

# T5 を用いた感情要約文の生成

福田芳樹<sup>1</sup> 伊東嗣功<sup>2</sup> 石井雅樹<sup>2</sup> 堂坂浩二<sup>2</sup>

<sup>1</sup> 秋田県立大学 大学院システム科学技術研究科 電子情報システム学専攻

<sup>2</sup> 秋田県立大学 システム科学技術学部 情報工学科

{m22b009,dohsaka,ishii,hidekatsu\_ito}@akita-pu.ac.jp

## 概要

人間同士のコミュニケーションの活性化を狙いとして、会話ロボットが文章の書き手の感情を推定するとともに、推定結果の判断理由を生成する手法に関する研究を行っている。本研究では、判断理由として書き手の感情が生起した要因に着目し、感情的な出来事について書かれた文章から書き手の感情の生起要因を含む要約文（感情要約文）を生成する手法を提案する。T5 を用いた提案手法の方が、ベースラインにおける SBERT を用いた抽出型要約と比較して高い ROUGE スコアを示し、人間の作成した要約文とほぼ変わらない文を生成した。さらに、T5 は感情推定を同時に行う場合においても同様の結果を示し、入力文章に感情タグを入れることで生成文の感情表現を制御できることを示した。

## 1 はじめに

人間同士のコミュニケーションを活性化する会話ロボットの研究が行われており [1, 2]、我々は人間同士の感情コミュニケーションでは感情のやり取りが重要な役割を担う点 [3, 4, 5] に着目し、文章と絵を用いて人間同士の感情コミュニケーションを活性化する会話ロボットの研究を進めている [6, 7, 8]。会話ロボットが共感や発話によって人間同士の感情コミュニケーションを促進できれば、心理セラピー効果が得られる [3] だけでなく、良好な対人関係の構築に役立ち [4]、社会交流の活性化につながるものが期待される。また、社会交流の活性化は認知症の予防にも効果的である [5]。

会話ロボットが絵から感情を推定し応答する研究では、感情推定の結果だけでなく、ロボットがその感情であると判断した理由も一緒に応答することで、人間同士のコミュニケーションがさらに活性化することが示唆された [7]。一方で、人間は、書き手の感情を推定する根拠として文章の内容を用いるこ

とが自然である。このため、会話ロボットが絵だけでなく文章も判断理由に用いることができれば、応答の幅が広がり、妥当性や信頼性の向上によって、人間同士のコミュニケーションをより活性化することができると考えられる。

そこで、我々は会話ロボットの感情判断理由として書き手の感情が生起する要因に着目し、感情的な出来事について書かれた文章から書き手の感情生起要因を含む要約文（以後、感情要約文）を生成する手法を提案する。具体的には、「昨日、とても嬉しいことがありました。当選確率の低いライブに当選しました！！最高にうれしい！！速攻で有給の申請を入れました。」のように感情的に書かれた文章から、「当選確率の低いライブに当選したから嬉しい」のような要約文を生成する。

会話ロボットと感情生起要因に関連する研究に、感情生起要因を利用した感情推定 [9, 10] がある。本研究は基本的に入力文章の感情は推定済みとして、入力感情の生起要因に着目した要約文を生成する点に特徴がある。

以下において、2 節で提案手法とデータセットの構築方法を述べ、3 節で評価の方法と結果について示す。最後に、4 節で結果についてまとめる。

## 2 提案手法

### 2.1 諸定義

本研究における感情、感情表現、感情要約文に関連研究 [9, 10, 8] を参考に次のように定義する。

**感情** Plutchik の基本 8 感情 [11] の {喜び, 悲しみ, 怒り, 恐れ, 嫌悪, 信頼, 予期, 驚き} とする。

**感情表現** 中村の感情表現辞書 [12] をもとに作成された感情強度辞書 [13] のうち、幸村の定義 [14] を満たす述語とする。

**感情要約文** 感情生起要因は「当選確率の低いライ

喜びの文章 1			
文番号	文	要約に用いた出来事	主な出来事
1	体重が減ってる！	☑	●
2	結婚式まであと1か月！！	☐	○
3	この調子で目標体重まで痩せるぞ！！	☐	○
4	心も体もなんだか調子が良くて最高！！	☐	○
書き手が感情を抱いた出来事の要約		接続表現	感情表現
(書き手は)	体重が減った	から	嬉しい

図1 喜びのアノテーション済み文章例

に当選したので「嬉しい」のように感情表現に原因・理由の副詞節を作る接続表現を伴って従属接続できるものとする。この文型を感情生起要因文型と呼び、この文型を満たす要約文を感情要約文と呼ぶ。

## 2.2 提案手法の概要

近年、様々な自然言語処理タスクを Text-To-Text で解くトランスフォーマーモデル T5 が発表され、当時の GLUE や SuperGLUE などの様々なタスクで SoTA を達成した [15]。本研究では T5 を用いて文章の感情要約文を生成する。

## 2.3 データセットの作成

学習・評価に用いるためのデータセットを構築する。事前に書き手の感情が推定済みの文章を収集した。まず、複数人のクラウドワーカーがあるテーマと感情をもとに 4, 5 文からなる文章を作成し、次に、1つの文章につき 5 人の独立したクラウドワーカーが書き手の感情を推定した。本研究では、これらのうち 3 人以上の感情が一致した文章を使用し、多数派感情を文章の書き手の感情とする。このように作成した 1212 件の各文章に対して、下記に示す作業を行う。また、1つの文章に 3 人分の作業データを付けるものとし、1 人分の作業例を図 1 に示す。

**作業 1** 書き手の感情が生起した要因となる出来事を感情生起要因文型に沿って簡潔に要約する。

**作業 2** 作業 1 に用いた出来事（感情生起要因）の含む文をすべて選択する。

**作業 3** 作業 2 で選択した文のうち、最も主な出来事（感情生起要因）を含む文を 1 つ選択する。

このデータセットを train : dev : test = 8 : 1 : 1 に分割して学習・評価データとする。

表 1 T5 の学習データ例

input	<喜び>体重が減ってる！結婚式まであと1か月！！この調子で目標体重まで痩せるぞ！！心も体も何だか調子が良くて最高！！
target	体重が減ったから嬉しい。

## 2.4 T5 のファインチューニング

事前学習済み T5 はウェブ上に公開されているモデル<sup>1)</sup>を利用し、HuggingFace 社の Transformers[16] を用いてファインチューニングする。T5 は Text-To-Text のモデルであるため、学習に用いる input と target はテキストである。本研究では表 1 に示すように、input を接頭辞に感情タグ (例:<喜び>) を付けた感情的な出来事について書かれた文章とし、target を作業 1 で作成された文とする。これらを用いて、約 3 千件のデータを max\_input\_length=128, max\_target\_length=32, batch=32, epochs=10 の設定で学習する。

## 3 評価

### 3.1 評価方法

T5 を利用した提案手法における生成文を、次節で述べるベースラインにおける生成文と 2.3 節で作成した作業者の要約文を基準に評価する。評価指標には ROUGE-N-F[17] を利用し、評価データには学習に未使用の 126 文章を用いる。

### 3.2 ベースライン

我々はこれまでに感情と文章を入力することで 2.3 節の作業 3 を推定する手法を提案してきた [8]。具体的には、文章  $S$  とその書き手の感情  $e$  が与えられたときに、感情  $e$  が生起する原因となった主な出来事を含む文番号  $idx$  を推定する手法を提案した。このとき、式 (1) に示すように、文番号  $idx$  は入力各文  $S^i$  と事前に自動収集された感情生起要因文型を満たす事例全件  $EP_e^j$  の文ベクトルによって計算されるコサイン類似度行列の最大値によって決まる。

$$idx = \arg \max_i \left( \max_j \left( \text{CosSim}(S^i, EP_e^j) \right) \right) \quad (1)$$

しかし、正解率は低く、推定される文は感情生起要因文型を満たすとは限らないという 2 つの大きな問題点がある。これらの問題点を解決した手法をベースラインとする。

1) <https://huggingface.co/sonois/t5-base-japanese>

**正解率の向上** 正解率が低い原因は、自動収集した感情生起要因事例のノイズや別の意味の事例と高い類似度を持つことに起因する。ノイズに関しては、事例の自動収集時の形態素解析器を KNP[18]に変更してタグ<節-機能: 原因・理由>と係り受け情報を用いて感情生起要因文型を検知することで大幅に削減した。また、文の特徴量を改善するために SBERT[19] を利用する。SBERT は良質な文ベクトルを生成するために、BERT 族を効率的にファインチューニングするモデルである。本研究では事前学習済みモデルに東北大学の公開している BERT<sup>2)</sup> を、損失関数に TripletLoss(anchor 文に作業 1 の文, positive 文に作業 3 の文, negative 文に作業 2 以外の文)を用いて T5 と同様の学習データで学習した (batch=32, epochs=10)。

**要約生成ヒューリスティクス** 2 つ目の問題点を解決するために、文を感情生起要因文型を満たす形に変形する手法を提案する。式 (1) で推定された文 (以後、選択文) に対して、次のヒューリスティクスを適用し感情要約文を作成する。

1. 選択文中に感情生起要因文型がある場合は「【文中の感情生起要因】【文中の接続表現】【文中の感情表現】」を出力。
2. 上記以外で入力文章中に感情表現が含まれる場合は「【選択文の文末過去形】【基本接続表現】【文章中の感情表現】」を出力。
3. 上記以外の場合「【選択文の文末過去形】【基本接続表現】【基本感情表現】」を出力。

ここで、基本表現は 2.3 節で構築したデータセットで最も使用された接続表現「ので」と感情表現「(喜び) 嬉しい, (悲しみ) 悲しい, (怒り) 腹が立つ, (恐れ) 不安だ, (信頼) 信頼している, (嫌悪) 嫌だ, (驚き) 驚いた, (予期) 興味がある」を利用する。

### 3.3 結果と考察

#### 3.3.1 ベースラインの事前実験

表 2 に先行研究を改良したベースラインの作業 3 の推定結果を示す。ここで、 $E[X]$  は作業 3 を正解する期待値, normal はノイズ除去済みの事例データベース約 3.5 万件<sup>3)</sup>を用いたシステムの結果, ideal は作業 1 の結果を normal で用いた事例データベー

2) <https://huggingface.co/cl-tohoku/bert-base-japanese-whole-word-masking>

3) 対象コーパスの一部に不満調査データセット [20] を含む

表 2 主な感情生起要因を含む文の推定正解率

model	normal	ideal	$E[X]$
BERT	0.40	0.59	0.42
SBERT	0.48	<b>0.81</b>	

表 3 生成要約文と参照文の ROUGE スコアの比較

model	ROUGE-N-f		
	N=1	N=2	N=L
users	0.53	0.31	0.47
baseline -predicted task3	0.36	0.17	0.31
-correct task3	0.46	0.27	0.41
T5 -no emotion tag	<b>0.53</b>	0.33	0.47
-add emotion tag	<b>0.53</b>	<b>0.34</b>	<b>0.49</b>

スに追加した結果である。normal の結果を見ると、SBERT を用いたシステムは先行研究と同じ状況である BERT を用いたシステムよりほんの少し向上しているが、依然として期待値とほぼ同じ値である。一方で ideal の結果を見ると、SBERT を用いたシステムの正解率は非常に高いことが分かる。このことから、コサイン類似度と感情生起要因事例データベースを用いた主な感情生起要因の含む文の推定では、事例データベース内に類似する事例がある場合に文の特徴量化手法として SBERT を用いることは有効であると言える。今後、特に記載のない限り SBERT-normal の結果を用いるものとする。

#### 3.3.2 要約文の生成結果と考察

表 3 に各手法における生成文の ROUGE スコアを示す。なお、users は作業者間のスコア平均である。具体的には、ある 1 人の作業者の要約文を生成文、他の 2 人の要約文を参照文として、すべての組み合わせにおいて ROUGE スコアを計算したときのスコアの平均値である。このスコアは作業者間の要約文の単語一致度を示し、システムのスコアがこの値に近ければ近いほどシステムは作業者の作成した要約文と同じ単語を生成していると言える。

**ベースラインの結果の考察** baseline-predicted task3 は先行研究を改善した手法の結果 (表 2 の SBERT-normal の結果に対応) であり、-correct task3 は主な感情生起要因の推定は行わずに作業者の作業 3 の結果をそのまま 3.2 節で定義した要約生成ヒューリスティクスに入力した場合である。

表 3 を見ると、主な感情生起要因の推定の正解率が低い状態ではスコアが低く、仮に主な感情生起要因の推定で完全に正答した場合 (-correct task3 の結

表4 入力文章と各システムの生成例

model	ROUGE-L-f	生成文 (参照文)
恐れのある文章		
		20年間勤めた会社で、大規模なリストラが行われている。年齢的に、自分も対象となる可能性があり心配している。同期からも上司に呼び出されて、通知を受けたという話を聞いた。毎日、不安を抱えながら仕事をしている。
users	0.73	20年間勤めた会社で大規模なリストラが行われ、年齢的に自分も対象になる可能性があるから不安だ。
baseline-predicted task3	0.36	年齢的に、自分も対象となる可能性があり心配していたので心配する。
T5-add emotion tag	0.81	20年間勤めた会社で大規模なリストラが行われ、自分も対象となる可能性があるため不安だ。

果)でも T5 や作業員間のスコアには届かないことが示された。

具体的な結果を考察したところ、作業員たちが作業3で選択した文をそのまま要約文に使用している場合にスコアが高くなり、作業2で選択した文に含まれる単語を利用した場合や、同義語や類義語で置換する場合にスコアが低くなる傾向があった。表4のbaselineに生成結果の一例を示す。表4に示すように式(1)によって選択された文の文末に感情表現が入っている場合、生成文の感情表現が重複することがあった。

**T5の結果の考察** T5-add emotion tagは表1のように感情タグを文頭に付けた場合であり、-no emotion tagはそれが無い場合である。

表3を見ると、どちらの場合においても、作業員間のスコアとほぼ変わらない。このことからT5は作業員間の作成した要約文とほぼ同じ単語を用いて要約文を生成していると言える。

具体的な結果を考察したところ、-add emotion tagでは意味の通らない文はほとんどなかった。表3のT5のROUGEスコアが作業員間のスコアよりも高くなるのは、表4に示すように、T5の生成文がある作業員の参照文とほぼ同じ文を生成していることに起因する。また、表3の-no emotion tagが-add emotion tagよりも若干低い結果になっているのは、入力文章の感情とは異なる感情表現を生成しているためであった。このため、T5-no emotion tagの生成文には、文法的には正しいが意味は通らない文が稀であった。

表5 T5の生成文を用いた感情推定の正解率

タグ	一致人数		
	3	4	5
あり	0.83	0.85	0.88
なし	0.48	0.45	0.90

そこで、不適切な感情表現を生成する現象に着目し、T5の生成文に対して感情表現辞書を用いた単純な感情推定を行った。このとき、表5に感情推定の正解率を文章の多数派感情の一致人数別にまとめたものを示す。表を見ると、感情タグがない場合は一致人数が少なくなるにつれて正解率が下がっていることが分かる。一方で、感情タグがある場合は一致人数にかかわらず正解率は高い。このことから、人間が見ても書き手の感情が分かりづらい場合はT5も感情を当てにくいことが分かった。また、感情タグを入れることで出力文の感情を制御することが可能であることも示唆された。

## 4 おわりに

会話ロボットの感情の判断理由として文章中の感情生起要因に着目し、文章の感情生起要因を含む要約文(感情要約文)を生成する手法を提案した。具体的には、感情的な出来事について書かれた文章を用いて感情要約文を生成するようにT5をファインチューニングする方法とそのためのデータセットの構築方法を提案した。ベースラインとして、SBERTを用いて入力文章の中から主な感情生起要因の含む文を選択し、ヒューリスティクスによって感情要約文を作成する手法を定義した。提案手法と比較したところ、ベースラインは主な感情生起要因の含む文の選択精度がROUGEスコアに大きく関わるを示した。また、ヒューリスティクスに正解文を入力した場合でもT5のROUGEスコアには届かないことを示した。T5では学習データが少ないにもかかわらず、ROUGEスコアが高い自然な文を生成した。一方で、入力文章に感情タグを入れない場合、書き手の感情が人間でも理解しにくい文章はT5でも適切な感情表現を生成できない傾向があった。しかし、感情タグを入れれば適切な感情表現を生成できることが分かった。

今後の課題として、提案手法に基づいて判断理由を応答する会話ロボットを構築し、実際に人間同士のコミュニケーションを促進できるか実証実験を行う必要があると考えられる。

## 謝辞

本研究は科研費（16K00355）の助成を受けたものである。また、株式会社 Insight Tech が国立情報学研究所の協力により研究目的で提供している「不満調査データセット」を利用した。

## 参考文献

- [1] Kohji Dohsaka, Ryota Asai, Ryuichiro Higashinaka, Yasuhiro Minami, and Eisaku Maeda. Effects of conversational agents on activation of communication in thought-evoking multi-party dialogues. **IEICE Transactions on Information and Systems**, Vol. E97.D, No. 8, pp. 2147–2156, 2014.
- [2] Shinya Fujie, Yoichi Matsuyama, Hikaru Taniyama, and Tetsunori Kobayashi. Conversation robot participating in and activating a group communication. In **INTERSPEECH**, pp. 264–267, 2009.
- [3] Jonathan R Carryer and Leslie S Greenberg. Optimal levels of emotional arousal in experiential therapy of depression. **Journal of consulting and clinical psychology**, Vol. 78, No. 2, pp. 190–199, 2010.
- [4] Ronald E Riggio. Assessment of basic social skills. **Journal of Personality and social Psychology**, Vol. 51, No. 3, pp. 649–660, 1986.
- [5] Laura Fratiglioni, Stephanie Paillard-Borg, and Bengt Winblad. An active and socially integrated lifestyle in late life might protect against dementia. **The Lancet Neurology**, Vol. 3, No. 6, pp. 343–353, 2004.
- [6] 木下圭, 堂坂浩二, 石井雅樹, 伊東嗣功. 文書と絵からの感情認識手法の開発. 言語処理学会 第 25 回年次大会, pp. 117–120, 2019.
- [7] 鈴木哲司, 堂坂浩二. コミュニケーション活性化を目的とした絵からの感情判断理由の生成. 秋田県立大学ウェブジャーナル B(研究成果部門), No. 5, pp. 24–31, 2018.
- [8] 福田芳樹, 堂坂浩二, 石井雅樹, 伊東嗣功. BERT を利用した感情生起要因推定手法の検討. 人工知能学会 第 34 回全国大会論文集 (4Rin182), 2020.
- [9] 松本和幸, 三品賢一, 任福継, 黒岩真吾. 感情生起事象文型パターンに基づいた会話文からの感情推定手法. 自然言語処理, Vol. 14, No. 3, pp. 239–271, 2007.
- [10] 徳久良子, 乾健太郎, 松本裕治. Web から獲得した感情生起要因コーパスに基づく感情推定. 情報処理学会論文誌, Vol. 50, No. 4, pp. 1365–1374, apr 2009.
- [11] Robert Plutchik. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. **American scientist**, Vol. 89, No. 4, pp. 344–350, 2001.
- [12] 中村明. 感情表現辞典. 東京出版, 1993.
- [13] 赤堀立樹, 堂坂浩二, 伊東嗣功, 石井雅樹. 感情強度辞書を利用した日本語ツイートの感情強度の推定. 人工知能学会 第 33 回全国大会論文集 (3Rin214), 2019.
- [14] 幸村幸夫. 日本語のシンタクスと意味. くろしお出版, 1982.
- [15] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. **Journal of Machine Learning Research**, Vol. 21, No. 140, pp. 1–67, 2020.
- [16] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Transformers: State-of-the-art natural language processing. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations**, pp. 38–45, Online, October 2020. Association for Computational Linguistics.
- [17] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In **Text Summarization Branches Out**, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.
- [18] 笹野遼平, 黒橋禎夫. 大規模格フレームを用いた識別モデルに基づく日本語ゼロ照応解析. 情報処理学会論文誌, Vol. 52, No. 12, pp. 3328–3337, 2011.
- [19] Nils Reimers and Iryna Gurevych. Making monolingual sentence embeddings multilingual using knowledge distillation. In **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing**. Association for Computational Linguistics, 11 2020.
- [20] 三澤賢佑, 成田和弥, 田内真惟人, 中島正成, 黒橋禎夫. 定量調査のための意見調査コーパス構築への取り組み. 言語処理学会第 23 回年次大会 発表論文集, pp. 1014–1017, 2017.