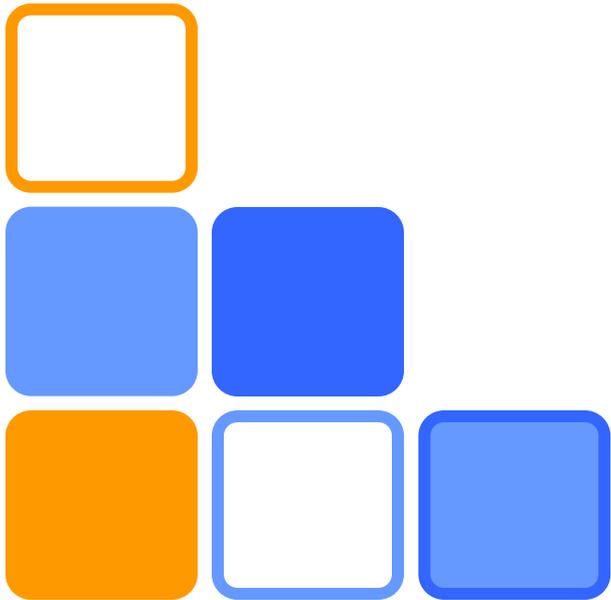
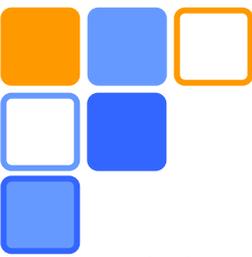


映画に対する実況ツイートの 感情抽出手法の提案

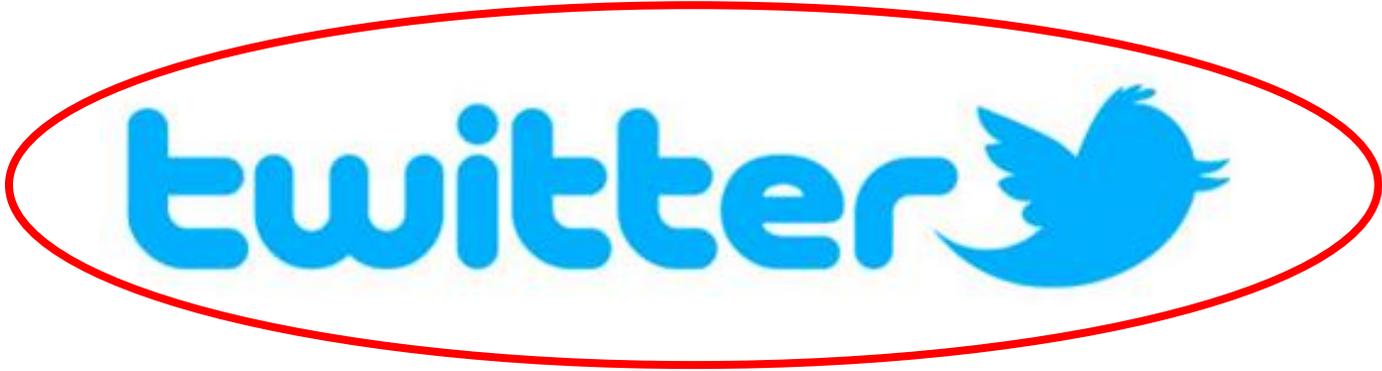


☆若井 祐樹(甲南大学)
熊本 忠彦(千葉工業大学)
灘本 明代(甲南大学)



背景

近年、TwitterやFacebookといったソーシャルメディアが発展している

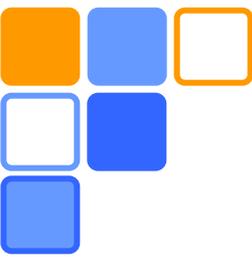


twitter



気軽に発信できることが利点

Twitterで自分が思っていることや
感じていることをその場で発しているのでは



背景

テレビ番組を視聴しながらの実況ツイート

スポーツ

やったー！よかった…。星野さん、宙に舞えてよかった！

ドラマ

法律が佳境を迎えてテンションおかしくなってる。

映画

なんか悲しいお話。でも意味がよく分からない。
三番目の弟がハリーのお父さんとか？

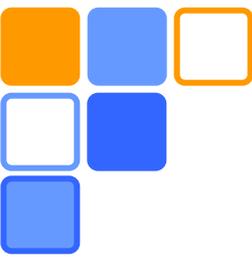
実況ツイート

番組出演者やキャラクターのセリフ

番組に対する自分の意見や感想

視聴者の感情が多く含まれる

実況ツイートから感情を抽出する手法を提案

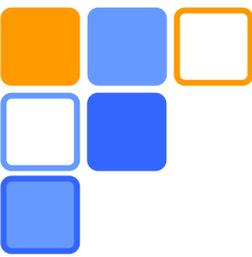


実況ツイートから感情を抽出する利点

時系列毎に実況ツイートを見て
感情を抽出することで・・・

- ◆番組に対する感情の流れを読み取ることができる
- ◆感情の大きさから盛り上がりシーンを探索することができる





目的

実況ツイートからユーザの感情を抽出する

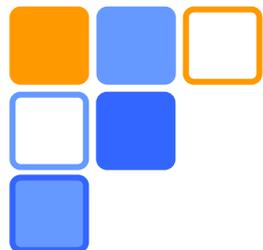
テレビで放送されている映画の実況ツイートに着目

- ◆ 一般の文章を対象として
 - 既存辞書を用いた実況ツイートの感情抽出

- ◆ Twitter特有表現を考慮した

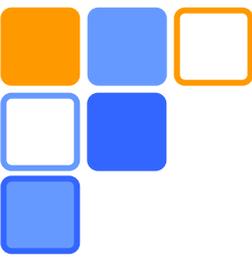
- 顔文字 (^o^)
- 若井祐樹, 熊本忠彦, 灘本明代, ツイートの感情抽出の為の顔文字の役割分類, 第2回Web インテリジェンスとインタラクション研究会(ARG SIG-WI2), 2013.

- 繰り返し表現 (よっしゃああああ)



①既存辞書を用いた実況ツイートの感情抽出

②Twitter特有表現を考慮した感情抽出

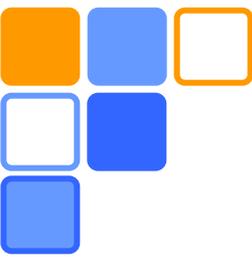


既存の感情辞書について

感情に関する辞書はいくつか存在する

①感情表現語辞書

②感情表現辞典



感情表現語辞書

熊本ら(千葉工大)が提案・構築した感情辞書
過去5年間の新聞記事データベースに現れる任意の単語とあらかじめ
定義されている感情語群との共起関係から構築している

反義語関係にある2つの感情語から構成されている→感情軸

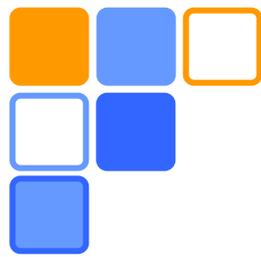
悲しい⇔楽しい

怒り⇔うれしい

緊迫⇔のどか

単語	-3.0 感情軸1 3.0 悲しい⇔楽しい	-3.0 感情軸2 3.0 怒り⇔うれしい	-3.0 感情軸3 3.0 緊迫⇔のどか
初受賞	2.172 楽しい	3.000 うれしい	1.848 のどか
勝つ	1.236	1.68	-1.878
死刑	-2.694	-2.514	-1.23
拘束する	悲しい -2.646	怒り -2.382	緊迫 -3.000

感情値算出手法(感情表現語辞書)



ダメだ、もう既に泣きそう

ダメだ/、/もう/既に/
泣き/そう

悲しい⇔楽しい

ダメだ	-0.48517
既に	-0.24417
泣き	-1.70719
合計	-2.04034

悲しい⇔楽しい

$-2.04034 / 3 = -0.68011$

怒り⇔嬉しい 少し悲しい

$-1.99762 / 3 = -0.66587$

緊迫⇔のどか 少し怒り

$-1.26347 / 3 = -0.42116$

少し緊迫

映画の実況ツイート

形態素解析

感情語と感情値を抽出

文の感情を決定
感情語毎の感情値の合計
感情語数

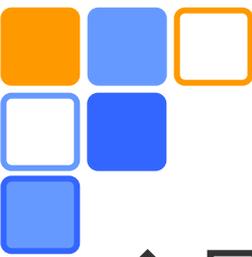
次の
ツイートへ

辞書毎に
処理

感情表現語辞書
悲しい⇔楽しい

感情表現語辞書
怒り⇔嬉しい

感情表現語辞書
緊迫⇔のどか



感情表現語辞書を用いた 感情抽出の実験

◆ 目的

ツイートの感情抽出における問題点の洗い出し

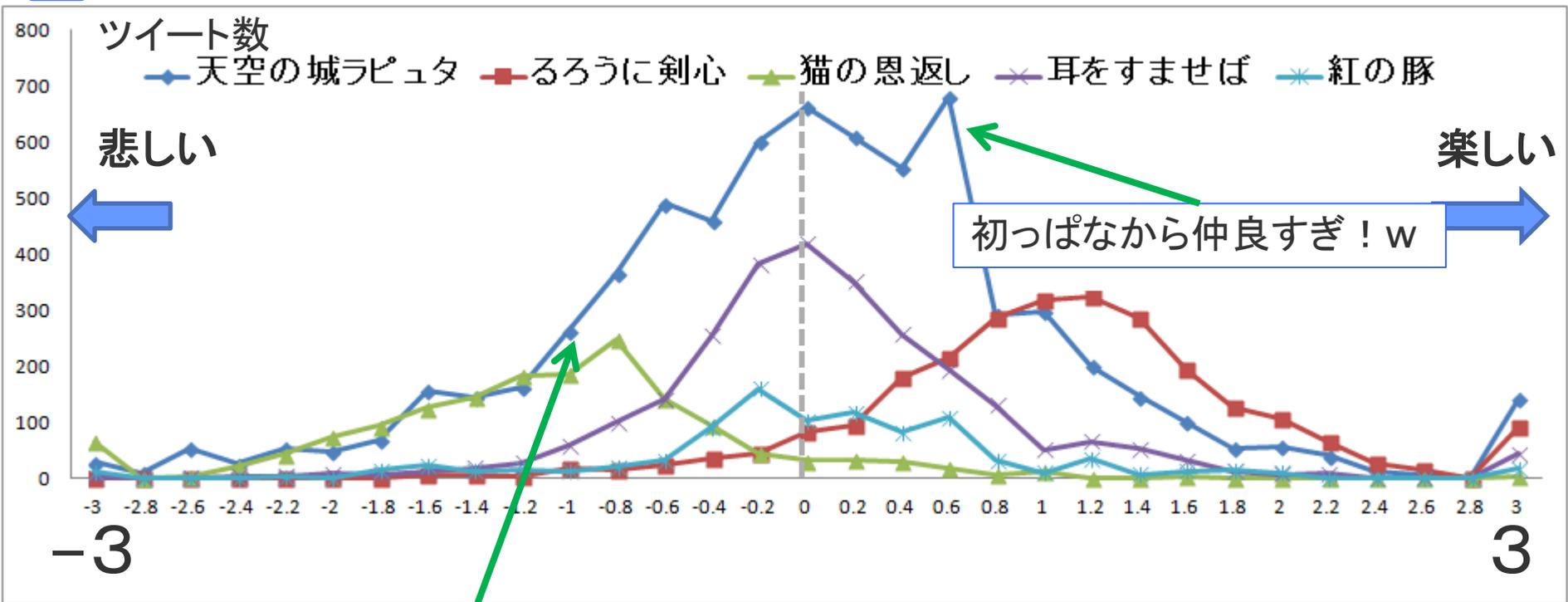
収集データ: テレビで放送された映画5作品の実況ツイート

映画名	ツイート数
天空の城ラピュタ	8124
るろうに剣心	2616
猫の恩返し	1666
耳をすませば	2742
紅の豚	1084

◆ 実験方法

- ① 収集したツイートからリプライ(@)・リツイート(RT)・ハッシュタグ(#)・URLを削除
- ② 熊本らが構築した感情表現語辞書を用いた感情抽出手法により
各々のツイートの感情値を求める

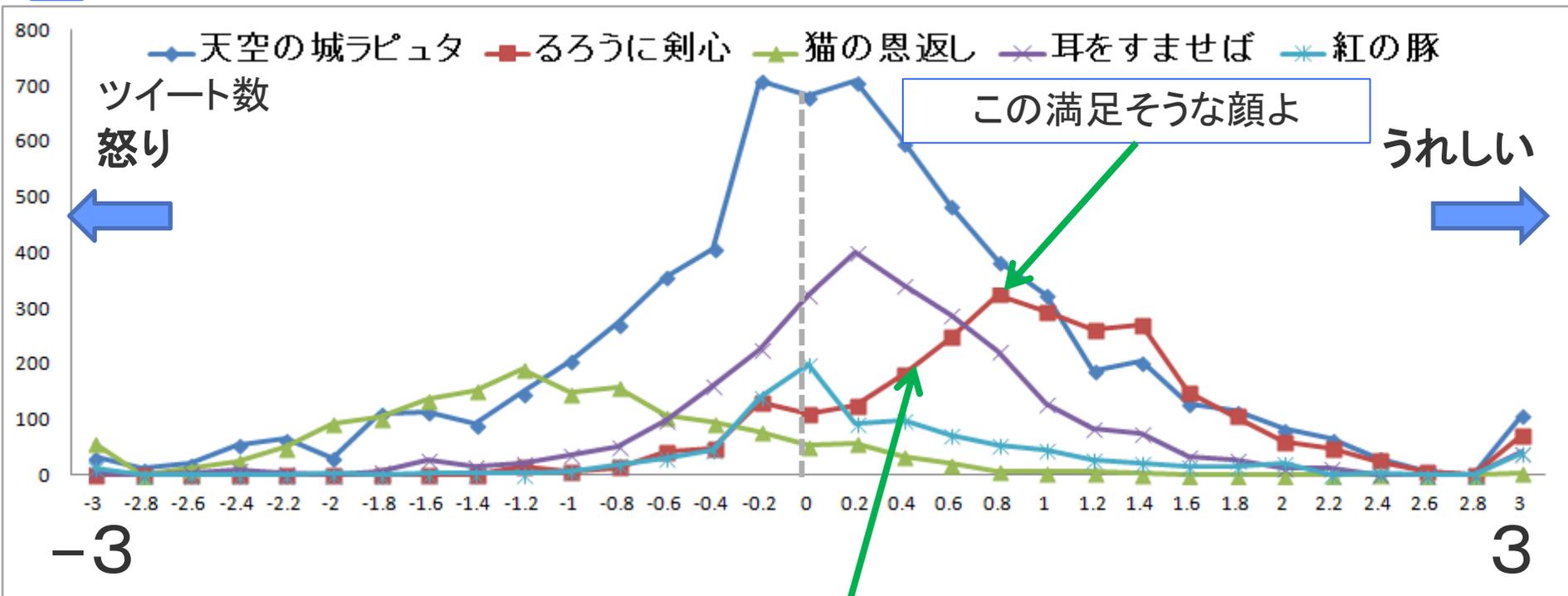
実験結果(悲しい⇔楽しい)



何気にドーラ格好いい(°▽°)

楽しそうな感情のイメージなのに
悲しいにふれてしまっている

実験結果(怒り⇔うれしい)

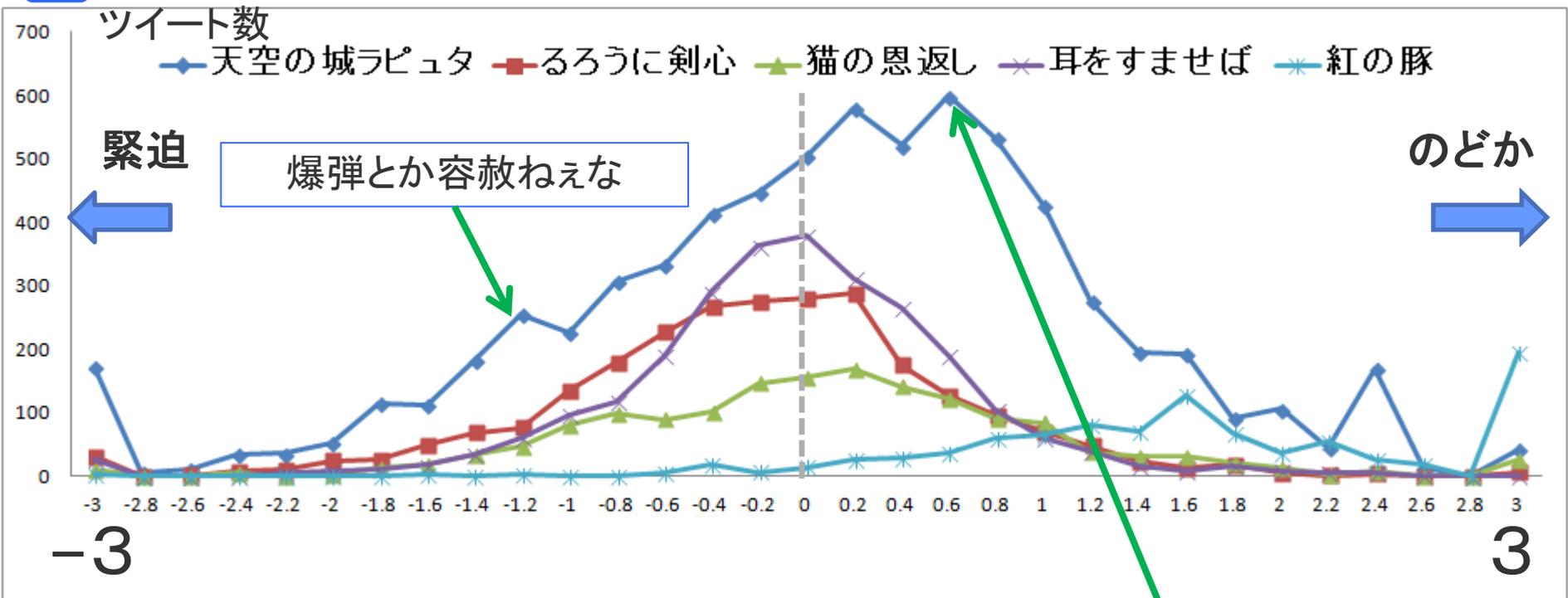


-3

3

値は嬉しいになっているが
イメージができない

実験結果(のどか⇔緊迫)



うわああああ

のどかのような
落ち着いたイメージが欠けている

実験結果(全感情軸)

ツイート数
悲しい

楽しい

3

3

怒り

うれしい

3

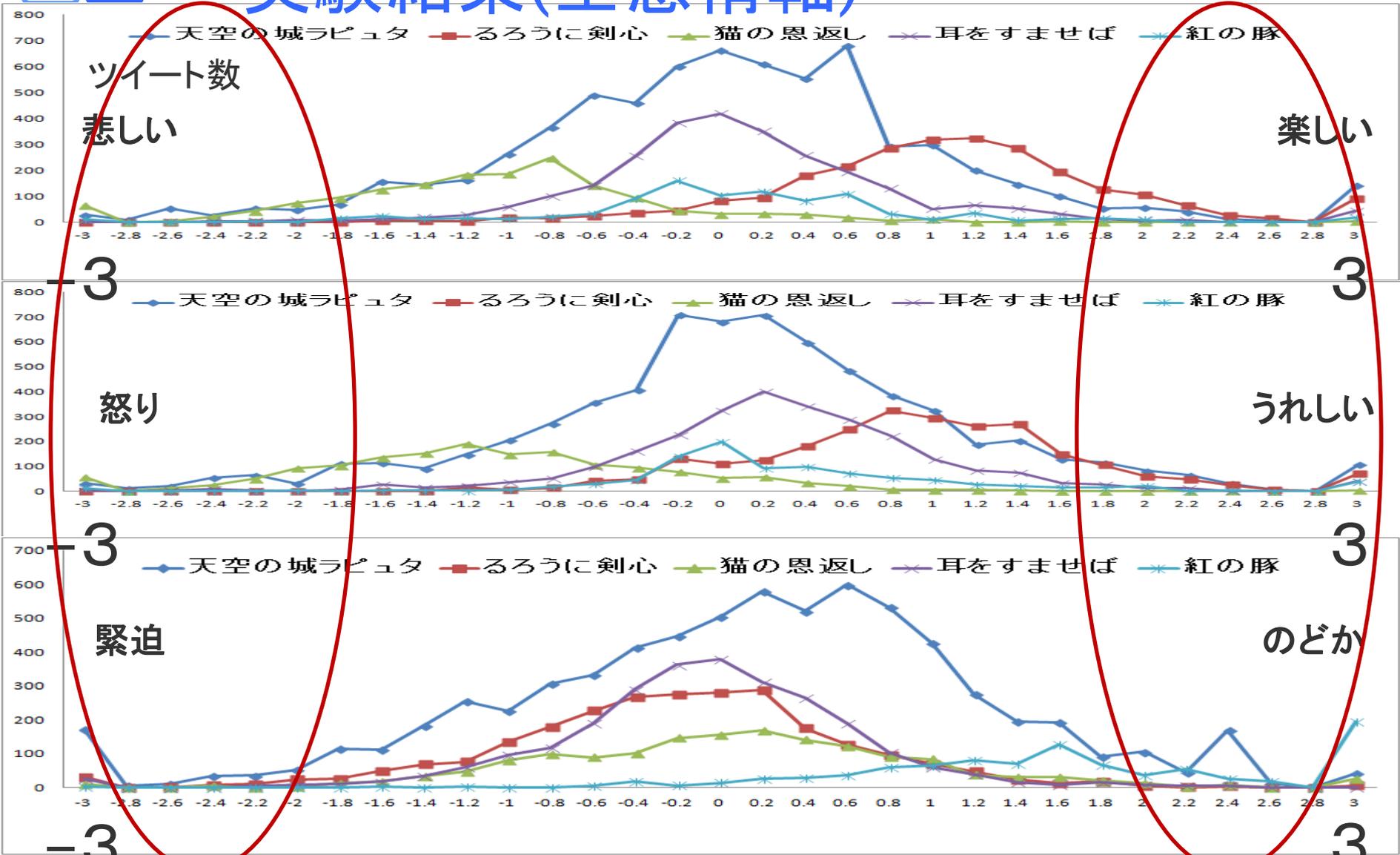
3

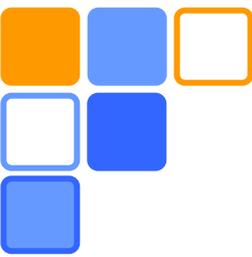
緊迫

のどか

-3

3





うまくいかなかった原因

感情値が顕著に表さなかった

実験で用いたツイートは感情表現語辞書では対応できない語句が多数存在していた

感情表現語辞書の不足

解決策

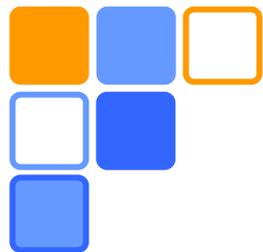
① 感情表現語辞書の拡張

Twitterでよく使われる語句に対応できるように辞書を拡張する

② 感情軸の拡張

様々な感情に関する語句から感情を抽出する

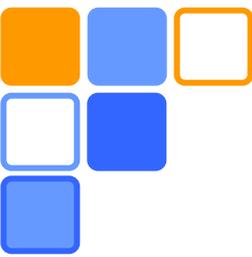
→ 10軸からなる感情表現辞典



既存の感情辞書

①感情表現語辞書

②感情表現辞典



感情表現辞典

中村明の感情表現辞典は小説の用例を10軸の感情に分類

喜

哀

怒

怖

好

厭

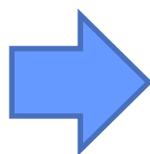
昂

驚

安

恥

感情	単語	感情	単語
喜	楽しい, 嬉しい, 心が引かれる	怒	怒る, 甚だしい, むっとする
哀	悲しい, 痛い, じいんと来る	怖	怖い, 震える, 気味悪い
恥	恥ずかしい, 照れる, 穴に入りたい	好	恋しい, 愛しい, 敬意を表する
厭	暗い, 憎い, 愛想をつかす	昂	苛立つ, 感情, 心が張り詰める
安	のんびり, すっきり, 気を鎮める	驚	ぼんやり, 歓喜, 目を丸くする



実況ツイートの感情の抽出を行い
どれだけ対応できるのか検証する

感情値算出手法(感情表現辞典)



映画の実況ツイート

猫の照れ顔可愛い。

猫/の/照れ/顔/可愛い/。

照れる→(恥)
可愛い→(好)

次の
ツイートへ

辞書毎に
処理

感情表現辞典
(喜)

-
-
-

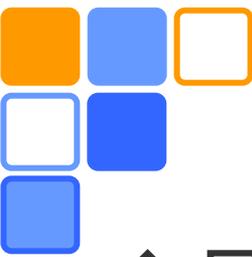
感情表現辞典
(恥)

形態素解析

感情語とその感情を取得

取得した感情語に対する
感情の出現頻度
→ 1ツイートの感情値

	喜	好	安	哀	厭	怖	怒	恥	昂	驚
照れる	0	0	0	0	0	0	0	1	0	0
可愛い	0	1	0	0	0	0	0	0	0	0
感情値	0	1	0	0	0	0	0	1	0	0



感情表現辞典を用いた 感情抽出の実験

◆目的

- 映画の実況ツイートにどれだけ対応が可能か
- ツイートの感情抽出における問題点の洗い出し

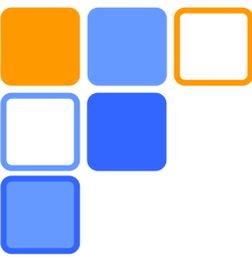
◆収集データ:

- 前実験同様 映画5作品の実況ツイート

映画名	ツイート数
天空の城ラピュタ	8124
るろうに剣心	2616
猫の恩返し	1666
耳をすませば	2742
紅の豚	1084

◆実験方法

- ① 収集したツイートからリプライ(@)・リツイート(RT)・ハッシュタグ(#)・URLを削除
- ② 感情表現辞典を用いた感情抽出手法により感情抽出を行う



実験結果

映画名	感情がとれたツイートの割合
天空の城ラピュタ	10.7%
るろうに剣心	14.7%
猫の恩返し	10.4%
耳をすませば	13.3%
紅の豚	10.7%

楽しむ	そわそわ	ショック	愛でる
可愛い	嫌う	鬱陶しい	むかつく
困る	にやにや	眺める	呆気ない

ツイート内に含まれている
感情表現辞典に収録されている単語

284 単語(全3136単語)

うまくいかなかった原因

感情	単語	感情	単語
喜	楽しい, 嬉しい, 心が引かれる	怒	怒る, 甚だしい, むっとする
哀	悲しい, 痛い, じいんと来る	怖	怖い, 震える, 気味悪い
恥	恥ずかしい, 照れる, 穴に入りたい	好	恋しい, 愛しい, 敬意を表する
厭	暗い, 憎い, 愛想をつかす	昂	苛立つ, 感情, 心が張り詰める
安	のんびり, すっきり, 気を鎮める	驚	ぼんやり, 歓喜, 目を丸くする

複合語や文で感情を表現している語句が多数

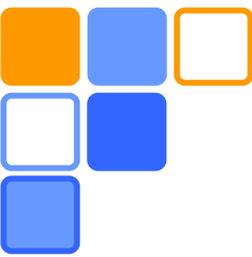


形態素解析を行った際に該当されなかった

Twitterでよく使われている特有な表現

顔文字

繰り返し表現



うまくいかなかった原因

① 顔文字

ジブリはおばさんがかっこいい(´ー`)

(((:°Д°)))がカクブルブル

顔文字を使って、自分の感情を表現している

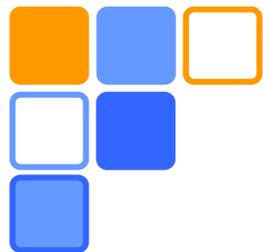
② 繰り返し表現

くるううううううう

きったああああああああああああああああああ

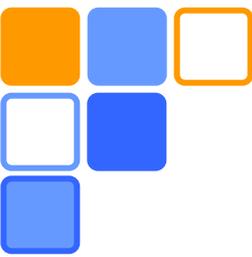
叫ぶように感情を表現している

繰り返し表現を焦点に当て 感情抽出を行う



①既存辞書を用いた実況ツイートの感情抽出

②Twitter特有表現を考慮した感情抽出



Twitter特有表現

叫喚フレーズ

「きたああああ！」のように
崩れた表記をして母音を繰り返す表現

浅井洋樹, 秋岡明香, 山名早人. きたああああああああああああああああ！！！！11: マイクロブログを用いたことにより教師なし叫喚フレーズ抽出, 第5回データ工学と情報マネジメントに関するフォーラム(DEIM Forum2013), A4-4, 2013.



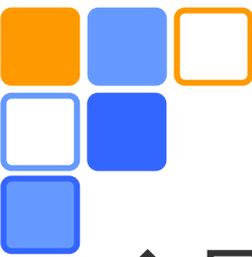
Twitter特有表現

◆ 定義

- 同じ母音を3つ以上繰り返してある
- 大文字・小文字を区別しないもの

ツイート例

猫ばばするなああああああ！！！！



Twitter特有表現に対するユーザ実験

◆ 目的

Twitter特有表現が使われていること
どのような感情の変化があるのか

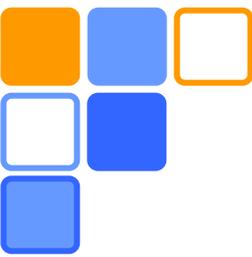
◆ 被験者: 11名

◆ 実験データ

Twitter特有表現がある実況ツイート50ツイート

◆ 実験方法

- ① Twitter特有表現の部分を削除したツイートを見て感情表現辞典の10軸の感情から最大3つまで選び感情の強さを10点満点で評価する
- ② Twitter特有表現のあるツイートを見て①と同様に評価してもらう



評価方法

$$TW_{i,j} = \frac{1}{h} \sum_{i=1}^h \frac{(R-NR)}{10}$$

$TW_{i,j}$: ある感情軸*i*におけるあるツイート*j*の評価値

h : 被験者の人数

R : ツイート*j*における*Twitter*特有表現がある場合のある感情軸*i*におけるツイートを評価した点数

NR : ツイート*j*における*Twitter*特有表現がない場合のある感情軸*i*におけるツイートを評価した点数

$\begin{cases} TW_{i,j} > 0: \textit{Twitter} \text{ 特有表現によって感情をより強めている} \\ TW_{i,j} < 0: \textit{Twitter} \text{ 特有表現によって感情をより弱めている} \end{cases}$

実験結果と考察

感情	感情(強)	感情(弱)	感情変化なし
喜	67%	2%	31%
哀	35%	10%	55%
怒	27%	2%	71%
怖	29%	10%	61%
好	80%	2%	18%
厭	33%	14%	53%
昂	100%	0%	0%
驚	84%	4%	12%
安	8%	41%	51%
恥	14%	0%	86%

〇〇かわいいよお

〇〇かわいいよおおおおおおお

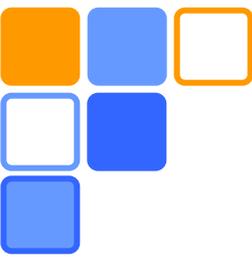
気持ちが高揚していることを表現

→ 「安」の感情が弱める

◆Twitter特有表現

- 感情を強調する機能がある
- 「安」は反対に弱めている

→ 感情値算出を行う



まとめと今後の課題

◆まとめ

- 映画の実況ツイートに絞り、感情の抽出を行った
- 既存の感情辞書を用いて、ツイートの感情抽出する場合の問題点の洗い出し
- Twitter特有表現である繰り返しに着目したユーザ実験による感情分析を行った

◆今後の課題

- Twitter特有表現の定量化
- ツイートに多く存在する未定義語の対策
- 実況ツイートに特化した感情語辞書の構築
- 実況ツイートにおける感情の変化等の分析