

# 能動学習を利用した未知語アノテーションの検討

黒澤 瞭佑<sup>†</sup> 酒井 哲也<sup>†</sup>

<sup>†</sup> 早稲田大学基幹理工学研究科情報理工・情報通信専攻 〒169-8555 東京都新宿区大久保 3-4-1

E-mail: †kuro-one@akane.waseda.jp, †tetsuyasakai@acm.org

あらまし Twitter を対象とした感情解析を扱う従来研究の多くでは、ツイートに含まれる未知語が解析精度向上の妨げとなっている。未知語の感情を獲得する方法としては、人力で一つ一つアノテーションを行うという方法が挙げられるが、それには膨大な時間的・人的コストが掛かると同時に、SNS の強みである情報の即時性が失われてしまうという問題がある。本研究は、能動的アノテーション手法によって優先的にアノテーションを行う必要のある未知語を含むツイートを厳選して感情推定を行うことで、情報の即時性を維持したまま、低コストでツイートの感情を推定し、機械学習に利用可能な推定結果を得る手法を提案する。

キーワード 自然言語処理, 感情解析, 能動学習, テキストマイニング, Twitter

## 1. はじめに

自然言語処理をはじめとする機械学習の分野では、ニューラルネットワークの技術を利用した研究が盛んに行われており、意味解析や構文解析、文書分類問題などの多くの分野で高精度の結果を残している [1]。その自然言語処理の一分野である意見分析や感情分析といった、テキストデータから書き手の意見や感情を抽出する研究においても、ニューラルネットワークを利用した解析が実現している [2] [3] [4]。

しかし、いずれの研究に関しても、モデルを学習するためには膨大なデータとそれに応じた計算時間、それを実現する強力な計算機を必要としており、大規模データや計算機が確保できない場合には想定した成果を得ることが出来ないという問題がある。また、学習の質や精度は、多くの場合、使用するデータの質や偏り、カテゴリなどに大きく依存するため、全ての文章に対して汎用性のあるモデルの作成は難しい。

特に、Twitter などの SNS サービスにおけるテキストデータを用いた研究には、省略語や短縮語、未知語などの一般的なでない言葉の出現や、投稿文字数の制限などによる崩し言葉や文法構造の乱れ、各語に対するデータ量や文法構造が一定でないなどのために、汎用性だけでなく、同じようなテキストデータを入力しても芳しい精度が得られないという問題がある [5] [7] [8]。また、学習に必要なラベルが付与されているデータが少なく、教師あり学習においては十分な性能を持つモデルを作成することが出来ないことも問題となっている。

一方で、以前から行われているニューラルネットワークを用いない研究には、テキストデータを単語レベルのベクトルとして教師あり学習を用いた分類を行うもの [5] や、単語に感情やその度合いを表す数値などのラベルを付与してテキスト全体の解析を行う研究などが存在する [6] [7] [8] が、新語や未知語の誤分解や単語情報の不足・欠損などが解析精度向上の妨げとなっていると述べられている。

そこで、本研究では未知語の感情に着目し、未知語を含むテ

キストデータへの感情アノテーションを、能動的ラベリングを用いて行うことでツイート全体の感情を機械学習に利用可能な形で推定する手法を提案する。

能動学習とは、機械学習の際に大量のラベル付きデータを用意出来ない場合、選択的に人手によるアノテーションを行うことで精度向上を図る教師あり学習の方法のひとつで、学習データが少ない場合に効果を発揮するものである [13]。この手法では人手によるアノテーションの前にアノテートすべきデータの選定を行うため、精度向上の結果に寄与するデータを効率良く見つけ出すことが可能である。

Twitter の未知語を含むテキストデータの感情ラベル付与にこれを適用することによって、テキスト全体の感情推定を行うモデルの精度向上を見込むことが出来ると考えられる。よって本研究では、未知語を含むテキストデータへのラベリングを、能動的アノテーション手法によって高効率・低コストでのポジティブ・ネガティブ判定（極性判定）を行うことで、機械学習に利用可能な推定結果を得ることを目標とする。

## 2. 関連研究

### 2.1 ニューラルネットワークを用いた感情解析

ニューラルネットワークを用いた感情解析の研究には以下のようなものがある。

Zhao ら [2] の研究では、Amazon のカスタマーレビューを基に、カスタマーレビューの評価の星の数で極性ラベルを付与した弱教師ありの事前学習を行った後、アノテータによって極性ラベルが付与されたデータを用いて fine tuning した LSTM でのレビュードキュメント単位での極性判定を実現していた。

Chen ら [3] の研究では、Zhao ら [2] の研究と同様に、商品のカスタマーレビューを対象にした LSTM を用いたドキュメント単位の極性判定を行った。その際、ユーザの好みや商品の特徴に Attention し、学習することで判定精度の改善に成功している。

Ye ら [4] の研究は、センテンスを単語レベルに分割して事前

学習した SentiNet と CNN を組み合わせることで、あらかじめ極性ラベルが付与されている SemEval のデータや映画のレビューの極性判定を行い、ベースラインであるシンプルな CNN の結果を上回った。

いずれの研究においても、あらかじめラベルの付与されたデータセットを用いた学習を前提としており、データにラベルの無い状態でこれらの手法を用いることは出来ないと言える。

## 2.2 感情語辞書や単語ベクトルを用いた感情解析

感情語辞書や単語ベクトルを用いた感情解析の研究には以下のようなものがある。

松林ら [5] の研究では、Word2Vec [18] からツイートの特徴ベクトルを生成し、喜怒哀楽の感情の推定を行った。また、堀宮ら [6] の研究では、ツイートに対する返信（リプライ）に着目し、人手によってラベル付けされたデータを用いて TF-IDF などを用いて特徴量を算出し、分類器の学習を行うことで 6 種類の感情の推定を試みた。さらに、村石ら [7] の研究では、感情語辞書と Yahoo!知恵袋のデータを利用してツイートの感情値とカテゴリを推定し、それと対になる感情を誘発させる行動を推薦する研究を行った。

山本ら [8] の研究では、ツイートの感情解析のために、文書内の単語間共起関係から感情語辞書を生成し、さらに新たな素性として Twitter 特有の顔文字に対して感情を付与した顔文字辞書を利用して感情推定を行った。

従来研究で得られた成果物を用いたものや、従来手法を用いて新たに生成したなどの違いはあるが、いずれも Twitter や Yahoo!知恵袋と言った多くの不定形なデータの感情解析をするために、感情語辞書を用いた研究を行っていることが分かる。

## 2.3 能動学習を利用した関連研究

能動学習を利用した研究として、Yang と Loog [12] の疑似アノテータを用いたクラス分類に関するものがある。まず少量のラベル付きデータから能動学習の学習器を基にラベル無しデータの中からラベル付けすべき候補を抽出し、選ばれた候補とその他のラベル無しデータの最近傍法の値を最小化する最終候補を 1 つ選択する。その後、選ばれた最終候補に複数の疑似アノテータを用いてランダムなラベリングを施したものを再び能動学習器を用いて候補を選んでいくというサイクルを作ることで、すべてのデータをラベリングするのではなく、ラベルの必要なもののみ選択することが可能となっている。また、選ばれた候補は、アノテーション可能な分だけ人間のアノテータに渡すことでラベル付きデータとなり、更なる精度向上に貢献している。

以上の研究により、未知語を含むツイートの感情推定においても、能動学習を用いてラベルの必要なデータのみを抽出することや、疑似アノテーションによってラベリングすることが有効な手段であると考えられる。

## 3. 提案手法

本研究では未知語を含むツイートに対して適切な感情推定を行うために、未知語を含むツイートに対して適切な感情アノ

テーションを行うことを目的として、以下の 3 つの手順からなる手法を提案する。図 1 は以下の手順を図式的に示した概要図である。

1 つ目の手順では、入力されたツイートを形態素に分割し、ツイート全体の感情を感情極性値として数値化する。この段階では未知語の感情は考慮せず、次の手順で機械学習を行うための前段階として、ツイート全体を通しての感情を機械的に推定する自動アノテーションを行う。

2 つ目の手順では、前の操作で得た結果を基に感情のラベル付けを行い、単語ベクトルを用いた機械学習をもってツイートの感情ラベルを予測するモデルを作成する。モデルは入力されたツイートに対してツイートの感情ラベルを予測し、その予測確率を算出するものである。

3 つ目の手順では、前の操作で得られたモデルを用いて、未知のツイートに対してラベル予測を行う。ここでは予測値が低いものに着目し、そこに未知語が含まれていれば、未知語の存在が正しい予測を妨げたと判断し、2 番目に予測値の高い感情をツイートの感情とするアノテーションを行う。

以上の手順の詳細を以下の項にて示す。

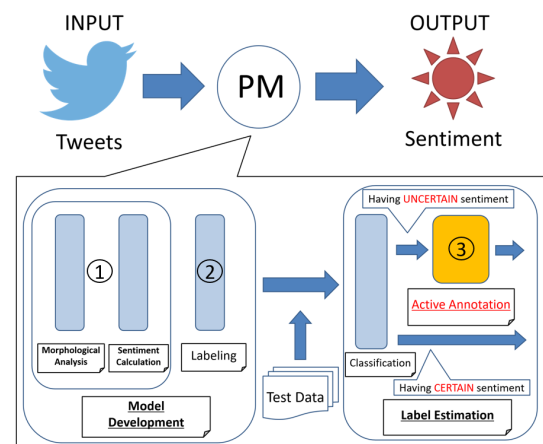


図 1 概要図

### 3.1 感情語辞書による自動アノテーション

本項では、感情語辞書を使ったツイートの自動アノテーションについて説明する。

ここでは、ツイート全体の感情はそのツイートに含まれる感情語の感情ラベルによって決定されていると仮定し、単語レベルに分解されたツイートから感情語辞書に含まれる単語を抽出することで、感情を推定するものとする。感情語辞書とは、単語に対して感情のラベルや値が与えられているものの集合を指し、ここでは先行研究で得られた 2 種類の感情語辞書を紹介する。

1 つ目は、高村ら [9] の研究で得られた単語感情極性対応表である。単語感情極性対応表には約 55,000 語の名詞、動詞、副詞、形容詞に対し、語彙ネットワークを用いて算出された  $-1.0$  から  $1.0$  までの値が感情極性値として付与されている。この場合、値が  $-1.0$  に近い程ネガティブで  $1.0$  に近いほどポジティブ

1	優れる:すぐれる:動詞:1
2	良い:よい:形容詞:0.999995
3	喜ぶ:よるこぶ:動詞:0.999979
4	褒める:ほめる:動詞:0.999979
5	めでたい:めでたい:形容詞:0.999645
6	賢い:かしこい:形容詞:0.999406
7	善い:いい:形容詞:0.999314
8	適す:てきす:動詞:0.999295
9	天晴:あっぱれ:名詞:0.999267
10	祝う:いわう:動詞:0.999122
11	功績:こうせき:名詞:0.999104
12	賞:しょう:名詞:0.998943
13	嬉しい:うれしい:形容詞:0.998871
14	喜び:よろこび:名詞:0.998861
15	才知:さいち:名詞:0.998771
16	徳:とく:名詞:0.998745
17	才能:さいのう:名詞:0.998699
18	素晴らしい:すばらしい:形容詞:0.998617
19	芳しい:かんばしい:形容詞:0.998578
20	称える:たたえる:動詞:0.998576

図 2 単語感情極性対応表の一例

な単語であると判断される。この感情語辞書は語彙数は多いが、一般的な語に関しては高精度で極性判定されているが、極性値が自動算出されたものであるため、あまり一般的でない語などに関しては正常な値でない場合があることが確認されている。

2つ目は、小林ら[10]、東山ら[11]の研究で得られた日本語評価極性辞書である。日本語評価極性辞書には約 5,000 語の用言と約 8,500 語の名詞および複合名詞に対しポジティブ、ネガティブのラベルが付与されている。こちらは単語感情極性対応表と比較して一般的な語が多く、付けられた感情ラベルは全て人手でアノテートされたものであるので信頼度の高いデータと言えるが、データ量が少ないため、正しい感情推定が出来ない場合があるという欠点がある。

本研究では、では、1つ目の感情語辞書である単語感情極性対応表を用いてツイートのアノテーションを行う。まず、ツイートを形態素解析ツールである MeCab を用いて単語レベルに分解する。形態素解析に使用した辞書は、新語や未知語に多く対応した mecab-ipadic-NEologd [17] を用いた。続いて、各単語の活用形を一般形に変換し、単語感情極性対応表に記載のある単語の形と揃えることで、より多くの単語に対して正しい感情値が得られるようにした。また、単語に感情値を与える際には下記の3つのルールを適用した。

1. 「かなり」「すごい」「とても」の単語の直後の単語には1.2 倍の感情値を与える。
2. 「ない」は否定の助動詞である場合、直前の単語の感情極性値に -1.0 倍の感情値を与える。
3. 後に出てくる単語ほど、ツイート全体の感情が表れやすいと仮定し、ツイート内の単語数に応じた重みを付与する。

上記のルールを適用し、ツイート中の各単語について感情極

1	ネガ (経験) →あがく
2	ネガ (経験) →あきらめる
3	ネガ (経験) →あきる
4	ネガ (経験) →あぎれる
5	ネガ (経験) →あぎれる た
6	ネガ (経験) →あせる
7	ネガ (経験) →あなどる
8	ネガ (経験) →あやしむ
9	ネガ (経験) →あやぶむ
10	ネガ (経験) →あやまる
11	ネガ (経験) →あらしう
12	ネガ (経験) →あわてる
13	ネガ (経験) →いたたまれない
14	ネガ (経験) →いたむ
15	ネガ (経験) →いつわる
16	ネガ (経験) →いらだたしい
17	ネガ (経験) →いら立つ
18	ネガ (経験) →うつうつ
19	ネガ (経験) →うなだれる
20	ネガ (経験) →うなる

図 3 日本語評価極性辞書 (用言編) の一例

140	あいさつ→e→~する (行為)
141	あいだ→e→~である・になる (状態) 客観
142	あいつら→e→~である・になる (状態) 客観
143	あいまい→n→~である・になる (評価・感情) 主観
144	あいまいさ→e→~がある・高まる (存在・性質)
145	あえぎ→e→~がある・高まる (存在・性質)
146	あかし→e→~がある・高まる (存在・性質)
147	あからさま→n→~である・になる (評価・感情) 主観
148	あきらめ→n→~がある・高まる (存在・性質)
149	あく→e→~がある・高まる (存在・性質)
150	あく抜け→p→~する (出来事)
151	あこがれ→p→~がある・高まる (存在・性質)
152	あこ→e→~である・になる (状態) 客観
153	あぎやか→p→~である・になる (状態) 客観
154	あし→e→~がある・高まる (存在・性質)
155	あせも→n→~である・になる (状態) 客観
156	あしたち→e→~である・になる (状態) 客観
157	あたたかさ→p→~がある・高まる (存在・性質)
158	あだ→n→~である・になる (評価・感情) 主観
159	あだおろそか→n→~である・になる (評価・感情) 主観
160	あだおろそか・あだやおろそか→n→~である・になる (評価・感情) 主観

図 4 日本語評価極性辞書 (名詞編) の一例

性値を計算した後、その総和を各ツイートの感情値とした。

### 3.2 機械学習用ラベルの選定およびテストデータのラベル推定

本項では 3.1 項で得られた感情値を基に機械学習用のラベルを決定する。

3.1 項で得られた感情値は未知語の感情を考慮していないが、「感情値が一定値以上のものは未知語を含んでいてもツイート本来の感情は変化しない」と仮定すると、閾値を設定することでツイートの感情極性がポジティブ、ネガティブ、ニュートラルのいずれかに分類することが出来る。分類されたツイートは属するクラスの特徴を持つということが出来るため、これらのクラスをラベルとした機械学習を行いモデルを作成することで、未知のツイートの持つ感情を推定することが可能となる。機械学習用のラベルは、\_label\_Positive, \_label\_Neutral, \_label\_Negative の3種とし、モデルの学習には skip gram のアーキテクチャを利用した fastText [19] を用いた。

fastText とは、対象となる単語の周辺語の共起確率を計算してベクトル化することで単語分散表現を学習する自然言語処



fastText によるモデル作成には、形態素に分割されたツイートとそれに対応するラベルが入力として必要になるため、3.1 項で作成したシステムで形態素に分けたツイートと 4.1 項で作成したラベルのペアを作成した。

モデル学習用に作成した入力データのラベリングの例を以下の図 6 に示す。

```

1 |label_Positive, ありがとう いつも 応援 する て くれる みんなの おかげ だね
2 |label_Positive, かんたん さん 制作 したので ずかん さんの 好き は 止まる ない
3 |label_Negative, よろしく お願ひ する ぞう
4 |label_Negative, おめでたう ごさる ます
5 |label_Negative, おげの さん さん さん さん さん
6 |label_Positive, 今年 も 宜しく お願ひ する ます
7 |label_Neutral, 遠く て も いい の 贈り物 た 姿 を さらす て も いい の ぶんぶん あはは
8 |label_Neutral, ちう 運を 争う の ちう 争う と お願ひ する たい たい の に ぶー
9 |label_Negative, SMTA は する ないん ずか
10 |label_Negative, おげ さん さん さん さん さん さん さん さん さん さん さん さん
11 |label_Positive, の 際 際
12 |label_Negative, まっしゅ あん
13 |label_Negative, まっしゅ あん さん さん さん さん さん さん さん さん さん さん さん
14 |label_Neutral, おれ 運 出る た わんた 終わる て る た
15 |label_Negative, おめでたう ごさる ます
16 |label_Neutral, おげの さん さん さん さん さん さん さん さん さん さん さん さん
17 |label_Neutral, 今年 も よろしく お願ひ します わ ご主人様
18 |label_Neutral, マリス する て 早く と 運の 日 運に 奥さん 長期 で 実家 帰る から 仕事 次第 で いらう ても 参加 する ます ぞ
19 |label_Positive, 白と 黒 で やる て る た 笑
20 |label_Negative, やまいて 様 こんばんは

```

図 6 閾値によるツイートのラベリングの実行例

### 4.3 未知語を含むツイートに対するアノテーションの実験

4.2 項で作成したモデルを用いて評価用データのラベルの予測を行う。評価用データは、2 人のアノテータによってラベル付けされた全 2,269 件のツイートの内、両者のラベルが一致した 1,622 件のツイートとした。

各アノテータがラベルを付与した全 2,269 件のツイートおよび一致した 1,622 件のツイートに付与されたラベルの内訳を以下の表 2 に示す。

表 2 アノテータによるラベル付け結果と一致したラベルの数

	Negative	Neutral	Positive	ALL
annotator_0	625	556	1,088	2,269
annotator_1	598	645	1,026	2,269
Matched	451	344	827	1,622

両方のアノテータの意見が一致した 1,622 件のツイートに対して、3.1 項で作成したモデルのみを用いて予測した手法 (System\_A) の他に、ラベルの予測確率の低いツイートに対して機械的に予測ラベルを入れ替えた手法 (System\_B) と未知語を含むツイートに対して能動的に予測ラベルを入れ替え手法 (System\_C) に対しての実験を行った。今回は、1 番目のラベルの予測確率が 0.6 未満だったものに限り予測ラベルの入れ替えの対象とした。機械的に予測ラベルを入れ替えたツイートは 115 件、能動的に予測ラベルを入れ替えたツイートは 20 件となった。また、利用したモデルは全て Model\_3 とした。

## 5. 結果と考察

### 5.1 実験結果と評価

各システムによる評価用データの正解ラベルの予測結果を適合率 (Precision)、再現率 (Recall)、F 値 (F-measure) で評価したものを以下に示した。モデルのみを用いた予測結果を表 3 に、機械的に予測ラベルを入れ替えたものを表 4 に、未知語を含むツイートに対して能動的に予測ラベルを入れ替えたものを表 5 にそれぞれ示した。また、ベースラインとしてランダムにラベリングしたラベル予測結果の評価を表 6 に示した。

表 3 モデルによるラベル予測結果 (System\_A)

	Precision	Recall	F-measure
Negative	0.3270	0.3059	0.6370
Neutral	0.7228	0.0756	0.4160
Positive	0.4503	0.1212	0.5033

表 4 予測確率に基づく機械的ラベル予測結果 (System\_B)

	Precision	Recall	F-measure
Negative	0.3245	0.2342	0.6477
Neutral	0.7073	0.0756	0.4135
Positive	0.4449	0.1143	0.5048

表 5 予測確率に基づく能動的ラベル予測結果 (System\_C)

	Precision	Recall	F-measure
Negative	0.3330	0.2935	0.6455
Neutral	0.7339	0.0785	0.4184
Positive	0.4581	0.1239	0.5077

表 6 ランダムラベルによるラベル予測結果 (System\_D)

	Precision	Recall	F-measure
Negative	0.2580	0.2218	0.5019
Neutral	0.3237	0.3430	0.3180
Positive	0.2871	0.2694	0.3893

結果を Tukey HSD 検定を用いて評価する [20] [21]。

TukeyHSD とは 2 つ以上のシステム間の統計的な有意差を検定する多重比較法のひとつで、各システムによる評価値が同様のシステムから出力されたと仮定し、各評価値を行列化したものの中で行をランダムに並べ替えた際の各列の平均を求め、その最大値と最小値との差の分布からシステムの p 値を算出する検定手法である。これにより、任意のシステム間の統計的有意差を調べることが出来る。

各システム ABCD は、提案システムおよびベースラインに以下の表 7 のように対応している。

表 7 各システム名と実システムの対応表

System_A	モデルによる予測
System_B	確率に基づく機械的予測
System_C	確率に基づく能動的予測
System_D	ランダムな予測

各システム間の Tukey HSD 検定の結果を図 7 に示す。

```

Multiple Comparison of Means - Tukey HSD,FWER=0.05
-----
group1 group2 meandiff lower upper reject
-----
A B -0.0055 -0.0498 0.0387 False
A C 0.0049 -0.0393 0.0492 False
A D -0.0863 -0.1305 -0.0421 True
B C 0.0105 -0.0337 0.0547 False
B D -0.0808 -0.125 -0.0365 True
C D -0.0912 -0.1355 -0.047 True
-----

```

図 7 TukeyHSD による検定結果



## 5.2 考察

### 5.2.1 混合行列による評価と考察

4.3項の結果より、提案手法を用いたラベル予測は、ベースラインとして設定したランダムに予測した予測結果をいずれも上回る結果となった。また、提案手法では作成したモデルのみを利用したものより、予測確率の低いラベルに対して能動的にアノテーションを行ったものの方が良い結果を残していることが分かる。予測確率の低いツイート全てに対して機械的にラベルを入れ替えるものが、モデルのみの予測結果を下回った結果としては、未知語が含まれていないツイートや未知語が含まれていてもツイート全体の感情には寄与しないものも含めて全てラベルを入れ替えてしまっているからであると推測できる。

機械的なラベリング手法と能動的なラベリング手法の結果を比較すると、ネガティブなツイートの判定を除き、能動的なラベリング手法がF値において機械的なラベリング手法を上回っていることが分かる。機械的に入れ替えたラベルが115件、能動的に入れ替えたラベルが20件であったことを考慮すると、少ないアノテーション量で精度向上に貢献していると言えるため、効率の良いアノテーションが出来たと考えられる。

また、正解のデータに対してシステムがどのようなラベルを予測しているかを調べるため、各システムにおける正解ラベルと予測ラベルの混合行列を生成した。生成した各システムの混合行列を以下の表8、表9、表10、表11に示す。

表8 モデルによるラベル予測の混合行列

	P-Negative	P-Neutral	P-Positive
T-Negative	326	23	102
T-Neutral	224	26	94
T-Positive	447	36	344

表9 予測確率に基づく機械的ラベル予測の混合行列

	P-Negative	P-Neutral	P-Positive
T-Negative	319	37	95
T-Neutral	227	26	91
T-Positive	437	48	342

表10 予測確率に基づく能動的ラベル予測の混合行列

	P-Negative	P-Neutral	P-Positive
T-Negative	331	22	98
T-Neutral	225	27	92
T-Positive	438	43	346

表11 ランダムラベルによるラベル予測の混合行列

	P-Negative	P-Neutral	P-Positive
T-Negative	146	146	141
T-Neutral	110	113	121
T-Positive	301	247	279

各混合行列を見ると、提案手法におけるネガティブなツイートに関しては真であるツイートが多く、正解に対して良く判定

出来ていることが分かる。しかし、ニュートラルやポジティブに関しては真と判定されたものが他に比べて少ないため、偽の判定をしていることが多いことが分かる。特にニュートラルに関しては本来はニュートラルであるのにネガティブと判定されたものがほとんどを占めていることが分かる。これらの誤推定は、感情語辞書による自動アノテーションと閾値を用いた学習用ラベルの決定の手順に原因があると考えられるため、以下で考察していく。

### 5.2.2 感情語辞書を用いた自動アノテーションとラベリングへの考察

感情語辞書を用いた自動アノテーションの前処理として単語に対する感情値付与のために行った形態素解析では、全ての形態素を基本形に直す、助動詞「ない」とその前の動詞を連結する、記号や絵文字などの削除などでツイートを整形する処理を行った。多くのツイートはその前処理によって正しく形態素に分割されていたが、人の名前や呼び名、漢字を用いずにひらがなで書かれた単語、基本形が正しく登録されていない未知語などに関しては誤分解が確認された。単語の誤分解によっては正しい感情語の取得が出来ない場合があるため、今後は単語の連結確率を考慮した形態素解析を行う、基本形を持たない単語はあらかじめ対象単語を用いるなど、誤分解を防ぐためのルールの追加することや、単語ではなく文字レベルの分割から単語の区切れ目を学習して単語を獲得するなどの処理を検討する必要があると言える。

Label	Score	Label	Score
_label__Positive	0.508758	_label__Negative	0.387849
_label__Neutral	0.487460	_label__Negative	0.278193
_label__Neutral	0.378372	_label__Positive	0.351782

図8 ツイートの誤分解の例

感情語辞書による自動アノテーションでは、感情語辞書をそのまま用いて感情値を算出したものより感情値の散らばりが少なくなったことによってそれぞれのツイートの感情の閾値を推定しやすくなったという点では効果があると考えられる。しかし、ヒストグラム中の最頻値である $-0.06$ から $-0.07$ の付近にも、それ以外の感情と判定できるツイートが見られたため、今回設定したルールを用いても感情語辞書内の数値と人間の感覚で知覚する感情との差を大きく埋めるには至らなかったと考えられる。設定するルールを増やすことやツイート内での感情値の重みの付け方を変化させることで改善することも可能であると考えられるが、今回の手法では感情語辞書の内容が結果に大きく影響するため、より人間の感性に近い感情語辞書を使用すること出来ればより劇的に結果を改善できると言える。

閾値を用いた学習用ラベル決定では、自動アノテーション結果を用いて閾値を設定して付与するラベルを決定し、それを用いた学習を行った。その際に表1に示したように閾値を変化させて数種類のモデルを作成したが、本実験ではF値を参照した予測結果が一番良いモデルとしてModel.3を採用してラベル予測をした結果、表8、表9、表10、表11のようにNeutralと

判定されるツイートの割合が低いことが分かった。

そこで本実験に採用したモデル以外を用いた予測結果から同様に混合行列を求めたところ、以下の表 12, 表 13 が得られた。

表 12 モデルによるラベル予測の混合行列 (Model\_1)

	P-Negative	P-Neutral	P-Positive
T-Negative	239	153	59
T-Neutral	181	113	50
T-Positive	345	220	262

表 13 モデルによるラベル予測の混合行列 (Model\_2)

	P-Negative	P-Neutral	P-Positive
T-Negative	292	73	86
T-Neutral	217	51	76
T-Positive	411	101	315

以上の表 12, 表 13 と 5.1 項の混合行列の表 8 とを比較すると、ニュートラルと判定できるツイートが増加することが分かった。また、Model\_1 に関しては、真のニュートラルが他のラベルと判定される割合が、他の 2 つのモデルと比べて高いことも分かった。ニュートラルと判断される閾値の範囲を広げることでこの結果が得られたため、これにより、感情語辞書による自動アノテーションによる結果は、ニュートラルな感情を持つツイートを収集する際にはある程度有効であるということが示された。

しかし、いずれの結果においてもポジティブ・ネガティブであるツイートの真の推定割合は低いため、それらの感情を持つツイートをより多く検出するための手法が必要となり、それを実現するためには、人間の知識を利用した能動的なアノテーションをより効果的な形で行うことが重要であると言える。

### 5.2.3 fastText の学習パラメータ設定と結果への考察

本実験でモデル生成の際に使用した学習設定パラメータは、学習回数である epoch を 1000 回と設定し、あとはデフォルト設定のままであった。そこで、設定したモデルの学習の際に設定した学習オプションを変化させた場合、結果にどのように影響するか調査した。

学習時にどのパラメータを設定して学習したかを以下の表 14 にまとめた。設定した閾値は全て本実験で利用した Model\_3 と同様のものとし、epoch 数を変更する Model\_X 以外は epoch 数 1000 回の設定をそのまま用いるとする。

表 14 パラメータのデフォルト値と各モデルの設定値

	Model_X	Model_Y	Model_Z
default	-epoch 5	-minCount 1	-dim 100
set	-epoch 100	-minCount 10	-dim 300

以下に各モデルの適合率 (precision), 再現率 (Recall),  $F$  値 (F-measure) をまとめた表 15, 表 16, 表 17 を示す。

以下には予測結果と正解ラベルの内訳である混合行列の表 18, 表 19, 表 20 を示す。

以上の結果より、学習パラメータの設定では推定結果が大き

表 15 Model\_X のラベル予測結果

	Precision	Recall	F-measure
Negative	0.3274	0.3038	0.6391
Neutral	0.7339	0.0698	0.4111
Positive	0.4528	0.1135	0.5004

表 16 Model\_Y のラベル予測結果

	Precision	Recall	F-measure
Negative	0.3202	0.3488	0.6402
Neutral	0.7361	0.0436	0.4196
Positive	0.4462	0.0775	0.5069

表 17 Model\_Z のラベル予測結果

	Precision	Recall	F-measure
Negative	0.3263	0.2941	0.6375
Neutral	0.7228	0.0727	0.4148
Positive	0.4497	0.1166	0.5026

表 18 Model\_X の混合行列

	P-Negative	P-Neutral	P-Positive
T-Negative	331	22	98
T-Neutral	226	24	94
T-Positive	454	33	340

表 19 Model\_Y の混合行列

	P-Negative	P-Neutral	P-Positive
T-Negative	332	14	105
T-Neutral	239	15	90
T-Positive	466	14	347

表 20 Model\_Z の混合行列

	P-Negative	P-Neutral	P-Positive
T-Negative	326	24	101
T-Neutral	225	25	94
T-Positive	448	36	343

く変化することはないことが分かる。これは元々の学習データ自体が推定結果に大きく影響していることを表している。ただし、設定回数未満出現する単語を切り捨てる minCount のパラメータは、感情を推定するのに必要な単語を切り捨ててしまっているため、他に比べて推定精度が下がっていることが確認できる。また、このパラメータ設定では本研究で着目している未知語の出現をなかったことにしてしまうため、今後の研究の際にも用いる必要がないと言える。

### 5.2.4 Tukey HSD 検定に対する考察

表 7 に示した Tukey HSD による検定では、提案モデルによるラベル予測 System\_A, 確率に基づく機械的ラベル予測 System\_B, 確率に基づく能動的ラベル予測 System\_C のいずれもランダムなラベル予測 System\_D において 5%水準で統計的に有意であるという結果が得られた。しかし、提案手法である System\_A, System\_B, System\_C 間ではいずれにおいても統計的有意差を得ることが出来なかった。

この理由としては、Tukey HSD 検定自身がシステム間におい

て統計的有意差を持つと結論付けることに対してかなり保守的な検定手法であるということに加え、ラベル推定の対象となった1,622件のツイートの内、System\_Bでは115件、System\_Cでは20件と入れ替えたラベルがごく少数に限られていたため、推定結果に対しての寄与率が低く、結果として有意差が得られなかったものと考えられる。

## 6. 結論と今後の課題

本研究ではツイートを対象とした感情推定を行うために、感情が付与されていない故に推定の妨げとなる未知語に着目し、未知語を含むツイートに能動的なアノテーションを行う手法を提案した。ツイート全体の感情を推定する上で用いた感情語辞書と設定したルールによる感情値推定手法および閾値を用いた学習ラベル作成は、ニュートラルなツイートを判定するには一定の効果があったが、ポジティブ・ネガティブなツイートを判定するには、更なるルール設定の必要性や、より人間の感性に近い感情語辞書を学習に用いる必要があると言える。未知語を含むツイートの能動的ラベリング手法に関しては、予測確率の低いものに限定して行ったため予測結果に大きく貢献することは無かったが、少ないアノテーションであっても予測結果を向上させることが出来た。今回は能動的アノテーションがどの程度予測結果に影響するかの知見を得ることが出来たため、今後はこれらのアノテーション結果を再度機械学習の学習データ内に取り入れて学習することで、人間の感性を学習したより良いモデルを作り出すことが出来るかどうか検証する必要があると考えられる。また、今回の研究のように感情語辞書を用いた感情推定研究の他に、深層学習を利用した感情推定手法などに能動学習を取り入れ、推定精度向上に寄与するか検証することで更なる展望を期待できると考えられる。

## 文 献

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, L., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I., "Attention is All You Need," arXiv, <https://arxiv.org/abs/1706.03762v5>, 6 December, 2017.
- [2] Zhao, W., Guan, Z., Chen, L., He, X., Cai, D., Wang, B., Wang, Q., "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis," IEEE Transactions on Knowledge and Data Engineering, Vol.30, No.1, pp.185-197, January 2018.
- [3] Chen, H., Sun, M., Tu, C., Lin, Y., Liu, Z., "Natural Sentiment Classification with User and Product Attention," Conference on Empirical Methods in Natural Language Processing, pp.1650-1659, November 2016.
- [4] Ye, Z., Li, F., Baldwin, T., "Encoding Sentiment Information into Word Vectors for Sentiment Analysis," Proceedings of the 27th International Conference on Computational Linguistics, pp.997-1007, August 2018.
- [5] 松林圭, 五味京祐, 古川和折, 松尾祐佳, 松原良和, 日諸マルセロ優次, 中村拓哉, 山下晃弘, 松林勝志, "Twitter上に投稿された文章に基づく感情推定手法とその応用に関する検討," 情報処理学会第78回全国大会, 2016.
- [6] 堀宮ありさ, 板野遼平, 佐藤晴彦, 小山聡, 栗原正仁, 沼澤政信, "Twitterにおける発話者へのリプライを用いたユーザ感情推定手法," DEIM Forum 2012 F2-1, 2012.
- [7] 村石将嗣, 北山大輔, "ツイートの感情推定に基づく対になる

- 感情を誘発する行動の推薦手法," DEIM Forum 2016, P2-2, 2016.
- [8] 山本湧輝, 熊本忠彦, 灘本明代, "Twitter特有表現を考慮したツイートの多次元感情抽出手法の提案," 情報処理学会関西支部 支部大会, 2014.
- [9] 高村大也, 乾考司, 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 Vol. 47 No. 2, pp.627-637, 2006.
- [10] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, "意見抽出のための評価表現の収集," 自然言語処理, Vol. 12, No. 3, pp.203-222, 2005.
- [11] 東山昌彦, 乾健太郎, 松本裕治, "述語の選択選好性に着目した名詞評価極性の獲得," 言語処理学会第14回年次大会論文集, pp.584-587, 2008.
- [12] Yang, Y., Loog, M., "Single Shot Active Learning using Pseudo Annotators," arXiv, <https://arxiv.org/abs/1805.06660v1>, 17 May, 2018.
- [13] Settles, B., "Active Learning Literature Survey," Computer Science Technical Report 1648, University of Wisconsin-Madison, Updated on 26 January, 2010.
- [14] Joulin, A., Grave, E., Bojanowski, P., Mikolov, T., "Bag of Tricks for Efficient Text Classification," arXiv, <https://arxiv.org/pdf/1607.01759v2.pdf>, 7 July, 2016.
- [15] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T., "Enriching Word Vectors with Subword Information," arXiv, <https://arxiv.org/pdf/1607.04606v1.pdf>, 15 July, 2016.
- [16] "MeCab: Yet Another Part-of-Speech and Morphological Analyzer," <http://taku910.github.io/mecab/>.
- [17] "mecab-ipadic-NEologd: Neologism dictionary for MeCab," <https://github.com/neologd/mecab-ipadic-neologd/blob/master/README.ja.md>.
- [18] "Word2Vec," <https://deeplearning4j.org/ja/ja-word2vec>.
- [19] "fastText," <https://fasttext.cc/>
- [20] 酒井哲也, "情報アクセス評価方法論 検索エンジン評価のために," コロナ社, pp.158-166, 2015.
- [21] Sakai, T., "Laboratory Experiments in Information Retrieval: Sample Sizes, Effect Sizes, and Statistical Power," Springer, 2018.