

Transportation Mode Classification from Mobile GPS Data

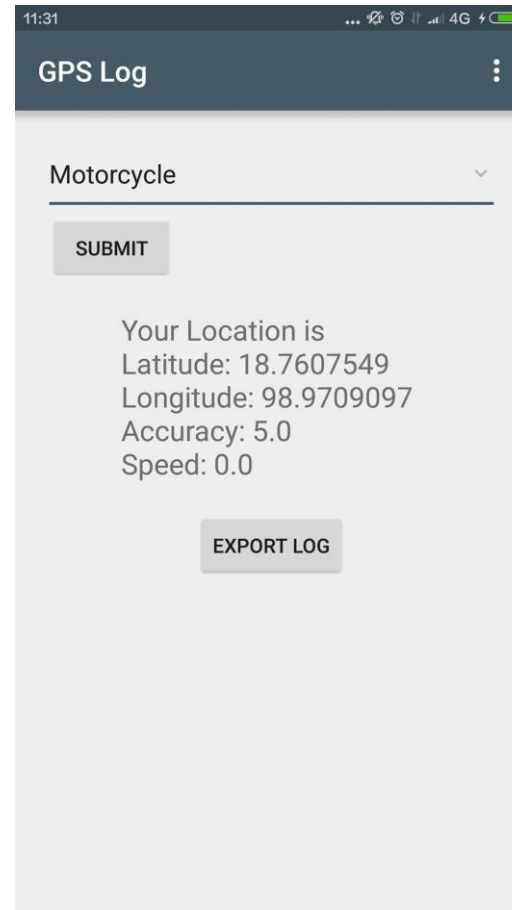
Introduction

ในปัจจุบัน เทคโนโลยีที่เจริญเติบโตขึ้น โดยเฉพาะอย่างยิ่งทางด้านอินเทอร์เน็ต และได้มีการเก็บข้อมูลเพื่อมาวิเคราะห์เพื่อศึกษาความสัมพันธ์ของข้อมูล เช่น ระบบแนะนำการเดินทาง แนะนำสินค้า รวมไปถึงการวิเคราะห์เพื่อจัดการบริหารเมืองให้ดียิ่งขึ้น จึงได้มีแนวคิดที่จะนำข้อมูล GPS จากโทรศัพท์มือถือซึ่งได้จากการเดินทางในรูปแบบต่าง ๆ ของผู้ใช้งานมาทำการวิจัยและทดลองศึกษารูปแบบการเดินทางในรูปแบบต่าง ๆ จากข้อมูล GPS เพื่อสร้างระบบ การทำนายรูปแบบการเดินทางได้

Overview



Data Collection



Database

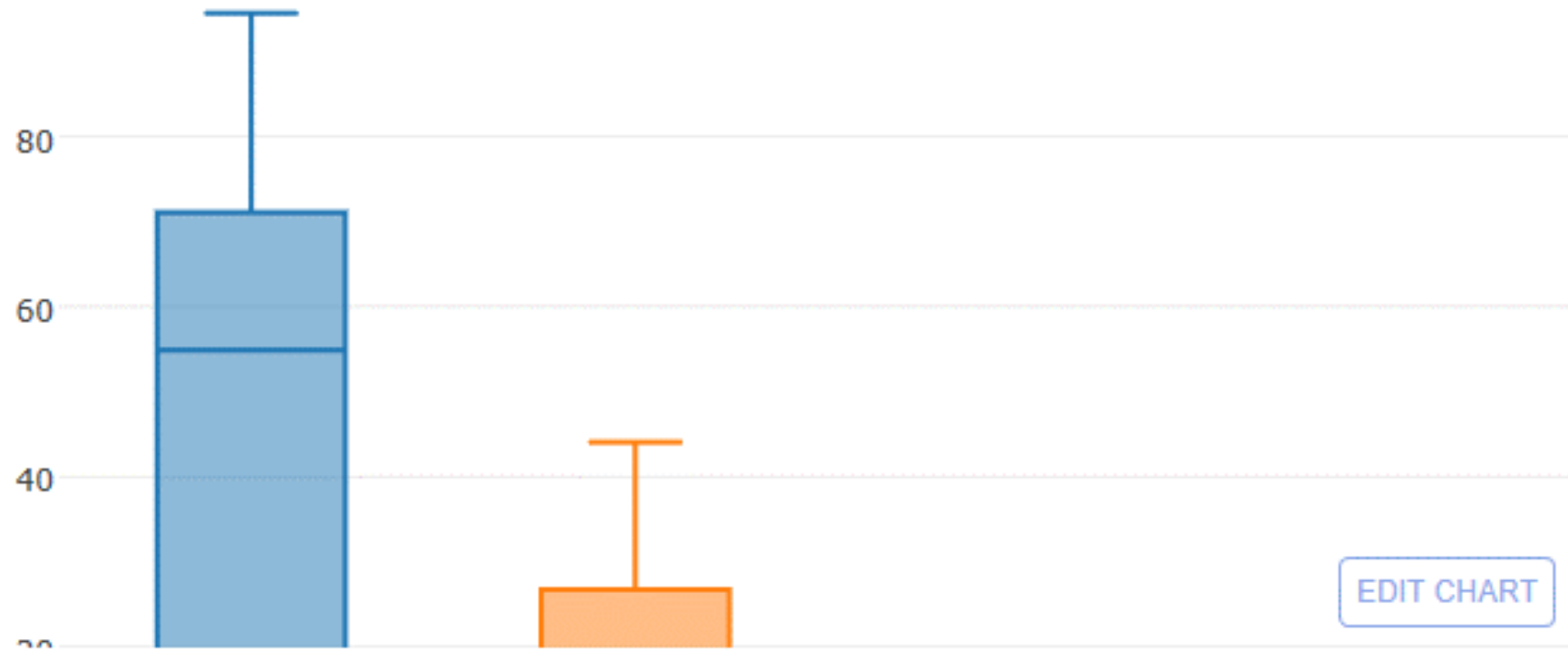
num	label	latitude	longitude	speed	accuracy	bering	time
1	Car	18.7905201	98.9551935	0.0	149.699005126953	0.0	1469804224297
1	Car	18.7905123	98.9552185	0.0	150.712005615234	0.0	1469804242743
1	Car	18.7904911	98.9552118	0.0	107.896003723145	0.0	1469804247636
1	Car	18.7904869	98.9552149	0.0	88.4010009765625	0.0	1469804252728
1	Car	18.7907486	98.9552814	0.0	159.0	0.0	1469804257763
1	Car	18.7907492	98.955281	0.0	96.0	0.0	1469804262000
1	Car	18.7907491	98.955281	0.0	96.0	0.0	1469804265000
1	Car	18.7907491	98.955281	0.0	96.0	0.0	1469804269000
1	Car	18.7907492	98.955281	0.0	78.0	0.0	1469804275000
1	Car	18.7907492	98.955281	0.0	96.0	0.0	1469804279000
1	Car	18.7907489	98.955281	0.0	50.0	0.0	1469804283000
1	Car	18.7907492	98.955281	0.0	40.0	0.0	1469804287000
1	Car	18.790752	98.9552815	0.0	96.0	0.0	1469804290000
1	Car	18.7907648	98.955284	0.0	96.0	0.0	1469804293000
1	Car	18.7907699	98.9552849	0.0	96.0	0.0	1469804296000
1	Car	18.7907701	98.9552849	0.0	96.0	0.0	1469804299000
1	Car	18.7908016	98.9553717	17.1000003814697	96.0	35.0	1469804302000
1	Car	18.7909567	98.9554188	18.8999996185303	96.0	29.0	1469804305000
1	Car	18.7917345	98.9549993	18.8999996185303	55.0	9.0	1469804309000
1	Car	18.7918839	98.9551315	18.8999996185303	43.0	30.0	1469804313000
1	Car	18.7920252	98.9552137	19.7999992370606	35.0	27.0	1469804316000
1	Car	18.7921047	98.9554562	17.1000003814697	29.0	30.0	1469804319000

ข้อมูลแสดงระบบกำหนดตำแหน่งบนโลก (GPS) ของการเดินทางแบบต่าง ๆ

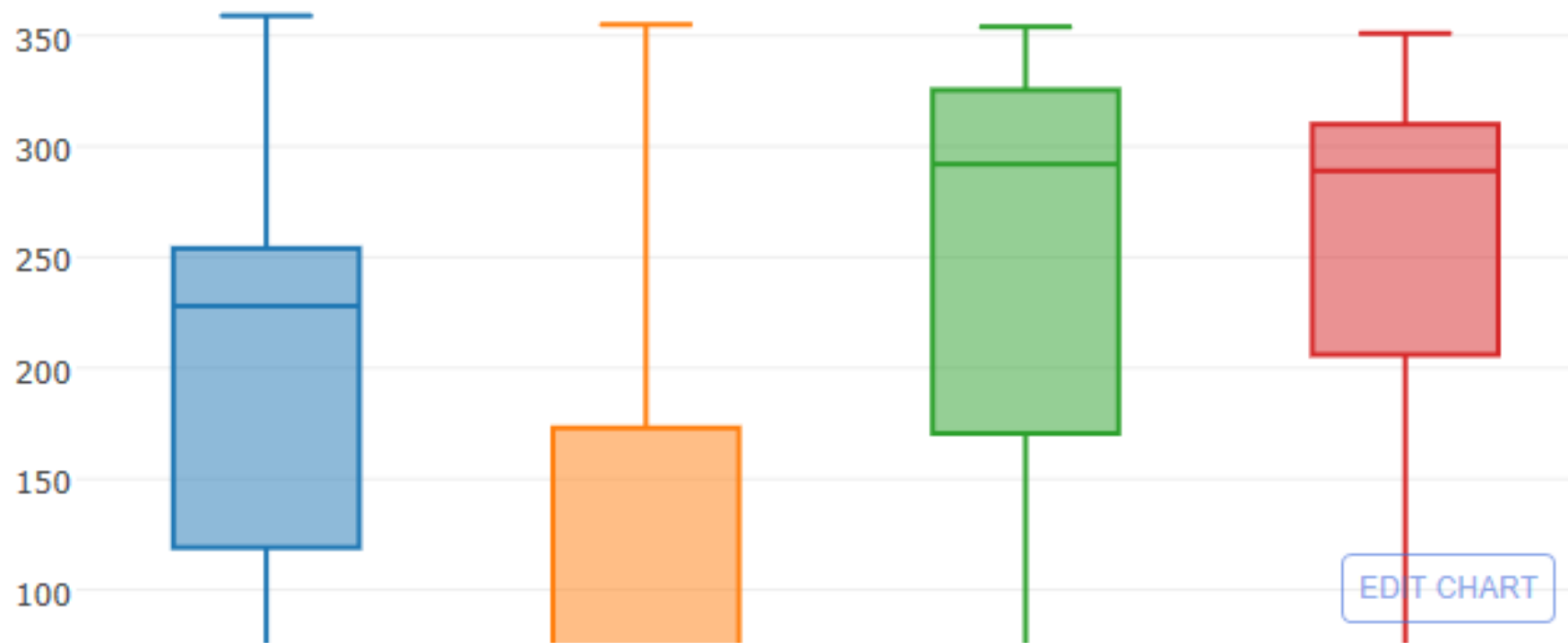
โดยมีข้อมูลที่ใช้ในการวิเคราะห์ดังนี้

- | | |
|-------------------------|--------------------------|
| 1. ละติจูด (Latitude) | 4. ความเร็ว (Speed) |
| 2. ลองจิจูด (Longitude) | 5. ความแม่นยำ (Accuracy) |
| 3. เวลา (Timestamp) | 6. มุม (Angle) |

Data



Data



Data Preprocessing

1. ลบค่าผิดปกติออกจากข้อมูล (Outliers)

1.1 ข้อมูล GPS ที่มีลำดับเวลาไม่ถูกต้อง

1.2 ข้อมูล GPS ที่มีความเร็วผิดปกติ โดยอ้างอิงจากแผนภาพกล่อง

1.3 ข้อมูล GPS ที่มีทิศทางแตกต่างจากกลุ่มข้อมูล



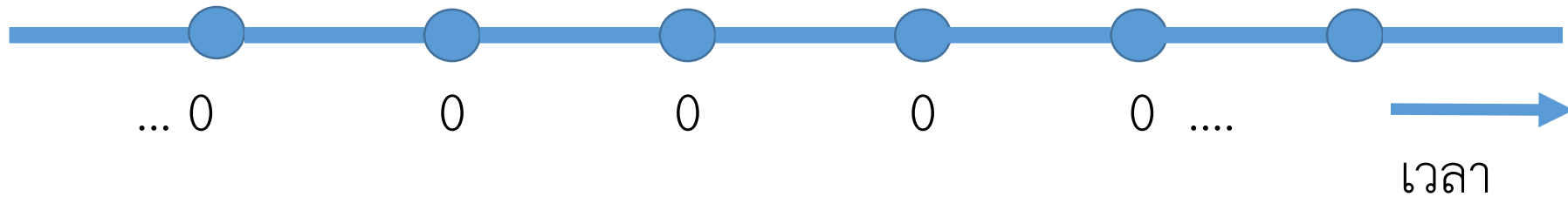
speed_plot.html



bearing_plot.html

Data Preprocessing

2. ความเร็ว



จะทำการลบจุดดังกล่าวออกเพื่อลดจำนวนข้อมูล ทำให้ประมวลผลได้เร็วขึ้น

3. ความแม่นยำ กรณีมากกว่า 200 ขึ้นไปทำการลบข้อมูลจุดนั้นออก

(We define accuracy as the radius of 68% confidence.)

Feature Extraction

การสกัดคุณลักษณะ

1. HCR(Heading Change Rate)[1][3] อัตราการเปลี่ยนแปลงทิศทางการเคลื่อนที่ต่อระยะทาง
2. SR(Stop Rate)[1][3] อัตราการหยุดต่อระยะทาง
3. VCR(Velocity Change Rate)[1][3] อัตราการเปลี่ยนความเร็วต่อระยะทาง
4. Walk or Non-walk [4] ตรวจจับการเคลื่อนที่เดิน หรือ ไม่เดิน
5. Stay point [4] จุดที่หยุดพัก
6. Average Speed [7] ความเร็วเฉลี่ย

Feature Extraction

การสกัดคุณลักษณะ

8. Average Acceleration[7] อัตราเร่งเฉลี่ย

9. *Speed deviation การเบี่ยงเบนของความเร็ว

10. *Bearing deviation การเบี่ยงเบนของทิศทาง

11. *Sliding-Window การเลื่อนหน้าต่าง เพื่อประสิทธิภาพในการสกัดคุณลักษณะ

Modeling

ทดลองและทำการสร้างโมเดลรูปแบบต่าง ๆ

1. Naïve Bayes
2. Decision Tree
3. Random Forest
4. Neural Networks

References

[1] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, Wei-Ying Ma, “Understanding Mobility Based on GPS Data”, Proceedings of the 10th ACM conference on Ubiquitous Computing, 2008

เป้าหมาย: เพื่อวิเคราะห์ข้อมูลการเดินทางในรูปแบบต่างๆ

ลักษณะข้อมูล: ข้อมูล GPS โดยเก็บข้อมูลทุก ๆ 2 วินาที มีการเก็บข้อมูลจาก 65 คนรวมระยะเวลาทั้งหมด 10 เดือนและผู้ทำการเก็บข้อมูลต้องทำ label ด้วย

วิธีการ: โดยสกัดคุณลักษณะจากข้อมูลออกมาดังนี้ Heading change rate, Stop rate, Velocity change rate โดยมี Decision tree เป็นโมเดลในการทดลอง

ผลที่ได้: ผลเฉลี่ยความถูกต้องรูปแบบการเดินทางดังนี้ เดิน, ขับรถ, รถบัส, จักรยาน เป็น 76.2%

References

[2] Sasank Reddy, Jeff Burke, Deborah Estrin, Mark Hansen, Mani Srivastava, “Determining Transportation Mode On Mobile Phones”, ISWC’08 Proceeding of the 12th IEEE International Symposium on Wearable Computers, 2008

เป้าหมาย: ระบุรูปแบบการเดินทางจากข้อมูลโทรศัพท์ที่ได้

ลักษณะของข้อมูล: GPS, Accelerometer, ตำแหน่งของโทรศัพท์มือถือบนตัวผู้ใช้นี้ ข้อมูล, ออก, แชน, กระเป๋าเป้, กระเป๋ากางเกง และมี Label รูปแบบการเดินทางต่าง ๆ โดยใช้ข้อมูลจาก 6 คนโดยในแต่ละครั้งจะเก็บข้อมูลยาว 8 นาทีในแต่ละแบบรวมข้อมูลทั้งหมด 20 ชั่วโมง

วิธีการ: โดยสกัดคุณลักษณะจากข้อมูลดังนี้ ความแปรปรวน และ พลังงาน จากผลรวมจากการทำ FFT ของ accelerometer และ ความเร็วจาก GPS และนำไปทดลองกับโมเดลต่าง ๆ ดังนี้ Naïve Bayes, Decision Tree, k-Nearest Neighbor, Support Vector Machine, Hidden Markov

References

และทดลองแยกชุดข้อมูลตามตำแหน่งที่แตกต่างกันด้วย

ผลที่ได้: ทุกโมเดลได้เกิน 90% ยกเว้น Naïve Bayes และตำแหน่งของโทรศัพท์มีผลทำให้ค่าความแม่นยำแตกต่างกันประมาณ 1.2%

[3] Yu Zheng, Like Liu, Longhao Wang, Xing Xie, “Learning Transportation Mode from Raw GPS Data for Geographic Applications on the web”, WWW’08 Proceedings of the 17th international conference on World Wide Web, 2008

ในงานชิ้นนี้เป็นผู้ทำเดียวกับ [1] ซึ่งมีการใช้ข้อมูลเดียวกัน และ โมเดลเหมือนกัน แต่เพิ่มวิธีการแบ่งข้อมูลออกเป็นส่วนๆ(Segmentation) โดยแบ่ง 2 แบบคือ ตามระยะทาง 100, 150, 200 เมตร ตามระยะเวลา 60, 90, 120 วินาที

References

[4] Apichon Witayangkurn, Teerayut Horanont, Yoshihide Sekimoto, “Trip Reconstruction and Transportation Mode Extraction on Low Data Rate GPS Data from Mobile Phone”, International Conference on Computers in Urban Planning and Urban Management, CUPUM 2013

เป้าหมาย: เพื่อวิเคราะห์ข้อมูลการเดินทางในรูปแบบต่าง ๆ

ลักษณะข้อมูล: ข้อมูล GPS โดยเก็บข้อมูลทุก ๆ 5 นาที มีการเก็บข้อมูลจาก 100 คนรวมในระยะ
ระยะเวลา 1 เดือน

วิธีการ: โดยขั้นต้นมีการเตรียมข้อมูลดังนี้ Stay point โดยในงานนี้ได้ใช้ Threshold ในการจำแนก
จุดพักดังนี้ เวลา > 14 นาที, รัศมี 196 เมตร

References

ถัดมาคือ Change point ก็คือจะแยก Walk กับ non-walk โดยมีเงื่อนไขคือหากความเร็วน้อยกว่า 100 เมตร/นาที่ ถือว่าเป็นการเดิน การสกัดคุณลักษณะข้อมูลมีดังนี้ ระยะทางรวม, ระยะเวลารวม, ความเร็วต่ำสุด สูงสุด เฉลี่ย ความเร่งสูงสุด อัตราการเปลี่ยนความเร็ว และ เปอเซ็น ของพิกัด GPS อยู่บนถนน หรือ รางรถไฟ (ในประเทศไทยมีการเก็บข้อมูลพิกัดของถนน และ รางรถไฟ) และนำไปทดสอบกับโมเดลที่ใช้ Random Forest

ผลที่ได้: ผลเฉลี่ยความถูกต้องรูปแบบการเดินทางดังนี้ อยู่กับที่, เดิน, ขับรถ, จักรยาน, รถไฟ เป็น 84.17%

References

[5] Xin Gao, Jie Tian, Guiling Wang, “ ” Detection of Transportation Mode Based on Smartphones for Reducing Distracted Driving, MobiCom’14 Proceedings of the 20th annual international conference on Mobile computing and networking, 2014

เป้าหมาย: เพื่อตรวจจับลักษณะการเดินทาง วิ่ง, เดิน, ขับรถ, รถขนส่งมวลชน เพื่อนำไปสร้างแอปพลิเคชัน ล็อคหน้าจอหากมีการขับรถ

ลักษณะข้อมูล: โดยการเก็บข้อมูลคือ 5 นาทีของผู้ใช้แต่ละคน วิ่ง, เดิน ของแต่ละคนจะมีการเก็บแยกเนื่องจากขึ้นงานให้เหตุผลว่าแต่ละคนนั้นมีความเร็วในการเดินและวิ่งที่แตกต่างกัน แต่ในข้อมูลของการขับรถ และ รถขนส่งมวลชน จะมีการใช้ร่วมกันมีการเก็บ GPS , Accelerometer, Magnetometer

References

วิธีการ: ใช้ Threshold เป็นหลักในการแยกเดิน วิ่ง หากความเร็วเกินวิ่ง จะนำมาแยกเปรียบเทียบกับ เส้นทางกับการขับรถยนต์ และ นั่งรถขนส่งมวลชน

ผลที่ได้: สร้างเป็นแอปพลิเคชันล๊อคหน้าจอบ ขณะขับรถ

[6] Angel J. Lopez, Daniel Ochoa, Sidharta Gautama, “Detecting Changes of Transportation Mode by Using Classification Data”, Information Fusion 18th International Conference, 2014

เป้าหมาย: เพื่อวิเคราะห์การเปลี่ยนโหมดในการเดินทาง

ลักษณะข้อมูล: โดยใช้ข้อมูลจาก GHAC(Google Human Activity Classifications) โดยข้อมูลนี้มี GPS, Label รูปแบบการเดินทางเพื่อใช้เป็น Ground truth ในการตรวจสอบความถูกต้อง

References

วิธีการ: ขั้นต้นทำการลบข้อมูลที่ไม่มี label และ ข้อมูลที่ไม่ระบุจุดสิ้นสุด ทำ Sliding windows เพื่อสร้างค่าเฉลี่ย และนำมาสร้างเป็นเวกเตอร์ กล่าวคืออนุมานว่า กิจกรรมที่แตกต่างกันจะมีมุมที่แตกต่างกันออกไป

ผลที่ได้: ความแม่นยำของการเปลี่ยนโหมดคือ 90%

[7] Leon Stenneth, Ouri Wolfson, Philip S. Yu, Bo Zu, “Transportation Mode Detecting using Mobile Phones and GIS Information”, GIS’11 Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems

References

เป้าหมาย: เพื่อวิเคราะห์ข้อมูลการเดินทางในรูปแบบต่าง ๆ แบบมีโครงข่ายของ การขนส่ง (Transportation network คือ ข้อมูลโดยละเอียดของเส้นทาง ไกล่ป้ายรถบัส ทางรถราง)

ลักษณะข้อมูล: ข้อมูล GPS ทุก ๆ 15 วินาทีและมี การทำ Label 3 สัปดาห์ จากกลุ่มผู้ใช้ 9 คน

วิธีการ: โดยสกัดคุณลักษณะจากข้อมูลออกมาดังนี้ ค่าเฉลี่ยความแม่นยำของ GPS, ความเร็วเฉลี่ย, การเปลี่ยนทิศทางเฉลี่ย, ความเร่งเฉลี่ย, ความใกล้ตำแหน่งจอดรถบัส, ความใกล้ราง, อัตราการใกล้จุดจอดรถบัส และมีการนำไปทดสอบกับโมเดล Naïve Bayes, Bayesian Network, Decision Tree, Random Forest, Multilayer perceptron โดยแบ่งเป็น 2 กลุ่มคือ 1. ไม่มี Transportation network 2. มี Transportation network

ผลที่ได้: แบบที่ 1 มีความแม่นยำเฉลี่ย 70% แบบที่ 2 มีความแม่นยำเฉลี่ย 90%

Conclusion

จากการอ่านผลงานที่มีมาก่อนทำให้ได้แนวคิด ในการออกแบบโมเดล การเตรียมข้อมูล รวมไปถึง การเก็บข้อมูล และ การตรวจสอบระหว่างผลจากโมเดล กับ ค่าจริงที่ได้จากการบันทึกมา

แสดงให้เห็นถึงมีแนวคิดวิเคราะห์จากข้อมูลที่คล้ายๆกัน แต่แตกต่างกัน เพื่อหาวิธีที่เหมาะสมกับ สถานที่ กับ สภาพแวดล้อม ในข้อมูลที่จำกัด อย่างใน [7] มีการนำ Transportation network มาใช้ซึ่งเป็นของประเทศสหรัฐอเมริกา ซึ่งได้มาการเก็บข้อมูลแผนที่การเดินทางอย่างละเอียด โดยมีจุดของป้ายรถเมล์ และคิดว่าน่าจะมีค่าอื่น ๆ อีกที่ไม่ถูกกล่าวถึง เมื่อนำข้อมูลดังกล่าวมาประยุกต์ใช้ ด้วยให้ผลความแม่นยำได้มากขึ้นกว่าเดิมมาก