商品推薦系統

一基於知識的推薦系統(|+||)

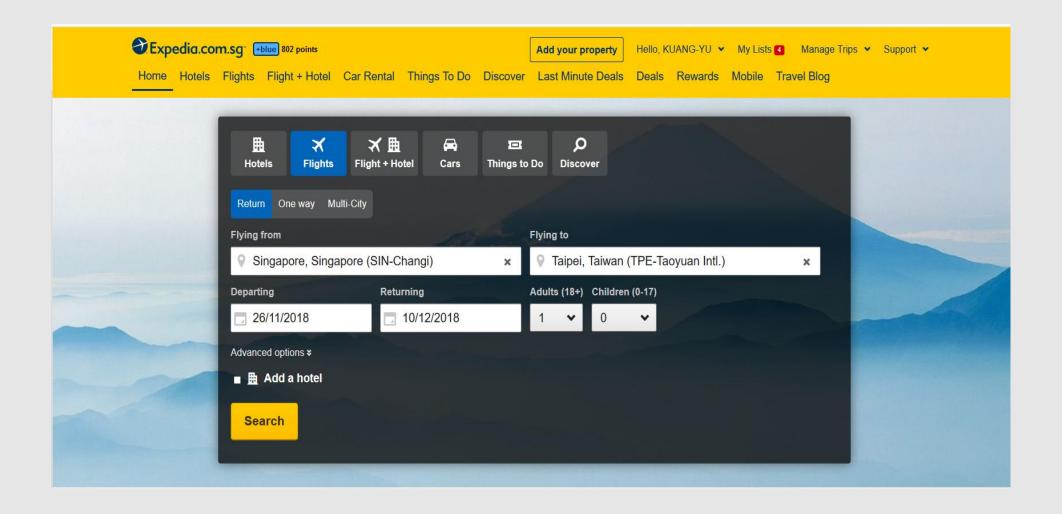
Lecture 4

講師:周光宇博士

2018.08







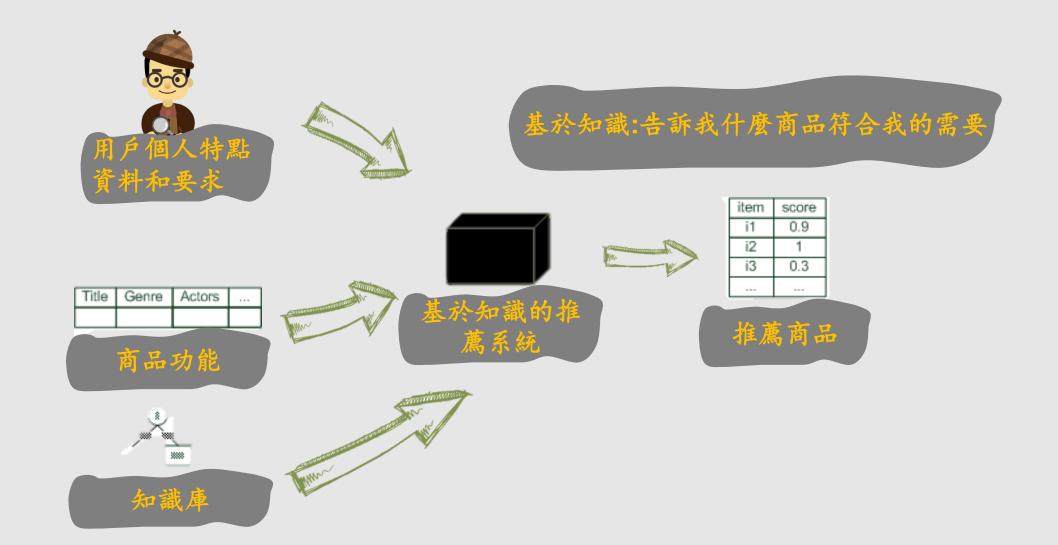


簡介

- 基於內容和協同過濾的系統都需要大量有關過去購買和評級體驗的數據。如果數據量有限,就會發生冷啟動問題。
- 此外,這些方法並不適用於房地產、汽車、旅遊要求、交友、金融服務 或昂貴奢侈品等高度客制化商品的領域。這樣的物品消費一般很少,通 常很難有足夠的評價和最新的數據。在許多情況下,商品領域可能很複 雜,只有少數幾個特定屬性的實例。
- 還有,如果商品功能隨時間而變化,如手機、電腦等,因為過去的評分不能使用,這些方法也不適用。
- 基於知識的推薦系統被設計來處理這些情況。它依賴於明確地徵求用戶對這些商品的要求,讓用戶探索固有的複雜產品空間,並了解各種選擇之間可用的權衡。



基於知識的推薦系統





複雜商品的例子: 數位相機

id	price(\$)	mpix	opt-zoom	LCD-size	movies	sound	waterproof
p1	148	8.0	4×	2.5	no	no	yes
p2	182	8.0	5x	2.7	yes	yes	no
<i>p</i> 3	189	8.0	10x	2.5	yes	yes	no
<i>p</i> 4	196	10.0	12x	2.7	yes	no	yes
<i>p</i> 5	151	7.1	3x	3.0	yes	Yes	no
<i>p</i> 6	199	9.0	3x	3.0	yes	yes	no
<i>p</i> 7	259	10.0	3x	3.0	yes	yes	no
p8	278	9.1	10x	3.0	yes	yes	yes



複雜商品的例子: 購買房屋

ltem-ld	Beds.	Baths.	Locality	Туре	Floor Area	Price
1	3	2	Bronx	Townhouse	1600	220.000
2	5	2	Chappaqua	Split-level	3600	973,000
3	4	2	Yorktown	Ranch	2600	630,000
4	2	1	Yorktown	Condo	1500	220,000
5	4	2	Ossining	Colonial	2700	430,000





推薦系統的比較

推薦方法	概念目標	輸入
協同過濾	一個基於我和其他人對於某些商品的評價/購買行動的合作格式	用户評分和社群評分
基於內容	基於我過去的評價/購買行動和商品內容的格式	用戶評分和商品屬性
基於知識	一個基於我對這些商品的具體要求的格式	用戶特定的要求, 商品屬性和領域知識



基於知識的推薦系統的類型

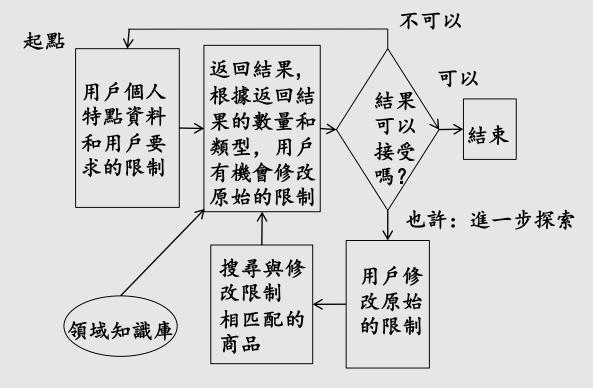
- 基於限制的(Constraint-Based)推薦系統:在這些系統中,用戶通常在商品屬性/功能上指定需求或限制(例如,下限或上限)。此外,特定於領域的規則用於將用戶要求或特點與商品屬性/功能相匹配,或者將用戶特點與商品屬性/功能相關聯。根據返回結果的數量和類型,用戶有機會修改原始需求。
- 基於案例的(Case-Based)推薦系統:在這些系統中,用戶將特定情況或商品指定為目標案例或定位點。然後,搜尋與這些目標案例相似的商品。 返回的結果通常用作新的目標案例,用戶可以進行一些互動式的評估和修改。
- 在這兩種情況下,系統都為用戶提供了改變指定要求的機會。 但是,兩種情況的做法是不同的。



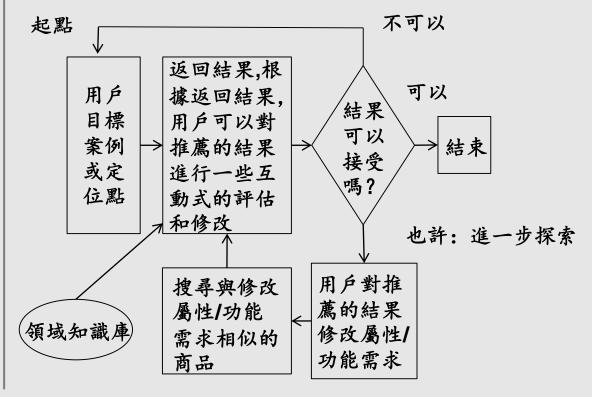
基於知識的推薦系統中的互動過程



基於限制



國 基於案例





基於限制的推薦系統

- 基於限制的推薦系統允許用戶硬性指定商品屬性的要求或約束條件。
- 使用一組領域知識的規則來將用戶需求與商品屬性相匹配。
- 用戶不是領域專家,可能不能總是用描述商品的相同屬性來指定他們的查詢;因此,需要額外的一套領域知識的規則來對應用戶需求和商品屬性。
- 例如: "保守型投資"可以對應到"國債投資"(用戶需求), "單身"可以對應到"臥室數量 < 3"(用戶個人資料)。



基於限制的推薦問題

• 從此目錄中選擇符合用戶要求的項目

id	price(\$)	mpix	opt-zoom	LCD-size	movies	sound	waterproof
P ₁	148	8.0	$4\times$	2.5	no	no	yes
P_2	182	8.0	5×	2.7	yes	yes	no
P_3	189	8.0	10×	2.5	yes	yes	no
P_4	196	10.0	12×	2.7	yes	no	yes
P_5	151	7.1	$3 \times$	3.0	yes	yes	no
P_6	199	9.0	$3 \times$	3.0	yes	yes	no
P ₇	259	10.0	$3 \times$	3.0	yes	yes	no
P ₈	278	9.1	10×	3.0	yes	yes	yes

- 例如, 用户的要求可以是
 - "價格應該低於300美元"
 - "相機應該適合戶外運動攝影"





- 有三種主要數據的輸入類型:
 - 描述用戶的固有特性(例如人口統計個人特點資料,風險概況等)和商品中特定的限制要求(例如最低的臥室數量)的屬性。
 - 一知識庫將客戶特性/需求對應到各種商品屬性的知識庫。知識庫來源於公開可用的訊息、領域專家、過去的經驗或歷史數據集的數據挖掘。
 - 商品目錄包含所有商品的列表以及相對應的商品屬性。
- 如果不同的用戶指定相同的輸入,他們將得到相同的結果。 這個 特點對大多數基於知識的系統來說是很常見的。

一個基於約束推薦系統的用戶界面示例



EXAMPLE OF HYPOTHETICAL CONSTRAINT-BASED INTERFACE FOR HOME BUYING (constraint-example.com) [ENTRY POINT]						
I WOULD LIKE TO BUY A HOUSE SATISFYING THE FOLLOWING REQUIREMENTS:						
MIN. BR	MAX. BR	MIN. BATH	MAX. BATH			
MIN. PRICE	MAX. PRICE SUBMIT	HOME STYLE SEARCH	ZIP CODE			

Source: Recommender Systems – The Textbook



返回相關結果 (步驟1)

- 對於用戶指定的每個限制要求(或個人特性),將檢查它是否與知識庫某些規則的預設條件相匹配。如果存在這樣的匹配,則該規則的結果條件則被視為有效的選擇條件。
- 例如,如果客户在其個人特性和偏好中指定了Family-Size = 6 和 Zip Code = 10547,則可能檢測到 Family-Size = 6 觸發以下規則:
 - Family-Size >= 5 => Min-Bedrooms >= 3
 - Family-Size >= 5 => Min-Bathrooms >= 2
 - Min-Bedrooms >= 3 => Price >= \$100,000
- 這些規則作為限制添加到用戶的需求中。



返回相關結果(步驟2和步驟3)

- 這些擴展的需求被用來構成一個一般的數據庫查詢。 這代表了一個計算商品目錄上限制交集的傳統數據庫查詢:
 - $(Bedrooms \ge 3) \cap (Bathrooms \ge 2) \cap (Price \ge \$100,000) \cap (ZipCode = 10547)$
- 該方法基本上將所有用戶特性和需求/限制對應到商品領域中的屬性/功能限制。
- 然後使用此查詢來找出與用戶需求相關目錄中的實例。
- 然後將找出的商品列表呈現給用戶。 用戶可以進一步修改她的要求以獲得更精確的推薦。
- 在整個過程中,起始的探索和進一步的修改要求可以導致用戶找到有用的推薦。一般來說,這是無法只靠她自己的探索來發現。

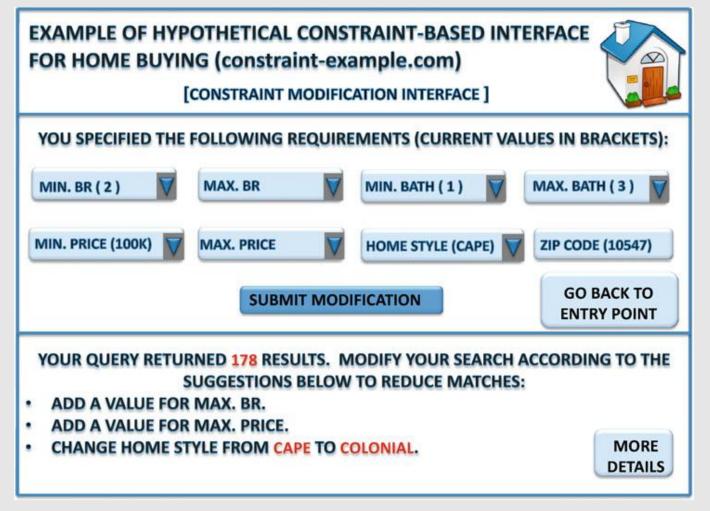


處理缺少輸入值的方法

- 如果用戶無法指定與商品屬性相關的所有值,則有幾種可能的解決方案:
 - 系統可以使未指定的屬性不受限制, 並僅基於指定的限制產生結果。
 - 可以向用戶建議預置值(default values)以提供指導,並讓用戶決定是使用,修改還是忽略這些值。
- 預置值是可以學習的。歷史數據集的平均值可以用作預置值。

在基於約束的推薦系統中用於處理太多查詢結果的用戶界面示例





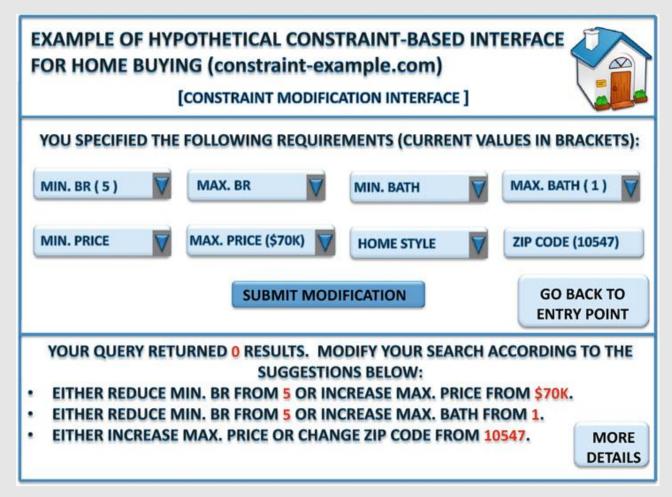
匹配商品的排序



- 有許多根據用戶要求對商品進行排序的自然方法。
- 最簡單的方法是讓用戶指定一個商品的單一屬性來排列匹配商品。
- · 一個常用的方法是使用效用函數(Utility Functions) 對匹配的商品進行排名。例如,效用函數可以被定義為各個屬性效用的加權函數。
- 讓 $V = (v_1...v_d)$ 定義為匹配商品屬性的向量值。每個屬性分配一個權重 w_j ,同時,它有一個由效用函數定義的貢獻 $f_i(v_i)$ 。
- 然後, 匹配商品的效用 $U(\bar{V})$ 由下列公式給出: $U(\bar{V}) = \sum_{j=1}^{d} w_{j}. f_{j}(v_{j})$
- 設計一個好的效用函數的通常需要特定的領域知識,或從過去用戶互動的數據中學習。

在基於約束的推薦系統中用於處理無效查詢結果用戶界面的示例







處理不可滿足的需求

id	price(€)	mpix	opt-zoom	LCD-size	movies	sound	waterproof
P ₁	148	8.0	4×	2.5	no	no	yes
P_2	182	8.0	5×	2.7	yes	yes	no
P_3	189	8.0	10×	2.5	yes	yes	no
P_4	196	10.0	12×	2.7	yes	no	yes
P ₅	151	7.1	$3 \times$	3.0	yes	yes	no
P ₆	199	9.0	$3 \times$	3.0	yes	yes	no
P ₇	259	10.0	$3 \times$	3.0	yes	yes	no
P ₈	278	9.1	10×	3.0	yes	yes	yes

在這個例子中,任何P = {p1, p2, p3, p4, p5, p6, p7, p8} 不能滿足下列一組的要求 REQ = {r1: price <= 150, r2: opt-zoom = 5x, r3: sound = yes, r4: waterproof = yes}





- 最小診斷集是一組最小的用戶需求,其放鬆(調整)將允許系統找出至少少一個可行的推薦。
- Given $P = \{p_1, p_2, \ldots, p_n\}$. and $REQ = \{r_1, r_2, \ldots, r_m\}$ where the recommendation set is empty. A diagnosis is a minimal set of elements $\{r_1, r_2, \ldots, r_k\} = d \subseteq REQ$ that have to be repaired in order to restore consistency with the given product assortment so at least one solution can be found.
- 在我們的例子中, $CS_1 = \{r_1, r_2\}$, $CS_2 = \{r_2, r_3\}$, and $CS_3 = \{r_1, r_4\}$ 是最小的診斷集





- MinRelax 算法用於找出最小診斷的完整集合。
- This algorithm calculates, for each item p_i ∈ P and the requirements in REQ, a corresponding product-specific relaxation PSX. PSX is a minimal diagnosis d ∈ D (the set of all minimal diagnoses) if there is no set r such that r ⊂ PSX.
- 在我們的例子中, $D = \{\{r_2, r_3\}, \{r_1, r_4\}, \{r_1, r_2\}\}$

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8
r_1 : <i>price</i> <= 150	1	0	0	0	0	0	0	0
r_2 : opt-zoom = $5x$	0	1	0	0	0	0	0	0
r_3 : sound = yes	0	1	1	0	1	1	1	1
r_4 : waterproof = yes	1	0	0	1	0	0	0	1



基於案例的推薦系統

- 在基於案例的推薦系統中,我們用相似性函數來找出與指定目標(或案例)相似的例子。例如,用戶可以指定目標案例或地點,臥室的數量和期望的價格來作為指定目標的一組屬性。
- 與基於限制的系統不同,基於案例的推薦系統對這些屬性沒有強制限制 (例如最小值或最大值)。
- 相似性函數用於找出與用戶指定的目標最相似的例子。如果沒有找到完全匹配用戶指定的目標,則使用相似度函數來找出與用戶需求盡可能相似的商品。
- 在相似的商品被排列與呈現之後,可以使用互動評估方法(Critiquing)來確 定最終的結果。

一個基於案例的推薦系統中用戶界面的示例



EXAMPLE OF HYPOTHE INTERFACE FOR HOME			N Company
NUMBER OF BR PRICE RANGE	NUMBER OF BATH ZIP CODE SUBMIT SEARCH	HOME STYLE	FEATURES:
I WOULD LIKE TO BUY AN H	OUSE JUST LIKE THE ONE AT MOHEGAN LAKE	T THE FOLLOWING	ADDRESS:

Source: Recommender
Systems – The Textbook





- 相似性指標: 必須將指定目標各種屬性的重要性恰當地納入相似性函數中, 以使系統有效地工作。
- 評估方法(Critiquing Methods): 可以使用各種不同的評估方法 來支持不同特點的探勘目標。

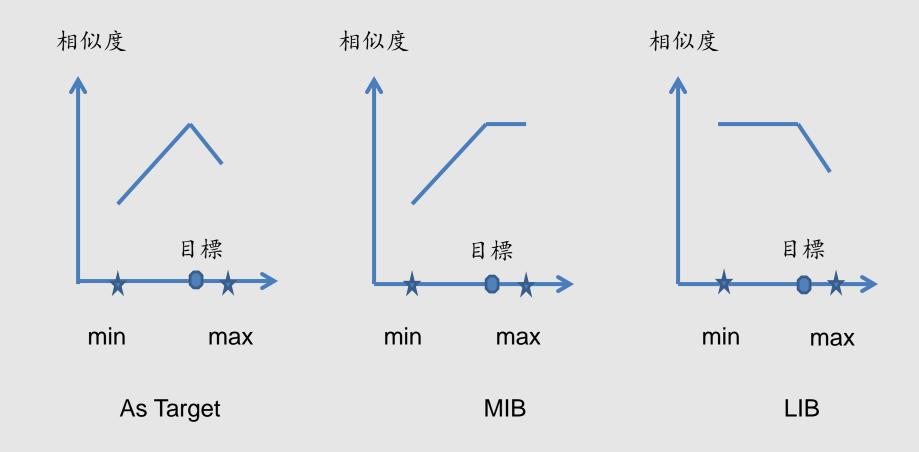
基於案例的推薦系統



- 使用目標案例其他商品之間的相似度來找出商品
- 使用距離相似度 $similarity(p, REQ) = \frac{\sum_{r \in REQ} w_r * sim(p, r)}{\sum_{r \in REQ} w_r}$
- 定義
 - sim (p, r)是目標案例和其他商品之間屬性r的距離
 - w_r 是用户要求屬性r的重要性權重
- 在現實世界中, 客戶想要
 - 最大化某些屬性。即相機的解析度,"越多越好"(MIB-More Is Better)
 - 最小化某些屬性。即相機的價格,"越少越好"(LIB-Less Is Better)



計算相似度





計算相似度

· 在MIB屬性的情況下, p和r之間的局部相似度計算如下:

Sim(p, r) =
$$\frac{\emptyset_r(p) - \min(r)}{\max(r) - \min(r)}$$
 or

$$Sim(p, r) = 1 - \frac{|\emptyset_r(p) - r|}{\max(r) - \min(r)} + \alpha \cdot I(\emptyset_r(p) > r) \cdot \frac{|\emptyset_r(p) - r|}{\max(r) - \min(r)}$$

· 在LIB屬性的情況下, p和r之間的局部相似性計算如下:

Sim(p, r) =
$$\frac{max(r) - \emptyset_r(p)}{max(r) - min(r)}$$
 or

$$Sim(p, r) = 1 - \frac{|\emptyset_r(p) - r|}{\max(r) - \min(r)} + \alpha \cdot I(\emptyset_r(p) < r) \cdot \frac{|\emptyset_r(p) - r|}{\max(r) - \min(r)}$$

• 在某些情况下,相似性應完全基於與最初定義的要求的距離:

Sim(p, r) = 1 -
$$\frac{|\emptyset_r(p) - r|}{\max(r) - \min(r)}$$

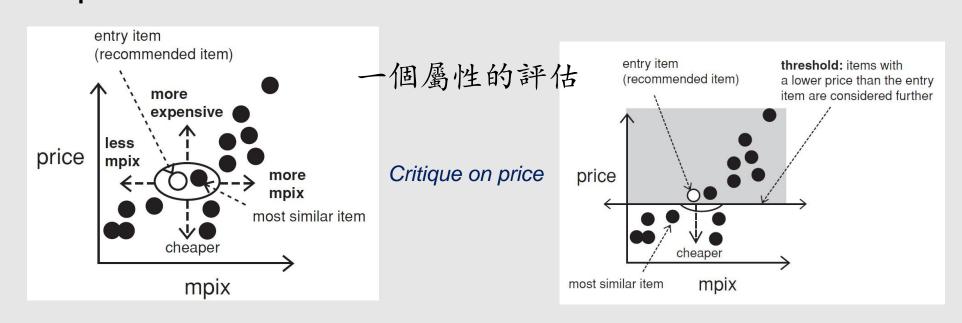
Where $\emptyset_r(p)$ is the *item attribute value*.

For example, $\emptyset_{mpix}(p1) = 8.0$

與基於案例的推薦系統互動



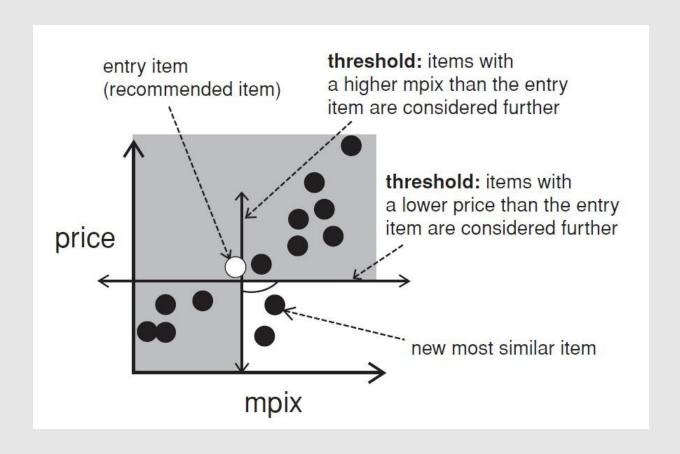
- 顧客可能不知道他們在尋找什麼
- 評估是支持這種導航的有效方式
- · 客戶指定其當前商品(輸入商品)不滿足的變更請求(價格或mpix)





多屬性評估

• 操作多個屬性的評估可以提高推薦對話的效率





EXAMPLE OF HYPOTHETICAL CASE-BASED RECOMMENDATION INTERFACE FOR HOME BUYING (critique-example.com)

[SIMPLE CRITIQUING INTERFACE]

YOU SPECIFIED THE FOLLOWING TARGET:

812 SCENIC DRIVE, MOHEGAN LAKE, NY

YOUR TOP RECOMMENDATION IS:

742 SCENIC DRIVE, MOHEGAN LAKE, NY

WE RECOMMEND THIS HOUSE BECAUSE: IT HAS SIMILAR BEDROOMS, BATHROOMS, LOCALITY, PRICE RANGE, AND HOME STYLE AS YOUR TARGET

I WOULD LIKE TO BUY A HOUSE SIMILAR TO THE TOP RECOMMENDATION BUT WITH ONE OF THE FOLLOWING CHANGES:

NUMBER OF BR	SUBMIT CHANGE	NUMBER OF BATH	SUBMIT CHANGE
PRICE RANGE	SUBMIT CHANGE	ZIP CODE	SUBMIT CHANGE
HOME STYLE	SUBMIT CHANGE	SEE OTHER RESULTS	GO BACK TO ENTRY POINT

(a) Simple critiquing by directly modifying feature values

EXAMPLE OF HYPOTHETICAL CASE-BASED RECOMMENDATION INTERFACE FOR HOME BUYING (critique-example.com)

[SIMPLE CRITIQUING INTERFACE]

YOU SPECIFIED THE FOLLOWING TARGET:

812 SCENIC DRIVE, MOHEGAN LAKE, NY

YOUR TOP RECOMMENDATION IS: 742 SCENIC DRIVE, MOHEGAN LAKE, NY

WE RECOMMEND THIS HOUSE BECAUSE: IT HAS SIMILAR BEDROOMS, BATHROOMS, LOCALITY, PRICE RANGE, AND HOME STYLE AS YOUR TARGET

I WOULD LIKE TO BUY A HOUSE SIMILAR TO THE TOP RECOMMENDATION
BUT WITH ONE OF THE FOLLOWING CHANGES:

UMBER OF BR	MORE	LESS	NUMBER OF BATH	MORE	LESS	
RICE RANGE	MORE	LESS	EXP	LORE NEAR	BY ZIP CODES	
	EXPLORE RE	LATED STYLES	SEE OI	THER	GO BACK T	

(b) The conversational style of directional critiques

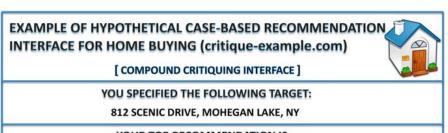
RESULTS

ENTRY POINT



Source: Recommender Systems – The Textbook





YOUR TOP RECOMMENDATION IS: 742 SCENIC DRIVE, MOHEGAN LAKE, NY WE RECOMMEND THIS HOUSE BECAUSE: IT HAS SIMILAR BEDROOMS, BATHROOMS, LOCALITY, PRICE RANGE, AND HOME STYLE AS YOUR TARGET I WOULD LIKE TO BUY A HOUSE SIMILAR TO THE TOP RECOMMENDATION BUT WITH ONE OR MORE OF THE FOLLOWING CHANGES: NUMBER OF BATH NUMBER OF BR **HOME STYLE** PRICE RANGE ZIP CODE **SEE OTHER GO BACK TO** SUBMIT SEARCH RESULTS **ENTRY POINT**

(a) Compound critiquing by modifying multiple feature values

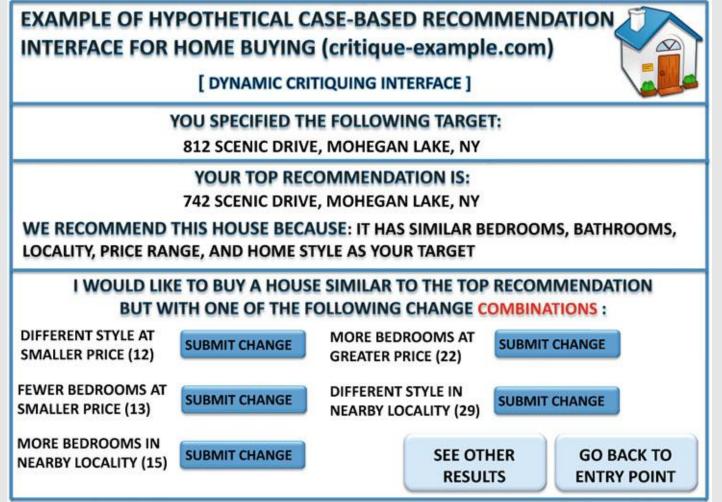




Source: Recommender Systems - The Textbook

用於在基於案例的推薦系統中進行動態評估的用戶界面





Source: Recommender Systems – The Textbook



評估中的解釋

- 建議在評估過程中提供解釋,因為它有助於用戶更好地理解訊息空間和各種商品屬性/功能之間的權衡。
- 基於評估的系統其主要危險是用戶漫無目的的在推薦系統中蜿蜒而未能夠成功地找到他們想要的東西。添加解釋大大降低了這種可能性。



總結



基於知識的推薦系統

- ◆ 基於限制的推薦系統
- ◆ 基於案例的推薦系統



限制性

- ◆ 領域知識獲取的成本
 - 來自領域專家
 - 來自用戶
 - 來自網絡資源
- ◆ 模型的準確性
 - 非常精準的偏好模型需要許多用戶交 互週期(User Interaction Cycle)
- ◆ 獨立性假設可以被挑戰
 - 偏好並不總是彼此獨立





推薦閱讀:

基於效用理論的壽險保單多準則推薦系統

- Multi-criteria Recommender System for Life Insurance Plans based on Utility Theory by Akshay Hinduja and Manju Pandey
- http://www.indjst.org/index.php/indjst/article/viewFile/111376/79552

先進的推薦系統

- □情境敏感的推薦系統
 - (Context-Sensitive Recommender Systems)
- □社群和以信任為中心的推薦系統
 - (Social and Trust-Centric Recommender Systems)
- □防攻擊推薦系統
 - (Attack-Resistant Recommender Systems)
- □群組推薦系統
 - (Group Recommender Systems)
- □計算廣告
 - (Computational Advertising)

情境敏感的推薦系統: 概述



- 情境相關的推薦系統是根據附加的情境具體情況來定制其推薦。
- 情境的一些例子是:
 - 時間: 平日, 週末, 假期, 四季等
 - 地點: 當地餐館, 電影院, 娛樂場所等
 - 社交信息,如用户的朋友,標籤,社交圈子,活動夥伴等
- 在情境敏感推薦系統中, U x I x C 映射到相應的的可能評級, U, I和C分別表示用戶, 商品和情境。
- 在一個推薦應用程序中可以使用多種類型的情境。 例如, 時間和地點一起考慮。 因此, 可能有多個情境維度。

解決方案



- · 情境預過濾: 根據相關的情境預過濾評級。只有符合情境的評級被保留為推薦系統使用。
- 情境後過濾:首先在整個全球評級集上執行推薦。然後根據相關的情境過濾。只有符合情境的推薦被保留和排列。
- 在模型中包含情境:情境信息被直接合併到預測函數中。

Group Recommender Systems 群組推薦系統





我知道瑪麗,保羅,邁克爾和珍妮的個別品味,作為一個群體我應該如何向他們推 薦什麼電視節目?

群組推薦系統



- 在群組推薦系統中, 商品被用戶群組而不是單個用戶使用; 因此, 建議必須根據群組的組成進行調整。
- 常見的應用是電影,電視,音樂和旅遊領域,這些活動通常由一群人一起出席。
- 由於用戶可以由於社交現象(如情緒傳染和一致性)而受 到其他用戶的影響,因此使用直接的平均值向群組推薦商 品項目的方法往往效果不佳。
- 因此, 能夠在建模過程中使用社會現象至關重要。

社交現象



- 情感傳染: 不同用戶的滿意度可能會相互影響。 例如,如果一群用戶正在一起看電影,而一些群組的成員不喜歡這部電影,這可能對其他用戶產生傳染效應。 因此,群組的最終體驗可能與平均評分所代表的大不相同。
- 一致性:它與情感傳染的概念密切相關,因為用戶表達的意見會相互影響。然而,它和社會傳染稍有不同,因為用戶要有意識地有同樣的意見,因為他們的意見將受同伴的行為影響而無意識地改變。

在群組環境下的基於協同過濾和基於內容系統



- 就用於建立群組推薦的方法而言,基於協同過濾和基於內容的系統通常十分相似。
- 一般方法包括以下兩個步驟:
 - 為每個用戶獨立計算出商品的預測評分。
 - -對於每個商品,透過一個預測評分的集合函數,將來自該群組的 各個成員的評分彙總為單個群組組評分。然後根據彙總的預測評 分對所有商品項目進行排名。





- 最小痛苦的策略: 用於該群組的總評分是該組中任何成員的最低評分。 其基本思想是防止社會傳染和一致性的負面影響。
- 加權平均值: 用於該群組的總評分是該組成員評分的加權平均值。成員的加權重通常用於模擬特定類型的情況, 以防止極度不喜歡或不可行。
- 排除痛苦的平均值:用於該群組的總評分是在排除具有最低評分的個人的評分之後評分的平均值。基本的想法是最大限度地享受愉快的體驗,因為愉快的體驗也和不愉快的體驗相同,可能會感染。

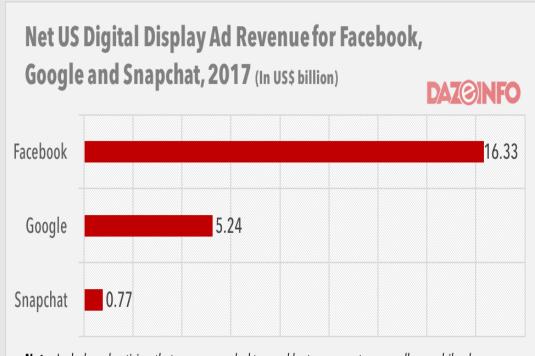
計算廣告



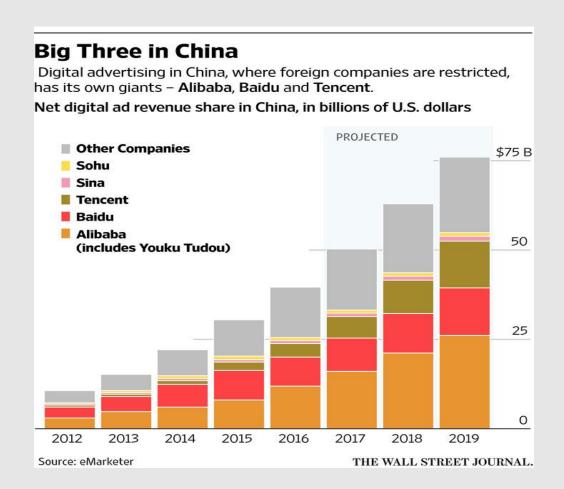
- 計算廣告是最近提出的一個術語,它使用各種計算方法來做內容有針對性的廣告
- · 計算廣告所面對的核心問題是: 定位最適合情境的廣告。情境 涉及內容(查詢,網頁內容),用戶訊息和位置訊息。因此, 來自推薦系統的許多想法直接用於計算廣告。
- 基於內容的定位實例包括贊助搜索和基於情境的廣告。在這兩種情況下,都會在特定情境下向用戶推薦廣告(類似於某個商品)。
- 放置內容相關的廣告具有雙重優勢。首先,用戶對該主題的直接興趣可以被利用,這反過來又增加了用戶探索廣告的機會。 其次,它帶來更好的用戶體驗。



廣告收入的機會



Note: Includes advertising that appears on desktop and laptop computers as well as mobile phones, tablets and other internet-connected devices. net ad revenue afetr companies pay traffic acquisition cost (TAC) to partner sites; includes banner ads and other. **Source:** eMarketer, March 2017



線上廣告



- Ad Network
- Ad Exchange

