

傳統與結合 LDA 之協同過濾推薦法 在新穎性和相關性指標之比較

統碩一 涂筱宜

目錄

01. 前言

02. 協同過濾及LDA介紹

03. 研究方法

04. 實驗結果

05. 結論



01.

前言

推薦系統

推薦系統 (Recommendation System) 是一種協助使用者過濾資訊的系統，透過分析客戶的潛在偏好來推薦相關的資訊與產品給使用者，偏好來自於使用者的個人資訊，其中包含喜好、評分、點閱紀錄。

推薦系統

- 過去推薦系統的研究普遍都是以針對改善推薦相關性，但單憑相關性並不能夠進一步來評估推薦系統的實際效用。
 - 近年來許多研究提出衡量推薦不同層面的指標，例如：多樣性(Diversity)、驚奇性(Surprise)、新穎性(Novelty)和偶然性 (Serendipity)等。
-

研究目的

- 認為傳統的協同過濾有不足之處，**使用 LDA 模型可以提升相關性的推薦之外，對新穎性的推薦也有所幫助**，並利用 LDA 推薦使用者從未關注的項目。
 - 新穎性指的是推薦使用者可能關注過或已經購買過之外的冷門項目。
-

02.

協同過濾及LDA介紹

協同過濾

傳統協同過濾(Collaborative Filtering) 是目前推薦系統最普遍使用的技術，**利用項目之間的關聯性或相似興趣的群體作為分析基準**，進一步計算相似性來推薦使用者感興趣的物品。

協同過濾 (Collaborative Filtering)



User Based

以**使用者為基礎**的協同過濾，依**分析相似興趣或相似行為的使用者為同群體**，進一步計算相鄰的其他使用者所對項目的評分，來進行推薦給使用者可能感興趣的項目。



Item Based

以**項目為基礎**的協同過濾，**依據項目之間的相似度**，無須由使用者過去的歷史資訊，尋找相鄰使用者。

協同過濾缺點

- 傳統協同過濾著重於找出最相似的項目或群體，並未考慮項目內容的潛在語意，造成推薦結果的相關性降低。
 - 另一方面是考量使用者對項目評分高低產生推薦，但評分低的項目容易被忽略，造成推薦的新穎性降低。
-

LDA

- 隱含迪利克雷分布 (**Latent Dirichlet Allocation, LDA**) 是一種廣泛使用的機率主題模型，是一種**非監督式學習**的演算法。
 - LDA主題模型是通過對每個文件進行塑模，**探勘出文件中的潛在主題**，集中每篇文件的主題按照機率分布的形式產生。
-

結合LDA的協同過濾

結合LDA主題模型的協同過濾能夠彌補傳統協同過濾的缺點，LDA能夠從項目內容中找出潛在語意的主題分布，藉由推薦相似主題之項目，可避免推薦以熱門項目為主的情形，達到相關性和新穎性的推薦。



03.

研究方法

相關性的評估指標

研究使用平均絕對誤差 (**Mean Absolute Error, MAE**) 作為相關性的衡量指標，是用來比對預測分數以及實際評分之間的差距。

$$MAE = \frac{\sum_{u=1}^N |P_{u,i} - R_{u,i}|}{N}$$

MAE 值越低，表示誤差較小，代表有越佳的推薦相關性。

新穎性的評估指標

研究採用三種新穎性指標來觀察推薦系統的成果

(1) 是基於自信息的新穎性方法視為 N1

$$I_a = \log_2(u/k_a)$$

其中 k_a/u 為任意一位使用者選到 a 項目的機率

u 為使用者的總數

新穎性的評估指標

研究採用三種新穎性指標來觀察推薦系統的成果

(2) 第二種是發現為基礎的項目新穎性視為 N2

$$\text{nov}(i|\theta) = 1 - p(\text{seen}|i, \theta)$$

$$p(\text{seen}|i, r) \sim \frac{|i|}{|\mathcal{U}|} = \frac{|\{u \in \mathcal{U} | r(u, i) \neq \emptyset\}|}{|\mathcal{U}|}$$

其中 $p(\text{seen}|i, \theta)$ 為使用者在 θ 的情境下會看到項目 i 的機率

新穎性的評估指標

研究採用三種新穎性指標來觀察推薦系統的成果

(3) 第三種是距離為基礎的項目新穎性視為 N3

$$nov(i|\theta) = \frac{\sum_{j \in \theta} p(\text{choose}|j, \theta, i)d(i|j)}{\sum_{j \in \theta} p(\text{choose}|j, \theta, i)}$$

其中 $p(\text{choose}|j, \theta, i)$ 為使用者已選擇項目 i 的情況下，
在 θ 的情境下，使用者會選擇項目 j 的機率



04.

實驗結果

相關性指標的表現

Models	MAE
Item-Based	1.3559
User-Based	0.7607***
LDA-IB	0.8291
LDA-UB	0.8427
備註：* $p < 0.1$ ；** $p < 0.05$ ；*** $p < 0.01$	

新穎性指標的表現

Models	N1	N2	N3	Sum of Novelty
Item-Based	8.396***	0.990***	0.481***	9.867***
User-Based	2.907	0.847	0.462	4.217
LDA-IB	5.869	0.817	0.431	7.117
LDA-UB	3.439	0.887	0.463	4.789

備註：* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$



05.

結論

- 結合 LDA 之 User-Based 的推薦法比傳統協同過濾 User-Based 的推薦法，不只有提升新穎性的推薦，也犧牲較小相關性達到更好的推薦。
 - 在相關性的表現中傳統協同過濾 Item-Based 的推薦法是最差，在結合 LDA 之 Item-Based 的推薦法大幅改善，得到更好的相關性但在新穎性表現而是降低。
-

從兩種指標不同的面向分析，更有助於分析其推薦結果，依據結果建議使用者選擇適當的推薦方法，獲得使用者想要的推薦。

The image features a light blue background with decorative wavy lines in the top corners and a horizontal line at the bottom. The text "THANK YOU" is centered in a bold, black, serif font.

THANK YOU