

การใช้สื่อสังคมออนไลน์ เพื่อคำนวณสภาพการจราจรในเมืองใหญ่

ธนเดช บุญวรพงศ์ และ สุภกิจ นุตยะกุล

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบังฯ กรุงเทพฯ

Emails: tanadechb@gmail.com, supakit@it.kmitl.ac.th

บทคัดย่อ

งานวิจัยนี้ได้สนใจในส่วนของวิธีการที่จะนำข้อมูลจากสื่อสังคมออนไลน์ เช่น ทวิตเตอร์ มาใช้ในการวิเคราะห์และแปลงสภาพ เพื่อให้เป็นข้อมูลการจราจรในรูปแบบอื่นๆ อาทิเช่น เส้นสีการจราจรที่แสดงผลที่บ่งบอกสถานะต่างๆระบบแผนที่หรือระบบนำทางต่างๆ อาทิเช่น Google map [9] หรือ Longdo map [23] เพื่อช่วยให้ข้อมูลการจราจรมีประสิทธิภาพมากยิ่งขึ้น ซึ่งแหล่งข้อมูลที่เป็นสื่อสังคมออนไลน์ที่มีผู้ติดตามอันดับต้นๆ และมีบัญชีผู้ใช้ที่นิยมกระจายข่าวสารข้อมูลการจราจรเช่น js100radio : จส100 [3], fm91trafficpro : สวพ 91 [4] จากข้อมูลที่กำลังมาข้างต้น เมื่อเราได้ข้อมูลมาแล้ว ต้องทำการวิเคราะห์และแปลงสภาพข้อมูลเหล่านั้น โดยวิธีการ สังเคราะห์ได้ใช้ TLexs [22] เป็นต้นแบบในการวิเคราะห์ข้อมูลที่อยู่ในรูปแบบของประโยคให้เป็นคำที่มนุษย์เข้าใจได้ง่าย และใช้ HMM Model [20] ในการแปลงสภาพข้อมูลเหล่านั้นออกมาเป็นข้อมูลจราจรที่ต้องการ

คำสำคัญ – ปัญหาการจราจร; สื่อสังคมออนไลน์; สมาร์ทโฟน; ระบบขนส่งมวลชน

Abstract

This paper proposed a method to extract information from Social Network, such as, Twitter. The information is analyzed and transformed to another format (traffic information) that used for Google map [9] or Longdo map [23]. The purpose is to improve the effectiveness of traffic information. The main sources of traffic information on twitter that have a thousand followers in Thailand are js100radio [3] and fm91trafficpro [4]. The proposed analysis method is TLexs [22] to create a prototype model for extracting sentences that are understandable by human. The transformation method is HMM Model [20] to transform traffic information that we expected.

Keywords – Traffic Jam; Social Network; Smartphone; Transit System

1. บทนำ

ปัญหาการจราจรภายในกรุงเทพมหานคร นับเป็นปัญหาหนึ่งที่รัฐบาล และกรุงเทพมหานคร ได้พยายามศึกษาปัญหา และหาวิธีการแก้ไขอย่างยั่งยืนมาหลายยุค หลายสมัย ทั้งการก่อสร้างระบบขนส่งมวลชน อาทิเช่น รถไฟฟ้าบีทีเอส และรถไฟฟ้าใต้ดินเอ็มอาร์ที หรือ การพัฒนาระบบขนส่งรูปแบบใหม่ อาทิเช่น รถบัสอาร์ที แต่ถึงกระนั้น ก็ยังไม่มีวิธีการใดที่สามารถแก้ไขปัญหาได้อย่างชัดเจน [1] ซึ่งเกิดจากหลายๆปัจจัย อาทิเช่น 1) ปัญหาความไม่สมดุลกันของโครงสร้างระบบการขนส่งที่มุ่งใช้ถนนมากเกินไป 2) โครงสร้างระบบการขนส่งที่ใช้พลังงานสัดส่วนสูงสุดแต่ขาดประสิทธิภาพ 3) ปัญหาความแออัดของการจราจรในเขตกรุงเทพมหานครและเมืองใหญ่อื่นๆ 4) ปัญหาด้านกฎหมายและความหย่อนหยานในการบังคับควบคุมให้มีการปฏิบัติตามกฎหมาย โดยจากการประเมินของกองบัญชาการตำรวจนครบาล [1] เมื่อปี 2557 พบว่าปัญหาการจราจรภายในกรุงเทพมหานครจะมีแนวโน้มรุนแรงมากยิ่งขึ้น ทั้งนี้เป็นผลเนื่องจากการความต้องการในการเดินทางของประชาชนนั้นมีมากขึ้นนั่นเอง

ซึ่งผลจากปัญหาที่การจราจรนั้น พบว่าเป็น หนึ่งในปัจจัยที่มีผลต่อพฤติกรรมของมนุษย์ในการเดินทางไปยังที่ใดที่หนึ่งอีกด้วย [2] โดยในปี 2009 S.Priyawat และสมาชิก ได้ทำการศึกษาเพื่อตรวจสอบอิทธิพลของตัวแปรทัศนคติที่มีต่อการยอมรับมาตรการเก็บค่าใช้ถนน พบว่าการตระหนักถึงปัญหาการจราจรที่เกิดขึ้นในระดับบุคคลและสังคม เป็นส่วนหนึ่งของการยอมรับนี้ด้วย กล่าวคือปัญหาการจราจรมีผลต่อพฤติกรรมมนุษย์ที่จะเลือกว่าจะเดินทางหรือไม่ ในสถานะที่การจราจรติดขัดในเวลานั้น

จากปัญหาการจราจรนี้เอง ทำให้เมื่อประชากรต้องการเดินทางไปยังที่ใดที่หนึ่ง จะต้องหาข้อมูลเส้นทางการจราจรก่อนที่จะออกเดินทาง ซึ่งรูปแบบของการค้นหาข้อมูลการจราจรก็มีหลายประเภท อาทิเช่น การค้นหาข้อมูลการจราจรจากการฟังวิทยุ ที่ให้ข้อมูลข่าวสารด้านการจราจร จากการที่ได้รับข้อมูลการจราจรจากเจ้าหน้าที่ที่อำนวยความสะดวกอยู่บริเวณนั้น รวมถึงเปิดให้สมาชิกที่ฟัง สามารถโทรเข้าไปรายงานสภาพการจราจรในจุดที่มีปัญหาอีกด้วย อาทิเช่น [3] จส100, [4] สวพ.เอฟเอ็ม 91 อีกวิธีการหนึ่งคือ การค้นหาข้อมูลผ่านทางอินเทอร์เน็ต ซึ่งในปัจจุบันนี้อัตราการใช้อินเตอร์ของประชากรไทยนั้น เพิ่มสูงขึ้นมาก จะเห็นได้จากอัตราการเข้าถึงอินเทอร์เน็ตจากรายงานของสำนักงานสถิติแห่งชาติ [5] พบว่าจากเดิมในปีพ.ศ.2552-2556 มีอัตราใช้อินเทอร์เน็ตในประเทศไทย แต่เดิมประมาณร้อยละ 20.1 และได้เพิ่มขึ้นเป็นร้อยละ 28.9 ของกลุ่มตัวอย่าง ทำให้บ่งชี้ได้ว่าอินเทอร์เน็ตเป็นสิ่งที่ได้รับความนิยม ประกอบกับ ช่องทางการเข้าถึงอินเทอร์เน็ตได้มีมากขึ้น แต่เดิมสามารถเข้าถึงได้โดยผ่านทางคอมพิวเตอร์ แต่ในปัจจุบันสามารถเข้าถึงได้ในอุปกรณ์ที่เรียกว่า สมาร์ทโฟน [6] ซึ่งปัจจุบันสมาร์ทโฟนได้รับความนิยมอย่างมาก [7] ซึ่งจากจุดเด่นของสมาร์ทโฟนนั่น ทำให้เกิดการใช้งานที่หลากหลายมากขึ้น [8] ซึ่งจากงานวิจัยนี้จะเห็นได้ว่าพฤติกรรมการใช้งานสมาร์ทโฟนของประชากรนั้น จะแบ่งเป็น 2 ประเภทหลักๆคือ การใช้งานเว็บไซต์บนสมาร์ทโฟน และการใช้งานแอปพลิเคชันบนสมาร์ทโฟนมาก ซึ่งในส่วนของการใช้งานแอปพลิเคชันบนสมาร์ทโฟนนั่น ซึ่งจากประโยชน์ของอินเทอร์เน็ตนั้น ทำให้สามารถค้นหาข้อมูลการจราจรจากแหล่งต่างๆขึ้นอีกด้วย อาทิเช่น

Google Map [9] ที่เป็นบริการที่แผนที่รวมถึงข้อมูลการจราจรได้อีกด้วย และ Longdo map [23] ที่เป็นบริการให้แผนที่รวมถึงข้อมูลการจราจรที่เป็นบริษัทเอกชนสัญชาติไทย ที่สามารถให้ข้อมูลการจราจรเช่นเดียวกับ Google Map

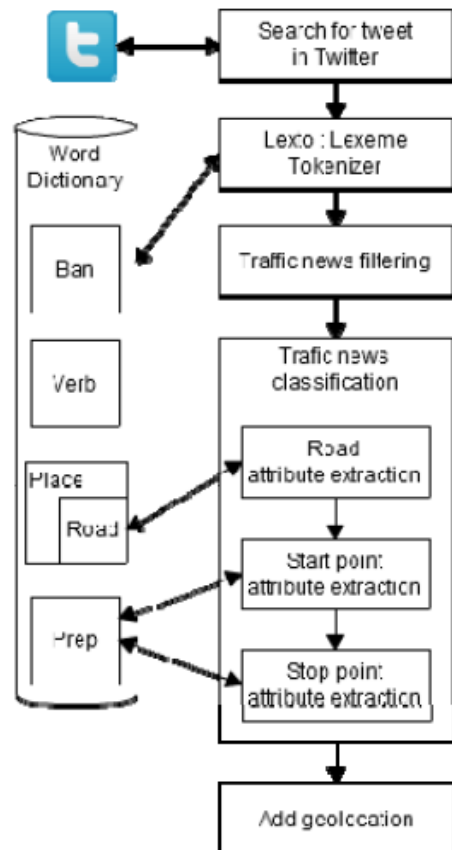
ซึ่งจากข้อมูลทีกล่าวมาข้างต้นทั้งหมดผู้วิจัยมีความสนใจที่จะใช้ข้อมูลการจราจรจากสื่อสังคมออนไลน์มาวิเคราะห์รวมถึงแปลงข้อมูลเพื่อให้เป็นข้อมูลการจราจรที่อยู่ในรูปแบบที่สามารถนำไปใช้ในรูปบบอื่นๆ อาทิเช่น ข้อมูลการจราจรในรูปแบบแผนที่เช่น Google Map [9] หรือ Longdo Map [23] เพื่อให้ข้อมูลการจราจรมีความสมบูรณ์มากยิ่งขึ้น

2. เทคนิคการดึงข้อมูลจากสื่อสังคมออนไลน์มาวิเคราะห์การจราจร

2.1 Social-based Traffic Information

Extraction and Classification [10]

ในงานวิจัยชิ้นนี้ ใช้การใช้ประโยชน์จากข้อมูลจราจร ผ่านเครือข่ายสื่อสังคมออนไลน์ ซึ่งคอนเซ็ปของงานวิจัยนี้มี 3 ขั้นตอนหลักได้แก่ รวบรวม, สรุปลผล, เผยแพร่ ข้อมูลการจราจร โดยใช้เทคนิค Syntactic Analysis ในการวิเคราะห์ข้อความที่ส่งกันผ่านโปรแกรมสื่อสังคมออนไลน์ที่เรียกว่า Twitter “ทวิตเตอร์” [11] โดยใช้โปรแกรมที่เรียกว่า LexTo “เล็กซ์โต” : โปรแกรมตัดคำสำหรับภาษาไทย [12] โดยมีกระบวนการดังรูปภาพที่ 1



รูปภาพที่ 1 : กระบวนการทำงานของการนำข้อความจากโปรแกรมสื่อสังคมออนไลน์ทวิตเตอร์ แปลงเป็น ข้อมูลการจราจร

จากรูปภาพที่ 1 มีรายละเอียดดังนี้

A. Search for traffic tweets in twitter

เป็นขั้นตอนที่กำหนดขอบเขตของการค้นหาข้อมูลที่เราอยากได้ ในกรณีนี้เช่นคำว่า “รถติด” และ “อุบัติเหตุ” เป็นขอบเขตของการค้นหา

B. Text tokenization using Lexo and

Dictionary

Dictionary เป็นขั้นตอนที่ใช้พจนานุกรม (dictionary) ในการวิเคราะห์และสังเคราะห์ข้อมูลที่ได้มา

จาก twitter โดยแบ่งออกเป็น 4 ประเภทคือ 1) สถานที่ 2) กริยา 3) คำต้องห้าม 4)คำบุพบท

Lexto เป็นพจนานุกรมที่ใช้วิเคราะห์คำไทยที่สร้างโดย Sansarn [18] ซึ่งสามารถใช้เพื่อวิเคราะห์ข้อมูลการจราจรได้

C. Traffic news filtering

เป็นขั้นตอนที่วิเคราะห์ว่าทวิตเตอร์ไหนสามารถใช้งานเป็นข้อมูลการจราจรที่จริงได้

D. Traffic news classification

โดยแบ่งประเภทของข้อมูลการจราจรที่ได้มาจากทวิตเตอร์นี้เป็น 2 ประเภท 1) พ้อยท์ โดยพ้อยท์หมายความว่า เป็นจุดที่เกิดเหตุการณ์ หรือเกิดเหตุขึ้น 2) ลิ้งค์ เป็นข้อมูลที่ตั้งจุดเริ่มต้น ไปถึง จุดสิ้นสุด เช่น รถติด เป็นต้น ซึ่งในตารางที่ 1 นั้น เป็นกระบวนการจัดเรียงข้อมูลการจราจรที่ได้มานั้นนำมาแปลงข้อมูลจากโปรแกรมทวิตเตอร์เป็นข้อมูลการจราจร ซึ่งการจัดเรียงข้อมูลนี้มี 2 เงื่อนไข ดังนี้

- 1) ข้อความที่นำมาใช้จะต้องมีอย่างน้อย 2 คำ คือ ชื่อของสถานที่ และ ประเภทของคำกริยา
- 2) ข้อความที่นำมาใช้จะต้องไม่มีคำต้องห้าม เช่น ประโยคที่เป็นคำถาม

Messages	Place	Verb	Ban
รถติด	-	รถติด	-
ถนนดินแดงรถติดหรือเปล่าครับ	ถนนดินแดง	รถติด	หรือเปล่า
แยกแควรายมีอุบัติเหตุจุดขึ้นกับรถรางแล้วมีคนเจ็บ	แยกแควราย	ชน	-

ตารางที่ 1 : ตัวอย่างของข้อมูลการจราจรที่ผ่านกระบวนการแล้ว

ในขั้นตอนต่อไปต้องทำการวิเคราะห์ให้เป็น ถนน, จุดเริ่มต้น,จุดสิ้นสุด เพื่อที่จะนำมาวิเคราะห์ข้อมูลการจราจรดังตารางที่ 2

Message	Road	Start	End	Category
สุขุมวิทขาออก ข้ามแยกบางนาทางหน้า กรมอุตสาหกรรม	สุขุมวิท	แยกบางนา	กรมอุตสาหกรรม	Link
แจ้งวัฒนะฝนตกแล้ว จรัล	แจ้งวัฒนะ	-	-	Point
ส.กรุงธนฯเข้า ท้ายสะพานลอยในถนน สีรินทว	ส. กรุงธนฯ	-	ถนน สีรินทว	Point
ส.กรุงธนฯ มุ่งหน้า อนุสาวรีย์ รถโล่ง	ส. กรุงธนฯ	อนุสาวรีย์	-	Point

ตารางที่ 2 : ตัวอย่างข้อมูลที่ได้จากการวิเคราะห์

E. Add geo-location

หลังจากที่ได้ข้อมูลทั้งหมดมาแล้วนั้น ก็ทำการเพิ่มในส่วนของ แลตติจูด และ ลองจิจูดของพื้นที่ ที่ได้ข้อมูลมาเพื่อจำลองพื้นที่ของข้อมูลการจราจรบนแผนที่ว่าแสดงสภาพการจราจรเป็นอย่างไร โดยแหล่งข้อมูลของข้อความผ่านโปรแกรม ทวิตเตอร์นั้น มาจาก 3 แหล่งข้อมูลดังตารางที่ 3

FM91TRAFFICPRO	JS100RADIO	GUEST
14%	7%	79%

ตารางที่ 3 : จำนวนของข้อความที่นำมาใช้ในงานวิจัยเปรียบเทียบโดยผู้ใช้งาน

งานวิจัยชิ้นนี้เป็นงานวิจัยที่บ่งบอกได้ว่าข้อมูลการจราจรจากสื่อสังคมออนไลน์ สามารถเป็นอีกแหล่งข้อมูลหนึ่งที่เชื่อถือและสามารถนำมาวิเคราะห์หรือสังเคราะห์ให้อยู่ในรูปแบบอื่นๆที่สามารถเข้าใจได้ง่าย ซึ่งแหล่งข้อมูลสื่อสังคมออนไลน์ที่งานวิจัยนี้ยกตัวอย่าง

ขึ้นมา คือ ทวิตเตอร์ แต่เทคนิคที่งานวิจัยนี้ยังสามารถนำไปประยุกต์ใช้กับสื่อสังคมออนไลน์ประเภทนี้ได้อีกด้วย

2.2 Study Traffic Conditions by Analyzing Foursquare and Instagram Data [13]

โดยงานวิจัยชิ้นนี้ใช้การวิเคราะห์ข้อมูลจากโปรแกรมสื่อสังคมออนไลน์ 2 ประเภท คือ Foursquare [14] และ Instagram [15] ซึ่งเป็นโปรแกรมสื่อสังคมออนไลน์ที่ใช้การ “เช็คอิน” และนำข้อมูลนั้น มาวิเคราะห์เพื่อที่จะนำข้อมูลว่าข้อมูลจากการเช็คอินนี้สามารถนำมาเพื่อพยากรณ์การเปลี่ยนแปลงของสภาพการจราจรในปัจจุบันได้หรือไม่

ซึ่งในการทดสอบสมมุติฐานนี้ ได้ใช้เมืองแมตฮันตัน, นิวยอร์ก ประเทศสหรัฐอเมริกา เป็นเมืองที่ไว้ใช้เก็บข้อมูลเพื่อทดสอบสมมุติฐานนี้ โดยใช้ข้อมูลจาก 2 ส่วน คือ (1) ข้อมูลจากการเช็คอินจาก Instagram และ foursquare (2) ข้อมูลการจราจรจาก Bing maps [16]

โดยในขั้นตอนของการทำงานนั้นได้แบ่งออกเป็น 5 ขั้นตอน ดังนี้

1) สรุปล่วงเวลาที่ใช้ในการเช็คอินเป็น 3 ชั่วโมง ต่อ 1 ช่วง 1 วันมีทั้งหมด 24 ชั่วโมง ดังนั้นจึงสามารถแบ่งได้เป็นทั้งหมด 8 ช่วงเวลา อย่างไรก็ตาม ไม่สามารถสรุปลข้อมูลจากการจราจรให้เป็นช่วงเวลาได้ แต่ข้อมูลการจราจรสามารถแบ่งได้ตามเส้นสีการจราจร สีเขียว คือ การจราจรสามารถทำได้อย่างสะดวก สีเหลือง คือ การจราจรติดขัดปานกลาง สีแดง คือ การจราจรติดขัดจากข้อมูลที่ทาง Bing Maps ให้นั้น สามารถแบ่งค่าตัวแปรของการจราจรได้ออกเป็น 3 ตัวแปร ดังนี้

- Bing = 1 เมื่อค่าเฉลี่ยของการจราจรเฉพาะสถานที่น้อยกว่า 1 (เส้นสีแดงแสดงการจราจร สีเขียว)
- Bing = 2 เมื่อค่าเฉลี่ยของการจราจรเฉพาะสถานที่มากกว่า 1 แต่ไม่เกิน 1.5 (เส้นสีแสดงการจราจร สีเหลือง)
- Bing = 3 เมื่อค่าเฉลี่ยของการจราจรเฉพาะสถานที่มากกว่า 1.5 (เส้นสีแสดงการจราจร สีแดง)

2) สำหรับในแต่ละส่วนของถนน สามารถคำนวณได้จากค่าเบี่ยงเบนมาตรฐานของการเช็คอิน ข้อมูลเหล่านี้จะใช้บนแผนที่ โดยแบ่งเป็นหมวดหมู่ 1, 2 หรือ 3 ตามประเภทที่แบ่งไว้

3) สำหรับช่วงเวลาที่แบ่งออกเป็น 8 ช่วงเวลา (1-8) มีขั้นตอนดังนี้ ขั้นตอนที่ 1 ตรวจสอบการจราจรตามประเภท (1, 2 หรือ 3) ขั้นตอนที่ 2 คำนวณจำนวนของการเช็คอินว่ามีส่วนไหนอยู่บนถนน และ ขั้นตอนที่ 3 จากจำนวนของการเช็คอินในส่วนถนน นั้นหมายถึง ค่ามาตรฐานที่ได้จากการคำนวณในขั้นตอนที่ 2 จะทำให้ผลของการเช็คอินสรุปได้ตามด้านล่างนี้

- ถ้าจำนวนของการเช็คอินในถนนมีค่าเท่ากับค่าความน่าเชื่อถือ นั้นหมายถึง ค่า Bing = 2
- ถ้าจำนวนของการเช็คอินในถนนมีค่ามากกว่าค่าความน่าเชื่อถือ นั้นหมายถึง ค่า Bing = 3
- ถ้านอกเหนือสองค่าที่กล่าวมา นั้นหมายถึง ค่า Bing = 1

4) หลังจากวิเคราะห์ข้อมูลทั้งหมดแล้วจึงหาความสัมพันธ์ระหว่าง ประเภทของการจราจร และ ประเภทของการเช็คอิน ดังนี้

- **Group 1:** จำนวนการเช็คอินมีค่าน้อยกว่าค่าของปริมาณการจราจร นั้นหมายถึง จำนวนของการเช็คอินมีน้อย แต่การจราจรมีความแออัด สำหรับกรณีนี้ การจราจรมีค่า $Bing = 2$ หรือ 3 และจำนวนของการเช็คอินมีน้อยกว่า
- **Group 2:** จำนวนการเช็คอินมีค่ามากกว่าค่าของปริมาณการจราจร นั้นหมายถึง ค่าของการเช็คอินมีมาก ค่าของการจราจรมี้อยู่ เมื่อ การค่าของการจราจรมีค่า $Bing = 1$ หรือ 2 และ จำนวนของการเช็คอินมีมาก
- **Group 3:** จำนวนการเช็คอินมีค่าน้อยกว่า หรือ เท่ากับค่าของปริมาณการจราจร ในการที่รวบรวมข้อมูลจาก Group 1 และ Group 3
- **Group 4:** จำนวนการเช็คอินมีค่าน้อยกว่า หรือ เท่ากับค่าของปริมาณการจราจร ในการที่รวบรวมข้อมูลจาก Group 2 และ Group 3
- **Group 5:** จำนวนการเช็คอินเท่ากับค่าของปริมาณการจราจร นั้นหมายถึง จำนวนของการเช็คอิน กับ การจราจรมีค่าเท่ากัน ซึ่งจะหมายถึงมีสัมพันธ์ค่าเท่ากับ 1

5) ในขั้นตอนสุดท้ายสรุปข้อมูลตามช่วงเวลาเป็นตัวแปร และแบ่งเวลาที่จะนำมาวิเคราะห์เป็นวันทำงาน (วันจันทร์-วันศุกร์) และ วันหยุด (วันเสาร์ และ วันอาทิตย์) โดยสำหรับ 2 ช่วงข้อมูลถูกแบ่ง โดยแบ่งเป็น ความถี่ของการเช็คอินใน 24 ชั่วโมง และ ความถี่ของความหนาแน่นของการจราจร

สมมุติฐานของเราใช้ 2 ฟังก์ชันในการปรับ δ นาที่ จากนั้นเรานำเสนอสมการ โดยที่ y_t คือค่าของการจราจร

ในเวลา t และ y^*_{k} คือค่าของการเช็คอินในเวลา k . เพื่อหาผลลัพธ์ของค่า δ ซึ่งคำนวณจากความแตกต่างของผลรวม ระหว่างค่าการจราจร y ในเวลา และค่าการเช็คอิน y ในเวลาที่ $t - \delta$ เป็นสมการด้านล่าง [13]

$$D(\delta) = \sum_t (|y_t - y^*_{(t-\delta)}|)$$

งานวิจัยชิ้นนี้เป็นงานวิจัยที่นำข้อมูลสื่อสังคมออนไลน์จากทาง Foursquare และ Instagram มาวิเคราะห์ในวิธีการที่แตกต่างกัน เพื่อที่จะนำข้อมูลนั้นไปแสดงผลในรูปแบบของการจราจรบนแผนที่ ซึ่งการแปลงข้อมูลให้อยู่การจราจรบนแผนที่นั้น ทำให้ข้อมูลการจราจรได้แสดงอยู่ในรูปแบบที่สามารถเข้าใจได้ง่ายมากยิ่งขึ้น อีกทั้งข้อมูลการจากสื่อสังคมออนไลน์ยังช่วยเพิ่มความสมบูรณ์ของข้อมูลการจราจรให้มีความละเอียดมากขึ้นอีกด้วย

3. เทคนิคในการแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลจราจร

3.1 Evolving HMM for ranking Twitter influence [19]

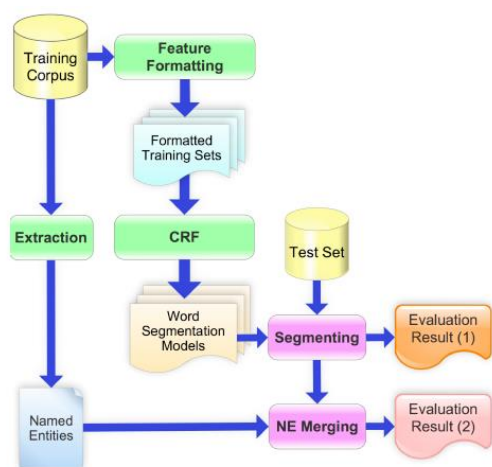
งานวิจัยชิ้นนี้ได้ทำการศึกษาในเรื่องของอิทธิพลของบัญชีผู้ใช้ในเครือข่าย twitter ซึ่งปัจจุบันนี้ทวีตเตอร์ได้เข้ามามีบทบาทในแง่ของ Digital Marketing ที่นำข้อมูลไปวิเคราะห์ หรือ การโฆษณาชวนเชื่อของกลุ่มผู้ก่อการร้าย ซึ่ง HMM (Hidden Markov Model) [20] เป็นวิธีการที่จะบันทึกและวิเคราะห์พฤติกรรมที่เกิดขึ้นเหล่านั้น

ซึ่งงานวิจัยนี้ได้ใช้วิธีการ evolutionary ที่มีพื้นฐานจาก HMM Model ในการวิเคราะห์ข้อมูล ซึ่ง วิธีการ

งานวิจัยชิ้นนี้ได้สนใจในเรื่องของการใช้ HMM Model ในการวิเคราะห์ข้อมูลจากสื่อสังคมออนไลน์ จะเห็นว่า HMM Model สามารถปรับเปลี่ยนรูปแบบในการรับข้อมูลได้ขึ้นอยู่กับผู้วิจัยสนใจจะนำข้อมูลชนิดใดมาใช้กับ HMM Model เพื่อวิเคราะห์

3.2 TLexs: Thai Lexeme Analyser Based on the Conditional Random Fields

งานวิจัยชิ้นนี้ได้นำเอา Conditional Random Fields (CRFs) มาสอนการแยกคำแล้วเก็บไว้ในคลังข้อมูล โดยมีวิธีการแยกคำตามรูปภาพที่ 2



รูปภาพที่ 2 : กระบวนการทำงานของการสอนประมวลผลเก็บ
ข้อมูลในคลังข้อมูล

จากรูปภาพด้านบน ในขั้นตอนแรกจะเป็นการเตรียมข้อมูลให้อยู่ในรูปแบบชุดข้อมูล จากนั้นจึงแบ่งชุดข้อมูลออกเป็น 3 ชุด ได้แก่ 1.char (ตัวอักษรทั้งหมด) 2.char-type (แบ่งจากประเภทของตัวอักษรเป็น 10 ประเภท) 3.combined (ใช้ทั้งตัวอักษรและรูปแบบของตัวอักษร)

สำหรับ char-type ได้แบ่งออกเป็น 10 ประเภท ตามตารางที่ 4

Tag	Type	Value
c	Consonant characters which can be assigned as the word ending character	ก ข ค ด ต ถ ท ธ น บ ป ผ ฝ พ ฟ ภ ก มี ฬ ฉ ล ช ย ส ห ส ฟ อ
n	Consonant characters which cannot be assigned as the word ending character	ท ฑ ฒ ฌ ญ ฮ ย
v	Vowel characters which are not allowed to begin a word	ะ - - - - - , ั ิ ุ ู ึ ื
w	Vowel characters which are allowed to begin a word	เ แ อ โ ใ ใ
t	Tonal characters	" " " " "
s	Symbol characters	" ๗ .
d	Digit characters	0-9
q	Quote characters	"' " "
p	Space character within a word	-
o	Other characters	a-z A-Z

ตารางที่ 4 : Character-type (char-type)

จากนั้นในการสอนคลังข้อมูลจึงได้มีการแปลงข้อมูลให้อยู่ในรูปแบบชุดข้อมูลในการสอนคลังข้อมูล ในรูปภาพที่ 3 เป็นตัวอย่างของการแปลงข้อความให้อยู่ในรูปแบบของ combined โดยคอลัมน์แรกเป็น char ในคอลัมน์ที่สองเป็น char-type และในคอลัมน์ที่สามเป็นการประกาศลำดับชั้นที่คาดคะเนได้ โดยแบ่งเป็น 2 ประเภท คือ word-beginning character (B) และ intra-word character (I) จากนั้นใช้การแบ่งค่าที่มีพื้นฐานจาก CRFs

CRFs algorithm ได้เรียนรู้จากการลำดับของ
สัญลักษณ์ที่เป็นไปได้จากการตรวจสอบการประกาศ
ลำดับชั้น ซึ่งการแยกค่าเหล่านี้สามารถทดสอบได้ด้วยการ
ทดสอบข้อมูลที่เป็นชุด

จ	C	B
ก	V	I
ค	C	B
ด	C	I
น	C	I
อ	n	I
บ	C	I
ป	V	I
ม	C	I
ล	w	B
ร	C	I
ช	v	I
อ	t	I
ง	C	I
ข	C	I
น	q	B
ท	C	B
า	C	I
ว	C	B
:	v	I
ค	o	B
	C	B

รูปภาพที่ 3 : ตัวอย่างของข้อมูลตัวอักษรในรูปแบบของ combined

ซึ่งจากการสอนคลังข้อมูลให้เก็บข้อมูลขนาดใหญ่โดยเรียกว่า name entities (NEs) โดยรูปภาพที่ 4 ได้แสดงถึงตัวอย่างของ NEs ใน 3 รูปแบบที่แตกต่างกัน คือ บุคคล,บริษัท และ สถานที่ ในรูปภาพที่ 4

NE type	Returned segments	Correct answer
Person	(หลวงตามหาบัว) (ญาณสัมปันโน)	(หลวงตามหาบัว ญาณสัมปันโน)
Person	(นายสุรินทร์) (พิศสุวรรณ)	(นายสุรินทร์ พิศสุวรรณ)
Organization	(กรมทรัพย์สินทางปัญญา)	(กรมทรัพย์สินทางปัญญา)
Organization	(สถานีวิทยุโทรทัศน์แห่งประเทศไทย)	(สถานีวิทยุโทรทัศน์แห่งประเทศไทย)
Place	(ศูนย์การค้าสยามพารากอน)	(ศูนย์การค้าสยามพารากอน)

รูปภาพที่ 4 : ตัวอย่างของ NEs ที่บันทึกลงในคลังข้อมูล

งานวิจัยนี้ได้นำเสนอถึงวิธีการที่จะจำแนกประโยคให้ออกเป็นในรูปแบบของคำ ซึ่งในกระบวนการที่จำแนกประโยคให้อยู่ในรูปแบบของคำนั้น ใช้การแบ่งประเภทของข้อมูลให้ออกเป็นรูปแบบที่กำหนดไว้ จากนั้นจึงใช้ CRFs ในการสอนคลังข้อมูลให้รู้จักคำต่างๆ หลังจากทีคลังข้อมูลได้เรียนรู้คำต่างๆแล้ว ก็จะสามารถจำแนกประโยคต่างๆออกมาเป็นคำได้นั่นเอง

4.สรุป

จุดประสงค์ของบทความนี้ต้องการจะนำเสนอวิธีการที่จะนำเอาข้อมูลสื่อสังคมออนไลน์มาใช้ในการจราจร โดยแบ่งออกเป็น 2 ส่วน 1) เทคนิคในการนำข้อมูลสื่อสังคมออนไลน์มาวิเคราะห์การจราจร และ 2) เทคนิคในการแปลงข้อมูลให้อยู่ในรูปแบบข้อมูลจราจร จากข้อ 1) จะเห็นได้ว่าข้อมูลที่ได้จากสื่อสังคมออนไลน์นั้น สามารถนำมา จำแนก วิเคราะห์ และ ปรับใช้ เพื่อช่วยแก้ไขปัญหาการจราจรได้ในระดับหนึ่ง ซึ่งในส่วนของงานวิจัย [11] นั้นได้กล่าวไว้ว่า ข้อมูลที่ได้จากทวีตเตอร์นั้น เป็นข้อมูลที่ช่วยในการจำแนกข้อมูลการจราจรได้ อีกทั้งยังสามารถช่วยให้ผู้ที่พบเห็นสามารถวางแผนและหลีกเลี่ยงจุดที่การจราจรคับคั่งได้ ซึ่งจากงานวิจัยนี้ได้รับรู้ไว้ว่า ข้อมูลจากสื่อสังคมออนไลน์อื่นๆ ซึ่งสามารถใช้ขั้นตอนวิธีที่คล้ายกัน กับสื่อสังคมออนไลน์ชนิดอื่นๆได้ อาทิเช่น Facebook [17] ที่มีลักษณะการใช้งานคล้ายคลึงกับ ทวิตเตอร์ ซึ่งการวิเคราะห์ข้อมูลด้านสื่อสังคมออนไลน์เพื่อจำแนกข้อมูลการจราจรนั้น ไม่ได้จำกัดเพียงแต่การใช้ข้อมูลในลักษณะตัวอักษรเท่านั้น

ซึ่งจากงานวิจัย [13] พบว่าการใช้ข้อมูลจากการเช็คอินในการวิเคราะห์การจราจรเช่นกัน ซึ่งจากข้อมูลที่งานวิจัยนี้ได้ค้นพบ พบว่าข้อมูลจากการเช็คอินจากสื่อสังคมออนไลน์ทั้ง 2 โปรแกรมนั้น สามารถนำมาวิเคราะห์ข้อมูลการจราจรได้ โดยเป็นการตรวจจับการเช็คอินที่อยู่บนถนน แล้วนำข้อมูลนั้น มาเปรียบเทียบกับปริมาณการจราจรที่เก็บรวบรวมมาได้จาก Bing Maps [16] ซึ่งในผลของงานวิจัยนั้นพบว่า การเช็คอินนั้นสามารถบ่งบอกปริมาณการจราจรในบริเวณนั้นตามหลัง 36 นาทีจากปริมาณการจราจรที่เก็บรวบรวมมาได้จาก Bing

Maps ซึ่งผู้วิจัยเชื่อว่าเวลา 36 นาทีนั้น คือเวลาที่เกิดขึ้นจากความแออัดของจราจรและเวลาที่ผู้ที่เดินทาง เดินทางมาถึงจุดเช็คอิน

จากงานวิจัยทั้งสองงานแรกนั้นถึงแม้ว่าข้อมูลด้านสื่อสังคมออนไลน์สามารถช่วยในการให้ข้อมูลการจราจรได้ แต่ทั้งนี้ก็ยังมีความน่าสนใจในเรื่องของ ความน่าเชื่อถือของข้อมูล เนื่องจากข้อมูลของด้านสื่อสังคมออนไลน์เป็นข้อมูลที่ต้องใช้ความพิจารณาอย่างรอบคอบก่อนที่จะตัดสินใจเชื่อถือข้อมูลนั้น ยกตัวอย่างเช่น ข้อมูลที่ได้จากทวิตเตอร์ของ จส100 จำเป็นต้องมีเจ้าหน้าที่ในการยืนยันข้อมูลการจราจรอีกขั้นหนึ่งก่อน จึงจะเผยแพร่ข้อมูลการจราจรนั้นต่อในสื่อสังคมออนไลน์อีกทีหนึ่ง ดังนั้นข้อมูลของสื่อสังคมออนไลน์ที่จะนำมาใช้เป็นข้อมูลการจราจรหลักเลยนั้น ยังคงต้องมีการทำวิจัยเพิ่มเติมก่อนที่จะนำข้อมูลดังกล่าวไปใช้งานจริง

จากข้อ 2) จากการที่ได้ข้อมูลจากสื่อสังคมออนไลน์มาแล้ว การแปลงข้อมูลให้อยู่ในรูปแบบของการจราจรนั้น จากงานวิจัย [23] เมื่อได้ข้อมูลจากสื่อสังคมออนไลน์มาแล้วนั้น ในการแปลงข้อมูลของสื่อสังคมออนไลน์ที่อยู่ในรูปแบบของประโยค ออกเป็นคำ หรือตัวอักษร การที่แปลงข้อมูลให้อยู่ในรูปแบบตัวอักษรได้นั้น ทำให้ง่ายต่อการที่จะนำข้อมูลไปประมวลผลต่อในกระบวนการใดๆก็ตาม โดยกระบวนการในการประมวลผลต่อ นอกจากใช้การสอนคลังข้อมูลดังที่งานวิจัยนี้ได้กล่าวไว้ ยังสามารถใช้วิธีต่างๆได้อีก อาทิเช่น ใช้ HMM Model [22] ในการประมวลผลได้อีกด้วย ซึ่งจากงานวิจัยที่กล่าวไว้ โดยปกติ HMM Model ได้ใช้ Baum Welch algorithm ในการกำหนดการกระหายและเปลี่ยนความน่าจะเป็นของข้อมูลเป็นการใช้ evolutionary algorithm ในการประมวล

แทน ซึ่งทางผู้เขียนบทความได้มองว่า มีความเป็นไปได้ที่จะใช้กระบวนการหรือรูปแบบอื่นๆ ในการประมวลผลได้

เอกสารอ้างอิง

- [1] กองบัญชาการตำรวจนครบาล. “แผนงานการพัฒนาและแก้ไขปัญหาการจราจรในกรุงเทพมหานคร”. 2557
- [2] ผศ.ดร.สุรเมศวร์ พิริยะพัฒน์. “ทฤษฎีพฤติกรรมและการประยุกต์สำหรับการวิเคราะห์ พฤติกรรมการเดินทาง ในประเทศไทย” วิศวกรรมสาร ปีที่ 63 ฉบับที่ 6 (พฤศจิกายน-ธันวาคม 2553). หน้า 59-70
- [3] จส 100 [Online] Available: <http://guru.sanook.com/2310/>
- [4] สวพ.เอฟเอ็ม 91 [Online] Available: <http://www.fm91bkk.com/home91/index.php/2011-12-29-10-39-40/2-2011-12-29-11-07-23>
- [5] สำนักงานสถิติแห่งชาติ กระทรวงเทคโนโลยีสารสนเทศและการสื่อสาร. “สรุปผลที่สำคัญ สํารวจการมีการใช้เทคโนโลยีสารสนเทศและการสื่อสาร”. 2556
- [6] สมาร์ทโฟนคืออะไร แท็บเล็ต-แพ็บเล็ต ต่างกันอย่างไร? [Online] Available: <http://news.siamphone.com/news-14121.html>
- [7] ธมนวรรณ กัญญาหัตถ์ และ ศรัณยพงศ์ เทียงธรรม. “ความพึงพอใจต่อผู้บริโภคในเขตกรุงเทพมหานคร ที่มีต่อประโยชน์ของสมาร์ทโฟน”. (วิทยานิพนธ์บริหารธุรกิจมหาบัณฑิต มหาวิทยาลัยกรุงเทพ, 2554).
- [8] Users spend 2 hours-plus a day toying with mobile apps (2014, April 1) [Online] Available: <http://www.cnet.com/news/users->

- spend-two-hours-plus-a-day-toying-with-mobile-apps/#ftag=CAD590a51e
- [9] About us, Google Map [Online] Available: <https://www.google.com/maps/about/>
- [10] Napong Wanichayapong, Wasawat Ptuthipunyaskul, Wasan Pattara-Atikom and Pimwadee Chaivalit, "Social-Based Traffic Information Extraction and Classification" in 11th International Conference on ITS Telecommunications, St. Petersburg, Russia, Aug 2011, pp 107-112
- [11] About us, Twitter [Online] Available: <https://about.twitter.com>
- [12] Lexto, Thai Lexeme Tokenizer [Online] Available : <http://www.sansarn.com/lexto>
- [13] Anna Izabel J.Tostes, Thiago H.Silva, Fátima Duarte-Figueiredo "Study Traffic Conditions by Analyzing Foursquare and Instagram Data" in 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks, New York, USA, 2014, pp 17-24
- [14] About us, Foursquare [Online] Available: <https://foursquare.com/about>
- [15] About us, Instagram [Online] Available: <https://instagram.com/about/us/>
- [16] The Bing Maps Platform [Online] Available: <http://www.microsoft.com/maps/>
- [17] About us, Facebook [Online] Available: <https://www.facebook.com/facebook>
- [18] Sansarn, Search engine with Thai lexeme tokenization by NECTEC. <http://www.sansarn.com>
- [19] Ali Ahmed Thwerani, Dr.Sayeed Ghani "Evolving HMM for ranking Twitter influence" in International Conference on Information and Communication Technologies (ICICT), 2015
- [20] Lawrence R. Rabiner "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition" in Proceeding of the IEEE, Volume 77, No.2, February 1989.
- [21] Goldberg, Dacia E. Genetic Algorithms. Pearson Education India, 2006.
- [22] Choochart Haruechaiyasak, Sarawoot Kongyoung "TLexs: Thai Lexeme Analyser Based on the Conditional Random Fields" Proc. 8th International Symposium on Natural Language Processing, 2009.
- [23] Longdo Map [Online] Available: <https://map.longdo.com>