

6Q-03

深層学習を用いた VOC に基づいた 顧客の退会危険度推定

川崎雄大^{†1} 櫻井義尚^{†1} 櫻井恵里子^{†2} 鶴田節夫^{†3}
 明治大学^{†1} 西武文理大学^{†2} 東京電機大学^{†3}

1. 研究背景

Amazon や楽天をはじめとする EC(Electronic Commerce) サイトが近年需要を高めており、市場規模と利用者数の増加が進んでいる。

一方で、利用者の増加に伴って EC サイトの負担が大きくなるのがカスタマーサポートにおける顧客対応である。アメリカンエクスプレス社が行った消費者調査によると、「今後5年間で、顧客サービスにおいて最も改善してほしい項目は何ですか?」という問いに対して、日本での回答は「より素早く要望に対応すること」が 39%、「オンライン・モバイルでの問題解決ができること」が 21%で上位2項目を占めた。

そこで、本稿では機械学習によって退会危険者を分類し、カスタマーサポートに寄せられた大量のメールから早期返答が求められるものを発見することを目指した。

2. 先行研究

2.1 カスタマーサポートの満足度研究

本研究が対象とするカスタマーサポートに関連する研究として Park らによる研究[3]が挙げられる。Park らは、カスタマーサポートで用いられる利用者とオペレータのオンラインチャットセッションを対象として、利用者の満足度に深く影響する要素を分析した。オンラインセッションで用いられた単語の感情極性と、セッションのメタデータ(長さ、発話回数)を入力とし、セッションの満足度を予測するランダムフォレストモデルを作成した結果、セッションの内容を表す単語の感情極性がより分類に寄与することがわかった。

この結果をもとに、本稿ではカスタマーサポートに寄せられたメールテキストを分類に用いる情報として選択した。

2.2 深層学習によるテキスト分類

本研究ではテキストデータをもとに分類を行う。そのため的手法として、Kim の研究[4]を先行研究として参考にし

た。この研究では、テキストのトピック分類問題に対して畳み込みニューラルネットワーク(Convolutional Neural Network, CNN)を用いた手法を提案し、従来 CNN よりテキスト分類に適しているとされていたリカレントニューラ

ルネットワークやサポートベクターマシンといった手法よりも高い分類精度を実現できること映画のレビューテキストの分類問題において実証した。

3. 提案手法

本研究で提案するのは、畳み込みニューラルネットワークを用いた、メールテキストに基づく退会危険性の有無を二値分類することである。メールごとに、受信後に送信者(顧客)が退会したか否かを正解ラベルとして付与したものを畳み込みニューラルネットワークの学習に用いる。本章では提案手法のフローについて述べる(図1)

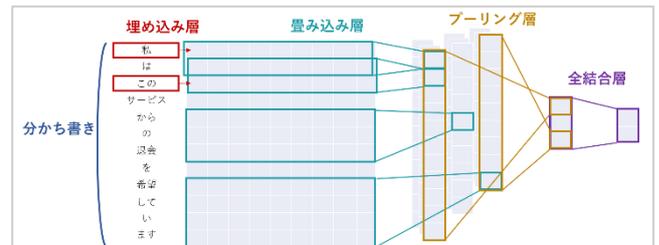


図1 提案手法のフロー

3.1 前処理

テキストには前処理を施す。入力として用いるメール文書には一般に、内容と直接関係ない形式的な内容が存在する。例えば、署名や連絡先などの情報や、過去のやり取りを返信に含む場合の引用文がそれに該当する。これらの情報は、畳み込みニューラルネットワークの入力として単語をベクトル化する際にノイズとなる。そのため本研究では、そのノイズ要素を除去するために、一般に署名を囲っていることが多い「-」や「*」の記号で囲まれた範囲や、引用文で用いられる「>」を冒頭にもつ行を削除するプログラムを作成・適用した。

また、日本語でテキストを畳み込みニューラルネットワークに入力できる形式に変換するために、形態素解析ツール MeCab[5]を用いて分かち書きを行った。

3.2 テキストのベクトル化

分かち書きをした文書の各単語を一次元ベクトルに変換する。文書は単語ベクトルの連続した行列となり、それが畳み込みニューラルネットワークへの入力形式となる。

Predicting Unsubscription from VOC using Deep Learning
^{†1} YUTA KAWASAKI, Meiji University
^{†1} YOSHITAKA SAKURAI, Meiji University

^{†2} ERIKO SAKURAI, Bunri University of Hospitality
^{†3} SETSUO TSURUTA, Tokyo Denki University

単語のベクトル化には Skip-gram モデル[6]を用いる。Skip-gram とは、one-hot 表現された単語を入力として、文書中でその単語の周辺語を推測するモデルである。モデルは入力層、中間層、出力層の三層からなるニューラルネットワークであり、入力層と中間層間の重みがそれぞれの単語を表現するベクトルとして利用できる(図2)。

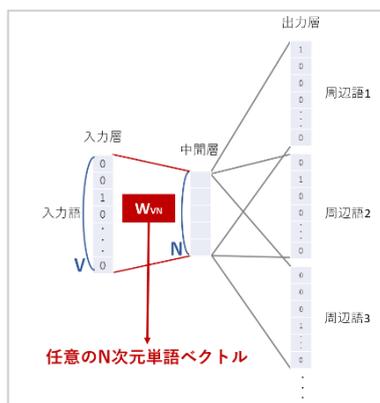


図2 Skip-gram モデル

3.3 畳み込みニューラルネットワーク(CNN)

畳み込みニューラルネットワークは以下の三層からなる。

- 1) 畳み込み層
- 2) プーリング層
- 3) 全結合層

畳み込み層では、複数の単語ベクトルにフィルタをかける演算を行う。フィルタは、任意の単語数×単語のベクトル化次元数の行列で、1単語ずつフィルタの位置をスライドしながら、単語の埋め込み表現行列とフィルタ行列の各セルで乗算を行う。

プーリング層では、畳み込み層処理後の行列に対して、マックスプーリングを行う。その結果、各フィルタの特徴と最も一致した値を並べたベクトルが得られる。

全結合層は、プーリング層の出力結果数のノードと、分類クラス数のノード間に重みを設けて結合したものである。今回は退会危険者/会員の二値分類であるために、2クラスでの分類として実装を行った。

4. 実験

4.1 データ

メールテキストデータは、共同研究企業の EC サイトへ寄せられたメール 25,000 件を用いた。そのうち、メール受信後に退会に至ったメール 554 件に合わせて、退会に至らなかったメールをアンダーサンプリングし 1108 件のメールを用いた。そのうち、1000 件を学習用、108 件をテスト用データセットとした。

4.2 メタデータによる分類との比較

本研究で行うメール文書による退会危険者分類と比較として、会員ステータスを用いた分類を行った。この実験は年齢や性別といった顧客のもつ属性と、メール文書中で交わされるテキストがもつ情報のどちらがより会員の退会意向と結びついているかを検証する目的で行う。会員ステータスは「年齢」「性別」「累計購入金額」「直近の年間累計

購入金額」「購入回数」「定期購入回数」「紹介入会の有無」「サイト主催のサロンイベントへの参加回数」の 8 ステータスを使用する。手法はロジスティック回帰とサポートベクターマシンの 2 つの手法にて実験した。

4.3 結果

メール文書による予測と比較して行った、会員ステータスによる分類精度との比較は以下ようになった。

表 1 会員ステータスによる予測と提案手法の比較

手法	分類精度(Accuracy)
ロジスティック回帰	0.572
SVM	0.728
CNN (提案手法)	0.792

5. 結論

本研究では、EC サイト退会につながる可能性の高いメールを分類するため、カスタマーサポートへ寄せられたメール文書に基づいた、畳み込みニューラルネットワークによる分類モデルを提案した。その結果、会員ステータスによる分類精度を上回る精度で退会危険者を分類することが可能であることが分かった。

6. 展望

サポートベクターマシンによる実験から、会員ステータスによっても比較的容易に高い精度のモデルを作成することが分かったため、メール文書と会員ステータスの組み合わせによる分類モデルを提案することを目指す。加えて、今回学習に用いたデータ数が、非常に少ない退会者数に合わせた結果 1000 件程度となった。そのため、豊富に存在する会員のメール文書を利用するため、畳み込みニューラルネットワークモデルの不均衡データへの適用を考えた拡張を今後の目標とする。

参考文献

- [1] Kunwoo Park, Jaewoo Kim, “Mining the Minds of Customers from Online Chat Logs”, CIKM '15 Proceedings of the 24th ACM International on Conference on Information and Knowledge Management pp.1879-1882
- [2] Yoon Kim, “Convolutional Neural Networks for Sentence Classification”, arXiv. <https://arxiv.org/abs/1408.5882> (2018 年 1 月 28 日参照)
- [3] “Mecab: : Yet Another Part-of-Speech and Morphological Analyzer” , <http://taku910.github.io/mecab/> (2018 年 1 月 3 日参照)
- [4] Tomas Mikolov, Ilya Sutskever, Kai Chen, et al, “Distributed Representations of Words and Phrases and their Compositionality”, arXiv. <https://arxiv.org/pdf/1310.4546.pdf>