

# SVMを用いた 感情抽出に基づく フォロワー推薦

---

甲南大学 知能情報学部 知能情報学科  
灘本研究室 11371002 秋山和寛

# 背景



# 背景

## ユーザをフォローする基準

膨大なユーザの中から  
フォローしたくなるようなユーザを自力で探し出すのは困難

友人

共通の趣味

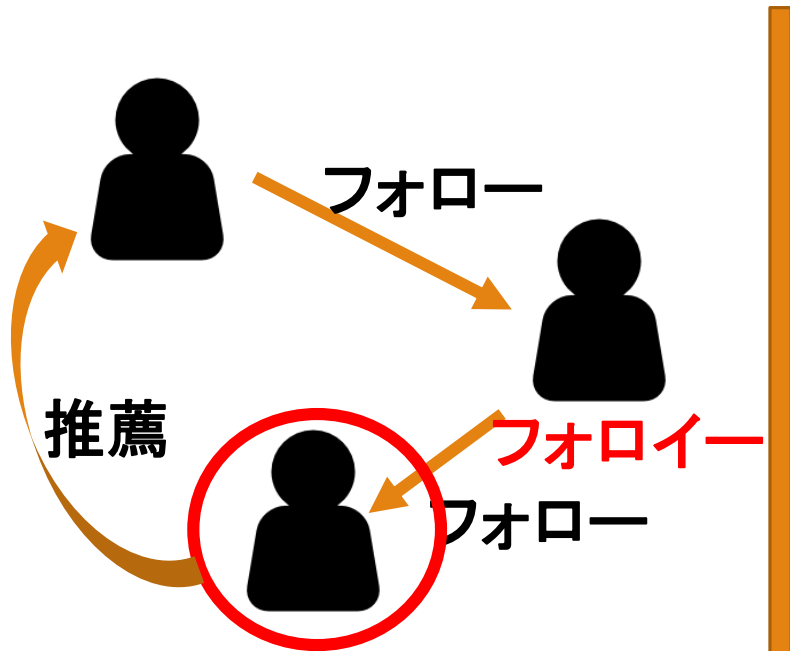
情報発信者

有名人

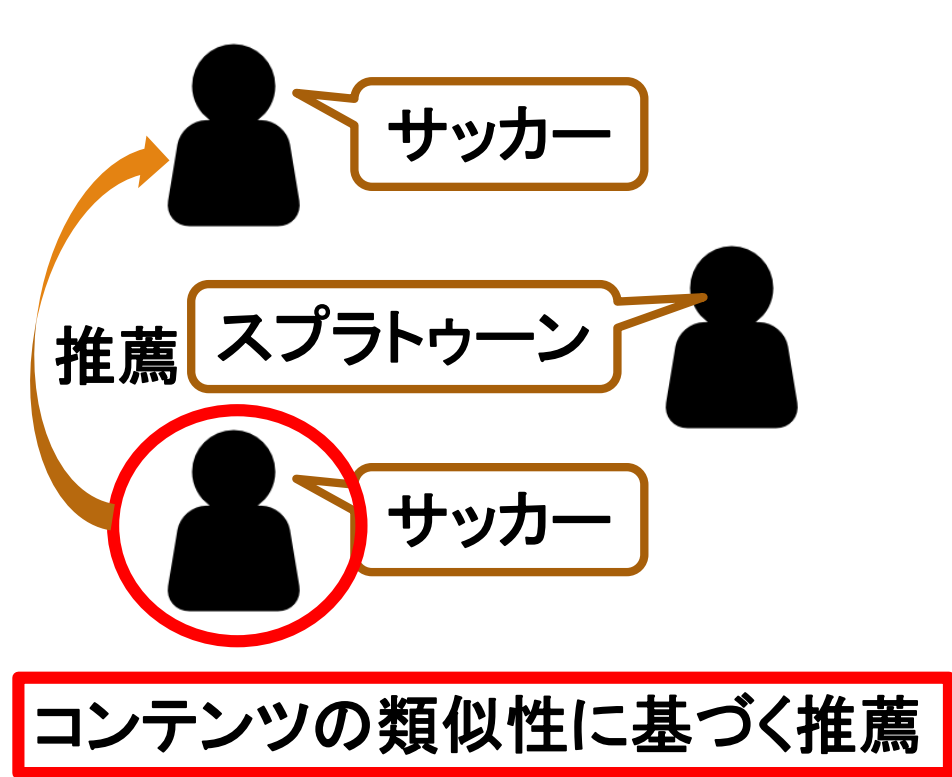
etc

# 背景

## フォロイー推薦



ネットワーク関係に基づく推薦



コンテンツの類似性に基づく推薦

# 背景

## 研究

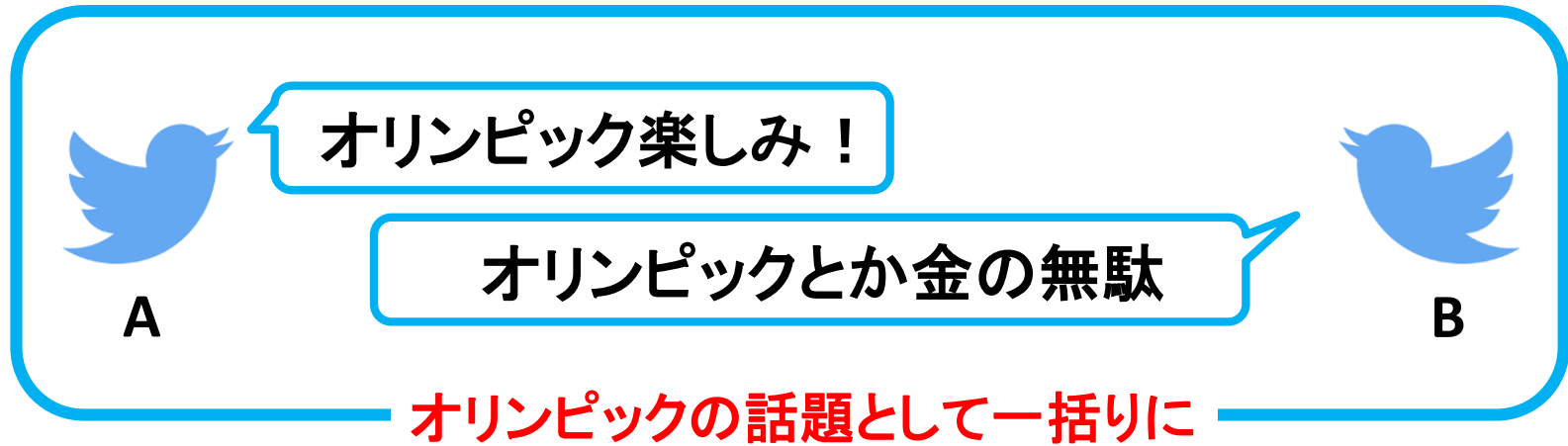
- 協調フ
- ツイート
- フォロイ

これらの手法では  
話題に対する感情を考慮していない

- クラスタリングを用いてツイートから話題を抽出し、共通または類似の話題があるかに基づいて推薦する



# 背景



適切な推薦であるとは言い難い

## 山本ら[1]の研究

- 共通の話題とその話題に関する感情を考慮したフォロワー推薦を提案
- 感情の類似性の抽出の精度が十分であるとは言い難い

[1] Yuki Yamamoto, Tadahiko Kumamoto, Akiyo Nadamoto,  
“Followee Recommendation Based on Topic Extraction and Sentiment Analysis from Tweets,”  
International Conference on Information Integration and Web-based Applications & Services(iiWAS2015), pp. 215-225, 2015.

# 目的

SVMを用いた感情抽出に基づく  
フォロイー推薦手法の提案



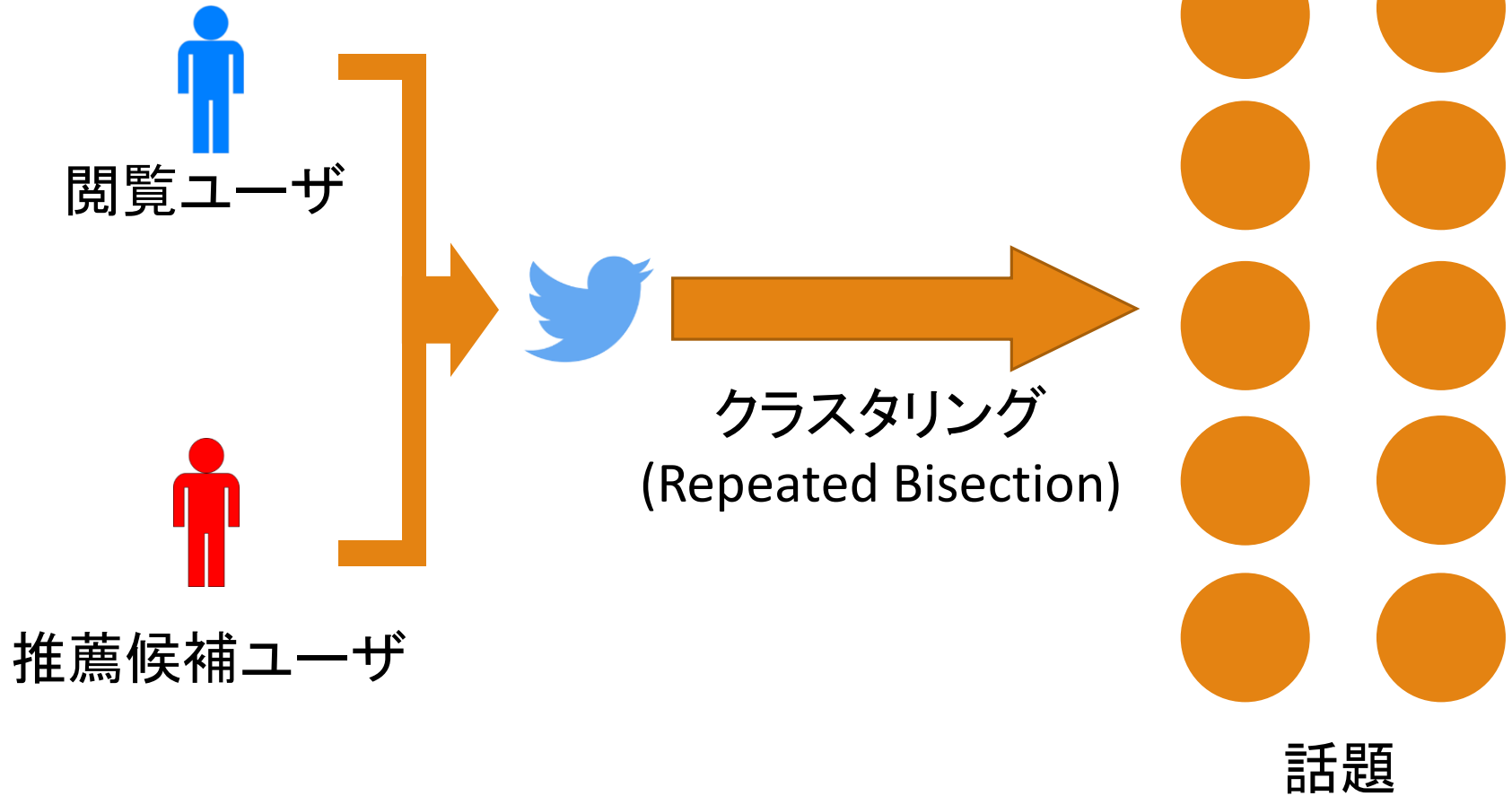


# 提案手法の流れ

1. 共通の話題の抽出
2. SVMによる感情の分類
3. 感情の類似性の抽出



# 話題のクラスタリング

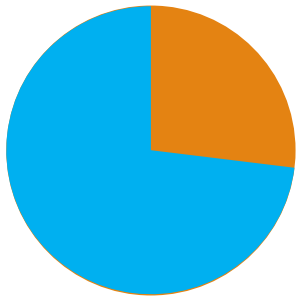


# 共通の話題の抽出

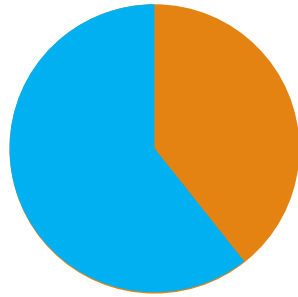
- 両ユーザのツイート数の比率
- ツイートの凝集性

# 両ユーザのツイート数の比率

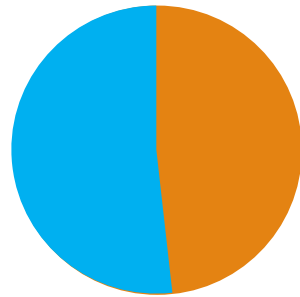
■ 閲覧ユーザ  
■ 推薦候補ユーザ



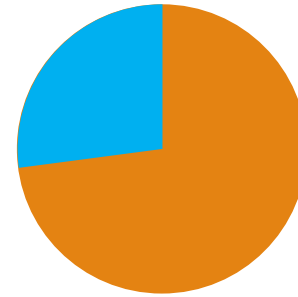
話題1



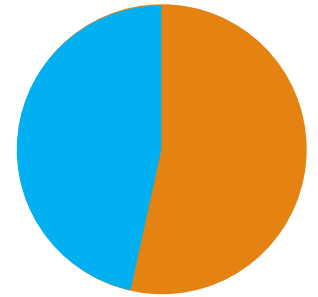
話題2



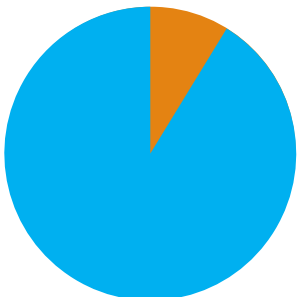
話題3



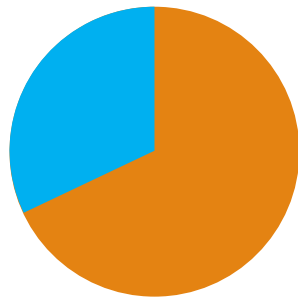
話題4



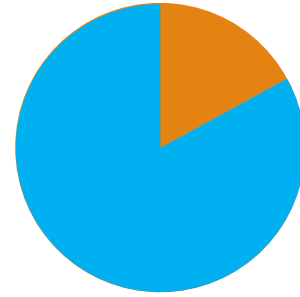
話題5



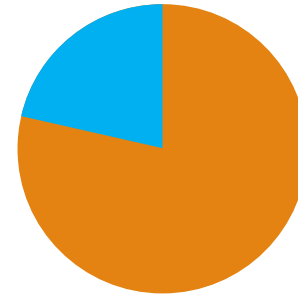
話題6



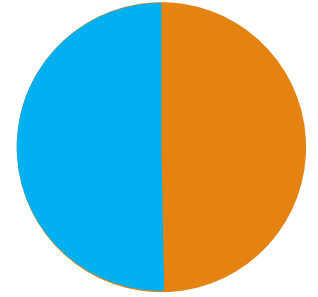
話題7



話題8



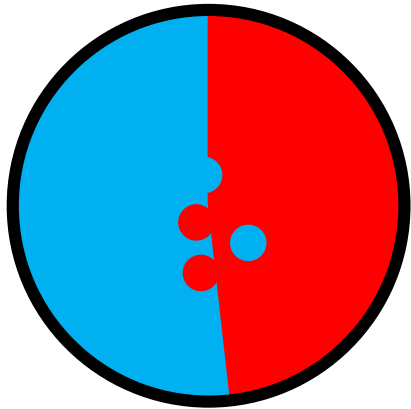
話題9



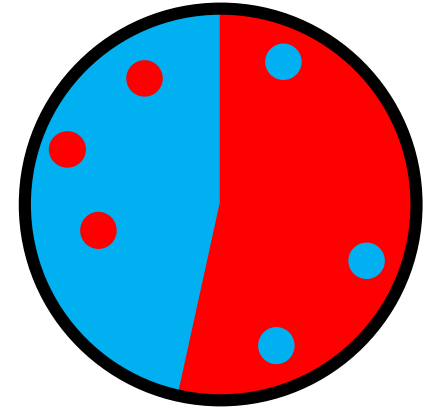
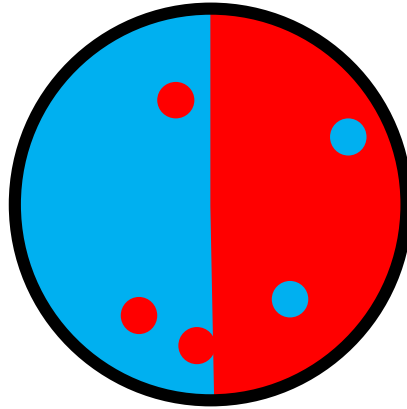
話題10

# ツイートの凝集性

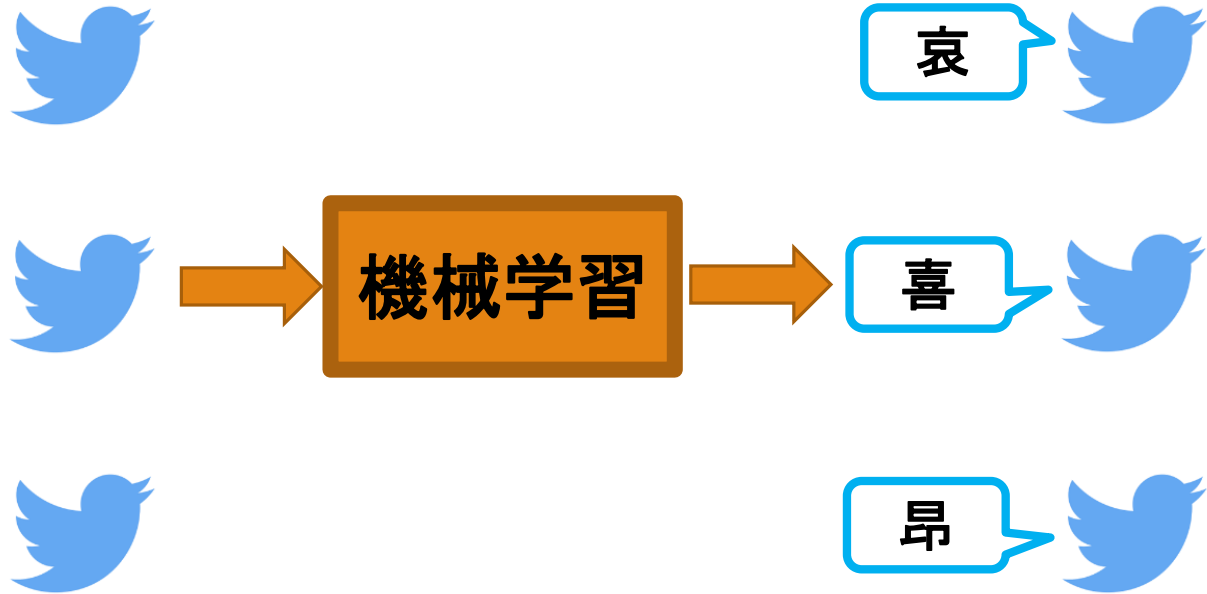
- 閲覧ユーザ
- 推薦候補ユーザ



共通話題クラスタ



# 感情の分類



## 8の感情

|     |   |   |   |   |     |   |   |
|-----|---|---|---|---|-----|---|---|
| 喜・好 | 安 | 昂 | 哀 | 怖 | 怒・厭 | 驚 | 恥 |
|-----|---|---|---|---|-----|---|---|

# 機械学習

- SVMを使用
- カーネルはガウスクーネルを使用

# 機械学習

- 教師データには我々の行ったアンケートにより、感情毎に分類されたツイートを用いる

- 25件のツイート
- 10人の被験者
- 各軸5段階評価



58セット

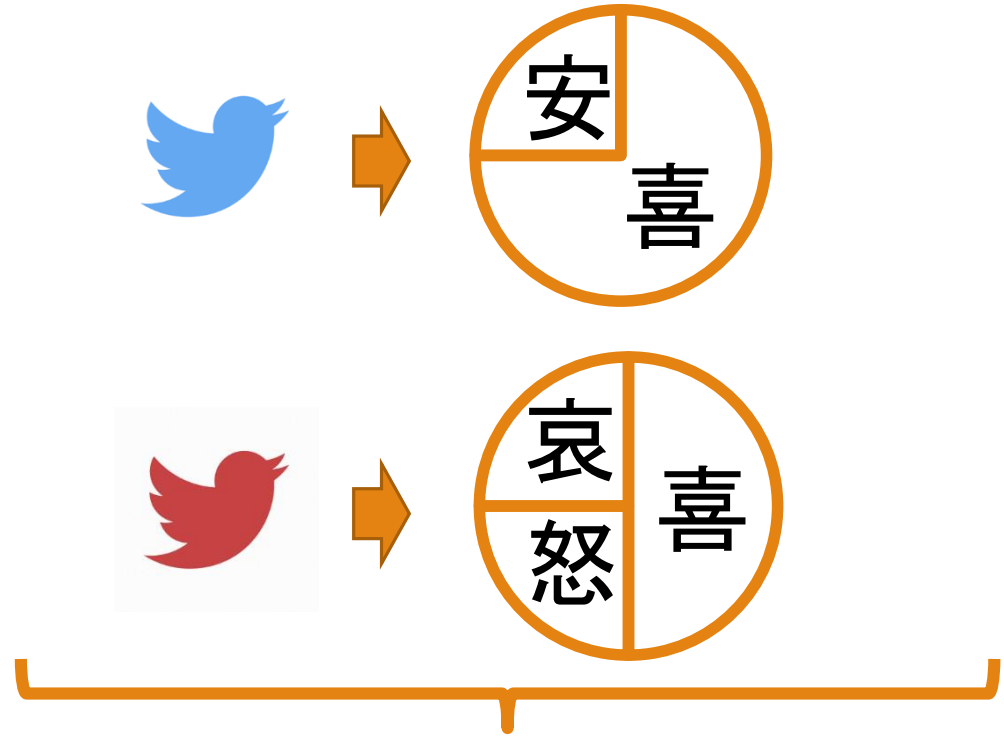
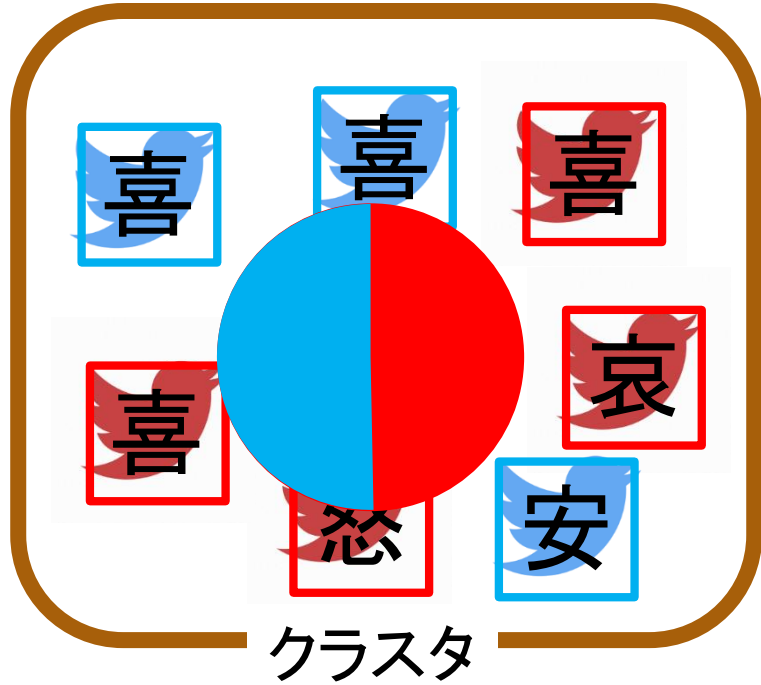


1450ツイート



# 機械学習

- 素性は全教師データに対し形態素解析を行い、  
得られた名詞・動詞・形容詞を対象とし、  
各々の単語が各ツイートに含まれているか否かとする



ユーザ間の類似度

JS-divergence

## JS-divergence

2つの確率分布間の差異を測る尺度

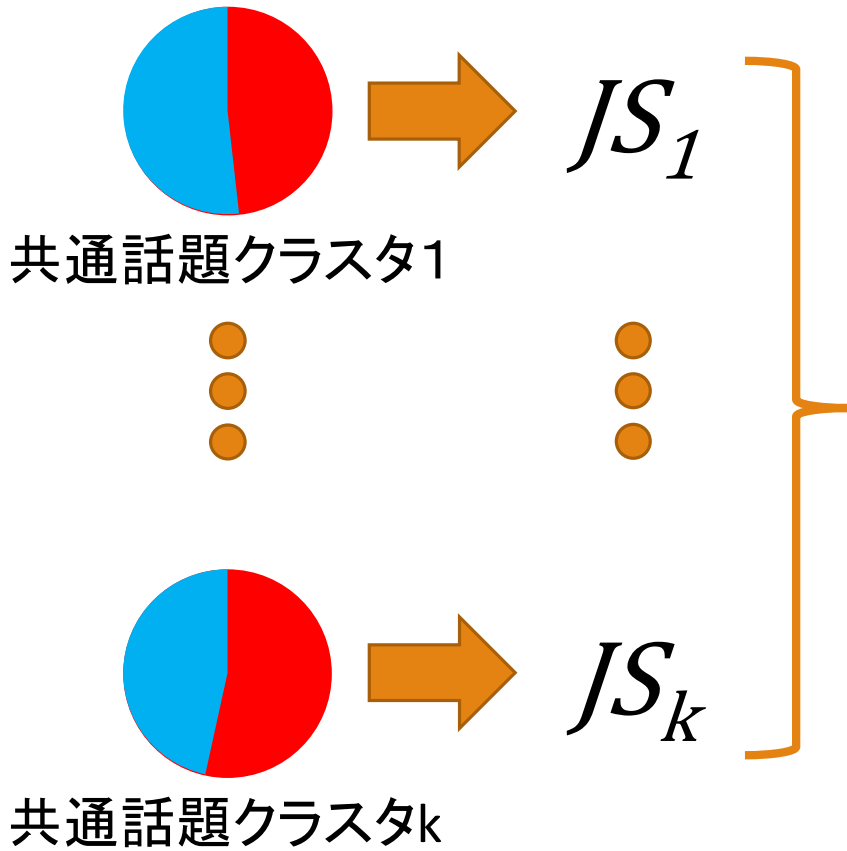
$$JS = \frac{1}{2}KL(P||M) + \frac{1}{2}KL(Q||M)$$

$$KL(P||Q) = \sum_{i=1}^8 P(i) \log_2 \frac{P(i)}{Q(i)}$$

$$M_i = \frac{1}{2}P(i) + Q(i)$$

P: 閲覧ユーザの8軸の感情それぞれの割合

Q: 推薦候補ユーザの8軸の感情それぞれの割合



各共通話題クラスタの類似度の和が  
全推薦候補ユーザの類似度の  
平均未満の場合、推薦候補ユーザを  
フォロイー候補として推薦

# 実験

- 被験者：4名の大学生
- 被験者が自身のffユーザ200名の中からフォローしたいと評価したユーザを正解データとし、適合率、再現率、F値を求める

| ユーザ  | 適合率   | 再現率   | F値    |
|------|-------|-------|-------|
| ユーザ1 | 0.072 | 0.450 | 0.125 |
| ユーザ2 | 0.106 | 0.667 | 0.183 |
| ユーザ3 | 0.129 | 0.653 | 0.250 |
| ユーザ4 | —     | —     | —     |

## 考察

- ユーザ1-3の適合率が低い  
→機械学習の分類精度が低いため
- ユーザ1の値が特に低い  
→様々なユーザと繋がりがあるので話題の偏りが少なく、  
個々のユーザとの類似性が低くなっているため
- ユーザ4  
→本人のつぶやきと推薦候補のつぶやきが全く異なっており、  
共通話題クラスタが抽出されなかった

# 実験結果

## 良い例

- サッカー好きのユーザに、サッカーに関するツイートをしており、気が合うユーザを推薦した

## 悪い例

- 声優に興味のないユーザに声優のアカウントを推薦した
  - 閲覧ユーザのフォロワーに声優に興味のあるユーザがいた
  - “予約”という語が両者のツイートに共通して出現していた

# 今後の課題

- 機械学習の精度向上
- ユーザのツイート・フォロー傾向を考慮
- 感情の強弱の考慮