

# 語義の曖昧性を考慮した極性判定

三浦 康秀 榎 茂之 服部 圭悟 大熊 智子  
富士ゼロックス株式会社 コミュニケーション技術研究所

{yasuhide.miura, sakaki.shigeyuki, keigo.hattori,  
ohkuma.tomoko}@fujixerox.co.jp

## 1 はじめに

近年、インターネット上で個人が自由に情報発信を行うソーシャルメディアが急速に拡大し、様々な事柄に対する個人の意見が活発に発信されるようになった。これら個人の意見の中で関心の高い情報の一つとして、製品・サービス等の評価情報が上げられる。評価情報は利用者が事前に良し悪しを判断するのに利用できるだけでなく、提供者である企業・公共団体等にとっても今後の改善等において有益な情報となる。このため、評価情報を抽出し整理する技術は評判分析等の名称で従来より活発に研究されてきた [5, 12].

評判分析を行うときによく用いられる言語リソースとして、極性もしくは評価表現の辞書が上げられる。例を上げると、ツイートの極性判定をタスクとした SemEval 2013 Task 2 [7] において多くの参加者が MPQA, SentiWordNet [2] 等を利用している。もっとも、評判分析における言語リソースの利用は単純化されていることも多い。例えば、SentiWordNet には同義語、上位語、下位語等の関係が WordNet [4] の情報として含まれているが、極性判定に利用される際には、ポジティブ語・ネガティブ語の出現回数、テキスト中のポジティブ語・ネガティブ語のスコアの合計等の単純な素性としてしばしば利用される。

本稿では、極性判定において言語リソースの語義を語義曖昧性を考慮した上で利用した場合の効果について述べる。具体的には、WordNet の synset<sup>1</sup> および SentiWordNet の synset 毎の極性スコアを極性判定の素性として利用した場合の効果述べる。WordNet の synset をテキスト中の単語に結び付けるには、lemma (見出し語) 毎に定義された多数の synset の曖昧性を解消する必要がある。例えば英語の “like” という単語であれば、WordNet には日本語の “好む” に該当するもの、“似ている” に該当するもの等の 11 の

synset が定義されている。文脈に応じて変わる語義の曖昧性を解消するための様々な手法は長年研究されており [8], 近年では WordNet を対象に語義の曖昧性を解消することも多い。本稿で提案する手法において WordNet の synset を利用するときには、語義曖昧性の解消を行った上で素性として利用する。本稿で強調する提案手法の特徴は以下の 2 点である。

1. WordNet の語義を曖昧性を考慮した上で利用することにより、極性判定の性能が向上する。
2. 既存の WordNet の多言語対応の成果を用いることにより、提案手法を容易に別言語に適用できる。

本稿では、2 章で提案手法である WordNet の語義の曖昧性を考慮した極性判定手法を述べる。3 章では、提案手法の効果を確認するために行った評価実験について述べる。4 章では、実験結果の考察を述べる。5 章では、まとめおよび今後の展望を述べる。

## 2 手法

### 2.1 語義曖昧性解消手法

語義の曖昧性解消手法としては、知識源のグラフ構造を用いた Agirre らの Personalized PageRank に基づく手法 [1] を用いる。Agirre らの手法では以下の式 1 に基づく PageRank ベクトル  $\mathbf{Pr}$  を計算する。

$$\mathbf{Pr} = cM\mathbf{Pr} + (1 - c)\mathbf{v} \quad (1)$$

式中の  $c$  は減衰パラメータであり、0-1 の間の値に設定する。 $M$  はノード間の状態遷移確率行列であり、ノード  $j$  からノード  $i$  への状態遷移確率  $M_{j,i}$  は、リンクが存在すれば  $i$  の外向き次数 (outdegree)  $d_i$  に応じた  $1/d_i$ , 存在しなければ 0 に設定する。 $\mathbf{v}$  は正規化ベクトルであり、通常の PageRank では一律な値を用いるが、Personalized PageRank では曖昧性解消の文脈に出現する語に対応するノードにバイアスが掛かるように設定する。

<sup>1</sup>類義語の集合として表現されている、WordNet における語義識別子。

Agirre らの手法において WordNet[4] を知識源として用いるときには、ノードには synset を用い、WordNet の関係が定義されたノード間にはエッジを設定する。語義の曖昧性解消の実行時には、WordNet 全体より構成されるグラフ構造を  $G$  とすると、まずテキスト中に含まれる語と対応が付く synset より構成される部分グラフ  $G_D \subseteq G$  を抽出する。次に  $G_D$  に対して式 1 を一定回数繰り返し、結果として得られるノードのランクを対応する synset の曖昧性解消スコアとして採用する。

## 2.2 提案手法

提案手法では、教師あり機械学習手法に基づく極性判定を行う。極性情報の教師信号 (ポジティブ, ネガティブ, ニュートラル等) が付与されたテキストから素性を抽出し、それらを Support Vector Machine (SVM) で学習する。テキストから抽出する素性としては以下を用いる。

**UNI** テキスト中に出現する単語の unigram をバイナリ素性として用いる。単純であるが効果の高い素性であることが知られており、3章の評価実験においても UNI のみの設定をベースラインとして用いる。

**POS** 単語の unigram と品詞の組み合わせをバイナリ素性として用いる。例を上げると、形容詞 (JJ) の “like” であれば “like-JJ” になる。Agirre らの手法では語義曖昧性解消のときに品詞情報を用いるため、3章の評価実験において品詞情報単独の効果を確認するために用いる。

**WT** テキスト中に出現する単語を対象に語義曖昧性解消を行った結果を 0-1 の値を取る素性として用いる。図 1 に WT 素性の例を示す。図中の例文であれば、WordNet に lemma が存在する単語は “like” および “video” であり、語義曖昧性解消を実行するとそれらの synset (01824736-v, 01824736-n 等) がスコア付きで得られる。WT 素性では、これら synset とスコアのペアを素性として抽出する。

**WP** WT 素性の素性値に、語義曖昧性解消スコアと SentiWordNet[2] のポジティブ極性スコアを掛け合わせた素性を抽出する。例を上げると、図 1 の “01824736-n” (日本語では “好む” に該当) であれば、SentiWordNet のポジティブ極性スコアは 0.125 であるため、 $0.442313 \times 0.125 = 0.0552891$  の素性値が設定される。

**WN** WT 素性の素性値に、語義曖昧性解消スコアと SentiWordNet のネガティブ極性スコアを掛け合

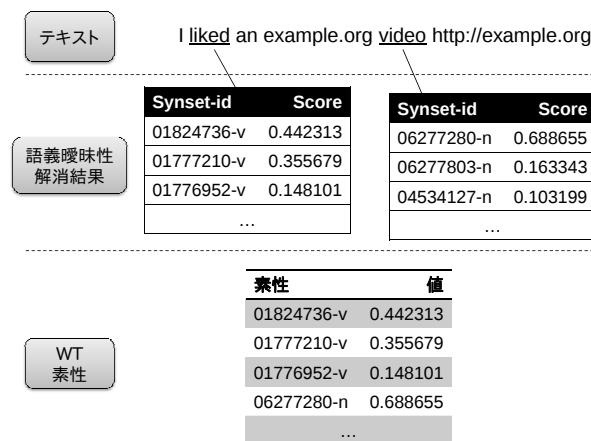


図 1: WN 素性の例。

わせた素性を抽出する。

**1BEST** 単語毎に、語義曖昧性解消結果のスコアが最上位の synset をバイナリ素性として用いる。図 1 の例であれば、“like” の “01824736-v” と “video” の “06277280-n” を抽出する。WT と類似した素性であるが、1BEST では最上位の synset のみを正しい語義と決定し、バイナリ素性として抽出している。

**3BEST** 単語毎に、語義曖昧性解消結果のスコアが上位 3 つまでの synset をバイナリ素性として用いる。

## 3 実験

### 3.1 評価データ

本実験では、4 種類の評価データを対象に提案手法の効果を確認した。以下にそれぞれの概要を述べる。

**SemEval-2013** SemEval-2013 Task2[7] の Sub-task B: Message Polarity Clarification で用いられた 8015 の英語のツイートデータ<sup>2</sup>。データ中の “objective” ラベルはニュートラル極性とみなした。ラベルの内訳は、ポジティブ: 2968, ネガティブ: 1190, ニュートラル: 3857 である。

**MovieReview** Pang ら [10] による 2000 件の映画レビューデータ。ラベルの内訳は、ポジティブ: 1000, ネガティブ: 1000 である。

**JaTweet** 著者らが極性情報を付与した 1204 の日本語ツイートデータ。ラベルの内訳は、ポジティブ: 360, ネガティブ: 328, ニュートラル: 516 である。文献 [6] で述べている極性判定の評価データ

<sup>2</sup>これは Training データと Development データの約 70% に該当する。SemEval-2013 Task2 ではツイート本文のデータは配布されておらず、この数は著者らが取得できた全データである。

から、2人のアノテータが“ポジティブ”，“ネガティブ”，“ニュートラル”のいずれかの極性に合意したツイート抽出した。

**NTCIR-7-MOAT** NTCIR-7 MOAT[11]の日本語セクションのデータから、2人以上のアノテータが合意した6897文。非意見文(not opinionated)はニュートラル極性とみなした。ラベルの内訳は、ポジティブ:111, ネガティブ:333, ニュートラル:6453である。

## 3.2 設定

### 3.2.1 英語

2.2節で述べた素性を抽出するには、テキストから単語の同定、品詞タグ付け、lemmaの取得が必要になる。英語データ(SemEval-2013, MovieReview)の処理には、Stanford POS Tagger 3.2.0<sup>3</sup>を用いた。語義曖昧性解消には、Agirreらの手法[1]を実装したUKB 2.0<sup>4</sup>を用いた。UKBの設定には、辞書にWordNet 3.0、知識源にはWNet30+glossを用い、語義曖昧性アルゴリズムとしてはppr\_w2wをデフォルトパラメータで用いた。SVMの実装にはliblinear-java 1.92<sup>5</sup>をデフォルトパラメータで用いた。

### 3.2.2 日本語

日本語データ(JaTweet, NTCIR-7-MOAT)を対象としたときには、単語の同定、品詞タグ付け、lemmaの取得には、日本語の形態素解析器のKuromoji<sup>6</sup>を辞書をIPADICに設定して用いた。語義曖昧性解消には、UKBの辞書に日本語WordNet 1.1[3]を用い、それ以外の設定は英語と同じものを用いた。SVMの実装および設定には英語と同じものを用いた。

## 3.3 実験1

提案手法をSemEval-2013データに対して適用し、5交差検定でその性能を評価した。7種類の素性の組み合わせについての結果を表1に示す。表中のBaselineはUNI素性のみを用いた設定であり、それ以外の設定ではUNIに“素性”カラムに記された素性を追加した。P, R, F, Aはそれぞれ、適合率、再現率、 $F_1$ 値、accuracyを意味する。WTから3BESTの設定についてはBaselineに対してシャッフル検定[9]を行っており、\*の付いた値は $p \leq 0.10$ 、\*\*の付いた値は $p \leq 0.05$ の有意水準を満たしたことを意味する。最

も高い性能はWT+WP+WNの設定で得られており、Baselineに対して $F_1$ 値で0.9-1.6ポイント、accuracyで約1.1%の有意な性能向上が得られた。

## 3.4 実験2

残りの3つの評価データに対して、5交差検定でBaselineと実験1で最も高い性能が得られたWT+WP+WNの性能を評価した。表2にMovieReview、表3にJaTweet、表4にNTCIR-7-MOATの結果を示す。 $F_1$ 値およびaccuracyについては、3つの評価データ全てについて精度向上が得られた。もっとも、SemEval-2013と同様の $p \leq 0.05$ の有意水準で性能向上が確認できたのは、MovieReviewのポジティブの $F_1$ 値、JaTweetのネガティブの適合率、NTCIR-7-MOATのニュートラルの再現率、 $F_1$ 値、accuracyのみであった。

## 4 考察

評価実験によりWordNetの語義が極性判定の性能向上に寄与することが示された。また、2.2節で述べた語義曖昧性解消結果が品詞情報を含むことについては、3.3節のPOS設定が示すように品詞情報の導入のみでは極性判定の性能向上は確認できなかった。このため、WordNetの語義の導入による性能向上は品詞情報による効果ではないと推測できる。

WordNetの語義の導入手法については、複数の手法を3.3節の実験1で検証した。このとき語義曖昧性の情報をスコア付きで導入した場合(WT, WP, WN)の効果が、上位結果をバイナリ素性で導入した場合(1BEST, 3BEST)を上回っている。この要因としては、語義の曖昧性を文脈から完全に解決するのが難しい点、また語義曖昧性解消手法の精度が不十分である点が上げられる。特に2点目の語義曖昧性解消手法の精度については、WordNetのような大規模な言語リソースを対象とした語義曖昧性解消問題は難しいことが知られており[8]、今回上位結果を導入した素性はノイズを多く含んでいた可能性が考えられる。

日本語WordNetのlemmaを用いた日本語対応については、英語のツイートデータ(SemEval-2013)で得られた有意な性能向上は、日本語のツイートデータ(JaTweet)では確認できなかった。この要因には、データの収集手法の違い、データの規模の違い、日本語WordNetがエラーを含むこと[3]、テキストからのlemmaの抽出が形態素解析器に依存すること等が考えられるが、その詳細は今後検討する必要がある。

<sup>3</sup><http://nlp.stanford.edu/downloads/tagger.shtml>

<sup>4</sup><http://ixa2.si.ehu.es/ukb/>

<sup>5</sup><http://liblinear.bwaldvogel.de/>

<sup>6</sup><http://www.atilika.org/>

素性	ポジティブ			ネガティブ			ニュートラル			A
	P	R	F	P	R	F	P	R	F	
Baseline	0.634	0.655	0.647	0.441	0.399	0.419	0.680	0.687	0.684	0.632
POS	0.631	0.652	0.641	0.431	0.394	0.412	0.677	0.678	0.677	0.626
WT	0.650**	0.660	0.655**	0.454*	0.409	0.431	0.685	0.698**	0.691**	0.641**
WP+WN	0.644**	0.657	0.650*	0.453**	0.413**	0.432**	0.684**	0.692**	0.688**	0.638**
WT+WP+WN	0.650**	0.661	0.656**	0.459**	0.413*	0.435**	0.687*	0.699**	0.693**	0.643**
1BEST	0.646	0.652	0.649	0.438	0.411	0.424	0.680	0.688	0.684	0.633
3BEST	0.646	0.662	0.654*	0.451	0.405	0.427	0.683	0.692	0.687	0.638

表 1: SemEval-2013 を対象とした評価実験の結果.

素性	ポジティブ			ネガティブ			A
	P	R	F	P	R	F	
Baseline	0.851	0.849	0.850	0.849	0.851	0.850	0.850
WT+WP+WN	0.857	0.860*	0.858**	0.859*	0.856	0.858*	0.858*

表 2: MovieReview を対象とした評価実験の結果.

素性	ポジティブ			ネガティブ			ニュートラル			A
	P	R	F	P	R	F	P	R	F	
Baseline	0.591	0.625	0.607	0.497	0.506	0.502	0.566	0.537	0.551	0.555
WT+WP+WN	0.591	0.625	0.607	0.522**	0.506	0.514	0.574	0.562*	0.568*	0.566

表 3: JaTweet を対象とした評価実験の結果.

素性	ポジティブ			ネガティブ			ニュートラル			A
	P	R	F	P	R	F	P	R	F	
Baseline	0.114	0.117	0.116	0.133	0.201	0.160	0.944	0.918	0.931	0.871
WT+WP+WN	0.138	0.144	0.141	0.136	0.186	0.157	0.943	0.925**	0.934**	0.876**

表 4: NTCIR-7-MOAT を対象とした評価実験の結果.

## 5 おわりに

提案手法により, WordNet の語義を曖昧性を考慮しつつ導入することにより,  $F_1$  値で 0.8–1.6 ポイント, accuracy で 0.5–1.1% の有意な性能向上が確認できた. また, 日本語においても日本語 WordNet の lemma 情報を用いることにより, 英語と同様の語義曖昧性を考慮した処理が実現できることを示した.

今後の展望としては, WordNet の同義語, 上位語, 下位語等の関係をより直接的に導入することを検討している. 今回導入した WordNet の語義は語義曖昧性解消結果をそのまま導入する単純な素性であり, より詳細な情報を利用することによりさらなる精度の向上が実現できる可能性がある.

## 参考文献

- [1] E. Agirre and A. Soroa. Personalizing PageRank for word sense disambiguation. In *Proc. of EACL 2009*, pp. 33–41, 2009.
- [2] S. Baccianella, A. Esuli, and F. Sebastiani. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proc. of LREC 2010*, pp. 2200–2204, 2010.
- [3] F. Bond, H. Isahara, S. Fujita, K. Uchimoto, T. Kuribayashi, and K. Kanzaki. Enhancing the Japanese WordNet. In *Proc. of ALR-7*, pp. 1–8, 2009.
- [4] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
- [5] B. Liu. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
- [6] Y. Miura, K. Hattori, T. Ohkuma, and H. Masuichi. Topic modeling with sentiment clues and relaxed labeling schema. In *Proc. of SAAIP 2013*, pp. 6–14, 2013.
- [7] P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and T. Wilson. SemEval-2013 task 2: Sentiment analysis in Twitter. In *Proc. of SemEval-2013*, pp. 312–320, 2013.
- [8] R. Navigli. Word sense disambiguation: A survey. *ACM Computing Surveys*, Vol. 4(2), No. 10, pp. 1–69, 2009.
- [9] E. Noreen. *Computer-Intensive Methods for Testing Hypotheses: An Introduction*. John Wiley and Sons, Inc., 1989.
- [10] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proc. of ACL 2004*, pp. 271–278, 2004.
- [11] Y. Seki, D. Evans, L. Ku, L. Sun, H. Chen, and N. Kando. Overview of multilingual opinion analysis task at NTCIR-7. In *Proc. of NTCIR-7*, pp. 185–203, 2008.
- [12] 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(3), pp. 201–241, 2006.