

การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์ OPINION MINING FROM ONLINE SOCIAL NETWORKS

กานดา แผล้วฒนากุล

มหาวิทยาลัยนอร์ทกรุงเทพ

ดร. ปราโมทย์ ลือนาม

ผู้ช่วยศาสตราจารย์ประจำสถาบันบัณฑิตพัฒนบริหารศาสตร์

บทคัดย่อ

ความคิดเห็นและข้อเสนอแนะที่แพร่หลายจำนวนมากในเครือข่ายสังคมออนไลน์เป็นประโยชน์ต่อธุรกิจ ทำให้ทราบความรู้สึกของผู้บริโภคทั้งในแง่บวกและลบ ต่อคุณลักษณะของผลิตภัณฑ์ในแต่ละด้าน อันจะนำไปสู่การปรับปรุงคุณภาพของสินค้าหรือบริการ การวิเคราะห์ความคิดเห็นจำนวนมากมหาศาลได้อย่างรวดเร็วสามารถทำได้โดยอาศัยระบบคอมพิวเตอร์ที่ประมวลผลชุดขั้นตอนวิธี การทำเหมืองความคิดเห็น (opinion mining) แบ่งเป็น 2 ขั้นตอนหลัก ได้แก่ (1) ขั้นตอนการแทนข้อความให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถนำไปประมวลผลได้ และ (2) ขั้นตอนการจำแนกประเภทความคิดเห็นออกเป็นขั้ว ความคิดเห็นเชิงบวก เชิงลบ หรือเป็นกลาง ทำให้ได้ข้อมูลที่จะช่วยเพิ่มความสามารถในการแข่งขัน ลดระยะเวลาและค่าใช้จ่ายในการสำรวจตลาด รวมถึงปรับกลยุทธ์สนองตอบต่อความต้องการของผู้บริโภคที่เปลี่ยนแปลงอย่างรวดเร็ว บทความนี้กล่าวถึงแนวคิดพื้นฐาน และการประยุกต์ใช้เหมืองความคิดเห็น รวมทั้งทบทวนวรรณกรรมที่เกี่ยวข้อง พร้อมคำวิจารณ์ และข้อเสนอแนะในการศึกษาวิจัย

คำสำคัญ: การวิเคราะห์เหมืองความคิดเห็น การจำแนกข้อความ เครือข่ายสังคมออนไลน์

ABSTRACT

Nowadays, opinions on products and services on social networking sites, such as Facebook and Twitter, are considered to be beneficial for all businesses. They were proved to be very helpful as they can help businesses know customer needs and then can prompt response to them. Opinion Mining has two major tasks: (1) transforming of opinions into manageable and computable forms, and (2) classifying the opinions into negative, positive or neutral groups. The method can not only save time and reduce effort for evaluating such large amounts of data but also help businesses improve their quality of products and services. This article reviews concepts and applications of opinion mining. Prior related studies, suggestion and criticism are also provided.

Keywords: Opinion Mining, Text Classification, Social Networks

บทนำ

ปัจจุบันผู้บริโภคในยุคดิจิทัลเริ่มเปลี่ยนพฤติกรรมการติดต่อสื่อสารกับธุรกิจ จากช่องทางเดิมอย่างโทรศัพท์ จดหมาย มาใช้เครือข่ายสังคมออนไลน์ อาทิ เฟสบุ๊ค ทวิตเตอร์ ในการแสดงความคิดเห็น (opinion) และข้อเสนอแนะ (suggestion) ต่อสินค้าหรือบริการ รวมถึงสอบถามปัญหาต่างๆ ผ่านทางเครือข่ายสังคมออนไลน์มากขึ้น ซึ่งข้อความดังกล่าวมีอิทธิพลอย่างมากต่อผู้ใช้เครือข่ายสังคมออนไลน์ รวมถึงภาคธุรกิจที่ต้องการเป็นผู้นำตลาดในยุคที่มีการแข่งขันสูงเช่นนี้ ภาคธุรกิจจำเป็นต้องคำนึงถึงความพึงพอใจของลูกค้าเป็นสำคัญ ต้องติดตามความคิดเห็น (brand monitoring) รับฟังข้อเสนอแนะของผู้บริโภคอย่างสม่ำเสมอ เพื่อแก้ไขปรับปรุง และเพิ่มประสิทธิภาพการบริหารลูกค้าสัมพันธ์ โดยเฉพาะอย่างยิ่งข้อความแสดงความคิดเห็นเชิงลบที่มีต่อสินค้าหรือบริการจะถูกกระจายตัวอย่างรวดเร็วซึ่งส่งผลเสียอย่างมากต่อธุรกิจ นอกจากนี้ยังรวมถึงข้อคำถามหรือข้อสงสัยของผู้บริโภคที่ต้องการคำตอบจากผู้เกี่ยวข้อง หากข้อความเชิงลบ ข้อเสนอแนะและข้อคำถามต่างๆ ได้รับการตอบสนอง จะเป็นอีกช่องทางหนึ่งในการเข้าถึงผู้บริโภค ซึ่งสามารถช่วงชิงความได้เปรียบเชิงการแข่งขันจากคู่แข่งในตลาดได้ แต่ทว่าข้อความที่แสดงความคิดเห็นบนเครือข่ายสังคมออนไลน์มีอยู่จำนวนมาก หลากหลายเว็บไซต์ การค้นหาความคิดเห็น ข้อเสนอแนะและข้อคำถามที่เกี่ยวข้องกับเรื่องที่สนใจทำได้ยาก ต้องเสียทั้งเวลาในการติดตามและค่าใช้จ่ายสำหรับคนที่คอยติดตามความคิดเห็น ดังนั้นจะอย่างไรให้ภาคธุรกิจสามารถติดตามและตอบสนองต่อความคิดเห็นต่างๆ ได้อย่างรวดเร็ว

การวิเคราะห์ความคิดเห็นบนเครือข่ายสังคมออนไลน์โดยใช้ความสามารถของคอมพิวเตอร์สามารถช่วยลดปัญหาตามที่กล่าวมาได้ แต่ความท้าทายหนึ่งสำหรับการวิเคราะห์ความคิดเห็นคือ ข้อความแสดงความคิดเห็นส่วนใหญ่ที่อยู่บนอินเทอร์เน็ตนิยมใช้ภาษาที่มีโครงสร้างประโยคที่ไม่แน่นอน (unstructured data) หรือเป็นภาษาธรรมชาติ (natural language) ไม่ถูกต้องตามหลักไวยากรณ์ทางภาษา ทำให้ยากต่อการวิเคราะห์งานวิจัยด้านการวิเคราะห์ความคิดเห็น จึงได้นำเทคนิคการวิเคราะห์เหมืองข้อความ (text mining) และการประมวล

ภาษาธรรมชาติ (Natural Language Processing: NLP) มาประยุกต์ใช้ เพื่อวิเคราะห์ความคิดเห็นของผู้บริโภค ซึ่งเรียกว่าการวิเคราะห์เหมืองความคิดเห็น (opinion mining) หรือการวิเคราะห์ความรู้สึก (sentiment analysis) ในงานวิจัยของ Peter (2002), Mingqing Hu and Bing Liu (2004), Bo Pang and Lillian Lee (2008), Alisa Kongthong, Niran Angkawat tanawit, Chatchawal Sangkeettrakarn, Pornpimon Palingoon and Choochart Haruechaiyasak (2010) และ Khampol Sukhum, Supot Nitsuwat and Choochart Haruechaiyasak (2011) ได้นำเสนอการวิเคราะห์ความพึงพอใจของผู้บริโภคในหัวข้อเรื่องที่สนใจ อาทิ สินค้า บริการ การเมือง ซึ่งบทความนี้จะอธิบายแนวคิด ทฤษฎี เทคนิคต่าง ๆ ที่เกี่ยวข้อง รวมถึงกระบวนการในการวิเคราะห์การสร้างแบบจำลอง และแสดงตัวอย่างจากงานศึกษาวิจัยที่เกี่ยวข้องกับการวิเคราะห์เหมืองความคิดเห็น

การวิเคราะห์เหมืองความคิดเห็น

การวิเคราะห์เหมืองความคิดเห็นเป็นกระบวนการวิเคราะห์ข้อคิดเห็นที่ผู้บริโภคนำเสนอในรูปของข้อความ ซึ่งพบได้บนเครือข่ายสังคมออนไลน์ เพื่อให้ทราบถึงความพึงพอใจในแง่บวกหรือลบ โดยกระบวนการดังกล่าวได้นำความรู้ทางด้านการค้นหาลักษณะแฝงของข้อมูล (knowledge discovery) หรือเรียกอีกชื่อหนึ่งว่าการวิเคราะห์เหมืองข้อมูล (data mining) มาประยุกต์ใช้ แต่เนื่องจากข้อมูลที่นำมาวิเคราะห์เหมืองความคิดเห็นนั้นมีลักษณะเป็นข้อความ กระบวนการค้นหาลักษณะแฝงจึงถูกเรียกว่าการวิเคราะห์เหมืองข้อความ (text mining) ดังนั้นการวิเคราะห์เหมืองความคิดเห็นจึงจำเป็นต้องนำความรู้ด้านการวิเคราะห์เหมืองข้อความมาใช้ นอกจากนี้ข้อความที่นำมาวิเคราะห์มีลักษณะโครงสร้างไม่แน่นอน ซึ่งแตกต่างกันตามธรรมชาติของการเรียนรู้ของสมองมนุษย์แต่ละคน (กนกวรรณ เขียววรรณ, 2555) เรียกว่าภาษาธรรมชาติ จึงต้องนำความรู้ในเรื่องการประมวลผลภาษาธรรมชาติมาใช้ร่วมด้วย

ดังนั้นในบทความนี้จะกล่าวถึงทฤษฎีที่ใช้ในการวิเคราะห์เหมืองความคิดเห็น 2 เรื่องหลักคือ (1) การประมวลผลภาษาธรรมชาติ และ (2) การวิเคราะห์เหมืองข้อความ

การประมวลผลภาษาธรรมชาติ

การประมวลผลภาษาธรรมชาติ คือการแปลความจากภาษาธรรมชาติที่มนุษย์ใช้สื่อสารกันให้อยู่ในรูปแบบที่เป็นโครงสร้าง (structured data) ที่เครื่องคอมพิวเตอร์สามารถเข้าใจได้ 2 แนวทาง (ยีน ภูววรรณ, 2535) คือ แนวทางการศึกษาและทำความเข้าใจกับโครงสร้างทางภาษาศาสตร์ และอีกแนวทางคืออาศัยความรู้ด้านปัญญาประดิษฐ์ โดยการแทนความรู้ (knowledge representation) ด้วยคลังคำ (corpus) ซึ่งจะกล่าวในรายละเอียดต่อไป

สำหรับข้อความภาษาไทยจำเป็นต้องตัดประโยคออกเป็นคำก่อน เนื่องจากภาษาไทยมีรูปแบบการเขียนคำยาวต่อกัน ไม่มีการเว้นวรรคระหว่างคำดังเช่นภาษาอังกฤษ โปรแกรมตัดคำสำหรับข้อความภาษาไทยที่เปิดให้ดาวน์โหลดได้ฟรี ซึ่งพัฒนาโดยหน่วยปฏิบัติการวิทยาการมนุษย์ภาษา ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) คือ โปรแกรมตัดคำ LexTo (Thai Lexeme Tokenizer) และ TLexs (Thai Lexeme Analyser) เมื่อข้อความถูกตัดคำออกมาด้วยโปรแกรมแล้ว จึงนำคำมาวิเคราะห์และแทนข้อความให้อยู่ในรูปแบบที่เป็นโครงสร้าง ด้วยแนวทางใดแนวทางหนึ่งหรือทั้งสองแนวทางร่วมกัน

การประมวลผลภาษาธรรมชาติโดยอาศัยโครงสร้างทางภาษาศาสตร์ ที่ประกอบด้วยหน่วยย่อยที่เล็กที่สุดคือหน่วยคำ (morpheme) ซึ่งอาจมีหรือไม่มี ความหมายก็ได้ หน่วยคำหลายหน่วยคำประกอบกันเป็นคำ (word) ที่มีความหมาย คำหลายคำประกอบกันเป็นวลี (phrase) ซึ่งแบ่งเป็นนามวลี (Noun Phrase: NP) และกริยาวลี (Verb Phrase: VP) วลีหลายวลีประกอบกันเป็นประโยค (Sentence: S) ซึ่งอธิบายได้ด้วยไวยากรณ์ (grammar) ตามหลักภาษาศาสตร์ หรือเรียกว่าสูตรไวยากรณ์ (grammar formalism) สูตรไวยากรณ์ที่นิยมใช้มากที่สุด ได้แก่ไวยากรณ์แบบไม่พึ่งบริบท (Context Free Grammar: CFG) กล่าวคือไม่พิจารณาถึงความหมายของคำ แต่จะอาศัยหน้าที่ของคำเป็นหลัก คือ คำนาม (N) คำสรรพนาม (PRON) คำกริยา (V) คำกริยาช่วย (AUX) คำคุณศัพท์ (ADJ) คำกริยาวิเศษณ์ (ADV) และคำบุพบท (PREP) แล้วจึงอธิบายตามกลุ่มคำนาม (นามวลี) และ

กลุ่มคำกริยา (กริยาวลี) ในการแบ่งประโยคออกเป็น ส่วนดังนี้ (สมนึก สินธุปวน, 2546)

$$S = NP + VP$$

$$NP = N \mid N + (ADJ) + (ADV) + (PP) \mid PRON$$

$$VP = V \mid V + (ADV) \mid AUX + V \mid VP + NP$$

$$PP = PREP + NP \mid PREP + VP \mid PREP + NP + VP$$

การประมวลผลภาษาธรรมชาติโดยอาศัยการแทนความรู้ด้วยคลังคำ เนื่องจากภาษาไทยมีรูปแบบไม่ตายตัว การวิเคราะห์ด้วยโครงสร้างทางภาษาศาสตร์ไม่สามารถวิเคราะห์ประโยคได้ทั้งหมด ประกอบกับภาษาไทยมีคำเกิดขึ้นใหม่มากมาย และบางคำเป็นคำเจาะจงเฉพาะหัวข้อใดหัวข้อหนึ่ง จึงต้องใช้วิธีการรวบรวมคำจากเอกสาร ประโยค จากแหล่งข้อมูลที่เกี่ยวข้อง โดยให้ผู้เชี่ยวชาญ (domain experts) พิจารณาและนำมาเก็บไว้ในคลังคำ (corpus) ปัจจุบันมีหลายหน่วยงานที่เก็บรวบรวมคลังคำภาษาไทยเพื่อใช้ในการวิจัย และขยายผลในเชิงพาณิชย์ ได้แก่ ORCHID Corpus ของ NECTEC ซึ่งรวบรวมคำศัพท์มาจากบทความนิตยสาร และกำกับหน้าที่ของคำไว้แล้ว Lexitron Corpus ของ NECTEC และ NAIst Corpus ที่รวบรวมคำศัพท์โดยมหาวิทยาลัยเกษตรศาสตร์ เป็นต้น

การวิเคราะห์เหมืองข้อความ

ข้อมูลแบ่งออกเป็น 2 รูปแบบคือ (1) ข้อมูลที่เป็นโครงสร้าง (structured data) ซึ่งการประมวลผลข้อมูลที่เป็นโครงสร้างเรียกว่าการวิเคราะห์เหมืองข้อมูล และ (2) ข้อมูลที่ไม่เป็นโครงสร้าง (unstructured data) หรือไม่มีโครงสร้างที่ชัดเจน (implicit structured data) ซึ่งส่วนใหญ่ มักอยู่ในรูปแบบของข้อความหรือภาษาธรรมชาติ และเรียกกระบวนการวิเคราะห์ข้อความว่าการวิเคราะห์เหมืองข้อความ (text mining) โดยเป็นการนำความรู้ด้านการวิเคราะห์เหมืองข้อมูลมาประยุกต์ใช้

หน่วยงาน Cross Industry Standard Process for Data Mining (CRISP DM) ได้นำเสนอกระบวนการวิเคราะห์เหมืองข้อมูล อันประกอบด้วยขั้นตอนดังต่อไปนี้

- (1) การทำความเข้าใจกับธุรกิจและระบุปัญหาของงาน (business understanding) (2) การรวบรวม

ข้อมูลและพิจารณาความถูกต้อง เหมาะสมของข้อมูล (data understanding) (3) การเตรียมข้อมูล (data preparation) (4) การสร้างแบบจำลองการวิเคราะห์ข้อมูล (modeling) (5) การประเมินหรือวัดประสิทธิภาพของแบบจำลองการวิเคราะห์ข้อมูล (evaluation) และ (6) การนำผลลัพธ์ที่ได้จากการวิเคราะห์ไปใช้งานจริง (development) สำหรับกระบวนการวิเคราะห์เหมืองข้อความมีกระบวนการเพิ่มเติมจากการกระบวนการวิเคราะห์เหมืองข้อมูลในขั้นตอนที่ 3 คือการเตรียมข้อมูล เนื่องจากข้อความมีลักษณะไม่เป็นโครงสร้าง ดังนั้นในขั้นตอนการเตรียมข้อมูลจำเป็นต้องแปลงรูปแบบของข้อความที่ไม่มีโครงสร้างให้เป็นโครงสร้างที่แน่นอนก่อน (Ronen and James, 2007) เพื่อให้คอมพิวเตอร์สามารถนำไปประมวลผลต่อไป และเพื่อให้ได้โครงสร้างข้อมูลที่เหมาะสมสำหรับขั้นตอนวิธี (algorithm) ที่จะใช้ในการวิเคราะห์ต่อไป

การเตรียมข้อมูล

กระบวนการเตรียมข้อมูลสำหรับวิเคราะห์เหมืองข้อความ ประกอบด้วยกระบวนการหลัก 3 กระบวนการ ได้แก่ การเลือกคุณลักษณะ (feature selection) การกลั่นกรองข้อความ (text cleaning) และ การแทนข้อความ (text representation) โดยอาศัยความรู้จากการประมวลผลภาษาธรรมชาติ มีรายละเอียดดังนี้

(1) กระบวนการเลือกคุณลักษณะเป็นกระบวนการเบื้องต้นในการเลือกคำสำคัญ เทคนิคพื้นฐานที่นิยมใช้คือการตัดคำที่ไม่มีนัยสำคัญออก คำที่ไม่มีนัยสำคัญ (stop words) ได้แก่ คำบุพบท คำสันธาน สรรพนาม ลักษณะนาม ตัวเลข รวมถึงคำลงท้ายประโยคในภาษาไทย (Khampol Sukhum, Supot Nitsuwat and Choochart Haruechaiyasa, 2011)

(2) การกลั่นกรองข้อความ เป็นกระบวนการแก้ไขคำผิดให้ถูกต้อง และแก้ไขคำซ้ำซ้อนหรือคำที่มีความหมายเดียวกัน (synonym word) ให้เป็นคำเดียวกัน รวมถึงแก้ไขข้อความบางส่วนที่ขาดหายไปสมบูรณ์ ดังประโยคตัวอย่าง เช่น “ฉันชอบคนอ่านข่าวรายกานนี้มาก น่าจะออกอากาศเวลาหัวค่ำกว่าเดิมหน่อย” ในขั้นตอนนี้จะแก้ไขคำผิด คือคำว่า “รายกาน” แก้ไขเป็น “รายการ” และ “ออกอากาศ” แก้ไขเป็น “ออกอากาศ” และคำที่มี

ความหมายเหมือนกันแต่ใช้คำต่างกันจะแก้ไขให้เป็นคำเดียว เช่น คำว่า “คนอ่านข่าว” กับ “ผู้ประกาศข่าว” มีความหมายเหมือนกันจึงแก้ไขเป็น “ผู้ประกาศข่าว” เป็นต้น

(3) การแทนข้อความเป็นกระบวนการแทนข้อความให้อยู่ในรูปแบบที่มีโครงสร้างแน่นอน มีเทคนิคที่นิยมใช้คือ เวกเตอร์สเปซโมเดล (Vector Space Model: VSM) เป็นวิธีการหนึ่งในการแทนข้อความให้มีโครงสร้างแบบพีเจอร์เวกเตอร์ (feature vector) เพื่อให้คอมพิวเตอร์สามารถนำไปประมวลผลต่อไป โดยมีวิธีการแทนข้อความดังนี้ (Atorn Nuntiyagul, 2006)

- การแทนข้อความด้วยถ่วงคำ (Bag-of-word) เป็นวิธีการที่ง่ายและสะดวกที่สุด โดยแทนข้อความด้วยคำที่รวบรวมไว้แล้วในคลังคำ หากมีคำในถ่วงคำเกิดขึ้นในข้อความที่นำมาวิเคราะห์ จะแทนค่าด้วย 1 แต่หากในข้อความไม่มีคำที่กำหนดไว้ในถ่วงคำเกิดขึ้น จะแทนค่าด้วย 0

- การแทนข้อความด้วยค่าความถี่การเกิดขึ้นของคำ (Term Frequency: TF)

- การแทนค่าคำด้วยค่าความถี่คำ-ค่าส่วนกลับความถี่เอกสารที่เกิดคำ (Term Frequency – Inverse Document Frequency: TF-IDF) เนื่องจากวิธีการแทนข้อความด้วยค่าความถี่การเกิดขึ้นของคำเพียงอย่างเดียว ไม่สามารถจำแนกข้อความได้ดีพอ เพราะค่าความถี่ของคำสูง หมายถึง คำนั้นมีโอกาสเกิดขึ้นในหลายข้อความพร้อมกัน ซึ่งไม่มีประโยชน์ต่อการวิเคราะห์เพื่อจำแนกประเภทข้อความ (เช่น จำแนกความคิดเห็นเป็นข้อความเชิงบวกหรือลบ) ดังนั้น Salton and Buckley (1988) จึงนำเสนอวิธีการแทนค่าคำด้วยค่าความถี่คำ-ค่าส่วนกลับความถี่เอกสารที่เกิดคำ ซึ่งคำนึงถึงความถี่ของคำจากทั้งเอกสารด้วย ถือเป็นวิธีที่ได้รับความนิยมมากที่สุด เนื่องจากมีวิธีคำนวณไม่ซับซ้อนและมีประสิทธิภาพสูง

การสร้างแบบจำลองการวิเคราะห์ข้อมูล

แบบจำลองการวิเคราะห์ข้อมูลถูกสร้างจากการเรียนรู้รูปแบบข้อมูลที่มีอยู่และรู้ผลเฉลยแล้ว เรียกว่าชุดข้อมูลเรียนรู้ (training set) เพื่อนำไปวิเคราะห์ อธิบาย หรือทำนายรูปแบบของข้อมูลที่ยังไม่เกิดขึ้นหรือยังไม่รู้ผลเฉลย เรียกว่าชุดข้อมูลทดสอบ (test set) ซึ่งการสร้างแบบจำลองการวิเคราะห์ข้อมูลแบ่งเป็น 2 เทคนิคหลักคือ (1) การ

เรียนรู้แบบมีผู้สอน (supervised learning) มีวิธีที่ใช้ เช่น การจำแนกข้อมูล (classification) สำหรับกระบวนการวิเคราะห์เหมืองข้อความเรียกกระบวนการดังกล่าวว่า การจำแนกข้อความ (text classification หรือ text categorization) ยังมีปริมาณชุดข้อมูลเรียนรู้มาก จะยิ่งทำให้แบบจำลองการจำแนกข้อความที่มีความถูกต้องสูง แต่ระยะเวลาที่เครื่องใช้ในการสร้างแบบจำลองก็มากตามไปด้วยอัลกอริทึมสำหรับการจำแนกข้อความ เช่น ต้นไม้ตัดสินใจ นาอ์ฟเบย์ และซัพพอร์ทเวกเตอร์แมชชีน เป็นต้น และ (2) การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) มีวิธีที่ใช้ เช่น การจัดกลุ่มข้อมูล (clustering) ในงานวิจัยของ Thorsten (1998) และกานดา แก้วพัฒนากุล (2556) ได้ทดสอบเพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการจำแนกข้อความด้วยอัลกอริทึมต้นไม้ตัดสินใจ นาอ์ฟเบย์ และซัพพอร์ทเวกเตอร์แมชชีน ซึ่งพบว่าซัพพอร์ทเวกเตอร์แมชชีนมีประสิทธิภาพการจำแนกข้อความดีกว่าวิธีอื่น

การประยุกต์ใช้การวิเคราะห์เหมืองความคิดเห็น

ข้อความบนเครือข่ายสังคมออนไลน์แบ่งเป็น 3 ประเภท ได้แก่ ข้อเท็จจริง ความคิดเห็น และข้อเสนอแนะ โดย (1) ข้อเท็จจริงหมายถึงข้อความหรือเหตุการณ์ที่เป็นมาหรือที่เป็นอยู่ตามจริง (2) ความคิดเห็นหมายถึงความเห็นหรือข้อวินิจฉัยหรือความเชื่อที่แสดงออกตามความเห็น รู้ คิด และ (3) ข้อเสนอแนะหมายถึงข้อคิดเห็นเชิงแนะนำที่เสนอเพื่อพิจารณา ชี้แจงให้ทำหรือปฏิบัติตาม (กานดา แก้วพัฒนากุล, 2555) ดังตัวอย่าง “เมื่อคืนได้ดูรายการพื้นที่ชีวิต ไม่ชอบพิธีกรเลย อยากให้ปรับปรุงรายการใช้ภาษาหน่อย” จากประโยคตัวอย่างสามารถจำแนกประเภทของบทวิจารณ์ได้ ดังนี้

ข้อเท็จจริง : “เมื่อคืนได้ดูรายการพื้นที่ชีวิต”

ความคิดเห็น : “ไม่ชอบพิธีกรเลย”

ข้อเสนอแนะ : “อยากให้ปรับปรุงรายการใช้ภาษาหน่อย”

ธุรกิจในปัจจุบันให้ความสำคัญกับความคิดเห็นของผู้บริโภคบนเครือข่ายสังคมออนไลน์มากขึ้น แต่ปัญหาคือบางความคิดเห็นมีข้อความแสดงความคิดเห็นค่อนข้างยาว และมีเพียงบางประโยคเท่านั้นที่ผู้อ่านให้ความสนใจ และข้อความความคิดเห็นของแต่ละบุคคลมีความหลากหลายใน

แต่ละมุมมองที่แตกต่างกันไป จึงเสียเวลามากหากต้องอ่านทุกความคิดเห็นว่าจะสรุปความคิดเห็นในมุมมองเฉพาะเรื่องที่น่าสนใจได้ หลายธุรกิจจึงได้นำวิธีการวิเคราะห์เหมืองความคิดเห็นไปประยุกต์ใช้ เพื่อสรุปความพึงพอใจของผู้บริโภคหรือแม้กระทั่งผู้บริโภคเองที่ค้นหาความพึงพอใจของบุคคลอื่นที่แสดงความคิดเห็นไว้บนเครือข่ายสังคมออนไลน์

การวิเคราะห์เหมืองความคิดเห็นเป็นกระบวนการอัตโนมัติเพื่อตรวจสอบทัศนคติของผู้เขียนในหัวข้อเรื่องใดเรื่องหนึ่ง โดยแบ่งทัศนคติหรือข้อความความคิดเห็น (polarity) เป็นข้อความความคิดเห็นเชิงบวก ลบ และเป็นกลาง ซึ่งแบ่งระดับการวิเคราะห์ออกเป็น 3 ระดับคือ (1) ระดับเอกสาร (document level) เป็นการวิเคราะห์ข้อความแสดงความคิดเห็นจากทั้งเอกสาร, D, มาสรุปแยกข้อความความคิดเห็นเป็น ขั้วบวก ลบ หรือเป็นกลาง (2) ระดับประโยค (sentence level) ใน 1 เอกสารประกอบด้วยหลายประโยค, S, โดยกำหนดให้ $D = \{S_1, S_2, S_3, \dots, S_n\}$ เมื่อ n คือ จำนวนประโยคทั้งหมดในเอกสาร การวิเคราะห์ในระดับประโยคจึงเป็นการวิเคราะห์ทีละประโยคแล้วแยกข้อความความคิดเห็นเป็น ขั้วบวก ลบ หรือเป็นกลาง และ (3) ระดับคุณลักษณะ (feature level) เป็นการวิเคราะห์ข้อความความคิดเห็นโดยแยกตามคุณลักษณะ (feature หรือ aspect) ของหัวข้อที่ถูกนำมาแสดงความคิดเห็น (object หรือ entity) แล้วจึงนำมาแยกข้อความความคิดเห็นเป็น ขั้วบวก ลบ หรือเป็นกลาง

Turney (2002) ได้นำเสนอการวิเคราะห์ความคิดเห็นในระดับเอกสาร โดยใช้คำวิเศษณ์หรือวลีวิเศษณ์ในการบ่งชี้ข้อความเห็น ได้แก่ คำคุณศัพท์ (adjective) และคำกริยาวิเศษณ์ (adverb) และใช้คำนามหรือนามวลีในการบ่งชี้หัวข้อที่สนใจ (objects) และสรุปข้อความความคิดเห็นของเอกสารเป็นเชิงบวก (thumbs up) หรือเชิงลบ (thumbs down) ด้วยวิธีการนับจำนวนค่าที่แสดงข้อความความคิดเห็น ขั้วบวกและขั้วลบ หากข้อความความคิดเห็นมีจำนวนมากกว่าจะสรุปเป็นข้อความความคิดเห็นของเอกสารนั้น แต่เนื่องจากในหนึ่งเอกสารประกอบไปด้วยหลายประโยคการวิเคราะห์ในระดับเอกสารจึงไม่มีความละเอียดมากพอ Wiebe & Hwa (2005), Kim; Liu และคณะ (2004) จึงนำเสนอวิธีการวิเคราะห์ความคิดเห็นในระดับประโยค และจัดกลุ่มประโยคเป็นความคิดเห็นเป็นเชิงบวก หรือเชิงลบ

ในปี 2008 Liu และคณะนำเสนอว่าวัตถุหรือหัวข้อที่ผู้แสดงความคิดเห็นกล่าวถึงประกอบด้วยหลาย

คุณลักษณะ เช่น กล้องถ่ายรูป มีคุณลักษณะที่หลากหลาย ไม่ว่าจะเป็น จอภาพ น้ำหนัก ฟังก์ชันการใช้งาน เป็นต้น ดังนั้นการวิเคราะห์ในระดับเอกสารและประโยคเป็นการสรุปอย่างกว้างๆ ไม่สามารถทราบผู้แสดงความคิดเห็น กล่าวถึงคุณลักษณะใดของหัวข้อที่สนใจ Liu และคณะ (2008) จึงนำเสนอการวิเคราะห์เหมือนความคิดเห็นในระดับคุณลักษณะ (feature level) ซึ่งทำให้ผู้อ่านทราบว่าผู้แสดงความคิดเห็นมีทัศนคติเชิงบวก ลบ หรือเป็นกลางต่อคุณลักษณะของหัวข้อที่สนใจ

Bing Liu (2011) ได้นำเสนอองค์ประกอบของข้อความแสดงความคิดเห็น อันประกอบด้วย 5 ส่วน ดังนี้ (1) หัวข้อที่สนใจ (object หรือ entity) เช่น สินค้า บริการ การเมือง หรือเหตุการณ์ที่สนใจ เป็นต้น (2) คุณลักษณะที่มีต่อหัวข้อที่สนใจ (feature หรือ aspect) (3) ผู้แสดงความคิดเห็น (opinion holder) (4) เวลาที่แสดงความคิดเห็น และ (5) ขั้วความคิดเห็นที่มีต่อหัวข้อหรือคุณลักษณะที่สนใจ ซึ่งสามารถอธิบายองค์ประกอบของข้อความแสดงความคิดเห็นได้ดังสมการ

	$(e_j, a_{jk}, SO_{ijk}, h_i, t_i)$	
โดยที่	e_j	คือ หัวข้อที่ถูกนำมาแสดงความคิดเห็นลำดับที่ j
	a_{jk}	คือ คุณลักษณะของหัวข้อ e_j ลำดับที่ k
	h_i	คือ ชื่อผู้แสดงความคิดเห็น ต่อหัวข้อ e_j และคุณลักษณะ a_{jk} ลำดับที่ i
	t_i	คือ เวลาที่แสดงความคิดเห็น ต่อหัวข้อ e_j และ คุณลักษณะ a_{jk} ลำดับที่ i
	SO_{ijk}	คือ ขั้วความคิดเห็นเชิงบวก (+), ลบ (-) หรือเป็นกลาง (neutral) ต่อหัวข้อ e_j และคุณลักษณะ a_{jk} ของผู้แสดงความคิดเห็น h_i ณ เวลา t_i

ดังตัวอย่าง *Abc123 on 5-1-2008 "I bought an iPhone a few days ago. It is such a nice phone. The touch screen is really cool. The voice quality is clear too. It is much better than my old Blackberry, which was a terrible phone and so difficult to type with its tiny keys. However, my mother was mad with me as I did not tell her before I bought the phone. She also thought the phone was too expensive,..."*

สามารถอธิบายองค์ประกอบของข้อความแสดงความคิดเห็นได้ดังนี้

(iPhone, GENERAL, +, Abc123, 5-1-2008)

(iPhone, touch_screen, +, Abc123, 5-1-2008)

...

จากนั้นจึงนำมาสรุปความคิดเห็นที่มีต่อหัวข้อที่สนใจ ได้ดังภาพที่ 1 หัวข้อที่สนใจคือ Digital camera 1 โดยแยกตามคุณลักษณะของกล้องถ่ายรูป เช่น คุณภาพของภาพถ่าย (picture quality) มีจำนวนผู้แสดงความคิดเห็นเชิงบวก 256 ความคิดเห็น และจำนวนขั้วความคิดเห็นเชิงลบ 6 ความคิดเห็น และคุณลักษณะด้านขนาดของกล้อง (size) มีจำนวนผู้แสดงความคิดเห็นเชิงบวก 134 ความคิดเห็น และจำนวนขั้วความคิดเห็นเชิงลบ 10

ความคิดเห็น นอกจากนั้นแล้วยังสามารถอธิบายด้วยรูปภาพเพื่อให้เข้าใจได้ดียิ่งขึ้น ดังภาพที่ 2 ภาพบนแสดงหัวข้อที่สนใจคือ Cell Phone 1 แกน x แทนคุณลักษณะด้านต่างๆ ดังนี้ voice, screen, battery, size และ weight และแกน y แทนจำนวนของขั้วความคิดเห็นเชิงบวกและลบ ส่วนภาพที่ 2 ด้านล่างเป็นการเปรียบเทียบ 2 หัวข้อที่สนใจ คือ Cell Phone 1 และ Cell Phone 2 แบ่งตามคุณลักษณะด้านต่างๆ

Digital_camera 1:

Feature: picture quality:

Positive: 253

<individual review sentence>

Negative: 6

<individual review sentence>

Feature: size:

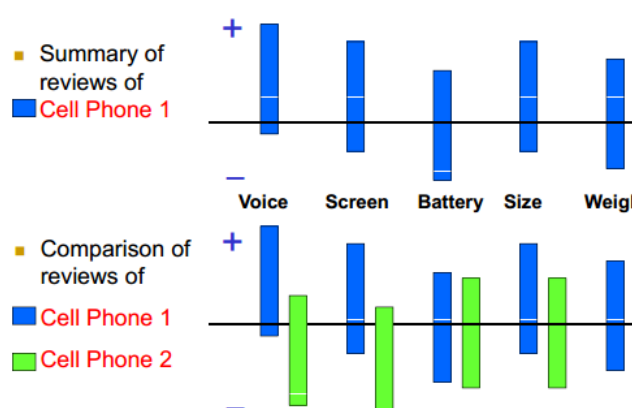
Positive: 134

<individual review sentence>

Negative: 10

<individual review sentence>

...

ภาพที่ 1 ตัวอย่างผลลัพธ์ที่ได้จากการจำแนกความคิดเห็น**ภาพที่ 2** ตัวอย่างผลลัพธ์แสดงความเห็นเชิงบวกและลบ และการเปรียบเทียบความคิดเห็นจาก 2 หัวข้อ

สำหรับการวิเคราะห์ความคิดเห็นภาษาไทย จะมีขั้นตอนวิธีที่ย่งยากกว่าภาษาอังกฤษเนื่องจากภาษาไทยมีรูปแบบการเขียนที่ไม่ตายตัว แนวทางการประมวลผลภาษาธรรมชาติที่นิยมใช้คือวิเคราะห์จากโครงสร้างทางภาษาศาสตร์ ร่วมกับการแทนความรู้ด้วยคลังคำ โดย ชูชาติ หฤไชยะศักดิ์; อลิสา คงทน (2553) และวรรณฎา วรณศรี (2553) ได้นำเสนอวิธีการวิเคราะห์เหมือนความคิดเห็นสำหรับภาษาไทยของผู้ใช้บริการโรงแรม โดยมีขั้นตอนดังนี้

1. การทำความเข้าใจกับธุรกิจและปัญหาของงาน คือวิเคราะห์ถึงความต้องการข้อมูลสารสนเทศที่จะนำไปใช้ประโยชน์ เช่น การแบ่งคุณลักษณะของหัวข้อที่สนใจ

2. การรวบรวมข้อความแสดงความคิดเห็นของผู้ใช้บริการโรงแรมที่อยู่บนเว็บไซต์ agoda.com และแบ่งข้อมูลออกเป็น 2 ชุด คือ (1) ชุดข้อมูลเรียนรู้ และ (2) ชุดข้อมูลทดสอบ โดยให้ผู้เชี่ยวชาญจำแนกชุดข้อมูลเรียนรู้ออกเป็นข้อความคิดเห็นเชิงบวกและลบ แบ่งตามหัวข้อและคุณลักษณะ

3. การเตรียมข้อมูล แบ่งเป็น 3 กระบวนการ คือ (1) เลือกคุณลักษณะสำคัญ ผู้เชี่ยวชาญสร้างคลังคำสำหรับวิเคราะห์ความคิดเห็นที่เกี่ยวข้องกับโรงแรมและเก็บไว้ในคลังคำเพื่อเป็นแหล่งความรู้ทางภาษา (language resources) (2) กลั่นกรองข้อความ และ (3) แทนข้อความในขั้นตอนนี้จะทำการแบ่งเอกสารออกเป็นประโยค และตัดประโยคออกเป็นคำ จากนั้นจึงกำกับคำด้วยหน้าที่ของคำตามคลังคำที่เก็บไว้ในแหล่งความรู้ทางภาษา ดังภาพที่ 3 เป็นการกำกับคำคุณลักษณะ และคำแสดงข้อความคิดเห็น และขั้นตอนสุดท้ายคือการแทนคำด้วยค่า TF-IDF

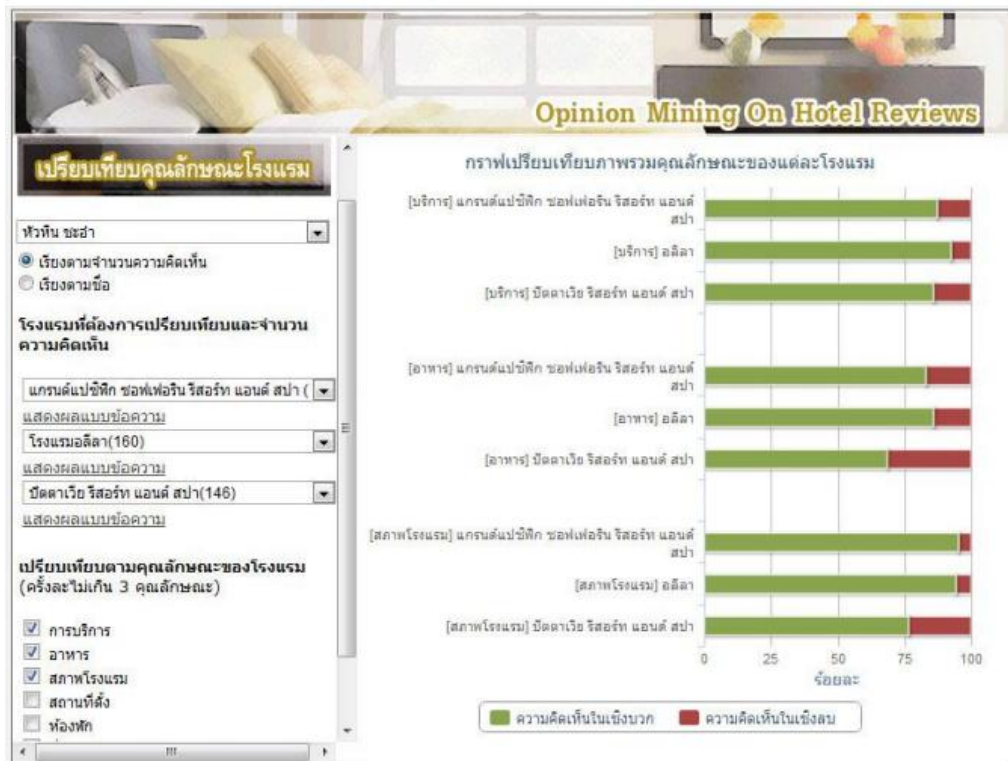
4. การสร้างแบบจำลองการวิเคราะห์ข้อมูลจากชุดข้อมูลเรียนรู้เพื่อจำแนกข้อความคิดเห็นเป็นเชิงบวกหรือลบ สามารถใช้โปรแกรม Weka หรือ RapidMiner ในการสร้างแบบจำลองได้

5. การประเมินประสิทธิภาพของแบบจำลองการวิเคราะห์ข้อมูล โดยนำชุดข้อมูลทดสอบมาจำแนกข้อความคิดเห็นด้วยแบบจำลองการวิเคราะห์ข้อมูล และให้ผู้เชี่ยวชาญตรวจสอบความถูกต้องอีกครั้ง

6. การนำแบบจำลองการวิเคราะห์ข้อมูลไปใช้งานจริง โดยผลลัพธ์ที่ได้จากการวิเคราะห์แสดงดังภาพที่ 4 โดยแสดงเป็นกราฟเปรียบเทียบภาพรวมคุณลักษณะของแต่ละโรงแรม ดังภาพเป็นการเปรียบเทียบ 3 โรงแรม แบ่งตามคุณลักษณะ 3 ด้าน คือ บริการ อาหาร และสภาพโรงแรม โดยที่กราฟสี่เหลี่ยมแสดงข้อความคิดเห็นเชิงบวกและสี่เหลี่ยมแสดงข้อความคิดเห็นเชิงลบ



ภาพที่ 3 ตัวอย่างการกำกับค่าคุณลักษณะและค่าแสดงข้อความความคิดเห็น



ภาพที่ 4 ตัวอย่างการผลลัพธ์ที่ได้จากการวิเคราะห์เหมืองความคิดเห็น

ที่มา: วรรณญา วรรณศรี (2553)

จากเนื้อหาที่ได้กล่าวมาทั้งหมด สามารถสังเกตได้ว่า การวิเคราะห์เหมืองความคิดเห็นเป็นการวิเคราะห์ความคิดเห็นทางตรงเท่านั้น กล่าวคือมีค่าแสดงข้อความความคิดเห็นที่ชัดเจน แต่สำหรับข้อเสนอแนะ ตัวอย่างเช่น “รายการนี้มีแต่พิธีการเก่ง ๆ แต่คิดว่าควรปรับปรุงเรื่องการให้บริการให้ถูกต้องอีกนิด” ประโยคที่มีค่าแสดงข้อความความคิดเห็นที่ชัดเจนคือ “พิธีการเก่ง” แต่ข้อเสนอแนะเรื่อง “ควรปรับปรุงการให้บริการ” กลับไม่ถูกนำมาวิเคราะห์ ทั้ง ๆ ที่ข้อเสนอแนะของผู้บริโภคเป็นสิ่งจำเป็นยิ่งที่ธุรกิจควรพิจารณาเพื่อนำไปปรับปรุงคุณภาพ ซึ่งในงานวิจัยของ Amar Viswanathan, Prasanna Venkatesh, Bintu Vasudevan, Rajesh Balakrishnan and Lokendra Shastri (2011) กล่าวว่า 20-30% ของความคิดเห็นบน

เครือข่ายสังคมออนไลน์จะมีข้อเสนอแนะซ่อนอยู่ ดังนั้น ผู้เขียนจึงมีความเห็นว่า น่าจะเป็นประโยชน์หากสามารถสกัดคุณค่าจากข้อเสนอแนะเหล่านี้ได้

สรุป

ปัจจุบันธุรกิจให้ความสำคัญกับความคิดเห็นของผู้บริโภคบนเครือข่ายสังคมออนไลน์มากขึ้น แต่เนื่องจากความคิดเห็นจำนวนมากหาศาลบนเครือข่ายสังคมออนไลน์ ภาคธุรกิจจึงนำเทคนิคการวิเคราะห์เหมืองความคิดเห็นมาประยุกต์ใช้เพื่อวิเคราะห์ความพึงพอใจของผู้บริโภคทั้งในแง่บวกและแง่ลบที่มีต่อคุณลักษณะของหัวข้อที่สนใจในแต่ละด้านได้อย่างรวดเร็ว อาทิ หน้าจอของมือถือ หน้าหลักของ

กล้องถ่ายรูป การบริการของโรงแรม ระยะทางของสถานที่ รวมถึงพิธีกรในรายการโทรทัศน์ เป็นต้น ซึ่งช่วยให้ธุรกิจทราบความรู้สึกของผู้บริโภคเพื่อนำไปปรับกลยุทธ์ รวมถึงการปรับปรุงคุณภาพสินค้าและบริการของธุรกิจต่อไป

การวิเคราะห์เหมืองความคิดเห็นบนเครือข่ายสังคมออนไลน์แบ่งออกเป็น 2 กระบวนการหลัก คือ (1) ขั้นตอนการแทนข้อความให้อยู่ในรูปแบบที่คอมพิวเตอร์สามารถนำไปประมวลผลได้ โดยนำความรู้ด้านการประมวลผลภาษาธรรมชาติมาประยุกต์ใช้ และ (2) ขั้นตอนการจำแนกความคิดเห็นออกเป็นข้อความเชิงบวกหรือลบ โดยนำความรู้ด้านการทำเหมืองข้อความมาประยุกต์ใช้ ซึ่งกระบวนการนี้เป็นการสร้างแบบจำลองเพื่อจำแนกข้อความความคิดเห็นของชุดข้อมูลทดสอบ

กระบวนการวิเคราะห์เหมืองความคิดเห็นนอกจากจะวิเคราะห์ความพึงพอใจของผู้บริโภคเชิงบวก

หรือลบแล้ว สามารถปรับปรุงไปประยุกต์ใช้กับแนวทางการวิเคราะห์ข้อความบนเครือข่ายสังคมออนไลน์อื่นๆ ได้อีก ดังเช่น การรับฟังข้อเสนอแนะของผู้บริโภคว่าต้องการให้ธุรกิจปรับปรุงเรื่องใดและแก้ไขไปในแนวทางใดหรือข้อเสนอแนะต่อแนวทางการแก้ไขปัญหาในวิกฤตการณ์ต่างๆ ตัวอย่างคือวิกฤตการณ์น้ำท่วมใหญ่ ผู้ใช้เครือข่ายสังคมออนไลน์ต่างต้องการทราบปัญหารวมถึงเสนอแนวทางการแก้ไข แต่ปัญหาที่เกิดขึ้นคือข้อความจำนวนมากสับสนมีความหลากหลาย แบ่งย่อยได้หลายหมวดหมู่ ทั้งวิธีการแก้ไขและสถานที่เกิดเหตุ จึงควรมีวิธีการจัดการกับข้อเสนอแนะเพื่อให้ได้ข้อสรุปที่สามารถนำไปใช้งานได้อย่างเป็นระบบและทันต่อสถานการณ์ ซึ่งแนวทางการวิเคราะห์ข้อเสนอแนะเพื่อสกัดหาแนวทางแก้ไขปัญหาถูกเรียกว่าการวิเคราะห์เหมืองข้อเสนอแนะ (Suggestion Mining)

บรรณานุกรม

- กานดา แฉ้วนกุล และปราโมทย์ ลีโอนาม. (2555). *การวิเคราะห์เหมืองข้อเสนอแนะจากบทวิจารณ์รายการโทรทัศน์*. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ วิทยาศาสตรบัณฑิต) สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- กนกวรรณ เขียนวรรณ. (2555). *การประมวลผลภาษาธรรมชาติ*. ค้นวันที่ 3 ธันวาคม 2555 จาก www.mbs.mut.ac.th/paper/pdf/29.pdf
- วรัญญา วรรณศรี. (2553). *ระบบวิเคราะห์ข้อความแสดงความเห็นสำหรับโรงแรม*. ค้นวันที่ 13 มีนาคม 2554 จาก <http://thailang.nectec.or.th/halloffame/>
- สมนึก สีนุพาน. (2546). *การวิเคราะห์กระจายคำในประโยคภาษาไทย โดยการโปรแกรมเชิงเจเนติก*. (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ วิทยาศาสตรบัณฑิต) สถาบันบัณฑิตพัฒนบริหารศาสตร์.
- Ellen Riloff, Janyce Wiebe and William Phillips. (2005). *Exploiting Subjectivity Classification to Improve Information Extraction*, **AAAI-2005**, pp.1106-1111.
- Gerard Salton and Christopher Buckley. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*. Vol. 24, No. 5, pp. 513-523.
- Hu Mingqing and Liu Bing. (2004). Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Retrieved March 13, 2011 from <http://www.informatik.uni-trier.de/~ley/db/conf/kdd/kdd2004.html>
- IBM Software Business Analytics. (2012). *CRISP DM*. Retrieved February 10, 2012 from <http://www.ibm.com>
- Joachims, Thorsten. (1998). Text Categorization with Support Vector Machines: Learning with Many Relevant Features. In *Proceedings of the 10th European Conference on Machine Learning (ECML'98)*. Claire Nedellec and Céline Rouveirol, eds. London: Springer-Verlag. pp. 137-142.
- Kongthon Alisa, Angkawattanawit Niran, Sangkeettrakarn Chatchawal, Palingoon Pornpimon and Haruechaiyasak Choochart. (2010). Using an Opinion Mining Approach to Exploit Web Content in Order to Improve

- Customer Relationship Management. *Proceedings of Technology Management for Global Economic Growth*. Retrieved December 10, 2011 from IEEE Xplore Digital Library.
- Liu Bing. (2011). Tutorial Sentiment Analysis and Opinion Mining. *UIC College of Engineering*. Retrieved March 1, 2012 from <http://www.cs.uic.edu/~liub/>
- Nuntiyagul Atorn. (2006). *Text Categorization & Retrieval for Thai Item Bank using Patterned Keyword in phrase (PKIP)*. Doctoral dissertation, Mahidol University.
- Pang Bo and Lee Lillian. (2008). Opinion Mining and Sentiment Analysis. *Journal of the ACM*. 2 (January): 1-135. Retrieved December 10, 2011 from ACM Digital Library.
- Phawattanakul Kanda and Luenam Pramote. (2013). Suggestion Mining and Knowledge Construction from Thai Television Program Reviews. In *Proceedings of the International Multi Conference of Engineers and Computer Scientists*. Hongkong. pp.307-312.
- Sukhum Khampol, Nitsuwat Supot and Haruechaiyasak Choochart. (2011). Opinion Detection in Thai Political News Columns Based on Subjectivity Analysis. *Information Technology Journal*. 14 (July): 27-31. Retrieved March 1, 2012 from <http://suanpalm3.kmutnb.ac.th/journal>
- Turney, Peter D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. Assoc. for Computational Linguistics. Pennsylvania: Association for Computational Linguistics Stroudsburg. pp.417-424.
- Vishwanath, J. and Aishwarya, S. (2011). User Suggestions Extraction from customer Reviews A Sentiment Analysis approach. *International Journal on Computer Science and Engineering (IJCSE)*. 3 (March): 1203-1206.
- Viswanathan Amar, Venkatesh Prasanna, Vasudevan Bintu, Balakrishnan Rajesh and Shastri Lokendra. (2011). Suggestion Mining from Customer Reviews. *AMCIS 2011 PROCEEDINGS*. Retrieved February 10, 2012 from AIS Electronic Library (AISeL)