

行政対話システムにおける 検索エンジン TSUBAKI と BERT を併用した FAQ 検索の高度化

坂田 亘[†] 柴田 知秀^{‡§} 田中 リベカ^{‡§} 黒橋 禎夫^{‡§¶}

[†]LINE 株式会社

[‡]京都大学

[§]科学技術振興機構 CREST [¶]国立情報学研究所 CRIS

wataru.sakata@linecorp.com {shibata, tanaka, kuro}@nlp.ist.i.kyoto-u.ac.jp

1 はじめに

対話システムが徐々に我々の生活に浸透してきており、企業サイトへの導入のみならず、社会課題解決に向けた活用に期待が高まっている。我々はコミュニティの活性化や行政サービスの円滑化を目的とし、対話システムをベースとした行政と市民の新たなコミュニケーションチャネルの構築を目指している [8]。コミュニケーションチャネルの核となるのは行政サービスに関する問い合わせへの自動応答であり、行政サイト上の「よくある質問」(FAQ) を知識源として応答する。

一般に、ユーザからの問い合わせ(クエリと呼ぶ)と FAQ の各 QA の表現にはズレがあり、柔軟なマッチングが課題となる。我々の行政対話システムでは自動獲得した同義語を用いた柔軟なマッチングが行える検索エンジン TSUBAKI [3] を用いている。QA の各 Q を 1 文書とみなしてクエリとの柔軟なマッチングを行い、クエリに対してスコアの高い Q を含む QA をシステムの応答とする。しかし、記号マッチングでは解消できない表現のズレが存在する。

クエリと FAQ のマッチングにおいて使えるもう一つの手がかりはクエリと FAQ の A の関連度である。行政サイトには FAQ リストを提供している自治体も多いことから Q と A のペアは比較的容易に入手でき、ここから Q と A の関連度を学習することができる。Q と A の関連度を学習できれば、クエリと各 QA の A との関連度を計算し関連度が高い QA をシステム発話とすることができる。

Yahoo!知恵袋などのコミュニティ QA サイトの QA ペアを情報源として質問応答や対話を行う研究は多くあるが ([6][7])、本研究では Q と A の関連度の学習に近年提案された BERT [1] を用いる。BERT は様々な問題に応用可能な強力なモデルであり、GLUE や SQuAD などの多くのタスクで SOTA を更新してい

る。本研究の場合、Q と A を入力として Q と A の関連度を学習する。

本論文では、検索エンジン TSUBAKI によるクエリと Q の類似度計算と、BERT によるクエリと A の関連度計算を併用することにより、クエリと FAQ の柔軟なマッチングを行う手法を提案する。提案手法の概要を図 1 に示す。ユーザ発話(クエリ)に対してクエリと Q の類似度とクエリと A の関連度を併用し、対象自治体 FAQ のリストから適切な FAQ を検索し、システム発話とする。学習には全国自治体の FAQ ページから収集した QA ペアを用いる。

クラウドソーシングを利用して行政への質問を収集し実験を行ったところ、提案手法はベースラインよりも高いスコアを示すことを確認した。

2 提案手法

2.1 TSUBAKI によるクエリ・Q の類似度計算

TSUBAKI は OKAPI BM25 [2] をベースとしたモデルである。単語単位でのインデキシングの他に係り受け単位でインデキシングしており、構造的言語処理による文書検索を可能にしている。自動獲得した同義語を用いた柔軟なマッチングが可能である。本研究で利用するにあたり、行政ドメインにおける同義語を獲得し、拡充している [9]。

2.2 BERT によるクエリ・A の関連度計算

2.2.1 BERT

BERT は入力文の強力な encoding を行える Transformer [5] をベースとし、大規模な生テキストで pre-training し、各タスクで fine-tuning するモデルで文分類問題、文ペア分類、質問応答などのタスクに適

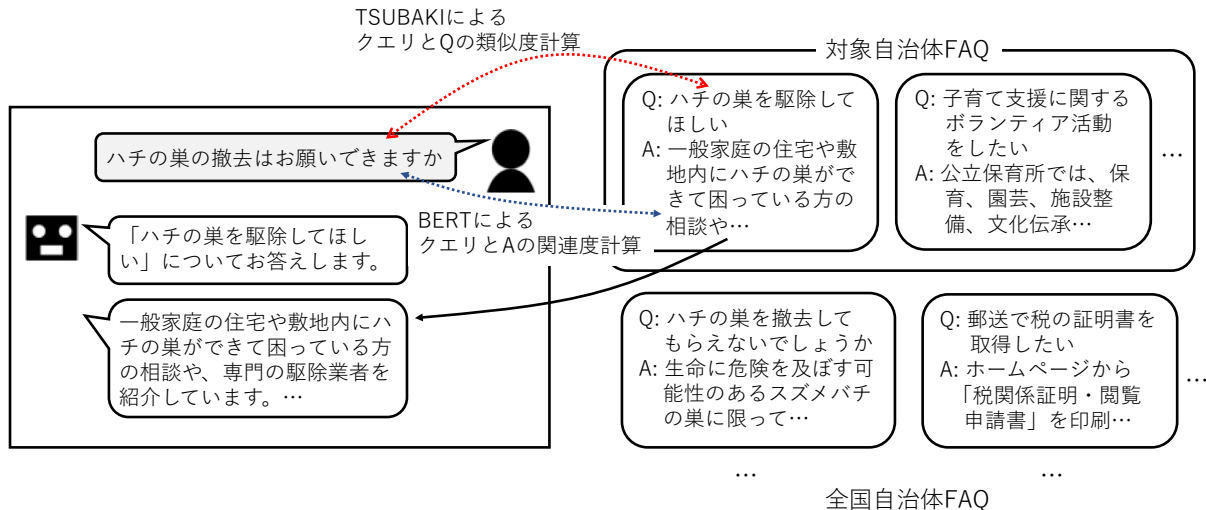


図 1: 提案手法の概要

用されている。大規模コーパスでの pre-training により、タスクのデータ量が少ない場合でも高い精度を達成することができる。本研究で扱うのは質問と回答に対する文ペア分類となる。質問と回答を連結して入力とし Transformer を適用することにより、質問と回答の関連性を賢く捉えることができる。

2.2.2 クエリ・A の関連度計算

ある自治体 G_i の FAQ ページに QA ペアが M_i 個存在するとき、各 QA ペアを $\{Q_{i,1}, A_{i,1}\}, \{Q_{i,2}, A_{i,2}\}, \dots, \{Q_{i,M_i}, A_{i,M_i}\}$ とする。学習は全自治体の全 QA ペア $Q_{i,j}, A_{i,j}$ を正例とし、 $Q_{i,j}$ に対して $A_{i',j'}$ をランダムに選択することにより生成した負例を学習データとして BERT により二値分類する。具体的には Q と A の関連度を $score(Q, A)$ と定義すると、正例に対しては $score(Q_{i,j}, A_{i,j})$ が 1 になるように、負例に対しては $score(Q_{i,j}, A_{i',j'})$ が 0 になるように学習する。

検索時は検索対象の行政自治体 (G_t とする) の QA ペアを用い、すべての QA ペアに対してユーザーのクエリ q と回答のスコア $score(q, A_{t,j})$ ($j = 1, \dots, M_t$) を計算し、スコア上位を検索結果とする。

2.3 TSUBAKI と BERT の併用

上記で導入した TSUBAKI によるクエリと Q の類似度計算と、BERT によるクエリと A の関連度計算を併用することにより、ロバストで柔軟なマッチングを実現する。併用の方法は TSUBAKI のスコアと BERT のスコアを単純に加算し、ランキングを統合するものである。もう一つの方法として、TSUBAKI のスコアが高い場合はクエリと Q で重複している語が多く正解である可能性が特に高いので、そのような QA を優

先してランキングし、残りを単純統合するものが考えられる。

具体的には TSUBAKI、BERT それぞれが出力したスコア上位 10 件を以下の 2 つの方法で統合する。

単純統合 TSUBAKI、BERT の検索結果上位 10 件のスコアを加算し、ランキングする。ただし、予備実験の結果を踏まえ TSUBAKI の検索結果のみに含まれているものを除外する。

TSUBAKI 優先統合 TSUBAKI の検索結果上位 10 件の中にスコアが閾値 α 以上のものがあり、それが BERT の上位 10 件にも含まれている場合、それらを優先的に 1 位から順にランキングし、残りの候補に対しては単純統合と同様の方法でスコアを加算しランキングする。

スコアを統合するにあたり、TSUBAKI のスコアはクエリが長いほど高くなる傾向にあるので、(クエリ中の自立語の数) $\times k_1 +$ (クエリ中の係り受けの数) $\times k_2$ を満点とみなし、この値で割って正規化する¹。

3 実験

我々が実証実験を行っている尼崎市を対象自治体とし実験・評価を行った。尼崎市の行政ページには QA ペアが約 1,700 件存在する。

3.1 評価セット

クラウドソーシングを用いて行政自治体への質問(クエリ)を収集した。クエリの例を表 1 に示す。このようなクエリが 443 件得られた。次に、各クエリに対

¹BERT が出力するスコアは 0 から 1 までの値をとるので正規化の必要はない。

戸籍謄本の発行をして欲しいのですが、代金はいくらかかりますか？
 母子健康手帳を発行して頂きたいのですが、手続きには何が必要ですか？
 新築住宅を購入することを考えているのですが、減額措置について教えて欲しいです。
 センタープールのファン送迎バスはどの駅から出ていますか

表 1: クラウドソーシングで収集したクエリの例

して TSUBAKI、BERT が 10 件出力し、1 件ずつに対してユーザのクエリに関係しているかどうかについて以下の 4 分類を人手で付与した。

- A 正しい情報が書いてある。
- B 関連情報が書いてある。
- C トピックは同一であるが関連情報が書かれていない。
- D 関係ない情報が書いてある。

TSUBAKI、BERT の出力の中に正しく回答できている QA が見つからない質問については、キーワードを変えながら検索を行い関連している QA を見つけ上記と同様の分類を付与した。関連している QA が一つも見つからない回答は評価セットから除外した。その結果、338 クエリに対する評価セットが得られた。

3.2 実験設定

評価尺度として、上位 1 件に対する精度 P@1、MRR (Mean Reciprocal Rank)、MAP (Mean Average Precision)、nDCG (normalized Discounted Cumulative Gain) を用いた。MRR、MAP、nDCG ではそれぞれ上位 10 件を対象に計算を行った。P@1、MAP、MRR はカテゴリ A、B、C すべてを正解とした。nDCG ではカテゴリ A、B、C の評価レベルをそれぞれ 3、2、1 とした。

BERT の pre-training には日本語 Wikipedia すべて (約 1,800 万文) を利用し、fine-tuning には収集した行政自治体 21 市の QA ペア計 2 万件を用いた。入力テキストには形態素解析器 Juman++ [4] を適用し単語分割を行った。

BERT の比較手法として BiLSTM を用いたモデルを採用した。Q と A それぞれを BiLSTM に入力し、Q、A の forward と backward の層を全て連結し、MLP に通すことにより 2 値分類を行った。入力層では word2vec で pretraining したものを初期値として利用した。

BERT、BiLSTM モデルともに 1 つの正例に対して 24 の負例をランダムにサンプルして学習した。2.3 節で説明した TSUBAKI のスコアの正規化の際の係数

	P@1	MAP	MRR	NDCG
TSUBAKI	0.488	0.535	0.571	0.478
Bi-LSTM	0.349	0.439	0.471	0.429
BERT	0.494	0.568	0.616	0.583
提案手法 (単純統合)	0.627	0.644	0.701	0.634
提案手法 (T 優先)	0.642	0.656	0.712	0.644

表 2: 実験結果 (提案手法の “T 優先” は TSUBAKI 優先統合を示す)

は $k_1 = 4, k_2 = 2$ とし、TSUBAKI 優先統合の閾値 α は 0.3 とした。

3.3 実験結果と考察

実験結果を表 2 に示す。まず、クエリと A の関連度のみを考慮する手法を比べると、BERT は Bi-LSTM を大きく上回っており、このタスクでの BERT の有効性を示すことができた。TSUBAKI と BERT を比べると P@1 の値はほぼ同じであるが、他の尺度では BERT が上回っており、BERT は多くの正解を返すことができていることがわかる。そして、提案手法は TSUBAKI または BERT 単体を上回っており、統合する有効性を示せた。2 種類の統合の方法を比べると、TSUBAKI を優先した統合がわずかに上回っていることがわかる。

評価例を表 3 に示す。1 つ目の例では TSUBAKI は「相談」がマッチしているだけのキャリア相談に関する QA を出力してしまっているが、BERT は正しい QA を出力できている。

2 つ目と 3 つ目はいずれも提案手法で 1 位が正解している例である。3 つ目の例については提案手法 (T 優先) で正解しているが、提案手法 (単純統合) では不正解となった例である。これは TSUBAKI で「住民票」や「どこで行う」などがマッチしていることから確信度が高く、優先して採用したことによって正解できたものであり、単純に統合すると誤ったものが 1 位となっていた。

4 つ目の例はいずれの手法でもうまく回答できなかったものである。BERT ではクエリ中の「楽器」と A 中の「音楽」の関連性を捉えられているが誤った QA を返している。正しい回答として「Q: 青少年センターの館内施設の案内・利用時間・休館日について」という項目があり、A 部分に青少年センターには音楽室があること、市民はセンター内の施設を利用できることが書かれてある。これを回答するためには A 部分のさらなる深い理解が必要である。

クエリ	TSUBAKI	BERT	提案手法 (T 優先)
職場のパワハラ相談を受けている窓口はありますか?	1 × Q: キャリア相談を受けたいのですが A: 相談場所。尼崎市竹谷町2丁目183番地出屋敷リベル3階...	1 ○ Q: 仕事関係の問題で弁護士に相談したい A: 賃金、解雇、労働災害など専門的かつ高度な労働問題について...	1 ○ Q: 弁護士の労働相談を受けることはできるのか A: 賃金、解雇、労働災害など専門的かつ高度な労働問題について...
免許の更新はどこで行えばいいですか?	1 × Q: 医療従事者の免許申請 (新規、籍訂正・書換え、再交付) はどこで行っていますか A: 医師、歯科医師、保健師についての免許申請を...	1 × Q: 【高齢者バス特別乗車証】バス券の更新には、本人がいかないとだめですか A: 原則、ご本人で申請手続きをしてください。...	1 ○ Q: 運転免許証の更新の手続きについて教えてください。 A: 運転免許証の更新手続きにつきましては...
	2 ○ Q: 運転免許証の更新の手続きについて教えてください A: 運転免許証の更新手続きにつきましては...	2 × Q: バス券の更新申請は代理人でもできますか A: 原則、ご本人で申請手続きをしてください...	2 × Q: バス券の更新申請は代理人でもできますか A: 原則、ご本人で申請手続きをしてください...
住民票が必要ですが、時間外とか土日取ることは出来ますか?	1 ○ Q: 住民票は何処 (どこ) で何時 (いつ) 取れますか A: ... 住民票は平日と土曜日に取得することができます...	1 × Q: マイナンバー付きの住民票は代理人でも取得できますか? A: マイナンバー付きの住民票は、本人または同一世帯に...	1 ○ Q: 住民票は何処 (どこ) で何時 (いつ) 取れますか A: ... 住民票は平日と土曜日に取得することができます...
	2 × Q: 住民票や戸籍謄抄本を第三者にとられることがある... A: 正当な理由があれば第三者からの請求が法により認め...	2 × Q: 住民票の記載内容を電話で問合せすることができますか。 A: お電話では住民票の内容はお答えできません...	2 × Q: マイナンバー付きの住民票は代理人でも取得できますか A: マイナンバー付きの住民票は、本人または同一世帯員に...
楽器の練習に使える場所はありますか?	1 × Q: 期日前投票所の場所はどこですか A: 期日前投票所は市内に3カ所あります...	1 × Q: 尼崎市少年音楽隊について教えてください A: 「尼崎市少年音楽隊」には、合唱隊、吹奏楽隊、...	1 × Q: 尼崎市少年音楽隊について教えてください A: 「尼崎市少年音楽隊」には、合唱隊、吹奏楽隊、...

表 3: 評価例 (表中の○×はラベル ABC を○とみなしたものである)

4 おわりに

本論文では検索エンジン TSUBAKI によるクエリと Q の類似度計算と、BERT によるクエリと A の関連度計算を併用することにより、クエリと FAQ の柔軟なマッチングを行う手法を提案し、それぞれを単独で用いるよりも高い精度を達成できることを示した。今後の課題としては前の発話を参照し適切に文脈を考慮することなどがあげられる。

謝辞

本研究は科学技術振興機構 CREST 「知識に基づく構造的言語処理の確立と知識インフラの構築」の支援のもとで行われた。行政 FAQ サイトのデータを提供していただいた東北大学乾・鈴木研究室の山口健史氏に感謝します。

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. *arXiv preprint arXiv:1810.04805*, 2018.

[2] Stephen E. Robertson, Steve Walker, Micheline Hancock-Beaulieu, Aarron Gull, and Marianna Lau. Okapi at trec. In *Text REtrieval Conference*, 1992.

[3] Keiji Shinzato, Tomohide Shibata, Daisuke Kawahara, Chikara Hashimoto, and Sadao Kurohashi. TSUBAKI: An open search engine infrastructure for developing new information access methodology. In *IJCNLP2008*, pp. 189–196, 2008.

[4] Arseny Tolmachev, Daisuke Kawahara, and Sadao Kurohashi. Juman++: A morphological analysis toolkit for scriptio continua. In *EMNLP2018*, pp. 54–59, Brussels, Belgium, 2018.

[5] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS2017*, pp. 5998–6008. 2017.

[6] Yu Wu, Wei Wu, Ming Zhou, and Zhoujun Li. Sequential match network: A new architecture for multi-turn response selection in retrieval-based chatbots. In *ACL2017*, 2017.

[7] Zhao Yan, Nan Duan, Peng Chen, Ming Zhou, Jian-she Zhou, and Zhoujun Li. Building task-oriented dialogue systems for online shopping. In *AAAI2017*, pp. 516–525, 2017.

[8] 田中リベカ, 坂田亘, 柴田知秀, 黒橋禎夫, 橋本泰一. 対話ポットをベースとした行政と市民の新たなコミュニケーションチャンネルの構築. 情報処理学会 第 81 回年次大会, 2019.

[9] 坂田亘, 柴田知秀, 黒橋禎夫. 行政サイトの FAQ ページを利用した対話フローチャートの自動生成とそれに基づく対話ポットの構築. 言語処理学会 第 24 回年次大会, pp. 65–68, 2018.