

# CRF を用いたレビューにおける節単位毎の感情推定

秋山 和寛<sup>†</sup> 三澤 賢祐<sup>††</sup> 成田 和弥<sup>††</sup> 熊本 忠彦<sup>†††</sup> 灘本 明代<sup>††††</sup>

<sup>†</sup> 甲南大学大学院 自然科学研究科 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

<sup>††</sup> 株式会社 Insight Tech 〒163-1333 東京都新宿区西新宿 6-5-1 新宿アイランドタワー

<sup>†††</sup> 千葉工業大学 情報科学部 〒275-0016 千葉県習志野市津田沼 2-17-1

<sup>††††</sup> 甲南大学 知能情報学部 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: <sup>†</sup>m1724002@s.konan-u.ac.jp, <sup>††</sup>{kensuke\_mitsuzawa,kazuya\_narita}@insight-tech.co.jp,

<sup>†††</sup>kumamoto@net.it-chiba.ac.jp, <sup>††††</sup>namamoto@konan-u.ac.jp

あらまし 企業において顧客の意見は重要な価値を持つが、日々寄せられる顧客の意見には重要なものもあればそうでないものもある。企業はそれら玉石混濁なデータの分析を行い、その中から対応優先度の高いものを抽出し顧客の要望を知るとともに企業運営に活かしている。しかしながら企業に寄せられる顧客の意見は膨大であり、また一つ一つの意見の中にも重要な部分とそうでない部分が混在していることがある。このことから投稿された意見に対し自動で感情の極性分類を実施できれば、情報抽出に大きく寄与できると考える。本論文ではそのはじめの一步として、投稿された顧客の意見を節単位に分割し、CRF を用いて各節の極性を推定する。

キーワード 感情推定, レビュー, CRF

## 1. はじめに

近年インターネット上には様々なレビューサイトが普及している。それに伴い、人々は購入した商品や受けたサービスについての感想や意見を自由に投稿している。また、メールフォームなどを用いた企業への直接の意見投稿も依然として行われており、それらの数は膨大である。これら投稿された意見や感想にはポジティブなものだけでなく、不満や改善要求などが書かれたネガティブなものもある。このネガティブな情報に対して素早く対応することができれば、問題を初期のうちに解決できる可能性が高くなる。商品やサービスの提供元である企業（以下、まとめて“サービス提供企業”と呼ぶ）はそれら玉石混濁なデータを用いて、提供している商品やサービスに問題が発生していないかの確認や発生している問題の原因の分析を行っている。また優先的に取り組むべき課題の評価などを行い商品やサービスの提供に活かしている。例えば Amazon<sup>(注1)</sup>、楽天<sup>(注2)</sup>のようなレビューサイトには、日々利用者からの商品やサービスの肯定的、否定的な意見が寄せられている。また、不満買取センター<sup>(注3)</sup>というサービスでは、幅広いトピックに対するユーザの不滿意見が寄せられている。いずれのサービスにおいても、投稿された情報を分析すれば商品やサービスの改善に繋がられる可能性がある。

しかしながら投稿される顧客の意見や感想は膨大であり、サービス提供企業にとって重要なものとそうでないものがあるだけでなく、一つ一つの投稿にも重要な部分とそうでない部分が混在していることがあり、これらを人手で分類することは困難で

ある。

このことから我々は投稿文に対し自動で感情推定を行うことが可能となれば、情報抽出の手間を減らすことができると考える。また、感情推定を行う範囲に関しては、情報抽出の手間を減らすという点から、それ単体で意味を持ち、尚且つできるだけ簡潔な部分であることが望ましいと考えられる。そのため、文の構造において語や句の集合によって1つの意味を形作っている節の部分を対象とする。このように節単位で感情の推定を行うことができれば、投稿文の文長が長くとも注目すべき箇所を素早く見つけることが可能となる。これにより、少ない手間で情報を取得できるユーザインターフェースを提供することが可能となる。

しかしながら、投稿文の記述内容はサイトドメイン毎に、また、商品やサービス毎に特徴的に書かれている。その為に、1つの感情推定手法で様々なサイトドメインにおける様々な商品やサービスの感情分析が可能かどうかは不明である。

本論文では、様々なサイトドメインへの投稿文からの情報抽出の際の負担軽減を目的として、投稿文を節単位に分割し学習器を用いて各節の感情の推定を行う。そして、この感情推定の学習器が複数のサイトドメインや複数の商品やサービスに有効であるかどうかを検証する。具体的には、各節の感情推定には構造学習により系列ラベリングを解く手法である条件付き確率場 (CRF) [1] を用い、感情推定を行う節だけでなくその周辺の節との関係性も考慮する。さらに、CRF を用いた感情分析が複数のサイトドメインに有効であるかどうかの評価実験として、3種類の実験を行い、どのような学習器の汎用性が高いかの分析を行う。

以下、2. 章では関連研究について述べる。3. 章では CRF による感情推定について述べ、4. 章では実験により感情推定の精度及び汎用性を測る。最後に 5. 章では本論文のまとめと今後

(注1) : <https://www.amazon.co.jp/>

(注2) : <https://www.rakuten.co.jp/>

(注3) : <http://fumankaitori.com/>

の課題について述べる。

## 2. 関連研究

“Amazon”や“TripAdvisor”<sup>(注4)</sup>といったレビューサイトの普及に伴い、レビューから有益な情報を抽出するためにレビューの分析に関する研究が数多く行われている。以下でそれらについて説明する。

Rustamov ら [2] は主観的な部分には意見が書かれているという仮定のもと、2つの機械学習によるアプローチから文レベルでの主観部分抽出を行っている。Paltoglou ら [3] は主観性及び極性の2つの値で各文を特徴付け、機械学習による分類を行っている。田村ら [4] はユーザ及び商品のトピックに加え、その商品に付けられた評点を加えたトピックモデルの構築を提案している。これらの研究では投稿文全体や投稿文中の各文を対象としているのに対し、本研究では節を対象とした感情推定を行っている点が異なる。

また、レビューを分析する上で本文中の内容に対する投稿者の感情分析を行うことはとても重要である。鈴木ら [5] はラベル付きデータに加えラベルなしデータを用いた学習を行うことで単体で感情を持つ語のみならず、形容詞のような前後の文脈を考慮しなければ感情がわからない語にも対応した評価表現抽出手法を提案している。Lin ら [6] はレビュー本文の背景にある感情を仮定しトピックモデルを構築している。Choi ら [7] は単語間の相互作用による極性の生成に着目し、否定語が及ぼす影響を考慮した表現レベルでの感情分析を行っている。これらに対し、本研究ではCRFを用いて周辺の節の情報も考慮して対象となる節の感情分析を行うという点が異なる。

## 3. CRF を用いた感情の推定

本研究では、投稿された投稿文から情報を抽出する際の負担軽減を目的として、投稿文を節単位に分割し、各節に対して感情推定を行う。感情推定を行う際には、投稿文中における文法や文脈などの節同士の関係性も考慮する必要があると考えられる。そのため、構造学習により系列ラベリングを解く手法であるCRFを用いて各節の感情を推定する。学習にはCRFの実装であるCRFsuite [8]を用いる。

### 3.1 投稿文の節単位への分割

投稿文中の各節を対象として感情推定を行う前処理として投稿文を節単位に分割する。節への分割には黒橋らの作成した日本語構文・格・照応解析システム (KNP) [9] を使用する。具体的には、まず投稿文に対して形態素解析エンジン Juman<sup>(注5)</sup>を用いて形態素解析し、出力結果に対し knp を用いて構文解析を行う。解析結果から投稿文内の節情報を取得し、それに基づき投稿文を分割する。尚、ある節の長さが5文字以下である場合には次の節と併合する。また、最後の節の長さが5文字以下である場合には1つ前の節と併合する。

例えば、「ゴム手袋の中に水が入ると気持ち悪いし、なかなか

乾かない。だけどひっくり返すのが大変」という投稿文の場合、「ゴム手袋の中に水が入ると」「気持ち悪いし」「なかなか乾かない」「だけどひっくり返すのが大変」という節に区切り、これらの節を用いる。

### 3.2 CRF を用いた感情の推定

#### 使用データ

CRFの学習に用いる教師データ及び評価実験に用いる評価用データには我々の行ったアンケートにより節毎に感情を付与された投稿文を用いる。アンケートにはクラウドソーシングを用いる。またアンケートに用いるデータとして、不満買取センターへの投稿文から1500件、及びAmazon内でレビュー内容の傾向が違ふと考えられる日用品、家電機器の各カテゴリから3000件を用いる。不満買取センターはジャンルを問わず投稿された、利用者が日々感じた不満に関する投稿の中からランダムに取得したものを使用する。Amazonは「トイレトペーパー」、「洗剤」、「防虫剤」、「ノートPC」、「ロボット掃除機」、「スマートフォン」の商品レビューよりそれぞれランダムに500件ずつ取得したものを使用する。具体的には、KNPにより分割された各投稿文の節合計33587件に対し10人の作業員による感情評価を行った。作業員は各節を見て、ポジティブ、ネガティブ、ニュートラルのいずれに当てはまるかを評価した。10人の作業員のうち6人以上の評価がいずれかの感情軸に集中した場合、その感情を該当の節の感情とする。また、いずれの感情軸への評価も過半数に達しなかった場合は明確な感情を持たないとして該当の節の感情はニュートラルとする。アンケートによる感情ラベル付与の例を表1に、ラベル付与の結果及び商品ごとのアンケート作業員のカッパ係数を表2に示す。尚、ラベル付与結果における不満買取センターのポジティブ節数が不満投稿全体の2%と極めて低い値になっているが、これは不満買取センターがユーザの様々な不満に関する投稿を受け付けるサービスとして運営されているためである。

#### 素性

素性には対象節中に含まれる各語の原形、品詞、極性値及び極性反転語の数を対象とする。尚、ここでの極性反転語とは「しかし」や「だが」といった、後の節の感情である極性に影響を与える語のことを指す。本研究では国立国語研究所<sup>(注6)</sup>より公開されている分類語彙表増補改訂版データベース [10] の反対の意味を持つ接続詞、接続助詞39語を対象とする。また、語の極性値の導出には高村ら [11] が作成した単語感情極性対応表を用いる。具体的には、ある語の極性値  $x$  を単語感情極性対応表の値  $n$  に応じて以下のように決定する。

$$x = \begin{cases} 2 & (n > 0.5) \\ 1 & (0.5 > n > 0) \\ -1 & (0 > n > -0.5) \\ -2 & (n < -0.5) \end{cases}$$

(注4) : <https://www.tripadvisor.jp/>

(注5) : <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN>

(注6) : <https://www.ninjal.ac.jp/>

表 1 節毎のラベルの付与例

投稿文「この掃除機は性能が良いのだが、音がうるさいので、音が静かなモーターにしてほしい」

節	ラベル
この掃除機は性能が良いのだが、	ポジティブ
音がうるさいので、	ネガティブ
音が静かなモーターにしてほしい	ニュートラル

表 2 ラベル付与結果

データ			ポジティブ節数	ネガティブ節数	ニュートラル節数	カッパ係数
サイトドメイン	カテゴリ	商品				
不満買取センター	不満	不満投稿	69 (2%)	2257 (50%)	2152 (48%)	0.435
Amazon	日用品	トイレトペーパー	955 (31%)	442 (14%)	1721 (55%)	0.473
Amazon	日用品	洗剤	1234 (28%)	700 (16%)	2492 (56%)	0.504
Amazon	日用品	防虫剤	777 (24%)	370 (12%)	2492 (64%)	0.460
Amazon	家電機器	ノート PC	1054 (21%)	858 (17%)	3187 (62%)	0.464
Amazon	家電機器	ロボット掃除機	1922 (25%)	1051 (14%)	4628 (61%)	0.479
Amazon	家電機器	スマートフォン	1118 (20%)	1086 (19%)	3505 (61%)	0.513

#### 4. 同一分類器による複数サイトドメインの評価実験

CRF による感情推定が、本研究の目的の 1 つであるサイトドメインの違いを超えた分類器の有効性を示すかどうかを検証するための評価実験を行った。

実験は以下の 3 つの評価毎に行った。

##### 実験 1 同一サイトドメインにおける学習

ある 1 つのサイトドメインの投稿文のみを教師データとし、同じサイトドメインの投稿文を分類する。

##### 実験 2 異なるサイトドメインによる学習

ある 1 つのサイトドメインの投稿文のみを教師データとし、違うサイトドメインの投稿文を分類する。

##### 実験 3 複数サイトドメインの混合による学習

複数のサイトドメインの投稿文を混ぜたデータを教師データとし、他のサイトドメインの投稿文を分類する。

それぞれの実験に対し評価用データにラベル付けされた感情を正解として、CRF により分類されたポジティブ、ネガティブ、ニュートラルそれぞれの適合率、再現率、F 値を求めた。

#### 4.1 実験 1：同一サイトドメインによる学習

##### 実験 1：実験条件

同一サイトドメインの投稿文の感情を正しく推定できるかどうかの確認を目的とし、ある 1 つのサイトドメインの投稿文のみで学習を行い、同じサイトドメインの他の投稿文の感情分類を行った。具体的には以下の 2 つの実験を行った。

##### ● 実験 1-1

あるサイトドメインから、商品に関係なくランダムに抽出した

投稿文のみを教師データとし、同じサイトドメインの他の投稿文を分類する。

具体的には、不満買取センターからランダムに抽出した投稿文 1500 件中 1200 件を学習に用い、残りの 300 件の投稿文を評価用データとして用いた。

##### ● 実験 1-2

あるサイトドメインのある商品の投稿文のみを教師データとし、同一サイトドメインの他の商品の投稿文を分類する。

具体的には以下の 4 種類のデータを用いた。

##### － データ 1

Amazon の日用品カテゴリの投稿文 1200 件を学習に用い、日用品カテゴリの投稿文 300 件を評価用データとして用いた。

##### － データ 2

Amazon の家電機器カテゴリの投稿文 1200 件を学習に用い、家電機器カテゴリの投稿文 300 件を評価用データとして用いた。

##### － データ 3

Amazon の日用品カテゴリの投稿文 1200 件を学習に用い、家電機器カテゴリの投稿文 300 件を評価用データとして用いた。

##### － データ 4

Amazon の家電機器カテゴリの投稿文 1200 件を学習に用い、日用品カテゴリの投稿文 300 件を評価用データとして用いた。

#### 実験 1：結果と考察

実験 1 の結果を表 3 に示す。実験 1-1 ではネガティブ、ニュー

トラルの結果が高めの値を示しているのに対し、ポジティブは適合率こそ高いものの再現率、F 値は低いものとなっている。これは表 2 の感情毎の節数の割合からも読み取れるように、不満買取センターへ投稿される投稿文にはポジティブな内容の部分が少ないため、学習用データ評価データともにポジティブな節の数が少なく、分類がうまくできなかったと考えられる。

また、実験 1-2 においてはデータ 1, 2 はどれも高い値を示しているのに対し、データ 3, 4 はネガティブの値が低いものとなっている。

これは同じ Amazon の投稿文であっても日用品では「臭い」や「汚れ」、家電機器では「エラー」や「バッテリー (の容量などに対する不満)」というようにカテゴリや商品によって投稿文中でネガティブな意味として使用される語に違いが見られるため、商品が異なるとこのような評価の語が異なるので、うまく学習できなかったと考える。

## 4.2 実験 2: 異なるサイトドメインによる学習

### 実験 2: 実験条件

異なるサイトドメイン間での投稿文の特徴の違いが投稿文の感情推定にもたらす影響を測ることを目的とし実験を行った。本実験はある 1 つのサイトドメインの投稿文のみを教師データとし、異なるサイトドメインの投稿文の感情推定を行った。具体的には以下の 2 つの実験を行った。

#### ● 実験 2-1

不満買取センターの投稿文 1200 件を学習に用い、以下の 2 種類のデータを評価データとして用いた。

##### – 実験 2-1: データ 1

不満買取センターの投稿文 1200 件を学習に用い、Amazon の日用品カテゴリの投稿文 300 件を評価用データとして用いた。

##### – 実験 2-1: データ 2

不満買取センターの投稿文 1200 件を学習に用い、Amazon の家電機器カテゴリの投稿文 300 件を評価用データとして用いた。

#### ● 実験 2-2

以下の 2 種類のデータを学習データとして用い、不満買取センターの投稿文 300 件を評価に用いた。

##### – 実験 2-2: データ 1

Amazon の日用品カテゴリの投稿文 1200 件を学習に用い、不満買取センターの投稿文 300 件を評価用データとして用いた。

##### – 実験 2-2: データ 2

Amazon の家電機器カテゴリの投稿文 1200 件を学習に用い、不満買取センターの投稿文 300 件を評価用データとして用いた。

### 実験 2: 結果と考察

実験 2 の結果を表 4 に示すとともに結果に対する考察を述べる。

実験 2-1 ではどちらのデータでもポジティブは適合率が高く再現率が低いという結果になっている。これは不満買取センターでのポジティブな表現には「良い」や「嬉しい」、「好き」などが多く、これらの単語が含まれている評価対象の Amazon

の節に対してはうまく分類することが出来た。しかしながらこれら以外にも「優しい」や「おすすめ」、「安心」などといったポジティブな表現も Amazon 内には多く存在しており、それらに対応ができなかった。このように両方のサイトに共通して出現している感情を示す語に対しては感情推定ができており適合率が高くなった。しかしながら、一方のサイトのみに多く出現している感情を示す語については感情推定ができておらず、再現率が低い結果となった。また、ネガティブの値も低い。これは不満買取センターのネガティブな投稿文は「悪い」や「嫌だ」というように直接的な表現で書かれていることが多い。それに対し、Amazon では「台所洗剤ってすぐなくなる」や「AC アダプタがでかい」のように間接的であったり、文脈から読み取るようなネガティブな表現も多く存在している。そのため、このようなサイトドメイン間での投稿内容の違いに対してうまく感情推定をすることができなかったためだと考えられる。

実験 2-2 ではどちらのデータでもポジティブの値が低い、これは評価対象である不満買取センターにおいてポジティブな節の数が少なかったためと考えられる。また、ネガティブの値も低いものとなっているが、これは「もっと楽に〜したい」や「〜だったら良いのに」の「楽」や「良い」といったようなネガティブ節中のポジティブな意味を持つ語に反応して、ポジティブと判断されたためだと考えられる。これらの結果から、1 つのサイトドメインのデータによる学習では異なるサイトの感情推定があまりできないことが判明した。これにより、教師データを複数のサイトにするか、もしくは感情推定の手法を改良するかが必要であることが推測できる。

## 4.3 実験 3: 複数サイトドメイン混合による学習

### 実験 3: 実験条件

これまでの実験により、1 つのサイトドメインの投稿文のみを教師データとして学習を行った場合、その投稿文の特徴によっては感情推定がうまくいかないことが判明した。そこで、感情の割合や書かれ方の違う投稿文を併せた学習による感情推定の汎用性への影響を調べることを目的とする実験を行った。本実験では複数のサイトドメインのデータを混ぜて学習し、他のサイトドメインの投稿文の感情推定を行った。具体的には以下の 2 つの実験を行った。

#### ● 実験 3-1

以下の 2 種類のデータを用いて評価を行った。

##### – 実験 3-1: データ 1

不満買取センターと Amazon 日用品カテゴリの投稿文各 600 件を教師データとし、Amazon の家電機器カテゴリの投稿文 300 件を評価用データとして用いた。

##### – 実験 3-1: データ 2

不満買取センターと Amazon の家電機器カテゴリの投稿文各 600 件を教師データとし、Amazon の日用品カテゴリの投稿文 300 件を評価用データとして用いた。

#### ● 実験 3-2

以下の 2 種類のデータを用いて評価を行った。

##### – 実験 3-2: データ 1

不満買取センターと Amazon の日用品カテゴリの投稿文

表 3 実験 1: 同一サイトドメインによる学習結果

評価実験	使用データ	ラベル	適合率		再現率		F 値
			精度	データ数	精度	データ数	
実験 1-1	学習データ: 不満買取センター 評価データ: 不満買取センター	ポジティブ	0.750	6/8	0.400	6/15	0.522
		ネガティブ	0.877	392/447	0.879	392/446	0.878
		ニュートラル	0.869	393/452	0.881	393/446	0.875
実験 1-2 (データ 1)	学習データ: Amazon(日用品) 評価データ: Amazon(日用品)	ポジティブ	0.893	634/710	0.868	634/730	0.881
		ネガティブ	0.905	296/327	0.706	296/419	0.794
		ニュートラル	0.879	1363/1551	0.947	1363/1439	0.912
実験 1-2 (データ 2)	学習データ: Amazon(家電機器) 評価データ: Amazon(家電機器)	ポジティブ	0.887	583/657	0.878	583/664	0.883
		ネガティブ	0.883	386/437	0.737	386/437	0.803
		ニュートラル	0.908	1926/2121	0.950	1926/2027	0.929
実験 1-2 (データ 3)	学習データ: Amazon(日用品) 評価データ: Amazon(家電機器)	ポジティブ	0.661	362/548	0.583	362/621	0.619
		ネガティブ	0.531	113/213	0.208	113/543	0.299
		ニュートラル	0.721	1645/2282	0.875	1645/1879	0.791
実験 1-2 (データ 4)	学習データ: Amazon(家電機器) 評価データ: Amazon(日用品)	ポジティブ	0.802	422/526	0.555	422/760	0.656
		ネガティブ	0.516	144/279	0.383	144/376	0.440
		ニュートラル	0.723	1279/1768	0.890	1279/1437	0.798

表 4 実験 2: 異なるサイトドメインによる学習結果

評価実験	使用データ	ラベル	適合率		再現率		F 値
			精度	データ数	精度	データ数	
実験 2-1 (データ 1)	学習データ: 不満買取センター 評価データ: Amazon(日用品)	ポジティブ	0.882	15/17	0.021	15/720	0.041
		ネガティブ	0.266	171/642	0.424	171/403	0.327
		ニュートラル	0.629	1184/1883	0.834	1184/1419	0.717
実験 2-1 (データ 2)	学習データ: 不満買取センター 評価データ: Amazon(家電機器)	ポジティブ	0.800	8/10	0.013	8/612	0.026
		ネガティブ	0.317	255/805	0.468	255/545	0.378
		ニュートラル	0.687	1545/2250	0.810	1545/1908	0.743
実験 2-2 (データ 1)	学習データ: Amazon(日用品) 評価データ: 不満買取センター	ポジティブ	0.110	10/91	0.500	10/20	0.180
		ネガティブ	0.825	99/120	0.204	99/485	0.327
		ニュートラル	0.551	426/773	0.889	426/479	0.681
実験 2-2 (データ 2)	学習データ: Amazon(家電機器) 評価データ: 不満買取センター	ポジティブ	0.136	6/44	0.429	6/14	0.207
		ネガティブ	0.811	185/228	0.377	185/491	0.515
		ニュートラル	0.586	414/707	0.873	414/474	0.701

各 600 件を教師データとし、不満買取センターの投稿文 300 件を評価用データとして用いた。

#### 実験 3-2: データ 2

不満買取センターと Amazon の家電機器カテゴリの投稿文各 600 件を教師データとし、不満買取センターの投稿文 300 件を評価用データとして用いた。

#### 実験 3: 結果と考察

実験 3 の結果を表 5 に示すとともに結果に対する考察を述べる。

不満買取センターのデータのみで学習を行った実験 2-1 の結果と比較して、実験 3-1 ではポジティブの値が上がっている。実際に実験で推定された感情を確認してみても、実験 2-1 で正しくポジティブだと推定することが出来ていなかった「大変満足しています」や「おすすめの一品です」といったデータを正しくポジティブと推定することが出来た。表 6 に実験 2-1 と比較して正しくポジティブだと推定することができるようになった例を示す。これはポジティブな表現が豊富な Amazon のデータを加えて学習することで、不満買取センターのデータに不足していたポジティブな表現を補うことが出来たためだと考えら

れる。

一方で、Amazon のデータのみで学習を行った実験 1-2 のデータ 3, 4 の結果と比較すると、ポジティブ、ネガティブともに値はほぼ変化していない。これは不満買取センターにおけるポジティブなデータが少ないため、評価用データでもポジティブのデータ数が少なかったということが挙げられる。またそれに加えて実験 2 の考察でも述べているように、不満買取センターのポジティブなデータによく現れるのは「好き」や「嬉しい」などの表現であるが、これらは Amazon のデータ内にも既に多く存在している。このように学習データに新たに加えた不満買取センターのデータにポジティブな感情を示す新たな語が少なかったことから、不満買取センターのデータを学習データに加えても学習結果が大きくは変わらず、そのため値の変化も少なかったと考えられる。またネガティブに関しては実験 2 より、不満買取センターと Amazon のネガティブな表現には違いがある。このことからネガティブの表現が異なっている不満買取センターのデータを学習に加えても Amazon のネガティブなデータの感情推定には寄与しなかったと考えられる。

表 5 実験 3：複数サイトドメイン混合による学習結果

評価実験	使用データ	ラベル	適合率		再現率		F 値
			精度	データ数	精度	データ数	
実験 3-1 (データ 1)	学習データ：不満買取センター, Amazon(日用品) 評価データ：Amazon(家電機器)	ポジティブ	0.661	349/528	0.497	349/702	0.567
		ネガティブ	0.492	150/305	0.281	150/534	0.358
		ニュートラル	0.736	1862/2530	0.875	1862/2127	0.800
実験 3-1 (データ 2)	学習データ：不満買取センター, Amazon(家電機器) 評価データ：Amazon(日用品)	ポジティブ	0.780	322/413	0.435	322/740	0.559
		ネガティブ	0.489	198/405	0.466	198/425	0.477
		ニュートラル	0.702	1295/1845	0.864	1295/1498	0.775
実験 3-2 (データ 1)	学習データ：不満買取センター, Amazon(日用品) 評価データ：不満買取センター	ポジティブ	0.462	6/13	0.545	6/11	0.500
		ネガティブ	0.869	370/426	0.799	370/463	0.832
		ニュートラル	0.795	372/468	0.859	372/433	0.826
実験 3-2 (データ 2)	学習データ：不満買取センター, Amazon(家電機器) 評価データ：不満買取センター	ポジティブ	0.476	10/21	0.476	10/21	0.476
		ネガティブ	0.855	342/400	0.729	342/469	0.787
		ニュートラル	0.754	365/478	0.864	365/452	0.805

実験 3-2 においても、不満買取センターのデータのみの学習で感情推定を行った実験 1-1 に比べてポジティブの再現率が上がっている。表 7 に実験 1-1 と比較して正しくポジティブだと推定することができるようになった例を示す。これは実験 3-1 と同様に Amazon のデータを加えて学習を行うことで、不満買取センター単体の学習では対応できなかったポジティブな表現に対応できるようになったためだと考えられる。しかしながら、一方で適合率は下がってしまっている。これは実験 1-1 ではポジティブと推定されたデータ数は 8 件だったのに対して、実験 3-2 のデータ 1 では 13 件、データ 2 では 21 件のデータがポジティブであると推定されているように、ポジティブな学習データが増えた結果、誤判断も増えてしまったためだと考えられる。

また、Amazon のデータのみで学習を行った実験 2-2 と比較するとポジティブ、ネガティブの値がともに上がっている。ポジティブに関しては適合率が大きく上がっているが、これは評価データである不満投稿データのポジティブな表現は限られている。そのため同様の傾向を持つデータを学習に加えたことにより、実験 2 のデータ 1, 2 ではそれぞれ 91 件、44 件だったポジティブと判断されたデータが実験 3 ではデータ 1 が 13 件、データ 2 では 21 件まで減少しているように、誤判断の数を減らすことができたためと考えられる。ネガティブに関しては再現率が上がっているが、これは実験 2 の考察でも述べているように、不満買取センターと Amazon のネガティブな表現には違いがある。このことから Amazon のデータのみの学習ではうまく感情推定ができなかった不満買取センターのネガティブな表現を、学習データに不満買取センターのデータを加えることで推定できるようになったためと考えられる。

#### 4.4 総 評

以上の結果をまとめると、実験 1 の各結果より、同一のサイトドメインの投稿文での感情推定はある程度の精度を出すことはできるが、投稿文の傾向により推定する感情に得手不得手が生じてしまうことがあるといえる。また実験 2 では、同様に 1 つのサイトドメインのデータを用いた学習のみでは投稿文の特徴が違う他のサイトドメインの投稿文を感情推定するのが難しいということが判明した。これら実験 1, 実験 2 の結果より、

1 つのサイトドメインのみを学習に用いては感情推定の汎用性を得られないということがわかる。また実験 3 の結果より、複数のサイトドメインの投稿文を併せて学習することで不足していた感情表現の情報を補うことができ、ある程度の精度を得ることが出来た。このことから、複数のサイトドメインの投稿文を学習に用いることで、ある程度の汎用性を持つ学習器を作成することができるといえる。しかしながら、今回値が格段に良くなったというわけではなく、また十分な比較を行うことができたとも言えないということから、今後はどのようなデータを混合するのが最も汎用性が高いのか、また混合の割合はどの程度が良いのかなどという点についてもより深く研究する必要があると考える。また今回評価に用いた不満買取センターデータにおいて、ポジティブとそれ以外の感情とのデータ数に大きな差があったため、評価用データとして不満買取センターのポジティブなデータを増やす必要がある。

## 5. まとめと今後の課題

本論文では、様々なサイトドメインへの投稿文からの情報抽出の際の負担軽減を目的として、投稿文を節毎に分割し、CRF を用い各節に対し感情推定を行いその精度を測った。結果として、CRF の学習には複数のサイトドメインのデータを用いることによりある程度汎用性の高い感情推定を行うことができるということが判明した。

今後の課題としては、データの混合に適するデータの調査やデータの混合割合の最適値の追求及びデータ数の増量、また他の機械学習やルールベースによる感情推定との比較などを行っていきたい。

## 謝 辞

本論文の一部は JSPS 科研費 17K00430, 16K07973 及び、私学助成金（大学間連携研究補助金）の助成によるものである。ここに記して謹んで感謝の意を表する。

## 文 献

- [1] J. Lafferty, S. McCallum, F. Pereira, “Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data”, International Conference on Machine

表 6 実験 3-1 で新たに正しくポジティブだと推定できた例

節の内容	実験 2-1 での推定結果	実験 3-1 での推定結果
大変満足しています	ネガティブ	ポジティブ
購入の決め手でした	ネガティブ	ポジティブ
おすすめできます	ニュートラル	ポジティブ
比較的安く	ネガティブ	ポジティブ
なにより気に入ったのが	ニュートラル	ポジティブ

表 7 実験 3-2 で新たに正しくポジティブだと推定できた例

節の内容	実験 1-1 での推定結果	実験 3-2 での推定結果
おいしかったので	ニュートラル	ポジティブ
店舗内は涼しくて快適です	ネガティブ	ポジティブ
先生方も環境も最高だと思って	ニュートラル	ポジティブ
遅くまで商品も揃っているし	ネガティブ	ポジティブ
わかりやすい	ネガティブ	ポジティブ

Learning, 2001.

- [2] S. Rustamov, M. Clements, “Sentence-Level Subjectivity Detection Using Neuro-Fuzzy Models”, Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp.108–114, 2013.
- [3] G. Paltglou, M. Thelwall, “More than Bag-of-Words: Sentence-based Document Representation for Sentiment Analysis”, Proceedings of Recent Advances in Natural Language Processing, pp.546–552, 2013.
- [4] 田村一樹, 吉川大弘, 古橋武, “評点付きレビュー文書を対象としたトピックモデルの構築に関する検討”, 情報処理学会論文誌 Vol.56 No.3, pp.1013–1027, 2015.
- [5] 鈴木泰裕, 高村大也, 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, 2004.
- [6] C. Lin, Y. He, “Joint sentiment/topic model for sentiment analysis”, Proceedings of the 18th ACM conference on Information and knowledge management, pp.375–384, 2009.
- [7] Y. Choi, C. Cardie, “Learning with compositional semantics as structural inference for subsentential sentiment analysis”, Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp.793–801, 2008.
- [8] N. Okazaki, “CRFsuite A fast implementation of Conditional Random Fields (CRFs)”, 2011, <http://www.chokkan.org/software/crfsuite/> [Online; accessed 5-January-2018].
- [9] S. Kurohashi, “日本語構文・格・照応解析システム KNP”, 2016, <http://nlp.ist.i.kyoto-u.ac.jp/?KNP> [Online; accessed 5-January-2018].
- [10] 国立国語研究所 (2004), “分類語彙表増補改訂版データベース (ver.1.0)”, 2015.
- [11] 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.02 pp. 627–637, 2006.