粒状性を考慮したラフ集合ベースの混合多項分布型共クラスタリングに基づく協調フィルタリング

人間情報システム研究グループ
BGA22050　毛利 憲竜

構成

1. はじめに
2. 準備
   - クラスタリング（HCM, RCM, RSCM）
   - 共クラスタリング（HCCMM, RCCMM, RSCCMM）
3. RSCCMMに基づく協調フィルタリング（RSCCMM-CF）
4. 数値実験
5. おわりに

はじめに

推薦システム
ユーザー毎に，好ましいコンテンツを推薦する

https://www.amazon.co.jp/Prime-Video

クラスタリング

類似した対象同士からなるクラスターを抽出することにより，データを自動的に分類・要約する手法

- 容易な実装
- 効率的な計算
- メモリの削減

クラスタリングに基づく協調フィルタリング(CF)

嗜好パターンの類似したユーザーをクラスターとして抽出し，ユーザーの帰属するクラスター内で嗜好度の高いコンテンツを推薦

はじめに

ラフ共クラスタリング

- ユーザーとアイテム間の関係を表す共起情報
  → 関係の強いユーザーとアイテムの組に着目する**共クラスタリング**
- 人間の感覚に基づく不確実性を含むデータ
  → ラフ集合理論に基づく**ラフクラスタリング**

従来法: RCCMM-CF

- 共クラスタリングにラフ集合理論の観点を導入した**ラフ共クラスタリング**(RCCMM)に基づくCF
- ラフ集合理論の重要概念である粒状性を考慮していない

提案法: RSCCMM-CF

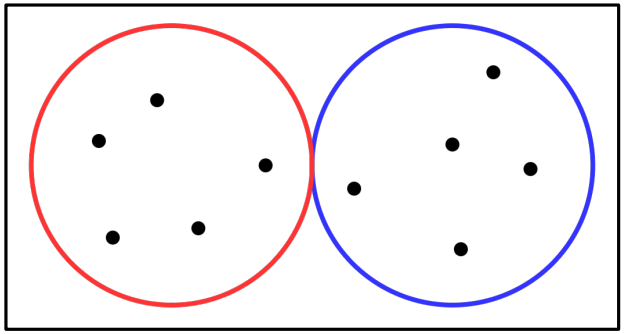
- **粒状性**を考慮したラフ共クラスタリング(RSCCMM)をCFに応用
- 実データを用いた数値実験により，推薦性能・粒状性の効果を検証する

構成

1. はじめに
2. 準備
   - クラスタリング（HCM, RCM, RSCM）
   - 共クラスタリング（HCCMM, RCCMM, RSCCMM）
3. RSCCMMに基づく協調フィルタリング（RSCCMM-CF）
4. 数値実験
5. おわりに

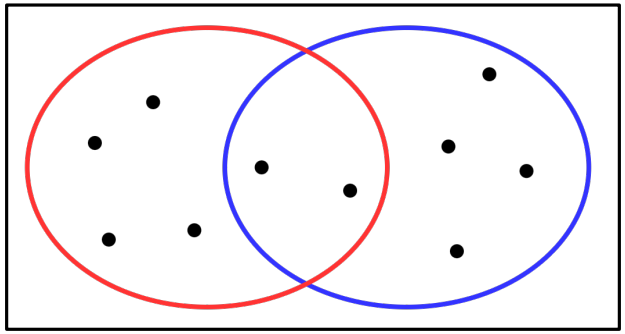
クラスタリング（HCM, RCM, RSCM）

- 各対象はクラスター中心と対象との距離尺度をもとに帰属を決定
- 対象のメンバシップとクラスター中心の更新を繰り返してクラスターを抽出



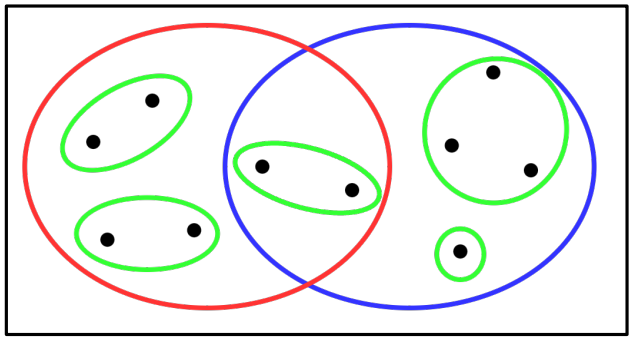
Hard C-Means (HCM)
[MacQueen, 1967]

- 各対象は**排他的**に一つのクラスターのみに帰属する



Rough C-Means (RCM)
[Lingras and West, 2004]

- 対象が複数のクラスターに帰属でき，クラスターの**オーバーラップ**を表現できる



Rough Set C-Means (RSCM)
[Ubukata et al, 2018]

- 対象空間を**粒状化**し，粒単位でクラスターへの帰属を判定する

構成

1. はじめに
2. 準備
   - クラスタリング（HCM, RCM, RSCM）
   - **共クラスタリング（HCCMM, RCCMM, RSCCMM）**
3. RSCCMMに基づく協調フィルタリング（RSCCMM-CF）
4. 数値実験
5. おわりに

Hard Co-Clustering induced by Multinomial Mixture models (HCCMM) [Ubukata et al, 2020]

- 共クラスタリング: **対象×項目の共起関係データ**において，
  関連性の強い対象と項目の組からなる**共クラスター**を抽出する

共起関係データ

	項目1	項目2	項目3	項目4	項目5
対象1	1	0	1	0	0
対象2	1	1	0	0	0
対象3	1	1	1	0	0
対象4	0	0	1	1	0
対象5	0	0	1	0	1

共クラスター1

共クラスター2

対象メンバシップ

各対象の最も親近性の高いクラスターへの帰属を表す

	対象1	対象2	対象3	対象4	対象5
共クラスター1	**1**	**1**	**1**	0	0
共クラスター2	0	0	0	**1**	**1**

項目メンバシップ

各クラスター内での項目の重要度を表す

	項目1	項目2	項目3	項目4	項目5
共クラスター1	1/2	1/4	1/4	0	0
共クラスター2	0	0	1/2	1/4	1/4

Rough CCMM (RCCMM) [Ubukata et al, 2020]

- HCCMM法における，ハードな対象割り当ての条件を緩和し，
  オーバーラップ構造をもつ共クラスターを抽出する手法

共起関係データ

	項目1	項目2	項目3	項目4	項目5
対象1	1	0	1	0	0
対象2	1	1	0	0	0
対象3	1	1	1	1	0
対象4	0	0	1	1	0
対象5	0	0	1	0	1

共クラスター1

共クラスター2

対象の正規化メンバシップ

各対象の比較的類似度の高いクラスターへの帰属を表す

	対象1	対象2	対象3	対象4	対象5
共クラスター1	1	1	**1/2**	0	0
共クラスター2	0	0	**1/2**	1	1

項目のメンバシップ

各クラスター内での項目の重要度を表す

	項目1	項目2	項目3	項目4	項目5
共クラスター1	3/7	2/7	2/7	0	0
共クラスター2	0	0	1/2	1/4	1/4

Rough Set CCMM (RSCCMM) [Ubukata et al, 2021]

- 対象の二項関係を設定し，対象空間を粒状化する

共起関係データ：
対象間の類似性に基づき二項関係を設定

	項目1	項目2	項目3	項目4	項目5
対象1	1	0	1	0	0
対象2	1	1	0	0	0
対象3	1	1	1	0	0
対象4	0	0	1	1	0
対象5	0	0	1	0	1

二項関係がある

二項関係がある

共起関係データ：粒状化

	項目1	項目2	項目3	項目4	項目5
対象1	1	0	1	0	0
対象2	1	1	0	0	0
対象3	1	1	1	0	0
対象4	0	0	1	1	0
対象5	0	0	1	0	1

Rough Set CCMM (RSCCMM) [Ubukata et al, 2021]

- HCCMM法によって，暫定共クラスターを抽出し，粒単位で共クラスターへの帰属を判定する

共起関係データ: ハードな分割

	項目1	項目2	項目3	項目4	項目5
対象1	1	1	1	0	0
対象2	1	1	0	0	0
対象3	1	0	1	0	0
対象4	0	0	1	1	0
対象5	0	0	1	0	1

共クラスター1

共クラスター2

共起関係データ: 粒単位での帰属判定

	項目1	項目2	項目3	項目4	項目5
対象1	1	1	1	0	0
対象2	1	1	0	0	0
対象3	1	0	1	0	0
対象4	0	0	1	1	0
対象5	0	0	1	0	1

共クラスター1

共クラスター2

構成

1. はじめに
2. 準備
   - クラスタリング（HCM, RCM, RSCM）
   - 共クラスタリング（HCCMM, RCCMM, RSCCMM）
3. **RSCCMMに基づく協調フィルタリング（RSCCMM-CF）**
4. 数値実験
5. おわりに

RSCCMMに基づく協調フィルタリング(RSCCMM-CF)

協調フィルタリングが対象とするデータ

- ユーザー数n，アイテム数mとした $n \times m$ の共起関係行列 $R = \{r_{ij}\}$

RSCCMM法によるラフ共クラスタリング

- RSCCMM法を用いて，対象空間の粒状性を考慮した
  ラフ共クラスタリングを行い，共クラスター構造を抽出

協調フィルタリングへの応用

- RSCCMM法によるクラスタリングによって抽出された，
  それぞれの共クラスター内で嗜好度の高いコンテンツを推薦

負の交差エントロピーに基づく二項関係

- 対象間の二項関係を形成する類似度尺度として，**負の交差エントロピー**を用いる
- 交差エントロピー：2つの確率分布の非類似度尺度

共起関係データ

	アイテム1	アイテム2	アイテム3	アイテム4	アイテム5
ユーザー1	1	1	1	0	0
ユーザー2	1	1	0	0	0
ユーザー3	1	0	1	0	0
ユーザー4	0	0	1	1	0
ユーザー5	0	0	1	0	1

各行を正規化し，
各ユーザーの特徴を
確率分布として捉える

ユーザーの確率分布

	アイテム1	アイテム2	アイテム3	アイテム4	アイテム5
ユーザー1	1/3	1/3	1/3	0	0
ユーザー2	1/2	1/2	0	0	0
ユーザー3	1/2	0	1/2	0	0
ユーザー4	0	0	1/2	1/2	0
ユーザー5	0	0	1/2	0	1/2

負の交差エントロピーに基づく二項関係

正規化された共起関係行列:

$\tilde{X} = \{\tilde{x}_{ij}\}$

	アイテム 1	アイテム 2	アイテム 3	アイテム 4	アイテム 5
ユーザー 1	1/3	1/3	1/3	0	0
ユーザー 2	1/2	1/2	0	0	0
ユーザー 3	1/2	0	1/2	0	0
ユーザー 4	0	0	1/2	1/2	0
ユーザー 5	0	0	1/2	0	1/2

- 対象iと対象tの類似度

$$S_{it}^{CE} = \sum_{j=1}^{m} \tilde{x}_{tj} \log \tilde{x}_{ij}$$

- 類似度の閾値処理で二項関係を設定

$$R_{it}^{CE} = \begin{cases} 1 & \left(S_{it}^{CE} \geq \delta\right) \\ 0 & (\text{otherwise}) \end{cases}$$

二項関係: R（例）

	ユーザー 1	ユーザー 2	ユーザー 3	ユーザー 4	ユーザー 5
ユーザー 1	1	0	0	0	0
ユーザー 2	0	1	1	0	0
ユーザー 3	0	1	1	0	0
ユーザー 4	0	0	0	1	1
ユーザー 5	0	0	0	1	1

負の交差エントロピーは負の値を取り，対象間の類似性が認められないほど，類似度は小さくなる

RSCCMM-CFのアルゴリズム

Step 1.
共起関係行列$R = \{r_{ij}\}$に
RSCCMM法を適用し，正規化
ユーザーメンバシップ $\tilde{u}_{ci}$ ，
アイテムメンバシップ w_{cj}を算出

Step 2.
アイテムjのユーザーiに対する
推薦度$\hat{r}_{ij}$ を算出

$$\hat{r}_{ij} = \sum_{c=1}^{C} \tilde{u}_{ci} w_{cj}$$

Step 3.
$\hat{r}_{ij}$が閾値η以上の時，推薦を行う

共起関係行列 $R = \{r_{ij}\}$

	アイテム1	アイテム2	アイテム3	アイテム4	アイテム5
ユーザー1	1	0	1	0	0
ユーザー2	1	1	0	0	0
ユーザー3	1	1	1	0	0
ユーザー4	0	0	1	1	0
ユーザー5	0	0	1	0	1

共クラスター1（ユーザー1〜3 × アイテム1〜3）
共クラスター2（ユーザー2〜5 × アイテム2〜3）

正規化ユーザーメンバシップ $\tilde{u}_{ci}$

	ユーザー1	ユーザー2	ユーザー3	ユーザー4	ユーザー5
共クラスター1	1	1/2	1/2	0	0
共クラスター2	0	1/2	1/2	1	1

アイテムメンバシップ w_{cj}

	アイテム1	アイテム2	アイテム3	アイテム4	アイテム5
共クラスター1	3/7	2/7	2/7	0	0
共クラスター2	0	1/4	1/2	1/8	1/8

推薦度 $\hat{r}_{ij}$

	アイテム1	アイテム2	アイテム3	アイテム4	アイテム5
ユーザー1	3/14	15/56	15/56	0	0
ユーザー2	3/14	15/56	11/28	1/16	1/16
ユーザー3	3/7	2/7	2/7	1/16	1/16
ユーザー4	0	1/4	1/2	1/8	1/8
ユーザー5	0	1/4	1/2	1/8	1/8

$\eta = \frac{1}{8}$

推薦の有無

	アイテム1	アイテム2	アイテム3	アイテム4	アイテム5
ユーザー1	1	1	1	0	0
ユーザー2	1	1	1	0	0
ユーザー3	1	1	1	0	0
ユーザー4	0	1	1	1	1
ユーザー5	0	1	1	1	1

構成

1. はじめに
2. 準備
   - クラスタリング（HCM, RCM, RSCM）
   - 共クラスタリング（HCCMM, RCCMM, RSCCMM）
3. RSCCMMに基づく協調フィルタリング（RSCCMM-CF）
4. **数値実験**
5. おわりに

使用データ: NEEDS-SCAN/PANEL

2000年の調査対象の996世帯が家電等を含む18種類の製品を
所有しているか否かを記録したデータ

- **オリジナルデータ**:

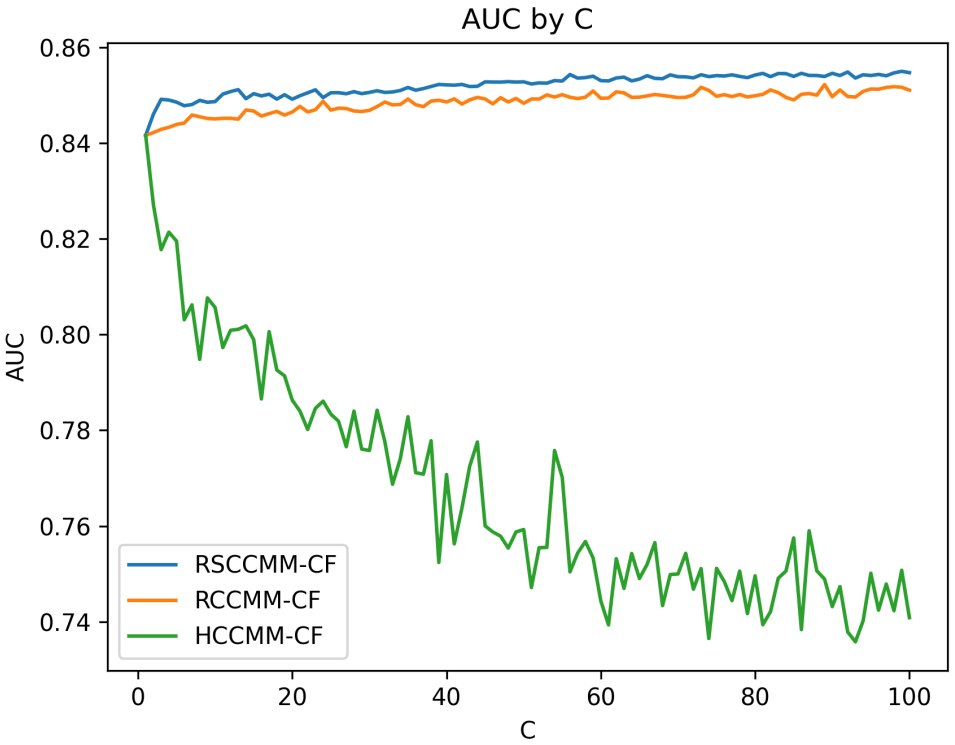
  行列要素 r_{ij} は世帯 i が製品 j を所有している場合に1，所有していなければ0

- **トレーニングデータ**:

  テストデータとして，$r_{ij} = 1$ の要素の中から1000要素選択し，
  全て0に置き換えたデータ

- **評価指標: ROC-AUC**
  - 推薦の閾値を種々変化させ，
    偽陽性率に対する真陽性率をプロットした曲線の下部の面積
  - ランダムな推薦の場合，$AUC = 0.5$ 程度になり，
    1.0 に近づくほど推薦性能が良いとされる

結果: RSCCMM-CF vs RCCMM-CF, HCCMM-CF
(NEEDS-SCAN/PANEL)



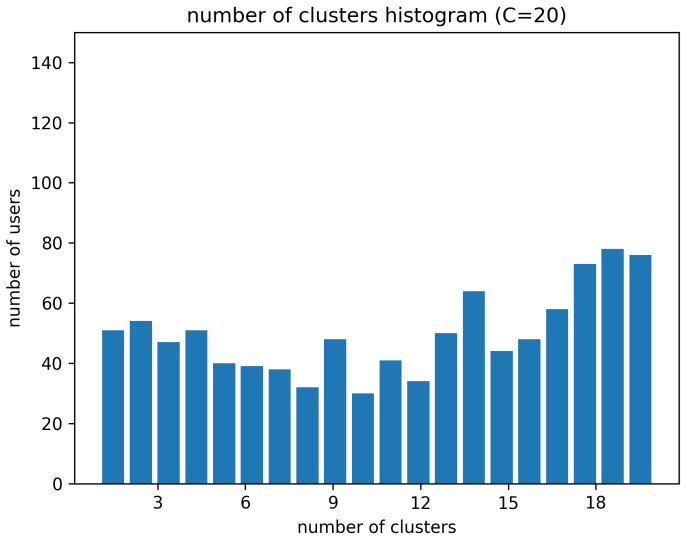
図．クラスター数によるAUCの変化
(**RSCCMM-CF**, **RCCMM-CF**, **HCCMM-CF**)

- 各結果はランダムな初期値に基づく10回試行の平均値
- **RCCMM-CF**はオーバーラップ度合いを調節するパラメータα, βを，**RSCCMM-CF**はδを種々変化させ，最大値を採用

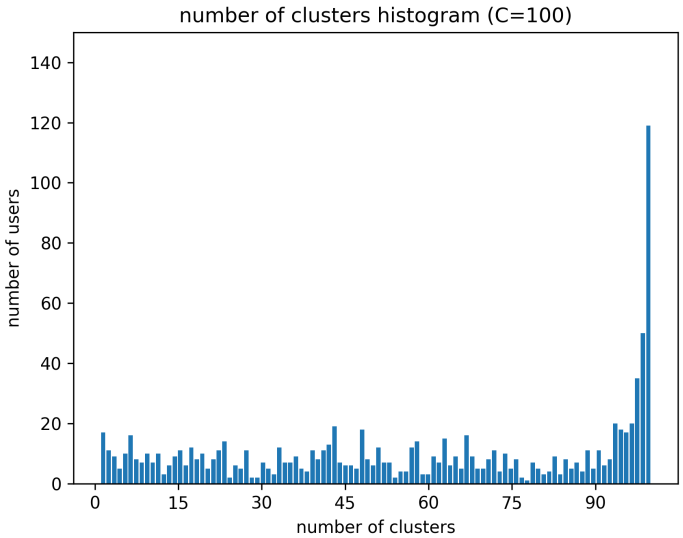
✓ クラスター数の増加に伴い，**RCCMM-CF**と**RSCCMM-CF**はAUCが向上

✓ いずれのクラスター数においても，提案法である**RSCCMM-CF**が高いAUC(推薦性能)を達成

結果: RSCCMM-CFにおけるクラスターオーバラップの変化



図．帰属クラスター数のヒストグラム（$C = 20$）



図．帰属クラスター数のヒストグラム（$C = 100$）

- $C = 100$ ではすべてのクラスターに帰属するユーザーの割合（オーバーラップ）が大きい
- いずれの結果においても，一部のクラスターに帰属するユーザーが一定割合存在する

提案法は重複の多いデータにおいて，平均的な嗜好パターンのユーザーと特殊な嗜好パターンに二分するように作用し，推薦性能を維持できる

使用データ: MovieLens-100k（ML）

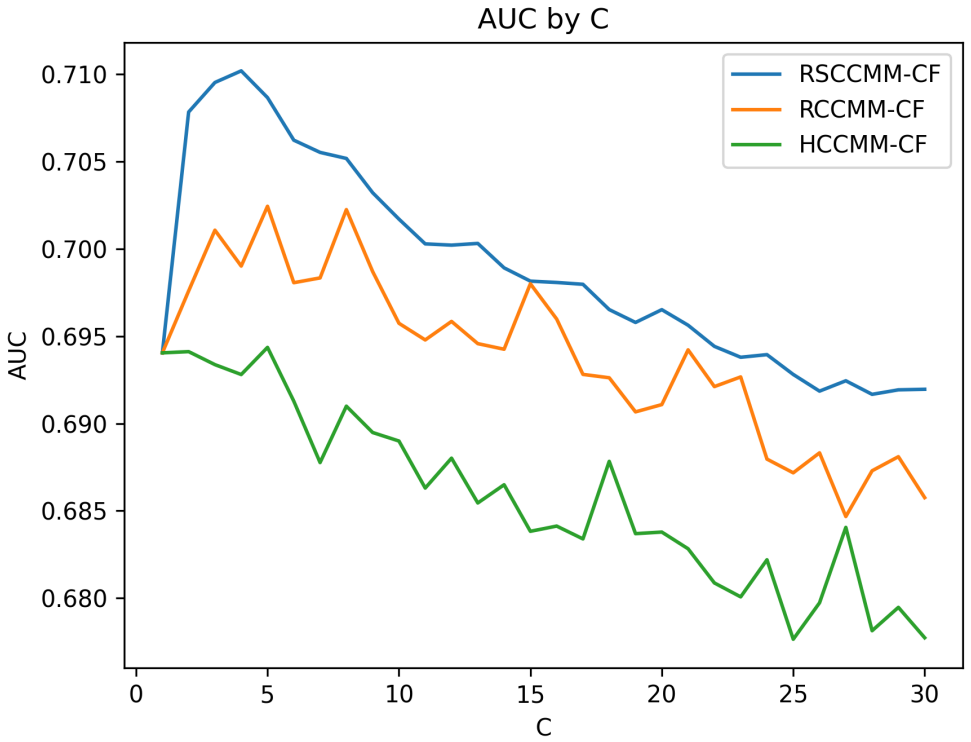
ミネソタ大学のチームによって発表された100,000個の映画の5段階評価を含むデータ

→ 以下の2種類の前処理を加え，二値化を行ったMLデータセットと二値化を行わないMLデータセットを作成し，各々の約10%をテストデータとして未評価値に変換.

データ行列 $R = \{r_{ij}\}$ の前処理の違い

	二値化を行ったML	**二値化を行わないML**
データの抽出	30以上の映画を評価した690ユーザーと，50以上のユーザーに評価された30種類の映画を抽出.	データの抽出は行わなず，すべてのデータを使用
二値化	ユーザーiの映画jに対する評価が **4以上の場合** $r_{ij} = 1$, **3以下の場合** $r_{ij} = 0$.	元の5段階評価のまま使用
未評価値	$r_{ij} = 0.5$に変換.	各ユーザーの**平均評価値**で補完

結果: RSCCMM-CF vs RCCMM-CF, HCCMM-CF
(二値化を行ったMovieLens-100k)



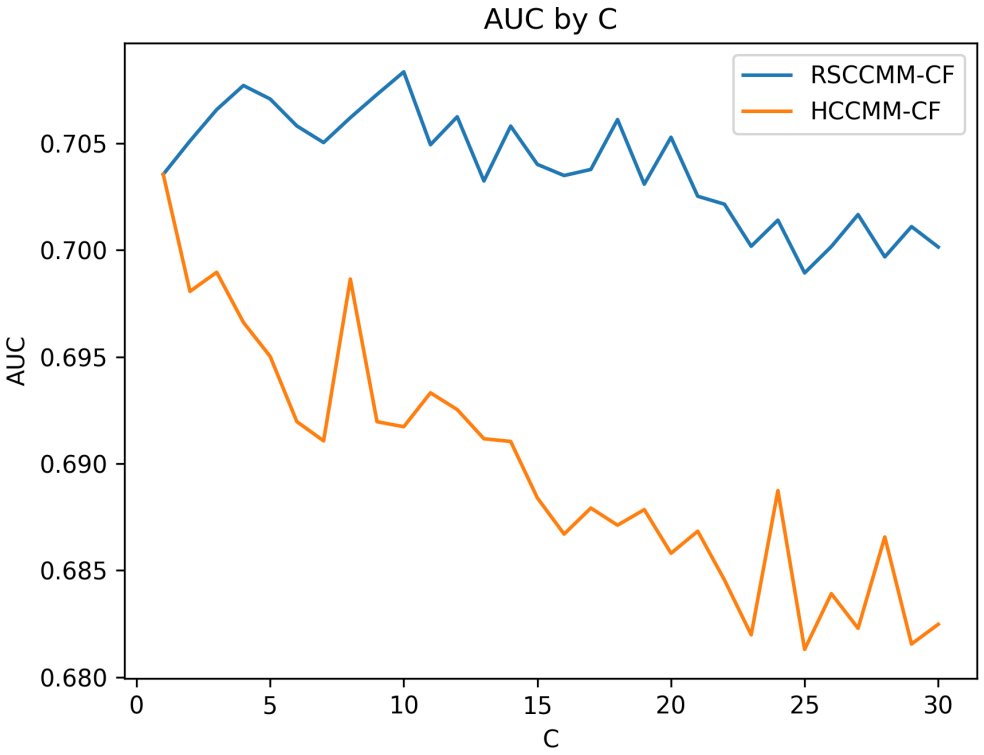
図．クラスター数によるAUCの変化
(**RSCCMM-CF**, **RCCMM-CF**, **HCCMM-CF**)

- 各結果はランダムな初期値に基づく10回試行の平均値
- **RCCMM-CF**はオーバーラップ度合いを調節するパラメータα, βを，**RSCCMM-CF**は δ を種々変化させ，最大値を採用

✓ クラスター数の増加に伴い，全ての手法においてAUCが向上し，最大となったのちに低下

✓ いずれのクラスター数においても，提案法である**RSCCMM-CF**が高いAUCを達成

結果: RSCCMM-CF vs HCCMM-CF
(二値化を行わないMovieLens-100k)



図．クラスター数によるAUCの変化
(**RSCCMM-CF**, **HCCMM-CF**)

- 各結果はランダムな初期値に基づく10回試行の平均値
- **RSCCMM-CF**は δ を種々変化させ，最大値を採用
  (RCCMM-CFは$C = 1$での結果から向上が見られなかったため省略)

✓ クラスター数の増加に伴い，提案法のAUCが向上し，最大となったのちに低下していく

✓ いずれのクラスター数においても，提案法である**RSCCMM-CF**が高いAUCを達成

結果: MovieLens-100kデータの前処理による違い

表: 2種類の前処理によるMLデータセットにおける提案法(RSCCMM-CF)の各AUC

	二値化を行ったML	**二値化を行わないML**
初期クラスター数 $C = 1$でのAUC	0.6940	**0.7036**
	0.0168	**0.0047**
最大AUC (クラスター数)	**0.7108** (C=4)	0.7083 (C=10)

- $C = 1$での結果は，**二値化を行わないML**での結果が高い
- 最大AUCでは**二値化を行ったML**での結果が高い
- 変化量を比較すると，**二値化を行ったML**の方が大きなAUCの向上が見られる

提案法は各ユーザー内でアイテム同士の違いがはっきり現れているデータ（前処理）ほど効果が現れやすい

おわりに

粒状性を考慮したラフ集合ベースの混合多項分布型共クラスタリングに基づく協調フィルタリングを提案し，実世界のデータを用いた数値実験を行うことによって，推薦性能・粒状性の効果を検証した

結果

- 3種類のデータセットにおいて，パラメータを適切に設定することで，従来法と比較して高い推薦性能が得られることを確認

  →ラフ集合理論に基づく粒状性の考慮は共起関係データの協調フィルタリングタスクにおいて有効であることが示唆された
- 各データセット（前処理）における結果の分析から、手法にあった提案法の挙動やより効果のあるデータの特徴を確認

補足: 各データの視覚化

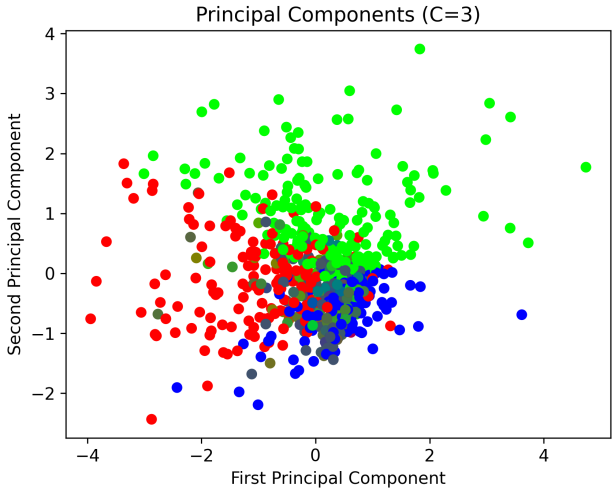


図. 二値化を行ったMLデータセットの視覚化

- 第一主成分：映画に高い評価をつける傾向があると左側
- 第二主成分：評価した映画が少ないと下側

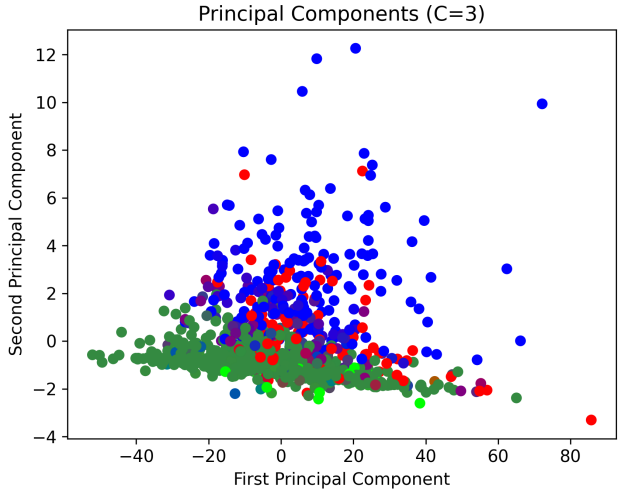


図. 二値化を行わないMLデータセットの視覚化

- 第一主成分：映画に高い評価をつける傾向があると左側
- 第二主成分：評価した映画が少ないと下側

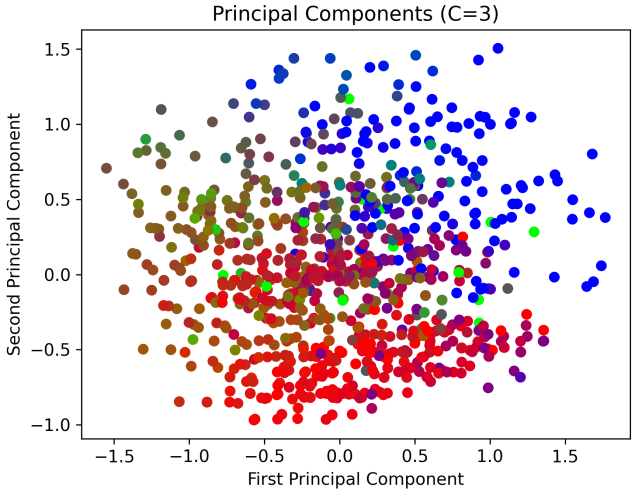


図. NEEDS-SCAN/PANELデータセットの視覚化

- 第一主成分：製品の種類が多いと左側，少ないと右側
- 第二主成分：大規模世帯が下側，小規模世帯が上側

補足：論文抜粋（FCCMM，HCCMM）

- FCCMMの目的関数（参考文献[7]）

$$\text{max.} \quad J_{FCCMM}^{simple} = \sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ci} r_{ij} \log w_{cj} + \lambda_u \sum_{c=1}^{C}\sum_{i=1}^{n} u_{ci} \log \frac{\alpha_c}{u_{ci}},$$

$$\text{s.t.} \quad u_{ci}, w_{cj}, \alpha_c \in (0,1], \forall c, i, j,$$

$$\sum_{c=1}^{C} u_{ci} = 1, \forall i, \sum_{j=1}^{m} w_{cj} = 1, \forall c, \sum_{c=1}^{C} \alpha_c = 1,$$

MMMsのファジィ度の設定を備えたFCCMMの目的関数からファジィ化項を削除することで，HCCMMの目的関数が得られる。

- HCCMMの目的関数（p-3/6）

$$\text{max.} \quad J_{\text{HCCMM}} = \sum_{c=1}^{C}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ci} r_{ij} \log w_{cj}, \quad (11)$$

$$\text{s.t.} \quad u_{ci} \in \{0,1\}, w_{cj} \in (0,1], \forall c, i, j, \quad (12)$$

$$\sum_{c=1}^{C} u_{ci} = 1, \forall i, \ \sum_{j=1}^{m} w_{cj} = 1, \forall c. \quad (13)$$

クラスター c と対象 i の類似度 s_{ci} を次式で定める．

$$s_{ci} = \sum_{j=1}^{m} r_{ij} \log w_{cj}. \quad (14)$$

ここで，$s_{ci} \leq 0$ となる点に注意する．s_{ci} が大きいほど類似度が大きい．

補足: 論文抜粋（RCCMM）

- RCCMM（thesis p4）

Step 1 クラスター数C, クラスターのオーバーラップ度合いを調節するパラメータ α $(\alpha \geq 1)$, β $(\beta \leq 0)$ を設定する.

Step 2 ランダムに C 個の対象をサンプリングし，項目メンバシップ w_{cj} を (15) 式によって初期化する.

$$w_{cj} = \frac{r_{cj}}{\sum_{l=1}^{m} r_{cl}}.$$

$$\tilde{u}_{ci} = \frac{\overline{u}_{ci}}{\sum_{l=1}^{C} \overline{u}_{li}}.$$

Step 2 ランダムに C 個の対象をサンプリングし，項目メンバシップ w_{cj} を (15) 式によって初期化する.

Step 3 対象 i のクラスター c の上エリアに対するメンバシップ $\overline{u}_{ci}$ を以下の式で計算し，(6) 式で正規化メンバシップを求める.

$$\overline{u}_{ci} = \begin{cases} 1 & (s_{ci} \geq \alpha s_i^{\max} + \beta), \\ 0 & (\text{otherwise}). \end{cases} \tag{19}$$

Step 4 項目メンバシップ w_{cj} を更新する.

$$w_{cj} = \frac{\sum_{i=1}^{n} \tilde{u}_{ci} r_{ij}}{\sum_{l=1}^{m} \sum_{i=1}^{n} \tilde{u}_{ci} r_{il}}. \tag{20}$$

Step 5 $\overline{u}_{ci}$ に変化がなくなるまで **Step 3-4** を繰り返す.

補足: 論文抜粋（負の交差エントロピー）

3.1　二項関係の設定

対象空間を粒状化するため，対象間の二項関係を設定する．まず，二項関係を構成するための対象間の類似度を定義する．共起関係データに適した類似度を考えるため，HCCMM 法における対象 i とクラスター c の類似度 ((14) 式) を参考に，対象 i と対象 t の類似度 S_{it} を定義する．項目メンバシップは混合多項分布から派生したものであるため，確率分布を基礎とした類似度を考える．そこで，各対象について共起情報の総和が 1 となるように正規化した $\tilde{r}_{ij}$ を考慮し，各対象の共起情報を確率分布として捉える:

$$\tilde{r}_{ij} = \frac{r_{ij}}{\sum_{l=1}^{m} r_{il}}. \tag{23}$$

対象 i と対象 t の類似度 S_{it} を下記のように定義する:

$$S_{it} = \sum_{j=1}^{m} \tilde{r}_{tj} \log \tilde{r}_{ij}. \tag{24}$$

これは，負の交差エントロピーとみなせる．交差エントロピーは 2 つの確率分布が離れている度合いを示すため，S_{it} は類似度とみなせる．類似度 S_{it} に基づき，二項関係を以下のように設定する:

$$R_{it} = \begin{cases} 1 & (S_{it} \geq \delta)\,, \\ 0 & (\text{otherwise})\,. \end{cases} \tag{25}$$

ここで，$\delta \leq 0$ はラフさを調節するパラメータであり，δ が小さいほど粗い粒状化となり，上近似が拡大し，クラスターのオーバーラップが大きくなる．一般に，S_{it} は非対称であり，R_{it} は対称性を満たさない．

補足：論文抜粋（RSCCMM abstract p1）

2. RSCCMM 法に基づく協調フィルタリング

2.1. RSCCMM 法

RSCCMM 法は対象空間の粒状性を考慮したラフ共クラスタリング手法である．対象空間 U を二項関係 $R \subseteq U \times U$ によって粒状化し，対象のクラスターへの帰属を粒ごとに判定することで，帰属の不確実性を取り扱い，クラスターのオーバーラップを実現する．R の設定の仕方によって多様な粒状化がなされ，多様な分類が可能となる．

対象 i と項目 j の共起度を r_{ij}，対象 i の共クラスター c に対するメンバシップを u_{ci}，項目 j の共クラスター c に対するメンバシップを w_{cj}，対象数を n，項目数を m として，RSCCMM 法のアルゴリズムを以下に示す．

Step 1 クラスター数 C，二項関係 $R \subseteq U \times U$ を設定する．

Step 2 項目メンバシップ w_{cj} を次のように初期化する．ランダムに C 個の対象をサンプリングし，それぞれ総和が 1 となるように正規化する．

$$w_{cj} = \frac{r_{cj}}{\sum_{l=1}^{m} r_{cl}}. \tag{1}$$

Step 3 クラスター c と対象 i の類似度を s_{ci} とし，対象 i のクラスター c に対するメンバシップ u_{ci} を，最も類似度の大きいクラスターとの類似度 $s_i^{\max}$ に基づいて計算

$$s_{ci} = \sum_{j=1}^{m} r_{ij} \log w_{cj}, \tag{2}$$

$$s_i^{\max} = \max_{1 \le c \le C} s_{ci}, \tag{3}$$

$$u_{ci} = \begin{cases} 1 & (s_{ci} \ge s_i^{\max}), \\ 0 & (\text{otherwise}). \end{cases} \tag{4}$$

Step 4 対象 i のクラスター c に対するラフメンバシップ値 μ_{ci}^R と上近似に対するメンバシップ $\overline{u}_{ci}$，正規化メンバシップ値 $\tilde{u}_{ci}$ を順に以下の式で計算する．

$$\mu_{ci}^R = \frac{\sum_{t=1}^{n} R_{it} u_{ct}}{\sum_{t=1}^{n} R_{it}} \tag{5}$$

$$\overline{u}_{ci} = \begin{cases} 1 & (\mu_{ci}^R > 0), \\ 0 & (\text{otherwise}). \end{cases} \tag{6}$$

$$\tilde{u}_{ci} = \frac{\overline{u}_{ci}}{\sum_{l=1}^{C} \overline{u}_{li}}. \tag{7}$$

Step 5 項目メンバシップ w_{cj} を以下の式で更新する．

$$w_{cj} = \frac{\sum_{i=1}^{n} \tilde{u}_{ci} r_{ij}}{\sum_{l=1}^{m} \sum_{i=1}^{n} \tilde{u}_{ci} r_{il}}. \tag{8}$$

Step 6 u_{ci} に変化がなくなるまで **Step 3-5** を繰り返す．