

日本語文章からの感情推定難易度判定

Estimating the Degree of Difficulty to Extract Writer's Emotion in Japanese Sentences

山下 紗苗^{*1} 上 泰^{*1} 奥村 紀之^{*2}
Sanae YAMASHITA Yasushi KAMI Noriyuki OKUMURA

^{*1}明石工業高等専門学校 ^{*2}大手前大学
National Institute of Technology, Akashi College Otemae University

Existing emotion estimation system always estimates some emotion from sentences. However, some sentences are difficult to estimate by humans. In order for the dialogue system to estimate emotion like humans, we think that it should also give results that "emotion estimation is difficult" as humans do. Therefore, we aim to construct a system to calculate the estimation difficulty level when estimating the author's emotions in Japanese sentences. Emotion estimation difficulty was judged by combination of classification by classifier, whether there is negative expression, emotional expression or not. F_1 value exceeded 80%.

1. はじめに

SNS の普及とともに自分の書いた文章が他人に読まれる機会も増えているが、同時に意図しない投稿の広がりや誹謗中傷なども増加している。SNS の投稿には感情が含まれていることが多く、投稿に含まれる感情を分析できればこれらの問題を未然に防げる可能性がある [松林 16]。

既存の感情推定システムは、文章から何かしらの感情を推定できることが前提となっている。しかし実際は人間にとって感情推定しにくい文章も少なからずある [山下 18]。我々は、対話システムがより人間らしい対応をするためには、システムは全ての文章について何らかの感情推定結果を出すのではなく、人間と同様に「感情推定が困難である」といった結果も出すべきと考える。そのため本論文では、日本語文章から筆者の感情を推定する際の感情推定難易度を求めるシステムを構築する。システムは複数の判断基準を組み合わせたものにする。例えば否定表現が含まれていれば、そのデータにおける感情推定難易度は高いと考えられる。また、感情表現が含まれていれば、そのデータにおける感情推定難易度は低いと考えられる。したがってシステムによる難易度判定は、否定表現の有無、感情表現の有無、分類器による予測を組み合わせで行うこととする。

2. 関連研究

2.1 節では既存の研究で用いられている感情分類方法を紹介する。2.2 節では否定表現の有無を判定する方法を紹介する。

2.1 感情の分類方法

感情の分類には様々な種類がある。感情表現辞典では感情を表現する単語を喜、怒、哀、怖、恥、好、厭、昂、安、驚の 10 クラスに分類しており、Ptaszynski ら [Ptaszynski 17] による感情表現解析システム ML-Ask はこの分類を採用している。心理学の分野で用いられる感情モデルをもとにするケースもある。長谷川 [長谷川 14] らは Plutchik の感情の輪 [Plutchik 80] における基本感情を用いた。この感情モデルには 8 つの基本感情があり、強さによってそれぞれ 3 段階の感情を持つ。例えば喜び軸なら弱い順に平穏、喜び、恍惚である。基本感情は喜

び、悲しみ、信頼、嫌悪、怒り、恐れ、期待、驚きであり、喜び-悲しみ、信頼-嫌悪、怒り-恐れ、期待-驚きのようにペアとなって互いに反対の感情を持つとされる。この反対の感情があるという特性は、システムとの親和性が高いことなどから感情の分類に採用されやすい。本論文でも Plutchik の感情の輪をもとにした感情分類を採用した。

2.2 日本語文章における否定表現有無の判定

否定表現検出に関する研究の多くは、単語の品詞情報を必要としている。品詞情報を得るためには形態素解析器を用いることが多いが、辞書にない否定表現は検出不可能であり、常に文法的に正しい品詞推定や単語区切りができるとは限らない。我々は日本語文章に否定表現が含まれるかどうかを、品詞情報のみに頼らず判定する方法について検証した [山下 2 18]。その結果、単語の品詞情報と基本形によるマッチング、代表的な否定表現の出現位置、否定極性表現の出現位置を素性として学習したナイーブベイズが有用であるとわかった。

3. 感情表現検出実験

本実験は、感情表現を含むと判断する基準の獲得を目的とする。3.1 節で実験手法、3.2 節で実験結果と考察を述べる。

3.1 実験手法

感情を連想させる表現を感情表現と定義する。本実験では、既存の感情解析システム ML-Ask^{*1} の内部で用いられている 2,100 個の感情表現をベースにする。ML-Ask で用いられている感情表現が存在する全ての感情表現を網羅しているとは限らないため、Word2Vec^{*2} を用いて単語間類似度 (Cosine 類似度) を求め、ML-Ask の感情表現と類似度の高い単語があればそれも感情表現とみなす。以下に判定手順を示す。本実験では、この類似度スコア θ を決定することを考える。

1. ML-Ask の感情表現を感情表現リストとする
2. 評価対象の文章を MeCab^{*3} で単語区切りにし、出現単語の集合を得る

^{*1} <http://arakilab.media.eng.hokudai.ac.jp/~ptaszynski/repository/mlask.htm>

^{*2} 「日本語 Wikipedia エンティティベクトル」http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector

^{*3} システム辞書は mecab-ipadic-NEologd (v0.0.6)

連絡先: 奥村紀之, 大手前大学現代社会学部, 兵庫県伊丹市稲野町 2-2-2, TEL 072-770-6334, FAX 072-770-6916, noriyuki@otemae.ac.jp

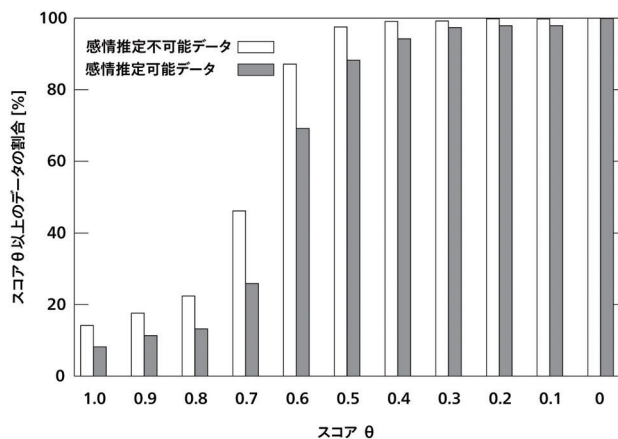


図 1: 感情表現が含まれているデータの割合

- 感情表現リストの単語と出現単語全ての組み合わせについて類似度を求め、最高類似度をその文章のスコアとする
 - スコアが θ 以上の文章は感情表現を含むと判定する
- 被験者 26 人がツイートの感情アノテーションを行い [山下 18], 25 % 以上の人アノテーションに迷ったものを感情推定不可能なデータ、そうでないものを感情推定可能なデータとする。

3.2 実験結果と考察

感情推定不可能データ、感情推定可能データのそれぞれについて、感情表現を含むとみなすデータの割合を図 1 に示す。

感情表現は、感情推定可能なデータよりも感情推定不可能なデータに含まれている割合が高かった。感情推定不可能なデータでスコアが最大 (1.0) になった例としてえーこれは自画自賛しても怒られない、公開待ちなのが惜しいがある。この例では、感情表現怒るを含むため推定難易度が低いとみなせるが、否定表現を含むため推定難易度が高いデータともいえる。

スコアが 0.7 以上の文章には@[ユーザ名] ええじゃろ～、るっぴーかいぎ、試験と丸被りの予感がするし今年は諦めかなあなどがあり、概ね感情表現が含まれていると判断した。実用の際は、感情推定不可能データの多くを事前に何らかの方法で除去してから感情表現の有無を判定することを考える。

4. 分類器による判定実験

本実験は、3 種類の分類器による感情推定難易度判定を目的とする。本章では 4.1 節で実験手法、4.2 節で実験結果を述べ、4.3 節でベースラインとの比較を行う。

4.1 実験手法

分類器を用いない、類似度による難易度判定をベースラインとする。分類器として、Word2Vec を素性とした SVM, CNN, LSTM の 3 種類を作成する。名詞、助詞、動詞、助動詞、記号、形容詞、副詞、感動詞、連体詞、フィラー、接続詞の 11 品詞からいくつかの品詞を除外して素性を作成することで、分類に有用な品詞を実験的に求める。

学習及び評価には、3 章で使用したデータと同じものを用いる。全データ 998 件のうち、感情推定不可能なものは 635 件、感情推定可能なものは 373 件ある。

4.1.1 ベースライン

評価対象データと感情推定難易度の高いデータとの文間類似度を求め、類似度がある値以上であれば、その評価対象データは感情推定難易度が高いとみなす。文間類似度の算出には Word Mover's Distance を用いる。

4.1.2 Word2Vec を素性とした SVM

学習データを MeCab で単語区切りにし、Word2Vec の 200 次元の単語ベクトルを得る。データを構成する単語ベクトルの和を文章ベクトルとし、この 200 次元の文章ベクトルを SVM の素性にする。10 分割交差検定し、評価を行う。

4.1.3 Word2Vec を素性とした CNN

CNN によりテキスト分類を実装した例として、Kim のモデル [Kim 14] が挙げられる。本手法では Kim の提案するモデルを参考に、図 2 のように構成する。埋め込み層では SVM と同様にして単語ごとに 200 次元の文章ベクトルを作成する。畳み込み層では 3×200 サイズのフィルタを 128 個用意し、それぞれで畳み込みを行う。その後プーリング層で 1 フィルタごとにプーリングを行い、128 個のニューロンを出力する。最後に全結合した後 Softmax 関数により各クラスに属する確率を出力する。ドロップアウト率は 0.0~1.0 まで 0.1 ずつ値を変えて試行し、最も良いスコアのものを採用する。

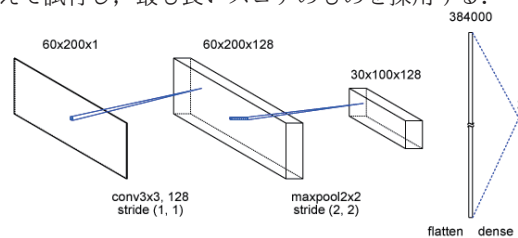


図 2: 本実験で構築する CNN のモデル例

4.1.4 Word2Vec を素性とした LSTM

ネットワークは図 3 のように構築する。埋め込み層では SVM, CNN と同じ手順で文章ベクトルを作成する。次に、隠れ層数 100 の LSTM 層を設ける。全結合層以降の扱いは CNN と同様である。

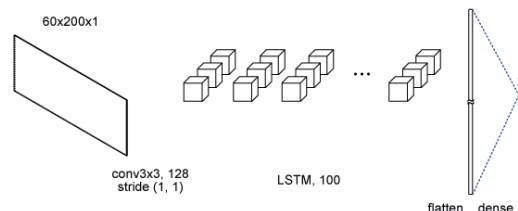


図 3: 本実験で構築する LSTM のモデル例

4.2 実験結果

ベースラインと分類器 3 種類の性能を比較する。システムが難易度低と判定したもののうち、実際に感情推定可能であるものがどれだけ含まれているかを重視するため、性能指標には Precision を重視する F_β 値 ($\beta = 0.4$) を用いる。混同行列は (TP FN FP TN) の並びで表示した。TP は感情推定不可能であり難易度高と予測されたデータ、FN は感情推定不可能であるが難易度低と予測されたデータ、FP は感情推定可能であるが難易度高と予測されたデータ、TN は感情推定可能であり難易度低と予測されたデータである。

4.2.1 ベースライン

最高精度 F_β における類似度と各種評価値を表 1 に示す。ほとんどのデータが感情推定難易度高と予測されている。

4.2.2 Word2Vec を素性とした SVM

SVM において特定の品詞を除外した素性で学習した際の評価値を表 2 に示す。全品詞を用いた素性で学習した結果と比較して、特に助詞、形容詞、副詞の 3 品詞をそれぞれ除外した素性で学習した結果の F_β 値が上がっている。したがって、この 3 品詞は SVM で感情推定難易度判定を行う際の素性と

して有効ではないと考えられる。

助詞を除外した素性で学習した結果、評価データ 1 つあたりの平均文字数は 39.7 文字、中央値は 33.0 文字であった。感情推定不可能なデータで難易度低と判定されたデータについては、平均文字数は 44.2 文字、中央値は 21.0 文字であり、宣伝を意図した長い文章のデータが目立った。また、感情推定可能なデータで難易度低と判定されたデータの 4 割強が画像やサイトの URL を含んでいたが、そうでないデータが URL を含んでいる割合は 1 割未満にとどまった。

4.2.3 Word2Vec を素性とした CNN

CNN において特定の品詞を除外した素性で学習した際の評価値を表 3 に示す。感動詞を除外した素性で学習した結果、感情推定可能なデータで分類器に難易度高と判定されたデータには括弧^{*4}が 1 つも含まれていなかった。また SVM 同様、感情推定不可能なデータで分類器に難易度低と判定されたデータは、他と比べて URL を含んでいることが多かった。

4.3 Word2Vec を素性とした LSTM

LSTM において特定の品詞を除外した素性で学習した際の評価値を表 4 に示す。動詞と助動詞を除外した素性で学習した結果、評価データ 1 つあたりの平均文字数は 40.4 文字、中央値は 32.5 文字であった。感情推定不可能なデータで難易度低と判定されたデータについては、平均文字数は 32.1 文字、中央値は 17.5 文字であり、SVM と同様に中央値が低くなる傾向が見られた。また SVM や CNN と同様に、感情推定不可能なデータで分類器に難易度低と判定されたデータは他と比べて URL を含んでいることが多かった。

助動詞は原作に甲東園出てきてた気がする、くやしいのでお気に入りフォントに変えてやったぜ [URL]、月がわりときれいですねなど、過去、完了、断定などの意味で使われることが多く、それ自身からは感情が想起されにくい。感情推定の材料になりにくい。しかし、否定表現のないなど一部の助動詞を含む文章は感情推定難易度が高いと考えられるため、助動詞全てを除外して良いとも限らない。

表 1: ベースラインの評価

Recall	Precision	F_β	混同行列
0.901	0.519	0.552	(336 37 311 62)

表 2: Word2Vec を素性とした SVM の評価

除外した品詞	Recall	Precision	F_β	混同行列
除外なし	0.545	0.552	0.551	(48 40 39 55)
名詞	0.482	0.562	0.549	(41 44 32 51)
助詞	0.636	0.644	0.643	(56 32 31 63)
動詞	0.568	0.593	0.590	(54 41 37 50)
助動詞	0.620	0.516	0.528	(49 30 46 57)
記号	0.516	0.605	0.591	(49 46 32 54)
形容詞	0.593	0.622	0.618	(51 35 31 65)
副詞	0.470	0.644	0.613	(47 53 26 56)
感動詞	0.535	0.590	0.582	(46 40 32 63)
連体詞	0.609	0.582	0.586	(53 34 38 57)
フィラー	0.525	0.602	0.590	(53 48 35 46)
接続詞	0.556	0.610	0.602	(50 40 32 60)
助詞, 形容詞	0.633	0.538	0.549	(57 33 49 43)
助詞, 副詞	0.550	0.647	0.632	(55 45 30 52)
形容詞, 副詞	0.711	0.602	0.615	(59 24 39 60)

表 3: Word2Vec を素性とした CNN の評価

除外した品詞	Recall	Precision	F_β	混同行列
除外なし	0.500	0.676	0.645	(46 46 22 68)
名詞	0.735	0.581	0.598	(61 22 44 41)
助詞	0.667	0.667	0.667	(66 33 33 50)
動詞	0.535	0.662	0.641	(53 46 27 56)
助動詞	0.740	0.640	0.652	(71 25 40 46)
記号	0.60	0.698	0.682	(60 40 26 55)
形容詞	0.772	0.617	0.635	(71 21 44 46)
副詞	0.710	0.660	0.666	(66 27 34 55)
感動詞	0.787	0.914	0.894	(74 20 7 80)
連体詞	0.729	0.642	0.653	(70 26 39 47)
フィラー	0.605	0.571	0.576	(52 34 39 57)
接続詞	0.823	0.612	0.635	(79 17 50 36)

表 4: Word2Vec を素性とした LSTM の評価

除外した品詞	Recall	Precision	F_β	混同行列
除外なし	0.788	0.670	0.684	(67 18 33 64)
名詞	0.642	0.612	0.616	(52 29 33 54)
助詞	0.753	0.615	0.631	(64 21 40 57)
動詞	0.609	0.709	0.693	(56 36 23 67)
助動詞	0.545	0.714	0.685	(55 46 22 59)
記号	0.681	0.627	0.634	(64 30 38 49)
形容詞	0.698	0.615	0.625	(67 29 42 44)
副詞	0.529	0.672	0.648	(45 40 22 75)
感動詞	0.591	0.658	0.648	(52 36 27 66)
連体詞	0.653	0.711	0.702	(64 34 26 58)
フィラー	0.625	0.640	0.637	(55 33 31 63)
接続詞	0.624	0.624	0.624	(53 32 32 65)
動詞, 助動詞	0.717	0.724	0.723	(71 28 17 56)

4.3.1 各手法の比較

4.3.1 項～4.3.4 項で示した結果を抜粋し、表 5 で比較する。 F_β 値より、分類器による感情推定難易度の判定には感動詞を除外した素性で学習した CNN を用いる。

ベースラインである単語間類似度を用いた手法では、ほとんどのデータを難易度高と判定している。単語間類似度をもとに感情推定難易度を判定した場合は難易度高に判定されやすいと予想される。一方で単語分散表現をもとに判定した場合は、感情推定不可能なデータも可能なデータも 6 割以上が正しく判定されており、判定に偏りはないといえる。特に CNN では、感情推定可能なデータを難易度低と判断できた割合は 9 割強にものぼる。

表 5: ベースラインと各手法のスコア比較

手法 (除外した品詞)	Recall	Precision	F_β
ベースライン	0.901	0.519	0.552
SVM (助詞)	0.636	0.644	0.643
CNN (感動詞)	0.787	0.914	0.894
LSTM (動詞, 助動詞)	0.717	0.724	0.723

5. 感情推定難易度判断システムの構築

3 章では感情表現の有無、4 章では分類器によって感情推定難易度の判断が可能になった。本章では、これらの判断手法を組み合わせる感情推定難易度判定システムを構築する。5.1 節でシステムの構成、5.2 節でシステムの評価について述べる。

*4 (), (), { }, 「 」, [] など

表 6: システムの評価

組み合わせ	Accuracy	Recall	Precision	F_1	F_β	混同行列
否定表現	0.268	0.241	0.951	0.384	0.676	(58 183 3 10)
感情表現	0.575	0.593	0.935	0.726	0.866	(143 98 10 3)
分類器	0.646	0.660	0.952	0.779	0.897	(159 82 8 5)
否定表現 + 感情表現	0.689	0.718	0.940	0.814	0.902	(173 68 11 2)
否定表現 + 分類器	0.740	0.768	0.949	0.849	0.919	(185 56 10 3)
感情表現 + 否定表現	0.154	0.116	0.933	0.207	0.474	(28 213 2 11)
感情表現 + 分類器	0.413	0.411	0.934	0.571	0.794	(99 142 7 6)
分類器 + 否定表現	0.173	0.133	0.970	0.234	0.519	(32 209 1 12)
分類器 + 感情表現	0.413	0.411	0.934	0.571	0.794	(99 142 7 6)
否定表現 + 感情表現 + 分類器	0.579	0.598	0.935	0.729	0.867	(144 97 10 3)
否定表現 + 分類器 + 感情表現	0.579	0.598	0.935	0.729	0.867	(144 97 10 3)
感情表現 + 否定表現 + 分類器	0.465	0.473	0.927	0.626	0.819	(114 127 9 4)
感情表現 + 分類器 + 否定表現	0.102	0.054	1.000	0.102	0.292	(13 228 0 13)
分類器 + 否定表現 + 感情表現	0.484	0.490	0.937	0.643	0.832	(118 123 8 5)
分類器 + 感情表現 + 否定表現	0.102	0.054	1.000	0.102	0.292	(13 228 0 13)

5.1 システムの構成

本システムは、日本語文章を受け取り、感情推定難易度として低か高のどちらかを返すものとする。システム内部では、否定表現の有無、感情表現の有無、分類器による予測を組み合わせて感情推定難易度を判断する。否定表現を含むかどうかの判定には、で提案したナイーブベイズを用いる [山下 2 18]。否定表現を含む文章は感情推定難易度が高いと考えられるため、ナイーブベイズによって否定表現を含むと判定された文章の難易度は高になる。感情表現を含むかどうかは、3 章で提案したように、感情表現リストとの単語間類似度が 0.7 以上であるか否かで判定する。感情表現を含む文章は感情推定難易度が低いと考えられるため、ここで感情表現を含むと判定された文章の感情推定難易度は低になる。分類器による予測では、4 章で提案した分類器を用いて感情推定難易度を判定する。

5.2 システムの評価

評価には、書き手（筆者）のことを知っている人物 8 人が筆者のツイート 254 件を感情アノテーションしたものを用いる。このデータは、各実験で用いた学習及び評価データには含まれていない。3 章で使用したデータ同様、感情推定不可能なデータと感情推定可能なデータに分ける。

評価結果を表 6 に示す。否定表現の有無と分類器の 2 段階で判定したものが最も良いスコアであった。感情推定難易度が高いデータの 7 割は正しい判断ができていない。感情表現を判定に組み込んだもので、実際には難易度が高いにもかかわらずシステムでは低いと判定されているデータ (FN) の特徴の一つに～したいを含むというものがある。これは書き手の希望を表す表現だが、通常は実際にそれを行うことでどのような感情を得られるのかが書かれないため、感情表現が正しく検出されにくい。実際には難易度が低いにもかかわらずシステムでは高いと判定されているデータ (FP) には、感情表現として認められなかった単語を含むものがあつた。感情表現として認められなかった単語には神、たのしいなどがある。例えば神は喜びや信頼を表現する単語として用いられている。本研究で構築した感情表現検出システムにおける神と類似度の高い感情表現には恩寵、慈悲などがあるが、類似度が閾値 0.7 を下回っているため感情表現とは認められなかった。また、平仮名表記たのしいは漢字表記の楽しいと意味が変わらないにもかかわらず、この 2 単語の類似度は 0.51 となっており、直感に反する結果が得られた。

6. おわりに

本テーマに先行研究は存在しないため精度の比較はできないが、感情推定難易度が高いデータは 7 割が正しく判断されている。一方で感情推定難易度が低いデータは 8 割が誤判定されており、感情表現を含むが検出に失敗した文、感情が含まれていない文は判定が難しいと分かった。

今後は本実験をベースラインとし、誤判定した文章の特徴やアノテーション時間をシステムに組み込むことで精度向上をはかる。また筆者の書いた文章だけでなく、他の人によって書かれた文章も学習データに用いることで汎用性向上を目指す。

謝辞

本研究は JSPS 科研費 18K11455 の助成を受けている。

参考文献

- [松林 16] 松林圭, 五味京祐, 古川和祈, 松尾祐佳, 松原良和, 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 回全国大会講演論文集 (1), pp.79-80, 2016.
- [山下 1 18] 山下紗苗, 上泰, 加藤恵梨, 酒井健, 奥村紀之, 人手による感情ラベル付けにおける応答時間に着目した感情推定難易度の評価, 第 13 回テキストアナリティクス・シンポジウム
- [Ptaszynski 17] Ptaszynski, M., Dybala, P., Rzepka, R., Araki, K., & Masui, F.: ML-Ask: Open Source Affect Analysis Software for Textual Input in Japanese. Journal of Open Research Software, 5(1), 2017.
- [長谷川 14] 長谷川貴之, 鍛冶伸裕, 吉永直樹, 豊田正史, オンライン上の対話における聞き手の感情の予測と喚起, 人工知能学会論文誌 29 (1), pp.90-99, 2014.
- [Plutchik 80] Plutchik, R, A General Psychoevolutionary Theory of Emotion, in Emotion: Theory, research, and experience: Vol. 1. Theories of emotion, pp.333, New York: Academic, 1980.
- [山下 2 18] 山下紗苗, 上泰, 奥村紀之, 品詞情報とルールベースによる否定表現有無の判定, 言語処理学会第 25 回年次大会発表論文集 (発表予定のためページ数未定)
- [Kim 14] KIM, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882, 2014.