

手描きスケッチの類似度

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、極めて創作性が高い創作物の形態である。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-rnn

sketch-rnn 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 描き順データ

Google が公開する描き順データセットである Quick-Draw! [1] 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 cos 類似度

コサイン類似度は、そのまま、ベクトル同士の成す角度の近さを表現するため、三角関数の普通のコサインの通り、1 に近ければ類似しており、0 に近ければ似ていないことになる。

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

2.4 akaze

類似画像検索は任意の画像に対して類似している画像を探すことである。本研究では、類似画像検索のアルゴリズムとして Accelerated KAZE (AKAZE) [2] とヒストグラム比較を用いた。AKAZE は特徴点およびその特徴量を抽出するアルゴリズムの一つであり、拡大縮小や回転に強いという利点がある。本研究における類似度の計算には特徴点マッチングをし、距離の平均を取るという手法を用いた。特徴点マッチングでは距離を算出しているため同一画像の場合、類似度は 0 となり、値が小さくなるほど類似度が高いことを示す。ヒストグラムは横軸を画素値、縦軸を画素値の出現頻度と設定したグラフである。ヒストグラムから画像のコントラスト、明るさ、画素値の分布を確認することができる。ヒストグラム比較では主に色を用いた類似画像検索が可能となっている。

3 提案手法

3.1 絵描き歌

絵描き歌生成システムについて、藤井ら [] の先行研究では、～であった。

この問題を「スケッチの類似度比較」「スケッチの分割」「クラスラベルから歌詞生成」の 3 つに分割した。

絵描き歌では、似ている別のスケッチを合成することでプログレッシブ?的に目的のスケッチを完成させる。そこで目的のスケッチを「全体スケッチ」、元のスケッチの一部を取り出したものを「部分スケッチ」とし、部分スケッチの和が全体スケッチとなるような分割方法を実験する。

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「スケッチの分割」では、求めた類似度指標によって最適な分割を求め、歌詞生成のための情報として部分スケッチのクラスラベルを獲得する。

「クラスラベルから歌詞生成」では、すでに得たクラスラベルやその位置情報から、絵描き歌の歌詞を言語モデルを利用して生成する。

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部分スケッチを求め、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。

表 1: 実験 1 の設定

	name
step	150,000
class	345
latent	(128,)
learning rate	0.0000
learning rate	0.0000
learning rate	0.0000
learning rate	0.0000
learning rate	0.0000

ここに数式

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。

ここに数式

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を

ここに数式

とし、式 ~ でスケッチの類似度指標を評価する。時間の都合上～は 1 とした。

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全て "sheep" クラスで揃えた。

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。

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 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、Sketch RNN ノミの場

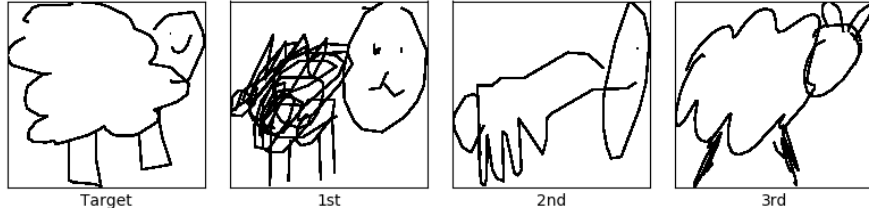


図 1: (手法 A) Sketch RNN



図 2: (手法 B) AKAZE

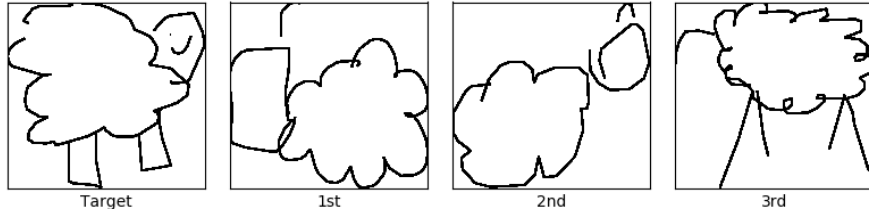


図 3: (手法 C) Sketch RNN + AKAZE

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．

最後に両方を混合した指標は，AKAZE のみの場合と大きな変化はないが 2 番目のスケッチは片方のみの場合ではなかったスケッチで，羊毛と体の向きが類似しており 比較がうまくいく場合もあることが確認された．

AKAZE の影響が強いことと Sketch RNN の性能が高くないことから，係数 a を調整する必要がある．

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実際に部分スケッチを求めた．

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を 1 つ分割した．

5.1 設定

分割方法はスケッチのストローク列の中で 1 箇所分割するように部分スケッチを得た．

データベースの 345 クラスの代表スケッチ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．

主観では体を ”bush” や ”cloud”，足を ”coffee cup” と比較してほしかったため，単純な部分スケッチが選



図 4: 分割候補となる部分スケッチ

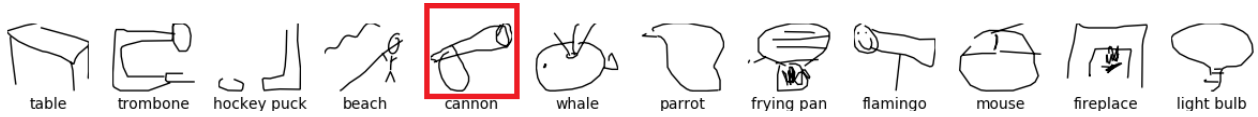


図 5: Sketch RNN の推定結果

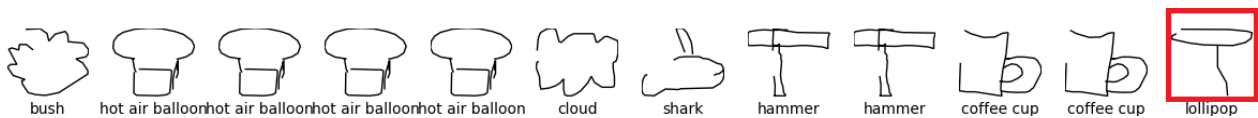


図 6: AKAZE の推定結果

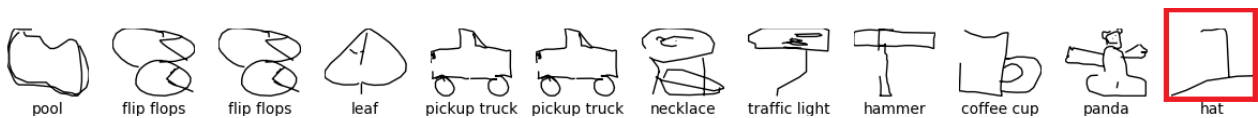


図 7: Sketch RNN + AKAZE の推定結果

ばれにくいような補正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～実験をし、類似度指標の考察をした。

絵描き歌用のデータセットを整備すること。

分割したクラスに対して言語モデルを通して、絵描き歌の歌詞を生成することが課題としてあげられる。

ここにテキストを入力してください。ここにテキストを入力してください。ここにテキストを入力してください。ここにテキストを入力してください。ここにテキストを入力してください。ここにテキストを入力してください。ここにテキストを入力してください。

参考文献

- [1] 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.

- [2] 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.