

Twitter からの
消費者ニーズ抽出手法に関する研究

筑波大学
図書館情報メディア研究科

2018年3月
川島崇秀

目次

第 1 章	序論	1
1.1	背景	1
1.2	本研究の目的	2
1.3	本論文の構成	2
第 2 章	関連研究	3
2.1	口コミ分析・要望抽出に関する研究	3
2.2	Distant Supervision を用いた情報抽出に関する研究	4
2.3	単語の分散表現学習に関する研究	4
2.4	深層学習を用いたテキスト分類に関する研究	6
2.5	本研究の位置づけ	7
第 3 章	要望の定義	8
第 4 章	要望の抽出手法	11
4.1	提案手法の概要	11
4.2	要望表現辞書の作成	12
4.3	教師データの収集	13
4.3.1	手がかり表現を含むツイートの抽出	13
4.3.2	手がかり表現が文末から 3-gram 以内に出現するツイートの抽出	13
4.4	CNN による分類器の構築	14
第 5 章	評価実験	16
5.1	評価対象	16
5.2	評価方法	18
5.3	実験設定	18
5.3.1	ハイパーパラメータの設定	18

5.3.2	実装	18
5.4	実験結果・考察	19
5.5	追加実験	20
5.5.1	「要望の対象が異なる」ツイートを要望に含めた場合との精度変化を 比較	20
5.5.2	回答一致率毎の精度変化を比較	21
第 6 章	結論	23
	謝辞	24
	参考文献	25

目次

2.1	CBoW モデル [1]	5
2.2	Skip-gram モデル [1]	5
2.3	畳み込みニューラルネットワーク [2]	6
4.1	提案手法の概要	12
4.2	CNN モデルの構造 [3]	15
5.1	データセット (i) を対象とした回答一致率毎の精度変化	21
5.2	データセット (ii) を対象とした回答一致率毎の精度変化	22

表目次

5.1	教師データの収集に使用するゲームタイトル一覧	17
5.2	実験結果	19
5.3	要望の対象が異なるツイートを要望に含めた場合の分類結果	21

第 1 章

序論

1.1 背景

近年、ソーシャルメディアの普及により、誰でも簡単に情報発信ができるようになった。ソーシャルメディア上では、製品やサービスに関する口コミが日々大量に投稿されている。こういった背景から、企業においてソーシャルメディア上の口コミ情報を市場調査や反響測定といったマーケティング活動に活かそうという試みが注目されている [4]。特にソーシャルメディア上の口コミ情報を分析し、トレンド予測、ブランド・商品に対する評判の理解や改善に活かす試みは、ソーシャルリスニングと呼ばれており、従来のマーケティング調査では難しかった、消費者が日常の中で発する自然な声を聞くことができるといったメリットがある。

ソーシャルメディアの中でも近年顕著な普及を見せる Twitter は、リアルタイム性の高さ・ユーザの多様性・投稿量の多さから分析対象として大きな注目を集めている [5]。しかし、Twitter 上では、日々何千万ものクチコミ情報が秒単位で行われており、これらの投稿を人手で分析することには膨大なコストが掛かる。したがって、Twitter 上の投稿をビジネス活動に活用していくためには、自動で口コミ情報の抽出を行い、要約を行うなどの分析を自動化するツールの開発が必要不可欠である。

Twitter 上の投稿を自動で抽出する研究としては、センチメント分析が知られている [6]。センチメント分析とは、ユーザが書くレビューに対して、肯定的なものと否定的なものの 2 つのカテゴリに分類する手法である。しかし、この手法では、投稿を感情という視点で肯定的か否定的かの 2 値に分類する為、感情を含まないがビジネス活動に於いて価値のある情報を抽出することが困難である。例えば、要望を含む投稿などは消費者のニーズを直接的に表す重要な情報であると考えられるが、従来の手法で分類することは難しい。

そこで本研究では、「要望」という点に着目して、消費者の要望を含むツイートの抽出を試みる。従来の手法で抽出の対象としていなかった要望を含むツイートの抽出を行い、それらのデータをビジネス活動に活かしていくことで、サービスの品質改善や新規事業の創造といった

活動の支援につなげていくことが期待できる。

しかしながら，Twitter 上の投稿から商品やサービスに関する要望を含むツイートの抽出をする際に課題となるのは，Twitter 上の投稿における多様な文章表現である．Twitter 上には文法的に崩れた表現や多様な文章表現が非常に多く存在している．それ故，従来の提案手法を適用した場合，これらの多様な表現への対応が困難であった．

1.2 本研究の目的

本研究の目的は，企業の消費者ニーズの理解を支援するために，従来手法と比較して，より高い精度での要望抽出を実現することにある．最終的な展望としては，消費者ニーズの理解を支援することによって，サービスの品質改善や新規事業の創造といった，ビジネス活動の促進に繋げていくことが期待される．そこで本研究では，要望を含むツイートの抽出に多層ニューラルネットワークの一つである畳み込みニューラルネットワークを適用することで，従来手法と比較してより高い精度での抽出を試みる．また，学習データの収集に半教師あり学習の一つである「Distant Supervision」の考えを適用することで，より低コストな要望表現の抽出方法を提案する．

1.3 本論文の構成

本論文の構成は次の通りである．第2章で関連研究について概観し，本研究の位置づけを明確にする．第3章で本研究で対象とする要望の定義を行い，4章で提案手法を詳述する．第5章で評価した結果を述べ，考察する．最後に6章でまとめと今後の課題を示す．

第 2 章

関連研究

2.1 口コミ分析・要望抽出に関する研究

口コミ分析に関する研究としては、weblog を対象としたものが主流であった。鈴木ら [7] は、半教師あり機械学習の手法を用いて、Weblog から評価情報の抽出を行った。奥村ら [8] は、weblog を定期的に監視し、評価情報を自動抽出するシステムを開発した。

Twitter 上の口コミ情報の抽出に関する研究も盛んに行われている。Twitter 上の投稿に対してセンチメント分析を行った研究としては野畑ら [6] の研究が挙げられる。野畑らは教師あり機械学習の手法を用いて Twitter 上の投稿をポジティブなものとネガティブなものとの 2 つのカテゴリに分類した。

要望文の抽出に関する研究ではアンケートの自由記述欄から意見や要望を抽出する試みが行われている。山本ら [9] は、アンケートの自由回答欄から要望を抽出する手法として、自由回答の記述の何文目に要望が書かれる傾向にあるのかを分析し、要望文を自動抽出する手法を提案した。大塚ら [10] は自由回答アンケートにおいて間接的な要求を抽出するための基準として「～てほしい」に言い換え可能か否かという基準を提案し、機械学習の手法を用いて要求を含むテキストを抽出している。

また、Twitter 上から要望を含む投稿を抽出する研究としては、栗原ら [11] の研究が挙げられる。栗原らは Twitter 上から地方自治体に関する要望を含む投稿の抽出を試みている。手法としては、あらかじめ作成した要望表現の特徴を含む辞書を用いたパターンマッチングを用いている。山本 [12] らは、Twitter を用いて生活に関連する単語からなる辞書を作成し、特定の地域の要望を含む、生活情報を抽出する手法を提案した。これに対して筆者ら [13] は、ソーシャルゲームに関するツイートを対象とした要望抽出に、機械学習のアルゴリズムである SVM を用いた手法を提案している。

要求表現の言語学的な定義に関する試みも行われている。大森 [14] は要求表現の定義として、「命令」、「依頼」、「禁止」、「誘いかけ」、「希望」、「当為非断定」、「希望非断定」の態度

を帯びる文は要求文であるとした。大森はさらに、要求表現の文法的な特徴として、「～しろ」「～たい」「～ほしい」といった文末表現を挙げている。

2.2 Distant Supervision を用いた情報抽出に関する研究

Web からの情報抽出のタスクに Distant Supervision を用いた研究も行われている。Distant Supervision[15] とは、半教師あり学習の手法の一つであり、知識ベースから取得した少数の手がかり表現を用いることで、半自動的な教師データの収集を可能にする学習方法である。M. Mintz[15] らは Web テキストからの関係性抽出のタスクに Distant Supervision の考え方を適用し、少量の知識ベースから大量の学習データを収集している。この際、知識ベースとして FreeBase を用いている。三浦ら [16] は、Twitter の投稿に対するセンチメント分析のタスクに Distant Supervision の考え方を適用することで、教師あり機械学習の低コスト化に成功している。この際、学習データ収集に用いる手がかり表現として顔文字を用いている。山本ら [17] は、Web 上のニュース記事からの企業間関係抽出のタスクに Distant Supervision を適用している。ここでは、あらかじめ作成した教師データから、その判断の決め手となった語を抽出し、手がかり表現としている。

これらの研究では Web 上からの情報抽出に関するタスクにおいて、分類性能を低下させることなく、教師あり機械学習の低コスト化に成功している。これらの結果から、適切な手がかり表現の定義が可能であれば、情報抽出のタスクにおいて「Distant Supervision」の考えを適用することで、低コストで教師データを取得できると考えられる。

2.3 単語の分散表現学習に関する研究

分散表現とは、単語を固定長のベクトルで表現する情報表現の 1 つである。分散表現を学習するタスクは表現学習と呼ばれ、言語の意味的な類似度を捉える手法が近年注目を集めている。表現学習の代表的な手法としては、Continuous Bag-of-Words(CBOW) モデル (図 2.1) [18] や Skip-gram モデル (図 2.2)[19] が挙げられる。これらのモデルでは、テキスト内の単語を周辺単語から予測する単語予測のタスクを行い、大量のテキストから ニューラルネットワークで学習することで、中間層における各単語の重みを抽出し、単語に対する概念ベクトルの獲得に成功している。これらの手法によって獲得したベクトル表現を、テキスト分類タスクにおける素性として用いることで、単語の意味的な類似度を考慮した予測モデルの構築が期待できる。

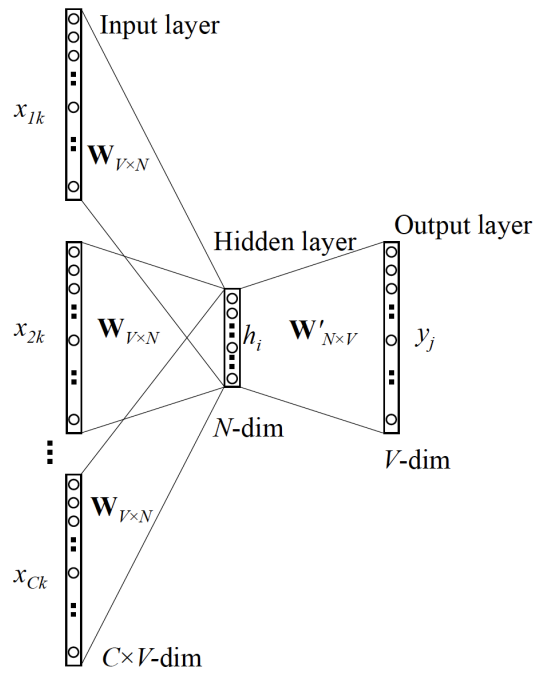


図 2.1 CBoW モデル [1]

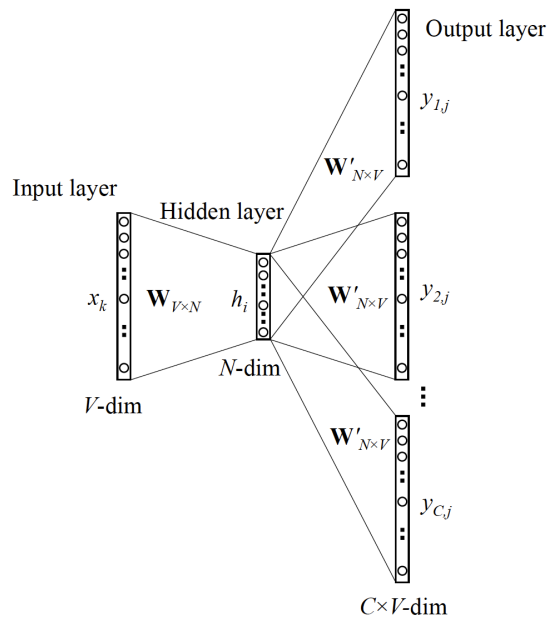


図 2.2 Skip-gram モデル [1]

2.4 深層学習を用いたテキスト分類に関する研究

深層学習とは、多層ニューラルネットワークを用いた機械学習の手法の一つである。深層学習は、画像認識の分野において高い識別性能を示したことから注目され、現在では、音声認識・自然言語処理など様々な分野に於いて研究されている。テキスト分類のタスクでは、特に多層ニューラルネットワークの一つである畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) を用いた手法が精力的に研究されている。

CNN(図 2.3) とは、人間の視覚野を模倣して作られたニューロン間の結合を局所化するネットワークである。畳み込み層とプーリング層、全結合層から構成される順伝播型ニューラルネットワークで、誤差逆伝播法により学習する。CNN をテキスト分類に応用した研究としては、Kim[3] らの研究が挙げられる。Kim らは、Word2Vec により事前学習した単語の分散表現を入力とし、畳み込み層 1 層と全結合層 1 層とによって構成されたシンプルな CNN を用いて、評判分析含む複数のデータセットを評価しており、良好な結果を得られている。Zhang[20] らは、文字レベルの畳み込みニューラルネットワークを用いてテキスト分類を行っており、100 万件程度の文書データを用いた場合に於ける有効性を確認している。

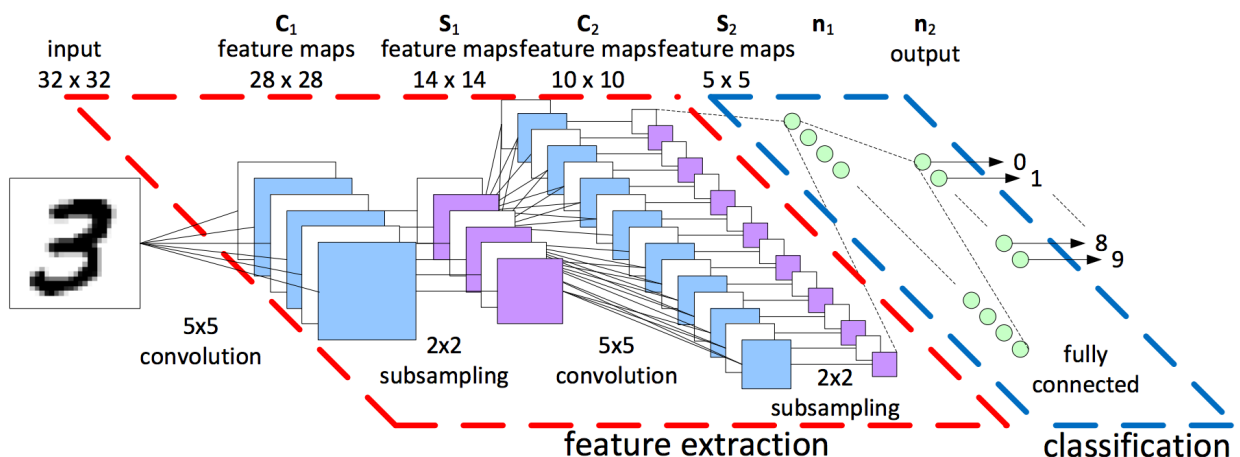


図 2.3 畳み込みニューラルネットワーク [2]

2.5 本研究の位置づけ

本研究では、多層ニューラルネットワークの一つである畳み込みニューラルネットワークを用いることによって、Twitter 上から要望文を対象とした自動抽出を行う。また、半教師あり学習の手法の一つである「Distant Supervision」を用いることによって、少数の手がかり表現から半自動的に学習データを収集する。そのため、要望という視点から Twitter 上のクチコミ情報を抽出するという点で、感情という視点から投稿の抽出を試みた野畑らや三浦らの研究と異なる。また要望の抽出手法においても、深層学習の手法を用いるという点で、従来手法とは異なる。

本研究は、Twitter 上の要望文の抽出というタスクに対して、深層学習の手法を適用し、学習データの収集に半教師あり学習の手法の一つである「Distant Supervision」を用いることで、低コストかつ高精度な要望抽出を試みた点に新規性があると考えられる。

第3章

要望の定義

関連研究で述べた大森と栗原らの論文を参考に、本研究で扱う要望を定義する。大森は要求とは、文に描かれている事態であるとし、次の4つの条件を満たす表現であるとした。

- 条件1 当該事態は、それを捉える当事者にとって望ましい事態である
- 条件2 当該事態は、当該当事者にとって未実現である
- 条件3 当該事態の実現主体として、個人、集団、組織など、意志を持つ主体が存在する
- 条件4 当該当事者は、当該事態の実現を、当該事態の実現主体に求めている

大森はさらに、直接的な要求を表す文章は、「命令」、「依頼」、「禁止」、「誘いかけ」、「希望」、「当為」、「当为非断定」、「希望非断定」のいずれかの態度を帯びるとし、それぞれの文法的な特徴を明らかにしている。

また、栗原らは大塚らの研究を参考に、「～てほしい」「～てください」「～てくれ」といった、日本語母語話者のほとんどが「要求」と判断できる表現を「直接要求」表現とし、「～べき」「～がベストだと思う」「が必要」といった、「～てほしい」に言い換え可能な表現を「要求意図」表現とした。栗原らはさらに、Twitter上の投稿は自由回答アンケートと異なりユーザの独り言や愚痴が投稿される傾向があることに注目し、直接要求や要求意図に当てはまらない場合でも、その内容が要望の動機になる否定的なテキストを「不満」と定義し、「直接要求」「要求意図」「不満」の3つに該当するテキストを要望と定義した。

本研究では以上の先行研究を踏まえ「命令」、「依頼」、「禁止」、「誘いかけ」、「希望」、「当為」、「当为非断定」、「希望非断定」の態度を帯びる表現と、これらに該当しないが要望の動機となる否定的な表現を「不満」とし、まとめて要望と定義する。

命令

相手が意志的に制御できる動作を、相手に強制する表現

例 1) つまらん心配はしないで早く行け

例 2) はやくバグ修正しろ

依頼

相手の意志を尊重して、相手にある動作をするよう頼む表現

例 3) あなたはやく帰ってきてちょうだい

例 4) ちょっと、その婆さんに会ってみてくれないか?

禁止

相手にある動作をしないこと、あるいは、ある事態が生じないように努力することを命令する表現

例 5) そういうことに、やたら興味を持つな

例 6) いちいちアップデートすんな

誘いかけ

聞き手に、話し手と同様の行動をとるように要求する表現

例 7) やりましょう、松田さん熊谷さん

例 8) 一緒にゲームしましょう!

希望

話し手自身に関わる事態の実現を希望する、あるいは他者がある事態を実現することを希望する表現

例 9) 千葉へ行ってもらいたい

例 10) 早く返金して欲しい

希望非断定

希望の態度を断定することを控える表現

例 11) 音楽というコンテンツを手に入れたら、通勤の時に電車で iPod やその他携帯音楽プレイヤーで聴きたいかもしれない

例 12) Windows も Mac も辞書データをひっくるめて月額制でお安くしておきますよという、プレミアムコースを作ってもらいたいかもしれない

当為

ある事態が望ましいとか、必要だ、というように事態の当否を述べる当為の態度のうち、「～べきだ」、「～なければならない」のような述語の基本形をとって表される表現

例 13) 日本は早急に貿易黒字を減らすべきだ

例 14) 君は、あの時彼と別れるべきだった

当為非断定

当為の態度を断定することを控える表現

例 15) 日本は早急に貿易黒字を減らすべきだろう

例 16) 君は積極的になったほうがいいかもしれない。

不満

「命令」、「依頼」、「禁止」、「誘いかけ」、「希望」、「当為」、「当為非断定」、「希望非断定」に該当しないが、要望の動機となる否定的な表現

例 17) 横浜市営地下鉄の始発遅い、最悪

例 18) 市役所の対応悪いわ

第 4 章

要望の抽出手法

本性ではまず、本提案手法の概要を説明する。次に、4.2 節では、教師データの収集に用いる要望表現辞書の作成方法について述べる。4.3 節では、4.2 節で作成した要望表現辞書を用いて、教師データを収集する手順について説明する。4.4 節では、CNN を用いた分類器の構築手法について詳述する。

4.1 提案手法の概要

本研究における提案手法の枠組みを図 4.1 に示す。まず、①のステップにおいて、要望抽出の対象となる商品名/サービス名を含むツイートの収集を行う。次に収集したツイートに対して後述する要望表現辞書と n-gram 判定の 2 段階の処理によって教師データの抽出を行う。その後、および②のステップにおいて教師データから素性の構築と学習を行い、分類器を生成する。ステップ③、において構築した分類器を用いて要望を含む投稿の抽出を行う。

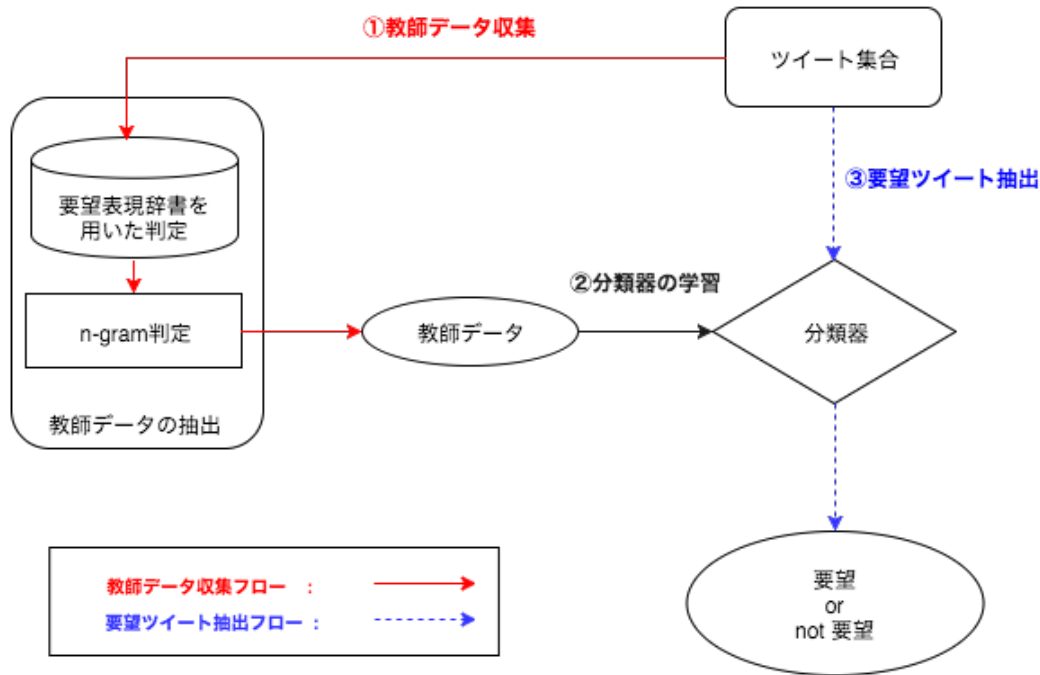


図 4.1 提案手法の概要

4.2 要望表現辞書の作成

Distant Supervision の手法を用いた教師データの収集では、予め、教師データの判別の手がかりとなる表現を決定しておき、それらの表現的な特徴を含むデータを収集することによって、半自動的な教師データの収集を可能にする。Distant Supervision の手法を用いて教師データを収集するには、まず、教師データの特徴を決定する必要がある。以降、本論文では、教師データの特徴となる表現のことを手がかり表現と呼称する。

先行研究において要望を含む文には「～しろ」「～たい」「～ほしい」といった特徴的な文末表現が出現することが知られている。本研究では、大森らの論文中に記述されている要望文の特徴表現リストを参考に、合計 19 個の特徴的な表現を定義し、手がかり表現とする。これらの手がかり表現をデータベースに格納することで、要望表現辞書とした。

本研究で使用する手がかり表現一覧

(命令) ～しろ, ～して

(依頼) ～ください, ～下さい, 動詞否定形 + でくれ, ～てくれ, ～て頂戴 (ちょうだい)

(禁止) ～するな, ～やるな

(誘いかけ) ～しよう

(希望/希望非断定) 動詞連用形 + たい, 動詞テ形 + 欲 (ほ) しい, ～たい, ほしい

(当為/当為非断定) ～べき

(不満) ～過ぎ, ～にくい, ～づらい, ～ない

4.3 教師データの収集

4.2 節で作成した要望表現辞書を用いて, 教師データを収集する. 要望表現辞書を用いて教師データを収集するに当たって重要となるのは, 手がかり表現の出現位置である. 要望表現辞書に定義した手がかり表現は文末に出現する傾向が極めて高く, 文末以外で出現する場合は, 要望とはならない可能性が高い. そこで本研究では, より高い精度で教師データを収集する為に, 手がかり表現の出現位置を考慮した, 以下の2段階の処理で教師データの収集を行う.

4.3.1 手がかり表現を含むツイートの抽出

4.2 節で作成した要望表現辞書を用いて, TwitterAPI から対象の商品名/サービス名を含み, かつ手がかり表現を含むツイートを抽出する. この際, 以下の除去対象となる語を含むツイートはスパム/ボットによる投稿である可能性が高いとして, 収集の対象外とした.

除去対象となる語の例

定期, 拡散, 速報, 価格, 最安, 料金, Line, Amazon, DM, ゲット, iTunes, 彼氏

4.3.2 手がかり表現が文末から 3-gram 以内に出現するツイートの抽出

先行研究において, 要望を含む投稿には, 文末に特徴的な表現が現れることが確認されている [14]. Twitter 上の手がかり表現を含む要望を収集した結果, 「～し/て/欲しい/なあ」や「～しろ/よ/アホ」など, 手がかり表現が, 文末から 3-gram 以内に出現する傾向が確認された.

この結果を元に, 4.3.1 節で抽出したツイートを文単位に分割し, 文末から 3-gram 以内に手がかり出現するものを選択し, 教師データとした. ツイートの文単位への分割には, 句点・空白・顔文字といった表現を区切り文字として使用し, これらの表現が出現した位置でツイートを分割し, 1 文と見なしている. また, 3-gram 以内に出現するかの判定には, 形態素解析

器である Mecab を使用した。MeCab を用いてツイートを形態素に分解し、手ごかり表現が、文末の形態素から 3 形態素以内の距離に存在しているかを確認することで判断している。

4.4 CNN による分類器の構築

4.3 節で収集した教師データを用いて分類器の構築を行う。分類器のアルゴリズムには CNN を用い、ネットワーク構造は Kim[3] らが提案したモデルを参考に行っている。図 4.2 に本提案手法におけるニューラルネットワークの構造を示す。畳み込み層 1 層、プーリング層 1 層のシンプルな構造を持ち、入力層では、 n 単語からなる文章を入力とし、各ラベル (本研究では要望または非要望の 2 値) の確率を出力としている。入力文に含まれる i 番目の単語の k 次元の分散表現を $x_i \in \mathbb{R}^k$ と表すと、文 $x_{1:n} \in \mathbb{R}^{n \times k}$ は以下の式で表される。

$$x_{1:n} = x_1 \oplus x_2 \oplus \dots \oplus x_n, \quad (4.1)$$

i 番目の単語から h 単語分のフレーズ $x_{i:i+h-1}$ に対して、 h 単語のウィンドウサイズを持つフィルタ $w \in \mathbb{R}^{h \times k}$ をかけ、以下の式で表される非線形関数に通すことで、特徴 c_i を計算する。

$$c_i = f(w \cdot x_{i:i+h-1} + b). \quad (4.2)$$

これより特徴マップ $c = [c_1, c_2, \dots, c_{n-h+1}]$ が得られる。さらにプーリング層 (Max Pooling) において、特徴マップ c の中から最大値を抜き出し、これをフィルタの特徴とする。また、本提案手法では、複数のフィルタを用意することで、複数の特徴を用いている。次に、これらの特徴群を全結合層において結合し、ソフトマックス出力層への入力とすることで最終的な出力値を得る。最後に予測値と実測値の誤差を計算し、誤差逆伝播法により重みを更新する。

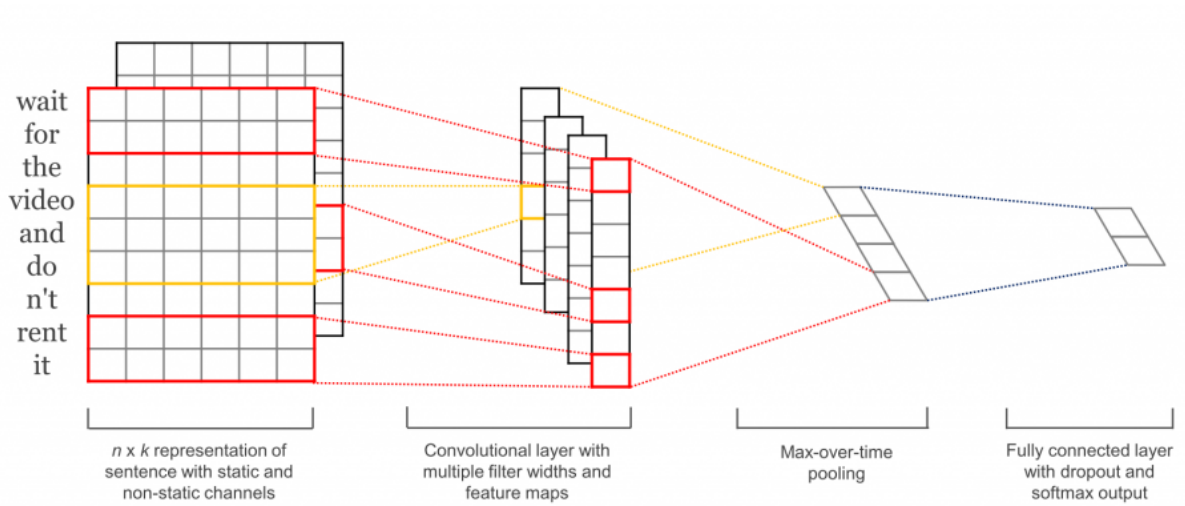


図 4.2 CNN モデルの構造 [3]

第 5 章

評価実験

5.1 評価対象

ソーシャルゲームは、近年急速な普及を見せており、多くの要望が Twitter 上に投稿されている。そこで本研究ではソーシャルゲームに関する投稿を評価対象とした。また、分類器を構築する際に使用する教師データとして、表 5.1 に示すソーシャルゲーム 10 タイトルに関する投稿を使用する。

評価用データは、2015 年 8 月から 2 ヶ月間に投稿されたツイートを以下の 2 通りの方法で収集した。この際、4.3.1 節で除去対象としたキーワードを含むツイートの除去を事前に行っている。

データセット (i) 教師データで使ったゲームタイトルを使用する

教師データとして使ったソーシャルゲーム 10 タイトルのタイトル名を含むツイートを各 200 件ずつ、合計 1000 件のツイートを収集した。この際、ゲームの公式 Twitter アカウントに対するリプライも収集対象としている。結果、要望を含む投稿は 76 件得られた。

データセット (ii) 教師データで使ったゲームタイトルを使用しない

教師データとして使っていないソーシャルゲームタイトル「白猫プロジェクト（白プロ、#白猫）」を含むツイートを 1000 件収集した。この際、ゲームの公式 Twitter アカウント（@wcat __ project）に対するリプライも収集対象としている。結果、要望を含む投稿は 79 件得られた。

以上の手法により収集したツイートに対して、クラウドソーシングサービスのランサーズを用いて、要望のラベルを人手で付与した。各ツイート毎に 5 名の参加者に回答してもらい、もっとも一致率の高い解答を正解ラベルとして付与した。この際、回答の質を向上させる為

に，100 ツイート毎に解答難度の低いダミーデータを用意し，ダミーデータへの回答を誤った参加者の解答を事前に除去している．回答者の判別の一致度を示す k 係数は，手法 (i) で 0.468，手法 (ii) で 0.548 となり，5 人の解答はおおむね一致していることが分かる．本研究では，以上の手順により作成したラベル付きツイートを評価用データとして使用する．

表 5.1 教師データの収集に使用するゲームタイトル一覧

ゲームタイトル	検索時に追加で使ったキーワード
パズル&ドラゴンズ	パズドラ, @pad __ sexy
モンスターストライク	モンスト, @monst __ mixi
刀剣乱舞	とうらぶ, @TOUKEN __ STAFF
艦隊これくしょん	艦これ, @KanColle __ STAFF
アイドルマスターシンデレラガールズ	モバマス, # imas __ cg, アイマス CG
ラブライブ! スクールアイドルフェスティバル	スクフェス, @lovelive __ SIF
Fate/Grand Order	fatego, FATEGO, @fgoproject, FateGo
ブレイブフロンティア	ブレフロ, @brave __ pr
魔法使いと黒猫のウィズ	黒ウィズ, 黒猫のウィズ, @colopl __ quiz
グランブルーファンタジー	グラブル, @gbf __ vee, @granbluefantasy

5.2 評価方法

提案手法と従来法の分類精度の比較を行うことによって、本手法の有効性を検証する。提案手法の有効性を検証するには、抽出したツイート集合がどれだけ正解しているかという正確性と、抽出した記事集合が全ての正解のうち、どれだけ正解を含んでいるかという網羅性の2つの観点からの評価が必要となる。本論文では、正確性を適合率 (precision), 網羅性を再現率 (recall), 適合率と再現率の調和平均である F 値 (F-measure) によって提案手法の抽出精度を評価する。それぞれの計算方法について、以下に示す。

$$precision = \frac{\text{抽出した正解ツイート数}}{\text{抽出したツイート数}} \quad (5.1)$$

$$recall = \frac{\text{抽出した正解ツイート数}}{\text{全ての正解ツイート数}} \quad (5.2)$$

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \quad (5.3)$$

5.3 実験設定

5.3.1 ハイパーパラメータの設定

CNN 学習時における目的関数には交差エントロピー誤差, 最適化には Adam を用いる。

ハイパーパラメータは, Kim らの手法 [3] や Zhang らの研究 [21] を参考にしつつ, グリッドサーチを行い, 以下の値に設定した。

- Embeddings 層における重みの初期化には, 事前に word2vec よって 3 万件のツイートを学習させることで獲得した 100 次元の単語分散表現を用いる
- 全結合層で dropout(dropout 率 0.5) を行う
- 畳み込み層におけるフィルタサイズには, [3,4,5] を用いる
- 各フィルタサイズ毎のフィルタ数は 100 枚使用
- ミニバッチサイズを 128 とし, 10 エポック学習を行う

5.3.2 実装

提案手法の実装

CNN の実装には, Python の深層学習用ライブラリである Keras を使用した。また, Keras のバックエンドには TensorFlow を用いている。

テキストの分かち書きには Mecab を使用しており，word2vec による分散表現の学習には，Python のトピックモデリング用ライブラリである gensim を用いた

比較手法の実装

比較手法として，栗原らの辞書ベース分類器 (DB)[11] と，筆者らが提案した SVM 分類器 (SVM+BoW)[13] を用いる．SVM 分類器の実装には Python の機械学習ライブラリである scikit-learn を使用した．また，SVM のカーネルには線形カーネルを使用している．パラメタの設定では，C パラメタの値を 1.0 から 10.0 まで 0.1 刻みで試して調整を行い，1.0 という値に設定している．素性には，単語の出現頻度などによって文書をベクトルで表現する形式である Bag of Words(BoW) を用いる．ただし，品詞が，名詞・動詞・形容詞・形容動詞・副詞・助動詞 のいずれかに該当しない単語は除外した．また，素性の構築時には 文章内において一定以上の割合で出現する単語を頻出語として除去している．

辞書ベース分類器の実装に当たっては，栗原らの論文を参考に，大塚らの論文中に記述されている要望表現リストを辞書の作成に使用した．データセット (i)，データセット (ii) の手順により収集した評価用データに対して，作成した辞書内の要望表現とマッチするかどうかの判定を行い，一致する場合に要望を含む投稿であると判定する．

5.4 実験結果・考察

実験結果を表 5.2 に示す．

表 5.2 実験結果

		適合率	再現率	F 値
データセット (i)	提案手法	0.26	0.46	0.34
	SVM+BoW	0.24	0.42	0.31
	DB	0.2	0.06	0.12
データセット (ii)	提案手法	0.34	0.61	0.43
	SVM+BoW	0.29	0.62	0.40
	DB	0.19	0.09	0.12

表 5.2 より，いずれの手法の評価データを用いた場合でも，ベースライン手法と比較して，適合率，F 値が向上していることが分かる．この要因の一つとしては，分類器の構築に単語分散表現と CNN を用いる事で，単語の意味的な類似度と文脈情報が考慮された結果，予測モデルの識別精度が向上した為であると考えられる．

また、データセット (i)、データセット (ii) を比較すると、教師データの収集時に使用していない未知のゲームタイトルに関しても、教師データの収集時と同様のゲームタイトルを使用したデータセット (i) と同等以上の分類性能を発揮していることが確認できる。逆に、データセット (i) よりもデータセット (ii) の場合に高い分類性能を発揮している。この要因としては、データセット (i)、データセット (ii) で収集したツイートの判別難易度の違いが考えられる。評価データのラベル付けを行った際の k 係数の値は、データセット (i) で 0.468、データセット (ii) で 0.548 となっており、データセット (i) で評価者の回答一致率が低いことが確認できる。このことからデータセット (i) の評価データは、人手でも判別の難易度の高いツイートを多く含んでいる可能性がある。

また、適合率に関しては、再現率と比較して大きく差が開く結果となった。この要因としては、まず、要望の対象が異なるツイートの判別難易度の高さが考えられる。今回の実験の抽出対象は、3章で挙げた条件に加えて、対象のソーシャルゲームに関する要望であるという2つの要件を満たすツイートに限定している。その結果、センチメント分析のように単純な2値分類のタスクよりも判別難易度が高くなると考えられる。これにより、分類精度が大きく低下してしまう可能性が高い。次に考えられる要因としては、Twitter 上の要望の人手による判別難易度の高さが挙げられる。今回使用した評価データセットは、いずれの手法を用いて収集した場合でも、5人の回答が一致しないケースが多く見受けられた。Twitter 上の投稿は、文章が省略されていたり、日本語が崩れていることが多い。加えて本研究で要望の対象としている不満に関するツイートは、暗にユーザの要望を示すあいまいな表現が多い為、より判別を困難にしていることが推測される。人手でも評価が一致しないツイートを抽出する事は困難である為、分類精度を低下させている可能性が高い。

5.5 追加実験

実験結果より得られた仮説を検証する為に、追加で以下の実験を行った。

5.5.1 「要望の対象が異なる」ツイートを要望に含めた場合との精度変化を比較

要望の対象が異なるツイートの判別難易度の高さが分類精度に与える影響を検証する為に、要望の対象が異なるツイートを抽出対象に含めた場合と含めなかった場合で、分類精度の変化を分析した。表 5.3 より、いずれの方法で収集した評価データに対しても、要望の対象が異なるツイートを抽出対象に含めた場合に、適合率・F 値が上昇していることが確認できる。この要因としては、分類器が、要望の対象が異なるツイートを、要望を含むツイートとして誤判定してしまっている事が考えられる。要望の対象を考慮した上で、より高い精度での分類を実現

する為には、要望の対象をより正確に識別する手法が必要であると考えられる。

表 5.3 要望の対象が異なるツイートを要望に含めた場合の分類結果

		適合率	再現率	F 値
データセット (i)	含める	0.39	0.44	0.42
	含めない	0.26	0.46	0.34
データセット (ii)	含める	0.56	0.67	0.61
	含めない	0.34	0.61	0.43

5.5.2 回答一致率毎の精度変化を比較

人手による判別の難易度が分類精度に与える影響を検証する為に、回答の一致数毎に評価データを用いて分類を行った。実験結果を図 5.1, 図 5.2 に示す。図 5.1, 図 5.2 からいずれの手法で収集した評価データにおいても、回答の一致数が 5 の時に最も F 値が高くなり、一致数が 3 の時に最も F 値が低くなる事が分かる。この結果より、人手による判別の難易度の高さが、分類精度に影響を与えていることが確認できる。また、一致数が 5 のツイートのみを使用した場合と一致数が 3 のツイートのみを使用した場合の F 値の差を比較すると、データセット (i) では、 -0.08 、手法 (ii) では -0.1 となっており、極端な精度の低下は見られなかった。以上の結果より、本手法の人手による判別難易度の高いデータに対するロバスト性が確認できる。

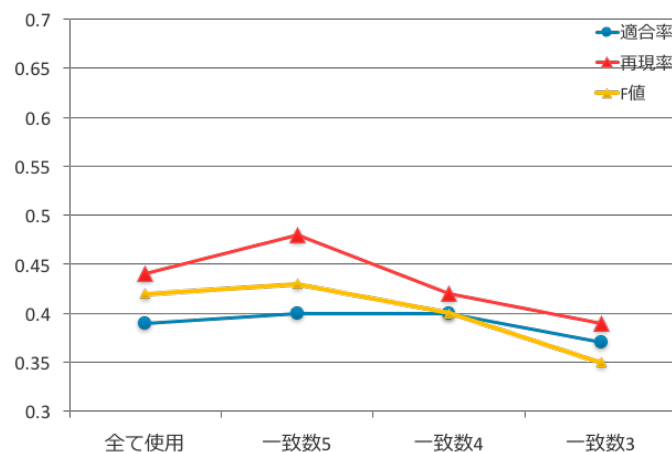


図 5.1 データセット (i) を対象とした回答一致率毎の精度変化

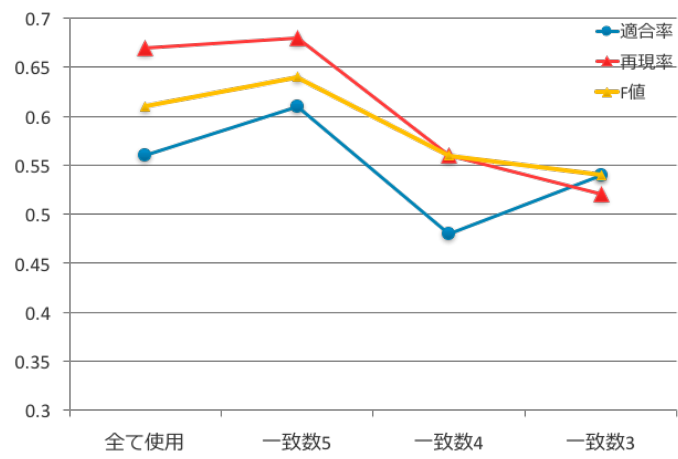


図 5.2 データセット (ii) を対象とした回答一致率毎の精度変化

第6章

結論

本論文では、Twitter 上からより高い精度で要望に関する投稿を抽出することを目的に、Twitter から消費者の要望を含む投稿の抽出手法を提案した。本手法では、要望表現の抽出に多層ニューラルネットワークの一つである畳み込みニューラルネットワークを用いる事で、従来手法と比較して、より高い精度での要望抽出に取り組んだ。また、教師データの収集に半教師あり学習の一つである Distant Supervision を適用することで、低コストな機械学習の実現を試みた。

評価実験では、ソーシャルゲームに関する投稿を対象とし、提案手法と従来手法の分類精度の比較を行うことによって、本手法の有効性を検証した。教師データの収集時に使用したゲームタイトル名を使用する場合と、使用しない場合の2通りの方法でソーシャルゲームに関する投稿を各1000件ずつ収集し、人手でラベル付けを行ったものを評価用データとして用意した。評価データに対して、構築した分類器を用いて分類を行った結果、いずれの方法で収集した評価用データに対しても、適合率、F 値において提案手法が高い評価を示し、有効性を確認できた。追加実験により、要望の対象が異なるツイートの判別難易度の高さが、分類精度にネガティブな影響を与えていることが検証された。また、回答一致率毎の精度変化を比較することで、人手による判別難易度の高い問題に対する本手法のロバスト性が確認できた。今後の課題としては、要望の対象が異なるツイートに対する識別精度の向上が挙げられる。

謝辞

指導教員である手塚太郎准教授は，研究の細部にわたり懇切丁寧な指導をしてくださいました．若林啓助教授には，副指導教員として研究について熱心なご指導を頂きました．心より感謝いたします．また，日常の議論を通じて多くの知識や示唆を頂いた手塚研究室の皆様にも感謝します．

参考文献

- [1] Fei Tian, Hanjun Dai, Jiang Bian, Bin Gao, Rui Zhang, Enhong Chen, and Tie-Yan Liu. A probabilistic model for learning multi-prototype word embeddings. In *COLING*, pages 151–160, 2014.
- [2] Maurice Peemen, Bart Mesman, and Henk Corporaal. Speed sign detection and recognition by convolutional neural networks.
- [3] Yoon Kim. Convolutional neural networks for sentence classification. In *In EMNLP*. Citeseer, 2014.
- [4] 萩原雅之. 次世代マーケティングリサーチ. ソフトバンククリエイティブ, 2011.
- [5] 奥村学. マイクロブログマイニングの現在 (第3回集合知シンポジウム). 電子情報通信学会技術研究報告. *NLC, 言語理解とコミュニケーション*, 111(427):19–24, jan 2012.
- [6] 野畑周, 内藤弘朗, and 清水徹. ヤフージャパンのリアルタイム検索における感情分析 (言語理解とコミュニケーション) – (第5回テキストマイニング・シンポジウム). 電子情報通信学会技術研究報告. *NLC, 言語理解とコミュニケーション*, 114(211):31–35, sep 2014.
- [7] 鈴木泰裕, 高村大也, and 奥村学. Weblog を対象とした評価表現抽出. 人工知能学会セマンティックウェブとオントロジー研究会 (*SIG-SW&ONT*), A401-02, 2004.
- [8] 奥村学, 南野朋之, 藤木稔明, and 鈴木泰裕. blog ページの自動収集と監視に基づくテキストマイニング. 人工知能学会セマンティックウェブとオントロジー研究会 (*SIG-SW&ONT*), A401-01, 2004.
- [9] 山本瑞樹, 乾孝司, 高村大也, 丸元聡子, and 大塚裕子. 文章構造を考慮した自由回答意見からの要望抽出. 言語処理学会第12回年次大会, 2006.
- [10] 大塚裕子, 内山将夫, and 井佐原均. 自由回答アンケートにおける要求意図判定基準. 自然言語処理, 11(2):21–66, 2004.
- [11] 栗原理聡, 佐々木彬, 松田耕史, 岡崎直観, and 乾健太郎. Twitter を利用した地域ごとの要望抽出. 第29回人工知能学会全国大会, pages 1–4, 2015.
- [12] 山本修平 and 佐藤哲司. Twitter からの実生活情報の抽出法の提案. 第4回データ工学と情報マネジメントに関するフォーラム (*DEIM Forum 2012*), F3-4, 2012.

- [13] 川島 崇秀, 佐藤 哲司, and 神門 典子. Twitter からの消費者ニーズの抽出手法に関する提案. 第 4 回データ工学と情報マネジメントに関するフォーラム *DEIM2016* 論文集, B5-1, 2016.
- [14] 大森晃. 要求抽出のための言語学的基礎論: 要求概念の定義, および要求の態度 (データベース, 一般論文). *情報科学技術フォーラム講演論文集*, 8(2):167–174, aug 2009.
- [15] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 1003–1011, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.
- [16] 三浦康秀, 服部圭悟, 大熊智子, and 増市博. Distant supervision による感性トピックの抽出. *富士ゼロックス テクニカルレポート*, (23):72–80, 2014.
- [17] 山本彩奈, 宮村祐一, 中田康太, and 岡本昌之. Deepdive を用いた web ニュース記事からの企業間関係抽出. *DICOMO シンポジウム 2015*, pages 172–179, 2015.
- [18] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [19] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
- [20] Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Advances in neural information processing systems*, pages 649–657, 2015.
- [21] Ye Zhang and Byron Wallace. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. *arXiv preprint arXiv:1510.03820*, 2015.