

単語の極性を埋め込んだ分散表現

中村 拓[†] 田 然[†] 乾 健太郎^{†‡}
[†]東北大学 [‡]理化学研究所 AIP センター

{tnakamura, tianran, inui}@ecei.tohoku.ac.jp

1 はじめに

周辺文脈を予測する単語分散表現は、類義性やアナロジーの判定において有効性を示しているが、単語の対義関係や感情極性を分散表現から読み取することは難しいとされている [10, 12]. これは、単語が対義や極性が反転している時でも、似たような文脈で出現することが多いからであるが、一方で対義性や極性の反転によって生じる共起語の変化も確かに存在する [3, 9]. よって、辞書情報などの外部知識を教師信号とし、周辺文脈を利用して対義や極性を汎化して学習するのは魅力的な研究課題である。

本研究では、単語の極性を線形関数で分離できるような分散表現モデルを2つ提案し、単語の極性分類と対義語識別に対する汎化性能を評価した。結果、(i) 辞書情報に応じて単語ベクトルを初期化するだけで高い汎化性能が得られること、(ii) 通常の単語ベクトルを用いて対義語を識別するのは極めて難しいが、辞書情報を取り入れて同時学習することで、ある程度対義性を学習できることを発見した。

2 関連研究

単語の類義性はベクトルのコサイン距離で測れるが、対義語も距離の近い所に現れてくることが大きな問題とされてきた [8]. この問題に直接取り組む手法として、辞書情報を利用し同義語を近づけ、対義語を遠ざける学習がよく研究されたが [2, 10, 11], 距離を直接変える手法は単語ベクトルの性質を大きく変えてしまう。実際、訓練時に周辺文脈を予測する項のウェイトより距離を変える項のウェイトを遥かに大きく設定する必要があり [10], 学習されたベクトルはもはや周辺文脈の情報を忠実に取り入れたとは言えない。このように訓練されたベクトルが果たして対義性を汎化できたことになるかについては疑問視する見方もあり、実際、評価時に使われたデータの90%以上は「対義語の同義語も対義語」という簡単な推論で訓練データから導き出せるという指摘もある [1]. このような背景から、本研究では距離を変える学習ではなく、線形関数によって極性を識別するような学習を検証した (図1). 提案手

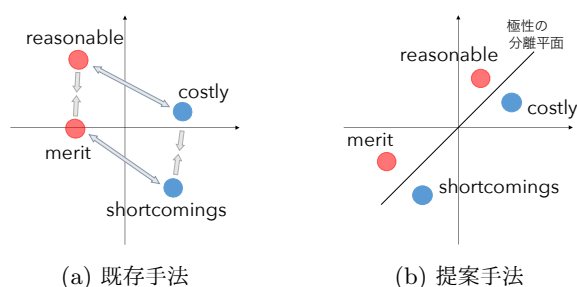


図1: 距離を変える (a) と線形関数で分離する (b)

法では、線形関数で極性を分離しながら意味の類似する単語のベクトルを近づけることができるので、これは既存手法よりも緩い制約である。

一方、対義語と同じく対照的な意味を捉える感情極性に対しては、「ポジティブ・ネガティブ」という自然な軸が存在しているため、「線形関数で分離する」という考えは古くからあり、感情極性を識別する分散表現にも成功例が多い [14, 5, 15]. 本研究は、まず感情極性の埋め込みに提案手法を検証し、そこで得られた経験に対義語判定に応用し、感情極性の場合と比較する。感情極性の埋め込みについては、Voら [16] が本研究と似た提案をしているが、Voらの提案は感情極性だけを学習した2次元の短いベクトルであるに対し、我々の提案は極性情報を分散的に単語ベクトルの全ての次元に保存し、線形関数をもってそれを取り出すものである。これは、単語ベクトルの部分空間に極性や辞書の情報を写像する [6, 13] などの研究とも視点が異なる。

3 提案モデル

いくつかの単語対に対しては、極性辞書によって単語の極性が定義されている。本研究では極性辞書による極性既知の単語をシードと呼び、極性の教師情報として用いる。単語の極性を分散表現学習に取り入れる手法として、(i) シード単語の初期化を極性情報に応じて行う、(ii) シードの極性を分離するための同時学習を行う2つのモデルを提案する。対義語のシードについては、ポジティブ・ネガティブのような軸が存在しないため、対義語辞書の単語に対しランダムにラベルを振り分ける。

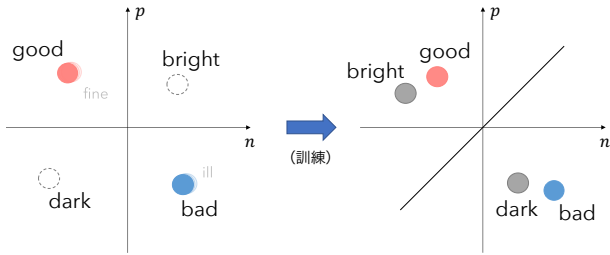


図 2: 初期化モデル

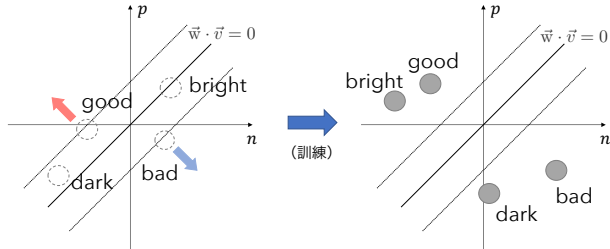


図 3: 同時学習モデル

3.1 初期化モデル

シード単語の極性情報を単語ベクトルの初期化で取り入れるこの提案は、Vo らの研究 [16] に倣い各単語のベクトルをネガティブな成分 \mathbf{n} とポジティブな成分 \mathbf{p} に分け、それぞれ前半と後半の次元に対応すると仮定する。モデルの学習を通して、単語の極性がこれらの次元に分散して保存されることを期待する。極性のネガティブなシードに対しては $[\mathbf{n} : \mathbf{p}] = [1, \dots, 1, -1, \dots, -1]$ 、ポジティブなシードに対しては $[\mathbf{n} : \mathbf{p}] = [-1, \dots, -1, 1, \dots, 1]$ となるように単語ベクトル学習時の初期化を行う。各ベクトルは正規化し、シード以外の単語についてはランダムに初期化する。

3.2 同時学習モデル

周辺文脈との共起から単語分散表現を学習する代表的な手法の一つとして Skip-gram with Negative Sampling (SGNS) [7] がある。提案手法では、文脈共起に基づく学習に、SGNS モデルに L2 正則化項を加えた式 1 を最小化すべき目的関数とする分散表現モデルを用いる。本稿では以降 SGNS モデルと表記する。

$$-\sum_{t \in V} \sum_{c \in C_t} \left(\ln \sigma(\vec{v}_t \cdot \vec{v}_c) + \sum_{k=1}^K \ln \sigma(-\vec{v}_t \cdot \vec{v}_c) \right) - \lambda \|\vec{v}_t\|^2 - \lambda \|\vec{v}_c\|^2 \quad (1)$$

ここで、 V は単語集合、 C_t は単語 t の文脈、 \vec{v}_t は単語 t のベクトル、 \vec{v}_c は文脈に出現する単語 c のベクトル (文脈ベクトル)、 K は負例サンプリング数、 \vec{v}_t は学習データの単語ユニグラム分布からランダムにサンプルした擬似負例単語のベクトル、 λ は正則化項のパラメータである。

シード単語の極性を識別するための目的関数として、

SVM のヒンジ損失関数 (式 2) を用いる。

$$\frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i(\vec{w} \cdot \vec{v}_i - b)) \quad (2)$$

ここで、 n はシード単語数、 y_i はシード単語の極性クラスラベル (ポジティブ 1, ネガティブ -1)、 \vec{w} は係数ベクトル、 \vec{v}_i は単語または文脈ベクトル、 b はバイアス項である。シード単語の極性分離と、周辺文脈共起からの学習を同時に行うため、係数 \vec{w} を固定して、単語または文脈ベクトル \vec{v}_i をパラメータとみる。係数 \vec{w} は前半次元を -1, 後半次元を 1 として正規化したベクトルを用いる。 d 次元の単語ベクトルを学習するとき、係数ベクトルは

$$\vec{w} = \frac{\overbrace{[-1, \dots, -1]}^{d/2}, \overbrace{[1, \dots, 1]}^{d/2}}{\|\vec{w}\|^2}$$

となる。バイアス $b = 0$ とし、全体として式 1 と式 2 の和 (3 式) を目的関数として最小化する。

$$-\sum_{t \in V} \sum_{c \in C_t} \left(\ln \sigma(\vec{v}_t \cdot \vec{v}_c) + \sum_{k=1}^K \ln \sigma(-\vec{v}_t \cdot \vec{v}_c) \right) - \lambda \|\vec{v}_t\|^2 - \lambda \|\vec{v}_c\|^2 + \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i \vec{w} \cdot \vec{v}_i) \quad (3)$$

SGNS の確率的勾配降下法 (SGD) による単語ベクトル更新と同時に、式 2 を最小化するための更新を行う。初期化は全ての単語についてランダムに行う。

4 実験 1: 感情極性分類

各モデルで学習した単語ベクトルを用い、単語の感情極性二値分類 (ポジティブ・ネガティブ) の正解率によって評価を行なった。

4.1 実験設定

周辺文脈との共起情報学習データとして、English Wikipedia を用いた。出現単語数は約 16 億語、出現頻度 150 以上の単語を語彙とし語彙数は約 15 万語である。

感情極性の教師データ (シード) として、MPQA[17] からタイプが strongly subjective で、ポジティブまたはネガティブいずれかのラベルが付与された単語のうち、Wikipedia から作成した語彙内の 3499 語 (ポジティブ 1255 語、ネガティブ 2244 語) を用いた。

評価データとして、Opinion Lexicon[4] から、ポジティブまたはネガティブのどちらかで、シードに用いた単語と重複しない 2281 語 (ポジティブ 763 語、ネガティブ 1518 語)*1 を用いた。

*1 多数クラス (ネガティブ) が 66.54% となった。

表 1: 識別関数による単語感情極性の分類正解率

		全体	シード近傍
初期化モデル	target	75.31	89.46 (2012 語)
	context	87.59	-
同時学習モデル	target	87.29	90.19 (1947 語)
	context	88.20	-

表 2: SGNS のベクトルを特徴量とした SVM による単語感情極性の分類正解率

		open test	closed test
SVM-SGNS	target	91.10	99.54
	context	91.36	99.51

各モデルについて、ベクトルの次元数は 200 次元、共起をカウントする文脈の窓幅は前後 3 語、負例サンプル*2数は 3、正則化項のパラメータ $\lambda = 1/16384$ とした。SGD のエポック数は 1 とした。

4.2 評価方法

単語極性の識別関数を、学習した各ベクトル $\vec{v}_i = (v_{i1}, \dots, v_{id})$ のポジティブな成分の和とネガティブな成分の和の差と定義する。単語または文脈ベクトル \vec{v}_i に対する識別関数は

$$\sum_{j=d/2}^d v_{ij} - \sum_{j=1}^{d/2} v_{ij}$$

となり、式 2, 3 における $\vec{w} \cdot \vec{v}_i$ に対応する。識別関数の値が正の時ポジティブ、負の時ネガティブを表す。単語 (target) ベクトル、文脈 (context) ベクトルをそれぞれ使い、各々の識別関数によって評価データで単語の感情極性分類を行う。

制約の効果の範囲を確かめるため、シード単語の近傍 (いずれかのシード単語とのコサイン類似度が 0.5 より大きい) 語のみについても分類を行なった。

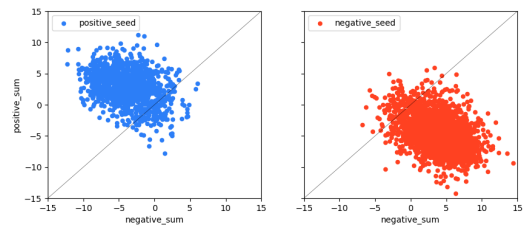
ベースラインとして、SGNS モデルで学習したベクトルの各次元値を特徴量に、シード単語の極性ラベルを教師として分類器 (SVM) の訓練を行い、評価データで分類 (open test) を行う。SVM は RBF カーネルを用い、単語、文脈ベクトルともに正規化する。分類器が感情極性の訓練データの情報を十分に捉えていることを確認するため、訓練データにおいても評価 (closed test) を行う。

4.3 結果・分析

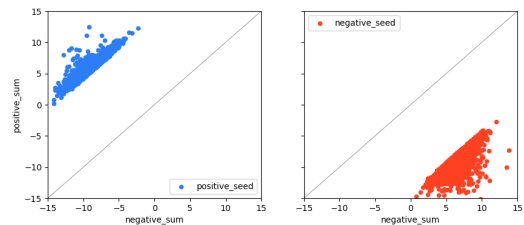
感情極性分類の結果を表 1, 表 2 に示す。表 1 のシード近傍については、評価データ全体で評価した場合と比較して変化があった単語 (target) ベクトルの識別関数を用いた結果のみ示す。

SGNS モデルの分散表現を特徴量に、シード単語を教師データとして訓練した SVM が最も高い識別性能

*2 サンプリングには単語のユニグラム確率の 0.75 乗を用いた。



(a) 初期化モデル



(b) 同時学習モデル

図 4: シード文脈ベクトルに対する識別関数値の分布

を示した。初期化モデルはシード近傍において、評価データ全体で分類を行なった場合と比べ正解率が高く、シード単語に近い単語については感情極性の特徴を捉えていると考えられる。同時学習モデルは単語ベクトル、文脈ベクトルの識別関数を用いた場合ともに評価データ全体においても SVM に匹敵する分類正解率となり、初期化モデルに比べ極性の汎化能力が高いと推測される。また、両提案モデルの結果に共通して、文脈ベクトルの識別関数の方が単語ベクトルの識別関数に比べて高い識別性能を示し、文脈ベクトルに極性の違いが強く現れる傾向がみられた。

単語の極性を線形分離するように保持すると仮定した成分が各提案モデルでどう学習されているか確かめるため、極性の違いをより保持できると考えられる文脈ベクトルについて、シードのポジティブ成分の和、ネガティブ成分の和をそれぞれ縦軸、横軸として図 4 のようにプロットした。各図の左側 (青) がポジティブ、右側 (赤) がネガティブなシードを表し、図中の斜線は識別関数が 0 となる、すなわち式 2, 3 における $\vec{w} \cdot \vec{v}_i$ が $\vec{w} \cdot \vec{v}_i = 0$ となる境界を表す。極性のクラスラベルはポジティブを 1, ネガティブを -1 としていることから、境界の上側がポジティブ、下側がネガティブな領域とみることができる。シードの極性に応じて初期化を行うだけでも、ある程度極性を分離でき、同時学習モデルでは、さらに識別面に関してポジティブな成分とネガティブな成分がよく分離されるように学習できていると言える。

5 実験 2: 対義語判定

感情極性以外の単語の極性分離にモデルが応用可能か調査するため予備的な対義語判定実験を行い、対義

表 3: 対義語ペア判定の再現率

	open test	closed test
初期化モデル	52.85	72.41
同時学習モデル	44.11	91.72
SVM-SGNS	1.02	48.78

語の正例ペアを学習した分散表現を用いて対義語と判断できた割合を再現率として評価した。

5.1 実験設定

周辺文脈との共起情報学習には、4.1 章と同じ Wikipedia データを用いた。対義語の教師データには、WordNet 及び Roget's Thesaurus から作成された公開データ*3を用い、評価データを除く語彙内の 11496 語について同一ラベルに対義語が入らないようにランダムに 1, -1 のラベルを付与し (1:5897 語, -1:5599 語) シードとした。

対義語ペア判定の評価には、上記公開データで分割されている評価データからペアの両単語が共起情報の学習データ語彙に含まれ、シードに含まれない 492 の対義語ペアを用いた。

5.2 評価方法

初期化モデルおよび同時学習モデルで学習したベクトルの識別関数を用いて、与えられた単語ペアが対義関係にあるかの判定を行う。感情極性分類の結果 (4.3 章) から、識別性能が高かった文脈ベクトルの識別関数を用い、ベースラインとして SGNS で学習した単語ベクトルを特徴量、シード単語のラベルを教師に訓練した SVM (RBF カーネル) を用いる。

評価データの対義語ペアのうち、ペアの各単語で識別関数の符号 (ベースラインの SVM では予測ラベルの符号) が異なるペアの数を正解数、評価データの対義語ペア全体に占める割合を再現率として評価する (open test)。訓練データを作成した対義語ペアにおいても、同様に評価 (closed test) を行う。

5.3 結果・分析

対義語ペア判定の結果を表 3 に示す。対義語ペアの判定に関しては、ベースラインに設定した SVM ではほとんど正しく判定できていない。初期化モデルで学習した単語ベクトルの識別関数を用いた場合の再現率が最も高く、同時学習モデルは評価データにおいては初期化モデルに次ぐ結果となり、訓練データに用いた対義語ペアにおける評価では初期化モデルよりも判定性能が高かった。シード情報を用いた初期化だけでも単語の対義性に関する幾許かの知識を分散表現に埋め込める可能性を示唆していると考えられる。

*3 <https://github.com/antonyms/AntonymPipeline/tree/master/data>

6 おわりに

周辺単語の文脈と、外部知識として辞書情報を用い、極性を線形関数で分離できるように埋め込んだ単語分散表現を学習する 2 つのモデルを提案した。

感情極性分類による評価を通して、既知の単語極性を用いた初期化を行うモデルが予想以上に高い識別性能を持ち、極性分離の学習を分散表現の学習と同時にを行うモデルは初期化モデルに比べ極性汎化性能を持つことがわかった。また、対義語ペアの判定評価から、対義関係は感情極性の分離よりも難しく、手法改善の余地が大きいことが確認された。

今後、対義語も区別できるような分散表現の実現を目指し、シードラベルの付与方法見直しや、同時学習用に追加の次元を設けるアプローチなどを検討したい。

謝辞

本研究は JSPS 科研費 15H01702, NTT コミュニケーション科学基礎研究所の支援を受けたものである。

参考文献

- [1] Z. CHEN, W. LIN, Q. CHEN, X. CHEN, S. WEI, H. JIANG, AND X. ZHU, *Revisiting Word Embedding for Contrasting Meaning*, in ACL, 2015.
- [2] M. FARUQUI, J. DODGE, S. K. JAUHAR, C. DYER, E. HOVY, AND N. A. SMITH, *Retrofitting Word Vectors to Semantic Lexicons*, in NAACL, 2015.
- [3] S. FENG, J. S. KANG, P. KUZNETSOVA, AND Y. CHOI, *Connotation Lexicon: A Dash of Sentiment Beneath the Surface Meaning*, in ACL, 2013.
- [4] M. HU AND B. LIU, *Mining and Summarizing Customer Reviews*, in KDD '04, 2004.
- [5] M. IYYER, V. MANJUNATHA, J. BOYD-GRABER, AND H. DAUMÉ III, *Deep Unordered Composition Rivals Syntactic Methods for Text Classification*, in ACL, 2015.
- [6] A. KUMAR AND J. ARAKI, *Incorporating Relational Knowledge into Word Representations using Subspace Regularization*, in ACL, 2016.
- [7] T. MIKOLOV, I. SUTSKEVER, K. CHEN, G. CORRADO, AND J. DEAN, *Distributed representations of words and phrases and their compositionality*, in NIPS, 2013.
- [8] S. M. MOHAMMAD, B. J. DORR, G. HIRST, AND P. D. TURNEY, *Computing Lexical Contrast*, Computational Linguistics, 39 (2013).
- [9] K. A. NGUYEN, S. SCHULTE IM WALDE, AND N. T. VU, *Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction*, in ACL, 2016.
- [10] M. ONO, M. MIWA, AND Y. SASAKI, *Word Embedding-based Antonym Detection using Thesauri and Distributional Information*, NAACL, (2015).
- [11] N. T. PHAM, A. LAZARIDOU, AND M. BARONI, *A Multitask Objective to Inject Lexical Contrast into Distributional Semantics*, in ACL, 2015.
- [12] S. ROTHE AND H. SCHÜTZE, *AutoExtend: Extending Word Embeddings to Embeddings for Synsets and Lexemes*, in ACL, 2015.
- [13] S. ROTHE AND H. SCHÜTZE, *Word Embedding Calculus in Meaningful Ultradense Subspaces*, in ACL, 2016.
- [14] R. SOCHER, B. HUVAL, C. D. MANNING, AND A. Y. NG, *Semantic Compositionality through Recursive Matrix-Vector Spaces*, in EMNLP, 2012.
- [15] D. TANG, F. WEI, B. QIN, N. YANG, T. LIU, AND M. ZHOU, *Sentiment Embeddings with Applications to Sentiment Analysis*, IEEE Transactions on Knowledge and Data Engineering, (2016).
- [16] D. T. VO AND Y. ZHANG, *Don't Count, Predict! An Automatic Approach to Learning Sentiment Lexicons for Short Text*, in ACL, 2016.
- [17] T. WILSON, J. WIEBE, AND P. HOFFMANN, *Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis*, in HLT-EMNLP, 2005.