

# QR 決済事業者に関する満足度評価を目的とした ツイート分類手法の提案

干場敦<sup>1</sup> 正田備也<sup>1</sup>

**概要**：本論文は、QR 決済サービス事業者に関する満足度を、ツイートを用いて評価するための分類器の構築方法について提案する。モデルの構築に際しては、3つの機械学習モデルに加え、BERT による分類器も構築し、分類性能の比較を行った。ツイートコーパスから、満足度に関するツイートを収集するために、オリジナルの辞書を作成し、候補となるツイートを抽出した。本論文では、抽出したツイートに対し、広告や宣伝目的のツイートを取り除くための分類器、満足度に関する内容を含んでいるか否かを分類するための分類器、満足と不満足进行分类するための分類器という3種類の分類器を通すモデルを提案する。分類器毎に各アルゴリズムの性能の比較を行ったところ、いずれもBERT による分類器の性能が、他の機械学習手法を用いたモデルの性能よりも優れていることが明らかになった。

**キーワード**：機械学習、深層学習、自然言語処理、BERT、QR 決済

## Tweet Classification for Evaluating Customer Satisfaction on QR Payment Service Providers

ATSUSHI HOSHIBA<sup>†1</sup> TOMONARI MASADA<sup>†2</sup>

### 1. はじめに

本研究は、Twitter の投稿データを使って、PayPay や LINE Pay といった国内の QR コードを使った決済サービス(以下 QR 決済サービスと言う)の満足度の評価ができる分類器の構築を行うことを目的とする。日本の QR 決済サービス市場は 2018 年 10 月に PayPay が誕生したのを契機に、決済手段の 1 つとしての地位を確立するに至っている。

2020 年 12 月に矢野経済研究所が発表した QR 決済サービスの市場規模の推移と予測は図 1 の通りである。2019 年の消費増税の際に行われたキャッシュレス還元事業や、新型コロナウイルスによる接触を伴わない決済形態への関心の高まりを受け、右肩上がりの成長を遂げており、今後もこのトレンドが続くと思われる。

PayPay のサービス開始以降、大手通信会社等がこのマーケットに参入しており、熾烈なシェアの獲得競争が行われている。2020 年 6 月に QR 決済事業の先駆的存在であった Origami がサービスを終了し、メルペイに統合されたことから、当面は参入と統合配合が繰り返される出入りの激しい業界となることが予想される。

QR 決済はスマートフォンを用いて行うのが一般的であり、SNS との親和性も高いと考え、関連するツイートを分析し満足度を調査することができれば、今後の市場の動向を占う一つの材料となるのではないかと考え、本研究テーマを選ぶに至った。

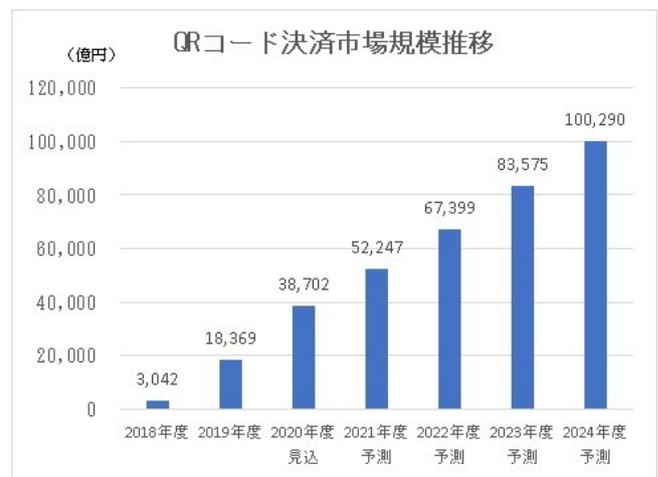


図 1 QR 決済事業の市場予測  
(矢野経済研究所ホームページ[1]をもとに作成)

### 2. 関連研究

日本語のツイートを対象に、ある事象に関連する肯定または否定的な投稿を抽出・分析する研究を紹介する。

王ら[2] は、ラーメン店に関する評判情報を含むツイートを、分類器を用いて抽出し、感情辞書を使って評判を数値化している。しかし、この研究は、以下の 2 点について課題が残るものとなっている。1 点目は、何を以って評価情報とするかについて明らかにされておらず、評価が筆者

<sup>1</sup> 立教大学大学院 人工知能科学研究科  
Graduate School of Artificial Intelligence and Science, Rikkyo University.

の主観によるラベル付のみに依拠していると思われる点である。2点目は、分類器の構築に当たっては、線形カーネルSVM、多項モデルナイーブベイズ、ベルヌーイモデルナイーブベイズを用い、アンサンブル学習を行っているが、各モデルのF値がいずれも0.7以下となっており、分類精度に改善の余地が大きい点である。

感情分析とは目的は異なるが、通信障害を判定するためにツイートを分析する研究も行われている。その中には、障害に関する内容を含みそうなツイートをコーパスから抽出するために辞書を作成し、分類器を用いて内容の仕分けを試みているものなど、感情分析においても有用だと思われるものが存在する。

池田ら[3]は、投稿内容から通信サービスに関する否定的な表現を、否定表現辞書を使って抽出し、抽出された投稿を時系列分析することで、投稿数の遷移から異常事態を検知している。否定表現辞書の構築に際しては、各通信会社名で投稿を収集し、それを既存の汎用辞書を使って分類している。そして、分類された否定的な投稿からAICを使って否定的な投稿に特徴的な表現を獲得することで辞書化している。しかし、この方法は単純なキーワードでのマッチングとなっていることから、ネットワーク障害に関係のない否定的な表現、例えば「A社の苦情窓口に繋がらない」といったものを含むことになっており、これらを仕分けのために人力での精査を必要としている。

池田ら[3]の研究をもとに書かれた山田ら[4]の研究では、ツイートを抽出する際に辞書を用いるのは同様だが、実際に障害が発生した際の投稿内容から肯定/否定的な表現を抽出している。具体的には、障害が発生した特定の日時の投稿を収集し、TF-IDFを用いて重要度の高い表現を辞書として利用をしている。辞書を用いて抽出された投稿から、教師あり学習を利用してサービスに関する肯定/否定的な表現を含んでいないものを除外した上で、PrecisionとRecallを用いて分類の精度の評価を行っている。この研究については、玉ら[2]、池田ら[3]の研究よりも使用される辞書の精度は高いと考えられるが、処理コストの小ささからLIBLINEARのライブラリからしかモデルを選択しておらず、この点において改善の余地があると思われる。

### 3. 提案手法

#### 3.1 提案手法の要点

今回の提案する手法は、以下の点において、関連研究との差別化を図っている。

- ① 極性辞書を使わず、QR決済サービスと類似するサービスに関する口コミサイトの情報を収集し、TF-IDFを用いてオリジナルの辞書を使って感情分析を行った点。
- ② 2018年に発表され、11個の自然言語処理タスクでSOTAを記録した汎用言語表現モデルであるBERT[5]を

ベースにした分類器も構築し、3つの機械学習モデル(ロジスティック回帰、ベルヌーイモデルナイーブベイズ、カーネル化SVM)との精度の比較を行った点。

#### 3.2 提案手法の概要

QR決済サービスに関するツイートの満足度を評価する分類器の構築にあたって、本研究で提案する手法の概要は図2の通りである。各決済事業者名が含まれるツイートのコーパスを用意した上で、オリジナルの辞書を作成し、3つの分類器を通すことで、ツイートをふるいにかけていく構成としている。

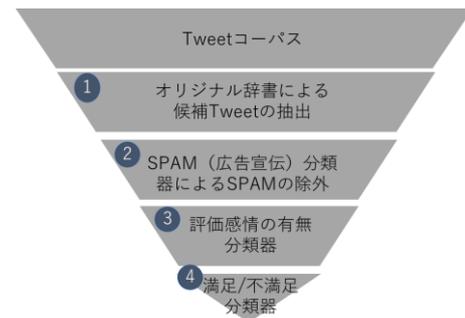


図2 提案手法の概要図

図2に付された番号は、提案システムを構築する手順であり、具体的には下記の流れとなっている。

- ① 極性辞書を使わず、類似サービスの口コミサイトから満足と不満足に関する辞書を作成。
- ② 辞書に基づいてコーパスからツイートを抽出し、そこからさらに宣伝や広告といった不要なツイート(以下spamと呼ぶ)を判別する分類器を構築。
- ③ ②により、ふるいにかけてられたツイートを対象に、満足と不満足に該当するツイートかそれ以外かを判別する分類器を構築。
- ④ ③の結果残った、満足か不満足のいずれかを含むツイートを対象に、満足と不満足を判別する分類器を構築。

#### 3.3 対象データと前処理

分類器の構築に際し使用したデータは、5つの国内の代表的なQR決済サービス名(PayPay, LINE Pay,メルペイ, d払い, au Pay)を含む約50万件のツイートである。機械学習モデルの構築に際しては、ツイートに対して形態素解析エンジンのMeCab[6]を用いて形態素解析ならびに表1の前処理を行った上で学習を実施した。

表 1 前処理の概要

前処理	内容
数字の置き換え	ツイート中の数字を全て 0 置換
特定品詞の削除	助詞, 助動詞, 接続詞, 記号の削除
活用形の原形化	活用がある品詞の原形化

### 3.4 オリジナル辞書の作成

ツイート内容から満足や不満足を表すものかを判定するにあたり、オリジナル辞書を作成し、辞書に該当する言葉を 2 つ以上含むものを満足、不満足を表すツイートの候補とした。汎用的な極性辞書を使わなかったのは、QR 決済サービスにおける満足を考えたときに、「便利」や「使い勝手」という言葉が思い浮かんだが、それを汎用的な極性辞書として知られる単語勘定極性対応表[7]を使って調べてみると、ネガティブな極性を持っていたため、正しく評価できない懸念があると感じたためである。

オリジナルな辞書の作成にあたっては、QR 決済と同じ決済サービスに該当するクレジットカードに関して、Web サイト「価格.com」に各カードの 5 段階の評点と口コミが掲載されていたため、そちらの情報を収集することによって行った。具体的には、2020 年 9 月時点で投稿されていたクレジットカードに関する口コミデータ 4,886 件のうち、評点が 5 点の 1,110 件が満足を表す口コミ、評点が 1 点の 1,123 件が不満足を表す口コミと定義し、それぞれのカテゴリで TF-IDF 値を計算することで、満足、不満足を表す特徴的な単語を求め、辞書にした。作成した辞書の内容は表 2、表 3 の通りである。

表 2 満足を表す辞書

簡単	好き	大きい	かなり	アップ
溜まる	感じ	機能	一般	管理
高め	レビュー	便利	選択	いただける
結構	コンビニ	魅力	だいたい	充実
好み	色々	欲しい	使い勝手	補償
稼げる	実質	有効	おすすめ	盛り
通常	優待	比べる	重宝	ほぼ
安い	一番	安心	たまる	基本
満足	拡充	嬉しい	最大	気に入る
最高	実績	シンプル	それなり	独自
十分	やすい	貯める		

表 3 不満足を表す辞書

最悪	全く	突然	終了	分かる
解約	勝手	繋がる	もらう	もらえる
条件	結局	作成	上限	中止
問い合わせ	停止	説明	最低	不正
注文	もう	こんな	問い合わせる	

学習データとした約 50 万件のツイートのうち、オリジナル辞書に記載の単語を 2 語以上含むツイートを、満足もしくは不満足を表すツイートの候補と定義し、抽出をしたところ、1,519 件のツイートが該当したので、これをもとに分類器の構築を行った。

### 3.5 機械学習による分類器の構築

#### 3.5.1 spam 分類器

3.2 で述べたように、本研究では 3 つの分類器を通すことで満足度の評価を行う構成としており、1 つ目として spam を判別する分類器の構築を行った。spam を除外する理由としては、オリジナルの辞書を使用した 1,519 件のツイートを目視で確認したところ、spam に該当するものが 452 件見つかったため、満足と不満足を分類する上での障害となると考えたためである。

機械学習ライブラリの Scikit Learn を用い、ロジスティック回帰、ベルヌーイモデルナイーブベイズ、カーネル化 SVM の 3 種類の手法で、人力でラベル付けを行った 1,519 件のツイートをインプットとしてモデルを構築した。

説明変数には、形態素解析をしたツイートのテキストを Scikit Learn の CountVectorizer を使ってベクトル化したものと、ツイートを投稿したアカウントのフォロー数とフォロワー数を用いた。

インプットデータを学習用とテスト用に分割し、学習用データで交差検証を行った結果、最も Accuracy が高かったパラメータでテストデータの評価を行った。

#### 3.5.2 評価感情の有無に関する分類器

2 つ目の分類器は、1,519 件の候補ツイート群から spam を取り除いた 1,066 件を対象に、満足や不満足と言った評価に関する感情を表しているものと、そうでないものを分類するためのものである。

本研究で示す提案手法を試す以前に、1,066 件を対象に満足/不満足/それ以外の 3 クラス分類を行う分類器の構築も試みたが、各クラスに属するデータ数が多いとは言えないためか、いずれのモデルの Accuracy, Precision, Recall も 0.8 を下回った。2 クラス分類を 2 回行った方が、エラーの学習や性能の安定性が増すのではと考え、提案手法を実施するに至った。

インプットデータが違うだけで、使ったライブラリや説明変数、学習及び評価の方法は spam 分類器と同様である。

### 3.5.3 満足/不満足のカテゴリ

3 つ目の分類器は満足と不満足を分類するものである。

3.5.2 で用いた 1,066 件から評価感情が含まれていないツイート 722 件を除外した 328 件を対象に、分類器の構築を行った。

こちらにも使ったライブラリや説明変数、学習及び評価の方法は spam 分類器と同様であり、インプットデータのみが異なる。

### 3.6 BERT を用いた分類器の構築

3.5 で記載した機械学習による分類器に加え、BERT を用いて 3 つの分類器の構築も行った。

#### 3.6.1 BERT について

BERT は 2018 年に Google AI Language により発表された自然言語処理モデルで、Transformer[8]の Encoder 部分を使用している。表 4 に示す、2 種類の教師なしの双方向の事前学習を行うことで大量のデータから単語の埋め込み表現を獲得し、そちらで獲得した重みを初期値として解決したいタスクに応じて教師ありのデータでファインチューニングを行うという 2 段階の構造になっているのが特徴である。BERT は 11 個の自然言語処理タスクで、SOTA を記録し、その性能の高さと、転移学習のし易さにより話題となった。

表 4 BERT で用いられた事前学習

事前学習の種類	内容
Masked Language Model	入力文の 15%をランダムにマスキングし、マスクされた文字を予測するタスク
Next Sentence Prediction	2 つの入力文が、連続する文か否かの判別を行うタスク

#### 3.6.2 分類器の構築

日本語に対応した BERT モデルとして東北大学の乾研究室が作成し bert-base-japanese-whole-word-masking[9]を使用した。spam の有無、評価感情の有無、満足と不満足のカテゴリを構築するにあたり、対象としたツイートについては、3.5 と同様であるが、フォロワー数やフォロー数とは説明変数として使用せず、ツイート本文のみをインプットとして用いた。

## 4. 評価実験結果

3 つ分類器の評価性能の比較結果は、図 3~5 の通りである。いずれの分類器においても BERT の分類器の性能が高いことが見て取れる。また、表 5 は、spam を取り除いた後のツイートに対して提案手法で示す 2 クラスの分類器を 2 回

組み合わせることによる分類結果と、提案手法とは別に行った 3 クラス分類を 1 回実施することによる分類結果の比較である。

評価感情の有無に関する分類と満足と不満足のカテゴリに関しては他の機械学習モデルとの性能差は顕著であるが、BERT では単語同士の前後関係を考慮することができるため、それが分類をする上で重要な要素となっていることが推察できる。

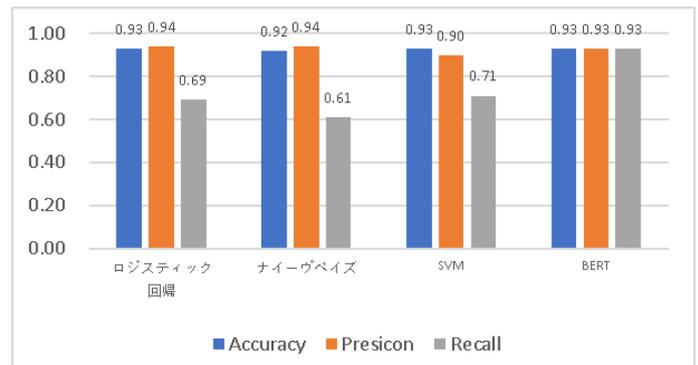


図 3 spam 分類器の分類結果

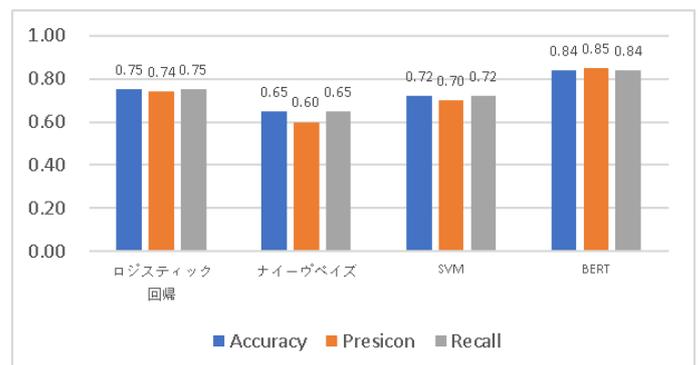


図 4 評価感情の有無の分類器の分類結果

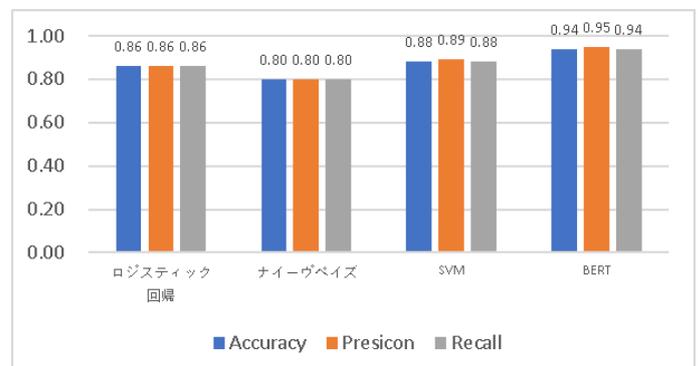


図 5 満足/不満足のカテゴリの分類器の分類結果

表 5 提案手法と 3 クラス分類の比較

	3クラス分類			提案手法 (2クラス分類器×2)					
				評価感情有/無 分類器			満足/不満足 分類器		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
ロジスティック回帰	0.63	0.70	0.63	0.75	0.74	0.75	0.86	0.86	0.86
ナイーブベイズ	0.66	0.68	0.66	0.65	0.60	0.65	0.80	0.80	0.80
SVM	0.64	0.66	0.64	0.72	0.70	0.72	0.88	0.89	0.88
<b>BERT</b>	<b>0.73</b>	<b>0.75</b>	<b>0.73</b>	<b>0.84</b>	<b>0.85</b>	<b>0.84</b>	<b>0.94</b>	<b>0.95</b>	<b>0.94</b>

## 5. おわりに

本研究では、QR 決済事業者の満足度を、ツイートを用いて評価するために 3 つの分類器を通す手法を提案した。3 つの機械学習モデルと BERT を使用したモデルの計 4 つの手法の比較を行ったが、いずれの分類においても BERT を用いたものの性能が一番良いことが明らかになった。また、spam を取り除いた後の満足/不満足/それ以外の分類においては、3 クラス分類を行うのではなく、2 回に分けて 2 クラス分類を行った方が分類の精度が良くなることも明らかになった。

BERT を用いたモデルにおいて、分類に失敗したツイートの一例を表 6 に示す。分類に失敗したツイートを見ると「やばい」や「強い」というように、使用する場面によって意味が変わる言葉が含まれているツイートが存在することが分かった。これらを正確に分類するためには、学習データの増大が必要であると考え、引き続きツイートを収集するなどし、性能の向上を図っていきたい。

表 6 BERT モデルで分類に失敗した例

分類器	分類に失敗したツイート例
評価感情有無 の分類器	<ul style="list-style-type: none"> <li>三井不動産の d 払い 20%還元めっちゃ強い。某アウトレットパーク行ってるけど、半額のものからさらに 20%還元になるから桁が 1 つ小さくなる会計になる</li> <li>誰か楽天とか PayPay モールのキャンペーンでもらえるポイントを計算してくれるスプレッドシート作ってないかな。条件分けできると嬉しい</li> </ul>
満足と不満足 の分類器	<ul style="list-style-type: none"> <li>PayPay やばいよなコンビニヘビーユーザーだから結構返ってくる</li> <li>あと LINE Pay 毎月のクーポン 10 枚あるけど使い勝手悪いの多いから営業頑張って欲しいのと、%じゃないのに戻してほしいのと、あとクーポン出すのめっちゃ時間かかるからたどり着くまで時間かかるから場所考えて欲しい..</li> </ul>

**謝辞** 本論文の執筆にあたり、熱心なご指導をいただいた正田備也教授に感謝の意を表します。

## 参考文献

- [1] 矢野経済研究所：QR コード決済市場に関する調査を実施  
[https://www.yano.co.jp/press-release/show/press\\_id/2598](https://www.yano.co.jp/press-release/show/press_id/2598)
- [2] 王博, 服部隆志, 萩野達也：Twitter におけるロコミ情報の抽出と分析. 情報処理学会第 81 回全国大会論文集 2019(1), pp. 139-140 (2019)
- [3] 池田和史, 服部元, 小野智弘, 麻生英樹：Twitter 解析による通信品質低下傾向の早期検出手法の提案. 情報科学技術フォーラム講演論文集 Vol.11, No.2, pp.101-102 (2012)
- [4] 山田尚, 落合桂一, 横井靖弘, 神山剛, 鳥居大祐：機械学習を用いたツイート解析と統計的異常検知による 通信障害検出システム 情報処理学会論文誌 コンシューマ・デバイス&システム Vol.10, No.2, pp.1-10 (2020)
- [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.
- [6] MeCab : Yet Another Part-of-Speech and Morphological Analyzer.  
<http://taku910.github.io/mecab/>
- [7] 高村大也, 乾孝司, 奥村学：スピンモデルによる単語の感情極性抽出. 情報処理学会論文誌ジャーナル Vol.47, No.02, pp. 627-637 (2006)
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin. Attention is All You Need. In Proceedings of NeurIPS'17.
- [9] 事前学習済みモデル bert-base-japanese-whole-word-masking  
<https://github.com/cl-tohoku/bert-japanese>