

การจำลองเครือข่ายซับซ้อนและการทำนายผลวิดีโอออนไลน์ ด้วยเทคนิคเหมืองข้อมูล

Represents to Complex Networks and Prediction of VDO Online with Data Mining

นราธิป ทองปาน และ สุริยวรรณ แจ่มจิตร

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏมหาสารคาม และ สาขาวิชาเทคโนโลยีสารสนเทศ
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏกาญจนบุรี

บทคัดย่อ

ในการใช้งานผ่านระบบอินเทอร์เน็ต มีรูปแบบการให้บริการที่หลากหลายช่องทาง ในการติดต่อสื่อสาร ข้อมูลถึงกัน โดยใช้สื่อสังคมออนไลน์ที่เข้าถึงผู้คนทุกระดับ ทำให้เกิดเครือข่ายทางสังคม (social Network) ในลักษณะของการแบ่งปัน เช่น รูปภาพ วิดีโอออนไลน์ผ่านทางเฟสบุ๊ค (facebook) ทวิตเตอร์ (twitter) และยูทูบ (youtube) ซึ่งเป็นสื่อสังคมที่นิยมในปัจจุบัน ในงานวิจัยนี้ผู้วิจัยมีวัตถุประสงค์เพื่อ 1) แสดงความสัมพันธ์ของสื่อ สังคมนำเสนอในรูปแบบเครือข่ายซับซ้อน (complex network) 2) นำเสนอจัดกลุ่มผู้ใช้สื่อสังคม (social Media) โดยใช้เทคนิคเหมืองข้อมูลแบบ SOM และ 3) หาเทคนิคต้นไม้ตัดสินใจที่ให้ค่าการพยากรณ์ในระดับการศึกษา ถูกต้องสูงสุด แล้วนำผลที่ได้ไปแปลงเป็นกฎการตัดสินใจเพื่อประยุกต์ใช้ในการทำตลาดเฉพาะกลุ่ม ข้อมูลที่ใช้ในการ วิเคราะห์ได้แก่ ข้อมูลของผู้ใช้สื่อวิดีโอออนไลน์จาก pewinternet.org จำนวน 1,003 คน เครื่องมือที่ใช้ในการ วิเคราะห์ข้อมูล ได้แก่ โปรแกรม Weka, Orange Canvas และ Pajek ผลการวิจัยพบว่า ความสัมพันธ์ของผู้ใช้สื่อ สังคมด้านวิดีโอออนไลน์ เป็นรูปแบบเครือข่ายซับซ้อนมีลักษณะการเชื่อมโยงแบบ Small World การจัดกลุ่มผู้ใช้ สื่อสังคม โดยใช้เทคนิคเหมืองข้อมูลแบบ SOM แสดงให้เห็นว่าผู้ใช้ที่มีระดับการศึกษาสูงและมีการจ้างงานแบบเต็มเวลา จะมีการดูวิดีโอออนไลน์ผ่านทางสังคมออนไลน์มากที่สุด และต้นไม้ตัดสินใจเทคนิคแบบ ID3 มีค่าความถูกต้องในการ พยากรณ์ระดับการศึกษาสูงสุดเท่ากับ 98.7179 % และทำการแปลงต้นไม้ตัดสินใจ ID3 จะได้กฎจำนวน 12 กฎ

คำสำคัญ: วิดีโอออนไลน์, เครือข่ายสังคม, ทฤษฎีกราฟ, เหมืองข้อมูล

ABSTRACT

In order to work through the internet, there are a variety of service delivery models to communicate the information to each other by using social media to reach people in all levels and leading to social networking such as sharing photos and watching video online through the Facebook, Twitter and YouTube which are popular nowadays. Therefore, in this paper, the users information from social media related to online video with the aim to 1) represent to complex network using Pajek program, 2) classify social media using SOM technique, and 3) find out a decision tree techniques to classify the high accuracy and using decision tree to generate a rule of decision applied to the niche market. The data set from website of pewinternet.org consisted of 1,003 people. The instruments of the research were Weka, Orange Canvas and Pajek. The findings found that the relation between social media users presents complex networks in the form of a "small world". A cluster of users with high level education was employed full time and watching video online through social media the most. The decision trees with ID3 technique contributed to the highest accuracy at 98.7179 % and 12 rule base of ID3 technique.

Keywords: Video Online, Social Network, Graph Theory, Data Mining

บทนำ

ปัจจุบันการใช้งานผ่านระบบเครือข่ายอินเทอร์เน็ต เป็นที่นิยมใช้กันอย่างแพร่หลาย มีรูปแบบการให้บริการที่หลากหลาย สามารถใช้งานได้ง่ายและสะดวกรวดเร็ว การใช้สื่อสังคมออนไลน์ ในการสื่อสารข้อมูลถึงกัน สื่อสังคมออนไลน์ เป็นรูปแบบการสื่อสารข้อมูลที่เข้าถึงผู้ใช้งานได้ทุกระดับ ทำให้เกิดเครือข่ายทางสังคม (social network) ในลักษณะของการแบ่งปัน เช่น รูปภาพ วิดีโอออนไลน์ผ่านทางเฟสบุ๊ค (facebook) ทวิตเตอร์ (twitter) และ ยูทูบ (youtube) เครือข่ายสังคม (social network) สามารถอธิบายลักษณะความสัมพันธ์ทางสังคมระหว่างบุคคลและปฏิสัมพันธ์ระหว่างสมาชิกเครือข่าย โดยใช้ทฤษฎีกราฟ[1] แทนความสัมพันธ์ทางสังคม เพื่อเป็นตัวแทนของข้อมูล ซึ่งทฤษฎีกราฟสามารถทำให้มองเห็นในสิ่งที่มีความซับซ้อนมากขึ้น แต่ในการแก้ปัญหาบางประเภทจะมีลักษณะที่ซับซ้อนและจำเพาะ จึงต้องเลือกวิธีที่เหมาะสม เพื่ออธิบายคุณสมบัติต่าง ๆ ของเครือข่ายและสมาชิก เป็นต้น

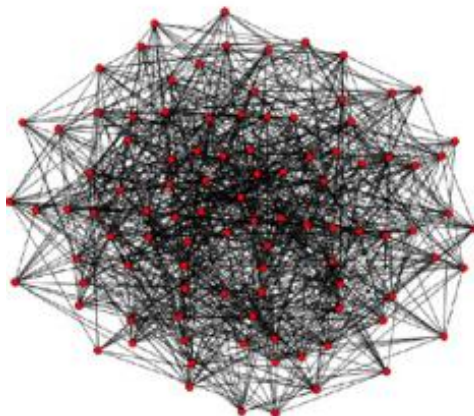
อย่างไรก็ตาม ในการอธิบายคุณสมบัติของเครือข่ายสังคม ในงานวิจัยนี้ ผู้วิจัยได้ใช้เทคนิคเหมืองข้อมูล [2] เป็นกระบวนการเพื่อค้นหารูปแบบที่เหมาะสมและความสัมพันธ์จากข้อมูลที่มีอยู่จำนวนมากมายที่ซ่อนอยู่ในชุดข้อมูลของผู้ใช้สื่อวิดีโอออนไลน์[3] ในงานวิจัยนี้ผู้วิจัยได้นำเสนอ 1) เพื่อหาความสัมพันธ์ของสื่อสังคมนำเสนอในรูปแบบเครือข่ายซับซ้อน 2) เพื่อนำเสนอจัดกลุ่มผู้ใช้สื่อสังคม โดยใช้เทคนิคเหมืองข้อมูลแบบ SOM และ 3) หาเทคนิคค้นไม่ตัดสินใจที่ให้ค่าการพยากรณ์ในระดับการศึกษาถูกต้องสูงสุด แล้วนำผลที่ได้ไปแปลงเป็นกฎการตัดสินใจเพื่อประยุกต์ใช้ในประชาสัมพันธ์ หรือ การตลาดเฉพาะกลุ่ม

วรรณกรรมที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาทฤษฎี และงานวิจัยต่างๆ ที่เกี่ยวข้อง เพื่อใช้ในการศึกษาวิธีการ ในการดำเนินงานวิจัย โดยมีงานวิจัยที่เกี่ยวข้องดังนี้

1. เครือข่ายทางสังคม

เครือข่ายทางสังคม หมายถึง กลุ่มของสมาชิกเครือข่าย (Nodes) ที่มีความเกี่ยวข้องเชื่อมโยงกัน โดยมีความสัมพันธ์ และปฏิสัมพันธ์ทางสังคมในรูปแบบต่าง ๆ เช่น ความเป็นเพื่อน การพูดคุย การติดต่อสื่อสาร และสื่อสังคมออนไลน์

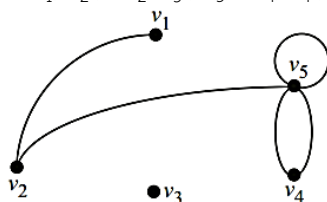


ภาพที่ 1 เครือข่ายซับซ้อน [4]

สื่อทางสังคม (social media) เป็นเครื่องมือหรือรูปแบบจากเว็บ 2.0 ที่การถ่ายทอดจากการเขียนและการส่งผ่านข้อมูลผ่านทางเว็บไซต์ เช่น รูปภาพ ข้อความ วิดีโอ และ เสียง สามารถนำมาใช้ในหลากหลายรูปแบบผ่านทางเครื่องมือบริการสื่อสังคมออนไลน์ (social network service : SNS) เช่น เฟสบุ๊ก (facebook) เนื่องจาก [5] ได้นำเสนอการตรวจสอบโครงสร้างของชุมชนขนาดใหญ่และการวิเคราะห์ข้อมูลของเฟสบุ๊ก ใน [6] จึงได้นำเสนอชุดข้อมูล (data set) เครือข่ายสังคมแบบใหม่โดยใช้เฟสบุ๊ก, ทวิตเตอร์ (twitter) โดย [7] ได้ใช้วิธีการดึงข้อมูล (crawler) เพื่อวิเคราะห์ข้อมูลของทวิตเตอร์ของหัวข้อที่มีแนวโน้มสูงสุดและรายงานเกี่ยวกับพฤติกรรมชั่วขณะของพวกผู้ใช้และการมีส่วนร่วมของผู้ใช้ และ ยูทูบ (youtube) ใน [8] ได้นำเสนอและการศึกษาเกี่ยวกับการวัดสถิติในการใช้งานของวิดีโอ YouTube เป็นต้น

2. ทฤษฎีกราฟ

การใช้ทฤษฎีกราฟ [1] แทนความสัมพันธ์ทางสังคม เพื่อเป็นตัวแทนของข้อมูล ซึ่งทฤษฎีกราฟสามารถทำให้มองเห็นในสิ่งที่มีรูปแบบที่หลากหลายมากขึ้น ซึ่งกราฟ (graph) [9] ประกอบด้วยจุดยอด (vertices) และเส้นเชื่อม (edge) เขียนสัญลักษณ์ได้ $G = (V, E)$ โดยที่ V เป็นเซตของจุดยอด และ E เป็นเซตของเส้นเชื่อม โดยมี $V = [v_1, v_2, v_3, v_4, v_5]$ และ $E = [(v_1, v_2), (v_2, v_5), v_3, (v_4, v_4), (v_5, v_5), (v_5, v_4)]$



ภาพที่ 2 แสดงส่วนประกอบของกราฟ [1]

3. โปรแกรม Pajek

เป็นโปรแกรมใช้เพื่อวิเคราะห์เครือข่ายทางสังคม (SNA: Social Network Analysis) [10] ในการแสดงภาพการเชื่อมต่อของเครือข่ายขนาดใหญ่ (large network) ที่มีหลายพันส่วนหรือแม้กระทั่งนับล้านจุด ในกลุ่มต่าง ๆ ในเครือข่าย เพื่ออธิบายลักษณะความสัมพันธ์ทางสังคมระหว่างบุคคลและปฏิสัมพันธ์ระหว่างสมาชิกเครือข่าย [4] ซึ่งมี [11] นำเสนอเครือข่ายที่สามารถนำเข้าได้อย่างง่ายดายในการทำ Social Action จากความหลากหลายของรูปแบบที่นิยม เช่น Pajek และ GraphML

4. เทคนิคเหมืองข้อมูล

เป็นกระบวนการเพื่อค้นหารูปแบบที่เหมาะสมและความสัมพันธ์จากข้อมูลที่มีอยู่จำนวนมากมายที่ซ่อนอยู่ในชุดข้อมูลโดยอาศัยหลักสถิติ การรู้จำ และการเรียนรู้ของเครื่อง (machines learning) ซึ่งเทคนิคที่ใช้ในการทำเหมืองข้อมูล [2] แบ่งออกเป็น 2 กลุ่ม ดังนี้ 1. การจัดกลุ่ม (clustering) เป็นกระบวนการจัดกลุ่มข้อมูลที่มีคุณสมบัติคล้ายกัน หรือ ต่างกันมากที่สุด เป็นรูปแบบการเรียนรู้ข้อมูลโดยไม่ต้องอาศัยชุดฝึกสอน (unsupervised learning) ได้แก่ SOM (self-organizing maps) ใน [12] มีได้ศึกษาถึงคุณสมบัติการจัดกลุ่มให้เหมาะสมกับเงื่อนไขที่ต้องการ ควรคำนึงคุณสมบัติที่ส่งผลถึงประสิทธิภาพและประสิทธิภาพในการจัดกลุ่ม เช่น ข้อมูลขนาดใหญ่ รูปร่างของกลุ่มข้อมูล มิติสูง ซับซ้อน ที่ และ 2. การจำแนก (classification) เป็นกระบวนการสร้างโมเดลด้วยการจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ (class) โดยอาศัยชุดฝึกสอนนำมาช่วยในการเรียนรู้ข้อมูล เพื่อไปทำนายข้อมูลในชุดแบบทดสอบ (test data) ที่มีการระบุกลุ่มข้อมูล ได้แก่ ต้นไม้ตัดสินใจ (decision tree) มีลักษณะคล้ายโครงสร้างต้นไม้ โดยมีโหนดแสดงคุณลักษณะ (attribute) กิ่งต้นไม้แสดงถึงเงื่อนไขในการทดสอบ และโหนดใบ (leaf node) แสดงถึงผลลัพธ์ของข้อมูล ซึ่งใน [13] ทำวิจัยเกี่ยวกับ web-base Decision Support System

(DSS) ใช้ C4.5 เพราะเป็นวิธีที่ได้ถูกในการตัดสินใจตามที่ต้องการใช้งาน และเป็นวิธีที่แก้ไขได้ง่าย นอกจากนี้ [14] มีการวิเคราะห์รูปแบบการบุกรุกบนเครือข่าย เปรียบเทียบกับ 3 เทคนิค ได้แก่ เทคนิค J48 หรือ C4.5, เทคนิค JRip Rule และเทคนิค Neural Network ผลการพยากรณ์พบว่า เทคนิค J48 มีค่าความถูกต้องสูงสุด เท่ากับ 99.998% และ ใน [15] แสดงการเปรียบเทียบประสิทธิภาพการจำแนกแบบ Naïve Bayesian, Instance-Based Learner, NBTree, Id3SimpleCART, SMO และ J48 พบว่าวิธีการของ NBTree ให้ผลการจำแนกที่ดีที่สุดคือ 90.20%

จากงานวิจัยได้ศึกษา พบว่าเทคนิคต้นไม้ตัดสินใจที่ให้ประสิทธิภาพสูงในการจำแนกข้อมูลสูงและนิยมใช้ คือ เทคนิค C4.5, ID3 และ NBTree ดังนั้นในงานวิจัยนี้ได้เลือกใช้ 3 เทคนิคข้างต้นมาเปรียบเทียบและเลือกเทคนิคที่ให้ค่าความถูกต้อง (Accuracy) สูงสุด มีค่าความคลาดเคลื่อนระหว่างค่าจริงกับค่าที่ใช้พยากรณ์ต่ำ (Root Mean Squared Error: RMSE) เนื่องจากค่าพยากรณ์มีค่าเข้าใกล้ค่าจริง

5. โปรแกรม Orange canvas

เป็นโปรแกรมด้านการทำเหมืองข้อมูล [16] โดยสร้างแบบจำลอง ใช้งานง่าย และเป็นโปรแกรมฟรี (open source) รองรับทุกแพลตฟอร์มในงานวิจัยนี้ใช้โปรแกรม Orange canvas รุ่น 2.7 เพื่อใช้หา SOM (self-organizing maps) เนื่องจากมีผลลัพธ์แสดงในรูปแบบกราฟิกที่สวยงาม

6. โปรแกรมเวก้า (Weka)

การทำงานจะคล้ายโปรแกรม Orange canvas แต่โปรแกรมนี้มีจุดด้อยในการแสดงผลลัพธ์ ด้านกราฟิกไม่สวย แต่เป็นที่นิยมใช้อย่างแพร่หลาย เนื่องจากมีคำสั่งให้เลือกใช้อย่างครบถ้วน และสามารถเขียนฟังก์ชันเพิ่มในโปรแกรมได้ โดยในงานวิจัยนี้ใช้โปรแกรมเวก้า รุ่น 3.6.9

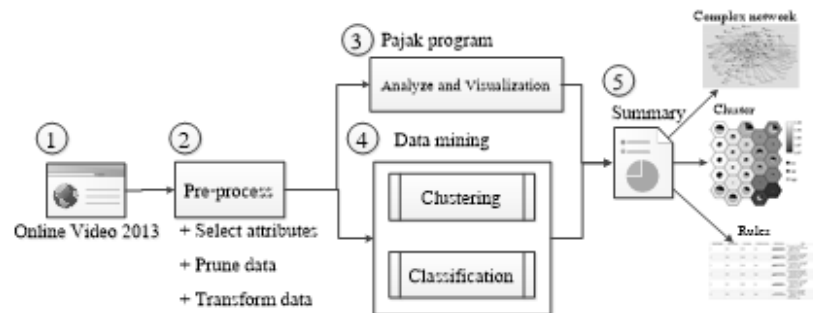
7. ประยุกต์ใช้การทำตลาดเฉพาะกลุ่มลูกค้า

ผู้วิจัยได้ศึกษาข้อมูลจาก Pew Internet & American Life Project [17] เผยแพร่ข้อมูลผลการสำรวจกิจกรรมที่ผู้ใช้ในสื่อออนไลน์ ในเดือนกรกฎาคม 2013 พบว่ากลุ่มผู้ใช้ชาวอเมริกันมีกิจกรรมอัปโหลด (upload) และ โพสต์วิดีโอ (post videos) เพิ่มขึ้นเป็นสองเท่าในระยะเวลา 4 ปี จาก 14% เป็น 31% ด้วยความนิยมที่เพิ่มขึ้น และยังมีการศึกษากิจกรรมการแบ่งปันวิดีโอ [8, 18] ซึ่งเห็นถึงความสัมพันธ์ของวิดีโอมีลักษณะเครือข่ายแบบ small world network และ ชี้ให้เห็นถึงความสัมพันธ์ใกล้เคียงกัน สอดคล้องกันระหว่างวิดีโอที่มีมาก (Strong correlation) ซึ่งทำให้ผู้วิจัยสนใจในการศึกษาถึงการเครือข่ายของวิดีโอในสังคมออนไลน์ ในการทำตลาดเฉพาะกลุ่มลูกค้า [19] (Niche Market) เป็นกำหนดการทำธุรกิจที่มีกลุ่มเป้าหมายที่เฉพาะเจาะจง เพื่อกลุ่มลูกค้าที่เล็กลงไป มีการแข่งขันน้อยกว่าตลาดใหญ่ๆ ที่มุ่งเน้นคนจำนวนมาก (mass market) ดังนั้นในการทำตลาดเฉพาะกลุ่มลูกค้า มีคุณลักษณะที่เฉพาะเจาะจงของสินค้าและบริการรวมทั้งการที่ธุรกิจสามารถมุ่งเน้นตอบสนองความต้องการของตลาดส่วนน้อยนี้ได้เต็มที่ โดยลูกค้าที่มีรสนิยมและความต้องการคล้ายคลึงกัน เพื่อนำไปประยุกต์ใช้ในการให้บริการหรือจำหน่ายสินค้าในตลาดเฉพาะกลุ่มต่อไป

การดำเนินการวิจัย

ในงานวิจัยนี้ผู้วิจัยได้ดำเนินการดังนี้

ขั้นที่ 1 ผู้วิจัยได้ทำการศึกษาข้อมูลจาก Online Video 2013 [17] เกี่ยวกับผู้ใช้อินเทอร์เน็ตในการทำกิจกรรมวิดีโอออนไลน์ เช่น การดูวิดีโอออนไลน์ โดยข้อมูลที่ศึกษามาจากการสัมภาษณ์ทางโทรศัพท์จากกลุ่มตัวแทนของผู้ใหญ่ชาวอเมริกันจำนวน 1,003 คน ที่อาศัยในสหรัฐอเมริกา โดย Princeton Survey Research Associates International (PSRAI) เป็นผู้เก็บข้อมูลระหว่าง 25-28 กรกฎาคม 2013 แสดงขั้นตอนการดำเนินงานดังนี้



ภาพที่ 3 แสดงขั้นตอนการดำเนินงาน

ขั้นที่ 2 กระบวนการประมวลผล ในขั้นแรกในการคัดเลือกแอททริบิวต์ที่ใช้ในการวิจัย ผู้วิจัยได้ทำการตัดข้อมูลที่ใช้ไม่ได้ออก (Prune Data) ซึ่งทำให้ได้ข้อมูล ประกอบด้วยแอททริบิวต์จำนวน 6 แอททริบิวต์ จากนั้นดำเนินการแปลงข้อมูล (Transform data) อยู่ในรูปแบบไฟล์โปรแกรม Pajek และโปรแกรม Orange สำหรับนำไปประมวลผล ดังนี้

- 1) ช่องทางการเข้าถึงวิดีโอ ได้แก่ ผ่านทางสังคมออนไลน์โดยใช้คอมพิวเตอร์ (use_FB) เช่น เฟสบุ๊ก และผ่านโปรแกรมประยุกต์ของโทรศัพท์มือถือ (use_MB) เช่น Vine,
- 2) อายุ (age) ประกอบด้วย 4 ช่วงดังนี้ 18-29 ปี (1), 30-49 ปี (2), 50-64 ปี (3) และ 65 ปีขึ้นไป (4)
- 3) พื้นที่ผู้ใช้สังคมออนไลน์ (region) ประกอบด้วย 4 พื้นที่ดังนี้ NORTH-EAST (1), MID-WEST (2), SOUTH (3), WEST (4)
- 4) เพศ (sex) ประกอบด้วยเพศชาย (1) และเพศหญิง (2)
- 5) การศึกษา (edu) ประกอบด้วย 3 ระดับ คือ ระดับต่ำ (low) หมายถึงจบการศึกษาต่ำกว่ามัธยมศึกษาหรือเทียบเท่า ระดับกลาง (mid) หมายถึงระดับอนุปริญญาหรือเทียบเท่า และระดับสูง (high) หมายถึงระดับปริญญาตรีขึ้นไป
- 6) ประเภทวิดีโอ (VOD01) และการจ้างงาน (employ) ประกอบด้วย 3 ระดับคือการจ้างแบบเต็มเวลา (Employed full-time) การจ้างบางส่วน (Employed part-time) และไม่มีงานทำ (Not employed)

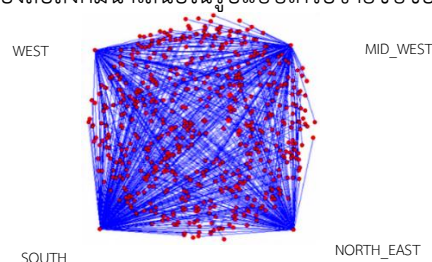
ขั้นที่ 3 ใช้โปรแกรม Pajek เพื่อวิเคราะห์ลักษณะความสัมพันธ์ทางสังคมระหว่างกิจกรรมของผู้ใช้สื่อสังคมออนไลน์ แสดงในรูปแบบกราฟ

ขั้นที่ 4 ใช้โปรแกรม Orange เพื่อทำการจัดกลุ่ม และด้วยใช้โปรแกรม weka เพื่อจำแนกข้อมูล โดยเลือกชุดทดสอบเป็นชุดฝึกสอน (Use training set)

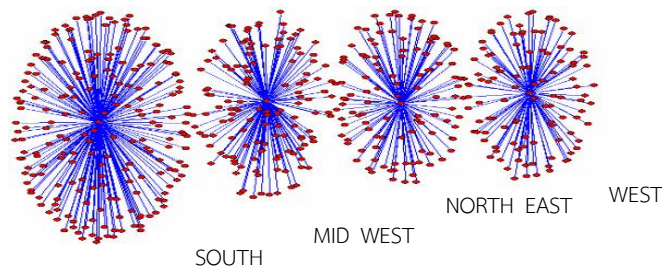
ผลการวิจัย

ในการวิเคราะห์ข้อมูล ผู้วิจัยได้แบ่งเป็น 3 ส่วน มีรายละเอียดดังนี้

ส่วนที่ 1 แสดงความสัมพันธ์ของสื่อสังคมนำเสนอในรูปแบบเครือข่ายซับซ้อน (Complex network)



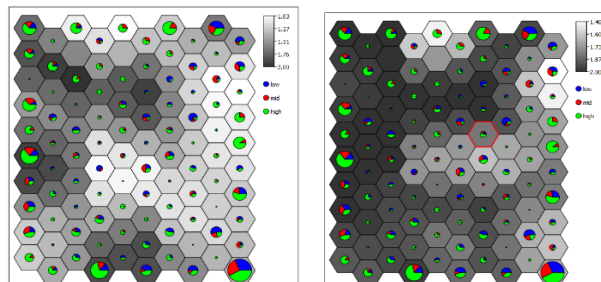
ภาพที่ 4 (ก) แบ่งตามชนิดวิดีโอ ระดับการศึกษา ช่องทางการดูวิดีโอ การจ้างงาน และเพศ



ภาพที่ 4 (ข) แบ่งตามพื้นที่เป็นกลุ่ม (cluster)

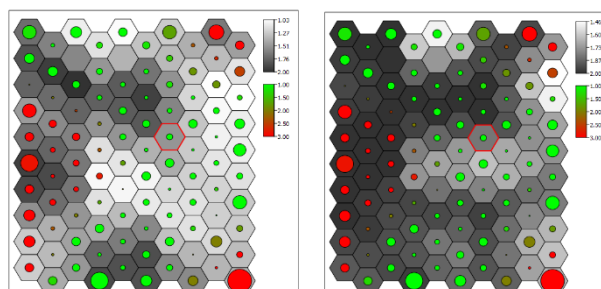
จากภาพที่ 4 (ก) เห็นได้ว่าเครือข่ายซับซ้อน ทำให้เห็นการเชื่อมโยงระหว่างกลุ่มข้อมูลได้ชัดเจนและเป็นที่น่าสนใจ จากภาพที่ 4 (ข) แสดงเครือข่ายซับซ้อนแบ่งตามพื้นที่ของผู้ใช้โดยกลุ่มที่ดูวิดีโอออนไลน์มากที่สุดคือพื้นที่ SOUTH

ส่วนที่ 2 ความสัมพันธ์ระหว่างช่องทางการดูวิดีโอ – ระดับการศึกษา แสดงจัดกลุ่มโดยใช้เทคนิคเหมือนข้อมูลแบบ SOM ดังภาพที่ 5

(ก) ผ่านทางสังคมออนไลน์
(Computer PC)(ข) ผ่านทางโปรแกรมประยุกต์
(Mobile Application)

ภาพที่ 5 ความสัมพันธ์ระหว่างช่องทางการดูวิดีโอ – ระดับการศึกษา

จากภาพที่ 5 พบว่าผู้ใช้ที่มีการศึกษาระดับสูงเป็นกลุ่มที่มากที่สุด ในการดูวิดีโอผ่านทางสังคมออนไลน์ มากกว่าผ่านทางโปรแกรมประยุกต์ของโทรศัพท์มือถือ

(ก) ผ่านทางสังคมออนไลน์
(Computer PC)(ข) ผ่านทางโปรแกรมประยุกต์
(Mobile Application)

ภาพที่ 6 ความสัมพันธ์ระหว่างช่องทางการดูวิดีโอ – การจ้างงาน

จากภาพที่ 6 พบว่า พบว่าผู้ใช้ที่มีการจ้างงานแบบเต็มเวลาเป็นกลุ่มที่มีมากที่สุด โดยส่วนใหญ่ดูวิดีโอออนไลน์ผ่านทางสังคมออนไลน์มากกว่าผ่านทางโปรแกรมประยุกต์ของโทรศัพท์มือถือ สรุปได้จากภาพที่ 5 และภาพที่ 6 พบว่าผู้ใช้ที่มีระดับการศึกษาสูงและมีการจ้างงานแบบเต็มเวลาเป็นกลุ่มที่มีมากที่สุด โดยดูวิดีโอออนไลน์ผ่านทางสังคมออนไลน์ เช่น เฟสบุ๊ก

ส่วนที่ 3 แสดงเทคนิคต้นไม้ตัดสินใจที่ให้ค่าการพยากรณ์ในระดับการศึกษาถูกต้องตารางที่ 1

ตารางที่ 1 เปรียบเทียบผลลัพธ์เทคนิคต้นไม้ตัดสินใจ

Decision tree	C4.5 EBP	C4.5 REP	ID3	NBTree
Accuracy	70.8791%	67.7656%	98.7179%	79.304%
RMSE	0.3748	32.2344	0.0684	0.3066
Build model (seconds)	0.07	0.01	0.01	3.25

จากตารางที่ 1 พบว่าเทคนิค ID3 ให้ค่าความถูกต้องสูงสุด เท่ากับ 98.7179 % และมีค่า RMSE น้อยที่สุดเท่ากับ 0.0684

สรุป

ในงานวิจัยนี้ได้ผลการวิจัย พบว่า 1) ความสัมพันธ์ของผู้ใช้วิดีโอออนไลน์ เป็นรูปแบบเครือข่ายซับซ้อน มีลักษณะการเชื่อมโยงแบบ Small World อธิบายได้ว่า กลุ่มผู้ใช้งานที่อยู่ในพื้นที่เดียวกันหรือเขตพื้นที่ใกล้เคียง จะมีโอกาสสนใจกิจกรรมสื่อสังคมออนไลน์เหมือนกันมากกว่าคนที่อยู่พื้นที่ห่างไกลกัน หรือ นอกเขตกัน 2) การจัดกลุ่มความสัมพันธ์ระหว่างช่องทางการดูวิดีโอ – ระดับการศึกษา และความสัมพันธ์ระหว่างช่องทางการดูวิดีโอ – การจ้างงาน โดยใช้แบบ SOM พบว่า กลุ่มผู้ใช้ที่มีระดับการศึกษาสูง และมีการจ้างงานแบบเต็มเวลาเป็นกลุ่มที่มีมากที่สุด โดยช่องทางการดูวิดีโอผ่านทางสังคมออนไลน์ 3) เทคนิคต้นไม้ตัดสินใจ ID3 ให้ค่าการทำนายระดับการศึกษาถูกต้องสูงสุดเท่ากับ 98.7179 % และทำการแปลงต้นไม้ตัดสินใจ ID3 จะได้กฎจำนวน 12 กฎ ที่ได้เป็นกฎการทำนายระดับการศึกษา เพื่อนำไปประยุกต์เครือข่ายของวิดีโอในสังคมออนไลน์ ในการตลาดเฉพาะกลุ่มลูกค้า (Niche Market)

ขอขอบคุณ

แหล่งข้อมูลรายการผลการของกลุ่มผู้ใช้งานสื่อสังคมออนไลน์ จาก Pew Internet & American Life Project [17] ที่อนุเคราะห์ข้อมูลในงานวิจัยนี้

เอกสารอ้างอิง

- [1] Christensen, C and Albert, R. (2007). "Using graph concepts to understand the organization of complex systems," International Journal of Bifurcation and Chaos, vol. 17, pp. 2201-2214.
- [2] Jiawei Han, M. K. (2006). "Data Mining: Concepts and Technique," ed: San Francisco, Morgan Kaufmann
- [3] Purcell, K. Online Video Available: <http://pewinternet.org/Reports/2013/Online-Video.aspx>, 2013.
- [4] Hoppe, B. and Reinelt, C. "Social network analysis and the evaluation of leadership networks," The Leadership Quarterly, vol. 21, pp. 600-619,8//2010.
- [5] Ferrara, E. "A large-scale community structure analysis in Facebook," EPJ Data Science, vol. 1, pp. 1-30, 2012.
- [6] Lewis, K. Kaufman, J. Gonzalez, M. Wimmer, A. and Christakis. N. "Tastes, ties, and time: A new social network dataset using Facebook.com," Social Networks, vol. 30, pp. 330-342, 10//2008.
- [7] Kwak, H. Lee, C. Park, H. and Moon. S. "What is Twitter, a social network or a news media?," presented at the Proceedings of the 19th international conference on World wide web, Raleigh, North Carolina, USA, 2010.
- [8] Xu, C. Dale, C. and Jiangchuan, L. "Statistics and Social Network of YouTube Videos," in Quality of Service, 2008. IWQoS 2008. 16th International Workshop on, 2008, pp. 229-238.
- [9] Ruohonen, K. GRAPH THEORY, 2013.
- [10] Batagelj, V. M. Andrej. "Pajek " Version 2.05, 2011.
- [11] Perer, A. and Shneiderman, B. "Balancing Systematic and Flexible Exploration of Social Networks," Visualization and Computer Graphics, IEEE Transactions on, vol. 12, pp. 693-700, 2006.
- [12] Rui, X. and Wunsch, D. II, "Survey of clustering algorithms," Neural Networks, IEEE Transactions on, vol. 16, pp. 645-678, 2005.
- [13] Govindaraju, R. and Mahardhika, D. "Web-based Decision Support System using C4.5 decision tree algorithm," in Advanced Computer Science and Information System (ICACSIS), International Conference on, 2011, pp. 395-400, 2011.
- [14] จิโรจน์ ภาคศิริ, "การวิเคราะห์รูปแบบการบุกรุกข้อมูลบนเครือข่าย โดยใช้เทคนิคดาต้าไมนิง. (2550). " วารสารเทคโนโลยีสารสนเทศ. ปีที่ 3 ฉบับที่ 6 (ก.ค.- ธ.ค. 2550), vol. 3, 40-46, 2550
- [15] Cufoglu, A. Lohi, M. and Madani. K. "A comparative study of selected classifiers with classification accuracy in user profiling," in Computer Science and Information Engineering, 2009 WRI World Congress on, 2009, pp. 708-712.
- [16] Demšar, J. Zupan, B. Leban, G. and Curk. T. "Orange: From Experimental Machine Learning to Interactive Data Mining," in Knowledge Discovery in Databases: PKDD 2004. vol. 3202, J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 537-539.
- [17] Purcell, K. Online Video 2013. Available: <http://pewinternet.org/Reports/2013/Online-Video.aspx>
- [18] Xu, C. Haitao, L. and Jiangchuan, L. "Video sharing propagation in social networks: Measurement, modeling, and analysis," in INFOCOM, 2013 Proceedings IEEE, 2013, pp. 45-49.
- [19] online. Niche Market. Available: <http://www.doothaithai.com/read.php?tid=159.html>, 2014