การเปรียบเทียบประสิทธิภาพการตรวจจับการทุจริตด้วยการเพิ่มชื่อเสียงบนเว็บไซต์การประมูลออนไลน์ระหว่างเอนโทรปีของ Shannon และนาอีฟเบย์

Performance Comparison for Inflated Reputation Fraud Detection in Online Auction Website between Shannon's Entropy and Naïve Bayes

ดร.ลักษมี โขมโนทัย

โปรแกรมวิชาบริหารธุรกิจ (คอมพิวเตอร์ธุรกิจ) คณะวิทยาการจัดการ มหาวิทยาลัยราชภัฏนครราชสีมา
E-mail: laksamee.k@nrru.ac.th

บทคัดย่อ : การทุจริตการประมูลออนไลน์เป็นปัญหาที่สำคัญและมีความวิกฤตต่อเว็บไซต์การประมูลออนไลน์ ในงานวิจัยที่ผ่านมาได้มีการนำเสนอแนวคิดในการคำนวณความหลากหลายของเพื่อนบ้านโดยใช้วิธีเอนโทรปีของ Shannon เพื่อใช้ในการจำแนก ผู้ใช้งานที่มีเจตนาทุจริต (Fraudster) ออกจากผู้ใช้งานทั่วไปที่ไม่มีเจตนาร้ายแอบแฝง (Legitimate user) ในงานวิจัยฉบับนี้นำเสนอวิธีการคำนวณความหลากหลายของเพื่อนบ้านด้วยวิธีความน่าจะเป็นของนาอีฟเบย์ จากนั้นทำการทดสอบประสิทธิภาพในการจำแนก Fraudster กับข้อมูลการทำธุรกรรมการซื้อ-ขายที่เกิดขึ้นจริงบนเว็บไซต์การประมูลออนไลน์ จากผลการทดสอบประสิทธิภาพแสดงให้เห็นว่าการคำนวณความหลากหลายของเพื่อนบ้านทั้งสองวิธีมีประสิทธิภาพในการจำแนก Fraudster ที่ใกล้เคียงกัน

คำสำคัญ : การทุจริตด้วยการเพิ่มชื่อเสียง, การตรวจจับการทุจริต, นาอีฟเบย์, การประมูลออนไลน์

ABSTRACT : Online auction fraud is a crucial problem in online auction website. Recently, Shannon entropy has been proposed for calculating the neighbor diversity to classify fraudsters from legitimate users. In this study, we propose to use Naïve Bayes as different method for quantifying the neighbor diversity. We then apply the neighbor diversity based on Naïve Bayes for classifying fraudsters. The dataset used in this study was collected from a real world auction website. The experimental results present that the classification performance of these two neighbor diversities are similar.

Keywords : Inflated reputation fraud, Fraudster detection, Naïve Bayes, Online auction

1. บทนำ

ในภาวการณ์ปัจจุบันที่เทคโนโลยีอินเทอร์เน็ตมีความเจริญก้าวหน้าส่งผลให้การชอปปิ้งออนไลน์และการประมูลออนไลน์ได้รับความนิยมอย่างแพร่หลายในหมู่ผู้ใช้งานอินเทอร์เน็ต เนื่องจากผู้คนสามารถค้นหาและเลือกซื้อสินค้าประเภทต่างๆ ที่ตรงกับความต้องการได้ในราคาถูกเมื่อเปรียบเทียบกับร้านค้าจริงที่มีหน้าร้าน (Physical store) ทั้งนี้เพราะผู้จำหน่ายสินค้าที่เลือกใช้ช่องทางการประมูลออนไลน์เป็นช่องทางในการจำหน่ายมีการเสนอขายสินค้าในราคาที่ต่ำกว่าที่จำหน่ายทั่วไปตามท้องตลาด ด้วยเหตุนี้จึงทำให้เว็บไซต์ดังกล่าวสามารถจูงใจให้ผู้ที่สนใจในการชอปปิ้งที่เป็นผู้ใช้งานทั่วไปและไม่มีเจตนาร้ายแอบแฝง (Legitimate user) อีกทั้งยังเป็นที่ดึงดูดผู้ที่มีวัตถุประสงค์ที่จะหลอกลวงผู้ใช้งานคนอื่น (Fraudster) เข้าร่วมในการดำเนินธุรกรรมบนเว็บไซต์ดังกล่าว ในแผนการหลอกลวงผู้ใช้งานคนอื่นๆ Fraudster จะสร้างรายการธุรกรรมปลอมเพื่อใช้ในการดำเนินการซื้อ-ขายกับ Legitimate user โดยเฉพาะผู้ใช้งานที่เป็นผู้ซื้อ ทั้งนี้ตัวอย่างความเสียหายที่อาจเกิดขึ้นกับผู้ซื้อภายหลังจากการดำเนินธุรกรรมเสร็จสิ้น เช่น ได้รับผลิตภัณฑ์ที่ไม่ใช่ของแท้ (Non genuine product) ชำระค่าสินค้าแล้วแต่ไม่ได้รับสินค้าตามกำหนด หรือ ได้รับสินค้าที่มีความแตกต่างจากที่ได้สั่งซื้อ เพื่อเป็นการแก้ไขและป้องกันการเกิดปัญหาดังกล่าว โดยส่วนใหญ่แล้วเว็บไซต์ชอปปิ้งออนไลน์และเว็บไซต์การประมูลออนไลน์จะมีการจัดเตรียมระบบการใช้ชื่อเสียง (Reputation system) ไว้เพื่อช่วยเหลือผู้ใช้งาน ทั้งนี้ Reputation system เป็นระบบที่ช่วยให้ผู้ใช้งานสามารถดำเนินการตัดสินใจได้ว่าจะซื้อหรือไม่ซื้อสินค้าจากผู้ขายคนดังกล่าว การทำงานของ Reputation system นั้นผู้ที่ดำเนินธุรกรรมการซื้อ-ขายทั้งในฐานะที่เป็นผู้ซื้อและผู้ขายจะถูกร้องขอให้กำหนดระดับคะแนนให้แก่คู่ค้าแต่ละฝ่ายภายหลังจากการดำเนินธุรกรรมการซื้อ-ขายได้สิ้นสุดลง ความมีชื่อเสียง (Reputation) ของผู้ใช้งานแต่ละคนจะขึ้นอยู่กับคะแนนที่ถูกคำนวณจากระดับคะแนนที่ได้มาจากประวัติการดำเนินธุรกรรม ดังนั้นผู้ใช้งานที่มีคะแนนชื่อเสียง (Reputation score) อยู่ในระดับสูงจะได้รับความเชื่อถืออย่างมากจากผู้ใช้งาน โดยเฉพาะอย่างยิ่งผู้ใช้งานที่เป็นผู้ซื้อ ดังนั้นผู้ขายที่มีชื่อเสียงดีจึงมีแนวโน้มที่จะจูงใจลูกค้าและสามารถเพิ่มยอดขายได้มากยิ่งขึ้น [1]

ดังที่ได้กล่าวมาแล้วว่า Reputation score เกิดจากการคำนวณคะแนนที่เกิดจากการดำเนินธุรกรรมการ-ซื้อขายในอดีตซึ่งต้องเป็นคะแนนที่อยู่ในระดับดี ทั้งนี้การได้มาซึ่งชื่อเสียงในระดับสูง (High reputation) จำเป็นต้องอาศัยเวลาในการรวบรวมคะแนนตามเกณฑ์ดังกล่าว อย่างไรก็ตาม Fraudster สามารถสร้างชื่อเสียงให้กับตนเองได้อย่างรวดเร็วด้วยการดำเนินธุรกรรมการซื้อ-ขายปลอมด้วยกันเองภายในกลุ่มผู้สมรู้ร่วมคิด จากนั้นจึงมอบระดับคะแนนที่ดีให้ซึ่งกันและกันภายหลังจากการดำเนินธุรกรรมได้เสร็จสิ้นลง Fraudster ที่ดำเนินการทุจริตในลักษณะนี้จะถูกเรียกว่า “การทุจริตด้วยการเพิ่มชื่อเสียง” (Inflated reputation fraud) [2] ซึ่งการทุจริตในลักษณะนี้เป็นภัยอย่างร้ายแรงต่อเว็บไซต์การประมูลออนไลน์รวมทั้งธุรกิจออนไลน์ในลักษณะอื่นๆ สำหรับการตรวจจับ Inflated reputation fraud บนเว็บไซต์การประมูลออนไลน์นั้นมีอยู่หลายรูปแบบ จากงานวิจัยที่ผ่านมามีผู้วิจัยหลาย

ท่านใช้แนวคิดของกราฟเครือข่าย (Network graph) ในการตรวจจับ Fraudster ที่ทำงานร่วมกับผู้สมรู้ร่วมคิดในการพยายามเพิ่ม (Boost up) Reputation score [2-6] จากแนวคิดการใช้งาน Network graph ดังกล่าวทำให้พบว่าการใช้วิธีการวิเคราะห์เครือข่ายสังคม (Social Network Analysis : SNA) เป็นวิธีการที่สามารถใช้ในการตรวจจับ Fraudster และกลุ่มของผู้สมรู้ร่วมคิดได้อย่างมีประสิทธิภาพ [2, 5, 6]

ในงานวิจัยของ Lin และ Khomnotai [7] ได้นำเสนอแนวคิดเกี่ยวกับ “ความหลากหลายของเพื่อนบ้าน” (Neighbor Diversity : ND) ในการตรวจจับ Inflated reputation fraud โดยที่ความหลากหลายดังกล่าวของผู้ใช้งานเว็บไซต์การประมูลออนไลน์แต่ละคนจะถูกคำนวณจากความหลากหลายของคู่ค้าที่มีการดำเนินธุรกรรมการซื้อ-ขายกับผู้ใช้งานเว็บไซต์การประมูลออนไลน์ ทั้งนี้ในงานวิจัยดังกล่าวได้นำวิธีการเอนโทรปีของ Shannon (Shannon entropy) [8] มาใช้ในการคำนวณ ND ผลลัพธ์จากงานวิจัยได้แสดงให้เห็นว่า ND ที่ถูกคำนวณจากจำนวนระดับคะแนนที่ได้รับจากการทำธุรกรรม (Neighbor Diversity on the number of received ratings : NDr) มีประสิทธิภาพสูงกว่าวิธีการที่ใช้ k-core และ/หรือ Center Weight (CW) ที่ได้ถูกนำเสนอไว้ในงานวิจัยที่ผ่านมาก่อนหน้านี้ [2, 5, 6]

ทั้งนี้วิธีการเอนโทรปีของ Shannon (Shannon entropy) [8] ได้ถูกนำมาใช้ในการคำนวณ ND ในงานวิจัยดังกล่าว ผลลัพธ์จากงานวิจัยได้แสดงให้เห็นว่า ND ที่ถูกคำนวณจากจำนวนระดับคะแนนที่ได้รับจากการทำธุรกรรม (Neighbor Diversity on the number of received ratings : NDr) มีประสิทธิภาพสูงกว่าวิธีการที่ใช้ k-core และ/หรือ Center Weight (CW) ที่ได้ถูกนำเสนอไว้ในงานวิจัยที่ผ่านมา [2, 5, 6]

สำหรับงานวิจัยฉบับนี้ผู้วิจัยได้นำเสนอการคำนวณความหลากหลายของเพื่อนบ้านโดยวิธีความน่าจะเป็นของนาอีฟเบย์ (Naïve Bayes) และทำการทดสอบประสิทธิภาพในการจำแนก Fraudster เพื่อเปรียบเทียบกับการคำนวณความหลากหลายของเพื่อนบ้านโดยวิธีวิธีการเอนโทรปีของ Shannon

เนื้อหาของงานวิจัยฉบับนี้ได้แบ่งออกเป็นหัวข้อย่อยดังต่อไปนี้ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 3 การดำเนินงานวิจัย 4 การตั้งค่าการทดลองและผลการทดลอง 5 อภิปรายผลการทดลอง และสรุปผลการทดลองในหัวข้อที่ 6

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในส่วนนี้ประกอบด้วยทฤษฎี และงานวิจัยที่ผ่านมาที่เกี่ยวข้องกับการตรวจจับ Inflated reputation fraud จากเว็บไซต์การประมูลออนไลน์ ซึ่งมีรายละเอียดดังนี้

2.1 ทฤษฎีการวิเคราะห์เครือข่ายสังคม (Social Network Analysis)

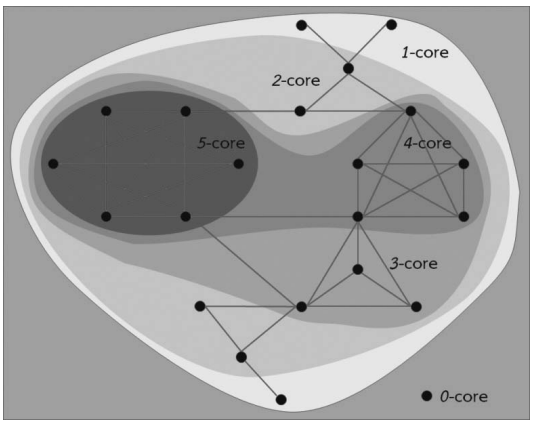
เครือข่ายสังคมเป็นแบบจำลองทางคณิตศาสตร์ที่ใช้ทฤษฎีกราฟ (Graph theory) มาประยุกต์ในการอธิบายความสัมพันธ์ระหว่างผู้ดำเนินการ (Actor) และสารสนเทศที่เกิดขึ้นภายในเครือข่าย ซึ่งภายในกราฟประกอบด้วยเซตของโหนด (Node) หรือ จุด (Vertices) ซึ่งเป็นตัวแทนของผู้ดำเนินการ และเส้น (Line) หรือ ขอบ (Edge) เป็นตัวแทนของความสัมพันธ์ระหว่างผู้ดำเนินการภายในเครือข่าย [9-11]

สำหรับ SNA เป็นการวิเคราะห์เพื่อทำความเข้าใจถึงพฤติกรรมของมนุษย์ในด้านการปฏิสัมพันธ์กับบุคคล ชุมชน หรือ องค์กรที่อยู่ในเครือข่ายสังคม โดยเฉพาะอย่างยิ่งการตรวจสอบรูปแบบความสัมพันธ์ของบุคคลภายในกลุ่ม [9, 10] ดังนั้น SNA จึงถูกนำมาใช้ในการศึกษาวิจัยเป็นจำนวนมาก

SNA เป็นการใช้กราฟในการศึกษาความโครงสร้างของสัมพันธ์ระหว่างสมาชิกภายในชุมชน [5, 12, 13] ซึ่งความสัมพันธ์ดังกล่าวสามารถแบ่งออกได้เป็น 2 ประเภท คือ กลุ่ม (Group) และ เครือข่าย (Network) [14] ทั้งนี้ Group เป็นเซตของบุคคลที่มีลักษณะของความสัมพันธ์แบบใกล้ชิด มีระยะห่างไม่มาก และมีการปฏิสัมพันธ์กันบ่อยครั้ง ในขณะที่บุคคลที่อยู่ภายใน Network จะมีระยะห่างกันมาก มีการปฏิสัมพันธ์กันต่ำกว่า มีความสัมพันธ์ที่หลากหลายลักษณะ และมีความคล่องตัวสูง ทั้งนี้การเชื่อมโยงระหว่างสมาชิกเป็นการแสดงถึงความสัมพันธ์ของ Group หรือ Network ภายในเครือข่ายสังคมซึ่งสามารถมีได้หลายลักษณะ การเชื่อมโยงระหว่างสมาชิกช่วยให้ผู้ใช้งานเข้าใจถึงผลกระทบของโครงสร้างทางสังคมที่มีต่อสมาชิกแต่ละคนตลอดจนสมาชิกทั้งหมดของชุมชน [2] ในการศึกษาทางด้าน SNA นั้นแนวคิดเกี่ยวกับการใช้งาน *k*-core ในการตรวจสอบกลุ่มบุคคลที่มีความสัมพันธ์กันอย่างเหนียวแน่น (Cohesive group) เพื่อตรวจจับ Fraudster มีความเป็นไปได้มากกว่าแนวคิดอื่น เช่น Component, Clique และ *k*-plex [2, 6] เนื่องจากแต่ละ Node ภายในกราฟเครือข่ายสังคมแสดงถึงสมาชิกแต่ละคน และแต่ละ Edge ที่เชื่อมโยงทั้ง 2 Node เข้าด้วยกันแสดงความสัมพันธ์ที่เกิดขึ้นระหว่างบุคคลทั้ง 2 คน ดังนั้นแนวคิดของ k-core สามารถนิยามได้ดังนี้:

กำหนดให้ $G = (V, E)$ หมายถึงกราฟเครือข่ายสังคม โดยที่ V เป็นเซตของ Node และ E เป็นเซตของ Edge กราฟย่อย $H = (W, E/W)$ ที่เกิดจากเซตของ $W \subseteq V$ คือ *k*-core ถ้าดีกรีของ v ไม่น้อยกว่า k สำหรับทุกๆ $v \in V$ และ H เป็นกราฟย่อยสูงสุดด้วยคุณสมบัตินี้

ทั้งนี้ Node อาจจะเป็นองค์ประกอบอยู่ภายในหลายกราฟย่อย *k*-core โดยแต่ละกราฟย่อยดังกล่าวจะมีค่า k ที่แตกต่างกัน คุณสมบัติ *k*-core ของ Node คือค่า k ที่สูงที่สุดของกราฟย่อย *k*-core ทั้งหมดที่ Node ดังกล่าวเป็นสมาชิก [15, 16] สำหรับเครือข่ายที่ประกอบด้วยกราฟย่อย *k*-core ที่มีค่า k ที่แตกต่างกันสามารถแสดงได้ดังภาพที่ 1



ภาพที่ 1 0, 1, 2, 3, 4 และ 5 core

2.2 เอนโทรปีของ Shannon (Shannon entropy)

Shannon entropy เป็นแนวคิดในการใช้การกระจายความน่าจะเป็นควบคู่กับจำนวน

ข้อมูลของเหตุการณ์ในการคำนวณความแม่นยำของสารสนเทศในทฤษฎีสารสนเทศ (Information theory) [8] ที่ได้รับความนิยมนำไปใช้ในการคำนวณความหลากหลายของกลุ่ม เช่น การเรียนรู้ของเครื่องจักร (Machine learning) [17] การจัดการด้านการลงทุน (Portfolio management) [18, 19] นอกจากนี้ยังถูกนำมาใช้ในงานวิจัยฉบับนี้เพื่อการคำนวณ ND ของผู้ใช้งานเว็บไซต์การประมูลออนไลน์แต่ละคน

2.3 นาอีฟเบย์ (Naïve Bayes)

Naïve Bayes เป็นตัวจำแนกทางสถิติ (Statistical classifier) ที่ใช้หลักการของทฤษฎีเบย์ (Bayes' theorem) เป็นพื้นฐาน โดยเป็นการนำหลักการทางสถิติในเรื่องของความน่าจะเป็นมาใช้ในการจำแนกข้อมูล Naïve Bayes มีสามารถในการณ์ความน่าจะเป็นของสมาชิกภายในคลาสได้ เช่น ความน่าจะเป็นที่ว่าของข้อมูลว่าควรจะถูกจำแนกไว้เป็นคลาสใด [20] ทั้งนี้ Naïve Bayes มีสมมติฐานที่ว่าผลกระทบของค่าแอตทริบิวต์ในคลาสที่กำหนดจะมีความเป็นอิสระจากค่าที่ปรากฏภายในแอตทริบิวต์อื่น

2.4 การทำเหมืองข้อมูล (Data Mining and Application)

เนื่องจากการทำเหมืองข้อมูลช่วยให้ผู้ใช้งานสามารถทำการวิเคราะห์และระบุความสัมพันธ์ของข้อมูลได้ง่าย ดังนั้นจึงทำให้การทำเหมืองข้อมูลกลายเป็นเครื่องมือที่ได้รับความนิยมนำไปใช้ในงานด้านการตรวจจับการทุจริตเป็นจำนวนมาก เช่น การตรวจจับการทุจริตบัตรเครดิต (Credit card fraud) [21] การตรวจจับการเสนอราคาหน้าม้าในการประมูลออนไลน์ (Bid shilling) [22] รวมทั้งการทุจริตในการประมูลออนไลน์ (Online auction fraud) [6, 23, 24] โดยทางเทคนิคแล้ว การทำเหมืองข้อมูลเป็นกระบวนการที่ใช้อัลกอริทึมในการเรียนรู้ของเครื่องเพื่อสกัดรูปแบบของข้อมูล หรือ ความสัมพันธ์ระหว่างเซตของแอตทริบิวต์จากชุดข้อมูลที่มีขนาดใหญ่ [25, 26] ซึ่งสามารถแบ่งออกได้เป็นสองประเภทคือ การเรียนรู้แบบมีผู้สอน (Supervised learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning) โดยการเรียนรู้แบบมีผู้สอนหรือการจำแนก (Classification) นั้นการเรียนรู้จะได้มากจากผลลัพธ์ที่เกิดขึ้นจริงของแต่ละตัวอย่างในชุดข้อมูลฝึกฝน โดยมีวัตถุประสงค์หลักเพื่อทำนายคลาส (Class) ของข้อมูลที่ยังไม่ถูกระบุคลาสด้วยความถูกต้องสูงสุด [20, 25]

ในงานวิจัยฉบับนี้ใช้โปรแกรม Weka ในการทำเหมืองข้อมูลโดยเลือกใช้อัลกอริทึม Decision Tree J48 (J48), Neural Network (NN) และ Support Vector Machine (SVM) ซึ่งเป็นอัลกอริทึมสำหรับการจำแนกซึ่งเป็นวิธีการการเรียนรู้แบบมีผู้สอน และเนื่องจากวิธีการ cross-validation เป็นวิธีการคัดเลือกข้อมูลแบบสุ่มเพื่อนำมาใช้เป็นข้อมูลฝึกฝน และข้อมูลทดสอบ ดังนั้นในงานวิจัยฉบับนี้จึงใช้วิธีการ 10-fold cross-validation ในการประมวลผลการทดลอง

2.5 งานวิจัยที่เกี่ยวข้อง

การตรวจจับ Inflated reputation fraud เป็นปัญหาที่ร้ายแรงต่อเว็บไซต์การประมูลออนไลน์และเว็บไซต์ชอปปิ้งออนไลน์อื่นๆ จากงานวิจัยที่ผ่านมาได้มีผู้นำเสนอวิธีการที่ใช้ในการตรวจจับการทุจริตในลักษณะดังกล่าวนี้ไว้หลายวิธี บางวิธีนำเสนอการใช้คุณลักษณะ

ต่างๆ ที่ได้มาจากประวัติการดำเนินธุรกรรมการซื้อ-ขายของผู้ใช้งาน [3, 23] เช่น จำนวนรวม (Summary) ค่าเฉลี่ย (Average) และค่าเบี่ยงเบนมาตรฐาน (Standard deviation) ของราคาขายหรือราคาซื้อของสินค้าในช่วงระยะเวลาหนึ่ง เป็นต้น สำหรับอีกวิธีหนึ่งที่ได้รับความนิยมคือการใช้ SNA ในการตรวจจับกลุ่มผู้ที่กระทำการทุจริตบนเว็บไซต์การประมูลออนไลน์ [2-6, 13, 27-31]

2.5.1 Wang และ Chiu [2] นำเสนอการใช้ k-core และ core/periphery ratio ซึ่งเป็นคุณสมบัติของ SNA ในจำแนกความแตกต่างระหว่าง Fraudster และ Legitimate user ทั้งนี้ core/periphery ratio คือจำนวน Node ที่ปรากฏใน *k*-core หารด้วยจำนวน Node ที่อยู่ติดกับ *k*-core จากผลการศึกษาพบว่า *k*-core และ core/periphery ratio สามารถตรวจจับ Inflated reputation fraud ได้อย่างมีประสิทธิภาพ นอกจากนี้ผลการทดสอบยังแสดงให้เห็นว่า Fraudster มักจะปรากฏใน *k-core* โดยที่ $k \geq 2$

2.5.2 งานวิจัยในภายหลังของ Wang และ Chiu [5] ได้แนะนำให้ใช้ *k-core* ร่วมกับ Center Weight (CW) เนื่องจากการใช้เพียง *k*-core อย่างเดียวส่งผลให้ Precision มีค่าต่ำ ดังนั้น CW จึงถูกนำมาใช้เพื่อทำการปรับปรุงค่า Precision ทั้งนี้ในงานวิจัยฉบับดังกล่าวได้ประยุกต์ใช้อัลกอริทึม Robbery ในการคำนวณค่า CW ของแต่ละ Node โดยในชั้นแรกจะมีการการกำหนดค่าน้ำหนัก (Weight) ของ Node โดยคำนวณจากดีกรีของแต่ละ Node จากนั้น Node ที่มีน้ำหนักมากที่สุดซึ่งถือว่าเป็น Node ที่มีความแข็งแกร่งจะทำการชิงน้ำหนักจาก Node ที่อยู่ติดกันหากมีน้ำหนักน้อยกว่า และจะดำเนินการซ้ำต่อไปเรื่อยๆ ในที่สุด Node ที่มีน้ำหนักมากกว่า 0 จะถือว่าเป็น Node ศูนย์กลาง (Center node) และ Node อื่นๆ ที่อยู่ติดกันจะมีน้ำหนักเป็น 0 ดังนั้น Node ที่มีน้ำหนักมากกว่า 0 จึงเป็นศูนย์กลางของ *k*-core และ Node ดังกล่าวส่วนใหญ่จะถูกจำแนกเป็น Fraudster ผลการทดสอบที่เกิดจากการใช้งาน *k*-core ร่วมกับ CW สามารถปรับปรุงค่า Precision ให้สูงขึ้น อย่างไรก็ตามการใช้งานร่วมกันดังกล่าวส่งผลกระทบต่อการลดลงของค่า Recall

2.5.3 Chiu และ คณะ [6] ได้นำเสนอการคัดแยกบัญชีที่มีความผิดปกติ (Abnormal account) จากข้อมูลการประมูลออนไลน์ที่เกิดขึ้นจริงบนเว็บไซต์ Yahoo! โดยใช้ตัวชี้วัดทางเครือข่ายสังคม (Social network metric) ได้แก่ normalized betweenness, *k*-core, *k*-plex, *n*-cliques, และ degree ในการสร้าง Classifier เพื่อใช้ในการจำแนก Fraudster ออกจาก Legitimate user จากผลการทดสอบพบว่าท่ามกลางบัญชี (Account) ที่มีความผิดปกตินั้นสามารถระบุ Fraudster ได้ด้วยการใช้ normalized betweenness, *k*-core โดยที่ $k \geq 6$, และ *k*-plex ที่ประกอบด้วย 5 Node โดยที่ k = 2

2.5.4 Lin และ Khomnotai [7] ได้นำเสนอแนวคิดเกี่ยวกับ ND เพื่อทำการปรับปรุงค่า Precision และค่า Recall ตามที่ได้กล่าวไปแล้วในเบื้องต้นว่า Fraudster มักจะดำเนินธุรกรรมกับผู้สมรู้ร่วมคิดในการที่จะ Boost up Reputation score ของตน ดังนั้นคุณลักษณะบางประการของผู้สมรู้ร่วมคิดและ Fraudster อาจจะมีความคล้ายกัน และ ND ของเพื่อนบ้านของ Fraudster มีแนวโน้มที่จะมีความหลาก

หลายต่ำ ผลการทดสอบจากแนวคิดนี้แสดงให้เห็นว่า ND ที่คำนวณจากจำนวนระดับคะแนนที่ได้รับจากการทำธุรกรรมสามารถจำแนก Fraudster ออกจาก Legitimate user ได้อย่างมีประสิทธิภาพ ทั้งนี้เพื่อนบ้านของแต่ละ Trader สามารถนิยามได้ดังนี้:

กำหนดให้ x แทนแต่ละ Trader สำหรับเพื่อนบ้านของ x คือ Trader ที่มีการทำธุรกรรมการซื้อ-ขายกับ x และมอบคะแนนประเมินให้แก่ x อย่างน้อย 1 คะแนน

โดยทั่วไปแล้ว Fraudster ที่ต้องการจะเพิ่มระดับของ Reputation score ในระยะเวลาอันรวดเร็วมักจะต้องมีการดำเนินธุรกรรมการซื้อ-ขายในที่มีปริมาณมาก ซึ่งการดำเนินธุรกรรมลักษณะนี้ในระยะเวลาที่จำกัดจำเป็นต้องอาศัยการดำเนินธุรกรรมภายในกลุ่มผู้ทุจริตด้วยกันเอง หลังจากสิ้นสุดการดำเนินธุรกรรมจึงมีการมอบระดับคะแนนที่ดีให้ซึ่งกันและกัน เพราะเหตุนี้จึงมีหลายวิธีที่มีการประยุกต์ใช้ SNA ในการตรวจจับ Inflated reputation fraud โดยใช้การวิเคราะห์เครือข่ายที่สร้างขึ้นจากรายการธุรกรรมการซื้อ-ขายสินค้า เพื่อค้นหากลุ่มของผู้ใช้งานที่มีการติดต่อสัมพันธ์กันอย่างเหนียวแน่น (Cohesive group) ทั้งนี้ในงานวิจัยที่เกี่ยวข้องกับการใช้งาน SNA มักจะนำเอาคุณลักษณะต่างๆ ของ SNA เช่น *k*-plex, clique, betweenness และ k-core มาใช้ในการตรวจสอบ Cohesive group เพื่อตรวจจับ Fraudster ผลการทดสอบจากงานวิจัยดังกล่าวพบว่า *k*-core เป็นคุณลักษณะของ SNA ที่มีประสิทธิภาพมากที่สุดสำหรับการตรวจจับ Fraudster [2, 6]

3. การดำเนินงานวิจัย

3.1 การเก็บรวบรวมข้อมูล

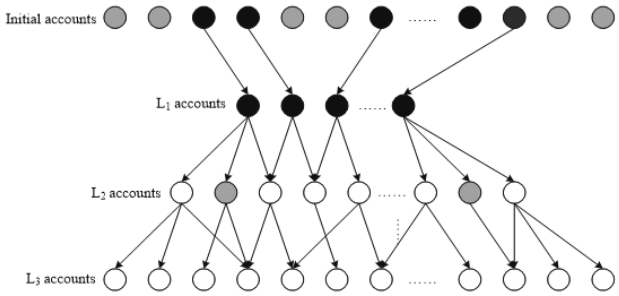
ในการเก็บรวบรวมข้อมูลเพื่อนำมาใช้ในงานวิจัยฉบับนี้ผู้วิจัยได้ทำการพัฒนาโปรแกรมสำหรับเก็บรวบรวมข้อมูลจากเว็บไซต์ (Web crawler) เพื่อเก็บรวบรวมข้อมูลการทำธุรกรรมการซื้อ-ขายที่เกิดขึ้นจริงบนเว็บไซต์ Ruten (www.ruten.com.tw) ซึ่งเป็นหนึ่งในเว็บไซต์การประมูลออนไลน์ที่ใหญ่ที่สุดในประเทศไต้หวัน โดยเป็นการร่วมทุนระหว่าง eBay และ PChome Online [13] เนื่องจาก Ruten จะทำการประกาศรายชื่อ Account ของ Trader ที่ถูกระงับการใช้งานบนเว็บไซต์พร้อมด้วยเหตุผลของถูกการระงับการใช้งานของ Trader ดังกล่าว จากนั้นรายชื่อเหล่านี้จะถูกใช้เป็น Account เริ่มต้นเพื่อใช้ในการรวบรวมข้อมูลเพิ่มขึ้นในลักษณะเดียวกันกับวิธีการรวบรวมข้อมูลที่ถูกนำเสนอในงานวิจัยที่ผ่านมา [5, 6] ทั้งนี้รายละเอียดของการรวบรวมข้อมูลมีดังต่อไปนี้:

รวบรวมข้อมูลรายชื่อ Account ของ Trader ที่ถูกระงับการใช้งานในระหว่างเดือนกรกฎาคม 2013 รวมทั้งสิ้น 9,168 Account ที่ถูกระงับการใช้งาน จากนั้นทำการตรวจสอบเหตุผลการถูกระงับของแต่ละ Account ด้วยมือ (Manually examined) เพื่อคัดเลือกเฉพาะผู้ที่ได้รับคะแนนประเมินอย่างน้อย 1 คะแนนและถูกระงับการใช้งานด้วยเหตุผลที่เกี่ยวข้องกับการทุจริต เช่น การปลอมแปลงการประมูล การมีระดับคะแนนประเมินที่เกินจริง การขายสินค้าปลอม (ละเมิดลิขสิทธิ์) และการที่ผู้ขายได้รับชำระเงินแล้วแต่ไม่จัดส่งสินค้า ภายหลังจากการคัดเลือกมี Account ที่ถูกระงับการใช้งานที่ตรง

ตามเงื่อนไขอยู่ 932 Account ซึ่งจะถูกกำหนดให้เป็น L1 แต่เนื่องจากมี 1 Account ที่ได้รับการทำการคืนสถานะให้กลับเป็นปกติในระหว่างเดือนตุลาคม 2013 ดังนั้น L1 จึงประกอบด้วย Fraudster จำนวน 931 Account และ Legitimate user จำนวน 1 Account

ในขั้นตอนการรวบรวมข้อมูลเพิ่มเติมนั้น L1 จะถูกใช้เป็น Account เริ่มต้นในการรวบรวมข้อมูลเพิ่มเติม เนื่องจากคะแนนประเมินที่ Account ในกลุ่ม L_1 ได้รับนั้นจะทำให้สามารถทราบถึงข้อมูลที่เกี่ยวข้องกับ Trader ผู้ที่มอบคะแนนประเมินให้แก่ Account ในกลุ่ม L_1 ดังกล่าว ภายหลังจากเสร็จสิ้นการรวบรวมข้อมูลเพิ่มเติมพบว่ามี Account ใหม่ที่ได้รับคะแนนประเมินอย่างน้อย 1 คะแนนเพิ่มขึ้นจำนวน 3,475 Account ทั้งนี้ Account ใหม่ดังกล่าวได้ถูกกำหนดให้เป็น L_2 อย่างไรก็ตาม Account ในกลุ่ม L_2 ประกอบด้วยผู้ที่ถูกระงับการใช้งานจำนวน 149 Account และถูกกำหนดให้เป็น Fraudster เนื่องจากผู้ที่ถูกระงับการใช้งานเหล่านี้เป็นผู้ที่มอบคะแนนประเมินบวกให้กับ Account ในกลุ่ม L_1

ในขั้นตอนที่สามของการรวบรวมข้อมูลเพิ่มเติมนี้ Account ในกลุ่ม L_2 จะถูกใช้เป็น Account เริ่มต้นเพื่อทำการรวบรวมข้อมูลเพิ่มเติม โดยการใช้คะแนนประเมินที่ได้รับมาในลักษณะเช่นเดียวกันกับการใช้ Account ในกลุ่ม L_1 และภายหลังจากเสร็จสิ้นกระบวนการรวบรวมข้อมูลในขั้นตอนนี้แล้วพบว่ามี Account ใหม่ที่ได้รับคะแนนประเมินอย่างน้อย 1 คะแนนเพิ่มขึ้นจำนวน 233,169 Account และถูกกำหนดให้เป็น L_3 ทั้งนี้ความสัมพันธ์ระหว่าง Account ที่อยู่ในกลุ่ม L_1, L_2 และ L_3 แสดงได้ดังภาพที่ 2



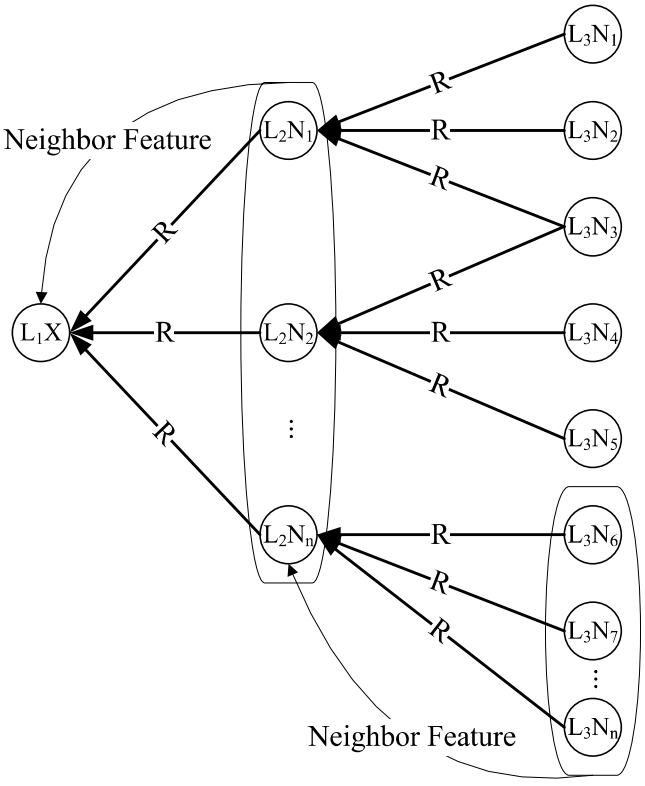
ภาพที่ 2 ลำดับชั้นความสัมพันธ์ของ Account ที่อยู่ในกลุ่ม L_1, L_2 และ L_3

3.2 การสร้างเครือข่ายสังคม

ในงานวิจัยฉบับนี้ใช้วิธีการสร้างเครือข่ายสังคมจากประวัติรายการธุรกรรมซื้อ-ขายของแต่ละ Account เพื่อใช้ในการตรวจจับ Inflated reputation fraud ซึ่งเป็นวิธีการที่มีลักษณะเช่นเดียวกันกับที่ใช้ในงานวิจัยที่ผ่านมา [4-6, 31] จากนั้นทำการคำนวณ *k*-core, CW และ ND ของแต่ละ Account เนื่องจาก *k*-core ถูกพบว่ามีประสิทธิภาพในการตรวจจับ Fraudster [2, 6] และ CW ถูกพบว่าสามารถช่วยเพิ่มค่า Precision ให้สูงขึ้นมากกว่าการใช้เพียง *k*-core [5] ทั้งนี้โปรแกรม Pajek เวอร์ชัน 3.13 [32] ได้ถูกนำมาใช้ในการสร้างกราฟเครือข่ายสังคมรวมทั้งคำนวณค่าคุณลักษณะต่างๆ ของเครือข่ายสังคม ได้แก่ *k*-core และ CW เนื่องจาก Pajek ได้ถูกพัฒนาขึ้นมาเพื่อใช้ในการจัดการเครือข่ายสังคมที่มีขนาดใหญ่ [10, 32] นอกจากนี้ยังเป็นซอฟต์แวร์ที่อนุญาตให้สามารถนำไปใช้ในงานที่ไม่ใช่เชิงพาณิชย์ได้โดยไม่เสียค่าใช้จ่าย

ภายในกราฟเครือข่ายสังคมที่ถูกสร้างขึ้นจากการทำธุรกรรมการซื้อ-ขาย แต่ละ Node แสดงถึงผู้แต่ละ Account ที่ใช้งานเว็บไซต์การประมูลออนไลน์ และแต่ละ Edge ที่เชื่อมโยงทั้ง 2 Node เข้าด้วยกันแสดงถึงรายการธุรกรรมการซื้อ-ขายที่เกิดขึ้นระหว่าง Account ทั้ง 2 โดยที่แต่ละ Account มีการมอบคะแนนให้ซึ่งกันและกันภายหลังจากเสร็จสิ้นการดำเนินธุรกรรมการซื้อ-ขาย และเพื่อลดความซับซ้อนของเครือข่าย ดังนั้นถ้าระดับคะแนนระหว่าง Account ทั้ง 2 มีมากกว่า 1 คะแนน Edge เกิดขึ้นซ้ำระหว่างทั้ง 2 Node จะไม่ถูกเพิ่มเข้าไปในเครือข่าย สำหรับการสร้างเครือข่ายรายการธุรกรรมการซื้อ-ขายภายในงานวิจัยฉบับนี้นั้น Account ทั้งหมดที่อยู่ในกลุ่ม L_1 L_2 และ L_3 จะถูกเพิ่มเข้าไปในเครือข่ายรายการธุรกรรมการซื้อ-ขายดังกล่าว จากนั้นจึงทำการเพิ่ม Edge ซึ่งเป็นระดับคะแนนที่ได้รับมาจาก Account ที่อยู่ในกลุ่ม L_1 และ L_2 เข้าไปในกราฟเครือข่ายดังกล่าว ภายหลังจากเสร็จสิ้นการสร้างเครือข่ายแล้วพบว่าภายในเครือข่ายประกอบด้วย 237,576 (932 + 3475 + 233,169) Node และ 348,259 Edge ทั้งนี้เครือข่ายที่ถูกสร้างขึ้นนี้ไม่ครอบคลุมถึง Edge ของระดับคะแนนที่เกิดขึ้นภายหลังจากวันที่ 31 กรกฎาคม 2013

ข้อมูลที่ถูกรวบรวมข้อมูลโดยการใช้ Web crawler มีความครอบคลุมถึงคุณลักษณะต่างๆ ของเพื่อนบ้าน (Neighbor feature) ได้แก่ จำนวนระดับคะแนนที่ได้รับจากการทำธุรกรรม (The number of received ratings) จำนวนรายการธุรกรรมที่ถูกยกเลิก (The number of cancelled transactions) และวันที่เริ่มต้นใช้งานเว็บไซต์การประมูลออนไลน์ (Joined date) ของแต่ละ Account ดังนั้นภายหลังจากเครือข่ายถูกสร้างขึ้นอย่างสมบูรณ์แล้วเครือข่ายดังกล่าวจะถูกนำไปใช้ในการคำนวณค่า *k*-core, CW และ ND ของแต่ละ Account ที่อยู่ในกลุ่ม L_1 และ L_2 เนื่องจาก Account ที่อยู่ในกลุ่ม L_3 ถูกนำมาใช้ในการสร้างเครือข่ายเพียงเพื่อใช้ในการคำนวณ ND ของ Account ที่อยู่ในกลุ่ม L_1 และ L_2 เท่านั้นแต่ไม่มีความครอบคลุมถึงเพื่อนบ้านของ Account ที่อยู่ในกลุ่ม L_3 ทั้งนี้ความสัมพันธ์ระหว่าง Account และเพื่อนบ้านภายในเครือข่ายสังคมสามารถแสดงได้ดังภาพที่ 3



ภาพที่ 3 ความสัมพันธ์ระหว่าง Account และเพื่อนบ้าน

3.3 ความหลากหลายของเพื่อนบ้าน (Neighbor Diversity)

ในขั้นตอนการคำนวณ ND ของแต่ละ Account แอตทริบิวต์ (Attribute) *attr* จะถูกเลือกเพื่อนำไปใช้ในการคำนวณ จากนั้น Account ทั้งหมดถูกแบ่งออกเป็นระดับชั้นต่างๆ ตาม *attr* ที่ถูกเลือกมาคำนวณ โดยกำหนดให้ x หมายถึงแต่ละ Account เพื่อนบ้านของ *x* คือ Account ที่ทำธุรกรรมการซื้อ-ขายกับ *x* และมอบคะแนนประเมินให้กับ *x* อย่างน้อย 1 คะแนน กำหนดให้ *r* คือจำนวนคะแนนที่แต่ละ Account ได้รับจากการประเมินภายหลังจากเสร็จสิ้นการทำธุรกรรมการซื้อ-ขาย ถ้า $0 \le r$ แล้ว Account จะถูกจัดไว้ในระดับชั้นที่ 1 ถ้า $50 \times 2^{i-2} \le r \le 50 \times 2^{i-1}$ แล้ว Account จะถูกจัดไว้ในระดับชั้นที่ *i* โดยที่ $i > 1$ กำหนดให้ *n* เป็นจำนวนของระดับชั้นและ pi(x) เป็นสัดส่วนของจำนวนเพื่อนบ้านของ *x* ที่ถูกจัดไว้ในระดับชั้นที่ *i*-th ดังนั้นสามารถกำหนดเงื่อนไขเป็นสมการทางคณิตศาสตร์ได้ดังนี้

$$0 \le p_i(x) \le 1, \text{ for i=1 to n} \quad (1)$$

$$\sum_{i=1}^{n} p_i(x) = 1 \quad (2)$$

3.3.1 คำนวณความหลากหลายของเพื่อนบ้านด้วยวิธีเอนโทรปีของ Shannon (Neighbor Diversity Based On Shannon's Entropy) ในการคำนวณ ND ของ *attr* โดยใช้ Shannon's Entropy แทนด้วย ND_S^{attr} ดังนั้นในการคำนวณ ND ของแต่ละแอตทริบิวต์ *attr* สามารถกำหนดเป็นสมการทางคณิตศาสตร์ได้ดังนี้

$$ND_S^{attr} = -\sum_{i=1}^{n} p_i(x) log_2(p_i(x)) \quad (3)$$

3.3.2 คำนวณความหลากหลายของเพื่อนบ้านด้วยวิธี Naïve Bayes (Neighbor Diversity Based On Naïve Bayes) ในการคำนวณ ND ของ *attr* โดยใช้ Naïve Bayes แทนด้วย ND_{NB}^{attr} ดังนั้นในการคำนวณ ND ของแต่ละแอตทริบิวต์ *attr* สามารถกำหนดเป็นสมการทางคณิตศาสตร์ได้ดังนี้

$$ND_{NB}^{attr} = \prod_{i=1}^{n} p_i(x) \quad (4)$$

4. การตั้งค่าการทดสอบและผลการทดสอบ

ชุดข้อมูลที่ถูกในการทำการเปรียบเทียบประสิทธิภาพของ ND_S^{attr} และ ND_{NB}^{attr} ในงานวิจัยฉบับนี้เป็นข้อมูลชุดเดียวกันกับที่ใช้ในงานวิจัยของ Lin และ Khomnotai [7] ซึ่งเป็นข้อมูลการทำธุรกรรมการซื้อ-ขายที่เกิดขึ้นจริงที่ได้ถูกรวบรวมจากเว็บไซต์การประมูลออนไลน์ Ruten (www.ruten.com.tw) ดังรายละเอียดที่ได้อธิบายไว้แล้วในตอนต้น ชุดข้อมูลดังกล่าวนี้ประกอบด้วย 4,407 Account โดยที่ 1,080 Account เป็น Fraudster และ 3,327 Account เป็น Legitimate user

4.1 การตั้งค่าการทดลอง

ภายหลังจากเสร็จสิ้นการรวบรวมข้อมูลแต่ละ Account จะมีค่า จำนวนระดับคะแนนที่ได้รับจากการทำธุรกรรม จำนวนรายการธุรกรรมที่ถูกยกเลิก และวันที่เริ่มต้นใช้งานเว็บไซต์การประมูลออนไลน์ *k*-core, CW, ND_S^{attr} และ ND_{NB}^{attr} เนื่องจาก จำนวนระดับคะแนนที่ได้รับจากการทำธุรกรรม เป็นแอตทริบิวต์ที่มีประสิทธิภาพสูงสุดในการตรวจจับ Fraudster ในงานวิจัยของ Lin และ Khomnotai [7] ดังนั้นผู้วิจัยได้คัดเลือกแอตทริบิวต์ดังกล่าวเป็นแอตทริบิวต์ *attr*

เพื่อนำมาใช้ในการคำนวณ ND_S^{attr} และ ND_{NB}^{attr} ในงานวิจัยฉบับนี้ สำหรับการทดสอบการตรวจจับ Inflated reputation fraud Fraudster ผู้วิจัยเลือกใช้โปรแกรม Weka ซึ่งเป็นซอฟต์แวร์ในการทำเหมืองข้อมูลที่สามารถนำมาใช้งานได้โดยไม่เสียค่าใช้จ่าย [26] และเลือกใช้อัลกอริทึมในการจำแนกจำนวนสามอัลกอริทึมได้แก่ J48 decision tree (J48), Neural Networks (NN), และ Support Vector Machine (SVM) จากนั้นเลือกการประมวลผลเป็นแบบ 10-fold cross-validation ทั้งนี้ผลลัพธ์จากการทดสอบสามารถแสดงได้ดังตาราง 1-ตาราง 6 ทั้งนี้ NDS แสดงถึง ND ของจำนวนระดับคะแนนที่ได้รับจากการทำธุรกรรมที่ถูกคำนวณโดยวิธี Shannon entropy และ ND_{NB} แสดงถึง ND ของจำนวนระดับคะแนนที่ได้รับจากการทำธุรกรรมที่ถูกคำนวณโดยวิธีความน่าจะเป็นของ Naïve Byes

4.2 ผลการทดสอบประสิทธิภาพ

ผลการทดสอบประกอบด้วยสองส่วน ส่วนแรกใช้เพียง ND หรือ Naïve ในการสร้าง Classifier ผลการทดสอบการตรวจจับ Inflated reputation fraud โดยใช้อัลกอริทึม J48, NN และ SVM แสดงได้ดัง ตาราง 1 ตาราง 2 และตาราง 3 ตามลำดับ ทั้งนี้ผลการทดสอบที่ดีที่สุดที่ของแต่ละอัลกอริทึมจะถูกแสดงเป็นตัวหนา

ตารางที่ 1 ประสิทธิภาพของ J48 ในการตรวจจับ Fraudster โดยการใช้ ND_S หรือ ND_{NB}

Input	Accuracy (%)	Recall	Precision	F1-measure
ND_S	84.1843	**0.8019**	0.6420	**0.7131**
ND_{NB}	**84.2523**	0.7880	**0.6467**	0.7104

ตารางที่ 2 ประสิทธิภาพของ Neural Network ในการตรวจจับ Fraudster โดยการใช้ ND_S หรือ ND_{NB}

Input	Accuracy (%)	Recall	Precision	F_1-measure
ND_S	**83.1405**	**0.7620**	**0.6287**	**0.6890**
ND_{NB}	80.1225	0.6611	0.5833	0.6198

ตารางที่ 3 ประสิทธิภาพของ Support Vector Machine ในการตรวจจับ Fraudster โดยการใช้ ND_S หรือ ND_{NB}

Input	Accuracy (%)	Recall	Precision	F_1-measure
ND_S	**83.1405**	**0.7306**	**0.6358**	**0.6799**
ND_{NB}	82.0286	0.6648	0.6254	0.6445

ในการทดสอบส่วนที่สองใช้ทั้ง *k*-core และ CW ร่วมกับ ND_S หรือ ND_{NB} ในการสร้าง Classifier เนื่องจากในงานวิจัยที่ผ่านมาแสดงให้เห็นว่าการใช้งาน *k*-core ร่วมกับ CW ช่วยเพิ่มประสิทธิภาพในการตรวจจับ Fraudster [5] ผลการทดสอบการตรวจจับ Inflated reputation fraud โดยใช้อัลกอริทึม J48, NN และ SVM แสดงได้ดัง ตาราง 4 ตาราง 5 และตาราง 6 ตามลำดับ ทั้งนี้ผลการทดสอบที่ดีที่สุดที่ของแต่ละอัลกอริทึมจะถูกแสดงเป็นตัวหนา

ตารางที่ 4 ประสิทธิภาพของ J48 ในการตรวจจับ Fraudster โดยการใช้ *k*-core และ CW ร่วมกับ ND_S หรือ ND_{NB}

Input	Accuracy (%)	Recall	Precision	F_1-measure
k-core+ CW+ND_S	**85.8180**	**0.8731**	0.6590	**0.7511**
k-core+ CW+ND_{NB}	85.4323	0.8380	**0.6596**	0.7382

ตารางที่ 5 ประสิทธิภาพของ Neural Network ในการตรวจจับ Fraudster โดยการใช้ *k*-core และ CW ร่วมกับ ND_S หรือ ND_{NB}

Input	Accuracy (%)	Recall	Precision	F_1-measure
k-core+ CW+ND_S	**83.7758**	**0.7787**	**0.6386**	**0.7017**
k-core+ CW+ND_{NB}	82.3690	0.6685	0.6328	0.6502

ตารางที่ 6 ประสิทธิภาพของ Support Vector Machine ในการตรวจจับ Fraudster โดยการใช้ k-core และ CW ร่วมกับ ND_S หรือ ND_{NB}

Input	Accuracy (%)	Recall	Precision	F_1-measure
k-core+ CW+ND_S	**84.4112**	**0.7685**	**0.6551**	**0.7073**
k-core+ CW+ND_{NB}	82.9816	0.7407	0.6299	0.6809

5. อภิปรายผลการทดลอง

เมื่อทำการเปรียบเทียบผลการทดสอบในส่วนที่สองกับผลการทดสอบในส่วนที่หนึ่งพบว่าเมื่อมีการทำงานร่วมกับ k-core และ CW ช่วยปรับปรุงประสิทธิภาพให้การจำแนกได้เพียงเล็กน้อยเท่านั้น ทั้งนี้ค่าความแม่นตรงที่เพิ่มขึ้นเกิดจากการใช้งานอัลกอริทึม J48 (อยู่ระหว่าง 1.1800% และ 1.6337%) NN (อยู่ระหว่าง 0.6353% และ 2.2465%) และ SVM (อยู่ระหว่าง 0.6353% และ 0.9530%)

6. สรุปผล

งานวิจัยฉบับนี้ผู้วิจัยได้ประยุกต์แนวคิดเกี่ยวกับความหลากหลายของเพื่อนบ้านของผู้ใช้งานแต่ละคน เพื่อตรวจจับ Inflated reputation fraud เว็บไซต์ประมูลออนไลน์ ทั้งนี้ผู้วิจัยเลือกใช้ Shannon entropy และความน่าจะเป็นของ Naïve Byes ในการคำนวณความหลากหลายของเพื่อนบ้าน เพื่อเป็นการศึกษาถึงประสิทธิภาพในการตรวจจับ Fraudster ของทั้งสองวิธี จากผลการทดสอบประสิทธิภาพในการจำแนกพบว่าการคำนวณความหลากหลายของเพื่อนบ้านด้วยวิธี Shannon entropy มีประสิทธิภาพในการจำแนกได้ดีกว่าการคำนวณความหลากหลายของเพื่อนบ้านด้วยวิธีความน่าจะเป็นของ Naïve Byes เพียงเล็กน้อย ซึ่งแสดงให้เห็นว่าวิธีการคำนวณความหลากหลายของเพื่อนบ้านทั้งสองวิธีมีประสิทธิภาพในการจำแนก Fraudster ที่ใกล้เคียงกัน

7. กิตติกรรมประกาศ

ในการดำเนินงานวิจัยฉบับนี้ได้รับการสนับสนุนเงินทุนวิจัยจาก The National Science Council of Taiwan (R.O.C.) ภายใต้ Grant 102-2221-E-155-034-MY3 ขอขอบคุณ Dr. Jun-Lin Lin Associate Professor in Department of Information Management, Yuan Ze University, Taiwan (R.O.C.) ที่ได้กรุณาให้คำแนะนำในการดำเนินงานวิจัยจนงานวิจัยนี้สำเร็จลุล่วง

8. บรรณานุกรม

(1) D. Houser and J. Wooders, "Reputation in Auctions: Theory, and Evidence from eBay," Journal of Economics & Management Strategy, vol. 15, pp. 353-369, 2006.

(2) J. C. Wang and C. Q. Chiu, "Detecting Online Auction Inflated-Reputation Behaviors Using Social Network Analysis," in Annual Conference of The North American Association for Computational Social and Organizational Science (Online Publication, http://www.casos.cs.cmu.edu/events/conferences/2005/2005_proceedings/Wang.pdf). Notre Dame, Indiana, USA, 2005.

(3) D. H. Chau, S. Pandit, and C. Faloutsos, "Detecting Fraudulent Personalities in Networks of Online auctioneers," in Proceedings of the 10th European conference on Principle and Practice of Knowledge Discovery in Databases: PKDD 2006, Berlin, Germany, 2006, pp. 103-114.

(4) S. Pandit, D. H. Chau, S. Wang, and C. Faloutsos, "Netprobe: A Fast and Scalable System for Fraud Detection in Online Auction Networks," in Proceedings of the 16th international conference on World Wide Web, Banff, Alberta, Canada, 2007, pp. pp.201-210.

(5) J. C. Wang and C. C. Chiu, "Recommending Trusted Online Auction Sellers Using Social Network Analysis," Expert Systems with Applications, vol. 34, pp. 1666-1679, 2008.

(6) C. Chiu, Y. Ku, T. Lie, and Y. Chen, "Internet Auction Fraud Detection Using Social Network Analysis and Classification Tree Approaches," International Journal of Electronic Commerce, vol. 15, pp. 123-147, 2011.

(7) J.-L. Lin and L. Khomnotai, "Using Neighbor Diversity to Detect Fraudsters in Online Auctions," Entropy, vol. 16, pp. 2629-2641, 2014.

(8) C. E. Shannon, "A Mathematical Theory of Communication," The Bell System Technical Journal, vol. 27, pp. 379-423, 1948.

(9) L. C. Freeman, The Development of Social Network Analysis: A Study in the Sociology of Science. Canada: Empirical Press Vancouver, 2004.

(10) W. d. Nooy, M. Andrej, and V. Batagelj, Explore Social Network Analysis with Pajek. New York: Cambridge University Press, 2005.

(11) M. Zhang, "Social Network Analysis: History, Concepts, and Research," in Handbook of Social Network Technologies and Applications, B. Furht, Ed., ed: Springer US, 2010, pp. 3-21.

(12) S. P. Borgatti. (1998, 28/07/2013). What is Social Network Analysis? Available: *http://www.analytictech.com/networks/whatis.htm*

(13) C. H. Yu and S. J. Lin, "Fuzzy Rule Optimization for Online Auction Frauds Detection Based on Genetic Algorithm," Electronic Commerce Research, vol. 13, pp. 169-182, 2013/05/01 2013.

(14) L. Garton, C. Haythornthwaite, and B. Wellman, "Studying Online Social Networks," Journal of Computer-Mediated Communication, vol. 3, 1997.

(15) V. Batagelj and M. ZaveZšnik, "An O(m) Algorithm for Cores Decomposition of Networks," arXiv preprint cs/0310049, pp. 1-10, 2003.

(16) V. Batagelj and M. ZaverŠnik, "Fast Algorithms for Determining (Generalized) Core Groups in Social Networks," Advances in Data Analysis and Classification, vol. 5, pp. 129-145, July 2011.

(17) J. R. Quinlan, C4.5: Programs for Machine Learning: Morgan Kaufmann Publishers Inc., 1993.

(18) I. Usta and Y. M. Kantar, "Mean-Variance-Skewness-Entropy Measures: A Multi-Objective Approach for Portfolio Selection," Entropy, vol. 13, pp. 117-133, 2013/10/28 2011.

(19) J.-L. Lin, "On the Diversity Constraints for Portfolio Optimization," Entropy, vol. 15, pp. 4607-4621, October 2013.

(20) J. Han, M. Kamber, and J. Pei, Data Mining : Concepts and Techniques, 3rd ed ed. USA: Morgan Kaufmann, 2012.

(21) Y. Sahin and E. Duman, "Detecting Credit Card Fraud by Decision Trees and Support Vector Machines," in The International MultiConference of Engineers and Computer Scientists 2011, Hong Kong, 2011, pp. pp.442-447.

(22) H. S. Shah, N. R. Joshi, A. Sureka, and P. R. Wurman, "Mining eBay: Bidding Strategies and Shill Detection," in *WEBKDD 2002 - Mining Web Data for Discovering Usage Patterns and Profiles*. vol. 2703, O. Za©ane, J. Srivastava, M. Spiliopoulou, and B. Masand, Eds., ed: Springer Berlin Heidelberg, 2003, pp. 17-34.

(23) D. H. Chau and C. Faloutsos, "Fraud Detection in Electronic Auction," presented at the European Web Mining Forum, held as part of ECML/PKDD, Porto, Portugal, 2005.

(24) W.-H. Chang and J.-S. Chang, "An Effective Early Fraud Detection Method for Online auctions," Electronic Commerce Research and Applications, vol. 11, pp. 346-360, July–August 2012.

(25) F. Gorunescu, Data Mining: Concepts, Models and Techniques: Springer, 2011.

(26) I. H. Witten, F. Eibe, and M. A. Hall, Data Mining : Practical Machine Learning Tools and Techniques, 3rd ed. USA: Morgan Kaufmann, 2011.

(27) M. Morzy, "New Algorithms for Mining the Reputation of Participants of Online Auctions," in *Internet and Network Economics*. vol. 3828, X. Deng and Y. Ye, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 112-121.

(28) M. Morzy, "Cluster-Based Analysis and Recommendation of Sellers in Online Auction," Computer Systems Science and Engineering, vol. 22, pp. 279-287, 2007.

(29) Z. Bin, Z. Yi, and C. Faloutsos, "Toward a Comprehensive Model in Internet Auction Fraud Detection," in Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, 2008, pp. 79-79.

(30) C. H. Yu and S.-J. Lin, "Web Crawling and Filtering for On-line Auctions from a Social Network Perspective," Information Systems and e-Business Management, vol. 10, pp. 201-218, 2012/06/01 2012.

(31) S.-J. Lin, Y.-Y. Jheng, and C.-H. Yu, "Combining Ranking Concept and Social Network Analysis to Detect Collusive Groups in Online Auctions," Expert Systems with Applications, vol. 39, pp. 9079-9086, 2012.

(32) V. Batagelj and A. Mrvar. (2013, Pajek and Pajek XXL – Program for analysis and visualization of very large networks: reference manual (version 3.13).