

CSC662 ปฏิบัติการ ๑๒

การวิเคราะห์การเกาะกลุ่มในซอฟต์แวร์ Weka

เขียนโดย ผศ. ดร. กรุง สีนอกกรมย์สราญ

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

เนื้อหาที่ครอบคลุม

- แนะนำซอฟต์แวร์ Weka
- ฐานข้อมูลเชิงสัมพันธ์และการเตรียมข้อมูล
- สถิติและการทำเหมืองข้อมูลแบบอธิบาย
- การทำเหมืองข้อมูลแบบกฎเชื่อมโยง
- การทำเหมืองข้อมูลแบบจำแนกประเภท
- การทำเหมืองข้อมูลแบบการวิเคราะห์การเกาะกลุ่ม
- การใช้ผังการไหลของความรู้ Knowledge flow

2

การวิเคราะห์การเกาะกลุ่ม (Cluster analysis)

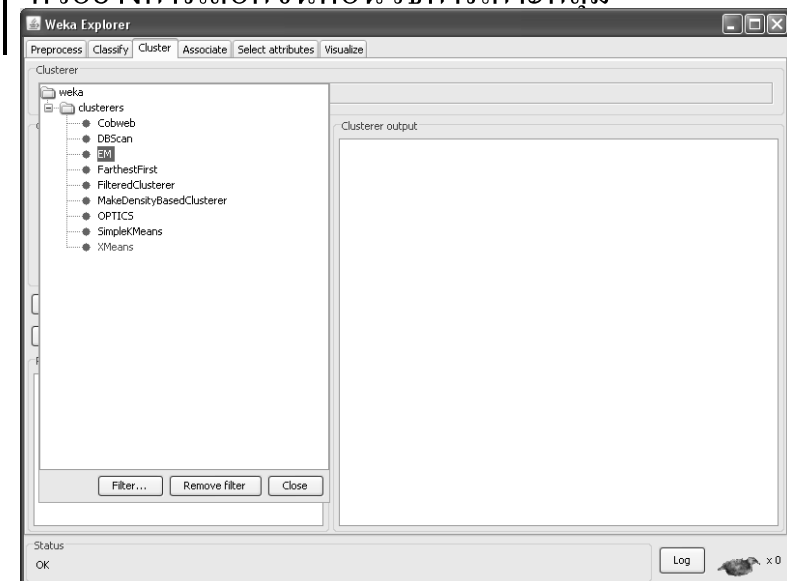
ขั้นตอนวิธีการวิเคราะห์การเกาะกลุ่มใน Weka

- Cobweb ใช้หลักการจัดจำแนกโดยต้นไม้และความน่าจะเป็น
- DBScan วิธีการเกาะกลุ่มโดยใช้ความหนาแน่น
- EM การเกาะกลุ่มโดยใช้ค่าคาดคะเนที่มากที่สุด
- FarthestFirst การเกาะกลุ่มโดยเลือกตัวที่ไกลที่สุดก่อน
- Filtered Cluster การเกาะกลุ่มที่ผ่านการกรองก่อน
- MakeDensityBasedClusterer วิธีการเกาะกลุ่มที่ใช้หลายวิธีรวมกัน
- OPTICS วิธีการเกาะกลุ่มโดยใช้หลักความหนาแน่น
- SimpleKMeans วิธีการเกาะกลุ่มแบบแบ่งกัน โดยใช้ค่าเฉลี่ยหรือฐานนิยม
- XMeans วิธีการเกาะกลุ่มที่ไม่ต้องกำหนดค่า K

3

การวิเคราะห์การเกาะกลุ่ม (Cluster analysis)

ตัวอย่างการเลือกขั้นตอนวิธีการเกาะกลุ่ม



เพิ่ม weather.arff

@relation weather

- @attribute outlook {sunny, overcast, rainy}
- @attribute temperature real
- @attribute humidity real
- @attribute windy {TRUE, FALSE}
- @attribute play {yes, no}

@data

sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no

5

การวิเคราะห์การเกาะกลุ่ม (Cluster analysis)

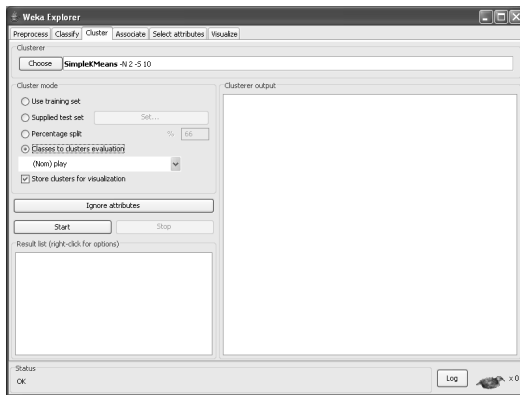
การเตรียมเพิ่มข้อมูล

- กำหนดลักษณะประจำที่มีค่าแตกต่างกันทั้งหมดออกก่อน เช่น คีย์หลัก (Primary key)
- กำหนดลักษณะประจำที่มีค่าเพียงค่าเดียวออก
- ลักษณะประจำที่ใช้อาจเป็น Numeric หรือ Categorical ก็ได้
- ในกรณีที่ต้องการเปรียบเทียบการเกาะกลุ่มกับลักษณะประจำเป้าหมายที่น่าสนใจ ให้เลือก Classes to cluster evaluation ในการวิเคราะห์ซอฟต์แวร์ Weka จะไม่ใช่ลักษณะประจำเป้าหมายในการคิดระยะ

6

การวิเคราะห์การเกาะกลุ่ม (Cluster analysis)

ขั้นตอนการเกาะกลุ่มโดย K-means

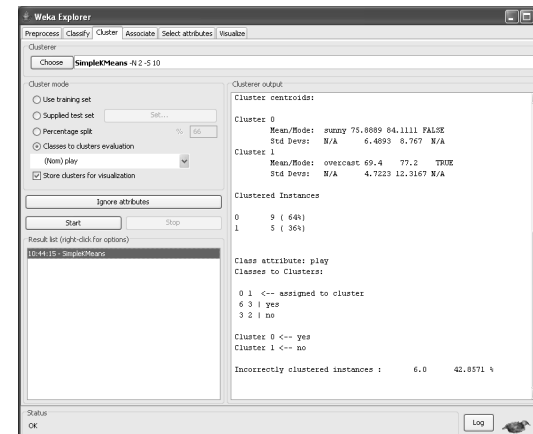


- เปิด โมดูล Explorer ของซอฟต์แวร์ Weka
- เปิดเพิ่ม weather.arff
- เลือกแถบ Cluster
- เลือก SimpleKMeans ในกล่อง Clusterer

7

การวิเคราะห์การเกาะกลุ่ม (Cluster analysis)

ผลลัพธ์ที่ได้จาก SimpleKMeans

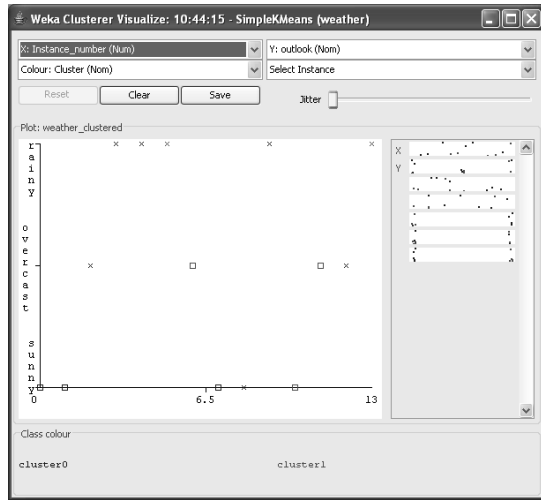


- เลือก Classes to clusters evaluation เพื่อเปรียบเทียบผลที่ได้จากการเกาะกลุ่มกับคลาสเป้าหมายที่ต้องการ
- ลักษณะประจำที่อยู่เหนือปุ่ม start ต้องเป็นคลาสเป้าหมายที่ต้องการ
- กดปุ่ม Start จะได้ผลลัพธ์ดังรูปด้านซ้าย

8

การวิเคราะห์การเกาะกลุ่ม (Cluster analysis)

ผลที่ได้จากการเลือก Visualize cluster assignments



- เลือก Visualize cluster assignments โดยกดเมาส์ปุ่มขวาจากผลลัพธ์ในกล่อง Result list
- กราฟแสดงผลของการกำหนดกลุ่มได้ดังภาพด้านขวา

สรุป

- การทำเหมืองข้อมูลแบบการวิเคราะห์การเกาะกลุ่มมีขั้นตอนวิธีให้เลือกใช้อยู่ 9 แบบสำหรับ Weka 3.5.3
- ถ้าผู้ใช้ต้องการเปรียบเทียบระหว่างกลุ่มกับคลาส เลือกใช้ Classes to cluster evaluation
- เราสามารถดูด้วยภาพนามธรรมโดยเลือก Visualize cluster assignment