

次元圧縮法を導入したクエリに基づくバイクラスタリング 情報推薦への応用

著者	武内 充, 三浦 功輝, 岡田 吉史
雑誌名	計測自動制御学会システム・情報部門学術講演会講演論文集
巻	2015
発行年	2015-11-18
URL	http://hdl.handle.net/10258/3860

次元圧縮法を導入したクエリに基づくバイクラスタリング 情報推薦への応用

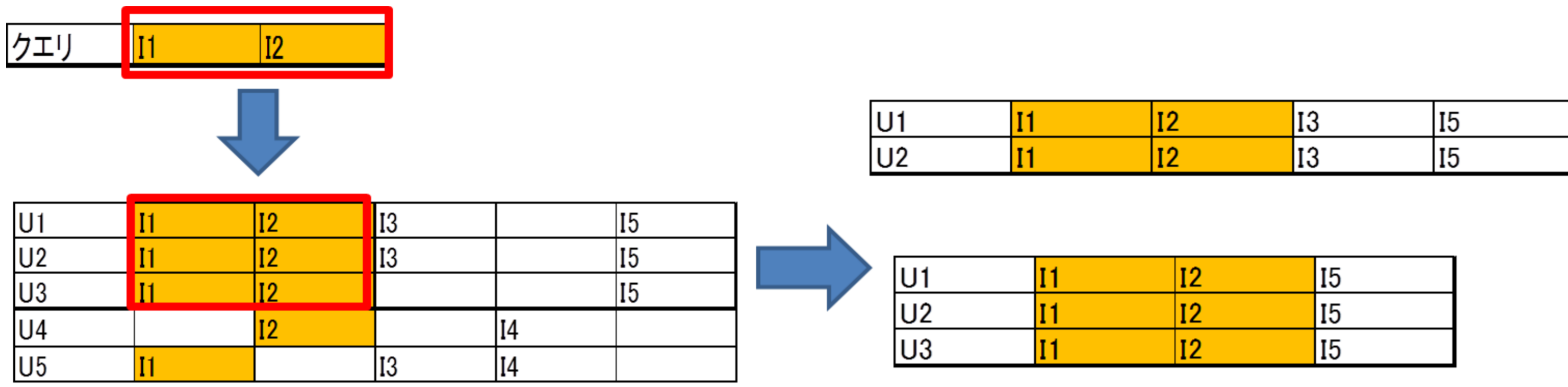
著者	武内 充, 三浦 功輝, 岡田 吉史
雑誌名	計測自動制御学会システム・情報部門学術講演会講演論文集
巻	2015
発行年	2015-11-18
URL	http://hdl.handle.net/10258/3860

次元圧縮法を導入したクエリに基づくバイクラスタリング -情報推薦への応用-

○武内充、三浦功輝、岡田吉史(室蘭工業大学)

研究背景

- 情報推薦の手法の一つとして協調フィルタリング法が挙げられる
 - バイクラスタリング法(NBCF)
 - アイテムベース法
 - ユーザーベース法
- 以前、我々の研究室ではクエリに基づくバイクラスタリングを用いた協調フィルタリング法(以下、前手法)を提案した



データベースの縮減

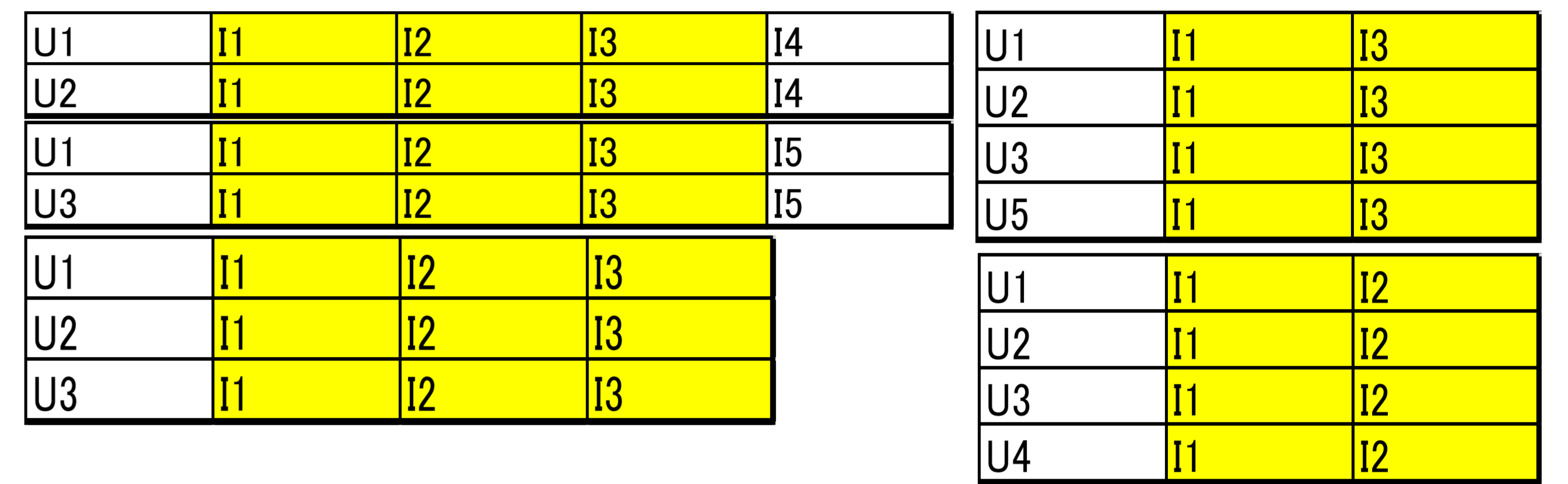
クエリを全て評価しているユーザのみでバイクラスタを生成するため、バイクラスタ抽出時間を大幅に短縮することが出来た

*横山直也, 岡田吉史, "クエリに基づくバイクラスタリングを用いた協調フィルタリング法", 日本感性工学会 生命ソフトウェアシンポジウム2014, 2014.

目的

- 前手法では類似したアイテムを多く含むバイクラスタが大量に抽出される

U1	I1	I2	I3	I4	I5
U2	I1	I2	I3	I4	
U3	I1	I2	I3		I5
U4	I1	I2			
U5	I1		I3		



- ユーザからの評価が類似したアイテムを融合することでバイクラスタの抽出数を削減

バイクラスタ抽出時間を削減する

方法

STEP1:縮減データベースの作成

1. ユーザ(アイテム)の融合
評価件数(被評価件数)が最も多い未融合のユーザ(アイテム)を親ユーザ(親アイテム), それ以外のユーザ(アイテム)を子ユーザ(子アイテム)とする
2. 融合条件を満たした子ユーザ(子アイテム)を親ユーザ(親アイテム)に融合する
3. 全てのユーザ(アイテム)が親ユーザ(親アイテム)になるまで繰り返す

図はトランザクションデータベースからアイテムを融合することで縮減データベースを作成している

本手法では、ユーザ、アイテムの順にデータベースを縮減する方法をUI, アイテム、ユーザの順に縮減する方法をIUとする

ユーザの融合条件

- 子ユーザが未融合
- 親ユーザと子ユーザが共に評価しているアイテムの評価値が全て一致
- 親ユーザの評価件数の $\alpha\%$ 以上のアイテムを子ユーザが評価

1 トランザクションデータベース

U1	I1	I2	I3	I4	I5	1
U2	1	1	1	1	1	1
U3	1	0	1	1	1	0
U4	1	0	0	1	1	0
U5	0	0	0	1	1	0

U1	I1	I2	I4	I5	1
U2	1	1	1	1	1
U3	1	0	1	1	0
U4	1	0	0	1	0
U5	0	0	0	1	0

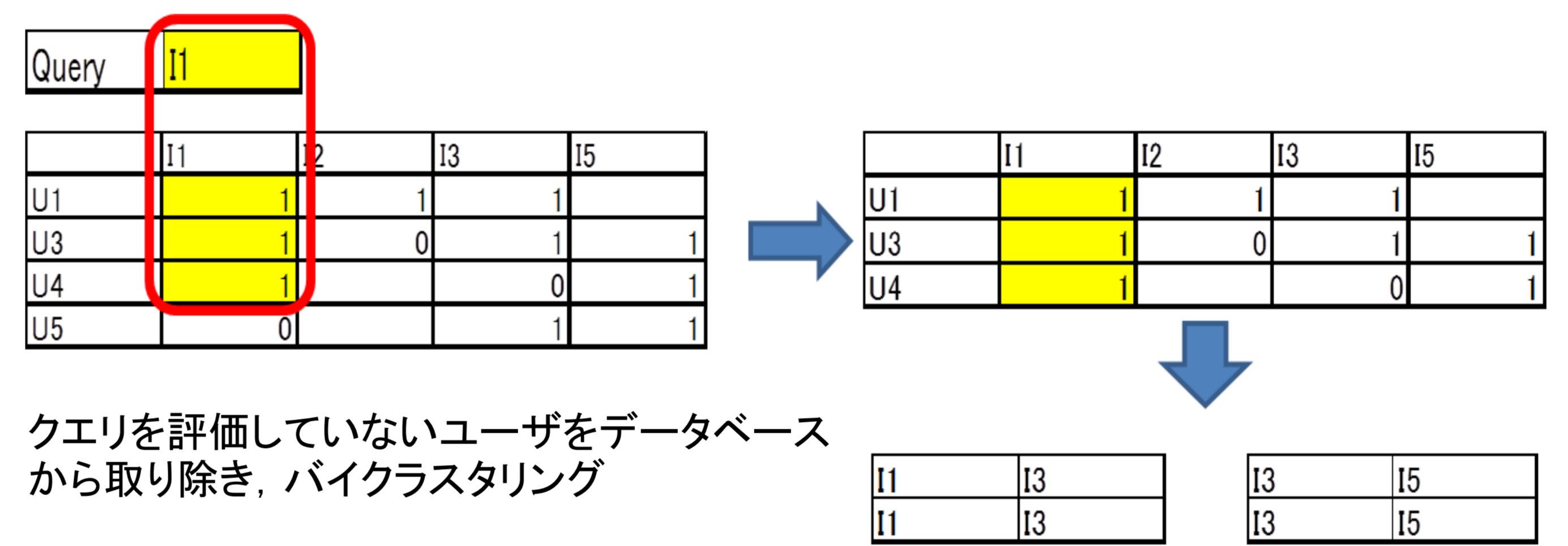
縮減データベース

U1	I1	I2	I4	0
U2	1	1	1	1
U3	1	0	0	1
U4	1	0	0	1
U5	0	0	0	1

アイテムの融合条件

- 子アイテムが未融合
- 親アイテムと子アイテムを共に評価しているユーザの評価値が一致
- 親アイテムを評価しているユーザの $\alpha\%$ 以上が子アイテムを評価

STEP2:クエリに基づくバイクラスタリング



クエリを評価していないユーザをデータベースから取り除き、バイクラスタリング

STEP3:推薦スコアの決定

抽出したバイクラスタを元に各アイテムの推薦スコアを決定
アイテム*i*の推薦スコアScore(*i*)は下式で定義される

$$Score(i) = \sum_{b_i} \frac{|I_q \cap I_{b_i}|}{|I_{b_i}|} \times |U_{b_i}|$$

I_q :クエリとして入力されたアイテムの集合
 I_{b_i} :バイクラスタ b_i に含まれるアイテムの集合
 U_{b_i} :バイクラスタ b_i に含まれるユーザの集合

実験

実験内容

- 推薦精度の検証(ランキング上位N位を使用)
- バイクラスタ抽出時間の比較

データセット

- 映画のベンチマークデータセットを使用
 - ユーザ数: 900
 - アイテム数: 1682
 - (1人の)最低評価件数: 10件
 - 評価値: 1~5

バイクラスタ設定パラメータ

- 最小アイテム数: 4
- 最小ユーザ数
 - 本手法(UI): 15
 - 本手法(IU): 6
 - 前手法: 2

融合条件

- $\alpha = 50$

評価指標

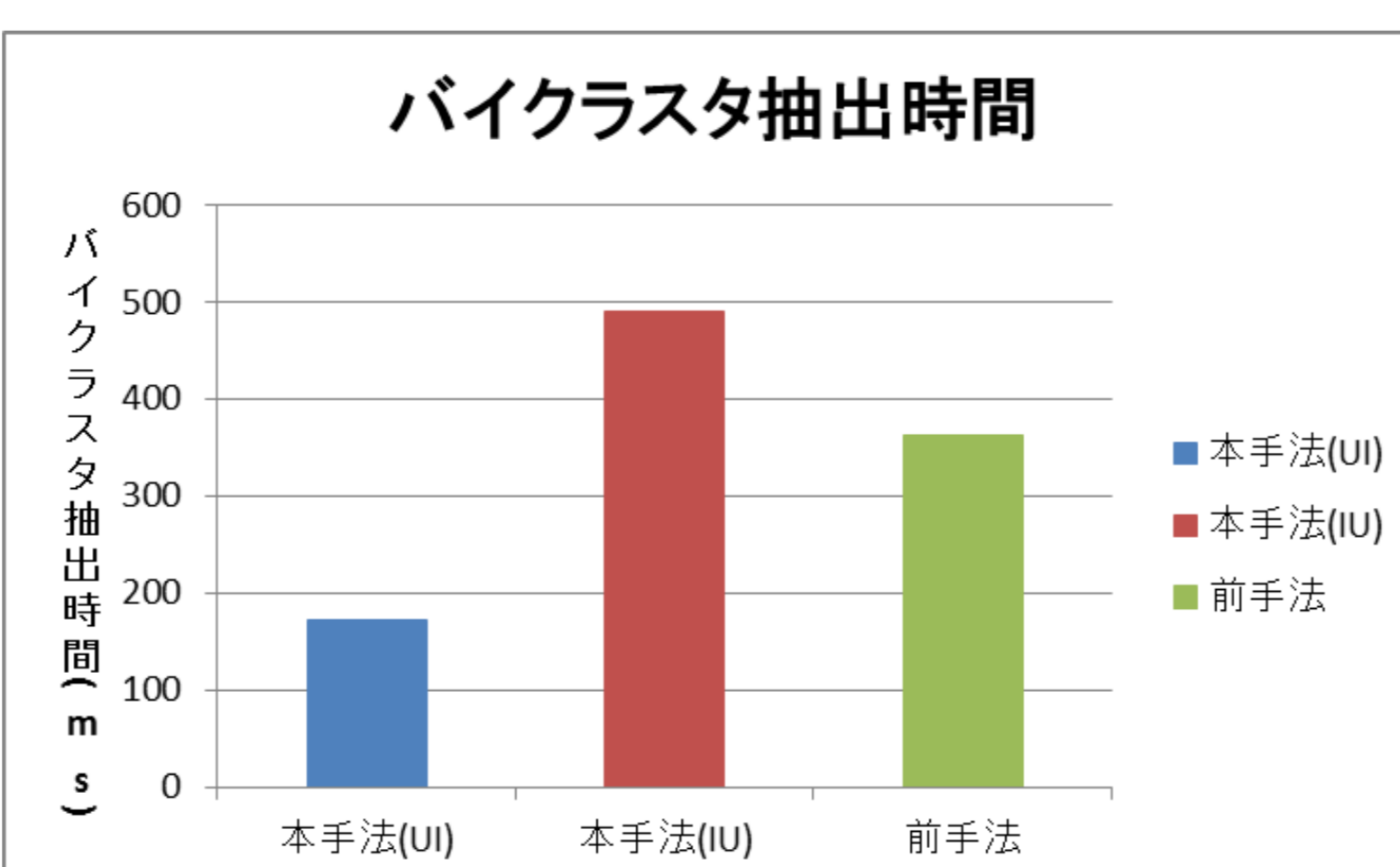
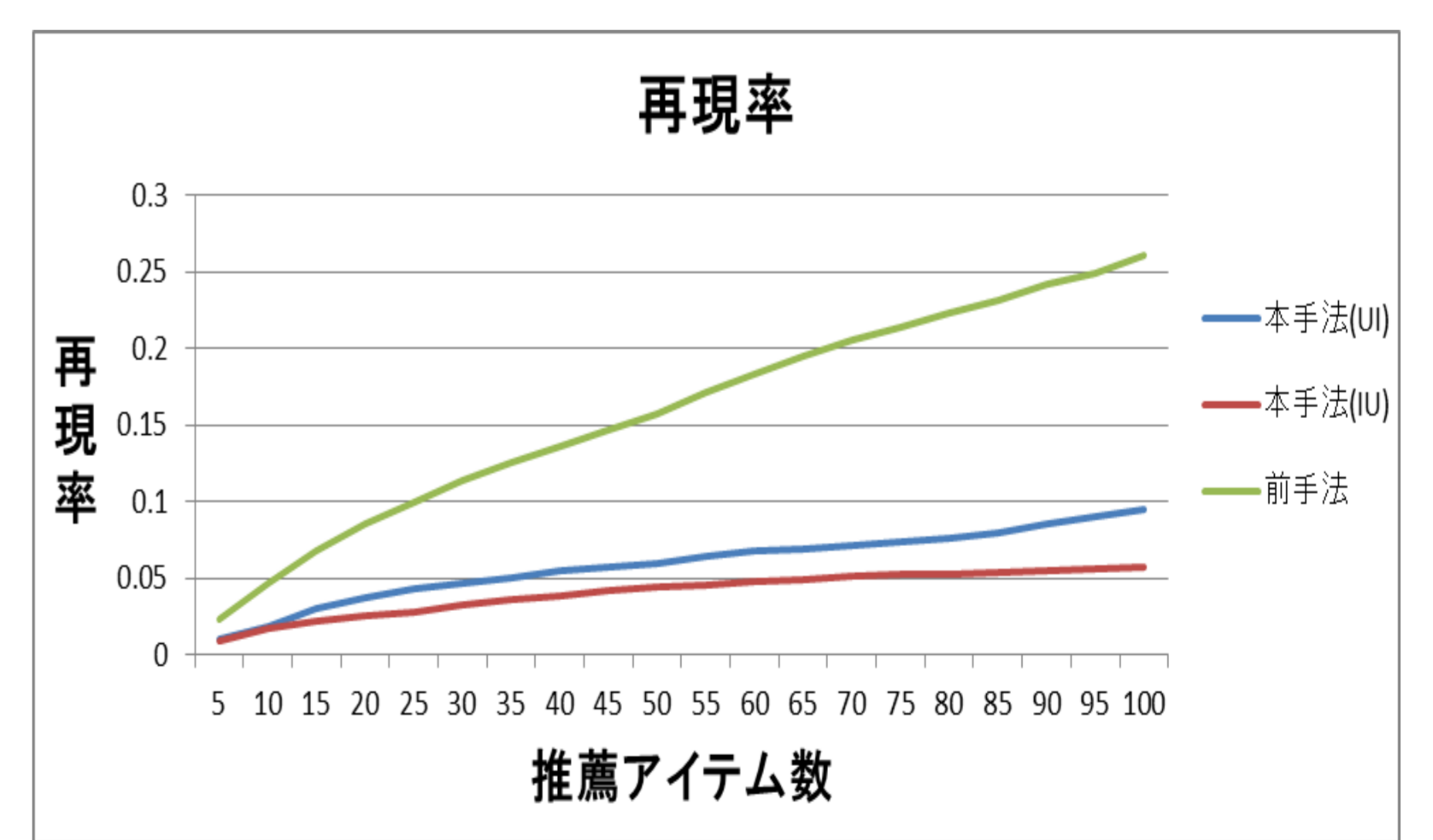
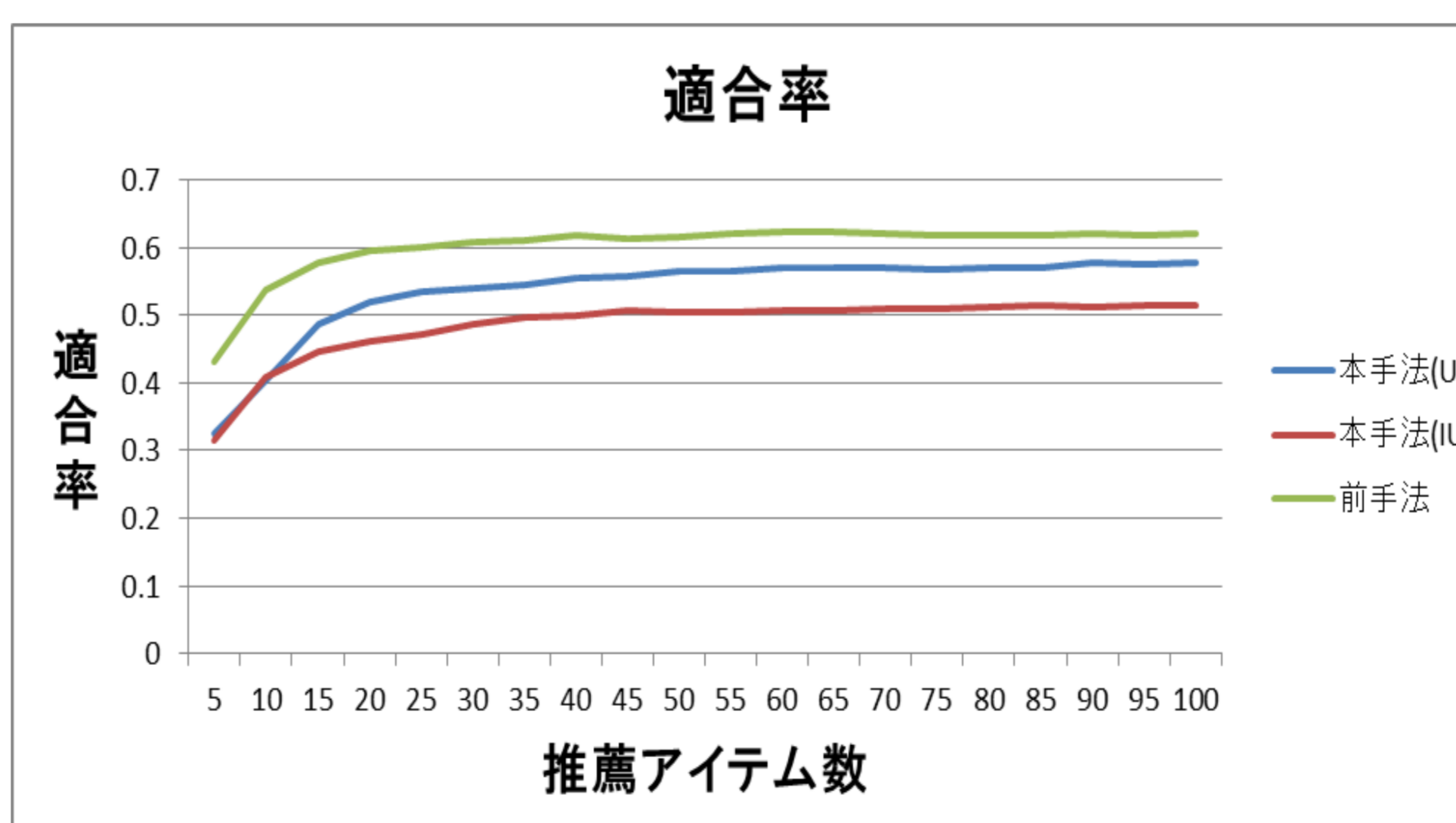
- 適合率(Precision), 再現率(Recall)

$$Precision = \frac{|A \cap B|}{|B|} \quad Recall = \frac{|A \cap B|}{|A|}$$

A: ユーザが高評価しているアイテムの集合

B: ユーザが評価している推薦されたアイテムの集合

結果・まとめ



本手法(UI)では、前手法と比べてバイクラスタ抽出時間を削減することができた
- データベースの縮減を正しく行うことができた

本手法は前手法に比べ、適合率、再現率が低下した
- 縮減によって、有用なデータも多く取り除いてしまったと考えられる
- 500個以上のアイテムを融合している親アイテムが存在する

今後の課題

- バイクラスタ設定パラメータの調整
- 融合条件の調整
- 別のデータセットへの適用