

SNS上の文章を用いた 感情辞書構築の有効性検証

大澤 泰我[†]

マツキン ケネスジェームス[†]
[†]東京情報大学

永井 保夫[†]

研究背景

- 近年コミュニケーションロボットの需要が高まっている
- 対話システムと人との自然なコミュニケーションの実現が求められている
- 人との自然な対話の実現において、感情情報は必要不可欠

対話システムの問題点

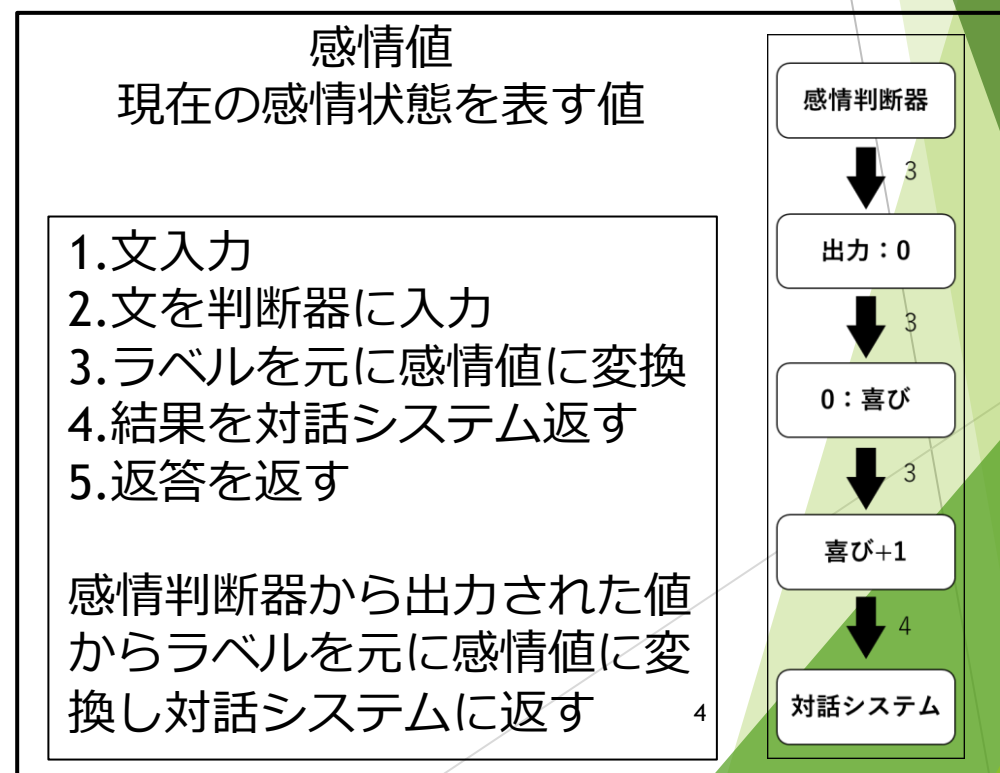
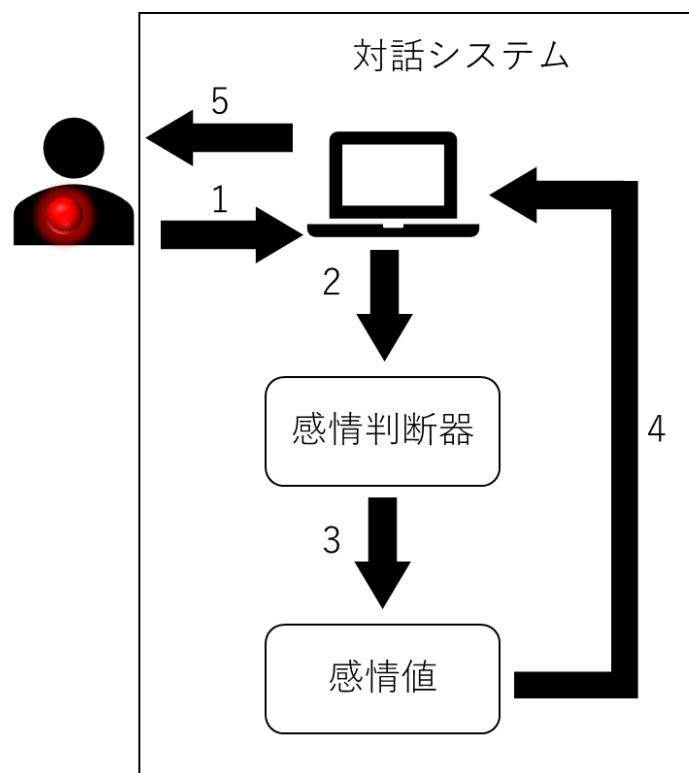
- 対話の際に感じる心理的抵抗感や発話が機械的
- 解決する方法の一つとして、感情を考慮した対話の実現が考えられている[1][2]

[1]河野 泉,池邊 亮志,和氣 早苗,上窪 真一,岩沢 透,西村 健士.感情表現を用いた対話システムEDSの開発(1): システム概要と感情モデル.情報処理学会研究報告, “HI,ヒューマンインタフェース研究会報告 89”, pp.43-48,(2000)

[2]河野 泉,池邊 亮志,和氣 早苗,上窪 真一,岩沢 透,西村 健士.感情表現を用いた対話システムEDSの開発(2): システム概要と感情モデル.情報処理学会研究報告, “HI,ヒューマンインタフェース研究会報告 89”, pp.49-56,(2000)

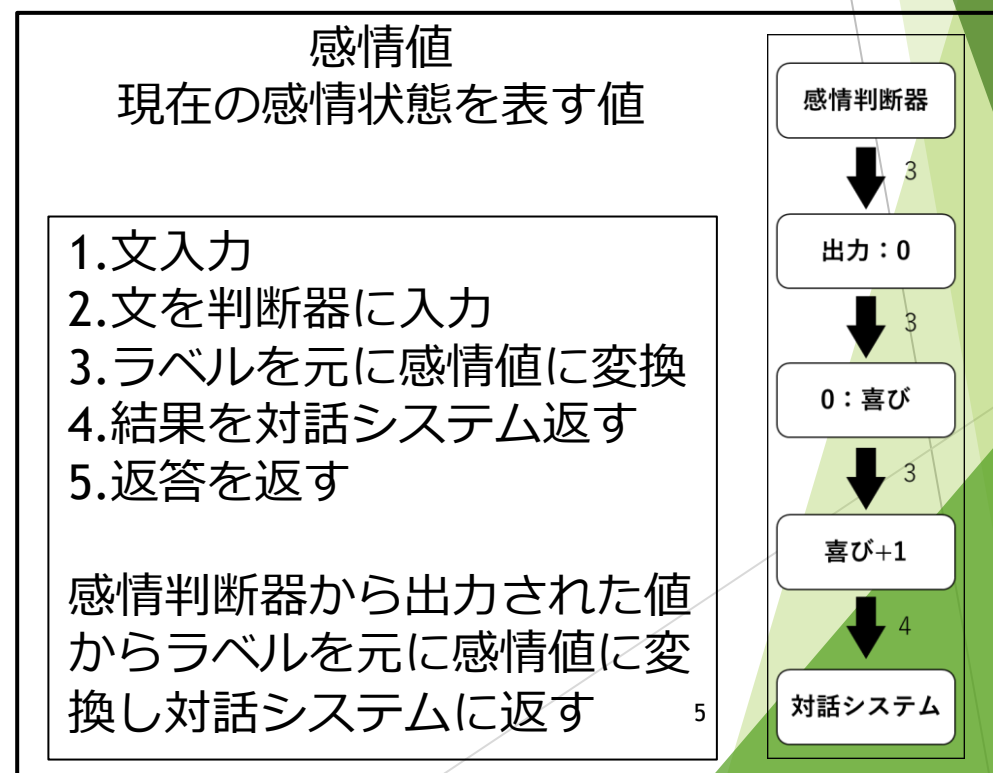
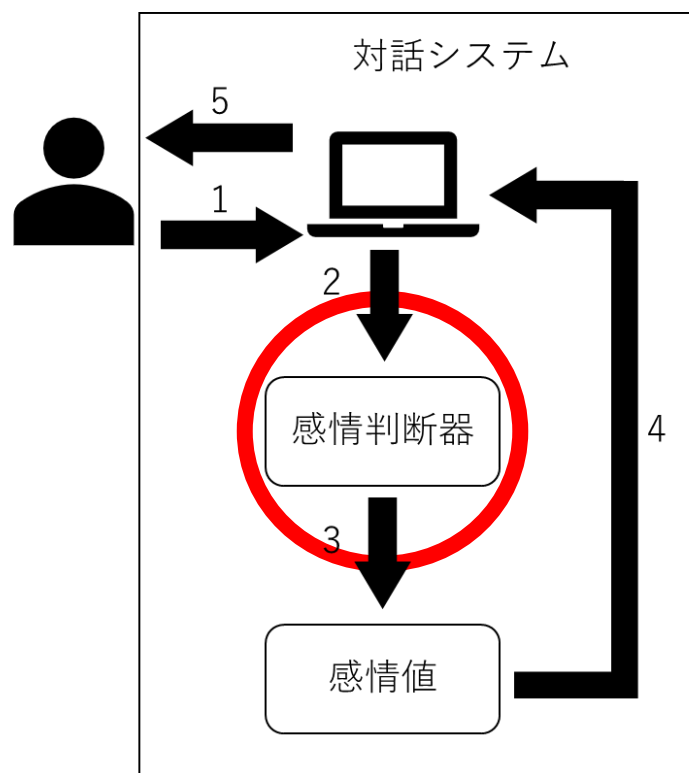
感情を考慮する対話システム

- 感情判断器を組み込むことで、対話システムは感情値の違いによる文章生成を行う



感情を考慮する対話システム

- 感情判断器を組み込むことで、対話システムは感情値の違いによる文章生成を行う



研究目的

- 対話システムから受け取った入力文から感情を判断し、対話システムに感情値を返す
- データセットを作りLSTMを用いて感情判断器を作成する

感情辞書の現状と問題

	関連研究1[3]	関連研究2[4]	本研究[5]
作成方法	人手によるラベリング	人手によるラベリング	自動ラベリング
	専門家3人が決定したラベルの平均	学生20人が決定したi番目に多いラベルの割合	文章に含まれる顔文字を元にラベリング
件数	約2千	約1千	約7万
辞書内容	ラベル, ルビ, 単語	ラベル, 文章, 意図	ラベル, 文章

- 感情語の辞書を作るのには、多くの時間と労力を必要とする
- 辞書は特定の感情モデルに依存する

[3] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka, Kenji Araki. A system for affect analysis of utterances in Japanese supported with web mining. 日本知能情報ファジイ学会誌, vol.21, No.2, pp.194-213 (2009)

[4] 山本裕司, 吉村枝里子, 土屋誠司, 渡部広一. 会話文章からの話者の意図・感情判断. 情報処理学会研究報告, vol.2010-ICS-158, No.2 (2010)

[5] 大澤泰我, ケネス マッキン, 永井保夫. 対話システムのための感情判断器の検討. 情報処理学会第82回全国大会. 1S-06, (2020).

感情辞書の現状と問題

	関連研究1[3]	関連研究2[4]	本研究[5]
作成方法	人手によるラベリング	人手によるラベリング	自動ラベリング
	専門家3人が決定したラベルの平均	学生20人が決定したi番目に多いラベルの割合	文章に含まれる顔文字を元にラベリング
件数	約2千	約1千	約7万
辞書内容	ラベル, ルビ, 単語	ラベル, 文章, 意図	ラベル, 文章

- 感情語の辞書を作るのには、多くの時間と労力を必要とする
- 辞書は特定の感情モデルに依存する

[3] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka, Kenji Araki. A system for affect analysis of utterances in Japanese supported with web mining. 日本知能情報ファジイ学会誌, vol.21, No.2, pp.194-213 (2009)

[4] 山本裕司, 吉村枝里子, 土屋誠司, 渡部広一. 会話文章からの話者の意図・感情判断. 情報処理学会研究報告, vol.2010-ICS-158, No.2 (2010)

[5] 大澤泰我, ケネス マッキン, 永井保夫. 対話システムのための感情判断器の検討. 情報処理学会第82回全国大会. 1S-06, (2020).

感情辞書の構築手法

データセットの例

ありがとうございます！(^^)

(@_@) (T_T) Σ(° A° ;)
＼(^o^)/
(-_-;) (*´▽`*)

Twitterキーワード検索APIで
顔文字のついた文を収集、各パラ
メータ用のファイルにまとめる

Twitter

約7万9千件

喜び

悲しみ

信頼

嫌悪

恐れ

怒り

驚き

予期

データイメージ

0	喜び	$(\cong \nabla \cong)$ $(\wedge \wedge)$
1	信頼	$\varepsilon - (' \vee ' *)$ $(; ' ? ')$
2	恐れ	$\text{ガ}^* \text{クガ}^* \text{ク}(((; ^\circ \text{D}^\circ))) \text{フ}^* \text{ル}^* \text{ル}$ $\Sigma (^\circ \Delta^\circ)$
3	驚き	$\Sigma (=^\circ \omega^\circ =;) \text{マジ} !?$ $\Sigma (^\circ \text{D}^\circ) \text{アア} !?$
4	悲しみ	$(/ _ \cdot \cdot) \text{シクシク}$ $(\rho^\circ \cap^\circ) \text{グ ス}$
5	嫌悪	$(' _ \succ ') \text{フーン}$ $(/ _ - ;) \text{ハア...}$
6	怒り	$\text{むっ} \cdot \cdot \cdot (\overline{} \cap \overline{} \#$ $\text{ムカムカ} \cdots ((\circ (-^\circ - ;)$
7	期待	$((\circ (^\wedge \nabla ^\wedge) \circ)) \text{わくわく}$ $\text{♪} ((\circ (\text{〃} \frown \nabla \frown \text{〃}) \circ)) \text{♪} \text{ド}^* \text{キド}^* \text{キ}$

チューニングイメージ	
原文	# test @you ありがとうございます！(^.^) http://test.com
加工後	ありがとう ございます！

データの作成

例：喜びの顔文字「(^-^)」が含まれ文章を取得

デーセツ

Twitter

「ありがとうございます！(^-^)」

0,息子に『はい、ママ、金[★]取得
ル』って保護者カード苦[★]に[★]お[★]は[★]て
もらった。

「0,ありがとうございます！」

4,昨日食材買い込んだのに米[★]か[★]ない[★]

形態素解析

「0,ありがとう/ござい/ます/！」

7,学校が終わってサン[★]
期待を味わった後また学校[★]に[★]行く[★]
のか...？私はクリスマス[★]は[★]学校休[★]
むぜ...

「0,ありがとう ございます！」

0	喜び
1	信頼
2	恐れ
3	驚き
4	悲しみ
5	嫌悪
6	怒り
7	期待

辞書構築自動化のメリット

- 辞書を自動で構築できること
 - 従来手法は、高コストであり、個人差が出る
 - また、辞書の再構築や部分修正といったメンテナンスが容易ではない[6]
- 任意の感情をその都度抽出できること
 - 推定可能な感情の種類を限定した方法では、システム用途が限られてしまう

研究の概要[7]

- データセット
 - TwitterAPIから集めたオリジナルの辞書
- 形態素解析
 - MeCab
- 学習
 - LSTMネットワーク

感情辞書構築の有効性検証

- 感情を含む文章を実際に人が分類・決定し、文章データの正解ラベルと人が判断した感情表現が一致しているかの検証
- アンケート形式
 - 感情辞書から各感情表現につき2件ずつランダムに文章を選ぶ
 - 正解ラベルを伏せて、アンケート調査を行う

	喜び	信頼	恐れ	驚き	悲しみ	嫌悪	怒り	期待	合計
件数	2	2	2	2	2	2	2	2	16

検証手法

正解ラベル	文章データ
喜び	タグ反応ありがとうございます
喜び	同い年の人少ないので私も嬉しいです！
信頼	午前中に用事があるけどお昼配信には間に合うようだからホッとしています。
信頼	とにかくエアコンが比較動いてくれてホッとしている
恐れ	どんどん狩られるんだろうなあ…
恐れ	今日は日本語能力試験を受けます
驚き	なんか電車1時間位遅れてるんやけどマジ？
驚き	今日色々あって遅刻で学校行くんよ
悲しみ	見に行きたかったー！
悲しみ	もっと韓国語出来たらたくさん話せるのに
嫌悪	今はわざわざ、赤の他人の赤ちゃん写真を延々とスクロールして見られるくらい
嫌悪	みんなに生きる希望がある事を知らしめています！
怒り	このままだと気持ちも身体も壊れてしまうわ
怒り	電車でおっさんに意地悪された。
期待	今日はみんなにとってどんな日だったかなあ～
期待	頑張っってね楽しみにしてる～！

タグ反応ありがとうございます

この文章は、どの感情に当てはまると思いますか？

喜び：希望が達成された時や、やさしさを感じた時の爽やかな気持ち
信頼：心配することなく、信じて安心できる気持ち
恐れ：害悪や危険な事柄に対して、逃避したいと感じる気持ち
驚き：予期しない事象を体験した時の瞬間的な感情
悲しみ：物事がうまくいかなかった時や、大切なものを失ったときに感じる残念な気持ち
嫌悪：憎み嫌い、不快に感じる気持ち
怒り：侮辱されたり傷つけられたりした時に起こる不愉快な気持ち
期待：事柄が自分の思い道理になることを望む気持ち

☐ 喜び

☐ 信頼

☐ 恐れ

☐ 驚き

☐ 悲しみ

☐ 嫌悪

☐ 怒り

☐ 期待

回答結果

文章データ

喜び

タグ反応ありがとうございます

悲しみ

同い年の人少ないので私も嬉しいです！

一致数

アンケートの回答が、
文章データに対応する
正解ラベルと一致した数

同い年の人少ないので私も嬉しいです！

この文章は、どの感情に当てはまると思いますか？

喜び：希望が達成された時や、やさしさを感じた時の爽やかな気持ち

検証結果

	回答数	一致数	一致率
喜び	46	40	87.0%
信頼	47	8	17.0%
恐れ	46	23	50.0%
驚き	46	15	32.6%
悲しみ	46	31	67.4%
嫌悪	46	2	4.3%
怒り	47	12	25.5%
期待	46	35	76.1%

アンケート問題数	回答者数	合計回答件数
16問	25人	370件

回答数

各感情表現2件ずつの
文章データに対しての
回答数

一致数

アンケートの回答が、
文章データに対応する
正解ラベルと一致した数

一致率

回答数と一致数の比率

考察

	回答数	一致数	一致率
喜び	46	40	87.0%
信頼	47	8	17.0%
恐れ	46	23	50.0%
驚き	46	15	32.6%
悲しみ	46	31	67.4%
嫌悪	46	2	4.3%
怒り	47	12	25.5%
期待	46	35	76.1%

アンケート問題数	回答者数	合計回答件数
16問	25人	370件

回答数

各感情表現2件ずつの
文章データに対しての
回答数

一致数

アンケートの回答が、
文章データに対応する
正解ラベルと一致した数

一致率

回答数と一致数の比率

考察

	回答数	一致数	一致率
喜び	46	40	87.0%
信頼	47	8	17.0%
恐れ	46	23	50.0%
驚き	46	15	32.6%
悲しみ	46	31	67.4%
嫌悪	46	2	4.3%
怒り	47	12	25.5%
期待	46	35	76.1%

考察

	回答数	一致数	一致率
喜び	46	40	87.0%
信頼	47	8	17.0%
恐れ	46	23	50.0%
驚き	46	15	32.6%
悲しみ	46	31	67.4%
嫌悪	46	2	4.3%
怒り	47	12	25.5%
期待	46	35	76.1%

考察

	回答数	一致数	一致率
喜び	46	40	87.0%
信頼	47	8	17.0%
恐れ	46	23	50.0%
驚き	46	15	32.6%
悲しみ	46	31	67.4%
嫌悪	46	2	4.3%
怒り	47	12	25.5%
期待	46	35	76.1%

- Twitterの文章データから感情表現を決定する際に使う顔文字の選択が重要

結論

- Twitterからの顔文字を用いた感情辞書の構築は、文章データから感情表現を決定する際に使う顔文字の選択が重要である

結論

- Twitterからの顔文字を用いた感情辞書の構築は、文章データから感情表現を決定する際に使う顔文字の選択が重要である

一致率	感情表現			
50%以上	喜び	恐れ	悲しみ	期待
50%以下	信頼	驚き	嫌悪	怒り

今後の予定

- 顔文字を用いた感情辞書構築における顔文字と文章データに含まれる感情の差異について

0	喜び	$(\geq \nabla \leq)$ $(\wedge \wedge)$
1	信頼	$\varepsilon - (' \vee ` *)$ $(; ' ? ')$
2	恐れ	$\text{カ}^{\circ} \text{ク}^{\circ} \text{カ}^{\circ} \text{ク}^{\circ} (((;^{\circ} \text{D}^{\circ}))) \text{フ}^{\circ} \text{ル}^{\circ} \text{フ}^{\circ} \text{ル}^{\circ}$ $\Sigma (^{\circ} \Delta^{\circ})$
3	驚き	$\Sigma (=^{\circ} \omega^{\circ} = ;) \text{マジ} ! ?$ $\Sigma (^{\circ} \text{D}^{\circ}) \text{アア} ! ?$
4	悲しみ	$(/ _ \cdot ,) \text{シクシク}$ $(\rho ^{\circ} \cap ^{\circ}) \text{グ スン}$
5	嫌悪	$(' _ \succ ') \text{フーン}$ $(/ _ - ;) \text{ハア...}$
6	怒り	$\text{むっ} \cdot \cdot \cdot (\overline{\cap} \overline{\#}$ $\text{ムカムカ} \cdots ((\text{o} (- ^{\circ} - ;)$
7	期待	$((\text{o} (^{\circ} \nabla ^{\circ}) \text{o})) \text{わくわく}$ $\text{♪} ((\text{O} (\text{〃} \frown \nabla \frown \text{〃}) \text{O})) \text{♪} \text{ドキドキ}$

参考文献

- [1] 河野 泉,池邊 亮志,和氣 早苗,上窪 真一,岩沢 透,西村 健士.感情表現を用いた対話システムEDSの開発(1): システム概要と感情モデル.情報処理学会研究報告, “HI, ヒューマンインタフェース研究会報告 89”, pp.43-48,(2000) .
- [2] 河野 泉,池邊 亮志,和氣 早苗,上窪 真一,岩沢 透,西村 健士.感情表現を用いた対話システムEDSの開発(2): システム概要と感情モデル.情報処理学会研究報告, “HI, ヒューマンインタフェース研究会報告 89”, pp.49-56,(2000) .
- [3] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka, Kenji Araki. A system for affect analysis of utterances in Japanese supported with web mining. 日本知能情報フアジイ学会誌, vol.21, No.2, pp.194-213 (2009)
- [4] 山本裕司, 吉村枝里子, 土屋誠司, 渡部広一. 会話文章からの話者の意図・感情判断. 情報処理学会研究報告, vol.2010-ICS-158, No.2 (2010)

参考文献

- [5] 大澤泰我,ケネス マッキン,永井保夫.対話システムのための感情判断器の検討.情報処理学会第82回全国大会.1S-06,(2020).
- [6] 熊本忠彦,田中克己.Web ニュース記事からの喜怒哀楽抽出.情報処理学会研究報告,pp.15-20.NL-165, (2005).
- [7] 大澤泰我,ケネス マッキン,永井保夫.文脈から感情を分析するための感情判断器の検討.第 19 回情報科学技術フォーラム.F-026,(2020).
- [8] 大西智佳,奥村紀之.感情判断のための顔文字の感情極性辞書の構築.情報処理学会第76回全国大会,4P-2 (2014)
- [9] 熊本忠彦,田中克己.Web ニュース記事からの喜怒哀楽抽出.情報処理学会研究報告,pp.15-20.NL-165, (2005).
- [10] 顔文字辞典 - Simeji.<https://simeji.me/blog/顔文字-一覧/kaomoji>

