画像キャプションの質的評価に向けた文の心像性推定手法の検討

梅村 和紀† カストナー マークアウレル† 井手 一郎† 川西 康友†
平山 高嗣† 道満 恵介‡ 出口 大輔† 村瀬 洋†
名古屋大学† 中京大学‡
{umemurak, kastnerm}@murase.is.i.nagoya-u.ac.jp
{ide, kawanishi, murase}@i.nagoya-u.ac.jp
{takatsugu.hirayama, ddeguchi}@nagoya-u.jp
kdoman@sist.chukyo-u.jp

1 はじめに

近年，コンピュータビジョン技術や自然言語処理技術の発達に伴い，画像キャプショニングに関する様々な研究 [1] が行われ，その技術は目ざましく発達している．これらの技術により自動生成されたキャプションは画像内の具体的事象の描写能力に長けており，画像内容に基づく検索のための索引付けとしては有効であるものの，例えばニュース記事中の画像のキャプションとしては必ずしも適切ではない．

このような技術的背景に基づき，我々は用途に応じた適切なキャプショニングを実現することを目指し，キャプションが元の画像をどれほど思い浮かべやすいものであるかに着目している．ここで，キャプションの評価指標としては Paivio ら [2] が定義した，心的イメージの喚起しやすさを表す単語属性である心像性（imageability）を採用する．心像性が高いキャプションから想起されるイメージは具体的で一意的なものであると考えられる．一方，心像性が低いキャプションから想起されるイメージは抽象的で多様なものであると考えられる．与えられたキャプションの心像性を定量的に評価できれば，用途に応じて適切な具体度／抽象度のキャプションを選択，生成できるようになると考えている．

そこで本研究では，キャプションの質的評価に向けて，文の心像性を推定することを目的とする．本発表ではその手法を提案し，有効性を検証する．

以降，2 節で関連研究を紹介し，3 節で文の心像性の推定手法を提案する．次に 4 節では，推定した心像性の評価実験とニュース記事のキャプションの分析を行なう．最後に 5 節で本発表をむすび，今後の課題について述べる．

2 関連研究

Mathews ら [3] は，画像を的確に描写した上で，“positive” や “negative” といった印象をふまえたキャプショニングを行なった．Gan ら [4] は “humorous” や “romantic” といった人間の感情を含めた魅力的なキャプショニングを行なった．

Paivio ら [2] は，心理言語学において単語の心像性（imageability）を定義した. 具体的には,「単語の意味に対応する種々の感覚イメージの想い浮かべ易さを表す主観的評定値」と定義されている [5].

田中ら [6] は，Web 文書の具体度の評価と，具体的な Web ページを検索するためのクエリ推薦の手法を提案した．彼らは，単語の心像性と具象性に着目し，それらの平均値として単語の具体度を定義した．段落の具体度は含まれる単語の具体度の平均値とし，文書の具体度は段落の具体度の最大値とした．この手法では，具体度を算出する際に単語間の関係性を考慮していないため，修飾された際に増加すべき語の具体度が増加しないことがあるという問題点がある．

3 文の心像性推定手法

本節では，文の心像性の推定手法を提案する．処理手順として，文を構文解析し，構文木の構造に基づいて文の心像性スコアを算出する．なお，以下の説明及び実験では英文を想定する．

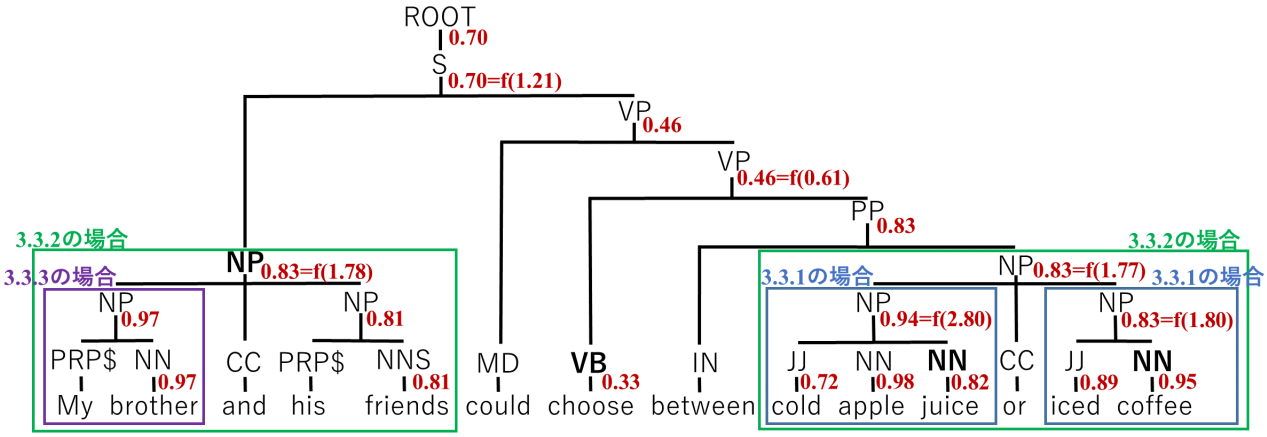

図 1: 文の心像性スコアの算出例（太字のノードはメインノード，赤字は各ノードで算出されたスコア）

3.1 単語の心像性

まず，文中の単語の心像性の検索について述べる．ここでは，Reilly ら [7] と Cortese ら [8] が作成した心像性辞書と MRC Psycholinguistic Database（MRCDB）[9] のうち動詞に関する心像性辞書を組み合わせて作成した辞書を利用する．この辞書には，重複しない 5,544 語の英単語に対して心像性の値が付与されている．これらは被験者実験により定められた，Likert 尺度に基づく [100,700] の値をとる．ここでは処理の都合上，これらの値を [0,1] に正規化し，単語の心像性スコアとする．

一般に，文中の単語は文法規則に従い活用している．そのため，各単語を原形に戻してから辞書を検索する．具体的には，NLTK[1] を用いて，小文字化，ステミング，レンマ化の処理を順に施し，その各段階で辞書を検索し，最初にヒットした時の心像性スコアを単語の心像性スコアとする．これにより，少しでも文中に存在する語形に近いスコアを利用する．

また，例外として，数詞と序数詞のスコアは最大値 1 とした．また，助動詞，冠詞，記号など，単体では意味をなさない単語は，ストップワードとしてスコアを与えず，その後の処理でも無視することとした．ただし，構文解析時にはこれらのストップワードも含めて処理をする．なお，これらの単語を除いた心像性辞書に存在しない語をスコア未知語と呼ぶ．

[1] http://www.nltk.org/

3.2 構文解析

次に，構文解析方法とその結果得られた構文木について述べる．ここでは，構文解析器として，Stanford CoreNLP [10] を利用した．得られた構文木は多分木であり，その葉ノードは全て，形態素解析結果の形態素をもつ．葉ノード以外の全てのノードは，品詞名や構成素名をもち，それらをラベルと呼ぶ．また，根ノードから同じ深さにあるノード同士が共通の親ノードをもつ場合，それらのノードを兄弟ノードと呼ぶ．以降，生成された構文木は全て正しいものとして利用する．

3.3 文の心像性の算出

最後に，生成された構文木に基づいて，文の心像性スコアを算出する方法について述べる．本手法では，構文木の葉ノードからボトムアップに心像性スコアを組み合わせて，文全体の心像性を算出する．最後に [0,1] に正規化されたスコアを元の [100,700] に戻し，文の心像性とする．なお，スコア未知語については無視する．

心像性スコア算出の例を図 1 に示す．各ノードに付与された数値は，提案手法により算出されたスコアである．

3.3.1 修飾関係にある兄弟ノードの場合

まず，兄弟ノード同士が修飾関係にある場合は，修飾語より被修飾語のスコアが変化すると考えられる．

表 1: 手法・設問パターン毎の正解率

パターン	1	2	3	合計
提案手法	100 % (5/5)	100 % (5/5)	80 % (4/5)	93 % (14/15)
比較手法 1：平均値	0 % (0/5)	100 % (5/5)	80 % (4/5)	60 % (9/15)
比較手法 2：最大値	− (0/0)	100 % (2/2)	80 % (4/5)	86 % (6/ 7)

そのため，被修飾語をメインノードと呼び，修飾語の心像性によりメインノードの心像性を増加させる．

メインノードの決定方法は 3 通りある．1 つ目は，名詞の兄弟ノードが 1 つ存在する場合に，そのノードをメインノードとする．2 つ目は，名詞の兄弟ノードが複数存在する場合に，最後の名詞のノードをメインノードとする．3 つ目は，名詞の兄弟ノードが 1 つも存在しない場合，最初のノードをメインノードとする．以上のように，修飾関係にある兄弟ノードの間でスコアを算出し，親ノードの心像性スコアとする．このとき，心像性スコア I を式 1 により求める．x_i $(i = 1, \ldots, n \,|\, i \neq m)$ は兄弟ノードの心像性スコア，x_m はメインノードの心像性スコアとする．ここで，n はメインノードを含めた兄弟ノードの数である．式 1 では，メインノード以外の全ての兄弟ノードについて，その心像性スコアに 1 を足したものと，メインノードの心像性スコアの積を算出している．この値は 1 を超える可能性があるので，式 2 の関数を適用することで，[0,1] に正規化する．

$$I = f\Big(x_m \prod_{i=1(\neq m)}^{n} (x_i + 1)\Big) \quad (1)$$

$$f(x) = 1 - e^{-x} \quad (2)$$

3.3.2 並列な兄弟ノードの場合

兄弟ノードに “and” や “or” などの等位接続詞を含む場合，他の全てのノードのスコアの和を算出し，その値に式 2 の関数を適用した値を親ノードの心像性スコアとする．

3.3.3 兄弟ノードがない場合

兄弟ノードがない場合はそのノードのスコアをそのまま親ノードのスコアとする．

4 実験

4.1 推定した心像性の評価実験

4.1.1 実験概要

提案手法の妥当性を検証するために評価実験を行なった．心像性の定義を示した上で，本実験のために作成した英文対を提示し，被験者に「どちらの文の方が心像性が高いと思うか」という設問に回答させた．被験者は英語を母語としない，20 代の男女 12 名である．実験で用いた文はスコア未知語と固有名詞を含まないものとした．設問は，(1) 修飾語の有無，(2) 単語の入れ替え，(3) 複数の単語や節ごとの入れ替えの 3 パターンの設問を各 5 問，合計 15 問用意した．

比較手法として，文中に含まれる単語の心像性の平均値による推定と，最大値による推定を行なった．ここで，いずれの手法においても，対となる文の両者のスコアが等しい場合には，評価から除外した．各手法において，過半数の被験者が選択した文と各手法によるスコアが高かった文との一致数を正解数とした．

4.1.2 実験結果

各手法，各設問パターン毎の正解率を表 1 に示す．比較手法に比べ，提案した手法の正解率が上回っており，提案手法の有効性を確認した．

また，設問パターン 1 については，平均値による推定手法では全設問で不正解だったが，提案手法による推定では全設問で正解だった．その理由として，ある修飾語が加わった場合，一般に心像性は増加すると考えられるが，平均値による算出では減少することがあるためである．

また，最大値による推定手法では，その最大値が各文で等しいことがよくあり，その場合には文間でスコアを比較することができなかった．

表 2: データセット内の単語における検証結果

単語の種類	のべ語数	正味語数
スコア既知語	15,718 語	4,311 語
スコア未知語	7,101 語	2,898 語
ストップワード語	29,489 語	300 語
数詞	2,080 語	272 語
固有名詞	9,021 語	5,842 語
合計	63,409 語	13,623 語

4.2 ニュース記事のキャプションの分析

実際のニュース記事において，キャプション中の語のうち心像性辞書に含まれるものを調べた．分析には，雑誌記事や新聞記事等からなる Breaking News データセット [11] を利用した．

分析結果を表 2 に示す．スコア既知語とは心像性辞書に含まれている語を指し，スコア未知語とはストップワードと数詞を除いた，心像性辞書に含まれていない語を指す．固有名詞を含めて全体の約 6 割もの語が心像性辞書に含まれていないことが分かった．今後の課題として，これらの単語の心像性を推定する手法を検討する必要がある．

5 おわりに

本発表では，文の構造に基づいて，その心像性を推定する手法を提案した．被験者実験により，推定した心像性の妥当性を確認した．

今後の課題として，スコア未知語の心像性推定が挙げられる．

謝辞

本研究の一部は，科学研究費補助金による．

参考文献

[1] S. Bai and S. An. A survey on automatic image caption generation. *Neurocomputing*, Vol. 311, pp. 291–304, 2018.

[2] A. Paivio, J. C. Yuille, and S. A. Madigan. Concreteness, imagery, and meaningfulness values for 925 nouns. *J. Exp. Psycho.*, Vol. 76, No. 1, pp. 1–25, 1968.

[3] A. P. Mathews, L. Xie, and X. He. Senticap: Generating image descriptions with sentiments. In *Proc. 30th AAAI Conf. on Artificial Intelligence*, pp. 3574–3580, 2016.

[4] C. Gan, Z. Gan, X. He, J. Gao, and L. Deng. Stylenet: Generating attractive visual captions with styles. In *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3137–3146, 2017.

[5] 佐久間尚子, 伊集院睦雄, 伏見貴夫, 辰巳格, 田中正之, 天野成昭, 近藤公久. 単語心像性評価における表記の影響. 日本心理学会第 72 回大会発表論文集, p. 791, 2008.

[6] 田中伸弥, アダムヤトフト, 田中克己. Web ページの具体度の評価と具体的な Web ページのためのクエリ推薦. 第 4 回データ工学と情報マネジメントに関するフォーラム論文集, No. D7-4, pp. 1–8, 2012.

[7] J. Reilly and J. Kean. Formal distinctiveness of high-and low-imageability nouns: Analyses and theoretical implications. *Cogn. Sci.*, Vol. 31, No. 1, pp. 157–168, 2007.

[8] M. J. Cortese and A. Fugett. Imageability ratings for 3,000 monosyllabic words. *Behav. Res. Methods, Instrum., Comput.*, Vol. 36, No. 3, pp. 384–387, 2004.

[9] M. Wilson. MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behav. Res. Methods, Instrum., Comput.*, Vol. 20, No. 1, pp. 6–10, 1988.

[10] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proc. 52nd Annual Meeting of ACL: System Demonstrations*, pp. 55–60, 2014.

[11] A. Ramisa, F. Yan, F. Moreno-Noguer, and K. Mikolajczyk. The BreakingNews dataset. In *Proc. 6th Workshop on Vision and Language*, pp. 38–39, 2017.