การระบุขั้วความคิดเห็นในระดับคุณลักษณะของสมาร์ตโฟนบนเฟซบุ๊ก

Feature-Based Polarity Specification of SmartPhone on Facebook

มงคล แสนสุข[1*], พนิดา ทรงรัมย์[2] และ พัฒนพงษ์ ชมภูวิเศษ[3]

Mongkol Saensuk[1*] Panida Songram[2], and Phatthanaphong Chompoowiset[3]

วิทยาการคอมพิวเตอร์ วิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม

Computer Science Informatics Mahasarakham University

E-Mail: Mongkolsaensuk@gmail.com[1*], Panida.s@msu.ac.th[2], Phatthanaphong.c@msu.ac.th[3]

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการวิเคราะห์คุณลักษณะและจำแนกความคิดเห็นเป็นเชิงบวกและเชิงลบเกี่ยวกับคุณสมบัติของสมาร์ตโฟน โดยนำเสนอวิธีการระบุขั้วความคิดเห็นของสมาร์ตโฟนโดยใช้เทคนิควิธีการใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น ซึ่งวิธีการที่นำเสนอจะพิจารณาระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น เพื่อระบุว่าคำระบุขั้วความคิดเห็นที่ปรากฏในข้อความคิดเห็นเป็นของคำคุณลักษณะใด โดยใช้ API ของเว็บไซต์เฟซบุ๊ก ในการรวบรวมข้อความคิดเห็นในเพจต่าง ๆ ของสมาร์ตโฟน 3 รุ่น ได้แก่ Samsung, Oppo และ Asus

ผลการวิจัยพบว่า การระบุขั้วความคิดเห็นของสมาร์ตโฟนโดยใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็นมีประสิทธิภาพมากที่สุดคือ ค่าความแม่นยำ 75.15% ค่าความระลึก 78.40% และค่าประสิทธิภาพโดยรวม 76.74% เมื่อเทียบกับการจำแนกด้วยกฎ การจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ

คำสำคัญ: คำระบุขั้วความคิดเห็น, คุณสมบัติของสมาร์ตโฟน, เหมืองความคิดเห็น

Abstract

This paper presents a technique to develop of features analysis and classification of positive and negative opinions towards the feature words of smart-phone and designate the opinion polarities of smartphones using a distance-based approach. The proposed technique considers the spatial distance between the pre-define smartphone features with respect to an associated polarity. Data is collected from the comments appearing on Facebook pages of three major smartphone brands (i.e. (i) Samsung, (ii) Oppo and (iii) Asus in Thai.

The experiments show that the proposed technique is superior to the base-lines and provides a promising result and precision rate is 75.15% recall rate is 78.40% and f-measure rate is 76.74% of the result. Opinion Classification is carried out based on the smartphone features. A rule-based and 3-word-offset technique are compared.

Keywords: Polarity Words, Feature Words of Smartphones, Opinion Mining

บทนำ

ปัจจุบันโทรศัพท์เคลื่อนที่แบบสมาร์ตโฟนช่วยอำนวยประโยชน์ในชีวิตประจำวันของคนไทยมากขึ้น เช่น ใช้ในการถ่ายภาพ ใช้ในการสื่อสารและค้นคว้าหาข้อมูลต่าง ๆ วางแผนการเดินทาง รวมทั้งการทำธุรกรรมทางการเงิน ซึ่งการใช้งานเหล่านี้เกิดขึ้นจากคุณสมบัติที่มีในสมาร์ตโฟน เช่น กล้องถ่ายภาพ แอปพลิเคชัน หน่วยความจำ การเชื่อมต่อกับโลกออนไลน์ รายงานการวิจัยตลาดของบริษัท Insight Express ประเทศสหรัฐอเมริการะบุว่า คนไทยมีพฤติกรรมการใช้งานของสมาร์ตโฟนจนติดเป็นนิสัยถึง 98% [1] และมีการใช้งานสมาร์ตโฟนเพิ่มมากขึ้นถึง 29.1% โดยมียอดซื้อเพิ่มขึ้นจาก 7.1 ล้านเครื่อง เป็น 8 ล้านเครื่อง [2] ดังนั้นจึงทำให้บริษัทที่ผลิตและจัดจำหน่ายสมาร์ตโฟนในแต่ละองค์กรแข่งขันเพื่อที่จะนำสมาร์ตโฟนของตัวเองเข้าสู่ตลาด โดยใช้ช่องทางสื่อสารต่าง ๆ ในการโฆษณา เช่น สื่อสิ่งพิมพ์ ทีวี เว็บไซต์เครือข่ายสังคมออนไลน์ โดยเฉพาะเว็บไซต์เฟซบุ๊ก (www.facebook.com) ซึ่งเป็นเว็บไซต์เครือข่ายสังคมออนไลน์ที่ได้รับความนิยมและเข้าถึงผู้ใช้สูงสุดในปัจจุบัน [3] ดังนั้นผู้ผลิตและผู้จัดจำหน่ายสมาร์ตโฟนจึงได้ทำการโฆษณาสมาร์ตโฟนของตนเองในเว็บไซต์เฟซบุ๊ก และเว็บไซต์เฟซบุ๊กยังทำให้ผู้ผลิต ผู้จำหน่ายและลูกค้ามีปฏิสัมพันธ์กันอย่างง่ายขึ้น เช่น ผู้มีประสบการณ์ใช้งานหรือผู้ที่สนใจในสมาร์ตโฟน ก็จะทำการแสดงความคิดเห็นเชิงทัศนคติ ที่มีต่อสมาร์ตโฟนรุ่นที่สนใจ ดังนั้นเมื่อมีผู้สนใจจำนวนมากจึงทำให้เกิดความคิดเห็นต่อสมาร์ตโฟนจำนวนมากเช่นกัน ซึ่งข้อความคิดเห็นคุณสมบัติของสมาร์ตโฟน เช่น กล้องถ่ายภาพ แบตเตอรี่ หน้าจอ ขนาด การเชื่อมต่อ และราคา เป็นข้อความคิดเห็นที่ส่งผลกระทบต่อ ผู้ที่กำลังจะตัดสินใจซื้อหรือลูกค้าในอนาคต ผู้ผลิต และผู้จัดจำหน่ายได้ เช่น ถ้ามีการแสดงความคิดเห็นในเชิงลบที่มีต่อคุณลักษณะที่สำคัญของสมาร์ตโฟนจำนวนมาก เช่น “กล้องแย่” “ถ่ายรูปห่วย” “กล้องไม่ชัด” จะทำให้ผู้ที่กำลังจะเป็นลูกค้าไม่ตัดสินใจซื้อสมาร์ตโฟนรุ่นนั้น หรือถ้ามีการแสดงความคิดเห็นในเชิงบวกต่อคุณสมบัติของสมาร์ตโฟนจำนวนมากเช่น “ถ่ายรูปสวยมาก” “ภาพถ่ายชัดแจ๋ว” “กล้องแจ่มสุด ๆ” ก็จะทำให้ส่งผลด้านบวกต่อผู้ที่กำลังจะเป็นลูกค้า นอกจากนี้แล้วผู้ผลิต ผู้จัดจำหน่าย สามารถนำความคิดเห็นที่มีต่อสมาร์ตโฟนของตนไปกำหนดปัญหา ปรับปรุงและพัฒนาให้ตรงกับความต้องการของผู้ใช้งาน ก็จะช่วยให้ธุรกิจประสบความสำเร็จได้ [4] ดังนั้นข้อความคิดเห็นของผู้ที่สนใจที่มีต่อสมาร์ตโฟนบนเว็บไซต์เครือข่ายสังคมออนไลน์มีความสำคัญเป็นอย่างมาก เนื่องจากความคิดเห็นมีจำนวนมากและการใช้คำต่าง ๆ ที่แตกต่างกันในการเขียนความคิดเห็น ทำให้ยากที่จะอ่านและวิเคราะห์ทัศนคติของผู้แสดงความคิดเห็นในแต่ละคุณลักษณะ

ผู้วิจัยจึงเกิดแนวคิดที่จะนำความคิดเห็นเหล่านี้มาวิเคราะห์และศึกษา เพื่อหาแนวทางในการระบุว่าผู้แสดงความคิดเห็นมีความคิดเห็นเป็นเชิงบวกหรือเชิงลบต่อคุณลักษณะที่สนใจ โดยการวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็นในแต่ละความคิดเห็น ซึ่งจะสามารถระบุได้ว่า ในแต่ละคุณลักษณะของสมาร์ตโฟนมีทัศนคติในด้านใด เช่น กล้องเป็นเชิงบวก หน้าจอเชิงบวก แบตเตอรี่เชิงลบ เครื่องเชิงลบ

1. วัตถุประสงค์การวิจัย

เพื่อพัฒนาวิธีการวิเคราะห์คุณลักษณะและจำแนกความคิดเห็นเป็นเชิงบวกและเชิงลบเกี่ยวกับคุณสมบัติของสมาร์ตโฟน

2. งานวิจัยที่เกี่ยวข้อง

วรัญญา วรรณศรี และเด่นดวง ประดับสุวรรณ [5] ได้รวบรวมความคิดเห็นจาก www.agoda.co.th นำมาวิเคราะห์และพัฒนาเป็นระบบจำแนกความคิดเห็นของโรงแรม โดยการตัดข้อความภาษาไทยใช้เทคนิคนิพจน์ปรกติ (Regular Expression) ร่วมกับโปรแกรม LexToPlus ในคัดเลือกคำคุณลักษณะ ซึ่งเป็นคุณสมบัติของโรงแรม ได้กำหนดขึ้นมาเป็น 6 คุณลักษณะคือ การบริการ อาหาร โรงแรม สถานที่ตั้ง ห้องพัก และราคา และได้สร้างคลังคำที่มีความหมายเหมือนคำคุณลักษณะ เพื่อมาช่วยในการสกัดคำคุณลักษณะ และทำการวิเคราะห์

ความคิดเห็นในระดับคุณลักษณะ โดยนำข้อความในประโยคความคิดเห็นไปเปรียบเทียบกับคลังคำคุณลักษณะและคำระบุขั้วความคิดเห็น และจำแนกความคิดเห็นเป็นเชิงบวก เชิงลบ และเป็นกลาง ของแต่ละคุณลักษณะ เช่น “ห้องสะอาดมาก” จะประกอบไปด้วย คำคุณลักษณะ “ห้อง” คำระบุขั้วความคิดเห็น “สะอาด” คำระบุปริมาณ “มาก” และวัดประสิทธิภาพการสกัดคำคุณลักษณะ พบว่า Precision 76.87%, Recall 83.63%, F-Measure 80.11% และสกัดคำระบุความคิดเห็น Precision 74.51%, Recall 84.74%, F-Measure 79.30%

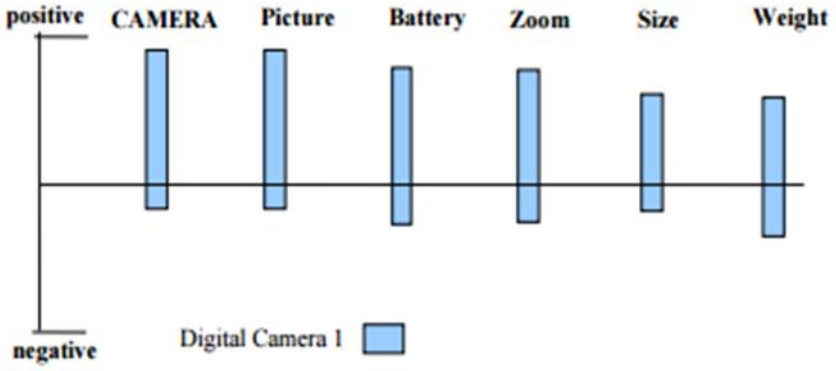
นริศร์ พรหมบุตร [6] นำเสนองานวิจัยในการจำแนกและสรุปค่าความพึงพอใจของความคิดเห็น โดยใช้ความคิดเห็นที่มีต่อโทรศัพท์มือถือในเว็บไซต์ www.siamphone.com มีวิธีการดังนี้ ขั้นตอนที่ 1 คัดเลือกความคิดเห็นเฉพาะที่เกี่ยวข้องกับโทรศัพท์มือถือ โดยมีเงื่อนไขการคัดเลือกคือ คัดเลือกตามคีย์เวิร์ด (Keyword) และคำคุณลักษณะต่าง ๆ เป็นต้น ขั้นตอนที่ 2 ตัดคำด้วยอัลกอริทึม Longest Matching ด้วยโปรแกรม Lexto – Thai Lexeme Tokenizer ขั้นตอนที่ 3 นำข้อความที่ได้ไปกำหนดโครงสร้างและตรวจสอบโครงสร้างทางไวยากรณ์ตามหลักภาษาไทย เพื่อหาข้อความที่เป็นคำคุณลักษณะ และข้อความแสดงความพึงพอใจตามหลักไวยากรณ์ทางภาษา เช่น “จอใหญ่” จอ คือ NF หรือคำคุณลักษณะ และใหญ่ คือ ADJ หรือข้อความแสดงถึงความพึงพอใจ ขั้นตอนที่ 4 หาค่าความพึงพอใจ โดยกำหนดค่าความถึงพอใจเป็น 3 ระดับ คือ ดี ไม่ดี และพอใช้ โดยระดับความพึงพอใจ ดี อาจประกอบไปด้วยหลายข้อความ เช่น เจ๋ง สุดยอด คุ้มค่า อึด ฯลฯ ระดับความพึงพอใจ ไม่ดี ประกอบไปด้วยหลายคำ เช่น แย่ ห่วย เจ้ง ฯลฯ จากนั้น นำคำ ADJ ดังกล่าวมา ตรวจสอบกับฐานข้อมูลความพึงพอใจระดับต่าง ๆ เพื่อกำหนดระดับความถึงพอใจดังกล่าว ผลลัพธ์ของงานวิจัยจะแสดงในรูปของแต่ละคำคุณลักษณะ และอธิบายถึงปริมาณข้อความความระดับความพึงพอใจ เช่น คุณลักษณะรูปทรง ประกอบไปด้วย ดี 26 ข้อความ พอใช้ 0 ข้อความ ไม่ดี 1 ข้อความ และทำการวัดประสิทธิภาพด้วย Recall และ Precision เฉลี่ยทุกคุณลักษณะคือ 60% และ 77% ตามลำดับ

Wang and Song. [7] ได้รวบรวมความคิดเห็นเกี่ยวกับอุปกรณ์อิเล็กทรอนิกส์ (Electric Devices) ที่เป็นภาษาจีน จำนวน 662 ความคิดเห็น ประกอบไปด้วยความคิดเห็นในเชิงบวก จำนวน 356 ความคิดเห็น และความคิดเห็นเชิงลบจำนวน 306 ความคิดเห็น ความคิดเห็นที่เกี่ยวกับโรงแรม (Hotel) จำนวน 398 ความคิดเห็น เป็นความคิดเห็นเชิงบวกจำนวน 180 ความคิดเห็น และความคิดเห็นที่เป็นเชิงลบจำนวน 218 ความคิดเห็น และความคิดเห็นเกี่ยวกับเกี่ยวกับวารสารออนไลน์ (E-Journal) 297 ความคิดเห็น ซึ่งเป็นความคิดเห็นที่เป็นเชิงบวกจำนวน 139 ความคิดเห็น และความคิดเห็นเชิงลบ 158 ความคิดเห็น โดยนำเสนอวิธีการถ่วงน้ำหนักคำขยายกริยา High Adverb of Degree Count (HADC) ในข้อความคิดเห็น เพื่อปรับปรุงค่าของคำแสดงความรู้สึกให้ชัดเจน รวมถึงการตรวจสอบจากสัญลักษณ์ท้ายประโยคด้วย เช่น “โรงแรมสกปรกมาก !” “โรงแรมสกปรก ?” ซึ่งได้พิจารณาความสัมพันธ์ระหว่างคำสำคัญจากปัจจัยต่างๆ ประกอบด้วย ti = (f,o,omc,gms,nmc,p) โดยที่ ti คือ ความคิดเห็นที่ 1 ที่ถูกระบุคลาสแล้ว f คือ คุณลักษณะ o คือ คำระบุขั้วความคิดเห็นของคำคุณลักษณะนั้น omc คือ จำนวนของคำระบุขั้วความคิดเห็นที่ถูกแก้ไขทั้งหมด เช่น การเช็คระดับความชอบ ได้แก่ “excessively”, “too”, “ultra”, gms คือ ค่าเฉลี่ยของคะแนนการแก้ไขของคำระบุขั้วความคิดเห็น o, nmc คือ จำนวนคำคัดค้านของคำระบุขั้วความคิดเห็น o และ p คือ ลักษณะของประโยคว่าเป็นประโยคบอกเล่าหรือประโยคคำถาม โดยกำหนดค่าคะแนนคำระบุขั้วความคิดเห็นเป็น 0-2 ผลการวัดประสิทธิภาพ ด้วยเทคนิควิธีในการถ่วงน้ำหนักที่นำเสนอ มีค่าความถูกต้องคือของ Electric devices 89.14%, Hotel 91.90% และ E-journal 88.88%

Liu [8] นำเสนองานวิจัยเกี่ยวกับการทำเหมืองความคิดเห็นโดยกล่าวว่า การทำเหมืองความคิดเห็นคือการสกัดเอาคุณลักษณะและส่วนประกอบของความคิดเห็นที่มีต่อวัตถุที่สนใจในแต่ละเอกสารหรือความคิดเห็น และตรวจสอบความคิดเห็นนั้นว่าเป็นเชิงบวก เชิงลบ หรือเป็นกลาง ในการทดลองซึ่งใช้ชุดข้อมูล Product Reviews โดยคัดเลือกเอาเฉพาะความคิดเห็นที่เกี่ยวกับกล้องดิจิทัล หรือ Digital_Camera_1 จำนวน 132 ความคิดเห็น และใช้คุณลักษณะของกล้องดิจิทัล (Feature Based) เป็นคุณลักษณะ (Feature) จากนั้นจำแนกความคิดเห็นเป็นเชิงบวกและเชิงลบในแต่ละคุณลักษณะ โดยคุณลักษณะของกล้องประกอบไปด้วย 1) ตัวกล้อง 2) รูปภาพ

3) แบตเตอรี่ 4) การซูม 5) ขนาด 6) น้ำหนัก ซึ่งในการจำแนกโดยการนำคุณลักษณะเหล่านี้มาแทนคุณลักษณะนี้ Liu ได้กล่าวถึงปัญหา 3 ข้อ ดังนี้ 1) การสกัดคุณลักษณะของวัตถุจากข้อความที่มีอยู่ในความคิดเห็นจะทำได้อย่างไร

2) จะกำหนดขั้วความคิดเห็นในแต่ละคุณลักษณะว่าเป็นความคิดเห็นในเชิงบวกและเชิงลบ หรือเป็นกลางได้อย่างไร 3) ปัญหาการจัดกลุ่มคำที่มีความหมายคล้ายกับคำคุณลักษณะในความคิดเห็นที่แตกต่างกันจะใช้คำหรือวลีที่แตกต่างกันในการเขียนความคิดเห็นต่อคุณลักษณะเดียวกัน เช่น “ภาพถ่ายนี้สวยจัง”หรือ “รูปนี้สวยจัง” เพื่อให้ได้ผลลัพธ์ที่ได้จากการจำแนกในลักษณะดังกล่าว Liu จึงนำเสนอวิธีการสรุปความคิดเห็น (Opinion Summary) โดยมีตัวอย่างผลลัพธ์ที่ได้จากการจำแนกดังภาพที่ 1



ภาพที่ 1 ผลลัพธ์ที่ได้จากการจำแนกความคิดเห็นของ Digital_Camera_1 [8]

จากภาพที่ 1 จะเห็นได้ว่า ผลลัพธ์ที่ได้ในแต่ละคุณลักษณะมีการสรุปรวมความคิดเห็นเป็นเชิงบวกและเชิงลบและ Liu ได้ยกตัวอย่างประโยคคือ “The Picture Quality of this Camera is Amazing” โดยคุณลักษณะของประโยคนี้คือ Picture Quality และมีความคิดเห็นเป็นเชิงบวกต่อคุณลักษณะนี้ คือคำว่า Amazing และได้กล่าวถึงเทคนิคที่สามารถวิเคราะห์แยกเอาคุณลักษณะและคำแสดงถึงความรู้สึกจากงานวิจัยของ Liu และ Cheng [9] ซึ่งสามารถแก้ปัญหานี้ได้โดยการนำความคิดเห็นที่ได้เข้าสู่การทำ Path-of-speech tagging (POS tags) โดยการหาประธานของประโยคซึ่งเป็นคำนาม (N) คือคำว่า Picture quality และคำที่ระบุขั้วความคิดเห็นหรือคำคุณศัพท์ (Adj) คำว่า amazing

Mukherjee และ Bhattacharyya [10] นำเสนองานวิจัยเกี่ยวกับ การวิเคราะห์ความรู้สึกแบบเจาะจงสำหรับความคิดเห็นของสินค้า ในงานวิจัยนี้ได้มุ่งเน้นไปที่ การระบุโดเมนของประโยคความคิดเห็นและการระบุว่าคำระบุขั้วความคิดเห็นเป็นของคุณลักษณะใด ในขั้นตอนการกำหนดปัญหาของงานวิจัยนี้ได้กำหนดปัญหาของงานออกเป็น 3 อย่าง ได้แก่ 1) การสกัดคำคุณลักษณะ เพื่อจะตรวจสอบว่าประโยคความคิดเห็นนั้นเป็นความคิดเห็นของโดเมนอะไร 2) การสกัดคำระบุขั้วความคิดเห็น โดยสกัดจากขั้นตอนการทำ POSTagging และในงานวิจัยนี้ระบุว่า คำระบุขั้วความคิดเห็นไม่ได้มีแค่คำคุณศัพท์อย่างเดียว (Adjective) แต่ยังรวมไปถึงคำนาม (Noun) คำกริยา (Verb) และคำขยายคำนาม (Adverb) 3) การจำแนกคำระบุขั้วความคิดเห็นเป็นขั้วบวกและขั้วลบในแต่ละคุณลักษณะ เช่น ไม่เลว ไม่แย่ เป็นเชิงบวก ขั้นตอนการดำเนินงานวิจัยมีดังนี้ 1) การสกัดคุณลักษณะ โดยแบ่งออกเป็น 2 วิธีได้แก่ การสกัดคุณลักษณะแบบไม่ทราบโดเมน และแบบทราบโดเมน วิธีที่ 1 แบบไม่ทราบโดเมนคือในประโยคความคิดเห็นดังกล่าวไม่ถูกระบุป้ายของโดเมน (Domain Label) ไว้ ดังนั้นคุณลักษณะที่สกัดมาได้จากถูกนำไปเปรียบเทียบกับข้อมูลเบื้องต้น (Prior) ของแต่ละโดเมนเพื่อหาว่าประโยคความคิดเห็นอยู่ในโดเมนอะไร วิธีที่ 2 แบบทราบโดเมน จะทำการลดคุณลักษณะคือ คำนาม ที่ไม่เกี่ยวข้องออกไป 2) การสกัดความสัมพันธ์ ในขั้นตอนนี้

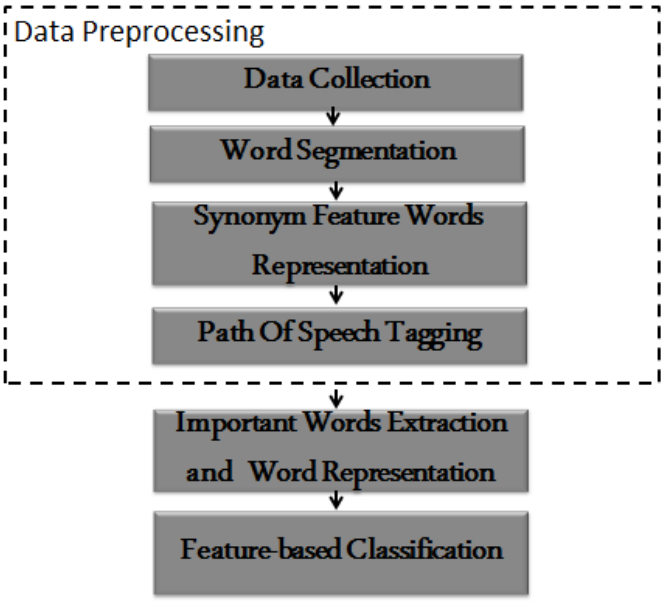
จะทำการคำนวณอย่างง่ายจากระยะห่างของคำต่าง ๆ ที่สกัดออกมาได้ โดยแบ่งเป็น 2 วิธี ได้แก่ วิธีที่ 1 ความสัมพันธ์เพื่อนบ้านโดยตรง (Direct Neighbor Relation) โดยวัดระยะห่างของคำคุณลักษณะกับคำระบุขั้วความคิดเห็นโดยประโยคนั้นต้องไม่มีคำหยุด วิธีที่ 2 ความสัมพันธ์แบบไม่เป็นอิสระ (Dependency Relation) โดยจะวัดระยะห่างจากชนิดของคำที่อยู่ใกล้คำนาม เช่น คำคุณศัพท์ คำกริยา คำขยายคำนาม เพื่อหาคำระบุขั้วความคิดเห็นของคำคุณลักษณะนั้น 3) การแทนค่าลงในกราฟ (Graph Representation) คือ การนำคำสำคัญที่ได้จากการสกัดความสัมพันธ์ มาแทนค่าลงในกราฟ 4) การสกัดความไม่เป็นอิสระ (Dependency Extraction) โดยมีเทคนิควิธีในการสกัดคือ การจัดกลุ่ม โดยการวัดระยะห่างระหว่างคำสำคัญโดยใช้ระยะของคำเป็นระยะห่างของเส้น ถ้าระยะห่างระหว่างคำคุณลักษณะกับคำระบุขั้วความคิดเห็นนั้นมีน้อยกว่าระยะห่างกับคุณลักษณะอื่น จะถือว่าคำระบุขั้วความคิดเห็นเป็นของคุณลักษณะดังกล่าว 5) จำแนกความคิดเห็นของคุณลักษณะที่ได้จากการสกัด (Classification of Extracted Features) ในขั้นตอนนี้จะสรุปว่า แต่ละคุณลักษณะมีความคิดเห็นเป็นขั้วบวกหรือขั้วลบ โดยแบ่งวิธีการจำแนกออกเป็น 2 แบบ คือ 1) การจำแนกแบบการใช้กฎ (Rule Based Classification) วิธีนี้จะใช้พจนานุกรมในการจำแนก 2) การจำแนกโดยใช้เทคนิคเครื่องเรียนรู้ (Supervised Classification) โดยการจำแนกเฉพาะคำระบุขั้วความคิดเห็นในแต่ละประโยค ตัวอย่างเทคนิคในการจำแนก เช่น ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) 6) การเรียนรู้พารามิเตอร์ (Learning Parameters) คือ การเรียนรู้ค่าพารามิเตอร์จากข้อมูลที่มีอยู่ โดยทำการจัดกลุ่ม 2 กลุ่ม พบว่าค่า เธรทโชว์ที่ทำให้เกิดความถูกต้องสูงคือ 3 โดยมีความถูกต้องคือ 69.29% ในการทดลองของงานวิจัยนี้ใช้ชุดข้อมูล จำนวน 2 ชุด โดยชุดที่ 1 มี 500 ความคิดเห็น โดยสกัดจาก 3 โดเมน ได้แก่ Labtops, Camera และ Printers และชุดที่ 2 จำนวน 2500 ความคิดเห็น ในโดเมนที่ต่างกัน ได้แก่ antivirus, camera, dvd, ipod, music player, router, mobile และอื่นๆ โดยแต่ละประโยคมีลาเบล ค่าความถูกต้องรวมของข้อมูลชุดที่สองเมื่อเทียบกับ Baseline1 และ Baseline2 คือ Baseline1 50.35% Baseline2 58.93% วิธีการที่นำเสนอคือ 70% จากนั้นทำการทดลองการจำแนกแบบ rule-based ในชุดข้อมูลที่ 1 สำหรับวิเคราะห์เฉพาะความรู้สึกว่าเป็นเชิงบวกหรือเชิงลบ พบว่า มีความถูกต้อง 80.98% และ Baseline1 คือ 68.75% และ Baseline2 คือ 61.10% ส่วนการจำแนกแบบเครื่องเรียนรู้กับโดเมน Mobile และ Camera ด้วยเทคนิควิธี Support Vector Machine ค่าความถูกต้องของ Baseline1 (mobile) คือ 51.42% Baseline1 (camera) คือ 50% วิธีการที่นำเสนอคือ mobile 83.82% และ camera 86.99%

วิธีดำเนินการวิจัย

1. เครื่องมือการวิจัย

1.1 Graph API Facebook Version 2.5
1.2 ฐานข้อมูล MySQL และ ภาษา PHP
1.3 เครื่องมือ Matlab ใช้ในการทดลองวิเคราะห์จำแนกความคิดเห็นและวัดประสิทธิภาพงานวิจัย
1.4 เครื่องมือในการวัดประสิทธิภาพ Precision Recall F-Measure

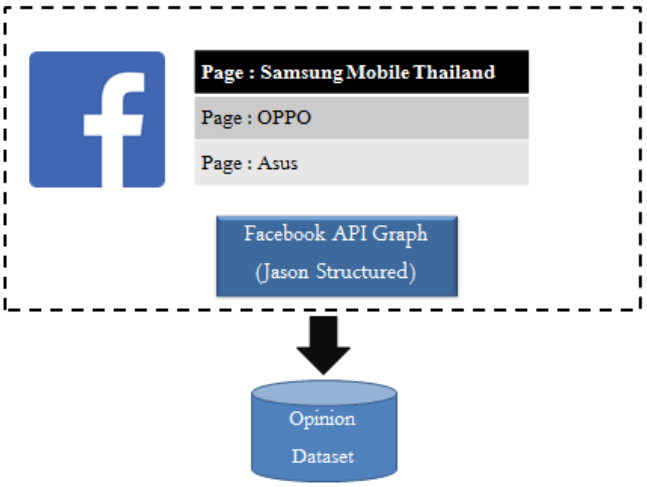
2. ขั้นตอนดำเนินการโดยรวมในการระบุขั้วความคิดเห็น แสดงดังภาพที่ 1



ภาพที่ 1 ขั้นตอนการระบุขั้วความคิดเห็น

1.1 การเตรียมข้อมูล (Data Preprocessing) คือขั้นตอนของการจัดการข้อมูลให้อยู่ในรูปแบบที่สามารถนำไปวิเคราะห์ได้ซึ่งประกอบไปด้วย

1.1.1 การรวบรวมข้อมูล (Data Collection) คือขั้นตอนในการรวบรวมข้อมูลหรือข้อความคิดเห็นภาษาไทย ในหน้าเพจของสมาร์ตโฟน ประกอบไปด้วย Samsung Mobile Thailand, Oppo และ Asus โดยรวบรวมผ่าน Graph-API ของเฟซบุ๊กซึ่งอยู่ในโครงสร้างแบบ Jason ซึ่งเป็นชุดข้อมูลในการทดลอง และเก็บไว้ในฐานข้อมูล MySQL ในภาพที่ 2 การเก็บรวบรวมข้อความคิดเห็น



ภาพที่ 2 การเก็บรวบรวมข้อความคิดเห็น

และได้นำข้อความคิดเห็นที่ได้ทำการเก็บรวมรวมมาระบุคลาส ซึ่งให้ผู้เชี่ยวชาญด้านที่มีความรู้เกี่ยวกับสมาร์ตโฟนและการอ่านภาษาไทยเป็นผู้ระบุคลาส จำนวน 5 คน และคัดเลือกเอาเฉพาะข้อความคิดเห็นที่ผู้เชี่ยวชาญมีความคิดเห็นตรงกันอย่างน้อย 3 คนขึ้นไป มาใช้เป็นข้อมูลในการทดลองดังตารางที่ 1

ตารางที่ 1 สัญลักษณ์ที่ใช้แทนคำสำคัญ

ลำดับ	ชื่อเพจในเว็บไซต์เฟซบุ๊ก	จำนวนข้อความคิดเห็น
1	Samsung mobile Thailand	497
2	OPPO	499
3	ASUS	299

1.1.2 การตัดคำ (Word Segmentation) การตัดคำ คือ การแบ่งคำด้วยเทคนิควิธีต่าง ๆ จากข้อความโดยมีวิธีการแตกต่างกันไป ในงานวิจัยนี้ได้ใช้เทคนิควิธี การตัดคำแบบคำยาวที่สุด (Longest Matching) ร่วมกับพจนานุกรมตัดคำ Lexitron ของ NECTEC [11] และเพิ่มพจนานุกรมคำคุณลักษณะและคำที่มีความหมายเหมือนคำคุณลักษณะลงไปในพจนานุกรมตัดคำด้วย ดังนั้นคำที่ได้ในแต่ละประโยคความคิดเห็นจะสามารถค้นหาคำคุณลักษณะที่เป็นคำศัพท์ใหม่ได้พบในประโยคความคิดเห็นอย่างสมบูรณ์

1.1.3 การแทนที่คำที่มีความหมายเหมือนคำคุณลักษณะ (Synonym Feature Words Representation) คือ จัดกลุ่มคำที่มีความหมายเหมือนคำคุณลักษณะเดียวกันให้เป็นคำคุณลักษณะเดียวกัน เช่น “กล้อง” “ถ่ายรูป” “ถ่ายภาพ” “จับภาพ” ซึ่งคำเหล่านี้ในข้อความคิดเห็นจะถูกแทนที่ด้วยคำว่า “กล้องถ่ายรูป” ทุกคำ โดยมีการนำข้อความคิดเห็นมาค้นหากับพจนานุกรมคำคุณลักษณะและคำที่มีความหมายเหมือนคำคุณลักษณะ ถ้าตรงกับคำที่มีความหมายเหมือนคำคุณลักษณะใดให้แทนด้วยคำคุณลักษณะนั้น ในขั้นตอนนี้จะช่วยให้การระบุคำคุณลักษณะและการวิเคราะห์คุณลักษณะรวดเร็วยิ่งขึ้น

1.1.4 การกำกับหน้าที่ของคำ (POS Tagging) คือ การระบุคำที่พบในประโยคความคิดเห็นเป็นคำชนิดใด เช่น “กล้องถ่ายรูป” เป็นคำนาม “ใช้” เป็น คำกริยา และ “ดี” เป็น คำคุณศัพท์ ซึ่งในขั้นตอนนี้จะใช้พจนานุกรม Lexitron taged และเพิ่มคำคุณลักษณะและชนิดของคำคุณลักษณะ โดยมีขั้นตอนคือ นำข้อความคิดเห็นมาค้นหากับพจนานุกรม Lexitron taged ส่วนที่เป็นคำศัพท์และถ้าตรงกับคำศัพท์ใดให้แทนด้วยชนิดคำของคำศัพท์นั้น ถ้าไม่ตรงกับคำใดให้แทนด้วย “0”

1.2 การสกัดและการแทนที่คำสำคัญ (Important Words Extraction and Word Representation) ในขั้นตอนนี้ได้ทำการสกัดคำคุณลักษณะ คำระบุขั้วความคิดเห็น คำนิเสธหรือคำปฏิเสธ และคำเชื่อมประโยค [12], [13] เพื่อนำไปวิเคราะห์หรือจำแนกในระดับคุณลักษณะ โดยมีขั้นตอนดังนี้

1.2.1 การสกัดคุณลักษณะ จะสกัดเอาเฉพาะทำศัพท์ที่มีชนิดของคำเป็น คำนาม (N) คำนามวลี (NP) เช่น “กล้องถ่ายรูป” “แบตเตอรี่” “สี” เป็นต้น โดยมีวิธีการดังนี้ นำข้อความคิดเห็นที่ผ่านขั้นตอนการกำกับหน้าที่ของคำมาคัดเลือกเอาเฉพาะคำนามและคำนามวลีในประโยค เช่น “กล้องถ่ายรูป|N |สวย|V |มาก|ADV” จากข้อความคิดเห็นจะสกัดเอาเฉพาะคำว่า “กล้องถ่ายภาพ” ซึ่งเป็นคำนาม

1.2.2 การสกัดคำระบุขั้วความคิดเห็น จะสกัดเอาเฉพาะคำศัพท์ที่มีชนิดของคำเป็น คำกริยา (V) เช่น “สวย” “หมด” ชนิดคำขยายนาม (ADJ) เช่น “คมชัด” “ดี” “แย่” ชนิดคำขยายกริยา (ADV) เช่น “เลย” “แล้ว” เช่น “กล้องถ่ายรูป|N |สวย|V |มาก|ADV” จากข้อความคิดเห็นจะสกัดเอาเฉพาะคำว่า “สวย” และ “มาก” ซึ่งเป็นคำกริยาและคำขยายกริยา และจากการสังเกตชนิดคำต่าง ๆ ในพจนานุกรม Lexitron tagged พบว่า มีอีก 1 ชนิดคำที่มีคำระบุขั้วความคิดเห็นแฝงอยู่จำนวนมาก คือชนิดคำที่ไม่รู้จัก หรือ Unknow : (UNK) ดังนั้นจึงได้ทำการสกัดชนิดคำที่ไม่รู้จัก (UNK) เช่น “ยอดเยี่ยม” “สุดยอด”

3) การสกัดคำเชิงปฏิเสธ (NEG) เช่น “ไม่” “แต่” ตัวอย่างข้อความคิดเห็นเช่น “กล้องถ่ายรูป|N |ไม่|NEG |สวย|V” จากข้อความคิดเห็นจะสกัดเอาเฉพาะคำว่า “ไม่” ซึ่งเป็นคำเชิงปฏิเสธ

4) การสกัดชนิดคำเชื่อม (CONJ) เช่น “และ” “กับ” ตัวอย่างข้อความคิดเห็น เช่น “กล้องถ่ายรูป|N | และ|CONJ |เครื่อง|N |สวย|V” จากข้อความคิดเห็นจะสกัดเอาเฉพาะคำว่า “และ” ซึ่งเป็นคำเชื่อมจากนั้นแทนที่คำต่าง ๆ ดังตารางที่ 2

ตารางที่ 2 สัญลักษณ์ที่ใช้แทนคำสำคัญ

คำและชนิดของคำ	สัญลักษณ์ที่ใช้แทน
คำคุณลักษณะ	ตัวเลขลำดับคุณลักษณะ
คำระบุขั้วความคิดเห็นเชิงบวก	+
คำระบุขั้วความคิดเห็นเชิงลบ	-
คำเชื่อม	<<
คำเชิงปฏิเสธ	!

โดยนำข้อความคิดเห็นมาค้นหากับชนิดคำที่ต้องการ เมื่อพบคำที่ตรงกันให้แทนด้วยสัญลักษณ์แทนคำสำคัญ ถ้าไม่ตรงกับชนิดใดให้แทนด้วย “0”

1.4 การจำแนกความคิดเห็นในระดับคุณลักษณะ (Feature-based Classification) คือ การวิเคราะห์แต่ละข้อความคิดเห็นว่า ในแต่ละคุณลักษณะมีขั้วความคิดเห็นเป็นเชิงบวกหรือเชิงลบ แต่การระบุขั้วความคิดเห็นว่าเป็นเชิงบวกหรือเชิงลบของแต่ละคุณลักษณะเป็นไปได้ยาก เนื่องจากผู้แสดงความคิดเห็นอาจจะเขียนข้อความคิดเห็นไว้ตำแหน่งในข้อความคิดเห็นก็ได้ หรืออาจจะอยู่ด้านหน้าหรือด้านหลังคำคุณลักษณะนั้นก็ได้ เช่น “กล้องชัดดีมาก แต่แบตเตอรี่ไม่ทน” จากตัวอย่างประโยคนี้พบว่า มีการเขียนคำระบุขั้วความคิดเห็นคือคำว่า “ดี” และ “ไม่ทน” ไว้ด้านหลังของคำคุณลักษณะ “กล้อง” และ “แบตเตอรี่” ตามลำดับ แต่ผู้แสดงความคิดเห็นอาจเขียนคำระบุขั้วความคิดเห็นไว้ด้านหน้าของคำคุณลักษณะก็ได้ เช่น “ดีมากเลยกล้อง แต่แบตเตอรี่ไม่ทน” จากตัวอย่างประโยคข้อความคิดเห็นนี้ พบว่าคำระบุขั้วความคิดเห็น “ดี” ถูกเขียนไว้ด้านหน้าของคำคุณลักษณะ ดังนั้นในงานวิจัยนี้ได้นำเสนอการจำแนกความคิดเห็นโดยวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น ซึ่งเป็นเทคนิควิธีที่ยืดหยุ่นในการระบุคำระบุขั้วความคิดเห็นของแต่ละคุณลักษณะ และได้ทำการเปรียบเทียบกับเทคนิควิธีเดิมคือการจำแนกความคิดเห็นโดยการใช้กฎ และการจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ มาระบุขั้วความคิดเห็น

การจำแนกความคิดเห็นด้วยเทคนิคต่างๆ จำเป็นต้องนำข้อความคิดเห็นมาผ่านกฎเพื่อให้การระบุขั้วความคิดเห็นทำได้ง่ายขึ้นโดยทั้ง 3 วิธี ได้ใช้กฎในการระบุขั้วความคิดเห็น ดังตารางที่ 3

ตารางที่ 3 กฎที่ใช้ในการกระบุขั้วความคิดเห็น

ลำดับกฎ	การสกัดคำคุณลักษณะและการแทนคำคุณลักษณะ	ผลลัพธ์
1	เครื่องเสื่อมเร็ว = |13|-|+|	|13|-|
2	เครื่องดีดี = |14|+|+|	|14|+|
3	กล้องแย่แย่ = |1|-|-|	|1|-|
4	กล้องไม่สวย = |1|!|+|	|1|-|
5	จอละเอียดน้อย = |14|+|-|	|14|-|
6	เครื่องกล้องดี = |13|1|+|	|13|+|1|+
7	เครื่องและกล้องดี = |13|<<|1|+|	|13|+|1|+

เมื่อนำทุกข้อความคิดเห็นมาผ่านกฎในการระบุขั้วความคิดเห็นเหล่านี้แล้ว จะใช้เทคนิควิธีการจำแนกความคิดเห็น 3 วิธีการในการเปรียบเทียบการจำแนกความคิดเห็น

1) การจำแนกความคิดเห็นโดยการใช้กฎ [7] คือ การจำแนกว่าคำระบุขั้วความคิดเห็นที่เป็นเชิงบวกหรือเชิงลบที่ปรากฏอยู่ด้านหลังของคำคุณลักษณะใด จะกำหนดว่าเป็นคำระบุขั้วความคิดเห็นของคำคุณลักษณะนั้น เนื่องจากข้อความคิดเห็นภาษาไทย ส่วนใหญ่จะเขียนคำระบุขั้วความคิดเห็นไว้ด้านหลังของคำคุณลักษณะ เช่น “กล้องถ่ายรูปสวยมาก” ซึ่งหมายความว่า “กล้องถ่ายรูป→เชิงบวก” “แบตเตอรี่ทน” ซึ่งหมายความว่า “แบตเตอรี่→เชิงบวก” “สีแดงสวยมาก” ซึ่งหมายความว่า “สี→เชิงบวก”

2) การจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ [14] คือ การจำแนกว่าคำระบุขั้วความคิดเห็นที่ปรากฏอยู่หน้าคำคุณลักษณะและหลังคำคุณลักษณะ 3 คำ เป็นขั้วบวกหรือขั้วลบและคำระบุขั้วความคิดเห็นเป็นของคุณลักษณะใด การระบุขั้วความคิดเห็นด้วยกฎที่ใช้ในการระบุขั้วความคิดเห็นในตารางที่ 2 ข้อ 1-4 เพื่อระบุขั้วความคิดเห็นให้ถูกต้อง และถ้าในระยะ 3 คำทั้งด้านหน้าและด้านหลังของคำคุณลักษณะมีคำระบุขั้วความคิดเห็นในเชิงบวกมากกว่า ก็จะระบุว่าคุณลักษณะนั้นมีความเห็นในเชิงบวก แต่ถ้าในระยะ 3 คำทั้งด้านหน้าและด้านหลัง มีคำระบุขั้วความคิดเห็นในเชิงลบมากกว่า ก็จะระบุว่าคุณลักษณะนั้นมีความคิดเห็นในเชิงลบ เช่น “เครื่องนี้สวยมากแต่ไม่ดีอะกล้องไม่ชัด” ในคุณลักษณะ “เครื่อง” จะถูกตัดออกมา 3 คำด้านหน้าและด้านหลังจะได้คำว่า “นี้-สวย-มาก” และนับคำเชิงบวกเชิงลบใน 3 คำนี้ ซึ่งจะพบคำเชิงบวกคือ “สวย” 1 คำ ดังนั้นจึงสรุปว่า คำคุณลักษณะ “เครื่อง” เป็นเชิงบวก ส่วนคุณลักษณะ “กล้อง” ก็จะถูกตัดคำโดยรอบหน้าหลังมา 3 คำเช่นกัน โดยมี 3 คำด้านหน้าคือ “แต่-ไม่ดี-อะ” และ 3 คำด้านหลังคือ “ไม่-ชัด” ทำการนำคำเชิงบวกเชิงลบเหล่านี้ พบว่า มีคำเชิงลบ 2 คำ คือ “ไม่ดีและไม่ชัด” จึงสรุปว่า คุณลักษณะ “กล้อง” เป็นเชิงลบ เป็นต้น

3) การจำแนกความคิดเห็นโดยการวัดระยะห่างคำสำคัญ มีแนวคิดมากจากคำคุณลักษณะทุกคำในประโยคความคิดเห็นควรถูกนำมาพิจารณาถึงคำระบุขั้วความคิดเห็นว่ามีขั้วความคิดเห็นที่อยู่ขั้วใด และคำระบุขั้วความคิดเห็นที่อยู่ใกล้คำคุณลักษณะใดจะถือว่าเป็นของคำคุณลักษณะนั้น โดยนำข้อความคิดเห็นมาผ่านกฎในการระบุขั้วความคิดเห็นในตารางที่ 2 ข้อที่ 1-4 เพื่อระบุคำระบุขั้วความคิดเห็นให้ถูกต้อง จากนั้นวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น เช่น “เครื่องนี้สวยมากแต่ไม่ดีอะกล้องไม่ชัด” จากคำคุณลักษณะ “เครื่อง” มีระยะห่างกับคำระบุขั้วความคิดเห็นทั้งหมดในข้อความคิดเห็น ได้แก่ “สวย-ไม่ดี-ไม่ชัด” พบว่า “สวย” อยู่ใกล้คำคุณลักษณะ “เครื่อง” มากที่สุด และคำว่า “สวย” จึงระบุว่าคุณลักษณะ “เครื่อง” เป็นเชิงบวก ส่วนคุณลักษณะ “กล้อง” ก็จะนำมาวัดระยะห่างกับคุณระบุขั้วความคิดเห็นทุกคำเช่นกัน “สวย-ไม่ดี-ไม่ชัด” ซึ่งพบว่าคำว่า “ไม่ชัด” อยู่ใกล้คำคุณลักษณะ “กล้อง” มากที่สุด และจะสรุปว่า คุณลักษณะ “กล้อง” เป็นเชิงลบ เป็นต้น

1.5 การวัดประสิทธิภาพ (Performance)

ในงานวิจัยนี้ได้ทำการวัดประสิทธิภาพในการจำแนกความคิดเห็นในระดับคุณลักษณะด้วยวิธีการต่างๆ คือ ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และค่าประสิทธิภาพโดยรวม (F-Measure) [15] ซึ่งค่า Precision จะเป็นค่าที่ระบุได้ว่า การทำนายว่าข้อความคิดเห็นเป็นเชิงบวกหรือเชิงลบและการทำนายว่าคำระบุขั้วความคิดเห็นเป็นของคำคุณลักษณะมีความแม่นยำหรือไม่ และค่า Recall เป็นค่าที่ระบุได้ว่า ทำนายว่าข้อความคิดเห็นเป็นเชิงบวกหรือเชิงลบ และการทำนายว่าคำระบุขั้วความคิดเห็นเป็นของคำคุณลักษณะได้ถูกต้องเท่าไหร่จากค่าความจริงของผลเฉลย ส่วน F-Measure เป็นการวัดประสิทธิภาพโดยรวมของการจำแนก งานวิจัยนี้ไม่ได้วัดประสิทธิภาพค่าความถูกต้อง (Accuracy) เนื่องจากเป็นการใช้วิธีการวัดประสิทธิภาพเหมือนกันกับค่าความระลึก

ค่าความแม่นยำ $(Precision)$

$$= \frac{\text{จำนวนที่ทำนายถูกต้อง}}{\text{จำนวนผลลัพธ์ที่ได้จากการทำนายทั้งหมด}}$$

ค่าความระลึก (Recall) $$= \frac{\text{จำนวนที่ทำนายถูกต้อง}}{\text{จำนวนผลเฉลยทั้งหมด}}$$

ค่าประสิทธิภาพโดยรวม $(\text{F} - \text{Measure})$ $$= \frac{2 \text{ x Precision x Recall}}{\text{Precision} + Recall}$$

4. การวิเคราะห์ข้อมูล

การทดลองนี้ได้ทำการวัดประสิทธิภาพ โดยหาค่า ความแม่นยำ (Precision) ความระลึก (Recall) และประสิทธิภาพโดยรวม (F-Measure) โดยมีผลลัพธ์ที่ได้จากการทดลองดังนี้

ผลการวิจัย

จากการจำแนกความคิดเห็นโดยเปรียบเทียบกันทั้ง 3 เทคนิควิธี โดยวิธีที่ 1 คือ จำแนกด้วยกฎ วิธีที่ 2 คือ จำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ และวิธีที่ 3 คือ การวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น ซึ่งเป็นวิธีที่ได้นำเสนอในงานวิจัยนี้ โดยมีผลลัพธ์ดังตารางที่ 4 เปรียบเทียบประสิทธิภาพการทดลอง

ตารางที่ 4 เปรียบเทียบประสิทธิภาพการทดลอง

ประสิทธิภาพ	วิธีที่ 1	วิธีที่ 2	วิธีที่ 3
Precision	79.12%	81.88%	75.15%
Recall	68.40%	67.39%	78.40%
F-Measure	73.37%	73.93%	76.74%

จากตารางที่ 4 แสดงให้เห็นถึงประสิทธิภาพของแต่ละการทดลองดังนี้ การทดลองวิธีที่ 1 ค่า Precision คือ 79.12% และค่า Recall คือ 68.40% และค่า F-Measure คือ 73.37% การทดลองวิธีที่ 2 ค่า Precision คือ 81.88% และค่า Recall คือ 67.39% และค่า F-Measure คือ 73.93% การทดลองวิธีที่ 3 ค่า Precision คือ 75.15% และค่า Recall คือ 78.40% และค่า F-Measure คือ 76.74%

อภิปรายผลการวิจัย

จากผลการทดลองพบว่า ประสิทธิภาพความแม่นยำ เป็นผลมาจากจำนวนเหตุการณ์ของการทำนายที่ถูกต้องหารด้วยจำนวนเหตุการณ์ของการทำนายทั้งหมด ดังนั้นจำนวนเหตุการณ์ของการทำนายทั้งหมดจึงมีผลต่อค่าความแม่นยำโดยตรง จากการทดลองของจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ คือ 81.90% ซึ่งสูงกว่า จำแนกด้วยกฎคือ 79.03% และการวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น คือ 75.09% เป็นผลมาจากการจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและ

ด้านหลังอย่างละ 3 คำ มีจำนวนเหตุการณ์ที่สามารถระบุคุณลักษณะและขั้วความคิดเห็นน้อยกว่าการจำแนกโดยใช้กฎ และการจำแนกโดยใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น เนื่องจากในประโยคความคิดเห็นไม่มีคำระบุขั้วความคิดเห็นอยู่ใกล้คำคุณลักษณะนั้นเลย เพราะประโยคความคิดเห็นภาษาไทยไม่มีรูปแบบการเขียนที่เป็นอิสระไม่มีรูปแบบตายตัว เช่น “กล้อง เครื่อง แบตเตอรี่ ไวไฟและเสียงห่วย” จะทำให้ไม่พบคำระบุขั้วความคิดเห็นของคำคุณลักษณะ “กล้อง” “เครื่อง” “แบตเตอรี่” ได้ และทำให้จำนวนเหตุการณ์ของการทำนายมีจำนวนน้อยลง หรือมีความแม่นยำเพิ่มขึ้น ส่วนการจำแนกความคิดเห็นโดยการใช้กฎ มีประสิทธิภาพความแม่นยำสูงกว่า การจำแนกความคิดเห็นโดยใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น เป็นผลมาจากจำนวนเหตุการณ์ในการทำนายน้อยกว่าเช่นกัน เพราะการจำแนกความคิดเห็นโดยใช้กฎไม่สามารถระบุขั้วความคิดเห็นของคุณลักษณะของประโยคที่เขียนคำระบุขั้วความคิดเห็นทั้งด้านหน้าและด้านหลังคำคุณลักษณะได้ เช่น “ใช้ดีเครื่องนี้แต่สีไม่สวย” “สวยมากสีนี้เครื่องก็ดีอะ” เป็นต้น ประโยคความคิดเห็นเหล่านี้คำระบุขั้วความคิดเห็นไม่ได้อยู่ด้านหน้าหรือด้านหลังของคำคุณลักษณะอย่างใดอย่างหนึ่ง ส่วนใหญ่แล้วข้อความภาษาไทยเขียนคำระบุขั้วความคิดเห็นไว้ด้านหลังของคำคุณลักษณะ เช่น “เครื่องดี” “กล้องถ่ายสวย” “แบตเตอรี่ทน” ดังนั้นจึงเลือกเอาคำระบุขั้วความคิดเห็นที่อยู่ด้านหลังของคำคุณลักษณะ ซึ่งเป็นผลทำให้จำนวนเหตุการณ์ของการทำนายทั้งหมดของการจำแนกโดยใช้กฎน้อยกว่า และในทางตรงกันข้ามเทคนิคการจำแนกความคิดเห็นโดยใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น ทุกคำคุณลักษณะจะถูกนำไปพิจารณาถึงแม้คำคุณลักษณะดังกล่าวไม่มีคำระบุขั้วความคิดเห็น หรือเขียนคำระบุขั้วความคิดเห็นผิด เช่น “กล้องก็ธรรมดา แต่แบตเตอรี่อย่างทน” “วิ๊ง ๆ ในความเป็นกล้อง แต่แบตเตอรี่แย่” จากประโยคความคิดเห็นแรก คำว่า “ธรรมดา” ไม่ได้สื่อความหมายในด้านบวกหรือด้านลบ แต่คำว่า “ทน” เป็นคำในเชิงบวก ก็จะถูกวัดระยะห่างและถูกเลือกเป็นคำระบุขั้วความคิดเห็นของคุณลักษณะ “กล้อง” เนื่องจากไม่มีคำระบุขั้วความคิดเห็นที่มีระยะห่างน้อยกว่าคำนี้ และประโยคความคิดเห็นที่ 2 คำว่า “วิ๊ง ๆ” เป็นคำแสลง หมายถึง “ถ่ายรูปสวย” เป็นคำเชิงบวกของคุณลักษณะ “กล้อง” แต่ไม่สามารถตรวจพบคำดังกล่าว ดังนั้นจึงนำคำว่า “แย่” เป็นคำระบุขั้วความคิดเห็นในเชิงลบมาพิจารณา และทำให้การทำนายคุณลักษณะ “กล้อง” เป็นเชิงลบด้วย ดังนั้นจำนวนเหตุการณ์การทำนายของการจำแนกโดยใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็นจึงมีจำนวนมากกว่า

ประสิทธิภาพความระลึก เป็นผลมาจากจำนวนเหตุการณ์ทำนายถูกต้องหารด้วยเหตุการณ์ความเป็นจริงของผลเฉลยทั้งหมด การวัดประสิทธิภาพด้วยเทคนิควิธีนี้จะเกี่ยวข้องกับค่าความจริงหรือผลเฉลยโดยตรง ดังนั้นค่าประสิทธิภาพความระลึกจึงเป็นค่าประสิทธิภาพที่มีความสำคัญ คือ 78.69% ซึ่งสูงกว่าการจำแนกด้วยกฎ โดยมีค่าความระลึก และการวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น คือ 69.17% และสูงกว่าการจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ คือ 68.04% ดังนั้นจึงสรุปว่าการจำแนกความคิดเห็นโดยการวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็นมีประสิทธิภาพความถูกต้องสูงกว่า

ประสิทธิภาพโดยรวม การวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น มีค่าประสิทธิภาพโดยรวม 76.85% สูงกว่าการจำแนกด้วยกฎ โดยมีค่าประสิทธิภาพโดยรวมคือ 73.77% และการจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ โดยมีค่าประสิทธิภาพโดยรวมคือ 74.32% เนื่องจากจำนวนเหตุการณ์การทำนายในการการวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็นมีจำนวนมากกว่าการจำแนกด้วยกฎ และการจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ และการการวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น มีการทำนายผิดพลาดน้อยกว่าการจำแนกด้วยกฎ และการจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ เป็นผลมาจากเทคนิคการจำแนกความคิดเห็นโดยใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น มีการระบุคุณลักษณะและขั้วความคิดเห็นความคิดเห็นอย่างยืดหยุ่น สามารถระบุคุณลักษณะและขั้วความคิดเห็นประโยคความคิดเห็นที่เขียนคำระบุขั้วความคิดเห็นไว้ทั้งด้านหน้าและด้านหลังของคำคุณลักษณะได้ทำให้มีจำนวนเหตุการณ์ทำนายได้ถูกต้องสูงกว่าการจำแนกความคิดเห็นโดยใช้กฎ ตัวอย่างประโยคความคิดเห็น

เช่น “สวยมากครับผมใช้สีขาวอยู่กล้องชัดมาก” “Mirror5 ถ่ายแล้วภาพนี้สวยมาก” เป็นต้น จากตัวอย่างประโยคความคิดเห็นที่ 1 คำว่า “สวยมากผมใช้สีขาว” ซึ่งกล่าวถึงคุณลักษณะของสมาร์ตโฟนคือ “สี” และมีความคิดเห็นเป็นเชิงบวก แต่มีการเขียนคำระบุขั้วความคิดเห็นไว้ด้านหน้าของคำคุณลักษณะ การจำแนกโดยใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็นสามารถระบุคำระบุขั้วความคิดเห็นที่อยู่ด้านหน้าของคำคุณลักษณะได้โดยการใช้ระยะห่าง ดังนั้นประสิทธิภาพความระลึกและค่าประสิทธิภาพโดยรวมของการจำแนกโดยใช้ระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น จึงมีประสิทธิภาพสูงกว่าการจำแนกโดยใช้กฎและการจำแนกความคิดเห็นโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ เนื่องจากสามารถระบุคุณลักษณะและขั้วความคิดเห็นเหตุการณ์ได้อย่างถูกต้องและยืดหยุ่นสูงกว่าเทคนิควิธีอื่น

จากผลการทดลองกับข้อมูลความคิดเห็นของแต่ละยี่ห้อของสมาร์ตโฟนพบว่า การจำแนกความคิดเห็นโดยวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็น เป็นวิธีที่นำเสนอมีประสิทธิภาพมากกว่าการจำแนกโดยใช้กฎ [11] และการจำแนกโดยตัดคำที่อยู่ใกล้คำคุณลักษณะทั้งด้านหน้าและด้านหลังอย่างละ 3 คำ [10] เนื่องจากข้อความคิดเห็นภาษาไทยที่เขียนบนเว็บไซต์เครือข่ายสังคมออนไลน์เป็นรูปแบบการเขียนที่เป็นอิสระไม่มีโครงสร้างทางภาษาชัดเจน การจำแนกโดยการวัดระยะห่างระหว่างคำคุณลักษณะและคำระบุขั้วความคิดเห็นเป็นเทคนิควิธีที่มีการพิจารณาทุกคำในประโยคความคิดเห็น จึงมีประสิทธิภาพสูงกว่า

ข้อเสนอแนะ

การวิเคราะห์ความคิดเห็นจะมีประสิทธิภาพสูงขึ้นได้จะต้องทำการเพิ่มคำศัพท์คำแสลง หรือคำที่แสดงถึงความรู้สึก ที่เป็นคำศัพท์ใหม่ๆ ให้ก้าวทันภาษาเขียนของผู้แสดงความคิดเห็น จะทำให้เทคนิคการสกัดคำคุณลักษณะละคำระบุขั้วความคิดเห็นมีประสิทธิภาพมากขึ้น และยังมีบางคำศัพท์เป็นคำที่แสดงถึงความรู้สึกและคุณลักษณะในตัว เช่น “หนัก” หมายถึง คุณสมบัติด้าน “น้ำหนัก” และเป็นการแสดงความรู้สึกในด้านลบ และ“ค้าง” หมายถึง “เครื่องค้าง” เป็นคำที่แสดงถึงคุณสมบัติเกี่ยวกับเครื่องในด้านลบ ดังนั้นการสกัดคำสำคัญต่าง ๆ จึงมีส่วนสำคัญในการวิเคราะห์ข้อความคิดเห็น และพัฒนาวิธีการจำแนกความคิดเห็นในระดับคำคุณลักษณะ โดยการระบุคำระบุขั้วความคิดเห็นของคำคุณลักษณะให้มีประสิทธิภาพมากยิ่งขึ้น

กิตติกรรมประกาศ

ขอขอบคุณกองส่งเสริมการวิจัยและบริการวิชาการ มหาวิทยาลัยมหาสารคาม ที่ให้ทุนสนับสนุนการทำวิจัย

เอกสารอ้างอิง

[1] Cisco. (2014). *Online Knowledge*. Retrieved from http://www.thailandonlineexpo.com/news/detail/190

[2] กองบรรณาธิการ PositioningMagazine. (2556). *พฤติกรรมการใช้ smartphones ของคนไทย*. สืบค้นจาก http://www.positioning mag .com/content/คนไทยใช้สมาร์ตโฟน-ถ่ายรูป-ฟังเพลง-มากกว่าเข้าเน็ท-ติดแบรนด์-ฟังก์ชั่นต้องจัดเต็ม.

[3] วิยะดา ฐิติมัชฌิมา. (2553). เครือข่ายสังคมออนไลน์: แนวโน้ม ปรากฏการณ์ และจริยธรรม. *วารสารนักบริหาร*, 30(4), 150-156

[4] Liu B. (2010). *Sentiment Analysis and Subjectivity*. Handbook of Natural Language Processing. 1-39

[5] วรัญญา วรรณศรี, และเด่นดวง ประดับสุวรรณ. (2553). *ระบบวิเคราะห์ข้อความแสดงความคิดเห็นสำหรับโรงแรม*. กรุงเทพฯ: มหาวิทยาลัยธรรมศาสตร์.

[6] นริศร์ พรหมบุตร. (2550). *การทำไมน์นิ่งความคิดเห็นเกี่ยวกับสินค้า: กรณีศึกษาโทรศัพท์มือถือ.* (สารนิพนธ์ปริญญามหาบัณฑิต, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ.

[7] Wang H, Liu L & Song W. (2014). Feature-based Sentiment Analysis Approach for Product Reviews. *Journal of software.* 9(2). 274-279.

[8] Lui, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lecturer on Human Language Technology, 5(1), 1-67

[9] Lui B, Hu M, & Cheng J. (2005). Opinion Observer: Analyzing and Comparing Opinions on the Web. In Proceedings of the 14th *International Conference on World Wide Web.* pp. 342-351. ACM.

[10] Mukherjee S, & Bhattacharyya P (2012). *Feature Specific Sentiment Analysis for Product Reviews*. Computational Linguistics and Intelligent Text Processing. Lecture Note in Computer Science, pp. 475-487.

[11] NECTEC. (2558). *โปรแกรมตัดคำสำหรับข้อความภาษาไทย LexTo Thai Lexeme Tokenizer*. Retrieved from http://www.sansarn.com/lexto/

[12] มหาวิทยาลัยราชภัฏมหาสารคาม. (2558). *รายงานผลการดำเนินงานการส่งเสริมเผยแพร่และพัฒนาต่อยอดสื่อ eDLTV*. มหาสารคาม: ผู้แต่ง.

[13] Ravi K. V. & Raghuveer K. (2012). Web User Opinion Analysis for Product Features Extraction and Opinion Summarization. *International Journal of Web & Semantic Technology*, 3(4), 69-82.

[14] Aravindan S. & Ekbal A. (2014, December). Feature Extraction and Opinion Mining in Online Product Reviews. *In International Conference on Information Technology*, pp. 94-99. IEEE.

[15] พลอยพรรณ สอนสุวิทย์. (2560). การพัฒนาขั้นตอนวิธีในระบบตรวจจับการบุกรุกทางเครือข่ายด้วยเอดาบูทเอ็มวัน. *วารสารวิชาการการจัดการเทคโนโลยีสารสนเทศและนวัตกรรม.* 4(2). 158-166.