โครงการ
การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ การแนะนำร้านอาหารแบบหลายเกณฑ์โดยโครงข่ายประสาทเทียมแบบวนซ้ำ
Restaurant Multicriteria Recommendation based on Recurrent Neural Network

ชื่อนิสิต นางสาวกชพรรณ สมสา 5833602423
นางสาวณัฐสินี ศรีเวทย์บดี 5833627123

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์
สาขาวิชา วิทยาการคอมพิวเตอร์

ปีการศึกษา 2561

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของโครงงานทางวิชาการที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของโครงงานทางวิชาการที่ส่งผ่านทางคณะที่สังกัด
The abstract and full text of senior projects in Chulalongkorn University Intellectual Repository(CUIR)
are the senior project authors' files submitted through the faculty.

การแนะนำร้านอาหารแบบหลายเกณฑ์โดยโครงข่ายประสาทเทียมแบบวนซ้ำ

Restaurant Multicriteria Recommendation based on Recurrent Neural Network

นางสาวกชพรรณ สมสา
นางสาวณัฐสินี ศรีเวทย์บดี

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2561
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Restaurant Multicriteria Recommendation based on Recurrent Neural Network

Kotchapun Somsa

Natsinee Srivetbordee

A Project Submitted in Partial Fulfillment of the Requirements

for the Degree of Bachelor of Science Program in Computer Science

Department of Mathematics and Computer Science

Faculty of Science Chulalongkorn University

Academic Year 2018

Copyright of Chulalongkorn University

หัวข้อโครงงาน	การแนะนำร้านอาหารแบบหลายเกณฑ์โดยโครงข่ายประสาทเทียมแบบวนซ้ำ
โดย	นางสาวกชพรรณ สมสา
	นางสาวณัฐสินี ศรีเวทย์บดี
สาขาวิชา	วิทยาการคอมพิวเตอร์
อาจารย์ที่ปรึกษาโครงงานหลัก	ผู้ช่วยศาสตราจารย์ ดร. ศรันญา มณีโรจน์

ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับโครงงานฉบับนี้เป็นส่วนหนึ่ง ของการศึกษาตามหลักสูตรปริญญาบัณฑิต ในรายวิชา 2301499 โครงงานวิทยาศาสตร์ (Senior Project)

..... หัวหน้าภาควิชาคณิตศาสตร์

(ศาสตราจารย์ ดร.กฤษณะ เนียมมณี) และวิทยาการคอมพิวเตอร์

คณะกรรมการสอบโครงงาน

..... อาจารย์ที่ปรึกษาโครงงานหลัก

(ผู้ช่วยศาสตราจารย์ ดร. ศรันญา มณีโรจน์)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.มนนัทธ์ พงษ์พานิช)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. เก่ง วิบูลย์ธัญญ์)

กชพรรณ สมสา, ณัฐสินี ศรีเวทย์บดี: การแนะนำร้านอาหารแบบหลายเกณฑ์โดยโครงข่ายประสาทเทียมแบบวนซ้ำ. (Recurrent Multicriteria Recommendation based on Recurrent Neural Network) อ.ที่ปรึกษาโครงงานหลัก: ผศ.ดร.ศรันญา มณีโรจน์, 37 หน้า.

ระบบแนะนำถูกสร้างขึ้นมาเพื่อช่วยแนะนำสินค้า ผลิตภัณฑ์ ตลอดจนบริการต่าง ๆ ให้กับผู้ใช้ระบบ และเพื่อให้การแนะนำนั้นตรงกับความต้องการของผู้ใช้ระบบแต่ละคน จึงมีการคิดค้นระบบแนะนำวิธีการต่าง ๆ ขึ้นมากมาย ซึ่งระบบแนะนำหลักที่ใช้ในปัจจุบันคือ Collaborative Filtering (CF) และ Content-Based Filtering (CBF) โดย CF คือการแนะนำที่ดูจากการกระทำของผู้ใช้ที่เราสนใจและผู้ใช้คนอื่น ๆ ในระบบที่มีความสัมพันธ์บางอย่างกับผู้ใช้นั้น เช่น ความชอบ รสนิยม พฤติกรรมต่างๆ ที่คล้ายคลึงกับผู้ใช้ที่เราสนใจ โดยผู้ใช้อื่นที่มีความคล้ายคลึงกับผู้ใช้ที่เราสนใจ เรียกแทนว่า “เพื่อนบ้าน (Neighbor)” จากนั้นจึงนำรายการข้อมูลต่างๆ จากเพื่อนของผู้ใช้ มาแนะนำให้กับผู้ใช้ ดังนั้นการแนะนำแบบนี้จึงทำให้ได้ข้อมูลที่มีความหลากหลายในการแนะนำ ส่วน CBF คือการแนะนำที่ดูจากพฤติกรรมในอดีตของผู้ใช้ที่เราสนใจเป็นหลัก โดยจะเก็บข้อมูลนั้นมาวิเคราะห์ และแนะนำข้อมูลต่าง ๆ ที่มีความคล้ายคลึงกันกับข้อมูลนั้น มาแนะนำให้แก่ผู้ใช้ ดังนั้นการแนะนำวิธีนี้จึงสามารถแนะนำข้อมูลได้ตรงตามความชอบของผู้ใช้ ในปัจจุบัน ระบบแนะนำมีหลายโดเมน เช่น ภาพยนตร์ เพลง โรงแรม ร้านอาหาร เป็นต้น ซึ่งมีงานวิจัยเกี่ยวกับโดเมนร้านอาหาร ซึ่งอาจอยู่ในรูปแบบ CF หรือ CBF ถูกนำมาใช้ร่วมกับโครงข่ายประสาทเทียมแบบวนซ้ำเท่านั้น ซึ่ง CF มีข้อเสียคือข้อมูลขาเข้าต้องมีจำนวนมากเพียงพอที่จะนำมาเป็นฐานของข้อมูลเพื่อใช้ในการแนะนำ และ CBF มีข้อเสียคือข้อมูลขาออกของการแนะนำดังกล่าวจะมีแต่ข้อมูลที่มีความคล้ายคลึงกันกับความพึงพอใจในอดีตของผู้ใช้ทำให้ข้อมูลขาออกจะไม่มีความหลากหลาย จึงไม่เกิดการแนะนำวัตถุแบบใหม่ให้กับผู้ใช้เป้าหมาย จากเหตุผลข้างต้นจะสังเกตได้ว่าจากงานวิจัยที่ผ่านมาพบว่ายังไม่มีงานวิจัยที่นำทั้งสองรูปแบบมารวมกันในรูปแบบการแนะนำร้านอาหาร ดังนั้นงานวิจัยนี้จึงเลือกใช้วิธีการแนะนำร้านอาหาร โดยมีการประมวลผลของการกรองร่วมและการกรองแบบอิงเนื้อหาร่วมกันบนโครงข่ายประสาทเทียมแบบวนซ้ำ โดยใช้ข้อมูลวิจารณ์และลักษณะเฉพาะของวัตถุแบบหลายเกณฑ์ จากผลการทดลองการรวมขั้นตอนระหว่างข้อมูลวิจารณ์และลักษณะเฉพาะของวัตถุแบบหลายเกณฑ์ ให้ผลลัพธ์ที่ดีกว่าลักษณะเฉพาะของวัตถุแบบหลายเกณฑ์เพียงอย่างเดียว

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์ ลายมือชื่อนิสิต .....

ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาการคอมพิวเตอร์ ลายมือชื่อ อ.ที่ปรึกษาโครงงานหลัก .....

ปีการศึกษา 2561

5833602423, 5833627123: MAJOR COMPUTER SCIENCE
KEYWORDS: RECOMMENDER SYSTEM, COLLABORATIVE FILTERING (CF), CONTENT-BASED FILTERING (CBF), REVIEWS, ITEM-BASED

KOTCHAPUN SOMSA, NATSINEE SRIVETBORDEE: Recurrent Multicriteria Recommendation based on Recurrent Neural Network. ADVISOR: ASST. PROF. SARANYA MANEEROJ, Ph.D., 37 pp.

A recommender system or a recommendation system are utilized for recommend items for users including movies, books, research articles and products in general. There are many methods of recommender system. In present, we use Collaborative Filtering (CF) and Content-Based Filtering (CBF). Collaborative Filtering is a method of prediction making about the interests of a user by preferences collecting or taste information from many users that related to user called “NEIGHBORS”. The standard method of CF is known as Nearest Neighborhood algorithm. CF produces more several recommendations. Content-Based Filtering is a method of items recommendation based on a comparison between the content of items and user profile. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. The user profile is represented with the same terms and built up by analyzing historical behavior to the content of items which have been done by the user. Recommender systems currently have various domains including movies, music, hotels and restaurants. There are many restaurants researches only using one of two methods (CF or CBF) for recurrent neuron network. Data sparsity is seen as key disadvantage of CF because it should have enough informations to recommend and limited content analysis is a disadvantage of CBF leads to no various of outputs. According to previous reasons, we propose to use both of CF and CBF in restaurant recommender system by using recurrent neuron network and the inputs are

reviews and multicriteria of item-based. As a result of experiment, an integration of them has more effective than using only multicriteria of item-based.

Department: Mathematics and Computer Science Student's Signature Kotchaporn Sumsa

Student's Signature Natsinee Srivetcharapee

Field of Study: Computer Science Advisor's Signature Sarunya

Academic Year: 2018

กิตติกรรมประกาศ

งานวิจัยเรื่องการแนะนำร้านอาหารแบบหลายเกณฑ์โดยโครงข่ายประสาทเทียมแบบวนซ้ำสำเร็จลุล่วงไปได้ด้วยดีเพราะได้รับการสนับสนุน และช่วยเหลือจากคณาจารย์ และบุคคลต่างๆต่อไปนี้

ขอขอบพระคุณอาจารย์ที่ปรึกษา ผศ.ดร.ศรันญา มณีโรจน์ ที่คอยให้คำปรึกษา การสนับสนุนและคำแนะนำแก่คณะผู้จัดทำเป็นอย่างดีมาตลอดระยะเวลาวิจัย

ขอขอบพระคุณรุ่นพี่ท็อป พี่ฟลา พี่พีช พี่โค้กพี่ย้งยี้ ศิษย์เก่าที่ให้คำแนะนำในเรื่องเกี่ยวกับรูปแบบของโครงข่ายประสาทเทียมแบบวนซ้ำ และช่วยเหลือคณะผู้จัดทำในการศึกษาอัลกอริทึมรูปแบบต่างๆ

ขอขอบคุณครอบครัว และเพื่อนๆ ทุกคนที่คอยให้คำปรึกษาและแนะนำการแก้ปัญหา รวมไปถึงแนวทางการศึกษาหาข้อมูลในงานวิจัยชิ้นนี้ รวมถึงยังเป็นกำลังใจตลอดการทำงานจนกระทั่งงานวิจัยนี้ประสบความสำเร็จไปได้ด้วยดี

คณะผู้จัดทำ

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... จ

บทคัดย่อภาษาอังกฤษ ..... ช

กิตติกรรมประกาศ ..... ฌ

สารบัญ ..... ญ

สารบัญตาราง ..... ฏ

สารบัญภาพ ..... ฐ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและเหตุผลการวิจัย 1

1.2 วัตถุประสงค์ของการวิจัย 3

1.3 ขอบเขตการวิจัย 3

1.4 ขั้นตอนการวิจัย 3

1.5 ประโยชน์ที่คาดว่าจะได้รับ 4

1.6 โครงสร้างของรายงาน 4

บทที่ 2 งานวิจัยที่เกี่ยวข้อง 5

2.1 Recommendation System (RS) 5

2.2 Multi-Criteria Recommendation System 7

2.3 Classification 8

2.4 Recurrent Neural Network 9

2.5 Recommendation System กับ ร้านอาหาร 10

บทที่ 3 วิธีการวิจัย 13

3.1 การเตรียมข้อมูล 14

3.2 โครงข่ายประสาทเทียมแบบวนซ้ำกับข้อมูลวิจารณ์ร้านอาหารของเพื่อนบ้าน (CF) 15

3.3 โครงข่ายประสาทเทียมแบบวนซ้ำกับลักษณะเฉพาะของวัตถุ (CBF) 16
3.4 รวบรวมผลลัพธ์จากสองขั้นตอน 17
บทที่ 4 ผลการวิจัย 18
4.1 Data set 18
4.2 วิธีที่ใช้ในการประเมิน 18
4.3 ผลการทดลอง 20
บทที่ 5 ข้อสรุปและข้อเสนอแนะ 23
5.1 ข้อสรุป 23
รายการอ้างอิง 24
ภาคผนวก ก แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal ปีการศึกษา 2561 . 26
ประวัติผู้เขียน 32

สารบัญตาราง

หน้า

ตารางที่ 4.1 แสดงค่าสถิติต่างๆ ที่ใช้ในการทดลอง..... 18

สารบัญภาพ

หน้า

รูปที่ 2.1 User – Item rating matrix..... 6
รูปที่ 2.2 User-item multi criteria rating matrix.....7
รูปที่ 2.3.1 โครงสร้างของโครงข่ายประสาทเทียม..... 8
รูปที่ 2.4 โครงสร้างของโครงข่ายประสาทเทียมแบบวนซ้ำ.....10
รูปที่ 2.5 โครงสร้างการทำงานในการสร้างข้อความใหม่.....11
รูปที่ 3.1 ขั้นตอนการทำงานของการวิจัย.....13
รูปที่ 3.1.1 ตัวอย่างของข้อมูลวิจารณ์.....14
รูปที่ 3.1.2 ตัวอย่างของข้อมูลร้านอาหาร.....14
รูปที่ 3.1.3 ตัวอย่างของข้อมูลผู้ใช้.....15
รูปที่ 3.2.1 รูปแบบการทำงานของข้อมูลวิจารณ์ของร้านอาหาร.....15
รูปที่ 3.2.2 การทำนายความพึงพอใจของผู้ใช้เป้าหมายโดยอิงจากเพื่อนบ้าน.....16
รูปที่ 3.3 โครงข่ายประสาทเทียมแบบวนซ้ำกับลักษณะเฉพาะของวัตถุ.....17
รูปที่ 4.3.1 แสดงผลลัพธ์ความคล้ายคลึงกันของคะแนนความพึงพอใจและร้านอาหารจากข้อมูลผู้ใช้ 10 คน ในฝั่ง CF และ CBF 20
รูปที่ 4.3.2 แสดงผลลัพธ์ความคล้ายคลึงกันของคะแนนความพึงพอใจและร้านอาหารจากข้อมูลผู้ใช้ 10 คน ในฝั่ง CBF.....21
รูปที่ 4.3.3 แสดงผลลัพธ์ความคล้ายคลึงกันของคะแนนความพึงพอใจและร้านอาหารจากข้อมูลผู้ใช้ 156 คน ในฝั่ง CF และ CBF.....21
รูปที่ 4.3.4 แสดงผลลัพธ์ความคล้ายคลึงกันของคะแนนความพึงพอใจและร้านอาหารจากข้อมูลผู้ใช้ 156 คน ในฝั่ง CBF.....22
รูปที่ 5.1 แสดงภาพรวมการทำงานทั้งหมด.....23

บทที่ 1
บทนำ

1.1 ความเป็นมาและเหตุผลการวิจัย

ระบบแนะนำ (Recommendation System) หมายถึง ระบบที่แนะนำข้อมูลผลิตภัณฑ์ หรือผู้คน ซึ่งต่อไปจะถูกเรียกว่าวัตถุ (Item) ให้กับผู้ใช้ระบบโดยอ้างอิงจากสมมติฐานการเรียนรู้ข้อมูลความชอบหรือความต้องการ ณ ขณะนั้นของผู้ใช้ โดยระบบแนะนำเป็นหัวข้องานวิจัยที่สำคัญ และมีประโยชน์มากมาย เช่น การเพิ่มยอดขายให้กับธุรกิจ หรือเข้าใจสิ่งที่ผู้ใช้งานต้องการจริงๆ ในปัจจุบัน ระบบแนะนำเป็นระบบที่สำคัญในแอพพลิเคชัน หรือเว็บไซต์ที่ให้บริการออนไลน์ เช่น Wongnai, Food Panda, Line man เป็นต้น เทคนิคหลักของระบบแนะนำประกอบด้วย 2 วิธี คือ การกรองร่วม (Collaborative Filtering) และการกรองแบบอิงเนื้อหา (Content-Based-Filtering)

การกรองร่วม (Collaborative Filtering) เป็นวิธีหนึ่งที่นิยมใช้ในระบบแนะนำ โดยหาผู้ใช้งานจากข้อมูลที่ใช้ฝึกฝนซึ่งมีการให้รูปแบบของคะแนนความพึงพอใจที่คล้ายคลึงกับผู้ใช้งานเป้าหมายแล้วใช้ความนิยมของผู้ใช้งานที่คล้ายกันช่วยในการทำนาย โดยเราเรียกผู้ใช้งานที่คล้ายคลึงกับผู้ใช้งานเป้าหมายนั้นว่า เพื่อนบ้าน (Neighbors) สำหรับขั้นตอนแบบหนึ่งของการกรองร่วม เนื่องจากรับข้อมูลคะแนนความพึงพอใจในรูปแบบเมทริกซ์แล้วหาผู้ใช้คนอื่นๆ k คน ที่มีความพึงพอใจใกล้เคียงกันมากที่สุด (k-Nearest Neighbors) เพื่อหาผู้ใช้งานที่มีการให้คะแนนความพึงพอใจคล้าย ๆกันหลังจากนั้นหากต้องการทำนายคะแนนความพึงพอใจของผู้ใช้งานเป้าหมาย ให้ใช้สมการที่ 1 ในการทำนายคะแนนความพึงพอใจของผู้ใช้งานคนนั้นต่อวัตถุที่ต้องการโดย u คือผู้ใช้งานที่ต้องการจะทำนายคะแนนความพึงพอใจ, i คือวัตถุ, v หมายถึงเพื่อนบ้าน, N_u คือชุดข้อมูลเพื่อนบ้านของผู้ใช้งาน u, w_{uv} คือ ความเหมือนระหว่างผู้ใช้งาน u และเพื่อนบ้าน v, r_{vi} คือคะแนนความพึงพอใจที่เพื่อนบ้าน v บนวัตถุ i และ r_{ui} คือคะแนนความพึงพอใจที่ทำนายได้ของผู้ใช้งาน u บนวัตถุ i

$$\hat{r}_{ui} = \frac{\sum v \in N_u \, w_{uv} r_{vi}}{\sum v \in N_u \, |w_{uv}|} \qquad (1)$$

การกรองแบบอิงเนื้อหา (Content-Based Filtering) เป็นอีกวิธีหนึ่งที่นิยมใช้ในระบบแนะนำ โดยใช้พื้นฐานข้อมูลจากรายละเอียดของวัตถุที่ผู้ใช้เป้าหมายเคยให้ความพึงพอใจไว้ในอดีต ทำให้ได้ประวัติการใช้งานวัตถุของผู้ใช้เป้าหมาย (User Profile) จากนั้นจะนำประวัติดังกล่าวไป

คำนวณหาค่าความคล้ายคลึงกันกับรายละเอียดข้อมูลของวัตถุอื่นๆแล้วดูวัตถุที่มีความสัมพันธ์กันมากที่สุดกับประวัติการใช้งานมาแนะนำให้กับผู้ใช้เป้าหมาย

เนื่องจากระบบแนะนำเป็นการใช้เทคนิคการจำแนกข้อมูล (Classification) อย่างหนึ่ง ซึ่งมีป้ายกำกับ (Label) เพื่อบอกประเภท โดยพื้นฐานแล้วเราต้องมีชุดข้อมูลหนึ่งก่อน โดยชุดข้อมูลนี้ได้จากการเก็บตัวอย่างจริง แล้วส่งให้คอมพิวเตอร์เพื่อทำการเรียนรู้ที่จะแยกประเภทโดยอาศัยการเรียนรู้จากข้อมูลที่มีอยู่ จากนั้นจะนำข้อมูลปัจจุบันมาทดสอบและดูว่าข้อมูลที่ถูกทดสอบสามารถแยกประเภทได้ถูกต้อง ตามประเภทของข้อมูลจริง ด้วยเหตุนี้ โครงข่ายประสาทเทียมก็เป็นการใช้เทคนิคการจำแนกข้อมูลเช่นเดียวกัน จึงมีผู้วิจัยอื่นนำโครงข่ายประสาทเทียม (Neural Network) มาประยุกต์บนระบบแนะนำ โดยแต่เดิมนั้นกระบวนการกรองร่วมและการกรองแบบอิงเนื้อหาได้นำข้อมูลของผู้ใช้ (Users) และข้อมูลของวัตถุ (Items) มาเป็นข้อมูลขาเข้าในการดำเนินการ จากนั้นมีผู้วิจัยได้เริ่มนำเทคนิคการฝังตัว (Embedding) [1] เข้ามาใช้ในโครงข่ายประสาทเทียมซึ่งจะจัดคุณลักษณะของผู้ใช้ (User Characteristic) และคุณลักษณะของวัตถุ (Item Characteristic) เป็นเวกเตอร์ แล้วนำไปเป็นข้อมูลขาเข้าของระบบแนะนำ เพื่อให้มีประสิทธิภาพที่ดีขึ้น

ในโครงข่ายประสาทเทียมประกอบไปด้วย โครงข่ายประสาทเทียมตามแบบแผน (Conventional Neural Network), โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) [2] และ โครงข่ายประสาทเทียมแบบวนซ้ำสองทาง (Bidirectional Recurrent Neural Network) ซึ่งแบบโครงข่ายประสาทเทียมดั้งเดิม คือการนำข้อมูลขาเข้า (Inputs) มาวิเคราะห์ โดยนำไปเรียนรู้เชิงลึกในชั้นระหว่างกลาง (Hidden layer) เพื่อให้ได้ผลลัพธ์ (Outputs) ออกมาในรูปแบบง่ายๆ เช่น ชอบ หรือ ไม่ชอบ เป็นต้น เนื่องจากในปัจจุบันความพึงพอใจของผู้ใช้ที่มีต่อวัตถุมีการเปลี่ยนแปลงตลอดเวลา

เพราะฉะนั้นระบบแนะนำที่ใช้รูปแบบโครงข่ายประสาทเทียมแบบดั้งเดิมจึงไม่สามารถวิเคราะห์ข้อมูลในรูปแบบที่เปลี่ยนแปลงตามลำดับเวลาเพื่อบอกความสัมพันธ์ระหว่างข้อมูลขาเข้าตั้งแต่อดีตจนถึงปัจจุบันได้ ดังนั้นสำหรับการทำงานที่มีข้อมูลขาเข้าเป็นอนุกรมชุดรูปแบบ (Series) จึงเหมาะกับการใช้โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neuron Network) เช่น เดือนมิถุนายนผู้ใช้ A รับประทานร้านส้มตำ เดือนกรกฎาคมรับประทานร้านพิซซ่า เดือนสิงหาคมรับประทานร้านติ่มซำ ซึ่งมีการเก็บรวบรวมข้อมูลแบบอนุกรมมาจนถึงปัจจุบันแล้วนำมาแนะนำว่าร้านถัดไปควรรับประทานอาหารร้านใด

ในปัจจุบัน ระบบแนะนำมีหลายโดเมน เช่น ภาพยนตร์ เพลง โรงแรม ร้านอาหาร เป็นต้น โดเมนร้านอาหารถูกนำมาใช้ในโครงข่ายประสาทเทียมแบบวนซ้ำ ซึ่งอาจอยู่ในรูปแบบของการกรองร่วมหรือการกรองแบบอิงเนื้อหา ซึ่งรูปแบบการกรองร่วมจะสามารถแนะนำวัตถุให้กับผู้ใช้เป้าหมาย (Target User) โดยมีการอ้างอิงจากเพื่อนบ้านที่มีความพึงพอใจคล้ายคลึงกัน และ การกรองแบบอิงเนื้อหา เป็นการแนะนำวัตถุที่มีความคล้ายคลึงกันกับประวัติการใช้งานวัตถุของผู้ใช้

เป้าหมาย แต่มีข้อเสียคือข้อมูลขาออกของการแนะนำดังกล่าวจะมีแต่ข้อมูลที่มีคล้ายคลึงกันกับความพึงพอใจในอดีตของผู้ใช้ทำให้ข้อมูลขาออกจะไม่มีความหลากหลาย จึงไม่เกิดการแนะนำวัตถุแบบใหม่ให้กับผู้ใช้เป้าหมาย จากเหตุผลข้างต้นจะสังเกตได้ว่าจากงานวิจัยที่ผ่านมาพบว่ายังไม่มีงานวิจัยที่นำทั้งสองรูปแบบมารวมกันในรูปแบบการแนะนำร้านอาหาร ผู้วิจัยจึงเสนอรูปแบบโดยเลือกใช้โดเมนร้านอาหารที่มีข้อมูลขาเข้า 2 อย่างคือ การเขียนวิจารณ์ร้านอาหาร (Reviews) ซึ่งจะนำมาใช้กับการกรองร่วม และ การฝังวัตถุ (Item Embedding) นำมาใช้กับการกรองแบบอิงเนื้อหา ในการนำสองวิธีมารวมกันนั้น ผู้วิจัยจะใช้วิธี กลไกความสนใจ (Attention Mechanism) [3] ที่มีการถ่วงน้ำหนัก (Weight) กับข้อมูลขาออกของวิธีการกรองร่วม แล้วนำไปเป็นข้อมูลขาเข้าของวิธีการกรองแบบอิงเนื้อหาเพื่อแนะนำร้านอาหารให้กับผู้เป้าใช้หมายที่จะรับประทานเป็นร้านถัดไป ดังนั้นผู้วิจัยคาดว่าการรวมรูปแบบทั้งสองสามารถให้ผลลัพธ์ที่ดียิ่งขึ้น

1.2 วัตถุประสงค์ของการวิจัย

เพื่อเสนอขั้นตอนวิธีใหม่ที่สามารถรวมวิธีการกรองร่วมและการกรองแบบอิงเนื้อหาเพื่อทำงานบนระบบแนะนำร้านอาหารโดยโครงข่ายประสาทเทียมแบบวนซ้ำ

1.3 ขอบเขตการวิจัย

ข้อมูลที่นำมาทดสอบเป็นชุดข้อมูล ที่มีการวิจารณ์ร้านอาหารจากผู้ใช้ไม่เกิน 1,000 รายการ

1.4 ขั้นตอนการวิจัย

1. ศึกษางานวิจัยและบทความทางวิชาการที่เกี่ยวข้องกับระบบแนะนำ
2. ระบุปัญหาและข้อเสียที่เกิดขึ้นจากงานวิจัยก่อนหน้า
3. ออกแบบและวิเคราะห์วิธีการแก้ปัญหาของระบบแนะนำ
4. พัฒนาและทดสอบความถูกต้องของระบบแนะนำที่ได้นำเสนอ
5. ทำการทดลองเพื่อวัดประสิทธิภาพของระบบที่ได้นำเสนอ
6. วิเคราะห์และอภิปรายผลการทดลองพร้อมทั้งนำผลการอภิปรายไปปรับปรุง ระบบแนะนำที่ได้นำเสนอให้ดีขึ้น
7. จัดทำเอกสารต่าง ๆที่เกี่ยวข้อง เช่น แบบเสนอหัวข้อโครงงาน รูปเล่มโครงงานวิทยาศาสตร์ระดับปริญญาบัณฑิต และเอกสารประกอบการนำเสนอ

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ประโยชน์ในด้านความรู้และประสบการณ์ต่อตัวนิสิตเอง
   1.1. ได้เรียนรู้ทฤษฎีและการปฏิบัติของการสร้างขั้นตอนวิธีใหม่สำหรับทำระบบแนะนำ
   1.2. ได้นำความรู้ในสาขาวิชาวิทยาการคอมพิวเตอร์ในการแก้ปัญหากับระบบแนะนำ
2. ประโยชน์ในด้านการแก้ไขปัญหาของสังคมหรือสภาพแวดล้อม
   2.1. ได้พัฒนาและต่อยอดความรู้ใหม่ซึ่งมีประโยชน์ต่อวงการวิจัย
   2.2. ได้พัฒนาองค์ความรู้ซึ่งจะสามารถนำไปประยุกต์ใช้กับธุรกิจในอนาคต

1.6 โครงสร้างของรายงาน

บทที่ 2 จะกล่าวถึงระบบแนะนำซึ่งประกอบไปด้วยการกรองร่วม (Collaborative Filtering) และการกรองแบบอิงเนื้อหา (Content-Based Filtering) จากนั้นจะกล่าวถึงการแนะนำแบบหลายเกณฑ์ การจำแนกข้อมูล โครงข่ายประสาทเทียม โครงข่ายประสาทเทียมแบบวนซ้ำ และงานวิจัยที่เกี่ยวข้องกับงาน ตามลำดับ

บทที่ 3 จะกล่าวถึงวิธีการวิจัย ซึ่งจะประกอบไปด้วยการเตรียมข้อมูล โครงข่ายประสาทเทียมแบบวนซ้ำกับข้อมูลวิจารณ์ร้านอาหารของเพื่อนบ้าน โดยมีหัวข้อย่อยคือ การนำข้อมูลวิจารณ์ของร้านอาหารและความพึงพอใจทั้งหมดมาสร้างโมเดล และการทำนายความพึงพอใจของผู้ใช้เป้าหมาย โดยได้นำคะแนนความพึงพอใจของเพื่อนบ้านมาอ้างอิง ลำดับต่อมาคือ โครงข่ายประสาทเทียมแบบวนซ้ำกับลักษณะเฉพาะของวัตถุ และรวบรวมผลลัพธ์จากสองขั้นตอน ตามลำดับ

บทที่ 4 จะกล่าวถึงผลการวิจัย ซึ่งประกอบไปด้วยขอบเขตของข้อมูลทั้งหมดที่นำมาใช้งานวิจัย วิธีที่ใช้ในการประเมินงานวิจัย และผลการทดลองที่ได้จากงานวิจัย ตามลำดับ

บทที่ 5 จะกล่าวถึงข้อสรุปและผลัพธ์ที่ได้จากงานวิจัย

บทที่ 2
งานวิจัยที่เกี่ยวข้อง

ในบทนี้ ส่วนแรกจะมีการแนะนำระบบแนะนำ หรือ Recommendation System (RS) โดยจะกล่าวถึงแนวคิดที่ใช้ในการสร้างระแบบแนะนำ จากนั้นจะกล่าวถึงเทคนิคต่างๆ รวมถึงการกรองแบบอิงเนื้อหา (Content-Based Filtering) และ การกรองร่วม (Collaborative Filtering) ส่วนที่สองเราจะกล่าวถึงการจัดกลุ่มประเภทของข้อมูล (Classification) ซึ่งโครงข่ายประสาทเทียม (Neuron Network) ได้นำเทคนิคดังกล่าวมาใช้ โดยมีโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neuron Network) เป็นประเภทหนึ่งของโครงข่ายประสาทเทียมซึ่งเราจะกล่าวในส่วนที่สาม ส่วนสุดท้ายจะกล่าวถึงการนำ Recommendation System มาใช้กับโดเมนร้านอาหาร

2.1 Recommendation System (RS)

ระบบแนะนำ หรือ Recommendation system เป็นระบบที่ถูกสร้างขึ้นเพื่อทำนายคะแนนความพึงพอใจ (Rating) ที่ผู้ใช้มีให้ต่อข้อมูลวัตถุ จากนั้นจึงนำสิ่งของเหล่านั้นมาแนะนำให้ผู้ใช้งานระบบอันเป็นการช่วยกรองเฉพาะข้อมูลที่ผู้ใช้ระบบสนใจซึ่งทำให้เกิดความสะดวกสบายและประหยัดเวลาในการรับชมสิ่งของต่างๆ ประกอบด้วย 2 เทคนิคหลัก คือ การกรองแบบอิงเนื้อหา (Content-Based Filtering) และ การกรองร่วม (Collaborative Filtering)

2.1.1 Content-Based Filtering (CBF)

การกรองแบบอิงเนื้อหา หรือ Content-Based Filtering (CBF) เป็นรูปแบบหนึ่งที่นิยมใช้ในระบบแนะนำ โดยจะแนะนำรายการที่คล้ายกับรายการที่ผู้ใช้ที่ชอบในอดีต ตัวอย่างเช่นผู้ใช้ A ให้คะแนนความพึงพอใจกับชุดของรายการ { i_1, i_2, i_3, i_4 } ในระดับคะแนนความพึงพอใจสูง ดังนั้นระบบเรียนรู้ว่าผู้ใช้ A ชอบรายการเหล่านี้แล้วแนะนำรายการอื่น ๆ ที่คล้ายกับชุดรายการที่ผู้ใช้ A พึงพอใจให้กับผู้ใช้ A ขั้นแรกระบบจะวิเคราะห์ชุดของคะแนนความพึงพอใจของผู้ใช้เพื่อสร้างโปรไฟล์ของผู้ใช้ซึ่งแสดงถึงความสนใจของผู้ใช้ที่มีต่อวัตถุ โดยปกติสามารถถูกแทนด้วยเวกเตอร์ จากนั้นระบบจะค้นหาความคล้ายคลึงกันระหว่างโปรไฟล์ของผู้ใช้และข้อมูลวัตถุอื่น ๆ มีหลากหลายวิธีการในการวัดความคล้ายคลึงกัน [4]

ข้อดีของวิธีแนะนำแบบ CBF คือความเป็นอิสระของผู้ใช้ CBF ใช้เฉพาะข้อมูลของผู้ใช้ที่ใช้งานอยู่ในการคาดการณ์ ดังนั้นระบบสามารถแนะนำรายการให้กับผู้ใช้ที่ใช้งานอยู่แม้ว่าจะมีผู้ใช้เพียงรายเดียวในฐานข้อมูลหรือรายการเป้าหมายที่ไม่ได้รับการให้คะแนนความพึงพอใจโดยผู้ใช้

รายใดก่อนหน้านี้ อย่างไรก็ตาม CBF ก็มีข้อจำกัด เนื่องจาก CBF จะแนะนำรายการที่เจาะจงให้กับผู้ใช้เพราะอิงจากประวัติการใช้งาน ทำให้ไม่สามารถแนะนำรายการที่แตกต่างไปจากเดิมให้ผู้ใช้ได้

2.1.2 Collaborative Filtering (CF)

การกรองร่วม หรือ Collaborative Filtering (CF) เป็นวิธีหนึ่งที่นิยมใช้ในระบบแนะนำ และเป็นเทคนิคที่ช่วยแก้ไขปัญหาข้อมูลที่เกิดขึ้นโดยบังเอิญ ซึ่งเกิดขึ้นโดยใช้ชุดของคะแนนความพึงพอใจของผู้ใช้รายอื่นเพื่อแนะนำข้อมูลวัตถุให้กับผู้ใช้ที่ใช้งานอยู่ ขั้นตอนหลักของ CF ประกอบด้วยการค้นหาเพื่อนบ้านและการทำนายคะแนนความพึงพอใจ ก่อนอื่นจะค้นหาผู้ใช้ที่มีลักษณะคล้ายกับผู้ใช้ที่ใช้งานอยู่ซึ่งเรียกว่าเพื่อนบ้าน จากนั้นชุดคะแนนความพึงพอใจของเพื่อนบ้านจะถูกใช้เพื่อแนะนำผู้ใช้ที่ใช้งานอยู่ เนื่องจากข้อมูลผู้ใช้อื่นมีความหลากหลายมากกว่าข้อมูลผู้ใช้ที่ใช้งานอยู่ CF สามารถแก้ไขปัญหา serendipitous ข้อมูลขาเข้าของ CF คือเมทริกซ์การจัดคะแนนความพึงพอใจของผู้ใช้ต่อวัตถุดังแสดงในรูปที่ 2.1

	i_1	i_2	i_3	i_4	i_5
u_1	5	-	4	-	2
u_2	4	5	-	4	3
u_3	5	5	4	3	2
u_4	4	-	-	4	-
u_5	2	5	3	-	1

ภาพที่ 2.1 User-item rating matrix

จากรูปที่ 2.1 ผู้ใช้ห้าคนให้คะแนนห้ารายการ สมมติว่าผู้ใช้ u_1 เป็นผู้ใช้ที่ใช้งานอยู่ ในการค้นหาเพื่อนบ้านจะต้องหาความคล้ายคลึงกันระหว่างผู้ใช้ u_1 กับผู้ใช้รายอื่นจากรายการที่ได้รับการให้คะแนนความพึงพอใจ ตัวอย่างเช่นรายการที่มีการจัดอันดับร่วมกันของผู้ใช้ u_1 และผู้ใช้ u_2 คือ { i_1, i_5 } ความคล้ายคลึงกันระหว่างผู้ใช้สามารถวัดได้โดยความเหมือนโคไซน์ [4] หรือความสัมพันธ์ของเพียร์สัน [4]

ให้ u หมายถึงผู้ใช้ที่ใช้งาน, v หมายถึงผู้ใช้อื่น, i หมายถึงรายการเป้าหมาย, r_{ui} หมายถึงการให้คะแนนความพึงพอใจของผู้ใช้งาน u บนวัตถุ i, r_{vi} หมายถึงการให้คะแนนความพึงพอใจของผู้ใช้งาน v บนวัตถุ i, $\bar{r}_u$ หมายถึงคะแนนเฉลี่ยที่ผู้ใช้ u ให้คะแนนคะแนนความพึงพอใจในทุกๆรายการ และ $\bar{r}_v$ หมายถึงคะแนนเฉลี่ยที่ผู้ใช้ v ให้คะแนนคะแนนความพึงพอใจในทุกๆรายการ

การคำนวณ Cosine similarity ระหว่างผู้ใช้ u และ v สามารถทำได้ดังสมการ (1):

$$cosine\ (u,v) = \frac{\sum_i r_{ui} r_{vi}}{\sqrt{\sum_i r_{ui}^2}\sqrt{\sum_i r_{vi}^2}} \quad (1)$$

การคำนวณ Pearson's correlation ระหว่างผู้ใช้ u และ v สามารถทำได้ดังสมการ (2):

$$Pearson\ (u,v) = \frac{\sum_{i \in I_u \cap I_v}(r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v}(r_{ui} - \bar{r}_u)^2}\sqrt{\sum_{i \in I_u \cap I_v}(r_{vi} - \bar{r}_v)^2}} \quad (2)$$

ในขั้นตอนการทำนายคะแนนความพึงพอใจ CF สร้างการทำนายคะแนนความพึงพอใจโดยใช้ชุดข้อมูลของเพื่อนบ้าน วิธีที่นิยมมากที่สุดคือผลรวมถ่วงน้ำหนัก ให้ N_u แสดงถึงชุดข้อมูลเพื่อนบ้านของผู้ใช้ u, w_{uv} หมายถึงความคล้ายคลึงกันระหว่างผู้ใช้ u และเพื่อนบ้าน v, การคาดการณ์คะแนนความพึงพอใจสำหรับรายการ i ของผู้ใช้ที่ใช้งาน u สามารถคำนวณดังสมการ (3):

$$\hat{r}_{ui} = \frac{\sum v \in N_u\ w_{uv} r_{vi}}{\sum v \in N_u\ |w_{uv}|} \quad (3)$$

ข้อดีของวิธีแนะนำแบบ CF คือ ไม่ต้องการคำอธิบายเกี่ยวกับวัตถุ และสามารถแนะนำวัตถุได้หลากหลายหมวดมากกว่า แต่วิธีการแนะนำแบบ CF นั้นก็มีข้อเสียคือ ระบบต้องการข้อมูลในปริมาณที่มากพอสมควรจึงจะเริ่มทำการแนะนำได้ บ่อยครั้งจะทำการแนะนำวัตถุที่กำลังเป็นที่นิยมซึ่งวัตถุเหล่านั้นอาจจะไม่ตรงกับรสนิยมของผู้ใช้ นอกจากนี้ ในระบบต้องมีข้อมูลจำนวนมากพอสมควรจึงจะเริ่มแนะนำได้

CBF และ CF สามารถแนะนำรายการให้กับผู้ใช้ที่ใช้งานอยู่โดยใช้การให้คะแนนความพึงพอใจแบบเกณฑ์เดียวเพื่อทำการคาดการณ์ แต่การใช้การวัดแบบเกณฑ์เดียวนั้นไม่ดีพอที่จะแสดงความชอบของผู้ใช้

2.2 Multi-Criteria Recommendation System

ระบบแนะนำแบบหลายเกณฑ์ หรือ Multi-Criteria Recommendation System (MC) ช่วยให้ผู้ใช้แสดงความพึงพอใจได้มากกว่าหนึ่งด้าน โดยระบบสามารถแนะนำข้อมูลจากการแสดงความพึงพอใจของผู้ให้บริการได้อย่างถูกต้องและลึกซึ้ง ดังนั้นระบบแนะนำแบบหลายเกณฑ์สามารถทำให้การแนะนมีคุณภาพที่ดีขึ้น ผู้ใช้สามารถระบุความพึงพอใจของพวกเขาสำหรับวัตถุหนึ่งๆ ดังแสดงในรูปที่ 2.2:

	i_1	i_2	i_3	i_4
u_1	(5,5,3,4)	(1,3,3,3)	-	(1,2,2,2)
u_2	-	(3,3,1,3)	(2,1,1,2)	(3,3,1,3)

รูปที่ 2.2 User-item multi criteria rating matrix

จากรูปที่ 2.2 แสดงเมทริกซ์การแสดงความพึงพอใจแบบหลายเงื่อนไขระหว่างผู้ใช้ { u_1, u_2 } และรายการ { i_1, i_2, i_3, i_4 } ผู้ใช้ให้คะแนนความพึงพอใจรายการย่อยสามข้อและคะแนนความพึงพอใจโดยรวมหนึ่งรายการ พิจารณาการให้คะแนนที่ผู้ใช้สองคนให้คะแนนรายการ i_2 ผู้ใช้ u_1 และผู้ใช้

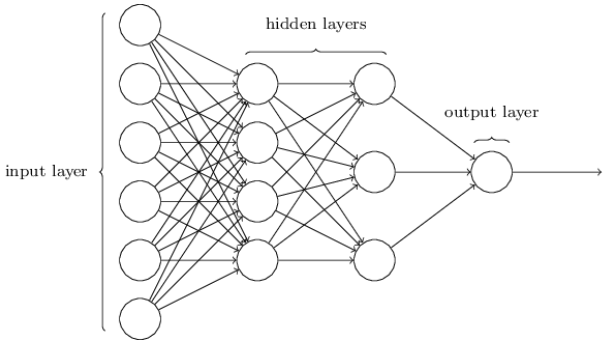
u_2 มีคะแนนเดียวกันในการให้คะแนนความพึงพอใจโดยรวมคือ 3 หากเราใช้การจัดอันดับโดยรวมเพียงอย่างเดียวดูเหมือนว่าผู้ใช้ u_1 และผู้ใช้ u_2 ให้คะแนนความพึงพอใจในระดับเดียวกัน หากเราพิจารณารายการย่อยอื่น ๆ ผู้ใช้ u_1 ให้คะแนนความพึงพอใจคือ 3 สำหรับรายการย่อยที่สองและสาม แต่ให้คะแนนความพึงพอใจสำหรับรายการย่อยแรกคือ 1 และผู้ใช้ u_2 ให้คะแนนความพึงพอใจรายการย่อยแรกและรายการย่อยที่สองคือ 3 แต่ให้คะแนนความพึงพอใจในรายการย่อยที่ 3 คือ 1 แสดงให้เห็นว่าแม้ว่าพวกเขาจะมีคะแนนโดยรวมที่เหมือนกัน แต่ก็ไม่ได้หมายความว่าพวกเขามีความพึงพอใจต่อสิ่งเดียวกันเหมือนกัน

2.3 Classification

การจำแนกข้อมูล หรือ Classification เป็นรูปแบบที่ระบบแนะนำนำมาใช้โดยมีป้ายกำกับ (Label) เพื่อบอกประเภท โดยพื้นฐานแล้วเราต้องมีชุดข้อมูลหนึ่งก่อน โดยชุดข้อมูลนี้ได้จากการเก็บตัวอย่างจริง แล้วส่งให้คอมพิวเตอร์เพื่อทำการเรียนรู้ที่จะแยกประเภทโดยอาศัยการเรียนรู้จากข้อมูลที่มีอยู่ จากนั้นจะนำข้อมูลปัจจุบันมาทดสอบและดูว่าข้อมูลที่ถูกทดสอบสามารถแยกประเภทได้ถูกต้อง ตามประเภทของข้อมูลจริง ด้วยเหตุนี้ โครงข่ายประสาทเทียมก็เป็นการใช้เทคนิคการจำแนกข้อมูลเช่นเดียวกัน

2.3.1 Neural Network

โครงข่ายประสาทเทียม หรือ Neural Network (NN) มีการทำงานคล้ายเซลล์ประสาทภายในสมองของมนุษย์ โดยเครือข่ายประสาทเทียมจะทำงานแบบขนานและมีการจัดเรียงเป็นชั้น ซึ่งโหนดซ้ายสุดในเครือข่ายเรียกว่า ชั้นข้อมูลขาเข้า (Input Layer) และเส้นประสาทภายในชั้นเรียกว่าเซลล์ประสาทขาเข้า โหนดขวาสุดเรียกว่าชั้นข้อมูลขาออกจะมีเซลล์ประสาทขาออก (Output Layer) ซึ่งในกรณีนี้จะมีเซลล์ประสาทขาออกเพียงเซลล์เดียว ชั้นกลางเรียกว่าชั้นที่ซ่อน (Hidden Layer) เนื่องจากเซลล์ประสาทในเลเยอร์นี้ไม่ใช่อินพุตหรือเอาท์พุต ดังรูปที่ 2.3.1



รูป 2.3.1 โครงสร้างของโครงข่ายประสาทเทียม

การทำงานของเครือข่ายประสาทเทียม เมื่อมีข้อมูลขาเข้า (Input) เข้ามายังเครือข่ายประสาท (Network) จะนำข้อมูลนำเข้ามาคูณกับค่าน้ำหนัก (Weight) ของแต่ละขา ผลที่ได้จากข้อมูลขาเข้า

ทุกๆขาของโครงข่ายประสาทเทียมจะนำมารวมกับและนำมาเปรียบเทียบกับระดับการกระตุ้นต่ำสุดที่ทำให้เกิดการตอบสนอง (Threshold) ที่กำหนดไว้ ถ้าผลรวมมีค่ามากกว่าค่าที่ทำให้เกิดการตอบสนอง โครงข่ายประสาทเทียมจะส่งข้อมูลขาออก เพื่อนำไปใช้เป็นข้อมูลขาเข้าในโครงข่ายประสาทเทียมอื่นๆ ที่เชื่อมกันในโครงข่ายจนได้เป็นผลลัพธ์ที่ต้องการ หลังจากนั้นจะต้องมีการตรวจสอบผลลัพธ์ที่ได้จากระบบเพื่อหาความคลาดเคลื่อน หากค่าความคลาดเคลื่อนที่ได้จากเป้าหมายมีมากเกินไปจะต้องนำไปปรับนักหนักการเรียนรู้ใหม่ (Weight) แบบแพร่กระจายย้อนกลับ (Back-propagation) ซึ่งจะเป็นการปรับน้ำหนักความ ความคลาดเคลื่อนจากชั้นข้อมูลขาออกไปยังชั้นก่อนหน้า และทำการปรับน้ำหนักไปจนถึงชั้นข้อมูลขาเข้า ซึ่งการเรียนรู้แบบนี้จะต้องอาศัยการทำซ้ำหลายครั้งจนกว่าจะได้ข้อมูลขาออกตามที่กำหนดหรือได้ค่าความคลาดเคลื่อนน้อยจนพอยอมรับได้ ซึ่งจำนวนรอบในการทำซ้ำนั้นจะขึ้นกับความยากง่ายของปัญหา ขนาดข้อมูล รวมไปถึงจำนวนชั้นของโครงสร้างของโครงข่ายที่เราสร้างไว้ด้วย

ความโดดเด่นของโครงข่ายประสาทเทียมคือ จะมีการปรับค่าเมื่อเริ่มเรียนรู้และระหว่างการดำเนินการเกี่ยวกับข้อมูล รูปแบบการเรียนรู้ขั้นพื้นฐานส่วนใหญ่เน้นที่การให้น้ำหนักของข้อมูลขาเข้าซึ่งเป็นวิธีที่แต่ละโหนดให้น้ำหนักความสำคัญของการป้อนข้อมูลแต่ละชั้น ข้อมูลที่มีค่าน้ำหนักที่สูงจะนำไปสู่การให้คำตอบที่ถูกต้อง ซึ่งโครงข่ายประสาทเทียมประกอบด้วยหลายรูปแบบ เช่น โครงข่ายประสาทเทียมตามแบบแผน (Convolutional Neural Network) เป็นที่นิยมในปัจจุบัน โดยเฉพาะบริเวณที่มีการทำงานเกี่ยวข้องกับการรับรู้ภาพ อัลกอริทึมเครือข่ายประสาทประเภทนี้มีการนำมาใช้ในแอปพลิเคชั่นเกี่ยวกับรูปภาพ เช่น การจดจำใบหน้าการแปลงข้อความและการประมวลผลภาษาธรรมชาติ เป็นต้น อีกหนึ่งรูปแบบที่จะกล่าวคือโครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) โครงข่ายนี้จะมีความซับซ้อนมากขึ้น โดยมีรูปแบบการเรียนรู้เชิงลึกแบบวนซ้ำ และบันทึกผลลัพธ์ของโหนดการประมวลผลหลังจากนั้นจะนำผลลัพธ์กลับเข้าไปในโมเดลเพื่อทำการเรียนรู้ใหม่อีกครั้ง

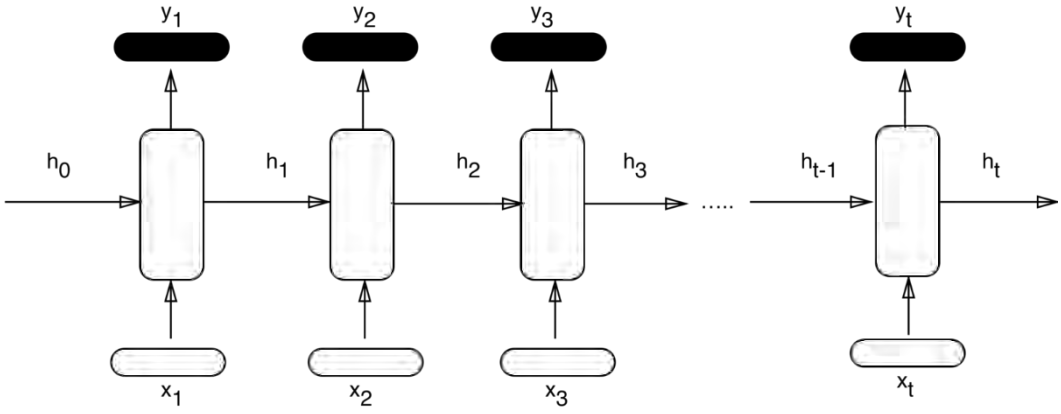
2.4 Recurrent Neural Network

โครงข่ายประสาทเทียมแบบวนซ้ำ หรือ Recurrent Neural Network (RNN) [5] เป็นโครงข่ายที่ประกอบไปด้วยโหนดอย่างน้อยหนึ่งโหนดที่มีการเชื่อมต่อแบบไหลเป็นวงกลมได้ สามารถช่วยให้โครงข่ายทำการประมวลผลชั่วคราว และเรียนรู้ได้ตามลำดับ

โครงสร้างของโครงข่ายประสาทเทียมแบบวนซ้ำมีหลากหลายรูปแบบ ซึ่งรูปแบบมาตรฐานคือ Multi-Layer Perceptron (MLP) โดยจะมีส่วนข้อมูลขาเข้า (Input) ชั้นข้อมูลซ่อน (Hidden Unit) และส่วนข้อมูลขาออก (Output) ทั้งนี้สามารถเพิ่มจำนวนของ MLP และสามารถนำมาใช้ประโยชน์ในด้านหน่วยความจำ โดยโครงสร้างของโครงข่ายประสาทเทียมแบบวนซ้ำจะนำรูปแบบของ MLP มาเรียงต่อกัน มีเซลล์ประสาทเชื่อมต่อกันหลายๆเซลล์ นำข้อมูลขาเข้าในชั้นปัจจุบัน และข้อมูล

จากชั้นข้อมูลซ่อนก่อนหน้า มาเป็นข้อมูลขาเข้าในชั้นปัจจุบัน เพื่อประมวลผลก่อนให้เป็นข้อมูลขาออก ดังรูปที่ 2.4

ความแตกต่างที่พิเศษขึ้นมาจาก Neural Network ธรรมดาก็คือ ภายในชั้นข้อมูลซ่อนนั้นจะมีการประมวลผลซ้ำหลายๆครั้ง ดังนั้น RNNs จะสามารถใช้เป็นหน่วยความจำภายใน เพื่อนำไปประมวลผลตามลำดับของข้อมูลขาเข้า



รูป 2.4 โครงสร้างของโครงข่ายประสาทเทียมแบบบนซ้ำ

2.5 Recommendation System กับ ร้านอาหาร

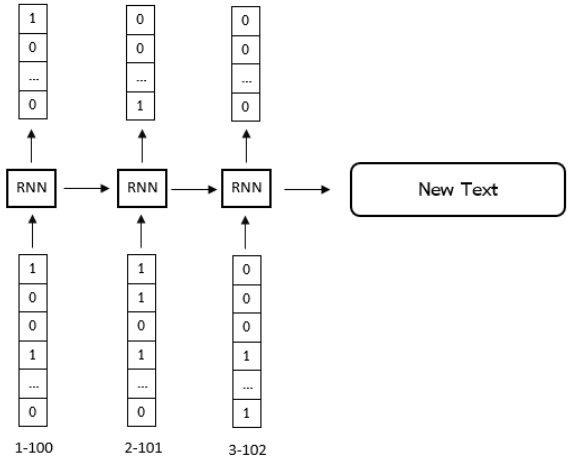
2.5.1 A Recurrent Neural Network Based Recommendation System

David Zhan Liu and Gurbir Singh [2] เป็นงานวิจัยเกี่ยวกับการแนะนำร้านอาหารให้กับผู้ใช้เป้าหมาย โดยงานวิจัยจะนำข้อมูลวิจารณ์ร้านอาหารหนึ่งร้านมาให้โครงข่ายประสาทเทียมแบบวนซ้ำทำการเรียนรู้ และปรับค่า ก่อนที่จะให้ค่า threshold ออกมาเพื่อใช้เป็นค่าเปรียบเทียบในการบ่งบอกความชอบของผู้ใช้ในรอบถัดไป หลังจากนั้นจะทำการค้นหาเพื่อนบ้านของผู้ใช้เป้าหมาย เมื่อได้เพื่อนบ้านของผู้ใช้มาแล้วจะนำข้อมูลวิจารณ์ของเพื่อนบ้านไปประมวลผลอีกครั้งในโมเดลเดิมซึ่งการประมวลผลในครั้งนี้ จะใช้ข้อมูลวิจารณ์ของเพื่อนบ้านเป็นข้อมูลขาเข้า เมื่อทำการประมวลผลในโมเดลเสร็จเรียบร้อย จะได้ค่า threshold ออกมาอีกหนึ่งค่า เพื่อนำไปเทียบกับค่า threshold ในรอบแรก ถ้าค่า threshold ในรอบหลังมีค่ามากกว่าหรือเท่ากับค่า threshold ในรอบแรกแสดงว่าเพื่อนบ้านพึงพอใจร้านอาหารนี้ และผู้ใช้เป้าหมายน่าจะพึงพอใจร้านอาหารนี้เช่นกัน แต่ถ้าค่า threshold ในรอบหลังมีค่าน้อยกว่าค่า threshold ในรอบแรกแสดงว่าเพื่อนบ้านไม่พึงพอใจร้านอาหารนี้ และผู้ใช้เป้าหมายน่าจะไม่พึงพอใจร้านอาหารนี้เช่นกัน

งานวิจัยนี้ได้ทำการพิจารณาข้อมูลทีละร้านเพื่อหาว่าผู้ใช้เป้าหมายจะพึงพอใจหรือไม่พึงพอใจร้านอาหารดังกล่าว โดยนำข้อมูลวิจารณ์ (reviews) เป็นข้อมูลขาเข้าในการสร้างโมเดล และนำข้อมูลวิจารณ์ของเพื่อนบ้านมาอ้างอิงความพึงพอใจให้กับผู้ใช้เป้าหมาย ดังนั้น งานวิจัยนี้จึงจัดเป็นงานที่เกี่ยวข้องกับการกรองร่วมโดยใช้โครงข่ายประสาทเทียมแบบวนซ้ำ

2.5.2 Implementing a Recurrent Neural Network for text generation using Keras

Sagar Jaiswal [7] เป็นงานวิจัยที่เกี่ยวกับการสร้างข้อความใหม่จากการเรียนรู้ข้อมูลในบทประพันธ์ของเชคสเปียร์ ในงานนี้จะนำข้อความที่อยู่ในบทประพันธ์ของเชคสเปียร์มาทำการเรียนรู้โดยแปลงข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปเรียนรู้กับโครงข่ายประสาทเทียมแบบวนซ้ำได้ จากนั้นจะสร้างพจนานุกรมของแต่ละอักขระโดยแทนด้วยตัวเลข เช่น a แทนด้วย 1, b แทนด้วย 2 เป็นต้น ต่อมาจะจัดเก็บข้อมูลในรูปข้อมูลขาเข้าและข้อมูลขาออก ซึ่งข้อมูลขาเข้าจะถูกในตัวแปร x ในรูปแบบของ array ที่มีขนาด 1x100 และข้อมูลขาออกจะถูกเก็บในตัวแปร y ในรูปแบบของ array ที่มีขนาด 1xจำนวนข้อมูลในพจนานุกรม ข้อมูลขาเข้าจะรับอักขระ 100 ตัวแรกตามลำดับเวลา เมื่อทำการวนซ้ำจะใช้ 100 อักขระถัดไป ส่วนข้อมูลขาออกจะเก็บเพียง 1 อักขระถัดไปจากข้อมูลขาเข้า เช่น เมื่อข้อมูลขาเข้าเป็นอักขระที่ 1-100 ข้อมูลขาออกจะเป็นอักขระตัวที่ 101 เป็นต้น ซึ่งอักขระทั้งหมดจะถูกจัดเก็บในรูปของตัวเลขจำนวนเต็มตามพจนานุกรมที่สอดคล้องกับอักขระนั้น ลำดับต่อไปจะเตรียมข้อมูลนำเข้าเพื่อนำไปใช้ในการเรียนรู้ ซึ่งมีการกำหนดรูปแบบข้อมูลขาเข้าเป็นอาเรย์ของอักขระตามลำดับเวลา และข้อมูลขาออกจะถูกแปลงเป็น one-hot vector ในรูปแบบ 0 และ 1 โดยที่ 1 จะเกิดขึ้นในตำแหน่งที่มีอักขระเกิดขึ้นจริงเท่านั้น ตัวอย่างเช่น เมื่อเรามี “recommend to eat” เป็นข้อมูลขาเข้า และสมมติให้ประมวลผลทีละ 5 อักขระ จะได้ว่า x คือ [18,5,3,15,13] จะได้ว่า เป็นข้อมูลขาออก ดังนั้น y คือ [0,0,...,1,...,0,0] โดยตำแหน่งที่เป็น 1 คือตำแหน่งของอักขระ “m” ที่ตรงกับพจนานุกรม เป็นต้น และนำไปสร้างรูปแบบการเรียนรู้ในโครงข่ายประสาทเทียมแบบวนซ้ำ เมื่อสร้างรูปแบบการเรียนรู้เรียบร้อยแล้วจะทำการสร้างข้อความใหม่โดยเริ่มจากการสุ่มอักขระทำให้ได้ประโยคในรูปแบบของจำนวนเต็ม โดยกำหนดความยาวของอักขระไว้ที่ 500 อักขระ จากนั้นจะนำไปเรียนรู้ในรูปแบบที่สร้าง เมื่อได้รับผลการทำนายออกมาแล้วจะทำการแปลงจำนวนเต็มกลับไปเป็นตัวอักษร ดังแสดงในรูปที่ 2.5



รูป 2.5 โครงสร้างการทำงานในการสร้างข้อความใหม่

งานนี้ได้ทำการพิจารณาข้อความที่อยู่ในบทประพันธ์ของเชคสเปียร์ โดยใช้อักขระทีละ 100 อักขระเป็นข้อมูลขาเข้า และนำอักขระดังกล่าวมาทำให้เป็นลักษณะเฉพาะตามลำดับเวลา โดยให้ข้อมูลขาออกเป็นอักขระถัดไปของข้อมูลขาเข้า ก่อนนำไปทำการสร้างรูปแบบของการเรียนรู้ โดยทำนายการเรียนรู้จากรูปแบบอักขระที่เคยเกิดขึ้นในอดีต ดังนั้น งานวิจัยนี้จึงจัดเป็นงานที่เกี่ยวข้องกับการกรองแบบอิงเนื้อหาโดยใช้โครงข่ายประสาทเทียมแบบวนซ้ำ

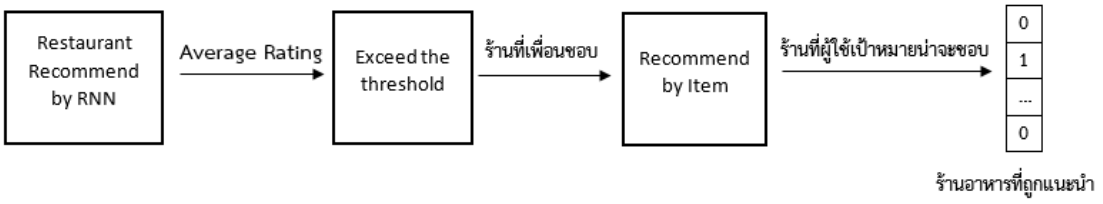
ในปัจจุบัน ระบบแนะนำมีหลายโดเมน เช่น ภาพยนตร์ เพลง โรงแรม ร้านอาหาร เป็นต้น โดเมนร้านอาหารถูกนำมาใช้ในโครงข่ายประสาทเทียมแบบวนซ้ำ ซึ่งอาจอยู่ในรูปแบบของการกรองร่วม (CF) หรือ การกรองแบบอิงเนื้อหา (CBF) ซึ่งรูปแบบการกรองร่วมจะสามารถแนะนำวัตถุให้กับผู้ใช้เป้าหมาย (Target User) โดยมีการอ้างอิงจากเพื่อนบ้านที่มีความพึงพอใจคล้ายคลึงกัน แต่มีข้อเสียคือข้อมูลขาเข้าต้องมีจำนวนมากเพียงพอที่จะนำมาเป็นฐานของข้อมูลเพื่อใช้ในการแนะนำ และการกรองแบบอิงเนื้อหา เป็นการแนะนำวัตถุที่มีความคล้ายคลึงกันกับประวัติการใช้งานวัตถุของผู้ใช้เป้าหมาย แต่มีข้อเสียคือข้อมูลขาออกของการแนะนำดังกล่าวจะมีแต่ข้อมูลที่มีความคล้ายคลึงกันกับความพึงพอใจในอดีตของผู้ใช้ทำให้ข้อมูลขาออกจะไม่มีความหลากหลาย จึงไม่เกิดการแนะนำวัตถุแบบใหม่ให้กับผู้ใช้เป้าหมาย

จากที่ได้กล่าวมาข้างต้นนั้น จะสังเกตได้ว่าที่ผ่านมามีผู้วิจัยได้เสนอการแนะนำร้านอาหารด้วย Collaborative Filtering (CF) หรือ Content-Based Filtering (CBF) โดยมีข้อมูลขาเข้าเป็น Review หรือ Item Embedding เพียงแบบใดแบบหนึ่งเท่านั้น และยังไม่มีผู้วิจัยคนใดนำสองเทคนิคนี้มาทำงานร่วมกันในรูปแบบของการแนะนำร้านอาหาร

บทที่ 3
วิธีการวิจัย

จากความสามารถในการวิเคราะห์ข้อมูลที่เปลี่ยนแปลงตามลำดับเวลาของ Recurrent Neural Network จึงมีผู้วิจัยที่นำ Recurrent Neural Network ไปใช้เป็นเครื่องมือวิเคราะห์ข้อมูลในงาน Recommendation เป็นจำนวนมาก แต่จะพบว่างานวิจัยที่ผ่านมาจะใช้ Recurrent Neural Network ได้นำมาใช้ร่วมกับ Collaborative Filtering โดยมีข้อมูลขาเข้าเป็นข้อมูลวิจารณ์ (Reviews) หรือใช้ร่วมกับ Content-Based Filtering โดยมีข้อมูลขาเข้าเป็นวัตถุ (Items) เพียงรูปแบบใดรูปแบบหนึ่งเท่านั้น แต่เมื่อได้นำข้อดีของทั้งสองรูปแบบมาพิจารณาแล้วพบว่า Content-Based Filtering มีข้อดีคือไม่ต้องการข้อมูลเป็นจำนวนมาก และสามารถแนะนำได้อย่างรวดเร็ว ส่วน Collaborative Filtering สามารถแนะนำวัตถุได้หลากหลายหมวดมากกว่า เพื่อทำให้ประสิทธิภาพในการแนะนำนั้นดีขึ้นจากการที่นำข้อดีของทั้งสองวิธีมาปกปิดข้อด้อยของกันและกัน และช่วยประสานการทำงานให้มีประสิทธิภาพมากยิ่งขึ้น

โครงการวิจัยนี้จึงเสนอขั้นตอนวิจัยที่สามารถแนะนำร้านอาหารที่ผู้ใช้น่าจะสนใจโดยทำการแนะนำจากข้อมูลวิจารณ์ร้านอาหารและคะแนนความพึงพอใจของเพื่อนบ้านที่มีต่อร้านอาหารผ่านกระบวนการของโครงข่ายประสาทเทียมแบบวนซ้ำ มารวมกับการแนะนำวัตถุ โดยมีวัตถุประสงค์ คือ สามารถรวมขั้นตอนวิจัยใหม่ที่แนะนำร้านอาหารให้กับผู้ใช้โดยใช้กระบวนการของโครงข่ายประสาทเทียมแบบวนซ้ำ โดยฝั่ง CF จะมีข้อมูลขาเข้าเป็นข้อมูลวิจารณ์ (reviews) และฝั่ง CBF จะมีข้อมูลขาเข้าเป็นลักษณะเฉพาะของร้านอาหาร (Item characteristic) สำหรับขั้นตอนการวิจัยจะเริ่มจาก 1) การเตรียมข้อมูล 2) โครงข่ายประสาทเทียมแบบวนซ้ำกับข้อมูลวิจารณ์ร้านอาหารของเพื่อนบ้าน (CF) 3) โครงข่ายประสาทเทียมแบบวนซ้ำกับลักษณะเฉพาะของวัตถุ (CBF) 4) รวบรวมผลลัพธ์จากสองขั้นตอน ดังแสดงในรูปที่ 3



รูป 3.1 ขั้นตอนการทำงานของการวิจัย

3.1 การเตรียมข้อมูล

ข้อมูลที่ใช้ในงานวิจัย คือข้อมูลของผู้ใช้ ทั้งหมด 156 คน ข้อมูลวิจารณ์ของผู้ใช้ทั้งหมด 9,360 ข้อมูล และร้านอาหารจำนวน 7,192 ร้าน ดังแสดงในรูปที่ 3.1.1 จะนำข้อมูล date ไปใช้ในการเรียงตามลำดับเวลา text และ star ใช้ในการประมวลผลข้อมูลฝั่ง CF รูปที่ 3.1.2 นำ attributes ไปจัดให้อยู่ในรูปที่ข้อมูลขาเข้าที่ประกอบด้วย 0 และ 1 business_id และ name ใช้ในการประมวลผลข้อมูลฝั่ง CBF และ 3.1.3 friend ใช้ในการค้นหาเพื่อนบ้านของ user และ user_id ใช้เป็น key หลักในการ mapping ข้อมูล จากนั้นจะนำข้อมูลมารวมเป็นส่วนประกอบเดียวกัน โดยเชื่อมกันด้วย business_id และ user_id ในลำดับต่อไปจะทำการจัดรูปข้อมูลให้อยู่ในรูปข้อมูลขาเข้าที่นำไปประมวลผลได้ โดยขั้นแรก เราจะนำข้อมูลวิจารณ์มาทำการลบคำที่ไม่มีประโยชน์ต่อการประมวลผลออก ก่อนจะนำไปเป็นข้อมูลขาเข้าของฝั่ง Collaborative Filtering ในส่วนถัดไปจะจัดรูปคุณลักษณะของร้านต่างๆให้อยู่รูป 0 และ 1 เพื่อนำไปเป็นข้อมูลขาเข้าของฝั่ง Content-Based Filtering และทำการประมวลผลต่อไป

	business_id	cool	date	funny	review_id	stars	text	useful	user_id
0	iCQpiavjjPzJ5_3gPD5Ebg	0	2011-02-25	0	x7mDIiDB3jEiPGPHOmDzyw	2	The pizza was okay. Not the best I've had. I p...	0	msQe1u7Z_XuqjGoqhB0J5g
1	pomGBqfbxcqPv14c3XH-ZQ	0	2012-11-13	0	dDl8zu1vWPdKGihJrwQbpw	5	I love this place! My fiance And I go here atl...	0	msQe1u7Z_XuqjGoqhB0J5g
2	jtQARsP6P-LbkyjbO1qNGg	1	2014-10-23	1	LZp4UX5zK3e-c5ZGSeo3kA	1	Terrible. Dry corn bread. Rib tips were all fa...	3	msQe1u7Z_XuqjGoqhB0J5g
3	elqbBhBfElMNSrjFqW3now	0	2011-02-25	0	Er4NBWCmCD4nM8_p1GRdow	2	Back in 2005-2007 this place was my FAVORITE t...	2	msQe1u7Z_XuqjGoqhB0J5g
4	Ums3gaP2qM3W1XcA5r6SsQ	0	2014-09-05	0	jsDu6QEJHbwP2Blom1PLCA	5	Delicious healthy food. The steak is amazing. ...	0	msQe1u7Z_XuqjGoqhB0J5g

รูป 3.1.1 ตัวอย่างของข้อมูลวิจารณ์

	attributes	business_id	categories	is_open	name	neighborhood	review_count	stars
0	{'BikeParking': 'False', 'BusinessAcceptsCredi...	Apn5Q_b6Nz61Tq4XzPdf9A	Tours, Breweries, Pizza, Restaurants, Food, Ho...	1	Minhas Micro Brewery		24	4.0
1	{'Alcohol': 'none', 'BikeParking': 'False', 'B...	AjEbIBw6ZFfln7ePHha9PA	Chicken Wings, Burgers, Caterers, Street Vendo...	0	CK'S BBQ & Catering		3	4.5
2	{'Alcohol': 'beer_and_wine', 'Ambience': '{'ro...	O8S5hYJ1SMc8fA4QBtVujA	Breakfast & Brunch, Restaurants, French, Sandw...	0	La Bastringue	Rosemont-La Petite-Patrie	5	4.0
3	{'BikeParking': 'True', 'BusinessAcceptsCredit...	45bWSZtniwPRiqlivpS8Og	Coffee & Tea, Food	1	The Coffee Bean & Tea Leaf		63	4.0
4	{'BusinessParking': '{'garage': False, 'street...	9A2quhZLyWk0akUetBd8hQ	Food, Bakeries	0	Bnc Cake House	Koreatown	7	4.0
5	{'Alcohol': 'none', 'BusinessAcceptsCreditCard...	6OuOZAok8ikONMS_T3EzXg	Restaurants, Thai	1	Thai One On	Ridgewood	7	2.0

รูป 3.1.2 ตัวอย่างของข้อมูลของร้านอาหาร

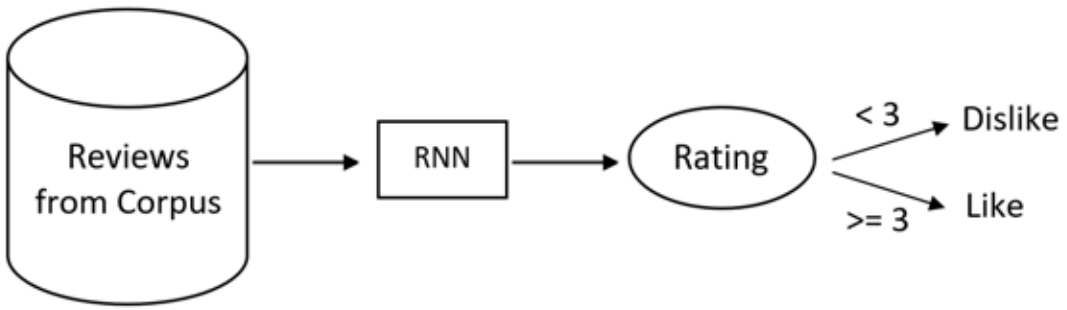
average_stars	friends	name	review_count	user_id
3.89	YwaKGmRNnSa3R3N4Hf9jLw, NDjDKV7IWeqS_kvdINS9zw...	Yolie	8	B3xwUPXMlCb7_5IEDlq4Uw
4.25	PhZlPdJsoaEX2SWKPnfTvw	Anthony	8	Pc5vc0RuoCa8YU6hD-y3Ow
1.00	RVyNRipJXhEQko_eztWYyg, SaGyqyKgE06i7Kzi3rF5gA...	Tisharra	1	dz1RKDdHziTy0uw7VtXeXQ
5.00	zudO7GgOaNbqlGDUhNQVgQ, gDy6k3AqlkcN4cElkWnx_A...	Colleen	2	6OmHL9dwt5RmKwhGFgjvAA
4.50	gMxSAPGVR-ubiG9wMUYSKg, WHy102b0fS_UTwmYFoNBHQ...	Aaron	5	tGbVIKlZAPGrt3VjTn7gRA
4.67	KtuTspkRE2ienARgs89Ndw, MOmoao5si-N3nCjVef3vxA...	James	15	0EocYOZhpDTYqqO_uYbw5A
1.50	1xgrHL7Ah9uEvzrtLCTYIg, T2Gunr0D50EEgqguEk9ssw...	Bonnie	2	W_V8_Id9tpOcZAG1hxcLxg
5.00	0AiZbmQUMyoo4cGbX-8iUA, 81_lyihbg-M2QpmJ9g3WAQ...	Amy	2	1x_2iGnd7j5jcB7VjmFZvQ
3.00	mAYv0wxFK6BJej5zAVaSTg, ghqycqgHlUw_N0aS5vRtMQ...	Ash	2	6Gu9vLSSuo9N5o3xRQcO5g
3.97	-VmuNoYI6-Wl0kLrUZbLSg, W-qd_DCuv5QhfvA4JpjYcg...	Bryan	148	ucj0RT3ZsWHFXQssLq-wmQ
5.00	DN9FYA-ppH3WRYpw4KV5ow, 9Rd4CT7psFf9nbU5uAmOTA...	Amanda	1	_a7BES74d_thQfpNBuk8Gw

รูป 3.1.3 ตัวอย่างของข้อมูลของผู้ใช้

3.2 โครงข่ายประสาทเทียมแบบวนซ้ำกับข้อมูลวิจารณ์ร้านอาหารของเพื่อนบ้าน (CF)

3.2.1 การนำข้อมูลวิจารณ์ของร้านอาหารและความพึงพอใจทั้งหมดมาสร้างโมเดล

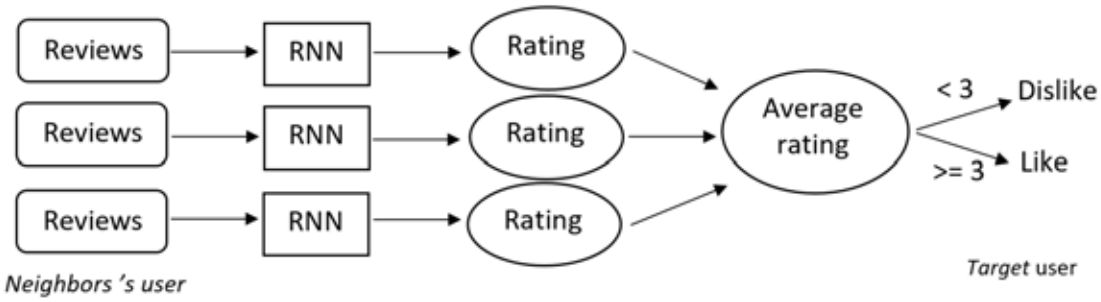
ในขั้นแรกจะนำข้อมูลวิจารณ์ทั้งหมดมาเป็นข้อมูลขาเข้า และให้ความพึงพอใจของแต่ละข้อมูลวิจารณ์เป็นข้อมูลขาออกเพื่อทำการเรียนรู้ ผ่านการทำงานของโครงข่ายประสาทเทียมแบบวนซ้ำ ยกตัวอย่างเช่น จากรูปที่ 3.2.1 นำข้อมูลวิจารณ์ทั้งหมดในคลังข้อมูลมาทำการเรียนรู้ ผ่านโครงข่ายประสาทเทียมแบบวนซ้ำ ซึ่งเราสามารถทำนายคะแนนความพึงพอใจของข้อมูลวิจารณ์ของร้านอาหารจากโมเดลนี้ได้ โดยเราจะกำหนดให้ค่า threshold คือ 3 หากได้รับคะแนนความพึงพอใจมากกว่าหรือเท่ากับ 3 หมายถึง พึงพอใจ แต่หากคะแนนความพึงพอใจน้อยกว่า 3 หมายถึง ไม่พึงพอใจ ดังแสดงในรูปที่ 3.2.1 เพื่อทำการสร้างโมเดลที่สามารถตรวจสอบการแสดงระดับความพึงพอใจของข้อมูลวิจารณ์



รูป 3.2.1 รูปแบบการทำงานของข้อมูลวิจารณ์ของร้านอาหาร

3.2.2 การทำนายความพึงพอใจของผู้ใช้เป้าหมายโดยได้นำคะแนนความพึงพอใจของเพื่อนบ้านมาอ้างอิง

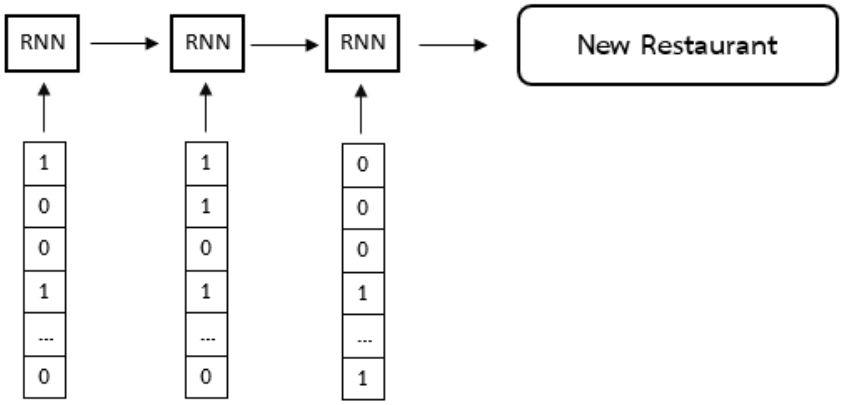
ในขั้นตอนนี้จะนำหลักเกณฑ์ของ CF มาใช้ในการค้นหาเพื่อนบ้าน ขั้นถัดมานำผู้ใช้ไปค้นหาเพื่อนบ้านโดยจะทำการสุ่มเพื่อนบ้านมาจำนวนหนึ่งจากจำนวนเพื่อนบ้านทั้งหมดที่ผู้ใช้เป้าหมายมี เพื่อนำข้อมูลวิจารณ์มาทำการดำเนินการในโครงข่ายประสาทเทียมแบบวนซ้ำที่ได้ทำการเรียนรู้มาแล้วในขั้นตอนแรก หลังจากดำเนินการแล้วจะได้รับคะแนนความพึงพอใจของเพื่อนบ้านแต่ละคน จากนั้นนำไปหาค่าเฉลี่ยของคะแนนความพึงพอใจ ก่อนทำการเปรียบเทียบเงื่อนไข ถ้าคะแนนความพึงพอใจของเพื่อนบ้านมีค่าน้อยกว่า 3 แสดงว่า ผู้ใช้เป้าหมายไม่น่าจะชื่นชอบร้านที่ถูกวิจารณ์โดยเพื่อนบ้านดังกล่าว เราจะทำการค้นหาเพื่อนบ้านกลุ่มใหม่จากเพื่อนบ้านทั้งหมดที่ผู้ใช้เป้าหมายมีและทำการเรียนรู้ซ้ำ แต่ถ้าคะแนนความพึงพอใจของเพื่อนบ้านมีค่ามากกว่าหรือเท่ากับ 3 แสดงว่าผู้ใช้เป้าหมายน่าจะชื่นชอบร้านที่ถูกรีวิวโดยเพื่อนบ้านดังกล่าว ดังแสดงในรูป 3.2.2 จึงจะส่งข้อมูลนี้ไปเป็นข้อมูลขาเข้าของส่วนถัดไป ซึ่งได้นำหลักเกณฑ์ของ CBF มาทำนายร้านอาหารถัดไปจากลักษณะที่คล้ายคลึงกัน



รูป 3.2.2 การทำนายความพึงพอใจของผู้ใช้เป้าหมายโดยอิงจากเพื่อนบ้าน

3.3 โครงข่ายประสาทเทียมแบบวนซ้ำกับลักษณะเฉพาะของวัตถุ (CBF)

ในขั้นนี้จะนำข้อมูลแบบหลายเกณฑ์คุณลักษณะของร้านอาหารที่ประกอบไปด้วย ที่จอดรถ ระดับเสียง การจ่ายเงินด้วยบัตรเครดิต เป็นต้น โดยที่ผู้ใช้แต่ละรายได้ทำการวิจารณ์ตามลำดับเวลา โดยให้ข้อมูลขาเข้าเป็นลักษณะเฉพาะของร้านอาหาร (Item Characteristic) 5 ร้านแรก ที่อยู่ในลักษณะของ array ประกอบด้วยข้อมูลที่มีค่า 0 หรือ 1 และมีข้อมูลขาออกเป็นร้านอาหารถัดไปจาก 5 ร้านนั้น จากนั้นจะทำการแปลงข้อมูลขาออกของร้านอาหารให้อยู่ในรูปแบบของ one-hot vector ในลำดับต่อมาทำการสร้างโมเดลการเรียนรู้ด้วยโครงข่ายประสาทเทียมแบบวนซ้ำ และทำนายร้านอาหารถัดไปให้กับผู้ใช้ ดังแสดงในรูปที่ 3.3



รูปที่ 3.3 โครงข่ายประสาทเทียมแบบวนซ้ำกับลักษณะเฉพาะของวัตถุ

3.4 รวบรวมผลลัพธ์จากสองขั้นตอน

ในขั้นสุดท้ายจะนำข้อมูลวิจารณ์ของเพื่อนบ้านที่ได้จากการประมวลผลจากข้อมูลขาออกฝั่ง CF ที่มีคะแนนความพึงพอใจเฉลี่ยมากกว่าหรือเท่ากับ 3 มาเป็นข้อมูลขาเข้าของฝั่ง CBF โดยจะนำลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่ได้จากในข้อมูลวิจารณ์ของฝั่ง CF มาทำการดำเนินการในโมเดลของฝั่ง CBF เพื่อแนะนำร้านถัดไปที่ผู้ใช้เป้าหมายน่าจะชอบและไปใช้บริการ

บทที่ 4
ผลการวิจัย

ในบทนี้จะแสดงผลการทดสอบประสิทธิภาพของระบบ โดยทำการเทียบผลการทดสอบระหว่าง การใช้ RNN ร่วมกับ CF และ CBF ที่มี ข้อมูลวิจารณ์และลักษณะข้อมูลเฉพาะแบบหลายเกณฑ์เป็นข้อมูลขาเข้า กับ การใช้ RNN ร่วมกับ CBF ที่มีลักษณะข้อมูลเฉพาะแบบหลายเกณฑ์เป็นข้อมูลขาเข้า เพียงอย่างเดียว

4.1 Data set

ข้อมูลที่ใช้ในงานวิจัย คือ ข้อมูลจาก Yelp Dataset มีข้อมูลของผู้ใช้ทั้งหมด 156 คน ,ข้อมูลร้านอาหารทั้งหมด 7192 ร้าน และข้อมูลวิจารณ์ร้านอาหาร 9360 ข้อมูล โดยแบ่งเป็น training set ในส่วน CF ใช้ข้อมูลวิจารณ์ ทั้งหมด 9360 ข้อมูลและในส่วน CBF ใช้ร้านอาหาร 7192 ร้าน และเป็น test set ในส่วน CF ใช้ ข้อมูลวิจารณ์ 300 ข้อมูลต่อผู้ใช้ 1 คน โดยมีผู้ใช้ทั้งหมด 156 คน จึงนำไปสู่การใช้ร้านอาหาร 300 ร้านต่อผู้ใช้ 1 คน โดยมีผู้ใช้ทั้งหมด 156 คน สำหรับส่วน CBF ในลำดับถัดมา สำหรับรายละเอียดเพิ่มเติมสามารถดูได้ในตารางที่ 4.1

ตารางที่ 4.1 แสดงค่าสถิติต่าง ๆที่ใช้ในการทดลอง

จำนวนผู้ใช้	156
จำนวนข้อมูลวิจารณ์	9360
จำนวนคะแนนความพึงพอใจ	9360
จำนวนร้านอาหาร	7192

สำหรับข้อมูลแบบหลายเกณฑ์ ซึ่งเป็นลักษณะของข้อมูลร้านอาหาร ประกอบด้วยข้อมูล 14 ชนิด เช่น ที่จอดจักรยาน, ระดับความดังของเสียง, ไวไฟ, การบริการจัดส่ง, ที่นั่งภายนอกร้าน เป็นต้น

4.2 วิธีที่ใช้ในการประเมิน

การเปรียบเทียบประสิทธิภาพของผลการทดลองระหว่าง 2 วิธี คือ 1.) วิธีการใช้ RNN ร่วมกับ CF และ CBF 2.) วิธีการใช้ RNN ร่วมกับ CBF เพียงอย่างเดียว โดยใช้ Cosine Similarity เป็นวิธีในการประเมิน

4.2.1 Cosine Similarity

Cosine Similarity เป็นวิธีการเปรียบเทียบความคล้ายคลึงกันของเวกเตอร์สองเวกเตอร์โดยใช้วิธีการ Dot product เพื่อหาสัมประสิทธิ์ความคล้ายคลึงกัน (Similarity Coefficient) ดังสมการที่ 5 [8]

$$\cos(u, v) = \frac{\overrightarrow{I_u} \cdot \overrightarrow{I_v}}{|\overrightarrow{I_u}| \cdot |\overrightarrow{I_v}|} \quad (5)$$

จากสมการที่ 5 ให้ $\overrightarrow{I_u}$, $\overrightarrow{I_v}$ คือ เวกเตอร์ของลักษณะวัตถุเฉพาะแบบหลายเกณฑ์ของข้อมูลขาเข้า u และข้อมูลขาออก v

จากวิธีข้างต้นพบว่าการวัดความคล้ายคลึงกันสามารถวัดได้สองรูปแบบ เพื่อใช้วัดความคล้ายคลึงระหว่างคะแนนความพึงพอใจที่ทำนายได้กับคะแนนความพึงพอใจที่ผู้ใช้เคยให้ และวัดความคล้ายคลึงระหว่างร้านอาหารที่ทำนายได้กับร้านอาหารที่ผู้ใช้เคยรับประทาน

จากวิธีข้างต้นพบว่าสามารถหาค่าสัมประสิทธิ์ความคล้ายคลึงกันได้สองรูปแบบ เพื่อที่จะวัดความคล้ายคลึงระหว่างคะแนนความพึงพอใจที่ทำนายได้กับคะแนนความพึงพอใจที่ผู้ใช้เคยให้ ดังสมการที่ 6 และวัดความคล้ายคลึงระหว่างร้านอาหารที่ทำนายได้กับร้านอาหารที่ผู้ใช้เคยรับประทาน ดังสมการที่ 7

$$Sim_1 = \frac{\sum_{i=1}^{n}(\cos(u_i,v)\times r_{u_i})}{\sum_{i=1}^{n} r_{u_i}} \quad (6)$$

$$Sim_2 = \frac{\sum_{i=1}^{n}(\cos(u_i,v)\times r_{u_i})}{\sum_{i=1}^{n} \cos(u_i,v)} \quad (7)$$

การคำนวณ Sim_1

เพื่อหาค่าความคล้ายคลึงกันของคะแนนความพึงพอใจที่ผู้ใช้เคยให้กับคะแนนความพึงพอใจที่ทำนายได้ แบ่งเป็น 2 ส่วน คือ

1. ในส่วนที่มีการทำงานร่วมกันของ CF และ CBF จะนำลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่เป็นข้อมูลขาออกของฝั่ง CBF ของผู้ใช้คนหนึ่งมาหาค่าความคล้ายคลึงระหว่างผลลัพธ์ของผู้ใช้ กับลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่ผู้ใช้ดังกล่าวได้ทำการเขียนข้อมูลวิจารณ์ จากนั้นจะนำไปคูณกับคะแนนความพึงพอใจที่ทำนายได้จากการทำ CF หารด้วยค่าเฉลี่ยของคะแนนความพึงพอใจทั้งหมด

2. ในส่วนที่มีการทำงานของ CBF เพียงอย่างเดียว จะนำลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่เป็นข้อมูลขาออกฝั่ง CBF ของผู้ใช้คนหนึ่งมาหาค่าความคล้ายคลึงระหว่างผลลัพธ์ของผู้ใช้ กับลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่ผู้ใช้ดังกล่าวได้ทำการเขียนข้อมูลวิจารณ์

จากนั้นจะนำไปคูณกับคะแนนความพึงพอใจที่ผู้ใช้ได้แสดงไว้ แล้วหารด้วยค่าเฉลี่ยของคะแนนความพึงพอใจทั้งหมด

การคำนวณ Sim_2

เพื่อหาค่าความคล้ายคลึงกันของร้านอาหารที่ผู้ใช้เคยไปรับประทานกับร้านอาหารที่ทำนายได้ แบ่งเป็น 2 ส่วน คือ

1. ในส่วนที่มีการทำงานร่วมกันของ CF และ CBF จะนำลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่เป็นข้อมูลขาออกของฝั่ง CBF ของผู้ใช้คนหนึ่งมาหาค่าความคล้ายคลึงระหว่างผลลัพธ์ของผู้ใช้ กับลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่ผู้ใช้ดังกล่าวได้ทำการเขียนข้อมูลวิจารณ์ จากนั้นจะนำไปคูณกับคะแนนความพึงพอใจที่ทำนายได้จากการทำ CF หารด้วยค่าเฉลี่ยของความคล้ายคลึงกันของผลลัพธ์

2. ในส่วนที่มีการทำงานของ CBF เพียงอย่างเดียว จะนำลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่เป็นข้อมูลขาออกฝั่ง CBF ของผู้ใช้คนหนึ่งมาหาค่าความคล้ายคลึงระหว่างผลลัพธ์ของผู้ใช้ กับลักษณะเฉพาะแบบหลายเกณฑ์ของร้านอาหารที่ผู้ใช้ดังกล่าวได้ทำการเขียนข้อมูลวิจารณ์ จากนั้นจะนำไปคูณกับคะแนนความพึงพอใจที่ผู้ใช้ได้แสดงไว้ แล้วหารด้วยค่าเฉลี่ยของความคล้ายคลึงกันของผลลัพธ์

4.3 ผลการทดลอง

ในหัวข้อนี้ผลลัพธ์ที่ได้จาก Restaurant Multicriteria Recommendation based on Recurrent Neural Network ซึ่งเป็นผลลัพธ์จากการใช้ RNN ร่วมกับการทำงาน CF และ CBF โดยมีข้อมูลขาเข้าคือ ข้อมูลวิจารณ์ในส่วนของ CF และ ลักษณะข้อมูลเฉพาะแบบหลายเกณฑ์ ในส่วนของ CBF โดยเปรียบเทียบกับ การใช้ RNN ร่วมกับการทำงานของ CBF โดยมีข้อมูลขาเข้าคือ ลักษณะข้อมูลเฉพาะแบบหลายเกณฑ์เพียงอย่างดียว

จากข้อมูลผู้ใช้ทั้งหมด 156 คน เราจะนำผลลัพธ์มาแสดงเบื้องต้นโดยมีผู้ใช้เพียง 10 คน แต่ละผลการทดลองมีลักษณะดังรูปที่ 4.3

user_id	CF+CBF	avg_sumrate	avg_sumsim
-zSnLWbNZtNoKnUpothzrg	0109 Dessert & Chocolate	0.748620809	3.739693903
-znBbADZqQ_30bWe6o6rKg	Lucille's Smokehouse Bar-B-Que	0.75275441	3.577733869
0R4juIjFh3TFiWDCanGY9w	Wah Fu Chinese Restaurant	0.738960985	3.910365095
0k0PQctiUZWdtdBdDSp2_w	Charlotte Coffee Station	0.687284851	3.654572318
0xiX3PVkh2WoqZb423AcjA	0109 Dessert & Chocolate	0.702898866	3.866883031
1phzsrzdfHi84bwxBCpu_A	0109 Dessert & Chocolate	0.736559539	3.615965583
1v4x10Twahc1xa2DDat3KQ	Raising Cane's	0.697366821	3.768732614
2LaXC_AW4I0EBU9FhzpOgg	0109 Dessert & Chocolate	0.725821005	3.64830699
2VhvnM9TxiowYKPXirmgvQ	Raising Cane's	0.719315479	3.670503855
3ItZMmWMLA-enY2yQY0tAA	Strip Liquor	0.696284709	3.930958772
	(CF+CBF): AVG	**0.720586747**	**3.738371603**

รูปที่ 4.3.1 แสดงผลลัพธ์ความคล้ายคลึงกันของคะแนนความพึงพอใจและร้านอาหารจากข้อมูลผู้ใช้ 10 คน ในฝั่ง CF และ CBF

user_id	CBF	avg_sumrate	avg_sumsim
-zSnLWbNZtNoKnUpothzrg	Borgo Trattoria	0.632557817	3.622009491
-znBbADZqQ_30bWe6o6rKg	Scarpetta	0.796455096	2.890698517
0R4juIjFh3TFiWDCanGY9w	Sami's Gyros	0.772941911	2.90968106
0k0PQctiUZWdtdBdDSp2_w	Valhalla Pub and Eatery	0.65964376	3.582431059
0xiX3PVkh2WoqZb423AcjA	Dutch Bros. Coffee	0.641316079	3.203650442
1phzsrzdfHi84bwxBCpu_A	The Peppermill Restaurant & Fi	0.684327487	3.951496766
1v4x10Twahc1xa2DDat3KQ	Scarpetta	0.707109326	3.911156736
2LaXC_AW4I0EBU9FhzpOgg	Beijing Noodle No. 9	0.696588726	3.653625883
2Vhvnm9TxiowYKPXirmgvQ	Taco Bell	0.614205247	3.805482893
3ItZMmWMLA-enY2yQY0tAA	424 Walnut	0.822747789	3.136243894
	(CBF): AVG	0.702789324	3.466647674

รูปที่ 4.3.2 แสดงผลลัพธ์ความคล้ายคลึงกันของคะแนนความพึงพอใจและร้านอาหารจากข้อมูลผู้ใช้ 10 คน ในฝั่ง CBF

เมื่อทำการทดสอบครบทั้ง 156 คน ให้ผลลัพธ์ดังนี้

1	user_id	CF+CBF	avg_sumrate	avg_sumsim
120	r0WMBsH20aquSDGKo5_oRA	U Food Grill	0.72836187	3.61024291
121	Rw1QE99aazERB_z3S8D8nA	Weeds Cafe	0.718231348	3.72187518
122	rZQCd47n7OwPd71igVX6Og	Bob's Diner	0.780854192	3.706788346
123	rzTHDO7RFgZ6gQbKuxmYnw	Einstein Bros Bagels	0.761179336	3.962437181
124	s-591-mtIyP7F1Lffw98jw	The Shout House	0.714259028	3.758665155
125	s-QuErK1oRlAzT5qnO6FOQ	Fleming's Prime Steakhouse & Wine E	0.750477216	3.766756726
126	s7XIFcwZnhEUMzSZ4EiAHA	Suzie B's	0.725641389	3.780945968
127	sRrl6zoyLYRTVlhdn25hhQ	KINKA IZAKAYA BLOOR	0.694195141	3.863850877
128	Su2kadHdmAVhM-_1TIlhSQ	Sharetea	0.762454256	3.752193656
129	SX7xBRFGfrEjfXDqw0d-ew	Juliette & Chocolat	0.709936123	3.786745404
130	TFxeEvpjMNQ3AWL49iMwtA	Sweet Addiction	0.723017189	3.925562271
131	tG24pgYOAyEONNhgmhVO-Q	BarChef	0.710091286	3.661436654
132	tmvYYqXiz3I235x-SSOeTw	Raising Cane's	0.784166991	3.680228413
133	uLWWpi62JJLnTh7UciJFHg	Spotted Donkey Cantina, el Pedregal	0.720978707	3.845929176
134	UNSUk0fGl4vhMziQT4qiaA	Juliette & Chocolat	0.741536979	3.913636969
135	uNxXIIFzjHm5r62UTyVjrQ	Subway	0.733411106	3.531413739
136	UPEa9tPhUAk18qTHal_d1g	Hash House A Go Go	0.747616333	3.770457877
137	UYp9lazOwsIIaLwVrDrAQw	Jac & Ray	0.731003296	3.726474291
138	Vd24R1OnKzw8WjHLrrDTtw	Jarrod's Coffee, Tea & Gallery	0.766382978	3.940238346
139	VXbi8ndacawnd6kidQF8dw	St. Louis Bar & Grill	0.741563058	3.813998707
140	W92S95QwdQERvx7lvxmeKg	Pizza Mart	0.745519788	3.54052909
141	W_pPF_bKVNtA8m7lRqdZdw	Stacy's Smokehouse	0.712452787	3.850624032
142	WgBoWF5e63SsqJlpbOxOzA	Pieology Pizzeria	0.723992586	3.841999974
143	WpZTKVU09--Fb2qFneqkZg	Trader Joe's	0.77126485	3.601204663
144	wyODhqTQFfBum-B98JLLHw	The Works	0.771926572	3.755485101
145	x_Men2agHTy6ao9YGFQv-A	Subway	0.723685347	3.633266442
146	XKkKEn_mo19vrM0vVIXiEQ	Roundabout Brewery	0.715796033	3.73405707
147	XMN0nIXjW_wEyKQZCIyIYQ	Pizza Mart	0.71406065	3.81771743
148	Xoqo8B_5bZ_JvmSXt644Uw	Geppetto Cafe	0.708866642	3.909954356
149	xqddl659Tkw6cpb4BXCx0g	Sugar Bowl	0.733679716	3.879423412
150	XWld6N9X2-WdvhvsKcaWwg	Roundabout Brewery	0.716724734	3.835975707
151	y-OW396z2nqGyyklOLnVjw	Chew & Chug	0.776892678	4.195703598
152	YdS7NVnEhK0gTypX2kMc9A	Bob's East Side Deli	0.7546256	3.883619643
153	Yuyy16nuHylNKZn9I9Vx-Q	The Farm At South Mountain	0.70530591	3.742514881
154	ZEjXrUKwBECS6MMB2hDwOA	Cold Stone Creamery	0.71318639	3.708558445
155	zh7sSdnM3CefK8RBRWahZQ	Suzie B's	0.748144113	3.660558198
156	zJH5RdZsRKOMtx3W0-Ba2Q	Berdena's	0.727948111	3.987074445
157	Zt045qHBdc5LbqH_V16ixw	Vons	0.708030291	3.725859207
158		AVG (CF+CBF)	0.730847279	3.73000887

รูปที่ 4.3.3 แสดงผลลัพธ์ความคล้ายคลึงกันของคะแนนความพึงพอใจและร้านอาหารจากข้อมูลผู้ใช้ 156 คน ในฝั่ง CF และ CBF

1	user_id	CBF	avg_sumrate	avg_sumsim
120	r0WMBsH20aquSDGKo5_oRA	Fragrant Bakery	0.605250383	3.461787687
121	Rw1QE99aazERB_z3S8D8nA	Robertos Taco Shop	0.776500649	3.624293392
122	rZQCd47n7OwPd71igVX6Og	Asian Picnic	0.723617106	3.558987382
123	rzTHDO7RFgZ6gQbKuxmYnw	Due Forni	0.68070136	3.660476963
124	s-591-mtIyP7F1Lffw98jw	Stan's Northfield Bakery	0.717794517	4.149753976
125	s-QuErK1oRlAzT5qnO6FOQ	Le Vieux Velo	0.628351002	3.651485266
126	s7XIFcwZnhEUMzSZ4EiAHA	Noodle Shop	0.714298485	3.238979853
127	sRrl6zoyLYRTVlhdn25hhQ	Olive Garden Italian Restaurant	0.303447798	4.077478922
128	Su2kadHdmAVhM-_1TIlhSQ	McGurkee's Italian Kitchen	0.812217489	3.746942744
129	SX7xBRFGfrEjfXDqw0d-ew	Pieology Pizzeria	0.727099001	3.449772097
130	TFxeEvpjMNQ3AWL49iMwtA	Sugar Marmalade	0.640031968	3.817262557
131	tG24pgYOAyEONNhgmhVO-Q	Zoe's	0.692915824	3.281707989
132	tmvYYqXiz3I235x-SSOeTw	Panera Bread	0.520264017	4.091054084
133	uLWWpi62JJLnTh7UciJFHg	ACE Hotel Pittsburgh	0.735256104	4.015457464
134	UNSUk0fGl4vhMziQT4qiaA	Salara Dessert Lounge & Bistro	0.716067601	3.400041164
135	uNxXIIFzjHm5r62UTyVjrQ	The Shout House	0.727216468	4.343921714
136	UPEa9tPhUAk18qTHal_d1g	Chili's	0.760591021	3.701392179
137	UYp9lazOwsIIaLwVrDrAQw	Friendship Perk And Brew	0.764777491	3.928091664
138	Vd24R1OnKzw8WjHLrrDTtw	Pieology Pizzeria	0.681051983	4.10224963
139	VXbi8ndacawnd6kidQF8dw	Sugar Marmalade	0.647884199	3.742500832
140	W92S95QwdQERvx7lvxmeKg	The Farm At South Mountain	0.749848684	3.877309361
141	W_pPF_bKVNtA8m7lRqdZdw	Chili's	0.787091395	4.509899798
142	WgBoWF5e63SsqJlpbOxOzA	BLAM!Wich	0.577645695	3.859151786
143	WpZTKVU09--Fb2qFneqkZg	Victor	0.779705888	3.428802168
144	wyODhqTQFfBum-B98JLLHw	Jeni's Splendid Ice Creams	0.696877557	3.864497649
145	x_Men2agHTy6ao9YGFQv-A	Seasons Buffet	0.781576689	4.10937919
146	XKkKEn_mo19vrM0vVIXiEQ	Stockyards Restaurant	0.778637629	3.204416404
147	XMN0nIXjW_wEyKQZCIyIYQ	Jocks & Jills Sports Grille	0.751043535	4.210209707
148	Xoqo8B_5bZ_JvmSXt644Uw	Woodlot Restaurant	0.702076789	3.497829675
149	xqddl659Tkw6cpb4BXCx0g	Cotto Italian Comfort Food	0.716866179	3.036584598
150	XWld6N9X2-WdvhvsKcaWwg	La Casa Blanca	0.690490705	3.321572705
151	y-OW396z2nqGyyklOLnVjw	Cocina Mendoza	0.7308568	3.692447108
152	YdS7NVnEhK0gTypX2kMc9A	The Macaron Tea Room	0.726287386	3.811305469
153	Yuyy16nuHylNKZn9I9Vx-Q	Caffe Artigiano	0.724540121	3.971139735
154	ZEjXrUKwBECS6MMB2hDwOA	The Farm At South Mountain	0.60759427	3.378277556
155	zh7sSdnM3CefK8RBRWahZQ	Japaneiro	0.740283902	3.975915611
156	zJH5RdZsRKOMtx3W0-Ba2Q	Olive Garden Italian Restaurant	0.586178391	3.696422958
157	Zt045qHBdc5LbqH_V16ixw	Cold Stone Creamery	0.440945136	3.768214654
158		**AVG (CBF)**	**0.712835694**	**3.704685725**

รูปที่ 4.3.4 แสดงผลลัพธ์ความคล้ายคลึงกันของคะแนนความพึงพอใจและร้านอาหารจากข้อมูลผู้ใช้ 156 คน ในฝั่ง CBF

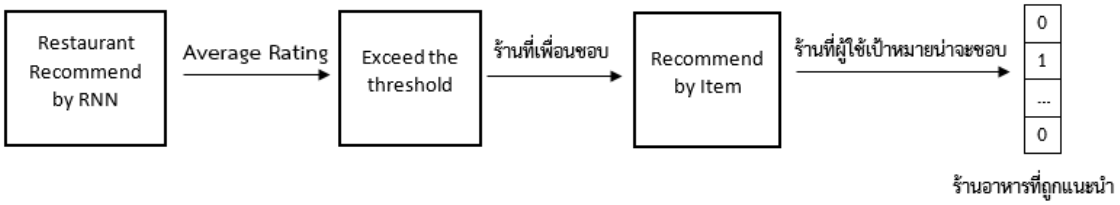
จะเห็นได้ว่า การทำงานร่วมกันระหว่าง CF และ CBF มีค่าสัมประสิทธิ์ความคล้ายคลึงกันของคะแนนความพึงพอใจที่ทำนายได้กับคะแนนความพึงพอใจที่ผู้ใช้เคยให้ประมาณ 0.73 และค่าสัมประสิทธิ์ความคล้ายคลึงกันของร้านอาหารที่ทำนายได้กับร้านอาหารที่ผู้ใช้เคยรับประทานประมาณ 3.73 สำหรับการทำงานของ CBF มีค่าสัมประสิทธิ์ความคล้ายคลึงกันของคะแนนความพึงพอใจที่ทำนายได้กับคะแนนความพึงพอใจที่ผู้ใช้เคยให้ประมาณ 0.71 และ ค่าสัมประสิทธิ์ความคล้ายคลึงกันของร้านอาหารที่ทำนายได้กับร้านอาหารที่ผู้ใช้เคยรับประทานประมาณ 3.70 ซึ่งสามารถกล่าวได้ว่าการทำนายแบบ CF และ CBF มีค่าสัมประสิทธิ์ของความคล้ายคลึงกันมากกว่าการทำนายแบบ CBF เพียงอย่างเดียว

บทที่ 5
ข้อสรุปและข้อเสนอแนะ

จากการทดลองในบทที่ 4 Restaurant Multicriteria Recommendation based on Recurrent Neural Network มีข้อสรุปดังนี้

5.1 ข้อสรุป

โครงการนี้ได้นำเสนองานวิจัยระบบแนะนำร้านอาหารโดยโดยมีการประมวลผลของ Collaborative Filtering และ Content-Based Filtering ร่วมกันบนRecurrent Neuron Nertwork โดยใช้ Reviews และ Multicriteria item-based จากผลการทดลองการรวมขั้นตอนระหว่าง Reviews ในส่วนของ Collaborative Filtering และ Multicriteria item-based ในส่วน Content-Based Filtering ให้ผลลัพธ์ที่ดีกว่า Multicriteria item-based ในส่วน Content-Based Filtering เพียงอย่างเดียว



รูปที่ 5.1 แสดงภาพรวมการทำงานทั้งหมด

รายการอ้างอิง

[1] Tim Donkers, Benedikt Loepp, Jürgen Zieler. **Sequential User-based Recurrent Neural Network Recommendations**, University of Duisburg-Essen Duisburg, Germany, 2017.

[2] David Zhan Liu, Gurbir Singh **A Recurrent Neural Network based Recommendations System**. University of Stanford, 2017.

[3] Wenjie Pei, Jie Yang, Zhu Sun, **Interacting Attention-gated Recurrent Networks for Recommendation**., Delft University of Technology Delft, The Netherlands, 2017.

[4] Sarwar, B., et al., **Item-Based Collaborative Filtering Recommendation Algorithms**, **Proceedings of the 10th international conference on World Wide Web**. 2001, ACM: Hong Kong, Hong Kong. p. 285-295.

[5] Shuai Zhang, Lina Yao, Aixin Sun, Yi Tay, **Deep Learning based Recommender System: A Survey and New Perspectives**. 2018, ACM: University of New South Wales, Nanyang Technological University. p. 18-21.

[6] Keattisak Chankaew, Supot Nitsuwat, **A Comparison of Prediction Palm Oil Time Series Using Polynomial Artificial Neural Network and Artificial Neural Network, Proceedings of the 12th graduate research conference.** 2011, Khon Koen University.

[7] Sagar Jaiswal, **Keras-Recurrent-Neural-Network-Python** [Online]**.** 2017, Available from:https://github.com/sagar448/Keras-Recurrent-Neural-Network-Python/blob/master/KERAS%20Recurrent%20Neural%20Networks.py [2019, January 12]

[8] Faisal Rahutomo*, Teruaki Kitasuka, and Masayoshi Aritsugi, **Semantic Cosine Similarity.** 2011, Graduate School of Science and Technology, Kumamoto University.

ภาคผนวก

ภาคผนวก ก
แบบเสนอหัวข้อโครงงาน รายวิชา 2301399 Project Proposal
ปีการศึกษา 2561

ชื่อโครงงาน (ภาษาไทย) การแนะนำร้านอาหารแบบหลายเกณฑ์โดยโครงข่ายประสาทเทียมแบบวนซ้ำ

ชื่อโครงงาน (ภาษาอังกฤษ) Restaurant Multicriteria Recommendation based on Recurrent Neural Network

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ศรันญา มณีโรจน์

ผู้ดำเนินการ 1. นางสาวกชพรรณ สมสา เลขประจำตัวนิสิต 5833602423

2. นางสาวณัฐสินี ศรีเวทย์บดี เลขประจำตัวนิสิต 5833627123

สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

หลักการและเหตุผล

ระบบแนะนำ (Recommendation System) หมายถึง ระบบที่แนะนำข้อมูล ผลิตภัณฑ์ หรือผู้คน ซึ่งต่อไปจะถูกเรียกว่าวัตถุ (Item) ให้กับผู้ใช้ระบบโดยอ้างอิงจากสมมติฐานการเรียนรู้ข้อมูลความชอบหรือความต้องการ ณ ขณะนั้นของผู้ใช้ โดยระบบแนะนำเป็นหัวข้องานวิจัยที่สำคัญ และมีประโยชน์มากมาย เช่น การเพิ่มยอดขายให้กับธุรกิจ หรือเข้าใจสิ่งที่ผู้ใช้งานต้องการจริงๆ ในปัจจุบันระบบแนะนำเป็นระบบที่สำคัญในแอพพลิเคชัน หรือเว็บไซต์ที่ให้บริการออนไลน์ เช่น Wongnai, Food Panda, Line man เป็นต้น เทคนิคหลักของระบบแนะนำประกอบด้วย 2 วิธี คือ การกรองร่วม (Collaborative Filtering) และการกรองแบบอิงเนื้อหา (Content-Based-Filtering)

การกรองร่วม (Collaborative Filtering) เป็นวิธีหนึ่งที่นิยมใช้ในระบบแนะนำ โดยหาผู้ใช้งานจากข้อมูลที่ใช้ฝึกฝนซึ่งมีการให้รูปแบบของคะแนนความพึงพอใจที่คล้ายคลึงกับผู้ใช้งานเป้าหมายแล้วใช้ความนิยมของผู้ใช้งานที่คล้ายกันช่วยในการทำนาย โดยเราเรียกผู้ใช้งานที่คล้ายคลึงกับผู้ใช้งานเป้าหมายนั้นว่า เพื่อนบ้าน (Neighbors) สำหรับขั้นตอนแบบหนึ่งของการกรองร่วมเนื่องจากรับข้อมูลคะแนนความพึงพอใจในรูปแบบเมทริกซ์แล้วหาผู้ใช้คนอื่นๆ k คน ที่มีความพึง

พอใจใกล้เคียงกันมากที่สุด (k-Nearest Neighbors) เพื่อหาผู้ใช้งานที่มีการให้คะแนนความพึงพอใจคล้าย ๆกันหลังจากนั้นหากต้องการทำนายคะแนนความพึงพอใจของผู้ใช้งานเป้าหมาย ให้ใช้สมการที่ 1 ในการทำนายคะแนนความพึงพอใจของผู้ใช้งานคนนั้นต่อวัตถุที่ต้องการโดย u คือผู้ใช้งานที่ต้องการจะทำนายคะแนน, i คือวัตถุ, v หมายถึงเพื่อนบ้าน, N_u คือชุดข้อมูลเพื่อนบ้านของผู้ใช้งาน u, w_{uv} คือ ความเหมือนระหว่างผู้ใช้งาน u และเพื่อนบ้าน v, r_{vi} คือคะแนนที่เพื่อนบ้าน v บนวัตถุ i และ r_{ui} คือคะแนนความพึงพอใจที่ทำนายได้ของผู้ใช้งาน u บนวัตถุ i

$$\hat{r}_{ui} = \frac{\sum v \in N_u \, w_{uv} r_{vi}}{\sum v \in N_u \, |w_{uv}|} \qquad (1)$$

การกรองแบบอิงเนื้อหา (Content-Based Filtering) เป็นอีกวิธีหนึ่งที่นิยมใช้ในระบบแนะนำ โดยใช้พื้นฐานข้อมูลจากรายละเอียดของวัตถุที่ผู้ใช้เป้าหมายเคยให้ความพึงพอใจไว้ในอดีต ทำให้ได้ประวัติการใช้งานวัตถุของผู้ใช้เป้าหมาย (User Profile) จากนั้นจะนำประวัติดังกล่าวไปคำนวณหาค่าความคล้ายคลึงกันกับรายละเอียดข้อมูลของวัตถุอื่นๆแล้วดูวัตถุที่มีความสัมพันธ์กันมากที่สุดกับประวัติการใช้งานมาแนะนำให้กับผู้ใช้เป้าหมาย

เนื่องจากระบบแนะนำเป็นการใช้เทคนิคการจำแนกข้อมูล (Classification) อย่างหนึ่ง ซึ่งมีป้ายกำกับ (Label) เพื่อบอกประเภท โดยเริ่มแรกจะนำข้อมูลในอดีตมาให้ระบบเรียนรู้เพื่อกำกับประเภทของข้อมูล จากนั้นจะนำข้อมูลปัจจุบันมาทดสอบโดยเปรียบเทียบกับประเภทของข้อมูลในอดีต ซึ่งการจำแนกประเภทนั้นสามารถบอกได้โดยอาศัยการเทียบเคียงกับข้อมูลตัวที่อยู่ใกล้ที่สุด ด้วยเหตุนี้ โครงข่ายประสาทเทียมก็เป็นการใช้เทคนิคการจำแนกข้อมูลเช่นเดียวกัน จึงมีผู้วิจัยอื่นนำโครงข่ายประสาทเทียม (Neural Network) มาประยุกต์บนระบบแนะนำ โดยแต่เดิมนั้นกระบวนการกรองร่วมและการกรองแบบอิงเนื้อหาได้นำข้อมูลของผู้ใช้ (Users) และข้อมูลของวัตถุ (Items) มาเป็นข้อมูลขาเข้าในการดำเนินการ จากนั้นมีผู้วิจัยได้เริ่มนำเทคนิคการฝังตัว (Embedding) [1] เข้ามาใช้ในโครงข่ายประสาทเทียมซึ่งจะจัดคุณลักษณะของผู้ใช้ (User Characteristic) และคุณลักษณะของวัตถุ (Item Characteristic) เป็นเวกเตอร์ แล้วนำไปเป็นข้อมูลขาเข้าของระบบแนะนำ เพื่อให้มีประสิทธิภาพที่ดีขึ้น

ในโครงข่ายประสาทเทียมประกอบไปด้วย โครงข่ายประสาทเทียมตามแบบแผน (Conventional Neural Network), โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neural Network) [2] และ โครงข่ายประสาทเทียมแบบวนซ้ำสองทาง (Bidirectional Recurrent Neural Network) ซึ่งแบบโครงข่ายประสาทเทียมดั้งเดิม คือการนำข้อมูลขาเข้า (Inputs) มาวิเคราะห์ โดยนำไปเรียนรู้เชิงลึกในชั้นระหว่างกลาง (Hidden layer) เพื่อให้ได้ผลลัพธ์ (Outputs) ออกมาในรูปแบบง่ายๆ เช่น ชอบ หรือ ไม่ชอบ เป็นต้น เนื่องจากในปัจจุบันความพึงพอใจของผู้ใช้ที่มีต่อวัตถุมีการเปลี่ยนแปลงตลอดเวลา

เพราะฉะนั้นระบบแนะนำที่ใช้รูปแบบโครงข่ายประสาทเทียมแบบดั้งเดิมจึงไม่สามารถวิเคราะห์ข้อมูลในรูปแบบที่เปลี่ยนแปลงตามลำดับเวลาเพื่อบอกความสัมพันธ์ระหว่างข้อมูลขาเข้าตั้งแต่อดีตจนถึงปัจจุบันได้ ดังนั้นสำหรับการทำงานที่มีข้อมูลขาเข้าเป็นอนุกรมชุดรูปแบบ (Series) จึงเหมาะกับการใช้โครงข่ายประสาทเทียมแบบวนซ้ำ (Recurrent Neuron Network) เช่น เดือนมิถุนายนผู้ใช้ A รับประทานร้านส้มตำ เดือนกรกฎาคมรับประทานร้านพิซซ่า เดือนสิงหาคมรับประทานร้านติ่มซำ ซึ่งมีการเก็บรวบรวมข้อมูลแบบอนุกรมมาจนถึงปัจจุบันแล้วนำมาแนะนำว่าร้านถัดไปควรรับประทานอาหารร้านใด

ในปัจจุบัน ระบบแนะนำมีหลายโดเมน เช่น ภาพยนตร์ เพลง โรงแรม ร้านอาหาร เป็นต้น โดเมนร้านอาหารถูกนำมาใช้ในโครงข่ายประสาทเทียมแบบวนซ้ำ ซึ่งอาจอยู่ในรูปแบบของการกรองร่วมหรือการกรองแบบอิงเนื้อหา ซึ่งรูปแบบการกรองร่วมจะสามารถแนะนำวัตถุให้กับผู้ใช้เป้าหมาย (Target User) โดยมีการอ้างอิงจากเพื่อนบ้านที่มีความพึงพอใจคล้ายคลึงกัน และ การกรองแบบอิงเนื้อหา เป็นการแนะนำวัตถุที่มีความคล้ายคลึงกันกับประวัติการใช้งานวัตถุของผู้ใช้เป้าหมาย แต่มีข้อเสียคือข้อมูลขาออกของการแนะนำดังกล่าวจะมีแต่ข้อมูลที่มีความคล้ายคลึงกันกับความพึงพอใจในอดีตของผู้ใช้ทำให้ข้อมูลขาออกจะไม่มีความหลากหลาย จึงไม่เกิดการแนะนำวัตถุแบบใหม่ให้กับผู้ใช้เป้าหมาย จากเหตุผลข้างต้นจะสังเกตได้ว่าจากงานวิจัยที่ผ่านมาพบว่ายังไม่มีงานวิจัยที่นำทั้งสองรูปแบบมารวมกันในรูปแบบการแนะนำร้านอาหาร ผู้วิจัยจึงเสนอรูปแบบโดยเลือกใช้โดเมนร้านอาหารที่มีข้อมูลขาเข้า 2 อย่างคือ การเขียนวิจารณ์ร้านอาหาร (Reviews) ซึ่งจะนำมาใช้กับการกรองร่วม และ การฝังวัตถุ (Item Embedding) นำมาใช้กับการกรองแบบอิงเนื้อหา ในการนำสองวิธีมารวมกันนั้น ผู้วิจัยจะใช้วิธี กลไกความสนใจ (Attention Mechanism) [3] ที่มีการถ่วงน้ำหนัก (Weight) กับข้อมูลขาออกของวิธีการกรองร่วม แล้วนำไปเป็นข้อมูลขาเข้าของวิธีการกรองแบบอิงเนื้อหาเพื่อแนะนำร้านอาหารให้กับผู้เป้าใช้หมายที่จะรับประทานเป็นร้านถัดไป ดังนั้นผู้วิจัยคาดว่าการรวมรูปแบบทั้งสองสามารถให้ผลลัพธ์ที่ดียิ่งขึ้น

วัตถุประสงค์

เพื่อเสนอขั้นตอนวิธีใหม่ที่สามารถรวมวิธีการกรองร่วมและการกรองแบบอิงเนื้อหาเพื่อทำงานบนระบบแนะนำร้านอาหารโดยโครงข่ายประสาทเทียมแบบวนซ้ำ

ขอบเขตของโครงงาน

ข้อมูลที่นำมาทดสอบเป็นชุดข้อมูล ที่มีการวิจารณ์ร้านอาหารจากผู้ใช้ไม่เกิน 1,000 รายการ

วิธีการดำเนินงาน

1. ศึกษางานวิจัยและบทความทางวิชาการที่เกี่ยวข้องกับระบบแนะนำ
2. ระบุปัญหาและข้อเสียที่เกิดขึ้นจากงานวิจัยก่อนหน้า
3. ออกแบบและวิเคราะห์วิธีการแก้ปัญหาของระบบแนะนำ
4. พัฒนาและทดสอบความถูกต้องของระบบแนะนำที่ได้นำเสนอ
5. ทำการทดลองเพื่อวัดประสิทธิภาพของระบบที่ได้นำเสนอ
6. วิเคราะห์และอภิปรายผลการทดลองพร้อมทั้งนำผลการอภิปรายไปปรับปรุงระบบแนะนำที่ได้นำเสนอให้ดีขึ้น
7. จัดทำเอกสารต่าง ๆที่เกี่ยวข้อง เช่น แบบเสนอหัวข้อโครงงาน รูปเล่มโครงงานวิทยาศาสตร์ระดับปริญญาบัณฑิต และเอกสารประกอบการนำเสนอ

จากวิธีการดำเนินงานที่กล่าวมา สามารถเขียนเป็นแผนภูมิแกนต์ (Gantt Chart) ได้ดังนี้

ขั้นตอนการดำเนินงาน	ปี 2560							ปี2561			
	มิ.ย.	ก.ค.	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1 .ศึกษางานวิจัยและบทความทางวิชาการที่เกี่ยวข้อง											
2. ระบุปัญหาและข้อเสียที่เกิดขึ้นจากงานวิจัยก่อนหน้า											
3. ออกแบบและวิเคราะห์วิธีการแก้ปัญหาของระบบแนะนำ											
4. พัฒนาและทดสอบความถูกต้องของระบบแนะนำที่ได้นำเสนอ											
5. ทำการทดลองเพื่อวัดประสิทธิภาพของระบบที่ได้นำเสนอ											
6. วิเคราะห์ อภิปรายผลการทดลองและปรับปรุงระบบแนะนำที่ได้นำเสนอ											
7.จัดทำเอกสารต่าง ๆที่เกี่ยวข้อง											

ประโยชน์ที่คาดว่าจะได้รับ

1. ประโยชน์ในด้านความรู้และประสบการณ์ต่อตัวนิสิตเอง
    1.1. ได้เรียนรู้ทฤษฎีและการปฏิบัติของการสร้างขั้นตอนวิธีใหม่สำหรับทำระบบแนะนำ
    1.2. ได้นำความรู้ในสาขาวิชาวิทยาการคอมพิวเตอร์ในการแก้ปัญหากับระบบแนะนำ
2. ประโยชน์ในด้านการแก้ไขปัญหาของสังคมหรือสภาพแวดล้อม
    2.1. ได้พัฒนาและต่อยอดความรู้ใหม่ซึ่งมีประโยชน์ต่อวงการวิจัย
    2.2. ได้พัฒนาองค์ความรู้ซึ่งจะสามารถนำไปประยุกต์ใช้กับธุรกิจในอนาคต

อุปกรณ์และเครื่องมือที่ใช้

1. ฮาร์ดแวร์
    1.1 เครื่องคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการ Window 8 มีหน่วยประมวลผล 2.66 GHz Intel Core Pentium หน่วยความจำหลัก 2 GB
    1.2 เครื่องคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการ Window 8 มีหน่วยประมวลผล 2.40 GHz AMD A8-7200P Radeon R5 หน่วยความจำหลัก 4 GB
2. ซอฟต์แวร์
    2.1 Jupiter Notebook
    2.2 MySQL Database Version 5.7

งบประมาณ

1.	หมึกเติมสีดำ Brother LC-38BK	1 ตลับ	540	บาท
2.	WIRELESS MOUSE M185	2 ชิ้น	1398	บาท
3.	WD My Passport 2017 1 TB	2 ชิ้น	2980	บาท
4.	กระดาษ A4 Double A	1 ชุด	120	บาท
		รวม	5038	บาท

หมายเหตุ: งบประมาณอาจมีการเปลี่ยนแปลงตามความเหมาะสม และราคาดังกล่าวอ้างอิงจาก http://www.logitech.com, http://www.wddevice.com และ http://www.officemate.co.th

เอกสารอ้างอิง

[1] Tim Donkers.; Benedikt Loepp.; Jürgen Zieler. Sequential User-based Recurrent Neural Network Recommendations., University of Duisburg-Essen Duisburg,
Germany, 2017.

[2] David Zhan Liu, Gurbir Singh A Recurrent Neural Network based Recommendations System. University of Standford, 2017.

[3] Wenjie Pei, Jie Yang, Zhu Sun Interacting Attention-gated Recurrent Networks for Recommendation., Delft University of Technology Delft, The Netherlands, 2017

ประวัติผู้เขียน

นางสาวกชพรรณ สมสา
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
Email: Kaopode.bq@hotmail.com

นางสาวณัฐสินี ศรีเวทย์บดี
ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
Email: mintnamo@hotmail.com