)]}{d [H(V_j|V_i)]} < 0 \tag{3.11}$$

Next, we have

$$H(V_j|V_i) = \sum_{v \in \mathbb{K}} p(V_i = v) H(V_j|V_i = v) \tag{3.12}$$

and

$$\frac{p(|v_{j,u_1} - v_{j,u_2}| < \epsilon | |v_{i,u_1} - v_{i,u_2}| < \epsilon)}{\sum_{v \in \mathbb{K}} p(V_i = v)^2 p(|v_{j,u_1} - v_{j,u_2}| < \epsilon | |v_{i,u_1} - v_{i,u_2}| < \epsilon)} > \frac{\sum_{v \in \mathbb{K}} p(V_i = v)^2}{\sum_{v \in \mathbb{K}} p(V_i = v)^2} \tag{3.13}$$

where \mathbb{K} is the set of all discrete votes. From eq. (3.12) and eq. (3.13) we can easily derive ineq. (3.11). Therefore, ineq.(3.9) holds. \square

The above theorem clearly shows that large mutual information between the feature and the target means a high dependency between them. Therefore, it encourages us to propose mutual information as a weighting method in collaborative filtering.

$$w_j = I(V_j;V_i) \tag{3.14}$$

where V_j and V_i are the votes on product j and target product i respectively. According to eq. (3.8), we use the following equation to estimate the mutual information between two products:

$$I(V_j;V_i) = H(V_j) + H(V_i) - H(V_j, V_i) \tag{3.15}$$

where $H(V_j, V_i)$ is the joint entropy between two products. Since not all the consumers have voted for the two products, calculation is done over the overlap. If the average number of overlapping consumers between two products is n , and there are totally m products in the training data set, the computational complexity for calculating the mutual information between all pairs of products is $O(nm^2)$.

4. Selecting Relevant Instances

An interesting question is, since the number of recorded consumers is explosively increasing, how to speed up the prediction? To respond to this challenge, we propose a method to reduce the training data set by selecting only highly relevant instances. In our application the instance is the consumers in the preference database.

In collaborative filtering algorithm, the computational complexity is linear with respect to the number of consumers who cast a vote to the predicted product (n in eq. 2.1). Therefore, one way to speed up the process of recommendation is to reduce the number of consumers for every target product in the training data set. This can be done through random sampling or data focussing techniques[3][4]. However these methods have the problem that the quality of the prediction is reduced due to the loss of information.

We propose a data reduction method that can even improve the quality of the prediction. Intuitively, this data reduction works as follows: First, we pick up the consumers who have given votes to many products because of their low sparsity and clear profile. Secondly, we wish to select consumers whose votes are mainly over dependant products, since those products can provide more accurate information to infer a consumer's preference. Based on the above analysis, we use the following measure to rank the *relevance* of consumer i to target product r :

$$R_{i,r} = \log n_i \cdot \frac{\sum_{M \in M_i, r \in M} I(V_j;V_i)}{n_i - 1} \tag{3.15}$$

where n_i is the number of the votes cast by i . M_i is the set of products voted by i . For every product in the training data set, we rank consumers who cast a vote to the product according to the relevance (eq. 3.15) and only the top $k\%$ of the ranking list will be used in the prediction. The rest $(1-k)\%$ will be removed from the training data set. In this case, the *selection rate* is $k\%$ and the *reduction rate* is $(1-k)\%$.

5. Experimental Evaluation

In this section, we report results of an experimental evaluation of our proposed techniques. We describe the data set used, the experimental methodology, as well as

the performance improvement compared with traditional techniques.

5.1 The EachMovie Database

We ran experiments using data from the *EachMovie* collaborative filtering service, which was part of a research project at the Systems Research Center of Digital Equipment Corporation. The database contains votes from 72,916 users on 1,628 movies. User votes were recorded on a numeric six-point scale (We transfer it to 0, 1, 2, 3, 4, and 5).

Although data from 72,916 users is available, we restrict our analysis to 35,527 users who gave at least 20 votes over the totally 1623 movies. For those users whose vote number is less than 20, since their profiles are unclear, it is hard to be used in evaluation. Moreover, to speed up our experiments, we randomly selected 10,000 users from the 35,527 users and divided them into a training set (8000 users) and a test set (2000 users).

5.2 Metrics and Methodology

Since we are interested in a system that can accurately predict a consumer's vote on a specific product, we use the mean absolute error (MAE), where the error is the value of the differences between the actual vote and the predicted vote, to evaluate the quality of prediction. This metric has been widely used in previous work[1], [5], [6] and [10].

As in [1], we also employ two protocols, *All but One*, and *Given K*. In the first class, we randomly hide an existing vote for each test consumer, and try to predict its value given all the other votes the consumer has voted on. The *All but One* experiments investigate the algorithms' performance when given as much data as possible from each test consumer, and are indicative of what might be expected of the algorithms under steady state usage where the database has accumulated a fair amount of data about a particular consumer. The second protocol, *Given K*, randomly select *K* votes from each test consumer as the observed votes, and then attempts to predict the remaining votes. It looks at consumers with less data available, and examines the performance of the algorithms when there is relatively little known about an active consumer. Its results show the performance of algorithms during the startup period, when a consumer is new to a particular collaborative filtering recommender.

5.3 Results

As shown in Fig.2, we investigate the accuracy of collaborative filtering using different feature weighting methods. The experiments were conducted for training set with 200, 500, 1000, 2000, 5000 and finally 8000

consumers. Our result show that mutual information based weighting achieves the best accuracy, yielding an improvement of about 5% compared to the standard method without feature weighting. Entropy based method, on the other hand obtain only a slight improvement. This can be explained by the fact that the variance of entropy across movies is not very large. We also find that weighting by the inverse user frequency even reduces the accuracy of prediction.

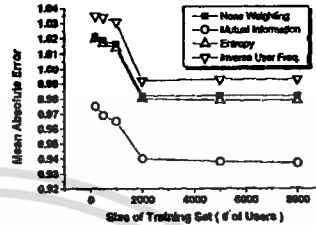


Figure 2. All but One results of feature weighting in different training sizes

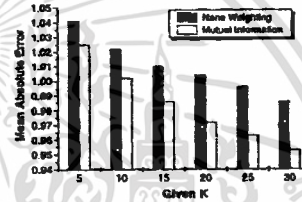


Figure 3. Given K results of mutual information weighting method

Fig.3 shows results under the protocols of *Given 5, 10, 15, 20, 25, 30*. In the six cases, mutual information weighting results in an improved accuracy. The improvement of MAE varies from 1.5% to 4.5%. The results indicate the more we know about the active consumer, the more improvement can be achieved by our weighting scheme.

We also evaluated the performance of our method of selecting relevant instances. The outcomes are given in Fig. 4 and Fig. 5. As described in section 4, we sort consumers in descending order of their relevance to each movie, and select highly relevant consumers for the prediction in different selection rates of 3.13%, 6.25%, 12.5%, 25%, 62.5% and 100%. The results are compared with the outcomes of random sampling. The proposed method outperforms random sampling in accuracy, and

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

the combination with mutual information based feature weighting results in further 4–5% improvement of mean absolute error. As shown in Figure 5, the computational complexity is linear with respect to the number of consumers in the training data set. For example, if we select 6.25% of the training size, the average prediction time for each vote is reduced from 399 ms to 27.5 ms.

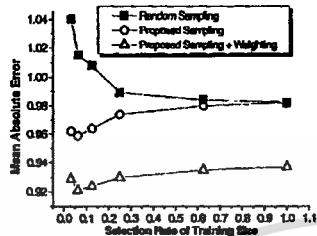


Figure 4. All But One performance for different selection rates

Moreover, Figure 4 shows that there is an optimal selection rate with respect to the accuracy, which is 6.25%. To conclude, we can achieve over 6% improvement in accuracy by using only 6.25% of the whole training data set while at the same time reducing the runtime by a factor of 15. We think it is due to the existence of irrelevant consumers in the whole training set and those irrelevant consumers are the noise for the target product.

6. Conclusion

In this paper, we present different feature weighting methods to improve the accuracy of the memory-based collaborative filtering algorithm. Furthermore, we introduce a relevance measure of an instance to the target and propose to reduce the size of training data set by selecting only highly relevant instances. We give an empirical evaluation of different feature weighting methods. Our results show that mutual information achieves the best accuracy. Our data reduction method can significantly reduce the size of the training data set and speed up the collaborative filtering algorithm. The most interesting fact is that our method even achieves an improvement on the accuracy of about 6% at a reduction rate of 94%, while random sampling decreases the accuracy by 4%.

Our result shows that feature weighting and selecting relevant instances are very promising methods for data mining. In the future, we will apply our methods

to model-based collaborative filtering algorithms. We will also investigate the performance of our methods to other applications such as web page usage mining and text mining.

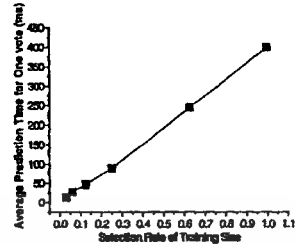


Figure 5. Prediction time for different selection rates

7. References

- [1] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, 1998.
- [2] G. Deco, and D. Obradovic, *An Information-Theoretic Approach to Neural Computing*, Springer-Verlag Inc., New York, 1996.
- [3] M. Ester, H.-P. Kriegel, and X. Xu, "Knowledge Discovery in Large Spatial Databases: Focusing Techniques for Efficient Class Identification", In *Proc. of 4th Int. Symp. on Large Spatial Databases*, Portland, ME, 1995, also in *Lecture Notes in Computer Science, Vol. 951*, Springer, 1995, pp.67-82.
- [4] M. Ester, H.-P. Kriegel, and X. Xu, "A Database Interface for Clustering in Large Spatial Databases", In *Proc. 1st Int. Conf. on Knowledge Discovery and Data Mining (KDD95)*, Montreal, Canada, 1995, pp. 94-99.
- [5] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and evaluating choices in a Virtual Community of Use", In *Proceedings of CHI95*.
- [6] F. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", In *Proceedings of the 1994 Computer Supported Collaborative Work Conference*.
- [7] G. Salton, and M. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, 1983.
- [8] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Analysis of Recommender Algorithms for E-Commerce", In *Proceedings of ACM E-Commerce 2000 Conference*.
- [9] C. E. Shannon, "A Mathematical Theory of Communication", *Bell Sys. Tech. Journal*, vol. 27, 1948
- [10] U. Shardanand, and P. Maes, "Social Information filtering Algorithms for Automating 'Word of Mouth'", In *Proceedings of CHI95*.

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายสุกชัย อัสวชินเทพกุล
วัน เดือน ปีเกิด	9 มีนาคม 2521 ที่กรุงเทพมหานคร
ประวัติการศึกษา	2532 ประถมศึกษา โรงเรียนชุมชนบึงบา 2535 มัธยมศึกษาตอนต้น สาขาคณิตศาสตร์-อังกฤษ โรงเรียนหนองเสือวิทยาคม 2538 มัธยมศึกษาตอนปลาย สาขาวิทยาศาสตร์-คณิตศาสตร์ โรงเรียนรัฐรัตน์ 2545 วิทยาศาสตร์บัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยรามคำแหง



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้