

対話行為における少数派タグ分類のための多段階分類法 A Multi-step Classification Method Considering Minority Tags in The Dialogue Act Classification

宮城 孝明¹⁾ 遠藤 聡志²⁾
Takaaki Miyagi Satoshi Endo

1 はじめに

近年、企業活動において対話システムを導入することが一般的になっている。この対話システムをタスク指向型対話システムと呼び、ある目的を達成するための対話システムである。一方でユーザと雑談するだけの対話システムを非タスク指向型対話システムと呼ぶ。対話システム構築に統計手法やルールベース手法が用いられた。Vinyals ら [1] は、単純なモデルと大規模な会話データセットで簡単な会話だが生成を可能にした。非タスク指向型対話システムは、自由な話題でユーザを楽しませるの事を目的としており、対話システムに欠かせない重要な役割である [2]。自由な対話システムの実現には、対話内容を把握し適切な応答を生成する必要がある。対話内容の理解や応答文生成のための情報 [3][4] として対話行為が用いられる。対話行為は発話者の発話の「内容」や「意図」、「役割」を表しており、その種類分けを対話行為タグと呼ぶ [13]。対話行為分類は、発話や応答に対しての対話行為を分類する問題である。応答の対話行為分類を行い、その対話行為タグを応答文生成に応用する研究 [3] もある。また、Cervone ら [9] や Kumar ら [10] の研究では、応答生成時に対話行為を用いる事の有効性を示しており、応答の対話行為タグを分類することで、応答文生成の精度向上に期待できる。そのため、本研究では応答の対話行為分類問題を扱う。

昨年、我々が [8] 示した応答の対話行為分類結果では偏った分類予測となってしまう。本研究で使用している Switchboard Dialog Act Corpus[5](以下、SwDA) データセットでは、「Statement」タグと「Understanding」タグが全体の 7 割を占めており、分類結果もほとんどがこの 2 つとなってしまう。本研究ではこの問題を解決するために、多段階分類とコスト考慮型学習の 2 つの手法を提案する。多段階分類では、多数派タグとその他のタグを階層的に繰り返し分類する事で、少数派タグが分類できる事を期待する。コスト考慮型学習は、少数派タグの損失値を大きくする事で、少数派タグを多数派タグに誤分類する事を減らせると期待する。2 つの手法を応用するため、損失関数の重みの調整、データセットや分類器を変更した。これらを通し偏りが無い分類が行えるのか検証する。

2 関連技術

2.1 多段階分類

多段階分類は、複数の分類器を重ね段階応じて適切なクラス分類を行う手法である。本研究では、多数派タグ

- 1) 琉球大学大学院理工学研究科工学専攻知能情報プログラム, Graduate School of Engineering and Science, University of the Ryukyus
- 2) 琉球大学工学部工学科知能情報コース, Computer Science and Intelligent Systems, University of the Ryukyus

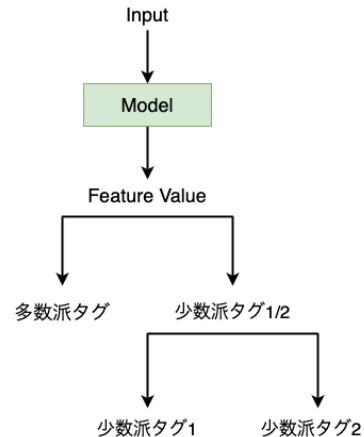


図1 本研究で行う多段階分類の概要

とグルーピングしたタグの二値分類を段階的に繰り返し行い、多数派タグの特徴量を把握し分類を行えるようにする。そして、最終層では少数派タグだけの分類を行い、少数派タグの特徴量を把握できる事を目指す。図1では、本研究で行う多段階分類の例を示す。分類するタグを多数派タグと少数派タグ 1/2 とする。1 段階目は多数派タグと少数派タグ 1/2 の二値分類を行い、2 段階目は少数派タグ1、少数派タグ2の二値分類を行う。

2.2 コスト考慮型学習

一般的な分類問題では誤分類による損失値は全て同等と扱っているが、現実では「不良品を良品と分類、良品を不良品と分類」などの分類問題も存在し、損失値を同等に扱えない場合がある。「不良品を良品と分類」した場合、不良品を良品として扱ってしまう恐れがある。そのため、不良品を良品と誤分類した場合、損失値を増加させる事で誤分類を起きにくくする。このような場合にコスト考慮型学習が用いられている。コスト考慮型学習では、各クラスの誤分類時のコスト設定を行い学習を通して、損失値が小さくなるようにする手法である。そのため、コスト考慮型は偏ったクラスに重みの調整を加える事ができる。よって、不均衡なデータセットの問題に対応するための提案手法の一つとして扱い、多段階分類と比較しながら議論する。

3 先行研究

3.1 応答の対話行為分類

応答の対話行為分類は、入力された発話文に対しての応答文の対話行為タグを予測する問題である。大原ら [14] は単語系列を処理する Recurrent Neural Network(以下、RNN) と発話系列を処理する RNN を組み合わせた階層型 RNN を用いて解いた。田中ら [11] は階層型 RNN に対して対話行為系列、対話行為系列を考慮する層を提案し、前後の対話行為の関係性を学習させ精度向上した。我々は、田中らの対して単語系列と発話系列の箇所自己注意機構を用いた Combination

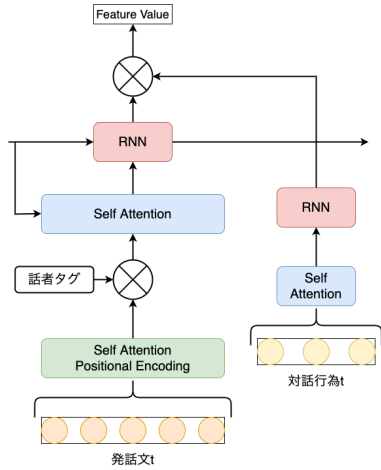


図 2 Cmb Attention モデルの概要

Attention(以下, Cmb Attention) モデルを提案した [8]. Cmb Attention モデルを図 2 に示す. 自己注意機構を用いる事で, 重要な単語同士や発話同士の関係性を考慮した分類を可能にし, 田中らのモデルより精度が向上した. しかし, 不均衡なデータセットのため分類が偏ってしまい, 全てのタグをバランス良く分類できない問題が発生した. そのため, 偏りのあるタグに対して重みを付け加えたり, そのタグの特徴量を捉え分割できるようにする事で問題解決を図る.

4 提案手法

少数派タグの特徴表現を獲得し, 分類が行えるようにするため, 多段階分類とコスト考慮型学習を行う. 以下のセクションでは, 本研究で提案した多段階分類とコスト考慮型学習を示す.

4.1 多段階分類

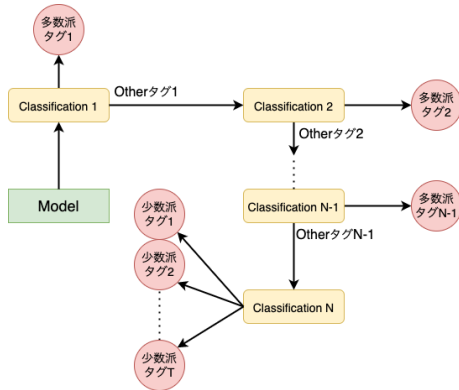


図 3 本研究の多段階分類手法

本研究での多段階分類手法の概要を図 3 に示す. 図 3 では, 分類器は N 個, Other タグは $N-1$ 個, 多数派タグは $N-1$ 個, 少数派タグは T 個ある. そして, Other タグ 1 は多数派タグ 2 から $N-1$ と全ての少数派タグをまとめたものであり, Other タグ 2 は多数派タグ 3 から $N-1$ と全ての少数派タグをまとめたもの, Other タグ $N-1$ は全ての少数派タグをまとめたものとなる. 分類の動作では, 多数派タグが残っているまでは段階的に複数回の二値分類を行い, 少数派タグのみの場合多値分類を行う. 図 3 では, 分類器 1 から $N-1$ までは二値分類, 分類器 N で多値分類を行う. 次に, データセットと分類器の提案を示す.

4.1.1 多段階用のデータセット構成

表 1 各段階のデータセットの割合

タグ名(割合)	何段階目	段階毎の割合
Statement(51.1%)/Other1	1 段階目	51.1%/48.9%
Understanding(26.4%)/Other2	2 段階目	54.1%/45.9%
Uninterpretable(12.4%)/Other3	3 段階目	55.1%/44.9%
Question(6.2%)/Agreement(2.7%)/Apology(0.5%)	4 段階目	61.6%/27.1%/4.8%
Greeting(0.3%)/Other(0.2%)/Directive(0.2%)	4 段階目	3.0%/1.9%/1.6%

本研究で使用したデータセット SwDA は, 2 者の電話会話を文字に書き起こしている. また, 本来は対話行為タグ数は 42 個となっているが, 我々の研究では 9 個にまとめた. 図 3 の分類を行うため, データセットを新たに修正し, 各タグ同士のバランスを調整した. 表 1 は, 各段階毎のデータセットのタグとその割合, 段階毎のタグ同士の割合を示す. 多数派タグは全体数をクラス数で割った値より多いタグと定義し, 少数派タグはそれ以下とする. そして, 1 段階目から 3 段階目まで多数派タグとその他タグとし, 4 段階目では少数派タグのみのデータセットを作成した.

4.1.2 分類器

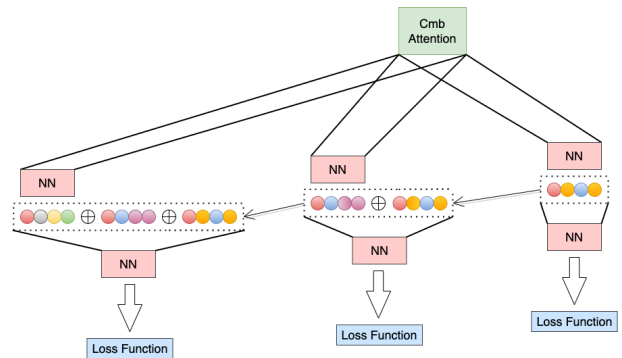


図 4 分類器の概要

全てをクラスを同一で扱う Flat な分類では, 全てのクラスを分類することである. そのため, Flat な分類では多段階分類問題を解く事ができない. 本章では, 多段階分類を行えるための分類器を提案する. 段階的に分類を行うために, 上位層の表現情報を下位層に受け渡す事が必要だと考えられるため, Dehong Gao ら [6] が提案した手法を用いた. 図 4 では, 3 つの分類器を重ねた分類器の概要を示す. Cmb Attention が出力した特徴量表現を各段階毎の Neural Network(以下, NN) に入力し, 段階毎の表現情報を獲得する. そして, 現在の階層と前の階層の表現情報を連結させ, 次の NN に入力し, 前の階層の表現を考慮した特徴表現を得る. 各段階で出力した特徴表現から損失値を求め, 段階毎に加算しその値を最終的な損失値とする.

4.2 コスト考慮型学習

本研究で使用しているデータセットは不均衡なデータセットのため, モデルの分類ではほとんどが多数派タグになってしまう. そこで, タグによって損失値を変動させ, 予測しにくいタグの学習の重みを大きくする. 使用した損失関数は CrossEntropyLoss であり, 各タグに対しての重みを設定した. 式.1 は, クラスの重みを含めた損失関数の式となる. これにより, 各タグに対して重み設定を行い損失値を変動させる.

$$\text{loss}(x, \text{class}) = -w[\text{class}] \log\left(\frac{\exp(x)[\text{class}]}{\sum_j \exp(x[j])}\right) \quad (1)$$

また、各クラスの重みでは

- コスト考慮型学習 v1: $\log(\text{全体数}/\text{各タグ数})$
多数派タグの誤分類コストと比べ、少数派タグの誤分類コストを上げるため、全体数を各タグ数で割った。しかし、そのままではコスト値が膨大なため、対数関数を用い扱いやすい値にした。
- コスト考慮型学習 v2: 1.0, 1.0, 2.0, 2.0, 3.0, 3.0, 4.0, 4.0, 4.0
「Statement」タグと「Understading」タグの損失値は増減させずに、他のタグを多い順に一定間隔でコスト値を大きくした。検証段階で、Recall や F 値に向上が見られたコスト値を示した。今後も更なる改善が見られるか随時変更し検証する。

5 実験

本実験では、提案モデルを学習させその精度評価を行う。提案モデルの有効性を示すため、baseline として Cmb Attention モデルを用いた。評価手法では、Accuracy と F 値, Recall, Precision を用いる。今回の実験は予測精度が重要な点となるため、Recall に焦点をあてる。

5.1 実験設定

本実験で使用したハイパーパラメータは以下となる。

- 単語 Embedding の次元数: 256
- 自己注意機構/RNN の隠れ層の次元数: 512
- 対話行為 Embedding/隠れ層の次元数: 128

また、語彙のサイズを 20000 とし、未知の単語が出現した場合は「UNK」と設定した。そして、多段階分類の損失関数は 1 から 3 段階目までは Binary Cross Entropy, 4 段階目では Cross Entropy とした。最適化関数は、Adam(学習率 0.00005)を用いた。本実験では、訓練データ、検証データ、評価データの比率を 8:1:1 とし、訓練回数を 30epoch で行いその際の検証データが最小損失値になったモデルの重みパラメータを評価データに使用した。

5.2 結果と考察

5.2.1 各モデルの比較結果

表 2 各モデルの分類結果

モデル	Accuracy	F 値	Precision	Recall
CmbAttention	59.53%	19.24%	27.24%	19.95%
コスト考慮型学習 v1	52.52%	21.17%	22.39%	22.83%
コスト考慮型学習 v2	57.94%	21.17%	24.49%	21.77%
多段階分類	47.34%	15.22%	20.73%	15.56%

表 2 は各モデルの分類結果を示している。コスト考慮型学習は精度向上した部分もあるが、多段階分類は全ての評価において低下した。コスト考慮型学習は、期待していた Recall は 2.3% 近く向上しており、さらに F 値も微細だが向上した。このことから、コスト考慮型学習は不均衡なデータセットでも上手く対応した学習が可能であると考えられる。なぜ、多段階分類が不均衡なデータセットに対応できないかその原因の調査するため、各段階の精度比較と混同行列を求めた。

表 3 各層の分類結果

評価手法	第 1 段階	第 2 段階	第 3 段階	第 4 段階
Accuracy	58.45%	33.27%	3.20%	1.41%
F 値	58.42%	22.41%	3.96%	5.66%
Precision	58.42%	38.08%	46.10%	25.39%
Recall	58.42%	20.95%	2.08%	3.63%

5.2.2 多段階分類の結果の考察

各段階での分類状況を示す事で低下の原因を把握する。表 3 では、各段階毎の分類結果である。第 1 段階目は半数以上が正解となり比較的に高い結果となっている。しかし、第 2 段階目以降から急激に低下し、第 4 段階目では Accuracy は 1.41% であり Recall は 3.63% である。段階を追う毎に精度が落ちている事がわかる。段階毎の誤分類の影響を受けているのか調査するため、各タグの分類予測と正解を確認する混同行列を求めた。

図 5 は、多段階分類の混同行列である。第 1 段階目の分類では、少数派タグである「Question」タグ、「Apology」タグ、「Other」タグ、「Directive」タグの半数以上が「Statement」タグと誤分類している。また、第 2 段階目でも「Understading」タグと誤分類しているタグもある。そのため、第 4 段階目の分類を行う前に残っているタグが少なく、少数派タグの Recall 向上につながらない結果になった。多段階分類によって、多数派タグの特徴量を把握し、分類しやすくする事が目的だったが、結果的には Cmb Attention よりも低くなった。多数派タグの特徴量の中に少数派の特徴量も含まれていたため、分類結果が良くならなかったと考えられる。

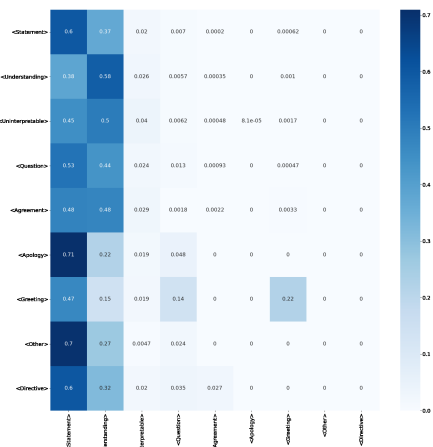


図 5 多段階分類の混同行列

5.2.3 コスト考慮型学習の結果と考察

コスト考慮型学習について各タグの Recall を求め表 4 に示す。少数派タグの分類は CmbAttention と比べ、どちらも向上している。損失値に対して重みを乗算させる事で、これまで分類しにくかったタグでさえ分類が可能になった。つまり、損失値を大きくする事でそのタグの重要性を高め、分類の際に重要となる特徴量を獲得できていたからだと考えられる。少数派タグである「Question」タグや「Agreement」タグ、「Apology」タグはそれぞれ向上の余地があると見えるため、それぞれの各タグの

重み調整を行う必要がある。一方の「Greeting」タグは、30%程度付近となっているためこれ以上の向上は難しい。また、「Directive」タグは 0.10%だが向上しているが、「Other」タグは 0.00%と向上の余地が見られない。全体の特徴量から全てのタグの分類精度が向上するか特徴量を可視化し判断を行う。

表 4 各タグの Recall

タグ	CmbAttention	コスト考慮型学習 v1	コスト考慮型学習 v2
Statement	75.27%	58.31%	72.85%
Understanding	76.30%	79.13%	69.30%
Uninterpretable	7.47%	11.12%	17.79%
Question	0.20%	5.24%	1.20%
Agreement	4.01%	6.32%	8.14%
Apology	0.00%	16.63%	0.31%
Greeting	16.34%	30.09%	26.20%
Other	0.00%	0.00%	0.00%
Directive	0.00%	0.00%	0.10%

図 6 は、コスト考慮型学習 v1 で求めた対話特徴量を tSNE[12] によって可視化した結果である。この結果より、多数派タグは纏まっているように見えるが少数派タグは散り散りとなっており、多数派タグに隠れてしまっている。そのため、各タグに重みをつけ学習を行っても、タグの特徴量を捉える事にも限りがあると考えられる。また、多段階分類では複数回の分割を通して、ごく一部のタグのみしか分類できないため、多段階分類では精度が低下したと考えられる。さらなる精度向上を目標とするためには、新たなデータをモデルに入力として与える必要がある。発話文の感情表現によって発話者の心理状態を把握することが挙げられる。しかし、発話文のみでは限界があるため、別案として外部情報を取得しそれを考慮する手法が考えられる。例えば、発話者の表情や仕草、ペルソナなどから発話者の思考を取り入れる。これらの有効な研究について今後調査を行い、実験に取り入れていきたい。

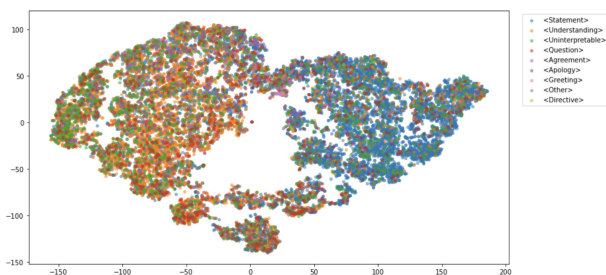


図 6 tSNE による対話特徴量の可視化

6 今後の展望

本研究では応答の対話行為分類の問題となる少数派タグの分類にくさを解決すべく多段階分類とコスト考慮型学習を提案し、その有効性について検証した。実験の結果、多段階分類は全ての評価指標で低下し、コスト考慮型学習は Recall と F 値の向上を確認した。多段階分類

の精度低下の原因として、多数派タグの特徴量を上手く捉えきれず分割できなかった事が挙げられる。tSNE でも確認したように、各タグが一つに纏まらず散り散りになっている事が大きな要因だと考えられる。コスト考慮型学習では、偏ったデータセットに対して有効的に働く事が確認できたため、今後この手法について調査する。しかし、コスト考慮型学習にも限界があるため、更なる向上のため発話者情報について有効性の検討を行う。

参考文献

- [1] Vinyals, Oriol, and Quoc Le. "A neural conversational model." arXiv preprint arXiv:1506.05869 (2015).
- [2] 小林峻也, and 萩原将文. "ユーザの嗜好や人間関係を考慮する非タスク指向型対話システム." 人工知能学会論文誌 (2016): DSF-502.
- [3] Zhao, Tiancheng, Ran Zhao, and Maxine Eskenazi. "Learning discourse-level diversity for neural dialog models using conditional variational autoencoders." arXiv preprint arXiv:1703.10960 (2017).
- [4] 河野誠也, 吉野幸一郎, and 中村哲. "対話行為を用いた制御可能なニューラル対話モデルの検討." 人工知能学会全国大会論文集 第 33 回 (2019). 一般社団法人 人工知能学会, 2019.
- [5] Jurafsky, Daniel, and Elizabeth Shriberg. "Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13 Daniel Jurafsky*, Elizabeth Shriberg+, and Debra Biasca** University of Colorado at Boulder &+ SRI International." (1997).
- [6] Gao, Dehong, et al. "Deep hierarchical classification for category prediction in e-commerce system." arXiv preprint arXiv:2005.06692 (2020).
- [7] Raheja, Vipul, and Joel Tetreault. "Dialogue act classification with context-aware self-attention." arXiv preprint arXiv:1904.02594 (2019).
- [8] 宮城孝明, and 遠藤聡志. "自己注意機構を用いた応答の対話行為推定." IEICE Conferences Archives. The Institute of Electronics, Information and Communication Engineers, 2021.
- [9] Cervone, Alessandra, Evgeny Stepanov, and Giuseppe Ricciardi. "Coherence models for dialogue." arXiv preprint arXiv:1806.08044 (2018).
- [10] Kumar, Harshit, Arvind Agarwal, and Sachindra Joshi. "Dialogue-act-driven conversation model: An experimental study." Proceedings of the 27th International Conference on Computational Linguistics. 2018.
- [11] 田中昂志, 高山隼矢, and 荒瀬由紀. "対話システムにおける履歴を考慮した応答の対話行為推定." 人工知能学会全国大会論文集 第 33 回全国大会 (2019). 一般社団法人 人工知能学会,
- [12] Van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." Journal of machine learning research 9.11 (2008).
- [13] Boyer, Kristy, et al. "Dialogue act modeling in a complex task-oriented domain." Proceedings of the SIGDIAL 2010 Conference. 2010.
- [14] 大原康平, 佐藤翔悦, 吉永直樹, 豊田正史, 喜連川優. 階層型 RNN を用いた対話における応答の対話行為予測. 言語処理学会第 24 回年次大会講演論文集, 2018.