Vol.2018 No.9,B1-3
2019/2/21

ロバスト性の向上に向けた複数回クラスタリングによる協調フィルタリング手法の提案

平野瑞己[†1]　張建偉[†1]

通販サイトやレビューサイトなどで，多数のユーザの評価を元にアイテムを推薦するために協調フィルタリングが広く用いられている．しかし，この手法にはいくつかの問題が存在し，その一つとして特定のアイテムの評価値を歪める攻撃の存在が挙げられる．本研究では対象データにクラスタリングを複数回施し，クラスタごとに評価値を予測することで攻撃に対する影響を軽減しつつ予測を行う手法を提案する．また，実際の評価値と予測値との誤差を用いた評価手法を用いて，予測手法の有用性を調査し，さらに攻撃前後の誤差を比較することによって攻撃に対するロバスト性を調査する．

1. はじめに

近年では通販サイトやレビューサイトなどが広く利用されているが，一般にサイト内で取り扱うアイテム数や，利用するユーザ数は膨大である．そのためユーザが求めるアイテムを提供するための推薦システムは，利用するユーザ、サイト運営者双方にとって有用である．その中でも本研究での研究対象である協調フィルタリングは推薦システムに用いられる代表的な手法の一つであり，ユーザがアイテムに行った評価そのものをユーザの特徴として扱えるという利点が存在する．

しかし，協調フィルタリングにはいくつかの欠点があり，その一つに推薦結果を歪める攻撃[1)]が挙げられる．協調フィルタリングを用いる推薦システムでは，特定のアイテムに高評価，あるいは低評価をつけた攻撃ユーザデータを複数追加することで，推薦結果を歪めることができる．この欠点の解消は推薦システムによる推薦の信頼性を高めるために重要である．

協調フィルタリングでは自分と似たユーザを探し出し，それらのユーザが好むアイテムを推薦するため，協調フィルタリングを行う際，あらかじめ似ているユーザをクラスタ分けしておくことで予測精度が向上すると予想できる．しかし，クラスタを細分化しすぎることでかえってこの攻撃の影響を増大させることも考えられる．そこで本研究では分割されたクラスタに対して再度クラスタリングを行い，クラスタ内のデータ数を多くした後に評価値の予測を行う予測手法を提案する．また，本予測手法を評価するために，実際の評価値との誤差を用いる評価手法を提案し，さらに攻撃前後での予測誤差を用いて，攻撃に対するロバスト性を調査する．

†1 岩手大学
Iwate University

2. 基本概念

本項では研究対象となるシステムや攻撃の概念について解説する．

2.1 推薦システム

1 節でも述べたように，一般的な通販サイトなどでは多種多様かつ膨大な数のアイテムを取り扱っているため，サイトを利用するユーザが自力で求めるアイテムを探し出すのは非常に困難である．そのためにユーザが求めるアイテムを推薦するシステムはユーザに対してはもちろん，サイトの利便性を高める点において，サイト運営者に対しても有用である．

推薦システムに用いられる代表的な手法は 2 つあり，それぞれ内容ベースフィルタリングと協調フィルタリングである．内容ベースフィルタリングは，あらかじめサイト内で扱うアイテムについての特徴をデータとして管理し，利用者から得た嗜好データに合うものを推薦する手法である．一方の協調フィルタリングは,「自分と似た嗜好のユーザが好むアイテムは，自分も好むだろう」という仮定を元に，アイテムに対しての評価から自分と似たユーザを探し出し，それらのユーザが好むアイテムを推薦する手法である．

本研究では協調フィルタリングを取り扱う．

2.2 協調フィルタリング

協調フィルタリングではアイテムの特徴を全く用いず，ユーザのアイテムへの評価のみから類似ユーザを探すことができる．協調フィルタリングを用いた評価予測の簡単な例を示す．

表 1 は 4 人のユーザ，5 つのアイテムについて，最低 1，最高 5 の 5 段階評価を行っていることを表している．空欄の場合はユーザは対象アイテムに評価を行っていない．このときにユーザ 1 がアイテム 5 をどのように評価するかを予測したい．

協調フィルタリングはユーザの評価から類似ユーザを見つけ出す手法であるため，評価予測の際はまず各ユーザが同一アイテムに付けた評価値を比較する．今回の例の場合ユーザ 2 がほぼ同じ評価である半面，ユーザ 3 とは評価値の差が大きい．ユーザ 4 とは似通った評価値をつける部分もあるが，大きく異なる評価値をつける部分もあるため，ユーザ 1 と一

表 1　ユーザのアイテムに対する評価例

	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1	5	2	3	5	**?**
ユーザ2	4		3	4	5
ユーザ3	2	5		1	2
ユーザ4	5	2	1	1	3

番よく似たユーザはユーザ 2 と見ることができる．実際に類似のユーザを見つける時には，互いに評価を行っている全てのアイテムからユーザ評価の相関係数を計算し，相関係数の高いものを類似ユーザとする．

類似ユーザを見つけだした後は，対象アイテムの評価値を参照する．これは前項で述べたとおり，自分と似たユーザが高評価するアイテムは自分も高評価し，逆に似たユーザが低評価するアイテムは自分も低評価するだろうという仮定から来ている．ユーザ 2 はアイテム 5 に最高評価である「5」の評価をしているため，ユーザ 1 もアイテムを気に入るだろうと予測でき，アイテム 5 はユーザ 1 に推薦すべきアイテムとなる．実際に予測する際には，類似のユーザを複数人見つけ出し，彼らの評価の類似度により評価に重み付けして予測値を算出する．詳細については 3.4 節にて述べる．

2.3　協調フィルタリングへの攻撃

協調フィルタリングはユーザの評価のみで推薦を行うことができるという特徴があるが，この特徴による欠点として，例えば誰も評価していないアイテムは推薦できない，ユーザの評価したアイテム数が少ない場合に正確な評価ができない (cold-start 問題) といったものが挙げられる．本研究で採り上げる攻撃もまた，この特徴による問題の 1 つである．

例えば，同様の商品を取り扱う A 社と B 社があり，A 社の商品が人気になる，つまりは多くのユーザに高評価されることは，B 社にとっては好ましくないことであろう．このとき B 社は先述の協調フィルタリングの特徴を悪用し，A 社の商品の評価を下げる，もしくは B 社の商品の評価を上げるようなデータを持つユーザを作り出すことによって，協調フィルタリングによる推薦に影響を与えることができる．

表 2 は 2.2 節で示した表 1 に，攻撃用のユーザを追加したもので，このユーザは類似ユーザに選ばれた際にアイテム 5 の評価が低くなるように，アイテム 5 の評価が最低値の 1 となっている．

表 2　ユーザからの攻撃の例

	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1	5	2	3	5	**?**
ユーザ2	4		3	4	5
ユーザ3	2	5		1	2
ユーザ4	5	2	1	1	3
攻撃ユーザ	5	2	3	5	1

このデータを用いてユーザ 1 のアイテム 5 の評価を予測すると，アイテム 5 以外の評価がすべて一致しているため，前節で類似ユーザに選ばれたユーザ 2 に代わり，攻撃ユーザが選ばれることになる．するとユーザ 1 のアイテム 5 への予測値は攻撃ユーザが評価した「1」となり，これは前節での予測とは逆に推薦すべきでないアイテムであると予測されることを示している．攻撃によって評価が歪められることは推薦システムの評価の信頼性に関わるため，この攻撃への対策は重要となる．

3. 予 測 手 法

3.1　研究の趣旨

2.2 節で述べたように，協調フィルタリングはユーザに推薦すべきアイテムを，そのユーザに似ているユーザがアイテムに行った評価から予測するシステムである．そのため予測の際にはあらかじめ似ているユーザ同士を同一クラスタに分類をしておくことで予測精度の向上が期待でき，また，攻撃ユーザが同一のクラスタに分類されれば，そのほかのデータは攻撃による影響を受けずに済む．

しかし，攻撃ユーザが多くの一般のユーザに似ている場合，ユーザの分類を行うことでかえって攻撃の影響を増大させることも考えられる．そこで本研究では，分割されたクラスタに対して再度クラスタリングを行いクラスタ内のデータを大きくすることによって，精度を維持，もしくは向上させつつ攻撃の影響を抑えることを目指す．

3.2　予測手法の概要

予測手法ではデータセットにクラスタリングを行い，ユーザをいくつかのクラスタに分割した後，分割したクラスタ内のデータからクラスタの代表点を計算し，代表点を用いて再度クラスタリングを行うことで，分割したクラスタを結合する．そしてクラスタ内でユーザの

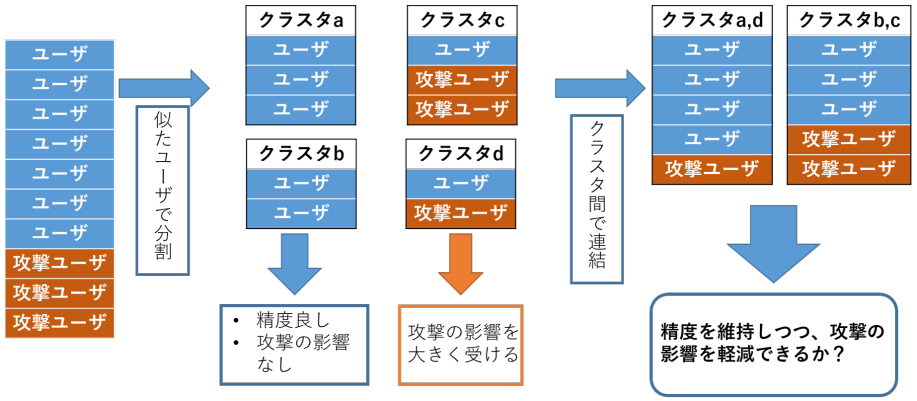

図 1 予測手法の流れ

類似度を計算し，その類似度とユーザが付けた評価を用いて評価値を予測する．予測手法の流れを図 1 に示す．

3.3 クラスタリング

推薦システム内のユーザを類似のグループに分けるため，データセットにクラスタリングを行う．はじめに K 平均法クラスタリング[2] でユーザをクラスタに分類し，次に各クラスタに所属するデータの代表点を用いて再度クラスタリングを行う．この時推薦システム内のあるクラスタ C に帰属するユーザを $U_i(i = 1, 2, \ldots, m)$，アイテムを $I_j(j = 1, 2, \ldots, n)$，ユーザ U_i がアイテム I_j に行った評価スコアを R_{ij} としたとき，クラスタ $C = (c_1, \ldots, c_j, \ldots, c_n)$ における c_j は，以下のように表せる．

$$c_j = \frac{\sum_{i=1}^{m} R_{ij}}{m} \tag{1}$$

また，クラスタ数は 1 回目は 20-100，2 回目は 2-(1 回目のクラスタ数の半分) としている．これは 1 回目のクラスタリングである程度のクラスタ数を確保し，2 回目のクラスタリングでクラスタサイズが大きくなることを担保するためである．

3.4 予測値の計算

ユーザデータに対して協調フィルタリングによるユーザ間類似度を計算し，さらに計算した類似度から各アイテムの評価の予測値を計算する．

類似度の計算にはピアソンの積率相関係数を用いる．データセット内のユーザ U_a, U_b がアイテム I_j に付けた評価を R_{aj}, R_{bj}，ユーザがつけた評価値の平均を $\bar{R_a}, \bar{R_b}$ とすると，ユーザ U_a, U_b のピアソン積率相関係数は以下のようになる．

$$r_{ab} = \frac{\sum_{j=1}^{n}(R_{aj} - \bar{R_a})(R_{bj} - \bar{R_b})}{\sqrt{\sum_{j=1}^{n}(R_{aj} - \bar{R_a})^2}\sqrt{\sum_{j=1}^{n}(R_{bj} - \bar{R_b})^2}} \tag{2}$$

また，ユーザ U_i のアイテム I_j に対する予測値 P_{ij} は，類似度の高い上位 l 人のユーザ $U_k(k = 1, \ldots, l)$ がアイテム I_j に付けた評価値 R_{kj} に類似度による重み付けを行い，以下のように計算する．

$$P_{ij} = \frac{\sum_{k=1}^{l} r_{ki} R_{kj}}{\sum_{k=1}^{l} r_{ki}} \tag{3}$$

この予測値の計算はクラスタリングを行わない場合は全てのデータ，クラスタリングを行う場合は同一クラスタに属するデータのみを用いて行う．

4. 評 価 手 法

4.1 評価手法の概要

本研究の目標は予測評価の精度を維持，もしくは向上させつつ，攻撃による影響を抑えることである．そのために実際に評価が行われている部分に協調フィルタリングによる評価値予測を行い，実際の評価値と予測値との誤差を計測することで予測手法の精度を評価する．また，攻撃前後でそれぞれ誤差を計測した後に，さらにその誤差の差分を計算することで，攻撃に対するロバスト性を分析する．このときの誤差の差分は，攻撃前後での協調フィルタリングの予測値の変化に等しい．

4.2 誤 差 評 価

協調フィルタリングによる評価予測を行うとき，既に評価されている部分に予測を行うと，その部分で予測が行われていないと仮定した状態で評価値予測を行うため，実際の評価値と予測値との誤差を測定することができる．

そこで，はじめに攻撃前の実際の評価値と予測値との誤差を測定し，評価する．この誤差評価の流れを図 2 に示す．この測定を行う理由としては，推薦システムは攻撃を追加していない状態でも高精度な予測が行われる必要があることと，攻撃の影響を調査するために攻撃前後の予測値の変化を用いることが挙げられる．

誤差の指標としては MAE，RMSE を用いる．これらの指標は予測値 P と実際の評価値 R，評価が行われているデータ数 N を用いて以下のように表せる．

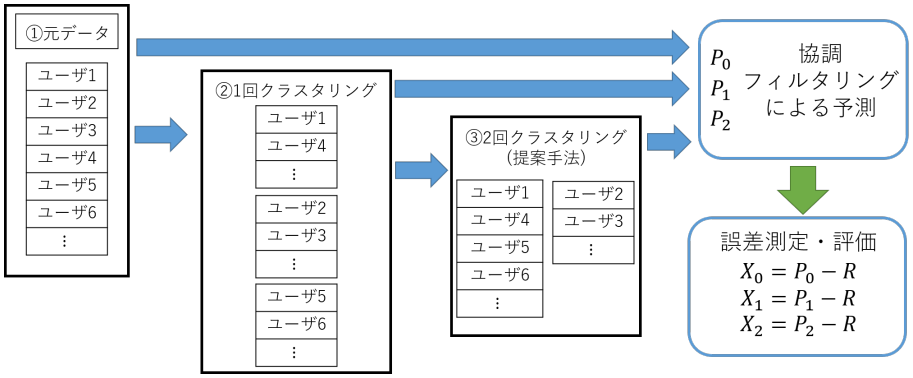

図 2 誤差評価 (攻撃前) の流れ

$$MAE = \frac{\sum_{k=1}^{N} |P_k - R_k|}{N} \tag{4}$$

$$RMES = \sqrt{\frac{\sum_{k=1}^{N} (P_k - R_k)^2}{N}} \tag{5}$$

また，例えば 1 回目のクラスタ数が 100 だった場合，2 回目のクラスタ数は 2 から 50 までと幅がある．よって，比較のために 2 回クラスタリング時は誤差の平均値，最大値，最小値をそれぞれ取っている．

4.3 攻撃に対するロバスト性評価

次に攻撃を追加したデータセットを作成し，誤差評価と同様に実際の評価値と予測値との誤差を測定した後，攻撃前後の誤差の差分 (攻撃による予測値の変化) から攻撃に対するロバスト性を評価する．ロバスト性評価の流れを図 3 に示す．

協調フィルタリングへの攻撃は Mobasher らの研究[3] により定義されている以下のものを用いた．

ランダム攻撃 ランダムに選んだアイテムに，ランダムな評価値を与える．

平均攻撃 ランダムに選んだアイテムに，アイテムの評価の平均値を与える．

バンドワゴン攻撃 人気なアイテムに高評価を与える．またそれとは別にランダムに選んだアイテムにランダムな評価値を与える．

本研究では特定アイテムの評価を上げる攻撃を対象とし，攻撃をユーザデータに追加する際には，攻撃対象のアイテムを無作為に抽出し最高評価を与えた．また攻撃ユーザの数 (アタックサイズ) と，攻撃ユーザが攻撃対象以外に評価するアイテムの数 (フィラーサイズ) を変更しつつ追加した．

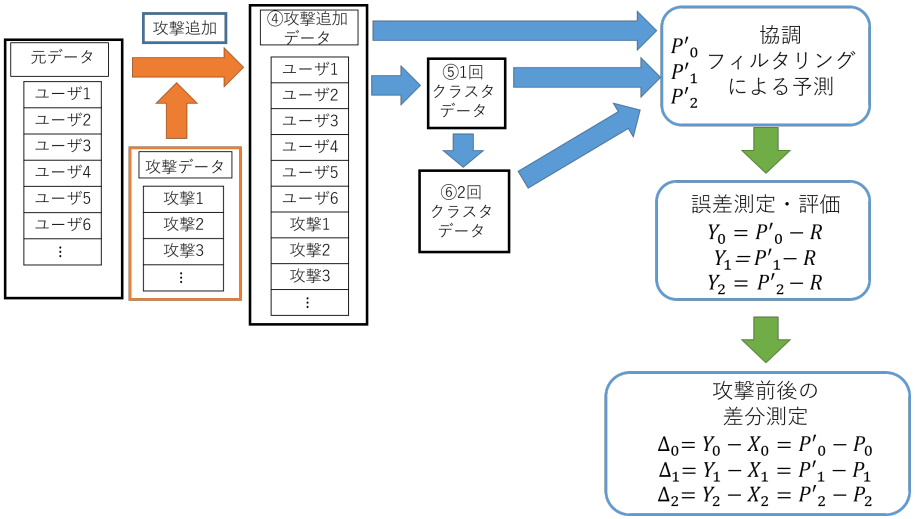

図 3 ロバスト性評価 (攻撃前後の誤差差分) の流れ

最後に，攻撃前後の評価値と予測値との誤差の差分を計算する．この差分は攻撃による評価の変化を表すため，差分が小さいほど攻撃に対するロバスト性が高いと見ることができる．また，4.2 節で示したように本研究では実際の評価値と予測値との差を測定するが，データセット内の評価値は変化しないため，攻撃前後の誤差の差分は攻撃による予測値の変化量の平均に等しい．

5. 評 価 実 験

5.1 使用データセット・実験設定

本実験では Movielens100K[⋆1] をデータセットとして用いる．このデータセットには 943 のユーザが 1,682 の映画アイテムに対して，1-5 までの範囲で付けた評価が 10 万件含まれている．また，1 人のユーザは最低 20 の映画について評価を行っている．

⋆1 https://grouplens.org/datasets/movielens/100k

攻撃に対するロバスト性を測定する際には，すべてのアイテムからランダムに 50 のアイテムを攻撃対象として選び，最高評価を付けた．その後，4.3 節に示した攻撃ごとに攻撃対象に選ばれなかったアイテムを選び，攻撃プロファイルを作成した．

5.2 攻撃前の誤差評価

本研究では攻撃の追加前後の誤差とそれらの差分をクラスタリングを行う回数ごとに測定しているため，実験結果は表 3 のように 9 種類に分類される．実験では，はじめに攻撃前の予測誤差を計測し，クラスタリングの回数ごとに比較を行った．これは表 3 の X_0 から X_2 に当たる．指標 MAE の結果を表 4，指標 RMSE の結果を表 5 に示す．この実験ではクラスタリングなし，1 回クラスタリング，2 回クラスタリングのそれぞれの場合で協調フィルタリングによる評価を行い，実際の評価値との誤差を計算する．その後，それぞれの誤差の数値を比較し，1 番小さいものを太字，2 番目に小さいものを下線で表記している．

結果を見ると，どちらの指標においても 1 回クラスタリング時の誤差が小さく，次に 2 回クラスタリング，クラスタリングなしと続いた．また，2 回クラスタリングの誤差の最大値は常にクラスタリングなしの誤差より小さいが，誤差の最小値でも 1 回クラスタリングの誤差より下回ることはなかった．これはクラスタ数が多いほど予測精度が向上するという予想通りの結果である．

5.3 攻撃後の誤差評価

次に 4.3 節に示した攻撃用のユーザデータを追加し，協調フィルタリングでの予測値の誤差を測定した．これは表 3 の Y_0 から Y_2 に当たる．表 6 から表 11 にそれぞれの攻撃と指標で測定した誤差を示す．実験前はクラスタ数を大きくするほど攻撃の影響を受けやすいという予想から，2 回クラスタリングを行う手法の誤差が小さくなると予想していた．しかし，結果はすべての攻撃，指標において 1 回クラスタリングを行う手法の誤差より下回るものはなかった．また，この誤差はクラスタ数が多いほど小さくなる傾向が見られる．このことから，協調フィルタリングの予測誤差を小さくすることにクラスタ数が大きく関係していると

表 3 実験の分類

	クラスタリングなし	**1回クラスタリング**	**2回クラスタリング**
攻撃なし	$X_0 = P_0 - R$	$X_1 = P_1 - R$	$X_2 = P_2 - R$
攻撃あり	$Y_0 = P'_0 - R$	$Y_1 = P'_1 - R$	$Y_2 = P'_2 - R$
攻撃前後の差分	$\Delta_0 = Y_0 - X_0 = P'_0 - P_0$	$\Delta_1 = Y_1 - X_1 = P'_1 - P_1$	$\Delta_0 = Y_2 - X_2 = P'_2 - P_2$

表 4 誤差測定：攻撃前：MAE

クラスタなし	**1回クラスタリング**		**2回クラスタリング(提案手法)**			
MAE (X_0)	クラスタ数	MAE (X_1)	クラスタ数	MAE(X_2)		
				平均値	最大値	最小値
0.613	20	**0.455**	2-10	0.569	0.606	0.530
0.613	30	**0.431**	2-15	0.575	0.602	0.546
0.613	40	**0.399**	2-20	0.560	0.600	0.533
0.613	50	**0.377**	2-25	0.558	0.603	0.528
0.613	60	**0.362**	2-30	0.541	0.608	0.496
0.613	70	**0.353**	2-35	0.533	0.606	0.489
0.613	80	**0.325**	2-40	0.518	0.606	0.471
0.613	90	**0.327**	2-45	0.508	0.602	0.456
0.613	100	**0.311**	2-50	0.504	0.606	0.446

表 5 誤差測定：攻撃前：RMSE

クラスタなし	**1回クラスタリング**		**2回クラスタリング(提案手法)**			
RMSE (X_0)	クラスタ数	RMSE (X_1)	クラスタ数	RMSE(X_2)		
				平均値	最大値	最小値
0.771	20	**0.609**	2-10	0.728	0.767	0.689
0.771	30	**0.588**	2-15	0.737	0.760	0.711
0.771	40	**0.561**	2-20	0.724	0.758	0.704
0.771	50	**0.536**	2-25	0.721	0.761	0.696
0.771	60	**0.528**	2-30	0.706	0.767	0.668
0.771	70	**0.524**	2-35	0.699	0.765	0.662
0.771	80	**0.492**	2-40	0.684	0.764	0.643
0.771	90	**0.505**	2-45	0.675	0.759	0.635
0.771	100	**0.493**	2-50	0.673	0.765	0.625

考えられる．

5.4 攻撃に対するロバスト性評価

続いて攻撃後の予測誤差を計測し，攻撃前の予測誤差との差分を取り比較を行った．これは表 3 の Δ_0 から Δ_2 の部分に当たる．この実験において，クラスタリングを 2 回行う手法での誤差差分の最小値がクラスタリングなし，1 回クラスタリング両方の誤差差分より小さくなることがある．その場合には最小値の項に太字と破線で表記している．表 12，13 はそれぞれランダム攻撃における指標 MAE，RMSE の差分である．ランダム攻撃では 1 回クラスタリングでのクラスタ数が小さい場合 (クラスタ数=20) はクラスタなしの誤差差分が小さくなるが，クラスタ数が大きくなると (クラスタ数=50,100) クラスタリングを用いる手法の誤差差分が小さくなる傾向が見られた．2 回クラスタリング手法の誤差の差分の平均値は，多くの場合で 1 回クラスタリング手法より大きくなるが，RMSE 指標での測定結果には最も小さい誤差差分となるものも存在する．また，平均値で誤差の差分が小さくならずとも，最小値では 1 番誤差の差分が小さくなるものが多数存在する．これは適切なクラスタ数を設定することによって，提案手法を用いてロバスト性を向上できることを示している．最大値については，例えばアタックサイズ 5%，フィラーサイズ 20%，1 回クラスタリング時のクラスタ数 100 のときの誤差差分を比較すると，指標 MAE ではクラスタなしのときで 0.168，提案手法で 0.153 と小さくなっている．また，指標 RMSE でもクラスタなしの誤差差分 0.138 に対して，提案手法の誤差差分は 0.124 と小さくなっている．2 回クラスタリング手法の誤差差分の最大値が，クラスタリングなしの誤差差分より小さくなる場合，クラスタ数を吟味しなくとも，攻撃の影響を軽減することができる可能性を示している．

表 14，15 はそれぞれ平均攻撃における指標 MAE，RMSE の差分である．平均攻撃は他 2 つの攻撃と比べ誤差の差分が小さくなっている．また，1 回クラスタリングでのクラスタ数を増やしたときに，誤差の差分が大きくなる傾向が見られた．2 回クラスタリングの誤差差分の平均値は，どちらの指標においても少なくとも他 2 つのどちらかの手法の差分より小さく (下線部)，両方の手法より小さいもの (太字) も複数存在する．また，最小値について見るとほぼすべての場合で他の誤差差分より小さくなっている．このことから，2 回クラスタリングを行う手法は平均攻撃に対して有効であるといえる．

表 16，17 はそれぞれバンドワゴン攻撃における指標 MAE，RMSE の差分である．バンドワゴン攻撃では，ほぼすべての場合でクラスタリングを用いる手法の誤差差分がクラスタなしの場合の誤差差分より小さくなる．クラスタリングを用いる手法同士で比較すると，多くの場合で 1 回クラスタリングの誤差差分のほうが小さいが，攻撃のアタックサイズや

表 6 誤差測定：ランダム攻撃：MAE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		MAE (Y_0)	クラスタ数	MAE (Y_1)	クラスタ数	MAE(Y_2)		
						平均値	最大値	最小値
5	5	0.651	20	**0.497**	2-10	0.619	0.642	0.588
5	5	0.651	50	**0.400**	2-25	0.584	0.641	0.542
5	5	0.651	100	**0.331**	2-50	0.532	0.645	0.470
5	20	0.780	20	**0.631**	2-10	0.751	0.775	0.719
5	20	0.780	50	**0.518**	2-25	0.716	0.772	0.665
5	20	0.780	100	**0.383**	2-50	0.631	0.748	0.555
20	5	0.776	20	**0.621**	2-10	0.713	0.769	0.713
20	5	0.776	50	**0.503**	2-25	0.710	0.771	0.671
20	5	0.776	100	**0.442**	2-50	0.656	0.770	0.590
20	20	1.328	20	**1.189**	2-10	1.276	1.333	1.276
20	20	1.328	50	**1.074**	2-25	1.282	1.333	1.235
20	20	1.328	100	**0.879**	2-50	1.200	1.316	1.117

表 7 誤差測定：ランダム攻撃：RMSE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		RMSE (Y_0)	クラスタ数	RMSE (Y_1)	クラスタ数	RMSE(Y_2)		
						平均値	最大値	最小値
5	5	0.801	20	**0.650**	2-10	0.772	0.792	0.742
5	5	0.801	50	**0.563**	2-25	0.739	0.790	0.703
5	5	0.801	100	**0.511**	2-50	0.694	0.795	0.643
5	20	0.910	20	**0.779**	2-10	0.885	0.905	0.854
5	20	0.910	50	**0.680**	2-25	0.855	0.903	0.808
5	20	0.910	100	**0.553**	2-50	0.771	0.870	0.708
20	5	0.903	20	**0.766**	2-10	0.878	0.898	0.852
20	5	0.903	50	**0.662**	2-25	0.851	0.898	0.821
20	5	0.903	100	**0.640**	2-50	0.810	0.898	0.763
20	20	1.289	20	**1.222**	2-10	1.285	1.303	1.270
20	20	1.289	50	**1.149**	2-25	1.274	1.303	1.246
20	20	1.289	100	**1.008**	2-50	1.209	1.278	1.161

表 8　誤差測定：平均攻撃：MAE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		MAE (Y_0)	クラスタ数	MAE (Y_1)	クラスタ数	MAE(Y_2)		
						平均値	最大値	最小値
5	5	0.630	20	**0.459**	2-10	0.583	0.620	0.539
5	5	0.630	50	**0.405**	2-25	0.562	0.623	0.529
5	5	0.630	100	**0.321**	2-50	0.517	0.619	0.464
5	20	0.649	20	**0.468**	2-10	0.602	0.638	0.557
5	20	0.649	50	**0.423**	2-25	0.582	0.613	0.558
5	20	0.649	100	**0.369**	2-50	0.555	0.653	0.500
20	5	0.657	20	**0.468**	2-10	0.584	0.645	0.542
20	5	0.657	50	**0.393**	2-25	0.560	0.643	0.516
20	5	0.657	100	**0.304**	2-50	0.515	0.651	0.447
20	20	0.661	20	**0.489**	2-10	0.602	0.613	0.575
20	20	0.661	50	**0.453**	2-25	0.598	0.617	0.582
20	20	0.661	100	**0.436**	2-50	0.584	0.669	0.561

表 9　誤差測定：平均攻撃：RMSE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		RMSE (Y_0)	クラスタ数	RMSE (Y_1)	クラスタ数	RMSE(Y_2)		
						平均値	最大値	最小値
5	5	0.787	20	**0.612**	2-10	0.741	0.777	0.698
5	5	0.787	50	**0.573**	2-25	0.724	0.779	0.697
5	5	0.787	100	**0.504**	2-50	0.684	0.773	0.644
5	20	0.807	20	**0.624**	2-10	0.761	0.794	0.718
5	20	0.807	50	**0.583**	2-25	0.744	0.771	0.720
5	20	0.807	100	**0.539**	2-50	0.715	0.803	0.665
20	5	0.815	20	**0.621**	2-10	0.742	0.801	0.701
20	5	0.815	50	**0.555**	2-25	0.724	0.798	0.687
20	5	0.815	100	**0.479**	2-50	0.682	0.804	0.627
20	20	0.845	20	**0.644**	2-10	0.762	0.771	0.733
20	20	0.845	50	**0.606**	2-25	0.759	0.782	0.742
20	20	0.845	100	**0.600**	2-50	0.744	0.835	0.721

表 10　誤差測定：バンドワゴン攻撃：MAE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		MAE (Y_0)	クラスタ数	MAE (Y_1)	クラスタ数	MAE(Y_2)		
						平均値	最大値	最小値
5	5	0.651	20	**0.481**	2-10	0.600	0.642	0.563
5	5	0.651	50	**0.404**	2-25	0.582	0.642	0.547
5	5	0.651	100	**0.321**	2-50	0.532	0.645	0.466
5	20	0.774	20	**0.600**	2-10	0.727	0.754	0.686
5	20	0.774	50	**0.495**	2-25	0.697	0.754	0.653
5	20	0.774	100	**0.374**	2-50	0.639	0.761	0.560
20	5	0.763	20	**0.585**	2-10	0.718	0.754	0.680
20	5	0.763	50	**0.466**	2-25	0.692	0.754	0.647
20	5	0.763	100	**0.398**	2-50	0.635	0.759	0.558
20	20	1.263	20	**1.055**	2-10	1.225	1.249	1.186
20	20	1.263	50	**0.816**	2-25	1.177	1.249	1.110
20	20	1.263	100	**0.631**	2-50	1.098	1.262	1.000

表 11　誤差測定：バンドワゴン攻撃：RMSE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		RMSE (Y_0)	クラスタ数	RMSE (Y_1)	クラスタ数	RMSE(Y_2)		
						平均値	最大値	最小値
5	5	0.801	20	**0.633**	2-10	0.752	0.793	0.717
5	5	0.801	50	**0.567**	2-25	0.739	0.792	0.709
5	5	0.801	100	**0.501**	2-50	0.695	0.794	0.639
5	20	0.909	20	**0.749**	2-10	0.866	0.890	0.827
5	20	0.909	50	**0.652**	2-25	0.841	0.890	0.802
5	20	0.909	100	**0.541**	2-50	0.787	0.889	0.719
20	5	0.893	20	**0.730**	2-10	0.853	0.884	0.820
20	5	0.893	50	**0.622**	2-25	0.834	0.887	0.800
20	5	0.893	100	**0.579**	2-50	0.787	0.886	0.728
20	20	1.255	20	**1.105**	2-10	1.242	1.257	1.217
20	20	1.255	50	**0.900**	2-25	1.200	1.257	1.156
20	20	1.255	100	**0.745**	2-50	1.130	1.253	1.057

Vol.2018 No.9,B1-3
2019/2/21

表 12 ロバスト性分析：ランダム攻撃：MAE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		差分 (Δ_0)	クラスタ数	差分 (Δ_1)	クラスタ数	差分(Δ_2)		
						平均値	最大値	最小値
5	5	**0.039**	20	0.042	2-10	0.050	0.066	**0.033**
5	5	0.039	50	**0.023**	2-25	0.026	0.047	**0.001**
5	5	0.039	100	**0.020**	2-50	0.028	0.047	**-0.001**
5	20	**0.168**	20	0.176	2-10	0.182	0.199	**0.166**
5	20	0.168	50	**0.142**	2-25	0.158	0.181	**0.120**
5	20	0.168	100	**0.072**	2-50	0.127	0.153	0.097
20	5	**0.163**	20	0.167	2-10	0.176	0.188	**0.161**
20	5	0.163	50	**0.126**	2-25	0.152	0.170	0.129
20	5	0.163	100	**0.130**	2-50	0.152	0.183	**0.122**
20	20	**0.715**	20	0.735	2-10	0.739	0.751	0.724
20	20	0.715	50	**0.698**	2-25	0.724	0.749	**0.698**
20	20	0.715	100	**0.568**	2-50	0.696	0.739	0.669

表 13 ロバスト性分析：ランダム攻撃：RMSE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		差分 (Δ_0)	クラスタ数	差分 (Δ_1)	クラスタ数	差分(Δ_2)		
						平均値	最大値	最小値
5	5	**0.030**	20	0.041	2-10	0.045	0.065	**0.023**
5	5	0.030	50	0.027	2-25	**0.018**	0.046	-0.011
5	5	0.030	100	**0.019**	2-50	0.021	0.044	**-0.012**
5	20	**0.138**	20	0.170	2-10	0.157	0.177	**0.138**
5	20	0.138	50	0.143	2-25	**0.134**	0.161	0.094
5	20	0.138	100	**0.060**	2-50	0.099	0.124	0.069
20	5	**0.131**	20	0.157	2-10	0.150	0.164	**0.129**
20	5	0.131	50	**0.126**	2-25	0.130	0.148	**0.108**
20	5	**0.131**	100	0.148	2-50	0.137	0.167	**0.110**
20	20	**0.518**	20	0.613	2-10	0.558	0.581	0.534
20	20	**0.518**	50	0.613	2-25	0.553	0.573	0.529
20	20	0.518	100	**0.516**	2-50	0.537	0.567	**0.503**

表 14 ロバスト性分析：平均攻撃：MAE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		差分 (Δ_0)	クラスタ数	差分 (Δ_1)	クラスタ数	差分(Δ_2)		
						平均値	最大値	最小値
5	5	0.017	20	**0.004**	2-10	0.014	0.019	0.008
5	5	0.017	50	0.028	2-25	**0.005**	0.028	-0.017
5	5	0.017	100	**0.010**	2-50	0.013	0.037	**-0.006**
5	20	0.036	20	**0.013**	2-10	0.033	0.049	0.021
5	20	0.036	50	0.046	2-25	**0.025**	0.047	0.004
5	20	**0.036**	100	0.057	2-50	0.051	0.099	**0.007**
20	5	0.045	20	**0.013**	2-10	0.015	0.039	**-0.001**
20	5	0.045	50	0.016	2-25	**0.003**	0.040	-0.028
20	5	0.045	100	**-0.007**	2-50	0.011	0.052	**-0.012**
20	20	0.049	20	0.035	2-10	**0.033**	0.058	0.003
20	20	0.049	50	0.077	2-25	**0.041**	0.079	0.009
20	20	**0.049**	100	0.125	2-50	0.079	0.124	**0.021**

表 15 ロバスト性分析：平均攻撃：RMSE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		差分 (Δ_0)	クラスタ数	差分 (Δ_1)	クラスタ数	差分(Δ_2)		
						平均値	最大値	最小値
5	5	0.016	20	**0.004**	2-10	0.013	0.021	**0.003**
5	5	0.016	50	0.037	2-25	**0.003**	0.028	-0.017
5	5	0.016	100	0.012	2-50	**0.011**	0.040	-0.007
5	20	0.035	20	**0.015**	2-10	0.033	0.048	0.021
5	20	0.035	50	0.047	2-25	**0.023**	0.045	0.003
5	20	**0.035**	100	0.046	2-50	0.042	0.084	**0.002**
20	5	0.044	20	**0.012**	2-10	0.014	0.036	**-0.004**
20	5	0.044	50	0.020	2-25	**0.003**	0.037	-0.028
20	5	0.044	100	**-0.013**	2-50	0.009	0.050	**-0.015**
20	20	0.074	20	0.035	2-10	**0.034**	0.063	0.001
20	20	0.074	50	0.070	2-25	**0.038**	0.083	0.009
20	20	0.074	100	0.107	2-50	**0.071**	0.105	0.019

表 16　ロバスト性分析：バンドワゴン攻撃：MAE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		差分 (Δ_0)	クラスタ数	差分 (Δ_1)	クラスタ数	差分(Δ_2)		
						平均値	最大値	最小値
5	5	0.038	20	**0.027**	2-10	0.031	0.037	**0.022**
5	5	0.038	50	0.028	2-25	**0.024**	0.048	-0.010
5	5	0.038	100	**0.010**	2-50	0.028	0.052	**-0.001**
5	20	0.162	20	0.145	2-10	**0.084**	0.100	0.069
5	20	0.162	50	**0.118**	2-25	0.139	0.156	**0.116**
5	20	0.162	100	**0.063**	2-50	0.135	0.160	0.106
20	5	0.150	20	**0.131**	2-10	0.148	0.157	0.134
20	5	0.150	50	**0.089**	2-25	0.134	0.156	0.102
20	5	0.150	100	**0.087**	2-50	0.131	0.160	0.112
20	20	0.650	20	**0.601**	2-10	0.656	0.665	0.640
20	20	0.650	50	**0.439**	2-25	0.619	0.648	0.565
20	20	0.650	100	**0.320**	2-50	0.594	0.656	0.548

表 17　ロバスト性分析：バンドワゴン攻撃：RMSE

アタックサイズ (%)	フィラーサイズ (%)	クラスタなし	1回クラスタリング		2回クラスタリング(提案手法)			
		差分 (Δ_0)	クラスタ数	差分 (Δ_1)	クラスタ数	差分(Δ_2)		
						平均値	最大値	最小値
5	5	0.029	20	**0.024**	2-10	0.024	0.031	**0.009**
5	5	0.029	50	0.031	2-25	**0.018**	0.046	-0.018
5	5	0.029	100	**0.009**	2-50	0.022	0.051	**-0.008**
5	20	**0.137**	20	0.141	2-10	0.138	0.154	**0.117**
5	20	0.137	50	**0.117**	2-25	0.120	0.138	**0.096**
5	20	0.137	100	**0.048**	2-50	0.114	0.142	0.083
20	5	0.122	20	**0.121**	2-10	0.126	0.140	**0.104**
20	5	0.122	50	**0.086**	2-25	0.113	0.138	**0.084**
20	5	0.122	100	**0.086**	2-50	0.114	0.147	0.091
20	20	**0.484**	20	0.497	2-10	0.514	0.531	0.487
20	20	0.484	50	**0.364**	2-25	0.479	0.508	0.440
20	20	0.484	100	**0.253**	2-50	0.458	0.494	0.424

フィラーサイズが小さいとき，2 回クラスタリングの誤差差分の最小値が 1 回クラスタリングの誤差差分より小さくなるものが多数存在する．また，2 回クラスタリングの誤差差分の最大値の中には，クラスタなしの誤差差分を下回るものが存在する．

6. 関 連 研 究

協調フィルタリングに対する攻撃の概念は O'Mahony[1] によって提言され，以降研究者たちは攻撃の検出やロバスト性の強化について研究している．日本においては神嶌[4] が論文中で「サクラ攻撃」として紹介している．

協調フィルタリングへの攻撃に対するアプローチとしては，大きく攻撃ユーザを検出する方法と攻撃に対してロバストなシステムの構築に分けられる．攻撃ユーザを検出する方法としては統計量を用いる方法[5]，攻撃検出用の指標を用いる方法[6]，クラスタリングを用いる手法などがある．クラスタリングを用いる手法では定期的にクラスタリングを行い，クラスタの代表点に大きな変化があった場合に攻撃が追加されたとする方法[7] や，クラスタリングを行った結果，サイズが小さくなったクラスタを攻撃用ユーザクラスタとみなす方法[8] がある．一方で攻撃に対してロバストなシステムの構築については攻撃検出用の指標を組み合わせて用いる方法[9] や SVD ベース協調フィルタリング[10] 等が存在する．

本研究の類似研究としては Mobasher[11] によるものがある．Mobasher は攻撃に対して PLSA ベース協調フィルタリングが有効であることを示す際，本研究で用いている K 平均法を用いた比較実験を行っている．Mobasher の研究では本研究で取り上げたランダム攻撃に対して実験を行っておらず，またクラスタリングを 2 回行うことで既存のクラスタを結合するといった手法は用いていない．

7. 結　　論

本研究ではユーザデータにクラスタリングを 2 度施し，分類されたクラスタ内で予測を行う手法を提案した．また，手法の評価として，実測値と予測値の誤差及び攻撃前後での誤差の差分を用いることを提案し，実際に計測を行い，攻撃に対する影響の大きさを調査した．それによりクラスタリングを 2 度行う予測手法は，特に平均攻撃に対して攻撃の影響を抑えることに有用であるという結果を得ることができた．

今後の課題として，予測手法については他のクラスタリング手法や類似度の計算法の使用，評価手法については本研究で取り上げていない攻撃の追加や，評価に用いた指標の吟味が挙げられる．

Vol.2018 No.9,B1-3
2019/2/21

参 考 文 献

1) Michael P. O'Mahony, Neil J. Hurley, Guenole C. M. Silvestre, Promoting recommendations: an attack on collaborative filtering :DEXA 2002, pp. 494-503, 2002.
2) J. MacQueen, Some methods for classification and analysis of multivariate observations, BSMSP, Vol.1,pp. 281-297, 1967.
3) Bamshad Mobasher, Robin D. Burke, Runa Bhaumik, Chad Williams, Towards trustworthy recommender systems: an analysis of attack models and algorithm robustness :ACM TOIT,Vol.7, No.4, Article No.23, 2007.
4) 神嶌　敏弘，推薦システムのアルゴリズム (3) :人工知能学会誌，2008 年 3 月号.
5) Runa Bhaumik, Chad Williams, Bamshad Mobasher, Robin D. Burke, Securing collaborative filtering against malicious attacks through anomaly detection :AAAI ITWP 2006.
6) Paul-Alexandru Chirita, Wolfgang Nejdl, Cristian Zamfir, Preventing shilling attacks in online recommender systems :WISM 2005, pp. 67-74, 2005.
7) Michael P. O'Mahony, Neil J. Hurley, Guenole C. M. Silvestre, Collaborative filtering-safe and sound? :ISMIS 2003, pp. 506-510, 2003.
8) Runa Bhaumik, Bamshad Mobasher, Robin D. Burke, A clustering approach to unsupervised attack detection in collaborative recommender systems : ICDM 2011, pp. 181-187, 2011.
9) Chad Williams, Runa Bhaumik, Bamshad Mobasher, Robin D. Burke, Jeff J. Sandvig:, Detection of obfuscated attacks in collaborative recommender systems :ECAI 2006, pp. 19-23, 2006.
10) Fuguo Zhang, Shenghua Xu, Analysis of trust-based e-commerce recommender systems under recommendation attacks: ISDPE 2007, pp. 385-390, 2007.
11) Bamshad Mobasher, Robin D. Burke, Jeff J. Sandvig, Model-Based collaborative filtering as a defense against profile injection attacks :AAAI 2006, pp 1388-1393, 2006.