1.บทคัดย่อ

ระบบแนะนำรายการโทรทัศน์ของประเทศสหรัฐอเมริกานั้น เป็นระบบที่ช่วยในการตัดสินใจของผู้บริโภคในการรับชมสื่อให้เข้ากับความต้องการของตัวเอง รวมทั้งเหมาะสมกับ ช่วงวัยอีกด้วย เพราะในประเทศสหรัฐอเมริกานั้นมีรายการโทรทัศน์ที่มากมายและหลากหลาย ซึ่งในบางครั้งผู้บริโภคอาจจะไม่ได้รับชมรายการที่ต้องการ หรือไม่รู้จักรายการใดที่เหมาะสมกับความต้องการเลย แต่ต้องการคำแนะนำ บทความนี้นำเสนอรูปแบบของระบบแนะนำรายการโทรทัศน์ โดยอาศัยรูปภาพและอัลกอริธึมของระบบแนะนำแบบ เค-มีน คลัสเตอร์ริ่ง (K-mean Clusttering)

คำสำคัญ : ระบบแนะนำ เค-มีน คลิสเตอร์ริ่ง

1.บทนำ

ในปัจจุบันระบบแนะนำ (Recommender System) ถูกนำมาใช้เป็นอย่างมากในเรื่องต่างๆ ไม่ว่าจะเป็นเรื่อง ความสนใจ การทำงาน การเรียน และในด้านความบันเทิง ซึ่งระบบตัดสินใจนั้นจะต้องมีกลุ่มของข้อมูล (Data Set) ที่มีความแม่นยำและคลาดเคลื่อนน้อยที่สุด โดยระบบแนะนำนั้นผู้ใช้จะต้องกรอกข้อมูลที่ระบบแนะนำต้องการให้ครบถ้วนและต้องใกล้เคียงหรือตรงกับผู้ใช้มากที่สุด เพื่อให้การประมวลผลออกมานั้นมีความใกล้เคียงกับความต้องการของผู้ใช้ระบบแนะนำมากที่สุด ซึ่งในบางครั้งการประมวลผลของระบบแนะนำก็อาจจะคลาดเคลื่อนหรือผิดพลาดซึ่งบางทีก็มาจากตัวระบบแนะนำเอง หรือบางทีก็มาจากผู้ใช้ที่อาจจะกรอกข้อมูลไม่ครบถ้วน หรือไม่ถูกต้องก็เป็นได้

โดยในระบบแนะนำนั้นมักใช้วิธีการ K-mean ซึ่งเป็นอัลกอริธึมรูปแบบหนึ่งที่มีการนำมาใช้กับระบบแนะนำ <มีต่อ แต่ยังคิดไม่ออก>

2.เอกสารหรืองานวิจัยที่เกี่ยวข้อง

2.1 วิธีการ เคมีน

เคมีน คลัสเตอร์ริ่ง อัลกอริธึมนั้น ถูกพัฒนาขึ้นโดย เจ.แม๊คควีน แล้วจากนั้น เจ.เอ. ฮาร์ติงเก้น และ เอ็ม.เอ. หว่องก็ได้รับช่วงนำไปพัฒนาต่อ โดยพื้นฐานของวิธีการ เคมีนนั้นจะแบ่งแยกกลุ่มของข้อมูลโดยใช้ คุณสมบัติแปลงเป็น เลขจำนวน เค (k) ซึ่งเป็นจำนวนของกลุ่ม และ เค (k) นั้นจะต้องเป็นเลขจำนวนเต็มบวก ซึ่งเมื่อเราได้ทำการรวมค่าเรียบร้อยแล้ว ก็จะเกิด จุดศูนย์กลางของคลัสเตอร์ (Cluster Centroid)

(ภาพแผนผังอัลกอริธึม)

2.2 แผนผังการทำงาน

ขั้นที่ 1 นั้นเริ่มต้นการตัดสินใจของอัลกอริธึมโดยที่ดูจากจำนวนค่า เค ซึ่งมีค่าเท่ากับจำนวนของคลัสเตอร์

ขั้นที่ 2 นำพาร์ทิชั่น (Partition) เริ่มต้นที่มีหน้าที่ในการแยกประเภทข้อมูล ลงไปเก็บไว้ที่ เค คลัสเตอร์ ซึ่งเราสามารถทำการทดสอบแบบสุ่มหรือกระบวนการที่เป็นไปตามระบบได้โดยสามารถทำตามดังนี้ :

2.1 นำค่าที่จะใช้ทดสอบค่าแรกของเค ให้เป็น คลัสเตอร์องค์ประกอบเดี่ยว (Single-element clusters)

2.2 แล้วกำหนดค่าที่เหลือให้เป็นคลัสเตอร์ที่มีความใกล้เคียงกับจุดศูนย์กลาง แล้วคำนวณจุดศูนย์กลาง (Centroid) อีกครั้งเพื่อให้ได้คลัสเตอร์

ขั้นที่ 3 ให้นำตัวอย่างที่เห็นได้จากภาพรวมนั้น แล้วคำนวณระยะห่างของคลัสเตอร์จากจุดศูนย์กลาง (Centroid) ถ้าเกิดว่าตัวอย่างนั้นที่กำลังอยู่ภายในคลัสเตอร์นั้นไม่ได้อยู่ใกล้กลับจุดศูนย์กลาง (Centroid) ให้ทำการเปลี่ยนข้อมูลภายในคลัสเตอร์นั้นแล้วอัพเดท (update) จุดศูนย์กลาง (Centroid) ของคลัสเตอร์ตัวนั้นๆ เพื่อที่จะได้รับตัวอย่างตัวใหม่และคลัสเตอร์ที่สูญเสียไป

ขั้นที่ 4 ให้ทำตามขั้น 3 ใหม่ไปเรื่อยๆจนกว่าจะเกิดการบรรจบกัน (Convergence) หรือจนกว่าค่าที่ใช้ทดสอบนั้นไม่มีการกำหนดค่าเข้ามาใหม่อีกครั้ง