

絵描き歌生成のための手描きスケッチ類似度

1 はじめに

近年、機械学習の発展を背景に人工知能 (Artificial Intelligence: AI) が注目を浴びている。AI は単純なパターン認識では人間の能力を凌駕する性能を示す一方で、人間の情緒や感性に関する分野、特に人間の創作物への AI の適用はいまだ困難である。創作物の中でも、スケッチは、年齢、国籍、文化にかかわらず、幅広く共感可能な表現であるため重要度が高い。このため、感性を取り扱う分野での重要な AI の課題の 1 つとされてきた。また人工知能で画像データを扱う問題の場合、画像データの形式は画素の集合で高次元なラスタ形式のことが一般に多く、次元が少なく描き順の情報も持てるベクター形式での研究は多くない。

本研究ではベクター形式のイラストのスケッチデータを用いて、創作物の 1 つである絵描き歌の生成を目指す。絵描き歌は画像、言語、音楽など様々なモードを含むため、極めて創作性を要する問題である。先行研究では単純なスケッチに対して、オブジェクトと位置関係を示す文章の生成に成功している。本研究では先行研究の手法を改良しより複雑なスケッチについて、絵描き歌らしい情緒的な歌詞を生成することを目標とし、改良したスケッチ分割手法を提案する。

2 要素技術

2.1 Sketch Recurrent Neural Network

Sketch Recurrent Neural Network (Sketch RNN) [1] とは深層学習を用いて、人間がイラストを描くときの一連の描き順を学習することでユーザのイラストを補完、推測する深層学習を利用したモデルである。Sketch RNN の学習モデルは Variational Autoencoder (VAE) を基に構成されている。入力スケッチの時系列を含む描き順を意味する行列であり、潜在ベクトルがエンコーダの出力となる。また、デコーダの出力は入力と同じ、画像を生成できる行列となる。

2.2 Quick Draw データセット

Google が公開する描き順データセットである Quick-Draw! [2] の形式を使用する。この形式では、1 つの手描きスケッチを構成する描き順データは行列で保持される。描き順データの行方向は手描きスケッチの描き始めから描き終わりまでのペンの状態に関する情報を時系列に沿って保持する。また、描き順データの列方向はある時刻 t におけるペンの状態を保持しており、 x 座標の相対値、 y 座標の相対値、ストロークの描き終わりのブール値をそれぞれベクトルで持っている。

2.3 Cosine Similarity

コサイン類似度 (Cosine Similarity) は、2 つのベクトルの類似性を表す尺度である。ベクトルがなす角がコサイン類似度となるため、1 のとき完全に類似し、0 のとき無関係となる。

$$\cos(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|} \quad (1)$$

2.4 Accelerated KAZE

Accelerated KAZE (AKAZE) [3] は任意の画像に対して類似している画像を探す手法である。AKAZE は特徴点検出を行い画像の拡大や縮小回転に対して不変となっている KAZE をもとにしており、KAZE と同様に高い認識精度を持ちつつ計算に必要な時間を大幅に短縮している。類似度指標としては、マッチングした特徴点の距離平均を用いた。したがって 0 のとき完全に一致し、無関係のとき画像サイズに近くなる。

3 提案手法

3.1 絵描き歌

藤井ら [4] による絵描き歌生成の先行研究ではこの問題を 3 つに分割していた。

まず「スケッチの類似度比較」では、2 つのスケッチ同士の類似度を求める場合の適当な指標を定める。

次に「クラスラベル列の取得」では、求めた類似度指標で 1 ストロークごとにデータベースと比較し、最も近いクラスラベルの順列を取得する。

最後に「歌詞生成」では、クラスラベルや位置情報から、絵描き指示を生成的に構築していた。

本研究ではまず類似度指標を見直し、スケッチデータで考えられる 2 つの途中経過と最終状態について、新たにそれぞれを改良した指標を導入する。

また絵描き歌では一般に似ている別のオブジェクトを合成することで目的のスケッチを完成させるので、2 つ目のクラスラベル列の取得についてスケッチの分割方法を改良する。先行研究では 1 ストロークごとに分割していたが、性能の向上のため意味のあるオブジェクトごとに分割する新たなスケッチ分割手法を提案する。

3.2 手法

本研究では絵描き歌生成のためにまず「スケッチの類似度比較」と「スケッチの分割」を実験した。それぞれを実験 1, 実験 2 とする。

実験 1 ではスケッチの類似度指標となる 3 つの候補について、スケッチの類似度を比較しその結果を定性的に評価した。

実験 2 では実験 1 で求めた指標を用いて実際にスケッチを分割し、結果を比較した。

4 実験 1

実験 1 では最適なスケッチの類似度指標を求める。スケッチでは線を書く順番の「途中経過」と完成した「最終状態」を合わせた 2 つの状態があり、この 2 つの状態を比較する必要がある。

途中経過はスケッチの時系列データを扱うことが出来る Sketch RNN を利用した。Sketch RNN は Variational Autoencoder (VAE) であるため、スケッチの時系列データを入力することで潜在ベクトル z を得ることが出来る。この z のコサイン類似度を途中経過の類似度指標とした。 $S(x, y)$ を類似度関数, $E(x)$ をエンコードとすると, S_a は次式。

$$S_a(U, V) = \frac{E(U) \cdot E(V)}{|E(U)| |E(V)|} \quad (2)$$

次に最終状態はスケッチを画像化し、画像比較手法 AKAZE で比較することにした。先行研究では画質評価手法 Structural Similarity (SSIM) が用いられてい

表 1: 実験 1 の設定

Parameters	
learning step	150,000
learning rate	0.001
batch size	100
latent shape	(128,)
decode temperature	0.01

たが, AKAZE は特徴点マッチングをするため画像の内容を比較できる。

$$S_b(U, V) = A \circ C_1 \circ \text{Sigmoid} \circ C_2(U, V) \quad (3)$$

ここで A は AKAZE 類似度, C は S_a とスケールを合わせるための補正関数で, $C_1(x) = \frac{120-x}{10}$, $C_2(x) = \frac{x+1}{2}$ である。

これらを足し合わせ, 途中経過と最終状態の両方を考慮する混合指標 S_c を

$$S_c(U, V) = \frac{S_a(U, V) + S_b(U, V)}{2} \quad (4)$$

とし, 式 2 ~ 4 でスケッチの類似度指標を評価する。

4.1 実験設定

スケッチデータセット QuickDraw から、対象とするスケッチとこれと比較するスケッチ 10 枚を取り出し、3 つの類似度指標で実験した。取り出したスケッチはいずれも "sheep" クラスである。

実験の設定は表 1 に示す。Sketch RNN のパラメータは事前に長時間学習したものを使用した。

4.2 結果

図 1, 2, 3 に実験 1 の結果を示す。左から対象のスケッチと類似度が高いと判断された上位 3 つのスケッチである。

手法 A の Sketch RNN のみの場合は、描き順を考慮しているため最終状態はばらばらとなった。しかし胴体と頭、足を持った羊が右を向いていることから、描き順という点で評価すると妥当であると言える。

次に手法 B の AKAZE のみの場合は、羊毛の表現がなされているスケッチが選ばれており、手法 A の場合と比較して極めて似ていることがわかる。一方で体の部位や向きなどの描き順は無視されている。



図 1: (手法 A) Sketch RNN

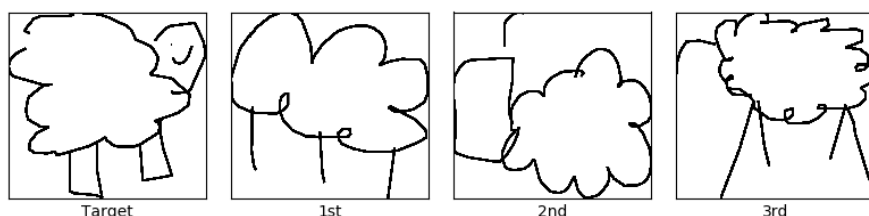


図 2: (手法 B) AKAZE

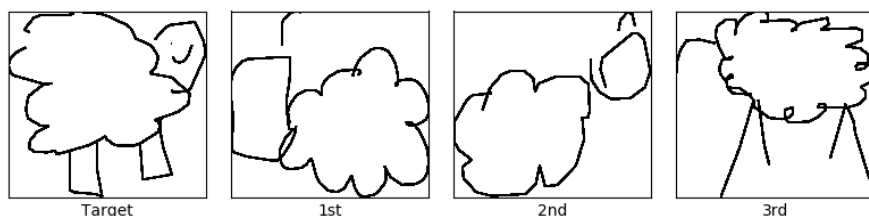


図 3: (手法 C) Sketch RNN + AKAZE

最後に手法 C の両方を混合した指標は、AKAZE のみの手法 B と大きな変化はないが結果の上位 2 番目のスケッチはどちらか片方のみの場合ではなかったスケッチで、羊毛と体の向きも類似しており 比較がうまくいく場合もあることが確認された。

5 実験 2: スケッチの分割

実験 2 では実験 1 で求めた指標を用いて分割実験をした。対象のスケッチに含まれる一部を取り出したものを「部分スケッチ」と呼ぶことにする。

実験 1 の 3 つの類似指標それぞれで、部分スケッチに最も近い データベース中のクラスを求め、対象のスケッチから最尤の部分スケッチのクラスラベルを得た。

5.1 設定

分割方法はスケッチのストローク列の中で 1 箇所を分割し、2 つのスケッチに分けるように部分スケッチを得た。したがって部分スケッチのストロークは順番が保たれる。

データベースの 345 クラスの代表スケッチ 1 枚から最も類似度の高いものを部分スケッチの推定クラスとし、全体で最大の類似度となった分割を求める。

実験 1 同様に "Sheep" クラスを分割対象として選んだ。このスケッチは 7 画であるため、候補は 12 個となった。

5.2 結果

図 4 ～ 7 に実験 2 の推定結果を示す。図 4 はストローク列の 1 箇所でも分割される可能性がある部分スケッチの候補である。図 5 ～ 7 は各類似度におけるそれぞれの推定する類似スケッチとなる。図中の赤枠は類似度最大の部分スケッチである。

AKAZE 類似度を利用している手法の部分スケッチが一致した。最も似ているのは Sketch RNN + AKAZE の "hat" クラスと思われる。途中経過と最終状態を両方考慮することで、類似度の性能が上がっている可能性がある。定量的に評価するためアンケート調査を行いたい。

他の部分スケッチについてみると、体を "bush" や "cloud" , 足を "coffee cup" と比較するなど主観的



図 4: 分割候補となる部分スケッチ

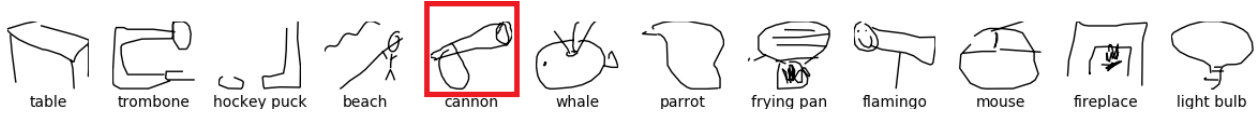


図 5: 手法 A の Sketch RNN の推定結果

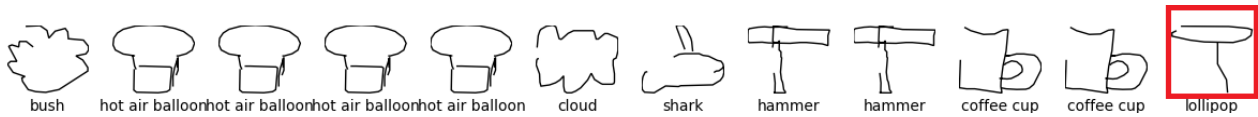


図 6: 手法 B の AKAZE の推定結果

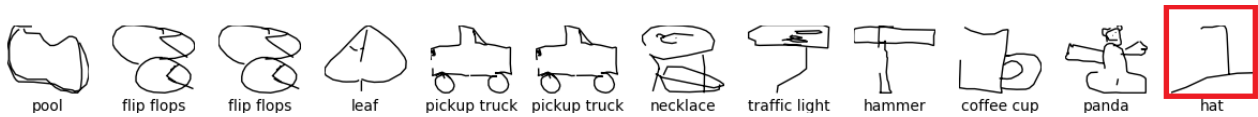


図 7: 手法 C の Sketch RNN + AKAZE の推定結果

に極めてマッチした推定結果もあったが、これらが選ばれることはなかった。このようなマッチしていると思われる部分スケッチが選ばれるように、類似度指標の補正などの改良が必要であると考えられる。

6 まとめと今後の課題

書き順を含むスケッチデータに対して絵描き歌生成のための新たな類似度指標とスケッチの分割方法の提案をした。

類似度指標では、AKAZE を用いた最終状態を入力する手法が良い結果となったが、Sketch RNN のパラメータの性能が低かったことも原因として挙げられる。

スケッチの分割方法では、先行研究ではなかったオブジェクトベースの視点を取り入れ、一部では妥当な部分スケッチが得られた。

本研究では定性的な評価のみであったため、定量的な評価としてアンケート調査をしたいと考えている。

今後の課題は本研究で得られたクラスラベルをもとに、言語モデルを用いた絵描き歌の歌詞の生成をすることである。

また画像情報と言語情報がセットとなっているような新たな絵描き歌用のデータセットを整備することが

課題としてあげられる。

参考文献

- [1] David Ha and Douglas Eck. A Neural Representation of Sketch Drawings. *ArXiv e-prints*, April 2017.
- [2] Salman Cheema, Sumit Gulwani, and Joseph LaViola. Quickdraw: Improving drawing experience for geometric diagrams. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1037–1064, New York, NY, USA, 2012. ACM.
- [3] Pablo F Alcantarilla and T Solutions. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell.*, 34(7):1281–1298, 2011.
- [4] 森 直樹 藤井 涼佑. 手描きスケッチの比較手法を用いた絵描き歌自動生成システムの提案. *修士学位論文*, 2019.