

センチメント特徴量を用いたスパム判定機械学習向けのレビュー データベクトル変換モデル

大槻明¹, 川村雅義²

概要: 本研究では, スпам判定機械学習用のレビューデータベクトル変換モデルについて提案する. レビュー文はテキストデータでありカテゴリカルデータであるため, そのままでは機械学習に適用することが難しい. そこで, 本研究ではレビューデータから **bigram** を抽出しセンチメント分析を行うことで, 抽出した **bigram** にセンチメント特徴量を付与することでレビューデータをベクトルデータに変換するモデルについて提案する. 本モデルは, テキストデータをベクトルデータに変換するモデルであるため, スпам判定以外にも様々な機械学習に応用することが可能である.

キーワード: テキストマイニング, センチメント分析, スпам判定, 機械学習

1. はじめに

スパム判定モデルに関する先行研究は, 表 2 に示すように, 「レビューの類似性」, 「レーティング行動の類似性」, 「レビューの信憑性」といった基準を元にスパム判定を行う研究が多い. Jindal[1]や Mukherjee[2]のモデルでは, レビューの類似性からスパム判定を行っており, Jaccard 類似度係数が 90%以上のときに複製レビューとみなしてスパム判定している. また, Lim[3]のモデルでは, レーティング行動の類似性からスパム判定を行っており, サクラチェックカー[4]はでレビューの信憑性やレビュー行動の信憑性を元にスパム判定を行っている.

最近では, Ott[5]や岡山[6]の研究のように, SVM を用いてスパムレビューを機械学習することによりスパム判定を行うモデルも提案されている. レビューテキストはカテゴリカルデータであるため, そのままでは機械学習に適用することが難しい. そこで Ott のモデルでは, LIWC (Linguistic Inquiry and. Word Count)[7]を用いてレビューテキストデータをベクトルデータに変換したうえで機械学習に適用している. しかし, センチメント特徴量を用いてレビューデータをベクトルデータに変換するモデルは未だ提案されていない.

ゆえに, 本研究は, レビューテキストから **bigram** を抽出し, 抽出した **bigram** にセンチメント特徴量を付与することでレビューデータをベクトルデータに変換するモデルについて提案する.

2. 提案モデル

2.1 モデルの全体像

本提案モデルの全体像イメージを図 1 に示す. 図 1 の text mining model (左側) の箇所がレビューテキストから

bigram を抽出し, 抽出した **bigram** 単位でセンチメント特徴量を付与しているイメージである. そして, Machine Learning Model (右側) の箇所が, text mining model で出力するセンチメント特徴量が付与された **bigram** を単位として ($1\text{Review} = \text{bigram}_1, \text{bigram}_2, \sim, \text{bigram}_n$), 機械学習及び予測を行う 3 層ニューラルネットワークモデルのイメージである.

2.2 Text mining model におけるセンチメント分析

センチメント分析には, 東京工業大学奥村・高村研究室が公開している単語感情極性対応表[8]を用いた. text mining model の出力, つまり, センチメント特徴量が付与された **bigram** の出力イメージは図 1 及び表 1 に示すとおりであるが, このセンチメント特徴量が付与されるまでの流れは下記①～②に示すとおりである.

- ① まず, 抽出された **bigram** の単語ペアに対して合成極性値を与える. 合成極性値とは, 例えば **bigram** (問題-無い) が与えられた時に, 「問題」と「無い」にそれぞれセンチメント特徴量を付与してそれらを合計 (問題-0.716755+無い-0.54371=-1.260465) したものである.
- ② 次に, 合成極性値を用いて個々のレビューの特徴ベクトルを作成する. 合成極性値の単語ペアに一致する単語ペアがあればそのレビュー中での「合成極性値×頻度」とし, なければ 0 とする. これは, 合成極性値の単語ペアがそのレビューの中で 2 回以上登場する可能性があるための措置である. そして, この「合成極性値×頻度」のスコアが, 表 1 の列に設定されている **bigram_n** のセンチメント特徴量スコアとなる.

そして, 表 1 が最終的な text mining model の出力イメージであるが, 行項目が 1 行 1 レビューを表し, 列項目が 1

1 日本大学

2 MK 未来ソフトウェア

レビューに含まれる bigram_n のセンチメント特徴量スコアを表す。そして一番右の spam 列が、スパムレビューであれば 1 を、スパムレビューでなければ 0 をそれぞれ表す。この表 1 のデータをインプットデータとして Machine Learning Model で機械学習を行う。なお、スパム判定の基準は、次節で述べるようにサクラチェッカーを参考に設定した。

表 1. text mining model の出力イメージ

	bigram_1	bigram_2	...	bigram_n	spam
Review ₁	-0.3472	-4.4725		-0.6945	1
Review ₂	1.9999	1.2604		2.9999	1
...					
Review _n	-0.4367	0.9999		0.6473	0

2.3 スпам判定基準

本研究では、サクラチェッカーにおけるさくら度が 90% 以上で、かつ評価が 1 または 5 のレビューをスパムレビューとし、さくら度が 10% 以下で評価が 1 または 5 のレビューを非スパムレビューとして設定した。つまり、サクラチェッカーにおけるスパム確定が 90% 以上の商品について、さらに極端なレーティングがされているレビューにはサクラが多いだろうと考えるアルゴリズムである。2020/6 時点でサクラチェッカーサイトを調べたところ、双方約 2 万件ずつ合計 4 万件程度存在していた。

なお、本研究の主眼はスパム判定の基準自体を提案することではなく、スパムレビューの特徴を機械学習するためにテキストデータをベクトルデータに変換するためのモデルを提案することであるため、スパム判定の基準自体は既存のものを応用することとしている。

2.4 Machine Learning Model における機械学習及び予測

図 1 の右側の Machine Learning Model の箇所が、2.2 節の text mining model で出力したセンチメント特徴量が付与された bigram を単位として (1Review = $\text{bigram}_1, \text{bigram}_2, \dots, \text{bigram}_n$)、機械学習及び予測を行う 3 層ニューラルネットワーク (NN) モデルのイメージである。この 3 層 NN のニューロン数の構成を表 2 に示す。

表 2. 3 層 NN のニューロン数の構成

入力層->	中間層 1 ->	中間層 2 ->	出力層
5000	128	256	2

なお、この 3 層 NN のモデルはあくまで一例であり、これを 4 層以上に多層化したり、誤差逆伝播法を適用するなど、様々な応用例が考えられる。また、NN 以外にも、表 1 のデータはロジスティック回帰分析にも応用できるため、

今後の課題として、評価実験を通じて、text mining model によって出力されたベクトルデータに最適な機械学習及び予測手法について検討していきたいと考える。

3. おわりに

本研究では、スパム判定機械学習用のレビューデータベクトル変換モデルについて提案した。レビュー文はテキストデータでありカテゴリカルデータであるため、そのままでは機械学習に適用することが難しいが、本研究によってカテゴリカルデータをベクトルデータに変換する一つの手法が確立した。今後は、text mining model によって出力されたベクトルデータを、機械学習及び予測する手法について研究していきたい。さらに、現在は Transformer[9] や BERT[10] を使ったテキスト分類の事例が提案されているため、これら先行・関連手法との比較検証を通じて本研究の有効性について確認していきたい。

表 2. スパム判定モデルの先行研究比較

	Jindal[1]	Mukherjee[2]	Lim[3]	サクラチェッカー[4]	Ott[5]	岡山[6]
1 スパム判定の基準	レビューの類似性	レビューの類似性 レターティング行動の類似性	レターティング行動の類似性	レビューの信憑性 レビュー行動の信憑性	レビューの類似性	レビューの類似性
2 対象データ	<ul style="list-style-type: none"> Amazon レビュー 580 万レビュー 214 万レビュー 670 万商品の情報 	<ul style="list-style-type: none"> Amazon レビュー 109,518 レビュー 53,469 レビュー 39,392 商品の情報 	<ul style="list-style-type: none"> Amazon レビュー 50 人のレビュー 	Amazon レビュー	<ul style="list-style-type: none"> TripAdvisor (旅行レビューサイト) 5 つ星の評価を受けているホテル 20 件のレビュー 	<ul style="list-style-type: none"> 口コミサイト Yelp 144,072 件の Venue 4,153,150 件のレビュー
3 スパム判定モデル	二文書の Shingle の集合 (共起語) 間の Jaccard 類似度係数が 90% 以上 のときに複製とみなし、スパムとして次の 3 つのタイプに分類 <ul style="list-style-type: none"> タイプ 1 偽りの意見 タイプ 2 (対象製品ではなく) ブランドへの意見 タイプ 3 (広告や、質問とその回答のような) レビューでないもの 	個人向けのパラメータとして、個人が書き込んだレビュー間の類似度と製品に与える星の数の傾向などを元に 2 項関係モデルによりスパムグループかどうかを判定	レターティング行動パターンの類似性からスパムを判定	下記の項目を機械学習し、サクラかどうかを判定 <ul style="list-style-type: none"> 価格 製品 ショッピング情報・地域 ショッピングレビュー レビュー分布、日付 レビュー&レビュー 怪しい日本語のサクラレビュー ライバル社からの悪評 (従業員レビュー)	ホテルに対するスパムレビューを負例、スパムではないレビューを正例とし、品詞 (POS), LIWC (The Linguistic Inquiry and Word Count), N グラムを対象として、ナイーブベイズ (NB) とサポートベクターマシン (SVM) を用いて、スパムかどうかを判定する分類器を学習	Fake News Challenge のデータと政治に関するニュースを SVM により学習することで、フェイクニュース分類器を作成

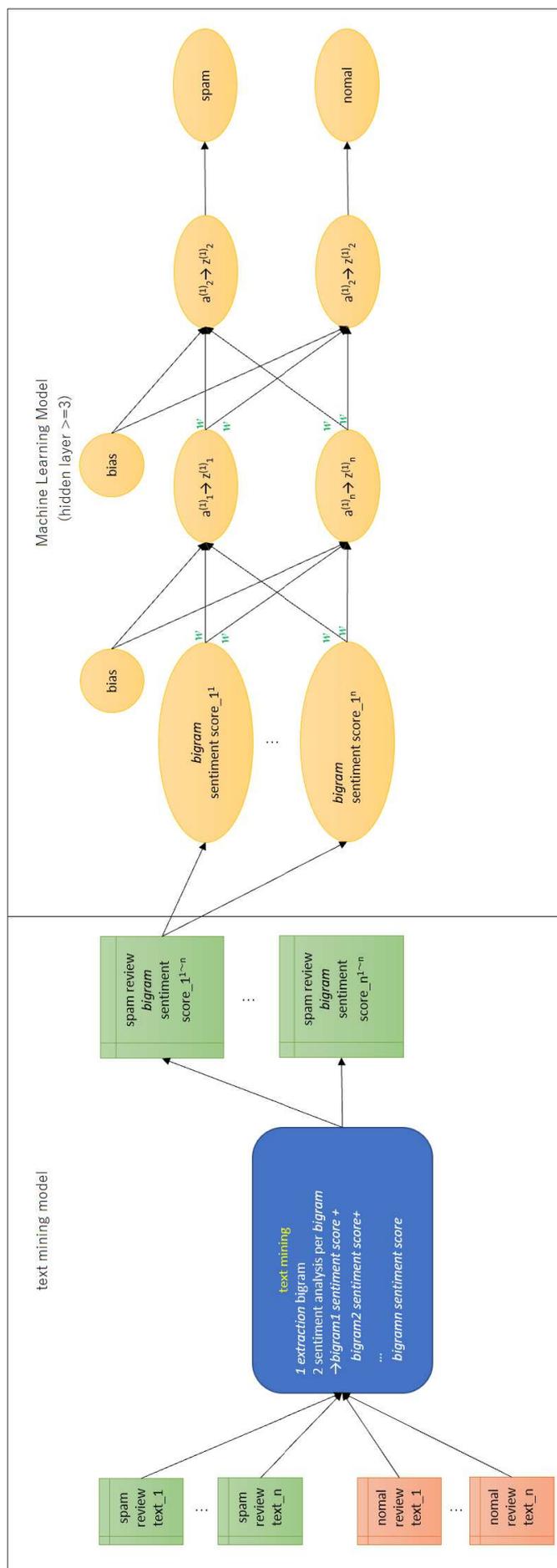


図 1. 本提案モデルの全体像イメージ

参考文献

- [1]Nitin Jindal and Bing Liu :Opinion Spam and Analysis, In Proceedings of the 2008 International Conference on Web Search and Data MiningFebruary 2008 Pages 219–230, 2008.
- [2] Arjun Mukherjee, Bing Liu, and Natalie Glance :Spotting fake reviewer groups in consumer reviews, Proceedings of the 21st international conference on World Wide Web, pp.91–200, 2012.
- [3]LIM Ee Peng, NGUYEN Viet-An, JINDAL Nitin, LIU Bing and LAUW Hady Wirawan : Detecting Product Review Spammers using Rating Behaviors, Proceedings of the 19th ACM International Conference on Information and Knowledge Management: October 26-30, 2010.
- [4]サクラチェッカー :<https://sakura-checker.jp/article/faq/>.(2020/6/21 閲覧済み)
- [5]Myle Ott, Yejin Choi, Claire Cardie, Jeffrey T. Hancock :Finding Deceptive Opinion Spam by Any Stretch of the Imagination, Proceedings of ACL 2011: HLT, pp.309-319, 2011.
- [6]岡山光平, 石川博, 廣田雅春 :フェイクニュース分類器を用いた口コミサイトのレビューの分析, DEIM Forum 2018, pp.3-5.
- [7]LIWC : <https://liwc.wpengine.com/>.(2020/6/21 閲覧済み)
- [8]単語感情極性対応表 :http://www.lr.pi.titech.ac.jp/~takamura/pubs/pn_ja.dic. 2020/6/21 閲覧済み.
- [9]Self-Attention を利用したテキスト分類 :https://www.inoue-kobo.com/ai_ml/text-classification-by-selfattention/index.html . 2020/6/30 閲覧済み.
- [10]日本語 pre-trained models を fine tuning して分類問題を解く : <https://qiita.com/kenta1984/items/7f3a5d859a15b20657f3>. 2020/6/30 閲覧済み.