บทที่ 3

วิธีการดำเนินการวิจัย

ในบทนี้ได้นำเสนอวิธีการดำเนินการวิจัยซึ่งจะประกอบไปด้วย ขอบเขตการทดลอง เครื่องมือและซอฟแวร์ที่ใช้สำหรับการทดลอง การออกแบบการทดลองและวิธีการดำเนินการทดลองในแต่ละขั้นตอนตามลำดับ เพื่อใช้สำหรับเป็นแนวทางในการทำวิจัย และนำเสนอกระบวนการในการพัฒนาวิธีการประมาณเวลาในการเดินทางจากข้อมูลป้ายจราจรอัจฉริยะ

3.1 ขอบเขตการทดลอง

ประกอบไปด้วย เส้นทางที่ใช้สำหรับการทดลองและเก็บข้อมูล ประเภทของยานพาหนะ ชนิดข้อมูลที่เก็บ วันและเวลาที่เก็บข้อมูล

เส้นทางที่ใช้ในการทดลองและเก็บข้อมูล ผู้วิจัยเลือกใช้เส้นทาง ดังนี้

- ถนนพญาไท จากถนนรางน้ำ ถึง แยกพญาไท Link ID 108 ระยะทาง 0.739 กิโลเมตร
- ถนนพญาไท จากแยกพญาไท ถึง แยกราชเทวี Link ID 102 ระยะทาง 0.543 กิโลเมตร
- ถนนพญาไท จากแยกราชเทวี ถึง แยกปทุมวัน Link ID 2209 ระยะทาง 0.673 กิโลเมตร
- ถนนสีสม จากหลังป้าย26(สีลม) ถึง แยกรารมย์ Link ID 2617 ระยะทาง 0.959 กิโลเมตร
- ถนนสีสม จากแยกรารมย์ ถึง แยกสุรศักดิ์ Link ID 2615 ระยะทาง 0.995 กิโลเมตร
- ถนนสีสม จากแยกสุรศักดิ์ ถึง แยกบางรัก Link ID 2609 ระยะทาง 0.283 กิโลเมตร
- ถนนพหลโยธิน จากหลังป้าย(ลาดพร้าว2)ถึง แยกลาดพร้าว Link ID 410 ระยะทาง 0.533 กิโลเมตร
- ถนนพหลโยธิน จากแยกลาดพร้าว ถึง ถนนวิภาวดี Link ID 215 ระยะทาง 0.221 กิโลเมตร
- ถนนพหลโยธิน จากถนนวิภาวดี ถึง แยกสะพานควาย Link ID 603 ระยะทาง 2.984 กิโลเมตร

- ถนนพหลโยธิน จากเซ็นทรัลลาดพร้าว ถึง แยกรัชโยธิน Link ID 608 ระยะทาง 1.192 กิโลเมตร
- ถนนพหลโยธิน จากแยกรัชโยธิน ถึง แยกเกษตร Link ID 608 ระยะทาง 1.532 กิโลเมตร

โดยที่ Link ID เป็นการกำหนดจาก ฐานข้อมูลของศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ห้องปฏิบัติการระบบขนส่งและจราจรอัจฉริยะ (ITS)

สำหรับเกณฑ์ในการเลือกเส้นทาง ที่ใช้ทำการเก็บข้อมูลเวลาในการเดินทางคือ จะต้องเป็นเส้นทางที่มีข้อมูลระดับความติดขัดจากป้ายจราจรอัจฉริยะ และเป็นเส้นทางที่มีการเปลี่ยนแปลงของระดับความติดขัด ซึ่งเส้นทางที่เลือกจะเป็นเส้นทางที่มีความสำคัญ เป็นที่ตั้งของแหล่งธุรกิจการค้าขนาดใหญ่หรือเป็นเส้นทางที่เชื่อมต่อเพื่อนำไปสู่องค์กรต่างๆ เช่น โรงพยาบาล มหาวิทยาลัย สนามกีฬา เป็นต้น ทั้ง 4 เส้นทาง 11 Link ทำการเก็บข้อมูลทั้งหมดจำนวน Link ละ 35 ครั้ง

ยานพาหนะ ในการทดลอง ใช้ยานพาหนะประเภท รถยนต์ส่วนบุคคล

ข้อมูลที่เก็บ แบ่งออกเป็น 3 ชนิด คือ

1) ข้อมูลพิกัดของยานพาหนะ และเวลา ตามมาตรฐาน NMEA 0183 Version 2.3 ซึ่งจะใช้อุปกรณ์ GPS เชื่อมต่อกับคอมพิวเตอร์โน้ตบุ๊ก และโทรศัพท์มือถือ ด้วยสัญญาณบลูทูธ

2) เก็บข้อมูลเวลาในการเดินทางผ่านจุดที่กำหนดเป็นจุดเริ่มต้น ถึงจุดสิ้นสุดของเส้นทาง แต่ละ Link ID

3) ระดับความติดขัดจากป้ายจราจรอัจฉริยะ ณ เวลาเดินทาง ซึ่งข้อมูลระดับความติดขัดจะได้จากการตั้งกล้องดิจิตอลถ่ายภาพป้ายจราจรอัจฉริยะ และข้อมูลระดับความติดขัดจากฐานข้อมูลของศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ห้องปฏิบัติการระบบขนส่งและจราจรอัจฉริยะ (ITS)

วันและเวลาที่เก็บข้อมูล สำหรับการทดลอง วันและเวลาที่เก็บข้อมูล คือ

วันจันทร์ - วันศุกร์ ตั้งแต่เวลา 06.30 -20.30 น.

วันเสาร์ - วันอาทิตย์ ตั้งแต่เวลา 10.00 - 20.00 น.

เหตุผลที่เวลาในการเก็บข้อมูลในวันเสาร์ และวันอาทิตย์จะเริ่มต้นตั้งแต่เวลา 10.00 น. เพราะว่าในช่วงเวลาตอนเช้าของทั้งสองวัน ระดับความติดขัดจะมีการเปลี่ยนแปลงน้อยมาก ถ้าออกไปเก็บก็จะได้แต่ระดับความติดขัดสีเขียว ผู้วิจัยจึงเลือกใช้เวลาการเก็บข้อมูลเริ่มต้นตั้งแต่เวลาดังกล่าว

3.2 เครื่องมือและซอฟแวร์ที่ใช้

3.2.1 โทรศัพท์เคลื่อนที่ ในการทดลองใช้โทรศัพท์เคลื่อนที่ ที่เป็นระบบปฏิบัติการ Symbian จำนวน 3 เครื่อง

ภาพที่ 3.1
แสดงภาพโทรศัพท์เคลื่อนที่ระบบปฏิบัติการ Symbian

3.2.2 อุปกรณ์นำร่องสัญญาณดาวเทียม หรืออุปกรณ์ GPS ที่สามารถเชื่อมต่อสัญญาณบลูทูธ (Bluetooth GPS) จำนวน 4 เครื่อง

ภาพที่ 3.2
แสดงภาพอุปกรณ์ GPS ประเภทเชื่อมต่อสัญญาณบลูทูธ

3.2.3 นาฬิกา เป็นนาฬิกาสำหรับดูเวลา จำนวน 1 เครื่อง

3.2.4 กล้องดิจิตอลสำหรับบันทึกข้อมูล ระดับความติดขัดจากป้ายจราจรอัจฉริยะ

3.2.5 คอมพิวเตอร์ที่ใช้สำหรับเก็บข้อมูล เป็นคอมพิวเตอร์โน้ตบุ๊ก ระบบปฏิบัติการ Windows XP Home

ซอฟแวร์ที่ใช้สำหรับการวิจัย แบ่งออกเป็น 2 ประเภท

1) ซอฟแวร์สำหรับโทรศัพท์เคลื่อนที่ Cellular Probe ใช้สำหรับเก็บข้อมูลอุปกรณ์ GPS ที่เชื่อมต่อกับโทรศัพท์เคลื่อนที่

2)ซอฟแวร์สำหรับคอมพิวเตอร์มี 4 โปรแกรม

- โปรแกรม Weka เวอร์ชัน 3.6.0 สำหรับประมวลผลทางสถิติ
- โปรแกรม PostgreSQL 8.3 ติดต่อฐานข้อมูล
- โปรแกรม Quantum GIS แสดงผลข้อมูล GPS

- โปรแกรม GPS Diagnostics ใช้สำหรับเก็บข้อมูลจาก อุปกรณ์ GPS ที่เชื่อมต่อโน้ตบุ๊ก

3.3 การออกแบบการทดลอง

จากสมมติฐานงานวิจัยคือ ป้ายจราจรอัจฉริยะ สามารถใช้ในการประมาณเวลาในการเดินทางได้ ดังนั้นผู้วิจัยจึงออกแบบการทดลอง โดยทำการเก็บข้อมูลเวลาการเดินทางจริง โดยเก็บข้อมูลบนเส้นทางที่มีการใช้งานในชีวิตประจำวัน โดยใช้รถยนต์ส่วนตัว ในช่วงเวลาต่าง ๆ ของวัน ทำการเก็บข้อมูลเวลาการเดินทางจากอุปกรณ์ GPS และเก็บข้อมูลจากลักษณะการทำงานของป้ายจราจรอัจฉริยะคือระดับความติดขัด ซึ่งข้อมูลระดับความติดขัดจะได้จาก 1)ผู้วิจัยทำการตั้งกล้องดิจิตอลถ่ายภาพป้ายจราจรอัจฉริยะ 2)ข้อมูลระดับความติดขัดจากฐานข้อมูลของศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ห้องปฏิบัติการระบบขนส่งและจราจรอัจฉริยะ (ITS) ซึ่งข้อมูลระดับความติดขัดจะตรงกับเวลาการเดินทางจริง หลังจากนั้นจึงนำข้อมูลทั้งสองคือ เวลาการเดินทางและระดับความติดขัด เข้าสู่เทคนิคที่หนึ่งคือ กระบวนการวิเคราะห์ความถดถอยเชิงเส้น และเทคนิคที่สองคือข่ายงานประสาทเทียม เพื่อหาวิธีการหรือรูปแบบการประมาณเวลาในการเดินทางจากทั้งสองเทคนิค แล้วนำผลลัพธ์ มาเปรียบเทียบหาประสิทธิภาพที่ดีที่สุดของแต่ละเทคนิค แต่ด้วยในทางปฏิบัติ มีข้อจำกัดของปริมาณชุดข้อมูลเวลาการเดินทาง ผู้วิจัยไม่สามารถเก็บข้อมูลเวลาการเดินทางได้มากเพราะข้อจำกัดจากหลายด้าน จึงเก็บได้แค่เพียง 35 ครั้งในทุก ๆ เส้นทาง ซึ่งจำนวนชุดข้อมูลขนาดนี้ ผู้วิจัยจึงเลือกใช้เทคนิครูปแบบสำหรับทั้ง 2 เทคนิคคือ Leave one out Cross validation (Mitchell, 1997)

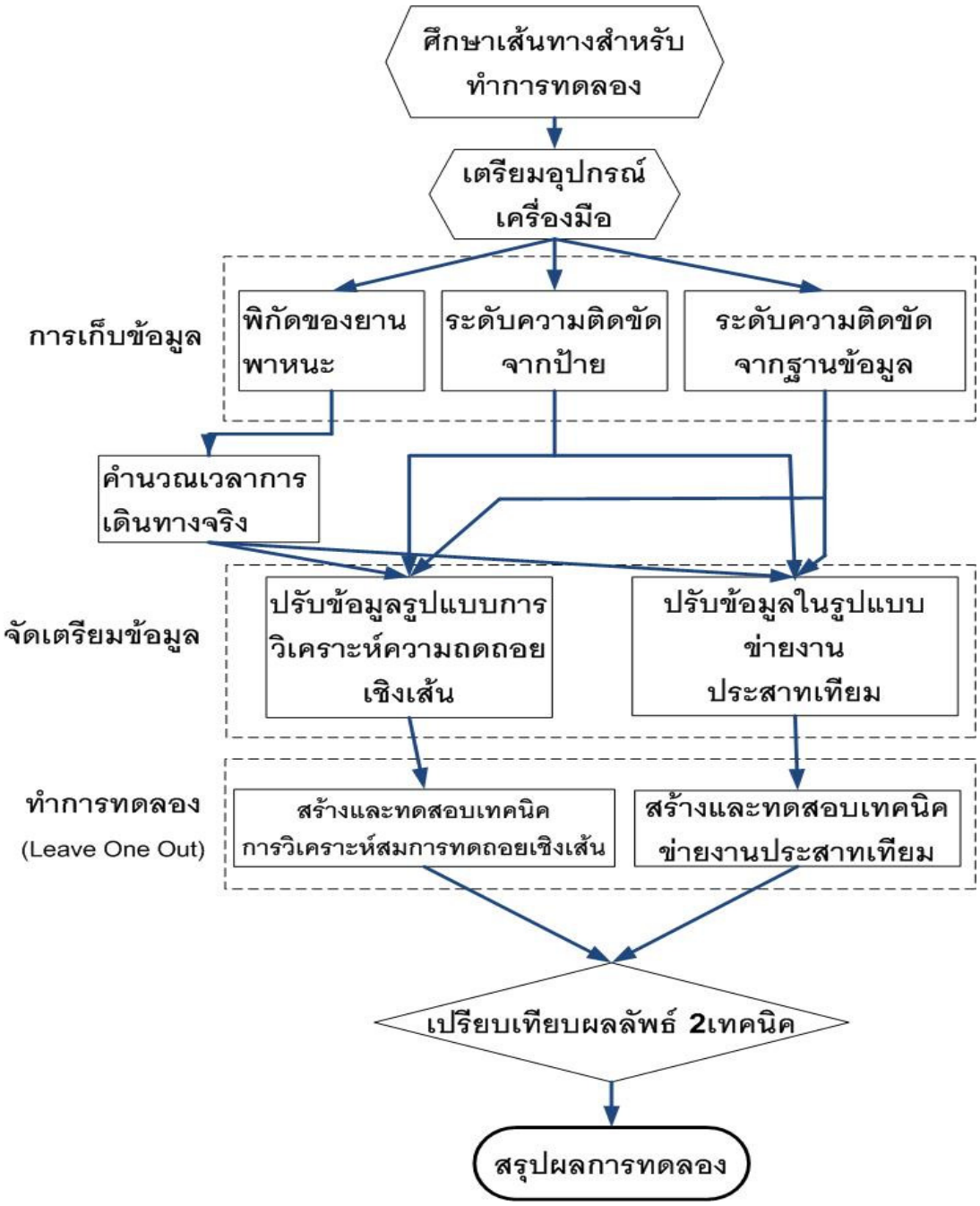
3.4 โครงสร้างวิธีทำการทดลอง

เริ่มต้นจากเก็บข้อมูลเวลาการเดินทาง พร้อมกับระดับความติดขัด ณ เวลาการเดินทางซึ่งแหล่งที่มาได้อธิบายในหัวข้อ 3.3 โดยระดับความติดขัดจากป้ายจราจรอัจฉริยะกับระดับความติดขัดของฐานข้อมูลจาก NECTEC นั้น มีความแตกต่างกัน คือ ระดับความติดขัดจากป้ายจราจรอัจฉริยะ มีความถี่ในการ Update เร็วกว่า ขณะที่ระดับความติดขัดจากฐานข้อมูลของ NECTEC จะมีความถี่ในการอัพเดทเฉลี่ย 5 นาที เนื่องจากต้นทางของข้อมูลระดับความติดขัด จะมาจากบริษัท Genius Joint Venture ซึ่งเป็นบริษัทลูกของ บริษัท Forth Corp Public ส่งข้อมูลมายังทางกรุงเทพมหานคร แล้วจึงส่งต่อมายัง NECTEC อีกที และทางกรุงเทพมหานครได้ให้เหตุผลว่าทำไมถึงต้อง Update ทุก 5 นาที เนื่องจากเครือข่ายอินเทอร์เน็ตของ บริษัท Genius ยังไม่ดีพอ (อยู่ในขั้นทดลอง) ถ้าเครือข่าอินเตอร์เน็ตดีกว่านี้ ก็จะสามารถอัพเดตได้ถี่กว่าปัจจุบันได้

แต่ข้อมูลระดับความติดขัดจากฐานข้อมูลของ NECTEC จะสามารถทราบถึงเวลาของสถานะระดับความติดขัด ว่าเป็นสถานะนี้มานานเท่าไหร่ ซึ่งข้อมูลระดับความติดขัดจากทั้งสองแหล่งข้อมูล มีข้อดีข้อเสียต่างกัน ผู้วิจัยจึงเลือกใช้ข้อมูลทั้งสองชุดในการทดลองแล้วเลือกเอาผลลัพธ์ที่ดีที่สุด หลังจากที่ได้ชุดข้อมูลครบ จึงเข้าสู่กระบวนการจัดเตรียมข้อมูลให้อยู่ในรูปแบบของทั้ง 2 เทคนิค ต่อจากนั้นจึงนำผลลัพธ์ที่ได้มาเปรียบเทียบ และในขั้นตอนสุดท้ายคือทำการสรุปผลการทดลอง แสดงดังภาพที่ 3.3

ภาพที่ 3.3

แสดงโครงสร้างวิธีทำการทดลอง



3.5 วิธีดำเนินการทดลอง

ขั้นตอนที่ 1 ศึกษาเส้นทาง สำหรับทำการทดลองและเก็บข้อมูล โดยจุดเริ่มต้นและจุดสิ้นสุดของเส้นทางได้ถูกกำหนดจากฐานข้อมูล ต่อจากนั้นทำการบันทึกข้อมูลพิกัดซึ่งจะได้ตำแหน่งเริ่มต้น และสิ้นสุดของถนน 2 ตำแหน่ง ของถนนหนึ่ง Link เพื่อจะเป็นเส้นอ้างอิงสำหรับไว้ใช้หาค่าเวลาในการเดินทางที่แท้จริง ซึ่งในทางปฏิบัติ จุดเริ่มต้นและจุดสิ้นสุดถูกกำหนดไว้แล้ว ซึ่งดูได้จากฐานข้อมูลของ NECTEC

ขั้นตอนที่ 2 จัดเตรียมอุปกรณ์เครื่องมือพร้อมติดตั้งซอฟแวร์ให้พร้อมสมบูรณ์ หลังจากนั้นปรับเวลาคอมพิวเตอร์ และอุปกรณ์โทรศัพท์เคลื่อนที่ ที่เชื่อมต่อกับอุปกรณ์ GPS และนาฬิกาให้ตรงกัน

ขั้นตอนที่ 3 ทำการเก็บข้อมูลทั้งหมด 11 Link ตลอดทั้งวัน โดยสำหรับอุปกรณ์ GPS ที่ได้เชื่อมต่อกับคอมพิวเตอร์โน้ตบุ๊กจะเก็บข้อมูลผ่านโปรแกรม GPS Diagnostics และโทรศัพท์เคลื่อนที่ทั้ง 2 ตัว จะเก็บข้อมูลผ่านโปรแกรม Cellular Probe ซึ่งข้อมูลที่เก็บเป็นข้อมูลของตำแหน่งยานพาหนะที่เคลื่อนที่ผ่านเส้นทางในการทดลอง ซึ่งจะอยู่ในรูปของประโยค GPRMC ตามมาตรฐาน NMEA 0813 แสดงตัวอย่างดังภาพที่ 3.4 ทั้งนี้เมื่อรถยนต์เคลื่อนที่ผ่านจุดเริ่มต้นและจุดสิ้นสุดของเส้นทาง โดยจะทำการบันทึกค่าเวลาจากนาฬิกาด้วย

จำนวนชุดข้อมูลที่ใช้ในการทดลอง คือ 35 ชุด ทั้งหมด 11 Link ซึ่งการจำนวนชุดข้อมูลที่เหมาะสมทางสถิติจะคำนวณได้จากสมการ

$$n = \left(2z_{\alpha/2}\frac{\sigma}{w}\right)^2 \qquad (3.1)$$

แต่เนื่องจากข้อจำกัดทางด้านเวลา และค่าใช้จ่ายในการเก็บข้อมูลผู้วิจัยจึงต้องหาจุดที่เหมาะสมที่สุดของจำนวนข้อมูลคือเท่ากับ 35 ชุดข้อมูล

ภาพที่ 3.4

ความหมายของประโยค GPRMCตามมาตรฐาน NMEA 0183

- **RMC**

$GPRMC,hhmmss.ss,A,llll.ll,a,yyyyy.yy,a,x.x,x.x,ddmmyy,x.x,a*hh

RMC = Recommended Minimum Specific GPS/TRANSIT Data

1 = UTC of position fix
2 = Data status (V=navigation receiver warning)
3 = Latitude of fix
4 = N or S
5 = Longitude of fix
6 = E or W
7 = Speed over ground in knots
8 = Track made good in degrees True
9 = UT date
10 = Magnetic variation degrees (Easterly var. subtracts from true course)
11 = E or W
12 = Checksum

ข้อมูลที่ใช้จากรูปแบบ GPRMC มีดังนี้

ตำแหน่งลำดับที่ 1 UTC (Universal Time Coordinate) เป็นเวลามาตรฐานที่ Prime Meridian (Longitude 0 องศา) ที่ลากผ่านเมือง Greenwich ประเทศอังกฤษ ดังนั้นจึงเรียกว่า GMT (Greenwich Mean Time) เวลา UTC ใช้เป็นเวลามาตรฐานของวงการดาราศาสตร์ อุตุนิยมวิทยา การเดินเรือ เป็นต้น เวลาในท้องถิ่นในแต่ละประเทศจะเทียบกับเวลา UTC นี้ โดยการบวกหรือลบกับ UTC Offset สำหรับแต่ละพื้นที่ โดยเวลาในประเทศไทยเท่ากับ UTC+7 (UTC Offset ประเทศไทยเท่ากับ +07 ชั่วโมง ตำแหน่งลำดับที่ 2 มี 2 สถานะคือ A = data valid และ V = data not valid ในที่นี้ข้อมูลที่มีสถานะ A ตำแหน่งลำดับที่ 3 เป็นตำแหน่งของ Latitude ของอุปกรณ์ GPS ในสถานะยืนยันตำแหน่ง (fix) ตำแหน่งลำดับที่ 5 เป็นตำแหน่งของ Longitude ของอุปกรณ์ GPS ในสถานะยืนยันตำแหน่ง (fix)

ขั้นตอนที่ 4 ทำการปรับข้อมูลให้อยู่ในรูปแบบเพื่อ การวิเคราะห์ความถดถอยเชิงเส้น โดยเริ่มจากการหาค่าเวลาในการเดินทางที่แท้จริง

การหาค่าเวลาในการเดินทางที่แท้จริง (True Travel Time)(ไพโรจน์ พาสพิษณุ, 2550) ในแต่ละครั้งที่ยานพาหนะเคลื่อนที่ผ่านเส้นทางที่กำหนด เมื่อเคลื่อนที่ผ่านจุดที่กำหนดในขั้นตอนที่ 1 ผู้เก็บข้อมูลจะบันทึกเวลาที่เคลื่อนที่ผ่าน (Timestamp) ซึ่งทำให้ได้ค่าเวลา 2 ค่า คือ เวลาก่อนเข้า Link และสิ้นสุด Link แต่ค่าเวลาดังกล่าวอาจคลาดเคลื่อนเพราะเป็นการสังเกตของผู้เก็บข้อมูลขณะที่อยู่บนยานพาหนะที่เคลื่อนที่ จึงจำเป็นต้องตรวจสอบโดยการนำค่าเวลาทั้ง 2 ไปเปรียบเทียบกับเวลาและตำแหน่งในประโยค GPRMC

ต่อจากนั้นทำการระบุตำแหน่งนั้นลงบนแผนที่จำลองเพื่อดูว่าอยู่ใกล้เส้นอ้างอิงมากที่สุดหรือไม่ หากยังไม่ถูกต้องให้เลื่อนไปยังตำแหน่งถัดไป และทำซ้ำจนกระทั่งได้ตำแหน่งที่ถูกต้อง ซึ่งจะนำไปสู่เวลาที่ยานพาหนะเคลื่อนที่ผ่านเส้นอ้างอิงที่ถูกต้องด้วยเช่นกัน ท้ายที่สุดเมื่อได้ค่าเวลาที่ถูกต้อง ณ จุดเริ่มต้นของ Link ไปถึงจุดสิ้นสุดของ Link เมื่อนำมาหาผลต่าง ก็จะทำให้ได้เวลาในการเดินทางที่แท้จริง

การปรับข้อมูลให้อยู่ในรูปแบบสำหรับการวิเคราะห์ความถดถอยเชิงเส้น

ทำหลังจากได้ข้อมูลเวลาการเดินทางที่เป็นเวลาที่แท้จริง และข้อมูลระดับความติดขัด ณ เวลาเดินทางเรียบร้อยแล้วจัดเรียงไว้ในคอลัมน์ที่ 1 ส่วนคอลัมน์ที่ 2 คือระดับความติดขัดเฉลี่ยแสดงดังภาพที่ 3.5 เพื่อเตรียมเข้าสู่ขั้นตอนถัดไปคือการวิเคราะห์ความถดถอยเชิงเส้น

เนื่องจากในส่วนของข้อมูลระดับความติดขัด เป็นตัวอักษรภาษาอังกฤษคือ H, M, L จำเป็นต้องทำการแปลงค่าเป็นตัวเลข เพื่อความสะดวกสำหรับการวิเคราะห์ ดังนั้นระดับความติดขัด H เท่ากับ 100, M เท่ากับ 50 และ L เท่ากับ 0 หลังจากนั้นจึงนำเอาตัวเลขที่แสดงถึงระดับความติดขัดมาหาค่าเฉลี่ยในแต่ละเส้นทาง

ภาพที่ 3.5

แสดงการปรับข้อมูลให้อยู่ในรูปแบบการวิเคราะห์ความถดถอยเชิงเส้น

Congestion Level (Average)	Travel Time (3 Links) (sec.)	
50	1260	ตัวอย่างการเดินทางครั้งที่ 1
83.3	1870	
0	553	
0	578	
.	.	
.	.	
.	.	
50	1047	ตัวอย่างการเดินทางครั้งสุดท้าย

การปรับข้อมูลให้อยู่ในรูปแบบสำหรับการป้อนข่ายงานประสาทเทียม

หลังจากได้ชุดข้อมูลทั้ง 35 ชุดในแต่ละเส้นทางแล้ว จึงทำการแปลงข้อมูลให้อยู่ในรูปแบบของไฟล์นามสกุล arff ซึ่งเป็นไฟล์สำหรับประมวลผลข่ายงานประสาทเทียมผ่านโปรแกรม Weka 3.6.0 โดยภายในไฟล์จะเป็นไฟล์สำหรับสร้าง และทดสอบข่ายงานประสาทเทียม ตามรูปแบบ Leave one out Cross Validation โดยแบ่งข้อมูลออกตามจำนวนชุดข้อมูล ใช้ส่วนที่เป็น n-1 สำหรับสอน โดยที่ n คือจำนวนชุดข้อมูล และอีก 1 ส่วนสำหรับทดสอบ จากนั้นนำผลลัพธ์ที่ได้นำไปทดสอบหาค่าความผิดพลาด ขั้นตอนของการใช้โปรแกรม Weka ดูได้จากภาคผนวก ก แต่เนื่องจากโหนดอินพุตสามารถใส่ระดับความติดขัดเป็นตัวอักษรได้เลยจึงไม่ต้องทำการแปลงเป็นตัวเลขเหมือนเทคนิคการวิเคราะห์ความถดถอยเชิงเส้น

ภาพที่ 3.6

แสดงการปรับข้อมูลให้อยู่ในรูปแบบสำหรับป้อนข่ายงานประสาทเทียม

Link ID (108)	Time Last Status(sec.)	Link ID (102)	Time Last Status(sec.)	Link ID (2209)	Time Last Status(sec.)	Travel Time (sec.)	
H	7200	M	7200	L	7200	1200	ตัวอย่างครั้งที่ 1
M	7200	M	601	L	7200	1439	
M	7200	M	7200	L	7200	1847	
H	3900	M	7200	L	7200	1021	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
L	7200	L	7200	L	7200	428	ตัวอย่างครั้งสุดท้าย

ขั้นตอนที่ 5 สร้างและทดสอบหาค่าผลลัพธ์การวิเคราะห์สมการถดถอยเชิงเส้น

การสร้างและทดสอบการวิเคราะห์สมการถดถอยเชิงเส้น

สำหรับสมการความสัมพันธ์เชิงเส้นของงานวิจัยนี้จะอยู่ในรูปของ

$$\text{Travel Time} = a\,(\text{Sum Congestion Level}) + b \qquad (3.2)$$

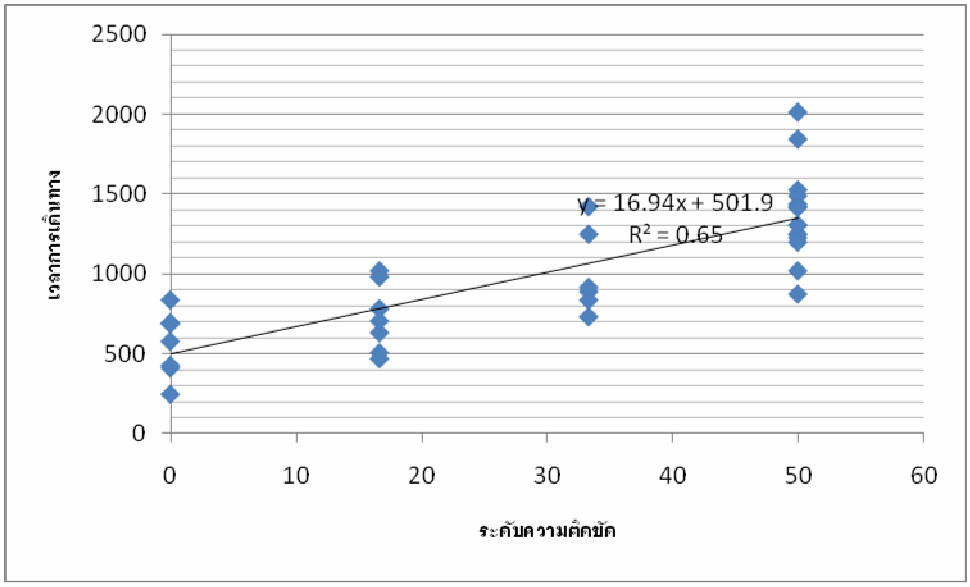
โดยที่

a คือ ความชันของเส้นตรง ซึ่งเป็นค่าที่แสดงถึงอัตราการเปลี่ยนแปลงของเวลาในการเดินทาง เมื่อผลรวมสะสมของระดับความติดขัดเปลี่ยนแปลงไป 1 หน่วย

b คือ ค่าส่วนตัดแกน Y หรือค่าเวลาในการเดินทางเมื่อ Sum Congestion Level ค่าเป็นศูนย์ โดยภายหลังจากการนำข้อมูลทั้ง 2 ชุด คือเวลาการเดินทาง กับ ระดับความติดขัดที่จัดเตรียมไว้ เข้าสู่การวิเคราะห์สมการการถดถอยเชิงเส้น ซึ่งผู้วิจัยได้ใช้โปรแกรม Weka เวอร์ชั่น 3.6.0 ทำการคำนวณ ผลลัพธ์ที่ได้คือ

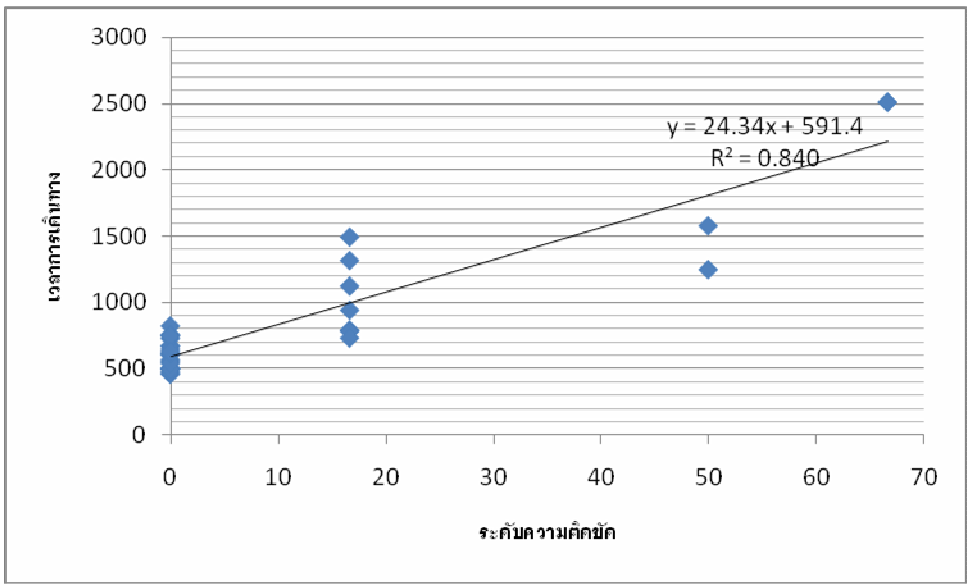
ภาพที่ 3.7

กราฟแสดงค่า R Square เส้นทางที่ 1



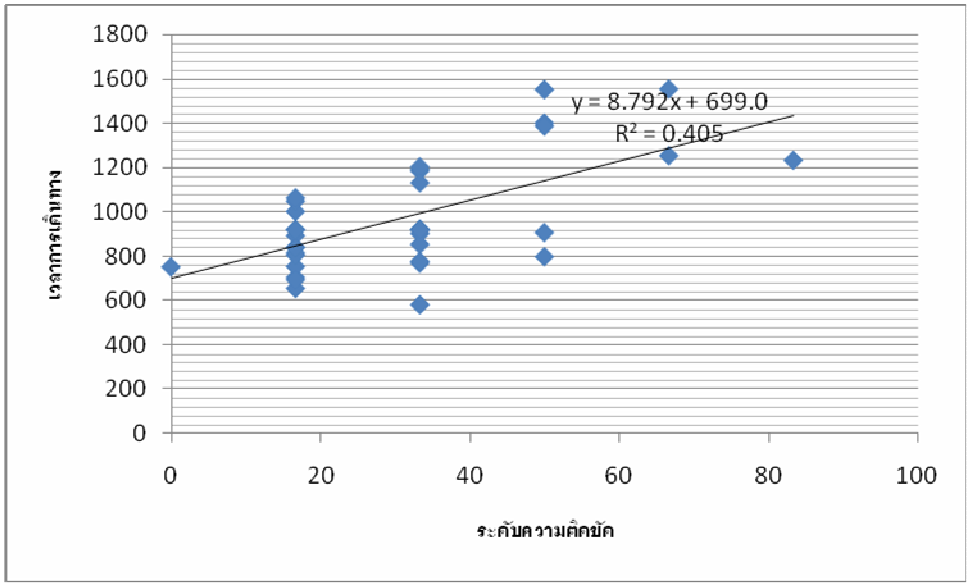
ภาพที่ 3.8

กราฟแสดงค่า R Square เส้นทางที่ 2



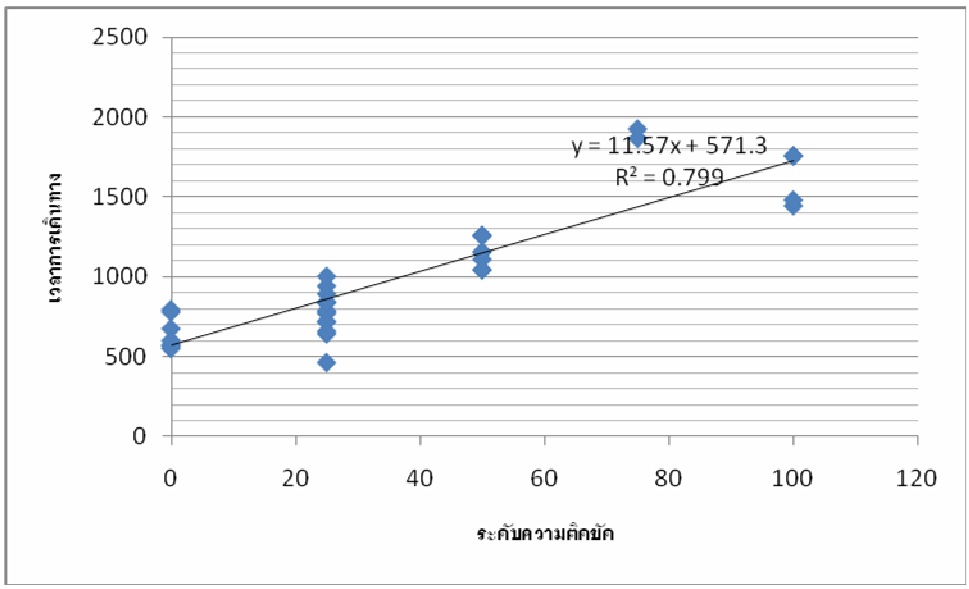
ภาพที่ 3.9

กราฟแสดงค่า R Square เส้นทางที่ 3



ภาพที่ 3.10

กราฟแสดงค่า R Square เส้นทางที่ 4



ภาพที่ 3.11

ตัวอย่างตารางค่าประมาณเวลาการเดินทางจากสมการถดถอยเชิงเส้น

เส้นทางที่ 1Link ID 108, 102, 2209

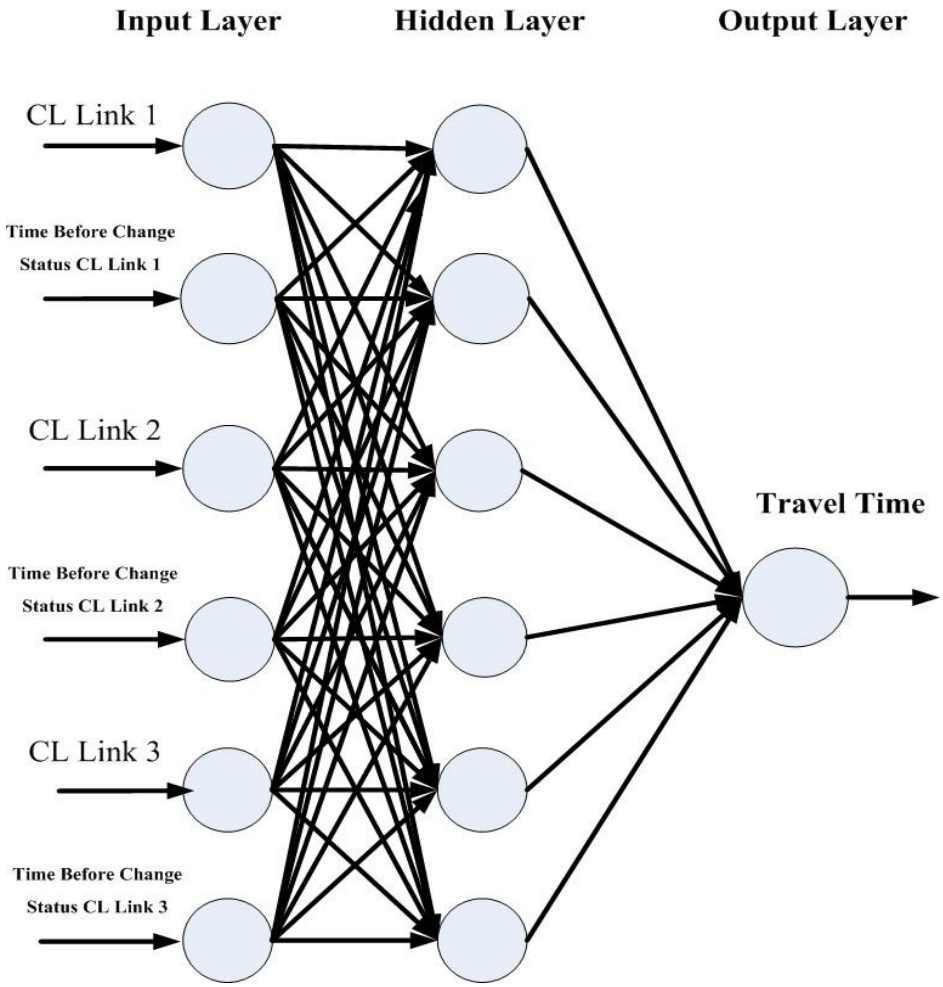
108	102	2209	เวลาประมาณ (3 Links)
L	L	L	0:07:42
L	L	M	0:12:27
L	L	H	0:17:12
L	M	L	0:11:31
L	M	M	0:16:17
L	M	H	0:21:02
L	H	L	0:15:21
.	.	.	.
.	.	.	.
.	.	.	.
H	H	M	0:30:32
H	H	H	0:35:18

ตัวอย่างการคำนวณจากสมการ การวิเคราะห์ถดถอยเชิงเส้น ผู้วิจัยได้ทำการคำนวณทุกเหตุการณ์ (Combination) ที่มีความน่าจะเป็นที่จะเกิดทั้งหมด 35 เหตุการณ์ จาก 3 Links 1 เส้นทาง ซึ่งการเจริญเติบโตของตัวเลข จะอยู่ในรูปแบบ 3^n โดยที่ n คือจำนวน Links และเวลาประมาณที่ได้ คำนวณจากสมการวิเคราะห์การถดถอยเชิงเส้นดังนี้

$$y = 16.944x + 501.99 \quad (3.3)$$

ขั้นตอนที่ 6 สร้างและทดสอบข่ายงานประสาทเทียมโดยใช้ข้อมูลทั้งหมด 35 ชุด ต่อ Link โดยเลือกใช้แนวทาง Leave one out Cross Validation เพราะจำนวนตัวอย่างที่มีน้อย โดยโครงสร้างข่ายงานประสาทเทียมแสดงดังภาพที่ 3.12

ภาพที่ 3.12
แสดงโครงสร้างข่ายงานประสาทเทียมเอาต์พุตเวลาการเดินทาง



การสร้างและทดสอบข่ายงานประสาทเทียม

ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) (Rumelhart และ McClelland, 1986) ได้ทำการสร้างคุณสมบัติพิเศษที่เพิ่มขึ้นคือ ทำให้ข่ายงานประสาทเทียมสร้างพื้นผิวการตัดสินใจแบบไม่เป็นเชิงเส้นได้ (Nonlinear Decision Surface) โดยที่สามารถแบ่งแยกตัวอย่างได้ดี กว่าพื้นผิวการตัดสินใจแบบเชิงเส้น (Linearly Decision Surface) อัลกอริทึมแบบแพร่กระจายย้อนกลับนั้น ได้ใช้กฎการเรียนรู้คือ กฎเดลต้า (Delta Rule) มีข้อดีที่การเรียนรู้จะลู่เข้าสู่ระนาบหลายลักษณะ ทำให้ค่าผิดพลาดเกิดขึ้นน้อย และใช้หลักการเคลื่อนที่ลงตามความชัน (Gradient Descent) กฎเดลต้าจะหาเวกเตอร์น้ำหนักที่ให้ค่าผิดพลาดของตัวอย่างสอนน้อยที่สุด โดยการคำนวณทางคณิตศาสตร์ จึงต้องใช้ฟังก์ชันกระตุ้นที่สามารถหาอนุพันธ์ได้ เช่น ฟังก์ชันเชิงเส้น (Linear Function) หรือ ฟังก์ชันซิกมอยด์ (Sigmoid Function)

โครงสร้างข่ายงานประสาทเทียมสำหรับการวิจัย มีองค์ประกอบต่าง ๆ ดังนี้

Input Layer ประกอบไปด้วยข้อมูลสำคัญที่ได้จากการเก็บข้อมูล โดยงานวิจัยนี้ใช้เวลาการเดินทาง และความเร็ว แต่ละ Link ที่ปรากฏบนเส้นทางในการทดลอง 1 โหนด เท่ากับ 1 Link พร้อมกับระดับความติดขัด

Hidden Layer เป็นชั้นที่เป็นส่วนสำคัญสำหรับ ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับ มีหน้าที่ปรับค่าน้ำหนักผิดพลาดจากข้อมูลสอน ที่เรียกว่า การเรียนรู้ของข่ายงานประสาทเทียม จากงานวิจัยส่วนใหญ่จะใช้จำนวนโหนด ในชั้นซ่อนเท่ากับจำนวนโหนดของ input node รวมกับจำนวนของ output node จากนั้นหารสอง หรือเท่ากับจำนวน input node แต่สำหรับการวิจัยนี้ได้ทำการทดลองแบบลองผิดลองถูก หาจำนวน Hidden Layer ที่เหมาะสมที่สุด

Output Layer เป็นชั้นที่แสดงผลลัพธ์ จากการเรียนรู้ของข่ายงานประสาทเทียม ซึ่งการวิจัยนี้ใช้ เวลาในการเดินทางกับความเร็ว

จากการสร้างข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับมีโครงสร้างดังภาพที่ 3.12 ส่วนของ input layer ใช้จำนวนโหนด เท่ากับจำนวนเซลล์ที่เกิดขึ้นบนเส้นทางนั้น ส่วนของ hidden layer ในแต่ละโหนด ใช้ฟังก์ชันกระตุ้นซิกมอยด์ (Sigmoid Transfer Function) แสดงดังสมการที่ 3.4 ทำหน้าที่แปลงค่า input layer ให้เป็นแบบไม่เชิงเส้น (Nonlinearity) เพื่อให้

สอดคล้องกับคุณสมบัติของข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับ โดยจำนวนโหนดที่ใช้ถ้าใช้แบบจำนวนของ input node รวมกับจำนวนของ output node จากนั้นหารสอง ผลการสอนข่ายงานประสาทเทียมมีค่าความผิดพลาดสูง จึงเลือกใช้จำนวนโหนดเท่ากับจำนวน input node ซึ่งการเรียนรู้ของข่ายงานประสาทเทียมจะเรียนรู้ได้ดีกว่าจำนวนโหนดน้อย หรือสามารถสร้างระนาบหลายมิติการเรียนรู้ที่ซับซ้อนมากขึ้น แต่จะทำให้ข่ายงานประสาทเทียมมีความซับซ้อนสูงตามไปด้วย และจะทำให้ใช้เวลาการเรียนรู้นาน และอาจเกิดปัญหาการเข้ากันเกินไป (Overfitting Problem) (Mitchell, 1997:108-111) ดังนั้นจึงเลือกจำนวนโหนดในชั้น hidden layer ตามผลลัพธ์ที่ดีที่สุด

$$Y = \frac{1}{1+e^{-x}} \tag{3.4}$$

Output layer ในการวิจัยนี้ คือ เวลาในการเดินทางและความเร็ว ซึ่งประกอบไปด้วยโหนด ที่ใช้ฟังก์ชันกระตุ้นเชิงเส้น (Linear Transfer Function) แสดงดังสมการที่ 3.5 ทำหน้าที่ส่งผลรวมค่าผลลัพธ์ที่ได้ของแต่ละโหนดจาก Hidden layer ส่งต่อไปยัง Output layer โดยไม่เปลี่ยนแปลงค่าของข้อมูล หรือค่า α ตามสมการที่ 3.5 ซึ่งมีค่าเท่ากับ 1

$$Y = \alpha X \tag{3.5}$$

การทดลองแบบลองผิดลองถูกจะมีการปรับค่าพารามิเตอร์ที่สำคัญได้แก่ อัตราการเรียนรู้ (Learning Rate) ค่าโมเมนตัม (Momentum) และจำนวนรอบที่ต้องการให้ข่ายงานประสาทเทียมเรียนรู้ (Epoch) โดยการเลือกปรับค่าพารามิเตอร์ต่าง ๆ เป็นเรื่องยาก ต้องใช้การลองผิดลองถูก (Trial & Error) ค่าอัตราการเรียนรู้ จะเป็นค่าคงที่ มีค่าอยู่ระหว่าง 0 และ 1 เป็นค่าที่แสดงถึงอัตราการปรับค่าความผิดพลาดการสอนข่ายงานประสาทเทียมในแต่ละรอบ ค่าที่สูงใกล้ค่า 1 ข่ายงานประสาทเทียมจะใช้เวลาการเรียนรู้เร็ว แต่อาจจะทำให้ผลการเรียนรู้ที่ได้ออกมาไม่ดี ค่าโมเมนตัม เป็นค่าคงที่ มีค่าอยู่ระหว่าง 0 และ 1 เป็นค่าที่เป็นตัวคูณของปริมาณค่าความผิดพลาดจากการสอนข่ายงานประสาทเทียมในรอบก่อนหน้านี้ ผลคูณที่ได้จะถูกนำไปเพิ่มให้กับค่าความผิดพลาดจากการสอนข่ายงานประสาทเทียมในรอบปัจจุบัน เป็นการเร่งการเรียนรู้

และปรับค่าความผิดพลาดได้มากขึ้น แต่ใช้เวลาในการเรียนรู้นาน การวิจัยนี้จะใช้การปรับค่าพารามิเตอร์ดังนี้

Learning Rate เริ่มจากค่า 0.5 แล้วลดค่าลงทีละ 1
Momentum เริ่มจากค่า 0.1 แล้วเพิ่มค่าขึ้นทีละ 0.1 ไม่เกิน 0.4
Epochs เริ่มจากค่า 500 แล้วเพิ่มค่าทีละ 500 Epochs

นอกจากนี้ผู้วิจัยได้เลือกใช้เทคนิคแบบ Leave one out Cross-Validation สำหรับจำนวนข้อมูลที่มีน้อย (Mitchell, 1997) โดยแบ่งข้อมูลออกเป็น n ส่วน ตามจำนวนอินพุต และใช้ส่วนที่เป็น n-1 สำหรับสอน โดยที่ n คือจำนวนข้อมูลอินพุต และอีก 1 ส่วนสำหรับทดสอบ ผลลัพธ์สุดท้ายจะถูกนำมาหาค่าเฉลี่ย เงื่อนไขการหยุดการสอนข่ายงานประสาทเทียมพิจารณาจากครบรอบตามจำนวน Epoch หรือค่าความผิดพลาดจากชุดข้อมูลสอนมีค่าต่ำและผลการประมาณค่าจากตัวอย่างทดสอบมีค่าต่ำ หรือเมื่อผลการประมาณค่าไม่แตกต่างไปจากเดิม

ขั้นตอนที่ 7 เปรียบเทียบผลลัพธ์เทคนิคการวิเคราะห์สมการถดถอย กับเทคนิคข่ายงานประสาทเทียม ในรูปแบบแยกแต่ละเส้นทางกับรวมเส้นทาง (อธิบายในหัวข้อถัดไป) ต่อจากนั้นทำการคำนวณค่าทางสถิติ จากนั้นจึงสรุปผล

การแปลงค่าเวลาการเดินทางเป็นความเร็ว

จากผลการทดลอง เนื่องจากงานวิจัยนี้ผู้วิจัยได้ตั้งสมมติฐานในข้อที่ 2 ว่าวิธีการใช้เทคนิคข่ายงานประสาทเทียม สำหรับการประมาณเวลาในการเดินทางมีความถูกต้อง มากกว่าแนวทางการวิเคราะห์ความถดถอยเชิงเส้น แต่ผลการทดลองที่ออกมาไม่เป็นไปตามสมมติฐานใน 3 เส้นทางที่ทำการทดลอง ผู้วิจัยจึงพิจารณาถึงจำนวนชุดข้อมูลที่ทำการทดลอง 35 ชุดนั้นน้อยไปประกอบกับ เทคนิคการวิเคราะห์ความถดถอยเชิงเส้น มีระดับความติดขัดเป็นสัดส่วนเชิงเส้นกับเวลาในการเดินทางจริง ดังนั้นการประมาณเวลาในการเดินทางจากการวิเคราะห์สมการถดถอยเชิงเส้นจึงให้ค่าความผิดพลาดที่ต่ำ ดังนั้นผู้วิจัยจึงทำการทดลองเพิ่มเติม โดยทำการแปลงค่าเวลาการเดินทาง จากเวลาการเดินทางเป็น ความเร็ว เพื่อที่จะสามารถนำชุดข้อมูลจากทุกเส้นทางมารวมกันได้ จึงทำให้มีจำนวนชุดข้อมูลเพิ่มขึ้น จาก 35 ชุดของแต่ละเส้นทาง เป็น 140 ชุด

และผู้วิจัยมีความประสงค์ ที่จะต้องการใช้รูปแบบการประมาณเวลาในการเดินทางกับในทุกเส้นทางที่มีคุณลักษณะของถนนที่อยู่ในเมืองใกล้เคียงกันและมีข้อมูลระดับความติดขัดไม่ใช่เฉพาะเจาะจงให้การทดลองทำการประมาณเวลาการเดินทาง เพียงแต่ในเส้นทางที่ทำการเก็บข้อมูล จึงทำการตั้งสมมติฐานเพิ่มเติมอีกว่า ถ้าเปลี่ยนเอาต์พุตจากเวลาการเดินทาง มาเป็นค่าความเร็ว โดยทำการคำนวณแปลง จากค่าเดิมคือ เวลาการเดินทาง เปลี่ยนเป็นค่าความเร็ว ก็จะสามารถนำไปประยุกต์ใช้ได้กับหลายเส้นทาง

สำหรับวิธีดำเนินการทดลองใน **ขั้นตอนที่** 1 - 3 ยังคงทำเหมือนกับการทดลองดังกล่าวข้างต้น คือ ทำการศึกษาเส้นทาง จัดเตรียมอุปกรณ์เครื่องมือ และทำการเก็บข้อมูล ต่อจากนั้นก่อนที่จะเข้าสู่ขั้นตอนที่ 4 จะต้องทำการคำนวณ แปลงค่าเวลาการเดินทาง เป็นค่าความเร็ว ซึ่งจะคำนวณได้จากสูตร

$$v = \frac{s}{t} \tag{3.6}$$

โดยที่

v หมายถึง ความเร็ว

s หมายถึง ระยะทาง

t หมายถึง เวลาการเดินทาง

ขั้นตอนที่ 4

การปรับข้อมูลให้อยู่ในรูปแบบสำหรับการวิเคราะห์ความถดถอยเชิงเส้น ทำหลังจากได้ข้อมูลเป็นค่าความเร็วของเส้นทางทั้ง 4 เส้นทาง จากจำนวนข้อมูล 35 ครั้งในแต่ละเส้นทาง เมื่อแปลงเป็นค่าความเร็วแล้ว ก็จะสามารถนำข้อมูลค่าความเร็วมารวมกันได้ จำนวนข้อมูลทั้งหมดจึงเพิ่มเป็น 140 ครั้ง โดยค่าความเร็วจัดเรียงไว้ในคอลัมน์ที่ 1 ส่วนคอลัมน์ที่ 2 คือระดับความติดขัดเฉลี่ยแสดงดังภาพที่ 3.13 เพื่อเตรียมเข้าสู่ขั้นตอนถัดไปคือการวิเคราะห์ความถดถอยเชิงเส้น

การปรับข้อมูลให้อยู่ในรูปแบบสำหรับการป้อนข่ายงานประสาทเทียมหลังจากได้ชุดข้อมูลทั้ง 140จึงทำการแปลงข้อมูลให้อยู่ในรูปแบบของไฟล์นามสกุล arff ซึ่งเป็นไฟล์สำหรับ

ประมวลผลข่ายงานประสาทเทียมผ่านโปรแกรม Weka 3.6.0 โดยภายในไฟล์จะเป็นไฟล์สำหรับสร้าง และทดสอบข่ายงานประสาทเทียม ตามรูปแบบ Leave one out Cross Validation โดยแบ่งข้อมูลออกตามจำนวนอินพุต ใช้ส่วนที่เป็น n-1 สำหรับสอน โดยที่ n คือจำนวนอินพุต และอีก 1 ส่วนสำหรับทดสอบ จากนั้นนำผลลัพธ์ที่ได้นำไปทดสอบหาค่าความผิดพลาด ขั้นตอนของการใช้โปรแกรม Weka ดูได้จากภาคผนวก ก

ขั้นตอนที่ 5 สร้างและทดสอบหาค่าผลลัพธ์การวิเคราะห์สมการทดถอยเชิงเส้น และสำหรับสมการความสัมพันธ์เชิงเส้นของงานวิจัยนี้จะอยู่ในรูปของ

$$\text{Velocity} = a\,(\text{Sum Congestion Level}) + b \qquad (3.7)$$

a คือ ความชันของเส้นตรง ซึ่งเป็นค่าที่แสดงถึงอัตราการเปลี่ยนแปลงของค่าความเร็ว เมื่อผลรวมสะสมของระดับความติดขัดเปลี่ยนแปลงไป 1 หน่วย

b คือ ค่าส่วนตัดแกน Y หรือค่าเวลาในการเดินทางเมื่อ Sum Congestion Level เป็นศูนย์โดยภายหลังจากการนำข้อมูลทั้ง 2 ชุด คือค่าความเร็ว กับ ระดับความติดขัดที่จัดเตรียมไว้ เข้าสู่การวิเคราะห์สมการการถดถอยเชิงเส้น ซึ่งผู้วิจัยได้ใช้โปรแกรม Weka เวอร์ชั่น 3.6.0 ทำการคำนวณ

ภาพที่ 3.13

แสดงการปรับข้อมูลให้อยู่ในรูปแบบการวิเคราะห์ความถดถอยเชิงเส้น

Congestion Level (Average)	Velocity (km/h)	
50	5.869	ตัวอย่างครั้งที่ 1
50	4.895	
50	3.814	
33.33	9.609	
.	.	
.	.	
.	.	
16.66	9.371	ตัวอย่างครั้งสุดท้าย

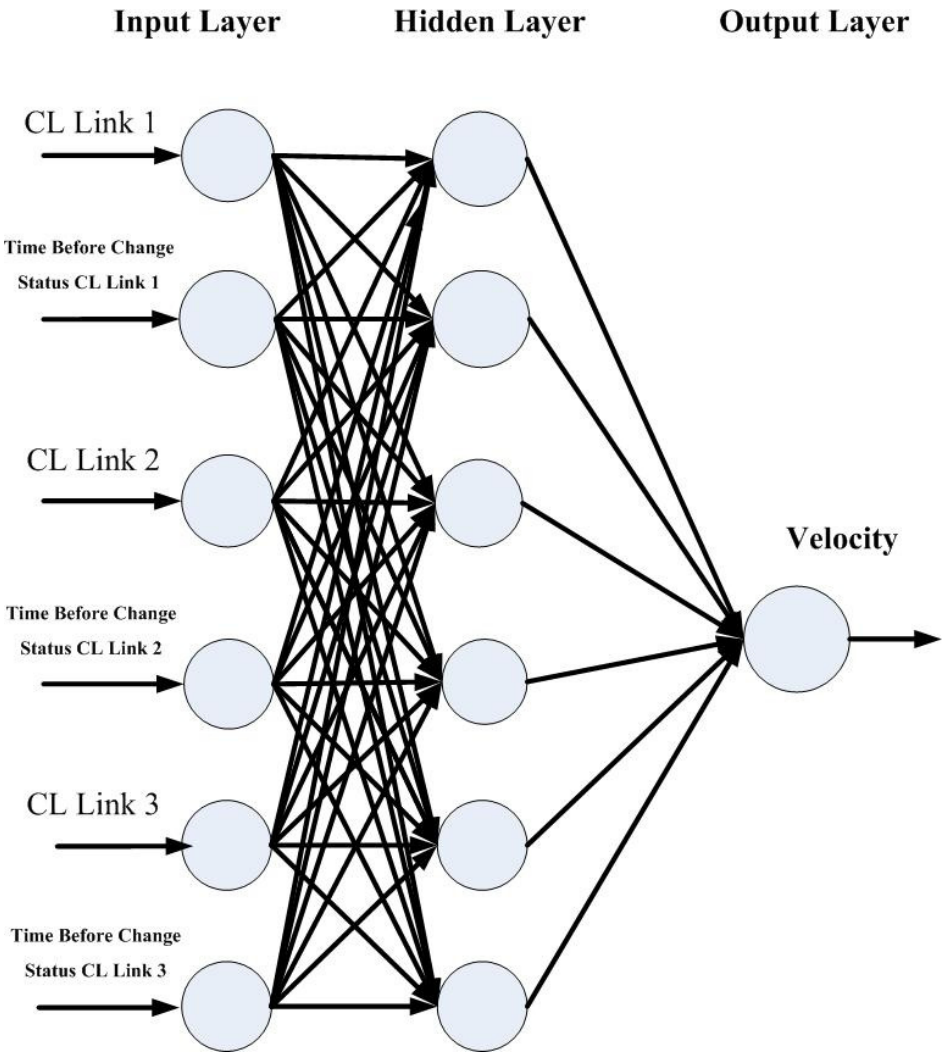
ขั้นตอนที่ 6 สร้างและทดสอบข่ายงานประสาทเทียมโดยใช้ข้อมูลทั้ง 140 ชุด ซึ่งค่าเอาต์พุตคือค่าความเร็ว ซึ่งคำนวณได้จากสมการที่ 3.6 แสดงการปรับข้อมูลให้อยู่ในรูปแบบข่ายงานประสาทเทียมในภาพที่ 3.14 และโครงสร้างข่ายงานประสาทเทียมแสดงดังภาพที่ 3.15

ภาพที่ 3.14

แสดงการปรับข้อมูลให้อยู่ในรูปแบบข่ายงานประสาทเทียม

Link 1	Time Last Status(sec.)	Link 2	Time Last Status(sec.)	Link3	Time Last Status(sec.)	Velocity (km/h)	
H	7200	M	7200	L	7200	5.8	ตัวอย่างครั้งที่ 1
M	7200	M	601	L	7200	9.6	
M	7200	M	7200	L	7200	8.4	
H	3900	M	7200	L	7200	5.6	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
L	7200	L	7200	L	7200	13.3	ตัวอย่างครั้งสุดท้าย

ภาพที่ 3.15
แสดงโครงสร้างข่ายงานประสาทเทียมเอาต์พุตความเร็ว



ขั้นตอนที่ 7 เปรียบเทียบผลลัพธ์เทคนิคการวิเคราะห์สมการถดถอย กับเทคนิคข่ายงานประสาทเทียม ในรูปแบบแยกแต่ละเส้นทาง กับรูปแบบรวมทุกเส้นทาง เพื่อให้นำไปประยุกต์ใช้ได้กับทุกเส้นทาง (Generalize) และคำนวณค่าทางสถิติ จากนั้นจึงสรุปผล