

自然言語入力を利用した半自動絵生成手法の検討

A Semiautomatic Picture Drawing Method by means of User's Natural Language Inputs

上野 未貴*¹ 森 直樹*¹ 松本 啓之亮*¹
Miki Ueno Naoki Mori Keinosuke Matsumoto

*¹大阪府立大学, 工学研究科
College of Engineering, Osaka Prefecture University

Recently, a semiautomatic picture drawing method is required lots of research fields. We have already proposed conversation system sharing picture information with users called *Pictgent*, and defined the picture model as XML format. *Pictgent* also needs effective semiautomatic picture drawing method to change a picture as various user's inputs. However, it is difficult to put adequate tags to each part of picture efficiently. In this study, we proposed a semiautomatic picture drawing twitter bot called *Cotopaint* in order to obtain the users natural language inputs for the specific picture. The picture drawing process by *Cotopaint* is shown to confirm the effectiveness of proposed methods.

1. はじめに

近年、ヒューマンエージェントインタラクションや教育、エンターテインメントといった研究分野で自由に絵を利用したいという要望が高まっている。クリエイティブコモンズなどで一定条件を満たした上で利用できる部品となる絵は配布されているが、その組合せによって一枚の絵を作る過程についてはデザイナーに一任されており、タッチの差異などの問題からも、依然として研究での絵利用への敷居は高い。また筆者らは絵の情報をユーザと共有して対話が可能なシステムとして絵情報共有型エージェント (Picture Information Shared Conversation Agent: *Pictgent*) [1] を提案している。その中で、使用する絵と絵の中の状況を表す言葉を結びつけた絵モデルを提案しており、最終的には対話の過程でこの絵モデルを自動生成し、絵コーパスとして公開することを考えている。また、この *Pictgent* の一部の機能である、ユーザの入力した文から簡単な絵を生成する Twitter Bot である、ことべいんと (@cotopaint) をプロトタイプとして稼働させている。

言語を入力として絵を生成するにあたって、関連する研究としては、発話者の非言語情報の音量や音程を数値化して絵を描くツール [2] や、言語情報としてオノマトペの意図の解明を目指し母音や子音などの音節の構成要素に対して属性ベクトルを設定し入力とすることでロボットのモーションを指示するシステム [3] などが提案されている。また、パーツ分けされた絵への情報付与の研究として、ピクトグラムへのセマンティクスの導入手法 [4] が提案されている。しかし、上記の従来研究は、人間の絵や動作に対する連想や概念といった感性情報をどのように付与すれば、客観性がある情報付与が可能か、という問題点を抱えている。一方、*Pictgent* ではシナリオに応じて絵の修飾をするが、シナリオ以外でもユーザの入力を反映して絵を生成する機能がある。シナリオに沿ったパターンについては、作成者が予想して拡張することができるが、シナリオに沿わない場合、予想外の入力となり、それに対して絵と言語のパターンを用意するのは人的コストがかかりすぎるといった問題がある。絵やパーツに対してユーザの想起する内容や対話内容を見ることができれば、パーツに対して十分な語を関連付けること

で、この問題を解決できる。そこで、本研究では、この問題を解決するため、ソーシャルコミュニケーションサービスを利用して、ユーザの自然言語入力をもとにして、半自動的に絵を生成をし、投稿してユーザの反応を見ることにより、絵に対してユーザが想起する過程を把握し、パターンを増加させたり、適切なタグ付けをする手法について検討する。

第 2. 章で *Pictgent* について説明し、第 3. 章で絵モデルについて述べる。その後、4. 章で本手法の概要を示し、第 5. 章で拡張と応用について述べ、第 6. 章で今後の課題を述べてまとめる。

2. *Pictgent*



図 1: *Pictgent* の GUI の概観

Pictgent とは、絵を提示して、その内容についてユーザと対話をするシステムである。絵と情報を対応付けたものを絵モデルと呼ぶ。*Pictgent* には 2 つの会話モードとして、シナリオモードと雑談モードがありシナリオ以外の入力にも対応することができる。

この 2 つのモードを使い分けることにより、両者の利点を活かして対話すること、および絵を用いることにより、言語だけの対話では得られない絵に対する感性的な入力を得ることや、自然に状況を限定して対話を実現することが大きな特徴である。*Pictgent* の適用対象は限定しないが、低年齢層のユーザ

表 1: 絵パーツ DB の一部

partid	sceneid	imagenam	filepath
1	1	コート, 上着	coat1-1.png
1	2	コート, 上着	coat1-2.png

イメージ名はカンマ区切りで複数保持可能



図 2: 絵 1-ベース 図 3: 絵 1-修飾後 図 4: パーツ

でも楽しんで使用できるアプリケーションとして成立するように構築する。以下、Pictgent を構成する各モジュールの概要を示す。

絵モジュール 絵を示すモジュールである。本モジュールは絵の内容を表すモデルを XML 形式で持つ。

シナリオ進行モジュール ユーザの各種情報を内部状態として保持し、また、ユーザの入力に応じて適切なシナリオの遷移を管理するモジュールである。

会話モジュール ユーザの入力から使用するモードを決定して、文整形を施し実際の出力をするモジュールである。

Pictgent では、既に子供向け科学イベントで 192 名の被験者実験をしており、絵モデルを対話システムに導入することで児童の心理状態を確認することに関して統計的に一定の効果を確認済みである。図 1 に Pictgent の GUI を示す。

3. 絵モデル

本章では絵の内容を表す絵モデルについて示す。絵モデルはオブジェクト指向モデリングに基づき構成する。現在は、特に主要な登場人物のみをオブジェクトとして定義したが、シナリオに直接的には関わらない人物や無生物をオブジェクトとすることも可能である。絵モデルは、XML 形式で保存され、必要に応じてモデルの編集も可能とする。

3.1 絵モデルの概要

表 1 にパーツの DB の例を、図 2 ~ 図 4 にパーツを利用した絵の修飾の例を示す。

絵モデルを使用する Pictgent では、シナリオの遷移によって大きく絵が変化する場合と、対話内容によって絵内で小さな変化をする場合がある。絵モデルの中で、その絵において使用できるパーツを明示し、表 1 の DB と照合して、場面遷移時に絵の修飾をして一枚の絵を表示する。一方、場面内においてユーザの入力を反映して絵内で、図 4 内の 2-2 の帽子を被せるなどの小さな変化をする場合もある。本研究では主に、後者の絵内での小さな変化の手法に焦点をあてる。

3.2 問題点

Pictgent ではできるだけ多くのパーツについてユーザからの入力を受け入れることを考えている。しかしながら、表 1 に示すような、絵パーツと言葉との対応のパターンが少ない段階では、ユーザの自由入力の種類の多さに対応することができず、登録されていない言葉に対してどのパーツを提示すれば

よいかかわらないという問題がある。しかし、この問題を解決するために、作成者が絵モデルや絵パーツを増加させる際に、絵を拡張するために新たな絵パーツを用意したり既存パーツに対して言葉やタグを付けるという定性的な作業をする場合に、そもそもどのような言葉やタグを付加すれば良いかという問題点がある。

3.3 半自動絵生成による改善

上記問題点を解決するため、ソーシャルコミュニケーションサービス上で、絵を半自動生成してユーザに提示し、ユーザの自由入力を見ることにより、絵や絵の変化からどのような内容を想起するかという過程のデータを集めることにより、既存パーツと言語の増加の順序、および、既存パーツや絵モデルにどういった付与すべき情報を調べることにより、上記問題を解決するため、その要素技術を整備した。

4. 半自動絵生成手法の概要

本研究では、まず半自動絵生成手法について述べる。なお、ここでの半自動生成という言葉は、ベースの絵やパーツについてはあらかじめ人手で用意するが、その組合せについては、ユーザの自由入力をもとにして自動的に絵を生成するという意味である。

4.1 Twitter での絵生成の流れ

現在、ユーザの自然言語入力をもととして、絵を生成する機能を実装したものを Twitter Bot として稼働させている。本節では、その基本手法について述べる。まず、絵を描画するにあたって、現在は、ベースレイヤと、描画レイヤという 2 種のレイヤを重ねあわせている。ベースレイヤでは、白紙または、指定画像を読み込む。一方、描画レイヤではユーザの入力を用いて絵を描きかえる。以下にこの Twitter Bot による半自動絵生成手法の手順を述べる。

1. 起動。Twitter Bot から絵の描画開始を伝えるツイートをする。時刻 $t = 0$ とする。
2. ベースレイヤに外部テキストファイルに記述した画像パスを参照しベースとなる絵を読み込む。
3. 固定部 m + 変動部 c 間待機する。
4. 待機中に Twitter Bot に対してメンションされたツイートを時系列順に取得し、ユーザのツイート内容を入力としてコマンドを選択し、描画レイヤにコマンドを実行して描画する。
5. 各ユーザに実行したコマンドの内容をリプライする。
6. ベースレイヤに描画レイヤを合成して画像を時刻 t の履歴として作成してローカルディスクに保存し、Twitter サーバ上に時刻 t の絵としてアップデートする。時刻 t の絵を時刻 $t = t + 1$ のベースレイヤとする。描画レイヤを初期化する。
7. 変動時間 c を乱数で決定し、時刻 $t = t + 1$ とし、3. に戻って繰り返す。

4.2 描画コマンド

描画内容を表すベクトル $d = (p, c, n, r)$ を定義する。以下で、各要素について示す。

座標 $p(x, y)$ で表される描画レイヤ上の点。

色 $c(r, g, b)$ で表される RGB の 3 要素を持つ色。

回数 n 描画回数。

半径 r 単位は px。

表 2: d の要素と言葉の対応

要素	言葉	値
座標	「上」	$p = (0, a)$
	「下」	$p = (0, S/2) + (0, a)$
	「右」	$p = (S/2, 0) + (a, 0)$
	「左」	$p = (a, 0)$
	「真ん中」	$p = (S/2 - a, S/2 - a) + (a, a)$
色	登録している色名	要素を一意に決定
	それ以外	一様乱数で要素を決定
回数	1 ~ 9 回	各数に対応して n を決定
	「いっぱい」「たくさん」	$n = 20$ (現在)

- S : キャンパスの大きさ. $x-y$ 座標の最大値. 単位は px. 図 5 の例では, 400.
- a : $[0, L]$ の一様乱数.
- L : 乱数の最大値. 定義域 $0 \leq L \leq S/2$ を満たす任意の値.



図 5: キャンパス 1

以下に, 描画コマンドとして現在使用できる機能を示す.

1. d をもとに, 座標 p を中心として, 半径 r で色 c で n 個の図形を描画する.
2. p 中の要素である色 c を変更する.
3. 画像パーツを p を始点として配置する.
4. 一つ前の操作を戻す.
5. 描画レイヤを初期化する.
6. 描画レイヤとベースレイヤを保存して Twitter サーバ上にアップデートする.
7. d 中の要素である c の色で自由に線を引く.

ユーザはツイートの内容で 1.~6. を実行することが可能である. 1.~6. に加えて, 5. 章で示すマスターのみキャンパスに直接 7. の操作が可能としている. 画像パーツとは予め用意した透過 PNG 形式の画像であり, 外部ファイルで特定の言葉との対応付けを指定可能である.

4.3 自然言語との対応

現在は, 描画コマンドおよび d の各要素はユーザの入力とあらかじめ登録したキーワードによって決まる. 一つの入力内に複数のコマンドにマッチするキーワードが含まれている場合は, 作成者側で指定した優先順位やコマンドの組合せをもとに複数実行する場合と, 一つだけ実行する場合がある. 現在, 実装済みの言葉と要素の主要なものを表 2 に示す.

5. マスター権限による絵の操作

ローカルクライアント側の PC には 4.1 節の手順の 6. のタイミングで自動的にユーザのツイート内容が反映されるキャン



図 6: キャンパス 2, GUI とタイムラインの例



図 7: キャンパス 3, GUI とタイムラインの例



図 8: キャンパス 4, GUI とタイムラインの例

パスが表示されている. また, GUI として描画コマンドを実行するボタンを用意しており, キャンパス上でドラッグ&ドロップすることで線を引くことが可能である. つまり, ローカルクライアント上でマスター権限をもつ人物が描画コマンドを実行することにより, サーバ側に反映することができる. ローカルクライアントの操作権を持つ人物をマスター, ソーシャル上の人物をユーザと呼ぶ. 前者と後者は一対多の関係にあるものとする. 本章では, 4. 章のソーシャル上のユーザのツイート内容を反映した描画に加え, マスターの介入する場合の描画手法とその応用を述べる.

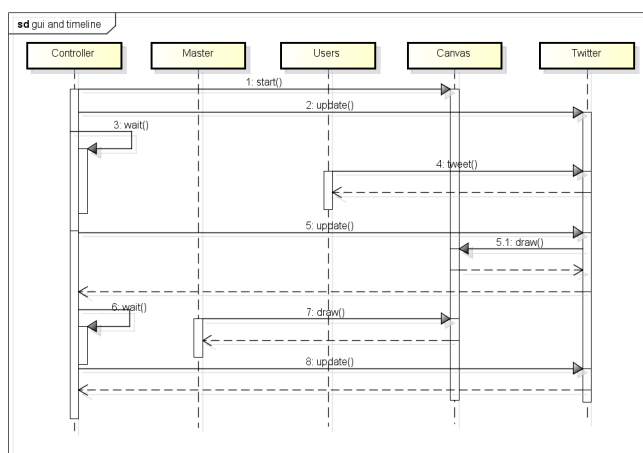


図 9: シーケンス図

5.1 マスターの介在する半自動絵生成手法

図 5 にベースのキャンバスと $x-y$ 座標軸を示す。図 6 ~ 図 8 にキャンバスの移り変わりやツイート内容の例を示す。右側がキャンバスの GUI, 左側が Twitter のタイムラインの例を示す。新たに描画された部分を赤枠で囲んでいる。なお, 例の図中には userA と user B というユーザの Twitter アカウントと, bot のツイッターアカウントである cotopaint というアカウントが存在する。また, 図中にはないが, マスターのアカウントも存在する。

図 6 中の userA の「丸を右上に描いてお花増えたらいいな」という入力に対しては「丸」という形を表す言葉から 4.2 節の描画コマンド 1. を「右」「上」という言葉から, d の p の要素の x, y をそれぞれ乱数で決めており, 図中赤枠で囲まれた中の丸が反映して描画された結果である。

図 9 に図 6, 図 7 に対応した実行の流れをシーケンス図で示す。また, 以下に, その動作の流れをシーケンス図内の番号と対応付けて示す。

1. 起動。ベース画像を読み込んで, キャンバスをローカルクライアントに表示する。
2. お絵描きを始めることを Twitter 上に投稿する。
3. 待機。
4. userA が Twitter Bot 宛てにツイート (メンション) する。
5. 3. で待機中の間にメンションされた内容を取得し, キャンバスに描画して反映し, Twitter 上にも投稿する。また, ことべいんさんから userA に反映内容をメンションする。図 6 に示す。
6. 待機。
7. マスターがキャンバスに描画する。
8. 描画した内容を反映して Twitter 上に投稿し, マスターが絵を描いていることをツイートする。図 7 に示す。

なお, この場合, Twitter Bot はマスターとユーザのインタラクションの間で話題を提示する調整役のような役目を果たす。

5.2 応用

このローカルクライアントの操作機能を使うことにより, マスターの加えた操作内容とユーザのツイートの対応を把握することにより, 自然言語の入力内容と絵の内容の対応をとり, 絵モデルの構成を検討する。複数ユーザが直接的に絵を描画し

たり, チャットをすることができる既存のお絵描きチャットのようなシステムとは異なり, あくまで直接的な描画ができるのはマスターだけであり, ユーザはツイートという自然言語入力を通して絵への変化を加えるという制限により, ユーザ同士の自然言語での入力が描画コマンドに直接関係する内容になるように促す。

本手法をもとにした応用と拡張として考えていることを以下に示す。

絵に関する問い合わせ コマンドの実行を通知するリプライに絵の変化に対して, どう感じたかを問い合わせる内容を追加することにより, コマンドとツイートという言葉をも自動的に対応付けたい。また, 絵モデルには絵についてのユーザの対話ログも合わせて対応付けることにより, 絵の辞書である絵コーパスを作りたいと考えている。そこで, 絵全体で「どういう絵か?」「何をしているか?」などと Twitter Bot もしくはマスターから問い合わせ情報を集める。

リアルタイムでのマスターの操作 図 8 が示すように絵の変化に対してのユーザ同士のツイート内容を参考にしたり, マスターの Twitter アカウントでどのような絵を書いて欲しいか, また次にどのような場面が予想されるかといったことをツイートし, ユーザに意見を求め, 次の絵をマスターが描くことで, 絵を生成するとともに, 絵から想起する情報や期待するシナリオを合わせて考えることにより, 絵モデルの生成および状態遷移のある Pictgent のシナリオを本手法を使って収集したログをもとにして考える。

6. まとめと今後の課題

本研究では, 自然言語を入力として, 絵を半自動生成する手法とソーシャルコミュニケーションサービスを使った拡張について述べた。今後の課題として, 本システムを利用したユーザとのインタラクションを通じて, 言語と絵の対応付けについて定量的な考察をし, 絵の半自動生成手法の改善および作成した絵モデルを一般公開することを目指す。

謝辞

なお, 本研究は一部, 日本学術振興会科学研究補助金基盤研究 (C) (課題番号 22500208) および, リバネス研究費 DeNA 賞の補助を得て行われたものである。

参考文献

- [1] Miki Ueno, Naoki Mori, Keinosuke Matsumoto: Picture Information Shared Conversation Agent: Pictgent, Distributed Computing and Artificial Intelligence Advances in Intelligent and Soft Computing, Volume 151, 91-94, DOI: 10.1007/978-3-642-28765-7_12 (2012)
- [2] 神原啓介, 塚田浩二: オノマトペン, 日本ソフトウェア科学会 論文誌 (コンピュータソフトウェア), Vol.27, No.1, pp.48-55 (2010)
- [3] 小松 孝徳, 秋山 広美: ユーザの直感的表現を支援するオノマトベ表現システム, 電子情報通信学会論文誌. A, 基礎・境界 J92-A(11), 752-763, (2009)
- [4] Kazunari Ito, Motohiro Matsuda, Martin J. Durst and Koiti Hasida: SVG Pictograms with Natural Language Based and Semantic Information, SVG Open (2007)