

画像からユーモアを想起させる説明文章の生成

1 はじめに

私が、奈良先端科学技術大学院大学 (以下 NAIST) で取り組みたい研究テーマは「画像からユーモアを想起させる文章の生成」である。本稿では、この研究テーマの研究背景及び目的、先行研究、そして提案手法について述べる。

2 研究の背景、目的

近年 Apple の Siri や Softbank の Pepper など、ヒトと対話的なインタラクションを行う対話システムが数多く提案されている。日常生活で用いられるシステムやエンターテインメント性が要求されるシステムではユーザーを飽きさせない、対話継続欲求の高いデザインを採用する必要がある。宮澤らは人同士のインタラクションを参考に、毎回の対話において「次回も続けたい」と感じる要因を分析することで、対話継続欲求の高いシステムのデザイン方法の確立を目指してきた [7]。分析の結果、「対話において相手の発話行動を限定しないこと」、「相手の話を聞いている実感を与えること」、「人工物のことを生かした意外性の高いユーモアを使うこと」が有効であることが明らかになった。

以上の知見を参考に、藤倉らはユーモア応答を用いることでユーザーの対話継続欲求の向上を目指した [1]。しかし、ユーザーの言語情報に基づきユーモア発話を生成する試みは、ユーザーが発話を辞めた時点で閉ざされてしまう。ユーモアは言語情報のみならず、視覚情報などからも生成可能である。例えば、友人が誰かのモノマネをしている時に、「ツッコミ」を入れる場面が考えられる。視覚情報を下にユーモア発話を生成すれば、ユーザーの言語情報に限定されないユーモアが生成可能であると想定される。さらに視覚情報を用いてシステムがユーモアを自ずから生成することができれば、ユーザーが発話を行う必要がないため、ユーザーがシステムの利用をやめてしまってもユーモア生成の可能性が閉ざされない。システムによって生成されたユーモアに対して、ユーザーが面白さを感じることができれば、システムが再利用される可能性が残されるだろう。そこで本研究では、画像からユーモアを想起させる文章を生成することで、ユーザーの対話継続欲求を向上させることを目指す。図 1 に、システムがユーモア文章を生成する流れを示す。初めに、システムがユーザーなどの写真を撮影する。次にシステムが、撮影した画像から「立派なサボテンが微笑んでいますね！」などのユーモア文章を生成し、音声、またはテキストデータとしてユーザーに提示する。最後に、ユーザーが提示されたユーモアを受容する。ユーザーがユーモアを受容した際、面白さを感じることができれば、ユーザーの対話継続欲求が高まると想定される。

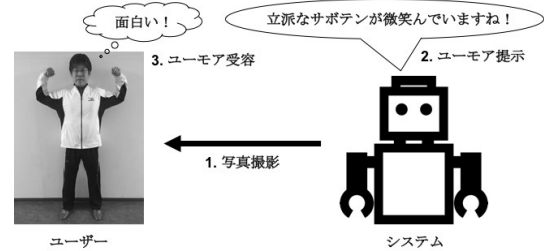


図 1: システムによるユーモア文章生成の流れ

3 先行研究

ユーモアに関わる認知的メカニズムを説明する理論として「不適合理論 [3]」がある。「不適合理論」には、ユーモアの生成因を不適合そのものに求める「不適合モデル [3]」や、その解決に求める「不適合-解決モデル [5]」がある。「不適合モデル」は、期待の心象と実際の刺激が異なることにより発生する不適合そのものによってユーモアが引き起こされると考えるモデルである。例えば、一発芸やナンセンスなジョークなどが「不適合モデル」に対して当てはまりが良いとされる。一方、「不適合-解決モデル」は、不適合に対して論理的な脈絡を発見することにより解決の過程が導かれ、その結果ユーモアが引き起こされると考えるモデルである。例えば、漫才の「ボケ」に対する「ツッコミ」はボケが生じさせた不適合を解決する手がかりを与える役割を持っている。「不適合-解決モデル」は解決という概念の導入により、不適合モデルに比べ、多くのユーモア現象を説明できる [8]。また、曖昧さ耐性の低い個人ほど、「不適合モデル」より「不適合-解決モデル」を好むという結果が明らかになっている [4]。

ユーモア現象は、単語間類似度の観点からも説明できる。Kao らは、Ambiguity と Distinctiveness の観点から駄洒落の面白さを説明しており、実験の結果から置換される対象の単語と置換後の単語が同じ文脈で発生しやすく、かつ置換される単語と置換後の単語間の関連性が小さいほどユーモアを受容性が向上すると示した [2]。従って、単語間の類似度は不適合の強さを表していると想定される。

4 提案手法

前節で述べたように、「不適合-解決モデル」は「不適合モデル」と比較すると、より多くのユーモア現象を説明することができ、かつより多くの人にユーモアとして受容されると想定される。そこで、本研究では、「不適

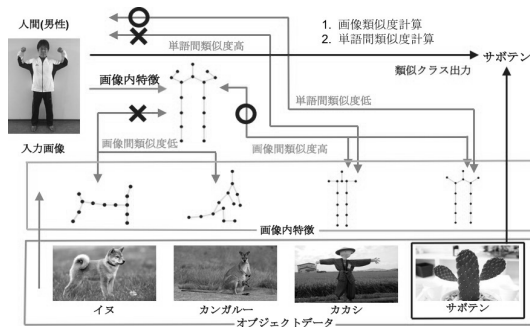


図 2: 類似度計算の過程

合-解決モデル」に基づき、画像からユーモア文章を生成することを目指す。単語間の類似度は不適合の強さを表していると想定されるため [2]、生成されるユーモア文章における単語と画像に写る対象を表す単語の関連度が小さいほどユーモアの受容性が向上すると想定される。しかし、単語間の類似度を考慮するだけでは、解決の過程が導かれないため、「不適合-解決モデル」を当てはめたユーモアとして受容されないだろう。そこで、画像間類似度を計算することで解決の過程を導く。画像に写る対象と生成される文章における単語を表す画像の画像間類似度が高ければ、両者が類似しているという観点から解決の過程が導かれると想定される。つまり、生成される単語と画像に写る物体の単語間類似度が低いことにより不適合が生じ、それらの画像間類似度が高いことにより解決の過程が導かれる。

本研究において提案する類似度計算の過程を図 2 に示す。初めに、画像を入力し、システムによって予め用意されたオブジェクトデータと画像間類似度を比較し、入力画像に写る物体と類似する物体を選択する。オブジェクトデータは、「イヌ」、「カンガルー」、「カカシ」、「サボテン」など様々な物体の画像と、その画像に対応する名詞のペアから構成される。画像間類似度が高いほど、解決の過程が容易になると想定されるため、入力画像とオブジェクトデータを比較した際、画像間類似度が高いと想定される「カカシ」、「サボテン」を選択する。次に、単語間類似度を計算する。単語間類似度が低いほど、不適合が強まると想定されるため、画像間類似度の計算過程で洗濯された「カカシ」、「サボテン」のうち、入力画像を表す名詞である「人間」との単語間類似度が低いと想定される「サボテン」を最終的に選択する。

画像間類似度を計算方法として、CNN(Convolutional Neural Network)を用いる。CNN は分類タスク用に学習済みのモデルを利用し、最終層手前の Feed Forward 層から取得される特徴ベクトルを用いて \cos 類似度によって類似度計算を行う。単語間類似度の計算方法についても、画像間類似度の計算と同様に、単語を特徴ベクトルに変換し、 \cos 類似度を用いることによって類似度計算を行う。

上述の手法を用いて文章内の単語を他のオブジェクトと置換することで、文章がユーモアとして受容されると想定される。

提案手法では、Neural Image Caption(NIC)[6]を用いて、上述の手法を取り入れたモデルの構築を目指す。NIC とは、画像に写る状況を説明文章を生成する手法である。NIC により生成された文章の主語と類似度計算

により導出された単語を置換した結果を、ユーモア文章として扱う。NIC により生成された文章に対して、単語の置換を行うことで、画像と文章のミスマッチにより、より不適合が強まると想定される。

5 まとめ

本稿では、私が NAIST で取り組みたい研究テーマである「画像データからユーモアを想起させる説明文章の生成」の研究概要について述べた。研究の流れとしては初めに、本研究で提案するユーモア文章の生成手法の検証及び必要に応じて改良を行う。次に、提案手法に基づき、システムを設計、開発する。最後に、システムによって生成されたユーモア文章がユーザーにユーモアとして受容されるかどうか検討するため、ユーモア評価実験を行い、実験結果からシステムの妥当性を検証する。本研究では、画像間類似度が大きく、単語間類似度が小さい単語を文章内に用いることで、ユーモアとして受容されると想定している。従って、ユーモア評価実験において、画像間類似度及び単語間類似度の変化によってユーモアの受容性がどのように変化するか検討する。

参考文献

- [1] Shohei Fujikura, Yoshito Ogawa, and Hideaki Kikuchi. Humor utterance generation for non-task-oriented dialogue systems. In *Proceedings of the 3rd International Conference on Human-Agent Interaction, HAI '15*, pp. 171–173. ACM, 2015.
- [2] Justine T. Kao, Roger Levy, and Noah D. Goodman. A computational model of linguistic humor in puns. *Cognitive Science*, 2015.
- [3] Goran Nerhardt. Humor and inclination to laugh: emotional reactions to stimuli of different divergence from a range of expectancy. *Scandinavian Journal of Psychology*, Vol. 11, No. 3, p. 185195, 1970.
- [4] W. Ruch and F Hehl. Intolerance of ambiguity as a factor in the appreciation of humour. *Personality and Individual Differences*, Vol. 4, No. 5, pp. 443–449, 1983.
- [5] Jerry M. Suls. A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis. In *The psychology of humor*, pp. 81–100. Academic Press Inc, 1972.
- [6] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *CoRR*, Vol. abs/1411.4555, 2014.
- [7] 宮澤幸希, 常世徹, 榊井祐介, 松尾智信, 菊池英明. 音声対話システムにおける継続欲求の高いインタラクションの要因. *電子情報通信学会論文誌*, Vol. J-95-A, No. 1, pp. 27–36, 2012.
- [8] 伊藤大幸. ユーモアの生起過程における論理的不適合および構造的適合の役割. *認知科学*, Vol. 17, No. 2, pp. 297–312, 2010.