

令和3年11月17日

報道機関 各位

国立大学法人 電気通信大学

オノマトペで質感を表現する人工知能を開発

【ポイント】

- * つるつる、さらさらといったオノマトペで質感を表現するAIを開発
- * 画像に写っているモノの質感を100人の被験者がオノマトペで表現
- * オノマトペにディープラーニングを適用した研究は世界でも珍しい

【概要】

電気通信大学大学院情報理工学研究科、および人工知能先端研究センターの坂本真樹教授らは、モノの質感をつるつる、さらさらといったオノマトペ（擬音語・擬態語の総称）^[1]で表現する人工知能（AI）を開発しました。

1,946枚の画像に写っているモノの質感を100人の被験者にオノマトペで表現してもらったデータを使って、深層学習（ディープラーニング）のモデルを作りました。1枚の画像に対して複数のオノマトペを正解データとして学習させることで、人によって質感の印象が異なるという曖昧さを考慮したモデルの作成に成功しました。

ディープラーニングをはじめとするAIは、一般に質感のような曖昧な学習は苦手とされており、オノマトペにディープラーニングを用いたという点でも世界でまれにみる研究成果といえます。

成果は国際学術誌「Frontiers in Psychology」に掲載されました。

【背景】

現在、人の神経細胞をモデルにしたニューラルネットワークをベースとする畳み込みニューラルネットワーク（Convolutional Neural Network:CNN）が、一般物体認識の分野で注目されています。その中でも、近年はより多層の構造を持つ深層畳み込みニューラルネットワーク（Deep Convolutional Neural Network:DCNN）が頻繁に用いられています。DCNNには画像の特徴量を学習の過程で自動で抽出できる利点があり、モノの質感を認識する際の課題である「着眼点人がにより異なる」ケースでも適用することができます。しかしながら、DCNNをはじめとするAIはそもそも質感のような曖昧な学習は苦手としており、学習の手法に工夫が必要でした。

そこで、オノマトペを使って質感の表現を試みました。オノマトペには、音韻の一音一音に印象が結びつき、共通の印象を与える「音象徴性^[2]」が強く現れる傾向があります。これを利用すると、印象を定量化できるため、オノマトペ1語でも多様かつ微細な表現ができるようになります。例えば、形容詞は表現の種類が膨大にあり、また副詞でも「とても」「かなり」といったように似たような意味を持つ表現が数多くありますが、オノマトペを使えばこれらを包含した表現が1語で可能になります。

【手法】

本研究では、DCNN を使ってオノマトペが持つ音象徴性に着目し、質感を表現するAI を作製することを目指しました。図1に示すように、入力を画像、出力を音韻とし、オノマトペによる正解データを学習させます。

実験では、素材画像の研究として一般的なフリッカー材料データベース (Flickr Material Database:FMD) 画像から切り出した、繊維、ガラス、金属、プラスチック、水、葉、革、紙、石、木の10個の 카테고리素材に関する1,946枚の画像と、これらの画像に対応づけた30,138語のオノマトペを実験データとして用いました。オノマトペは、画像に写っているモノの質感を100人の被験者が表現したものです。

この実験データの中から、実験に適したオノマトペとして全体の97.7%にあたる29,443語を選びました。選定したオノマトペは、「さーさー」「ふわふわ」といった「1・2モーラ型」のみで、ローマ字50音表記に沿った音韻だけを使用します(「ふぁ」「てい」などは除外する)。さらに、日本語として適切な表現だけを採用しています(「んめんめ」などは除外)。オノマトペは下表に示す73次元の音韻ベクトルで表現しています。

概要図



図1 学習モデルの概要

【さらさら】

	反復		1モーラ母音			1モーラ子音			1モーラ特殊語尾	2モーラ母音		2モーラ子音		2モーラ特殊語尾
番号	0	1	2	...	6	...	12	...	34~36	37~41	42~68	...	72	
要素	Re	noRe	a(1)	...	o(1)	...	s(1)	L(2)	
値	1	0	1	...	0	...	1	0	

図2 オノマトペの音韻ベクトルの例

学習モデルの構成は、1,000種類の物体認識のための16層DCNNモデルであるVGG16(図3)をベースとし、出力層をオノマトペの音韻ベクトルとしました。一般に、一つの画像に対して、質感を表現する正解オノマトペは一意に定まらず、被験者により自由回答されるオノマトペも多様な

りえます（図4）。この点を考慮し、各画像に対して複数回答されたオノマトペの音韻ベクトルの平均を正解音韻ベクトルとして学習させました。

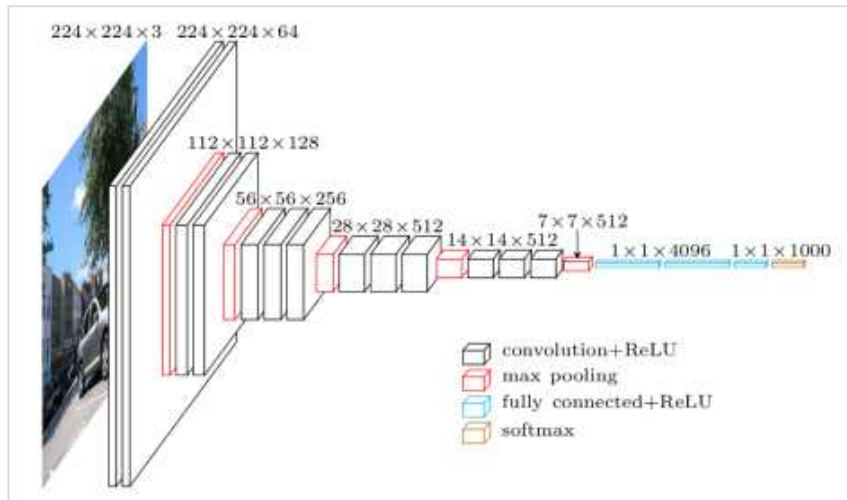


図3 ベースとした16層DCNNモデル「VGG16」

学習手法

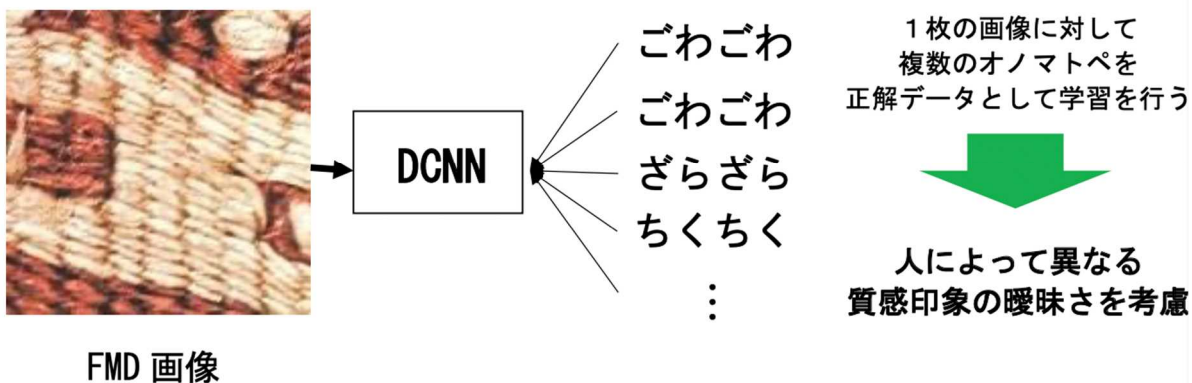


図4 質感の多様なオノマトペ表現を考慮した学習手法

【成果】

1枚の画像に対して複数のオノマトペを正解データとして学習させることで、人によって質感の印象が異なるという曖昧さを考慮したDCNNのモデルを作成することができました。このモデルによって出力された音象徴単語は、約80%の正解率を持つことを確認しています。

【今後の期待】

モノの質感をオノマトペによって表現するDCNNを使ったモデルを作りました。DCNNをはじめとするAIは、質感のような曖昧な学習は苦手とされており、オノマトペにDCNNを用いたという点でも画期的な成果といえます。

人間のように質感を表現できるコンピュータが実現すれば、人とロボットが共存するといわれる将来、例えば、ロボットが目の不自由な人に質感を教えるといったことが可能になると期待されます。

(論文情報)

雑誌名 : 「Frontiers in Psychology」

論文タイトル : Computer Vision System for Expressing Texture Using Sound-Symbolic Words

著者 : Koichi Yamagata, Jinhwan Kwon, Takuya Kawashima, Wataru Shimoda, Maki Sakamoto

DOI 番号 : 10.1109/CVPR.2015.7298970

(外部資金情報)

本研究は科学研究費補助金による支援を受けて行われました。(MEXT Grant Nos. JP23135510 and JP25135713 および JSPS KAKENHI Grant No. JP15H05922)。

(用語説明)

[1] オノマトペ : 擬音語・擬態語の総称

[2] 音象徴性 : 言語音が触覚や視覚など何らかの感覚的印象と結びつく現象

【連絡先】

<研究内容に関すること>

電気通信大学大学院 情報理工学研究科

教授 坂本真樹

E-Mail: maki.sakamoto@uec.ac.jp

<報道に関すること>

電気通信大学 総務企画課 広報係

Tel : 042-443-5019 Fax : 042-443-5887

E-Mail : kouhou-k@office.uec.ac.jp