

การเติมเต็มค่าข้อมูล GPS ที่ขาดหาย

GPS-Data Imputation



ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ปริญญานิพนธ์ปีการศึกษา 2563

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การเติมเต็มค่าข้อมูล GPS ที่ขาดหาย

GPS-DATA IMPUTATION

ผู้จัดทำ

1. นายพีรณชัย น้อมแนบ รหัสนักศึกษา 60010731
2. นายสวัสดิ์ บุญยฤทธิกิจ รหัสนักศึกษา 60010915



อาจารย์ที่ปรึกษา

(ผศ. ดร.รัฐชัย ชาวอุทัย)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

การเติมเต็มค่าข้อมูล GPS ที่ขาดหาย

นายพีรคนย์ น้อมแนบ 60010731
นายวสวัตต์ บุญยฤทธิกิจ 60010915
ผศ. ดร.รัฐชัย ชาวอุทัย อาจารย์ที่ปรึกษา
ปีการศึกษา 2563

บทคัดย่อ

การวิเคราะห์ข้อมูล GPS ของยานพาหนะเพื่อนำไปใช้งานในด้านการขนส่งหรือการจราจรนั้น จำเป็นต้องใช้ข้อมูลที่สมบูรณ์ แต่มีบางเหตุการณ์ที่ทำให้การจับเก็บข้อมูล GPS มีการขาดหาย ไม่ว่าจะปัญหานั้นจะมาจากอุปกรณ์ GPS ของยานพาหนะหรือเครือข่ายที่ใช้รับส่งสัญญาณ GPS มีปัญหา ดังนั้น การทำให้ข้อมูล GPS สมบูรณ์ก่อน จะช่วยในวิเคราะห์ได้ง่ายขึ้น

งานวิจัย การเติมเต็มค่าข้อมูล GPS ที่ขาดหาย ทำขึ้นเพื่อหาวิธีเติมข้อมูล GPS ที่ขาดหายไปในช่วง 10 นาทีขึ้นไป จะทดลองกับถนนรูปแบบทางตรง ทางโค้ง ทางแยกเช่น ทางสี่แยก, ทางสามแยก และ ทางเลี้ยวเมืองหรือบายพาส (Bypass) ซึ่งข้อมูล GPS ที่เติมในช่วงนั้นจะอ้างอิงกับข้อมูลถนนและข้อมูลรถยนต์ในอดีตในประเทศไทย โดยกระบวนการเติมข้อมูล GPS จะมีวิธีต่าง ๆ ได้แก่ 1) เติมด้วยข้อมูลจุดอ้างอิงถนน 100 เมตร, 2) เติมด้วยถนนอ้างอิงที่ได้จากการทำ K-Means Clustering, 3) เติมด้วยการสร้างจุดใหม่โดยอ้างอิงจากความเร็วรถและมุม Bearing นอกจากนี้จะมีการปรับปรุงวิธีเหล่านี้ในแบบต่าง ๆ อย่างการ Projection ข้อมูลจาก K-Means Clustering ให้อยู่บนถนนหรือการใช้ความเร็วจากการเทียบอัตราส่วน ส่วนทางบายพาสจะเปรียบเทียบทางที่นิยมเพื่อนำไปใช้ในการเติมจุดต่อไป และผลการเปรียบเทียบแต่ละ โมเดลในถนนแต่ละรูปแบบ โมเดลที่เติมด้วยข้อมูลจุดถนน 100 เมตรจะให้ผลที่ดีที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

GPS-Data Imputation

Mr. Peeradon	Nomnaeb	60010731
Mr. Wasawat	Boonyarittikit	60010915
Asst.Prof.Dr. Rathachai	Chawuthai	Advisor
Academic Year 2020		

ABSTRACT

A complete GPS dataset is required to analyze GPS data for the purpose of solving logistics or traffic problems. However, there are some circumstances that render the collected GPS data incomplete, such as GPS devices that do not function normally, data transmission networks that fail, and so on. As a result, filling in missing GPS data before using it will improve the accuracy of the analysis.

GPS-Data Imputation is a study that seeks methods to impute GPS data in missing intervals using road data and historical vehicle data of Thailand. In this study, four types of roads were chosen to be tested: straight roads, curve roads, intersections or crossroads, and bypasses. To fill missing GPS data, there are three main methods: 1) impute GPS data by 100m reference road, 2) impute GPS data by reference road from K-Means Clustering, and 3) impute GPS data by creating new GPS data from vehicle speed and Bearing angle. Furthermore, there are two methods that deviate from the main one, such as projecting a series of reference road from K-Means Clustering to lie on actual road and adjusting speed based on the proportion of speed in the traffic stream at a given time. There is a particular section for bypass that choose the route before filling the predicted point in missing intervals. When comparing each method, the best result is obtained by imputing GPS data by a 100m reference road.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

II

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

กิตติกรรมประกาศ

ปริญญาบัตรฉบับนี้สามารถสำเร็จได้ด้วยการช่วยเหลือจากบุคคลที่มีพระคุณหลาย ๆ ท่านที่ทำให้ทรัพยากร คำปรึกษาด้านวิชาการที่เป็นประโยชน์ จึงขอขอบพระคุณในการช่วยเหลือต่าง ๆ ณ ที่นี้

ผู้วิจัยขอขอบพระคุณ ผศ. ดร. รัฐชัย ชาวอุทัย อาจารย์ที่ปรึกษาของโครงการ ผู้ให้แนวทางในการดำเนินโครงการ ความรู้ที่จำเป็นต่อโครงการและช่วยตรวจพิจารณาโครงการ และสุดท้ายผู้วิจัยขอขอบพระคุณครอบครัว และคณาจารย์ภาควิศวกรรมคอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังทุกท่าน ที่ให้การสนับสนุนในด้านต่าง ๆ ทำให้ปริญญาบัตรฉบับนี้สำเร็จได้

ผู้วิจัยหวังว่าปริญญาบัตรฉบับนี้จะมีประโยชน์ต่อผู้ที่สนใจไม่มากนักน้อย หากมีข้อผิดพลาดไม่ว่าด้านใด ผู้วิจัยขออภัยในข้อผิดพลาดนั้นมา ณ ที่นี้ด้วย

พีรณย์ น้อมแนบ
วสวัตต์ บุญฤทธิกิจ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

III

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	I
บทคัดย่อภาษาอังกฤษ.....	II
สารบัญ.....	IV
สารบัญตาราง.....	VI
สารบัญรูปภาพ.....	VII
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาของปัญหา.....	1
1.2 วัตถุประสงค์ของโครงการ.....	2
1.3 ขอบเขตของโครงการ.....	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ.....	2
1.5 ขั้นตอนการดำเนินงาน.....	3
บทที่ 2 ทฤษฎีที่เกี่ยวข้อง.....	6
2.1 รายละเอียดของข้อมูลถนน.....	6
2.2 รายละเอียดของ GPS.....	13
2.3 งานวิจัยที่เกี่ยวข้อง.....	14
2.4 ทฤษฎีที่เกี่ยวข้อง.....	18
2.5 เครื่องมือที่ใช้.....	22
บทที่ 3 การออกแบบและพัฒนาระบบ.....	24
3.1 โมเดลการหาจังหวัดจากตำแหน่งของข้อมูล GPS.....	24
3.2 โมเดลการหาจุดอ้างอิงบนถนนในระยะ 100 เมตร.....	25
3.3 โมเดลการเติมเต็มตำแหน่ง GPS.....	27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

	หน้า
3.4 ฟังก์ชันในการเติมเต็มตำแหน่ง GPS ในช่วงข้อมูลที่หาย.....	44
บทที่ 4 การทดลองและผลการทดลอง.....	48
4.1 ผลการหาจังหวัดจากพิกัด GPS.....	48
4.2 ผลการหาจุดอ้างอิงบนถนนทุก 100 เมตร.....	49
4.3 ผลที่ได้จากโมเดลในการเติมจุดถนนที่ขาดหาย.....	51
บทที่ 5 สรุปผลและปัญหาของโครงการ.....	64
5.1 สรุปผล.....	64
5.2 ปัญหา อุปสรรค และแนวทางแก้ไข.....	65
5.3 สิ่งที่จะพัฒนาต่อในวิชาโครงการ 2.....	66
บรรณานุกรม.....	67
ภาคผนวก ก.....	69
ภาคผนวก ข.....	74

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง	หน้า
1.1 แผนการดำเนินการ โครงการงาน (ภาคเรียนที่1).....	3
1.2 แผนการดำเนินการ โครงการงาน (ภาคเรียนที่2).....	5
4.1 ผลเปรียบเทียบค่าต่าง ๆ ของแต่ละโมเดลที่ใช้หาจังหวัดของพิกัด GPS.....	49
4.2 ค่าเฉลี่ยระยะห่างของข้อมูลถนนตัวแทนที่สร้างด้วย K-Means กับข้อมูลถนนจริง.....	51
4.3 เปรียบเทียบผล MAE ของระยะทางและความเร็วของแต่ละ โมเดลกับทางตรง.....	53
4.4 เปรียบเทียบผล MAE ของระยะทางและความเร็วของแต่ละ โมเดลกับทางโค้ง.....	54
4.5 เปรียบเทียบผล MAE ของระยะทางและความเร็วของแต่ละ โมเดลกับทางแยก.....	55
4.6 เปรียบเทียบผล MAE ของระยะทางและความเร็วของแต่ละ โมเดลกับทางบายพาส.....	56
4.7 ผลการเปรียบเทียบระยะทาง, เวลาและจำนวนรถ ในจังหวัดขอนแก่น.....	57
4.8 ผลการเปรียบเทียบระยะทาง, เวลาและจำนวนรถ ในจังหวัดสระบุรี.....	58

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

VI

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

สารบัญรูปภาพ

รูปภาพ	หน้า
2.1 ป้ายทางหลวงพิเศษระหว่างเมือง.....	7
2.2 ป้ายทางหลวงแผ่นดินทั้ง 4 ระดับ.....	7
2.3 ตัวอย่างข้อมูลรถบรรทุก.....	9
2.4 ตัวอย่างข้อมูลรถแท็กซี่.....	10
2.5 ตัวอย่างข้อมูลถนนทางหลวง.....	11
2.6 ตัวอย่างข้อมูลถนนทางหลวงที่แปลงจาก shapefile เป็น csv file.....	12
2.7 ตัวอย่างข้อมูลหลักกิโลเมตร.....	12
2.8 กราฟตัวอย่างของ Logistic Regression.....	20
3.1 ตัวอย่างของการตี Grid.....	25
3.2 ขั้นตอนการสร้างข้อมูลถนนที่แต่ละจุดห่างกัน 100 เมตร.....	26
3.3 ข้อมูลที่มีช่วงเวลาที่ห่างกันมากกว่าหรือเท่ากับ 30 นาที.....	28
3.4 ตัวอย่างข้อมูล GPS ที่เต็มช่วงที่หายแล้ว.....	29
3.5 ตัวอย่างตารางข้อมูลอ้างอิงของถนน 100 เมตร.....	30
3.6 ขั้นตอนการสร้างตารางความเร็ว.....	32
3.7 ขั้นตอนทำงานในการเติมข้อมูล GPS ของโมเดลที่ 1.....	34
3.8 การนำเส้น A สร้างใหม่ลงบนเส้น B.....	36
3.9 ขั้นตอนทำงานในการเติมข้อมูล GPS ของโมเดลที่ 2.....	38
3.10 ขั้นตอนทำงานในการเติมข้อมูลของ โมเดลที่ 2 โดยขยายส่วนก่อนเติมข้อมูลในช่วงที่หาย.....	40
3.11 ขั้นตอนทำงานในการเติมข้อมูล GPS ของโมเดลที่ 3.....	42
3.12 ตัวอย่างถนนรูปแบบทางตรง.....	45
3.13 ตัวอย่างถนนรูปแบบทางโค้ง.....	45
3.14 ตัวอย่างถนนรูปแบบทางแยก.....	46
3.15 ตัวอย่างถนนรูปแบบทางบายพาส.....	46
4.1 ข้อมูลจุดอ้างอิงถนนที่ยังไม่ได้ทำพิคคี่ให้ห่างกันทุก 100 เมตร.....	50
4.2 ข้อมูลจุดอ้างอิงถนนที่ทำพิคคี่ให้อยู่ห่างกันทุก 100 เมตร.....	50

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับใช้ในเชิงวิชาการเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูปภาพ (ต่อ)

รูปภาพ	หน้า
4.3 ถนนอ้างอิงจากโมเดล K-Means ก่อนทำ Projection ลงบนถนน	51
4.4 ถนนอ้างอิงจากโมเดล K-Means หลังทำ Projection ลงบนถนน	52
4.5 ข้อมูลรถที่ผ่านทางบายพาสในระยะเวลา 1 เดือนในจังหวัดขอนแก่น	57
4.6 ข้อมูลรถที่ผ่านทางบายพาสในระยะเวลา 1 เดือนในจังหวัดสระบุรี	58
4.7 จำนวนค่า K ที่ใช้และระยะห่างของพิกัดที่ทำนายกับพิกัดจริงกับรูปแบบถนนทางตรง	59
4.8 จำนวนค่า K ใช้และเวลาของโมเดลที่ขณะทำนายผลกับรูปแบบถนนทางตรง	59
4.9 จำนวนค่า K ที่ใช้และระยะห่างของพิกัดที่ทำนายกับพิกัดจริงกับรูปแบบถนนทางโค้ง	60
4.10 จำนวนค่า K ใช้และเวลาของโมเดลที่ขณะทำนายผลกับรูปแบบถนนทางโค้ง	60
4.11 จำนวนค่า K ที่ใช้และระยะห่างของพิกัดที่ทำนายกับพิกัดจริงกับรูปแบบถนนทางแยก	61
4.12 จำนวนค่า K ใช้และเวลาของโมเดลที่ขณะทำนายผลกับรูปแบบถนนทางแยก	61
4.13 จำนวนค่า K ใช้และเวลาของโมเดลที่ใช้ขณะทำนายผลกับรูปแบบถนนทางบายพาส	62
4.14 จำนวนค่า K ใช้และเวลาของโมเดลที่ขณะทำนายผลกับรูปแบบถนนทางบายพาส	63
ก.1 ส่วนของถนน 205 ที่อยู่ในพื้นที่จังหวัดนครราชสีมา	69
ก.2 ส่วนของถนน 1 ที่อยู่ในพื้นที่จังหวัดนครสวรรค์	70
ก.3 ส่วนของถนน 2, 207, 2160 ที่อยู่ในพื้นที่จังหวัดนครราชสีมา	70
ก.4 ส่วนของถนน 214, 226, 293 ที่อยู่ในพื้นที่จังหวัดสุรินทร์	71
ก.5 ส่วนของถนน 4, 37 ที่อยู่ในพื้นที่จังหวัดประจวบคีรีขันธ์	72
ก.6 ส่วนของถนน 2, 202 ที่อยู่ในพื้นที่จังหวัดนครราชสีมา	73
ก.7 ส่วนของถนน 208, 230 ที่อยู่ในพื้นที่จังหวัดขอนแก่น	73
ข.1 ตัวอย่างเพิ่มเติมที่ 1 ของถนนรูปแบบทางตรง	74
ข.2 ตัวอย่างเพิ่มเติมที่ 2 ของถนนรูปแบบทางตรง	74
ข.3 ตัวอย่างเพิ่มเติมที่ 1 ของถนนรูปแบบทางโค้ง	75
ข.4 ตัวอย่างเพิ่มเติมที่ 2 ของถนนรูปแบบทางโค้ง	75
ข.5 ตัวอย่างเพิ่มเติมที่ 1 ของถนนรูปแบบทางแยก	76
ข.6 ตัวอย่างเพิ่มเติมที่ 1 ของถนนรูปแบบทางบายพาส	76

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

ในบทนี้จะกล่าวถึงจุดเริ่มต้นของโครงการ โดยเริ่มตั้งแต่ความเป็นมาของปัญหา วัตถุประสงค์และขอบเขตของโครงการ ประโยชน์ที่คาดว่าจะได้รับจากโครงการ และแผนดำเนินโครงการ

1.1 ความเป็นมาของปัญหา

ปัจจุบันการใช้งานการจราจรบนท้องถนนเป็นที่นิยมของคนจำนวนมากเนื่องจากการเดินทางที่มีความสะดวกสบาย และสามารถเข้าถึงได้ง่ายสำหรับทุกกลุ่มคน เพื่อที่จะอำนวยความสะดวกสำหรับการเดินทางให้ไปถึงจุดหมายได้อย่างรวดเร็วและแม่นยำ จึงมีความต้องการในการจัดการจราจรให้เกิดประสิทธิภาพสูงสุดจึงได้ มีการนำระบบที่สามารถระบุตำแหน่งต่าง ๆ ทั่วโลก หรือที่เรียกว่า Global Positioning System (GPS) มาใช้งานในการคำนวณหาเส้นทาง เพื่อให้ได้เส้นทางที่สั้นที่สุดหรือเส้นทางที่ใช้ระยะเวลาในการเดินทางน้อยที่สุด ทั้งในกลุ่มบุคคลธรรมดาที่ต้องการเดินทางและในกลุ่มธุรกิจการขนส่งต่าง ๆ เพื่อลดต้นทุนในการขนส่งสินค้า และยังมีเทคโนโลยีบางประเภทที่ต้องการความสมบูรณ์ของข้อมูล GPS เพื่อทำงานให้ได้เต็มประสิทธิภาพ เช่น VANET (Vehicular ad hoc networks) และ DTN (Delay/disruption Tolerant Networks) แต่การนำระบบ GPS มาใช้งานในการวิเคราะห์หรืออื่น ๆ นั้น ในระหว่างการใช้งานอาจจะมีปัญหาจากข้อมูลหรืออุปกรณ์ได้ และปัญหาที่สำคัญอย่างหนึ่งที่ทำให้การคำนวณเส้นทางนั้น เกิดความผิดพลาดได้ก็คือการขาดหายไปของข้อมูล GPS

ถึงแม้ว่าปัจจุบันระบบ GPS นั้นได้ครอบคลุมทุกพื้นที่ทั่วโลก แต่การขาดหายไปของ GPS อาจเกิดได้จากหลายสาเหตุ เช่น เกิดปัญหาที่อุปกรณ์ GPS บนพาหนะทำให้มีการส่งสัญญาณกลับมาที่เซิร์ฟเวอร์รับสัญญาณไม่ต่อเนื่อง เซิร์ฟเวอร์รับสัญญาณที่จัดเก็บข้อมูลมีปัญหาทำให้ข้อมูลที่ต้องการถูกจัดเก็บในช่วงเวลานั้นหายไปหรือขัดข้อง เครือข่ายอินเทอร์เน็ตที่ใช้รับส่งข้อมูล GPS มีปัญหาทำให้ข้อมูลขาดหายไป และนอกเหนือจากสาเหตุอาจจะเกิดจากตัวคนขับปิด GPS ด้วยตัวเอง พื้นที่พาหนะอยู่เป็นบริเวณที่อับสัญญาณ ซึ่งทางผู้จัดทำได้เห็นความสำคัญของปัญหาการขาดหายไปของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ข้อมูล GPS นี้ และได้ศึกษาแนวทางในการแก้ปัญหาโดยได้นำความรู้ในด้านการวิเคราะห์ข้อมูล (Data Analytics) มาปรับใช้งานเพื่อเติมส่วนที่หายไปของข้อมูล GPS

โครงการการเติมเต็มค่าข้อมูล GPS ที่ขาดหาย (GPS-Data Imputation) นี้จึงถูกจัดทำขึ้น เพื่อรวบรวมและศึกษาข้อมูล GPS ของรถที่ใช้ในการจราจรบนท้องถนน และทำการทำนายข้อมูลของ GPS ในส่วนที่หายไปโดยจะทำการทำนายทั้งพิกัดของ GPS ที่หายไปและความเร็วของรถ ณ ตำแหน่งพิกัดนั้นด้วย เพื่อจะได้นำข้อมูลของ GPS ที่สมบูรณ์ครบถ้วนนี้ไปใช้ในการคำนวณ หรือวิเคราะห์สิ่งต่างๆ ให้เกิดความถูกต้องและแม่นยำขึ้น เพื่อให้เกิดประโยชน์ต่อนักพัฒนาในการต่อไป

1.2 วัตถุประสงค์ของโครงการ

- 1) เพื่อออกแบบโครงสร้างข้อมูลสำหรับจัดเก็บข้อมูลถนนและการจราจร
- 2) เพื่อวิเคราะห์หารูปแบบเชิงเวลาของปริมาณจราจรและอัตราการไหลของแต่ละจุดบนท้องถนน
- 3) เพื่อทำนายและเติมเต็มจุด GPS ที่หายไปบนถนน

1.3 ขอบเขตของโครงการ

- 1) พื้นที่ที่นำมาศึกษา ได้แก่ ถนนทางหลวงแผ่นดินสายหลัก 11 สาย และทางเลี้ยว
- 2) ประเภทรถที่นำข้อมูลมาศึกษา ได้แก่ รถบรรทุก, รถโดยสารไม่ประจำทาง, รถแท็กซี่
- 3) ช่วงเวลาที่นำมาศึกษาเป็นเวลา 4 ปี คือ ปีพุทธศักราชที่ 2560, 2561, 2562, 2563
- 4) ค่า MAE (Mean Absolute Error) ของระยะห่างที่ผิดพลาดของจุดที่เติมเต็มไม่เกิน 1 กิโลเมตรและความเร็วไม่เกิน 10 กิโลเมตรต่อชั่วโมง

1.4 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ทำโมเดลที่สามารถเติมเต็มข้อมูลพิกัด GPS ที่ขาดหายไปได้
- 2) ได้ข้อมูลพิกัด GPS ที่สมบูรณ์ขึ้นจากการเติมช่วงที่หายไป เพื่อนำข้อมูลไปวิเคราะห์ในแต่ละด้าน เช่น พฤติกรรมของคนขับรถในการเลือกเส้นทาง สถานที่ท่องเที่ยวที่คนนิยมไปในช่วงเวลาหนึ่ง เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- 3) เข้าใจการใช้งานอัลกอริทึมและการใช้งานสมการทางคณิตศาสตร์ต่างๆที่เกี่ยวข้องกับพิกัด GPS เช่น การหาระยะทางระหว่างพิกัด GPS เป็นต้น
- 4) ทำความรู้จักข้อมูลต่าง ๆ ของถนนในประเทศไทย เช่น ระบบหมายเลข และรูปแบบของถนน

1.5 ขั้นตอนการดำเนินงาน

- 1) ทำโมเดลในการหาจังหวัดของข้อมูลพิกัด GPS
- 2) ทำข้อมูลหลักถนน 100 เมตรจากข้อมูลถนนในประเทศไทยที่มี
- 3) ทำตารางบอกความเร็วรถยนต์ในแต่ละจุดบนถนน โดยแบ่งที่ละ 1 ชั่วโมง, 30 นาที, 15 นาที
- 4) เตรียมข้อมูลสำหรับทดสอบ โดยใช้ข้อมูลที่มีเวลาห่างกันตั้งแต่ 10 -30 นาทีขึ้นไป
- 5) ศึกษางานวิจัยอื่นที่เกี่ยวข้อง
- 6) ทำโมเดลทำนายจุดในช่วงที่หายไปบนถนน ในวิธีต่าง ๆ รวมถึงโมเดลที่อ้างอิงจากงานวิจัย
- 7) เปรียบเทียบการวัดผลจากแต่ละโมเดล
- 8) วิเคราะห์และเลือกแนวทางพัฒนา โมเดลที่ให้ผลลัพธ์ที่ดีกว่า
- 9) ทำโมเดลทำนายถนนรูปแบบบายพาส และวัดผลที่ได้
- 10) ทำฟังก์ชันสำหรับใช้เติมจุดในช่วงที่หายไปบนถนน

ตารางที่ 1.1 แผนการดำเนินการโครงการ (ภาคเรียนที่1)

หัวข้อกิจกรรม	สิงหาคม				กันยายน				ตุลาคม					พฤศจิกายน				
	1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4	
หาจังหวัดของพิกัด GPS																		
ทำข้อมูลถนนที่แต่ละจุดห่างกัน 100 เมตร																		
ทำตารางบอกความเร็วของแต่ละจุดบนถนน																		

เอกสารนี้เป็นเอกสารลิขสิทธิ์สงวนไว้สำหรับการใช้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามเผยแพร่ต่อสาธารณะ และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

หัวข้อกิจกรรม	สิงหาคม				กันยายน				ตุลาคม					พฤศจิกายน			
	1	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4
เตรียมข้อมูลสำหรับทดสอบที่มีช่วงที่หายไปมากกว่า 30 นาที																	
ทำโมเดลสร้างจุดในช่วงเวลาที่ห่างกัน 30 นาทีขึ้นไป																	
วัดความถูกต้องของตำแหน่งจุดที่สร้างกับข้อมูลจริงๆ																	
ศึกษางานวิจัยอื่นที่เกี่ยวข้อง																	
สร้างโมเดลของงานวิจัยที่จะนำมาเปรียบเทียบ																	
นำโมเดลที่สร้างไปเปรียบเทียบผลกับงานวิจัยที่ได้ศึกษา																	
วิเคราะห์และเลือกแนวทางพัฒนาโมเดลที่ให้ผลลัพธ์ที่ดีกว่า																	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตารางที่ 1.2 แผนการดำเนินการโครงการ (ภาคเรียนที่2)

หัวข้อกิจกรรม	มกราคม				กุมภาพันธ์				มีนาคม				
	1	2	3	4	1	2	3	4	1	2	3	4	5
ปรับปรุงให้โมเดลทำนาย ความเร็วได้ใกล้เคียงมาก ขึ้น													
ทำโมเดลเติมจุดสำหรับ ถนนบายพาส													
ปรับปรุงโมเดลสำหรับ ถนนบายพาส													
ทำเซอร์วิสฟังก์ชันในการ เติมจุด													

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงรายละเอียดของข้อมูลที่ใช้ในโครงการนี้ เช่น ระบบตัวเลขของถนนหรือความหมายข้อมูลของรถประเภทต่าง ๆ รายละเอียดของงานวิจัยที่นำมาใช้อ้างอิงและเปรียบเทียบผลการทดลอง และทฤษฎีกับเครื่องมือที่เกี่ยวข้องที่คณะผู้จัดทำได้ศึกษาค้นคว้าเพื่อนำมาประกอบโครงการนี้ ซึ่งสามารถดูได้ตามหัวข้อย่อต่อไปนี้

2.1 รายละเอียดของข้อมูลถนน

ในหัวข้อนี้จะอธิบายระบบตัวเลขของหมายเลขถนนทางหลวงในประเทศไทย และรายละเอียดของข้อมูลหลักถนนและข้อมูลรถยนต์ที่ใช้ โดยจะแบ่งเนื้อหาตามหัวข้อย่อต่อไปนี้

2.1.1 ระบบตัวเลขของถนนในประเทศไทย

ระบบตัวเลขทางหลวงในประเทศไทยแบ่งออกเป็น 5 ประเภทหลัก ได้แก่ ทางหลวงพิเศษ, ทางหลวงแผ่นดิน, ทางหลวงชนบท, ทางหลวงท้องถิ่น และทางหลวงสัมปทาน

2.1.1.1 ทางหลวงพิเศษ

ทางหลวงพิเศษ หรือ ทางหลวงพิเศษระหว่างเมือง คือทางหลวงที่มีจุดประสงค์เพื่อให้การจราจรผ่านได้ตลอดอย่างรวดเร็วเป็นพิเศษ มีกรมทางหลวงเป็นผู้จัดการเรื่องการสร้าง การขยายถนน การบำรุงรักษา รวมถึงมีการควบคุมจุดเข้าออกทางหลวงของรถยนต์ และจัดเก็บค่าผ่านทาง (ป้ายสีน้ำเงิน)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูปที่ 2.1 ป้ายทางหลวงพิเศษระหว่างเมือง
ที่มา <https://www.wikiwand.com/th/ทางหลวงในประเทศไทย>

2.1.1.2 ทางหลวงแผ่นดิน

ทางหลวงแผ่นดิน เป็นเส้นทางสาธารณะสายหลักที่เป็นโครงข่ายเชื่อมระหว่างภาค จังหวัด อำเภอ ไปจนถึงสถานที่ที่สำคัญต่างๆ เพื่อให้การเดินทางและขนส่งเป็นไปได้อย่างสะดวกยิ่งขึ้น โดยมีกรมทางหลวงในสังกัดกระทรวงคมนาคม เป็นผู้รับผิดชอบและดูแลโครงการ



ก)



ข)



ค)



ง)

รูปที่ 2.2 ป้ายทางหลวงแผ่นดินทั้ง 4 ระดับ

- ก) ทางหลวงแผ่นดิน 1 หลัก
- ข) ทางหลวงแผ่นดิน 2 หลัก
- ค) ทางหลวงแผ่นดิน 3 หลัก
- ง) ทางหลวงแผ่นดิน 4 หลัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงที่มาของเอกสารทุกครั้งที่มีการนำไปใช้
ที่มา <https://www.wikiwand.com/th/ทางหลวงในประเทศไทย>

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.1.1.3 ทางหลวงชนบท

ทางหลวงชนบทจะมีกรมทางหลวงชนบทเป็นผู้จัดการก่อสร้าง ขยาย บูรณะ และบำรุงรักษา และอธิบดีกรมทางหลวงชนบทเป็นผู้จัดให้ลงทะเบียนไว้เป็นทางหลวงที่กรมทางหลวงชนบท

การตั้งชื่อทางหลวงชนบท จะอ้างอิงจากหมายเลขทางหลวงแผ่นดิน หรือชื่อหมู่บ้าน หรือชื่อสถานที่สำคัญ โดยที่บริเวณนั้นเป็นจุดเริ่มต้นและจุดสิ้นสุดทางหลวงนั้นจึงนำมาตั้งชื่อถนน เมื่อมีการพัฒนาระบบโครงข่ายมากขึ้น การใช้ชื่อทางหลวงอย่างเดียวอาจจะก่อให้เกิดการสับสน และไม่สามารถทราบว่าทางหลวงนั้นอยู่ในจังหวัดใด จึงมีการนำรหัสของแต่ละสายทางเข้ามาเป็นตัวบอกถึงที่ตั้ง และลำดับของสายทาง ซึ่งรหัสสายทางของทางหลวงชนบทประกอบด้วยตัวอักษรย่อของจังหวัด 2 ตัว และตัวเลข 4 ตัว มาใช้กำกับชื่อของแต่ละทางหลวงชนบท

2.1.1.4 ทางหลวงท้องถิ่น

ทางหลวงท้องถิ่น คือ ทางหลวงที่องค์กรปกครองส่วนท้องถิ่นเป็นผู้ดำเนินการก่อสร้าง ขยาย บูรณะ และบำรุงรักษา และได้ผู้ว่าราชการจังหวัดเป็นผู้จัดให้ลงทะเบียนไว้เป็นทางหลวงท้องถิ่นไว้ ณ ศาลากลางจังหวัด มีการตั้งชื่อรหัสสายทางจะตัวอักษรไทย 3 ตัว แล้วตามด้วยตัวเลข 5 หลัก

2.1.1.5 ทางหลวงสัมปทาน

ทางหลวงสัมปทาน คือ ทางหลวงที่กรมทางหลวงได้ให้เอกชนถือสัมปทานไว้ และได้ลงทะเบียนไว้เป็นทางหลวงสัมปทาน ซึ่งอธิบดีกรมทางหลวงเป็นผู้จัดให้ลงทะเบียนไว้ ณ กรมทางหลวง ในสรุพนियามความหมายของทางหลวงสัมปทานจากพระราชบัญญัติทางหลวงสัมปทาน พุทธศักราช 2542 คือทางหลวงที่รัฐให้สัมปทานแก่เอกชนในการสร้าง บำรุงรักษา โดยสามารถเก็บค่าใช้ทาง

ทางหลวงสัมปทาน จะมีระบบหมายเลขทางหลวงเหมือนกับทางหลวงแผ่นดิน หรือทางหลวงพิเศษระหว่างเมืองอย่างไรอย่างหนึ่ง ขึ้นอยู่กับลักษณะของทางหลวงสัมปทานนั้นว่า มีลักษณะเป็นทางหลวงแผ่นดินหรือทางหลวงพิเศษระหว่างเมือง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.1.2 รายละเอียดของข้อมูลที่ใช้

2.1.2.1 ข้อมูลของรถบรรทุกและรถโดยสาร

time_stamp	vid	lat	lon	speed	unit_type	city
01/01/2019 0:00	MDI	14.569848	100.9812	54	7	Saraburi
01/01/2019 0:00	MDM	14.576211	100.76735	16	8	Saraburi
01/01/2019 0:00	MDC	14.597548	100.896776	66	3	Saraburi
01/01/2019 0:00	MDM	14.31745	100.838073	56	7	Saraburi
01/01/2019 0:00	MDI	14.304551	100.82953	74	6	Saraburi
01/01/2019 0:00	MDA	14.681817	100.894565	54	7	Saraburi
01/01/2019 0:00	MDM	14.520075	100.91513	44	6	Saraburi

รูป 2.3 ตัวอย่างข้อมูลรถบรรทุก

จัดเก็บอยู่ในรูปแบบของ zip โดยจะมีรายละเอียดข้อมูลดังต่อไปนี้

- คอลัมน์ที่ 1 time_stamp หมายถึง วันที่และเวลาที่เก็บข้อมูลในรูปแบบ ปี-เดือน-วันที่ ชั่วโมง:นาที:วินาที
- คอลัมน์ที่ 2 unit_id หรือ vid หมายถึง Vehicle ID หรือชุดข้อความประจำตัวของรถแต่ละคัน ซึ่งจะไม่ซ้ำกัน
- คอลัมน์ที่ 3 lat หมายถึง พิกัด Latitude ซึ่งอยู่ในรูปแบบของ องศาศนิยม หรือ Decimal Degree
- คอลัมน์ที่ 4 lon หมายถึง พิกัด Longitude ซึ่งอยู่ในรูปแบบของ องศาศนิยม หรือ Decimal Degree
- คอลัมน์ที่ 5 speed หมายถึง ความเร็วของรถ มีหน่วยเป็นกิโลเมตรต่อชั่วโมง
- คอลัมน์ที่ 6 unit_type หมายถึง ตัวเลขแทนประเภทของรถโดยจะใช้ 1, 3 ถึง 9 ซึ่งมีความหมายดังนี้
 - 1 หมายถึง รถโดยสารประจำทาง
 - 3 หมายถึง รถโดยสารไม่ประจำทาง
 - 4 หมายถึง รถโดยสารส่วนบุคคล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- 5, 8, 9 หมายถึง รถบรรทุกส่วนบุคคล
- 6, 7 หมายถึง รถบรรทุกไม่ประจำทาง

2.1.2.2 ข้อมูลของรถแท็กซี่

time_stamp	vid	lat	lon	speed	heading	for_hire_light	engine_acc	gpsvalid
01/03/2017 0:00	IzDPC	14.86668	101.71465	110	75	0	1	1
01/03/2017 0:00	4q9GM	15.08166	102.22885	54	59	0	1	1
01/03/2017 0:00	04kLx	14.65818	101.37789	111	83	1	1	1
01/03/2017 0:01	IzDPC	14.86972	101.7269	65	75	0	1	1
01/03/2017 0:01	4q9GM	15.08592	102.23616	54	58	0	1	1
01/03/2017 0:02	04kLx	14.66385	101.41121	14	8	1	1	1
01/03/2017 0:02	jhlMfb	14.7734	101.52162	79	193	1	1	1
01/03/2017 0:02	IzDPC	14.87034	101.73684	79	92	0	1	1

รูป 2.4 ตัวอย่างข้อมูลรถแท็กซี่

จัดเก็บในรูปแบบของ tar.bz2 โดยจะมีรายละเอียดข้อมูลดังนี้

- คอลัมน์ที่ 1 vid หมายถึง Vehicle ID หรือชุดข้อความประจำตัวของรถแท็กซี่แต่ละคัน ซึ่งจะไม่ซ้ำกันในแต่ละคัน
- คอลัมน์ที่ 2 gpsvalid หมายถึง ตัวรับส่ง GPS ของรถแท็กซี่ทำงานปกติ
- คอลัมน์ที่ 3 lat หมายถึง พิกัด Latitude ซึ่งอยู่ในรูปแบบของ องศาศนิยม หรือ Decimal Degree โดยจะมีทศนิยมไม่เกิน 5 ตำแหน่ง
- คอลัมน์ที่ 4 lon หมายถึง พิกัด Longitude ซึ่งอยู่ในรูปแบบของ องศาศนิยม หรือ Decimal Degree โดยจะมีทศนิยมไม่เกิน 5 ตำแหน่ง
- คอลัมน์ที่ 5 timestamp หมายถึง วันที่และเวลาที่รถแท็กซี่ส่งสัญญาณ GPS ไป ข้อมูลอยู่ในรูปแบบ ปี-เดือน-วันที่ ชั่วโมง:นาที:วินาที
- คอลัมน์ที่ 6 speed หมายถึง ความเร็วของรถแท็กซี่ที่เวลานั้น หน่วยกิโลเมตรต่อ ชั่วโมง
- คอลัมน์ที่ 7 heading หมายถึง ทิศทางที่รถเคลื่อนที่ไป (0-360) โดยทิศเหนือคือ 0

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาค้นคว้าเท่านั้น เมื่ออนุญาตเห็นใบเซปประโชยชนดานการค้ำ
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- คอลัมน์ที่ 8 for_hire_light หมายถึง ไฟบอกสถานะมีผู้โดยสารหรือไม่ โดย 1 คือ ไฟติดและมีผู้โดยสาร, 0 คือ ไฟดับและมีผู้โดยสาร
- คอลัมน์ที่ 9 engine_acc หมายถึง เครื่องยนต์รถกำลังทำงาน ซึ่ง 1 คือ เครื่องยนต์กำลังทำงานและเก็บข้อมูลทุก 1 นาที, 0 คือ เครื่องยนต์ดับและเก็บข้อมูลทุก 3 นาที

2.1.2.3 ข้อมูลของถนน

ข้อมูลถนนแต่เดิมถูกจัดเก็บไว้ในรูปแบบ shapefile และผู้จัดทำได้แปลงข้อมูลนี้ให้อยู่ในรูปแบบ csv เพื่อให้ง่ายต่อการใช้งาน

ROAD_CODE	ROAD_NAME	SECTION_ID	ROAD_ID	CODE	NAME	KM_START	KM_END	LENGTH	DISTRICT	C	REMARK	REVISION	STATUS	LENGTH_GPS	ADMIT_DATE	SEC_CODE	LAYER_ID	LAYER	geometry
0	3309	...	5491	5143	201	...	6000	21800	15.8	416	...	7	A	16.083	01/01/2011 0:00	3309-0201	1	section	LINestring (100.5551478632897 14.119858512758
1	3901	...	5493	5300	102	...	19972	31872	11.9	419	...	7	A	12.782	01/01/2011 0:00	3901-0102	1	section	LINestring (100.4097419643706 13.790814409682
2	3412	...	5448	5214	101	...	0	8943	8.943	413	...	7	A	8.943	01/01/2011 0:00	3412-0101	1	section	LINestring (100.5419322135922 14.350482907547
3	3632	...	5466	5737	100	...	0	100	0.1	413	...	7	A	0.107	01/01/2011 0:00	3632-0100	1	section	LINestring (100.4858417351827 14.362391923248
4	354	...	5494	5713	100	...	0	945	0.945	417	...	7	A	0.699	01/01/2011 0:00	354-0100	1	section	LINestring (100.7195106627391 13.665966674758
5	3901	...	5900	5300	101	...	0	19972	19.972	419	...	7	A	19.195	01/01/2011 0:00	3901-0101	1	section	LINestring (100.4063260408377 13.675957513274
6	347	...	3164	4342	101	...	0	15881	15.881	416	...	7	A	15.882	01/01/2011 0:00	0347-0101	1	section	LINestring (100.5500018387161 14.014786593935
7	3508	...	5503	3271	100	...	0	4300	4.3	416	...	7	A	4.317	01/01/2011 0:00	3508-0100	1	section	LINestring (100.6071906079019 13.948444812992
8	3591	...	5504	5637	101	...	0	800	0.8	416	...	7	A	0.786	01/01/2011 0:00	3591-0101	1	section	LINestring (100.5698778302436 14.118317339465
9	3454	...	1987	5239	400	...	86609	111609	25	413	...	7	A	24.972	01/01/2011 0:00	3454-0400	1	section	LINestring (100.3663772468349 14.495185179623

รูป 2.5 ตัวอย่างข้อมูลถนนทางหลวงในรูปแบบ shapefile

ในส่วน of shapefile ข้อมูลที่ใช้จะมีอยู่ 2 คอลัมน์ คือ ROAD_CODE หรือหมายเลขถนน และ geometry ซึ่งจะเก็บ LINestring ของถนนไว้ในรูปแบบคู่อันดับของ ลองจิจูด, ละติจูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

```
rd;rd_type;polyline
3309;DOH;[(14.1198585127587055,100.5551478632897044),(14.1198921538952149,100.5551333992303853),
3901;DOH;[(13.7908144096850425,100.4097419643706246),(13.7901031209355995,100.4096894267690203),
3412;DOH;[(14.3504829075473808,100.5419322135922044),(14.3508412394895242,100.5420259172366002),
3632;DOH;[(14.3623919232485751,100.4858417351827455),(14.3622799676644384,100.4858004828622029),
354;DOH;[(13.6659666747585220,100.7195106627390970),(13.6656511995044667,100.7255658255450186)]
3901;DOH;[(13.6759575132748434,100.4063260408377261),(13.6751576221377018,100.4062720705277059),
0347;DOH;[(14.0147865939335610,100.5500018387160850),(14.0149880739550294,100.5501266651174177),
3508;DOH;[(13.9484448129921894,100.6071906079018987),(13.9484597620263262,100.6071886013174463),
3591;DOH;[(14.1183173394656372,100.5698778302435983),(14.1184241509916450,100.5697410941146330),
3454;DOH;[(14.4951851796235118,100.3663772468349435),(14.4951330999996859,100.3663711944368089),
0303;DOH;[(13.6861384498892438,100.4867763000834771),(13.6859890500171737,100.4869498998120321),
0308;DOH;[(14.1823605811365265,100.6137811525154433),(14.1833593679052630,100.6124561864895810),
3470;DOH;[(14.4517933167467927,100.7255688916307719),(14.4518553399937115,100.7256239178146728),
3902;DOH;[(14.1895914508565610,100.6365215099976211),(14.1893816355541489,100.6367222604639835),
3415;DOH;[(13.7455413732865619,100.2298910553207918),(13.7456938317293389,100.2298887725686996),
3702;DOH;[(13.6196652295770466,100.9220266169890152),(13.6195909160105995,100.9221453250450367),
3702;DOH;[(13.6764919988991522,100.8365260524178524),(13.6751330355046026,100.8372763074782341),
3702;DOH;[(13.6659227406546044,100.8499169438759822),(13.6658728198347070,100.8499949038274508),
```

รูป 2.6 ตัวอย่างข้อมูลถนนทางหลวงที่แปลงจาก shapefile เป็น csv file

ข้อมูลที่แปลงมาจาก shapefile จะถูกเก็บไว้ 3 คอลัมน์

- rd คือ หมายเลขถนน
- rd_type คือ ประเภทของถนน
- polyline คือ พิกัดละติจูดและลองจิจูดของถนนซึ่งจะเก็บไว้ในอาร์เรย์ (Array) ของ Tuple ของพิกัดละติจูด, ลองจิจูด

rd	km	lat	lon
11	132	16.15467658	100.6225711
11	133	16.16319491	100.6257612
11	134	16.17225845	100.6265008
11	135	16.18133104	100.62716
11	136	16.19040296	100.627838
11	136	16.19040296	100.627838
11	137	16.19947594	100.6285041
11	138	16.20855004	100.6291732
11	139	16.21762626	100.6298019

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

รูป 2.7 ตัวอย่างข้อมูลหลักกิโลเมตร

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

โดยที่แต่ละคอดัมนี่มีความหมายดังนี้

- rd คือ หมายเลขถนน
- km คือ เลขหลักกิโลเมตร
- lat , lon คือ พิกัดละติจูดและลองจิจูดของหลักกิโลเมตร

2.2 รายละเอียดของ GPS

ในหัวข้อนี้จะอธิบายว่า GPS คืออะไร และประกอบจากอะไรบ้าง

2.2.1 GPS คืออะไร

GPS คือ ระบบระบุตำแหน่งบนพื้นโลก ย่อมาจากคำว่า Global Positioning System ซึ่งจะประกอบไปด้วย 3 ส่วนหลัก คือ

1. ส่วนอวกาศ จะมีหรือข่ายดาวเทียมหลักจาก 3 ประเทศ คือ อเมริกา รัสเซีย ยุโรป

เครือข่ายของอเมริกา ชื่อ NAVSTAR (Navigation Satellite Timing and Ranging GPS) มีดาวเทียมที่ใช้จริงและสำรอง บริหารงานโดย Department of Defenses

เครือข่ายของยุโรป ชื่อ Galileo บริหารงานโดย ESA หรือ European Satellite Agency

เครือข่ายของรัสเซีย ชื่อ GLONASS หรือ Global Navigation Satellite บริหาร โดย Russia VKS (Russia Military Space Force)

ณ วันที่ค้นข้อมูล รัฐบาลสหรัฐอเมริกาได้เปิดให้ภาคประชาชนทั่วโลกใช้ข้อมูลจากดาวเทียมของทางอเมริกาได้ฟรี เฉพาะข้อมูลที่ไม่ส่งผลกระทบต่อความมั่นคงของอเมริกา โดยจะมีระยะผิดพลาดของความแม่นยำประมาณบวก / ลบ 10 เมตร

2. ส่วนควบคุม จะมีแค่สถานีภาคพื้นดิน โดยมีสถานีใหญ่อยู่ที่ประเทศอเมริกา และศูนย์ควบคุมย่อยอีก 5 จุด ในภูมิภาคต่าง ๆ ทั่วโลก
3. ส่วนผู้ใช้งาน ผู้ใช้งานสัญญาณดาวเทียมจะต้องมีเครื่องรับสัญญาณที่รับคลื่นและแปลรหัสจากดาวเทียมเพื่อนำสัญญาณนั้นมาใช้งานในรูปแบบต่างๆ อย่างเหมาะสม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.3 งานวิจัยที่เกี่ยวข้อง

ในหัวข้อจะกล่าวถึงงานวิจัยที่ได้ทำการศึกษาและนำมาพัฒนาต่อ ดังหัวข้อย่อยต่อไปนี้

2.3.1 Filling the Gaps of Vehicular Mobility Traces

งานวิจัยนี้ถูกจัดทำขึ้นโดย Fabrício A. Silva , Clayson Celes , Azzedine Boukerche , Linnyer B. Ruiz , Antonio A. F. Loureiro โดยงานวิจัยนี้ถูกคิดค้นมาเพื่อนำข้อมูลไปใช้ใน VANET (Vehicular ad hoc networks) และ DTN (Delay/disruption Tolerant Networks) ซึ่ง 2 เทคโนโลยีนี้จำเป็นต้องใช้ข้อมูลที่มีความต่อเนื่อง ไม่มีการสูญหายเพื่อให้ได้ผลลัพธ์ที่มีประสิทธิภาพที่ดีที่สุด โดยข้อมูลในที่นี้หมายถึงตำแหน่งพิกัดของ GPS ของรถแท็กซี่ ซึ่งในงานวิจัยนี้ได้เสนอวิธีแก้ไข ปัญหา โดยใช้อัลกอริทึมอยู่ 2 ส่วน ส่วนที่ 1 ใช้อัลกอริทึมประเภท Clustering ในการหาพิกัดที่จะนำไปเติม ข้อมูลที่หายไป โดยจะบันทึกเก็บลงใน RefSys และในส่วนที่ 2 จะนำ RefSys ไปเติมแทนที่ ส่วนข้อมูลที่หายไป ซึ่งในงานวิจัยนี้ได้ทำการทดลองกับพื้นที่ถนนที่แตกต่างกัน โดยจะเลือกพื้นที่มา 3 ภูมิภาค ได้แก่ 1) San Francisco, USA 2) Rome, Italy 3) Shanghai and Beijing, China

ข้อมูลที่ใช้ในงานวิจัยนี้จะใช้ข้อมูล GPS ของรถแท็กซี่ โดยข้อมูลของ San Francisco, USA มีอยู่ทั้งหมด 11,219,955 จุด จากรถ 536 คัน, ข้อมูลของ Rome, Italy มีอยู่ทั้งหมด 21,817,851 จุด จากรถ 316 คัน และข้อมูลของ Shanghai, China มีอยู่ทั้งหมด 6,075,587 จุด จากรถ 4,316 คัน

อัลกอริทึมในส่วนที่ 1 Cluster-based Reference System ใช้ในการหา RefSys ซึ่งจะเป็นข้อมูล อ้างอิงตำแหน่งถนนที่จะนำไปเติมในช่วงข้อมูลที่หายไป โดยจะใช้วิธี K-Means Clustering Method ในการหา ซึ่งอาจจะมีปัญหาในการกำหนดค่า K ของ Cluster ทางผู้วิจัยได้เสนอการใช้ Elbow Method ในการแก้ปัญหานี้ จะใช้วิธีในการหาดังนี้

- Input คือ ข้อมูลของรถที่ได้เก็บไว้ทั้งหมด (Raw Data) และ จำนวน Clusters (K)
- ทำการ Clustering Raw Data โดยจะใช้ (raw_data , K)
- นำจุด Centroids ของ Clusters มาเก็บไว้ใน RefSys
- Output คือ RefSys ซึ่งจะนำไปใช้ในอัลกอริทึมในส่วนที่ 2 ต่อไป

โดย K-Means Clustering Method นี้จะให้ $O(nkdi)$ เมื่อ n คือ จำนวน sample, k คือ จำนวน Cluster, i คือ จำนวนรอบในการทำ Clustering และ d คือ จำนวน Dimensions ซึ่งในที่นี้มี 2 Dimensions คือ Latitude, Longitude

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

อัลกอริทึมในส่วนที่ 2 Calibration Method ใช้ในการหาช่วงข้อมูลที่หายไปและทำการเติมช่วงที่หายไปนั้นด้วย RefSys โดยจะมีการกำหนดค่าดังต่อไปนี้

min_d คือ ระยะห่างที่น้อยที่สุดที่จะถือว่าช่วงนั้นหายไป

max_d คือ ระยะห่างที่มากที่สุดที่จะถือว่าช่วงนั้นหายไป

time_d คือ เวลาที่น้อยที่สุดที่จะถือว่าช่วงนั้นหายไป โดยจะใช้วิธีการดังนี้

- Input คือ ข้อมูลของรถคันที่ต้องการนำมาเติมช่องว่าง T (columns: time_stamp, vid, lat, lon), RefSys (columns: lat, lon), min_d, max_d, time_d
- Output คือ T' ซึ่งจะมีข้อมูลเส้นทางของรถที่ไม่มีช่องว่าง
- เก็บ T[1] ไว้ใน T'
- วง loop เพื่อหาช่วงที่หายไป ซึ่งจะเช็คจากระยะห่างระหว่างจุดว่าอยู่ในช่วง min_d, max_d และ เวลาที่ต่างกันมากกว่า time_d หรือไม่ ถ้าอยู่ในช่วงนั้นก็ถือว่ามิจุดหายไป และทำการเก็บจุดที่เริ่มหายและจุดสิ้นสุดของการหายเอาไว้ใน Pp, Pn ตามลำดับ ในกรณีที่ไม่มีจุดหายไปก็จะเก็บจุดนั้นไว้ใน T' และทำการเช็คจุดถัดไป
- หา RefSys ที่อยู่ระหว่าง Pp, Pn
- นำ RefSys ที่อยู่ใกล้กับ Pp ที่สุด ไปเทียบมุมระหว่าง Pp, Pn เพื่อหาจุดที่จะนำไปเติมใน T' ในทิศทางที่ถูกต้อง โดยจะเก็บจุดไว้ใน List L ก่อน ถ้ามุมที่เทียบไปคนละทางกันก็จะหา RefSys ที่อยู่ใกล้กับจุด Pp ใหม่
- ทำซ้ำจนกว่าจะเติมจุดในช่วงที่ว่างครบโดยการอัปเดต Pp ใหม่เรื่อยๆ และทำการลบ RefSys ที่ผ่านมาแล้วทิ้งเพื่อป้องกันการวนลูปซ้ำ
- หลังจากหาจุดที่อยู่ระหว่างช่องว่างครบแล้วก็จะนำจุดใน L ไปใส่ใน T'

โดยใน Calibration Method นี้ทางผู้วิจัยบทความนี้ได้กำหนด min_d = 50m, max_d = 500m และ time+d = 1200s. ซึ่งจะได้ $O(NofT \cdot NofC^2)$ โดยที่ NofC คือ จะนวนจุดของ RefSys ที่อยู่ระหว่างช่องว่างนั้น และ NofT คือ ระยะระหว่างช่องว่างนั้น

ซึ่งเมื่อทำทั้ง 2 อัลกอริทึมนี้แล้วจะได้พิกัด Latitude, Longitude แต่ยังคงขาด Timestamp ซึ่งในการหา Timestamp จะใช้สมการในการหาดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

$$a^* \cdot t = p_p \cdot t + \frac{(p_n \cdot t - p_p \cdot t) \cdot d(p_p, a^*) \cdot \left| \frac{\vec{p_p a^*}}{p_p p_n} \right|}{d(p_p, p_n) \cdot \left| \frac{\vec{p_p a^*}}{p_p p_n} \right| \cdot \left| \frac{\vec{p_p p_n}}{p_p p_n} \right|} \quad (2.1)$$

แต่ละตัวแปรมีความหมายดังนี้

- a^* หมายถึง Timestamp ณ ตำแหน่งนั้น (ตำแหน่งที่ต้องการหา Timestamp)
- p_p หมายถึง Timestamp ณ ตำแหน่ง p_p
- p_n หมายถึง Timestamp ณ ตำแหน่ง p_n
- a^* หมายถึง จุดของ RefSys ที่อยู่ใกล้กับ p_p มากที่สุด

โดยงานวิจัยนี้สามารถเติมเต็มพิกัดที่หายไปได้ใกล้เคียงกับข้อมูลจริง แต่ก็ยังมีปัญหาที่เกิดขึ้น คือ บางตำแหน่งพิกัดของจุดที่เดิมมีตำแหน่งที่ไม่ตรงกับถนนเมื่อเทียบจากแผนที่ เนื่องจากพิกัดที่นำมาเติมได้มาจากจุด Centroids ของแต่ละ Clusters ในกรณีที่มีข้อมูลของรถอยู่นอกถนนเป็นจำนวนมากหรือบริเวณนั้น เป็นจุดพักรถหรือปั้มน้ำมันก็จะทำให้ตำแหน่งของจุด Centroids ออกนอกถนนไปด้วย

2.3.2 Completing Truck's Missing Trajectory Based on The Historical Information

งานวิจัยนี้ถูกจัดทำขึ้น โดย Guicheng Shen, Cailin Zhang, Baicheng Tang โดยจุดประสงค์ในการเริ่มงานวิจัยนี้คือ ในการวิเคราะห์ข้อมูลพิกัด GPS นั้นจำเป็นต้องการความสมบูรณ์ของข้อมูลพิกัด GPS ทำให้ปัญหาการหายไปของข้อมูลพิกัดเป็นปัญหาที่สำคัญ งานวิจัยนี้จึงเสนอวิธีการในการหาเส้นทางที่รถบรรทุกจากข้อมูลพิกัดรถบรรทุกในอดีต โดยจะแบ่งวิธีเป็น 2 ส่วน ส่วนที่ 1 คือการหาเส้นทางทุกเส้นทางที่เป็นไปได้จากช่วงพิกัด GPS ที่หายไปและ ส่วนที่ 2 คือใช้อัลกอริทึม K-Modes ในการเส้นทางที่รถส่วนใหญ่ใช้ เพื่อนำไปแทนที่บนช่วงพิกัด GPS ที่หายไป

ข้อมูลที่ใช้เป็นข้อมูลการเดินทางของรถบรรทุกใน Jiangxi, China เป็นช่วงตั้งแต่ 17 กันยายน 2555 ถึง 9 ตุลาคม 2562 โดยมีข้อมูลทั้งหมดมากกว่า 960,000 ข้อมูล ซึ่งงานวิจัยนี้ได้กำหนดว่า ถ้าช่วงเวลาของ พิกัด GPS 2 จุดห่างเกิน 30 วินาทีจะหมายความว่า เป็นช่วงพิกัด GPS ที่หายไป

สูตรหาระยะห่างของพิกัด GPS 2 จุดที่ใช้ในงานวิจัยนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

$$a = \cos(x_2 - x_1) \quad (2.2)$$

$$b = \cos(x_1) * \cos(x_2) \quad (2.3)$$

$$c = \cos(y_2 - y_1) \quad (2.4)$$

$$S = \sqrt{2} * R * \arcsin(\sqrt{1 - a + b - b * c}) \quad (2.5)$$

โดยความหมายแต่ละตัวแปรมีดังนี้

x_1 = ละติจูดของจุดก่อนช่วงที่สัญญาณจะหาย, x_2 = ละติจูดของจุดหลังช่วงที่สัญญาณหาย

y_1 = ลองจิจูดของจุดก่อนช่วงที่สัญญาณจะหาย, y_2 = ลองจิจูดของจุดหลังช่วงที่สัญญาณหาย

R = รัศมีของโลก ประมาณ 637,500 เมตร

ส่วนที่ในการหาเส้นทางที่เป็นไปได้ทั้งหมดจาก 2 จุดพิกัดนั้น โดยกำหนดให้ M คือจุดก่อนช่วงพิกัด GPS ที่จะหายไป และ N คือจุดหลังช่วงพิกัด GPS ที่จะหายไป และใช้ระยะ 50 เมตรรอบจุด M, N เพราะเป็นตัวเลขที่อ้างอิงถึงสภาพถนนที่ใช้ โดยวิธีการจะมีดังนี้

- สร้างวงกลมรัศมี 50 เมตร โดยใช้จุด M และหาจุดที่อยู่ในรัศมีวงกลม 50 เมตรที่สร้างไว้
- ขยับไปจุดที่เจอว่าอยู่ในวงกลมทุกจุดโดยติดบอกไว้ว่ามาจากจุด M และสร้างวงกลมรัศมี 50 เมตร เพื่อ หาจุดภายในวงและขยับไปเรื่อย ๆ จนไม่เจอจุดอื่นในวงกลมที่สร้าง
- สร้างวงกลมรัศมี 50 เมตร โดยใช้จุด N และหาจุดที่อยู่ในรัศมีวงกลม 50 เมตรที่สร้างไว้
- ขยับไปจุดที่เจอว่าอยู่ในวงกลมทุกจุดโดยติดบอกไว้ว่ามาจากจุด N และสร้างวงกลมรัศมี 50 เมตรเพื่อ หาจุดภายในวงและขยับไปเรื่อย ๆ จนไม่เจอจุดอื่นในวงกลมที่สร้าง
- หาเส้นทางที่เชื่อมระหว่าง M และ N โดยหาจุดที่มีจุดข้างๆมาจาก M และ N จากที่ติดไว้

โดยผลลัพธ์ที่ได้คือเส้นทางทั้งหมดระหว่างจุด M และ N ต่อไปจะใช้สูตรหาระยะห่างที่กล่าวไว้ข้างต้นหาระยะห่างของทุกเส้นทางและหาระยะเวลาที่ใช้เพื่อนำไปวาดกราฟสำหรับใช้ K-Modes

ส่วนการหาเส้นทางที่คนขับรถบรรทุกส่วนใหญ่ใช้ จะใช้อัลกอริทึม K-Modes ในการจำแนกกลุ่มโดยแบ่งตามค่าระยะห่างกับระยะเวลาที่คำนวณไว้ของแต่ละเส้นทาง หลังจากที่ใช้ K-Modes

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

จะได้กลุ่มที่มีจำนวนเส้นทางเยอะที่สุดเป็นเส้นทางที่เป็นไปได้มากที่สุดในการนำไปแทนที่ในช่วงข้อมูลที่หายไประหว่างจุด M และจุด N

ปัญหาของงานวิจัยนี้ คือ ตัวงานวิจัยไม่มีการกล่าวถึงว่าเส้นทางจริงๆที่นำไปแทนที่ช่วงที่หายไปนั้น จะใช้จากการเลือกเส้นทางเส้นในกลุ่มเส้นทางหรือว่าหาค่ากลาง (Centroid) ของทั้งกลุ่มมาใช้ ทำให้ไม่ทราบว่าควรจะใช้ค่าอะไร จึงไม่สามารถนำมาวัดผลเพื่อเปรียบเทียบผลลัพธ์ได้

2.4 ทฤษฎีที่เกี่ยวข้อง

ในหัวข้อนี้จะเป็นรายละเอียดของทฤษฎีที่เกี่ยวข้องและสูตรทางคณิตศาสตร์ที่มีการนำมาใช้ โดยมีรายละเอียดตามหัวข้อย่อยต่อไปนี้

2.4.1 Haversine Formula

สูตรทางคณิตศาสตร์ Haversine เป็นสูตรหาระยะห่างระหว่าง 2 จุดบนทรงกลม ซึ่งเมื่อวัดกับโลกจะใช้ค่าละติจูดและลองจิจูดในการคำนวณ โดยมีใช้สูตรดังนี้

$$d = 2r \arcsin \left(\sqrt{h(\phi_2 - \phi_1) + \cos(\phi_2) h(\psi_2 - \psi_1)} \right) \quad (2.6)$$

$$d = 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\psi_2 - \psi_1}{2} \right)} \right) \quad (2.7)$$

เมื่อ $h(\theta)$ คือ haversine function : $h(\theta) = \sin^2 \left(\frac{\theta}{2} \right) = \frac{1 - \cos(\theta)}{2}$

d คือ ระยะห่าง

r คือ รัศมีของโลก มีค่าประมาณ 6371.0 กิโลเมตร

ϕ_1, ϕ_2 คือ latitude ของจุด 1, 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

$\psi_1 - \psi_2$ คือ longitude ของจุด 1, 2

2.4.2 Bearing Angle

ระบบ Bearings จะบอกทิศของเส้นที่ลากตามความโค้งของโลกจากจุดเริ่มต้นไปยังจุดปลายทาง ด้วยค่ามุมที่วัดเทียบกับเส้นเมริเดียนที่ผ่านจุดเริ่มต้นนั้นไปทางทิศตะวันออกหรือตะวันตก เมื่อจุดปลายทางอยู่ทางทิศเหนือ ค่ามุมจะอ้างอิงกับทิศเหนือ ส่วนเมื่อจุดปลายทางอยู่ทางทิศใต้ ค่ามุมจะอ้างอิงกับทิศใต้ โดยบอกทิศอ้างอิงว่าเป็นเหนือหรือใต้ ตามด้วยค่ามุมและทิศของจุดปลายทางนั้นว่าหันไปด้านตะวันออกหรือตะวันตก

$$\theta = \text{atan2}(\sin\Delta\lambda \cdot \cos\varphi_2, \cos\varphi_1 \cdot \sin\varphi_2 - \sin\varphi_1 \cdot \cos\varphi_2 \cdot \cos\Delta\lambda) \quad (2.8)$$

เมื่อ φ_1, λ_1 หมายถึง จุดเริ่มต้น, φ_2, λ_2 หมายถึง จุดสุดท้าย, $\Delta\lambda$ หมายถึง longitude ที่ต่างกัน

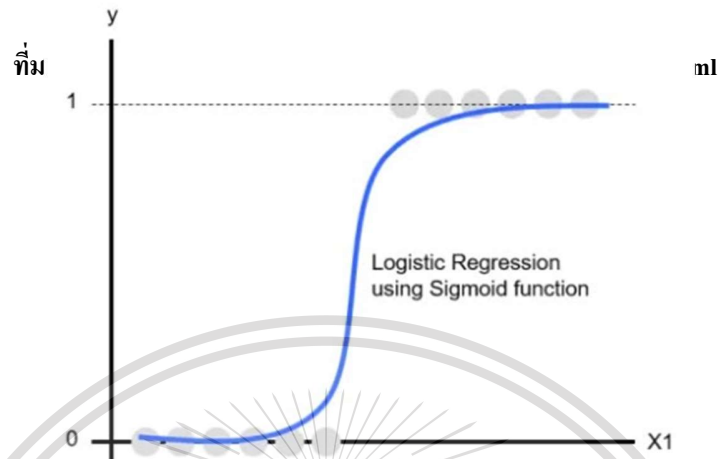
2.4.3 Logistic Regression

เป็น 1 ในโมเดลของการเรียนรู้ของเครื่อง (Machine Learning) ที่ใช้ในการจำแนกข้อมูลหรือ Classification โดยใช้ Sigmoid Function มาช่วยในการทำนาย ซึ่งจะสามารถทำนายได้เพียง 2 กลุ่ม แต่ในกรณีที่น่ามาใช้ในการทำนายกับข้อมูลที่มีหลายกลุ่มจะใช้เทคนิค Classifier Combination ในการคาดการณ์ผลที่ได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 2.8 ตัวอย่างของ Logistic Regression

2.4.4 Naive Bayes

เป็น 1 ในโมเดลของการเรียนรู้ของเครื่อง (Machine Learning) ที่ใช้ในการจำแนกข้อมูลหรือ Classification โดยจะใช้ทฤษฎีในเรื่องความน่าจะเป็นช่วยในการทำนายผล

$$P(c|x) \text{ Posterior Probability} = \frac{P(x|c) \text{ Likelihood} P(c) \text{ Class Prior Probability}}{P(x) \text{ Predictor Prior Probability}} \quad (2.9)$$

โดยจะมีความหมายดังนี้

- $P(c|x)$ Posterior Probability คือ ความน่าจะเป็นที่ข้อมูลที่มีแอตทริบิวต์เป็น x จะมีคลาส C
- $P(x|c)$ Likelihood คือ ความน่าจะเป็นที่ข้อมูลที่มีคลาส C และมีแอตทริบิวต์ x
- $P(c)$ Class Prior Probability คือ จำนวน Class ที่อาจจะเกิดขึ้น/จำนวน Class ทั้งหมดหรือความน่าจะเป็นของ Class C
- $P(x)$ Predictor Prior Probability คือ จำนวน Attribute ทั้งหมด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.4.5 Decision Tree

เป็น 1 ในโมเดลของการเรียนรู้ของเครื่อง (Machine Learning) ที่ใช้ในการจำแนกข้อมูลหรือ Classification โดยใช้หลักการตัดสินใจแบบบนลงล่าง ซึ่งจะมีลักษณะคล้ายกับต้นไม้ โดยจะเริ่มพิจารณาจากรากหรือที่เรียก Root Node ที่อยู่ด้านบนก่อนแล้วจึงแตก Branch แยกไปยัง Node ลูกที่อยู่ถัดลงไปเรื่อย ๆ ในการสร้าง Decision Tree นั้นจะใช้หลักการ Information Gain หรือ Gini มาช่วยในการเลือกลำดับการตัดสินใจของ Attribute ซึ่งจะใช้ If-Then-Else ในการตัดสินใจ

2.4.6 Deep Learning

เป็น 1 ในโมเดลของการเรียนรู้ของเครื่อง (Machine Learning) ที่เลียนแบบการทำงานจากระบบเซลล์ประสาทในสมองของมนุษย์ (Neural Network) โดย Deep Learning จะมีส่วนประกอบอยู่ 3 ชั้นหลัก ๆ คือ Input Layer, Hidden Layer, Output Layer ซึ่งภายใน Hidden Layer สามารถประกอบไปด้วย Neural Network ตั้งแต่ 0 ชั้นขึ้นไปได้ และภายใน Neural Network เหล่านั้นจะมี Kernel และ Activation Function ที่จะช่วยในการทำนายผลและส่งค่าไปชั้นถัดไป

2.4.7 K-Means Classification

เป็น 1 ในโมเดลของการเรียนรู้ของเครื่อง (Machine Learning) ที่ใช้ในการแบ่งกลุ่มข้อมูลหรือ Clustering ซึ่งนับเป็น Unsupervised Learning โดยใช้หลักการที่ว่าจุดที่ต้องการทำนายอยู่ใกล้กับข้อมูลกลุ่มใดมากที่สุด จะถือว่าเป็น 1 ในสมาชิกกลุ่มนั้น ซึ่งในการหาว่าจุดที่จะทำนายอยู่ใกล้กับกลุ่มไหนจะทำโดยการสร้างตัวแทนหรือจุดศูนย์กลางของกลุ่มนั้นและหาระยะห่างระหว่างจุดข้อมูลที่ต้องการหาว่าเป็นสมาชิกของกลุ่มใดกับตัวแทนของกลุ่ม โดยจำเป็นต้องกำหนดจำนวนกลุ่มก่อนที่จะเริ่มทำนาย

2.4.8 Mean Absolute Error

$$Error = Actual - Predicted \quad (2.10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Error| \quad (2.11)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

โดยนิยามเรียกเป็นชื่อย่อว่า MAE หรือเรียกอีกชื่อหนึ่งว่า L1 Loss ค่า MAE ใช้ Absolute มาช่วยทำให้ค่า Error กลายเป็นบวก สูตรคำนวณจึงเป็นการนำค่า Error มาใส่ Absolute ก่อนที่จะนำมาหาค่าเฉลี่ยของ Error และค่านี้จะมีความอ่อนไหวกับ Outlier น้อยกว่า MSE และ RMSE เนื่องจากการนำค่า Error มาใส่ Absolute เฉย ๆ ซึ่งหากพบว่าข้อมูลมี Outlier อยู่เยอะ ค่านี้ก็อาจเหมาะสมกว่าในการนำไปใช้ประเมินประสิทธิภาพของโมเดล

2.5 เครื่องมือที่ใช้

2.5.1 Scikit-learn

Scikit-learn เป็นไลบรารีของการเรียนรู้ของเครื่องที่เป็นโอเพ่นซอร์ส โดยสามารถใช้ได้กับการเรียนรู้แบบมีผู้สอนและการเรียนรู้แบบไม่มีผู้สอน Scikit-learn มีเครื่องมือที่หลากหลายสำหรับการวัดค่าว่าโครงสร้างมีการเหมาะสมกับข้อมูล การจัดการข้อมูลก่อนนำไปใช้ การเลือกโครงสร้าง การวัดความถูกต้องของผลลัพธ์ และอีกหลาย ๆ ด้านนอกเหนือจากนี้

2.5.2 Shapely

Shapely เป็น Python Package ของ BSD-license (Berkeley Software Distribution License) ที่ใช้จัดการและวิเคราะห์ระนาบของวัตถุรูปทรงเรขาคณิต โดยมีพื้นฐานมาจากไลบรารี GEOS (Geometry Engine - Open Source) และ JTS (JTS Topology Suite) ที่มีการใช้กันแพร่หลาย นอกจากนี้ Shapely ไม่ได้ยึดติดกับรูปแบบของข้อมูลหรือระบบพิกัดของข้อมูล แต่สามารถใช้ร่วมกับ Package ที่ข้อมูลนั้นอยู่ได้เลย

2.5.3 PostgreSQL

เป็นระบบจัดการฐานข้อมูลแบบ Open Source พัฒนาโดย PostgreSQL Global Development Group โดยอยู่บนพื้นฐานของภาษา SQL และมีคุณสมบัติอื่น ๆ เพิ่มเติมอีกมากมาย โดยมีลักษณะเด่นในการจัดเก็บข้อมูลที่มีขนาดใหญ่และซับซ้อน รวมถึงสามารถปรับขนาดได้อย่างมีประสิทธิภาพ และยังมีผู้ใช้งานเป็นจำนวนมาก และสามารถใช้งานได้ด้วยภาษาทางคอมพิวเตอร์ได้หลากหลายภาษา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2.5.4 PostGIS

เป็นส่วนขยายมาจาก PostgreSQL โดยจะการเพิ่มการทำงานให้สามารถใช้งานกับข้อมูลทางภูมิศาสตร์ได้ รวมถึงการสืบค้นข้อมูลของพิกัดทางภูมิศาสตร์ที่อยู่ใน SQL

2.5.5 pgAdmin

เป็นเครื่องมือในการใช้งาน PostgreSQL ที่เป็นที่นิยมใช้งานมากที่สุดเนื่องจากมีฟังก์ชันให้เลือกใช้งานได้อย่างหลากหลาย และยังเป็น Open Source และยังสามารถใช้งานได้กับทุกระบบปฏิบัติการพื้นฐาน ไม่ว่าจะเป็น Linux, Unix, MacOS หรือ Windows

2.5.6 Microsoft Power BI

เป็นโปรแกรมที่ใช้ช่วยวิเคราะห์ (Analytic) และสรุปผลออกมาเป็นรูปแบบต่าง ๆ (Visualization) เช่น กราฟแท่ง กราฟพาย แผนภูมิ เป็นต้น โดยผู้ใช้ไม่จำเป็นต้องมีความรู้ด้านเทคนิคมากนัก ซึ่งช่วยในการตัดสินใจและกำหนดกลยุทธ์



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 3

การออกแบบและพัฒนาระบบ

ในบทนี้จะกล่าวถึง 4 หัวข้อหลักที่ทำการพัฒนา คือ 3.1) โมเดลหาจังหวัดจากตำแหน่ง GPS 3.2) โมเดลการหาจุดอ้างอิงบนถนนทุก ๆ 100 เมตร 3.3) อัลกอริทึมและโมเดลเติมเต็มตำแหน่ง GPS และ 3.4) ฟังก์ชันเติมเต็มตำแหน่ง GPS

ส่วนของการหาจังหวัดจากตำแหน่ง GPS จะนำข้อมูล Polygon ของจังหวัดและข้อมูลตำแหน่ง GPS มาใช้ในการหาจังหวัด โดยใช้ Machine Learning Model, ส่วนของข้อมูลถนน 100 เมตร จะนำเอา Shapefile ของถนนและข้อมูลหลักกิโล มาใช้ในการหาจุดอ้างอิงบนถนนทุก ๆ 100 เมตร, ส่วนของการเติมเต็มข้อมูล GPS จะประกอบไปด้วยโมเดลที่ใช้ในการเติมเต็มตำแหน่ง GPS อยู่ 4 โมเดลและการวัดผลเปรียบเทียบแต่ละโมเดล โดยจะวัดจากตำแหน่งและความเร็วที่ทำนาย และในที่สุดท้าย ส่วนของฟังก์ชันจะเป็นการนำเอาโมเดลที่แม่นยำมากที่สุดมาใช้ในการสร้างฟังก์ชันเพื่อให้ผู้ใช้งานสามารถนำโมเดลไปใช้งานและปรับปรุงต่อได้

3.1 โมเดลการหาจังหวัดจากตำแหน่งของข้อมูล GPS

ในการหาตำแหน่งจากข้อมูล GPS ว่าอยู่ในจังหวัดไหน จะช่วยในการจำกัดขอบเขตของข้อมูลที่ต้องใช้โดยคัดเฉพาะพื้นที่จังหวัดที่เกี่ยวข้อง ทำให้ลดเวลาที่ต้องใช้ประมวลผลจากข้อมูลทั่วประเทศไทยเหลือแค่พื้นที่จังหวัดของจุดที่สนใจ ซึ่งมีวิธีการในการทำอยู่สองส่วน ส่วนที่ 1 จะทำการกำหนดพิกัดของจังหวัดหรือการสร้างข้อมูลของจังหวัด และส่วนที่ 2 จะนำพิกัดของจังหวัดไปทำนายพิกัดของ GPS ว่าอยู่ในจังหวัดไหน

3.1.1 การทำงานส่วนที่ 1 สร้างข้อมูล GPS ของจังหวัดในไทย

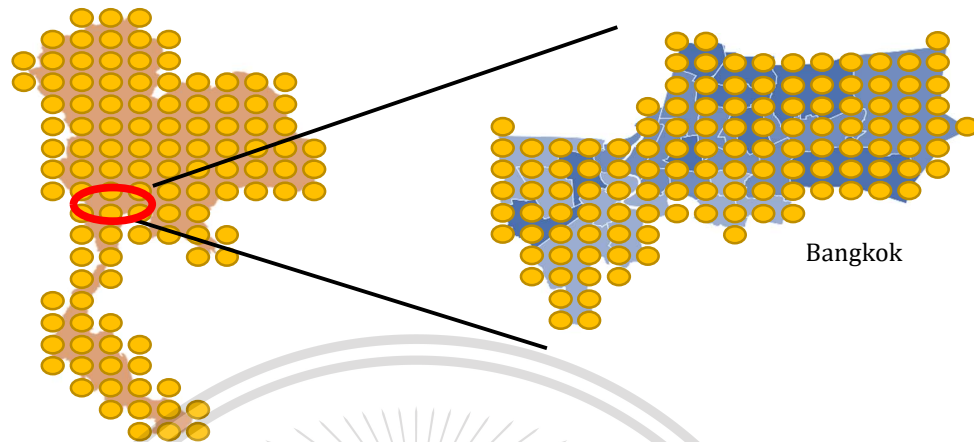
ทำการเชื่อมต่อกับฐานข้อมูลและอ่าน shape file ของจังหวัดที่เป็นไฟล์ Polygon ที่อยู่ใน PostgreSQL ลงในโปรแกรมภาษา Python จากนั้นทำการสร้าง Grid ที่อยู่ในประเทศไทย โดยให้พิกัดละติจูด ลองจิจูดมีระยะห่าง 1 กิโลเมตรจากข้อมูลพิกัดอื่นรอบข้าง และนำข้อมูลของจังหวัดและอำเภอที่ได้จาก polygon มาใส่ลงไปในข้อมูล

เอกสารนี้เป็นเอกสารที่จัดทำขึ้นเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูปที่ 3.1 ตัวอย่างของการตี Grid

3.1.2 การทำงานส่วนที่ 2 การฝึกสอนโมเดล

นำข้อมูลที่ได้มาทำการฝึกสอนโมเดล โดยในการทดลองจะใช้โมเดลประเภท Classification ในการทำนายเช่น Logistic Regression, Naive Bayes, Decision Tree และ Deep Learning และเลือกนำเอาโมเดลที่ดีที่สุดมาใช้งาน โดยให้ Input เป็น ละติจูด ลองจิจูด และ Output เป็นจังหวัด

3.1.3 การวัดผลเปรียบเทียบแต่ละโมเดล

ใช้ความแม่นยำ (Correctly Classified), เวลาที่ใช้ในการ Train ทั้ง Training Set, ปริมาณ Memory ของ Object ของ Model และเวลาที่ใช้ในการ Test เฉลี่ยต่อ Location

3.2 โมเดลการหาจุดอ้างอิงบนถนนในระยะ 100 เมตร

เป็นการสร้างข้อมูลเก็บจุดอ้างอิงถนนทุก 100 เมตร เพื่อลดการประมวลผลจากข้อมูลดิบที่เก็บตามตำแหน่งถนนและเพื่ออ้างอิงระยะห่างของแต่ละจุดได้เท่า ๆ กัน โดยสามารถสรุปขั้นตอนการทำงานดังนี้

- แปลงข้อมูลถนนจาก Shapefile ให้เป็น CSV และกำหนดหมายเลขประจำตัวของแต่ละส่วนของถนนหรือ Road ID (rid)

- หาระยะห่างระหว่างจุด และบวกสะสมให้ถึง 100 เมตร

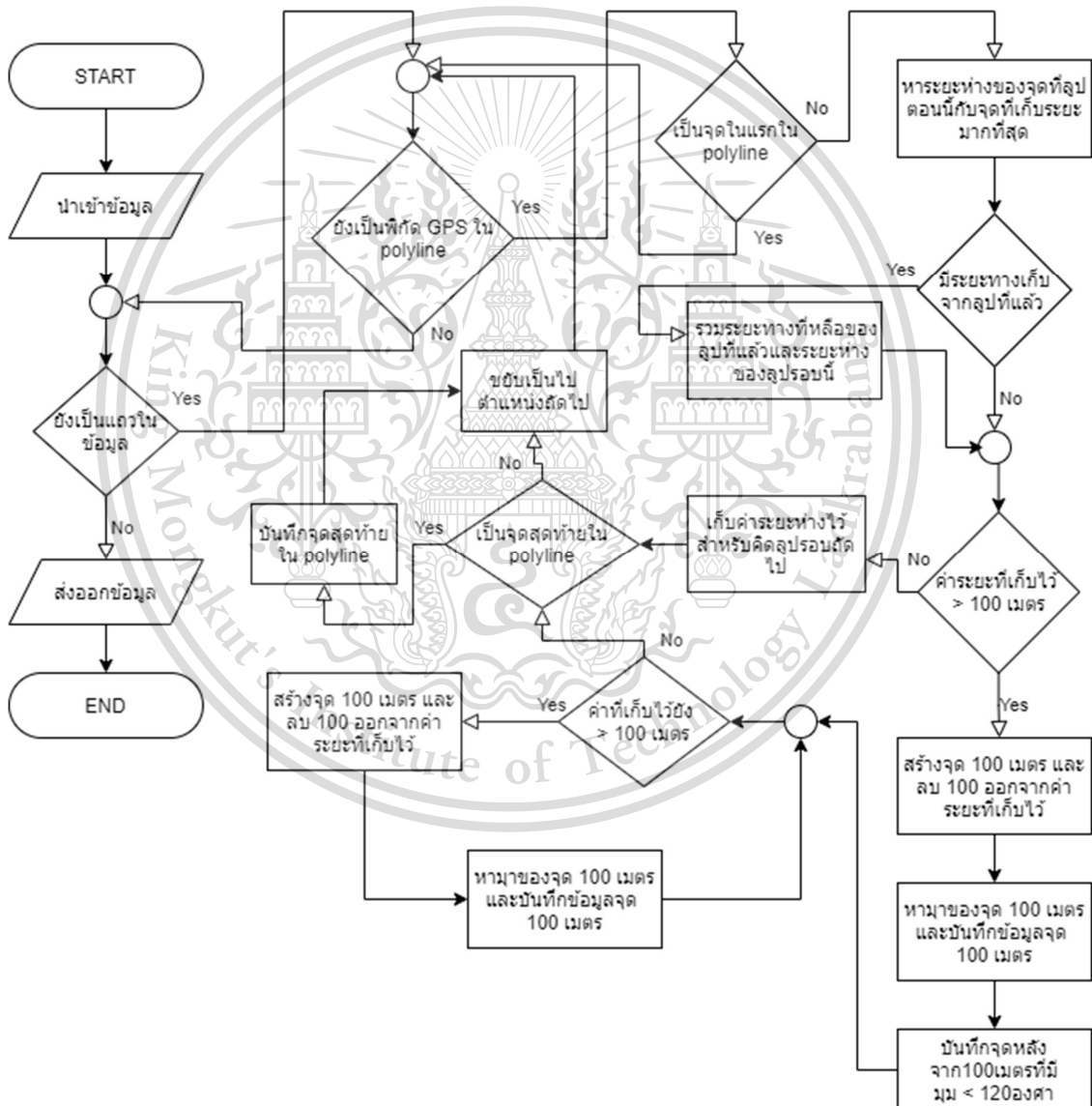
เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์และใช้เฉพาะเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- เก็บค่าทุก ๆ 100 เมตร โดยการสร้างจุดขึ้นใหม่ที่ระยะ 100 เมตร และอ้างอิงทิศทางจากค่ามุม Bearing ตามสมการที่ 2.4.2
- เก็บค่าจุดโค้งของถนน เมื่อวัดมุมเทียบกับจุดก่อนหน้าและจุดถัดไปได้้น้อยกว่า 120 องศา
- เมื่อสิ้นสุดข้อมูลถนนที่มี จะเก็บจุดสุดท้ายไว้และเปลี่ยนไปทำส่วนของถนน (rid) ถัดไป

แผนภาพ Flowchart ด้านล่าง คือ รายละเอียดการทำงานเพิ่มเติมของส่วนนี้



รูป 3.2 ขั้นตอนการสร้างข้อมูลถนนที่แต่ละจุดห่างกัน 100 เมตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

โดยข้อมูลที่ได้หลังจากแปลงเป็น .csv แล้วมีรายละเอียดดังต่อไปนี้

- คอลัมน์ที่ 1 rid หมายถึง หมายเลขประจำตัวของแต่ละส่วนถนนหรือ Road ID ซึ่งถนน 1 หมายเลขสามารถมีได้หลาย Road ID
- คอลัมน์ที่ 2 rd หมายถึง หมายเลขถนน หรือ Road Code
- คอลัมน์ที่ 3 km หมายถึง ตำแหน่งของจุดอ้างอิงถนนที่เก็บค่าทุก ๆ 100 เมตร โดยจุดเริ่มถนนเป็น 0
- คอลัมน์ที่ 4 type หมายถึง ประเภทของ km โดยจะแบ่งออกเป็น 3 ประเภทดังนี้
 - ประเภทที่ 1 คือ 100m หมายถึง ข้อมูลจุดอ้างอิงถนนเมื่อเก็บระยะได้ครบทุก ๆ 100 เมตร
 - ประเภทที่ 2 คือ 1km หมายถึง ข้อมูลจุดอ้างอิงถนนเมื่อเก็บระยะได้ครบทุก ๆ 1 กิโลเมตร
 - ประเภทที่ 3 คือ bend หมายถึง ข้อมูลถนน ณ พิกัดตำแหน่งนั้นทำมุมกับจุดก่อนหน้า และจุดถัดไปน้อยกว่า 120 องศา
- คอลัมน์ที่ 5 lat หมายถึง พิกัด Latitude ซึ่งอยู่ในรูปแบบของ องศาศนิยม หรือ Decimal Degree
- คอลัมน์ที่ 6 lon หมายถึง พิกัด Longitude ซึ่งอยู่ในรูปแบบของ องศาศนิยม หรือ Decimal Degree
- คอลัมน์ที่ 7 latx หมายถึง พิกัด Latitude ซึ่งอยู่ในรูปแบบของ องศาศนิยม หรือ Decimal Degree ที่มีศนิยม 3 ตำแหน่ง
- คอลัมน์ที่ 8 lonx หมายถึง พิกัด Longitude ซึ่งอยู่ในรูปแบบของ องศาศนิยม หรือ Decimal Degree ที่มีศนิยม 3 ตำแหน่ง
- คอลัมน์ที่ 9 angle หมายถึง มุมของพิกัดที่เทียบระหว่างพิกัดก่อนหน้ากับพิกัดถัดไป ในกรณีที่พิกัดนั้นเป็นพิกัดแรกหรือพิกัดสุดท้ายของ rid จะมีมุม 180 องศา

3.3 โมเดลการเติมเต็มตำแหน่ง GPS

เป็นโมเดลที่ใช้ในการเติมเต็มช่วงข้อมูล GPS ของรถยนต์ที่หายไปในช่วงเวลา 10, 20, 30 นาที

3.3.1 ข้อมูลที่ต้องการในการนำเข้าสู่ระบบและข้อมูลส่งออกจากระบบ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ข้อมูลที่จะนำเข้าสู่ระบบจะประกอบไปด้วยข้อมูลที่มีช่องว่างของพิกัดขาดหายไปที่ต้องการนำมาเติมจุด ข้อมูลที่ส่งออกจะเป็นข้อมูลที่ผ่านการเติมพิกัดที่ขาดหายไปเสร็จแล้ว

3.3.1.1 ข้อมูลที่นำเข้าสู่ระบบ (Input)

ข้อมูลที่นำเข้าสู่ระบบ จะเป็นข้อมูลพิกัด GPS ของรถที่มีการขาดหายไปของข้อมูล โดยโครงการนี้จะเติมข้อมูลเมื่อมีพิกัด GPS ขาดหายไปเกิน 10, 20, 30 นาที ซึ่งข้อมูลที่นำเข้าจะประกอบไปด้วยข้อมูลพื้นฐานต่างๆของ GPS ได้แก่ เวลาที่ GPS ของรถส่งสัญญาณมา ข้อความประจำตัวของรถ ละติจูดของรถที่เวลานั้น ลองจิจูดของรถที่เวลานั้น ความเร็วของรถที่เวลานั้น และประเภทของรถ

time_stamp	vid	lat	lon	speed	heading	hire
28/12/2019 0:54	rtJHM	14.15068	101.3148	98	89	1
28/12/2019 0:55	rtJHM	14.15024	101.32846	95	101	1
28/12/2019 0:57	rtJHM	14.14302	101.35335	92	128	1
28/12/2019 1:08	rtJHM	14.13054	101.47435	80	99	1
28/12/2019 1:09	rtJHM	14.12756	101.48366	94	108	1

รูป 3.3 ตัวอย่างข้อมูลที่มีช่วงเวลาที่ห่างกันมากกว่าหรือเท่ากับ 30 นาที

3.3.1.2 ข้อมูลส่งออกจากระบบ (Output)

ข้อมูลส่งออกจากระบบ จะเป็นข้อมูลพิกัด GPS ที่เติมลงในช่วงที่หายไปแล้ว ซึ่งจะส่งออกข้อมูลเวลาที่ GPS ของรถที่ทำนายขึ้น, ข้อความประจำตัวของรถ, ละติจูดของรถที่เวลานั้น, ลองจิจูดของรถที่เวลานั้น, ความเร็วของรถที่เวลานั้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

	time_stamp	vid	lat	lon	speed	
0	28/12/2019 0:57	[REDACTED]	rtJHM	14.14138058	101.3554796	56
1	28/12/2019 0:58	[REDACTED]	rtJHM	14.13718869	101.361588	56
2	28/12/2019 0:58	[REDACTED]	rtJHM	14.13526135	101.3681884	56
3	28/12/2019 0:59	[REDACTED]	rtJHM	14.13377518	101.3733493	56
4	28/12/2019 0:59	[REDACTED]	rtJHM	14.1355267	101.3788803	56

รูป 3.4 ตัวอย่างข้อมูล GPS ที่เต็มช่วงที่หายแล้ว

3.3.2 รายละเอียดข้อมูลที่ใช้พัฒนาโมเดล

ข้อมูลที่น่ามาสร้างและพัฒนาระบบจะประกอบด้วย 2 ส่วน คือ ข้อมูลพิกัดของรถทั้งหมดที่มีและข้อมูลของถนน

3.3.2.1 ข้อมูลพิกัดของรถ

ข้อมูลพิกัดของรถที่ได้เก็บไว้จะเป็นข้อมูลพิกัด GPS ของรถทั้งหมดที่ผ่านช่วงถนนที่กำหนดไว้ ซึ่งจะประกอบด้วยข้อมูลต่าง ๆ ได้แก่ เวลาที่ GPS ของรถส่งสัญญาณมา ข้อความประจำตัวของรถ ละติจูดของรถที่เวลานั้น ลองจิจูดของรถที่เวลานั้น ความเร็วของรถที่เวลานั้น และประเภทของรถ โดยข้อมูลพิกัดของรถที่จะนำมาสร้างและพัฒนาระบบ นั้นจะถูกแบ่งออกเป็นสองส่วน โดยส่วนที่ 1 จะถูกนำมาใช้ในการฝึกสอนระบบ และส่วนที่ 2 จะนำมาใช้ในการทดสอบระบบเพื่อวัดผลและพัฒนาระบบต่อไป โดยแต่ละส่วนมีรายละเอียดดังนี้

3.3.2.1.1 ส่วนที่ 1 การฝึกสอนโมเดล

เป็นส่วนที่ถูกใช้ในการฝึกสอนโมเดล จะใช้ข้อมูลทั้งหมด 3 ปี ได้แก่ ปี พุทธศักราช 2560, 2561, 2562

3.3.2.1.2 ส่วนที่ 2 การทดสอบโมเดล

เป็นส่วนที่จะนำมาใช้ในการทดสอบระบบ จะใช้ข้อมูลทั้งหมด 1 เดือนในปี พุทธศักราช 2563

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

3.3.2.2 ข้อมูลอ้างอิงถนนในระยะ 100 เมตร

โดยที่ข้อมูลถนนแต่เดิมถูกเก็บในรูปแบบของ Shapefile (.shp) โดยภายในจะประกอบด้วยข้อมูลต่าง ๆ ของถนนซึ่งในโครงการนี้เลือกนำข้อมูลมาใช้อยู่ 2 ข้อมูลนั่นก็คือ หมายเลขของถนนและ Geometry หรือชุดข้อมูลพิกัดทางภูมิศาสตร์ ต่อไปจะเปลี่ยนรูปแบบของข้อมูลให้เป็นเก็บที่ละจุดพิกัดภูมิศาสตร์ โดยถ้าเป็นส่วนถนนเดียวกัน แต่ละพิกัดจะห่างกันทุก 100 เมตร รายละเอียดข้อมูลต่าง ๆ จะอยู่ในบทที่ 2.1.2.3 โดยข้อมูลของช่วงถนนที่เลือกมาพัฒนาจะเป็นรูปแบบที่ทางผู้จัดทำสนใจ คือ ถนนทางตรง, ถนนทางโค้ง, ถนนทางบายพาส (Bypass) หรือถนนเลี้ยวเมือง

rid	rd	km	ptype	lat	lon	latx	lonx	angle
3344	354	0	100m	13.66597	100.7195	13.666	100.72	180
3344	354	0.1	100m	13.66592	100.7204	13.666	100.72	180
3344	354	0.2	100m	13.66587	100.7214	13.666	100.721	180
3344	354	0.3	100m	13.66582	100.7223	13.666	100.722	180
3344	354	0.4	100m	13.66577	100.7232	13.666	100.723	180
3344	354	0.5	100m	13.66573	100.7241	13.666	100.724	180
3344	354	0.6	100m	13.66568	100.7251	13.666	100.725	180
3344	354	0.655	bend	13.66565	100.7256	13.666	100.726	87
3344	354	0.66	100m	13.66565	100.7256	13.666	100.726	180
3346	347	0	100m	14.01479	100.55	14.015	100.55	180
3346	347	0.1	100m	14.01559	100.5504	14.016	100.55	178
3346	347	0.2	100m	14.0164	100.5508	14.016	100.551	179
3346	347	0.3	100m	14.01721	100.5512	14.017	100.551	180
3346	347	0.4	100m	14.01801	100.5516	14.018	100.552	178
3346	347	0.5	100m	14.01883	100.552	14.019	100.552	179

รูป 3.5 ตัวอย่างข้อมูลอ้างอิงของถนน 100 เมตร

3.3.3 โมเดลและอัลกอริทึมที่ใช้พัฒนาโมเดล

โมเดลที่ใช้พัฒนามีอยู่ทั้งหมด 5 โมเดลหลักและ 1 อัลกอริทึม ซึ่งสามารถแบ่งได้ตามนี้ อัลกอริทึมในการสร้างตารางความเร็ว และโมเดลที่พัฒนาขึ้น 5 โมเดลแบ่งเป็นโมเดลที่ใช้เดิมพิกัด GPS ที่ขาดหายไป 4 โมเดล และโมเดลที่ใช้ในการเลือกเส้นทางสำหรับทางบายพาส 1 โมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

3.3.3.1 อัลกอริทึมการสร้างตารางความเร็วของรถที่แต่ละพิกัดของถนนในแต่ละช่วงเวลา

โมเดลนี้จะนำข้อมูลจุดอ้างอิงถนนที่เก็บทุก ๆ 100 เมตรและข้อมูลตำแหน่งพิกัดของรถทั้งหมดที่มี ใช้สร้างตารางความเร็วโดยจะสร้างตารางความเร็วแบ่งออกเป็นช่วงละ 15 นาที, 30 นาที และ 1 ชั่วโมง โดยในการสร้างตารางความเร็วจะแบ่งกลุ่มข้อมูลพิกัดของรถตามข้อมูลจุดอ้างอิงถนนว่าแต่ละจุดมีรถวิ่งผ่านกี่คัน ซึ่งจะอิงตามเวลาที่รถผ่านจุดนั้นด้วย หลังจากที่ได้จำนวนรถแล้วก็จะนำค่าเฉลี่ยของความเร็วของรถที่ผ่านพิกัดนั้นทั้งหมดมาเติมลงในตารางความเร็วตามช่วงเวลาที่กำหนด หลังจากที่ได้เติมข้อมูลทั้งหมดแล้วในกรณีที่มีช่วงของตารางความเร็วที่ว่างอยู่ จะใช้วิธีในการหาความเร็วอยู่ 2 วิธี

- วิธีที่ 1 ใช้การหาค่าเฉลี่ยของความเร็วที่อยู่ในพิกัดก่อนหน้าและถัดไป
- วิธีที่ 2 ใช้การหาความเร็วตามความชันที่ลดลงหรือเพิ่มขึ้นของพิกัดที่อยู่ก่อนหน้าและถัดไป

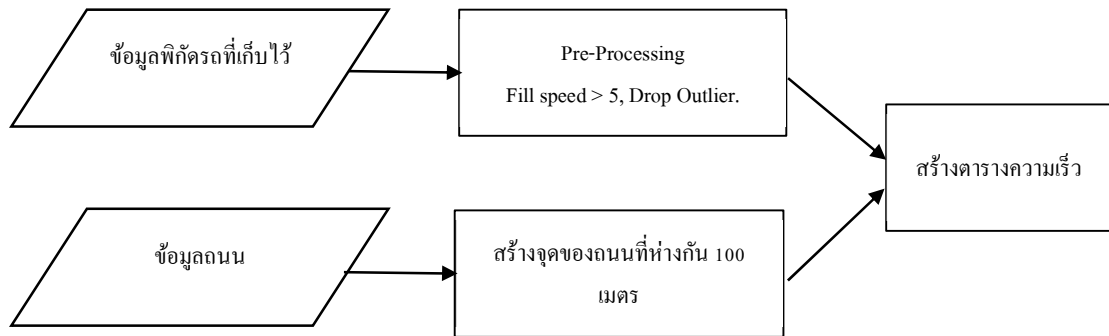
โดยมีรายละเอียดขั้นตอนการทำงานดังนี้

- ข้อมูลนำเข้าแบ่งออกเป็น 2 ส่วน ส่วนที่ 1 คือข้อมูลพิกัดของรถทั้งหมดที่เก็บไว้ ส่วนที่ 2 คือ ข้อมูลถนน
- ส่วนที่ 1 ข้อมูลพิกัดของรถทั้งหมดที่เก็บไว้ จะนำมาเข้าสู่กระบวนการ Pre-Processing ก่อนโดยการเลือกเฉพาะพิกัดที่มีความเร็วมากกว่า 2 กิโลเมตรต่อชั่วโมง และทำการลบข้อมูล Outlier
- ส่วนที่ 2 ข้อมูลถนน จะนำมาสร้างจุดของถนนที่ห่างกัน 100 เมตรตามหัวข้อที่ 3.2
- หลังจากนั้นจะนำเอาข้อมูลทั้ง 2 ส่วนมาสร้างตารางความเร็วตามวิธีที่ได้อธิบายไปในข้างต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 3.6 ขั้นตอนการสร้างตารางความเร็ว

3.3.3.2 โมเดลที่ 1 เติมข้อมูล GPS ด้วยข้อมูลอ้างอิงถนน 100 เมตร

เป็นการเติมข้อมูล GPS ที่หายไป โดยใช้ข้อมูลจุดอ้างอิงถนน 100 เมตรมาช่วยในการเติม ซึ่งจะใช้จุดพิกัดที่ได้จากตารางความเร็วเติมแทนที่ในช่วงข้อมูลที่หายไปของข้อมูลนำเข้า ในขั้นตอนที่ 1 จะทำการหาจุดเริ่มต้นและจุดสุดท้ายของช่วงเวลาที่ขาดหายไป และเมื่อเจอช่วง 30 นาทีที่มีพิกัดหายไป จะนำข้อมูลของตารางความเร็วที่มีพิกัดและเวลาที่ตรงหรือใกล้เคียงกับแต่ละจุดในช่วงนั้นมาเติม โดยจะทำการเติมจุดที่อยู่ใกล้กับจุดเริ่มต้นของข้อมูลที่หายาก่อน หลังจากนั้นก็จะทำการเติมจุดถัดไปจนครบ ในขั้นตอนนี้จะเติมข้อมูลเพียง ชุดข้อมูลประจำตัวของรถหรือ Vehicle ID ละติจูด ลองจิจูดและความเร็ว ซึ่งจะเห็นว่าข้อมูลที่ต้องการยังขาดข้อมูลที่สำคัญอีกส่วนคือเวลา ในการเพิ่มเวลานั้นจะทำการเติมจากการคำนวณด้วย สมการที่ 3.1

$$a^* \cdot t = p_p \cdot t + \frac{(p_n \cdot t - p_p \cdot t) \cdot d(p_p, a^*) \cdot \left| \frac{\vec{p_p a^*}}{p_p p_n} \right|}{d(p_p, p_n) \cdot \left| \frac{\vec{p_p a^*}}{p_p a^*} \right| \cdot \left| \frac{\vec{p_p p_n}}{p_p p_n} \right|} \quad (3.1)$$

โดยจะมีขั้นตอนการทำงานเมื่อเจอข้อมูลที่หาย 30 นาทีดังนี้

- เริ่มต้นจากการนำเข้าข้อมูลที่ต้องการเติมพิกัดที่หายไป โดยจะหาระยะห่างและเวลาที่แต่ละคู่พิกัดที่อยู่ตำแหน่งติดกันเพื่อหาช่วงของพิกัดที่ขาดหาย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

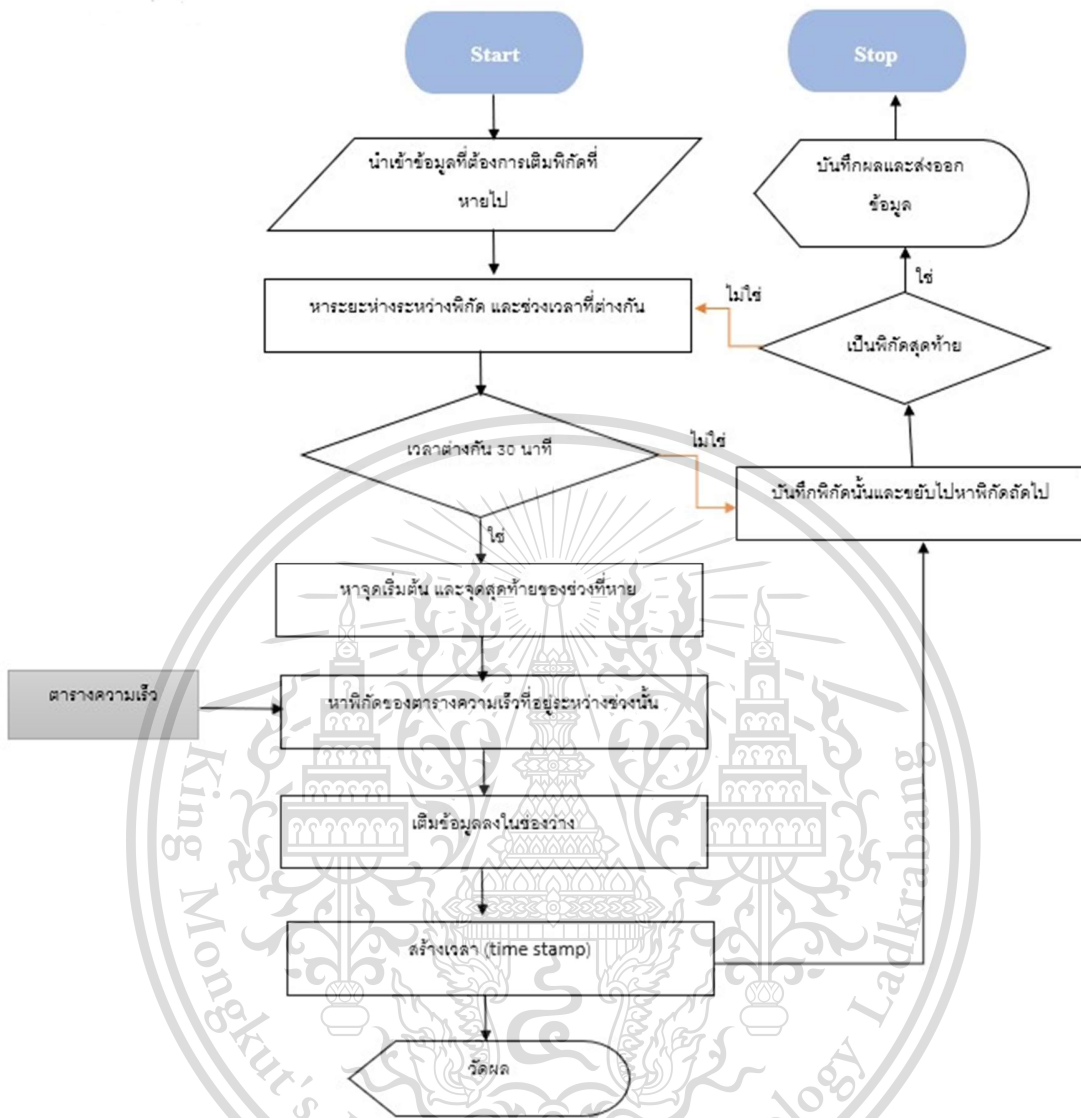
- ตรวจสอบว่ามีช่วงที่พิกัดขาดหายหรือไม่ โดยในการทดลองจะกำหนดระยะห่างที่น้อยที่สุดที่จะเติมคือ 3 กิโลเมตร และเวลาที่น้อยที่สุดที่จะถือว่าเป็นช่วงพิกัดขาดหายคือ 1800 วินาทีหรือ 30 นาที
- ถ้าพบว่ามีช่วงพิกัดที่ขาดหายไป จะทำการหาจุดที่เริ่มหายและจุดที่สิ้นสุดการหาย และทำการหารระยะห่างระหว่างจุดทั้งสองเก็บไว้ในตัวแปร d และทำการหาจุดศูนย์กลางระหว่างทั้ง 2 พิกัดนั้นด้วย
- นำข้อมูลของตารางความเร็วที่ได้สร้างไว้ก่อนแล้ว มาหาว่ามีพิกัดใดบ้างในตารางความเร็วที่ห่างจากจุดศูนย์กลางเป็นระยะ $d/2$ แล้วนำมาเก็บไว้ในตัวแปร C
- เติมข้อมูลลงในช่องว่าง โดยนำเอาพิกัดที่อยู่ในตัวแปร C มาเติม โดยถ้าพิกัดที่เริ่มหายกับสิ้นสุดการหายอยู่บนถนนเส้นเดียวกันจะนำเอาเฉพาะพิกัดที่อยู่บนถนนเส้นนั้นมาเติม แต่ถ้าพิกัดที่เริ่มหายกับสิ้นสุดการหายอยู่บนถนนคนละเส้นกันจะนำเอาพิกัดที่อยู่บนถนนเส้นที่พิกัดเริ่มหายอยู่มาเติมก่อนหลังจากนั้นจะนำเอาพิกัดที่อยู่บนถนนที่จุดสิ้นสุดอยู่มาเติมต่อท้าย โดยจะเลือกเอาจุดที่อยู่ใกล้กับพิกัดเริ่มหายมาเติมก่อนและจึงขยับไปเรื่อย ๆ
- สร้างข้อมูลเวลาที่เติมด้วยการคำนวณตามสมการที่ 3.1
- หลังจากนั้นจะทำการวัดผลและบันทึกพิกัดที่เติม แล้วหาช่วงของพิกัดที่ขาดหายอื่นถัดไปจนถึงพิกัดสุดท้ายของข้อมูล

โดยขั้นตอนส่วนนี้สามารถทำเป็น Flowchart ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 3.7 ขั้นตอนทำงานในการเติมข้อมูล GPS ของโมเดลที่ 1

3.3.3.3 โมเดลที่ 2 เติมข้อมูล GPS ด้วยข้อมูลอ้างอิงถนนจาก K-Means Clustering

เป็นการทำนายโดยใช้ข้อมูลพิกัดของรถทั้งหมดที่จัดเก็บไว้และข้อมูลถนนมาช่วยในการทำนายโดยใช้อัลกอริทึมอยู่ทั้งหมด 2 อัลกอริทึม ได้แก่ อัลกอริทึมที่ 1 นำข้อมูลพิกัดของรถทั้งหมดมาหาจุดอ้างอิงของพิกัดโดยใช้ K-Means Clustering Method เพื่อเอาจุดศูนย์กลาง (Centroids) ของแต่ละกลุ่ม (Cluster) มาอ้างอิงแทนถนน และอัลกอริทึมที่ 2 นำข้อมูลอ้างอิงที่ได้มาเติมลงในช่องว่างของข้อมูลนำเข้าโดยจะทำการปรับพิกัดของจุดอ้างอิงให้อยู่บนเส้นถนนก่อนทำการเติมพิกัด

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์และสงวนสิทธิ์ในเนื้อหา ไม่อนุญาตให้นำไปใช้ในการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

3.3.3.3.1 อัลกอริทึมที่ 1 หาจุดอ้างอิงถนน (RefSys)

นำข้อมูลพิกัดของรถทั้งหมดที่จัดเก็บไว้ที่อยู่ในช่วงของพิกัดที่หายไปมาสร้าง K-Means Clustering ในการหาข้อมูลจุดอ้างอิงถนน โดยจะใช้ Centroids ที่ได้จาก K-Means Clustering มาเป็นจุดอ้างอิงถนนที่จะนำไปเติมในช่องว่างของข้อมูลต่อไป โดยข้อมูลจุดอ้างอิงได้จะประกอบไปด้วยละติจูด ลองจิจูดของแต่ละพิกัด โดยอัลกอริทึมนี้อาจมีปัญหาอยู่ที่จะใช้ค่า K เท่าไหร่ถึงจะเหมาะสม โดยจากการทดลองในโครงการนี้ใช้ค่า K เท่ากับ 140 จะเหมาะสมที่สุดสำหรับถนนที่มีความยาว 90 กิโลเมตรและมีจำนวนจุดทั้งหมด 19,735 จุด โดยทดสอบจากรถทั้งหมด 46 คัน

3.3.3.3.2 อัลกอริทึมที่ 2 เติมพิกัดลงในช่องว่างของข้อมูล

ในอัลกอริทึมนี้จะทำการหาจุดเริ่มต้นและจุดสุดท้ายของช่วงเวลาที่ขาดหายไป เมื่อเจอช่วงที่มีพิกัดหายไป จะนำข้อมูลของจุดอ้างอิงถนนที่อยู่ในช่วงนั้นมาเติม โดยจะทำการเติมจุดที่อยู่ใกล้กับจุดเริ่มต้นของข้อมูลที่หายก่อน ในระหว่างการเติมจะทำการวัดมุมระหว่างจุดก่อนทำการเติม เพื่อป้องกันการเติมข้อมูลผิดทิศทางในกรณีที่จุดทำมุมกันเกิน 90 องศาจะถือว่า จุดอ้างอิงจุดนั้นอยู่คนละทิศทางกับเส้นทางที่หายไป หลังจากนั้นก็จะทำการเติมจุดถัดไปจนครบ อัลกอริทึมนี้จะเติมข้อมูลของพิกัดเพียง ชุดข้อความประจำตัวของรถหรือ Vehicle ID ละติจูด ลองจิจูด ซึ่งยังขาดความเร็วและเวลาที่รถวิ่งผ่านแต่ละจุดบนถนน ในส่วนของความเร็วจะทำการทำนายโดยใช้ข้อมูลของตารางความเร็วเข้ามาช่วย โดยจะถือว่าจุดนั้นอยู่ใกล้กับจุดไหนในตารางความเร็วมากที่สุด โดยจะเทียบกับเวลาด้วย ซึ่งในส่วนของความเร็วจะทำการเติมโดยใช้ สมการดังต่อไปนี้ในการเติม

$$a^* \cdot t = p_p \cdot t + \frac{(p_n \cdot t - p_p \cdot t) \cdot d(p_p, a^*) \cdot \left| \frac{\vec{p_p a^*}}{p_p p_n} \right|}{d(p_p, p_n) \cdot \left| \frac{\vec{p_p a^*}}{p_p a^*} \right| \cdot \left| \frac{\vec{p_p p_n}}{p_p p_n} \right|} \quad (3.2)$$

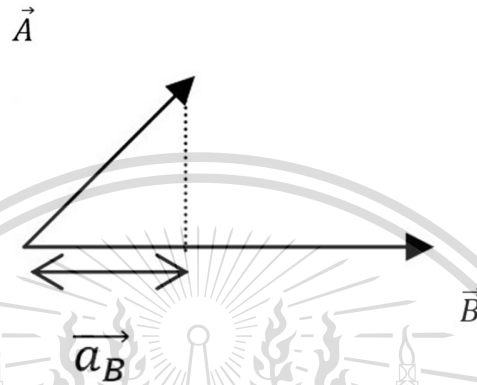
เนื่องจากจุดที่ได้จากการทำนายเป็นจุดอ้างอิงถนนที่เกิดจาก Centroids ของ K-Means Clustering ซึ่งมีผลให้บางจุดอาจมีตำแหน่งที่ไม่ได้อยู่บนถนนเมื่อเทียบกับแผนที่ ทางผู้จัดทำจึงใช้วิธีการ Projection จุดที่อยู่บนถนนให้อยู่บนถนนโดยใช้ข้อมูลของถนนมาช่วย โดยจะใช้สมการดังต่อไปนี้ในการปรับจุดให้อยู่บนถนน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

$$\vec{a}_B = \frac{\vec{A} \cdot \vec{B}}{|\vec{B}|} \quad (3.3)$$



รูป 3.8 การนำเส้น A สร้างใหม่ลงบนเส้น B

ที่มา http://www.tatc.ac.th/files/09011219194805_11052617172949.pdf

รายละเอียดขั้นตอนในการทำงานเมื่อเจอข้อมูลที่หาย 30 นาทีมีดังนี้

- เริ่มต้นจากการนำเข้าข้อมูลที่ต้องการเติมพิกัดที่หายไป โดยจะหาระยะห่างและเวลาที่แต่ละพิกัดห่างกันเพื่อหาช่วงของพิกัดที่หาย
- ตรวจสอบว่ามีช่วงที่พิกัดขาดหายหรือไม่ โดยกำหนดว่าถ้ามีค่าระยะห่างและเวลาตามนี้จะถือว่าเป็นช่วงที่หายไป ค่าระยะห่างเกิน 3 กิโลเมตร และช่วงเวลาต่างกันเกิน 1800 วินาทีหรือ 30 นาที
- ถ้าพบว่ามีช่วงที่ขาดหายไป จะทำการหาระยะห่างระหว่างทั้ง 2 จุดเก็บไว้ในตัวแปร d และทำการหาจุดศูนย์กลางระหว่างทั้ง 2 พิกัดนั้นด้วย
- นำข้อมูลพิกัดของรถที่มีโดยคัดเฉพาะคันที่ผ่านช่วงที่หายมาทำโมเดล K-Means Clustering เพื่อเก็บ Centroid มาเป็นตัวแปร RefSys
- นำข้อมูลของ RefSys ที่ได้สร้างเอาไว้มาหาว่ามีพิกัดใดบ้างใน RefSys ที่ห่างจากจุดศูนย์กลางเป็นระยะ d/2 แล้วนำมาเก็บไว้ในตัวแปร C

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น เมื่อนำไปใช้ประโยชน์ด้านการค้าไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- เติมข้อมูลพิกัดจาก RefSys ลงในช่วงที่หายไปนั้น
- สร้างข้อมูลเวลาของจุดอ้างอิงโดยคำนวณจากสมการ 3.2
- เติมความเร็วของจุดนั้นในช่วงเวลาที่ตรงกัน โดยใช้ข้อมูลจากตารางความเร็ว
- บันทึกพิกัดนั้นและขยับหาช่วงที่พิกัดขาดหายไปจนถึงจุดสุดท้าย
- ทำการ Projection พิกัดที่ทำนายได้ให้อยู่บนถนน โดยใช้สมการ 3.3
- แสดงผลและทำการวัดผล

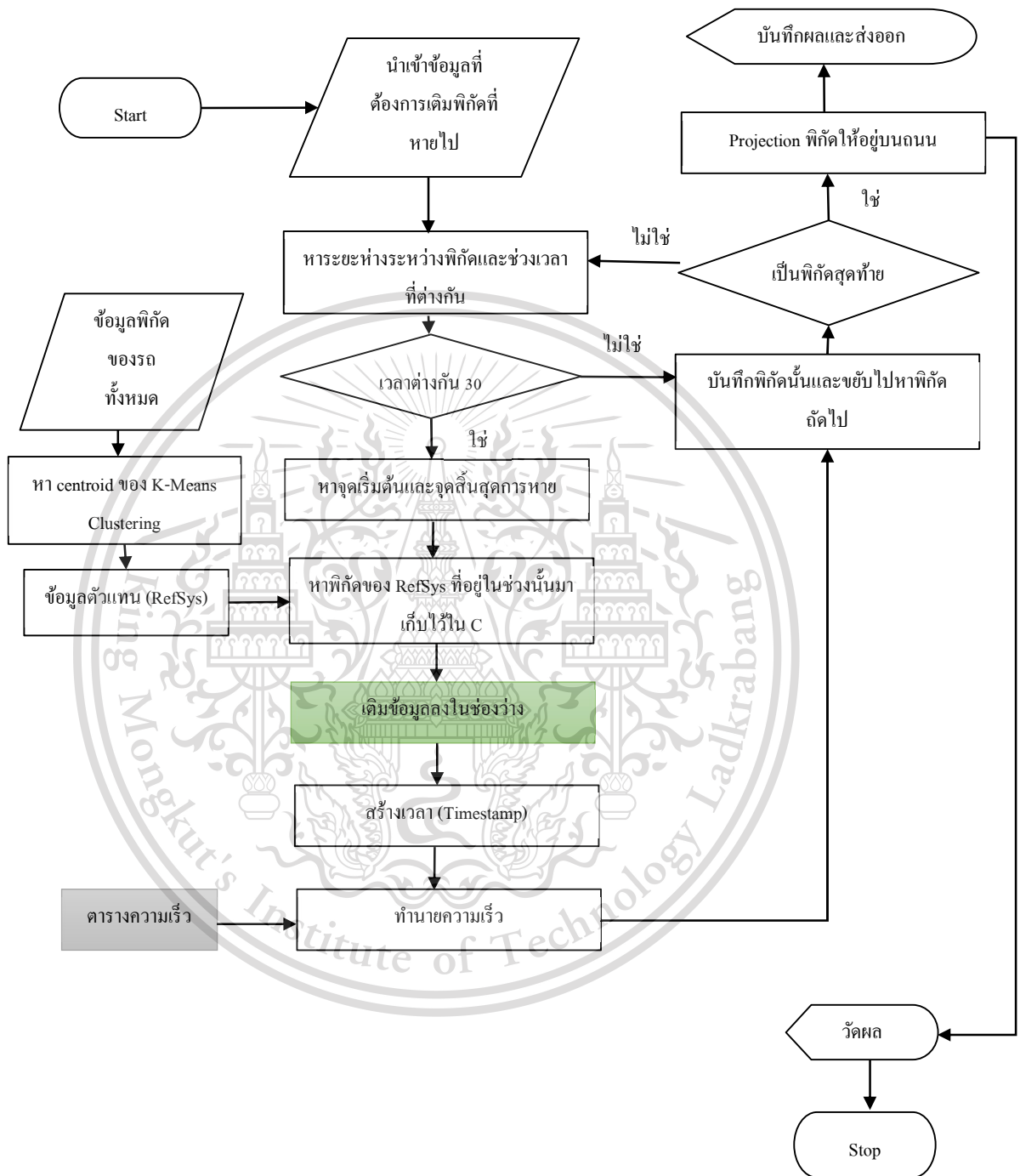
ซึ่งสามารถเขียนเป็น Flowchart ดังนี้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 3.9 ขั้นตอนทำงานในการเติมข้อมูล GPS ของโมเดลที่ 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รายละเอียดเพิ่มเติมขั้นตอนทำงานในการเติมข้อมูลของอัลกอริทึมที่ 2 โดยขยายส่วนก่อนเติมข้อมูลในช่วงที่หาย จะมีดังนี้

- หาจุดใน C ที่อยู่ใกล้กับจุดเริ่มต้นมากที่สุด แล้วนำจุดนั้นมาหามุมระหว่างจุดเริ่มกับจุดสิ้นสุดของช่วงที่หาย เพื่อดูว่าไปในทิศทางเดียวกันไหม
- ตรวจสอบว่ามุมน้อยกว่า 90 องศาหรือไม่ ถ้าไม่ใช่ จะลบจุดนั้นออกจาก C เพราะถือว่าจุดนั้นไม่ได้มีเส้นทางที่จะไปหาจุดสิ้นสุด
- ถ้ามุมน้อยกว่า 90 องศาบันทึกจุดนั้นและนำจุดนั้นเปลี่ยนเป็นจุดเริ่มต้นเพื่อหาพิกัดต่อไป และลบจุดนั้นออกจาก C เพื่อป้องกันการวนลูปซ้ำไม่จบ
- ทำวนซ้ำจน C ไม่มีข้อมูลเหลืออยู่แล้วทำการเติมพิกัดที่เก็บไว้ลงในช่วงที่หาย

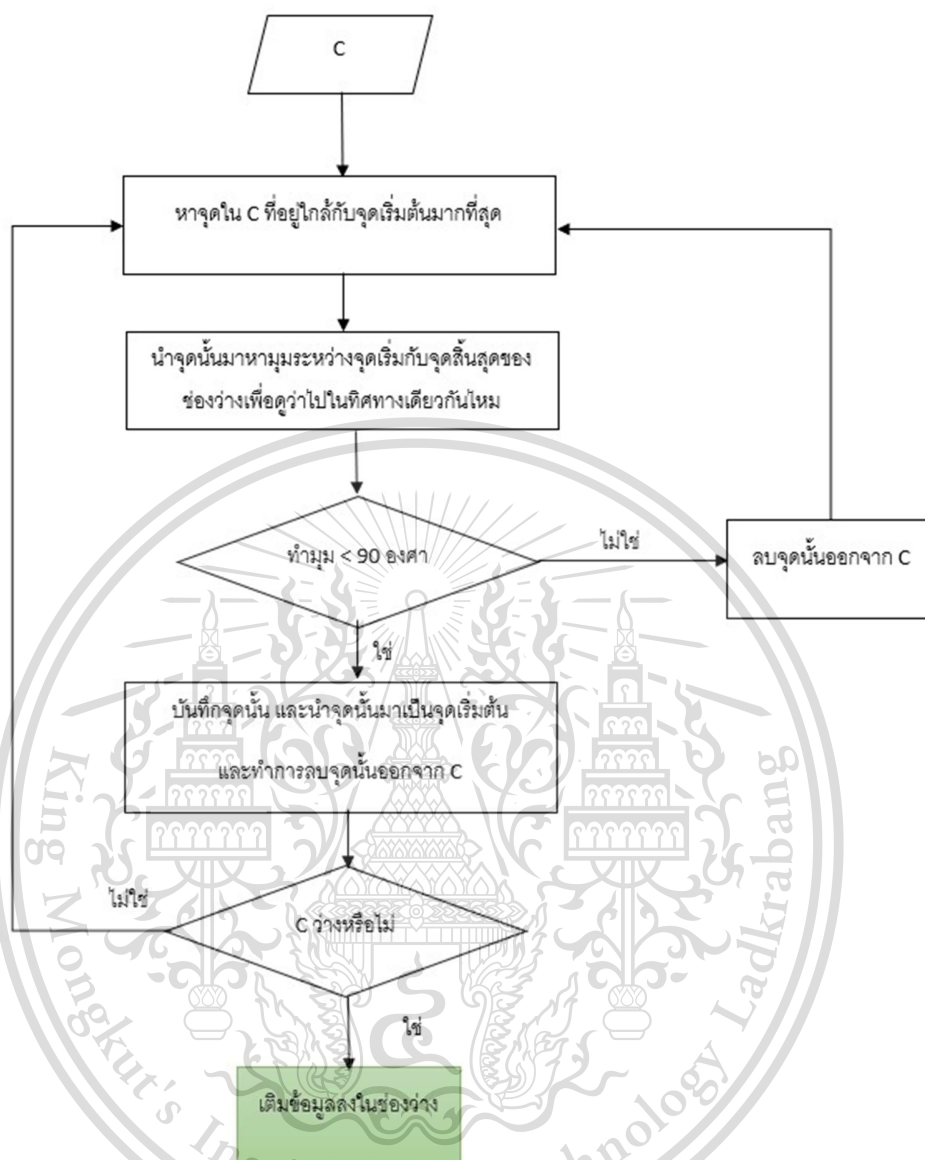
มี Flowchart ดังนี้



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 3.10 ขั้นตอนทำงานในการเติมข้อมูลของโมเดลที่ 2 โดยขยายส่วนก่อนเติมข้อมูลในช่วงที่หาย

3.3.3.4 โมเดลที่ 3 เติมข้อมูล GPS ด้วยการสร้างจุดตามค่าระยะห่างและมุม Bearing

เป็นการเติมข้อมูล GPS ด้วยการสร้างจุดขึ้นมาใหม่จากข้อมูลที่มี โดยจะใช้ค่าละติจูด, ค่าลองจิจูด, ค่ามุม Bearing ที่อิงมาจากข้อมูลถนน และค่าระยะห่างจากจุด GPS นี้กับจุดที่จะสร้างใหม่ โดยจะเอาระยะห่างมาจากตารางความเร็ว ซึ่งแปลงจากความเร็วเป็นระยะทาง โดยจะคำนวณจุดใหม่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

$$\varphi_2 = \arcsin(\sin \varphi_1 \cdot \cos \delta + \cos \varphi_1 \cdot \sin \delta \cdot \cos \theta) \quad (3.4)$$

$$\lambda_2 = \lambda_1 + \arctan_2(\sin \theta \cdot \sin \delta \cdot \cos \varphi_1, \cos \delta - \sin \varphi_1 \cdot \sin \varphi_2) \quad (3.5)$$

โดยแต่ละตัวแปรมีความหมายดังนี้

φ คือ ละติจูด, λ คือ ลองจิจูด,

θ คือ มุม Bearing, δ คือ ระยะห่างเชิงมุมที่เกิด d/R

d คือ ระยะห่างที่ทำได้ในหน่วยกิโลเมตร

R คือ รัศมีของโลกมีค่าประมาณ 6,371 กิโลเมตร

โมเดลนี้จะมีรายละเอียดขั้นตอนเมื่อเจอข้อมูลที่หาย 30 นาทีดังนี้

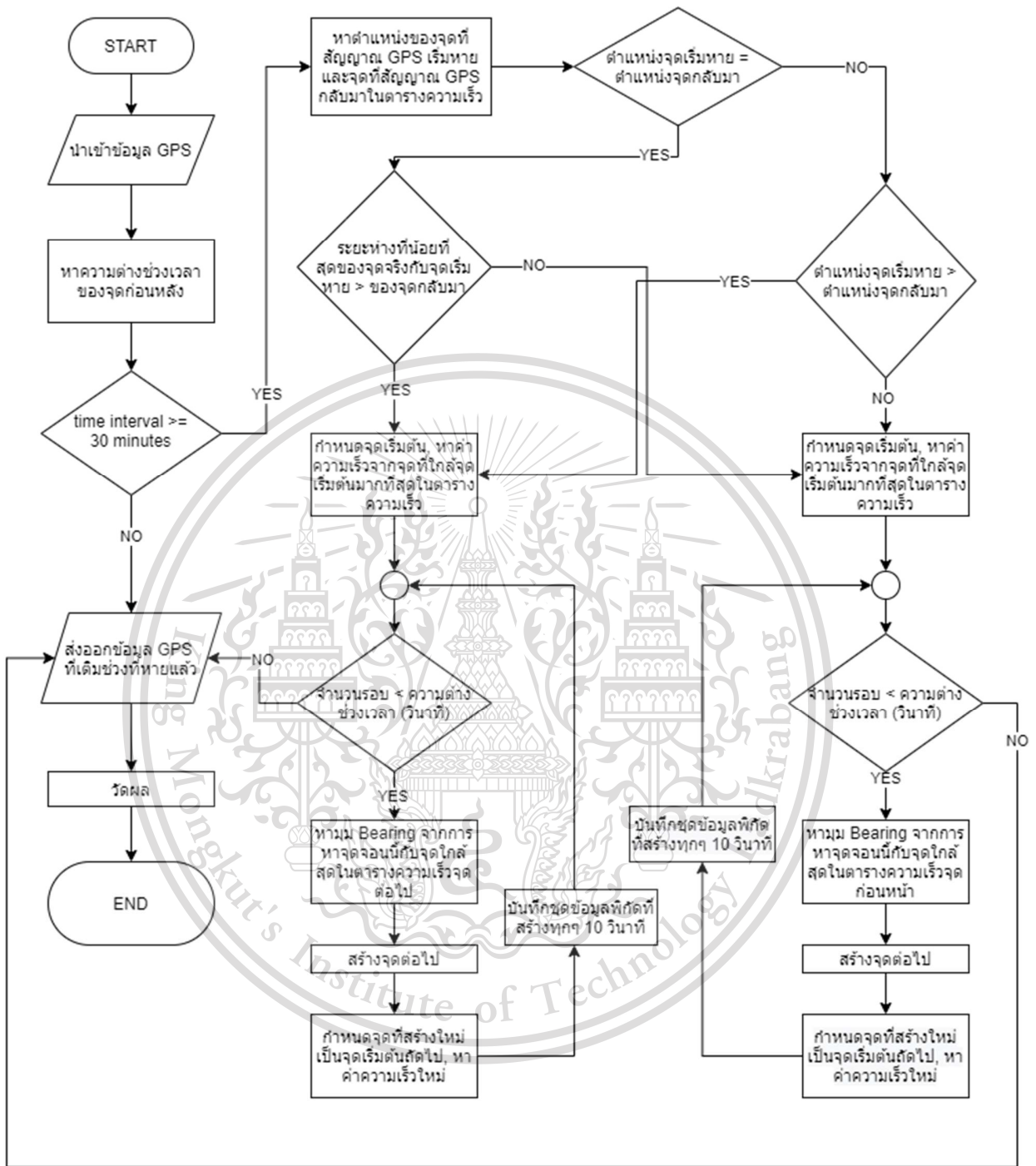
- หาข้อมูลพิกัด GPS 2 จุดติดกันที่มีเวลาห่างกันมากกว่า 30 นาที
- หาว่ารถกำลังวิ่งไปที่ด้านไหนของถนน โดยเทียบว่าจุดที่เริ่มสัญญาณหายกับจุดที่สัญญาณกลับมาปกติ ใกล้จุดไหนบ้างในตารางความเร็ว
- หามุม Bearing เทียบระหว่างจุดข้าง ๆ ทั้ง 2 โดยเรียงลำดับตามข้อมูลพิกัดในตารางเร็ว และด้านของถนนที่หาก่อนหน้านี้
- กำหนดค่าเริ่มต้นอย่างค่าพิกัดเริ่มต้น เวลาเริ่มต้น ระยะห่างจากตารางความเร็ว โดยการแปลงความเร็วเป็นระยะทาง มุม Bearing ที่หาไว้แล้ว
- สร้างจุดใหม่ทุก ๆ 1 วินาที และจะเก็บเป็นค่าที่จะใช้ส่งออกระบบทุก ๆ 10 วินาที โดยจะเปลี่ยนจุดเริ่มต้นจากจุดแรกที่สัญญาณหายเป็นจุดที่สร้างใหม่ที่ 10 วินาที

ในรูปด้านล่างจะเป็นขั้นตอนในการทำงานของโมเดลนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 3.11 ขั้นตอนทำงานในการเติมข้อมูล GPS ของโมเดลที่ 3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3.3.5 โมเดลที่ 4 เติมข้อมูล GPS ด้วยข้อมูลอ้างอิงถนน 100 เมตร, การเทียบอัตราส่วนพิกัด

โมเดลนี้เป็นการนำเอาโมเดลที่ 1 มาปรับปรุงในส่วนของการหาความเร็วของแต่ละจุด GPS ที่หายไป โดยใช้อัตราส่วนระหว่างพิกัดที่ห่างจริงเทียบกับเวลาและความเร็วโดยใช้สมการดังนี้

$$V = \frac{S}{T} \quad (3.6)$$

โดยแต่ละตัวแปรมีความหมายดังนี้

- V = ความเร็ว
- S = ระยะทาง
- T = เวลา

$$V_{new} = V + V_{ratio} \quad (3.7)$$

$$V_{ratio} = \frac{V - \text{mean}(V)}{\text{mean}(V)} \quad (3.8)$$

โดยแต่ละตัวแปรมีความหมายดังนี้

- V_{new} = ความเร็วใหม่
- Mean(V) = ค่าเฉลี่ยของความเร็ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

3.3.3.6 โมเดลที่ 5 เติมข้อมูล GPS กับรูปแบบถนนบายพาส

ในการเติมพิกัด GPS ของรถที่วิ่งผ่านถนนบายพาสนั้น จะแบ่งออกเป็น 2 ส่วน ส่วนที่ 1 จะเป็นการหาเส้นทางที่รถน่าจะวิ่งผ่าน และส่วนที่ 2 จะเป็นการเติมเต็มจุด GPS ที่หายไปบนเส้นทางนั้น

- ส่วนที่ 1 การหาเส้นทางที่เป็นไปได้ของรถ จะเลือกเอาเส้นทางที่มีรถผ่านมากที่สุด เป็นทางที่รถคันที่ต้องการหาวิ่งผ่าน
- ส่วนที่ 2 การเติมเต็มจุดที่หายไป จะเลือกใช้โมเดลในหัวข้อที่ 3.3.3.2 มาใช้ในการเติมเต็มพิกัดของทางบายพาสเนื่องจากมีความแม่นยำมากที่สุด เมื่อเทียบกับโมเดลอื่น

3.3.4 การวัดผลเปรียบเทียบแต่ละโมเดล

ใช้ Mean Absolute Error ในการวัดความผิดพลาดระหว่างจุดพิกัด GPS จริงๆกับจุดพิกัด GPS ที่เติมเต็มเข้ามาที่มีเวลาต่างกันน้อยที่สุด ทั้งระยะห่างและความเร็วที่จุดพิกัดนั้น โดยระยะห่างไม่เกิน 1 กิโลเมตรและความเร็วไม่เกิน 10 กิโลเมตรต่อชั่วโมง

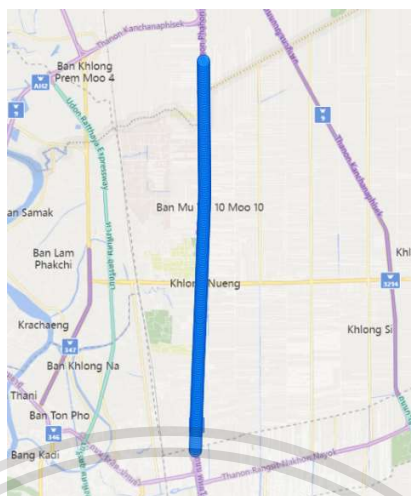
ซึ่งการทดลองนั้นจะทดลองกับรูปแบบถนน 4 รูปแบบที่ผู้จัดทำประเมินว่าเป็นประเภทเดียวกันจากถนนทั้งหมด 20 เส้น โดยแต่ละรูปแบบมีดังนี้

- รูปแบบถนนทางตรง จะวัดจากในช่วงถนนที่สนใจจะไม่มีช่วงที่ทำมุมน้อยกว่า 120 องศา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

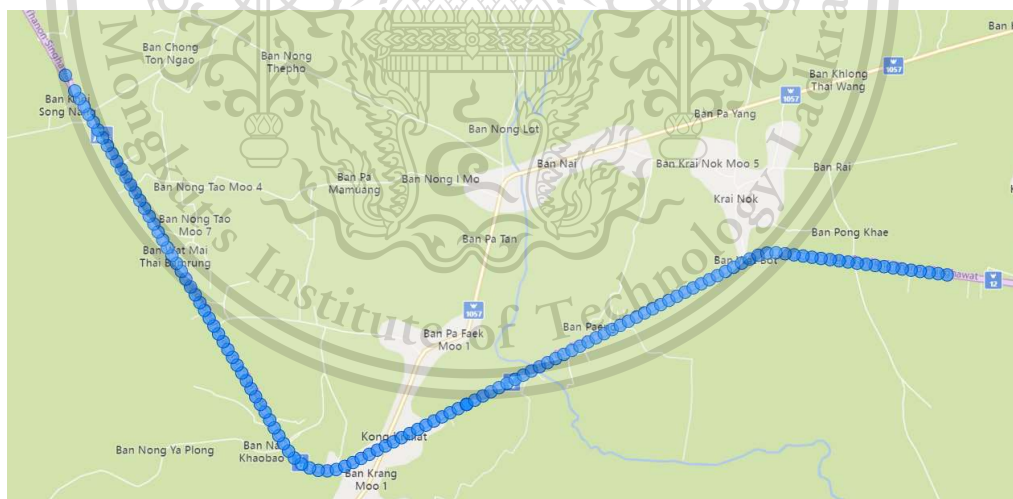
This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 3.12 ตัวอย่างถนนรูปแบบทางตรง

- รูปแบบถนนทางโค้ง จะวัดจากในช่วงถนนที่สนใจจะมีช่วงที่ทำมุมน้อยกว่า 120 องศา



รูป 3.13 ตัวอย่างถนนรูปแบบทางโค้ง

- รูปแบบถนนทางแยก จะมีลักษณะเป็น 3 แยก หรือ 4 แยก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

มีรูปถนนเพิ่มเติมในภาคผนวก ข

3.4 ฟังก์ชันในการเติมเต็มตำแหน่ง GPS ในช่วงข้อมูลที่หาย

ฟังก์ชันในการเติมเต็มตำแหน่ง GPS แบ่งเป็น 2 ส่วนหลัก คือ ส่วนที่ 1 ใช้หาเส้นทางที่เป็นไปได้ จาก 2 เส้นทาง ใช้เฉพาะกับทางบายพาสและส่วนที่ 2 ใช้เติมข้อมูล GPS ลงบนช่วงที่หายไป อินพุตที่ใช้ ในฟังก์ชันทั้ง 2 ส่วน คือ ข้อมูล GPS ที่หายไปและช่วงเวลาที่ต้องการเติม โดยยึดเอาโมเดลจากหัวข้อ 3.3.3.2 เป็นโมเดลในการเติม GPS เนื่องจากเมื่อเทียบผลแต่ละ โมเดล จุด GPS ที่ได้โมเดลนี้ใกล้เคียงกับ จุด GPS จริงมากที่สุด

3.4.1 ส่วนที่ 1 ใช้หาเส้นทางที่เป็นไปได้ ใช้เฉพาะกับทางบายพาส

ในส่วนนี้จำเป็นต้องใช้ข้อมูล GPS ของรถที่มีอยู่แล้วมาใช้คำนวณหาเส้นทาง โดยข้อมูล GPS ของรถจะคัดเฉพาะรถที่วิ่งในจังหวัดเดียวกับข้อมูล GPS ของรถที่หายไปจากอัลกอริทึมในหัวข้อ 3.1 เพื่อลดขนาดของข้อมูล แล้วนำมาสร้างเป็นข้อมูลที่ใช้ในการเลือกเส้นทาง โดยจะข้อมูล GPS ของรถที่มีอยู่แล้วและข้อมูล GPS ของรถที่จะนำมาทดสอบ โดยจะมีขั้นตอนทำงานดังนี้

- นำข้อมูล GPS ของรถที่มีอยู่แล้วมาหาพิกัดที่มีจังหวัดเหมือนกับข้อมูลที่จะนำมาทดสอบ
- หลังจากคัดเฉพาะข้อมูลที่อยู่จังหวัดเดียวกันแล้วจะเลือกเอาเฉพาะรถที่วิ่งผ่านจุดต้นทาง , จุดปลายทางเดียวกับช่วงที่หายไปของรถที่จะนำมาทดสอบ
- หลังจากนั้นจะนำเอารถที่เลือกไว้ทั้งหมดมาเทียบกับเส้นทางทั้งหมด เพื่อหาว่ารถที่วิ่ง ในช่วงนี้จะใช้เส้นทางไหนบ้าง

เมื่อได้เส้นทางกับจำนวนรถที่วิ่งผ่าน จะสามารถตัดสินใจได้ว่าเส้นทางไหนเป็นที่นิยมที่สุดและจะเลือก เส้นทางนั้นมาใช้เป็นเส้นทางที่จะเติมจุด GPS ในส่วนฟังก์ชันที่เติมจุด GPS ต่อไป

3.4.2 ส่วนที่ 2 ใช้เติมจุด GPS ในช่วงที่ข้อมูลหายไป

ในส่วนนี้จะทำการเติมจุดที่หายด้วย โมเดลจากหัวข้อที่ 3.3.3.2

เนื่องจากในการเติมจุดให้ข้อมูล GPS บนถนนรูปแบบอื่น ๆ ที่ไม่ใช่ บายพาส เช่น ทางตรง ทางโค้ง ทางแยก ไม่มีส่วนที่ต้องใช้ข้อมูล GPS ของรถที่มีอยู่แล้ว จึงไม่จำเป็นต้องใช้ส่วนที่ 1 ที่ใช้เวลานาน ทำให้การฟังก์ชันนี้เหมาะที่จะใช้กับถนนรูปแบบอื่น ๆ ที่ไม่ใช่ บายพาส

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อใช้ในการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงผลที่ได้จากการทดลองแบ่งออกเป็น 4 ส่วน ได้แก่ การหาจังหวัดของพิกัด GPS, การหาจุดอ้างอิงบนถนนในระยะ 100 เมตร, ผลที่ได้จากแต่ละโมเดลในการเติมจุด GPS ที่ขาดหาย, การใช้งานฟังก์ชัน

4.1 ผลการหาจังหวัดจากพิกัด GPS

การทดลองหาจังหวัดของพิกัด GPS จะใช้ทั้งวิธี Hashing Search และ Machine Learning Model โดยทดลองใช้ Hashing Search อยู่ 2 วิธี คือ Python Dictionary และ Pandas Dataframe ใช้ Machine Learning Model อยู่ 4 วิธี ได้แก่ Logistic Regression, Naive Bayes, Decision Tree, Deep Learning ผลที่ได้จากโมเดลที่ใช้ในการทำงานมีดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 4.1 ผลเปรียบเทียบค่าต่าง ๆ ของแต่ละโมเดลที่ใช้หาจังหวัดของพิกัด GPS

ชุดข้อมูล ที่ใช้ ฝึกสอน	ชุดข้อมูล ที่ใช้ ทดสอบ	วิธีการ	เวลาในการ ฝึกสอนทั้ง ชุดข้อมูล	ปริมาณ Memory ของ Object ของ Model	เวลาในการทดสอบ โมเดล (เฉลี่ยต่อ 1 ตำแหน่ง)	ความแม่นยำ ในการแยก ข้อมูล ทดสอบ
city-1km	taxi-city-2	Python Dictionary	1.22 sec	69,555,770	16.80/15,581,042 sec	0.98
		Pandas Dataframe	0.60 sec	223,990,846	70,028.03/15,581,942 sec	0.98
		Logistic Regression	3,728.45 sec	1,670	216.20/15,581,041 sec	0.08
		Naive Bayes	2.86 sec	11,880	419.13/15,581,041 sec	0.90
		Decision Tree	2.88 sec	1,373	65.99/15,581,041 sec	0.99
		Deep Learning	7,976.69 sec	1,197,815	718.17/15,581,041 sec	0.84

จากตาราง 4.1 โมเดลที่เลือกมาใช้จะเป็นโมเดลที่ได้จากวิธี Decision Tree เนื่องจากมีผลการทำนายที่แม่นยำ และใช้พื้นที่ Memory น้อย แม้ว่าเวลาที่ใช้ในการทำนายจะใช้มากกว่าวิธี Python Dictionary เล็กน้อย แต่เมื่อเทียบกับการใช้พื้นที่ Memory แล้วสามารถมองข้ามส่วนนี้ไปได้ ซึ่งความลึกหรือ Depth ของ Decision Tree ที่ได้ เท่ากับ 18 โดยใช้ข้อมูลในการทดสอบ 15,581,041 พิกัด

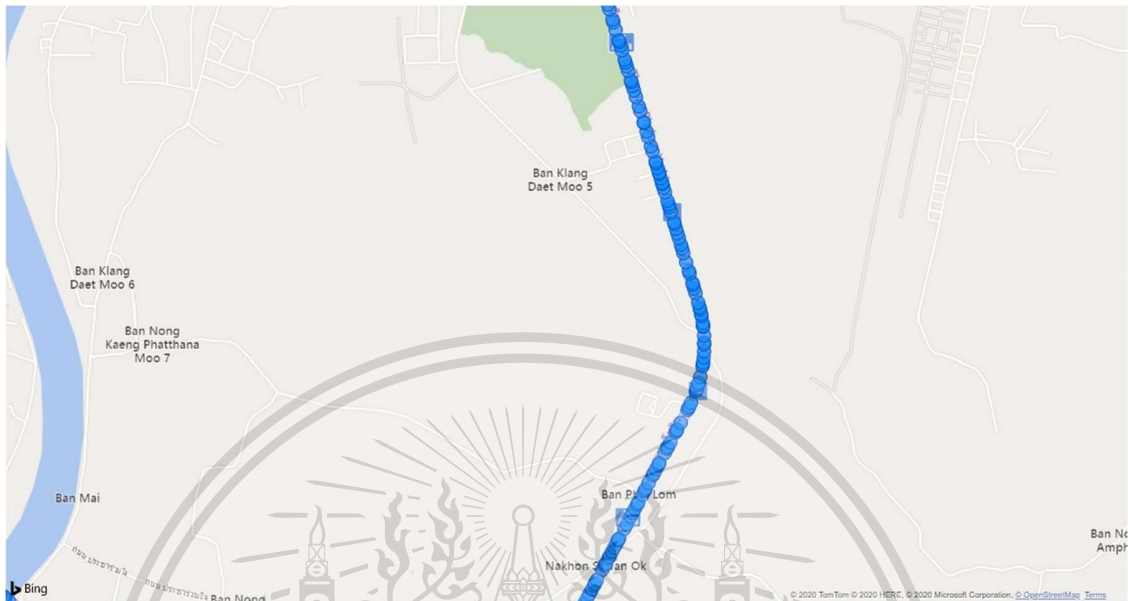
4.2 ผลการหาจุดอ้างอิงบนถนนทุก 100 เมตร

ผลที่ได้จากทำข้อมูลจุดอ้างอิงทุก 100 เมตร คือ ข้อมูลพิกัดของถนนที่แต่ละพิกัดจะอยู่ห่างกัน 100 เมตร โดยสามารถดูตัวอย่างถนนที่ยังไม่ผ่านการเก็บพิกัดและเก็บพิกัดตามรูป 4.1 และรูป 4.2

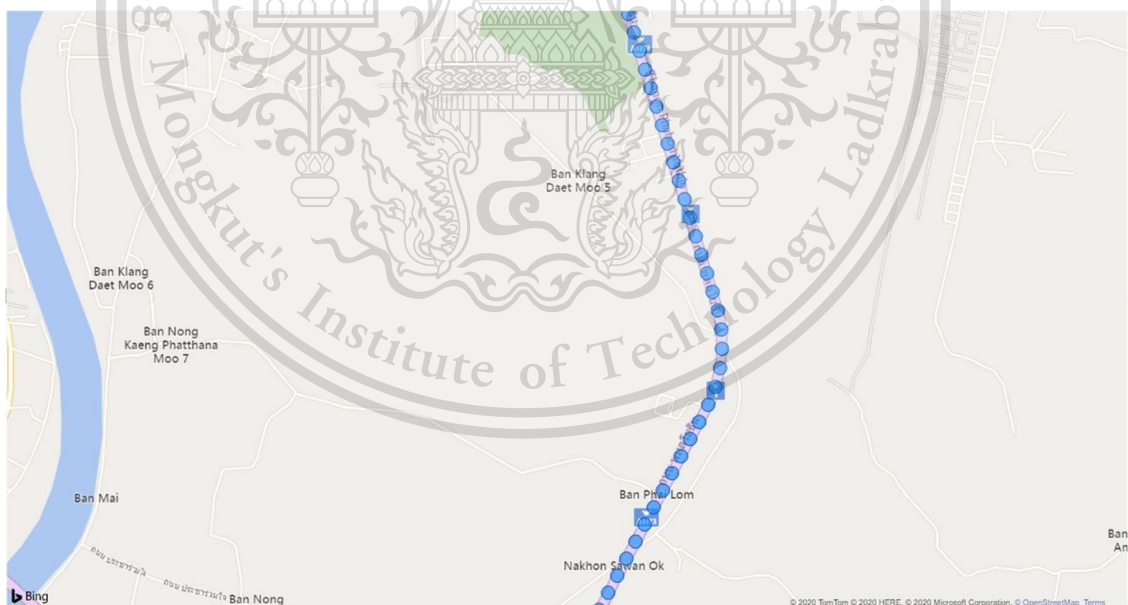
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 4.1 ข้อมูลจุดอ้างอิงถนนที่ยังไม่ได้ทำฟิสิกส์ให้ห่างกันทุก 100 เมตร



รูป 4.2 ข้อมูลจุดอ้างอิงถนนที่ทำฟิสิกส์ให้อยู่ห่างกันทุก 100 เมตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

4.3 ผลที่ได้จากโมเดลในการเติมจุดถนนที่ขาดหาย

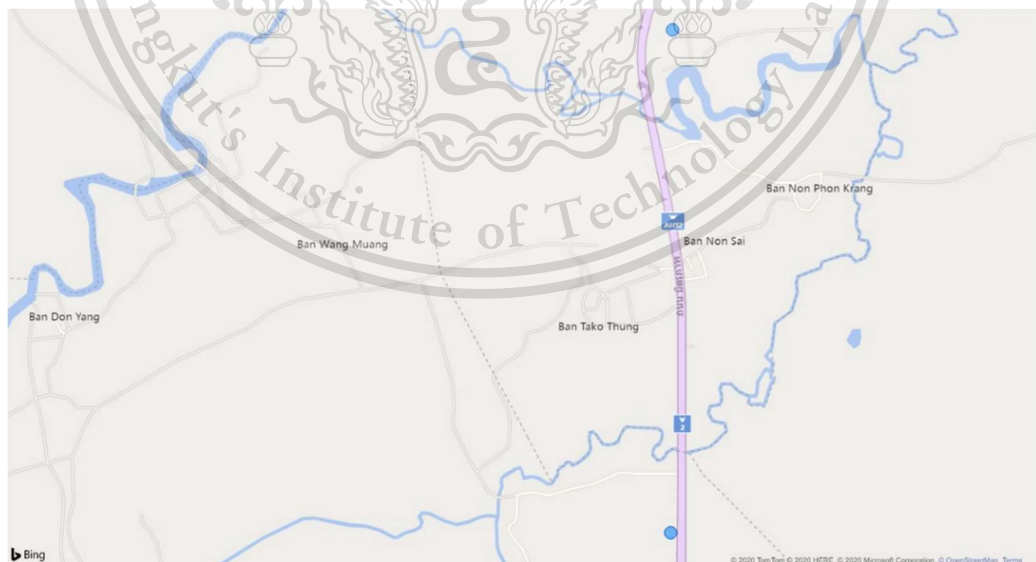
โดยโมเดลทั้งหมดจะมี 6 โมเดล ได้แก่ โมเดลในการเติมเต็มจุด GPS ที่ขาดหายมีอยู่ทั้งหมด 4 โมเดล, โมเดลในการเลือกเส้นทางสำหรับบายพาส 1 โมเดล และโมเดลของบทความที่นำมาอ้างอิง 1 โมเดล โดยแต่ละโมเดลจะทำการทดลองกับเส้นทาง 4 รูปแบบ ได้แก่ ทางตรง ทางโค้ง ทางแยกและบายพาส

4.3.1 ผลของ K-Means ที่ไม่มีการ Projection ให้พิกัดอยู่บนถนนและผลหลังจาก Projection แล้ว

ตาราง 4.2 ค่าเฉลี่ยระยะห่างของข้อมูลถนนตัวแทนที่สร้างด้วย K-Means กับข้อมูลถนนจริง

อัลกอริทึม	ระยะห่างจากจุดไปที่ถนน (กิโลเมตร)
K-Means	0.0155

โดย dist_{road} หมายถึงระยะห่างระหว่างถนนกับจุดที่ทำนายได้ ซึ่งมีค่าเฉลี่ยอยู่ที่ 15 เมตรต่อจุด และหลังจากการทำการ Projection แล้วจุดจะอยู่บนถนนทั้งหมด

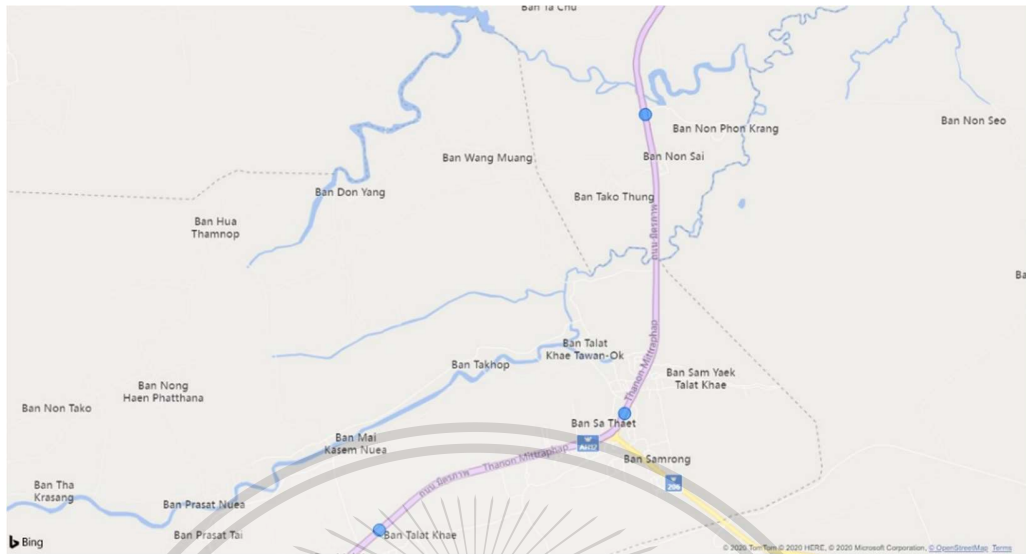


รูป 4.3 ถนนอ้างอิงจากโมเดล K-Means ก่อนทำ Projection ลงบนถนน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 4.4 ถนนอ้างอิงจากโมเดล K-Means หลังทำ Projection ลงบนถนน

4.3.2 ผลการเปรียบเทียบโมเดลที่ทดลองในทางตรง

โดยในการทดลองใช้โมเดลทั้งหมด 6 แบบ ได้แก่ K-Means (จากบทความที่นำมาอ้างอิง), K-Means Projection (จากหัวข้อที่ 3.3.3.3), 100m road (จากหัวข้อที่ 3.3.3.2), Create point with speed (mean) (จากหัวข้อที่ 3.3.3.3 โดยใช้ตารางความเร็วที่สร้างจากค่าเฉลี่ย มาใช้ในการสร้างจุดที่หายไป), Create point with speed (slope) (จากหัวข้อที่ 3.3.3.3 โดยใช้ตารางความเร็วที่สร้างจากค่าชันของจุดที่อยู่ก่อนและหลัง มาใช้ในการสร้างจุดที่หายไป), 100 m + ratio (จากหัวข้อที่ 3.3.3.4)

ในการทดลองนี้ได้เลือกถนนถนนหมายเลข 24 ที่ผ่านจังหวัดนครราชสีมา โดยมีรายละเอียดดังนี้

- ระยะทางที่ใช้ทั้งหมด 33 กิโลเมตร
- ประเภทของข้อมูลรถที่ใช้ คือ รถแท็กซี่
- ข้อมูลฝึกสอนอยู่ในช่วงปีพุทธศักราช 2560 ถึง 2562 เป็นเวลา 2 ปี มีจำนวนรถอยู่ทั้งหมด 493 คัน เป็นแถวพิกัด 9,098 พิกัด
- ข้อมูลทดสอบอยู่ในปีพุทธศักราช 2563 เป็นเวลา 1 เดือน มีจำนวนรถอยู่ทั้งหมด 20 คัน
- ค่า K ที่ใช้กับ K-Means Clustering คือ 16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ตาราง 4.3 เปรียบเทียบผล MAE ของระยะทางและความเร็วของแต่ละโมเดลกับทางตรง

อัลกอริทึม	ค่า Mean Absolute Error	
	ระยะทาง (km)	ความเร็ว (km/hr)
K-Means	0.674	15.360
K-Means Projection	0.669	15.406
100m road	0.328	15.283
Create point with speed (mean)	1.907	16.841
Create point with speed (slope)	1.915	17.799
100 m + ratio	0.548	13.338

จากค่าที่ได้มาข้างต้นจะเห็นได้ว่า ในเรื่องระยะห่างระหว่างจุดโมเดลที่ให้ค่าแม่นยำที่สุด คือ โมเดล 100m road ส่วนเรื่องความเร็วเป็นโมเดล 100m + ratio ที่แม่นยำที่สุด จากตัวอย่างในตารางข้างต้นระยะห่างระหว่างจุดที่หายเฉลี่ย 33 กิโลเมตร โมเดล 100m road และโมเดล 100m + ratio จะต้องสร้างจุดมากถึง 330 จุด แต่ในกรณีของอัลกอริทึม K-Means Projection จะสร้างเพียง 16 จุดตามค่า K ที่ใช้คือ 16 และอีกทั้งความแม่นยำในการทำนายของอัลกอริทึม 100m road กับ K-Means Projection ต่างกันไม่มากนักและพิกัดที่ได้อยู่บนถนนทั้งคู่ ดังนั้นอัลกอริทึม K-Means Projection จึงน่าสนใจมากกว่าที่จะใช้ในการทำนายพิกัดในถนนเส้นตรง

4.3.3 ผลการเปรียบเทียบโมเดลที่ทดลองในทางโค้ง

โดยในการทดลองใช้โมเดลทั้งหมด 6 แบบได้แก่ K-Means (จากบทความที่นำมาอ้างอิง) , K-Means Projection (จากหัวข้อที่ 3.3.3.2), 100m road (จากหัวข้อที่ 3.3.3.1), Create point with speed (mean) (จากหัวข้อที่ 3.3.3.4 ตารางความเร็วที่ใช้จะสร้างจากค่าเฉลี่ย), Create point with speed (slope) (จากหัวข้อที่ 3.3.3.4 ตารางความเร็วที่ใช้จะสร้างจากค่าชันของจุดที่อยู่ก่อนและหลัง), 100 m + ratio (จากหัวข้อที่ 3.3.3.5)

ในการทดลองได้เลือกถนนถนนหมายเลข 33 ที่ผ่านจังหวัดปราจีนบุรี โดยมีรายละเอียดดังนี้

- ระยะทางที่ใช้ทั้งหมด 17 กิโลเมตร

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ที่ใช้นั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- ข้อมูลฝึกสอนอยู่ในช่วงปีพุทธศักราช 2560 ถึง 2562 เป็นเวลา 2 ปี มีจำนวนรถอยู่ทั้งหมด 276 คัน เป็นแถวพิกัด 4,057 พิกัด
- ข้อมูลทดสอบอยู่ในปีพุทธศักราช 2563 เป็นเวลา 1 เดือน มีจำนวนรถอยู่ทั้งหมด 17 คัน
- ค่า K ที่ใช้กับ K-Means Clustering คือ 26

ตาราง 4.4 เปรียบเทียบผล MAE ของระยะทางและความเร็วของแต่ละโมเดลกับทางโค้ง

อัลกอริทึม	ค่า Mean Absolute Error	
	ระยะทาง (km)	ความเร็ว (km/hr)
K-Means	0.401	14.587
K-Means Projection	0.390	14.559
100m road	0.310	14.617
Create point with speed (mean)	0.739	13.583
Create point with speed (slope)	0.749	14.005
100 m + ratio	0.527	13.501

จากค่าที่ได้มาข้างต้นจะเห็นได้ว่าโมเดล 100m road แม่นยำมากที่สุด ในเรื่องของระยะทางระหว่างจุด และโมเดล 100m + ratio แม่นยำที่สุดในเรื่องความเร็ว ผลที่ได้ใกล้เคียงกับทางตรง ดังนั้นโมเดลที่ใช้สำหรับทางตรงและทางโค้งสามารถใช้งานร่วมกันได้

4.3.4 ผลการเปรียบเทียบโมเดลที่ทดลองในทางแยก

โดยในการทดลองใช้โมเดลทั้งหมด 6 แบบได้แก่ K-Means (จากบทความที่นำมาอ้างอิง), K-Means Projection (จากหัวข้อที่ 3.3.3.2), 100m road (จากหัวข้อที่ 3.3.3.1), Create point with speed (mean) (จากหัวข้อที่ 3.3.3.4 ตารางความเร็วที่ใช้จะสร้างจากค่าเฉลี่ย), Create point with speed (slope) (จากหัวข้อที่ 3.3.3.4 ตารางความเร็วที่ใช้จะสร้างจากค่าชันของจุดที่อยู่ก่อนและหลัง), 100 m + ratio (จากหัวข้อที่ 3.3.3.5)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ในการทดลองได้เลือกถนนถนนหมายเลข 2 และ 207 ที่ผ่านจังหวัดนครราชสีมา โดยมีรายละเอียดดังนี้

- ระยะทางที่ใช้ทั้งหมด 10 กิโลเมตร
- ประเภทของข้อมูลรถที่ใช้ คือ รถแท็กซี่
- ข้อมูลฝึกสอนอยู่ในช่วงปีพุทธศักราช 2560 ถึง 2562 เป็นเวลา 2 ปี มีจำนวนรถอยู่ทั้งหมด 298 คัน เป็นแถวพิกัด 2,703 พิกัด
- ข้อมูลทดสอบอยู่ในปีพุทธศักราช 2563 เป็นเวลา 1 เดือน มีจำนวนรถอยู่ทั้งหมด 14 คัน
- ค่า K ที่ใช้กับ K-Means Clustering คือ 20

ตาราง 4.5 เปรียบเทียบผล MAE ของระยะทางและความเร็วของแต่ละโมเดลกับทางแยก

อัลกอริทึม	ค่า Mean Absolute Error	
	ระยะทาง (km)	ความเร็ว (km/hr)
K-Means	0.550	19.872
K-Means Projection	0.547	20.297
100m road	0.455	22.115
Create point with speed (mean)	1.096	28.244
Create point with speed (slope)	1.123	27.460
100 m + ratio	0.939	25.453

จากค่าที่ได้มาข้างต้นจะเห็นได้ว่าในเรื่องระยะทางระหว่างจุด โมเดล 100m road ให้ผล MAE แม่นยำมากที่สุด และในด้านความเร็วจะเป็นโมเดล K-Means ที่ใกล้เคียงที่สุด

4.3.5 ผลการทดลองที่แสดงการเปรียบเทียบโมเดลที่ทดลองในทางบายพาส

โดยในการทดลองใช้โมเดลทั้งหมด 7 แบบ ได้แก่ K-Means (จากบทความที่นำมาอ้างอิง), K-Means Projection (จากหัวข้อที่ 3.3.3.2), 100m road (จากหัวข้อที่ 3.3.3.1), Create point with speed (mean) (จากหัวข้อที่ 3.3.3.4 ตารางความเร็วที่ใช้จะสร้างจากค่าเฉลี่ย), Create point with speed (slope)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่าจะกรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

(จากหัวข้อที่ 3.3.3.4 ตารางความเร็วที่ใช้จะสร้างจากค่าชั้นของจุดที่อยู่ก่อนและหลัง), 100 m + ratio (จากหัวข้อที่ 3.3.3.5), บายพาส (จากหัวข้อที่ 3.3.3.6)

ในการทดลองได้เลือกถนนถนนหมายเลข 2 และ 204 ที่ผ่านจังหวัดนครราชสีมา โดยมีรายละเอียดดังนี้

- ระยะทางที่ใช้ทั้งหมด 29 กิโลเมตร
- ประเภทของข้อมูลรถที่ใช้ คือ รถแท็กซี่
- ข้อมูลฝึกสอนอยู่ในช่วงปีพุทธศักราช 2560 ถึง 2562 เป็นเวลา 2 ปี มีจำนวนรถอยู่ทั้งหมด 660 คัน เป็นแถวพิกัด 14,321 พิกัด
- ข้อมูลทดสอบอยู่ในปีพุทธศักราช 2563 เป็นเวลา 1 เดือน มีจำนวนรถอยู่ทั้งหมด 32 คัน
- ค่า K ที่ใช้กับ K-Means Clustering คือ 22

ตาราง 4.6 เปรียบเทียบผล MAE ของระยะทางและความเร็วของแต่ละโมเดลกับทางบายพาส

อัลกอริทึม	ค่า Mean Absolute Error	
	ระยะทาง (km)	ความเร็ว (km/hr)
K-Means	3.915	22.187
K-Means Projection	3.863	21.632
100m road	4.815	17.293
Create point with speed (mean)	6.265	22.795
Create point with speed (slope)	6.266	26.64
100 m + ratio	5.344	21.679
Bypass	3.846	17.293

ในการทดลองจะเห็นได้ว่าโมเดลที่ใช้กับทางตรง ทางโค้ง ทางแยกนั้น ไม่สามารถนำมาใช้งานกับทางบายพาสได้เนื่องจาก ต้องเลือกทางที่รถจะผ่านด้วยซึ่งในทางบายพาสจะมีเส้นทางที่รถผ่านได้หลายเส้น จึงต้องมีโมเดลในการเลือกเส้นทางก่อนที่จะทำการเติมจุด

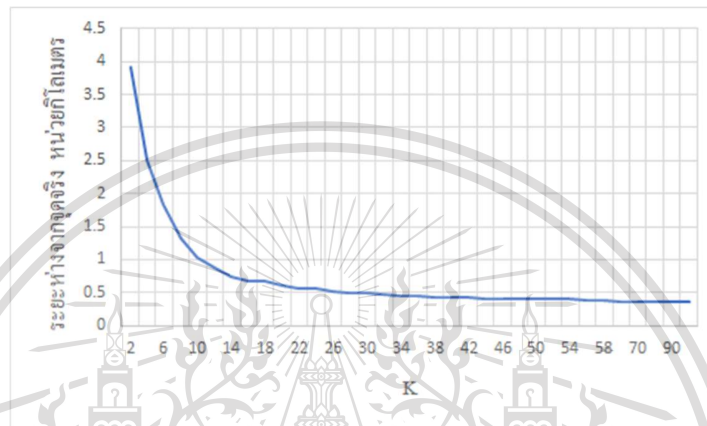
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

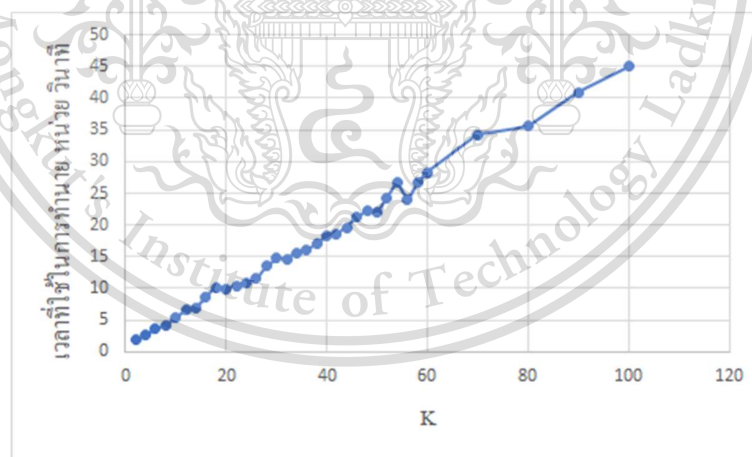
Forbidden to modify the content, and cite the document when use.

4.3.6.1 ค่า K ที่เหมาะสมกับทางตรง

ใช้ Elbow method ในการหาค่า K ที่เหมาะสม จะได้ผลลัพธ์ตามกราฟตัวอย่างด้านล่าง ซึ่งสร้างจากรถ 20 คัน ที่มีพิกัดที่หายไปเป็นระยะเวลา 20 นาที ระยะทางที่หายไปเฉลี่ย 33 กิโลเมตรต่อคัน



รูป 4.7 จำนวนค่า K ที่ใช้และระยะห่างของพิกัดที่ทำนายกับพิกัดจริงกับรูปแบบถนนทางตรง



รูป 4.8 จำนวนค่า K ใช้และเวลาของโมเดลที่ใช้ขณะทำนายผลกับรูปแบบถนนทางตรง

จากกราฟจะสรุปได้ว่าค่า K ที่เหมาะสมกับ K-Means คือ 16 สำหรับถนนที่มีลักษณะ
 เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
 เป็นทางตรง

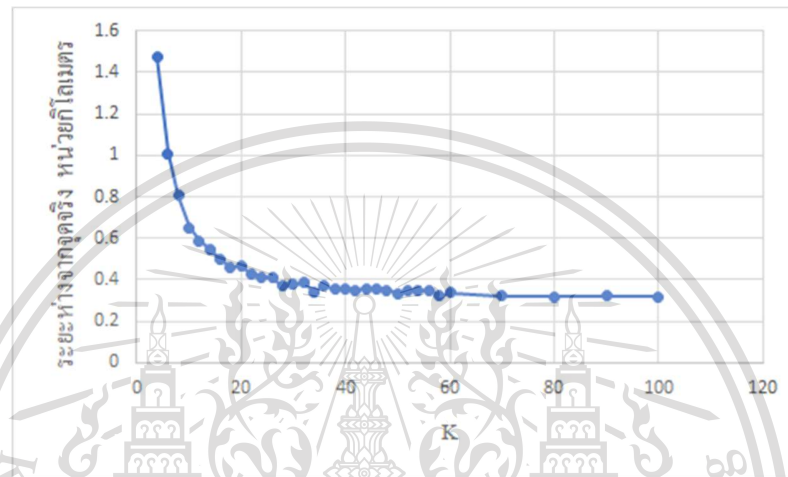
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

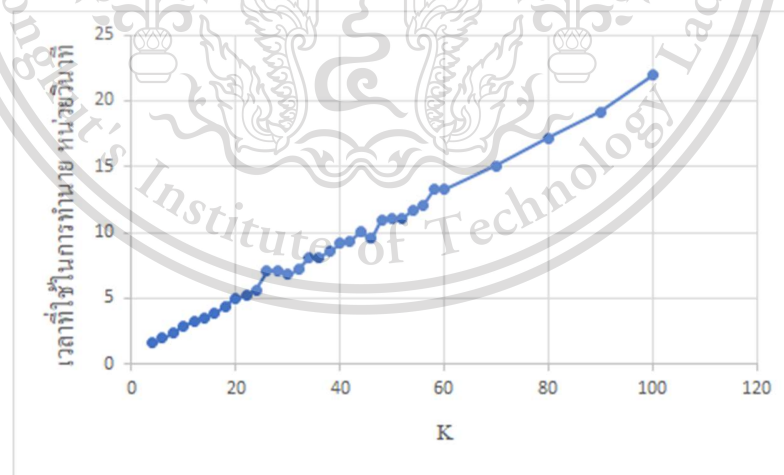
Forbidden to modify the content, and cite the document when use.

4.3.6.2 ค่า K ที่เหมาะสมกับทางโค้ง

ใช้ Elbow method ในการหาค่า K ที่เหมาะสม จะได้ผลลัพธ์ตามกราฟตัวอย่างด้านล่าง ซึ่งสร้างจากรถ 17 คัน ที่มีพิกัดที่หายไปเป็นระยะเวลา 20 นาที ระยะทางที่หายไปเฉลี่ย 17 กิโลเมตรต่อคัน



รูป 4.9 จำนวนค่า K ที่ใช้และระยะห่างของพิกัดที่ทำนายกับพิกัดจริงกับรูปแบบถนนทางโค้ง



รูป 4.10 จำนวนค่า K ใช้และเวลาของโมเดลที่ใช้ขณะทำนายผลกับรูปแบบถนนทางโค้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

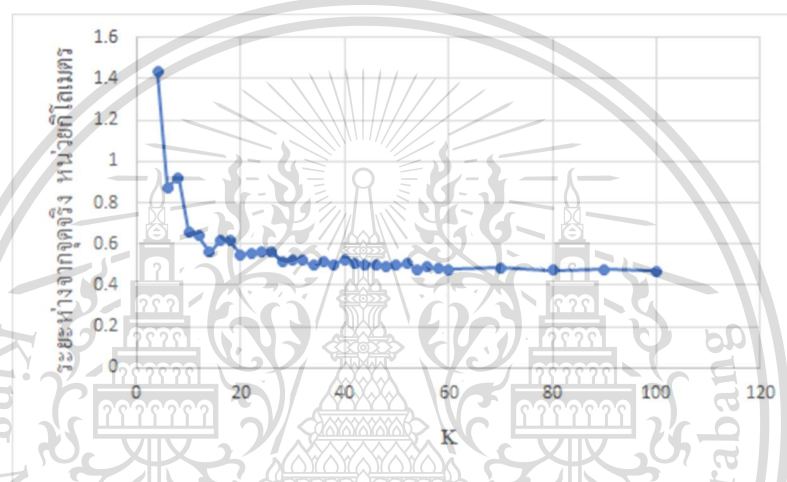
This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

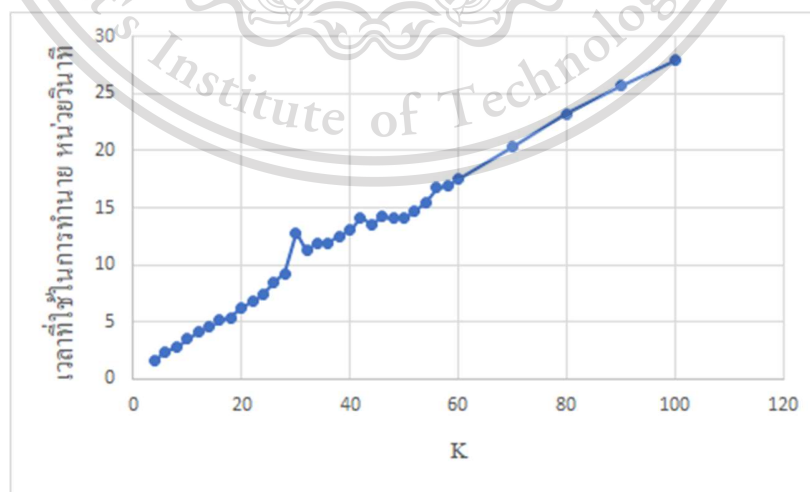
จากกราฟจะสรุปได้ว่าค่า K ที่เหมาะสมกับ K-Means คือ 26 สำหรับถนนที่มีลักษณะเป็นทางโค้ง

4.3.6.3 ค่า K ที่เหมาะสมกับทางแยก

ใช้ Elbow method ในการหาค่า K ที่เหมาะสม จะได้ผลลัพธ์ตามกราฟตัวอย่างด้านล่าง ซึ่งสร้างจากรถ 14 คัน ที่มีพิกัดที่หายไปเป็นระยะเวลา 10 นาที ระยะทางที่หายเฉลี่ย 10 กิโลเมตรต่อคัน



รูป 4.11 จำนวนค่า K ที่ใช้และระยะห่างของพิกัดที่ทำนายกับพิกัดจริงกับรูปแบบถนนทางแยก



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้เฉพาะในวงการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

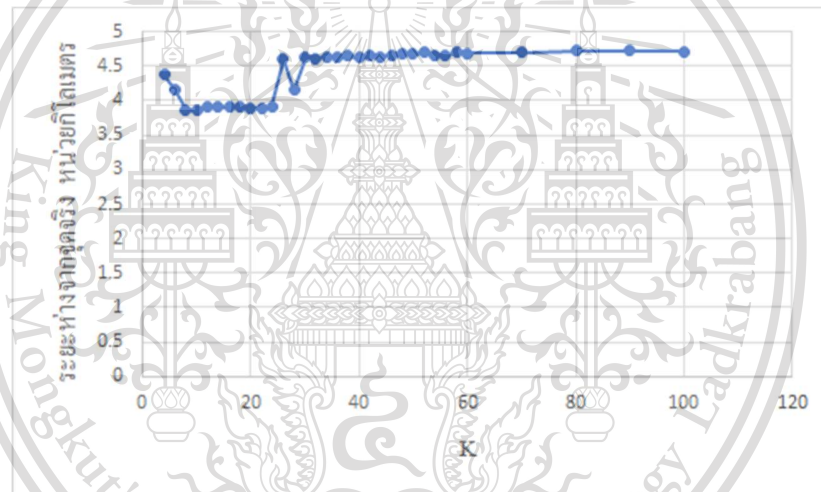
This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

จากกราฟจะสรุปได้ว่าค่า K ที่เหมาะสมกับ K-Means คือ 20 สำหรับถนนที่มีลักษณะเป็นทางแยก

4.3.5.4 ค่า K ที่เหมาะสมกับทางบายพาส

ใช้ Elbow method ในการหาค่า K ที่เหมาะสม จะได้ผลลัพธ์ตามกราฟตัวอย่างด้านล่าง ซึ่งสร้างจากรถ 31 คัน ที่มีพิกัดที่หายไปเป็นระยะเวลา 20 นาที ระยะทางที่หายไปเฉลี่ย 24 กิโลเมตรต่อคัน

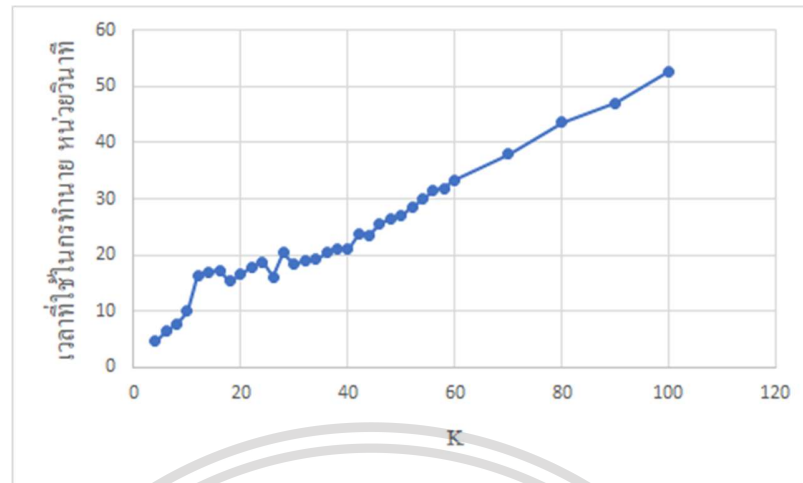


รูป 4.13 จำนวนค่า K ที่ใช้และระยะห่างของพิกัดที่ทำนายกับพิกัดจริงกับรูปแบบถนนทางบายพาส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.



รูป 4.14 จำนวนค่า K ใช้และเวลาของโมเดลที่ใช้ขณะทำนายผลกับรูปแบบถนนทางบายพาส

เนื่องจากกราฟที่ได้ไม่สามารถใช้ Elbow method ในการหาค่า K ได้ ดังนั้นในการทดลองนี้จึงเลือกใช้ค่า K ที่ให้ค่าระยะห่างจากจุดจริงน้อยที่สุดนั่นก็คือ 22

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บทที่ 5

สรุปผลและปัญหาของโครงการ

ในบทนี้จะเป็นส่วนสรุปผลการพัฒนาทั้งหมดที่ได้ทดลองมาตลอดปีการศึกษา ปัญหาที่ทางผู้จัดทำได้เจอระหว่างการทำงานและแนวทางที่ได้ใช้แก้ไข และแนวทางสำหรับผู้ที่จะนำโครงการนี้ไปพัฒนาต่อ

5.1 บทสรุปของโครงการ

สิ่งที่พัฒนาในโครงการนี้หลัก ๆ สามารถแยกได้เป็น 4 ส่วนดังนี้

5.1.1 การหาจังหวัดของพิกัด GPS

โมเดลที่ใช้หาจังหวัดของพิกัด GPS ที่ให้ความแม่นยำที่สูงและไม่เสียจำนวนพื้นที่ในการเก็บตัวโมเดล คือ Decision Tree แม้ว่าความเร็วในการสร้างโมเดลจะช้ากว่าโมเดลอื่นอยู่บ้างก็ตาม ในส่วนจำนวนความชันของ Decision Tree คือ 18

5.1.2 ข้อมูลถนน 100 เมตร

ผลลัพธ์ที่ได้คือข้อมูลจุดอ้างอิงถนนที่มีระยะห่างของแต่ละพิกัดเป็น 100 เมตร และมีการเก็บจุดพิกัดที่วัดมุมโค้งได้น้อยกว่า 120 องศา เพื่อบอกว่าที่ตำแหน่งของจุดตรงนั้นเป็นทางโค้ง

5.1.3 การเติมจุดถนนที่ขาดหาย

โมเดลที่เลือกในการเติมจุด GPS ที่ขาดหายในถนนที่มีรูปแบบเป็นทางตรง คือ K-Means Projection แม้ว่าผลลัพธ์ที่ได้จะแยกว่าการใช้ถนน 100 เมตรเข้าไปแทน แต่ผลลัพธ์ที่ได้จาก K-Means Projection ไม่ได้ต่างกันมาก สิ่งที่ทำให้เลือก K-Means Projection คือขนาดพื้นที่ในการจัดเก็บโมเดล เพราะข้อมูล 100 เมตรใช้จำนวนจุดที่มากต่อถนน 1 เส้น ในขณะที่ K-Means Projection ใช้เพียง 15 จุด ในกรณีที่ใช้ค่า K เท่ากับ 15 ซึ่งในทางที่มีลักษณะทางตรง ทางโค้ง หรือทางแยก สามารถใช้ K-Means Projection ในการทำนายได้ แต่ในส่วนของถนนทางโค้งจะต้องนำข้อมูลของรถมาเลือกเส้นทางก่อน

เอกสารนี้เป็นเอกสารที่จะเติมจุด GPS ลงไปสำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

5.1.4 การสร้างฟังก์ชันในการทำนายจุด

ในการทำนายจุดจะเลือกใช้โมเดล 100 เมตร มาใช้ในการสร้างเนื่องจากโมเดล 100 เมตร จะใช้แค่ข้อมูลถนน ไม่จำเป็นต้องใช้ข้อมูลของการวิ่งของรถทั้งหมดที่มี จึงทำให้ประหยัดพื้นที่ในการจัดเก็บและประหยัดเวลาในการสร้างอีกด้วย และส่วนของความเร็วจะใช้ความเร็วเฉลี่ยของรถคันนั้นแทนการสร้างตารางความเร็วเพื่อลดขนาดข้อมูลที่ใส่และประหยัดเวลา ซึ่งฟังก์ชันนี้จะใช้ได้กับข้อมูล GPS ที่อยู่บนถนนที่เป็นทางตรง ทางโค้ง และทางแยกเท่านั้น ในกรณีที่ต้องการนำไปใช้กับทางบายพาสจะต้องมีฟังก์ชันในการหาถนนที่รถนิยมวิ่งผ่านก่อนจึงจะสามารถนำมาทำนายจุดต่อได้

5.2 ปัญหา อุปสรรค และแนวทางแก้ไข

- 1) ฐานข้อมูลถนนที่นำเข้าระบบมีส่วนที่ไม่ครบถ้วน เช่น ถนนบางเส้นไม่มีในฐานข้อมูล บางเส้นมีส่วนที่ขาดช่วง การเก็บข้อมูลถนนส่วนเดิมซ้ำ จึงเลือกถนนเส้นที่มีข้อมูลสมบูรณ์มาใช้งาน
- 2) ในช่วงถนนที่เลือกมาทดลอง มีบางจุดที่พอดูข้อมูลการตำแหน่งของรถยนต์แล้วมีจำนวนค่าน้อยกว่าที่คาดไว้ ทำให้ต้องขยายช่วงถนนหรือเปลี่ยนช่วงถนนที่ใช้
- 3) ในช่วงถนนที่เลือกมาทดลอง ในบางช่วงเวลาไม่มีรถวิ่งอยู่เลยจึงทำให้การทำตารางความเร็วที่ต้องการความเร็วของรถเป็นรายชั่วโมงหรือเวลาอื่นมีข้อมูลว่างหรือเป็น 0 ในช่วงนั้น จึงได้ใช้ข้อมูลรถในช่วงเวลาอื่นแทน โดยแบ่งออกเป็น 2 วิธี โดยตามค่าเฉลี่ยของทั้งวันกับเทียบช่วงเวลาหน้าหลังเพื่อปรับเพิ่มลดแบบขึ้นบันได
- 4) ข้อมูลรถบรรทุกที่ใช้มีการรับส่งสัญญาณ GPS ที่ไม่ต่อเนื่อง ทำให้มีผลต่อการวัดผลระยะห่างของจุดจริงและจุดที่เติมเต็มเข้ามาใหม่ ทำให้อาจจะใช้ชุดข้อมูลของยานพาหนะประเภทอื่นแทน
- 5) ข้อมูลรถแท็กซี่ ขณะมีผู้โดยสารมักจะวิ่งเส้นทางเดิม ๆ ทำให้ต้องเปลี่ยนถนนหรือเพิ่มช่วงเวลาที่ต้องการศึกษา
- 6) จำนวนข้อมูลที่ผ่านถนนบายพาสที่สนใจมีจำนวนน้อยจึงทำให้ข้อมูลไม่เพียงพอในหาข้อแตกต่างที่ชัดเจนในการเลือกตัดสินใจเส้นทางที่จะเติมจุด GPS

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

5.3 แนวทางการพัฒนาต่อ

- 1) ทาวิธีในการเก็บโมเดลที่จะนำมาใช้ในการเลือกเส้นทางขายพาสหรือลดขนาดข้อมูลฝึกสอนเพื่อลดเวลาในการประมวลผล
- 2) เพิ่มข้อมูลรถเพื่อสร้างตารางความเร็วให้มีความแม่นยำมากขึ้น
- 3) พัฒนาให้โมเดลสามารถใช้ได้กับถนนทุกรูปแบบ
- 4) พัฒนาให้สามารถนำโมเดลไปใช้กับถนนท้องถิ่นได้
- 5) ศึกษาพฤติกรรมของรถที่มีการเลี้ยวกลับ



เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

บรรณานุกรม

Fabrício A. Silva, Clayson Celes, Azzedine Boukerche, Linnyer B. Ruiz and Antonio A. F. Loureiro. 2015. Filling the Gaps of Vehicular Mobility Traces. 47–54. in **MSWiM '15: Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems**. New York : Association for Computing Machinery.

Guicheng Shen, Cailin Zhang and Baicheng Tang. 2015. Completing Truck's Missing Trajectory Based on The Historical Information. in **2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)**. Beijing : Beijing Wuzi University

กรมทางหลวง. 2560. ระบบหมายเลขทางหลวง. [Online]. Available :
<http://www.doh.go.th/content/page/page/119>

ทางหลวงในประเทศไทย. [Online]. Available : <https://www.wikiwand.com/th/ทางหลวงในประเทศไทย>

Movable Type. **Calculate distance, bearing and more between Latitude/Longitude points**.

[Online]. Available : <https://www.movable-type.co.uk/scripts/latlong.html>

Dr. Surames Piriawat. **Angle, Bearing, and Azimuth**. [Online]. Available :

[http://www.surames.com/images/column_1227454931/Unit%205-](http://www.surames.com/images/column_1227454931/Unit%205-Bearing%20and%20azimuth%20angles.pdf)

[Bearing%20and%20azimuth%20angles.pdf](http://www.surames.com/images/column_1227454931/Unit%205-Bearing%20and%20azimuth%20angles.pdf)

เอกสารนี้เป็นเอกสารที่สงวนลิขสิทธิ์ไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

scikit-learn. **Clustering**. [Online]. Available : <https://scikit-learn.org/stable/modules/clustering.html>

The PostgreSQL Global Development Group. **PostgreSQL**. [Online]. Available :
<https://www.postgresql.org/about/>

Spatial and Geographic objects for PostgreSQL. [Online]. Available : <https://postgis.net>

PgAdmin PostgreSQL Tools. [Online]. Available : <https://www.pgadmin.org>

PgAdmin PostgreSQL Tools. PostgreSQL. [Online]. Available : <https://www.pgadmin.org>

Global5 Thailand. 2549. **ความรู้ทั่วไปเกี่ยวกับ GPS**. [Online]. Available :
<https://www.global5thailand.com/thai/gps.htm>

PCP. 2562. **MAE, MSE, RMSE — เลือกใช้ยังไงดีมาลองดูที่ความหมาย**. [Online]. Available :
<https://medium.com/c-g-datacommunity/mse-rmse-mae-เลือกใช้ยังไงดีมาลองดูที่ความหมาย-17b37b0b14b3>

David Iseminger, olprod. 2564. **Power BI Desktop คืออะไร**. [Online]. Available :
<https://docs.microsoft.com/th-th/power-bi/fundamentals/desktop-what-is-desktop>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

ภาคผนวก ก

รูปแบบถนนที่พบในประเทศไทย

ก. 1 รูปแบบถนนที่พบประเทศไทย

จากที่ได้ทำการสำรวจแผนที่ของประเทศไทยมา พบรูปแบบถนนดังนี้

- 1) รูปแบบถนนที่ไม่มีโค้งน้อยกว่า 120 องศา



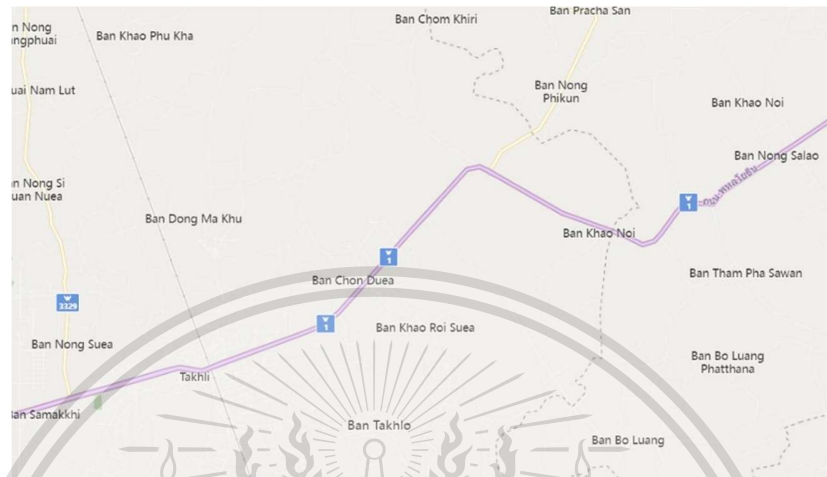
รูป ก.1 ส่วนของถนน 205 ที่อยู่ในพื้นที่จังหวัดนครราชสีมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

2) รูปแบบถนนที่มีทางโค้งน้อยกว่า 120 องศา



รูป ก.2 ส่วนของถนน 1 ที่อยู่ในพื้นที่จังหวัดนครสวรรค์

3) รูปแบบถนนที่มีแยกเป็นตัว Y



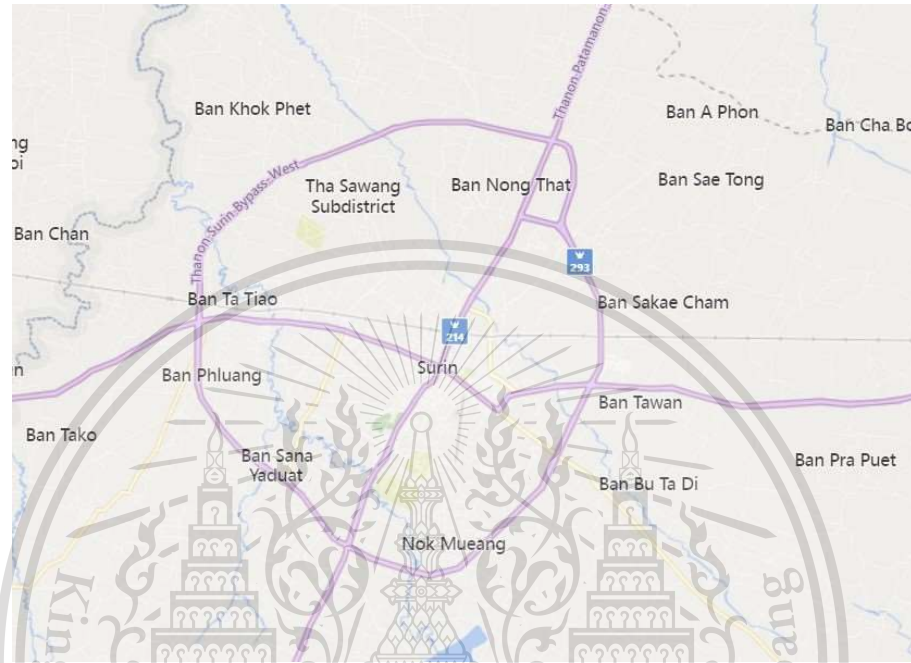
รูป ก.3 ส่วนของถนน 2, 207, 2160 ที่อยู่ในพื้นที่จังหวัดนครราชสีมา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

4) รูปแบบถนนที่เป็น 4 แยกและมีวงเวียนตัดผ่าน



รูป ก.4 ส่วนของถนน 214, 226, 293 ที่อยู่ในพื้นที่จังหวัดสุรินทร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

5) รูปแบบถนนบายพาส (Bypass) หรือถนนเลี่ยงเมือง



รูป ก.5 ส่วนของถนน 4, 37 ที่อยู่ในพื้นที่จังหวัดประจวบคีรีขันธ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

6) รูปแบบถนนที่เป็น 4 แยก



รูป ก.6 ส่วนของถนน 2, 202 ที่อยู่ในพื้นที่จังหวัดนครราชสีมา

7) รูปแบบถนนที่เป็น 3 แยก



รูป ก.7 ส่วนของถนน 208, 230 ที่อยู่ในพื้นที่จังหวัดขอนแก่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับใช้ในการเรียนการสอนและการวิจัยเท่านั้นที่จะนำเอกสารนี้ไปใช้ประโยชน์ด้านการค้า

ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

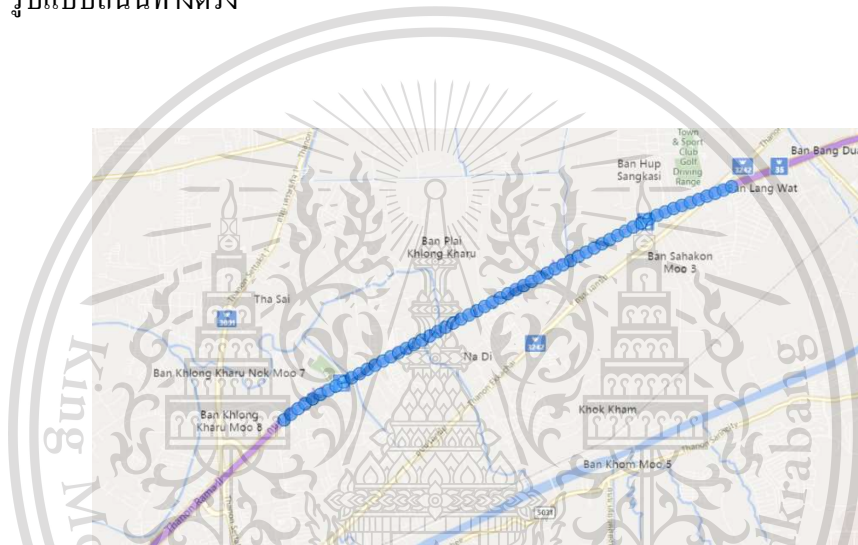
Forbidden to modify the content, and cite the document when use.

ภาคผนวก ข

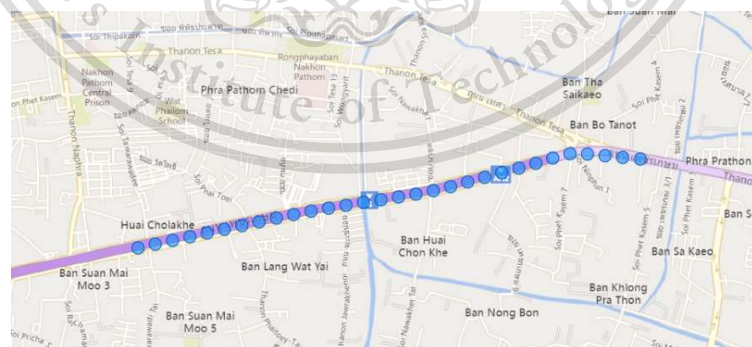
รูปแบบถนนที่ทดลองเพิ่มเติม

ข. 1 รูปแบบถนนที่ใช้ทดลองในการเติมข้อมูล GPS

- รูปแบบถนนทางตรง



รูป ข.1 ตัวอย่างเพิ่มเติมที่ 1 ของถนนรูปแบบทางตรง



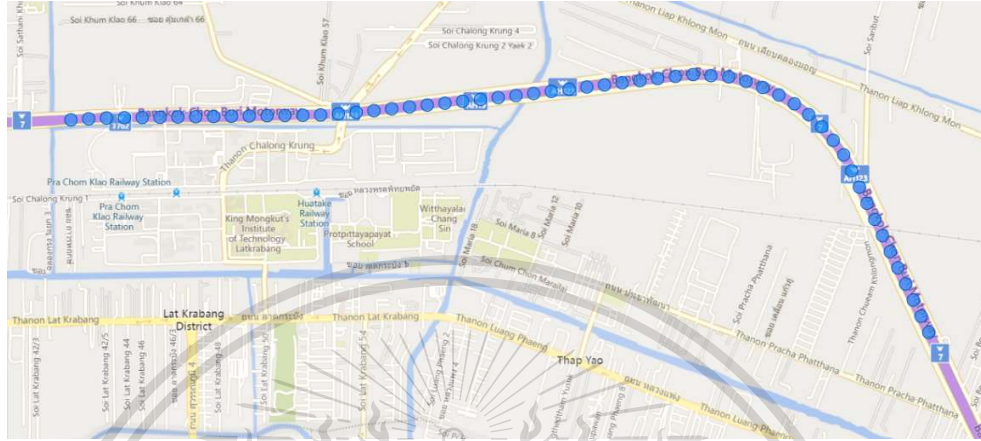
รูป ข.2 ตัวอย่างเพิ่มเติมที่ 2 ของถนนรูปแบบทางตรง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

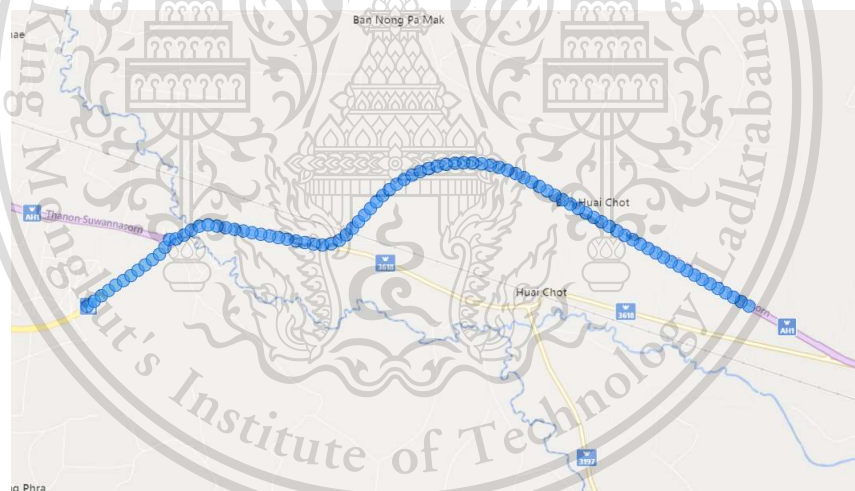
This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.

- รูปแบบถนนทางโค้ง



รูป ข.3 ตัวอย่างเพิ่มเติมที่ 1 ของถนนรูปแบบทางโค้ง



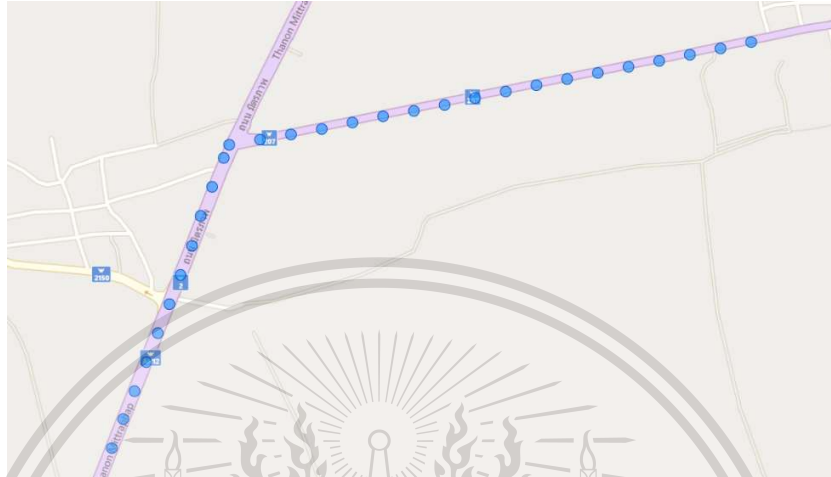
รูป ข.4 ตัวอย่างเพิ่มเติมที่ 2 ของถนนรูปแบบทางโค้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ตัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

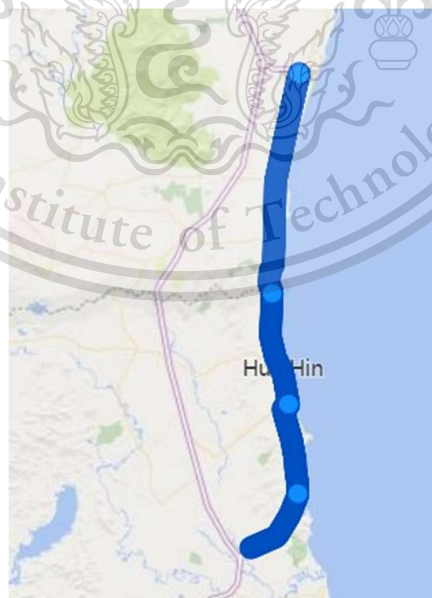
Forbidden to modify the content, and cite the document when use.

- รูปแบบถนนทางแยก



รูป ข.5 ตัวอย่างเพิ่มเติมที่ 1 ของถนนรูปแบบทางแยก

- รูปแบบถนนทางบายพาส



รูป ข.6 ตัวอย่างเพิ่มเติมที่ 1 ของถนนรูปแบบทางบายพาส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่นิยมนำไปใช้ประโยชน์ด้านการค้า
ไม่ว่ากรณีใดๆทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

This material is reserved for educational use only, not allowed for commercial use.

Forbidden to modify the content, and cite the document when use.