บทที่ 4

วิธีการวิจัย (Methodology)

4.1 ข้อมูลความเร็วรถ (Road Traffic Speed Data)

ข้อมูลความเร็วรถที่นำมาทำการคาดการณ์หาแนวโน้ม เป็นข้อมูลที่ได้มาจากการประมวลผลภาพวีดีโอ ที่ได้จากการกล้องวงจรปิด (CCTV) ของการทางพิเศษแห่งประเทศไทย

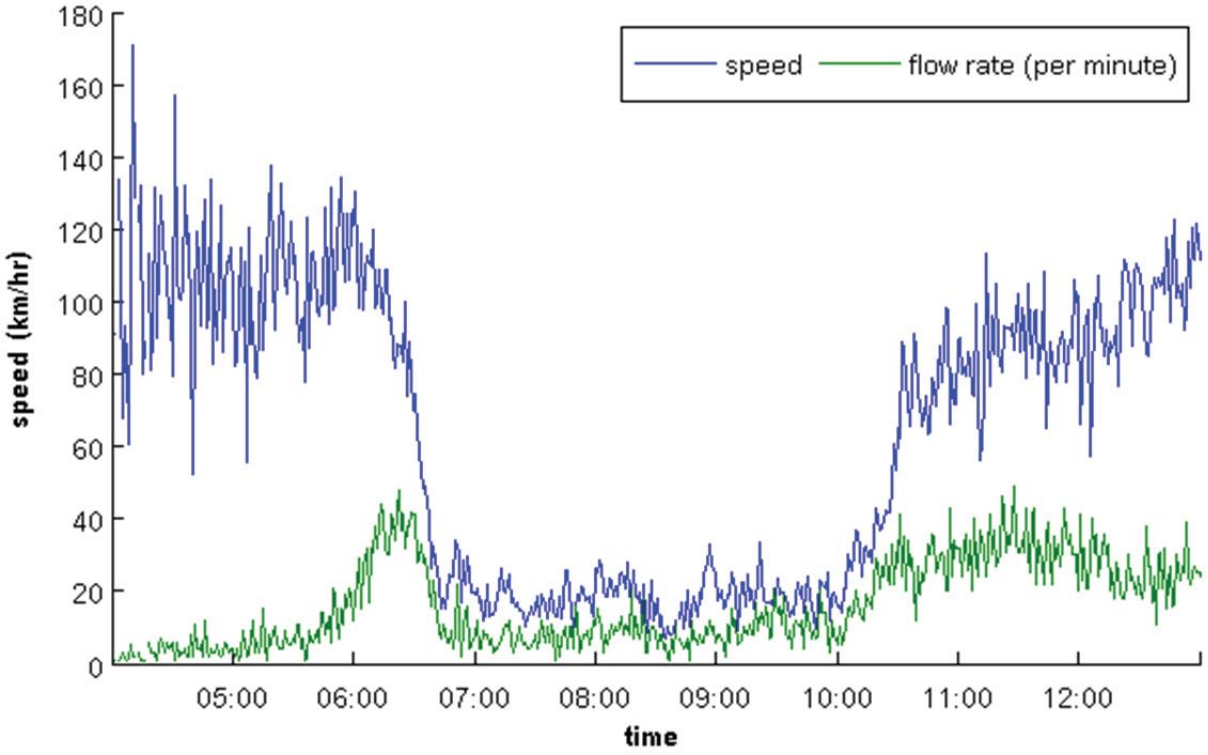
ภาพที่ 8

ตัวอย่างภาพที่ได้จากกล้องวงจรปิด (CCTV) ที่ติดตั้งบนทางด่วน

หลังจากการประมวลผลภาพจะได้ข้อมูล 2 ชนิด คือข้อมูลความเร็ว (speed) และข้อมูลปริมาณรถ (flow rate) ดังภาพที่ 9

ภาพที่ 9

กราฟแสดงข้อมูลความเร็วและข้อมูลปริมาณรถในหนึ่งช่วงเวลา



ข้อมูลความเร็ว (speed) คือ ข้อมูลที่บอกถึงความเร็วของรถที่แล่นผ่าน ในหน่วยกิโลเมตรต่อชั่วโมง ความเร็วของรถที่ได้จากการประมวลผลวีดีโอนี้จะมีค่าอยู่ระหว่าง 1 ถึง 180 กิโลเมตรต่อชั่วโมง

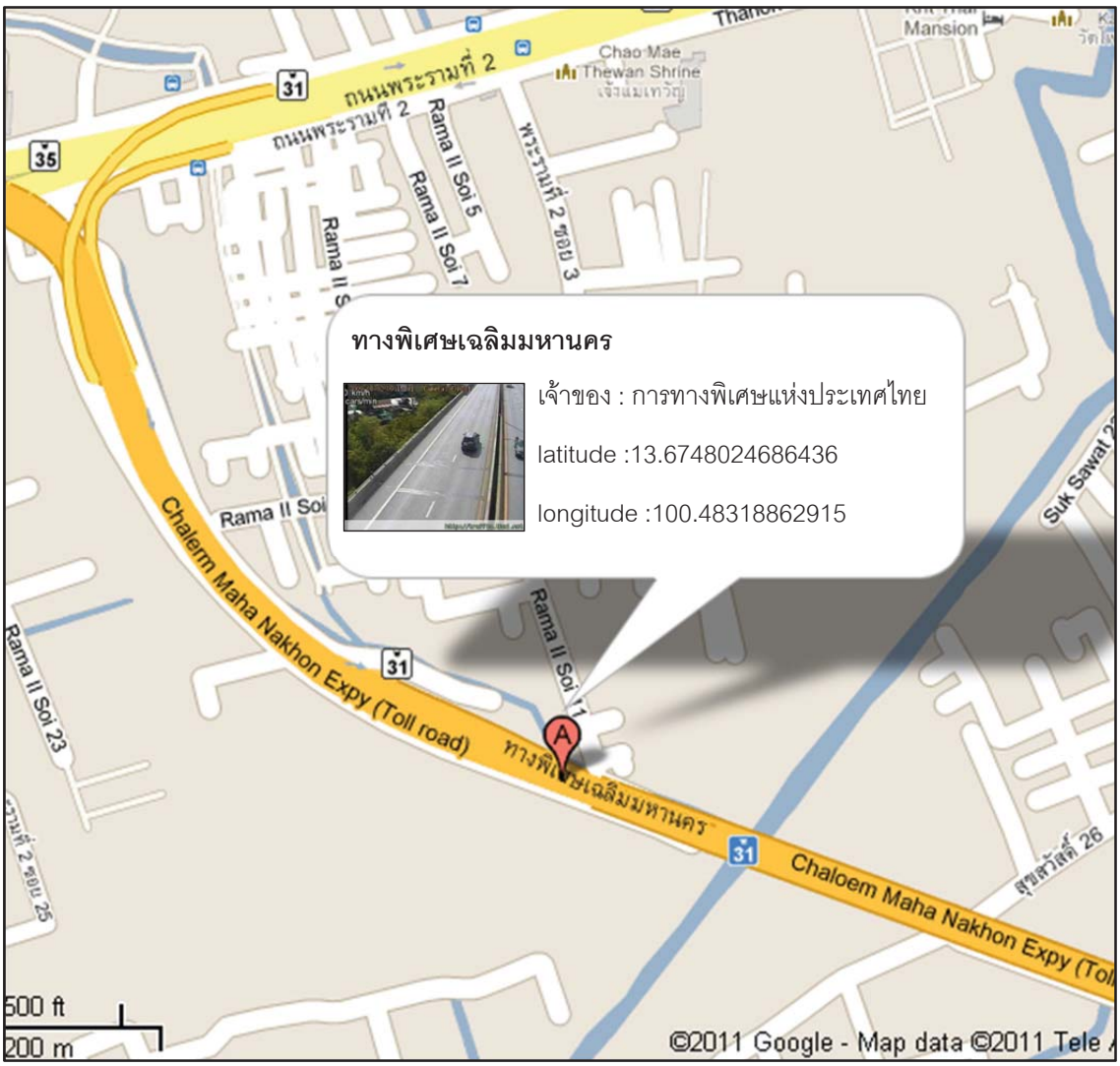
ข้อมูลปริมาณรถ (flow rate) คือ ข้อมูลที่บอกถึงจำนวนรถที่วิ่งผ่านในหนึ่งหน่วยเวลา มีหน่วยเป็นคันต่อนาที

จากภาพที่ 9 ช่วง 06:00 ถึง 07:00 จะเห็นได้ว่าปริมาณรถที่ไหลผ่านต่อนาทีมีจำนวนเพิ่มมากขึ้น และความเร็วของรถที่แล่นผ่านเริ่มมีความเร็วลดต่ำลง แสดงถึงการเริ่มเข้าสู่สภาพจราจรที่ติดขัด ช่วง 07:00 ถึง 10:00 จะเห็นได้ว่าความเร็วของรถที่แล่นผ่านจะไม่เกิน 40 km/hr และจำนวนรถที่ไหลผ่านมีปริมาณน้อย แสดงให้เห็นถึงสภาพการจราจรที่ติดขัด ช่วง 10:00 ถึง 10:30 ความเร็วของรถที่แล่นผ่านมีความเร็วเพิ่มขึ้น จำนวนของรถที่วิ่งผ่านก็มีมากขึ้น แสดงให้เห็นถึงสภาพการจราจรเริ่มคลี่คลายจากติดขัดเป็นปกติ

ข้อมูลความเร็วรถที่นำมาทำการทดลอง เป็นข้อมูลจากกล้อง CCTV ที่ติดตั้งอยู่บนทางพิเศษเฉลิมมหานคร ขาเข้ากรุงเทพมหานคร ในภาพที่ 10 ที่ได้เก็บไว้ตั้งแต่วันที่ 1 กรกฎาคม 2010 จนถึงวันที่ 30 กันยายน 2010 เป็นจำนวนทั้งหมด 92 วัน

ภาพที่ 10

สถานที่ติดตั้งกล้องที่นำมาทำการทดลอง



4.2 ขั้นตอนการทดลอง

ขั้นตอนในการดำเนินการทดลองมีทั้งหมด 4 ขั้นตอนหลัก คือ

1. การปรับให้เรียบ และแบ่งข้อมูลทำ Training data set และ Testing data set
2. การเลือกขนาดของข้อมูลที่จะใช้วัดความคล้าย และขนาดของข้อมูลที่จะใช้ในการคาดการณ์แนวโน้ม
3. การแปลงข้อมูลความเร็วเป็นข้อมูลลำดับ
4. ทดสอบคาดการณ์แนวโน้ม

4.2.1 การปรับให้เรียบ และแบ่งข้อมูลทำ Training data set และ Testing data set

ลักษณะของข้อมูลความเร็วที่ได้มาจะเห็นได้ว่ามีการแกว่งขึ้นลงของข้อมูล (noise) ซึ่งต้องทำการปรับให้เรียบ (smoothing) ก่อนนำข้อมูลเหล่านี้ไปใช้ในการทำการคาดการณ์ วิธีที่ใช้ในการปรับเรียบคือวิธีการหาค่าเฉลี่ยเคลื่อนที่ (moving average) ขนาดของการหาค่าเฉลี่ย (window) ที่ใช้ได้เลือกจากการทดลองทำการปรับเรียบแล้วเพิ่มขนาดของ window แล้วนำค่า Root Mean Square Error (RMS) ที่ได้จากการปรับเรียบข้อมูลความเร็วใน 1 วัน มาทำการเปรียบเทียบ ซึ่งได้ผลของค่า RMS ดังตารางที่ 1

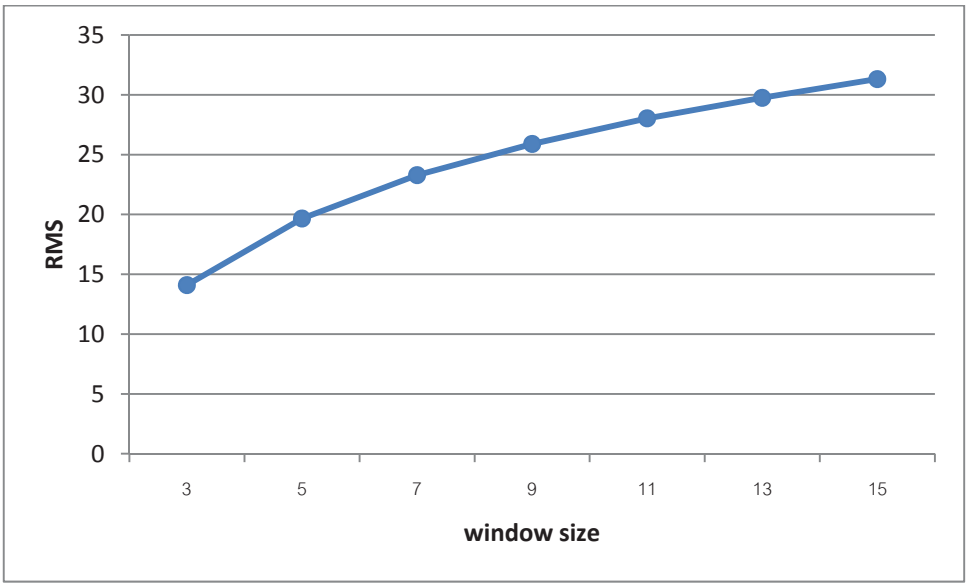
ตารางที่ 1

ค่า RMS ของการปรับให้เรียบด้วยขนาดของ window ที่ต่างกัน

series		RMS
raw speed data	smooth (w=3)	14.1035
raw speed data	smooth (w=5)	19.6614
raw speed data	smooth (w=7)	23.2828
raw speed data	smooth (w=9)	25.8914
raw speed data	smooth (w=11)	28.0362
raw speed data	smooth (w=13)	29.7545
raw speed data	smooth (w=15)	31.319

ภาพที่ 11

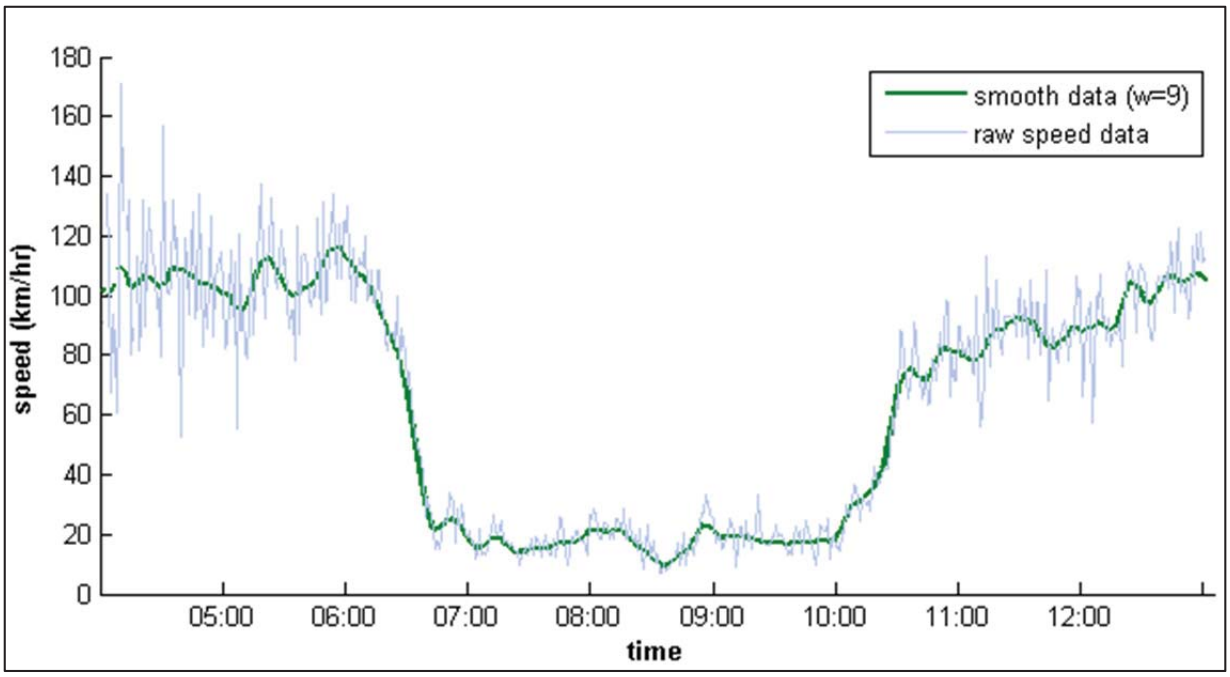
แสดงค่า RMS ของการปรับให้เรียบด้วยขนาดของ window ที่ต่างกัน



จากภาพที่ 11 จะเห็นได้ว่าค่า RMS จากการเทียบระหว่างก่อนและหลังการปรับให้เรียบจะมีแนวโน้มลู่เข้าสู่ค่าคงที่เมื่อขนาดของ window มีขนาดกว้างขึ้น โดยความแตกต่างของค่า RMS ที่ขนาดของ window เท่ากับ 7 และ 9 นาที เท่ากับ 2.6086 และความแตกต่างของค่า RMS ที่ขนาดของ window เท่ากับ 9 นาที และ11 นาที เท่ากับ 2.1448 จะเห็นได้ว่าค่า RMS นั้นเริ่มมีความแตกต่างที่คงที่ ดังนั้นจึงได้เลือกขนาดของ window เท่ากับ 9 นาที ซึ่งเป็นจุดที่ค่า RMS เริ่มมีความแตกต่างที่คงที่ และเป็นขนาดเวลาที่ไม่นานมากเกินไป ซึ่งผลของการปรับเรียบด้วย window เท่ากับ 9 นาที ให้ผลดังภาพที่ 12

ภาพที่ 12

กราฟแสดงข้อมูลหลังการปรับให้เรียบด้วยขนาด window เท่ากับ 9 นาที



หลังจากทำการปรับให้เรียบแล้ว ข้อมูลความเร็วรถที่ได้แสดงถึงแนวโน้ม (trend) และรูปแบบของการเปลี่ยนแปลงของข้อมูลความเร็วรถได้ชัดเจนกว่าเดิม

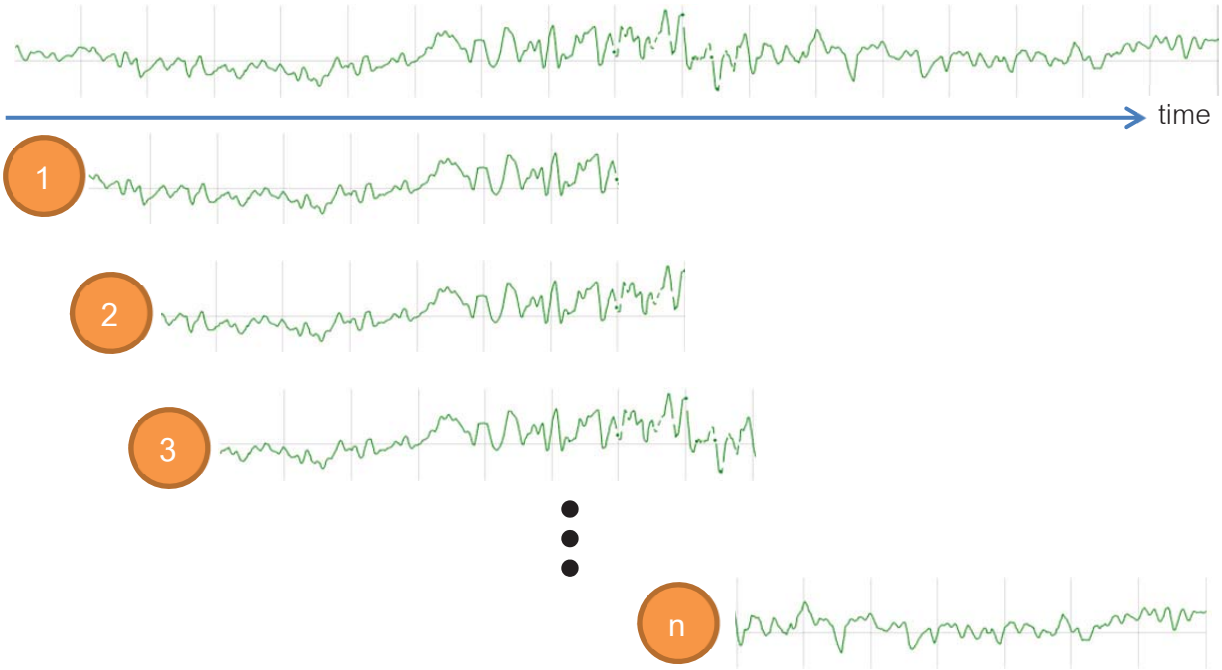
สำหรับการแบ่งข้อมูลเพื่อทำเป็น training data set และข้อมูลสำหรับ testing data set นั้น ได้ใช้วิธีการแบ่งข้อมูลที่มีออกเป็นอัตราส่วน 70 ต่อ 30 โดยเป็น training data set 70% และ testing data set 30% ซึ่งข้อมูลที่ถูกแบ่งนี้จะแบ่งตามลำดับเวลา ไม่มีการสลับที่ข้อมูลเพราะข้อมูลสภาพจราจรนั้นมีเวลาเข้ามาเกี่ยวข้อง โดยปกติแล้วข้อมูล testing data set นั้นจะเกิดขึ้นหลังข้อมูล training data set เสมอ

4.2.2 การเลือกขนาดของข้อมูลที่จะใช้วัดความคล้าย และขนาดของข้อมูลที่จะใช้ในการคาดการณ์แนวโน้ม

ในขั้นตอนนี้ข้อมูลความเร็วรถที่ได้ทำการปรับให้เรียบแล้วจะถูกนำไปทำการแบ่งข้อมูลออกเป็นชิ้นย่อยๆ แบบเลื่อนไหล (sliding) เพื่อเป็นการเตรียมข้อมูลสำหรับนำไปทำการวัดความคล้าย และการคาดการณ์ดังภาพที่ 13

ภาพที่ 13

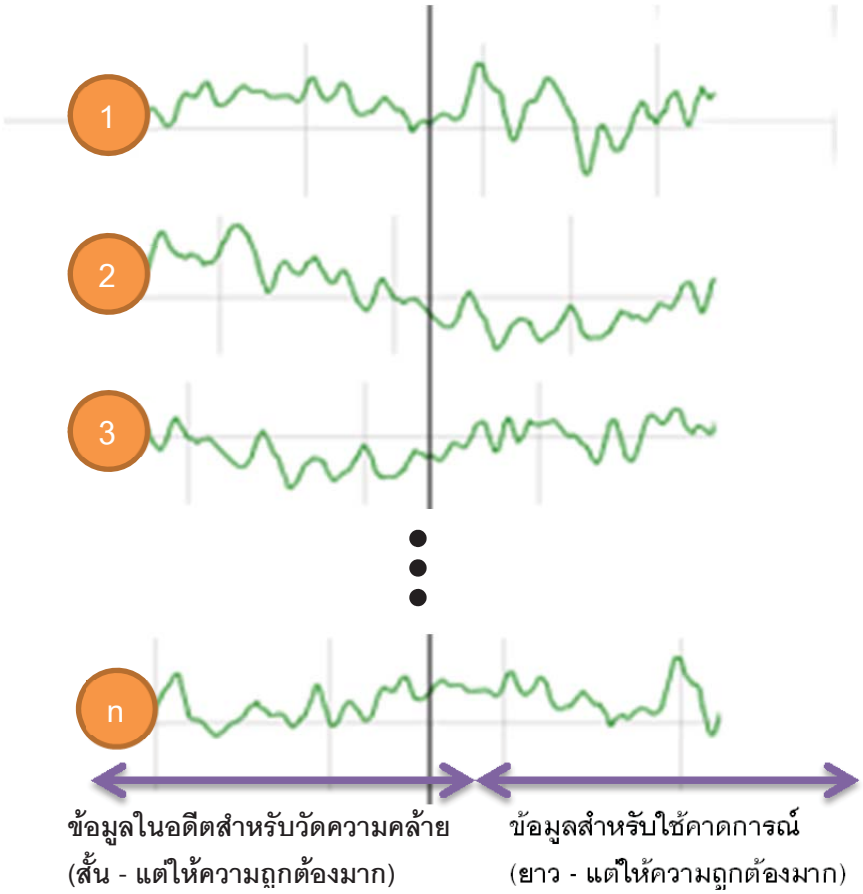
แบ่งข้อมูลออกเป็นชิ้นๆแบบเลื่อนไหล เพื่อเตรียมสำหรับการวัดความคล้ายและการคาดการณ์



สำหรับขนาดของข้อมูลที่ทำการตัดแบ่งเพื่อเตรียมข้อมูลสำหรับการวัดความคล้ายจะขึ้นอยู่กับสองส่วนคือ ส่วนของข้อมูลที่ใช้สำหรับการวัดความคล้าย และส่วนของข้อมูลที่ใช้ในการคาดการณ์ ดังภาพที่ 14 ซึ่งความยาวของทั้งสองส่วนนั้นส่งผลโดยตรงต่อผลของการคาดการณ์ โดยส่วนที่ใช้สำหรับการวัดความคล้ายนั้นจะต้องสั้นที่สุดเท่าที่จะสั้นได้ เนื่องจากข้อมูลนี้จะต้องนำไปผ่านกระบวนการวัดความคล้ายถ้าหากข้อมูลนี้มีความยาวมาก ก็จะทำให้สิ้นเปลืองเวลาที่ใช้วัดความคล้าย และจะต้องเป็นความยาวที่ส่งผลต่อผลของการคาดการณ์ โดยให้ความถูกต้องของการคาดการณ์มากที่สุดด้วย สำหรับส่วนที่เป็นข้อมูลที่ใช้สำหรับการคาดการณ์ จะต้องเป็นข้อมูลที่ยาวที่สุดเท่าที่จะนำมาใช้ในการคาดการณ์ได้ เนื่องจากเราต้องการการคาดการณ์ที่ให้ผลดีที่สุดและไกลที่สุดเท่าที่จะเป็นไปได้

ภาพที่ 14

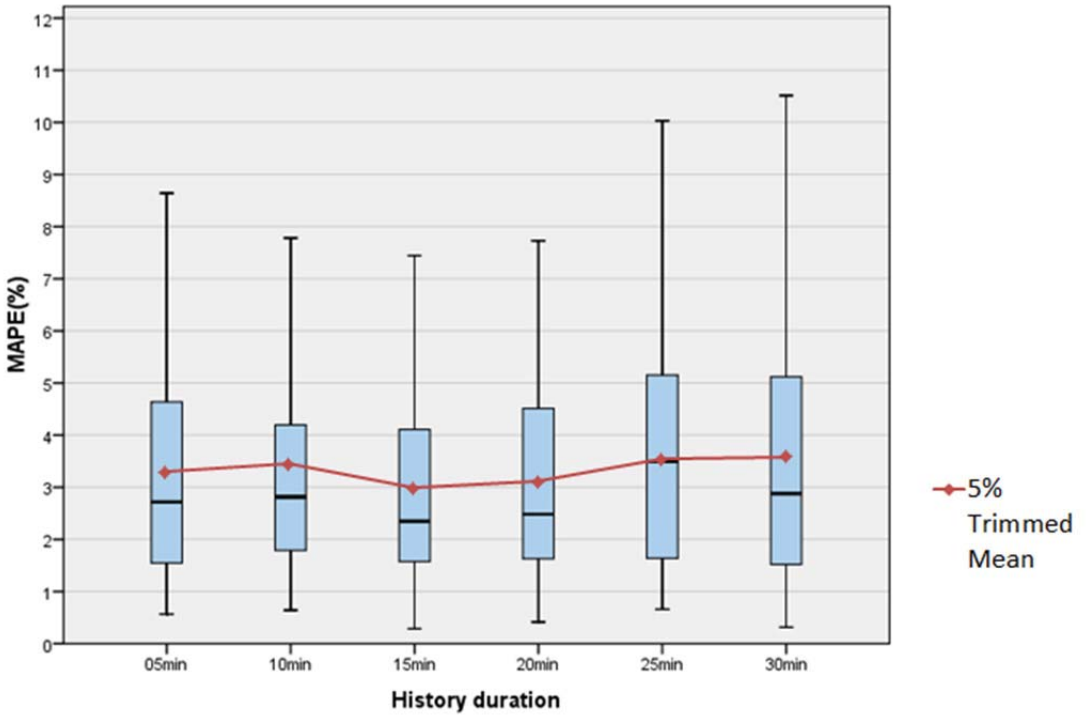
ขนาดของข้อมูลที่ทำการตัด



จากคำถามของขนาดข้อมูลที่ใช้สำหรับวัดความคล้ายและการคาดการณ์ที่เกิดขึ้น จึงได้ทำการทดลองเพื่อหาขนาดของข้อมูลที่เหมาะสมที่สุด ในขั้นแรกได้ทำการกำหนดขนาดของเวลาสำหรับการคาดการณ์ไว้ที่ 10 นาที เพื่อทดลองปรับระยะเวลาในส่วนของการวัดความคล้ายเพื่อดูว่าสามารถที่จะให้ผลการคาดการณ์ที่ดีที่สุดที่ระยะเวลาเท่าไร สาเหตุที่ทำการเลือกเวลาสำหรับคาดการณ์ในอนาคตเท่ากับ 10 นาที เพราะว่า เหตุการณ์ที่เกิดขึ้นบนถนนจะส่งผลต่อสภาพการจราจรในอนาคตไม่ต่ำกว่า 5 นาที ดังนั้น 10 นาทีจึงเป็นเวลาที่เหมาะสมไม่สั้นหรือนานจนเกินไปเหมาะสำหรับทดสอบการคาดการณ์ ในการทดลองนี้จะใช้วิธีการของ Dynamic Time Warping สำหรับการวัดความคล้ายข้อมูลความเร็วรถที่อยู่ในรูปแบบอนุกรมเวลา (time series) โดยผลการทดลองแสดงอยู่ในภาพที่ 15

ภาพที่ 15

Box plot แสดงค่า MAPE ที่ได้หลังจากคาดการณ์แนวโน้มของข้อมูลความเร็วรถใน 10 นาทีข้างหน้าโดยใช้ข้อมูลในอดีตที่มีความยาวแตกต่างกัน

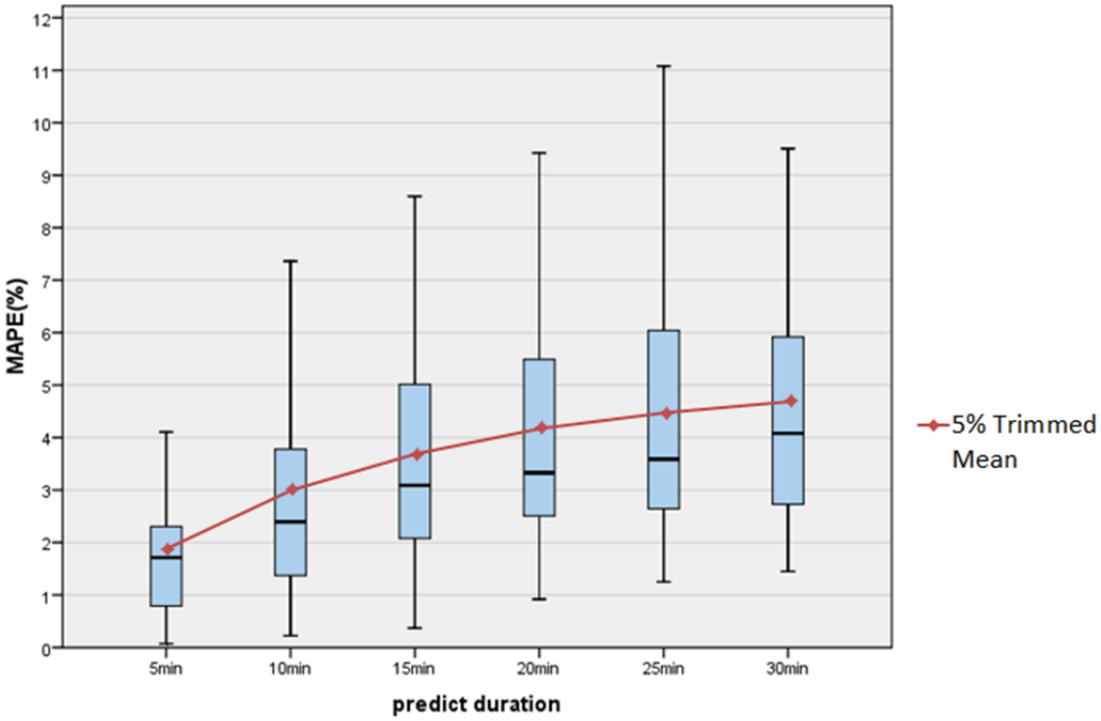


จากภาพที่ 15 จะเห็นได้ว่าความยาวของข้อมูลในอดีตที่ส่งผลต่อความถูกต้องของการคาดการณ์มากที่สุดอยู่ที่ 15 นาที (ค่า mean และ median ต่ำกว่าระยะเวลาอื่นๆ)

สำหรับข้อมูลที่ใช้สำหรับทำการคาดการณ์นั้น ได้ทำการทดลองเพื่อทดสอบว่า ถ้าเราใช้ข้อมูลความเร็วรถในอดีตเป็นเวลา 15 นาที จะสามารถให้ความถูกต้องในการคาดการณ์แนวโน้มในอนาคตเป็นอย่างไร ผลการทดลองแสดงอยู่ในภาพที่ 16

ภาพที่ 16

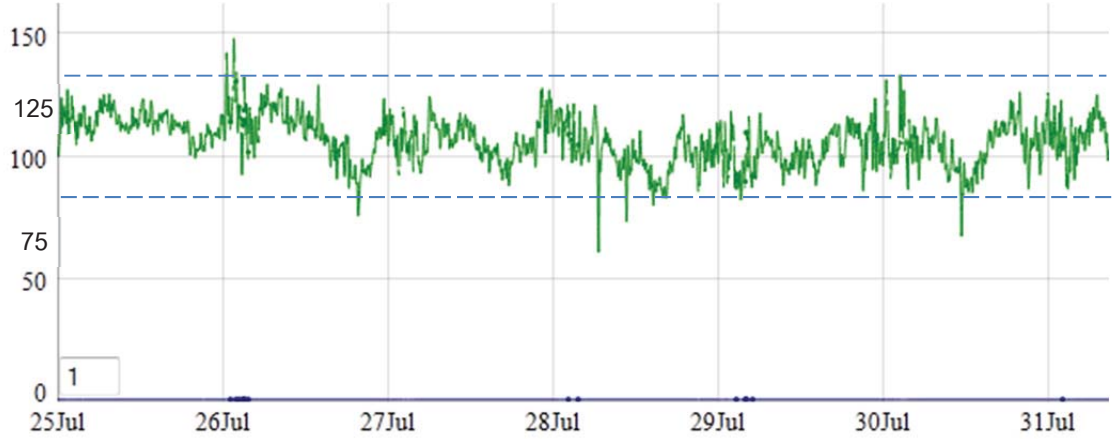
ค่า MAPE ที่ได้จากการคาดการณ์แนวโน้มข้อมูลในอนาคตที่เวลาต่างๆ



จากรูปจะเห็นได้ว่าค่า MAPE ที่ได้จากการทดลองคาดการณ์โดยใช้ข้อมูลในการวัดความคล้ายเป็นเวลา 15 นาทีนั้น เมื่อเพิ่มระยะเวลาในการคาดการณ์ขึ้นไปเรื่อยๆ จะทำให้เกิดค่า MAPE ที่สูงขึ้นเรื่อยๆ และจะเห็นว่าค่า MAPE ที่ได้นั้นลู่เข้าสู่ค่าคงที่เนื่องมาจาก ความเร็วของรถที่แล่นอยู่บนทางด่วนนั้นจะแตกต่างจากความเร็วรถที่แล่นอยู่บนทางปกติ จะสังเกตได้จากภาพที่ 17 ช่วงความเร็วของรถที่แล่นบนทางด่วนจะแคบกว่าช่วงความเร็วของรถที่แล่นบนทางปกติ

ภาพที่ 17

กราฟแสดงความเร็วรถตั้งแต่วันที่ 25 ถึง 31 มิถุนายน 2553



จากภาพที่ 17 จะเห็นได้ว่าข้อมูลความเร็วของรถที่แล่นบนทางด่วนตลอดทั้งวันนั้น ส่วนใหญ่จะวิ่งด้วยความเร็วไม่ต่ำกว่า 80 km/hr มีบางช่วงเวลาที่วิ่งต่ำกว่า 80 km/hr แต่เป็นช่วงเวลาที่สั้น และข้อมูลความเร็วรถส่วนใหญ่จะอยู่ระหว่าง 80 ถึง 120 อาจจะมีเร็วกว่า 120 km/hr บ้าง แต่เป็นส่วนน้อย สาเหตุที่ข้อมูลความเร็วรถบนทางด่วนเป็นแบบนี้เพราะว่า คนที่ใช้ทางด่วนมักจะเป็นคนที่ต้องการเดินทางให้ถึงที่หมายเร็วที่สุดโดยหลีกเลี่ยงรถที่ติดอยู่บนทางปกติ และบนทางด่วนจะไม่มีใครสามารถที่จะจอดรถเพื่อทำธุระตามข้างทางเหมือนทางปกติได้นอกจากรถเสีย คนส่วนใหญ่จึงขับรถด้วยความเร็วไม่ต่ำกว่า 80 km/hr และจะขับไม่เกิน 120 km/hr เนื่องมาจากคนขับจะคำนึงถึงความปลอดภัยในการเดินทาง รวมทั้งรถส่วนใหญ่นั้นมีอายุการใช้งานมานานด้วย

สำหรับวิทยานิพนธ์นี้จะเลือกเวลาที่ใช้ในการคาดการณ์ที่ 10 นาทีมาทำการทดลอง เนื่องจากทางพิเศษเฉลิมมหานคร มีจุดติดตั้งกล้องตลอดทางทั้งหมด 11 จุด ดังแสดงในภาพที่ 18 โดยมีระยะทางประมาณ 13 km จากภาพที่ 17 รถส่วนใหญ่ที่วิ่งบนทางเส้นนี้จะวิ่งด้วยความเร็วต่ำสุดประมาณ 80 km/hr ถ้าจะต้องเดินทางจากจุด A ไปยังจุด K ดังภาพที่ 18 จะต้องใช้เวลาในการเดินทางประมาณ 9.75 หรือ 10 นาทีเป็นอย่างน้อย

ภาพที่ 18

แสดงจุดติดตั้งกล้อง CCTV บนทางพิเศษเฉลิมมหานคร

4.2.3 การแปลงข้อมูลความเร็วเป็นข้อมูลลำดับ

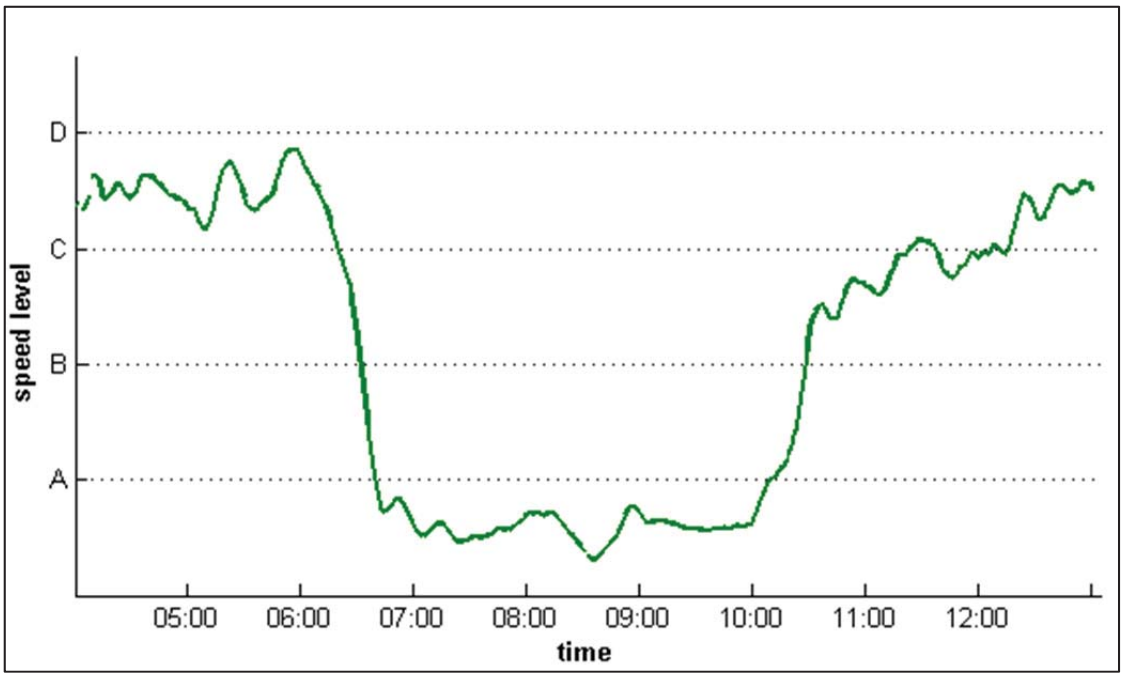
โดยปกติสามารถที่จะนำข้อมูลที่แบ่งออกเป็นชิ้นๆ ในขั้นตอนที่แล้ว ไปทำการเทียบความคล้าย โดยการใช้วิธีการของ Dynamic Time Warping เพื่อทำการคาดการณ์แนวโน้มของความเร็วที่จะเกิดขึ้นในอนาคตได้ แต่ปัญหาที่เกิดขึ้นคือ การใช้เวลาจำนวนมากในการคำนวณความคล้ายในแต่ละครั้ง

จากปัญหาการใช้เวลาจำนวนมากในการคำนวณความคล้าย วิทยานิพนธ์นี้จึงได้เสนอให้มีการแปลงรูปข้อมูลในส่วนที่ต้องใช้วัดความคล้าย จากข้อมูลที่อยู่ในรูปแบบอนุกรมเวลา (time series) เป็นข้อมูลแบบลำดับ (sequence) ตามกระบวนการที่เสนอในภาพที่ 7

การแปลงข้อมูลความเร็วรถที่ได้ทำการปรับให้เรียบแล้วให้เป็นข้อมูลลำดับ (quantization) ทำได้โดยการแบ่งความเร็วรถออกเป็นช่วงเท่าๆกัน แล้วแทนค่าช่วงความเร็วรถด้วยสัญลักษณ์ ทำให้เราสามารถที่จะแปลงข้อมูลความเร็วรถที่เป็นตัวเลขให้กลายเป็นข้อมูลอักขระที่แสดงถึงช่วงความเร็วรถ ดังภาพที่ 19

ภาพที่ 19

การใช้ตัวอักษรแทนช่วงความเร็ว โดยแบ่งขนาดแต่ละช่วงเท่าๆกัน



สำหรับขนาดของช่วงความเร็วรถที่ทำการแบ่งนั้นจะเป็นตัวแปรหนึ่งในการทดลอง เพราะช่วงความเร็วรถที่ต่างกันจะส่งผลต่อความถูกต้องของการคาดการณ์ และขนาดของข้อมูลลำดับที่ได้หลังจากการรวมข้อมูลที่ซ้ำเข้าด้วยกัน ถ้าหากทำการกำหนดช่วงที่แคบ ก็จะทำให้ลดขนาดของข้อมูลลำดับได้น้อย แต่ยังคงลักษณะของแบบข้อมูลได้เหมือนกับข้อมูลที่ยังไม่ได้ทำการแปลง ถ้าหากกำหนดช่วงของความเร็วรถที่กว้าง ก็จะสามารถที่จะลดขนาดของข้อมูลได้มาก แต่แบบข้อมูลลำดับระดับความเร็วที่เกิดขึ้นจะมีลักษณะที่เปลี่ยนแปลงไปจากข้อมูลก่อนทำการแปลง

ภาพที่ 20

แสดงตัวอย่างลำดับของข้อมูลความเร็วหลังจากการแปลง

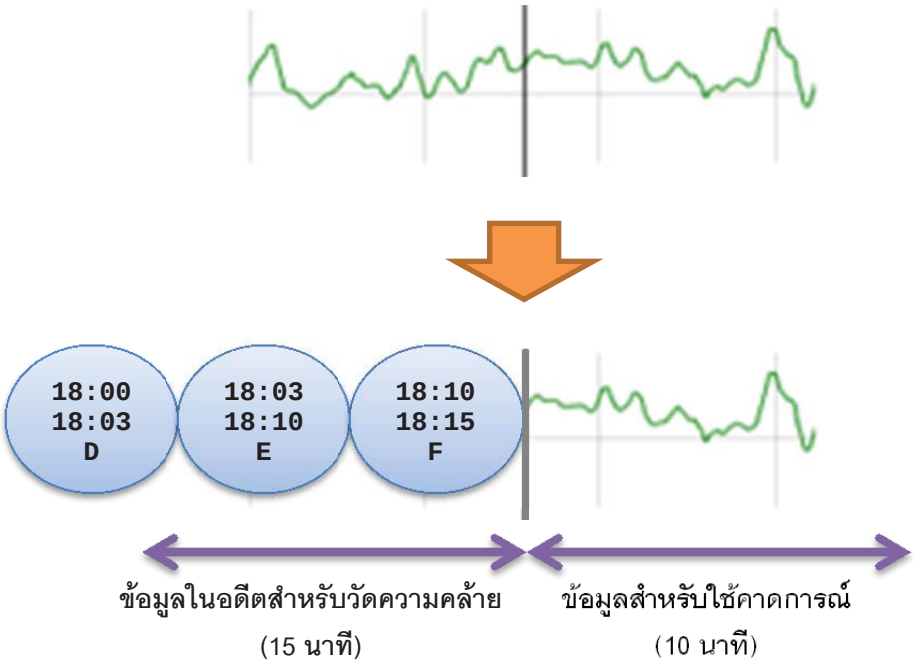
D D D D D C C C B B B A A A A A A A A A A B B B B C C

จากภาพที่ 20 ลำดับระดับความเร็วเหล่านี้สามารถที่จะทำการวัดความคล้ายได้จากวิธีการ edit distance ซึ่งเป็นวิธีการวัดความคล้ายของแบบข้อมูลลำดับได้ แต่เวลาที่ใช้ในการ

คำนวณจะเท่ากับวิธีการของ Dynamic Time Warping ซึ่งไม่ได้ช่วยลดเวลาที่ใช้ในการคำนวณ แต่ถ้าหากจัดข้อมูลนี้ให้อยู่ในรูปแบบของข้อมูลลำดับระดับความเร็วรถที่ได้นิยามไว้ก่อนหน้านี้ สามารถที่จะลดเวลาที่ใช้ในการคำนวณลงได้

ภาพที่ 21

แปลงข้อมูลความเร็วที่จะทำการวัดความคล้ายเป็นข้อมูลลำดับ แต่คงข้อมูลสำหรับคาดการณ์แนวโน้มไว้เหมือนเดิม

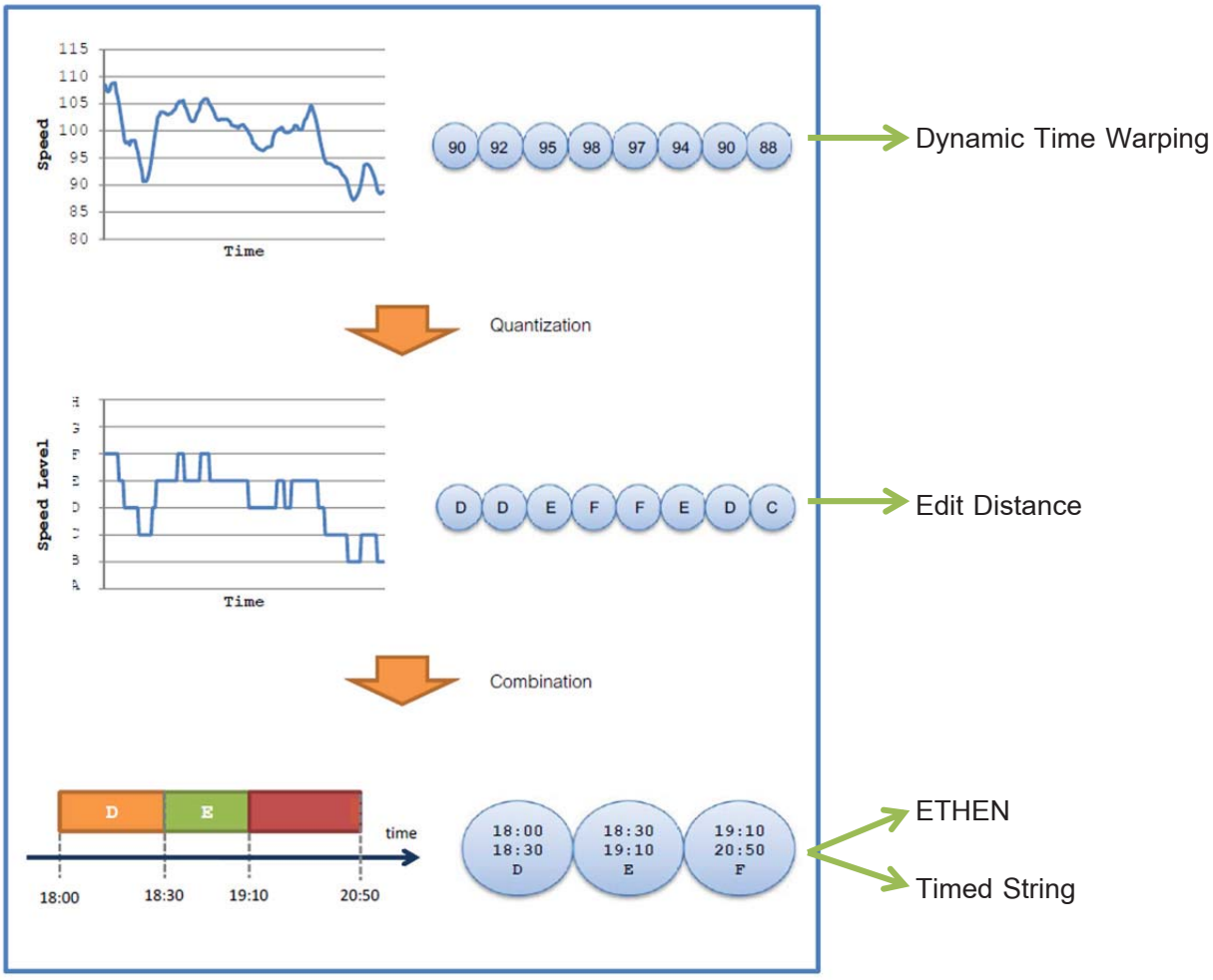


4.2.4 ทดสอบคาดการณ์แนวโน้ม

จากขั้นตอนที่แล้ว เราจะได้ training data set และ testing data set โดยในแต่ละวิธีที่ใช้ในการวัดความคล้ายนั้นจะใช้วัดข้อมูลที่ได้จากในแต่ละขั้นตอนที่ทำการแปลงข้อมูล เพื่อเป็นการหาว่าวิธีการแปลงข้อมูลไปแล้วนั้นส่งผลต่อความถูกต้องไปในทิศทางใด ดังแสดงใน ภาพที่ 22

ภาพที่ 22

แสดงชนิดของข้อมูลและวิธีการวัดความคล้ายที่รองรับ



สำหรับการทดลองในการคาดการณ์แนวโน้มจะประกอบด้วยขั้นตอนย่อยอีก 4 ขั้นตอนคือ

1. การหาความคล้ายระหว่าง testing data set แต่ละตัวกับ training data set ทั้งหมด ในขั้นตอนนี้เป็นขั้นตอนสำหรับการค้นหาข้อมูล training data set ที่มีลักษณะเหมือนกับข้อมูลที่ใช้ test มากที่สุด และขั้นตอนนี้จะทำการวัดความคล้ายแบบจับคู่กับข้อมูล training data set ทั้งหมด เวลาที่ต้องใช้ในขั้นตอนนี้จึงมากกว่าขั้นตอนอื่นๆ

2. เรียงลำดับผลการวัดความคล้าย หลังจากทำการวัดความคล้ายกับข้อมูลที่อยู่ใน training data set ทุกตัวแล้วก็จะทำการเรียงลำดับข้อมูล เพื่อหาข้อมูลที่มีความคล้ายกับข้อมูลที่ใช้ Test มากที่สุด
3. ทำการคาดการณ์แนวโน้ม (trend) โดยใช้วิธีการ Dynamic Time Warping, Edit Distance, Timed String และ วิธีการวัดความคล้ายของลำดับความเร็วที่เสนอ (ETHEN) นำข้อมูลในอนาคตของข้อมูลที่ได้จากการเรียงลำดับผลการวัดความคล้ายมา 3 ตัว แล้วทำการเฉลี่ยข้อมูลความเร็ว
4. นำแนวโน้ม (trend) ที่ได้จากการคาดการณ์ไปทำการหาค่าความผิดพลาด โดยจะใช้วิธีการวัดค่าความผิดพลาดของอนุกรมเวลา (time series) ที่นิยมใช้กันคือ Mean Absolute Percentage Error (MAPE) ซึ่งจะให้ค่าความ error ออกมาเป็น % ของความผิดพลาด