九州大学学術情報リポジトリ
Kyushu University Institutional Repository

感情Tagを用いた感情学習及びLSTM,GRUの比較実験：デスクトップマスコットのAIエージェント化に向けて

弓場, 邦哲
九州大学大学院システム情報科学府情報知能工学専攻

伊東, 栄典
九州大学情報基盤研究開発センター

https://hdl.handle.net/2324/2552966

出版情報：電子情報通信学会技術研究報告. AI, 人工知能と知識処理. 119 (317), pp.31-36, 2019-11-21. The Institute of Electronic, Information and Communication Engineers
バージョン：
権利関係：利用は著作権の範囲内に限られます

一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
IEICE Technical Report

感情 Tag を用いた感情学習及び LSTM,GRU の比較実験
－デスクトップマスコットの AI エージェント化に向けて－

弓場　邦哲†　　伊東 栄典‡

† 九州大学システム情報科学府　〒819-0395 福岡市西区元岡 744
‡ 九州大学情報基盤研究開発センター　〒819-0395 福岡市西区元岡 744
E-mail:　† k.yuba.615@s.kyushu-u.ac.jp,　‡ ito.eisuke.523@m.kyushu-u.ac.jp

あらまし　対話プログラムは、GRU やサーバーによる学習の普及により、大きくその精度を上昇させた。しかし、その個性は限られており、ユーザー毎に好みのプログラムを提供するサービスは存在しない。そこで、ユーザーごとに好みのキャラクターで作成されていたデスクトップマスコットに対話技術を組み込むことで解決を図ることにした。その手始めとして、感情学習と対話技術で多用される LSTM と GRU の比較実験を行った。
キーワード　デスクトップマスコット,感情学習,GRU,LSTM,Seq2Seq

Emotion Learning from Emotion Tags and Comparative Experience with LSTM and GRU
－Heading toward Sublimation from Desktop Mascot to AI Agent－

Kuniaki YUBA†　and　Eisuke ITO‡

† Graduate School of ISEE, Kyushu University　744 Motooka, Nishi-ku, Fukuoka, 819-0395 Japan
‡ Research Institute for IT, Kyushu University　744 Motooka, Nishi-ku, Fukuoka, 819-0395 Japan
E-mail:　† k.yuba.615@s.kyushu-u.ac.jp,　‡ ito.eisuke.523@m.kyushu-u.ac.jp

Abstract Dialogue Programs have improved their performances by using GPU and Servers. But, their characters are low diversity and there aren't services to create their programs for users. So, we pay attention to Desktop Mascots that were created for users and high diversity and try to solve this problem by incorporating dialogue technology into these Desktop Mascot.
As a starting point, we decided to conduct Emotional learnings and Comparative Experiences LSTM and GRU.
Keywords　Desktop Mascot, Emotional Learning, LSTM, GRU, Seq2Seq

1. 概略
1.1 背景
デスクトップマスコットという旧時代のデスクトップアプリが存在する。かつて一時期はやったものだが、スマートフォンの台頭などにより衰退しつつある。
また、ChatBot という技術が存在する。これはサーバーや GPU を用いて会話機能を学習することで、高速学習を行うのが一般的であり、端的に言えば一部の人間にしか作成できない。また、そのエージェントの口調や性格も悪く言えば、当たり障りのない性格口調に設定されており、多様性に欠いている。一方でサーバーや GPU での学習によるものには及ばないものの、CPU で学習可能である。では、学習にかける時間が長くない Bot が存在すると誰でも作りやすい対話 Bot ができるのではないか。
後述するデスクトップマスコットは GUI を持つ。そして、簡単な事務機能を持つこともある。これに既存の AI 技術や ChatBot 技術を導入することで、Text ベースではない、”身体”や”表情”を持った AI エージェントとしてデスクトップマスコットを昇華することができよう。
本発表は、その足掛かりとして、
①デスクトップマスコットの表情を変化させるための簡単な感情学習手法の探求
②すべて LSTM で実装していたものを LSTM より計算数の少ない GRU に組み替えたことで、学習時間、学習精度がどう変化したかの比較
この 2 つを実行することを目的としたものである。
第 2 章では先行研究・既存技術について説明し、第 3 章では感情 Tag を用いた感情学習について述べる。第 4～6 章では、今回行う GRU と LSTM に RNN セルを用いたときの学習及び学習結果の差についての比較実験について内容とその結果を述べる。
よって第 1 章 2 項では、背景やこの後述べる内容に

This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.
Copyright ©2019　by　IEICE

ついての前提知識を記述する。

1.2 デスクトップマスコット

2000～2010 年代にかけて、文書作成ソフトや PC ゲームの付録、商用 PC ソフトなど、PC(主に Windows)業界において、発展そして、スマートフォンの普及により、衰退していったデスクトップマスコットというものがある。

デスクトップマスコットは、明確な定義がない。参考として、ACG（これ以降アニメ・コミック・ゲームのことをさす）業界の単語・用語について、ユーザーが自由に編集して情報を更新されているニコニコ大百科には、次のように定義されている。

“デスクトップマスコットとは,パソコンやモバイル機器の表示画面（デスクトップ）上に常駐する、様々な形状（漫画・ゲーム・アニメのキャラクターの形など）のウィンドウを利用したアプリケーションである“[1]

本研究で目指すデスクトップマスコットについて述べる。外部から情報(ニュースや天気等)を得ることを除き、スタンドアローンで動く。画面上でキャラクターの画像が表示され、吹き出しなどにキャラクターのセリフが表示される。また、開発環境として、windows のデスクトップアプリとして機能、および学習を行うため、WPF (Windows Platform Foundation) と C#言語でモデルの学習、プログラムの動作を開発した。

2. 先行研究及び既存技術

本論文で使用する単語を以下の通りに定義する。

(1)対話データ

ユーザーの発話(Utterance)とプログラム側の応答(Response)を 1 組の対話とするデータ。

(2)基礎学習コーパス

応答した人物及びキャラクターが固定されていない対話データを集めて作成された、基礎的な応答を学習するための対話データ群。

(3)転移学習コーパス

応答文を特定のキャラクターや人物だけに絞った対話データのみを集めた、特定の場面での応答や口調の取得等、転移学習に使用するための対話データ群。

2.1 対話学習モデル

2.1.1 Seq2Seq

Vinyels らが提案した Sequence to Sequence (Seq2Seq) [2]は大きく Encorder RNN と Decorder RNN の 2 つの部分によって構成されている。Encorder に翻訳する前の文章、Decorde に翻訳後の文章を時系列データとして学習させることで、機械翻訳で使用されていたが、Vinyals らにより一問一答形式の会話でも使用可能であると報告される [3]と Chatbot などに使用されるようになった。

さらに、Seq2Seq は、Encorder の入力文の最初の部分が反映されづらいために、長文に弱いという欠点を持っていたが、Bahdaunau ら [4]が Attention モデルを挟むことでその解決を試みた。

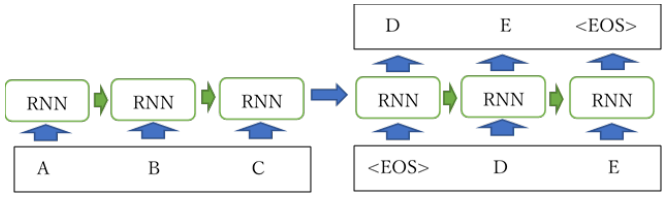

図 1　Seq2Seq

2.1.2 HRED

HRED は Seq2Seq モデルの一問一答形式では解答可能であるものの、文脈を考慮して解答ができないという欠点がある。これを克服するために、Encorder RNN から Decorder RNN に渡す間に、Context RNN を挟むことで、文脈を考慮した解答ができるように Serban らが改良したモデルである。 [5]

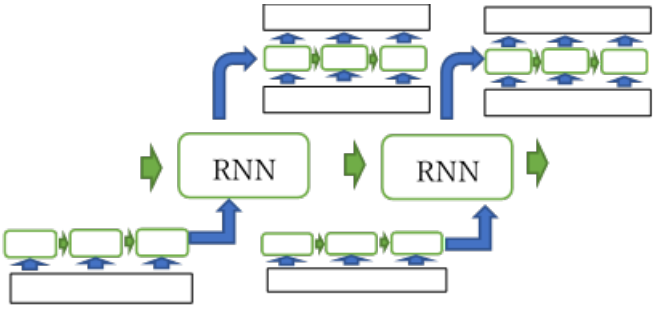

図 2　HRED

2.1.3 VHRED

HRED や Seq2Seq は一つの Utterance に対して、ひとつの応答しか返せないという欠点が存在した。

そこで Serban らはさらに、Context 層に確率的なノイズ(潜在変数)を加えることで、多様な出力を可能とする Latent Variable Hierarchical Recurrent Encoder-Decorder (VHRED) を発表した。 [6]

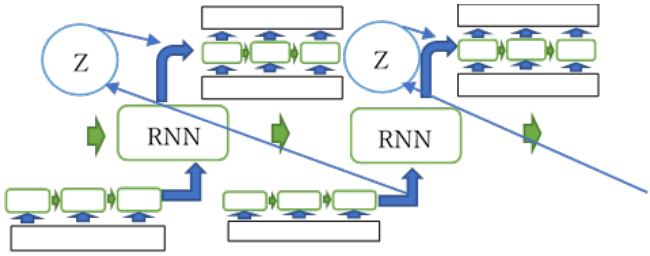

図 3　VHRED

2.2 口調・個性の取得

Seq2seq モデルの応答時の個性を付与する手法を述べる。赤間らは、基礎学習コーパスで学習し、Seq2Seq モデルの単語辞書から基礎学習コーパスの低頻出単語と転移学習コーパスの高頻出単語を交換する。そのうえで、転移学習コーパスで転移学習を行う手法を提案している**エラー! 参照元が見つかりません。**。実験結果より、転移学習に使用するコーパスが少なくともその効果は十分に得られることが報告されている。

また VHRED での文脈に対応した応答取得法として、和田らは、基礎学習データを Seq2Seq モデルで学習したうえで、その Encorder,Decorder モデルの値をそのまま用い、転移学習コーパスを転移学習させる際に Context 層を学習させる手法を提案している。 [7]

3. 実験 1 ・感情学習

デスクトップマスコットの GUI を持つという text ベースの bot には存在しない能力を引き出すうえで、文章に応じて、表情差分を変更することで、プログラムが表情豊かに表現しているように動作させる必要があった。

そこで対話データのプログラム側の応答文の文末に“<嬉>”といった感情を示す Tag を付与し、感情 Tag ごと学習させることで、感情の発露を学習させることを試みた。なお転移学習には赤間らの手法を用いた。単語データは基礎学習コーパスの学習時、基礎学習コーパスに含まれる単語の 1 回のみコーパス内に存在する単語のみを除去したものを使用し単語 ID を付与し、転移学習時は上記単語に含まれない転移学習コーパスの単語と上記の低頻出単語をすべて入れ替えた。

またデータの前処理として Mecab [8]を用いて、文章の分かち書きを行った。その際、人名名詞が連続する際は、人名名詞の一部分が通常名詞として分割学習されるのを防ぐために結合して一つの名詞としてデータを作成した。そして、発話文について、“。” や “!” などの記号は SMS やチャットを行う際、省略されることが多い。このため記号は省略をした。

3.1 使用データ・モデル

(1)モデル

表 1　モデル構成表

層	詳細
Encorder 層	3 層 Bidirectional LSTM
Decorder 層	3 層 UniDirectional LSTM
入力次元数	128 次元
隠れ次元数	100 次元
Attention	Self-Attention

(2)学習コーパス

学習データは、雑談対話コーパスとゲームのイベント文を用いて作成した。

表 2　学習コーパス

コーパス	文章数
基礎学習コーパス	4000 文
転移学習コーパス	500 文

また感情 Tag の分散は以下の通りになった。

表 3　学習コーパスの感情 Tag 分布

data	共感	嬉	感動	困惑	照れ	不満
基礎	81.4%	7.3%	2.4%	7.2%	0.8%	0.7%
転移	33.3%	30.0%	12.4%	16.5%	4.3%	1.3%

基礎学習コーパスは 3500 文を雑談対話コーパスで作成したために、<共感>が多くなってしまった。一方、転移学習コーパスは特定のキャラクターのイベント文から作成したために、比較的に感情 Tag は分散した。

(3)評価・実験結果

評価は、4 人の被験者に基礎学習コーパスでのみ学習したもの(モデル 1)と転移学習コーパスで転移学習コーパスを学習したモデル(モデル 2)の 2 モデルをキャラクターの再現性(口調)、感情の付与の正当性(感情)、対話・成文の妥当性(対話)の 3 点について、5 段階評価で評価をしてもらった。

評価結果は以下のとおりである。

表 4　モデル評価

Model	感情	口調	対話
モデル 1	2	3	2.25
モデル 2	3.5	4	3.5

①対話

モデル 1 は棄却率が低いものの、棄却された単語が応答結果に<unknown>として表示されることが多く、結果として対話・成文の妥当性が低くなってしまった。

一方、モデル２は棄却率が 31.3%にもかかわらず、<unknown>はほとんど表示されなかった。

表 5　単語数及び棄却数

モデル	総単語数	棄却数	棄却率
モデル１	3286	572	14.8%
モデル２	3286	1499	31.3%

②感情

モデル１の学習データは、<共感>のデータが占める割合が高く、感情 Tag も<共感>が高頻度で出力されていたために、感情の評価が低くなった。

モデル２は、<共感>の占める割合が低く、分散しているために、感情の評価も上がった。

③口調

赤間らの実験結果と通り、口調の取得はうまくいったと見受けられる。また、感情の表出も口調の一端として学習したために、感情の評価も向上したと受け取ることができる。

4. 実験２・RNN セルの比較

4.1 実験内容

次に 3 章の転移学習コーパスの文量を 36 文増やし、データを 2 倍にした以下のコーパスを用いて， GRU と LSTM によるモデルを作成した。学習に用いた文章を流し込んだ時の出力データと元データとのレーベンシュタイン距離を用いて両者を転移学習後に学習データと比較する。

表 6　学習コーパス

コーパス	文章数
基礎学習コーパス	4000 文
転移学習コーパス	1074 文

4.2 使用モデル

(1)モデル L(LSTM)

表 7　モデル L 構成表

データ	詳細
Encorder 層	3 層 Bidirectional LSTM
Decorder 層	3 層 UniDirectional LSTM
入力次元数	128 次元
隠れ状態の次元数	100 次元
Attention	Self-Attention

(2)モデル G(GRU)

表 8　モデル G 構成表

データ	詳細
Encorder 層	3 層 Bidirectional GRU
Decorder 層	3 層 UniDirectional GRU
入力次元数	128 次元
隠れ状態の次元数	100 次元
Attention	Self-Attention

なお GRU 層の Decorder 層は,Cho らの GRU の実装[9]をもとに、Forget Gate の R[t](U[h]h[t-1]+C[h]c)を以下のように変更した。なお◉は要素積、W,U は重みを示す。

Recet Gate:

$R[t]=\sigma(W[r]X[t]+U[r]h[t-1]+C[r]c)$

WriteGate:

$Z[t]=\sigma(W[z]X[t]+U[z]h[t-1]+C[z]c)$

ForGet Gate:

$H[t]=Tanh(W[h]X[t]+\underline{R[t]◉(U[h]h[t-1])+C[h]c)}$

Output Gate:

$h[t]=Z[t]h[t-1]+(1-Z[t])H[t]$

5. 実験結果

5.1　学習時間

4000 文を 5 回基礎学習した。単語を辞書に登録してから 5 回目の学習後、モデルを保存するまでの学習時間は以下のようになった。

表 9　学習時間

モデル	5 回総計	1 回平均
モデル L	4h54m	59m
モデル G	4h27m	53m

一方、結合やクロス算など計算回数は LSTM が 25 回、GRU は 23 回であることから、計算回数に依拠していると思われる。

表 10　計算ブロック数

RNN	Tanh	σ	＋	×	◉	1-
LSTM	2	3	9	11(8)	0	0
GRU	1	2	9	9(6)	1(1)	1

()内は重みとの計算回数

5.2 学習精度

テスト用の 30 文を用いて正答率についてレーベンシュタイン距離を用いて測定した。結果は以下に示す。

表 11　学習精度

	LSTM	GRU
正答率	30%	23%
平均距離	9.43	12.9
文意が通る	6	4

LSTM が GRU より精度が良い。また、検査をしていくうえで、惜しい文（テストデータとは答えが違うものの、文意が通るもの）が存在した。惜しい文の数も LSTM が GRU を上回った。

5.3 モデルサイズ

モデル自体のサイズを以下に示す。

表 12　モデルサイズ

モデル	サイズ
モデル L	42,137kB
モデル G	36,786kB

LSTM1 セルの総次元数は 91600 次元、GRU1 セルの総次元数は 68700 次元である。保存している次元数の差がサイズに反映されたのだろう。

6.考察

Seq2Seq の発展形とされる HRED や VHRED では LSTM の代わりに GRU がよく使用されている。そこで LSTM と GRU との学習精度差を確認してみた。計算ブロック数が多い LSTM の方が精度は良いようである。しかし精度を上げるために学習数を増やすとなると第 5 章第 1 項の学習時間より GRU を用いる方が適切だろうか。

７．まとめ・展望

今回はデスクトップマスコットの AI エージェント化のために用いる技術の基礎実験として、感情 Tag を応答文の文末に代入して学習する感情学習、そして LSTM と GRU の 2 RNNCell を用いた比較実験を行った。

GRU は LSTM に比べて精度が少し劣るが、その分、パラメーターの出入力数や学習時間、モデルの大きさなど組み込みで扱う上で有利な点も存在する。今回は実装できなかったが、GRU によって作成した HRED や Attention を Self-Attention ではなく、Transformer [10]などで使用される MultiHead-Attention に交換することで、会話の精度を GPU に頼ることなく、上昇できるように努力したい。

またデスクトップマスコットを AI エージェント化するには、対話技術以外にも事務処理命令を対話と分離して処理する必要が存在する。よって、既存の文書分別アルゴリズムなど他にも実装しなければならないものが、多々ある。複数の機械学習やディープランニング技術を織り交ぜて、ユーザーに寄り添うユーザー専用のデスクトップマスコットの作成に尽力したい。

文　献

[1] “デスクトップマスコットとは(デスクトップマスコットとは)[単語記事]-ニコニコ大百科,” ドワンゴ, 16 1 2009. [オンライン]. Available: https://dic.nicovideo.jp/a/デスクトップマスコット. [アクセス日: 7 11 2019]

[2] Ilya Sutskever,Oriol Vinyals,Quoc V. Le, “Sequence to Sequence Learning,” Proc. of EMNLP’14, pp. 1724-1734, 2014.

[3] Quoc V. Le, Oriol Vinyals, “A Neural Conversational Model,” arXiv:1506.05869, 2015

[4] Dzmitry Bahdanau, KyungHyun Cho, Yoshua Bengio, “NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE,” arXiv:1409.0473, 2016.

[5] Iulian V. Serban, Alessandro Sordoni, Yoshua Bengio, Aaron Courville, Joelle Pineau, “Building End-To-End Dialogue Systems,” Proc. AAAI’16, pp.3776-3783, 2016.

[6] Iulian Vlad Serban, Alessandro Sordoni, Ryan Lowe, Laurent Charlin, Joelle Pineau, “A Hierarchical Latent Variable Encoder-Decoder Model for Generating Dialogues,” Proc. AAAI’17, pp. 3295-3301, 2017.

[7] 萩原将文, 和田翔, “転移学習を用いた階層型潜在変数付きエンコーダ・デコーダによる自動相談システム,” 日本感性工学会論文誌 Vol.18 No.4, 2019.

[8] MeCab: Yet Another Part-of-Speech and Morphological Analyzer, 2013.

[9] Kyunghyun Cho, Bart van Merrienboer Caglar Gulcehre,Dzmitry Bahdanau,Fethi Bougares,Holger Schwenk,Yoshua Bengio, “Learning Phrase Representations using RNN Encoder-Decoder,” 2014

[10] Ashish Vaswani, Noam Shazeer, Niki

Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, “Attention Is All You Need,” 2017.

[11] 赤間 玲奈, 稲田 和明,小林 颯介,佐藤 祥多,乾 健太郎, “転移学習を用いた対話応答のスタイル制御,” 言語処理学会 第 23 回年次大会, 2017.