ระบบแนะนำร้านอาหารในประเทศไทยแบบผสมด้วยการเรียนรู้เชิงลึก

น.ส.อภิสรา แซ่ลิ้ม

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Deep Hybrid Restaurant Recommender System in Thailand

Miss Apisara Saelim

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Department of Computer Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2021

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	ระบบแนะนำร้านอาหารในประเทศไทยแบบผสมด้วยการเรียนรู้เชิงลึก
โดย	น.ส.อภิสรา แซ่ลิ้ม
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ดร.พีรพล เวทีกูล)

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

..... กรรมการ
(อาจารย์ ดร.พิตติพล คันธวัฒน์)

..... กรรมการภายนอกมหาวิทยาลัย
(ผู้ช่วยศาสตราจารย์ ดร.อนันตพร หรรษคุณาฒัย)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

อภิสรา แซ่ลิ้ม : ระบบแนะนำร้านอาหารในประเทศไทยแบบผสมด้วยการเรียนรู้เชิงลึก. (Deep Hybrid Restaurant Recommender System in Thailand) อ.ที่ปรึกษาหลัก : ศ. ดร.บุญเสริม กิจศิริกุล

ในยุคที่ข้อมูลมากมายมหาศาล ระบบผู้แนะนำมีบทบาทสำคัญอย่างมากตราบใดที่ผู้บริโภคยังคงมีความต้องการใช้ข้อมูลและส่งข้อมูลมากขึ้น ธุรกิจจำนวนมากได้ใช้ระบบผู้แนะนำเพื่อเป็นตัวช่วยให้แก่ผู้ใช้ได้ทำการค้นหาสินค้าหรือรายการโดยอ้างอิงจากข้อมูลการบริโภคของผู้ใช้ที่ผ่านมา โครงข่ายประสาทเชิงลึกได้แสดงให้เห็นผลลัพธ์ที่น่าสนใจในหลากหลายสาขาวิชา รวมถึงระบบผู้แนะนำในช่วงไม่กี่ปีที่ผ่านมา อย่างไรก็ตาม การศึกษาดังกล่าวได้ละเว้นการใช้ข้อมูลเสริมในแบบจำลอง ดังนั้น ในงานวิจัยนี้จึงนำเสนอระบบการแนะนำเชิงลึกด้วยโครงข่ายประสาท ซึ่งประกอบด้วยการกรองการทำงานร่วมกันเชิงลึกเพื่อเรียนรู้ปัจจัยแฝงของการโต้ตอบของผู้ใช้และสินค้า และเพิ่มประสิทธิภาพของแบบจำลองด้วยข้อมูลคุณลักษณะของสินค้าโดยใช้เพอร์เซปตรอนแบบหลายชั้น และรวมทั้งสองแบบจำลองนี้เข้าด้วยกัน เรียกว่า แบบจำลอง DNNRecs นอกเหนือจากโครงสร้างของแบบจำลองแล้ว ยังได้มีการนำเสนอวิธีการทำวิศวกรรมคุณลักษณะเพื่อสร้างคุณลักษณะใหม่จากข้อความวิจารณ์โดยใช้เทคนิค tf-idf งานวิจัยนี้ได้ใช้ชุดข้อมูลที่เกิดขึ้นจริงในประเทศไทยและแสดงให้เห็นถึงประสิทธิภาพของแบบจำลองที่นำเสนอ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่อนิสิต .....
ปีการศึกษา	2564	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6270312721 : MAJOR COMPUTER SCIENCE

KEYWORD: Top-N recommender systems, Auxiliary Information, Information Retrieval, Deep Neural Network

Apisara Saelim : Deep Hybrid Restaurant Recommender System in Thailand. Advisor: Prof. BOONSERM KIJSIRIKUL

In the age of flooded information, Recommender Systems play a crucial role as long as consumers consume more content and submit more data. Many businesses have implemented Recommender Systems to assist users find items based on their previous interactions. Deep neural networks have demonstrated promising results in a variety of disciplines, including recommendation systems in the past few years. However, such studies ignore auxiliary information input. In this work, we propose a deep recommendation system with neural networks that consists of deep collaborative filtering to learn user and item interaction latent factor and multi-layer perceptrons to enrich the performance with textual information and combines these two sub-models, called DNNRecs. Apart from our model framework, we also contribute a feature engineering method to create new features from review text by using technique tf-idf. Extensive experiments on one real-life dataset in Thailand demonstrate the effectiveness of the proposed model.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study: Computer Science Student's Signature .....

Academic Year: 2021 Advisor's Signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ได้รับความอนุเคราะห์จากศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ได้ให้คำปรึกษา แนะนำ และข้อคิดเห็นต่าง ๆ อันเป็นประโยชน์อย่างยิ่งในการทำวิทยานิพนธ์ รวมถึงวิธีการตีพิมพ์ผลงานทางวิชาการและการเขียนวิทยานิพนธ์ อีกทั้งยังช่วยตรวจสอบแก้ไขข้อผิดพลาดต่าง ๆ ผู้วิจัยขอขอบพระคุณด้วยความเคารพเป็นอย่างสูง

ขอขอบพระคุณ รองศาสตราจารย์ ดร.พีรพล เวทีกูล อาจารย์ ดร.พิตติพล คันธวัฒน์ และผู้ช่วยศาสตราจารย์อนันตพร หรรษคุณาฒัย กรรมการสอบวิทยานิพนธ์ที่ได้เสียสละเวลาอันมีค่าชี้ให้เห็นจุดที่สามารถปรับปรุง แนวทาง และคำแนะนำที่เป็นประโยชน์ในการทำวิทยานิพนธ์ฉบับนี้

นอกจากนี้ ขอขอบคุณบริษัท วงใน มีเดีย จำกัด สำหรับชุดข้อมูลในการทำการทดลองในวิทยานิพนธ์ฉบับนี้

สุดท้ายนี้ผู้วิจัยขอขอบคุณผู้เกี่ยวข้องทุกท่านที่มิได้กล่าวมาข้างต้นที่คอยให้ความช่วยเหลือ ทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยหวังเป็นอย่างยิ่งว่าวิทยานิพนธ์ฉบับนี้จะเป็นประโยชน์ต่อผู้สนใจไม่มากก็น้อย

อภิสรา แซ่ลิ้ม

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

..... ค

บทคัดย่อภาษาไทย ..... ค

..... ง

บทคัดย่อภาษาอังกฤษ ..... ง

กิตติกรรมประกาศ ..... จ

สารบัญ ..... ฉ

สารบัญตาราง ..... ฌ

สารบัญรูปภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1. ความเป็นมาและความสำคัญของปัญหา 1

1.2. วัตถุประสงค์ของงานวิจัย 2

1.3. ขอบเขตการวิจัย 2

1.4. ประโยชน์ที่คาดว่าจะได้รับ 2

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง 4

2.1. ระบบแนะนำ 4

2.1.1. การกรองแบบพิจารณาเนื้อหา 4

2.1.2. การกรองแบบพึ่งพาผู้ใช้ร่วม 4

2.1.3. ระบบแนะนำแบบผสม 4

2.2. นิวรอลเน็ตเวิร์ก 5

2.2.1. เพอร์เซปตรอน 5

2.2.2. ฟังก์ชันกระตุ้น 6

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.2.3. ฟังก์ชันต้นทุน.....6
2.3. นิวรอลเน็ตเวิร์กคอนโวลูชัน.....6
2.3.1. ชั้นคอนโวลูชัน.....7
2.3.2. ผังคุณลักษณะ.....7
2.3.3. ชั้นการรวม.....7
2.3.4. ชั้นการเชื่อมโยงเต็มรูปแบบ.....8
2.4. การแยกตัวประกอบเมทริกซ์.....8
2.5. เพอร์เซปตรอนหลายชั้น.....9
2.6. การวัดประสิทธิภาพ.....9
2.7. เทคนิคการวิเคราะห์ค้นหาคำสำคัญของข้อมูลด้วยวิธี TF-IDF.....10
บทที่ 3 งานวิจัยที่เกี่ยวข้อง.....11
3.1. งานวิจัยในกลุ่มการกรองการทำงานร่วมกันเชิงลึก.....11
3.2. งานวิจัยในกลุ่มแบบจำลองผสมเชิงลึก.....13
บทที่ 4 แนวคิดและวิธีการดำเนินงาน.....15
4.1. แบบจำลองแบบพึ่งพาผู้ใช้ร่วม.....15
4.2. แบบจำลองแบบผสม.....15
บทที่ 5 การออกแบบและพัฒนาแบบจำลอง.....17
5.1. ชุดข้อมูล.....17
5.2. การทำความสะอาดข้อมูล.....18
5.2.1. การจัดการข้อมูลซ้ำ.....18
5.2.2. การจัดการข้อมูลผิดปกติ.....19
5.2.3. การจัดการค่าว่าง.....19
5.2.4. การนำอักษรที่ไม่เกี่ยวข้องออก.....19

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.2.5. การตัดคำภาษาไทย....20
5.2.6. การเข้ารหัสข้อมูล....20
5.2.7. การปรับช่วงขอบเขตข้อมูล....21
5.3. การเตรียมชุดข้อมูลสำหรับการทดลอง....21
5.4. การสร้างข้อมูลคุณลักษณะเสริม....22
5.5. สถาปัตยกรรมแบบจำลอง....23
5.6. ผลการทดลอง....24
บทที่ 6 สรุปผลการวิจัย....29
6.1. สรุปผลการวิจัย....29
บรรณานุกรม....30
ประวัติผู้เขียน....33

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

หน้า

ตารางที่ 1 ลักษณะข้อมูลในชุดข้อมูลร้านอาหาร 17

ตารางที่ 2 ลักษณะข้อมูลในชุดข้อมูลการให้คะแนนและคำวิจารณ์.....18

ตารางที่ 3 ตัวอย่างอักษรและสัญลักษณ์ที่ไม่เกี่ยวข้อง.....20

ตารางที่ 4 คุณลักษณะชุดข้อมูล.....21

ตารางที่ 5 เปรียบเทียบประสิทธิภาพของแบบจำลองที่นำเสนอต่อแบบจำลองพื้นฐานด้วยเมทริกซ์ HR@10 และ NDCG@10 26

ตารางที่ 6 เปรียบเทียบประสิทธิภาพของแบบจำลองที่นำเสนอระหว่างแบบจำลองที่มีการฝึกอบรมจากแบบจำลองย่อยก่อนหน้าต่อแบบจำลองที่เริ่มการฝึกอบรมตั้งแต่ต้นด้วยเมทริกซ์ HR@10 และ NDCG@10.....26

ตารางที่ 7 ผลการทดลองที่มีความไวต่อ hyper-parameter ระหว่าง latent factor และ hidden layer.27

ตารางที่ 8 ตัวอย่างผลการแนะนำร้านอาหาร 10 อันดับแรก เมื่อ HR@10 เท่ากับ 1.....27

ตารางที่ 9 ตัวอย่างผลการแนะนำร้านอาหาร 10 อันดับแรก เมื่อ HR@10 เท่ากับ 0.....27

ตารางที่ 10 เปรียบเทียบประสิทธิภาพของแบบจำลองเมื่อจำนวนร้านอาหารที่แนะนำต่างกัน 28

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญรูปภาพ

หน้า

รูปที่ 1 โครงสร้างเพอร์เซปตรอนเมื่อแต่ละข้อมูลรับเข้ามีค่าน้ำหนัก 5
รูปที่ 2 โครงสร้างนิวรอลเน็ตเวิร์กคอนโวลูชัน 7
รูปที่ 3 ชั้นการรวม 8
รูปที่ 4 โครงสร้างแบบจำลอง Neural collaborative filtering 11
รูปที่ 5 โครงสร้างแบบจำลอง Outer Product-based NCF 12
รูปที่ 6 โครงสร้างแบบจำลอง ConvNCF 12
รูปที่ 7 โครงสร้างแบบจำลอง DeepCF 13
รูปที่ 8 แนวคิดแบบจำลอง 15
รูปที่ 9 กระบวนการทดสอบแบบจำลอง 17
รูปที่ 10 ข้อมูลร้านอาหารที่มีข้อมูลซ้ำซ้อน 19
รูปที่ 11 ตัวอย่างคำในพจนานุกรมคำศัพท์เกี่ยวกับร้านอาหาร 20
รูปที่ 12 ตัวอย่างการเตรียมชุดข้อมูลฝึกฝน 22
รูปที่ 13 การสร้างข้อมูลคุณลักษณะเสริมจากคำวิจารณ์ด้วยเทคนิค tf-idf 23
รูปที่ 14 สถาปัตยกรรมแบบจำลอง 23

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 1
บทนำ

1.1. ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันผู้ใช้บริการทุกคนต้องเผชิญกับทางเลือกที่หลากหลาย ตัวอย่างเช่น ผู้ใช้บริการบางคนกำลังมองหาหนังสือสักเล่มเพื่ออ่านโดยไม่มีความคิดที่เฉพาะเจาะจงเกี่ยวกับสิ่งที่ผู้ใช้บริการต้องการ มีความเป็นไปได้อย่างมากที่เขาเหล่านั้นจะต้องเสียเวลาไปกับการค้นหาข้อมูลจากอินเตอร์เน็ตหรือเว็บไซต์ต่าง ๆ เพื่อมองหาคำแนะนำจากผู้ใช้บริการคนอื่น ๆ แต่ถ้าหากมีเว็บไซต์หรือแอปพลิเคชันที่สามารถแนะนำสินค้าหรือบริการให้กับผู้ใช้บริการเหล่านี้ได้อย่างตรงกับความต้องการของผู้ใช้บริการ โดยอ้างอิงจากประวัติข้อมูลที่ผู้ใช้บริการเคยไปใช้บริการหรือแสดงความคิดเห็น นอกจากที่ผู้ใช้บริการจะไม่ต้องเสียเวลาในการค้นหาแล้ว ระบบแนะนำยังสามารถแนะนำสินค้าหรือบริการที่ตรงกับรสนิยมของผู้ใช้บริการได้อีกด้วย

ธุรกิจออนไลน์จำนวนมากอาศัยความคิดเห็นและการให้คะแนนของลูกค้าโดยเฉพาะความคิดเห็นที่ชัดเจนอย่างเช่น การให้คะแนน มีความสำคัญอย่างยิ่งในอุตสาหกรรมบันเทิงและอีคอมเมิร์ซที่การมีส่วนร่วมของลูกค้าทั้งหมดได้รับผลกระทบจากการให้คะแนนเหล่านี้ ตัวอย่างเช่น Netflix อาศัยข้อมูลการให้คะแนนดังกล่าวเพื่อพัฒนาระบบแนะนำและเพิ่มประสิทธิภาพในการแนะนำภาพยนตร์หรือรายการทีวีที่ดีที่สุดที่เหมาะสมและตรงกับรสนิยมของผู้ใช้บริการมากที่สุด และนี่คือสิ่งที่ระบบแนะนำทำ ทุกวันนี้มีหลากหลายธุรกิจที่เป็นที่รู้จักได้นำระบบแนะนำนี้ไปใช้ในการช่วยแนะนำผู้ใช้บริการของธุรกิจของเขาเหล่านั้น เช่น Amazon, Netflix หรือ YouTube เป็นต้น

ธุรกิจร้านอาหารเป็นอีกหนึ่งธุรกิจที่คนส่วนมากชอบเขียนคำวิจารณ์เกี่ยวกับร้านอาหารที่เคยได้ไปใช้บริการเพื่อเป็นการแบ่งปันประสบการณ์ของผู้ใช้บริการ ไม่ว่าผู้ใช้บริการจะชอบหรือไม่ก็ตาม ตัวอย่างเช่น เว็บไซต์วงใน (www.wongnai.com) ซึ่งเป็นเว็บไซต์บริการเครือข่ายสังคม โดยผู้ใช้เว็บไซต์สามารถค้นหาบริการ ร้านอาหาร ข้อมูล และคำวิจารณ์จากสมาชิกคนอื่นได้ เพื่อมองหาบริการหรือร้านอาหารที่น่าสนใจและใช้ในการตัดสินใจเลือกว่าจะไปรับบริการหรือร้านอาหารนั้น ๆ หรือไม่ ในปี 2009 บริษัท Netflix ได้มีการจัดการแข่งขัน The Netflix Prize ขึ้น เพื่อค้นหาอัลกอริทึมผู้แนะนำที่สามารถปรับปรุงระบบผู้แนะนำที่มีอยู่ของ Netflix โดย Netflix สามารถสร้างมูลค่าทางเศรษฐกิจได้เป็นอย่างมากจากระบบผู้แนะนำ มีรายงานว่า 80% ของผู้ใช้งานเลือกวิดีโอ Netflix เป็นผลมาจากคำแนะนำส่วนบุคคล [1]

The Netflix Prize มอบข้อมูลและสิ่งจูงใจสำหรับนักวิจัย ซึ่งนำไปสู่การปรับปรุงที่สำคัญในการนำวิธีการแยกตัวประกอบเมทริกซ์ (Matrix Factorization) ไปใช้กับระบบผู้แนะนำ วิธีการแยกตัวของเมทริกซ์ เช่น การแยกตัวของค่าเอกพจน์ (Singular Value Decomposition หรือ SVD) ที่ได้มีการนำมาใช้งานก่อนหน้านี้ แต่ในระหว่างและหลังจากที่การแข่งขันจบลง ได้มีการนำวิธีดังกล่าวมาใช้มากขึ้นและได้รับการปรับปรุงอย่างมากสำหรับการกรองร่วม (Collaborative filtering) ในตอนท้ายของการแข่งขันทีม BellKor's Pragmatic Chaos ซึ่งเป็นการรวมตัวกันของ 2 ทีม คือ ทีมของ KorBell และทีมของ Big Chaos ชนะการแข่งขัน ได้รับรางวัลใหญ่มูลค่า 1 ล้านเหรียญ จึงทำให้การแยกตัวประกอบ (Matrix Factorization) เป็นวิธีการที่รู้จักกันอย่างแพร่หลายในระบบผู้แนะนำตามแบบจำลองการกรองร่วมกัน (Collaborative filtering) ซึ่งในปี 2009 Koren และคณะ [2] ได้ตีพิมพ์บทความ

เทคนิคการแยกตัวประกอบเมทริกซ์สำหรับระบบผู้แนะนำใน IEEE โดยสรุปเทคนิคที่ใช้ในการแยกตัวประกอบเมทริกซ์และวิธีที่สามารถนำไปประยุกต์ใช้กับงานระบบแนะนำได้

อย่างไรก็ตามในช่วงหลายปีที่ผ่านมาการเรียนรู้เชิงลึกได้รับความสนใจอย่างมากในงานวิจัยหลายสาขา จึงทำให้มีหลายงานวิจัยเกี่ยวกับระบบผู้แนะนำแบบเรียนรู้เชิงลึก ได้นำการเรียนรู้เชิงลึกไปใช้งานและแสดงให้เห็นถึงประสิทธิภาพที่ดีขึ้นของระบบผู้แนะนำเมื่อนำไปใช้กับการดึงข้อมูล เช่น งานวิจัย [3-5] เสนอวิธีการใช้การเรียนรู้เชิงลึกเพื่อเรียนรู้ความสัมพันธ์ระหว่างผู้ใช้บริการและวัตถุจากข้อมูลมาใช้แทนวิธีการแยกตัวประกอบ เช่น การใช้วิธีเพอร์เซปตรอนหลายชั้น (Multi-layer Perceptron) และ โครงข่ายประสาทเทียบแบบคอนโวลูชัน (Convolution Neural Network หรือ CNN) จากผลการทดลองพบว่าการใช้การเรียนรู้เชิงลึกช่วยพัฒนาประสิทธิภาพของระบบผู้แนะนำได้ดีขึ้นกว่าแบบดั้งเดิม งานวิจัย [6, 7] นำเสนอวิธีการนำการเรียนรู้เชิงลึกมาใช้ในการตีความจากคำวิจารณ์เพื่อนำไปใช้ในการทำนายการให้คะแนนของผู้ใช้บริการต่อวัตถุต่าง ๆ ซึ่งจากผลการทดลองพบว่าแบบจำลองนี้สามารถช่วยแก้ปัญหาความเบาบางของข้อมูลและมีประสิทธิภาพในการทำนายการให้คะแนนจากคำวิจารณ์

ดังนั้นจึงมีแนวคิดที่จะนำเสนอแบบจำลองการพัฒนาระบบแนะนำร้านอาหารในประเทศไทยบนชุดข้อมูลภาษาไทย ที่มีความเบาบางของข้อมูลการให้คะแนนเป็นจำนวนมาก และข้อความวิจารณ์ภาษาไทยที่มีรูปแบบและโครงสร้างทางภาษาที่ซับซ้อนและแตกต่างจากภาษาอื่น ๆ ให้ได้ผลลัพธ์ที่มีประสิทธิภาพมากขึ้นจากชุดข้อมูลที่มีจำกัด โดยจะนำเสนอวิธีการสกัดคุณลักษณะของผู้ใช้บริการและร้านอาหารจากข้อความบทวิจารณ์ เพื่อนำมาเป็นข้อมูลคุณลักษณะใช้ร่วมกับแบบจำลองแบบพึ่งพาผู้อื่นร่วมจากงานวิจัย [5] โดยการวัดประสิทธิภาพจะวัดจากความแม่นยำว่าร้านอาหารที่ผู้ใช้บริการไปใช้บริการจากชุดข้อมูลทดสอบอยู่ใน Top N Recommend ร้านอาหารที่ระบบแนะนำหรือไม่

1.2. วัตถุประสงค์ของงานวิจัย

เพื่อนำเสนอแนวทางพัฒนาระบบแนะนำด้วยวิธีการเรียนรู้เชิงลึก โดยมุ่งเน้นให้แบบจำลองสามารถวิเคราะห์คุณลักษณะสำคัญจากข้อมูลการให้คะแนนและคำวิจารณ์ของผู้ที่เคยไปใช้บริการที่ร้านอาหารต่าง ๆ และพิจารณาความสัมพันธ์ของแต่ละคุณลักษณะในอดีต เพื่อเพิ่มประสิทธิภาพในการแนะนำให้มีค่าความแม่นยำมากที่สุด เมื่อเทียบกับแบบจำลองพื้นฐาน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.3. ขอบเขตการวิจัย

1.3.1. ชุดข้อมูลที่จะใช้ในงานวิจัยนี้ คือ ข้อมูลการให้คะแนนร้านอาหารและคำวิจารณ์ของลูกค้าที่มีต่อร้านอาหารที่นำมาจากเว็บไซต์วงใน (www.wongnai.com)

1.3.2. ชุดข้อมูลนี้ได้รับการสนับสนุนจาก บริษัท วงใน มีเดีย จำกัด ซึ่งเป็นผู้ให้บริการเว็บไซต์วงใน ซึ่งเป็นเว็บไซต์บริการเครือข่ายสังคม โดยผู้ใช้เว็บไซต์สามารถค้นหาบริการ ร้านอาหาร ข้อมูล และคำวิจารณ์จากสมาชิกคนอื่นได้

1.3.3. สร้างแบบจำลองระบบผู้แนะนำร้านอาหารจากข้อมูลการให้คะแนนของผู้ใช้บริการและข้อความบทวิจารณ์

1.3.4. สร้างแบบจำลองระบบผู้แนะนำร้านอาหารจากแบบจำลองการเรียนรู้เชิงลึก

1.4. ประโยชน์ที่คาดว่าจะได้รับ

1.4.1. พัฒนาระบบผู้แนะนำโดยใช้การสร้างแบบจำลองด้วยการเรียนรู้เชิงลึก

1.4.2. ปรับปรุงประสิทธิภาพของระบบผู้แนะนำด้วยนิวรอลเน็ตเวิร์กเชิงลึก

1.4.3. สกัดคุณลักษณะของร้านอาหารและผู้ใช้บริการจากข้อความบทวิจารณ์

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

2.1. ระบบแนะนำ

ระบบแนะนำ (Recommender System) คือ ระบบที่ช่วยแนะนำสินค้าหรือบริการ ซึ่งต่อไปนี้จะเรียกว่าสินค้า (Item) ให้กับลูกค้าหรือผู้ใช้บริการ โดยวิเคราะห์จากประวัติการใช้งานของผู้ใช้บริการในอดีต หรือวิเคราะห์จากข้อมูลการใช้งานของผู้ใช้บริการคนอื่น ๆ ที่มีคุณลักษณะที่คล้ายคลึงกับผู้ใช้บริการ ณ ขณะนั้น โดยตัวอย่างของ Application ที่ใช้ระบบแนะนำสินค้าให้กับผู้ใช้บริการ ได้แก่ เว็บไซต์ Amazon.com หรือ YouTube [8]

เบื้องหลังของการทำงานของระบบแนะนำนี้ จะประกอบไปด้วยหลากหลายแบบจำลอง ขึ้นอยู่กับข้อมูลผู้ใช้บริการและจุดประสงค์ที่ต้องการจากระบบแนะนำ โดยแบบจำลองส่วนใหญ่ที่มักจะถูกใช้งานมีอยู่ด้วยกัน 3 ประเภท ได้แก่ การกรองแบบพิจารณาเนื้อหา การกรองแบบพึ่งพาผู้ใช้ร่วม และระบบแนะนำแบบผสม

2.1.1. การกรองแบบพิจารณาเนื้อหา

เป็นรูปแบบจำลองที่เรียนรู้คุณลักษณะของสินค้าที่จะแนะนำ และจะแนะนำสิ่งที่มีคุณลักษณะที่คล้ายคลึงกับคุณลักษณะของผู้ใช้บริการ รวมถึงคุณลักษณะของสินค้าที่ผู้ใช้บริการเคยใช้หรือชอบ เช่น YouTube จะแนะนำคลิปวิดีโอที่มีคุณลักษณะหรือเนื้อหาที่คล้ายคลึงกับวิดีโอที่ผู้ใช้บริการเคยดูมาก่อนหน้านี้ ดังนั้นการใช้รูปแบบจำลองนี้ ระบบจะต้องมีข้อมูลคุณลักษณะของสินค้าหรือบริการ เช่น กลุ่ม/ประเภท ราคา หรือคำอธิบายตัวสินค้า เป็นต้น

2.1.2. การกรองแบบพึ่งพาผู้ใช้ร่วม

เป็นรูปแบบจำลองที่เรียนรู้จากพฤติกรรมของผู้ใช้บริการกับผู้ใช้บริการคนอื่น ๆ ที่มีความคล้ายคลึงกับผู้ใช้บริการ เช่น ถ้า Bob ชอบสินค้า A,B,C และ Alice ชอบสินค้า A,B,D ดังนั้น สินค้า C จะถูกแนะนำให้กับ Alice และสินค้า D จะถูกแนะนำให้กับ Bob เป็นต้น ดังนั้นการใช้ระบบนี้จะต้องมีข้อมูลการให้คะแนนของสินค้าหรือบริการของผู้ใช้แต่ละคนที่ผ่านมา โดยลักษณะการให้คะแนนมี 2 ประเภท [9] ได้แก่

- Explicit rating เป็นการให้คะแนนที่ชัดเจน เช่น การให้คะแนน หรือการกดให้ดาว 1-5
- Implicit rating เป็นการให้คะแนนโดยนัย เช่น การคำนวณคะแนนจากการกดเข้าไปดูสินค้าหรือกดถูกใจ

ซึ่งข้อมูลการให้คะแนนนี้ สามารถนำมาใช้ในการพัฒนาระบบแนะนำได้ 2 ประเภท

1. Memory based รูปแบบจำลองนี้จะมุ่งเน้นไปที่การหาความสัมพันธ์หรือความคล้ายระหว่างผู้ใช้บริการหรือวัตถุจากข้อมูลการให้คะแนน ตัวอย่างเช่น การใช้ Nearest Neighbor ในการทำนายการให้คะแนน จากการวิเคราะห์ข้อมูลจากผู้ใช้บริการ (User-based) หรือ สินค้า (Item-based) ที่ใกล้เคียงกัน

2. Model based จะใช้เทคนิคของการเรียนรู้ของเครื่อง (Machine Learning) เพื่อหา User embedding และ Item embedding มาทำนายคะแนนที่ผู้ใช้บริการจะให้กับวัตถุ

2.1.3. ระบบแนะนำแบบผสม

เป็นรูปแบบจำลองที่จะรวมเทคนิคแนะนำสองอย่างขึ้นไปเพื่อให้ได้ประสิทธิภาพที่ดีขึ้น จากการใช้ข้อดีของแต่ละแบบจำลองมาทำงานร่วมกัน โดยทั่วไปแบบจำลองแบบพึ่งพาผู้ใช้ร่วม จะทำงานร่วมกันกับเทคนิคอื่น ๆ เพื่อพยายามหลีกเลี่ยงปัญหา [10] ตัวอย่างเช่น การนำแบบจำลองแบบพิจารณาเนื้อหา มาทำงานร่วมกับแบบจำลองแบบพึ่งพาผู้ใช้ร่วม เพื่อช่วยแก้ปัญหาหากมีผู้ใช้บริการเข้ามาใหม่และยังมีประวัติข้อมูลการให้คะแนนในอดีตไม่เพียงพอ แบบจำลองพิจารณาเนื้อหา ก็จะสามารถมาช่วยแก้ไขปัญหานี้ได้

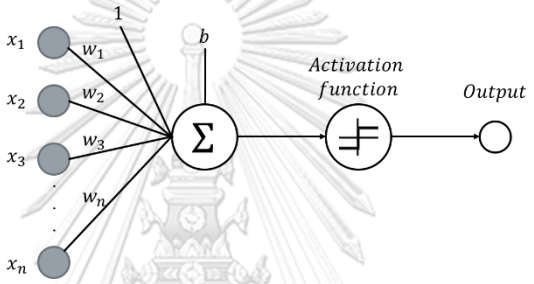
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2. นิวรอลเน็ตเวิร์ก

นิวรอลเน็ตเวิร์ก (Neural Network) หรือ โครงข่ายประสาทเทียม ถูกออกแบบมาให้เสมือนโครงข่ายประสาทสมอง ซึ่งมีวัตถุประสงค์ให้คอมพิวเตอร์สามารถทำงานได้แบบเดียวกับสมองมนุษย์ คือ การคิดและจดจำ โดยนิวรอลเน็ตเวิร์กสามารถเรียนรู้ได้จากชุดข้อมูลฝึกสอนที่ถูกป้อนเข้าไปเพื่อนำไปใช้ทำนายข้อมูลทดสอบที่ไม่เคยเห็นมาก่อน โดยในหัวข้อนี้จะมีหัวข้อย่อยที่กล่าวถึงการทำงานของนิวรอลเน็ตเวิร์กหลัก ๆ ดังต่อไปนี้ เพอร์เซปตรอน ฟังก์ชันกระตุ้น และฟังก์ชันต้นทุน

2.2.1. เพอร์เซปตรอน

เพอร์เซปตรอน (Perceptron) เป็นหน่วยย่อยที่สุดของนิวรอลเน็ตเวิร์กที่จำลองลักษณะของเซลล์ประสาท (Neuron) ของมนุษย์ โดยทำหน้าที่นำข้อมูลรับเข้ามาคำนวณหาผลรวมเชิงเส้น (Linear combination) เพื่อให้ได้ข้อมูลส่งออกที่เป็นค่าคงที่ และมีผลลัพธ์แตกต่างกันไปตามฟังก์ชันกระตุ้น (Activation function) ดังรูปที่ 1



รูปที่ 1 โครงสร้างเพอร์เซปตรอนเมื่อแต่ละข้อมูลรับเข้ามีค่าน้ำหนัก

เราสามารถแทนฟังก์ชันของเพอร์เซปตรอนที่จำแนกผลลัพธ์ออกเป็นสองกลุ่มด้วย $f(x)$ โดยที่ x เป็นข้อมูลรับเข้า และ $\hat{y}$ เป็นข้อมูลส่งออก โดยการแสดงการคำนวณสามารถเขียนได้ดังสมการที่ (1)

$$\hat{y} = f(x) = \begin{cases} 1 \; if \; \sum_{i=1}^{m} w_i x_i + b > 0 \\ 0 \qquad otherwise \end{cases} \tag{1}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

กำหนดให้ w คือ น้ำหนัก (weight)

b คือ ไบแอส (bias)

m คือ จำนวนทั้งหมดของข้อมูลรับเข้า

เนื่องจากเพอร์เซปตรอนไม่ได้จำแนกชุดข้อมูลตัวอย่างออกมาได้อย่างถูกต้องทุกตัว จึงจำเป็นต้องมีกระบวนการเรียนรู้ของเพอร์เซปตรอน เพื่อให้เพอร์เซปตรอนมีการปรับน้ำหนักในระหว่างการเรียนรู้ ส่งผลให้สามารถจำแนกตัวอย่างได้ถูกต้องมากขึ้น กำหนดให้ชุดข้อมูลตัวอย่างแทนด้วย x และผลลัพธ์จริงแทนด้วย y สามารถเขียนสมการการเรียนรู้ของเพอร์เซปตรอนได้ดังสมการที่ (2) และ (3)

$$w_i \leftarrow w_i + \Delta w_i \tag{2}$$

$$\Delta w_i = \alpha(\hat{y} - y)x_i \tag{3}$$

เมื่อ α เป็นอัตราการเรียนรู้ (Learning rate) ที่บ่งบอกถึงการเรียนรู้ของเพอร์เซปตรอนในแต่ละรอบว่ามีการเปลี่ยนแปลงน้ำหนักเมื่อเทียบกับอัตราส่วนผลลัพธ์ไปมากเท่าใด ถ้าอัตราการเรียนรู้มีค่ามากเกินไปเพอร์เซปตรอน

จะเรียนรู้ได้เร็ว แต่อาจไม่สำเร็จ แต่ถ้ามีค่าน้อย การปรับน้ำหนักสามารถทำได้อย่างละเอียดแต่จะใช้เวลาในการเรียนรู้นาน

2.2.2. ฟังก์ชันกระตุ้น

สำหรับข้อมูลส่งออกของเพอร์เซปตรอนตามปกติแล้วผลลัพธ์ที่ได้จะมีลักษณะเป็นเชิงเส้น ทำให้ไม่สามารถตอบปัญหาที่ซับซ้อนได้ จึงจำเป็นต้องมีการใช้ฟังก์ชันกระตุ้น (Activation function) เพื่อให้นิวรอลเน็ตเวิร์กสามารถตอบโจทย์ปัญหาได้หลากหลายรูปแบบ โดยฟังก์ชันกระตุ้นเปรียบเสมือนการทำงานของสมองขณะเรียนรู้ที่เซลล์ประสาทจะถูกกระตุ้นเพื่อส่งสัญญาณไปยังเซลล์อื่น ๆ เพื่อให้ประมวลผลต่อไป สำหรับฟังก์ชันกระตุ้นที่ใช้ในงานวิจัยนี้ ได้แก่ ฟังก์ชันเรคติไฟต์เชิงเส้น (Rectified Linear Unit : ReLU function) เป็นฟังก์ชันที่ให้ผลลัพธ์มีค่ามากกว่าหรือเท่ากับ 0 โดยสมการของฟังก์ชันเรคติไฟต์เชิงเส้นสามารถเขียนสมการคำนวณได้ดังสมการที่ (4)

$$ReLU(x) = \begin{cases} 0 \; if \; z < 0 \\ z \; if \; z \geq 0 \end{cases} \tag{4}$$

2.2.3. ฟังก์ชันต้นทุน

ฟังก์ชันต้นทุน (Cost function หรือ Loss function) เป็นฟังก์ชันที่แสดงถึงต้นทุนของนิวรอลเน็ตเวิร์ก โดยในกระบวนการเรียนรู้ของนิวรอลเน็ตเวิร์ก จะมีการปรับค่าน้ำหนักโดยมีจุดประสงค์เพื่อลดค่าของฟังก์ชันต้นทุน

กำหนดให้ J คือ ฟังก์ชันต้นทุน

n คือ จำนวนของข้อมูลทั้งหมดที่ใช้เรียนรู้

y_i คือ ผลลัพธ์จริงของข้อมูลชุดที่ i

$\hat{y}$ คือ ผลลัพธ์ที่ได้มาจากการทำนายชุดที่ i

สามารถเขียนสมการของฟังก์ชันต้นทุนที่นิยมได้ดังต่อไปนี้

1. ค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Squared Error : MSE)

$$J = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{5}$$

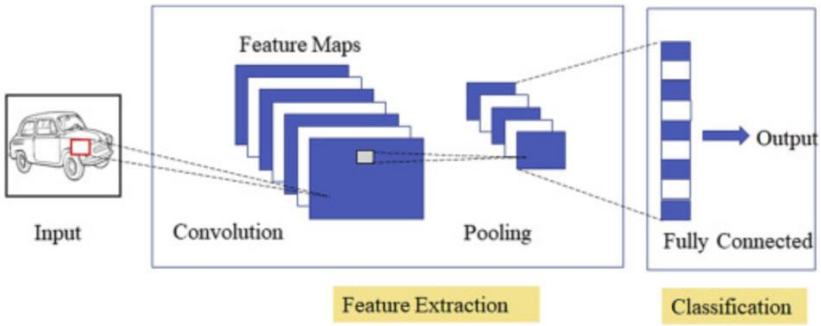
2. ค่าเฉลี่ยครอสเอนโทรปีแบบทวีภาค (Binary Cross-entropy)

$$J = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i) \tag{6}$$

2.3. นิวรอลเน็ตเวิร์กคอนโวลูชัน

นิวรอลเน็ตเวิร์กคอนโวลูชัน (Convolution Neural Network) ถูกนำมาใช้อย่างแพร่หลายในงานวิจัยที่เกี่ยวข้องกับการจำแนกรูปภาพเนื่องจากมีลักษณะพิเศษในการสกัดหาคุณลักษณะพิเศษของรูปภาพได้ ซึ่งภายหลังได้มีการถูกนำมาใช้สกัดคุณลักษณะพิเศษกับข้อมูลประเภทอื่น ๆ โดยอ้างอิงจากรูปแบบที่ใช้กับรูปภาพ โดยการแปลงข้อมูลรับเข้า (Input) ให้เป็นเวกเตอร์ (Vector)

เมื่อข้อมูลรับเข้า เข้าสู่นิวรอลเน็ตเวิร์กคอนโวลูชันจะผ่านตัวกรอง (Filter) เพื่อดำเนินการสร้างฟีเจอร์แมพ (Feature map) ซึ่งการกระทำดังกล่าวคือการสกัดคุณลักษณะพิเศษของข้อมูล โดยโครงสร้างของนิวรอลเน็ตเวิร์กคอนโวลูชันสามารถเขียนได้ดังรูปที่ 2



รูปที่ 2 โครงสร้างนิวรอลเน็ตเวิร์กคอนโวลูชัน

[ที่มา: https://www.sciencedirect.com/topics/engineering/convolutional-neural-network]

ส่วนประกอบของนิวรอลเน็ตเวิร์กคอนโวลูชันมีดังนี้

2.3.1. ชั้นคอนโวลูชัน

เมื่อข้อมูลรับเข้า เข้าสู่ชั้นคอนโวลูชัน (Convolution layer) ชั้นนี้จะทำหน้าที่หาฟีเจอร์จากกลุ่มข้อมูลรับเข้า ซึ่งจะพิจารณาข้อมูลรับเข้าเป็น เมทริกซ์ จากนั้นใช้วิธีการดอทเมทริกซ์กับตัวกรอง โดยน้ำหนักของตัวกรอง จะเป็นน้ำหนักที่มีการใช้ร่วมกันในทุก ๆ การทำคอนโวลูชันของข้อมูลรับเข้า

2.3.2. ผังคุณลักษณะ

ผังคุณลักษณะ (Feature Maps) เป็นสิ่งที่ได้จากตัวกรอง (Filter) ที่ช่วยดึงคุณลักษณะที่ใช้ในการรู้จำตัววัตถุออก โดยปกติตัวกรองอันหนึ่งจะดึงคุณลักษณะที่สนใจออกมาได้หนึ่งอย่าง จึงจำเป็นต้องมีตัวกรองหลายตัว เพื่อหาคุณลักษณะทางพื้นที่หลายอย่างประกอบกัน โดยวิธีการทำผังคุณลักษณะด้วยตัวกรองประกอบด้วย 2 ตัวแปร คือ ก้าวย่าง (Stride) และ การขยายความ (Padding)

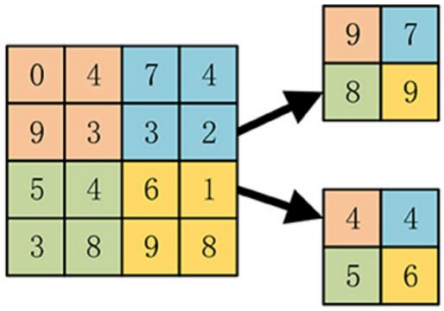
1. ก้าวย่าง (Stride) เป็นตัวกำหนดว่าจะเลื่อนตัวกรองไปด้วยสเตป (Step) เท่าไหร่ ซึ่งถ้าหากกำหนดค่าของก้าวย่างที่มากขึ้น จะทำให้ได้ผังคุณลักษณะที่มีขนาดเล็กลง
2. การขยายความ (Padding) เป็นการเติมเลขศูนย์หรือค่าต่าง ๆ เข้าไป เพื่อให้ผังคุณลักษณะที่ได้ยังคงมีขนาดเท่ากับข้อมูลขาเข้า (Input) เมื่อถูกผ่านชั้นนิวรอลเน็ตเวิร์กคอนโวลูชัน

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.3.3. ชั้นการรวม

ชั้นการรวม (Pooling layer) เป็นชั้นที่ทำหน้าที่ในการสกัดเอาส่วนที่สำคัญที่สุดของข้อมูล และเพิ่มประสิทธิภาพในการประมวลผลให้รวดเร็วยิ่งขึ้น โดยกลไกของชั้นการรวมสามารถทำได้ 2 วิธี คือ การรวมค่าสูงสุด (Max Pooling) และ การรวมค่าเฉลี่ย (Average Pooling)

1. การรวมค่าสูงสุด (Max Pooling) จะสกัดเอาเฉพาะค่าสูงสุดของตาราง (Grid) มาเก็บไว้ในข้อมูลขาออก (Output) ซึ่งเป็นรูปแบบที่นิยมใช้บ่อย
2. การรวมค่าเฉลี่ย (Average Pooling) เป็นการหาค่าเฉลี่ยของตาราง (Grid) มาเก็บไว้ในข้อมูลขาออก (Output)



รูปที่ 3 ชั้นการรวม

[ที่มา: https://www.frontiersin.org/articles/10.3389/fnins.2018.00818/full]

2.3.4. ชั้นการเชื่อมโยงเต็มรูปแบบ

ชั้นเชื่อมโยงเต็มรูปแบบ (Fully connected layer) เป็นชั้นสุดท้ายของโครงข่ายประสาทเทียมคอนโวลูชัน หลังจากที่ดำเนินการประกอบชั้นคอนโวลูชันและชั้นการรวมจำนวนหนึ่ง โดยชั้นนี้จะประกอบไปด้วยเพอร์เซปตรอนหลาย ๆ ตัว โดยที่เพอร์เซปตรอนแต่ละตัวจะมีเส้นเชื่อมกับเพอร์เซปตรอนทุกตัวในชั้นก่อนหน้า และเพอร์เซปตรอนทุกตัวในชั้นถัดไป ทำให้สามารถคำนวณการป้อนไปข้างหน้าและการแพร่กระจายย้อนกลับได้

2.4. การแยกตัวประกอบเมทริกซ์

การกรองการทำงานร่วมกันมักใช้ในรูปแบบของการแยกตัวประกอบของเมทริกซ์ (Matrix Factorization) โดยแสดงถึงผู้ใช้และรายการต่าง ๆ เป็นเวกเตอร์คุณลักษณะแฝง การโต้ตอบระหว่างผู้ใช้กับวัตถุ อาจอธิบายได้โดยใช้ผลิตภัณฑ์การคูณ (dot product) ของเวกเตอร์แฝงระหว่างผู้ใช้และวัตถุในพื้นที่คุณลักษณะ

การแยกตัวประกอบเมทริกซ์ เชื่อมโยงผู้ใช้บริการหนึ่งรายกับวัตถุทั้งหมดด้วยเวกเตอร์มูลค่าจริงของปัจจัยแฝง โดยที่ p_u และ q_i หมายถึง ปัจจัยแฝงสำหรับผู้ใช้ u และวัตถุ i โดย MF จะประมาณการโต้ตอบ y_{ui} ด้วยผลิตภัณฑ์ภายใน (inner product) ของ p_u และ q_i ดังสมการ (7)

$$\hat{y}_{ui} = f(u, i \mid p_{u,} q_i) = p_u^T q_i = \sum_{k=1}^{K} p_{uk}, q_{ki} \quad (7)$$

โดยที่ $\hat{y}_{(u,i)}$ คือ คะแนนการทำนาย

$p_{(u)}$ คือ เวกเตอร์แฝงของผู้ใช้ u

$q_{(i)}$ คือ เวกเตอร์แฝงของรายการ i

K คือ มิติของพื้นที่แฝง

ทุกมิติของพื้นที่แฝงนั้นเป็นอิสระจากมิติอื่น ๆ และ การแยกตัวประกอบเมทริกซ์จะจำลองการทำงานร่วมกันระหว่างผู้ใช้และเวกเตอร์แฝงของวัตถุ การแยกตัวประกอบเมทริกซ์ เป็นแบบจำลองเชิงเส้นของการวิเคราะห์องค์ประกอบแฝง เนื่องจากมิติของพื้นที่แฝงนั้นเชื่อมโยงเชิงเส้นกับน้ำหนักเท่ากัน

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.5. เพอร์เซปตรอนหลายชั้น

เพอร์เซปตรอน เป็นหน่วยย่อยที่สุดของโครงข่ายประสาทเทียมที่จำลองลักษณะของเซลล์ประสาทมนุษย์โดยนำอินพุตและชุดค่าผสมเชิงเส้นการคำนวณ เพื่อรับข้อมูลเอาต์พุตคงที่และมีผลลัพธ์ที่แตกต่างกันขึ้นอยู่กับฟังก์ชันการเปิดใช้งาน

เราสามารถแทนที่ฟังก์ชันของเพอร์เซปตรอนที่แบ่งผลลัพธ์ออกเป็นสองกลุ่มด้วย $f_{(x)}$ โดยที่ x คือ อินพุต และ $\hat{y}$ คือ เอาต์พุต โดยแสดงการคำนวณได้ดังสมการ (8)

$$\hat{y} = f_{(x)} = \begin{cases} 1, if \ \sum_{i=1}^{m} w_i x_i + b > 0 \\ 0, otherwise \end{cases} \quad (8)$$

โดยที่ w คือ น้ำหนัก

b คือ ค่าความเอนเอียง (bias)

m คือ อินพุตทั้งหมด

เมื่อพูดถึงการเรียนรู้การโต้ตอบระหว่างผู้ใช้กับรายการ เทคนิคเพอร์เซปตรอนหลายชั้น (Multi-Layer Perceptron) มีความยืดหยุ่นและไม่เป็นเชิงเส้นอย่างมาก สามารถเขียนการคำนวณได้ดังสมการ (9)

$$z_1 = \phi_1(p_u, q_i) = \begin{bmatrix} p_u \\ q_i \end{bmatrix}, \quad (9)$$

$$\phi_2(z_1) = a_2(W_2^T z_1 + b_2),$$

.....

$$\phi_L(z_{L-1}) = a_L(W_L^T z_{L-1} + b_L),$$

$$\hat{y}_{ui} = \sigma(h^T \phi_L(z_{L-1})),$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.6. การวัดประสิทธิภาพ

สำหรับงานวิจัยนี้จะแบ่งรายการร้านอาหารที่ผู้ใช้ไปครั้งล่าสุดเป็นชุดข้อมูลทดสอบ และสำหรับข้อมูลก่อนหน้าจะใช้เป็นข้อมูลฝึกฝน โดยค่าที่ใช้วัดประสิทธิภาพของระบบแนะนำนี้ จะพิจารณาร่วมกันทั้งหมด 2 ค่า

2.6.1. อัตราการตี (Hit Ratio) เป็นการพิจารณาความถูกต้องของรายการร้านอาหารที่ระบบแนะนำให้กับผู้ใช้จำนวน k รายการ (HR@k) สามารถคำนวณได้จากสมการ (10)

$$Hit\ Ratio = \frac{\sum_{i=1}^{N} HR@k}{N} \quad (10)$$

โดยที่ k คือ จำนวนรายการร้านอาหารที่ระบบจะทำการแนะนำ

N คือ จำนวนข้อมูลทดสอบทั้งหมด

2.6.2. ค่าปรับลดอันดับสะสมปกติ (NDCG หรือ Normalized Discounted Cumulative Gain) เป็นการพิจารณาความถูกต้องของรายการร้านอาหารที่ระบบแนะนำให้กับผู้ใช้ โดยพิจารณาจากลำดับในระบบแนะนำที่ตรงกับร้านอาหารในชุดทดสอบ หมายสามารถคำนวณได้จากสมการ (11)

$$NDCG = \frac{\sum_{i=1}^{N} \frac{\log(2)}{\log(rank+2)}}{N} \quad (11)$$

โดยที่ $rank$ คือ ลำดับของร้านอาหารในระบบแนะนำที่ตรงกับร้านอาหารชุดทดสอบ

N คือ จำนวนข้อมูลทดสอบทั้งหมด

2.7. เทคนิคการวิเคราะห์ค้นหาคำสำคัญของข้อมูลด้วยวิธี TF-IDF

เป็นเทคนิคที่ใช้ในการพิจารณาองค์ประกอบของคำภายในประโยคหรือเอกสารเป็นหลัก โดยจะไม่นำลำดับของคำภายในเอกสารมาใช้ประกอบการวิเคราะห์ด้วย เทคนิคนี้ประกอบด้วย 2 องค์ประกอบ ได้แก่

2.7.1. Term Frequency (TF) เป็นค่าที่บ่งบอกความถี่ของคำแต่ละคำที่ปรากฏในเอกสารเอกสารหนึ่ง โดยคิดคำนวณจากการนำจำนวนครั้งที่คำนั้น ๆ ปรากฏในเอกสารมาหารด้วยจำนวนคำทั้งหมดในเอกสาร สามารถคำนวณได้จากสมการ (12)

$$TF\ (\text{ของคำคำหนึ่ง}) = \frac{\text{จำนวนของคำนั้น ๆ ในเอกสาร}}{\text{จำนวนของคำทั้งหมดในเอกสาร}} \quad (12)$$

2.7.2. Inverse document Frequency (IDF) เป็นการคำนวณค่าน้ำหนัก (weight) ความสำคัญของแต่ละคำ โดยคำที่พบเจอได้บ่อย ๆ ในหลาย ๆ เอกสาร จะมีค่า IDF ต่ำ ซึ่งบ่งบอกว่าคำเหล่านั้นไม่มีความสำคัญต่อเอกสารนั้น ๆ สามารถคำนวณได้จากสมการ (13)

$$IDF\ (\text{ของคำคำหนึ่ง}) = \log\left(\frac{\text{จำนวนเอกสารทั้งหมดที่ใช้พิจารณา}}{\text{จำนวนเอกสารที่มีคำคำนั้นปรากฏอยู่}}\right) \quad (13)$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3
งานวิจัยที่เกี่ยวข้อง

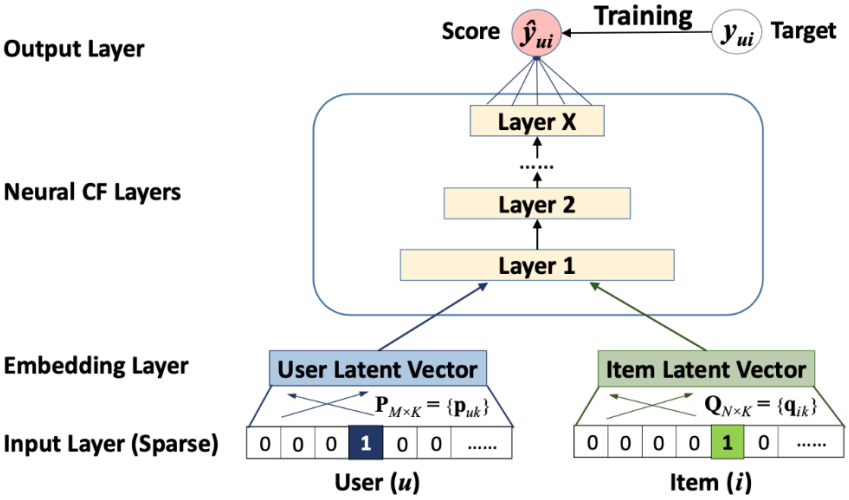
งานวิจัยที่เกี่ยวข้องในงานวิจัยนี้ ได้ทำการศึกษาใน 2 หัวข้อ คือ งานวิจัยในกลุ่มการกรองการทำงานร่วมกันเชิงลึก (Deep Collaborative Filtering) และงานวิจัยในกลุ่มแบบจำลองผสมเชิงลึก (Deep Hybrid Model)

3.1. งานวิจัยในกลุ่มการกรองการทำงานร่วมกันเชิงลึก

โครงข่ายประสาทเทียมแบบลึกสามารถสร้างแบบจำลองความไม่เป็นเชิงเส้นในข้อมูลด้วยการกระตุ้นแบบไม่เชิงเส้น เช่น relu, sigmoid, tanh เป็นต้น คุณสมบัตินี้ทำให้สามารถจับรูปแบบการโต้ตอบรายการผู้ใช้ที่ซับซ้อนได้

3.1.1. งานวิจัยของ He, Xiangnan และคณะ ใช้เพอร์เซปตรอนในการเรียนรู้การปฏิสัมพันธ์ระหว่างผู้ใช้กับสินค้าแทนวิธีการทำ Matrix Factorization แบบดั้งเดิม

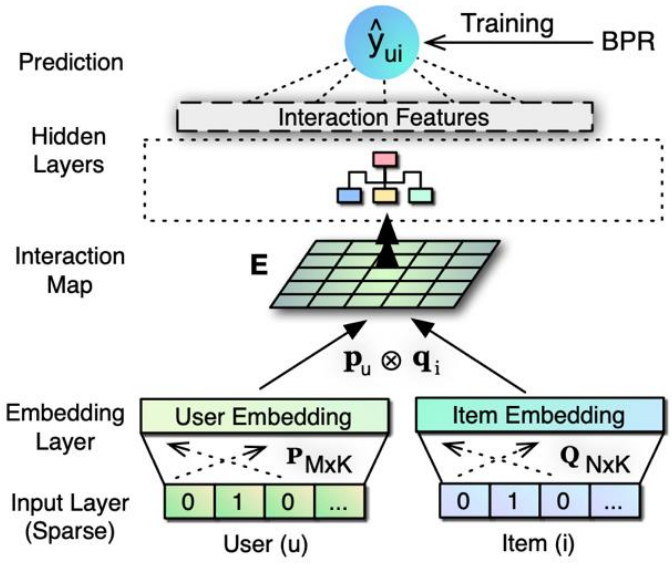
งานวิจัย ในปี 2017 [3] นำเสนอแบบจำลองการแยกตัวประกอบเมทริกซ์ของระบบประสาทแบบใหม่ โดยเป็นการรวมแนวคิดการแยกตัวประกอบเมทริกซ์แบบดั้งเดิมกับเพอร์เซปตรอนหลายชั้นเข้าด้วยกัน ทำให้แบบจำลองนี้มีทั้งจุดแข็งของความเป็นเชิงเส้นของการแยกตัวประกอบเมทริกซ์และความไม่เป็นเชิงเส้นของเพอร์เซปตรอนหลายชั้นสำหรับการสร้างแบบจำลองโครงสร้างแฝงของผู้ใช้และวัตถุ ดังรูปที่ 4 โดยงานวิจัยนี้ได้ทำการทดลองบนชุดข้อมูล MovieLens และ Pinterest และใช้วิธีการแบ่งชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบแบบ Leave-one-out โดยการใช้ข้อมูลการโต้ตอบครั้งสุดท้ายของแต่ละผู้ใช้เป็นชุดข้อมูลทดสอบ และชุดข้อมูลที่เหลือเป็นชุดข้อมูลฝึกฝน และวัดประสิทธิภาพของแบบจำลองระบบแนะนำจากค่า HR@k และ NDCG@k จากผลการทดลองพบว่าการใช้เพอร์เซปตรอนหลายชั้นช่วยเพิ่มประสิทธิภาพของระบบผู้แนะนำเพิ่มขึ้นจากวิธีการแยกตัวประกอบเมทริกซ์แบบดั้งเดิม



รูปที่ 4 โครงสร้างแบบจำลอง Neural collaborative filtering
[ที่มา: https://dl.acm.org/doi/pdf/10.1145/3038912.3052569]

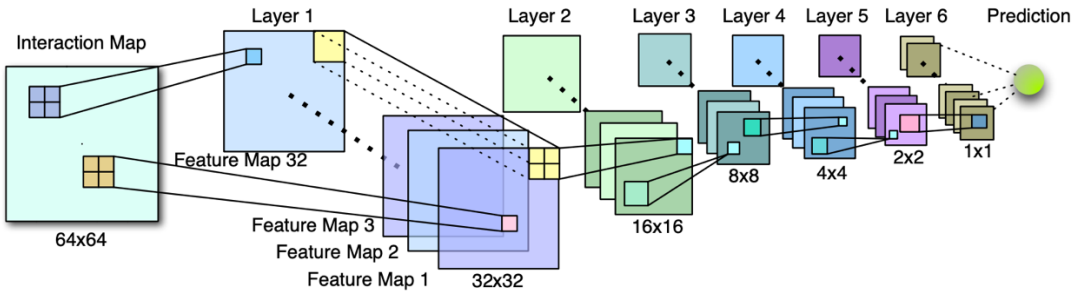
3.1.2. งานวิจัยของ He, Xiangnan และคณะ ใช้สถาปัตยกรรมโครงข่ายประสาทเชิงลึกในการทำการกรองแบบทำงานร่วมกัน (Collaborative filtering)

งานวิจัย [5] ในปี 2018 นำเสนอแบบจำลอง ConvNCF ซึ่งเป็นสถาปัตยกรรมใหม่ที่พัฒนามาจากแบบจำลอง NCF โดยมีแนวคิดคือ การใช้ผลิตภัณฑ์ภายนอก (Outer Product) บนชั้นการฝัง เพื่อสร้างแบบจำลองรูปแบบการโต้ตอบระหว่างผู้ใช้และสินค้า ดังรูปที่ 5 จากนั้นได้มีการนำเน็ตเวิร์กคอนโวลูชันหลายชั้นมาใช้กับผลลัพธ์ของผลิตภัณฑ์ภายนอกและจับความสัมพันธ์ระดับสูงระหว่างมิติการฝัง ดังรูปที่ 6 จากการทดลองพบว่า การใช้ผลิตภัณฑ์ภายนอกมีการเข้ารหัสสัญญาณมากกว่าการแยกตัวประกอบเมทริกซ์โดยทำหน้าที่ในการหาความสัมพันธ์ระหว่างมิติการฝังที่แตกต่างกัน อีกทั้งยังเก็บเฉพาะข้อมูลดั้งเดิมในชั้นการฝังโดยไม่ต้องสร้างแบบจำลองความสัมพันธ์ใด ๆ และในส่วนของการใช้เน็ตเวิร์กคอนโวลูชันหลายชั้นจะช่วยให้สร้างแบบจำลองที่ลึกกว่าเพอร์เซปตรอนหลายชั้นได้ง่ายและเป็นประโยชน์ต่อการเรียนรู้ความสัมพันธ์ลำดับสูงระหว่างมิติการฝัง



รูปที่ 5 โครงสร้างแบบจำลอง Outer Product-based NCF
[ที่มา: https://arxiv.org/pdf/1808.03912.pdf]

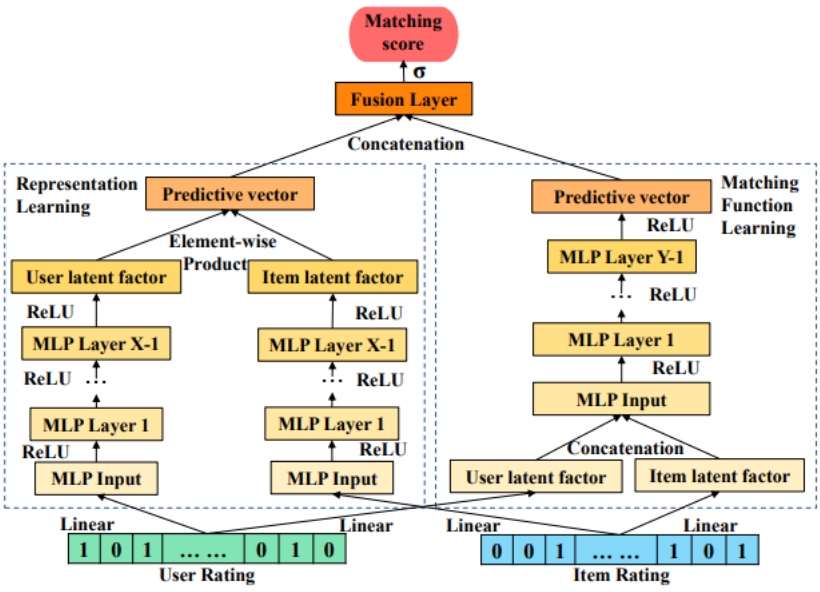
CHULALONGKORN UNIVERSITY



รูปที่ 6 โครงสร้างแบบจำลอง ConvNCF
[ที่มา: https://arxiv.org/pdf/1808.03912.pdf]

3.1.3. งานวิจัยของ Zhi-Hong, Deng และคณะ ผสมผสานวิธีการกรองการทำงานร่วมกันโดยอิงจากการเรียนรู้แทนค่าและการเรียนรู้ฟังก์ชันการจับคู่

งานวิจัย [11] ในปี 2019 นำเสนอแบบจำลอง DeepCF เป็นสถาปัตยกรรมที่ใช้เครือข่ายประสาทเทียมแบบสองเส้นทาง ดังรูปที่ 7 เพื่อแยกตัวประกอบเมทริกซ์การให้คะแนนและเรียนรู้การแทนค่า ได้แก่ เครือข่ายการเรียนรู้การเป็นตัวแทน (Representation Learning) และ เครือข่ายการเรียนรู้ฟังก์ชันการจับคู่ (Matching Function Learning) สำหรับเครือข่ายการเรียนรู้การเป็นตัวแทนจะพยายามนำข้อมูลผู้ใช้และวัตถุลงในพื้นที่การแสดงเดียวกัน ในกรณีนี้ ยิ่งการแสดงผู้ใช้และวัตถุมีความคล้ายคลังกันมากเท่าใด ระดับการจับคู่ก็จะสูงขึ้น และสำหรับเครือข่ายการเรียนรู้ฟังก์ชันการจับคู่จะพยายามเรียนรู้โดยตรงเพื่อจับคู่วัตถุผู้ใช้เข้ากับฟังก์ชันการจับคู่ที่ซับซ้อนของคะแนนที่ตรงกัน หลังจากนั้นจะทำการนำผลลัพธ์ของทั้งสองเครือข่ายรวมเข้าด้วยกันแบบการเชื่อมเชื่อมโยงเต็มรูปแบบ จากนั้นแบบจำลองจะทำการคำนวณคะแนนจับคู่ ของผู้ใช้กับแต่ละวัตถุ โดยงานวิจัยนี้ได้ทำการทดลองบนชุดข้อมูลทั่วไปที่หลากหลาย ได้แก่ ml-1m lastfm AMusic และ AToy สำหรับวิธีการแบ่งชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ จะนำข้อมูลการโต้ตอบครั้งสุดท้ายของแต่ละผู้ใช้เป็นชุดข้อมูลทดสอบ และชุดข้อมูลที่เหลือเป็นชุดข้อมูลฝึกฝน และวัดประสิทธิภาพของระบบแนะนำจากค่า HR@k และ NDCG@k จากผลการทดลองพบว่าแบบจำลอง DeepCF นี้สามารถเพิ่มประสิทธิภาพการทำงานแบบกรองร่วมได้ดียิ่งขึ้น



รูปที่ 7 โครงสร้างแบบจำลอง DeepCF

[ที่มา: https://ojs.aaai.org/index.php/AAAI/article/view/3769/3647]

3.2. งานวิจัยในกลุ่มแบบจำลองผสมเชิงลึก

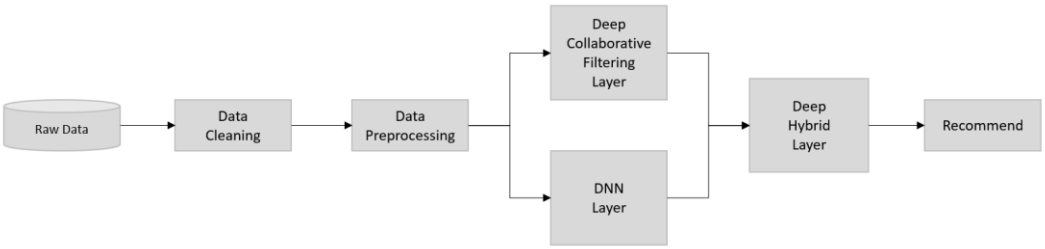
งานวิจัยของ Çakır, Muhammet และคณะ ในปี (2019) [12] นำเสนอแบบจำลองผสมเชิงลึกบนชุดข้อมูลการรับสมัครงานของเว็บไซต์หนึ่งในประเทศตุรกี ซึ่งแบบจำลองประกอบด้วยวิธีการแยกตัวประกอบ (Matrix Factorization) และโครงข่ายประสาทเทียมเชิงลึก (Deep Neural Network หรือ DNN) โดยวิธีการแยกตัวประกอบจะใช้เทคนิคการฝังตัว (Embedding) มีข้อมูลเข้าเป็นข้อมูลการโต้ตอบระหว่างผู้ใช้และวัตถุในการหาตัว

ประกอบแฝง (Latent factor) สำหรับโครงข่ายประสาทเทียมเชิงลึกจะถูกนำมาใช้ในการวิเคราะห์ข้อมูลเสริม เช่น คุณสมบัติของการประกาศรับสมัครงานและผู้สมัคร โดยข้อมูลเสริมนี้จะมีทั้งข้อมูลที่เป็นแบบต่อเนื่องและหมวดหมู่ สำหรับข้อมูลที่เป็นแบบต่อเนื่องจะถูกนำมาทำการนอร์มอไรซ์ (Normalize) ก่อนที่จะถูกนำเข้าสู่โครงข่ายประสาทเทียม จากนั้นข้อมูลเสริมทั้งหมดจะถูกนำเข้าสู่ชั้นรวมแบบต่อเนื่อง และในชั้นสุดท้ายข้อมูลตัวประกอบแฝงและข้อมูลเสริมจะถูกนำมาวิเคราะห์ร่วมกันเพื่อทำการแนะนำรายการตำแหน่งงานที่ได้รับการทำนายการให้คะแนนมากที่สุดสิบอันดับ (Top-10 Recommend) ให้แก่ผู้สมัคร และแนะนำนำผู้สมัครที่ได้รับการทำนายการให้คะแนนมากที่สุดสิบอันดับให้แก่ผู้รับสมัครงาน จากผลการทดลองพบว่าแบบจำลองนี้มีประสิทธิภาพการตี (Hit Ratio) ในการแนะนำตำแหน่งงานให้แก่ผู้สมัครในสิบอันดับแรก (HR@10) ถึง 85.5% และประสิทธิภาพการตี ในการแนะนำผู้สมัครให้แก่ตำแหน่งงานที่เปิดรับในสิบอันดับแรกถึง 79.5%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 4
แนวคิดและวิธีการดำเนินงาน

งานวิจัยนี้นำเสนอการพัฒนาแบบจำลองระบบแนะนำร้านอาหารในประเทศไทยบนชุดข้อมูลภาษาไทย ที่มีความเบาบางของข้อมูลการให้คะแนนเป็นจำนวนมาก และข้อความวิจารณ์ภาษาไทยที่มีรูปแบบและโครงสร้างทางภาษาที่ซับซ้อนและแตกต่างจากภาษาอื่น ๆ ให้ได้ผลลัพธ์ที่มีประสิทธิภาพมากขึ้น และเนื่องจากชุดข้อมูลที่มีจำกัด จึงมีแนวคิดที่จะนำข้อมูลต่าง ๆ ที่มีในชุดข้อมูลมาผ่านขั้นตอนการทำวิศวกรรมข้อมูล เช่น การสกัดคุณลักษณะของผู้ใช้บริการและร้านอาหารจากข้อความวิจารณ์ หรือ การทำข้อมูลสถิติต่าง ๆ และทำการเลือกคุณลักษณะที่มีผลต่อประสิทธิภาพของแบบจำลอง เพื่อพัฒนาแบบจำลองให้มีประสิทธิภาพมากยิ่งขึ้น



รูปที่ 8 แนวคิดแบบจำลอง

โดยแบ่งอธิบายแนวคิดของแบบจำลองแต่ละแบบดังต่อไปนี้

4.1. แบบจำลองแบบพึ่งพาผู้ใช้ร่วม

สำหรับแบบจำลองนี้นำแนวคิดจากงานวิจัย [11] มาทำการพัฒนา โดยทำการสร้างแบบจำลอง Collaborative filtering แบบใช้ข้อมูลโดยอ้อม (Implicit interaction)

แบบจำลองจะทำการฝึกฝนด้วยการใช้ข้อมูลแบบ implicit กล่าวคือ จากชุดข้อมูลที่ผู้ใช้บริการมีการให้คะแนนแต่ละวัตถุ แบบจำลองจะไม่สนใจว่าผู้ใช้บริการให้คะแนนเท่าไหร่ แต่จะสนใจแค่ว่าผู้ใช้บริการคนนั้น ๆ เคยให้คะแนนหรือไม่ ซึ่งอาจจะหมายถึงผู้ใช้บริการเคยไปใช้บริการหรือซื้อวัตถุนั้น ๆ โดยก่อนที่จะนำข้อมูลไปทำการฝึกฝน จะทำการแปลงข้อมูลจากข้อมูลการให้คะแนนเป็นคลาส 1 หรือ 0 ก่อน โดยที่ 1 หมายถึง ผู้ใช้บริการเคยไปใช้บริการหรือซื้อวัตถุและทำการให้คะแนนกับวัตถุ และ 0 หมายถึง ผู้ใช้บริการไม่เคยไปใช้บริการหรือซื้อวัตถุและก็ไม่เคยให้คะแนนเลย โดยที่แบบจำลองนี้จะทำการฝึกฝนแบบจำลองแบบหมวดหมู่ (Classification) และจะทำการทำนายค่าความน่าจะเป็นให้กับแต่ละวัตถุที่ผู้ใช้บริการไม่เคยให้คะแนนว่าผู้ใช้บริการมีความน่าจะเป็นว่าจะไปใช้บริการหรือซื้อวัตถุไหนมากที่สุด 10 อันดับแรก

4.2. แบบจำลองแบบผสม

แนวคิดของแบบจำลองนี้ได้มาจากงานวิจัย [12] ที่ได้มีการนำข้อมูลคุณลักษณะของผู้สมัครงานและตำแหน่งงานที่เปิดรับสมัครมาใช้เป็นข้อมูลคุณลักษณะให้กับแบบจำลอง สำหรับงานวิจัยนี้จึงได้มีแนวคิดที่จะนำคุณลักษณะของผู้ใช้บริการและร้านอาหารมาเป็นข้อมูลขาเข้าให้กับแบบจำลอง เพื่อใช้ในการฝึกฝนแบบจำลอง โดยจะใช้ทั้ง

ข้อมูลคุณลักษณะของร้านอาหารที่มีอยู่แล้ว ร่วมกับข้อมูลคุณลักษณะอื่น ๆ ที่จะทำการสร้างขึ้นมาจากขั้นตอนการทำวิศวกรรมข้อมูล (Feature Engineering) ดังนี้

4.2.1. ข้อมูลคุณลักษณะของร้านอาหารที่มีอยู่แล้ว เช่น ประเภทร้านอาหาร ระดับราคา พื้นที่สำหรับจอดรถ บริการส่ง เป็นต้น

4.2.2. ทำการสกัดคุณลักษณะของผู้ใช้บริการและร้านอาหารจากข้อความบทวิจารณ์ เนื่องจากมีแนวคิดว่าข้อความวิจารณ์ที่ผู้ใช้บริการได้เขียนถึงร้านอาหารนั้น น่าจะมีข้อมูลที่เฉพาะเจาะจงที่พูดถึงร้านอาหารนั้น ๆ และอาจจะยังสามารถแสดงถึงความสนใจในด้านต่าง ๆ ของผู้ใช้บริการได้อีกด้วย เช่น ผู้ใช้บริการบางคนอาจจะสนใจในเรื่องของการให้บริการ แต่บางคนก็อาจจะสนใจในเรื่องของราคา เป็นต้น และเนื่องจากข้อความวิจารณ์นั้นมีทั้งในด้านบวกและด้านลบ จึงมีวิธีการในการสกัดดังนี้

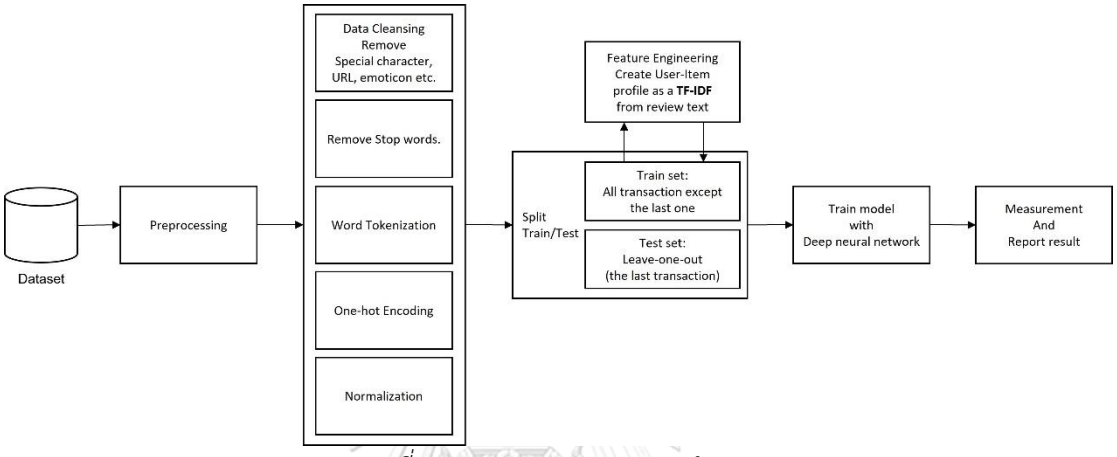
1. นำแต่ละข้อความมาทำการสกัดคำด้วยวิธีการ TF-IDF
2. นำค่า TF-IDF ของแต่ละข้อความที่ได้มาทำการรวมหรือเฉลี่ยให้กับผู้ใช้บริการและร้านอาหารนั้นๆ
3. ทำการคัดเลือกคุณลักษณะ (Feature selection) เนื่องจากคำสำคัญที่ได้จากการสกัดจากวิธีการ TF-IDF นั้นมีจำนวนมาก และบางคำอาจจะไม่มีผลต่อการพิจารณาเกี่ยวกับธุรกิจร้านอาหาร จึงจะได้หาวิธีการในการคัดเลือกคุณลักษณะให้ดียิ่งขึ้น

4.2.3. ทำการสร้างข้อมูลคุณลักษณะจากความรู้พื้นฐาน (Knowledge based) ที่มีทางด้านธุรกิจร้านอาหาร เพื่อนำมาเป็นข้อมูลคุณลักษณะขาเข้าให้กับแบบจำลอง ได้แก่ ข้อมูลคุณลักษณะที่บอกถึงจำนวนครั้งของการให้คะแนนของแต่ละร้านอาหารในช่วงเวลาหนึ่ง (temporal window) เนื่องจากถ้าหากพิจารณาจากธรรมชาติของผู้ใช้บริการร้านอาหารแล้ว ในบางครั้งผู้ใช้บริการอาจจะไปร้านอาหารนั้น ๆ เนื่องจากเพราะเป็นร้านอาหารเปิดใหม่หรืออยู่ในกระแส จึงทำให้ผู้ใช้บริการตัดสินใจไปใช้บริการร้านอาหารนั้น ๆ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 5
การออกแบบและพัฒนาแบบจำลอง

หัวข้อนี้จะกล่าวถึงรายละเอียดของแต่ละขั้นตอนการทดลอง ดังรูปที่ 9 ได้แก่ ชุดข้อมูลที่ใช้ในการทดลอง วิธีการเตรียมข้อมูลก่อนนำมาใช้ วิธีการเตรียมชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบ วิธีการสร้างคุณลักษณะเสริมจากชุดข้อมูลที่มี โครงสร้างของแบบจำลอง วิธีการวัดผล และผลการทดลอง



รูปที่ 9 กระบวนการทดสอบแบบจำลอง

5.1. ชุดข้อมูล

ชุดข้อมูลที่นำมาใช้ในวิทยานิพนธ์นี้ เป็นข้อมูลการให้คะแนนและคำวิจารณ์ภาษาไทย ที่มีต่อร้านอาหารที่ผู้ใช้บริการเคยไปใช้บริการ โดยได้รับการสนับสนุนจาก บริษัท วงใน มีเดีย จำกัด ชุดข้อมูลการให้คะแนนและคำวิจารณ์ โดยจะแบ่งข้อมูลเป็นชุดข้อมูลฝึกสอน (Training data) และ ชุดข้อมูลทดสอบ (Test data) โดยชุดข้อมูลประกอบด้วย 2 ส่วน

5.1.1. ชุดข้อมูลร้านอาหาร ประกอบด้วยคุณลักษณะดังตารางที่ 1

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 1 ลักษณะข้อมูลในชุดข้อมูลร้านอาหาร

ชื่อคอลัมน์	คำอธิบาย
restaurant_id	รหัสประจำร้านอาหาร ตัวอย่าง ‘D7325BB085FEB5784B5FF0E8D7E3BA63858CE089’
restaurant_name	ชื่อร้านอาหาร ตัวอย่าง ‘OSAKA Buffet’ , ‘ชาตรามือ’
restaurant_type	ประเภทร้านอาหาร ตัวอย่าง ‘Japanese’ , ‘Cafe/Coffee Shop’
city	จังหวัดที่ตั้งของร้านอาหาร ตัวอย่าง ‘Bangkok’
area	พื้นที่ตั้งของร้านอาหาร ตัวอย่าง ‘Pathum Wan’ , ‘Don Mueang’

price	ระดับราคาของอาหาร มีทั้งหมด 5 ระดับ โดย ระดับ 1 หมายถึง ราคาอาหารต่ำกว่า 100 บาท และ ระดับ 5 หมายถึง ราคาอาหารตั้งแต่ 1000 บาทขึ้นไป
parking_car	คุณลักษณะที่บอกว่าร้านอาหารมีที่จอดรถหรือไม่ โดยที่ True หมายถึง มีที่จอดรถ False หมายถึง ไม่มีที่จอดรถ
delivery	คุณลักษณะที่บอกว่าร้านอาหารมีบริการส่งหรือไม่ โดยที่ True หมายถึง มีบริการส่ง False หมายถึง ไม่มีบริการส่ง
wifi	คุณลักษณะที่บอกว่าร้านอาหารมีบริการไวไฟหรือไม่ โดยที่ True หมายถึง มีไวไฟ False หมายถึง ไม่มีไวไฟ
alcohol	คุณลักษณะที่บอกว่าร้านอาหารมีเครื่องดื่มที่มีแอลกอฮอร์ขายหรือไม่ โดยที่ True หมายถึง มีเครื่องดื่มที่มีแอลกอฮอร์ขาย False หมายถึง ไม่มีเครื่องดื่มที่มีแอลกอฮอร์ขาย

5.1.2. ชุดข้อมูลการให้คะแนนและคำวิจารณ์ของผู้ใช้บริการ ประกอบด้วยคุณลักษณะดังตารางที่ 2

ตารางที่ 2 ลักษณะข้อมูลในชุดข้อมูลการให้คะแนนและคำวิจารณ์

ชื่อคอลัมน์	คำอธิบาย
user_id	รหัสประจำผู้ใช้บริการ ตัวอย่าง ‘00078584646DEC3D2B4A42064EE09895A427B59F’
restaurant_id	รหัสประจำร้านอาหาร
rating	คะแนน ที่ผู้ใช้บริการให้แก่ร้านอาหารนั้น ๆ ขนาดการให้คะแนนมีทั้งหมด 5 ระดับ 1-5 โดยที่ 1 หมายถึง พึงพอใจน้อยที่สุด และ 5 หมายถึง พึงพอใจมากที่สุด
review_text	คำวิจารณ์ ที่ผู้ใช้บริการเขียนแสดงความรู้สึกที่มีต่อร้านอาหารนั้น ๆ
reviewed_at	วันที่และเวลา ที่ผู้ใช้บริการให้คะแนนและเขียนคำวิจารณ์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.2. การทำความสะอาดข้อมูล

5.2.1. การจัดการข้อมูลซ้ำ

จากการตรวจสอบชุดข้อมูลร้านอาหารพบว่า บางรหัสประจำร้านอาหารที่ต่างกัน มีชื่อร้านอาหารและคุณลักษณะที่เหมือนกันทุกประการ แต่ต่างกันที่ตัวอักษรเล็กหรือใหญ่ในชื่อร้านอาหาร ดังรูปที่ 10

restaurant_id	restaurant_name	restaurant_type	branch	city	area	price	parking_car	delivery	wifi	alcohol
A9F7CB7724B13837F76BF6A9FD44972D2AC85B3E	JOHA Korean Restaurant	Korean	NaN	Bangkok	Phayathai	3.0	False	NaN	False	True
A000A135AC5572B333DD3885BEC5C4A7F3543B47	Joha Korean Restaurant	Korean	NaN	Bangkok	Phayathai	3.0	False	NaN	False	True

รูปที่ 10 ข้อมูลร้านอาหารที่มีข้อมูลซ้ำซ้อน

ดังนั้นจึงทำการจัดการข้อมูลซ้ำด้วยการแปลงข้อมูลชื่อร้านอาหารให้เป็นตัวอักษรเล็กเหมือนกันทั้งหมด และนำรหัสประจำร้านอาหารทั้ง 2 รหัสไปทำการตรวจสอบข้อมูลการให้คะแนนในชุดข้อมูลการให้คะแนน เพื่อตรวจสอบว่ารหัสประจำร้านอาหารอันไหนมีจำนวนการให้คะแนนจากผู้ใช้บริการมากที่สุด แล้วรหัสประจำร้านอาหารนั้นจะถูกเก็บไว้ และจะนำไปแทนที่อีกรหัสประจำร้านอาหารหนึ่งที่มีจำนวนการให้คะแนนจากผู้ใช้บริการที่น้อยกว่าในชุดข้อมูลการให้คะแนน เพื่อเป็นการลดข้อมูลซ้ำซ้อนในชุดข้อมูลร้านอาหาร

5.2.2. การจัดการข้อมูลผิดปกติ

จากการตรวจสอบชุดข้อมูลการให้คะแนนและคำวิจารณ์ พบว่า มีผู้ใช้บริการส่วนหนึ่งมีข้อมูลการให้คะแนนและคำวิจารณ์มากผิดปกติ (Outlier) เนื่องจากผู้ใช้บริการเหล่านี้มีอาชีพเป็นผู้เขียนบทวิจารณ์ร้านอาหารต่าง ๆ ซึ่งในบางครั้งที่ผู้ใช้บริการเหล่านี้ไปใช้บริการร้านอาหาร อาจจะเป็นการถูกจ้างให้ไปโดยที่ไม่ใช่ความต้องการที่แท้จริงของผู้ใช้บริการเหล่านี้ ดังนั้น เพื่อไม่ให้ข้อมูลมีความลำเอียง (Bias) และเหลือเพียงข้อมูลจากผู้ใช้บริการที่แท้จริง ข้อมูลรายการให้คะแนนและคำวิจารณ์ในลักษณะนี้จึงจะถูกกำจัดออกจากชุดข้อมูล

5.2.3. การจัดการค่าว่าง

จากการตรวจสอบชุดข้อมูลร้านอาหาร พบบางคอลัมน์ที่มีค่าว่าง ซึ่งแต่ละคอลัมน์มีชนิดของข้อมูลที่แตกต่างกัน จึงมีการจัดการค่าว่างที่ต่างกันดังนี้

1. คอลัมน์ที่มีชนิดข้อมูลแบบสายอักขระ (String) จะถูกจัดการค่าว่าง โดยการแทนค่าว่างด้วย ‘unknown’
2. คอลัมน์ที่มีชนิดข้อมูลแบบตรรกะ (Boolean) จะถูกจัดการค่าว่างโดยการแทนค่าว่างด้วย ตรรกะ False

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.2.4. การนำอักษรที่ไม่เกี่ยวข้องออก

การกำจัดอักษรหรือสัญลักษณ์ที่ไม่สำคัญออกจากคำวิจารณ์เพื่อเพิ่มประสิทธิภาพต่อการแยกคุณลักษณะของร้านอาหารและผู้ใช้บริการ เนื่องจากชุดข้อมูลร้านอาหารและชุดข้อมูลการให้คะแนนนั้นไม่มีข้อมูลคุณลักษณะของร้านอาหารและผู้ใช้บริการที่เพียงพอต่อการทำแบบจำลอง

อักษรหรือสัญลักษณ์ที่ไม่สำคัญต่อการแยกคุณลักษณะ เช่น ตัวอักษรพิเศษ (Special characters), ตัวเลข (Numbers), อีโมจิ (Emoji), ยูอาร์แอล (URL) ดังแสดงตัวอย่างในตารางที่ 3

ตารางที่ 3 ตัวอย่างอักษรและสัญลักษณ์ที่ไม่เกี่ยวข้อง

ชนิดตัวอักษรและสัญลักษณ์ที่ไม่เกี่ยวข้อง	ตัวอย่างของตัวอักษรและสัญลักษณ์ที่ไม่เกี่ยวข้อง
ตัวอักษรพิเศษ	() # ! , . $ & \ / ?
ตัวเลข	0 1 2 3
อีโมจิ	
ยูอาร์แอล	http://www.facebook.com/NewKungPaChim

5.2.5. การตัดคำภาษาไทย

เนื่องจากประโยคในภาษาไทยไม่มีการแบ่งคำที่ชัดเจน จึงจำเป็นต้องมีขั้นตอนการตัดแยกคำจากประโยค (Word Segmentation) ซึ่งในงานวิจัยนี้เลือกใช้เครื่องมือ word tokenizer จากไลบรารี่ PyThaiNLP และเนื่องจากข้อความวิจารณ์เกี่ยวกับร้านอาหารนั้นค่อนข้างมีคำเฉพาะเกี่ยวกับอาหารเป็นส่วนมาก ในงานวิจัยนี้จึงได้ทำการสร้างพจนานุกรมขึ้นมาเอง โดยใช้คำภาษาไทยทั่วไปจากพจนานุกรม thai_words จากไลบรารี่ PyThaiNLP และพจนานุกรมเกี่ยวกับอาหารโดยเฉพาะซึ่งเผยแพร่โดยบริษัท วงใน มีเดีย จำกัด (https://github.com/wongnai/wongnai-corpus/blob/master/search/food_dictionary.txt) ตัวอย่างคำใน food dictionary ดังรูปที่ 11

โรตีทุเรียน
ตำร้อยเอ็ด
ส้นคอย่างแจ่ว
เกี๊ยวกุ้งทอดครีมสลัด
ส้มตำปูนึ่ง
สายไหมเย็น
บะหมี่ปูหมูแดง
เมนูเด็ก
น้ำเต้าหู้ใส่เนื้อเฉาก๊วยทรงเครื่อง
ส้มตำกุ้งสดสาหร่าย
ปลาลิ้นหมาทอดน้ำปลา
ข้าวมันไก่ทอดไรซ์เบอรี่
ไอติมมันม่วง
เนื้อปลาเก๋าต้มผักเสฉวน
archino
อาหารที่ตักมาค่ะ
โกโกสูตรเข้มข้น
มอคค่าเฟรปปูชิโน่
ซูชิเป็ดรมควัน
ไส้มะพร้าวอ่อน
ซัลม่อนแซ่บ
ชอคโกแลตครีมชิพปั่น
แกงเนื้อออสเตเลีย
หมึกสดชิ้นใหญ่
สปาเก็ตตี้ผงกระหรี่

รูปที่ 11 ตัวอย่างคำในพจนานุกรมคำศัพท์เกี่ยวกับร้านอาหาร

5.2.6. การเข้ารหัสข้อมูล

การเข้ารหัสข้อมูล (One-hot Encoding) เป็นขั้นตอนหนึ่งในการเตรียมชุดข้อมูลให้พร้อมสำหรับการเป็นข้อมูลขาเข้าของแบบจำลอง โดยจะทำการแปลงข้อมูลที่มีลักษณะแบบคุณลักษณะ (Categorical) ให้เป็นข้อมูล

ตัวเลข (Numerical) คือ ตัวเลข 0 หรือ 1 ตามค่าของข้อมูล เช่น คุณลักษณะบริการที่จอดรถ สำหรับค่า True จะถูกแปลงเป็นเลข 1 และค่า False จะถูกแปลงเป็นเลข 0

5.2.7. การปรับช่วงขอบเขตข้อมูล

เนื่องจากข้อมูลมีความหลากหลาย ทั้งชนิดข้อมูล รูปแบบข้อมูล และขอบเขตของข้อมูล จึงจำเป็นต้องทำการปรับช่วงขอบเขตข้อมูล (Normalization) ก่อนที่จะป้อนข้อมูลให้กับแบบจำลอง โดยในงานนี้จะใช้วิธีการปรับขอบเขตแบบ Min-Max ที่จะปรับช่วงข้อมูลให้อยู่ในช่วง 0 ถึง 1 ดังสมการ (14)

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{14}$$

5.3. การเตรียมชุดข้อมูลสำหรับการทดลอง

สำหรับชุดข้อมูลที่ใช้ในการทดลองงานวิจัยนี้ ได้ทำการกรองข้อมูลมาจากผู้ใช้บริการที่มีการให้คะแนนจำนวนอย่างน้อย 20 รายการขึ้นไป และจะต้องเป็นผู้ใช้บริการที่มีการให้คะแนนและเขียนวิจารณ์ให้กับร้านอาหารจำนวนอย่างน้อย 5 ร้านอาหารขึ้นไป โดยมีลักษณะข้อมูลดังตารางที่ 4

ตารางที่ 4 คุณลักษณะชุดข้อมูล

Dataset	# Rating	# User	# Restaurant
Rating	158116	2017	4271

จากนั้นทำการเตรียมชุดข้อมูลฝึกฝนและชุดข้อมูลทดสอบสำหรับการทดลอง ด้วยวิธีการ Leave-one-out คือ จะนำการโต้ตอบระหว่างผู้ใช้และร้านอาหารในครั้งล่าสุดของแต่ละผู้ใช้มาเป็นข้อมูลชุดทดสอบและข้อมูลที่เหลือทั้งหมดเป็นชุดข้อมูลฝึกฝน

5.3.1. ชุดข้อมูลทดสอบ (Test Dataset) มีวิธีการเตรียมชุดข้อมูลดังนี้

1. นำการโต้ตอบระหว่างผู้ใช้และร้านอาหารในครั้งล่าสุดของแต่ละผู้ใช้แยกออกมาจากชุดข้อมูลทั้งหมด
2. สร้างรายการร้านอาหารจำนวน 99 รายการ สำหรับแต่ละผู้ใช้บริการ โดยการสุ่มจากรายการร้านอาหารที่ผู้ใช้ไม่เคยไปใช้บริการในชุดข้อมูลฝึกฝน ต่อไปนี้จะเรียกว่า ชุดข้อมูลทดสอบที่ผู้ใช้บริการไม่เคยไป (test negative)
3. ทำการเพิ่มรายการร้านอาหารล่าสุดที่อยู่ในชุดข้อมูลทดสอบที่ได้จากข้อ 1. เข้าไปในชุดข้อมูล test negative รวมเป็น 100 รายการ โดยประกอบไปด้วย รายการร้านอาหารที่ผู้ใช้บริการไม่เคยไปใช้บริการจำนวน 99 รายการ และ ร้านอาหารที่ผู้ใช้ไปใช้บริการครั้งล่าสุดจำนวน 1 รายการ

5.3.2. ชุดข้อมูลตรวจสอบ (Validation Dataset) มีวิธีการเตรียมชุดข้อมูลดังนี้

1. แบ่งข้อมูลจำนวน 20% จากชุดข้อมูลทดสอบที่ได้เตรียมไว้จากหัวข้อ 5.3.1.

5.3.3. ชุดข้อมูลฝึกฝน (Training Dataset) มีวิธีการเตรียมชุดข้อมูลดังนี้

2. นำรายการโต้ตอบระหว่างผู้ใช้และร้านอาหารที่เหลือทั้งหมด ยกเว้นรายการล่าสุดนำมาเป็นชุดข้อมูลฝึกฝน
3. เนื่องจากงานวิจัยนี้เป็นการทดลองแบบใช้ข้อมูล implicit โดยไม่คำนึงถึงคะแนนการให้คะแนนของผู้ใช้บริการ แต่จะสนใจเพียงว่าผู้ใช้บริการเคยไปใช้บริการร้านอาหารหรือไม่ ถ้าหากผู้ใช้บริการเคยไปใช้บริการจะให้มีค่าเป็น 1 (positive) และสำหรับร้านอาหารที่ผู้ใช้ไม่เคยไปใช้บริการจะให้มีค่าเป็น 0 (negative)
4. ทำการแปลงข้อมูลจากชุดข้อมูลฝึกฝนทั้งหมดด้วยอัตราส่วน 1 positive ต่อ 4 negative โดย 4 negative รายการนั้นจะมาจากการสุ่มร้านอาหารที่ผู้ใช้บริการไม่เคยไปใช้บริการในชุดข้อมูลฝึกฝน ตัวอย่างดังรูปที่ 12

Userid	Restaurantid	Rating
0	2300	3
1	1986	3

→

Userid	Restaurantid	Visited
0	2300	1
1	1986	1

→

Userid	Restaurantid	Visited
0	**2300**	**1**
0	1	0
0	2	0
0	1986	0
0	4	0
1	**1986**	**1**
1	5	0
1	2300	0
1	7	0
1	8	0

รูปที่ 12 ตัวอย่างการเตรียมชุดข้อมูลฝึกฝน

5.4. การสร้างข้อมูลคุณลักษณะเสริม

เพื่อเพิ่มประสิทธิภาพให้แก่แบบจำลอง งานวิจัยนี้ได้ทำการสร้างข้อมูลคุณลักษณะเสริม 2 คุณลักษณะ ได้แก่

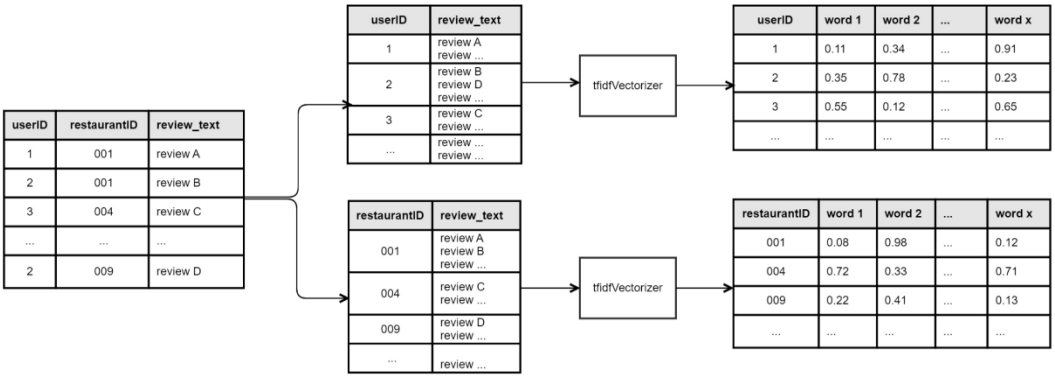
5.4.1. การสร้างคุณลักษณะเสริมจากคำวิจารณ์ด้วยวิธีการ tf-idf

เนื่องจากชุดข้อมูลที่ใช้ในงานวิจัยนั้นมีเพียงข้อมูลคุณลักษณะของร้านอาหาร แต่สำหรับข้อมูลของผู้ใช้บริการนั้นมีเพียงข้อมูลรหัสผู้ใช้บริการเท่านั้น ผู้วิจัยจึงได้ทำการดึงคุณลักษณะของผู้ใช้บริการจากข้อความวิจารณ์ที่ผู้ใช้บริการได้ทำการเขียนให้กับร้านอาหารที่เคยไปใช้บริการ เนื่องจากคำวิจารณ์เหล่านี้อาจจะสามารถบอกถึงความสนใจที่ผู้ใช้บริการมีต่อร้านอาหารและยังสามารถบอกถึงคุณลักษณะอื่น ๆ ของร้านอาหารที่นอกเหนือจากคุณลักษณะที่มีอยู่ในชุดข้อมูลอยู่แล้วอีกด้วย โดยจะทำการสร้างเวกเตอร์สถิติของคำสำหรับแต่ละผู้ใช้บริการและร้านอาหาร ดังรูปที่ 13 โดยมีวิธีการดังนี้

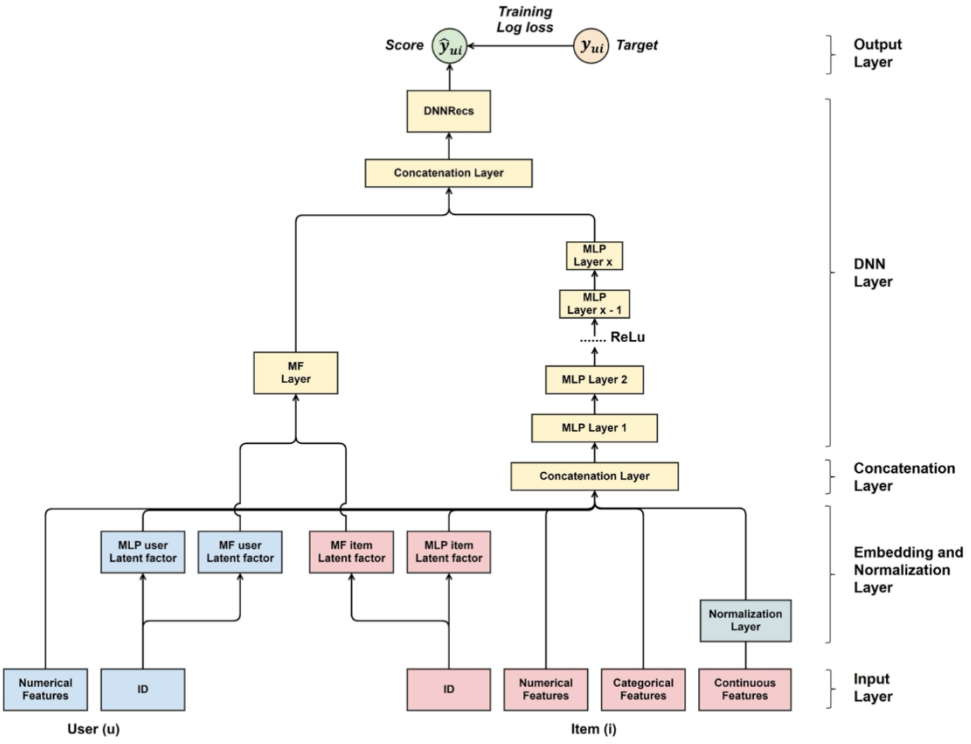
1. นำข้อความวิจารณ์ที่ผ่านการทำความสะอาดข้อมูลเรียบร้อยแล้ว มาทำการตัดคำโดยใช้ฟังก์ชัน word_tokenize จากไลบรารี่ PyThaiNLP ในการตัดคำ และได้ทำการสร้างพจนานุกรมเกี่ยวกับอาหารโดยเฉพาะ ดังที่กล่าวในหัวข้อ 5.2.5 เพื่อใช้ในการตัดคำ
2. นำคำที่ตัดได้ทั้งหมดของแต่ละผู้ใช้บริการมารวมกัน ในทางเดียวกัน นำคำที่ตัดได้ทั้งหมดของแต่ละร้านอาหารมารวมกัน
3. นำคำทั้งหมดที่ตัดได้ของแต่ละผู้ใช้บริการและร้านอาหารไปผ่านกระบวนการสร้างเวกเตอร์สถิติด้วยเทคนิค tf-idf เพื่อสร้างเป็นเวกเตอร์สถิติของคำสำหรับแต่ละผู้ใช้บริการและร้านอาหาร

5.4.2. การสร้างคุณลักษณะเสริมจากจำนวนการเข้าใช้บริการของแต่ละร้านอาหาร

สร้างคุณลักษณะเสริมจากจำนวนการเข้าใช้บริการของแต่ละร้านอาหาร โดยการนำจำนวนการให้คะแนนของร้านอาหารที่ผ่านมาใน 30 วันย้อนหลัง เพื่อใช้เป็นข้อมูลลักษณะเสริมอีกหนึ่งอย่างสำหรับร้านอาหาร ซึ่งสถิตินี้อาจตีความได้ว่า ร้านอาหารนั้น ๆ อาจจะกำลังได้รับความนิยมในช่วง 30 วันที่ผ่านมา จึงทำให้มีผู้ไปใช้บริการเป็นจำนวนมาก



รูปที่ 13 การสร้างข้อมูลคุณลักษณะเสริมจากคำวิจารณ์ด้วยเทคนิค tf-idf



รูปที่ 14 สถาปัตยกรรมแบบจำลอง

5.5. สถาปัตยกรรมแบบจำลอง

รูปที่ 14 เป็นสถาปัตยกรรมของแบบจำลองเรียนรู้การโต้ตอบระหว่างรายการผู้ใช้กับรายการร้านอาหาร และข้อมูลเสริมของผู้ใช้และร้านอาหาร โดยข้อมูลเสริมของร้านอาหาร หมายถึง ข้อมูลคุณลักษณะของร้านอาหาร ได้แก่

บริการที่จอดรถ การชำระเงินด้วยบัตรเครดิต บริการจัดส่ง สินค้าแอลกอฮอร์ และ บริการไวไฟ สำหรับข้อมูลเสริมของผู้ใช้บริการ ได้แก่ สถิติของคำในข้อความวิจารณ์ของผู้ใช้ที่ถูกสร้างขึ้นโดยใช้เทคนิค tf-idf ดังรูปที่ 13 เพื่อใช้เป็นคุณสมบัติของผู้ใช้บริการ

โดยแบบจำลองนี้จะประกอบไปด้วยเทคนิคการคำนวนสองเทคนิคเข้าด้วยกัน ได้แก่ เทคนิคการแยกตัวประกอบเมทริกซ์และเคอร์เนลเชิงเส้น เพื่อใช้ในการหาการคุณลักษณะแฝงของผู้ใช้และรายการร้านอาหาร ซึ่งจะได้ผลลัพธ์ออกมาในรูปแบบเวกเตอร์คุณลักษณะแฝง โดยมีข้อมูลเข้าเป็นรหัสผู้ใช้ และรหัสร้านอาหาร และอีกหนึ่งเทคนิค คือ การเรียนรู้เชิงลึกแบบเพอร์เซปตรอนหลายชั้น เพื่อแยกคุณลักษณะผู้ใช้และรายการ โดยมีข้อมูลเข้าเป็นรหัสผู้ใช้ รหัสร้านอาหาร และคุณลักษณะของทั้งผู้ใช้และร้านอาหาร และในขั้นตอนสุดท้าย จะใช้แบบจำลองโครงข่ายประสาทเทียมเชิงลึกในการทำนายความน่าจะเป็นของรายการร้านอาหารในชุดข้อมูลทดสอบ โดยแต่ละเทคนิคที่ใช้ในแบบจำลองนี้สามารถอธิบายโดยละเอียดได้ดังนี้

ทางด้านซ้าย (MF Layer) ของแบบจำลอง เป็นเทคนิคการแยกตัวประกอบเมทริกซ์ที่ให้ผลลัพธ์เป็นเวกเตอร์คุณลักษณะแฝงของผู้ใช้และร้านอาหาร โดยใช้ผลิตภัณฑ์การคูณ (dot product) เพื่อเรียนรู้เวกเตอร์แฝงของผู้ใช้และร้านอาหารในพื้นที่คุณลักษณะ เพื่ออธิบายการโต้ตอบระหว่างผู้ใช้และร้านอาหาร ด้วยฟังก์ชันกระตุ้นแบบไม่เชิงเส้น โดยมีข้อมูลเข้าเป็นรหัสผู้ใช้และรหัสร้านอาหาร

ทางด้านขวา (MLP Layer) ของแบบจำลอง ใช้เทคนิคเพอร์เซปตรอนหลายชั้น โดยมีการเพิ่มข้อมูลขาเข้าให้แก่แบบจำลองเพิ่มเติมเกี่ยวกับผู้ใช้และร้านอาหาร เช่น เวกเตอร์สถิติของคำในข้อความวิจารณ์ของผู้ใช้และร้านอาหาร คุณลักษณะของร้านอาหาร ซึ่งนอกเหนือจากรหัสของผู้ใช้และร้านอาหาร จากนั้น MLP จะประมวลผลคุณลักษณะผลลัพธ์ดังสมการที่ 9 โดยแบบจำลองจะเรียนรู้การโต้ตอบระหว่าง p_u และ q_i

ในขั้นตอนสุดท้าย ผลลัพธ์จากทั้งสองเทคนิคจะถูกรวมเข้าด้วยกันในชั้นเชื่อมโยงก่อนจะถูกนำเข้าโครงข่ายประสาทเทียมเชิงลึกชั้นสุดท้ายเพื่อคำนวณค่าความน่าจะเป็น ดังสมการที่ 15 ของแต่ละร้านอาหารในชุดข้อมูลทดสอบสำหรับแต่ละผู้ใช้บริการ

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

$$\phi^{MF} = p_u^G \odot q_i^G, \tag{15}$$

$$\phi^{MLP} = a_L(W_L^T\left(a_{L-1}\left(\ldots a_2\left(W_2^T\begin{bmatrix}p_u^M\\q_i^M\end{bmatrix}+b_2\right)\ldots\right)\right)+b_L),$$

$$\hat{y}_{ui} = \sigma\left(h^T\begin{bmatrix}\phi^{MF}\\\phi^{MLP}\end{bmatrix}\right)$$

โดยสรุป แบบจำลองจะทำนายความน่าจะเป็นของร้านอาหารแต่ละรายการ และจัดอันดับร้านอาหาร 10 อันดับแรกให้ผู้ใช้เป็นผลลัพธ์ของแบบจำลองนี้

5.6. ผลการทดลอง

ในหัวข้อนี้จะกล่าวถึง ค่าพารามิเตอร์ที่ใช้ในการทดลอง วิธีการวัดผล แบบจำลองพื้นฐาน และประสิทธิภาพของของแบบจำลองที่นำเสนอ ดังนี้

5.6.1. การกำหนดค่าพารามิเตอร์ในแบบจำลอง

ในการทดสอบแบบจำลอง ได้มีการปรับค่าพารามิเตอร์ดังนี้ learning rate ในช่วง [0.00001,0.0001,0.001], batch size ในช่วง [32,64,128,256] และ latent factor ในช่วง [8,16,32,64]

5.6.2. วิธีการวัดผล

แบบจำลองจะทำการทำนายค่าความน่าจะเป็นให้กับรายการร้านอาหารจำนวน 100 รายการ ที่ถูกเตรียมไว้ในหัวข้อ 5.3.1 และจะแนะนำร้านอาหารที่มีค่าความน่าจะเป็นใน 10 อันดับแรกให้กับแต่ละผู้ใช้ โดยเมทริกซ์ที่ใช้ในการวัดผลแบบจำลองนั้นประกอบด้วย 2 เมทริกซ์ ได้แก่ อัตราตี (HR@10) และค่าปรับลดอันดับสะสมปกติ (NDCG@10)

5.6.3. แบบจำลองพื้นฐาน

แบบจำลองพื้นฐานที่ใช้ในการเปรียบเทียบประสิทธิภาพของแบบจำลองที่นำเสนอนี้ มีจำนวนทั้งหมด 3 แบบจำลอง ได้แก่

1. GMF เป็นแบบจำลองที่ใช้เคอร์เนลเชิงเส้นเพื่อจำลองการโต้ตอบของผู้ใช้และรายการในลักษณะเดียวกับที่ vanilla MF ทำ
2. MLP ใช้เพอร์เซปตรอนหลายชั้นในการคำนวณค่าโต้ตอบแบบไม่เป็นเชิงเส้นระหว่างผู้ใช้และร้านอาหาร และใช้ข้อมูลขาเข้าเพียงรหัสผู้ใช้และรหัสร้านอาหาร
3. NueMF เป็นแบบจำลองที่รวมเทคนิค GMF และ MLP เข้าด้วยกัน

5.6.4. ประสิทธิภาพของแบบจำลอง

จากผลการเปรียบเทียบประสิทธิภาพแบบจำลองในตารางที่ 5 ถึง 10 สามารถสรุปข้อสังเกตที่สำคัญได้ดังนี้

1. จากผลการทดลองในตารางที่ 5 พบว่า แบบจำลองที่นำเสนอมีประสิทธิภาพเหนือกว่าแบบจำลองพื้นฐานอย่างมีนัยสำคัญ โดยเฉพาะจำนวน latent factor ที่เพิ่มขึ้นยิ่งส่งผลให้ประสิทธิภาพของแบบจำลองดีขึ้นเช่นกัน สิ่งนี้แสดงให้เห็นถึงประสิทธิภาพของข้อเสนอว่าข้อมูลเสริมถูกใช้เป็นข้อมูลเข้าและทำให้แบบจำลองมีประสิทธิภาพมากขึ้น
2. จากผลการทดลองในตารางที่ 6 พบว่า แบบจำลองที่มีการฝึกอบรมต่อจากแบบจำลองย่อยที่ได้ผลที่ดีที่สุดก่อนหน้า (Continue Training) จะมีประสิทธิภาพดีกว่าแบบจำลองที่ต้องฝึกอบรมตั้งแต่ต้น (Beginning Training) ผลลัพธ์นี้แสดงให้เห็นถึงศักยภาพของการฝึกอบรมล่วงหน้า ซึ่งทำให้แน่ใจได้ว่า การแยกตัวประกอบเมทริกซ์และเพอร์เซปตรอนหลายชั้นเรียนรู้คุณลักษณะจากหลาย ๆ มิติ และทำให้แบบจำลองมีประสิทธิภาพที่ดีที่สุดได้
3. จากตัวอย่างผลการแนะนำร้านอาหาร 10 อันดับแรกในตารางที่ 8 พบว่า แบบจำลองสามารถแนะนำร้านอาหารในลำดับที่ 1 ได้ถูกต้องตรงกับชุดข้อมูลทดสอบสำหรับผู้ใช้บริการหมายเลข 12 และยังสามารถแนะนำร้านอาหารที่อยู่ในหมวดหมู่เดียวกับร้านอาหารในชุดข้อมูลทดสอบได้ถึง 50%
4. จากตัวอย่างผลการแนะนำร้านอาหาร 10 อันดับแรกในตารางที่ 9 พบว่า แบบจำลองไม่สามารถแนะนำร้านอาหารใน 10 อันดับแรกได้ถูกต้องตรงกับชุดข้อมูลทดสอบสำหรับผู้ใช้บริการหมายเลข 9 แต่ในขณะเดียวกันพบว่า แบบจำลองสามารถแนะนำร้านอาหารที่อยู่ในหมวดหมู่เดียวกับร้านอาหารในชุดข้อมูลทดสอบได้ถึง 60%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5. จากผลการทดลองในตารางที่ 10 พบว่า แบบจำลองมีประสิทธิภาพที่ดีมากขึ้น เมื่อค่า K หรือจำนวนร้านอาหารที่แบบจำลองทำการแนะนำนั้นมีค่ามากขึ้น ในทางกลับกัน แบบจำลองจะมีประสิทธภาพลดลงเมื่อค่า K หรือจำนวนร้านอาหารที่แบบจำลองต้องทำการแนะนำมีค่าลดลง

5.6.5. ค่าพารามิเตอร์ที่มีผลต่อประสิทธิภาพของแบบจำลอง

1. จำนวนของ Latent factor ผลจากชุดข้อมูลทดสอบ พบว่าแบบจำลองมีประสิทธิภาพที่ดีที่สุดเมื่อ latent factor มีจำนวน 32 จากตาราง 5 และ 6 จะพบว่าแบบจำลองจะยิ่งมีประสิทธิภาพมากยิ่งขึ้นเมื่อจำนวน latent factor มีจำนวนเพิ่มมากขึ้น

2. จำนวนของ Hidden layer นอกจากจำนวนของ latent factor แล้ว จำนวน hidden layer ก็อาจจะเป็นพารามิเตอร์ที่สำคัญอีกตัวหนึ่งเช่นกัน จากผลลัพธ์ตารางที่ 7 แบบจำลองที่นำเสนอให้ผลลัพธ์ที่ดีที่สุดเมื่อ hidden layer มีจำนวน 4 ชั้น

ตารางที่ 5 เปรียบเทียบประสิทธิภาพของแบบจำลองที่นำเสนอต่อแบบจำลองพื้นฐานด้วยเมทริกซ์ HR@10 และ NDCG@10

Latent Factors	Metrics	Baseline methods			Proposal method		% Improvement of NeuMF vs. DNNRecs
		GMF	MLP	NeuMF	Deep Content (MLP with aux)	DNNRecs	
8	HR	0.4056	0.4209	0.4145	0.4765	**0.4864**	7.2%
	NDCG	0.2315	0.2361	0.2370	**0.3104**	0.3085	7.2%
16	HR	0.4209	0.4170	0.4165	0.4789	**0.4829**	7.4%
	NDCG	0.2437	0.2349	0.2296	**0.3168**	0.3083	7.9%
32	HR	0.4075	0.4318	0.4363	0.4796	**0.5166**	12.3%
	NDCG	0.2286	0.2383	0.2443	0.3164	**0.3351**	11.7%
64	HR	0.4018	0.4184	0.4298	0.4814	**0.5151**	15.8%
	NDCG	0.1966	0.2431	0.2291	0.3157	**0.3382**	14.2%

จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 6 เปรียบเทียบประสิทธิภาพของแบบจำลองที่นำเสนอระหว่างแบบจำลองที่มีการฝึกอบรมจากแบบจำลองย่อยก่อนหน้าต่อแบบจำลองที่เริ่มการฝึกอบรมตั้งแต่ต้นด้วยเมทริกซ์ HR@10 และ NDCG@10

Latent Factors	Metrics	Proposal method		% Improvement
		Beginning Training	Continue Training	
8	HR	0.4864	**0.4968**	1.0%
	NDCG	0.3085	**0.3141**	0.6%
16	HR	0.4829	**0.5027**	2.0%
	NDCG	0.3083	**0.3257**	1.7%
32	HR	0.5166	**0.5300**	1.3%
	NDCG	0.3351	**0.3467**	1.2%
64	HR	0.5151	**0.5181**	0.3%
	NDCG	0.3382	**0.3426**	0.4%

ตารางที่ 7 ผลการทดลองที่มีความไวต่อ hyper-parameter ระหว่าง latent factor และ hidden layer ด้วยเมทริกซ์ HR@10

Hidden Layers	Latent factors			
	8	16	32	64
1	0.4512	0.4655	0.4700	0.4522
2	0.4626	0.4765	0.4631	**0.5171**
3	0.4626	0.4829	**0.5151**	**0.5181**
4	0.4968	**0.5027**	**0.5300**	**0.5121**

ตารางที่ 8 ตัวอย่างผลการแนะนำร้านอาหาร 10 อันดับแรก เมื่อ HR@10 เท่ากับ 1

Test dataset		Top 10 Recommended Result		
UserID	RestaurantID	RestaurantID	Restaurant Name	Restaurant Category
12	3345	**3345**	**letao**	**bakery,cake,dessert**
		2114	heekcaa	coffee,juice,soft drink,tea
		120	**rapl apple pie**	**bakery,cake**
		3473	four seasons	chinese
		3912	**harrods**	**bakery,cake,dessert,ice cream**
		3892	**ไอติมหม้อไฟยศเส**	**ice cream**
		960	burger & lobster	seafood
		4091	**yamagoya ramen**	**japanese,ramen**
		2152	uno mas	international
		1311	**ขนมไทยเสน่ห์ คาเฟ่ ย่านบางลำภู**	**dessert**
***** Restaurant's detail of test dataset [ID, Name, Category] = [3345, "letao", "bakery,cake,dessert"]**				

ตารางที่ 9 ตัวอย่างผลการแนะนำร้านอาหาร 10 อันดับแรก เมื่อ HR@10 เท่ากับ 0

Test dataset		Top 10 Recommended Result		
UserID	RestaurantID	RestaurantID	Restaurant Name	Restaurant Category
9	1288	**308**	**ก๋วยเตี๋ยวคั่วไก่**	**rice dish**
		2596	**ปลาอยู่เย็น**	**thai**
		3363	indian food	indians
		3241	ก๋วยเตี๋ยวเรืออยุธยา เจ๊จุ่ม	noodles
		1440	**แสนยอดโภชนา**	**thai**
		122	**กินเฮง ข้าวมันไก่ตอนสูตรโบราณ**	**rice dish**
		2058	ร่ำรวยโภชนา	chinese
		3681	baskin robbins	ice cream
		2464	**ส้มตำเจ๊แดง**	**north east,thai**
		345	**ขาหมูตรอกซุง**	**rice dish**
***** Restaurant's detail of test dataset [ID, Name, Category] = [1288, "ข้าวแกงแม่อ้อน สะพานหัน", "rice dish,thai"]**				

ตารางที่ 10 เปรียบเทียบประสิทธิภาพของแบบจำลองเมื่อจำนวนร้านอาหารที่แนะนำต่างกัน

K	HR@K	NDCG@K
5	0.4963	0.3288
10	0.5300	0.3467
15	0.6039	0.3663
20	0.6634	0.3806

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 6
สรุปผลการวิจัย

6.1. สรุปผลการวิจัย

งานวิจัยนี้ได้ทำการสำรวจประสิทธิภาพของการเรียนรู้เชิงลึกสำหรับการเรียนรู้การโต้ตอบระหว่างผู้ใช้และรายการร้านอาหาร ตลอดจนความเป็นไปได้ของข้อมูลเพิ่มเติมของผู้ใช้และร้านอาหาร อีกทั้งได้มีการสร้างคุณลักษณะใหม่จากข้อความวิจารณ์เพื่อใช้เป็นข้อมูลเข้าให้กับแบบจำลองอีกด้วย แบบจำลอง DNNRecs ประกอบด้วยการแยกตัวประกอบของเมทริกซ์และเพอร์เซปตรอนหลายชั้น พร้อมข้อมูลการให้คะแนนและข้อความวิจารณ์ร้านอาหารบนชุดข้อมูลจริงในประเทศไทย มีการสร้างคุณลักษณะของผู้ใช้และร้านอาหารจากข้อความวิจารณ์โดยการแปลงข้อความเป็นข้อมูลตัวเลขด้วยเทคนิค tf-idf และงานวิจัยที่นำเสนอนี้ได้แสดงถึงความสำคัญของข้อมูลเสริมที่สามารถนำไปใช้ในการปรับปรุงและพัฒนาประสิทธิภาพของแบบจำลองให้ดียิ่งขึ้น รวมทั้งการปรับค่าพารามิเตอร์ต่าง ๆ ยังสามารถเพิ่มประสิทธิภาพของแบบจำลองให้มากยิ่งขึ้นได้อีกด้วย

จากผลการเปรียบเทียบประสิทธิภาพแบบจำลองในตารางที่ 5 พบว่า แบบจำลองที่นำเสนอมีประสิทธิภาพเหนือกว่าแบบจำลองพื้นฐานอย่างมีนัยสำคัญ โดยแบบจำลองที่นำเสนอเป็นการรวมเทคนิคการแยกตัวประกอบเมทริกซ์และเพอร์เซปตรอนหลายชั้นเข้าด้วยกัน รวมถึงการนำข้อมูลเสริมมาเป็นข้อมูลเข้าให้แก่แบบจำลอง สิ่งนี้แสดงให้เห็นถึงประสิทธิภาพของข้อเสนอว่า การใช้ข้อมูลเสริมมาเป็นข้อมูลเข้าให้แก่แบบจำลองทำให้แบบจำลองมีประสิทธิภาพมากขึ้น

จากผลการเปรียบเทียบประสิทธิภาพแบบจำลองในตารางที่ 6 พบว่า การนำแบบจำลองย่อยที่มีการฝึกอบรมก่อนหน้ามาเป็นค่าเริ่มต้นให้กับแบบจำลองสุดท้ายแล้วทำการฝึกอบรมต่อนั้นสามารถทำให้แบบจำลองมีประสิทธิภาพดีขึ้นมากกว่าการฝึกอบรมแบบจำลองย่อยตั้งแต่ต้น เนื่องจากการแยกตัวประกอบเมทริกซ์และเพอร์เซปตรอนหลายชั้นเรียนรู้คุณลักษณะจากหลาย ๆ มิติ จนทำให้แบบจำลองมีประสิทธิภาพที่ดียิ่งขึ้น

จากผลการทดลองในตารางที่ 5 และ 7 พบว่า จำนวนของ latent factor และ hidden layer นั้นมีผลต่อประสิทธิภาพของแบบจำลองอย่างมีนัยสำคัญ โดยที่แบบจำลองจะมีประสิทธิภาพมากยิ่งขึ้นเมื่อ latent factor และ hidden layer มีจำนวนเพิ่มมากขึ้น

และจากตัวอย่างผลการแนะนำร้านอาหารใน 10 อันดับแรก พบว่า แบบจำลองสามารถแนะนำร้านอาหารที่อยู่ในหมวดหมู่เดียวกับร้านอาหารในชุดข้อมูลทดสอบได้ถึง 50-60 % แม้ว่าในบางครั้ง แบบจำลองจะไม่สามารถแนะนำร้านอาหารได้ถูกต้องตรงกับชุดข้อมูลทดสอบก็ตาม

6.2. ข้อเสนอแนะ

สำหรับการต่อยอดงานวิจัยนี้ เนื่องจากในปัจจุบันยังมีข้อมูลอีกมากมายในมิติต่าง ๆ ที่สามารถนำมาเป็นสมมุติฐานในการเชื่อมโยงพฤติกรรมที่บ่งบอกถึงลักษณะของผู้บริโภค รวมถึงเทคนิคขั้นสูงต่าง ๆ เกี่ยวกับโครงข่ายประสาทเชิงลึกอื่น ๆ ที่สามารถใช้ในการวิเคราะห์ข้อมูลและเพิ่มประสิทธิภาพในการทำนาย

บรรณานุกรม

1. Gomez-Uribe, C.A. and N. Hunt, *The netflix recommender system: Algorithms, business value, and innovation.* ACM Transactions on Management Information Systems (TMIS), 2015. **6**(4): p. 1-19.
2. Koren, Y., R. Bell, and C. Volinsky, *Matrix factorization techniques for recommender systems.* Computer, 2009. **42**(8): p. 30-37.
3. He, X., et al. *Neural collaborative filtering*. in *Proceedings of the 26th international conference on world wide web*. 2017.
4. Xue, H.-J., et al. *Deep matrix factorization models for recommender systems*. in *IJCAI*. 2017. Melbourne, Australia.
5. He, X., et al., *Outer product-based neural collaborative filtering.* arXiv preprint arXiv:1808.03912, 2018.
6. Zheng, L., V. Noroozi, and P.S. Yu. *Joint deep modeling of users and items using reviews for recommendation*. in *Proceedings of the tenth ACM international conference on web search and data mining*. 2017.
7. Catherine, R. and W. Cohen. *Transnets: Learning to transform for recommendation*. in *Proceedings of the eleventh ACM conference on recommender systems*. 2017.
8. Ricci, F., L. Rokach, and B. Shapira, *Recommender systems: introduction and challenges*, in *Recommender systems handbook*. 2015, Springer. p. 1-34.
9. Zigoris, P. and Y. Zhang. *Bayesian adaptive user profiling with explicit & implicit feedback*. in *Proceedings of the 15th ACM international conference on Information and knowledge management*. 2006.
10. Burke, R., *Hybrid recommender systems: Survey and experiments.* User modeling and user-adapted interaction, 2002. **12**(4): p. 331-370.
11. Deng, Z.-H., et al. *Deepcf: A unified framework of representation learning and matching function learning in recommender system*. in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019.
12. Çakır, M., Ş.G. Öğüdücü, and R. Tugay. *A deep hybrid model for recommendation systems*. in *International Conference of the Italian Association for Artificial Intelligence*.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2019. Springer.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียน

ชื่อ-สกุล อภิสรา แซ่ลิ้ม

วัน เดือน ปี เกิด 27 มิถุนายน 2534

สถานที่เกิด กรุงเทพมหานคร

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY