

自然言語処理モデル BERT を用いた社内問合せ業務の効率化

富島 僚 本郷 拓人 張 嵐

株式会社テブコシステムズ アジャイルセンター

1.はじめに

近年、当社の業務量の急増に伴い社内での問い合わせも増加している。この状況により、従業員の残業時間の増加や業務効率の低下を招いており、問題の解決が急務となっている。さらに、既存の問い合わせ用の IT ツールの使用料が費用として発生しており、これを削減したいという声が上がっていた。こういった背景で、近年自然言語処理のモデルで顕著な成果を上げている BERT[1]を活用したチャットボットの導入による問い合わせ業務の負担を緩和することを検討した。

本稿では、社内問い合わせ業務に対処するために、質問に適切に回答するモデルを開発する取り組みについて説明し、そのアプローチを示す。

2.社内問い合わせ業務効率化のニーズ

問い合わせ業務に以下の課題が顕在化している。

- ① 既存の IT ツールが質問に対して適切な回答ができない現象がしばしば発生しており、問い合わせ対応の遅延や正確な情報提供の難しさといった問題を引き起こしている
- ② IT ツールの使用料によるコスト増加
上記の課題に対処し、社内問い合わせ業務の効率化を図るため、BERT という汎用事前学習モデルを採用し、少量のデータでチューニングが可能で使用料を削減しつつ、より適切な回答を得るために、自社で BERT を用いた社内問い合わせ対応のモデルを開発した。

3.BERT を用いたモデル開発のアプローチ

BERT は汎用的な自然言語処理モデルではあるが、工夫をしなければ正解率は 70~80%にとどまる。より高い正解率を目指して、適切な回答を得るために以下の観点で工夫を試みた。

- ① ハイパーパラメータチューニング
- ② 教師データの増量
- ③ 他の事前学習モデルを選定

以下、上記のアプローチのポイントを説明する。

3.1 ハイパーパラメータチューニング

BERT with SentencePiece for Japanese text [2]のモデルに対して学習率を $2e-5$ に固定してバッチサイズを変化させた場合の正解率を 5 回計測した結果の平均を取ることで探索した。正解率の平均値が最大になるようにバッチサイズと学習率の組み合わせを調査した。その結果、学習率が $2e-5$ 、バッチサイズが 32 の組み合わせが平均正解率 0.84 を達成する最適な設定であることが判明した。学習率が $2e-5$ 、バッチサイズが 16 のときも平均正解率が 0.84 であったが、バッチサイズが 32 の正解率のほうが標準偏差より小さかった。

表 1.学習率を $2e-5$ に固定したバッチサイズごとの 5 回平均正解率

#	1	2	3	4	5	6	7
バッチサイズ	2	4	8	16	32	64	128
学習率	$2e-5$	$2e-5$	$2e-5$	$2e-5$	$2e-5$	$2e-5$	$2e-5$
正解率	0.67	0.76	0.81	0.84	0.84	0.83	-
補足							メモリエラーのため打ち止め

表 2.バッチサイズを 32 に固定した学習率ごとの 5 回平均正解率

#	1	2	3	4	5	6
学習率	$2e-5$	$3e-5$	$4e-5$	$5e-5$	$6e-5$	$7e-5$
バッチサイズ	32	32	32	32	32	32
正解率	0.84	0.824	0.822	0.826	0.822	0.82

3.2 教師データの増量

当初用意していた教師データよりも多くの量が必要であったため、以下の方法でデータの増量を試みた。

- ① Easy Data Augmentation(EDA)ライブラリの適用
EDA ライブラリを使用して、テキスト内でランダムに単語を同義語で置き換えることにより、データセットの多様化を実現できた。

Efficiency Enhancement of Internal Inquiry Operations
Using the BERT Natural Language Processing Model

Ryo Tomishima, Takuto Hongou, Lan Zhang
TEPCO Systems Corporation

- ② 逆翻訳
テキストを日本語→中国語→英語→日本語など、異なる言語に翻訳し、元の言語に戻すという逆翻訳手法により、新たなデータの生成と多様性を増やすことができた。
- ③ Easy Data Augmentation(EDA)ライブラリと逆翻訳の併用
EDA ライブラリと逆翻訳を組み合わせてデータ増量を行い、正解率は 92.8%までの向上できた。

3.3 他の事前学習モデルを比較

BERT には複数モデルが存在するため、当社にて以下のモデルで性能評価を行った。

表 3. モデルの比較

モデルの出所	ランダム再現性	正解率
Pretrained Japanese BERT models [4]	確保できる	77.4%
BERT with Sentence Piece for Japanese text [2]	確保できない	92.8%
NICT BERT 日本語 Pre-trained モデル [3]	確保できない	93.5%

上記の評価結果に基づいて、検証システムの構築では平均正解率が最も高い NICT が構築したモデルを採用する方向で考えている。再現性の確保ができないモデルであることが課題だが、正解率の標準偏差も比較することにより結果の安定性を評価する必要があると考えている。

4. BERT によるモデル構築に対する考察

BERT モデルの構築を通じて、以下の可能性が確認できた。

- ① 実用に耐えうる正解率が出せた
過去に使用した少量のデータでも正解率は 90%を超えるため、問い合わせ業務の質問に対する回答について、既存 IT システムを代替可能と確認できた。
 - ② モデルの性能改善が確認できた
ハイパーパラメータチューニングによる性能改善だけでなく、データ拡張による性能改善を確認できたことにより、システムを継続運用することによって貯まるデータで正解率を改善できる可能性が確認できた。
- 性能改善を含めた BERT モデルの開発は、社内問い合わせ業務で継続運用を可能にするための出発点である。これにより、現実的に収集可能なデータ量で社内業務を効率化し、運用中に追加で集まるデータを活用して更なる性能改善を図る

ことができる。

5. まとめ

本稿にて、筆者らの少量データを用いて社内問い合わせ業務の効率化を目指し、BERT モデルを開発するアプローチについて説明した。

現段階では開発途中でありながら、90%に上る正解率を達成しており、これによって現行の問い合わせ業務で対応すべき件数を 90%程度削減できると期待している。

今後は今回の結果を受けて以下を進めることを目指している。

- ① 社内問い合わせ業務の代替を目指した実証実験を行う
- ② 更なるデータを収集し、質問応答タスクの正解率を向上させる
- ③ 改良 BERT アルゴリズムの開発を行う

社内問い合わせ業務の効率化を実現するために、正解率を向上したモデルを開発するだけでなく、適切にモデルを選定し、再現性を高め、学習データを収集して正解率を向上させる実証実験を行うことが重要である。モデルの継続的改善を実現するために、実証実験から実用の段階においても社内のユーザーを巻き込み、システムの普及を促進し、ユーザーの要求を満たす非機能要件を作りこむことが必要だと考えている。

参考文献

1. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Oct 11, 2018
2. BERT with Sentence Piece for Japanese text, GitHub, Feb 15, 2021
<https://github.com/yoheikikuta/bert-japanese>
3. NICT BERT 日本語 Pre-trained モデル, Mar 3, 2020
<https://alaginrc.nict.go.jp/nict-bert/index.html>
4. Pretrained Japanese BERT models released, 東北大学自然言語処理研究グループ, Dec 13, 2019
<https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/>