



# なんとなく協調フィルタリング 順序応答に基づく推薦

神島 敏弘 (産業技術総合研究所)  
<http://www.kamishima.net/>  
人工知能学会 FAI & KBS研究会 (2004.3.3)

# 概要

利用者の嗜好を順序を用いて計測する  
協調フィルタリングによる推薦システムの提案

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



“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

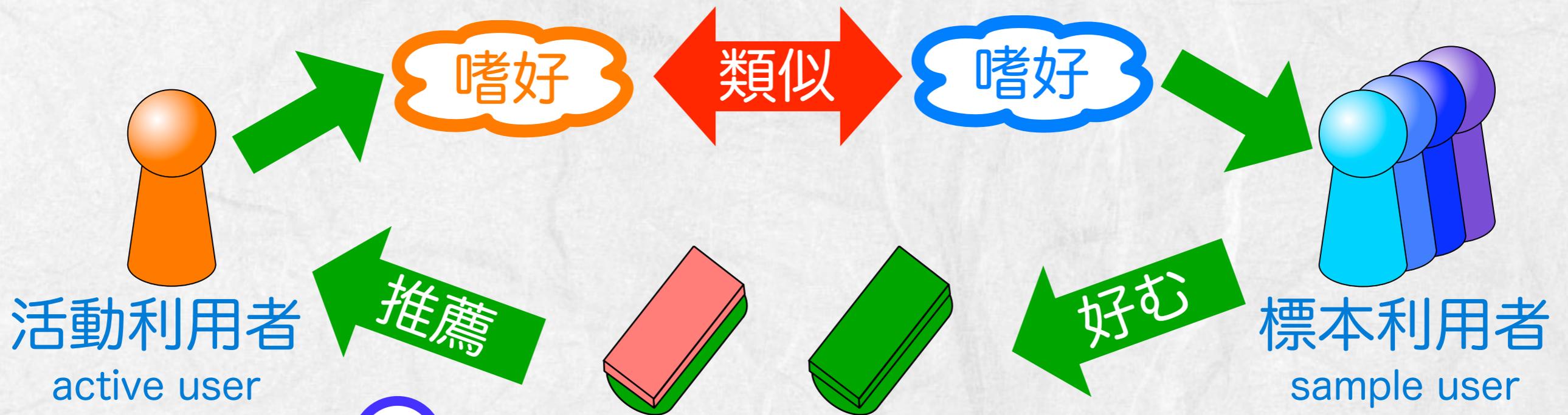
順序の導入により、さらに適切な推薦が可能になる

# 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



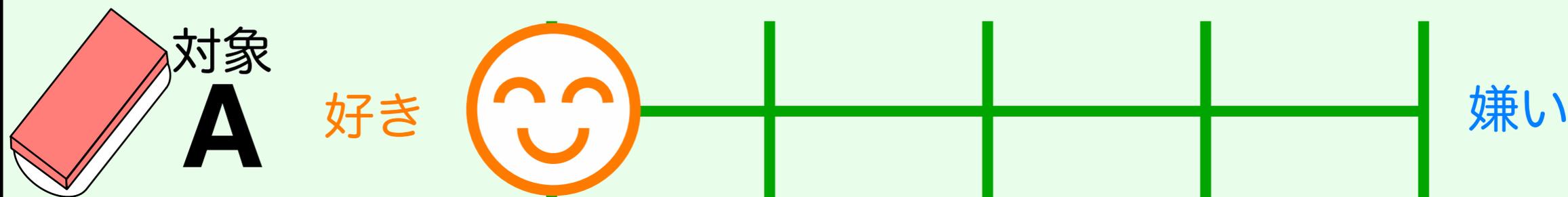
3 類似した嗜好を持つ標本利用者が好む対象を、活動利用者に推薦

# 順序による計測

## 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



# SD法の問題点(1)

## 1. SD法の非現実的な仮定

- ✳ 全被験者は尺度の全長に対する理解を共有  
異なる被験者の間で、その感覚を定量的に共有するのは困難  
順位法では相対的にしか評価しないので、絶対量は無関係
- ✳ 尺度の目盛りは等間隔  
被験者が嗜好の度合いを等間隔に分割するのは困難  
順位法では間隔は無視し、相対的な大小関係だけを扱う

## 2. 心理的な攪乱要因

例：Central Tendency Effect

尺度の中心に近い部分を使って回答する傾向

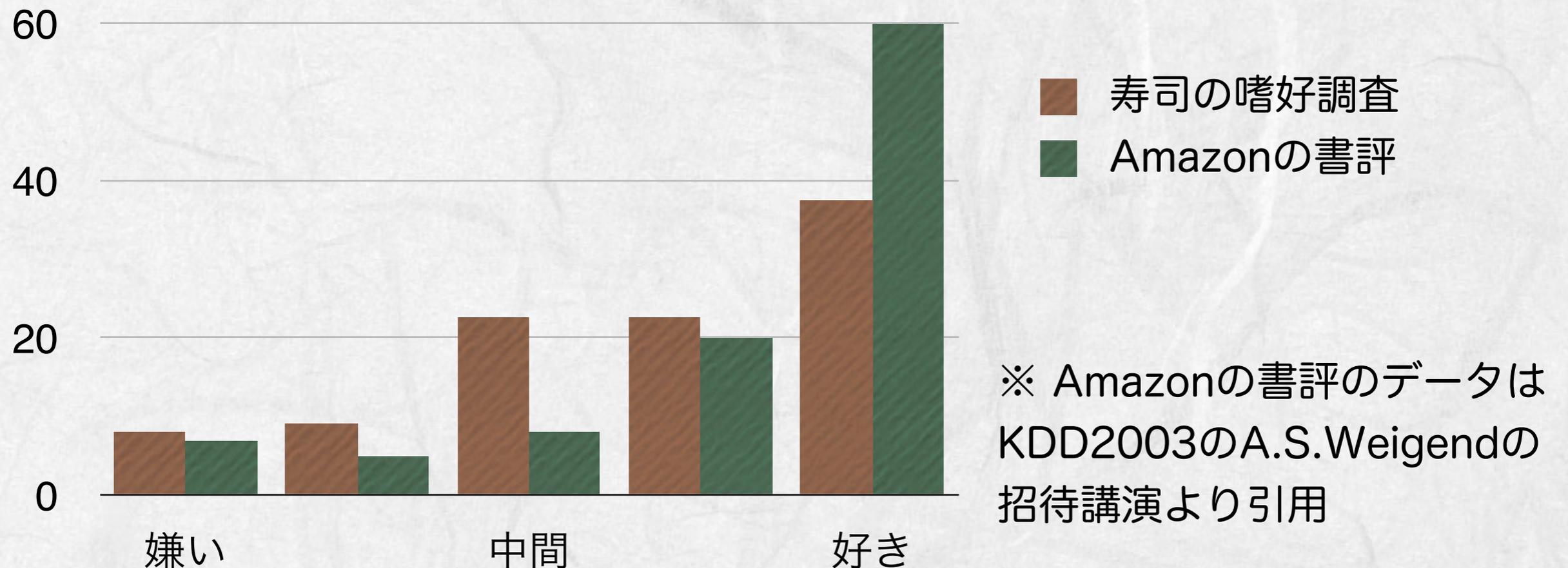
# SD法の問題点(2)

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



# GroupLensの方法 (注)

活動利用者と標本利用者iの間の類似度

$$\text{類似度 } i = \text{相関係数} \left( \begin{array}{l} \text{活動利用者の} \\ \text{対象jへの評価値} \end{array} \begin{array}{l} \text{標本利用者 } i \text{ の} \\ \text{対象jへの評価値} \end{array} \right)$$

※ 両方の利用者が評価した対象についてのみを考慮した相関係数

活動利用者の対象jの予測評価値

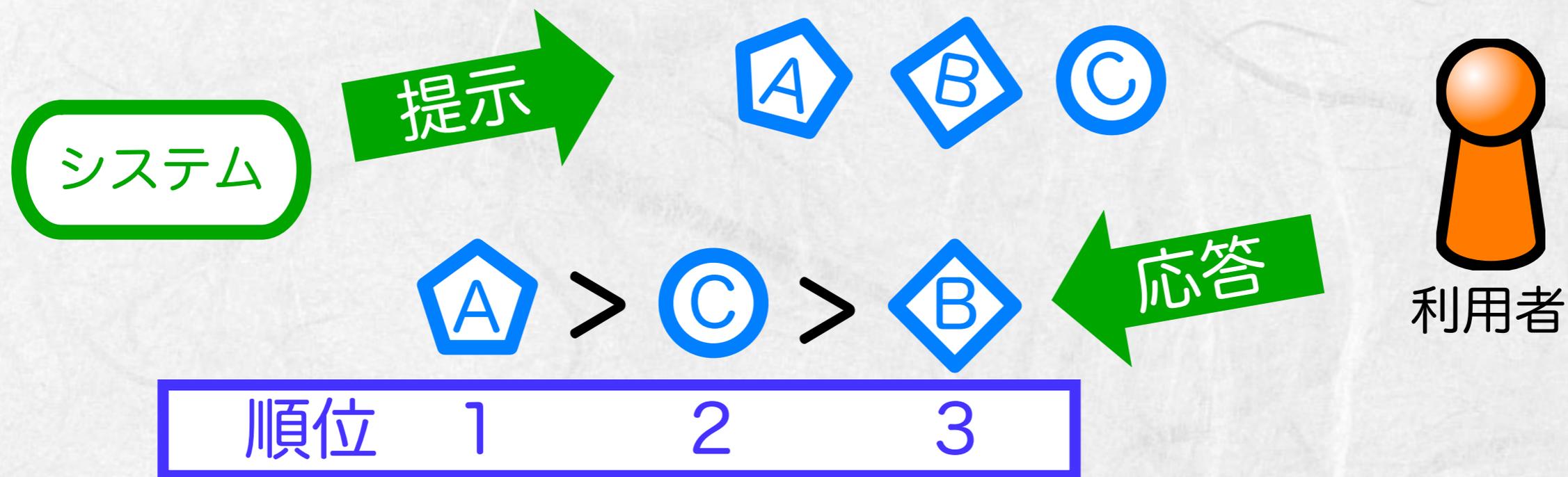
$$\text{正規化係数} \sum_{\text{利用者DB}} \text{類似度 } i \times \left( \begin{array}{l} \text{標本利用者 } i \text{ の} \\ \text{対象jへの評価値} \end{array} - \begin{array}{l} \text{標本利用者 } i \text{ の} \\ \text{平均評価値} \end{array} \right)$$

予測評価の高い対象を活動利用者に推薦

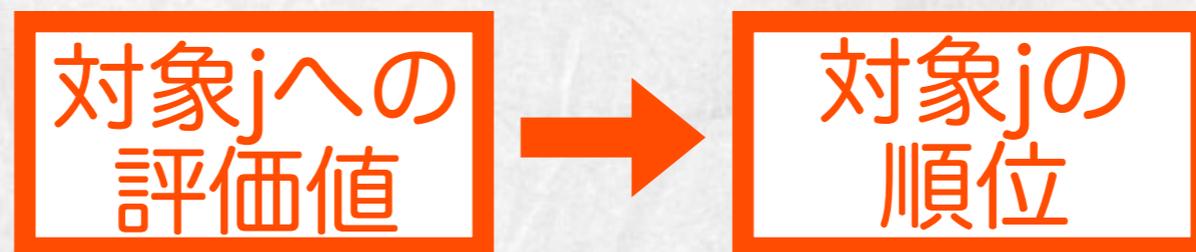
注: 現在は映画の推薦システム MovieLens ([www.movielens.org](http://www.movielens.org))

# メモリベース法

GroupLensでは利用者から対象の**評価値**を得るが  
なんとなく協調フィルタリングでは**順序**を得る

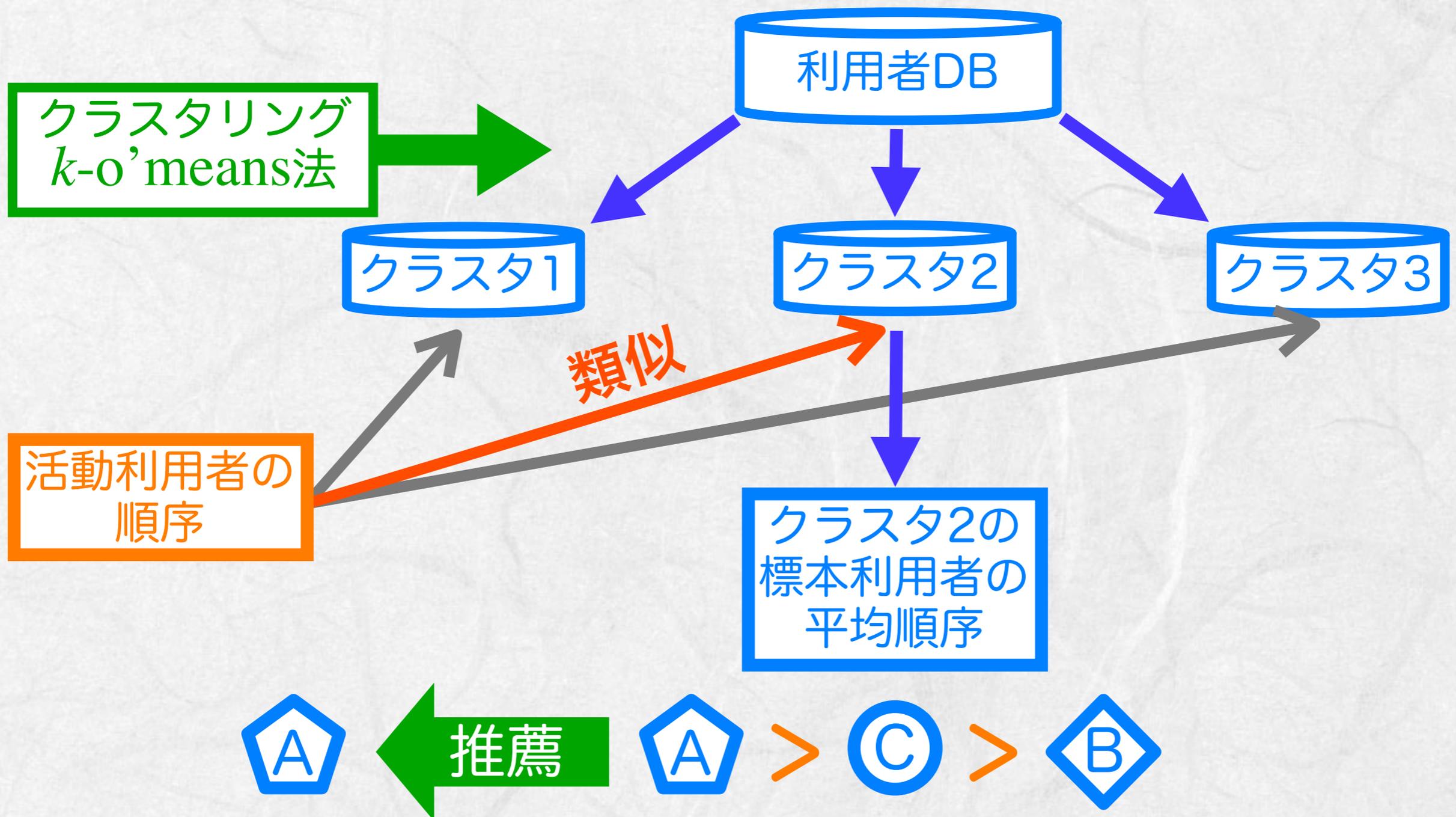


GroupLensの方法との相違は



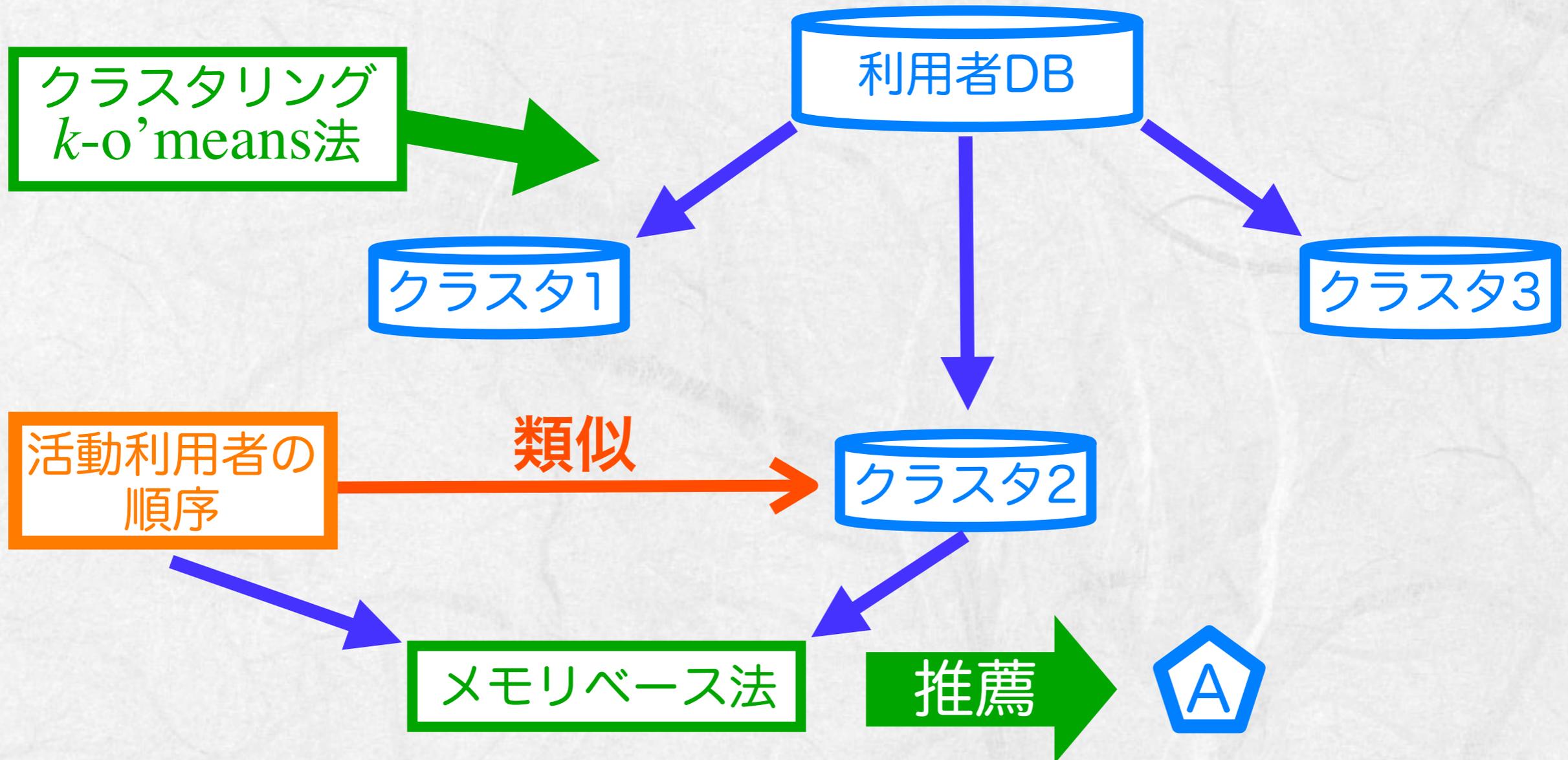
と置き換えた点のみ

# モデルベース法



嗜好が類似しているクラスタで好まれる対象を推薦

# 合成法



類似した標本利用者だけを使うメモリベース法

比較する標本利用者の数を減らせるので通常のメモリベースより高速  
クラスタリングによるハッシュ

# 寿司の嗜好調査データでの実験

## 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
- ✿ 訓練用対象集合はSD法でも評価

もう一度、あなたが好きな順に番号をつけてください。  
途中で、どのネタや番号を選んでいないか分からなくなったときには、「チェックする」ボタンを押すと、まだ選んでいない番号やネタが分かります。

チェックする

	1番	2番	3番	4番	5番	6番	7番	8番	9番	10番
とびこ	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>					
たい	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
とろ	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
まぐろ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>						
いくら	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>					
めんたいこ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
あおやぎ	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>				
しゃこ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
うなぎ	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
赤貝	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>						

終わったら押してください

『とびこ』 トビウオの卵  
『とろ』 まぐろの脂の多い部分  
『たい』 鯛  
『まぐろ』 鯖: 赤身の部分

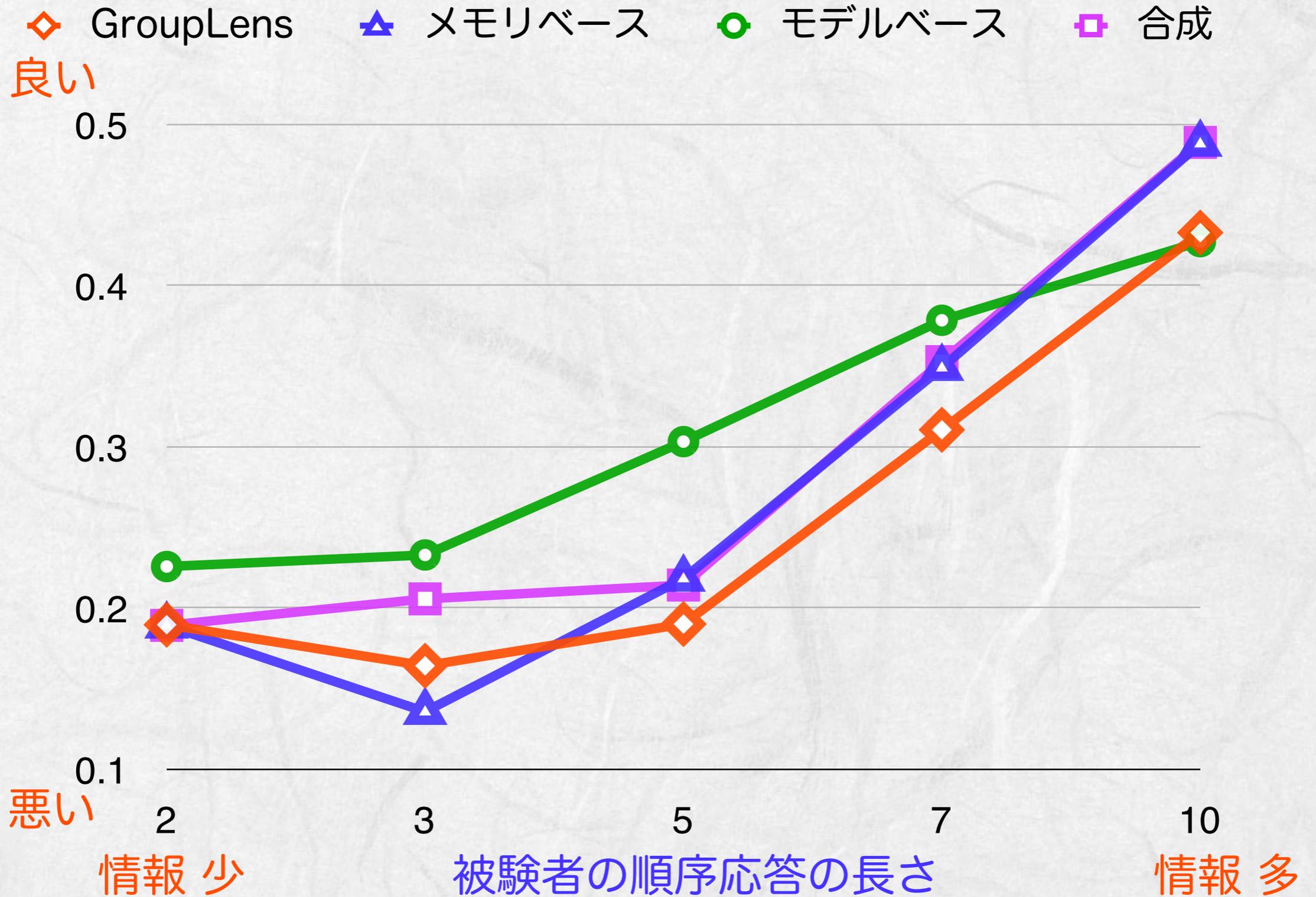
寿司の名前

順位を指定

訓練用対象集合の応答を利用者DBや活動利用者の嗜好と考え、テスト用対象集合の寿司に対する嗜好を予測し、予測の精度を測る

# 実験結果

予測順序と被験者の応答順序の間  
のSpearman's  $\rho$



# 結論と今後の予定

## 結論

- ✿ ひとりの利用者が評価する対象数が5個以上なら，SD法の代わりに順位法で嗜好を計測することで推薦の予測精度が向上
- ✿ モデルベース法は，利用者個人ではなくグループの嗜好を用いて予測するので，利用者の個人情報が少ない場合に有利
- ✿ クラスタ内の近傍の利用者の嗜好だけを用いた合成法では，予測精度を下げることなく高速化が可能

## 今後の予定

- ✿ 利用者あたりの評価対象数が少ない場合の予測精度の向上
- ✿ 複数の異なる対象集合に対して，一人の利用者が順序応答を与えた場合に適用可能な手法の開発