



# 情報中立推薦での中立性項の改良

神鳶 敏弘<sup>\*1</sup>, 赤穂 昭太郎<sup>\*1</sup>, 麻生 英樹<sup>\*1</sup>, 佐久間 淳<sup>\*2</sup>

<sup>\*1</sup>産業技術総合研究所, <sup>\*2</sup>筑波大学

第13回 情報論的学習理論と機械学習 (IBISML) 研究会, 2013.7.18

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

# 概要

## 推薦の中立・公平性

- \* フィルターバブル：個人化で偏った話題の情報にのみに接触
- \* システム運営者が、コンテンツ提供者を公平に扱う
- \* 法や契約により利用が禁止されている情報を排除して推薦する



## 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

## 推薦の中立性が必要な例

- \* フィルターバブル問題
- \* コンテンツ提供者の公平性

## 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

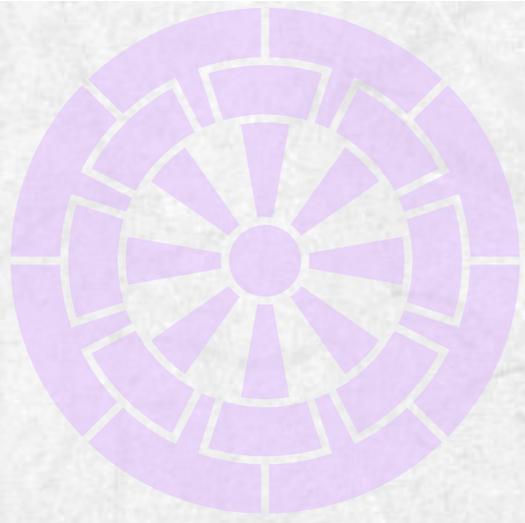
- \* 確率的行列分解子モデル, 情報中立性強化のための拡張
- \* 相互情報量とCalders&Verwerスコアに基づく中立性項

## 実験

- \* 予測精度と情報中立性のトレードオフ

## 関連研究

## まとめ



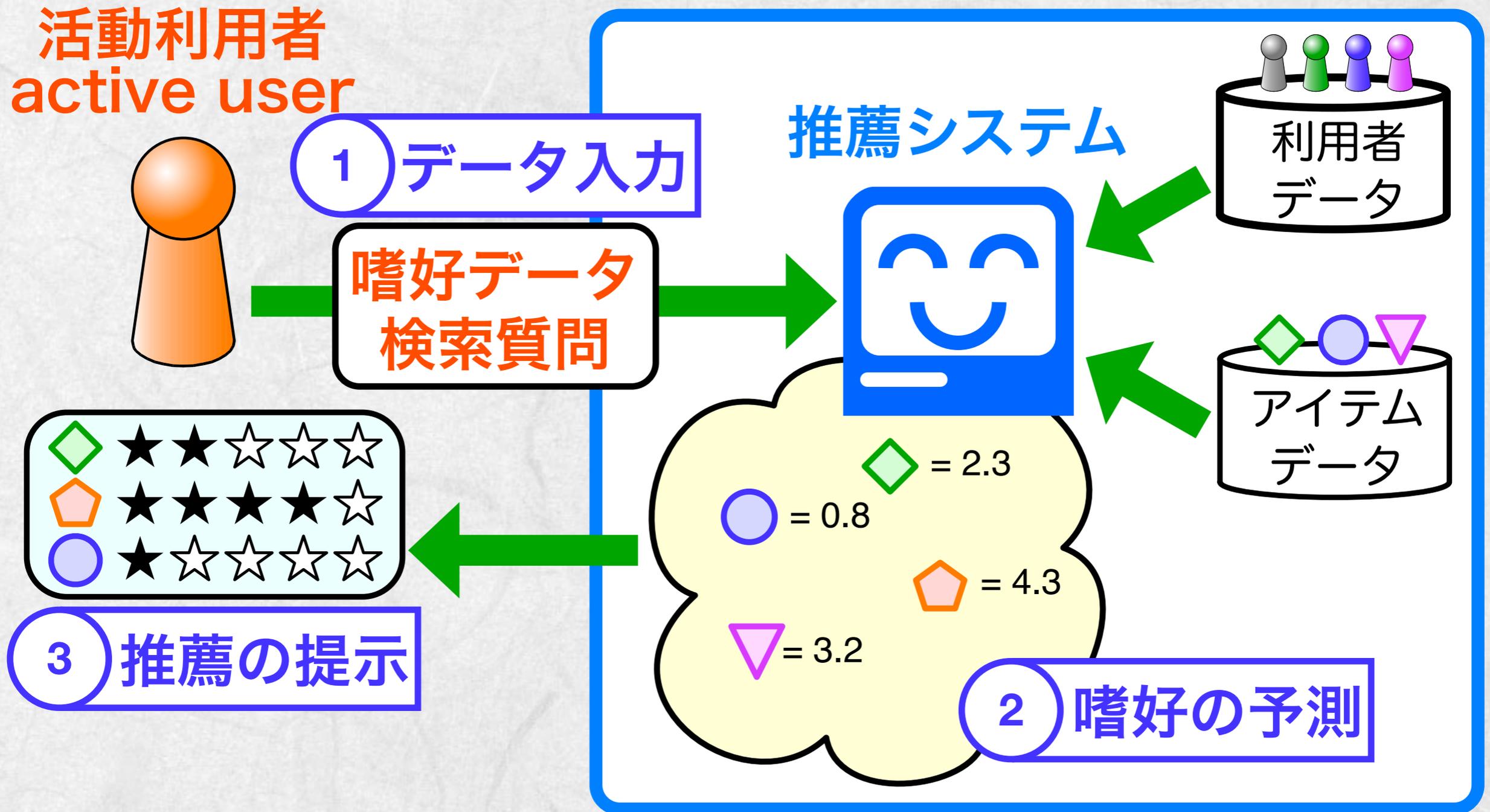
# 推薦の中立性が必要な例



# 推薦システムの実行過程

[Konstan+ 03]

## O-I-Pモデル (Output-Input-Process)



# フィルターバブル問題

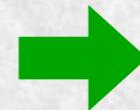
[TED Talk by Eli Pariser]

## 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



# 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のパネルでの対応策

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

- \* 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
- \* **利用者に視点を制御する手段を与える**



**利用者が指定した観点について中立性を保証した推薦**

Pariserの例：推薦される友人が保守派か進歩派かという観点については中立性を保証するが，他の点については個人化にゆだねる

# コンテンツ提供者の公平性

## 推薦システム運営者がコンテンツ提供者を公平に扱う必要性

### \* Google のランキング問題

2012年11月のBloombergの記事 (<http://bloom.bg/PPNEaS>) によれば、自身のサービスを競合企業と同種のサービスより上位に表示しているとの点についてFTCの調査を受けた

### \* コンテンツ提供者が顧客である場合

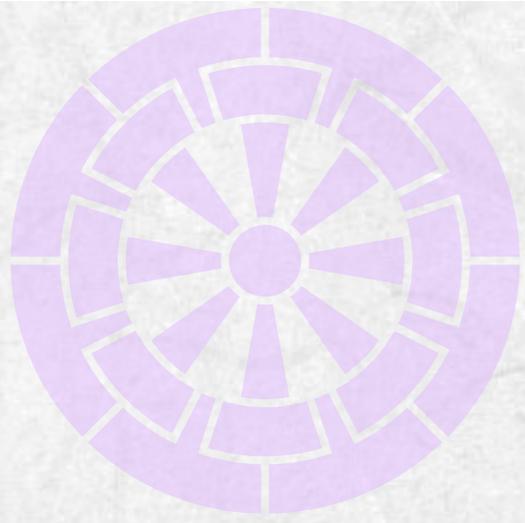
小売店をまとめたオンラインモールや、顧客が提供する情報を掲載する不動産・求職サイトでは、コンテンツ提供者の公平性に配慮が必要



現在は、無作為であることをもって公平としているが、より積極的に

## コンテンツ提供者が誰かという観点について中立性を保証した推薦

コンテンツ提供者が誰かという情報は無視して、コンテンツの他の情報にのみ基づいて推薦をする



# 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、2匹を区別できるルール数で類似度を測る  
 $2^n$ 匹のアヒルを分類するのに  $n$  個の二値特徴があり、これらの特徴を用いた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2匹のアヒルを区別できるルールの数は同じ

このとき、任意の2匹のアヒルを識別するブール式の数は、アヒルが醜いかどうかによらず常に  $2^{n-2}$  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と分類されている  
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、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：分類には特定の側面を重視する必要



**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いものに分類することが推薦

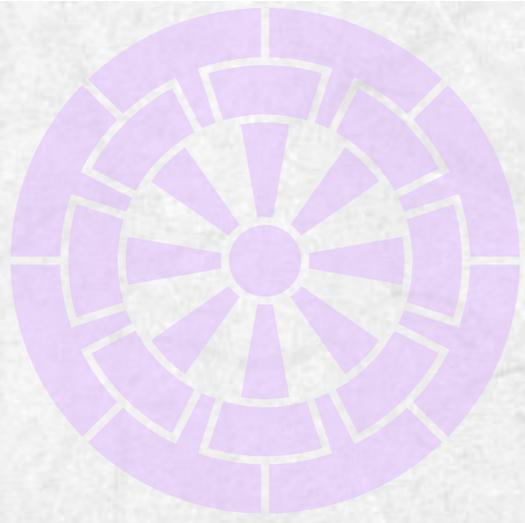
**情報中立な推薦**

**ある指定した観点に対する中立性を強化した推薦**

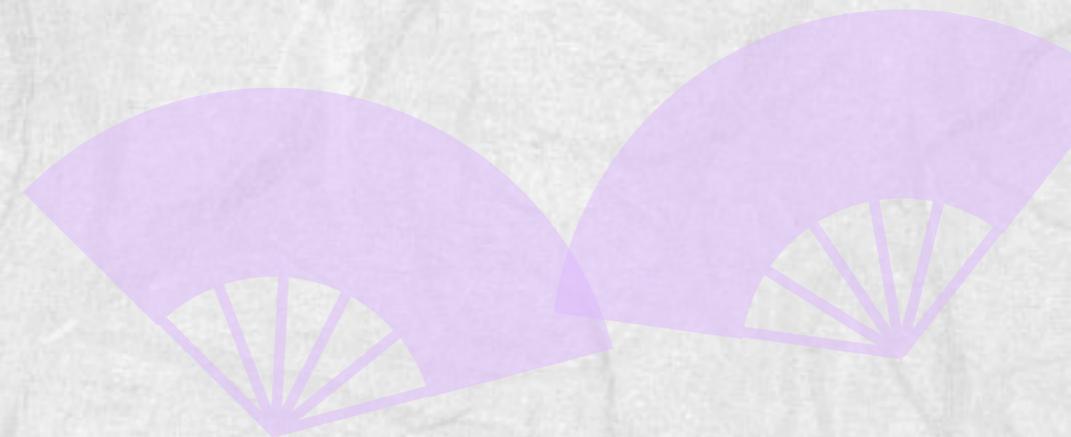
特定の観点以外の特徴は扱いに差があってもよい

**直感的には…**

- ✳ 指定した観点の情報は、推薦結果に影響を与えない
- ✳ 指定した観点の値がどんな値でも、他の条件が同じなら推薦結果は同じになる



# 情報中立推薦システム



# 確率的行列分解モデル

[Salakhutdinov 08, 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

確率的行列分解モデル (probabilistic matrix factorization model)  
行列分解型の推薦モデルの基本形で、他のモデルの原型になっている

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

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

利用者依存バイアス

利用者  $x_i$ , アイテム  $y_i$ , 評価値  $r_i$  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

# 視点特徴

## 視点特徴変数 (viewpoint feature) $V$

- \* 利用者などが指定する視点を表す特徴
- \* 利用者  $x$  とアイテム  $y$  の片方, もしくは両方に依存して値が決まる

### \* フィルターバブルの場合

Pariser の Facebook の例の場合, 友人が保守的か革新的かを視点特徴変数とすると, この観点については公平

### \* コンテンツ提供者の公平性

コンテンツを誰が提供しているかを視点特徴とすると, コンテンツ提供者を公平に扱える

### \* その他

利用者の性別, 映画の公開年, アイテムの評価時刻など

簡単にするため, ここでは  $V \in \{0, 1\}$  の二値変数に限定

# 中立性項と目的関数

**中立性項**  $\text{neutral}(R, V)$  : 推薦の中立性を評価する項

- \* 予測評価と視点特徴の両方に依存
- \* 推薦の中立性が高いほど、すなわち視点特徴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

確率的行列分解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

$$\sum_D (r_i - \hat{r}(x_i, y_i, v_i))^2 + \eta \text{neutral}(R, V) + \lambda \|\Theta\|_2^2$$

二乗損失関数

中立性関数

L<sub>2</sub>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

# 損失関数の変更

## 確率的行列分解モデルの変更

推薦結果の補正ができるように、視点特徴への依存性を導入

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

視点特徴の各値ごとに潜在変数モデルを作り、視点特徴の値に応じてモデルを選択する

# 情報中立推薦システム

推薦の中立性：視点特徴の値に推薦結果が影響されないこと



$\Pr[R | V] = \Pr[R]$ ：推薦結果  $R$  と視点特徴  $V$  が統計的に独立



中立性関数 =  $R$  と  $V$  の統計的独立性の評価指標



相互情報量

$$I(R; V) = \sum_{R, V} \Pr[R, V] \log \frac{\Pr[R|V]}{\Pr[R]}$$

従来法：解析的微分不能・非効率



Caldars&Verwerスコア

$$\| \Pr[R|V = 0] - \Pr[R|V = 1] \|$$

提案法：解析的微分可能・効率的

# 相互情報量

$$\begin{aligned} -I(R; V) &= -\sum_{R, V} \Pr[R, V] \log \frac{\Pr[R|V]}{\Pr[R]} \\ &\approx -\frac{1}{|\mathcal{D}|} \sum_{(r, v) \in \mathcal{D}} \log \frac{\Pr[r|v]}{\Pr[r]} \\ &= -\frac{1}{|\mathcal{D}|} \sum_{(r, v) \in \mathcal{D}} \log \frac{\Pr[r|v]}{\sum_{v' \in V} \Pr[r|v'] \Pr[v']} \end{aligned}$$

Pr[v]の計算は  
容易だが  
混合分布の  
計算は面倒

## Pr[r|v]の計算

$$\Pr[r|v] = \sum_{X, Y} \Pr[X, Y] \Pr[r|X, Y, v] = \frac{1}{|\mathcal{D}|} \sum_{(x, y) \in \mathcal{D}} \Pr[r|x, y, v]$$

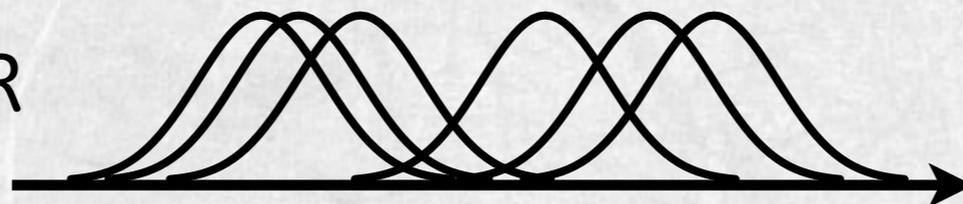
確率的行列分解モデル：  $\Pr[r|x, y, v] \sim \text{Normal}(\hat{r}(x, y, v), \sigma^2)$

# 相互情報量

$$\Pr[r|v] = \frac{1}{|\mathcal{D}|} \sum_{(x,y) \in \mathcal{D}} \Pr[r|x,y,v]$$

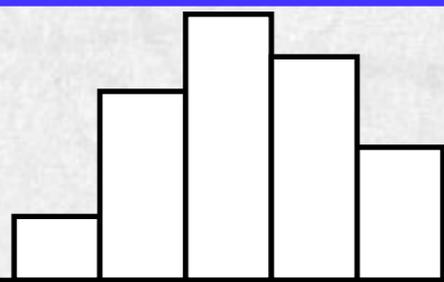
データ数と同じ  
要素数の  
混合正規分布

評価スコア R



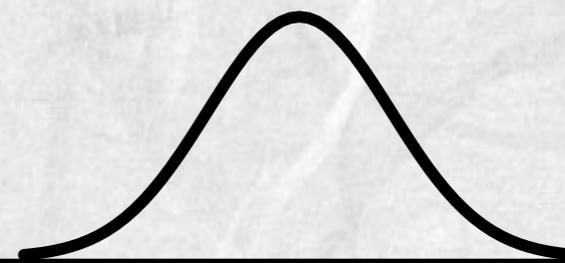
複雑なので（強引に）置き換え

mi-hist



ヒストグラムで表現

mi-normal



正規分布1個で表現

解析的に微分できないので目的関数の数値最適化は非効率

# Calders&Verwerのスコア(CVスコア)

視点特徴の値が異なる場合の  $R$  の分布を近づける

$$- \| \Pr[R|V=0] - \Pr[R|V=1] \|$$

分布の近づけ方は2種類

m-match

$$- (\text{Mean}_{\mathcal{D}(0)}[\hat{r}] - \text{Mean}_{\mathcal{D}(1)}[\hat{r}])^2$$

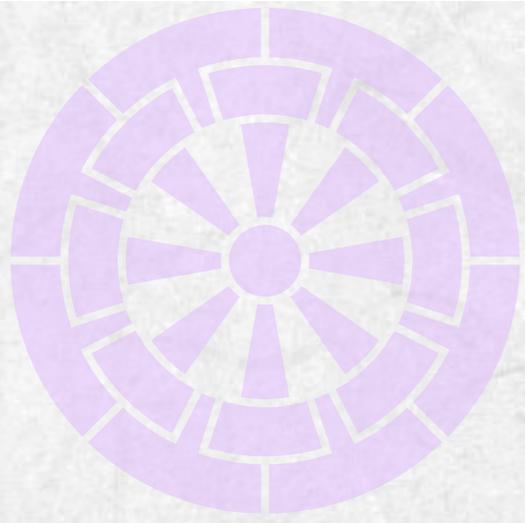
$V=0$  と  $1$  のときの  
予測スコアの平均を合わせる

r-match

$$- \sum_{(x,y) \in \mathcal{D}} (\hat{r}(x,y,0) - \hat{r}(x,y,1))^2$$

$V$  の実際の値とは無関係に  
 $V$  が  $0$  でも  $1$  でも  
同じ予測スコアになるように

解析的に微分できるので目的関数の数値最適化は効率的



# 実験結果



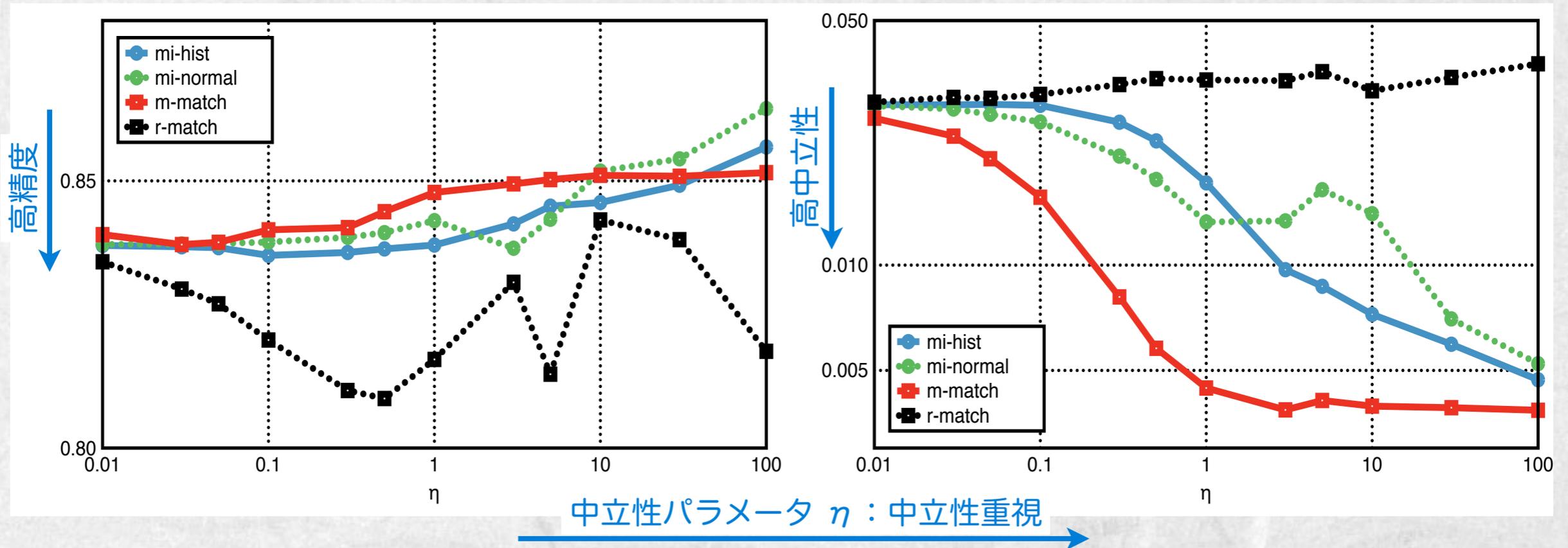
# 実験条件

- \* Movielens 100k データのうち、9409個の評価を抽出して実験
  - \* mi-hist/mi-normal の最適化で、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つたため
- \* 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
- \* 実験に使った視点特徴：**Genderの方が元から中立性が高い**
  - \* **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**（予測評価値と視点特徴の正規化相互情報量）

# 実験結果：Year視点

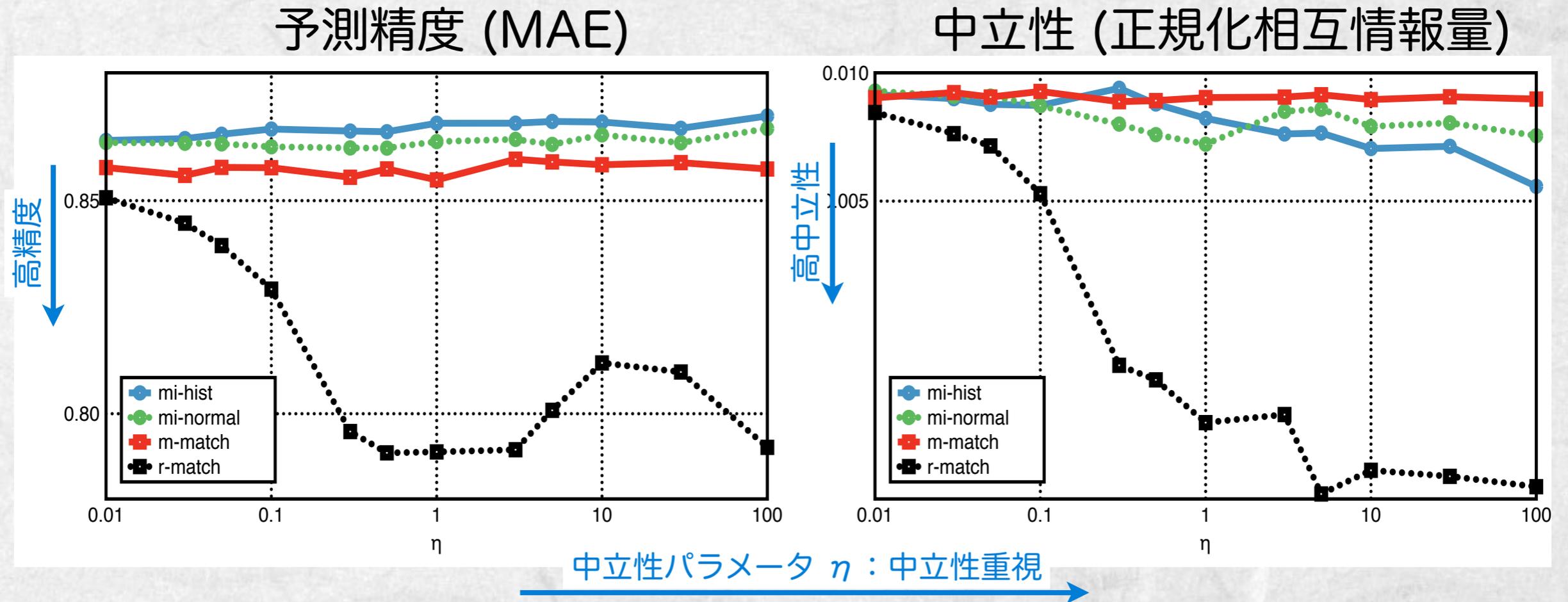
予測精度 (MAE)

中立性 (正規化相互情報量)

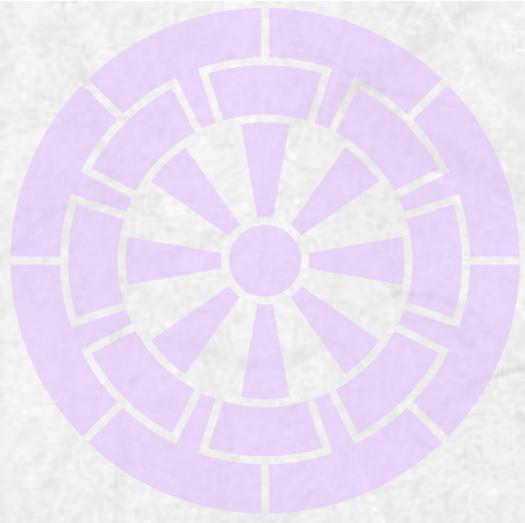


- \* r-match 以外は， $\eta$ を増やして中立性を重視すると，予測精度はそれほど深刻ではない一方で，中立性は急速に減少している
- \* r-match のふるまいを説明することは難しい．最適化の問題か，実在しない状態を扱ったことによる問題か不明

# 実験結果：Gender視点



- \* r-match 以外は,  $\eta$ を増やして中立性を重視してもあまり変化はない. 元々この視点では $V=0$ と $1$ の平均評価値の差は小さい
- \* r-match のふるまいを説明することはやはり難しい. 中立性が向上してるのは分布の2次以上のモーメントの影響の可能性.



# 関連研究



# 推薦の多様性

[Ziegler+ 05]

## 話題の分散化 (Topic Diversification)

利用者と嗜好が一致するアイテムを選びつつ，推薦リストから類似したアイテムは排除することで，推薦アイテムの多様性を確保

類似度関数  $c$  : アイテムの内容に基づいてアイテム間の類似度を測る  
アイテムの階層的な分類があるとき，その構造木中でのホップ数の逆数などで測る

### 多様性

アイテムが推薦リスト中や推薦時系列上にのアイテムが互いに類似していない

推薦結果間の関係

### 中立性

視点特徴の情報を排除するだけで，推薦アイテム間の類似性は関係ない

推薦結果と視点の関係

# プライバシー保護データマイニング

評価値変数  $R$  と視点特徴  $V$  の独立性

評価値変数  $R$  と視点特徴  $V$  の相互情報量が 0



プライバシー保護データマイニングの観点からの解釈

評価値  $R$  の予測値を知られても、視点特徴  $V$  の情報が漏洩しない

# 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した
- ＊効率的に計算可能な中立化項を提案

## 今後の予定

- ＊他の独立性指標の採用で，効率性と中立化性能の両立
- ＊評価値予測ではなく，アイテムの推薦での中立化手法の開発
- ＊pLSI / LDA など生成系の推薦モデルの情報中立化

# 実験コードを公開しています

まだ前のバージョンですが，更新予定です

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

## 謝辞

- \* Movielens データを公開している Grouplens research lab に感謝する
- \* 本研究はJSPS科研費 16700157, 21500154, 23240043, 24500194, および 2550094 の助成を受けたものである