写真とイラストから否定を認識する：人間と深層学習モデルの比較
Recognizing negation from photographs/illustrations: comparing humans and deep learning models

佐藤 有理[†]，峯島 宏次[‡]，植田 一博[†]
Yuri Sato, Koji Mineshima, Kazuhiro Ueda
[†] 東京大学，[‡] 慶應義塾大学
The University of Tokyo, Keio University
satoyuri0@gmail.com

概要

視覚表現は否定を描くことができるだろうか．この問題を，写真とコミックイラストの実世界視覚表現のデータ分析を用いて検討する．まず，画像キャプショニング課題を用いた実験により，従来の見解に反して，一部の視覚表現が否定を表現できることを示す．さらに，画像が否定を表現できる理由を分析するために，否定に関連する画像の分類課題を用いた実験を行い，機械学習（深層学習 CNN）と人間のパフォーマンスを比較する．その結果，人は画像には直接描かれていない背景知識や常識を利用して否定を認識することを議論する．

キーワード：否定，視覚表現，写真，イラスト， 実世界データ，機械学習

1. はじめに

否定表現は，人の思考やコミュニケーションにおいて重要な役割を果たしている．自然言語では「電車が来ている」から「電車が来ていない」のように否定文を作ることで，何らかの事態の否定を比較的容易に表現することができる．同様に，記号論理の言語やプログラミング言語では，否定は入力の真理値を反転させる真理関数の演算子（¬）として捉えられ，一定の役割を果たす．こうした言語表現における否定の意味や使用については，これまでに言語学や論理学に関連する分野 [6, 8, 9, 17] において研究が蓄積されてきた．

言語表現と比べて，写真やイラストなどの視覚表現において否定を表現するのはそれほど簡単ではない．例えば，「電車が来ない」ことを知らせるために，電車が写っていない駅のプラットホームの写真を送ったとしよう．受け手は送り手の意図通りに情報を理解できるだろうか．おそらくこの情報伝達はテキストによるものほど確実ではないだろう．写真や絵，地図のような非言語的表現によって否定を描くことはできないという見解は，前期ウィトゲンシュタイン [18] に典型的に見られるように，分析哲学において広く支持されている [7, 5].

本研究の目的は，「否定を表現するものとして人が認識できる視覚表現はあるのか」という問いを検討することである．視覚表現に関する実証研究としては，統制された評価実験を行うためにトップダウンで設計・作成されたタイプの表現を扱う図的推論や論理図の実験研究がある [14]．これに対し本研究は，科学の領域の外で自然に発生し，日常的な場面で人間の思考を表現し伝達するために使用されるタイプの視覚表現に焦点を当てる．すなわち，実際に人々に使われ，文化の中でデザインとして生き残っている実世界の視覚表現について収集と分析を行うデータ駆動型アプローチをとる．実世界の視覚表現のなかでも，写真画像とコミックイラストを題材とする [15, 16]．人が画像から読み取れることを説明するという画像キャプショニング課題を用いた実験により，視覚表現が否定を表現できるのか検証する．さらに，画像が否定を表現できる理由を分析するために，否定に関連する画像の分類課題を用いた実験を行い，機械学習（深層学習 CNN）のパフォーマンスと人間のパフォーマンスを比較する．最後に，否定認識における常識想起の役割について考察する．

2. 写真における否定

写真キャプション 機械学習用の写真画像データセット MS-COCO[11] とその 1 画像に 5 つキャプションを付与した日本語キャプション STAIR Captions [19] を使用した．否定表現「〜ない」を含むキャプションが 1 つ以上付された画像を選び，さらに否定に関連するかどうかのアノテーションとその理由説明（否定の対象）の両方が 3 名中 2 名以上一致した 65 枚を「否定画像」として抽出した．否定に関連しない「否定フリー画像」も同様に抽出した．

人間による写真分類 参加者（203 名，オンライン）に

は，写真画像が与えられ，それを否定を含む（否定を使うのが自然である・適切である）ものとそうでないものに分類するよう求めた．否定の典型例として「〜がない」「〜がいない」「〜でない」「〜できない」「〜が消えた」「〜が空っぽだ」「動かない」といった表現が教示された．まず（130 枚からランダムに抽出された）18 枚の画像分類課題を解き，正解・不正解が与えられた．それをもう一度繰り返したあと，新規のテスト課題が与えられた．人間の正答率は，否定画像の判定で 67.5%，否定フリー画像の判定で 73.6%だった．

機械による写真分類　否定画像と否定フリー画像を training 用，validation 用，test 用に分け，データ拡張のうえ使用した．VGG16 ファインチューニング付きの畳み込みニューラルネットワーク（CNN）モデルで学習を行った．この CNN モデル学習の結果，test の 20 枚の否定画像のうち 11 枚（55%）が否定と正しく分類された．否定フリーの画像の正答率も同じだった．

3. イラストにおける否定

写真画像のキャプション数の少なさ，（出版作品でないが故の）伝達意図の不明確さの改善を目指し，コミックイラストを題材としたデータ分析を行なった．

イラストキャプション　日本マンガデータセット Manga109 [1, 12] と合わせて，傑作マンガ作品（主要な漫画賞受賞者による主要な漫画賞受賞作品）をリストアップした．手塚治虫『火の鳥』など 131 作品を含めた．否定のアノテーションをつけ，3 名中 2 名以上一致したに 111 画像を「否定画像」として抽出した．「否定フリー画像」も同様に抽出した．459 名の参加者に対して，画像（言語部分は削除）から読み取れることを説明するように求め，結果として画像 1 枚につき平均 23.95 件のキャプションを得た．とくに 17 枚の画像については，コマ一つの状態かつ効果線などの特殊記号 [4] のない状態で，否定画像における否定句の出現頻度が否定フリー画像のそれよりも有意に多かった．これは純粋な視覚的要素のみで否定を表現することが可能である事例として捉えられる．

人間によるイラスト分類　上記の 17 枚と 10%有意の 1 枚を加えた計 18 枚の否定画像と同数の否定フリー画像を用いた．人間（205 名参加）の正答率は，否定画像の判定で平均 84.3%，否定フリー画像で 84.1%だった．

機械によるイラスト分類　18 枚の画像について，1 枚を test 用に外して，残りを training と validation 用にして CNN 学習を繰り返した．結果，18 枚の否定画像のうち平均 61.1%が正しく否定画像と分類された．また，否定フリー画像の正答率も同じく 61.1%だった．

4. 考察：否定認識と常識想起

「否定を表現するものとして人が認識できる視覚表現はあるのか」という問いに対して，これまでの画像キャプショニングの結果から「ある」と答えることができる．とくに，十分な数のキャプションを収集したコミックイラストについては，視覚的要素だけで否定を表現できるケースがあることをより確実に示すことができた．では，画像のどのような特徴が否定認識を可能にしているのだろうか．図 1 は，人間の否定分類で正答率の高かった写真画像である．例えば，否定画像 (a_2) の正答率は 95%であり，キャプションやアノテーターの理由記述は「部屋に家具がない」というものだった．この記述の対象は，画像に直接描かれていないもの（家具）であり，それを復元するには，背景知識（通常，部屋には家具がある）の利用がうかがえる．また図 2（a_1）のように，正しく否定分類されたイラスト画像（正答率 100%）についても，「トイレには通常，紙がある」という背景知識の利用がうかがえる．このように，画像に直接描かれていない知識や常識を使って否定情報を認識していることが想像される．深層学習モデルにおける全体正答率は，写真画像とコミックイラストとも，モデルが否定画像と否定フリー画像を一般に分類できないということを示していた．今回の機械学習実験で使用した程度の訓練データの大きさでは，背景知識にあたるものまでは得ることができず，画像的特徴のみから否定を自動的に分類することは困難であったと考えられる．このことは，機械学習一般において背景知識や常識を扱うことが困難であるという見解 [3, 10] と整合的である．

これまで考察した否定認識と常識想起の関係について，否定と認識されやすい画像は，そうでない画像よりも，(画像内容と矛盾するところの) 常識が想起されやすいという仮説を検証する追加実験を実施した．否定分類と否定フリー分類で正答率の高かった上位 5 画像 (85%以上) を写真・イラストごとに用いた (計 20 画像)．画像の表す状況の適切な説明となるように，「ふつう＿＿＿＿＿のに，画像のような状況になっている」という空欄を埋めることを求めた．この空欄に対する記述が特に必要でないと考える場合には，「なし」と書くように指示した．例題として，ギザのスフィンクス像の画像が与えられ，解答例として，「ふつうは顔に鼻があるのに，画像のような「顔に鼻がない」状況になっている」という例が与えられた．参加者は 30 名で，オンライン実験として行われた．結果として，10 画像のうちの常識記述の出現回数は，否定画像で 7.97 回，否定フリー画像で 1.76 回だった．「な

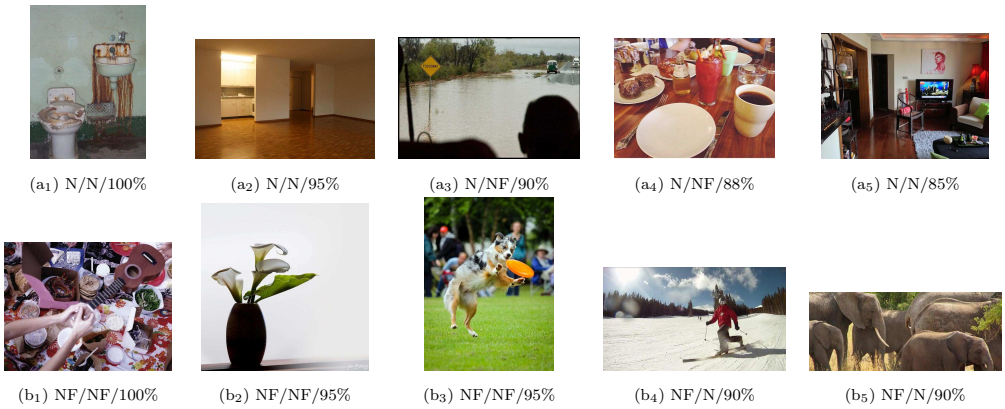

図 1　写真画像の分類結果，正答率の高い上位 5 つの事例．N/NF/90%” は，正解ラベルが「否定 (N)」，CNN による予測ラベルが「否定フリー (NF)」，人間の平均正答率が 90% ということを示す．

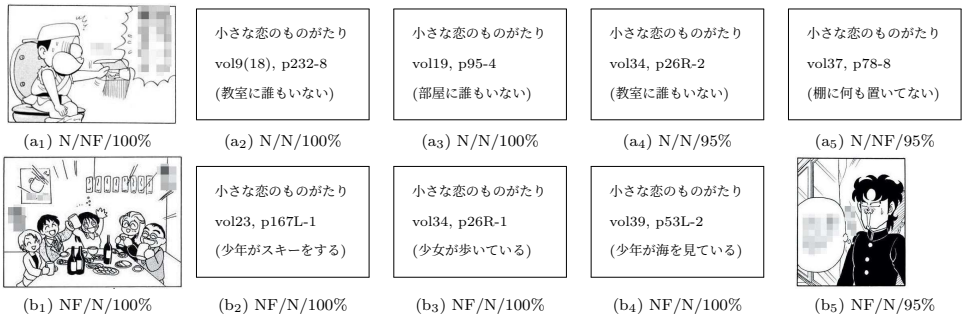

図 2　コミックイラスト画像の分類結果，正答率の高い上位 5 つの事例．利用許諾のある Manga109 データセット所収作品以外のコミックイラスト a_2-a_5，b_2-b_4 は，著作権の都合上作品の該当箇所情報のみ記載．

し」という回答については，否定画像で 0.86 回， 否定フリー画像で 4.28 回だった．否定画像と否定フリー画像の比較においていずれも有意差があった．この結果は，画像内情報だけでなく画像外の常識も利用することで，否定が認識されることを支持するものとして理解することができる．

今後，他の機械学習・深層学習モデル（例えば，アテンション機構を入れたもの）においてもさらに分析を行い，機械学習との対比によって人間の否定認識の特徴を明らかにしていきたいと考えている．また，否定以外にも，人の意図など，一般に視覚的に表現されにくいとされる情報が指摘されており [2]，それらの認識についても研究を進めていく計画である．

Appendix

否定画像として使用したコミック作品のリスト

あくはむ, 新居さとし, 講談社
天晴れ！カッポーレ, 菅野博之, 徳間書店
ありさ 2, 八神健, 角川書店
ぶらり鉄扇捕物帳, 佐佐木あつし, 集英社
デュアルジャスティス, 竹山祐右, 東京三世社

永遠のウィズ, みやうち沙矢, 講談社
はるかリフレイン, 伊藤伸平, 白泉社
ハイスクール！奇面組 vol20, 新沢基栄, 集英社
ラブひな vol14, 赤松健, 講談社
魔法使い養成専門マジックスター学院, 南澤久佳, 東京三世社
燃える！お兄さん vol19, 佐藤正, 集英社
むこうきずのチョンボ, みなもと太郎, 講談社
OL ランチ, さんりようこ, 小学館
ぴかる 元気です！, 栗城祥子, 小学館
しまっていこうぜ！ vol1, 吉森みき男, 秋田書店
タップ君の探偵室, ふくやまけいこ, 徳間書店
てんしのはねとアクマのシッポ, 霧賀ユキ, 東京三世社
徹さん, 川口憲吾, 東京三世社
東洋綺談, トオジョオミホ, 講談社
翼の記憶, 佐藤晴美, 朝日ソノラマ
うちの猫’ ず日記, がぁさん, 東京三世社
（以上，Manga109 より）
ブラックジャック vol05 06 12 13 14 22, 手塚治虫, 秋田書店
ブッダ vol01 03 12 13, 手塚治虫, 潮出版社
小さな恋のものがたり vol01 02 08 12 15 16 18 19 20 22 23 25 32 34 37 38 39 40 41 42, みつはしちかこ, 学研/立風書房
サイボーグ 009, 石ノ森章太郎, 秋田書店
火の鳥 vol02 05 08 09 手塚治虫, 朝日新聞出版
のたり松太郎 vol03 04 07 13 16 18, ちばてつや, 小学館
おれは鉄兵 vol01 10 12 14, ちばてつや, 講談社
テレビくん, 水木しげる, 中央公論新社
（以上，傑作リストより）

謝辞

本研究は JSPS 科研費 JP20K12782（代表：佐藤）の助成を受けたものです.

文献

[1] Aizawa, K., Fujimoto, A., Otsubo, A., Ogawa, T., Matsui, Y., Tsubota, K., & Ikuta, H. (2020). Building a manga dataset “Manga109” with annotations for multimedia applications. *IEEE MultiMedia*, *27*, 8–18.

[2] Alikhani, M., & Stone, M. (2019). “Caption” as a coherence relation: Evidence and implications. In *Proceedings of SiVL@NAACL 2019* (pp. 58-67), ACL.

[3] Bernardi, R., Cakici, R., Elliott, D., Erdem, A., Erdem, E., Ikizler-Cinbis, N., Keller, F., Muscat, A., & Plank, B. (2016). Automatic description generation from images: A survey of models, datasets, and evaluation measures. *Journal of Artificial Intelligence Research*, *55*, 409-442.

[4] Cohn, N. (2013). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images*. Bloomsbury.（中澤潤 (訳) (2020) マンガの認知科学：ビジュアル言語で読み解くその世界 北大路書房）

[5] Crane, T. (2009). Is perception a propositional attitude? *The Philosophical Quarterly*, *59*, 452–469.

[6] Déprez, V., & Espinal, M.T. (eds.) (2020). *The Oxford Handbook of Negation*. Oxford University Press.

[7] Heck, R. (2007). Are there different kinds of content?. In B.P. McLaughlin & J.D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind* (pp. 117–138). Blackwell.

[8] Horn, L. R. (1989). *A Natural History of Negation*. University of Chicago Press.（河上誓作 (監訳) (2018) 否定の博物誌 ひつじ書房）

[9] 加藤泰彦, 今仁生美, 吉村あき子 (編) (2010). 否定と言語理論　開拓社.

[10] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*, E253.

[11] Lin, T.Y, Maire, M., Belongie, S., Hays, J., & Perona, P. (2014). Microsoft COCO: Common objects in context. In *Proceedings of ECCV 2014, LNCS 8693* (pp.740–755). Springer.

[12] Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., & Aizawa, K. (2017). Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, *76*, 21811–21838.

[13] Sato, Y., & Mineshima, K. (2020). Depicting negative information in photographs, videos, and comics: a preliminary analysis. In *Proceedings of Diagrams 2020, LNAI vol.12169* (pp.485–489). Springer.

[14] 佐藤有理，峯島宏次 (2021). 論理の図形表現. 認知科学 *28*, 139-152.

[15] Sato, Y., Mineshima, K., & Ueda, K. (2021). Visual representation of negation: Real world data analysis on comic image design. To appear in *CogSci 2021*, 7 pages. Preprint at arXiv:2105.10131

[16] Sato, Y., & Mineshima, K. (2021). Can humans and machines classify photographs as depicting negation? To appear in *Diagrams 2021*, 4 pages. Springer.

[17] Wansing, H. (Ed.). (1996). *Negation: A Notion in Focus*. Walter de Gruyter.

[18] Wittgenstein, L. (1914/1984). *Notebooks 1914–1916.*, Anscombe, G.E.M. & von Wright, G.H. (Eds.). University of Chicago Press.

[19] Yoshikawa, Y., Shigeto, Y., & Takeuchi, A. (2017). STAIR captions: Constructing a large-scale Japanese image caption dataset. In *Proceedings of ACL 2017*, pp. 417–421.