

チャットボット運用における新規問い合わせ候補の抽出

戸田 隆道 友松 祐太 杉山 雅和

株式会社 AI Shift

{toda_takamichi, tomomatsu_yuta, sugiyama_masakazu}@cyberagent.co.jp

1 はじめに

近年カスタマーサポートの分野においてチャットボットによるサポートが注目を集めており、メールにはない返信のリアルタイム性や電話と比べた際の気軽さから、現在多くの企業で導入されている。

チャットボットの初期設計や運用改善では問い合わせ内容と回答のペアをチャットボットに登録していく必要があるが、多くの場合この作業は対話ログを参照しながら人手で行っている。我々はこの作業の負担を減らすため、対話ログのクラスタリングを行い、類似する問い合わせをまとめている。加えて重要度スコアを計算し、まとめられた問い合わせの中で代表的なものを抽出している。これにより作成の手間が大幅に減ることに加え、問い合わせのボリュームゾーンもわかるので、頻出する問い合わせから優先度をつけて回答ペアの作成に取り組むことができる。

本稿では我々の取り組んでいるクラスタからの新規問い合わせ抽出手法について紹介する

2 関連研究

対話をクラスタリングする研究として東中らの研究 [1] が挙げられる。この研究はフライト予約などの特定のドメインを対象にしたタスク指向対話におけるスロットを埋めるのに必要な対話をクラスタリングするが、本稿で対象とする対話は一問一答で解決へ導くための問い合わせ候補を抽出するという点で異なっている。

文章の重要度スコアを計算する手法として Jagadeesh らの研究 [2] が挙げられる。こちらは文章要約のために文の位置や動詞の有無などをスコアリングし、情報の過不足が無いよう文章を抽出する手法で、類似文章から代表分を抽出するという点で我々の手法と異なっている。

3 手法紹介

本章ではクラスタからの新規問い合わせ候補の抽出手法について紹介する。

3.1 クラスタリング

代表的な文章特徴量の表現方法である TF-IDF を特徴量とする階層クラスタリングを行う。これにより意味的に似ている文章のまとまりを作ることができ、クラスタサイズからそのクラスタに類する問い合わせの頻度を推測することができる。

3.2 重要度スコア計算

クラスタリングの特徴量として使用した TF-IDF は文章に含まれる単語の重要度スコアとして利用することができる。しかし TF-IDF はレアワードに対して重要度スコアが高くなってしまいう傾向にあるため、類似する文章から代表文を得る場合、レアワードが含まれる文章の重要度スコアを極端に高くしてしまう恐れがある。そこで我々は、クラスタ内での単語の出現率を計算し、出現率の低い (クラスタ内の半分未満) 単語を有する文章には出現数の逆数のペナルティを与えるような重要度スコア計算を考えた。これにより、ノイズに左右されにくい代表文の抽出が期待できる。Algorithm 1 にその計算を示す。

4 実験

実際にチャットボット導入時に使用したログデータを 1000 件サンプリングし、それらに対して提案手法を用いて問い合わせ抽出を行った。なおデータ取り扱いの都合上、例示する際には固有名詞にマスキングを行い、一部の文脈に変更を加えている。

Algorithm 1 ペナルティ付き重要度スコア計算

Require: 各文章ごとに含まれる単語リスト, クラスタサイズ, クラスタ内全単語, TF-IDF

```
for  $i = 0 \dots$  各文章ごとに含まれる単語リスト.length do
  words = 各文章ごとに含まれる単語リスト [i]
  penalty = []
  for  $j = 0 \dots$  words.length do
    w = words[j]
    count = 0
    for  $k = 0 \dots$  クラスタ内全単語.length do
      if  $w ==$  クラスタ内全単語 [k] then
        count += 1
      end if
    end for
    if count <  $\frac{\text{クラスタサイズ}}{2}$  then
      penalty[j] = count
    else
      penalty[j] = 0
    end if
  end for
  重要度スコア =  $\max(TF - IDF) * \frac{1}{\sum(\text{penalty}) + 1}$ 
end for
```

4.1 重要度スコア計算の比較

重要度スコア計算の比較手法として、単純に TF-IDF の最大値を文章の重要度スコアとして取るものに加え、TF-IDF に文章長を考慮した BM25[3] の最大値と EmbedRank[4] を比較する。EmbedRank は文章の分散表現を利用し、教師なしで文書からキーフレーズを抽出する手法であるが、今回はクラスター内の発話群を結合したものを文書、各発話を 1 つのフレーズとして適用する。分散表現は BERT[5] を利用して抽出した。

4.2 結果

上記のデータに対して 3.1 項で述べたクラスタリングを行った結果、282 のクラスタに分かれた。一例として、表 1 にあるクラスタにまとめられた発話群を 2 つ示す。

ID	発話
1-1	注文のキャンセルをお願いします。
1-2	次回からの注文を全てキャンセルです。
1-3	前に注文しています。キャンセルして下さい!
1-4	よくわからないのに、[不満内容] キャンセルをお願いします
1-5	[製品名] を使った後、[不満内容]。合わないようなので全ての商品の注文をキャンセルさせてほしいのです。よろしくお願ひ致します。
1-6	[製品名] を注文後、キャンペーン価格に関する問い合わせをしましたら、回答記述で貴社から今回の注文を一方的にキャンセルされました。[不満内容]。注文キャンセルを取り消して頂けますでしょうか?
2-1	支払い方法変更
2-2	別の支払い方法を希望
2-3	支払い方法変更お願いいたします
2-4	デビットカード支払いはできないのでしょうか 別の支払い方法はないのでしょうか?
2-5	クレジットカードでの支払いでしたが、フィッシング詐欺にあつてクレジットカードを停止しました。支払い方法を変更してコンビニ支払いにしたいので支払い方法をお知らせください。
2-6	お世話になっております。次回の [製品名] のお支払い方法をクレジットカードにしたいのですが、[不満内容] その先がわかりません。お返事お願いします。[氏名]

表 1: クラスタにまとめられた発話群の例

ID 1-*には注文のキャンセル、ID 2-*には支払い方法の変更に関する問い合わせがまとめられている事がわかる。(ID1-6のようにクラスタ内容と異なる発話が含まれているが本稿では扱わない)

これらの文章の重要度スコアを計算したものを表 2 に示す。

ID 1-*では TF-IDF の最大値を取るだけで代表文として適切な発話の重要度スコアが高くなっているが、ID2-*では、ID2-5 のフィッシング詐欺というレアワードの影響で TF-IDF の重要度スコアが高くなってしまっている。EmbedRank だと ID1-6 や ID2-5 の発話のように、多くの情報が含まれた発話の重要度スコアが高くなってしまふ傾向にあつた。BM25 は全体的に悪い結果となつてしまつた。これは基本的にチャツ

ID	TF-IDF	BM25	EmbedRank	提案手法
1-1	0.6798	0.2303	0.9181	0.6798
1-2	0.6771	0.3311	0.9331	0.2257
1-3	0.5276	0.2369	0.9445	0.1758
1-4	0.4602	0.1955	0.9630	0.0657
1-5	0.4398	0.2184	0.9661	0.0314
1-6	0.4211	0.2442	0.9663	0.02340
2-1	0.6049	0.2840	0.6728	0.6049
2-2	0.6123	0.4054	0.8687	0.3061
2-3	0.5135	0.2514	0.9188	0.1711
2-4	0.5145	0.4134	0.9147	0.0857
2-5	0.6244	0.2929	0.9418	0.0780
2-6	0.3126	0.2491	0.9480	0.0156

表 2: 問い合わせ抽出のための重要度スコア計算

トボットによる対話は短文が標準的であり、文章長の考慮はあまり意味のないものであったことが考えられる。今回の例では ID 1-*では ID1-1 の発話、ID 2-*では ID 2-1 または ID2-2 の発話が適切であると考えられ、我々の手法ではこの 2 つの発話の重要度スコアが高くなっていることがわかる。

5 おわりに

本稿では、チャットボットをカスタマーサポートに適用する際に必要な回答作成の負担を減らすための問い合わせ抽出手法を紹介した。今後は、回答が含まれているデータ (例えばメールや有人対応など) から、問い合わせだけでなく、それに対する回答の抽出にも取り組むことでさらなる効率化を目指す。加えて、現在のクラスタリング方法では ID 1-6 のような、異なった内容の発話がクラスタに含まれてしまう場合があるため、問い合わせ箇所を抽出してからクラスタリングを行うなどのアルゴリズムの改善にも努める。

参考文献

- [1] 東中竜一郎, 川前徳章, 貞光九月, 南泰浩, 目黒豊美, 堂坂浩二, 稲垣博人: 対話行為設計のための発話クラスタリング, 言語・音声理解と対話処理研究会, Vol. 63, pp. 37-42 (2011 年)
- [2] Jagadeesh J, Prasad Pingali, Vasudeva Varma: Sentence Extraction Based Single Document

Summarization, *Workshop on Document Summarization, 19th and 20th March, 2005, IIT Allahabad*

- [3] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze.: An Introduction to Information Retrieval, *Cambridge University Press, 2009* p. 233.
- [4] Kamil Bennani-Smires, Claudiu Musat, Andreea Hossmann, Michael Baeriswyl, Martin Jaggin: Simple Unsupervised Keyphrase Extraction using Sentence Embeddings, *Proceedings of the 22nd Conference on Computational Natural Language Learning*, 221-229
- [5] 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.*