

# ドメイン別ユーザプロフィールの 構築と情報推薦への応用

---

尾崎研究室

鈴木陽介



# 目次

- はじめに、目的
- 提案手法
  - ✓概要
  - ✓ドメイン別ユーザプロファイルの構築
  - ✓情報推薦への応用
- 評価実験
- まとめ、今後の課題

# 協調フィルタリング

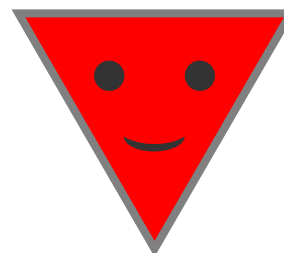


桜道家美味しかったー





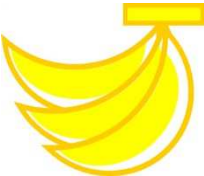


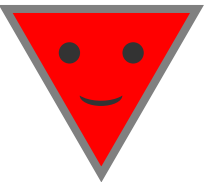
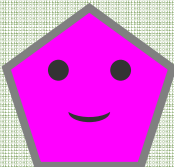
せい家を推薦

桜道家美味しいねー



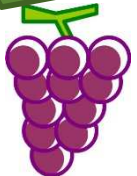

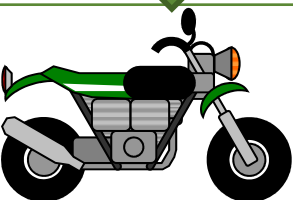


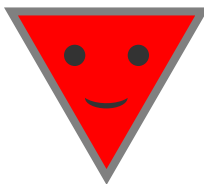
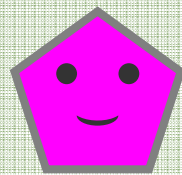
せい家おすすめー

# 協調フィルタリング

				
 類似	5	4	1	4
	4	5	2	?
	2	1	5	2

ユーザの嗜好プロフィールによって「口コミ」を再現

# ドメイン別のプロファイル

				
	5	4	1	4
	4	5	2	?
	2	1	5	2

# Twitter上の専門家ユーザー



**TSUTAYA**

♪待望の新作

「24 リブ・アナザー・デイ」(・V・)  
TSUTAYAで予約購入すると20%オフ♪

★詳しくはコチラ  
[inkto.jp/1DRHE78](http://inkto.jp/1DRHE78)

ドメインの専門家  
ドメインに特化した投稿多



**Seiichi Tani**

tani · 17時間

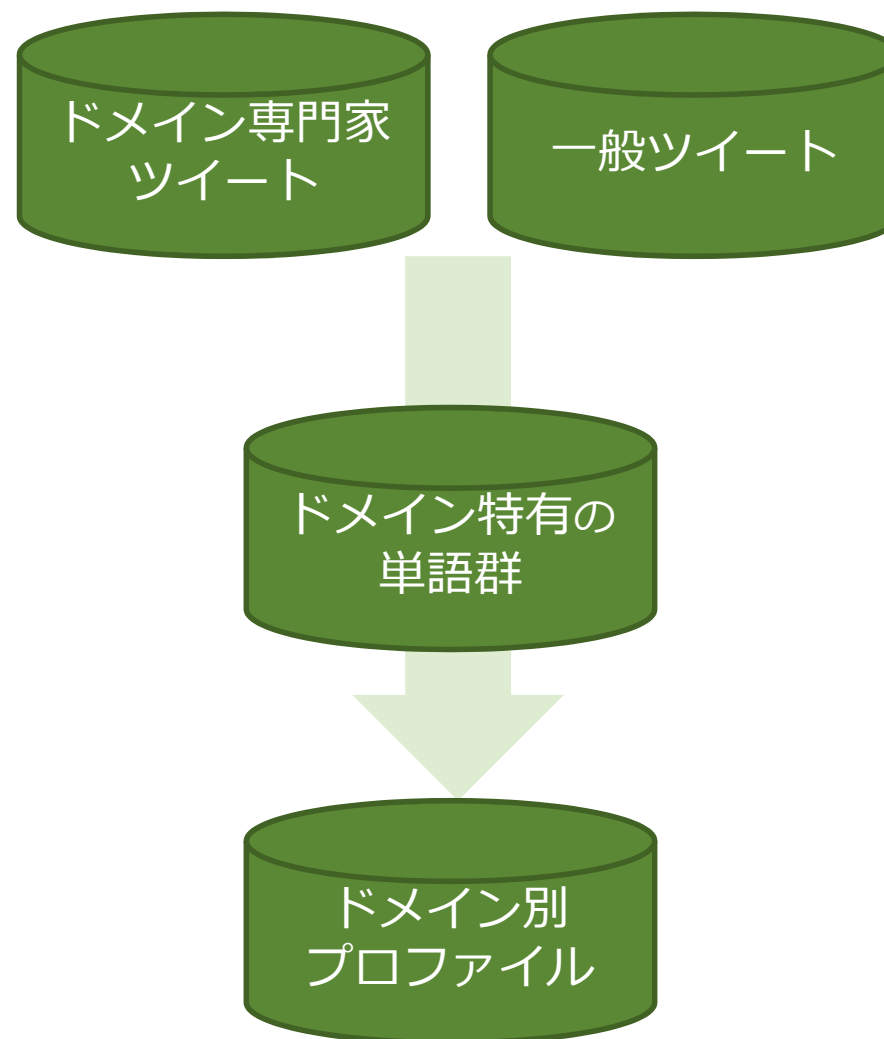
ちょっと遅れて、移動開始。

一般的なユーザー  
普段の出来事や感じたこと

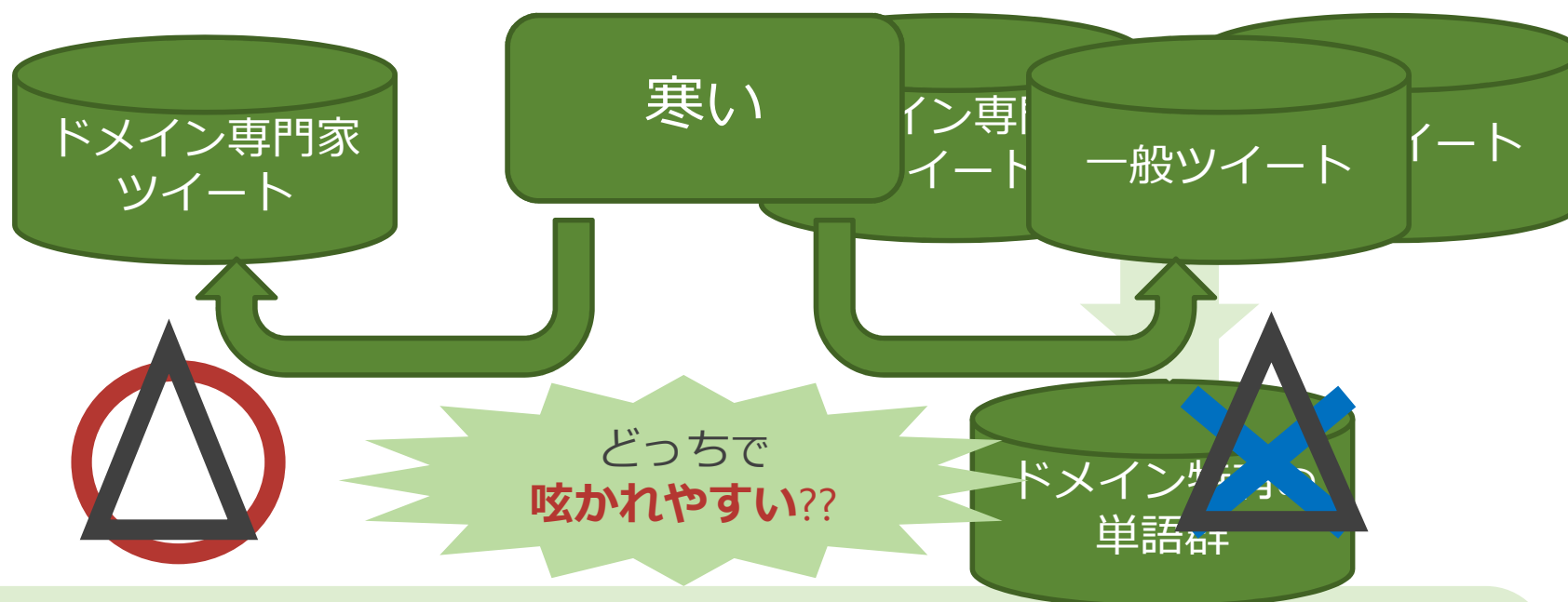
# 提案手法の概要

## ドメイン別 ユーザプロフィール

	野球	メジャー	...
$u_i$	4	9	...



# ドメイン特有単語の抽出：オッズ比



オッズ比 =  $\frac{P_{野球}^d}{1 - P_{野球}^d} \bigg/ \frac{P_{野球}^s}{1 - P_{野球}^s}$

ドメイン別 プロファイル

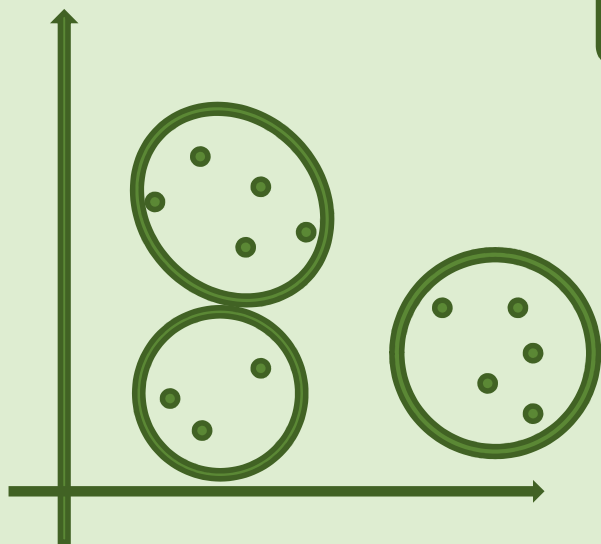


# ドメイン別辞書の次元縮約

	$w_1$	$w_2$	...	$w_n$
$u_i$	3	0	...	0

次元が大きすぎ!!  
プロフィールが疎に!!

クラスタリング



類似

ケータイ

スマホ

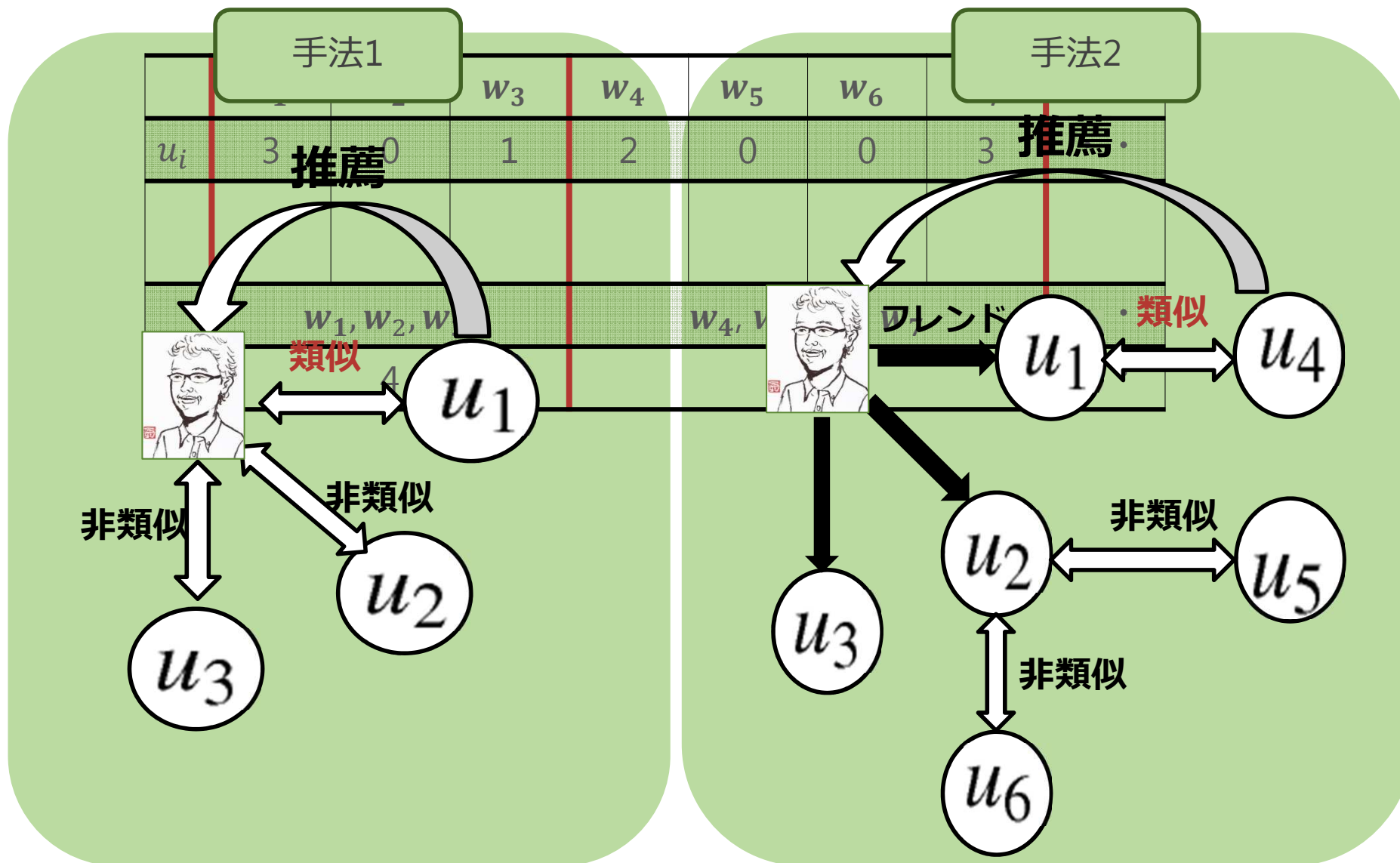
共起する単語

メール, line, 充電

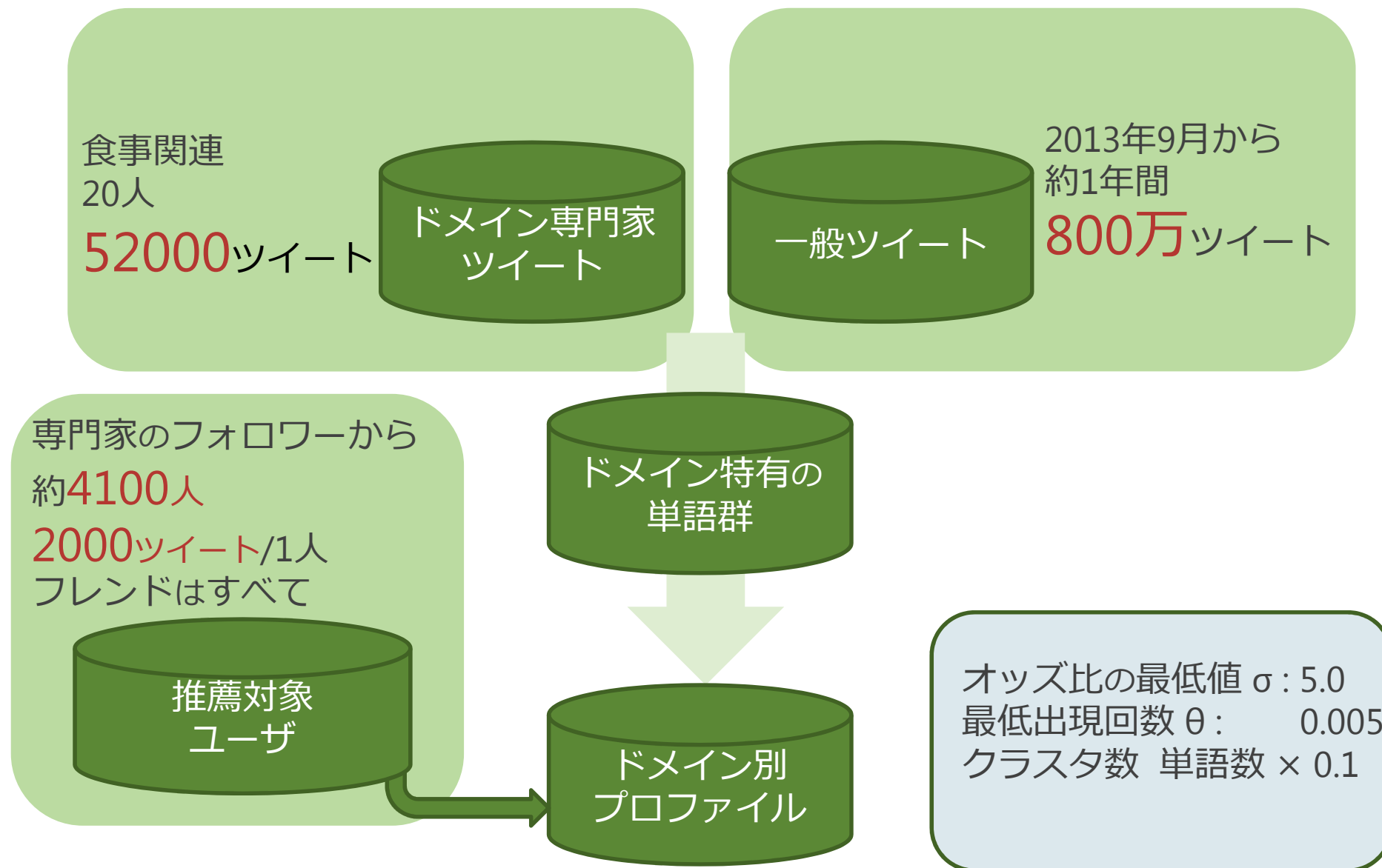
共起の回数

	$w_1$	$w_2$	$w_3$
$w_1$	0	3	1
$w_2$	3	0	7
$w_3$	1	7	0

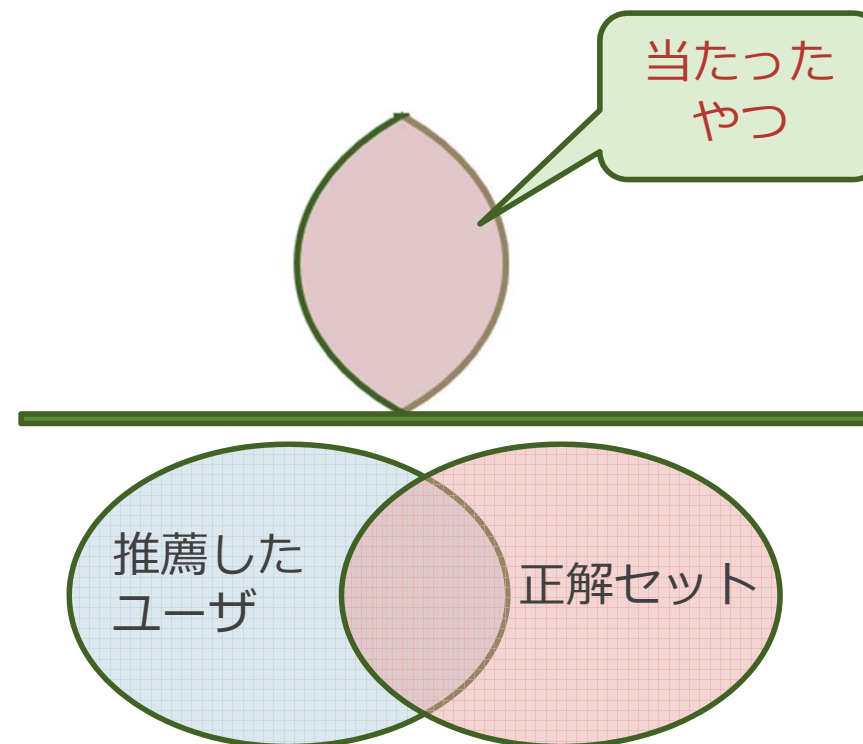
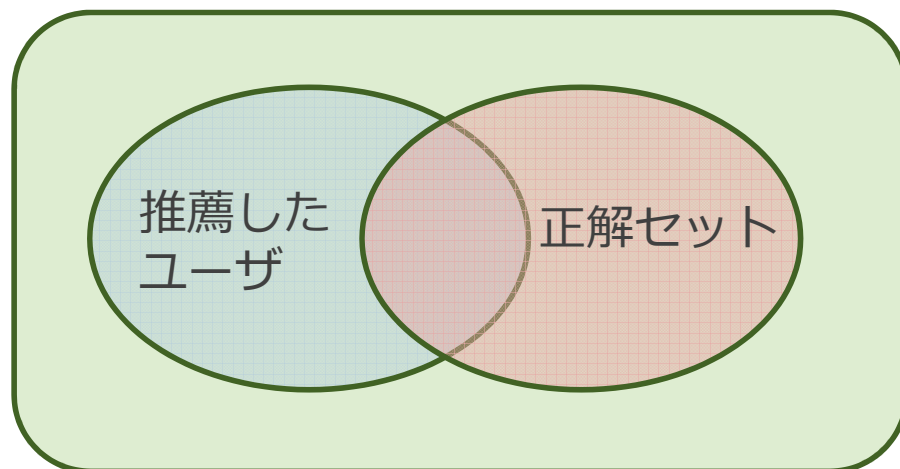
# ユーザ推薦への応用



# 評価実験：データ準備



# 適合率 (precision)



precision : 手法1, 2

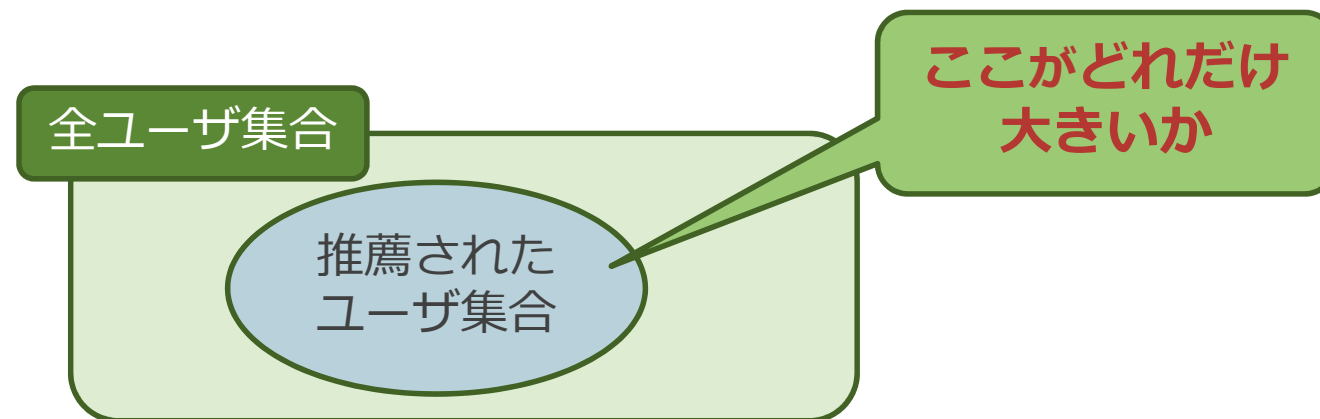
最大値	0.1
最低値	0.0
平均	0.0

**低い!!**

Twitterの特徴のため  
リアルな友達とつながりやすい

↓  
コミュニティの問題  
リンク予測

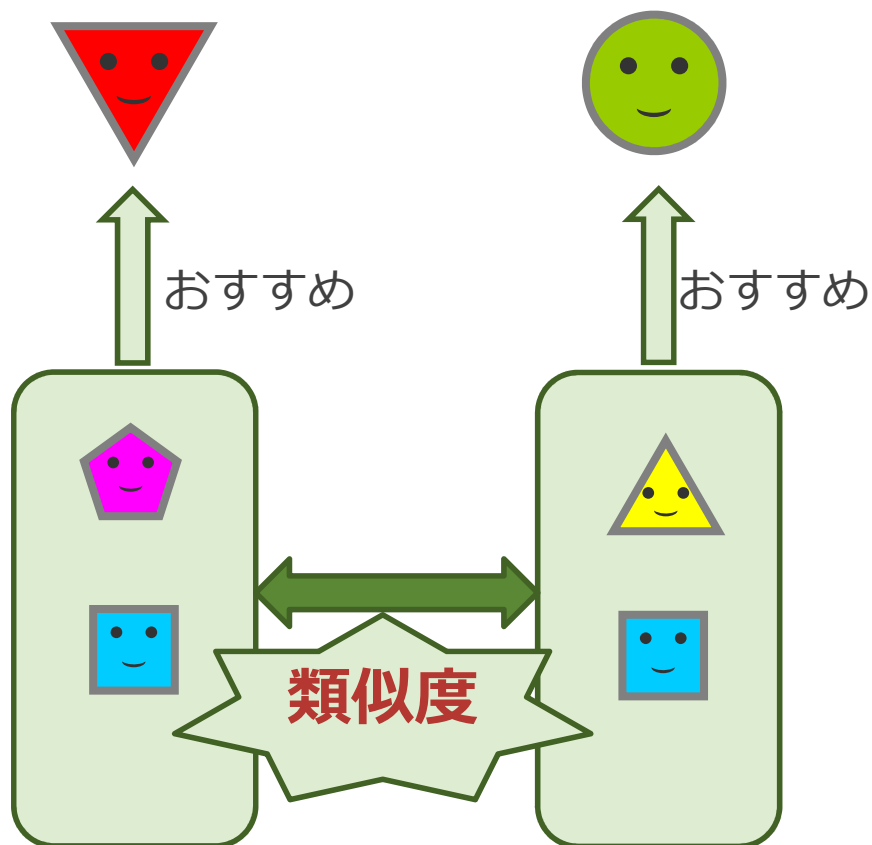
## 多様性：凝集多様性(Aggregate diversity)



		手法1	
		クラスタリング	
		有	無
オッズ比	有	0.01	0.06
	無	0.05	0.27

		手法1	
		クラスタリング	
		有	無
オッズ比	有	0.02	0.09
	無	0.09	0.58

# 多様性：ユーザ間相違度(Inter-user diversity)



手法1

		クラスタリング	
		有	無
オッズ比	有	0.01	0.06
	無	0.05	0.27

手法1

		クラスタリング	
		有	無
オッズ比	有	0.02	0.09
	無	0.09	0.58

# まとめ、今後の課題

## ●まとめ

- ✓高精度な推薦に向けた、ドメイン別のユーザプロフィール構築
- ✓Twitterにおけるユーザ推薦

## ●今後の課題

- ✓k-meansクラスタリングではなくLDAの利用
- ✓パラメタ値の自動設定
- ✓推薦リストの偏りの改善
- ✓ドメイン別ユーザプロフィールの複合的利用