

画像に対する発話を対象とした形容詞獲得手法

内田 ゆず 荒木 健治
 Yuzu Uchida Kenji Araki
 {yuzu,araki}@media.eng.hokudai.ac.jp

北海道大学大学院 情報科学研究科
 Graduate School of Information Science and Technology, Hokkaido University

1. はじめに

人間の幼児は語彙獲得において、名詞を他の品詞よりも早く、多く獲得することがわかっている [1,2]. したがって、幼児はある程度の語彙を獲得した後に、形容詞と名詞を組み合わせた発話を行うようになり、より複雑で豊かな表現が可能になると考えられる。

我々はこれまでに、幼児の言語獲得能力を手がかりに、語彙や文法の知識を持たない状態から名詞概念を獲得するシステム SINCA (System for Noun Concepts Acquisition from utterances for Image) の構築を行った [3]. SINCA は画像に関する内容の発話 (書き起こしテキスト) を入力として得ることで、名詞概念 (画像に対するラベル) を獲得することが可能である。

語彙獲得の次のステップとして、我々は本稿で名詞概念獲得の手法と同様に語彙や文法の知識を与えずに形容詞を獲得するアルゴリズムを提案する. 本システムの入力は実画像とその画像に対するユーザの発話であるので、画像処理の必要性を考慮して、獲得対象は色を表す形容詞に限定した。

本稿では、画像に対する発話からその画像に含まれる事物の色を表す形容詞を獲得する手法について述べる。

2. システムの概要

本システムの処理の流れを図 1 に示す. また、本章では、それぞれの処理について簡単に述べる。

2.1 入力

入力は画像と文の対である. 入力画像は Web カメラ (USB-CAMCHAT2/アイ・オー・データ機器. 有効画素数: 30 万画素) からキャプチャされた画像 (以降画像 P と呼ぶ), 入力文は画像 P を見せながらユーザが幼児に話かける発話 1 文 (以降文 S と呼ぶ) である. 入力画像は、ユーザが自由に被写体を選び撮影するものである. 入力文は全てひらがなで表記され、入力文に形態素解析などの前処理は一切施されない. ひらがな表記を用いてい

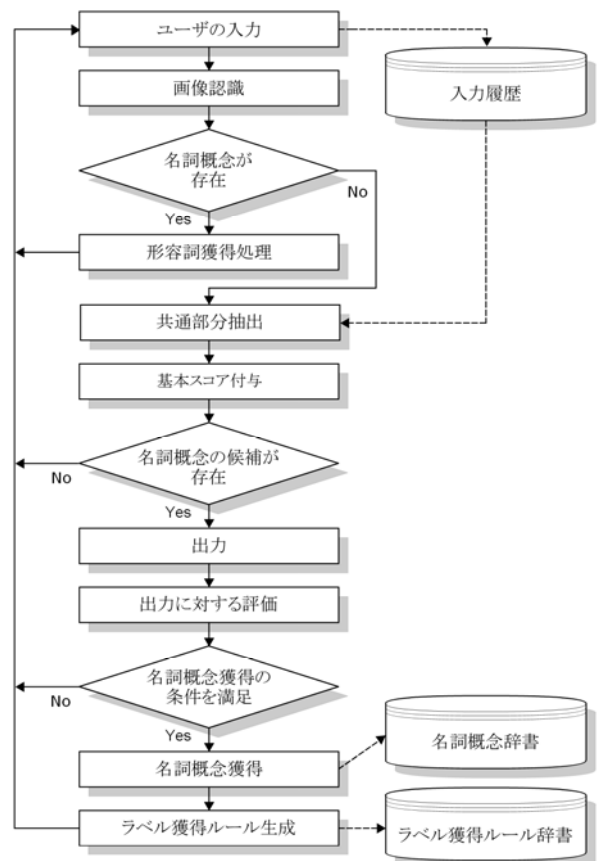


図 1 処理の流れ

るのは、ユーザによって表記に揺れが生じることと、入力された文字列自体に意味が含まれてしまうことを避けるためである. また、形態素解析などを行わないのは、幼児が正確な品詞分割などの能力を持っていないと考えるためである。

2.2 画像認識

過去に同じ被写体が写った画像が入力されたかどうかを判断する. ここでは、エボリューション・ロボティクス社¹の ERSP3.1 (Evolution Robotics Software Platform) に含まれる”ERSP ビジョン”を

¹ <http://www.evolution.com/products/ersp/>

用いた。ERSP 3.1は、ロボット製品の作成を目的とした総合開発プラットフォームで、照明や物体の位置が管理されていない現実的な環境の中でも、ロボットや装置が2次元と3次元の物体を認識することができる画像認識ツールである。

2.3 共通部分抽出

システムは入力を得ると、過去に画像 P とともに入力された文と文 S を比較して、字面が一致する文字列を切り出す。この切り出された文字列を共通部分と呼ぶ。これ以降の処理で共通部分は、画像 P に対応するラベルの候補として扱われる。

2.4 基本スコア付与

抽出された共通部分には基本スコアが付与される。基本スコアとは、その共通部分のラベルとしての確からしさを表した値であり、出現頻度が高く、文字数が多く、他の画像と共に出現することのない共通部分ほど高いスコアを与えられる。基本スコアの計算式は式(1)のようになる。

$$SCORE = \alpha \times \frac{F}{PN} \times \sqrt{L} \dots \dots (1)$$

式(1)で、 α は共通部分が他の画像とともに出現している場合スコアを減少させるようにはたらく係数、 F は共通部分が同一画像と共に出現した頻度、 PN は画像の出現回数、 L は共通部分の文字数である。対象の共通部分が他の画像と共に出現している場合、 α は 0.6 とし、そうでない場合は 1.0 とした。

2.5 出力

2.4 で述べた方法で求めた基本スコアが閾値 θ を超えた共通部分は、画像 P に含まれる事物のラベルに適している可能性が高いと判断され、テキストで出力される。

2.6 出力に対する評価

システムの出力に対してユーザは次の3つのキーワードのうち、最も相応しいものを選び、入力する。

- ・「じょうず」：ラベルとして適切である
- ・「おいしい」：ラベルとしては適切でないが意味はわかる
- ・「ちがうよ」：意味がわからない

幼児がこれらのキーワードを完全に理解するとは考えられないが、実際には、大人の表情や声の調子で感じ取ることのできる情報は多い。本手法ではこれらの情報の代わりにキーワードを用いることとする。

ユーザの反応によってその共通部分のスコアの再計算を行う。具体的には、基本スコアに係数 β を乗ずる。係数 β は、予備実験から、ユーザの評価

が「じょうず」の場合は 1.5、「おいしい」の場合は 0.8、「ちがうよ」の場合は 0.2 とした。

2.7 名詞概念獲得

「入力」から「ユーザの評価」の処理を繰り返した結果、再計算されたスコアが閾値 T を超え、さらに「じょうず」という評価を得たことがある共通部分は画像 P のラベルとして獲得される。

2.8 ラベル獲得ルール生成

ラベル獲得ルールとは、再帰的な名詞獲得を行うためのルールである。人間は過去に得た知識を活用し、より効率的に学習を進めていく。本手法ではそのような再帰的な学習を次のようにして実現している。

システムが文字列 S1 をある事物に関する正しいラベルとして獲得すると、その事物に関する過去の

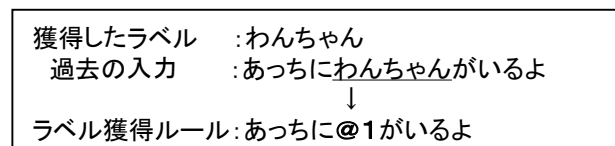


図 2 ラベル獲得ルールの生成例

入力文のうち、文字列 S1 を含む文から、ラベル獲得ルールを生成する。ラベル獲得ルールとは、図 2 のようにラベルの部分を変数とすることで、入力文を抽象化したものである。次に、生成したラベル獲得ルールに合致する入力文があった場合、変数部 @1 に相当する部分を切り出し、スコアを上昇させる。

これは、幼児は様々な表現を聞いているうちに、どのような表現がラベルを示すものなのかを学習して、より効率的に学習を進めていると考え、その様子をモデル化したものである。実際に、語彙爆発期の幼児は、一度大人が事物を指して言葉を発するのを聞いただけで、正しくその言葉を使うことができる(即時マッピング)ことが知られている[5]。ラベル獲得ルールを生成することで、本システムでもこれと似た現象を再現することができると考えられる。本システムでもこれと似た現象を再現することができると考えられる。

2.9 形容詞獲得処理

画像 P に対するラベルが既に獲得されている場合、形容詞獲得の処理を行う。

2.9.1 色の同定

画像 P に含まれる事物の色を判定する。まず、あらかじめ撮影しておいた背景と画像 P の間の背景差分をとり、差分全体を対象の事物と判断する。次に、得られた差分データに最も多く含まれる色を

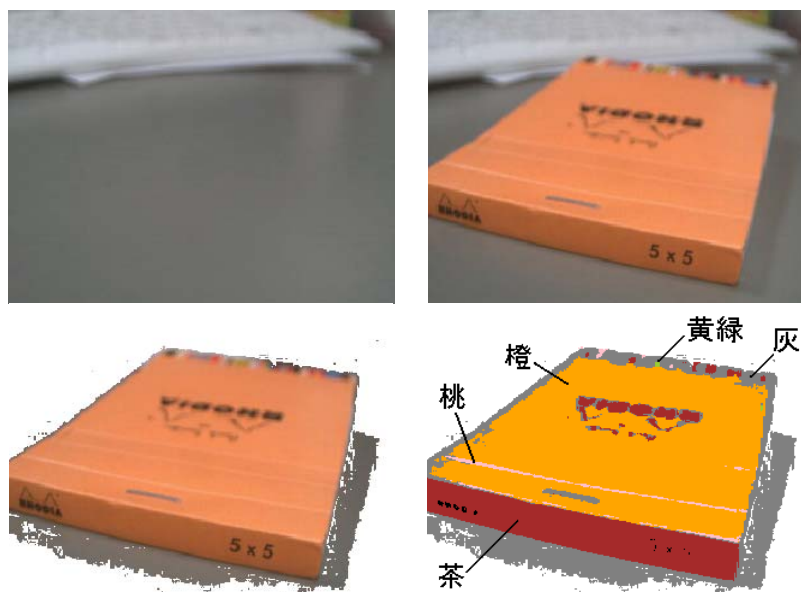


図3 色の同定の例
(上段左：背景 上段右：入力画像 下段左：背景差分 下段右：減色処理後)

その事物の色と定義する。しかし、得られた差分データの色情報は24ビットで表わされており、そのまま使用するのは困難である。そこで、各画素ごとに一般的に使われる13色(赤, 青, 黄, 緑, 黄緑, 茶, 橙, 青, 水色, 桃色, 灰, 黒, 白)とのユークリッド距離を計算し、最も距離が近い色に近似することで、減色処理を行った。得られた背景差分、減色後の画像の例を図3に示す。

2.9.2 形容詞獲得

画像Pに含まれる事物の色を同定したのち、過去に入力された同じ色を持つ画像とともに入力された文と入力文Sを比較する。双方の文に含まれるラベルは削除した状態で、名詞概念獲得のための共通部分抽出と同様に、字面が一致する部分を抽出し、形容詞の候補とする。

抽出された共通部分には式(2)で表わされるスコアを付与する。Fは共通部分が同一画像と共に出現した頻度、Lは共通部分の文字数である。

$$SCORE_{adj} = F \times \sqrt{L} \dots \dots (2)$$

名詞概念獲得とは異なり、形容詞はある画像に対して一意に決定されるものではないため、高スコアを得られたいくつかの共通部分を獲得すべきである。実験を重ね、閾値を決定することが今後の課題である。

4. 実験

提示された画像に対して「画像に含まれる事物を見せながら赤ちゃんに話しかける」ことを想定した発話を回答するアンケートを実施し、入力文を収

集する。この際、回答には事物の色に関する表現を含めるように指示する。このようにして収集した発話を本システムに入力し、色を表す形容詞の獲得が可能であるか評価する実験を行う予定である。

5. まとめ

名詞概念獲得システム SINCA と同様に、語彙や文法の知識を持たない状態から形容詞を獲得する手法を提案した。現状では獲得する対象を、色を表す形容詞に限定しているが、今後は幅広い形容詞を獲得可能なシステム構築を目指したい。

参考文献

- [1] 但馬香里: 幼児における一語発話の獲得について—1歳10ヶ月から2歳0ヶ月児の3人の幼児による観察報告, 東京工芸大学工学部紀要, Vol.27, No.2, pp.59-64, 1997.
- [2] WAXMAN, S.R., & BOOTH, A. E.: Seeing pink elephants: Fourteen-month-olds' interpretations of novel nouns and adjectives. Cognitive Psychology, 43, pp. 217-247.
- [3] 内田ゆず, 荒木健治: 画像に対する発話を対象とした名詞概念獲得システム SINCA, 知能と情報(日本知能情報ファジィ学会誌), Vol.20, No.5, pp.3-13, 2008.
- [4] 今井むつみ: ことばの学習のパラドックス, 共立出版, 1997.