

ความสัมพันธ์ระหว่างทวิตเตอร์และตราสารการเงิน

นนท์ คณิงสุขเกษม และ อีรพงศ์ ลีลานุภาพ

คณะเทคโนโลยีสารสนเทศ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง กรุงเทพฯ

Emails: s6606011@kmitl.ac.th, teerapong@it.kmitl.ac.th

บทคัดย่อ

ทวิตเตอร์เป็นที่นิยมใช้กันอย่างแพร่หลาย ข้อมูลสารสนเทศจากทวิตเตอร์มีจำนวนมหาศาลและยังคงเพิ่มขึ้นในแต่ละวัน ดังนั้นทวิตเตอร์จึงเป็นแหล่งข้อมูลที่สำคัญซึ่งสามารถนำมาใช้ในการศึกษาในแง่มุมต่างๆ โดยงานวิจัยนี้มุ่งเน้นศึกษาความสัมพันธ์ระหว่างข้อมูลสารสนเทศบนทวิตเตอร์กับตราสารการเงิน ประโยชน์ที่เป็นไปได้จากการหาความสัมพันธ์นี้สามารถนำมาเป็นตัวช่วยในการทำนายการเคลื่อนไหวของตราสารการเงิน การศึกษาในเรื่องนี้จำเป็นต้องอาศัยทฤษฎีทั้งจากทางด้านการเงิน และเทคโนโลยีสารสนเทศ ผู้เขียนจึงมีความประสงค์ที่จะอภิปรายข้อมูลเชิงทฤษฎีจากทั้งสองด้าน และงานวิจัยที่เกี่ยวข้อง โดยบทความนี้จะกล่าวถึงทฤษฎีทางการเงินที่เกี่ยวข้องและทฤษฎีที่สัมพันธ์กับพฤติกรรมทางสังคมออนไลน์ในทวิตเตอร์ รวมทั้งงานวิจัยเกี่ยวกับการหาค่าความสัมพันธ์ระหว่างทวิตเตอร์และตราสารการเงิน

คำสำคัญ – ทวิตเตอร์; ไมโครบล็อก; ตราสารการเงิน; หุ้น

ABSTRACT

Due to the popularity of Twitter, the massive volume of Twitter data has become an interesting source for studying in various aspects. This paper focuses on studying the correlations between Twitter data and financial instruments. We expect the efficacy of the correlations in predicting movements in financial markets. This study requires knowledge of finance and information technology. Accordingly, we intend to conduct the discussion of the related financial theories, online social behavior in Twitter and some of the related works about correlation between Twitter data and movement of the financial markets or any financial instrument.

Keywords – Twitter; Microblog; Financial instrument; Stock

1. บทนำ

ทวิตเตอร์ได้รับการกล่าวว่าเป็นระบบไมโครบล็อกที่โด่งดังที่สุดในขณะนี้ โดยมีผู้ใช้งานเฉลี่ย 215 ล้านคนต่อเดือน และมากกว่า 100 ล้านคนต่อวัน ซึ่งทำให้มีการส่งข้อความ

สั้นหรือที่เรียกกันว่า “ทวิต” เป็นจำนวนมากถึง 500 ล้านทวิตต่อวันโดยประมาณ [1]

ทวิตเตอร์ถูกนำมาใช้งานในหลายจุดประสงค์อย่างกว้างขวาง เช่น การใช้งานส่วนตัว การใช้งานด้านธุรกิจ และการใช้งานลับแบบกลุ่มเฉพาะ หรือแม้กระทั่งการใช้

ในการโฆษณาสินค้า นอกจากนี้ผู้ใช้ส่วนใหญ่ยังใช้ทวิตเตอร์เป็นสื่อกลางในการรับรู้ข่าวสารต่างๆ แทนการรับรู้ข่าวสารในรูปแบบเดิมๆ ด้วย

โดย Kwak *et al.* [2] ศึกษาพบว่ามากกว่า 85% ของทวิตเตอร์เป็นข้อความที่กล่าวถึงข่าวพาดหัว หรือข้อความที่เกี่ยวข้องกับข่าวสารต่างๆ รวมทั้งมีรายงานจาก Petrovic *et al.* [3] มาสนับสนุนด้วยว่าข่าวสารจากทางทวิตเตอร์นั้นมีความฉับไวกว่าข่าวสารที่ได้จากแหล่งข่าวแบบดั้งเดิม เช่น หนังสือพิมพ์ออนไลน์ สื่อวิทยุ และโทรทัศน์

เนื่องจากลักษณะเด่นของทวิตเตอร์ในแง่ของความแพร่หลายและความเร็วในการรับส่งข้อความ ทำให้ทวิตเตอร์กลายเป็นแหล่งข้อมูลของข่าวสารที่สามารถรายงานข้อมูลได้อย่างทันเหตุการณ์ โดยมีงานวิจัยหลายงานพยายามนำทวิตเตอร์มาใช้เพื่อตรวจจับเรื่องราว ข่าว หรือเหตุการณ์ต่างๆ ที่เกิดขึ้น และนำข้อมูลเหล่านั้นมาใช้ประโยชน์โดยทันที [4,5]

นอกจากนี้ส่วนหนึ่งของข้อมูลข่าวสารที่ได้จากทวิตเตอร์ คือ ข้อมูลข่าวสารที่เกี่ยวข้องกับบริษัทที่อยู่ในตลาดหลักทรัพย์ เช่น คำแนะนำซื้อหรือขายหลักทรัพย์จากนักลงทุนที่มีชื่อเสียง และข่าวการควบรวมกิจการของบริษัทในตลาดหลักทรัพย์ เป็นต้น โดยหากข้อความในทวิตมีการกล่าวถึงบริษัทเหล่านั้น ผู้ส่งข้อความมักจะรวมเอาชื่อย่อในตลาดหลักทรัพย์ของบริษัทนั้นไว้ในทวิตด้วย และนิยมใส่เครื่องหมายดอลลาร์ (\$) ไว้ข้างหน้าเพื่อระบุว่าตัวย่อนั้นคือชื่อย่อในตลาดหลักทรัพย์ ซึ่งต่อมาทางทวิตเตอร์ก็ได้แปลงการใส่เครื่องหมายดอลลาร์เป็นเครื่องหมายในการติดตามค้นหาทวิตที่มีอักษรย่อของบริษัทในตลาดหลักทรัพย์นั้นๆ [6]

ด้วยสาเหตุที่กล่าวมาข้างต้น ทวิตเตอร์จึงเป็นแหล่งข้อมูลที่น่าสนใจต่อการนำข้อมูลสารสนเทศที่เกี่ยวกับตราสารการเงินทั้งหมดมาศึกษาและวิจัยร่วมกับสมมติฐานตลาดที่มีประสิทธิภาพ (Efficient Market

Hypothesis) เนื่องจาก Fama [7] กล่าวว่า สมมติฐานตลาดที่มีประสิทธิภาพเป็นเพียงข้อสมมติพื้นฐาน “ราคาของตราสารการเงินทั้งหลายสะท้อนถึงข้อมูลสารสนเทศทั้งหมดที่เกี่ยวข้องกับตราสารการเงินนั้นๆ อย่างสมบูรณ์แล้ว” หรืออีกนัยหนึ่งคือ นักลงทุนไม่สามารถชนะตลาดได้โดยการเรียนรู้ทุกอย่างเกี่ยวกับตราสารการเงินนั้นๆ เพราะข้อมูลสารสนเทศทั้งหมดที่เกี่ยวกับตราสารการเงินนั้นๆ ได้ถูกสะท้อนออกมาในรูปของราคาของตราสารการเงินนั้นๆ เรียบร้อยแล้ว

มีงานวิจัยจำนวนหนึ่งศึกษาหาสหสัมพันธ์ (correlation) ระหว่างข้อมูลสารสนเทศทางการเงินที่ได้จากทวิตเตอร์และการเคลื่อนไหวของราคาของตราสารการเงินต่างๆ [8,9] ซึ่ง Ruiz *et al.* [8] ได้รายงานผลว่าข้อมูลสารสนเทศที่ได้จากทวิตเตอร์นั้นมีค่าสหสัมพันธ์ร่วมกับปริมาณการซื้อขายของหุ้นเป็นอย่างมาก และมีความสัมพันธ์กับการเคลื่อนไหวของราคาหุ้นบ้าง Mao *et al.* [9] แสดงให้เห็นว่าจำนวนของทวิตในแต่ละวันมีความสัมพันธ์กันกับการเคลื่อนไหวของราคาหุ้นเช่นกัน และยังสรุปเพิ่มเติมอีกด้วยว่า ข้อมูลที่ได้จากทวิตเตอร์ดูเหมือนจะเป็นประโยชน์ต่อการทำนายการเคลื่อนไหวของราคาหุ้น ซึ่งจะขัดแย้งกับสมมติฐานตลาดที่มีประสิทธิภาพ แต่ไปสนับสนุนทฤษฎีจิตวิทยาการเงิน (Behavioral Finance) ซึ่งศึกษาเศรษฐศาสตร์การเงินในแง่พฤติกรรมของมนุษย์

ทฤษฎีจิตวิทยาการเงินจะแสดงให้เห็นว่าการตัดสินใจทางการเงินของนักกลุมนั้นอาจจะได้รับอิทธิพลมาจากการรับรู้ที่ผิดพลาดหรือจากอารมณ์และความรู้สึกได้ [10] ซึ่งในด้านของทวิตเตอร์ก็มีการศึกษาทางด้านอารมณ์จากข้อความทวิตด้วยเช่นกัน และมีการศึกษาด้วยว่าอารมณ์ที่ได้จากข้อความในทวิตนั้นสามารถทำนายการเคลื่อนไหวของตลาดหลักทรัพย์ได้ [11]

2. ทฤษฎีทางการเงินที่เกี่ยวข้อง

สมมติฐานตลาดที่มีประสิทธิภาพและทฤษฎีจิตวิทยาการเงินยังเป็นเรื่องที่มีการถกเถียงกันอยู่จนถึงปัจจุบัน การถกเถียงนี้ยังสามารถเห็นได้จากการมอบรางวัลโนเบล สาขาเศรษฐศาสตร์ (Nobel Memorial Prize in Economic Sciences) ในปี 2013 ซึ่งจะเห็นว่าแม้ Shiller ผู้นำสำคัญของการศึกษาเศรษฐศาสตร์ในแนวทฤษฎีจิตวิทยาการเงิน ซึ่งเป็นทฤษฎีที่ขัดแย้งกับสมมติฐานของ Fama ผู้ซึ่งคิดสมมติฐานตลาดที่มีประสิทธิภาพมาโดยตลอด ยังได้รับมอบรางวัลโนเบลในสาขานี้ร่วมกัน

2.1. สมมติฐานตลาดที่มีประสิทธิภาพ

(Efficient market hypothesis)

สมมติฐานตลาดที่มีประสิทธิภาพนั้นกล่าวว่า ข้อมูลสารสนเทศต่างๆแพร่กระจายอย่างรวดเร็วและแทบจะทันทีทันใดในตลาดการเงิน ซึ่งส่งผลกระทบต่อราคาของตราสารการเงินที่เกี่ยวข้องโดยทันที และด้วยเหตุนี้จึงทำให้เป็นไปได้ยากที่จะสามารถชนะตลาดได้ตลอดเวลา [12] โดยตลาดที่มีประสิทธิภาพสามารถแบ่งได้เป็น 3 แบบ [13,14] ได้แก่

1) **Weak form** เป็นสมมติฐานที่กล่าวว่า ราคาของตราสารการเงินจะเป็นตัวสะท้อนข้อมูลสารสนเทศทั้งหมดที่ได้จากข้อมูลการซื้อขายของตราสารการเงินนั้นโดยสมบูรณ์แล้ว ตัวอย่างของข้อมูลการซื้อขายของตราสารการเงิน เช่น ข้อมูลราคา และปริมาณการซื้อขายย้อนหลังของหุ้น เป็นต้น ในทางปฏิบัติเราอาจสามารถกล่าวได้ว่านักลงทุนไม่ควรจะคาดหวังว่าการวิเคราะห์ทางเทคนิคจะสามารถช่วยทำนายราคาของตราสารการเงิน หรือช่วยเลือกตราสารการเงินเพื่อทำกำไรได้

2) **Semi-strong form** เป็นสมมติฐานที่กล่าวว่า ข้อมูลสารสนเทศที่เกี่ยวข้องกับตราสารการเงินที่ถูกเผยแพร่ออกสู่สาธารณชนทั้งหมดจะถูกสะท้อนอยู่ในรูปของราคา

ของตราสารการเงินนั้นๆเรียบร้อยแล้ว ซึ่งข้อมูลสารสนเทศเหล่านั้นจะได้มาจาก ข้อมูลพื้นฐานของบริษัท ผลัดกันของบริษัท คุณภาพการบริหารงานของผู้บริหาร และข้อมูลอื่นอีกมากมายที่นักลงทุนสามารถเข้าถึงได้ ซึ่งข้อมูลเหล่านี้รวมถึงข้อมูลราคาย้อนหลังด้วย ซึ่งความหมายในทางปฏิบัติคือ นักลงทุนไม่ควรที่จะคาดหวังว่าแบบจำลองใดๆก็ตามที่สร้างจากเศรษฐศาสตร์มหภาค ข้อมูลพื้นฐานของบริษัท หรือแม้กระทั่งข้อมูลจากนักวิเคราะห์ทั้งหลาย จะสามารถช่วยเลือกตราสารการเงินเพื่อทำกำไรได้

3) **Strong form** เป็นสมมติฐานที่กล่าวว่า ราคาของตราสารการเงินจะเป็นตัวสะท้อนข้อมูลสารสนเทศทั้งหมดที่เกี่ยวข้องกับตราสารการเงินนั้นแล้ว ไม่ว่าข้อมูลนั้นจะถูกเปิดเผย หรือไม่เปิดเผยสู่สาธารณชนก็ตาม หรือในทางปฏิบัติคือ นักลงทุนไม่ควรที่จะคาดหวังว่าจะสามารถเลือกหุ้นเพื่อหากำไรจากข้อมูลวงในหรือแม้แต่ข้อมูลจากเจ้าของบริษัทโดยตรงก็ตาม

2.2. ทฤษฎีจิตวิทยาการเงิน

(Behavioral finance)

Shiller [15] อธิบายว่าทฤษฎีจิตวิทยาการเงินนั้นคือ ทฤษฎีทางการเงินจากมุมมองทางสังคมศาสตร์ ซึ่งเป็นการรวมทฤษฎีทางการเงินกับศาสตร์ทางด้านจิตวิทยาและสังคมวิทยาเข้าด้วยกัน

Bodie *et al.* [13] สรุปไว้ว่า สมมติฐานของทฤษฎีจิตวิทยาการเงิน คือ ทฤษฎีทางการเงินแบบดั้งเดิมไม่สนใจว่าคนทั่วไปจริงๆแล้วจะตัดสินใจอย่างไร และไม่สนใจว่าคนเหล่านั้นจะสร้างการเปลี่ยนแปลงได้อย่างไร Bodie *et al.* ยังกล่าวอีกว่า มีจำนวนนักเศรษฐศาสตร์จำนวนมากขึ้นเรื่อยๆที่ทำการแปลผลของงานวิจัยที่ผิดปกติไปจากปกติแต่ไปสอดคล้องกับบางส่วนของทฤษฎีความไม่สมเหตุสมผล (irrationalities) ซึ่งดูเหมือนว่าจะเป็น การแสดงลักษณะของบุคคลที่ตัดสินใจในเรื่องยากๆ

ความไม่สมเหตุสมผลนี้สามารถแบ่งได้เป็น 2 ประเภทคือ 1) นักลงทุนเหล่านั้นไม่สามารถประมวลผลข้อมูลสารสนเทศได้อย่างถูกต้องเสมอ ซึ่งทำให้ความสามารถในการสรุปการแจกแจงความน่าจะเป็นเกี่ยวกับผลตอบแทนในอนาคตผิดพลาดไปด้วย และ 2) แม้นักลงทุนจะได้รับข้อมูลการแจกแจงความน่าจะเป็นของผลตอบแทนที่ถูกต้องแล้วก็ตาม แต่บ่อยครั้งพวกเขาก็ยังตัดสินใจไม่ตีพ็อตที่จะสอดคล้องกับข้อมูลที่ได้รับ หรือแม้กระทั่งตัดสินใจอย่างไม่เป็นระบบ

ซึ่ง Damodaran [16] สรุปว่าความไม่แน่นอนเหล่านั้นทำให้เกิดยุคใหม่สำหรับการศึกษาทางการเงินที่ถูกเรียกว่า ทฤษฎีจิตวิทยาทางการเงิน โดยนักวิจัยจำนวนมากได้ใช้หลักฐานที่รวบรวมได้จากการทดลองทางจิตวิทยาการลงทุนเพื่อที่จะวิเคราะห์ว่านักลงทุนตอบสนองกับข้อมูลสารสนเทศอย่างไร เพื่อให้สามารถทำนายว่าราคาจะเปลี่ยนแปลงไปอย่างไร โดยผลลัพธ์ที่ได้ออกมา คือ การตอบสนองของนักลงทุนนั้นชัดเจน แต่การทำนายกลับได้ผลที่ไม่ชัดเจน ซึ่งผลที่ไม่ชัดเจนนั้นสามารถอธิบายได้ด้วยทฤษฎีจิตวิทยาการเงิน

3. ทฤษฎีและงานวิจัยที่เกี่ยวข้องกับ

พฤติกรรมทางสังคมออนไลน์ในทวีตเตอร์

Aarts *et al.* [17] ได้รวบรวมงานวิจัยที่เกี่ยวข้องกับพฤติกรรมต่างๆ บนทวีตเตอร์และแยกแยะมันออกเป็น 3 กลุ่ม ดังต่อไปนี้

3.1. บุคลิกของตัวบุคคล (Actor characteristics)

บุคลิกของตัวบุคคลในสังคมออนไลน์สามารถแบ่งย่อยเป็นสองมุมมองกว้างๆ ได้แก่ มุมมองในเรื่องอิทธิพลของข้อมูลสารสนเทศที่มีอิทธิพลต่อตัวบุคคล และมุมมองในเรื่องพฤติกรรมของตัวบุคคลที่ปรับตัวให้เข้ากับสถานการณ์ปกติหรือสถานการณ์ที่ถูกกดดันจากสังคม

ข้อมูลสารสนเทศจะมีอิทธิพลมากกว่าหากข้อมูลสารสนเทศนั้นมาจากบุคคลที่เราเห็นว่ามีความคิดเห็นในแนวทางเดียวกับเรา [18] และดูเหมือนว่าทฤษฎีส่วนใหญ่ที่เกี่ยวข้องกับอิทธิพลทางสังคม จะมีความเห็นไปในทำนองเดียวกันว่า อิทธิพลในกลุ่มทางสังคมนั้นขึ้นอยู่กับกระบวนการตัดสินใจต่อข้อมูลสารสนเทศของแต่ละบุคคลอย่างมีเหตุผล หรือความกดดันที่มีต่อการปรับตัวให้เข้ากับสังคมนั้น แต่ Postmes [19] แสดงให้เห็นว่าในส่วนของการทฤษฎีพฤติกรรมมวลชน (crowd behavior) นั้นจะแตกต่างกันออกไป กล่าวคือ มวลชนสามารถทำให้สมาชิกในกลุ่มคิดอย่างไม่มีเหตุผลได้ และเขายังกล่าวอีกว่าเหตุผลที่ทำให้เกิดการรวมกลุ่มกันทางพฤติกรรมเป็นเพราะแต่ละคนมีความต้องการอะไรที่มากขึ้นหรือน้อยลงในเรื่องเหมือนกัน

มีแบบจำลองจำนวนหนึ่งที่ถูกนำเสนอมาเพื่อทำนายอิทธิพลของตัวบุคคลที่มีต่อระบบสังคมออนไลน์ โดยดูจากลักษณะของบุคคลนั้น ตัวอย่างเช่น Bakshy *et al.* [20] ทำนายโดยใช้ลักษณะของบุคคลดังนี้ จำนวนผู้ที่ติดตามทวีตจากบุคคลนั้นอยู่ (follower), จำนวนเพื่อนในระบบ, จำนวนทวีตที่บุคคลนั้นเคยส่งเข้าระบบ และระยะเวลาที่เข้าร่วมในระบบ แต่เขาก็พบว่าการทำนายนั้นเชื่อถือไม่ได้ และ Mehta *et al.* [21] พบว่าจำนวนของผู้ติดตามทวีตนั้นมีผลต่อการทำนายมากกว่าจำนวนการส่งต่อข้อความทวีตนั้น (retweet)

3.2. บุคลิกของข้อความ

(Message characteristics)

บุคลิกของข้อความเป็นลักษณะสำคัญที่ทำให้มีการกระจายและการแพร่หลายของข้อมูลสารสนเทศ ซึ่งข้อมูลที่ดึงดูดความสนใจจะมีความเป็นไปได้ที่จะถูกเผยแพร่ต่อออกไปมากกว่าข้อความที่ไม่มีบุคลิกเด่น ซึ่ง Berger [22] กล่าวว่า ในอดีตมีการสรุปว่าข้อความที่ดึงดูดความสนใจคือ ข้อความที่ทำให้เกิดความวิตกกังวลขึ้น

ซึ่งเป็นความรู้สึกทางด้านลบ แต่มีการถกเถียงกันว่า ข้อความเชิงบวกก็มีการส่งต่อกันอย่างแพร่หลายได้เช่นกัน

แต่ Berger พบว่าข้อความที่มีการใช้สัญลักษณ์แสดงอารมณ์ (Emoticons) ที่มีลักษณะที่เร้าอารมณ์สูงทั้งด้านบวกและลบ จะช่วยกระตุ้นให้มีการแบ่งปันข้อความนั้นมากกว่าข้อความที่มีการใช้สัญลักษณ์แสดงอารมณ์ที่มีลักษณะที่เร้าอารมณ์ต่ำ และ Bakshy *et al.* [20] พบว่าเนื้อหาของข้อความที่ถูกประเมินว่าน่าสนใจกว่า จะมีการส่งต่อข้อความออกไปมากกว่า ในขณะที่ Lehmann *et al.* [23] แสดงให้เห็นว่า กฎของการออกแรงน้อยที่สุด (principle of least effort) ร่วมกับความสนใจที่มีอยู่จำกัด เป็นส่วนสำคัญของพฤติกรรมการส่งต่อทวิต ซึ่งพวกเขาสังเกตเห็นว่าผู้ใช้จะส่งต่อข้อความก็ต่อเมื่อข้อความนั้นปรากฏให้ผู้ใช้เห็นง่ายที่สุด เช่นข้อความที่แสดงอยู่บริเวณบนสุดของข้อความทั้งหมด

3.3. บุคลิกของเครือข่าย

(Network characteristics)

บุคลิกของเครือข่ายช่วยให้เข้าใจลักษณะเฉพาะของความสัมพันธ์ระหว่างผู้ใช้ทั้งหมด แทนที่จะเป็นเพียงการจำกัดอยู่แค่ลักษณะเฉพาะของผู้ใช้คนใดคนหนึ่งเท่านั้น โดยบุคลิกของเครือข่ายในทวิตเตอร์จะเป็นการกระจายข้อความออกเป็นทอดๆ (cascades) และมีโครงสร้างคล้ายกับโครงสร้างต้นไม้ (tree-like)

Weng *et al.* [24] แสดงให้เห็นว่า บุคลิกของเครือข่ายที่มีการติดต่อกันระหว่างกันในหมู่เพื่อนบนทวิตเตอร์สามารถอธิบายได้ด้วย หลักแห่งความคล้ายคลึง (homophily) ซึ่งก็คือเมื่อผู้ใช้ทวิตเตอร์ลงชื่อเป็นผู้ติดตามทวิตของเพื่อนคนหนึ่งแล้ว เพื่อนคนนั้นก็มักจะลงชื่อติดตามทวิตกลับ เมื่อพวกเขามีเรื่องที่สนใจคล้ายคลึงกัน Gonçalves [25] พบว่า ผู้ใช้สามารถมีความสัมพันธ์อย่างมั่นคงในทวิตเตอร์ได้สูงสุดที่ 100 - 200 คน และจากการศึกษาของ Ardon *et al.* [26] พบว่า หัวข้อที่จะเป็นที่นิยม คือ หัวข้อที่เกิดจากเรื่องที่ไม่ปะติดปะต่อกัน

หลายๆเรื่องที่ใช้หลายๆคนร่วมกันอภิปรายจนเกิดข้อสรุปแล้วรวมกันเป็นเรื่องใหญ่ๆเรื่องหนึ่ง และพวกเขายังได้ข้อสรุปอีกว่า ทวิตเตอร์เป็นสื่อกลางของรูปแบบประชาธิปไตยบางส่วน ซึ่งเห็นได้จากหัวข้อที่เป็นที่นิยม ซึ่งส่วนมากแล้วจะเป็นข้อความที่เกิดจากผู้ที่มีผู้ติดตามจำนวนมาก แต่อย่างไรก็ตาม สำหรับหัวข้อที่จะกลายเป็นที่สนใจอาจเป็นหัวข้อที่มาจากคนที่ไม่มีชื่อเสียงหรือบุคคลทั่วไป ส่วนการศึกษาของ Kwak *et al.* [2] นั้นเป็นการศึกษาในเชิงปริมาณ โดยได้ผลลัพธ์ออกมาว่า จำนวนผู้ตามทวิตไม่ได้กระจายตัวแบบยกกำลัง แต่มีระยะการกระจายตัวอย่างมีประสิทธิภาพ (เส้นผ่านศูนย์กลาง) อยู่ที่ 4.12 ทอด และมีเพียงแค่ 22.1% ของผู้ใช้ที่ติดตามและโต้ตอบทวิตซึ่งกันและกัน พวกเขาพบว่าผู้ใช้มีความสัมพันธ์ที่เชื่อมถึงกันบนทวิตเตอร์อย่างมากเพียง 10 คนเท่านั้น แต่มีการกระจายแบบหางยาว (long tail distribution) ได้ถึงหลายร้อยคน

4. งานวิจัยเกี่ยวกับการหาความสัมพันธ์และอิทธิพลที่เกิดขึ้นระหว่างทวิตเตอร์และตราสารการเงิน

จากที่ได้กล่าวมาข้างต้นจะเห็นได้ว่า ทฤษฎีทางการเงินที่เกี่ยวข้องกับเรื่องข้อมูลสารสนเทศและพฤติกรรมของนักลงทุน มีความสอดคล้องและเกี่ยวข้องกันกับทฤษฎีทางเทคโนโลยีสารสนเทศที่เกี่ยวข้องกับทวิตเตอร์จำนวนมาก และทำให้เกิดการวิจัยหาค่าสหสัมพันธ์กันระหว่างข้อมูลสารสนเทศทางการเงินที่ได้จากทวิตเตอร์และข้อมูลสารสนเทศจากตลาดตราสารการเงิน

ซึ่งก่อนหน้านี้ทวิตเตอร์จะกำเนิดขึ้นมา Antweiler *et al.* [27] ได้ทำการศึกษาและพบว่าข้อมูลสารสนเทศเกี่ยวกับหุ้นที่ได้จากข้อความตามกระดานข่าวในอินเทอร์เน็ตมีผลกระทบต่อตลาด และยังพบอีกว่าข้อความเหล่านั้นช่วยในการทำนายความผันผวนของตลาดได้ ซึ่งผลกระทบต่อผลกำไรที่ได้จากการซื้อขายหุ้น

ตามการทำนายนั้นสูงอย่างมีนัยสำคัญทางสถิติ แต่ก็ยังมีผลน้อยในเชิงเศรษฐกิจ

นอกจากนี้ Nofsinger [10] ยังกล่าวว่า อารมณ์ความรู้สึกในสังคมนั้นมีผลกระทบต่อเศรษฐศาสตร์การเงิน และระดับของการมองโลกในแง่ดีหรือร้ายในสังคมถูกสะท้อนออกมาให้เห็นได้จากอารมณ์ของผู้ตัดสินใจในทางการเงิน และเพราะอารมณ์เหล่านั้นสัมพันธ์ร่วมกับผู้ที่มีส่วนร่วมในทางเศรษฐกิจ จึงนำไปสู่สมมติฐานที่สำคัญ 3 แบบ ได้แก่

1) อารมณ์ความรู้สึกในสังคมที่มีอิทธิพลต่อรูปแบบการตัดสินใจของผู้บริโภค นักลงทุน หรือแม้กระทั่งผู้จัดการบริษัท ในลักษณะที่คล้ายคลึงกัน และอารมณ์ทางสังคมที่สุดขั้วนั้นเป็นลักษณะที่เกิดขึ้นจากการลงทุนหรือกิจกรรมทางธุรกิจที่มองโลกในแง่ดีหรือร้ายมากเกินไป

2) เนื่องจากการซื้อขายหุ้นนั้นมีทั้งเรื่องอารมณ์และเรื่องประสิทธิภาพของตลาดเข้ามาเกี่ยวข้อง ตลาดหลักทรัพย์เองจึงเป็นตัววัดอารมณ์ทางสังคมได้โดยตรง

3) ด้วยเหตุผลที่ลักษณะทั่วไปและลักษณะพิเศษของกิจกรรมทางธุรกิจนั้นจะเป็นไปตามอารมณ์ทางสังคมแทนที่จะเป็นผู้นำอารมณ์ทางสังคม จึงทำให้แนวโน้มของการเคลื่อนไหวในตลาดหลักทรัพย์สามารถช่วยทำนายกิจกรรมทางเศรษฐศาสตร์และการเงินในอนาคตได้

หลังจากที่ทวิตเตอร์ได้ถูกเริ่มใช้งานและเป็นที่นิยมกันอย่างแพร่หลาย จึงเป็นสาเหตุหนึ่งที่ทำให้ข้อมูลสารสนเทศเกี่ยวกับการเงินกระจายสู่สาธารณชนง่ายและแพร่หลายมากขึ้น ทำให้มีการศึกษาด้านการหาความสัมพันธ์และอิทธิพลของอารมณ์ความรู้สึกในสังคมเพิ่มมากขึ้นโดยมุ่งเน้นไปที่ความสัมพันธ์ที่เกิดขึ้นกับทวิตเตอร์เป็นหลัก แต่ยังมีการวิจัยที่ได้รับการยอมรับและตีพิมพ์ในวารสารและการประชุมวิชาการที่มีคุณภาพระดับดีเยี่ยมไม่มากนัก ดังตัวอย่างต่อไปนี้

4.1. Correlation Financial Time Series with Micro-Blogging Activity

Ruiz *et al.* [8] ทำการเก็บและกรองข้อความทวิตที่เกี่ยวข้องกับบริษัทที่อยู่ในตลาดหลักทรัพย์ที่มีชื่ออยู่ในดัชนี S&P500 จำนวน 150 บริษัท (ได้จากการสุ่มซื้อบริษัทขึ้นมา) ซึ่งพวกเขากรองทวิตโดยใช้นิพจน์ปกติ (regular expression) สร้างกฎออกมาเป็นชุดของเงื่อนไข เพื่อนำมาหาชื่อของบริษัท รวมทั้งชื่อย่อหลักทรัพย์ของบริษัท และแฮชแท็ก (hashtag, '#') ที่เกี่ยวข้องกับบริษัท และทำการปรับนิพจน์ปกตินั้นจากการทดลองกับข้อความทวิตที่สุ่มมา 30 ข้อความ จนกว่าจะได้ข้อความที่เกี่ยวข้องเกิน 50% ขึ้นไป และตัดบริษัทที่ไม่สามารถหาวิธีปรับนิพจน์ปกติได้ออกไป เช่น Apple ซึ่งเป็นบริษัทที่มักจะมีชื่ออยู่ในข้อความโฆษณา (spam) จำนวนมาก

ในส่วนของคุณลักษณะเฉพาะที่พวกเขาสกัดออกมาจากทวิตเพื่อที่จะเอามาใช้หาค่าสหสัมพันธ์กับอนุกรมเวลาของตราสารทุนนั้น สามารถแบ่งได้เป็น 2 กลุ่ม โดยลักษณะเฉพาะในกลุ่มแรกจะเป็นลักษณะที่เกิดจากการวัดกิจกรรมที่เกิดขึ้นทั้งหมดในระบบทวิตเตอร์ เช่น จำนวนทวิต, จำนวนการส่งต่อข้อความทวิต, จำนวนของเพื่อนของผู้ที่ทวิต, จำนวนทวิตที่มี URL รวมอยู่ในข้อความ, จำนวนแฮชแท็กที่ถูกใช้ในทวิต และจำนวนทวิตที่พูดถึงผู้ใช้คนอื่น เป็นต้น ส่วนลักษณะเฉพาะในกลุ่มที่สองจะเป็นลักษณะที่เกิดจากการวัดคุณสมบัติของการมีปฏิสัมพันธ์กันระหว่างผู้ใช้ (induced interaction graph) เช่น จำนวนผู้ใช้ และจำนวนผู้ใช้ที่มีการติดต่อกันทั้งแบบทางเดียวและสองทาง เป็นต้น

พวกเขาใช้ ค่าสัมประสิทธิ์สหสัมพันธ์ไขว้ (Cross-Correlation Coefficient, CCF) ในการประเมินความสัมพันธ์ของตัวแปรต่างๆ ในแต่ละช่วงเวลา (time lags) โดยพวกเขาหาค่าสหสัมพันธ์ระหว่างลักษณะเฉพาะต่างๆ จากทั้ง 2 กลุ่ม กับ 1) ปริมาณการซื้อขายของหุ้น และ 2) ราคาหุ้น ผลลัพธ์ที่ได้คือ จำนวนของส่วนที่มีการ

เชื่อมต่อกัน มีค่าสหสัมพันธ์สูงสุดกับทั้งปริมาณการซื้อขายของหุ้นและราคาหุ้น และค่าสหสัมพันธ์ระหว่างลักษณะเฉพาะต่างๆกับปริมาณการซื้อขายของหุ้น และราคาหุ้น โดยคิดกรณีที่ $lag = 0$ มีการเรียงอันดับที่ใกล้เคียงกัน และเมื่อ $lag = -1$ ลักษณะเฉพาะเหล่านั้นมีค่าเป็นบวกซึ่งแสดงถึงว่าลักษณะเฉพาะเหล่านั้นมีความสามารถในการทำนายปริมาณการซื้อขายของหุ้นและราคาหุ้นในวันถัดไป แต่พบว่าการทดสอบกับราคาหุ้นจะได้ค่าสหสัมพันธ์ที่น้อยมาก

จากการทดสอบหาค่าสหสัมพันธ์กับลักษณะเฉพาะต่างๆ กับทั้ง 50 บริษัทแยกกัน แล้วแยกตามประเภทของบริษัท พบว่าค่าสหสัมพันธ์กับลักษณะเฉพาะบางอย่างมีค่าสูงกับเฉพาะบางบริษัทเท่านั้น จึงทำการหาปัจจัยด้วยการแบ่งบริษัททั้ง 50 บริษัทเป็น 3 กลุ่มคือ ต่ำ, กลาง และสูง ตามการจัดอันดับของตัวบ่งชี้ทางการเงินทั้ง 5 (Current Ratio, Gross Profit, Enterprise Value/EBITDA, PEG Ratio, Float and Beta) โดยผลลัพธ์ที่ได้คือ ค่าสหสัมพันธ์สูงกับบริษัทที่มีหนี้ต่ำไม่ว่าบริษัทนั้นจะมีตัวบ่งชี้ทางการเงินที่ดีหรือไม่ ซึ่งอาจจะเกี่ยวข้องกับหุ้นที่ถูกคาดหวังให้ราคาขึ้นอย่างรวดเร็วหรืออาจจะกำลังถูกเลือกสำหรับการถือสถานะขาย นอกจากนี้ข้อความทวิตยังมีค่าสหสัมพันธ์ที่สูงขึ้นกับหุ้นสำหรับบริษัทที่มีค่าเบต้าสูงและมูลค่ารวมหุ้นที่มีให้ซื้อขายในตลาดต่ำ ซึ่งสื่อให้เห็นว่ากิจกรรมบนทวิตเตอร์นั้นมีค่าสหสัมพันธ์ที่สูงกับปริมาณการซื้อขายของหุ้นของบริษัทที่มีสถานะการเงินผันผวน

พวกเขายังได้ลองนำหุ้นที่มีมูลค่าตามราคาตลาดสูงสุด 20 ตัวมารวมกันแล้วแปลงเป็นดัชนีเพื่อมาคำนวณค่าสหสัมพันธ์ และได้ผลว่าค่าสหสัมพันธ์กับปริมาณการซื้อขายนั้นน้อยกว่าเมื่อคิดแยกเป็นบริษัท ซึ่งสาเหตุน่าจะมาจากการหักลบกันของการเพิ่มขึ้นของกิจกรรมของบางบริษัทกับการไม่มีกิจกรรมของอีกบางบริษัท และที่สำคัญ

ค่าความแปรปรวนของจำนวนของส่วนประกอบที่เชื่อมกันมีค่าน้อยกว่าที่คิดแบบแยกบริษัทด้วยเช่นกัน

พวกเขาได้ทดลองเพิ่มจำนวนข้อความทวิตที่นำมาวิเคราะห์โดยลดข้อจำกัดในการกรองทวิต ซึ่งทำให้ได้ข้อความทวิตเพิ่มขึ้นถึง 150 เท่า แต่ในจำนวนนั้นก็มีข้อความขยะเพิ่มขึ้นด้วย และเมื่อทำการวัดค่าสหสัมพันธ์แล้วพบว่า ถึงแม้จะได้ข้อความมาวิเคราะห์จำนวนมากขึ้น แต่มันกลับเป็นการเพิ่มตัวรบกวนมากกว่าจะเป็นประโยชน์จึงทำให้ค่าสหสัมพันธ์น้อยลง

เมื่อได้ผลจากการหาค่าสหสัมพันธ์แล้วพวกเขาทำการศึกษาเพิ่มเติมว่า เมื่อนำค่าสหสัมพันธ์ระหว่างลักษณะเฉพาะของทวิตกับราคาหุ้นมาใช้ร่วมกับกลยุทธ์ในการซื้อขายหุ้นจะช่วยให้สามารถทำนายราคาหุ้นได้แม่นยำขึ้นหรือไม่ โดยพวกเขากำหนดพฤติกรรมของนักลงทุนให้เป็นดังนี้ 1) ตอนเริ่มต้นนักลงทุนมีเงินต้น C_0 2) ในทุกๆเช้าของทุกวัน t เขาจะซื้อหุ้นด้วยเงินทั้งหมดที่มีเวลานั้น C_t โดยนักลงทุนใช้กลยุทธ์หลายอย่างในการเลือกหุ้นที่จะซื้อและแต่ละหุ้นจำนวนเท่าใด โดยเป็นหุ้นที่ถูกสุ่มมาจากหุ้นในดัชนี S&P500 3) นักลงทุนถือหุ้นไว้ทั้งวัน 4) นักลงทุนขายหุ้นทั้งหมดเมื่อเวลาปิดตลาดของวันที่ t และเงินที่ได้จากการขายทั้งหมดจะเป็นเงินต้นใหม่ C_{t+1} และนำกลับไปใช้ในขั้นตอนที่ 2 จนถึงวันสุดท้ายของการจำลอง 5) เปรียบเทียบกับการลงทุนปกติ โดยกลยุทธ์ที่ให้นักลงทุนใช้มี 4 ประเภทคือ 1) แบบสุ่มเลือกหุ้นในทุกๆเช้าและแบ่งจำนวนเงินเท่าๆกันในการซื้อหุ้นแต่ละบริษัท 2) เลือกตามดัชนีการเงินใดการเงินหนึ่ง เช่นเลือกจากมูลค่าตามราคาตลาด หรือ จำนวนหนี้สินทั้งหมด เป็นต้น และแบ่งจำนวนเงินเท่าๆกันในการซื้อหุ้นแต่ละบริษัท 3) ใช้การถดถอยแบบอัตโนมัติ (AR(5)) มาทำนายราคา 4) พวกเขาเสนอให้ใช้การถดถอยแบบอัตโนมัติเชิงเวกเตอร์ (VAR) โดยใช้ข้อมูลราคาและลักษณะเฉพาะของข้อความทวิตเข้าไปร่วมในสมการด้วย โดยแต่ละแบบจำลองถูกฝึกโดยวิธีกำลังสองน้อยที่สุด

(Ordinary Least Squares, OLS) โดยผลลัพธ์ที่ได้คือ การใช้กลยุทธ์ที่พวกเขาเสนอโดยใช้ข้อมูลร่วมกับ ลักษณะเฉพาะนั้นให้ผลที่ดีที่สุดและเป็นเพียงกลยุทธ์เดียว ที่ให้ผลลัพธ์เป็นบวกอีกด้วย โดยจำนวนส่วนประกอบที่ ติดต่อกันเป็นลักษณะเฉพาะที่ให้ผลลัพธ์ที่ดีที่สุด

โดยสรุปพวกเขาพบว่า ลักษณะเฉพาะของทวิตเตอร์มี ความสัมพันธ์กับปริมาณการซื้อขายหลักทรัพย์มากกว่า การเคลื่อนไหวของราคาของหลักทรัพย์ แต่อย่างไรก็ตาม ข้อมูลสารสนเทศที่ได้จากทวิตเตอร์สามารถช่วยเป็นตัว ผลักดันให้กลยุทธ์ในการซื้อขายหลักทรัพย์มีประสิทธิภาพ มากขึ้น

4.2. Twitter mood predicts the stock market

Bollen *et al.* [11] ทำการวิจัยเพิ่มเติมจากเศรษฐศาสตร์ เชิงพฤติกรรม (Behavioral economics) ซึ่งกล่าวถึง ผลกระทบทางด้านอารมณ์ที่มีต่อพฤติกรรมของรายบุคคล โดยพวกเขาทำการทดลองหาว่า อารมณ์ที่ได้จาก ทวิตเตอร์สามารถใช้ในการทำนายการเคลื่อนไหวของ ตลาดหลักทรัพย์ได้หรือไม่ ซึ่งมันสามารถบอกเป็นนัยได้ว่า อารมณ์มีผลกระทบต่อพฤติกรรมในสังคมในวงกว้าง หรือไม่ ซึ่งสอดคล้องกับทฤษฎีจิตวิทยาการเงิน (Behavioral Finance)

โดย Bollen *et al.* ใช้ทวิตที่ได้มาจากการสุ่มจาก ทวิตเตอร์ตั้งแต่วันที่ 28 กุมภาพันธ์ 2551 ถึงวันที่ 19 ธันวาคม 2551 จำนวนเกือบ 10 ล้านทวิต และนำ ข้อความทวิตทั้งหมดมาวิเคราะห์ร่วมกับอนุกรมเวลาของ ราคาปิดของดัชนีตลาดหลักทรัพย์ดาวน์โจน (Dow Jones Index) ผ่านการวิจัย 3 ขั้นตอนดังนี้

1) ระบุอารมณ์ของแต่ละทวิต โดยพวกเขาเลือกใช้ เครื่องมือ 2 ประเภทเพื่อประเมินอารมณ์จากข้อความใน ทวิต คือ OpinionFinder ซึ่งจะระบุข้อความนั้นแสดง อารมณ์ทางด้านบวกหรือลบ และ Google-Profile of Mood States (GPOMS) ซึ่งจะไม่ระบุออกมาโดยตรงว่า

ข้อความนั้นเป็นอารมณ์ไหนแต่จะให้เป็นการมาทั้งหมด 6 มิติ ได้แก่ สงบ (Calm), ตื่นตัว (Alert), มั่นใจ (Sure), มี ชีวิตชีวา (Vital), อ่อนโยน (Kind) และมีความสุข (Happy) ซึ่งเมื่อรวมค่าทางอารมณ์จากทั้งสองเครื่องมือเข้าด้วยกัน ทำให้ 1 ข้อความจะมีค่าทางอารมณ์ทั้งหมด 7 ค่า ซึ่งเอา มาคำนวณร่วมกับดัชนีตลาดดาวน์โจน

พวกเขาทำการตรวจสอบด้วยว่าค่าที่ได้จากเครื่องมือ ทั้งสองถูกต้องและสัมพันธ์กับอารมณ์ที่เกิดขึ้นจริงหรือไม่ โดยการเก็บข้อมูลจากข้อความทวิตในช่วงที่มีเหตุการณ์ สำคัญ 2 เหตุการณ์คือ เลือกตั้งประธานาธิบดี และวัน ขอบคุณพระเจ้า โดย OF สามารถระบุอารมณ์ของ เหตุการณ์ทั้งสองได้ด้วยการเพิ่มขึ้นของอารมณ์บวก เล็กน้อยเป็นระยะเวลาช่วงสั้นๆ และ GPOMS สามารถ แสดงอารมณ์ที่หลากหลายที่มีต่อเหตุการณ์เลือกตั้งได้ถึง 3 วันช่วงเลือกตั้ง และเพิ่มขึ้นมากของอารมณ์มีความสุข ในวันขอบคุณพระเจ้า

นอกจากนี้พวกเขายังทดสอบความสัมพันธ์ระหว่าง อารมณ์ต่างๆทั้งหมด 7 อารมณ์ที่ได้จากทั้งสองเครื่องมือ โดยใช้การวิเคราะห์การถดถอยเชิงเส้นตรงแบบพหุแล้ว พบว่าอารมณ์จาก OF สัมพันธ์กับ Sure, Vital และ Happy จาก GPOMS อย่างมีนัยสำคัญ แต่ไม่สัมพันธ์กับ อารมณ์ Calm, Alert และ Kind จาก GPOMS ซึ่งทำให้ สรุปได้ว่าอารมณ์จากทั้งสองเครื่องมือมีส่วนที่ทับซ้อนกัน บ้างแต่ไม่ทุกมิติ โดย GPOMS ให้ได้หลาหม่มมมมากกว่า OF

2) หาความสัมพันธ์ (Causality) ระหว่างค่าอารมณ์ ทั้ง 7 กับการเคลื่อนไหวของดัชนีตลาดดาวน์โจน โดยใช้ การวิเคราะห์ความสัมพันธ์แบบเป็นเหตุเป็นผลของแกรง เจอร์ (Granger causality) โดยมีเพียงอารมณ์ Calm จาก GPOMS เพียงอารมณ์เดียวเท่านั้นที่สามารถช่วยทำนาย ได้อย่างมีนัยสำคัญทางสถิติ แต่ก็ยังมีบางเหตุการณ์ที่ ตลาดเคลื่อนไหวซึ่งเป็นเหตุการณ์ที่เกิดขึ้นอย่างไม่คาดคิดจึง ไม่สามารถคาดการณ์โดยอารมณ์ของมวลชนได้

3) ใช้แบบจำลองโครงข่ายประสาทที่จัดการได้ด้วยตัวเองแบบฟัซซี่ (Self-Organizing Fuzzy Neural Network model, SOFNN) ซึ่งเป็นแบบจำลองหนึ่งที่น่าสนใจในการทำนายการเคลื่อนไหวของดัชนีดาวโจนส์ โดยทำการทดสอบจากข้อมูล 2 ชุดคือ i) ค่าดัชนีดาวโจนส์ ในช่วง 3 วันก่อนหน้า และ ii) ค่าดัชนีร่วมกับค่าอารมณ์ต่างๆที่ได้มาจากขั้นตอนแรก มาเปรียบเทียบกับเมื่อนำค่าอารมณ์ต่างๆมาใช้ร่วมกันในแบบจำลองจะสามารถช่วยปรับปรุงการทำนายให้ถูกต้องมากขึ้นหรือไม่

ผลจากการวิเคราะห์หาความสัมพันธ์แบบแรงจูงใจระบุว่ามิเพียง อารมณ์สงบ (Calm) และบางระดับของ มีความสุข (Happy) เท่านั้นที่มีความสัมพันธ์กับดัชนีดาวโจนส์ ซึ่งเมื่อนำค่าทางอารมณ์ต่างๆไปทดสอบร่วมกับแบบจำลอง SOFNN แล้วได้ผลลัพธ์ดังนี้คือ มิเพียงค่าของอารมณ์สงบ (Calm) เท่านั้นที่สามารถช่วยให้การทำนายได้ผลที่ถูกต้องมากขึ้น ซึ่งก็คือได้ค่ามากที่สุดเมื่อเทียบกับการใช้ค่าอารมณ์อื่นๆ โดยมีนสามารถช่วยลดค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ (Mean Absolute Percentage Error) ให้เหลือเพียง 1.83% จาก 2.13% และช่วยให้ค่าความถูกต้องเพิ่มขึ้นจาก 73.3% เป็น 86.7% และเป็นที่น่าสนใจว่า เมื่อนำค่าอารมณ์ที่มีความสุขมาคิดรวมด้วยกับค่าของอารมณ์สงบ ทำให้ช่วยลดค่าเฉลี่ยความคลาดเคลื่อนสมบูรณ์ได้มากอย่างมีนัยสำคัญ และค่าความถูกต้องยังคงมีค่าสูงอยู่ที่ 80% ซึ่งสูงกว่าการใช้แบบจำลองแบบไม่มีค่าอารมณ์ใดๆมาช่วย

5. การอภิปราย

จากงานวิจัยที่ผ่านมาสามารถสังเกตเห็นได้ว่างานหลักของการหาความสัมพันธ์ระหว่างทวีตเตอร์และตราสารการเงินเกี่ยวข้องกับ

1) การศึกษาการเก็บและกรองข้อมูลเฉพาะส่วนที่ต้องการจากทวีตทั้งหมด รวมถึงวิธีการแยกแยะทวีตตามลักษณะเฉพาะต่างๆ

2) การศึกษาลักษณะเฉพาะของทวีตที่สามารถนำมาใช้ในการหาค่าสหสัมพันธ์ได้ เช่น จำนวนการส่งต่อของแต่ละทวีต หรือแม้กระทั่งอารมณ์ของแต่ละทวีต เป็นต้น และการศึกษาวิธีในการคำนวณเพื่อนำค่าของลักษณะเฉพาะนั้นมาใช้ประโยชน์

3) วิธีคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ซึ่งมีให้เลือกอยู่ด้วยกันหลายวิธี เช่น Pearson's correlation coefficient และ Kendall tau rank correlation coefficient เป็นต้น เพื่อนำมาคำนวณและเปรียบเทียบค่าสหสัมพันธ์ หรือเลือกวิธีการวิเคราะห์การถดถอย เพื่อหาว่า ลักษณะเฉพาะต่างๆของทวีตเตอร์สามารถช่วยในการทำนายความเคลื่อนไหวของตลาดการเงินได้หรือไม่และอย่างไร

โดยส่วนที่ยังมีปัญหามากที่สุดเป็นส่วนของการเก็บกรองและแยกแยะทวีต เนื่องจากทวีตเตอร์อนุญาตให้รับข้อมูลโดยผ่าน Twitter API เพียงแค่ 1% ของทวีตเตอร์ทั้งหมดเท่านั้น นอกจากนี้ข้อมูล 1% ที่ได้มายังแตกต่างจากข้อมูลที่ได้รับการสุ่มแบบปกติอีกด้วย [28] ซึ่งอาจจะมีความกระทบต่อความถูกต้องของผลการทดลองได้ จึงต้องหาวิธีการกรองข้อมูลให้มีประสิทธิภาพมากที่สุด และถึงแม้ว่าจะเก็บข้อมูลมาได้เพียง 1% แต่ก็เพียง 1% จากประมาณ 500 ล้านทวีตต่อวัน ซึ่งเป็นข้อมูลจำนวนมากมหาศาลมาก รวมทั้งข้อความส่วนใหญ่ที่ได้จากทวีตเตอร์จะเป็นข้อความที่ไม่ใช่โครงสร้างทางภาษาอย่างเป็นทางการ และด้วยเหตุผลที่ทวีตเตอร์จำกัดให้ใช้ได้แค่ 140 ตัวอักษรต่อทวีต จึงทำให้มีการใช้ตัวย่อ และคำแสดงอยู่เป็นจำนวนมาก จึงทำให้ยากต่อการนำมาวิเคราะห์เพื่อแยกแยะทวีตที่มีลักษณะเฉพาะที่แตกต่างกัน

แต่การกรองและแยกแยะนั้นก็ยังไม่มีประสิทธิภาพเพียงพอที่จะสามารถกรองและแยกแยะทวีตที่ต้องการออกมาได้ถูกต้องทั้งหมดดังตัวอย่างในหัวข้อ 4.1 ที่ไม่สามารถนำบริษัท Apple ซึ่งเป็นบริษัทที่มีมูลค่าในตลาดสูงมากมาวิเคราะห์ได้ ถึงแม้ว่าจะมีงานวิจัยจำนวนหนึ่งที่

มีการนำเทคนิคการประมวลผลภาษาธรรมชาติ (Natural Language Processing) [29,30,31] มาจำแนกชื่อเฉพาะจากข้อความในทวิต ซึ่งมีส่วนช่วยให้สามารถกรองและแยกแยะทวิตได้ดีขึ้น แต่ก็ยังไม่ดีพอที่จะสามารถนำมาใช้กรองและแยกแยะทวิตที่เกี่ยวข้องกับหลักทรัพย์หรือแยกแยะทวิตที่มีลักษณะเฉพาะที่ต้องการได้ทั้งหมด แต่อย่างไรก็ตาม การนำตัวกรองสามัญ (Naïve Bayes) และการจำแนกด้วยซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) [32] รวมถึงการใช้เครื่องมือเวก้า (Weka) [33] อาจจะช่วยให้ออกมามีประสิทธิภาพดีขึ้นได้อีก

จะเห็นได้ว่าลักษณะเฉพาะที่ใช้ในตัวอย่างงานวิจัยตัวอย่างในหัวข้อ 4.1 นั้นยังไม่ครอบคลุมทั้งหมด ยังมีลักษณะเฉพาะอื่นอีกจำนวนมากจากหัวข้อที่ 3 ที่ยังไม่ได้ถูกนำมาใช้ร่วมกับการวิเคราะห์หาความสัมพันธ์กับการเคลื่อนไหวของตลาดการเงิน

ในส่วนของการวิเคราะห์ความสัมพันธ์ นอกจากใช้การหาค่าสหสัมพันธ์และการวิเคราะห์การถดถอย (Regression Analysis) ยังสามารถใช้ฟังก์ชันความคล้ายคลึง (Similarity functions) [34] เช่น Kullback-Leibler divergence, Jensen-Shannon divergence และ Skew divergence เป็นต้น มาวิเคราะห์ร่วมด้วย นอกจากนั้นยังสามารถใช้วิธีอื่นๆเข้ามาช่วยได้ด้วยได้อีกหลายวิธี เช่น การทำเหมืองข้อมูลที่ใช้กฎความสัมพันธ์ สำหรับข้อมูลเชิงโครงสร้างเข้ามาช่วยวิเคราะห์ [35]

นอกจากนี้แม้ว่าจะมีผลจากการวิจัยบางส่วนที่ดูเหมือนจะมีผลสอดคล้องกับสมมติฐานตลาดที่มีประสิทธิภาพ และบางส่วนของดูเหมือนจะมีผลสอดคล้องกับทฤษฎีจิตวิทยาการเงิน ซึ่งก็มีความเป็นไปได้ที่จะสามารถนำไปใช้ในการโต้แย้งในงานวิจัยทางการเงินได้ แต่ก็ยังไม่มีข้อสรุปหรือการโต้แย้งในด้านนี้จากทางฝั่งของเทคโนโลยีสารสนเทศอย่างจริงจัง

7. สรุป

บทความนี้เป็นบทความที่รวบรวมทฤษฎีที่เกี่ยวข้องกับการหาความสัมพันธ์ระหว่างข้อมูลสารสนเทศที่ได้จากทวิตเตอร์ รวมทั้งลักษณะเฉพาะต่างๆของทวิต กับความเคลื่อนไหวของตราสารทางการเงิน เช่น ปริมาณการซื้อขาย และราคาของหลักทรัพย์ เป็นต้น โดยจะเห็นได้ว่าทฤษฎีทางการเงินที่กล่าวถึงเรื่องข้อมูลสารสนเทศ และพฤติกรรมทางสังคมทางการเงินนั้น สอดคล้องกับทฤษฎีและผลการวิจัยจากข้อมูลสารสนเทศจากทวิตเตอร์และพฤติกรรมในสังคมของผู้ใช้ทวิตเตอร์ ซึ่งถูกนำมาใช้ศึกษาความสัมพันธ์และอิทธิพลที่เกิดขึ้นระหว่างทวิตเตอร์และตราสารการเงิน จากการศึกษาพบว่ายังมีปัญหาอยู่อีกมาก โดยเฉพาะอย่างยิ่งในส่วนของการเก็บ การกรอง และการแยกแยะทวิตทั้งหมด และยังมีส่วนของลักษณะเฉพาะต่างๆในทวิตเตอร์อีกมากมายที่ยังไม่มีการศึกษา รวมทั้งการใช้ทฤษฎีต่างๆที่นำมาใช้วิเคราะห์ความสัมพันธ์ของทวิตเตอร์ที่มีกับตลาดตราสารการเงิน และทฤษฎีที่นำมาใช้ทดสอบสมมติฐานที่ว่าทวิตเตอร์สามารถช่วยในเรื่องการทำนายการเคลื่อนไหวของตราสารการเงิน ยังนำมาใช้ได้ไม่ครอบคลุมทุกกรณี

กิตติกรรมประกาศ

บทความนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือจาก รศ. ดร. พรฤดี เนติโสภาคกุล ที่ให้ความรู้และคำแนะนำในด้านการประมวลผลภาษาธรรมชาติ รศ.ดร. วรพจน์ กรีสระเดช ที่ให้ความรู้และคำแนะนำทางด้านการทำเหมืองข้อมูล และ ดร. สุภารัตน์ ดันทะนงศักดิ์กุล ที่ให้ความรู้และคำแนะนำเกี่ยวกับทฤษฎีทางการเงิน ผู้ศึกษาจึงขอกราบขอบพระคุณเป็นอย่างสูง

เอกสารอ้างอิง

- [1] Form S-1 Registration Statement of Twitter Inc., the United States Securities and Exchange Commission, Oct. 3, 2013. [Online]. Available: <http://www.sec.gov/Archives/edgar/data/1418091/000119312513390321/d564001ds1.htm> [Accessed: Mar. 10, 2014]
- [2] H. Kwak, C. Lee, H. Park, and S. Moon, "What is twitter, a social network or a news media?" in Proceedings of the 19th International Conference on World Wide Web, ser. WWW '10. ACM, pp. 591–600., 2010
- [3] S. Petrovic, et al., "Can twitter replace newswire for breaking news," in Seventh International AAAI Conference on Weblogs and Social Media, 2013.
- [4] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes twitter users: Real-time event detection by social sensors," in Proceedings of the 19th International Conference on World Wide Web. ACM, pp. 851– 860, 2010
- [5] S. Vieweg et al, "Microblogging during two natural hazards events: What twitter may contribute to situational awareness," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 1079–1088., 2010
- [6] Erin Kim, "Twitter unveils 'cashtags' to track stock symbols." Internet: <http://money.cnn.com/2012/07/31/technology/twitter-cashtag/index.htm>, Jul. 31, 2012. [Accessed: Mar. 10, 2014]
- [7] E. F. Fama, "Efficient capital markets: II," The journal of finance, vol.46, no. 5, pp. 1575–1617, 1991.
- [8] E. J. Ruiz et al, "Correlating financial time series with micro-blogging activity," in Proceedings of the Fifth ACM International Conference on Web Search and Data Mining. ACM, pp. 513–522., 2012
- [9] Y. Mao et al, "Correlating S&P 500 stocks with twitter data," in Proceedings of the First ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research. ACM, pp. 69–72., 2012
- [10] J. R. Nofsinger, "Social mood and financial economics," Journal of Behavioral Finance, vol. 6, no. 3, pp. 144–160, 2005.
- [11] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," Journal of Computational Science, vol. 2, no. 1, pp. 1 – 8, 2011.
- [12] W. R. Lasher, Financial management: A practical approach. Thomson South-Western, 2008.
- [13] Z. Bodie, A. Kane, and A. J. Marcus, Investments 7th International edition. Published by McGraw-Hill/Irwin, a business unit of the McGraw-Hill Companies, 2008.
- [14] L. Chincarini and D. Kim, Quantitative equity portfolio management: an active approach to portfolio construction and management. McGraw Hill Professional, 2006.
- [15] R. J. Shiller, "From efficient markets theory to behavioral finance," Journal of economic perspectives, pp. 83–104, 2003.

- [16] A. Damodaran, *Investment valuation: Tools and techniques for determining the value of any asset*. John Wiley & Sons, 2012.
- [17] O. Aarts et al. "Online social behavior in twitter: A literature review," in *Data Mining Workshops (ICDMW)*, 2012 IEEE 12th International Conference on, pp. 739–746, Dec 2012
- [18] J. C. Turner, *Social influence*. Thomson Brooks/Cole Publishing Co, 1991
- [19] T. Postmes, "The psychological dimensions of collective action, online," in *The Oxford handbook of internet psychology*, A. Joinson, K. McKenna, T. Postmes, and U. Reips, Eds. Oxford University Press, ch. 12, pp. 165–184, 2007
- [20] E. Bakshy et al., "Everyone's an influencer: Quantifying influence on twitter," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*. ACM, pp. 65–74, 2011
- [21] R. Mehta et al., "Sentiment analysis and influence tracking using twitter," *International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE)*, vol. 1, no. 2, 2012.
- [22] J. Berger, "Arousal increases social transmission of information," *Psychological Science*, vol. 22, no. 7, pp. 891–893, 2011.
- [23] J. Lehmann et al., "Dynamical classes of collective attention in twitter," in *Proceedings of the 21st International Conference on World Wide Web*. ACM, pp. 251–260, 2012
- [24] J. Weng et al., "Twitterrank: Finding topic-sensitive influential twitterers," in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*. ACM, pp. 261–270, 2010
- [25] B. Gonçalves, N. Perra, and A. Vespignani, "Modeling users' activity on twitter networks: Validation of Dunbar's number," *PLoS ONE*, vol. 6, no. 8, Aug. 2011.
- [26] S. Ardon et al., "Spatio-temporal analysis of topic popularity in twitter," *CoRR*, vol. abs/1111.2904, 2011.
- [27] W. Antweiler and M. Z. Frank, "Is all that talk just noise? The information content of internet stock message boards," *The Journal of Finance*, vol. 59, no. 3, pp. 1259–1294, 2004.
- [28] F. Morstatter et al., "Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose." in *Seventh International AAAI Conference on Weblogs and Social Media*, 2013.
- [29] X. Liu and M. Zhou, "Two-stage NER for tweets with clustering," *Information Processing & Management*, vol. 49, no. 1, pp. 264 – 273, 2013.
- [30] C. Li et al., "TwiNER: Named entity recognition in targeted twitter stream," in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, pp. 721–730., 2012
- [31] N. Kanungsukkasem, P. Netisopakul, and T. Leelanupab, "Recognition of NASDAQ stock symbols in tweets," in *2014 6th International Conference on Knowledge and Smart Technology (KST2014)*, Jan. 2014.

- [32] N. Collier and S. Doan, "Syndromic classification of twitter messages," in *Electronic Healthcare*, vol. 91, pp. 186–195., 2012
- [33] I. H. Witten, E. Frank, and A. Mark, "Hall. 2011. Data Mining: Practical machine learning tools and techniques," 2011.
- [34] L. Lee. Measures of Distributional Similarity. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics (ACL'99)*, pages 25–32, 1999.
- [35] L. Singh et al., "An algorithm for constrained association rule mining in semi-structured data," in *Methodologies for Knowledge Discovery and Data Mining*, N. Zhong and L. Zhou, Eds. Springer Berlin Heidelberg, vol. 1574, pp. 148–158, 1999.