

Linked Open Data を利用した単語極性予想による レビュー感情分析

山元 悠太*†

古崎 晃司†

駒谷 和範†

* 大阪大学 工学部電子情報工学科

† 大阪大学 産業科学研究所

{yamamoto, kozaki}@ei.sanken.osaka-u.ac.jp

komatani@sanken.osaka-u.ac.jp

1 はじめに

文章からの感情分析は、文章から筆者の考えを把握し、同時に筆者が抱く感情を分析することを目的とする技術である [1] [2]。対話システムにおける発話内容、SNS での発言、レビューなど、多様な文章を対象とした感情分析の利用が期待できる。

中でも、様々な製品やサービスに対し、実際に利用した消費者がその感想や評価を述べた文章であるレビューは、感情分析の対象としてニーズが高い。なぜならレビューは消費者にとっては、購入・利用するかどうかの意思決定を行うための情報源となり、商品を提供する企業にとっては、評価を把握し、改善につなげるための情報源となるためである。レビューからの感情分析により、評価を得ている点・批判されている点を自動的に分析するシステムが実現できれば、サービス改善の効率上昇を図ることができる。

本研究ではレビュー中で表されている感情を分析し、良い感情を表す Positive(肯定的;Pos), 批判的な感情を表す Negative(否定的;Neg), そしてその両方に当てはまらない Even(中立), の3種類の感情極性(以下, 極性)に分類する手法を開発する。

多数の単語(及び単語列)に対して、その単語が持つ極性を定義したデータベースを極性辞書と呼ぶ。この極性辞書を最も基本的な情報源とした感情分析が一般的であるが、この方法には課題が残っている。その一つが、極性辞書で定義されていない極性未知単語が文章中に出現することである。それらの極性未知単語の中に Pos や Neg の極性を持つ(と考えられる)単語が存在すると、それを考慮することで文章全体の極性に影響する可能性がある。

2 研究の全体像

2.1 アプローチ

この論文では、極性未知単語の存在に対する解決法を提案する。極性が未知であるという問題を解決するためには、極性未知単語の極性の予想を行う必要がある。本研究では機械学習を使って極性予想を行う。また、以降は本研究に置いて分析対象となるレビュー等の文書を一般化してテキストと呼ぶ。

テキストの感情分析を行う手法は、単語単位で見ていくものが主流である。テキストを単語に分解し、各単語の極性を極性辞書を用いて調べ、その結果を全体について総合的に判断することでテキストの極性を決める(図1)。本研究においても同様に、テキストに含まれる単語の極性をもとに、全体の極性を決めるアプローチを採るが、極性の定義されていない単語が出現した場合にその極性を予想するという処理を追加する。

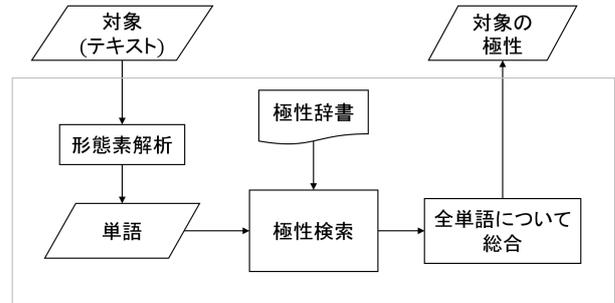


図1: 極性辞書を利用した感情分析

2.2 使用するデータ

本研究で分析対象としたレビューデータは、楽天株式会社と国立情報学研究所が協力して研究者に提供している楽天データセット¹の中にある、以下の2つのレビューデータセットである。

楽天トラベルレビューデータ

多数の旅館・ホテルに対するレビュー情報(本文、ホテル情報、5段階評価など)が約340万件分含まれるデータセット

文単位極性ラベル付きレビューデータ

楽天トラベルレビューデータのうち1000件について、文単位、計5209文について、2人のアノテータによって極性ラベルが付与されたデータセット

文単位の極性ラベルは複数種類(褒め、苦情、要求、ニュートラル、評価なし、保留)があるが、褒めをPos、苦情と要求をNeg、その他をEvenとして3極性に統合した。また、アノテータによって付与結果が異なる場合もあり、この場合も1ラベルへの統合を行った。一方がPos/Neg、もう一方がEvenであればEvenでない方に統合し、一方がPos、もう一方がNegとなっている場合はEvenとした。

また、極性辞書は、日本語評価極性辞書²を利用する。この辞書では、名詞(単語)についてPos・Even・Neg [3]、及び動詞・形容詞・形容動詞といった用言(単語・単語列)についてPos・Neg [4]の極性が定義されている。なお、この辞書は形容動詞も用言としているが、形容動詞は単語としては語幹と語尾に分かれる。本研究では語幹名詞だけを極性を持つ用言として扱うこととする。

¹<https://www.nii.ac.jp/dsc/idr/rakuten/rakuten.html>

²<http://www.cl.ecei.tohoku.ac.jp/index.php?Open%20Resources%2FJapanese%20Sentiment%20Polarity%20Dictionary>

3 単語極性予想

極性未知単語の極性を知るためには、その単語についての意味を表す情報が必要である。本研究では、情報を取得する方法を、単語の文法的分類に応じて2種類に分け、それぞれに対して単語の極性を予想する手法を提案する。

3.1 概要文章を利用した名詞極性予想

名詞に対しては「概要」の文章を利用して極性を予想する。名詞の中でも固有名詞はレビュー中に多く出現するが、その大多数が極性を定義されていない。また、名詞はそれぞれ固有の意味を持ち、似た意味を持つ単語が少ない傾向がある。そこで今回は単語の意味を端的に表した概要文章を利用する。概要文章を取得するリソースとして、自由に利用できる LOD(Linked Open Data) の一つである DBpedia³ を用いる。DBpedia から対象単語の概要文章を取得できた場合は、そこから機械学習用の素性(表1)を抽出する。

表 1: 名詞極性予想の素性

	文の数	
cs	全単語数	
cw	固有名詞・代名詞の数 (関連研究 [6] で使われている)	
csn		
cp	Pos 語数	各単語の極性は極性辞書で検索する
ce	Even 語数	
cn	Neg 語数	
cr	文節内での単語	
	極性反転回数	

これらの素性を基にして SVM(Support Vector Machine) による機械学習を行い、単語の極性予測を行う分類器を構築する。学習プログラムの実装には Weka⁴ を利用する。学習(正解)用データとして、極性辞書で極性が定義されている単語を対象とし、定義された極性を正解とする。単語に対して上記の方法で7つの素性を抽出し、正解の極性と共に入力して名詞用の分類器を構築する。そして極性未知の名詞に対しては、同様に概要を取得・素性を抽出して分類器にかけることで極性を予想する。

3.2 類義語情報を利用した用言極性予想

用言の場合は、似た意味を持つもの(類義語)が比較的多い傾向にあるため、これを利用する。類義語情報を取得するリソースとしては、様々な単語間の関係を取得できる WordNet [5] を用いる。対象単語の類義語(一つとは限らず、多い場合は百個単位で存在する)を WordNet から取得できた場合は、そこから素性(表2)を抽出する。

表 2: 用言極性予想の素性

	全類義語数	
cw		
cp	Pos 語数	各単語の極性は極性辞書で検索する
ce	Even 語数	
cn	Neg 語数	

³Web 上の百科事典である Wikipedia を元に構築された LOD(<http://ja.dbpedia.org/>).

⁴<https://www.cs.waikato.ac.nz/ml/weka/>

概要文章を用いる場合と同様に、極性辞書で定義されている単語を正解データとし、各単語の類義語情報を WordNet から取得する。そこから上記の4つの素性を抽出し、定義された正解の極性と共に入力して用言用の分類器を構築する。そして極性未知の用言についても素性を取得・抽出し、分類器にかけることで極性を予想する。

3.3 単語極性予想のまとめ

提案手法である極性未知単語の極性予想の概略図は図2のようになる。事前に名詞用・用言用の分類器を構築しておき、テキスト処理中に極性未知単語があればその品詞分類に応じて LOD から素性を取得し、分類器によって予想極性を決定する。また、この単語極性予想を導入した場合のテキスト感情分析の概略図は図3のようになる。

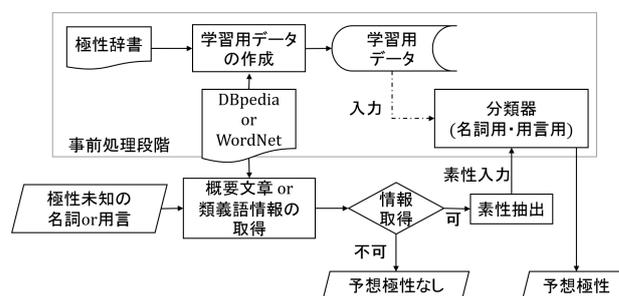


図 2: 極性未知単語の極性予想

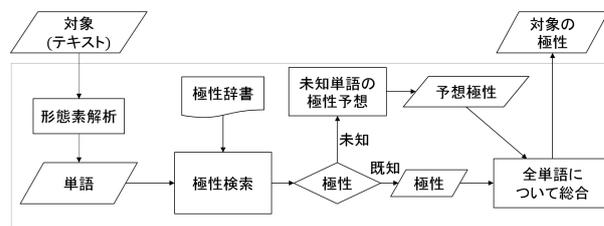


図 3: 単語極性予想を導入した感情分析

単語極性予想を導入しない場合は、この図から極性未知の場合の処理が省かれることになる。

4 評価実験

構築した分類器の性能や、提案手法によるテキスト感情分析の効果を調べるために、評価実験を行った。なお、この実験では、最初の処理として形態素解析器である Mecab⁵ を利用してテキストを単語へ分解した。

4.1 単語極性予想の性能

3章で提案した極性未知単語の極性予想(単語極性予想)の性能を評価する。なお、本研究で用いた極性辞書に含まれる用言自体は5,281項目だが、Evenの正解データを持たない(2.2節参照)。そこで、用言でも Evenの予想を可能とするため、WordNet で類義語情報を取得できる名詞1,066項目を参考用の学習データとして利用した。この結果、名詞極性予想が適用できる3,069項目、用言極性予想ができる6,347項目から素性を取

⁵<http://taku910.github.io/mecab/>

得し、各分類器用の学習データを作成した。これらのデータを用いて分類器ごとに10分割交差検証を行い、分類器の性能評価を行った。その結果を表3に示す。

表3: 単語極性分類器の性能

名詞極性分類器			
評価対象	Precision	Recall	F-Measure
Positive	0.667	0.613	0.639
Even	0.706	0.747	0.726
Negative	0.822	0.808	0.815
加重平均	0.741	0.747	0.747

用言極性分類器			
評価対象	Precision	Recall	F-Measure
Positive	0.790	0.342	0.477
Even	0.406	0.846	0.548
Negative	0.884	0.550	0.678
加重平均	0.714	0.575	0.579

評価指標は,Pos,Even,Neg各極性,及び各項目数に応じて重みをつけた平均(加重平均⁶)についての適合率(Precision;P),再現率(Recall;R),F値(F-Measure;F)を比較することで評価した。これらの指標は,ある極性をTrue,他の2極性をFalseとすることで,3極性それぞれについて算出している。Pos,Even,Negの各行は極性ごとの性能,加重平均は全体的な性能を示している。表3の結果より,極性予想の性能は用言極性分類器と比較して,全体的に名詞極性予想器の方が高いことが分かる。

4.2 単語極性予想による拡張規模

極性辞書のみを用いた場合と比べ,極性予想を取り入れることで極性を判定できる単語数がどれだけ増えるか,を確認する。調査の対象は2.2節で挙げた,文単位での極性ラベルを持つデータのレビューとする。レビュー中に出現する名詞や用言について(同じ単語が出現した場合は重複して数えない),総数,極性辞書に記載されている数,極性予想が可能な数,をそれぞれ調べ,比較した。結果を表4に示す。

表4: 拡張規模実験の結果

	名詞	用言
レビュー中の総数	4,657	917
極性辞書で定義されている	632 (13.6%)	170 (18.5%)
極性未知	4,025(86.4%)	747(81.5%)
極性予想可	695(14.9%)	570(62.2%)
辞書で定義	1337	740
+極性予想可	(28.7%)	(80.7%)

各セルの括弧内の数値は,各列の「レビュー中の総数」に対する割合を表している。また,極性予想可の行は極性未知の単語のうち,提案手法によって極性予想が可能となった極性未知単語の数を表している。この表より,名詞については約2倍,用言については約4倍の数の単語の極性を判定できるようになったことが分かる。

⁶(加重平均) = $\frac{1}{(\text{全項目数})} \sum_{p \in \{\text{Pos, Even, Neg}\}} (\text{極性 } p \text{ の数値}) * (\text{極性 } p \text{ の項目数})$

4.3 レビュー感情分析の性能

提案手法をレビューの感情分析に用いた際の性能を調べるために,単語の極性予想の適用有無による性能差を比較した。

4.3.1 実験方法

実際のレビューに対してベースライン手法と提案手法の2つの方法で感情分析を行い,正解となる極性と一致しているかを確認した。評価に用いるレビューは文単位の極性ラベルを持たない楽天トラベルデータセットからランダムに選択したレビュー3000件(どちらの手法でも同じ3000件を用いる)とした。また,その正解となる極性として,各レビューに付随した総合評価を用いた。この総合評価は1(悪い)から5(良い)の5段階で付けられており,この値が1か2であればNeg,3であればEven,4か5であればPosを正解極性とした。即ち,各レビューはテキストと正解極性ラベルを持つ。この性能比較実験において,比較対象となる手法は以下の2つである。

単語極性予想なし(ベースライン手法)

極性未知単語の極性予想(単語極性予想)を使わず,極性辞書で定義された単語だけを考慮する。

単語極性予想あり(提案手法)

単語極性予想を使って極性を考慮する単語数を増やす。ベースライン手法とは,極性を考慮する単語の範囲だけが異なる。

そして,テキスト全体の極性は,テキスト中のPos,Neg各語数をカウントし,その多数決を取ることで決定した。PosとNegの語数が同じ場合のみEvenとした。分析結果のConfusion Matrixを表5に,各極性,および加重平均についてのPrecision,Recall,F-Measureを表6に示す。

表5: 各手法のレビュー感情分析結果
単語極性予想なし(ベースライン)

		予想極性			計
		Pos	Even	Neg	
正解極性	Pos	1836	457	187	2480
	Even	185	88	85	358
	Neg	54	34	74	162
	計	2075	579	346	3000

単語極性予想あり

		予想極性			計
		Pos	Even	Neg	
正解極性	Pos	2036	285	159	2480
	Even	203	80	75	358
	Neg	50	21	91	162
	計	2289	386	325	3000

表5より,正解極性がPosであるレビューが8割以上であることが分かる。これはデータセットに含まれるレビュー約340万件のうち約250万件がPosであることによるもので,今後,各極性の件数を揃えて同様の実験を行う予定である。

表 6: 各手法の性能評価

	Positive			Even			Negative			加重平均		
	P	R	F	P	R	F	P	R	F	P	R	F
ベースライン手法	0.89	0.74	0.81	0.15	0.25	0.19	0.21	0.46	0.29	0.76	0.67	0.70
提案手法	0.89	0.82	0.85	0.21	0.22	0.22	0.28	0.46	0.29	0.77	0.74	0.75

P:Precision, R:Recall, F:F-measure

表 5.6 より、ベースライン手法に比べて、若干ではあるが提案手法の性能が高い (加重平均の F-Measure が 0.05 だけ高い) ことが分かる。

4.4 考察

単語極性予想の性能を評価する実験では (表 3), 極性未知単語に対する有効な極性予想方法が開発できたと言える。用言極性分類器は名詞極性分類器に比べて性能が低い、これは Even 分類を可能とするための名詞素性を追加したことが一因と考えられる。現状では、分類に用いる素性が少ないため、各極性ごとに重み付きの多数決を取っているような状態にあり、学習方法の改良が必要と思われる。この点については、5 章にて述べる文脈情報の考慮によって改良を行いたいと考えている。

単語極性予想の拡張規模については (表 4), 極性未知単語の量的な問題が改善されるようになったと言える。しかし名詞に関しては依然として極性未知のものが 8 割以上と非常に多い。レビュー中に出現する名詞の総数は非常に多く、今回の方法では、概要文章が取得できない割合が高いことが原因の 1 つである。しかし中には「100 円」「1 階」のように数字を含むものなど、単語レベルで極性を予想することにあまり意味がないと思われるものが含まれる。一方、「チェックイン (チェックインの誤字)」や「ハブラシ (歯ブラシの表記揺れ)」のような、誤字脱字や表記揺れによって正しい予想ができなくなるものも多数あり、こちらについては別の表記揺れ訂正システムなどを適用しなければならない問題である。

最後に、単語極性予想をレビュー感情分析に組み込むことで、若干の性能上昇が認められた (表 6)。レビュー極性予想はどちらの場合でも Pos と判定されるケースが多かったが、この理由としては、「…したほうがいい」「…すればいい」のように、「いい」をはじめとする Pos 単語が批判的な意味合いの文脈でも出現しやすいのではないかと考えられる。また、総合評価が Neg でも、ホテル側をフォローする意味で「…は良かった」という評価がされているものが多く、これも Pos 判定の増加に影響していると考えられる。

5 まとめと今後の課題

本研究では、極性辞書を利用したテキスト感情分析に、極性未知単語の極性予想を取り入れた。提案手法によって、感情分析において極性を考慮できる単語の数が増加し、感情分析自体の性能上昇が認められた。

今後の課題として、文脈による単語極性の変化を考慮したいと考えている。同じ単語であっても、単語が現れる前後の文脈によってその極性が変化する場合がある。例えば、「騒音が小さい」という文の「小さい」は Positive な意味合いとなる。しかし、「部屋が小さい」と

いう文では「小さい」が Negative な意味合いとなる。このようなケースに対応することを目指した極性予想方法を検討している。

また、テキストの感情分析を行う上で、単語単位では感情を十分に考えられないと仮定し、単語よりも大きい単位として 評価情報 を導入する研究がある [7] [8]。評価情報は主に、ある対象と、それに対する評価単語の組で表される。評価表現の導入によってテキストに含まれる感情をより広い視野で考えることが可能となる。現在、この評価情報を単語極性予想と組み合わせた際の性能変化についての調査を進めている。

また、修辞表現による極性変化も考えられる。例えば、逆接・願望・比較・皮肉のような、多彩な表現パターンを持つ修辞表現による単語や意見の極性が反転である。現時点ではこうした表現の検出までは考慮できていないため、今後の重要な課題としていきたい。

謝辞

本研究では、楽天株式会社が国立情報学研究所の協力により研究目的で提供されている「楽天公開データ」を利用した。ここに感謝の意を表す。

参考文献

- [1] B. Pang & L. Lee: Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval, Vol.2, Nos. 1-2, pp. 1-135 (2008).
- [2] 乾 孝司 & 奥村 学: テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol. 13, No. 3, pp.201-241 (2006).
- [3] 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).
- [4] 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).
- [5] F. Bond, T. Baldwin, R. Fothergill & K. Uchimoto: Japanese SemCor: A Sense-tagged Corpus of Japanese, Proceedings of the 6th International Global Wordnet Conference, pp. 56-63, 2012.
- [6] J. Karoui, F. B. Zitoun, V. Moriceau, N. Aussenac-Gilles, & L. H. Belguith: Towards a contextual pragmatic model to detect irony in tweets, 53rd Annual Meeting of the ACL-IJCNLP, Proceedings of the Conference 2, pp. 644-650, 2015.
- [7] T. Nasukawa & J. Yi: Sentiment analysis: Capturing favorability using natural language processing, Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP 2003, pp. 70-77 (2003).
- [8] 飯田 龍, 小林のぞみ, 乾 健太郎, 松本 裕治, 立石 健二 & 福島 俊一: 意見抽出を目的とした機械学習による属性-評価値対同定, 情報処理学会研究報告自然言語処理, Vol. 1, pp. 21-28 (2005).