

圖窮景獻 – Picture Your Way

Scatter / Gather 結合 LDA 與 K-Means 建構圖片景點推薦系統

指導老師：杜逸寧

專題組員：蔡孟恬、蘇奕華、杜孟澐、張家語、王鈺婷、施建宏

• 摘要

本研究是一個以圖片為主軸的推薦系統，使用了政府開放平台上的景點資料集「景點－觀光資訊資料庫」，配合時下流行的社群軟體－Instagram中的照片，結合多個演算法，並設計數個實驗驗證可行性。此系統可將景點依不同類型分群，融合圖片的美觀排序，最終得以快速找出使用者適合的景點類型進而做推薦。。

• 研究目的

本研究欲探討如何從海量的資料集中「快速」及「準確」的提供預測結果，為了解決此問題而研發出的新演算法，提出一個較為創新的推薦方法並且能夠快速地命中使用者潛在之喜好景點。

• 研究方法

第一部分 –

本研究地毯式搜集政府開放資料集中的景點於Instagram上的圖片，同時利用Jieba斷詞演算法，對資料集中的景點簡述進行預處理以利於後續的建模，得到景點與圖片的資料庫。

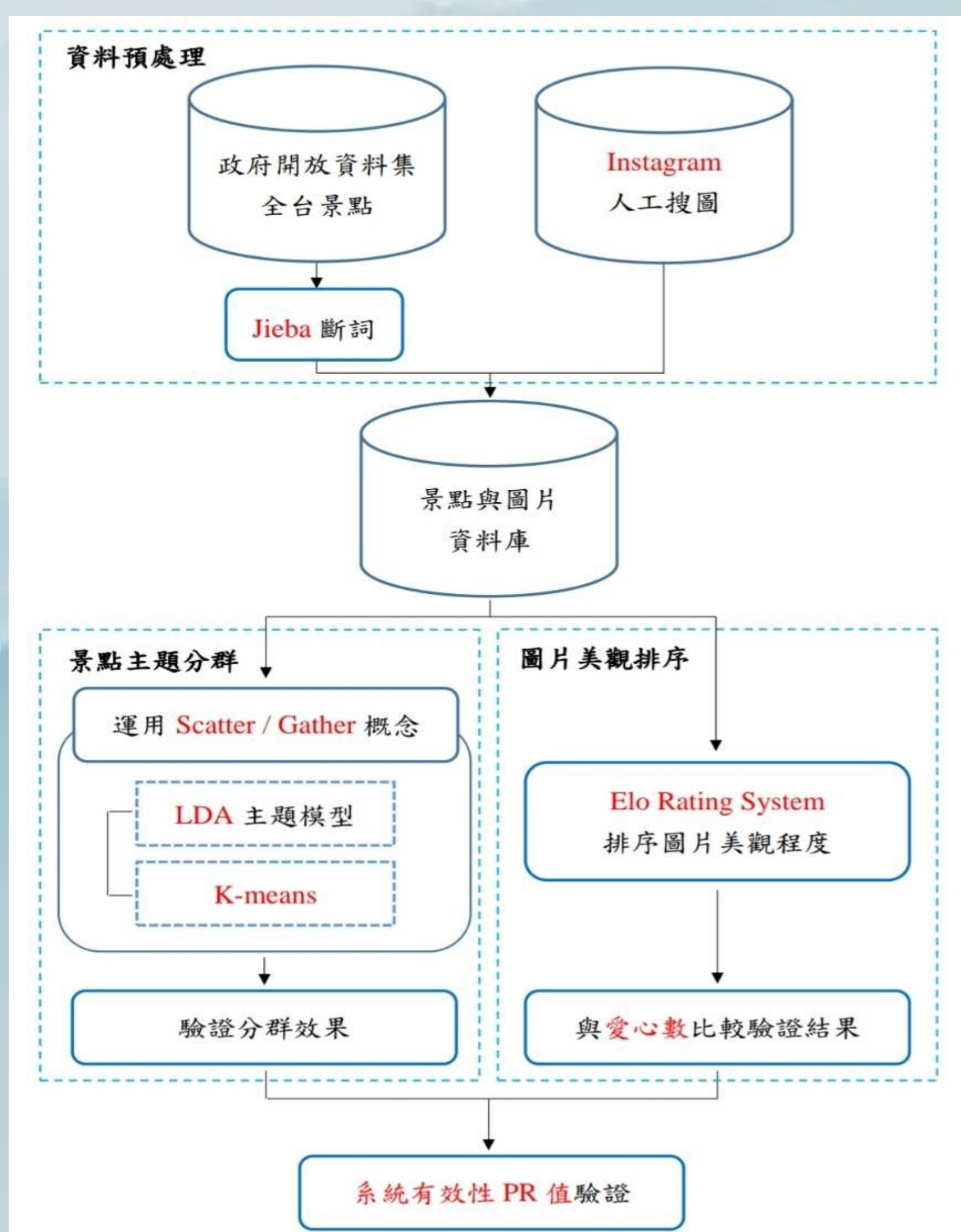
第二部分 –

將Scatter/Gather的概念運用在景點分群，並融合Latent Dirichlet Allocation主題模型與K-Means分群法作為其分群過程使用的演算法。從LDA主題模型挖掘各景點的潛在主題，並萃取各景點的LDA主題機率分布作為K-Means分群的變數。

除了景點類型外，景點美觀亦為社會大眾選擇景點時的重要因素，因此透過Elo Rating System反覆訓練景點圖片賦予美觀排序，並設計實驗收集問卷，將結果進行Spearman相關分析，結果證明Elo Rating System相較於Instagram中圖片本身的愛心數量來得可行。

第三部分 –

設計了一電腦版系統問卷蒐集實驗數據，計算系統有效性PR值，用實際的數據推測本系統之實用程度。



• 結論與建議

本研究所使用的概念並非只可運用在景點推薦系統而已，圖片的視覺影響一直以來都相當直接，大眾多可僅僅經由圖片的展示引發興趣，因此不論是在旅館民宿與餐廳美食等觀光產業上都能以此相同架構進行快速推薦。本系統的應用亦可更加精緻化，起初為了顯示演算法的效用並無限制縣市，而加上地區篩選的功能並非難事，甚至可以在未來做更多的優化。

• 結果與討論

雖然圖片具有相當直觀的特性，但現今的技術仍然無法完善地將圖片以其內容主題進行快速分群，為了達到此效果，本研究分析每個景點背後的簡述卻使用其圖片代為呈現，以此提供使用者圖片分群的效果。與以往直接對圖片進行各種分析、色調比對甚至內容偵測等不同，在這種種方法皆效果不佳的情況下，本研究大膽地在文字與圖片做了鏈結並進行分析，並為驗證整體概念以及演算法的可行性設計數個實驗。

• 簡要文獻：

Bester, D. W., & Maltitz, M. J. (2013). *Introducing Momentum to the Elo Rating System* (Master's thesis). University of the Free State: Department of Mathematical Statistics and Actuarial Science, 5-30.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.