

交渉成否検知におけるファインチューニングモデル

渡邊賢[†]東京農工大学 工学部 知能情報システム工学科[†]藤田桂英[‡]東京農工大学 工学研究院 先端情報科学部門[‡]

1 はじめに

交渉は、日常生活やビジネスなど様々な場面で必要となる重要な活動である。しかし、交渉そのものや背景の複雑さから、互いの利益を最大化できるような合理的な合意を得ることは難しい。そこで、交渉者が合理的な合意を得るために、人間同士の交渉の支援についての研究が取り組まれている [2] が、交渉成否検知および予測に関する研究はまだ少ない。

本論文では、交渉ダイアログから交渉の成否を検知する交渉成否検知や交渉の初期段階でその成否を予測する交渉成否予測に取り組む。アノテーションされている交渉ダイアログデータは規模が小さいことから、既存研究 [2] にファインチューニングモデルを導入することに効果があると考えられる。そこで、交渉成否検知及び予測のタスクにおいて、学習時間を短縮し、その性能を向上させるために、ファインチューニングを用いたモデルを提案する。

2 交渉成否検知および予測タスク

交渉ダイアログ D は、 $n \in \mathbb{N}$ ターンの発話 $\{s_1, s_2, \dots, s_n\}$ からなる。ただし、発話 s は一人の交渉者によるもので、1つ以上の文章を含む。 D が与えられたときに、 D の要素数を w 、交渉進行率を r としたときに、 D の冒頭 $[wr]$ 要素からなるダイアログ D' を考える。 D' が与えられたときに、その交渉が成功しているか失敗しているかを判定する。成功している場合には0を、失敗している場合には1をラベル付けする。 $r = 1.0$ のときを交

渉成否検知と呼び、 $0.0 < r < 1.0$ のときを交渉成否予測と呼ぶ。

3 提案手法

サイズの大きなデータセットで事前学習を行なったモデルに対してファインチューニングを行うことで性能と学習効率の向上を目指す。学習モデルには、以下の3つを用いる。

- テキストベースの BERT[1] (**BERT**). BERT モデルでは、uncased(大文字と小文字を区別しない)の $BERT_{LARGE}$ を使用した。
- 対話行為ベース [2] を考慮した GRU[3] に Self-Attention 機構 [4] を無効にしたモデル (**GRU**).
- 対話行為ベース [2] を考慮した GRU[3] に Self-Attention 機構 [4] を有効にしたモデル (**GRU_{Att}**).

事前学習: 交渉成否検知のためにアノテーションされたデータは、CRAIGSLISTBARGAIN (CB), DEALORNODEAL (DN) があり、これらのコーパスのうち、最もダイアログ数及び語彙数が大きいのは CB である。そこで、CB によりモデルの事前学習に DN をテストに使用する。交渉進行率が 100%、25% の場合について、五分割の交差検証を行い、最も ROC-AUC が高かったモデルを事前学習済みモデルとした。

ファインチューニング: 事前学習に CB を用いたため、DN でファインチューニングを行う。交渉成否検知と予測の両方のタスクを対象とするため、交渉進行率が 1.0 の場合 (100%)、交渉進行率が 0.25 の場合 (25%) の二つの場合を対象とする。本手法では、訓練データに対して、テストデータと同じ交渉進行率のデータを用いる。これは学習

Fine-Tuning Models in Breakdown Detection in Negotiation Dialogues

[†]Faculty of Engineering, Tokyo University of Agriculture and Technology

[‡]Institute of Engineering, Tokyo University of Agriculture and Technology

表 1: 交渉成否検知・交渉成否予測の実験結果

交渉進行率 r		1.0		0.25	
モデル		ROC-AUC	epoch	ROC-AUC	epoch
GRU	TD	0.835	9.2	0.736	8.0
	FT	0.835	6.4	0.728	7.2
GRU _{Att}	TD	0.722	8.8	0.666	6.4
	FT	0.732	7.2	0.732	8.0
BERT	TD	0.850	8.4	0.645	8.4
	FT	0.855	6.6	0.688	5.8

時とテスト時のタスク内容が異なると、性能の低下が発生するためである。

4 評価実験

実験設定: 五分割の交差検証を行い、各手法の検知及び予測の性能とエポック数を比較する。性能は、ROC-AUC から判断する。エポック数については、上限を 10 とし、検証データによる評価で F1 スコアが三回更新されなかった時点でのエポック数の平均値とする。ファインチューニングを行なったモデル (FT) と比較するために、対象データセットのみを学習したモデル (TD) についても五分割の交差検証を行い、検出および予測性能とエポック数を計測した。

実験結果と考察: 表 1 は、交渉成否検知・交渉成否予測（交渉進行率 0.25 および 1.0）における各モデルの ROC-AUC と平均 epoch 数を示している。交渉成否検知（交渉進行率が 1.0 の場合）について、GRU, GRU_{Att}, BERT のいずれにおいても、ファインチューニングを導入した FT と導入していない TD で ROC-AUC に大きな差は出ていない。その一方で、エポック数が、TD と比較して FT では減少したことが確認できる。これらの実験結果から、交渉成否検知においては、学習時間の短縮の面ではファインチューニングの効果がある一方で、検出性能の面ではファインチューニングの効果はないといえる。検出性能の面で効果がなかった理由としては、DN が、十分な規模のデータ件数を持っていたために、対象データセットのみの学習で可能な限りの性能が出せるようになっていることが考えられる。

表 1 から、交渉進行率が 0.25 の場合について、GRU_{Att} と BERT の二つでは TD と比べ FT の ROC-AUC は上昇しており、GRU と BERT の二

つではエポック数は減少していることが確認できる。これらの実験結果から、交渉進行率が 0.25 の交渉成否予測においては、予測性能と学習時間の短縮の両方の面でファインチューニングの効果があるといえる。予測性能と学習時間の短縮の両方の面で効果が見られた理由としては、交渉成否予測が交渉成否検知と比べて難易度の高いタスクであり、DN のみでは十分に学習できなかったことが考えられる。

また、交渉進行率が 0.25 の場合、テキストベースの BERT モデルに比べて、対話行為ベースの GRU モデルの方が高い性能を示した。これは、GRU モデルは対話行為フローの冒頭が一致すれば、対話行為フローを参照することで、より正確に最終的な交渉成否が予測できるためと考えられる。

5 まとめ

本論文では、交渉ダイアログからその成否を検知するタスクにおいて、サイズの大きなデータセットを用いて事前学習を行ったモデルに対して別のデータセットでファインチューニングを行うことで、対象データセットのみの学習に比べて学習時間を短縮することができることを示した。また、交渉の成否を予測するモデルに対しては、予測性能の向上に効果があることを明らかにした。

参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT-2019*, Vol.1, pp.4171–4186, 2019.
- [2] Atsuki Yamaguchi, Kosui Iwasa, and Katsuhide Fujita. Dialogue act-based breakdown detection in negotiation dialogues. In *Proceedings of EACL-2019*, pp.745–757, 2021.
- [3] Kyunghyun Cho, Bart van Merriënboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078, 2014.
- [4] Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of ACL-2016*, Vol.2, pp.207–212, 2016.