



なんとなく協調フィルタリング 複数の順序応答に基づく推薦

神鳶 敏弘

産業技術総合研究所

第19回 人工知能学会 全国大会 (2004.6.15-17)

概要

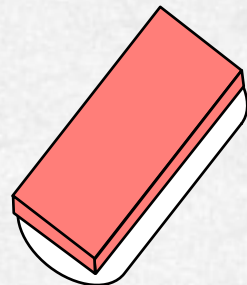
利用者の嗜好を順位法で計測する
協調フィルタリング手法の拡張

順序：何らかの基準で対象を整理したもの

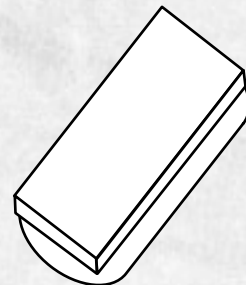
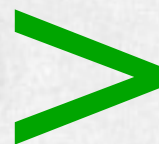
例：Aさんが好き[基準]な順に寿司[対象]を整理



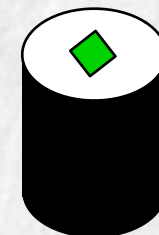
好き



とろ



いか



かっぱ巻



嫌い

“いか” より “とろ” が好き

しかし、どれくらい好きかは不明
順序は主観的な変量の計測に適す

複数個の順序応答を同時に利用できるように拡張

協調フィルタリング

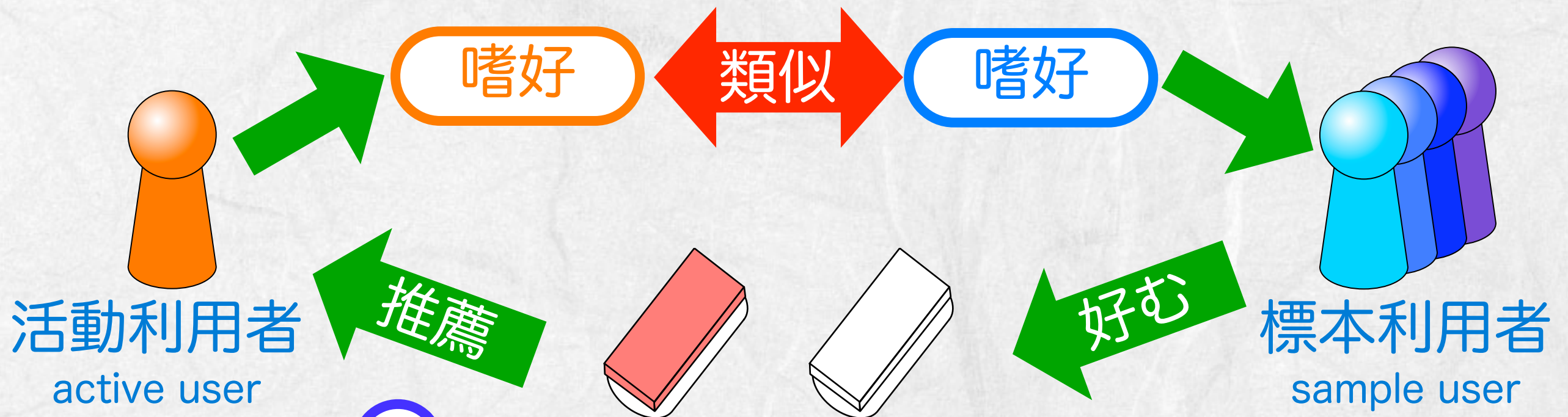
「口コミ情報」を用いて利用者が好む対象を見つける方法

1

活動利用者の嗜好をシステムに提示

2

利用者DBから、活動利用者と類似した嗜好を持つ標本利用者を検索



3

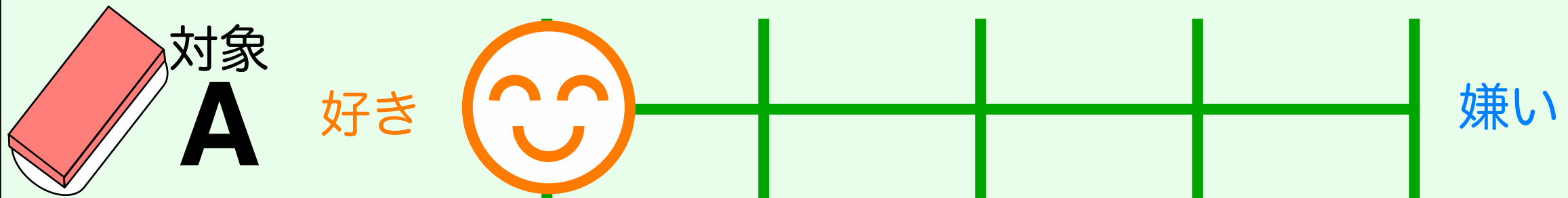
類似した嗜好を持つ標本利用者が好む対象を、活動利用者に推薦

SD法と順位法

Semantic Differential 法 (SD法)

両端を対義語で表した尺度を用いる

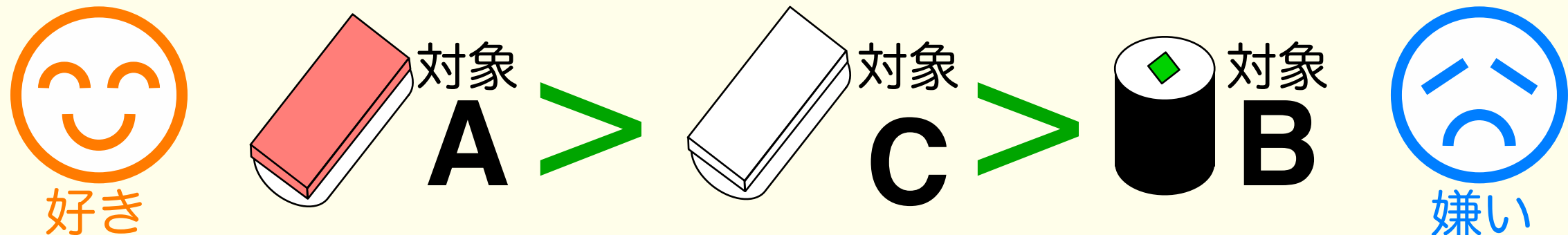
例：被験者が 対象A を好きならば，尺度の「好き」を選ぶ



順位法

計測する度合いの強さの順に対象を整列

例：被験者は 対象A が最も好きで，対象B が最も嫌い



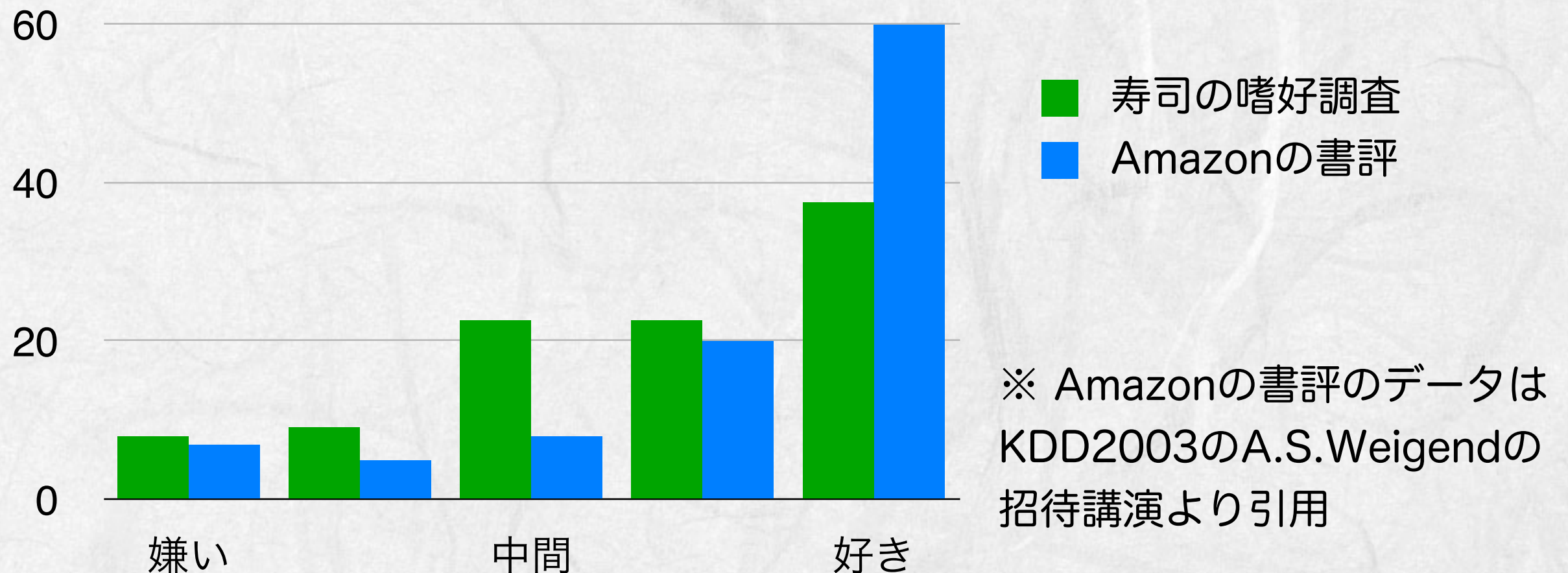
なぜ順序を使うのか？

SD法の仮定 → 理想的なスコアの分布は対称で単峰



現実の分布はかなり歪んだ分布

5段階のSD法で、被験者が選択した値の分布の例



問題点

順位法の採用によって予測精度の向上に成功！



順位法は万能か？

100個の対象について嗜好を調査する場合

SD法

100個の商品について
5段階のスコア付け

可 能

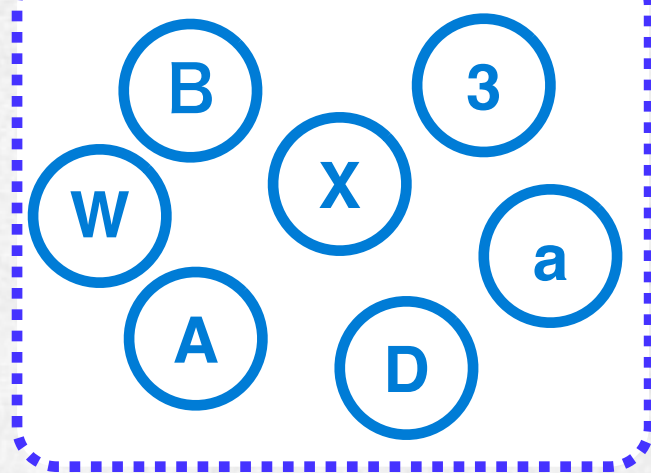
順位法

100個の商品を好きな
ものから順に整列する

無 理

複数の順序応答

多数の対象



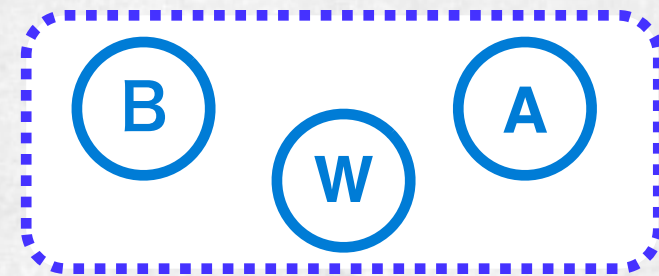
利用者



一度に全ての対象を整理

無理

少数の対象をサンプリング



利用者



サンプリングした
対象だけを整理

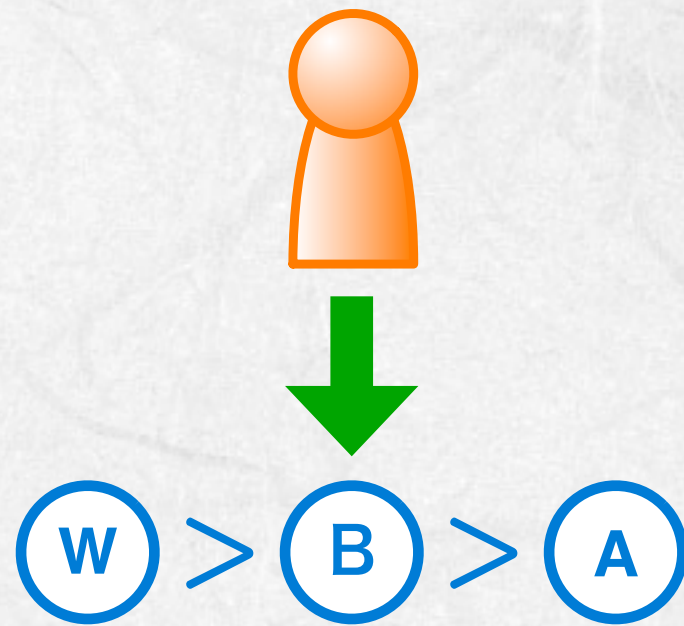
サンプリングと整理を反復

可能

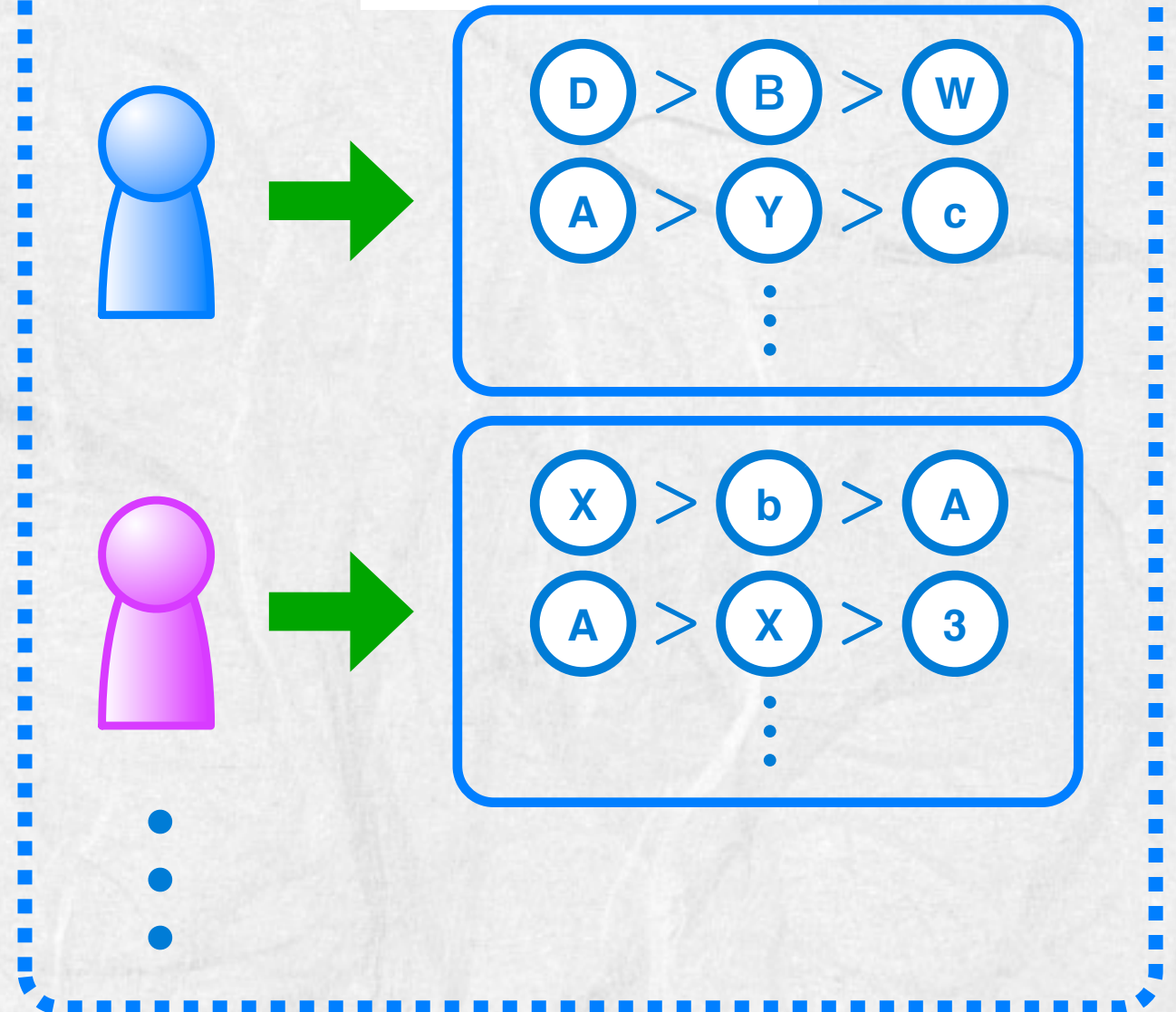
単複なんとなくCF

なんとなくCFでは嗜好パターンを順序応答で計測

活動利用者



標本利用者



一人から一つの順序応答 一人から複数の順序応答を許す

※従来法は単単なんとなくCFと呼んで区別する

単純相関法（単単の場合）

GoupLensの方法で類似度に順位相関を使う
利用者間の類似度で重み付けした平均順位を利用

活動利用者と標本利用者iの間の類似度

$$\text{類似度 } i = \text{順位相関} \left(\begin{array}{c} \text{活動利用者の} \\ \text{順序の対象jの順位} \end{array}, \begin{array}{c} \text{標本利用者iの} \\ \text{順序の対象jの順位} \end{array} \right)$$

活動利用者の対象jの予測評価値

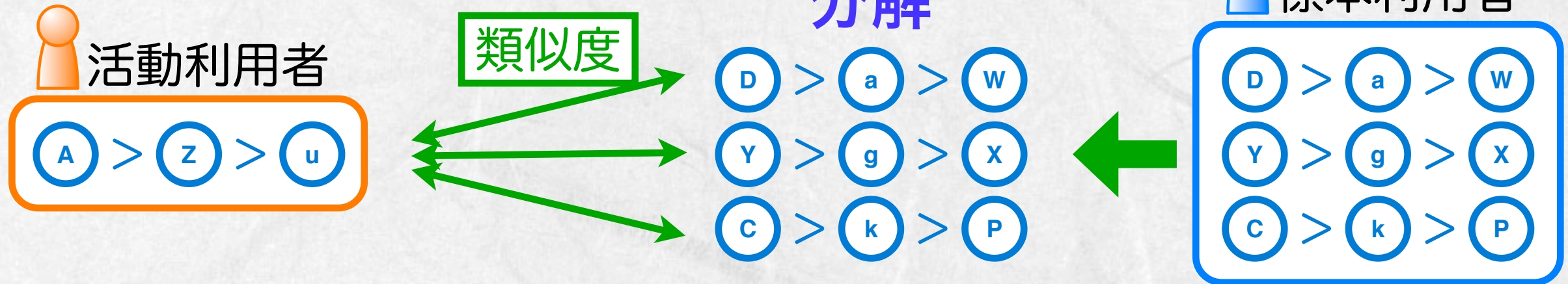
$$\begin{array}{c} \text{正規化} \\ \text{係数} \end{array} \sum_{\text{利用者DB}} \text{類似度 } i \times \left(\begin{array}{c} \text{標本利用者iの} \\ \text{順序の対象jの順位} \end{array} - \begin{array}{c} \text{標本利用者iの} \\ \text{平均順位} \end{array} \right)$$

予測評価の高い対象を活動利用者に推薦

単純相関法（単複の場合）

活動利用者と標本利用者の類似度の計測方法

併合単純相関法



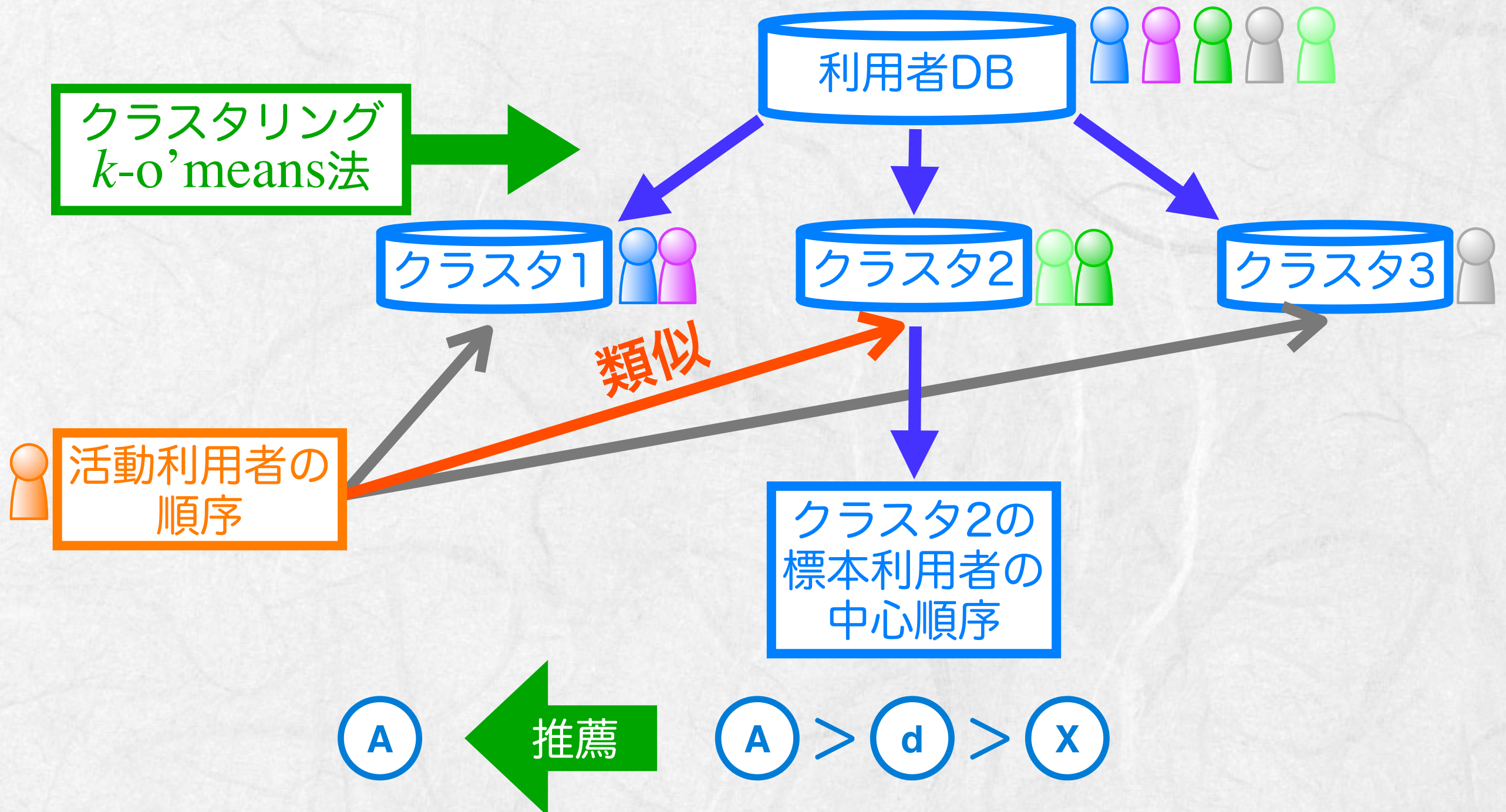
※ 応答順序が同じ標本利用者に由来するという情報を破棄

中心順序単純相関法



※ 中心順序は集合中の順序との距離の総和が最小になる順序

クラスタリング法 (単単の場合)



嗜好が類似しているクラスタで好まれる対象を推薦

クラスタリング法 (単複の場合)

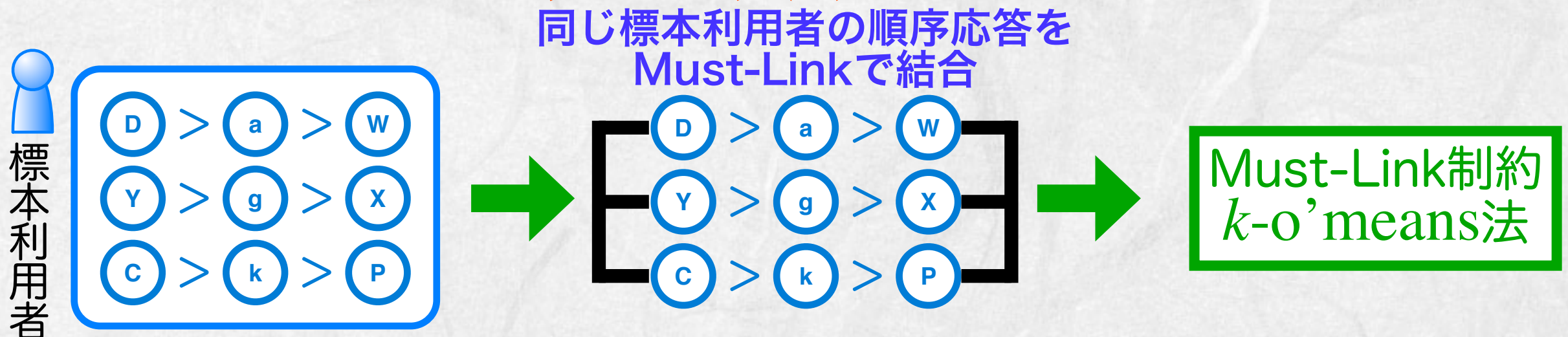
標本利用者の順序応答をクラスタリングする方法

併合クラスタリング法



※ 応答順序が同じ標本利用者に由来するという情報を破棄

Must-Linkクラスタリング法



※ Must-Link制約: このリンクで結合された対象は同じクラスに分類

実験

順位の入力フォーム

- ✿ 寿司は全部で100種
- ✿ 総被験者数は5000人
- ✿ 訓練用対象集合：被験者ごとに寿司を10種ランダムに抽出
- ✿ テスト用対象集合：全被験者共通
- ✿ 訓練用&テスト用対象集合の寿司をそれぞれ順位法で評価

	1番	2番	3番	4番	5番	6番	7番	8番	9番	10番
とびこ	☐	☐	☐	☐	☐	☐	☒	☐	☐	☐
たい	☐	☐	☒	☐	☐	☐	☐	☐	☐	☐
とろ	☐	☒	☐	☐	☐	☐	☐	☐	☐	☐
まぐろ	☐	☐	☐	☐	☐	☐	☐	☐	☐	☒
いくら	☐	☐	☐	☐	☐	☐	☐	☒	☐	☐
めんたいこ	☐	☐	☐	☒	☐	☐	☐	☐	☐	☐
あおやぎ	☐	☐	☐	☐	☐	☒	☐	☐	☐	☐
しゃこ	☐	☐	☐	☐	☒	☐	☐	☐	☐	☐
うなぎ	☒	☐	☐	☐	☐	☐	☐	☐	☐	☐
赤貝	☐	☐	☐	☐	☐	☐	☐	☐	☒	☐

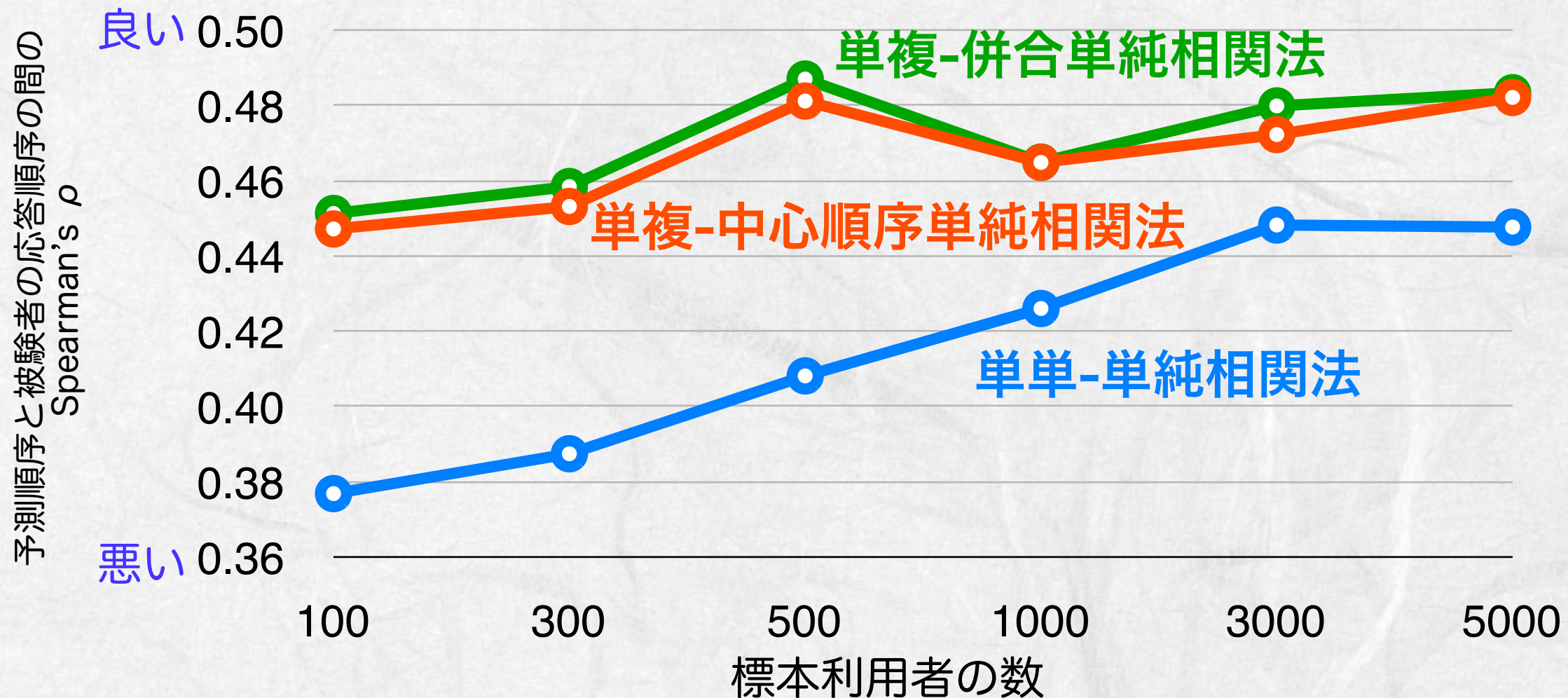
寿司の名前

順位を指定

一つの応答順序からサンプリングして
複数の応答順序を擬似的に生成

結果 (単純相関法)

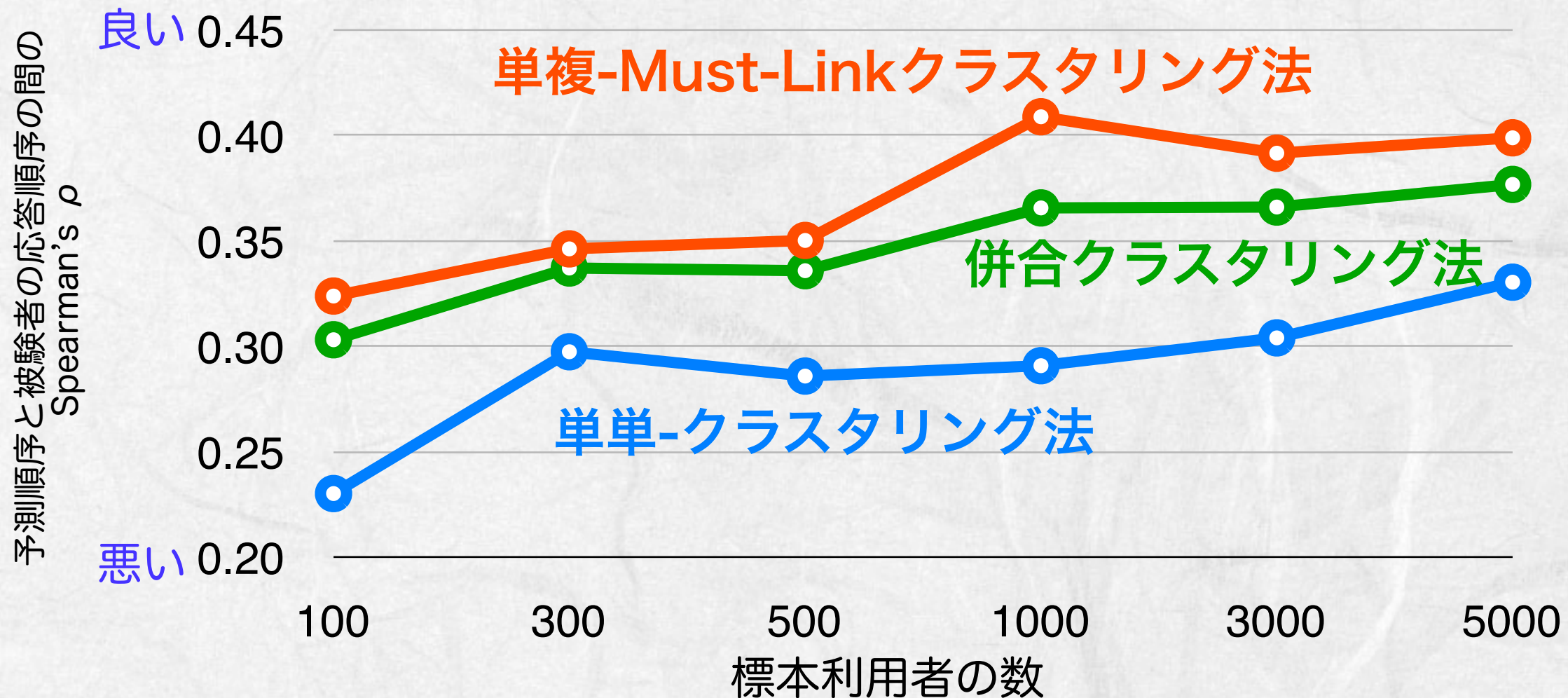
長さ5の順序応答1個の単単と，2個用いた単複の場合



- ✿ 複数の順序を利用することで，推定精度は向上
- ✿ 中心順序単純相関法は併合単純相関法と同等の精度で，計算量は単単の場合とほぼ同等で高速

結果 (クラスタリング法)

長さ5の順序応答1個の単単と，2個用いた単複の場合



- ✳ 複数の順序を利用することで，推定精度は向上
- ✳ Must-Linkクラスタリング法は，順序応答がどの標本利用者に由来するのかという情報を利用するので，併合クラスタリング法より推定精度が良かった

複複なんとなくCF

標本利用者だけでなく，活動利用者も複数の順序応答
単複なんとなくCFをサブルーチンとして利用

事前中心化



中心化



中心順序

単複
なんとなくCF



事後中心化



単複
なんとなくCF



個別の予測順序

中心化



結果

応答順序の長さ=5, サンプル数=5000

応答順序の数は活動利用者も標本利用者も2個ずつ

	中心順序 単純相関法	Must-Link クラスタリング法
事前中心化	0.475	0.402
事後中心化	0.462	0.424

- ✳ 単純相関法では高速な事前中心化がよい
- ✳ クラスタリング法では事後中心化がよい

まとめ

まとめ

- ✳ 従来のなんとなくCFでは、対象が多い場合に、利用者の嗜好データを獲得できない問題があった。これに対して、各利用者から複数の順序応答を獲得する枠組みを提案。
- ✳ この拡張なんとなくCFのための手法を開発し、複数の順序応答を用いて推薦精度を向上できることを確認

今後の予定

- ✳ 複数なんとなくCFへの拡張
- ✳ 内容に基づく推薦との統合