



情報中立推薦システム

神畠 敏弘^{*1}, 赤穂 昭太郎^{*1}, 麻生 英樹^{*1}, 佐久間 淳^{*2}

^{*1}産業技術総合研究所, ^{*2}筑波大学 / 科学技術振興機構

2012年度人工知能学会全国大会 (第26回) @ 山口市, 2012.6.12-15

<http://www.kamishima.net/>



概要

フィルターバブル問題

個人化技術の普及により生じたと、Eli Pariser が主張する問題

- ✳ 関心のある話題のみに集中し，多様な話題を知る機会を失う
- ✳ 関心のない話題は，どんなに重大であっても除外される



情報中立推薦システム

みにくいアヒルの子の定理などから，絶対的に中立な推薦は不可能



利用者などが指定した，特定の視点に対して，推薦結果ができるだけ中立性を保つように配慮する推薦システム

目次

フィルターバブル問題

- ✳ FacebookとGoogle検索の例
- ✳ Pariserの主張とRecSys2011での議論

推薦における情報中立性

- ✳ みにくいあひるの子の定理, 特定の視点に対する中立性

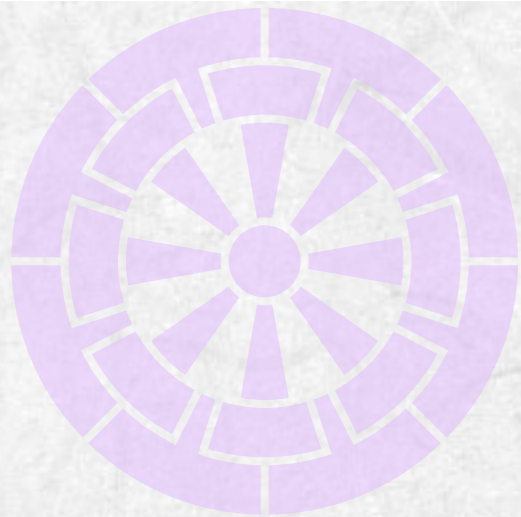
情報中立推薦システム

- ✳ 潜在因子モデル, 情報中立性強化のための拡張

実験

- ✳ 予測精度と情報中立性のトレードオフ

まとめ



フィルターバブル問題



フィルターバブル問題

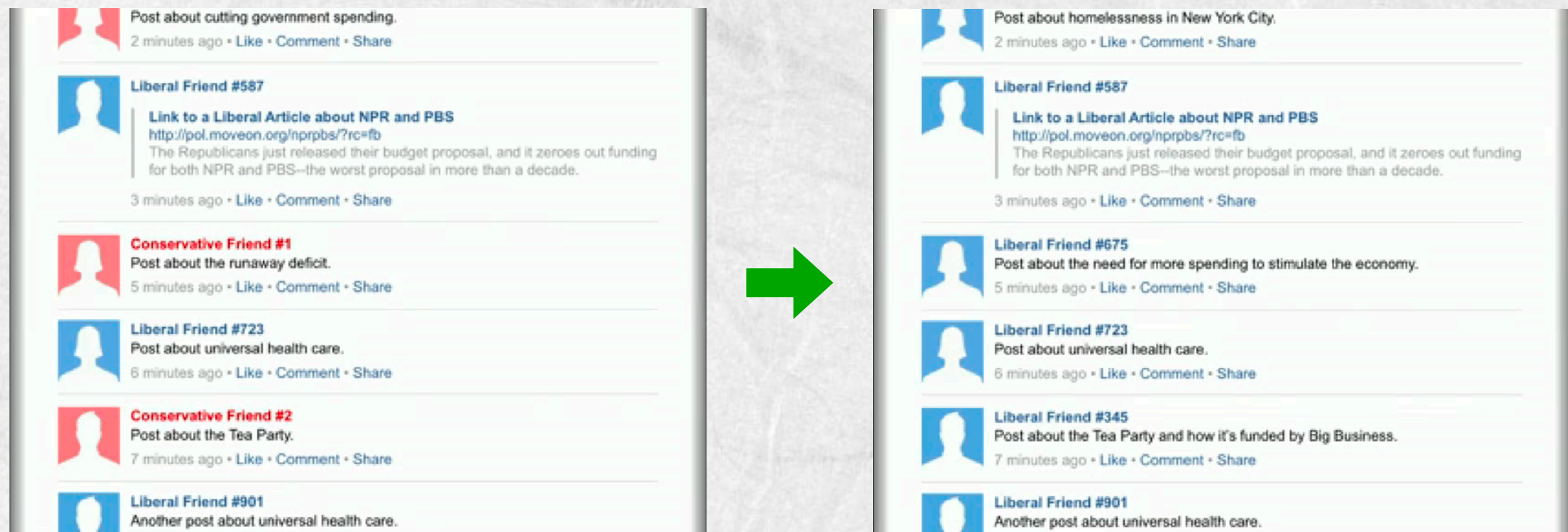
[TED Talk by Eli Pariser]

フィルターバブル問題

個人化技術の普及により，ネット利用者が多様な情報源や視点に接する機会が減らされているとの Eli Pariser による主張

<http://www.thefilterbubble.com/>

進歩派の友人を増やしていると，私に断りなく
FaceBook の友人推薦リストから保守派が消されてしまった



Googleで，重大事件でも関心の対象外なら，検索されない事例も

フィルターバブルの議論

[RecSys 2011 Panel on the Filter Bubble]

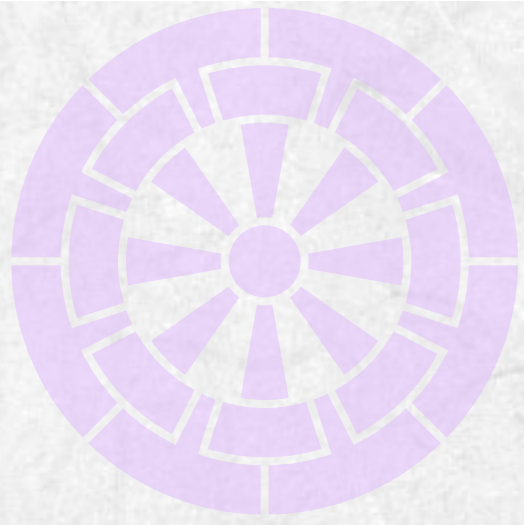
Pariser の主張

- ✳ 個人化技術により，利用者が多様な経験をする機会が減少
- ✳ 社会での情報共有が困難になり，合意形成が困難になる

RecSys 2011のパネルでの議論

- ✳ 本質的なトレードオフである
 - ✳ あらゆる観点に対して**絶対的に中立的な選択は存在しない**
 - ✳ 全ての情報を網羅することはできず，**フィルタリング技術は必須**
- ✳ この問題に対して，推薦システムはどう改善すべきか？
 - ✳ どういう基準で選択しているのかという透明性・説明を確保
 - ✳ 利用者が現在求めるもののだけでなく，長期的多様性も考慮
 - ✳ アイテムごとの個別判断ではなく，推薦リスト全体の良さを追求

<http://acmrecsys.wordpress.com/2011/10/25/panel-on-the-filter-bubble/>



推薦における中立性



みにくいアヒルの子の定理

[Watanabe 69]

みにくいアヒルの子と普通のアヒルの子とを分類する

可能な分類ルールのうち、二匹を区別できるルール数で類似度を測る
分類するのに十分の数の二値特徴があり、それらの特徴のブール式で
アヒルを識別する



みにくいアヒルと普通のアヒルを区別できるルールと
任意の二匹のアヒルを区別できるルールの数は同じ

n 匹のアヒルがいるとき、任意の二匹のアヒルを識別するブール式の
数は、アヒルの選び方によらず常に 2^{n-2} 個となる



みにくいアヒルの子と普通のアヒルの子を区別できない！

みにくいアヒルの子も、普通のアヒルの子も互いに類似度は同じ

みにくいアヒルの子の定理

[Watanabe 69]

しかしみにくいアヒルの子はみにくい
なぜそんなことが起きるのか？



分類ルールの数だけで、ルールの内容を考慮していない

- ✳ ルールに含まれる特徴の違いを考慮しない
(体全全体の色が違ってても、ちょっと曲がった羽根があるだけでも
同じ扱い)
- ✳ ルールに含まれる項の数など、その複雑さを考慮しない



分類をするためには、対象の特定の側面や、特定の分類ルールを重視
することを必然的に伴う

情報中立な推薦

みにくいあひるの子の定理

対象のある側面を重視しないと分類できない



絶対的に中立な推薦は本質的に不可能

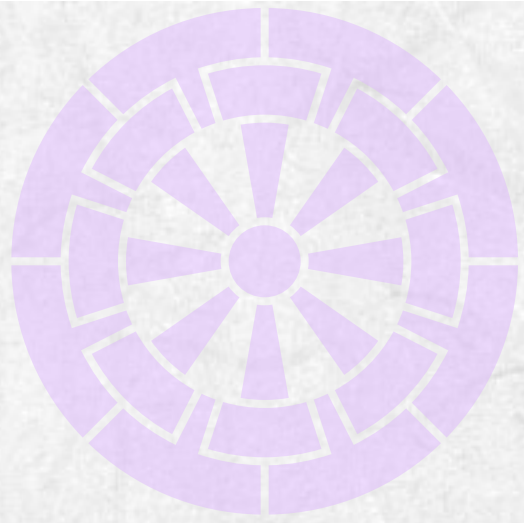
関心のあるものとなないものに分類することが推薦

情報中立な推薦

ある特定の観点に対する中立性を強化した推薦

特定の観点以外の特徴は扱いに差があってもよい

例：Pariser の友人推薦の例で、推薦される友人が保守・革新のどちらかの中立性は確保するが、居住地や年齢などの偏りは許す



情報中立推薦システム



潜在因子モデル

[Koren 08]

評価値予測タスク：利用者がアイテムにつけるであろう評価値を予測

潜在因子モデル (latent factor model)：行列分解型の推薦モデルの基本形で、多くの拡張手法の元になっている手法

利用者 x のアイテム y についての予測評価値

全体バイアス 利用者依存バイアス 利用者・アイテムの
交差効果

$$\hat{s}(x, y) = \mu + b_x + c_y + \mathbf{p}_x \mathbf{q}_y^\top$$

利用者依存バイアス

利用者 x_i , アイテム y_i , 評価値 s_i を集めた訓練事例集合に対して、 L_2 正則化項付きの二乗誤差損失を最小化してパラメータを求める

情報中立推薦システム

視点変数 v ：利用者などが指定する視点を表す変数

※ 予稿中では目的変数 t と呼んでいるが，クラス変数との誤解をさけるため変更

- ★ 視点変数の値に推薦結果が影響されないことが中立性
- ★ ここでは 0/1 の二値変数に限定
- ★ 利用者とアイテムの片方か，両方に依存して確定的に値が決まる

例：利用者の性別，アイテムの製造年，アイテムの評価時刻

潜在変数モデルの変更

推薦結果の補正ができるように，視点変数への依存性を導入

$$\hat{s}(x, y, v) = \mu^{(v)} + b_x^{(v)} + c_y^{(v)} + \mathbf{p}_x^{(v)} \mathbf{q}_y^{(v)\top}$$

視点変数の各値ごとに潜在変数モデルを作り，視点変数の値に応じてモデルを選択する

情報中立推薦システム

中立性関数 $\text{neutral}(\hat{s}, v)$: 推薦の中立性を評価する関数

- 予測評価値と視点変数の値の両方に依存
- 推薦の中立性が高いほど，すなわち視点変数の予測評価値への影響が小さいほど大きな値を出力する

潜在因子モデルを拡張した情報中立推薦モデルの目的関数

中立性パラメータ
中立性と予測精度のバランスを調整

正則化
パラメータ

$$\sum_D (s_i - \hat{s}(x_i, y_i, v_i))^2 - \eta \text{neutral}(\hat{s}(x_i, y_i, v_i), v_i) + \lambda \|\Theta\|_2^2$$

二乗損失関数

中立性関数

L₂正則化項

この目的関数を最小化するようにパラメータを学習

情報中立推薦システム

推薦の中立性：視点変数の値に推薦結果が影響されないこと



中立性関数 = 負の相互情報量

中立性を統計的独立性とみなし予測評価値と視点変数の相互情報量を中立性関数として採用

相互情報量の計算には $\Pr[\hat{s}|v]$ の計算が必要



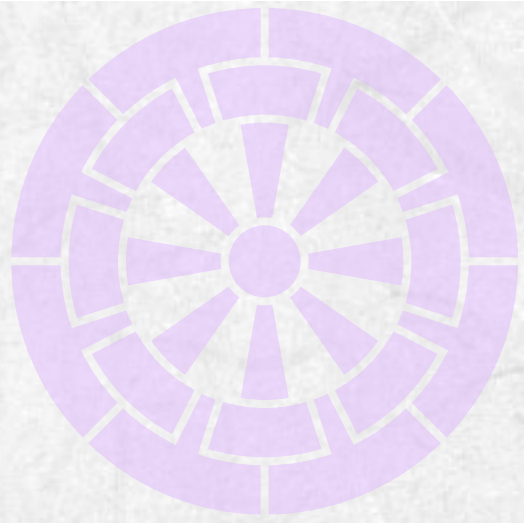
分布にヒストグラムモデルを導入



目的関数の勾配を解析的に計算できなくなった



勾配を使わずに実行できる Powell法 で目的関数を最小化



実験結果

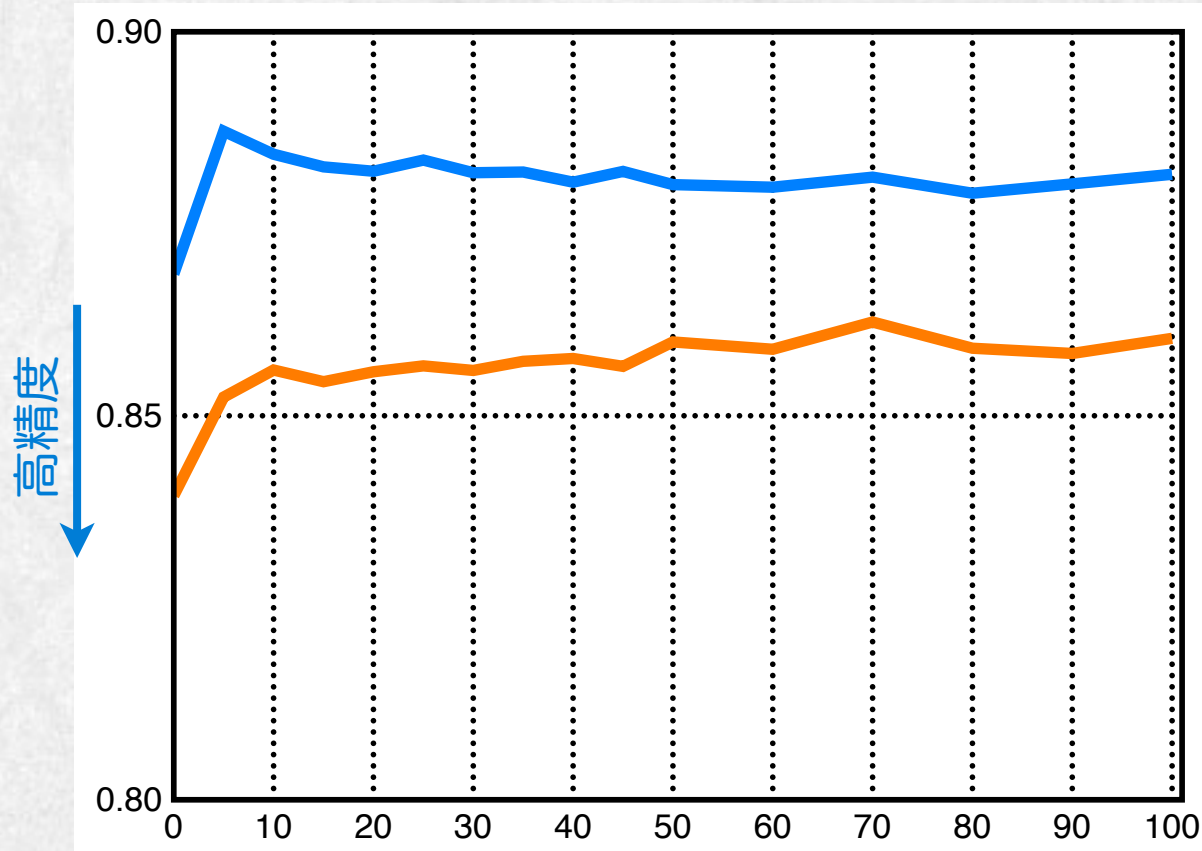


実験条件

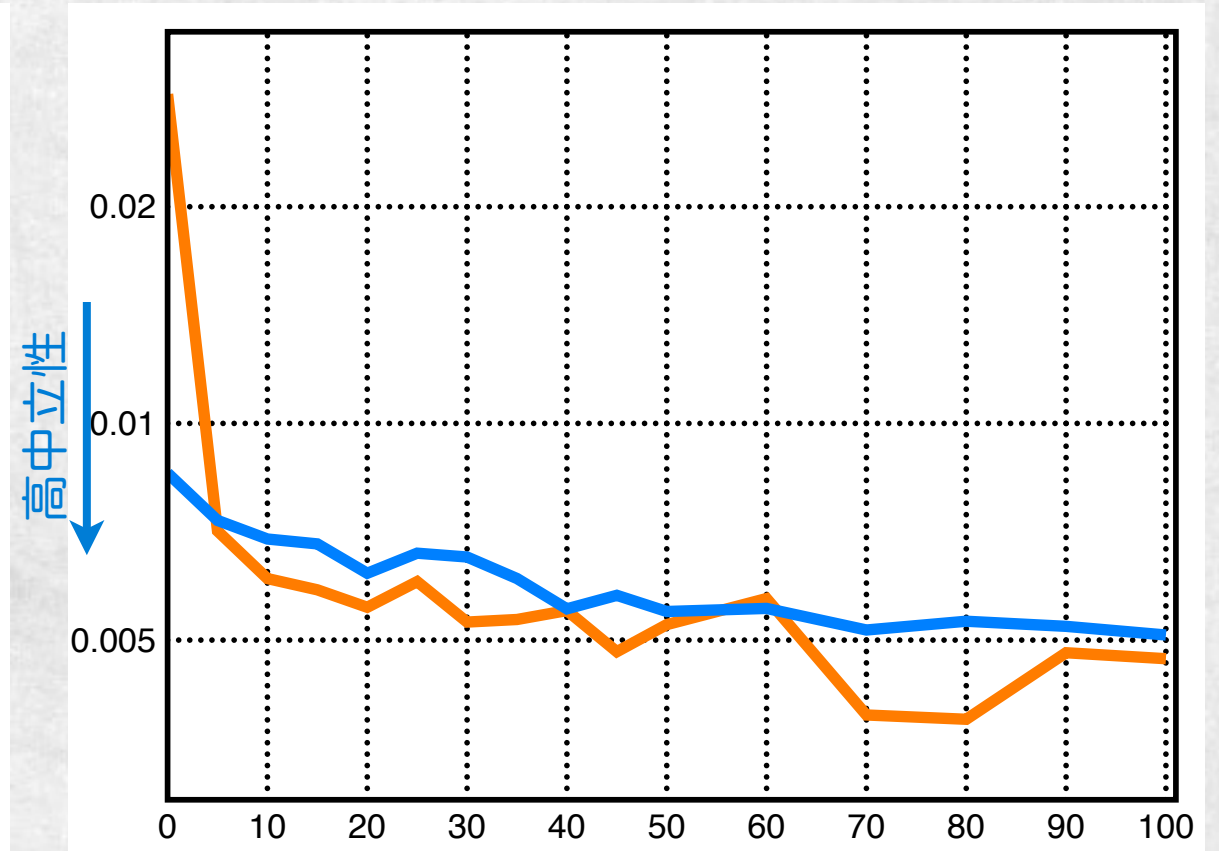
- ★ Movielens 100k データのうち、9409個の評価を抽出して実験
 - ★ Powell法は、目的関数の評価回数が多く大規模データの処理はできなかったため
 - ★ 潜在因子数 $K=1$ ，正則化パラメータ $\lambda = 0.01$
 - ★ 5分割交差確認
 - ★ 実験に使った視点変数
 - ★ **Year**：映画の公開年が1990年以降かどうか
 - ★ **Gender**：評価者の性別
 - ★ 評価尺度
 - ★ 予測精度：**MAE**（平均絶対誤差）
 - ★ 中立性：**正規化MI**（予測評価値と視点変数の正規化相互情報量）
- ※ 最適化の終了条件などの条件の変更のため予稿と実験結果は異なる

実験結果

予測精度 (MAE)



中立性 (正規化相互情報量)



中立性パラメータ η : 中立性重視

視点変数 : **Year**=映画の公開年, **Gender**=利用者の性別

- * η を増やして中立性を重視すると, 予測精度は若干上昇し, 中立性は急速に減少している → 効率よく中立性を確保できている
- * 視点変数の違いによって効果は異なる

まとめ

本発表の寄与

- ✳ みにくいアヒルの子の定理に基づく考察に基づき，推薦における中立性を定式化
- ✳ 推薦の中立性を強化する情報中立推薦システムを開発
- ✳ 提案アルゴリズムが予測精度をそれほど低下させることなく，中立性を強化できることを実験的に示す

今後の予定

- ✳ 現在の情報中立推薦システムにはスケーラビリティの問題
 - ✳ 解析的に勾配を計算できる目的関数の定式化が必要
- ✳ 独立成分分析で使われている尖度などの他の独立性指標の採用
- ✳ pLSI / LDA など生成系の推薦モデルの情報中立化