

単語埋め込み技術の違いによる日本語ツイートの感情分類精度の比較実験

Comparative experiments on sentiment classification of Japanese tweets using distributed representation

佐藤 一輝 尾崎 知伸
Kazuki Sato Tomonobu Ozaki

日本大学文理学部情報科学科

Department of Information Science, College of Human and Sciences, Nihon University

In this paper, we report the results of sentiment classification of Japanese tweets using distributed representation. In the experiments, six existing methods are employed to obtain a distributed representation of each tweet, and classifiers are built by K-NN, Random Forest and Support Vector Machine, respectively. We evaluate each distributed representation from the viewpoint of classification accuracy on eleven emotions.

1. はじめに

Twitter とは、ツイートと呼ばれる 140 文字以内の短文を投稿できるサービスである。ブログや他の SNS と比べて気軽に投稿が行えるという性質から、ツイートにはユーザの感情が現れやすいと考えられる。

本研究では、ツイートに対する感情推定の問題を取り上げる。一般にツイートは 1 つの文が短くかつ文法的に崩れたものが多いため、形態素解析が難しく、単純に感情を表す単語（感情語）とのマッチングを行うだけでは十分な推定精度を得ることができない。また、既存研究の多くは感情極性（正の感情か、負の感情か）の推定を対象としており、喜怒哀楽を含むより詳細な感情分類を対象とした研究は必ずしも多くない。本研究では、学習に用いるツイートの表現方法や学習アルゴリズムの違いがどの程度感情推定に影響するのか、また、感情の違いによって推定の難易度がどの程度異なるのかを明らかにするため、感情表現辞典 [1] に記載されている 10 感情（喜、好、安、哀、厭、怖、怒、恥、昂、驚）を対象とし、分散表現技術を含む 6 種のベクトル化技術と 3 種の機械学習手法を用いて感情推定（感情分類）実験を行った。

2. 関連研究

日本語のテキストに対して感情分析を行う研究は数多く行われている。堀宮ら [2] は Twitter においてユーザの発言に対する反応であるリプライを利用して投稿者の感情を推定する手法を提案している。Ekman の定義した基本 6 感情「幸福感」、「驚き」、「恐れ」、「悲しみ」、「怒り」、「嫌悪」を対象に、2-gram を素性とした TF-IDF 法を用いて特徴ベクトルを生成し、SVM を用いて分類器を構築している。また三好ら [3] は、ツイートの顔文字を構成する記号の出現頻度や組み合わせに着目し、記号の感情別出現率を算出した後、顔文字の感情を「喜び」、「哀しみ」、「怒り」、「驚き」、「照れくささ」の 5 種類に分類している。顔文字によるツイートの極性分類も行っており、「喜び」、「照れくささ」を正の感情、「悲しみ」、「怒り」、「驚き」を負の感情として、ツイートに出現した顔文字の出現回数を基に、ツイートの極性を推定している。池上らの研究 [4] では、顔文字と顔文字を除いたツイート文の 2 つの感情極性の組み合わせ

がツイート文の感情極性に与える効果を分析しており、顔文字の役割を考慮してツイートの感情極性を推定する手法を提案している。その際、感情表現辞典 [1] を利用し、正負を表す 2 次元ベクトルの感情極性値を割り当てた感情語辞書を作成し、顔文字を除いたツイート文の感情極性を推定している。山本らの研究 [5] では、顔文字に加え、Twitter 特有の表現として「やったあああああ」など語尾の母音を 3 回以上繰り返す表現を考慮した感情抽出手法を提案している。感情推定のために定量化された感情語辞書を作成し、ツイート文と Twitter 特有表現の感情を、10 種類に分類している。

これらの研究に対し、本研究では、ツイートの表現方法や対象とする感情の違いが、推定精度に与える違いを考察することを主な目的としている。

3. 感情分類実験

3.1 データセット

本研究では、2015 年に投稿されたツイートの内、感情表現辞典 [1] に掲載されている感情語を含むツイートを対象とした。ツイートの取得には、Twitter Streaming API^{*1} を用いた。前処理としてリプライ、リツイート、url を含むツイートを省き、ハッシュタグを取り除いたうえで、各ツイートに対して、「喜」、「好」、「安」、「哀」、「厭」、「恐」、「怒」、「恥」、「昂」、「驚」の 10 種類の感情と 10 種類以外の感情を含む「その他」から各ツイートが含む感情（正解クラス）の付与を行った。なお、1 つの感情（正解クラス）につき、感情語が表す感情とツイートが含む感情が一致するツイートを 120 ツイート、一致しないツイートを 120 ツイート準備した。結果として得られた 2400 ツイートのクラス分布を表 1 に示す。

3.2 ツイートのベクトル化とパラメタ設定

本論文では、Bag-of-Words, TF-IDF, Word2Vec, Paragraph Vector, ScoreSent2Vec, Tweet2Vec の 6 種類の既存手法を用いてツイートのベクトル表現を獲得した。以下それぞれの概要を示す。

Bag-of-Words

実際の実験では、形態素解析器 MeCab を使用し、名詞と形容詞、動詞を抽出することで Bag-of-Words によるベクトル表現を得た。

連絡先: 尾崎 知伸, 日本大学 文理学部 情報科学科, 〒156-8550 東京都世田谷区桜上水 3-25-40, tozaki@chs.nihon-u.ac.jp

*1 <https://dev.twitter.com/streaming/overview>

TF-IDF

TF-IDF 法は、文書内の単語の出現頻度を表す指標 TF と、その単語が現れる文書頻度の逆数である IDF の積である。すなわち、下式によって、文書 d における単語 t の値を求める。

$$TF_{t,d} = \frac{n_{t,d}}{\sum_k n_k}, \quad IDF_t = \log 2 \left(\frac{|D|}{DF_t} \right)$$

ここで、 $n_{t,d}$ は文書 d 中の単語 t の出現数、 $\sum_k n_k$ は文書 d 内に出現する全単語の出現数、 $|D|$ は全文書数、 DF_t は単語 t が出現する文書数である。Bag-of-Words と同様に形態素解析器 MeCab を使用し、名詞と形容詞、動詞を抽出した上で、各ツイートにおける各語の TF-IDF 値を計算した。

Word2Vec

Word2Vec[6] とは、単語を数百次元の固定長のベクトルとして表現する手法である。隠れ層と出力層の 2 層からなるニューラルネットワークであり、実際には CBOW(Continuous Bag of Words) モデルと Skip-gram モデルの 2 種類のニューラルネットワーク構造が提案されている。本実験では CBOW モデルを使用した。CBOW モデルは、前後の単語を利用して、 t 番目の単語 $w(t)$ を推定する。すなわち、 $w(t)$ の前後の単語 $w(t-c), \dots, w(t-1), w(t+1), \dots, w(t+c)$ を入力とし、対象となる単語 $w(t)$ を出力するニューラルネットワークである。

MeCab を使用して得られる、分かち書きされた単語を Word2Vec の CBOW モデルを用いてベクトル化した。パラメータとしてベクトルの次元数を 100, 200, 300, 400, 500 の 5 通り、window size を 6, 7, 8, 9, 10 の 5 通り、ネガティブサンプリングに使用する単語数を 0, 5, 10, 15, 20 の 5 通り、iteration を 5, 10, 15, 20 の 4 通りの計 500 通りのパラメータ設定それぞれで単語の分散表現を得た。なお、Word2Vec を使用して得られる単語に対する分散表現を利用して、(単語の集合・系列である) ツイートをベクトル化する方法として、単語ベクトルの各要素に対し、平均値、最大値、中央値の 3 種類を用いている。

Paragraph Vector

Paragraph Vector[7] とは、Word2Vec の拡張であり、(単語ではなく) 文書を数百次元の固定長ベクトルとして表現する手法である。Word2Vec と同様に、CBOW モデルを拡張した PV-DM(Paragraph Vector with Distributed Memory) モデルと Skip-gram モデルを拡張した PV-DBOW(Paragraph Vector with Distributed Bag of Words) モデルの 2 種類のニューラルネットワーク構造が提案されている。PV-DM モデルでは、対象となる単語 $w(t)$ を推定するために、前後の単語 $w(t-c), \dots, w(t-1), w(t+1), \dots, w(t+c)$ に加え、文書固有の ID を利用する。

本実験では PV-DM モデルを使用し、分かち書きされたツイートを対象に分散表現を獲得した。パラメータとしてベクトルの次元数を 100, 200, 300, 400 の 4 通り、window size を

6, 7, 8, 9, 10 の 5 通り、出現回数が少ない単語を破棄する min count を 3, 4, 5, 6, 7 の 5 通り、ネガティブサンプリングに使用する単語数を 5, 10, 15, 20 の 4 通り、iteration を 20 とした、計 400 通りのパラメータ設定それぞれでツイートの分散表現を得た。

ScoreSent2Vec

ScoreSent2Vec[8, 9] とは、Paragraph Vector の拡張であり、低次元の数値ベクトルとして与えられる付随情報 (score vector) を同時に予測するニューラルネットワークである。Paragraph Vector の PV-DM モデルを拡張した SPV-DM (Scored Paragraph Vector with Distributed Memory) モデルと PV-DBOW モデルを拡張した SPV-DBOW (Scored Paragraph Vector with Distributed Bag-of-Words) モデルの 2 種類のネットワーク構造が提案されている。本実験で使用した SPV-DM モデルは、PV-DM モデルと同様に文書固有の ID から、対象となる単語 $w(t)$ と score vector の値を同時に予測する問題をニューラルネットワークに学習させることで、付随情報を考慮した文書のベクトル表現を獲得する。

実験では、パラメータとしてベクトルの次元数を 100, 200, 300, 400 の 4 通り、window size を 6, 7, 8, 9, 10 の 5 通り、出現回数が少ない単語を破棄する min count を 3, 4, 5, 6, 7 の 5 通り、ネガティブサンプリングに使用する単語数を 5, 10, 15, 20 の 4 通り、iteration を 20 とした、計 400 通りのパラメータ設定それぞれでツイートの分散表現を得た。また、付随情報 (score vector) として感情ごとの感情語の出現数、ツイートの文字数、出現した単語の極性値の和の 3 種類を使用した。ここで極性値の計算には [10] の単語感情極性対応表を使用した。

Tweet2Vec

Tweet2Vec[11] とは、ツイートからハッシュタグを予測するニューラルネットワークである。時系列データを対象とした LSTM(Long Short Term Memory) の 1 つである GRU(Gated Recurrent Unit) を用いてツイートを文字単位で学習する。ツイートを前から処理をする GRU と後ろから処理をする GRU の 2 つの出力を用いてツイートをベクトル化する。Tweet2Vec は単語単位ではなく、文字単位で学習するため、表記ゆれや単語の崩れが多くみられる SNS の投稿の処理に向いていると考えられる。また、日本語は 1 文字当たりの情報量が英語より多いと考えられるため、Tweet2Vec は日本語テキストの処理に向いていると考えられる。

Tweet2Vec は本来ハッシュタグの予測を目的としているが、今回はハッシュタグの代わりに感情語を用いた。またパラメータとして、ベクトルの次元数を 400, 500, 600 の 3 通り、文字のルックアップテーブルの次元数を 100, 150, 200, 250, 300 の 5 通り、生成する正規乱数の標準偏差を 0.1, 0.01 の 2 通り、iteration を 20, 30, 40 の 3 通り、計 90 通りのパラメータ設定それぞれでツイートの分散表現を得た。

3.3 実験結果と考察

実験には、python の機械学習ライブラリ scikit-learn で提供されている 3 種の学習アルゴリズム、k 近傍法、ランダムフォレスト、SVM (one-against-rest 法) を用いた。なお、学習時におけるパラメータはデフォルトのものを使用している。

各パラメータ設定に対してそれぞれ 6 分割交差検定を用いてモデルの F 値を評価した。各ベクトル化技術及び学習アルゴリズムにおいて、F 値に関する最良値を表 2 に示す。表中の ScoreSent2Vec(len) はツイートの文字数、ScoreSent2Vec(sentiment words) は出現した感情語の個数、Score-

表 1: 感情ラベル内訳

感情の種類	出現数	感情の種類	出現数
喜	152	恐	140
好	184	怒	135
安	158	恥	133
哀	185	昂	170
厭	200	驚	169
		その他	774

表 2: 実験結果

ベクトル化技術	喜	好	安	哀	厭	恐	怒	恥	昂	驚	その他
---------	---	---	---	---	---	---	---	---	---	---	-----

K 近傍法の実験結果

Bag-of-Words	0.077	0.138	0.155	0.048	0.084	0.088	0.08	0.109	0.178	0.044	0.355
TF-IDF	0.273	0.204	0.296	0.280	0.218	0.177	0.267	0.417	0.18	0.268	0.405
Word2Vec(average)	0.082	0.165	0.046	0.111	0.103	0.021	0.072	0.065	0.069	0.012	0.446
Word2Vec(median)	0.076	0.132	0.116	0.063	0.145	0.052	0.077	0.034	0.082	0.09	0.456
Word2Vec(max)	0.129	0.212	0.051	0.156	0.129	0.049	0.099	0.012	0.041	0.052	0.462
ParagraphVector	0.085	0.038	0.06	0.112	0.118	0.032	0.022	0.062	0.056	0.033	0.396
ScoreSent2Vec(len)	0.082	0.077	0.115	0.172	0.207	0.079	0.066	0.134	0.091	0.157	0.516
ScoreSent2Vec(sentiment words)	0.674	0.659	0.813	0.638	0.492	0.838	0.906	0.851	0.712	0.746	0.691
ScoreSent2Vec(polarity value)	0.112	0.095	0.109	0.152	0.204	0.047	0.056	0.022	0.087	0.035	0.443
Tweet2Vec	0.158	0.161	0.152	0.129	0.106	0.107	0.128	0.272	0.092	0.182	0.404

ランダムフォレストの実験結果

Bag-of-Words	0.23	0.207	0.29	0.215	0.133	0.158	0.192	0.491	0.143	0.228	0.485
TF-IDF	0.231	0.243	0.178	0.145	0.119	0.178	0.199	0.516	0.105	0.131	0.494
Word2Vec(average)	0.158	0.083	0.039	0.098	0.098	0.078	0.109	0.131	0.118	0.106	0.472
Word2Vec(median)	0.09	0.158	0.084	0.085	0.1	0.039	0.143	0.075	0.073	0.06	0.488
Word2Vec(max)	0.104	0.213	0.073	0.111	0.146	0.109	0.078	0.043	0.041	0.052	0.494
ParagraphVector	0.067	0.082	0.063	0.126	0.102	0.059	0.067	0.043	0.06	0.038	0.436
ScoreSent2Vec(len)	0.229	0.214	0.247	0.185	0.247	0.067	0.195	0.221	0.199	0.293	0.477
ScoreSent2Vec(sentiment words)	0.733	0.692	0.84	0.697	0.548	0.875	0.901	0.892	0.706	0.757	0.707
ScoreSent2Vec(polarity value)	0.102	0.132	0.067	0.094	0.181	0.029	0.029	0.102	0.113	0.042	0.402
Tweet2Vec	0.085	0.079	0.047	0.103	0.109	0.04	0.069	0.052	0.06	0.054	0.458

SVMの実験結果

Bag-of-Words	0.294	0.272	0.307	0.319	0.253	0.345	0.317	0.545	0.212	0.358	0.499
TF-IDF	0.241	0.228	0.285	0.332	0.186	0.279	0.271	0.487	0.201	0.261	0.493
Word2Vec(average)	0.202	0.121	0.136	0.183	0.049	0.028	0.163	0.186	0.133	0.214	0.511
Word2Vec(median)	0.109	0.13	0.138	0.111	0.056	0.061	0.28	0.156	0.106	0.14	0.499
Word2Vec(max)	0.12	0.118	0.064	0.126	0.082	0.078	0.264	0.127	0.141	0.08	0.515
ParagraphVector	0.04	0.046	0.045	0.084	0.108	0.01	0.082	0.027	0.062	0.022	0.38
ScoreSent2Vec(len)	0.221	0.225	0.178	0.229	0.087	0.108	0.052	0.27	0.221	0.28	0.509
ScoreSent2Vec(sentiment words)	0.729	0.704	0.854	0.755	0.608	0.848	0.891	0.916	0.779	0.776	0.726
ScoreSent2Vec(polarity value)	0.089	0.038	0.033	0.03	0	0	0.014	0.028	0	0.022	0.48
Tweet2Vec	0.07	0.07	0.112	0.106	0.151	0.09	0.18	0.211	0.075	0.277	0.491

ScoreSent2Vec(polarity value) は出現した単語の感情極性値の和をそれぞれ付随情報として使用したものである。

ベクトル化手法に着目すると、Paragraph Vector より Bag-of-Words や TF-IDF の F 値が高い値となった。データセットが少なかつたため、学習が十分ではなかつたこと、感情語をもとに感情ラベルを付与したため、感情語の有無が直接現れる Bag-of-Words や TF-IDF の方がうまく分類できたと考えられる。同様に、付随情報として、出現した感情語の個数を使用した ScoreSent2Vec は、ツイートの文字数や単語の極性値の和を使用するよりも高い結果となった。

Tweet2Vec が日本語テキストの処理に向いていると予想したが、必ずしも高い精度を得ることができなかった。これはデータセットの少なさに加えて、Tweet2Vec は感情語を分類するニューラルネットワークを学習するため、感情語とツイートの感情が一致しないツイートの感情分類がうまくできていないと考えられる。

次に、感情に着目して精度を考察する。「その他」は他の感情に比べると高い精度が得られている。これはデータセット中

に「その他」が多く含まれるためだと考えられる。一方、各ベクトル化手法等で得られる F 値を感情ごとに平均すると、k 近傍法では、「恥」が高く、次いで「安」、「好」、「哀」が高い結果となった。この結果は、ランダムフォレストや SVM でも同様であり、ランダムフォレストでは「恥」が最も精度が良く、「好」、「喜」、「怒」がそれに続き、一方 SVM では精度の高い順に「恥」、「怒」、「驚」、「哀」となった。また、どの学習アルゴリズムを用いても、「恐」、「昂」に対して低い推定結果が得られた。

4. まとめ

本研究では、日本語ツイートの感情分類を対象に、11 の感情、6 種のベクトル化手法及び 3 種の学習アルゴリズムによる精度の違いを考察した。

今後の課題として、感情の分類数を考えることがあげられる。データセットが少なく、感情語を含むツイートのみ感情ラベルを付与していたため、感情語を含まないツイートに対して感情ラベル付与および、データセットの増量が必要であ

る。また、感情ラベルの付与の手間を省くため、半教師付きでの学習を目標としたい。

参考文献

- [1] 中村明: 感情表現辞典, 東京堂出版 (1993)
- [2] 堀宮ありさ, 坂野遼平, 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 (2012)
- [3] 三好辰明, 太田学: ツイートに出現する顔文字等の文字と記号に着目した感情分類, DEIM Forum 2013 (2013)
- [4] 池川知里, 新妻弘崇, 太田学: 顔文字の役割を利用したツイートの感情極性推定, DEIM Forum 2014 (2014)
- [5] 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, 2014 年度情報処理学会関西支部 支部大会 (2014)
- [6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado and Jeffrey Dean: Distributed Representations of Words and Phrases and their Compositionality, *Advances in Neural Information Processing Systems* (2013)
- [7] Quoc V. Le and Tomas Mikolov : Distributed Representations of Sentences and Documents, *CoRR*, (2014)
- [8] 橋戸拓也, 新妻弘崇, 太田学: ParagraphVector への追加情報の効率的な埋め込み, 情報処理学会研究報告, Vol. 2015-DBS-162, No. 11, pp. 1-8 (2015)
- [9] 橋戸拓也, 新妻弘崇, 太田学: Paragraph Vector1 へ埋め込む有効な付随情報の検討, DEIM Forum 2016 (2016)
- [10] 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.2, pp.627-637 (2006)
- [11] Bhuwan Dhingra, Zhong Zhou, Dylan Fitzpatrick, Michael Muehl and William W. Cohen.: Tweet2Vec: Character-Based Distributed Representations for Social Media, *CoRR*, (2016)