6S-03

SANOVA RNN: 低頻度な対話行為の特徴を考慮する対話行為推定モデル

泉春乃† 加藤昇平†‡

†名古屋工業大学 大学院工学研究科情報工学専攻
‡名古屋工業大学 情報科学フロンティア研究院

1 はじめに

雑談可能な対話システムにおける対話内容の理解や自然な応答生成のための重要な技術の1つに，対話行為推定がある．対話行為推定とは，対話内のある1文の内容を対話行為から選択することである．対話行為とは話者の発話の意図であり，本研究では『質問 (YesNo)』や『あいづち』など，JAIST タグ付き自由対話コーパス [1] に定義された 9 種で構成される．もし対話システムがユーザ発話の『応答 (YesNo)』を推定できれば，その発話が以前の質問や確認の発話への返答であることを推測可能となり，対話システムは発話の話題に関して掘り下げることができる．対話行為を推定する利点として，このような文脈に沿う応答生成への活用が挙げられる．

対話行為推定の従来研究としては，Bi-RNN を用いた Bothe ら [2] や，JAIST タグ付き自由対話コーパスを定義した福岡ら [1] などが挙げられる．これら従来研究から，過去の発話内容は対話行為推定における重要な特徴の 1 つであることが推察できる．また雑談対話において各対話行為である発話の出現頻度は対話行為ごとに大きく異なるため，従来の推定手法では頻度の高い対話行為に偏って推定する傾向がある．

実際に対話システムにおける応答生成の手がかりとして対話行為推定を導入する場合，推定結果が偏ることは生成される応答が偏る原因となり得る．そこで我々は，頻度の低い対話行為の特徴を捉えることで様々な対話行為を推定する手法を提案している．本稿では提案手法による実験結果について考察を行う．

2 対話行為推定手法

我々の提案している対話行為推定手法は，2 層の RNN，9 種の 2 値分類層および Attention 機構を組み合わせたネットワークモデルである Self-Attention Networks One-Versus-All Recurrent Neural Networks (SANOVA RNN) である．

会話内の 1 文を 1 発話，連続した発話の系列を 1 対話として，ある 1 対話における t 番目の発話 s_t の対話行為推定を目的とする．ここでは，1 発話の対話行為推定に用いる連続した過去の発話の総数を文脈長 L と表現する．図 1 に SANOVA RNN のネットワーク構造を示す．本手法では，ネットワークにおける各層の学習を下層から順に分割して行う．本手法における RNN には Bidirectional LSTM を，本手法における Attention 機構には Additive Attention を用いる．

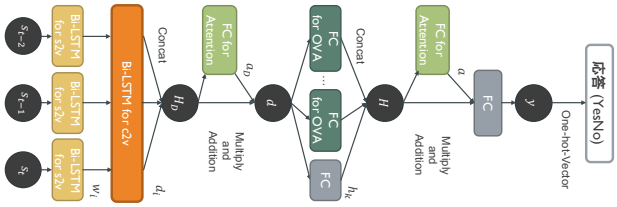

図 1: SANOVA RNN

2.1 学習：Bi-LSTM for sentence2vec

発話 s_t のみから対話行為を推定するネットワーク (以下 Bottom とする) の学習を行う．Bottom とは，入力層，図 1 における **Bi-LSTM for s2v** および出力層とする 2 層の Fully Connected (FC) から構築された 4 層の Neural Networks (NN) である．発話 s_t を形態素などで分かち書きし，word2vec により変換した分散表現を入力として Bottom を学習させることで，Bi-LSTM for s2v は入力された 1 発話の内容から対話行為をある程度推定できるように特徴を捉えたベクトルを出力すると考える．

2.2 学習：Bi-LSTM for context2vec

任意の文脈長の発話を用いて対話行為を推定するネットワーク (以下 Top とする) の学習を行う．Top とは，入力層，図 1 における **Bi-LSTM for c2v**，**FC for Attention**，**FC** および 1 層 FC から構築される 5 層 NN である．学習済みの Bi-LSTM for s2v を用いて変換した $s_{max(1,t-L+1)}$ から s_t までの各発話を入力として Top を学習させることで，Top は Bottom では考慮できなかった過去の発話の特徴を推定に用いることが可能となる．また FC for Attention から出力される Attention Weight(図 1 中 a_D) は，推定における各発話の重要度とみなせる．

2.3 学習：One-Versus-All

発話 s_t がある対話行為であるか否かを推定する 2 値分類層（One-Versus-All）の学習を対話行為数分行う．(図 1 中 FC for OVA）学習には 2 層の FC を用いて，入力は図 1 中 d，出力は対応する対話行為に対する 1 次元の確率とする．各層に対応した対話行為である発話とそうでない発話を同数用いた学習により，全発話を用いた学習における低頻度な対話行為の特徴が埋もれるという問題を軽減する．

2.4 学習：Attention and output

2 値分類層第 1 層の各出力 u 次元ベクトルおよび Top の出力層第 1 層の出力 u 次元ベクトルを結合した $10 \times u$ 次元行列 H を入力として，10 次元ベクトルの Attention Weight を学習する．学習した Attention Weight を用

SANOVA RNN: Dialogue Act Classification Considering Rare Utterances

Haruno IZUMI†, Shohei KATO†‡
†Dept. of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology
‡Frontier Research Institute for Information Science, Nagoya Institute of Technology
†‡Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan
{izumi, shohey}@katolab.nitech.ac.jp

表 1: 単語分散表現の比較

	分かち書き器	学習ファイル	語彙数	F1-score
wiki	MeCab	2.6 GB	576,521	0.613
twitter	Sentencepiece	0.5 GB	34,611	0.635
hottoSNS	MeCab, Juman	63.4 GB	2,067,629	0.611

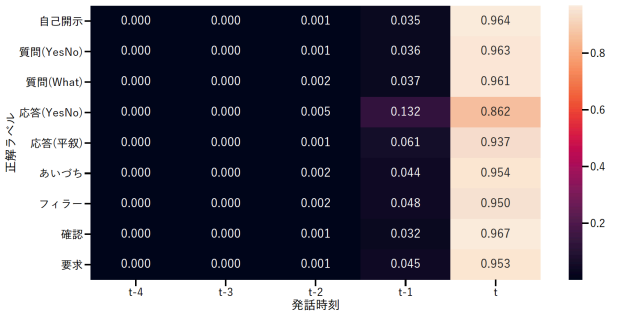

図 2: 各発話への Attention Weight の平均

いて行列 H の重み付き和をとった u 次元ベクトルから対話行為を推定する．上記の学習により，入力された発話によって注目する層を選択することが可能となり，過去の発話を考慮した上で対話行為の出現頻度に左右され難い推定を行うことができると考える．

3 実験

3.1 実験設定

モデルの学習及び実験のための対話行為タグが付与された雑談対話コーパスとして，福岡らの提案した JAIST タグ付き自由対話コーパス [1] を用いる．コーパスに収録されている 92,020 発話の対話行為をそれぞれ推定することを目的とする．全 97 対話から発話数ができるだけ均一になるよう 19 対話または 20 対話ずつ抽出することにより，対話データの 5 分割を行ったものを用いて，5 分割交差検証を実施した．

3.2 単語分散表現の比較

入力に用いる単語分散表現として，以下の 3 種を比較する．1) **wiki**: 日本語 Wikipedia の全記事から学習した word2vec．2) **twitter**: 2019 年 4 月に投稿されたツイートの一部から作成した Sentencepiece および同データから学習した word2vec．3) **hottoSNS**: 株式会社ホットリンクにより作成された，SNS データ等 Web ページから学習した word2vec．

表 1 に，各分散表現の学習のための分かち書き手法，ファイルサイズ，word2vec の語彙数と提案モデル ($L = 5$) による推定性能を示す．各 F1-score に対応のある t 検定を実施したところ，twitter モデルと他 2 手法間において有意な差 ($p < 0.01$) が確認された．よって以降の実験には twitter モデルを用いる．

3.3 対話行為推定に必要な文脈帳の考察

図 2 に，発話 s_t の推定における各時刻の発話への Attention Weight の平均を正解ラベルごとに示す．この図より，特に応答 (YesNo) である発話を分類する際は過去の発話が重要であり，またどの対話行為も時刻 t から $t-2$ までを用いて分類していることがわかる．よって実験における文脈長 $L = 5$ は妥当である．

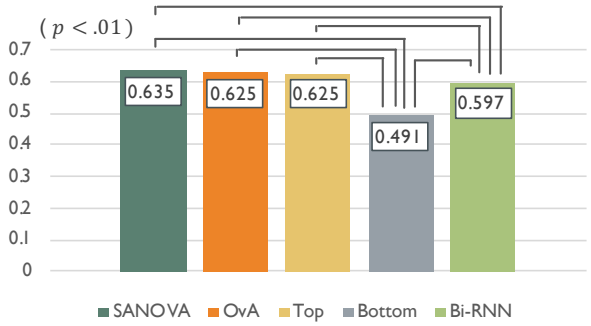

図 3: 各手法における F1-score

3.4 対話行為推定性能のモデル間比較

提案手法 (以下 SANOVA) の対話行為推定性能を検証するために，比較手法には以下の 4 つを用いる．

- **Bi-RNN**: Bothe ら [2] 提案．本稿 Top における Bi-LSTM を Bi-RNN に置き換えたモデル．
- **Bottom**: 1 発話から対話行為を推定する．(2.1 節)
- **Top**: 文脈を考慮して対話行為を推定する．(2.2 節)
- **OVA**: 各 2 値分類層の出力を当該対話行為の確率とみなし，最大を推定結果とする．(2.3 節)

それぞれの手法について F1-score を算出し，対応のある t 検定を実施した．図 3 にその結果を示す．

Bottom は時刻 t の発話推定時に当該発話のみを用いるが，他の手法は時刻 $t-4$ から t の発話を用いる．また Bottom は他手法と比較すると，F1-score が有意に低い．これより過去の発話を用いることは対話行為推定に効果的であることが確認された．

Bi-RNN は Bottom 以外の他 3 手法と比較すると，F1-score が有意に低い．3 手法は Bi-RNN と異なり，単語および文脈を理解する層に Bi-LSTM を組み込んでいる．よって Bi-LSTM は Bi-RNN よりも対話行為推定モデルに有用だと考えられる．

Top，OVA と SANOVA における F1-score を比較すると，SANOVA が他 2 手法をわずかに上回る．SANOVA は Top および OvA どちらの出力も考慮できる手法であるため，本研究の目的である頻度の低い対話行為の推定性能を向上させながら，対話全体では最も高い推定性能になったと考えられる．

4 おわりに

本稿では，提案手法 SANOVA RNN が出現頻度の低い対話行為の特徴をよく捉えながら，対話全体でも高い推定性能をもつことを確認した．今後はロボットアバターのジェスチャのための対話行為推定モデルに本手法を採用することにより，テキストチャットにおけるリアルタイムのロボットジェスチャを実現する．

参考文献

[1] 福岡知隆, 白井清昭：対話行為に固有の特徴を考慮した自由対話システムにおける対話行為推定, 自然言語処理, Vol. 24, No. 4, pp. 523–547 (2017).

[2] Bothe, C., Magg, S., Weber, C. and Wermter, S.: Conversational Analysis Using Utterance-level Attention-based Bidirectional Recurrent Neural Networks, *Interspeech 2018* (2018).