木構造とグラフ構造を用いた オンライン議論における談話行為の分類

三浦 康秀^{†,††}・狩野 竜示[†]・谷口 元樹[†]・谷口 友紀[†]・
三沢翔太郎[†]・大熊 智子[†]

本稿ではオンライン議論における談話行為を分類するモデルを提案する。提案モデルでは談話行為を分類するために、ニューラルネットワークを用いて議論のパターンを学習する。談話行為の分類において議論のパターンを取り入れる重要性は既存の研究においても確認されているが、対象としている議論に併せたパターン素性を設計する必要があった。提案モデルではパターン素性をを用いずに、木構造およびグラフ構造を学習する層を用いて議論のパターンを学習する。提案モデルを Reddit の談話行為を分類するタスクで評価したところ、従来手法と比較して Accuracy で 1.5%, F_1 値で 2.2 ポイントの性能向上を確認した。また、提案モデル内の木構造学習層およびグラフ構造学習層間の相互作用を確認するため、提案手法の中間層を注意機構を通じて分析した。

キーワード：談話行為, 木構造, グラフ構造, ニューラルネットワーク, 議論分析

Discourse Act Classification Using Discussion Patterns with Neural Networks

YASUhide MIURA^{†,††}, RYUJI KANO[†], MOTOKI TANIGUCHI[†],
TOMOKI TANIGUCHI[†], SHOTARO MISAWA[†] and TOMOKO OHKUMA[†]

We proposed a model that classifies discussion discourse acts using discussion patterns with neural networks. Several attempts have been made in earlier works to analyze texts that are used in various discussions. The importance of discussion patterns has been explored in those works but their methods required a sophisticated design to combine pattern features with a classifier. Our model introduces tree learning approaches and a graph learning approach to capture discussion patterns without pattern features. In an evaluation to classify discussion discourse acts in Reddit, the model achieved improvements of 1.5% in accuracy and 2.2 in F_1 score compared to the previous best model. We further analyzed the model using an attention mechanism to inspect interactions among different learning approaches.

Key Words: *Discourse Act, Tree Structure, Graph Structure, Neural Networks, Discussion Analysis*

[†] 富士ゼロックス株式会社, Fuji Xerox Co., Ltd.

^{††} 東京工業大学, Tokyo Institute of Technology

1 はじめに

近年、ソーシャルニュースサイトや討論ポータルの発展に伴い、様々な話題がオンライン上で議論されるようになった。これら議論は世の中の貴重な意見を含んでいるが、分析には関連する複数の投稿・発言の内容を理解する必要がある。これまでも対話行為の分析 (Stolcke, Ries, Coccaro, Shriberg, Bates, Jurafsky, Taylor, Martin, Van Ess-Dykema, and Meteer 2000; Bunt, Alexandersson, Carletta, Choe, Fang, Hasida, Lee, Petukhova, Popescu-Belis, Romary, Soria, and Traum 2010) を発展させ、議論を談話行為に基づいて分析するアプローチが提案されてきた。議論における談話行為の自動的な分類は、情報アクセスや要約の改善に寄与できると考えられている。このため、電子メール (Cohen, Carvalho, and Mitchell 2004; Carvalho and Cohen 2005, 2006; Hu, Passonneau, and Rambow 2009; Omuya, Prabhakaran, and Rambow 2013)、ニュースグループ (Wang, Joshi, and Rose 2007)、技術電子掲示板 (Kim, Wang, and Baldwin 2010b; Wang, Lui, Kim, Nivre, and Baldwin 2011; Bhatia, Biyani, and Mitra 2012; Liu, Baldwin, and Cohn 2017)、ソーシャルニュース (Zhang, Culbertson, and Paritosh 2017) 等の談話行為・対話行為が既存の研究で対象とされてきた。

議論において談話行為・対話行為を分類する際には、議論のパターンを取り入れることの重要性がたびたび指摘されてきた。投稿間の関係・リンク (Carvalho and Cohen 2005; Hu et al. 2009)、投稿の位置・深さ (Wang et al. 2007; Kim, Cavedon, and Baldwin 2010a; Kim et al. 2010b; Wang et al. 2011; Bhatia et al. 2012; Zhang et al. 2017; Liu et al. 2017) 等のパターン情報は、確率的なグラフィカルモデル、構造学習モデル、系列学習モデル等と組み合わせて利用されてきた。これらアプローチは談話行為の分類において有効性を示したが、分類モデルにパターン情報を取り入れるためにタスク依存のパターン素性を設計する必要があった。本稿では議論のパターンをニューラルネットワークを用いて取り入れるモデルを提案する。近年、ニューラルネットワークを用いて木構造 (Socher, Lin, Ng, and Manning 2011; Socher, Karpathy, Le, Manning, and Ng 2014; Tai, Socher, and Manning 2015) やグラフ構造 (Defferrard, Bresson, and Vandergheynst 2016; Kipf and Welling 2017) を学習する有効性が示されている。提案モデルではパターン素性を設計せずに、木構造学習層とグラフ構造学習層を用いて議論のパターンを学習する。

既存の研究では様々な対象の談話行為・対話行為が分類されてきたが、本稿では Reddit¹の談話行為の分類に提案モデルを適用する。Reddit は大規模なソーシャルニュースサイトであり、数多くのトピックについて日々議論が行われている。議論はスレッド単位で行われ、トピックを提供する最初の投稿および投稿に対する返信の連鎖で構成される。提案モデルの評価では Zhang et al. (2017) の 9 種類の談話行為を対象にする。9 種類の談話行為は *Answer*, *Elaboration*,

---

¹ <https://www.reddit.com/>

図 1 Reddit を対象とした談話行為の例. 各丸はスレッド中の投稿, 赤色は最初の投稿, 青色は返信の投稿, 矢印は返信を表している. 最初の質問 (*Question*) が答え (*Answer*) とユーモア (*Humor*) の返信を受けており, 答えの一つはさらに謝辞 (*Appreciation*) と追加情報 (*Elaboration*) の返信を受けている.

Question, Appreciation, Agreement, Disagreement, Humor, Announcement, Negative Reaction であり, 図 1 にこれら談話行為の例を示す. 本稿では次の二つの理由で Reddit を対象とした. 第一に, Reddit 上での議論は投稿をノード, 返信関係をエッジとした木構造およびグラフ構造として表すことができる. 第二に, 公開されている Reddit の談話行為が付与されたコーパスは大規模であり, ニューラルネットワークを用いて木構造やグラフ構造を学習するのに適している. 本稿の貢献には以下の三点が挙げられる:

- (1) 投稿間の構造に対応した, 木構造学習層とグラフ構造学習層を含むモデルを提案する.
  - (2) 談話行為の分類性能において, 提案モデルが従来のパターン素性と系列学習を組み合わせたモデルを上回ることを示す.
  - (3) 提案モデルの中間層を注意機構を通じて分析し, 談話行為の分類に有効な構造を確認する.
- 本稿の以降の章では次の内容を述べる. 2 章で提案モデルの関連研究を紹介し, モデルの詳細を 3 章で述べる. 4 章で提案モデルを用いた評価実験を報告し, 結果を 5 章で考察する. 最後に 6 章では本稿をまとめ, さらに今後の展望を述べる.

2 関連研究

2.1 談話行為・対話行為の分類

議論を対象として, 談話行為・対話行為を自動的に分類する手法は従来より研究されてきた. 初期の研究としては, 特に電子メールおよび技術電子掲示板が対象となることが多かった. Cohen et al. (2004) は電子メールに対話行為をアノテーションし, テキスト素性を用いた分類器を構築した. この研究は Carvalho and Cohen (2005) によって関係素性を用いたコレクティブ分類モデル, Carvalho and Cohen (2006) によって N-gram 素性の拡張が行われた. Hu et al. (2009) も

電子メールに対話行為をアノテーションし、構造学習モデルを用いた分類を試みた。この研究は Omuya et al. (2013) によってクラス毎の素性の最適化およびカスケード分類器を用いた拡張が行われた。Kim et al. (2010b) は技術電子掲示板の投稿に対話行為をタグ付けし、パターン素性と系列学習器を用いた分類実験を実施した。この研究はのちに、Wang et al. (2011) によって対話行為の分類とリンク関係の同時学習、Liu et al. (2017) によって外部記憶を持つ系列学習器の拡張が行われた。Bhatia et al. (2012) も技術電子掲示板の投稿に対話行為をタグ付けし、パターン素性を含む様々な素性を用いた分類器を評価した。

議論を対象とした談話行為・対話行為の分類は、電子メール・技術電子掲示板以外を対象としても研究されてきた。Wang et al. (2007) はニュースグループの節を対象にパターン素性と系列学習モデルを用いた議論カテゴリの分類を評価した。Kim et al. (2010a) はチャットの対話行為を分類するためにパターン素性を用いた系列学習モデルを探索した。Zhang, Gao, and Li (2011) は短文投稿を対話行為でアノートし、単語素性と文字素性を用いた分類モデルを学習した。Zhang, Gao, and Li (2012) は同じ短文投稿データを用いて、ラベル伝搬に基づく半教師あり学習モデル等を探索した。Ferschke, Gurevych, and Chebotar (2012) は対話行為が付与された Wikipedia ノートページのコーパスを作成し、テキスト素性とパターン素性を用いた分類器を学習した。Zhang et al. (2017) は Reddit のスレッドを談話行為でアノートし、パターン素性を含む様々な素性を用いた系列学習モデルを設計した。

2.2 Reddit の自動分析

大量のテキストおよび画像コンテンツから構成される Reddit を対象に、世の中の意見を分析する研究が数多く行われてきた。Reddit を対象とした自動分析の中でも、人気度を表す karma score（ポジティブ投票とネガティブ投票の差）の分析は広く研究されてきた。Jaech, Zayats, Fang, Ostendorf, and Hajishirzi (2015) は人気度のランキングタスクを提案し、ペアワイズ分類器を用いたランキングを行った。Wei, Liu, and Li (2016) は人気度を用いて投稿を説得力に応じてランキングするシステムを設計した。He, Ostendorf, He, Chen, Gao, Li, and Deng (2016) は動的に人気度を推定する深層強化学習の構成を提案し、この研究はさらに He, Ostendorf, and He (2017) によって 2 段階の処理に拡張された。Hessel, Lee, and Mimno (2017) は人気度の推定タスクにおいてマルチモーダル素性が有効であることを示した。Cheng, Fang, and Ostendorf (2017) は関連要素を因子として結び付けるニューラルモデルを用いて人気度を推定し、文書埋め込みに基づくモデルを上回る性能を実現した。Zayats and Ostendorf (2018) は人気度を推定するのに適したグラフ構造のニューラルモデルを提案し、投稿を独立に扱うモデルを上回る性能を実現した。

Reddit を対象とした自動分析では、人気度以外に着目した研究も数多く行われている。本稿が対象とする談話行為の分析 (Zhang et al. 2017) もその一つであり、他にも様々なタスクが

検討されている。Buntain and Golbeck (2014) は回答を投稿する人の社会的な役割に着目し、それらを教師あり学習で分類した。Tan, Niculae, Danescu-Niculescu-Mizil, and Lee (2016) は ChangeMyView subreddit を対象とし、説得性の高い投稿の明確化および意見の適応性を分類するモデルを構築した。Lim, Carman, and Wong (2017) は投稿者の専門性を投稿内容に非依存な手法を用いて探索した。

2.3 提案モデルとの比較

提案モデルでは議論のパターンをニューラルネットワークに基づく木構造学習層とグラフ構造学習層を用いて学習する。これは従来用いられてきたパターン素性を用いずに、テキスト情報と議論のパターンを同時に学習するアプローチである。また、談話行為・対話行為の分類では従来は系列学習モデルが用いられることが多かった (Wang et al. 2007; Kim et al. 2010a, 2010b; Wang et al. 2011; Zhang et al. 2017; Liu et al. 2017)。Carvalho and Cohen (2005) は確率的なグラフィカルモデルを用いたが、関係素性と事前学習したテキスト素性に基づく分類器を前提としている。提案モデルを Reddit の自動分析を行うモデルと比較すると、類似したアプローチが Zayats and Ostendorf (2018) で用いられている。Zayats and Ostendorf (2018) では Long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) を拡張し、階層および時系列情報をグラフ構造として学習するモデルを提案した。しかし、このモデルは Reddit の人気度を効率的に学習するための設計を含んでいる。Reddit の人気度は投稿時間や投稿者の属性に強く相関することが知られている (Jaech et al. 2015)。人気度の推定モデルを談話行為の推定にそのまま適用しても、4 章で詳細を述べる実験で示すように性能は限定的である。

3 提案モデル

提案モデルである Tree-LSTM GCN Hybrid の概要を図 2 に示す。Tree-LSTM GCN Hybrid ではまずスレッド中の投稿を LSTM と最大プーリングでエンコード (Comment Encoder) し、投稿の特徴量を得る。次に、投稿間の関係を木構造の LSTM 処理 (Parent-Branch Tree-LSTM, Child-Sum Tree-LSTM) およびグラフ構造の畳み込み処理 (GCN) で取り入れた投稿の特徴量を得る。投稿の特徴量にはさらに注意機構 (Self-Attention) を適用し、最終的には全結合層でラベルに接続 (Label Classifier) する。

Tree-LSTM GCN Hybrid では、既存の LSTM、木構造の LSTM、グラフ構造の畳み込みの処理を一部変更した上で組み合わせて、議論の分析に効果的なモデルを提案している。返信元から返信先に展開される議論のフローを取り入れるために、LSTM の 1 ステップ前の参照処理を修正した Parent-Branch Tree-LSTM を用意した。返信先から返信元に展開される議論のフローを取り入れるために、Tai et al. (2015) の Child-Sum Tree-LSTM を用意した。返信の方向に依

図 2 提案モデル Tree-LSTM GCN Hybrid の概要. このモデルではスレッドに含まれる投稿を Comment Encoder でエンコードし, それらを木構造学習層 (Parent-Branch Tree-LSTM, Child-Sum Tree-LSTM) およびグラフ構造学習層 (GCN) で処理する.

存しない周辺で展開される議論のフローを取り入れるために, Defferrard et al. (2016) の GCN を用意した. 本章の以降では, 各構成要素の詳細について述べる.

Comment Encoder

Comment Encoder では投稿単位の処理を実現するために, 投稿中の単語から投稿の特徴量を求める. タイトルテキスト $\mathbf{a}_{title}$ と投稿テキスト $\mathbf{a}_{1...I}$ からなる系列データを入力とする. 各テキスト中の単語は, Embedding Layer の埋め込み行列 $\mathbf{E}$ によって $\mathbf{x}_{title}$ および $\mathbf{x}_{1...I}$ に変換する. 変換した入力, 双方向 LSTM²により以下の状態遷移関数に基づき処理する:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i) \quad (1)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o) \quad (2)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f) \quad (3)$$

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c) \quad (4)$$

$$\mathbf{c}_t = \mathbf{i}_t \odot \tilde{\mathbf{c}}_t + \mathbf{f}_t \odot \mathbf{c}_{t-1} \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (6)$$

ここで $\mathbf{i}_t$ は入力ゲート, $\mathbf{o}_t$ は出力ゲート, $\mathbf{f}_t$ は忘却ゲート, $\tilde{\mathbf{c}}_t$ はメモリセルゲート, $\mathbf{c}_t$ はメモリセルの状態, $\mathbf{h}_t$ は隠れ状態, $\mathbf{W}_*$ と $\mathbf{U}_*$ は重み行列, $\mathbf{b}_*$ はバイアス項, σ はロジスティックシグモイド関数, $\odot$ は要素積の演算子である. 双方向 LSTM の出力は結合し, 時間方向の最大プーリング(Max Pooling)により投稿の特徴量 $\mathbf{m} = \max(\vec{\mathbf{h}} \parallel \overleftarrow{\mathbf{h}})$ を得る. ここで, $\vec{\mathbf{h}}$ は $\mathbf{h}$ の順方向での処理結果, $\overleftarrow{\mathbf{h}}$ は $\mathbf{h}$ の逆方向での処理結果, $\parallel$ はテンソルの結合を意味する. また, 最初の投稿については, タイトルの出力 ($\mathbf{h}_{title}$) と投稿の出力 ($\mathbf{h}_1$) の要素和 ($\oplus$) を $\vec{\mathbf{h}}_{title} \parallel \overleftarrow{\mathbf{h}}_{title} \oplus \vec{\mathbf{h}}_1 \parallel \overleftarrow{\mathbf{h}}_1$ のようにして, 最大プーリングの前に求める.

Parent-Branch Tree-LSTM

Parent-Branch Tree-LSTM では親ノード (返信元) の情報を用いた談話行為の推定を可能にする. 投稿の特徴量は根ノードから葉ノードへと処理する. LSTM の 1 時刻前の隠れ状態 $\mathbf{h}_{t-1}$ (式 1-4) を親ノードの隠れ状態 $\mathbf{h}_{parent}$ に置き換えて, 以下の式に基づき処理する:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{parent} + \mathbf{b}_i) \quad (7)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{parent} + \mathbf{b}_o) \quad (8)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{parent} + \mathbf{b}_f) \quad (9)$$

² タイトルを持つ最初の投稿と以降の投稿を別途処理できるように, 最初の投稿用の LSTM_s と返信の投稿用の LSTM_c を用意した.

$$\tilde{\mathbf{c}}_t = \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{parent} + \mathbf{b}_c) \quad (10)$$

なお、投稿には複数の子ノード（返信）がありえるため、 $\mathbf{h}_{parent}$ は複数の投稿で共有されることがある。また、以降の処理では $\mathbf{h}$ を Parent-Branch Tree-LSTM が出力する投稿の特徴量 $\mathbf{r}_P$ として用いる。

Child-Sum Tree-LSTM

Child-Sum Tree-LSTM では子ノード（返信）の情報を用いた談話行為の推定を可能にする。投稿の特徴量は葉ノードから根ノードへと Child-Sum Tree-LSTM(Tai et al. 2015) で処理する。具体的には、式 1, 2, 4 に含まれる 1 時刻前の隠れ状態 $\mathbf{h}_{t-1}$ を子ノードの隠れ状態 $\tilde{\mathbf{h}}_t$ に置き換え、忘却ゲート $\mathbf{f}_t$ （式 3）とメモリセルゲート $\mathbf{c}_t$ （式 5）を以下の式に置き換え処理する：

$$\tilde{\mathbf{h}}_t = \sum_k \mathbf{h}_k \quad (11)$$

$$\mathbf{f}_{tk} = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_k + \mathbf{b}_f) \quad (12)$$

$$\mathbf{c}_t = \mathbf{i}_t \odot \tilde{\mathbf{c}}_t + \sum_k \mathbf{f}_{tk} \odot \mathbf{c}_k \quad (13)$$

ここで $\mathbf{h}_k$ は子ノードの隠れ状態、 $\mathbf{c}_k$ は子ノードのメモリセル状態である。また、以降の処理では $\mathbf{h}$ を Child-Sum Tree-LSTM が出力する投稿の特徴量 $\mathbf{r}_C$ として用いる。

GCN

GCN では周辺ノード（返信元および返信先）の情報を用いた談話行為の推定を可能にする。各ノードの周辺ノードを畳み込みフィルタでグラフ上の畳み込みとして取り込む。具体的には、畳み込みフィルタをチェビシェフ展開で近似する手法 (Hammond, Vandergheynst, and Gribonval 2011) に基づく、グラフ畳み込み処理 (Defferrard et al. 2016) を用いる。投稿の特徴 $\mathbf{h}_l$ を入力として、以下の状態遷移関数で処理する：

$$\mathbf{h}_{l+1} = \mathbf{U} g_\theta(\mathbf{\Lambda}) \mathbf{U}^T \mathbf{h}_l \quad (14)$$

$$g_\theta(\mathbf{\Lambda}) = \sum_k^{K-1} \theta_k T_k(\tilde{\mathbf{\Lambda}}) \quad (15)$$

ここで l は層の番号、 $\mathbf{U}$ は入力グラフにおける正規化されたラプラシアンの一階基底、 $\theta_k \in \mathbb{R}^K$ はチェビシェフ係数のベクトル、 $T_k(\tilde{\mathbf{\Lambda}})$ は k 次のチェビシェフ多項式である。グラフ上の畳み込み処理は連続的に適用することにより多層化でき、提案モデルでは式 16 のようにグラフ畳み込み処理 2 層と正規化線形関数 (ReLU) を組み合わせた投稿の特徴量 $\mathbf{r}_G$ を得る：

$$\mathbf{r}_G = \mathbf{U} g_\theta(\mathbf{\Lambda}) \mathbf{U}^T \text{ReLU} \left(\mathbf{U} g_\theta(\mathbf{\Lambda}) \mathbf{U}^T \mathbf{m} \right) \quad (16)$$

Self-Attention

Parent-Branch Tree-LSTM (P), Child-Sum Tree-LSTM (C), GCN (G) に対応する $j \in \{P, C, G\}$ 番目の特徴量を, 自己注意方法の注意機構 (Yang, Yang, Dyer, He, Smola, and Hovy 2016; Lin, Feng, dos Santos, Yu, Xiang, Zhou, and Bengio 2017)(Self-Attention) で統合する. 注意機構で統合した i 番目の投稿の特徴量 $\mathbf{s}_i$ は, 議論のフローを取り入れた特徴量 $\mathbf{r}_{ji}$ の重み付き和として以下のように求める:

$$\mathbf{s}_i = \sum_{j \in \{P, C, G\}} \alpha_{ji} \mathbf{r}_{ji} \quad (17)$$

$$\alpha_{ji} = \frac{\exp(\mathbf{v}_\alpha^T \mathbf{u}_{ji})}{\sum_{j' \in \{P, C, G\}} \exp(\mathbf{v}_\alpha^T \mathbf{u}_{j'i})} \quad (18)$$

$$\mathbf{u}_{ji} = \tanh(\mathbf{W}_\alpha \mathbf{r}_{ji} + \mathbf{b}_\alpha) \quad (19)$$

ここで $\mathbf{v}_\alpha$ は重みベクトル, $\mathbf{W}_\alpha$ は重み行列, $\mathbf{b}_\alpha$ はバイアス項である.

Label Classifier

各投稿の特徴量 $\mathbf{s}_i$ を, 全ての i で共有された全結合層とソフトマックス関数を用いて以下のようなラベル出力 $\sigma(\mathbf{y}_i)$ を得る:

$$\mathbf{y}_i = \mathbf{W}_d \mathbf{s}_i + \mathbf{b}_d \quad (20)$$

$$\sigma(\mathbf{y}_i) = \frac{\exp(\mathbf{y}_i)}{\sum_{i'} \exp(\mathbf{y}_{i'})} \quad (21)$$

ここで $\mathbf{W}_d$ は重み行列, $\mathbf{b}_d$ はバイアス項である.

4 実験

4.1 ベースライン

提案手法のベースラインとして, ルールに基づくモデル (Rule 5-ACTS), 従来のパターン素性と系列学習を組み合わせたモデル (CRF Vote), 2 種類のニューラルモデル (LSTM-CRF Vote, Greaph-LSTM) を用意した.

Rule 5-ACTS

機械学習に基づかないベースラインとして、単純なルールに基づく分類器 Rule 5-ACTS を用意した。スレッド中の投稿 c は Algorithm 1 に基づいてラベル l に分類される。なお、この分類器では投稿が *Agreement*, *Disagreement*, *Humor*, *Negative Reaction* の4種類のラベルに分類されることはない。

CRF Vote

Zhang et al. (2017) で最も良い性能を示した、条件付き確率場 (CRF) に基づくモデルを CRF Vote として実装した。このモデルではスレッドを分解して、根ノードから各葉ノードへの系列を独立して処理する。内容語、句読点、パターン、著者、スレッド、コミュニティの素性を CRF で学習する。素性の中でもパターン素性の大きな効果がアブレーションテストにより確認されている。なお、スレッドを系列に分解して処理するため、一つの投稿に対して複数の異なるラベルを分類することがある。この際には、投稿のラベルは分類したラベルの多数決によって決定する。

Algorithm 1 Rule 5-ACTS

```

if  $c$  is 最初の投稿 then
  if  $c$  が疑問符 “?” を含む then
     $l \leftarrow \textit{Question}$ 
  else
     $l \leftarrow \textit{Announcement}$ 
  end if
else { $c$  は最初の投稿ではない}
  if  $c$  が疑問符 “?” を含む then
     $l \leftarrow \textit{Question}$ 
  else
    if  $c$  が “thank”, “thanks”, “thx”, “thxs”, “tks” のいずれかをを含む then
       $l \leftarrow \textit{Appreciation}$ 
    else
      if  $c$  の上位投稿に Question がある then
         $l \leftarrow \textit{Answer}$ 
      else
         $l \leftarrow \textit{Elaboration}$ 
      end if
    end if
  end if
end if

```

LSTM-CRF Vote

Comment Encoder (3 章) に LSTM 層と CRF 層を組み合わせ、CRF Vote をニューラルモデルに拡張した LSTM-CRF Vote を実装した。系列学習タスクにおいて、LSTM と CRF を組み合わせるアプローチが有効であることは知られている (Huang, Xu, and Yu 2015; Lample, Ballesteros, Subramanian, Kawakami, and Dyer 2016; Ma and Hovy 2016)。また、CRF Vote と同様にスレッドは根ノードから葉ノードへの系列に分解する。このモデルではまず各系列を Comment Encoder で処理し、各投稿の特徴量を得る。次に LSTM 層により投稿間の依存関係を導入し、さらに CRF 層でラベル間の相関関係を導入する。また、複数のラベルが分類された投稿については、CRF Vote と同様に多数決でラベルを決定する。

Graph-LSTM

グラフ構造に対応した LSTM (Zayats and Ostendorf 2018) を Graph-LSTM として実装した。スレッド上の投稿間の親子関係および兄弟関係を処理できるように、LSTM を拡張して親子間の接続を扱う忘却ゲートと兄弟間の接続を扱う忘却ゲートを用意している。また、投稿中の単語分散表現の平均に加えて、グラフ上の位置や返信数といったパターン素性を組み合わせた入力を用意している。このモデルは談話行為の分類を目的として設計されていないが、Reddit 投稿の人気度を推定する際には優れた性能を発揮している。

4.2 データ・評価手法

提案モデルとベースラインモデルの評価には Zhang et al. (2017) のデータ³を用いた。データは Reddit の 9,438 スレッド、115,827 投稿、2,837 コミュニティ (subreddit) より構成され、各投稿には表 1 に示される 10 種類の談話行為が 3 人のアノテータによりアノテートされている。Zhang et al. (2017) の設定を踏襲し、アノテータ間の多数決でラベルを決定できない投稿および多数決で決定したラベルが *Other* の投稿を除外した。結果として、評価には 9,131 スレッド、98,865 投稿のデータを用いた。各談話行為のラベルの数、それらの最初の投稿の割合、Krippendorff's Alpha による一緻度を表 2 に示す。

すべてのモデルは 10 交差検定で評価し、評価尺度としては Accuracy, Precision, Recall, F_1 値を用いた。ニューラルモデルについては、開発データを用意するために 10 交差検定ではデータを 8:1:1 の割合で訓練:開発:テストに分割した。テスト時の性能を測定する際には、開発データで最良の F_1 値が得られた学習結果を利用した。

³ <https://github.com/google-research-datasets/coarse-discourse>

表 1 各談話行為とその概要

談話行為	概要
Answer	事前の投稿に出現した質問もしくは依頼に対する回答.
Elaboration	事前の投稿に対しての追加情報.
Question	質問もしくは依頼.
Appreciation	事前の投稿に対しての感謝, 興奮, 賞賛等.
Agreement	事前の投稿に対しての何らかの合意.
Disagreement	事前の投稿に対しての修正, 批判, 対比, 反対等.
Humor	冗談, しゃれ, 皮肉等.
Announcement	ニュース, 引用, 物語, 見解, 批評, 洞察等.
Negative Reaction	事前の投稿に対しての攻撃, 嘲り, 嫌悪, 怒り等.
Other	上記の 9 種類の談話行為に該当しない.

表 2 談話行為の数, 最初の投稿の割合, Krippendorff's Alpha による一致度

談話行為	投稿数	最初の投稿の割合	Krippendorff's Alpha
Answer	40,723	0.00%	0.785
Elaboration	18,513	0.00%	0.401
Question	17,105	41.32%	0.823
Appreciation	8,380	0.00%	0.611
Agreement	4,815	0.00%	0.426
Disagreement	3,263	0.00%	0.383
Humor	2,291	2.71%	0.343
Announcement	2,002	100.00%	0.732
Negative Reaction	1,773	0.00%	0.330
全体	98,865	—	0.645

4.3 モデルの設定

単語分散表現の事前学習

ニューラルモデルで用いる単語分散表現は, 2006–2016 の Reddit ダンプデータ^{4,5}をサンプリングしたデータで事前学習した. サンプリングでは, 実験データ中に表れる 2,837 コミュニティ (4.2 節) に属する, 約 2.3 億の投稿を取得した. 事前学習では word2vec (Mikolov, Sutskever, Chen, Corrado, and Dean 2013) を skip-gram アルゴリズムで次元 = 100, 学習率 = 0.025, ウィンドウ幅 = 5, ネガティブサンプル数 = 5, エポック = 5 の設定で用いた.

各層のユニット数と最大単語数

提案モデルやベースラインモデルには, ユニット数のパラメータを設定できる層が含まれてい

⁴ https://bigquery.cloud.google.com/dataset/fh-bigquery:reddit_posts

⁵ https://bigquery.cloud.google.com/dataset/fh-bigquery:reddit_comments

る. LSTM のユニット数としては, $LSTM_S$ と $LSTM_C$ については 300 に, LSTM-CRF Vote については 600 に設定した. Parent-Branch Tree-LSTM, Child-Sum Tree-LSTM, GCN のユニット数には 600 を設定した. また, Reddit の投稿には数千語にもおよぶ長い投稿が存在する. 効率化のために, スレッドの最初の投稿については最大 400 語, 返信の投稿については最大 100 語のみを利用した.

最適化手法

LSTM-CRF Vote, Graph-LSTM, Tree-LSTM GCN Hybrid の最適化には確率的勾配降下法を用いた. 目的関数には LSTM-CRF Vote は CRF のスコアを用い, Graph-LSTM と Tree-LSTM GCN Hybrid は交差エントロピーを用いた. 確率的勾配降下法のパラメータとしては, 学習率は $\{0.01, 0.1\}$ から開発データで良い性能を示した値を選択し, モメンタム = 0.9, 勾配クリッピング = 3.0 を用いた. また, 過学習を避けるためにドロップアウト (Srivastava, Hinton, Krizhevsky, Sutskever, and Salakhutdinov 2014) を割合 0.5 で LSTM 層と GCN の中間層に導入した.

4.4 実験結果

表 3 に各モデルの評価結果を示す. また, 提案モデルである Tree-LSTM GCN Hybrid については, 一つの構成要素のみを用いたモデルを提案モデル (構成要素名) として用意している. 結果から, Tree-LSTM GCN Hybrid が従来の CRF Vote を Accuracy で 1.5%, F_1 値で 2.2 上回る性能を示している. LSTM-CRF Vote も CRF Vote を上回る性能を示しており, ニューラルモデルへの単純な拡張およびニューラルモデルにより議論のパターンを学習する有効性が示唆される⁶. Graph-LSTM は CRF Vote に及ばない性能しか得られておらず, Reddit を対象として

表 3 提案モデルとベースラインモデルの Accuracy, Precision, Recall, F_1 値

モデル	Accuracy	Precision	Recall	F_1 値
Rule 5-ACTS	0.590	0.513	0.590	0.536
CRF Vote (Zhang et al. 2017)	0.764	0.752	0.764	0.748
LSTM-CRF Vote	0.769	0.757	0.769	0.759
Graph-LSTM (Zayats and Ostendorf 2018)	0.738	0.720	0.738	0.724
提案モデル (Parent-Branch Tree-LSTM)	0.772	0.760	0.772	0.763
提案モデル (Child-Sum Tree-LSTM)	0.631	0.609	0.631	0.611
提案モデル (GCN)	0.767	0.755	0.767	0.756
提案モデル (Tree-LSTM GCN Hybrid)	0.779	0.768	0.779	0.770

Precision, Recall, F_1 値については, 9 つの談話行為をデータ数に応じて重み付けした平均値である. 太字は各評価指標で最大の値を示している.

⁶ すべての評価指標において, Tree-LSTM GCN Hybrid と LSTM-CRF Vote 間に有意水準 5% で統計的な有意差があることを Fisher-Pitmann Permutation Test で確認している. また, LSTM-CRF Vote と CRF Vote 間にお

いても人気度の推定モデルを談話行為の分類にそのまま用いるのは難しいことが分かる。Rule 5-ACTS では他のモデルを大きく下回る性能になっており、単純なルールではこのタスクに対応しきれないことを示している。なお、計算時間については、Tree-LSTM GCN Hybrid は 10 交差検定の評価を完了するのに NVIDIA Titan X gpu を 1 個用いて約 21 時間掛かった。ニューラルモデルは一般的に計算負荷が高く、本実験においても Intel Core i7 cpu を 1 コア用いて約 4 時間掛かった CRF Vote の約 5 倍の時間が掛かっている。

提案モデルの単一構成要素を用いた結果については、構成要素ごとに異なる傾向が得られた。Parent-Branch Tree-LSTM が最も高い性能、GCN が Parent-Branch Tree-LSTM を僅かに下回る性能、Child-Sum Tree-LSTM が他二つより大幅に低い性能を示した。Reddit の議論では根ノードから葉ノードへと返信の連鎖で議論が進む。このため、親ノードを参照する構成要素が高い性能を示しているこれらの結果は直感に合う。また、GCN では 2 層の畳み込み層により、長さ 2 以内のノードを参照している。GCN で Parent-Branch Tree-LSTM に近い性能が得られており、談話行為の分類においては遠いノードの影響は僅かであると考えられる。

5 考察

5.1 木構造・グラフ構造の有効な組み合わせ

提案モデルでは三つの構成要素を注意機構を用いて統合している。Tree-LSTM GCN Hybrid がこれら構成要素をどのように組み合わせているかを分析するために、注意機構が出力する確率値を分析した。図 3a に、すべての投稿についての確率値から求めた確率密度関数を示す。最

図 3 注意機構の確率値から求めた確率密度関数

いても、Accuracy については有意水準 10%, Precision, Recall, F_1 値については有意水準 5% で統計的な有意差があることを確認している。

も優先される構成要素は Parent-Branch Tree-LSTM であり, 4.4 節の実験結果で単独で最も高い性能が得られた構成要素でもある, しかし, 2 番目に優先される構成要素は, 単独では低い性能を示した Child-Sum Tree-LSTM となった. このため, 木構造やグラフ構造の効果は, 単独で用いた場合と組み合わせた場合では必ずしも一致しないことが分かる. 図 3b に最初の投稿に限定した場合の確率密度関数を示す. 最初の投稿には親ノードがなく, 全体の場合と比較して Child-Sum Tree-LSTM と GCN が優先されている.

注意機構の確率値をさらに返信の数ごとに場合分けした結果を図 4 に示す. 返信がないか少ない場合は, 求めた確率密度関数は全体と似た傾向を示した (図 4a). 返信が中程度の場合には, Child-Sum Tree-LSTM と GCN のピークの位置が Parent-Branch Tree-LSTM に近くなった (図 4b). 返信数が大きい場合には, GCN のピークの位置が Child-Sum Tree-LSTM を上回る結果となった (図 4c, 4d). GCN の結果から, 返信数が多ければ周辺の投稿のみで談話行為の分類が十分に行えることが伺える.

図 4 返信の数ごとに分割した確率密度関数

5.2 ニューラルモデルの強み

提案モデルのより詳細な効果を確認するために、談話行為毎の性能を確認した。表 4 に Rule 5-ACTS, CRF Vote, LSTM-CRF Vote, Parent-Branch Tree-LSTM, Tree-LSTM GCN Hybrid の談話行為毎の F_1 値を示す。結果として、ニューラルモデルがほぼすべての談話行為について、従来モデルと比較して同等以上の性能を達成している。特に *Disagreement*, *Humor*, *Negative Reaction* については、CRF Vote から大きく性能が向上した。これらの談話行為は他と比べて発生頻度が小さく（表 2）、ニューラルモデルが低頻度の談話行為の特徴をうまく捉えていることを示唆している。

Announcement においては、複数の構成要素を組み合わせたモデル (Tree-LSTM GCN Hybrid) が 3.2 ポイントの F_1 値の向上を示した。*Announcement* は最初の投稿にしか表れない特徴があり（表 2）、5.1 節の分析でも Child-Sum Tree-LSTM と GCN が最初の投稿において優先されていた。提案モデルが改善を示さなかった唯一の談話行為としては、*Appreciation* が挙げられる。表 5 に示される *Appreciation* の Precision, Recall, F_1 値を確認したところ、Precision において

表 4 提案モデルとベースラインモデルの各談話行為についての F_1 値

談話行為	Rule 5-ACTS	CRF Vote	LSTM-CRF Vote	Parent-Branch Tree-LSTM	Tree-LSTM GCN Hybrid
Answer	0.685	0.884	0.892	0.897	0.900
Elaboration	0.325	0.647	0.657	0.660	0.667
Question	0.754	0.853	0.869	0.870	0.877
Appreciation	0.575	0.730	0.704	0.713	0.720
Agreement	0.000	0.460	0.458	0.462	0.475
Disagreement	0.000	0.219	0.280	0.259	0.299
Humor	0.000	0.180	0.285	0.269	0.290
Announcement	0.679	0.787	0.779	0.787	0.819
Negative Reaction	0.000	0.189	0.280	0.302	0.313

太字は各談話行為で最大の値を示している。

表 5 *Appreciation* についての提案モデルとベースラインモデルの Precision, Recall, F_1 値

モデル	Precision	Recall	F_1 値
Rule 5-ACTS	0.836	0.439	0.575
CRF Vote (Zhang et al. 2017)	0.794	0.676	0.730
LSTM-CRF Vote	0.721	0.689	0.704
提案モデル (Parent-Branch Tree-LSTM)	0.722	0.703	0.713
提案モデル (Tree-LSTM GCN Hybrid)	0.750	0.693	0.720

太字は各評価指標で最大の値を示している。

は Rule 5-ACTS および CRF Vote が優れた性能を示していた。この結果から、強い言語的な手掛かり（例，“thank”）が存在する談話行為については、ルールやタスク依存の素性を含むモデルに強みがあることが伺える。

CRF Vote に対して提案モデル (Tree-LSTM GCN Hybrid) の推定結果が改善した例を図 5 に示す。a の ID2 ではスポーツ選手の怪我に対して意見を述べているが、誤って Elaboration と推定されていた談話行為が正しく Negative Reaction と推定されるようになった。提案モデルにより、ID2 の投稿は親投稿 (ID1) を補足しているのではなく、親投稿に対しての個人的かつ否定的な見解であることを判定できていることが確認できる。また、b の ID3 では国ごとの製品価格の

図 5 CRF Vote に対して Tree-LSTM GCN Hybrid の推定結果が改善した例。例中では、人名を “PERSON”，具体的な URL を “URL”，製品名を “PRODUCT”，数値を “XXXX” に置き換えている。また、例中の CRF Vote, Tree-LSTM GCN Hybrid はそれぞれのモデルの推定結果である。

違いについて意見を述べているが、誤って Answer と推定されていた談話行為が正しく Humor と推定されるようになった。提案モデルにより、ID3 の投稿は親投稿 (ID1) の質問に答えているというより、親投稿の内容に即した冗談であることを判定できていることが確認できる。

6 おわりに

本稿では議論のパターンをニューラルネットワークを用いて学習し、談話行為を分類するモデルを提案した。提案モデルでは 3 種類の構成要素 (Parent-Branch Tree-LSTM, Child-Sum Tree-LSTM, GCN) を用意し、議論のパターンを学習することを試みた。評価実験により、提案モデルは従来モデルと比較して Accuracy で 1.5%, F_1 値で 2.2 の性能向上を確認した。また、3 種類の構成要素間の関係を注意機構の確率値を通じて分析した。今後の課題としては、提案モデルを Reddit 以外の議論に対して適用することを検討している。Reddit には大規模な投稿データを容易に入手できるという利点があるが、投稿時間や投稿者の属性等のテキスト以外の要素が投稿に大きな影響を与えている。議論を分析するための柔軟なアーキテクチャを探索するためにも、今後は提案モデルの拡張をより一層進めたい。

謝 辞

東京工業大学奥村・高村研究室の皆様には、ニューラルネットワークを用いて木構造やグラフ構造を学習する手法について、議論を通じて様々なアドバイスを頂きました。ここに感謝の意を表します。また、本論文は The 27th International Conference on Computational Linguistics (COLING 2018) に採択された、Integrating Tree Structures and Graph Structures with Neural Networks to Classify Discussion Discourse Acts (Miura, Kano, Taniguchi, Taniguchi, Misawa, and Ohkuma 2018) を日本語で書き直し内容を追加したものです。

参考文献

- Bhatia, S., Biyani, P., and Mitra, P. (2012). “Classifying User Messages For Managing Web Forum Data.” In *Proceedings of the 15th International workshop on the Web and Databases*.
Bunt, H., Alexandersson, J., Carletta, J., Choe, J.-W., Fang, A. C., Hasida, K., Lee, K., Petukhova, V., Popescu-Belis, A., Romary, L., Soria, C., and Traum, D. (2010). “Towards an ISO Standard for Dialogue Act Annotation.” In *Proceedings of the 7th Conference on International Language Resources and Evaluation*, pp. 2548–2555.
Buntain, C. and Golbeck, J. (2014). “Identifying Social Roles in Reddit Using Network Struc-

- ture.” In *Proceedings of the Workshop on Modeling Social Media: Mining Big Data in Social Media and the Web*, pp. 615–620.
- Carvalho, V. and Cohen, W. (2006). “Improving “Email Speech Acts” Analysis via N-gram Selection.” In *Proceedings of the Analyzing Conversations in Text and Speech*, pp. 35–41.
- Carvalho, V. R. and Cohen, W. W. (2005). “On the Collective Classification of Email “Speech Act”.” In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 345–352.
- Cheng, H., Fang, H., and Ostendorf, M. (2017). “A Factored Neural Network Model for Characterizing Online Discussions in Vector Space.” In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2296–2306.
- Cohen, W. W., Carvalho, V. R., and Mitchell, T. M. (2004). “Learning to Classify Email into “Speech Acts”.” In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 309–316.
- Defferrard, M., Bresson, X., and Vandergheynst, P. (2016). “Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering.” In *Advances in Neural Information Processing Systems 29*, pp. 3844–3852. Curran Associates, Inc.
- Ferschke, O., Gurevych, I., and Chebotar, Y. (2012). “Behind the Article: Recognizing Dialog Acts in Wikipedia Talk Pages.” In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 777–786.
- Hammond, D. K., Vandergheynst, P., and Gribonval, R. (2011). “Wavelets on Graphs via Spectral Graph Theory.” *Applied and Computational Harmonic Analysis*, **30** (2), pp. 129–150.
- He, J., Ostendorf, M., and He, X. (2017). “Reinforcement Learning with External Knowledge and Two-Stage Q-functions for Predicting Popular Reddit Threads.” *arXiv preprint arXiv:1704.06217*.
- He, J., Ostendorf, M., He, X., Chen, J., Gao, J., Li, L., and Deng, L. (2016). “Deep Reinforcement Learning with a Combinatorial Action Space for Predicting Popular Reddit Threads.” In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1838–1848.
- Hessel, J., Lee, L., and Mimno, D. (2017). “Cats and Captions vs. Creators and the Clock: Comparing Multimodal Content to Context in Predicting Relative Popularity.” In *Proceedings of the 26th International Conference on World Wide Web*, pp. 927–936.
- Hochreiter, S. and Schmidhuber, J. (1997). “Long Short-Term Memory.” *Neural Computation*, **9** (8), pp. 1735–1780.
- Hu, J., Passonneau, R., and Rambow, O. (2009). “Contrasting the Interaction Structure of an

- Email and a Telephone Corpus: A Machine Learning Approach to Annotation of Dialogue Function Units.” In *Proceedings of the SIGDIAL 2009 Conference*, pp. 357–366.
- Huang, Z., Xu, W., and Yu, K. (2015). “Bidirectional LSTM-CRF Models for Sequence Tagging.” *arXiv preprint arXiv:1508.01991*.
- Jaech, A., Zayats, V., Fang, H., Ostendorf, M., and Hajishirzi, H. (2015). “Talking to the Crowd: What Do People React to in Online Discussions?” In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 2026–2031.
- Kim, S. N., Cavedon, L., and Baldwin, T. (2010a). “Classifying Dialogue Acts in One-on-One Live Chats.” In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 862–871.
- Kim, S. N., Wang, L., and Baldwin, T. (2010b). “Tagging and Linking Web Forum Posts.” In *Proceedings of the 14th Conference on Computational Natural Language Learning*, pp. 192–202.
- Kipf, T. N. and Welling, M. (2017). “Semi-Supervised Classification with Graph Convolutional Networks.” In *Proceedings of the 5th International Conference on Learning Representations*.
- Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., and Dyer, C. (2016). “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.
- Lim, W. H., Carman, M. J., and Wong, S.-M. J. (2017). “Estimating Relative User Expertise for Content Quality Prediction on Reddit.” In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pp. 55–64.
- Lin, Z., Feng, M., dos Santos, C. N., Yu, M., Xiang, B., Zhou, B., and Bengio, Y. (2017). “A Structured Self-attentive Sentence Embedding.” In *Proceedings of the 5th International Conference on Learning Representations*.
- Liu, F., Baldwin, T., and Cohn, T. (2017). “Capturing Long-range Contextual Dependencies with Memory-enhanced Conditional Random Fields.” In *Proceedings of the 8th International Joint Conference on Natural Language Processing*, pp. 555–565.
- Ma, X. and Hovy, E. (2016). “End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1064–1074.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). “Distributed Representations of Words and Phrases and their Compositionality.” In *Advances in Neural Information Processing Systems 26*, pp. 3111–3119.
- Miura, Y., Kano, R., Taniguchi, M., Taniguchi, T., Misawa, S., and Ohkuma, T. (2018). “Inte-

- grating Tree Structures and Graph Structures with Neural Networks to Classify Discussion Discourse Acts.” In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3806–3818.
- Omuya, A., Prabhakaran, V., and Rambow, O. (2013). “Improving the Quality of Minority Class Identification in Dialog Act Tagging.” In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 802–807.
- Socher, R., Karpathy, A., Le, Q. V., Manning, C. D., and Ng, A. Y. (2014). “Grounded Compositional Semantics for Finding and Describing Images with Sentences.” *Transactions of the Association for Computational Linguistics*, **2**, pp. 207–218.
- Socher, R., Lin, C. C.-Y., Ng, A. Y., and Manning, C. D. (2011). “Parsing Natural Scenes and Natural Language with Recursive Neural Networks.” In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, pp. 129–136.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). “Dropout: A Simple Way to Prevent Neural Networks from Overfitting.” *Journal of Machine Learning Research*, **15**, pp. 1929–1958.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R. A., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., and Meteer, M. (2000). “Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech.” *Computational Linguistics*, **26** (3), pp. 339–373.
- Tai, K. S., Socher, R., and Manning, C. D. (2015). “Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks.” In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1556–1566.
- Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., and Lee, L. (2016). “Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions.” In *Proceedings of the 25th International Conference on World Wide Web*, pp. 613–624.
- Wang, L., Lui, M., Kim, S. N., Nivre, J., and Baldwin, T. (2011). “Predicting Thread Discourse Structure over Technical Web Forums.” In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 13–25.
- Wang, Y.-C., Joshi, M., and Rose, C. (2007). “A Feature Based Approach to Leveraging Context for Classifying Newsgroup Style Discussion Segments.” In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 73–76.
- Wei, Z., Liu, Y., and Li, Y. (2016). “Is This Post Persuasive? Ranking Argumentative Com-

- ments in Online Forum.” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 195–200.
- Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., and Hovy, E. (2016). “Hierarchical Attention Networks for Document Classification.” In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 1480–1489.
- Zayats, V. and Ostendorf, M. (2018). “Conversation Modeling on Reddit Using a Graph-Structured LSTM.” *Transactions of the Association for Computational Linguistics*, **6**, pp. 121–132.
- Zhang, A., Culbertson, B., and Paritosh, P. (2017). “Characterizing Online Discussion Using Coarse Discourse Sequences.” In *Proceedings of the 11th AAAI International Conference on Web and Social Media*, pp. 357–366.
- Zhang, R., Gao, D., and Li, W. (2011). “What Are Tweeters Doing: Recognizing Speech Acts in Twitter.” In *Proceedings of the AAAI-11 Workshop on Analyzing Microtext*, pp. 86–91.
- Zhang, R., Gao, D., and Li, W. (2012). “Towards Scalable Speech Act Recognition in Twitter: Tackling Insufficient Training Data.” In *Proceedings of the Workshop on Semantic Analysis in Social Media*, pp. 18–27.

略歴

三浦 康秀：2004 年電気通信大学大学院電気通信学研究科修士課程修了。同年、富士ゼロックス株式会社入社。現在、コミュニケーション技術研究所において、自然言語処理の研究開発に従事。また併せて、2016 年より東京工業大学工学院情報通信系博士後期課程に在学中。言語処理学会会員。

狩野 竜示：2013 年東京大学情報理工学系研究科修士課程修了。同年、富士ゼロックス株式会社入社。現在、コミュニケーション技術研究所において、自然言語処理の研究開発に従事。

谷口 元樹：2009 年北海道大学大学院情報科学研究科修士課程修了。同年、富士ゼロックス株式会社入社。現在、コミュニケーション技術研究所において、自然言語処理の研究開発に従事。

谷口 友紀：2004 年神戸大学大学院自然科学研究科修士課程修了。同年、富士ゼロックス株式会社入社。現在、コミュニケーション技術研究所において、自然言語処理の研究開発に従事。

三沢翔太郎：2016 年早稲田大学大学院創造理工学研究科修士課程修了。同年、富士ゼロックス株式会社入社。現在、コミュニケーション技術研究所におい

て, 自然言語処理の研究開発に従事.

大熊 智子: 1994 年東京女子大学文理学部卒業. 1996 年慶應義塾大学政策・メディア研究科修士課程修了. 同年, 富士ゼロックス株式会社入社. 現在, コミュニケーション技術研究所において, 自然言語処理と情報推薦の研究に従事. 言語処理学会編集委員. 博士 (学術).

(2018 年 7 月 26 日 受付)

(2018 年 10 月 29 日 再受付)

(2018 年 12 月 10 日 採録)