

はじめに

ユーザーベース協
調フィルタリング

関係性マイニング

提案手法

評価実験

おわりに

関係性マイニングと協調フィルタリングを 用いた情報推薦手法

滝沢 光介

富山県立大学 電子・情報工学科

June 11, 2021

協調フィルタリングの現状

ユーザーベース協調フィルタリングは、ユーザー間の類似度を用いて推薦を行う手法であるが、推薦の精度と多様性がトレードオフの関係になる問題が挙げられる。

本研究の目的

関係性マイニングを用いてユーザーの感情的な価値判断の傾向を読み取り、ユーザーの選好パターンを解析し、それを協調フィルタリングに反映させることで推薦精度と多様性の両立を図る。

ユーザーベース協調フィルタリングとは

ユーザー間の類似度を求め、類似度の高いユーザー抽出し、対象ユーザーの未評価アイテムに対する評価値予測を行う手法。

ユーザ	アイテム					
	A	B	C	D	E	F
ユーザ1	3	5	1	*	5	*
ユーザ2	3	4	*	1	5	5
ユーザ3	4	*	2	4	4	3
ユーザ4	2	4	2	5	4	3
ユーザ5	*	4	4	3	*	1

Table 1: 評価行列

ピアソン相関係数

$$sim(a, b) = \frac{\sum_{p \in P_{a,b}} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P_{a,b}} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P_{a,b}} (r_{b,p} - \bar{r}_b)^2}} \quad (1)$$

a : ユーザー a

b : ユーザー b

P : ユーザー集合

$r_{i,j}$: ユーザー i によるアイテム j への評価値

評価値の予測

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)} \quad (2)$$

N : 最近傍ユーザー集合

関係性マイニングとは

ラフ集合論に基づき, 属性間の相互関係を表す新たな属性を相互関係表にまとめ, 項目間の比較により特徴を抽出する手法.

情報表と決定表

ラフ集合論において, データは情報表または決定表で表され, 以下のよう定義される.

$$S = (U, AT, \{V_a | a \in AT\}, R_{AT}, \rho) \quad (3)$$

U : 対象の有限集合

AT : 属性の有限集合

V_a : 属性 a の集合

$R_{AT} = \{\{R_a\} | a \in AT\}$: 各属性 $a \in AT$ の集合 V_a で定義された二項関係の集合 $\{R_a\}$ の集合

$\rho(U * AT \rightarrow V)$: 対象 $x \in U$ の属性 $a \in AT$ における値

$\rho(x, a) \in V_a$ を表す

関係性マイニングの定義 (2)

6/19

相互関係決定表

属性の関係性を表す新たな属性を追加し, それらをまとめた表を相互関係決定表と呼び, 以下の式で定義される.

$$S_{int} = (U, AT_{int}, V \cup \{0, 1\}, R_{int}, \rho_{int}) \quad (4)$$

$$R_{int} = R_{AT} \cup \{ \{ R_{a_i * b_i} \} | \exists a_i, b_i \in C \} \cup \{ \{ = \} | \text{For each } aRb \} \quad (5)$$

$R_{a_i * b_i}$ は属性 a_i, b_i の属性値の関係を表す集合である.

$$AT_{int} = AT \cap \{ aRb | \exists R \in R_{a*b}, R(a, b) \neq \emptyset \} \quad (6)$$

$$R(a, b) = \{ x \in U | (\rho(x, a), \rho(x, b)) \in R \} \quad (7)$$

関係性マイニングの定義 (3)

7/19

相互関係決定表 (2)

$\rho_{int} : U * AT_{int} \rightarrow V \cup \{0, 1\}$ は以下の式で定義され, 対象ユーザーの関係性属性を含む全ての属性における属性値を表す.

$$\rho_{int}(x, c) = \begin{cases} \rho(x, c), & \text{if } c \in AT, \\ 1, & c = aRb \text{ and } x \in R(a, b), \\ 0, & c = aRb \text{ and } x \notin R(a, b), \end{cases} \quad (8)$$

はじめに

ユーザーベース協
調フィルタリング

関係性マイニング

提案手法

評価実験

おわりに

step1:関係性属性の作成

ユーザーの評価済みアイテムからアイテムの対 (a, b) を n 個選択する. このとき, アイテム b の評価値がアイテム a の評価値より閾値 Dif 以上にならないようにする.

$$Dif = \frac{[\max(Range) - \min(Range)]}{2} \quad (9)$$

得られた対について各ユーザーに対する関係性属性の値を以下の式で定義する.

$$\rho_{int}(u, a \leq_{Dif} b) = \begin{cases} 0, & \text{if } \rho(u, b) - \rho(u, a) < Dif, \\ 1, & \rho(u, b) - \rho(u, a) \geq Dif, \\ *, & \rho(u, a) = * \vee \rho(u, b) = * \end{cases} \quad (10)$$

選好パタンの抽出例

対象ユーザーの評価の差が大きいアイテムの関係性を表し、他のユーザーと比較することで類似度を測る。

ユーザ	アイテム					
	A	B	C	D	E	F
ユーザ1	3	5	1	*	5	*
ユーザ2	3	4	*	1	5	5
ユーザ3	4	*	2	4	4	3
ユーザ4	2	4	2	5	4	3
ユーザ5	*	4	4	3	*	1

Table 2: 評価値行列

ユーザ	$A \leq_2 B$	$A \leq_2 E$	$C \leq_2 A$	$C \leq_2 B$	$C \leq_2 E$
ユーザ1	1	1	1	1	1
ユーザ2	0	1	*	*	*
ユーザ3	*	0	1	*	1
ユーザ4	1	1	0	1	1
ユーザ5	*	*	*	0	*

Table 3: 選好パターン

提案システムの流れ (2)

10/19

step2: ユーザー類似度の計算

類似度として対象ユーザーと他のユーザーの間で関係性属性の値が一致する割合を用いる.

$$sim(x, u) = \frac{|\{a \leq_{Dif} b \in AT_{int} | \rho(x, a \leq_{Dif} b) = \rho(u, a \leq_{Dif} b)\}|}{n} \quad (11)$$

step3: 予測評価値の計算

類似度の高いユーザー k 人を最近傍ユーザーとし, 式 (2) を用いて予測評価値を求める.

はじめに

ユーザーベース協
調フィルタリング

関係性マイニング

提案手法

評価実験

おわりに

使用データセット

MovieLens100k データセット, *MovieLens1m* データセット, *Jester* の3つのデータセットを使用する. それぞれのデータセットに対して, 一般的なユーザーベース協調フィルタリングと提案手法を実装し比較を行う.

データセット	ユーザ数	アイテム数	評価件数
MovieLens_100k	943	1682	100000
MovieLens_1m	6040	3952	1000209
Jester	24983	100	1810455

Table 4: 使用するデータセット

MAE

推薦制度の評価指標として,*MAE* を利用する.*MAE* は以下の式で定義される.

$$MAE = \frac{1}{|U|} \sum_{u \in U} \left(\frac{1}{|O_u|} \sum_{i \in O_u} |p_{u,i} - r_{u,i}| \right) \quad (12)$$

$p_{u,i}$: アイテムの予測評価値

$r_{u,i}$: 実際の評価値

O_u : 対象ユーザーの評価済みアイテムの関係性属性の作成に用いられなかったアイテムの集合

Coverage

推薦の多様性を表す評価指標として, *Coverage* を利用する.
Coverage は以下の式で定義される.

$$Coverage = \frac{1}{|U|} \sum_{u \in U} (100 * \frac{|I_u^c \cap Z_u|}{|I_u^c|}) \quad (13)$$

I_u^c : 対象ユーザーの未評価アイテムの集合

Z_u : 最近傍ユーザーが評価を与えているアイテムの集合

実験条件

各データセットについてそれぞれの手法を用いて最近傍ユーザーが 20 人, 30 人, 40 人, 50 人と変化させた際の MAE と $Coverage$ を求める。また, *MovieLens* では関係性属性の個数を 10, 20 と変化させ, *Jester* では関係性属性の数を 5, 10 個の比較を行った。

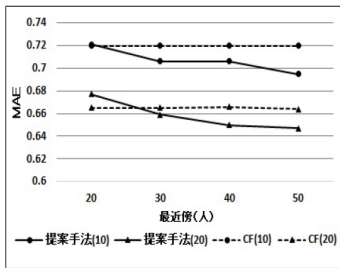


図 1: MovieLens100k での MAE 比較

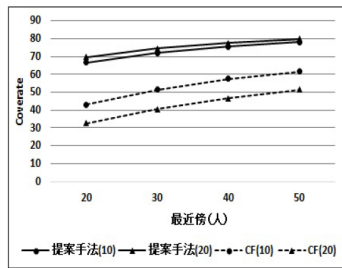


図 2: Coverage 比較

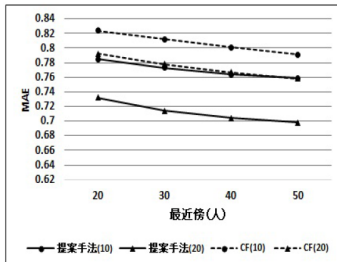


図 3: MovieLens1m での MAE 比較

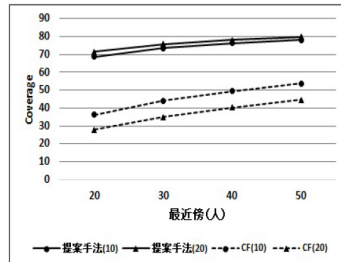


図 4: Coverage 比較

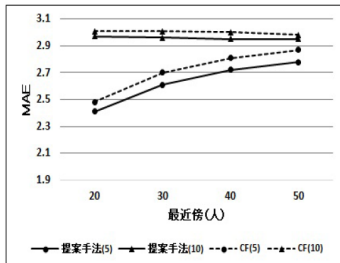


図 5: Jester での MAE 比較

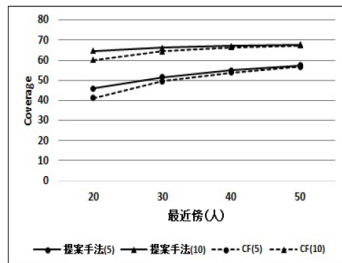


図 6: Coverage 比較

適合率と再現率

2つの評価指標以外にも, *MovieLens100k* データセットに対しては, 適合率 (*Precision*) と再現率 (*Recall*) の評価も行った. 適合率と再現率は以下の式で表せる.

$$Precision = \frac{TP}{TP + FP} \quad (14)$$

$$Precision = \frac{TP}{TP + TN} \quad (15)$$

TP: 対象ユーザーに対して推薦されたアイテムで対象ユーザーが実際に高評価を与えているアイテムの個数

FP: 対象ユーザーに対して推薦されたアイテムで対象ユーザーが低評価あるいは未評価のアイテム個数

TN: 対象ユーザーに推薦されなかったアイテムで対象ユーザーが高評価を与えているアイテムの個数

実験条件

それぞれの推薦手法で, 推薦アイテムの個数を 30,50,70,90 と変化させ適合率と再現率を求めた. 最近傍ユーザーの人数は 50 人で固定する.

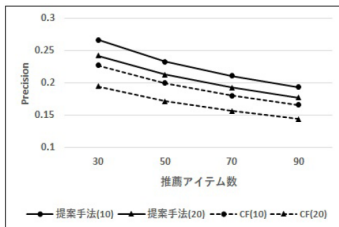


図 7: 適合率の比較

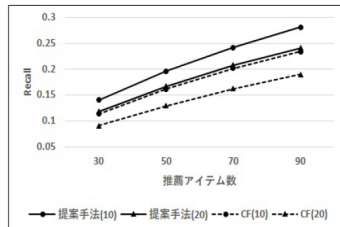


図 8: 再現率の比較

実装したこと

- 1 多様性と推薦精度を両立させた情報推薦システムの実装
- 2 提案手法のユーザーベース協調フィルタリングに対する有効性の検証

今後の課題

推薦精度と多様性を両立させた手法は様々であり, それらの手法との比較が必要