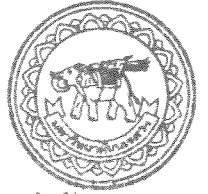


อธินันตนาการ



สำนักหอสมุด



รายงานวิจัยฉบับสมบูรณ์

การพัฒนาระบบแนะนำการซื้อจักรยานเชิงท่องเที่ยวบนพื้นฐานรูปแบบ
การซื้อเฉพาะบุคคลด้วยเทคโนโลยีอุปกรณ์ไอทีสวมใส่

ผู้ช่วยศาสตราจารย์ ดร.เกรียงศักดิ์ เตมีย์

สำนักหอสมุด มหาวิทยาลัยนครสวรรค์

วันลงทะเบียน... 4 ก.พ. 2565

เลขทะเบียน... 1048497

เลขเรียกหนังสือ... 2 GV

10104

ก769ร

2563

เดือน กันยายน ปี 2563

สัญญาเลขที่ R2561B074

รายงานวิจัยฉบับสมบูรณ์

การพัฒนาระบบแนะนำการขึ้นจักรยานเชิงท่องเที่ยวบนพื้นฐานรูปแบบ
การขึ้นเฉพาะบุคคลด้วยเทคโนโลยีอุปกรณ์ไอทีสวมใส่



ผู้ช่วยศาสตราจารย์ ดร.เกรียงศักดิ์ เตมีย์
ภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
คณะวิทยาศาสตร์ มหาวิทยาลัยรัตนนคร

สนับสนุนโดย
งบประมาณเงินแผ่นดิน
ปีงบประมาณ 2561

กิตติกรรมประกาศ

โครงการวิจัยนี้ได้รับทุนอุดหนุนวิจัยจากมหาวิทยาลัยขอนแก่นให้ทุนอุดหนุนการวิจัยจากงบประมาณเงินแผ่นดิน ประจำปีงบประมาณ พ.ศ. 2561 สัญญาเลขที่ R2561B074 เรื่อง การพัฒนาระบบแนะนำการชี้จักรยานเชิงท่องเที่ยวบนพื้นฐานรูปแบบการชี้เฉพาะบุคคลด้วยเทคโนโลยีอุปกรณ์ไอทีสวมใส่ โดยได้งบประมาณแผ่นดิน ในวงเงิน 2,229,700.00 บาท (สองล้านสองแสนสองหมื่นเก้าพันเจ็ดร้อยบาทถ้วน) และงบประมาณจากภาคเอกชน บริษัท เทเลคอร์ป จำกัด ในวงเงิน 560,800 บาท (ห้าแสนหกหมื่นแปดร้อยบาทถ้วน)

ขอขอบคุณภาควิชาวิทยาการคอมพิวเตอร์และเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร ที่ให้การสนับสนุนวัสดุอุปกรณ์ รวมถึงสถานที่ในการทำวิจัย

ประโยชน์อันพึงมีจากรายงานวิจัยฉบับนี้ ผู้จัดทำขอมอบให้กับผู้มีพระคุณทุกท่าน หากรายงานฉบับนี้มีข้อผิดพลาดประการใดผู้จัดทำขอน้อมรับไว้เพียงผู้เดียว



เกรียงศักดิ์ เตมีย

Executive Summary

จากผลสรุปงานวิจัยด้านระบบสารสนเทศที่สนับสนุนการท่องเที่ยวที่มีอยู่เดิมนั้น จะพบว่าส่วนใหญ่ยังไม่พิจารณาถึงมุมมองด้านกิจกรรมสันตนาการมากนัก คณะผู้วิจัยจึงได้กำหนดขอบเขตปัญหาการวิจัย โดยมุ่งเน้นที่กิจกรรมการชี้จรรย์านรายบุคคลบนพื้นฐานรูปแบบการชี้เฉพาะบุคคล เป็นปัญหาในการวิจัย เนื่องจากเป็นกลุ่มเป้าหมายใหม่ที่มีกำลังทรัพย์และมีการเติบโตขึ้นอย่างต่อเนื่อง ดังนั้นการวิจัยเพื่อพัฒนาระบบแนะนำการชี้จรรย์านเชิงท่องเที่ยวบนพื้นฐานรูปแบบการชี้เฉพาะบุคคล ด้วยเทคโนโลยีอุปกรณ์ไอทีที่สวมใส่นี้ ประกอบด้วย

1. การรวบรวมข้อมูลการชี้จรรย์านจากผู้ให้บริการจัดเก็บข้อมูลการชี้จรรย์านต่างๆ เพื่อนำมาจำแนกและจัดกลุ่มรูปแบบการชี้จรรย์านที่เฉพาะบุคคลแบบต่างๆ ซึ่งอุปกรณ์ไอทีที่แบบสวมใส่สามารถบันทึกข้อมูลเหล่านี้แบบทันทีทันใด โดยผู้ชี้จรรย์านสามารถจัดเก็บข้อมูลดังกล่าวเข้าสู่ระบบจัดเก็บข้อมูล
2. การจำแนกและจัดกลุ่มลักษณะของเส้นทางจากแหล่งท่องเที่ยวสำหรับการชี้จรรย์าน และรูปแบบการชี้เฉพาะบุคคล เพื่อเป็นส่วนสำหรับระบบแนะนำ โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning)
3. ระบบการแนะนำเส้นทางจรรย์านตามความต้องการของผู้ใช้ และพิกัดของเส้นทางที่ได้รับคความนิยม

ซึ่งผลลัพธ์ที่มีผลการประเมินประสิทธิภาพในการจำแนกและจัดกลุ่มที่ดีที่สุดจะถูกนำมาประยุกต์ใช้ในกระบวนการแนะนำเส้นทางที่เหมาะสมกับการชี้จรรย์านเฉพาะบุคคลต่อไป โดยผู้ใช้มีความพึงพอใจต่อระบบที่พัฒนาขึ้นอยู่ในระดับดี

บทนำ

การวิจัยเพื่อพัฒนาระบบแนะนำการช้อปปิ้งออนไลน์เชิงท่องเที่ยวบนพื้นฐานรูปแบบการช้อปปิ้งเฉพาะบุคคล ด้วยเทคโนโลยีอุปกรณ์ไอทีที่สวมใส่นี้ประกอบด้วยการรวบรวมข้อมูลการช้อปปิ้งจากผู้ใช้บริการจัดเก็บข้อมูลการช้อปปิ้งต่างๆ เพื่อนำมาจำแนกและจัดกลุ่มรูปแบบการช้อปปิ้งที่เฉพาะบุคคลแบบต่างๆ เพื่อเป็นส่วนในการแนะนำตามความต้องการของผู้ใช้ และพิกัดของเส้นทางที่ได้รับความนิยม โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งผลลัพธ์ที่มีผลการประเมินประสิทธิภาพในการจำแนกและจัดกลุ่มที่ดีที่สุดจะถูกนำมาประยุกต์ใช้ในกระบวนการแนะนำเส้นทางที่เหมาะสมกับการช้อปปิ้งเฉพาะบุคคลต่อไป โดยผู้ใช้มีความพึงพอใจต่อระบบที่พัฒนาขึ้นอยู่ในระดับดี



Abstract

This research project proposes a Recreation Cycling Recommendation System based on Personalized Cycling Style with Wearable Technology. The research methodology begins with a process of data collection from activities' data providers. The collected data was cleansed and transformed into an analysis-ready dataset. The machine learning techniques were applied with data set for route pattern recognition. Then, the algorithm's result as route patterns were included in the recommendation algorithm for presenting cycling routes that meet user preferences. The proposed system was additionally evaluated by the user satisfaction survey, which its result was leveled as "good".



สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทสรุปผู้บริหาร	ข
บทคัดย่อ	ค
Abstract	ง
สารบัญ	จ
บทที่	
1 บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
วัตถุประสงค์ของการศึกษา	4
ทฤษฎีและกรอบแนวคิดของงานวิจัย	5
วิธีการดำเนินการวิจัย	6
ขอบเขตการดำเนินงาน	7
ประโยชน์ที่คาดว่าจะได้รับ	7
2 ทฤษฎีพื้นฐานและงานวิจัยที่เกี่ยวข้อง	8
ความหมายของระบบแนะนำ	8
ข้อพิจารณาในการพัฒนาระบบแนะนำ	9
การให้คะแนนต่อสินค้า	10
ประเภทของระบบแนะนำ	11
3 Trajectory Clustering by GPS tracking dataset using QuickBundles	17
Problem Statements and Preliminaries.	17
Experiment	19
Route Clustering	22
Discussion and Conclusion	24
4 การแนะนำเส้นทางในการขี่จักรยานเฉพาะบุคคล	26
การจัดเตรียมข้อมูลผู้ขี่จักรยาน	26
การแนะนำเส้นทางเฉพาะบุคคล	29
ผลการทดลอง	30
5 สรุปผลการทดลอง อธิบายผลและข้อเสนอแนะ	32

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันการท่องเที่ยวเชิงสันทนาการได้รับความนิยมอย่างแพร่หลายทั่วโลก เนื่องจากมีกิจกรรมสันทนาการที่มีส่วนช่วยส่งเสริมสุขภาพร่างกาย และช่วยให้จิตใจได้ผ่อนคลายในเวลาเดียวกัน โดยนาย John Tribe (Tribe, 2011) ได้ให้คำจำกัดความของการท่องเที่ยวเชิงสันทนาการว่าเป็นการท่องเที่ยวที่มีกิจกรรมที่ทำหายเข้ามาเกี่ยวข้อง เช่น การปีนหน้าผา ตกปลา ตกอล์ฟ เป็นต้น ซึ่งการท่องเที่ยวด้วยการขี่จักรยานเป็นกิจกรรมที่ทำหายชนิดหนึ่งที่ได้รับนิยาม มีการจัดกิจกรรมที่หลากหลายประเภท เช่น การขี่จักรยานท่องเที่ยวในเส้นทางธรรมชาติ หรือการแข่งขันกีฬาจักรยาน เป็นต้น โดยในประเทศไทยนั้น กิจกรรมขี่จักรยานถือเป็นการท่องเที่ยวเชิงสันทนาการหนึ่ง ที่ได้รับความนิยมเป็นอย่างมาก เนื่องจากมีภูมิอากาศที่เหมาะสม เอื้ออำนวยให้สามารถขี่จักรยานได้ตลอดทั้งปี เป็นสาเหตุให้ชาวต่างชาตินิยมเข้ามาจัดกิจกรรมขี่จักรยานในประเทศไทย เช่น ทีมนักกีฬาจักรยานชาติญี่ปุ่นเลือกมาเก็บตัวฝึกซ้อมที่จังหวัดเชียงรายเพื่อหลีกเลี่ยงสภาพอากาศอันเลวร้าย ในช่วงฤดูหนาวของแต่ละปี กลุ่มผู้ประกอบการท่องเที่ยวที่จัดกิจกรรมการขี่จักรยานตามเส้นทางธรรมชาติต่างๆ ในจังหวัดเชียงใหม่ เป็นต้น รวมไปถึงภาคธุรกิจที่เล็งเห็นแนวโน้มการเติบโตของตลาดนักท่องเที่ยวที่ต้องการพัฒนาศักยภาพของตนเองสู่ความเป็นเลิศ เช่น วิทยาลัยบูรณะ สปอร์ตไฮเทค จังหวัดภูเก็ต ที่เปิดประสบการณ์ฝึกซ้อมไตรกีฬาด้วยโปรแกรมการฝึกซ้อมไตรกีฬาระยะสั้น จากคณะผู้ฝึกสอนที่มีประสบการณ์ชาวไทยและชาวต่างประเทศ โดยมีภูมิประเทศที่สวยงามของจังหวัดภูเก็ตเป็นที่ตั้ง เป็นต้น จากสถานการณ์ดังกล่าว พบว่ามีความสอดคล้องกับแผนการขับเคลื่อนยุทธศาสตร์ท่องเที่ยว ปี 2558-2560 เรื่องการ พัฒนาสินค้าและบริการ ข้อที่ 2 การพัฒนารายสาขา ข้อย่อยที่ 2 การท่องเที่ยวเชิงกีฬา และยังเป็นการยกระดับภาพลักษณ์การท่องเที่ยวให้เป็น “Quality Leisure Destination” โดยเป็นการแสวงหาดูแลกลุ่มใหม่ คณะผู้วิจัยจึงเห็นว่าสมควรที่จะพัฒนาระบบสนับสนุนการท่องเที่ยวเชิงสันทนาการดังกล่าว โดยระบบแนะนำการท่องเที่ยวเชิงสันทนาการควรตอบสนองต่อความต้องการของนักท่องเที่ยวเชิงสันทนาการโดยตรง จึงได้รวบรวมและสรุปผลงานวิจัยด้านระบบแนะนำการท่องเที่ยวที่น่าสนใจต่างๆ โดยมีผลสรุปดังนี้

ระบบแนะนำการท่องเที่ยวแบบ Personalization techniques (Gao, Liu, & Wu, 2010) มีเป้าหมายที่จะทำการปรับข้อมูลเกี่ยวกับการท่องเที่ยวให้เหมาะสมกับผู้ใช้ ตามความต้องการและรสนิยมของผู้ใช้ และ สร้างระบบแนะนำ (Adomavicius & Tuzhilin, 2005), โดยใช้การคัดกรอง

ข้อมูลที่ไม่เกี่ยวข้องออกไป เพื่อให้ได้ข้อมูลที่เฉพาะบุคคลมากที่สุด สำหรับผู้ใช้แต่ละคน สำหรับในด้านการท่องเที่ยว มีระบบแนะนำการเดินทาง (Ricci, 2002) ที่มีเป้าหมายในการพยายามจับคู่ลักษณะการเดินทางท่องเที่ยวและกิจกรรมต่างให้ตรงกับความต้องการของผู้ใช้ ซึ่งระบบดังกล่าวจะมีประโยชน์มากถ้าระบบสามารถวิเคราะห์ความต้องการของผู้ใช้ได้เอง โดยทั้งทางตรงและทางอ้อม (Sieg, Mobasher, & Burke, 2007) ข้อมูลทางตรงนั้นอาจได้มาจากผู้ใช้เอง เช่น การระบุข้อมูลความชอบโดยการกรอกแบบฟอร์ม และโปรแกรมให้ผู้ใช้พิมพ์โต้ตอบกับระบบเหมือนระบบข้อความสำหรับข้อมูลทางอ้อมได้มาจากการวิเคราะห์พฤติกรรมของผู้ใช้เอง

สำหรับระบบแนะนำการท่องเที่ยวบางระบบ ไม่เพียงแต่นำความต้องการของผู้ใช้เท่านั้น แต่ยังวิเคราะห์ บริบท (context) ผู้ใช้ระหว่างการเดินทาง (Abowd et al., 1999) เมื่อผู้ใช้เดินทางถึงเป้าหมายในระดับหนึ่ง สามารถปรับค่าความต้องการ และระบบปรับแผนการเดินทางแบบ real time ซึ่งจะมีประโยชน์มากและด้วยเทคโนโลยีในปัจจุบัน ถ้าผู้ใช้ ปรับความต้องการผ่าน Mobile Application ซึ่งข้อมูลบริบทที่รวบรวมไปถึง ที่ตั้ง เวลาที่ใช้โปรแกรม เพื่อเทียบเคียงถึงสภาพอากาศ และนำไปปรับข้อมูลการท่องเที่ยวใหม่ให้ผู้ใช้ (Lamsfus, Alzua-Sorzabal, Martin, Salvador, & Usandizaga, 2009)

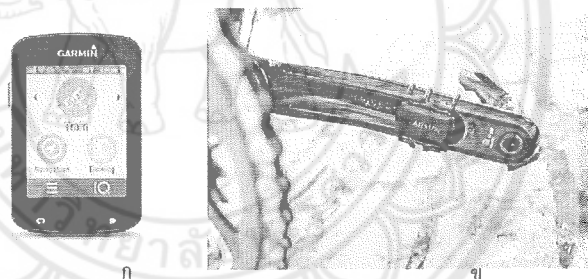
ระบบแนะนำการท่องเที่ยวนั้น ส่วนมากจะสนใจเรื่องการจัดการท่องเที่ยวให้ได้ตรงกับรสนิยมหรือความชอบโดยอาศัยข้อมูลที่รวบรวมทั้งทางตรงและทางอ้อม แต่ไม่ได้นำข้อมูลเรื่องสุขภาพของนักท่องเที่ยวมาพิจารณา มีบางระบบที่พิจารณาถึงสุขภาพของนักท่องเที่ยวร่วมด้วย (Agapito et al., 2014) สร้างระบบแนะนำอาหารท้องถิ่นที่เหมาะสมกับนักท่องเที่ยวเพื่อเป้าหมายให้มีสุขภาพที่ดี รวมถึงจัดอาหารเหมาะสมสำหรับนักท่องเที่ยวที่มีโรคประจำตัว แนะนำสถานที่ท่องเที่ยวพร้อมอาหารท้องถิ่น ช่วยเพิ่มความสะดวกในการเดินทางไปยังที่ไม่เคยรู้จัก และไม่ทราบว่ามีผลกระทบกับสุขภาพอย่างไร

ซึ่งหากพิจารณาจากผลสรุปงานวิจัยด้านระบบสารสนเทศที่สนับสนุนการท่องเที่ยวที่มีอยู่เดิมนั้น จะพบว่าส่วนใหญ่ยังไม่พิจารณาถึงมุมมองด้านกิจกรรมสันทนาการมากนัก คณะผู้วิจัยจึงได้กำหนดขอบเขตปัญหาการวิจัย โดยมุ่งเน้นที่กิจกรรมการช้อปปิ้งรายบุคคลเป็นปัญหาในการวิจัย เนื่องจากเป็นกลุ่มเป้าหมายใหม่ที่มีกำลังทรัพย์และมีการเติบโตขึ้นอย่างต่อเนื่อง

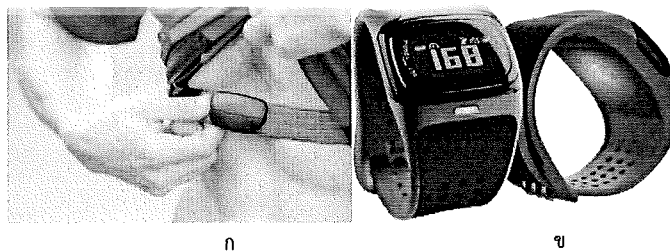
คณะผู้วิจัยตั้งสมมุติฐานว่าในมุมมองของผู้ใช้กิจกรรมรายบุคคลนั้น ส่วนใหญ่มักชื่นชอบเส้นทางช้อปปิ้งที่แตกต่างกันตามลักษณะรูปร่างและความชื่นชอบส่วนบุคคล เช่น ผู้ที่มีรูปร่างผอมเพรียว อาจจะชื่นชอบรูปแบบการช้อปปิ้งไปตามเส้นทางชัน เนื่องจากน้ำหนักน้อย จึงช้อปปิ้งเส้นทางชันได้ ดีกว่ากลุ่มคนที่มีรูปร่างใหญ่ ในทางตรงกันข้ามผู้ที่มีมวลกล้ามเนื้อเยอะ มักชอบเส้นทางที่มีราบเรียบเพราะมีมวลกล้ามเนื้อเป็นปัจจัยที่มีผลต่อความเร็วในทางราบมากกว่า เป็นต้น ดังนั้นผู้ใช้กิจกรรมที่ต้องการท่องเที่ยวไปเปิดประสบการณ์ใหม่โดยการช้อปปิ้งในเส้นทางอื่นๆที่ห่างไกลจาก

ถิ่นฐานของตนเอง จึงต้องการคำแนะนำถึงเส้นทางการขี่จักรยานที่เหมาะสมกับรูปแบบการขี่จักรยานของตนเอง มิใช่เพียงคำแนะนำสถานที่ท่องเที่ยวโดยทั่วไปที่ระบบแนะนำที่ท่องเที่ยวต่างๆ ให้บริการอยู่

ข้อมูลหนึ่งที่ช่วยสนับสนุนการจำแนกรูปแบบการขี่จักรยานรายบุคคลคือ ข้อมูลประวัติการขี่จักรยาน ซึ่งผู้ขี่จักรยานในปัจจุบัน มักใช้อุปกรณ์ไอทีในการบันทึกข้อมูลการขี่จักรยานแต่ละครั้ง โดยอุปกรณ์ไอทีดังกล่าว มีทั้งที่เป็นแบบที่ติดตั้งที่ตัวรถจักรยานหรือแบบสวมใส่ ตัวอย่างอุปกรณ์ไอทีที่ติดตั้งที่ตัวจักรยานได้แก่ Cyclometer (ภาพที่ 1.1 ก) ซึ่งเป็นชุดอุปกรณ์ที่ประกอบด้วยตัวรับรู้ชนิดต่างๆ เช่น ตัวรับรู้ GPS สำหรับข้อมูลความเร็ว พิกัดตำแหน่ง และความลาดชันของเส้นทาง ตัวรับรู้ Thermometer สำหรับข้อมูลอุณหภูมิ และตัวรับรู้ Cadence meter (ภาพที่ 1 ข) สำหรับข้อมูลความถี่ของรอบขา เป็นต้น ส่วนตัวอย่างอุปกรณ์ไอทีแบบสวมใส่ได้แก่ ตัวรับรู้อัตราการเต้นของหัวใจ (ภาพที่ 1.2 ก และ ข) ซึ่งมีรูปลักษณะ 2 แบบ คือแบบสายคาดอก ที่ใช้รับสัญญาณไฟฟ้าจากร่างกายโดยใช้หลักการเดียวกันกับที่ใช้ในการทำ EKG (Walraven, 2010) ในทางการแพทย์ และแบบนาฬิกาข้อมือ ข้อมูลที่ใช้แสงจะทำงานโดยอาศัยการเปลี่ยนแปลงของอัตราการดูดกลืนแสงของผิวหนังซึ่งแปรผันตามจังหวะการเต้นของหัวใจ เป็นต้น



ภาพที่ 1.1 อุปกรณ์ไอทีแบบติดตั้งที่ตัวจักรยาน ก) ตัวรับรู้ Cadence meter ข) Cyclometer



ภาพที่ 1.2 ตัวรับรู้อัตราการเต้นหัวใจ ก) แบบสายคาดอก ข) แบบนาฬิกาข้อมือ

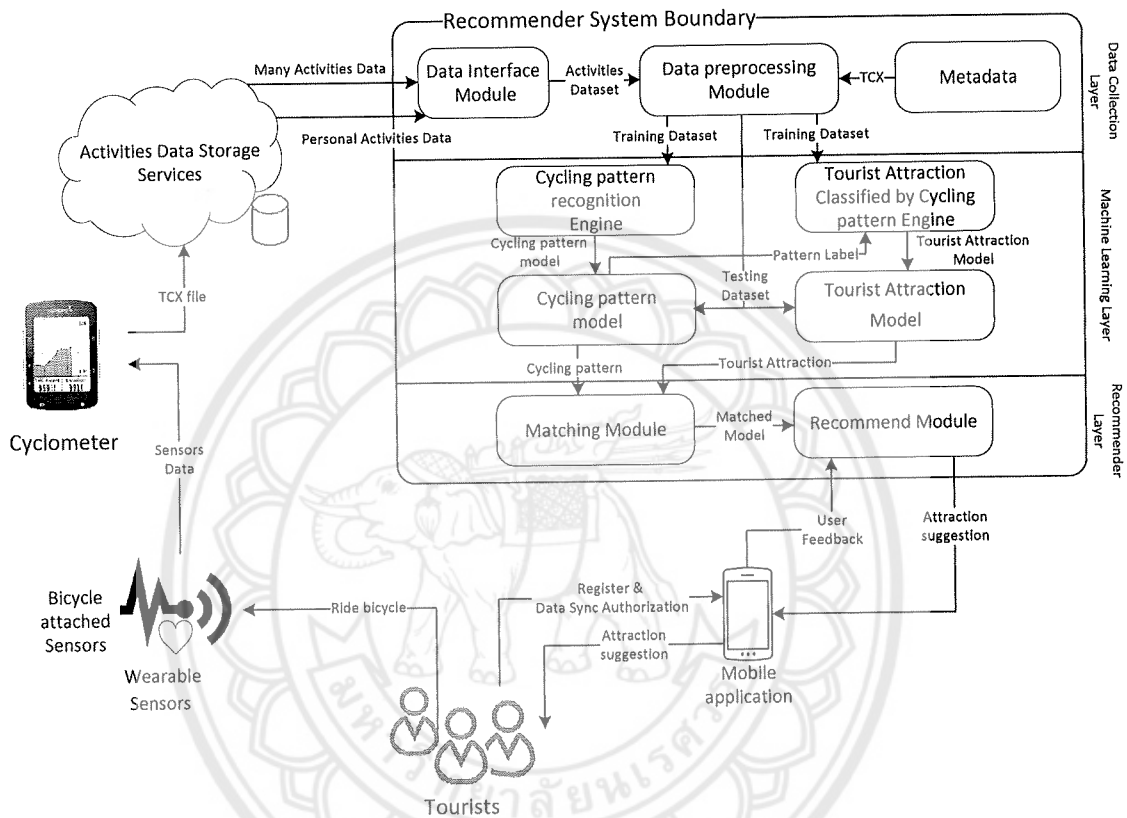
ซึ่งอุปกรณ์ไอทีแบบสวมใส่สามารถบันทึกข้อมูลเหล่านี้แบบทันทีทันใด โดยผู้ใช้จักรยานสามารถจัดเก็บข้อมูลดังกล่าวเข้าสู่ระบบจัดเก็บข้อมูล โดยผู้ใช้จักรยานสามารถเลือกใช้ได้หลากหลาย อาทิเช่น บริการจากผู้ผลิตอุปกรณ์ไอทีสวมใส่โดยตรง และบริการที่เป็นมาตรฐานเปิดให้ผู้ใช้งานอุปกรณ์ไอทีสวมใส่ต่างๆสามารถส่งข้อมูลเข้ามาจัดเก็บไว้ได้ เป็นต้น ดังนั้น บริการจัดเก็บข้อมูลเหล่านี้จึงถือเป็นแหล่งข้อมูลสำคัญที่สามารถเชื่อมโยงเพื่อนำเข้าข้อมูลมาใช้ในการจำแนกรูปแบบการขี่จักรยานเฉพาะบุคคล ซึ่งสามารถประยุกต์ใช้ในการจำแนกกลุ่มและแนะนำเส้นทางที่เหมาะสมกับรูปแบบการขี่จักรยานเฉพาะบุคคล

การวิจัยเพื่อพัฒนาระบบแนะนำการขี่จักรยานเชิงท่องเที่ยวบนพื้นฐานรูปแบบการขี่เฉพาะบุคคล ด้วยเทคโนโลยีอุปกรณ์ไอทีสวมใส่นี้ จึงประกอบด้วยการรวบรวมข้อมูลการขี่จักรยานจากผู้ใช้บริการจัดเก็บข้อมูลการขี่จักรยานต่างๆ เพื่อนำมาจำแนกและจัดกลุ่มรูปแบบการขี่จักรยานที่เฉพาะบุคคลแบบต่างๆ เพื่อเป็นส่วนในการแนะนำตามความต้องการของผู้ใช้ และพิกัดของเส้นทางที่ได้รับคความนิยม โดยใช้เทคนิคการเรียนรู้ของเครื่อง (Machine Learning) ซึ่งผลลัพธ์ที่มีผลการประเมินประสิทธิภาพในการจำแนกและจัดกลุ่มที่ดีที่สุดจะถูกนำมาประยุกต์ใช้ในกระบวนการแนะนำเส้นทางที่เหมาะสมกับการขี่จักรยานเฉพาะบุคคลต่อไป โดยผู้ใช้มีความพึงพอใจต่อระบบที่พัฒนาขึ้นอยู่ในระดับดี

วัตถุประสงค์ของการศึกษา

1. เพื่อพัฒนาระบบการรวบรวมข้อมูลการขี่จักรยานจากอุปกรณ์ไอทีสวมใส่ของนักท่องเที่ยว
2. เพื่อศึกษาและพัฒนาแบบจำลองเพื่อใช้ในการเรียนรู้และรู้จำรูปแบบการขี่จักรยานจาประวัติการขี่จักรยานของนักท่องเที่ยว
3. เพื่อพัฒนาระบบแนะนำเส้นทางท่องเที่ยวให้สอดคล้องกับพฤติกรรมการขี่จักรยานเฉพาะบุคคล
4. เพื่อประเมินความพึงพอใจของนักท่องเที่ยวที่ชื่นชอบการขี่จักรยานที่มีต่อระบบฯ

ทฤษฎีและกรอบแนวคิดของงานวิจัย
วิธีการดำเนินงาน



ภาพที่ 1.3 รูปกรอบแนวคิดโครงการวิจัย

พัฒนาระบบแนะนำเส้นทางท่องเที่ยวให้สอดคล้องกับพฤติกรรมการขี่จักรยานเฉพาะบุคคล เป็นระบบแนะนำชนิดหนึ่งที่ให้ผลลัพธ์เป็นการแนะนำเส้นทางท่องเที่ยวเชิงสันทนการให้แก่ผู้ใช้ โดยใช้รูปแบบในการขี่จักรยานรายบุคคลเป็นแนวทางในการเสนอแนะ โดยระบบดังกล่าวมีส่วนประกอบที่สำคัญ 3 ส่วนได้แก่

1. ส่วนการรวบรวมข้อมูล โดยประกอบด้วยตัวรับรู้ข้อมูลที่เกี่ยวข้องกับการขี่จักรยาน อันได้แก่ อุปกรณ์ไอทีแบบสวมใส่เพื่อตรวจจับสัญญาณชีพจร และตัวรับรู้อื่นๆเพื่อตรวจจับสัญญาณจากอื่นๆ จากการขี่จักรยาน เช่น ความถี่ของรอบขาปั่น ความเร็ว ระยะทาง พิกัดตำแหน่ง ความลาดชันของเส้นทาง อุณหภูมิ เป็นต้น ซึ่งตัวรับรู้จะส่งสัญญาณที่ตรวจจับได้เข้าสู่หน่วยประมวลผลสำหรับจักรยาน

(Cyclometer) ซึ่งทำหน้าที่รวบรวมสัญญาณจาก sensor ต่างๆ แล้วจัดเก็บในรูปแบบโครงสร้างข้อมูลมาตรฐาน โดยในงานวิจัยนี้ใช้ Training Center XML (TCX) ซึ่งเป็นโครงสร้างข้อมูลที่ได้รับ ความนิยมแพร่หลายในการแลกเปลี่ยนข้อมูลระหว่างกลุ่มอุปกรณ์บันทึกข้อมูลออกกำลังกาย เมื่อผู้ใช้ จบการขี่จักรยานและบันทึกข้อมูลการขี่แต่ละครั้ง ข้อมูลสัญญาณต่างๆจากตัวรับรู้จะถูกรวบรวมเป็น ไฟล์ TCX แล้ว upload เข้าสู่บริการจัดเก็บข้อมูลการออกกำลังกาย (Activities Data Storage Services) เพื่อรวบรวมประวัติการขี่จักรยานทั้งหมดที่ผ่านมาของผู้ใช้แต่ละคน

2. ส่วนการสร้างตัวแบบรู้จำรูปแบบการขี่จักรยานรายบุคคล ส่วนนี้ประกอบด้วยส่วนการเชื่อมโยง ข้อมูล (Data Interface Module) เข้ากับแหล่งข้อมูลประวัติการขี่จักรยานรายบุคคลจากบริการ จัดเก็บข้อมูลการออกกำลังกายต่างๆ โดยข้อมูลที่เชื่อมโยงมาได้จะถูก preprocess ให้มีโครงสร้าง ข้อมูล (Metadata) ที่เหมาะสมเพื่อสร้างชุดข้อมูลสำหรับฝึกสอนให้แก่ส่วนการรู้จำรูปแบบการขี่ จักรยานโดยเครื่อง (Cycling pattern recognition engine) โดยผลลัพธ์ในการเรียนรู้ของส่วนการ รู้จำรูปแบบการขี่จักรยานโดยเครื่องคือตัวแบบรูปแบบการขี่จักรยาน (Cycling pattern model) ซึ่ง ตัวแบบรูปแบบการขี่จักรยานที่เครื่องเรียนรู้ได้นั้นจะถูกนำไปใช้ในการจัดกลุ่มของเส้นทางท่องเที่ยวที่ น่าสนใจ (Tourist attraction) ที่เหมาะสมกับรูปแบบการขี่จักรยานรายบุคคล

3. ระบบเสนอแนะเส้นทางท่องเที่ยวที่น่าสนใจ (Recommender System) เป็นส่วนที่น่าผลลัพธ์ ของตัวแบบรูปแบบการขี่จักรยานรายบุคคลและเส้นทางท่องเที่ยวที่น่าสนใจ มาผนวกเข้าด้วยกันเกิด เป็นระบบที่แนะนำเส้นทางขี่จักรยานที่เหมาะสมกับรูปแบบการขี่จักรยานรายบุคคล โดยระบบ เสนอแนะดังกล่าวจะติดต่อกับผู้ใช้งานผ่านแอปพลิเคชันบนโทรศัพท์มือถือ โดยมีความสามารถในการ แนะนำเส้นทางท่องเที่ยวที่น่าสนใจในบริเวณใกล้เคียงกับพิกัดตำแหน่งปัจจุบันของผู้ใช้ หรือพิกัด ตำแหน่งที่ผู้ใช้สนใจ และเมื่อผู้ใช้จักรยานบนเส้นทางที่ระบบแนะนำเสร็จแล้ว ข้อมูลการขี่จักรยาน ดังกล่าวจะถูก sync เข้าสู่ระบบ recommender เพื่อประเมินคุณภาพการแนะนำเส้นทางท่องเที่ยว ในครั้งนั้นๆ แล้วนำไปปรับปรุงวิธีการแนะนำเส้นทางครั้งต่อไป

วิธีการดำเนินการวิจัย

ทำการศึกษาการทำงานของระบบไฟฟ้าขนาดเล็กแยกโดดโดยใช้ตัวควบคุมแบบพีไอ (PI) โดยตัวควบคุมพีไอจำเป็นต้องมีค่าพารามิเตอร์ที่เหมาะสมเพื่อใช้ในการลดค่าการเบี่ยงเบนความถี่ใน ระบบไฟฟ้าขนาดเล็กแยกโดด ดังนั้นผู้วิจัยได้ดำเนินการศึกษาวิธีการหาค่าพารามิเตอร์ที่เหมาะสมของ ตัวควบคุมพีไอด้วยวิธีกลุ่มอนุภาค และใช้โปรแกรม MATLAB สร้างแบบจำลองเชิงเส้นของระบบ ไฟฟ้าขนาดเล็กแยกโดดเพื่อแสดงให้เห็นว่าค่าพารามิเตอร์ที่ได้นั้นลดการเบี่ยงเบนความถี่ลง

ขอบเขตการดำเนินงาน

ขอบเขตด้านข้อมูล

1. ประชากรกลุ่มเป้าหมายคือผู้ซึ่งจักรยาน โดยแบ่งกลุ่มด้วยเพศ และ ช่วงอายุต่างๆ ระหว่าง 15 - 60 ปี
2. ข้อมูลเป็นข้อมูลการขี่จักรยาน โดยมีช่วงเวลาระหว่างปี 2555-2559

ขอบเขตด้านพื้นที่

1. การสำรวจข้อมูลจะทำการสำรวจและเก็บรวบรวมข้อมูลในพื้นที่จังหวัด พิษณุโลก ลำปาง พะเยา เชียงราย และเชียงใหม่

ขอบเขตด้านการทำงานของระบบ

1. ระบบสามารถแนะนำเส้นทางท่องเที่ยวด้วยการรูปแบบการขี่จักรยานรายบุคคล โดยเส้นทาง อยู่บริเวณในพื้นที่ที่สำรวจ ข้อมูล
2. ระบบสามารถแสดงข้อมูลเส้นทางท่องเที่ยวด้วยการขี่จักรยานในภาพรวมและข้อมูลเชิงลึกให้กับผู้ใช้เพื่อประกอบการตัดสินใจในเลือกเส้นทางท่องเที่ยวที่สนใจได้
3. ระบบสามารถแสดงผลประโยชน์ต่อสุขภาพในมุมมองด้านการพัฒนาศักยภาพร่างกายที่ได้จากการท่องเที่ยวแต่ละครั้ง เพื่อเป็นการกระตุ้นให้ผู้ใช้เกิดการปรับเปลี่ยนพฤติกรรมสุขภาพ ในทิศทางที่ดีขึ้นต่อไป

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อสนับสนุนการท่องเที่ยวเชิงสันตนาการด้วยการขี่จักรยาน ตอบสนองต่อความต้องการของกลุ่มนักท่องเที่ยวเชิงสันตนาการที่ขยายตัวเพิ่มขึ้น
2. เพื่อดึงดูดนักท่องเที่ยวชาวไทยและชาวต่างประเทศให้เข้ามาท่องเที่ยวเชิงสันตนาการในประเทศมากยิ่งขึ้น
3. เพื่อกระตุ้นเศรษฐกิจด้านการท่องเที่ยวเชิงสันตนาการด้วยการขี่จักรยาน สนับสนุนให้เกิดการพัฒนาคุณภาพของสินค้าและบริการด้านการท่องเที่ยวเชิงสันตนาการให้มีคุณภาพเพิ่มมากขึ้น

บทที่ 2

ทฤษฎีพื้นฐานและงานวิจัยที่เกี่ยวข้อง

การใช้งานข้อมูลในปัจจุบัน มีข้อมูลเกิดขึ้นจำนวนมากในหลากหลายรูปแบบจากแหล่งข้อมูลหลายๆ แหล่ง เช่น สมาร์ททีวี เว็บไซต์ อีเมล สื่อสังคมออนไลน์ ทำให้ผู้ใช้งานไม่สามารถประมวลผลข้อมูลที่หลากหลายดังกล่าวได้อย่างมีประสิทธิภาพ โดยปัญหาดังกล่าวเรียกว่าข้อมูลเกิน (Information Overload) ซึ่งปัญหาข้อมูลเกินมีผลต่อการตัดสินใจในการคัดเลือกข้อมูลที่ต้องการหรือผู้ใช้งานได้รับข้อมูลที่ไม่ได้เกี่ยวข้องหรือไม่ตรงกับความต้องการ ทำให้ยากต่อการเลือกใช้ข้อมูลหรืออาจทำให้มองข้ามข้อมูลที่สำคัญไปได้ เป็นเหตุให้เกิดความผิดพลาดในการตัดสินใจได้ ซึ่งการนำระบบแนะนำมาใช้ (Recommender System) เป็นวิธีการหนึ่งที่สามารถจัดการข้อมูลจำนวนมากเพื่อประมวลผลและแสดงสารสนเทศที่เกี่ยวข้องกับความต้องการให้แก่ผู้ใช้ได้ (Tulasi K., Nathaniel D. และ Jennifer L., 2017)

ความหมายของระบบแนะนำ

ระบบแนะนำเป็นแอปพลิเคชันที่นำมาใช้ในการจัดการปัญหาข้อมูลเกินได้ โดยสามารถค้นหาและแนะนำข้อมูลที่เกี่ยวข้องกับสิ่งที่ผู้ใช้งานต้องการ ระบบแนะนำสามารถสร้างรายการแนะนำสินค้าเฉพาะบุคคลให้แก่ผู้ใช้งานตามความชอบ และยังสามารถแนะนำสินค้าใหม่ ๆ ที่เหมาะสม โดยลักษณะการทำงานของระบบแนะนำจะใช้ประโยชน์จากคุณลักษณะข้อมูลของสินค้า เช่น สี ขนาด ราคา แหล่งผลิต หรือข้อมูลจากผู้ใช้ เช่น การให้คะแนนความพึงพอใจต่อสินค้าที่เคยซื้อ ความสนใจ ความชอบไม่ชอบ เพศ อายุ และยังมีการใช้ข้อมูลที่ได้จากการสำรวจจากกลุ่มประชากร (Demographics) ปัจจุบันระบบแนะนำได้รับความนิยมและถูกใช้เป็นกลไกในการแนะนำสินค้าและบริการให้แก่ผู้ใช้งานเป็นอย่างมาก เช่น การแนะนำสินค้าของเว็บ Amazon.com การแนะนำภาพยนตร์ของ Netflix การแนะนำรายการข่าวสาร บริการทางการเงิน รวมถึงทวิตเตอร์ นอกจากนี้ระบบแนะนำยังสนับสนุนการทำงานที่หลากหลายทำให้ผู้ใช้ให้บริการต่าง ๆ สามารถเข้าใจความต้องการของลูกค้า และสามารถนำไปใช้งานในการให้บริการเพื่อเพิ่มยอดขาย และตอบสนองความพึงพอใจให้แก่ลูกค้าได้มากขึ้น (Tulasi K., Nathaniel D. และ Jennifer L., 2017)

การให้คะแนนต่อสินค้า

คะแนนของสินค้า (Rating) ในระบบแนะนำนั้น เป็นสิ่งที่สามารถนำมาใช้พิจารณาความชอบ/ไม่ชอบของลูกค้าที่มีต่อสินค้าหรือบริการได้ โดยระบบแนะนำจะนำข้อมูลคะแนนของสินค้าที่เก็บไว้มาพิจารณาในการสร้างรายการแนะนำสินค้าให้แก่ผู้ใช้งาน โดยลักษณะของการให้คะแนน มี 2 รูปแบบ ได้แก่ (Sirinart Tangruamsub, 2019)

1. คะแนนแบบชัดเจน (Explicit Rating)

คะแนนแบบชัดเจนนั้น ได้มาจากการที่ผู้ใช้งานเป็นผู้ให้คะแนนต่อสินค้าหรือบริการโดยตรง ด้วยการให้คะแนนผ่านช่องทางที่จัดเตรียมไว้ให้ เช่น คะแนนความชอบ 8/10 หมายถึง ผู้ใช้งานชื่นชอบ/พึงพอใจต่อสินค้ามาก และคะแนนความชอบ 1/10 หมายถึงลูกค้าไม่ชอบ/ไม่พึงพอใจต่อสินค้า โดยคะแนนแบบชัดเจนนั้น ทำให้สามารถทราบได้ทันทีว่าผู้ใช้งานชอบ หรือไม่ชอบสินค้า และยังสามารถบอกระดับความชอบได้จากคะแนนโดยตรง ดังแสดงในภาพที่ 1 แต่ผู้ใช้งานส่วนใหญ่ ไม่ได้ซื้อสินค้าหรือบริการทั้งหมดที่มีในระบบ หรือผู้ใช้งานบางคนไม่ได้ให้คะแนนต่อสินค้าที่ซื้อ ดังนั้น หากระบบมีสินค้าหรือบริการจำนวนมากอาจมีสินค้าจำนวนที่ไม่ถูกให้คะแนนจากผู้ใช้งาน ซึ่งอาจก่อให้เกิดปัญหาความเบาบางของข้อมูลได้ (Sparsity)



Long-lasting power

1.5 volt power for everyday devices

10-year shelf life

Anti-leak, anti-corrosion design

AmazonBasics 48 Pack AA High-Performance Alkaline Batteries, 10-Year Shelf Life, Easy to Open Value Pack

Visit the AmazonBasics Store

ดูข้อมูลเพิ่มเติม 67,660 ratings
364 คำถามที่พบบ่อย

Amazon's Choice for "amazonbasics"

ภาพที่ 2.1 การให้คะแนนต่อสินค้าแบบชัดเจน

ที่มา : (Amazon, 2020)

2. คะแนนแบบโดยนัย (Implicit Rating)

คะแนนแบบโดยนัย ได้มาจากช่องทางอื่นนอกเหนือจากการให้คะแนนโดยตรงจากผู้ใช้งาน เช่น จำนวนครั้งที่คลิกเข้าไปดูสินค้า การตัดสินใจเลือกซื้อสินค้า หรือการเลือกสินค้านั้นไว้เป็นสินค้าที่ชื่นชอบ หรือสินค้าที่อยากได้ ซึ่งคะแนนแบบโดยนัยนี้ มักมีปริมาณข้อมูลที่มากกว่าคะแนนแบบชัดเจน

แจ้ง แต่คะแนนแบบโดยนัยไม่สามารถเก็บข้อมูลตอบกลับเชิงลบได้ (Negative Feedback) เช่น หากผู้ใช้งานไม่ได้คลิกดูสินค้า ไม่สามารถบอกได้ว่า ผู้ใช้งานไม่ชอบสินค้า แต่อาจเป็นเพราะมีสินค้าชนิดนี้ อยู่แล้ว หรืออาจเพราะไม่เห็นสินค้าชนิดนี้ถูกแนะนำขึ้นมา และการพิจารณาคะแนนโดยนัยจากการเลือกซื้อสินค้านั้น อาจไม่สามารถระบุความชอบที่แท้จริงของผู้ใช้งานได้ ในกรณีผู้ใช้งานอาจเลือกซื้อสินค้านั้นให้แก่ผู้อื่น

ประเภทของระบบแนะนำ

ระบบแนะนำ ถูกแบ่งประเภทดังนี้ (Charu C, 2016; Tulasi K., Nathaniel D. และ Jennifer L., 2017)

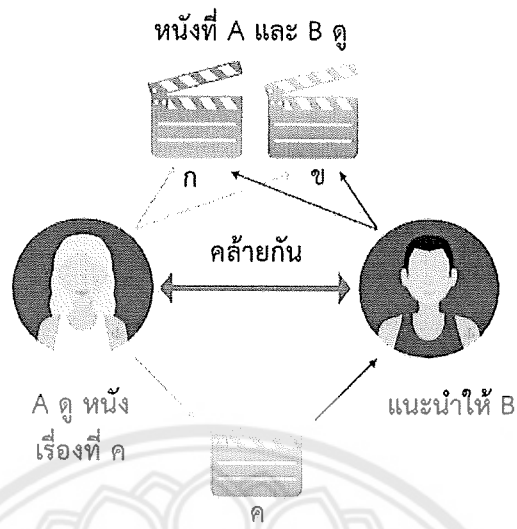
1. Collaborative Filtering Recommender System

ระบบแนะนำแบบ Collaborative Filtering Recommender System นั้น จะทำการแนะนำรายการสินค้าโดยพิจารณาจากข้อมูลของผู้ใช้งานหลาย ๆ คน เพื่อคาดเดาความชื่นชอบ (Rating) ของผู้ใช้งานคนอื่น โดยอาศัยแนวคิดคือ ผู้ใช้งานที่มีความคล้ายคลึงกันจะถูกจัดไว้ในกลุ่มเดียวกัน และมักจะชอบรายการสินค้าที่คล้ายกัน ดังนั้น ระบบจะแนะนำสินค้าให้แก่ผู้ใช้งานอีกคนในกลุ่มเดียวกันด้วย และจะใช้คะแนนของสินค้าเป็นข้อมูลหลักในการพิจารณาในการสร้างรายการแนะนำ โดยระบบแนะนำแบบ Collaborative Filtering Recommender System สามารถแบ่งวิธีการทำงานได้ 2 แบบ ได้แก่ memory-based และ model-based (Charu C, 2016; Rocca, 2019; Sirinart Tangruamsub, 2019)

1.1 memory-based

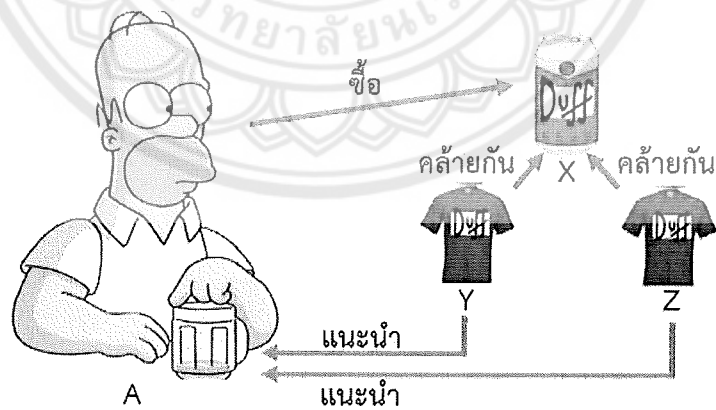
วิธีการทำงานแบบ memory-based เป็นวิธีการสร้างรายการแนะนำโดยการคำนวณความคล้ายคลึงของข้อมูลที่เก็บไว้ในระบบเพื่อสร้างรายการแนะนำ โดยสามารถทำได้ 2 วิธี ได้แก่

1.1.1 user-based collaboration filtering จะทำการคาดเดาคะแนนที่ให้ต่อสินค้าในระบบโดยพิจารณาจากผู้ที่มีความคล้ายคลึงกัน เช่น หากผู้ใช้งาน A และ B ชื่นชอบหนังสือที่ ก และ ข เหมือนกัน ระบบจะพิจารณาว่า A และ B อยู่ในกลุ่มเดียวกัน เมื่อ A ดูหนังสือที่ ค ระบบจะแนะนำหนังสือที่ ค ให้แก่นาย B ด้วย ดังแสดงในภาพที่ 2.2



ภาพที่ 2.2 การแนะนำสินค้าโดยวิธีการทำงานแบบ user-based collaboration filtering
ที่มา : ปรับปรุงจาก (Emma Grimaldi, 2018)

1.1.2 item-based collaboration filtering จะทำการคาดเดาคะแนนที่ให้ต่อสินค้าในระบบโดยการพิจารณาจากสินค้าที่มีความคล้ายคลึงกับสินค้าที่ผู้ใช้งานเคยซื้อในอดีต เช่น หากผู้ใช้งาน A ชื่นชอบสินค้า X และสินค้า X มีความคล้ายคลึงกับสินค้า Y และ Z ระบบจะแนะนำสินค้า Y และ Z ให้แก่ผู้ใช้งาน A ด้วย ดังแสดงในภาพที่ 2.3



ภาพที่ 2.3 การแนะนำสินค้าโดยวิธีการทำงานแบบ item-based collaboration filtering
ที่มา : ปรับปรุงจาก (Handaru, 2016)

ในการเก็บข้อมูลเพื่อใช้ในการสร้างรายการแนะนำด้วยวิธีการแบบ memory-based นั้น จะเก็บข้อมูลในรูปแบบเมตริกซ์ และทำการคำนวณหาความคล้ายกันของผู้ใช้งานหรือสินค้า ดังแสดงในภาพที่ 2.4 เป็นการเก็บข้อมูลคะแนนที่ผู้ใช้งาน A-E ให้แก่โรงแรมหมายเลข 1-5 และเมื่อเปรียบเทียบความคล้ายแล้ว พบว่า ผู้ใช้งาน B มีความคล้ายกับผู้ใช้งาน D ดังนั้น จะสามารถคาดเดาคะแนนที่ผู้ใช้งาน D ต่อโรงแรม 4 ได้

	1	2	3	4	5
A	7.00	6.00	6.80		8.00
B	8.30	6.00	8.00	7.40	7.00
C		7.00		8.00	9.50
D	8.00	6.00	7.75	?	7.50
E		6.50	5.50	7.75	

ภาพที่ 2.4 การเปรียบเทียบความคล้ายเพื่อคาดเดาการให้คะแนนต่อสินค้า
ที่มา : (Sirinart Tangruamsub, 2019)

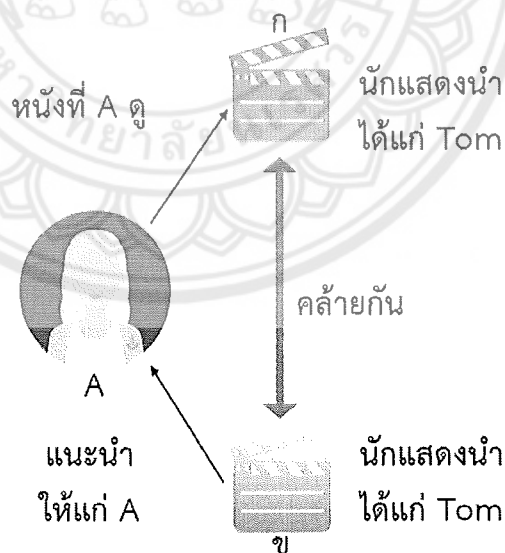
วิธีการทำงานแบบ memory-based นั้นทำงานจากการใช้ข้อมูลคะแนนที่เก็บไว้และนำมาสามารถสร้างรายการแนะนำได้โดยการใช้ฟังก์ชันในการคำนวณความคล้ายของข้อมูลนั้นได้โดยตรง เช่น ผู้ใช้งาน หรือสินค้า แต่เนื่องจากวิธีการนำข้อมูลคะแนนมาใช้ แต่ผู้ใช้ส่วนมากไม่ได้ซื้อสินค้าทั้งหมดที่มีในระบบ หรือไม่ได้ทำการให้คะแนนต่อสินค้าครบทุกรายการ ทำให้เกิดปัญหาความเบาบางของข้อมูล (Sparsity) ที่จะนำมาใช้เป็นข้อมูลในการคำนวณความคล้ายเพื่อสร้างรายการแนะนำ และหากมีข้อมูลจำนวนมาก อาจเกิดปัญหาในการใช้งานทรัพยากรและเวลาในการคำนวณความคล้ายของข้อมูลเนื่องจากข้อมูลที่เก็บไว้ในระบบมีจำนวนมาก (Scalability) นอกจากนี้ ในกรณีของลูกค้าหรือสินค้ารายการใหม่ จะยังไม่มีข้อมูลการให้คะแนน ระบบจะไม่มีข้อมูลมากพอที่จะนำมาสร้างรายการแนะนำได้ (Cold-start)

1.2 model-based

วิธีการทำงานแบบ model-based เป็นวิธีการสร้างรายการแนะนำโดยการใช้หลักการเรียนรู้ของเครื่องจักร (Machine Learning) และการทำดาต้าไมนิ่ง (Data Mining) มาสกัดหารูปแบบในการให้คะแนนของผู้ใช้งานต่อสินค้า และนำรูปแบบที่ได้ไปใช้ในการทำนายคะแนนที่ผู้ใช้มีต่อสินค้าอื่น เช่น เทคนิคการเรียนรู้ของเครื่องจักร จะนำข้อมูลมาทำการสอนและเรียนรู้รูปแบบการให้คะแนนของผู้ใช้งานต่อสินค้า และนำรูปแบบที่ได้จากการเรียนรู้มาทำนายค่าคะแนนของสินค้าที่เหลือเพื่อสร้างรายการแนะนำให้แก่ผู้ใช้งาน

2. Content-based Filtering Recommender System

ระบบแนะนำแบบ Content-based Filtering Recommender System นั้น ไม่ได้พิจารณาจากข้อมูลจากผู้ใช้งานคนอื่น ๆ แต่จะทำการแนะนำรายการสินค้าโดยพิจารณาจากข้อมูลในอดีตของผู้ใช้งานคนนั้นโดยตรงว่าเคยซื้อสินค้าหรือบริการใดบ้าง โดยใช้ข้อมูลคุณลักษณะของสินค้าที่เคยซื้อในอดีต เช่น ขนาด สี ราคา ประเภท เป็นต้น มาทำการพิจารณาแนะนำสินค้าที่มีคุณลักษณะคล้ายกับสินค้าที่ผู้ใช้งานเคยซื้อ เช่น ผู้ใช้งาน A เคยดูหนังเรื่อง ก ที่มีนักแสดงนำชื่อ Tom ระบบแนะนำจะหนังเรื่อง ข ที่มีนักแสดงนำชื่อ Tom ให้แก่ A ด้วย ดังแสดงในภาพที่ 2.5 (Charu C, 2016; Sirinart Tangruamsub, 2019)



ภาพที่ 2.5 การทำงานของระบบแนะนำแบบ Content-based Filtering Recommender System

ที่มา : ปรับปรุงจาก (Emma Grimaldi, 2018)

ระบบแนะนำ Content-based Filtering Recommender System สามารถทำงานได้ดีในการแนะนำรายการสินค้าใหม่ให้แก่ผู้ใช้งานถึงแม้ว่าสินค้ารายการนั้นจะไม่มีคะแนนจากผู้ใช้งานในอดีต เนื่องจากระบบแนะนำแบบนี้สามารถพิจารณาความสัมพันธ์ของคุณลักษณะของสินค้าใหม่ที่คล้ายคลึงกับสินค้าที่มีคะแนนอยู่แล้วเพื่อนำมาสร้างรายการแนะนำได้ แต่เนื่องจากการทำงานของระบบแนะนำแบบ Content-based Filtering Recommender System นั้น พิจารณาข้อมูลของสินค้าจากคุณลักษณะของสินค้าที่ผู้ใช้งานเคยซื้อในอดีต ดังนั้น สินค้าที่ไม่ได้มีคุณลักษณะดังกล่าว จะไม่ถูกแนะนำให้แก่ผู้ใช้งานซึ่งส่งผลให้รายการสินค้าที่แนะนำให้แก่ผู้ใช้งานนั้นไม่แตกต่างจากสินค้าที่ผู้ใช้งานเคยซื้อมากนัก ทำให้รายการแนะนำนั้นขาดความหลากหลาย (Diversity) นอกจากนี้ระบบแนะนำจะไม่สามารถแนะนำสินค้าให้แก่ผู้ใช้งานรายใหม่ได้เนื่องจากไม่มีประวัติการซื้อสินค้าเก็บไว้ในระบบ (Charu C, 2016; Juthaporn Vipatpakpaiboon, 2019)

3. Knowledge-based Recommender System

ระบบแนะนำแบบ Knowledge-based Recommender System เป็นระบบแนะนำที่ไม่ได้ใช้ข้อมูลคะแนน (Rating) ของสินค้าในการสร้างรายการแนะนำ แต่ทำงานโดยการใช้ฐานองค์ความรู้เฉพาะด้าน และนำมาสร้างกฎความสัมพันธ์ของสินค้าหรือบริการ และสร้างรายการแนะนำจากคุณลักษณะของสินค้าหรือบริการนั้น ๆ ที่ตรงกับความต้องการของผู้ใช้งาน โดยใช้คำค้น และเงื่อนไขในการค้นหาของผู้ใช้งานไปพิจารณาร่วมกับกฎที่สร้างขึ้นจากฐานองค์ความรู้ จากนั้นระบบจะทำการแนะนำสินค้าที่ตรงกับคำค้นที่ผู้ใช้งานค้นหา และสินค้าที่มีคุณลักษณะคล้ายกับสิ่งที่ผู้ใช้งานต้องการ (Tulasi K., Nathaniel D. และ Jennifer L., 2017) เช่น ผู้ใช้งานสืบค้นข้อมูลบ้านพัก โดยการใส่ข้อมูลที่ต้องการได้แก่ ราคา จำนวนห้องน้ำ จำนวนห้องนอน ขนาดพื้นที่ จากนั้นเว็บไซต์จะแสดงรายการบ้านตามรายการค้นหา และยังสามารถแนะนำรายการบ้านที่มีคุณลักษณะอื่นเพิ่มเติมให้แก่ผู้ใช้งานได้ โดยระบบแนะนำแบบ Knowledge-based Recommender System เหมาะสมกับการซื้อสินค้าที่ไม่ได้ซื้อเป็นประจำ เช่น บ้าน รถ บริการสินเชื่อ ซึ่งสินค้าประเภทนี้จะมีปริมาณข้อมูลคะแนนไม่มากนัก และสินค้าบางประเภทอาจมีช่วงอายุที่สั้น ไม่เหมาะแก่การนำมาสร้างรายการแนะนำ เช่น รถยนต์ เมื่อเวลาผ่านไป รถยนต์รุ่นที่แนะนำให้แก่ผู้ใช้งานนั้นอาจเลิกผลิตไปแล้ว เป็นต้น และเนื่องจากระบบแนะนำประเภทนี้ไม่ได้ใช้ข้อมูลคะแนนในการสร้างรายการแนะนำ จึงทำให้หลีกเลี่ยงปัญหาในกรณีของลูกค้านำหรือสินค้ารายการใหม่ได้ (Cold-start) (Charu C, 2016; Jackson Wu, 2019)

4. Hybrid Recommender System

ระบบแนะนำแบบ Hybrid Recommender System เป็นระบบแนะนำที่ผสมผสานวิธีการทำงานของระบบแนะนำตั้งแต่ 2 วิธีขึ้นไปใช้ร่วมกันในการสร้างรายการแนะนำ โดยมีวัตถุประสงค์เพื่อลดข้อจำกัดของระบบแนะนำในแบบต่าง ๆ เช่น การนำเทคนิคการเรียนรู้ของเครื่องจักร เพื่อแก้ปัญหาคอมพิวเตอร์ของข้อมูล โดยทำการทำนายคะแนนของสินค้าร่วมกับวิธีการแบบ Content-based Filtering Recommender System เพื่อลดปัญหาที่ระบบจะไม่แนะนำสินค้าที่ไม่มีคะแนน โดยการพิจารณาแนะนำสินค้าจากคุณลักษณะของสินค้าร่วมกันเป็นต้น (Tulasi K., Nathaniel D. และ Jennifer L., 2017



Chapter III

Trajectory Clustering by GPS tracking dataset using QuickBundles

Trajectory clustering is considered as analytical methods in trajectory data mining for many modern applications such as a tourist recommendation system. Our research project is a route recommending system for cycling tourists, and we aim to retrieve all their route patterns by adopting GPS-based trajectories dataset. This dataset is informative and high volume, thus, we need clustering methods that are effective and fast in computational time. This paper proposes methods to collect, preprocess, and analyze the GPS tracking data from the existing online service for cyclists. QuickBundles is the primary technique we use to allocate route items into clusters, and also propose cluster validation methods based on silhouette coefficient. The results have shown that QuickBundles has the capability of clustering the trajectory data in terms of computational speed and validity. This approach can benefit online systems, by recommending personalized routes for the cycling tourists.

Problem Statements and Preliminaries.

Trajectory data is recorded in different forms according to the types of sensors, objects movement, or tracking purposes. In this paper, we focused on geographical trajectory data, which is represented as a sequence of n points in geographical space as the following equation:

$$T = (t_1, t_2, \dots, t_n) \quad (1)$$

where t_i is a tracking point which is represented as a combination of coordination and timestamp that forms the sub equation:

$$t_i = (x_i, y_i, z_i, ts_i) \quad (2)$$

where x_i , y_i , z_i , and ts_i are latitude, longitude, altitude, and timestamp, respectively. This data sequence is an essential standard for several data schemes in GPS tracking devices, while some of the data schemes add other relevant information, such as velocity and cumulative distance.

1. Dataset. The cycling workout data that contains GPS track points were explored and selected from online data services, <https://ridewithgps.com/>. This research used selection criterion as follows:

- 1) The selected data must be recorded as TCX file.
- 2) Every Track point in the selected files must have altitude data.
- 3) The route length in the selected files must range between 3 and 180 km.

After previous selection criteria were applied, selected files were converted into raw dataset using a GPSTrace application [8]. However, this data had defects such as missing values and/or sensor reading errors. Therefore, preprocessing is required to deal with the defects in the raw dataset.

2. Data Preprocessing. The dataset usually contains some errors and/or outliers is caused by the hardware. Therefore, to rectify these problems, the dataset was preprocessed by the following methods:

- 1) Unifying the GPS track point's time intervals to 1 second.
- 2) To correct the outlier and missing values, we used the Linear Kalman Filter technique [9].
- 3) All elevation data was altered by resetting the initial track points to zero.

3. Cluster Analysis using QuickBundles. QB is a clustering technique for classifying routes by their elevation profile, and its tractography simplification of a complex Diffusion MRI dataset provides informative clusters in seconds. The tractography is sequential of points in R^3 can be categorized as a form of trajectory data. Therefore, QB is a clustering technique that may also be able to clusters trajectory data. The primary reason why QB successfully classifies trajectory data is that it uses the simple symmetric distance function, which is called the minimum average direct-flip (MDF) distance as equation 3 to 5.

$$d_{\text{direct}}(s, t) = d(s, t) = \frac{1}{K} \sum_{i=1}^K |s_i - t_i| \quad (3)$$

$$d_{\text{flipped}}(s, t) = d(s, t^F) = d(s^F, t) \quad (4)$$

$$\text{MDF}(s, t) = \min(d_{\text{direct}}(s, t), d_{\text{flipped}}(s, t)) \quad (5)$$

where a trajectory $s = [s_1, s_2, \dots, s_k]$ consists k sequential tract points. s^F is a flipped version on s where $s^F = [s_k, s_{k-1}, \dots, s_2, s_1]$. $|x-y|$ denotes the Euclidean distance between two points x and y . the direct distance $d_{\text{direct}}(s, t)$ between two trajectories s and t is the mean of Euclidean distance between corresponding track points. The main advantages of the MDF distance supports QB to have substantial high performance by taking into account both direct and flipped trajectories which are

demonstrated in the Computation Time section.

4. Cluster Validation. Since our work is unsupervised, clustered data without labels are unknown, therefore, cluster validation is considered an essential method to confirm the clustering technique by using the silhouette coefficient [10]. Let i be any object in the clustering and A its corresponding cluster. Then

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} \Delta(i, j) \quad (6)$$

measures the average distance of i to all other objects in cluster A . We then compute each cluster $C \neq A$

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} \Delta(i, j) \quad (7)$$

to quantify the distance to other ones. The minimum value is,

$$b(i) = \min_{C \neq A} d(i, C) \quad (8)$$

that gives the distance of i to the second-best cluster, therefore, the silhouette value $s(i)$ of i is then defined as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (9)$$

This value in the range of $[-1; 1]$ quantifies how well object i fits into cluster A . If the $s(i)$ is negative, the object is placed in the wrong cluster, and if it is around 0, the object is between the clusters, while if it is positive, the object is placed in the correct cluster.

Experiment.

In this section, the preprocessed dataset that is a result of methods in prior sections is used for route clustering. The experiment goal is to classify the routes dataset according to their elevation profiles by applying QB.

Since the length of the trajectory item possibly influence the QB's result [7], the route data set was pre-classified by length. Therefore, the differentiation of trajectory had less affect. We use k-Means to classify routes into 3 groups: short-distance, medium-distance, and long-distance ones. After trajectory data items were grouped by their route length, each item in each group had their altitude profiles classified using QB clustering.

However, the primary parameters that influence the validity of QB clustering are their distance threshold by varying their range by 10, 20, 50, 100, 200 mts, and vary the number of clusters from 2 to 6. To evaluate the validity of clustering, we applied the silhouette coefficient [10] to each group with similar route lengths. We aimed to find the best parameter schemes which presented the best results by clustering validity. We focused on the parameter scheme with a high average silhouette coefficient.

Table 3 Average silhouette coefficient for classification of a medium-distance group

QB Threshold	Number of clusters				
	2	3	4	5	6
20	0.6814	0.5826	0.4564	0.3317	0.2557
30	0.6814	0.5826	0.4564	0.3317	0.3347
50	0.6814	0.5826	0.4564	0.4709	0.4459
100	0.6814	0.5935	0.5268	0.5425	0.5740
200	0.6991	<u>0.7034</u>	0.7034	0.7034	0.7034

Table 4 Average silhouette coefficient classification of a long-distance group

QB Threshold	Number of clusters				
	2	3	4	5	6
20	0.3772	0.2860	0.3406	0.2590	0.2212
30	0.3772	0.2860	0.3406	0.2590	0.2212
50	0.3772	0.2860	0.3406	0.2590	0.2212
100	0.3772	0.2860	0.4229	0.2590	0.2212
200	0.3772	<u>0.4407</u>	0.4229	0.4229	0.4229

For the short-distance group, the silhouette coefficient score corresponds to the varied QB thresholds and the varied number of clusters, are demonstrated in Table 2. When the QB threshold is 200, the average silhouette coefficient value for each cluster is higher than all other parameters. The lowest number of clusters is preferred because it classifies the items effectively, while minimizing computation time by comparing items distance against a lesser number of centroids. Therefore, we considered the QB threshold =200 and the number of clusters = 2 as the optimal parameters for the short-distance group.

For the medium-distance group, the silhouette coefficient score corresponds to the varied QB threshold and the varied number of clusters, which are demonstrated in Table 3. The average silhouette coefficient for QB threshold = 200 in all clusters are the highest among other parameter configurations. We selected QB threshold = 200 and the number of clusters = 3 as the best parameter configuration because it is the lowest number of clusters that have the highest average silhouette coefficient values.

For the long-distance group, as illustrated in Table 4, the parameter configuration that has the highest average silhouette coefficient consists of the QB threshold at 200 and 3 clusters.

3. Route Clustering.

After we achieved the optimal parameter for QB clustering, we applied it to the data set for all groups, and were divided according to their length and elevation profile.

For the short-distance group, 303 routes were divided into 2 clusters according to their elevation profiles, and the centroids of the clusters that correspond to each category are illustrated in Figure 1A.

Short routes elevation in the 1st cluster range between -227.88 to 316.43 meters above sea-level while another cluster have their routes elevation range between 9.10 to 836.69 meters above sea-level, as illustrated in Figure 3.2A. It shows that most of routes elevation data in 1st cluster are relative closer to each other at sea-level while routes elevation in 2nd cluster are distributed among above sea-level altitude.



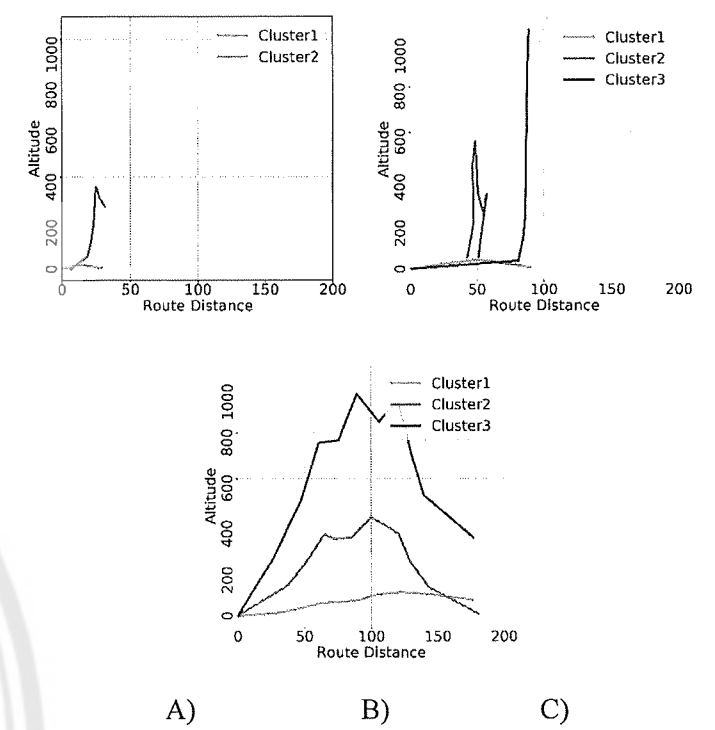


Figure 3.1. A) Cluster centroids of short-distance groups, B) Cluster centroids of medium-distance groups, C) Cluster centroids of long-distance groups

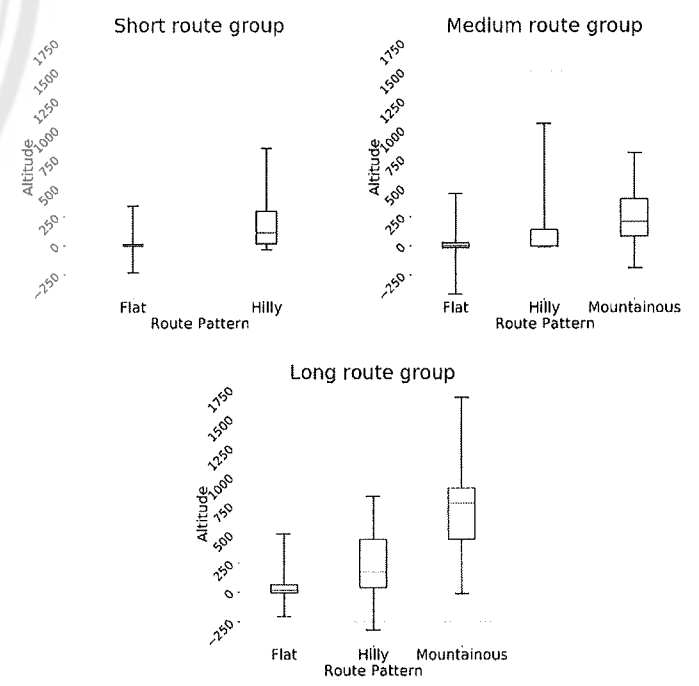


Figure 3.2. The box plot of routes elevation by each cluster in the short distance group

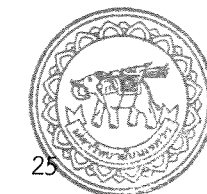
For the medium-distance group, 124 routes were divided into 3 clusters according to their elevation profiles, and the centroids of the clusters correspond to each category, are illustrated in Figure 3.1B.

The 1st cluster of the medium-distance group consists of 89 routes with their altitudes ranging from -417.80 to 429.40 meters. The 2nd cluster has 34 routes and altitudes ranging from -181.56 to 765.33 meters above sea-level. The last cluster consists of only 1 route, and its altitude ranges between -9.11 to 836.69 meters. As illustrated in Figure 3.2B, routes in the 1st cluster mostly remain in sea-level while routes in the 2nd cluster distributes on higher altitude. Finally, the route elevation in the last cluster distributes among limited range of above sea-level but consists the highest altitude point in comparison to other routes.

In the long-distance group, 54 routes were divided into 3 clusters according to their elevation profiles, and the centroids that corresponded to each category, are illustrated in Figure 1C.



๒ ๐๗
๑๐๖๕
๑๗๖๘
๒๕๖๕
๑๐๔๘๔๙๗



สำนักหอสมุด

The 1st cluster consists of 24 routes with their altitudes ranging from -319.14 to 818.41 meters. The 2nd cluster consists of 23 routes with varying altitudes between -191.80 to 481.48 meters. The 3rd cluster consists of 7 routes with altitude ranges from 0 to 1576.68 meters. Route elevation distribution in each clusters were illustrated in Figure 2C. The 2nd cluster consists routes that mostly locate at sea-level while others consists routes that locate above sea-level. However, routes in the 1st cluster locate at lower altitude than routes in the 3rd cluster.

Discussion and Conclusion.

The QB technique has shown that it is an effective clustering method for GPS datasets. QB threshold and the number of clusters, are the primary parameters that influence the validity of route clustering based on their elevation profile. The centroid of each cluster showed some distinctive patterns that can be described as,

1) The flat route pattern. Route elevation profiles in this cluster usually show sustained horizontal levels with some minor changes that raises the lowest level in comparison to other clusters.

2) The hilly route pattern. The elevation profile of routes in this cluster share a similar pattern which is a combination of flat and hilly terrain, and the raised altitude is higher than the 1st cluster but lower than the 3rd one.

3) The mountainous route pattern. Overall, routes in this pattern consist of high altitudes that are greater than either of the others. This trajectory pattern usually start at a low altitude, but does not define the finish elevation.

The description of classified routes conforms to the cycling expert's opinion that categorizes the routes into 3 types, flat, hilly, and mountainous. Therefore, these proposed techniques have the capabilities to retrieve route patterns by classifying them according to elevation automatically. Our methods are considered as a useful tool for recommending routes to cyclists that are similar to their preferences.

In addition, the reason for QB's high performance is the algorithm's complexity. The best case is when QB's complexity is linear time $O(N)$ with the number of items N while the worst-case is $O(N^2)$ when every cluster contains only 1 item [7]. The average case is $O(MN)$, where M is the number of clusters, however, M is typically lower than N ; thus, we can neglect it and denote it as $O(N)$ because it is common in the complexity theory. In this experiments, QB takes 250 milliseconds to analyze and classify the route elevations and total distance of the entire dataset. The processing was done on Intel® Core™ i5-8350U CPU at 1.8GHz, 8GB of RAM, and the Windows 10 Home Edition 64bit.

This study proposed methods to cluster trajectory data by using QuickBundles as the primary technique. The selected data was recorded by GPS-devices and stored on open-access services, which we collected using the defined criteria. The collected data was preprocessed to correct the missing values, outliers, and error sensor readings.

We classified the routes by performing the following 2 steps; 1) grouping routes into 3 groups by their total distance using k-Means, 2) classifying routes by their elevation profiles using QuickBundles. The classification results were measured the clustering validity by using the silhouette coefficient. QuickBundles successfully cluster route data according to their elevation profiles. The results of the proposed techniques agreed with the expert's opinions. These results show that QuickBundles are effective methods for retrieving route patterns from GPS stored data, and recommending them to cyclists.

The suggested further research can be enrich route patterns with additional features, such as the route's turns and route's terrain. These enriched route patterns can be used to proposed tourism recommendation systems for different tourists who emphasize the routes by different route's characteristics.



บทที่ 4

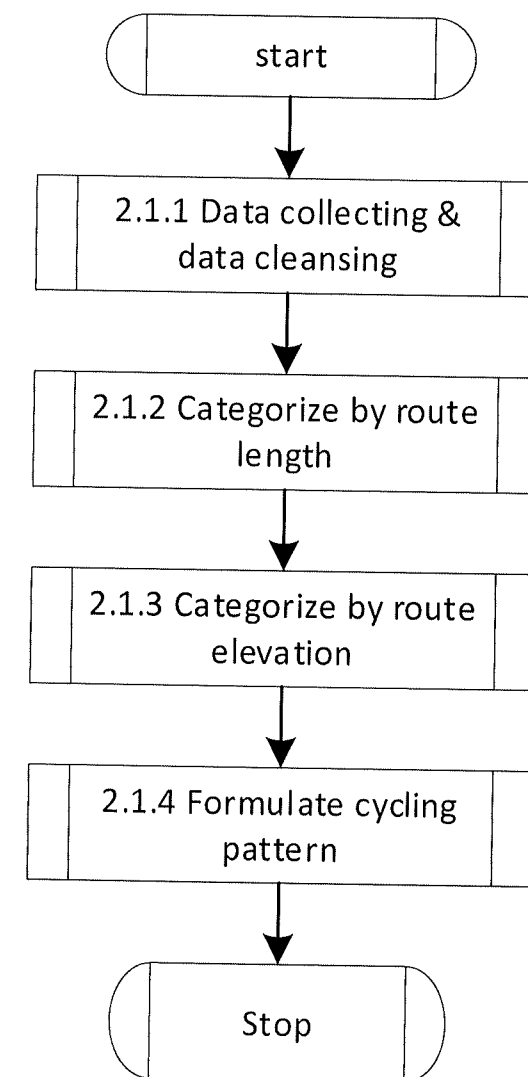
การแนะนำเส้นทางในการขี่จักรยานเฉพาะบุคคล

หลังจากได้กระบวนการในการจัดกลุ่มเส้นทางจักรยาน โดยใช้รูปแบบของความสูงของเส้นทางเป็นตัวแบบในการจัดกลุ่มดังที่ได้กล่าวในบทที่ผ่านมา เราสามารถนำผลลัพธ์ของกระบวนการดังกล่าวมาใช้ในการแนะนำเส้นทางในการขี่จักรยานเฉพาะบุคคลต่อไป โดยในบทนี้ จะกล่าวถึงกระบวนการต่างๆ ในการเลือกเส้นทางจักรยานที่เหมาะสมกับคุณลักษณะของแต่ละบุคคล ดังมีรายละเอียดที่จะกล่าวในหัวข้อต่างๆ ได้แก่ วิธีการทดลอง ผลการทดลอง อภิปรายผล และสรุปผล

การจัดเตรียมข้อมูลผู้ขี่จักรยาน

ในโครงการวิจัยนี้ ต้องการแนะนำเส้นทางที่เหมาะสมกับพฤติกรรมการขี่ของผู้ใช้ จึงจำเป็นต้องรวบรวมประวัติการขี่จักรยานของผู้ใช้ ว่ามีการขึ้นเส้นทางรูปแบบใด โดยใช้ตัวแบบที่ได้จากการจัดกลุ่มเส้นทางในบทก่อนหน้านี้ มาช่วยวิเคราะห์ว่า นักขี่จักรยานแต่ละคนมีพฤติกรรมในการขี่จักรยานในเส้นทางแต่ละกลุ่มเป็นสัดส่วนเท่าใด จากนั้นจึงนำมาสร้างเป็นข้อมูลพฤติกรรมของนักขี่จักรยานรายบุคคลต่อไป ซึ่งมีรายละเอียดดังภาพที่ 4.1



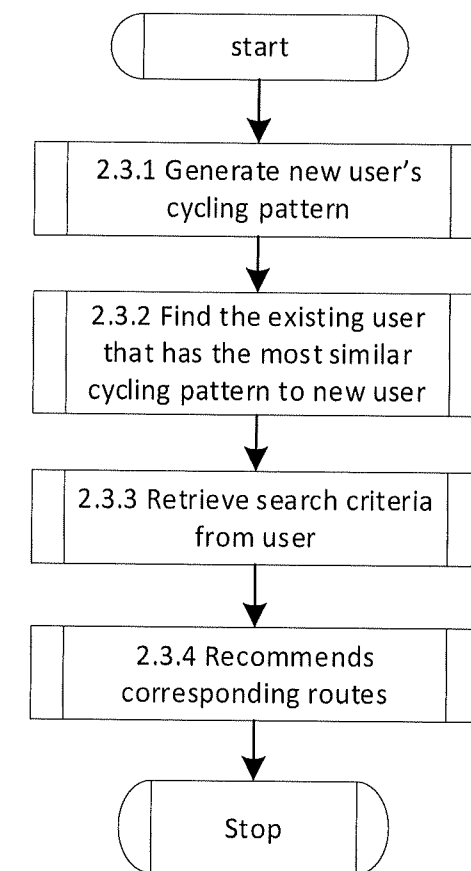


ภาพที่ 4.1 ลำดับการทำงานของกระบวนการจัดเตรียมข้อมูล

1. รวบรวมประวัติการขี่ของนักขี่จักรยานเฉพาะบุคคลจากแหล่งข้อมูลที่เปิดเผยได้เช่น Strava.com, Garmin Connect เป็นต้น โดยข้อมูลที่รวบรวมได้จะอยู่ในรูปแบบไฟล์ gpx เราจึงจำเป็นต้องผ่านกระบวนการในการทำความสะอาดข้อมูลต่างๆ เพื่อขจัดข้อมูลจะลดประสิทธิภาพของการจัดกลุ่ม ซึ่งใช้วิธีการแบบเดียวกับการเตรียมข้อมูลในบทที่ 3
2. นำเส้นทางทั้งหมดที่ผ่านการทำความสะอาดข้อมูลแล้ว มาจัดกลุ่มตามความยาวของเส้นทาง โดยใช้ตัวแบบในการจัดกลุ่มตามความยาวในบทที่ 3

การแนะนำเส้นทางเฉพาะบุคคล

ในโครงการนี้ นำเสนอแนวทางการแนะนำเส้นทางให้แก่ผู้ใช้ โดยแก้ปัญหาเรื่อง Cold Start โดยนำเทคนิค Collaborative filtering โดยสร้างรูปแบบพฤติกรรมการขี่จักรยานของผู้ใช้จากประวัติการขี่จักรยานของผู้ใช้เอง มาสร้างเป็นข้อมูลรูปแบบพฤติกรรมการขี่จักรยานของผู้ใช้รายนั้น แล้วนำไปหาผู้ใช้ที่มีพฤติกรรมใกล้เคียงกัน แล้วนำเสนอเส้นทางที่เหมาะสม โดยพิจารณาเงื่อนไขการค้นหาค้นหาอื่นๆ ณ ขณะนั้น โดยให้กำหนดให้ผลการสืบค้นอันดับต้นๆ เป็นเส้นทางที่ได้รับความนิยมชื่นชอบจากผู้ใช้รายอื่นที่มีพฤติกรรมใกล้เคียงกับผู้ใช้รายนั้น เพื่อแนะนำเส้นทางที่เหมาะสมกับความต้องการและพฤติกรรมของผู้ใช้ โดยมีรายละเอียดตามลำดับการทำงานในภาพที่ 4.2



ภาพที่ 4.2 ลำดับการทำงานสำหรับการแนะนำเส้นทางให้แก่ผู้ใช้

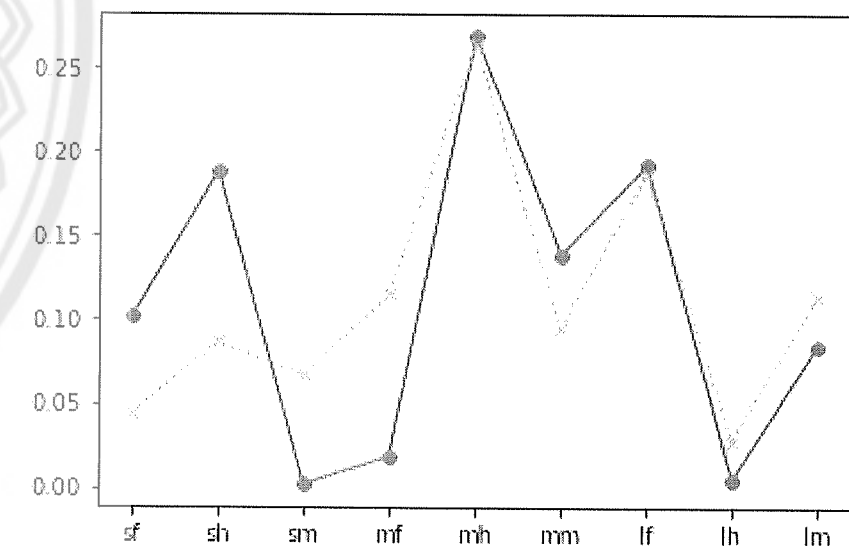
1. เมื่อมีผู้ใช้รายใหม่เข้ามาในระบบ ระบบจะทำการสืบค้นพฤติกรรมการขี่จักรยานจากประวัติการขี่ย้อนหลัง ซึ่งผู้ใช้ต้องอนุญาตให้เก็บข้อมูลประวัติจากแหล่งข้อมูลที่เข้าถึงได้ เช่น Strava หรือ Garmin connect เป็นต้น เมื่อทำการรวบรวมข้อมูลได้แล้ว ระบบจะ

ดำเนินการจัดเตรียมข้อมูลผู้ใช้จักรยาน ดังรายละเอียดที่ได้กล่าวในข้อที่ 4.1. เพื่อสร้างข้อมูลตัวแทนพฤติกรรมผู้ใช้จักรยาน ให้แก่ผู้ใช้รายใหม่นี้

2. ค้นหาผู้ใช้รายเดิมในฐานะข้อมูลที่เกิดจากการขี่จักรยานใกล้เคียงกับพฤติกรรมผู้ใช้จักรยานของผู้ใช้รายใหม่มากที่สุด
3. ผู้ใช้กรอกข้อมูลบริบทเพื่อค้นหาเส้นทาง ได้แก่ พิกัดตำแหน่งที่ต้องการไปท่องเที่ยว และขอบเขตของระยะทางที่ต้องการขี่จักรยาน และขอบเขตของความสูงชันของเส้นทางที่ต้องการขี่จักรยาน
4. ระบบสืบค้นเส้นทางตามเงื่อนไขต่างๆที่ผู้ใช้ระบุไว้ในข้อที่ 2.3.3 โดยจัดเรียงเส้นทางที่มีอยู่ในหมวดหมู่ที่ผู้ใช้มีพฤติกรรมขี่ในสัดส่วนที่สูงที่สุดขึ้นมาเป็นอันดับต้น

ผลการทดลอง

จากการจัดเตรียมข้อมูลผู้ใช้จักรยานตามแนวทางที่ 2.1 เราสามารถจัดเตรียมฐานข้อมูลพฤติกรรมผู้ใช้จักรยานของผู้ใช้จำนวน 50 คน แล้วนำมาวิเคราะห์ว่าผู้ใช้แต่ละคนมีพฤติกรรมใกล้เคียงกับผู้ใช้จักรยานคนใดในฐานะข้อมูล โดยใช้วิธีหาค่า similarity ด้วย Euclidean distance ดังตัวอย่างในภาพที่ 4.3



ภาพที่ 4.3 ตัวอย่างพฤติกรรมผู้ใช้จักรยานของผู้ใช้ที่มีความใกล้เคียงกัน

จากภาพที่ 4.3 กำหนดให้เส้นทึบคือข้อมูลพฤติกรรมของผู้ใช้รายใหม่ และเส้นประคือข้อมูลพฤติกรรมของผู้ใช้ที่มีพฤติกรรมใกล้เคียงกับผู้ใช้รายใหม่มากที่สุด ซึ่งมีค่าสัดส่วนในการขี่ของการขี่บน

เส้นทางประเภทต่างๆที่ใกล้เคียงกัน ซึ่งเมื่อวัดด้วย Euclidean distance แล้วมีค่าน้อยที่สุดเมื่อเปรียบเทียบกับพฤติกรรมของผู้ใช้รายอื่น ๆ จึงสามารถนำข้อมูลของผู้ใช้ที่มีพฤติกรรมใกล้เคียงกับผู้ใช้รายใหม่นี้ มาช่วยในการแนะนำเส้นทางให้แก่ผู้ใช้รายใหม่ต่อไป



สรุปผลการทดลอง อธิบายผลและข้อเสนอแนะ

คณะผู้วิจัยได้กำหนดขอบเขตปัญหาการวิจัยโดยมุ่งเน้นที่การแนะนำเส้นทางการขี่จักรยานรายบุคคลบนพื้นฐานรูปแบบการขี่เฉพาะบุคคลเป็นปัญหาในการวิจัย ดังนั้นการวิจัยเพื่อพัฒนาระบบแนะนำการขี่จักรยานเชิงท่องเที่ยวบนพื้นฐานรูปแบบการขี่เฉพาะบุคคล ด้วยเทคโนโลยีอุปกรณ์ไอที สวมใส่นี้ประกอบด้วย

1. การรวบรวมข้อมูลการขี่จักรยานจากอุปกรณ์ไอทีสวมใส่ของนักท่องเที่ยว และลักษณะของเส้นทางจากแหล่งท่องเที่ยวสำหรับการขี่จักรยาน
2. ระบบจัดกลุ่มเส้นทางการขี่จักรยาน ซึ่งผลลัพธ์ที่มีผลการประเมินประสิทธิภาพในการจำแนกและจัดกลุ่มที่ดีที่สุดจะถูกนำมาประยุกต์ใช้ในกระบวนการแนะนำเส้นทางที่เหมาะสมกับการขี่จักรยานเฉพาะบุคคล
3. ระบบแนะนำเส้นทางท่องเที่ยวให้สอดคล้องกับพฤติกรรมการขี่จักรยานเฉพาะบุคคล ตามความต้องการของผู้ใช้ และพิกัดของเส้นทางที่ได้รับความนิยม

จากแนวทางการออกแบบดังกล่าว ช่วยให้ระบบสามารถแนะนำเส้นทางการขี่จักรยานที่เหมาะสมกับพฤติกรรมการขี่จักรยานมากขึ้น และด้วยเหตุผลที่ผู้ใช้ได้ขี่จักรยานบนเส้นทางที่มีคุณลักษณะที่ใกล้เคียงกับเส้นทางที่เคยขี่มาก่อน ก็จะช่วยให้เกิดความเพลิดเพลินกับการขี่จักรยานในได้มากขึ้น ซึ่งหากขยายฐานข้อมูลให้ครอบคลุมพื้นที่มากยิ่งขึ้น ก็จะช่วยให้ผู้ใช้งานสามารถค้นพบเส้นทางขี่จักรยานในพื้นที่ต่างๆ ช่วยเพิ่มประสบการณ์ท่องเที่ยวเชิงสันทนาการที่ดี และมีส่วนช่วยกระตุ้นการท่องเที่ยวในพื้นที่ใหม่ๆได้

สรุปผลโครงการวิจัยนี้ได้นำเสนอแนวทางในการแนะนำเส้นทางการขี่จักรยานเฉพาะบุคคล โดยมีกระบวนการในการจัดเก็บข้อมูลประวัติการขี่จักรยาน ตลอดจนเตรียมข้อมูลที่จัดเก็บมาให้พร้อมสำหรับการวิเคราะห์ข้อมูล จากนั้นนำข้อมูลที่ผ่านการเตรียมมาจำแนกกลุ่มด้วยตัวแบบต่างๆ ได้แก่ตัวแบบความใกล้เคียงของเส้นทาง และตัวแบบความสูงของเส้นทาง จากนั้นนำผลการจำแนกมาจัดเรียงข้อมูลใหม่เพื่อใช้เป็นข้อมูลพฤติกรรมการขี่จักรยานของผู้ใช้รายบุคคล ซึ่งข้อมูลพฤติกรรมการขี่จักรยานของผู้ใช้รายบุคคลจะถูกนำไปเปรียบเทียบกับผู้ใช้รายอื่นๆ เพื่อค้นหาว่าผู้ใช้รายใหม่ มีพฤติกรรมใกล้เคียง

กับผู้ใช้รายใดในฐานะข้อมูล จากนั้นจึงนำเสนอเส้นทางตามเงื่อนไขการสืบค้นของผู้ใช้รายนั้น โดยจัดอันดับตามความชื่นชอบของผู้ใช้ในฐานะข้อมูลที่มีพฤติกรรมใกล้เคียงกับผู้ใช้รายนั้นๆ ต่อไป



บรรณานุกรม

- Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggle, P. (1999). Towards a better understanding of context and context-awareness. In International Symposium on Handheld and Ubiquitous Computing (pp. 304–307). Springer.
- Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734–749. <https://doi.org/10.1109/TKDE.2005.99>
- Agapito, G., Calabrese, B., Carè, I., Falcone, D., Guzzi, P. H., Ielpo, N., ... Cannataro, M. (2014). Profiling basic health information of tourists: Towards a recommendation system for the adaptive delivery of medical certified nutrition contents. In High Performance Computing & Simulation (HPCS), 2014 International Conference on (pp. 616–620). IEEE.
- Amazon. (2020). **AmazonBasics 48 Pack AA High-Performance Alkaline Batteries, 10-Year Shelf Life, Easy to Open Value Pack.** เข้าถึงเมื่อ 18 สิงหาคม 2563, จาก https://www.amazon.com/AmazonBasics-Performance-Alkaline-Batteries-Count/dp/B00MNV8E0C/ref=sr_1_1?dchild=1&keywords=amazonbasics&pf_rd_p=9349ffb9-3aaa-476f-8532-6a4a5c3da3e7&pf_rd_r=VPRA49MAYMQRKX4QGNHY&qid=1597681610&sr=8-1.
- Baptiste Rocca. (2019). **Introduction to recommender systems.** เข้าถึงเมื่อ 7 สิงหาคม 2563, จาก <https://towardsdatascience.com/introduction-to-recommender-systems-6c66cf15ada>
- Bushra Alhijawi, Ghazi Al-Naymat, Nadim Obeid, Arafat Awajan. (2019). **Mitigating the Effect of Data Sparsity: A Case Study on Collaborative Filtering Recommender System.** 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), Amman, Jordan, Jordan.
- Charu C. Aggarwal. (2016). **Recommender Systems : The Textbook.** NY, USA: Springer International Publishing.

- Emma Grimaldi. (2018). How to build a content-based movie recommender system with Natural Language Processing. เข้าถึงเมื่อ 7 สิงหาคม 2563, จาก <https://towardsdatascience.com/how-to-build-from-scratch-a-content-based-movie-recommender-with-natural-language-processing-25ad400eb243>.
- Gao, M., Liu, K., & Wu, Z. (2010). Personalisation in Web Computing and Informatics: Theories, Techniques, Applications, and Future Research. *Information Systems Frontiers*, 12(5), 607–629. <https://doi.org/10.1007/s10796-009-9199-3>
- Handaru. (2016). Build Recommendation Engine Using Graph. เข้าถึงเมื่อ 10 สิงหาคม 2563, จาก <https://medium.com/tiket-com-dev-team/build-recommendation-engine-using-graph-cbd6d8732e46>.
- Jackson Wu. (2019). Knowledge-Based Recommender Systems: An Overview. เข้าถึงเมื่อ 1 กรกฎาคม 2563, จาก <https://medium.com/@jwu2/knowledge-based-recommender-systems-an-overview-536b63721dba>.
- Juthaporn Vipatpakpaiboon. (2019). มาทำความรู้จัก ระบบแนะนำ (Recommendation System) ว่าคืออะไรกัน. เข้าถึงเมื่อ 7 สิงหาคม 2563, จาก <http://bigdataexperience.org/introduction-to-recommendation-system/>.
- Lamsfus, C., Alzua-Sorzabal, A., Martin, D., Salvador, Z., & Usandizaga, A. (2009). Human-centric Ontology-based Context Modelling in Tourism. In *KEOD* (pp. 424–434).
- Ricci, F. (2002). Travel recommender systems. *IEEE Intelligent Systems*, 17(6), 55–57.
- Sieg, A., Mobasher, B., & Burke, R. D. (2007). Learning Ontology-Based User Profiles: A Semantic Approach to Personalized Web Search. *IEEE Intelligent Informatics Bulletin*, 8(1), 7–18.
- Sirinart Tangruamsub. (2019). Recommendation System (แบบสรุปไม่ค่อยสั้น). เข้าถึงเมื่อ 10 สิงหาคม 2563, จาก <https://medium.com/@sinart.t/recommendation-system-แบบสรุปเอาเอง-ce6246f49754>.
- Tulasi K. Paradarami, Nathaniel D. Bastian, และ Jennifer L. Wightman. (2017). A hybrid recommender system using artificial neural networks. *Expert Systems with Applications*, 83, 300-313. <http://www.sciencedirect.com/science/article/pii/B9780080890500000250>

Walraven, G. (2010). Basic Arrhythmias, 7th Edition (7 edition). Upper Saddle River, NJ: Prentice Hall.





ภาคผนวก

Design Framework for Building a Recreation Cycling Recommendation System based on Personalized Cycling Style with Wearable Technology

Surapong Wiriya^{a,b,*}, Nattapon Kumyaito^c, Kreangsak Tamee^{a,d,*}

^aDepartment of Computer Science and Information Technology, Faculty of Science, Naresuan University, Phisanulok 65000, Thailand

^bMajor of Computer Multimedia, Faculty of Science and Technology, Chaopraya University, Nakhon Sawan 60240, Thailand

^cDepartment of Information Technology, School of Information and Communication Technology, University of Phayao, Phayao 56000, Thailand

^dResearch Center for Academic Excellence in Nonlinear Analysis and Optimization, Naresuan University, Phisanulok 65000, Thailand

*E-mail address: surapongw58@email.nu.ac.th (Surapong Wiriya), kreangsakt@nu.ac.th (Kreangsak Tamee)

Abstract: This paper presented a framework for building a recreation cycling recommendation system based on personalized cycling style with wearable technology. This framework is applied to recreation cycling recommendation system using data from wearable IT equipment. For example, frequency data of cycle the legs, velocity, distance, coordination, slope of path and temperature are built cycling pattern model and tourist attraction model by machine learning technique for recommending route, consists of distance, slopes and enjoyment. The travel information cost distance from home or hotel and also predictions variables also can be used to make the decisions in cycling, for example rainy, humidity, temperature and windy.

Keywords: recreation cycling, recommendation system, personalized cycling style, wearable technology

1. Introduction

Recently, recreation tourism is popular in the world. Because recreation activity helps healthcare and mind relaxation. John Tribe (Tribe, 2011) definition of recreation tourism, travel has an extreme activity, for example, climbing, fishing and golfing. In 2016, (UNWTO, 2017) International tourist arrivals grew by 3.9% to reach a total of 1,235 million worldwide, an increase of 46 million over the previous year. Asia growth in 2016 with 9% increase in international arrivals. International tourism receipts grew by 2.6% in real terms with total earnings in the destinations estimated at US\$ 1,220 billion worldwide. For Thailand, (Vanhalleyk, 2017) the first quarter of 2017 was the arrival of 9,194,057 tourists, which was a modest increase of 1.72% compared with the previous year (9,038,893 arrivals). The second quarter was and increase of 7.63% compared with 2016 (From 7,553,001 to 7,129,583). The third quarter saw an increasing in visitors from 8,246,701 to 8,777,142 (6.43%). (WTTC, 2017) The total contribution of Travel & Tourism to GDP was USD82.5bn, 20.6% of GDP in 2016, and is forecast to rise by 9.4% in 2017, and to rise by 6.5% pa to USD169.9bn, 31.7% of GDP in 2027. (Thaihealth, 2015) Thai people have been alerted to all kinds of health and physical activities, especially running and cycling. The data has shown that Thai people are more popular in cycling for health and recreation tourism, and more are seen in the cycling tour of the many natural routes.

Tourism Recommendation Systems have become very popular applications for organizing and planning tourist trips. A recommender system for tourism industry (Mehrbakhsh, Karamollah, Mohsen, & Vahid, 2017) introduces a new introduction system using forecasting and dimension reduction and grouping methods using data from TripAdvisor. (Leila & David, 2015) Adaptive Tourist Recommendation System (ATRS), which can change its behavior to adapt to the new situation. This system

collects information about the traveler, for example, accommodation and places. Type of transportation, restaurants, cafes, shopping centers, medical, pharmacies and hospital information and the travel are produced from different resources. It offered personalized recommendation based on the collected information. The sensor is also used for Cyclist Experience Mapping (Shane, et al., 2009) by creating a cycling map of the cycling experience. The environment is carbon dioxide, Volume level, and the traffic is tight. We use them to analyze the route for the right path and measure cyclist performance and fitness measurement by sending data from sensor via wireless access point and mobile phone. (Sasank, Katie, & Gleb, 2010) Build a system that helps guide the best route by using data from cyclists through the GPS to store speed, accelerometer used for road surface analysis, and use a camera to create Geo-tagged.

From the above mentioned text. It is evidenced that the tourism industry is growing every year. On the other hand, the promotions of public health care. Organizing activities that make people want to do an exercise with more bicycles. The numbers of people turn into cycling for health and to travel in various places increased. There are some systems that consider the health of tourists. (Agapito, et al., 2014) Create a system that recommends local food that is good for tourists to achieve the better health. In addition, the sensor is used to collect the data to analyze the best route for cyclists. The information obtained based on the sensor, for example, speed, surface conditions, through storage on the site. In this paper, we see the perspective of individual cyclists. Most often, they enjoy different cycling routes based on their personal appearances and personal preferences. The data is collected in the analysis will focus on finding the route for an individual, and it based on health considerations. Using Basics

data, temperature data and heart rate, are used to provide the route for cyclists to recreate tourism.

This paper was organized as follows: section 2 Trend Technology. In section 3 we propose Recreation Cycling Recommendation System (RCRS) concepts. Finally in section 4 Conclusion.

2. Sensing Technology

One of the information that supports the classification of personal cycling is cycling history information which current cyclists often use IT equipment to record its cycling data by the IT equipment. There is either a bicycle mount or a wearable. Examples, IT equipment installed on the bike are the Cyclometer (Fig. 1a), a set of devices that incorporates different types of receivers, such as GPS for positional velocity data and slope path, Thermometer recognition for temperature and Cadence meter (Fig. 1b) for frequency data of cycle the legs, wearable IT equipment include: Heart rate perception (Fig. 2a and 2b), which is characterized by two types of strains the electrical signal from the body. The same principle applies to the EKG (Walraven, 2010) in medical practice and a light-based watch will work by changing the rate of light absorbed by the skin, which is varies with the heartbeat.

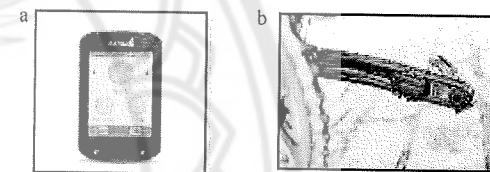


Fig. 1 - (a) Cyclometer (b) Cadence meter

source: a) garmin (2017), b) racingbikeshop (2017)

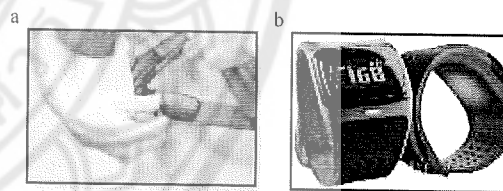


Fig. 2 - (a) hung on the breast (b) watch

source: a) hetloopcentrum (2017), b) pinterest (2017)

The wearable IT equipment can be used to record these data immediately. Cyclists can store such data into a storage system. Cyclists can choose from a variety of services, for example, directly from IT equipment manufacturers and the standard service open to the wearable IT

equipment can be sent to data store, i.e. Therefore, these data storage services are important sources that can be linked to import data into classification of personalized cycling which this can be used to classify the groups and recommend routes which are suitable with the specific cycling patterns.

Framework that built a recreation cycling recommendation system based on personalized cycling style with wearable technology. It included a collection of cycling data from various cycling storage providers to identify and group personalized cycling style as a guide to user needs and the coordinates of the popular routes.

3. Recreation Cycling Recommendation System (RCRS) Concept

The goal of this section is to set the fourth a design framework consisting of Data Acquisition Layer, Data Collection Layer, Machine

Learning Layer and Recommender Layer (Fig. 3)

3.1 Data Acquisition Layer

The Data Acquisition Layer contains receiver about cycling data from wearable IT equipment for stretching of the pulse and other perceptual data to detect other signals from the cycling, for example, frequency data of cycle the legs, velocity, distance, coordinate, slope of path and temperature, i.e. The receiver sends the detected signal to the cyclometer, which collects signals from the sensors and stores it in a standardized data structure. This framework uses the Training Center XML (TCX), a popular data structure for exchanging data between exercise equipment when the user finishes cycling and records the cycling information. The signals from the receiver are compiled into a TCX file and then uploaded into the activity data storage services to collect the entire cycling history of each user.

1) *Wearable Sensors* (Subhas, 2015) have become very popular in many applications for example medical, entertainment, security, and commercial fields. Wearable sensor are useful in (Yongjin, Kyuchang, & Changseok, 2014), (Liming, Jesse, & Chris, 2012) human activity recognition, (Tang, 2015) monitoring and improving sports performance, (Sasank, Katie, & Gleb, 2010), (Shane, et al., 2009) find good routes for cyclist, etc. Generally, (Liming, Jesse, & Chris, 2012) wearable sensor refers to sensors that are positioned directly or indirectly on human body. They can be used to collect information for example body position and movement, pulse, heart rate, and body heat. Wearable sensors are used to detect cycling activities. It consists of: a) *Accelerometer sensors* (Liming, Jesse, & Chris, 2012) which are particularly effective in monitoring actions that involve in activity recognition, such as standing, sitting, walking, running and riding. In the research, accelerometer sensors have been used to learn how to build human activities. For example, (Reza & Amir, 2016) Construction worker's activity recognition; state, behavior, and

surrounding context using sensors on smartphone for capture body movements by collecting data using embedded accelerometer and gyroscope sensors. (Andrea , Angelo, & Stephen, 2015) Automatic detection of body position of a wearable sensor on five different parts of the body is ankle, thigh, hip, arm and wrist. The most accurate location is the installation of the sensor at the ankle for walking activity. *b) Global Positioning System (GPS)* (Liming, Jesse, & Chris, 2012) wearable sensor of monitoring location-based activities in open pervasive and mobile environment.

2) *TCX file*: Training Center XML (TCX) (Training Center XML, 2016) is information trades arrange presented in 2007 as a component of

tracks, however regards a track as an action instead of basically a progression of GPS focuses. TCX gives gauges to exchanging heart rate, running cadence, bicycle cadence, calories in the detailed track. It additionally gives synopsis information as laps.

3) *Data Storage*: Cloud Storage systems (Rekha, Bahman, Rodrigo, & Kenan, 2017) are now mature enough to handle a massive volume of heterogeneous and rapidly changing data that consist of a number of storage devices connected by the network. Recreation Cycling Recommendation System has to store personal activities data, for instance; frequency data of cycle the legs, velocity, distance, coordinate, slope of path and temperature,

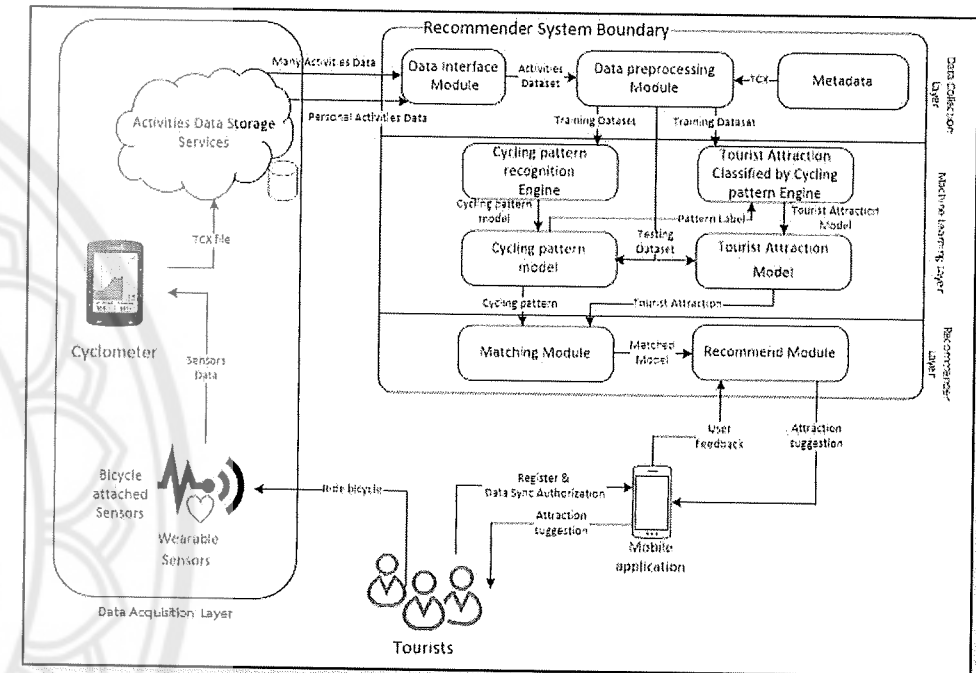


Fig. 3 Conceptual Framework

personal activities data are Training Center XML (TCX) format (Fig. 4), a popular data structure for exchanging data between exercise equipment when the user finishes cycling and records the cycling information and transfer to cloud storage system before sending to Data Collection Layer.

3.4 Recommender Layer

Recommender System is the result of personalizing cycling model and interesting travel routes. It is a system that recommends cycling suitable routes for personalizing cyclists. The Recommender system communicates with the user through mobile applications. It has the ability of recommended interesting travel routes in the neighbourhood of the user's current location coordinates or coordinates users' interested and when the user cycles on the route recommended. This cycling information will be synchronized with the recommender system to evaluate the quality of the route guidance, and then improve the route guidance.

1) Matching Module

The matching of the system is to obtain the cycling data of the cyclist in order to obtain the value of the transmission of the wearable IT equipment. The analysis compared with the cycling pattern model to determine the cycling pattern of cyclists. Then, the pattern is matched with the tourist attraction model that was created in the machine learning layer and then introduced the personalized cycling style.

2) Recommend Module

When the data is matched, the system will introduce the personalized cycling style. It has introduced route, consists of distance, slopes and enjoyment. The travel cost information, for example; distance from home or hotel, predictions variables that can be used to make decisions in cycling, during rainy, humidity, temperature and windy. Tourists can use information via Mobile application. Therefore, tourists connect the wearable IT equipment and Mobile Application then tourists need to register to application before receiving attraction suggestion.

4. Conclusion

In this paper, we proposed the conceptual framework that built a recreation cycling recommendation system based on personalized cycling style with wearable technology to receive the data from wearable IT equipment, building cycling model and tourist attraction model. The target tasks were cycling data of personal cyclist to matching with cycling model and tourist attraction model for helpful recommend route of personal cyclist. The solution is to set the fourth a design framework consisting of *Data Acquisition Layer* contains receiver about cycling, for example, wearable IT equipment for stretching of the pulse and other perceptual data to detect other signals from the cycling. *Data Collection Layer* includes a Data Interface Module to personalize cycling history resources from various fitness storage services. *Machine Learning Layer* build cycling model and tourist attraction model by using machine learning techniques, and *Recommender Layer* is the result of personalizing cycling model and interesting travel routes. It is the system

that recommends cycling suitable routes for personalizing cyclists. Resultantly, Cyclist could get meaningful information about recreation cycling.

Acknowledgements

This work was supported by the Department of Computer Science and Information Technology, Faculty of Science, Naresuan University and Chaopraya University, Thailand.

REFERENCES

- Agapito, G., Calabrese, B., Care, I., Falcone, D., Guzzi, P. H., Lelpo, N., & Cannataro, M. (2014). Profiling basic health information of tourists: Towards a recommendation system for the adaptive delivery of medical certified nutrition contents. *High Performance Computing & Simulation, 2014 International Conference* (pp. 616-620). IEEE.
- Andrea, M., Angelo, M., & Stephen, S. (2015). Accelerometry-based recognition of the placement sites of a wearable sensor. *Pervasive and Mobile Computing, 62-74*.
- Bing, L., Songshan, (. H., & Hui, F. (2017). An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tourism Management, 132-141*.
- Damianos, G., Charalampos, K., Konstantinos, M., & Grammati, P. (2014). Mobile recommender systems in tourism. *Journal of Network and Computer Application, 319-333*.
- Gao, M., Liu, K., & Wu, Z. (2010). Personalisation in Web Computing and Informatics: Theories, Techniques, Applications, and Future Research. *Information Systems Frontiers, 607-629*.
- Joan, B., Antonio, M., & Aida, V. (2014). Intelligent tourism recommender systems: A survey. *Expert Systems with Application, 7370-7389*.
- Junlin, Z., Samuel, O. W., & Haoziang, W. (2017). Intelligent computing system based on pattern recognition and data mining algorithms. *Sustainable Computing: Informatics and Systems*.
- Leila, E., & David, S. (2015). Adaptive tourist recommendation system: conceptual frameworks and implementations. *Vietnam J Comput Sci, 95-107*.

- Li, L., Yuxin, P., Shu, W., Ming, L., & Zigang, H. (2016). Complex activity recognition using time series pattern dictionary learned from ubiquitous sensors. *Information Sciences*, 41-57.
- Liming, C., Jesse, H., & Chris, D. (2012). Sensor-Based Activity Recognition. *IEEE Transactions on System, Man, and Cybernetics*, 790-808.
- Malley, B., Ramazzotti, D., & Wu, J. (2016). Data Pre-processing. In M. C. Data, *Secondary Analysis of Electronic Health Records* (pp. 115-141). Springer, Cham.
- Mehrbakhsh, N., Karamollah, B., Mohsen, R., & Vahid, R. (2017). A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques. *Computers & Industrial Engineering*, 357-368.
- Muhammad, A., & Muhammad, U. (2015). A Novel Framework for Aspect-based Opinion Classification for Tourist Places. *The Tenth International Conference on Digital Information Management (ICDIM 2015)* (pp. 1-9). Republic of Korea : IEEE.
- Peter, M., & Timothy, G. (2011). *The NIST Definition of Cloud Computing*. Gaithersburg: National Institute of Standards and Technology.
- Rekha, N., Bahman, J., Rodrigo, N., & Kenan, M. (2017). Cloud storage reliability for Big Data applications: A state of the art survey. *Journal of Network and Computer Application*, 35-74.
- Reza, A., & Amir, H. (2016). Smartphone-based construction worker's activity recognition and classification. *Automation in Construction*, 198-209.
- Sasank, R., Katie, S., & Gleb, D. (2010). Biketastic: Sensing and Mapping for Better Biking. *ACM Conference on Human Factors in Computing Systems* (pp. 1817-1820). USA: ACM.
- Sergio, R.-G., Bartosz, K., Salvador, G., & Michal, W. (2017). A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing*, 39-57.
- Shane, B. E., Emiliano, M., Nicholas, D. L., Ronald, A. P., Gahng-Seop, A., & Andrew, T. C. (2009). *BikeNet: A Mobile Sensing System for Cyclist Experience Mapping*. *ACM Transactions on Sensor Networks*, 6:2-6:39.
- Steven, V., Viktor, S., Olivier, J., Pieterjan, D. P., Jurgen, S., & Rik, V. W. (2013). Web-based Enrichment of Bike Sensor Data for Automatic Geo-annotation. *GEOCROWD* (pp. 63-70). USA: ACM.
- Subhas, C. M. (2015). Wearable Sensors for Human Activity Monitoring: A Review. *IEEE Sensors Journal*, 1321-1330.
- Tang, S. L. (2015). *Textiles for Sportswear*. UK: Elsevier.
- Thaihealth. (2015, August 17). *คนไทยปั่นจักรยานเพิ่มขึ้น 100 เปอร์เซ็นต์*. Retrieved from <http://www.thaihealth.or.th/Content/29147-คนไทยปั่นจักรยานเพิ่มขึ้น%20100%20เปอร์เซ็นต์.html>
- Thailandtourismcouncil. (2017). *Thailand Tourism Confidence Index*. Retrieved from http://www.thailandtourismcouncil.org/imgadmins/confidence_file/TCT_confidence_file_th_2016-09-27_08-10-17.pdf
- Tourism, D. o. (2017). *ระบบฐานข้อมูลแหล่งท่องเที่ยว*. Retrieved from <http://61.19.236.136:8090/dotr/statistic.php>
- Tousim Western Australia. (2008). *JUMPSTARTGUIDE*. Retrieved from <http://revistas.um.es/turismo/article/viewFile/221641/173181>
- Training Center XML. (2016, August 22). Retrieved from https://en.wikipedia.org/wiki/Training_Center_XML
- Tribe, J. (2011). *The Economics of Recreation, Leisure and Tourism*. Retrieved from <http://www.sciencedirect.com/science/article/pii/B9780080890500000250>
- UNWTO. (2017). *UNWTO Tourism Highlights 2017 Edition*. Retrieved from <https://www.e-unwto.org/doi/book/10.18111/9789284419029>
- Vanhaleweyk, G. (2017, November). *Tourism Statistics Thailand 2000-2017*. Retrieved from <http://www.thaiwebsites.com/tourism.asp>
- Walraven, G. (2010). *Basic Arrhythmias, 7th Edition (7 edition)*. Upper Saddle River, NJ: Prentice Hall.

WTTC, C. (2017). *TRAVEL&TOURISM ECONOMIC IMPACT 2017 THAILAND*. Retrieved from <https://www.wttc.org/-/media/files/reports/economic-impact-research/>

Yongjin, K., Kyuchang, K., & Changseok, B. (2014). Unsupervised learning for human activity recognition using smartphone sensors. *Expert System with Application*, 6067-6074.

Zulfiqar, A., Syed, K. S., & Waseem, S. (2017). Performance Analysis of Statistical Pattern Recognition Methods in KEEL. *International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017* (pp. 2022-2030). Marseille, France: Procedia Computer Science.



TRAJECTORY CLUSTERING BY GPS TRACKING DATASET USING QUICKBUNDLES

NATTAPON KUMYAITO¹ AND KREANGSAK TAMEE^{1,2}

¹Department of Computer Science and Information Technology

²Research Center for Academic Excellence in Nonlinear Analysis and Optimization
Naresuan University
Muang, Phitsanulok 65000, Thailand
{nattaponk; kreangsakt}@nu.ac.uk

Received March 2020; accepted June 2020

ABSTRACT. *Trajectory clustering is considered as analytical methods in trajectory data mining for many modern applications such as a tourist recommendation system. Our research project is a route recommending system for cycling tourists, and we aim to retrieve all their route patterns by adopting GPS-based trajectories dataset. This dataset is informative and high volume; thus, we need clustering methods that are effective and fast in computational time. This paper proposes methods to collect, preprocess, and analyze the GPS tracking data from the existing online service for cyclists. QuickBundles is the primary technique we use to allocate route items into clusters, and also propose cluster validation methods based on silhouette coefficient. The results have shown that QuickBundles has the capability of clustering the trajectory data in terms of computational speed and validity. This approach can benefit online systems, by recommending personalized routes for the cycling tourists.*

Keywords: Trajectory model, Data processing, Pattern classification, Recommendation system

1. **Introduction.** Nowadays, location-tracking devices have evolved and are widely adopted. These devices are capable of precisely tracking locations and record them on-the-fly to cloud services, with the most famous one being the Global Positioning System (GPS). GPS provides the advantage of both global and accuracy, via a group of satellites that orbit the earth broadcasting signals that are picked up by a system of receivers.

Li et al. [1] proposed that the use of GPS for tourism research was very beneficial for tourism research which matured from feasibility and usefulness, tourist behavior and tourism recommendation. Since our research project aimed to classify routes by analyzing GPS data, we reviewed other related works that analyzed routes for tourism. For instance, Du et al. [2] used the topic hierarchy of scenic spots and the features of scenic spots for travel route mining system. Jiang et al. [3] identified scenic hotspot by proposing density based clustering algorithm using the data of geographical coordinates and tour route data. Malik and Kim [4] proposed travel route optimization techniques that use tourist vehicle data. Hamid and Croock [5] predicted tourists' interested place using their smartphone's GPS trajectory. Nardini et al. [6] analyzed the trajectories of mobile users to initiate the recommender system for tourists.

This research aims to propose the route recommendation system for cyclists who need to discover riding routes that have similar characteristic to their riding history. This system must recognize the routes pattern from cyclists' activities data source that may be considered as big data. Therefore, we need effective methods to cluster vast amount of data into groups of route patterns. We adopted the QuickBundles (QB) [7] algorithm, due to its efficient computational complexity and simplification, which was successfully

applied to tractography clustering analysis in Diffusion MRI dataset that overcame the complexity of extensive data, and provided informative clusters in seconds. The details of the QB technique are described in Section 2.3.

The further details of trajectory clustering in this paper are presented in the following sections. Section 2, Problem Statements and Preliminaries, describes the main purpose of trajectory clustering problems and the methods required to prepare and analyze the dataset. Section 3, Experiment, describes the techniques for constructing trajectory clustering models and its results. Section 4, Discussion, presents the findings of this study. Finally, Section 5, Conclusions, will sum up the important issues and research opportunities.

2. Problem Statements and Preliminaries. Trajectory data is recorded in different forms according to the types of sensors, objects movement, or tracking purposes. In this paper, we focused on geographical trajectory data, which is represented as a sequence of n points in geographical space as the following equation:

$$T = (t_1, t_2, \dots, t_n) \quad (1)$$

where t_i is a tracking point which is represented as a combination of coordination and timestamp that forms the sub equation:

$$t_i = (x_i, y_i, z_i, ts_i) \quad (2)$$

where x_i , y_i , z_i , and ts_i are latitude, longitude, altitude, and timestamp, respectively. This data sequence is an essential standard for several data schemes in GPS tracking devices, while some of the data schemes add other relevant information, such as velocity and cumulative distance.

2.1. Dataset. The cycling workout data that contains GPS track points were explored and selected from online data services, <https://ridewithgps.com/>. This research used selection criteria as follows.

- 1) The selected data must be recorded as TCX file.
- 2) Every track point in the selected files must have altitude data.
- 3) The route length in the selected files must range between 3 and 180 km.

After previous selection criteria were applied, selected files were converted into raw dataset using a GPSBabel application [8]. However, this data had defects such as missing values and/or sensor reading errors. Therefore, preprocessing is required to deal with the defects in the raw dataset.

2.2. Data preprocessing. The dataset usually contains some errors and/or outliers caused by the hardware. Therefore, to rectify these problems, the dataset was preprocessed by the following methods.

- 1) Unify the GPS track point's time intervals to 1 second.
- 2) To correct the outlier and missing values, we used the linear Kalman filter technique [9].
- 3) All elevation data was altered by resetting the initial track points to zero.

2.3. Cluster analysis using QuickBundles. QB is a clustering technique for classifying routes by their elevation profile, and its tractography simplification of a complex Diffusion MRI dataset provides informative clusters in seconds. The tractography that is sequential of points in R^3 can be categorized as a form of trajectory data. Therefore, QB is a clustering technique that may also be able to cluster trajectory data. The primary reason why QB successfully classifies trajectory data is that it uses the simple summet

Equations (3) to (5).

$$d_{direct}(s, t) = d(s, t) = \frac{1}{K} \sum_{i=1}^K |s_i - t_i| \quad (3)$$

$$d_{flipped}(s, t) = d(s, t^F) = d(s^F, t) \quad (4)$$

$$MDF(s, t) = \min(d_{direct}(s, t), d_{flipped}(s, t)) \quad (5)$$

where a trajectory $s = [s_1, s_2, \dots, s_k]$ consists of k sequential track points. s^F is a flipped version on s where $s^F = [s_k, s_{k-1}, \dots, s_2, s_1]$. $|x - y|$ denotes the Euclidean distance between two points x and y . The direct distance $d_{direct}(s, t)$ between two trajectories s and t is the mean of Euclidean distance between corresponding track points. The main advantages of the MDF distance supports QB to have substantial high performance by taking account of both direct and flipped trajectories.

2.4. Cluster validation. Since our work is unsupervised, clustered data without labels are unknown, cluster validation is considered an essential method to confirm the clustering technique by using the silhouette coefficient [10]. Let i be any object in the clustering and A its corresponding cluster. Then

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} \Delta(i, j) \quad (6)$$

measures the average distance of i to all other objects in cluster A . We then compute each cluster $C \neq A$

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} \Delta(i, j) \quad (7)$$

to quantify the distance to other ones. The minimum value is

$$b(i) = \min_{C \neq A} d(i, C) \quad (8)$$

that gives the distance of i to the second-best cluster; therefore, the silhouette value $s(i)$ of i is then defined as

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (9)$$

This value in the range of $[-1; 1]$ quantifies how well object i fits into cluster A . If the $s(i)$ is negative, the object is placed in the wrong cluster, and if it is around 0, the object is between the clusters, while if it is positive, the object is placed in the correct cluster.

3. Experiment. In this section, the preprocessed dataset that is a result of methods in prior sections is used for route clustering. The experiment goal is to classify the routes dataset according to their elevation profiles by applying QB.

Since the length of the trajectory item possibly influences the QB's result [7], the route dataset was pre-classified by length. Therefore, the differentiation of trajectory had less effect. We use k-Means to classify routes into 3 groups: short-distance, medium-distance, and long-distance ones. After trajectory data items were grouped by their route length, each item in each group had their altitude profiles classified using QB clustering.

However, the primary parameters that influence the validity of QB clustering are their distance threshold by varying their range by 10, 20, 50, 100, 200 mts, and vary the number of clusters from 2 to 6. To evaluate the validity of clustering, we applied the silhouette coefficient [10] to each group with similar route lengths. We aimed to find the best parameter schemes which presented the best results by clustering validity. We focused on the parameter scheme with a high average silhouette coefficient.

3.1. Route-length grouping using k-Means. Since this research is part of the route recommendation system for cyclists, the route distances are a primary factor that they need to consider according to their physical fitness. Instead of recommending routes without distances, to recommend ones that meet the cyclists preferred distances is more useful. In addition, the QB technique is sensitive to the varying lengths of the trajectories in the dataset. Therefore, they need to be classified by their total distances. Our classification criteria were applied due to our expert's suggestion that the appropriate distance categories should be termed as, short, medium, and long. Therefore, this paper divides the data of 481 trajectory items into 3 categories of route length by applying k-Means where $k = 3$, and the results are illustrated in Table 1.

TABLE 1. Classification dataset by route length

Group	Shortest route (km)	Longest route (km)	Number of routes
Short	1.078	59.130	303
Medium	60.005	133.918	124
Long	134.609	270.926	54

3.2. Cluster analysis. In the QB clustering technique, the primary parameters that influence clustering validity are the QB thresholds and the defined number of clusters. In Table 2 to Table 4, we demonstrate the average silhouette coefficient for the classification

TABLE 2. Average silhouette coefficient for classification of a short-distance group

QB threshold	Number of clusters				
	2	3	4	5	6
20	0.7455	0.5144	0.2650	0.1527	0.1635
30	0.7455	0.5144	0.2722	0.2792	0.2786
50	0.8601	0.6572	0.6282	0.6263	0.6308
100	0.8751	0.8346	0.8207	0.7925	0.7925
200	0.8834	0.8834	0.8834	0.8834	0.8834

TABLE 3. Average silhouette coefficient for classification of a medium-distance group

QB threshold	Number of clusters				
	2	3	4	5	6
20	0.6814	0.5826	0.4564	0.3317	0.2557
30	0.6814	0.5826	0.4564	0.3317	0.3347
50	0.6814	0.5826	0.4564	0.4709	0.4459
100	0.6814	0.5935	0.5268	0.5425	0.5740
200	0.6991	0.7034	0.7034	0.7034	0.7034

TABLE 4. Average silhouette coefficient classification of a long-distance group

QB threshold	Number of clusters				
	2	3	4	5	6
20	0.3772	0.2860	0.3406	0.2590	0.2212
30	0.3772	0.2860	0.3406	0.2590	0.2212
50	0.3772	0.2860	0.3406	0.2590	0.2212
100	0.3772	0.2860	0.4229	0.2590	0.2212

of short, medium, and long-distance routes which corresponds to the QB thresholds and the number of clusters.

For the short-distance group, the silhouette coefficient scores corresponding to the varied QB thresholds and the varied number of clusters are demonstrated in Table 2. When the QB threshold is 200, the average silhouette coefficient value for each cluster is higher than all other parameters. The lowest number of clusters is preferred because it classifies the items effectively, while minimizing computation time by comparing items distance against a less number of centroids. Therefore, we considered the QB threshold = 200 and the number of clusters = 2 as the optimal parameters for the short-distance group.

For the medium-distance group, the silhouette coefficient scores corresponding to the varied QB threshold and the varied number of clusters are demonstrated in Table 3. The average silhouette coefficient for QB threshold = 200 in all clusters is the highest among other parameter configurations. We selected QB threshold = 200 and the number of clusters = 3 as the best parameter configuration because it is the lowest number of clusters that have the highest average silhouette coefficient values.

For the long-distance group, as illustrated in Table 4, the parameter configuration that has the highest average silhouette coefficient consists of the QB threshold at 200 and 3 clusters.

3.3. Route clustering. After we achieved the optimal parameter for QB clustering, we applied it to the dataset for all groups, and routes were divided according to their length and elevation profile.

For the short-distance group, 303 routes were divided into 2 clusters according to their elevation profiles, and the centroids of the clusters that correspond to each category are illustrated in Figure 1(a).

Short routes elevation in the 1st cluster ranges between -227.88 to 316.43 meters above sea-level while the other cluster has their routes elevation range between 9.10 to 836.69 meters above sea-level, as illustrated in Figure 2(a). It shows that most of routes elevation data in the 1st cluster are relatively closer to each other at sea-level while routes elevation in the 2nd cluster is distributed among above sea-level altitude.

For the medium-distance group, 124 routes were divided into 3 clusters according to their elevation profiles, and the centroids of the clusters corresponding to each category are illustrated in Figure 1(b).

The 1st cluster of the medium-distance group consists of 89 routes with their altitudes ranging from -417.80 to 429.40 meters. The 2nd cluster has 34 routes and altitudes ranging from -181.56 to 765.33 meters above sea-level. The last cluster consists of only 1 route, and its altitude ranges between -9.11 to 836.69 meters. As illustrated in Figure 2(b), routes in the 1st cluster mostly remain in sea-level while routes in the 2nd cluster distribute on higher altitude. Finally, the route elevation in the last cluster distributes among limited range of above sea-level but consists of the highest altitude point in comparison to other routes.

In the long-distance group, 54 routes were divided into 3 clusters according to their elevation profiles, and the centroids that corresponded to each category, are illustrated in Figure 1(c).

The 1st cluster consists of 24 routes with their altitudes ranging from -319.14 to 818.4 meters. The 2nd cluster consists of 23 routes with varying altitudes between -191.80 to 481.48 meters. The 3rd cluster consists of 7 routes with altitude ranges from 0 to 1576.68 meters. Route elevation distribution in each cluster was illustrated in Figure 2(c). The 2nd cluster consists routes that mostly locate at sea-level while others consists of routes that locate above sea level. However, routes in the 1st cluster locate at lower altitude

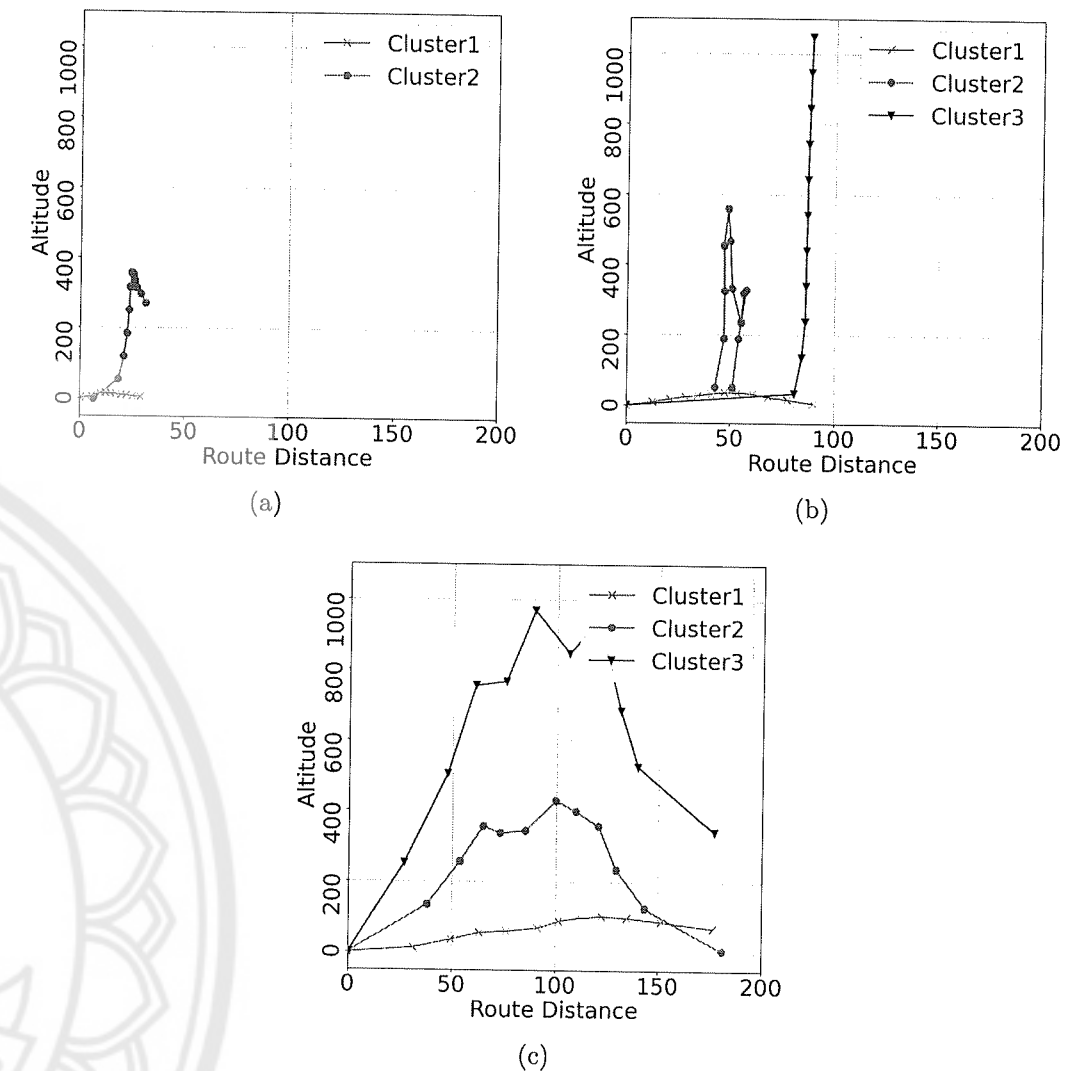


FIGURE 1. (a) Cluster centroids of short-distance groups, (b) cluster centroids of medium-distance groups, and (c) cluster centroids of long-distance groups

4. **Discussion.** The QB technique has shown that it is an effective clustering method for GPS datasets. QB threshold and the number of clusters are the primary parameters that influence the validity of route clustering based on their elevation profile. The centroid of each cluster showed some distinctive patterns that can be described as the following.

1) The flat route pattern. Route elevation profiles in this cluster usually show sustained horizontal levels with some minor changes that raises the lowest level in comparison to other clusters.

2) The hilly route pattern. The elevation profile of routes in this cluster shares a similar pattern which is a combination of flat and hilly terrain, and the raised altitude is higher than the 1st cluster but lower than the 3rd one.

3) The mountainous route pattern. Overall, routes in this pattern consist of high altitudes that are greater than either of the others. This trajectory pattern usually starts at a low altitude, but does not define the finish elevation.

The description of classified routes conforms to the cycling expert's opinion that categorizes the routes into 3 types: flat, hilly, and mountainous. Therefore, these proposed techniques have the capabilities to retrieve route patterns by classifying them according to elevation automatically. Our methods are considered as a useful tool for recommending

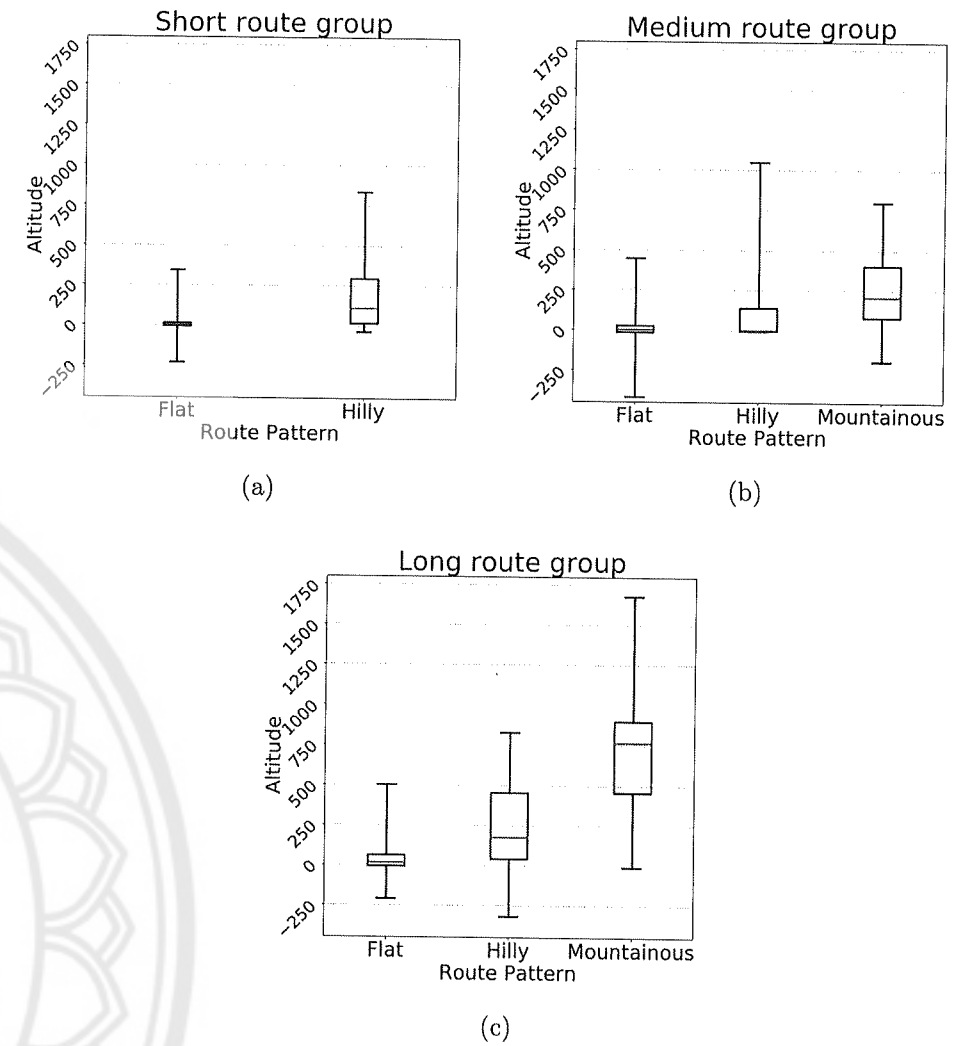


FIGURE 2. (a) The box plot for routes elevation by short-distance group, (b) the box plot for routes elevation by medium-distance group, and (c) the box plot for routes elevation by long-distance group

In addition, the reason for QB's high performance is the algorithm's complexity. The best case is when QB's complexity is linear time $O(N)$ with the number of items N while the worst-case is $O(N^2)$ when every cluster contains only 1 item [7]. The average case is $O(MN)$, where M is the number of clusters, but M is typically lower than N ; thus, we can neglect it and denote it as $O(N)$ because it is common in the complexity theory. In this experiment, QB takes 250 milliseconds to analyze and classify the route elevations and total distance of the entire dataset. The processing was done on Intel®Core™ i5-8350U CPU at 1.8GHz, 8GB of RAM, and the Windows 10 Home Edition 64bit.

5. Conclusion. This study proposed methods to cluster trajectory data by using QuickBundles as the primary technique. The selected data was recorded by GPS-devices and stored on open-access services, which we collected using the defined criteria. The collected data was preprocessed to correct the missing values, outliers, and error sensor readings.

We classified the routes by performing the following 2 steps: 1) grouping routes into 3 groups by their total distance using k-Means; 2) classifying routes by their elevation profiles using QuickBundles. The clustering validity was measured by using the silhouette coefficient method. QuickBundles successfully clusters route data according to their elevation profiles.

These results show that QuickBundles is an effective method for retrieving route patterns from GPS stored data, and recommending them to cyclists.

The suggested further research can be to enrich route patterns with additional features, such as the route's turns and route's terrain. These enriched route patterns can be used to propose tourism recommendation systems for different tourists who emphasize the routes by different routes' characteristics.

Acknowledgement. This research was funded by the National Research Council of Thailand under Grant R2561B074. In addition, many thanks to Mr. Peter G. Barton of the Naresuan University Language Centre for his editing assistance and advice on English expression in this document.

REFERENCES

- [1] J. Li, L. Xu, L. Tang, S. Wang and L. Li, Big data in tourism research: A literature review, *Tour. Manag.*, vol.68, pp.301-323, doi: 10.1016/j.tourman.2018.03.009, 2018.
- [2] S. Du, H. Zhang, H. Xu, J. Yang and O. Tu, To make the travel healthier: A new tourism personalized route recommendation algorithm, *J. Ambient Intell. Humaniz. Comput.*, vol.10, no.9, pp.3551-3562, doi: 10.1007/s12652-018-1081-z, 2019.
- [3] Z. A. Jiang, M. Wang and Y. Chen, Path recommendation based on geographic coordinates and trajectory data, *J. Commun.*, vol.38, no.5, pp.165-171, 2017.
- [4] S. Malik and D. Kim, Optimal travel route recommendation mechanism based on neural networks and particle swarm optimization for efficient tourism using tourist vehicular data, *Sustainability*, vol.11, no.12, doi: 10.3390/su11123357, 2019.
- [5] R. A. Hamid and M. S. Croock, A developed GPS trajectories data management system for predicting tourists' POI, *Telkonnika*, vol.18, no.1, pp.124-132, 2020.
- [6] F. M. Nardini, S. Orlando, R. Perego, A. Raffaetà, C. Renso and C. Silvestri, Analysing trajectories of mobile users: From data warehouses to recommender systems, in *A Comprehensive Guide through the Italian Database Research over the Last 25 Years*, S. Flesca, S. Greco, E. Masciari and D. Saccà (eds.), Cham, Springer International Publishing, 2018.
- [7] E. Garyfallidis, M. Brett, M. M. Correia, G. B. Williams and I. Nimmo-Smith, QuickBundles, a method for tractography simplification, *Front. Neurosci.*, vol.6, doi: 10.3389/fnins.2012.00175, 2012.
- [8] R. Lipe, *GPSBabel, Free Software for GPS Data Conversion and Transfer*, 2010.
- [9] P. L. Houtekamer and H. L. Mitchell, Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, vol.126, no.3, pp.796-811, 1998.
- [10] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, vol.20, pp.53-65, doi: 10.1016/0377-0427(87)90125-7, 1987.