# 聞き手の感情を喚起する発話応答生成

長谷川 貴之 †　鍛治 伸裕 ‡　吉永 直樹 ‡　豊田 正史 ‡
† 東京大学 大学院情報理工学系研究科
‡ 東京大学 生産技術研究所
{hasegawa, kaji, ynaga, toyoda}@tkl.iis.u-tokyo.ac.jp

## 1　はじめに

我々は常日頃から相手の感情を意識しながら発話を行っている．例えば，落ち込んでいる相手には励ましの言葉をかけるよう意識したり，一般的に相手の感情を害するような発言は避けるよう意識したりしているはずである．そのため，計算機による発話生成においても，聞き手の感情をモデルに組み入れることは自然な発話を生成する上で重要であると考えられる．しかしながら，これまで発話生成という分野自身が依然として未成熟であるという事情もあり，発話生成において聞き手の感情をどのようにモデル化すべきかという問題に対して，十分な知見の蓄積は行われていない．

本論文では，与えられた発話に対し，その発話者に喜びや悲しみのような特定の感情（目的感情）を喚起させる応答を生成するタスクを提案し，それを実現するための方法を議論する．例えば，目的感情を喜びとした場合,「3 日間高熱が出ているの」という入力に対しては「すぐに良くなるといいね」のような，また悲しみとした場合は「風邪が治るまで来ないでね」のような応答を生成することを目標とする (図 1)．このようなタスクは対話システムのモジュールとしてだけではなく，オンラインコミュニケーション環境における予測入力システムの開発 [1] につながると期待される．

我々は上記のタスクに対して，Ritter ら [2] が提案した統計的手法をベースとする手法を適用する．Ritter らは統計的機械翻訳の枠組みを応答生成に応用し，応答（図 1 の右側の発話）を入力発話（図 1 の左側の発話）に対する翻訳とみなし応答生成器を学習している．我々も基本的には Ritter らと同様のアプローチを採用するが，特定の感情を喚起する応答生成という問題を，翻訳におけるドメイン適応の問題とみなし，各感情に適応させたモデル学習を行う．モデルの学習には，SNS 上の対話データの 1 つである Twitter のツイートに対して半自動的に感情タグを付与することにより構築した大規模な訓練データを用いた．

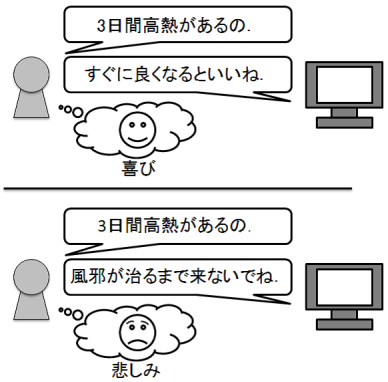


図 1: 与えられた発話に対するシステムの応答例．これらの応答はそれぞれ，喜び，悲しみを発話者に喚起している．

評価実験では，30 億ツイートから構築したコーパスを利用して応答生成器を学習し，作業者によって作成したテストデータで評価した．感情を考慮せずに学習したモデルに対して特定の感情に特化したモデルを適応させることにより，良い結果となることを示す．

## 2　感情タグ付き対話コーパス

教師あり学習を適用する上では，どのように大規模かつ信頼性の高い訓練コーパスを作るかが問題となる．本節では，SNS から大規模な感情タグ付き対話コーパスを構築する方法について述べる．はじめに，対話コーパスのソースとして，Twitter から発話を収集する．次に，感情表現を手がかりとすることで発話に発話者の感情を自動的にタグ付けする．

まず Twitter REST API[1] を利用することにより発話（ツイート）をクロールする．クロールされたデータは，2011/3〜2012/5 までの期間に，約 70 万ユーザーによってツイートされた 30 億の発話で構成されてい

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

| 感情 | 感情表現 |
|---|---|
| 怒り | いらいら，腹が立つ，怒る |
| 期待 | 楽しみ，期待する，わくわく |
| 嫌悪 | 嫌う，うんざり，不快 |
| 恐れ | 恐れる，心配，怖い |
| 喜び | 嬉しい，幸せ，感激 |
| 悲しみ | 悲しい，寂しい |
| 驚き | 驚く，びっくり，Σ（・□・;） |
| 受容 | 安心，頼りになる，ほっとする |

表 1: 感情表現の例

る．次に，Twitter 特有の表現を削除，または変更することで，クロールされた発話をクリーニングする．具体的には，引用を示すマークである RT や QT を含むツイートは削除し，URL の文字列を含むツイートでは URL 文字列を ‘URL’ に置き換え，@user_name は削除する．

次に，前処理した発話から 2 ユーザーによる連続した応答のやりとりを対話として抽出する（表 2）．具体的は，Twitter REST API によって提供される ‘in_reply_to_status_id’ のフィールドを利用し，ある発話に対して応答している発話を同定する．対話に関わる発話数は 5 億，発話と応答のペア（発話ペア）が約 3 億 6 千万，復元された対話数は 1.7 億だった．

このようにして得られた対話データの発話に対して，少数の感情表現を手がかりとすることで発話者の感情をタグ付けする．本研究では，Plutchik [3] が定めた 8 つの基本感情（怒り，期待，嫌悪，恐れ，喜び，悲しみ，驚き，受容）を感情カテゴリとして採用し，各感情カテゴリにつき平均 10 個の感情表現を人手で用意した．表 1 は使用した感情表現の例である．

教師あり学習においてはタグ付けの精度が重要となるため，感情表現を含み，なおかつ，以下の 2 つの条件を満たす発話をタグ付与の対象とする．

1. 感情表現が自立語を修飾していない
2. 感情表現が否定，仮定，命令，疑問，譲歩，引用の表現と共起しない

例えば,「私は怒った父親が怖い」では怒りのタグは付かない（1 つ目の条件により排除される）．また,「明日は遠足だから雨が降ったら悲しいなぁ」も悲しみのタグは付かない（2 つ目の条件により排除される）．この文は可能性に基づく言及となっており，悲しみのタグを付けることは不適切である．条件に使われた表現は，否定（ない，ぬ），仮定（たら），疑問（？）などである．

表 3 に，感情がタグ付けされた発話数とその精度を示す．タグ付けの精度はランダムに選んだ各感情カテゴリ 100 発話を人手で調査した．精度は怒りと受容以外の感情カテゴリでは 95%以上であった．

| 発話 | 感情 |
|---|---|
| **A**: 一緒に夕食にいかない？ | |
| **B**: すみません．38 度の熱があるため行けません． | |
| **A**: え!?そうなの!?すぐに良くなるといいね． | 驚き |
| **B**: ありがとう．そういってくれて嬉しいよ． | 喜び |

表 2: 感情付き対話の例: 最初の列は 2 ユーザーによる一連の発話であり，2 列めは発話にタグ付けされた発話者の感情を示している．

| 感情 | 精度 | 発話数 |
|---|---|---|
| 怒り | 94% | 197,756 |
| 期待 | 99% | 2,346,350 |
| 嫌悪 | 97% | 337,135 |
| 恐れ | 96% | 1,927,557 |
| 喜び | 95% | 2,247,105 |
| 悲しみ | 96% | 533,931 |
| 驚き | 97% | 830,372 |
| 受容 | 92% | 337,301 |

表 3: タグ付け精度と感情がタグ付けされた発話数

# 3 提案手法

本節では，目的感情を喚起する応答生成手法について述べる．目的感情は Plutchik [3] が定めた基本 8 感情とした．3.1 節では，Ritter ら [2] が提案した統計的応答生成手法について述べる．3.2 節では，ターゲットユーザーの目的感情を喚起させる応答を生成するために応答生成器のモデルを各感情に適応させる方法を述べる．

## 3.1 統計的応答生成

我々は Ritter らの研究 [2] と同様に，統計的機械翻訳の枠組みを利用することにより，与えられた発話に対して応答を生成する．具体的には，この枠組みでは応答を入力発話に対する翻訳とみなす．一般的な機械翻訳システムと同様に，モデルは発話と応答のペア（発話ペア）から機械翻訳ツールを利用することで学習する．

応答生成器を構成する翻訳モデルと言語モデルの学習には GIZA++ [4] と SRILM [5] を利用した．このうち，翻訳モデルについては，後処理として次のように翻訳テーブルをフィルタリングした．GIZA++は対話データに直接適応されたとき，同じ単語を含むフレーズペアを発見しやすいため，そのまま利用するとオウム返しの応答が生成されやすいことが経験的に知られている [2]．そこで，この現象を避けるために，フレー

ズ同士が部分文字列の関係にあるフレーズをテーブルから取り除いた．

得られた言語モデルと翻訳モデルを用いて，与えられた発話に対して適切な応答を生成するために Moses デコーダーを用いた [6]．フレーズの配置は応答の適切さとは相関が強くないため，機械翻訳の場合とは異なり並び替えモデルを用いない [2]．

### 3.2 モデル適応

本節では，3.1 節の枠組みを用いて，応答の内容をコントロールする手法について説明する．

本稿では，応答の内容をコントロールする 1 つの手段として，モデル適応のアプローチを試みる．2 節で構築した感情タグ付き対話コーパスを利用して，8 つの感情に適応した翻訳モデルと言語モデルを学習する．具体的には，各感情 $e$ に対して，感情 $e$ がタグ付けされた発話の直前の発話ペアから翻訳モデルの学習に利用される．表 2 の対話を例として見てみると，最初の 2 つの発話は，驚きの感情を喚起させる応答生成器の翻訳モデルを学習する．その一方で，2 番目の発話は言語モデルの学習に使われる．

この単純なアプローチでは，データスパースネス問題の影響で上手く応答を生成することが難しい．なぜなら，感情タグ付き対話コーパスにおいては，全ての発話に感情がタグ付けされているわけではないため，モデルの学習に使うことができる発話数は全コーパスと比較すると少なくなってしまうからである．この問題に対処するために感情に特化した適応モデルに加えて，感情を考慮せずに全コーパスで学習した一般モデルも併用する．翻訳モデル・言語モデルを，それぞれ全コーパスから学習した翻訳モデル・言語モデルと線形補間により組み合わせる．

翻訳モデルの線形補間には Moses 付属のスクリプト `tmcombine.py` [7]，言語モデルの線形補間には SRILM をそれぞれ利用した．翻訳モデルから得られる 4 つの素性（2 つのフレーズ翻訳確率と 2 つの語彙重み）全てに対して，同じ重み $\alpha$ ($0.0 \leq \alpha \leq 1.0$) を利用した．言語モデルにおける重みは $\beta$ ($0.0 \leq \beta \leq 1.0$) とした．

この 2 つのパラメータ $\alpha$ と $\beta$ は適応モデルの強さをコントロールする．$\alpha$（または $\beta$）= 1.0 のときは，適応モデルのみを用いることに相当する．また，$\alpha$（または $\beta$）= 0.0 のときは，一般モデルのみを用いることに相当する．$\alpha$ と $\beta$ が共に 0.0 のときは，3.1 節で述べた感情を考慮しないモデルと等しくなる．

| | |
|---|---|
| 適応なし | 0.47 |
| 翻訳モデルのみ適応 ($\alpha = 0.6$) | 0.50 |
| 言語モデルのみ適応 ($\beta = 0.8$) | 0.70 |
| 提案手法 ($\alpha = 0.8, \beta = 0.8$) | **0.94** |

表 4: BLEU スコアの比較

## 4 評価実験

### 4.1 テストデータ

我々は 5 人の作業者にテストデータの作成を依頼した．テストデータ中の対話は，2 つの発話から成り，与えられた発話に対してその発話者に特定の感情（目的感情）を喚起させるように作業者が書いた発話（応答）である．

テストデータ作成の具体的な流れは以下の通りである．まず各作業者に 80 発話ずつ与え，それに対する応答を作成してもらうよう依頼した（各感情ごと 10 発話）．この与えられた 80 発話は各作業者で重なることはない．作業者の負担を減らすために，発話は感情タグ付き対話コーパスから提供し，作業者には応答できそうな発話だけを選んでもらった．この選んだ発話は学習コーパスから削除した．結果として，400 発話・応答ペア（= 80 × 5）を得られた．それぞれの発話に対して，さらに他の 2 人に応答を作成してもらった．これにより，1200 発話（= 400 × 3）が得られた．

### 4.2 実験結果

モデルを学習するために，感情タグ付き対話コーパスの 3 億 6 千万の発話ペアを使った．しかしながら，計算機資源の制限から一般モデルの翻訳モデルの学習は約 200 万発話ペアを使った．言語モデルの学習には 3 億 6 千万の応答全てを使って学習した．適応モデルの翻訳モデルと言語モデルは各感情に対応する発話ペア全てで学習した．

評価には人手で作成したテストデータを利用した．システムはテストデータ中の発話と目的感情が与えられると，それに対する応答を生成する．応答生成の結果は BLEU [8] によって評価した．

表 4 は提案手法の 4 つの BLEU スコアを比較している．最初の列はモデル適応を行わないベースラインを表している．これは提案手法の特別なケース（$\alpha = \beta = 0.0$）にあたる．2 番めと 3 番めの列は，翻訳モデルと言語モデルのみに適応モデルを用いた手法である．最後の列は提案手法であり，翻訳モデルと言語モデルの両方が適応モデルである．最後の 3 つは線形補間の重

| | |
|---|---|
| **目的感情**: | 喜び |
| **入力**: | 写真コンクールで優秀賞とったーーーー！<br>2 月 7 日表彰式！ |
| 適応なし: | そーなの？今日はコンクール前に体験 |
| 提案手法: | おお，おめでとうございます !!!!! |
| **目的感情**: | 受容 |
| **入力**: | 放送で前に深夜放送で出てきた人なんだよ w<br>というか，地震あったの？大丈夫？ |
| 適応なし: | 私もみたのですが..! |
| 提案手法: | 結構ゆれたけど大丈夫よ |
| **目的感情**: | 期待 |
| **入力**: | やろやろ！<br>2,3,4 日くらいからへべレケちゃいましょ！ |
| 適応なし: | へべレケなときにやるか（苦笑） |
| 提案手法: | やりますか？酩酊楽しみ楽しみー！ |

表 5: モデル適応なしの手法と提案手法の実験結果

み $\alpha$ と $\beta$ が $\{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$ のグリッドサーチによって最適な重みに調整されている．

この結果からモデル適応が目的感情を喚起するための応答生成に寄与することが確認された．BLEU スコアは 0.47 から 0.94 に向上した．また，翻訳モデル，または言語モデルの一方を適応させただけでは充分でないことを示している．最も高い BLEU スコアは，両方のモデルを目的感情に適応させたときに達していた．

表 5 はモデル適応を行わず生成した結果とモデル適応を行なって生成した結果の例である．最初の 2 つの例では，提案手法が目的感情である喜びと受容を喚起する応答を上手く生成している．これらの例からモデル適応は祝福や安心させるフレーズに大きな確率を与えているものと考えられる．最後の例では，システムは期待の目的感情を喚起させることに成功している．この例では，応答の発話者（システム）は期待の感情を持っていて，これが聞き手の感情に影響を与えていると解釈できる．興味深いことに実際の会話でも同様の現象が報告されている [9].

## 5 関連研究

対話における自動応答生成において，近年では統計的なアプローチが試みられている [2] しかし，我々が知る限り，現時点においてユーザーの感情をモデル化した統計的応答生成システムはない．

読み手の感情を喚起する文を生成する研究としては，ジョークやユーモアのあるテキストを生成する研究がある [10, 11]．これらの研究は特定の感情を聞き手に喚起するという点で我々の研究に似ているが，1 つの感情に特化している．これに対して，我々が提案する手法は 1 つの感情に特化しない汎用性のある手法である．

翻訳モデルや言語モデルの線形補間は機械翻訳のドメイン適応として広く使われている [7]．しかしながら，応答生成において適応した研究はなく，本研究はモデル適応が応答生成においても有効であることを示した初めての研究となる．

## 6 おわりに

本稿では，聞き手の感情を喚起する応答を生成する手法を提案した．提案手法は，モデル適応を用いることで，翻訳モデルと言語モデルを特定の感情に特化したモデルに適応させた．今後の課題は人手により生成した文の日本語らしさ，感情の喚起の有無を評価することである．また，さらに正確な応答生成を実現するためには，発話 1 つだけを見て応答を生成するのではなく，文脈を考慮した応答生成手法が求められる．

## 参考文献


[1] B. Pang and S. Ravi. Revisiting the predictability of language: Response completion in social media. In *Proceedings of EMNLP*, pp. 1489–1499, 2012.

[2] A. Ritter, C. Cherry, and WB. Dolan. Data-driven response generation in social media. In *Proceedings of EMNLP*, pp. 583–593, 2011.

[3] R. Plutchik. A general psychoevolutionary theory of emotion. In *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion*, pp. 3–33. New York: Academic, 1980.

[4] FJ. Och and H. Ney. A systematic comparison of various statistical alignment models. *Computational Linguistics*, Vol. 29, No. 1, pp. 19–51, 2003.

[5] A. Stolcke. SRILM - an extensible language modeling toolkit. In *Proceedings of ICSLP*, pp. 901–904, 2002.

[6] Moses: Open Source Toolkit for Statistical Machine Translation. http://www.statmt.org/moses/.

[7] R. Sennrich. Perplexity minimization for translation model domain adaptation in statistical machine translation. In *Proceedings of EACL*, pp. 539–549, 2012.

[8] K. Papineni, S. Roukos, T. Ward, and W. Zhu. BLEU: a method for automatic evaluation of machine translation. In *Proceedings of ACL*, pp. 311–318, 2002.

[9] S. Kim, J. Bak, and AH. Oh. Do you feel what I feel? social aspects of emotions in Twitter conversations. In *Proceedings of ICWSM*, pp. 495–498, 2012.

[10] P. Dybala, M. Ptaszynski, J. Maciejewski, M. Takahashi, R. Rzepka, and K. Araki. Multiagent system for joke generation: Humor and emotions combined in human-agent conversation. *Journal of Ambient Intelligence and Smart Environments*, Vol. 2, No. 1, pp. 31–48, 2010.

[11] I. Labtov and H. Lipson. Humor as circuits in semantic networks. In *Proceedings of ACL (Short Papers)*, pp. 150–155, 2012.