

対話システムにおける顔文字の学習

中村純平* 池田剛* 乾伸雄* 小谷善行*

概要

顔文字は文字対話で感情を表現するのに有効である。顔文字は書き手が感情や対話の成り行きから決めるが、その決め方の実際のメカニズムははっきりしていない。本稿では、インターネットでのチャットログなどから対話に適した顔文字を学習する手法を提案する。顔文字の学習モデルは、3階層のニューラルネットモデルによるバックプロパゲーションで行った。各感情パラメータと行動カテゴリを入力とし、各顔文字のパーツ(口, 目, 手, オプション)を出力とした。その結果、学習誤差は0.3となり、クローズドデータに関しては約92%、オープンデータに関しては約50%の許容解を得ることができた。本稿の手法で顔文字のパーツと感情・行動要素の間の関係を獲得できた。

Learning Face Mark for Natural Language Dialogue System

Junpei NAKAMURA* Takeshi IKEDA* Nobuo INUI* Yoshiyuki KOTANI*

Abstract

The face mark is effective for expressing emotions in a text dialogue system. Although each face mark can be determined from histories of emotional elements and activity elements of writers, this real mechanism has not clarified. Hence, we propose a method of learning face mark for a natural language dialogue system using chat dialogue data in the Internet, etc. We use a back propagation error learning of a three-layer neural net to learn a model of face mark. For this neural net, the inputs neurons express emotional parameters and activity categories, and the outputs neurons express parts of face mark that includes mouth, eyes, arms, and optional things. Experimental results showed that, the learning error was 0.3, and we could get the performance approximately 92 % permissible value for the learning set of dialogues and approximately 50 % for the evaluation set of dialogues. It showed that our method expressed the system acquired the information of relationship between parts of face mark and emotional, activity elements well.

1. はじめに

人々が会話をを行うとき、会話の内容の他に表情や声の抑揚などのノンバーバル情報で話者

*東京農工大学 情報コミュニケーション工学科(Tokyo University of Agriculture and Technology Department of Computer, Information and Communication Sciences)

の感情を推測することができる。しかし、文面だけのコミュニケーションである場合、相手の感情を推測するのは難しい。インターネットや携帯電話が普及した現在、メールやチャットなどで、他者と文面だけのコミュニケーションをとる機会が増加した。それに伴って、文字や記

号で自己の感情を表現する工夫がなされた。その工夫の一つが顔文字である。実際に、顔文字は文面だけのコミュニケーションにおける感情の表現に非常に有効であると言える。

一方、人間とコンピュータが対話する対話システムでは、音声対話などでノンバーバル情報を含む対話の研究が行われてきている。しかし、文字だけで行う対話システムも未だ多く存在する。顔文字を用いることで、そのような文字だけで行う対話システムでも感情を表現することができる。しかしながら、顔文字は主に感情とその他の行動的要素で決定され、その使用法は人によって異なる。そこで本稿では、一般的な顔文字生成を行うため、顔文字とそれらの要素との関係を獲得する。そして、対話事例から顔文字を学習し、対話システムの出力文に適した顔文字を決定することを試みた。また、その学習した結果を用いて出力文に顔文字を付加する対話システムを実装した。

2. 対話システムと顔文字

顔文字は、感情的な情報(感性情報)と行動的な情報を含んでいる。つまり、顔文字は感情要素と行動要素で構成されると言える。したがって、対話システムに顔文字を出力させるためには、感情要素と行動要素を持たせる必要がある。本稿では、感情要素を得るために感情モデル[2]を、行動要素を得るために Dialogue Act (以下 DA)[1][3]を用いる。なお、使用する感情は[2]と同じく、喜び、おかしさ、詫び、悲しみ、疑問、怒り、焦り、意味不明、驚き、親しみ、不満、恐れの 12 種類であり、DA は、[1]で定義される挨拶、別れ、意見、意志、事実説明、理由、wh 疑問、yn 疑問、確認、依頼、提案、肯定的返事、否定的返事、熟考、詫び、驚き、感謝の 17 種類である。

3. ニューラルネットによる顔文字の学習

前章で述べたとおり、顔文字は感情要素(パラメータ)と DA などの行動要素(カテゴリ)によって決定される。つまり、式(1)のように表現できる。

$$F \cong f(E, S) \quad (1)$$

F: 顔文字

E: 感情要素

S: 行動要素(DA など)

しかし、式(1)の関数 f は人によって異なり、一概に決定することはできない。したがって、顔文字を含む発話事例から顔文字と感情値・DA の関係を調べる必要がある。本稿では、出力顔文字の決定は感情ベクトル値と DA のパターンマッチングであると考え、ニューラルネット(BP 法)でそれらの関係を学習させた。

3.1 顔文字のパーツ化

顔文字をそのままニューラルネットの出力として割り当てると、出力の数が膨大になってしまうばかりか、教師データに無い顔文字は出力できない。したがって、本稿では顔文字を口、目、手、オプションの 4 パーツに分け、それぞれのパーツを出力とすることで、このような問題を回避した。本稿では、それぞれのパーツを 4 ビットで表現した(表 3.1-1~3.1-4)。

表 3.1-1 口パーツと対応ビット

口	ビット表現	例
	0000	(^^)
o	0001	(*^o^*)
-	0010	(-_-)
-	0011	(^-^)/~
.	0100	(^.^)
◇	0101	(.◇.)

▽	0110	(◦▽◦)
∇	0111	∖(´∇`)ノ
x	1000	(° x°)
^	1001	(^ ^)
○	1010	(TOT)
ω	1011	Σ(◦ω◦)ノ
□	1100	Σ(□ □ ;)
Д	1101	(° Д°)
△	1110	(ノ`△´)ノ
フーン	1111	(´_´)フーン

表 3.1-2 目パーツと対応ビット

目	ビット表現	例
^ ^	0000	(^ ^)/
^ , ^	0001	(^ - ^)
´ , `	0010	(´ Д `;)
◦ , ◦	0011	(◦▽◦)
□ , □	0100	(□ - □)
> , <	0101	(>_<)
::	0110	(; ;)
m , m	0111	m(_)m
- , -	1000	(- -)
T , T	1001	(TT)
? , ?	1010	(? - ?)
@ , @	1011	(@_@)
° , °	1100	(° x°)
ノ´ , ´ノ	1101	(ノ`△´)ノ
¬ , ¬	1110	(¬ _ ¬)
≥ , ≤	1111	(≥ ∇ ≤)

表 3.1-3 手パーツと対応ビット

手	ビット表現	例
	0000	(^ ^)
∖ , /	0001	∖(´∇`)ノ

∖ , /	0010	∖(´-´)
∖ , /	0011	∖(^-^)/
, A	0100	(; ^ ^ A
r ,	0101	r(^ ^;)
, /	0110	(^ ^)/
m , m	0111	m(_)m
, ポイ ^ ◦	1000	(◦◦)ポイ ^ ◦
ノ´ , ´ノ	1001	(ノ`△´)ノ
, ♪	1010	(◦_◦♪
¬ , ¬	1011	¬(´-´)¬
, ^ ,	1100	(; ^ ^) ^ ,
, /	1101	(^ ^)/
ノ´ , ´ノ	1110	Σ(◦ω◦)ノ
, ^ ,	1111	^ (^ ;

表 3.1-4 オプションパーツと対応ビット

オプション	ビット表現	例
	0000	(^ ^)
, * ,	0001	(* ^ * ♪
, ; ,	0010	(° °;)
, ; ,	0011	(; ¬ Д ¬)
, - ,	0100	(- ^ O ^ -)
Σ , ; ,	0101	Σ(□ □ ;)
, (; ,) ,	0110	((; ° Д °))
f ,	0111	f(^ ^)
, ; , ; ,	1000	(^ ^ ; ; ;
, ; , ; , ; ,	1001	(^ ^ ; ; ; ; ;
, ; , ; , ; ,	1010	(; - _ -;)
, フーン	1011	(´_´)フーン
, ◦ O (1100	(´-´)◦ O (
, # , # ,	1101	(#´ o `#)
, zzz	1110	(- -)zzz
, ?	1111	(◦◇◦)?

なお、表 3.1-3、3.1-4 の顔文字パーツにおける“,”は、顔文字の位置情報を表している(図 3.1)。

O1, H1 (O2, H2 ,E1 ,M ,E2 ,H3 ,O3) H4, O4	
M	:口
E1,E2	:目
H1,H2,H3,H4:	手
O1,O2,O3,O4:	オプション
例	口:“ω” 目:“・・” 手:“ , ,ノノ”
	オプション:“Σ, ,; , ”
	→ Σ(・ω・ノノ)

図 3.1 顔文字パーツの位置情報と表現

上記のように、本稿では顔文字をパーツに分けて表現した。このため、図 3.1 の位置情報で再現できない顔文字((・_・)(・_・)フルフル など)や、括弧文字((笑)など)は学習の対象にしていない。

3.2 学習データの準備

学習のための事例データには、インターネットでのチャットやインスタントメッセージャーから収集したログのうち、顔文字が付いているものだけを用いた。また、事例の各発話に話者の感情値を筆者らが推測して付加した。このときの感情値は、なし(0.0)、低(0.3)、中(0.6)、高(1.0)の 4 段階で表現した。感情値の推測は発話文の内容や顔文字を同時に参考にして行った。また、それぞれの発話に DA 情報を付加した。DA 情報は 5 ビットで表現した。最後に、顔文字をパーツに分けビット化した。パーツ分けできないものは、似た顔文字で代用した(表 3.2.1)が、似た顔文字が無いものはそのままテストデータ用に保持した。用意した総事例数は 900 である。事例中の顔文字(括弧文字は

含まない)の使用頻度、パーツごとの使用頻度の上位 5 位までをそれぞれ表 3.2.2、表 3.2.3-1 ~表 3.2.3-4 に示す。なお、顔文字の種類は 119 種であった。

表 3.2.1

パーツ分けできない顔文字と代用顔文字の例

元顔文字	代用	元顔文字	代用
(笑)	(^-^)	(泣)	(T-T)
(爆)	(^-^;;;)	(苦笑)	(^^;;;)
ヽ(´∞`)ノ	ヽ(´-`)ノ	ㄣ(´~` ㄣ	ㄣ(´-` ㄣ

表 3.2.2 事例顔文字使用頻度(上位 5 位)

順位	顔文字	使用回数	割合(%)
1	(^^;	132	14.7
2	(^^;;;)	67	7.44
3	(^^)	50	5.56
4	(--;	24	2.7
4	(;_;	24	2.7

表 3.2.3-1 事例顔文字口パーツ使用頻度

順位	パーツ	使用回数	割合(%)
1	なし	24	20.17
2	—	22	18.49
3	—	16	13.45
3	。	7	5.882
5	.	5	4.202

表 3.2.3-2 事例顔文字目パーツ使用頻度

順位	パーツ	使用回数	割合(%)
1	^ ^	39	32.77
2	´ `	12	10.08
3	・ ・	10	8.403
3	--	10	8.403
5	° °	8	6.723

表 3.2.3-3 事例顔文字手パーツ使用頻度

順位	パーツ	使用回数	割合(%)
1	なし	79	66.39
2	.../	9	7.563
3	.../~	7	5.882
3	\...ノ	4	3.361
5	┌...┐	3	2.521

表 3.2.3-4 事例顔文字オプション使用頻度

順位	パーツ	使用回数	割合(%)
1	なし	74	62.18
2	...;	10	8.403
3	...;	6	5.042
3	...*	5	4.202
5	Σ...;	3	2.521

3.3 ニューラルネットのモデル

本稿で作成したニューラルネットは、入力 17(感情 12個+DA5ビット)ユニット 出力 16(4ビット×4個)ユニット、中間 34 ユニットの 3 層ネットワークであり(図 4.3)、事例データの顔文字パーツビットを教師値とするバックプロパゲーション法[4]で学習させる。

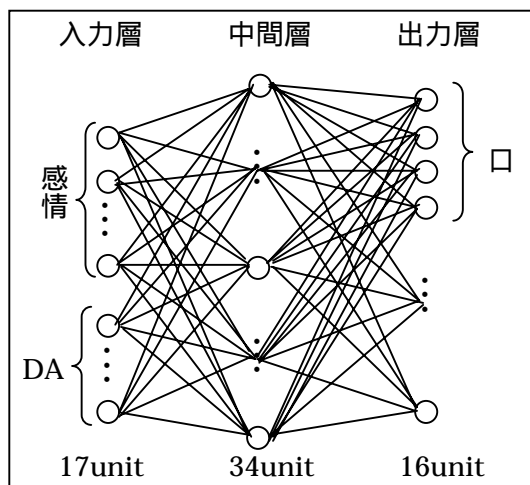


図 3.3 学習に用いたニューラルネット

4. ニューラルネットの学習結果と対話システムによる出力

前章で示したニューラルネットモデルを用いて顔文字の学習を行った。学習データは、3.2 節で作成した事例データから顔文字をパーツ分けできたもの 656 個(ランダムに抽出)を用いた。学習回数は 20000 回で、誤差はほぼ 0.3 に収束した。

学習した重みを用いて、クローズドデータ、オープンデータ、パーツ分けできなかったデータそれぞれに対して顔文字を出力させた。また、ニューラルネットを対話システム[2]に組み込み、実際に対話を行った。以下にそれぞれの結果を示す。

4.1 ニューラルネット学習の評価

クローズデータ(656 個)、オープンデータ(129 個)、パーツ化できないデータ(38 個)に対する出力を次に示す(表 4.1)。顔文字の評価は、教師と完全に一致したもの、顔文字は異なるが、意味的に問題がなく許容できるもの、顔文字が近い・印象が似ているが、許容はできない(感情的な意味合いが異なる)もの、まったく似ていないものの 4 つに筆者らが分類した。

表 4.1 学習結果

	クローズ		オープン		パーツ化できない	
	数	率	数	率	数	率
完全一致	513	78.2	13	10.08		
許容	90	13.72	49	37.98	20	52.63
近い,似た印象	30	4.573	40	31.01	8	21.05
似てない	23	3.506	27	20.93	10	26.32
計	656		129		38	

クローズデータの場合は、完全一致が 78% であり、許容できる顔文字を含めると、約 92% が適切な顔文字が出力されていることがわかる。オープンデータの場合は、完全一致は 10% と少なかったものの、許容できる顔文字を含めると約 50% が適切な出力ができています。パーツ化できないデータは、許容できる顔文字が 52% もあり、良好な結果であると考えています。

4.2 対話システムでの出力結果

感情モデルを含んだ事例ベース自由対話システム[2]にニューラルネットワークを組み込んで実際に動作させた。次にその一部を示す(図 4.2-1~4.2-4)。なお、図中の“<あなた>”で始まる行がユーザの発話で、“<カシス>”で始まる行がシステムの発話である。

```

<あなた> こんにちは (一一)
<カシス> じゃあ、明後日ね (一▽)
<あなた> お、できたね (^;
<カシス> えっと、どういうこと? Σ(° 皿° ;)(1)

```

図 4.2-1 事例にない顔文字の出力

```

<カシス> さぼったら友達無くすんじゃない? (^)
<あなた> 余計なお世話だボケが(° 皿°)
<カシス> お前の方がアホじゃー (;_>;) (2)
<あなた> 何だその顔は? (^;
<カシス> コンピュータです。(° ▽° ;)(3)
<あなた> いい返事だ(^▽^)/
<カシス> 私はコンピュータです。(;-_-;)
<あなた> さっきも言ったじゃん(一一;
<カシス> (;_-_-;) (4)

```

図 4.2-2 顔文字だけの出力

```

<カシス> 乗物が楽しかった。(^-^ポイ^°) (5)
<あなた> 怒っているのか...?(一□一;)
<カシス> 死ぬ。(^-^ポイ^°) (6)
<あなた> うわー...お前はひどい奴だなあ...
Σ(一□一;)
<カシス> (-ノ_>一)/ (7)

```

図 4.2-3 誤学習の出力

```

<カシス> 分からないなら別にいい。(一_>一)
<あなた> 勝手な奴だなあ(一一;
<カシス> 馬鹿は頭が良いのですか?(一一一) (9)
<あなた> 馬鹿は馬鹿だから馬鹿なのです。
(一一一)
<カシス> 大好き。(;^▽^;) (10)
<あなた> ありがとう\(^ ^)
<カシス> いえいえ(一▽)

```

図 4.2-4 顔文字としてふさわしくない出力

対話システムの実行結果を見ると、図 4.2-1 の(1)のような事例に存在しない顔文字を出力することがある。そのような顔文字は、ほとんどの場合が顔文字として成立しているが、中には図 4.2-4 の(9)のように顔文字なのか一目ではわからないようなものも出力されてしまった。図 4.2-2 の(2)などもよく分からない顔文字であると言える(暗黙のルールではないが、一般的に“;”を目(涙目)として使用した場合、汗を表す“;”は使用されない)。加えて、図 4.2-3 の(7)のように、顔文字としては成立するが、何を表しているのかよくわからない顔文字も出力された。また、図 4.2-3 の(5)のような誤学習も見受けられた。ただし、(6)のような使われ方の場合は、顔文字の使い方として間違っているとは言えない(笑いながら怒っている状態であると推測できる)。

従来の対話システムは、ユーザの入力に対する応答として、文章を返す。そのため、文章を

返さないという応答はありえない。しかし、顔文字が使用されるとき、顔文字だけの発話というものが頻繁に見かける。そこで、本稿での対話システムでは、敢えて文章を返さない応答を持たせ、顔文字だけを応答として返すようにした。そのときの顔文字は、「DA なし」という DA カテゴリと感情値から決定される。図 4.2-1～4 中の(4)、(7)がそれにあたる。

また、ランダムに顔文字を付加した例を次に示し(図 4.2-5)、比較する。

<あなた>	こんにちは(^▽^)/
<カシス>	おはようございます。今日はいい天気ですね。 (□□;))
<あなた>	まあ、そうですね(^_^)
<カシス>	クリスマスプレゼントはどんなのが欲しいの? (・ー・)/
<あなた>	別に・・・(ー;)
<カシス>	アホ(・_・)
<あなた>	なんだと(°Д°)
<カシス>	ここです。(ノ△^)/ ナニー
<あなた>	なにが? Σ(□□;))
<カシス>	嫌いな人とも付き合わなきゃいけないんだ。m (_)m
<あなた>	そうかもね(・_・)
<カシス>	もうね、アホかと。バカかと。(・_・)/ ハイ

図 4.2-5 ランダムな顔文字出力例

5. 考察

本章では次の事柄を考察する。

- ・対話事例から事例データへの変換作業
- ・ニューラルネット学習の重みと出力結果
- ・対話システムとの融合による利点・欠点

5.1 事例データ作成についての考察

対話事例を事例データに変化する作業はほぼ手作業で行った。特に、感情値の付加は人手

でやらざるを得なかった。なぜなら、話者の感情を推測するには、顔文字から感情を推測することの他に、発話の内容やその背景を理解しなければならないからである。しかし、人手でやるということは、それだけミスが生じやすいとも言える。今回の事例データの全てが正しいとは言いがたい。例えば、図 4.2-3 の(5)、(6)で使われている“ポイ °”という手パーツは、怒りの感情以外では使われることはない。もし、ニューラルネットによる学習が局所解に陥ってなければ、事例データの方が間違っていると推測するしかない。今後何らかの方法でミスの少ない事例を作成もしくは収集する必要がある。

5.2 学習結果についての考察

前章で示したとおり、クローズドデータでは約 92%、オープンデータでは約 50%の有効な出力を得た。しかし、実際に文章と共に出力された場合は、図 4.2-3 の(6)のように文の内容次第で適切と判断することもできる。その点をふまえると、この結果は数値以上に良い結果なのではないかと思っている。また、オープンデータの完全一致が約 10%であり、許容な顔文字が約 40%であることから、少しでも入力が異なれば違う顔文字を出力しやすいということがわかる。したがって、本稿で提案した手法は発話文に適した顔文字を出力することにおいて有効であると言える。

しかしながら、顔文字が感性情報を含んでいるため、統計的な方法による精度の向上は難しく、いたずらに事例を増やすだけでは効果はないと考えている。

5.3 対話システムの出力の考察

顔文字を対話システムの入出力に付加することで、人間味溢れた対話を実現できる他、文

章からではわからないことをイメージできる。例えば、図 4.2-2 の(3)である。もしこれが文章だけであつたら、ただ事実を述べているだけでも感じられる。しかし、“(° ▽° ;)” という顔文字が付加されることで、皮肉たっぷりの応答をしているように感じられる。また、図 4.2-3 の(6)のように、発話文と異なった印象を与えることもできる。

一方、顔文字として成立していない出力や、意味のわからない顔文字の出力なども見受けられる。これは、感情の値と各パーツとの関係が一部で正確に獲得できていないことを意味している。この問題を解決するためには、感情モデルの改善と、よりミスが少ない事例による学習のやり直しが必要であると考えている。その他にも、ニューラルネットモデルの改良(4層構造にする、中間層の数を変化させるなど)を行うことで改善される可能性もある。しかし、システムとの対話ログを全体的に見ると、明らかに不適切な出力というのは少なく事例にない顔文字も多く出力されていた。また、ランダムと比較すると明らかに出力文章とあつた顔文字が出力されていることがわかる。

最後に、顔文字だけの出力に関する考察を述べる。確かに、顔文字だけの文章を出力することによって、より人間らしい応答ができるようになった。しかし、対話システムは事例ベース自由対話システムであり、システムとの発話は事例として保持される。したがって、システムが文章では何も応答しないという事例が増えてしまい、顔文字だけの応答が増える可能性もある。本システムを使って対話事例を収集する場合には、システムの応答は事例データとして使わないなどの対策が必要であろう。

6. おわりに

本稿では、対話システムで有効な顔文字を出力するために、顔文字付き対話ログから顔文字を学習する手法を提案し実装した。その結果、クローズドデータで約 92%の適した顔文字を出力し、オープンデータにおいても約 50%の適切な出力を得た。また、感情モデルを含む対話システムに組み込むことで、システムの感情を柔軟にユーザに伝えることができ、事例に存在しない顔文字をも出力することができた。

参考文献

- [1] 江部利明：Dialogue Act 情報を用いた事例ベース自由対話システム，東京農工大学大学院工学研究科 2000 年度修士学論文，2001
- [2] 中村純平 他：感情モデルと DialogueAct を用いて自由対話システムの応答文に合った顔文字を出力する手法，情報処理学会第 64 回全国大会(2)，pp.25-26,2002
- [3] Andreas Stolcke, Elizabeth Shriberg 他：Dialogue Act Modeling for Conversational Speech, 1998 AAAI Spring Symposium, AAAI Press, pp.98-105, 1998
- [4] 安西祐一郎：岩波講座ソフトウェア科学 16 認識と学習，岩波書店，1989