

การระบุตำแหน่งภายในอาคารโดยการวิเคราะห์สัญญาณ WiFi ด้วยเทคโนโลยีการเรียนรู้ของเครื่อง

Indoor Localization through WiFi Fingerprint using Machine Learning

ปฏิพล ศรีชุมสิน และ จันตรี ผลประเสริฐ

บทคัดย่อ

Location-Based Service (LBS) เป็นบริการที่ใช้ประโยชน์จากเทคโนโลยีไร้สายที่ทำให้คนสามารถที่จะระบุตำแหน่งที่อยู่ของผู้ใช้ได้ ผ่านทางโทรศัพท์มือถือ ซึ่งจะให้บริการออกมาในรูปแบบของ Software Application โดยที่เทคโนโลยี LBS จะเป็นการนำตำแหน่งของคนที่อยู่ภายในอาคารมาทำให้เกิดประโยชน์ ในปัจจุบันมีธุรกิจให้ความสนใจว่า LBS ไปใช้งานกันอย่างแพร่หลาย โดยส่วนใหญ่ที่คนนิยมใช้เทคโนโลยี LBS กันมากที่สุดคือการติดตามตำแหน่งของลูกค้าที่มาซื้อสินค้าภายในห้างสรรพสินค้า โดยการที่ผู้ประกอบการจะทำการส่งข้อมูลไปให้กับลูกค้าเพื่อที่จะให้ข้อมูลเพิ่มเติมต่างๆ แต่การที่จะทำให้เทคโนโลยี LBS มีความถูกต้อง จำเป็นต้องอาศัยเทคโนโลยีการระบุตำแหน่งที่มีความแม่นยำซึ่งถ้าภายนอกอาคาร(Outdoor) สามารถที่จะใช้เทคโนโลยี Global Positioning System (GPS)มาใช้ในการระบุตำแหน่งได้ แต่ถ้าเป็นภายในอาคาร จำเป็นต้องมีเทคโนโลยีการระบุตำแหน่งภายในอาคาร (Indoor localization) ที่มีความแม่นยำ งานวิจัยนี้เสนอการใช้เทคนิคการเรียนรู้ของเครื่องจักรในการระบุตำแหน่งของโทรศัพท์มือถือจากสัญญาณ WiFi ภายในอาคาร จากการทดลองพบว่าโมเดล K Nearest Neighbor (KNN) ให้ความแม่นยำมากที่สุดในการทำนาย โดยมีความถูกต้องสูงถึง 97% ในการระบุตำแหน่งในรูปของชั้นในอาคารที่โทรศัพท์มือถืออยู่ อย่างไรก็ตามการระบุพิกัดของโทรศัพท์มือถือในรูปของละติจูดและลองจิจูดยังมีความคลาดเคลื่อนสูงอยู่ซึ่งมีสาเหตุหลักมาจากความแปรปรวนของสัญญาณ WiFi ภายในอาคารซึ่งแนวทางแก้ไขที่เป็นไปได้คือการกรองสัญญาณ WiFi ก่อนนำเข้าโมเดลการเรียนรู้ของเครื่องจักร ซึ่งแนวทางการกรองสัญญาณด้วยเทคโนโลยีทาง signal processing ก็สามารถทำได้หลายรูปแบบเช่นการใช้ Kalman filtering เป็นต้น หรือ การใช้สัญญาณแบบอื่น เช่น สัญญาณ Beacon หรืออุปกรณ์ Radio Frequency Identification (RFID) มาช่วยในการระบุตำแหน่งให้มีความแม่นยำมากขึ้นต่อไป อย่างไรก็ตามการใช้สัญญาณ beacon หรือ RFID นั้นมีข้อเสียที่จะต้องจ่ายค่าจากการติดตั้งอุปกรณ์เหล่านี้เพิ่มเติม ซึ่งต่างจากการใช้สัญญาณ Wifi ในการระบุตำแหน่งของโทรศัพท์มือถือในอาคารที่โดยทั่วไปแล้วจะมีอุปกรณ์ Wifi AP ติดตั้งเพื่อใช้ในการเชื่อมต่อเครือข่ายอินเทอร์เน็ตอยู่แล้ว ดังนั้นประเด็นเหล่านี้จึงต้องนำมาพิจารณาเพิ่มเติมในส่วนของคุณค่าจากการติดตั้งอุปกรณ์เพิ่มเติมเหล่านี้

1. ที่มีความสำคัญ

Location-Based Service (LBS) เป็นบริการที่นำเทคโนโลยีไร้สายมาใช้ประโยชน์เพื่อทำการระบุตำแหน่งที่อยู่ของ Smart Phone ได้อย่างแม่นยำ ซึ่งจะให้บริการออกมาในรูปแบบของ Software Application โดยที่ LBS คือเทคโนโลยีที่ใช้การนำตำแหน่งของคนที่อยู่ภายในอาคารมาต่อยอดให้เกิดประโยชน์ ซึ่งในปัจจุบันมีธุรกิจให้ความสนใจว่า LBS ไปใช้งานกันอย่างแพร่หลาย โดยส่วนใหญ่ที่คนนิยมใช้เทคโนโลยี LBS กันมากที่สุด คือ การติดตามตำแหน่งของลูกค้าที่มาซื้อสินค้าภายในห้างสรรพสินค้า โดยการที่ผู้ประกอบการจะทำการส่งข้อมูลไปให้กับลูกค้าเพื่อที่จะให้ข้อมูลสนับสนุนการค้าเพิ่มเติม และการวิเคราะห์ตำแหน่งการติดป้าย

โฆษณาในห้างสรรพสินค้าเพื่อให้ทราบตำแหน่งใดควรเก็บค่าโฆษณาเท่าใดตามจำนวนผู้คนที่เดินผ่านบริเวณนั้น จะเห็นได้ว่าการระบุตำแหน่งภายในอาคารมีประโยชน์แก่ธุรกิจในปัจจุบันเป็นอย่างมาก ดังนั้น LBS ที่มีประสิทธิภาพจะต้องมีระบบระบุตำแหน่งภายในอาคารที่แม่นยำ โดยสำหรับการระบุตำแหน่งภายนอกอาคาร ซึ่งใช้บริการ Global Navigation Satellite System (GNSS) เช่น GPS ได้ถูกสร้างขึ้นและใช้งานมาแล้วกว่า 20 ปี ซึ่งเป็นสิ่งที่ขาดไม่ได้ในระบบนำทางภายนอกอาคาร อย่างไรก็ตาม ผู้คนใช้เวลาส่วนใหญ่อยู่ภายในอาคาร ซึ่งไม่สามารถใช้บริการ GNSS ได้ เนื่องจากสัญญาณของดาวเทียมไม่สามารถทะลุผ่านสิ่งก่อสร้างได้ และความแม่นยำของ GPS ไม่เพียงพอที่จะระบุตำแหน่งระดับห้องภายในอาคารได้โดยปัจจุบันได้มีการพัฒนาระบบระบุตำแหน่งภายในอาคารขึ้นมาหลายระบบ ซึ่งจำแนกออกได้เป็น 2 ลักษณะ คือ Infrastructure-based Systems และ Infrastructure-free Systems ซึ่งแบ่งตาม ระบบที่ต้องใช้อุปกรณ์ฮาร์ดแวร์ที่ต้องติดตั้งภายในอาคาร โดยอุปกรณ์ฮาร์ดแวร์เหล่านี้ถูกออกแบบมาใช้ในการระบุตำแหน่งภายในอาคารโดยเฉพาะ เช่น การติดตั้งตัวส่งสัญญาณ (Beacon Indoor Tracking) ทำให้มีค่าใช้จ่ายในการติดตั้งสูง ซึ่งต่างกับระบบประเภท Infrastructure-free ซึ่งไม่มีค่าใช้จ่าย โดยจะใช้โครงสร้างเดิมภายในอาคาร เช่น เครือข่ายอินเทอร์เน็ตไร้สาย ในการให้บริการตำแหน่ง อย่างไรก็ตามระบบประเภท Infrastructure-based มักจะให้ความแม่นยำสูงกว่าเนื่องจากมีการใช้อุปกรณ์ฮาร์ดแวร์เพิ่มเติม อย่างไรก็ตามเมื่อ Wireless Local Area Networks (WLANs) และอุปกรณ์เคลื่อนที่ (Mobile Devices) เช่น สมาร์ทโฟน มีการใช้งานอย่างแพร่หลาย สามารถพบการทำงานได้อยู่ทุกที่ ทำให้การระบุตำแหน่งภายในอาคารแบบ Infrastructure-free โดยมีประสิทธิภาพมากขึ้น โดยวิธีที่ได้รับความนิยมมากคือการใช้สัญญาณ WiFi (WiFi Fingerprint) ซึ่งเป็นการเก็บข้อมูลความแรงสัญญาณ WiFi ของ Access point ในบริเวณนั้นมาใช้ในการระบุตำแหน่ง ซึ่งวิธีนี้ไม่ต้องติดตั้งอุปกรณ์ฮาร์ดแวร์เพิ่มเติม เนื่องจากเป็นการใช้โครงสร้างของระบบ WLANs ภายในอาคาร ที่ติดตั้งอยู่แล้ว อย่างไรก็ตาม WLANs ไม่ได้ถูกออกแบบมาเพื่อการระบุตำแหน่ง ดังนั้นอุปสรรคที่สำคัญของวิธีนี้คือ สภาพแวดล้อมภายในอาคาร การกระจายของสัญญาณภายในอาคารยากแก่การทำนาย นอกจากนี้ การระบุตำแหน่งภายในอาคารโดยใช้ WLANs ผู้ใช้จะต้องพกพาโทรศัพท์ไว้กับตัว ซึ่งการเคลื่อนไหว หรือความแตกต่างของรูปแบบการถือสมาร์ตโฟนก็เป็นปัจจัยที่สำคัญ ที่จะทำให้ค่าความแรงของสัญญาณ WiFi แตกต่างกันไป ส่งผลให้การระบุตำแหน่งด้วยวิธี Wifi fingerprint ยังมีความคลาดเคลื่อนข้างสูง

ต่อมามีการใช้การเรียนรู้ของเครื่องจักร (Machine Learning) ซึ่งเป็นอัลกอริทึมที่สามารถเรียนรู้ข้อมูลได้ ถูกใช้ในการแก้ปัญหาประเภท การจดจำรูปแบบ (Pattern Recognition), การแบ่งแยกข้อมูล (Data Classification), การทำนาย (Prediction) และได้มีการนำการเรียนรู้ของเครื่องจักรมาใช้ในการวิเคราะห์ WiFi Fingerprint อย่างไรก็ตาม จะต้องมีการกระบวนการ Data Pre-processing ซึ่งเป็นการลดสิ่งผิดปกติในข้อมูลและเตรียมข้อมูลให้พร้อมก่อนนำไปเรียนรู้ เช่น กระบวนการลบค่าที่ไม่ถูกต้องหรือสูญหาย (Missing Values), การทำให้ข้อมูลเป็น

แบบการแจกแจงปกติ (Normal Distribution), การแปลงข้อมูล (Data transformation) ซึ่งต้องอาศัยการทำ Data Visualization เพื่อให้ทราบถึงปัญหาหรือสิ่งผิดปกติในข้อมูล ทำให้สามารถปรับค่าพารามิเตอร์ (Fine-tuning) ให้เหมาะสมได้ ก่อนนำไปเรียนรู้ด้วยอัลกอริทึมการเรียนรู้ของเครื่องจักรเพื่อเรียนรู้ข้อมูลที่ผ่านมามีงานวิจัยที่ใช้เทคนิคการเรียนรู้ของเครื่องจักรไปใช้ในการวิเคราะห์ความแรงของสัญญาณ WiFi เพื่อระบุตำแหน่งของโทรศัพท์มือถืออยู่มากมาย แต่ผลที่ได้ก็ยังไม่แม่นยำมากนักเนื่องมาจากสาเหตุหลายๆประการ อาทิ เช่น ความไม่แน่นอนของสัญญาณ WiFi สัญญาณรบกวนจากสภาพแวดล้อม เป็นต้น

งานวิจัยนี้นำเสนอการใช้เทคโนโลยีการเรียนรู้ของเครื่องจักรในการวิเคราะห์สัญญาณ WiFi เพื่อทำการระบุตำแหน่งของอุปกรณ์รับสัญญาณภายในพื้นที่ในอาคาร โดยเรศึกษการใช้โมเดลการเรียนรู้ของเครื่องจักรแบบ KNN, Random Forest และ Adaboost ในการวิเคราะห์สัญญาณ WiFi การวัดประสิทธิภาพของโมเดลนั้นใช้ Confusion Matrix และ Accuracy

วัตถุประสงค์ของการวิจัย

1. ศึกษาการใช้เทคโนโลยีการเรียนรู้ของเครื่องจักรมาทำการวิเคราะห์สัญญาณ WiFi เพื่อใช้ในการระบุตำแหน่งในอาคาร
2. เพื่อศึกษาและเปรียบเทียบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องจักรแบบต่างๆในการเอามาประยุกต์ใช้ระบุตำแหน่งโทรศัพท์มือถือ

ขอบเขตของการวิจัย

1. ใช้ข้อมูลสัญญาณ WiFi ในรูปของ Received Signal Strength Indicator (RSSI) ที่วัดได้ภายในอาคารมาระบุตำแหน่งของสัญญาณมือถือ
2. ใช้ UJIIndoor dataset เป็นฐานข้อมูลในการทดสอบประสิทธิภาพของเทคนิคการเรียนรู้ของเครื่องจักร
3. ใช้โมเดลการเรียนรู้ของเครื่องจักรในการวิเคราะห์สัญญาณ WiFi เพื่อระบุตำแหน่งของโทรศัพท์มือถือเป็นชั้นในอาคาร 4. วิเคราะห์ประสิทธิภาพในรูปของ Accuracy และ Confusion Matrix

2. แนวคิด ทฤษฎี และ เทคโนโลยีที่เกี่ยวข้อง

Location Based Service (LBS)

Location-Based Service (LBS) เป็นบริการที่ใช้ประโยชน์จากเทคโนโลยีไร้สาย ที่ทำให้คนสามารถที่จะระบุตำแหน่งที่อยู่ของผู้ใช้ได้ ผ่านทางโทรศัพท์มือถือ ซึ่งจะให้บริการออกมาในรูปของ Software Application โดยที่เทคโนโลยี LBS จะเป็นการนำตำแหน่งของคนที่อยู่ภายในอาคารมาต่อยอดให้เกิดประโยชน์ โดยในปัจจุบันมีองค์กรให้ความสนใจ LBS ไปใช้งานกันอย่างแพร่หลาย โดยส่วนใหญ่ที่คนนิยมใช้เทคโนโลยี LBS กันมากที่สุด คือการติดตามตำแหน่งของลูกค้าที่มาซื้อสินค้าภายในห้างสรรพสินค้า โดยการที่ผู้ประกอบการจะทำการส่งข้อมูลไปให้กับลูกค้า เพื่อที่จะให้ข้อมูลเพิ่มเติมต่างๆ ยกตัวอย่างเช่น Macy ซึ่งเป็นห้างสรรพสินค้าชั้นนำของ New York ได้ทำการ Track การใช้งานเว็บไซต์ของลูกค้า เมื่อลูกค้ารายหนึ่งแสดงพฤติกรรมชัดเจนว่ามีความสนใจในรองเท้า Brand Name คู่หนึ่งแต่กลับไม่ได้ทำการซื้อผ่านเว็บไซต์ ต่อมาลูกค้ารายนั้นได้เดินทางเข้าไปในแผนกรองเท้าของห้างสรรพสินค้า Macy ซึ่ง Macy สามารถทำการ Track ลูกค้าได้ผ่านโทรศัพท์มือถือจากแผนกรองเท้าภายในห้างสรรพสินค้าของ Macy ซึ่งมันจะสามารถนำเสนอส่วนลดในรองเท้าคู่ที่ลูกค้าสนใจอย่างอัตโนมัติ ผ่านทาง SMS หรือผ่านคำแนะนำของพนักงานในร้านค้า เพื่อกระตุ้นความสนใจ ให้ลูกค้าซื้อรองเท้าคู่นั้นในที่สุด ซึ่งวิธีการคือ ถ้าลูกค้าสนใจที่จะซื้อรองเท้า และผู้ประกอบการสามารถที่จะใช้เทคโนโลยี LBS ติดตามตำแหน่งของลูกค้าได้ โดยผู้ประกอบการจะทำการส่งข้อมูล

รองเท้าไปให้กับลูกค้าผ่านช่องทางต่างๆ เพื่อที่จะเป็นการกระตุ้นให้ลูกค้าอยากที่จะซื้อสินค้าเพิ่มขึ้น อย่างไรก็ตามการที่จะนำเทคโนโลยี LBS มาใช้งาน จำเป็นต้องมีเทคโนโลยีการระบุตำแหน่งที่แม่นยำ ซึ่งถ้าภายนอกอาคาร (Outdoor) สามารถที่จะใช้เทคโนโลยี GPS มาใช้ในการระบุตำแหน่งได้ แต่ถ้าเป็นภายในอาคาร จำเป็นต้องมีเทคโนโลยีการระบุตำแหน่งภายในอาคาร (Indoor localization) ที่มีความแม่นยำ ซึ่งเป็นปัญหาที่ค่อนข้างท้าทาย ว่าทำอย่างไรให้การระบุตำแหน่งนั้นมีความถูกต้องแม่นยำ และมีประสิทธิภาพมากที่สุด

เทคโนโลยี Wi-Fi

Wi-Fi ย่อมาจาก Wireless Fidelity เป็นเทคโนโลยีเครือข่ายไร้สาย ภายใต้เทคโนโลยีการสื่อสารตามมาตรฐาน IEEE 802.11 โดยเป็นมาตรฐานที่ถูกอนุมัติให้ใช้จาก IEEE (The Institute of Electrical and Electronics Engineers) เพื่อให้อุปกรณ์คอมพิวเตอร์สามารถสื่อสารกันได้บนมาตรฐานการทำงานแบบเดียวกัน สำหรับเทคโนโลยีเครือข่ายไร้สายนี้จะใช้คลื่นความถี่ Radio Frequency (RF) และคลื่นความถี่อื่นพาเรตในการรับส่งข้อมูล คลื่นความถี่วิทยุของเครือข่ายไร้สายจึงสามารถทะลุทะลวงกำแพง หรือสิ่งกีดขวางได้ทำให้การใช้งานบนเครือข่ายไร้สายมีความคล่องตัวและสะดวกสบายมากขึ้น โดยสามารถเชื่อมต่อเข้าสู่เครือข่ายได้ทุกที่มีสัญญาณ ข้อมูลจะถูกส่งผ่านคลื่นวิทยุความถี่ 2.4 GHz หรือ 5 GHz ในปัจจุบันนิยมใช้ Wi-Fi เพื่อต่อกับอินเทอร์เน็ต โดยอุปกรณ์พกพาต่างๆ เช่น คอมพิวเตอร์ส่วนบุคคล, เครื่องเล่นเกม, โทรศัพท์มือถือ, แท็บเล็ต เป็นต้น โดยที่สามารถเชื่อมต่ออินเทอร์เน็ตได้ผ่านอุปกรณ์ที่เรียกว่าแอคเซสพอยต์ (Access Point) หรือ ฮอตสปอต (Hotspot) ซึ่งเทคโนโลยี Wi-Fi ก็ยังคงมีข้อจำกัดในการใช้งานอยู่หลายประการ ไม่ว่าจะเป็นเรื่องของความปลอดภัย การใช้งานร่วมกันของอุปกรณ์ต่างๆ และความยากลำบากในการทำให้กับผู้ให้บริการเครือข่าย อย่างไรก็ตาม อัตราเร็วในการรับส่งข้อมูลผ่านอุปกรณ์ Access Point ของผู้ใช้งานแต่ละคนอาจมีค่าไม่เท่ากัน ขึ้นอยู่กับสภาพแวดล้อมในการใช้งาน และจำนวนผู้ใช้งานที่แบ่งกันรับส่งข้อมูลผ่านอุปกรณ์ Access Point ร่วมกัน นอกจากนั้นยังขึ้นอยู่กับรูปแบบในการรับส่งข้อมูลของแต่ละคนอีกด้วย

เนื่องจากสัญญาณ Wi-Fi เป็นสัญญาณที่มีอยู่แล้วภายในอาคาร เราสามารถนำอุปกรณ์ Access Point ที่มีอยู่แล้วมาใช้ประโยชน์ โดยการที่นำมาช่วยในการระบุตำแหน่งภายในอาคาร (Indoor Localization) วิธีการคือ ใช้ค่าความแรงของสัญญาณ Wi-Fi (RSSI) ในการที่จะเชื่อมโยงเข้ากับตำแหน่งของคนที่อยู่ภายในอาคารได้ ซึ่งวิธีการเก็บข้อมูลที่นิยมใช้กันก็คือ การเก็บข้อมูลแบบ Fingerprinting เพราะแต่ละตำแหน่ง สัญญาณ Wi-Fi มีความแรงของสัญญาณที่แตกต่างกัน ดังนั้นการใช้สัญญาณ Wi-Fi โดยการเก็บข้อมูลแบบ Fingerprinting จึงสามารถที่จะนำมาใช้ในการระบุตำแหน่งได้

Fingerprinting

Fingerprinting ได้รับความนิยมเป็นอย่างมากในการนำมาใช้เพื่อระบุตำแหน่งภายในอาคาร โดยหลักการทำงานของมันเป็นคือ การนำค่าความแรงของสัญญาณ Wi-Fi (RSSI of Wi-Fi) ที่วัดได้ ณ ตำแหน่งของผู้ใช้งานอยู่ขณะนั้น มาทำการเปรียบเทียบกับฐานข้อมูลที่มีการวัดค่า RSSI ไว้ก่อนหน้านี้แล้ว เพื่อที่จะทำการระบุตำแหน่งให้ตรงกับพิกัดเป้าหมายได้อย่างถูกต้องและเกิดความคลาดเคลื่อนน้อยที่สุด ข้อดีอย่างหนึ่งของวิธีการนี้คือ ไม่จำเป็นต้องใช้อุปกรณ์ที่มีราคาสูงซึ่งเหมาะสมกับการใช้งานภายในอาคาร Fingerprinting เป็นเทคนิคหนึ่งที่ใช้ในการระบุตำแหน่งซึ่งมีลักษณะการทำงานแบ่งออกเป็น 2 ขั้นตอน คือ

ระยะออฟไลน์ (Offline Phase) เป็นระยะเก็บข้อมูล

วิธีการคือ จะทำการกำหนดตำแหน่ง (Grid Point) ให้ครอบคลุมพื้นที่ที่เราต้องการจะทำการระบุตำแหน่ง จากนั้นทำการเก็บค่าข้อมูลสัญญาณ RSSI

จากหลายๆ Access Point ในบริเวณที่เราต้องการ ซึ่งเทคนิค Fingerprinting [1],[2] ข้อดีคือ ได้ค่าที่เป็นไปได้ใกล้เคียงกับการเก็บข้อมูลจริง แต่ข้อเสียคือ ใช้เวลาในการเก็บข้อมูลมาก

ระยะออนไลน์ (Online Phase) เป็นระยะที่ใช้ในการทดสอบ

วิธีการคือ ทำการวัดค่าสัญญาณ RSSI ที่ตำแหน่ง ณ ขณะนั้นแล้วนำค่าที่ได้มาทำการเปรียบเทียบกับค่าที่เราเก็บไว้ในระยะเก็บข้อมูล เมื่อทำการเปรียบเทียบแล้วจะได้ตำแหน่งของผู้ใช้ออกมา ข้อเสียของวิธีนี้คือ เมื่อสภาพแวดล้อมของสถานที่เปลี่ยนไป เช่น การเคลื่อนย้ายเฟอร์นิเจอร์ภายในห้องจะทำให้ Fingerprint ของตำแหน่งนั้นเปลี่ยนไป ซึ่งจำเป็นต้องมีการเก็บข้อมูลใหม่ทุกครั้ง และมีความไม่แน่นอนของสัญญาณ เช่น เมื่อมีผู้คนเดินในพื้นที่จะทำให้กระทบต่อความแรงสัญญาณ ทำให้ข้อมูลที่เก็บได้ อาจไม่แม่นยำเมื่อนำมาใช้งานจริง

Machine Learning

การเรียนรู้ด้วยเครื่องจักร หรือ Machine Learning คือศาสตร์ที่ยึดหยุ่นและสามารถปรับใช้ให้เข้ากับลักษณะของชุดข้อมูลที่น่าวิเคราะห์และสร้างกระบวนการทำนายผลตามลักษณะโมเดลซึ่งข้อมูลที่จะนำไปทำนายผลตามจุดประสงค์ในการวิเคราะห์ข้อมูลก็ต้องถูกปรับแต่งให้ตรงคุณลักษณะตามเป้าหมายนั้นก่อนด้วย เช่น การวิเคราะห์ข้อมูลหุ้นเพื่อทำนายผลกระทบทางเศรษฐกิจ การวิเคราะห์ข้อมูลยอดlike ของโฆษณาเพื่อกำหนดทิศทางการตลาดจากภาพที่3กระบวนการเรียนรู้ด้วยเครื่องจักรเป็นการศึกษาความสัมพันธ์ของข้อมูลแล้วนำไปเข้ากระบวนการคัดกรองตามแต่ละโมเดลที่มีคุณสมบัติเฉพาะเพื่อทำการวิเคราะห์แล้วส่งผลวิเคราะห์กลับมาใช้ตามวัตถุประสงค์ที่ตั้งไว้ซึ่งการวิเคราะห์นั้นไม่จำเป็นต้องทำแค่ครั้งเดียว โดยผลการวิเคราะห์ที่ได้สามารถบอกความเป็นไปได้ตามวัตถุประสงค์ที่เราตั้งไว้ในการทดสอบซึ่งมันจะชี้ทิศทางของกระบวนการดำเนินงานในขั้นตอนต่อไปได้นั่นเอง เริ่มต้นที่ข้อมูลที่จะนำมาใช้ทำการเรียนรู้ด้วยเครื่องจักรนั้นเป็นข้อมูลที่เก็บรวบรวมมาโดยมีวัตถุประสงค์ในการรวบรวมไว้ล่วงหน้าก่อนที่ข้อมูลจะเกิดขึ้นจริง หรือเป็นข้อมูลที่เกิดขึ้นจากเหตุการณ์จริงที่ผ่านมาแล้ว โดยหากเป็นข้อมูลที่เกิดจากเหตุการณ์จริงที่ผ่านมาแล้วก็ต้องดูว่าข้อมูลถูกแบ่งออกเป็นกลุ่มหรือ cluster หรือเปล่า เช่น ข้อมูลนักเรียนชาย นักเรียนหญิง ถ้าข้อมูลนั้นสามารถจำแนกเป็นกลุ่มได้ชัดเจนก็นำไปสู่กระบวนการหาค่าเฉลี่ยรวม หรือ K-Means โดย K คือจำนวนกลุ่มที่นำมาคิดค่าเฉลี่ยรวมแต่หากข้อมูลยังไม่ถูกแบ่งกลุ่มก็ต้องทำการแบ่งกลุ่มตามลำดับขั้นของข้อมูลถ้าข้อมูลที่จะนำมาเรียนรู้ด้วยเครื่องจักรนั้นเป็นข้อมูลที่รวบรวมเพื่อวัตถุประสงค์เฉพาะที่บ่งบอกข้อมูลทาง ปริมาณ และ คุณภาพ ได้ใหม่ โดยหากไม่มีข้อมูลส่วนที่สามารถบ่งบอกทางปริมาณและคุณภาพได้ก็ต้องดูต่อว่าชุดข้อมูลมีคุณลักษณะเชิงเส้นหรือ linear form ด้วยหรือเปล่าซึ่งหากไม่มีก็ต้องดูว่าข้อมูลมีคุณลักษณะทางความน่าจะเป็นหรือ probability distribution หรือเปล่าโดยหากไม่มีก็หมายความว่าข้อมูลแต่ละrecordนั้นเป็นอิสระต่อกันดังนั้นจะใช้โมเดล KNN แต่ถ้าข้อมูลมีคุณสมบัติทางความน่าจะเป็นก็จะใช้โมเดล Naive Bays classification ส่วนกรณีชุดข้อมูลมีคุณสมบัติเชิงเส้นด้วยก็จะใช้โมเดล Logistic regression แต่หากข้อมูลมีคุณลักษณะทาง ปริมาณ และ คุณภาพ ก็เป็นปัญหาจำพวก Regression โดยต้องดูต่อว่าข้อมูลมีความสัมพันธ์เชิงเส้นระหว่างกันไหมโดยหากมีก็ใช้โมเดล Linear regression แต่หากไม่มีก็ใช้ โมเดล Generalized Linear model หรือ GLM และหลังจากได้เรียนรู้รูปแบบการวิเคราะห์ชุดข้อมูลก่อนจะนำชุดข้อมูลไปทำการวิเคราะห์ด้วยการเรียนรู้ด้วยเครื่องจักรแล้วก็ต้องมาเรียนรู้ถึงความแตกต่างของชุดข้อมูลแต่ละแบบ การเรียนรู้ของเครื่องจักรสามารถแบ่งแบบคร่าวๆ ได้สองแบบคือ

1.Supervised Learning เทคนิคการเรียนรู้ของเครื่องประเภทการเรียนรู้แบบมีผู้สอน มีหลักการเรียนรู้จากข้อมูลนำเข้าที่จัดเตรียมไว้ก่อนหน้า ซึ่งประกอบด้วยข้อมูล

คุณลักษณะ(Feature)และข้อมูลคำตอบ (Label) จากนั้นจึงทำการทำนายผลลัพธ์ (Output Data) จากสิ่งที่ได้เรียนรู้ด้วยลักษณะ ในบางครั้งการทำนายนี้ทำให้ในขั้นตอนการเรียนรู้ต้องอาศัยข้อมูลจำนวนค่อนข้างมากเพื่อให้ได้ผลการทำนายที่แม่นยำมากขึ้น เทคนิคการเรียนรู้แบบมีผู้สอนที่นิยม ได้แก่ การจำแนกประเภท (Classification) และการถดถอย (Regression) โดยข้อมูลที่ใช้สำหรับการเรียนรู้จะต้องมีคุณสมบัติที่สนับสนุนการเรียนรู้และการแก้ปัญหาเรียกว่า Feature Vector ในขั้นตอนการเรียนรู้ของเครื่อง (Training) นั้นจะดำเนินการโดยใช้อัลกอริทึม (Algorithm)ที่เหมาะสมกับลักษณะของปัญหาที่ต้องการแก้ไข จนกระทั่งได้โมเดลที่ดีที่สุดสำหรับสร้างเป็นแบบจำลองดังแสดง หลังจากนั้นขั้นตอนการเรียนรู้ แบบจำลองที่ได้จะถูกนำมาประยุกต์ใช้งาน โดยการนำชุดข้อมูลอีกหนึ่งชุด เรียกว่า Test data เข้าสู่แบบจำลอง เพื่อทำนายผลลัพธ์

2.Unsupervised Learning เทคนิคการเรียนรู้ของเครื่องประเภทการเรียนรู้แบบไม่มีผู้สอน มีหลักการเรียนรู้จากข้อมูลนำเข้าที่มีเพียงข้อมูลคุณลักษณะเท่านั้น โดยการเรียนรู้และการทำนายผลนั้นจะอาศัยการจำแนกและสร้างรูปแบบจากข้อมูลนำเข้า โดยจะไม่มีการกำหนดผลลัพธ์ที่ได้ล่วงหน้า เทคนิคที่นิยม ได้แก่การลดมิติของข้อมูล (Dimensionality Reduction) และการจัดกลุ่ม(Clustering)

Classification

เป็นเทคนิคหนึ่งของการเรียนรู้ของเครื่องแบบมีผู้สอน โดยกระบวนการเรียนรู้นั้นจำเป็นต้องมีข้อมูลคำตอบ ซึ่งจะแบ่งข้อมูลคำตอบออกเป็น Class ต่างๆ สำหรับหลักการการทำงานของแบบจำลองจะแบ่งข้อมูลออกเป็น 2 ส่วน คือชุดข้อมูลการเรียนรู้ และชุดข้อมูลทดสอบซึ่งในงานวิจัยนี้แบ่งข้อมูลดังกล่าวในอัตราส่วน 80 : 20 หรือตามความเหมาะสม จากนั้นนำข้อมูลชุดการเรียนรู้ป้อนเข้าสู่ระบบเพื่อเข้าสู่ขั้นตอนการเรียนรู้โดยใช้อัลกอริทึมที่กำหนด จนกระทั่งได้ผลลัพธ์ออกมาเป็นแบบจำลองการทำนาย จากนั้นนำข้อมูลชุดทดสอบซึ่งเป็นข้อมูลที่เตรียมไว้สำหรับทดสอบประสิทธิภาพการทำงานของแบบจำลองมาใช้ในการทำนายผลและพิจารณาผลลัพธ์ที่ได้

รายละเอียดแบบจำลองที่เราจะใช้จำแนก

1. Random Forest เป็นการแบ่งข้อมูลออกจากกันเป็นกลุ่มโดยจับกลุ่มข้อมูลตามคุณลักษณะที่เกี่ยวข้องเนื่องกันไว้ด้วยกันโดยใช้เทคนิค Decision Tree ตามจำนวนกลุ่มที่ต้องการแบ่งแยกโดย เอาข้อมูลที่มีคุณลักษณะคล้ายกันหรือเหมือนกันมารวมกัน และหากมีต้นไม้มากๆ ต้นจะเรียกว่าป่า หรือ Forest โดยการทำการสุ่มค่า ณ จุดที่ Decision Tree มีความลึกหรือความสูงมากที่สุดจะเรียกว่าค่า K หรือ node สีแดงดังรูป K = 3 แล้วนำมาหาผลรวมแล้วเอาค่าที่สุ่มมาเปรียบเทียบเพื่อหาค่าเฉลี่ยที่ใกล้เคียงแล้วระบุชุดข้อมูลจากต้นไม้นั้นที่มีค่าเฉลี่ยใกล้เคียงค่าเฉลี่ยมากที่สุดนั้นคือสาเหตุทำให้ไม่จึงเรียก Random Forest เป็นการจำแนกข้อมูลออกเป็นจำนวน n ชุด โดยแต่ละชุดข้อมูลที่จำแนกออกมาได้อยู่ในรูป Decision Tree แล้วจะนำค่าความลึกที่ K ในแต่ละต้นไม้มารวมกันแล้วหาค่าเฉลี่ยก่อนนำไปเข้าสู่กระบวนการทำนายผล
2. K Nearest Neighbors (KNN) เป็นอัลกอริทึมที่พัฒนาโดยมีคุณลักษณะตรงตามตัวแบบจาก Euclidean Distance formula โดยจะทำการค้นหาข้อมูลตัวที่อยู่ใกล้กับข้อมูลหลักที่เข้าเทียบเคียงมากที่สุดออกมาโดยจะมีการค้นหาจำนวนข้อมูลเป็นจำนวน K ข้อมูล โดยจะเพิ่มระยะการค้นหาหากขึ้นๆจนกว่าจะครบจำนวน K ข้อมูล จากนั้นจะระบุกลุ่มข้อมูลที่ต้องการที่อยู่ใกล้ตัวเทียบเคียงออกมา เพราะบางครั้งอาจเจอข้อมูลที่ไม่สามารถจัดเข้ากลุ่มข้อมูลที่ต้องการได้ เพราะมีคุณลักษณะที่แตกต่างไปจากคุณลักษณะของข้อมูลที่เราจะมาส่วนใหญ่

3. Adaboost เป็น Algorithm ที่พัฒนาจากแนวคิดที่เกี่ยวกับการนำความน่าจะเป็นมาใช้เป็นตัวชี้วัดโดยจะพยายามทำการเพิ่มค่าของความน่าจะเป็นของชุดข้อมูลที่มีค่าความน่าจะเป็นน้อยให้มีค่าเพิ่มขึ้นแล้วค่อยทำการ classification โดย หากสามารถทำการเพิ่มค่าความน่าจะเป็นของชุดข้อมูลที่น่ามาทำนายค่าได้แล้วจึงทำการ classification ต่ออีกในกรณี ที่ความน่าจะเป็นที่หาออกมาได้มีค่าสูงกว่าค่าความน่าจะเป็นรอบก่อนหน้า และ หยุด ในกรณีที่ค่าความน่าจะเป็นที่หาออกมาได้มีค่าต่ำกว่าค่าความน่าจะเป็นที่หาได้ในรอบก่อนหน้าและนำผลที่ได้ไปทำการทำนายต่อไป

แนวคิดค่าน้ำหนักหรือค่าความน่าจะเป็นที่ใช้ในการชี้วัดข้อมูลของ AdaBoost

1. สมการที่ 1

$$f(x) = \sum_{t=1}^T t (\alpha_t h_t(x)) = \text{strong model (1)}$$

= ผลรวมของผลคูณของน้ำหนักกับ weak model(h)

ที่ชุดข้อมูล X ณ ข้อมูลที่ t

โดยหลังจาก classifier แล้วหากยังได้ weak model ก็จะทำการเพิ่มน้ำหนักหรือค่าสมการที่ 2 โดยที่

$$\alpha_t = \left(\frac{1}{2}\right) \ln(\epsilon_t)$$

ϵ_t คือ Error ณ ข้อมูลที่ t, α_t คือค่าน้ำหนักของข้อมูล ณ เวลาที่ t

งานวิจัยที่เกี่ยวข้อง

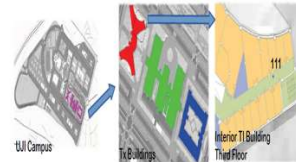
จากการที่ได้ศึกษางานวิจัยที่เกี่ยวข้องกับกระบวนการระบุตำแหน่งภายในอาคาร ผู้จัดทำจะขอยกตัวอย่างมาบางส่วน เพื่อศึกษาและนำมาปรับใช้กับโครงงานได้ดังต่อไปนี้

1. UJIIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint-based Indoor Localization Problems ได้เสนอและจัดทำฐานข้อมูล UJIIndoorLoc เพื่อให้เป็นฐานข้อมูลสาธารณะที่ใช้งานวิจัยทางด้านกระบวนการระบุตำแหน่งภายในอาคาร โดยใช้ WiFi Fingerprint และเพื่อเป็นฐานข้อมูลกลางในการเปรียบเทียบประสิทธิภาพระหว่างงานวิจัยต่างๆ
2. Low-effort place recognition with WiFi fingerprints using deep learning ได้เสนอโครงสร้างของ DNN ซึ่งประกอบด้วย Stacked Autoencoder และ Classifier เพื่อใช้ในการแบ่งแยกอาคารและชั้น ซึ่งมีนพุด คือ WiFi Fingerprint จำนวน 520 ค่า ซึ่งจะถูกลอยข้อมูลเหลือ 64 ค่าและจะถูกส่งต่อไปยัง Classifier เพื่อแบ่งแยกชั้นและอาคาร ผลการเปรียบเทียบประสิทธิภาพระหว่างโครงสร้างแบบต่างๆ จะพบว่าการใช้ SAE จะช่วยให้มีผล Testing Accuracy ดีขึ้น และที่โครงสร้างที่ใช้ SAE (256-128-64) + Classifier (128-128) จะค่า Testing Accuracy สูงที่สุด คือ ประมาณ 0.91 โดยงานนี้จะเป็นการแบ่งแยกอาคารและชั้นเท่านั้น ซึ่งโครงสร้างของ DNN ที่ใช้ จะไม่มีผลลัพธ์เป็นพิกัดตำแหน่ง
3. A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting ได้เสนอโครงสร้างของ DNN ซึ่งประกอบด้วย Autoencoder และ Classifier เพื่อใช้ในการแบ่งแยกอาคารชั้น และพิกัดตำแหน่ง (Latitude , Longitude) โดยในส่วน of Classifier ใช้ Loss Function คือ Binary Crossentropy ซึ่งใช้ในการทำ Multi-label Classifications และมีการใช้ Adam Optimizer ซึ่งเป็นวิธีที่มีประสิทธิภาพในการปรับค่า Weight ของ DNN เพื่อให้ค่าตอบดูเข้าค่าจริงมากที่สุด ในโครงสร้างนี้ใช้อัตรา Dropout ที่ 0.2 เพื่อป้องกันไม่ให้ DNN เกิด Overfitting
4. งานวิจัยเรื่อง User Localization in Complex Environments by Multimodal Combination of GPS, Wi-Fi, RFID and Pedometer Technologies ได้เสนอแนวคิดในการใช้เทคโนโลยี Multimodal (GPS, RFID, Pedometer) กับการใช้ Wi-Fi เพียงอย่างเดียวเพื่อทำการเปรียบเทียบประสิทธิภาพของการระบุตำแหน่ง โดยใช้วิธีการเก็บข้อมูลแบบ Grid Points เป็นหลัก ซึ่งผลการทดลองพบว่าความ

คลาดเคลื่อนเฉลี่ยของ Multimodal อยู่ที่ประมาณ 2.18 เมตร ส่วน Wi-Fi only ค่าความคลาดเคลื่อนเฉลี่ยอยู่ที่ประมาณ 9.99 เมตรดังนั้นจากผลการทดลองดังกล่าวสามารถสรุปได้ว่าการใช้เทคโนโลยีแบบ Multimodal มีความคลาดเคลื่อนต่ำกว่าการใช้ Wi-Fi เพียงอย่างเดียว

3. การพัฒนางานวิจัย

ขั้นตอนเตรียม Dataset ให้พร้อมใช้งาน



ภาพประกอบ 1 แสดงพื้นที่ 3 อาคารของ University Jaume I, Spain ที่มา : [ออนไลน์]

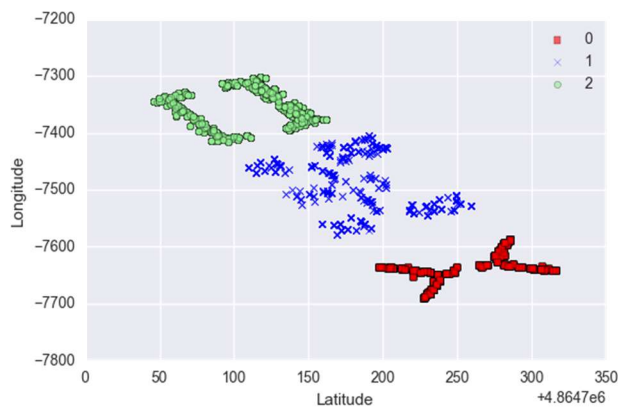
<http://ieeexplore.ieee.org/document/7275492>

UJIIndoor เป็นฐานข้อมูลสาธารณะดังรูปที่ 1 ประกอบด้วย 3 อาคาร แต่ละอาคารมี 4-5 ชั้น และมีจำนวนข้อมูล Training (90%), Validation (10%) : 19,937 (ไฟล์ train.csv) ฐานข้อมูลนี้เก็บข้อมูลโดยผู้ใช้งานจำนวน 20 คน โดยใช้สมาร์ตโฟนจำนวน 25 เครื่องในการเก็บค่า Fingerprint ซึ่งเป็นความแรงสัญญาณ WiFi จาก Access Point ในแต่ละตำแหน่งภายในแต่ละอาคาร และมีข้อมูลเป็น Sparse Data เนื่องจาก Fingerprint ส่วนใหญ่มีข้อมูล RSSI ของ Access Point ประมาณ 20 ค่า จาก 520 APs ซึ่งสามารถใช้ Stacked Autoencoder เพื่อใช้ในการลดเวลาในการวิเคราะห์ข้อมูลได้ ข้อมูลใน UJIIndoorLoc Dataset ประกอบด้วยค่า RSSI ของ WiFi Access Point ตั้งแต่ WAP1 จนถึง WAP520 ซึ่งเป็นค่าความแรงสัญญาณ WiFi มีหน่วยเป็น dB โดยที่ค่า 100 หมายถึงไม่สามารถรับสัญญาณของ Access Point นั้นได้ ค่า Longitude และ Latitude คือค่าพิกัดตำแหน่งของตำแหน่งที่เก็บข้อมูล ค่า SpaceID คือ ค่ารหัสตำแหน่งซึ่งจะมีการจากแบ่งพื้นที่ภายในชั้นนั้นๆ ค่า Floor คือเลขระบุแต่ละชั้น มีค่าตั้งแต่ 0 ถึง 4 ตามแต่ละอาคาร ค่า Building คือเลขระบุอาคาร ซึ่งอาคารของ University Jaume I ที่ใช้สร้างฐานข้อมูลนี้มี 3 อาคาร ดังนั้นจะใช้ค่า 0 ถึง 2 ในการระบุแต่ละอาคาร โดยข้อมูลใน dataset เป็นข้อมูลการเข้าถึง Internet ผ่านตัว WAPs ที่ติดตั้งอยู่ในระบบ Infrastructure โดยเอาข้อมูลการ access เข้าสู่อินเทอร์เน็ตของโทรศัพท์มือถือ ณ เวลาที่เคลื่อนที่ผ่าน WAP แต่ละจุดมาใช้ โดย WAPs ถูกติดตั้งไว้ในระบบ Infrastructure ที่ติดตั้งอยู่ในตึกสามตึกใน มหาวิทยาลัย Jaume ในประเทศ Spain โดยแต่ละตึกมีจำนวนชั้นอย่างน้อย 4 ชั้น มีพื้นที่อย่างน้อย 110.000 ตารางเมตร. โดย ข้อมูลที่สามารถใช้ในการ เอามาทำ classification เช่น อาคารและชั้น ที่จับสัญญาณได้ , ข้อมูลที่เอามาทำ regression เช่น ลองจิจูด และ ละติจูด ที่จับสัญญาณได้ ข้อมูล ชุดนี้ถูกการบันทึกเมื่อปี 2013 โดย ใช้ผู้เก็บข้อมูล อย่างน้อย 20 คนและโทรศัพท์มือถือ smartphone ชนิด OS แบบ Android อย่างน้อย 25 เครื่อง. โดยข้อมูลที่ เก็บได้ มีจำนวน 19937 records (trainingData.csv file) และ 1111 test records (validationData.csv file). โดยในหนึ่ง record ประกอบด้วย 529 attributes เช่น WiFi fingerprint, จุดที่จับสัญญาณได้ และข้อมูล อุปกรณ์ที่ใช้ทดลอง และ ตำแหน่ง WAP (Wireless Access Point หรือตัวกระจายสัญญาณ) โดย ชื่อคอลัมน์จะเป็นชื่อของตัวกระจายสัญญาณไวไฟจาก ตัวที่หนึ่ง ถึง ตัวที่ห้าร้อยสี่สิบเก้า ส่วนเลขแถวเป็นตัวเลขลำดับที่ของการวิ่งของข้อมูลทีผ่านระบบเครือข่ายสัญญาณไวไฟ โดยจากงานวิจัยจะทำการ Cleaning dataset โดย แบ่งเป็น train และ test โดยใช้ column ที่ 0 ถึง 520 เป็น train และ 521 ถึง 526 เป็น test และแปลงค่าใน record ที่ access point มีค่าเดิมเป็น 100 ให้เป็น NAN เพราะตั้งค่า 100 ไว้เป็นค่า default กรณีไม่มีโทรศัพท์มือถือเครื่องใด access ผ่าน WAP ดังนั้น และเมื่อได้ข้อมูลชุด train และ test แล้วก็ยังไม่สามารถนำข้อมูลไปใช้ใน Machine Learning Model ได้ในทันทีเนื่องจากข้อมูลที่ได้มามีค่าความ error อยู่สูงจึงต้อง

ทำการปรับแต่งข้อมูลตามขั้นตอนทางสถิติ ก่อนโดยทั้ง การกำจัดข้อมูล ที่ไม่ได้อยู่ในลักษณะ Normalization โดยใช้กระบวนการ skewness เพื่อทำการกำจัดข้อมูล ที่ไม่ได้ อยู่ในขอบเขตNormalization ออกไป และ กระบวนการ kurtosis เพื่อกำจัด ข้อมูลที่มีค่า ที่สูงหรือ ต่ำใดง ผิดปกติออกไป จากนั้นจึงเข้ากระบวนการ box cox อีกครั้งเพื่อกำจัดข้อมูล ส่วนที่มี จำนวน ซ้ำกันมากๆ ออกไป จากนั้นจึงได้ข้อมูลพร้อมใช้ แล้วทำการแบ่งข้อมูลออกเป็น 6 files แล้วจึงนำเข้าสู่ Machine Learning Model เป็นการต่อไป

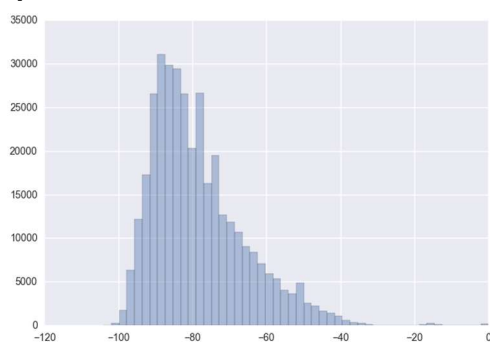
การทำ Exploratory Data Analysis และ Pre-processing

ในการวิเคราะห์ข้อมูลเบื้องต้นเราได้ทำการวิเคราะห์ข้อมูลความแรงสัญญาณไวไฟ เทียบกับตำแหน่งละติจูดและลองจิจูดดังแสดงในรูปที่ 2



ภาพประกอบ 2 การกระจายของชุดข้อมูลการ access ของตัว WAPs

จากภาพที่ 3 เป็นการแสดงให้เห็นถึงการกระจายกันของข้อมูลการ access ของ WAPs ในแต่ละอาคาร โดยใช้ Latitude และ Longitude เป็นตัวชี้วัด โดยแสดงเป็น สีเหลี่ยม สีแดงสำหรับอาคารแรก และ กากบาทสีน้ำเงินสำหรับอาคารที่สอง และ วงกลมสีเขียว สำหรับอาคารที่สาม ซึ่งจากรูปจะเห็นได้ว่าความแรงของสัญญาณไวไฟของแต่ละตึกมีการแยกกันตามแต่ละตึกอย่างชัดเจน ไม่มีสัญญาณที่กระจายข้ามอาคารมาได้ ต่อมาเมื่อทำการวิเคราะห์การกระจายตัวของข้อมูล (RSSI) ดังแสดงในรูปที่ 3 พบว่าข้อมูลส่วนใหญ่จะกระจายไปอยู่ทางซ้ายของการกระจายของข้อมูล (right-skewed) มีการกระจายตัวไม่เป็น normal distribution



ภาพประกอบ 3 การกระจายตัวของข้อมูล RSSI

ในส่วนของการหาโมเดลเพื่อวิเคราะห์ตำแหน่งของ smart phone จากค่า WiFi RSSI นั้น เราจะทำการทดสอบประสิทธิภาพของโมเดลการเรียนรู้ของเครื่องจักรทั้งหมด 3 แบบคือ KNN, Random Forest และ Adaboost โมเดล โดยจะทดสอบประสิทธิภาพของทั้ง 3 โมเดลนี้ในรูปของ Confusion matrix และ accuracy

4. ผลการวิจัย

ผู้วิจัยได้ทำการทดลองหาประสิทธิภาพของโมเดลการเรียนรู้ของเครื่องจักรหลายๆแบบในการระบุตำแหน่งของโทรศัพท์มือถือจากสัญญาณ WiFi โดยใช้ชุดข้อมูล UJIindoor โดยวัดประสิทธิภาพในรูปของ Accuracy และ Confusion Matrix ของชุดข้อมูลแยกตามแต่ละ BUILDINGID ซึ่งยืนยันไปในทิศทางเดียวกับ

งานวิจัย[1]และผลของงานวิจัยนี้ได้กล่าวถึงในย่อหน้าก่อนหน้านี้ โดย

BUILDINGID=0 ด้วย model Random Forest มี ค่า Accuracy เท่ากับ 0.7874, และ Confusion Matrix ของตึก 0 ของโมเดล Random Forest, adaBoost และ KNN ดังแสดงในภาพประกอบ 4, 5 และ 6 ตามลำดับ

			PREDICTED LABEL				
			Floor 0	Floor 1	Floor 2	Floor 3	Floor 4
A	L						
C	A	Floor 0	108	102	1	0	N
T	B	Floor 1	0	251	14	1	N
U	E	Floor 2	0	0	290	1	N
A	L	Floor 3	0	0	0	282	N
L	S	Floor 4	N	N	N	N	N

ภาพประกอบ 4 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล Random Forest ของตึก0และด้วย model K Nearest Neighbors มีค่า Accuracy เท่ากับ 0.9977, และ Confusion Matrix คือ

			PREDICTED LABEL				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	221	0	0	0	N
T	B	FLOOR 1	0	266	0	0	N
U	E	FLOOR 2	0	0	291	0	N
A	L	FLOOR 3	0	0	0	282	N
L	S	FLOOR 4	N	N	N	N	N

ภาพประกอบ 5 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล KNN ของตึก0 และด้วย model AdaBoost มีค่า Accuracy เท่ากับ 0.7788,และ Confusion Matrix คือ

			PREDICTED LABEL				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	211	0	0	0	N
T	B	FLOOR 1	0	0	266	0	N
U	E	FLOOR 2	0	0	291	0	N
A	L	FLOOR 3	0	0	0	282	N
L	S	FLOOR 4	N	N	N	N	N

ภาพประกอบ 6 Confusion Matrix ที่ได้จากผลการทำนายโดยรวมด้วยโมเดล AdaBoost ของตึก0

โดย BUILDINGID=1 ด้วย model Random Forest มี ค่า Accuracy เท่ากับ 0.7874,และ Confusion Matrix คือ

			PREDICTED LABEL				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	240	15	0	0	N
T	B	FLOOR 1	19	274	7	0	N
U	E	FLOOR 2	0	2	293	0	N
A	L	FLOOR 3	0	40	150	0	N
L	S	FLOOR 4	N	N	N	N	N

และด้วย model K Nearest Neighbors มีค่า Accuracy เท่ากับ 0.9977, และ Confusion Matrix คือ

			PREDICTED				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	225	0	0	0	N
T	B	FLOOR 1	0	299	1	0	N
U	E	FLOOR 2	0	1	293	1	N
A	L	FLOOR 3	0	0	0	190	N
L	S	FLOOR 4	N	N	N	N	N

และด้วย model AdaBoost มีค่า Accuracy เท่ากับ 0.7788, และ Confusion

Matrix คือ

			PREDICTED CLASS				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	255	0	0	0	N
T	B	FLOOR 1	0	300	0	0	N
U	E	FLOOR 2	0	295	0	0	N
A	L	FLOOR 3	0	0	0	190	N
L	S	FLOOR 4	N	N	N	N	N

ภาพประกอบ 7 Confusion Matrix ที่ได้จากการทำนายโดยรวมด้วยโมเดล AdaBoost ของตึก1

โดย BUILDINGID=2 ด้วย model Random Forest มี ค่า Accuracy เท่ากับ 0.7874, และ Confusion Matrix คือ

			PREDICTED CLASS				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	365	7	0	1	0
T	B	FLOOR 1	24	396	0	13	0
U	E	FLOOR 2	0	69	42	201	0
A	L	FLOOR 3	0	0	0	558	0
L	S	FLOOR 4	0	0	0	165	58

และด้วย model K Nearest Neighbors มีค่า Accuracy เท่ากับ 0.9977,

และ Confusion Matrix คือ

			PREDICTED CLASS				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	373	0	0	0	0
T	B	FLOOR 1	1	432	0	0	0
U	E	FLOOR 2	0	3	308	1	0
A	L	FLOOR 3	0	0	3	558	0
L	S	FLOOR 4	1	0	0	0	222

และด้วย model AdaBoost มีค่า Accuracy เท่ากับ 0.7788, และ Confusion

Matrix คือ

			PREDICTED CLASS				
			FLOOR 0	FLOOR 1	FLOOR 2	FLOOR 3	FLOOR 4
A	L						
C	A	FLOOR 0	373	0	0	0	0
T	B	FLOOR 1	0	433	0	0	0
U	E	FLOOR 2	0	312	0	0	0
A	L	FLOOR 3	0	0	0	558	0
L	S	FLOOR 4	0	0	0	0	223

จากการวิเคราะห์ผลการทดลองจะเห็นว่าโมเดล KNN มีประสิทธิภาพดีที่สุดในการทำนายตำแหน่งของโทรศัพท์มือถือจากสัญญาณ WiFi โดยมีประสิทธิภาพดี

ที่สุดในการทำนายตำแหน่งของอาคาร 0, 1 และ 2 ตามลำดับ โดยผลจากโมเดล Random Forest และ Adaboost ก็มีแนวโน้มคล้ายๆกันซึ่งสาเหตุที่น่าจะเป็นไปได้คือสัญญาณ WiFi ที่อาคาร 2 มีความไม่แน่นอนสูงทำให้ผลลัพธ์ที่ได้มีความคลาดเคลื่อนเยอะ ดังนั้นอาจจะต้องมีการประมวลผลสัญญาณเพื่อลดความไม่แน่นอนของสัญญาณก่อนจะนำเข้ามาใส่ในโมเดลการเรียนรู้ของเครื่องจักรต่อไป

5. สรุปและแนวทางการพัฒนาที่จะทำเพิ่มเติม

งานวิจัยนี้เสนอการใช้เทคนิคการเรียนรู้ของเครื่องจักรในการระบุตำแหน่งของโทรศัพท์มือถือจากสัญญาณ WiFi ภายในอาคาร จากการทดลองพบว่าโมเดล KNN ให้ความแม่นยำมากที่สุดในการทำนาย โดยมีความถูกต้องสูงถึง 97% ในการระบุตำแหน่งในรูปของชั้นในอาคารที่โทรศัพท์มือถืออยู่ อย่างไรก็ตามการระบุพิกัดของโทรศัพท์มือถือในรูปของละติจูดและลองจิจูดยังมีความคลาดเคลื่อนสูงอยู่ซึ่งมีสาเหตุหลักมาจากความแปรปรวนของสัญญาณ WiFi ภายในอาคารซึ่งแนวทางแก้ไขที่เป็นไปได้คือการกรองสัญญาณ WiFi ก่อนนำเข้ามาในโมเดลการเรียนรู้ของเครื่องจักร ซึ่งแนวทางการกรองสัญญาณด้วยเทคโนโลยีทาง signal processing ก็สามารถทำได้หลายรูปแบบเช่นการใช้ Kalman filtering เป็นต้นหรือ การใช้สัญญาณแบบอื่น เช่น สัญญาณ Beacon หรืออุปกรณ์ Radio Frequency Identification (RFID) มาช่วยในการระบุตำแหน่งให้มีความแม่นยำมากขึ้นต่อไป อย่างไรก็ตามการใช้สัญญาณ beacon หรือ RFID นั้นมีข้อเสียว่าต้องมี cost จากการติดตั้งอุปกรณ์เหล่านี้เพิ่มเติม ซึ่งต่างจากการใช้สัญญาณ Wifi ในการระบุตำแหน่งของโทรศัพท์มือถือในอาคารที่โดยทั่วไปแล้วจะมีอุปกรณ์ Wifi AP ติดตั้งเพื่อใช้ในการเชื่อมต่อเครือข่ายอินเทอร์เน็ตอยู่แล้ว ดังนั้นประเด็นเหล่านี้จึงต้องนำมาพิจารณาเพิ่มเติมในส่วนของคุณค่าจากการติดตั้งอุปกรณ์เพิ่มเติมเหล่านี้ การพัฒนาที่น่าจะทำเพิ่ม

ผู้วิจัยมีแนวทางในการต่อยอดความรู้ที่ได้จากการวิจัยนี้ดังต่อไปนี้

1. ศึกษาเทคโนโลยีทาง signal processing หรือการประมวลผลสัญญาณ เพื่อลดผลของสัญญาณรบกวนและลดความแปรปรวนของสัญญาณก่อนจะนำเข้าสู่โมเดลการเรียนรู้ของเครื่องจักรต่อไป
2. ทดลองใช้โมเดลการเรียนรู้ของเครื่องจักรแบบอื่นที่มีประสิทธิภาพดีกว่า KNN อาทิ เช่น XGBoost ในการวิเคราะห์สัญญาณ WiFi เพื่อระบุตำแหน่งของโทรศัพท์มือถือภายในอาคาร
3. เก็บข้อมูลสัญญาณ WiFi เพิ่มเติมเพื่อให้มีข้อมูลจำนวนมากขึ้นเพื่อให้โมเดลการเรียนรู้ของเครื่องจักรมีข้อมูลที่เรียนรู้เพื่อให้สามารถระบุตำแหน่งของโทรศัพท์มือถือจากสัญญาณ WiFi ได้แม่นยำขึ้น
4. ลองใช้โมเดลการเรียนรู้ของเครื่องจักรแบบไม่มีผู้สอน (Unsupervised learning) เช่น K-Means มาทำการจัดกลุ่มของสัญญาณ Wifi ก่อนจะนำเข้าสู่โมเดลการเรียนรู้ของเครื่องแบบมีผู้สอน เพราะการจัดกลุ่มอาจช่วยให้ข้อมูลมีการจำแนกประเภทอย่างชัดเจนและเป็นระบบมากขึ้นส่งผลให้โมเดลการเรียนรู้ของเครื่องจักรแบบมีผู้สอนสามารถทำการจำแนกสัญญาณได้ดีขึ้น

6. เอกสารอ้างอิง

- [1] Feng, Y., et al. Improved AdaBoost-based fingerprint algorithm for WiFi indoor localization. in 2014 IEEE 7th joint international information technology and artificial intelligence conference. 2014. IEEE.
- [2] Rusli, M.E., et al. An improved indoor positioning algorithm based on rssi-trilateration technique for internet of things (iot). in 2016 International Conference on Computer and Communication Engineering (ICCCCE). 2016. IEEE.

- [3] Torres-Sospedra, J., et al. UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. in 2014 international conference on indoor positioning and indoor navigation (IPIN). 2014. IEEE.
- [4] Xu, X., Y. Tang, and S. Li. Indoor localization based on hybrid Wi-Fi hotspots. in 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN). 2017. IEEE.
- [5] Improved AdaBoost-based fingerprint algorithm for WiFi indoor localization , 2014 IEEE 7th joint international information technology and artificial intelligence conference
- [6] An improved indoor positioning algorithm based on rssi-trilateration technique for internet of things , 2016 International Conference on Computer and Communication Engineering (ICCCE).
- [7] UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems , 2014 international conference on indoor positioning and indoor navigation (IPIN)
- [8] Indoor localization based on hybrid Wi-Fi hotspots , 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)
- [9] Sharan Naribole , A Machine Learning Approach to Wi-Fi Fingerprint based Localization,2014<https://github.com/sharan-Torres-Sospedra,2014#1>}naribole/wlan_localization/blob/master/UJIIndoorLoc-machine-learning.ipynb