# ニューラルネットワークによる文への絵文字装飾

篠田 悠斗　　島岡 聖世　　乾 健太郎
東北大学 工学部電気情報物理工学科　　東北大学 情報学研究科
tohoku_shinoda@yahoo.co.jp,{simaokasonse,inui}@ecei.tohoku.ac.jp

## 1 はじめに

ソーシャルメディアにおいて感情や視覚的概念などを簡潔に表現する手段として絵文字の使用が普及している [2]。文化庁の調査 [7] によれば、全国 16 歳以上の男女のうち、絵文字を見たことがあると回答した人は 85.1%、絵文字を使うことがあると回答した人は 56.1%であり、絵文字が多くの人に触れるまでに普及し、半数以上の人々に使用されていることがわかる。

特に、マイクロブログやチャットアプリなど口語的でカジュアルなコミュニケーションが頻繁に行われる媒体では、絵文字は表現に華やかさや感情的な起伏を付け加える手段として欠かせないものになった。

このように絵文字の使用が普及するにつれ、一般的な文字変換システムを使うよりもさらに手軽に絵文字を使用する需要が高まっていると考えられる。

そこで本論文では絵文字を用いた文を手軽に生み出す手法として、絵文字を使用しない文を入力として受け取り、それに対して文意に合致した絵文字を挿入することで文の絵文字による装飾を行う、再帰的ニューラルネットワークに基づくモデルを提案する。

実験では、ツイッター API により集められた絵文字を含むツイートのコーパスを用いて学習を行った。学習後のモデルを使って実際に幾つかの文を絵文字により装飾し、本手法の有効性を定性的に確かめた。

## 2 モデル

### 2.1 問題設定

分かち書き処理により得られた長さ $n$ の形態素の系列 $w_1, w_2, ..., w_n$ が入力として与えられる。出力として、長さ $m$ の絵文字の系列 $e_1, e_2, ..., e_m$ とそれらを入力された形態素系列のどこに挿入するかを表す位置系列 $p_1, p_2, ..., p_m$ を返す。例を図 1 に示す。

入力：ピザ 食べ た よ
出力：(🍕,2) (😋,5) (😎,5)
↓
ピザ🍕食べたよ😋😎

図 1: 入力文に絵文字を装飾する様子

### 2.2 機械学習モデル

機械学習のモデルとして Seq2Seq[5] を適用する。提案モデルの全体像を図 2 に示す。入力された形態素系列 $w_1, w_2, ..., w_n$ を、形態素から $d_w$ 次元実数ベクトルへの写像 $u_w$ を用いて、エンベディング系列 $u_w(w_1), u_w(w_2), ..., u_w(w_n)$ に変換する。エンベディング系列はエンコーダと呼ばれる再帰的ニューラルネットワークにより処理され、エンコーダの $d_{enc}$ 次元の中間層の状態ベクトル系列 $enc_1, enc_2, ..., enc_n$ が得られる。中間層の $t$ 番目の状態ベクトルは以下のような再帰式で定められる。

$$enc_t = rnn\_cell(u_w(w_t), enc_{t-1}; \theta_{enc}) \quad (1)$$

ここで $rnn\_cell$ は $t-1$ 番目の中間層の値と $t$ 番目の入力から $t$ 番目の中間層の値を計算するモジュールであり、本研究では GRU セル [1] を使用した。$\theta_{enc}$ はモジュールのパラメタを表す。エンコーダの中間層の初期状態はゼロベクトルとした。

次にデコーダと呼ばれる再帰的ニューラルネットワークを用いて絵文字とその挿入位置の系列を生成する。デコーダの中間層の初期状態はエンコーダの中間層系列の最後の要素 $enc_n$ とする。

$t$ 番目の絵文字 $e_t$ と挿入位置 $p_t$ の生成確率の計算方法は以下の通りである。まず、その直前の絵文字 $e_{t-1}$ および挿入位置 $p_{t-1}$ がそれぞれ写像 $u_e, u_p$ により、$d_e$ および $d_p$ 次元のエンベディング $u_e(e_{t-1}), u_p(p_{t-1})$ に変換される。これら 2 つのエンベディングの連結ベ

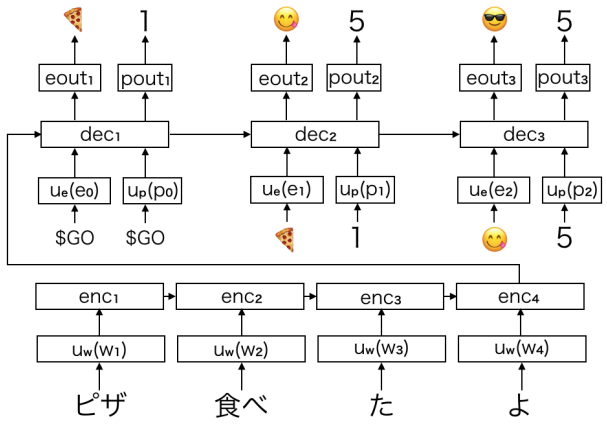


図 2: Seq2Seq に基づく絵文字装飾のモデル

クトル $[u_e(e_{t-1}), u_p(p_{t-1})]$ と直前の中間層の値を用いて、次の中間層の値が計算される。

$$dec_t = rnn\_cell([u_e(e_{t-1}), u_p(p_{t-1})], dec_{t-1}; \theta_{dec}) \quad (2)$$

ただし、$t = 1$ の場合、直前の絵文字と挿入位置が存在しないので、代わりに出力開始を表す特別な記号 $\$GO$ を $e_0$ および $p_0$ として使う。中間層の状態ベクトル $dec_t$ を入力として $t$ 番目の絵文字 $e_t$ と挿入位置 $p_t$ の生成確率を表すソフトマックス層 $eout_t$ および $pout_t$ が計算される。

$$eout_t = softmax(W_e dec_t + b_e) \quad (3)$$

$$pout_t = softmax(W_p dec_t + b_p) \quad (4)$$

ただし、絵文字の種類数を $d_{eout}$ として $W_e$ は $d_{eout} \times d_{dec}$ 次元の重み行列であり、$b_e$ は $d_{eout}$ 次元のバイアスベクトルである。また、挿入位置の種類数を $d_{pout}$ として $W_p$ は $d_{pout} \times d_{dec}$ 次元の重み行列であり、$b_p$ は $d_{pout}$ 次元のバイアスベクトルである。

損失関数として、絵文字と挿入位置のそれぞれのソフトマックス層の出力に対して、データから計算される対数尤度の単純平均を用いる。

# 3 実験

## 3.1 データセット

機械学習のためのコーパスとして、ツイッター API[1] を用いて 2015 年から 2016 年までのツイートのうち、絵文字が 5 回以上使用されている 763 万件のツイートを集めた。このうちの 700 万件を訓練データセットとして用い、残りの 63 万件を学習の終了時点を決定するためのバリデーションデータセットとした。絵文字の抽出は正規表現のパタンマッチングにより行い、分かち書きには形態素解析ツールである MeCab[4] を利用した。付録ではツイッターにおける絵文字の使用に関して、さらなる分析を行っている。

[1] https://dev.twitter.com/overview/api

## 3.2 ハイパーパラメタと実装

形態素の語彙数は 200000 とした。入力系列長 $n = 60$、出力系列長 $m = 8$ とした。学習データのインスタンスの系列長がこれらのハイパーパラメタよりも大きい場合は、はみ出した部分系列は無視し、逆に小さい場合は足りない部分系列を空白を表す記号 $\$PAD$ で埋めた。絵文字の出力層の次元は $d_{eout} = 752$、挿入位置の出力層の次元は $d_{pout} = 61$、エンコーダとデコーダの GRU セルの中間層の次元は $d_h = 300$、形態素、絵文字、挿入位置から、エンベディングへの写像 $u_w, u_e, u_p$ は実数ベクトルを値とする辞書として実装し、各エンベディングは乱数で初期化し、学習時の可変パラメタに含めた。エンベディングの次元はそれぞれ $d_w = 300, d_e = 200, d_p = 100$ とした。誤差逆伝播と確率的最急降下法に基づく最適化を行い、バリデーションデータセットで測った損失が最小の時点で学習を終了した。

事前処理と 2 節で記述した機械学習モデルの実装を含んだソースコードはリポジトリ `https://github.com/spinofi/NeuralEmojiDecorator` に公開されている。

## 3.3 定性的評価

学習後のモデルを使って実際に幾つかの単純な文を絵文字により装飾した。出力系列の生成には貪欲法を用いた。つまり、各時点においてそれぞれのソフトマックス層の出力ベクトルの最大値に対応する絵文字と挿入位置を選んでいく。

装飾の例を図 3 に示す。出力例から観察されることは以下の通りである。

1. 文意に合致した適切な絵文字が生成されている。
2. 1 種類の絵文字や 2 種類の絵文字の組が繰り返し使われる傾向がある。
3. 全ての例において、絵文字は文末にのみ付け加えられている。

バスケ が やり たい です 🏀🏀🏀🏀🏀
ピザ パーティー なう 🍕🍕🍕🍕🍕🍕🍕🍕
マック なう 🍟🍔🍟🍔🍟🍔🍟🍔
よろしく お願い し ます ！ 🙏🙏🙏🙏🙏🙏🙏🙏
サンタ さん が たくさん い まし た 🎅🎅🎅🎅🎅🎅🎅
海 行き たい 🚣🚤🚣🚤🚣
朝 は コーヒー です ね ☕☕☕☕☕☕☕☕
眠い です ね 😪😪😪😪😪😪😪😪
おはよう ！ 😃😁😁🙆🙌😁😁🌄
めちゃくちゃ 感動 し まし た ！ 😭😭😭😭😭😭😭😭
紅葉 が 綺麗 です ね 🍁🍁🍁🍁🍁🍁🍁🍁
お 誕生 日 おめでとう ！ 🎈🎉🎁🎂🎉
素晴らしい です 👏👏👏👏👏👏👏👏

図 3: 提案モデルによる絵文字装飾の例

## 4 関連研究

萩原ら [6] は、モバイル用 Web サイトにおける絵文字の用法を意味的、機能的、装飾的の 3 つに分類している。意味的用法では絵文字が具体的な語の代替として利用され、機能的な用法では「検索」、「書き込み」、「投稿日時」など Web ページにおける特定の機能や情報を表すのに用いられる。これらに対し、装飾的な用法では絵文字は単に文を装飾する目的で用いられる。本論文の手法は、装飾的な用途で絵文字を使いたい場合に適していると考えられる。

Eisner ら [3] は分布的意味論に基づいた絵文字のエンベディングを作成し、これを素性として用いることで極性分析のタスクにおいて性能を向上させることができることを示した。学習済みの絵文字エンベディングはオープンソースとして公開されている [2] ため、今後の発展としてこの資源を提案モデルに組み入れることで性能を向上させることが考えられる。

## 5 結論

本論文では、絵文字の使用をより手軽にすることを目的として、文に対して文意に合致した絵文字を自動的に装飾する手法を提案した。提案手法は Seq2Seq の枠組みに基づき、形態素系列を入力とし、絵文字系列および絵文字を入力系列のどの位置に挿入するかを示す系列を出力とする。絵文字を含む 700 万件のツイートを訓練データとして用い学習を行なった結果、提案手法の有効性を定性的に確認することができた。

[2] https://github.com/uclmr/emoji2vec

課題として、長い文に対する性能が良くないことや、同じ絵文字が繰り返し使われたり、文の末尾にのみ絵文字が付け加えられるなど、装飾法が多様性に富むものではないことが挙げられる。

今後の発展としては、学習データの量を増やすこと、他の機械学習モデルを試すこと、出力系列の生成方法を変えることなどが挙げられる。特に、実験で観察された、文末にしか絵文字が付かない問題については、系列ラベリング問題のように入力文の各地点で絵文字を出力するようなモデルにより改善が可能であると考えている。

## 謝辞

Pontus Stenetorp 氏には本研究の全般に渡り有益なフィードバックをいただいた。本研究は東北大学工学部 情報知能システム総合学 科「Step-QI スクール」の支援を受けた。また、本研究は JSPS 科研費 15H01702 の助成を受けた。

## 参考文献

[1] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[2] Oxford Dictionaries. Oxford dictionaries word of the year 2015 is…, 2015.

[3] Ben Eisner, Tim Rocktäschel, Isabelle Augenstein, Matko Bošnjak, and Sebastian Riedel. emoji2vec: Learning emoji representations from their description. *arXiv preprint arXiv:1609.08359*, 2016.

[4] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *EMNLP*, Vol. 4, pp. 230–237, 2004.

[5] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pp. 3104–3112, 2014.

[6] 萩原正人, 水野貴明. モバイル検索システムのための絵文字に対する意味解析. 言語処理学会第 16 次年次大会発表論文集, pp. 567–570, 2010.

[7] 文化庁. 平成 27 年度「国語に関する世論調査」の結果の概要, 2016.

## 付録：ツイッターにおける絵文字使用

ここでは、絵文字のツイート中での使い方について調査する。特に着目するのは、複数個の絵文字を組み合わせる用法である。コーパスは 3.2 節述べたものの一部を使用した。

### 絵文字 n グラムの頻出頻度

絵文字の 1,2,3 グラムの出現頻度上位 20 個を図 4 に示す。ただし、2,3 グラムに関しては同一の絵文字が連続するものを除いた。

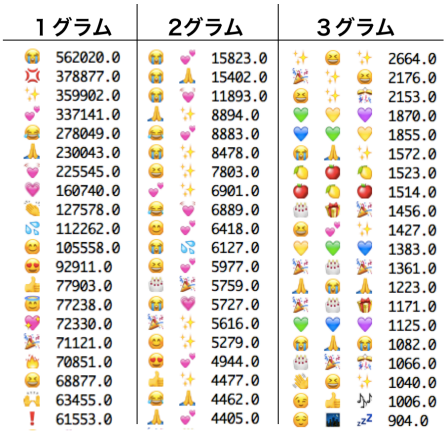


図 4: 絵文字 n グラムの頻出頻度上位 20 個

### 絵文字の順番が意味を持つか

絵文字 2 グラムの出現頻度上位 100 個において、2 グラムの順番を入れ替えた場合のそれぞれの出現頻度の比率を上位下位それぞれ 20 個について図 5 に示す。下位 20 個は絵文字同士を交換しても意味の変化が現れない組み合わせだと解釈できる。色違いのハートや果物の組み合わせなど、意味的に同質の絵文字を連続して使っていることがわかる。上位 20 個は 2 つの絵文字の順番が意味に重要な効果をもたらす例であると解釈出来る。例えば、顔とハート、汗、音符のそれぞれの組み合わせでは顔の表情が表現する感情を補うものとして組み合わせの後者の絵文字が使われていることがわかる。また、顔と手の組み合わせでは両者を同じ胴体に属すると見たとき自然なように表現されている。さらに、車と排気ガスの組み合わせでは、車の進行方向から見て自然な位置に排気ガスが置かれていることがわかる。

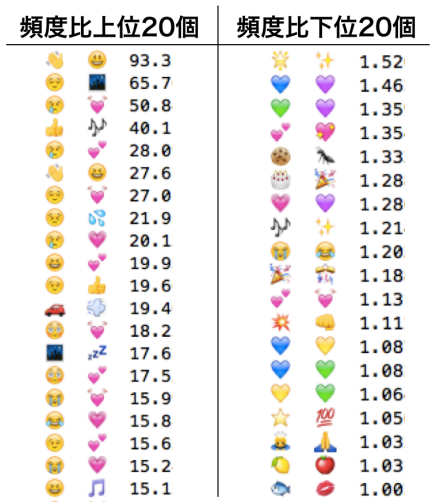


図 5: 順番を入れ替えた場合の 2 グラムの頻度比

### 高度な意味表現

複数の絵文字を使用することによって、個々の絵文字だけでは表現困難な高度な概念を表現する用法が観察された。絵文字 3 グラムの出現頻度上位 100 個の例から、そのような用法を選んだのが図 6 である。図 5、6 は複数個の絵文字の組み合わせが興味深い言語現象であることを示唆している。


図 6: 絵文字の組み合わせによる高度な意味表現