

Day 54 非監督式機器學習

非監督式機器學習簡介





周俊川

知識地圖非監督學習



機器學習概論 Introduction of Machine Learning

監督式學習 Supervised Learning

前處理 Processing 探索式 數據分析 Exploratory Data Analysis

特徵 工程 Feature Engineering 模型 選擇 Model selection

參數調整 Fine-tuning

集成 Ensemble 非監督式學習 Unsupervised Learning

> 分群 Clustering

降維 Dimension Reduction

非監督學習

Unsupervised learning

非監督簡介

分群 Clustering K-平均算法 K-Mean

階層分群法 Hierarchical Clustering

降類 Dimension Deduction 主成分分析PCA(Principal components analysis)

T分佈隨機近鄰嵌入 t-SNE



本日知識點目標

- 瞭解非監督式學習 (unsupervised learning)相關技術概要。
- 會解非監督式學習的應用場景。

什麼是非監督式學習?

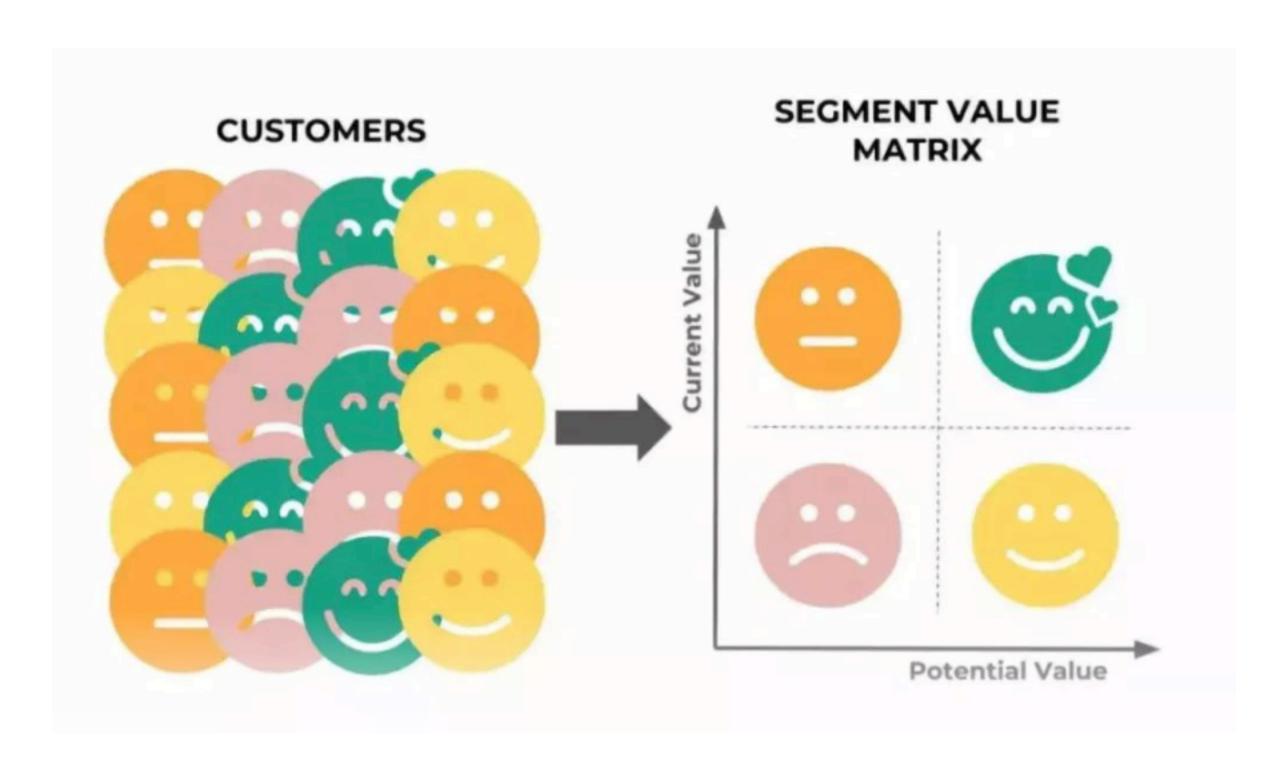


非監督學習允許我們在對結果無法預知時接近問題。非監督學習演算法只基於輸入資料找出模式。當我們無法確定尋找內容,或無標記(y)資料時,通常會用這個演算法,幫助我們了解資料模式。

應用案例 (一) :客戶分群



在資料沒有任何標記,或是問題還沒定義清楚前,可用分群的方式幫助理清資料特性。

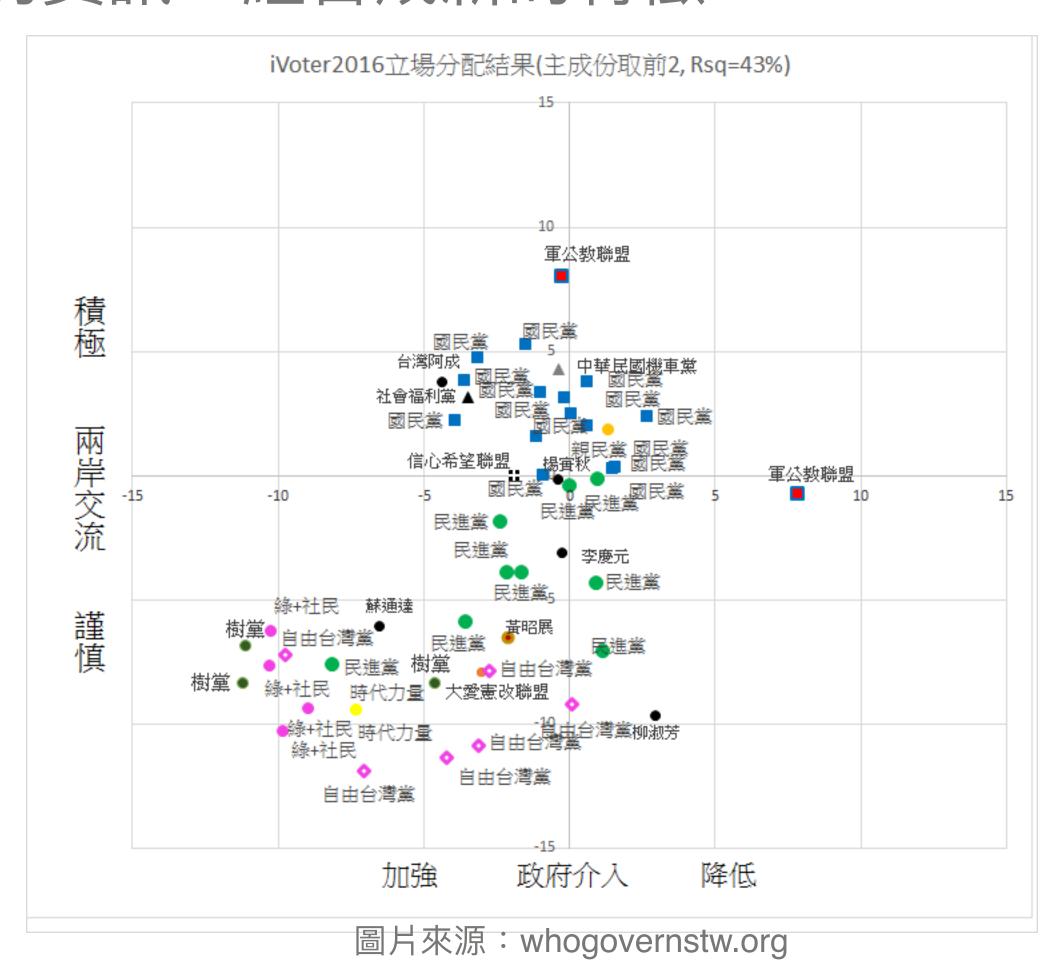


圖片來源:blog.csdn

應用案例 (二) :特徵抽象化



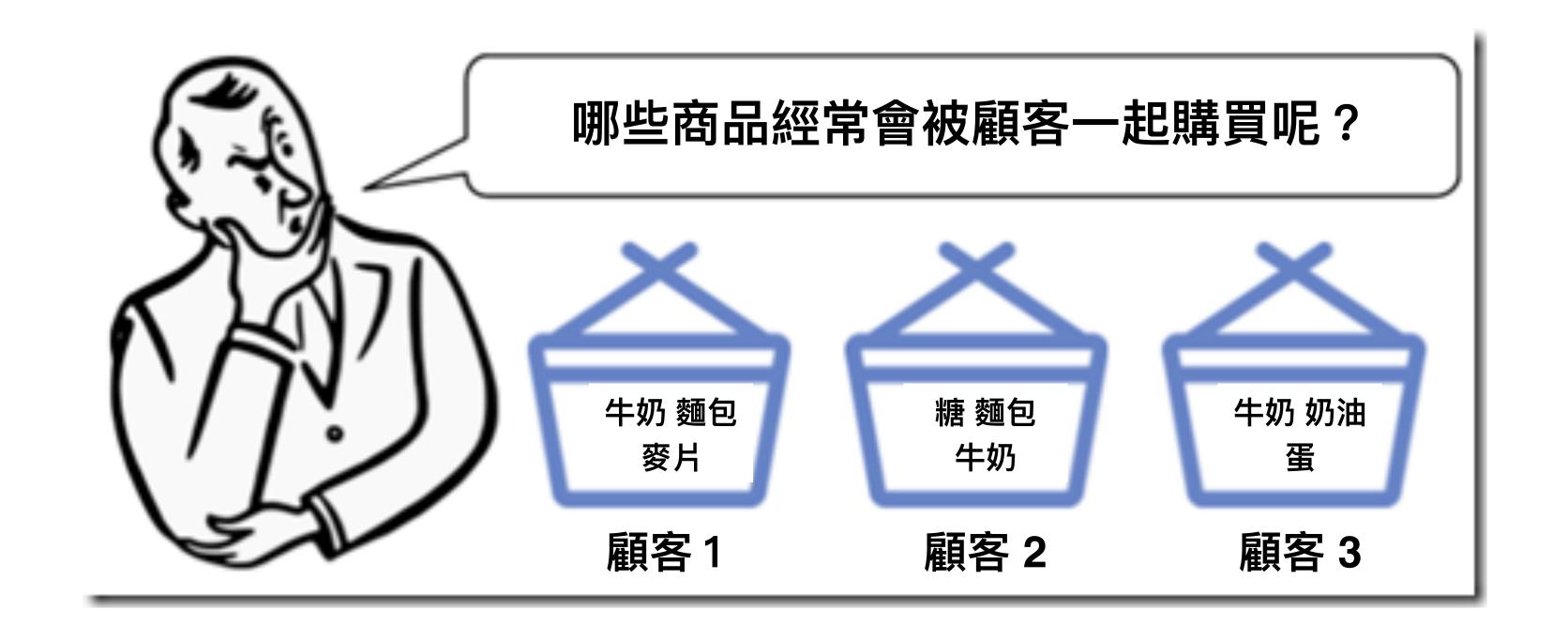
特徵數太多難於理解及呈現的情況下,藉由抽象化的技術幫助降低資料維度,同時不失去原有的資訊,組合成新的特徵。



應用案例 (三) :購物籃分析



資料探勘的經典案例,適用於線下或線上零售的商品組合推薦。



圖片來源:pulipuli.tumblr

應用案例 (四) :非結構化資料分析



非結構化資料如文字、影像等,可以藉由一些非監督式學習的技術,幫助呈現及描述資料。

· 主題模型 (topic model)



圖片來源:stackoverflow

非監督學習算法概要



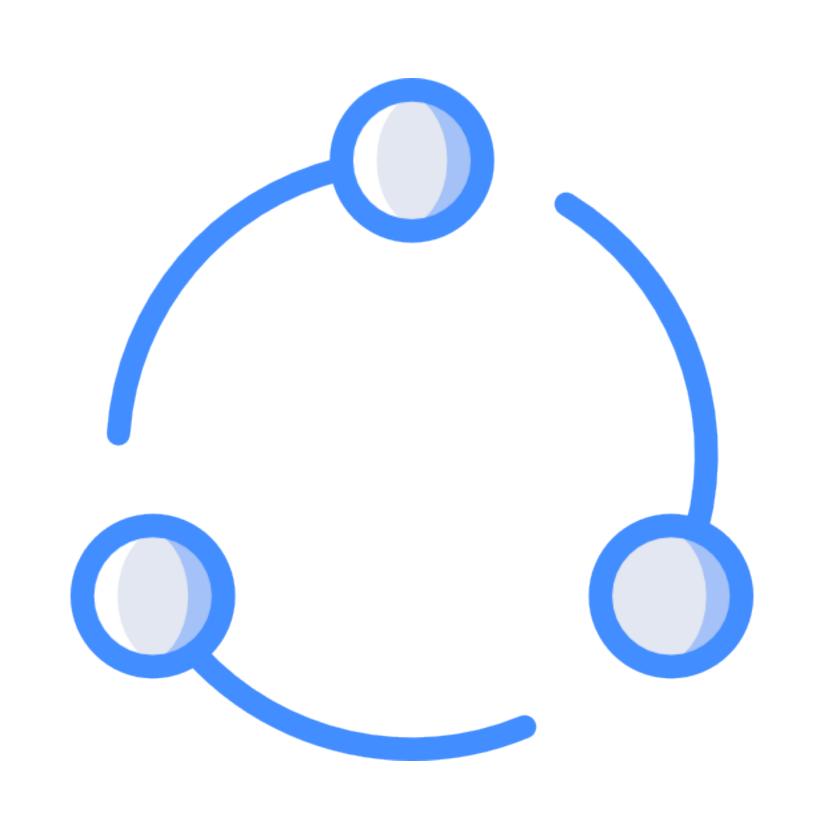
平 聚類分析: 尋找資料的隱藏模式

降低維度:特徵數太大且特徵間相關性高,以此方式縮減特徵維度

● 其他: 關聯法則 (購物籃分析)、異常值偵測、探索性資料分析等

重要知識點複習





- 在不清楚資料特性、問題定義、沒有標記的情況下,非監督式學習技術可以幫助我們理清資料脈絡
- 特徵數太龐大的情況下,非監督式學習可以幫助概念抽象化,用更簡潔的特徵描述資料
- 非監督式學習以聚類算法及降低維度算法爲主, 本課程也以這兩門技術進行探究



請跳出PDF至官網Sample Code&作業 開始解題

