

はじめに
網羅性
提案手法
評価実験
おわりに

ユーザの網羅性を反映したランキング手法 の提案

滝沢 光介

富山県立大学 電子・情報工学科

July 9, 2021

はじめに

2/14

ランキング手法の現状

現状のランキング手法では、ユーザーの動的な興味や人気が反映されにくいといった問題がある。

本研究の目的

網羅性という観点を考慮してランキング手法に協調フィルタリングを組み込ませることにより動的なランキング手法を提案し、ユーザのコンテンツ検索を効率化を図る。

提案手法

提案手法は協調フィルタリングを基にしたランキング手法であり、さらにユーザの網羅性を加えることで、ランキング手法の検索の効率化を図る。

ユーザの網羅性の定式化

3/14

網羅性

網羅性を全ユーザの検索履歴集合に対する、対象ユーザの検索履歴の情報損失度として表し、ユーザ u の網羅性を(1)式で定式化する。

$$\begin{aligned} w(u) &= w(h_u) = p(h_u | H) \propto p(H | h_u) p(h_u) \\ &= \prod_{d \in D} p(d | h_u)^{N_d} p(h_u) \end{aligned} \tag{1}$$

h_u : ユーザ u の検索履歴

H : 全ユーザの履歴集合 b

D : H に出現するコンテンツ集合

N_d : H における d の出現率

$P(h_u | H)$: h_u の H での出現率

はじめに

網羅性

提案手法

評価実験

おわりに

ユーザの網羅性の定式化(2)

4/14

はじめに

網羅性

提案手法

評価実験

おわりに

$p(h_u)$ と $p(d|h_u)$ を次のように推定する.

$$p(h_u) = \frac{1}{\text{全ユーザ数}} \quad p(d|h_u) = \frac{0.5 + n_{ud}}{0.5 * |D_u| + N_u} \quad (2)$$

n_{ud} : h_u における d の出現頻度

$|D_u|$: h_u におけるコンテンツの異なり数

N_u : h_u におけるコンテンツの出現頻度の合計

のことより、多くのユーザにアクセスされているコンテンツにより多くアクセスしているユーザは網羅性が高いと言える。この網羅性が高いユーザから協調フィルタリングへ使用する。

ユーザの網羅性の定式化(3)

5/14

はじめに
網羅性
提案手法
評価実験
おわりに

(1) 式は(3),(4)式と変形させることにより, KL ダイバージェンスに基づく情報損失に対応させることができ,(4)式を最大化させるような h_u は損失が少ないといえ, ユーザ u は網羅性が高いと言える.

$$w'(u) \equiv \frac{1}{|H|} \left(\log p(h_u) + \sum_{d \in H} N_d \log p(d|h_u) \right) \quad (3)$$

$$= \frac{1}{|H|} \log p(h_u) + \sum_{d \in H} p(d|H) \log p(d|h_u)$$

$$w'(u) = \frac{1}{|H|} \log p(h_u) + \sum_{d \in H} p(d|H) \log \frac{p(d|h_u)}{p(d|H)} \quad (4)$$

$$\equiv D(H|h_u) = D(U|u)$$

ランキングモデルの提案

6/14

提案モデル

図1に示すように、テキストフィルタリング、情報検索、協調フィルタリングはある確率を最大化する対象をユーザに提示しており、事後確率最大化問題となっている。

テキストフィルタリング:

$$\operatorname{argmax}_{c_k} p(c_k|d_j) \propto \operatorname{argmax}_{c_k} p(d_j|c_k)p(c_k) \quad (a)$$

c_k : カテゴリ d_j : テキスト

情報検索

$$\operatorname{argmax}_{d_j} p(d_j|q) \propto \operatorname{argmax}_{d_j} p(q|d_j)p(d_j) \quad (b)$$

d_j : テキスト q : クエリ

協調フィルタリング

$$\operatorname{argmax}_{d_j} p(d_j|h_u) \propto \operatorname{argmax}_{d_j} p(h_u|d_j)p(d_j) \quad (c)$$

d_j : コンテンツ h_u : ユーザの履歴

図1: テキストフィルタリング、情報検索、協調フィルタリングの関係性

提案システムの定式化(1)

7/14

定式化

テキストフィルタリング用いられる, *Naive Bayes* を協調フィルタリングに適用し, ランキングモデルを (5) 式で定義する.

$$\begin{aligned} \operatorname{argmax}_{d_i} P(d_i | \mu) &= \operatorname{argmax}_{d_i} P(d_i | h_{\mu}) \times \operatorname{argmax}_{d_i} P(h_{\mu} | d_i) P(d_i) \\ &= \operatorname{argmin}_{d_i} \left(\sum_{d_j \in h_{\mu}} n(d_j) \log P(d_j | d_i) + \log P(d_i) \right) \end{aligned} \quad (5)$$

提案手法の定式化(2)

8/14

式変形

網羅性を反映させるために,(5) 式にスムージングパラメータ λ を用いて (6) 式に変形する.

$$\begin{aligned}
 & \underset{d_i}{\operatorname{argmax}} P(d_i | u) \\
 &= \underset{d_i}{\operatorname{argmin}} \left(\left(\sum_{d_j \in h_L} n(d_j) (\lambda * \log P_s(d_j | d_i) + (1 - \lambda) * \log P(d_j)) \right) + \log P(d_i) \right) \\
 & \text{where } P_s(d_j | d_i) = \frac{|U_{sij}|}{|U_{si}|} \quad (6)
 \end{aligned}$$

λ の役割

λ は 0 から 1 の間の数字を表し, 網羅性の高いユーザの重みを決定している。 λ はユーザが調整可能であり, 自分の網羅性にあったフィルタリングを受けることが可能である.

実験概要

9/14

使用データセット

- *EachMovie* データセット

映画 1628 作品の 72916 人における評価. 評価は 0 から 1 の間の 0.2 間隔の 6 段階評価.

評価方法

各手法におけるリコメンデーションを *modifiedprecision* で評価する.

実験方法

ユーザ毎に評価が 0 でない映画を任意に抽出し, データの 25% をテストデータとし, 残りを学習データとして分け, これを 10 回行い, その 10 組のデータに対して実験を行う.

はじめに

網羅性

提案手法

評価実験

おわりに

評価実験(1)

10/14

ユーザの網羅性を考慮したランキング

提案するランキングモデルにおけるスムージングの効果を確かめるために、網羅性の高いユーザではなく全ユーザの履歴を用いて $top - N$ リコメンデーションの N とパラメータ λ を変化させ、*modifiedprecision* の観測を行った。

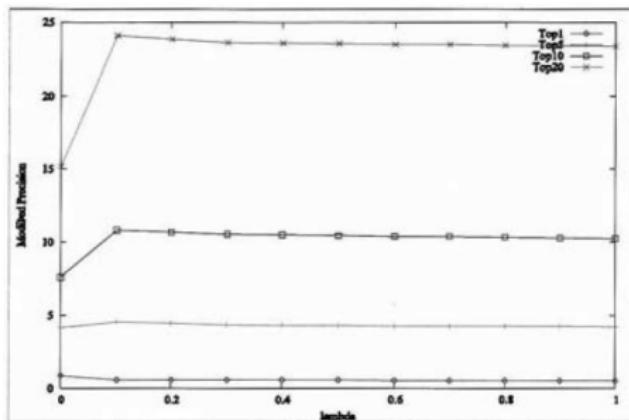


図 2: スムージングによる $top - N$ のサイズ毎の *modifiedprecision*

評価実験 (2)

11/14

ユーザの網羅性を考慮したランキング

*top – 20 リコメンデーション*を用いて行い、網羅性の高いユーザの履歴集合を用いた時と、全ユーザの検索履歴を用いた時の *modifidprecision* を比較する。

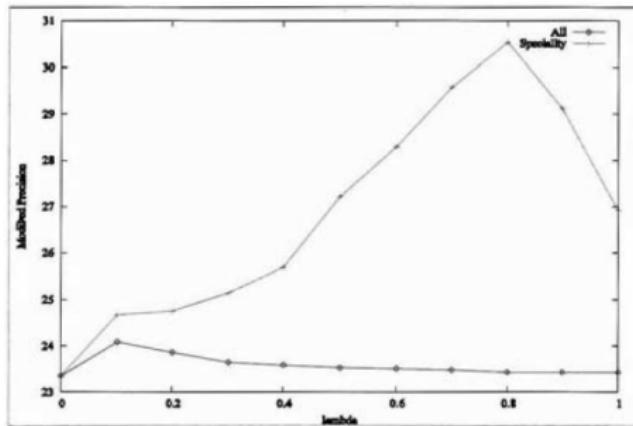


図 3: スムージング効果とユーザの網羅性の有無による *modifidprecision* の変化

評価実験(3)

12/14

映画の人気による変化

評価数の多い映画と少ない映画を選択し、それぞれの映画集合に対する *modifiedprecision* を調べた。

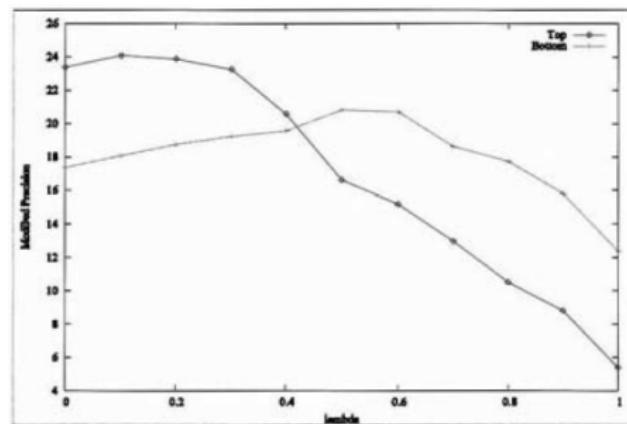


図 4: スムージング効果と映画の人気の有無による *modifiedprecision* の変化

評価実験(4)

従来手法との比較

従来手法との *modifiedprecision* の差を調べた。

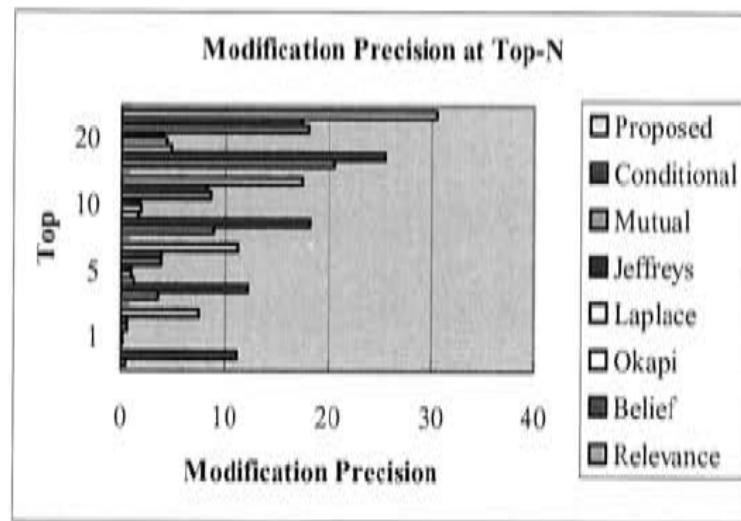


図 5: 従来手法との *modifiedprecision* の比較

はじめに

網羅性

提案手法

評価実験

おわりに

実装したこと

- ① 予測精度の高いランキングを実装した
- ② 出力においてユーザのフィードバックを可能とした

結論

ユーザの網羅性を考慮し,*Naive Bayes*に基づいた協調フィルタリングをすることで予測精度の向上に繋がった.