

# ニュース対話システムにおける感情音声合成のための ニュース記事の文に対する感情ラベルのアノテーションと識別

高津 弘明<sup>1</sup>安藤 涼太<sup>2</sup>松山 洋一<sup>1</sup>小林 哲則<sup>1</sup>早稲田大学<sup>1</sup>内外切抜通信社<sup>2</sup>

takatsu@pcl.cs.waseda.ac.jp, ando@naigaipc.co.jp,  
matsuyama@pcl.cs.waseda.ac.jp, koba@waseda.jp

## 1 はじめに

ニュースを伝える会話システム [高津 18] において、ニュースの内容に応じた感情音声合成を実現するための、ニュース記事の文に対する感情ラベルのアノテーション方法を検討し、それらを識別するモデルを提案する。

人間同士の会話では、相手と親しい関係性を構築（ラポール形成）する手段として、同調的な感情表出が有効だとされている [Clark 88]. このような観点から、近年では、感情表出を可能とする対話システム [Acosta 09, 石川 18, Lubis 18] が開発されている。これまでに、説得 [石川 18] やカウンセリング [Lubis 18] などのタスクにおいて、対話システムの感情表出の有効性が確認されてきた。我々が取り組んでいるニュース伝達タスク [高津 18] においても、継続利用性を向上させるためにユーザーとの関係構築が重要な課題となっている [高津 19b]. そこで、ラポール形成を促進する方法として、ニュースの内容に同調した感情表現を可能とする音声合成システムの開発を検討している。

もう一つ、ニュース伝達における感情音声合成に期待する効果として、理解の促進がある。我々は、談話構造上の発話の役割に基づいて重要な情報を強調するようなメリハリのある話し方を可能とする音声合成システム [高津 19a] を開発し、従来の一文単位の音声合成手法に比べて、内容が頭に入りやすい話し方になっていることを確認した。しかしながら、感情という面においては話し方が単調であるため、ニュースの内容に適した感情で情報を伝えることで、より頭に入りやすい話し方が実現できると考えている。

音声合成システムの感情の分類として、Russell の円環モデル [Russell 80] がしばしば用いられる [Hamada 14, Iida 03, Iwata 11]. 本研究では、ニュースの内容に基づく感情を、Russell の円環モデルの Valence 軸に基づき、「ポジティブ」「ニュートラル」「ネガティブ」に分類する。ニュース記事の各文に対して、これらの感情ラベルをアノテーションし、機械学習のためのデータセットを構築する。また、BERT [Devlin 19] と BiLSTM-CRF [Lample 16] を組み合わせたモデルにより感情ラベルを識別する手法を提案し、作成したデータセットを用いて提案手法の有効性を評価する。さらに、Human-in-the-loop [Munro 19] の枠組みで能動学習を繰り返すことで効率的にモデルの性能を改善できることを示す。

本稿の構成は次の通りである。2章で関連研究について述べる。3章でニュース記事の文に対する感情ラベルのアノテーション方法と作成したデータセットの統計について説明する。4章で提案モデルについて説明し、5章で作成したデータセットを用いてモデルの性能を評価した結果を報告する。6章で能動学習の効果を検証した結果を報告する。7章でまとめと今後の展望について述べる。

## 2 関連研究

テキストに含まれる評判や感情を分析・解析する技術は、評判分析や感情分析と呼ばれる。評判分析の研究の多くが、Twitter のツイートやレビュー文書などの主観的な意見を含むテキストを対象としている [Fan 18, Islam 19, Yin 19, Zhang 19]. これらの研究の多くが書き手視点で感情を分類している。

一方で、客観的な出来事に対する言及が主な内容であるニュース記事を対象とした研究では、読み手視点で感情が分類されることが多い。例えば、Lin らは、ニュースに対する読み手の感情を happy, angry, sad, surprised, heartwarming, awesome, bored, useful に分類し、ニュース記事を読んだユーザーが最も多く投票した感情を正解として SVM [Cortes 95] で識別した [Lin 07, Lin 08]. Li らは、読み手の感情を touching, empathy, boredom, anger, amusement, sadness, surprise, warmth に分類し、マルチラベル分類問題として定式化し、感情ごとに異なる文書の重みを推定するトピックモデルを提案した [Li 16]. Ciptadi らは、読み手の感情を proud, angry, sad, happy, afraid, amused, inspired, surprised に分類し、不均衡データの問題を緩和するために、SMOTE [Chawla 02] によるオーバーサンプリングの手法を適用し、ナイーブベイズ分類器やロジスティック回帰の分類性能が向上することを示した [Ciptadi 19].

このような記事全体を分類対象とした研究 [Ciptadi 19, Hui 17, Li 16, Lin 07, Lin 08, Liu 13, Wang 17] の他にも、見出しを分類対象とした研究 [Kirange 12, Strapparava 07] や、ニュース記事の文を分類対象とした研究 [Bhowmick 09, Bhowmick 10, Das 09, Li 15, Patil 12] がある。

文を対象とした研究として、例えば、Bhowmick らは、読み手の感情を anger, disgust, fear, happiness, sadness, surprise に分類し、ニュース記事の文に対して複数人でアノテーションを行い、一致率を評価した [Bhowmick 10]. その結果、anger と disgust を統合することで、surprise を除くことで一致率が向上することを確認した。また、モデルとして ADTboost.MH [Comite 03] を利用し、マルチラベルの分類性能を評価したところ、anger と disgust を統合し、surprise を除くことでモデルの性能も向上することを確認した。Li らは、一つの文が hate と angry のような似たラベルを持ちやすいというラベル依存性と、同じ文脈にある文は同じラベルを持ちやすいという文脈依存性を因子グラフでモデル化した [Li 15]. 実験の結果、文脈として文書や段落よりも隣接文を用いたときのモデルが最も良い性能を示すことを確認した。

ニュース記事ではなくレビュー文書を対象とした研究ではあるが、Zhang らは、文の系列ラベリング問題として定式化し、CRF [Lafferty 01] により感情ラベル (positive, negative, neutral) を識別する手法を提案した [Zhang 14]. また、能動学習の実験結果から、文書の前半分分の文の平均確率が最小であった文書に対してラベル付けしていく方法が最もモデル

表 1: データセットの統計 (P: ポジティブラベルの総数, O: ニュートラルラベルの総数, N: ネガティブラベルの総数)

	タイトル			文		
	P	O	N	P	O	N
スポーツ	65	10	71	555	399	564
テクノロジー	91	8	93	776	519	908
ビジネス	83	3	87	685	381	942
マーケット	42	5	58	326	165	514
経済・政治	55	6	62	492	383	455
国際・アジア	48	3	57	411	216	457
社会	114	4	98	860	559	699
地域	68	1	66	744	207	553

の精度向上に効果的であることを確認した。

我々も Zhang らと同様に、ニュース記事の各文の感情ラベルを識別する問題を、文の系列ラベリングの問題として捉え、BERT [Devlin 19] と BiLSTM-CRF [Lample 16] を組み合わせたモデルにより識別する手法を提案する。また、Human-in-the-loop [Munro 19] の枠組みで能動学習を繰り返すことで効率的なモデルの性能向上を図る。

### 3 感情ラベル付きニュース記事コーパス

ニュースの内容に基づく感情を、Russell の円環モデル [Russell 80] の Valence 軸に基づき、「ポジティブ」「ニュートラル」「ネガティブ」に分類する。ニュース記事の各文に対して、これらの感情ラベルをアノテーションすることで、機械学習のためのデータセットを構築した。

アノテーションは、日経新聞の 5 文から 20 文のニュース記事 (2017 年 6 月～2019 年 9 月) に対して、ウェブニュースのクリッピングの専門家が行った。アノテーターには、ルールベースの手法 (5.1 節参照) でカテゴリごとにランキングしたニュース記事のリストを提示し、リストの上位 (ポジティブであることが期待されるニュース)、中位 (ニュートラルであることが期待されるニュース)、下位 (ネガティブであることが期待されるニュース)、それぞれを均等にアノテーションさせた。また、多様な話題をカバーするために、似たような話題に対するアノテーションはなるべく避けるように指示した。

アノテーション作業では、まず、ニュース記事のタイトルを読み、記事の要旨を確認し、タイトルに対してアノテーションを行うことで、記事全体の内容や感情の傾向を理解する。次に、記事を全て読んだ後、一文目から順番に、各文に対して「ポジティブ」「ニュートラル」「ネガティブ」のいずれかのラベルを付与する。「ポジティブ」ラベルは、文脈を考慮した文の主題が良いこと、または、良い方向に向かっていることを示す内容である場合に付与する。例えば、社会貢献や市場拡大、権益獲得などが挙げられる。「ネガティブ」ラベルは、文脈を考慮した文の主題が悪いこと、または、悪い方向に向かっていることを示す内容である場合に付与する。例えば、業績悪化や不正行為、事件・事故などが挙げられる。「ニュートラル」ラベルは、ポジティブな内容もネガティブな内容も含まない場合に付与する。一つの文にポジティブな内容とネガティブな内容の両方が含まれることもあるが、今回は、そのような文を含む記事は対象外とした。

記事のカテゴリごとに各感情ラベルの統計を算出した結果を表 1 に示す。表は、タイトルおよび文に対してアノテーションした各感情ラベル (P: ポジティブ, O: ニュートラル, N: ネガティブ) の総数を表す。この表から、ニュートラルのラベルは他に比べて少ないが、ポジティブとネガティブのラベルはバランス良くアノテーションされていることが分かる。

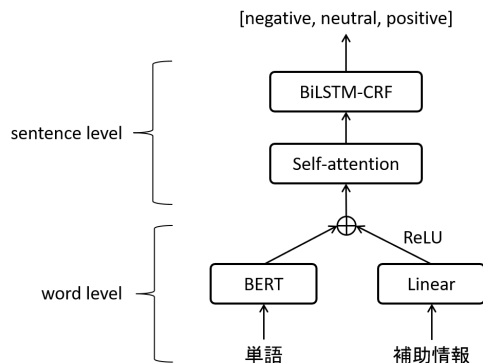


図 1: 提案モデル: BERT\_SA\_BiLSTM-CRF+

### 4 提案モデル

文の感情ラベルの識別問題を系列ラベリングの問題として定式化する。提案モデルの概略図を図 1 に示す。単語のエンコーダーには、BERT (Bidirectional Encoder Representations from Transformers) [Devlin 19] を使用した。BERT は、Transformer [Vaswani 17] のエンコーダーをベースとしたモデルである。BERT の事前学習モデルとして、柴田らが公開しているモデルを用いた [柴田 19]<sup>1</sup>。提案モデルでは、まず、各単語に対応する Transformer の最終層の埋め込み表現と各単語の補助情報 (品詞など) を結合した表現に対して Self-attention [Lin 17] を計算し、文のベクトル表現を得る。そして、得られた各文のベクトル表現を BiLSTM-CRF [Lample 16] に与え、前後の文の文脈情報とラベル間の依存関係を考慮したうえで、ターゲットの文のラベルを推定する。デコーディング時は、Viterbi アルゴリズムによってラベル系列を推定する。

補助情報としては、JUMAN++ (Ver.1.02) [Morita 15]<sup>2</sup> の形態素情報 (品詞大分類, 品詞細分類, 活用形, 活用型, カテゴリ, ドメイン), KNP (4.19) [Kawahara 06]<sup>3</sup> を適用して得られる固有表現クラスや係り受けの種類, 係り受け木の深さ, 係り元の文節数の他, 単語の TF, IDF, TF-IDF, 「」『』内の単語かどうか, 文頭からの文節位置, 記事における文および段落の位置, 記事のカテゴリ, 単語の感情極性値 (『日本語評価極性辞書』 [小林 05, 東山 08]<sup>4</sup>, 『単語感情極性対応表』 [Takamura 05]<sup>5</sup>, 『意見 (評価表現) 抽出ツール用モデル』<sup>6</sup> の評価表現辞書), 極性反転語か否か (『意見 (評価表現) 抽出ツール用モデル』の反転辞書) を用いた。

### 5 モデルの性能評価

作成したデータセットを用いて、提案モデルを評価する。

#### 5.1 実験設定

文分類の比較手法として、ラベルをランダムに選択するモデル (Random), データセットにおいて最頻出のラベル (つまり、ネガティブ) を選択するモデル (Mode), ルールベースの手法においてポジティブ, ニュートラル, ネガティブの閾値を最も性能が高くなるように調節したモデル (Rule-best),

<sup>1</sup><http://nlp.ist.i.kyoto-u.ac.jp/index.php?BERT> 日本語 Pre-trained モデル

<sup>2</sup><http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

<sup>3</sup><http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

<sup>4</sup><https://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resource>

<sup>5</sup>2F Japanese%20Sentiment%20Polarity%20Dictionary

<sup>6</sup>[http://www.lr.pi.titech.ac.jp/~takamura/pndic\\_ja.html](http://www.lr.pi.titech.ac.jp/~takamura/pndic_ja.html)

<sup>6</sup><https://alaginrc.nict.go.jp/li-outline.html#C-3>

表 2: モデルの性能比較

Task	Model	Macro F-measure	Overall Accuracy
Sentence Classification	Random	0.340	0.345
	Mode	0.183	0.380
	Rule-best	0.578	0.608
	SVM	0.608	0.645
	BERT	0.685	0.722
	BERT_SA+	0.700	0.734
Sequence Labeling	BiLSTM-CRF	0.670	0.723
	BERT_BiLSTM-CRF	0.738	0.774
	BiLSTM_SA_BiLSTM-CRF+	0.695	0.737
	BERT_SA_BiLSTM-CRF+	<b>0.771</b>	<b>0.805</b>

文の単語の Bag-of-Words を特徴量として学習した線形カーネルの SVM モデル (SVM), BERT の [CLS] に対応する Transformer の最終層の上に Linear 層を追加して Softmax を適用するモデル (BERT), BERT の各単語の Transformer の最終層と各単語の補助情報の埋め込み表現を結合したものに対して Self-attention を計算して得られる文のベクトル表現に Softmax を適用するモデル (BERT\_SA+) と比較した。ルールベースの手法では、単語の感情極性辞書 (4章参照) を用いて、極性反転も考慮したうえで、ポジティブ語とネガティブ語の出現頻度に基づいて文の感情極性値を計算している。

系列ラベリングの比較手法として、文の単語の Bag-of-Words を BiLSTM-CRF に入力して識別するモデル (BiLSTM-CRF), BERT の [CLS] に対応する Transformer の最終層の埋め込み表現を BiLSTM-CRF に入力して識別するモデル (BERT\_BiLSTM-CRF), 文の単語を入力とする BiLSTM の隠れ層に補助情報の埋め込み表現を結合したものに対して Self-attention を計算して得られる文のベクトル表現を BiLSTM-CRF に入力して識別するモデル (BiLSTM\_SA\_BiLSTM-CRF+) と比較した。

## 5.2 実験結果

評価尺度として Macro F-measure と Overall Accuracy を採用し、10 分割交差検定で評価した結果を表 2 に示す。この結果から、提案モデルが最も良い性能を示していることが分かる。また、単語の Bag-of-Words を特徴量としたモデルよりも BERT の埋め込み表現を用いたモデルの方が良い結果を示していることが分かる。さらに、補助情報を考慮することでモデルの性能が向上することが分かる。

## 6 能動学習の効果検証

Human-in-the-loop [Munro 19] の枠組みで、確信度が低い記事を優先的にアノテーションすることで、モデルの性能を効率的に改善できることを確認する。

### 6.1 実験設定

Viterbi アルゴリズムによるデコーディングで得られるスコアを文の数で割った値を記事の文系列に対するラベリングの確信度とする。データセットを訓練セット (全体の 3/4) とテストセット (全体の 1/4) に分割し、BERT\_SA\_BiLSTM-CRF+ の学習を行った。テストセットにおいて記事ごとに計算した Accuracy と最大値が 1 となるように正規化した確信度の散布図を図 2 に示す。この図から、確信度が高い記事ほど Accuracy が高い傾向にあることが見て取れる。また、ピアソンの積率相関係数は 0.576 であり、適度な相関が見られた。

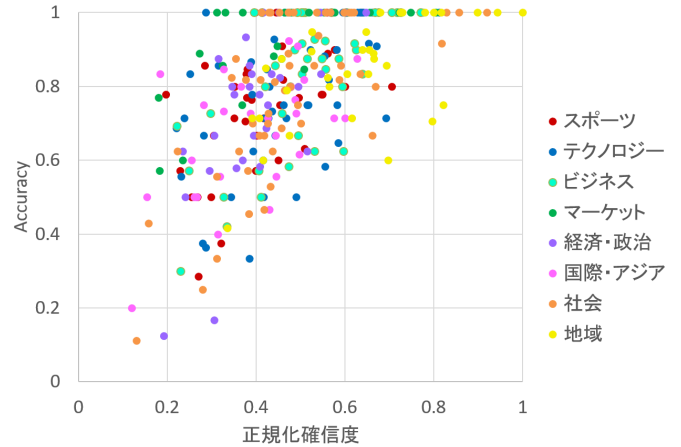


図 2: テストセットにおける確信度と Accuracy の散布図

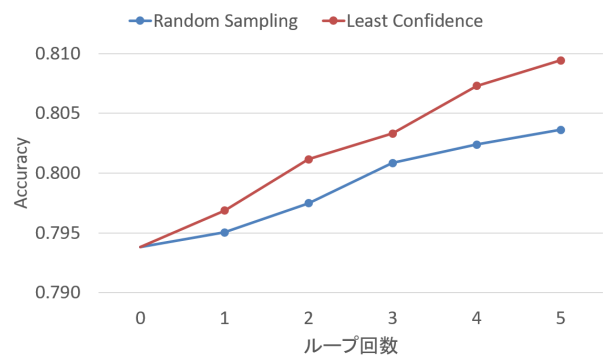


図 3: 能動学習のループ回数と Accuracy の変化

これらの結果から、確信度が低い記事を優先してアノテーションすることで、モデルの効率的な性能改善が期待できる。そこで、以下の 3 つのステップを繰り返すことによる Accuracy の変化を評価した。

- (1) データセットに含まれない、ラベルが未知の記事に対してモデルを適用し、カテゴリごとに確信度が低い順に記事をランキングする。
- (2) アノテーション条件にマッチする確信度が最も低い記事をカテゴリごとに一つずつ選択し、感情ラベルをアノテーションする。
- (3) アノテーションした記事を訓練セットに加えてモデルを再学習し、テストセットに対する性能を評価する。

比較手法として、確信度によらず、カテゴリごとにランダムに一つずつ選択した記事に対してアノテーションしたデータを加えて学習したモデルと比較した。

### 6.2 実験結果

確信度の低い記事にアノテーションしたデータを加えて学習したモデル (Least Confidence) とランダムに選んだ記事にアノテーションしたデータを加えて学習したモデル (Random Sampling) の、能動学習のループ回数に応じた Accuracy の変化をプロットしたグラフを図 3 に示す。

この結果から、確信度の低い記事にアノテーションしたデータを用いたモデルの方が、ランダムに選んだ記事にアノテーションしたデータを用いたモデルよりも、効率的に性能を改善できていることが分かる。

正解 予測  
 ■ positive → neutral  
 ■ neutral → positive  
 ■ negative → neutral  
 ■ neutral → negative  
 ■ positive → negative  
 ■ negative → positive

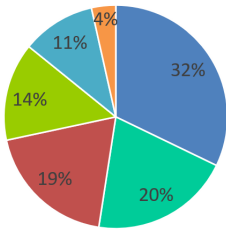


図 4: ループ前の誤り傾向

正解 予測  
 ■ neutral → positive  
 ■ positive → neutral  
 ■ negative → neutral  
 ■ neutral → negative  
 ■ negative → positive  
 ■ positive → negative

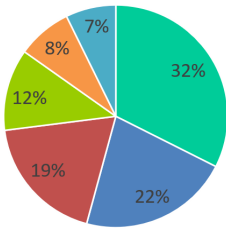


図 5: ループ後の誤り傾向

Accuracy が 0.794 のループ前のモデルの誤り傾向を図 4 に示す。Accuracy が 0.809 のループ後のモデルの誤り傾向を図 5 に示す。これらの結果を比較すると、Human-in-the-loop により、正解がポジティブの文をニュートラルと推定してしまう誤りは減らすことができた一方で、正解がニュートラルの文をポジティブと推定してしまう誤りが増えており、ポジティブとニュートラル間の識別が他に比べて難しいことが分かる。また、感情音声合成での利用を考えた場合、致命的となるのがポジティブとネガティブ間の誤りであるが、ポジティブな文をネガティブまたはネガティブな文をポジティブと推定してしまう誤りは、ループ前もループ後も誤りの 15% 程度に抑えられているが分かる。

## 7 おわりに

ニュースを伝える会話システム [高津 18] において、ニュースの内容に応じた感情音声合成を実現するため、ニュース記事の文に対して、感情ラベルのアノテーションを行い、機械学習のためのデータセットを構築した。また、感情ラベルを識別するモデルを提案し、作成したデータセットを用いて提案モデルの有効性を評価した。さらに、Human-in-the-loop [Munro 19] の枠組みで、確信度の低い記事を優先的にアノテーションすることで、モデルの性能を効率的に改善できることを確認した。

今後は、我々が開発した、談話構造上の発話の役割に基づいて重要な情報を強調するようなメリハリのある話し方を可能とする音声合成システム [高津 19a] に加え、本研究で提案したモデルにより推定した文の感情も考慮することで、感情的にもメリハリのある音声合成システムを開発するとともに、ニュース伝達タスクにおいて、感情を伴う話し方がラポール形成や理解の促進に効果があるか確認する。

## 参考文献

[Acosta 09] Acosta, J. C.: Using emotion to gain rapport in a spoken dialog system, in *Proceedings of the NAACL HLT Student Research Workshop and Doctoral Consortium*, pp. 49-54 (2009)

[Bhowmick 09] Bhowmick, P. K., Basu, A., Mitra, P., and Prasad, A.: Multi-label text classification approach for sentence level news emotion analysis, in *Proceedings of the 3rd International Conference on Pattern Recognition and Machine Intelligence*, pp. 261-266, (2009)

[Bhowmick 10] Bhowmick, P. K., Basu, A., and Mitra, P.: Classifying emotion in news sentences: When machine classification meets human classification, *International Journal on Computer Science and Engineering*, Vol. 2, No. 1, pp. 98-108 (2010)

[Chawla 02] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P.: SMOTE: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321-357 (2002)

[Ciptadi 19] Ciptadi and Girsang, A. S.: Emotion classification based on public opinion analysis on online news, *International Journal of Scientific & Technology Research*, Vol. 8, pp. 176-182 (2019)

[Clark 88] Clark, M. S. and Reis, H. T.: Interpersonal processes in close relationships, *Annual Review of Psychology*, Vol. 39, No. 1, pp. 609-672 (1988)

[Comite 03] Comite, F. D., Gilleron, R., Tommasi, M.: Learning multi-label alternating decision trees from texts and data, in *Proceedings of the 3rd International Conference on Machine Learning and Data Mining in Pattern Recognition*, pp. pp. 35-49 (2003)

[Cortes 95] Cortes, C. and Vapnik, V.: Support-vector networks, *Machine Learning*, Vol. 20, No. 3, pp. 273-297 (1995)

[Das 09] Das, D. and Bandyopadhyay, S.: Analyzing emotion in blog and news at word and sentence level, in *Proceedings of the 4th Indian International Conference on Artificial Intelligence*, pp. 1-13 (2009)

[Devlin 19] Devlin, J., Chang, M. W., Lee, K., and Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 4171-4186 (2019)

[Fan 18] Fan, F., Feng, Y., and Zhao, D.: Multi-grained attention network for aspect-level sentiment classification, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 3433-3442 (2018)

[Hamada 14] Hamada, Y., Elbarougy, R., and Akagi, M.: A method for emotional speech synthesis based on the position of emotional state in Valence-Activation space, in *Proceedings of the Signal and Information Processing Association Annual Summit and Conference*, pp. 1-7 (2014)

[東山 08] 東山昌彦, 乾健太郎, 松本裕治, 述語の選択適性に着目した名詞評価感性的獲得: 言語処理学会第 14 回年次大会論文集, pp. 584-587 (2008)

[Hui 17] Hui, J. L. O., Hoon, G. K., Zainon, W. M. N. W.: Effects of word class and text position in sentiment-based news classification, in *Proceedings of the 4th Information Systems International Conference*, pp. 77-85 (2017)

[Iida 03] Iida, A., Campbell, N., Higuchi, F., and Yasumura, M.: A corpus-based speech synthesis system with emotion, *Speech Communication*, Vol. 40, pp. 161-187 (2003)

[石川 18] 石川菜子, 水上雅博, 吉野幸一郎, Sakriani Sakti, 鈴木優, 中村哲: 感情表現を用いた説得対話システム, 人工知能学会論文誌, Vol. 33, No. 1, pp. 1-9 (2018)

[Islam 19] Islam, J., Mercer, R. E., and Xiao, L.: Multi-channel convolutional neural network for twitter emotion and sentiment recognition, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1355-1365 (2019)

[Iwata 11] Iwata, K. and Kobayashi, T.: Conversational speech synthesis system with communication situation dependent HMMs, in *Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop*, pp. 113-123 (2011)

[Kawahara 06] Kawahara, D. and Kurohashi, S.: A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis, in *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 176-183 (2006)

[Kirange 12] Kirange, D. K., Deshmukh, R. R.: Emotion classification of news headlines using SVM, *Asian Journal of Computer Science and Information Technology*, Vol. 5, No. 2, pp. 104-106 (2012)

[小林 05] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, 自然言語処理, Vol. 12, No. 3, pp. 203-222 (2005)

[Lafferty 01] Lafferty, J., McCallum, A., and Pereira, F.: Conditional random fields: probabilistic models for segmenting and labeling sequence data, in *Proceedings of the 18th International Conference on Machine Learning*, pp. 282-289 (2001)

[Lample 16] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C.: Neural architectures for named entity recognition, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 260-270 (2016)

[Li 15] Li, S., Huang, L., Wang, R., and Zhou, G.: Sentence-level emotion classification with label and context dependence, in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1045-1053 (2015)

[Li 16] Li, X., Xie, H., Rao, Y., and Chen, Y.: Weighted multi-label classification model for sentiment analysis of online news, in *Proceedings of the 2016 International Conference on Big Data and Smart Computing*, pp. 215-222 (2016)

[Lin 07] Lin, K. H.-Y., Yang, C., Chen, H.-H.: What emotions do news articles trigger in their readers?, in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 733-75 (2007)

[Lin 08] Lin, K. H.-Y., Yang, C., and Chen, H.-H.: Emotion classification of online news articles from the reader's perspective, in *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 220-226 (2008)

[Lin 17] Lin, Z., Feng, M., Santos, C.N., Yu, M., Xiang, B., Zhou, B., and Bengio, Y.: A structured self-attentive sentence embedding, in *Proceedings of the 5th International Conference on Learning Representations*, pp. 1-15 (2017)

[Liu 13] Liu, H., Li, S., Zhou, G., Huang, C.-R., and Li, P.: Joint modeling of news reader's and comment writer's emotions, in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pp. 511-515 (2013)

[Lubis 18] Lubis, N., Sakti, S., Yoshino, K., and Nakamura, S.: Unsupervised counselor dialog clustering for positive emotion elicitation in neural dialog system, in *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 161-170 (2018)

[Morita 15] Morita, H., Kawahara, D., and Kurohashi, S.: Morphological analysis for unsegmented languagusing recurrent neural network language model, in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2292-2297 (2015)

[Munro 19] Munro, R.: Human-in-the-loop machine learning, *Manning* (2019)

[Patil 12] Patil, S. S. and Chaudhari, A. P.: Classification of emotions from text using SVM based opinion mining, *International Journal of Computer Engineering & Technology*, Vol. 3, pp. 330-338 (2012)

[Russell 80] Russell, J. A.: A circumplex model of affect, *Journal of Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161-1178 (1980)

[柴田 19] 柴田知秀, 河原大輔, 黒橋橋夫: BERT による日本語構文解析の精度向上, 言語処理学会第 25 回年次大会発表論文集, pp. 205-208 (2019)

[Strappavara 07] Strappavara, C. and Mihalcea, R.: SemEval-2007 task 14: Affective text, in *Proceedings of the Fourth International Workshop on Semantic Evaluations*, pp. 70-74 (2007)

[Takamura 05] Takamura, H., Inui, T., and Okumura, M.: Extracting semantic orientations of words using spin model, in *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 133-140 (2005)

[高津 18] 高津弘明, 福岡維新, 藤江真也, 林良彦, 小林哲則: 意図の異なる多様な情報行動を可能とする音声対話システム, 人工知能学会論文誌, Vol. 33, No. 1, pp. 1-24 (2018)

[高津 19a] 高津弘明, 福岡維新, 藤江真也, 岩田和彦, 小林哲則: 会話によるニュース記事伝達のための音声合成, 人工知能学会論文誌, Vol. 34, No. 2, pp. 1-15 (2019)

[高津 19b] 高津弘明, 松山洋一, 本田裕, 藤江真也, 小林哲則: 会話によるニュース記事伝達のためのトリビアの獲得と活用, 人工知能学会第 33 回全国大会論文集, 3Rin2-29, pp. 1-4 (2019)

[Vaswani 17] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., and Kaiser, L.: Attention is all you need, in *Proceedings of the 31st Conference on Neural Information Processing System*, pp. 6000-6010 (2017)

[Wang 17] Wang, J.-H. and Liu, H.-Y.: Discovering reader's emotions triggered from news articles, in *Proceedings of the 4th Multidisciplinary International Social Networks Conference*, No. 45, pp. 1-7 (2017)

[Yin 19] Yin, D., Liu, X., Wu, X., and Chang, B.: A soft label strategy for target-level sentiment classification, in *Proceedings of the 10th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 6-15 (2019)

[Zhang 14] Zhang, K., Xie, Y., Yang, Y., Sun, A., Liu, H., and Choudhary, A.: Incorporating conditional random fields and active learning to improve sentiment identification, *Neural Networks*, Vol. 58, pp. 60-67 (2014)

[Zhang 19] Zhang, Y. and Zhang, Y.: Tree communication models for sentiment analysis, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 3518-3527 (2019)