2O3-OS-22b-1

ニューラルネットワークによる顔文字の原形推定

A method to estimate the Original Form of Kaomoji Using Neural Network

奥村 嶺 Rei Okumura
奥村 紀之 Noriyuki Okumura

明石工業高等専門学校 National Institute of Technology, Akashi College
明石工業高等専門学校 National Institute of Technology, Akashi College

In this paper, we describe estimation method of Kaomoji's original forms using neural networks. We constructed a neural network using only Bigram feature and one using combined Bigram feature, Trigram feature, SkipBigram feature. In addition, equilibrium data by restoration extraction is created and the other network evaluated. As an result, using a neural network shows a better tendency than the original estimation method based on the cosine similarity.

1. はじめに

光通信や 4G/LTE 通信などの高速インターネット普及に伴い，オンラインでのコミュニケーションは文字情報だけでなく，画像や音声，動画などのメディアも使用可能なサービスが増えてきている．しかし，依然として SNS 等における文字情報の必要性は高く，LINE でのメッセージのやり取りや Facebook のコメント，Twitter のリプライなど，文字列に依存したコミュニケーションは活発に交わされている．

その一方で，英語やドイツ語のような我々日本人にとっては自明と思われる内容までも明文化して情報を伝達する傾向のある言語とは異なり，日本語でのコミュニケーションでは，空気を読む，雰囲気に合わせるといった，察する能力が求められる．このようなハイコンテクスト（抽象的な表現を好む）言語であるとされる [1] 日本語と文字列のみのコミュニケーションは，両立が困難である．不用意な発言 1 つで炎上などのトラブルに発展してしまうこともある．これは，書き手の表情等の言葉以外の情報が正確に相手に伝わらず，（書き手の意思とは反する）字面通りの解釈をしてしまったり，言葉の奥にあるより深い情報を読み違えたりして起こる問題である．

日本語でのコミュニケーション，特に対面での会話では，話し手は多くを語らず，身振りや表情を交えることで聞き手に言葉以外の情報を伝える．そのため，オンラインでのコミュニケーションでは，文字の情報のみから全ての情報を読取る必要があり，書き手の負担が増大してしまう問題がある．書き手が過不足無く情報を記載することができれば良いが，そのようなことに慣れていない日本語話者に取っては極めて深刻な問題である．

本研究では，テキストベースのコミュニケーションで古くから用いられている顔文字に着目している．現在，顔文字の種類は 10 万種を超えており，今後も増加すると予想される．そのため，顔文字のリストを大量に保持するだけでは不十分であり，未知の顔文字に対するアプローチを見当する必要がある．さらに，顔文字は言語として成立している（品詞等のルールを定めることが可能な）文字列の解析とは異なり，ありとあらゆる記号列が構成要素となる．このことから，すべての顔文字に品詞等のルールを制定することが困難である．そこで，顔文字をある原形から派生したものであるとみなし，付随するパーツの傾向から確率的に顔文字が含んでいる情報を推定することが重要であると考えられる．

推定には，Cosine 類似度による手法や，サポートベクトルマシン（SVM）による手法など，多種多様なアプローチがある．本実験ではニューラルネットワークを用いた手法を扱う．予備実験を元に，無作為に復元抽出した学習データを使用しニューラルネットワークを構成した．さらに，学習データの不均衡を解消するため，復元抽出による均衡データを作成し評価した．

2. 関連研究

本研究に関連して，我々は顔文字へのアノテーションを進めている [2]．顔文字を包括的に扱い，確率的なモデルとして活用する為には，大規模な辞書が必要となる．顔文字は様々な記号から作られる可能性があるため，記号に着目して多元的に分析する必要がある．

また，テキストに含まれる顔文字の抽出に関する研究としては，Ptaszynski らの研究がある [3]．Ptaszynski らの CAO システムでは，顔文字における目-口-目の並び（Triplet）に着目した顔文字抽出を試みている．このシステムでは，1 万種の顔文字にデータベース規模で対応し，90%以上の精度で抽出可能だとしている．さらに自動拡張によって 300 万種に対応可能であると述べられており，これは顔文字の研究の中では非常に規模の大きいものである [4]．同じく，Tanaka らの研究では，機械学習によって顔文字を抽出を試みている [5]．また，文中からの顔文字抽出に関する研究として，Bedrick らの手法がある．Bedrick らは，隠れマルコフモデル (HMM) によって記号列を抽出し，確率的文脈自由文法 (PCFG) に基づく評価法によって顔文字の候補を選別する手法を提案している [6]．また，顔文字の分析に関しては，風間らの研究も参考になる [7]．

本稿での顔文字の原形の定義やそれに関する研究としては，奥村らの研究 [8] がある．また，同じ顔文字の原形を推定する関する研究に関しては，顔文字の原形推定にサポートベクトルマシン (SVM) を用いた研究 [9] が参考になる．

3. 顔文字の原形推定

先行研究 [2] より，顔文字の原形はおよそ 3,000 種類発見されている．また，顔文字と認識された任意の文字列とその原形のペアが 43,733 組分類されている．

連絡先: 奥村 嶺，明石工業高等専門学校，e1371@s.akashi.ac.jp

本稿では，顔文字と認識された任意の文字列（以下「原顔文字列」と略す）から，その顔文字の原形をニューラルネットワーク（Neural Network，以下「NN」と略す）によって推定する検証の進行状況を述べる．

実験環境としては，Python(3.5.2)，Chainer(1.19.0)，CUDA(8.0.44) を用いた．

3.1 Bigram 素性での精度計測

この検証では，予備実験により得られた知見を元に，Bigram 素性での学習を行う．NN の構造として，入力層 65 ユニット，中間層 1 層 10,000 ユニット，出力層 3,110 ユニットを構築した（図 1）．

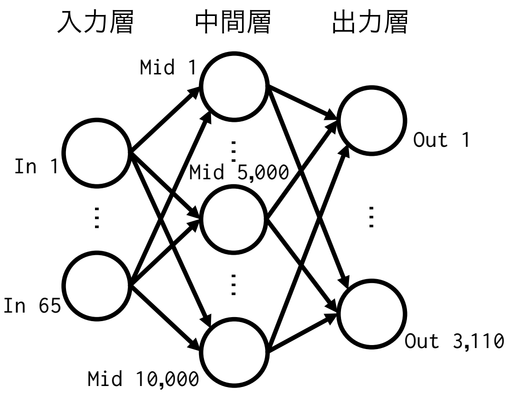

図 1: NN 構造の模式図

入力層が 65 ユニットで構成されている理由として，学習に使用した原顔文字列のうち，最大文字列長が 65 文字であることが挙げられる．また，中間層のユニット数は予備実験により決定している．この検証で使用するデータを以下表 1 に示す．

表 1: Bigram 素性での検証データ概要

データ	内容
学習データ	43,733 組 *1 のデータから復元抽出した 5,000 組のデータを 50 セット
評価データ	43,733 組のデータから復元抽出した 1,000 組のデータを 10 セット

表 1 を参照するとわかるように，学習データは 50 セット用意し，評価として，全ての NN で評価データ 10 セットを検証した．図 2 に各 NN における評価データ 10 セットの分類正解率を箱ひげ図で示す．

図 2 によると，Bigram 素性の分類正解率は，0.178 程度となっている．対して，予備実験では Bigram 素性の分類正解率は 0.692 であったが，これは評価データに学習データと同じものが多く含まれていた結果，高い精度となったと考えられる．

3.2 Bigram 素性，Trigram 素性，SkipBigram 素性複合での検証

この検証では，Bigram 素性，Trigram 素性，SkipBigram 素性 *2，の 3 つを同時に入力として与える．また，使用するデータを表 2 に示す．NN の構造は先の検証（図 1）と同様である．

*1　アノテーション済みの原顔文字列-原形の組

2　1 文字飛ばしの Bigram．例：素性分解前の文字列が (´ω`*) なら (´，*ω，´`，ω*，`)，のように分解される

表 2: 複数素性での検証データ概要

データ	備考
5,000 組の学習データ	43,733 組のデータから復元抽出した 5,000 組のデータ
10,000 組の学習データ	43,733 組のデータから復元抽出した 10,000 組のデータ
15,000 組の学習データ	43,733 組のデータから復元抽出した 15,000 組のデータ
評価データ	43,733 組のデータから復元抽出した 1,000 組のデータを 10 セット

上記の 3 つの学習データをそれぞれ別の NN で学習させる．検証結果を以下表 3 に示す．各分類正解率はそれぞれ上記の評価データ 10 セットの平均である．

表 3: Bigram 素性+Trigram 素性+SkipBigram 素性での学習

学習データ個数	分類正解率
5,000	0.213
10,000	0.312
15,000	0.391

表 3 より学習データの個数が多いほど精度が向上していることが見て取れる．

4. 学習データの偏りを緩和した検証

これまでの検証実験では，43,733 組の原顔文字列と原形のデータを用いた Ngram 要素の選定及び学習による原形推定を行ってきた．しかし，この 43,733 組の原顔文字列と原形のデータには，原形の種類に応じて原顔文字列の個数にばらつきがあり，100 種類以上の原顔文字列がある原形もあれば，1 つの原顔文字列しかない原形もある．このばらつきは，NN の学習データを作成する際に，選択される原形の種類に偏りが生じるような弊害となる．本節では，このばらつきを修正することで，NN での顔文字の原形推定の分類正解率を向上させられるか検証していく．

4.1 ばらつきの解消

この問題を解決するために，3,110 種類ある原形全てが 100 種類の原顔文字列を持つようなデータ（LargeKaomojiData）を作成した．作成にあたっては，全原形それぞれに対し復元抽出を行った．

4.2 学習データ，テストデータ，NN の設定

学習データは，上記の LargeKaomojiData 全てを用いる．評価データは，第 3.2 節で使用した評価データを引き続き用いる．

また，使用した NN は入力層 195 ユニット *3，中間層 1 層 10,000 ユニット，出力層 3,110 ユニットのものを使用している．活性化関数にはシグモイド関数：式 (1) と恒等関数を用い，過学習を防ぐために中間層から出力層の間に dropout を適用した．また，メモリ容量の観点および計算資源の制約によりバッチ勾配降下法は不可能でため，精度の面では不利ではあるが，データを分割して学習するミニバッチ勾配降下法を実施

*3　Bigram 素性，Trigram 素性，SkipBigram 素性の 3 種類を同時に入力するため，$65 \times 3 = 195$ ユニットとなっている

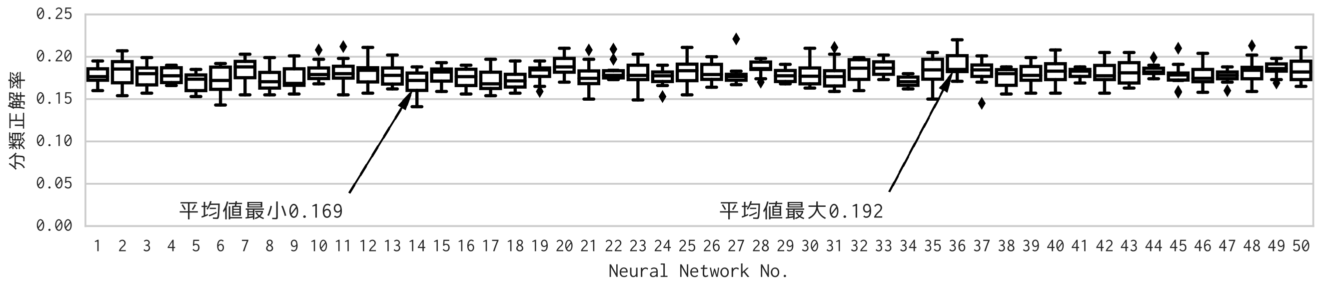

図 2: Bigram 素性での正解率

している．また，Optimaizer が用いる最適化アルゴリズムは Adam を選択している．

$$Sigmoid(x) = \frac{1}{1 + \exp(-x)} \quad (1)$$

4.3　評価

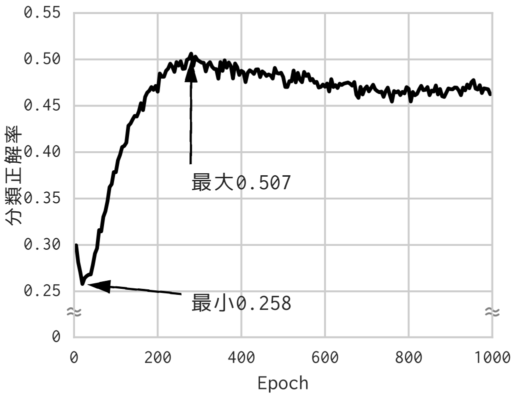

図 3: LargeKaomojiData での学習回数と分類正解率の変化

図 3 に学習回数と分類正解率（10 セットの平均）を示す．このデータからでは，過学習の有無や汎化の度合いを推し量る事はできない．そのため平均分類正解率が最大となる評価データ（最大正解率データ）での傾向調査を行った．表 4 に最大正解率データの傾向を示す．

表 4: 最大正解率データにおける正解数と学習データの関係

NN No.	学習データに存在		学習データに存在せず	
	正解	誤り	正解	誤り
1	496	267	12	225
2	475	318	7	200
3	503	285	13	199
4	508	282	10	200
5	493	261	14	232
6	502	279	15	204
7	476	301	11	212
8	489	294	9	208
9	503	289	11	197
10	501	272	14	213

さらに，より深く NN による原形推定の傾向を調査するため，先行研究 [2] で行われた Cosine 類似度による原形推定との比較を行った．その結果を表 5 に示す．

表 5: Cosine 類似度と NN の正解数の比較

NN No.	両方正解	NN 正解	Cosine 正解	両方誤り
1	272	236	228	264
2	259	223	246	272
3	251	265	209	275
4	277	241	222	260
5	251	256	227	266
6	271	246	205	278
7	250	237	244	269
8	257	241	239	263
9	262	252	225	261
10	245	270	240	245

また，Cosine 類似度と NN の原形推定可能な原顔文字列の方向性を吟味するため，Cosine 類似度と NN の両方で不正解だったデータの一部を表 6 に示す．この表には，原顔文字列，Cosine 類似度による分類，NN による分類，正例の原形に加え，以下に示す 2 つのデータが追加されている．

1. NN により (誤って) 分類された原形の原顔文字列のうち，学習データに含まれているもの（誤分類原顔文字列）

2. 原顔文字列と上記の誤分類原顔文字列との Cosine 類似度

4.4　考察

図 3 より，学習が進むにつれ分類正解率が上昇し，300Epoch ほど学習を進めた時点において最大 0.51 の分類正解率となっている．しかし，その後は分類正解率が僅かに低下し，その値は 0.45〜0.5 の間でほぼ停滞している．このグラフを見る限り，この NN 構成でこれ以上の学習を行うことによる分類正解率の向上は期待できない．

次に，表 4 より学習データに存在しないが正解しているものが極端に少ないことが読み取れ，過学習は起こっていないと言える．しかし，学習データに存在するにもかかわらず，不正解なものが 2 割ほど存在していることから，NN の構成次第ではさらなる精度向上の余地は十分にあると考えられる．

また，表 5 より，各正解数はほぼ均一であることが見て取れ，NN による原形推定は Cosine 類似度による原形推定と同等の精度であることがわかる．

表 6: Cosine 類似度と NN の原形推定可能な原顔文字列の比較

<table>
<tr><th>No.</th><th>NN による分類</th><th>原顔文字列</th><th>Cosine 類似度による分類</th><th>原形</th></tr>
<tr><td>1</td><td>(￣〓￣)</td><td>念力 <<<<<<(￣ w ￣) 人>>>>>> はぁぁぁ</td><td><><></td><td>(￣ w ￣)</td></tr>
<tr><td>2</td><td>(-ë-)</td><td>llllll(-_-;)llllll ずーん</td><td>[l_l]</td><td>(-_-)</td></tr>
<tr><td>3</td><td>(´∀‘)</td><td>ヒューー ストッ || || (|| ´ ρ ` ||) ‖ ‖タダイマァ</td><td>||</td><td>(´ρ‘)</td></tr>
<tr><td>4</td><td>(^.^)</td><td>:*.;” .*・;・^;・:＼(*0￥ 0*)/:・;^・;・*.” ;.*: 素敵♪</td><td>(*0*)</td><td>(0￥0)</td></tr>
</table>

<table>
<tr><th>No.</th><th>原顔文字列（※ 1）</th><th>誤分類原顔文字列（※ 2）</th><th>※ 1 と※ 2 の
Cosine 類似度</th></tr>
<tr><td>1</td><td>念力 <<<<<<(￣ w ￣) 人>>>>>> はぁぁぁ</td><td>念力 <<<<<<(￣〓￣) 人>>>>>> はぁぁぁ</td><td>0.973</td></tr>
<tr><td>2</td><td>llllll(-_-;)llllll ずーん</td><td>llllll(-ë-;)llllll ずーん</td><td>0.984</td></tr>
<tr><td>3</td><td>ヒューー ストッ || || (|| ´ ρ ` || ) ‖ ‖タダイマァ</td><td>ヒューー ストッ || || ( || ´ ∀ ` ||) ‖ ‖タダイマァ</td><td>0.975</td></tr>
<tr><td>4</td><td>:*.;” .*・;・^;・:＼(*0￥ 0*)/:・;^・;・*.” ;.*: 素敵♪</td><td>:*.;” .*・;・^;・:＼(*^.^*)/:・;^・;・*.” ;.*: 素敵♪</td><td>0.920</td></tr>
</table>

さらに，表 6 より，Cosine 類似度による分類では，<や>，l などの，大量に含まれる記号に惑わされて全く見当違いな原形を推定しているのがわかる．しかし，これらは原顔文字列に対応する原形には無関係であり，Cosine 類似度では推定が困難であることを示している．対して，NN による分類では，原顔文字列の原形の推定には失敗しているものの，推定された顔文字は極めて妥当なものだと言える．該当する原顔文字列（※ 1）と誤分類原顔文字列（※ 2）を比較すると，双方は (￣ w ￣) と (￣〓￣) や (-_-) と (-ë-;) などの，顔文字の顔と輪郭部分のみが異なる極めてよく似た文字列であることがわかる．これは，※ 1 と※ 2 の Cosine 類似度の値からも読み取れる．

5. まとめ

本稿では，顔文字の原形推定に関して，ニューラルネットワークでのアプローチの有用性について述べた．まず，予備実験として，学習データと評価データの明確な分離を行わず検証したところ，Bigram 素性を入力として与える推定結果が最も良い分類正解率となり，およそ 69%であった．

次に学習データと評価データを分別し，学習データを 50 セット用意することで傾向調査を行ったところ，分類正解率は 18%という結果となった．

また，複数の素性を組み合わせたり，学習データの偏りを緩和することによって，過学習を抑制しつつ，およそ 50%の分類正解率で類推が可能であることがわかった．

さらに，NN による分類は，先行研究にもある Cosine 類似度による分類と比較しても，同等の精度を保ちつつ，Cosine 類似度による分類が苦手とする llllll(-ë-;)llllll ずーんのような顔文字に対して優勢であることがわかった．NN のによる分類では，Cosine 類似度による分類のように，単に出現回数が多いだけの文字に影響されづらい．つまり，原顔文字列のより広い範囲を俯瞰することが可能であることが伺える．以上より，NN での原形推定の有用性については，十分期待できると言える．

今後は，NN の構成に必要なハイパーパラメータのチューニング実験を軸に，NN による原形推定の精度向上に努めていきたいと考えている．

謝辞

本研究は JSPS 科研費 15K21592 の助成を受けたものである．

参考文献

[1] Edward T. Hall. *Beyond Culture*. Anchor, 1976.

[2] 奥村紀之. 言語解析のための大規模顔文字辞書. 人工知能学会第 30 回全国大会, 2016.

[3] Michal Ptaszynski, Pawel Dybala, Rafal Rzepka, and Kenji Araki. Cao: A fully automatic emoticon analysis system based on theory of kinesics. *Affective Computing, IEEE Transactions*, Vol. 1, No. 1, pp. 46–59, Jan 2010.

[4] Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka, Kenji Araki, and Yoshio Momouchi. Speech, image, and language processing for human computer interaction - science of emoticons: Research framework and state of the art in analysis of kaomoji-type emoticons. *IGI Global*, Jan 2012.

[5] Yuki Tanaka, Hiroya Takamura, and Manabu Okumura. Extraction and classification of facemarks. In *Proceedings of the 10th International Conference on Intelligent User Interfaces*, IUI'05, pp. 28–34, New York, NY, USA, 2005. Association for Computing Machinery.

[6] Steven Bedrick, Russell Beckley, Brian Roark, and Richard Sproat. Robust kaomoji detection in twitter. In *Proceedings of the Second Workshop on Language in Social Media*, pp. 56–64, Montreal, Canada, Jun 2012. Association for Computational Linguistics.

[7] 風間一洋, 水木栄, 榊剛史. Twitter における顔文字を用いた感情分析の検討. 人工知能学会第 30 回全国大会, 2016. 3H3-OS-17a-4.

[8] 奥村紀之. 感情抽出のための顔文字の原形推定. 信学技報, Vol. 116, No. 78, pp. 1–4, 2016. NLC2016-1.

[9] 奥村紀之. 分類器による顔文字の原形推定. 信学技報, Vol. 116, No. 379, pp. 93–96, 2016. NLC2016-37.