การประยุกต์แนวคิดผู้เชี่ยวชาญเพื่อการแนะนำสินค้า

Applying Expert Concept for Recommending Items

ธนพล พุกเส็ง* และสุนันฑา สดสี

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าเจ้าพระนครเหนือ

ถนนประชาราษฎร์ 1 แขวงวงศ์สว่าง เขตบางซื่อ กรุงเทพมหานคร 10800

Thanaphon Phukseng* and Sunantha Sodsee

Faculty of Information Technology, King Mongkut's University of Technology North Bangkok,

Pracharat 1 Road, Wongsawang, Bangsue, Bangkok 10800

บทคัดย่อ

งานวิจัยนี้นำเสนอการประยุกต์ใช้แนวคิดผู้เชี่ยวชาญเพื่อสร้างการแนะนำสินค้า ซึ่งจะช่วยให้ผู้ใช้งานได้รับสินค้าที่มีความเหมาะสมและสอดคล้องต่อความต้องการ โดยพิจารณาความเป็นผู้เชี่ยวชาญจากปัจจัย 3 ด้าน ได้แก่ (1) ด้านการเชื่อมโยงในเครือข่าย โดยพิจารณาจากจำนวนการเชื่อมโยงที่เข้าสู่ผู้ใช้งานในเครือข่าย ซึ่งเป็นการแสดงถึงชื่อเสียงของผู้ใช้งานนั้น (2) ด้านความถี่ในการให้คะแนนสินค้า โดยพิจารณาจากการนับจำนวนครั้งในการให้คะแนนสินค้าของผู้ใช้งาน ซึ่งจำนวนครั้งในการให้คะแนนนั้นแสดงออกถึงความเป็นผู้เชี่ยวชาญ และ (3) ด้านความนิยมของผู้ใช้งานอื่นในระบบ ซึ่งเป็นการคิดค่าเฉลี่ยของการให้คะแนนผู้ใช้งานจากผู้ใช้งานอื่นในระบบ โดยผู้ใช้งานที่ได้รับคะแนนผู้ใช้งานสูงก็แสดงถึงมีความเชี่ยวชาญสูง และเพื่อเปรียบเทียบประสิทธิภาพในการทำนายการแนะนำของปัจจัยทั้ง 3 ด้าน งานวิจัยนี้จึงได้ใช้ชุดข้อมูล Epinion และทดลองเปรียบเทียบประสิทธิภาพความถูกต้อง (accuracy) และความผิดพลาด (error) ของการทำนาย ซึ่งผลการทดลองแสดงให้เห็นว่าการพิจารณาความเป็นผู้เชี่ยวชาญเมื่อคำนึงถึงความถี่ในการให้คะแนนสินค้าให้ผลลัพธ์ที่เหมาะสมสูงที่สุด โดยมีค่าความถูกต้อง (accuracy) ที่ร้อยละ 85.404 และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute error) เท่ากับ 0.146

คำสำคัญ : ระบบผู้แนะนำ; ผู้เชี่ยวชาญ; การเชื่อมโยงในเครือข่าย; ความถี่ในการให้คะแนน; ความนิยมของผู้ใช้งานอื่นในระบบ

Abstract

This research presents the applying expert concept consideration for recommending items. The proposed work aims to recommend optimal items to users, which are based on user preferences. Herein, this work applies 3 major expert identifications as following, (1) network topology: Considering a user network topology by concerning indegee of users, which is acting as user reputation level, (2) considering a frequency of item ratings, which is acting as a user-item

***ผู้รับผิดชอบบทความ :** thanaphon134@yahoo.com

expertise level, and (3) considering a user like by averaging like rating from other users, which is acting as user-like rating expertise level. For evaluation, the Epinion dataset is employed to test the proposed work focusing on accuracy and error of recommending. The results shows that the expert identification by considering frequency of item ratings applied in the recommendation process presents the highest percentage of accuracy and the lowest mean absolute error, which are 85.404 % and 0.146, respectively.

Keywords: recommender system; expert; social network; frequency of Item rating; user like

1. บทนำ

การเติบโตของเทคโนโลยีสารสนเทศในปัจจุบันทำให้ข้อมูลข่าวสารของสินค้าและบริการต่าง ๆ มีปริมาณเพิ่มขึ้นเป็นจำนวนมาก ที่เรียกว่า data overload ซึ่งส่งผลต่อพฤติกรรมการซื้อสินค้าของลูกค้าที่มีการค้นหาเปรียบเทียบสินค้าที่ตรงตามต้องการก่อนตัดสินใจซื้อสินค้า กล่าวคือ ลูกค้าจะต้องพิจารณาข้อมูลสินค้าที่มีอยู่เป็นจำนวนมากกว่าที่จะได้สินค้าตรงตามความต้องการ ในขณะเดียวกันผู้ขายก็ต้องการเสนอสินค้าที่เหมาะสมต่อความต้องการของลูกค้า ดังนั้นระบบผู้แนะนำจึงเกิดขึ้น โดยหมายถึงระบบที่นำเสนอสินค้าหรือบริการที่มีความเหมาะสมตรงกับความต้องการของลูกค้า [1] ซึ่งระบบผู้แนะนำนั้นมีหลากหลายเทคนิค เช่น content-base filtering เป็นการพิจารณาความคล้ายคลึงในเชิงเนื้อหา หรือ collaborative filtering ที่เป็นการพิจารณาในเชิงความสัมพันธ์ โดยอาศัยข้อมูลการให้คะแนนสินค้า (user-items matrix) เป็นต้น แต่ในการซื้อสินค้าหรือบริการของบุคคลทั่วไป มักจะรับฟังความคิดเห็นหรือข้อแนะนำจากบุคคล 2 กลุ่ม [2] ได้แก่ เพื่อน ซึ่งมีพฤติกรรม ความชื่นชอบที่คล้ายคลึงกัน และผู้เชี่ยวชาญซึ่งเป็นผู้ที่มีความชำนาญ หรือเชี่ยวชาญในด้านใดด้านหนึ่ง โดยเฉพาะถ้าหากลูกค้าไม่มีความรู้ในสินค้าเหล่านั้น การแนะนำจากผู้เชี่ยวชาญจะมีความสำคัญเป็นอย่างมาก ดังนั้นงานวิจัยนี้จึงมีวัตถุประสงค์เพื่อนำเสนอรูปแบบในการนำแนวคิดของผู้เชี่ยวชาญมาใช้ในการสร้างการแนะนำ ซึ่งการพิจารณาความเป็นผู้เชี่ยวชาญนั้นมีหลายลักษณะ [3,4] แต่สำหรับงานวิจัยนี้ที่ใช้ข้อมูลการให้คะแนนสินค้า ข้อมูลการให้คะแนนผู้ใช้งาน ร่วมกับข้อมูลเครือข่าย จะนำเสนอการพิจารณาความเป็นผู้เชี่ยวชาญจากปัจจัย 3 ด้าน ได้แก่ การเชื่อมโยงในเครือข่าย ความถี่ในการให้คะแนนสินค้า และความนิยมของผู้ใช้งานอื่นในระบบ และทดลองเปรียบเทียบเกี่ยวกับความผิดพลาด และความถูกต้องในการทำนาย เพื่อแสดงให้เห็นถึงประสิทธิภาพในการทำงานของปัจจัยแต่ละด้าน

สำหรับโครงสร้างของการนำเสนอในบทความประกอบด้วย ส่วนที่สอง ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ส่วนที่สามแนวคิดงานวิจัยที่นำเสนอ ส่วนต่อไป คือ การทดลองและผลการวิจัย และส่วนสุดท้ายสรุปผลการวิจัย

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในส่วนนี้จะนำเสนอทฤษฎีและงานวิจัยที่เกี่ยวข้องกับระบบผู้แนะนำรวมถึงการนำแนวคิดผู้เชี่ยวชาญมาใช้ในการสร้างการแนะนำ

2.1 ระบบผู้แนะนำ

ระบบผู้แนะนำ หมายถึง ซอฟต์แวร์ที่ช่วยสนับสนุนผู้ใช้งานในการระบุเลือกสินค้าที่ตรงกับความสนใจมากที่สุด [2] โดยทั่วไปแล้วสามารถแบ่งออกเป็น

3 เทคนิค ได้แก่ collaborative filtering, content-based filtering และ hybrid filtering [5]

collaborative filtering เป็นการพิจารณาความสัมพันธ์ระหว่างพฤติกรรมของผู้ใช้งาน เช่น ถ้าหากผู้ใช้งาน 2 คน มีความสนใจในเรื่องเดียวกันแล้ว ก็มีโอกาสที่จะเลือกชมภาพยนตร์ที่มีความคล้ายกัน [6] สำหรับข้อมูลที่ใช้โดยส่วนใหญ่จะอยู่ในรูปแบบของการให้คะแนนสินค้า สำหรับวิธีการนั้นสามารถแบ่งออกเป็น 2 กลุ่ม คือ model based เป็นการรวบรวมการให้คะแนนแล้วใช้โมเดลเพื่อให้เกิดการเรียนรู้ แล้วสร้างการทำนายผลคะแนน [2] และ memory based จะเป็นการระบุสิ่งที่สนใจจากการพิจารณาความคล้ายคลึงระหว่างผู้ใช้งาน โดยจะหาผู้ที่มีความคล้ายคลึงกันจากการให้คะแนน ซึ่งวิธีการ memory based จะมีความเหมาะสมมากกว่าในกรณีที่มีการเพิ่มสินค้าใหม่ [7]

content-based filtering เป็นการพิจารณาจากเนื้อความ หรือลักษณะของสิ่งเหล่านั้น โดยถ้าหากผู้ใช้งานมีความชื่นชอบในสิ่งใดสิ่งหนึ่งในอดีต ก็จะแนะนำสินค้าที่มีลักษณะคล้ายคลึงกับสิ่งนั้น [5] ซึ่งในรูปแบบนี้เองจะมีข้อจำกัดที่การกำหนดลักษณะของสินค้า ได้แก่ สินค้าสองชนิดแต่ต้องเก็บข้อมูลในลักษณะเดียวกัน หรือการกำหนดคุณลักษณะให้สัมพันธ์กับสินค้า เป็นต้น

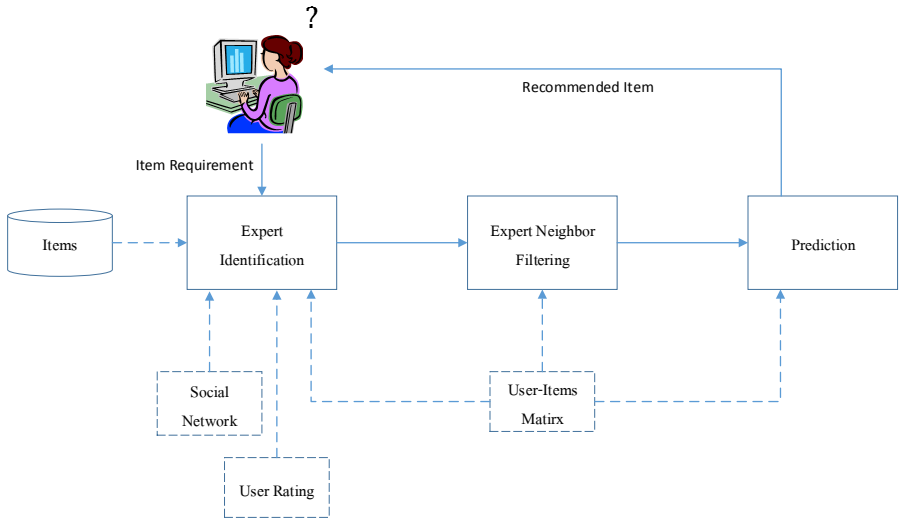
hybrid filtering เป็นการผสมผสานแนวทาง collaborative filtering เข้ากับวิธีการอื่นรวมถึง content-based filtering เพื่อให้ได้ผลลัพธ์ที่แตกต่างออกไป [1]

2.2 ผู้เชี่ยวชาญ

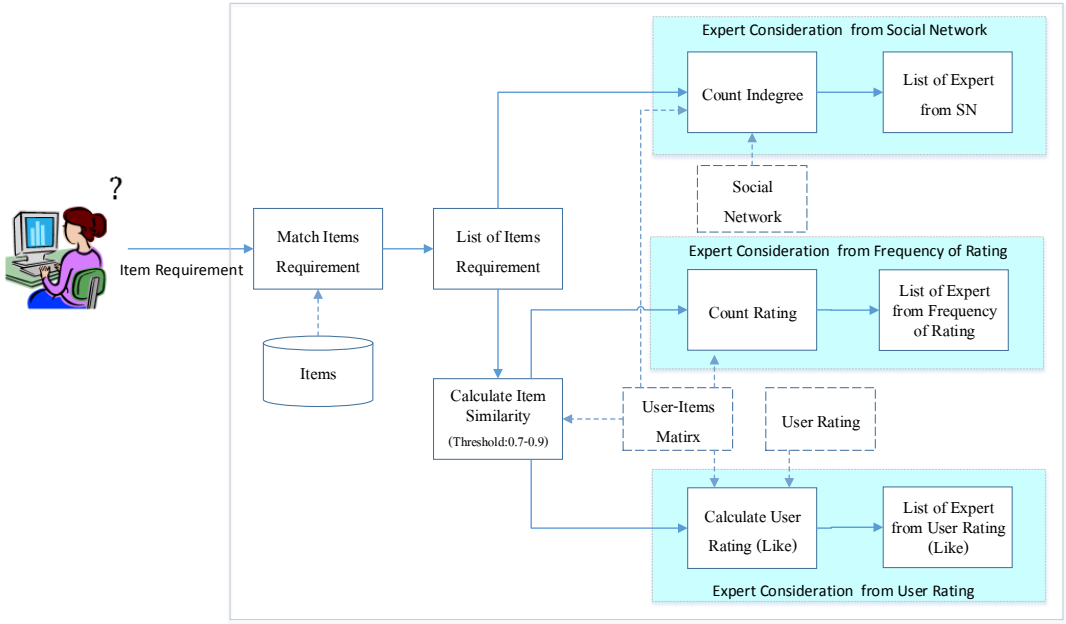
ระบบผู้แนะนำโดยทั่วไปจะให้ความสำคัญกับข้อมูลการให้คะแนนสินค้ามากกว่าความสัมพันธ์ระหว่างบุคคล แต่ในความเป็นจริงการเลือกซื้อสินค้าใด ๆ นั้น มักจะรับฟังความคิดเห็นจากบุคคล 2 กลุ่ม [2] คือ เพื่อน หมายถึง บุคคลที่มีรสนิยม ความชื่นชอบคล้ายคลึงกัน โดยเพื่อนมักมีอิทธิพลต่อการตัดสินใจของบุคคลต่าง ๆ และผู้เชี่ยวชาญ หมายถึง บุคคลที่มีความรู้ความชำนาญในด้านใดด้านหนึ่งที่จะส่งผลต่อการเลือกซื้อสินค้าของบุคคล [8] โดยเฉพาะถ้าหากเป็นสินค้าใหม่ คำแนะนำจากผู้เชี่ยวชาญจะมีความ สำคัญมากกว่าเพื่อนที่อาจไม่เคยได้ใช้สินค้าดังกล่าว ดังนั้นแนวคิดในการนำผู้เชี่ยวชาญมาใช้ในการสร้างการแนะนำจึงมีความน่าสนใจ ซึ่งงานวิจัยที่นำแนวคิดผู้เชี่ยวชาญมาใช้นั้นสามารถแบ่งเป็น 2 ลักษณะ คือ ในลักษณะแรกจะใช้เพื่อพิจารณาคุณสมบัติการเป็นผู้เชี่ยวชาญ [3,4,9,10] และอีกลักษณะคือการประยุกต์ใช้ในการพัฒนาระบบผู้แนะนำ [11-13] ซึ่งส่วนใหญ่จะใช้กับ content-based filtering มิได้นำมาใช้กับ collaborative filtering สำหรับงานวิจัยนี้ให้ความสนใจในการนำแนวคิดผู้เชี่ยวชาญมาใช้งานกับข้อมูลการให้คะแนนสินค้าในรูปแบบ collaborative filtering โดยจะพิจารณาความเป็นผู้เชี่ยวชาญจากปัจจัย 3 ด้าน คือ

2.2.1 การเชื่อมโยงในเครือข่าย เป็นการพิจารณาถึงชื่อเสียงของบุคคล (reputation) [1] โดยเมื่อมองการเชื่อมโยงการติดต่อในเครือข่ายด้วยทฤษฎีกราฟแล้ว บุคคลในเครือข่ายก็เปรียบเทียบได้กับโหนด (node) และการเชื่อมโยงติดต่อกันเทียบได้กับเส้นเชื่อม (link) ซึ่งจะใช้ค่าจำนวนการเชื่อมโยงที่เข้าสู่โหนด (indegree) เพื่อแทนชื่อเสียงของโหนด นั้น ๆ ได้ เนื่องจากบุคคลที่มีชื่อเสียงจะได้รับการติดตามจากบุคคลอื่น

2.2.2 ความถี่ในการให้คะแนน เป็นการแสดงออกถึงความเชี่ยวชาญของแต่ละบุคคล โดยผู้ใช้งานที่ให้ข้อมูลเข้าสู่ระบบบ่อยครั้งก็จะเป็นผู้ที่มีความเชี่ยวชาญสูงกว่า



รูปที่ 1 ภาพรวมกระบวนการทำงานของระบบผู้แนะนำจากประยุกต์ใช้แนวคิดผู้เชี่ยวชาญ



รูปที่ 2 กระบวนการระบุผู้ใช้งานที่เชี่ยวชาญจากข้อมูลเครือข่าย ความถี่ในการให้คะแนนและความนิยมของผู้ใช้งานอื่นในระบบ

2.2.3 ความนิยมของผู้ใช้งานอื่นในระบบ พิจารณาจากการได้รับคะแนนผู้ใช้งาน (user rating) จากบุคคลอื่นๆที่อยู่ในระบบ โดยบุคคลที่ได้รับคะแนนผู้ใช้งานสูงก็แสดงถึงมีความเชี่ยวชาญสูงเช่นเดียวกัน

3. แนวคิดงานวิจัยที่นำเสนอ

ส่วนนี้จะได้นำเสนอรูปแบบในการนำแนวคิดผู้เชี่ยวชาญมาใช้ในการสร้างการแนะนำ โดยประกอบด้วย 3 กระบวนการหลัก ได้แก่ การระบุผู้ใช้งานที่

เชี่ยวชาญ การคัดกรองผู้เชี่ยวชาญที่ใช้ในการสร้างการแนะนำ และการสร้างการแนะนำ สำหรับภาพรวมของกระบวนการสามารถแสดงได้ดังรูปที่ 1

3.1 การระบุผู้ใช้งานที่เชี่ยวชาญ (expert identification)

กระบวนการนี้เริ่มต้นจากการที่ได้ทราบรายการสินค้าที่สอดคล้องกับความต้องการของผู้ใช้งาน (match items requirement) จากนั้นจะเข้าสู่ขั้นตอนในการระบุผู้ใช้งานที่เชี่ยวชาญ โดยจะได้แบ่งการพิจารณาออกเป็น 2 ส่วน คือ ส่วนการพิจารณาจากข้อมูลเครือข่าย และส่วนการพิจารณาจากความถี่ในการให้คะแนนและความนิยมของผู้ใช้งานอื่นในระบบ โดยสามารถแสดงภาพรวมการทำงานได้ดังรูปที่ 2

3.1.1 ส่วนการพิจารณาผู้ใช้งานที่เชี่ยวชาญจากข้อมูลเครือข่าย (expert consideration from social network)

การทำงานในส่วนนี้จำเป็นที่จะต้องอาศัยข้อมูลการเชื่อมโยงในเครือข่ายผู้ใช้งาน โดยใช้ทฤษฎีกราฟเป็นแนวทางในการวิเคราะห์เครือข่าย [14] ซึ่งกราฟเครือข่ายผู้ใช้งาน G ประกอบด้วย $G = (V, E)$ ในที่นี้ V คือ เซตของผู้ใช้งาน และ E คือ เส้นแสดงการเชื่อมโยงผู้ใช้งาน และใช้หลักการดีกรี (degree) ในการวิเคราะห์ข้อมูล โดยกราฟที่ได้นั้นจะมีลักษณะเป็นกราฟแบบมีทิศทาง (directed graph) ซึ่งในงานวิจัยนี้ให้ความสนใจกับค่าจำนวนการเชื่อมโยงที่เข้าสู่ผู้ใช้งาน โดยแสดงเป็นสมการได้ดังนี้ [15]

$$C_d^-(a) = degree^-(a)$$

โดยที่ $degree^-(a) = |\{b \in V | (b, a) \in E\}|$ เป็นจำนวนของผู้ใช้งาน b ที่เชื่อมโยงเข้าสู่ผู้ใช้งาน a สำหรับการที่พิจารณาเฉพาะค่าจำนวนการเชื่อมโยงที่เข้าสู่ผู้ใช้งานนั้น เนื่องจากจำนวนของผู้ใช้งานที่พุ่งเข้านั้นสะท้อนถึงชื่อเสียงของบุคคลนั้น ซึ่งสื่อให้เห็นถึงการได้รับการมองให้เป็นผู้เชี่ยวชาญในเครือข่าย โดยกระบวนการนี้จะทำให้ได้รายการผู้ใช้งานที่มีความเชี่ยวชาญจากข้อมูลเครือข่าย

3.1.2 ส่วนการพิจารณาผู้ใช้งานที่เชี่ยวชาญจากความถี่ในการให้คะแนนและความนิยมของผู้ใช้งานอื่นในระบบ

การทำงานในส่วนนี้จะเกี่ยวข้องกับข้อมูลการให้คะแนนสินค้า โดยสามารถแบ่งการทำงานเป็น 3 ขั้นตอนย่อยดังต่อไปนี้

(1) คำนวณค่าความคล้ายคลึงของสินค้า (calculate item similarity) การให้คะแนนสินค้าของผู้ใช้งานนั้นมีการให้คะแนนที่หลากหลายสะท้อนถึงพฤติกรรมของบุคคลผู้นั้น โดยถ้าหากพิจารณาจากการให้คะแนนสินค้าทั้งหมดของผู้ใช้งานก็จะทำให้มีข้อมูลที่ได้จากสินค้าที่ไม่มีความเกี่ยวข้องกับสินค้าเป้าหมาย ดังนั้นในขั้นตอนแรกจึงได้มีการคำนวณค่าความคล้ายคลึงของสินค้าเป้าหมายกับสินค้าอื่น โดยอาศัยสมการ item-based adjusted cosine [16] เนื่องจากต้องการพิจารณาความสัมพันธ์ระหว่างสินค้า ดังนี้

$$IAdjCos_{xy} = \frac{\sum_{i=1}^{N_{xy}} (r_{ax,i} - \bar{r}_a) \times (r_{ay,i} - \bar{r}_a)}{\sqrt{\sum_{i=1}^{N_{xy}} (r_{ax,i} - \bar{r}_a)^2 \times \sum_{i=1}^{N_{xy}} (r_{ay,i} - \bar{r}_a)^2}}$$

โดยที่ $IAdjCos_{xy}$ หมายถึง ค่าความคล้ายคลึงของสินค้า x และ y ; r_{ax} และ r_{ay} หมายถึง การให้คะแนนสินค้าที่ผู้ใช้ a มีต่อสินค้า x และ y ; $\bar{r}_a$ หมายถึง ค่าเฉลี่ยของคะแนนที่ผู้ใช้ a มีต่อสินค้าทั้งหมด; N_{xy} หมายถึง เซตของผู้ใช้ที่มีการให้คะแนนทั้งสินค้า x และ y ซึ่งจากสมการดังกล่าวจะได้ค่าความคล้ายคลึงของสินค้าในช่วง [-1,1] จากขั้นตอนนี้จะได้ใช้ค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้า

(item similarity threshold) เพื่อเป็นการระบุสินค้าที่มีความใกล้เคียงกับสินค้าเป้าหมาย ดังนั้นผู้ใช้งานที่พิจารณาเป็นผู้เชี่ยวชาญก็จะต้องให้คะแนนสินค้าเป้าหมายและสินค้าที่มีความคล้ายคลึงกับสินค้าเป้าหมาย โดยจะให้ความสำคัญกับค่าความคล้ายคลึงของสินค้าที่มีค่าเป็นบวก ซึ่งค่าเทรชโฮลด์กำหนดไว้เป็น 0.7, 0.8 และ 0.9 ดังนั้นจากกระบวนการนี้จะได้รายการสินค้าที่จะนำเข้าสู่กระบวนการพิจารณาผู้ใช้งานที่เชี่ยวชาญจากความถี่ในการให้คะแนนและความนิยมของผู้ใช้งานอื่นในระบบ

(2) การพิจารณาผู้ใช้งานที่เชี่ยวชาญจากความถี่ในการให้คะแนน (expert consideration from frequency of rating) การทำงานในส่วนนี้จะเป็นการนับจำนวนครั้งที่ผู้ใช้งานได้ให้คะแนนในสินค้าเป้าหมายและสินค้าที่มีความใกล้เคียงกัน โดยแสดงเป็นสมการได้ดังนี้

$$Fre_a = \sum_{i=1}^{N} m_{a,i} \;\; ; \quad m_a = \begin{cases} 1 & Item \in \text{active item or similar items} \\ 0 & \text{other} \end{cases}$$

โดยที่ Fre_a หมายถึง ความถี่ในการให้คะแนนของผู้ใช้งาน a ; N หมายถึง จำนวนสินค้าทั้งหมด; active item หมายถึงสินค้าเป้าหมาย; similar items หมายถึง สินค้าที่คล้ายคลึงกับสินค้าเป้าหมาย โดยการดำเนินการในส่วนนี้มีพื้นฐานจากการที่ผู้เชี่ยวชาญจะเป็นผู้ที่เผยแพร่ข้อมูลให้กับบุคคลทั่วไป ดังนั้นกระบวนการนี้จะทำให้ได้รายการผู้ใช้งานที่มีความเชี่ยวชาญจากความถี่ในการให้คะแนน

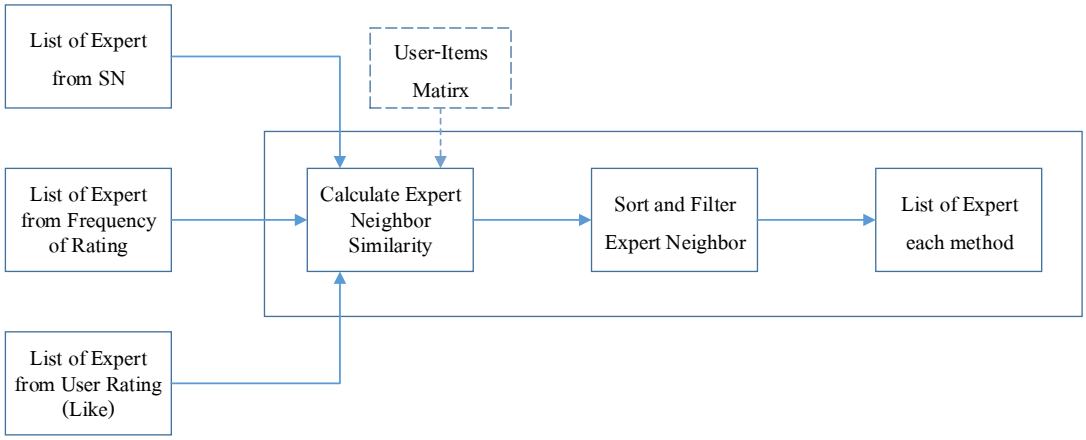
(3) การพิจารณาผู้ใช้งานที่เชี่ยวชาญจากความนิยมของผู้ใช้งานอื่นในระบบ (expert consideration from user rating) การทำงานในส่วนนี้จะคำนวณค่าการให้คะแนนผู้ใช้งาน โดยจะพิจารณาจากการให้คะแนนผู้ใช้งานที่มาจากการให้คะแนนสินค้าเป้าหมายและสินค้าที่มีความคล้ายคลึงกับสินค้าเป้าหมาย โดยแสดงเป็นสมการได้ดังนี้

$$AvgUserRating_a = \frac{\sum_{i=1}^{N_a} UR_{a,i}}{N_a}$$

โดยที่ $AvgUserRating_a$ หมายถึง ค่าเฉลี่ยของคะแนนผู้ใช้งาน a ; UR_a หมายถึง ค่าคะแนนที่ผู้ใช้งานอื่นในระบบให้กับผู้ใช้งาน a จากการให้คะแนนในสินค้าเป้าหมายและสินค้าที่คล้ายคลึงกับสินค้าเป้าหมาย; N_a คือ จำนวนครั้งในการให้คะแนนผู้ใช้งานต่อผู้ใช้งาน a จากผู้ใช้งานอื่นในระบบ สำหรับการดำเนินการในส่วนนี้มีพื้นฐานมาจากการได้รับคะแนนผู้ใช้งานจากผู้ใช้งานรายอื่นในระบบ เป็นการสะท้อนถึงความนิยมที่บุคคลอื่นมีต่อผู้ใช้งานรายนั้น ซึ่งเปรียบได้กับการเป็นผู้เชี่ยวชาญในสายตาของบุคคลอื่น ดังนั้นในกระบวนการนี้จะทำให้ได้รายการผู้ใช้งานที่เชี่ยวชาญจากความนิยมของผู้ใช้งาน (user like)

3.2 การคัดกรองผู้เชี่ยวชาญที่ใช้ในการสร้างการแนะนำ (expert neighbor filtering)

การดำเนินการของขั้นตอนที่ 3.1 จะได้รายการของผู้ใช้งานที่เชี่ยวชาญจากปัจจัยทั้ง 3 ด้าน และขั้นตอนนี้จะได้คำนวณค่าความคล้ายคลึงระหว่างผู้ใช้งานเป้าหมายและผู้ใช้งานที่มีความเชี่ยวชาญ โดยอาศัยข้อมูลการให้คะแนนสินค้า เนื่องจากผู้ใช้งานเป้าหมายแต่ละคนกับผู้ใช้งานที่เชี่ยวชาญอาจมีความชื่นชอบที่แตกต่างกัน โดยการคำนวณค่าความคล้ายคลึงนี้จะพิจารณาจากระยะห่าง (distance) ของการให้คะแนนระหว่างผู้ใช้งาน โดยให้ค่าความคล้ายคลึงระหว่างผู้ใช้งานอยู่ในช่วง [0,1] จากนั้นจึงจัดเรียงลำดับผู้ใช้งานที่เชี่ยวชาญของแต่ละด้าน โดยในการพิจารณาจากข้อมูลเครือข่ายจะจัดเรียงตามค่าจำนวนการเชื่อมโยงที่เข้าสู่ผู้ใช้งาน นอกจากนั้นจะเรียงลำดับตามความถี่ในการให้คะแนนและความนิยม



รูปที่ 3 กระบวนการคัดกรองผู้เชี่ยวชาญที่ใช้ในการสร้างการแนะนำ

ของผู้ใช้อื่นในระบบตามลำดับ จากนั้นจึงใช้ TOP-N method ในการเลือกผู้ใช้งานตามลำดับความเชี่ยวชาญ โดยสามารถแสดงรายละเอียดได้ดังรูปที่ 3

3.3 การสร้างการแนะนำ (prediction)

ขั้นตอนนี้เป็นการทำนายผลลัพธ์การให้คะแนนสินค้าของผู้ใช้งานเป้าหมายต่อสินค้าเป้าหมาย โดยอาศัยการให้คะแนนของผู้ใช้งานที่เชี่ยวชาญ โดยสามารถแสดงเป็นสมการได้ดังนี้

$$P_{ax} = \overline{R_a} + \frac{\sum_{i=1}^{N}(similarity_{ab,i} \times (R_{bx,i} - \overline{R_b}))}{\sum_{i=1}^{N}(similarity_{ab,i})}$$

โดยที่ a หมายถึง ผู้ใช้งานเป้าหมาย; b หมายถึง ผู้ใช้งานที่เชี่ยวชาญ; P_{ax} หมายถึง ผลการทำนายการให้คะแนนสินค้า x ของผู้ใช้งาน a ; R_{bx} เป็นค่าการให้คะแนนสินค้า x ของผู้ใช้ b ; $\overline{R_a}$ และ $\overline{R_b}$ แทนค่าเฉลี่ยการให้คะแนนสินค้าของผู้ใช้งาน a และ b ตามลำดับ; $similarity_{ab}$ หมายถึง ค่าความคล้ายคลึงระหว่างผู้ใช้งาน a และ b ; N แทนจำนวนของผู้ใช้งานที่มีการให้คะแนนสินค้าเป้าหมาย x

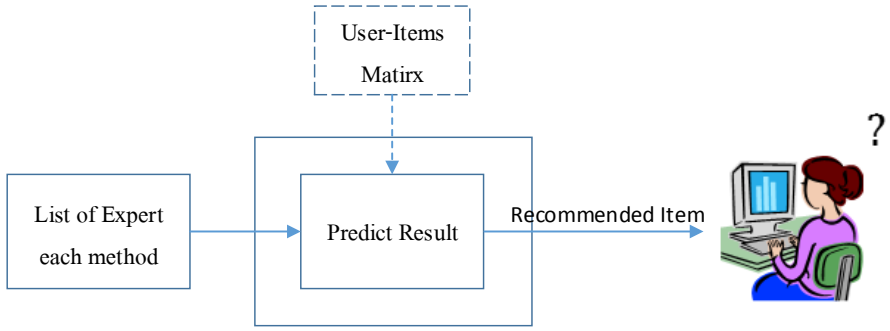
จากนั้นจะปรับค่า P_{ax} ให้อยู่ในช่วงที่ต้องการด้วยสมการ

$$P_{ax} = Min_{new} + \frac{(P_{ax} - Min_{old}) \times (Max_{new} - Min_{new})}{(Max_{old} - Min_{old})}$$

โดยที่ P_{ax} หมายถึง ผลลัพธ์การทำนายในการให้คะแนนสินค้าเป้าหมาย x ของผู้ใช้งาน a ที่ผ่านการปรับช่วงค่าคะแนน; Max_{new} และ Max_{old} หมายถึง ค่าที่มากที่สุดที่เป็นไปได้ในการให้คะแนนสินค้าในอัตราใหม่และอัตราเดิม; Min_{new} และ Min_{old} หมายถึง ค่าที่น้อยที่สุดที่เป็นไปได้ในการให้คะแนนสินค้าในอัตราใหม่และอัตราเดิมตามลำดับ ซึ่งผลลัพธ์ที่ได้จากขั้นตอนนี้เป็นการบ่งบอกถึงความเหมาะสมในการแนะนำสินค้าเป้าหมายดังกล่าวแก่ผู้ใช้งานนั่นเอง สำหรับภาพรวมของกระบวนการนี้แสดงได้ดังรูปที่ 4

4. การทดลองและผลการวิจัย

ส่วนนี้เป็นการนำแนวคิดที่ได้นำเสนอมาทดลอง และเปรียบเทียบความถูกต้องของการทำนายเพื่อแสดงถึงประสิทธิภาพในการทำงานที่ได้จากการระบุผู้เชี่ยวชาญจากแต่ละปัจจัย โดยแบ่งการนำเสนอ



รูปที่ 4 กระบวนการสร้างการแนะนำ

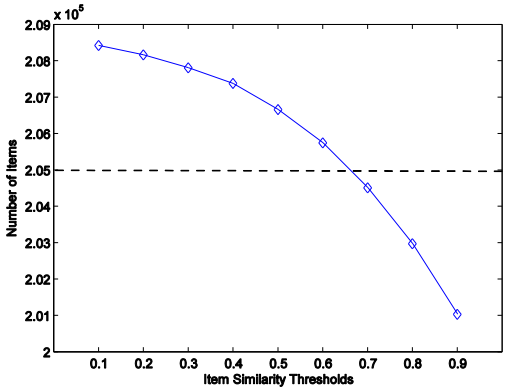
ออกเป็นการออกแบบการทดลอง และ ผลการทดลอง ซึ่งมีรายละเอียดดังนี้

4.1 การออกแบบการทดลอง

สำหรับการออกแบบการทดลองนั้น นอกจากรูปแบบที่ได้นำเสนอแล้ว จำเป็นที่จะต้องพิจารณาเลือกใช้และจัดเตรียมชุดข้อมูลในการทดลอง การกำหนดค่าและปัจจัยการทดลอง ตลอดจนเครื่องมือและวิธีการวัดผลการทดลอง ซึ่งมีดังต่อไปนี้

4.1.1 การเตรียมชุดข้อมูล จากรูปแบบที่นำเสนอ ข้อมูลที่ใช้จะต้องมีทั้งส่วนที่เป็นการเชื่อมโยงในเครือข่าย ข้อมูลการให้คะแนนสินค้าและข้อมูลการให้คะแนนผู้ใช้งาน ดังนั้นในงานวิจัยนี้จึงได้เลือกใช้ชุดข้อมูลจาก Epinion [17] ซึ่งเป็นข้อมูลที่ผู้ใช้งานได้ให้คะแนนสินค้าต่าง ๆ เช่น หนังสือ ภาพยนตร์ เพลง โดยมีการให้คะแนนสินค้าในช่วง 1 ถึง 5 มีการเชื่อมโยงเครือข่ายของผู้ใช้งานในลักษณะกราฟแบบมีทิศทาง และมีการให้คะแนนผู้ใช้งาน (helpfulness) ในช่วง 1 ถึง 5 โดยเมื่อทำความสะอาดข้อมูลแล้ว ทำให้ชุดข้อมูลประกอบไปด้วยผู้ใช้งาน 20,355 คน มีสินค้าจำนวน 30,738 ชิ้น และมีการให้คะแนนสินค้าจำนวน 532,937 รายการ คิดเป็นความหนาแน่นของข้อมูลร้อยละ 0.085 โดยที่จากข้อมูลดังกล่าวได้ใช้เทคนิค K-fold cross validation ในการแบ่งข้อมูลในการทดลอง ซึ่งได้แบ่งเป็น 5 ทบ (fold) แล้วในแต่ละทบได้สุ่มซ่อนข้อมูลการให้คะแนนเพื่อทำการทดลองเป็นจำนวนร้อยละ 10 ของข้อมูลในแต่ละทบ

4.1.2 การกำหนดค่าและปัจจัยการทดลอง จากรูปแบบที่นำเสนอนั้นจะมีการกำหนดค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าและการกำหนดวิธีวัดค่าความคล้ายคลึงระหว่างผู้ใช้งานซึ่งมีรายละเอียดดังนี้



รูปที่ 5 การเปรียบเทียบจำนวนสินค้าเมื่อกำหนดค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าแตกต่างกัน

(1) ค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้า จากสมการที่ใช้ในการคำนวณนั้นจะมีค่าความคล้ายคลึงของสินค้าอยู่ในช่วง [-1,1] แต่ต้องการพิจารณาเฉพาะค่าความคล้ายคลึงที่มีค่าเป็นบวก จึงได้ทดสอบกำหนดค่าเทรชโฮลด์ตั้งแต่ 0.1 ถึง 0.9 ซึ่งจาก

การทดสอบแล้วพบว่าค่าเทรชโฮลด์ของความคล้ายคลึงตั้งแต่ 0.7 ขึ้นไป มีความเหมาะสมทั้งในระดับความคล้ายคลึงของสินค้าและจำนวนสินค้าที่มีเพียงพอต่อการนำไปคำนวณผู้ใช้งานที่เชี่ยวชาญ ดังตัวอย่างกราฟที่แสดงในรูปที่ 5 ดังนั้นจึงได้กำหนดค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าที่ 0.7, 0.8 และ 0.9

(2) วิธีวัดค่าความคล้ายคลึงระหว่างผู้ใช้งาน ในการคำนวณค่าความคล้ายคลึงระหว่างผู้ใช้งานเป้าหมายกับผู้ใช้งานที่เชี่ยวชาญนั้น จะพิจารณาจากระยะห่างของการให้คะแนน ดังนั้นจึงได้เลือกใช้วิธีการคำนวณค่าความคล้ายคลึงจากระยะทาง โดยเลือกวิธีการที่เป็นที่นิยม [18] ได้แก่ Chebychev distance, Manhattan distance, Euclidean distance และ Minkowsky distance ดังสมการ

$$d(a,b) = \left(\sum_{i=1}^{N} \left| R_{a,i} - R_{b,i} \right|^{r} \right)^{1/r}$$

โดยที่ $d(a,b)$ หมายถึง ค่าระยะห่างของการให้คะแนนระหว่างผู้ใช้งานเป้าหมาย a กับผู้ใช้งานที่เชี่ยวชาญ b ด้วยวิธีการต่าง ๆ; R_a และ R_b หมายถึง ค่าคะแนนที่ผู้ใช้งานเป้าหมาย a กับผู้ใช้งานที่เชี่ยวชาญ b ได้ให้ไว้ในสินค้าชนิดเดียวกัน; N หมายถึง จำนวนของสินค้าชนิดเดียวกันที่ผู้ใช้งานทั้งสองคนได้ให้คะแนนไว้; กำหนดให้ r มีค่าเป็น 1, 2, 3, 4, 5, 10, 50 และ 100

จากนั้นจึงแปลงให้เป็นค่าความคล้ายคลึงระหว่างผู้ใช้งานด้วยสมการ [19]

$$sim(a,b) = \frac{1}{1 + d(a,b)}$$

โดยที่ $sim(a,b)$ หมายถึง ค่าความคล้ายคลึงระหว่างผู้ใช้งานเป้าหมาย a กับผู้ใช้งานที่เชี่ยวชาญ b ซึ่งมีค่าในช่วง [0,1]

4.1.3 เครื่องมือและวิธีการวัดผลการทดลอง สำหรับเครื่องมือที่ใช้ในการทดลองประกอบด้วยเครื่องคอมพิวเตอร์ที่มีหน่วยประมวลผลกลางเป็น Intel Corel7-4702MQ 2.20 GHz. หน่วยความจำขนาด 2GB ระบบปฏิบัติการ Microsoft Windows 8 และใช้โปรแกรม Matlab R2012a ร่วมกับ Microsoft SQL Server 2014 ในการทดลอง

ส่วนการวัดผลนั้นได้ใช้วิธีการวัดผลที่เป็นที่นิยม 2 กลุ่ม คือ ความผิดพลาดในการทำนาย และ ความถูกต้องในการทำนาย [20,21] ดังต่อไปนี้

(1) ความผิดพลาดในการทำนาย ใช้ 2 วิธีการ คือ ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute error, MAE) และค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (root mean square error, RMSE) โดยแสดงเป็นสมการได้ดังนี้

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |P_i - R_i|$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (P_i - R_i)^2}$$

โดยที่ P หมายถึง คะแนนที่ได้จากการทำนาย; R หมายถึง คะแนนจริงของสินค้า; N หมายถึง จำนวนสินค้าทั้งหมดที่ประเมินผล ซึ่งทั้ง 2 วิธีการ เป็นการวัดความผิดพลาดที่แตกต่างกันระหว่างค่าที่ได้จากการทำนายและค่าจริง โดยถ้าหากได้ค่าน้อยแสดงว่าวิธีที่นำเสนอสามารถทำนายได้ใกล้เคียงความเป็นจริง

(2) ความถูกต้องในการทำนาย ใช้ 3 วิธีการ คือ ค่าความถูกต้อง (accuracy) ค่าความแม่นยำ (precision) และค่าความระลึก (recall) โดยแสดงเป็นสมการได้ดังนี้

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

โดยที่ *TP* หมายถึง กรณีที่ผลการทำนายแนะนำสินค้านั้นและผู้ใช้งานได้เลือกสินค้านั้น; *FP* หมายถึง กรณีที่

ผลการทำนายแนะนำสินค้านั้น แต่ผู้ใช้งานไม่ได้เลือกสินค้านั้น; *TN* หมายถึง กรณีที่ผลการทำนายไม่แนะนำสินค้านั้น และผู้ใช้งานไม่ได้เลือกสินค้านั้น; *FN* หมายถึง กรณีที่ผลการทำนายไม่แนะนำสินค้านั้น แต่ผู้ใช้งานได้เลือกสินค้านั้น

สำหรับค่าความถูกต้องจะบอกถึงสัดส่วนความถูกต้องในการทำนายว่าใกล้เคียงกับความเป็นจริงเพียงใด ค่าความแม่นยำเป็นการพิจารณาค่าความแม่นยำของการทำนาย และค่าความระลึกเป็นการพิจารณาค่าความครบถ้วนของการทำนาย

4.2 ผลการทดลอง

ส่วนนี้จะได้นำเสนอผลการทดลอง โดยแสดงค่าเฉลี่ยผลการทดลองกับชุดข้อมูล Epinion ทั้ง 5 ทบ ซึ่งประกอบด้วย ผู้เชี่ยวชาญจากข้อมูลเครือข่าย (expert from social network, ESN) ผู้เชี่ยวชาญจากความถี่ในการให้คะแนน(expertfromfrequency, EF) และ ผู้เชี่ยวชาญจากความนิยมของผู้ใช้งานอื่นในระบบ (expert from user like, EL) ทั้งนี้ผู้เชี่ยวชาญจากความถี่ในการให้คะแนนและจากความนิยมของผู้ใช้งานอื่นนั้นมีการกำหนดค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าที่ 0.7, 0.8 และ 0.9 ซึ่งสามารถแสดงผลได้ดังตารางที่ 1 ถึงตารางที่ 5

ตารางที่ 1 ค่าเฉลี่ยการวัดค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute error, MAE)

Measure	ESN	EF (0.7)	EF (0.8)	EF (0.9)	EL (0.7)	EL (0.8)	EL (0.9)
Chebychev	0.147	**0.146**	**0.146**	**0.146**	0.148	0.147	0.147
Euclidean	0.147	**0.146**	**0.146**	**0.146**	0.148	0.148	0.147
Manhattan	0.147	**0.146**	**0.146**	**0.146**	0.148	0.148	0.147
Minkowsky r=3	0.147	**0.146**	**0.146**	**0.146**	0.148	0.147	0.147
Minkowsky r=4	0.147	**0.146**	**0.146**	**0.146**	0.148	0.147	0.147
Minkowsky r=5	0.147	**0.146**	**0.146**	**0.146**	0.148	0.148	0.147
Minkowsky r=10	0.147	**0.146**	**0.146**	**0.146**	0.148	0.147	0.147
Minkowsky r=50	0.147	**0.146**	**0.146**	**0.146**	0.148	0.147	0.147
Minkowsky r=100	0.147	**0.146**	**0.146**	**0.146**	0.148	0.147	0.147

ตารางที่ 2 ค่าเฉลี่ยการวัดค่ารากที่สองของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (root mean square error, RMSE)

Measure	ESN	EF (0.7)	EF (0.8)	EF (0.9)	EL (0.7)	EL (0.8)	EL (0.9)
Chebychev	0.383	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384
Euclidean	0.383	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384
Manhattan	0.384	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384
Minkowsky r=3	0.383	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384
Minkowsky r=4	0.383	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384
Minkowsky r=5	0.383	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384
Minkowsky r=10	0.383	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384
Minkowsky r=50	0.383	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384
Minkowsky r=100	0.383	**0.382**	**0.382**	**0.382**	0.384	0.384	0.384

ตารางที่ 3 ค่าเฉลี่ยการวัดค่าความถูกต้อง (accuracy)

Measure	ESN	EF (0.7)	EF (0.8)	EF (0.9)	EL (0.7)	EL (0.8)	EL (0.9)
Chebychev	85.294	85.401	85.401	85.396	85.238	85.252	85.267
Euclidean	85.295	85.397	85.394	85.393	85.231	85.247	85.264
Manhattan	85.287	85.387	85.385	85.382	85.217	85.239	85.253
Minkowsky r=3	85.297	**85.404**	**85.402**	**85.402**	85.236	85.251	85.268
Minkowsky r=4	85.298	85.401	85.401	85.399	85.236	85.250	85.269
Minkowsky r=5	85.296	85.401	85.401	85.398	85.236	85.249	85.267
Minkowsky r=10	85.297	85.400	85.401	85.398	85.239	85.252	85.269
Minkowsky r=50	85.294	85.401	85.401	85.397	85.238	85.252	85.267
Minkowsky r=100	85.294	85.401	85.401	85.396	85.238	85.252	85.267

ตารางที่ 4 ค่าเฉลี่ยการวัดค่าความแม่นยำ (precision)

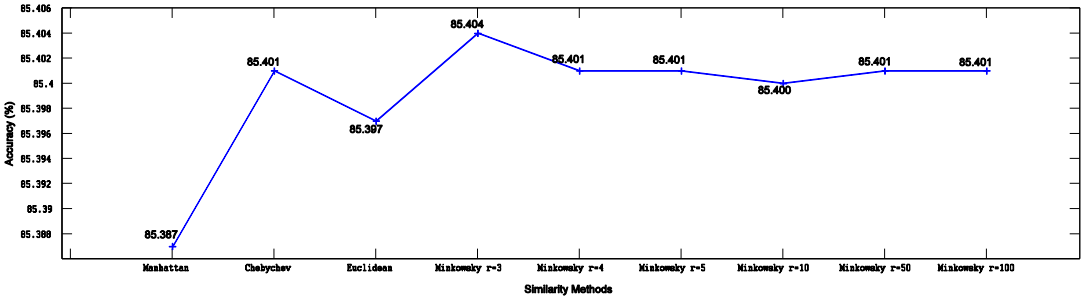
Measure	ESN	EF (0.7)	EF (0.8)	EF (0.9)	EL (0.7)	EL (0.8)	EL (0.9)
Chebychev	0.979	**0.984**	**0.984**	**0.984**	0.978	0.978	0.979
Euclidean	0.979	**0.984**	**0.984**	**0.984**	0.978	0.978	0.979
Manhattan	0.979	0.983	0.983	0.983	0.977	0.977	0.978
Minkowsky r=3	0.979	**0.984**	**0.984**	**0.984**	0.978	0.978	0.979
Minkowsky r=4	0.979	**0.984**	**0.984**	**0.984**	0.978	0.978	0.979
Minkowsky r=5	0.979	**0.984**	**0.984**	**0.984**	0.978	0.978	0.979
Minkowsky r=10	0.979	**0.984**	**0.984**	**0.984**	0.978	0.978	0.979
Minkowsky r=50	0.979	**0.984**	**0.984**	**0.984**	0.978	0.978	0.979
Minkowsky r=100	0.979	**0.984**	**0.984**	**0.984**	0.978	0.978	0.979

ตารางที่ 5 ค่าเฉลี่ยการวัดค่าความระลึก (recall)

Measure	ESN	EF (0.7)	EF (0.8)	EF (0.9)	EL (0.7)	EL (0.8)	EL (0.9)
Chebychev	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866
Euclidean	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866
Manhattan	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866
Minkowsky r=3	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866
Minkowsky r=4	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866
Minkowsky r=5	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866
Minkowsky r=10	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866
Minkowsky r=50	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866
Minkowsky r=100	0.866	0.864	0.864	0.864	**0.867**	**0.867**	0.866

4.2.1 พิจารณาผลความผิดพลาดในการทำนาย เมื่อพิจารณาประสิทธิภาพจากความผิดพลาดในการทำนายด้วยตารางที่ 1 และ 2 แล้ว เห็นได้ว่าค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยที่ได้นั้นมีความใกล้เคียงกันทั้งผลจาก ESN, EF และ EL คือ 0.14 แต่ถ้าพิจารณาให้ละเอียดแล้วพบว่าผลที่ได้จาก EF จะมีค่าความผิดพลาดน้อยที่สุด ทั้งนี้อาจเป็นไปได้จากปริมาณความถี่ในการให้คะแนนจะส่งผลต่อแนะนำสินค้า

4.2.2 พิจารณาผลความถูกต้องในการทำนาย เมื่อพิจารณาประสิทธิภาพจากความถูกต้องในการทำนาย ด้วยตารางที่ 3 ถึง 5 แล้ว เห็นได้ว่าค่าความถูกต้องอยู่ในระดับใกล้เคียงกันคือร้อยละ 85 ค่าความแม่นยำอยู่ในระดับ 0.97 ถึง 0.98 และค่าความระลึกอยู่ในระดับ 0.86 และเมื่อพิจารณาโดยละเอียดในทีละส่วนพบว่าสำหรับค่าความถูกต้องนั้นผลที่ได้จาก EF จะมีความถูกต้องสูงกว่ารูปแบบอื่น โดยเฉพาะเมื่อใช้วิธีคำนวณค่าความคล้ายคลึงระหว่างผู้ใช้งานเป็น Minkowsky distance ที่ r=3 และค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าเป็น 0.7 จะให้ค่าความถูกต้องสูงที่สุด ดังแสดงผลในรูปที่ 6 และสำหรับค่าความแม่นยำนั้น ผลลัพธ์ที่ได้จาก EF ก็ให้ค่าสูงสุดแทบทุกวิธีการคำนวณค่าความคล้ายระหว่างผู้ใช้งาน ส่วนค่าความระลึกจะให้ผลที่แตกต่างกัน คือ ผลลัพธ์ที่ได้จาก EF จะให้ค่าที่น้อยกว่ารูปแบบอื่น

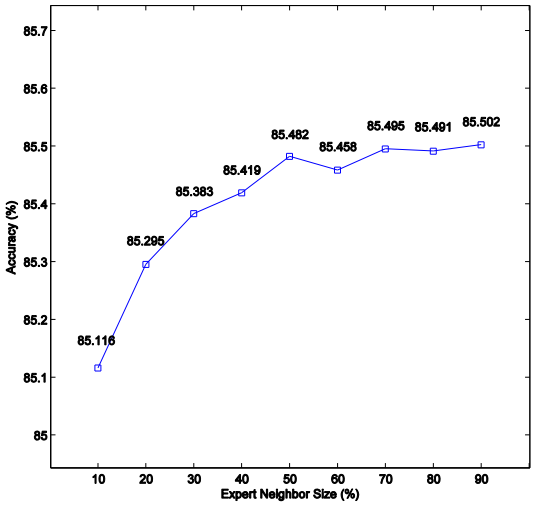


รูปที่ 6 การเปรียบเทียบค่าความถูกต้องขณะพิจารณาผู้ใช้งานที่เชี่ยวชาญจากความถี่ในการให้คะแนนด้วยวิธีวัดค่าความคล้ายคลึงระหว่างผู้ใช้งานที่แตกต่างกัน โดยที่มีค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าเป็น 0.7

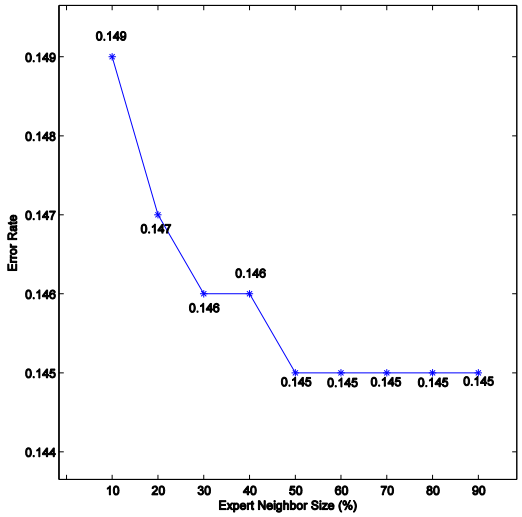
4.2.3 พิจารณาผลตามจำนวนของประชากร เมื่อพิจารณาผลลัพธ์ความผิดพลาดในการทำนายและความถูกต้องในการทำนายจะเห็นได้ว่ารูปแบบ EF ที่มีค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าที่ 0.7 และคำนวณค่าความคล้ายคลึงระหว่างผู้ใช้งานด้วยวิธี Minkowsky distance ที่ r=3 จะให้ผลลัพธ์ที่ดีกว่าวิธีการอื่น ดังนั้นจึงได้ใช้รูปแบบดังกล่าวพิจารณาในส่วนของค่าความถูกต้อง และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย เมื่อมีการกำหนดจำนวนประชากรที่แตกต่างกัน โดยแสดงผลได้ดังรูปที่ 7 และ 8 ซึ่งสำหรับค่าความถูกต้องที่ได้นั้น จะเห็นได้ว่ามีแนวโน้มเพิ่มขึ้นตามจำนวนของประชากร แต่ทั้งหมดก็อยู่ที่ ร้อยละ 85 ส่วนค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยนั้นจะลดลงตามจำนวนของประชากรที่เพิ่มขึ้น โดยหยุดนิ่งเมื่อประชากรมีขนาดร้อยละ 50 แต่ทั้งนี้ค่าที่ได้ก็จะอยู่ในระดับ 0.14

4.2.4 สรุปผลการทดลอง จากผลการทดลองที่ได้นำเสนอนั้นจะเห็นได้ว่าสามารถที่จะประยุกต์แนวคิดผู้เชี่ยวชาญมาใช้ในการสร้างการแนะนำได้ โดยผลลัพธ์ที่ได้นั้นมีค่าความถูกต้องที่

ร้อยละ 85 ค่าความแม่นยำอยู่ในระดับ 0.98 ค่าความระลึกอยู่ในระดับ 0.86 และค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยอยู่ที่ระดับ 0.14 ซึ่งเมื่อเปรียบเทียบกับงานวิจัยอื่นที่ใช้ชุดข้อมูลเดียวกันแต่อาจมีการจัดการข้อมูลต่างกันแล้วพบว่า งานวิจัยที่นำเสนอให้ผลลัพธ์ที่เหมาะสมกว่าในส่วนของค่าความแม่นยำและค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ย [7,22,23] นอกจากนี้เมื่อเปรียบเทียบรูปแบบที่ใช้พิจารณาผู้ใช้งานที่เชี่ยวชาญจากที่ได้นำเสนอไป พบว่าการพิจารณาผู้ใช้งานที่เชี่ยวชาญด้วยความถี่ของการให้คะแนนจะให้ผลลัพธ์ที่เหมาะสมที่สุด โดยเฉพาะในกรณีที่ค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าอยู่ที่ 0.7 ทั้งนี้เป็นไปได้จากการที่เมื่อพิจารณาสินค้าที่มีความคล้ายคลึงกันเป็นการแสดงถึงพฤติกรรมในการให้คะแนนของผู้ใช้งาน โดยเมื่อได้กำหนดค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าที่เหมาะสมก็จะช่วยในการคัดกรองสินค้าที่ส่งผลต่อการจัดลำดับผู้ใช้งานที่มีความเชี่ยวชาญด้วย และสำหรับวิธีการวัดค่าความคล้ายคลึงระหว่างผู้ใช้งานที่ได้นำเสนอนั้นให้ผลลัพธ์ที่ใกล้เคียงกัน แต่วิธี Minkowsky distance ที่ r=3 จะมีค่าความถูกต้องมากกว่าวิธีการอื่น ทั้งนี้อาจเนื่องมาจากรูปแบบของสมการ Minkowsky distance ที่ r=3 จะทำให้มีระยะห่างมากที่สุดในกลุ่มของ Minkowsky distance และเมื่อแปลงเป็นค่าความคล้ายคลึงแล้วก็ส่งผลทำให้ได้ผลลัพธ์ที่เหมาะสมกับชุดข้อมูลมากกว่า



รูปที่ 7 การเปรียบเทียบค่าความถูกต้องขณะพิจารณาผู้ใช้งานที่เชี่ยวชาญจากความถี่ในการให้คะแนน เมื่อมีจำนวนผู้ใช้งานที่เชี่ยวชาญแตกต่างกัน โดยที่มีค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าเป็น 0.7 และคำนวณค่าความคล้ายคลึงระหว่างผู้ใช้งานด้วยวิธี Minkowsky distance ที่ r=3



รูปที่ 8 การเปรียบเทียบค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยขณะพิจารณาผู้ใช้งานที่เชี่ยวชาญจากความถี่ในการให้คะแนน เมื่อมีจำนวนผู้ใช้งานที่เชี่ยวชาญแตกต่างกัน โดยที่มีค่าเทรชโฮลด์ของความคล้ายคลึงของสินค้าเป็น 0.7 และคำนวณค่าความคล้ายคลึงระหว่างผู้ใช้งานด้วยวิธี Minkowsky distance ที่ r=3

5. สรุปการวิจัย

งานวิจัยนี้ได้นำเสนอรูปแบบในการนำแนวคิดผู้เชี่ยวชาญมาใช้เพื่อสร้างการแนะนำสินค้า โดยอาศัยข้อมูลการให้คะแนนสินค้า ข้อมูลการให้คะแนนผู้ใช้งาน ร่วมกับข้อมูลเครือข่าย โดยได้นำเสนอการพิจารณาความเป็นผู้เชี่ยวชาญจากปัจจัย 3 ด้าน ได้แก่ ด้านการเชื่อมโยงในเครือข่าย ด้านความถี่ในการให้คะแนนสินค้า และด้านความนิยมของผู้ใช้งานอื่นในระบบ จากนั้นจึงได้พิจารณาความคล้ายคลึงระหว่างผู้ใช้งานเป้าหมายและผู้ใช้งานที่เชี่ยวชาญเพื่อสร้างการแนะนำ และได้ทดลองวัดประสิทธิภาพการทำนายการแนะนำจากแต่ละปัจจัยที่นำเสนอ ซึ่งจากผลการทดลองแสดงให้เห็นว่าปัจจัยทั้ง 3 ด้าน ให้ผลลัพธ์ใกล้เคียงกัน โดยมีความถูกต้องในการทำนายที่ร้อยละ 85 โดยเฉพาะเมื่อพิจารณาความเป็นผู้เชี่ยวชาญจากความถี่ในการให้คะแนนจะให้ผลลัพธ์ที่เหมาะสมมากที่สุด ซึ่งผลที่ได้จากการงานวิจัยนี้แสดงให้เห็นว่ารูปแบบที่นำเสนอนั้นสามารถนำไปปรับใช้งานได้โดยเฉพาะกับธุรกิจพาณิชย์อิเล็กทรอนิกส์ ที่จะช่วยนำเสนอสินค้าจากผู้ใช้งานที่เชี่ยวชาญซึ่งอาจมีความเหมาะสมต่อผู้ใช้งานเป้าหมายได้อย่างดียิ่งขึ้น สำหรับงานวิจัยในอนาคตนั้นจะได้นำข้อมูลประเภทสินค้ามาใช้ประกอบการพิจารณาผู้ใช้งานที่เชี่ยวชาญเพื่อปรับปรุงประสิทธิภาพการทำนาย

6. รายการอ้างอิง

[1] Bobadilla, J., Ortega, F., Hernando, A. and Gutiérrez, A., 2013, Recommender systems survey, Knowl. Based Syst. 46: 109-132.

[2] Sun, Z., Han, L., Huang, W., Wang, X., Zeng, X., Wang, M. and Yan, H., 2015, Recommender systems based on social networks, J. Syst. Softw. 99: 109-119.

[3] Afzal, M.T. and Maurer, H.A., 2011, Expertise Recommender System for Scientific Community, J. UCS. 17: 1529-1549.

[4] Davoodi, E., Kianmehr, K. and Afsharchi, M., 2013, A semantic social network-based expert recommender system, Appl. Intell. 39: 1-13.

[5] Li, Y.M., Hsiao, H.W. and Lee, Y.L., 2013, Recommending social network applications via social filtering mechanisms, Inf. Sci. 239: 18-30.

[6] Lin, C., Xie, R., Guan, X., Li, L. and Li, T., 2014, Personalized news recommendation via implicit social experts, Inf. Sci. 254: 1-18.

[7] Guo, G., Zhang, J. and Thalmann, D., 2014, Merging trust in collaborative filtering to alleviate data sparsity and cold start, Knowl. Based Syst. 57: 57-68.

[8] Jøsang, A., Ismail, R. and Boyd, C., 2007, A survey of trust and reputation systems for online service provision, Emerg. Issues Collab. Commer. 43: 618-644.

[9] Liao, Q.V., Wagner, C., Pirolli, P. and Fu, W.T., 2012, Understanding experts' and novices' expertise judgment of twitter users, pp. 2461-2464, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[10] Song, S., Lee, S., Park, S. and Lee, S., 2012, Determining user expertise for improving recommendation performance, p. 67, Proceedings of the 6 th International

Conference on Ubiquitous Information Management and Communication.

[11] Martin-Vicente, M.I., Gil-Solla, A., Ramos-Cabrer, M., Blanco-Fern´ndez, Y. and Lopez-Nores, M. 2011, Improving e-commerce collaborative recommendations by semantic inference of neighbors' practical expertise, pp. 9-14, The Sixth International Workshop on Semantic Media Adaptation and Personalization (SMAP).

[12] Li, Y.M., Liao, T.F. and Lai, C.Y., 2012, A social recommender mechanism for improving knowledge sharing in online forums, Large-Scale Distrib. Syst. Inf. Retr. 48: 978-994.

[13] Li, M., Liu, L. and Li, C.B., 2011, An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems, Expert Syst. Appl. 38: 8586-8596.

[14] Wu, D., Bao, L., Regan, A.C. and Talcott, C.L., 2013, Large-scale access scheduling in wireless mesh networks using social centrality, J. Parallel. Distrib. Comput. 73: 1049-1065.

[15] Molinero, X., Riquelme, F. and Serna, M., 2013, Power indices of influence games and new centrality measures for social networks, ArXiv Prepr. ArXiv13066929.

[16] Shambour, Q. and Lu, J., 2012, A trust-semantic fusion-based recommendation approach for e-business applications, Decis. Support Syst. 54: 768-780.

[17] Product Review Datasets: Epinions and Ciao, Available Source: http://www.public.asu.edu/~jtang20/datasetcode/truststudy.htm, November 12, 2015.

[18] D'Souza, J.F., Adams, C.K. and Reed, A., 2015, Validation of region-based crossover for clustering problems, Int. J. Imag. Robot. 15(2): 53-66.

[19] Winarko, E., Hartati, S. and Wardoyo, R., 2014, Improving the Prediction Accuracy of Multicriteria Collaborative Filtering by Combination Algorithms.

[20] Lika, B., Kolomvatsos, K. and Hadjiefthymiades, S., 2014, Facing the cold start problem in recommender systems, Expert Syst. Appl.,41: 2065-2073.

[21] Xie, F., Chen, Z., Shang, J. and Fox, G.C., 2014, Grey forecast model for accurate recommendation in presence of data sparsity and correlation, Knowl. Based Syst. 69: 179-190.

[22] Chen, C., Zeng, J., Zheng, X. and Chen, D., 2013, Recommender system based on social trust relationships, pp. 32-37, IEEE: The 10th International Conference on e-Business Engineering (ICEBE).

[23] Chen, C.C., Wan, Y.H., Chung, M.C. and Sun, Y.C., 2013, An effective recommenddation method for cold start new users using trust and distrust networks, Inf. Sci. 224: 19-36.