

筆順の類似度に関する評価モデルに対する考察

A Study on Evaluation Model on Stroke Order Similarity

藤井 涼佑^{*1} 森 直樹^{*1} 岡田 真^{*1}
 Ryosuke Fujii Naoki Mori Makoto Okada

^{*1}大阪府立大学工学研究科
 Graduate School of Engineering, Osaka Prefecture University

Since sketch is one of the most important creation by human, analysis of sketch drawing process by artificial intelligence is the essential topic. In this paper, we consider a computational model to evaluate the similarity of stroke order. We have proposed CASOOK in order to realize communication between computers and humans. In this research, we are trying to develop CASOOK-SR which is extended CASOOK adopting sketch-rnn. Since evaluating stroke order similarity is a difficult problem, and there was only a method to evaluate the similarity of stroke order by persons. In this paper, we define an objective comparable evaluation function on the similarity of stroke order. In our experiment, the correlation coefficient between the function and the similarity evaluation values by a person were 0.344 and 0.446. It is found that our calculation model is the same as human evaluation when an image made from a stroke order matrix is similar to another.

1. はじめに

近年、機械学習の発展を背景とした人工知能 (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が、一方で、人間の情緒や感性に関する分野への AI の適用、特に創作物への適用はまだまだ困難である。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 [Eitz 12]。著者らはこれまで計算機と人間とのコミュニケーションを実現するため、Creative Animating Sketchbook (CASOOK) [Ueno 15] を提案してきた。従来の研究において、CASOOK は筆順の情報を持たない静止画像のみを対象していたが、CASOOK の拡張として筆順を扱うために、深層学習モデルである sketch-rnn [Ha 17] を導入した Creative Animating Sketchbook with sketch-rnn (CASOOK-SR) を提案している。この手法では、筆順の決定プロセスが潜在意識下にあるという観点から、最終画像と同程度に人間の感性が筆順に含まれていることに着目している。また CASOOK-SR を実問題に適用するため、探し絵を自動生成するモジュールとして VSAPS を提案してきた。

しかしながら、過去の研究から生成された筆順を定量的に評価するために必要な類似性の比較は困難な課題の 1 つであり、人によって筆順の類似性を直接評価するしかなかった。そのため、本稿では、筆順の類似性に関する客観的で比較可能な評価モデルを定義し、その有効性を 2 種類の実験によって確認し、創作物としてのスケッチを人工知能で用いることの重要性を示す。

2. 要素技術

2.1 CASOOK

児童を対象としたユーザからのスケッチを入力として、スケッチに込められた感性情報を解析し、感性に合わせたアニメーションを付与するシステムとなっている。

2.2 CASOOK-SR

従来の CASOOK はユーザの入力したスケッチを画像情報のみで解析していた。これを拡張するために、CASOOK-SR では深層学習モデルである sketch-rnn を利用することで筆順情報と画像情報を解析し、筆順方法を活用した上で感性に合わせたマシンインタフェースが実現可能なシステムを目指す。

2.3 Sketch-rnn

sketch-rnn とは、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。図 1 に sketch-rnn の学習モデルを構成する Variational Autoencoder (VAE) [Pu 17] を示す。入力スケッチの時系列を含む筆順行列であり、潜在ベクトルがエンコーダの中間出力となる。また、最終出力は画像を生成可能とする筆順行列となる。

2.3.1 温度パラメータ τ について

sketch-rnn のパラメータである温度 τ は潜在空間の多様性に関して調整をする。温度 τ が 0 の場合出力は決定的、つまり、単一の特徴を強く反映した出力になる。

2.4 VSAPS

静止画像のみを扱う CASOOK に対して、筆順を導入する CASOOK-SR のコアモジュールとして探し絵に適用した Visual Search Automotive Production System (VSAPS) を提案した。VSAPS には大きな特徴が 2 つ存在する。1 つ目はユーザと計算機が対話できる構造を持っていることである。ユーザが描いたスケッチを元に探し絵の問題を作成後、ユーザは自分の描いたスケッチを VSAPS による類似画像の中から選択し、VSAPS がその正誤判定を確認する。2 つ目の特徴は筆順ベクトルを扱うことによって、静止画像の問題形式だけでなく、筆順を再現する動画の問題形式をユーザに提供することである。画像生成を畳み込みニューラルネットワーク (Convolutional Neural Network : CNN) [Simard 03] ベースの VAE ではなく、再帰型ニューラルネットワーク (Recurrent Neural Network : RNN) [Jain 99] ベースの sketch-rnn を使用することで VSAPS は時系列データの学習が可能である。

図 2 に VSAPS の画面の配置を示す。左側の部分にユーザが描いたスケッチを表示し、右側に探し絵の問題の選択肢を表

連絡先: 藤井 涼佑, 大阪府立大学, fujii@ss.cs.osakafu-u.ac.jp

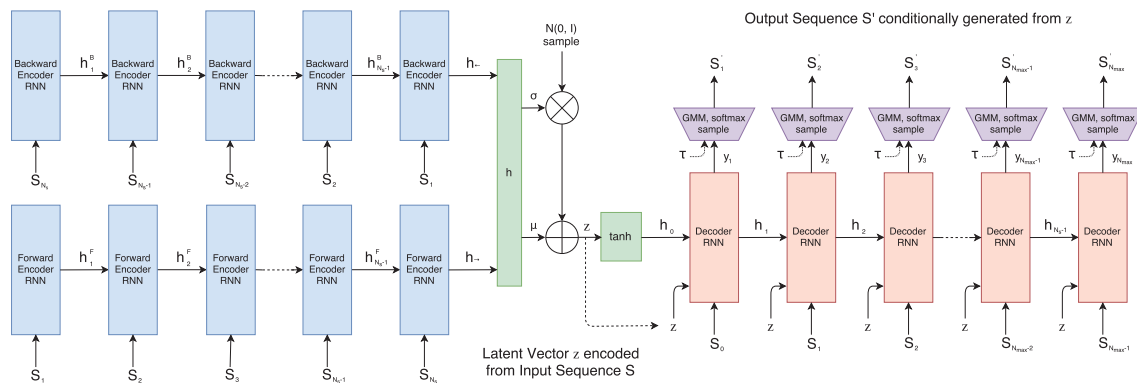


図 1: sketch-rnn モデル図

Where is your sketch?

Choose the same sketch in decode as your encode sketch!

encode	decode1	decode2	decode3	decode4	decode5
ユーザのスケッチ 表示部分					
	decode6	decode7	decode8	decode9	decode10
REPLAY	0m 26.67s				

図 2: VSAPS 画面構成

示す。ユーザは右側の選択肢から正解を選ぶことをゲームの趣旨とする。

3. 提案手法

3.1 評価モデル

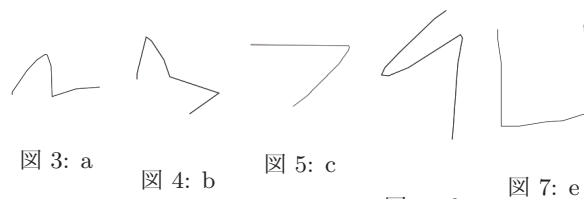
(1) 式に本稿で提案する指標を示す。ただし、 U, V は 1 枚のスケッチの筆順を含む行列で、それぞれのサイズは $t_U \times 3$, $t_V \times 3$ であり、 t_U や t_V は各スケッチの時間長を示し、3 は sketch-rnn がもつペンの状態ベクトルのサイズを示す。sketch-rnn では、ペンの x 座標とペンの y 座標、ペンの上下運動の有無を示す Bool 値をもつ。また、 K は任意の時間を持ち、 $0 < K \leq \min(l_U, l_V)$ を満たす値である。ここで、 l_U はスケッチの筆順行列 U の時間長を、 l_V はスケッチの筆順行列 V の時間長を示す。したがって、ある時刻までの筆順の総和である筆順ベクトルのコサイン類似度 [Singhal 01] の時間軸方向に対する平均値として定義する。

$$S(U, V) = \frac{1}{K} \sum_{k=1}^K \frac{(\sum_{i=1}^k U_i) \cdot (\sum_{i=1}^k V_i)}{(\|\sum_{i=1}^k U_i\|)(\|\sum_{i=1}^k V_i\|)} \quad (1)$$

4. 実験

4.1 目的

筆順の類似性を測る指標として、従来研究で使用されてきた人間の主観による定性的評価の代替となる、(1) 式の有効性を確認する。4.2 節では大きく異なる 2 つの筆順間での評価を扱



う。4.3 節では VSAPS を用いることで、似ている多数の筆順間での評価を扱う。

4.2 2 組の筆順の比較

4.2.1 手順

(1) 式で表される評価モデルが、2 つの筆順間の類似性に関する人間の主観的評価との相関を調べる。テストデータから選んだ 10 枚の画像を描き始め・中盤・描き終わりの 3 段階、つまり 30 種類の画像を、5 人の実験協力者に筆順の類似性に関して「非常に似ている」、「似ている」、「異なる」、「非常に異なる」の 4 段階で評価してもらう。5 人の評価を平均した値と (1) 式の値に関して相関係数を求めて、提案する指標の信頼度を確認する。尚、相関係数を計算するとき、実験協力者が評価する 4 段階について、「非常に似ている」は 4 に、「似ている」は 3 に、「異なる」は 2 に、「非常に異なる」は 1 に変換する。

図 3 ~ 7 に評価対象のスケッチの一例を示す。実験においては画像の比較ではなく、スケッチの筆順をアニメーションで再現したものを比較する。

4.2.2 結果

図 8 に人の評価値と (1) 式で表される評価モデルでえられる評価値の散布図を示す。図より、評価モデルが負の値を取るデータ 12 件はすべて人の主観的評価も非常に低いことがわかる。また、人の評価値が 2.5 以上で評価モデルの評価値が 0 以上であったデータは 19 件あった。

これらのデータを用いて相関係数を計算したところ、0.344 が得られた。したがって、人の評価と評価モデルは弱い正の相関があることを確認した。

4.2.3 考察

正の相関ではなく、弱い正の相関になった要因について考察する。主な原因は筆順行列の類似性と画像の形の類似性の関係性にあると考えている。

特に顕著な現象として、評価モデルによる評価値は高くなる傾向がある一方で、人の評価は低くなる傾向がある点であげ

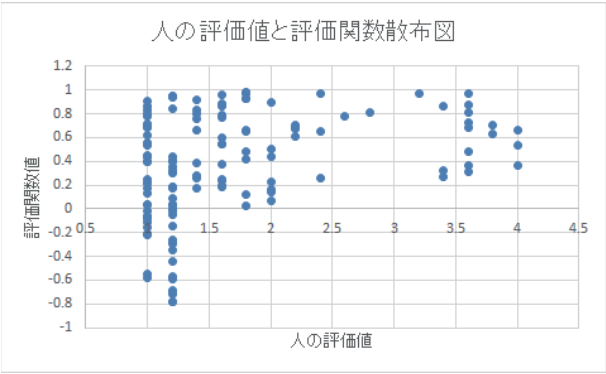


図 8: 人の評価と評価モデルの散布図

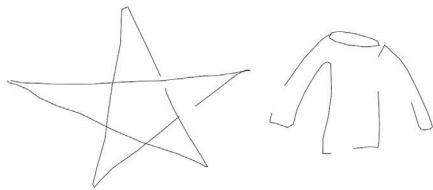


図 9: 実際に使用した画像ペア

られる。評価モデルにおいて、正の値は 135 件中 108 件あった。反対に人の評価において、2.5 以上は 135 件中 19 件であった。これらの現象は、評価モデルでは類似性の加点が多すぎることに起因すると考える。

これを確認するため、図 8 において左上のデータ、つまり、人の評価は低い一方で、評価モデルの値が高い画像のペアを示す。この画像ペアの評価モデルの値は 0.86 であった。一方で、人の評価値は 1.0 であり、「非常に異なる」をすべての実験協力者が回答していた。このデータから、筆順に関するコサイン類似度が高いにもかかわらず、全くの別画像の画像ペアの存在が確認された。

図 10 に図 9 における筆順の同時刻帯のベクトルの比較を示す。図 10 より、同時刻帯のベクトルの向きは図中の 1 ～ 7 番において一致していることがわかる。したがって、最終画像としてみた場合、2 つのスケッチが異なる一方で、筆順データとしてみた場合、類似性の高いスケッチが得られることもあると考えられる。

一方で、図 8 より、評価モデルの値が小さいすべてのデータは人の評価値も低いことが示された。したがって、評価モデルの値が小さければ人の評価は低いことがいえる。このため、提案した評価モデルは 2 の筆順の類似性が極端に低いケースには有効であると考えられる。

4.3 VSAPS を使った比較

4.3.1 手順

次に、VSAPS を使い、自動生成された筆順を使った探し絵の選択肢における類似性の評価をする。

4.3 節の実験では、4.2 節の実験で使ったデータと異なり、比較対象の筆順が非常に高い類似性をもつ。探し絵の類似度は、ユーザの入力した筆順と VSAPS が生成した筆順に対して (1) 式を適用した値を 9 組求め、それらの平均値とした。また、画像の類似具合を変更するパラメータとして、2.3 節で述べた τ の値を変動させた。

図 11 に τ が大きい、つまり、最終画像が非決定的な出力

表 1: VSAPS における類似性の主観的評価と客観的評価

画像 No.	主観的評価	客観的評価
1	2	0.8811
2	1	0.7270
3	1	0.9240
4	1	0.9517
5	3	0.9534
6	1	0.8996
7	1	0.8212
8	3	0.8385
9	1	0.8460
10	4	0.9810

において、評価モデルの値が小さい例を示す。また、図 12 に τ が小さい、つまり、最終画像が決定的な出力において、評価モデルの値が大きい例を示す。ただし、類似性の評価モデルの平均値はおよそ 0.80～0.99 の範囲の値を取った。図 11 より、正解である decode 9 以外の選択肢の筆順はいずれも丸みを帯びているため、人間の主観的にも類似性が低いといえる。一方で、図 12 では decode 4 ～ 10 の選択肢はほとんど区別がつかないほど類似性が高いことが確認される。これらの結果を裏付けるように、図 11 では、評価モデルの平均値は約 0.84 を示し、図 12 では、評価モデルの平均値は約 0.96 を示している。以上から、評価モデルの評価値と VSAPS の難易度は一定の相関があると推測される。これを前提として、探し絵の類似度を変更した場合、人による類似性評価と探し絵の類似度とどの程度相関があるか確認する。

このため、実験協力者に用意済みの探し絵を解いてもらい、探し絵の選択肢における類似性についてのアンケートを回答してもらった。評価項目には「非常に似ている」、「似ている」、「異なる」、「非常に異なる」の 4 つを準備した。ただし、4.3 章では類似性が高い中で、どれくらい異なるかを見ており、4.2 章の評価項目とは全く異なる評価スケールである。

4.3.2 結果

表 1 に 1 人の実験協力者のアンケート結果と評価モデルの値を示す。ただし、主観的評価が実験協力者の回答結果を指し、客観的評価が評価モデルの値を指す。また、主観的評価は、「非常に似ている」は 4 に、「似ている」は 3 に、「異なる」は 2 に、「非常に異なる」は 1 に変換した値となっている。主観的評価と客観的評価の相関係数は 0.446 を示した。したがって、主観的評価と客観的評価に正の相関があることを確認できた。

4.3.3 考察

主観的評価と客観的評価の相関係数は 0.4463 であった。したがって、VSAPS によって生成された筆順行列から構成される最終画像が類似している場合、(1) 式で与える評価値は非常に妥当性のあるモデルであると考えられる。

5. まとめと今後の課題

本稿では、筆順の類似性を測る客観的な指標を作るため、(1) 式の評価値と実験協力者による筆順の評価を比較した。データセットのテストデータからランダムに選んだ筆順の比較をした 4.2 節では、弱い正の相関がある 0.344 という相関係数を実験で確認した。VSAPS で自動生成した類似性の高い筆順同士の比較をした 4.3 節では、正の相関がある 0.446 という相関係数を実験で確認した。したがって、画像特徴における類似性の

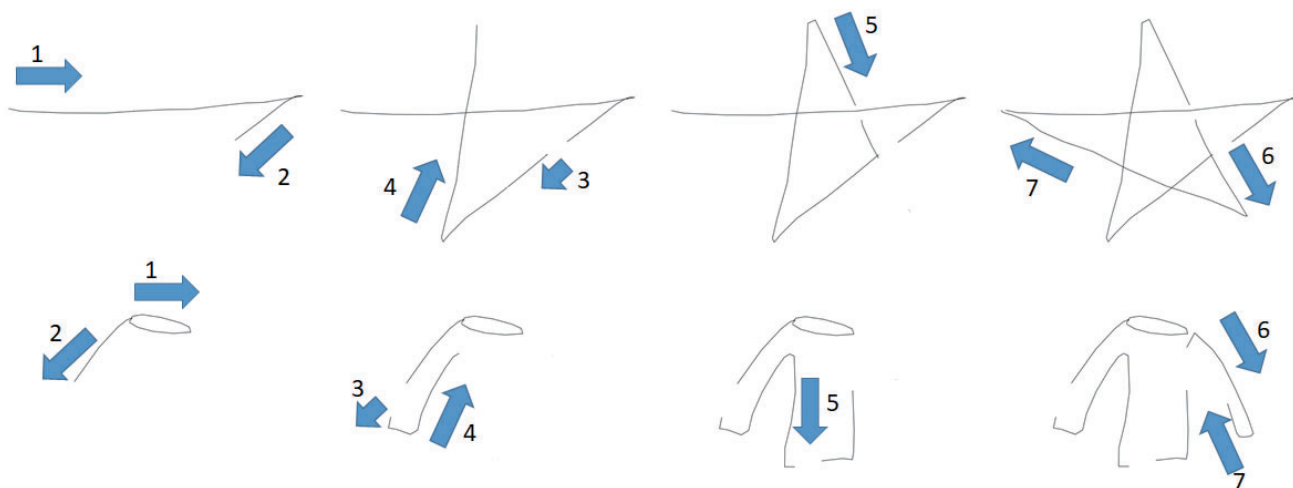


図 10: 筆順の部分ベクトル比較

Where is your sketch?

Choose the same sketch in decode as your encode sketch!

This question's level is 0.8425673606744719.

encode	decode1	decode2	decode3	decode4	decode5
	0.80191274630485 	0.8914049253191638 	0.8583438335643973 	0.7910954194224398 	0.8265369845115962
	0.8596210803674614 	0.8881432539518267 	0.9348854664443136 	0.999999948169839 	0.7331627473672067

図 11: 評価モデルの値が小さい例

高い画像ペアの筆順は (1) 式の評価値が有効であり、部分的に筆順の類似性を測る客観的な指標の作成ができたといえる。

今後の課題としては、まず 4.2 節, 4.3 節において、評価する人数が少ない点について改善が必要である。次に (1) 式の評価値は比較対象の最終画像ペアが大きく異なる場合にも適用できる補正項を必要としているについてより深い考察が必要である。最後に、人の創作物としてのスケッチを用いることで、人と人工知能の間にどのようなインタラクションが生じるかについて詳細な実験が必要である。

尚、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 26330282) の補助を得て行われたものである。

参考文献

- [Eitz 12] Eitz, M., Hays, J., and Alexa, M.: How Do Humans Sketch Objects?, *ACM Trans. Graph. (Proc. SIG-GRAPH)*, Vol. 31, No. 4, pp. 44:1–44:10 (2012)
- [Ha 17] Ha, D. and Eck, D.: A Neural Representation of Sketch Drawings, *CoRR*, Vol. abs/1704.03477, (2017)
- [Jain 99] Jain, L. C. and Medsker, L. R.: *Recurrent Neural Networks: Design and Applications*, CRC Press, Inc., Boca Raton, FL, USA, 1st edition (1999)

Where is your sketch?

Choose the same sketch in decode as your encode sketch!

This question's level is 0.9625661781652095.

encode	decode1	decode2	decode3	decode4	decode5
	0.899890842819132 	0.9145842597978872 	0.9705195546877092 	0.990840776664548 	0.9910460550610612
	0.9806659904921927 	1.0000000029075438 	0.9782431518159261 	0.963306011223115 	0.9739989580177679

図 12: 評価モデルの値が大きい例

- [Pu 17] Pu, Y., Gan, Z., Henao, R., Li, C., Han, S., and Carin, L.: VAE Learning via Stein Variational Gradient Descent, in Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. eds., *Advances in Neural Information Processing Systems 30*, pp. 4236–4245, Curran Associates, Inc. (2017)

- [Simard 03] Simard, P. Y., Steinkraus, D., and Platt, J. C.: Best practices for convolutional neural networks applied to visual document analysis., in *ICDAR*, Vol. 3, pp. 958–962 (2003)

- [Singhal 01] Singhal, A.: Modern Information Retrieval: A Brief Overview, *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, Vol. 24, No. 4, pp. 35–43 (2001)

- [Ueno 15] Ueno, M., Fukuda, K., Yasui, A., Mori, N., and Matsumoto, K.: CASOOK: Creative Animating Sketchbook, in Omatu, S., Malluhi, Q. M., Gonzalez, S. R., Bociwicz, G., Bucciarelli, E., Giulioni, G., and Iqba, F. eds., *Distributed Computing and Artificial Intelligence, 12th International Conference*, pp. 175–182, Cham (2015), Springer International Publishing