



ระบบการแนะนำสถานที่ภายใต้การวิเคราะห์พฤติกรรมของผู้ใช้



สุเมธ ดาราพิสุทธิ์

ดุขฎีนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต

สาขาวิชาวิทยาการข้อมูล

คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

2566

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

ระบบการแนะนำสถานที่ภายใต้การวิเคราะห์พฤติกรรมของผู้ใช้



สุเมธ ดาราพิสุทธิ

ดุชฎินิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต

สาขาวิชาวิทยาการข้อมูล

คณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

2566

ลิขสิทธิ์เป็นของมหาวิทยาลัยบูรพา

Location-based recommendation system based on user behavior analysis



A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR DOCTOR DEGREE OF PHILOSOPHY

IN DATA SCIENCE

FACULTY OF INFORMATICS

BURAPHA UNIVERSITY

2023

COPYRIGHT OF BURAPHA UNIVERSITY

คณะกรรมการควบคุมคุณิพนธ์และคณะกรรมการสอบคุณิพนธ์ได้พิจารณาคุณิพนธ์
นิพนธ์ของ สุเมธ ดาราพิสุทธิ์ ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปรัชญาคุณิพนธ์บัณฑิต สาขาวิชาวิทยาการข้อมูล ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมคุณิพนธ์

คณะกรรมการสอบคุณิพนธ์

อาจารย์ที่ปรึกษาหลัก

.....
(ผู้ช่วยศาสตราจารย์ ดร.โกเมศ อัมพวัน)

..... ประธาน
(รองศาสตราจารย์ ดร.อนุชิต จิตพัฒนกุล)

อาจารย์ที่ปรึกษาร่วม

.....
(รองศาสตราจารย์ ดร.ณัฐนนท์ ลีลาตระกูล)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.อุรวิรัฐ สุขสวัสดิ์ชน)

.....
(รองศาสตราจารย์ ดร.สุนิสา ริมเจริญ)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.โกเมศ อัมพวัน)

..... กรรมการ
(รองศาสตราจารย์ ดร.ณัฐนนท์ ลีลาตระกูล)

..... กรรมการ
(รองศาสตราจารย์ ดร.สุนิสา ริมเจริญ)

..... คณบดีคณะวิทยาการสารสนเทศ
(ผู้ช่วยศาสตราจารย์ ดร. กฤษณะ ชินสาร)

วันที่.....เดือน.....พ.ศ.....

บัณฑิตวิทยาลัย มหาวิทยาลัยบูรพา อนุมัติให้รับคุณิพนธ์ฉบับนี้เป็นส่วนหนึ่งของ
การศึกษาตามหลักสูตรปรัชญาคุณิพนธ์บัณฑิต สาขาวิชาวิทยาการข้อมูล ของมหาวิทยาลัยบูรพา

..... คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.วิทวัส แจ่มเอียด)

วันที่.....เดือน.....พ.ศ.....

62810103: สาขาวิชา: วิทยาการข้อมูล; ปร.ด. (วิทยาการข้อมูล)

คำสำคัญ: การแนะนำสถานที่, วิเคราะห์การเชื่อมโยง, วิธีการกรองร่วม, ความโดดเด่น
 สุเมธ ดาราพิสุทธิ์ : ระบบการแนะนำสถานที่ภายใต้การวิเคราะห์พฤติกรรมของผู้ใช้.

(Location-based recommendation system based on user behavior analysis)

คณะกรรมการควบคุมดัชนีพนธ์: โกเมศ อัมพวัน, ณัฐนนท์ ลีลาตระกูล, สุนิสา ริมเจริญ ปี พ.ศ.
 2566.

ระบบแนะนำสถานที่ที่มีบทบาทสำคัญในการสร้างรายการแนะนำสถานที่เฉพาะบุคคลให้กับผู้ใช้ เพื่อค้นหาสถานที่ที่น่าสนใจที่ตรงกับความต้องการของผู้ใช้ ส่งผลให้เมื่อผู้ใช้ได้ใช้บริการหรือเที่ยวชมในสถานที่ที่ชื่นชอบสามารถช่วยเพิ่มผลกำไรให้กับผู้ให้บริการ และเจ้าของแอปพลิเคชันแนะนำสถานที่ต่าง ๆ ได้ผลประโยชน์จากการโฆษณาสถานที่ตามมาด้วย อย่างไรก็ตาม การแนะนำสถานที่ยังคงประสบปัญหาความเบาบางของข้อมูล ปัญหาการเริ่มต้นได้ยากสำหรับผู้ใช้งานใหม่และปัญหาความน่าเบื่อของรายการแนะนำสถานที่ เพื่อแก้ไขปัญหาข้างต้น งานวิจัยนี้จึงได้นำเสนอขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ที่น่าสนใจที่เรียกว่า LACF-Rec3 (Link Analysis and Collaborative Filtering Recommendation system based on 3-behaviors) ซึ่งเป็นขั้นตอนวิธีการแบบผสมผสานของขั้นตอนวิธีการวิเคราะห์การเชื่อมโยงที่เรียกว่า HITS-3 และขั้นตอนวิธีการกรองร่วมที่เรียกว่า CF-3 โดยพิจารณาพฤติกรรมการเช็คอินของผู้ใช้และสถานที่ในสามแง่มุม กล่าวคือ ความถี่ในการเช็คอิน ความหลากหลายในการเช็คอิน และการเช็คอินซ้ำ ที่ซึ่งแบ่งการพิจารณาการสร้างรายการแนะนำสถานที่ใน 2 ประเภทผู้ใช้ คือ 1. ผู้ใช้ใหม่ ขั้นตอนวิธีการ HITS-3 จะถูกใช้ในการค้นหาความโดดเด่นของสถานที่โดยพิจารณาพฤติกรรมการเช็คอินของผู้ใช้และสถานที่ และทำการจัดอันดับตามค่าคะแนนความชอบตามแง่มุมที่โดดเด่นของสถานที่เพื่อสร้างรายการแนะนำสถานที่ให้กับผู้ใช้ใหม่ 2. ผู้ใช้ที่มีประวัติการเช็คอิน ขั้นตอนวิธี CF-3 จะใช้เทคนิคการปรับปรุงขั้นตอนวิธีการกรองร่วมโดยพิจารณาประวัติการเช็คอินของผู้ใช้ร่วมกับความโดดเด่นของสถานที่ในการสร้างรายการแนะนำสถานที่ จากผลการทดลองที่นำเสนอบนชุดข้อมูล Foursquare แสดงให้เห็นว่าขั้นตอนวิธีการ LACF-Rec3 มีประสิทธิภาพที่ดีกว่าขั้นตอนวิธีการเปรียบเทียบกับก่อนหน้าในด้านของความถูกต้องของรายการแนะนำและความถูกต้องของอันดับรายการแนะนำ

62810103: MAJOR: DATA SCIENCE; Ph.D. (DATA SCIENCE)

KEYWORDS: point-of-interest recommendations, link analysis, collaborative filtering, distinctiveness

SUMET DARAPISUT : LOCATION-BASED RECOMMENDATION SYSTEM BASED ON USER BEHAVIOR ANALYSIS. ADVISORY COMMITTEE: KOMATE AMPHAWAN, Ph.D. NUTTHANON LEELATHAKUL, Ph.D. SUNISA RIMCHAROEN, Ph.D. 2023.

Location-based recommender systems (LBRs) have exhibited significant potential in providing personalized recommendations for users. As a result, users' visits to their favorite places may contribute to an increase in profits for service providers. The owners of various location recommendation applications also benefit from placing advertisements. However, several challenges (such as data sparsity, the cold-start problem, and tedium problem) need to be addressed to develop more effective LBRs. In this paper, we propose a novel POI recommendation system, called LACF-Rec3, which employs a hybrid approach of link analysis (HITS-3) and collaborative filtering (CF-3) based on three visiting behaviors: frequency, variety, and repetition. Our method considers two user types for creating a list of recommended places. For cold-start users, HITS-3 identifies distinctive POIs based on user- and POI-visit patterns, ranks them accordingly, and recommends them to cold-start users. For existing users, CF-3 utilizes collaborative filtering based on their previous check-in history and POI distinctive aspects. Our experimental results conducted on a Foursquare dataset demonstrate that LACF-Rec3 outperforms prior methods in terms of recommendation accuracy, ranking precision, and matching ratio.

กิตติกรรมประกาศ

งานวิจัยนี้สามารถนำไปสู่ความสำเร็จได้ด้วยการมีที่ปรึกษาที่ดีอย่าง ผศ.ดร.โกเมศ อัมพวัน
รศ.ดร.ณัฐนนท์ ลีลาตระกูล

และรศ.ดร. สุนิสา ริมเจริญ ซึ่งคอยส่งเสริมและให้การสนับสนุนแนวทางการทำวิจัยและแนว
ทางการใช้ชีวิต

นอกจากนี้ขอขอบคุณการสนับสนุนทุนวิจัยจากคณะวิทยาการสารสนเทศ มหาวิทยาลัยบูรพา

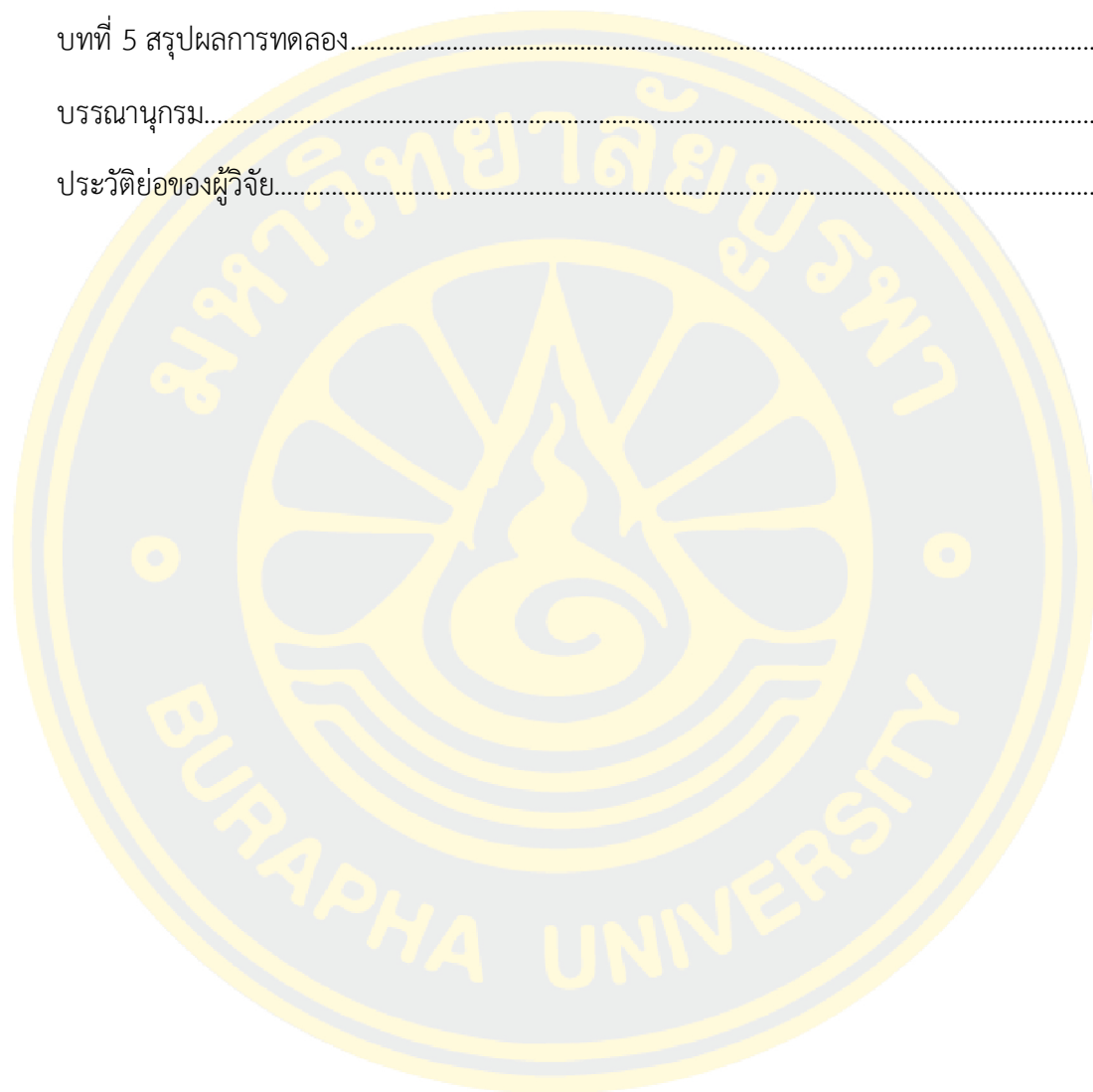
สุเมธ ดาราพิสุทธิ์



สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ญ
บทที่ 1 บทนำ.....	1
1.1 ที่มาและความสำคัญของงานวิจัย.....	1
1.2 วัตถุประสงค์ของงานวิจัย.....	4
1.3 ประโยชน์ที่คาดว่าจะได้รับ.....	5
1.4 ขอบเขตของงานวิจัย.....	5
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง.....	6
2.1 ระบบแนะนำสถานที่ (Location based recommender system).....	6
2.2 การสร้างรายการแนะนำสถานที่โดยขั้นตอนวิธีการกรองร่วม.....	7
2.3 การสร้างรายการแนะนำสถานที่โดยขั้นตอนวิธีวิเคราะห์การเชื่อมโยง.....	10
2.4 การวิเคราะห์การเปรียบเทียบขั้นตอนวิธีการก่อนหน้าและขั้นตอนวิธีการ LACF-Rec3.....	15
บทที่ 3 วิธีการที่นำเสนอ.....	18
3.1 นิยามพื้นฐาน.....	19
3.2 ขั้นตอนออฟไลน์.....	26
3.2 ขั้นตอนออนไลน์.....	30
บทที่ 4 ผลการทดลอง.....	34

4.1 การตั้งค่าการทดลองและการประเมินผลประสิทธิภาพการทดลอง.....	34
4.1.1 ชุดข้อมูลจัดเก็บประวัติการเช็คอินของผู้ใช้.....	34
4.2 ผลการทดลอง.....	40
บทที่ 5 สรุปผลการทดลอง.....	68
บรรณานุกรม.....	70
ประวัติย่อของผู้วิจัย.....	124



สารบัญตาราง

	หน้า
ตารางที่ 1 การวิเคราะห์การเปรียบเทียบระหว่างงานวิจัยก่อนหน้าและวิธีการที่นำเสนอ LACF-Rec3	16
ตารางที่ 2 คุณลักษณะของเมืองต่าง ๆ ที่ถูกใช้ในการทดลอง.....	35



สารบัญภาพ

	หน้า
ภาพที่ 1 ตัวอย่างของเครือข่ายสังคมออนไลน์ทางด้านสถานที่.....	6
ภาพที่ 2 ภาพรวมขั้นตอนวิธี POI recommendation framework using people's memory-based preferences and POI stickiness.....	10
ภาพที่ 3 ขั้นตอนวิธีการ Hyperlink-Induced Topic Search (HITS).....	11
ภาพที่ 4 ภาพรวมขั้นตอนวิธี CTD และ WHBPR.....	12
ภาพที่ 5 ภาพรวมขั้นตอนวิธี N-most interesting location-based recommender system ..	13
ภาพที่ 6 ภาพรวมขั้นตอนวิธี Tensor decomposition based collaborative filtering	14
ภาพที่ 7 ขั้นตอนวิธีการที่นำเสนอ LACF-Rec3	19
ภาพที่ 8 ตัวอย่างขั้นตอนวิธีออฟไลน์.....	28
ภาพที่ 9 ตัวอย่างขั้นตอนวิธีออนไลน์.....	32
ภาพที่ 10 ตัวอย่างของฐานข้อมูล Foursquare ที่ใช้ในการทดลอง.....	34
ภาพที่ 11 การสร้างขอบเขตการพิจารณาสถานที่ด้วยขั้นตอนวิธี MBR.....	38
ภาพที่ 12 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo.....	42
ภาพที่ 13 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda.....	43
ภาพที่ 14 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku	43
ภาพที่ 15 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato.....	44
ภาพที่ 16 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya.....	44

ภาพที่ 56 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการ แนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku	66
ภาพที่ 57 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการ แนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato	66
ภาพที่ 58 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการ แนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya	67
ภาพที่ 59 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการ แนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki	67



บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญของงานวิจัย

ในปัจจุบันเครือข่ายสังคมออนไลน์ได้รับความนิยมอย่างมาก รวมถึงเครือข่ายสังคมออนไลน์ทางด้านสถานที่ (Location based social network, LBSN) (Levandoski et al., 2012) (Yu & Chen, 2015) (Zhao et al., 2016) (Ding et al., 2018) (Xu et al., 2020) ที่จำนวนสถานที่ จำนวนผู้ใช้และจำนวนการเช็คอินมีจำนวนเพิ่มมากขึ้นอย่างรวดเร็ว ส่งผลให้ระบบแนะนำสถานที่ (Location-Based Recommender Systems, LBRS) เข้ามามีบทบาทสำคัญในการสร้างรายการแนะนำสถานที่เฉพาะบุคคลให้กับผู้ใช้ เช่น ร้านอาหาร ร้านกาแฟ ร้านพิซซา คาเฟ่ ร้านเบเกอรี่ บาร์ ห้างสรรพสินค้า และร้านสะดวกซื้อ เป็นต้น โดยการพิจารณาจากประวัติการเที่ยวชมในอดีตและปัจจัยต่าง ๆ ที่ส่งผลต่อการเที่ยวชมของผู้ใช้ เช่น ช่วงเวลา สถานที่ปัจจุบันของผู้ใช้ ความชอบส่วนบุคคล และหมวดหมู่ของสถานที่ เป็นต้น ซึ่งวัตถุประสงค์หลักของระบบแนะนำสถานที่ คือ การช่วยเพิ่มความพึงพอใจให้กับผู้ใช้ การปรับปรุงประสบการณ์การท่องเที่ยวของผู้ใช้และช่วยสนับสนุนการเติบโตของธุรกิจท้องถิ่นและการท่องเที่ยว

ดังนั้นระบบแนะนำสถานที่จึงได้รับความน่าสนใจอย่างมากจากทั้งนักวิจัยและในภาคธุรกิจ และถูกนำไปใช้ในหลากหลายงาน เช่น การท่องเที่ยว การขนส่งและการตลาด เป็นต้น โดยทางด้านการท่องเที่ยว ระบบแนะนำสถานที่สามารถช่วยนักท่องเที่ยวในการค้นหาสถานที่ที่น่าสนใจ กิจกรรมต่าง ๆ และร้านอาหารในพื้นที่ที่ไม่คุ้นเคย ทางด้านการขนส่ง ระบบแนะนำสถานที่สามารถช่วยเหลือผู้สัญจรและนักเดินทางในการวางแผนการเดินทางและระบุเส้นทางท่องเที่ยวที่สะดวกและมีประสิทธิภาพ ในด้านการตลาด ระบบแนะนำสถานที่สามารถช่วยองค์กรต่างๆ ในการโฆษณาผลิตภัณฑ์และบริการต่าง ๆ ต่อผู้ใช้โดยการพิจารณาจากสถานที่และความชอบของผู้ใช้นั้น ๆ

จากแพลตฟอร์มในเครือข่ายสังคมออนไลน์ทางด้านสถานที่ เช่น Foursquare Facebook place และ Yelp เป็นต้น ได้มีการแบ่งปันความรู้ ประสบการณ์และแลกเปลี่ยนความคิดเห็นในสถานที่ที่น่าสนใจ โดยได้ใช้ประโยชน์จากระบบแนะนำสถานที่ในการแก้ไขปัญหาข้อมูลสถานที่ที่มีจำนวนมาก (Information overload) โดยการคัดกรองสถานที่ที่คาดว่าผู้ใช้น่าจะชื่นชอบ จากมุมมองของธุรกิจต่าง ๆ ระบบแนะนำสถานที่สามารถสร้างผลกำไรให้แก่เจ้าของธุรกิจได้ อีกทั้งยังช่วยเพิ่มความภักดีต่อธุรกิจหรือแอปพลิเคชันของผู้ใช้ได้อีกด้วย แต่อย่างไรก็ตามระบบแนะนำสถานที่ยังมีหลายปัญหาที่ส่งผลต่อความถูกต้องของการสร้างรายการสถานที่ ได้แก่ ปัญหาความเบาบางของข้อมูล ปัญหาการเริ่มต้นยากของผู้ใช้และปัญหาความน่าเชื่อถือของรายการสถานที่ที่แนะนำ ซึ่งมีรายละเอียด ดังนี้

ปัญหาแรกคือความเบาบางของข้อมูล (data sparsity problem) เกิดจากสาเหตุที่อยู่ในสถานการณ์จริง จำนวนสถานที่ที่มีจำนวนมากและผู้ใช้มีแนวโน้มที่มีการเช็คอินสถานที่เพียงไม่กี่ครั้งในระแวกใกล้เคียงกับสถานที่ปัจจุบันของผู้ใช้ ส่งผลให้ข้อมูลการเช็คอินของผู้ใช้มีจำนวนน้อย ดังนั้นระบบแนะนำสถานที่จึงแนะนำได้ไม่ถูกต้องตามความต้องการของผู้ใช้เป้าหมาย

ปัญหาที่สองคือการเริ่มต้นยากของผู้ใช้ (user cold-start problem) เกิดจากสาเหตุที่ผู้ใช้ใหม่ ผู้ซึ่งไม่มีประวัติการเช็คอินในอดีต หรือมีจำนวนการเช็คอินน้อยครั้ง จึงไม่สามารถค้นหาความชอบของผู้ใช้ได้ โดยเฉพาะอย่างยิ่งในขั้นตอนวิธีการแนะนำสถานที่ที่จำเป็นต้องพิจารณาความคล้ายของประวัติการเช็คอินระหว่างผู้ใช้เป้าหมายและผู้ใช้คนอื่น ๆ ในระบบ ส่งผลต่อความถูกต้องของการสร้างรายการแนะนำสถานที่ให้กับผู้ใช้เป้าหมาย

ในการจัดการกับสองปัญหาดังกล่าว ขั้นตอนวิธีการวิเคราะห์การเชื่อมโยง (Link analysis-based algorithm, LA) เช่น ขั้นตอนวิธี HITS ซึ่งใช้โครงสร้างกราฟในการค้นหาสถานที่ที่น่าสนใจและผู้เชี่ยวชาญในพื้นที่ ขั้นตอนวิธีการนี้ใช้ประโยชน์จากองค์ความรู้และประสบการณ์ของผู้เชี่ยวชาญหรือผู้ใช้ในพื้นที่ในการสร้างการแนะนำสถานที่ให้กับผู้ใช้ใหม่ อย่างไรก็ตามถ้าผู้ใช้มีจำนวนการเช็คอินที่มากเพียงพอ ขั้นตอนวิธีการกรองร่วมสามารถให้ความถูกต้องในการสร้างรายการแนะนำสถานที่ที่ดีกว่าขั้นตอนวิธีการวิเคราะห์การเชื่อมโยง สำหรับขั้นตอนวิธีการกรองร่วม (Collaborative filtering algorithm) ที่ซึ่งพิจารณาประวัติการเช็คอินของผู้ใช้ในการสร้างรายการแนะนำสถานที่ เมื่อผู้ใช้มีจำนวนการเช็คอินที่มากแล้วขั้นตอนวิธีการนี้สามารถให้ความถูกต้องที่ดีกว่าขั้นตอนวิธีการวิเคราะห์การเชื่อมโยง แต่ในกรณีที่ผู้ใช้เป้าหมายคือผู้ใช้ใหม่หรือผู้ใช้ที่มีจำนวนการเช็คอินน้อยครั้ง การสร้างรายการแนะนำสถานที่ด้วยวิธีการวิเคราะห์การเชื่อมโยงสามารถให้ความถูกต้องที่ดีกว่าวิธีการกรองร่วม

สำหรับปัญหาที่สามคือความน่าเบื่อของรายการสถานที่ที่แนะนำ (Tedium problem) เกิดขึ้นเมื่อระบบแนะนำสถานที่โดยส่วนมากมีการพิจารณาเพียงความถี่ในการเช็คอินของผู้ใช้ ส่งผลให้โดยทั่วไปของรายการแนะนำประกอบด้วยสถานที่ที่ได้รับความนิยม สถานที่ที่มีชื่อเสียงหรือเป็นสถานที่ที่มีความโด่งดังซึ่งอาจไม่เหมาะสมสำหรับผู้ใช้ทุกคน เนื่องจากผู้ใช้อาจมีความชื่นชอบสถานที่ในหลากหลายแง่มุม เช่น บางกลุ่มผู้ใช้อาจชื่นชอบในการเที่ยวชมสถานที่ที่น่าสนใจแต่ไม่ได้มีชื่อเสียง บางกลุ่มผู้ใช้อาจมีการชื่นชอบในการเที่ยวชมสถานที่ใหม่อยู่เสมอ หรือบางกลุ่มผู้ใช้อาจมีความชื่นชอบในการกลับมาใช้บริการซ้ำในสถานที่ที่อาจไม่ได้เป็นที่นิยม เป็นต้น

ในการจัดการกับปัญหาที่กล่าวมาข้างต้น ในงานวิจัยนี้ได้นำเสนอขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ที่น่าสนใจที่ชื่อว่า LACF-Rec3 (Link Analysis and Collaborative Filtering Recommendation system based on 3-behaviors) โดยใช้ขั้นตอนวิธีการแบบผสมผสานของขั้นตอนวิธีการวิเคราะห์การเชื่อมโยงและขั้นตอนวิธีการกรองร่วม ที่ซึ่งพิจารณาพฤติกรรม

เช็คอินของผู้ใช้และสถานที่ในสามแง่มุม กล่าวคือ ความถี่ในการเช็คอิน ความหลากหลายในการเช็คอิน และการเช็คอินซ้ำ โดยขั้นตอนวิธี LACF-Rec3 จะทำการค้นหาความโดดเด่นของสถานที่ในสามแง่มุมดังกล่าว เพื่อเป็นโครงสร้างพื้นฐานในการสร้างรายการแนะนำสถานที่ที่น่าสนใจให้กับผู้ใช้ใหม่ และผู้ใช้ที่มีประวัติการเช็คอิน

สำหรับขั้นตอนการสร้างรายการแนะนำแบบโดยรวมสำหรับผู้ใช้ใหม่ ขั้นตอนวิธี HITS ซึ่งเป็นหนึ่งในขั้นตอนการวิเคราะห์การเชื่อมโยงที่ได้รับความนิยมได้ถูกพัฒนาต่อยอดเพื่อให้มีความถูกต้องของรายการแนะนำสถานที่ที่เพิ่มมากขึ้นที่เรียกว่า **ขั้นตอนวิธี HITS-3 (HITS based on 3 check-in behaviors)** โดยขั้นตอนวิธีการนี้ได้ค้นหาสถานที่ที่น่าสนใจด้วยการพิจารณาคุณลักษณะของสถานที่ ได้แก่ ความถี่ของการเช็คอิน ความหลากหลายของการเช็คอิน และการเช็คอินซ้ำของผู้ใช้ และสถานที่ ที่ซึ่งสถานที่ที่มีความถี่ในการเช็คอินสูงแสดงถึงสถานที่ที่ได้รับความนิยมและแสดงถึงสถานที่ที่เป็นกระแสตามสื่อสังคมออนไลน์ สำหรับสถานที่ที่มีความหลากหลายในการเช็คอินแสดงถึงเป็นสถานที่ที่มักเป็นจุดท่องเที่ยวและเป็นสถานที่ที่น่าสนใจภายในพื้นที่นั้น ๆ ซึ่งมีการดึงดูดผู้เข้าชมที่หลากหลายโดยไม่จำเป็นต้องมีการเที่ยวชมซ้ำ สำหรับสถานที่ที่มักมีการเที่ยวชมหรือใช้บริการซ้ำ แสดงถึงสถานที่ในชีวิตประจำวัน เช่น ร้านสะดวกซื้อ ห้างสรรพสินค้า ร้านอาหาร หรือเป็นสถานที่เฉพาะตามความชอบแต่ละบุคคล ได้แก่ สถานที่ออกกำลังกาย สนามฟุตบอล สนามบาสเกตบอลและบอร์ดเกมส์คาเฟ่ เป็นต้น

นอกจากนี้ขั้นตอนวิธี HITS-3 ได้มีการพิจารณาสถานที่ที่น่าสนใจ ด้วยการพิจารณาพฤติกรรมการเช็คอินของผู้ใช้ ได้แก่ ความถี่ในการเที่ยวชมของผู้ใช้ ความหลากหลายของสถานที่ที่เที่ยวชม และจำนวนสถานที่ที่ผู้ใช้เที่ยวชมซ้ำ ที่ซึ่งผู้ใช้ที่มีจำนวนการเที่ยวชมสูงแสดงถึงผู้ใช้ที่มีความกระตือรือร้น ชอบเที่ยวชมในสถานที่ต่าง ๆ และเป็นผู้ที่ชอบเข้าสังคมจึงมักมีการเที่ยวชมในสถานที่ที่ได้รับความนิยม สำหรับผู้ใช้ที่มีความหลากหลายของสถานที่ที่เที่ยวชมแสดงถึงผู้ใช้ที่ชื่นชอบเปิดประสบการณ์ใหม่ในการท่องเที่ยว โดยหลีกเลี่ยงการไปที่เดิมซ้ำ ๆ สำหรับผู้ใช้ที่มักมีจำนวนการเที่ยวชมหรือใช้บริการที่เดิมซ้ำ แสดงถึงความชื่นชอบเฉพาะต่อสถานที่นั้น ๆ และมักเป็นผู้ที่มีความภักดีต่อสถานที่นั้น ๆ ดังนั้นขั้นตอนวิธีที่นำเสนอ HITS-3 โดยพิจารณาความโดดเด่นและคะแนนความโดดเด่น (อธิบายรายละเอียดในบทที่ 3) ของทุกสถานที่และทุกผู้ใช้จะถูกนำไปพิจารณาในการสร้างรายการแนะนำสถานที่ที่น่าสนใจให้กับผู้ใช้

สำหรับขั้นตอนการสร้างรายการแนะนำแบบเฉพาะบุคคลสำหรับผู้ใช้ที่มีประวัติการเช็คอินสถานที่ที่น่าสนใจ แต่ละผู้ใช้จะถูกพิจารณาจากขั้นตอนวิธีการปรับปรุงวิธีการกรองร่วมที่เรียกว่า **CF-3 (Collaborative Filtering based on 3-distinctiveness)** เมื่อผู้ใช้ที่มีประวัติการเช็คอินร้องขอรายการแนะนำสถานที่ แล้วขั้นตอนวิธี CF-3 จะค้นหาผู้ใช้คนอื่น ๆ ในระบบผู้ซึ่งมีพฤติกรรมการเช็คอินสถานที่ที่คล้ายกับผู้ใช้เป้าหมายมากที่สุดโดยการคำนึงถึงความโดดเด่นของสถานที่แทนที่การ

พิจารณาความคล้ายของการเช็คอินสถานที่โดยตรง จากนั้นรายการแนะนำสถานที่ที่จะถูกสร้างโดยพิจารณาสถานที่ปัจจุบันของผู้ใช้เป้าหมาย

ในส่วนของ การประเมินผลประสิทธิภาพของขั้นตอนวิธี LACF-Rec3 งานวิจัยนี้ได้มีการประเมินผลในชุดข้อมูล Foursquare ซึ่งเป็นฐานข้อมูลจริงที่จัดเก็บประวัติการเช็คอินของผู้ใช้ จากการทดลองประสิทธิภาพความถูกต้องของรายการแนะนำสถานที่ด้วยขั้นตอนวิธี Precision และ Recall ความถูกต้องของอันดับรายการแนะนำสถานที่ด้วยขั้นตอนวิธี NDCG (Normalized Discounted Cumulative Gain) และการพิจารณาความชอบของผู้ใช้ด้วยขั้นตอนวิธี Matching Ratio แสดงให้เห็นว่าขั้นตอนวิธีที่นำเสนอ LACF-Rec3 ให้ความถูกต้องที่ดีกว่าขั้นตอนวิธีการเปรียบเทียบก่อนหน้าและจากผลการทดลองยังแสดงให้เห็นว่าขั้นตอนวิธี LACF-Rec3 สามารถแก้ปัญหาความเบาบางของข้อมูล ปัญหาการเริ่มต้นได้ยากและปัญหาความน่าเบื่อของสถานที่ที่แนะนำได้อย่างมีประสิทธิภาพ

1.2 วัตถุประสงค์ของงานวิจัย

1. เพื่อพัฒนาขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ที่น่าสนใจให้มีประสิทธิภาพที่เพิ่มขึ้นในด้านความถูกต้องที่ซึ่งสามารถแนะนำได้ทั้งผู้ใช้ใหม่และผู้ใช้ที่มีประวัติการเช็คอินโดยการวิเคราะห์พฤติกรรมเช็คอิน ได้แก่ ความถี่ในการเช็คอิน ความหลากหลายในการเช็คอิน และการเช็คอินซ้ำของผู้ใช้และสถานที่

2. เพื่อสร้างรายการแนะนำสถานที่ที่น่าสนใจให้กับผู้ใช้ โดยแบ่งได้เป็น 2 ประเภท คือ

2.1 รายการแนะนำสถานที่แบบโดยรวม จากการค้นหาสถานที่ที่มีความโดดเด่นเฉพาะพื้นที่สำหรับแนะนำให้กับผู้ใช้ใหม่

2.2 รายการแนะนำสถานที่เฉพาะบุคคล สำหรับผู้ใช้ที่มีประวัติการเช็คอินโดยการวิเคราะห์พฤติกรรมเช็คอินของความโดดเด่นของสถานที่ของผู้ใช้เป้าหมายกับผู้ใช้คนอื่นในระบบ เพื่อค้นหาสถานที่ที่คาดว่าผู้ใช้เป้าหมายชื่นชอบ

3. เพื่อเป็นขั้นตอนวิธีพื้นฐานในการพัฒนาระบบเทคโนโลยีสารสนเทศสำหรับการแนะนำสถานที่ที่น่าสนใจเฉพาะพื้นที่ให้กับผู้ใช้ใหม่และผู้ใช้ที่มีประวัติการเช็คอินที่มากเพียงพอเพื่อช่วยในการตัดสินใจ โดยการวิเคราะห์พฤติกรรมเช็คอินของผู้ใช้

4. เพื่อให้ผู้ที่มีความสนใจในแนวคิดที่นำเสนอสามารถศึกษาและประยุกต์ใช้ในงานวิจัยและประยุกต์ใช้กับระบบแนะนำสถานที่ของข้อมูลจริงได้

1.3 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ที่น่าสนใจที่มีความถูกต้องเพิ่มมากขึ้น ซึ่งสามารถสร้างรายการแนะนำสถานที่ได้ทั้งผู้ใช้ใหม่และผู้ใช้ที่มีประวัติการเช็คอินโดยการวิเคราะห์พฤติกรรมการเช็คอินของผู้ใช้
2. ได้องค์ความรู้ใหม่ของรายการแนะนำสถานที่แบบโดยรวมจากการค้นหาสถานที่ที่มีความโดดเด่นเฉพาะพื้นที่สำหรับการแนะนำให้กับผู้ใช้ใหม่ ที่ซึ่งสามารถแก้ไขปัญหาความเบาบางของข้อมูลและปัญหาการแนะนำสถานที่ให้กับผู้ใช้ใหม่ได้
3. ได้องค์ความรู้ใหม่ของรายการแนะนำสถานที่เฉพาะบุคคล สำหรับผู้ใช้ที่มีประวัติการเช็คอินโดยการวิเคราะห์พฤติกรรมการเช็คอินของผู้ใช้ ที่ซึ่งสามารถเพิ่มความถูกต้องให้กับรายการแนะนำสถานที่ได้และสามารถแก้ไขปัญหาความน่าเบื่อของสถานที่ที่แนะนำได้
4. ได้ขั้นตอนวิธีพื้นฐานในการพัฒนาระบบเทคโนโลยีสารสนเทศสำหรับการแนะนำสถานที่ที่น่าสนใจเฉพาะพื้นที่ให้กับผู้ใช้ใหม่และผู้ใช้ที่มีประวัติการเช็คอินเพื่อช่วยในการตัดสินใจโดยการวิเคราะห์พฤติกรรมการเช็คอินของผู้ใช้
5. ได้ผลงานวิจัยตีพิมพ์และเผยแพร่ในวารสารวิชาการ

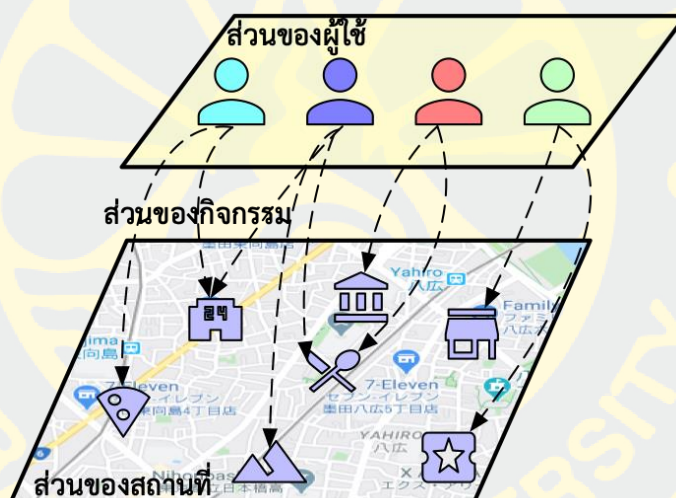
1.4 ขอบเขตของงานวิจัย

1. ฐานข้อมูลที่ใช้ในการทดลองของขั้นตอนวิธีการที่นำเสนอนี้ได้ใช้การจัดเก็บประวัติการเช็คอินของผู้ใช้จากเว็บไซต์ Foursquare ประกอบด้วย (Yang et al., 2014) 1. รหัสผู้ใช้ 2. รหัสสถานที่ 3. รหัสหมวดหมู่สถานที่ 4. หมวดหมู่สถานที่ 5. ละติจูด 6. ลองจิจูด 7. เวลาในการเช็คอิน 8. เวลา UTC โดยจัดเก็บในช่วง 12 เมษายน 2012 ถึง 16 กุมภาพันธ์ 2013 (ประมาณ 10 เดือน) โดยพิจารณาการสร้างรายการแนะนำสถานที่ใน 6 พื้นที่ของประเทศญี่ปุ่น คือ 1. เมือง Tokyo 2. เมือง Chiyoda 3. เมือง Shinjuku 4. เมือง Minato 5. เมือง Shibuya 6. เมือง Kawasaki
2. การแนะนำสถานที่ที่จะสามารถดำเนินการได้กับข้อมูลที่ประกอบด้วย ละติจูด (Latitude) ลองจิจูด (Longitude) ของสถานที่และหมวดหมู่ของสถานที่เท่านั้น
3. สถานที่ที่ถูกแนะนำให้กับผู้ใช้ประกอบด้วยหมวดหมู่ร้านอาหาร เช่น ร้านอาหาร ร้านอาหารญี่ปุ่น ร้านอาหารจีน ร้านกาแฟ ร้านเบเกอรี่ ร้านแซนวิช และร้านขนมหวาน เป็นต้น
4. การจำลองที่อยู่ปัจจุบันของผู้ใช้สำหรับกำหนดขอบเขตของการเลือกสถานที่เพื่อนำมาสร้างรายการแนะนำได้ใช้วิธีการ Minimum Bounding Rectangles (MBRs) (Ying et al., 2017)
5. การประเมินผลประสิทธิภาพของวิธีการที่นำเสนอนี้ได้ใช้วิธีการวัดความถูกต้องด้วยวิธี Precision วิธี Recall และวิธี Matching ratio และการวัดความถูกต้องของอันดับสถานที่ในรายการแนะนำด้วยวิธี Normalized Discounted Cumulative Gain (NDCG) (Wang et al., 2013)

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

สำหรับทฤษฎีและงานวิจัยที่เกี่ยวข้องเป็นการอธิบายถึงประวัติความเป็นมาและความสำคัญของระบบแนะนำสถานที่และรายละเอียดงานวิจัยก่อนหน้าเกี่ยวกับระบบแนะนำสถานที่ โดยทฤษฎีและงานวิจัยที่เกี่ยวข้องนี้ได้ทำการแบ่งวิธีการสร้างรายการแนะนำสถานที่ออกเป็น 2 หมวดหมู่ คือ 1. การสร้างรายการแนะนำสถานที่โดยพิจารณาประวัติการเช็คอินของผู้ใช้ 2. การสร้างรายการแนะนำสถานที่โดยพิจารณาผู้เชี่ยวชาญเฉพาะพื้นที่หรือผู้ใช้ภายในพื้นที่นั้น ๆ

2.1 ระบบแนะนำสถานที่ (Location based recommender system)



ภาพที่ 1 ตัวอย่างของเครือข่ายสังคมออนไลน์ทางด้านสถานที่

จากเทคโนโลยีอินเทอร์เน็ตและสมาร์ทโฟนที่ล้ำสมัยได้ส่งผลให้เครือข่ายสังคมออนไลน์ (Social network) ได้รับความนิยมจากผู้ใช้เป็นจำนวนมาก หนึ่งในเรื่องที่ได้ได้รับความนิยมอย่างมากที่สุดคือการแบ่งปันและแลกเปลี่ยนประสบการณ์ของสถานที่ที่ผู้ใช้ได้ทำการเยี่ยมชม เช่น ร้านอาหาร ห้างสรรพสินค้า ร้านกาแฟ ร้านเบเกอรี่ และร้านสะดวกซื้อ เป็นต้น ส่งผลให้ข้อมูลสถานที่ ข้อมูลผู้ใช้ และประวัติการเช็คอินของผู้ใช้และสถานที่เพิ่มขึ้นเป็นจำนวนมาก จากรายงานในเดือนมิถุนายน ปี ค.ศ. 2016 แอปพลิเคชัน Foursquare ได้มีการจัดเก็บสถิติในการเช็คอินของผู้ใช้ซึ่งมีมากกว่า 8 พันล้านเช็คอิน และสถานที่ที่ผู้ใช้ทำการเช็คอินมากกว่า 65 ล้านสถานที่รวมถึงผู้ใช้ในระบบซึ่งมากกว่า 55 ล้านผู้ใช้ในแต่ละเดือน (Zhao et al., 2016) และจากรายงานของแอปพลิเคชัน Yelp เปิดเผยว่ามีผู้ใช้งานต่อเดือนประมาณ 29 ล้านผู้ใช้ (Xu et al., 2020) จากข้อมูลดังกล่าวมาแสดงให้เห็นถึงความนิยมและความสนใจจากผู้ใช้งานจำนวนมาก ส่งผลให้ระบบแนะนำสถานที่ (Location

based recommender system) ได้เข้ามามีบทบาทสำคัญในการช่วยตัดสินใจ คัดกรองและแนะนำสถานที่ที่ผู้ใช้ชื่นชอบหรือคาดว่าจะเที่ยวชม ซึ่งระบบแนะนำสถานที่ที่สามารถแบ่งการสร้างรายการแนะนำออกเป็น 3 ส่วน คือ 1. การสร้างรายการแนะนำผู้ใช้ 2. การสร้างรายการแนะนำกิจกรรม 3. การสร้างรายการแนะนำสำหรับสถานที่ ดังภาพที่ 2-1 ที่ซึ่งการสร้างรายการแนะนำผู้ใช้คือการแนะนำผู้ใช้หรือกลุ่มของผู้ใช้ที่ซึ่งมีความคล้ายหรือความชอบใกล้เคียงกับผู้ใช้เป้าหมายหรือเป็นผู้ใช้ที่มีความเชี่ยวชาญหรือมีความรู้ในพื้นที่นั้น ๆ (Xiao et al., 2010) (Symeonidis et al., 2011) (Wang et al., 2014) สำหรับการสร้างรายการแนะนำกิจกรรมคือการแนะนำกิจกรรมที่ขึ้นกับช่วงเวลา ดังเช่น การเลือกซื้อสินค้าห้างสรรพสินค้าในช่วงบ่ายหรือวันหยุด การเที่ยวชมพิพิธภัณฑ์ในช่วงวันหยุดและการชมภาพยนตร์ในช่วงเวลากลางคืน เป็นต้น (V. Zheng et al., 2010) (V. W. Zheng et al., 2010) (Yang et al., 2014) สุดท้ายคือการสร้างรายการแนะนำสำหรับสถานที่ ซึ่งสามารถแบ่งรายละเอียดการแนะนำออกเป็นการแนะนำสถานที่เดียวให้กับผู้ใช้ และการแนะนำลำดับของสถานที่ให้กับผู้ใช้หรือทริป (Lim et al., 2018) (Lim et al., 2019)

2.2 การสร้างรายการแนะนำสถานที่โดยขั้นตอนวิธีการกรองร่วม

ในขั้นตอนวิธีการกรองร่วม (Collaborative filtering , CF)(Koren et al., 2021; Sarwar et al., 2001; Su & Khoshgoftaar, 2009) สำหรับระบบแนะนำเป็นขั้นตอนวิธีการสร้างรายการแนะนำแบบเฉพาะบุคคลซึ่งถูกใช้งานในหลายแอปพลิเคชัน เช่น สถานที่ ภาพยนตร์ เพลง ข่าว วิดีโอ และกิจกรรม เป็นต้น สามารถแบ่งได้เป็น 2 หมวดหมู่ คือ 1. ขั้นตอนวิธีการกรองร่วมโดยใช้โมเดล (model-based CF technique)(Baral & Li, 2018; Koren et al., 2021; Zhao et al., 2020) เป็นขั้นตอนโมเดลทางคณิตศาสตร์เพื่อใช้สำหรับฝึกฝนและทำนายค่าความชอบของผู้ใช้ วิธีที่นิยมมากที่สุดของขั้นตอนวิธีการกรองร่วมโดยใช้โมเดล คือ ขั้นตอนวิธีการแยกเมทริกซ์ (Matrix factorization, MF)(Koren et al., 2021) โดยเทคนิคการแยกเมทริกซ์จะใช้เวกเตอร์แทนรูปแบบการเช็คอินโดยสามารถแยกเป็นเมทริกซ์สำหรับผู้ใช้และรูปแบบแฝง และเมทริกซ์สำหรับสถานที่และรูปแบบแฝงเพื่อค้นหารูปแบบแฝงจากพฤติกรรมการเช็คอินของผู้ใช้และสถานที่ นอกจากนี้ขั้นตอนวิธีการกรองร่วมโดยใช้โมเดลยังประกอบด้วยโมเดลความน่าจะเป็น (Probabilistic model) และเทคนิคการลดมิติข้อมูล (Dimensionality reduction technique) เพื่อสร้างรายการแนะนำสำหรับผู้ใช้ 2. ขั้นตอนวิธีการกรองร่วมโดยใช้หน่วยความจำ (memory-based CF technique) (Sarwar et al., 2001) สามารถแบ่งได้เป็น 2 ขั้นตอนวิธีการย่อย คือ 1. ขั้นตอนวิธีการกรองร่วมโดยพิจารณาผู้ใช้เป็นหลัก (User-based CF)(Su & Khoshgoftaar, 2009) 2. ขั้นตอนวิธีการกรองร่วมโดยพิจารณาสถานที่เป็นหลัก (Item-based CF) ในส่วนของขั้นตอนวิธีการกรองร่วมโดยพิจารณาผู้ใช้เป็นหลักจะเป็นการค้นหาความคล้ายระหว่างผู้ใช้เป้าหมายกับผู้ใช้คนอื่น ๆ ในระบบโดยพิจารณา

ประวัติการเช็คอินของผู้ใช้เพื่อสร้างรายการแนะนำ ในส่วนของขั้นตอนวิธีการกรองร่วมโดยพิจารณาสถานที่เป็นหลักจะเป็นการค้นหาความคล้ายระหว่างสถานที่ที่ถูกเยี่ยมชมโดยผู้ใช้งานร่วมกันเพื่อสร้างรายการแนะนำสถานที่ สำหรับงานวิจัยที่พัฒนาและปรับปรุงขั้นตอนวิธีการกรองร่วมได้แก่

Yuan และคณะ (Yuan et al., 2013) ได้นำเสนอขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ให้กับผู้ใช้ โดยการประยุกต์ใช้วิธีการกรองร่วมโดยพิจารณาผู้ใช้เป็นหลักและพิจารณาช่วงเวลาในการเช็คอินของผู้ใช้ที่เรียกว่า time-aware POI recommendation ซึ่งวิธีการที่นำเสนอนี้สามารถทำงานได้ดีในข้อมูลที่มีความหนาแน่นเพียงพอ อย่างไรก็ตามโดยส่วนมากข้อมูลมักประสบปัญหาความเบาบางของข้อมูล (Sparsity problem) ดังนั้นผู้วิจัยจึงได้นำเสนอขั้นตอนวิธี Smoothing ด้วยพิจารณาช่วงเวลาที่ใกล้เคียงในการเช็คอินของผู้ใช้และพิจารณาการเช็คอินของสถานที่จากผู้ใช้ช่วงเวลาที่มีความคล้ายกับผู้ใช้เป้าหมายในการแก้ไขปัญหานี้ จากนั้นเทคนิค Bayes rules ได้ถูกนำเสนอในการค้นหาพฤติกรรมในการเช็คอินของผู้ใช้เพื่อสร้างรายการแนะนำสถานที่ จากขั้นตอนวิธีการที่นำเสนอนี้สามารถเพิ่มความถูกต้องและลดปัญหาความเบาบางของข้อมูลได้

Baral และ Li (Baral & Li, 2018) ได้นำเสนอการผสมผสานกันของสองโมเดลการสร้างรายการแนะนำสถานที่เฉพาะบุคคลโดยใช้ Ranking-based model และ Matrix factorization-based model ที่ซึ่งพิจารณาหลาย ๆ ปัจจัยที่ส่งผลต่อการเยี่ยมชมของสถานที่ ได้แก่ ข้อมูลสถานที่ เช่น พิกัดภูมิศาสตร์ หมวดหมู่ และช่วงเวลาในการเช็คอิน ข้อมูลผู้ใช้ เช่น ความคิดเห็น คำแนะนำ คะแนนที่ให้กับสถานที่ ระยะทางจากสถานที่ปัจจุบันของผู้ใช้และความสัมพันธ์ของเพื่อนระหว่างผู้ใช้ เป็นต้น เริ่มต้น มุมมองทางด้านความถี่จะถูกพิจารณาในการสร้างรายการแนะนำ จากนั้นขั้นตอนวิธีการดังกล่าวจะทำการผสมหลาย ๆ มุมมอง เช่น สองมุมมองของปัจจัยในการเช็คอินจะถูกพิจารณาในการสร้างรายการแนะนำ เช่น ความถี่ในการเช็คอินและช่วงเวลาในการเช็คอิน ความถี่ในการเช็คอินและความสัมพันธ์การเป็นเพื่อนในสื่อสังคมออนไลน์ ความถี่ในการเช็คอินและหมวดหมู่ของสถานที่และความถี่ในการเช็คอินและระยะทางของสถานที่ เป็นต้น

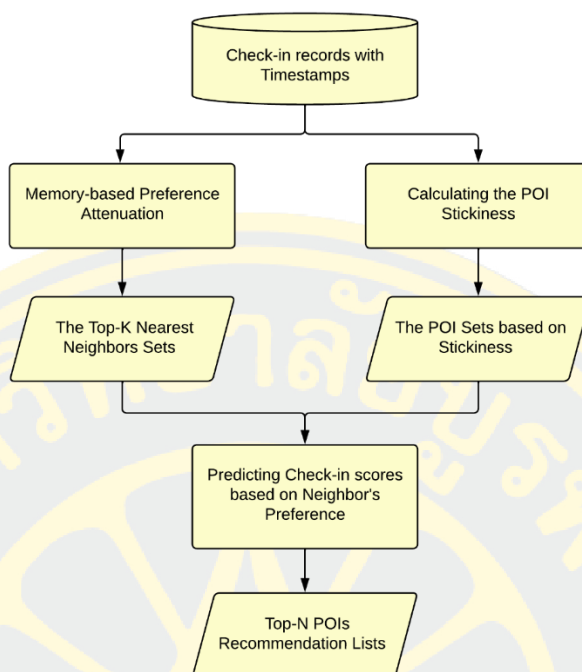
Zhao และคณะ (Zhao et al., 2020) นำเสนอขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ที่เรียกว่า Sentimental-Spatial POI Mining (SPM) โดยการพิจารณาความรู้สึกในการแสดงความคิดเห็นต่อสถานที่และคุณลักษณะต่างๆ ของสถานที่ที่ซึ่งได้จากผู้ใช้ที่มีการโพสต์พร้อมด้วยข้อความและสถานที่ในสื่อสังคมออนไลน์ที่ซึ่งในงานวิจัยนี้ได้ใช้การทดลองข้อมูลผู้ใช้จาก Sina Weibo โดยการประยุกต์ใช้ขั้นตอนวิธี Radial Basis Function (RBF) จากนั้นงานวิจัยนี้ได้นำเสนอขั้นตอนวิธี Sentimental-Spatial POI Recommendation (SPR) model โดยการพิจารณาความคล้ายระหว่างความรู้สึกในการแสดงความคิดเห็นต่อสถานที่และระยะห่างระหว่างสถานที่โดยใช้โมเดลการแยกเมทริกซ์ในการสร้างรายการแนะนำสถานที่เฉพาะบุคคล

ในปี 2017 Si และคณะ (Si et al., 2017) นำเสนอขั้นตอนวิธี CTF-ARA สำหรับสร้างรายการแนะนำสถานที่ โดยขั้นแรกได้ใช้ขั้นตอนวิธีการจัดกลุ่ม K-means สำหรับแบ่งประเภทผู้ใช้งานออกเป็นผู้ใช้ซึ่งมีการเช็คอินในสถานที่ต่าง ๆ บ่อยครั้ง (Active user) และผู้ใช้ซึ่งมีการเช็คอินในสถานที่ต่าง ๆ เพียงครั้งคราว (Inactive user) โดยผู้ใช้ที่เป็นประเภทมีการเช็คอินบ่อยครั้งจะมีการสร้างรายการแนะนำโดยการคำนึงถึงช่วงเวลาในการเช็คอินของผู้ใช้ร่วมกับการหาความคล้ายของสถานที่ใกล้เคียงในช่วงเวลาต่าง ๆ กับช่วงเวลาในปัจจุบันที่ต้องการสร้างรายการแนะนำเพื่อแก้ไขปัญหาความเบาบางของข้อมูล จากผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธีที่นำเสนอสามารถสร้างรายการแนะนำได้อย่างถูกต้องและแก้ไขปัญหาความเบาบางของข้อมูลได้

ปี 2019 ขั้นตอนวิธี memory-based POI preference attenuation model ถูกนำเสนอโดย Gan และ Gao (Gan & Gao, 2019) ขั้นตอนวิธีนี้สร้างรายการแนะนำสถานที่เฉพาะบุคคลโดยใช้ปรับปรุงขั้นตอนวิธีพื้นฐานจากขั้นตอนวิธีการกรองร่วม โดยก่อนสร้างเมตริกซ์การเช็คอินระหว่างผู้ใช้และสถานที่ วิธีการที่นำเสนอนี้ได้พิจารณาความชอบของผู้ใช้โดยให้ความสำคัญกับการเช็คอินใหม่ ๆ ของผู้ใช้ด้วยขั้นตอนวิธี memory-based POI preference attenuation model ซึ่งใช้แนวคิดของ Ebbinghaus forgetting curve จากนั้นจะทำการค้นหาความคล้ายระหว่างผู้ใช้โดยพิจารณาผู้ใช้เป็นหลักและประยุกต์ใช้ขั้นตอนวิธีการกรองร่วมในการสร้างรายการแนะนำสถานที่ให้กับผู้ใช้

Khazaei และ Alimohammadi (Khazaei & Alimohammadi, 2019) นำเสนอขั้นตอนวิธีการสร้างรายการแนะนำสถานที่โดยพิจารณาหลากหลายบริบทเกี่ยวข้องกับการเช็คอิน เรียกว่าขั้นตอนวิธี Context-aware group-oriented location recommendation system (CLGRW) สำหรับบริบทที่พิจารณา ได้แก่ บริบทของผู้ใช้ คือ ความสัมพันธ์ทางสังคมออนไลน์ของผู้ใช้และความชอบส่วนบุคคล บริบทของสถานที่ คือ หมวดยุค ความนิยม และระยะห่างทางพื้นที่ และบริบททางด้านสิ่งแวดล้อม คือ สภาพอากาศและวันของสัปดาห์ บริบทเหล่านี้จะถูกพิจารณาร่วมกันด้วยขั้นตอนวิธี random walk with restart (RWR) algorithm

Zhang และคณะ (Zhang et al., 2021) ได้แนะนำขั้นตอนวิธีการสร้างรายการแนะนำสถานที่โดยพัฒนาต่อยอดจากขั้นตอนวิธี memory-based POI preference attenuation model (Gan & Gao, 2019) ที่เรียกว่า Using users' memory-based preferences and the POI stickiness method (U-CF-Memory-Stickiness) โดยพิจารณาความสำคัญของการเช็คอินล่าสุดด้วยวิธี Ebbinghaus forgetting curve technique และเพิ่มการให้ความสำคัญของสถานที่ที่ถูกเยี่ยมชมซ้ำด้วยวิธี POI stickiness ทั้งสองเทคนิคจะถูกคำนวณร่วมกันด้วยขั้นตอนวิธีการกรองร่วม แสดงภาพรวมขั้นตอนวิธี POI recommendation framework using people's memory-based preferences and POI stickiness ดังภาพที่ 2



ภาพที่ 2 ภาพรวมขั้นตอนวิธี POI recommendation framework using people's memory-based preferences and POI stickiness

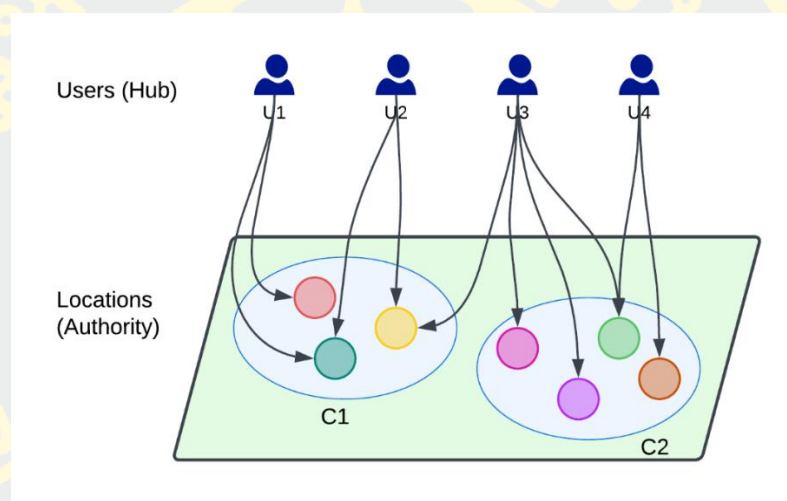
ล่าสุดในปี 2023 Zhou และคณะ (Zhou et al., 2023) นำเสนอขั้นตอนวิธี CULT-TF สร้างรายการแนะนำสถานที่เฉพาะบุคคล โดยพิจารณาข้อมูลบริบทระหว่างผู้ใช้ร่วมกับขั้นตอนวิธี Tensor factorization ในขั้นตอนแรกจะทำการจัดกลุ่มผู้ใช้ที่มีความรู้ของสถานที่ภายในพื้นที่และค้นหากลุ่มผู้ใช้ที่มีความชอบที่คล้ายกัน จากนั้นขั้นตอนวิธี Tensor factorization โดยใช้พื้นฐานโครงสร้าง U-L-T tensor ประกอบด้วย U แสดงถึงกิจกรรมของผู้ใช้ L แสดงถึงสถานที่ที่ได้รับความนิยมและ T แสดงถึงช่วงเวลาที่ได้รับคามนิยมในการเยี่ยมชมสถานที่ จากนั้นขั้นตอนวิธี Tucker factorization ร่วมกับขั้นตอนวิธี Least squares optimization จะถูกนำมาใช้ในการคำนวณเพื่อสร้างรายการแนะนำสถานที่ให้กับผู้ใช้

2.3 การสร้างรายการแนะนำสถานที่โดยขั้นตอนวิธีวิเคราะห์การเชื่อมโยง

การสร้างรายการแนะนำสถานที่โดยพิจารณาผู้ใช้ภายในพื้นที่นั้น ๆ เป็นขั้นตอนวิธีที่ใช้ความรู้และความเชี่ยวชาญของผู้ใช้ในพื้นที่เพื่อช่วยในการค้นหาสถานที่ที่สนใจ ที่ซึ่งแตกต่างจากงานวิจัยก่อนหน้านี้จำเป็นต้องใช้ประวัติการเช็คอินของผู้ใช้ ด้วยเหตุนี้ผู้ใช้ใหม่หรือนักท่องเที่ยวที่มีการเช็คอินน้อยครั้งหรือไม่มีประวัติการเช็คอินเลยจะส่งผลให้ความถูกต้องในการสร้างรายการแนะนำลดลง โดยการสร้างรายการแนะนำสถานที่โดยพิจารณาผู้ใช้ภายในพื้นที่หรือผู้เชี่ยวชาญในพื้นที่จะมี

ทั้งแบบการสร้างรายการแนะนำแบบโดยรวมและแบบเฉพาะบุคคลโดยการพิจารณาผู้ใช้ภายในพื้นที่ หรือผู้เชี่ยวชาญในพื้นที่แทนที่การพิจารณาผู้ใช้ทั้งหมดของระบบ ดังนี้

ในปี 2009 Zheng และคณะ (Zheng et al., 2009) นำเสนอขั้นตอนวิธีการสร้างรายการแนะนำสถานที่โดยใช้ขั้นตอนวิธีที่เรียกว่า Tree-Based Hierarchical Graph หรือ TBHG โดยพิจารณาการเช็คอินจากจีพีเอสของผู้ใช้ในหลาย ๆ ขนาดของพื้นที่โดยจัดเก็บเป็นลำดับชั้นแบบโครงสร้างต้นไม้และได้ทำการค้นหาผู้เชี่ยวชาญเฉพาะพื้นที่เพื่อสร้างรายการแนะนำสถานที่ที่ใช้การขั้นตอนวิธีการ HITS ดังภาพที่ 3 โดยขั้นตอนวิธี TBHG สามารถสร้างรายการแนะนำสถานที่ที่ถูกต้องกว่าระบบแนะนำสถานที่แบบดั้งเดิม



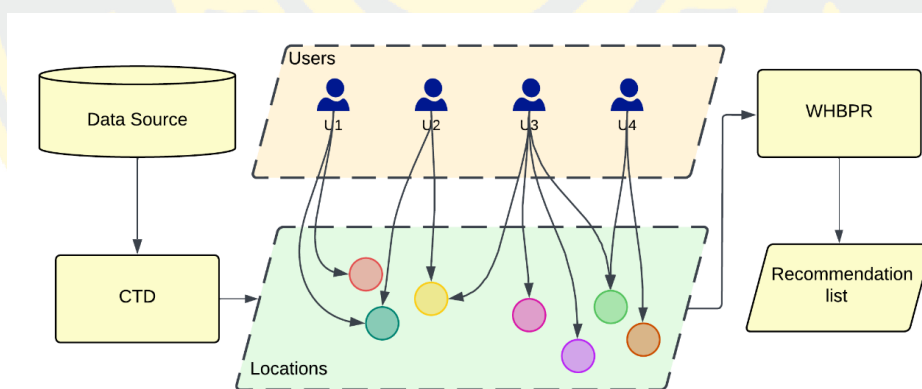
ภาพที่ 3 ขั้นตอนวิธีการ Hyperlink-Induced Topic Search (HITS)

ในปี 2012 Bao และคณะ (Bao et al., 2012) ได้นำเสนอระบบแนะนำสถานที่โดยพิจารณาความชอบของผู้ใช้ร่วมกับระยะทาง ที่ซึ่งมีการพิจารณาประวัติการเช็คอินของผู้ใช้ร่วมกับการค้นหาผู้เชี่ยวชาญเฉพาะพื้นที่ด้วยวิธี HITS ที่ซึ่งพิจารณาความถี่ของการเช็คอินของผู้ใช้และสถานที่ในแต่ละเมือง จากนั้นได้ปรับใช้ประวัติการเช็คอินในรูปแบบของโครงสร้างต้นไม้ที่เก็บรวบรวมหมวดหมู่ของการสถานที่ที่ผู้ใช้เช็คอินกับผู้เชี่ยวชาญเฉพาะพื้นที่ที่เรียกว่าโครงสร้าง Weight Category Hierarchy (WCH) โดยการหาความคล้ายระหว่างโครงสร้าง WCH แล้วประยุกต์ใช้ขั้นตอนวิธีการกรองร่วมในการทำนายค่าคะแนนความน่าสนใจของสถานที่ ที่ซึ่งวิธีการที่นำเสนอสามารถสร้างรายการแนะนำได้ดีกว่าขั้นตอนวิธีก่อนหน้า

ในปี 2013 Long และ Joshi (Long & Joshi, 2013) ได้ปรับปรุงขั้นตอนวิธี HITS สำหรับสร้างรายการแนะนำสถานที่ โดยพิจารณาความหลากหลายในการเช็คอินสถานที่ของผู้ใช้ด้วยการคำนวณค่า Entropy จากแนวคิดที่ว่าผู้ใช้ที่ไปหลายสถานที่มักมีความรู้หรือมีความเชี่ยวชาญ

มากกว่าผู้ใช้ที่ไปเพียงสถานที่เดียว อีกทั้งวิธีการที่นำเสนอนี้มีการพิจารณาความชอบของเพื่อนในสื่อสังคมออนไลน์ของผู้ใช้เป้าหมายอีกด้วย โดยข้อดีของวิธีการที่นำเสนอนี้คือไม่ต้องใช้ประวัติการเช็คอินของผู้ใช้เป้าหมายในการคำนวณ จากการทดลองในการสร้างรายการแนะนำสถานที่ในหมวดหมู่สถานบันเทิงยามค่ำคืนและหมวดหมู่ทั้งหมด แสดงให้เห็นว่าวิธีการที่นำเสนอนี้สามารถสร้างรายการแนะนำที่มีความถูกต้องมากกว่าขั้นตอนวิธี HITS แบบดั้งเดิม

Bagci และ Karagoz (Bagci & Karagoz, 2016) นำเสนอขั้นตอนวิธี Context-aware Location Recommendation with Random Walk หรือ CLoRW สำหรับสร้างรายการแนะนำสถานที่ โดยโครงสร้างหลักของวิธีคือการพัฒนาโครงสร้างกราฟที่ซึ่งแต่ละโหนดจะประกอบด้วยข้อมูลและบริบทของผู้ใช้ภายในระบบและประยุกต์ใช้ขั้นตอนวิธีการ Random-walk ร่วมกับพิจารณาบริบทต่าง ๆ ของผู้ใช้ เช่น ความสัมพันธ์กับเพื่อนในสื่อสังคมออนไลน์ ความชอบของผู้ใช้และสถานที่ปัจจุบันของผู้ใช้ และในขั้นตอนวิธีการนี้ยังมีการค้นหาผู้เชี่ยวชาญในพื้นที่ร่วมกับสถานที่ที่ได้รับความนิยมโดยใช้ขั้นตอนวิธี HITS แบบดั้งเดิม จากผลการทดลองแสดงให้เห็นว่าวิธีการที่ผู้วิจัยนำเสนอสามารถสร้างรายการแนะนำได้ถูกต้องมากกว่าวิธีการกรองร่วมแบบดั้งเดิมและยังสามารถแก้ไขปัญหาคอมพิวเตอร์ของข้อมูล รวมถึงไม่จำเป็นต้องทำการฝึกฝนโมเดลใหม่ทั้งหมดเมื่อมีผู้ใช้ใหม่เข้ามาในระบบ

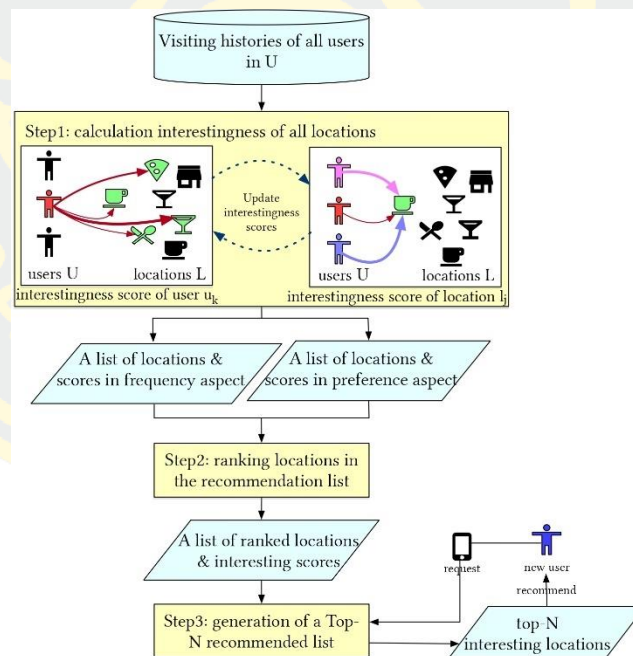


ภาพที่ 4 ภาพรวมขั้นตอนวิธี CTD และ WHBPR

ในปี 2017 ขั้นตอนวิธี Context-aware tensor decomposition (CTD) และขั้นตอนวิธี weighted HITS-based POI rating (WHBPR) ถูกนำเสนอโดย Ying และคณะ (Ying et al., 2017) เพื่อการสร้างรายการแนะนำสถานที่แบบเฉพาะบุคคล ด้วยขั้นตอนวิธีการนี้ได้แก้ไขปัญหาคอมพิวเตอร์ของข้อมูลด้วยการพิจารณาหมวดหมู่ของสถานที่แทนที่สถานที่ในขั้นตอนวิธี CTD และค้นหาสถานที่ที่น่าสนใจด้วยขั้นตอนวิธี WHBPR ซึ่งมีการพิจารณาความสัมพันธ์ของเพื่อนเพิ่มเติมภายใต้การพิจารณาขั้นตอนวิธีการแยกเมทริกซ์ จากนั้นได้ค้นหาสถานที่ที่น่าสนใจด้วยการประยุกต์ขั้นตอน

วิธีการ HITS ร่วมกับพิจารณาความคล้ายระหว่างพฤติกรรมของผู้ใช้ จากผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธี CTD ร่วมกับ WHBPR สามารถให้ความถูกต้องที่ดีกว่าวิธีการ HITS และขั้นตอนวิธีการกรองร่วมแบบดั้งเดิม แสดงภาพรวมขั้นตอนวิธี CTD และ WHBPR ดังภาพที่ 4

ปี 2020 ขั้นตอนวิธีการ N-most interesting location-based recommender system หรือเรียกว่าขั้นตอนวิธี NILR (Darapisut et al., 2022) ถูกนำเสนอเพื่อสร้างรายการแนะนำสถานที่ที่น่าสนใจสำหรับผู้ใช้งานใหม่และผู้ใช้ที่มีการเช็คอินน้อยครั้ง โดยการปรับปรุงขั้นตอนวิธีการ HITS ร่วมกับพิจารณาความถี่ในการเช็คอินที่สามารถหาได้จากจำนวนสถานที่ที่ผู้ใช้เช็คอินร่วมกับความชอบของผู้ใช้ซึ่งค้นหาจากการเช็คอินซ้ำของผู้ใช้ จากแนวคิดที่ว่าสถานที่ที่ได้รับความนิยมทางด้านความถี่ควรเกิดจากผู้ที่มีความถี่ในการเยี่ยมชมสถานที่ในพื้นที่บ่อยครั้ง และสถานที่ที่ได้รับความนิยมทางด้านความชอบควรเกิดจากผู้ใช้ที่มีการเช็คอินซ้ำในสถานที่ของพื้นที่ดังกล่าว จากนั้นเซทรายการสถานที่ของทั้งด้านความถี่และความชอบจะถูกจัดอันดับเพื่อสร้างเป็นรายการแนะนำสุดท้ายโดยการเปรียบเทียบระหว่างคะแนนความน่าสนใจในสองแง่มุมดังกล่าว แสดงภาพรวมขั้นตอนวิธี NILR ดังภาพที่ 5

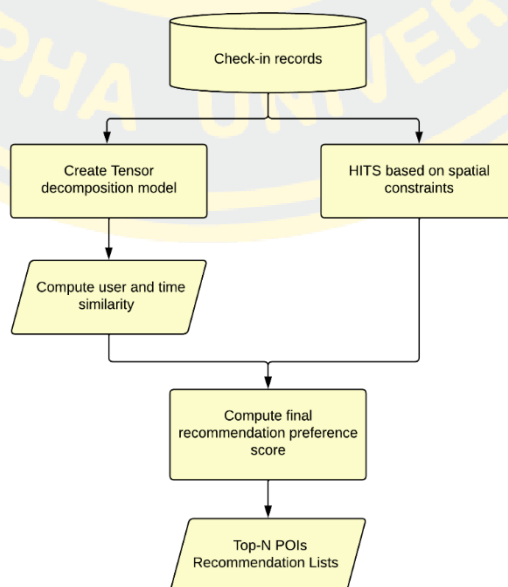


ภาพที่ 5 ภาพรวมขั้นตอนวิธี N-most interesting location-based recommender system

ในปี 2021 Sun และคณะ (Sun et al., 2021) นำเสนอขั้นตอนวิธี weighted HITS-based model สำหรับการสร้างรายการแนะนำสถานที่ โดยปรับปรุงขั้นตอนวิธี HITS เนื่องจากขั้นตอนวิธีการ HITS นั้นมีการสร้างรายการแนะนำสถานที่ที่น่าสนใจโดยพิจารณาจากความนิยมของผู้ใช้ใน

พื้นที่เพียงอย่างเดียว โดยไม่ได้พิจารณาถึงความนิยมของสถานที่นั้นว่าเกิดจากความน่าสนใจของสถานที่หรือเพราะว่าสถานที่นั้นมีความสะดวกในการเดินทาง ดังนั้นในงานวิจัยนี้จึงได้ปรับปรุงขั้นตอนวิธี HITS ที่เรียกว่าขั้นตอนวิธี weighted HITS โดยพิจารณา 2 คุณลักษณะคือความน่าสนใจของสถานที่และสถานที่ที่เป็นจุดขนส่งและทำการเพิ่มค่าน้ำหนักให้กับคุณลักษณะทั้งหมด 3 คุณลักษณะคือ 1. Popularity Weight คือการพิจารณาร้อยละของค่าความนิยมของสถานที่ 2. Transition Weight คือการพิจารณาร้อยละของจำนวนการขนส่งจากจุดขนส่งไปยังสถานที่ 3. Duration Weight คือการพิจารณาร้อยละของช่วงเวลาที่ใช้ในสถานที่ จากสมมติฐานที่ว่าสถานที่ที่น่าสนใจควรจะสามารถเข้าถึงได้ง่ายและมีจุดขนส่งหลายสถานที่ไปยังสถานที่ที่น่าสนใจนั้น และจุดขนส่งที่น่าสนใจและมีความง่าย สะดวกสบายควรสามารถไปยังสถานที่ที่น่าสนใจได้หลาย ๆ ที่ จากผลการทดลองในงานวิจัยนี้แสดงให้เห็นถึงวิธีการที่ผู้วิจัยนำเสนอสามารถให้ความถูกต้องที่ดีกว่าวิธีการ HITS แบบดั้งเดิม

ในปี 2021 ขั้นตอนวิธี Tensor decomposition based collaborative filtering (TDCF) ถูกนำเสนอโดย Yin และคณะ (Yin et al., 2021) สำหรับสร้างรายการแนะนำสถานที่เฉพาะบุคคล โดยการใช้ประโยชน์จากโครงสร้าง Tensor decomposition เพื่อพิจารณาความสัมพันธ์ของผู้ใช้และหมวดหมู่สถานที่ที่เช็คอินร่วมกับพิจารณาช่วงเวลาในการเช็คอินของผู้ใช้ ซึ่งสามารถแก้ไขปัญหาความเบาบางของข้อมูล จากนั้นได้พิจารณาความนิยมของสถานที่ด้วยขั้นตอนวิธี HITS ร่วมกับค่านิ่งถึงที่อยู่ปัจจุบันของผู้ใช้ จากผลการทดลองแสดงให้เห็นว่าวิธีการ TDCF สามารถให้ความถูกต้องที่ดีกว่าขั้นตอนวิธีการกรองร่วมแบบเดิม แสดงภาพรวมขั้นตอนวิธี Tensor decomposition based collaborative filtering ดังภาพที่ 6



ภาพที่ 6 ภาพรวมขั้นตอนวิธี Tensor decomposition based collaborative filtering

ล่าสุดขั้นตอนวิธี privacy-preserving time-aware recommendation หรือ PPTA-RM (Fan et al., 2023) ได้ถูกนำเสนอเพื่อสร้างรายการแนะนำสถานที่ถัดไปให้กับผู้ใช้ (Next POI recommendation) ที่ซึ่งขั้นตอนนี้จะทำการค้นหาความชอบโดยรวม (coarse-grained level) และความชอบโดยละเอียด (fine-grained level) สำหรับการพิจารณาความชอบโดยรวมคือการพิจารณาความชอบของหมวดหมู่ของผู้ใช้ด้วยการคำนึงถึงช่วงเวลาในการเช็คอินถัดไปโดยวิธี singular spectrum analysis (SSA) ร่วมกับวิธีการแยกแยะทริกซ์ สำหรับการพิจารณาความชอบโดยละเอียดจะคำนึงถึงระยะทางระหว่างสถานที่กับผู้ใช้ด้วยการปรับปรุงขั้นตอนวิธีการ HITS เพื่อสร้างรายการแนะนำสถานที่ต่อไป

2.4 การวิเคราะห์การเปรียบเทียบขั้นตอนวิธีการก่อนหน้าและขั้นตอนวิธีการ LACF-Rec3

หลายงานวิจัยก่อนหน้าได้มีการสร้างรายการแนะนำสถานที่โดยพัฒนาขั้นตอนวิธีการวิเคราะห์การเชื่อมโยงและขั้นตอนวิธีการกรองร่วม อย่างไรก็ตามงานวิจัยส่วนมากได้มีการพิจารณาเฉพาะความถี่ของการเช็คอินของผู้ใช้ ส่งผลให้เกิดปัญหาความน่าเบื่อของรายการแนะนำสถานที่เนื่องจากสถานที่ที่ถูกแนะนำด้วยความถี่ของการเช็คอินที่สูงจะมีความโดดเด่นมากกว่าสถานที่ที่น่าสนใจในมุมมองอื่น ๆ เช่น บางสถานที่ที่มีความน่าสนใจในมุมมองของความหลากหลายของผู้ใช้ที่เยี่ยมชมโดยไม่ได้มีความถี่ในการเยี่ยมชมที่สูงมากนัก หรือบางสถานที่ที่มีความน่าสนใจในมุมมองของการถูกใช้บริการซ้ำของผู้ใช้ เป็นต้น ดังนั้นงานวิจัยที่นำเสนอจึงได้มีการพิจารณาความโดดเด่นหลายแง่มุมของสถานที่และผู้ใช้ กล่าวคือ ความถี่ ความหลากหลายและการใช้บริการซ้ำ เพื่อค้นหาความชอบโดยรวมของผู้ใช้ นอกจากนี้งานวิจัยที่นำเสนอได้มีการพัฒนาขั้นตอนวิธีการสร้างรายการแนะนำเพื่อให้สามารถสร้างรายการแนะนำได้ทั้งผู้ใช้ใหม่และผู้ใช้ที่มีประวัติการเช็คอิน เพื่อแก้ไขปัญหาการเริ่มต้นยากของผู้ใช้และปัญหาความเบาบางของข้อมูล สำหรับผู้ใช้ใหม่ รายการแนะนำสถานที่ที่จะถูกแนะนำโดยการทำนายค่าคะแนนความชอบของสถานที่จากขั้นตอนวิธีการที่นำเสนอ HITS-3 สำหรับผู้ใช้ที่มีประวัติการเช็คอิน รายการแนะนำสถานที่ที่จะถูกแนะนำโดยการทำนายค่าคะแนนความชอบจากขั้นตอนวิธีการที่นำเสนอ CF-3 โดยใช้พื้นฐานความโดดเด่นของสถานที่ที่ได้จากขั้นตอน HITS-3 ในการค้นหาความชอบระหว่างผู้ใช้

ดังตารางที่ 2.1 แสดงการวิเคราะห์การเปรียบเทียบระหว่างงานวิจัยก่อนหน้าและวิธีการที่นำเสนอ LACF-Rec3 ทางผู้วิจัยได้สังเกตเห็นว่ามีหนึ่งงานวิจัย (Long & Joshi, 2013) ที่พิจารณาความหลากหลายของการเช็คอินในการสร้างรายการแนะนำสถานที่ และมีสองงานวิจัย (Darapisut et al., 2022; Zhang et al., 2021) ที่พิจารณาการเช็คอินซ้ำในการสร้างรายการแนะนำสถานที่ นอกจากนี้ ในงานวิจัยที่ใช้ขั้นตอนวิธีการแบบผสมผสานระหว่างขั้นตอนวิธีการวิเคราะห์การเชื่อมโยง

และขั้นตอนวิธีการกรองร่วมคือ (Bao et al., 2012; Yin et al., 2021; Ying et al., 2017) เพื่อสร้างรายการแนะนำสถานที่เฉพาะบุคคลให้กับผู้ใช้ แต่อย่างไรก็ตาม ในงานวิจัยก่อนหน้ายังไม่มียานวิจัยที่พิจารณาพฤติกรรมการเช็คอินในสามแง่มุม คือ ความถี่ในการเช็คอิน ความหลากหลายในการเช็คอิน และการเช็คอินซ้ำของผู้ใช้และสถานที่ ร่วมกับการพิจารณาความโดดเด่นของผู้ใช้และสถานที่ในแง่มุมดังกล่าวเพื่อสร้างรายการแนะนำสถานที่ให้กับผู้ใช้

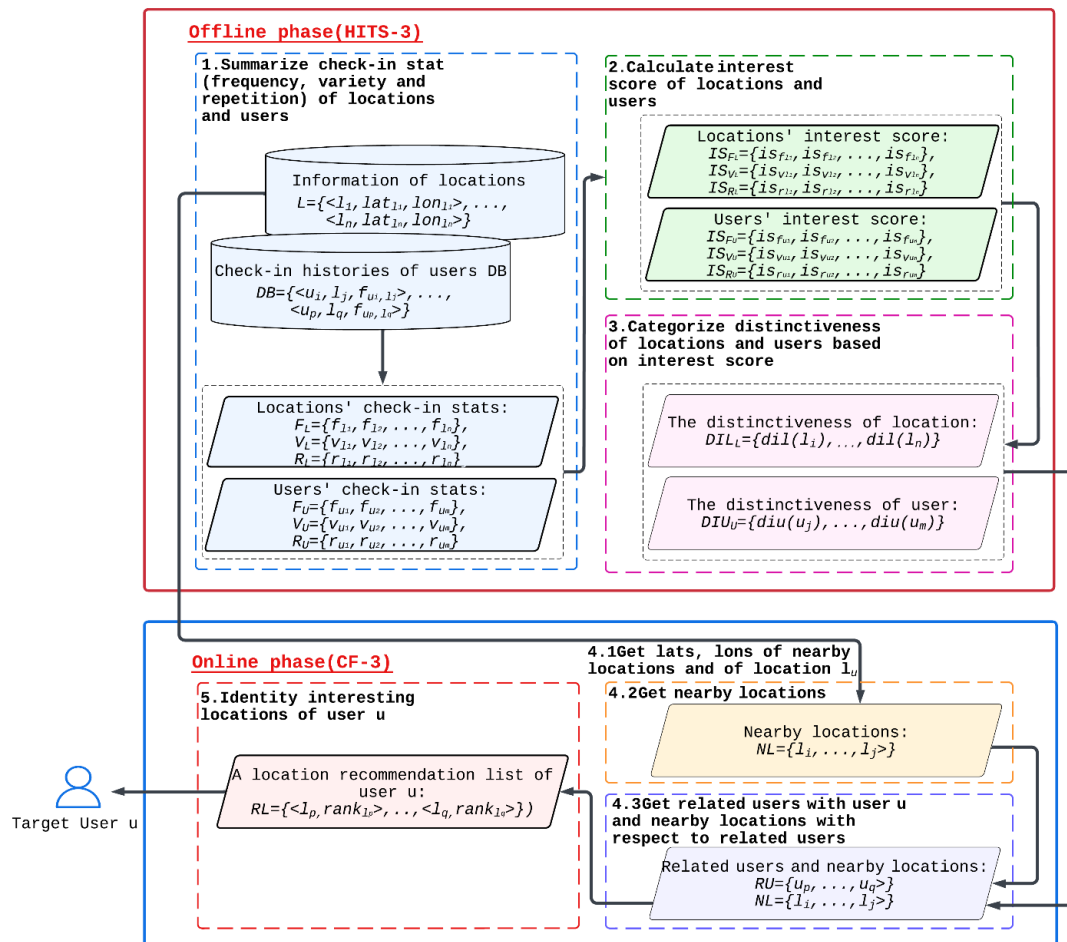
ตารางที่ 1 การวิเคราะห์การเปรียบเทียบระหว่างงานวิจัยก่อนหน้าและวิธีการที่นำเสนอ LACF-Rec3

งานวิจัย	วิธีการวิเคราะห์การเชื่อมโยง HITS		วิธีการกรองร่วม	คุณลักษณะ			
	สถานที่	ผู้ใช้		ความถี่	ความหลากหลาย	การเช็คอินซ้ำ	ความโดดเด่น
TBHG (Zheng et al., 2009)	✓	✓		✓			
LocPref (Bao et al., 2012)		✓	✓	✓			
ImpHITS (Long & Joshi, 2013)	✓			✓	✓		
CLoRW (Bagci & Karagoz, 2016)	✓	✓		✓			
TAP-F (Ying et al., 2017)	✓	✓	✓	✓			
NILR (Darapisut et al., 2022)	✓			✓		✓	
WHITS (Sun et al., 2021)	✓			✓			
TDCF (Yin et al., 2021)	✓		✓	✓			
PPTA-RM (Fan et al., 2023)	✓		✓	✓			
FCDST (Baral & Li, 2018)			✓	✓			
SPR (Zhao et al., 2020)			✓	✓			
timePOI (Yuan			✓	✓			

งานวิจัย	วิธีการวิเคราะห์การเชื่อมโยง HITS		วิธีการ กรองร่วม	คุณลักษณะ			
	สถานที่	ผู้ใช้		ความถี่	ความหลากหลาย	การ เข้าถึง ง่าย	ความ โดดเด่น
et al., 2013)							
CTF-ARA (Si et al., 2017)			✓	✓			
U-CF-M (Gan & Gao, 2019)			✓	✓			
CLGRW (Khazaei & Alimohammadi, 2019)			✓	✓			
U-CF-MS (Zhang et al., 2021)			✓	✓		✓	
CULT-TF (Zhou et al., 2023)			✓	✓			
LACF-Rec3	✓	✓	✓	✓	✓	✓	✓

บทที่ 3 วิธีการที่นำเสนอ

ในบทนี้จะกล่าวถึงนิยามพื้นฐานและรายละเอียดของขั้นตอนการสร้างรายการแนะนำสถานที่ โดยขั้นตอนวิธีการที่นำเสนอที่เรียกว่าระบบแนะนำสถานที่แบบผสมผสานด้วยวิธีการวิเคราะห์การเชื่อมโยงและขั้นตอนวิธีการกรองร่วมโดยพิจารณาพฤติกรรมการเช็คอินในสามแง่มุม ที่เรียกว่า LACF-Rec3 (Link Analysis and Collaborative Filtering Recommendation system based on 3-behaviors) สามารถแบ่งขั้นตอนวิธีการทำงานออกเป็น 2 ขั้นตอนคือ 1. ขั้นตอนแบบออฟไลน์ ที่ซึ่งเป็นการค้นหาพฤติกรรมการเช็คอินในสามแง่มุม คือ ความถี่ในการเช็คอิน ความหลากหลายในการเช็คอินและการเช็คอินซ้ำของผู้ใช้และสถานที่ จากนั้นค่าคะแนนความน่าสนใจของผู้ใช้และสถานที่จะถูกคำนวณจากขั้นตอนวิธี HITS-3 (HITS based on 3 check-in behaviors) แล้วความโดดเด่นของสถานที่และผู้ใช้จะถูกระบุเพื่อใช้เป็นพื้นฐานในการสร้างรายการแนะนำในขั้นตอนวิธีการแบบออนไลน์ต่อไป 2. ขั้นตอนแบบออนไลน์ เมื่อผู้ใช้เป้าหมายมีการร้องขอรายการแนะนำสถานที่ที่น่าสนใจ แล้วขั้นตอนวิธีการที่นำเสนอจะทำการค้นหาสถานที่และผู้ใช้ในสถานที่ปัจจุบันของผู้ใช้เป้าหมาย จากนั้นขั้นตอนวิธีการกรองร่วมที่เรียกว่า CF-3 (Collaborative Filtering based on 3-distinctiveness) จะถูกคำนวณโดยการค้นหาความคล้ายของผู้ใช้เป้าหมายและผู้ใช้ในพื้นที่ด้วยการพิจารณาความโดดเด่นของสถานที่ และในขั้นตอนสุดท้ายคือการคำนวณค่าคะแนนความน่าสนใจของสถานที่ จากนั้นทำการจัดอันดับเพื่อสร้างรายการแนะนำสถานที่ให้กับผู้ใช้เป้าหมาย แสดงภาพรวมของขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 ดังภาพที่ 7



ภาพที่ 7 ขั้นตอนวิธีการที่นำเสนอ LACF-Rec3

3.1 นิยามพื้นฐาน

กำหนดให้ $L_{DB} = \{ \langle l_1, lat_{l_1}, lon_{l_1} \rangle, \dots, \langle l_n, lat_{l_n}, lon_{l_n} \rangle \}$ เป็นเซตของ n สถานที่ โดยแต่ละสถานที่ l_j จะมีข้อมูลละติจูด (lat_{l_j}) และลองจิจูด (lon_{l_j}) และกำหนดให้เซตของผู้ใช้ $U = \{ u_1, u_2, \dots, u_m \}$ มีจำนวนทั้งหมด m ผู้ใช้ เป็นผู้ใช้ที่มีการเช็คอินอย่างน้อย 1 ครั้งในสถานที่ l_j ซึ่งอยู่ใน L_{DB} และประวัติการเช็คอินของผู้ใช้ u_i ที่สถานที่ l_j สามารถกำหนดเป็น $\langle u_i, l_j, f_{u_i, l_j} \rangle$ โดยที่ f_{u_i, l_j} คือความถี่ในการเช็คอินของผู้ใช้ u_i ที่สถานที่ l_j จากนั้นกำหนดให้ $DB = \{ \langle u_1, l_1, f_{u_1, l_1} \rangle, \dots, \langle u_m, l_n, f_{u_m, l_n} \rangle \}$ เป็นเซตของประวัติการเช็คอินของทุกผู้ใช้ และให้เซตของประวัติการเช็คอินของทุกผู้ใช้ที่สถานที่ $\langle l_p, lat_{l_p}, lon_{l_p} \rangle \in L_{DB}$ สามารถกำหนดได้เป็น $DB_{l_p} = \{ \langle u_1, l_p, f_{u_1, l_p} \rangle, \dots, \langle u_m, l_p, f_{u_m, l_p} \rangle \}$ และให้เซตของประวัติการเช็คอินของผู้ใช้ u_i ที่มีการเช็คอินในทุก

สถานที่ใน L_DB กำหนดได้เป็น $DB_{u_i} = \{ \langle u_i, l_1, f_{u_i, l_1} \rangle, \dots, \langle u_i, l_n, f_{u_i, l_n} \rangle \}$ และสถานที่ปัจจุบันของผู้ใช้เป้าหมาย u สามารถกำหนดเป็น $\langle lat_u, lon_u \rangle$

ในการสร้างรายการแนะนำสถานที่ให้กับผู้ใช้เป้าหมาย u ที่อยู่ที่สถานที่ปัจจุบัน $\langle lat_u, lon_u \rangle$ นั้น ฐานข้อมูล DB, DB_{l_p}, DB_{u_i} และ $\langle lat_u, lon_u \rangle$ ถูกพิจารณาเป็นข้อมูลนำเข้าแล้ววิธีการที่นำเสนอ LACF-Rec3 จะมีการสร้างรายการแนะนำสถานที่โดยพิจารณาพฤติกรรมการใช้คอินของผู้ใช้ กล่าวคือ ความถี่ในการใช้คอิน ความหลากหลายในการใช้คอินและการใช้คอินซ้ำของผู้ใช้ ดังนั้นนิยามพื้นฐานจะถูกอธิบายโดยละเอียด ดังนี้

นิยามที่ 3.1 ความถี่ของสถานที่และผู้ใช้

ความถี่ในการใช้คอินของสถานที่ l_p คือผลรวมของจำนวนครั้งที่ผู้ใช้ทั้งหมดในเซตผู้ใช้ U ใช้คอินที่สถานที่ l_p สามารถคำนวณดังสมการที่ 3.1

$$f(l_p) = \sum_{k=1}^m f_{u_k, l_p} \quad (3.1)$$

สำหรับความถี่ในการใช้คอินของผู้ใช้ u_j คือผลรวมของทุกสถานที่ที่ซึ่งผู้ใช้ u_j มีการใช้คอินสามารถคำนวณดังสมการที่ 3.2

$$f(u_j) = \sum_{k=1}^n f_{u_j, l_k} \quad (3.2)$$

นิยามที่ 3.2 ความหลากหลายของสถานที่และผู้ใช้

ความหลากหลายของผู้ใช้ที่มีการใช้คอินสถานที่ l_p คือผลรวมของจำนวนผู้ใช้ (u_k) ผู้ซึ่งที่มีการใช้คอินอย่างน้อย 1 ครั้ง ที่สถานที่ l_p ที่ถูกค้นหาจากฐานข้อมูล DB_{l_p} สามารถคำนวณดังสมการที่ 3.3

$$v(l_p) = \sum_{k=1}^m v a_k \text{ where } v a_k = \begin{cases} 1, & \text{if } f_{u_k, l_p} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3.3)$$

จากฐานข้อมูล สำหรับความหลากหลายของสถานที่ในการใช้คอินของผู้ใช้ u_j คือ ผลรวมของจำนวนสถานที่ที่ซึ่งผู้ใช้ u_j มีการใช้คอิน สามารถคำนวณดังสมการที่ 3.4

$$v(u_j) = \sum_{k=1}^n va_k \text{ where } va_k = \begin{cases} 1, & \text{if } f_{u_k, l_p} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3.4)$$

นิยามที่ 3.3 การเช็คอินซ้ำของสถานที่และผู้ใช้

สถานที่ l_p ที่ถูกเช็คอินซ้ำ คือ ผลรวมของผู้ใช้ผู้ซึ่งมีการเช็คอินสถานที่ l_p มากกว่า 1 ครั้ง โดยค้นหาจากฐานข้อมูล DB_{l_p} สามารถคำนวณดังสมการที่ 3.5

$$r(l_p) = \sum_{k=1}^m re_k \text{ where } re_k = \begin{cases} 1, & \text{if } f_{u_k, l_p} > 1 \\ 0, & \text{otherwise} \end{cases} \quad (3.5)$$

จำนวนสถานที่ที่ผู้ใช้ u_j มีการเช็คอินซ้ำ คือ ผลรวมของจำนวนสถานที่ที่ผู้ใช้ u_j มีการเช็คอินมากกว่า 1 ครั้ง โดยค้นหาจากฐานข้อมูล DB_{u_j} สามารถคำนวณดังสมการที่ 3.6

$$r(u_j) = \sum_{k=1}^n re_k \text{ where } re_k = \begin{cases} 1, & \text{if } f_{u_j, l_k} > 1 \\ 0, & \text{otherwise} \end{cases} \quad (3.6)$$

นิยามที่ 3.4 การปรับค่าคุณลักษณะ

กำหนดให้ $F_L = \{f(l_1), f(l_2), \dots, f(l_n)\}$ คือ เซตความถี่ในการเช็คอินของทุกสถานที่ใน L_DB เพื่อป้องกันการละเลยค่าที่มีจำนวนความถี่น้อย เนื่องจากส่วนมากผู้ใช้งานมีการเช็คอินเพียงไม่กี่ครั้ง แต่ในบางกลุ่มผู้ใช้งานมีการเช็คอินมาก ส่งผลให้ความถี่ของการเช็คอินมีจำนวนน้อยอาจถูกละเลยได้ และเพื่อลดปัญหาการแนะนำของระบบโดยแ่งมุ่มใด แ่งมุ่มหนึ่งเพียงอย่างเดียว ซึ่งแต่ละความถี่ในการเช็คอิน $f(l_p) \in F_L$ จะถูกปรับค่าให้อยู่ในช่วงคะแนน 0 ถึง 1 ดังสมการที่ 3.7

$$nf(l_p) = \begin{cases} 0.5 + \left((f(l_p) - avg) \times \frac{0.5}{max-avg} \right), & f(l_p) \geq avg \\ f(l_p) \times \frac{0.5}{avg}, & \text{otherwise} \end{cases} \quad (3.7)$$

ที่ avg คือ ค่าเฉลี่ยของความถี่ในการเช็คอินใน F_L หรือกำหนดได้ดัง $avg(f(l_1), f(l_2), \dots, f(l_n))$ และ max คือค่าที่มากที่สุดของความถี่ในการเช็คอินใน F_L หรือ $max(f(l_1), f(l_2), \dots, f(l_n))$

ในลักษณะเดียวกันจากสมการที่ 3.7 จะถูกปรับค่าคุณลักษณะในแต่ละความถี่ในการเช็คอินของผู้ใช้ $f(u_j) \in F_U = \{f(u_1), f(u_2), \dots, f(u_m)\}$ แต่ละความหลากหลายของผู้ใช้ที่มีการเช็คอินในสถานที่ $v(l_p) \in V_L = \{v(l_1), v(l_2), \dots, v(l_n)\}$ แต่ละความหลากหลายของสถานที่ในการเช็คอินของผู้ใช้ $v(u_j) \in V_U = \{v(u_1), v(u_2), \dots, v(u_m)\}$ แต่ละสถานที่ที่ถูกเช็คอินซ้ำ $r(l_p) \in R_L = \{r(l_1), r(l_2), \dots, r(l_n)\}$ และแต่ละจำนวนสถานที่ที่ผู้ใช้มีการเช็คอินซ้ำ $r(u_j) \in R_U = \{r(u_1), r(u_2), \dots, r(u_m)\}$

นิยามที่ 3.5 ค่าคะแนนความน่าสนใจของสถานที่และผู้ใช้

กำหนดให้ $ISF_L = \{isf(l_1), isf(l_2), \dots, isf(l_n)\}$ คือเซตค่าความน่าสนใจของสถานที่โดยพิจารณาความถี่เป็นหลักและ $ISF_U = \{isf(u_1), isf(u_2), \dots, isf(u_m)\}$ คือเซตค่าความน่าสนใจของผู้ใช้โดยพิจารณาความถี่เป็นหลัก เริ่มต้นค่าคะแนนความน่าสนใจของผู้ใช้ในเซทรายการ ISF_U จะถูกกำหนดค่าเท่ากับ 1 สำหรับค่าคะแนนความน่าสนใจโดยพิจารณาความถี่ $isf(l_p) \in ISF_L$ ของสถานที่ $l_p \in L_DB$ สามารถคำนวณได้จากผลรวมของค่าความน่าสนใจของผู้ใช้ที่มีการเช็คอินที่สถานที่ l_p ที่ซึ่งคำนวณได้จากผลรวมของผลคูณระหว่างความถี่ในการเช็คอินของแต่ละผู้ใช้ $f(u_q)$ และค่าคะแนนความน่าสนใจของผู้ใช้โดยคำนวณจากสมการที่ 3.9 $isf(u_q)$ ดังนั้น $isf(l_p)$ สามารถคำนวณได้ดังสมการที่ 3.8 และค่าคะแนนความน่าสนใจโดยพิจารณาความถี่ $isf(u_q) \in ISF_U$ ของผู้ใช้ $u_q \in U$ คือผลรวมของผลคูณระหว่างความถี่ในการเช็คอินของแต่ละสถานที่ $f(l_p)$ และค่าคะแนนความน่าสนใจของสถานที่ $isf(l_p)$ ที่ซึ่งคำนวณได้จากสมการที่ 3.8 ดังนั้น $isf(u_q)$ สามารถคำนวณได้ดังสมการที่ 3.9

$$isf(l_p) = \sum_{q=1}^m \begin{cases} f(u_q) \times isf(u_q), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (3.8)$$

$$isf(u_q) = \sum_{p=1}^n \begin{cases} f(l_p) \times isf(l_p), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (3.9)$$

สำหรับเซตค่าคะแนนความน่าสนใจโดยพิจารณาความหลากหลายในการเช็คอินของทุกสถานที่ สามารถกำหนดได้เป็น $ISV_L = \{isv(l_1), isv(l_2), \dots, isv(l_n)\}$ เซตค่าคะแนนความน่าสนใจโดยพิจารณาความหลากหลายในการเช็คอินของผู้ใช้สามารถกำหนดได้เป็น $ISV_U = \{isv(u_1), isv(u_2), \dots, isv(u_m)\}$ เซตค่าคะแนนความน่าสนใจโดยพิจารณาการเช็คอินซ้ำของทุกสถานที่สามารถกำหนดได้เป็น $ISR_L = \{isr(l_1), isr(l_2), \dots, isr(l_n)\}$ เซตค่าคะแนนความน่าสนใจโดยพิจารณาการเช็คอินซ้ำของผู้ใช้สามารถกำหนดได้เป็น $ISR_U = \{isr(u_1), isr(u_2), \dots, isr(u_m)\}$ แต่ละค่าคะแนนความน่าสนใจ สามารถคำนวณได้ตั้งสมการที่ 3.10, 3.11, 3.12 และ 3.13 เนื่องจากในขั้นตอนนี้เป็น การคำนวณซ้ำเพื่ออัปเดตค่าคะแนนความน่าสนใจของสถานที่และผู้ใช้ส่งผลให้ค่าคะแนนดังกล่าวมีจำนวนสูงขึ้นเรื่อย ๆ ดังนั้นขั้นตอนสุดท้ายคือการปรับค่าคะแนนด้วยวิธี L-2 normalization จะถูกใช้ในการคำนวณค่า $isf(l_p)$, $isf(u_q)$, $isv(l_p)$, $isv(u_q)$, $isf(l_p)$ และ $isf(u_q)$ เพื่อให้ค่าคู่เข้าจนค่าคะแนนความน่าสนใจไม่เปลี่ยนแปลง

$$isv(l_p) = \sum_{q=1}^m \begin{cases} v(u_q) \times isv(u_q), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (3.10)$$

$$isv(u_q) = \sum_{p=1}^n \begin{cases} v(l_p) \times isv(l_p), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (3.11)$$

$$isr(l_p) = \sum_{q=1}^m \begin{cases} r(u_q) \times isr(u_q), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (3.12)$$

$$isr(u_q) = \sum_{p=1}^n \begin{cases} r(l_p) \times isr(l_p), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (3.13)$$

นิยามที่ 3.6 ความโดดเด่นของสถานที่และผู้ใช้

ในการระบุความโดดเด่น $dil(l_p)$ ของแต่ละสถานที่ $l_p \in L_{DB}$ สามารถระบุได้จากค่าที่มากที่สุดเมื่อเปรียบเทียบระหว่างค่าคะแนนความน่าสนใจของสถานที่ด้านความถี่ในการเช็คอิน $isf(l_p)$ ค่าคะแนนความน่าสนใจของสถานที่ด้านความหลากหลายในการเช็คอิน $isv(l_p)$ ค่าคะแนนความน่าสนใจของสถานที่ด้านการเช็คอินซ้ำ $isr(l_p)$ เมื่อค่าที่มากที่สุด คือ $isf(l_p)$ แล้วค่า $dil(l_p)$ จะถูกกำหนดเป็น 1 ถ้าค่าที่มากที่สุดคือ $isv(l_p)$ แล้วค่า $dil(l_p)$ จะถูกกำหนดเป็น 2 และค่าที่มากที่สุดคือ $isr(l_p)$ แล้วค่า $dil(l_p)$ จะถูกกำหนดเป็น 3 แสดงดังสมการที่ 3.14

$$dil(l_p) = \begin{cases} 1, & \text{if } isf(l_p) = \max(isf(l_p), isv(l_p), isr(l_p)) \\ 2, & \text{if } isv(l_p) = \max(isf(l_p), isv(l_p), isr(l_p)) \\ 3, & \text{if } isr(l_p) = \max(isf(l_p), isv(l_p), isr(l_p)) \end{cases} \quad (3.14)$$

ทำนองเดียวกันในการระบุความโดดเด่นของแต่ละผู้ใช้ u_q ($diu(u_q)$) กำหนดได้ดังสมการที่ 3.15

$$diu(u_q) = \begin{cases} 1, & \text{if } isf(u_q) = \max(isf(u_q), isv(u_q), isr(u_q)) \\ 2, & \text{if } isv(u_q) = \max(isf(u_q), isv(u_q), isr(u_q)) \\ 3, & \text{if } isr(u_q) = \max(isf(u_q), isv(u_q), isr(u_q)) \end{cases} \quad (3.15)$$

นิยามที่ 3.7 ค่าคะแนนความโดดเด่นของผู้ใช้

กำหนดให้ $L_{u_q} = \{ \langle l_i, f_{u_q, l_i}, dil(l_i) \rangle, \dots, \langle l_j, f_{u_q, l_j}, dil(l_j) \rangle \}$ เป็นเซตที่จัดเก็บสถานที่ที่ถูกเช็คอินของผู้ใช้ $u_q \in U$, ความถี่ในการเช็คอิน และความโดดเด่นของสถานที่ และกำหนดให้ $L_{u_q}^1$ คือเซตของสถานที่ที่โดดเด่นทางด้านความถี่ในการเช็คอินที่ผู้ใช้ u_q เช็คอิน $L_{u_q}^2$ คือเซตของสถานที่ที่โดดเด่นทางด้านความหลากหลายในการเช็คอินที่ผู้ใช้ u_q เช็คอิน และ $L_{u_q}^3$ คือเซตของสถานที่ที่โดดเด่นทางด้านการเช็คอินซ้ำที่ผู้ใช้ u_q เช็คอิน โดยทั้งสามเซตดังกล่าวสามารถค้นหาได้จาก L_{u_q} ดังนั้นค่าคะแนนความโดดเด่นของผู้ใช้ u_q กำหนดเป็น $ds(u_q) = \langle dsf(u_q), dsv(u_q), dsr(u_q) \rangle$ ที่ซึ่งถูกระบุจากความถี่ในการเช็คอิน ความหลากหลายในการเช็คอินและการเช็คอินซ้ำ สามารถคำนวณได้จากสมการที่ 3.16 3.17 และ 3.18

$$dsf(u_q) = \sum_{i=1}^{|L_{u_q}^1|} f_{u_q, l_i} \quad (3.16)$$

$$dsv(u_q) = \sum_{j=1}^{|L_{u_q}^2|} f_{u_q, l_j} \quad (3.17)$$

$$dsr(u_q) = \sum_{k=1}^{|L_{u_q}^3|} f_{u_q, l_k} \quad (3.18)$$

นิยามที่ 3.8 ค่าคะแนนความคล้ายของความโดดเด่นของผู้ใช้

กำหนดให้ผู้ใช้ u_q ที่มีการเช็คอินในสถานที่ l_p ตั้งอยู่ในบริเวณใกล้กับที่อยู่ปัจจุบันของผู้ใช้เป้าหมาย u ดังนั้นค่าความคล้ายของพฤติกรรมเช็คอินระหว่างผู้ใช้ u_q และผู้ใช้เป้าหมาย u สามารถคำนวณได้จากการหาความคล้ายแบบโคไซน์ด้วยการเปรียบเทียบความโดดเด่นของแต่ละผู้ใช้ u_q ดังสมการที่ 3.19

$$sim(u, u_q) = \frac{(dsf(u) \times dsf(u_q)) + (dsv(u) \times dsv(u_q)) + (dsr(u) \times dsr(u_q))}{\sqrt{dsf(u)^2 + dsv(u)^2 + dsr(u)^2} \times \sqrt{dsf(u_q)^2 + dsv(u_q)^2 + dsr(u_q)^2}} \quad (3.19)$$

นิยามที่ 3.9 ค่าคะแนนความน่าสนใจของสถานที่โดยการพิจารณาผู้ใช้เกี่ยวข้อง

กำหนดให้ RU เป็นเซตของผู้ใช้เกี่ยวข้อง ผู้ซึ่งมีการเช็คอินอย่างน้อย 1 ครั้งในสถานที่ที่ตั้งอยู่บริเวณปัจจุบันของผู้ใช้เป้าหมาย u และกำหนดให้ SRU เป็นเซตของผู้ใช้เกี่ยวข้อง u_j ที่มีค่าความคล้าย $sim(u, u_j)$ มากกว่าค่าขีดแบ่ง δ โดยค่าคะแนนความน่าสนใจของสถานที่ l_p สามารถคำนวณได้โดยการพิจารณาข้อมูลการเช็คอินของผู้ใช้เกี่ยวข้อง คือ 1. ค่าความคล้ายระหว่างผู้ใช้เป้าหมาย u กับผู้ใช้เกี่ยวข้อง u_j ($sim(u, u_j)$) 2. ความถี่ในการเช็คอินของผู้ใช้ u_j ของสถานที่ l_p (f_{u_j, l_p}) และ 3. ค่าความโดดเด่นของผู้ใช้ u_j ตามค่าความโดดเด่นของสถานที่ $dil(l_p)$ ดังนั้นค่าคะแนนความน่าสนใจ $rs(l_p)$ ของสถานที่ l_p สามารถคำนวณดังสมการที่ 3.20

$$rs(l_p) = \sum_{j=1}^{|SRU|} (sim(u, u_j) \times f_{u_j, l_p} \times dsx(u_x)) \quad (3.20)$$

$$\text{โดยที่ } dsx(u_x) = \begin{cases} dsf(u_j), & \text{if } dil(l_p) = 1 \\ dsv(u_j), & \text{if } dil(l_p) = 2 \\ dsr(u_j), & \text{if } dil(l_p) = 3 \end{cases}$$

นิยามที่ 3.10 ค่าคะแนนความน่าสนใจของสถานที่โดยพิจารณาค่าความน่าสนใจของสถานที่

กำหนดให้ $NL = \{l_p, \dots, l_q\}$ เป็นเซตของสถานที่ที่อยู่ใกล้เคียงกับที่อยู่ปัจจุบันของผู้ใช้ เป้าหมาย u ดังนั้นค่าคะแนนความน่าสนใจของสถานที่ $l_p \in NL$ สามารถคำนวณได้จากค่าที่มากที่สุดระหว่างค่าคะแนนความน่าสนใจทางด้านความถี่การเช็คอิน ความหลากหลายการเช็คอินและการเช็คอินซ้ำ ดังสมการที่ 3.21

$$rs(l_p) = \max(isf(l_p), isv(l_p), isr(l_p)) \quad (3.21)$$

3.2 ขั้นตอนออฟไลน์

กำหนดให้ฐานข้อมูลที่จัดเก็บประวัติการเช็คอินของผู้ใช้ DB ของทุกผู้ใช้เป็นข้อมูลนำเข้า ซึ่งขั้นตอนออฟไลน์นี้มีเป้าหมายเพื่อค้นหาความโดดเด่นของผู้ใช้และสถานที่ในแง่ของความถี่ในการเช็คอิน ความหลากหลายในการเช็คอินและการเช็คอินซ้ำ โดยทั่วไปสถานที่ที่มีความโดดเด่นทางด้านความถี่ในการเช็คอินสูงรวมถึงจำนวนความหลากหลายและการเช็คอินซ้ำสูงด้วยแสดงถึงสถานที่ที่ได้รับความนิยมหรือเป็นสถานที่ที่เป็นกระแส สถานที่ซึ่งมีความโดดเด่นทางด้านความหลากหลายสูงแสดงถึงสถานที่ที่เป็นสถานที่ท่องเที่ยวหรือสถานที่ที่เป็นจุดเด่น (Landmark) ในพื้นที่นั้น ๆ ซึ่งผู้ใช้อัปเดตมีการเที่ยวชมเพียงครั้งเดียว สำหรับสถานที่ที่มีความโดดเด่นในการเช็คอินซ้ำสูงแสดงถึงสถานที่ที่ผู้ใช้อัปเดตมีการเที่ยวชมหรือใช้บริการในชีวิตประจำวัน เช่น ร้านสะดวกซื้อ ห้างสรรพสินค้า หรือเป็นสถานที่เฉพาะแต่ละผู้ใช้ เช่น สถานที่ออกกำลังกาย สนามบาสเก็ตบอล คาเฟ่บอร์ดเกมส์ เป็นต้น ในส่วนของผู้ใช้ สำหรับผู้ใช้ที่มีความโดดเด่นทางด้านความถี่ในการเช็คอินสูงแสดงถึงผู้ใช้ที่ชอบท่องเที่ยวและมักเป็นผู้ใช้ที่ชอบเล่นสังคมออนไลน์ สำหรับผู้ใช้ที่มีความโดดเด่นด้านความหลากหลายในการเช็คอินสูงแสดงถึงผู้ใช้ที่ชอบเปิดประสบการณ์การท่องเที่ยวใหม่ ๆ สุดท้ายผู้ใช้ที่มีความโดดเด่นทางด้านการเช็คอินซ้ำสูงแสดงถึงผู้ใช้ที่มีความชอบเฉพาะและมีความซื่อสัตย์ต่อสถานที่

ในบรรทัดที่ 1 และบรรทัดที่ 2 ของขั้นตอนวิธีที่ 1 ข้อมูลในการเช็คอินของแต่ละสถานที่ l_p และแต่ละผู้ใช้ u_j นั่นคือ ความถี่ในการเช็คอินของผู้ใช้ $f(u_j)$ และความถี่ที่ถูกเช็คอินของสถานที่ $f(l_p)$ ความหลากหลายของสถานที่ที่ผู้ใช้เช็คอิน $v(u_j)$ และความหลากหลายของผู้ใช้ที่เช็คอินสถานที่ $v(l_p)$ จำนวนสถานที่ที่ผู้ใช้เช็คอินซ้ำ $r(u_j)$ และจำนวนผู้ใช้ที่เช็คอินซ้ำต่อสถานที่ $r(l_p)$ จะถูกคำนวณตามนิยามที่ 3.1 นิยามที่ 3.2 และนิยามที่ 3.3 จากนั้นจะถูกจัดเก็บในเซต LCS และ UCS ต่อมาการปรับค่าคุณลักษณะดังกล่าวจะถูกคำนวณเพื่อลดการครอบงำของคุณลักษณะที่ซึ่ง

ความถี่ในการเช็คอินมักมีค่ามากกว่าหรือเท่ากับความหลากหลายในการเช็คอิน และความหลากหลายในการเช็คอินมีค่าที่มากกว่าค่าการเช็คอินซ้ำและเพื่อเป็นการเพิกเฉยต่อค่าคุณลักษณะที่มีจำนวนน้อย อย่างเช่นในกรณีที่สถานที่ที่ผู้ใช้ u_j มีการเที่ยวชมสถานที่ l_p 5 ครั้งและผู้ใช้ u_k มีการเที่ยวชมสถานที่ l_p 10 ครั้ง นั้นอาจหมายความว่าผู้ใช้ u_j มีความชอบในสถานที่ l_p เทียบเท่ากับผู้ใช้ u_k ถึงแม้จะมีจำนวนการเที่ยวชมน้อยกว่า ดังนั้น ในบรรทัดที่ 3 และบรรทัดที่ 4 ของขั้นตอนวิธีที่ 1 จึงเป็นการปรับค่าข้อมูลการเช็คอินของผู้ใช้และสถานที่ ดังนิยามที่ 3.4 ต่อมาขั้นตอนวิธี HITS-3 จะถูกใช้ในการคำนวณค่าคะแนนความน่าสนใจของผู้ใช้และสถานที่โดยพิจารณาข้อมูลการเช็คอินในแต่ละคุณลักษณะดังกล่าว ซึ่งมีความแตกต่างกับขั้นตอนวิธี HITS ที่มีการพิจารณาทางด้านความถี่ในการเช็คอินเพียงอย่างเดียว ดังนิยามที่ 3.5 ค่าคะแนนความน่าสนใจของแต่ละสถานที่ l_p สามารถนำเสนอได้ 3 แง่มุม คือ 1. ค่าความน่าสนใจของสถานที่โดยพิจารณาความถี่ ($isf(l_p)$) 2. ค่าความน่าสนใจของสถานที่โดยพิจารณาความหลากหลาย ($isv(l_p)$) และ 3. ค่าความน่าสนใจของสถานที่โดยพิจารณาการเช็คอินซ้ำ ($isr(l_p)$) ที่ซึ่งทั้ง 3 มุมมองดังกล่าวเมื่อคำนวณเสร็จแล้ว จะถูกจัดเก็บอยู่ในรูปแบบโครงสร้าง $\langle isf(l_p), isv(l_p), isr(l_p) \rangle$ และขณะเดียวกันค่าคะแนนความน่าสนใจของแต่ละผู้ใช้ u_i จะถูกจัดเก็บในรูปแบบ $\langle isf(u_i), isv(u_i), isr(u_i) \rangle$ ด้วย

ขั้นตอนวิธีที่นำเสนอ HITS-3 ค่าคะแนนความน่าสนใจของผู้ใช้และสถานที่ที่จะถูกคำนวณแบบทำซ้ำเมื่อการทำซ้ำในแต่ละรอบจะส่งผลให้ค่าคะแนนความน่าสนใจของผู้ใช้และสถานที่ที่มีค่าที่มากขึ้น ดังนั้นการปรับค่าด้วยขั้นตอนวิธี L2-Normalization แสดงขั้นตอนการทำงานดังบรรทัดที่ 5-8 ในขั้นตอนสุดท้ายตามบรรทัดที่ 6 และบรรทัดที่ 7 ค่าคะแนนความน่าสนใจที่มากที่สุดของแต่ละผู้ใช้และสถานที่ที่จะถูกคำนวณและระบุความโดดเด่นของทั้งผู้ใช้และสถานที่ ถ้าค่าคะแนนความน่าสนใจทางด้านความถี่ของสถานที่ l_p มีค่าที่มากที่สุด แสดงถึงสถานที่ l_p มีความโดดเด่นทางด้านความถี่ในการเช็คอิน นั่นคือ $dil(l_p)$ มีค่าเท่ากับ 1 ถ้า $dil(l_p)$ มีค่าเท่ากับ 2 แสดงว่าค่าคะแนนความน่าสนใจทางด้านความหลากหลายของสถานที่ l_p มีค่าที่มากที่สุด และในกรณีสุดท้าย $dil(l_p)$ มีค่าเท่ากับ 3 แสดงว่าค่าคะแนนความน่าสนใจทางการเช็คอินซ้ำของสถานที่ l_p มีค่าที่มากที่สุด ดังนั้นค่าคะแนนความน่าสนใจของผู้ใช้ u_i ($diu(u_i)$) จะถูกคำนวณในลักษณะเดียวกับการพิจารณาค่าคะแนนความน่าสนใจของสถานที่ โดยสรุปขั้นตอนวิธีทั้งหมดแสดงดังขั้นตอนวิธีที่ 1

Algorithm 1. Offline phase Algo**Input:**

A check-in database $DB = \{ \langle u_i, l_j, f_{u_i, l_j} \rangle, \dots, \langle u_p, l_q, f_{u_p, l_q} \rangle \}$

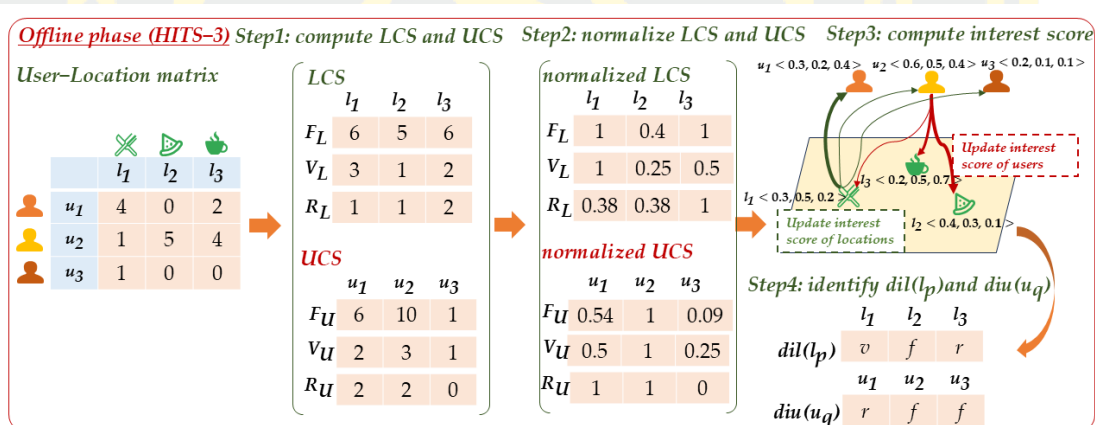
A location database $L_DB = \{ \langle l_1, lat_{l_1}, lon_{l_1} \rangle, \dots, \langle l_n, lat_{l_n}, lon_{l_n} \rangle \}$

Output:

Interest scores of locations and users $IS_{F_L}, IS_{F_U}, IS_{V_L}, IS_{V_U}, IS_{R_L}, IS_{R_U}$

A distinctiveness of each location $dil(l_p)$ and user $dii(u_q)$

- 1: compute location's check-in stats $LCS = \{ \langle f(l_1), v(l_1), r(l_1) \rangle, \dots, \langle f(l_n), v(l_n), r(l_n) \rangle \}$ (by Definitions 1–3)
- 2: compute user's check-in stats $UCS = \{ \langle f(u_1), v(u_1), r(u_1) \rangle, \dots, \langle f(u_m), v(u_m), r(u_m) \rangle \}$ (by Definitions 1–3)
- 3: normalize LCS (by Definition 4)
- 4: normalize UCS (by Definition 4)
- 5: **for** each t time **do**
- 6: compute interest score of locations and interest score of users, IS_{F_L} ,
- 7: $IS_{F_U}, IS_{V_L}, IS_{V_U}, IS_{R_L}, IS_{R_U}$ (by Definition 5)
- 8: **end for**
- 9: identify distinctiveness $dil(l_p)$ of each location in L_DB (by Definition 6)
- 10: identify distinctiveness $dii(u_q)$ of each user in U (by Definition 6)



ภาพที่ 8 ตัวอย่างขั้นตอนวิธีออฟไลน์

ดังภาพที่ 8 แสดงตัวอย่างขั้นตอนวิธีออฟไลน์ด้วยขั้นตอนวิธี HITS-3 จากประวัติการเช็คอินของผู้ใช้ ประกอบด้วยผู้ใช้งานจำนวน 3 ผู้ใช้ (u_1, u_2 และ u_3) และสถานที่จำนวน 3 สถานที่ (l_1, l_2 และ l_3) ผู้ใช้ u_1 มีการเช็คอินสถานที่ l_1 และสถานที่ l_3 จำนวน 4 ครั้งและ 2 ครั้งตามลำดับ ผู้ใช้ u_2 มีการเช็คอินสถานที่ l_1, l_2 และสถานที่ l_3 จำนวน 1 ครั้ง 5 ครั้งและ 4 ครั้งตามลำดับ และผู้ใช้ u_3 มีการเช็คอินสถานที่ l_1 จำนวน 1 ครั้ง

โดยขั้นตอนที่ 1 เซทค่าความถี่ของสถานที่ (F_L) ของสถานที่ l_1, l_2 และ l_3 จะถูกคำนวณดังเช่น $4 + 1 + 1 = 6, 0 + 5 + 0 = 5$ และ $4 + 2 + 0 = 6$ ตามลำดับ เซทค่าความหลากหลาย

ของผู้ใช้ที่เช็คอินสถานที่ (V_L) ของสถานที่ l_1, l_2 และ l_3 จะถูกคำนวณ ดังเช่น $1 + 1 + 1 = 3$, $0 + 1 + 0 = 1$ และ $1 + 1 + 0 = 2$ ตามลำดับ จากนั้นจำนวนการเช็คอินซ้ำของสถานที่ (R_L) ของสถานที่ l_1, l_2 และ l_3 จะถูกคำนวณ ดังเช่น $1 + 0 + 0 = 1$, $0 + 1 + 0 = 1$ และ $1 + 1 + 0 = 2$ ตามลำดับ ดังนั้น F_L, V_L และ R_L จะถูกจัดเก็บในเซต LCS ในทำนองเดียวกันเซตค่าความถี่ (F_U) ของผู้ใช้ u_1, u_2 และ u_3 จะถูกคำนวณ ดังเช่น $4 + 0 + 2 = 6$, $1 + 5 + 4 = 10$ และ $1 + 0 + 0 = 1$ ตามลำดับ เซตค่าความหลากหลายของสถานที่ที่ผู้ใช้เช็คอิน (V_U) ของผู้ใช้ u_1, u_2 และ u_3 จะถูกคำนวณ ดังเช่น $1 + 0 + 1 = 2$, $1 + 1 + 1 = 3$ และ $1 + 0 + 0 = 1$ ตามลำดับ จากนั้นเซตค่าการเช็คอินซ้ำ (V_U) ของผู้ใช้ u_1, u_2 และ u_3 จะถูกคำนวณ ดังเช่น $1 + 0 + 1 = 2$, $0 + 1 + 1 = 2$ และ $0 + 0 + 0 = 0$ ตามลำดับ ดังนั้น F_U, V_U และ R_U จะถูกจัดเก็บในเซต UCS ซึ่งสอดคล้องตามบรรทัดที่ 1 และ 2 ในขั้นตอนวิธีที่ 1

ในขั้นตอนที่ 2 ค่าความถี่ของสถานที่ l_1, l_2 และ l_3 จะถูกปรับค่า ดังเช่น $0.5 + \left((6-5.67) \times \frac{0.5}{6-5.67} \right) = 1$, $5 \times \frac{0.5}{5.67} = 0.4$ และ $0.5 + \left((6-5.67) \times \frac{0.5}{6-5.67} \right) = 1$ ตามลำดับ โดยค่าเฉลี่ยมีค่าเท่ากับ 5.67 และค่าที่มากที่สุดเท่ากับ 6 ในทำนองเดียวกัน ค่าความหลากหลายในการเช็คอินของผู้ใช้ของสถานที่ l_1, l_2 และ l_3 จะถูกปรับค่า ดังเช่น $0.5 + \left((3-2) \times \frac{0.5}{3-2} \right) = 1$, $1 \times \frac{0.5}{2} = 0.25$ และ $0.5 + \left((2-2) \times \frac{0.5}{3-2} \right) = 0.5$ ตามลำดับ โดยค่าเฉลี่ยมีค่าเท่ากับ 2 และค่าที่มากที่สุดเท่ากับ 3 ถัดไปค่าการเช็คอินซ้ำของสถานที่ l_1, l_2 และ l_3 จะถูกปรับค่า ดังเช่น $1 \times \frac{0.5}{1.3} = 0.38$, $1 \times \frac{0.5}{1.3} = 0.38$ และ $0.5 + \left((2-1.3) \times \frac{0.5}{2-1.3} \right) = 1$ ตามลำดับ โดยค่าเฉลี่ยมีค่าเท่ากับ 1.3 และค่าที่มากที่สุดเท่ากับ 2 สำหรับ ค่าความถี่ของผู้ใช้ u_1, u_2 และ u_3 จะถูกปรับค่า ดังเช่น $0.5 + \left((6-5.67) \times \frac{0.5}{10-5.67} \right) = 0.54$, $0.5 + \left((10-5.67) \times \frac{0.5}{10-5.67} \right) = 1$ และ $1 \times \frac{0.5}{5.67} = 0.09$ ตามลำดับ โดยค่าเฉลี่ยมีค่าเท่ากับ 5.67 และค่าที่มากที่สุดเท่ากับ 10 ค่าความหลากหลายของสถานที่ของผู้ใช้ u_1, u_2 และ u_3 จะถูกปรับค่า ดังเช่น $0.5 + \left((2-2) \times \frac{0.5}{3-2} \right) = 0.5$, $0.5 + \left((3-2) \times \frac{0.5}{3-2} \right) = 1$ และ $1 \times \frac{0.5}{2} = 0.25$ ตามลำดับ โดยค่าเฉลี่ยมีค่าเท่ากับ 2 และค่าที่มากที่สุดเท่ากับ 3 สุดท้ายค่าการเช็คอินซ้ำของผู้ใช้ u_1, u_2 และ u_3 จะถูกปรับค่า ดังเช่น

$0.5 + \left((2-2) \times \frac{0.5}{2-2} \right) = 1$, $0.5 + \left((2-2) \times \frac{0.5}{2-2} \right) = 1$ และ 0 ตามลำดับ โดยค่าเฉลี่ยมีค่าเท่ากับ 2 และค่าที่มากที่สุดเท่ากับ 2 ซึ่งสอดคล้องกับบรรทัดที่ 3 และ 4 ของขั้นตอนวิธีที่ 1

ในขั้นตอนที่ 3 ค่าคะแนนความน่าสนใจของสถานที่และผู้ใช้จะถูกคำนวณด้วยขั้นตอนวิธีการที่นำเสนอ HITS-3 ที่ซึ่งอธิบายดังบรรทัดที่ 5 ถึงบรรทัดที่ 8 ของขั้นตอนวิธีที่ 1 โดยที่ สามค่าคะแนนความน่าสนใจในแต่ละแง่มุมของแต่ละสถานที่จะถูกจัดเก็บในรูปแบบ $\langle isf(l_p), isv(l_p), isr(l_p) \rangle$ ที่ซึ่งค่าคะแนนความน่าสนใจของสถานที่ l_1 , l_2 และ l_3 จะถูกคำนวณ ดังเช่น $\langle 0.3, 0.5, 0.2 \rangle$, $\langle 0.4, 0.3, 0.1 \rangle$ และ $\langle 0.2, 0.5, 0.7 \rangle$ ตามลำดับ ดังนั้นค่าความโดดเด่นของสถานที่ l_p ($dil(l_p)$) จะถูกระบุตามบรรทัดที่ 9 ของขั้นตอนวิธีที่ 1 ดังนั้นค่าความโดดเด่นของสถานที่ l_1 คือโดดเด่นด้านความหลากหลายของผู้ใช้ สถานที่ l_2 คือโดดเด่นด้านความถี่ในการเช็คอินและสถานที่ l_3 คือโดดเด่นด้านการเช็คอินซ้ำของผู้ใช้ ทำนองเดียวกัน สามค่าคะแนนความน่าสนใจในแต่ละแง่มุมของแต่ละผู้ใช้จะถูกจัดเก็บในรูปแบบ $\langle isf(u_i), isv(u_i), isr(u_i) \rangle$ ที่ซึ่งค่าคะแนนความน่าสนใจของผู้ใช้ u_1 , u_2 และ u_3 จะถูกคำนวณ ดังเช่น $\langle 0.3, 0.2, 0.4 \rangle$, $\langle 0.6, 0.5, 0.4 \rangle$ และ $\langle 0.2, 0.1, 0.1 \rangle$ ตามลำดับ ดังนั้นค่าความโดดเด่นของแต่ละผู้ใช้ ($diu(u_q)$) จะถูกระบุตามบรรทัดที่ 10 ของขั้นตอนวิธีที่ 1 นั่นคือค่าความโดดเด่นของผู้ใช้ u_1 คือโดดเด่นด้านการเช็คอินซ้ำ ค่าความโดดเด่นของผู้ใช้ u_2 คือโดดเด่นด้านความถี่ในการเช็คอิน และค่าความโดดเด่นของผู้ใช้ u_3 คือโดดเด่นด้านความถี่ในการเช็คอินตามลำดับ

3.2 ขั้นตอนออนไลน์

สำหรับการสร้างรายการแนะนำสถานที่ให้กับผู้ใช้เป้าหมาย u สถานที่ปัจจุบันของผู้ใช้ดังกล่าว $\langle lat_u, lon_u \rangle$ จะถูกใช้ในการร้องขอรายการแนะนำสถานที่ ดังนั้นสถานที่ที่อยู่ใกล้เคียงกับสถานที่ปัจจุบันของผู้ใช้ในระยะ dt กิโลเมตร ที่ dt เป็นระยะที่มากที่สุดซึ่งถูกกำหนดโดยผู้ใช้เป้าหมายจะถูกค้นหาและสามารถกำหนดสถานที่ที่อยู่ใกล้เคียงเป็น $NL = \{l_p, \dots, l_q\}$ ซึ่งสอดคล้องกับบรรทัดที่ 1 ของขั้นตอนวิธีที่ 2 นอกจากนี้ผู้ใช้ผู้ซึ่งมีการเช็คอินอย่างน้อย 1 ครั้งในสถานที่ $l_p \in NL$ จะถูกระบุเป็นผู้ใช้แข่งขันที่ซึ่งจะถูกนำไปค้นหาผู้ใช้ที่มีพฤติกรรมเช็คอินกับผู้ใช้เป้าหมายต่อไป สามารถกำหนดเป็น $RU = \{u_i, \dots, u_j\}$ ซึ่งสอดคล้องกับบรรทัดที่ 2 ของขั้นตอนวิธีที่ 2 จากนั้นค่าคะแนนความโดดเด่นของสถานที่ทั้งสามแง่มุมในแต่ละผู้ใช้ $u_q \in RU$ นั้นคือ $ds(u_q)$ จะถูกคำนวณ โดยการพิจารณาความถี่ในการเช็คอิน ความหลากหลายในการเช็คอิน

และการเช็คอินซ้ำร่วมกับจำนวนสถานที่ที่ผู้ใช้มีการเช็คอิน ดังนิยามที่ 3.7 ความโดดเด่นของผู้ใช้ u_q โดยพิจารณาความถี่ ($dsf(u_q)$) สามารถคำนวณได้จากผลรวมความถี่ในการเช็คอินของผู้ใช้ u_q ที่สถานที่ l_j ซึ่งสถานที่ l_j มีความโดดเด่นทางด้านความถี่ในการเช็คอินหรือกล่าวคือ $dil(l_j)$ มีค่าเท่ากับ 1 สำหรับความโดดเด่นของผู้ใช้ u_q โดยพิจารณาความหลากหลาย ($dsv(u_q)$) สามารถคำนวณได้จากผลรวมความถี่ในการเช็คอินของผู้ใช้ u_q ที่สถานที่ l_j ซึ่งสถานที่ l_j มีความโดดเด่นทางด้านหลากหลายในการเช็คอินหรือกล่าวคือ $dil(l_j)$ มีค่าเท่ากับ 2 และความโดดเด่นของผู้ใช้ u_q โดยพิจารณาการเช็คอินซ้ำ ($dsl(u_q)$) สามารถคำนวณได้จากผลรวมความถี่ในการเช็คอินของผู้ใช้ u_q ที่สถานที่ l_j ซึ่งสถานที่ l_j มีความโดดเด่นทางด้านการเช็คอินซ้ำหรือกล่าวคือ $dil(l_j)$ มีค่าเท่ากับ 3 ดังนั้นค่าความคล้ายของความโดดเด่นสถานที่ที่ชื่นชอบ $sim(u, u_q)$ ของผู้ใช้เป้าหมาย u และผู้ใช้ $u_q \in RU$ จะถูกคำนวณตามนิยามที่ 3.8 เมื่อค่า $sim(u, u_q)$ มีค่าน้อยกว่าค่าขีดแบ่ง δ แล้วผู้ใช้ u_q จะถูกระบุเป็นผู้ใช้ที่ไม่เกี่ยวข้องกับผู้ใช้เป้าหมายหรือกล่าวคือผู้ใช้ u_q ไม่ได้จัดเก็บในเซต SRU ซึ่งสอดคล้องตามบรรทัดที่ 3 ของขั้นตอนวิธีที่ 2 เมื่อผู้ใช้เป้าหมาย u ไม่มีผู้ใช้ที่เกี่ยวข้องหรือกล่าวคือ $SRU = \emptyset$ แล้วผู้ใช้เป้าหมายจะถูกระบุเป็นผู้ใช้ใหม่ ตามการทำงานในบรรทัดที่ 4 ของขั้นตอนวิธีที่ 2

ดังนั้นการแนะนำสถานที่สำหรับผู้ใช้เป้าหมาย สามารถแบ่งได้เป็น 2 กรณี

1. เมื่อผู้ใช้เป้าหมาย u มีผู้ใช้ที่เกี่ยวข้อง กล่าวคือ $SRU \neq \emptyset$ ขั้นตอนวิธีการที่นำเสนอ CF-3 จะถูกใช้ในการสร้างรายการแนะนำสถานที่ สำหรับการแก้ไขปัญหาความเบาบางของข้อมูล ขั้นตอน CF-3 จะใช้การหาความคล้ายระหว่างผู้ใช้เป้าหมาย u และผู้ใช้ $u_q \in SRU$ โดยพิจารณาความโดดเด่นของสถานที่ l ที่ผู้ใช้ชื่นชอบ ($dil(l)$) นอกจากนี้ขั้นตอนวิธี CF-3 ยังคำนึงถึงค่าคะแนนความโดดเด่นของผู้ใช้ในแง่มุมมองของสถานที่นั้น ๆ ($dsx(u_q)$) ที่ซึ่งเปรียบเสมือนความเชี่ยวชาญของผู้ใช้เพื่อพิจารณาความน่าสนใจของสถานที่ โดยขั้นตอนวิธีนี้จะใช้สมมติฐานที่ว่าผู้ใช้ที่มีความชอบที่คล้ายกันมักมีการเที่ยวชมในสถานที่ที่มีความโดดเด่นแบบเดียวกันหรือคล้ายกัน สำหรับแต่ละสถานที่ ค่าคะแนนความน่าสนใจจะถูกคำนวณโดยพิจารณาค่าความคล้ายระหว่างผู้ใช้เป้าหมาย u และผู้ใช้ u_q ผู้ซึ่งเช็คอินในสถานที่ l_j ค่าความถี่ในการเช็คอินของผู้ใช้ u_q ที่สถานที่ l_j และค่าคะแนนความโดดเด่นของผู้ใช้ u_q ตามความโดดเด่นของสถานที่ $di(l_j)$ ตามนิยามที่ 3.9

2. เมื่อผู้ใช้เป้าหมาย u ไม่มีผู้ใช้ที่เกี่ยวข้อง กล่าวคือ $SRU = \emptyset$ สามารถระบุเป็นผู้ใช้ใหม่ ดังนั้นค่าคะแนนความน่าสนใจของแต่ละสถานที่ $l_j \in L_DB$ จะถูกคำนวณจากขั้นตอนวิธีที่นำเสนอ HITS-3 นั่นคือ $\langle isf(l_j), isv(l_j), isr(l_j) \rangle$ ที่ค่าคะแนนความน่าสนใจของสถานที่ l_j สามารถคำนวณจากค่าที่มากที่สุด เมื่อเปรียบเทียบกับในสามแง่มุม ตามนิยามที่ 3.10

หลังจากการพิจารณาค่าคะแนนความน่าสนใจของทุกสถานที่ทั้ง 2 กรณีดังกล่าว สถานที่ที่น่าสนใจจะถูกจัดอันดับจากมากไปหาน้อย N อันดับสำหรับสร้างเป็นรายการแนะนำสำหรับผู้ใช้งานเป้าหมาย แสดงดังบรรทัดที่ 5 ของขั้นตอนวิธีที่ 2

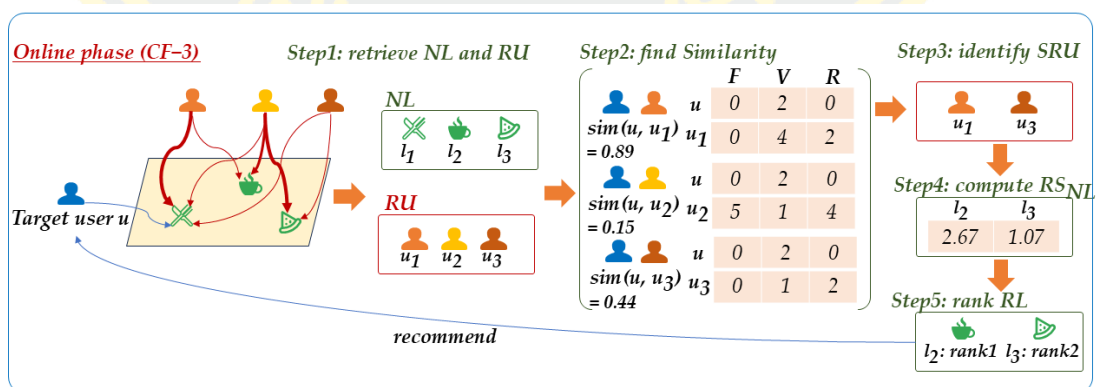
Algorithm 2. Online pphase Algo

Input:

- A check-in database $DB = \{ \langle u_i, l_j, f_{u_i, l_j} \rangle, \dots, \langle u_p, l_q, f_{u_p, l_q} \rangle \}$
 A location database $L_DB = \{ \langle l_1, lat_{l_1}, lon_{l_1} \rangle, \dots, \langle l_n, lat_{l_n}, lon_{l_n} \rangle \}$
 Interest scores of locations and users $IS_{F_L}, IS_{F_U}, IS_{V_L}, IS_{V_U}, IS_{R_L}, IS_{R_U}$
 A distinctiveness of each location $dil(l_p)$ and user $dIU(u_q)$
 A number of recommendations N
 A target user u with her current location $\langle lat_u, lon_u \rangle$
 A distance condition on recommended location, d
 A similarity threshold, δ

Output:

- A recommended list $RL = \{ \langle l_1, rs(l_1) \rangle, \langle l_2, rs(l_2) \rangle, \dots, \langle l_n, rs(l_n) \rangle \}$
 1: identify $NL = \{ l_p, \dots, l_q \}$, the set of locations nearby the current location $\langle lat_u, lon_u \rangle$ of u (within distance d)
 2: identify $RU = \{ u_i, \dots, u_j \}$, the set of relevant users visiting the location in NL at least once
 3: determine the distinctiveness score of each user, u_i , in RU and the similarity of u_i and u . Then, identify SRU by selecting the users from RU to be in SRU based on their distinctiveness with the similarity threshold δ (by Definitions 7 and 8)
 4: compute $RS_{NL} = \{ rs(l_p), \dots, rs(l_q) \}$ the set of recommendation scores of all locations in NL (if $RU \neq \emptyset$ applies Definition 9, otherwise, applies Definition 10)
 5: determine $RL = \{ \langle l_p, rank_{l_p} \rangle, \dots, \langle l_q, rank_{l_q} \rangle \}$, the set of N nearby locations with the associated ranks
 6: recommend RL to the target user u
-



ภาพที่ 9 ตัวอย่างขั้นตอนวิธีออนไลน์

ดังภาพที่ 9 แสดงตัวอย่างขั้นตอนการทำงานแบบออนไลน์ เมื่อผู้ใช้งานเป้าหมายมีการร้องขอรายการแนะนำสถานที่ ขั้นตอนวิธีการที่นำเสนอจะทำการค้นหาสถานที่ใกล้เคียงกับสถานที่ปัจจุบัน

ของผู้ใช้ (NL) นั่นคือ สถานที่ l_1 , l_2 และ l_3 และเซตของผู้ใช้ (RU) ที่มีการเช็คอินในสถานที่ใน NL นั่นคือผู้ใช้ u_1 , u_2 และ u_3 ซึ่งสอดคล้องกับบรรทัดที่ 1 และ 2 ในขั้นตอนวิธีที่ 2 สำหรับขั้นตอนที่ 2 ค่าความคล้ายระหว่างผู้ใช้เป้าหมาย u และผู้ใช้ที่อยู่ในเซต RU จะถูกคำนวณโดยการพิจารณาค่าคะแนนความโดดเด่นของสถานที่ในแต่ละผู้ใช้ ดังเช่น ค่าความคล้ายระหว่างผู้ใช้ u และผู้ใช้ u_1 ($sim(u, u_1)$) สามารถคำนวณได้โดย $\frac{(0 \times 0)(2 \times 4)(0 \times 2)}{\sqrt{0^2 + 2^2 + 0^2} \times \sqrt{0^2 + 4^2 + 2^2}}$ เท่ากับ 0.89 ค่าความคล้ายระหว่างผู้ใช้ u และผู้ใช้ u_2 ($sim(u, u_2)$) สามารถคำนวณได้โดย $\frac{(0 \times 5)(2 \times 1)(0 \times 4)}{\sqrt{0^2 + 2^2 + 0^2} \times \sqrt{5^2 + 1^2 + 4^2}}$ เท่ากับ 0.15 และค่าความคล้ายระหว่างผู้ใช้ u และผู้ใช้ u_3 ($sim(u, u_3)$) สามารถคำนวณได้โดย $\frac{(0 \times 0)(2 \times 1)(0 \times 2)}{\sqrt{0^2 + 2^2 + 0^2} \times \sqrt{0^2 + 1^2 + 2^2}}$ เท่ากับ 0.44 ตามลำดับ สำหรับขั้นตอนที่ 3 ถ้าค่าขีดแบ่งความคล้าย δ คือ 0.4 ดังนั้นผู้ใช้ u_2 ไม่ได้ถูกเลือกจัดเก็บในเซตของผู้ใช้ที่เกี่ยวข้อง SRU ดังบรรทัดที่ 3 ของขั้นตอนวิธีที่ 2 จากนั้นในขั้นตอนที่ 4 ค่าคะแนนความน่าสนใจของสถานที่ RS_{NL} จะถูกคำนวณค่าดังกรณีที่ 1 เนื่องจากเซต SRU ไม่เป็นเซตว่าง ดังนั้นค่าคะแนนความน่าสนใจของสถานที่ l_2 จะถูกคำนวณได้ดังนี้ $(0.89 \times 5 \times 0.6) + (0.44 \times 0 \times 0.2) = 2.67$ และค่าคะแนนความน่าสนใจของสถานที่ l_3 จะถูกคำนวณได้ $(0.89 \times 5 \times 0.6) + (0.44 \times 0 \times 0.2) = 2.67$, and $(0.89 \times 2 \times 0.6) + (0.44 \times 0 \times 0.2) = 1.07$ ในขั้นตอนสุดท้ายรายการแนะนำสถานที่สำหรับผู้ใช้เป้าหมาย u จะถูกจัดอันดับ นั่นคือสถานที่ l_2 คืออันดับ 1 และสถานที่ l_3 คืออันดับ 2 ตามลำดับ

บทที่ 4 ผลการทดลอง

สำหรับบทนี้ได้กล่าวถึงรายละเอียดของชุดข้อมูลที่จัดเก็บประวัติการเช็คอินของผู้ใช้ การตั้งค่าการทดลอง ขั้นตอนวิธีการประเมินผลและผลการทดลองของขั้นตอนวิธีการที่นำเสนอ และขั้นตอนการเปรียบเทียบประสิทธิภาพของขั้นตอนวิธีการที่นำเสนอ ซึ่งในการทดลองนี้ได้ทำการเปรียบเทียบกับขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ก่อนหน้า 4 ขั้นตอน ได้แก่ 1. ขั้นตอนวิธีการกรองร่วมโดยพิจารณาผู้ใช้เป็นหลัก (User-based collaborative filtering, UCF) 2. ขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ Personalized POI recommendation using memory-based preferences and POI stickiness (U-CF-MS) 3. ขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ Hypertext-induced topic-search-based recommendation (HITS) และ 4. ขั้นตอนวิธีการสร้างรายการแนะนำสถานที่ Tensor decomposition-based collaborative filtering (TDCF) สำหรับขั้นตอนการประเมินผลความถูกต้องของรายการแนะนำสถานที่ ประกอบด้วย 4 ขั้นตอน คือ 1. ขั้นตอนวิธี Precision 2. ขั้นตอนวิธี Recall 3. ขั้นตอนวิธี NDCG และ 4. ขั้นตอนวิธี Matching Ratio และสุดท้ายการสรุปผลการทดลองประสิทธิภาพความถูกต้องและการอภิปรายผลจะถูกนำเสนอ

4.1 การตั้งค่าการทดลองและการประเมินผลประสิทธิภาพการทดลอง

4.1.1 ชุดข้อมูลจัดเก็บประวัติการเช็คอินของผู้ใช้

ชุดข้อมูลที่ใช้ในการทดลองจะทำการจัดเก็บประวัติการเช็คอินของผู้ใช้จากแอปพลิเคชัน Foursquare (Yang et al., 2014) ประกอบด้วย 1. รหัสผู้ใช้ 2. รหัสสถานที่ 3. รหัสหมวดหมู่สถานที่ 4. หมวดหมู่สถานที่ 5. ละติจูด 6. ลองจิจูด 7. เวลาในการเช็คอิน 8. เวลา UTC โดยจัดเก็บในช่วง 12 เมษายน 2012 ถึง 16 กุมภาพันธ์ 2013 (ประมาณ 10 เดือน) แสดงตัวอย่างของฐานข้อมูล Foursquare ที่ใช้ในการทดลอง ดังภาพที่ 10

รหัสผู้ใช้	รหัสสถานที่	รหัสหมวดหมู่สถานที่	หมวดหมู่สถานที่	ละติจูด	ลองจิจูด	เวลาในการเช็คอิน	เวลา UTC
1386	4b64f0e3f964a5203cdb2ae3	4bf58dd8d48988d16e941735	Fast Food Restaurant	35.577	19.660	540	Tue Apr 03 21:33:17 +0000 2012
2243	4f534895e4b006151584808a	4bf58dd8d48988d1c4941735	Restaurant	35.627	39.714	40	Tue Apr 03 21:50:15 +0000 2012
1064	4b6fe136f964a520b4fe2ce3	4bf58dd8d48988d1e0931735	Coffee Shop	35.689	39.700	40	Tue Apr 03 22:13:21 +0000 2012
1248	4b72c558f964a520d2872de3	4bf58dd8d48988d111941735	Japanese Restaurant	35.652	39.544	540	Tue Apr 03 22:23:00 +0000 2012
1540	4b5829f4f964a520a44c28e3	4bf58dd8d48988d1e0931735	Coffee Shop	35.691	39.703	40	Tue Apr 03 22:23:18 +0000 2012
1964	4b5e4928f964a520c68729e3	4bf58dd8d48988d16e941735	Fast Food Restaurant	35.646	39.745	540	Tue Apr 03 22:25:30 +0000 2012
2243	4f534895e4b006151584808a	4bf58dd8d48988d1c4941735	Restaurant	35.627	39.7143	40	Tue Apr 03 22:26:33 +0000 2012
9655	4cde2fce41ed224b1e97d73c	4d4ae6fc7a7b7dea34424761	Fried Chicken Joint	35.789	39.661	540	Tue Apr 03 22:27:30 +0000 2012
1100	4b5a3b6cf964a52026b628e3	4bf58dd8d48988d16e941735	Fast Food Restaurant	35.574	139.658	540	Tue Apr 03 22:27:54 +0000 2012
1320	4df7423e483b96f73159a256	4bf58dd8d48988d111941735	Japanese Restaurant	35.598	39.6669	540	Tue Apr 03 22:29:18 +0000 2012

ภาพที่ 10 ตัวอย่างของฐานข้อมูล Foursquare ที่ใช้ในการทดลอง

ในประวัติการเช็คอินของผู้ใช้จะทำการจัดเก็บในหลายหมวดหมู่ของสถานที่ เช่น หมวดหมู่อาหาร (ได้แก่ ร้านอาหาร คาเฟ่ บาร์ เป็นต้น) หมวดหมู่การขนส่ง (ได้แก่ สถานีรถบัส สถานีรถไฟ สถานีรถไฟฟ้าใต้ดิน เป็นต้น) และหมวดหมู่กลางคืน (ได้แก่ สถานบันเทิงยามค่ำคืน เป็นต้น) อย่างไรก็ตาม ทางผู้วิจัยสังเกตได้ว่าจำนวนการเช็คอินโดยส่วนมากคือสถานที่ที่อยู่ในหมวดหมู่ขนส่ง ที่ซึ่งเป็นการเช็คอินที่เกี่ยวข้องกับกิจวัตรประจำวันของผู้ใช้ เช่น การเดินทางไปทำงานหรือไปโรงเรียนมากกว่าสถานที่ที่ไปในโอกาสพิเศษหรือผู้ใช้ที่มีการท่องเที่ยวเพื่อต้องการค้นหาสถานที่ใหม่ ๆ ดังนั้นในงานวิจัยนี้จึงเลือกสถานที่ที่เป็นหมวดหมู่อาหารเท่านั้นเพื่อมุ่งเน้นไปที่การแนะนำสถานที่ที่น่าสนใจที่เป็นการแนะนำประสบการณ์ใหม่ ๆ ให้กับผู้ใช้ โดยในการทดลองของขั้นตอนวิธีการที่นำเสนอได้มีการพิจารณาความโดดเด่นของสถานที่ที่ผู้ใช้ชื่นชอบที่ซึ่งแสดงถึงความเชี่ยวชาญในพื้นที่ของผู้ใช้ ดังนั้นชุดข้อมูลได้มีการแบ่งชุดข้อมูลออกเป็น 5 พื้นที่ที่ซึ่งมประวัติการเช็คอินของผู้ใช้มากที่สุดให้หมวดหมู่อาหาร ได้แก่ 1. เมือง Chiyoda 2. เมือง Shibuya 3. เมือง Minato 4. เมือง Shinjuku และเมือง 5. เมือง Kawasaki โดยใช้การดึงข้อมูลพื้นที่จาก Geocoding API¹ แสดงคุณลักษณะของเมืองต่าง ๆ ที่ถูกใช้ในการทดลอง ดังตารางที่ 2

ตารางที่ 2 คุณลักษณะของเมืองต่าง ๆ ที่ถูกใช้ในการทดลอง

พื้นที่	ผู้ใช้	สถานที่	จำนวนเช็คอินทั้งหมด	จำนวนเช็คอินต่อสถานที่	จำนวนเช็คอินต่อผู้ใช้
Chiyoda	1702	2826	15029	5.32	8.83
Shibuya	1352	2348	9051	3.85	6.69
Minato	1264	2556	8695	3.40	6.89
Shinjuku	1240	2222	8413	3.79	6.78
Kawasaki	537	1119	4259	3.81	7.93
รวมทั้งหมด	2169	11071	45447	4.11	20.95

ตารางที่ 2 แสดงถึงคุณลักษณะของฐานข้อมูลที่ใช้ในการทดลอง ประกอบด้วยผู้ใช้ทั้งหมดจำนวน 2,169 ผู้ใช้และสถานที่หมวดหมู่อาหารในเมืองโตเกียวทั้งหมดจำนวน 11,071 สถานที่ ที่ซึ่งแสดงให้เห็นว่าเมือง Chiyoda มีจำนวนการเช็คอินรวมสูงที่สุดรวมถึงค่าเฉลี่ยจำนวนการเช็คอินต่อ

¹ <https://developers.google.com/maps/documentation/geocoding/>

สถานที่และค่าเฉลี่ยจำนวนการเช็คอินต่อผู้ใช้ ขณะที่เมือง Shibuya Minato และ Shinjuku มีจำนวนการเช็คอินที่ใกล้เคียงกัน และสุดท้ายเมือง Kawasaki มีจำนวนการเช็คอินที่น้อยที่สุด

ในการทดลองที่นำเสนอนี้ ได้ประยุกต์ใช้ขั้นตอนวิธีการสุ่มข้อมูลแบบเที่ยงตรง 5 กลุ่ม (5-fold cross validation) โดยที่ 20% ของประวัติการเช็คอินของผู้ใช้จะถูกสุ่มเป็นข้อมูลทดสอบและ 80% จะถูกสุ่มเป็นข้อมูลสำหรับฝึกฝน

4.1.2 ขั้นตอนวิธีการที่ใช้ในการเปรียบเทียบ

เนื่องจากขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 เป็นขั้นตอนวิธีการแบบผสมผสานโดยใช้ขั้นตอนวิเคราะห์การเชื่อมโยงที่เรียกว่า HITS-3 และขั้นตอนการกรองร่วมที่เรียกว่า CF-3 ดังนั้นในงานวิจัยนี้จึงได้เปรียบเทียบกับวิธีการก่อนหน้าที่ซึ่งมีพื้นฐานการสร้างรายการแนะนำจากขั้นตอนวิธีการกรองร่วม คือ 1. ขั้นตอนวิธีการกรองร่วมโดยพิจารณาผู้ใช้เป็นหลัก (User-based collaborative filtering, UCF) (Yuan et al., 2013) ซึ่งเป็นขั้นตอนวิธีพื้นฐานของวิธีการกรองร่วมในการสร้างรายการแนะนำ 2. ขั้นตอนวิธี Personalized POI recommendation using memory-based preferences and POI stickiness (U-CF-MS) (Zhang et al., 2021) ซึ่งเป็นการปรับปรุงขั้นตอนวิธีการกรองร่วมล่าสุดโดยพิจารณาความถี่ในการเช็คอินร่วมกับการเช็คอินซ้ำ และขั้นตอนวิเคราะห์การเชื่อมโยง คือ 3. ขั้นตอนวิธี Hypertext-induced topic-search-based recommendation (HITS) (Bao et al., 2015) ซึ่งเป็นขั้นตอนวิธีพื้นฐานของขั้นตอนวิเคราะห์การเชื่อมโยง และ 4. ขั้นตอนวิธี Tensor decomposition-based collaborative filtering (TDCF) (Yin et al., 2021) ซึ่งเป็นวิธีล่าสุดของขั้นตอนวิธีการแบบผสมผสานโดยใช้ขั้นตอนวิเคราะห์การเชื่อมโยงและขั้นตอนการกรองร่วม รายละเอียดของขั้นตอนวิธีดังกล่าวและขั้นตอนวิธีที่นำเสนอถูกอธิบาย ดังนี้

1. ขั้นตอนวิธีการกรองร่วมโดยพิจารณาผู้ใช้เป็นหลัก (UCF) (Yuan et al., 2013) ขั้นตอนแรกของวิธีนี้จะทำการสร้างเมตริกซ์ความถี่ในการเช็คอินระหว่างผู้ใช้และสถานที่ จากนั้นการค้นหาความคล้ายระหว่างผู้ใช้เป้าหมายและผู้ใช้คนอื่น ๆ ในระบบโดยพิจารณาสถานที่ที่เช็คอินร่วมกันจะถูกคำนวณโดยใช้ขั้นตอนวิธีโคไซน์ (Cosine similarity) เพื่อสร้างการทำนายค่าคะแนนความชอบของสถานที่ต่อไป
2. ขั้นตอนวิธีการแนะนำสถานที่ Personalized POI recommendation using memory-based preferences and POI stickiness (U-CF-MS) (Zhang et al., 2021) ขั้นตอนวิธีการกรองข้อมูลร่วมโดยพิจารณาผู้ใช้เป็นหลักจะถูกปรับปรุงโดย

พิจารณาการเช็คคินล่าสุดของสถานที่และการเช็คคินซ้ำของสถานที่ที่เรียกว่าขั้นตอนวิธี POI stickiness ในขั้นแรกความถี่ในการเช็คคินจะถูกให้ค่าน้ำหนักด้วยขั้นตอนวิธี POI stickiness จากนั้นการหาความคล้ายด้วยวิธีโคไซน์จะถูกคำนวณเพื่อทำนายค่าสถานที่ที่น่าสนใจเพื่อสร้างรายการแนะนำสถานที่ให้กับผู้ใช้

3. ขั้นตอนวิธี Hypertext-induced topic-search-based recommendation (HITS) (Bao et al., 2015) ขั้นตอนที่แรกจะทำการสร้างโครงสร้างกราฟระหว่างผู้ใช้และสถานที่เพื่อค้นหาสถานที่ที่น่าสนใจ โดยสถานที่ที่น่าสนใจเกิดจากผลรวมค่าน้ำหนักของความถี่และค่าคะแนนของผู้ใช้ และผู้ใช้ที่มีความเชี่ยวชาญในพื้นที่เกิดจากผลรวมค่าน้ำหนักของความถี่และค่าคะแนนของสถานที่ที่ผู้ใช้คนนั้นเช็คคิน โดยขั้นตอนเหล่านี้จะเป็นกระบวนการทำซ้ำเพื่อสร้างรายการแนะนำสถานที่ให้กับผู้ใช้
4. ขั้นตอนวิธี Tensor decomposition-based collaborative filtering (TDCF) (Yin et al., 2021) ขั้นตอนวิธีแบบผสมผสานโดยในขั้นตอนการกรองร่วมโดยใช้วิธีการแยกเมทริกซ์ด้วยการพิจารณาผู้ใช้ หมวดยุทธศาสตร์ที่แทนที่การพิจารณาสถานที่โดยตรงและช่วงเวลาในการเช็คคิน และในส่วนของขั้นตอนวิเคราะห์การเชื่อมโยงโดยปรับปรุงขั้นตอนวิธี HITS ด้วยการพิจารณาความนิยมของสถานที่ร่วมกับระยะทางที่ผู้ใช้มักมีการเช็คคิน (Active area) จากนั้นทั้งสองวิธีจะถูกรวมค่าคะแนนความน่าสนใจของสถานที่ด้วยการให้ค่าถ่วงน้ำหนักเพื่อสร้างรายการแนะนำสถานที่ให้กับผู้ใช้
5. ขั้นตอนวิธีที่นำเสนอ LACF-Rec3 ขั้นตอนวิธีแบบผสมผสานโดยใช้การวิเคราะห์การเชื่อมโยงและขั้นตอนการกรองร่วม ซึ่งพิจารณาพฤติกรรมเช็คคินในสามแง่มุม คือ ความถี่ในการเช็คคิน ความหลากหลายในการเช็คคินและการเช็คคินซ้ำของทั้งผู้ใช้และสถานที่ โดยสร้างรายการแนะนำสถานที่ให้กับผู้ใช้ใหม่และผู้ใช้ที่มีประวัติการเช็คคิน สำหรับผู้ใช้ที่มีประวัติการเช็คคิน ขั้นตอนวิธี CF-3 ได้ทำการหาความคล้ายระหว่างความโดดเด่นของสถานที่ที่ผู้ใช้ชื่นชอบเพื่อระบุผู้ใช้ที่เกี่ยวข้อง และใช้ประโยชน์จากเซตของผู้ใช้ที่เกี่ยวข้องเพื่อสร้างรายการแนะนำเฉพาะบุคคลให้กับผู้ใช้ สำหรับผู้ใช้ใหม่ ขั้นตอนวิธี HITS-3 โดยพิจารณาพฤติกรรมเช็คคินทั้งสามแง่มุมดังกล่าวจะถูกใช้เพื่อค้นหาความโดดเด่นในแง่มุมต่าง ๆ ในแต่ละพื้นที่โดยไม่ต้องพิจารณาประวัติการเช็คคินของผู้ใช้เป้าหมาย

4.1.3 วิธีการประเมินผลประสิทธิภาพความถูกต้อง

ในงานวิจัยนี้ได้เปรียบเทียบประสิทธิภาพความถูกต้องในสองแง่มุม คือ ความถูกต้องของรายการแนะนำสถานที่และความถูกต้องของอันดับรายการแนะนำสถานที่ แต่อย่างไรก็ตามสถานที่ปัจจุบันของผู้ใช้เป้าหมายไม่ได้ถูกระบุในฐานข้อมูล ดังนั้นขั้นตอนแรกของการประเมินผลคือการระบุสถานที่ปัจจุบันของผู้ใช้ด้วยขั้นตอนวิธี Minimum Bounding Rectangle (MBR) (Bao et al., 2012) ซึ่งวิธีการนี้ได้สร้างขอบเขตการพิจารณาสถานที่สำหรับแนะนำให้กับผู้ใช้ด้วยสถานที่ทดสอบ (ขอบเขตการพิจารณาแสดงดั่งเส้นประที่ซึ่งได้จากสถานที่ A และสถานที่ B และสถานที่ปัจจุบันของผู้ใช้คือจุดสีเขียว) แสดงดั่งภาพที่ 11



ภาพที่ 11 การสร้างขอบเขตการพิจารณาสถานที่ด้วยขั้นตอนวิธี MBR

จากภาพแสดงให้เห็นถึงทุกสถานที่ภายในขอบเขตการพิจารณา (ทั้งสีแดงและสีดำ) ถูกระบุเป็นสถานที่แข่งขันเพื่อคัดเลือกเป็นสถานที่แนะนำสำหรับผู้ใช้ สถานที่สีแดงแสดงถึงสถานที่ทดสอบ (Ground truth locations) สถานที่สีดำแสดงถึงสถานที่ทั่วไป สถานที่ที่ถูกคาดสีเขียวคือสถานที่ที่ถูกแนะนำให้กับผู้ใช้ และสถานที่สีแดงที่ถูกคาดสีเขียวคือสถานที่ทดสอบซึ่งถูกแนะนำให้กับผู้ใช้แสดงถึงสถานที่ที่ถูกต้องในการแนะนำ (Recommended Ground Truth locations)

สำหรับวิธีการประเมินผลประสิทธิภาพความถูกต้องของรายการแนะนำสถานที่ในงานวิจัยนี้ได้ใช้ขั้นตอนวิธี Precision และขั้นตอนวิธี Recall (Baral & Li, 2018) (Zhang et al., 2021) โดยที่ขั้นตอนวิธี Precision คือการพิจารณาสัดส่วนของจำนวนสถานที่ทดสอบที่ถูกแนะนำ (number of Recommended Ground Truths) ต่อจำนวนรายการสถานที่ที่แนะนำ (number of Recommended POIs) และขั้นตอนวิธี Recall คือการพิจารณาสัดส่วนจำนวนสถานที่ทดสอบที่ถูก

แนะนำต่อจำนวนสถานที่ทดสอบทั้งหมด (number of Ground Truths) แสดงดังสมการที่ 4.1 และสมการที่ 4.2 ตามลำดับ

$$\text{Precision} = \frac{\text{number of Recommended Ground Truths}}{\text{number of Recommendations}} \quad (4.1)$$

$$\text{Recall} = \frac{\text{number of Recommended Ground Truths}}{\text{number of Ground Truths}} \quad (4.2)$$

สำหรับวิธีการประเมินผลประสิทธิภาพความถูกต้องของอันดับสถานที่ในรายการแนะนำโดยใช้ขั้นตอนวิธี NDCG (Normalized Discounted Cumulative Gain) (Wang et al., 2013) ที่ซึ่งเป็นการพิจารณาอันดับของสถานที่ทดสอบที่ถูกแนะนำ โดยการให้ค่าน้ำหนักกับอันดับของสถานที่ที่แนะนำถูก ซึ่งค่าน้ำหนักที่มีค่ามากแสดงถึงอันดับรายการที่แนะนำถูกในอันดับต้น ๆ และให้ค่าน้ำหนักที่ลดลงมาในอันดับที่แนะนำถูกในอันดับที่ต่ำกว่า แสดงดังสมการที่ 4.3 สมการที่ 4.4 และสมการที่ 4.5 ตามลำดับ

$$NDCG = \frac{DCG}{IDCG} \quad (4.3)$$

$$DCG = \sum_{i=1}^N \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (4.4)$$

$$IDCG = \sum_{i=1}^{|POI_{REL}|} \frac{2^{rel_i} - 1}{\log_2(i+1)} \quad (4.5)$$

กำหนดให้ N คือ จำนวนรายการสถานที่

rel_i คือ อันดับที่ i^{th} ของสถานที่ในรายการแนะนำ

POI_{REL} คือ สถานที่ตามอันดับที่ดีที่สุดในการแนะนำ

สำหรับตัววัดประสิทธิภาพระบบแนะนำสถานที่วิธีสุดท้าย ซึ่งเป็นวิธีการวัดที่ทางผู้วิจัยนำเสนอที่เรียกว่า Matching ratio ในการวัดคุณภาพของรายการแนะนำสถานที่เปรียบเทียบกับความชอบของผู้ใช้ จากปัญหาที่ว่าตัววัดประสิทธิภาพด้วยวิธี Precision และ Recall อาจไม่ได้สะท้อนประสิทธิภาพของระบบแนะนำในการทดลอง เนื่องจากชุดข้อมูลประวัติการเช็คอินที่ใช้ในการทดลองเป็นประวัติการเช็คอินที่ถูกจัดเก็บไว้ก่อนแล้ว ดังนั้นรายการแนะนำสถานที่ที่ถูกพัฒนาขึ้นจึงไม่ได้มีผลต่อผู้ใช้ทดสอบ กล่าวคือผู้ใช้ทดสอบไม่ได้รับรายการแนะนำสถานที่ที่แท้จริง ดังนั้นการวัดประสิทธิภาพด้วย Precision และ Recall กับสถานที่ทดสอบ (Ground truth POIs) ของผู้ใช้อาจจะไม่ได้แสดงถึงความชอบของผู้ใช้ เนื่องจากผู้ใช้ทดสอบไม่มีปฏิสัมพันธ์กับรายการแนะนำสถานที่และสถานที่ที่ผู้ใช้ชื่นชอบที่ปรากฏในรายการแนะนำสถานที่อาจไม่ได้เป็นสถานที่ทดสอบ ส่งผลให้สองตัววัดนี้อาจไม่ได้สะท้อนถึงประสิทธิภาพของระบบแนะนำ ภายใต้กรณีดังนี้ 1. ผู้ใช้ทดสอบอาจจะเยี่ยมชมสถานที่ที่ปรากฏในรายการแนะนำสถานที่แต่ไม่ได้เช็คอิน 2. สถานที่ที่แนะนำอาจจะเป็นสถานที่ที่ผู้ใช้ชื่นชอบแต่เมื่อไม่ได้มีการแนะนำที่แท้จริงแล้วส่งผลให้ผู้ใช้ไม่รู้ว่ามีส่วนที่ดังกล่าวอยู่

จากข้อจำกัดดังกล่าว ทางผู้วิจัยจึงได้นำเสนอขั้นตอนวิธี Matching ratio ในการวัดความชอบของผู้ใช้กับรายการแนะนำสถานที่ สามารถคำนวณได้จาก 1. จำนวนสถานที่ที่มีชนิดความโดดเด่นตรงกับความโดดเด่นของผู้ใช้ (number of POIs with matched distinctiveness) 2. จำนวนรายการแนะนำสถานที่ (number of Recommended POIs) ซึ่งอัตราส่วนของขั้นตอนวิธีนี้จะอยู่ในช่วง 0-100% ค่าที่มากแสดงถึงผู้ใช้ทดสอบมีความชอบต่อสถานที่ที่แนะนำโดยพิจารณาจากความโดดเด่นของสถานที่นั้น ๆ แสดงดังสมการที่ 4.7

$$\text{Matching ratio} = \frac{\text{number of POIs with matched distinctiveness}}{\text{number of Recommended POIs}} \times 100 \quad (4.7)$$

4.2 ผลการทดลอง

สำหรับผลการทดลอง ทางผู้วิจัยได้ทำการเปรียบเทียบขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 กับขั้นตอนวิธีการสร้างรายการแนะนำก่อนหน้า ในแต่ละจำนวนสถานที่ที่แนะนำ ประกอบด้วย อันดับ 5 สถานที่ที่น่าสนใจจนถึง 20 สถานที่ที่น่าสนใจสำหรับผู้เข้าเป้าหมายใน 6 พื้นที่การทดลอง ได้แก่ เมือง Tokyo 2. เมือง Chiyoda 3. เมือง Shinjuku 4. เมือง Minato 5. เมือง Shibuya 6. เมือง Kawasaki โดยการทดลองนี้ได้ทำการประเมินผลโดยแบ่งประเภทผู้ใช้เป็น ผู้ใช้ใหม่ และผู้ใช้ที่มีประวัติการเช็คอิน

4.2.1 การแนะนำสถานที่สำหรับผู้ใช้ใหม่

ในการประเมินผลสถานที่ที่แนะนำสำหรับผู้ใช้ใหม่ ที่ซึ่งการสร้างรายการแนะนำสถานที่ในผู้ใช้กลุ่มนี้จะมีความยากเนื่องจากผู้ใช้ไม่มีประวัติการเช็คอินหรือมีจำนวนการเช็คอินน้อยครั้ง ดังนั้นจึงไม่สามารถใช้ข้อมูลผู้ใช้ในการสร้างรายการแนะนำสถานที่เฉพาะบุคคลได้ ด้วยเหตุนี้ ขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 จึงได้มีการค้นหาสถานที่ที่น่าสนใจซึ่งตั้งอยู่ใกล้เคียงกับที่อยู่ปัจจุบันของผู้ใช้ โดยการพิจารณาความโดดเด่นของสถานที่ในหลาย ๆ แ่งมุม แสดงดังภาพที่ 12-15 จากผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธีการที่นำเสนอมีประสิทธิภาพความถูกต้องของรายการแนะนำที่ดีกว่าขั้นตอนวิธี HITS ที่ 63.30% และ 59.21% โดยใช้ขั้นตอนการวัดประสิทธิภาพ Precision และ Recall ตามลำดับ

ด้วยเหตุที่ว่า งานวิจัยส่วนมากได้มีการพัฒนาขั้นตอนวิธีการแนะนำสถานที่โดยพิจารณาเฉพาะความถี่ในการเช็คอินเป็นหลัก เช่นเดียวกับขั้นตอนวิธีการ HITS ส่งผลให้สถานที่ที่แนะนำมักเป็นสถานที่ที่โดดเด่นทางด้านความถี่เท่านั้น อย่างไรก็ตามผู้ใช้นั้นมีความชอบที่หลากหลายมากกว่าสถานที่ที่ได้รับความนิยมทางด้านความถี่เพียงอย่างเดียว โดยบางผู้ใช้นั้นมีความชื่นชอบในการเที่ยวชมสถานที่ในแ่งมุมอื่น ๆ เช่น การท่องเที่ยวสถานที่ตามผู้เชี่ยวชาญในพื้นที่ ที่ซึ่งอาจไม่ใช่สถานที่ที่มีความถี่สูง เป็นต้น

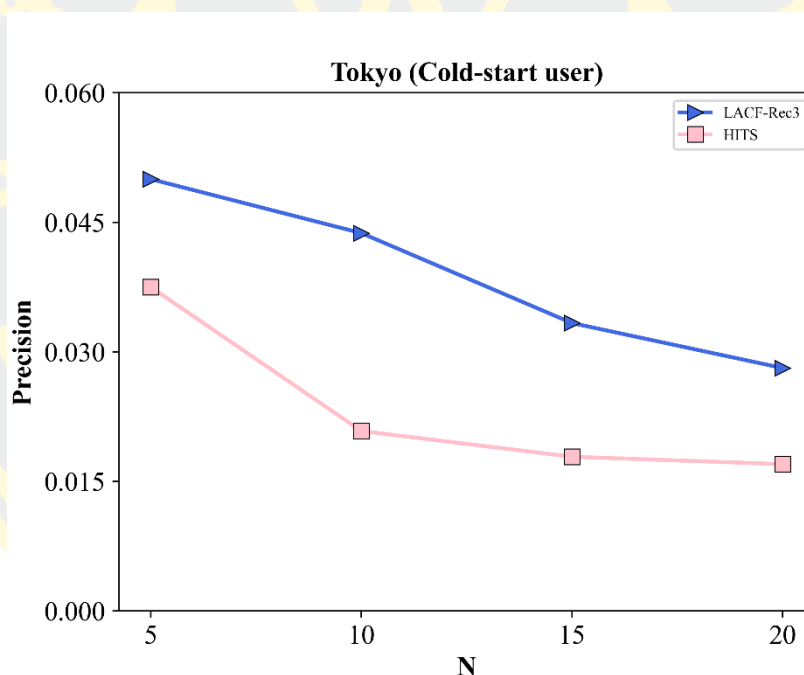
จากที่กล่าวมาแสดงให้เห็นว่าแต่ละผู้ใช้นั้นมีความชอบที่แตกต่างกัน ดังนั้นในการสร้างรายการแนะนำสถานที่ที่เหมาะสมที่สุดให้กับผู้ใช้ใหม่ซึ่งไม่สามารถค้นหาความชอบของผู้ใช้ได้ จึงควรประกอบด้วยสถานที่ที่มีความโดดเด่นในหลายแ่งมุมที่แตกต่างกันในรายการแนะนำสถานที่ เช่น สถานที่ที่มีความถี่ในการเช็คอินสูง หรือ สถานที่ที่มักถูกเที่ยวชมโดยผู้ใช้ที่หลากหลาย เป็นต้น

ดังนั้นขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 จึงได้มีการเลือกสถานที่ที่โดดเด่นในหลายแ่งมุมเพื่อสร้างรายการแนะนำสถานที่ให้ตรงกับความชอบของผู้ใช้มากที่สุด เช่น สถานที่ที่มีการเที่ยวชมสูงและมีจำนวนผู้เที่ยวชมที่หลากหลายสูงแต่มีการเที่ยวชมซ้ำน้อย สถานที่ที่มีการเที่ยวชมปานกลางแต่มีจำนวนผู้จำนวนผู้เที่ยวชมที่หลากหลายสูงและมีการเที่ยวชมซ้ำน้อยและสถานที่ที่มักมีจำนวนการเที่ยวชมซ้ำสูง เป็นต้น

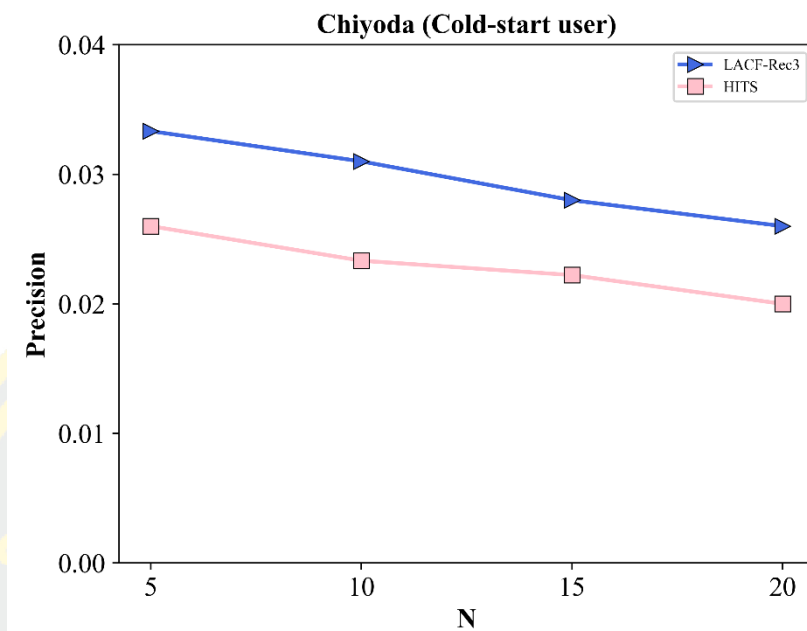
นอกจากนี้สถานที่ที่แนะนำได้ถูกเลือกในอันดับที่สูงกว่าขั้นตอนวิธีการก่อนหน้า โดยการประเมินผลอันดับของรายการแนะนำสถานที่โดยใช้วิธีการประเมินผล NDCG แสดงให้เห็นว่าขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 มีความถูกต้องในการจัดอันดับที่ดีกว่าขั้นตอนวิธี HITS ที่ 8.88%

สำหรับการประเมินผลด้วยวิธีการประเมินผล Matching ratio จากผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธีการที่นำเสนอมีความสามารถในการแนะนำได้ตรงตามความชอบของผู้ใช้ในช่วง

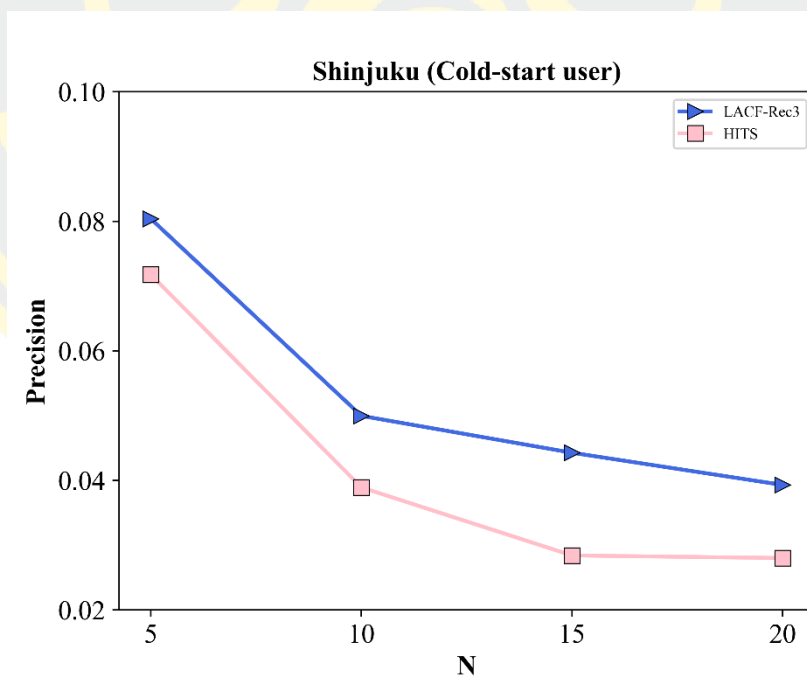
ประมาณ 25% ถึง 50% ที่ซึ่งมีความถูกต้องที่มากกว่าขั้นตอนวิธี U-CF-MS ขั้นตอนวิธี TDCF ขั้นตอนวิธี HITS และขั้นตอนวิธี UCF ที่ 521.51% 193.78% 17.82% และ 793.39% ตามลำดับ ที่ซึ่งในกรณีที่ดีที่สุด ขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 สามารถแนะนำได้ตรงตามความชอบของผู้ใช้ที่ประมาณ 50% ดังนั้นสถานที่แนะนำแต่ผู้ใช้ไม่ได้เกี่ยวข้องคือสถานที่ในแง่มุมอื่น ๆ ที่ไม่ตรงกับความต้องการของผู้ใช้เป้าหมาย ด้วยเหตุที่ขั้นตอนวิธีการที่นำเสนอโดยการใช้ขั้นตอนวิธีการ HITS-3 มีการระบุความโดดเด่นของสถานที่เพียงหนึ่งแง่มุม แต่ในบางสถานที่อาจมีความโดดเด่นในหลายแง่มุม เช่น บางสถานที่ควรมีความโดดเด่นทั้งด้านความถี่และการเข้าถึง ถ้าสถานที่ที่มีการเกี่ยวข้องโดยผู้ใช้งานจำนวนมากพร้อมกับมักมีการกลับมาใช้บริการซ้ำจำนวนมากด้วย เป็นต้น ดังนั้นงานในอนาคตจะมีการปรับปรุงขั้นตอนการระบุความโดดเด่นของสถานที่ด้วยการพิจารณาหลายแง่มุมความโดดเด่นของสถานที่พร้อมกัน



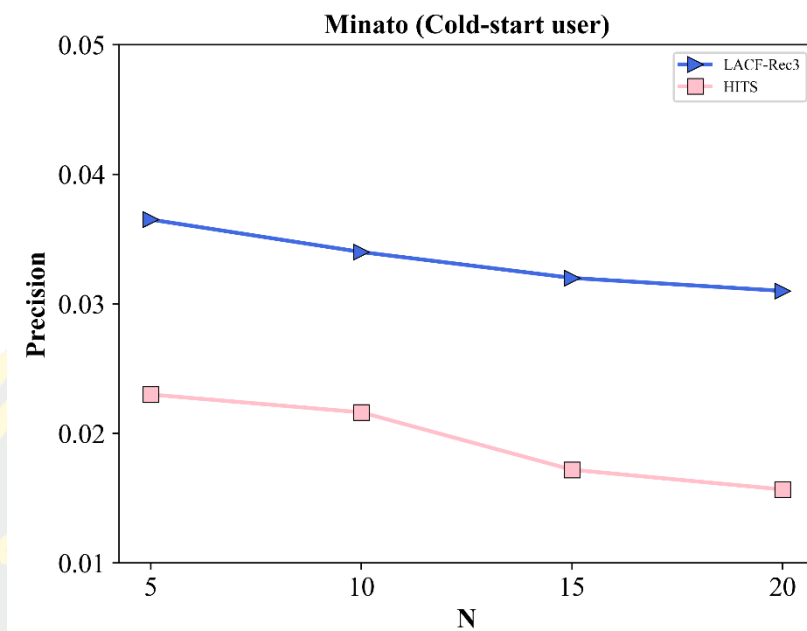
ภาพที่ 12 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo



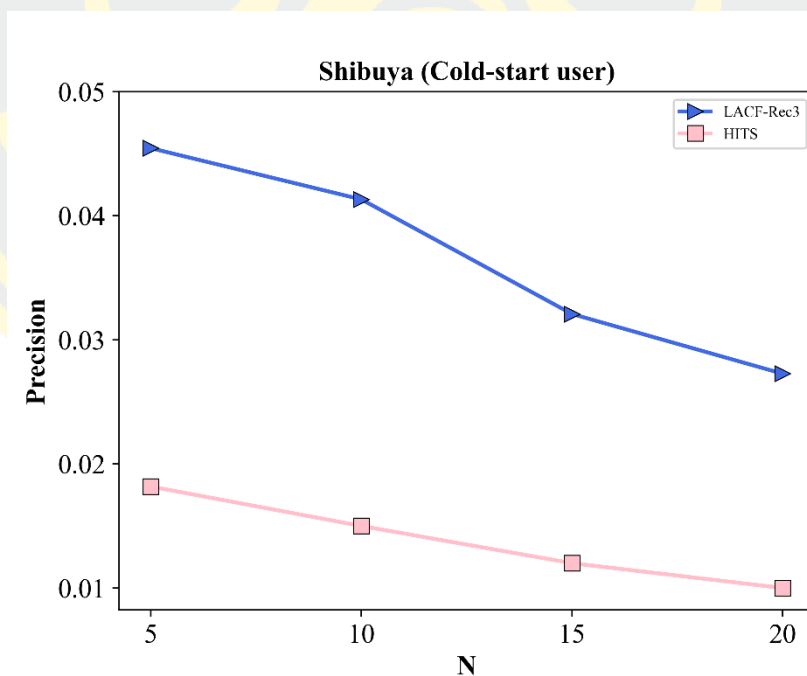
ภาพที่ 13 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda



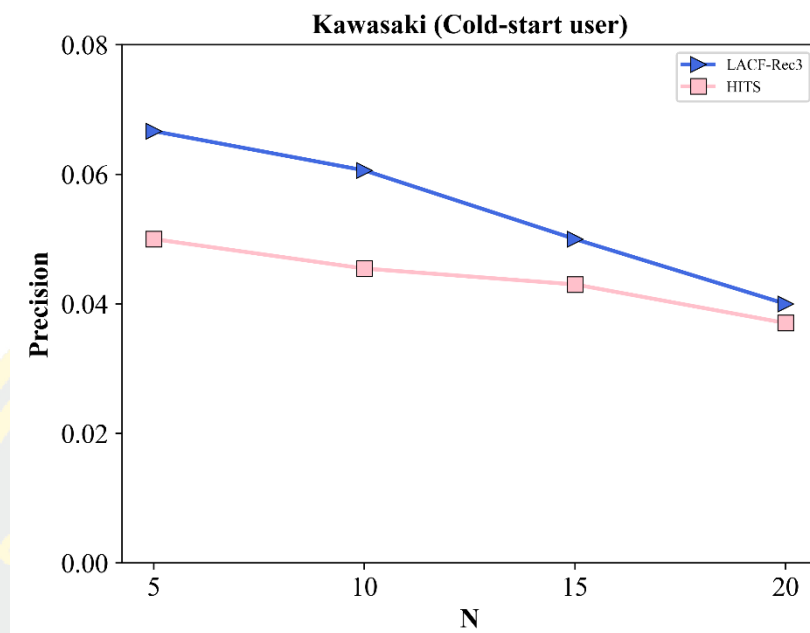
ภาพที่ 14 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku



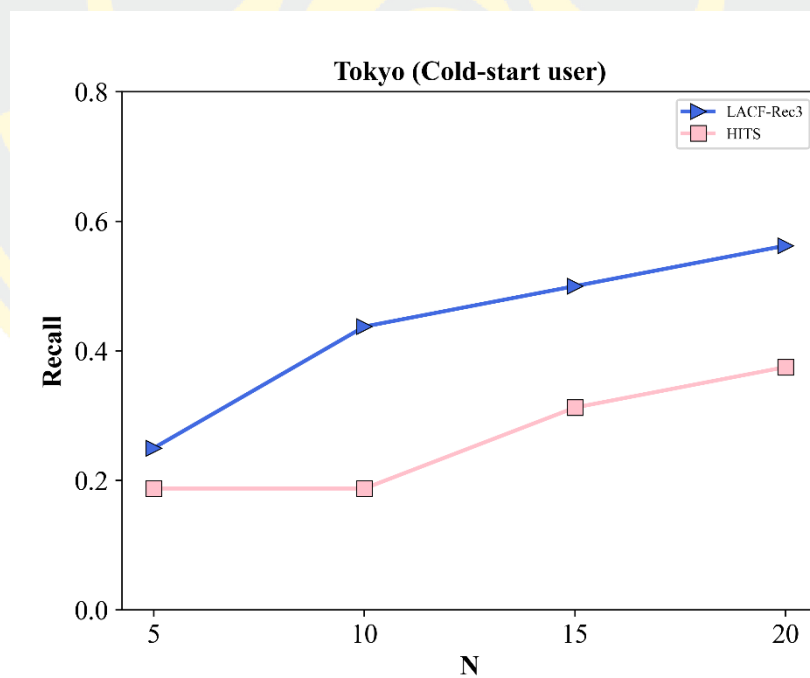
ภาพที่ 15 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato



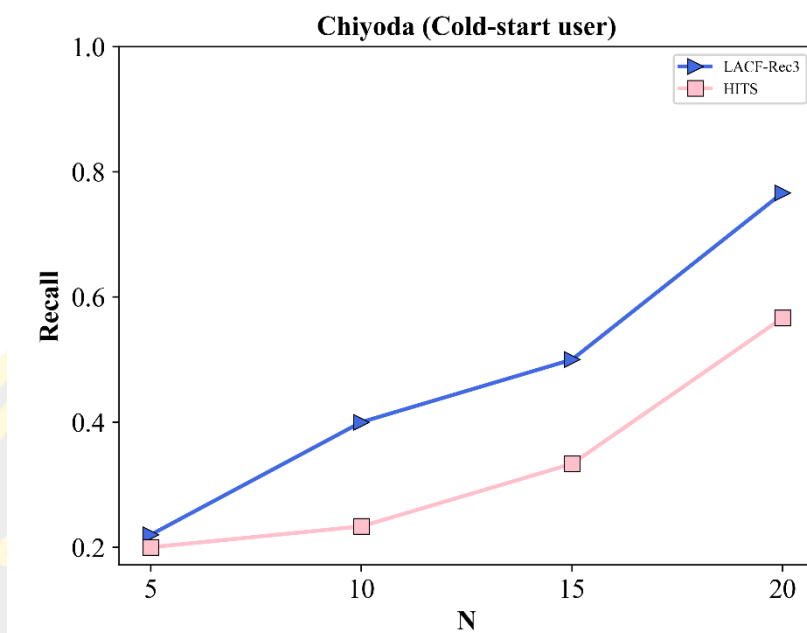
ภาพที่ 16 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya



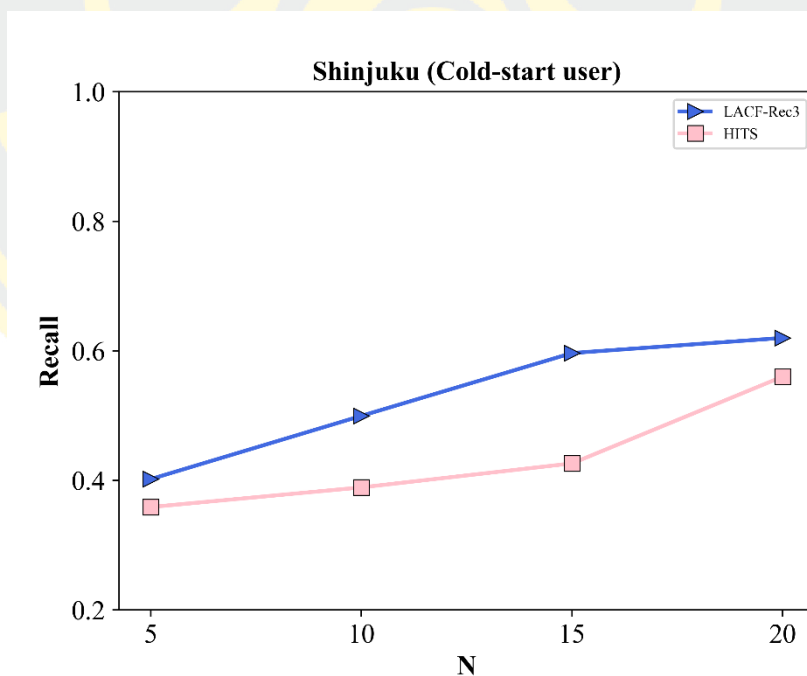
ภาพที่ 17 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki



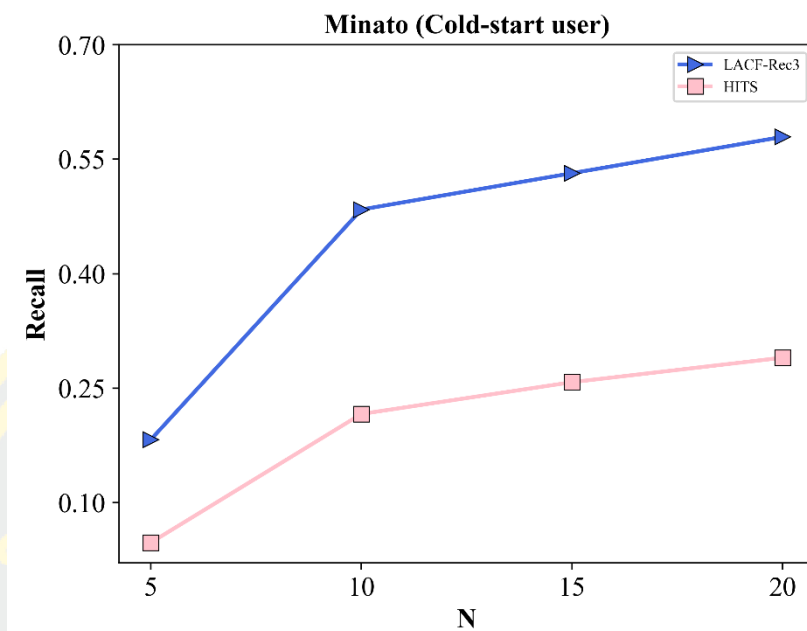
ภาพที่ 18 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo



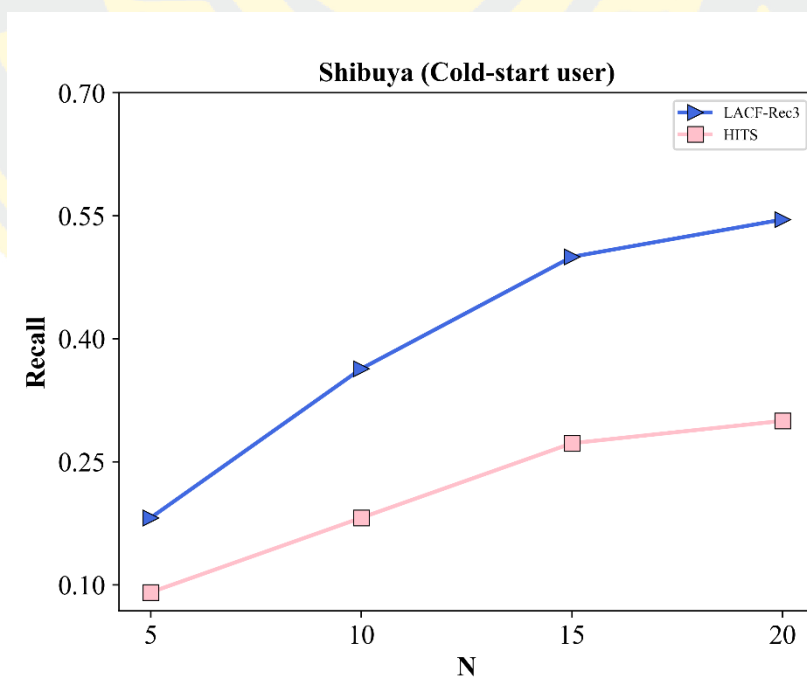
ภาพที่ 19 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda



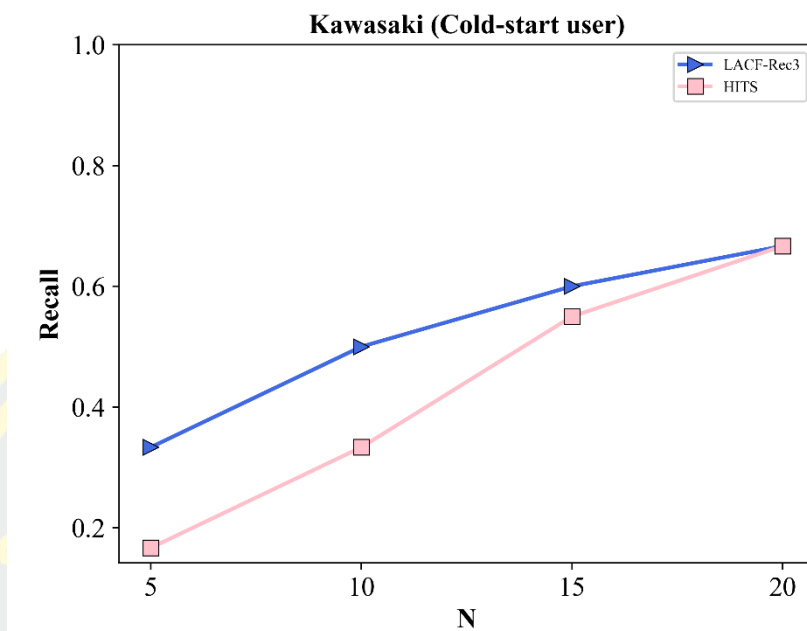
ภาพที่ 20 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku



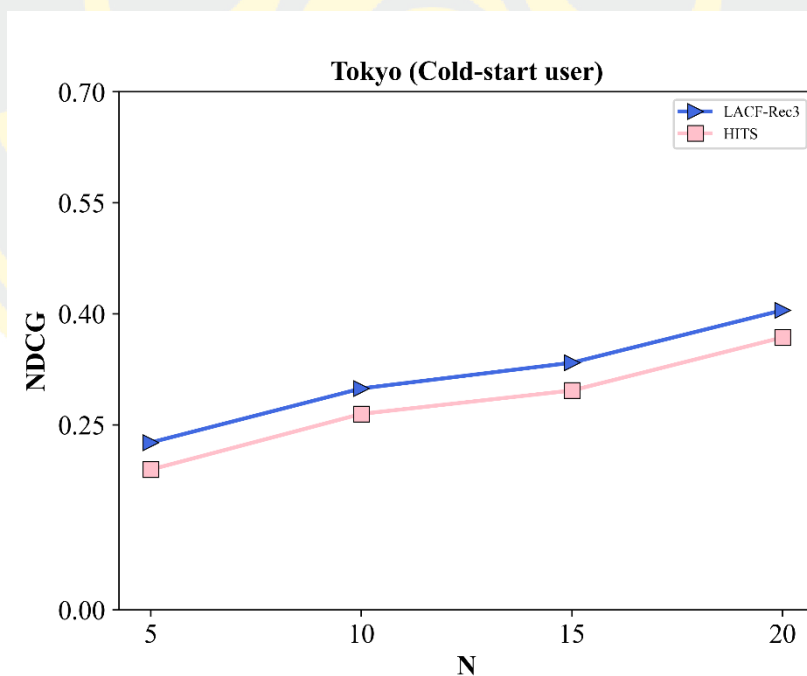
ภาพที่ 21 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato



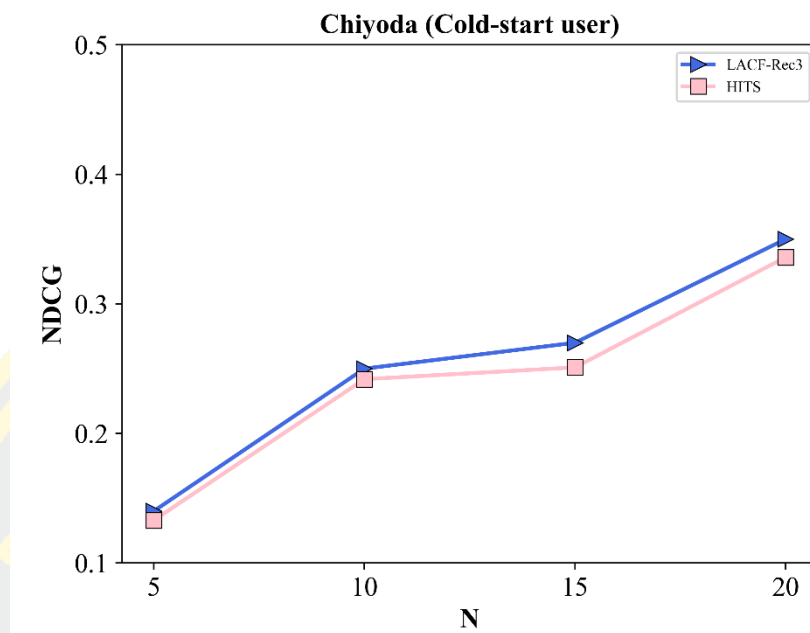
ภาพที่ 22 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya



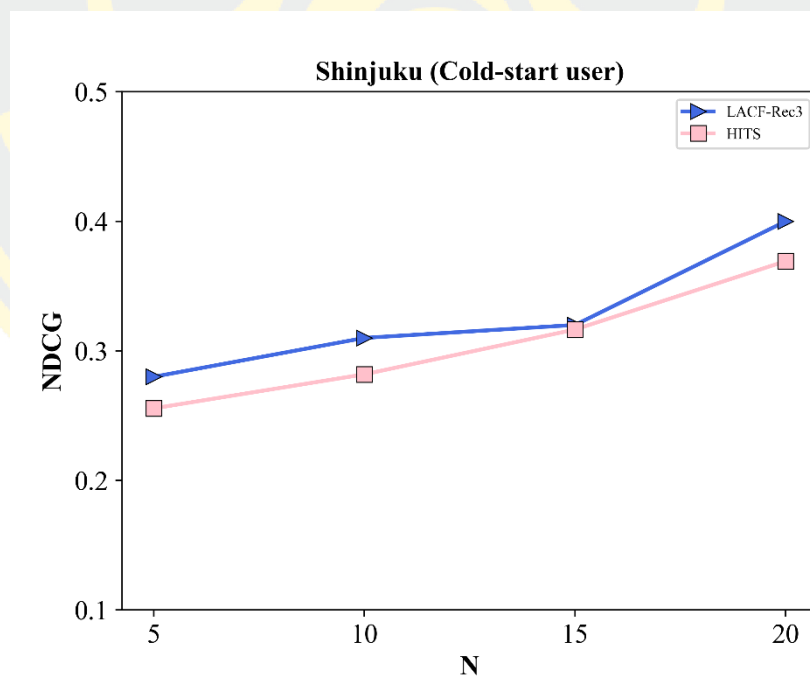
ภาพที่ 23 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki



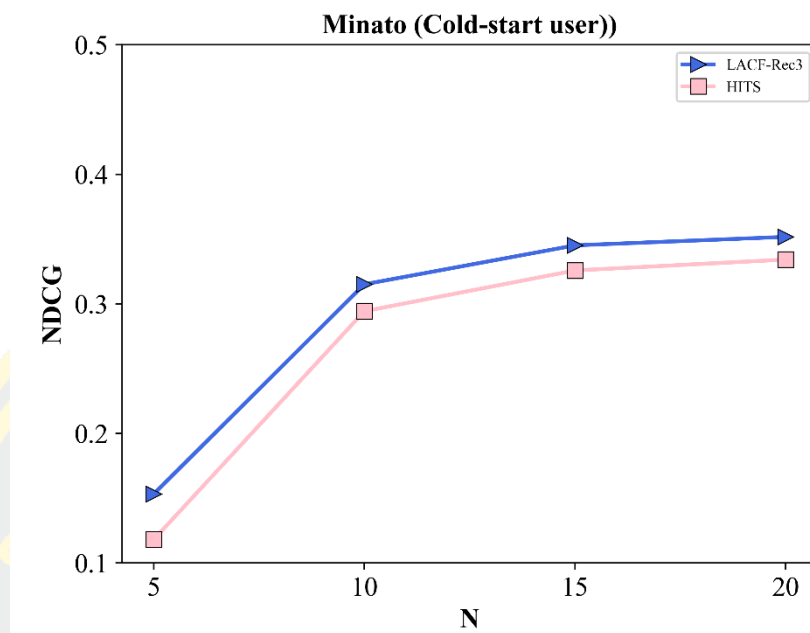
ภาพที่ 24 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo



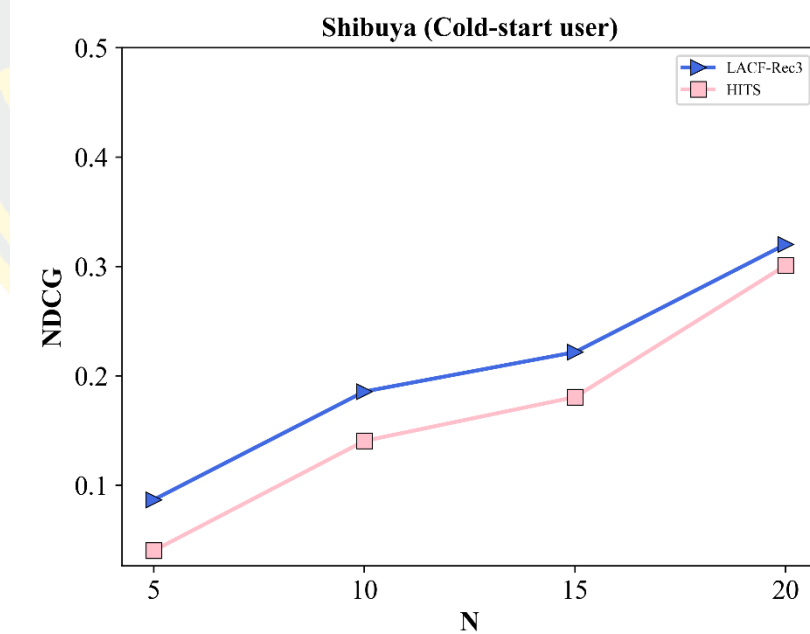
ภาพที่ 25 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda



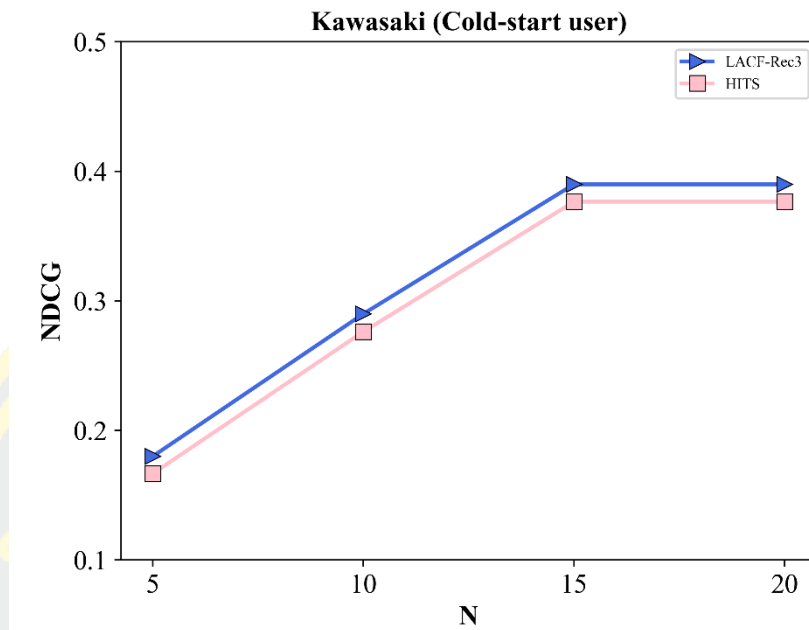
ภาพที่ 26 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku



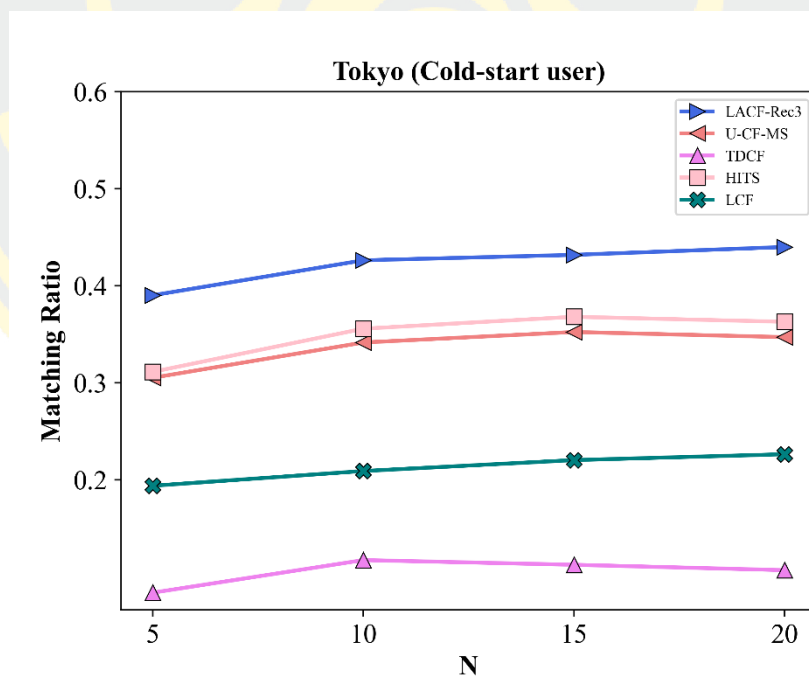
ภาพที่ 27 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato



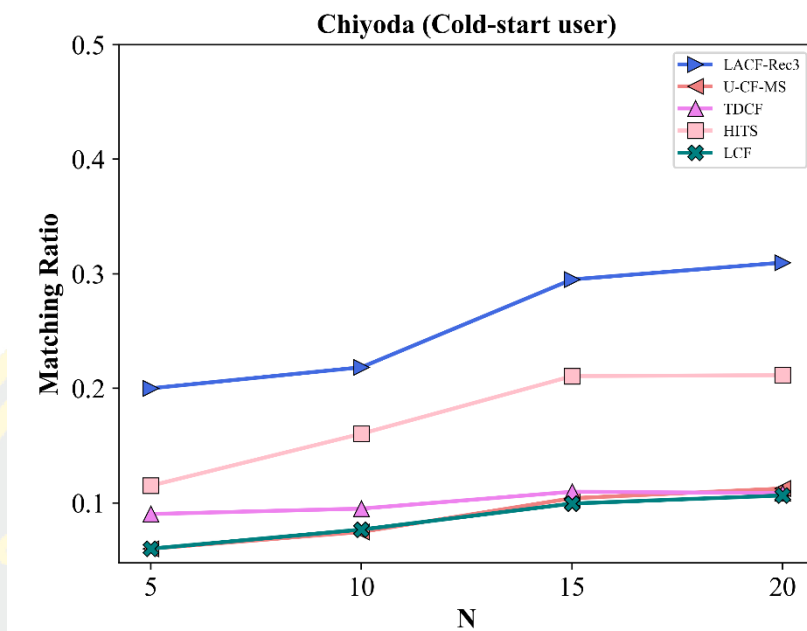
ภาพที่ 28 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya



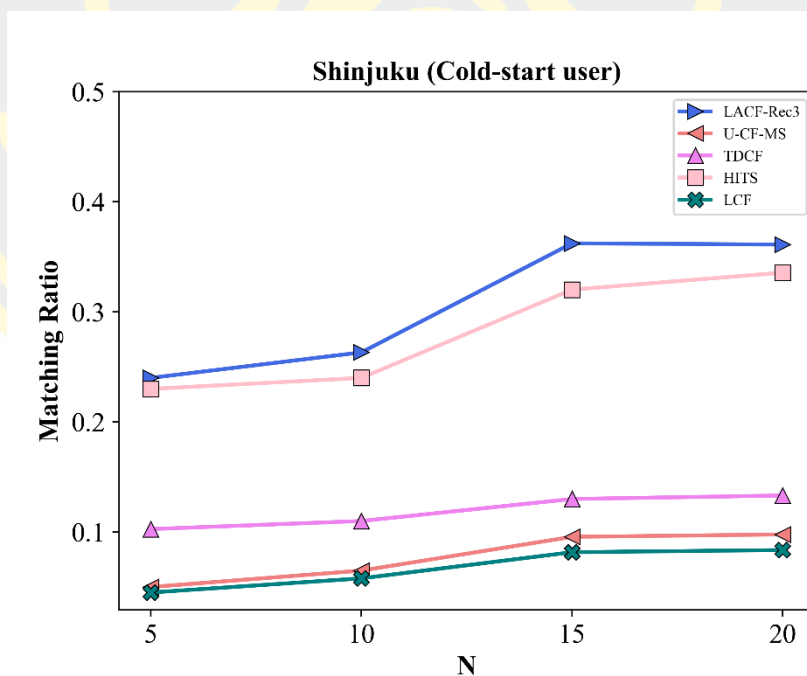
ภาพที่ 29 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki



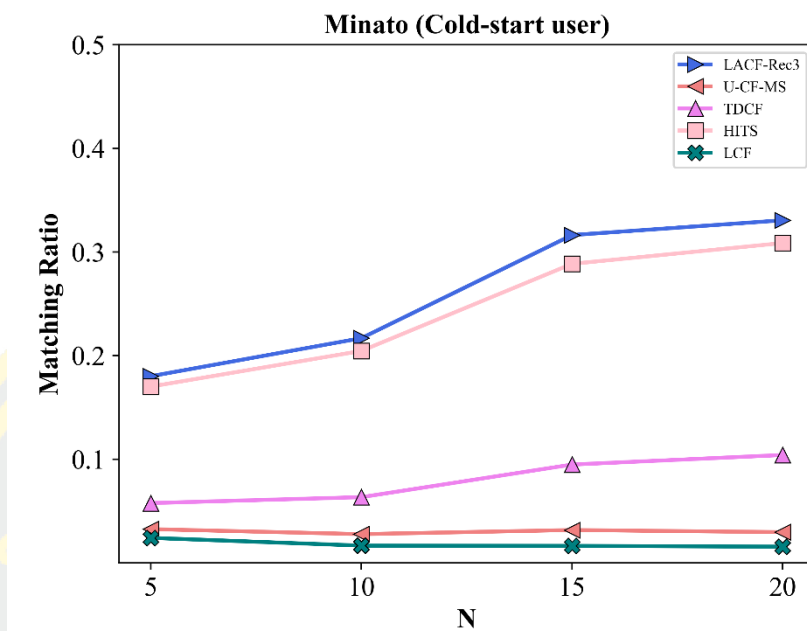
ภาพที่ 30 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo



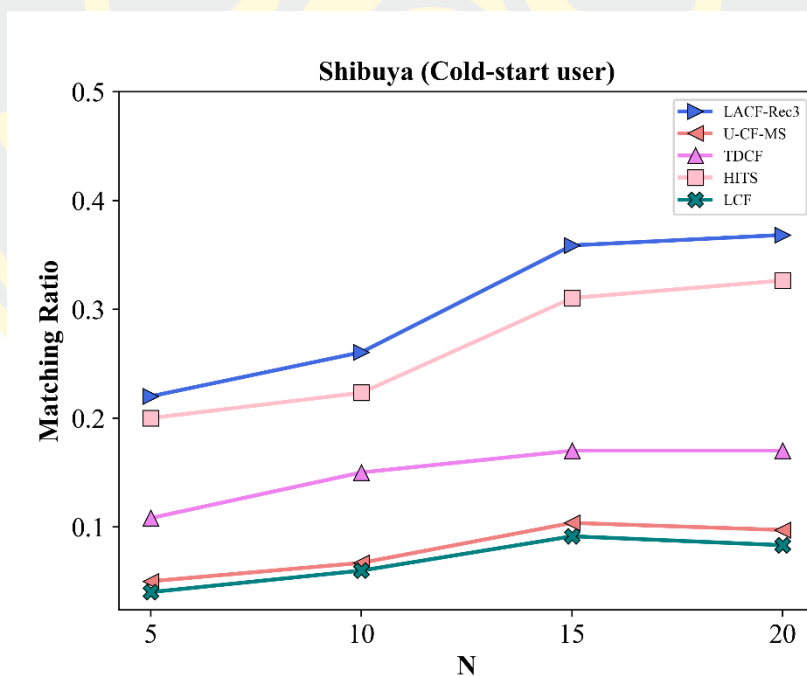
ภาพที่ 31 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda



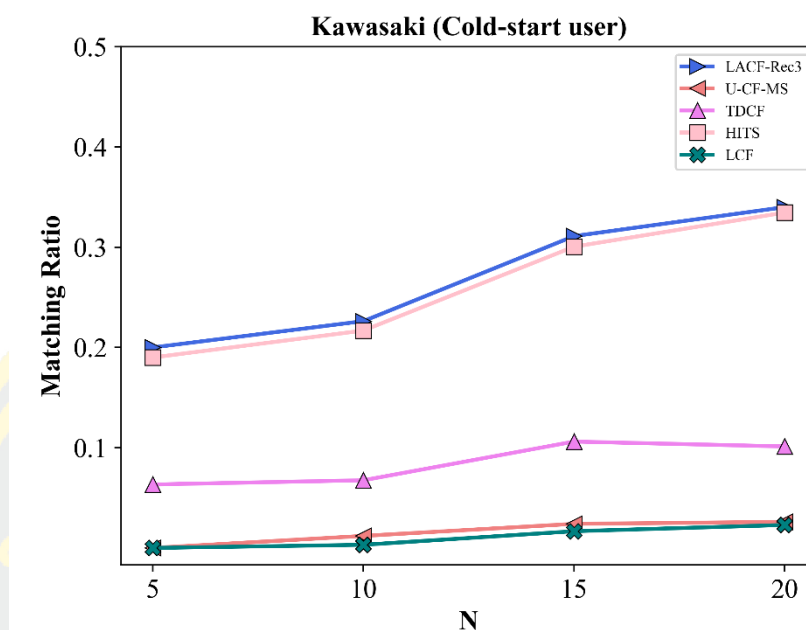
ภาพที่ 32 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku



ภาพที่ 33 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato



ภาพที่ 34 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya



ภาพที่ 35 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki

4.2.2 การแนะนำสถานที่สำหรับผู้ใช้ที่มีประวัติการเช็คอิน

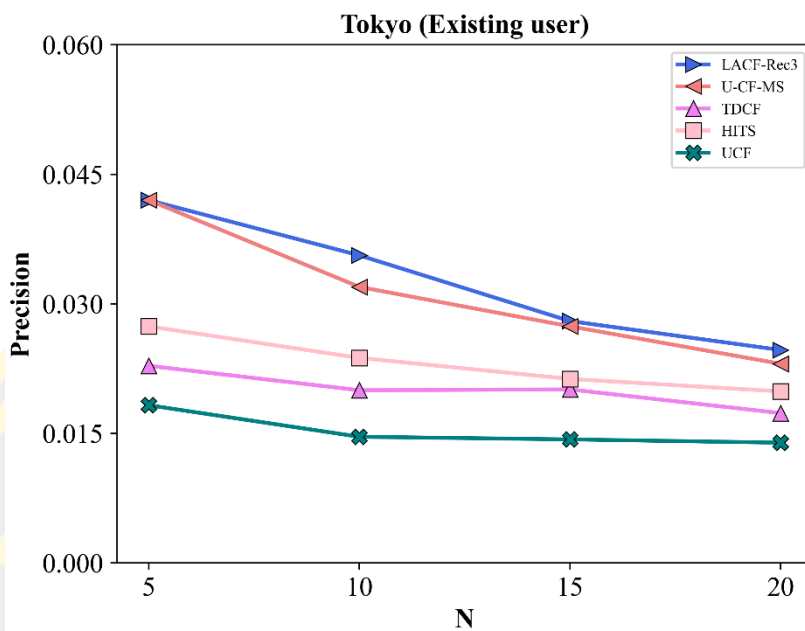
ในการประเมินผลสถานที่ที่แนะนำสำหรับผู้ใช้ที่มีประวัติการเช็คอิน แสดงผลการทดลองความถูกต้องในรายการแนะนำสถานที่ ดังภาพที่ 7-10 ขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 มีความถูกต้องในการแนะนำสถานที่ดีกว่าขั้นตอนวิธี U-CF-MS ขั้นตอนวิธี TDCF ขั้นตอนวิธี HITS และขั้นตอนวิธี UCF ด้วยการให้ขั้นตอนการประเมินผล Precision ที่ 60.35% 96.92% 116.59% และ 185.49% ตามลำดับ สำหรับขั้นตอนการประเมินผล Recall ขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 มีความถูกต้องในการแนะนำสถานที่ดีกว่าขั้นตอนวิธี U-CF-MS ขั้นตอนวิธี TDCF ขั้นตอนวิธี HITS และขั้นตอนวิธี UCF ที่ 61.44% 98.39% 93.23% และ 193.23% ตามลำดับ

ด้วยเหตุที่ว่าผู้ใช้เป้าหมายที่มีประวัติการเช็คอินส่งผลให้สามารถค้นหาสถานที่ที่น่าสนใจโดยการวิเคราะห์ความโดดเด่นของสถานที่ของผู้ใช้เป้าหมายในอดีต ขั้นตอนวิธีการที่นำเสนอได้ค้นหาความคล้ายระหว่างผู้ใช้เป้าหมายและผู้อื่น ในระบบด้วยการพิจารณาความโดดเด่นของสถานที่แทนที่การพิจารณาจากสถานที่โดยตรง เนื่องจากการพิจารณาสถานที่โดยตรงในชุดข้อมูลที่มีความเบาบาง ซึ่งในความเป็นจริงผู้ใช้อีกมีจำนวนการเช็คอินที่น้อย โดยมักมีการเช็คอินในสถานที่ที่มีความชอบหรือในสถานที่ที่อยู่ใ้ในระแวกใกล้เคียงกับที่อยู่ปัจจุบันของผู้ใช้เท่านั้น จึงค้นหาผู้ใช้ที่

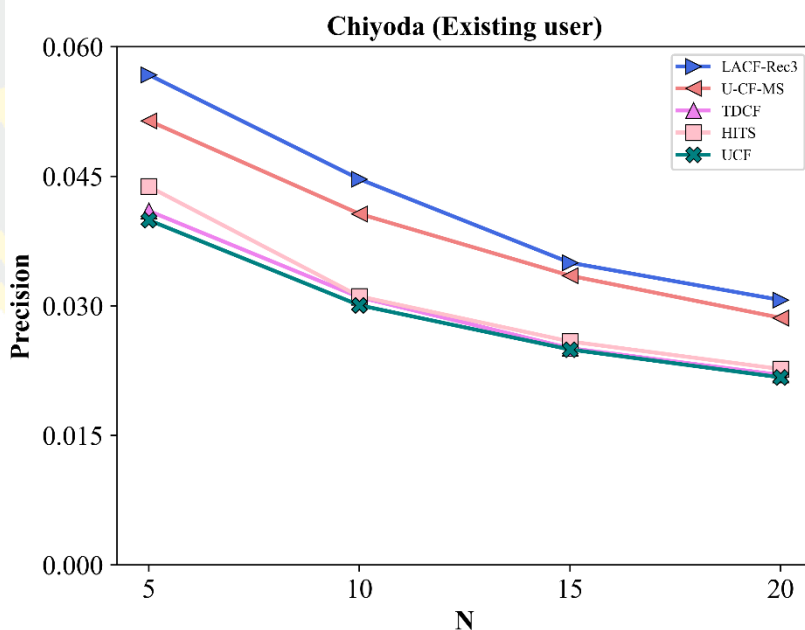
เกี่ยวข้องหรือผู้ใช้ที่มีพฤติกรรมการใช้คอนคัลกับผู้ใช้เป้าหมายได้จำนวนน้อยจากปัญหาความเบาบางของข้อมูล แต่การพิจารณาความคล้ายของความโดดเด่นของสถานที่ส่งผลให้ได้จำนวนผู้ใช้ที่เกี่ยวข้องมากกว่าและส่งผลไปยังสามารถแนะนำสถานที่ได้ตรงตามความชอบของผู้ใช้มากขึ้นด้วย

จากผลการทดลองความถูกต้องของรายการแนะนำสถานที่ ซึ่งมีความสอดคล้องกับความถูกต้องของอันดับการแนะนำสถานที่ โดยขั้นตอนวิธีการที่นำเสนอสามารถทำนายสถานที่ที่แนะนำได้อันดับที่สูงกว่าขั้นตอนวิธี U-CF-MS ขั้นตอนวิธี TDCF ขั้นตอนวิธี HITS และขั้นตอนวิธี UCF โดยใช้วิธีการประเมินผล NDCG ที่ 46.49% 26.17% 9.36% และ 137.66% ตามลำดับ

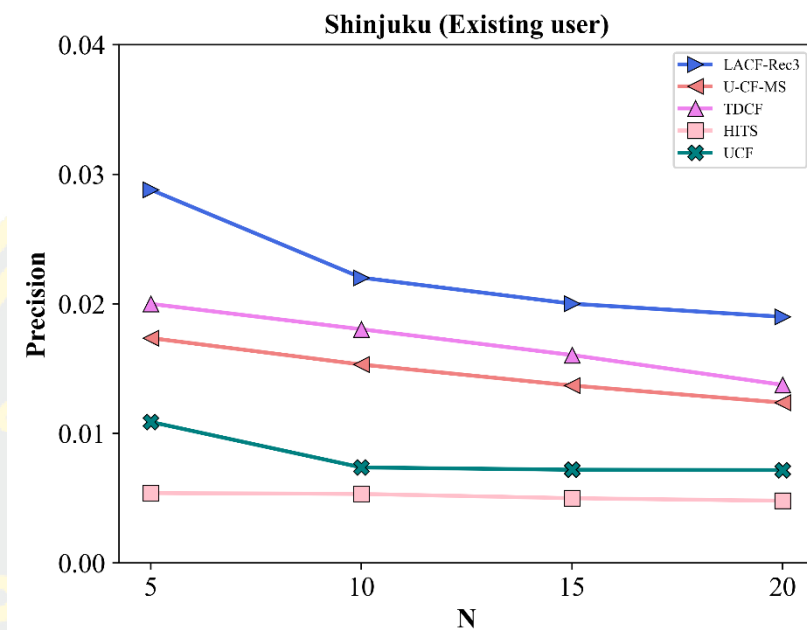
สำหรับการประเมินผลคุณภาพของรายการแนะนำสถานที่ด้วยวิธีประเมินผล Matching ration แสดงให้เห็นว่าขั้นตอนวิธีที่นำเสนอสามารถสืบค้นสถานที่ที่น่าสนใจซึ่งตรงตามความชอบของผู้ใช้ได้ถูกต้องมากกว่าขั้นตอนวิธี U-CF-MS ขั้นตอนวิธี TDCF ขั้นตอนวิธี HITS และขั้นตอนวิธี UCF ที่ 45.82% 159.83% 25.59% และ 205.21% ตามลำดับ เนื่องจากขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 โดยใช้ขั้นตอนวิธี CF-3 ได้พิจารณาสถานที่ที่แนะนำโดยพิจารณาความโดดเด่นในหลายแง่มุมของสถานที่ที่คาดว่าผู้ใช้ชื่นชอบ ซึ่งสามารถสืบค้นสถานที่ได้ดีกว่าขั้นตอนวิธีการก่อนหน้านี้ที่พิจารณาเพียงความถี่ในการเช็คอินเท่านั้น



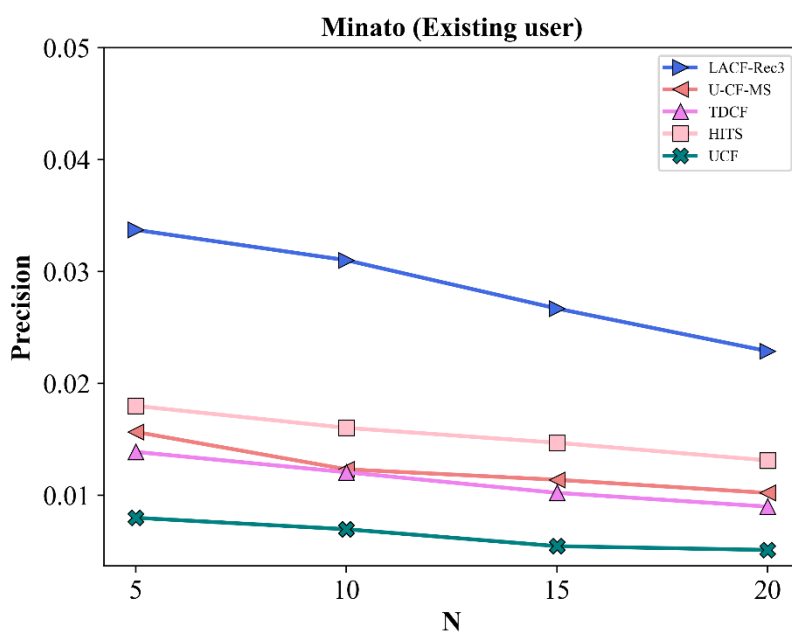
ภาพที่ 36 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo



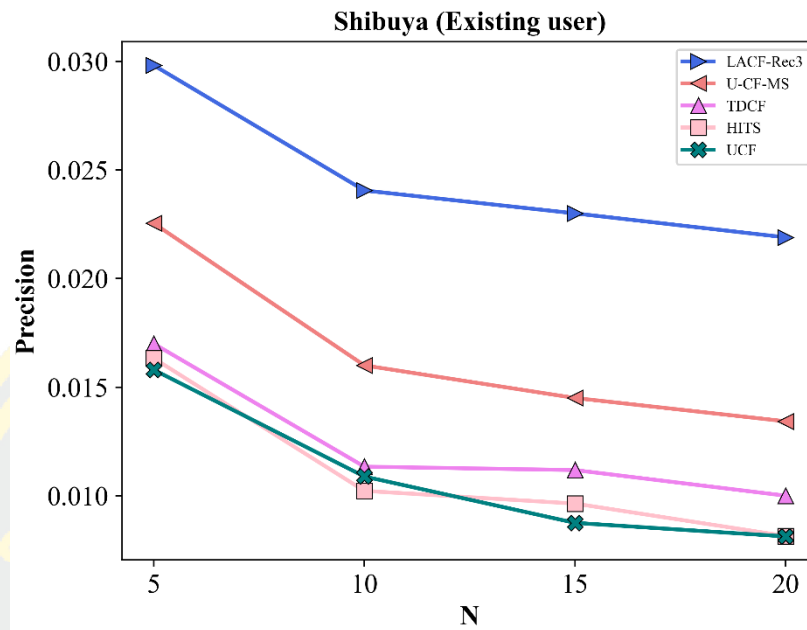
ภาพที่ 37 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda



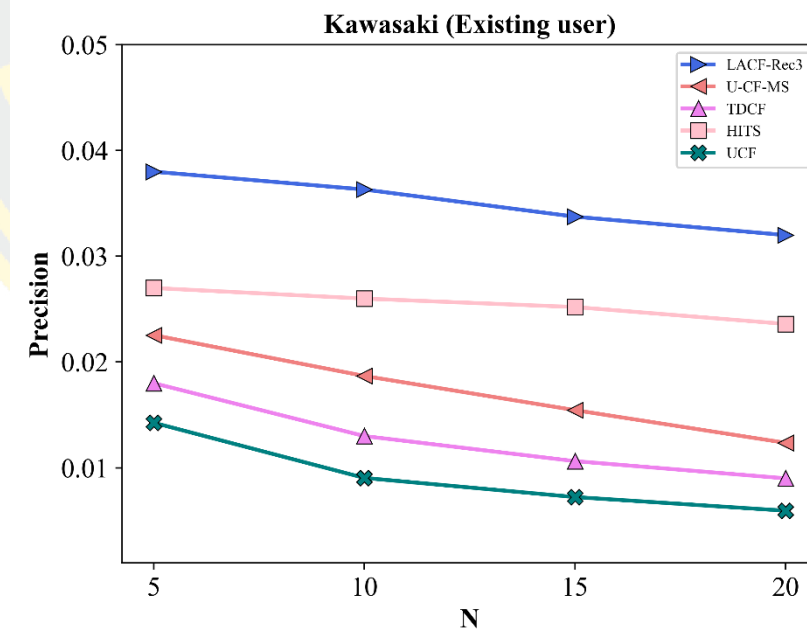
ภาพที่ 38 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku



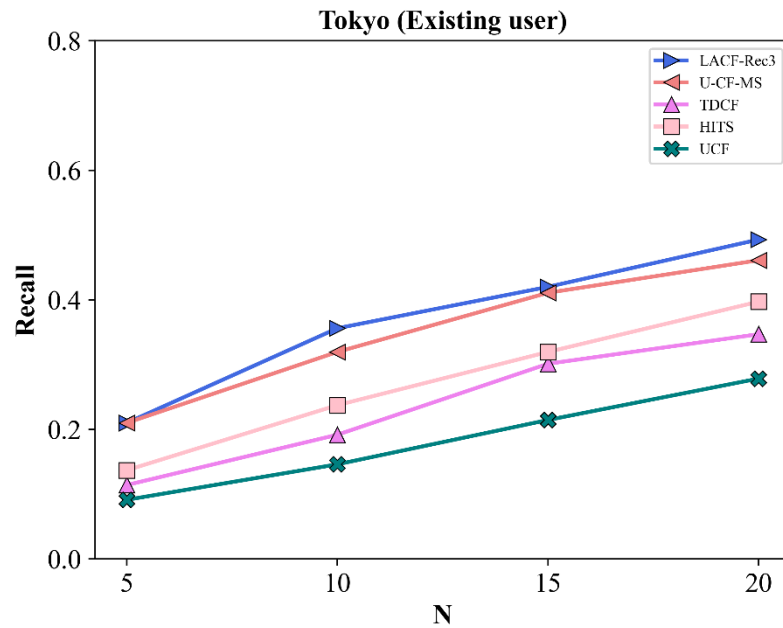
ภาพที่ 39 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato



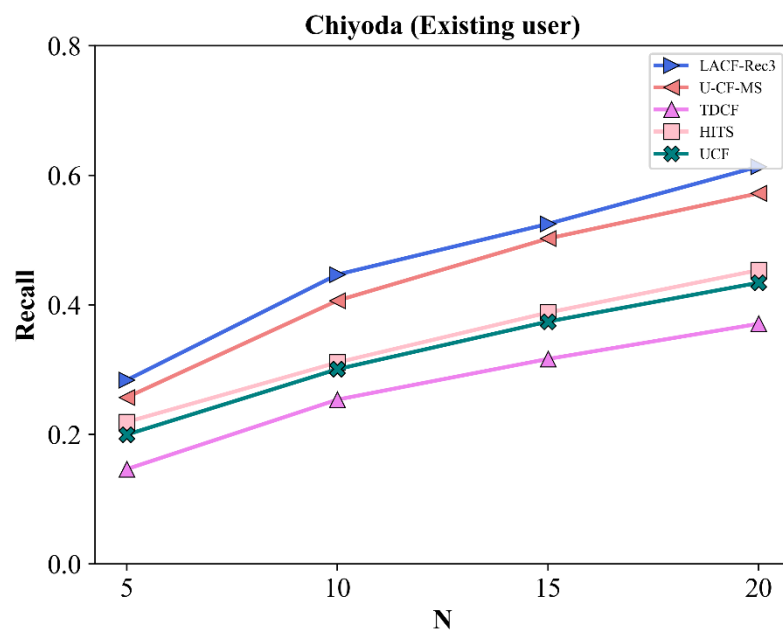
ภาพที่ 40 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya



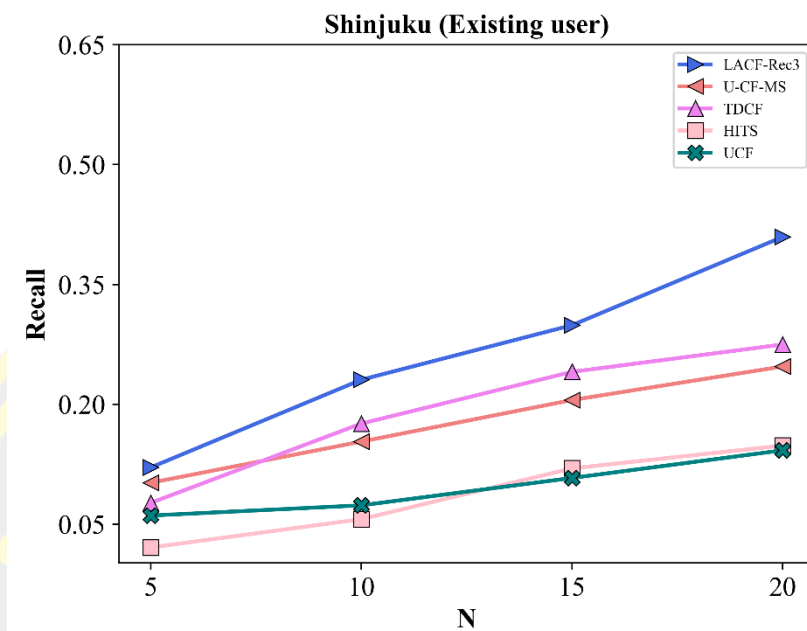
ภาพที่ 41 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Precision โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki



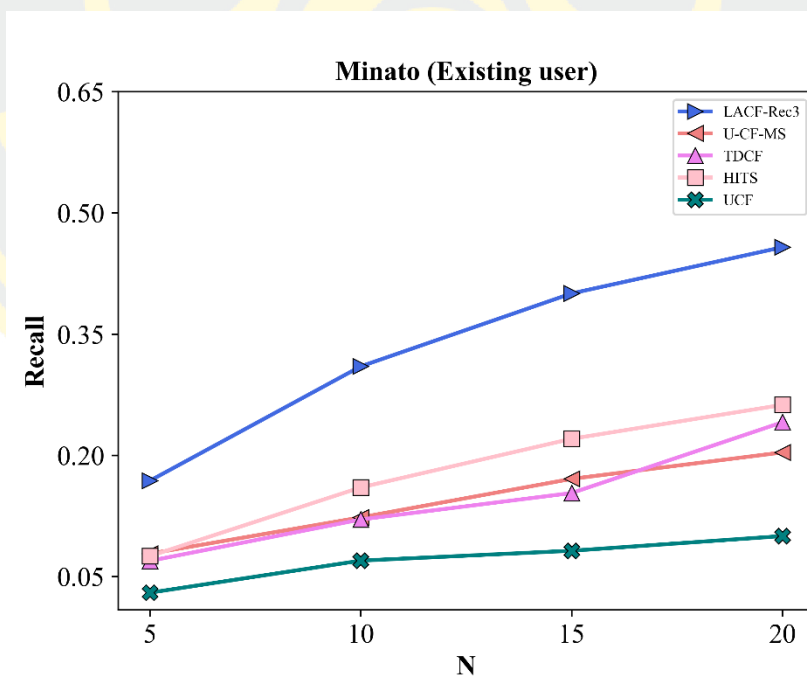
ภาพที่ 42 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo



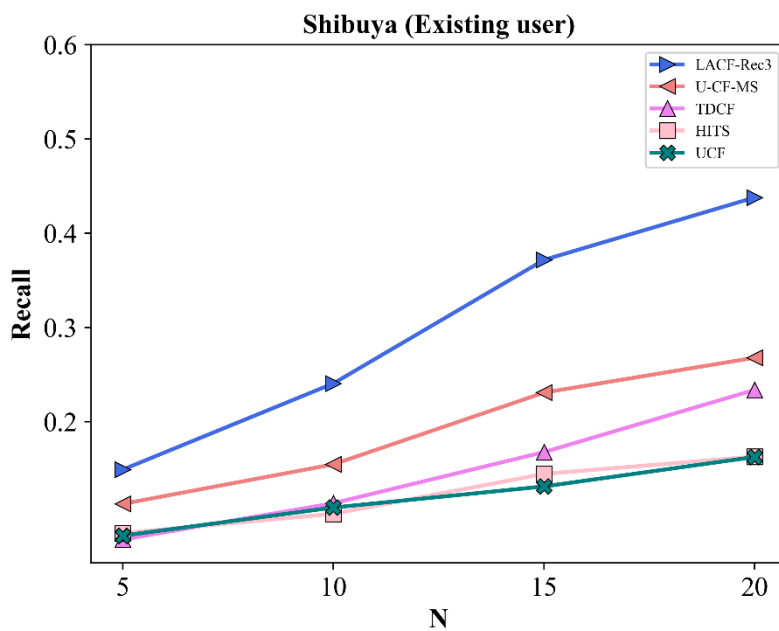
ภาพที่ 43 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda



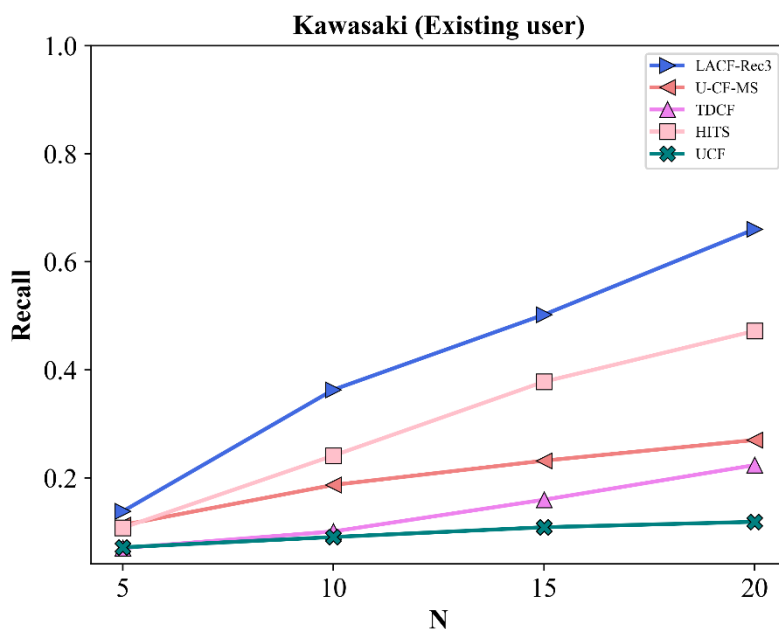
ภาพที่ 44 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku



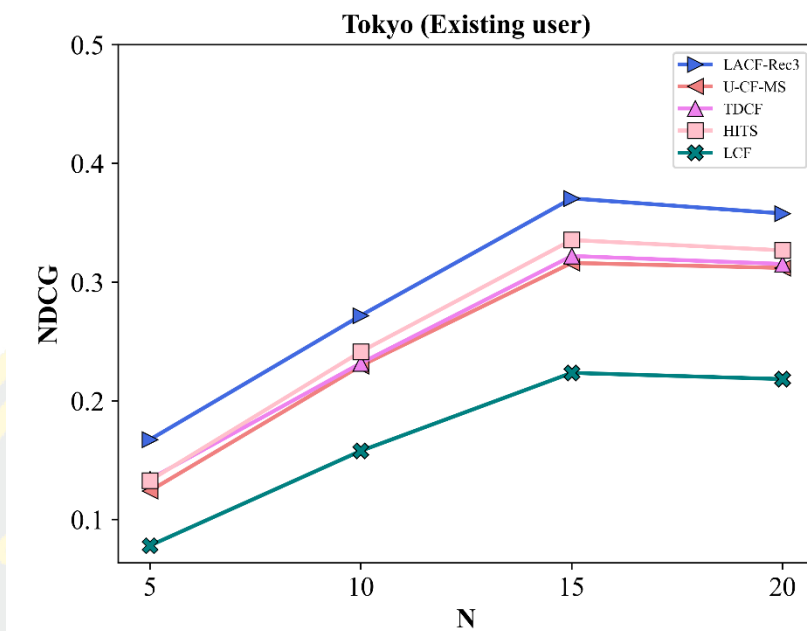
ภาพที่ 45 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato



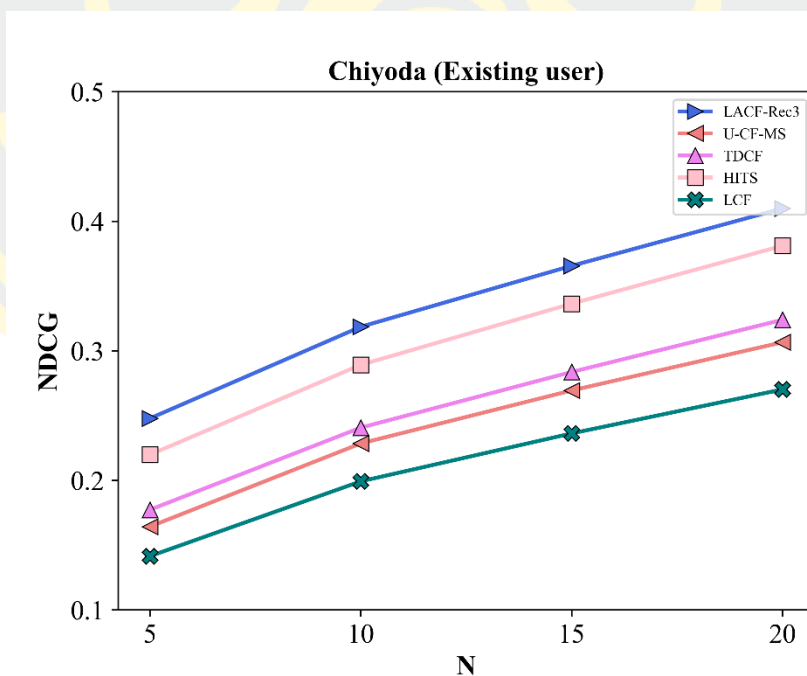
ภาพที่ 46 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya



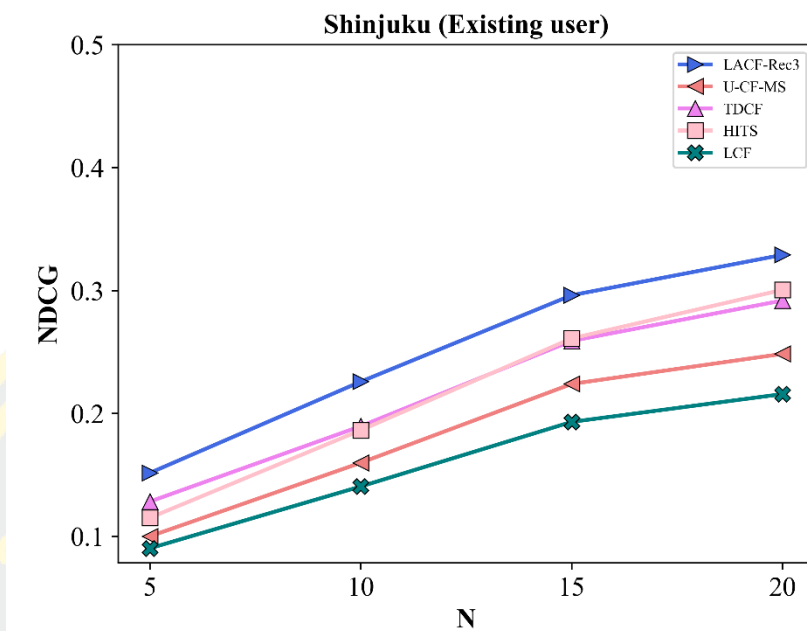
ภาพที่ 47 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Recall โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki



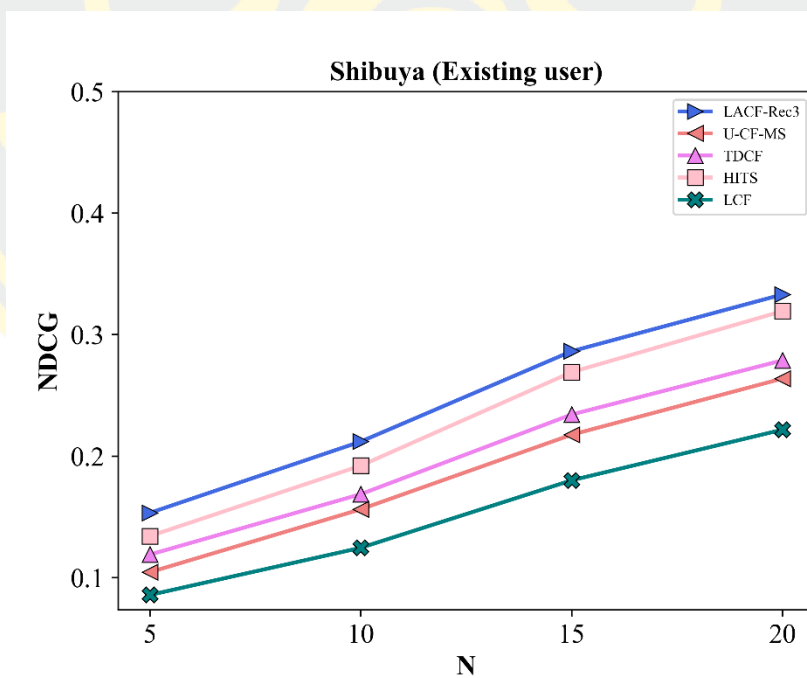
ภาพที่ 48 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo



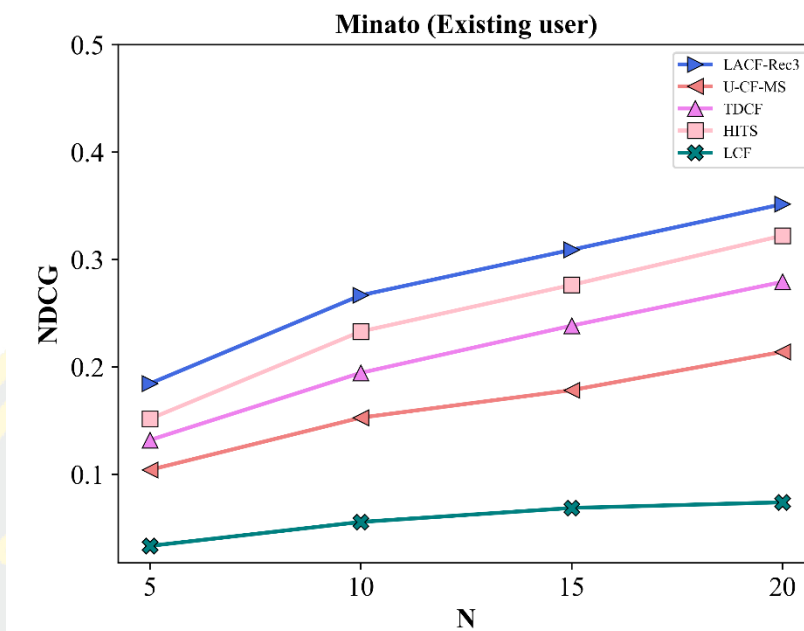
ภาพที่ 49 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda



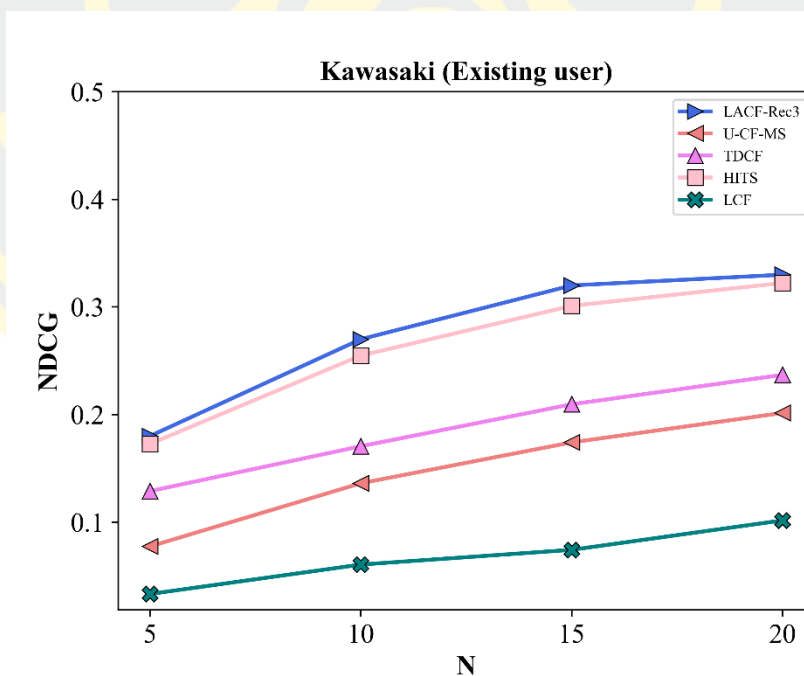
ภาพที่ 50 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku



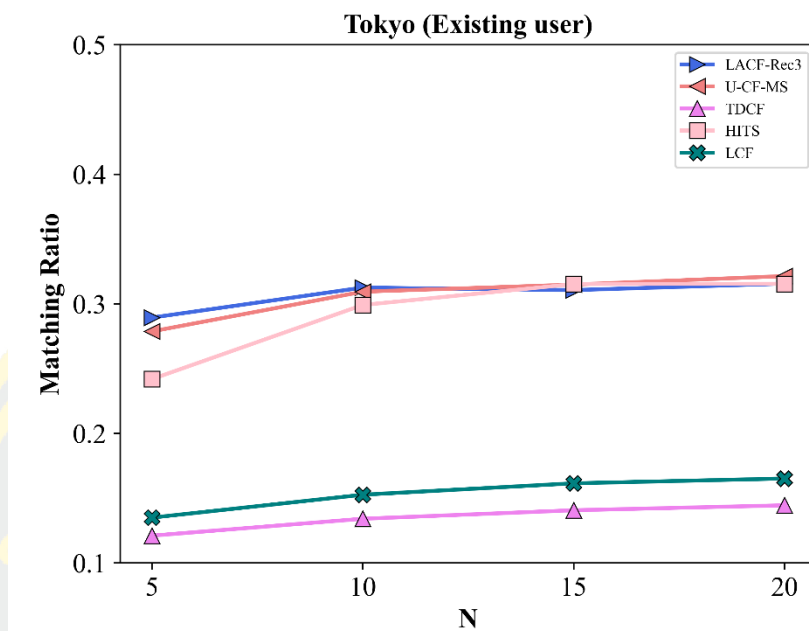
ภาพที่ 51 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya



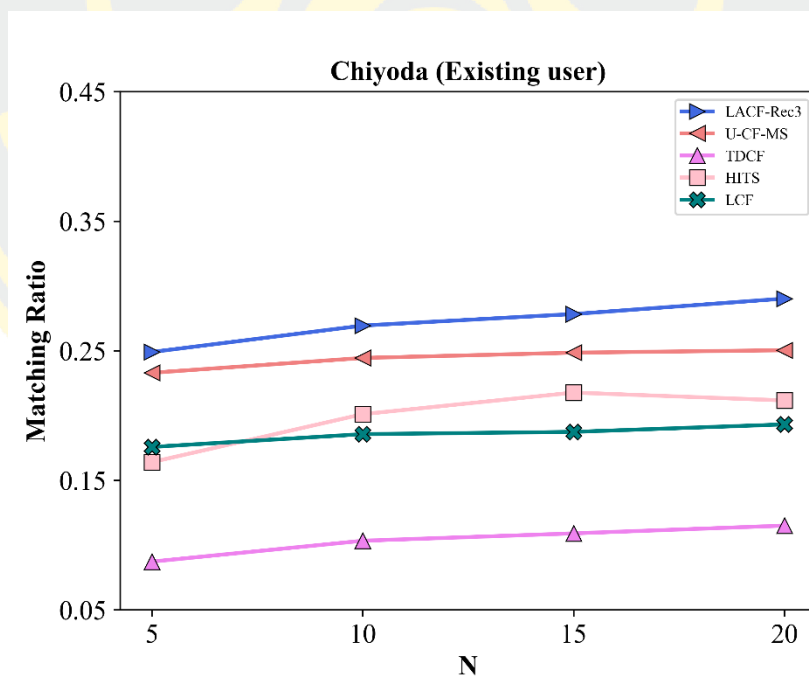
ภาพที่ 52 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato



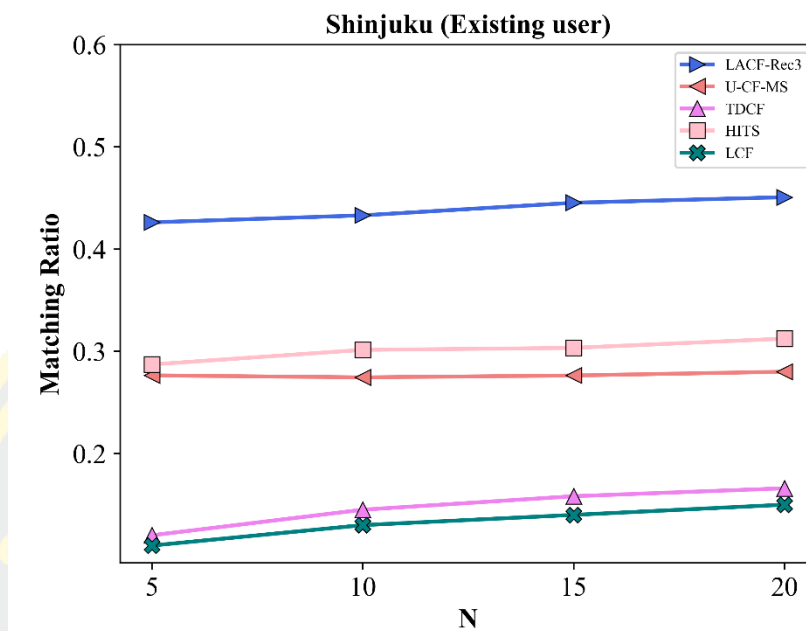
ภาพที่ 53 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ NDCG โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki



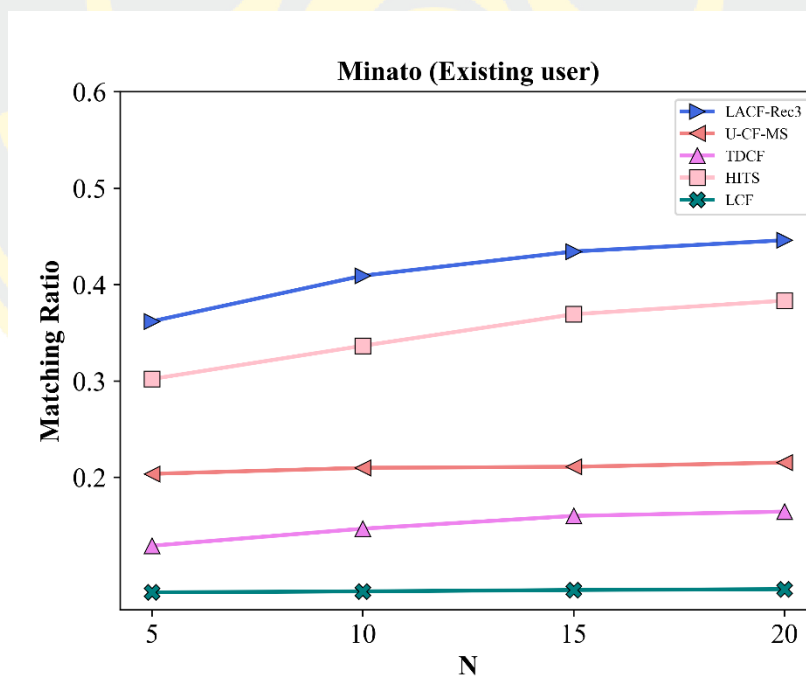
ภาพที่ 54 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Tokyo



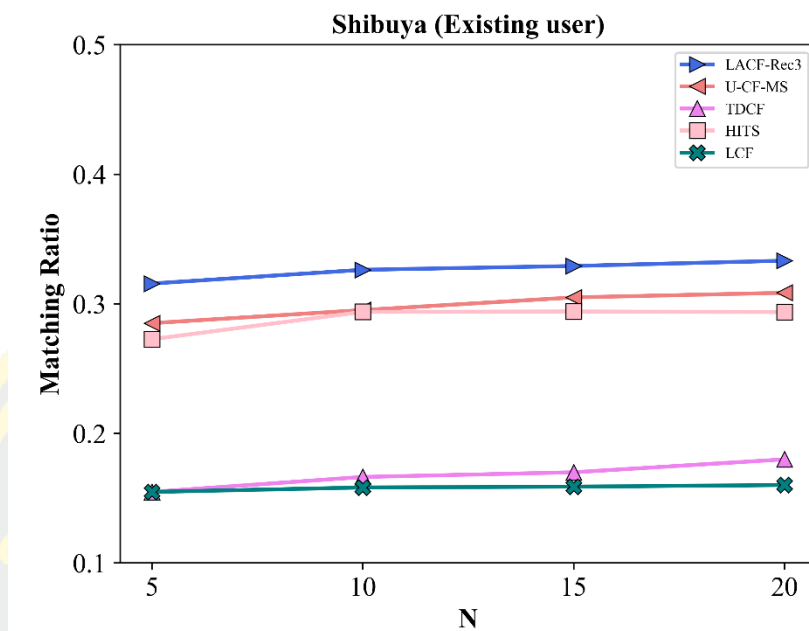
ภาพที่ 55 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Chiyoda



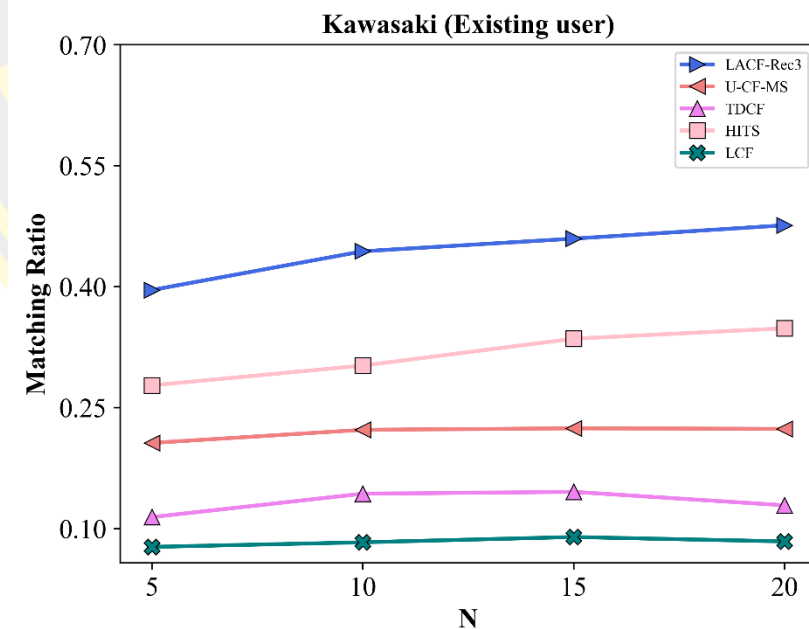
ภาพที่ 56 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shinjuku



ภาพที่ 57 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Minato



ภาพที่ 58 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Shibuya



ภาพที่ 59 ผลความถูกต้องโดยใช้วิธีการวัดประสิทธิภาพ Matching ratio โดยพิจารณารายการแนะนำที่ 5 ถึง 20 รายการสถานที่ของเมือง Kawasaki

บทที่ 5 สรุปผลการทดลอง

ในงานวิจัยนี้ได้นำเสนอระบบแนะนำสถานที่แบบผสมผสานที่เรียกว่า LACF-Rec3 (Link Analysis and Collaborative Filtering Recommendation system based on 3-behaviors) ระหว่างขั้นตอนวิเคราะห์ความเชื่อมโยงและขั้นตอนวิธีการกรองร่วม โดยการพิจารณาพฤติกรรม การเช็คอินของผู้ใช้และสถานที่ในสามแง่มุม ได้แก่ ความถี่ในการเช็คอิน ความหลากหลายในการเช็คอิน และการเช็คอินซ้ำ เพื่อสร้างรายการแนะนำสำหรับผู้ใหม่และผู้ที่มีประวัติการเช็คอิน การสร้างรายการแนะนำสถานที่สำหรับผู้ใหม่ ทางผู้วิจัยได้พัฒนาขั้นตอนวิธีการที่เรียกว่า HITS-3 (HITS based on three check-in behaviors) ซึ่งเป็นการปรับปรุงขั้นตอนวิเคราะห์ความเชื่อมโยงของขั้นตอนวิธีการ HITS เพื่อแก้ไขปัญหาผู้ใช้ใหม่และแก้ไขปัญหาความเบาบางของข้อมูล สำหรับผู้ใช้ที่ไม่มีประวัติการเช็คอินหรือมีการเช็คอินน้อยครั้ง ที่ซึ่งไม่จำเป็นต้องใช้ประวัติการเช็คอินในการค้นหาความชอบของผู้ใช้ โดยการพิจารณาความโดดเด่นของสถานที่ในสามแง่มุมดังกล่าวที่อยู่ในบริเวณใกล้เคียงกับที่อยู่ปัจจุบันของผู้ใช้ สำหรับการสร้างรายการแนะนำสถานที่ให้กับผู้ใช้ที่มีประวัติ การปรับปรุงขั้นตอนวิธีการกรองร่วมที่เรียกว่า CF-3 (Collaborative Filtering based on 3-distinctiveness) โดยการใช้ประโยชน์จากความโดดเด่นของทั้งสามแง่มุมดังกล่าวของสถานที่ได้จากขั้นตอนวิธีการ HITS-3 เพื่อค้นหาความคล้ายของพฤติกรรมเช็คอินบนความโดดเด่นของสถานที่ระหว่างผู้ใช้เป้าหมายและผู้ใช้คนอื่น ๆ ในระบบ จากนั้นการทำนายค่าคะแนนความชอบของสถานที่ จะถูกคำนวณจากความถี่ในการเช็คอิน ค่าความคล้ายของผู้ใช้ และความโดดเด่นของสถานที่ที่ผู้ใช้ชื่นชอบเพื่อสร้างรายการแนะนำสถานที่ให้กับผู้ใช้เป้าหมาย

สำหรับการประเมินผลประสิทธิภาพการทดลองของขั้นตอนวิธีที่นำเสนอ ทางผู้วิจัยได้มีการทดลองบนชุดข้อมูลประวัติการเช็คอินของผู้ใช้จริงจากแอปพลิเคชัน Foursquare โดยทดลองด้วยขั้นตอนวัดประสิทธิภาพ 4 วิธีการ ได้แก่ ขั้นตอนวิธีการ Precision และ Recall สำหรับประเมินผลความถูกต้องของรายการแนะนำสถานที่ ขั้นตอนวิธีการ NDCG ถูกใช้ในการประเมินผลความถูกต้องของอันดับรายการแนะนำสถานที่ และขั้นตอนสุดท้ายเป็นขั้นตอนวิธีการที่ทางผู้วิจัยได้นำเสนอคือขั้นตอนวิธี Matching ratio สำหรับประเมินผลความถูกต้องของรายการแนะนำสถานที่ โดยพิจารณาความชอบของผู้ใช้จากความโดดเด่นของสถานที่ที่ผู้ใช้เช็คอินกับความโดดเด่นของสถานที่บนรายการแนะนำ

สำหรับผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธีการที่นำเสนอ LACF-Rec3 มีประสิทธิภาพ ความถูกต้องของรายการแนะนำสถานที่และความถูกต้องของอันดับรายการแนะนำสถานที่ที่มากกว่า ขั้นตอนเปรียบเทียบก่อนหน้า เนื่องจากขั้นตอนที่นำเสนอ LACF-Rec3 สามารถค้นหาความชอบของผู้ใช้ที่มีความแตกต่างกันได้ดีกว่าขั้นตอนวิธีการก่อนหน้าโดยการพิจารณาความโดดเด่นของสถานที่ สำหรับการประเมินผลผู้ใช้ใหม่ ขั้นตอนวิธีการ LACF-Rec3 โดยใช้ขั้นตอนวิธี HITS-3 ในการค้นหาสถานที่ที่น่าสนใจสามารถสร้างรายการแนะนำสถานที่ที่มีความโดดเด่นในหลายแง่มุมจากพื้นที่ใกล้เคียงกับสถานที่ปัจจุบันของผู้ใช้จึงสามารถค้นหาความชอบของผู้ใช้ได้ดีกว่าขั้นตอนวิธีการก่อนหน้าที่มีการพิจารณาเพียงความถี่ในการเช็คอินเพียงอย่างเดียว สำหรับการประเมินผลผู้ใช้ที่มีประวัติการเช็คอิน โดยใช้ขั้นตอนวิธี CF-3 การค้นหาความคล้ายระหว่างผู้ใช้ด้วยความโดดเด่นของสถานที่แทนที่การพิจารณาสถานที่อย่างเดียวสามารถค้นหาผู้ใช้ที่เกี่ยวข้องได้มากกว่าและส่งผลให้ความถูกต้องของการสร้างรายการแนะนำสถานที่ที่ดีกว่าขั้นตอนวิธีการก่อนหน้า จากผลการทดลองแสดงให้เห็นว่าขั้นตอนวิธีการที่นำเสนอ มีประสิทธิภาพในการแก้ไขปัญหาความเบาบางของข้อมูล ปัญหาการเริ่มต้นได้ยากของผู้ใช้ใหม่และปัญหาความน่าเบื่อของรายการแนะนำสถานที่ได้

บรรณานุกรม

- Bagci, H., & Karagoz, P. (2016). Context-aware location recommendation by using a random walk-based approach. *Knowledge and Information Systems*, 47, 241-260.
- Bao, J., Zheng, Y., & Mokbel, M. F. (2012). Location-based and preference-aware recommendation using sparse geo-social networking data. Proceedings of the 20th international conference on advances in geographic information systems,
- Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. (2015). Recommendations in location-based social networks: a survey. *GeoInformatica*, 19, 525-565.
- Baral, R., & Li, T. (2018). Exploiting the roles of aspects in personalized POI recommender systems. *Data Mining and Knowledge Discovery*, 32, 320-343.
- Darapisut, S., Amphawan, K., Rimcharoen, S., & Leelathakul, N. (2022). N-Most Interesting Location-based Recommender System. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, 16(1), 84-99.
- Ding, Z., Li, X., Jiang, C., & Zhou, M. (2018). Objectives and state-of-the-art of location-based social network recommender systems. *Acm Computing Surveys (Csur)*, 51(1), 1-28.
- Gan, M., & Gao, L. (2019). Discovering memory-based preferences for POI recommendation in location-based social networks. *ISPRS International Journal of Geo-Information*, 8(6), 279.
- Koren, Y., Rendle, S., & Bell, R. (2021). Advances in collaborative filtering. *Recommender systems handbook*, 91-142.
- Levandoski, J. J., Sarwat, M., Eldawy, A., & Mokbel, M. F. (2012). Lars: A location-aware recommender system. 2012 IEEE 28th international conference on data engineering,
- Lim, K. H., Chan, J., Karunasekera, S., & Leckie, C. (2019). Tour recommendation and trip planning using location-based social media: A survey. *Knowledge and Information Systems*, 60, 1247-1275.
- Lim, K. H., Chan, J., Leckie, C., & Karunasekera, S. (2018). Personalized trip recommendation for tourists based on user interests, points of interest visit

- durations and visit recency. *Knowledge and Information Systems*, 54, 375-406.
- Long, X., & Joshi, J. (2013). A HITS-based POI recommendation algorithm for location-based social networks. Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining,
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-based collaborative filtering recommendation algorithms. Proceedings of the 10th international conference on World Wide Web,
- Si, Y., Zhang, F., & Liu, W. (2017). CTF-ARA: An adaptive method for POI recommendation based on check-in and temporal features. *Knowledge-Based Systems*, 128, 59-70.
- Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in artificial intelligence*, 2009.
- Sun, J., Matsushima, Y., & Ma, Q. (2021). Property Analysis of Stay Points for POI Recommendation. Database and Expert Systems Applications: 32nd International Conference, DEXA 2021, Virtual Event, September 27–30, 2021, Proceedings, Part I 32,
- Symeonidis, P., Papadimitriou, A., Manolopoulos, Y., Senkul, P., & Toroslu, I. (2011). Geo-social recommendations based on incremental tensor reduction and local path traversal. Proceedings of the 3rd acm sigspatial international workshop on location-based social networks,
- Wang, H., Li, G., & Feng, J. (2014). Group-based personalized location recommendation on social networks. Web Technologies and Applications: 16th Asia-Pacific Web Conference, APWeb 2014, Changsha, China, September 5-7, 2014. Proceedings 16,
- Wang, Y., Wang, L., Li, Y., He, D., & Liu, T.-Y. (2013). A theoretical analysis of NDCG type ranking measures. Conference on learning theory,
- Xiao, X., Zheng, Y., Luo, Q., & Xie, X. (2010). Finding similar users using category-based location history. Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems,
- Xu, S., Fu, X., Cao, J., Liu, B., & Wang, Z. (2020). Survey on user location prediction based on geo-social networking data. *World Wide Web*, 23(3), 1621-1664.
- Yang, D., Zhang, D., Zheng, V. W., & Yu, Z. (2014). Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Transactions on*

Systems, Man, and Cybernetics: Systems, 45(1), 129-142.

- Yin, M., Liu, Y., Zhou, X., & Sun, G. (2021). A tensor decomposition based collaborative filtering algorithm for time-aware POI recommendation in LBSN. *Multimedia Tools and Applications*, 80(30), 36215-36235.
- Ying, Y., Chen, L., & Chen, G. (2017). A temporal-aware POI recommendation system using context-aware tensor decomposition and weighted HITS. *Neurocomputing*, 242, 195-205.
- Yu, Y., & Chen, X. (2015). A survey of point-of-interest recommendation in location-based social networks. Workshops at the twenty-ninth AAAI conference on artificial intelligence,
- Yuan, Q., Cong, G., Ma, Z., Sun, A., & Thalmann, N. M.-. (2013). Time-aware point-of-interest recommendation. Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval,
- Zhang, H., Gan, M., & Sun, X. (2021). Incorporating memory-based preferences and point-of-interest stickiness into recommendations in location-based social networks. *ISPRS International Journal of Geo-Information*, 10(1), 36.
- Zhao, G., Lou, P., Qian, X., & Hou, X. (2020). Personalized location recommendation by fusing sentimental and spatial context. *Knowledge-Based Systems*, 196, 105849.
- Zhao, S., King, I., & Lyu, M. R. (2016). A survey of point-of-interest recommendation in location-based social networks. *arXiv preprint arXiv:1607.00647*.
- Zheng, V., Cao, B., Zheng, Y., Xie, X., & Yang, Q. (2010). Collaborative filtering meets mobile recommendation: A user-centered approach. Proceedings of the AAAI Conference on Artificial Intelligence,
- Zheng, V. W., Zheng, Y., Xie, X., & Yang, Q. (2010). Collaborative location and activity recommendations with gps history data. Proceedings of the 19th international conference on World wide web,
- Zheng, Y., Zhang, L., Xie, X., & Ma, W.-Y. (2009). Mining interesting locations and travel sequences from GPS trajectories. Proceedings of the 18th international conference on World wide web,





ภาคผนวก



ภาคผนวก ก

เอกสารรับรองผลการพิจารณาจริยธรรมการวิจัยในมนุษย์

เลขที่ IRB3-118/2564



เอกสารรับรองผลการพิจารณาจริยธรรมการวิจัยในมนุษย์
มหาวิทยาลัยบูรพา

คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ได้พิจารณาโครงการวิจัย

รหัสโครงการวิจัย : G-HS 090/2564

โครงการวิจัยเรื่อง : ระบบการแนะนำสถานที่ภายใต้การวิเคราะห์พฤติกรรมของผู้ใช้

หัวหน้าโครงการวิจัย : นายสุเมธ ตาราพิสุทธิ

หน่วยงานที่สังกัด : นิสิตระดับบัณฑิตศึกษา คณะวิทยาการสารสนเทศ

คณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา ได้พิจารณาแล้วเห็นว่า โครงการวิจัยดังกล่าวเป็นไปตามหลักการของจริยธรรมการวิจัยในมนุษย์ โดยที่ผู้วิจัยเคารพสิทธิและศักดิ์ศรีในความเป็นมนุษย์ ไม่มีการล่วงละเมิดสิทธิ สวัสดิภาพ และไม่ก่อให้เกิดอันตรายแก่ตัวผู้วิจัยและผู้เข้าร่วมโครงการวิจัย

จึงเห็นสมควรให้ดำเนินการวิจัยในขอบข่ายของโครงการวิจัยที่เสนอได้ (ดูตามเอกสารตรวจสอบ)

- | | |
|--|--|
| 1. แบบเสนอเพื่อขอรับการพิจารณาจริยธรรมการวิจัยในมนุษย์ | ฉบับที่ 1 วันที่ 2 เดือน ธันวาคม พ.ศ. 2564 |
| 2. เอกสารโครงการวิจัยฉบับภาษาไทย | ฉบับที่ 1 วันที่ 2 เดือน ธันวาคม พ.ศ. 2564 |
| 3. เอกสารชี้แจงผู้เข้าร่วมโครงการวิจัย | ฉบับที่ - วันที่ - เดือน - พ.ศ. - |
| 4. เอกสารแสดงความยินยอมของผู้เข้าร่วมโครงการวิจัย | ฉบับที่ - วันที่ - เดือน - พ.ศ. - |
| 5. เอกสารแสดงรายละเอียดเครื่องมือที่ใช้ในการวิจัย | ฉบับที่ - วันที่ - เดือน - พ.ศ. - |
| 6. เอกสารอื่นๆ | ฉบับที่ - วันที่ - เดือน - พ.ศ. - |

วันที่รับรอง : วันที่ 2 เดือน ธันวาคม พ.ศ. 2564

วันที่หมดอายุ : วันที่ 2 เดือน ธันวาคม พ.ศ. 2565

ลงนาม

(ผู้ช่วยศาสตราจารย์ แพทย์หญิงจรูญ แยมประทุม)

ประธานคณะกรรมการพิจารณาจริยธรรมการวิจัยในมนุษย์ มหาวิทยาลัยบูรพา

สำหรับโครงการวิจัย ระดับบัณฑิตศึกษา และระดับปริญญาตรี

ชุดที่ 3 (กลุ่มคลินิก/ วิทยาศาสตร์สุขภาพ/ วิทยาศาสตร์และเทคโนโลยี)



ภาคผนวก ข
เอกสารเผยแพร่ผลงานวิจัย



ISSN 2286 – 9131

Transactions

on Computer and Information Technology
Volume 16, No. 1, March 2022

Research Article

Performance analysis of downlink NOMA relying energy harvesting and full-duplex	<i>C.-B. Le, D.-T. Do</i>	1
An Enhancement of Network Reliability for Patient Monitoring System with IoT Rehabilitation Devices	<i>D. Buranapanichkit, N. Pornichalermpong, K. Thongnoo, K. Thongpull, K. Chetpattananondh, R. Duangsoithong, K. Sengchui and N. Jindapetch</i>	10
Distributed Region-Based Monitoring in Low-Power Listening Wireless Sensor Networks	<i>K. Pattamasirawat and C. Jaikaeo</i>	21
FFF: Fast Firewall Framework to Enhance Rule Verifying over High-speed Networks	<i>S. Khummanee, P. Songram and P. Pruksasri</i>	35
Exploiting Impact of Imperfect CSI on Performance of Backscatter-Aided NOMA System	<i>M.-S. V. Nguyen and D.-T. Do</i>	48
DSSF: Decision Support System to Detect and Solve Firewall Rule Anomalies based on a Probability Approach	<i>S. Khummanee, P. Chomphuwiset and P. Pruksasri</i>	56
ECHONET Lite Framework Based on Embedded Component Systems	<i>F. Qi, J. Y. Jiang, H. Oyama, H. Nagashima and T. Azumi</i>	74
N-Most Interesting Location-based Recommender System	<i>S. Darapisut, K. Amphawan, S. Rimcharoen and N. Leelathakul</i>	84
A Development of Visual-feedback Automatic Control for Robotic Manipulator	<i>H. Nakahara and K. Thongpull</i>	100
Spatial-Frequency Redistribution-based Saliency Region Detection for Thai Text Localisation	<i>T. Titjaroomroj, U. Suttapakti and W. Numsong</i>	109

Submission Guideline

In your manuscript, one author has been designated as the corresponding author with contact details:

- E-mail address
- Affiliation

Further considerations before your submission:

- Manuscript has been 'spell checked' and 'grammar checked'
- All references quoted in the Reference Section are cited in the body of the manuscript
- Permission for use of copyrighted materials from other sources must be declared
- The manuscript length must be at least 8 pages and no longer than 12 pages with two (2) columns.

Form more information visit: <https://ph01.tci-thaijo.org/index.php/ecti/author-guideline>

<https://ph01.tci-thaijo.org/index.php/ecti/>

ECTI



N-Most Interesting Location-based Recommender System

Sumet Darapisut¹ Komate Amphawan² Sunisa Rimcharoen³ and Nutthanon Leelathakul⁴

ABSTRACT

The popular and ubiquitous location-based social networks (LBSNs) appeal many users for sharing interesting locations with other users. As the collected data (such as users' profile, location's comment, and suggestion) become a lot larger in size, location-based recommender systems require more effective filters to be able to suggest potentially preferable locations to users. Location recommendation is more difficult and challenging especially if users have few or no check-in histories as a new user. Therefore, instead of depending on users' check-in histories, previous works focused on creating recommended location lists by leveraging the information given by other users who check in locations in each area. However, previous studies took into account only the frequencies of users for creating recommendation lists and have suffered from the cold-start problem where new users have few or none histories. As a result, the recommenders hardly suggest any locations matching the users' preference. In this paper, we propose an enhanced location-recommendation approach called N-most interesting location-based recommender system (NILR) to recommend interesting locations for new users. Our approach can be divided into three phases. First, the NILR discovers interesting locations by taking into account both the visiting frequencies and the preferences of users already in the system. Second, a ranking procedure is executed to create a final recommendation list based on two interestingness scores: one obtained from the HITS-based model (as adopted by [1] and [2]) and the other from our proposed method. Finally, we re-filter interesting locations based on the current location of the new user. Experimental results reveal the NILR can reach better precision, recall, average ranking and NDGC than HITS by 6%, 6%, 30% and 8% for Tokyo and 24%, 30%, 43% and 15% for New York dataset, respectively.

Article information:

Keywords: Location-based recommender systems, Non-profile users, Ranking

Article history:

Received: April 26, 2021

Revised: July 2, 2021

Accepted: November 4, 2021

Published: March 12, 2022

(Online)

DOI: 10.37936/ecti-cit.2022161.247546

1. INTRODUCTION

Location-based social networks (LBSNs) have been growing rapidly. This can be seen by counting the number of users and locations. This is possible because of the prevalence of Internet and mobile phone technologies, which facilitate the use of LBSNs [3], [4], [5], [6]. Users in social networks share locations together with corresponding opinions and suggestions for the locations they have visited in the real world. There are many popular LBSN applications such as Foursquare, Facebook Place, and Yelp. They allow users to share check-in data and add new locations. The sets of data (including user profiles, location profiles and relationships between users and locations)

collected by the applications become very large, making location recommendation more challenging.

To address the issue just mentioned, location-based recommender systems (LBRs) [7], [8], [9] play an important role in decision making, by filtering and recommending potentially preferable locations to users. Business owners might rely on LBRs to attract more new customers, resulting in more profit. Most LBRs perform efficiently in the case of dense user's check-ins or long visiting histories as it is relatively easy to capture users' preference. Nevertheless, in real-world scenarios, users tend to visit locations in their vicinity. They might have few check-ins and are unable to visit all locations, most of which are far

^{1,2,3,4}The authors are with Faculty of Informatics, Burapha University, Chonburi, Thailand 20130, E-mail: dear-sumet@gmail.com, komate@gmail.com, rsunisa@buu.ac.th and nutthanon@buu.ac.th

2.1 A location-based social network system

Location-based social network systems (LBSNs) [5], [6], [8], [9], have been popular in recent years due to the advances in mobile technologies and the Internet. Users can access LBSNs easily via smartphone applications, which allow users to share their locations. Also, they often share comments or suggestions regarding the interesting locations that they checked in, such as shopping malls, restaurants, coffee shops, bakeries, and convenience stores. According to Foursquare's report in June 2016, there were more than 8 billion check-ins each month at more than 65 million locations with over 55 million users in the world [6]. Yelp disclosed that there were about 29 million active users per month in the system [6]. Thanks to the large number of users and locations contributing to tremendous amount of available data, there are many extensive research works on location-based recommender systems (LBRs), which is one research area in the field of recommender systems. There are 3 main types of objects to be recommended [9]: users, locations, and activities, as shown in Fig. 1. Users (either travelers or local experts) are the ones who could influence their friends/followers or share similar preferences with others in interesting areas. Their common interests will be used as information to generate location recommendation lists and they can also be used as objects for recommendation [17], [18], [19]. There are two ways of recommending locations that users might prefer: a single location, and a sequence of places to visit, called a trip [20], [21], [22]. The activities are things that users do at a certain time in specific places. Interesting activities are recommended by considering popular activities of the users such as shopping in a mall in the afternoon, and watching drama in a theater at night [12], [23], [24].

Related to our work are location recommendations based on 1) users' experience and 2) local experts or other users in the system.

2.2 Location Recommendations based on users' experience

In 2013, Quan [20] presented a location recommendation algorithm that applied a collaborative filtering approach based on users' behavior. Furthermore, this work took into account the time context of a user. This is called a time-aware POI recommendation, and alleviates a data sparsity problem by considering temporal preference at other time slots of each user in conjunction with finding other users who have similar temporal preferences with a target user (who asks for a recommendation). A category-aware POI recommendation model [25] was proposed to capture user preferences by regarding location categories in transition patterns of a user. A matrix factorization was chosen for creating location recommendation lists. In addition, this method grouped users who

have similar check-in time periods and considered the users' current location. Daniel and Trevor [26] studied users' behaviors and locations in the Foursquare dataset such as frequent users' check-in patterns, time periods of returning to already-visited locations, distribution of checked-in locations, and checked-in location categories over time periods. As a result, this work could recommend potential locations where a user may visit later. In 2017, an adaptive POI recommendation approach, incorporating activities of each user in time context with spatial features called CTF-ARA, was presented [27]. The CTF-ARA method classified users into active and inactive users by using a k-means algorithm. Active users have a large number of check-ins. Time context is regarded for recommending locations to an active user. Inactive users are provided recommendations based on the locations without considering the time context. Jiuxin [28] proposed a location recommendation approach by analyzing users' check-in patterns. The patterns are categorized into three features: fine-grained time intervals, global popularity of locations, and users' personal preferences. Then a classification model is used for predicting the next check-in location of users. Next, a Point-of-Interest (POI) recommendation method was proposed [29]. The method considers three various aspects: 1) location attributes such as latitude, longitude, categories, and user's check-in time, 2) user attributes such as suggestions, reviews, and ratings that users provided, and 3) the other information such as distances of from POIs to user's location and users' social friendship. These aspects were regarded with a ranking-based and a matrix factorization-based method for generating POI recommendation lists. Mingxin and Ling [30] introduced a location recommendation method named a memory-based POI preference attenuation model by taking into account two aspects: user's movements, and a memory-based preference attenuation model by using Ebbinghaus's forgetting curve. These aspects were used to determine the similarity between the target user and others. Then, a collaborative filtering method based on the proposed similarity was used to create recommendation lists. Lastly, a user profile awareness recommendation algorithm (ISC-CF) was proposed [31] to recommend interesting locations and address the sparsity problems by taking into account reviews, user check-in histories, friend relationships, and the users' current location. Latent Dirichlet Allocation (LDA) and collaborative filtering were used to extract interesting locations visited by users and to recommend interesting locations, respectively.

The previous works mentioned above demand data about users' experiences to correctly recommend locations. They yield relatively low accuracy in the case where users are new to the system (i.e., their profile/activity histories have never been collected). In contrast, our work requires no such information.

away from the current location. Hence, recommending locations to a new user or a traveler from out of town is much more challenging. To cope with such scenarios, several works attempt to discover interesting locations based on visiting habits of local users.

Most existing works discover interesting locations based only on frequencies of check-in or visits [1], [2], [10], in which some locations may have a large number of visits of many users but have return visit counts of 0 (some users visit them only once and never return). In this case, the locations may be newly popular or they may be known as a must to be visited only once in a life time. On the other hand, some interesting but unpopular locations might have a moderate number of visits but users always return more than once. To the best of our knowledge, there exists no work that takes into account visiting frequency, locations' return visits, and each user's return counts to discover interesting locations. Additionally, LBRs have also suffered from the cold-start problem where new users have no or few check-in history entries causing recommenders to suggest hardly any locations that match the users' preference.

In this work, we present a novel location recommendation approach called N-most interesting location-based recommender system (NILR) to recommend interesting locations for new users (extended from [11]). Our NILR approach can be divided into three phases. First, the NILR can discover interesting locations based on previous users already in the system by taking into account frequency of visiting and preferences of users. The preferences of users can be considered by noting who visits a location more than once. Interesting locations and users have a synergistic relationship. Interesting locations are visited repeatedly and frequently by various users. Very knowledgeable users always visit and re-visit numerous interesting locations, possessing more insight and knowledge. Second, to rank locations in the recommendation list, we rank outstanding locations based on interestingness scores, frequencies, and preferences. In other words, users obtain a recommended location list, composed of locations selected effectively by considering diverse aspects instead of only one aspect as in traditional methods. Lastly, when a new user requests a nearby interesting location recommendation, our method filters out locations too far away from the user's current location and, based on the interestingness scores, sorts all locations in descending order. We evaluated our proposed method using a real-world dataset provided by Foursquare. The data set is composed of the check-in's information associated with five areas in Tokyo city: Chiyoda, Minato, Shinjuku, Shibuya and Chuo, and one area in New York city. The data is arranged by the number of locations in decreasing order [12]. For evaluating accuracies of lists recommended for new users, we simulate geospatial ranges, each of

which covers all the visited locations in each user's current area, by forming minimum bounding rectangles (MBRs) [1], [13]. The accuracies are measured in terms of precision, recall, ranking measures (*i.e.*, average ranking [14] and normalized discounted cumulative gain, NDCG [15]) and are compared with those of the HITS-based model. Experimental results reveal NILR performs effectively and efficiently, and it outperforms HITS in terms of accuracies and rankings.

The main contributions of our work can be summarized as follows:

- We propose a novel approach which can address the issue of cold start users, named N-most interesting location-based recommender system (NILR), to recommend interesting locations.
- To discover interesting locations, both check-in frequencies and preferences are considered as opposed to other traditional methods, which rely only on frequencies.
- We also propose a ranking algorithm to select outstanding locations for creating recommendation lists based on interestingness scores by taking into account both visitation frequencies to and preferences of users for locations.
- We evaluated our method on a real-world dataset, provided by Foursquare, that is an extensive dataset in LBSNs and we use four metrics including precision, recall, average ranking, and normalized discounted cumulative gain (NDCG).

2. RELATED WORK

This section provides background information and the motivation behind the location-based social network system and reviews existing location-based recommender systems. They can be divided into two categories: location recommendations based on user's experience, and recommendations based on choices made by local experts or other users in the system.

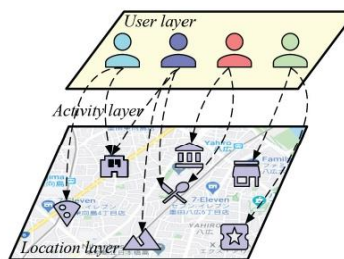


Fig. 1: An example of location-based social networks.

NILR is based on local experts or other users who are already in the system. Therefore, it is more suitable for recommending locations to users with few check-ins.

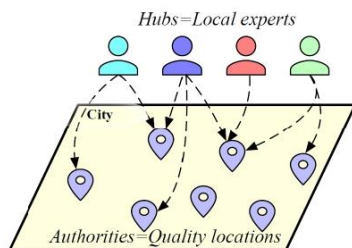


Fig.2: An example of the HITS-based model.

2.3 Location Recommendation based on local experts or other users in the system

Most of the location-based recommendation systems which consider local experts or other users in the system rely on the HITS (Hypertext Induced Topic Search) based model [32] that was originally used for ranking web pages in terms of authority scores and hub scores. An authority is a popular page linked by many hub pages which many authority pages liked. Several works applied this concept by mapping a meaning of the term authorities to locations and hubs to users. Thus, high authority scores mean interesting locations. Highly experienced users (with high hub scores) represent local experts, as shown in Fig. 2. Yu [10] proposed the Tree-Based Hierarchical Graph (TBHG). This work originally proposed the method of modeling travel-sequences of various users on a variety of geospatial sizes and discovering users who have high travelling experience (so called local experts) to recommend interesting locations by using the traditional HITS-based model. Later, Jia [1] presented location-based and preference-aware recommendation systems. The HITS-based model was applied for discovering local experts and then a Weighted Category Hierarchy (WCH) is created for all users. For generating recommendation lists, collaborative filtering (CF) is used for predicting interesting locations based on the similarity between the target user and local experts in the WCH. Xuelian and James [33] proposed a HITS-based POI recommendation algorithm regarding social relationships that focused on relationships among friends more than those among ordinary users. In addition, users, with high experiences from the HITS-based model are discovered by taking into account varieties of visited locations. Likewise, locations for recommending are considered by varieties of local experts measuring in terms of entropy. In 2016,

a context-aware location recommendation algorithm (CLoRW) [2] was proposed by using a graph model to represent popular locations, friendship relationships, and local experts. The popular locations based on other users in the system and local experts are figured out with the HITS-based model. Lastly, a random walk algorithm is used for generating location recommendation lists. A personalized successive POI recommendation approach was presented [34]. This work considered three factors to create recommendation list including successive behavior, locality behavior, and group preferences. A distance-weighted HITS algorithm is exploited to discover popular locations based on other users in the system. Later, a location recommendation system based on a Context-aware Tensor Decomposition (CTD) and a weighted HITS were presented [13]. This work suggested locations for an individual effectively. To capture users' preferences, the CTD method took into account temporal influences, including time context and location categories. The weighted HITS algorithm based on user friendships between a target user and other users was applied to generate recommendation lists. In 2019, a dynamic recommender system for suggesting shopping places on Foursquare was proposed [35]. This work searched for local experts in two aspects including experts in their location (called local authorities) and experts in query categories (called topic authorities). They also considered the social contacts of a target user. If the user has no social contacts, a global network graph is created. Otherwise, they created a privacy-aware network graph from the user's social contacts. Then, the network properties of the user are used as guidance to select appropriate algorithms (i.e., PageRank and Lazy random walk algorithms) to find local experts.

The aforementioned research works considered location recommendation lists based on local experts or other users in the system to improve the accuracy of recommendation lists. Nevertheless, several previous works in this field require users' profiles to determine matching local experts or other users in the system. Consequently, they might be unsuitable for new users or travelers without collected profiles. In [33], Long and Joshi consider users based on varieties of locations using entropy measures but this measure has high sensitiveness to check-in data. For example, if a user checks in every location once and another checks in almost every location more than once. The first user has higher entropy values than that of the latter, which may not be correct. In [2], [13], [35], they leverage social relationships to recommend locations. However, friends do not always share common preferences. Collaborative filtering based on users of similar preferences appears to yield more accurate recommendation lists. In [34], the work discovers interesting locations based on distances, which could take more computational time in the online phase

to calculate distances between users and locations. In [35], positive reviews are used to filter interesting locations. However, fake reviews might exist, and might cause biased recommendation lists. A lot of reviews on each shopping mall are needed for analyzing sentiment precisely. In addition, not all systems collect or provide reviews. Several previous works in this field require users' profiles for producing accurate recommendation lists, as shown in Table 1. They might be unsuitable for recommending locations to new users or travelers without collected profiles.

In our method, we focus on a location recommendation method based on other users in the system. Location recommendations are different from book or movie recommendations in that users always check in physical locations in their vicinity not far from their office or home, resulting in much sparser data of user experiences or profiles. Therefore, we propose an enhanced location recommendation method based on other users in the system which is more suitable for recommending locations to new users. This is achieved by taking into account visiting frequency, location return visits, and each user's return counts, instead of merely using frequencies. Revisiting represents a strong preference of each user for each location because he or she preferred to use the location's services more than once. Such locations can be implied as interesting locations. However, visiting frequencies of various users are also important to capture popular and interesting locations. We rank outstanding locations based on interesting scores by considering frequency of visiting and users' preferences (i.e., return visits).

3. OUR METHOD

In this section, we first describe basic definitions regarding the N-most interesting location-based recommendation problem. Then, the details of our proposed system, N-most interesting location-based recommender system (NILR), are introduced. The NILR considers a set of locations of food shops including restaurants, cafés, coffee, and desert shops. It measures the interestingness of each location by regarding both visiting frequencies and user preferences for visiting places. Lastly, the N-most interesting locations are recommended to new users unfamiliar with their current location.

3.1 Problem Statement

Let C be a target city which has m food shops (including restaurants, cafés, coffee, and desert shops) of which locations are represented as $L = \{l_1, l_2, \dots, l_m\}$. The location l_j of the j^{th} shop is associated with its latitude and longitude $\langle lat_l, lon_l \rangle$. The set of shops can be represented as $L = \{\langle l_1, lat_{l_1}, lon_{l_1} \rangle, \langle l_2, lat_{l_2}, lon_{l_2} \rangle, \dots, \langle l_q, lat_{l_q}, lon_{l_q} \rangle\}$. Subsequently, let $U = \{u_1, u_2, \dots, u_p\}$ be a set of p users

who have visited at least one shop in L . Whenever a user $u_k \in U$ goes to visit a shop $l_j \in L$, his visit is recorded as $v^{u_k, l_j} = \langle u_k, l_j, f \rangle$, where f denotes a frequency (i.e., the number of times the user u_k has visited the shop l_j). A database $DB = \{v_1, v_2, \dots, v_z\}$ contains an ordered set of z visiting records of all users in U who visited the shops in L .

The problem of location-based recommendation is to discover a set of N most interesting locations (of shops) for a new user nu_x (where $nu_x \notin U$) who are unfamiliar with their current location, and who ask for a list of recommended places, defined as $RL^{nu_x} = \{rl_1, rl_2, \dots, rl_N\}$, each of which is an interesting shop $l_j \in L^{nu_x}$ (with a high interestingness score) which the user nu_x should visit.

3.2 The proposed NILR system

Our NILR system takes visiting histories (e.g. check-in, sharing location, etc.) of users (already in the system) as input. Then, it recommends the N-most interesting locations by examining visiting histories of all users. As shown in Fig. 3, the NILR system consists of three main steps: 1) calculation of interestingness of all locations in L , 2) location ranking based on their interestingness, and 3) generation of the top-N most interesting locations for a specific user.

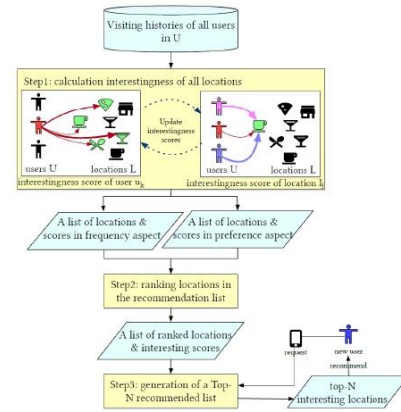


Fig.3: Our method.

3.2.1 Calculation the interestingness of all locations

To compute the interestingness of each location ($l_j \in L$), the visiting history of all users related to l_j is considered iteratively. The interestingness of the location l_j can be expressed by two scores: 1) the one based on its visiting frequency, and 2) the other

Table 1: Features for considering interesting locations based on local experts or other users in the system.

Research works	Features for considering interesting locations						
	Freq.	Variety	Revisit	Review	Distance	Friends	User's profile
[1]	✓	–	–	–	–	–	✓
[2]	✓	–	–	–	–	✓	✓
[10]	✓	–	–	–	–	–	✓
[13]	✓	✓	–	–	–	✓	✓
[33]	✓	✓	–	–	–	✓	✓
[34]	✓	–	–	–	✓	–	✓
[35]	–	–	–	✓	–	✓	–
NILR	✓	–	✓	–	–	–	–

based on visiting preference. These two values can be iteratively computed as described in the following definitions.

Definition 1: An interestingness score of the location l_j based on frequency of visits is calculated by letting $V^{l_j} = \{v_1, v_2, \dots, v_p\}$ be the ordered set of visiting records of all users at the location l_j . Each visiting record $v_x \in V^{l_j}$ such that $v_x = \langle u_x, l_j, f_x \rangle$ denotes the number of times, f_x , that the user u_x has visited the location l_j . The interestingness score of l_j based on the frequency of visits is determined by summing each user's visiting score (which is determined by multiplying frequency of visits at l_j by the user's interestingness score) as shown in Equation 1.

$$isf(l_j) = \sum_{x=1}^p (f_x \times isf(u_x)) \quad (1)$$

$isf(u_x)$ is u_x 's interestingness score (firstly initialized as 1), which is further updated as shown in Eq. 2.

Definition 2: An interestingness score of the user u_k is based on her visiting behavior. It can be computed by letting $V^{u_k} = \{v_1, v_2, \dots, v_q\}$ be the ordered set of visiting records of the user u_k at any locations in L . Each visiting record ($v_y \in V^{u_k}$) such that $v_y = \langle u_k, l_y, f_y \rangle$ indicates the number of times, f_y , that the user u_k has visited the location l_y . Then, the interestingness score of the user u_k based on his or her visiting behavior is equal to the summation of each location's visiting score which represents how many times he or she has ever visited (determined by multiplying her frequency of visits at l_y and the interestingness score of l_y), as shown in Equation 2.

$$isf(u_k) = \sum_{y=1}^q (f_y \times isf(l_y)) \quad (2)$$

$isf(l_y)$ is l_y 's interestingness score as computed by Eq. 1.

However, the values of $isf(l_j)$ and of $isf(u_k)$ grow after each calculation. It is necessary to perform an L2-normalization on the $isf(l_j)$ and $isf(u_k)$ (so that they eventually converge) using Equations 3 and 4.

$$isf(l_j) = \frac{isf(l_j)}{\sqrt{\sum_{y=1}^{|L|} isf(l_y)^2}} \quad (3)$$

$$isf(u_k) = \frac{isf(u_k)}{\sqrt{\sum_{x=1}^{|U|} isf(u_x)^2}} \quad (4)$$

On the other hand, to consider the preference of a user u_x for a location l_j , the number of revisits is considered and calculated. If the user u_x visits the shop l_j more than once, it can be said that the user u_x likes or prefers the shop. The level of a user's preference for a location is calculated as follows:

Definition 3: A preference of a user u_x for a location l_j is identified by either 0 or 1, expressing whether the user u_x repeatedly visits the location l_j , as shown in Equation 5.

$$pref(u_x, l_j) = \begin{cases} 1, & f_x > 1 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

f_x in $\langle u_x, l_j, f_x \rangle$ is the number of times that the user u_x has visited the location l_j .

Definition 4: A preference score of a user u_x is the number of locations that the user u_x has ever revisited, defined in Equation 6.

$$pref(u_x) = \sum_{j=1}^{|L|} pref(u_x, l_j) \quad (6)$$

Definition 5: An interestingness score of the location l_j based on preference of users can be calculated by letting $V^{l_j} = \{v_1, v_2, \dots, v_p\}$ be the ordered set of visiting records of all users at the location l_j . Each visiting record ($v_x \in V^{l_j}$) such that $v_x = \langle u_x, l_j, f_x \rangle$ denotes the number of times, f_x , that the user u_x visits the location l_j . The interestingness score of l_j based on preference of users is the summation of each user's visiting score (which is the product of the user's number of visits at l_j , the user's preference score, and the user's interestingness score based on preference), and is defined in Equation 7.

$$isp(l_j) = \sum_{x=1}^p (f_x \times pref(u_x) \times isp(u_x)) \quad (7)$$

$isp(u_x)$ is a user score based on his or her preference, firstly initialized as 1 and updated later as explained in Def. 7.

Definition 6: A preference score of a location l_j based on visiting of users is the number of users in U that prefer the location l_j (i.e. the users that visit l_j more than once) defined in Equation 8.

$$pref(l_j) = \sum_{x=1}^{|U|} pref(u_x, l_j) \quad (8)$$

Definition 7: An interestingness score of the user u_k based on preference can be calculated by letting $V^{u_k} = \{v_1, v_2, \dots, v_q\}$ be the ordered set of visiting records of the user u_k at any locations in L . Each visiting record ($v_y \in V^{u_k}$ such that $v_y = \langle u_k, l_y, f_y \rangle$) identifies the number of times, f_y , that the user u_k visits the location l_y . The interestingness score of the user u_k based on his or her location preference is determined by summing the preference scores of each location he or she has ever visited (which is the product of the user's frequency of visits at the location, the preference score of the location, and the interestingness score of l_y based on users' preference) as shown in Equation 9.

$$isp(u_k) = \sum_{y=1}^q (f_y \times pref(l_y) \times isp(l_y)) \quad (9)$$

$isp(l_y)$ is an interestingness score of the location l_y based on the preference of users, determined by Eq. 7.

After each calculation of the interestingness score $isp(l_j)$ for the location l_j and of the interestingness score $isp(u_k)$ for the user u_k , we perform L2-normalization as shown in Equations 10 and 11.

$$isp(l_j) = \frac{isp(l_j)}{\sqrt{\sum_{y=1}^{|L|} isp(l_y)^2}} \quad (10)$$

$$isp(u_k) = \frac{isp(u_k)}{\sqrt{\sum_{x=1}^{|U|} isp(u_x)^2}} \quad (11)$$

As mentioned above, the calculation of the interestingness of all locations in L is iterative. It is necessary to assign a number of iterations (t) to limit the number of computations. All the procedure details are shown in Algorithm 1. The user scores based on frequency and preference of visiting are initialized as 1. Then, for each iteration i of computation ($1 \leq i \leq t$), the interestingness score of each location $l_j \in L$ based on frequency and preference of visiting are first calculated by applying Eq. 1 and Eq. 7. Af-

ter that, the interestingness score of each user $u_k \in U$ based on his or her frequency and preference of visiting are then calculated. At the end of the process, each location's interestingness score is retained for further computation.

3.2.2 Ranking locations in the recommendation list

To identify interesting locations, the interestingness scores of the locations (based on frequency and preference of visiting calculated from the previous step) are considered. Then, the locations with higher interestingness scores (either on frequency or preference) are ranked with lower numbers. On the other hand, the locations with lower interestingness scores are ranked with higher numbers.

As detailed in the Ranking procedure of Algorithm 1, the ranking number is initialized to be 1. Then, the ranking process is performed repeatedly until all of locations are ranked. First, the maximum value of interestingness score on frequency and preference of visiting, max^{freq} and max^{pref} , are computed. Then, each location l_j having interestingness score $isp(l_j)$ equal to max^{freq} is identified and included into the set L^{freq} . Likewise, each location l_k having $isp(l_k) = max^{pref}$ is included into L^{pref} . Next, the values of max^{freq} and max^{pref} are compared to assign ranking numbers to locations in L^{freq} and L^{pref} in three cases:

1. If $max^{freq} = max^{pref}$, all locations either in L^{freq} or L^{pref} are ranked with the same ranking number.
2. If $max^{freq} < max^{pref}$, all locations in L^{pref} are ranked first and all locations in L^{freq} are ranked later.
3. If $max^{freq} > max^{pref}$, all locations in L^{freq} are ranked first and all locations in L^{pref} are ranked later.

After the ranking, all locations in L^{freq} and L^{pref} are removed from the set of (unranked) locations L and all of their interestingness scores are also removed from ISF^L and ISP^L . The details are shown in the *Ranking* procedure of Algorithm 1.

3.2.3 Generation of a Top-N recommended list

To recommend interesting locations to a new user nu_x , the user needs to give his current location, $loc(nu_x)$. Then, the system selects and ranks the N most interesting locations such that each location is not far from his current location (i.e., each has distance from $loc(nu_x)$ less than the prior-assigned maximum distance, z). The details are shown in the *TopNGeneration* procedure of Algorithm 1.

4. EXPERIMENTAL EVALUATION

In this section, we describe real-world datasets used in this experiment, the experiment settings, and performance measurements. We evaluate the accuracies of the recommendation lists and of the rankings of the top N most interesting locations obtained

Algorithm 1 N-most interesting location-based recommender system

Require: A visiting database, $DB = \{v_1, v_2, \dots, v_z\}$,
 A set of locations, $L = \{l_1, l_2, \dots, l_q\}$,
 A set of users, $U = \{u_1, u_2, \dots, u_p\}$,
 A number of iterations for processing, t
 Current location of a new-user nu , $loc(nu)$,
 A maximum distance (in km) between current location of the user and a recommended shop, z , and
 A number of shops to be recommended, N

Ensure: An order list of N -most interesting locations, $L^{nu} = \{l_y, \dots, l_z\}$

$ISF^t, ISP^t = \text{CalculationInterestingScore}(L, U, DB, t)$
 $IL = \text{Ranking}(L, ISF^t, ISP^t)$
 $L^{nu} = \text{TopNGeneration}(loc(nu), d, N, IL, L)$

*** Procedure $\text{CalculationInterestingScore}(L, U, DB, t)$

- $ISF^t = \{isf(u_k) | u_k \in U \wedge isf(u_k) = 1\}$
- $ISP^t = \{isp(u_k) | u_k \in U \wedge isp(u_k) = 1\}$

for all i^{th} iteration where $i \leq t$ do

- $ISF^i = \{isf(l_j) | l_j \in L \wedge isf(l_j) = \sum_{z=1}^p (f_z \times isf(u_k))\}$ (eq. 1)
- $ISP^i = \{isp(l_j) | l_j \in L \wedge isp(l_j) = \sum_{z=1}^p (f_z \times pref(u_k) \times isp(u_k))\}$ (eq. 7)
- $ISF^i = \{isf(u_k) | u_k \in U \wedge isf(u_k) = \sum_{y=1}^p (f_y \times isf(l_j))\}$ (eq. 2)
- $ISP^i = \{isp(u_k) | u_k \in U \wedge isp(u_k) = \sum_{y=1}^p (f_y \times pref(l_j) \times isp(l_j))\}$ (eq. 9)
- normalize interesting scores in $ISF^i, ISF^i, ISP^i, ISP^i$ (eq. 3, 4, 10 and 11)

end for

*** Procedure $\text{Ranking}(L, ISF^t, ISP^t)$

- $IL = \emptyset, rank = 1$

while $L \neq \emptyset$ do

- $max^{isf} = \max(isf(l_1), isf(l_2), \dots, isf(l_{|L|}))$
- $max^{isp} = \max(isp(l_1), isp(l_2), \dots, isp(l_{|L|}))$
- $L^{freq} = \{l_j \in L | isf(l_j) = max^{isf}\}$
- $L^{pref} = \{l_k \in L | isp(l_k) = max^{isp}\}$

if $max^{freq} = max^{pref}$

- $IL = IL \cup \{<l_j, rank> | l_j \in L^{freq}\}$
- $IL = IL \cup \{<l_k, rank> | l_k \in L^{pref}\}$

else if $max^{freq} > max^{pref}$

- $IL = IL \cup \{<l_j, rank> | l_j \in L^{freq}\}$
- $rank++$
- $IL = IL \cup \{<l_k, rank> | l_k \in L^{pref}\}$

else

- $IL = IL \cup \{<l_k, rank> | l_k \in L^{pref}\}$
- $rank++$
- $IL = IL \cup \{<l_j, rank> | l_j \in L^{freq}\}$

end if

- $ISF^i = ISF^i - \{isf(l_j) | l_j \in L^{freq}\}$
- $ISP^i = ISP^i - \{isp(l_k) | l_k \in L^{pref}\}$
- $L = L - (L^{freq} \cup L^{pref})$
- $rank++$

end while

*** Procedure $\text{TopNGeneration}(loc(nu), d, N, IL, L)$

- $L^{nu} = \emptyset, rank = 0$

for each location l_j in IL and $rank < N$ do

if $diff(loc(nu), loc(l_j)) \leq z$

- $L^{nu} = L^{nu} \cup \{l_k | l_k \in IL, rank_{l_k} = rank_{l_j}\}$
- $rank++$

end if

end for

from our recommendation system. The results are then compared with those of the HITS-based model. Lastly, we provide a performance analysis and related discussion.

4.1 Real-world datasets

For our experiments, we use a real-world dataset provided by Foursquare. The dataset [12] contains users' location-check-in histories. Each data entry consists of user ID, location ID, location category ID, a location category name, latitude, longitude, time zone offset, and UTC time collected from 12 April 2012 to 16 February 2013 (covering approximately 10 months). We focus only on locations in the most popular category (food category). Via Google Maps Geocoding API¹, we transform longitude and latitude data into the names of the physical areas. We then select five areas (including Chiyoda, Shibuya, Minato, Shinjuku and Chuo) in Tokyo and the New York area. These areas' associated data contain the highest frequency check-ins (i.e., the largest number of location check-ins) for this study. Since this dataset size is large and very sparse, we pre-process the dataset by discarding users who visited fewer than two locations, and locations each of which had fewer than two visitors. The remaining dataset is divided into a training dataset (80% of the total number of the remaining users) and a testing dataset (20%) for five-fold cross validation. In the testing phase, the users in the testing dataset are regarded as new users (to simulate cold-start users). Note that our method does not require using a new users' profile to determine potential locations. The characteristics of the dataset are shown in Table 2.



Fig.4: The example of current location simulation of a user.

4.2 Experiment settings and evaluation method

We compare our method with the HITS-based model (as adopted by [1] and [2]). Other than being classic, the HITS-based model shares the same target as ours. It aims to recommend locations to new users (who have no profile or check-in histories). The Chiyoda, Shinjuku, Minato, Shinjuku, Chuo and New York datasets are sparse, so we define that the

¹<https://developers.google.com/maps/documentation/geocoding/>

Table 2: The characteristics of the dataset

City	Users	Locations	Total check-ins	Check-ins per location	Check-ins per user
Chiyoda	958	827	7,283	8.81	7.60
Shibuya	581	719	4,107	5.71	7.10
Minato	519	719	3,836	5.34	7.39
Shinjuku	571	670	3,905	6.84	5.83
Chuo	243	325	1,645	5.06	6.77
New York	921	3,609	26,494	7.34	28.77

largest number of recommended locations as 10 (Top-N, where N is 10). From this, we consider recommending locations only to the users who have more than 10 candidate locations. In all experiments, a location in the testing dataset can be suggested only if at least one user had visited or checked in the location more than 1 time. As users' current locations are not specified in the dataset, we simulate them by creating each user's minimum bounding rectangle (MBR, shown as the dashed line in Fig. 4) ([1] and [13]). All locations located in a MBR are considered to determine whether or not they should be recommended to the user. In Fig. 4, we show an example of the current-location simulation using MBR in which the ground-truth locations are shown in red, and the recommended ones are shown in green. The algorithm performs perfectly only if it can recommend all the ground-truth locations (which are the ones the user visited at least once). The data related to ground-truth locations is not included in the training process.

For measuring the accuracy, we use four metrics reflecting the correctness of the recommended top-N locations. Specifically, we use precision, recall, average ranking, and NDCG, as shown in Equations 12, 13, 14, and 15 respectively.

$$Precision = \frac{\#correct_retrieved}{N} \quad (12)$$

$$Recall = \frac{\#correct_retrieved}{\#relevant} \quad (13)$$

$$AverageRanking = \frac{\sum rank}{\#testloc} \quad (14)$$

$$NDCG = \frac{DCG}{IDCG} \quad (15)$$

N is the number of recommendations. $\#correct_retrieved$ is the number of ground-truth locations retrieved correctly. $\#relevant$ is the number of all ground-truth locations. $\#testloc$ is the number of the ground-truth locations recommended by NILR or the HITS-based model. $rank$ is the position of the ground-truth locations in the recommendation list. DCG denotes the discounted cumulative gain (i.e., the position score of the ground-truth locations in the recommended list, [15]) $IDCG$ denotes the idealized

discounted cumulative gain (i.e., the perfect position score of the locations in the ideal recommendation list, [15])

4.3 Results and Discussion

Examining the experimental results, we analyze the accuracies of the recommendation lists in terms of precision, recall, average ranking, and NDCG against the varied number of recommendations. As shown in Figures 5 and 6, we compare the accuracies of our recommendation lists with those of the HITS-based model in terms of precision and recall, respectively. The performance comparisons measured by both metrics are similar. The precision and recall values obtained by our method are higher on average in all areas (except for the Shinjuku area). That anomaly occurs is because users in Shinjuku usually go to the shops of high check-in frequencies but have a low number of revisits, complying with the HITS-based model. In contrast, users in the other areas also users frequently revisited shops. Consequently, in the latter cases, our method outperforms the HITS-based model. Furthermore, we observe that the precision and recall values are quite low in our NILR approach and also the traditional HITS-based model as a result of the fact that we designed the offline experiments based on the check-in histories provided by Foursquare [12]. The evaluations were done by comparing the recommendation lists against the check-in histories (instead of being suggested to actual users). If we were to evaluate our method with real users, we would expect higher accuracy as the users would be able to look up and would be inclined to choose the locations suggested by the lists. In another case, the Foursquare users might actually have gone to certain locations being recommended by NILR but the subjects have forgotten to check-in via Foursquare. Those locations were excluded from the set of ground-truth locations in our experiments, resulting in the low precision and recall values. In summary, compared with the HITS-based model, our method yields the precision/recall values 6.37%/6.13%, and 24.43%/30.20% higher in Tokyo and New York, respectively. In terms of ranking ac-

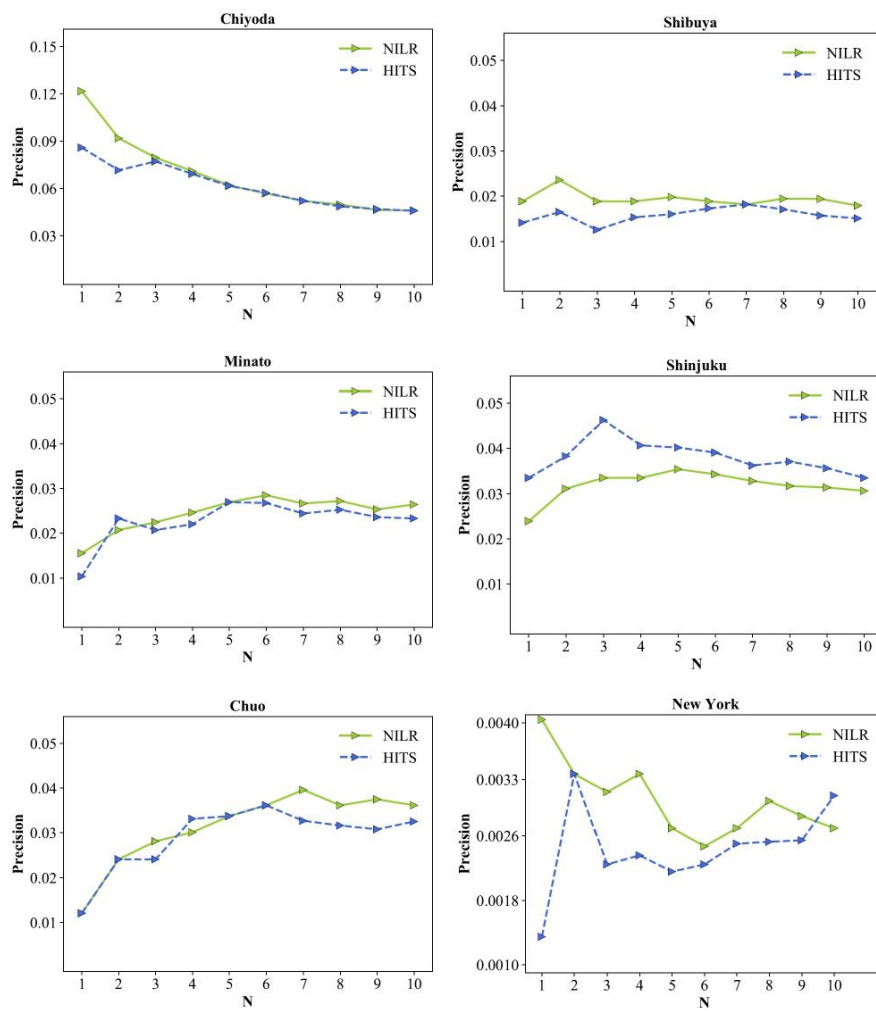


Fig.5: Precision metric w.r.t. Recommendation numbers.

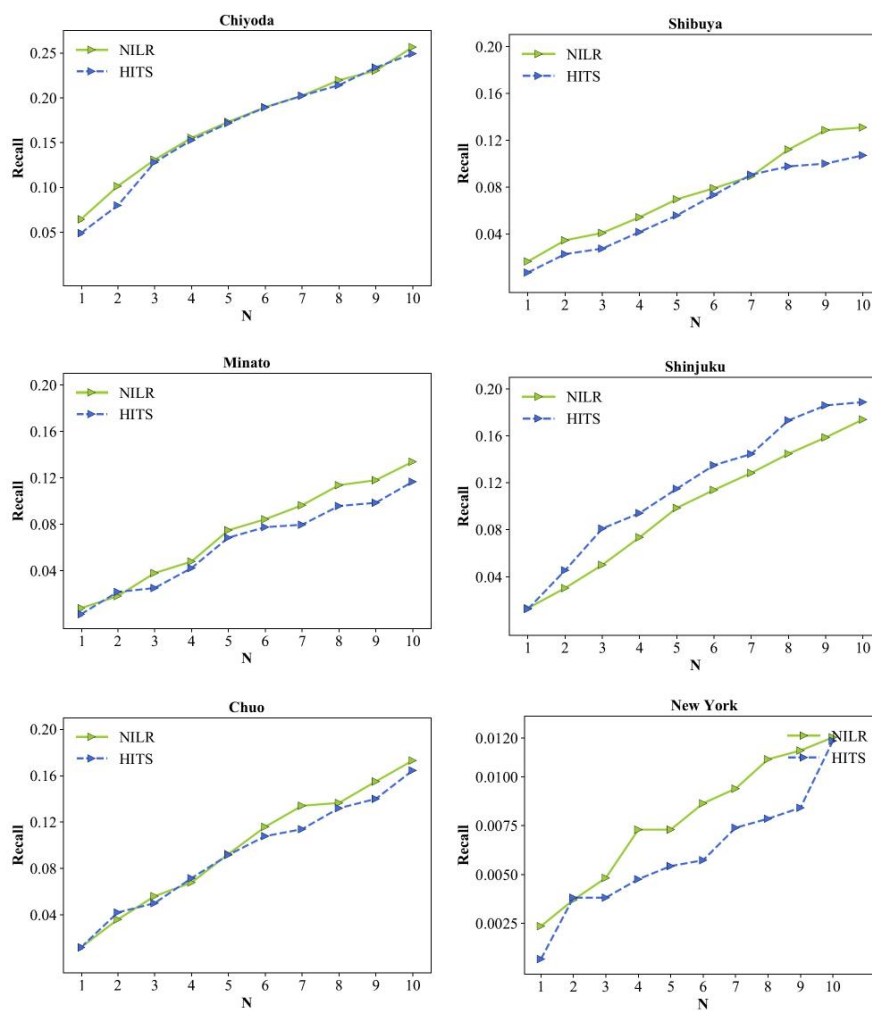


Fig.6: Recall metric w.r.t. Recommendation numbers.

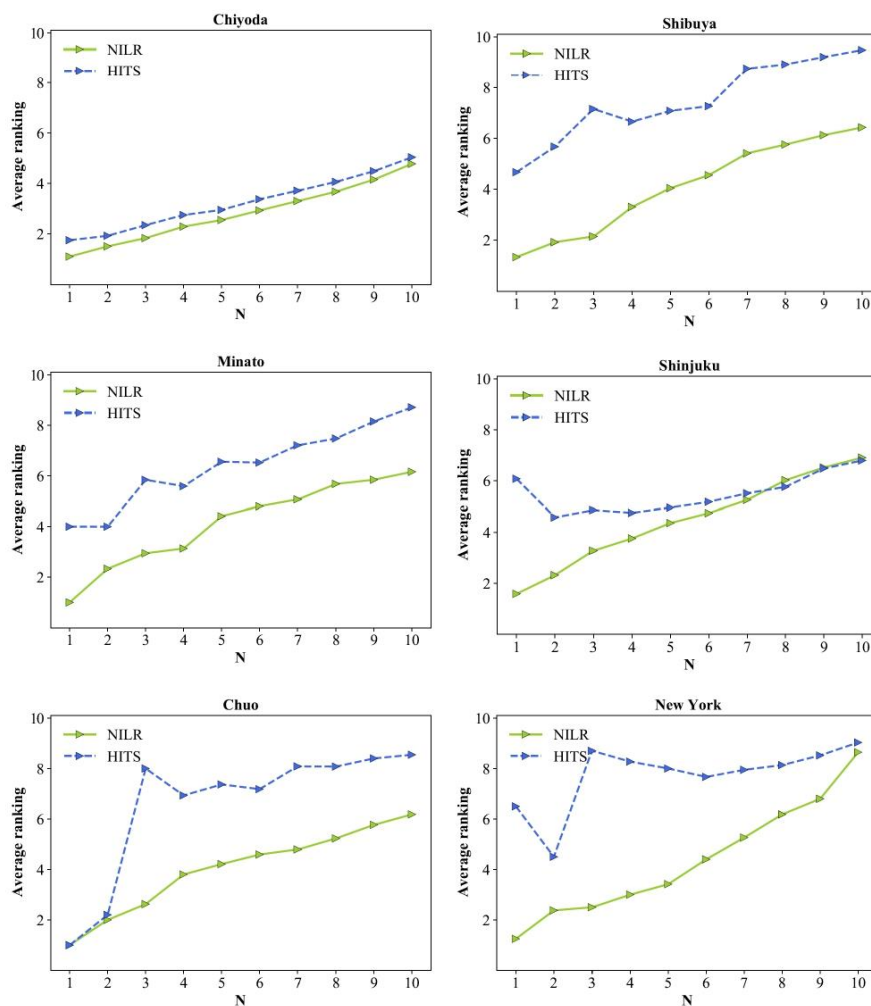


Fig. 7: Average ranking metric w.r.t. Recommendation numbers.

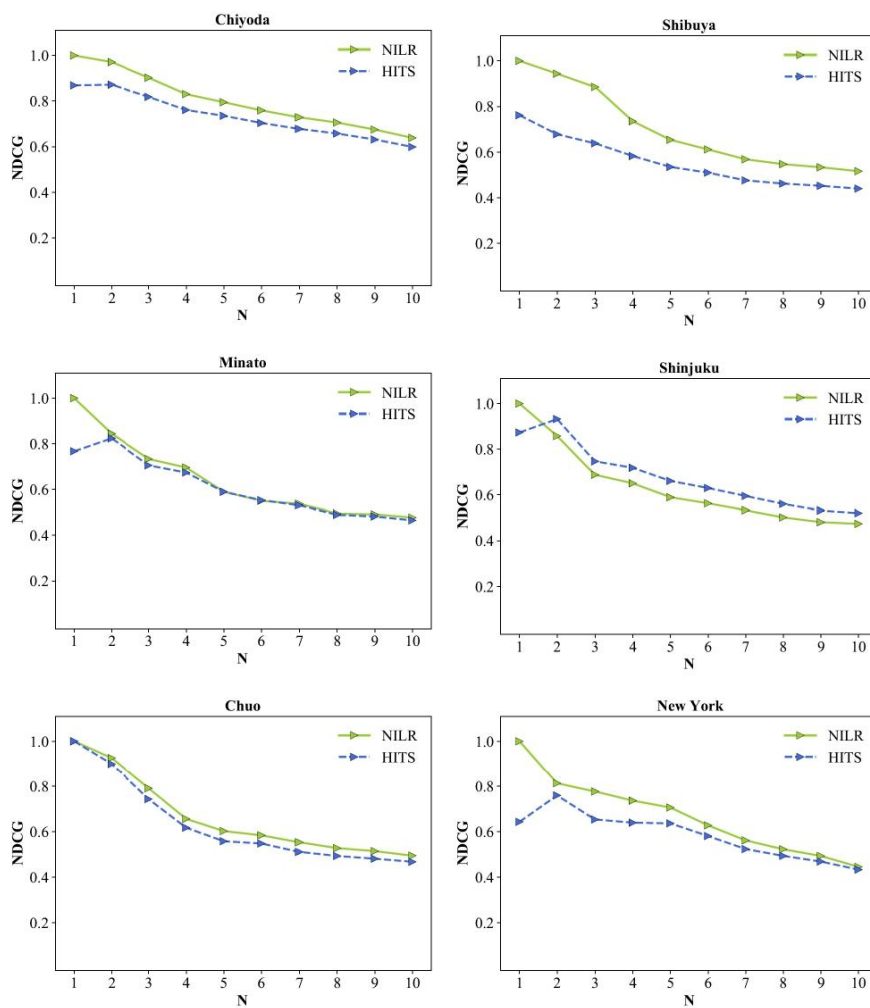


Fig.8: NDCG metric w.r.t. Recommendation numbers

accuracy, we show the values of average ranking of the recommendation lists in Fig. 7. We can see that our method ranks the ground-truth locations in all areas lower than those of the HITS-based model. The average ranking accuracy of our method is 30.27% and 43.28% more in Tokyo and New York, respectively. Finally, Fig. 8 shows the average ranking accuracy in terms of NDCG. Our method outperforms the HITS-based model by 8.01% and 14.60% in Tokyo and New York, respectively.

The comparison results measured in terms of precision, recall, and NDCG are similar. The HITS-based model outperforms ours (by giving lower ranks to the ground-truth locations) in the Shinjuku area. However, if we use the average ranking as the metric, the comparison results are different, especially in that area. This is due to the fact that calculating the average ranking is based on the average of non-normalized ranks, while determining the NDCG is based on ranking numbers normalized by log scale. In most of the cases in the Shinjuku area, the HITS-based model outperforms our method by just a little. However, there are a few cases in the area where our method outperforms the HITS-based model by a large margin. NDCG normalizes the differences with a log scale which makes the large gaps smaller. As a consequence, the HITS-based model outperforms ours on average in the Shinjuku area.

For the other areas, our method yields more accurate recommendation lists with better rankings. This is due to the fact that users often choose to visit locations revisited many times (defined as *locations based on preference* in Section 3) as opposed to the places visited many times by various people (defined as *locations based on frequency* in Section 3). Our method retrieves and recommends locations which are outstanding in both the frequency and preference aspects while the HITS-based model is good at retrieving locations outstanding only in the frequency aspect. In some cases where both methods can retrieve and recommend the same locations, when considering the preference aspect, our method assigns lower ranks to the ground truth locations.

5. CONCLUSION

In this work, we introduce an enhanced location-recommendation approach called N-most interesting location-based recommender system (NILR) to recommend interesting locations for new users to visit. Our method includes three phases: 1) interesting locations are extracted by regarding both visiting frequencies and users' preference, 2) interesting locations are ranked based on two interestingness scores: one obtained from the HITS-based model and the other from our proposed method, and 3) interesting locations are re-filtered based on the current location of each new user. Experimental results from a well-known real-world dataset prove our NILR method

outperforms the HITS-based model in terms of recommendation accuracy and ranking. Our method yields more accurate rankings than that of the HITS-based model by 30.27% and 43.28% for ranking recommended places in Tokyo city and New York respectively.

ACKNOWLEDGEMENT

This work was supported by Faculty of Informatics, Burapha University.

References

- [1] J. Bao, Y. Zheng and M. F. Mokbel, "Location-based and preference-aware recommendation using sparse geo-social networking data," in *Proceedings of the 20th international conference on advances in geographic information systems*, pp. 199-208, 2012.
- [2] H. Bagci and P. Karagoz, "Context-aware location recommendation by using a random walk-based approach," *Knowledge and Information Systems*, vol.47, no.2, pp.241-260, 2016.
- [3] Y. Zheng and X. Zhou, *Location-based social networks: Users*, pp. 243-276, 2011.
- [4] Y. Zheng, L. Capra, O. Wolfson and H. Yang, "Urban computing: Concepts, methodologies, and applications," *ACM Trans. Intell. Syst. Technol.*, vol.5, no.3, Sep 2014.
- [5] Y. Yu and X. Chen, "A survey of point-of-interest recommendation in location-based social networks," in *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
- [6] S. Xu, X. Fu, J. Cao, B. Liu and Z. Wang, "Survey on user location prediction based on geo-social networking data," *World Wide Web*, pp. 1-44, 2020.
- [7] D. Zhou, B. Wang, S. M. Rahimi and X. Wang, "A study of recommending locations on location-based social network by collaborative filtering," In Leila Kosseim and Diana Inkpen, editors, *Advances in Artificial Intelligence*, pages 255266, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- [8] J. J. Levandoski, M. Sarwat, A. Eldawy and M. F. Mokbel, "Lars: A location-aware recommender system," in *2012 IEEE 28th International Conference on Data Engineering*, pp. 450-461, 2012.
- [9] Zhijun Ding, Xiaolum Li, Changjun Jiang, and Mengchu Zhou. Objectives and state-of-the-art of location-based social network recommender systems. *ACM Comput. Surv.*, 51(1), January 2018.
- [10] Y. Zheng, L. Zhang, X. Xie and W.Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th international conference on World wide web*, pp. 791-800, 2009.
- [11] S. Darapisut, K. Amphawan, S. Rimcharoen and N. Leelathakul, "Nilr: N-most interesting

- location-based recommender system,” in *Conference on Smart Media and Applications*, 2020.
- [12] D. Yang, D. Zhang, V. W. Zheng and Z. Yu, “Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol.45, no.1, pp.129-142, 2014.
- [13] Y. Ying, L. Chen and Gencai Chen, “A temporal-aware poi recommendation system using context-aware tensor decomposition and weighted hits,” *Neurocomputing*, vol.242, pp.195-205, 2017.
- [14] R. Burke, *Hybrid Web Recommender Systems*, pp. 377-408, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.
- [15] C. Biancalana, F. Gasparetti, A. Micarelli and G. Sansonetti, “An approach to social recommendation for context-aware mobile services,” *ACM Trans. Intell. Syst. Technol.*, vol.4, no.1, pp.1-31, Feb. 2013.
- [16] S. Zhao, I. King and M. R. Lyu, “A survey of point-of-interest recommendation in location-based social networks,” *arXiv preprint*, arXiv:1607.00647, 2016.
- [17] X. Xiao, Y. Zheng, Q. Luo and X. Xie, “Finding similar users using category-based location history,” in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 442-445, New York, USA, 2010.
- [18] P. Symeonidis, A. Papadimitriou, Y. Manolopoulos, P. Senkul, and I. Toroshu, “Geo-social recommendations based on incremental tensor reduction and local path traversal,” in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pp. 89-96, New York, USA, 2011.
- [19] H. Wang, G. Li and J. Feng, “Group-based personalized location recommendation on social networks,” in *Lei Chen, Yan Jia, Timos Sellis, and Guanfeng Liu, editors, Web Technologies and Applications*, pp. 68-80, Cham, 2014. Springer International Publishing.
- [20] Q. Yuan, G. Cong, Z. Ma, A. Sun and N. Ma. Thalmann, “Time-aware point-of-interest recommendation,” in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '13*, pp. 363-372, New York, USA, 2013.
- [21] K. H. Lim, J. Chan, C. Leckie and S. Karunasekera, “Personalized trip recommendation for tourists based on user interests,” *points of interest visit durations and visit recency*, vol.54, no.2, pp. 375-406, February 2018.
- [22] K. H. Lim, J. Chan, S. Karunasekera and C. Leckie, “Tour recommendation and trip planning using location-based social media: a survey,” *Knowledge and Information Systems*, pp. 1-29, 2019.
- [23] V. W. Zheng, B. Cao, Y. Zheng, X. Xie and Q. Yang, “Collaborative filtering meets mobile recommendation: A user-centered approach,” in *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI'10*, pages 236-241. AAAI Press, 2010.
- [24] V. W. Zheng, Y. Zheng, X. Xie and Q. Yang, “Collaborative location and activity recommendations with gps history data,” in *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pp. 1029-1038, New York, USA, 2010.
- [25] X. Liu, Y. Liu, K. Aberer and C. Miao, “Personalized point-of-interest recommendation by mining users' preference transition,” in *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management, CIKM '13*, pp. 733-738, New York, USA, 2013.
- [26] D. Preotjuc-Pietro and T. Cohn, “Mining user behaviours: a study of check-in patterns in location based social networks,” in *Proceedings of the 5th annual ACM web science conference*, pp. 306-315, 2013.
- [27] Y. Si, F. Zhang and W. Liu, “Ctfara: An adaptive method for poi recommendation based on check-in and temporal features,” *Knowledge-Based Systems*, vol.128, pp.59-70, 2017.
- [28] J. Cao, S. Xu, X. Zhu, R. Lv and B. Liu, “Effective fine-grained location prediction based on user check-in pattern in lbsns,” *Journal of Network and Computer Applications*, vol.108, pp.64-75, 2018.
- [29] R. Baral and T. Li, “Exploiting the roles of aspects in personalized poi recommender systems,” *Data Mining and Knowledge Discovery*, vol.32, no.2, pp.320-343, 2018.
- [30] M. Gan and L. Gao, “Discovering memory-based preferences for poi recommendation in location-based social networks,” *ISPRS International Journal of Geo-Information*, vol.8, no.6, 279, 2019.
- [31] M. Xin, L. Wu and S. Li, “A user profile awareness service collaborative recommendation algorithm under lbn environment,” *International Journal of Cooperative Information Systems*, 28(03):1950008, 2019.
- [32] J. M. Kleinber., “Authoritative sources in a hyperlinked environment,” *Journal of the ACM (JACM)*, vol.46, no.5, pp.604-632, 1999.
- [33] X. Long and J. Joshi, “A hits-based poi recommendation algorithm for location-based social networks,” in *2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, pp. 642-647, 2013.
- [34] J. Chen, X. Li, W.K. Cheung and K. Li, A survey of point-of-interest recommendation in location-

based social networks Effective successive poi recommendation inferred with individual behavior and group preference," *Neurocomput.*, 210 pp.174-184, Oct. 2016.

- [35] D. Puspitaningrum, J. Fernando, E. Afriando, F. P. Utama, R. Rahmadini and Y. Pinata, "Finding Local Experts for Dynamic Recommendations Using Lazy Random Walk," *2019 7th International Conference on Cyber and IT Service Management (CITSM)*, pp. 1-6, 2019.



Komate Amphawan received BS and MS degrees in Computer Science in 2003 and 2006, respectively. Then, he received Ph.D. degree in Computer Engineering in 2011. Since 2011, he has been a faculty member in Faculty of Informatics, Burapha University, Thailand. His research interests include the area of data mining and knowledge discovery.



Sunisa Rimcharoen received a B.Sc. degree in computer science from Burapha University, Thailand, in 2003, an M.Sc. degree in computer science from Chulalongkorn University in 2005, and a Ph.D. degree in computer engineering from Chulalongkorn University in 2009.

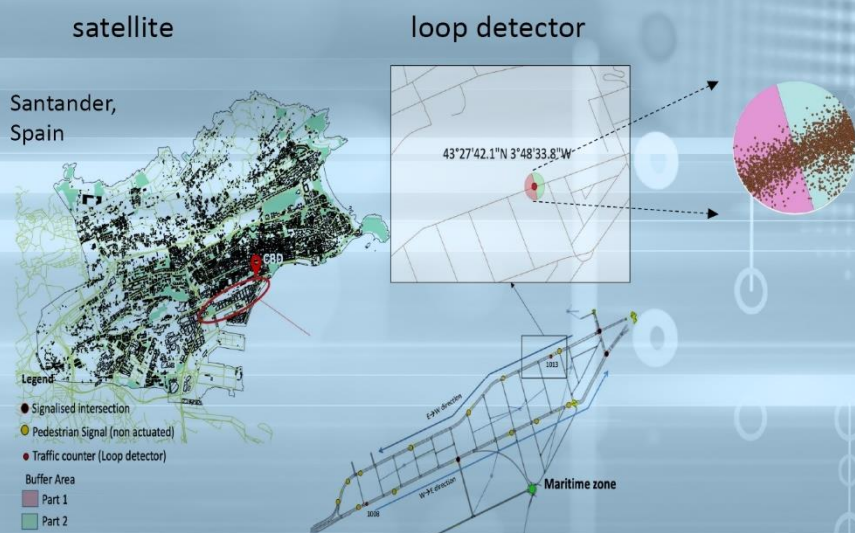
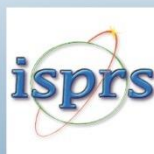
Dr. Sunisa has been a faculty member of the Faculty of Informatics, Burapha University, Chon Buri, Thailand, since 2009. Her current research interests include evolutionary computation, machine learning, prediction, and algorithmic trading.



Sumet Darapisut received BS and MS degrees in Computer Science in 2013 and 2016, respectively. Currently, he is working toward the PhD degree in the Faculty of Informatics, Burapha University, Thailand. His research interests include recommender system and data mining.



Nutthanon Leelathakul received a B.Eng. degree with first-class honors in electrical engineering from Kasetsart University, Bangkok, Thailand, in 1999, an M.Eng. degree in electrical and computer engineering from Cornell University in 2004, and a Ph.D. degree in electrical and computer engineering from Carnegie Mellon University in 2010. Dr. Nutthanon has been a faculty member of the Faculty of Informatics, Burapha University, Chon Buri, Thailand, since 2010. His current research interests include computer networking, security in IoT, deep learning, and data analytics.



Estimation of a Fundamental Diagram with Heterogeneous Data Sources in Santander

Volume 12 · Issue 10 | October 2023



mdpi.com/journal/ijgi
ISSN 2220-9964

Article: Deformation of High Rise Cooling Tower through Projection of Coordinates Resulted from Terrestrial Laser Scanner Observations onto a Vertical Plane

Ashraf A. A. Beshr, Ali M. Basha, Samir A. El-Madany and Fathi Abd El-Azeem
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 417; DOI: [10.3390/ijgi12100417](https://doi.org/10.3390/ijgi12100417)

Article: Identifying Urban Park Events through Computer Vision-Assisted Categorization of Publicly-Available Imagery

Yizhou Tan, Wenjing Li, Da Chen and Waishan Qiu
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 419; DOI: [10.3390/ijgi12100419](https://doi.org/10.3390/ijgi12100419)

Article: A Self-Attention Model for Next Location Prediction Based on Semantic Mining

Eric Hsueh-Chan Lu and You-Ru Lin
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 420; DOI: [10.3390/ijgi12100420](https://doi.org/10.3390/ijgi12100420)

Article: Enhancing the Understanding of the EU Gender Equality Index through Spatiotemporal Visualizations

Laya Targa, Silvia Rueda, Jose Vicente Riera, Sergio Casas and Cristina Portalés
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 421; DOI: [10.3390/ijgi12100421](https://doi.org/10.3390/ijgi12100421)

Article: Deep-Learning-Based Annotation Extraction Method for Chinese Scanned Maps

Xun Rao, Jiasheng Wang, Wenjing Ran, Mengzhu Sun and Zhe Zhao
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 422; DOI: [10.3390/ijgi12100422](https://doi.org/10.3390/ijgi12100422)

Article: The Seasonal Migrants Spatially Affect the Park Green Space Accessibility and Equity under Different Travel Modes: Evidence from Sanya, China

Wentong Yang, Yeqing Cheng, Chunru Xu and Jinping Zhang
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 423; DOI: [10.3390/ijgi12100423](https://doi.org/10.3390/ijgi12100423)

Article: A Knowledge-Guided Fusion Visualisation Method of Digital Twin Scenes for Mountain Highways

Ranran Tang, Jun Zhu, Ying Ren, Yongzhe Ding, Jianlin Wu, Yukun Guo and Yakun Xie
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 424; DOI: [10.3390/ijgi12100424](https://doi.org/10.3390/ijgi12100424)

Article: The Analyses of Land Use and Prevention in High-Density Main Urban Areas under the Constraint of Karst Ground Subsidence: Study of Wuhan City, China

Lin Gao, Yan Shi, Yang Qiu, Chuanming Ma and Aiguo Zhou
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 425; DOI: [10.3390/ijgi12100425](https://doi.org/10.3390/ijgi12100425)

Article: Extraction of Urban Road Boundary Points from Mobile Laser Scanning Data Based on Cuboid Voxel

Jingxue Wang, Xiao Dong and Guangwei Liu
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 426; DOI: [10.3390/ijgi12100426](https://doi.org/10.3390/ijgi12100426)

Article: Automated Generation of Room Usage Semantics from Point Cloud Data

Guoray Cai and Yimu Pan
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 427; DOI: [10.3390/ijgi12100427](https://doi.org/10.3390/ijgi12100427)

Article: The Study of Regional Innovation Network Structure: Evidence from the Yangtze River Delta Urban Agglomeration

Jie Chen, Liang Jiang, Ye Tian and Jing Luo
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 428; DOI: [10.3390/ijgi12100428](https://doi.org/10.3390/ijgi12100428)

Article: A Hybrid Discrete Artificial Bee Colony Algorithm Based on Label Similarity for Solving Point-Feature Label Placement Problem

Wen Cao, Jiaqi Xu, Yong Zhang, Siqi Zhao, Chu Xu and Xiaofeng Wu
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 429; DOI: [10.3390/ijgi12100429](https://doi.org/10.3390/ijgi12100429)

Article: Describing the Urban Jungle: A Multicriteria Urbanization Index for the Amazon

Pablo Cabrera-Barona, Denise Albán and Gustavo Durán
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 430; DOI: [10.3390/ijgi12100430](https://doi.org/10.3390/ijgi12100430)

Article: An Efficient and Accurate Convolution-Based Similarity Measure for Uncertain Trajectories

Guanyao Li, Xingdong Deng, Jianmin Xu, Yang Liu, Ji Zhang, Simin Xiong and Feng Gao
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 432; DOI: [10.3390/ijgi12100432](https://doi.org/10.3390/ijgi12100432)

Article: A Hybrid POI Recommendation System Combining Link Analysis and Collaborative Filtering Based on Various Visiting Behaviors

Sumet Darapisut, Komate Amphawan, Nutthanon Leelathakul and Sunisa Rimcharoen
ISPRS Int. J. Geo-Inf. **2023**, *12*(10), 431; DOI: [10.3390/ijgi12100431](https://doi.org/10.3390/ijgi12100431)

Article

A Hybrid POI Recommendation System Combining Link Analysis and Collaborative Filtering Based on Various Visiting Behaviors

 Sumet Darapisut, Komate Amphawan *, Nutthanon Leelathakul  and Sunisa Rimcharoen 

 Faculty of Informatics, Burapha University, Chonburi 20131, Thailand; dearsumet@gmail.com (S.D.);
 nutthanon@buu.ac.th (N.L.); rsunisa@buu.ac.th (S.R.)

* Correspondence: komate@gmail.com

Abstract: Location-based recommender systems (LBRs) have exhibited significant potential in providing personalized recommendations based on the user's geographic location and contextual factors such as time, personal preference, and location categories. However, several challenges (such as data sparsity, the cold-start problem, and tedium problem) need to be addressed to develop more effective LBRs. In this paper, we propose a novel POI recommendation system, called LACF-Rec3, which employs a hybrid approach of link analysis (HITS-3) and collaborative filtering (CF-3) based on three visiting behaviors: frequency, variety, and repetition. HITS-3 identifies distinctive POIs based on user- and POI-visit patterns, ranks them accordingly, and recommends them to cold-start users. For existing users, CF-3 utilizes collaborative filtering based on their previous check-in history and POI distinctive aspects. Our experimental results conducted on a Foursquare dataset demonstrate that LACF-Rec3 outperforms prior methods in terms of recommendation accuracy, ranking precision, and matching ratio. In addition, LACF-Rec3 effectively solves the challenges of data sparsity, the cold-start issue, and tedium problems for cold-start and existing users. These findings highlight the potential of LACF-Rec3 as a promising solution to the challenges encountered by LBRs.

Keywords: point-of-interest recommendations; link analysis; collaborative filtering; distinctiveness



Citation: Darapisut, S.; Amphawan, K.; Leelathakul, N.; Rimcharoen, S. A Hybrid POI Recommendation System Combining Link Analysis and Collaborative Filtering Based on Various Visiting Behaviors. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 431. <https://doi.org/10.3390/ijgi12100431>

Academic Editors: Wolfgang Kainz and Haosheng Huang

Received: 31 August 2023

Revised: 15 October 2023

Accepted: 20 October 2023

Published: 22 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As location-based data from mobile devices and social networks are increasingly available, location-based recommender systems (LBRs) [1–7] hold significant promise in providing personalized recommendations to users based on their locations and contextual factors. LBRs recommend points of interest (POIs) such as restaurants, museums, and shopping centers to users after considering the users' current location, historical check-in records, social context, and other relevant information. The primary objective of LBRs is to facilitate users in discovering captivating venues, improving their user experience, and contributing to the growth of local businesses and tourism.

LBRs have attracted significant attention from researchers and industry professionals due to their potential implications across diverse domains including tourism, transportation, and marketing. In tourism, for instance, LBRs help visitors explore novel and captivating attractions, events, and restaurants within unfamiliar cities. In transportation, LBRs can assist commuters and travelers plan their journeys and identify convenient and efficient routes. In marketing, LBRs can facilitate enterprises in promoting their products and services to potential customers based on their location and preferences.

Numerous location-based social network systems (LBSN) platforms including Foursquare, Yelp, and Facebook Places utilize POI recommendation systems to overcome the issue of information overload by identifying interesting venues and filtering out irrelevant options. From the enterprise perspective, these POI recommendation systems can yield substantial profit for proprietors because users can perceive their preferred locales, thereby augmenting user loyalty. Nonetheless, there are several challenges to POI recommender systems.

The first challenge lies in the issue of data sparsity. In real-world scenarios, the number of locations is growing rapidly, however, users tend to check-in only a few times within their immediate vicinity. Consequently, the user-POI matrix becomes sparse, causing difficulties in determining the users' preferences.

The second challenge is the cold-start problem encountered when users, who seek location recommendations (referred to as target users), lack a history of check-ins. In particular, traditional approaches rely on the similarity between places visited by the neighbors of users with existing check-in histories and those visited by target users. Hence, if the target users are new, the LBS struggles to effectively capture their location preferences.

To address these two challenges, link analysis-based (LA) algorithms such as the HITS (Hypertext Induced Topic Search) model [5,8,9] rely on graph techniques to extract quality POIs and local experts. These algorithms can leverage the knowledge and the experience of local experts within local areas (of which the associated data are non-sparse) to recommend potential locations for users without a history of check-ins (referred to as cold-start users). However, if there is a sufficient amount of user check-in data, CF methods are likely to outperform LA methods [10,11].

Collaborative filtering (CF) algorithms [6,12,13] rely on a user's check-in history to provide more accurate recommendations, especially when the number of check-ins of the user has increased. Nevertheless, in the case of the cold-start problem, LA algorithms tend to surpass CF algorithms [5,14,15].

The third challenge involves the tedium problem, where the majority of CF and LA algorithms rely only on the check-in frequency, often leading to the recommendation of predominantly popular places without considering personal preference. This tendency arises because the algorithms do not consider location diversity and novelty.

To tackle all three challenges, we propose a novel POI recommendation system called LACF-Rec3. This system utilizes a hybrid method combining link analysis and collaborative filtering. It is based on three visiting behaviors: frequency, variety, and repetition. LACF-Rec3 identifies interesting and distinctive POIs by analyzing the user- and POI-visit patterns, subsequently ranking them, and delivering recommendations to cold-start users. For users with existing check-in histories (referred to as existing users), LACF-Rec3 performs collaborative filtering based on their check-in history and the distinctive aspect of the POIs.

We finetuned a HITS-based model and called it HITS-3 (HITS based on three check-in behaviors) to discover interesting locales with distinctiveness (i.e., distinctive aspect) in all three location characteristics: visit frequency, user variety, and repeated check-ins. Locations with high check-in frequencies typically denote popular or trending destinations with a high number of visits. High-variety locations signify places visited by various individuals including tourist attractions and landmarks that draw a varied audience without recurring visits. Locations with the great number of repeat check-ins could be routine daily venues such as grocery stores or supermarkets as well as specific places tailored to particular users such as gyms, basketball courts, or board game cafes.

Furthermore, our HITS-3 model also discovers places with distinctiveness in all three user characteristics: check-in frequency, location variety, and the number of locations the user revisits. Users exhibiting high check-in frequency epitomize active and sociable individuals who frequently engage in location-based activities. Those with high location variety demonstrate an affinity for novel experiences and avoid revisiting the same places frequently. Users with a high number of locations revisited often possess specific preferences or display strong loyalty as dedicated patrons of specific locations.

Derived from the aforementioned HITS-based model, the distinctiveness and distinctive scores (explained in Section 3) of all locations and all users could be utilized to recommend locations for cold-start users. For target users with check-in histories, locations are recommended based on the interest scores calculated by our enhanced user-based collaborative filtering algorithm, called CF-3 (Collaborative Filtering based on 3-distinctiveness). Specifically, when a user already has a check-in history, CF-3 suggests locations based on

other users who exhibit similar distinctiveness and distinctive scores. Subsequently, CF-3 selects only interesting POIs that are in the vicinity of the target user's current location.

In this paper, we present the design and implementation of LACF-Rec3 (a combination of the HITS-3 and CF-3 methods), followed by a performance evaluation using a Foursquare dataset. Our experiments reveal that LACF-Rec3 outperforms the previous methods in terms of recommendation accuracy, ranking precision, and matching ratio. These results emphasize the potential of LACF-Rec3 as an effective solution to the challenges confronting LBRs.

In summary, the four main contributions of our novel POI recommendation system (LACF-Rec3) are as follows:

- We introduce a pioneering concept for capturing visiting behavior by considering three key characteristics, namely frequency, variety, and repetition. These are utilized to discover intriguing places and identify the distinctiveness of the POIs and users based on their visiting patterns.
- LACF-Rec3 aims to provide captivating POIs for both cold-start and existing users. Our LACF-Rec3 combines a new extended version of the link analysis approach with a novel collaborative filtering recommendation system.
 - a. We introduce HITS-3, which is an extended version of the link analysis recommendation technique. It integrates a HITS-based model and considers the distinctiveness of both users and POIs to generate the top-ranked POIs specifically for cold-start users.
 - b. We also propose a novel collaborative filtering technique, called CF-3, that considers the similarity in the users' distinctiveness to provide highly personalized recommendations for existing users.
- In addition, we propose a new metric, termed "Matching Ratio". The evaluation of LBRs performance typically relies on standard metrics such as precision and recall. Ideally, they should have been determined by comparing the list of recommended places with the list of locations where users have checked in after reviewing the recommendation. However, in the context of experiments, the locations suggested by the LBRs are not actually presented to users for selection. Hence, precision and recall are typically computed by comparing the list of recommended places with the list of checked-in locations, even though the users have not engaged with the recommended list. As a result, relying solely on precision and recall may not accurately capture the LBRs' recommendation accuracy. To overcome this limitation, we propose a novel evaluation metric that measures the matching between the recommended locations and the preferences of the target users. This new metric supplements the precision and recall measures, thereby enhancing the overall evaluation capabilities.
- We conduct experimental evaluations using a real-world Foursquare dataset. Our LACF-Rec3 method significantly outperforms other methods in terms of the recommendation accuracy and ranking accuracy, as demonstrated by the extensive evaluation experiment results in terms of precision, recall, and NDCG (normalized discounted cumulative gain) [16] metrics.

The remainder of this paper is structured as follows. Section 2 provides an overview of the related work in LBRs and outlines the main approaches and techniques employed in the current literature. Section 3 describes our proposed approach in detail. In Section 4, we provide insight into the experimental setup, the evaluation metrics, and the results of the experiments conducted on the real-world dataset. Section 5 discusses the results, the limitations, and the future directions of the proposed approach.

2. Related Work

Our proposed method generates a POI recommendation list for a cold-start user and an existing user based on the link-analysis technique and collaborative-filtering technique,

hence in this section, we review the existing POI recommendation systems that take advantage of the link-analysis technique and collaborative-filtering technique.

2.1. Link-Analysis Technique

Link-analysis techniques play a crucial role in recommender systems by leveraging the connections and relationships between users, items, or other contexts. In particular, a HITS-based model [8,17–19] is one of the link-analysis algorithms that is widely used in generating POI recommendations. The HITS-based model is a fundamental approach, designed to assess the significance of webpages interlinked to each other. The HITS algorithm assesses the authority scores and hub scores of web pages. The former measures the relevance and importance of a page's topics, while the latter represents a page's ability to link to other authority pages on the same topic. The algorithm iteratively computes both scores. An authority score of a web page, P , is the sum of the hub scores of the pages pointing to P , while P 's hub score is the sum of the authority scores of the pages P points to. In the context of LBSN, the scores of the POIs are determined in a manner analogous to the calculation of authority scores, while those of the users are computed using a method similar to the method for hub scores. For example, the first study of the application of the HITS algorithm to generate a POI recommendation list, called the tree-based hierarchical graph (TBHG), was presented by Zheng et al. [20]. The hub nodes according to the graph structure in the HITS algorithm are the users, and the authority nodes are groups of GPSs that can be addressed as locations. Bao et al. [5] proposed the location-based and preference-aware recommendation method. This approach generates recommendations for places by incorporating collaborative filtering and the HITS algorithm. The weight category hierarchy (WCH) is considered, which involves analyzing the frequency of the visited categories. Once experts in the area are identified, the WCH is further used to find similarities and generate a list of interesting recommendations for the target users. Long and Joshi [14] presented the HITS-based POI recommendation method. This approach improves the HITS algorithm by considering the diversity of check-ins by using entropy, based on the assumption that users with diverse check-ins may indicate their expertise in the area. In addition, this method takes into account the relationships with the user's friends in the social network. Bagci and Karagoz [21] presented Context-aware Location Recommendation with Random Walk method (CLoRW) to generate personalized POI recommendations. This method improves the random walk algorithm by taking into account local experts using traditional HITS methods, POI, the user's current POI, friendship relationship, POI popularity, etc. Ying et al. [15] proposed an approach to generate personalized POI recommendations, and aimed to address the sparsity problem by considering category-based replacement of locations in the CTD (context-aware tensor decomposition) process and by identifying interesting POIs using the WHBPR (weighted HITS-based POI rating) step with the additional consideration of friend relationships. In 2020, an N-most interesting location-based recommender system called NILR [9] was presented to generate a POI recommendation list for a cold-start user. The NILR considered both the frequency of visits and the user's preferences (i.e., the number of locations the user revisits). Then, the ranking procedure was applied to generate a final recommendation list. Sun et al. [22] presented a weighted HITS-based model algorithm to generate POI recommendation lists. This method recommends interesting places based on check-in frequency, number of transit points to the location, and time interval to visit in each location obtained by the improved HITS-based model to all active users (without considering their preferences). Yin et al. [13] proposed a tensor decomposition based collaborative filtering (TDCF) algorithm. The TDCF uses the tensor decomposition structure to consider the relationship of the users and check-in place categories along with the check-in interval of a user. The approach also aims to solve the sparsity problem by filling in the missing check-ins using the tensor decomposition technique. Then, the popularity of the location is determined using the HITS algorithm. The algorithm recommends popular locations to the user if they are close to the user's current location. Recently, a privacy-preserving

time-aware recommendation (PPTA-RM) technique was introduced [23]. This technique incorporated both coarse-grained and fine-grained recommendations to anticipate where users might go in the upcoming time slot. For the coarse-grained level, the method captured the users' preferences for POI categories using an extended matrix factorization technique and predicted the preferences using singular spectrum analysis (SSA). At the fine-grained level, a preferred location was discovered using an improved version of the hyperlink-induced topic search (HITS) algorithm.

2.2. Collaborative Filtering Technique

A collaborative filtering (CF) technique [24–26] is a technique in recommender systems for generating personalized recommendation lists in various domains such as places, movies, songs, news, videos, and events. It can be divided into two primary categories: model-based and memory-based techniques. Model-based CF techniques [26–28] employ mathematical models to learn and predict the user preferences. For example, matrix factorization was used in [26] to recognize latent patterns within user behavior and item characteristics, facilitating personalized recommendations. The matrix factorization techniques use vectors in a latent space to represent users and items, and decomposes the user–item interaction matrix into latent factors, thereby enabling the discovery of hidden patterns in user preferences and item characteristics. On the other hand, the memory-based CF techniques utilize the complete user–item interaction dataset to generate recommendations. By evaluating similarities between the users or items, these techniques identify the closest neighbors and offer personalized recommendations aligned with the user preferences. The memory-based CF technique can be further divided into user-based CF and item-based CF. In user-based CF, the systems [25] actively scan for other users with similar behaviors based on the historical data of the target user and generate recommendation lists. Similarly, item-based CF algorithms [24] search for similar items based on the consumption history of the target item and generates recommendation lists.

Earlier works involved the applications of the CF technique for location-based social networks (LBSNs) are as follows. Baral and Li [27] introduced two combined models for personalized POI recommendation: a ranking-based model and a matrix factorization-based model. They incorporated crucial factors such as visit frequency, social connections, time, location, and categorization into a unified recommendation framework. Zhao et al. [28] proposed a Sentimental-Spatial POI Mining (SPM) method by fusing sentimental and geographical attributes of locations. This work also proposed a Sentimental-Spatial POI Recommendation (SPR) model for personalized recommendations by considering sentiment similarity and geographical distance factors based on matrix factorization to generate a personalized location recommendation list. Yuan et al. [29] proposed a time-aware point-of-interest (POI) recommendation model by improving the user-based CF method. This method considered temporal and geographic information for the time context and solved the sparsity problem by using the smoothing technique to generate efficient POI recommendations. Si et al. [30] introduced an adaptive POI recommendation method called CTF-ARA by considering temporal features and user activity. Users are classified into active and inactive users based on their activity. For inactive users, the popularity of POIs is considered to find the similarity of users based on all time slots to generate POI recommendation lists. For active users, this method considers POI popularity based on sequential time slots to increase the recommendation accuracy of the list of recommended places for users. In 2019, a memory-based POI preference attenuation model algorithm was proposed by Gan and Gao [12]. This approach generated a list of personalized location recommendations for users based on the collaborative filtering method. This method determined the similarity of the user's check-in behavior with other users in the system and increased the importance of the user's recently visited locations using the Ebbinghaus forgetting curve technique. Khazaei and Alimohammadi [31] proposed a context-aware group-oriented location recommendation system (CLGRW) for LBSNs based on a random walk algorithm. CLGRW considers user contexts (e.g., social relationships, personal preferences), location

contexts (e.g., category, popularity, capacity, and spatial proximity), and environmental contexts (e.g., weather, day of the week). These contexts are based on a random walk with restart (RWR) algorithm. Zhang et al. [32] introduced the POI recommendation framework using the users' memory-based preferences and the POI stickiness method (U-CF-Memory-Stickiness), which is an improved version of the memory-based POI preference attenuation model. The U-CF-Memory-Stickiness assigns high scores to locations that the user has recently visited using the Ebbinghaus Forgetting Curve engine. In addition, this method also considers the revisit location with the POI stickiness method based on the collaborative filtering method. Recently, a CULT-TF method [33] was introduced to generate a personalized POI recommendation list. The method combined the contextual information of similar users into the tensor factorization model. A user clustering method was proposed to select active users with the greatest impact and influence. A U-L-T tensor was also proposed as the basis for creating a POI recommender system by considering the user activity, POI popularity, and time slot popularity.

2.3. Comparative Analysis of Baseline Approaches and Our Proposed Method (LACF-Rec3)

The aforementioned research works have provided POI recommendation lists primarily based on the link analysis technique and/or collaborative filtering technique. However, various earlier studies focused only on utilizing frequency. As a result, these location recommendations may suffer from a tedium problem, where locations are chosen predominantly based on their popularity, and not their variety and repetition rates. Therefore, we proposed the concept of location and user distinctiveness, in terms of frequency, variety, and repetition, to capture a broader spectrum of user preferences. In addition, we aimed to provide a tailored recommendation list for both cold-start and existing users by effectively coping with the cold-start and sparsity issues. For cold-start users, interesting locations are recommended by the proposed HITS-3 method, based on the distinctiveness of locations in their vicinity. For existing users, interesting venues are recommended by the CF-3 method.

Table 1 illustrates the comparative analysis between the baseline approaches and our proposed method. We observed that there was one prior research work [14] that recommended locations based only on user variety. Additionally, two previous research works [9,32] considered the check-in repetition. Prior hybrid recommendation systems [5,13,15,23] that had utilized both the link analysis model (i.e., HITS based model) and the collaborative filtering approach were utilized. However, there have been no research works that considered three check-in behaviors (i.e., frequency, variety, and repetition) while taking advantage of the distinctiveness of the users and locations to generate a POI recommendation list.

Table 1. Comparative analysis of the related works.

Works	Link Analysis		Collaborative Filtering (CF)	Features			
	HITS			Frequency	Variety	Repetition	Distinctiveness
	POI	Expert					
TBHG [20]	✓	✓		✓			
LocPref [5]		✓	✓	✓			
ImpHITS [14]	✓			✓	✓		
CLoRW [21]	✓	✓		✓			
TAP-F [15]	✓	✓	✓	✓			
NILR [9]	✓			✓		✓	
WHITS [22]	✓			✓			
TDCE [13]	✓		✓	✓			
PPTA-RM [23]	✓		✓	✓			
FCDST [27]			✓	✓			

Table 1. Cont.

Works	Link Analysis		Collaborative Filtering (CF)	Features			
	HITS			Frequency	Variety	Repetition	Distinctiveness
	POI	Expert					
SPR [28]			✓	✓			
timePOI [29]			✓	✓			
CTF-ARA [30]			✓	✓			
U-CF-M [12]			✓	✓			
CLGRW [31]			✓	✓			
U-CF-MS [32]			✓	✓		✓	
CULT-TF [33]			✓	✓			
LACF-Rec3	✓	✓	✓	✓	✓	✓	✓

3. Our Method

In this section, the basic notations and details of our proposed location-based recommendation system are described. As illustrated in Figure 1, our LACF-Rec3 method consists of two main phases: the offline and online phases. The first step within the offline phase involves the determination of frequency, variety, and repetition statistics related to the locations and users. Subsequently, interest scores are calculated for the locations and users using our extended HITS-based model (termed HITS-3), leveraging the statistics obtained from the preceding step. In the third step, distinctiveness and distinctive scores are identified for the users and locations.

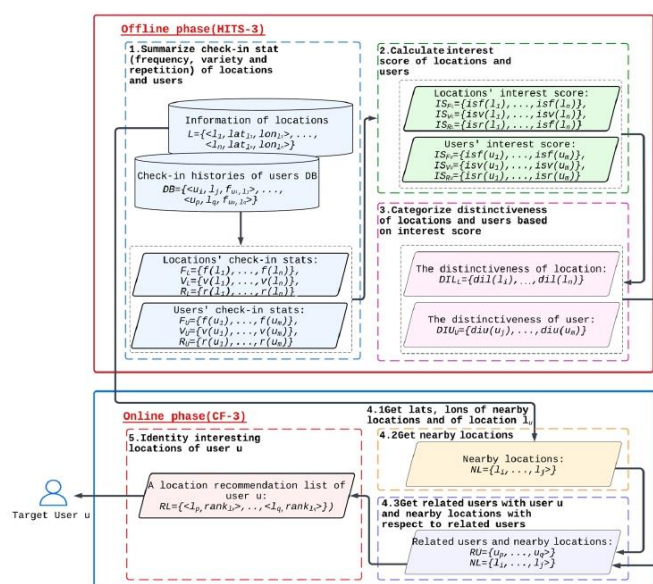


Figure 1. Our LACF-Rec3 method consists of two phases: offline phase with HITS-3, and online phase with CF-3.

In the online phase, upon the arrival of a target user's request for a POI recommendation list, our approach retrieves, from the database, locations proximate to the target user, u , as well as the set of users who have visited these locations at least once, RU . Next, our enhanced collaborative filtering technique (CF-3) determines the distinctiveness similarity between u and each user within RU . Finally, the locations with the highest recommendation scores are selected and ranked, constituting the top- N location list recommended for the target user.

3.1. Notation

Let $L_DB = \{ \langle l_1, lat_{l_1}, lon_{l_1} \rangle, \dots, \langle l_n, lat_{l_n}, lon_{l_n} \rangle \}$ be a set of all n locations where each location l_j is associated with its latitude (lat_{l_j}) and longitude (lon_{l_j}). A set $U = \{ u_1, u_2, \dots, u_m \}$ is a set of all users (i.e., m users) who checks in at least one location (l_j) in L_DB . A record of check-in history of a user u_i at a location l_j is represented by a 3-tuple $\langle u_i, l_j, f_{u_i, l_j} \rangle$ where f_{u_i, l_j} is the visit frequency of u_i at l_j (i.e., the number of times that u_i checks in at l_j). Let $DB = \{ \langle u_1, l_1, f_{u_1, l_1} \rangle, \dots, \langle u_m, l_n, f_{u_m, l_n} \rangle \}$ be the set of check-in histories of all users, $DB_{l_p} = \{ \langle u_1, l_p, f_{u_1, l_p} \rangle, \dots, \langle u_m, l_p, f_{u_m, l_p} \rangle \}$ be the set of the check-in histories of all users visiting the location $\langle l_p, lat_{l_p}, lon_{l_p} \rangle \in L_DB$, and $DB_{u_j} = \{ \langle u_j, l_1, f_{u_j, l_1} \rangle, \dots, \langle u_j, l_n, f_{u_j, l_n} \rangle \}$ be the set of check-in histories of a user u_j visiting all n locations in L_DB . The current location of a target user u is denoted by $\langle lat_u, lon_u \rangle$.

To provide a list of interesting locations to a target user u located at $\langle lat_u, lon_u \rangle$, the prior collected data of DB, DB_{l_p}, DB_{u_j} , and $\langle lat_u, lon_u \rangle$ are considered as input. Then, our LACF-Rec3 considers three visiting behaviors (i.e., the frequency, variety, and repetitions) of all users to identify interesting locations. All related notations can be described as follows.

Definition 1 (Frequency of locations and of users). The checked-in frequency of l_p is the total number of times that all m users in U check in at l_p , computed as follows.

$$f(l_p) = \sum_{k=1}^m f_{u_k, l_p} \quad (1)$$

Similarly, the check-in frequency of u_j is the total number of locations that the user u_j visited, computed as follows.

$$f(u_j) = \sum_{k=1}^n f_{u_j, l_k} \quad (2)$$

Definition 2 (Variety of locations and of users). The user variety of l_p is the number of individuals (u_k) who check in at least once at l_p retrieved from the database DB_{l_p} , defined as follows.

$$v(l_p) = \sum_{k=1}^m va_k \text{ where } va_k = \begin{cases} 1, & \text{if } f_{u_k, l_p} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Similarly, based on the database DB_{u_j} , the location variety of u_j is the number of locations u_j visited, defined as follows.

$$v(u_j) = \sum_{k=1}^n va_k \text{ where } va_k = \begin{cases} 1, & \text{if } f_{u_j, l_k} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

Definition 3 (Repetitions of locations and of users). Based on the database DB_{l_p} , the checked-in repetition of l_p is the number of users who have visited l_p more than once, defined as follows.

$$r(l_p) = \sum_{k=1}^m re_k \text{ where } re_k = \begin{cases} 1, & \text{if } f_{u_k, l_p} > 1 \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

Similarly, based on the database DB_{u_j} , the number of locations u_j revisited is the number of locations u_j visited more than once, defined as follows.

$$r(u_j) = \sum_{k=1}^n re_k \text{ where } re_k = \begin{cases} 1, & \text{if } f_{u_j, k} > 1 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

Definition 4 (Normalized characteristics). Let $F_L = \{f(l_1), f(l_2), \dots, f(l_n)\}$ be the set of checked-in frequency of locations in L_DB . To prevent small values being ignored, and avoid the dominance of one characteristic over the other two, each checked-in frequency $f(l_p) \in F_L$ is normalized as follows

$$nf(l_p) = \begin{cases} 0.5 + \left((f(l_p) - avg) \times \frac{0.5}{max - avg} \right), & f(l_p) \geq avg \\ f(l_p) \times \frac{0.5}{avg}, & \text{otherwise} \end{cases} \quad (7)$$

where avg is the average value of checked-in frequencies in F_L (i.e., $avg(f(l_1), f(l_2), \dots, f(l_n))$), and max is the maximum checked-in frequency in F_L (i.e., $max(f(l_1), f(l_2), \dots, f(l_n))$).

Similarly, Equation (7) is applied to normalize each check-in frequency of a user $f(u_j) \in F_U = \{f(u_1), f(u_2), \dots, f(u_m)\}$, each user variety of a location $v(l_p) \in V_L = \{v(l_1), v(l_2), \dots, v(l_n)\}$, each location variety of a user $v(u_j) \in V_U = \{v(u_1), v(u_2), \dots, v(u_m)\}$, each checked-in repetition of a location $r(l_p) \in R_L = \{r(l_1), r(l_2), \dots, r(l_n)\}$, and the number of locations u_j revisited $r(u_j) \in R_U = \{r(u_1), r(u_2), \dots, r(u_m)\}$.

Definition 5 (Interest score of locations and of users). Let $ISF_L = \{isf(l_1), isf(l_2), \dots, isf(l_n)\}$ and $ISF_U = \{isf(u_1), isf(u_2), \dots, isf(u_m)\}$ be the set of frequency-based interest scores of all locations and the set of all users based on their frequencies, respectively. All frequency-based interest scores in ISF_U are initialized to 1. The frequency-based interest score $isf(l_p) \in ISF_L$ of $l_p \in L_DB$ is the summation of all of the users' scores (which is the multiplication of each user's check-in frequency and prior-interest score) who have checked in at the location l_p . The frequency-based interest score $isf(u_q) \in ISF_U$ of $u_q \in U$ is subsequently computed by summarizing the score of locations that u_q has checked in. Both values can be defined as follows.

$$isf(l_p) = \sum_{q=1}^m \begin{cases} f(u_q) \times isf(u_q), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$isf(u_q) = \sum_{p=1}^n \begin{cases} f(l_p) \times isf(l_p), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Similarly, the variety-based and repetition-based interest score of all locations and all users can be defined as $ISV_L = \{isv(l_1), isv(l_2), \dots, isv(l_n)\}$, $ISV_U = \{isv(u_1), isv(u_2), \dots, isv(u_m)\}$, $ISR_L = \{isr(l_1), isr(l_2), \dots, isr(l_n)\}$ and $ISR_U = \{isr(u_1), isr(u_2), \dots, isr(u_m)\}$, respectively, where each interest score can be computed as follows.

$$isv(l_p) = \sum_{q=1}^m \begin{cases} v(u_q) \times isv(u_q), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

$$isv(u_q) = \sum_{p=1}^n \begin{cases} v(l_p) \times isv(l_p), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$isr(l_p) = \sum_{q=1}^m \begin{cases} r(u_q) \times isr(u_q), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

$$isr(u_q) = \sum_{p=1}^n \begin{cases} r(l_p) \times isr(l_p), & \text{if user } u_q \text{ has checked in at location } l_p \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

Finally, L-2 normalization is applied to $isf(l_p)$, $isf(u_q)$, $isv(l_p)$, $isv(u_q)$, $isr(l_p)$, and $isr(u_q)$.

Definition 6 (Distinctiveness of a location and of a user). To point out the distinctiveness of each location $l_p \in L_DB$, we identify the most outstanding aspect of l_p . The distinctiveness of l_p is 1 if the frequency-based interest score is greater than the other two. The distinctiveness of l_p is 2 and 3, if the variety-based and repetition-based interest score is greatest, respectively, illustrated as follows.

$$dil(l_p) = \begin{cases} 1, & \text{if } isf(l_p) = \max(isf(l_p), isv(l_p), isr(l_p)) \\ 2, & \text{if } isv(l_p) = \max(isf(l_p), isv(l_p), isr(l_p)) \\ 3, & \text{if } isr(l_p) = \max(isf(l_p), isv(l_p), isr(l_p)) \end{cases} \quad (14)$$

Similarly, we identify the most outstanding aspect of u_q . The distinctiveness of u_q is determined as follows.

$$diu(u_q) = \begin{cases} 1, & \text{if } isf(u_q) = \max(isf(u_q), isv(u_q), isr(u_q)) \\ 2, & \text{if } isv(u_q) = \max(isf(u_q), isv(u_q), isr(u_q)) \\ 3, & \text{if } isr(u_q) = \max(isf(u_q), isv(u_q), isr(u_q)) \end{cases} \quad (15)$$

Definition 7 (Distinctiveness score of a user). Let $L_{u_q} = \{ \langle l_p, f_{u_q, l_p}, dil(l_p) \rangle, \dots, \langle l_r, f_{u_q, l_r}, dil(l_r) \rangle \}$ be a set of 3-tuples, containing the name of locations checked-in by $u_q \in U$, check-in frequency, and the distinctiveness of the location. Also, let $L_{u_q}^1$, $L_{u_q}^2$, and $L_{u_q}^3$ be the set of 3-tuples in L_{u_q} , of which distinctiveness of the location is 1, 2, and 3, respectively. The distinctiveness score of u_q , $ds(u_q) = \langle dsf(u_q), dsv(u_q), dsr(u_q) \rangle$, indicates the distinctiveness of u_q on the aspects of check-in frequency, location variety, and the number of locations u_q revisited. ds_{u_q} is updated as follows.

$$dsf(u_q) = \sum_{i=1}^{|L_{u_q}^1|} f_{u_q, l_i} \quad (16)$$

$$dsv(u_q) = \sum_{j=1}^{|L_{u_q}^2|} f_{u_q, l_j} \quad (17)$$

$$dsr(u_q) = \sum_{k=1}^{|L_{u_q}^3|} f_{u_q, l_k} \quad (18)$$

Definition 8 (Similarity between users). Let u_q be the user who checked in at any location l_p nearby the current location of the target user u . Thus, to identify similarity in check-in behavior between u and u_q , the cosine similarity $sim(u, u_q)$ is determined by comparing their distinctiveness scores as follows.

$$sim(u, u_q) = \frac{(dsf(u) \times dsf(u_q)) + (dsv(u) \times dsv(u_q)) + (dsr(u) \times dsr(u_q))}{\sqrt{dsf(u)^2 + dsv(u)^2 + dsr(u)^2} \times \sqrt{dsf(u_q)^2 + dsv(u_q)^2 + dsr(u_q)^2}} \quad (19)$$

Definition 9 (A recommendation score of a location based on relevant users). Let RU be a set of relevant users who have checked in at least once at locations nearby the current location of the target user u . Let SRU be a set of relevant users, where u_j is the element of SRU if and only if $sim(u, u_j)$ is more than a threshold. A recommendation score of a nearby location l_p is determined by considering the visiting information associated with the relevant users including (i) the similarity between u and each relevant user u_j , $sim(u, u_j)$, (ii) the check-in frequency of u_j at the location l_p ,

f_{u_i, l_p} , and (iii) the distinctiveness score of u_j according to $dil(l_p)$. The recommendation score, $rs(l_p)$, of the location l_p is computed as follows.

$$rs(l_p) = \sum_{j=1}^{|SRU|} (sim(u, u_j) \times f_{u_i, l_p} \times dsx(u_j)), \quad (20)$$

$$\text{where } dsx(u_j) = \begin{cases} dsf(u_j), & \text{if } dil(l_p) = 1 \\ dsv(u_j), & \text{if } dil(l_p) = 2 \\ dsr(u_j), & \text{if } dil(l_p) = 3 \end{cases}$$

Definition 10 (A recommendation score of a location by location's interest scores). Let $NL = \{l_p, \dots, l_q\}$ be a set of locations nearby the current location of the target user u . The recommendation score for the location $l_p \in NL$ is determined as the maximum interest score among the frequency-based, variety-based, and repetition-based interest scores as follows.

$$rs(l_p) = \max(isf(l_p), isv(l_p), isr(l_p)) \quad (21)$$

According to the definitions given above, we divided the system into two phases: (i) the offline phase utilizes Definitions 1–6 to determine the distinctiveness of each user and location through our HITS-3 algorithm, and (ii) the online phase applies Definitions 7–10 to create a list of recommended POIs for the target users by employing our CF-3 algorithm.

3.2. Offline Phase

Given the check-in database DB of all users' visits to various locations as input, the offline phase aims to ascertain the distinctiveness of the users and locations in terms of the frequency, variety, and repetition of check-ins. Generally, locations with a high check-in frequency correspond to popular or trendy places, of which the number of visits or revisits is high. Locations with a high variety are typically tourist attractions and landmarks that most users or tourists visit only once. Locations with numerous repeated check-ins could be associated with places users frequently visited in daily life (such as grocery stores or supermarkets) or specific venues for particular users (such as gyms, basketball courts, board game cafes, etc.). Users with a high check-in frequency are likely to be active and sociable; those with diverse check-in locations tend to seek novel experiences and avoid monotony. Finally, users with high repeated check-ins might exhibit specific preferences or strong loyalty.

In the first and second lines of Algorithm 1, the check-in statistical characteristics for each location l_p and for each user u_j (i.e., frequency ($f(l_p)$ and $f(u_j)$), variety ($v(l_p)$ and $v(u_j)$), and repeatedly ($r(l_p)$ and $r(u_j)$)) are first calculated by Definitions 1–3 and then aggregated into the sets LCS and UCS (short for *Locations' Check-in Statistics* and *Users' Check-in Statistics*). To determine the distinctiveness of the locations and users, it is crucial to avoid the dominance of one characteristic over the others. Frequency values are usually higher than variety values, which are mostly greater than repeated check-in values. The third and fourth line of Algorithm 1 depict the normalization of each check-in statistic for each location and each user (by Definition 4), ensuring the absence of dominance. Specifically, we extended the HITS model (terms as *HITS-based on 3 check-in behaviors*, *HITS-3*) to take all three statistical characteristics (as opposed to the traditional frequency-based HITS model). This extension enables the computation of interest scores for the locations and users. As described in Definition 5, the interest score of each location l_p can be expressed as (i) the frequency-based interest score, $isf(l_p)$, (ii) variety-based interest score, $isv(l_p)$, and (iii) interest score based on repeated check-ins, $isr(l_p)$, which forms a 3-tuple $\langle isf(l_p), isv(l_p), isr(l_p) \rangle$. Similarly, each user u_i also has three interest scores in a 3-tuple form, $\langle isf(u_i), isv(u_i), isr(u_i) \rangle$. With the HITS-based model, these interest scores for the locations and users are iteratively computed. As the iteration progresses, the value of each interest score increases. As in the fifth to eighth line, we applied L2-normalization

to all of the interest scores to counteract exponential growth and allow for quicker score convergence. In the last step, in the sixth and seventh line of Algorithm 1, the highest interest score of each location and each user is determined and utilized to identify the distinctiveness of both the locations and the users. If the frequency-based interest score for the location l_p is the highest, it stands out for the check-in frequencies (i.e., $dil(l_p)$ is "1"). Similarly, $dil(l_p)$ is assigned "2" if the location's variety-based interest score is the highest, and "3" if its repetition-based score is the highest. The determination of $dii(u_q)$ follows a similar procedure. The complete algorithm for the offline phase is outlined in Algorithm 1.

Algorithm 1. Offline phase Algo

Input:

A check-in database $DB = \{ \langle u_i, l_j, f_{u_i, l_j} \rangle, \dots, \langle u_p, l_q, f_{u_p, l_q} \rangle \}$

A location database $L_DB = \{ \langle l_1, lat_{l_1}, lon_{l_1} \rangle, \dots, \langle l_n, lat_{l_n}, lon_{l_n} \rangle \}$

Output:

Interest scores of locations and users $IS_{F_L}, IS_{F_U}, IS_{V_L}, IS_{V_U}, IS_{R_L}, IS_{R_U}$

A distinctiveness of each location $dil(l_p)$ and user $dii(u_q)$

1: compute location's check-in stats $LCS = \{ \langle f(l_1), v(l_1), r(l_1) \rangle, \dots, \langle f(l_n), v(l_n), r(l_n) \rangle \}$ (by Definitions 1–3)

2: compute user's check-in stats $UCS = \{ \langle f(u_1), v(u_1), r(u_1) \rangle, \dots, \langle f(u_m), v(u_m), r(u_m) \rangle \}$ (by Definitions 1–3)

3: normalize LCS (by Definition 4)

4: normalize UCS (by Definition 4)

5: **for** each t time **do**

6: compute interest score of locations and interest score of users, IS_{F_L} ,

7: $IS_{F_U}, IS_{V_L}, IS_{V_U}, IS_{R_L}, IS_{R_U}$ (by Definition 5)

8: **end for**

9: identify distinctiveness $dil(l_p)$ of each location in L_DB (by Definition 6)

10: identify distinctiveness $dii(u_q)$ of each user in U (by Definition 6)

As shown in Figure 2, the check-in history from a prior collected check-in database contains three users (u_1 , u_2 and u_3) and three locations (l_1 , l_2 and l_3). User u_1 checks in at l_1 and l_3 four and two times, respectively. User u_2 checks in at l_1 , l_2 , and l_3 one, five, and four times, respectively. User u_3 checks in once at l_1 . In step 1, the visiting frequencies (F_L) of l_1 , l_2 , and l_3 are calculated as $4 + 1 + 1 = 6$, $0 + 5 + 0 = 5$, and $4 + 2 + 0 = 6$. The user varieties (V_L) of l_1 , l_2 , and l_3 are then computed as $1 + 1 + 1 = 3$, $0 + 1 + 0 = 1$, and $1 + 1 + 0 = 2$. Next, the number of repeated check-ins (R_L) of l_1 , l_2 , and l_3 , or the number of users that visit the locations more than once, are computed as $1 + 0 + 0 = 1$, $0 + 1 + 0 = 1$, and $1 + 1 + 0 = 2$, respectively. F_L , V_L , and R_L are stored in LCS . Similarly, the check-in frequencies (F_U) of u_1 , u_2 , and u_3 are computed as $4 + 0 + 2 = 6$, $1 + 5 + 4 = 10$, and $1 + 0 + 0 = 1$. The location varieties (V_U) of u_1 , u_2 , and u_3 are then computed as $1 + 0 + 1 = 2$, $1 + 1 + 1 = 3$, and $1 + 0 + 0 = 1$. Next, the number of revisits (R_U) of u_1 , u_2 and u_3 are counted as $1 + 0 + 1 = 2$, $0 + 1 + 1 = 2$, and $0 + 0 + 0 = 0$, respectively. F_U , V_U , and R_U are stored in UCS as shown in the first and second lines of Algorithm 1. In the second step, the visiting frequencies l_1 , l_2 , and l_3 are normalized as $0.5 + ((6 - 5.67) \times \frac{0.5}{6 - 5.67}) = 1$, $5 \times \frac{0.5}{5.67} = 0.4$, and $0.5 + ((6 - 5.67) \times \frac{0.5}{6 - 5.67}) = 1$, where avg is 5.67 and max is 6. Similarly, the user varieties of l_1 , l_2 , and l_3 are $0.5 + ((3 - 2) \times \frac{0.5}{3 - 2}) = 1$, $1 \times \frac{0.5}{2} = 0.5$, and $0.5 + ((2 - 2) \times \frac{0.5}{3 - 2}) = 0.5$, where avg is 2 and max is 3. Next, the numbers of repeated check-ins of l_1 , l_2 , and l_3 are computed as $1 \times \frac{0.5}{1.3} = 0.38$, $1 \times \frac{0.5}{1.3} = 0.38$, and $0.5 + ((2 - 1.3) \times \frac{0.5}{2 - 1.3}) = 1$, where avg is 1.3 and max is 2, respectively. The check-in frequencies of u_1 , u_2 , and u_3 are normalized as $0.5 + ((6 - 5.67) \times \frac{0.5}{10 - 5.67}) = 0.54$, $0.5 + ((10 - 5.67) \times \frac{0.5}{10 - 5.67}) = 1$, and $1 \times \frac{0.5}{5.67} = 0.09$, where avg is 5.67 and max is 10. The location varieties of u_1 , u_2 , and u_3 are normalized as $0.5 + ((2 - 2) \times \frac{0.5}{3 - 2}) = 0.5$, $0.5 + ((3 - 2) \times \frac{0.5}{3 - 2}) = 1$, and $1 \times \frac{0.5}{2} = 0.25$, where avg is 2 and max is 3. Finally, the number of revisits of u_1 , u_2 , and u_3 are normalized as $0.5 + ((2 - 2) \times \frac{0.5}{2 - 2}) = 1$, $0.5 + ((2 - 2) \times \frac{0.5}{2 - 2}) = 1$, and 0, where avg is 2 and max is 2,

respectively, as shown in the third and fourth lines of Algorithm 1. Next, in the third step, the interest score of the locations and users are computed using our HITS-3 model, as described in the fifth to eighth lines of Algorithm 1. Three interest scores of each location are specified in a 3-tuple form (i.e., $\langle isf(l_p), isv(l_p), isr(l_p) \rangle$). The interest scores of l_1 , l_2 , and l_3 are computed as $\langle 0.3, 0.5, 0.2 \rangle$, $\langle 0.4, 0.3, 0.1 \rangle$, and $\langle 0.2, 0.5, 0.7 \rangle$, respectively. The distinctiveness of each location ($dil(l_p)$) is determined according to the ninth line of Algorithm 1. Therefore, the distinctiveness of l_1 , l_2 , and l_3 are in the variety, frequency, and repetition aspects, respectively. Similarly, three interest scores of each user are also specified in a 3-tuple form (i.e., $\langle isf(u_i), isv(u_i), isr(u_i) \rangle$). The interest scores of u_1 , u_2 , and u_3 are $\langle 0.3, 0.2, 0.4 \rangle$, $\langle 0.6, 0.5, 0.4 \rangle$, and $\langle 0.2, 0.1, 0.1 \rangle$, respectively. The distinctiveness of each user ($dii(u_q)$) is determined according to the tenth line of Algorithm 1. The distinctiveness of u_1 , u_2 , and u_3 are in the repetition, frequency, and frequency aspects, respectively.

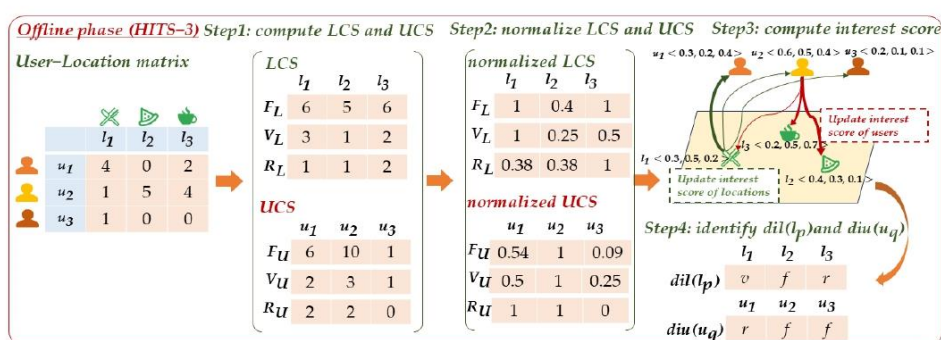


Figure 2. Example of the offline phase in the LACF-Rec3.

3.3. Online Phase

To recommend locations to a target user u , the user's current location, $\langle lat_u, lon_u \rangle$, is required along with the request. Subsequently, nearby locations within the distance of dt from u (the maximum distance specified by u), $NL = \{l_p, \dots, l_q\}$, are initially identified (as shown in the first line of Algorithm 2). The users who have checked in at least once at a nearby location $l_p \in NL$ (also referred to as relevant users), $RU = \{u_p, \dots, u_j\}$, are then retrieved (as shown in the second line of Algorithm 2). Next, the three distinctiveness scores of each user $u_q \in RU$, $ds(u_q)$, are computed on the aspects of check-in frequency, location variety, and the number of locations the user revisits, respectively (in accordance with Definition 7). The frequency-based distinctiveness, $dsf(u_q)$, of user u_q is determined by summing the visit frequency of u_q at l_j , which stands out in the aspect of checked-in frequency (i.e., $dil(l_j) = 1$). Similarly, the variety-based distinctiveness score, $dsv(u_q)$, of user u_q is calculated by summing the visit frequency of u_q at l_j , which stands out in the aspect of location variety (i.e., $dil(l_j) = 2$). Finally, the repetition-based distinctiveness score, $dsr(u_q)$, of user u_q is derived by summing the visit frequency of u_q at l_j , which stands out in the aspect of the number of revisits (i.e., $dil(l_j) = 3$).

Subsequently, the similarity, $sim(u, u_q)$, between the distinctiveness of the target user u and each user $u_q \in RU$ is calculated according to Definition 8. A user u_q is characterized as an "irrelevant user with respect to the target user u " and is not selected from RU to be in SRU if its similarity to u falls below the predefined similarity threshold δ (as stated in the third line of Algorithm 2). Note that the target user u will be considered as a cold-start user, if they lack a check-in history or if there is no "relevant user with respect to the target user u " (i.e., $SRU = \emptyset$).

Algorithm 2. Online pvhase Algo**Input:**

A check-in database $DB = \{ \langle u_i, l_j, f_{u_i, l_j} \rangle, \dots, \langle u_p, l_q, f_{u_p, l_q} \rangle \}$
 A location database $L_DB = \{ \langle l_1, lat_{l_1}, lon_{l_1} \rangle, \dots, \langle l_n, lat_{l_n}, lon_{l_n} \rangle \}$
 Interest scores of locations and users $IS_{F_u}, IS_{F_l}, IS_{V_u}, IS_{V_l}, IS_{R_u}, IS_{R_l}$
 A distinctiveness of each location $dil(l_p)$ and user $dii(u_q)$
 A number of recommendations N
 A target user u with her current location $\langle lat_u, lon_u \rangle$
 A distance condition on recommended location, d
 A similarity threshold, δ

Output:

A recommended list $RL = \{ \langle l_1, rs(l_1) \rangle, \langle l_2, rs(l_2) \rangle, \dots, \langle l_n, rs(l_n) \rangle \}$
 1: identify $NL = \{ l_p, \dots, l_q \}$, the set of locations nearby the current location $\langle lat_u, lon_u \rangle$ of u (within distance d)
 2: identify $RU = \{ u_i, \dots, u_j \}$, the set of relevant users visiting the location in NL at least once
 3: determine the distinctiveness score of each user, u_i , in RU and the similarity of u_i and u . Then, identify SRU by selecting the users from RU to be in SRU based on their distinctiveness with the similarity threshold δ (by Definitions 7 and 8)
 4: compute $RS_{NL} = \{ rs(l_p), \dots, rs(l_q) \}$ the set of recommendation scores of all locations in NL (if $RU \neq \emptyset$ applies Definition 9, otherwise, applies Definition 10)
 5: determine $RL = \{ \langle l_p, rank_{l_p} \rangle, \dots, \langle l_q, rank_{l_q} \rangle \}$, the set of N nearby locations with the associated ranks
 6: recommend RL to the target user u

Stated in the fourth line of Algorithm 2, the recommended locations for the target user, u , can be determined in two cases:

(Case 1) In the case that u is associated with relevant users ($SRU \neq \emptyset$), we extended collaborative filtering (named CF-3) to address the challenge of sparsity. Instead of the similarity in visited locations, CF-3 determines the resemblance between u and each $u_q \in SRU$ by considering the similarity of the distinctiveness of location l ($dil(l)$). CF-3 takes into account the distinctive scores of relevant users u_q ($dsx(u_q)$) in order to recommend intriguing locations. This approach is based on the assumption that u might share similar preferences and tends to visit at the same locations as their relevant users. For each checked-in location l_j , its recommendation score is calculated by considering the similarity between u and each user u_q (who check-in at l_j), the check-in frequency of user u_q at location l_j , and the distinctiveness score of u_q on the aspect of the distinctiveness $dil(l_j)$ of location l_j (by defined in Definition 9).

(Case 2) In the case where u has no relevant users (indicating a cold-start scenario), our method relies on the interest score associated with each location $l_j \in L_DB$ (i.e., $\langle isf(l_j), isv(l_j), isr(l_j) \rangle$ (provided by Algorithm 1)). Subsequently, the recommendation score of location l_j is determined by selecting the highest score among the three interest scores (as specified by Definition 10).

After considering all locations (whether in Case 1 or Case 2), only top- N locations with the highest recommendation scores are selected to be included in the recommendation list for the target user u , as outlined in the fifth line of Algorithm 2.

As depicted in Figure 3, when a target user u requests a list of recommended POIs, the system first retrieves the set of locations nearby her current location (NL) (i.e., l_1 , l_2 , and l_3), and the set of relevant users (RU) (i.e., u_1 , u_2 , and u_3). This process is described in the first and second lines of Algorithm 2. In the second step, the similarities of u and each user in RU are calculated based on the distinctiveness score. The similarity value of u and u_1 ($sim(u, u_1)$) is $\frac{(0 \times 0)(2 \times 4)(0 \times 2)}{\sqrt{0^2 + 2^2 + 0^2} \times \sqrt{0^2 + 4^2 + 2^2}} = 0.89$, $sim(u, u_2)$ is $\frac{(0 \times 5)(2 \times 1)(0 \times 4)}{\sqrt{0^2 + 2^2 + 0^2} \times \sqrt{5^2 + 1^2 + 4^2}} = 0.15$, and $sim(u, u_3)$ is $\frac{(0 \times 0)(2 \times 1)(0 \times 2)}{\sqrt{0^2 + 2^2 + 0^2} \times \sqrt{0^2 + 1^2 + 2^2}} = 0.44$. In the third step, if the similarity threshold δ is set to 0.4, u_2 is not selected to be in SRU , as outlined in the third line of Algorithm 2. Next, in the fourth step, the recommendation scores (RS_{NL}) of NL are computed. As a

result, the recommendation score of l_2 and l_3 are $(0.89 \times 5 \times 0.6) + (0.44 \times 0 \times 0.2) = 2.67$, and $(0.89 \times 2 \times 0.6) + (0.44 \times 0 \times 0.2) = 1.07$, respectively. In the final step, the set of N nearby locations (RL) is ranked, with l_3 holding the top rank and l_2 taking the second to be recommended to the target user u .

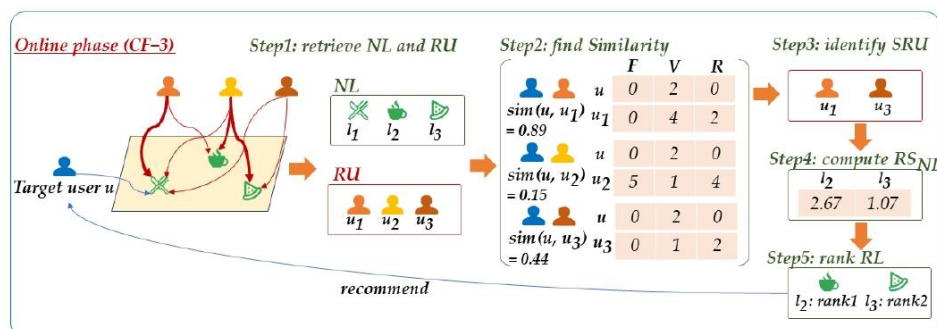


Figure 3. Example of the online phase in LACF-Rec3.

4. Experiments

In this section, we describe the experimental setup and evaluation methodologies. The details of the dataset used for these experiments are explained. We conducted a comparative analysis between our method and four baseline approaches including user-based collaborative filtering for POI recommendation (UCF), personalized POI recommendation using memory-based preferences and POI stickiness (U-CF-MS), hypertext-induced topic-search-based recommendation (HITS), and tensor decomposition-based collaborative filtering (TDCF). The evaluation method using the four metrics (i.e., precision, recall, NDCG, and matching ratio) is also explained. Finally, we report the experimental results and discuss the effectiveness of our proposed method.

4.1. Experiment Setup and Evaluation Methodologies

4.1.1. Dataset

To evaluate the performance of our proposed method, we employed a real-world dataset obtained from Foursquare, containing historical check-in records in Tokyo spanning from April 2012 to February 2013 (<https://sites.google.com/site/yangdingqi/home/foursquare-dataset> accessed on 1 August 2022) [34,35]. Within the dataset, a range of check-in activities were recorded across diverse categories including food establishments (e.g., restaurants, cafes, bars), transportation facilities (e.g., bus, train, and subway stations, airports), nightlife venues (e.g., nightclubs), and more. However, it was observed that a significant number of check-ins occurred at transportation sites, primarily associated with daily routines such as commuting to work or school, rather than places visited in special occasions or by users looking for new experiences. Based on this observation, we opted to utilize only check-in data associated with food-related locations. This choice allowed us to focus on the recommendation of novel experiences to users.

Each record of the check-in data pertains to a user who has checked in at a location and contains six fields: user identifier, location name, location category, latitude of location, longitude of location, and check-in timestamp. Our method is based on applying the users' expertise in certain geographical zones to create tailored recommendation lists for the target users. We thus categorized the locations into five areas (i.e., Chiyoda, Shibuya, Minato, Shinjuku, and Kawasaki) by using Google Maps Geocoding API [36]. These areas were subsequently incorporated into the users' check-in database.

Table 2 presents the information regarding the check-in activities of 2169 users at 11,071 food-related venues in Tokyo. Some users may have check-in activities at various cities. The table also shows the analysis of the check-in histories associated with each area. *Chiyoda* was the area with the highest values in terms of the check-in users, checked-in locations, total check-ins, average check-ins per location, and average check-ins per user. Those of *Shibuya*, *Minato*, and *Shinjuku* were closely aligned, and relatively smaller than those of *Chiyoda*. Finally, *Kawasaki* had the lowest values across these aspects.

Table 2. The characteristics of the dataset.

Location	Users (Existing Users/ Cold-Start Users)	Locations	Total Check-Ins	Check-In per Locations	Check-In per Users
Chiyoda	1702 (211/27)	2826	15,029	5.32	8.83
Shinjuku	1352(136/36)	2348	9051	3.85	6.69
Minato	1264 (115/27)	2556	8695	3.40	6.89
Shibuya	1240 (127/27)	2222	8413	3.79	6.78
Kawasaki	537 (52/9)	1119	4259	3.81	7.93
Total	2169 (356/40)	11,071	45,447	4.11	20.95

In our experiments, we employed a 5-fold cross-validation technique on the dataset, wherein 20% of the check-in histories were randomly selected as the testing dataset while the remaining 80% served as the training dataset.

4.1.2. Baseline Methods

As our method is a hybrid of link analysis and collaborative filtering techniques, we chose to compare it with the UCF [29] and U-CF-MS [32] methods because they are memory-based collaborative filtering techniques. UCF is a traditional and widely-employed collaborative filtering technique. U-CF-MS is a recently proposed user-based collaborative filtering approach geared toward individual POI recommendations, and it utilizes the number of repetitions, aligning with our proposed methodology.

Additionally, two more methods, HITS and TDCF, were also included in our comparative analysis. The HITS-based model [18] stands as the traditional and renowned algorithm developed for the task of discovering quality POI, based on a link analysis technique. Finally, TDCF [13] recommends the list of POIs using the most recently proposed hybrid method incorporating collaborative filtering and link analysis methods like our hybrid strategy. The overview of each method is described below:

User-based collaborative filtering for POI recommendation (UCF) [29]: This user-based collaborative filtering approach initially creates a user–location matrix, where each entry denotes the probability of a particular user visiting a particular POI. Recommendations are then generated by considering the similarity between the users' check-in histories.

Personalized POI recommendation using memory-based preferences and POI stickiness (U-CF-MS) [32]: This user-based collaborative filtering approach considers the recent check-in POIs of the users (memory-based) and the revisit behavior (repeated check-ins) of each POI (POI stickiness) to generate recommendations.

Hypertext-induced-topic-search-based recommendation (HITS) [18]: This link analysis-based approach initially constructs a user–POI network to determine the interest scores of POIs and subsequently recommend the top-N most interesting POIs within a geospatial distance.

Tensor decomposition based collaborative filtering (TDCF) [13]: This approach combines link analysis and collaborative filtering approaches considering the user check-in histories including user-id (username), location categories that the user checked in, and check-in time (called time slot). Additionally, its link-analysis component leverages the location popularity and distances between locations.

Our hybrid method (LACF-Rec3), which combines link analysis and collaborative filtering, considers three visiting behaviors to recommend locations for both existing and cold-start users. For existing users, we introduced a novel collaborative filtering technique (CF-3) that utilizes distinctiveness similarity to recommend POIs personalized to the target user, although they might not be popular among tourists. For cold-start users, an extended version of the HITS-based model (HITS-3) was proposed. This enhanced model incorporates all three visiting behaviors—frequency, variety, and repetition—to generate the top-ranked POIs. These recommended venues are not only characterized by a large number of visits but also by a broad range of individual visitors and a significant number of revisits.

4.1.3. Evaluation Methods

The performance of POI recommendation is examined through the assessment of three aspects: (i) accuracy, (ii) ranking recommendation, and (iii) matching of the recommended POIs with target users retrieved from the testing dataset.

Since the dataset does not include the current location of the target user, we employed the minimum bounding rectangle (MBR) technique [5,15] to create such information. The technique involves identifying a boundary of locations in the testing dataset, shown as the dash line (derived from location A and B) in Figure 4. The target user's current location is denoted by the green circle. All POIs within the boundary—whether depicted in red or black—can be considered as candidate POIs for recommendation purposes. The red POIs represent places where the target user from the testing dataset has previously checked in, referred to as ground truth POIs. The black POIs represent ordinary POIs located within the area but not visited by the target user. The red POIs with the green outline (termed recommended ground truth POIs) are the venues recommended to the target user, who has already visited them at least one. Similarly, the black ones outlined in green are the places recommended to the target user, although they have never visited the places. For instance, in Figure 4, there are a total of seven ground truth POIs. Assuming that the target user u is positioned within the green circle and requests a POI recommendation list, a recommender system recommends a list of the top five recommended POIs, consisting of three ground truth POIs and two ordinary POIs: three red and two black POIs with the green outline.



Figure 4. Recommendation effectiveness evaluation method using MBR.

Two common criteria (i.e., the precision and recall) (as in Equations (22) and (23)), [27,32] are adopted to evaluate the accuracy of recommendation. The precision is the ratio of

those recommended on the ground truth POIs to the total number of recommended items (denoted as N). The recall is the ratio of those recommended on the ground truth POIs to the total number of ground truth POIs.

$$\text{Precision} = \frac{\text{number of Recommended Ground Truths}}{\text{number of Recommended POIs}} \quad (22)$$

$$\text{Recall} = \frac{\text{number of Recommended Ground Truths}}{\text{number of Ground Truths}} \quad (23)$$

NDCG (normalized discounted cumulative gain) [16] is also adopted as a metric that evaluates the quality of the ranking of the recommended locations. Discounted cumulative gain (DCG) measures of ranking quality consider both the relevance of the recommended locations and their positions in the ranking. It is calculated by summing the relevance of each recommended location at each position in the ranking, but discounting the contributions of the locations that appear lower in the list. The ideal discounted cumulative gain (IDCG) represents the best possible DCG score that can be achieved for a given set of recommended locations. The NDCG is determined by the normalized DCG by IDCG, thus ranging from 0 to 1. The NDCG is formally given by Equations (24)–(26):

$$\text{NDCG} = \frac{\text{DCG}}{\text{IDCG}} \quad (24)$$

$$\text{DCG} = \sum_{i=1}^n \frac{2^{\text{rel}_i} - 1}{\log_2(i + 1)} \quad (25)$$

$$\text{IDCG} = \sum_{i=1}^{|\text{POI}_{REL}|} \frac{2^{\text{rel}_i} - 1}{\log_2(i + 1)} \quad (26)$$

where N denotes the number of POIs to be recommended, rel_i represents the relevancy of the i th POI in the recommended list, and POI_{REL} signifies the ideal list of POIs containing the top- N relevant POIs among all of the considered POIs.

For the last metric, the quality of the recommendation list is evaluated by the proposed matching ratio metric. Precision and recall, while commonly used, may not be able to holistically reflect the performance of recommendation systems within the context of the simulated experiments. This is due to the nature of the Foursquare data that were collected from users without the influence of any recommendation system. Consequently, the precision and recall calculations were based solely on the comparison of the recommended list with the list of locations where the target users have checked in, without actual user interaction with the recommendation system. To summarize, these two metrics may not entirely capture the system's performance under some circumstances, particularly when the system could suggest places that align with the preferences of the target users, but such places are not included as ground truth POIs in the Foursquare dataset. The reason behind their absence might be that (i) the target users might have visited these places but forgot to record the check-in on the platform, and (ii) the places were not known or had never been recommended to the target users for consideration.

Recognizing these limitations, we proposed the “matching ratio” as a metric to measure the extent to which POI recommendation lists align with the user preferences. It is determined by determining the ratio of (i) the number of POIs whose distinctiveness matches with the target user's distinctiveness to (ii) the number of recommended POIs, as shown in Equation (27). The ratio value is in the range of 0–100%: higher values indicate better matching between the recommended POIs and the users' preferences.

$$\text{Matching ratio} = \frac{\text{number of POIs with matched distinctiveness}}{\text{number of Recommended POIs}} \times 100 \quad (27)$$

4.2. Experimental Results and Discussion

In this part, we examine the performance of our LACF-Rec3 method by assessing the top-N recommended locations, where N ranges from 5 to 20, across six areas: Tokyo, Chiyoda, Shinjuku, Minato, Shibuya, and Kawasaki. We conducted a comparative analysis against the selected baseline methods.

4.2.1. Recommendation for Cold-Start Users

We started our examination by focusing on the recommendations for cold-start users. This group of users presents a distinct challenge because of the lack of a check-in history or relevant users. As a result, we have no advantage of leveraging user profiles to create personalized recommendations. In response to this challenge, we determined high-quality POIs in proximity to the user's current location by using the distinctiveness of the POI across multiple aspects, instead of the check-in history or relevant users. As shown in Figures 5–8, our proposed method has mostly higher precision and recall than HITS by 63.30% and 59.21%, respectively.

Previous studies have introduced POI recommendation algorithms that primarily focus on check-in frequency, resulting in recommended lists influenced by the popularity of locations. However, it is worth noting that users often possess diverse preferences beyond mere popularity. Certain individuals may be interested in other aspects such as the places that local experts often visit. In an ideal scenario, an optimal recommendation would include perfect destinations that excel across all aspects (i.e., being high popularity, visited by diverse individuals, and consistently revisited). However, most places could be distinctive only in specific aspects. For example, some locales might be highly frequently visited by numerous individuals, albeit a small number of revisits; thus, standing out in only two aspects. Others could attract visitors daily. Although its user variety is limited, it has high frequency rate and maintains a consistent base of returning visitors. Similarly, certain places could be a less frequently visited destination, but its visitor groups are diverse and has a loyal base of returning visitors. Our proposed method accommodates these various aspects of locations to provide recommendations that meet these multifaceted user preferences.

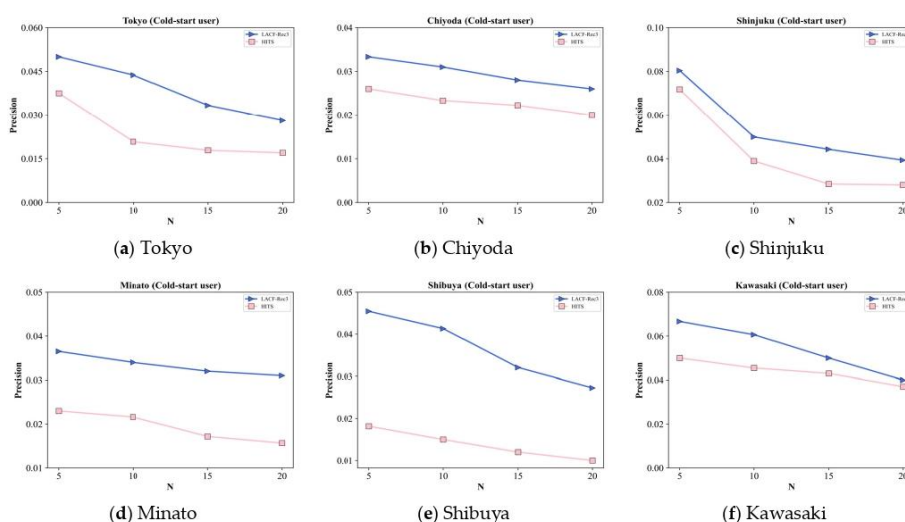


Figure 5. Precision metric for cold-start users with respect to the recommendation numbers.

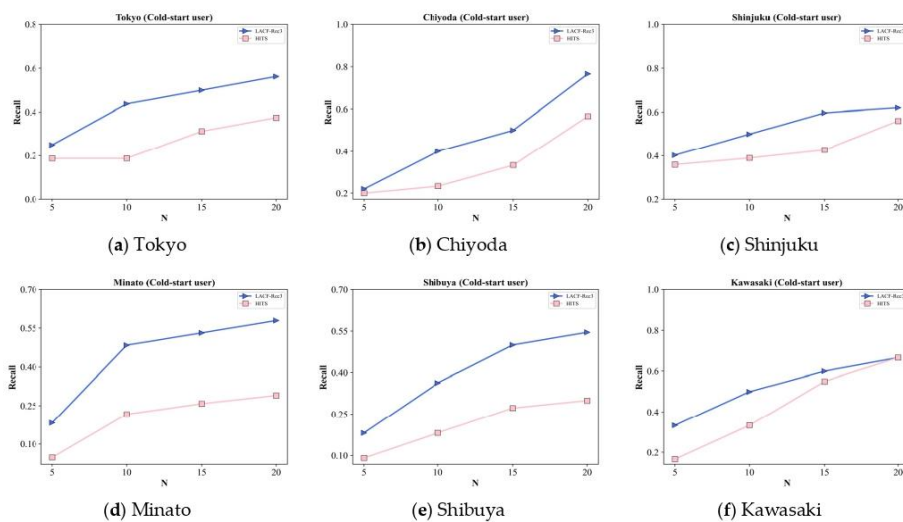


Figure 6. Recall metric for cold-start users with respect to the recommendation numbers.

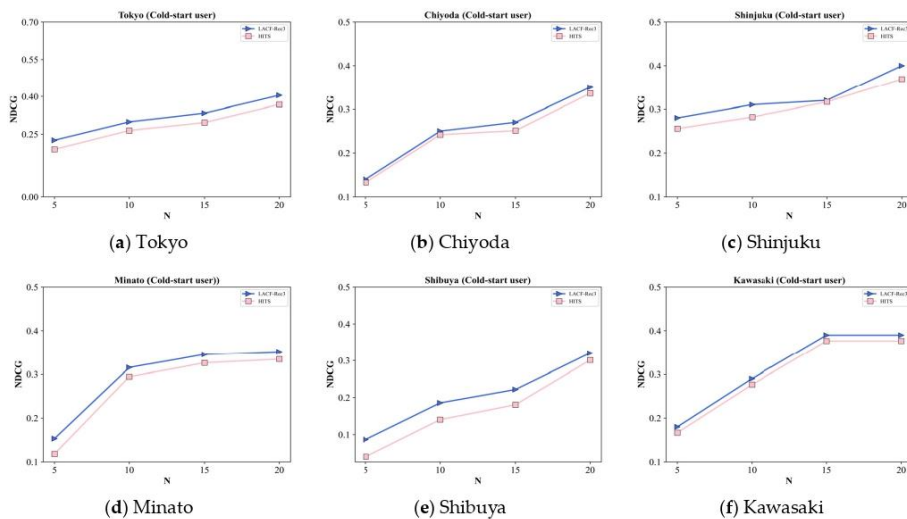


Figure 7. NDCG metric for cold-start users with respect to the recommendation numbers.

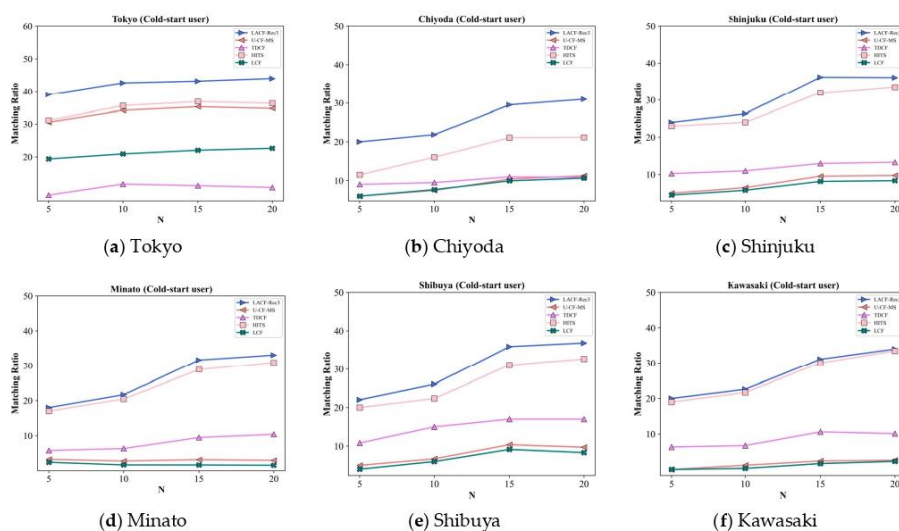


Figure 8. Matching ratio metric for cold-start users with respect to the recommendation numbers.

In addition, the ranks of the location recommendations selected by the user is higher than that of the previous method. The ranking accuracy evaluation using NDCG shows that the proposed algorithm was able to rank better than the HITS based model with 8.88%.

To evaluate the extent to which the location recommendations matched users' preferences, these were based on the matching ratio metric. According to the experimental results, our proposed method demonstrates the capability to recommend POIs that closely align with the user's preferences, particularly within the range of 25% to 50%. It outperforms U-CF-MS, TDCE, HITS, and UCF by 521.51%, 193.78%, 17.82%, and 793.39%, respectively.

Note that in the best case, our proposed method was able to provide recommendation lists that matched the users' preferences by 50%, meaning more than half of the recommended POIs are distinctive in the aspect different from those of the target users. This limitation stems from the fact that LACF-Rec3 determines the most distinctive aspect, which is then utilized by HIT-3, as opposed to simultaneously considering all three aspects. Our future work will focus on addressing this issue to bridge the gap.

In addition, we evaluated and compared our method with HITS as opposed to U-CF-MS, TDCE, and LCF. This choice stemmed from the inappropriateness of U-CF-MS and LCF for recommending locations for cold-start users, given their reliance on user profiles. Although TDCE is an extended version of HITS, it also requires a user profile to identify the active area of the target user. Hence, it cannot generate a POI recommendation list for cold-start users. The experimental results show that our method outperformed HITS in all of these areas. Especially in areas where cold-start users are prevalent such as Minato and Shibuya, our approach to generate recommendations is way more accurate than HITS. As HITS creates recommendations based solely on frequency, it is not suitable for such areas where users often visit diverse places and have a high rate of revisiting.

4.2.2. Recommendation for Existing Users

We investigated the efficiency of the recommendation list for existing users, and the experimental results are shown in Figures 9–12. Our LACF-Rec3 method demonstrated a significant improvement in terms of precision by 60.35%, 96.92%, 116.59%, and 185.49% when compared to U-CF-MS, TDCE, HITS, and UCF, respectively. Similarly, our method

also showed an improvement in terms of recall by approximately 61.44%, 98.39%, 93.23%, and 193.23% when compared to U-CF-MS, TDCF, HITS, and UCF, respectively.

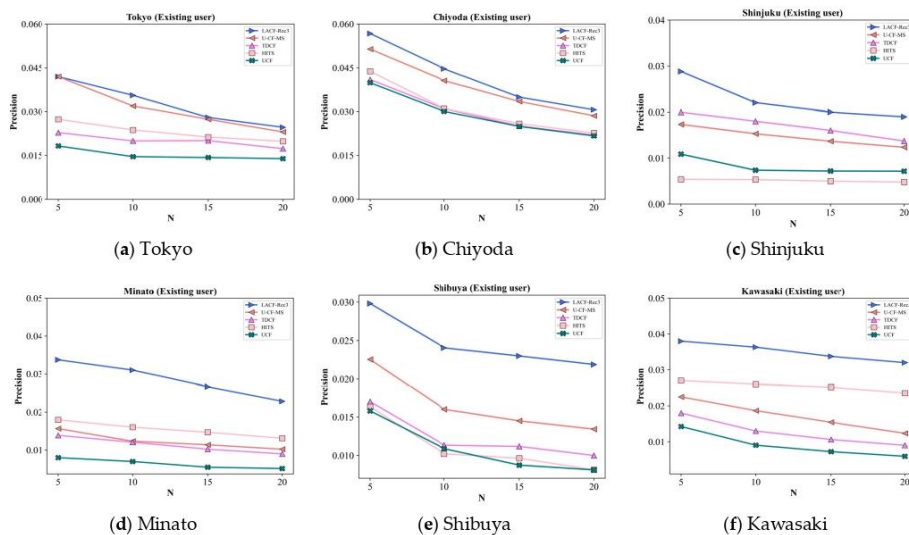


Figure 9. Precision metric for existing users with respect to the recommendation numbers.

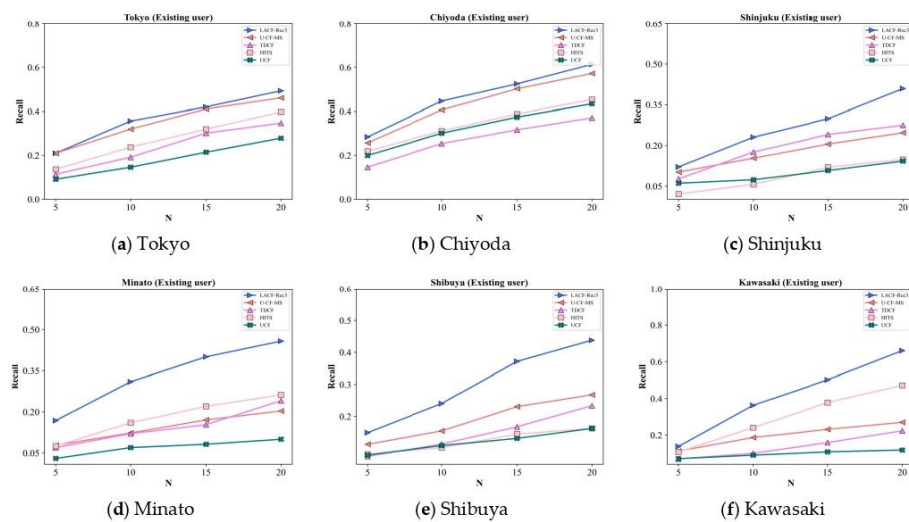


Figure 10. Recall metric for existing users with respect to the recommendation numbers.

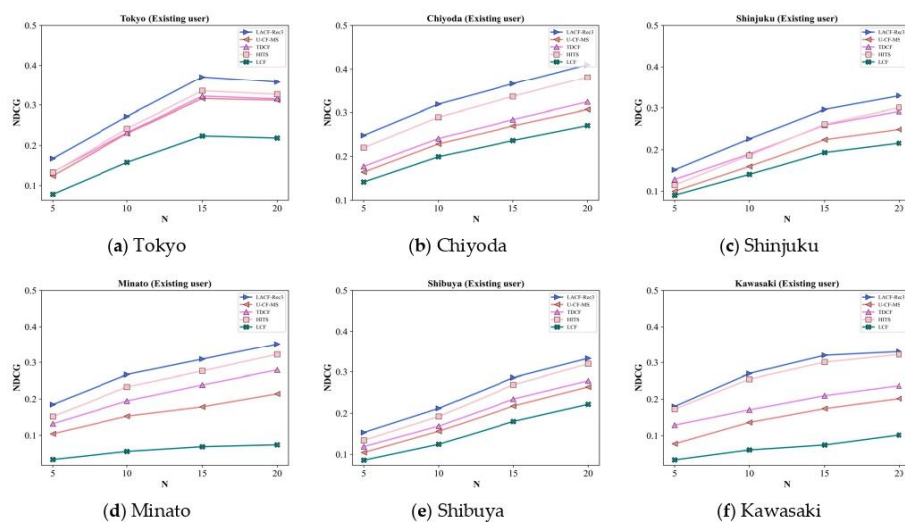


Figure 11. NDCG metric for existing users with respect to the recommendation numbers.

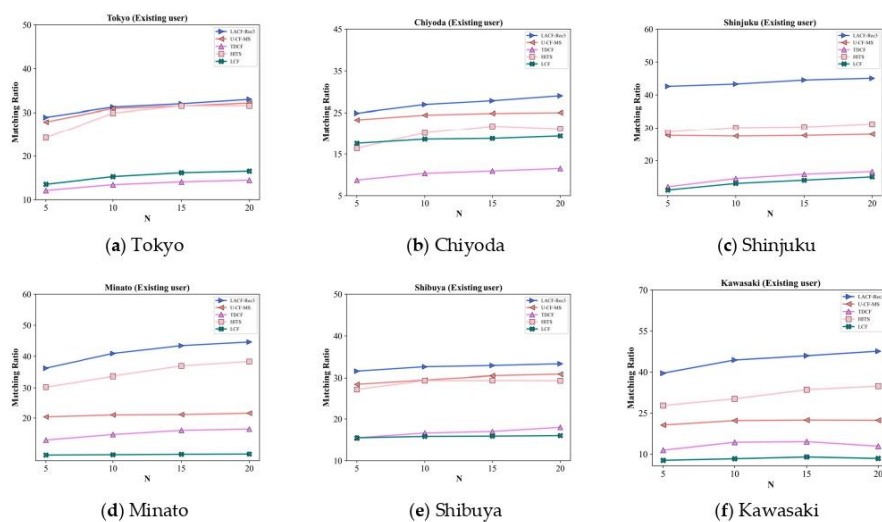


Figure 12. Matching ratio metric for existing users with respect to the recommendation numbers.

Since the target user has a check-in history, we can identify the target user's location interest by analyzing the distinctiveness of their favorite locations. Then, the distinctiveness of location associated with the target user can be compared with that of the other users, based on their check-in histories. Therefore, our proposed algorithm can identify users with similar interests to the target user more effectively than the other algorithms that directly consider only the locations that users have visited.

In addition, in real-world scenarios, users often have only a few check-ins at their favorite locations compared to a large number of locations. This data sparsity causes a difficulty in searching for other users who are similar to the target user. Our proposed method addresses this issue by recommending POIs based on the location distinctiveness related to relevant users, rather than comparing the list of locations visited by relevant users. As a consequence, our method is more capable of discovering relevant users with similar distinctiveness (and thus, similar preference) to the target user and is able to recommend lists of POIs that better align with the target users' preference.

Accordingly, the proposed method could generate more personalized POI recommendations than other algorithms, resulting in higher ranks of POIs in the user-selected recommendation list compared to the other methods. We observed that our proposed method outperformed U-CF-MS, TDCF, HITS, and UCF in terms of the NDCG metric by 46.49%, 26.17%, 9.36%, and 137.66%, respectively.

For the case of existing users, the evaluation of the matching ratio metric showed that our proposed algorithm provided better retrieval accuracy for POIs that matched the user preferences than that of U-CF-MS, TDCF, HITS, and UCF by 45.82%, 159.83%, 25.59%, and 205.21%, respectively. This is due to the fact that our method evaluates the distinctiveness by considering all three aspects, resulting in the ability to promote and recommend diverse types of POIs as well as avoid recommending popular yet common tourist destinations that could cause user disinterest.

Location recommendations may suffer from a tedium problem, where locations are chosen largely based on their popularity, and not on the variety and repetition rates. By considering the distinctiveness of locations based on the frequency, variety, and repetition, our POI recommendation list offers diverse and outstanding locations in each area. This helps users receive recommendations from multiple perspectives, thus solving the tedium problem.

Moreover, our analysis revealed that in areas with a large number of user check-ins (or a large number of existing users) such as Tokyo and Chiyoda, our LACF-Rec3 algorithm provided more accurate recommendations than those for areas with fewer user check-ins such as Shinjuku, Minato, Shibuya, and Kawasaki. More check-in data enabled our CF-3 to search for relevant users with similar check-in patterns more effectively. Therefore, our method excels in creating POI recommendation lists that align more accurately with the user preferences than the other methods.

In future work, we aim to focus on considering multiple aspects of distinctiveness among users and locations, as opposed to this work, where the distinctiveness of users and locations was based only on one aspect. This refinement becomes relevant in cases where users or locations exhibit multiple aspects of distinctiveness, leading to a possible decrease in the accuracy.

4.2.3. Privacy Issue Related to User Behavior Data Used in Recommendation Algorithms

The effectiveness of recommendation algorithms fundamentally relies on the acquisition and analysis of user behavior data such as the users' preferences, interests, or historical interactions to make personalized recommendations. Users are required to disclose information, creating a trade-off between utility and user privacy. While obtaining accurate recommendations is crucial, the sharing of personal information leads to the potential for privacy breaches, which may occur either deliberately (through snooping or hacking) or accidentally.

Although this study utilized a benchmark, which is masked and publicly available data, that did not contain sensitive information, it is essential to note that when the algorithm is deployed in production, users will need to provide information to access the desired features of our proposed method. Here, the privacy issue becomes a paramount concern. It is crucial to consider the legal and regulatory frameworks across all countries where the applications are utilized. Various countries have different data privacy regulations such as the European Union's General Data Protection Regulation (GDPR),

which imposes strict requirements on the collection, processing, and usage of user data. Non-compliance with these regulations can lead to substantial fines and legal repercussions. Obtaining informed consent from users for the collection and utilization of their data is also a critical legal obligation. To mitigate the potential privacy risks associated with the user behavior data, it is advisable to consider the adoption of existing privacy preserving recommendation methods such as cryptography-based approaches [37,38] or a two-stage privacy protection mechanism [39].

5. Conclusions

In this work, we proposed a novel hybrid POI recommendation system that combined link analysis and collaborative filtering, each of which was based on three visiting behaviors: frequency, variety, and repetition. Our method, called LACF-Rec3, focuses on the visiting characteristics of POIs and those of users to recommend POIs to existing and cold-start users. We introduced an extended version of the HITS-based model, called HITS-3, to generate top-ranked POIs for cold-start users. This model can handle not only the cold-start problem, but the sparsity problem by generating POI recommendation lists without using the user profile. To recommend interesting POIs for existing users, we introduced a novel collaborative filtering technique, called CF-3, which takes into consideration the distinctiveness of both the users (i.e., checked-in frequency, user variety, and checked-in repetitions) and POIs (checked-in frequency, user variety, and checked-in repetitions). CF-3 utilizes location distinctiveness, obtained from the HITS-3, to determine the similarity between the locations visited by target users and the ones visited by each relevant user. Then, we determined interesting locations based on the frequency of check-in, similarity of relevant users, and user distinctiveness to provide highly personalized recommendations. By considering distinctiveness, our method can help create diverse recommendation lists and address the tedium problem.

For our experimental evaluation on a real-world dataset, we used three well-known metrics: two for measuring accuracy (i.e., precision and recall) and the other for assessing the ranking accuracy (i.e., NDCG). In addition, this work proposed the matching ratio metric for evaluating the quality of recommendation lists by considering the distinctiveness of the target users in relation to the recommended POIs.

Our experimental results showed that our LACF-Rec3 method outperformed the baseline methods in terms of precision, recall, NDCG, and matching ratio. Our method effectively captured the diverse preferences of users by considering the distinctiveness of the POIs and users. For cold-start users, we recommended interesting locations based on location distinctiveness in their vicinity. As a result, our LACF-Rec3 method can generate diverse POI recommendation lists that match the preferences of target users.

For existing users, finding similarities in user behaviors can be facilitated by considering distinctiveness similarity rather than relying solely on location similarity. Furthermore, our method provides interesting locations to existing users based on their preferences.

The advantage of our LACF-Rec3 method lies in its ability to effectively capture the user preferences of target users, thereby generating POI recommendations. It is suitable for both cold-start users and existing users, relying on our proposed distinctiveness of locations and users, based on visiting frequency, variety, and repetition. Consequently, our method excels in the recommendation performance in terms of accuracy, ranking recommendation, and the alignment of recommended POIs with user preferences when compared to the baseline methods, which rely solely on frequency and/or repetition.

In future work, we plan to explore the characteristics of the POIs and users from multiple aspects (not just only one with the most distinctiveness). For example, some locations should be characterized by both the frequency and diversity of user check-ins, while some users may prefer visiting popular and frequently visited locations. In addition, we will attempt to evaluate the effectiveness of our method in terms of novelty and diversity. We also plan to incorporate novelty and diversity into the future version of our method to more efficiently cope with the tedium problem.

Author Contributions: Conceptualization, Sumet Darapisut, Komate Amphawan, Nutthanon Leelathakul and Sunisa Rimcharoen; Methodology, Komate Amphawan, Nutthanon Leelathakul and Sunisa Rimcharoen; Software, Sumet Darapisut; Validation, Sumet Darapisut and Komate Amphawan; Formal analysis, Komate Amphawan; Investigation, Komate Amphawan; Writing—original draft, Sumet Darapisut and Komate Amphawan; Writing—review & editing, Komate Amphawan, Nutthanon Leelathakul and Sunisa Rimcharoen; Visualization, Sumet Darapisut. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Faculty of Informatics, Burapha University, grant number 02/2565.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Zheng, Y.; Zhou, X. Location-based social networks: Users. In *Computing with Spatial Trajectories*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 243–276.
- Yu, Y.; Chen, X. A survey of point-of-interest recommendation in location-based social networks. In Proceedings of the Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015.
- Zhao, S.; King, I.; Lyu, M.R. A survey of point-of-interest recommendation in location-based social networks. *arXiv* **2016**, arXiv:1607.00647.
- Xu, S.; Fu, X.; Cao, J.; Liu, B.; Wang, Z. Survey on user location prediction based on geo-social networking data. *World Wide Web* **2020**, *23*, 1621–1664. [[CrossRef](#)]
- Bao, J.; Zheng, Y.; Mokbel, M.F. Location-based and preference-aware recommendation using sparse geo-social networking data. In Proceedings of the 20th International Conference on Advances in Geographic Information Systems, Redon Beach, CA, USA, 6–9 November 2012; pp. 199–208.
- Zhou, D.; Wang, B.; Rahimi, S.M.; Wang, X. A Study of Recommending Locations on Location-Based Social Network by Collaborative Filtering. In Proceedings of the Advances in Artificial Intelligence, Toronto, ON, Canada, 28–30 May 2012; Kosseim, L., Inkpen, D., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 255–266.
- Lim, K.H.; Chan, J.; Leckie, C.; Karunasekera, S. Personalized Trip Recommendation for Tourists Based on User Interests, Points of Interest Visit Durations and Visit Recency. *Knowl. Inf. Syst.* **2018**, *54*, 375–406. [[CrossRef](#)]
- Kleinberg, J.M. Authoritative sources in a hyperlinked environment. *J. ACM* **1999**, *46*, 604–632. [[CrossRef](#)]
- Darapisut, S.; Amphawan, K.; Rimcharoen, S.; Leelathakul, N. N-Most Interesting Location-based Recommender System. *ECTI Trans. Comput. Inf. Technol.* **2022**, *16*, 84–99. [[CrossRef](#)]
- Baral, R.; Wang, D.; Li, T.; Chen, S.C. GeoTeCS: Exploiting Geographical, Temporal, Categorical and Social Aspects for Personalized POI Recommendation (Invited Paper). In Proceedings of the 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI), Pittsburgh, PA, USA, 28–30 July 2016; pp. 94–101.
- Acharya, M.; Mohbey, K.K. Trust-aware spatial-temporal feature estimation for next POI recommendation in location-based social networks. *Soc. Netw. Anal. Min.* **2023**, *13*, 102. [[CrossRef](#)]
- Gan, M.; Gao, L. Discovering Memory-Based Preferences for POI Recommendation in Location-Based Social Networks. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 279. [[CrossRef](#)]
- Yin, M.; Liu, Y.; Zhou, X.; Sun, G. A Tensor Decomposition Based Collaborative Filtering Algorithm for Time-Aware POI Recommendation in LBSN. *Multimed. Tools Appl.* **2021**, *80*, 36215–36235. [[CrossRef](#)]
- Long, X.; Joshi, J. A HITS-based POI recommendation algorithm for Location-Based Social Networks. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013), Niagara Falls, ON, Canada, 25–28 August 2013; pp. 642–647.
- Ying, Y.; Chen, L.; Chen, G. A temporal-aware POI recommendation system using context-aware tensor decomposition and weighted HITS. *Neurocomputing* **2017**, *242*, 195–205. [[CrossRef](#)]
- Wang, Y.; Wang, L.; Li, Y.; He, D.; Liu, T.Y. A theoretical analysis of NDCG type ranking measures. In Proceedings of the Conference on Learning Theory, PMLR, Princeton, NJ, USA, 12–14 June 2013; pp. 25–54.
- Chakrabarti, S.; Dom, B.; Raghavan, P.; Rajagopalan, S.; Gibson, D.; Kleinberg, J. Automatic resource compilation by analyzing hyperlink structure and associated text. *Comput. Netw. ISDN Syst.* **1998**, *30*, 65–74. [[CrossRef](#)]
- Bao, J.; Zheng, Y.; Wilkie, D.; Mokbel, M. Recommendations in location-based social networks: A survey. *GeoInformatica* **2015**, *19*, 525–565. [[CrossRef](#)]
- Darapisut, S.; Amphawan, K.; Rimcharoen, S.; Leelathakul, N. NILR: N-Most Interesting Location-based Recommender System. In Proceedings of the Conference on Smart Media and Applications, Jeju, Republic of Korea, 17–19 September 2020.
- Zheng, Y.; Zhang, L.; Xie, X.; Ma, W.Y. Mining interesting locations and travel sequences from GPS trajectories. In Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, 20–24 April 2009; pp. 791–800.

21. Bagci, H.; Karagoz, P. Context-aware location recommendation by using a random walk-based approach. *Knowl. Inf. Syst.* **2016**, *47*, 241–260. [CrossRef]
22. Sun, J.; Matsushima, Y.; Ma, Q. Property Analysis of Stay Points for POI Recommendation. In Proceedings of the Database and Expert Systems Applications, Virtual Event, 27–30 September 2021; Strauss, C., Kotsis, G., Tjoa, A.M., Khalil, I., Eds.; Springer: Cham, Switzerland, 2021; pp. 374–379.
23. Fan, J.; Pan, C.; Geng, Y.; Li, S. A Privacy-Preserving Time-Aware Method for Next POI Recommendation. *Electronics* **2023**, *12*, 3208. [CrossRef]
24. Sarwar, B.; Karypis, G.; Konstan, J.; Riedl, J. Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th International Conference on World Wide Web, Hong Kong, China, 1–5 May 2001; pp. 285–295.
25. Su, X.; Khoshgoftar, T.M. A survey of collaborative filtering techniques. *Adv. Artif. Intell.* **2009**, *2009*, 421425. [CrossRef]
26. Koren, Y.; Rendle, S.; Bell, R. Advances in collaborative filtering. In *Recommender Systems Handbook*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 91–142.
27. Baral, R.; Li, T. Exploiting the roles of aspects in personalized POI recommender systems. *Data Min. Knowl. Discov.* **2018**, *32*, 320–343. [CrossRef]
28. Zhao, G.; Lou, P.; Qian, X.; Hou, X. Personalized location recommendation by fusing sentimental and spatial context. *Knowl.-Based Syst.* **2020**, *196*, 105849. [CrossRef]
29. Yuan, Q.; Cong, G.; Ma, Z.; Sun, A.; Thalmann, N.M. Time-Aware Point-of-Interest Recommendation. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA, 28 July–1 August 2013; SIGIR '13. pp. 363–372.
30. Si, Y.; Zhang, F.; Liu, W. CTF-ARA: An adaptive method for POI recommendation based on check-in and temporal features. *Knowl.-Based Syst.* **2017**, *128*, 59–70. [CrossRef]
31. Khazaei, E.; Alimohammadi, A. Context-Aware Group-Oriented Location Recommendation in Location-Based Social Networks. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 406. [CrossRef]
32. Zhang, H.; Gan, M.; Sun, X. Incorporating Memory-Based Preferences and Point-of-Interest Stickiness into Recommendations in Location-Based Social Networks. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 36. [CrossRef]
33. Zhou, Y.; Zhou, K.; Chen, S. Context-Aware Point-of-Interest Recommendation Based on Similar User Clustering and Tensor Factorization. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 145. [CrossRef]
34. Yang, D.; Zhang, D.; Zheng, V.W.; Yu, Z. Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs. *IEEE Trans. Syst. Man Cybern. Syst.* **2014**, *45*, 129–142. [CrossRef]
35. Yang, S.; Liu, J.; Zhao, K. GETNext: Trajectory Flow Map Enhanced Transformer for Next POI Recommendation. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, NY, USA, 11–15 July 2022; SIGIR '22. pp. 1144–1153.
36. Geocoding API. Available online: <https://developers.google.com/maps/documentation/geocoding/> (accessed on 1 August 2022).
37. Kim, J.; Koo, D.; Kim, Y.; Yoon, H.; Shin, J.; Kim, S. Efficient privacy-preserving matrix factorization for recommendation via fully homomorphic encryption. *ACM Trans. Priv. Secur.* **2018**, *21*, 1–30. [CrossRef]
38. Ge, Z.; Liu, X.; Li, Q.; Li, Y.; Guo, D. PrivItem2Vec: A privacy-preserving algorithm for top-N recommendation. *Int. J. Distrib. Sens. Netw.* **2021**, *17*. [CrossRef]
39. Xin, X.; Yang, J.; Wang, H.; Ma, J.; Ren, P.; Luo, H.; Shi, X.; Chen, Z.; Ren, Z. On the user behavior leakage from recommender system exposure. *ACM Trans. Inf. Syst.* **2023**, *41*, 1–25. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

