

営業日報における「お客様の声」分析と抽出

Customer Voice Extractions from Business Daily Reports

島田 敦子¹

Atsuko Shimada¹

¹株式会社リコー

¹ Ricoh COMPANY, LTD.

Abstract: 日々の営業活動報告はテキストデータで登録され、表現、内容、スタイルが既存の NLP システムの学習データセットに含まれていないため、解析して AI で活用したくてもできていない。本研究では、先行指標が記載されているため、ビジネス上価値があると認識している「お客様の声」文の解析をもとに定義を明確にし、学習データを作成した。最初に、お困りごとと課題の特徴をお客様の「事象」と「心象」を軸に定義した。「お困りごと」とは、現在の状態と理想の状態とにギャップがあり（事象）、そのギャップに対して困っている（心象）。「課題」とは、お困りごとを解消するための行動（事象）であり、行動に前向き（心象）。事象と心象の条件をクリアした文のみ学習データとして分類器を開発、営業日報からお客様の声文の抽出を試みた。テキストデータの AI 活用に手ごたえを感じる結果を得ることができた。

1. はじめに

日々の営業活動報告はテキストデータで登録され、表現、内容、スタイルが既存の NLP システムの学習データセットに含まれていないため、解析して AI で活用したくてもできていない。本研究では、テキストデータとして蓄積されている営業日報から、先行指標が記載されているため、ビジネス上価値があると認識している「お客様の声」の自動抽出を目的としている。

テキストデータは構造化データと異なり、書き手によって様々である。特に忙しいセールス・エージェント（以下 SA）が書く営業日報は短く、多くの当たり前の情報は省略された文としては不完全な表現がほとんどである。本研究では SA 自身の営業活動報告の中で、お客様より伺うことができた「お客様の声」を言語的に解析し、その特徴をもとに、お客様の困りごと文・課題文の伝える情報を明確に定義し、その定義をもとに、商材やソリューションの導入事例集のテキストにラベル付けを行った。

近年発表された汎用言語モデルにラベル付けしたデータを使って「お客様の声」文の文分類モデルを開発した。営業日報データでの結果は公開できないが、開発したモデルで livedoor ニュースコーパスの文を分類した結果を本稿で紹介する。

2. 関連研究

テキストデータからある特徴を有する文を抽出する研究は、NLP 領域で出が行われている。ニュースの記事より「意見文」抽出をする研究 [1]、ブログの記事より「意見文」の抽出・検索をする研究 [2]、Web 文書より「トラブル文」を抽出する研究 [3] などがあるが、多くは機械学習を用いた分類問題として特徴を有する文抽出を行った。本研究では汎用言語モデル BERT を使って単文の多値分類を行った。

「課題」抽出の研究では、特許明細から「従来技術課題」の抽出 [4]、「技術課題情報」の抽出 [5] があり、営業日報からは杉原らの「課題記述文」を抽出する研究 [6] がある。課題記述文の定義は各研究で異なっている。[6] において、従来研究における課題及び課題文の定義に関しては丁寧にまとめられている。本研究と従来研究の違いは、課題及び、課題文をお客様の心象に応じて、「お困りごと」と「課題」の二つに分割している点である。

3. 「お客様の声」の分析

SA が直接取得したお客様の声、特にビジネス上有益な先行指標が含まれている「お困りごと」と「課題」文を、お客様の事象と心象の 2 点を軸に定義する。最初に「事象」と「心象」について、本研究での意味を明記する。事象とは、「ある Event とその結

果」を指す。すべてのビジネス活動が Event であり、過去のあらゆる Event の結果の集合が現在の状態であるとする。心象とは、「心や意識の中に浮かぶ感覚やイメージ」を指す。

お客様の「お困りごと」と「課題」

ビジネスにおけるお客様の課題を定義するには、その前の段階である「問題」を明確にする必要がある。

本研究において、ビジネス上の「問題」とは「ある Event の結果とお客様が想定されていた結果のギャップ」を指す。問題という表現がネガティブなイメージを付与しがちだが、想定外の結果などのポジティブなギャップもここでは含まれる。理想的な結果と現実の結果の間にギャップがあるという事象に対して、お客様が困っている・不満足であるという心象が表現されたとき、その文を「お困りごと」文と判定する。

またお困りごととなっている問題(理想と現実のギャップ)がある状態を、何とかしたい・解消したいというお客様の心象が明確に表現された文をお客様の「課題」文と本研究では定義する。

お客様の理想は先行指標であり、ビジネスチャンスと直結する。日々蓄積される営業日報から上記の定義をもとにお客様のお困りごと文と課題文の自動抽出方法に関して次に述べる。

4. お困りごと文・課題文の自動抽出

本研究では、近年発表された汎用言語モデル BERT [7] を用い、お客様のお困りごと文と課題文を営業日報から言語的特徴を基に判定し、抽出する。

抽出方法

本研究では京都大学が発表している汎用言語モデル BERT の PreTrain モデルに、独自のお客様のお困りごと文・課題文をマニュアルでラベリングしたデータで Fine Tuning を行い、多値分類器(お困りごと文・課題文・その他の文を分類する分類器)を開発した。

データ

手作業でラベル付けしたデータは、弊社の商材・サービスの導入事例集を用いた。また Fine Tuning に使用したラベル付けしたデータの数は以下の通りである。

| 文分類カテゴリ | ラベルデータ数 |
|---------------|---------|
| お客様のお困りごと文 | 384 文 |
| お客様の課題文 | 311 文 |
| その他 (SA 活動報告) | 1,619 文 |

文分類モデルの実験と結果

本来は営業日報から、お困りごと文・課題文を分類・抽出を行っているが、本研究会では実際の営業日報データを分類した結果を公開できないため、テストデータとして livedoor ニュースコーパスの以下のテキストを使用した。

| No. | livedoor ニュースコーパス (家電チャンネル : http://news.livedoor.com/category/vender/kadench/) |
|-----|--|
| 1 | あなたは「正しい歯磨きの方法」を知っているだろうか。 |
| 2 | おそらく多くの人が小学生の時に歯の磨き方を習うのだが、時代によって「正しい方法」が変わっている。 |
| 3 | 歯磨きの仕方が悪いとせっかく磨いても歯垢が取れない。 |
| 4 | 歯垢は虫歯だけでなく、糖尿病の悪化や動脈硬化、心筋梗塞など、全身疾患の原因にもなるという。 |
| 5 | 大人になってからも最新の正しい歯磨き方法を是非習得したいものだ。 |
| 6 | ブラウン (P & G) は 2012 年 2 月 10 日より、電動歯ブラシ「ブラウンオーラル B」からプレミアムモデル『デンタブライド 5000』『デンタブライド 4000』を新発売する。 |

以下に開発した分類モデルによって、上記のデータを分類した結果を提示する。1 及び 2 の文には、明確な書き手の心象はない。よって分類モデルはお困り文でも課題文でもない判定している(それぞれの確信度を参照)。3、4 の文には事象及び心象の言語的特徴を有すため、分類モデルは「お困りごと」と判定している。文 3 の事象は「歯垢が取れない」、心象は「せっかく磨いても歯垢が取れない」ので、困るとなる。文 4 の事象は様々な「全身疾患の原因」になる、そして心象は「虫歯だけでなく」の「だけ」、及び「全身疾患の原因にも」の「にも」からネガティブな感覚を読み取ることができる。文 6 からは正しくない歯磨きの方法(現実)ではなく、「正しい歯磨きの方法を習得する」(理想)という、理想と現実のギャップを解消するための対策を前向きにおこないたいという事象と心象が読み取れるため、分類モデルは「課題文」と判定している。最後の文 6 は「電動歯ブラシを新発売する」という情報のみで書き手の心象は読み取れないので、モデルはお困りごととも課題とも判定していない。

(2012)

- [7] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.:
BERT: Pre-training of Deep Bidirectional Transformers
for Language Understanding, NAACL, Vol. 1, pp.
4171-4186, (2019).

| Text No. | 確信度 | |
|-------------|-------------------|------------------|
| | お困りごと文 | 課題文 |
| 1 | 0.3436632 | 0.60872 |
| 2 | 0.9062867 | 0.003379103 |
| 3 | 0.99576277 | 0.000890648 |
| 4 | 0.99131435 | 0.001663661 |
| 5 | 0.001380988 | 0.9935121 |
| 6 | 0.3436632 | 0.60872 |

5. まとめ

日々テキストデータで蓄積される営業日報から、その言語的な特徴をもとに、本研究ではお客様のお困りごとや課題を報告している文を自動抽出するモデルの開発を行った。今後はお客様の課題を更に分析し、事業所移転や基盤システムの変更などビジネス上大きな案件につながる課題を営業日報から抽出するためのラベルリングデータの作成とモデルの改良を行いたい。

参考文献

- [1] YU, H.: Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences, In Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 129-136, (2003)
- [2] FURUSE, O.: Opinion sentence search engine on open-domain blog, Proceedings of the 20th international joint conference on Artificial intelligence, pp. 2760-2765, (2007)
- [3] 丹治広樹:トラブルを表す文の Web からの抽出, 言語処理学会第 15 回年次大会発表論文集, pp. 140-143, (2009)
- [4] 西山莉紗: 特許公報を対象とした従来技術課題の抽出, 言語処理学会第 16 回年次大会発表論文集, pp. 82-85, (2010)
- [5] 坂井浩之: 特許明細書からの技術課題情報の抽出, 人工知能学会論文誌, pp. 531-540, (2009)
- [6] 杉原大悟, 大熊智子, 佐竹功次, 三浦康秀, 服部圭悟: 営業支援システム内に蓄積されたテキストデータからの課題記述文抽出, 情報通信学会技術研究報告言語理解とコミュニケーション 112(196), pp.7-12,